The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)
Vol. 21, No. 3, pp.81-89, Jun. 30, 2021. pISSN 2289-0238, eISSN 2289-0246

https://doi.org/10.7236/JIIBC.2021.21.3.81
JIIBC 2021-3-12

개선 된 SSD 기반 사과 감지 알고리즘

Apple Detection Algorithm based on an Improved SSD

정석용*, 이추담*, 왕욱비*, 진락*, 손진구*, 송정영**

Xilong Ding*, Qiutan Li*, Xufei Wang*, Le Chen*, Jinku Son*, Jeong-Young Song**

요 약 자연 조건에서 Apple 감지에는 가림 문제와 작은 대상 감지 어려움이 있다. 본 논문은 SSD 기반의 개선 된 모델을 제안한다. SSD 백본 네트워크 VGG16은 ResNet50 네트워크 모델로 대체되고 수용 필드 구조 RFB 구조가 도입되었다. RFB 모델은 작은 표적의 특징 정보를 증폭하고 작은 표적의 탐지 정확도를 향상시킨다. 유지해야 하는 정보를 필터링하기 위해 주의 메커니즘 (SE)과 결합하면 감지 대상의 의미 정보가 향상된다. 향상된 SSD 알고리즘은 VOC2007 데이터 세트에 대해 학습된다. SSD에 비해 개선 된 알고리즘은 폐색 및 작은 표적 탐지의 정확도를 3.4 % 및 3.9 % 향상 시켰다. 이 알고리즘은 오 탐지율과 누락된 감지율을 향상 시켰다. 본 논문에서 제안한 개선 된 알고리즘은 더 높은 효율성을 갖는다.

Abstract Under natural conditions, Apple detection has the problems of occlusion and small object detection difficulties. This paper proposes an improved model based on SSD. The SSD backbone network VGG16 is replaced with the ResNet50 network model, and the receptive field structure RFB structure is introduced. The RFB model amplifies the feature information of small objects and improves the detection accuracy of small objects. Combined with the attention mechanism (SE) to filter out the information that needs to be retained, the semantic information of the detection objectis enhanced. An improved SSD algorithm is trained on the VOC2007 data set. Compared with SSD, the improved algorithm has increased the accuracy of occlusion and small object detection by 3.4% and 3.9%. The algorithm has improved the false detection rate and missed detection rate. The improved algorithm proposed in this paper has higher efficiency.

Key Words : RFB, Attention Model, SSD, Apple detection, Objection detection, CNN

Ⅰ. Introduction

There are many published papers on the problem of fruit object detection. In many papers, the parameter design is to manually extract the object features in the image, such as common features such as color space and shape[1-2] and special features such as Graylevel Cooccurrence Matrix and HOG[3].Traditional methods can be used for image object recognition, but due to changes in lighting conditions, traditional design feature methods

*정회원, Weifang University of Science and Technology
**정회원, 배재대학교 컴퓨터 공학;
접수일자 2021년 12월 26일, 수정완료 2021년 4월 30일
게재확정일자 2021년 6월 4일

Received: 26 December, 2020 / Revised: 30 April, 2021 / Accepted: 4 June, 2021
**Corresponding Author: jysong@pcu.ac.kr
Dept: Computer engineering, Pai Chai University, Korea

affect the accuracy of detection. in recent years, Convolutional Neural Network (CNN) has been widely used in object classification and object detection[4-5] and has shown better robustness and accuracy in feature extraction and autonomous learning mechanisms[6] Convolutional neural network has made a landmark contribution to image recognition accuracy. CNN has a lot of research in agricultural product detection and recognition: Sa et al. use the Faster-RCNN model to detect sweet peppers in the image, which has the effect of improving accuracy;Yu,et al. use the Mask-RCNN model for detection And segment the strawberries in the greenhouse, so that the robot picks strawberries[7];Tian et al. used DenseNet model to improve YOLOV3,and integrated low-resolution feature layers to achieve apple growth monitoring and evaluation detection[8]; Koirala et al. Improved, studied the MangoYOLO network model, and improved the detection accuracy of mango[9]. Liu et al. proposed an SSD method, which has been well improved in terms of accuracy and real-time performance during object recognition and detection[10].

The paper takes Apple as the research object and improves the classic SSD algorithm to improve the ability of small object detection. In the feature extraction process, this paper extracts the features of the feature map output by the convolutional layer through the RFB receiving field, thereby reducing the feature loss. Attention mechanism is introduced to filter out the information that needs to be retained, and to enhance the semantic information of high-level feature maps. The SSD backbone network VGG16 is replaced with the deep residual network ResNet50.

II. Related Works

1. SSD Model

The SSD network adopts the idea of regression, which simplifies the complexity of network calculation and improves the real-time performance of the algorithm. SSD uses multi-scale object feature extraction to improve the robustness of objects of different scales. The basic network of SSD uses the VGG16 network. The two fully connected layers of VGG16 are replaced with convolutional layers, and the random inactivation layer and FC layer are removed, and the depth of the convolutional layer is increased. Conv8_2, Conv9_2, Conv10_2 are added at the back, Conv11_2 four convolutional layers. The input image size is 300×300. as shown in Figure1.

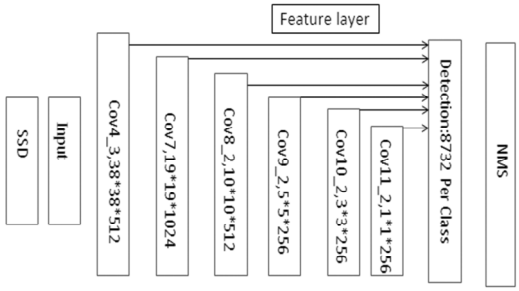

그림 1. SSD 모델
Fig. 1. SSD model

2. Residual Network

The deeper the network, the more obvious the phenomenon of gradient disappearance. He KM et al. proposed a residual network, which can realize identity mapping, that is, identity mapping. The most important part of Resident is shown in Figure3. The residual module shown. X is the input of the residual block, F(x) is the residual, and F(x) is the output after the first layer linear change and activation. Assuming that the desired network layer relationship is mapped to H(x), and any complex function can be approximately expressed by n nonlinear layers, the network can be fitted into another mapping F(x)=H(x)-x, So the original mapping becomes H(x)=F(x)+x, no additional parameters and computational complexity will be added. The deeper the network, the higher the accuracy, as shown in Figure2.

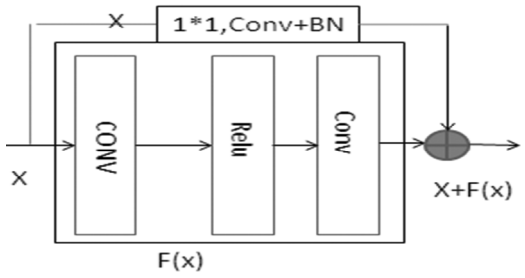

그림 2. 잔여 블록 구조 다이어그램
Fig. 2.Residual block structure diagram

Ⅲ. The algorithm of this Research

1. Improved SSD model

Under natural conditions, the low-level feature maps of traditional SSD contain less semantic information, and there are problems of missed detection and false detection of small objects. As the network depth increases, the accuracy of extracting features decreases. In order to improve the defects of the SSD model itself, an RA-SSD network model is proposed.

The improved SSD basic network model is ResNet50 as the basic network. The network has deeper layers and can extract deeper feature information. On the basis of ResNet, the original 4th residual block and the final average merge layer and fully connected layer are deleted. In Conv3_x and Conv4_x, the RFB model is introduced to increase the network width. The RFB model amplifies the feature information of small objects and improves the detection accuracy of small objects. The improved SSD adds Conv5_x, Conv6_x, Conv7_x and Conv8 four convolutional layers. The first three layers introduce an attention mechanism. The attention module is used to perform a global average pooling operation on the three feature maps, and the model obtains important information including image features.

The two convolutional layers at the back end of ResNet 50 and the last 4 convolutions of the model together form 6 convolutional layers as the detection object. Different predicted anchor frames are used, and the anchor frame size generation rules are the same as the original SSD. The improved SSD network model structure is shown in Figure5. Candidate frames of different sizes realize the detection of multi-scale object positions and categories, and finally produce the final detection results through non-maximum suppression (NMS).

2. Introducing the RFB model

In the improved SSD backbone network Resnet50, Conv3_x and Conv4_x introduce the RFB model, whose feature map sizes are 38×38×512 and 19×19×1024, respectively. The RFB model goes through two 1×1convs, and a part of the output Feature Map goes through 1×1conv, 3×3conv, 5×5conv, and then corresponds to 2×2conv with rate=1, 3×3cov, rate with rate=3, 3×3conv, and finally fusion features, after 1×1conv, the shortcut 1×1conv features are fused, and the prediction result is given through the activation function Relu, and the feature map scale remains unchanged. In the RFB structure, the feature information of the small object part is enlarged to improve the detection accuracy of the small object and the low-level feature map is enhanced to improve the semantic information of the feature map. The RFB structure is shown in Figure3.

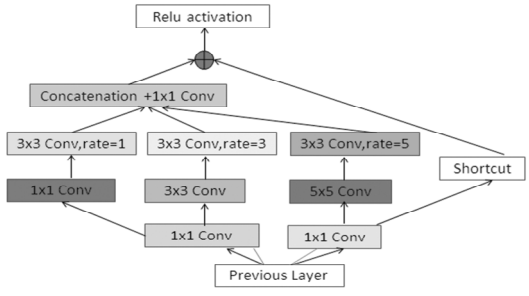

그림 3. 잔여 블록 구조 다이어그램
Fig. 3. RFB structure diagram

2. add channel attention mechanism

In order to further improve the accuracy of the algorithm, reduce the missed detection of the object, and strengthen the robustness of the algorithm, the improved SSD uses the idea of SeNet, this paper adds the attention mechanism to the feature fusion process. In the improved SSD backbone network Resnet50, Conv5_x, Conv6_x and Conv7_x introduce the attention model. The realization process of the attention model module is divided into three steps: squeeze, excitation, and Attention. The formula is as follows:

$$Y=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}X(i,j) \quad (1)$$

X represents a low-resolution and high-semantic information graph,H represents the input length, W represents the width, and C represents the number of channels. After the information graph is compressed, a one-dimensional array of length C is obtained.(i, j) means that there are horizontal and vertical coordinate points on the feature map of size H×W, and the output Y is a one-dimensional array of length C. The formula is a global average pooling operation. All the eigenvalues in each channel are averaged and then added. The activation process is to model the degree of correlation between each channel. The formula is as follows:

$$S=sigmoid(W_2 \cdot R(W_1 Y)) \quad (2)$$

Function R() is Relu() activation function. The dimension of W_1 is C'×C, and the dimension of W_2 is C×C'. In the text, C' defaults to C×1/4, and the final dimension of S is C×1×1. Then through the feature weighting operation: X'=X×S, the original input is sent to the improved SSD network of this paper for detection, and X is replaced with the feature X' obtained by the attention module, and added to the original network structure for object detection. This can enhance the focus on key channels.

Based on the improved SSD network, this paper uses Resnet50 to propose an RA-SSD model, as shown in Figure5. The convolutional layer structure added to the network model proposed in this paper is shown in Figure4.

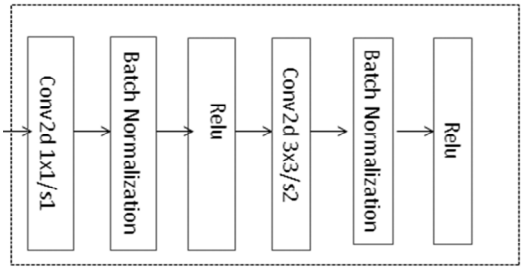

그림 4. 컨벌루션 레이어 구조 다이어그램
Fig. 4. Convolutional layer structure diagram

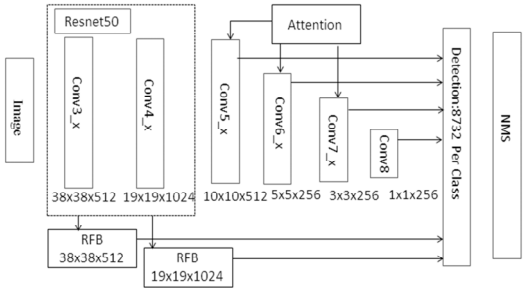

그림 5. RA-SSD 구조 다이어그램
Fig. 5. RA-SSD structure diagram

Ⅳ. Experimental results and analysis

1. Experimental software and hardware environment

The experiment uses the Pytorch deep learning framework for training and testing. The hardware configuration is Intel Core i7 CPU processor,32GB memory, NVIDIA TITAN Xp type 16GB GPU graphics card, operating system is Linux Ubuntu16.04, computing framework is CUDA10.0,neural network acceleration library for CUDNN7.5,the python programming language is used to realize the construction, training and verification of the network model of this paper.

2. Experimental data and preprocessing

The test image data collection device is a digital camera. In order to ensure the diversity of the samples, 731 apple images were collected under natural weather conditions. The image resolution is 2736×2736 pixels, the format is JPEG, and the distance to the photographed fruit is about 1 meter. In order to improve the calculation time of the algorithm, the pixel resolution of the image is adjusted to 300×300. After image enhancement, there are a total of 2924 images. Randomly select 585 images as the validation set, and select 2339 images in the training set. There is no overlap between the two images.

3. Experimental model training parameter settings

In order to improve efficiency, first use the pre-trained model on ImageNet to initialize the network parameters. Use transfer learning to train the model. The learning momentum is set to 0.9, the initial learning rate is set to 0.001, the weight decay is set to 0.0005, and the batch size is 32. The total number of training rounds is set to 300, and the verification period is set to 5000. When the accuracy of the model reaches convergence, the training is stopped, and the weights obtained through training are used as the initial weights. Targets identified by the network and areas with a probability of becoming Apple's objects greater than 0.7 are considered reserved areas.

4. Evaluation Index

The metric used in this paper is AP, It is used to evaluate the accuracy of model detection. Calculate the average accuracy of each category in the data set, the formula is show:

$$AP = \frac{1}{T}\sum_{n=1}^{k} M_n \times \frac{T_n}{n} \qquad (3)$$

T is the number of images in the dataset, and k is the total number of object objects. M_n=1, it means that the n is the detection object, otherwise $M_n = 0$. T_n represents the number of objects contained in n images that have been detected. Calculate the average value according to the formula. The larger the mAP value, the higher the accuracy of model detection.

5. Experimental comparison of model detection

The paper uses the VOC2007 data set as the experimental object to verify the average accuracy of the model. The VOC2007 data set includes 9963 pictures and a total of 20 objects. In order to verify the effect of RA-SSD network, this article uses SSD network without RFB module, SSD network without attention mechanism and RA-SSD network for prediction. AP is used as the standard. The experimental results are shown in Table1.

표 1. VOC2007에 따른 다양한 융합 방법의 테스트 결과
Table 1. test results of different under VOC2007

Algorithm	Dataset condition	AP %
RA-SSD(no RFB)	Same	78..1
RA-SSD (no Attention)	Same	79.5
RA-SSD	Same	81.4

This paper takes input 300×300 resolution pictures as the experimental object, and studies the calculation effects of different network models. Comparing different network SSD models with different picture sizes, the experimental results are shown in Table 3 and Figure7. In the experiments of occlusion and small object detection, the improved SSD model Apple has better detection accuracy. Apple suitable for small objects and occlusion conditions. This article needs to consider recall and accuracy in the process of Apple object recognition, so the F1 value is used to evaluate the recognition result.

$$F1 = \frac{2PR}{(P+R)} \quad (4) \qquad P = \frac{TP}{(TP+FP)} \quad (5)$$

$$R = \frac{TP}{(TP+FN)} \quad (6)$$

P is the accuracy rate, R is the recall rate; TP is the number of recognized apples, FP is the number of misrecognized apples, and FN is the number of unrecognized apples.

In order to screen out the network model with the best recognition effect, 3 kinds of networks were tested on 585 pictures. Compare the recognition results on the collection. The DSSD model is based on the backbone ResNet101 to improve the SSD. Therefore, the paper uses DSSD、SSD and RA-SSD for comparison. The results are shown in Table2. The result is that the F1 value based on RA-SSD is the highest, so the network model with the best recognition result is selected.

표 2. 다양한 SSD 네트워크의 감지 결과
Table 2. detection results of different model algorithm

Algorithm	Backbone network	F1 %
DSSD	ResNet101	87..25
SSD	VGG16	86.73
RA-SSD	ResNet50	89.56

6. Apple object recognition results and analysis

The paper uses RA-SSD algorithm compared with DSSD and SSD in small target detection and occluded target detection analysis.

(1) Recognition result small target detection analysis

The paper treats objects whose target area is smaller than 32×32 as small objects. The reason why small object detection is difficult: In the feature extraction process, the expression of the extracted features is weak. Therefore, different models will produce some missed detections and detection errors. Table3 and Figure6 show that compared with SSD, the algorithm improves the Average accuracy of small object detection by 3.9%. From Table3 concluded that compared with SSD algorithm, the algorithm reduces the false detection rate by 2.5% and the missing detection rate by 3%.. Compared with DSSD algorithm, this algorithm has certain advantages reduces false detection rate by 0.7% and missed detection rate by 1.7%.

표 3. 소형 표적 감지에서 서로 다른 알고리즘 비교
Table 3. Comparison of different algorithms in small object detection

Algorithm	False rate	missed rate	AP %
SSD	13.1	10.9	86.2
DSSD	11.3	9.6	87.4
RA-SSD	10.6	7.9	90.1

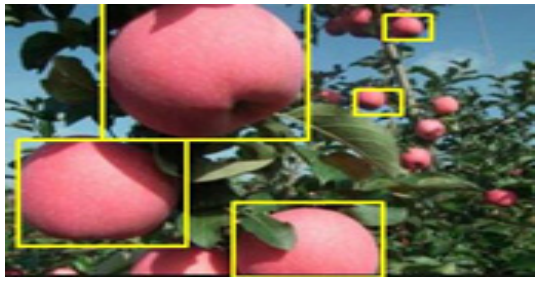

그림 6(a). SSD 작은 표적 감지
Fig. 6(a). SSD small object detection

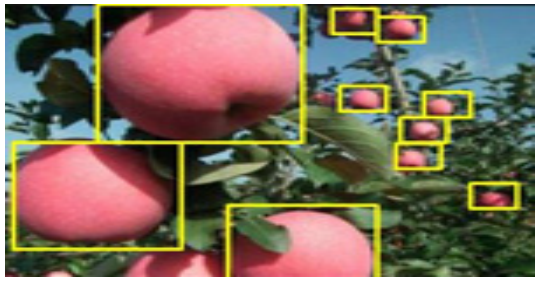

그림 6(b). DSSD 작은 표적 감지
Fig. 6(b). DSSD small object detection

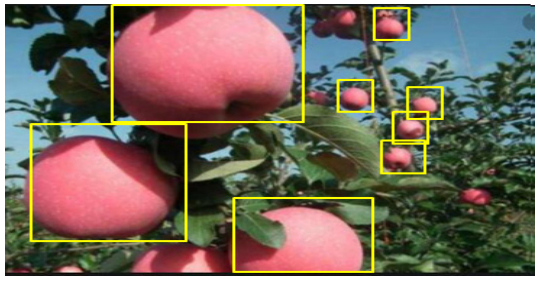

그림 6(c). RA-SSD 작은 표적 감지
Fig. 6(c). RA-SSD small object detection

(2) Recognition result occlusion target analysis

The algorithm texted the accuracy of apple detection in a occlusion environment. Occlusion increased the difficulty of fruit detection. The experiment uses occluded apple images for testing to evaluate the accuracy rate of the model. When the model tests that the occluded apple area is less than 1/3 of the entire apple area, the model detects the apple with an accuracy rate of 88.7%. Compared with the SSD and DSSD improved algorithm of detection Average accuracy under conditions of occlusion increased by 3.4% and 1.5%. As shown in Table4 and Figure7.

표 4 살짝 가린 상태에서 사과 비교
Table 4. Comparison of detection indexes of occluded apples

Algorithm	AP %
SSD	85.3
DSSD	87.2
RA-SSD	88.7

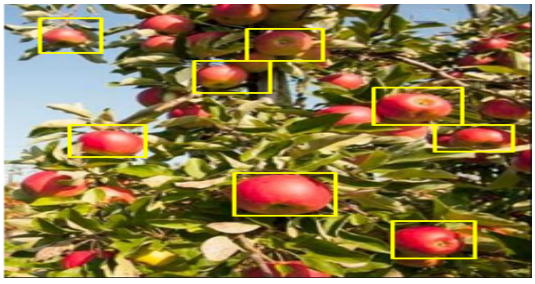

그림 7(a). SSD 폐색 물체 감지 결과
Fig. 7(a). SSD occlusion object detection results

그림 7(b). DSSD 폐색 물체 감지 결과
Fig. 7(b). DSSD occlusion object detection results

그림 7(c). RA-SSD 폐색 물체 감지 결과
Fig. 7(c). RA-SSD occlusion object detection results

V. Conclusion

This paper proposes a detection algorithm based on an improved SSD model. Replace the VGG16 of the classic SSD model with ResNet50, and use the RFB model to improve the accuracy of small object detection and improve missed objects. The attention mechanism is introduced to enhance the semantic information of high-level feature maps. Experimental data shows that for the detection of small objects, the accuracy of the RA-SSD model is 3.9% higher than that of SSD. For the detection of occluded objects, the accuracy of this model is 3.4% higher than that of SSD. Experiments verify the effectiveness of the improved method proposed in this paper.

References

[1] Gené-Mola J, et al, "Multimodal deep learning for Fuji apple detection using RGB-D cameras and their radiometric capabilities", Computers and Electronics in Agriculture, Vol. 162, pp. 689-698, July 2019. DOI: https://doi.org/10.1016/j.compag.2019.05.016.

[2] Underwood J P, Hung C, et al, "Mapping al-mond orchard canopy volume, flowers, fruit and yield using LiDAR and vision sensors", Computers and Electronics in Agriculture, Vol. 130, pp. 83-96, November 2016. DOI: https://doi.org/10.1016/j.compag.2016.09.014

[3] Zhou R, Damerow L, et al, "Using colour feature of cv 'Gala' apple fruits in an orchard in image processing to predict yield", Precision Agriculture, Vol. 13, No. 5, pp. 568-580, June 2012.

DOI: https://doi.org/10.1007/ s11119-012-9269-2

[4] Kestur R, Meduri A, Narasipura O, "MangoNet: A deep semantic segmentation architecture for a method to detect and count mangoes in an open orchard", Engineering Applications of Artificial Intelligence, Vol. 77, pp. 59-69, January 2019.
DOI: https://doi.org/10.1016/j.engappai.2018.09.011

[5] Szegedy C, Ioffe S, et al, "Inception-v4,inception-resnet and the impact of residual connections on learning", in Thirty-First AAAI Conference on Artificial Intelligence, pp. 1-12. February, 4-9, 2017.

[6] Liu W, Wang Z, et al, "A survey of deep neural network architectures and their applications", Neurocom-puting, Vol. 234, pp. 11-26, April 2017.
DOI: https://doi.org/10.1016/j.neucom.2016.12.038

[7] Sa I, Ge Z, et al. "Deepfruits: a fruit detection system using deep neural networks", Sensors, Vol. 16, No. 8, pp. 66-75, August 2016.
DOI: https://doi.org/10.3390/s16081222

[8] Kang, Tae-Wan, et al, "A Study on the Development of Driving Simulator for Improvement of Unmanned Vehicle Remote Control'", Journal of the Korea Academia-Industrial cooperation Society (JKAIS), Vol. 20, No. 6, pp. 86-94, June 2019.
DOI : https://doi.org/10.5762/KAIS.2019.20.6.86

[9] Eun-Gyu Ham, et al "Model Implementation of Reinforcement Learning for Trading Prediction Using Deep Q Network", The Journal of Korean Institute of Information Technology, Vol. 17, No. 4, pp. 1-8, April 2019.
DOI: 10.14801/jkiit.2019.17.4.1

[10] Soo-Mok Jung, "Advanced Pixel Value Prediction Algorithm using Edge Characteristics in Image", International Journal of Internet, Broadcasting and Communication, Vol. 12, No. 1, pp. 111-115, January 2020.
DOI: http://dx.doi.org/10.7236/IJIBC.2020.12.1.111

저 자 소 개

Xilong Ding(학생회원)

• Xilong Ding received the master's degree in Computer Application Technology form Ocean University of China in 2007. PhD student at Pai Chai University. His research interests include Machine learning and Computer vision technology.

Qiutan Li(학생회원)

• Qiutan Li graduated from Shandong Normal University in 2004 with a bachelor's degree in computer science and technology. He received the master degree in Computer Engineering from Shandong University of Science and Technology in 2009. He is currently studying for the computer engineer degree of Pai Chai University. His research interests Artificial Intelligence

Xufei Wang(학생회원)

• Xufei Wang received the master degree in mechanical engineering from Xinjiang University, China in 2007. He is pursuing a doctorate in computer engineering at Pai Chai University in Korea. His research interests include Self-driving, Machine learning.

Le Chen(학생회원)

• Le Chen received the master degree from Nanjing University of Technology in 2018. He is currently studying for the computer engineer degree of Pai Chai University. His research interests include Software Engineer, Artificial Intelligence..

Jinku Son(학생회원)

• Jinku Son received master degree in computer engineering from Waseda University. Japan in 2011. His research interests include AI, Big data, BPM.

Jeong Young Song(정회원)

- 1984.2: B.S. Degree, Computer Engineering, Hannam Univ. S. Korea
- 1992.3: M.S. Degree, Electrical Information and System, Waseda Univ., Japan.
- 1995.3: Ph.D. Degree, Electrical Information and System, Waseda Univ., Japan.
- 1995.3~1997.2: Computer Science, CheongUn Univ., Korea.
- 1997.3~Present: Computer Engineering, Professor, PaiChai Univ., S. Korea.
- 2011.9~2012.8: Invited Scholarship Professor, Department of EE(Electrical Engineering), ISU(Idaho State University), USA.
- Research Interests : Pattern Processing(Image, Speech, Character), Machine Learning. etc..