

Received January 27, 2022, accepted February 17, 2022, date of publication February 21, 2022, date of current version March 3, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3153116

# Alternative Collaborative Learning for Character Recognition in Low-Resolution Images

SUNGJIN LEE, JUN SEOK YUN, AND SEOK BONG YOO<sup>1</sup>, (Member, IEEE)

Department of Artificial Intelligence Convergence, Chonnam National University, Gwangju 61186, South Korea

Corresponding author: Seok Bong Yoo (sbyoo@jnu.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2020R1G1A1100798) and vehicle AI Convergence Research & Development Program through the National IT Industry Promotion Agency of Korea (NIPA) funded by the Ministry of Science and ICT (No. S0315-21-1001). This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-02068, Artificial Intelligence Innovation Hub).

**ABSTRACT** Character recognition in a single image is a technology utilized in various sensor platforms, such as smart parking and text-to-speech systems, and numerous studies are being conducted to improve its performance by experimenting with novel approaches. However, when low-quality images were inputted to a character recognition neural network for recognition, a difference in the resolution of the training image and low-quality image results in poor accuracy. To resolve this problem, this study proposes a collaborative trainable mechanism that integrates a global image feature extraction-based super-resolution neural network with a character recognition neural network. This collaborative trainable mechanism helps the character recognizer to be robust to inputs with varying quality in the real world. The alternative collaborative learning and character recognition performance test was conducted using the license plate image dataset among various character images, and the effectiveness of the proposed algorithm was verified using a performance test.

**INDEX TERMS** Collaborative learning, image super-resolution, character recognition, global image feature extraction.

## I. INTRODUCTION

Character recognition is a technique in computer vision that recognizes and classifies the regions of letters and digits to be identified inside a single image. This is a rather complicated process as it requires both detection and recognition of a particular character from a single image. A character recognition neural network detects a character area in an input image, detects the area, and adjusts its size according to the input. In this process, the quality of the output data deteriorates owing to the loss of resolution of the character area while cropping. There is a reduction in the recognition rate when low-quality legacy content is given as an input to a general character recognition neural network owing to the disparity in the resolution of the training data and the input image.

To solve this problem, researchers have used the super-resolution (SR) method to improve the quality of legacy content input to the character recognition neural network after

individual training of SR and character recognition neural networks. Our study focuses on improving the recognition rate of character recognition neural networks. There is a lack of practical research on improving the quality of low-quality legacy content and optimizing character recognition simultaneously through collaborative learning of the SR and character recognition neural networks. Accordingly, in this study, we designed a collaborative learnable neural network integrating SR and character recognition neural networks so that the character recognition performance is robust in data of various qualities. An alternative collaborative learning algorithm is utilized to effectively train the proposed neural network. The proposed alternative collaborative learning algorithm can be used to dramatically improve the recognition rate in various existing character recognition fields.

The contributions of this study can be summarized as follows:

- An alternative collaborative learning model for SR and character recognition neural networks was designed, and

The associate editor coordinating the review of this manuscript and approving it for publication was Yi Zhang<sup>1</sup>.

a training algorithm was proposed. We designed the collaborative learning neural network in two phases based on weight freezing.

- The proposed training algorithm scales the local patch-based feature extraction to a global scale. In terms of character recognition, using a global image including all numbers can utilize more information than using a patch with the number area cut off.
- Proposed framework guarantees stable recognition performance of collaborative learnable neural network through performance optimization based on alternative collaborative learning.

## II. RELATED WORKS

### A. SUPER-RESOLUTION (SR)

When character recognition is performed with a low-quality legacy content image, the input is enlarged with bicubic interpolation and the resolution decreases. The decrease in the recognition rate can be attributed to the reduced resolution of the image. Therefore, studies have to be conducted to convert low-quality legacy content images to high-resolution (HR) images to accelerate character recognition rates. To resolve this problem, a SR neural network was utilized to convert low-quality legacy content images to the same resolution as the images used for training the character recognition neural network. To analyze the performance of this particular framework, the peak signal-to-noise ratio (PSNR), which evaluates the difference between the original image and the image passed through the SR neural network, and structural similarity [1], which measures the similarity of structural information, are used. In addition, to solve various problems that exist in the real world using SR, various algorithms, such as weight reduction of the SR neural network and an algorithm of performing robust SR, regardless of training data, have been studied. JDNDMSR [2] determined the optimal weight by separately training the SR model, mosaic removal model, and noise removal model to perform each task when given the task of performing a noise, mosaic removal, and SR on one image. It proposed an algorithm to improve the image restoration performance of a model by performing collaborative learning that connects the three models into one model. DASR [3] proposed a model that shows a high-efficiency SR performance even in real-world applications. Existing SR neural networks assume bicubic interpolation as a blur kernel to train SR neural networks. Therefore, when creating low-quality legacy content, high-resolution images are reduced through bicubic interpolation. However, in a practical scenario, there are various blur kernels that create low-quality legacy content, which is the direct cause of poor performance. To solve this problem, the model was designed with five residual groups that converted low-quality legacy content with various blur kernels, and robustly trained on various blur kernels. Through this procedure, the proposed SR neural network shows outstanding performance even for low-quality legacy content practical contexts. SMSR [4] proposed a technique for skipping unnecessary operations by

training the spatial domain and the channel domain as a mask, which showed high energy efficiency. An adaptive target generator (ATG) [5] argued that when the existing SR neural network is trained by setting low-quality legacy content images as input data and HR as the correct output, the SR neural network aims to predict the neural network training process. but they proposed that this method limits SR neural network's potential. To solve this problem, the concept of an adaptive target generator that creates the best HR for SR neural networks was proposed and implemented. Through this algorithm, a neural network with a higher performance than the existing neural network in the blind SR task was proposed. SwinIR [6] proposed the SR baseline model for image restoration by using a swin transformer which performs good performance on high-level vision tasks. In specific, residual swin transformer blocks (RSTB) which is composed of several swin transformer layers with a residual connection. By this algorithm, SwinIR outperforms the state-of-the-art algorithm. DRN [7] improved the restoration ability by proposing the structure of a closed circuit inside the SR network and adding a dual regression loss function to the existing loss. MZSR [8] is the self-supervision method for restoring input images that were blurred under diverse blur conditions by training on various kernels. The difference between the referred SR methods of the related works and our proposed algorithm is that referred SR methods focused only on the reconstruction of the image and SR process with no character recognition point of view result in poor recognition performance. However, our proposed method can reconstruct low-quality legacy content to be more well-recognized than the previous SR method.

### B. CHARACTER RECOGNITION

The existing algorithms that are utilized for character recognition technology are as follows: YOLOv4 [9] improved performance by adding the following concepts to YOLOv3 [10] to implement an object recognition neural network with higher performance than YOLOv3 [10]. Initially, in YOLOv4 [9], the size of the neural network input image is fixed because of the fully connected layer, and the feature map from the last layer of the convolutional neural network is divided, averaged, and fixed to prevent distortion of the learned image. Spatial pyramid pooling (SPP) is a structure and an algorithm of shortening the information transfer path between the neural network layer in the front part that extracts low-level features and the neural network layer in the back part that extracts high-level features. The feature signals of low-level features are from the path aggregation network (PANet), a structure that reduces information loss of the entire feature map, and darknet [11], a neural network framework capable of training deep neural networks. This design is applied to the proposed model, which has a higher performance than the existing YOLOv3 [10]. An object recognition neural network with YOLOv5 [12] is a backbone model that extracts feature maps from images. In the existing DenseNet [13] structure, the

basic convolutional layer is divided and scheduled to solve the problem of training with gradient information reused in the forward and backpropagation processes. CSPNet [14] is used to achieve memory efficiency by reducing the amount of computation while maintaining accuracy by performing a convolution operation on only a certain part of the split part. By using bottleneck CSP, which applies the bottleneck structure that compresses the channel by performing the performance, we proposed a model with lower capacity and faster speed while maintaining similar performance to the existing YOLOv4 [9]. DetectoRs [15] proposed the following algorithm to prove that the design of the backbone model of the object recognition neural network is important. In the existing algorithm, the feature pyramid network (FPN), was used as a backbone model. The FPN extracts feature in a top-down manner after combining and utilizing each independent feature. DetectoRS [15] proposed a recursive feature pyramid (RFP) as a backbone model to improve performance by implementing an additional feedback network in the FPN structure to achieve a higher performance than the existing algorithm. Using this algorithm, DetectoRs [15] achieved state-of-the-art results on the coco test-dev data. TesseractOCR [16] engine was proposed to handle white-on-black text from a single image. It is one of the most precise available open-source OCR engines for character recognition. OpenALPR [17] is the recognition tool that is based on OpenCV and TesseractOCR to automatically recognize license plate characters. The difference between the referred character recognition methods and our proposed algorithm is that referred character recognition methods performance is decreased when the low legacy-content image is inputted because of the difference between image-quality of training image and test image. However, our proposed collaborative learning method is robust when the low legacy-content image is inputted, which is different in image quality.

### C. COMBINATION OF SR WITH OTHER TASK-DRIVEN NEURAL NETWORKS

Recent studies are conducted to improve the performance of character recognition. These studies can be divided into non-collaborative learning methods and collaborative learning methods.

In the non-collaborative learning methods, Lee *et al.* [18] proved that the SR technique can improve the character recognizer performance with low-quality legacy content images. And the SR technique was applied based on sequential data fusion and license plate detection using structural pattern features to improve the performance of the license plate recognition model in low-quality legacy content input images [19]. In addition, super-resolved recognition for low-quality legacy content was proposed with the data augmentation algorithm for left-right reversal [20]. This algorithm proposed a loss function that can increase data without the addition of data capacity by using the flip algorithm. Vasek *et al.* [21] proposed the convolutional neural network (CNN) method for

recognizing license plates in very LR videos. Lee *et al.* [22] proposed the SR method based on generative adversarial networks that can be applied in the license plate recognition challenged environments. However, these methods are not collaborative learning for character recognition, and they just connected parallel the SR network and character recognition. These methods do not contain any recognition term in the SR network, which is non-collaborative learning.

Unlike the above studies, there were various approaches to improve character recognition performance using collaborative learning methods. A transformer-based SR network (TBSRN) [23] was proposed a converged neural network that intensively restores only the scene text to solve the problem of a reduction in recognition rate that occurs when an image with text has low-quality legacy content. It performs SR so that the scene text recognizer can perform well through a pixel unit supervision module, location recognition module, and content recognition module. A pluggable SR unit (PlugNet) [24] was proposed to recognize scene texts which are in LR images with a novel degradation-aware scene text recognizer for collaborative learning. Hamdi *et al.* [25] proposed double generative adversarial networks for image enhancement. They performed SR training used for constructive license plate denoising and SR to increase the license plate recognition accuracy when LR image was used for recognition. Zhang *et al.* [26] proposed the multi-task generative adversarial network (MTGAN) which integrates the license plate SR and recognition in one collaborative learning framework. This method proposed a fully connected network as the generative adversarial network to integrate the knowledge of data distribution and license plate domain prior knowledge to generate HR license plate images. However, they did not propose the alternative collaborative training with two phases that can guarantee the training stability with rapid convergence and optimize the network parameters so that the loss function approaches the global minimum. To tackle this issue, in this paper, we propose alternative collaborative learning with two phases, Phase 1: SR network training with character recognition network freezing, Phase 2: character recognition network training with SR network freezing. Moreover, we propose a global feature image feature extraction method that can prevent character area cut off in the collaborative training process. It helps to utilize more character information than using a patch-based SR neural network. For this reason, our proposed method presents robust character recognition performance in low-resolution conditions.

In addition, to improve the performance of the other task, other studies utilized the collaborative learning method with SR neural network. IEN [27] performed full training of the object recognition neural network and the SR neural network so that object recognition is performed well even in low-quality legacy content images. SING [28] proposed the collaborative learning of SR neural network with person re-identification neural network. SING model is

constructed by connecting the SR and person re-identification neural networks in series to improve the performance of the re-identification neural network that measures whether persons in an image match in a low-resolution (LR) image. Wang *et al.* [29] proposed a step-by-step neural network evolution for solving very LR image recognition problems and getting a more advanced and superior neural network. Context-aware joint compression artifacts reduction and SR neural network (CAJNN) [30] was proposed the joint learning neural network for integrating both local and non-local features and generating an artifact-free HR image from the LR image, which was compressed with an arbitrary quality factor. Chen *et al.* [31] proposed an identity-aware face SR framework to recover identity-related textures which are beneficial in recovering identity information for face recognition. This method shows the superior performance of face recognition to other previous methods when the LR face image is inputted. Vo *et al.* [32] proposed the pyramid architecture with a SR (PSR) network to solve the problem of low accuracy of facial expression recognition in LR images. This method shows the high recognition accuracy of the facial expression in a single LR image. Nguyen *et al.* [33] proposed the SR architecture to generate the HR iris image from the LR iris image to improve iris recognition performance in the LR image. A semi-coupled locality constrained representation (SLR) [34] was proposed a semi-coupled dictionary learning scheme to promote the discriminative and representative capacity for face recognition and SR at the same time. This method also overcame the negative effects of one-to-many mapping with semi-coupled locality-constrained representation to enhance the consistency between LR and HR local manifold geometries. An enhanced AlexNet with SR and data augmentation (SRDA-AlexNet) [35] was proposed for face recognition in LR images. Xiao and Liu [36] proposed the SR-based traffic prohibitory sign recognition neural network to detect and classify various prohibitory. This method extracts traffic sign proposals on RGB images and filters the negative samples by utilizing the color and shape of prohibitory signs. Grm *et al.* [37] addressed the problem of SR in LR inputs in facial images. They proposed the novel face SR method that incorporates identity priors in the learning procedure by using a cascade of multiple SR models that progressively upscale the LR images using steps of scale factor  $\times 2$ . They addressed poor performance in LR image to recognition neural network. Zhang and Ling [38] proposed a supervised pixel-wise generative adversarial network (SPGAN) that can solve the recognition problem in the LR face image of  $16 \times 16$  patch size images. This method focused on whether each pixel of the generated SR face image is real or fake and achieved higher face recognition accuracy than some state-of-the-art methods. A joint SR and vehicle detection network (Joint-SRVDNET) [39] addressed the problem of aerial vehicle detection in SR images because of the lack of discriminative information. This method achieved better visual quality than the state-of-the-art methods for aerial SR with scale factor  $\times 4$  and improved the accuracy of

aerial vehicle detection. This method proposed the integration of detection and SR methods and not in terms of recognition, which can be useful in the real world. Ji *et al.* [40] proposed generative adversarial network frameworks to realize simultaneous SR convolutional neural network (SRCNN) for vehicle detection and the stunning success of deep CNN techniques. This method applied generative adversarial network frameworks to train SRCNN and vehicle detection networks in the collaborative learning manner. Ivan *et al.* [41] proposed a novel procedure for detecting small-scale objects on the road by using the SR method. This method proved that increasing the resolution of the images can improve object detection performance. Truong *et al.* [42] proposed the novel method of drone landing with deep learning-based SR reconstruction and marker detection on an image captured by the LR visible light camera to resolve the low performance of marker detection when the input image of the drone is the LR image. Unlike the above studies, we focus on character recognition performance improvement by using phase-wise collaborative learning.

### III. PROPOSED METHOD

#### A. OVERALL FRAMEWORK

An overall diagram of the proposed collaborative trainable neural network is shown in Fig. 1. The SR and character



FIGURE 1. An overall diagram of the proposed collaborative trainable neural network.

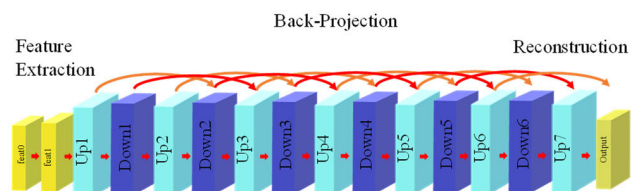


FIGURE 2. Diagram of SR neural network.

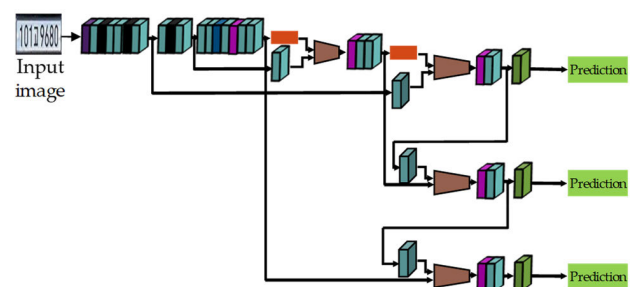


FIGURE 3. Diagram of character recognition neural network.

recognition neural networks are connected in succession. The SR neural network utilizes low-quality legacy content as input and converts it into an SR image, and the character recognition neural network detects characters in the SR image by using both the image and label data with character positions for the image. In this work, the DBPN [43] algorithm is benchmarked, and improved for license plate image SR, as presented in Fig. 2. In addition, the YOLOv5 [12] is benchmarked for the character recognition neural network, as shown in Fig. 3.

As shown in Fig. 2, we extracted the feature map of the low-quality legacy content image as input and passed the feature map through the up-blocks and down-blocks successively to obtain a feature map of 64 channels per block. The neural network concatenates the feature maps which were calculated for each block so that 448 channels of feature maps are delivered to the final output layer, and the SR RGB 3-channel image is obtained. Our goal is to train an SR neural network that performs SR for robust character recognition through these tasks. To perform collaborative learning of the model, we require the pre-trained weights of both neural networks. Therefore, it is necessary to train them before collaborative learning. In addition, the original HR image is down-sampled into low-quality legacy content, and the degraded low-quality legacy content is passed through an SR neural network to obtain an SR feature map. SR neural network through the difference between the gained SR feature map and HR feature map. To train and input the SR feature map derived through the SR neural network into the character recognition neural network in the state that it is a feature map rather than an image. As shown in Fig. 3, we adjusted the size of the SR feature map to the size  $(256 \times 256)$  of the character recognition neural network through bicubic interpolation. By inputting the image of this adjusted size into the character recognition neural network, the input image is passed to the focus layer. As shown in Fig. 4, the focus layer is a fast down-scaling layer with as little information loss as possible. It transforms the input image space into depth for speed. The transformed input feature map moves through the CBL and CSP1\_x layers. The CBL layer is an integrated layer consisting of convolution, batch normalization, and leaky rectified linear units. This layer performs the convolution operation and batch normalization for robust training. The CSP1\_x layer is a cross-stage partial network that achieves richer gradient combinations while maintaining low computation. By splitting the gradient flow, CSP1\_x brings down the computation of the network with the residual unit, which keeps the gradient of the network. The feature map was delivered through the spatial pyramid pooling (SPP) layer to bring about a fixed one-dimensional array for input to the fully connected layer. The upsample module performs upsampling of the feature map, which doubles the number of arrays in a feature map. Owing to the expansion of the feature map, the upsample module allows the detection of small objects.

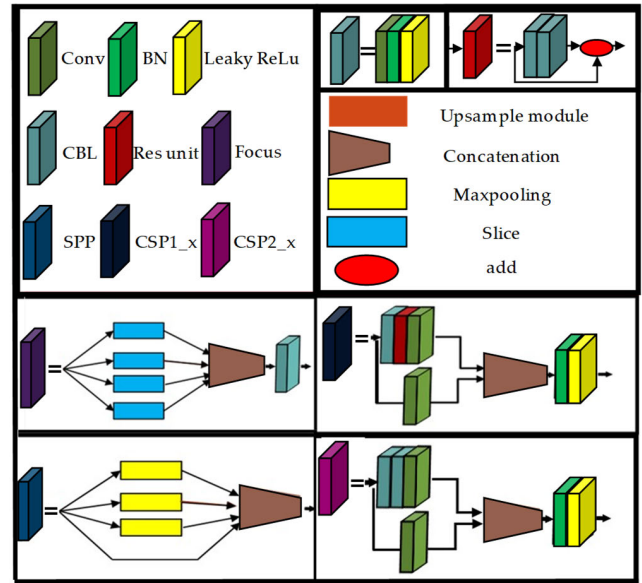


FIGURE 4. Detailed block of a character recognition neural network.

## B. COLLABORATIVE LEARNING FOR SR AND CHARACTER RECOGNITION

The collaborative learning flowchart for SR and the character recognition process is shown in Fig. 5. The first prediction is concatenated with the twice upsampled feature map and the first backbone layer's feature map. This prediction detects small objects in the input images. The second prediction is concatenated with a single upsampled feature map and the second backbone layer's feature map. This detects medium-sized objects in the input image. The third prediction is concatenated with a non-upsampled feature map, a twice upsampled feature map, and a twice convoluted feature map. This detects large objects in the input images. By obtaining these three sizes of character recognition feature maps, the character recognition neural network performs detection and recognition in a single image. The loss of the character recognition (localization, classification, and confidence losses) was obtained by comparing the output of the character recognition with the ground-truth feature map. The character recognition and SR losses are summed and back-propagated to the SR neural network. In this process, if the parameters of the character recognition neural network are changed, the correct collaborative learning cannot be performed. So, the character recognition neural network training was stopped. However, the loss can still be obtained. Through this, the SR neural network is trained to reduce both the SR and character recognition losses, and training is performed to obtain an SR weight with low losses and accurate character recognition.

As shown in Algorithm 1, the pseudocode capable of the integrated network inference was implemented. The inference time of the proposed collaborative algorithm was 3.2 ms

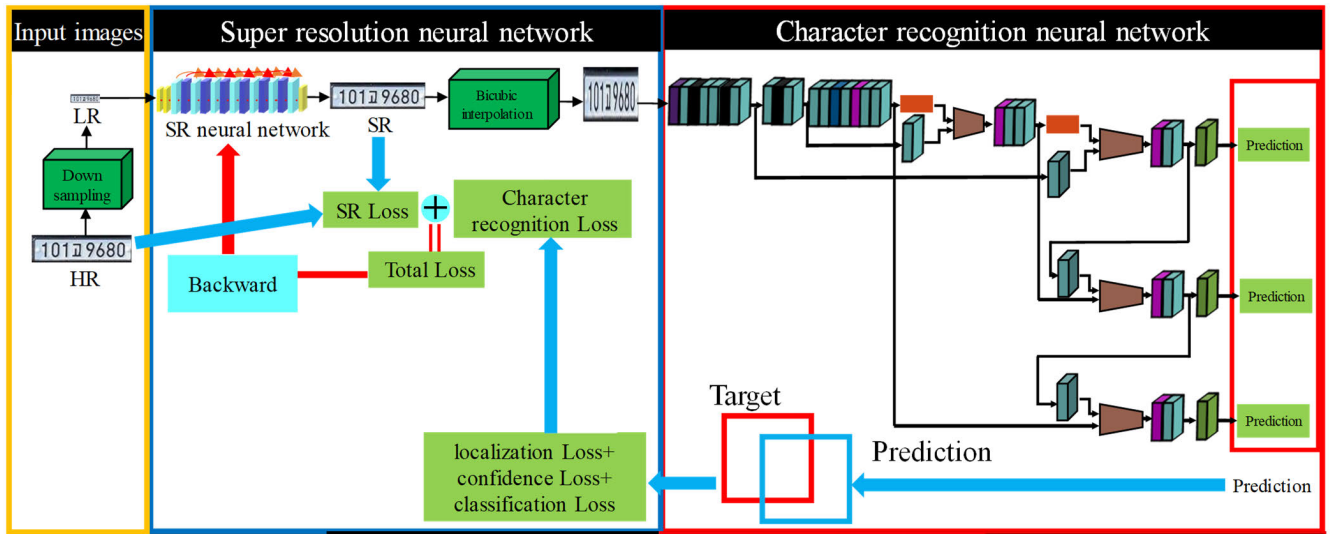


FIGURE 5. Flowchart of integrated network of SR neural network and character recognition neural network for collaborative learning.

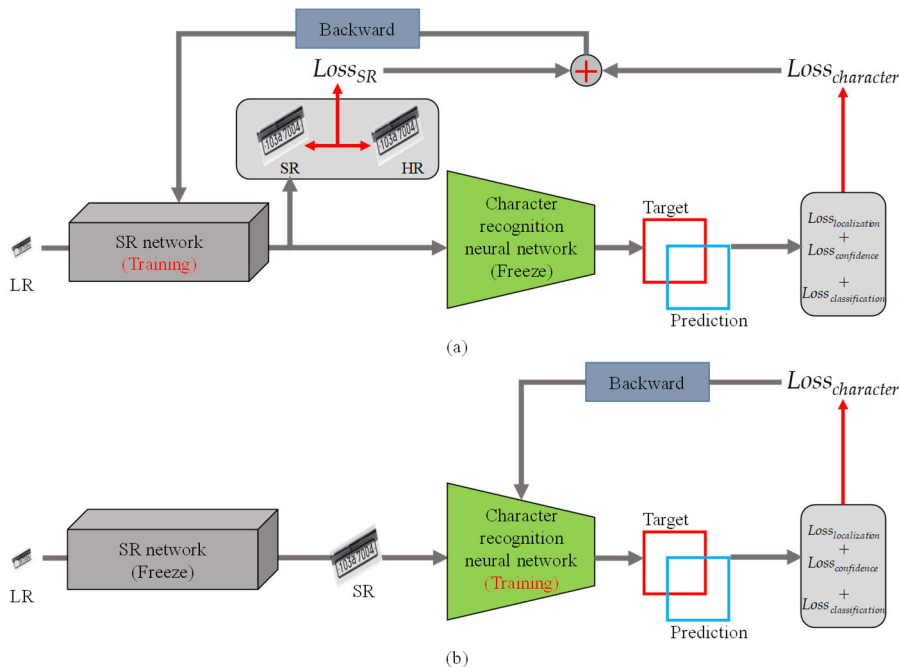


FIGURE 6. Flowchart of alternative collaborative learning in the training process. (a) Phase 1: SR neural network training while freezing character recognition neural network. (b) Phase 2: Character recognition neural network training while freezing SR neural network.

when a ground-truth HR image with a size of  $144 \times 96$  pixels as input. The input image was  $36 \times 24$  pixels in size scaled by a scale factor  $\times 4$ , and the character recognition neural network’s input size was  $256 \times 256$  pixels.

**C. ALTERNATIVE COLLABORATIVE TRAINING THORUGH WEIGHT FREEZING**

The SR neural network and character recognition networks are independently pre-trained before collaborative learning.

Then alternative collaborative learning is started using the pre-trained weight of the SR neural network and character recognition neural network. Alternative collaborative training through weight freezing was performed with phases 1 and 2. Phase 1 denotes the SR neural network training while freezing character recognition neural network, and Phase 2 denotes the character recognition neural network training while freezing SR neural network. To implement phases 1 and 2, the following process was performed. The training data used by

**Algorithm 1** The Pseudocode of the Proposed Integrated Network Inference

1: **Input:** low-quality legacy content image ( $\mathbf{L}$ ), number of iterations in the SR network ( $I$ )

2: **Output:** SR image ( $\mathbf{S}$ )

**do:**

3: set  $SR$  = SR neural network of proposed collaborative learning model

4: set  $CH$  = character recognition network of proposed collaborative learning model

5: set  $U$  = upsampling block in SR

6: set  $D$  = downsampling block in SR

**Step 1:** SR neural network

7: **For**  $k = 1, I$  **do**

8:  $\mathbf{S} = U(\mathbf{L})$

9:  $\mathbf{S} = D(\mathbf{S})$

10: **end for**

11:  $\mathbf{S} = U(\mathbf{S})$

12: **return**  $\mathbf{S}$

**Step 2:** Character recognition neural network

13:  $\mathbf{S} = \text{Bicubic\_interpolation}(\mathbf{S}, 256)$

14:  $\mathbf{S} = \text{Focus}(\mathbf{S})$

15:  $\mathbf{S} = \text{CSP1\_1}(\text{CBL}(\mathbf{S}))$

16:  $\mathbf{S} = \text{CSP1\_2}(\text{CBL}(\mathbf{S}))$

17:  $\mathbf{S\_1} = \text{CBL}(\mathbf{S})$

18:  $\mathbf{S\_2} = \text{CBL}(\text{CSP1\_1}(\text{CBL}(\mathbf{S\_1})))$

19:  $\mathbf{S\_3} = \text{CBL}(\text{CSP2\_1}(\text{CBL}(\text{SPP}(\text{CBL}(\text{CBL}(\mathbf{S\_2}))))))$

20:  $\text{concat\_1} = \text{concatenation}(\text{CBL}(\mathbf{S\_2}), \text{Upsample}(\mathbf{S\_3}))$

21:  $\text{concat\_2} = \text{concatenation}(\text{CBL}(\mathbf{S\_1}), \text{Upsample}(\text{CBL}(\text{CSP2\_2}(\text{concat\_1}))))$

22:  $\text{prediction\_1} = \text{Conv}(\text{CBL}(\text{CSP2\_3}(\text{concat\_2})))$

23:  $\text{concat\_3} = \text{concatenation}(\text{CBL}(\text{CSP2\_2}(\text{concat\_1})), \text{Conv}(\text{CBL}(\text{CSP2\_3}(\text{concat\_2}))))$

24:  $\text{prediction\_2} = \text{Conv}(\text{CBL}(\text{CSP2\_4}(\text{concat\_3})))$

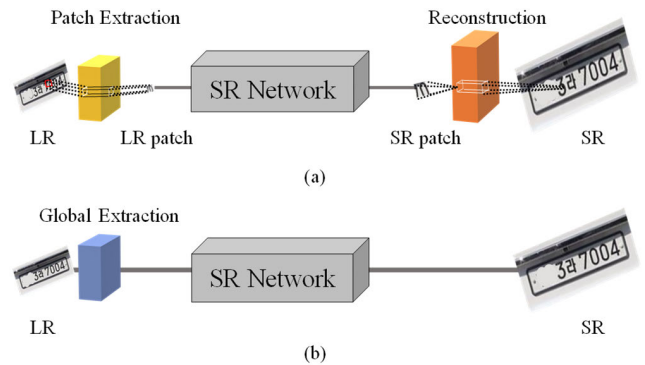
25:  $\text{concat\_4} = \text{concatenation}(\mathbf{S\_3}, \text{CBL}(\text{CSP2\_4}(\text{concat\_3})))$

26:  $\text{prediction\_3} = \text{Conv}(\text{CBL}(\text{CSP2\_5}(\text{concat\_4})))$

27: **Output** =  $\max\_confidence(\text{prediction\_1}, \text{prediction\_2}, \text{prediction\_3})$

the character recognition neural network is converted into LR images and passed through the pre-trained SR neural network to obtain SR images. These SR images are then used to further train the character recognition neural network. Through this design, the character recognition neural network trained with SR images performed better than the existing models. As shown in Algorithm 2, the pseudocode capable of alternative collaborative learning through weight freezing was implemented.

The character recognition neural network, which has been robustly trained in the existing HR domain, failed to correctly recognize low-quality legacy content images and SR images whose quality was improved through the SR neural network.



**FIGURE 7.** (a) Diagram of patch extraction-based SR. (b) Diagram of global image feature extraction-based SR.

To solve this problem, we propose an alternative collaborative learning algorithm that alternatively freezes the weights of the SR neural network and trains the character recognition neural network, as shown in Fig. 6. Fig. 6(a) shows the first phase for training the SR neural network by fixing the weights of the character recognition neural network, and Fig. 6(b) shows the second phase for training the character recognition neural network by fixing the weights of the SR neural network.

**D. SR ALGORITHM BASED ON GLOBAL IMAGE FEATURE EXTRACTION**

The existing SR algorithm uses batch normalization to expedite training. For batch normalization, the input global images should be connected and passed to the SR neural network. However, if the dimensions of the global images differ from each other, this procedure is not possible. To solve this problem, existing research [17] randomly extracts patches with local information from one global image, connects the extracted patches with the same width and height in batches, and inputs them into the SR neural network for batch normalization. This algorithm increases the training speed of the SR neural network. However, while testing the SR neural network trained with a patch unit of an actual image, a grid is generated from the image, and the generated grid degrades the performance of image restoration. Moreover, it is impossible to train the character recognition neural network with the patch image in the collaborative learning process. To solve this problem, training by inputting global images rather than image patches to the SR neural network, it was found that the SR algorithm based on global image feature extraction does not cause deterioration in image restoration ability caused by grids, resulting in improved performance. As shown in Fig. 7(a), the patch-extraction method cut images into the patch and put them into the SR neural network. As shown in Fig. 7(b), we propose a collaborative learning algorithm in which the entire image extracted with global extraction is input to the SR neural network to train.

### E. LOSS FUNCTION FOR THE COLLABORATIVE LEARNING ALGORITHM

The proposed collaborative trainable neural network calculates the loss of the SR and character recognition neural networks as follows:

$$Loss_{Total} = Loss_{SR} + Loss_{character} \times a. \quad (1)$$

The calculated losses are summed up and backpropagated to the SR neural network to implement a collaborative learning neural network, and the procedure for this is as follows. The loss of the SR neural network is obtained by separately obtaining the loss for 11,428 samples, whereas the loss of the character recognition neural network is obtained as the sum of the losses after obtaining all the character recognition losses of all images. Because of the difference of loss scale, the loss calculated from the character recognition neural network was divided into 11,428, and training was performed by multiplying the weight by  $\alpha$ . Equation (1) is the total training loss function, and the expansions are presented in (2) and (3).

### F. LOSS FUNCTION ANALYSIS

In this work, we analyze the loss function for the proposed model. The loss function of the SR neural network is defined as:

$$Loss_{SR} = \sum_{i=0}^N |HR_i - f(LR_i)| \quad (2)$$

where  $N$  is the number of training images,  $x_i$  is the  $i$ -th LR training image, is the SR result of  $x_i$ , and  $y_i$  is the  $i$ -th HR training image corresponding to the SR image  $f(x_i)$ . The loss function for character recognition consists of localization, confidence, and classification losses. The loss function is defined as follows:

$$\begin{aligned} & Loss_{character} \\ &= \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\ &+ \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + ((\sqrt{h} - \sqrt{\hat{h}})^2] \\ &+ \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} (C_i - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{noobj} (C_i - \hat{C}_i)^2 \\ &+ \sum_{i=0}^{S^2} 1_i^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2 \end{aligned} \quad (3)$$

where  $S^2$  is the grid cell for a particular character recognition. It takes a value of one if a character is recognized, and zero otherwise. The first and second terms denote the localization loss that calculates the error of the position of the bounding box, which estimates the location of the character in the

### Algorithm 2 The Pseudocode of the Alternative Collaborative Learning Through Weight Freezing

1: **Input:** low-quality legacy content image ( $L$ ), HR image ( $H$ ), number of training images ( $N$ ), label of character recognition ( $C$ )  
2: **Output:** SR weight ( $W_{SR}$ ), character recognition weight ( $W_{CH}$ )

do:

3: set  $SR$  = SR neural network of proposed collaborative learning model

4: set  $CH$  = character recognition neural network of proposed collaborative learning model

5: Freeze  $CH$  and Train  $SR$ .

(Phase 1) Step 1: SR loss calculation

6: For  $p=1, N$  do

7:  $S = SR(L)$

8:  $Loss_{SR} = |H - S|_1$  as in (2)

9: end for

(Phase 1) Step 2: character recognition loss calculation

10: For  $p=1, N$  do

11:  $S = Bicubic\_interpolation(S, 256)$

12:  $S = CH(S)$

13:  $Loss_{character} = Loss(S, C)$  as in (3)

14: end for

(Phase 1) Step 3: total loss calculation

15: For  $p=1, N$  do

16:  $Loss_{total} = Loss_{SR} + Loss_{character} \times a$  as in (1)

17: end for

(Phase 1) Step 4: SR weight update via total loss backpropagation

18: For  $p=1, N$  do

19:  $grad_{SR} = Backpropagate(SR, Loss_{total})$

20:  $W_{SR} \leftarrow W_{SR} - grad_{SR}$

21: return  $W_{SR}$

22: end for

23: Freeze  $SR$  and Train  $CH$ .

(Phase 2) Step 1: SR loss calculation

24: For  $p=1, N$  do

25:  $S = SR(L)$

26:  $Loss_{SR} = |H - S|_1$  as in (2)

27: end for

(Phase 2) Step 2: character recognition loss calculation

28: For  $p=1, N$  do

29:  $S = Bicubic\_interpolation(S, 256)$

30:  $S = CH(S)$

31:  $Loss_{character} = Loss(S, C)$  as in (3)

32: end for

(Phase 2) Step 3: CH weight update via character loss backpropagation

33: For  $p=1, N$  do

34:  $grad_{CH} = Backpropagate(CH, Loss_{character})$

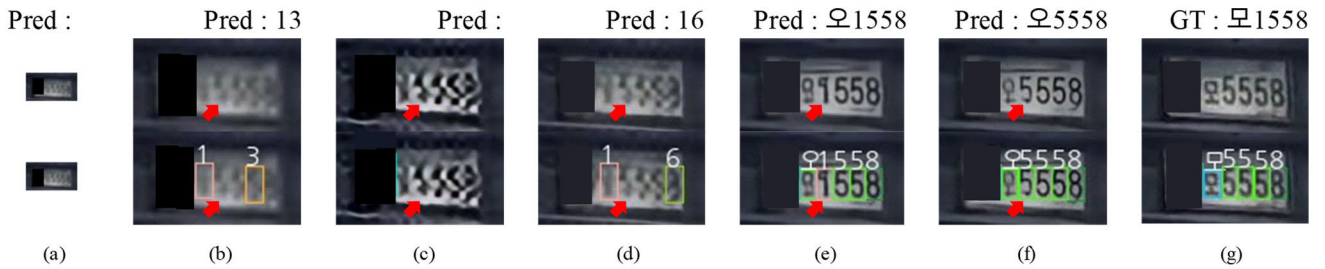
35:  $W_{CH} \leftarrow W_{CH} - grad_{CH}$

36: return  $W_{CH}$

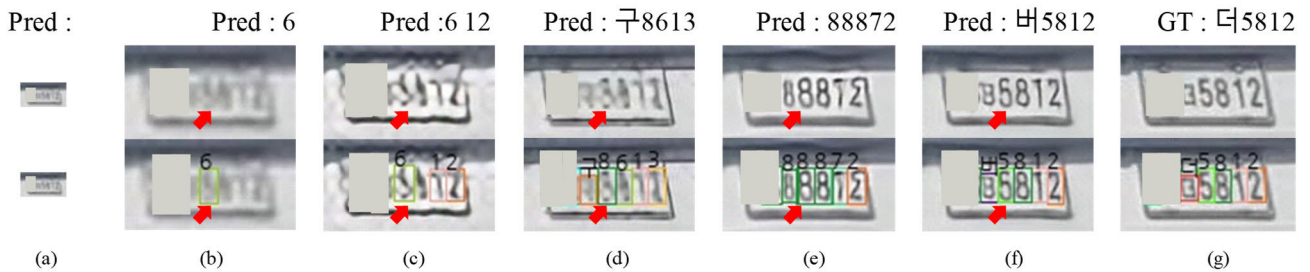
37: end for

image while performing character recognition.  $x_i$  denotes the  $i$ -th image's horizontal coordinate,  $y_i$  denotes the  $i$ -th image's vertical coordinate,  $w_i$  denotes the  $i$ -th image's width, and  $h_i$  denotes the  $i$ -th image's height.  $C_i$  denotes the  $i$ -th image's confidence of the class in the grid box, and  $p_i(c)$  denotes the  $i$ -th image's detected class's confidence. The character recognition neural network calculates the sum of squared errors for the  $x_i$ ,  $y_i$ ,  $w_i$ , and  $h_i$  of the predicted box. At this

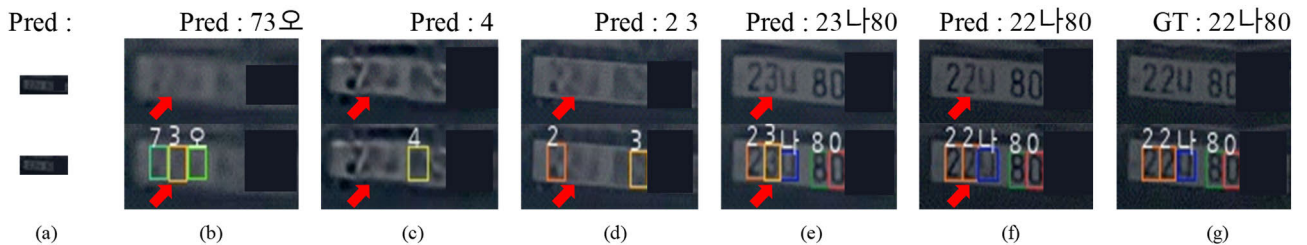




**FIGURE 8.** (a) LR input image ( $27 \times 15$ ) and the character recognition result of the LR input image. SR ( $\times 4$ ) and character recognition results of plate images on (b) bicubic image ( $108 \times 60$ ), (c) MZSR result, (d) DRN result, (e) Super-resolved recognition result, (f) collaborative learning result (ours), and (g) HR image. “GT” represents ground-truth and “Pred” represents the prediction result. Two characters were masked due to privacy policy.



**FIGURE 9.** (a) LR input image ( $24 \times 13$ ) and the character recognition result of the LR input image. SR ( $\times 4$ ) and character recognition results of plate images on (b) bicubic image ( $96 \times 52$ ), (c) MZSR result, (d) DRN result, (e) Super-resolved recognition result, (f) collaborative learning result (ours), and (g) HR image. “GT” represents ground-truth and “Pred” represents the prediction result. Two characters were masked due to privacy policy.



**FIGURE 10.** (a) LR input image ( $25 \times 10$ ) and the character recognition result of the LR input image. SR ( $\times 4$ ) and character recognition results of plate images on (b) bicubic image ( $100 \times 40$ ), (c) MZSR result, (d) DRN result, (e) Super-resolved recognition result, (f) collaborative learning result (ours), and (g) HR image. “GT” represents ground-truth and “Pred” represents the prediction result. Two characters were masked due to privacy policy.

time, because boxes with large widths and heights may have larger error values than smaller boxes, they are less affected by performing the square root operation for width and height. In (3),  $C_i$  denotes the confidence loss for the degree of certainty that the recognized character is a character while performing character recognition. This confidence value is a probability value between zero and one that is determined when a character is detected in the box. When no character was detected,  $\lambda_{noobj}$  was used.  $p_i(c)$  denotes a classification loss for classifying recognized characters. This loss function calculates the loss through the sum of squared errors of the probability that the recognized character is a specific class when a character is recognized. By adding all three loss functions, the character recognition neural network was trained.

#### IV. RESULTS

##### A. RESULTS OF COLLABORATIVE LEARNING (PAHSE 1)

In addition, 122 classes were defined with reference to the contents of a previous study [31] for character recognition neural network training. For effective performance of the collaborative learning, first, the SR and the character recognition neural networks were separately trained with the same 11,428 training set, and their performance was verified with the 1,999 samples in the validation set. The collaborative learning experiment was conducted as follows. Phase 1 of the collaborative learning SR neural network was performed in scale factor  $\times 3$  and scale factor  $\times 4$ , which was verified by measuring the PSNR and mAP of the validation set. We set the LR image-based character recognition network trained by LR images as the baseline. It is denoted as “LR recognition”.

In other SR-based recognition algorithms, the same character recognition network as our proposed method was commonly applied, as shown in Fig. 3. Since these algorithms are not collaborative learning-based methods but independent learning-based methods. Therefore, SR-based recognition algorithms were trained with HR images.

As shown in Table 1, the PSNR was increased by 4.03 dB, and the mean average precision (mAP) was increased by 3.8% points than the bicubic interpolation-based reconstruction. The PSNR was increased by 0.06 dB, and the mAP was increased by 2.9% points than the super-resolved recognition for scale factor  $\times 3$ . In Table 2, the PSNR was increased by 3.52 dB, and the mAP was increased by 9.7% points than the bicubic interpolation-based reconstruction. The PSNR was increased by 0.13 dB, and the mAP was increased 5.9% points than the super-resolved recognition for the scale factor  $\times 4$ . These results indicate that the proposed algorithm provides better reconstruction performance in terms of PSNR as well as gives higher recognition accuracy in terms of mAP compared to the existing SR-based reconstruction methods.

The result of the SR neural network was verified, indicating that the character recognition accuracy of the image restored through the proposed algorithm increased compared to that of the image restored by existing SR algorithms. Figs. 8–10 show the improvement of character recognition of collaborative learning compared with other SR-based algorithms [6–8, 19]. In Figs. 8–10, red arrows denote the region where the difference is highlighted. Fig. 8(e) shows that the super-resolved recognition [19] incorrectly recognizes the second character (“5”) as the character (“1”). While our proposed method’s result shows the correct recognition of the second character (“5”) as shown in Fig. 8(f). In Fig. 9, we can find that the super-resolved recognition result shows the incorrect recognition of the second character (“5”) as the character (“8”). On the contrary, our proposed method’s result shows that our method can recognize the second character (“5”) correctly. In Fig. 10, we can find that the super-resolved recognition result shows the wrong recognition of the second character (“2”) as the character (“3”). While our proposed method’s result shows that our method can recognize the second character (“2”) exactly. In addition, we note that the other SR-based recognitions do not provide satisfactory results compared with the proposed method as shown in Figs. 8–10.

## B. RESULTS OF COLLABORATIVE LEARNING (PHASE 2)

The algorithm in phase 2 is used for training the character recognition and SR neural network. First, the weights of the SR neural network were fixed, the SR image was obtained, and the character recognition neural network was trained using the SR image. This algorithm trains a character recognition neural network robustly on SR images, and the results for different scale factors ( $\times 3$ ,  $\times 4$ ) are presented in Tables 3 and 4. The collaborative learning results of the character recognition neural network show a higher performance for character recognition than the low-quality legacy content

**TABLE 1. Comparison between bicubic images ( $\times 3$ ), SR, collaborative learning, and HR images.**

PSNR & mAP on 1,999 license plate validation set		
Method	PSNR (dB)	mAP (%)
LR recognition	-	62.6
Bicubic	26.0	78.0
MZSR [8]	19.2	46.3
DRN [7]	-	-
SwinIR [6]	29.2	81.0
Super-resolved recognition [19]	30.0	78.9
Collaborative (ours)	30.1	81.8
HR image	-	94.2

**TABLE 2. Comparison between bicubic images ( $\times 4$ ), SR, collaborative learning, and HR images.**

PSNR & mAP on 1,999 license plate validation set		
Method	PSNR (dB)	mAP (%)
LR recognition	-	41.0
Bicubic	23.6	53.1
MZSR [8]	19.5	46.0
DRN [7]	25.8	56.6
SwinIR [6]	25.5	58.3
Super-resolved recognition [19]	27.0	56.9
Collaborative (ours)	27.2	62.8
HR image	-	94.2

image. To show the performance improvement comes from the collaborative training, we train the character recognition network in Phase 2 with the SR outputs of each method and verify their results. As shown in Tables 3 and 4, our collaborative training method presents that the mAP was increased by 4.5% than SwinIR in Table 3 and was increased by 9.6% than SwinIR in Table 4. Using our previous work [19], we fine-tuned our recognition network with our non-collaborative learning-based SR results. It is denoted as “Phase 2 of super-resolved recognition [19]” in Tables 3 and 4. Our collaborative learning method presents that the mAP was increased by 6.2% than the super-resolved recognition in Table 3 and was increased by 6.6% than super-resolved recognition in Table 4. These results demonstrate that our method is effective in improving recognition accuracy. Figs. 11 and 12 show an improvement in the collaborative learning of the SR neural network for different scale factors ( $\times 3$ ,  $\times 4$ ). We chose 200 typical license plates from the validation sets for the comparison of numeric class (0–9) and validated the performance of the character recognition in order to clearly show the comparison between collaborative learning results and super-resolved recognition results. The character recognition mAP results of a numeric class (0–9) for bicubic images ( $\times 3$ ), super-resolved recognitions, collaborative learning results, and HR images are presented from left to right in Fig. 11. We can find that mAP of collaborative learning results

**TABLE 3. Comparison of mAP between bicubic images ( $\times 3$ ), super-resolved recognition, collaborative learning, and HR images.**

mAP on 1,999 license plate validation set		
Method	Phase 1 mAP (%)	Phase 2 mAP (%)
LR recognition	62.6	-
Bicubic	78.0	82.3
MZSR [8]	46.3	57.5
DRN [7]	-	-
SwinIR [6]	81.0	83.1
Super-resolved recognition [19]	78.9	81.4
Collaborative (ours)	81.8	87.6
HR image	94.2	-

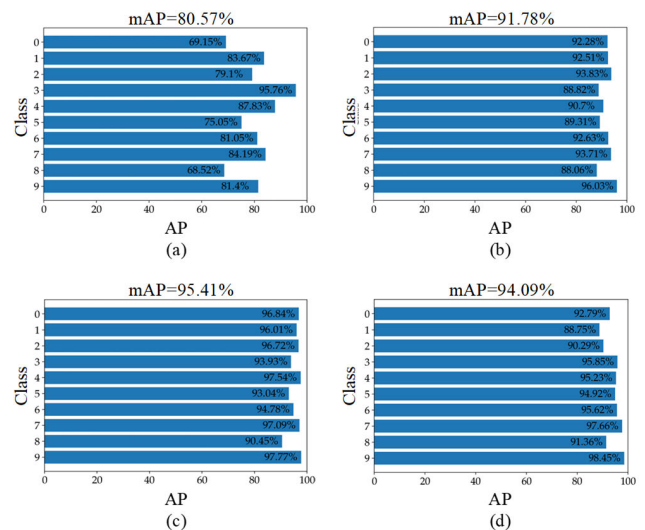
(95.41%) is 3.63% points higher than the mAP calculated for non-collaborative learning (91.78%) and 14.54% points higher than the mAP of bicubic images (80.57%) in Fig. 11. Fig. 12 demonstrates the character recognition mAP results of a numeric class (0–9) for bicubic images ( $\times 4$ ), super-resolved recognitions, collaborative learning results, and HR images, from left to right. we can know that mAP of collaborative learning results (81.10%) is 25.77% points higher than the mAP calculated for the non-collaborative learning (65.33%) and 43.37% points higher than the mAP of bicubic images (37.73%) in Fig. 12. This result shows that collaborative learning has the dominant effect on character recognition. As shown in Tables 3 and 4, these results demonstrate that the fine-tuning methods do not guarantee the best performance of the character recognition neural network because the loss function concerning character recognition was not considered during SR process. On the other hand, our collaborative learning method theoretically predominates in character recognition since our SR network was trained with character recognition loss. This brings character recognition performance improvement compared to the fine-tuning-based method. As shown in Figs. 8(f), 9(f), and 10(f), the proposed method enhances the character sharpness, character thickness, and character contrast, which are closely related character recognition components. For this reason, our proposed method outperforms fine-tuning methods in terms of character recognition. Meanwhile, unlike fine-tuning methods, our method performs one more training process (Phase 1). Therefore, our method approximately requires 1.3 times more training time than fine-tuning methods due to phase-wise training process. However, the inference time of our method is approximately same as the fine-tuning methods.

### C. RESULTS OF CHARACTER RECOGNITION

Table 5 and Fig. 13 demonstrate the performance of the proposed collaborative learning character recognition network. The combination of the SR and character recognition performance was tested with 1,999 LR images. Table 5 shows the comparison between the proposed collaborative learning character recognition network and other character recognition

**TABLE 4. Comparison of mAP between bicubic images ( $\times 4$ ), super-resolved recognition, collaborative learning, and HR images.**

mAP ON 1,999 LICENSE PLATE VALIDATION SET		
Method	Phase 1 mAP (%)	Phase 2 mAP (%)
LR recognition	41.0	-
Bicubic	53.1	61.9
MZSR [8]	46.0	59.8
DRN [7]	56.6	64.7
SwinIR [6]	58.3	65.1
Super-resolved recognition [19]	56.9	68.1
Collaborative (ours)	62.8	74.7
HR image	94.2	-

**FIGURE 11. mAP comparison results of plate images on (a) bicubic images ( $\times 3$ ), (b) super-resolved recognition, (c) collaborative learning results, and (d) HR images, respectively.**

algorithms. Since TesseractOCR [14] does not include the SR algorithm, we combine it with DBPN [43] for the recognition of LR images. The recognition performance of the proposed algorithm provides the highest mAP as shown in Table 5. The visual recognition comparison is also shown in Fig. 13. In Fig. 13, red character means the prediction of the character recognition neural network. OpenALPR [15] provides cloud demo service for a single image and its open-source code is not available. Instead of the quantitative comparison, we demonstrate visual recognition results on a sample image as shown in Fig. 13. While the proposed algorithm will recognize all characters from the license plate, the other algorithms cannot recognize the first character (“9”) or the third character (“ $\Xi$ ”) or the fourth character (“2”) or the fifth character (“4”) of the license plate exactly.

### V. DISCUSSION

In this section, we discuss the strengths and weaknesses of the proposed structure in detail. Our results show that

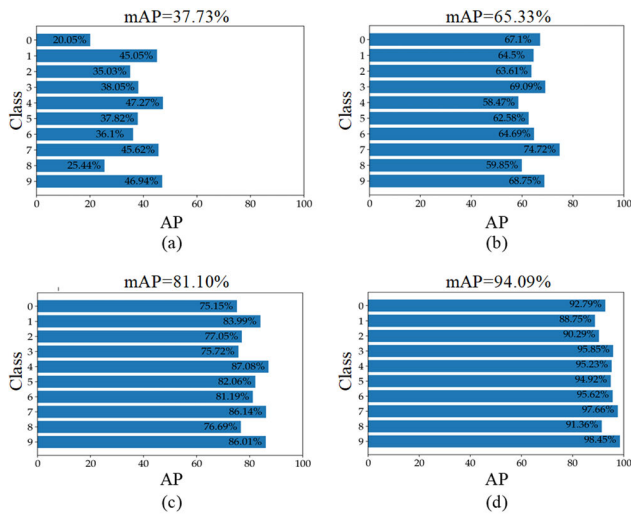


FIGURE 12. mAP comparison results of plate images on (a) bicubic images (x4), (b) super-resolved recognition, (c) collaborative learning results, and (d) HR images, respectively.

the proposed alternative collaborative learning algorithm has a positive impact on character recognition. However, some failure cases were shown for Korean letters by our proposed alternative collaborative learning network algorithm. These are the weakness of collaborative learning network algorithm, and these weaknesses can be overcome through data collection of Korean letter data. Several incorrect results for harsh conditions, such as dark and rotation situations, are shown in Figs. 14 and 15, respectively. The proposed collaborative learning algorithm makes wrong predictions in dark weather or when the license plate is rotated. This is a limitation of the proposed algorithm, and the limitation is expected to be resolved via the application of an additional data augmentation technique.

VI. ABLATION STUDY

A. EFFECTIVENESS OF SR ALGORITHM BASED ON THE GLOBAL IMAGE FEATURE EXTRACTION

In this section, we verify the effectiveness of the SR algorithm based on the global image feature extraction algorithm by comparison with the patch extraction-based algorithm. For a fair comparison, each algorithm was verified with 1,999 LR images. Table 6 shows the comparison between the global image feature extraction SR algorithm. and the patch-based extraction SR algorithm. As demonstrated in Table 6, it can be confirmed quantitatively that the proposed global image feature extraction has a higher reconstruction performance than the patch-based SR algorithm. The global image feature extraction SR algorithm can prevent the grid generation while the patch-based SR algorithm cannot prevent the grid generation, resulting in poor performance of image reconstruction from input LR image. The global image feature algorithm increased PSNR by 1.41 dB for scale factor x4 than patch-based SR algorithm. And the mAP of the global image-based



FIGURE 13. SR (x4) and character recognition results of plate images with (a) DBPN+TesseractOCR, (b) super-resolved recognition, (c) DBPN + OpenALPR, (d) collaborative learning result, and (e) HR image (80 x 60), respectively. "GT" represents ground-truth and "Pred" represents prediction results.

TABLE 5. Comparison of mAP of LR license plates for character recognition validation.

mAP on 1,999 license plate validation set	
Method	mAP (%)
DBPN [43] + TesseractOCR [14]	32.2
LR recognition	41.0
Super-resolved recognition [19]	56.9
Collaborative (ours)	74.7

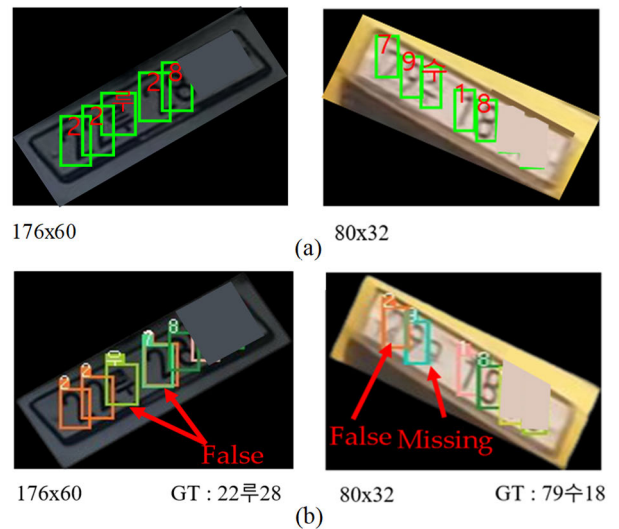
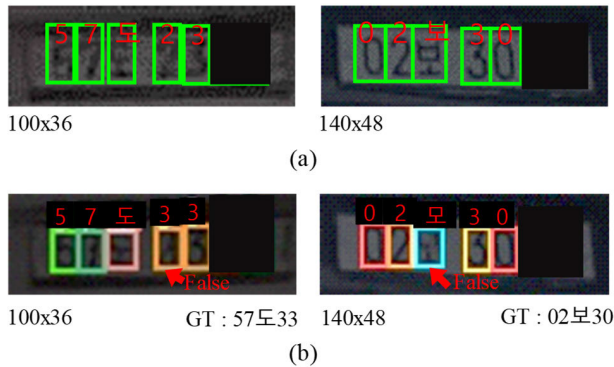


FIGURE 14. Failure examples of character recognition results in rotated condition. (a) Ground-truth labels on HR images and (b) the proposed collaborative learning SR (x4) and character recognition results obtained from LR images. Two characters were masked due to privacy policies.

feature extraction SR algorithm is 10.4% points higher than the mAP of the patch-based SR algorithm.

B. EFFECT OF THE LOSS FUNCTION

The effect according to the value of alpha in the loss function (1) is described in this section. In (1), the value of alpha scales the amount of character recognition loss that



**FIGURE 15.** Failure examples of character recognition results in dark condition. (a) Ground-truth labels on HR images and (b) the proposed collaborative learning SR ( $\times 4$ ) and character recognition results obtained from LR images. Two characters were masked due to privacy policies.

**TABLE 6.** PSNR and mAP on 1,999 license plate validation set.

PSNR on 1,999 license plate validation set		
Method	PSNR (dB)	mAP(%)
Patch-based SR algorithm	27.07	58.6
Global image-based feature extraction SR algorithm	28.48	69.0

**TABLE 7.** Comparison of the loss function of proposed method ( $\times 4$ ).

PSNR and mAP on 1,999 license plate validation set		
$\alpha$	PSNR (dB)	mAP (%)
1	23.2	41.2
0.5	24.7	51.3
0.1	25.4	55.2
0.01	27.2	62.8

backward to SR neural network. Note that the value of  $\alpha$  is the important parameter that determines the performance of the alternative collaborative learning process. If the value is set improperly, it causes the collaborative learning neural network training to be divergence. Table 7 shows the performance comparison according to the loss function by adjusting the value of  $\alpha$ . Among several values of  $\alpha$ , the collaborative learning with of  $\alpha = 0.01$  provides the best performance in terms of both PSNR and mAP. Through this ablation study about loss function, we set the value of  $\alpha$  to 0.01 in our experiment.

### C. COMPARISON OF ALTERNATIVE COLLABORATIVE LEARNING AND NON-ALTERNATIVE COLLABORATIVE LEARNING

To verify the performance of the proposed alternative collaborative learning, we perform the comparison of the

**TABLE 8.** Comparison between non-alternative collaborative learning ( $\times 4$ ) and alternative collaborative learning.

PSNR & mAP on 1,999 license plate validation set		
Method	PSNR (dB)	mAP (%)
Non-alternative collaborative learning	26.1	56.2
Alternative collaborative learning	27.2	74.7

alternative collaborative learning method and the non-alternative collaborative learning method by using 1,999 license plate validation set. Non-alternative collaborative learning denotes training SR neural network and character recognition neural network at once without step-wise weight freezing. The effectiveness of the alternative collaborative learning method can be derived from the results in Table 8. The PSNR and mAP of the proposed alternative collaborative learning method for scale factor  $\times 4$  is 27.2 dB and 62.8% on 1,999 license plate validation set, respectively. On the contrary, the performance of the non-alternative collaborative learning method is inferior to our proposed alternative collaborative learning method. Specifically, the PSNR of proposed alternative collaborative learning was increased by 1.1 dB, and the mAP was increased by 18.5% points compared to the non-alternative collaborative learning method for scale factor  $\times 4$ . This result demonstrates the effectiveness of our alternative collaborative learning method.

### VII. CONCLUSION

In this study, we designed a collaborative learning model that combines a character recognition neural network with an SR neural network with a novel loss function for the SR neural network and an alternative collaborative learning model to improve the recognition performance of the character recognition neural network. Specifically, we proposed the alternative collaborative learning method, which is consist of two phases, Phase 1: SR network training with character recognition network freezing, Phase 2: character recognition network training with SR network freezing. The stable recognition performance of the collaborative learnable neural network was verified by optimizing the recognition performance of the neural network step by step and efficiently training it. To evaluate the recognition performance of the proposed algorithm, we utilized quantitative evaluation scales such as mAP and PSNR, and a collaborative learnable neural network was implemented and tested to independently train the existing SR and character recognition neural networks. Therefore, the performance was higher than when character recognition alone was performed. Moreover, the proposed method outperforms the performance of other SR-based recognition methods in recognizing the characters in very low-resolution images. In addition, the proposed global image extraction technique for training SR neural networks has a higher

restoration ability than the existing local patch-based feature extraction technique. The loss function-based collaborative learning algorithm proposed in this study is expected to be a good reference when performing collaborative learning research in various image processing platforms.

## REFERENCES

- [1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [2] W. Xing and K. Egiazarian, "End-to-end learning for joint image demosaicing, denoising and super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Nashville, TN, USA, Jun. 2021, pp. 3507–3516.
- [3] Y. Wei, S. Gu, Y. Li, R. Timofte, L. Jin, and H. Song, "Unsupervised real-world image super resolution via domain-distance aware training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 13385–13394.
- [4] L. Wang, X. Dong, Y. Wang, X. Ying, Z. Lin, W. An, and Y. Guo, "Exploring sparsity in image super-resolution for efficient inference," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 4917–4926.
- [5] Y. Jo, S. Wug Oh, P. Vajda, and S. Joo Kim, "Tackling the ill-posedness of super-resolution through adaptive target generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 16236–16245.
- [6] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1833–1844.
- [7] Y. Guo, J. Chen, J. Wang, Q. Chen, J. Cao, Z. Deng, Y. Xu, and M. Tan, "Closed-loop matters: Dual regression networks for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 5407–5416.
- [8] J. W. Soh, S. Cho, and N. I. Cho, "Meta-transfer learning for zero-shot super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 3516–3525.
- [9] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [10] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
- [11] *DarkNet: Open Source Neural Networks in C*. Accessed: Oct. 2, 2021. [Online]. Available: <https://pjreddie.com/darknet>
- [12] *YOLOv5*. Accessed: Oct. 2, 2021. [Online]. Available: <https://github.com/ultralytics/yolov5>
- [13] G. Huang, Z. Liu, L. Maaten, and K. Q. Weinberger, "Densely connected convolution networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269.
- [14] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop (CVPR Workshop)*, Seattle, WA, USA, Jun. 2020, pp. 390–391.
- [15] S. Qiao, L.-C. Chen, and A. Yuille, "DetectoRS: Detecting objects with recursive feature pyramid and switchable atrous convolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 10213–10224.
- [16] *Tesseract Open Source OCR Engine*. Accessed: Oct. 25, 2021. [Online]. Available: <https://github.com/tesseract-ocr/tesseract>
- [17] *OpenALPR*. Accessed: Oct. 29, 2021. [Online]. Available: <https://www.openalpr.com/>
- [18] S. J. Lee, T. J. Kim, C. H. Lee, and S. B. Yoo, "Image super-resolution for improving object recognition accuracy," *J. Korea Inst. Inf. Commun. Eng.*, vol. 25, no. 6, pp. 774–784, Jun. 2021.
- [19] T.-Y. Song, Y.-H. Lee, M.-J. Kim, B.-H. Ku, and H.-S. Ko, "Fusion methods of license plate detection and super resolution for improving license plate recognition," *J. Korea Soc. Comput. Inf.*, vol. 16, no. 4, pp. 53–60, Apr. 2011.
- [20] S.-J. Lee and S. B. Yoo, "Super-resolved recognition of license plate characters," *Mathematics*, vol. 9, no. 19, p. 2494, Oct. 2021.
- [21] V. Vasek, V. Franc, and M. Urban, "License plate recognition and super-resolution from low-Resolution videos by convolutional neural networks," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Newcastle, U.K., Sep. 2018, p. 132.
- [22] Y. Lee, J. Yun, Y. Hong, J. Lee, and M. Jeon, "Accurate license plate recognition and super-resolution using a generative adversarial networks on traffic surveillance video," in *Proc. IEEE Int. Conf. Consum. Electron. Asia (ICCE-Asia)*, Las Vegas, NV, USA, Jun. 2018, pp. 1–4.
- [23] J. Chen, B. Li, and X. Xue, "Scene text telescope: Text-focused scene image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12026–12035.
- [24] Y. Mou, L. Tan, H. Yang, J. Chen, L. Liu, R. Yan, and Y. Huang, "PlugNet: Degradation aware scene text recognition supervised by a pluggable super-resolution unit," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, U.K., Aug. 2020, pp. 158–174.
- [25] A. Hamdi, Y. K. Chan, and V. C. Koo, "A new image enhancement and super resolution technique for license plate recognition," *Heliyon*, vol. 7, no.11, p. 8341, Nov. 2021.
- [26] M. Zhang, W. Liu, and H. Ma, "Joint license plate super-resolution and recognition in one multi-task GAN framework," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 1443–1447.
- [27] J. Seo and H. Park, "Object recognition in very low resolution images using deep collaborative learning," *IEEE Access*, vol. 7, pp. 134071–134082, 2019.
- [28] J. Jiao, W. S. Zheng, A. Wu, X. Zhu, and S. Gong, "Deep low-resolution person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, Feb. 2018, pp. 2–7.
- [29] Z. Wang, S. Chang, Y. Yang, D. Liu, and T. S. Huang, "Studying very low resolution recognition using deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1874–1883.
- [30] X. Xiang, Q. Lin, and J. P. Allebach, "Boosting high-level vision with joint compression artifacts reduction and super-resolution," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Milano, Italy, Jan. 2021, pp. 2390–2397.
- [31] J. Chen, J. Chen, Z. Wang, C. Liang, and C.-W. Lin, "Identity-aware face super-resolution for low-resolution face recognition," *IEEE Signal Process. Lett.*, vol. 27, pp. 645–649, 2020.
- [32] T.-H. Vo, G.-S. Lee, H.-J. Yang, and S.-H. Kim, "Pyramid with super resolution for In-the-Wild facial expression recognition," *IEEE Access*, vol. 8, pp. 131988–132001, 2020.
- [33] K. Nguyen, C. Fookes, S. Sridharan, and S. Denman, "Quality-driven super-resolution for less constrained iris recognition at a distance and on the move," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 4, pp. 1248–1258, Dec. 2011.
- [34] T. Lu, X. Chen, Y. Zhang, C. Chen, and Z. Xiong, "SLR: Semi-coupled locality constrained representation for very low resolution face recognition and super resolution," *IEEE Access*, vol. 6, pp. 56269–56281, 2018.
- [35] J. C. Tan, K. M. Lim, and C. P. Lee, "Enhanced AlexNet with super-resolution for low-resolution face recognition," in *Proc. 9th Int. Conf. Inf. Commun. Technol. (ICICT)*, Yogyakarta, Indonesia, Aug. 2021, pp. 302–306.
- [36] D. Xiao and L. Liu, "Super-Resolution-Based traffic prohibitory sign recognition," in *Proc. IEEE 21st Int. Conf. High Perform. Comput. Commun.; IEEE 17th Int. Conf. Smart City; IEEE 5th Int. Conf. Data Sci. Syst. (HPCC/SmartCity/DSS)*, Aug. 2019, pp. 2383–2388.
- [37] K. Grm, W. J. Scheirer, and V. Struc, "Face hallucination using cascaded super-resolution and identity priors," *IEEE Trans. Image Process.*, vol. 29, pp. 2150–2165, 2020.
- [38] M. Zhang and Q. Ling, "Supervised pixel-wise GAN for face super-resolution," *IEEE Trans. Multimedia*, vol. 23, pp. 1938–1950, 2021.
- [39] M. Mostofa, S. N. Ferdous, B. S. Riggan, and N. M. Nasrabadi, "Joint-SRVDNet: Joint super resolution and vehicle detection network," *IEEE Access*, vol. 8, pp. 82306–82319, 2020.
- [40] H. Ji, Z. Gao, T. Mei, and B. Ramesh, "Vehicle detection in remote sensing images leveraging on simultaneous super-resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 676–680, Apr. 2020.
- [41] I. García, R. Marcos Luque, and E. López, "Improved detection of small objects in road network sequences," 2021, *arXiv:2105.08416*.

- [42] N. Q. Truong, P. H. Nguyen, S. H. Nam, and K. R. Park, "Deep learning-based super-resolution reconstruction and marker detection for drone landing," *IEEE Access*, vol. 7, pp. 61639–61655, 2019.
- [43] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, Jun. 2018, pp. 1664–1673.
- [44] S. B. Yoo and M. Han, "Temporal matching prior network for vehicle license plate detection and recognition in videos," *ETRI J.*, vol. 42, no. 3, pp. 411–419, Feb. 2020.



**JUN SEOK YUN** received the B.S. degree in electronic engineering from Chosun University, Gwangju, South Korea, in 2020. He is currently pursuing the master's degree with the Department of Artificial Intelligence Convergence, Chonnam National University, Gwangju. His research interests include visual intelligence and computer vision.



**SEOK BONG YOO** (Member, IEEE) received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2009, and the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2011 and 2015, respectively. He was a Senior Engineer at Samsung Electronics, Suwon, South Korea, from 2015 to 2017, and a Senior Researcher at the Electronics and Telecommunications Research Institute, Daejeon, from 2017 to 2020. He is currently an Assistant Professor with Chonnam National University, Gwangju, South Korea. His research interests include visual intelligence, image processing, computer vision, and immersive media synthesis.



**SUNGJIN LEE** received the B.S. degree from the College of Artificial Intelligence Convergence, Chonnam National University, Gwangju, South Korea, in 2021, where he is currently pursuing the master's degree with the Department of Artificial Intelligence Convergence. His research interests include computer vision and machine learning technologies.

...