

Improved Loss Function for Defect Detection of Mobile Phone Screen

Zixuan Wang, Chengyuan Li, Yifan Zhang, Shuo Chang, Yilei Li and Fan Zhang
{ beiyowangzi@bupt.edu.cn, lcyan@bupt.edu.cn, zhangyf@bupt.edu.cn, changshuo@bupt.edu.cn,
liyilei@bupt.edu.cn, zhangfan2015@bupt.edu.cn }

Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunication, Beijing, P.R.China, 100876.
School of Information and Communication Engineering, Beijing Information Science and Technology University, Beijing, P.R.China, 100085.

Abstract. Due to the excellent feature learning and representation capabilities of deep learning, the method based on deep learning for mobile phone screen defect detection is gradually being applied to industrial detection. Nowadays, the cross-entropy loss commonly used in deep learning only focuses on the differences between different classes with less intra-class differences. It leads to poor discriminative ability of the model when the similarity between training samples is high. The contrastive loss reduces the intra-class variations and can distinguish between similar objects from different classes. Given the above analysis, we propose a Siamese network for mobile phone screen defect detection (SMSDD) using combined contrastive loss and cross-entropy loss, thereby enhancing the discriminative ability of model. Numerical results show that SMSDD achieves comparable performance.

Keywords: Defect Detection, Deep Learning, Cross-entropy Loss, Contrastive Loss, Siamese Network.

1 Introduction

Nowadays, mobile phones have become an indispensable part of daily life. The transparent glass screen of the mobile phone is the main window of human-machine interaction. If there are defects on the phone screen, it will affect the user experience directly. Therefore, to satisfy the high-quality and high-resolution requirements of the phone screen, screen defect detection is necessary.

Till now, the methods of mobile phone screen detection mainly rely on manual detection or traditional image processing. Traditional manual detection has shortcomings such as low speed, low accuracy and inconsistent detection standards, etc. Therefore, it can not adapt to the current industrial detection requirements [1]. Although methods based on image processing are faster, more reliable, and more accurate than manual detection, the feature extraction capability is unstable. Due to influences of illumination, camera noise, screen texture, etc., these methods may result in inaccurate detection results [2].

With the deepening of the research on Convolutional Neural Network (CNN), the method based on CNN has achieved superior results on various computer vision tasks [3-5],[8]. And methods based on CNN for mobile phone screen defect detection have become a good solution

[2],[6],[7]. The loss function is an important part for CNN. On one hand, it controls the ultimate goal of network optimization. On the other hand, it determines the learning direction of the network. The cross-entropy loss [9] is used to classify the images to a certain class generally. It mainly penalizes the inter-class loss, and there is no analysis of intra-class differences. In the screen defect detection, the area containing the defect accounts for a little proportion of the screen, mainly less than 0.1%. This leads to the similarity of some samples in defective and non-defective classes is high. Meanwhile, as shown in **Figure 1**, the defects have large differences in shape and small appearance similarity. So, such variations within the defects could overwhelm the variations within class differences and make screen defect detection challenging. This requires the model to consider the intra-class compactness while ensuring the inter-class separability. For the purpose, the contrastive loss [10] which mainly concentrates on reducing intra-class differences can meet the demand in the screen defect detection. So, we add the contrastive loss as a part of loss function of the network.

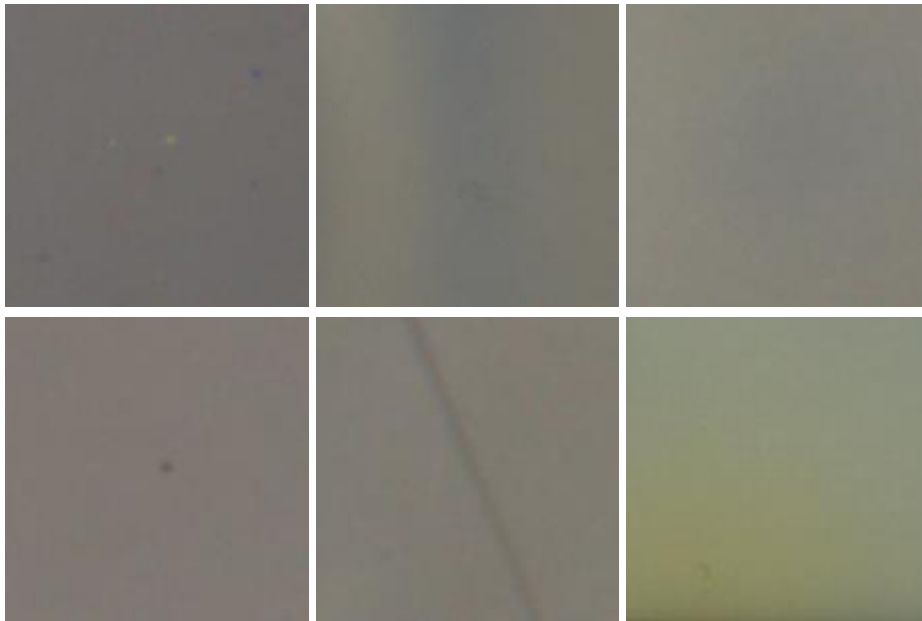


Fig. 1. Six different types of screen defects. They are various in area, shape, pixel value, etc.

In this paper, the mobile phone screen defect detection is transformed into binary classification, which only contains no-defective images and defective images. A Siamese network is proposed for mobile phone screen defect detection (SMSDD), which is supervised by the improved loss function that combines the cross-entropy loss with contrastive loss. Extensive experiments are conducted to prove the effectiveness of the proposed method on Mobile-Screen-Defect dataset.

The rest of this paper is organized as follows. Section 2 reviews the related work. In Section 3, details of the proposed SMSDD and the improved loss function are presented. In Section 4, the implementation details in model training are presented. The numerical results are provided in Section 5. Finally, Section 6 concludes the paper.

2 Related Work

2.1 Traditional Image Processing For Screen Defect Detection

Kim [11] et al. proposed a method that detected defects by reducing the threshold level of gray unevenness, which can detect most types of defects other than line defects. Lee and Yoo [12] et al. used the background subtraction method, with a two-dimensional curve to estimate the image background, to detect mura. With Fourier transform to eliminate the background interference of defects, Tsai and Tseng [14] et al. utilized a simple method of threshold processing to complete the image segmentation. In short, traditional image processing methods has high time complexity and weak robustness.

2.2 Convolutional Neural Network For Classification

With the rapid development of deep learning, many works for large-scale image classification have been proposed [13]. Krizhevsky [3] et al. introduced deep CNN into image classification for the first time. Simonyan and Zisserman proposed a very deep CNN in [15], and the performance was significantly improved by increasing the depth and using small convolutional filters of size 3×3 . The loss function commonly used in the classification system is the cross-entropy loss [9], which mainly penalizes the inter-class loss. However, defective and non-defective mobile phone screen samples have similar features. The variations within class differences are not enough obvious, where the result of classification is easy to be disturbed. So, only using the cross-entropy loss can not classify samples accurately. Therefore, it is not a wise choice to use a general image classification system for the mobile phone screen defect detection.

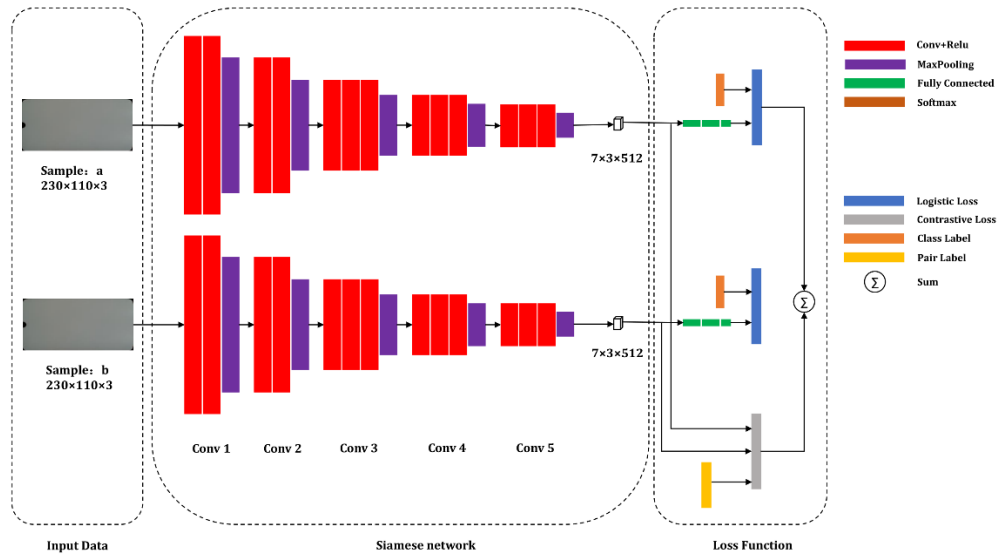


Fig. 2. The framework of SMSDD.

2.3 Contrastive Loss For CNN

The contrastive loss [10] is mainly used in Siamese framework. It can effectively reduce the intra-class variations by pulling feature vectors from the same instance together. Based on the contrastive loss, Yi Sun [16] et al. increased the inter-personal variations by drawing features extracted from different identities apart, while reduced the intra-personal variations by pulling features extracted from the same identity together in face recognition task. S. Chopra [17] et al. enhanced the robustness of the network to the nonlinear geometric transformation of the data by controlling the "semantic" distance between the paired data. Defect samples of mobile phone screen usually have great differences in area, pixel value, etc. Intra-class differences have great impact on classification [18] and we need to consider the intra-class compactness. In this case, the contrastive loss is helpful to enhance the classification ability of the model for the mobile phone screen defect detection.

3 The Proposed SMSDD Scheme

In this section, details of the proposed SMSDD are firstly described. Then, a brief introduction about contrastive loss and cross-entropy loss is given. Finally, We visualize the effects of the improved loss function.

3.1 Framework

The framework of SMSDD is depicted in **Figure 2**. The proposed network consists of two branches, and they share the same structure and weight. As shown in **Figure 2**, the backbone of Siamese Network is VGG16 [13] without fully connected (FC) layer. Samples a and b stand for input pair-data to update the Siamese network and the size is $230 \times 110 \times 3$. The class label is considered to be positive if there are no defects in the screen:

$$y = \begin{cases} 0, & \text{no - defective} \\ 1, & \text{defective} \end{cases} \quad (1)$$

The pair label is determined by whether a and b belong to the same class:

$$Y = \begin{cases} 1, & \text{the same class} \\ -1, & \text{other} \end{cases} \quad (2)$$

The loss function of the network is shown on the right side of **Figure 2**. For fully connected layer, it has two kinds of outputs, corresponding to the defective or no-defective image. The output of fully connected layer is used to calculate the cross-entropy loss. Under the supervision of the pair label, the outputs of Siamese Network are used to calculate the contrastive loss. Finally, the sum of the cross-entropy loss and contrastive loss is used to propagate gradients for updating the network parameters.

3.2 Loss Function

Cross-entropy Loss. Cross entropy loss is mainly used for classification tasks. In the screen defect detection task, there are two classes: no-defective and defective images. The corresponding cross-entropy loss is:

$$L_1 = -\frac{1}{N} \sum_{i=1}^N [y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)})] \quad (3)$$

where \hat{y} is the score of a single exemplar-candidate pair and y is the class label. N represents the total number of samples in the training set.

The cross-entropy loss mainly focuses on increasing the margin among candidates from different classes. Namely, there is no consideration about the intra-class variations, where the prediction result of the model is greatly reduced.

Contrastive Loss. We want the feature representations of deflections are close enough for positive pairs, and far away at least by m for negative pairs, where m is a hyper parameter, which means that we only consider paired data belonging to different classes with Euclidean distance less than m . Therefore, we employ the contrastive loss:

$$L_2 = \frac{1}{2N} \sum_{k=1}^N Y D_W^2 + (1 - Y) \max(m - D_W, 0)^2 \quad (4)$$

Where

$$D_W(X_1, X_2) = \|X_1 - X_2\|_2 = (\sum_{i=1}^P (X_1^i - X_2^i)^2)^{\frac{1}{2}} \quad (5)$$

D_W is the Euclidean distance between the sample feature X_1 and X_2 , where P is the eigenvector dimension. The variable Y stands for the pair label. When $Y = 1$, the two candidate samples (X_1 , X_2) are from the same class, and the loss function in equation (4) becomes equation (6):

$$L_S = \frac{1}{2N} \sum_{k=1}^N Y D_W^2 \quad (6)$$

So, if the Euclidean distance between the two samples in the feature space is large, the loss will be large. This means that the current model does not meet the requirement of intra-class compactness and needs to continue training. Otherwise, when $Y = 0$, the loss function is :

$$L_D = (1 - Y) \max(m - D_W)^2 \quad (7)$$

With the decrease of Euclidean distance D_W , the loss will be larger, which meets the requirement of inter-class separability.

3.3 Combined Loss

By integrating the contrastive loss into cross-entropy loss, a new loss function to supervise the SNMSDD is:

$$L = L_1 + \beta L_2 . \quad (8)$$

In equation (8), the hyperparameter β is a weight to balance the effect between the cross-entropy loss and contrastive loss.

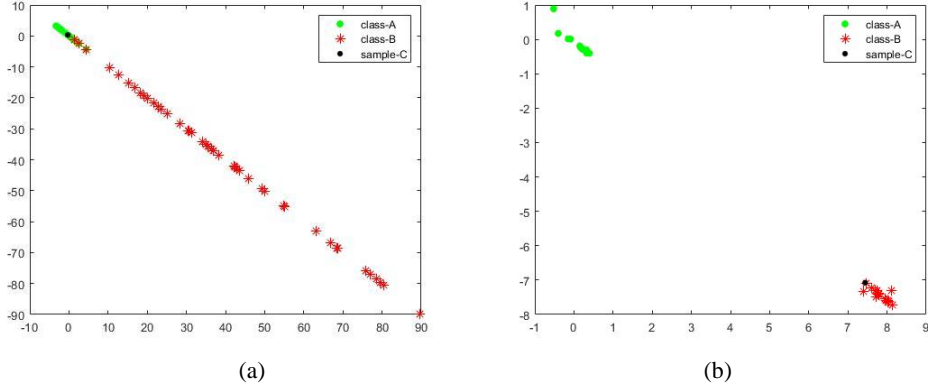


Fig. 3. Visualization of effects of the original and improved loss function.

Figure 3 visualizes effects of the original and improved loss function. For the sample C (black), although it belongs to class B (red), it has some same characteristics as class A (green). If we use the original loss function, its position in feature space is close to class A, which makes it difficult to classify. In **Figure 3(b)**, the improved loss function is used to control the distance between A and B in a certain range, which draws the sample C away from class A, making the model easy to distinguish similar samples.

4 Implementation Details

The proposed algorithm SMSDD is implemented in python on a Dell R7300 server with an NVidia 1080Ti GPU. The experiment is performed on Linux with a processor of Intel (R) Xeon (R) CPU E5-2620 v3 @ 2.40GHz and 21G RAM. The learning rate for model training is 0.01 in all epochs.

5 Experiments

5.1 Dataset and Preprocessing

Dataset. Screen images are taken by industrial cameras mounted on the production line as experimental data, and each image will be read and marked by experienced workers. We collected 1200 images, including 168 images of defective products and 1032 images of no-defective products.

Preprocessing. Due to the extensive black background in collected images, most of black border areas will be extracted during the feature extraction process, and the features of the mobile phone screen will be ignored. To avoid the influence of the background in the training,

the Hough transform [19] is used to find the edge points and the perspective transform is used to correct the region of interest (RoI) of screens. After the preprocessing, 80% of images are used as the training set for model training, 10% are used as the validation set, and the remaining 10% are used as the test set.

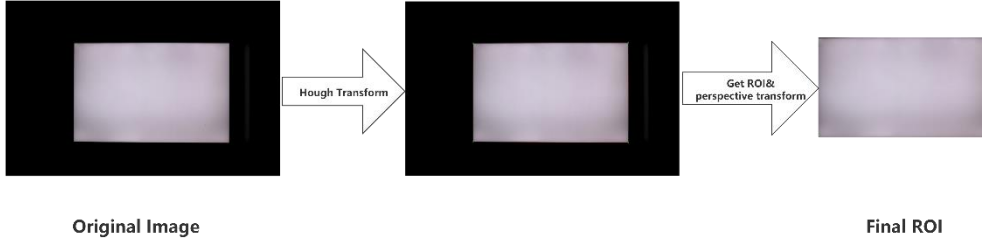


Fig. 4. The process of getting RoI.

Table 1. The loss function of different models.

Methods	Loss Function	Hyper-parameter
SMSDD-A	Cross-entropy Loss	$\beta = 0$
SMSDD-B	Cross-entropy Loss + β Contrastive Loss	$\beta = 0.1$
SMSDD-C	Cross-entropy Loss + β Contrastive Loss	$\beta = 0.01$
SMSDD-D	Cross-entropy Loss + β Contrastive Loss	$\beta = 0.001$
SMSDD-E	Cross-entropy Loss + β Contrastive Loss	$\beta = 0.0001$

5.2 Evaluation

To evaluate the performance of the proposed model, five contrastive experiments are carried out, as shown in Table 1.

Loss Evaluation. The comparison of training loss of different models are shown in **Figure 5**. SMSDD-A only use the cross-entropy loss to update the network parameters. Because it only penalizes the inter-class loss and has no consideration about the intra-class variations, the training result of SMSDD-A is inferior. Finally, the loss only converges to 0.23.

When $\beta = 0.1$, the cross-entropy loss plays a small role in model updating. This makes SMSDD-B unable to increase the margin among candidates from different classes correctly. According to the experimental result, SMSDD-B is the worst. Due to the appropriate values of β , SMSDD-C and SMSDD-D have a quick convergence speed and the final loss is low. They also perform well in validation and test sets, which can meet industry standards. Because β in SMSDD-E is small, the contrastive loss has less effect on the training process. Hence, the result of SMSDD-E is worse than the results of SMSDD-C and SMSDD-D, but better than the result of SMSDD-A.

Accuracy Evaluation. The results of accuracy based on different loss functions are in Table 2. Without the bells and whistles, our proposed SMSDD outperforms the original algorithm on validation and test sets in accuracy. Comparing to the model using only cross-entropy loss

(SMSDD-A), the SMSDD-C has made gains about 11% on validation set and 10% on test set in accuracy as the best choice of hyper-parameter β , respectively. This is due to the benefit of the contrastive loss, which is added to the loss function for updating the network parameters.

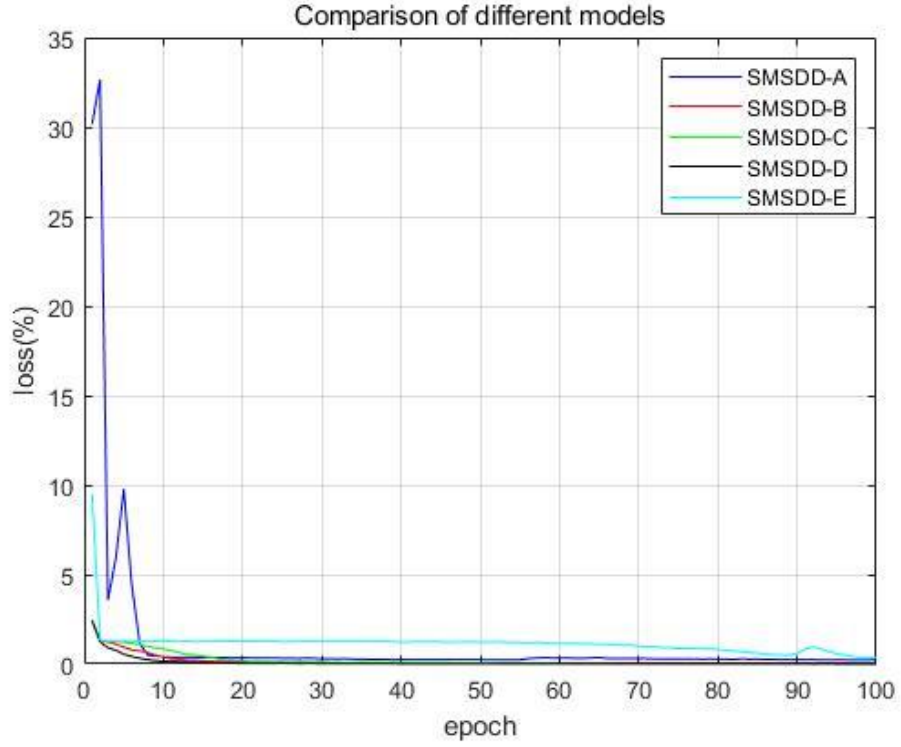


Fig. 5. Comparison of training loss of different models.

Table 2. Accuracy of different models on validation and test sets.

Methods	Valication Set(%)	Test Set(%)
SMSDD-A	85	83
SMSDD-B	79	74
SMSDD-C	96	93
SMSDD-D	95	91
SMSDD-E	91	88

6 Conclusion

In this paper, we introduce deep learning into mobile phone screen defect detection. Besides, a new loss function that combines the cross-entropy loss with the contrastive loss is proposed. The new loss considers the intra-class compactness and inter-class separability. Without the bells and whistles, the proposed algorithm outperforms the method supervised by

cross-entropy loss. In the future work, we plan to introduce the new loss function into multi-classification tasks.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant (61801052), in part by the National Key Research and Development Program of China under Grant (2019YFB1804400, 2018YFF0301202), and in part by the Beijing Natural Science Foundation under Grants (4202046).

References

- [1] R. Gao, W. RENXD: Research on mobile phone screen defect detection system based on image processing. *Measurement & Control Technology*. pp. 26-30 (2017)
- [2] J. Lei, X. Gao, Z. Feng, H. Qiu, and M. Song: Scale insensitive and focus driven mobile screen defect detection in industry. *Neurocomputing*. pp. 72-81 (2018)
- [3] G. E. Hinton , I. Sutskever, and A. Krizhevsky: Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*. pp. 1079-1105 (2012)
- [4] J. Long, E. Shelhamer, and T. Darrell: Fully convolutional networks for semantic segmentation. Vol. 39, pp. 640-651. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, US* (2015)
- [5] S. Ren, K. He, R. Girshick, and J. Sun: Faster r-cnn: Towards real-time object detection with region proposal networks. Vol. 39, pp. 1137-1149. *Advances in Neural Information Processing Systems, Canada* (2015)
- [6] Z.-C. Yuan, Z.-T. Zhang, H. Su, L. Zhang, F. Shen, and F. Zhang: Vision-based defect detection for mobile phone cover glass using deep neural networks. *International Journal of Precision Engineering and Manufacturing*. pp. 801-810 (2018)
- [7] C. Jian, J. Gao, and Y. Ao: Automatic surface defect detection for mobile phone screen glass based on machine vision. *Applied Soft Computing*. pp. 348-358 (2017)
- [8] Sai Huang, YuanYuan Yao, Zhiqing Wei et al.: Automatic Modulation Classification of Overlapped Sources Using Multiple Cumulants. *IEEE Transactions on Vehicular Technology*. pp. 6089-6101 (2017)
- [9] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein: A tutorial on the cross-entropy method. *Annals of operations research*. pp. 19-67 (2005)
- [10] R. Hadsell, S. Chopra, and Y. LeCun: Dimensionality reduction by learning an invariant mapping. Vol. 2, pp. 1735-1742. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, US* (2006)
- [11] W. S. Kim, D. M. Kwak, Y. C. Song, D. H. Choi, and K. H. Park: Detection of spot-type defects on liquid crystal display modules. Vol. 270, pp. 808-813. *Advances in Nondestructive Evaluation, US* (2004)
- [12] J. Y. Lee and S. I. Yoo: Automatic detection of region-mura defect in tft-lcd. *IEICE Transaction on Information and Systems*. pp. 2371-2378 (2004)
- [13] S. Huang, Y. Jiang, X. Qin et al.: Automatic Modulation Classification of Overlapped Sources Using Multi-Gene Genetic Programming With Structural Risk Minimization Principle. *IEEE Access*. pp. 48827-48839 (2018)
- [14] D.-M. Tsai, M.-C. Chen, W.-C. Li, and W.-Y. Chiu: A fast regularity measure for surface defect detection. *Machine Vision and Applications*. pp. 869-886 (2012)
- [15] K. Simonyan and A. Zisserman: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. (2014)
- [16] Y. Sun, Y. Chen, X. Wang, and X. Tang: Deep learning face representation by joint identification-verification. Vol. 2, pp. 1988-1996. *Advances in Neural Information Processing Systems* (2014)

- [17] S. Chopra, R. Hadsell, and Y. LeCun: Learning a similarity metric discriminatively, with application to face verification. Vol. 1, pp. 539-546. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, US (2005)
- [18] S. Huang, Y. Jiang, Y. Gao: Automatic Modulation Classification Using Contrastive Fully Convolutional Network. IEEE Wireless Communications Letters. pp. 1044-1047 (2019)
- [19] R. O. Duda and P. E. Hart: Use of the hough transformation to detect lines and curves in pictures. Commun. ACM. pp. 11-15 (1972)