ISSN 2466-135X
Vol.9, No.1

The 9th International Conference on

BIG DATA APPLICATIONS AND SERVICES (BIGDAS2021)

PROCEEDING

November 25-27, 2021
Jeju Island, Korea

Hosted by
Korea Big Data Service Society

THE KOREA
BIG DATA SERVICE SOCIETY
한국빅데이터서비스학회

CBNU
충북대학교

충북대학교
빅데이터연구소

Table of Contents

A Stride-based Action Analysis in the Pedestrian Dead Reckoning: Application to a Fireman Action

Seung-Ryol Maeng[1,1], Jeong-Bong You[2], Hyun-San Park[2], Sik-Yung Kim[2], Dae-Soo Kang[3],

[1] Div. of Computer Science, Kongju National University, S. Korea
[2] Div. of Electrical, Electronics, and Control Engineering, Kongju National University, S. Korea
[3] Div. of Information & Telecommunication Engineering, Kongju National University, S. Korea

Abstract. In the viewpoint of ethology, human actions are influenced from pychological or external causes. Particularly, length and tempo of stride are also changed according to environmantal situation. Presented in this paper is a stride-based action analysis. Walking steps captured with MEMS sensors are divided into a series of groups with the simillar length of stride. A walking step is included in the next group when the difference between its stride length and running average of the current group exceeds tollerence. It was simulated and plotted for verification. As some examples. we made ethological criteria for interpretation of human action and applied to fireman actions for post-analysis of a fire situation. Conclusively, the point source of fire and dangerous locations could be post-analyzed through our approach.

Keywords: dead reckoning, action analysis, segmentation of steps, big data, data visualization

1 Introduction

The dead reckoning is a navigation method in whch new position is estimated using current position and velocity. Recently, this is widely used for an indoor positioning of pedestrians with mobile devices. Under the premise of walking, the pedestrian dead reckoning(PDR) estimates the position of a pedestrian using the moving informations such as stride, velocity, and posture. Finally, the position is used for position-based services [1,2].

Principle of a PDR, in which sensors are attached on feet, is to detect steps from the sensed data and estimate the direction and length of strides. A stride means the distance betweens human steps. Stride and time provide the base for calculating the

1 Please note that the LNCS Editorial assumes that all authors have used the western naming convention, with given names preceding surnames. This determines the structure of the names in the running heads and the author index.

walking speed of a pedestrian. By integrating speed with azimuth, velocity and position can be calculated. MEMS devices, which are small, low-power, and low-price, are usually used to get steps and azimuths of a pedestrian.[2]

Since MEMS based PDR is a relative positioning, it tends to accumulate errors [1,4]. Thus, there have been many researches on error reduction. Nevertheless, PDR is useful in that infrastructures are unnecessary. Base of this technology is to detect human action using MEMS seonsors. If other information can be extracted from sensed information, the PDR becomes more useful.

In the viewpoint of ethology, human actions are influenced from pychological or external causes. Particularly, length and tempo of stride are also changed according to environmantal situation. We propose a method for anayzing human actions using stride. Walking steps are divided into a series of groups with the simillar length of stride and each group is interpreted with empirical ethology. This scheme will be applied to a fireman for post analysis of fire situation.

Next section illustrates the step detection. Our method is given in Section 3 and then experimental results in Section 4. Finally, we discuss future work in Section 5.

2 Step Detection from MEMS Sensors Data

A PDR with MEMS sensors is done with the four stages such as detection of a step, estimation of stride, calculation of azimuth, and calculation of speed in sequence. Details in each stage may be different according to where sensors are attached.[2] Attaching MEMS sensors to feet has an advantage in detecting steps.

Since our work is to analyze intention and situation of a pedestrian using stride, the step detection among four stages is most important. In detecting a step, digital output of an accelerometer is used. For detection of a step uses the fact that when setting foot on ground, the changing rate of accelerometer's output is regularly and repeatedly close to zero.

In fact, output of an accelerometer should include high frequency component because of ground slope, accelerometer angle, and sensor bias. To remove these noises and get the changing rate of acceleration with zero, the output signal is filtered with a low-pass filter and then a step can be detected in flat area. Then detected step provide a clue for calculating velocity through ingration of accelerometer output. For more details, refer to [2]. In their paper, they introduced walking frequency and output variance of an accelerometer and mentioned the relation of them to strides.

3 Step Segmentation and Its Visualization

We propose a three stages approach for interpretation of human actions as Figure 1. A s mentioned previously, the second and third stages are naturally solved because our work is based on stride as a metric for analysis and human step actions are empirically and ethologically interpreted. Now the first stage, segmentation of stage, remains only and depends on a property such as stride. Human steps has the property to be smoothl y connected although their purposes are different. Thus, to find the connection point b

etween segmentations is not easy.

A clip of action is defined as a regular sample to be extracted from feet action and r epresented with a tuple such as <time, position, direction>. At that time, they are store d a list like L = {v_1, v_2, v_3...v_m}. The list was sorted according to time, a member v_i is way ahead of v_{i+1}. If the clips of vi and v_{i+1} are actions for different purposes, they are fairly included in other segments. Now it is time to discuss a way to disconnect two cl ips into different segment.

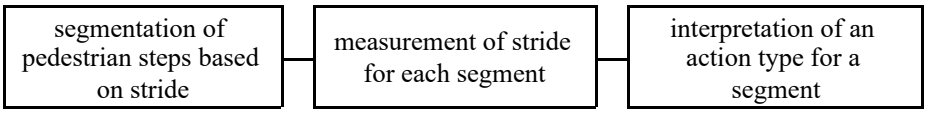


Figure 1. Stages of stride-based analysis

```
L = {v1, v2… vm} # list of strides for a fireman
Sk = {vj, vj+1,…vk}, for k<= m # the k−th sublist of L
Ak : the running average of walking speed for the sublist k
for vi, vi+1 ∈ L
   D = |vi – vi+1|
   if |D − Ak| <= e then
      Sk <− vi+1
   else
      Sk+1 <− vi+1
```

Figure 2. Segmentation algorithm of motion clips

Let A_k be the running average for the segment including the clip v_i and D be the sp eed difference of the clip v_i and v_{i+1}. If $|D – v_{i+1}|$ is bigger than tolerance e there is a str ong possibility that the clip v_{i+1} is included in a next segment. In short, the segmentati on process is given as Figure 2.

This way can be thought as a contrary concept of a motion graph to be used in the f ield of computer animation. A motion graph is a technique to generate a new motion c lip using a captured motion [3]. A pair of two independent motion clips with the most similarity is selected and linked together with a directed graph. A criterion of the simil arity is the distance between two motion clips.

4 Simulation Results and Its Application to Fireman Actions

Figure. 3 shows simulation results of stride segmentation. In simulation, about 500 strides was generated and classified into five groups by the similarity of stride. Figure 3(a) plots all strides generated for simulation time and Figure 3(b) does the average of strides by group, in which area of circles implies corresponding sample size. Conclusively, behavior of each group can be interpreted with different meanings such as running and walking.

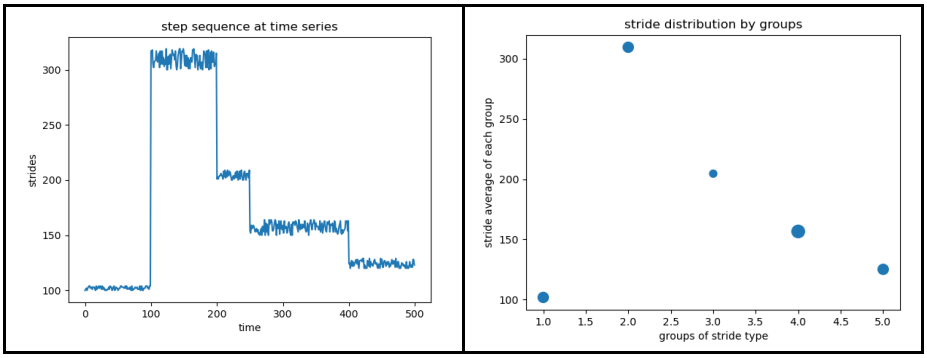


Figure 3. Simulation results of stride segmentation: (a) strides plotting (b) distribution of avera ge for each group

Now, let's consider the interpretation of walking actions. Naturally, urgent situation make a man run with big steps. Stride-based analysis of a pedestrian action depends on human behavior. For example, constant and periodical strides are interpreted as the walking or running. Table 1 shows the examples for some actions and its meanings. Without doubt, the length of stride is a basic information for calculating the moving distance for a given time.

Table 1. Stride types and their meanings

Actions	Meanings
constant and periodical strides	walking and running
non-constant strides	wondering
constant and slow strides	observing
non-constant and small strides	hurrying
stop or small striding	extinguishing

In the event of fire, the post analysis is important in terms of finding the point sour ce of the fire. At the same time, safety of a fireman in the fireplace should take preced ence over anything. If we ethologically define stride types for a fireman, where a fire man stayed, what he did, and when he ran can be easily analyzed through Figure 3.

5 Conclusions

In this work, an approach to step segmentation was proposed and its simulation results were shown. Human actions are influenced from pychological or external

causes. Dangerous situations make human walk faster. Although the PDR is for estimating the position, stride can be used to analyzed human actions. Main contribution of this paper is to segment walking steps of human into a series of groups with the simillar size and to ethologically interpret the meaning of each group. For an example, stride-based analysis can be used for post analysis of a fireman activity. Although our work has shown that stride-based approach is a method for analysis of human activity, more researches are needed for detailed and complicated analysis of human actions.

References

1. Seong Yun Cho, Jae Hong Lee, and Chan Gook Park: Cubature Point-based Location Domain Fusion of PDRs Mounted on Both Feet for Indoor Positioning of a Firefighter, J. of Institute of Control, Robotics and Systems, 26(8):630-636 (2020)
2. Seong Yun Cho, Seung Hyuck Shin, and Chan Gook Park: MEMS Sensor based Context-aware Technology of Walking Dynamics and Integrated Navigation for Seamless Pedestrian Navigation, Telecommunications Review, 19(1), 148-164, (2009)
3. Lucas Kovar, Michael Gleicher, and Frederic Pighin: Motion Graph, In: Proceedings of SIGGRAPH ‘02 (2002)
4. Jeongpyo Lee, Kyung-Eun Park, and Youngok Kim: A Step Detection Scheme for Positioning Technology based on Pedestrian Dead Reckoning, In: Proceedings of KAIS ‘21 Conference:1050-1052 (2021)

Pose Evaluation for Home Training Application Using Deep Learning

Yujung Shin, Seobin Kim, Kyung Chang Jeong, Aziz Nasridinov

Department of Computer Science, Chungbuk National University, Cheongju, South Korea

Abstract. With the popularization of remote offices and online education systems, people spend more time on smart devices. The incidence of musculoskeletal diseases caused by postural imbalance has risen. Therefore, the demand for Pilate exercises for joint and rehabilitation is rising. The existing exercise motion recognition system has an additional hardware purchase cost or does not have a real-time feedback function on an exercise posture. Therefore, in this study, we propose a smartphone-based system that can learn exercise patterns, then return real-time posture feedback. In this study, we tested the side jack and the wide squat in the Pilates movement. The accuracy of the side jack and wide squat is 92.4% and 91% accuracy, respectively.

Keywords: pose evaluation, deep learning, real-time feedback

1 Introduction

As the novel coronavirus that occurred in December 2019 spread worldwide, the World Health Organization declared a global pandemic of COVID-19 and released some national measures to slow the spread of the virus [1]. As the policy of keeping social distancing released, group exercise became impossible because the number of people using the gym was limited, and the outside activities decreased due to fear of infection. The incidence of musculoskeletal diseases is increasing because of people's inappropriate posture when using smart devices.

Pilate's exercise is effective for reducing musculoskeletal problems caused by inappropriate posture. Because it helps stabilize the body by correcting body alignment through the development of core muscles and deep respiratory muscles, which are the center of the body [2]. However, if we perform Pilate's exercise with an incorrect posture, we may not feel the effect and have a risk of injury. Therefore, the proposed system can help us exercise correctly and effectively according to real-time feedback from experts.

Although TV and Kinect-based applications provide the service, the additional hardware purchase cost is incurred, and these applications cannot provide real-time feedback. For overcoming these limitations, we propose a system that can learn Pilate's exercises and return real-time posture correction for movements using only a smartphone without purchasing additional devices. The motion recognition system of this paper is an algorithm that can compare motions with high accuracy even if body

dimensions are different between expert data and user data. As a result of experimenting on side jack and wide squat movements, the side jack showed 92.4% accuracy, and the wide squat showed 91% accuracy.

2 Related Study

2.1 Effects of Pilates on Posture Correction

Park [3] studied the effect of posture correction exercise on turtleneck syndrome. For 20 patients with turtleneck syndrome, posture correction exercise was given to 10 people in the experimental group and correct posture education to 10 people in the control group for 12 weeks. The results showed that significant differences occurred in both the experimental and control groups. However, significant differences in the cervical dysfunction index, cervical ROM, and pain perception only existed in the experimental group. Seo et al. [4] studied the effect of Pilates exercise on Cobb's angle of the lumbar spine in scoliosis patients. The experimental object is high school students with scoliosis. They will do Pilates exercises such as spine twisting, pelvic isometric, and jackknife extension in 10 weeks. As a result, Cobb's angle decreased statistically significantly by about 5 degrees. Kuo et al. [5] studied the effect of Pilates exercise on the sagittal spine posture in the elderly. There was a low-significant change in the thoracic kyphosis in 34 healthy elders over 60. However, there are no other significant changes observed because of the short study period.

2.2 Posture correction using Kinect sensor

Kim et al. [6] proposed a method for recognizing a moving posture by processing the skeletal information obtained from the Kinect sensor. After extracting the feature information of the human body region from which the background has been removed by using an image processing technique from the input image, each feature vector for human posture recognition is generated using Principle Component Analysis (PCA). A fuzzy classifier was designed using these feature values to recognize postures such as walking left and right, raising and lowering one hand, raising and lowering both arms, and raising and lowering feet. Kim et al. [7] proposed a system that allows users to exercise with the correct posture by comparing the real-time user's exercise posture obtained by Kinect with the correct posture. The received user's skeletal information is stored in the database in the form of a matrix, and the concordance of this and the posture stored in the existing database is calculated to evaluate the accuracy of the movement. Cho et al. [8] proposed a system for estimating movement posture using multiple Kinect sensors. By applying the Iterative Closest Point (ICP) algorithm to the skeletal information obtained from multiple Kinect sensors, the relative positions and directions between multiple Kinects are corrected, and the posture is estimated using the weighted average of the feature values. All 16 exercise postures required in the

exercise program, such as open arms, side leg lifts, and squats, can be recognized by this method.

2.3 Posture correction using wearable sensors

Kim et al. (2019) proposed a system that allows multiple users to receive posture correction through real-time communication between a wearable device and a Kinect sensor. An additional Kinect sensor can detect body parts that are not assigned with a wearable device and provide individual feedback by an algorithm that compares real-time posture with the posture that has been standard for users. Kim et al. (2017) proposed a system that can correct the user's exercise posture. Among Bluetooth Low Energy (BLE) communication methods, advertise mode allows multiple users to access at the same time. After determining whether to correct posture by applying the algorithm of each exercise to the measured acceleration and angle values, it provides voice feedback on posture for jumping rope, squat, and lunge exercises. Zhang et al. (2016) proposed a system that predicts the posture of body for a cyclist using wearable sensors. 3D posture information is obtained by attaching tri-axial gyroscope sensors to arms and legs, and additional information is obtained and utilized by wheel encoders and pedal crank encoders. Posture was predicted based on the Extended Kalman Filter, and both indoor and outdoor experiments showed excellent performance.

3 Proposed System

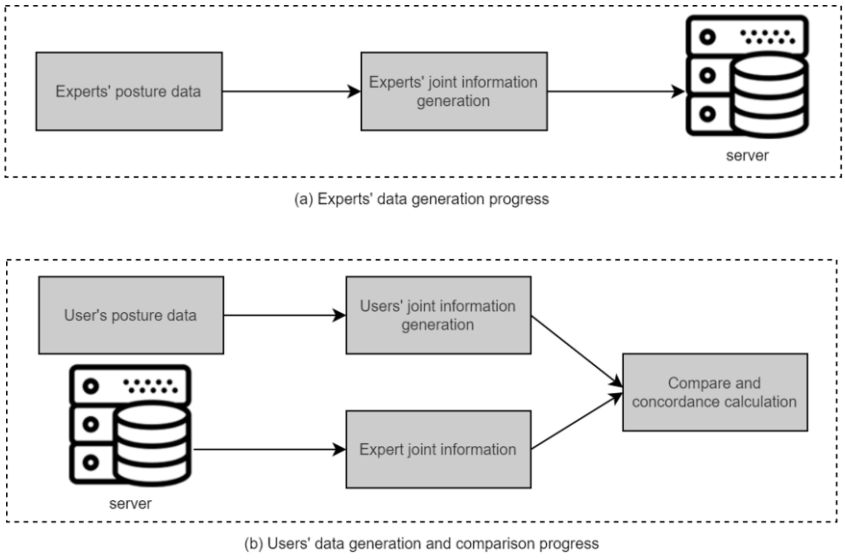


Figure 1. The work principle of real-time posture correction system

As shown in Figure 1, the proposed real-time posture correction system consists of expert data generation and user data generation and comparison. First, the expert joint information is extracted from the video of expert Pilate exercise posture. Then, it receives the user's movement motion through the smartphone camera, generates the user's joint information in real-time, compares it with the expert joint information stored in the server, and outputs the result.

3.1 Skeleton Extraction

Seventeen body position coordinates (x-coordinate, y-coordinate), as shown in Figure 2, predicted by PoseNet, were a vision model that can predict a person's posture. Their reliability scores are inserted into a matrix stored on the server for each frame of the video. Then, the concordance is calculated by comparing it with the user's joint information obtained in real-time.

ID	Part
0	Nose
1	Left eye
2	Right eye
3	Left ear
4	Right ear
5	Left shoulder
6	Right shoulder
7	Left elbow
8	Right elbow
9	Left wrist
10	Right wrist
11	Left pelvic area
12	Right pelvic area
13	Left knee
14	Right knee
15	Left ankle
16	Right ankle

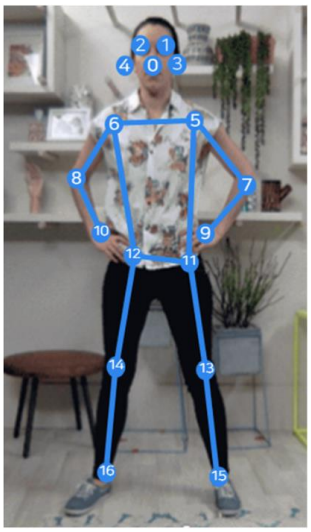


Figure 2. Joint positions predicted by PoseNet

3.2 Angle Calculation

As shown in Figure 3, if the body size of the expert and the user are different, even the same action will be recognized as another action due to the difference in coordinates.

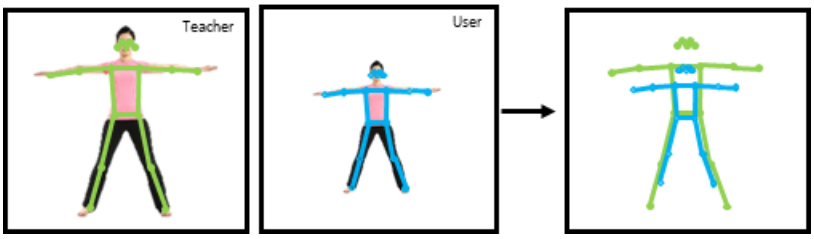


Figure 3. Problems encountered when comparing joint coordinate information between experts and users

To solve this problem, we devised an algorithm that can evaluate motion with minor error even when body dimensions are different by comparing angles of specific parts rather than simply comparing coordinate values of joints. The joint information predicted by PoseNet by applying the atan2 formula that calculates the absolute angles between two points and additionally measured 16 angles showed in Figure 4, then compares the joint angle information between experts and users and obtain the differences between them.

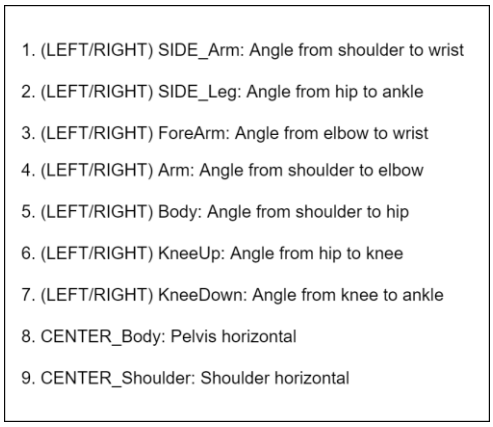


Figure 4. Additional angle information

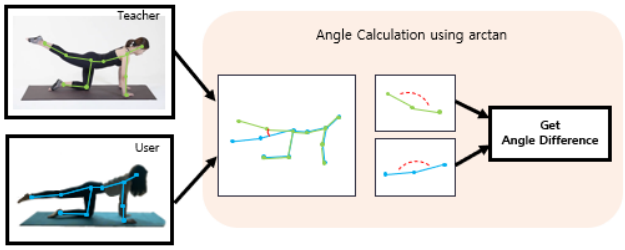


Figure 5. The method of angle comparison

3.3 Score Calculation

Evaluation criteria were proposed based on the difference between the joint angles of experts and users. If the error is less than or equal to 5 degrees, the evaluation result is 100 points. If the error is less than or equal to 5 degrees, the evaluation result is 90 points. If the error is less than or equal to 15 degrees, the evaluation result is 80 points. If the error is less than or equal to 20 degrees, the evaluation result is 70 points. If the error is higher than 20 degrees, the evaluation result is 50 points. The final score is for every 15 frames of the video, and the system can return real-time feedback of “Good” if it is 90 or more, “Normal” if it is 80 or more, and less than 90, and “Bad” if it is less than 80.

4 Results

The left side of Figure 6 is the [learning] stage, which is based on the current motion and provides real-time feedback until reaching the next classification motion to learn the basic skills for the corresponding exercise motion. On the right is the [Exercise] stage, it compares the user's postures with the expert's postures in real-time, and returns "Good", "Normal" and "Bad".

Figure 6. Launch POS: E app

Figure 7 shows the results of two experimenters performing side jack and wide squats. The accuracy of the side jack and wide squat was 92.4% and 91%, respectively. With the increase in the number of exercises, the user's awareness and accuracy of the posture become higher.

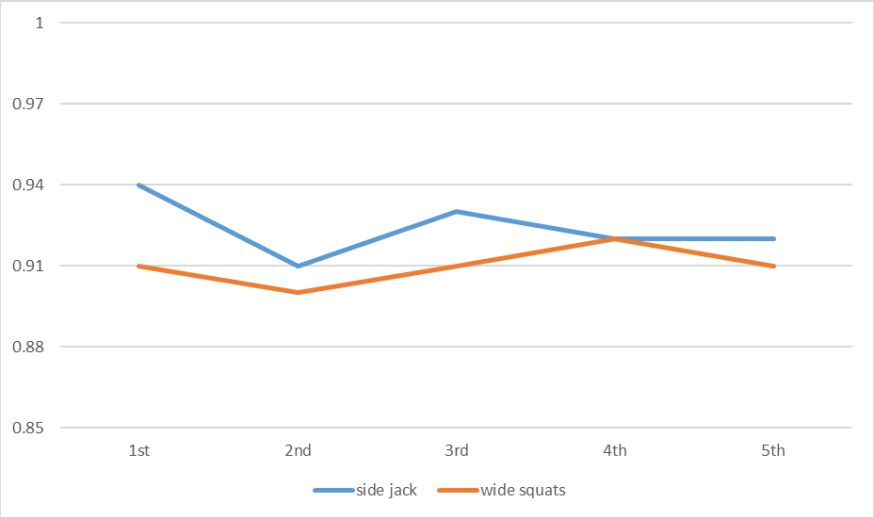


Figure 7. Experiment results

5 Conclusion

In this paper, we have proposed a real-time feedback system for posture correction during exercise using a smartphone. Joint and additional angle information are extracted from experts' videos by video processing and deep learning. In the future, we plan to extend the real-time posture correction system to the fields of rehabilitation, health, and dance. However, it is difficult for the camera to predict joints that are overlapped or covered, so it is necessary to use an additional sensor to perform 3D recognition that can solve the issue of the covered area and keep accurate evaluation.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1I1A3042145).

References

1. Hyun, A.H., Cho, J.Y., “Effects of Non-face-to-face Home Pilates Exercise for 8 Weeks in the Post-corona era on Body Composition, Abdominal Obesity, Pelvic Tilt and Muscle Strength, and Back Pain in Women After Childbirth”, Kinesiology 30.1 (2021): 61-69.
2. Hyun, A.H.; Cho, J.Y., “Effects of 12-weeks Pilates Mat Exercise on Body Composition, Delivery Confidence, and Neck Disability Index in Pregnant Women”. Sports Sci. 2019;(36):43-55.
3. Park, J.H., “Effects of Posture Correction Exercise Participation on Turtleneck Syndrome”, Graduate School of Sports Industry, Kookmin University Thesis, 2013
4. Seo, J.H., Hong, S.K., “The Pilates Effects about Scoliosis”, Journal of the Korean Society of Radiology, 2014
5. Kuo, Y.L., Tully, E.A., Galea, M.P., “Video Analysis of Sagittal Spinal Posture in Healthy Young and Older Adults”, Journal of Manipulative and Physiological Therapeutics, 2009
6. Kim J.K., Kim, S.K., J, Y.H., Park, J.B., “Human Posture Recognition Using Kinect Sensor”, The 43rd Summer Conference of the Korean Electrical Society, 2012
7. Kim J.W., Park, S.J., Min, K.Y., Lee, G.Y., “Virtual reality-based situational immersion English conversation learning system”, Journal of Convergence for Information Technology, 2017
8. Cho, Y.J., Park, K.S., “Design and Development of the Multiple Kinect Sensor-based Exercise Pose Estimation System”, Journal of the Korea Institute of Information and Communication Engineering, 2017
9. Kim, D.W., Nam, H.J., Lee, S.Y., Ham, Y.K., Seo, O.S., Lee, H.J., “Multi-User Home-Training Healthcare System Using Kinect Sensor and Wearable Devices”, The Journal of Korea Institute of Communications and Information Sciences, 2019
10. Kim, J.H., Lee, E.J., Jeong, S.H., Han, J.H., Cho, H.Y., Lee, H.J., “Self-Training Healthcare System Using Multi-Communication Wearable Devices”, The Korean Institute of Information Scientists and Engineers, 2017
11. Zhang, Y.Z., Chen, K., Yi, J.G., Liu, T., Pan, Q., “Whole-Body Pose Estimation in Human Bicycle Riding Using a Small Set of Wearable Sensors”, IEEE/ASME Transactions on Mechatronics, 2016

Survey on Piano Keyboard Recognition Research

So-Hyun Park[1], Yeon-Ji Kim[2], Da-Young Ji[3], Mi-Yeon Kim[4], Young-Ho Park[5, *]

[1] Big Data Using Research Center, Sookmyung Women's University,
[2] Department of Computer Engineering, Sookmyung Women's University,
[3] Dept. of Statistics, Sookmyung Women's University,
[4] Dept. of Airline Services & Tourism, Seoyeong University,
[5] Dept. of IT Engineering, Sookmyung Women's University,
{shpark, rladuswl01, dys621124, yhpark}sm.ac.kr, myk@seoyeong.ac.kr

Abstract. Recently, image processing studies in the field of arts and sports are emerging. For example, image information is used to recognize a musical instrument or to recognize and classify a player's behavior. In this study, research related to piano keyboard recognition is investigated and comparatively analyzed during the study of instrument recognition.

Keywords: Image Processing, Piano Keyboard Recognition, Object Recognition

1 Introduction

Recently, image processing studies in the field of arts and sports are emerging. For example, image information is used to recognize a musical instrument or to recognize and classify a player's behavior. More specifically, in order to recognize the piano keyboard among musical instruments, studies are underway to recognize the keyboard by applying image processing techniques to the performance image taken from above of the performer [1-6]. In this study, various keyboard recognition related studies are investigated. In addition, we intend to apply the hough algorithm and the sobel algorithm, which are mainly used in piano recognition research, to C3Pap, a piano performance image dataset.

The structure of the study is as follows. Chapter 2 introduces studies related to piano keyboard recognition. In Chapter 3, the hough algorithm and sobel algorithm, which are mainly used techniques for recognizing piano keys, are applied to C3Pap data, and the results are analyzed. Chapter 4 draws conclusions and discusses future research.

[1] Please note that the LNCS Editorial assumes that all authors have used the western naming convention, with given names preceding surnames. This determines the structure of the names in the running heads and the author index.

2 Piano Keyboard Recognition Research

This section introduces studies on recognizing the piano keyboard. To this end, Section 2.1 introduces the study of piano keyboard recognition. Section 2.2 introduces studies that recognized the piano keyboard from the side. Section 2.3 introduces the study of object recognition other than the piano keyboard.

2.1 Piano Keyboard Recognition Research

A method for detecting and tracking the piano keys being pressed in the piano performance video is developed. For this, image processing tools such as Sobel filter, Hough transform, morphology calculation, and region labeling are used [1]. Sobel Filter (Sobel Operator) is an edge detection algorithm based on the first derivative and can detect both vertical, horizontal, and diagonal directions [7]. Hough Line Transform is a technique that detects any shape if that shape can be represented in mathematical form in OpenCV, and is used to detect lines in an image [8]. Connected-component labeling (region labeling) is used to detect connected objects in binary images and can also process high-dimensional color images and data [9].

We propose a method to identify the piano keyboard in the video. By using binary threshold, Sobel operator and Hough transform algorithm, the keys of the piano keyboard were successfully identified in the webcam video image, and the tolerance to camera movement and occluded keys was demonstrated [2]. This study develops an interactive piano tutoring system that can observe students playing the piano and provide feedback on hand movements and musical accuracy. It consists of video and audio data, and the Pressed Key is detected in the video using the Inception V3 model [3]. Using the CNN model, the piano performance is analyzed based on only visual information. The black/white key is detected by applying the threshold in the Hough transform algorithm [4]. The keyboard is recognized from the video camera mounted on the piano keyboard, each key is recognized, and the Pressed Key is detected to perform Music Transcription. Hough line transform and connected components labeling algorithm are used for keyboard recognition [5].

It is provided in the current video, and it provides the latest video about the location, time and time of the current animation, time and animation. It can be recognized from different locations, and can use Hough Line Transform, Canny Edge Detector, Otsu's threshold algorithm [6]. Canny Edge Detector is a technology that detects edges in OpenCV [10]. Otsu's Thresholding Algorithm provides solid information about binarization [11].

2.3 Object Recognition Research

CompositionalNets are more powerful for classifying partially occluded objects than for object detection in the occlusion state of deep learning approaches such as Faster R-CNN [12]. We propose a new framework (Spatial Layout Model) for detecting multiple objects in a single image and inferring occlusion [13]. We propose a new Mask Generator based on Faster R-CNN as a shield recognition object detector

with enhanced features by simultaneously detecting occluded objects and fully utilizing spatial information [14]. We describe a novel framework based on boosting algorithms and cascade structure to efficiently detect faces/objects with occlusion [15].

3 Apply Piano Keyboard Recognition Algorithm to C3Pap Dataset

In this section, the hough algorithm and sobel algorithm, which are mainly used in piano recognition research, are applied to the piano performance image dataset C3Pap and the results are analyzed. For this, it is divided into a process of detecting piano keyboard using yolo v5 and then recognizing the keys in detail by applying an image processing algorithm.

3.1 Step 1 : Keyboard detection with Yolo v5

How to train piano keys using Yolov5 is as follows. After creating a custom dataset to learn the piano keyboard with the labelme [16] program, we trained it. During training, all types of images of C3Pap were used to label the keyboard [17]. The number of training data is 5729. It was trained using yolo5s.pt, which is the lightest model among Yolo v5 [18].

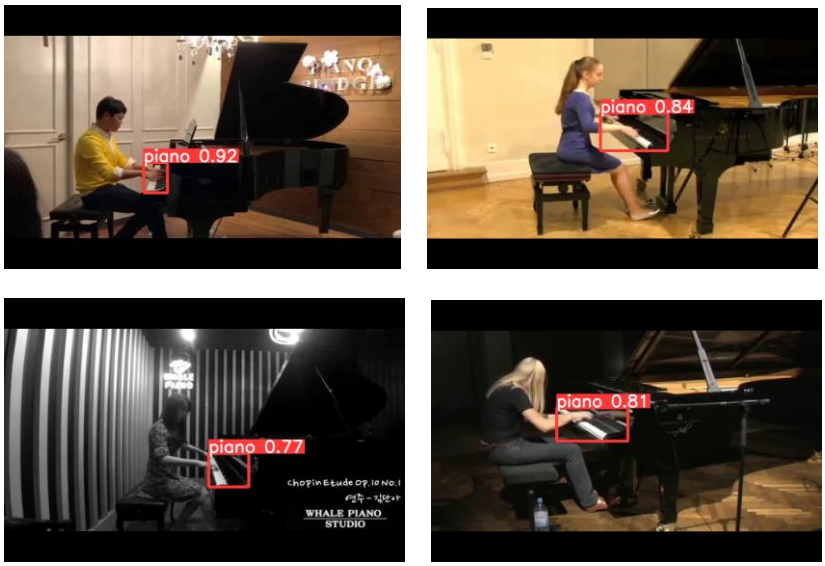


Fig. 1. Example of keyboard detection using Yolo v5

3.2 Step 2-3 : Piano keyboard recognition using image processing techniques

This section introduces a method of creating lines on the piano keyboard using video information in figure 2.

To this end, in Step 2, which has received the coordinates of the piano area recognized in Step 1, the remaining areas except for the corresponding coordinates are painted black. This is an operation to easily find only the outline of the white keys of the piano in Step 3. To do this, first, the coordinates of the piano area recognized in Step 1 are received. Then, the coordinate area is filled with white through the fillpoly function. The Fillpoly function receives the coordinates of an image and a polygon and draws a polygon on the corresponding image. When the bitwise_and operation is completed by displaying only the part painted in white and displaying the rest in black, only the part of the piano area marked in Step 1 remains, and the rest is displayed in black. Finally, in Step 3, the image is converted to grayscale through the cvtColor function as a pre-work to smoothly detect only the white area of the piano. In this case, the cvtColor function is a function that converts the color space of an image.

In Step 3, the work for detecting the white keyboard area of the piano is carried out. In this case, light gray to white was designated as the recognizable range. Then, through a mask, the area in the image where the color range exists is marked white, and the rest of the image is blackened. At this time, since a part of the keyboard may be displayed in black due to the hand covering the keyboard, an image expansion operation is performed to compensate for this. Next, the Canny algorithm is performed to extract the outline of the processed image. Finally, if the straight part of the extracted outline is detected using the Hough transform, the white keyboard part of the piano can be detected as a line. Then, it tries to detect a shape using the findContour function in the recognized white area. The findContour function is a function that finds the boundary of an area with the same intensity in an image. If multiple figures are detected, only one figure having the largest area is detected by comparing the area of each figure. After that, it is checked whether one detected figure is a rectangle, and if it is not a rectangle, the border of the figure is simplified until a rectangle is detected through the approxPolyDP function. In this case, the approxPolyDP function is a function that approximates a polygon by reducing the number of vertices according to the degree of simplification specified for a given polygon. Finally, if a rectangle is detected as a result of simplification, the four coordinate values that form the rectangle are received and lines are drawn in different colors on the front and back of the white keyboard of the piano.

Since C3Pap is actual performance video data, not in a limited environment, various variables exist. In case 1 and case 2 of figure 3, one line was detected instead of two lines. In addition, depending on the shooting angle, there are cases where the hand is larger than the piano, and even in this case, the line could not be detected properly.

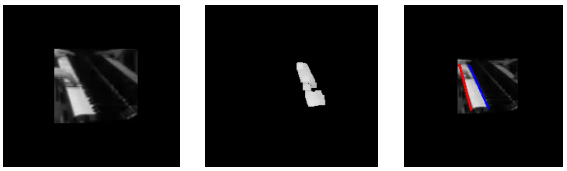


Fig. 2. Result of Creating lines on the piano keyboard using video information

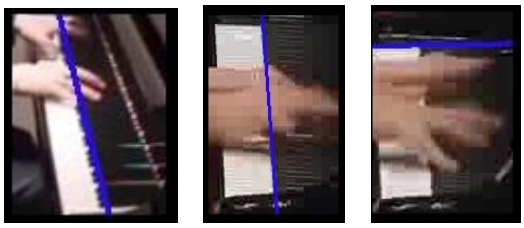


Fig. 3. Incorrectly generated lines example

4 Conclusion

In this paper, various keyboard recognition related studies were investigated. In addition, the hough algorithm and sobel algorithm, which are mainly used in the study of piano recognition, were applied to the piano performance image dataset, C3Pap. Unlike the existing performance image data set, C3Pap is an image data set taken from the side of the performer, so detection may not work well. In a future study, we intend to conduct a study to improve the keyboard recognition accuracy in the image of the side of the piano.

Acknowledgments. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology [NRF-2021R1C1C2004282].

References

1. P. Suteparuk, “Detection of piano keys pressed in video,” Dept. of Comput. Sci., Stanford Univ., Stanford, CA, USA, Tech. Rep., 2014.
2. A. Goodwin, R. D. Green, “Key detection for a virtual piano teacher,” 2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013), pp. 282-287, 2013.
3. J.W. Lee et al, “Observing pianist accuracy and form with computer vision,” 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2019.

4. S.J. Kang, J.Y. Kim, S.G. Yoon, “Virtual piano using computer vision,” arXiv preprint arXiv:1910.12539, 2019.
5. M. Akbari, H. Cheng, “Real-time piano music transcription based on computer vision,” IEEE Transactions on Multimedia, vol. 17, no. 12, pp. 2113-2121, 2015.
6. R. McCaffrey, “Piano Music Transcription Based on Computer Vision,” Dept. of Computer Science, University of Dublin, Trinity College, Diss, Master Dissertation, 2017.
7. Sobel Filter, https://docs.opencv.org/4.5.3/d5/d0f/tutorial_py_gradients.html
8. Hough Line Transform, https://docs.opencv.org/4.5.3/d3/de6/tutorial_js_houghlines.html
9. Connected-component labeling, https://en.wikipedia.org/wiki/Connected-component_labeling
10. Canny Edge Detector, https://docs.opencv.org/3.4.15/d7/de1/tutorial_js_canny.html
11. Otsu's Thresholding Algorithm, https://en.wikipedia.org/wiki/Otsu%27s_method
12. A. Wang, Y. Sun, A. Kortylewski, A. Yuille, “Robust object detection under occlusion with context-aware compositionalnets,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
13. Xiang. Y, Savarese. S, “Object detection by 3d aspectlets and occlusion reasoning,” Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013.
14. J. Xu, H. Wang, M. Xu, F. Yang, Y. Zhou, X. Yang, “Feature-Enhanced Occlusion Perception Object Detection for Smart Cities,” Wireless Communications and Mobile Computing, vol. 2021, 2021.
15. Y.Y. Lin, T.L. Liu, C.S. Fuh, “Fast object detection with occlusions,” European Conference on Computer Vision, Springer, Berlin, Heidelberg, 2004.
16. labelme, https://github.com/wkentaro/labelme
17. S.H. Park, Y.H. Park, “Audio-Visual Tensor Fusion Network for Piano Player Posture Classification,” Applied Sciences, vol.10, no.19, 2020.
18. yolo v5, https://github.com/ultralytics/yolov5

3D Convolutional Neural Network for Crowd Behavior Classification in Surveillance Videos

Jong-Hyeok Choi[1], Aziz Nasridinov[1,2], Yoo-Sung Kim[3,*]

[1] Bigdata Research Institute, Chungbuk National University, Cheongju, South Korea
[2] Department of Computer Science, Chungbuk National University, Cheongju, South Korea
[3] Department of Artificial Intelligence, Inha University, Incheon, South Korea

{leopard, aziz}@chungbuk.ac.kr, yskim@inha.ac.kr

Abstract. Classifying crowd behavior from videos is an ongoing challenge in computer vision because of its complexity. In particular, crowd behavior classification in a video surveillance system is necessary because it enables effective monitoring of various situations, such as dangerous situations. Due to this demand, recently, various studies are being conducted to perform such classification using 3D Convolutional Neural Networks (CNN). In this paper, for developing a new deep learning model, we will summarize the existing 3D CNN-based studies on crowd behavior classification and define the baseline performance through experiments.

Keywords: 3D CNN, Crowd Behavior Classification, Video Surveillance.

1 Introduction

In the computer vision field, various studies conducted to enable computers to understand human behavior. Among these studies, human behavior classification [1, 2] and crowd behavior classification [3-6] receive continuous attention because they can automatically and effectively monitor various situations, such as dangerous situations, using videos obtained from a video surveillance system. However, it was difficult to classify them due to the complexity of the unpredictable human actions and various interactions between humans [3]. Recently, various studies proposed significantly improving classification performance by solving complexity through deep learning methods to solve this problem [1-5]. These studies mainly start with image-based classification based on 2D Convolutional Neural Networks (CNN), and recently, video-based classification using 3D CNN has been proposed to analyze more complex situations accurately [2]. Among these studies, notable studies can be mentioned MoViNet [7] and SlowFast [8] proposed for human action classification, which showed very high accuracy from video datasets such as Kinetics [2,7]. However, it is not easy to apply these models to video surveillance systems. Because there are very few humans in the video dataset like the Kinetics used to train these

* Corresponding authors: Yoo-Sung Kim (email: yskim@inha.ac.kr).

models. However, in the real world, video surveillance includes many humans, and very complex interactions occur between them [3], so it is challenging to perform crowd behavior classification through existing methods.

In this paper, to solve the problem of existing crowd behavior classifications required in the video surveillance system with a deep learning-based method, the problems related to crowd behavior classification will be summarized through related studies. In addition, before proposing a new deep learning-based method, we will define the baseline performance required for deep learning model development using the existing 3D CNN through experiments.

2 Related Work

2.1 Dataset

The most important thing in solving computer vision problems based on deep learning is the dataset. However, unlike UCF-101, HMDB-51, and Kinetics, which are widely used datasets for human action classification [2,7], there were problems such as a small number of videos or the inability to diversify crowd behavior for data for crowd behavior classification. To solve these problems, Dupont et al. proposed the Crowd-11 dataset [3] that can define a total of 11 crowd behaviors based on the previously proposed crowd-related datasets and real-world videos obtained from YouTube, etc.

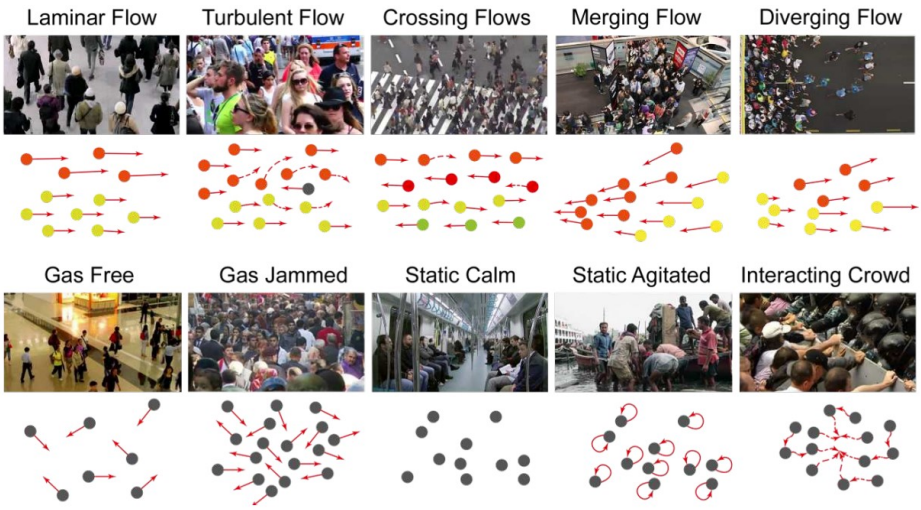


Fig. 1. Characteristics of each crowd behavior class in the Crowd-11 dataset [3].

Crowd-11 dataset defines a total of 11 crowd behavior classes, including no crowd, and Fig. 1 shows the differences between these crowd behavior classes. And Fig. 1, the colored dots represent the movements of a grouped crowd, and the gray dots represent the individual movements. Moreover, each crowd behavior class had

classified according to the various interactions between the crowds. Additionally, most videos that consist of the Crowd-11 dataset were obtained from various locations in the real world and consisted of 6272 video clips of less than 5 seconds. Due to these characteristics, the Crowd-11 dataset has received various attention as it is suitable for performing deep learning-based crowd behavior classification in surveillance videos.

2.2 3D Convolutional Neural Networks

Currently, most of the deep learning-based methods using the Crowd-11 dataset perform classification through 3D CNNs [3-5]. The method proposed with the Crowd-11 dataset used an early 3D CNN model called Convolutional 3D (C3D) [1] and showed high performance compared to the existing image classification models. After that, Inflated 3D ConvNet (I3D) [2], which showed good performance in human action classification by extending the GoogLeNet as a famous image classification model, was also applied to crowd behavior classification using the Crowd-11 dataset. In particular, I3D significantly improved the accuracy of human action classification by training RGB frames and optical flow frames on different I3D models and merging the inference results. Because these two-stream I3D supplements RGB and optical flow's disadvantages by merging spatial features through RGB and temporal features through optical flow. So, even with Crowd-11 dataset, the two-stream method performed the most accurate classification. However, the classification accuracy for turbulent flows, crossing flows, merging flows, and diverging flows, which characterize the behavior between grouped crowds, is not high, so the need for a new model that can infer the differences between them is very high [3, 4].

3 Experiment Result

This section evaluated the accuracy of crowd behavior classification using the I3D model through experiments. Because, in previous studies, they changed the structure of the I3D model for the experiment or hyperparameter setting and experimented in a very diverse way. Therefore, this paper conducts an experiment using the original I3D structure to obtain a more objective baseline performance. And for this purpose, the experimental environment and result are discussed in the following subsections.

3.1 Experimental Environment

In the case of the original I3D, the inference had performed using 64 RGB frames or 64 optical flow frames obtained from the same video, and the last classification layer performs classification as many as classes defined in the dataset. This paper uses this structure identically except the last classification layer for changing the number of classes to 11, which is the number of classes in the Crowd-11 dataset.

In the case of training and verification of I3D using the Crowd-11 dataset, training and verification are using only 5987 video clips consisting of 64 frames or more from

a total of 6146 video clips, excluding videos lost online. And, since the Crowd-11 dataset is not splitting into training and verification as official annotations, training and verification are performed by randomly splitting the whole crowd-11 dataset 8 to 2. In addition, since the performance of classification models using optical flow can vary greatly depending on the optical flow algorithm, we use the optical flow frame generated every frame through the traditional TV-L1 algorithm [9] for training and verification. And, in training, 64 frames extracted from video clips were resized to 256×256 and then resized to 224×224 with randomly crop and randomly flip, and in verification, only by resizing to 224×224 was performed.

Finally, the model has trained 100 epochs with a batch size of 16 on two Titan RTX through the SGD optimizer with a momentum of 0.9 and a base learning rate of 0.1. And we decayed the learning rate using a cosine annealing scheduler.

3.2 Experiment Result

Table 1 shows the classification accuracy of the I3D baseline model trained on the Crowd-11 dataset. At this time, since 64 frames were extracted from a random starting point in the video clip, the validation has been repeated five times using only RGB or optical flow and the two-stream.

Table 1. Classification accuracy of the I3D baseline model for the Crowd-11 dataset.

Variant	1	2	3	4	5	Average
RGB	**72.62%**	71.79%	71.95%	71.37%	71.87%	71.92%
Optical Flow	71.79%	**72.37%**	71.79%	71.20%	71.20%	71.67%
Two-Stream	77.38%	77.30%	76.96%	**77.80%**	77.46%	**77.38%**

As a result of the experiment, it confirmed that using two-stream showed the highest accuracy like the results of human action classification studies or crowd-11 related studies. And, since the overall classification accuracy also shows a promising result, developing a crowd behavior classification method in surveillance videos using 3D CNN seems possible.

4 Conclusion

In this paper, we have summarized existing deep learning-based crowd behavior classification methods to solve the crowd behavior classifications required in the video surveillance system. Also, we experimented with existing 3D CNN architectures using the Crowd-11 dataset, a large crowd scene video set, to define the baseline performance required to develop a new deep learning-based method.

In future work, we will apply SlowFast or MoViNet proposed after I3D to the Crowd-11 dataset and then propose a new deep learning-based method suitable for crowd behavior analysis through a new 3D CNN architecture.

Acknowledgments. This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-00203, Development of 5G-based Predictive Visual Security Technology for Preemptive Threat Response). This work was also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1I1A1A01041815).

References

1. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3d convolutional networks. In: IEEE International Conference on Computer Vision, pp. 4489-4497. IEEE (2015)
2. Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 6299-6308. IEEE (2017)
3. Dupont, C., Tobias, L., Luvison, B.: Crowd-11: A dataset for fine grained crowd behaviour analysis. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 9-16. IEEE (2017)
4. Bendali-Braham, M., Weber, J., Forestier, G., Idoumghar, L., Muller, P. A.: Transfer learning for the classification of video-recorded crowd movements. In: 11th International Symposium on Image and Signal Processing and Analysis, pp. 271-276. IEEE (2019).
5. Bendali-Braham, M., Weber, J., Forestier, G., Idoumghar, L., Muller, P. A.: Ensemble classification of video-recorded crowd movements. In: 12th International Symposium on Image and Signal Processing and Analysis, pp. 152-158. IEEE (2021)
6. Sreenu, G., Durai, M. S.: Intelligent video surveillance: a review through deep learning techniques for crowd analysis. Journal of Big Data, 6(1), 1-27 (2019)
7. Kondratyuk, D., Yuan, L., Li, Y., Zhang, L., Tan, M., Brown, M., Gong, B.: Movinets: Mobile video networks for efficient video recognition. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16020-16030. IEEE/CVF (2021)
8. Feichtenhofer, C., Fan, H., Malik, J., He, K.: Slowfast networks for video recognition. In: IEEE/CVF International Conference on Computer Vision, pp. 6202-6211 (2019)
9. Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime tv-l 1 optical flow. In: Joint pattern recognition symposium, pp. 214-223. Springer, Berlin, Heidelberg (2007)

An Accurate Extraction of Facial Meta-Information Using Selective Super Resolution from Crowd Images

Jieun Park[1], Yurim Kang[1], Yoo-Sung Kim[1,2*]

[1] Department of Information and Communication Engineering,
[2] Department of Artificial Intelligence,
Inha University
Incheon 22212, Korea

*yskim@inha.ac.kr

Abstract. An accurate extraction scheme of facial meta-information from low resolution crowd images is proposed. In order to detect crowd abnormal situations, extracting facial meta-information from crowd images is very helpful. However, since many crow images are of low resolutions, extracting facial meta-information is pretty difficult. To extract accurately facial meta-information from the low resolution crowd images, some face images which are not suitable to easily extract the meta-information should be classified and be improved the quality by a super resolution method. To confirm the feasibility of the proposed scheme, since gender of person can be regarded as very important facial meta-information, we compare the gender classification accuracies of using the proposed scheme and that of using the input crowd image itself. According to the experiment results, the proposed facial extraction scheme from crowd images using selective super resolution can improve the gender classification accuracy for crowd images than that for using the original crow images.

Keywords: Crowd abnormal detection, Low resolution crowd images, Facial meta-information, Selective super resolution, Gender classification

1 Introduction

Crowd analysis becomes important computer vision task because of its usability, including crowd movement analysis, abnormal behavior detection, and crowd size estimation[1]. As the urban population increases recently, crowd analysis and monitoring becomes more essential for public safety purposes. Crowd monitoring using intelligence video surveillance systems which can be encountered in daily life is increasing in many places such as subway stations, airports, and department stores even on the streets with a lot of people. In such a crowd situation, it is important to detect and prevent abnormal situations in advance through crowd monitoring. Therefore, accurate and rapid judgement of situation in crowd monitoring is required.

To recognize accurately and fast abnormal crow situations, using facial meta-information of person in crowd such as gender, wearing mask or sunglasses, and even

emotional expressions is very helpful. Of course, using these information can help to identify person in multi-camera environments. For example, if it is possible to analyze a face by accurately classifying gender, it will be helpful to understand a situation.

However, there are many difficulties to accurately extract facial meta-information from crowd images since they may be of low resolution and have heavy occlusions. In the case of face detection in a crowd with long distance from camera and many people, it is difficult to utilize it in future research because image resolution is not good as the size of the face decreases. Therefore, we propose a method to extract facial meta-information from low-resolution crowd images using a selective super resolution for improving the quality of the face images, so make them possible to use in future research. By solving low resolution problem occurring in crowd faces, the accuracy of gender classification can be improved, which can be contributed to crowd monitoring.

The rest of this paper is as follows. Section 2 describes the previous related works on the crowd dataset and on face detection model using convolutional neural network. In Section 3, the data set which is used for this study is introduced, and the proposed extraction scheme of facial meta-information from low-resolution images is discussed. In Section 4, the experiment to show the effectiveness of the proposed scheme is discussed. This paper is concluded with a short discussion of future studies in Section5.

2 Related Work

Crowd-11 dataset is widely used for crowd analysis researches. It classified crowd videos into 11 classes from considering various situations and places to classify abnormal behavior of crowds[1]. This includes crowd images in daily life and is close to the general CCTV images targeted in this paper. Therefore, some images from 8 classes suitable for this paper were selected except classes of 'Diverging Flow', 'Interacting Crowd', and 'No Crowd', which are difficult to identify crowd faces. Face detection tasks in crowd of real world have problems such as low resolution and heavy occlusion, unlike single person or refined situations. There is the widely used WIDER Face dataset for face detection in crowd[2]. In this dataset, considering various conditions such as pose, scale, and occlusion, it is classified into easy and hard situation according to the difficulty of face detection. In YOLO5Face([3]), they propose a method to improve the performance of face detection in hard situation close to the real world. It has high accuracy on very small faces, by adding a five-point landmark regression head into YOLOv5. Also, it belongs to the high real-time among object detection models. Therefore, we utilized YOLO5Face([3]) for face detection which is appropriate for the purpose of this study.

3 Extracting Facial Information by a Selective Super Resolution

3.1 Overall procedure for the extraction of facial meta-information

Fig. 1 shows the overall procedure for the extraction of facial meta-information from the crowd images. First, when crowd image comes in, face detection is performed by YOLO5Face([3]) as described in the previous section. Cropped face image from the crowd has very low resolution as the number of people increases or the size of the face decreases due to the distance from the camera. The lower resolution, the more difficult it is to identify the face, so it will be difficult to utilize in future research. Therefore, when the resolution of the face images is smaller than the threshold, we regard the face image as the hard case for extracting facial meta-information from the original image so that super resolution step is applied to improve the quality of the input. As the threshold value, after careful consideration of input data succeed by experiment we decided to use 48 * 48. At the last step we can extract the facial meta-information. In this paper, we try to extract gender information from the person's face.

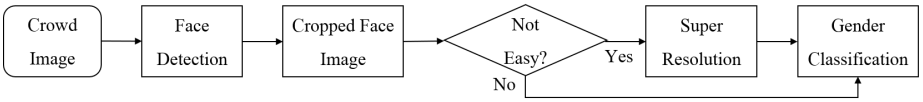


Fig. 1. Overall procedure for the extraction

3.2 Dataset

We need crowd dataset in real world for face detection and subsequent research in crowd analysis. Crowd data in real world means most like real situation, rather than a refined one. The most similar Crowd-11 dataset([1]) was used as our experiment, and some of them are selected in consideration of the angle and distance of camera. In other words, we included the distance enough to reveal the whole upper body of person and excluded the angle that interferes with face recognition due to heavy occlusion between people.

From Crowd-11 dataset([1]), we extracted total of 551 images by sampling 111 videos with an average length of 4 to 5 seconds at one frame per second as shown as Fig 2. To use the selected images in face detection and gender classification, ground truth data is required. We make annotation files by labeling the x and y coordinate values of face region and gender of the face in the images. There are 6,146 male faces and 4,984 female faces in total 11,130 faces.

Fig. 2. Examples of selected images for this study from Crowd-11 dataset

Within the selected crowd images, the performance of face detection and subsequent research is different because the size of the face. Since this paper focused on crowd face and analyze the experimental results according to the resolution of the face images, the selected images were classified according to the resolution of the image extracted from the face, not the resolution of the entire image.

A total of 11,130 faces were sorted according to the number of pixels, which means resolution. They are classified into three cases as shown in Table 1 according to the section where the number of images decreases sharply depending on the number of pixels. The case where the resolution of the face image itself is relatively large is easy case, the middle resolution is medium case, and the smallest is hard case. It is possible to identify the face in easy case to confirm with the naked eyes, while medium and hard cases have difficulties. The number of faces in hard case is the largest because there are more cases of small faces compared to large faces in crowd.

Case	Number of Face Images	Resolution
Easy Case	596	48x48 ~
Medium Case	5,102	29x29 ~ 48x48
Hard Case	5,432	~ 29x29

Table 1. Classification of the difficulty level of face images based on the resolutions

3.3 Face Detection

In this paper, face detection in crowd situations like CCTV video is required. Therefore, face detection was performed on the finally selected Crowd-11 [1] dataset using YOLO5Face([3]) model, which has excellent performance in the WIDER face dataset([2]) close to the real world.

Next processing step is conducted using face images extracted from the Crowd-11 dataset through face detection. In the case of the detected and enlarged face images of crowd, the resolution is not good. So, it is difficult to directly utilize original face images in future research. Therefore, among the studies to improve resolution, ESRGAN([4]) which has excellent perceptual improvement ability was applied to obtain better results in subsequent studies.

3.4 Super Resolution

Next step is conducted using face images extracted from the Crowd-11 dataset through face detection. In the case of the detected and enlarged face images of crowd, the resolution is not good as shown on the left of Fig 3, because there are many people, and the distance from camera is so long. So, it is difficult to directly utilize original face images in future research. Therefore, among the studies to improve resolution, ESRGAN([4]) which has excellent perceptual improvement ability was applied to obtain better results in subsequent studies. We got better resolution of face images by applying ESRGAN([4]) to hard and medium cases as shown on the right side of Fig 3.

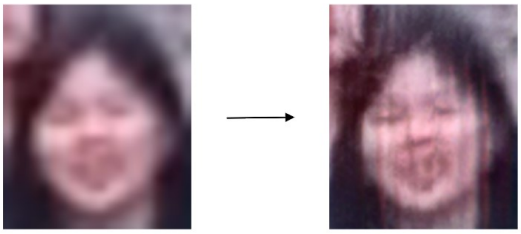


Fig. 3. Examples of images before and after ESRGAN application

3.5 Extraction of Gender Information as Facial Meta-Information

In post processing using detected crowd face images, increasing the accuracy of gender classification may help in detecting crowd anomaly behaviors. Therefore, we conducted a simple gender classification experiment to confirm an improvement of performance. In this case, Real-time Convolutional Neural Networks Gender Classification([5]) with good real-time performance was used in our experiment for application in video-surveillance.

4 Experiments

4.1 Face Detection

We conducted face detection using YOLO5Face Detector([3]) on the selected Crowd-11 dataset. To evaluate the accuracy of detection, if the detections over IOU(Intersection Over Union) > 0.5 is considered as the correct ones. Through this, we obtained about 80.6% of Recall and 71.1% of Accuracy. As the result of evaluation in Table 2, the performance of the easy case with the high resolution of the face images was the best, while the performance of the hard case with the low resolution was the worst.

Case	Recall	Accuracy
Easy Case	87.67%	84.41%
Medium Case	85.04%	80.32%
Hard Case	74.79%	59.99%
Total	80.6%	71.1%

Table 2. Performance evaluation by Face detection

4.2 Gender Classification

To improve the performance in gender classification, ESRGAN([4]) was applied to the face images of medium and hard cases, except for the easy case. The image after application has 4 times the resolution of the original face images. As the result of total evaluation in Table 3, the performance after application improved by 1%.

Case	Before ESRGAN	After ESRGAN
Easy Case	72.15%	-
Medium Case	68.86%	69.78%
Hard Case	65.24%	65.24%
Total	67.27%	68.20%

Table 3. Performance evaluation by gender classification before and after applying ESRGAN

5 Conclusions

In this paper, an accurate extraction scheme of facial meta-information from low resolution crowd images is proposed. In the proposed scheme, to accurately extract the facial meta-information even from the low-resolution crowd images, YOLO5Face detector which is well known as the good detector against small faces is used, and the

small images which are of lower than the threshold resolution determined by experiments are enhanced the quality by using ESRGAN. After that processing, the facial meta-information can be extracted more well. To confirm the feasibility of the proposed scheme, since gender of person can be regarded as very important facial meta-information, we compare the gender classification accuracies of using the proposed scheme and that of using the input crowd image itself. According to the experiment results, the proposed facial extraction scheme from crowd images using selective super resolution can improve the gender classification accuracy for crowd images than that for using the original crow images. This result shows the improvement of about 1% in gender classification. It is meaningful in that it tried to improve the performance in related studies by applying super resolution algorithm after face detection. Therefore, if an additional learning process in ESRGAN using train data optimize for face images is carried out, the possibility of development in future research is expected.

Acknowledgments. This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-00203, Development of 5G-based Predictive Visual Security Technology for Preemptive Threat Response). This work was also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1I1A1A01041815).

References

1. Crowd-11 : A Dataset for Fine Grained Crowd Behaviour Analysis, IEEE Conference on CVPRW (2017)
2. Yang, Shuo and Luo, Ping and Loy, Chen change and Tang, Xiaoou, WIDER FACE: A Face Detection Benchmark, IEEE conference on Computer Vision and Pattern Recognition (CVPR) (2016)
3. Delong Qi, Weijun tan, Qi Yao, Jingfeng Liu, YOLO5Face: Why Reinventing a Face Detector, arXiv preprint arXiv:2105.12931 (2021)
4. Xingtao Wang and others, ESRGAN: Enhanced Super-Resolution Generative Adversarial Network, arXiv preprint arXiv:1809.00219v2 (2018)
5. Octavio Arriaga and Matias Valdenegro-Toro and Paul Plöger, Real-time Convolutional Neural Networks for Emotion and Gender Classification. arXiv preprint arXiv:1710.07557 (2017)

AI-based Rehabilitation Support: A Concept Proposal

Jong-Hyeok Choi[1], Tserenpurev Chuluunsaikhan[2], Aziz Nasridinov[2,*]

[1] Bigdata Research Institute, Chungbuk National University, Cheongju, South Korea
[2] Department of Computer Science, Chungbuk National University, Cheongju, South Korea

{leopard, teo, aziz}@chungbuk.ac.kr

Abstract. Recently, the number of patients complaining of discomfort due to musculoskeletal disorders has been increasing. These patients can perform repetitive rehabilitation exercises, but their incorrect movements or poses worsen symptoms. In this short paper, we describe a concept of self-rehabilitation support using smartphones and artificial intelligence models.

Keywords: Rehabilitation, Pose estimation, Artificial Intelligence

1 Introduction

Recently, the number of patients complaining of discomfort due to musculoskeletal disorders caused by lack of exercise or excessive exercise has increased daily. These patients can return to society early through continuous and repetitive rehabilitation exercises. However, if the patient performs rehabilitation exercises alone, it may be less effective, or the symptoms may worsen due to incorrect movements or poses. Therefore, a visit to the hospital is necessary. However, considering COVID-19 or high hospital costs, it is not easy to receive continuous treatment. As a result, those called rehabilitation refugees have a negative impact on the national health insurance finances [1].

In this paper, to address this problem, we propose a concept that can support the self-rehabilitation exercise process of patients with mild symptoms based on pose estimation and pose analysis techniques using smartphones and the artificial intelligence (AI) model.

2 Concept Proposal

The process of the proposed method is as follows. When a patient with mild symptoms performs self-rehabilitation exercises, an AI-based pose estimation model first extracts the poses, and then analyzes these poses. In this case, the pose analysis method determines the correct pose by comparing it with pose information obtained

* Corresponding authors: Aziz Nasridinov (e-mail: aziz@chungbuk.ac.kr).

from the patient's rehabilitation treatment video. Here, if the patient takes a different pose, the pose analysis method judges it as an incorrect pose and provides feedback to the patient. Fig. 1 shows this overall process in the figure.

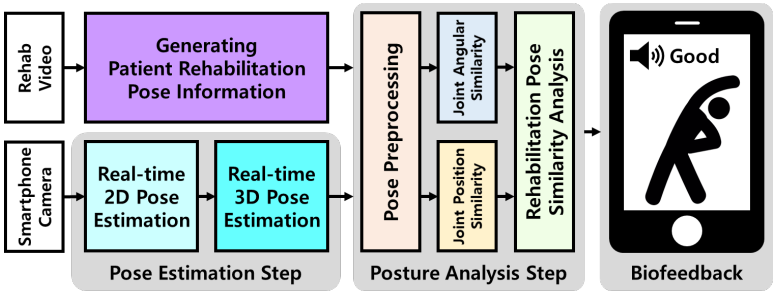


Fig. 1. The overall process of the proposed method.

There are several modules in the proposed conceptual framework, namely AI-based two-dimensional pose estimation model, pose analysis method, and audiovisual feedback method that can run on smartphones [2,3].

3 Conclusion

In this short paper, we have briefly described the self-rehabilitation support system using smartphones and artificial intelligence models. For this, we have explained the detailed flow of the proposed conceptual framework. In the future, we plan to build the proposed concept into an actual mobile application and implement it in real life to help patients with various muscular disorders.

Acknowledgments. This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1I1A1A01041815). This research was also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1I1A3042145).

References

1. When will the problem of rehabilitation refugees not receiving timely recovery treatment be resolved? Yonhap News Agency, https://www.yna.co.kr/view/AKR20170127005300017
2. Lee, J. J., Choi, J. H., Chuluunsaikhan, T., Nasridinov, A.: Pose evaluation for dance learning application using joint position and angular similarity. In: ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, pp. 67-70. ACM (2020)
3. Choi, J. H., Lee, J. J., Nasridinov, A.: Dance self-learning application and its dance pose evaluations. In: The 36th Annual ACM Symposium on Applied Computing, pp. 1037-1045. ACM (2021).

Through topic modeling based on core keyword extraction Newspaper Article Classification Model : Focusing on Korean Agricultural Newspaper Article Data

Mi-Seon Kim[1], Sang-hyun Choi [1*]

[1]Dept. Management Information System, Chungbuk National University, Cheongju, South Korea
sosgik96@naver.com, chois@cbnu.ac.kr

Abstract. In this paper, the article classification system based on the BERT model proposes a classification system that guarantees objectivity by analyzing keywords in the article as well as the title and content of the news article i[1]n text format and providing a document classification system. For unclassified documents in fields that do not have a common classification system, we propose a model that generates a middle classification by summarizing characteristics, automatically extracts document characteristics from newspaper articles, and uses them directly for document classification, and verifies its performance.

Keywords: BERT, Document Classification, Keyword Extraction, Text Mining, Unsupervised Learning,

1 Introduction

With the rapid development of information systems, it has become possible to systematically store and manage data related to news reports, such as news articles and multimedia data. As a result, domestic and foreign media companies have formed and operated a separate system for Internet article service, and are supporting users to use news services quickly, regardless of time and place, in addition to regular news coverage[1].

Typically, RSS (Really Simple Syndication), which is a service that allows users to see only the information they want to see a new article posted on a news service, or a tag that a reporter specifies a key keyword representing an article can be searched for as important information (tag), and the like. Due to the development of such information technology, news article pages of some portal sites provide a tag cloud function that previews popular or important content before users ask questions and provides information trends. However, most news article pages still list articles by date and field, and users can only read articles divided by category, but there is

* Corresponding Author

insufficient method to find information about other articles related to the article at a glance [2].

Therefore, the article classification system proposed in this paper analyzes news articles in text format, analyzes keywords in the articles, and provides a document classification system to implement a classification system that guarantees objectivity. Domestic news portals do not have a classification system that mainly deals with “agriculture”, so it is difficult to obtain specific information related to agriculture. is applied to implement a service to increase the usability of each exposure to the news portal. Through this, the overall limitations of the news classification information system to which key keywords are applied and future research directions are presented.

2 Related Research

2.1 Extraction of key keywords based on TF-IDF

Basically, the TF-IDF weight is a model that can determine the weight of a word within a document. The TF value reflects the importance of a word in the entire set of documents within a document, and the IDF value corrects the error of the TF value. Aspect plays an important role. In this paper, we take a similar approach to the TF-IDF weight model by measuring the relevance between keywords in an article, define them as core keywords, and verify by comparing those with and without core keywords.

2.2 Korean BERT Technical Document Classification Model

In a news article, the keywords appearing at the beginning of the document can be considered to be more relevant to the article, so a weight can be imposed on the position of the sentence in which the keywords appear in the document [2]. Therefore, the model used as an experiment in this thesis is a BERT-base Multilingual Cased model [3], and the machine reading comprehension task experiment can be performed only with fine-tuning using the pre-trained BERT Model in 103 languages including Korean. It is a model using the encoder of Transformer [4] with bidirectionality that can solve the long-term dependency problem and can learn the position of words in a sentence to achieve better performance.

3 Proposed Methods

3.1 Document Classification Process

The research model process of this paper is as follows.

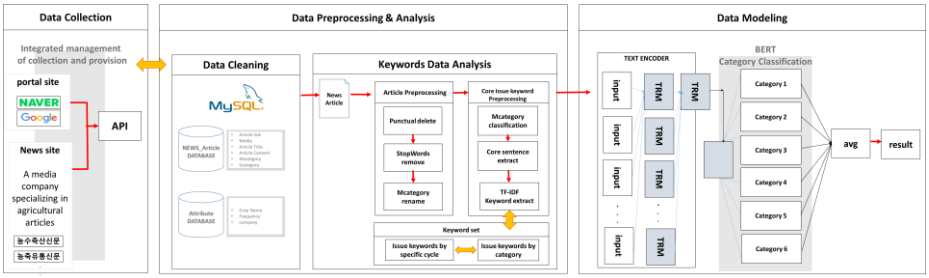


Fig. 1. Structure of data learning process

Data Collection. The data used in this paper is news text data related to crops through searches such as Naver/Google and selected 13 agricultural news portals classified into categories. A total of 126,945 articles were published as news from January 1, 2019 to June 30, 2021. To prevent overfitting, the model was trained by dividing the training data and the validation data in a ratio of 8:2 through hierarchical k-fold cross-validation.

Since each media company is classified into 56 sub-categories and 339 sub-categories divided by classification system, in this paper, the characteristics of each newspaper article were classified and composed of characteristics representing agricultural articles, and a classification system of 6 medium-class newspaper articles was constructed.

Table 1. Data classification system and label formation of experimental dataset

Category	Attribute	Label
Cooperative	Agricultural Cooperatives/	0
Crop livestock environment	Fertilizer/Pesticide/Feed/Drug	1
Distribution of crops and livestock	Local Wholesale Market/Consumer Consumer	2
Rural society	Rural society/Rural development/	3
Agricultural policy and economy	organizations/policy/	4
agricultural science	Forest Technology/Smart Farm/	5

Proposal Model. The baseline was set by performing fine-tuning with the BERT basic model. Each newspaper article has less than 1000 articles in the category to which it belongs, which corresponds to a class imbalance. The given problem is to classify whether an individual newspaper article belongs to each category out of a total of six subcategories. Therefore, it was modeled by generating a classifier for each subclass. Therefore, whether or not it belongs to individual intermediate classification becomes a binary classification problem, and the average loss was calculated for the BERT-based classification model for all crop article classification, and the loss value of individual documents was calculated and the parameters were updated.

As input data, each newspaper article is a data set of article id, publication date, article title, and article content, and all are listed in order and integrated. Each newspaper was treated as a single sentence. After making a newspaper article into one sentence, 10 words representing the article were designated as key keyword fields through TF-IDF-based key keyword extraction and applied as input factors at once.

4 Conclusion

The performance of the trained classification model was evaluated by calculating the F-score, precision, and recall for each document on the test data and averaging over the entire document. In the case of test data extracted with key keywords and applied as input data, 59.6% of the test data entered only as article content was 56.3%, suggesting that the model trained with key keywords showed predictive power above a certain level.

5 Acknowledgements

This paper has a limitation in that all information on a given technical document is processed as a single sentence. Future research tasks may include advanced research to improve the performance of classification models by individually processing and learning variables such as technology name, article content, and key keywords. In addition, if the model is trained by securing more training data, I think that a model with classification performance that is actually close to understanding the meaning of a human document can be obtained.

References

1. J. Devlinalet. (2018). BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding., arxiv.org/abs/1810.04805.
2. Ji-Hye KimJang, Hong-Jun Yoon, Han-Jun KimJae-Young. (2010). A related search technique for news articles using keyword relevance. Proceedings of the Korean Society of Information Science and Technology. (37(1C),54). the Korean Society of Information Science and Technology.

3. Kim Do-wooMyeong-wanKoo. (2016). A study on CNN-based Korean newspaper article classification using Doc2Vec. ″The Korean Language Information Society Conference″ (67-71).
4. LeeH., Park, Y.-H., & Lee, K. JG. (2020). Building a Korean Text Summarization Dataset Using News Articles of Social Media. KIPS Transactions on Software and Data Engineering. “KIPS Transactions on Software and Data Engineering,.”

Analysis of the Effect between Weather Conditions and Odor-Causing Substances using Machine Learning

Yeon-Ju Yu[1], Sang-Don Lee[2], Woo-Seok Choi[1,]
Ki-Yong Park[3], Sang-hyun Choi[2*]

[1]Dept. Bigdata, Chungbuk National University, Cheongju, South Korea
[2]Dept. Management Information System, Chungbuk National University, Cheongju, South Korea,
[3]Dept. Urban Engineering, Chungbuk National University, Cheongju, South Korea[1]
[1]{play981203,cdt3017}@gmail.com, [2]{chois,leesd}@cbnu.ac.kr, [3]pky3489@chungbuk.ac.kr

Abstract. As the demand for a high-quality residential environment for a more comfortable life increases along with the improvement of the economic level, interest in environmental issues is also increasing, and among them, environmental problems caused by bad odors are attracting attention as an important part. The purpose of this study was to find a significant relationship between weather conditions and odor-causing substances through the LightGBM regression model, a machine learning.

Keywords : Machine Learning, LightGBM, SHAP, Feature Importance

1 Introduction

As the demand for a high-quality residential environment for a more comfortable life increases along with the improvement of the economic level, interest in environmental issues is also increasing, and among them, environmental problems caused by bad odors are attracting attention as an important part[1]. In addition, the frequency of occurrence of odors with irregular and intermittent characteristics is a major cause of complaints[2]. In this study, the influence factors were analyzed by applying the machine learning technique to investigate the influence relationship between odor-causing substances and weather conditions. In particular, odor-causing substances, trimethylamine, a substance that causes fishy smell, was targeted. For the weather conditions, the weather information data of Cheongju-city was used.

2 Used Model

LightGBM is an ensemble model that uses boosting algorithms and has high accuracy and fast learning speed, which has been widely used in recent years. SHAP

[2] Corresponding author

is a new approach to explore the relationship between the input variable and the model's result value by calculating the SHAP value for each input variable. Therefore, SHAP can be used to interpret high importance input variables and how their predicted values change when the values of the input variables change[3].

3 Analysis

3.1 Collection data

In the case of weather information data, wind speed, precipitation, temperature, sunlight, solar radiation, and humidity data provided by Cheongju-city were used. All weather condition data were collected based on the daily average. As for the pollutant data, the observation concentration data of trimethylamine, a odor-causing substance, which is sensor data installed in Cheongju-city, was used. The data collection period is a total of two years from January 1, 2019 to December 31, 2020, and if there is a missing value in the dataset, the daily average value of the variable was obtained and replaced.

3.2 Research results

The purpose of this study is to create a LightGBM regression model to find variables that affect the odor-causing substance concentration. Based on this, we learned the LightGBM regression model using weather condition data such as precipitation, temperature, humidity, sunlight, solar radiation, and wind speed as independent variables. The dependent variable was set as the concentration data of trimethylamine.

To improve the learning performance of the model, the data were divided into training sets, validation sets, and test sets at an appropriate rate using Scikit-learn's train test split function.

The conditions for setting the hyperparameters of the model are learning rate: 0.1, max step: 3, boosting: gbdt, goal: regression, etc. Under these conditions, functional importance and SHAP values were derived after generating the LightGBM regression model.

Feature importance does not count negative influences. In addition, if multicollinearity exists between input variables, there is a problem that the result may be distorted, so the SHAP value is used to compensate for this. SHAP values are listed in ascending order of absolute value, meaning that the variable located at the top has a greater influence on the value of the dependent variable.

As a result of data analysis, it was found that the factor that had the greatest influence on the concentration of trimethylamine, a substance that causes odor, was temperature. Looking at the table below, the lowest temperature had the highest value in both feature importance and SHAP values. This means that the lowest temperature has the greatest effect on the trimethylamine concentration change. Conversely, precipitation and sum of sunlight rates did not significantly affect the concentration change.

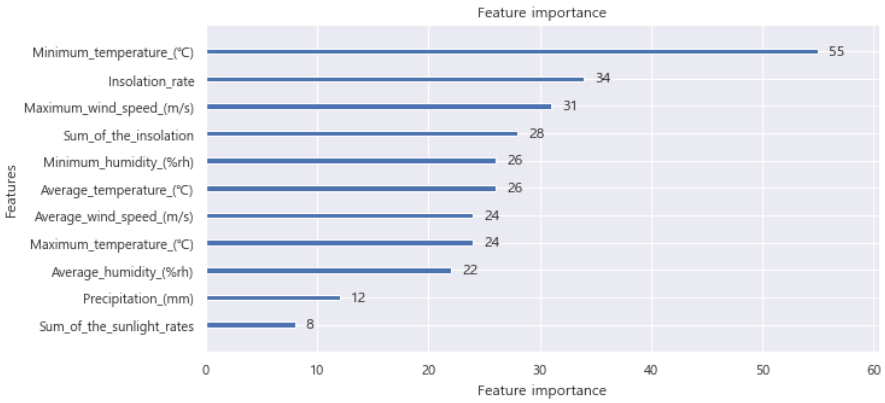


Fig. 1. Feature Importance

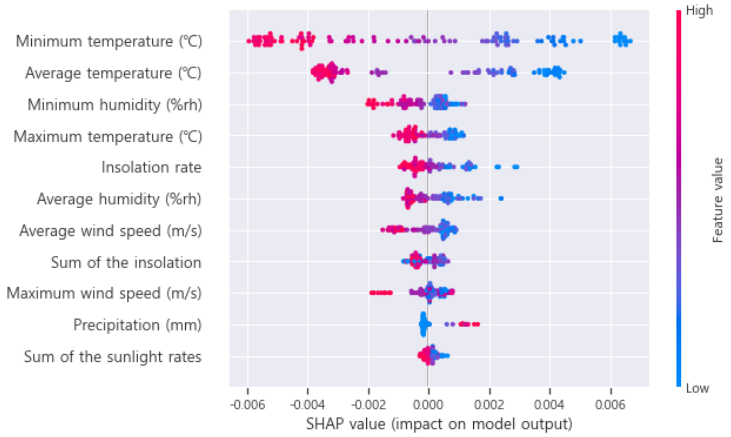


Fig. 2. SHAP value

4 Conclusion

Through this study, it was confirmed which weather conditions had a significant effect on pollutant concentration through SHAP values that calculate both positive and negative effects. In particular, it is meaningful in that it is possible to confirm the negative effects that are not well revealed on the surface based on the SHAP value. In this study, only weather condition data were used as independent variables, but more important associations are expected to be found by adding and analyzing more diverse regional environmental variables in the future.

Reference

1. Sang-woo Lee: Prediction of the effect of landfill odor on surrounding air quality - Case study of Cheongju Petition Metropolitan Landfill Site - Environmental Impact Assessment, vol. 21(5), pp. 695-705 (2012)
2. Kwang-Hee Lim: CALPUFF modeling of odors and suspended dust around the barn:centered on the temple.Chemical Engineering, vol. 57(1), pp. 90-104 (2019)
3. Sung-Woo Park, Jae-Wook Moon, Seung-Won Jung, Seung-Min Jung, In-Jun Hwang: Predicting the occurrence of explainable influenza based on SHAP using LightGBM. Collection of papers presented by the Korean Information Science Association, pp. 666-668 (2020)

Three-Level Classification of Crop Leaf Images Using Deep Learning

Borin Min[1], Gae-Ae Ryu, HyungChul Rah, Kwan-Hee Yoo[1]

[1] Dept. Of Computer Science, Chungbuk National University, Cheongju, South Korea
Department of Management Information Systems, Chungbuk National University, Cheongju 28644, Korea

{minborin, garyu, khyoo}@chungbuk.ac.kr
hrah@cbnu.ac.kr

Abstract. Classification of ground-captured plants provides a great choice to directly predict names and diseases from an input image of a crop. To solve the challenge causes, we try to preprocess the input crop large image to segment leaf instance object and background. We propose a deep learning approach using a convolutional neural network (CNN) to provide three-level classification: crop names, diseases, and diseases kinds. We applied the proposed method to 4 crop species and 11 disease types, which the results of SME (mean squared error) show 0.1762.

Keywords: Crop classification, three-level classification, ground-capture crop classification.

1 Introduction

Machine learning (ML) has evolved rapidly in recent decades, significantly since computer vision has incredibly outperformed in visual classification. Remarkably, crop species recognition and diseased detection are achieved by researchers [1–3]. In this work, we aim to identify three levels of an input crop which are crop classification: to distinguish four crop species—corn, cucumber, pepper, and strawberry using leaf; crop-diseased detection: to detect between normal (uninfected) and abnormal (infected disease) of crop species; crop-diseased classification: to distinguish disease species of crops—the ground-captured of crop plant in a field that is a major challenge in classification. In addition, there are disturbances—background such as grass and soil that come along with input could be covered area of interesting, which is the main impact to algorithm training cause confusion of prediction. Meanwhile, large-scale resolution of plant-captured could lead the algorithm to be tough to learn , optimize training accuracy, and be expensive, especially for small volume features such as disease. To respond to the addressed obstacle as mentioned, we proposed a deep learning approach called three-level classification, constructed by two phases—preprocessing data and classification.

Figure 1.1 representative twelve diseased from four crop kinds.

Figure 1.1 represents disease kinds, which are used for classification and detection. We have diseased corn (common rust, leaf streak, gray leaf, bacterial spot), cucumber (anthracnose, downy), pepper (mosaic, gangrene, bacterial spot), and strawberry (green blight, powdery mildew, spotted mite).

2 Related Work

Many researchers published similar studies to distinguish crop species, diseased leaf detection, and diseased recognition. Mohanty *et al.* [4] proposed a deep learning method to detect 38 disease classes of 26 species, input images were applied grayscale and segmentation, and then passed it to deep learning algorithms AlexNet[5] and GoogLeNet[6] for classification. Zhang *et al*. [7] proposed a deep learning method to specifically identify disease of cucumber leaf under field condition based on small sample size and deep convolutional neural network, GANs [8], a data generator deep learning algorithm were used to generate more sample data. Meanwhile, GrabCut is an image segmentation algorithm that could perform perfectly, which requires a bounding box manually. Be remarkable, and input images were used for model training are completely separated leaves. There was no complicated background such as grass and soil present in leaf captured data.

3 Proposed Method

In this approach, we design a specific architecture to narrow classification problems in crop species. The architecture consists of three-level classification models, which each level has its convolutional neural network (CNN), figure 3.1. Crop classification (L1) is the first classification with the Mask-RCNN model [9] to classify crop spices and instance segmentation. Based on the coordinated position of rectangle box (x_1, x_2, y1, y_2) and coordination of mask which can distribute instance from crop spices.

Disease detection (L2) and disease classification (L3) have disease detection and disease classification nodes. In each node contains a convolutional neural network model [6,10]. L2 performs binary classification among two classes: normal and diseased crops. Followed by L3, if there is a disease in L2, it will distinguish disease kinds.

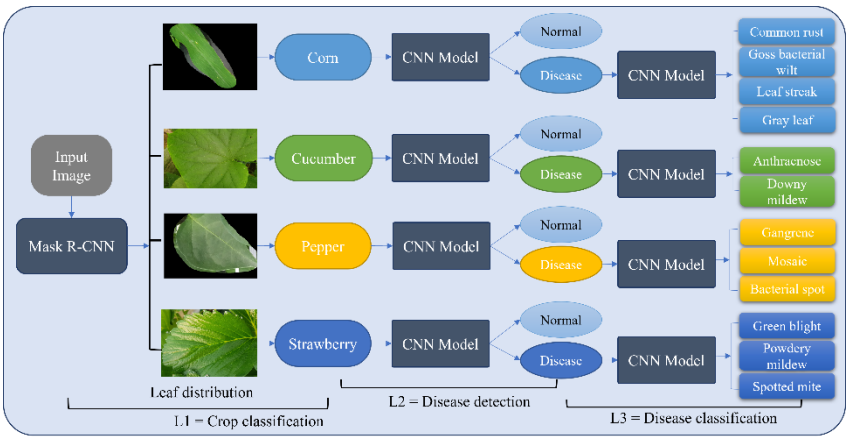


Figure 3.1 Three-level classification designs architecture.

Figure 3.1 shows that L1 attempt to predict C_i (crop classes *i*), the output of L1 is distributed before sending to L2. Which C_i represents for node in L2, each node C_i is responsible for classifying N and D (N= normal, D= disease). If the output of C_i is D, thus D is created as the node to classify d_i (d_i = represent of diseases), otherwise known as ending. CNN model is used to be continued train for L2 and L3 task. There are effectively designed CNN architecture like ResNet [10] and InceptionNet [6] that could be trained deeper layers, and it is acknowledged as the top-1 error rate in ILSVRC'15 for ResNet, while InceptionNet received a top-1 error rate in ILSVRC'14.

4 Experimental Results

We partially arrange the dataset into three parts: L1, L2, and L3 dataset. The L1 dataset represents L1 classification which annotates coordinated position with the label to crop images, which annotated corn, cucumber, pepper, and strawberry are 104, 102, 108, and 99 annotations. L2 and L3 datasets received images from the L1 classification's output and were grouped as a dataset by the node. A node of the L2 dataset represents a crop that distributes images by normal and disease, whereas the L3 dataset distributes images by disease name.

In L1, we used transfer learning of the pre-trained zoo model [11] based on Inception-ResNetV2 as the final classifier with 500 epochs. The final classification's loss is 0.2076. We also applied image augmentation and compared image generator using

GANs-like is Image-to-Image Translation with convolutional Adversarial Network and image rotation.

Tables 2—3. are the results of disease detection and classification using deep learning methods. We experienced changing training parameters to avoid overfitting, mainly applied image generators. Pix2pix [12], an image generator method, is used to synthesize images.

Table 1. Level 1: crop classification results by Mask R-CNN compared pre-trained model and from scratch learning.

Model	Crop Classification (L1)	Pre-Trained Weight	Classification Loss
Mask R-CNN	Crop Classification	Zoo Model	0.1047
		From scratch	0.1611

Table 2. Level 2: disease detection results by methods: InpcetionV3 and ResNet101-V2, and applied image generators.

Model	Image Generator	Accuracy (%)			
		Corn	Cucumber	Pepper	Strawberry
InceptionV3	Augmentation	97.49	94.57	88.18	93.88
InceptionV3	Augmentation + Pix2pix	96.40	98.13	95.47	97.47
ResNet101-V2	Augmentation	99.38	98.93	96.01	97.17
ResNet101-V2	Augmentation + Pix2pix	71.16	95.87	94.27	98.80

Table 3. Level 3: disease classification results by methods: InpcetionV3 and ResNet101-V2, and applied image generators.

Model	Image Generator	Accuracy (%)				MSE
		Corn	Cucumber	Pepper	Strawberry	
InceptionV3	Augmentation	97.49	94.57	88.18	93.88	0.6457
ResNet101-V2	Augmentation	99.13	98.50	73.96	97.69	0.1762

5 Conclusion and Future Work

In this paper, we proposed a deep learning method for the three-classification type crop: crop classification, diseased detection, and crop diseased classification. Apart from a large-scale input image, we first segmented the crop leaf image and classified crops (L1), then used the output from L2 as the input for the crop normal and disease detection (L2), followed by the disease classification (L3) to identify disease kinds. According to the various learning experience, we found ResNet101-v2 accomplished the highest result of MSE 0.1762. There is a drawback when designed to have multiple

classification models, which requires much training. However, a narrow class of learning could help model more convenience to the identification.

In the future, we would like to enhance how we could manage to build a single model for hierarchical classification and consume proper time.

Acknowledgment. This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry (IPET), funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (319003-01) and by the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2020-0-01462) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation).

References

1. Sharma, P., Hans, P., Gupta, S.C.: Classification Of Plant Leaf Diseases Using Machine Learning And Image Preprocessing Techniques. In: 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence). pp. 480–484. IEEE, Noida, India (2020). https://doi.org/10.1109/Confluence47617.2020.9057889.
2. Ajra, H., Nahar, Mst.K., Sarkar, L., Islam, Md.S.: Disease Detection of Plant Leaf using Image Processing and CNN with Preventive Measures. In: 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE). pp. 1–6. IEEE, Bangladesh (2020). https://doi.org/10.1109/ETCCE51779.2020.9350890.
3. Codizar, A.L., Solano, G.: Plant leaf recognition by venation and shape using artificial neural networks. In: 2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA). pp. 1–4. IEEE, Chalkidiki, Greece (2016). https://doi.org/10.1109/IISA.2016.7785361.
4. Mohanty, S.P., Hughes, D.P., Salathé, M.: Using Deep Learning for Image-Based Plant Disease Detection. Front. Plant Sci. 7, 1419 (2016). https://doi.org/10.3389/fpls.2016.01419.
5. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Commun. ACM. 60, 84–90 (2017). https://doi.org/10.1145/3065386.
6. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going Deeper with Convolutions. ArXiv14094842 Cs. (2014).
7. Zhang, J., Rao, Y., Man, C., Jiang, Z., Li, S.: Identification of cucumber leaf diseases using deep learning and small sample size for agricultural Internet of Things. Int. J. Distrib. Sens. Netw. 17, 155014772110074 (2021). https://doi.org/10.1177/15501477211007407.
8. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks. ArXiv14062661 Cs Stat. (2014).
9. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. ArXiv170306870 Cs. (2018).
10. Choi, Y., Kwun, H., Kim, D., Lee, E., Bae, H.: Residual Life Prediction for Induction Furnace by Sequential Encoder with s-Convolutional LSTM. Processes. 9, 1121 (2021). https://doi.org/10.3390/pr9071121.
11. http://download.tensorflow.org/models/object_detection/tf2/20200711/mask_rcnn_inception_resnet_v2_1024x1024_coco17_gpu-8.tar.gz.

12. Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-Image Translation with Conditional Adversarial Networks. ArXiv161107004 Cs. (2018).

Forecast of carry-in volumes of napa cabbage and Korean radish in market through statistical estimation of production areas

Min-Su Kim[1], Dong Jin[1], Jimin Lee[1], Helin Yin[1],
Yeong Hyeon Gu[1], Seong Joon Yoo[1,*]
,

[1] Department of Computer Science and Engineering,
Department of Convergence Engineering for Intelligent Drone, Sejong University, Korea
rlaalstn1504@naver.com, kimdoubled@gmail.com, leejeemin0608@gmail.com, yin-helin0608@gmail.com, yhgu@sejong.ac.kr, sjyoo@sejong.ac.kr

Abstract. The unstable supply and demand of agricultural products leads immediately to the instability of prices. If the carry-in volumes of korean radish and napa cabbage in market can appropriately predict the supply and demand as a critical indicator showing supply, it can be a great help in controlling the supply-demand status. In order to forecast the carry-in volumes in market, it is important to adequately reflect the climates of essential production areas. However, the main production areas of agricultural products are continuously altering due to climage changes. To overcome the problems resulting from climate changes, this study applied dynamic production area method of using meteorological data from three areas that were found to generate the highest monthly carry-in volume in the year immediately preceding the time point of the forecast. As a result of using fixed production areas and dynamic production areas respectively, the MAPE performance was found to be 2% higher on average when using dynamic production areas.

Keywords: Estimation of production areas, Output forecast, XGBoost

1 Introduction

Napa cabbage and Korean radish are important crops which account for a great part of farming and dietary life. Since these two crops have poor preservability, their supply-demand status is heavily dependent on the output. As the occurrence of abnormal climate changes is more frequent recently, there is a growing concern for the instability of supply and demand. Basically, the outputs of agricultural products are profoundly affected by the weather conditions. In this regard, the forecast of the outputs of agricultural products based on meteorological data could be a great help in controlling supply and demand. However, the output data of napa cabbage and

* corresponding author

Korean radish are provided with respect to season-specific cropping patterns once a year, making it difficult to identify the production time and area. On the other hand, the carry-in volumes of Korean radish and napa cabbage are considered appropriate as an indicator of their supply-demand status. This is because their production areas and carry-in volumes in market can be identified systematically according to date, and especially Korean radish and napa cabbage are quickly carried into market after harvest, due to poor preservability.

Since both napa cabbage and Korean radish are cultivated in various areas throughout four seasons, their supply-demand status in market is determined by the interactions with different production areas. Therefore, the supply-demand status of crops can be precisely predicted when the yields from different production areas are simultaneously considered.

However, the main production areas of crops are changing slightly every year due to climate changes and other factors. Accordingly, a complicated process that requires to keep track of production areas every year is necessary to accurately assess the relationship between meteorological data and output of agricultural products. Many previous studies involved selecting specific production areas and then limiting the subjects only to the varieties cultivated in these areas. To overcome the problems resulting from climate changes, this study applies dynamic production area method and selects the main production areas of crops. The dynamic production area method is a method of determining which area's meteorological data should be used in order to predict the carry-in volumes of crops. In other words, this method selects three areas that generate the highest monthly output on a monthly basis in the year immediately preceding the time point of the forecast. Respective forecast results are compared using MAPE values to verify whether forecast performance has been improved when using dynamic production areas compared to using fixed production areas.

The results of the analysis revealed that the performance was improved by approximately 2% on average when predicting the monthly carry-in volumes of crops in Garak Market [1] by using dynamic production areas compared to using fixed production areas. Consequently, the results of this study is expected to be helpful in predicting the supply and demand of agricultural products more accurately.

2 Related Works

In this chapter, previous studies to estimate yields of crops, such as output, yield per unit area, or carry-in volume in markets, based on weather conditions were examined.

Lee et al. [2] extracted the weather factors affecting crops per growth stage by studying the relationship between weather conditions and yield per unit area as to the highland napa cabbage crops cultivated in high lands of Gangwon-do Province and

then presented a forecast on future yield per unit area of highland napa cabbage considering the climate change issue.

Kwon et al. [3] utilized the meteorological information collected using the Ubiquitous Sensor Network (USN) to forecast the yield per unit area of highland napa cabbage based on weather conditions. Although precise meteorological data were used, research was conducted only on the highland napa cabbage cultivated in limited areas during a specific period. Therefore, the study has its limitations in that it is difficult to apply it in predicting supply and demand on the basis of national outputs.

Kim et al. [4] presented a forecast on the monthly carry-in volume of napa cabbage in Garak Market to predict supply and demand. In order to utilize the meteorological data, the different production areas of napa cabbage according to each season were designated and used differently per month, but there was a limitation in that the characteristics of the production areas altering slightly every year due to climate changes and other factors were not reflected.

3 Data & Algorithm

3.1 Data Preprocessing

To predict the monthly carry-in volumes of Korean radish and napa cabbage in Garak Market, the meteorological data of main production areas and the previous data of carry-in volumes in this market were used. For napa cabbage, the data of import and export from the Korea Customs Service were additionally used. The source and period of the data used are as shown in Table 1.

Table 1. Used data and source

	Data from Garak Market	**Meteorological data**	**Data of kimchi import and export**
Source	aT	Korea Meteorological Administration	Korea Customs Service
Period	January 2012 ~ November 2020	December 2014 ~ November 2020	November 2014 ~ November 2020
Cycle	day	day	month

Both meteorological data and carry-in volume data were processed into monthly data because they were made on a daily basis. In this case, the meteorological data

were processed as the mean monthly data, while the data of carry-in volumes and the data of kimchi import and export volumes were processed as the monthly output.

For the monthly output of each crop in Garak Market, data before 11 months and 23 months were generated as a derived variable. For napa cabbage, the data of kimchi import and export volumes from the Korea Customs Service were also used as an input variable. The variables used for the crop-specific forecast of carry-in volumes are tabulated as shown in Table 2.

Following the pre-treatment process, a total of 72 cases of data from December 2014 to November 2020 were generated for both crops. To assess the performance, the entire data were divided into training data (57 cases) and test data (15 cases) in the ratio of 8:2.

3.2 Method of Dynamic Production Area

In order to extract main production areas based on statistics, the settlement data from Garak Market were used. Since Garak Market, one of the largest markets in Korea, is where the crops are carried from and distributed to all over the country, it was determined that statistical production areas could be estimated.

The method to estimate statistical production areas involves extracting three areas with the highest monthly output by referring to the information of the areas producing the crops traded in Garak Market. In Garak Market, there are varieties not only produced domestically but also produced and processed overseas. Therefore, imported, salted, and processed varieties were excluded using the classification names of varieties in the settlement data from this market. Then by dividing the data on a monthly basis and ranking production areas in the order of decreasing monthly output, a total of three production areas with the highest monthly output were extracted.

In this way, it was possible to derive the production areas generating the highest output per month by using the previous data one year ago because the current settlement data for carry-in volumes did not exist. For example, the production areas selected in December last year could be utilized to select the production areas for December this year in order to reflect the most recent essential production areas.

To compare the performance of training and forecast using dynamic production areas with the pre-existing method, the comparators were selected based on the production area information released by the Korea Rural Development Administration, as presented in Table 3.

Table 2. Variables used for the crop-specific forecast of carry-in volumes

Crop	Variable category	Used variable
Korean Radish, Napa cabbage	Meteorological variables	Mean temperature, maximum temperature, minimum temperature, mean relative humidity, minimum relative humidity, precipitation, mean wind velocity, accumulated sunshine duration, typhoon watch, and typhoon warning
	Carry-in volume variables	Carry-in volume of napa cabbage in Garak Market this month, 1 year ago, 11 months ago, and 23 months ago
Napa cabbage	Other variables	Kimchi import and export volumes

Table 3. List of crop-specific fixed production areas used on a monthly basis

Month	Production area(South Korea)	Crop
1, 2, 3	Haenam, Jindo	Napa cabbage
	Seosan, Muan, Jeju	Korean radish
4, 5, 6	Seosan, Muan, Mungyeong	Napa cabbage
	Seosan, Muan, Mungyeong	Korean radish
7, 8, 9	Taebaek, Pyeongchang, Jeongseon	Napa cabbage
	Taebaek, Pyeongchang, Jeongseon	Korean radish
10, 11, 12	Dangjin, Hongseong, Gochang	Napa cabbage
	Dangjin, Hongseong, Asan	Korean radish

3.3 Machine Learning Algorithm

In this study, XGBoost algorithm was used to forecast the monthly carry-in volumes of Korean radish and napa cabbage in Garak Market, by referring to the outcome of the XGBoost algorithm used by Kim et al. [4], which was found to exhibit the most superior performance when comparing multiple algorithms with respect to the forecast of carry-in volumes of napa cabbage.

XGBoost is a tree-based ensemble model and can be utilized for both classification and regression problems, and provides excellent performance based on various optimization options. It is also an algorithm widely used in various fields due to its fast learning speed through parallel processing.

There are various hyperparameters in XGBoost, such as max_depth, learning_late, ubsample, and min_child_weight. In this study, various hyperparameters were optimized through grid search.

4 Experiments and Results

In this chapter, the performance between the experiment using meteorological data from selected dynamic production areas and the experiment using meteorological data from fixed production areas was compared to forecast the monthly carry-in volumes of Korean radish and napa cabbage in Garak Market.

4.1 Evaluation Measures

The test results were measured through the Mean Absolute Percentage Error (MAPE), and the MAPE is as shown in Equation 1.

$$MAPE = \frac{100}{n}\sum\frac{|Precision-Actual|}{Actual} \quad (1)$$

4.2 Results

It was found that the performance was improved by approximately 2% on average when tested using dynamic production areas compared to fixed production areas. The crop-specific detailed performances for each test set are presented in Table 4, and Figure 1 shows the status of estimated and actual values.

Table 4. Results of crop-specific tests of used production areas

Crop	Used production area	MAPE
Napa cabbage	Fixed production area	11.40%
	Dynamic production area	8.54%
Korean radish	Fixed production area	10.46%
	Dynamic production area	9.30%

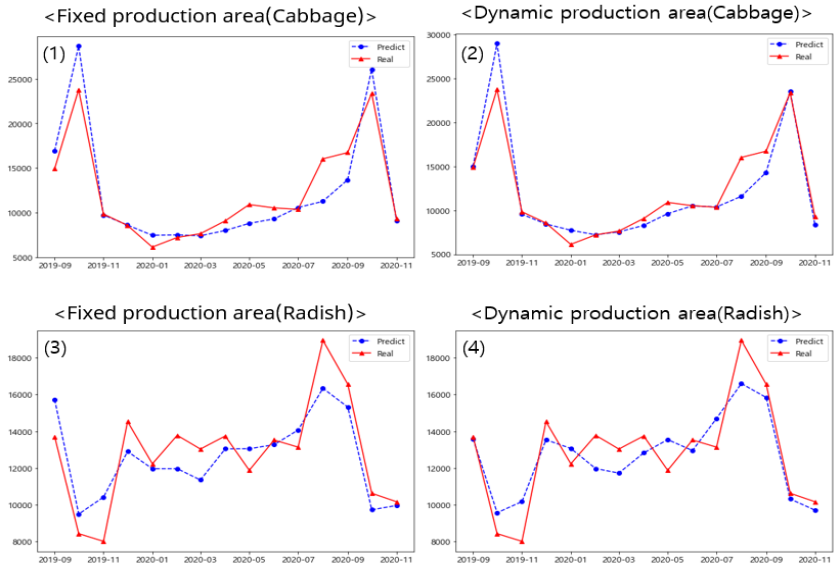


Fig 1. Comparison between actual and predicted value

5 Conclusions

In this study, the method of using three areas with the highest carry-in volume in the previous year as the main production areas was proposed in order to keep track of the slightly changing production areas, when estimating the carry-in volumes in markets. As a result, greater performance could be achieved when using the meteorological data from the main production areas selected in the way proposed in this study than using the meteorological data from the fixed production areas.

Further study is expected to derive a greater performance through a comparative analysis between multiple algorithms.

ACKNOWLEDGMENT

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2021-0-00755 / 20210007550012002, Dark data analysis technology for data scale and accuracy improvement) and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2019-0-00136, Development of AI-Convergence Technologies for Smart City Industry Productivity Innovation)

References

1. Garak market: https://www.garak.co.kr/main/main.do
2. Lee, S., & Heo, I. (2018). Impact of climate on yield of highland Chinese cabbage in Gangwon province, South Korea.
3. Kwon, T., Kim, R. Y., & Yoon, S. (2019). The Effect of Highland Weather and Soil Information on the Prediction of Chinese Cabbage Weight. Journal of Environmental Science International, 28(8), 701-707.
4. Kim, M. S, Jin, D., Lee, J. M, YIN, H., Gu, Y. H., Yoo, S. J. (2021). A Proposal for a Predictive Model of Cabbage Trading Volumes Using Machine learning. Korea Computer Congress 2021
5. Kang, T. H. (2007). An Analysis on the Relationship between Prices and Trading Volumes of Agricultural Products. The Korean Journal of Agricultural Economics.
6. Park, J. Y., & Park, Y. G. (2013). The development of Chinese cabbage and radish forecast models. M125. Korea Rural Economic Institute.
7. Kim, Y. S., Shim, K. M., Jung, M. P., & Choi, I. T. (2015). Study on meteorological factors affecting estimation of Chinese cabbage yield. Journal of Climate Change Research, 6, 277-281.
8. Suh, K., & Lee, J. J. (2004). Analysis of the Effect of Shipping Control depending on the Limited Storage Life of Agricultural Products. Journal of Korean Society of Rural Planning, 10(3), 53-58.
9. Ahn, B. I., Kim, S. Y., & Kim, B. Y. (2002). Trends and Causes of Spice Vegitable Price Instability in Korea. Korea Journal of Agricultural Economics Research, 43(1), 103-121.
10. Lim, C. H., Kim, G. S., Lee, E. J., Heo, S., Kim, T., Kim, Y. S., & Lee, W. K. (2016). Development on crop yield forecasting model for major vegetable crops using meteorological information of main production area. Journal of Climate Change Research, 7(2), 193-203.
11. Nam, K. H., & Choe, Y. C. (2017). An Dynamic Analysis on the Relationship among Prices, Trading Volumes, Import Volumes and Demand Using VAR-Focused on Cabbage, Onions, and Garlic. Journal of Agricultural Extension & Community Development, 24(1), 9-19.
12. Lee, D. S., & Kim, I. S. (2018). An Ex-post Analysis on the Price Stabilization Effects of Disposal Policy for Chinese Cabbage. The Korean Journal of Agricultural Economics.

Forecasting of irrigation by period using CNN - LSTM

Taeyang Kim[1],Wanhyun Cho[2], Myunghwan Na[3],

[1,2,3]Dept. Mathematics and Statistics, College of Natural Science, Chonnam National University,
Yongju-ro 77, Buk-ku, Gwangju, Korea
[1]ty7613@naver.com,[2]whcho@chonnam.ac.kr, [3]nmh@chonnam.ac.kr

Abstract. Soil humidity is one of the main factors influencing the growth of crops, and designating the irrigation period by accurately predicting the soil humidity while the crops are growing has a great influence on the increase in agricultural production. However, for accurate prediction of soil humidity, research on the relationship between soil humidity and environmental data is needed. In this study, onion data from Jeonnam and Gyeongnam regions provided by the Agricultural Technology Institute were used, and the observation period consists of hourly data from December 31, 2019 to June 5, 2021. It predicts soil humidity from May 1, 2021 to June 5, 2021 by learning data from December 31, 2019 to April 30, 2021. As a prediction model, a long short- and long-term memory (LSTM) and a CNN-LSTM are used. CNN-LSTM is a deep learning technique that shows excellent performance in large-capacity data processing, and aims to improve the accuracy of polarity analysis using a combination model of CNN(Convolution neural network) and LSTM(long-term memory) during deep learning. Through this, since there are several parameters, the relationship between the value and precision is considered to find the optimal combination and attempt to improve predictive performance such as accuracy. As a result, it was confirmed that the predictive power of the CNN-LSTM model was improved more than that of the LSTM model. It is expected that predictive power can be improved by using more diverse deep learning techniques as future research tasks.

Keywords:CNN-LSTM, Soil humidity, Time series data

1. Introduction

Agriculture includes a variety of activities. For example, it refers to the production or cultivation of crops and various activities related to the quality control of agricultural society. Among them, crop production and cultivation play an important role in our lives. As a result, research on crop production and cultivation is being actively conducted. The approach to crops is divided into two perspectives: production and cultivation factors. The analysis method varies considerably depending on where you focus. First of all, if you want to know under what good conditions crop cultivation is actively carried out, regression analysis using time series of various models is mainly conducted. Existing approaches include NWP, MA, ARIMA, ARIMA-X, and the like. However, if you want to predict the exact yield or output, you can use the recently

developed deep learning method. As a representative method, a method that can contain information on time series data such as RNN and LSTM is used. This study focused on soil humidity, one of the cultivation factors. Soil humidity is one of the main factors influencing the growth of crops, and designating the irrigation period by accurately predicting the soil humidity while the crops are growing has a great influence on the increase in agricultural production. However, for accurate prediction of soil humidity, research on the relationship between soil humidity and environmental data is needed.

2. Data & Analysis Methodology

As shown in the table below, onion data growth data and environmental data are for each farm by region. Onion growth data is divided into the ground and the basement. It is divided into leaf-cutting leaf-end test site leaf-cutting plant leaf-cutting plant leaf-cutting plant in the ground and job-seeking plant-cutting plant-cutting plant-cutting plant-cutting plant-day weight in the basement. Growth data in each region has the same variable.

Gyeongsangnam-do	Jeonnam	Remark
Average temperature (°C) Average temperature (°C) Precipitation (mm) Atmospheric pressure (kpa) Soil moisture (%) Insolation (MJ/m2) Soil EC (%)	Average temperature (°C) Average temperature (°C) Carbon dioxide (ppm) Soil moisture (%) Insolation (MJ/m2) Soil EC (%)	Gyeongnam : Changnyeong Hapcheon Hamyang Jeonnam : Muan Sinan Hampyeong.

Table. 1. This table shows the differences in environmental data variables by region. In order to perform the analysis, it was integrated into common variables and preprocessed.

2.1 Data

Environmental data did not have the same variables by region. It was necessary to unify the variables with atmospheric pressure and precipitation in Gyeongnam and carbon dioxide in Jeonnam. Each insufficient variable was combined using the Meteorological Administration data. The Jeonnam Agricultural Technology Institute measured environmental data every hour, 30 minutes, and 15 minutes using sensors. In order to analyze time series data, we had to unify the time unit, so preprocessing was performed every hour.

The number of data for each farm after preprocessing every hour. From here, it was found that the data of Kim Kwang-taek farms were very small. Later, in the analysis results, the number of data was very insufficient and there were missing values, so the analysis results were not good, so this farm did not use them.

2.2 Analysis Methodology

Several models such as SVM(Support Vector Machin) DNN(Deep Neural Network) RNN(Recurrence Neural Network) LSTM(Long Short Term Memory) LSTM(Long Short Term Memory) Autoencoder were used in the preliminary study. In this study, long short- and long-term memory (LSTM) and CNN-LSTM are used as predictive models. CNN-LSTM aims to improve accuracy using a combination model of CNN(Convolution Neural Network) and LSTM(Long Short Term Memory). The idea is that CNN first extracts important features of the variables and puts refined variables into the LSTM

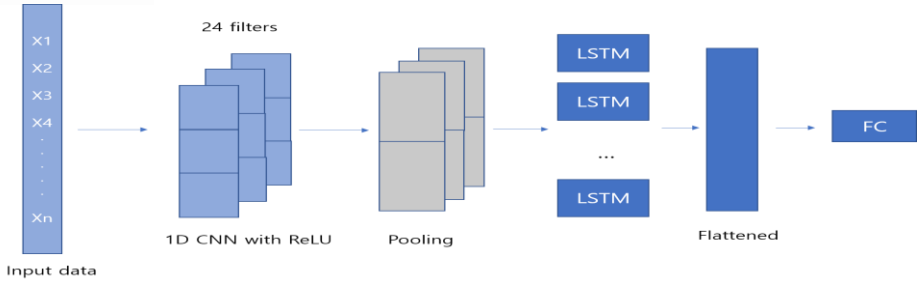


Fig. 1. CNN-LSTM Structure

In fact, if you check the reason in papers using CNN-LSTM, you can expect better predictive performance than when just using LSTM by filtering out noise from input data in the CNN model. The training set and test set were divided by 7 to 3 for each farm. Considering that it is generally 24 hours a day, kernel_size = 24 was set, the first 23 data points were used as predictors, and the last data points were used as target variables. For stride, 1 padding used the Rellu Tangent H function as the mutual padding activation function, and the optimizer used adam.

3. Analysis Results

In Figure 3, to tell you the lstm results first, it can be seen that most of the nine farms follow a similar trend.

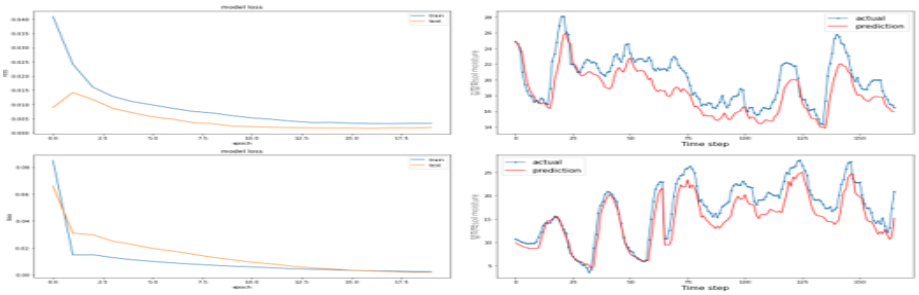


Fig. 2. This is the result of LSTM and x-axis and y-axis represent time and soil moisture.

This is the result of comparing LSTM and CNN-LSTM through In-gwang Kim farmhouse. Mse was used as the evaluation index, and as a result, it was confirmed that CNN-LSTM was better.

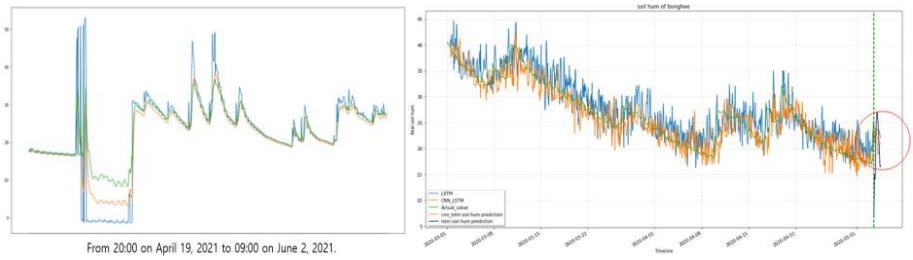


Fig. 3. The result(left) of CNN-LSTM(orange) and LSTM(green) in In-gwang Kim farmhouse and another result(right) of CNN-LSTM(orange) and LSTM(Blue) in Bong-hee Lee farmhouse, Gyeongsangnam-do and x-axis represents time and the y-axis represents soil moisture.

We predicted the last day by learning all days except the last day. In addition to this farmhouse, it was confirmed that CNN-LSTM reacts more sensitively to outliers with a narrower confidence interval than LSTM and predicts well.

4. Future Research

As such, it was confirmed that the CNN-LSTM model is good in that it purifies noise of input value compared to LSTM. However, if the analysis is performed with a larger amount of data than the amount of data you have now, you can expect better performance. Future research tasks include comparing predictive performance according to how to handle this part and using a better model in the data we received this time.

Acknowledgments. This work was supported by the Research Program of Rural Development Administration (Project No. PJ0153372021)

References

[1] CLEverReg: A CNN-LSTM based Linear Regression Technique for Temporal Fire Event Modelling(2019), Syed Adnan Yusuf & Abdul Samad & David James Garrity, Journal of Control Science and Engineering

[2] Prediction of Soil Moisture with Open Source Weather Data and Machine Learning Algorithms(2020), Young-bin Jang & Ik-hoon Jang & Young-chan Choe, Korean Journal of Agricultural and Forest Meteorology

[3] COVID-19 Pandemic Forecasting Using CNN-LSTM: A Hybrid Approach(2021), Zuhaira M. Zain & Nazik M. Alturki, Journal of Control Science and Engineering

Comparison of performance of various methods for segmenting Korean cattle regions from 3D images

Yunjeong Kang[1], Wanhyun Cho[2], Myung Hwan Na[3], Sangkyoon Kim[4]

[1,2,3] Dept. of Mathematics and Statistics, Chonnam National University,
Yongju-ro 77, Buk-ku, Gwangju, Korea
[4]Dept. of Electronics Engineering, Mokpo National University,
1666, Yeongsan-ro, Cheonggye-myeon, Muan-gun, Jeollanamdo, Korea

[1] yooon3220@naver.com, [2] whcho@chonnam.ac.kr, [3] nmh@chonnam.ac.kr,
[4]narciss76@mokpo.ac.kr

Abstract. In farms raising livestock, weight is used as an important indicator in many ways. At this time, when trying to measure the weight of the livestock, the livestock tends to lose weight due to stress or refusal to feed. Therefore, we need a method that can automatically measure the weight of livestock without using a scale. The first step in implementing this method is to segment the livestock from the input image. Therefore, in this paper, we would like to introduce the process of preprocessing the collected data to propose a model that can automatically segmented a body of livestock from the given 3D image. We collected RGB-D images at Korean cattle ranch in Gangwon-do used for analysis. Here, we considered the following four cases as a method of segmenting a Korean beef body from a given image. These are a method based on an RGB image, a method based on a depth image, a Grabcut method, and a combination of these, respectively. And to compare the performance of the considered partitioning methods, we used the Intersection over Union (IoU) measure. As a result of the analysis, using the combination of grabcut with depth, IoU showed the best performance of 0.8 or higher. In future research, we intend to build a system that can predict the weight of Korean cattle by extracting feature vectors from objects using grabcuts and depth-segmented images. In addition, since data on livestock such as pigs are also being collected in addition to Korean beef, a method for estimating the weight of various livestock using the developed model is also being considered.

Keywords: Livestock, Image processing, Cattle image segmentation, RGB-D image, Grabcut, Intersection over Union

1 Introduction

In farms raising livestock, the weight of livestock is used as an important indicator in many ways. The importance of livestock weight can be confirmed in two main aspects. First, we can check the current health or growth status of each livestock through the weight of the livestock. In generally, livestock refuse to feed or lose weight when they

are under stress or when diseases occur. It is also used as a way to check the current growth status, and you can use the age and weight of livestock to check whether they are growing appropriately and whether food is being paid according to the current growth status. Second, it can be used as a method to determine whether shipment is appropriate. Farms that produce beef determine the production grade or slaughter based on their weight. If you check the weight at regular intervals, you can determine the proper shipment of livestock.

However, the process of measuring the weight of large livestock is very labor-intensive. The existing method for checking the weight is to repeat the process of moving the livestock from the barn to the scale to check the weight and move back to the barn. In this process, livestock does not move according to their will, so resistance may occur and also stress may occur. People sometimes take a lot of time because they deal with many individuals. If an automation system that estimates weight is used, the farm can be operated much more efficiently than it is now.

Therefore, in this paper, to build a system for estimating the weight of livestock, we collected 3D images from Korean cattle farms and considered methods for segmenting the body of Korean cattle using these images.

2 Data Collection and Analysis Methodology

2.1 Data Collection

The data used for the analysis were collected at a Korean cattle ranch in Wonju, Gangwon-do. From June 2021 to the present, 15 Korean cattle of different ages are being measured and filmed once every two weeks. The collected image dataset consisted of 353 data sets by manually selecting 3~5 frames from the depth and color images captured using Intel RealSense depth cameras D455.

2.2 Analysis Methodology

The overall process of creating a system that can estimate the weight of livestock is shown in the Fig1 below. Separate the background and the object from the input image. Processing such as removing noise from the separated object is performed, and features are extracted. Various features can be selected based on shape, size, gradient, etc., and feature vectors are constructed using them. If it is suitable for a regression equation using the configured feature vector as a variable, the weight may be estimated based on the input image. In order for such a framework to be performed well, accurate size and weight estimation are possible only when the background and objects can be properly separated.

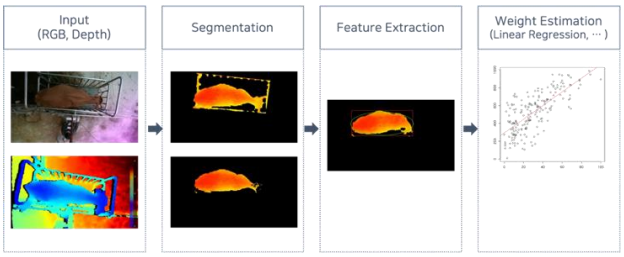


Fig. 1. If RGB image and Depth data are put together, the image is segmented. And the equation for estimating the weight is fitted using the feature vector extracted from the divided image.

In this study, we tried to apply several methods to this image segmentation process and compare its performance. Three methods were used to divide objects in the image. First, the HSV color space, which is mainly used in object detection, was used as a method using color. The RGB image was converted into an HSV image and a threshold was given to leave only pixels belonging to a specific color area. The range of Korean cattle colors set in this research ranged from 1 to 15.

$$f(x,y) = \begin{cases} 255, & if\ H_a \le Hue \le H_b \\ 0, & otherwise \end{cases} \quad (1)$$

The second method was a method using distance information, and depth information collected together with RGB images was used. By limiting the search range, only pixels belonging to the corresponding range were left. The depth range set in this study was 1.3m to 1.7m. If the depth of each pixel in the image falls within the selected upper and lower limits, a value of 255 is assigned; otherwise, 0 is assigned.

$$\mathrm{f(x,y)} = \begin{cases} 255 & if\ T_a \le d(x,y) \le T_b \\ 0 & otherwise \end{cases} \quad (2)$$

Here, the selection of the two lower bound (T_a) and upper bound (T_b) is made in two steps.

Finally, Grabcut algorithm was used. The Grabcut algorithm is a method of separating the foreground and the background based on the graph cut. It is a popular because a powerful interactive tool that does a good job segmenting the foreground and background. Grabcut repeats the process following. A user selects a rectangle box and created initial trimap. Pixels inside the rectangle are marked as unknown and outside will be background. Then Gaussian Mixture Models(GMMs) are created for initial foreground. Background and foreground pixels are aligned to the most likely Gaussian component. The GMMs are thrown away and then new GMMs are learned from pixel sets created in the previous set. Graph cut finds new boundaries between the background and foreground.

The Intersection over Union (IoU) was used as an index to compare the performance of the divided image. The IoU stands that the intersection of the correct answer image and the predicted image is obtained by dividing it into a combined part. If the IoU value exceeds 0.5, it is judged that the performance of the divided image is generally good.

3 Analysis Results

As a result of the analysis, when HSV or depth information was used alone, the average IoUwas very low at 0.3, and when Grabcut was used, it showed the best performance at 0.6. In case only colors are used, other objects with similar colors to Korean cattle are segmented together, and in case only distance information is used, objects with similar height to Korean cattle are segmented together, and thus, it could not show good performance(Fig 2).

Fig. 2. This is the result of segmentation using color (left) and distance (right) information.

When using a combination of image segmentation methods, when grabcut and depth information were used together, it showed the best performance of 0.8 or more and fps was about 1.4. In addition, when checking the overall IoU value, it can be seen that image segmentation is stably performed when Grabcut and depth information are used together.

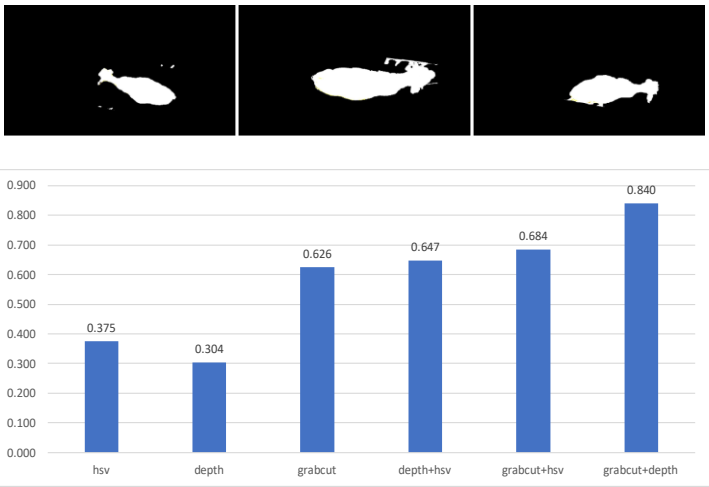


Fig. 3. The result of segmented image using Grabcut and depth information together (top). The graph below depicts an average IoU, which confirms that the performance when using multiple methods together is better than when using only one method.

4 Future Research Plans

Features can be extracted from the divided image to form a feature vector, and the weight estimation equation can be fitted using this. On average, Grabcut and depth information, which showed the best performance, will be used together to create an expression that can extract feature vectors and estimate weight from divided image images. In addition, since data on livestock such as pigs are also being collected in addition to Korean beef, a method for estimating the weight of various livestock using the developed model is also being considered.

Acknowledgments. This research was partially supported by a research grant from the National Research Foundation of Korea (Project No.: NRF-2020R1F1A1067066) and Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) and Smart Farm R&D Foundation (KosFarm) through Smart Farm Innovation Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) and Ministry of Science and ICT (MSIT), Rural Development Administration (RDA)(421017-04).

References

[1] D. Hema and Dr: S. Kannan, “Interactive Color Image Segmentation using HSV Color Space”, *Science and Technology Journal*, Vol. 7, 2020. http://doi.org/10.22232/stj.2019.07.01.05

[2] P. Ganesan, V. Rajini, B. S. Sathish and K. B. Shaik, "HSV color space based segmentation of region of interest in satellite images," *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, 2014, pp. 101-105, doi: 10.1109/ICCICCT.2014.6992938.

[3] Xiao Lin, Dalila Sánchez-Escobedo, Josep R. Casas and Montse Pardàs, “Depth Estimation and Semantic Segmentation from a Single RGB Image Using a Hybrid Convolutional Neural Network”, *Journals of Sensors*, vol. 19, No. 8, 2019. https://doi.org/10.3390/s19081795

[4] C. Rother, V. Kolmogorov, and A. Blake, “GrabCut: Interactive foreground extraction using iterated graph cuts”, *ACM Trans. Graph.*, vol. 23, pp. 309–314, 2004. https://doi.org/10.1145/1015706.1015720.

[5] Feilong Kang, Chunguang Wang, Jia Li, Zheying Zong, "A Multiobjective Piglet Image Segmentation Method Based on an Improved Noninteractive GrabCut Algorithm", *Advances in Multimedia*, vol. 2018, Article ID 1083876, 9 pages, 2018. https://doi.org/10.1155/2018/1083876

[6] Chang-bok Lee, Hyun-chong Cho, “Application of Deep Learning-based Image Segmentation Algorithm for Korean Cattle Weight Estimation”, *The Transactions of the Korean Institute of Electrical Engineers*, vol. 70, no. 9, pp. 1336-1344, 2021. https://dio.org/10.5370/KIEE.2021.70.9.1336

[7] B.Spatika Mira, T.Ravichandran, G.Yamuna, R.Durga, “A Review on Grabcut Algorithm”, *International Journal of Research in Advent Technology*, vol. 6, no. 10, pp. 194-200, 2018.

Forecasting Beef Price Using BERT Based Language Model with LSTM

Sooram Kang[1] , Wanhyun Cho[2], HyungChul Rah[3], Myung Hwan Na[4],

[1,2,4] Dept. Meth and Statistics, College of Natural Science, Chonnam National University, Yongju-ro 77, Buk-ku, Gwangju, Korea
[3]Research Institute of Veterinary Medicine, Chungbuk National University, 1, Chungdae-ro, Seowon-gu, Cheongju, Korea
[1] slkang2001@gmail.com, [2] whcho@chonnam.ac.kr, [3] rah.remnant@gmail.com, [4] nmh@chonnam.ac.kr

Abstract. Beef is one of Koreans' favorite meat, consumed a lot accordingly, and is one of the important factors in the domestic economy. As Korean economy grows and people's consumption power increases, demand for beef has continued to increase. However, the price of beef varies greatly when supply and demand imbalance occur due to temporary issues such as distribution, disease, or COVID-19. Therefore, related organizations continue to analyze and present trends in the price and supply of beef. However, since the analysis of related trends is conducted on a monthly or quarterly basis, immediate information acquisition is difficult, and its accuracy is insufficient. Therefore, this study aims to immediately reflects the issues related to beef and attempts to predict weekly unit prices while reflecting price fluctuations well. To reflect related issues, online articles are crawled to extract key sentences and used as a variable for the price prediction. To this end, a model that combines Kobert, one of BERT's Korean version models, and LSTM is proposed. In order to predict the the retail price of beef, online articles, economic indicators, and variables related to beef supply are used as exogenous variables. It is expected that short-term issues will be reflected through the proposed model to enable short-term price prediction, increase accuracy, and apply not only to beef but also to other livestock products or fields.

Keywords: beef price, BERT, Kobert, NLP, LSTM, price forecasting

1 Introduction

According to the Korea National Statistical Office, Korean per capita meat consumption increased from 11.3kg in 1980 to 53.9kg in 2018. Among them, beef consumption increased by 4.8% annually from 2.4 kg in 1980, and it reached 14.2 kg in 2018. In particular, the beef price is also increasing steeply along with the supply of beef as demand for home-cooked meals increases due to COVID-19. The beef price is one of the important factors in the household economy, but in the case of the retail price, the price deviation is large. Therefore, related organizations are publishing related trends monthly or quarterly, and several studies related to supply and demand

and the price prediction are being conducted. The following Figure 1 shows the trend of the retail price of the beef sirloin. It exceeded 100,000 won in 2021 from 50,000 won in 2011, and it has soared in 2020 with the COVID-19 epidemic.

However, related organization announces the beef supply and demand and price trends by quarterly or monthly. Therefore, it is necessary to take the price forecasting weekly or daily so that it can be used by actual economic actors, and to do so, it is necessary to quickly apply issues related to Korean cow and beef. To this end, in predicting the beef price, online articles were crawled and sentence expression through the Kobert model was used as a predictor. Online articles have the advantage of dealing with issues in a very early time along with the number of articles, and reliability can also be secured compared to social media. The prediction model used LSTM, a deep learning time series model.

Fig. 1. The trend of the retail price of the beef sirloin from January 2011 to September 2021

2 Data set and Method

For the analysis, data was collected between January 2011 and September 2021. In order to predict the retail price of beef sirloin, the number of Korean cattle raised, beef imports, CPI(Consumer Price Index), CCSI(Consumer Composite Sentiment Index) and the Kobert model's sentence expression for online articles crawled with keywords such as ‘Hanwoo’ and ‘Sogogi’, which means Korean cow and beef in Korean were used as exogenous variables. Online articles were crawled by Naver News and consisted only of articles from Yonhap news, which is a news agency of the state. Each data has a different time unit, so it was necessary to match it into one time unit. This study aims to predict the price on a weekly basis, so each data was converted on a weekly basis. In addition, in the case of the number of Korean cattle raised, the data before and after 2014 when the resume system was implemented is different, so the values from 2011 to 2013 were replaced by the average from 2014 to 2017.

LSTM(Long-Short Term Memory) is a model developed from the Recurrent Neural Network(RNN) and is mainly used in time series prediction and natural language processing. In traditional neural networks, output of the previous time point and the input of the current time point act independently but the RNN is characterized but RNN is characterized by a cyclic structure in which the output of the previous time point is input at the current time point. With this feature,it is widely used for time series natural language analysis. However, RNN decreased its learning ability due to the vanishing gradient problem. On the other hand, LSTM overcame this problem by learning long-term dependency through the gate inside the cell.

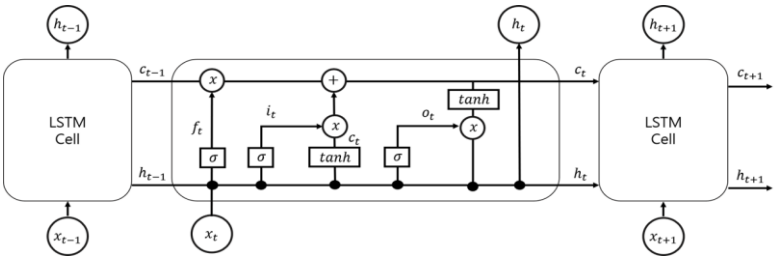


Fig. 2. The structure of LSTM

Kobert is a model learned Korean language and released by SK Brain on the structure of BERT. BERT is a pre-trained model unveiled by Google in 2018. It was pretrained with a vast amount of data and is used by additional training which is fine-tuning with labeled data on other tasks. The structure of the BERT is stacked with the encoder of the transformer. In the base version used in this study, a total of 12 encoders were stacked. Basically, BERT uses three embedding layers. WordPiece Embedding Layer tokenizes sentences and receives them as input of the model. Position Embedding Layer learns the order of tokenized words to learn location information. Segment embedding distinguishes between two sentences and handle tasks that require input of two sentences, such as QA problem.

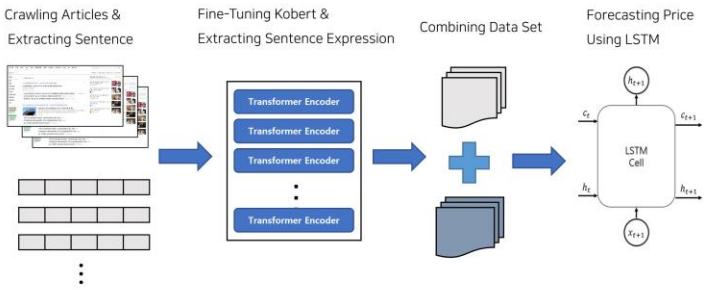


Fig. 3. The structure of proposed model

3 Experiment Result

In this study, the model proposed for predicting the retail price of beef sirloin is the structure of connecting Kobert and LSTM. Sentences containing 'Hanwoo' and 'Beef' from the collected articles were used as input of Kobert for fine-tuning the model. In general, after fine-tuning the BERT-based model with labeled sentence, tasks are performed through a classify layer at the end of the model, but in this study, only the expression of the model for the sentence after fine-tuning was used as input of LSTM. The extracted sentences were treated as one sentence per week. Therefore, sentence expressions can be said to be information on a week's online article about beef. The extracted sentence expressions were used as an exogenous variable of LSTM along with the number of Korean cattle raised, beef imports, CPI, and CCSI. And the train data and the test data were divided into 8 to 2.

The LSTM model learns twelve weeks of data and predicts the price after one week. The model with the best predictive power was presented by comparing models with window sizes of at least four weeks to up to twenty weeks. The model has one hidden layer and six hidden nodes. The predictive power was based on RMSE and MAP, and was found to be 6052 and 0.05, respectively. Since retail stores usually sell beef on a 100g basis, it can be said to have a deviation of about 605 won from the price of 100g.

Fig. 4. The prediction of model(orange) and real value(blue) of the price of beef sirloin

4 Conclusion

This study proposed the Kobert-LSTM model, which uses online articles as exogenous variables along with existing beef-related variables to reflect immediate issues in predicting beef prices. As a result of the experiment, the predictive power of the model with 12 weeks of window size, 1 hidden layer, and 8 hidden nodes was the highest. Through the proposed model, it is expected to be applicable not only to beef but also to various livestock products or crops. Through future research, it will be

possible to discover keywords that can reflect beef-related issues well and increase the predictive power of the model.

Acknowledgments. This research was partially supported by a research grant from the National Research Foundation of Korea (Project No.: NRF-2020R1F1A1067066) and Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) and Smart Farm R&D Foundation (KosFarm) through Smart Farm Innovation Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) and Ministry of Science and ICT (MSIT), Rural Development Administration (RDA)(421017-04) and Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ0153412021)" Rural Development Administration, Republic of Korea.

References

1. Seo, D., Shin, J.: An Analysis of the Effect of the Livestock Disease on Livestock Prices. Journal of Industrial Economics and Business. 27(6), 2431--2450 (2014)
2. Kim, D. H., Baek, S. W., Kim, S. H.: A Study on Prediction of Livestock Product Prices Using Intervention ARIMA Model - Focusing on Korean Beef and Imported Beef in the Corona 19 Pandemic Era -. The Korea Food Marketing Association. 38(1), 95--111 (2021)
3. Putri, R., Sukiyono, K., Sumartono, E.: Estimation of Indonesian Beef Price Forecasting Model. AGRITROPICA. 2(1), 46--55 (2019)
4. Chuluunsaikhan, T., Ryu, G., Yoo, K., Rah, H., Nasridinov, A.: Incorporating Deep Learning and News Topic Modeling for Forecasting Por Prices: The Case of South Korea. Agriculture. 10(11), 513; https://doi.org/10.3390/agriculture10110513 (2020)
5. Ryu, G., Rah, H., Nasridinov, A., Yoo, K.: Forecasts of the Amount Purchase Pork Meat by Using Structured and Unstructured Big Data: The Case of South Korea. Agriculture. 10(1), 21; https://doi.org/10.3390/agriculture10110513 (2020)
6. Shafqat, W.: A Hybrid Approach for Topic Discovery and Recommendations based on Topic Modeling and Deep Learning. Dissertation. Jeju National University, (2018)

Effects of cooking shows on the consumption of agricultural products: Focused on potatoes

HyungChul Rah[1.1], Hyeon-Woong Kim[2], Hyeon-Seok Ko[3], Jaehoon Shin[3], Yongbeen Cho[3], Aziz Nasridinov[1.2], Kwan-Hee Yoo[1.2]

[1.1] Research Institute of Veterinary Medicine,
[1.2] Department of Computer Sciences, Chungbuk National University, Cheongju, Korea
[2] Korea Rural Economic Institute, Naju, Korea
[3] Department of Digital Agriculture, Rural Development Administration, Jeonju, Korea
hrah@cbnu.ac.kr, economisthw@krei.re.kr, hsko@korea.kr, shinj@korea.kr, cho0yb@korea.kr, aziz@chungbuk.ac.kr, khyoo@cbnu.ac.kr

Abstract. Cooking shows on TV and YouTube channels have increased recently, and the influence of these broadcasts on food consumption is increasing. There were news articles about 'Baek Jong-won effect', in which the consumption of the agri-food mentioned on his cooking shows soared. In this study, we evaluated if 'Baek Jong-won effect' exists on the potato consumption, which Mr. Jong-won Baek broadcasted cooking recipes on TV and YouTube. After the potato recipe was broadcasted for the first time on TV show called HomeFoodRescue, the differences in the amount of money to purchase potatoes before and after the broadcast were estimated at 6 time points (3, 6, 9, 12, 24, and 36 months) by using the difference-in-differences method. The potato purchases at post-broadcast were less than those at pre-broadcast. No results were found suggesting the existence of 'Baek Jong-won effect' in the study.

Keywords: Baek Jong-won effect, Difference-in-Differences, Agri-Food, Potato

1 Introduction

Recently, cookbang or cooking shows on TV and YouTube channels have increased, and the influence of these broadcasts on food consumption is increasing [1, 2]. Mr. Jong-won Baek is one of the most popular chefs in Korea, and his cooking recipes have been famous, which have been aired on TV and YouTube and shared through blogs [3]. There have been several cases of being sold out after his advertising the agricultural products through the broadcasts he appeared on when agricultural product stocks were piled up due to overproduction or when prices of agricultural products plummeted [4]. There were several news articles about 'Baek Jong-won effect', in which the consumption of the mentioned agri-food soared [5]. In this study, we evaluated if 'Baek Jong-won effect' exists on the consumption of agricultural products.

2 Methods

In order to analyze 'Baek Jong-won effect' on agricultural products consumption, potatoes were selected among the candidates. Mr. Jong-won Baek showed potato cooking recipes four times (2015/06/15, 2016/05/10, 2017/02/21 and 2017/06/13) in the HomeFoodRescue (집밥백선생 in Korean), a cable TV show that started on 2015/05/19 and twice (2019/12/18 and 2019/12/23) in the Paik's Cuisine (백종원의 요리비책 in Korean), a YouTube channel by himself, which started on 2019/06/11.

After the potato cooking recipe was broadcasted for the first time on the HomeFoodRescue show on 2015/06/16, the differences in the amount of money to purchase potatoes before and after the broadcast were estimated at 6 time points (3, 6, 9, 12, 24, and 36 months) by using the difference-in-differences (DID) method. The potato consumption was measured by using data of amount of money to purchase of Agri-food consumers panel by Rural Development Administration. The data collection period was from up to 3 years before the first broadcast of potato cooking recipe on the HomeFoodRescue show to 3 years after the broadcast (2012/06/16 – 2018/06/15).

The structured data include the amount of money to purchase potatoes, potato retail and wholesale prices, wholesale market volume, annual and previous year production, cultivated area, and yield per unit area. The unstructured data were collected from TV including the number that potato cooking recipes were aired on the HomeFoodRescue show and from blogs including the number of blog posts mentioning HomeFoodRescue and potato cooking recipes (Table 1). The unstructured data described above were used as the independent variables. The amount of money to purchase potatoes was used as the dependent variable, and the control variables were the structured data described above except the amount of money to purchase potatoes.

Table 1. Status of unstructured data collected.

Feature Name	Number of data before up to 3 years of the first show	Number of data after up to 3 years of the first show
Number that potato cooking recipes were aired on the HomeFoodRescue show	0	28
Number that potatoes were mentioned on the HomeFoodRescue show	0	295
Number of potato purchases from the consumers panel data	9,470	9,122
Number of blog posts mentioning HomeFoodRescue and potato cooking recipes	38	2,315
Number of likes for blog posts mentioning HomeFoodRescue and potato cooking recipes	2	888
Number of comments on the blog posts mentioning HomeFoodRescue and potato cooking recipes	330	33,601
Number of blog posts mentioning potato cooking recipes	5,932	4,929
Number of likes for blog posts mentioning potato cooking recipes	4,278	3,225
Number of comments on the blog posts mentioning potato cooking recipes	127,111	77,731

Potential ‘Baek Jong-won effect’ on potato consumption was evaluated by using DID method. The DID method is useful to evaluate the effect of a policy or program by comparing before and after implementation of a policy or treatment as the concept of a control and an experimental group [6]. The treatment variable is a dummy variable indicating whether potential consumers were exposed to the potato recipes through blogs after the cooking recipe was first broadcasted on the HomeFoodRescue show. If the frequency of blogs mentioning “HomeFoodRescue” and “potato” during the post-broadcast periods is higher than that during the pre-broadcast periods, this variable is set to 1. If the blog frequency after the broadcast was lower than that before, a value of 0 was taken. Time variable is a dummy variable indicating if observation was made before or after June 16, 2015. A value of 1 was assigned to the periods from June 16, 2015 to the end of the study period, and a value of 0 is assigned to the periods from the beginning of the study period to June 15, 2015. For the seasonal fluctuations in the wholesale and retail prices, a value of 1 was assigned to the periods from April to July when the prices rise, and a value of 0 is to the periods from August to March.

3 Results

The amount of money to purchase potatoes were compared by using the DID method before and after the potato recipe was first broadcasted on the HomeFoodRescue program at 6 time points including 3, 6, 9, 12, 24, and 36 months after controlling for independent variables that could affect the money amount to purchase potatoes. Retail and wholesale potato prices, and the amount of money to purchase potatoes have shown a seasonal trend that potato prices (retail and wholesale) rise in April-July (Fig. 1).

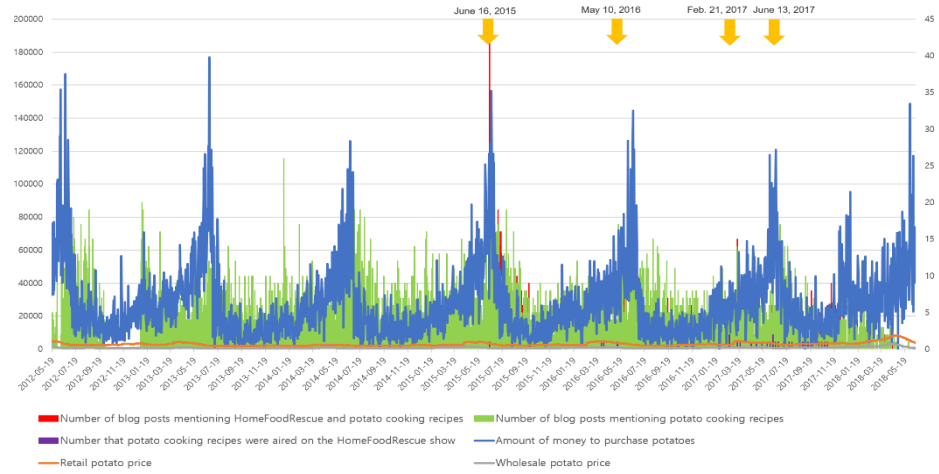


Figure 1. Trends of the money amount to purchase potatoes, retail and wholesale potato prices, potato cooking recipe-related blogs and TV shows during 2012-2018 with the dates of potato cooking recipes on HomeFoodRescue indicated by yellow arrows.

The results showed that the amount of money to purchase potatoes at 3 months after the broadcast was 4,705 won less than during the same period before the broadcast. The

differences of the amount of money to purchase potatoes were 2,939 won less at 6 months, 3,072 won less at 9 months, 6,088 won less at 12 months, 6,721 won less at 24 months, and 8,985 won less at 36 months than during the same period before the broadcast. The difference was statistically significant at 36 months alone (*p value* = 0.017). The post-broadcast potato purchases were less than the pre-broadcast potato purchases and the expected 'Baek Jong-won effect' was not observed.

4 Discussion

In order to evaluate 'Baek Jong-won effect' on agri-food consumption, the potatoes, which Mr. Jong-won Baek broadcasted cooking recipes for on a TV show called HomeFoodRescue, were analyzed for its consumption by using the DID method. Although there have been many news articles on 'Baek Jong-won effect,' this is the first paper to evaluate 'Baek Jong-won effect' on agri-food consumption [4, 5].

The results showed that the post-broadcast potato purchases were less than the pre-broadcast potato purchases from 3 to 36 months. The results do not support 'Baek Jong-won effect' on potato consumption. However, the difference in the amount of money to purchase potatoes between before and after was the smallest at 6 months and increased thereafter. Furthermore, it cannot be ruled out that the potato cooking recipes of 'HomeFoodRescue,' aired on a cable TV channel, had not much effect on the members of the consumers panel. For future study, it may be necessary to use larger population data such as data of Point-of-sale system from large retailer marts.

Acknowledgments. This work was supported by "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ0153412021)" Rural Development Administration, Republic of Korea, and partly by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant number: NRF-2020R1I1A1A01071884).

References

1. Sung, H.J., Kim, H.Y., Cho, M.: How Can TV Food Programs Be Used as an Effective Restaurant Marketing Tool? An Extension of ELM with Perceived Risk. Sustainability 12, 7131 (2020)
2. Jeon, C., Ji, Y.: A Study on Irrational Consumption Tendency According to Exposure of Video Contents of Mukbang(eating broadcasts) and Cookbang(cooking broadcasts). International Journal of Tourism Management and Sciene 36, 23-40 (2021)
3. MIN, K.-W.: From coffee cups to commercials, Baek Jong-won is everywhere. Korea JoongAng Daily, (2021)
4. Yoon, J.: 유통업계 강타한 '백종원 효과'... 지금은 '백종원시대'? (The 'Baek Jong-won Effect' that hit the retail industry... Now is the 'Baek Jong-won Era'?). 민중의 소리(Voice of the people), (2020)
5. Lee, I.: 백종원 효과'...한돈시장 훈풍되나('Baek Jong-won Effect'... Korean ork market is booming). 축산신문 (Chuksannews), (2021)
6. Lee, S.E., Kang, J.H.: The Effect of Tourist Attraction Exposure in a TV Program on the Tourist Volume: The Case of tvN "Over Flowers"Series. Korean Association of Applied Economics 20, 5-35 (2018)

Effectiveness Analysis of Smoothing Technique in Sentiment Analysis-based Pork Price Prediction

Yifan Zhu[1], Tserenpurev Chuluunsaikhan[2], HyungChul Rah[3], Aziz Nasridinov[2]

[1] Department of Big Data, Chungbuk National University,
Cheongju 28644, South Korea
[2] Department of Computer Science, Chungbuk National University,
Cheongju 28644, South Korea
[3] Department of Management Information System, Chungbuk National University,
Cheongju 28644, South Korea

Abstract. With the popularization of smartphones and social networking platforms, more and more news media have been born. These news media report real-time events. Therefore, researchers began to pay attention to the valuable information in the news. Many researchers performed news frequency analysis, keyword analysis, and sentiment analysis to dig the hidden information. In sentiment analysis, the Sentiment lexicon is a general tool used to calculate the sentiment value of news. However, when using the sentiment lexicon to calculate and smooth sentiment values, there is still a lack of a novel mechanism. In this paper, we performed sentiment analysis by the KOSAC sentiment lexicon. Next, we proposed a smoothing technique to smooth the sentiment value of news. At last, we predict the future price by the LSTM model. The results show that the smoothed sentiment value can reduce the error rates.

Keywords: News, Sentiment analysis, Smoothing technique, Price prediction

1 Introduction

In South Korea, pork is one of the people's favorite foods, so the price stability of pork is of great significance to the livestock industry in South Korea. However, the price of pork in South Korea is unstable, so pork producers and consumers have gotten lots of damages. Therefore, pork price prediction is very important for the Korean market. Traditional price predictions often use past prices to predict future prices, but this method cannot deal with unexpected events. In recent years, pork affected by diseases such as African Swine Fever (ASF) and foot-and-mouth disease (FMD), and the price fluctuated very sharply. News media is an important channel to reflect real-time events [1, 2]. Researchers analyze pork-related news to dig out the price trend information. Sentiment analysis can directly reflect the sentiment of news reporters on the price trend. In South Korea, Seoul National University built a sentiment lexicon that can analyze the polarity and intensity of text emotions in 2013 [3], but it still lacks a better way to use it.

In this study, we first use a web crawler to collect pork-related news and pork retail price. Then, use the KOSAC sentiment lexicon to perform sentiment analysis. We propose a preprocessing technique to smooth the sentiment value of news. Then the retail price of pork and the smoothed sentiment value are input to the LSTM model to predict the future pork prices.

2 Related Study

Chuluunsaikhan et al. [2] used Topic Modeling to analyze the pork-related news and predicted the pork price by LSTM. They collected news from Pigtimes website, and the daily retail price of pork from KAMIS website. In Topic Modeling, they preprocessed news and get several topics and keywords list of news by LDA. Then they calculated the values of all keywords by TF-IDF. At last, they put the values of all keywords, and pork retail price into statistics machine learning, and deep learning models. Their results showed there is a strong correlation between the keywords of news and pork retail price. Ryu et al. [4] collected structured data like agri-food consumers panel data, sales of pork meat, and production of pork meat. They also collected unstructured data like broadcast news, television program/shows, and blogs. After data collection, they combined structured data and unstructured data to train models. Then they used these models to forecast the demands of pork. And they proved when combine structured data and unstructured data, the forecasted results can be improved. Hyopil et al. [3] proposed the scheme of how to construct Korean sentiment corpus based on the Multi-Perspective-Question-Answering (MPQA). They extracted 8050 sentences from the articles in the news report as the data source. They followed the MPQA scheme to classify the private state into SEED, expressive types, subjectivity types. In subjectivity types, there are also other subtypes, like emotion, judgment, agreement, argument, intention, speculation. Emotion subtype also can be marked as emotion-positive, emotion-negative, and emotion-neutral. This lexicon can provide a great convenience for Korean sentiment analysis.

3 Proposed Method

3.1 Overall

As shown in Figure 1, we first collect pork-related news and pork retail price from PIGTIMES and KAMIS by the web crawler. Then we use the KOSAC sentiment lexicon to calculate the news sentiment value. After calculating the sentiment value, we smooth the news sentiment value by the proposed smoothing technique. Next, we input the smoothed sentiment value (N_1, N_2, …, N_T) and pork retail price (P_1, P_2, …, P_T) into the LSTM model to predict the future pork price. At last, we evaluate the model performance by root mean square deviation (RMSE) and mean absolute error (MAE).

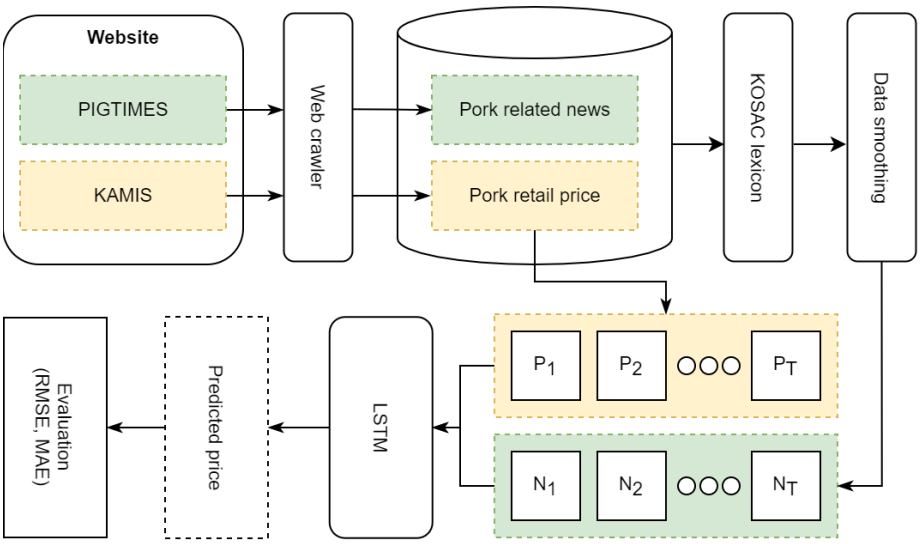


Figure 1. Overflow of proposed methods

3.2 Web crawler

We collect news from PIGTIMES [1] from January 2010 to December 2019. The object of our research is the pig, so we collect the news report that correlates with the pig. Therefore, we collected the news report from the PIGTIMES website. This website usually reports price trends, supply situations, demand situations, and government policy. We collect the pork retail price (2010.1.1 ~ 2019.12.31) from KAMIS [5]. KAMIS website releases prices for agriculture products and livestock products.

3.3 Calculate news sentiment value by KOSAC

KOSAC is built for capturing sentiment expressions and their patterns in Korean and representing their meaning to be interpretable for computer. We can download KOSAC online, and it contains some files like expressive-type, intensity, nested-order, polarity, subjectivity-polarity, and subjectivity-type. In our research, we used polarity.csv to analyze the sentiment value of every news. We can calculate the sentiment value of every morpheme in news by the polarity.csv file.

3.5 Smooth news sentiment value

To solve the irregular changes and outliers of news sentiment value, we will use the proposed smoothing technique to smooth the sentiment value of news and the formulation is:

$$S_t' = \begin{cases} S_{t-1}' + \dfrac{1 - e^{-|\text{diff}|}}{w(1 + e^{-|\text{diff}|})}, \text{if diff} \geq 0 \\ S_{t-1}' - \dfrac{1 - e^{-|\text{diff}|}}{w(1 + e^{-|\text{diff}|})}, \text{if diff} < 0 \end{cases} \quad (1)$$

where the sentiment value of first day in time-series table will be set as S'_t by default, S'_t is the smoothed sentiment value of day t, diff is calculated as: $S'_{t-1} - S'_t$

3.5 Construct machine learning and deep learning models

We constructed machine models and deep learning for training data. Random Forest (RF) is a flexible, easy-to-use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because of its simplicity and diversity. Gradient Boosting (GB) is an extremely popular machine learning algorithm that has proven successful across many domains and is one of the leading methods for winning Kaggle competitions. Extreme Gradient Boosting (XGB) is a decision-tree-based ensemble algorithm. It has a wide range of applications: regression, classification, etc. Long Short-term Memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections.

4 Result Analysis

In this section, we evaluate the performance of LSTM on raw and smoothed sentiment values. Next, we evaluated the prediction results of all models on raw and smoothed sentiment values.

4.1 Performance of raw sentiment value and smoothed sentiment value

Figure 2 shows the actual price and the price curve of LSTM with raw and smoothed sentiment values. We can see that LSTM with raw sentiment value can predict the overall price trend. But the price curve of LSTM with raw sentiment value is very steep from September 23, 2019, to October 7, 2019, and from October 24, 2019, to November 11, 2019. After smoothing the raw sentiment value, we can reduce the

outliers in the above two time periods, and the new price curve is more similar to the actual price. Therefore, we can conclude that the proposed smoothing technique can process news sentiment value effectively and improve the LSTM performance.

Figure 2. Predicted prices and Actual prices

4.2 Performance of all models

Figure 3 shows the RMSE of models with the proposed smoothing technique and models without the proposed smoothing technique. We can see that the smoothed sentiment value can reduce the errors of the three models except for Random Forest. The proposed smoothing technique is most effective for LSTM. Figure 4 shows the actual price curve and the prediction prices of all models on the raw and smoothed sentiment value. We can see that the prediction prices of the LSTM are the most stable and accurate. The GB prediction error is maximum, XGB, RF prediction results are most steep. After using the proposed smoothing technique, the prediction prices of all models have become more stable and accurate.

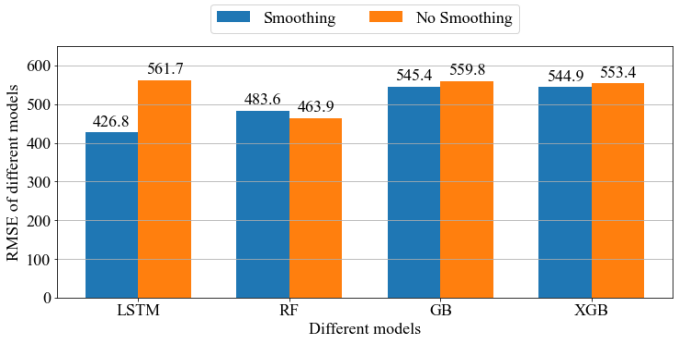


Figure 3. RMSE of models with raw and smoothed sentiment value

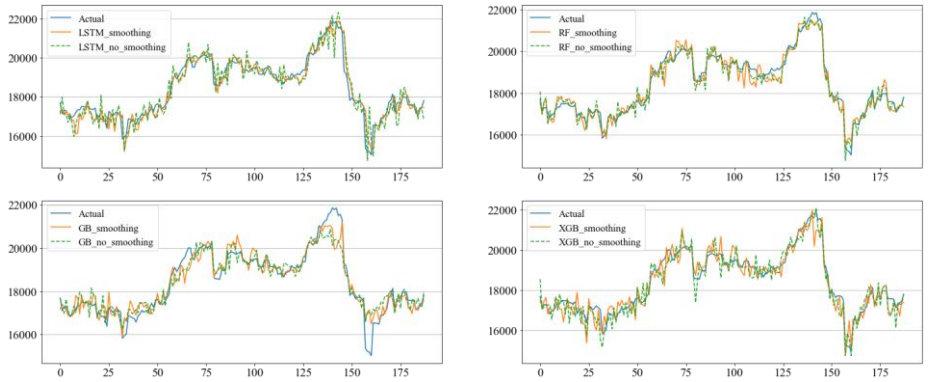


Figure 4. Actual prices and Predicted prices of all models

5 Conclusion

In this study, we collected pork-related news and pork retail price from PIGTIMES and KAMIS. Next, we used KOSAC to analysis to calculate the news sentiment value. Next, we proposed the smoothing technique to smooth the sentiment value. Finally, the smoothed sentiment value was input into LSTM to predict the future prices of pork. After smoothing the sentiment value, results could be more accurate than the raw sentiment value for LSTM, GB, and XGB models. Because of the advantages of the LSTM model in processing time series and nonlinear data, the performance of the LSTM model is better than the other three models.

This study also has some limitations. KOSAC is a universal sentiment lexicon, which is more suitable for the sentiment analysis of daily life. Therefore, we need to establish a domain-specific sentiment lexicon. Although PIGTIMES only publishes news related to pork, there is not much news related to pork price trends, so we need a mechanism to filter highly-quality news.

In the future, we will develop new data collection method to improve the quality of news. Next, we will try to automate the construction of sentiment lexicon to overcome the time-consuming and labor-cost of lexicon construction.

Acknowledgments. This research was funded by the "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ015341012021)" Rural Development Administration, Republic of Korea.

References

1. PIGTIMES: http://www.pigtimes.co.kr/

2. Chuluunsaikhan, T., Ryu, G. A., Yoo, K. H., Rah, H. C., Nasridinov, A., Incorporating Deep Learning and News Topic Modeling for Forecasting Pork Prices: the case of South Korea, Agriculture 2020, Volume 10.
3. Hyopil, S., Munhyong, K., Yu-Mi, J., Hayeon, J., Andrew, C., Annotation Scheme for Constructing Sentiment Corpus in Korean, in proceedings of the 26th Pacific Asia Conference on Language, Information and Computation, pages 181-190, 2012.
4. Ryu, G. A., Nasridinov, A., Rah, H. C., Yoo, K. H., Forecasts of the Amount Purchase Pork Meat by Using Structured and Unstructured Big Data, Agriculture 2020, Volume 10, Issue 1.
5. KAMIS: http://www.kamis.or.kr/customer/main/main.do

A Multidimensional Analysis System for Floor Impact Sound Reduction

Kyung-Hee Lee[1], Hyung-Min Cho[1], Hong-Seok Yang[2]

{lee.kyunghee, hchomin}@gmail.com, h.yang@lh.or.kr
[1] HealingSoft Co., Ltd., [2]Land and Housing Institute

Abstract. This study constructs a database in an easy-to-analyze form to accumulate floor impact sound data generated in an apartment building, and to support various analyses for reducing the noise. Based on this, the cause of a noise can be identified through data analysis from various aspects. Information on complex apartment structures and materials used in construction is also stored in the database to enable rich analysis. Users can conveniently manage and analyze noise data on the web portal and use it in business.

Keywords: Floor Impact Sound, Sound Database, Web Portal, Analysis

1. Introduction

In recent years, noise from middle floors in apartment buildings has emerged as a serious social problem. However, since the noise of the middle floor of an apartment building is determined by a combination of various variables such as the structure of the building and the materials used, accurate prediction is a complex problem. In this study, an easy-to-analyze database was built to accumulate floor impact sound data generated in apartment houses and to support various analyzes for noise reduction. Based on this, it is possible to identify the cause of noise through various aspects of data analysis. Information on complex apartment structures and materials used in construction is also stored in the database for rich analysis. Users can conveniently manage and analyze noise data from a web portal and use it in their business. When the built-up database is accumulated, it will become an important dataset for developing artificial intelligence prediction techniques using various variables. It is also designed to be highly scalable to accommodate a variety of future analytical needs.

2. System Architecture

This system uploads noise data related to floor impact sound measured in public housing as an Excel file or an employee can input the noise data one by one on the web into the database. All the data can be accumulated in a centralized database called a data lake. Users can analyze the accumulated database from various

perspectives by using multidimensional analysis. The system visualizes the analysis results and services them through an internet portal. Figure 1 shows the overall structure of this system.

The input data is largely divided into floor impact sound data measured by experts, structural data of public housing (floor structure or building structure, etc.), and material data indicating the materials used in the houses. In addition, drawings and certificates for materials are also stored together.

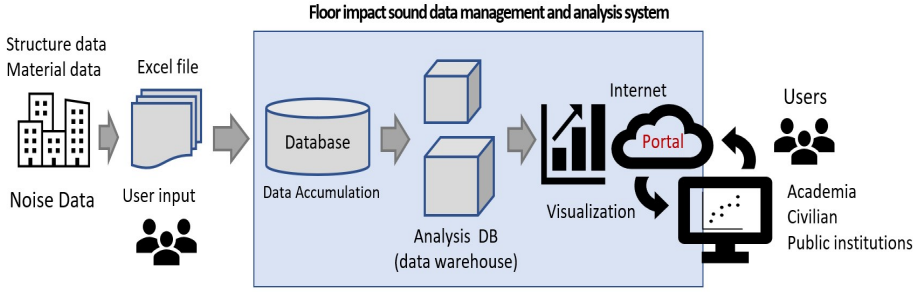


Fig. 1. Overall system architecture.

The database consists of a data lake that accumulates large input data and a data warehouse built for analysis from it. Currently, the noise data measured by experts is refined and then entered into the database in Excel or a record on the portal. In the future, if the installation of sensors for noise measurement in houses becomes common, data can be generated automatically in real-time. In this case, it is necessary to expand the data lake part by using big data technology, such as Nifi or ELK(Elastic Search, Logstash, and Kibana).

Analysis DB means data cubes or star schemas. Research is needed to find and reduce the cause of noise by analyzing the measured noise data from various perspectives. For this purpose, it is necessary to examine the change in noise by the building materials used, by the structure, or by a combination of these. A star schema is a suitable structure for performing this type of multidimensional analysis.

Visualization is a process of converting various analysis results into graph form so that users can intuitively judge. With the development of big data technology, numerous visualization techniques are being provided. In this study, Python-based Matplotlib and Seaborn, which are recently widely used open-source packages, were used to visualize the analysis results. Figure 2 shows some visualization results. Figure 2 is a visualization of the comparison between the noise measured in the slab layer and the noise in the finishing layer (lightweight on the left, heavyweight on the right). It can be seen from the two graphs that the noise of the finishing layer was reduced significantly.

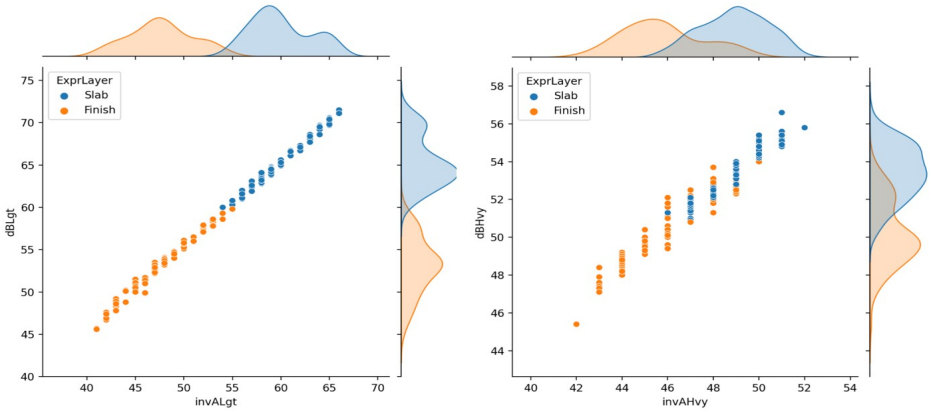


Fig. 2. Comparison of noise measured on the slab layer and the finishing layer.

Figure 3 is a visualization of slab floor noise (left) and finishing floor noise (right) by the bandwidth of the entire household. Each line represents the noise generated from a house. The graph shows the change in noise level by the house when the bandwidth is increased from 50 to 500. As shown in the graph on the left, overall dB noise (on Y-axis) from the slab layer is distributed in the range (33, 78), but the noise from the finishing layer in the right graph is reduced to the range (20, 75).

So far, the portal has provided 23 types of analysis results based on the accumulated data, and some of them are introduced in Figures 2 and 3. We expect that the number of analysis items will be increased by continuously accepting business needs. It also supports researchers to create various datasets easily and conveniently for their research purposes in the portal. Researchers can also access the database, develop, and register various analysis programs with Jupyter Notebook, and share them with others. We plan to develop a predictive model that can predict noise by introducing artificial intelligence techniques beyond multidimensional analysis.

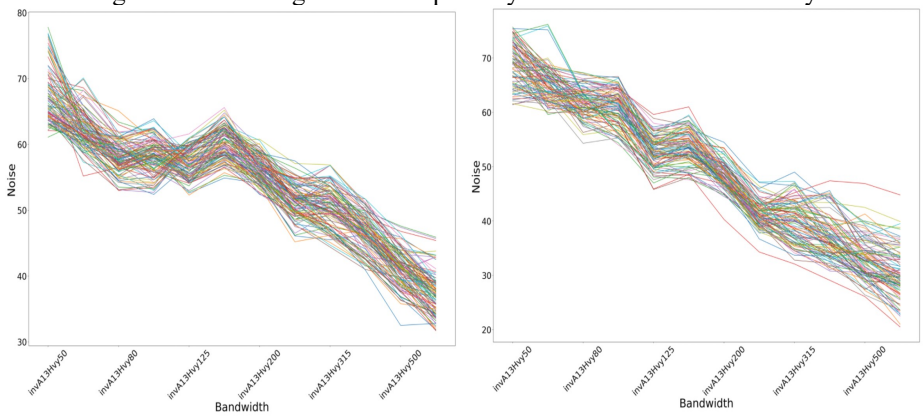


Fig. 3. Slab floor noise (left) and finishing floor noise (right) by bandwidths of all households.

3. Conclusion

Recently, along with the spread of public housing, the noise problem in the middle class is emerging as a social issue. As the middle floor noise, which is a concern for construction companies, is changed from the pre-approval method to the post-approval method, the problem of predicting the middle floor noise is emerging as an even more important task. In this study, a system for continuously collecting and analyzing noise data from the middle floor was established, and the analysis results were intuitively viewed on the portal using a visualization technique. In the future, noise data can be generated using various devices such as sensors, and it is necessary to expand so that advanced analysis and prediction techniques such as artificial intelligence can be serviced using this system.

Acknowledgment. This work is a part of the results of the 2020 regular project "Establishment of a technology sharing system for floor impact noise reduction and development of a customized model (R202002002)" supported by the Korea Land and Housing Corporation.

References

1. Hong-Suk Yang and Taemin Kim, Parametric Study on Floor Impact Sound with Design Factors of Layers Composing a Floating Floor in Multi-Residential Buildings, Trans. Korean Soc. Noise Vib. Eng., 30(2): 119~128, 2020 (https://doi.org/10.5050/KSNVE.2020.30.2.119)
2. Dae-Ho Mun, Analysis of Influence Factors of Floor Impact Sound in Residential Buildings, Ph. D. Thesis, Seoul National University, 2015.
3. A study on setting the direction of technology development for reducing the heavy floor impact sound of an apartment house, Research Report, Land & Housing Institute, 2013.
4. Factors affecting the inter-floor noise-blocking performance of apartment houses, Weekly Construction Technology Trend, Trend Report, 2013.
5. J. H. Kim, et al., Influence of Floor Dimension and Resilient Material on Heavy Impact Noise of Floating Floor System, Transactions of the Korean Society for Noise and Vibration Engineering, Vol. 27, No. 4, pp. 434~443, 2017.

A Study on Association between Metabolic Diseases and Food Consumption

Hong Jun Ho[1], Min ji Oh[1], Kyung Hee Lee[2], Wan Sup Cho [1]

[1]Dept. of Big Data, Chungbuk National University, Chungdae-ro, Seowon-gu, Cheongju-si Chungcheongbuk-do, Korea
[2]Data Department, Healingsoft Co., Ltd, #1206, HangangDaewoo Diobil, 115 Hangangro, Seoul, Korea
{ junho141201, minji.oh1020, lee.kyunghee }@gmail.com, { wscho }@chungbuk.ac.kr

Abstract. This research aims to identify the relationship between food and specific diseases, combining consumer panel data from RDA and heath checkup and medical treatment data from NHIS. The patient group was divided into 52 patients diagnosed with metabolic disease, and the control group was divided into 276 patients without metabolic disease diagnosis. It was determined that the interaction existed at the significance level α=0.05 as a result of testing the effect of two-way ANOVA interaction between the variable dividing the patient group and the control group and the purchased food.

Keywords: Metabolic disease, Food consumption, Two-way ANOVA

1. Introduction

This research aims to suggest food consumption characteristics necessary for preventing diseases by comparing and analyzing the food consumption of metabolic disease group and the control group by linking the consumer panel data of RDA and the medical treatment data of NHIS.

We created three types of variables: A class type of variables that divided the patient group and the control group, an agri-food-category type of variables that classified purchased food info 14 categories, and variables representing the average monthly consumption of food[1]. The technique used for the analysis is two-way ANOVA[2], which is an analysis of variance using two or more independent variables.

2. Research Method

Chapter 2 describes the data set used in the study and explains the main effects and interactions between the patient group and control group and food consumption patterns through the results of three tests of two-way ANOVA. The patient group is a group of 52 people who have been treated for metabolic diseases(type 2 diabetes, hypertensive disease, ischemic disease, obesity, dyslipidemia), and the control group

is a group of 276 people who are members of a different household than the patient group and have not been diagnosed with metabolic diseases.

Consumed foods in the data are divided into large, medium, and small categories according to the RDA consumer panel data classification criteria, but for smooth classification, food groups were reclassified based on the research[3]. By combining the health checkup data constructed by NHIS, the patient group and the control group were divided and compared.

2.1 Method and Experiment

First, the correlation on food categories for a total of 328 people was confirmed through exploratory data analysis. The correlation on noodles, bread, other proc food, and snack was high, and it was found that the relationship between vegetables and fruits was also high. According to the food patterns in which the correlation exists, it could be determined that the interaction between the patient group and the control group could also be identified.

Table 1. The Null Hypotheses

	The null hypotheses (H_0)	Statistics
1	There is no interaction between the class (patient group, control group) and the categories (consumption food groups).	F
2	There is no difference in the average monthly food consumption between the two groups (patient group, control group).	
3	There is no difference in the average monthly food consumption of 14 categories (food groups).	

A two-way ANOVA was conducted to determine the relationship between food consumption and metabolic disease. Two-way ANOVA is an analysis in which three pairs of null and alternative hypotheses are established. The hypotheses are as follows.

2.2 Result

In the experiment, all of the above three tests use the F-test statistic, and the F-value is the value obtained by dividing the between-group variance by the withon-group variance, and the difference is compared with the group variance.

In this study, the significance level α=0.05 was set, the F statistic was tested, and the explanation of the significance level is omitted.

	sum_sq	df	F	PR(>F)
C(category)	8337.168724	13.0	128.182037	2.596689e-296
C(Class)	14.366193	1.0	2.871400	9.023465e-02
C(category):C(Class)	111.850667	13.0	1.719678	5.048202e-02
Residual	22834.615529	4564.0	NaN	NaN

Fig.1. Two-way ANOVA statistics

As a result of interaction test 1, looking at the C(category):C(Class) row of fig.2, the F-value is 1.719678, and the significance probability value is 0.05048202, which is higher than the significance level α set in the study. This is the basis for rejecting the null hypothesis, and the first null hypothesis, which is class and category do not interact, is rejected. That is, there is a correlation between metabolic disease-related disease and food consumption under the significance level of 0.05, which dilutes the results of the other two tests.

3 . Conclusion

We obtained following conclusions from this research.

(1) There is an interaction between the 14 food group categories and the diagnosis of diseases related to metabolic diseases.

(2) The average monthly food consumption of people with metabolic diseases and those without diseases differs by food group.

Acknowledgement

This research was approved by the Institutional Review Board (IRB) at Ministry of Health and Welfare (IRB No.: P01-202012-21-010) to use the dataset. And this work was supported by Rural Development Administration grant funded by the Korea government (No. PJ01538032020, Research on the relationship between agri-food consumption and health)

Reference

1. Czekajło, Anna, et al. "Association between dietary patterns and metabolic syndrome in the selected population of Polish adults—results of the pure Poland study." European journal of public health 29.2 (2019): 335-340.
2. Jung, Mimi, et al. "Understanding and Interpretation of Interaction Effects in Multi-factor ANOVA Designs." KSME, 13.2 (2011): 1-15.
3. Jun-ho Hong, Jin-sil Kim, Yeon-ju Yu, Kyung-hee Lee, Wan-sup Cho. "Comparison of Housewives' Agricultural Food Consumption Characteristics by Age.", The Korea Journal of BigData, 6.1 (2021): 83-89.

An Active Youth Healthcare Management System based on Mydata and IoT Devices

Hyung-Min Cho[1], Nayeong Son[2], Kiyeol Ryu[3], Kyung-Hee Lee[1]

[1] Healingsoft/Bigdatalabs Co., Ltd.
[2] Chungcheongbuk-do Office of Education
[3] Department of Software and Computer Engineering, Ajou University
{[1]hchomin, [2]gamauji1004, [1]lee.kyunghee}@gmail.com, [3]kryu@ajou.ac.kr

Abstract. A mydata-based app that supports students' self-directed health management was proposed and a prototype system has been implemented. The proposed system supports school health teachers and students to collect, analyze and manage activity data in real-time using smartphone apps and smart bands. Based on this, close interaction between teachers and students is possible, contributing to the formation of healthy lifestyles for students.

Keywords: Mydata, Smartphone, Health bigdata, Analysis

1. Introduction

This study proposes a smartphone App that applies the concept of my health data for students' self-directed health management practices. By using the health management app, we can drive healthy living practices for selected risk group students; the students and teachers in a school can strengthen their self-directed health management capabilities through increased interactions based on a smartphone app and bigdata analysis. Students and teachers can scientifically lead a personalized healthy life through the analysis of each student's integrated health data. In addition, bio-signals generated from devices such as smartphones and smart bands are linked with existing health data to generate more useful decisions and customized services. Continuous interactions based on smartphone App and data analysis can lead the students to have a more healthy habit.

2. System Architecture

Fig.1 shows the system architecture for the target system. For high-risk students, physical examination data, health check-up data, medical records and medication information, meal log, bio-signals, counseling data, etc. are collected through ETL and saved in the database. The collected data is stored in a database (My Data) where each student manages his/her data. Students can upload or delete their data if they wish and can adjust the level of consent at any time. The collected data is linked by a

smartphone-based unique number and analyzed from various angles and provided as useful health information to students. In particular, school health teachers can support the formation of healthy lifestyles by conveniently integrating and managing high-risk students with smartphones and strengthening the interactions with students. In addition, accumulating health points and rewarding them motivates students for health activities. Accumulated data may be useful not only for the health researchers, police officers, public health teachers, but also the students. The usage log for the accumulated data can be stored in blockchain ledger for safer privacy.

Fig. 1 shows the system architecture. It consists of ETL (extraction, transformation, and loading), MyData database, and the App. for the users.

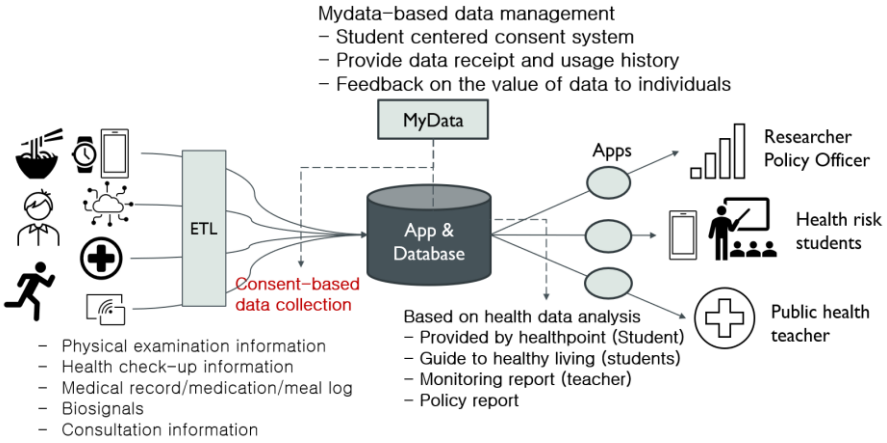


Fig. 1. The System Architecture

3. Prototype System

The prototype system consists of membership registration, consent management, health information collection including bio-signals, health information analysis module, and health activity display module. To implement the mydata principle, the level of consent is subdivided into four levels, and various consent management functions are provided: view my consent history, cancel consent, change consent, download consent details, etc. Health points are given to motivate students for health activities, and rewards are made when points are accumulated. The prototype system has been developed using Python, Keras, and Streamlit open source libraries.

4. Conclusion

A mydata based app system that supports students' self-directed health management was proposed and a prototype system has been implemented. Since the use of health information is strictly restricted in accordance with the Personal Information

Protection Act, the MyData principle was used to allow users to upload, change, and delete data based on their initiative. The proposed system supports school health teachers and students to collect, analyze and manage activity data in real-time using smartphone apps and smart bands. Based on this, close interaction between teachers and students is possible, contributing to the formation of healthy lifestyles for students.

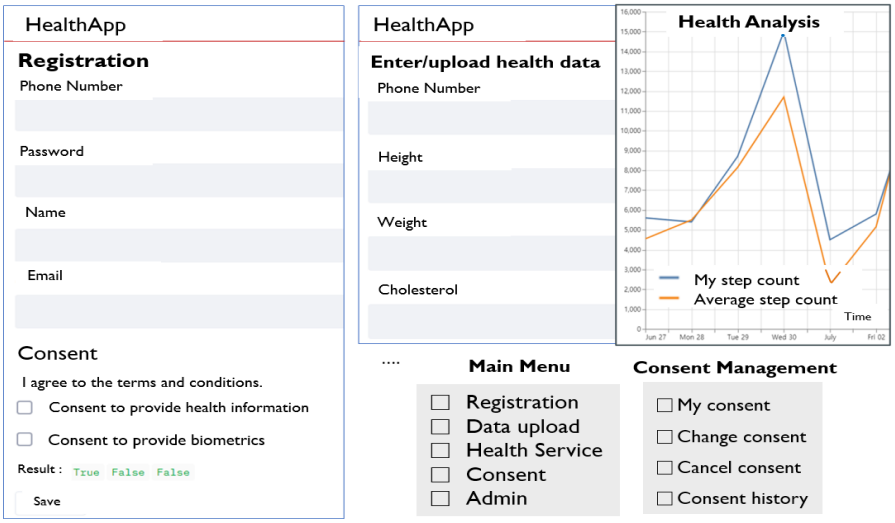


Fig.2. Main UIs of the Prototype System

Acknowledgement

This work was supported by Rural Development Administration grant funded by the Korean government (No. PJ01538032020, Research on the relationship between agri-food consumption and health)

Reference

1. Ministry of Education, Korea Educational Environment Protection Agency. “2019 Student Health Examination”. 2019.
2. Bong-Jeong Kim. “The Effect of Individual and School-Level Factors Affecting Adolescent Obesity: Application of Multilevel Analysis”. Journal of The Korean Data Analysis Society, 18(1), pp. 509-524. 2016.
3. Sukyung Kim, Saegyeol Choi, and Hyekyung Kim. “Multi-layered Factors Related to Adolescent Obesity: Analysis of Adolescent Health Behavior Survey Data for 2017-2019”. Journal of Health Education and Health Promotion, 38(1), pp. 13-24. 2021.
4. Yeo-Jin Lim, Won-Ok Oh, and Min-Hyeon Seok. “Health Behavior According to Obesity in Adolescents”. Child Health Nursing Research, 23(1), pp. 1-9. 2017.
5. Ministry of Food and Drug Safety. “The 15th (2019) Adolescent Health Behavior Survey Statistics”. 2019.
6. Korea Data Industry Promotion Agency. “My Data Service Guide”. 2019.

The Next Generation of Integrated Platform Approach for Smart Manufacturing

Jae Sung Kim[1], Saksonita Khoeurn[1], Saravit Soeng[1], Vungsovanreach Kong[1], Wan Sup Cho[2]

[1]Department of Big Data, Chungbuk National University, Chungdae-ro, Seowon-gu, Cheongju-si Chungcheongbuk-do, Korea

{comkjsb, saksonita, soengsaravit, kv.sovanreach }@gmail.com, [2]{ wscho }@chungbuk.ac.kr

Abstract. As the Industrial Internet of Things (IIoT) evolves, industrial device manufacturers start to pay attention to accepting the smart factory platforms. IIoT can significantly improve facility availability, product quality, and even productivity by predicting facility failures and product defects through real-time analysis by attaching sensors to various facilities. Up until now, Manufacturing Execution System (MES) solutions companies have collected and managed sensor data using relational databases as a branch concept of IIoT. Large amounts of sensor data are captured in a time-series manner, so it is possible to store relational databases traditionally. Still, the performance problems hassle the real-time analysis. Therefore, data migration and analysis in real-time are required. This paper aims to introduce an efficient and robust flow of access data that can solve the challenges of real-time decision-making by integrating data from relational databases captured by sensors into real-time analysis environments.

Keywords: Internet of Things**,** Data Flow, Data Collection, Data Storage, Smart Factory, Manufacturing Execution System.

1. Introduction

With the recent development of Information and Communication Technology (ICT) and the 4th Industrial Revolution, smart factories are drawing attention. Smart Factory refers to a factory that integrates the entire process, including planning, manufacturing, distributing, sales, and even facility maintenance with ICT applications. The technologies applied in the applications for smart factories include Big Data, Artificial Intelligence (AI), and the Internet of Things (IoT). These applications are developed to increase productivity and obtain customized manufacturing. Besides, MES has been widely used to adapt to these market changes. Most MES has been built and operated using the Relational Database (RDB) to store the attached sensor data. Manufacturers are attempting to expand sensor data and use it for real-time analysis to upgrade into smart factories. However, they have been using open-source tools such as RDB and Hadoop to retrieve sensor data as an extension concept of MES and store sensor data. As the number of sensors mounted on the factories has increased rapidly, data size has also increased. Still, companies are looking for alternatives since it is difficult to cover them with existing databases [1] effectively. Since a large amount of sensor data is captured in a time-series manner, the data must be collected and stored in a time-series

database. Also, they must be migrated to another database to analyze them in real-time. Notably, time-series databases are more suitable than existing RDB to store the data from IoT sensors. Using data migration traditionally can waste numerous resources, including time, labor, and cost. Therefore, data generated by the sensor must be automatically stored in different sources simultaneously to serve various purposes.

Hence, this paper aims to introduce an efficient and robust flow of access data captured by sensors from RDB to other targets databases in real-time using an open-source web-based tool. The proposed data flow enables the capabilities to build solutions that contribute to real-time decision-making for manufacturers and other related organizations.

The paper is organized as follows. The literature review on technology usage is discussed in section 2, while section 3 explains the proposed data flow. Section 4 illuminates the implementation and result of the proposed solution, and section 5 concludes the remarks.

2. Literature Review

This section is a brief literature review on each data flow process and the technologies mentioned in the paper.

2.1. Dataflow Processes

The very first stage is Data collection**.** Data collection is the process of collecting analog data in sensors from the physical world and transforming it into digital assets. The second process is data storage, which consists of storing and transforming the data into multi destinations. The last process is the data consuming services where end-users receive data system output focused on their decision-making requirements. Data output can be consumed by end-users or other application systems, including monitoring, analyzing, and additional business insights [2].

2.2. Tools and Database Management System

Microsoft SQL Server (MSSQL): a relational database management system (RDBMS) offers a variety of business intelligence, analytic application, and transactional processing in mostly corporative IT environments. It is known as one of the top three market-leading technologies in the database [3].

Time Series Database (TSDB): one kind of database that allows users to build real-time applications used in IoT, analytics, and cloud services with less coding and time. It was initially created to handle the vast volume and massive sources of time-stamped data generated by infrastructure, applications, and especially from sensors [4].

Apache NiFi: a web-based tool with data integration and data logistics for automating the activities of data cross systems. It offers real-time control that easily and seamlessly manages the data between destinations and sources [5].

Grafana: a platform is widely utilized in monitoring infrastructure and log analytics. It is mostly used to improve the corporate's operational efficiency. Its dashboards provide tracking users and events effortless as it automates the management, collection, and viewing of data [6].

Elasticsearch: Elasticsearch is a distributed, RESTful search and analytics engine capable of addressing a growing number of use cases. It centrally stores the data for fine-tuned relevancy, lightning-fast search, and powerful analytics that ease the scalability [7].

3. Proposed Data Flow

3.1. Motivation

As in the manufacturers' case, they build and utilize MES in an on-premise environment using an MSSQL database. Fig. 1. shows the data flow of MES. Data flow is within the scope of data collection, data storage, and data consumption services. Each flow requires a variety of technologies that companies can choose based on business requirements and expertise. It should also be able to be converted into insightful information for various business purposes.

Manufacturing data is stored in the MSSQL database as indicated by the data flow. In addition, sensor data installed on the equipment is stored in MSSQL for one day in a database and used for real-time monitoring. In manufacturers, as the number of sensors attached to facilities has increased rapidly, their data has also increased. It is not easy to process effectively with existing databases. To solve this problem, companies implemented a method to delete the data in one day and store records in the MSSQL database a day later. For sensor data, the MSSQL database, which is an RDB, has limitations in performance. When sensor data increases, bottlenecks in the existing

database cannot be avoided, and the manufacturers are considering limitations in cost and performance.

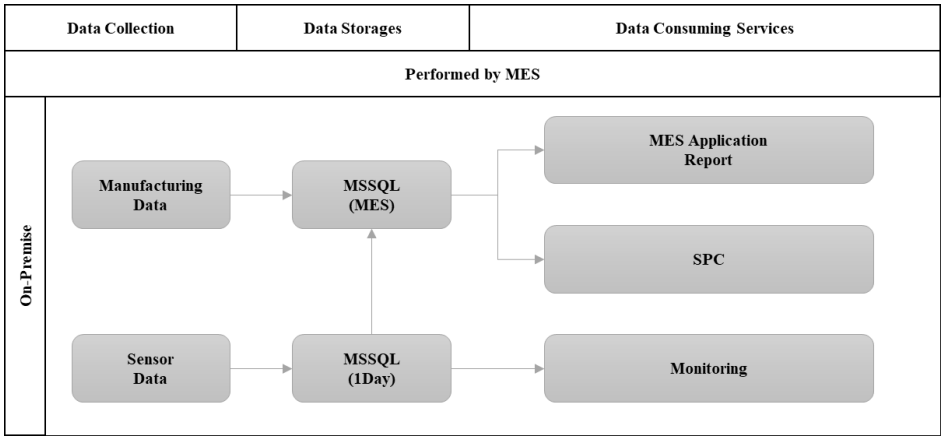


Fig. 1. Current Manufacturers' MES Data Flow

Manufacturers are considering applying a real-time analysis with the cloud platform to solve and upgrade the problems of existing MES solutions. They intend to predict and optimize process conditions through remote monitoring and AI real-time analysis in a cloud environment. Existing MSSQL databases must be automatically migrated to other database management systems such as TSDB and Elasticsearch to analyze real-time patterns using cloud platforms. The TSDB index is optimized for data accumulated over time. Therefore, as the data amount increases, and it shows fast processing speed over time. It is convenient with a retention policy that can automatically delete previous data. We propose a method that integrates the cloud systems connecting with the TSDB to upgrade into smart factories while still maintaining the manufacturers' MES systems.

3.2. Proposed Solution

In the proposed solution, sensor data is collected in real-time from the MSSQL database and automatically migrated to TSDB using Apache NiFi, as shown in Fig.2. After searching for existing database data, it is linked and integrated into TSDB and Elasticsearch in real-time according to the designed processes. Data stored in TSDB is used remotely for real-time monitoring and analysis, while integrated data linked to Elasticsearch is used for AI analysis and process optimization prediction. Using Apache NiFi, existing legacy system data can be integrated and linked to cloud platforms. Meanwhile, real-time analysis can identify problems with defective products or machines and control them not to move to the following process at some point to complete process optimization.

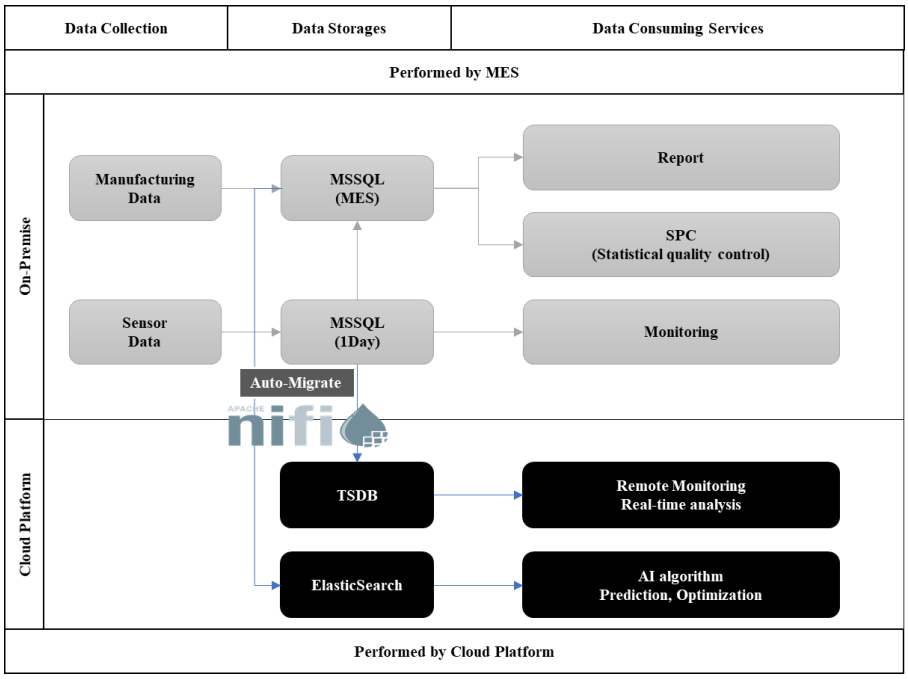


Fig. 2. Data flow architecture of the proposed solution

In this design process, Apache NiFi collects sensor data in real-time and integrates different databases. This tool provides a web-based, seamless user interface for data engineers to build robust data distribution processes [5]. The tool also supports configuration, which depends on the stream-based programming model, and provides highlights which contain the capability to perform inside groups, security using TLS encryption, extensibility. It also offers a highly configurable and modifiable data flow process that can modify data at runtime.

4 Implementation and Result

From the sensors, there are three kinds of databases to retrieve and put into the MSSQL, as shown in Fig.3. A_DATATABLE is used to store comprehensive sensor information such as its name, zone, location, description, and usage. DP_DATATABLE is created to store the pressures' data. Meanwhile, S_DATABALE is used to collect the temperature, humidity, and vibration from sensors. Notably, the value of the data is stored in the time-series type, which helps the users to understand the specific data change over time [8].

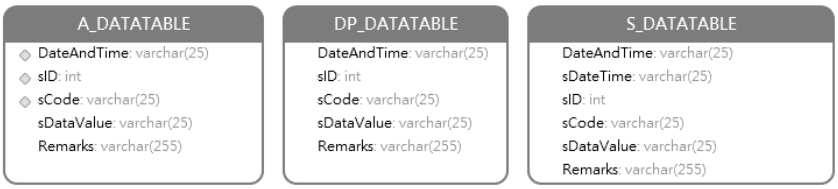


Fig. 3. Data tables from sensors stored in MSSQL

Each data table is designed with the same group process, as shown in Fig. 4. The first process is creating the query from MSSQL. The data was retrieved in the Binary Avro format by using the **AvroRecordSetWriter**. Since records retrieved are in batches, they need to be split using into multiple smaller **FlowFiles**. The next process converts the split Avro records into JSON and inserts the data right into the **Elasticsearch** in the following process. As TSDB can only record one value type, some fields are needed to convert its value type to put a record in the database.

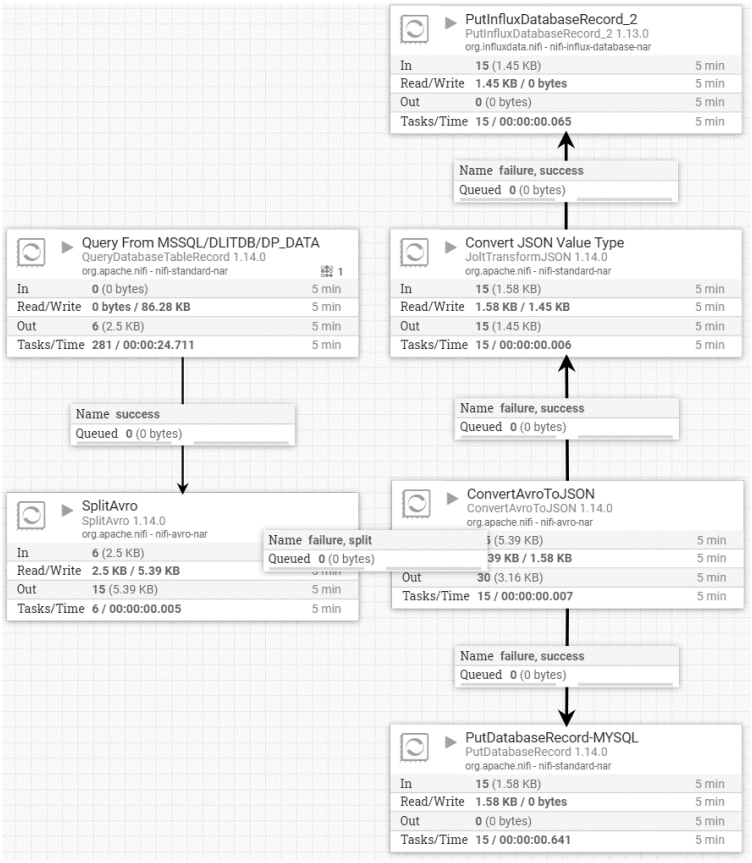


Fig. 4. Process of pulling data from DP_DATA in MSSQL into Elasticsearch and TSDB

5 Conclusion

This approach proposed Apache NiFi to solve the problems of large-capacity, real-time collected sensor data of existing MES solutions and solve the difficulties of real-time decision-making by integrating data into an environment capable of real-time analysis. Apache NiFi provides a standard data flow design that enables automatic data loading to multi destinations. Using this data flow in conjunction with the cloud computing platforms enables remote monitoring and real-time decision-making through web services in a cloud environment. The flow of linked and integrated data was displayed and visualized in a dashboard using Grafana. It was confirmed that the results showed satisfactory performance in time series pattern analysis and real-time monitoring of the collected data in seconds.

Fig. 5. The data access through Grafana

Acknowledgment. This study is based on research funding from the Ministry of Trade, Industry, and Energy and the Korea Institute of Industrial Technology Evaluation and Management (KEIT) in 2018 (No. 20001798).

References

1. Bloter, https://www.bloter.net/newsView/blt202003010004
2. Mathrani, S., Lai, X.: Big Data Analytic Framework for Organizational Leverage. Applied Sciences 11, 2340 (2021)
3. Microsoft, https://www.microsoft.com/en-us/sql-server/sql-server-downloads
4. InfluxData PLC., https://www.influxdata.com/time-series-database/
5. Apache, https://nifi.apache.org/index.html
6. Grafana Labs, https://grafana.com/
7. Elastic, https://www.elastic.co/elasticsearch/
8. Dix, P.: Why Time Series Matters for Metrics, Real-Time Analytics and Sensor Data. An Influxdata Technical Paper. InfluxData, https://www.influxdata.com/sensor-data-is-time-series-data/ (2021)

A Design and Implementation of Food-Safety Information System for Overseas Companies

Yang Guang[1], Kyung-Hee Lee[3],
Saksonita Khoeurn[2], Vungsovanreach Kong[2], Saravit Soeng[2], Wan-Sup Cho[1],

[1]Dept. of Big Data, Chungbuk National University, Chungdae-ro, Seowon-gu, Cheongju-si Chungcheongbuk-do, Korea
[2]Dept. of MIS, Chungbuk National University, Chungdae-ro, Seowon-gu, Cheongju-si Chungcheongbuk-do, Korea
[3]Data Department, Healingsoft Co., Ltd, Seoul, Korea
[1] yangguang1738@qq.com, [2] {saksonita, kv.sovanreach, soengsaravit, wscho63}@gmail.com, [3] lee.kyunghee@gmail.com

Abstract. We propose a design and construction plan for an overseas food safety big data system that continuously collects, stores, and manages address and food safety-related information. Considering that about half of overseas food companies registered in the Ministry of Food and Drug Safety are Chinese companies, our system is targeting China at first and then applied it to other countries.

Keywords: Address Verification, Food-Safety, Google Geocoding

1. Introduction

With the acceleration of the global economy, such as FTAs and e-commerce, the number of export and import-related companies is rapidly increasing. For example, about 100,000 overseas companies in 200 countries export food to Korea, and accurate information management for these companies is emerging as an important issue for food safety. The Ministry of Food and Drug Safety is accumulating and managing a variety of information to ensure the safety of imported food. However, the addresses registered with the Ministry of Food and Drug Safety contain many errors and often do not follow the standard address format. This address information error causes a delay in business processing and confusion in food safety management and is a major factor in reducing work efficiency. Accurate address information for overseas businesses that export food to Korea is the most basic information for food safety management. Furthermore, food safety-related SNSs or news about these companies or products are important information, and environmental information near the address of these companies is also important information that affects food safety.

In this paper, we propose a design and construction plan for an overseas food safety big data system that continuously collects, stores, and manages address and food safety-related information. Considering that about half of overseas food

companies registered in the Ministry of Food and Drug Safety are Chinese companies, our system is targeting China at first and then applied it to other countries.

2. System structure

2.1 Food Safety Data Types

We propose the 'Imported Food Safety Big Data System(: IFSB)', a big data system that continuously collects and accumulates various information related to the safety of imported food and analyzes it on the map to obtain direct insights. IFSB visualizes and provides comprehensive information on food safety on the map, such as the location of the company on the map, the existence of the company, surrounding environment information for the company, SNS, and news information, centering on the address of the foreign food company. Address verification of food companies has been done by a separate address verification system and utilizes various address databases and artificial intelligence technologies to improve accuracy and standardize address formats. Imported food safety information accumulated in the system so far is as follows.

- Overseas food companies information: 45,000 Chinese companies registered in the Ministry of Food and Drug Safety's Imported Food Information Maru (https://impfood.mfds.go.kr/)
- World real-time air quality information: https://aqicn.org/map/world/kr/
- Food safety-related SNS, news information: https://m.weibo.cn/

2.2 Food-Safety Information System (FSIS)

FSIS starts with the problem of verifying the English company name and English address of overseas food companies and making accurate address information into a database. Addresses of imported food companies can be accurately registered using various address DBs, and inaccurate addresses are corrected by suggesting an alternative address to the user. The next step is to display the address on the map so that the user can visually see the overseas business. In addition, various food safety information should be added centered on the address on the map. Typically, information related to food safety is added by analyzing SNS and news information about the company address. Currently, information is collected from China, but it is planned to expand to countries with large agricultural and food imports such as Japan, the United States, and Chile in the future. Then, environmental information that has a great impact on food safety is collected and visualized by overlaying it on the map. Figure 1 shows this process. Various information related to food safety is visualized on a map of China, and users can see various information related to food safety on the map.

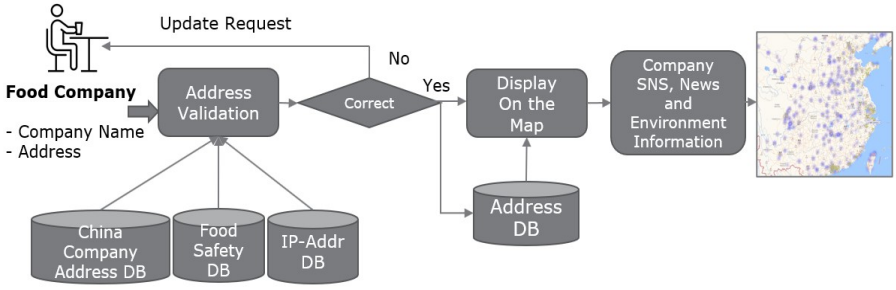


Fig. 1. Overall System Architecture.

2.3 System Implementation

FSIS(Food Safety Information System) verifies the addresses of about 45,000 Chinese food companies exporting to Korea and displays the addresses and food safety-related information overlaid on the map. The address verification system includes global address databases, fine dust information, and news/SNS information collection & analysis system, Google Geocoding API, Baidu map API. Figure 2 shows the UI screen of the system.

3. Conclusion

In this paper, we proposed FSIS(Food Safety Information System), a big data system to continuously strengthen the safety information of imported food. The system first verifies the address of a Chinese company that exports food to Korea, then the address is visualized on a map. Based on the qualified addresses, company information (company business registration certificate, company phone number, sales, etc.), environmental information, address environment-related news and SNS information were collected and visualized on the map. The user can see various information related to the safety of imported food in connection with the map so that the efficiency of food safety-related business can be improved.

Acknowledgment. This work was funded by the Ministry of Food and Drug Safety(MFDS, Korea)[Project Number: 21163MFDS517].

References

1. https://m.weibo.cn/
2. https://lbs.amap.com/
3. https://impfood.mfds.go.kr/

Fig. 2. User Interface of the System.

Automatic Classification of i-Ceramic Green Sheet Images using Deep Learning

In Joo[1], Kwan-Hee Yoo [1*]

[1] Dept. of Computer Science, Chungbuk National University, South Korea
{jooin95, khyoo}@chungbuk.ac.kr
*Corresponding Author

Abstract. Generally, an i-ceramic system is a ceramic industry manufacturing innovation platform using artificial intelligence. This platform has an artificial intelligence function that provides optimal production conditions for materials and production processes through big data analysis anytime, anywhere, and automatically inspects product defects. The casting slurry from the tape casting process in the ceramic processing processes is what we call green sheet. In the process, we used deep learning to inspect the number of defects in green sheets , we got a result over 98%. And it did not end there, and additionally, a technique was introduced to check where the defects are in one defective sheet. As a result, it is believed that high-quality ceramics can be produced by analyzing various environmental factors that affect the process.

Keywords: Greed Sheet, Mask-RCNN, U-Net, VGG16, Faster-CNN, Machine Learning, CNN, Image Detection

1 Introduction

Recently, smart factories are maximizing process efficiency and management through data collection and analysis. The i-ceramic operation also plays an important role in controlling high-quality products in the development of innovative ceramic materials[1]. One of the important issues in i-Ceramic is to classify defective products and good products according to each environmental factor.

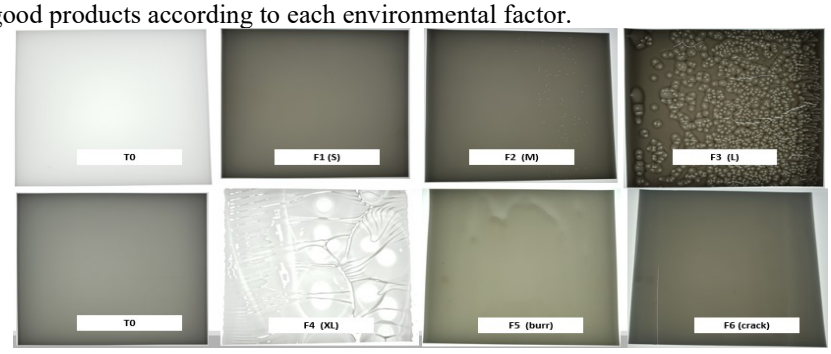


Fig. 1. Defects of i-ceramic images in ceramic-made products

Defect types in a green sheet include burrs, cracks, and bubbles as shown in Fig. 1. A total of 1780 14500x3000 images with several defects is collected. In the case of burr, it is a problem that occurs at the beginning and the end of the product, so it can be solved in the image preprocessing part. However, since cracks and bubbles are created in the middle of ceramic products, we adopted a method of managing them using deep learning.

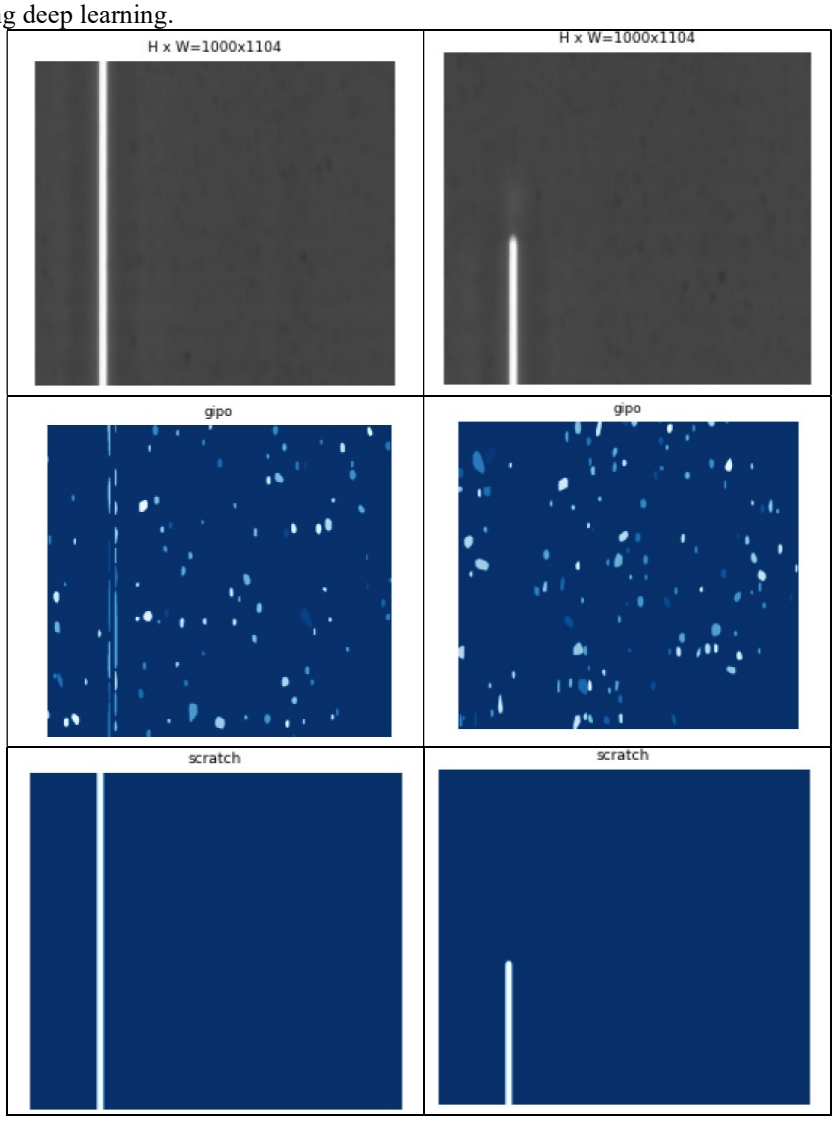


Fig. 2. There are three types of data types in the green sheet. One is the background of the blue part, and the second is the shape of the bubble displayed in the middle. The last one is the scratch in the photo on the right.

Fig. 2 separates actual scratches and bubbles. Even within one sheet, there is not only one defect state, but several defects can appear at once. We wanted to check not only the presence or absence of defects but also the number of defects per sheet.

There are a lot of researchers on them. First of all, Min, *et.al* [2] proposed a RESNET to detect discriminate defective and good products on ceramic images. And Suh *et. al.* [3] proposed a self-developed CNN (Convolutional Neural Networks) model to detect the determination of coating bolt. Even if the results are proposed, they cannot be applied yet in the real fields when the accuracy measure is considered. So, this paper proposes a hybrid CNN model of mask-RCNN (Region with Convolutional Neural Network) and Faster-RCNN that can detect the defective state of production products using deep learning technology that shows good performance in solving classification problems by utilizing images and images from various fields. Generally, they are very small defect elements that can only be resolved with the naked eye of a professional, but by reducing the feature-catching layer so that the features do not disappear, we have made better defect identification even in low-resolution images.

2 Proposed method and results

In this section, we propose a method for classifying defects in ceramic products according to the procedure shown in Fig. 3. Since the image sheet is too large at 14000 * 3000, we did an ROI job that cuts out the parts that are needed to normalize the image data. Then, to more accurately compare the features of the images, the images were divided into 12 equal parts and inserted into various models for comparison. The data set was divided into 80% and 20% percent for training and testing, respectively. We did this by comparing several models. The classification was carried out by changing various backbones in mask-RCNN and Faster-RCNN.

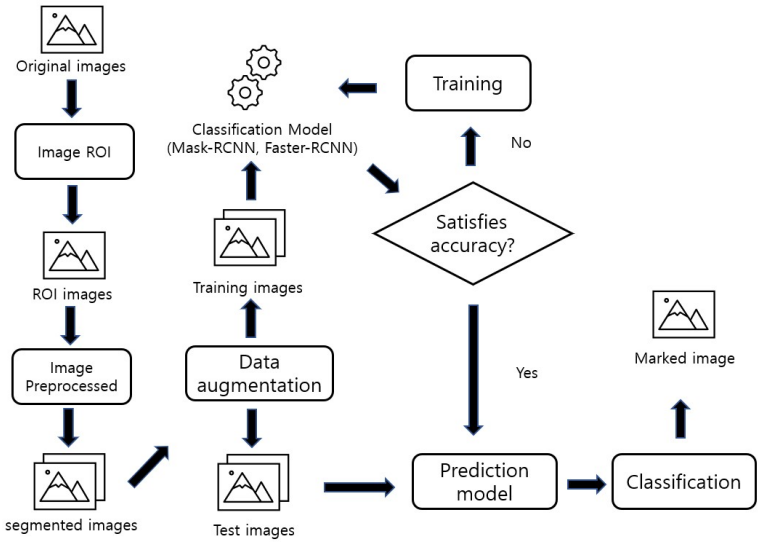
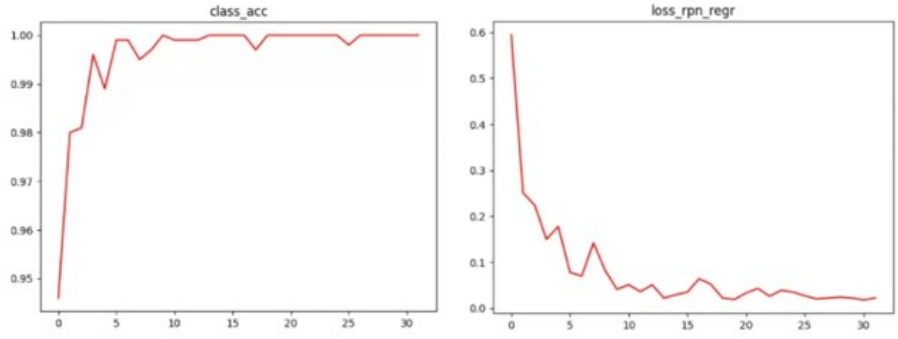


Fig. 3. The overall approach for detecting defects in i-ceramic images.

There are three characteristics of defects. The first type is the background, the second type is bubbles, and the third is scratches. Among the divided images with only a background are classified as good, and images with bubbles or scratches are classified as defective and features are marked. The correct answer rate is calculated by comparing the images classified in that way with actual data.

3 Conclusion and Future work

The criterion for good or bad is to measure the green sheet for air bubbles and scratches. When we apply the model, we can derive two main results. The first result is the classification of defective products and good products. As you can see in Fig 4, the model we used was able to catch up to 99% of the bad and good parts.



	precision	recall	f1-score	support
0	0.95	0.98	0.96	95
1	1.00	0.99	1.00	706
accuracy			0.99	801
macro avg	0.97	0.99	0.98	801
weighted avg	0.99	0.99	0.99	801

Fig 4. Classification report of predicts for defective and good products.

However, the performance of the model for locating defects is poor. Table 1 below shows the experimental results using some models. The scratches are large and the model holds it well overall. However, you can see that the bubbles are so small that they are not captured properly as the number of layers increases. If you look at the results, you can see that Mask-RCNN using ResNet101 backbone with many layers has the lowest accuracy, and Mask-RCNN using the other ResNet50 backbone has low accuracy. Among the models implemented so far, Faster-RCNN using the Vgg16 backbone has come out with the highest accuracy.

Table 1. A model that determines the location of the defect in the image caught as a defect

Back Bone	Model	accuracy (%)

ResNet101	Mask RCNN	34.58
ResNet50	Mask RCNN	41.64
Vgg16	Faster RCNN	77.30

To effectively distinguish small images, I think that we need to highlight features or find them using fewer layers. In the process, if we can distinguish not only good products from bad products but also the number of defects, we think we can more precisely control the environmental factors that make ceramics.

In the future, we will try to experiment again after highlighting the features using U Net. We will see how effectively we can detect air bubbles by lowering the number of layers in the model.

Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2020-0-01462) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation) and by the Ministry of Trade, Industry & Energy (MOTIE, Korea) under i-Ceramic manufacturing innovation platform technology development business. No.2004367, 'Development of cloud big data platform for the innovative manufacturing in the ceramic industry.

References

1. R. Jose and S. Ramakrisha, "Materials 4.0: Materials big data enabled materials discovery," Applied Material Today, vol.10, pp127-132, 2018.
2. Min, Borin, et al. "Abnormal Detection and Classification in i-Ceramic Images." 2020 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2020.
3. Sang-Won Suh, et al. " Development of Checker Switch Failure Detection System Using CNN Algorithm." Journal of the Korean Society of Mechanical Engineers 18.12 (2019): 38-44.
4. Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015): 91-99..
5. Ammirato, Phil, and Alexander C. Berg. "A mask-rcnn baseline for probabilistic object detection." arXiv preprint arXiv:1908.03621 (2019).

Estimation of Machine Health Stability using LSTM

Dimang Chhol, Kwan-Hee Yoo

Computer Science Department, Chungbuk National University, South Korea
{dimangchhol, khyoo}@chungbuk.ac.kr

Abstract. Forecasting the machine health stability(MHS) is important to know the overall stability of a machine in smart manufacturing systems. This paper explores to know the reflection of alarm signals to estimate the MHS for a machine. After collecting production and machine information, including alarm signals, we apply the Long Short-Term Memory(LSTM) as the deep learning technique to estimate MHS.

Keywords: Machine Health Stability, Alarm Signals, LSTM

1 Introduction

Machine Health Management has become a critical part of this industry 4.0 for the Smart manufacturing system [1]. Various techniques for predicting failures include a predictive model generator, a monitoring agent that predicts every 10 minutes, and a dashboard to visualize by given prediction [2]. Moreover, in the industrial machine, the machine controller was embedded with a different type of alarm. Therefore, it provides some benefits in using those sensors and machine states to report to the end-user regarding the diagnostic or the need of maintenances sometimes can put to some defined mode included the shut down of the machine to prevent the problem occurring [3].

This paper helps recognize equipment reliability in the production process to increase quality production in the factory. In addition, It applied one deep learning technique called the Long Short-Term Memory(LSTM) with the alarm signal and production machine data to predict the machine stability in the following production line.

2 Related Study

This section discovered the previous works that talked about predicting machine health stability using Machine Learning Techniques. According to [4], The author used four types of machine learning techniques which are Support Vector Machine(SVM), Ridge, Random Forest, and K-Nearest Neighbor (k-NN). They gathered data from the factory that created the part of the car in South Korea. The first main point of their research is to extract features artificially from the machine such as

state, alarm, and machine production result called not good(NG) product. The experiment result found that the SVM model is more stable than other models.

In [5], the author proposed two stages prediction process to predict product failure. The first stage is clustering. During this stage, since some features are correlated, they used the Principal Component Analysis(PCA) to reduce the dimension of the features. Moreover, they were then using k-means to group data. The second stage is to apply supervised learning techniques to predict the fault product in each cluster. The outcome of the final model demonstrates that the grid search is used to optimize the parameters of random forest classifiers. According to the author, with the performance metric, the AUROC(Area Under the Receiver Operating Characteristic Curve) score ranges from 0.5 (random guess) to one (perfect classification). Since the output score of 0.69 indicated the difficulty of this production failure prediction problem.

3 Proposed Methods

In Smart Manufacturing System, the Programmatic Logic Control(PLC) periodically sends data when some problems occur. In this research, the data such as production, machine information, and alarm from 2019 to 2020 from the production cars parts in South Korea. Related to alarm, we used data from the three main properties alarm data for use in this experiment are Alarm frequency($Freq_{alarm}$), Alarm variety($Variety_{alarm}$), and Alarm rate($Rate_{alarm}$). Based on [4] , we defined the Alarm rate by the following equation:

$$Rate_{alarm} = \frac{\sum_{i=1}^{N} Time_{alarm_i}}{interval \times Variety_{alarm}} \quad \textbf{(1)}$$

The duration of each alarm in seconds represents by $Time_{alarm_i}$ so that output will be between 0 and 1. Whereas the interval divides the data into chunks, and the interval is 30 minutes.

This study defines the machine health stability level from 0 to 1, as shown in Table 1.

Table 1. Machine Health Stability Level [4].

MHS	Quality
0.85 – 1.00	Excellent
0.75 – 0.85	Good
0.65 – 0.75	Average
0.5 – 0.65	Poor
0.0 – 0.5	Terrible

Our preprocessed dataset [4] contains 5716 records after applying the correlation analysis. We selected only essential features: non-active rate, alarm rate, not good

product rate, defective product rate, failure probability, failure frequency, individual rate(rate i), moving range rate(rate mr), process capability(cp) and process capability index(cpk). Thus, we train 70% of the data and Test 30% of our preprocessed data.

To predict the machine health stability, we construct a basic learning model with LSTM cells. The LSTM parameter used in this prediction is shown in the following Table 2.

Table 2. LSTM Experiment Parameter.

Parameter	Value
activation	"relu"
loss	"mse"
optimizer	"adam"
epochs	25

4 Result

After applying some implementation following part 3, we got the root mean square error score: 0.02 RMSE. Fig. 1 shows the result after applying LSTM techniques to our dataset. The Y-axis predicts machine health stability, and the X-axis is the data set from 2019 to 2020.

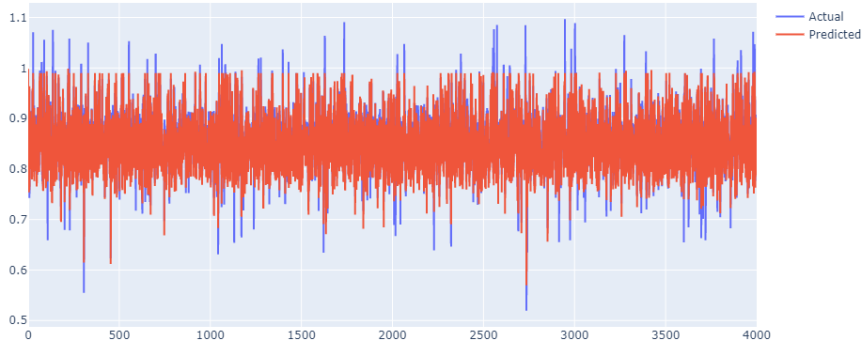


Fig. 1. The results of Predict Machine Health Stability.

After we test the same model compared to SVM, we see the result as in Table 3.

Table 3. LSTM compare to SVM.

Techniques	RMSE
LSTM	0.02
SVM (Linear)	0.0403

As we can see, the comparison result of Root Mean Square Error(RMSE) is lower than SVM (Linear). Therefore, since the LSTM technique has a lower result than SVM (Linear), we can conclude that the LSTM technique is a better estimator.

5 Conclusion and Future Work

This research study provides helpful info for forecasting machine health stability from the preprocessed data to create an LSTM model and then Calculate Root Mean Square Error (RMSE) as an estimator. Furthermore, the graph compares the actual and predicted stability with the deep learning technique using a simple model of LSTM. Moreover, by comparing with SVM (Linear), we found that LSTM has a better result. Finally, throughout this foundation result, We have an idea to extend further to apply it to more machines in the factory.

Acknowledgments. This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2020-0-01462) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation) and by the Ministry of Trade, Industry & Energy(MOTIE, Korea) under i-Ceramic manufacturing innovation platform technology development business. No.2004367, 'Development of cloud big data platform for the innovative manufacturing in ceramic industry.

References

1. Lee, G.Y., Kim, M., Quan, Y.J., Kim, M.S., Kim, T.J.Y., Yoon, H.S., Min, S., Kim, D.H., Mun, J.W., Oh, J.W., Choi, I.G., Kim, C.S., Chu, W.S., Yang, J., Bhandari, B., Lee, C.M., Ihn, J.B., Ahn, S. H.: Machine health management in smart factory: A review. J Mech Sci Technol 32**,** 987–1009 (2018)
2. Canizo, M., Onieva, E., Conde, A., Charramendieta, S., Trujillo, S.: Real-time predictive maintenance for wind turbines using Big Data frameworks, 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), pp. 70-77 (2017)
3. Langone, R., Alzate, C., Bey-Temsamani, A., Suykens J.A.K.: Alarm prediction in industrial machines using autoregressive LS-SVM models, 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), pp. 359-364 (2014)
4. Sadirbaev, B.: Prediction of Machine Health Stability in Smart Factory Systems using Machine Learning (2021)
5. Zhang, D., Xu, B., Wood, J.: Predict failures in production lines: A two-stage approach with clustering and supervised learning, 2016 IEEE International Conference on Big Data (Big Data), pp. 2070-2074 (2016)

Location Accuracy Measurement of Printing and Punching Products of Ceramic Smart Factory System

Sung-hoon Kim[1], Ga-Ae Ryu, Kwan-Hee Yoo
Dept. of Computer Science, Chungbuk National University, South Korea
{sidsid84 garyu,,khyoo}@chungbuk.ac.kr

Abstract. I-ceramic is a manufacturing innovation platform for the ceramic industry that combines the ceramic manufacturing process, big data, and artificial intelligence. The i-ceramic manufacturing process consists of material preparation, ball milling, tape casting, sheet cutting, punching, and printing in order. Punching & printing are the process of making holes and printing the sheet produced by the previous processes. In this process, products are repeatedly made, and quality inspections are conducted on these products. At this time, the important thing is positioning accuracy, and after measuring it, it is necessary to feed back the error value to readjust the printing and punching operation settings. To this end, in this paper, an effective method for measuring positional accuracy based on image processing is presented.

Keywords: i-Ceramic, printing, punching, location-accuracy, image processing

1 Introduction

A ceramic is any of the various hard, brittle, heat-resistant and corrosion-resistant materials made by shaping and then firing an inorganic, nonmetallic material, such as clay, at a high temperature.[1] Because fine ceramics have excellent electromagnetic, optical, and mechanical properties, they are used for the PCB(printed circuit board) of IC (integrated circuit) chips.

However, the fine wiring of the PCB is subjected to unstable electrochemical conditions in a high temperature/high humidity environment, causing many reliability problems [2]. An attempt was made by Y. Deng, *et.al* [3] in relation to this problem. To improve the high false alarm rate, which is a problem of the existing automatic optical inspection (AOI), they proposed a technique for classifying printing defects using CNN. They were trained and tested with 4526 pieces of data organized into 5 categories, and their model contains 5 convolutional layers and 3 fully-connected layers. This study yielded very positive results, lowering the false alarm rate from 97% to 34%. However, there is a limitation in that it lacks a specific approach to errors and merely searches for defects to give feedback on print operation settings.

Therefore, this study intends to identify the accuracy of the printing position by using image processing technology, derive the error in pixel units, and verify the printing state.

2 Proposed method

There are two types of images to be used for analysis: a ground-truth base image(gt_base_image) as a standard and an input image as an actual print, and each image consists of printed lines and a punched hole using a laser puncher. There are a total of 5 lines, each with a thickness of 130,110,90,70,50. The diameter of the hole has 5 sizes with the same ratio. And, there are four circular points, which are the benchmark points for the location, were measured in advance, and the input image and gt_base_image were compared. A μm to pixel ratio was also derived through a complete survey of the input images. (Assume that the benchmark circular point of input images is always correct).

After that, the input image is matched through transform based on the benchmark of the ground-truth base image (gt_base_image), which can be a blueprint for each line and hole.

Now, as can be seen in the following Figure 2-1, after a series of processes, the positioning accuracy and print status are judged in parallel.

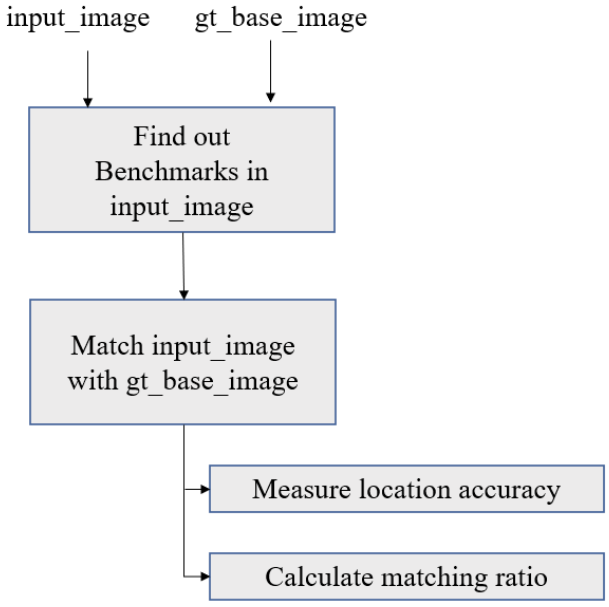


Figure 2-2 Analysis sequence in Measurement of location accuracy

2.1 Preprocessing of input images

gt_base_image consists of micrometers and input image consists of pixels. Since the analysis is performed in units of pixels, the calculation of the μm to pixel value must be preceded, and the μm to pixel value set in the preprocessing before analysis is maintained. In the following analysis processes, scaling is not used at all to prevent distortion. If the AOI size of the input image does not change, μm to pixel is also maintained.

2.2 Matching input_image process

The benchmark point is located at the upper left, upper right, lower left, and lower right of the outermost region. And, we found the coordinates of the benchmark points in the input image by using the Hough-gradient-circle detection method[4]. This algorithm traverses the edges of the circle and draws a straight line in the direction of the gradient of the pixel. As a result, the most accumulated pixel becomes the center point of the circle.

After matching the upper left benchmark of gt_base_image and input_image through translate, calculate the angle through arcTan operation with the upper right benchmark point. We adjust the input image to be horizontal by performing rotation to make the derived angle 0 degree.

2.3 Measure location accuracy

The gt_base_line shown in the Figure 2-3 is a line in the correct position, and the line of the actually printed in input_image does not exactly match this. Therefore, it is the measurement of positioning accuracy to find out how much error occurs.

There are three main methods to measure the location accuracy.

The first method of measuring the distance from the benchmark point to the both of edge spots will give inaccurate results because it is affected by the roughness of the printed line.

The method of measuring the center spot is also not used because it is possible only under the assumption that the line is symmetrical about the center.

The third method is to derive the position by performing template matching on the input image and the line of gt_base_image. Template matching is robust to rotation and scaling and is not affected by reconstruction algorithms [5]. Therefore, it is a well suited algorithm to perform while the size and angle of the image are fixed through the benchmark point, and it does not contradict the above(there is no scaling in the analysis process).

As a result of the process, it is possible to find the specific location with the smallest error with comparing gt_base_image. This specifies the position of the printed location.

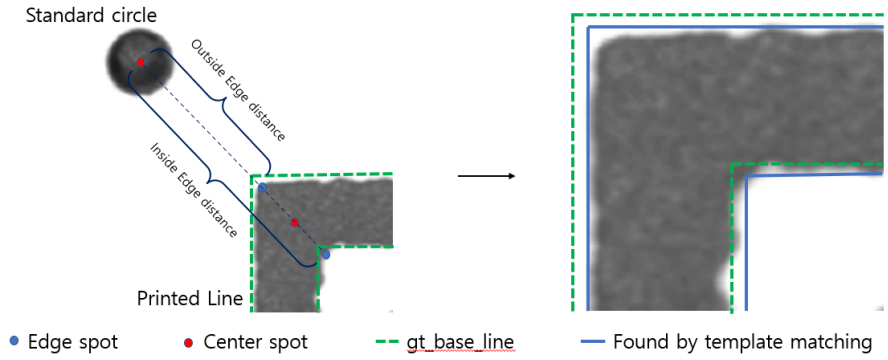


Figure 2-2 Method of finding location with printed line

2.4 Calculate matching ratio

Through the previous process, the printed location should be specified. And then, it was in a state where 1:1 comparison with the gt_base_image was possible. By using Mask, the overprinted part can be defined as “overprinted” compared to the image and the underprinted part can be defined as “lessprinted”. However, this operation was performed assuming that the positional error was 0, that is, the state was printed in the completely correct position. We have to divide and derive the error in that way, so that we can give the correct feedback to the print setting. Calculate the ratio of lessprinted and overprinted to the total number of pixels in the corrected print part of the gt_base_image, and use it as an index to express the Statement of printing.

3 Result

Table 3-1 Result of analysis

sample number	μm_to_pix	x_axis error	y_axis error	overprinted pixel	lessprinted pixel	over_ratio	less_ratio
1	3.700	7	-3	28	35681	0.02%	22.02%
2	3.703	6	-4	1707	16949	1.05%	10.46%
3	3.700	9	0	21	33890	0.01%	20.91%
4	3.704	5	1	1715	17500	1.06%	10.80%
…	…	…	…	…	…	…	…
17	3.700	2	-4	51	35464	0.03%	21.89%
18	3.699	1	-2	1498	15966	0.92%	9.85%
19	3.702	0	-2	18	31654	0.01%	19.53%
20	3.703	5	-6	4619	12440	2.85%	7.68%
21	3.698	1	-5	86	36172	0.05%	22.32%

The same analysis was performed for a total of 21 input images as shown in the . Table 3-2. With μm to pix, a fairly stable result is derived, and it is converging to 3.7. This can be interpreted as a correct operation in the detection of the benchmark circle. Regarding the positioning accuracy, an inconsistent error of less than 10 pixels is being measured, which seems to be due to the setting of the print or the difference in the position of the print sheet. It was measured that there were many lessprinted areas overall in the above 21 samples. In some cases, the printing line was cut off. In this case, attention is required because the path of electrons inside the chip is cut off.

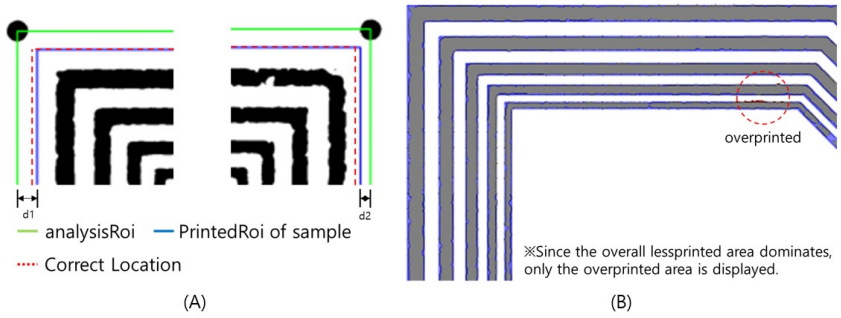


Figure 3-1 Example of result : location accuracy (A) and measurement of print statement (B)

Figure 3-2 is a example of Sample number 3, it was analyzed that it was printed at a position moved 9 pixels in the positive direction of the x-axis. In the above figure (Figure 3-3-(A), analysisRoi is the area set based on the benchmark circle analyzed in step 2-1, and printedRoi is the area in which the print area of the actual input image is detected including the blank space. Also, the correct location is the correct print area detected based on the gt_base_image. When this input image is enlarged, there is a difference in length between d1 and d2 so that it can be identified with the naked eye.

The difference is the distance from the correct location, and the directions are opposite. This figure is also the result of the same sample3. The part expressed in blue means lessprinted, and the part in red means overprinted.

4 Conclusion and future work

As a result of using template matching, it was possible to find the print position in pixels without being affected by the roughness of the print line.

This will present a standard that can increase the positional precision in the printing of the PCB.

However, the problem of assumptions still remained. The part that assumes that the benchmark point of the input image is “always correct” may cause distortion.

Therefore, if the number of data samples increases in the future, we plan to continue our research in the direction of improving location accuracy through learning of image processing deep learning without using a benchmark point.

Acknowledgments.

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2020-0-01462) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation) and by the Ministry of Trade, Industry & Energy(MOTIE, Korea) under i-Ceramic manufacturing innovation platform technology development business. No.2004367, 'Development of cloud big data platform for the innovative manufacturing in ceramic industry.

References

1. Heimann, Robert B. (16 April 2010). Classic and Advanced Ceramics: From Fundamentals to Applications, Preface. ISBN 9783527630189. Archived from the original on 10 December 2020. Retrieved 30 October 2020.
2. Seok-hwan Heo. "Biased HAST Prediction of printed circuit board life in tests."
3. Journal of the Korean Welding and Bonding Society 36.2 (2018): 1-6.
4. Y. Deng, A. Luo and M. Dai, "Building an Automatic Defect Verification System Using Deep Neural Network for PCB Defect Classification," 2018 4th International Conference on Frontiers of Signal Processing (ICFSP), 2018, pp. 145-149, doi: 10.1109/ICFSP.2018.8552045.
5. Russ, John C. The image processing handbook. CRC press, 2006.
6. Poonia, Poonam, Pawan K. Ajmera, and Vijayendra Shende. "Palmprint recognition using robust template matching." *Procedia Computer Science* 167 (2020): 727-736

Data Splitting for Prediction of Ceramics Processes using K-Mean Clustering

Jeong-Hun Kim[1], Zhu Yifan[2], Aziz Nasridinov[1]

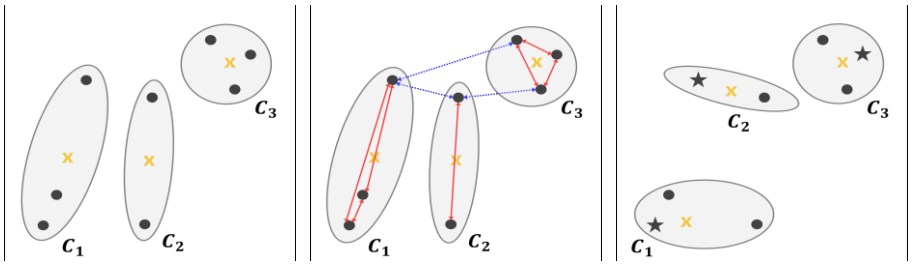
[1] Department of Computer Science, Chungbuk National University
Cheongju 28644, South Korea
[2] Department of Big Data, Chungbuk National University
Cheongju 28644, South Korea
{etyanue, yifan, aziz}@chungbuk.ac.kr

Abstract. Data splitting is one of the basic preprocessing steps for building prediction models in smart manufacturing applications. This approach is carried out before constructing the prediction models. The prediction models mainly use well-known algorithms such as random selection (RS) and Kennard-Stone (KS) for data splitting. However, these algorithms have limitations. Specifically, it is challenging to generate the training set of representative data samples that lead the prediction model to be appropriately learned as they tend to split the dataset to be biased. In the ceramic industry, we usually deal with a small amount of data due to the high complexity of the processes. Therefore, it is more difficult to split small datasets that derive from ceramic processes. This paper proposes an optimized centroid (OC) algorithm that improves data splitting for small datasets using K-Means clustering. We evaluate the OC algorithm by comparing the prediction accuracy with existing data splitting algorithms using a ceramic process dataset in the real-world.

Keywords: data splitting, small dataset, k-means, clustering, ceramics

1 Introduction

Data splitting is the process of dividing a dataset into two subsets, i.e., training and testing. The training set generated in the data splitting step is used to build prediction models. Therefore, the prediction models and their accuracy change according to the data splitting strategy [1]. Representative data splitting algorithms are Random Selection (RS) and Kennard-Stone (KS) [2]. The RS randomly splits a dataset. On the other hand, the KS splits the dataset uniformly by iteratively selecting pairs of the most dissimilar data samples based on the Euclidean distance. However, in these algorithms, if given data samples are biased, the result of data splitting is also biased, which may cause not proper training of the predictive model. In particular, this problem is critical for a small dataset collected from ceramic processes since the prediction model cannot learn on sparse data samples when the small dataset is biasedly partitioned [3]. In this paper, we propose a new data splitting algorithm, namely Optimized Centroid (OC).



(a) (b) (c)

Fig. 1. Example of Optimized Centroid (OC) algorithm. (a) Initial result of K-Means clustering; (b) Computation for optimized centroids; (c) Identification of representative data samples using the nearest neighbor.

To improve the accuracy of the prediction model, the OC generates a training set by identifying representative data samples from a small dataset using K-Means clustering. We evaluate the OC algorithm by comparing the accuracy of the prediction model with RS, KS, and Morais-Lima-Martin (MLM) [4] algorithms using a real-world ceramic dataset.

2 Proposed Method

Our OC algorithm consists of three steps. First, we utilize K-means clustering to find clusters in a dataset by the size of the training set. Each cluster has a centroid, where we assume that the centroid is the hypothetical data sample that best represents the cluster. However, since K-Means clustering does not always find optimal clusters, the identified centroids may not be representative data samples of the dataset. To find the optimal clusters, we further optimize K-Means clustering according to the basic principle of clustering. That is, clustering is based on the similarity between data samples, and the similarity between data samples of the same cluster (internal similarity) is maximized, and the similarity between data samples of different clusters (external similarity) is minimized. To achieve this purpose, we iteratively perform K-Means clustering to find optimal clusters in which the internal similarity is maximized, and the external similarity is minimized. After the K-Means clustering optimization step, we finally split the training and testing sets by considering the data sample that is the nearest neighbor of the centroid as a representative data sample. Fig. 1 shows a step-by-step example of the OC algorithm. Fig. 1(a) illustrates the initial K-Means clustering results for a given dataset. Here, the greater distance between the centroid of each cluster and the data samples, the greater error of the prediction model constructed by the centroid. Fig. 1(b) shows the process of optimizing K-Means clustering for the internal and external similarities based on the Euclidean distance. The result of optimizing K-Means clustering are shown in Fig. 1(c). Unlike the centroids identified in Fig. 1(a), the distance between the centroid and the data samples of the corresponding cluster is small. Therefore, it is possible to improve the error of the prediction model using the representative data samples identified from the optimized clustering result.

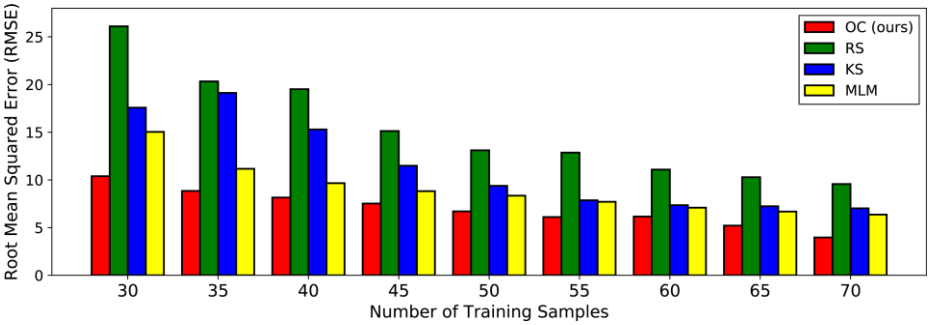


Fig. 2. The experimental results of each algorithm according to the number of training samples using the ceramic process dataset.

3 Experimental results

We conduct experiments to evaluate the effectiveness of the proposed Optimized Centroid (OC) algorithm on the predictive model. We compare the accuracy of the prediction model with existing data splitting algorithms Random Selection (RS), Kennard-Stone (KS), and Morais-Lima-Martin (MLM) using a real-world ceramic process dataset. The ceramic process dataset consists of 89 data samples collected in the green sheet process, each data sample having seven fields. The green sheet process consists of a ball mill machine that mixes ceramic slurry and a tape casting machine that manufactures a green sheet using this slurry. Specifically, the data fields include the amount of ceramic powder and organic additives used when mixing the ceramic slurry collected from the ball mill machine, and include the slurry viscosity, the casting speed, the height of blade gap, and the height of the ceramic slurry immediately after passing through the blade gap. collected from the tape casting machine. The model predicts the green sheet thickness. The accuracy of the prediction model is measured using the root mean squared error (RMSE).

Fig 2. shows the results of comparing the performance of data splitting algorithms for the number of training samples. All data splitting algorithms decrease the prediction error as the number of training samples increases. Our OC decreases the prediction error by about three times compared to existing algorithms. Specifically, OC has a relatively stable prediction error for a small dataset because the error reduction rate is the lowest according to the number of training samples. On the other hand, in KS, the prediction error is higher when the number of training samples is 35 than when the number of training samples is 30 because the data splitting is biased.

4 Conclusion

We have proposed a new data splitting algorithm called Optimized Centroid (OC) to improve the performance of prediction models in ceramic processes. We have evaluated the prediction accuracy of the proposed OC algorithm with existing data splitting algorithms using a ceramic process dataset in the real-world, which demonstrated the proposed method can achieve higher accuracy. In the future, we plan to evaluate the efficiency of the proposed algorithm on large datasets with multiple attributes.

Acknowledgments. This work is supported by the Ministry of Trade, Industry & Energy (MOTIE, Korea) under i-Ceramic manufacturing innovation platform technology development business. No. 2004367, 'Development of cloud big data platform for the innovative manufacturing in ceramic industry'.

References

1. Nguyen, Q.H., Ly, H.B., Ho, L.S., Al-Ansari, N., Le, H.V., Tran, V.Q., Prakash, I., Pham, B.T.: Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. Math. Probl. Eng. 2021, 1—15 (2021)
2. Kennard, R.W., Stone, L.A.: Computer aided design of experiments. Technometrics 11(1), 137—148 (1969)
3. Doan, Q.H., Le, T., Thai, D.K.: Optimization strategies of neural networks for impact damage classification of RC panels in a small dataset. Appl. Soft. Comput. 102, 1—18 (2021)
4. Morais, C.L., Santos, M.C., Lima, K.M., Martin, F.L.: Improving data splitting for classification applications in spectrochemical analyses employing a random-mutation Kennard-Stone algorithm approach. Bioinformatics 35(4), 5257—5263 (2019)

A Study for Defect and Classification of Ceramic Surface Images

SEONJONG BONG[1] , WONJUN LEE[2], SUYOUNG CHI[*],

[1 2] Korea University of Science and Technology, 217,Gajeong-ro,Youseong-gu, Daejeon, Korea,
(E-mail: sunjong@ust.ac.kr[1], wjdnjswns@ust.ac.kr[2])
[*] Electronics and Telecommunications Research Institute, Dajeon,305-370, Korea
(E-mail:chisy@etri.re.kr) *Corresponding author

Abstract. Machine learning is being applied to the entire industry and is being used in small and medium-sized manufacturing processes. In particular, many studies have been conducted on the detection of defects corresponding to simple intensive labor. If the experimental conditions are well made, good performance can be expected. We preprocessed the images based on these studies, classified good and defective products for ceramic products, and further classified the types of defects. After several experiments, the images were classified by selecting the sequential model of machine learning, and as a result, an accuracy of 98.5% was obtained.

Keywords: Ceramic surfaces, Machine Learning, Image Classification

1 Introduction

With machine learning being used across industries, machine learning is now a common word in manufacturing.[1] In particular, for the detection of defects on the surface of the product, if good/defective image data is sufficient, the performance of more than 95% can be achieved even using a simple machine learning technique. However, there are several reasons why it is difficult for small and medium-sized manufacturing companies to apply machine learning to their processes, one of which is that it is difficult to configure experimental conditions for collecting good/defective images.[2] We collected images of ceramic products using a high-resolution camera. These ceramic products are produced by pressing powder raw materials. If you classify and analyze various defects on the surface of this product, you can find out what kind of problem occurs in the press process. As a result of collecting and classifying images on the product surface, there are two types of defects. One of the two classes of defects is a crack and the other is contamination. Cracks are a phenomenon in which a part of the product falls off because the pressure of the pressure is not uniform as a whole. it will change color We classify images using machine learning techniques to classify good and defective products and identify the types of defects, and analyze the results.

2 Image preprocessing and learning model

2.1 Image preprocessing

The image of ceramic products has a resolution of 2240 * 2048 and a high-performance camera is used to collect product surface photos. In order to classify images of good and defective products by machine learning, some image processing was performed in advance. First, each data was well classified into good products, cracks, and Contamination. `

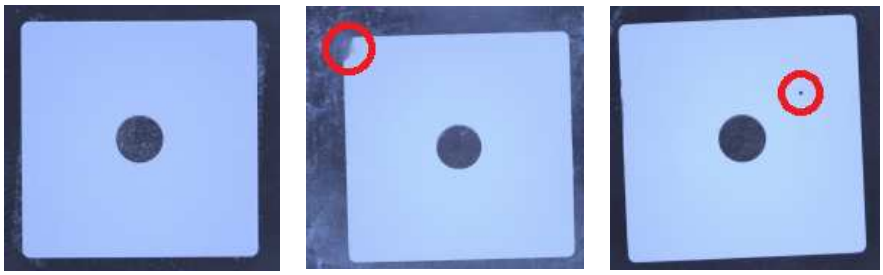
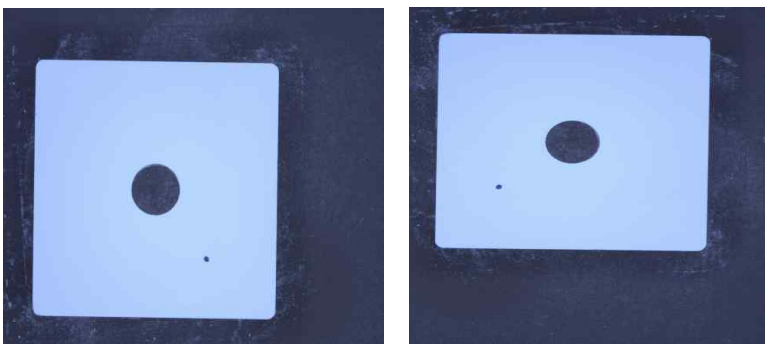


Fig. 1. Ceramic Product Images
(Left: Good, Middle: Crack, Right: Contamination)

The Fig.2 is a change in resolution.[3] High-resolution images have the advantage of being able to detect defective parts well, but also have disadvantages in that the amount of computation and density increase. It is necessary to classify good and defective products, and find a resolution that can classify each defect well. We conducted an experiment with images with resolutions of 10%, 20%, and 30% of the high-resolution image, and it was confirmed that the images (490, 410) with a resolution of 20% showed the best performance. Third, since we do not have many images of defective products, we had to train the model by augmenting the data as much as possible. For this purpose, horizontal and vertical flips were used among data augmentation techniques.



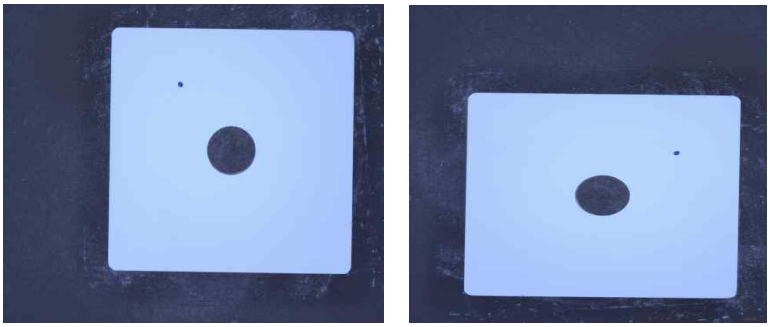


Fig. 2. Image Augmentation (Horizontal, Vertical)

2.2 Design of Learning Model

We created a classification model using a sequential model[4] using Keras Library provided by Tensorflow. The details of the model to be made with an image with a resolution of 20% are as follows. The compilation optimizer used 'Adam', and the provided 'SparseCategoricalCrossentropy' was used as the loss function.

Layer (type)	Output Shape	Param #
rescaling_1 (Rescaling)	(None, 490, 410, 3)	0
conv2d_3 (Conv2D)	(None, 490, 410, 16)	448
max_pooling2d_3 (MaxPooling2	(None, 245, 205, 16)	0
conv2d_4 (Conv2D)	(None, 245, 205, 32)	4640
max_pooling2d_4 (MaxPooling2	(None, 122, 102, 32)	0
conv2d_5 (Conv2D)	(None, 122, 102, 64)	18496
max_pooling2d_5 (MaxPooling2	(None, 61, 51, 64)	0
flatten_1 (Flatten)	(None, 199104)	0
dense_2 (Dense)	(None, 128)	25485440
dense_3 (Dense)	(None, 3)	387

Total params: 25,509,411
Trainable params: 25,509,411
Non-trainable params: 0

Fig. 3. Summary of Model

3 Experiment

The training dataset has 985 images consist of 785 good product images and 200 defective product images (100 Crack images, 100 Contamination images). This is the result of an experiment using images with 10% 20% 30% resolution using preprocessed images. Each epoch was learned 20 times.

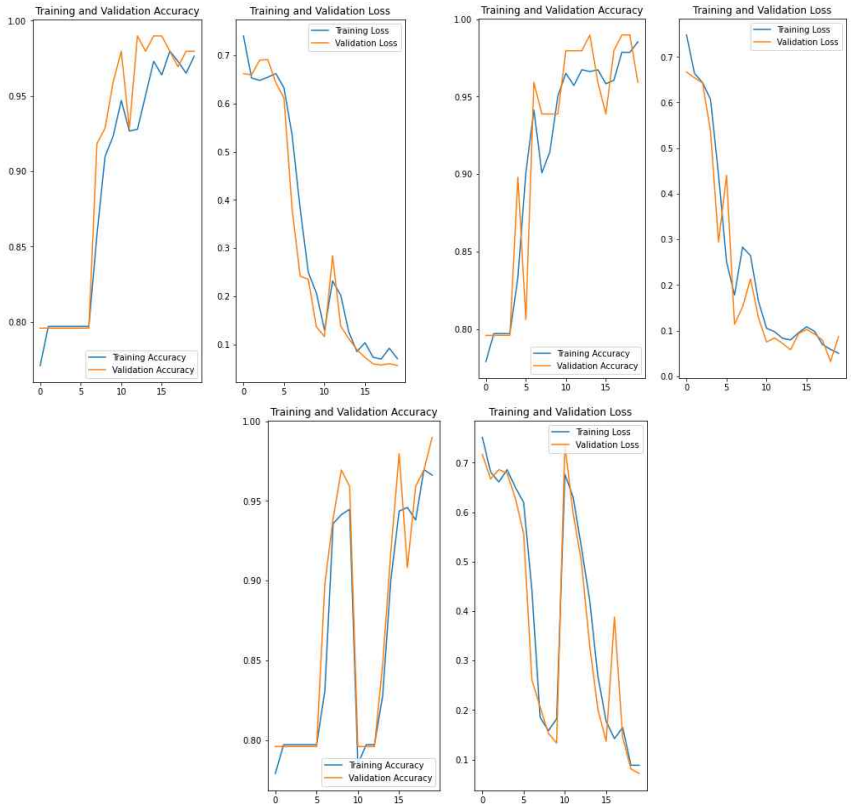


Fig. 3. Output of Learning

Table 1. accuracy of Learning

	10% Resolution (245*205)	20% Resolution (490*410)	30% Resolution (734*614)
Accuracy	97.6%	**98.5%**	96.6%
Validation Accuracy	97.9%	95.9%	98.9%

4 Conclusions

Methods for detecting defects on the surface of products and classifying good and defective products have been studied for a long time and are widely used in the field. We classified the surface of ceramic products by using the Sequential model of machine learning among various researched methods to classify the types of defects by dividing the surface of the ceramic product into good and defective products, and finally obtained the accuracy of 98.5 through several experiments. However, there is not enough learning image data, so we plan to conduct research using transfer learning[5] in the future to solve this problem.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Trade, Industry and Energy (MOTIE), Korea, under the i-Ceramic platform construction project- i-Ceramic manufacturing innovation platform No. 20004367, “Development of Cloud Big Data Platform for the Innovative Manufacturing in Ceramic Industry.”

References

1. Imoto, Kazunori, et al. "A CNN-based transfer learning method for defect classification in semiconductor manufacturing." 2018 international symposium on semiconductor manufacturing (ISSM). IEEE, 2018..
2. Elbehiery, H., A. Hefnawy, and M. Elewa. "Surface defects detection for ceramic tiles using image processing and morphological techniques." (2005).
3. Sabottke, Carl F., and Bradley M. Spieler. "The effect of image resolution on deep learning in radiography." Radiology: Artificial Intelligence 2.1 (2020): e190015.
4. Hutter, Frank, Holger H. Hoos, and Kevin Leyton-Brown. "Sequential model-based optimization for general algorithm configuration." International conference on learning and intelligent optimization. Springer, Berlin, Heidelberg, 2011.
5. Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." IEEE Transactions on knowledge and data engineering 22.10 (2009): 1345-1359..

Multi-target Learning on asymmetric U-Net for PNI boundary detection

Sangwon Lee[1], Youngjae Park[1], Jinhee Park[1], Giljin Jang[1], Hyemi Kim[2],

[1] School of Electronics Engineering, Kyoungpook National University,
80 Daehakro, Bukgu Daegu Republic of Koreas
{lsw0767, wim-c, pjhdrm, gjang}@knu.ac.kr
[2] Electronics and Telecommunications Research Institute,
Daejeon 34129, South Korea
miya0404@etri.re.kr

Abstract. Since perineural invasion (PNI) can have a poor prognosis, it is important to find out PNI as early as possible. However, detecting PNI in nerve cells from small scale optical images takes a lot of human efforts. Therefore, detecting perineural invasion with deep neural networks would be efficient and helpful in reducing human labor. In this work, we proposed an asymmetric network that can produce a segmentation map in level 2 resolution from level 0 scale. We also proposed a learning policy that utilizes nerves and tumors as targets for subtasks to help the network find PNI regions. Our method achieved about 0.42 for the evaluation metric over fifteen unseen whole slide images.

Keywords: Deep learning, CNN, Medical image segmentation

1 Introduction

Perineural invasion (PNI) is very important in diagnosing cancers because the cancer cells which are attached to the nerve can spread into surrounding normal tissue through the nerve [1]. PNI is hardly seen in non-malignant tumor cells, so pathologists use it as evidence of malignant tumors. Furthermore, PNI may result in a poor prognosis, so it's important to find out PNI as early as possible. However, detection of perineural invasion in nerve cells from small scale optical images takes a lot of efforts from the experts. Therefore, detecting perineural invasion with deep neural networks (DNNs) would be efficient and helpful in reducing human labor such as [2, 3, 4, 5].

In this work, we propose an optimized network with appropriate learning policy for the network that can make good use of all of provide annotation data. Based on the fact that PNI exist between nerve and tumor cell regions, we define the nerve and the tumor regions as sub-target and made the model to learn that in weak supervision. We design a network to produce a result in level 2 resolution, getting as much information as possible from level 0 resolution.

2 Method

2.1 Multi-Target Learning

The perineural invasion boundary, which is the target of this task, always appears with the tumor and nerve. Therefore, we hypothesized that if the model learns about the tumors and nerves, it can help find the invasion boundary. To do this, we utilize additional targets; nerves and tumors. Nerves are annotated in layer 1 as positive (exist) with local information, and in layer 3 and 4, annotated as negative (non-exist). But nerves also appear in layer 2. Similarly, in layers 1 and 4 there are tumors annotated as negative targets, whereas in layer 3 tumors are annotated as benign with no local information.

To take advantage of this variety of data with and without local information, we combined the supervised loss and the weakly supervised loss, layer by layer. Using model M to predict the nerve, boundary, and tumor in each channel, the supervised loss over nerves can be described as follow

$$L_{SN} = criterion\big(M(x_i)_j, y_{ij}\big), i \in \{1,3,4\}, j = 1 \quad (1)$$

where

x_i is the input patch from layer i

y_{ij} is the target patch for channel j, from layer i

$\mathrm{i} \in \mathrm{I} = \{1, 2, 3, 4\}$ is the index of annotation layer

$\mathrm{j} \in \mathrm{J} = \{1, 2, 3\}$ is the index of channel. Each channel represents nerve, PNI boundary and tumor, respectively.

The annotation layer index $i = 2$ is not used because there is no local information in y_{21}. Similarly, the supervised loss for tumors and boundary can be defined as follow

$$L_{ST} = criterion\big(M(x_i)_j, y_{ij}\big), i \in \{1,4\}, j = 3 \quad (2)$$

$$L_{SB} = criterion\big(M(x_i)_j, y_{ij}\big), i \in I, j = 2 \quad (3)$$

In the case of weakly supervised loss, we modified y to indicate the existence of nerve and tumor in layer 2. This can be done by adding y_{i2} to y_{i1} and y_{i3}. And we reduced the dimension of y by 2D-global-max function. We defined this modified target as $\hat{y}$ and then the weakly supervised losses for the nerve and tumor can be defined with global pooling function P

$$L_{WN} = criterion\left(P\big(M(x_i)\big)_j, \hat{y}_{ij}\right), i \in I, j = 1 \quad (4)$$

$$L_{WT} = criterion\left(P\big(M(x_i)\big)_j, \hat{y}_{ij}\right), i \in I, j = 3 \quad (5)$$

The final loss function can be defined as the sum of equation (1)~(5).

2.2 Network Architecture

A large dataset with three levels is used; level 0 is magnified to 16 times, level 1 is magnified to 4 times and level 2 is the original resolution. The output required for this challenge is the segmentation result in level 2, so conventional U-Net with symmetric structures could not perform well in this task.

Rather than employing the conventional structure, we optimized the structure for this task by making the model asymmetric. The encoder part of the network is same with a conventional one, but some decoder layers had been removed to keep the output size at level 2 resolution. Since we set the input size to level 0 resolution, the network has four more encoder layers than decoder layers.

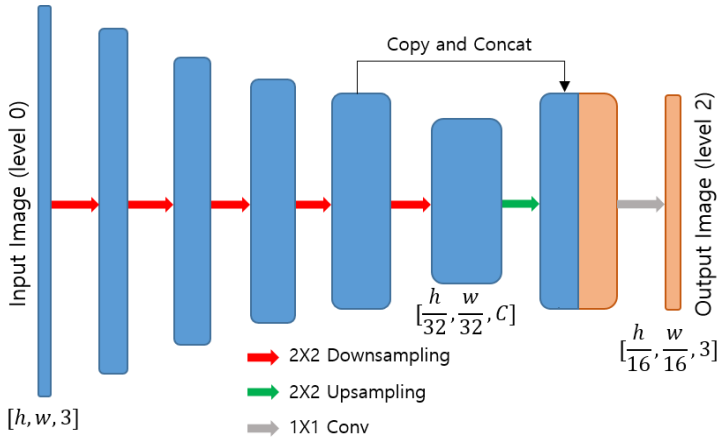


Fig. 1. The structure of proposed network. The input image is in level 0 resolution, so the model consists of five encoder layers and one decoder layer.

3 Results

For the experiment, we utilize [6] to implement our model with ImageNet pretrained efficientnet-b4 [7]. We trained our network with 135 whole slide images (from #1 to #45 for each organ). Then choose a submission model that performs best from the rest of whole slide images (from #46 to #50 for each organ). And we did some post-processing including thinning, thresholding and closing to create a clear line of thickness 1. Two types of thresholding, based on value and line length were performed. And all thresholding was done in line-wise, not in pixel-wise. The following figures show the raw model output, final result after post-processing and ground truth, from top to bottom, respectively.

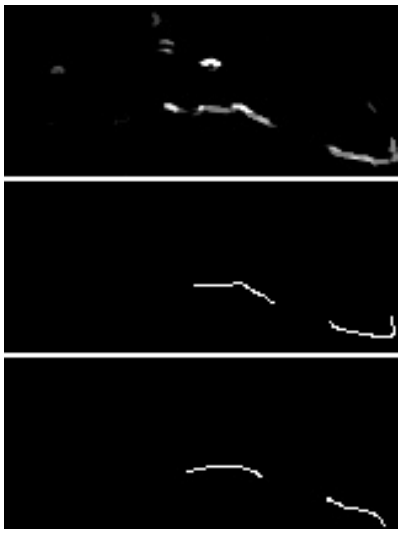


Fig. 2. Small part of result on pancreas #49

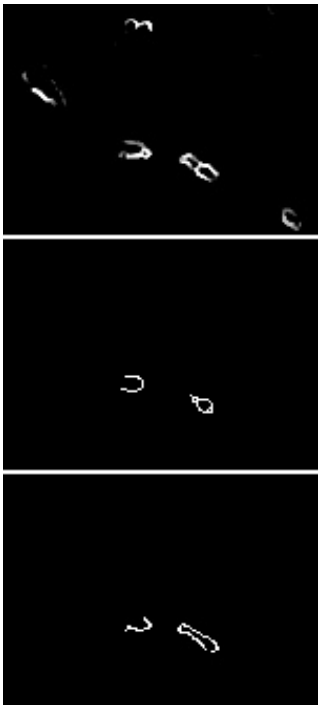


Fig. 3. Small part of result on prostate #46

Acknowledgment

This research is supported Year 2021 Copyright Technology R&D Program by Ministry of Culture, Sports and Tourism and Korea Copyright Commission (Project Name: Development of high-speed music search technology using deep learning, Project Number: 2021-hs-9500, Contribution Rate: 50%) and by the Technology Innovation Program (20016180, Forecast of overseas inflow of new infectious diseases and development of intelligent blocking technology) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea, Contribution Rate: 50%).

References

1. M Dunn, MB Morgan, TW Beer.: Perinerual invasion: identification, significance, and a standardization, Dermatologic surgery, 2009
2. MH Hesamian, W Jia, X He, P Kennedy.: Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges, Journal of Digital Imaging, 2019
3. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, IlyaSutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(56):1929–1958, 2014
4. Tongxue Zhou, Su Ruan, and Stephane Canu. A review: Deep learning for medical image segmentation using multi-modality fusion. Array, 3-4:100004, 2019
5. Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, NimaTajbakhsh, and Jianming Liang. UNet++: A nested U-Net architecture for medical image segmentation. CoRR,abs/1807.10165, 2018
6. Pavel Yakubovskiy. Segmentation models pytorch. https://github.com/qubvel/segmentation_models. pytorch, 2020.3
7. Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. CoRR,abs/1905.11946, 2019.

Deep learning for predicting the layer height in laser powder-feed metal additive manufacturing using big data

Muhammad Mu'az Imran[1,2], Young Kim[1], Gisun Jung[1], Azam Che Idris[2], Liyanage C. De Silva[2,3] and Yun Bae Kim[1],

[1] Department of Industrial Engineering, Sungkyunkwan University, Seo-bu Street 2066, Suwon, 16419, South Korea
[2] Faculty of Integrated Technologies, Universiti Brunei Darussalam, Bandar Seri Begawan, Brunei Darussalam
[3] School of Digital Science, Universiti Brunei Darussalam, Bandar Seri Begawan, Brunei Darussalam
{muazimran, lmjlguard, kimyb}@skku.edu
gsjung09@naver.com
{azam.idris, liyanage.silva}@ubd.edu.bn

Abstract. Despite the successful usage of directed energy deposition (DED) in the metal 3D printing industry, the reliability and repeatability of the process are still some of the major issues faced by the industry and academia. One of them is predicting the layer height, which is required for not only ensuring part geometrical accuracy and consistency but also its microstructural and mechanical properties. We have developed a novel deep learning process in DED additive manufacturing (AM) that has the potential to accurately predict the layer height of a complex geometry structure by considering laser power, scan speed, print angle and coordinate.

Keywords: Directed Energy Deposition, Big Data, LSTM, Prediction Model

1 Introduction

Directed energy deposition (DED) is one of the laser powder-feed metal additive manufacturing (LPFM-AM) methods that uses a highly concentrated energy beam (laser or electron) to simultaneously melt the substrate and the feeding material (powder or wire) at the same time.

Despite the successful usage of this method in the metal 3D printing industry, the reliability and repeatability of the process are still some of the major issues faced by the industry and academia. A high-quality product that is being generated by the process is difficult to obtain due to the multitude of process parameters involved such as laser power, scanning speed, powder feed rate, beam diameter, hatch spacing, dwell time, defocusing distance, stand-off distance, etc. The quality of the finished product can be viewed in terms of its geometrical accuracy—its deviation from the nominal dimension—also the microstructural and mechanical properties, and internal and external defects. To achieve a high-quality component, one would ideally need to

alter and optimize the process parameters corresponding to the physics of the process that keeps on changing dynamically and instantaneously.

To the authors' best knowledge, there have been no studies conducted utilizing a data-driven method to predict the layer height for complex geometries (i.e., overhang structure with varying angles) in DED. As most of the previous studies used trial-and-error methods experimented on either single-track, multi-track or thin-wall structures, alongside statistical methods for layer height prediction. Meanwhile, the usage of the deep learning model in DED is still scarce. Therefore, the motivation for the present work is to provide a novel framework in which the layer height can be predicted using a data-driven predictive model with high accuracy during DED under varying process parameters.

2 Related study

The traditional method relies on trial-and-error and design of experiment method, resulting in limited consideration of processing windows. A statistical method has been used to create a model based on the correlation between process parameters and the geometrical characteristics. Onwubolu [1] utilized multiple-regression method to find the correlation between the two variables by using an exponential relation. However, this method relies on the trial-and-error method, which is very time-consuming.

Meanwhile, the physics-based models namely analytical and numerical models are not reliable to predict the melt pool temperature and geometrical characteristics of the clad (solidified melt-pool). Although an analytical model is computationally faster than the numerical model, both methods are driven by a heap of assumptions leading to discrepancies with experimental results.

Machine learning (ML) methods on the other hand have been burgeoning in the literature due to their capability to predict or classify effectively for real-time process control and optimization while only limited knowledge of the underlying physics is required. Feenstra et al [2] used an artificial neural network (ANN) to elucidate the varying effects of process parameters on the melt pool geometrical characteristics, based on multiple single-track fabricated parts with varying process parameters.

Mozaffar et al [3] proposed a data-driven model using a Recurrent Neural Network (RNN) to predict the thermal history of any given point in a DED based on virtually generated data from an in-house finite element code, GAMMA. To test the robustness and generality of the proposed model, the author tested their model over two types of experiments—shorter and longer time span data, and non-trained dissimilar geometries—but yield poor outcomes.

Similarly, but with real-time experiments, Zhang et al [4] used Long short-term memory (LSTM) and XGBoost algorithms as part of their predictive models for estimating the melt pool temperature of a thin-wall fabricated part based on laser power and scanning speed. It was found out that LSTM exhibits a higher accuracy compared to XGBoost, but at an expense of computational speed. However, this method was not able to test its robustness with dissimilar geometries as [3].

3 Proposed framework

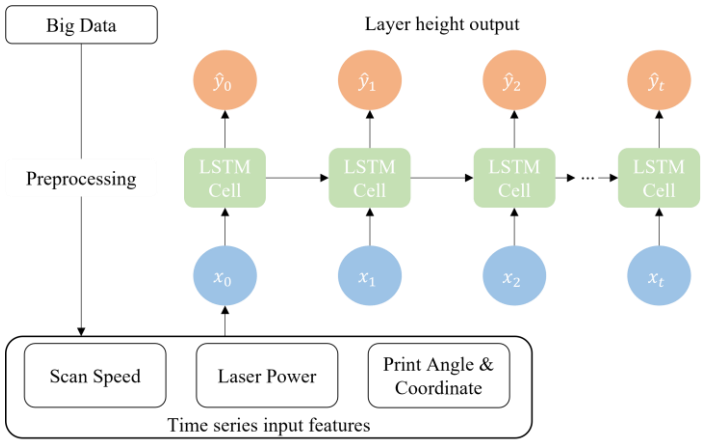


Fig.1. Proposed framework for predicting layer height in DED

In this section, we propose a framework tailored for predicting the layer height in DED as shown in Fig.1. Based on our data, which has time-series characteristics, LSTM is selected in this study due to its capability to learn by capturing the dependence between timesteps to effectively preserve the long-term effect of the sequential data developed by [5]. The basic structure of a typical LSTM consists of four main features, (1) Forget gate (decides what information should be kept or thrown away), (2) Store gate (decides what new information should be added to the cell state), (3) Cell state (carries the relevant information), and (4) Output gate (decides what the next hidden state should be).

The inputs of the LSTM architecture are as follows: (1) scan speed, (2) laser power, and (3) print angle and coordinate.

4 Future work

Moreover, the melt pool temperature is directly associated with the surface tension of the melt pool [6, 7]. As part of our future work, surface tension can be extracted from a mathematical model proposed by Su and Mills [8] and Li et al [9], which will be used as one of the input variables in accurately predicting melt pool temperature. Additionally, Wang et al [10] shed some light on the importance of surface tension towards the stability of melt pool, especially to support overhang structure. The authors added that the melt pool will collapse when the surface tension is less than the threshold value determined by the injected powder and its gas carrier component forces acting on the melt pool, $F_s < G + F_1$, as shown in Fig.2. Thus, by adding this feature, it is expected to predict the occurrence of surface defects on the fabricated

part. In some cases where high-precision engineering is required such as in aerospace or biomedical application, surface defects like irregular thickness and roughness are unacceptable since they can be detrimental to the mechanical properties of the fabricated part.

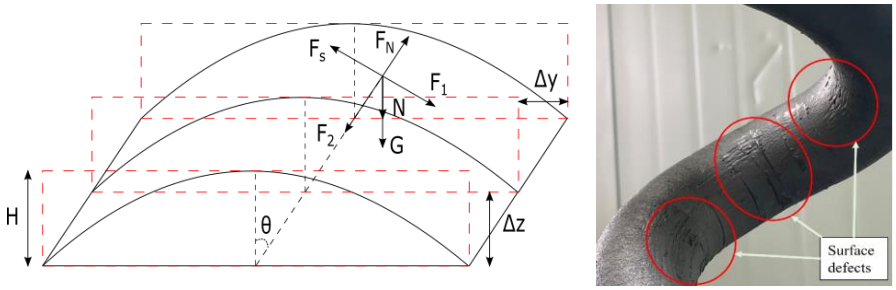


Fig.2. (a) A schematic of overhang structure, (b) Defects observed on the surface of the fabricated part with complex geometry.

References

1. Onwubolu, G.C., Davim, J.P., Oliveira, C., Cardoso, A.: Prediction of clad angle in laser cladding by powder using response surface methodology and scatter search. Opt. Laser Technol. 39, 1130–1134 (2007). https://doi.org/10.1016/j.optlastec.2006.09.008.
2. Feenstra, D.R., Molotnikov, A., Birbilis, N.: Utilisation of artificial neural networks to rationalise processing windows in directed energy deposition applications. Mater. Des. 198, 109342 (2021). https://doi.org/10.1016/j.matdes.2020.109342.
3. Mozaffar, M., Paul, A., Al-Bahrani, R., Wolff, S., Choudhary, A., Agrawal, A., Ehmann, K., Cao, J.: Data-driven prediction of the high-dimensional thermal history in directed energy deposition processes via recurrent neural networks. Manuf. Lett. 18, 35–39 (2018). https://doi.org/10.1016/j.mfglet.2018.10.002.
4. Zhang, Z., Liu, Z., Wu, D.: Prediction of melt pool temperature in directed energy deposition using machine learning. Addit. Manuf. 37, 101692 (2021). https://doi.org/10.1016/j.addma.2020.101692.
5. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. Neural Comput. 9, 1735–1780 (1997). https://doi.org/10.1162/NECO.1997.9.8.1735.
6. Wei, H.L., Mukherjee, T., Zhang, W., Zuback, J.S., Knapp, G.L., De, A., DebRoy, T.: Mechanistic models for additive manufacturing of metallic components. Prog. Mater. Sci. 116, (2021). https://doi.org/10.1016/j.pmatsci.2020.100703.
7. Kou, S.: Welding Metallurgy. Wiley (2003).
8. Su, Y., Mills, K.C., Dinsdale, A.: A model to calculate surface tension of commercial alloys. J. Mater. Sci. 40, 2185–2190 (2005). https://doi.org/10.1007/s10853-005-1930-y.
9. Li, Z., Mukai, K., Zeze, M., Mills, K.C.: Determination of the surface tension of liquid stainless steel. J. Mater. Sci. 40, 2191–2195 (2005).
10. Wang, X., Deng, D., Hu, Y., Ning, F., Wang, H., Cong, W., Zhang, H.: Overhang structure and accuracy in laser engineered net shaping of Fe-Cr steel. Opt. Laser Technol. 106, 357–365 (2018). https://doi.org/10.1016/j.optlastec.2018.04.015.

Trajectory Privacy Preservation by Using Deep Learning: Transformer-TrajGAN

Ellen S. Park[1], Hazel H. Kim[1], Suan Lee[2], Wookey Lee[3]

[1] Department of Industrial Security Governance, Inha University, South Korea
[2] Department of Computer Science, Semyung University, South Korea
[3] Department of Industrial Engineering, Inha University, South Korea
{sevenqkrtn,coco1833}@naver.com, suanlab@gmail.com, trinity@inha.ac.kr

Abstract. With the rapid development of mobile information and communication technology, the demand and supply of location-based services are increasing in many industries. Various trajectory data combining location data and spatio-temporal information are being used, leading to serious privacy violations for individuals and countries. To solve this problem, we propose a research methodology to generate synthetic trajectory data that protects personal privacy by transducing a deep learning model, transformer, with a generative adversarial network (GAN) approach. The proposed model preserves the spatial and temporal characteristics of trajectory data while protecting location data-based trajectory privacy.

Keywords: Deep learning, Generative Adversarial Network, Transformer, Trajectory, Privacy Protection

1 Introduction

Nowadays, location data is widely collected from smart devices with a built-in GPS (Global Positioning System). Trajectory data provides convenience by converging with various industries. However, if an individual can be identified through their trajectory data, serious privacy problems can arise. Therefore, it is necessary to use the trajectory information safely through privacy protection technology. According to [1, 2], trajectory data privacy protection methodology using k-anonymity and differential privacy was studied. However, since this approach focuses only on points, temporal characteristics are rarely considered, and data utility cannot be guaranteed. Our study proposes a solution by using Transformer-TrajGAN for trajectory data to preserve personal information to solve these problems..

2 Related Work

Previous studies conducted differential privacy studies using the k-anonymity model to protect individual trajectory data [1, 2, 3]. A model has also been studied to satisfy k-

anonymity using LFP-tree by converting the trajectory data into a vector [4, 5]. All of these protect trajectory privacy but confuse trajectory locations and give uncertainty in preserving privacy. To solve this problem, a methodology using the deep learning model LSTM and GAN was introduced [6]. This study proposes a model that protects trajectory privacy using a combination of the Transformer model, which has gained state-of-the-art results in trajectory data sequence processing and the GAN model.

3 Transformer-TrajGAN

Before generating the synthesis trajectory, we preprocess the trajectory data. The data preprocessing model performs the task of changing the original trajectory into a vector according to certain rules to fit the input of Transformer-TrajGAN.

3.1 Generator

- Input layer: The trajectory data that has been pre-processed and random noise are used as input values.
- Embedding layer: The location, day, time, location type are embedded as vectors using Multilayer Perceptron (MLP). In (1, 2, 3, 4), $e_n^{spatial}, e_n^{day}, e_n^{hour}, e_n^{type}$ are embedded vectors for each attribute. Δlat_n and Δlon_n are the latitude and longitude deviations of the nth trajectory point, and $v_n^{day}, v_n^{hour}, v_n^{type}$ are each one-hot vector for the day and time of the week of the nth trajectory point. $\emptyset^s, \emptyset^d, \emptyset^h, \emptyset^t$ are the MLPs using *ReLU*, and $W_{es}, W_{ed}, W_{eh}, W_{et}$ are the respective embedding weights for the MLP.

$$e_n^{spatial} = \emptyset^s(\Delta lat_n, \Delta lon_n; W_{es}) \quad (1)$$

$$e_n^{day} = \emptyset^d\left(v_n^{day}; W_{ed}\right) \quad (2)$$

$$e_n^{hour} = \emptyset^h(v_n^{hour}; W_{eh}) \quad (3)$$

$$e_n^{type} = \emptyset^t\left(v_n^{type}; W_{et}\right) \quad (4)$$

- Transformer layer: A transformer model is used to receive time series data with a specific time order as input and generate sequences in the same time step as the output.
- Decoding layer: Decodes the synthesized trajectory data from the output H of the Transformer layer. In (5, 6, 7, 8), $\Delta lat_n^{'}, \Delta lon_n^{'}$ is the latitude and longitude deviations of the nth synthetic trajectory point, and $v_n^{'day}, v_n^{'hour}, v_n^{'type}$ is one-hot vectors of day, time, type. D^s, D^d, D^h, D^t are dense layers that use *tanh* or *softmax* to decode location, day, time and type attributes. $W_{ds}, W_{dd}, W_{dh}, W_{dt}$ refer to decoding weights of the dense layer.

$$\left(\Delta lat_n^{'}, \Delta lon_n^{'}\right) = D^s(h_n; W_{ds}) \quad (5)$$

$$v_n^{'day} = D^d(h_n; W_{dd}) \quad (6)$$

$$v_n^{'hour} = D^h(h_n; W_{dh}) \quad (7)$$

$$v_n^{'type} = D^t(h_n; W_{dt}) \quad (8)$$

3.2 Discriminator

The discriminator serves to distinguish whether the preprocessed original trajectory data is the composite trajectory generated from the generator. Discriminator uses only preprocessed trajectory data excluding noise as input values and uses a many-to-one transformer model. In addition, binary classification is performed using *sigmoid* as an activation function in the dense layer, and in conclusion, one scalar value is output to classify whether it is an original or a synthetic trajectory.

4 Experiment

We used The Foursquare's NYC Check-in Dataset [6] and Taxi San Francisco Dataset. NYC Check-in data is check-in location information collected in New York for about 10 months, with a total of 193 users and 66,962 location data recorded. Taxi San Francisco data consists of GPS data collected over 30 days from about 5 million taxis in San Francisco, USA.

First, synthetic trajectory data was generated by our model using the NYC Check-in dataset. After that, the 'TUL (Trajectory-User Link) Algorithm' was performed to evaluate the effectiveness of privacy protection. The higher the algorithm's accuracy, the better the individual is identified, which means that the effectiveness of personal information protection is reduced. To measure the accuracy of the 'TUL Algorithm', ACC@1, ACC@5, Macro-P, and Macro-R were used [7]. When metrics are measured with original trajectory data, they all show high accuracy of about 0.9 [6]. However, the proposed model is 0.259 in ACC@1, ACC@5 is 0.495, Macro-P is 0.256, and Macro-R is 0.258. By reducing the accuracy of the TUL algorithm by about 0.6 to 0.7, the trajectory privacy protection effectiveness of the Transformer-TrajGAN model can be confirmed.

Second, the difference between the position points of the original trajectory and the synthetic trajectory was compared using the Taxi San Francisco data to confirm that the trajectory was well generated. We use Final Average Displacement (FAD, Equivalent Final Displacement Error [8]) and Mean Average Displacement (MAD, Equivalent Average Displacement Error [8]). In TABLE 1, our model shows FAD 0.294 and MAD 0.253. It can be seen that the proposed model performs better than when the trajectory was created using only a Transformer. In other words, both metrics show smaller values through the trajectory generation mechanism of GAN, which is difficult to distinguish whether it is real or fake.

Table 1. FAD and MAD for the Taxi San Francisco dataset.

Method	Data	FAD	MAD
Transformer	Taxi San Francisco	2.753	2.985
Transformer-TrajGAN	Taxi San Francisco	0.294	0.253

5 Conclusion

The use of location information is essential in using location-based services, but it can be abused and cause significant problems in terms of privacy protection. Our proposed model preserved privacy for personal location information by creating synthetic trajectory data using Transformer and GAN. Experimental results confirmed that the proposed model preserves privacy and has the performance of generating trajectories well. In addition, there is an advantage of having better data utility than the existing trajectory privacy protection model by reflecting the day and time. As future work, we plan to analyze the type of place in the synthesis trajectory of the proposed model. Through this, it can be safely used in data analysis tasks such as individual place preferences.

Acknowledgments. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2019S1A5C2A03081234).

References

1. Sweeney, L.: k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05), pp 557--570 (2002)
2. Dwork, C.: Differential privacy: A survey of results. In International conference on theory and applications of models of computation. pp. 1--19. Springer, Berlin, Heidelberg (2008)
3. Niu, B., Li, Q., Zhu, X., Cao, G., & Li, H.: Achieving k-anonymity in privacy-aware location-based services. In IEEE INFOCOM 2014-IEEE Conference on Computer Communications, pp. 754--762. (2014)
4. Eom, C. S. H., Lee, C. C., Lee, W., & Leung, C. K.: Effective privacy preserving data publishing by vectorization. Information Sciences, 527, pp 311--328. (2020)
5. Tojiboev, R., Lee, W., & Lee, C. C.: Adding noise trajectory for providing privacy in data publishing by vectorization. In 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 432--434. IEEE. (2020)
6. Rao, J., Gao, S., Kang, Y., & Huang, Q.: Lstm-trajgan: A deep learning approach to trajectory privacy protection. arXiv preprint arXiv:2006.10521. (2020)
7. May Petry, L., Leite Da Silva, C., Esuli, A., Renso, C., & Bogorny, V. MARC: a robust method for multiple-aspect trajectory classification via space, time, and semantic embeddings. International Journal of Geographical Information Science, 34(7), 1428-1450 (2020).
8. Su, H., Zhu, J., Dong, Y., & Zhang, B.: Forecast the Plausible Paths in Crowd Scenes. In IJCAI (Vol. 1, p. 2). (2017)

Improving Atrous Spatial Pyramid Pooling in DeepLabv3+ using the Attention mechanism for the diversity of polyps in semantic segmentation

Sy-Phuc Pham[1], Hyung-Jeong Yang[1,*], Duy-Phuong Dao[1], Guee-Sang Lee[1], Soo-Hyung Kim[1], Seok-Bong Yoo[1]
[1] Dept of Artificial Intelligence Convergence, Chonnam National University
61186, Gwangju, South Korea
*Corresponding Author: hjyang@jnu.ac.kr
{phamsyphuc123, duyphuongcri}@gmail.com, {gslee, shkim, sbyoo}@jnu.ac.kr

Abstract. Automatic polyps segmentation is a challenging task because of variations in the shape and size of polyps. In this paper, we proposed the model for the polyps segmentation based on DeepLabv3+ architecture, and the Attention mechanism can be solved the various shape and sizes of polyps. Our approach achieves competitive performance on MediaEval-Medico-2021 dataset, CVC-ClinicDB dataset, VIP-CUP18 dataset. The goal of this paper is to segment polyps in images taken from endoscopies. The results show that the performance of the proposed method improves than the recent method in segmentation.

Keywords: Deep learning, polyps segmentation, medical imaging.

1 Introduction

Deep neural networks are now the most powerful machine learning models in a wide variety of domains, from image analysis to natural language processing, and are extensively utilized in academia and industry. These developments have a huge potential in medical analysis and are popular in the medical area [1]. In recent years, some deep network architectures have been used in state-of-the-art (SOTA) for each task in medical image analysis. This paper focuses on image semantic segmentation with deep neural networks.

Colon polyps are growths on the lining of the colon and rectum. Colonoscopy is the standard procedure for the localization and removal of colorectal polyps. The use of polyps imaging is an important factor in diagnosis. From 2008 to 2016, the images in the dataset were gathered prospectively at a Norwegian hospital during standard clinical exams [2]. The MediaEval-Medico-2021 selected 1360 images for the challenge.

The technique of separating a picture into various sections is known as semantic image segmentation. Each region helps to identify different objects present in an image. Image segmentation is used to understand and interpret the image more efficiently. In this paper, we segmented the polyps given the two-dimensional (2D)

data. We proposed a model for automatically segmenting the polyps by utilizing DeepLabV3+ [3] with Attention [4] mechanism (DLV3SA) on 2D data. On semantic segmentation polyps, the proposed model has demonstrated great potential.

The content of this paper is organized as follows. Section 2 presents the summary of a few existing research closely related to semantic segmentation. In section 3, we discuss the proposed method. After that, details of the experiment and results are presented in section 4. Finally, the conclusions and future work is presented in section 5.

2 Related Works

V-Net [5] is the encoder-decoder architecture used to segment Magnetic Resonance Imaging (MRI) volumes. The aim of the authors in [5] is to extract features from the input and resolution the features with the appropriate stride. Like the architecture of U-Net [6], on the left of the V-Net is the compression path, on the right of the V-Net is the decompression path. The loss function has been used in this model based on the dice overlap coefficient (DOC). When we used DOC loss, we did not find the proper mix of foreground and background voxels. But for images that have more objects, the DOC function may not be good for the model. The model in [7] has been developed based on the U-Net [6]. For improving the convergence, the author eliminates bottlenecks [8] in the architecture and applies batch normalization [9]. This model has been used with a weight softmax loss function and achieved an average intersection over union of 0.863 in 3 folds. The limitation of this architecture, as shown in the result section of the paper, is that the segmentation of 2D images is not better.

U-Net can be used for automatic brain tumor detection [10]. U-Net is also used in 3D medical image segmentation [7]. U-Net architecture is based on the encoder-decoder structure that is used in semantic segmentation [6]. It consists of two paths in the architecture; the contraction is on the left, and the expansion is on the right. The mission of the contraction is to extract features; the benefit of the contraction is similar to an encoder. In the contraction, the input size is reduced from 572×572 to 32×32, and the depth is increased from 3 to 512. The expansion path includes each layer opposite to each layer in the contraction path. The upsampling process has been applied to increase the size of the layer. Finally, U-Net gets the prediction mask. The special in U-Net is the application of skip connection between two paths for each layer. The contraction path in the U-Net is not optimal for getting information about the polyps because of their various sizes.

ResUNet [11] is the encoder-decoder architecture with the U-Net backbone. Resiudal blocks to solve the problem of vanishing gradients in a deep network. The author introduced two architectures, ResUNet-a d6 and ResUNet-a d7. The encoder part in the ResUNet-a d6 includes 6 ResBlock, in the ResUNet-a d7 includes 7 ResBlock.

The encoder-decoder architecture has been used in Transformer [4] model. The attention mechanism is the core of the Transformer model. In semantic segmentation, an attention mechanism is used to learn to softly weight the multi-scale features with the location of each pixel [12].

3 Proposed Method

In this section, we first introduce the architecture of the proposed method. The proposed deep network includes two main, DeepLabv3+ and attention mechanism.

DeepLabv1 [13] and DeepLabv2 [14] have been developed by Liang Chen and George Papandreou. DeepLabv2 has changed little compared to DeepLabv1. The network achieved a good result on the Pascal VOC 2012. The architecture of DeepLab applied Atrous convolution instead of the Transposed convolution. The author applied Conditional Random Field (CRF) to calif the better prediction. The main part of DeepLab is as follows: Input→Deep Convolutional Neural Network (DCNN)→Score map→Bilinear interpolation→Fully connected CRF→Mask output. After using DCNN, the feature map was collected. The feature map has been scaled up by bilinear interpolation to return the input size. Boundaries are quoted well, and object shapes are further smoothed through the fully connected CRF. The key points in the DeepLab architecture are the Atrous convolution and fully connected CRF. It helps the model achieve good results in the Pascal dataset. The disadvantage of this version is the computation of the parameters in fully connected CRF.

To solve the polyp's various sizes in the dataset, we improve the Atrous Spatial Pyramid Pooling (ASPP) in DeepLabv3+ by adding batch normalization at the last feature before applying the Atrous convolution. The output goes through the convolution layer size 1×1 with a filters size is 256. When we set the high rate, the number of suitable filters decreases. A suitable filter can be applied to all feature areas, but it does not add padding with value zero. When we use the ASPP, the model has many filter sizes, and the number of suitable filters is bigger. To emphasize the feature area in the image, we proposed the ASPP combined with the attention mechanism. The size of the polyp is various size, using the attention to re-calibrate the feature map at each layer of the ASPP in order to prioritize more informative channels, as shown in Equation 1. We used the output of each layer in ASPP for the input of attention.

The proposed model was based on the Encoder-Decoder architecture, as shown in Figure 1. The network begins with the pre-trained ResNet [15] as the encoder, which is followed by the ASPP. The ASPP consists of Atrous convolution, which helps to encode multi-scale contextual information. The Atrous convolution layers for feature extraction instead of CNNs. The Atrous convolution layers have more special effects than CNNs, do not reduce the deep dimension of the feature map, but still, keep the number of parameters and the equivalent computational cost as CNNs [14]. So, the benefit of the Atrous convolution is increased output size compared to input. After each layer in ASPP, we applied the Attention to increase the performance of the ASPP. Attention mechanism was the solution of the small polyps segmentation and helped to improve the output from each layer in ASPP. After the attention block, the output from each layer in ASPP was the sum of the weighted values. This weight was the softmax function calculated based on the query and the corresponding key. The formula for calculating attention weight was called Scaled Dot-Product Attention, as shown in Equation (1). Before we concatenated the output from each layer in ASPP, this output went through the Squeeze and Excitation Network [16]. This network helps improve the feature representation by suppressing irrelevant features and enhancing the important features. Next, it is followed by a bilinear upsampling by a

factor of 4 and then concatenated with the low-level information from the encoder. A few 3×3 convolutions are applied, and again it is followed by a bilinear upsampling by a factor of 4. Finally, we got the prediction mask.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (1)$$

Where, Q is the queries, K is the keys of dimension d_k, V is the values of dimension d_v.

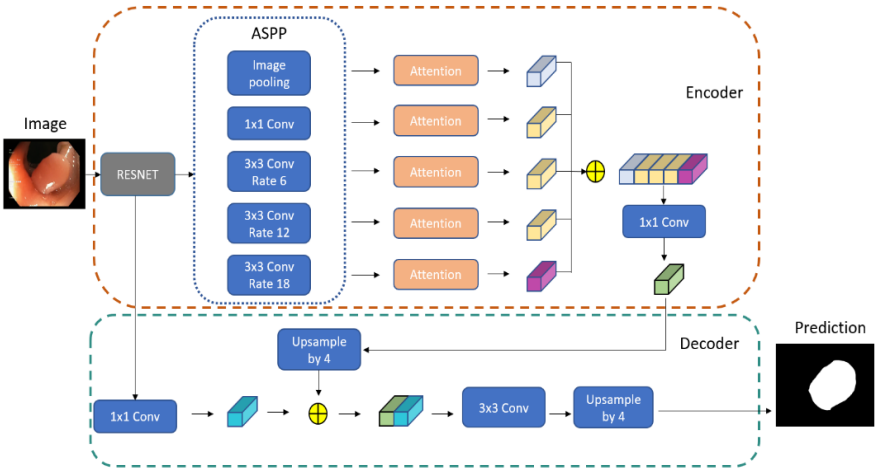


Fig. 1. Proposed DeepLabv3+ architecture with Attention (DLV3SA).

4 Experimental Result

We conducted the experiment with the MediaEval-Medico-2021 dataset. The dataset included 1360 images with various size ranging from 332×487 to 1920×1072. The train set has 1088 images (80%), and the valid set has 272 images (20%). About the polyps segmentation, polyps regions typically take up only a small portion of the entire image, and attention mechanism is one of the efforts to address small ROI segmentation. To improve the performance, we combine attention with the DeepLabv3+, suited for various regional segmentations. The loss function is an important factor in the segmentation task. We used Binary Cross-Entropy (BCE), as shown in Equation (2) [17].

$$L_{BCE}(y, \hat{y}) = -(ylog(\hat{y}) + (1 - y)\log(1 - \hat{y})) \quad (2)$$

Where, $\hat{y}$ is the predicted value by the prediction model, y is the ground truth.

To assess the effectiveness of the proposed model, we use Jaccard index [18], Precision [19], Recall [19], which are described in Equations (3, 4, 5) as evaluation metrics. At the semantic segmentation, we focus on the Jaccard index. The Jaccard index for the proposed model was much higher, as shown in Table 1 and Table 2. Figure 2 shows the prediction mask of 6 models, which are related in Table 1 and Table 2. To generalize the proposed model, we test its performance on three different medical datasets mentioned in Table 3. The results from the Table 3 give us the performance of the DLV3SA at the CVC_ClinicDB is better than U-Net. But in the remaining datasets, the U-Net model had superior results. The DLV3SA model was good for the datasets that contained these kinds of polyps. Because of the characteristics of endoscopic images, which are different from MRI and CT images, there are bright areas, and the method of execution of each type of image is different. It's also partly because we try to build a solution for the problem of semantic image segmentation for polyps as described in the proposed method. The attention mechanism is also better to help extract the features and make the weights change positively to predict the mask better than the DeepLabv3+ model.

The proposed model was implemented in Keras framework with Tensorflow backend. In this method, we try to train the proposed model with two sizes of an input image are $256 \times 256 \times 3$ and $512 \times 512 \times 3$. The result from the size $256 \times 256 \times 3$ was better. So, we chose the input size for DLV3SA model to be $256 \times 256 \times 3$ (image) and $256 \times 256 \times 1$ (mask), respectively. We used Adam optimizer with a learning rate of 0.0001 and a batch size of 4. The experiments were run on Amazon Web Services (AWS) server (NVIDIA T4 GPU).

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{3}$$

$$Precision = \frac{True\ positive}{True\ positive + False\ Posivtive} \tag{4}$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \tag{5}$$

Table 1. Performance on the MediaEval-Medico-2021 dataset with input size 256×256.

Model	Jaccard	Precision	Recall
U-Net [6]	0.63241	0.77203	0.77287
ResUnet [11]	0.61554	0.78453	0.73905
Model in [20]	0.7192	0.832	0.8236
U-Net with pre-trained MobileNetv2 [21]	0.74296	0.82357	**0.89366**
DeepLabv3+ with pre-trained ResNet [3]	0.75956	0.86585	0.86849
DLV3SA (proposed)	**0.76307**	**0.87481**	0.86433

Table 2. Performance on the MediaEval-Medico-2021 dataset with input size 512×512.

Model	Jaccard	Precision	Recall
U-Net [6]	0.547	0.76397	0.68804
ResUnet [11]	0.3722	0.54925	0.59317
Model in [20]	0.7333	0.8537	0.8341
U-Net with pre-trained MobileNetv2 [21]	0.71072	0.84906	0.8155
DeepLabv3+ with pre-trained ResNet [3]	0.73431	**0.86317**	0.8332
DLV3SA (proposed)	**0.76043**	0.85669	**0.88023**

Table 3. Spec of the datasets.

Dataset	Object	Number of images	Method
CVC-ClinicDB [22]	Polyps	612	Endoscopy
VIP-CUP18 [23]	Lung tumor	5204	Computed tomography (CT)
TCGA [24]	Brain tumor	1373	MRI

Table 4. Comparison between U-Net and DLV3SA on three datasets.

Network	Dataset	Jaccard	Precision	Recall
U-Net [6]	TCGA	**0.74247**	**0.85029**	0.82252
	CVC-ClinicDB	0.71827	0.8363	0.79322
	VIP-CUP18	**0.72147**	**0.83801**	**0.8113**
DLV3SA (proposed)	TCGA	0.7217	0.82582	**0.82672**
	CVC-ClinicDB	**0.80256**	**0.8975**	**0.87434**
	VIP-CUP18	0.67503	0.80585	0.76922

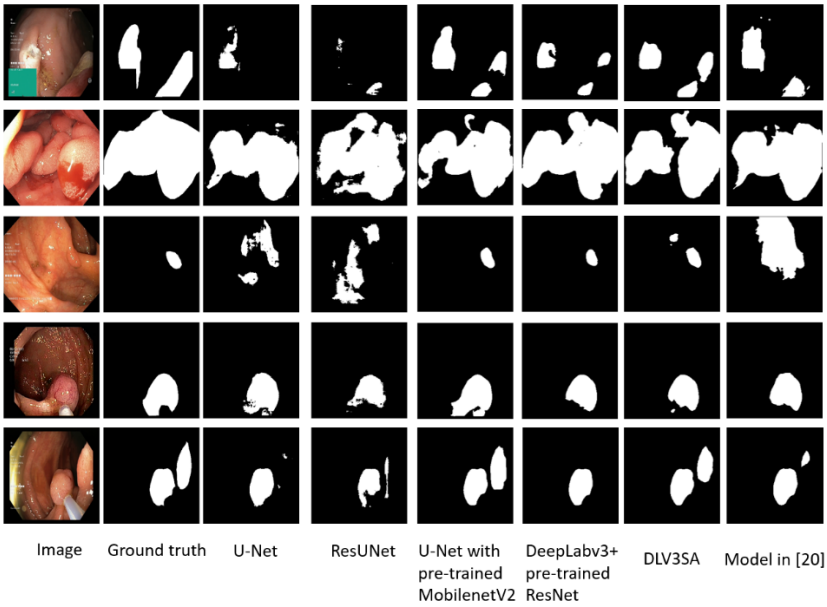


Fig. 2. Example results in the Medico dataset.

5 Conclusion

In this paper, the proposed model for polyps segmentation has been performed better than the other model in Table 1. The DLV3SA model can focus on the small region of the image by the contribution of Attention mechanism, and the advantage of ASPP better than the CNNs. The attention mechanism added in the ASPP improved the weight of the ASPP's output. Through the experiment, we found that it is possible to use DeepLabv3 architecture to segment medical images and in some cases achieve better results than specialize models for medical image segmentation such as U-Net. The future work includes improving the performance of this model by the enhance the Encoder path.

Acknowledgments. This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) & funded by the Korean government (MSIT). (NRF-2019M3E5D1A02067961).

References

1. A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, May 2019, doi: 10.1016/J.ZEMEDI.2018.11.002.
2. H. Borgli *et al.*, "HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Scientific Data 2020 7:1*, vol. 7, no. 1, pp. 1–14, Aug. 2020, doi: 10.1038/s41597-020-00622-y.
3. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11211 LNCS, pp. 833–851, Feb. 2018, Accessed: Oct. 30, 2021. [Online]. Available: https://arxiv.org/abs/1802.02611v3
4. A. Vaswani *et al.*, "Attention Is All You Need," *Advances in Neural Information Processing Systems*, vol. 2017-December, pp. 5999–6009, Jun. 2017, Accessed: Oct. 30, 2021. [Online]. Available: https://arxiv.org/abs/1706.03762v5
5. F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016*, pp. 565–571, Dec. 2016, doi: 10.1109/3DV.2016.79.
6. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241, May 2015, Accessed: Oct. 30, 2021. [Online]. Available: https://arxiv.org/abs/1505.04597v1
7. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9901 LNCS, pp. 424–432, Oct. 2016, doi: 10.1007/978-3-319-46723-8_49.
8. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision." pp. 2818–2826, 2016.

9. D. Tran, L. Bourdev, R. Fergus, … L. T.-P. of the, and undefined 2016, "Deep end2end voxel2voxel prediction," *cv-foundation.org*, Accessed: Nov. 04, 2021. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2016_workshops/w12/html/Tran_Deep_End2End_Voxel2Voxel_CVPR_2016_paper.html
10. H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-net based fully convolutional networks," *Communications in Computer and Information Science*, vol. 723, pp. 506–517, 2017, doi: 10.1007/978-3-319-60964-5_44.
11. F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, Apr. 2019, doi: 10.1016/j.isprsjprs.2020.01.013.
12. L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille, "Attention to Scale: Scale-Aware Semantic Image Segmentation." pp. 3640–3649, 2016.
13. L.-C. Chen, G. Papandreou, K. Murphy, and A. L. Yuille, "SEMANTIC IMAGE SEGMENTATION WITH DEEP CON-VOLUTIONAL NETS AND FULLY CONNECTED CRFS".
14. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, Jun. 2016, Accessed: Oct. 30, 2021. [Online]. Available: https://arxiv.org/abs/1606.00915v2
15. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, Dec. 2015, Accessed: Oct. 30, 2021. [Online]. Available: https://arxiv.org/abs/1512.03385v1
16. J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, Sep. 2017, Accessed: Oct. 30, 2021. [Online]. Available: https://arxiv.org/abs/1709.01507v4
17. S. Jadon, "A survey of loss functions for semantic segmentation," *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2020*, Jun. 2020, doi: 10.1109/cibcb48159.2020.9277638.
18. J. Bertels *et al.*, "Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory & Practice," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11765 LNCS, pp. 92–100, Nov. 2019, doi: 10.1007/978-3-030-32245-8_11.
19. D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," Oct. 2020, Accessed: Oct. 30, 2021. [Online]. Available: https://arxiv.org/abs/2010.16061v1
20. T. Khanh, D. Dao, N. Ho, H. Yang, E. B.-A. Sciences, and undefined 2020, "Enhancing u-net with spatial-channel attention gate for abnormal tissue segmentation in medical imaging," *mdpi.com*, doi: 10.3390/app10175729.
21. M. V. L. Branch and A. S. Carvalho, "Polyp Segmentation in Colonoscopy Images using U-Net-MobileNetV2," Mar. 2021, Accessed: Nov. 04, 2021. [Online]. Available: https://arxiv.org/abs/2103.15715v1
22. J. Bernal, F. Sánchez, … G. F.-E.-… M. I. and, and undefined 2015, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Elsevier*, Accessed: Nov. 04, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0895611115000567

23. 2018 IEEE Signal Processing Society Video and Image Processing (VIP) Cup. Available online: https://users.encs.concordia.ca/~i-sip/2018VIP-Cup/index.html (accessed on 19 May 2020)
24. M. Buda, A. Saha, and M. A. Mazurowski, "Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm," *Computers in Biology and Medicine*, vol. 109, pp. 218–225, Jun. 2019, doi: 10.1016/J.COMPBIOMED.2019.05.002.

Prediction of Solar Power Generation Based on CNN & RNN

Dong-kyu Yun[1], Ju-yeon Lee[1], Woo-seok Choi[1],
Aziz Nasridinov[3], Sang-hyun Choi[2*]

[1]Dept. Bigdata, Chungbuk National University, Cheongju, South Korea
[2]Dept. Management Information System, Chungbuk National University, Cheongju, South Korea
[3]Dept. Computer Science, Chungbuk National University, Cheongju, South Korea
{dongkyu.yun,yeony_yy,aziz,chois}@cbnu.ac.kr, {cdt3017}@naver.com

Abstract. Renewable energy is becoming important all over the world, and in particular, solar energy is spotlighted as an alternative energy that can be easily obtained anywhere. However, unlike fossil energy, solar energy lacks stability because electricity generated varies greatly depending on the weather. In order to overcome these limitations, studies to predict the amount of solar power generation are being actively conducted. In this study, we designed the amount of solar power generation prediction model based on CNN and RNN, and confirmed that deep learning-based power generation prediction is effective.

Keywords : Solar Power, Renewable energy, Deep Learning, Time Series

1 Introduction

According to the 『Communications Earth & Environment』, an international journal published by The Alfred Wegener Institute in Germany, Ingo Sarsgen, Greenland ice, which has the greatest impact on global sea level rise, melted 532 billion tons of ice in 2019, the highest loss ever[1].

Greenhouse gas generated by burning fossil fuels accelerates global warming, but as fossil fuels are also on the verge of depletion, the importance of developing and developing new and renewable energy is increasing[2]. Among them, solar energy that can be regenerated indefinitely using the sun is in the spotlight. The demand for accurate power generation prediction models has recently increased because such the amount of solar power generation fluctuates greatly depending on the location of the sun and external variables. Therefore, this study aims to predict the amount of solar power generation over time based on deep learning.

* Corresponding author

2 Data preprocessing

In this study, solar power and weather data were collected and used for about 4 years from 2016-12-13 to 2020-12-31. Each data has data collected every day from 0:00 to 23:00, and the data consists of a total of 35,487 rows.

Weather data, which are important data for predicting the amount of solar power generation, were collected using the Meteorological Administration API, and it was difficult to collect weather forecast data from the past, so the actual weather observation data were adjusted and used as weather forecast data.

Using Pearson Correlation Coefficient, power generation, temperature, humidity, dew point temperature, insolation, and total amount of cloud were used as variables.

The solar power plant data has a problem in that the power generation value is stored as '0' due to a system error and a network communication error. Since these values have a great influence on the prediction of power generation, sections recorded as '0' for more than 17 hours were removed. In addition, when it was recorded as a '0' value at peak time, which is the time between 10 and 14, it was also classified as an outlier and replaced with an average power generation value per hour. The missing values of weather observation data were calculated by interpolation.

After that, for model learning, Training Set, Validation Set, and Test Set of data were divided into a ratio of 7:2:1.

3 Analysis

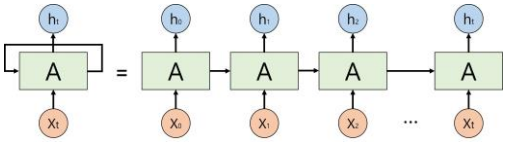


Fig. 1 RNN structure

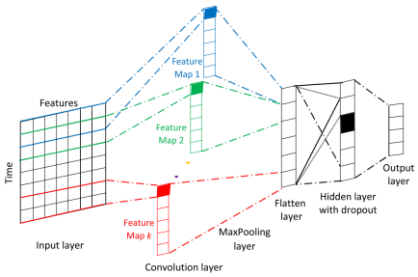


Fig. 2 One-dimensional CNN structure

The deep learning algorithm used in the analysis can be confirmed by <Fig. 1> and <Fig. 2>[3] above. RNN is an artificial neural network specialized in repetitive and sequential data learning, and CNN has a structure that extracts characteristics of data to identify patterns.

After model learning, we used this to designate variables from t-72 to t, which are data of the past 72 hours, as input data based on present time t, and predicted future time-specific generation from t+1 to t+24 through RNN and CNN models.

To this end, representative deep learning models, RNN and CNN, were used to build the amount of solar power generation prediction model, and the model was evaluated by selecting and adjusting hyperparameters.

As a result of model evaluation, the prediction accuracy of RNN and CNN is 85.85% and 87.97%, respectively, indicating the high prediction accuracy.

Table 1. Comparison table of prediction results by algorithm.

	Mean Forecast Error	Mean Forecast Accuracy
RNN	15.80 %	84.20 %
CNN	12.03 %	87.97 %

<Fig. 3> is a graph comparing the predicted power generation and actual power generation of RNN and CNN algorithms with randomly extracted 5-day data from the analysis results. Overall, it can be seen that the power generation prediction value of the CNN is more similar to the actual data than that of the RNN.

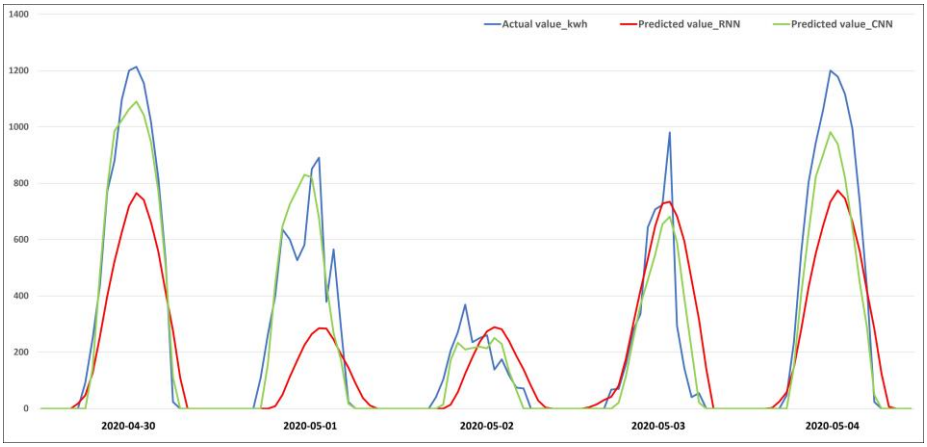


Fig. 3 Comparison graph of prediction results by algorithm.

4 Conclusion

In this study, deep learning algorithms such as CNN and RNN were applied to predict the amount of solar power generation, and it showed high accuracy with 85.85% RNN and 87.97% for CNN. This suggests that applying the deep learning method can be effective when predicting the amount of solar power generation. The analysis was carried out with a simple deep learning structure. If hyperparameter optimization or ensemble method is applied, better performance is expected. In future research, we intend to improve the model's performance by using an ensemble method that combines multiple deep learning models.

Reference

1. Michalea D. King, Ian M. Howat. et al.: Dynamic ice loss from the Greenland Ice Sheet driven by sustained glacier retreat. In: communications Earth & Environment Vol.1, No.1 (2020)
2. Mikael Höök, Tang Xu.: Depletion of fossil fuels and anthropogenic climate change. In: a review. Energy Policy Vol.52, pp.797-809 (2013)
3. Cheikhrouhou, Omar, et al.: One-dimensional CNN approach for ECG arrhythmia analysis in fog-cloud environments. IEEE Access 9, pp.103513-103523 (2021)

THE KOREA
BIG DATA SERVICE SOCIETY
한국빅데이터서비스학회

N13 404-2, Chungbuk University, Chungdae-ro 1
Seowon-Gu, Cheongju, Chungbuk 28644, Korea

Tel: 043-261-3636 / Fax: 043-266-3637

Email: kbigdataservice@gmail.com

Homepage: www.kbigdata.or.kr