Received August 22, 2019, accepted September 2, 2019, date of publication September 13, 2019, date of current version September 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2941005

Object Recognition in Very Low Resolution Images Using Deep Collaborative Learning

JEONGIN SEO AND HYEYOUNG PARK

School of Computer Science and Engineering, Kyungpook National University, Daegu 41566, South Korea

Corresponding author: Hyeyoung Park (hypark@knu.ac.kr)

This work was supported in part by the Institute for Information and Communications Technology Promotion (IITP) Grant funded by the Korean Government (MSIT) (Smart Summary Report Generation from Big Data Related to a Topic) under Grant 2016-0-00145, in part by the Institute for Information and Communications Technology Promotion (IITP) Grant funded by the Korean Government (MSIT) (Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding) under Grant 2016-0-00564, and in part by the BK21 Plus Project (SW Human Resource Development Program for Supporting Smart Life) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, South Korea, under Grant 21A20131600005.

ABSTRACT Although recent studies on object recognition using deep neural networks have reported remarkable performance, they have usually assumed that adequate object size and image resolution are available, which may not be guaranteed in real applications. This paper proposes a framework for recognizing objects in very low resolution images through the collaborative learning of two deep neural networks: image enhancement network and object recognition network. The proposed image enhancement network attempts to enhance extremely low resolution images into sharper and more informative images with the use of collaborative learning signals from the object recognition network. The object recognition network with trained weights for high resolution images actively participates in the learning of the image enhancement network. It also utilizes the output from the image enhancement network as augmented learning data to boost its recognition performance on very low resolution objects. Through experiments on various low resolution image benchmark datasets, we verified that the proposed method can improve the image reconstruction and classification performance.

INDEX TERMS Machine learning, object recognition, very low resolution recognition, image enhancement, deep neural networks, collaborative learning.

I. INTRODUCTION

Object recognition is one of the well-conquered problems in machine learning owing to the use of deep learning techniques [1]–[7]. After the success of AlexNet in the Imagenet large scale visual recognition challenge (ILSVRC) 2012 [1], the performance of object recognition using deep neural networks has improved rapidly and steadily. Whereas the networks in the early works were composed of limited numbers of layers owing to difficulties associated with the training process [2], various techniques such as the inception module [3] and the residual module [4]–[6] have been developed to resolve the difficulties [8]–[11]. In ILSVRC 2017, the squeeze-and-excitation network (SENet) achieved a top-5 error of 2.25% in classifying objects from 1000 classes and won the classification competition [7], which is astonishing performance considering that the top-5 error rate of humans is 5.1% [12].

Despite these prominent results, the focus on low-resolution object recognition has been weaker than that on high-resolution images. The average resolution of the images used in ILSVRC is 482×415 pixels [13]. Although those images contain backgrounds and multiple objects, they can retain sufficient information about each object, which enabled deep networks to extract rich visual features from them and achieve notable classification performance. However, there is no guarantee that the deep networks designed for high resolution object recognition can perform well when classifying extremely low resolution images, in which much of the useful object-related information is collapsed.

The very low resolution recognition problem deals with images of resolution lower than 16×16 pixels [14]–[20]. Although previous works have not dealt with this

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

topic seriously, it should not be neglected in real applications because the recognition of small objects in a large high resolution image is equivalent to low resolution problem. Considering this practical importance, numerous studies have attempted to recognize human faces from extremely low resolution images. These works related to human faces have tried to enhance low resolution images for representing facial components more clearly [14], [15] or to design new facial feature descriptors that are robust against to low resolution [16]. Deep learning techniques have further improved the recognition performance in LR facial images [17]–[19]. By contrast, there have been few works on low resolution object recognition, which involves further varieties of image features [19]–[22].

To improve the object recognition performance on low resolution images, image super-resolution (SR) techniques can be applied before the recognition step. The SR techniques based on deep neural networks [23]–[34] have outperformed traditional methods [35]. They have led to substantial improvements in image quality measures such as peak signal-to-noise ratio (PSNR) [36] and structural similarity (SSIM) index [37] and helped generate realistic high resolution images. However, the existing SR methods focus on enhancing the image quality of small patches, meaning they cannot necessarily extract useful object information from entire images. In addition, these methods do not consider semantic information, which means they may even reconstruct noise; this is undesirable from the viewpoint of object recognition. Therefore, to boost the recognition performance on low resolution images, it is necessary to develop more active enhancement methods by concentrating on the extraction of perceptually meaningful information.

With this consideration, we propose an integrated framework for object recognition in extremely LR images (8×8 pixels). This framework is based on the collaborative training of two deep neural networks: the image enhancement network (IEN) and the object recognition network. The IEN has been newly designed to enhance low resolution images into well-interpretable images that can be input into the object recognition network. The fundamental structure of the proposed IEN is inspired by various works on SR, and it includes additional convolutional blocks for extracting global context information, which is important for object recognition. In addition, we propose a new loss function for training the IEN by combining the typical super-resolution loss with additional losses associated with the collaborating object recognition network. The object recognition network is based on an existing well-trained model, and we propose systematic retraining strategies for this network that utilize the ability of the pre-trained network efficiently and augment its recognition performance on low resolution images. With these strategies, the object recognition network helps the IEN by providing training signals and uses the output of the trained IEN as additional training data for object recognition. Through the collaborative learning process, we expect the proposed model to achieve high recognition performance on low resolution and high resolution images.

The remainder of this paper is organized as follows. Section 2 describes related works on very low resolution object recognition and image super-resolution. The overall structure of the proposed networks and the training strategies employed in them are described in Section 3. Section 4 reports the experimental results and discusses the performance of the proposed method. Section 5 concludes the paper.

II. RELATED WORKS

A. VERY LOW RESOLUTION OBJECT RECOGNITION

Studies on very low resolution object recognition by using deep learning techniques are in an emergent state. Wang *et al.* [20] classified very low resolution objects by using deep learning and demonstrated the feasibility of using deep neural networks for the recognition of very low resolution objects, as well as for other recognition tasks, such as digits, faces, and fonts. However, with this approach, the networks used for each task must be trained from scratch with specific datasets, and its efficiency was not fully confirmed for common object recognition tasks with large datasets, such as ILSVRC.

Peng *et al.* [21] proposed to retrain a conventional object recognition network by using low resolution images. The retrained network exhibited marginally superior performance for low resolution images than that achieved with the typical training protocol, which uses high resolution and low resolution image simultaneously (*i.e.,* "mixed learning" in [21]). However, its performance on high resolution images was considerably poorer, implying that necessity to develop an elaborate training strategy for achieving good performance with images of various resolutions.

GenLR-Net [19] and resolution-aware convolutional neural network (RACNN) [22] also attempted to solve very low resolution object recognition problem. GenLR-Net performed well in classifying objects that were not trained by comparing the extracted features of low resolution images with the ones of high resolution images. The RACNN achieved superior classification performance than [21] owing to the attachment of SR layers ahead of an object recognition network. Although these preliminary works highlighted the necessity and feasibility of low resolution object recognition by using deep learning models, they considered only controlled variations [19] and did not fully examine image resolutions lower than 16×16 pixels [22].

The proposed training strategies for object recognition network have been designed to solve the abovementioned problems. By importing and fine-tuning a well-trained object recognition network, we attempt to achieve basic recognition performance on common high resolution objects. In addition, by actively involving the network in training the IEN, we try to ensure that the IEN can generate images suitable for object recognition. Finally, we use the images enhanced using the IEN for retraining the object recognition network to further improve its recognition performance on low resolution images.

B. IMAGE SUPER-RESOLUTION

Recent SR techniques based on deep learning have achieved high restoration performance in terms of the PSNR and the SSIM index [23]–[34]. Super-resolution convolutional neural network (SRCNN) [23] and VDSR [24] are pioneering SR techniques based on deep learning. In these techniques, convolutional layers are stacked to extract features from image patches [23], and the residual module and gradient clipping are adopted for efficiently training deep networks [24]. These developments have subsequently been applied to the enhanced deep super-resolution network (EDSR) [25] with refined and deeper structures. In addition, EDSR employed residual scaling [6], which controls the network output by multiplying it with a small scalar value. Haris *et al.* [26] proposed deep back-projection networks (DBPN) that enhance the reconstruction performance by iterating pairs of up- and down-sampling layers instead of merely stacking convolutional layers. These deep learning-based SR methods successfully recovered visual information from input images. However, the majority of deep learning-based SR methods have been designed to enhance the detailed appearance of small image patches in high resolution images. In other words, the existing deep learning-based SR methods cannot completely retain their reconstruction performance for extremely low resolution images. Moreover, during the patch-wise training process, the semantic information of a given image is typically neglected. On the contrary, the proposed IEN is designed to efficiently extract global perceptual features from extremely low resolution images and compensate for any gaps in information by means of collaborative learning with the object recognition network.

Furthermore, several studies have claimed that mean squared error (MSE) does not always guarantee higher reconstruction performance in terms of image quality for human perception [38]–[40]. Recently, a few studies on SR have employed perceptual loss based on visual information along with MSE to enhance reconstruction performance through the recovery of texture information [27]–[33]. Johnson *et al.* [28] suggested that perceptual loss helps improve image quality and generate human-acceptable images. EnhanceNet [29] indicated the limit of MSE in image restoration and the poor correlation between PSNR index and human perception, and the authors of EnhanceNet attempted to reconstruct image textures by using perceptual loss. SRGAN [30] is a SR technique based on the generative adversarial network (GAN), and it employs the perceptual loss obtained from the part of object recognition network. Although its PSNR and SSIM index are slightly lower than those of the other SR techniques that employ only MSE, SRGAN employs a sophisticated discriminator network to generate realistic images that can fool humans. This idea has succeeded to more complicated structures with additional loss functions [31], [32]. Voynov *et al.* [33] adopted perceptual loss from depth images to recover texture information. However, these techniques are less focused on practical applications tied to human perception, such as object recognition.

Haris *et al.* [34] combined their SR network (DBPN [26]) with a well-known object detection network (SSD [41]) and trained the SR network in an end-to-end manner by using the additive loss function representing the losses of the two networks. The proposed method takes a similar approach with a different purpose, namely, object recognition. In addition, instead of simple end-to-end-style combination, we present systematic training strategies and an integrated function covering multiple losses to achieve optimal object recognition performance.

III. PROPOSED METHOD

A. OVERALL STRUCTURE

The overall structure of the proposed model, as well as its training and inference, are illustrated in Fig. 1. As shown in Fig. 1(b), the entire model is composed of two networks: the IEN and the object recognition network. The proposed IEN has been newly designed to enhance low resolution images so that they can be used for object recognition. Unlike the conventional SR models that conduct patch-wise reconstruction by minimizing pixel-wise distance, the proposed IEN attempts to reconstruct object information in the entire image by maximizing perceptual fidelity. For the object recognition network, we have adopted a well-trained conventional model to benefit from its generalizability in the object recognition task. The processing flow in the inference stage is represented as a simple combination of the two networks. However, in the training stage, the two networks collaborate interactively to improve the performance of each network and achieve the ultimate goal.

To achieve the ultimate goal, which is to maximize the accuracy of object recognition on in very low resolution images, the training process is composed of three stages, as shown in Fig. 1(a). In training stage 1, a conventional object recognition network with well-trained parameters is imported and fine-tuned by using high resolution (HR) images from a specific dataset. In this fine-tuning stage, we freeze the early layers of the network to preserve the general feature extraction ability of the object recognition network, which is gained through learning using a large database. This first training stage yields an object recognition network with good recognition performance on the given specific HR images, and we use this network to create guiding signals for learning of the IEN in the second stage.

In training stage 2, the weight of the object recognition network is fixed and the IEN is trained in supervised manner by using input-output training samples and the object recognition network. As shown in Fig. 1(a), IEN takes low resolution (LR) images as inputs and generates enhanced LR (ELR) images of the same size. The outputs of IEN, the ELR images, are evaluated by four different types of loss functions. Because the target output of the IEN is HR images, the reconstruction and edge loss is computed by using the

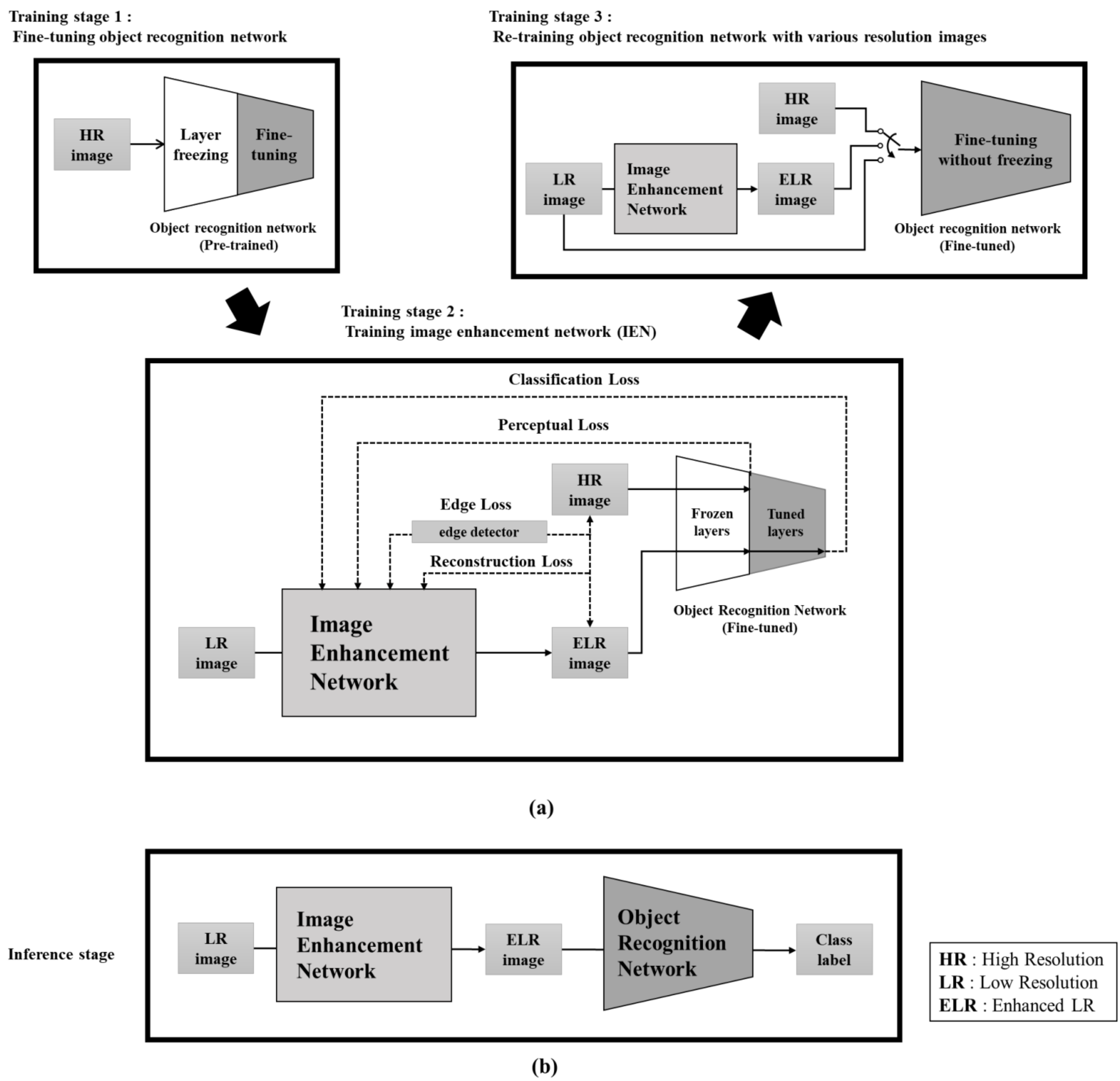

FIGURE 1. Overall structure of the proposed method. (a) training process and (b) inference process.

discrepancy between ELR and HR images. In addition, the classification and perceptual loss are computed by using the object recognition network that takes the ELR images as inputs. Using the compound loss signal, the IEN learns to reconstruct images by focusing on the information useful for object recognition. Detailed descriptions of the loss functions are given in the next subsection.

After completion of the learning of the IEN, we proceed to the third training stage. In training stage 3, the object recognition network is retrained with ELR images obtained from the trained IEN. To secure recognition performance on both HR and LR images, all HR, LR, and ELR images are input through data shuffling. The IEN is fixed in this stage, and all layers of the object recognition network are retrained without freezing to boost the object recognition ability on the given images of various resolutions.

After all three training stages are complete, the inference process can be applied to a new LR image. As shown in in Fig. 1(b), the IEN takes an LR image to generate an enhanced image, which is fed into the object recognition network as input. Then, the classification results are predicted through the object recognition network.

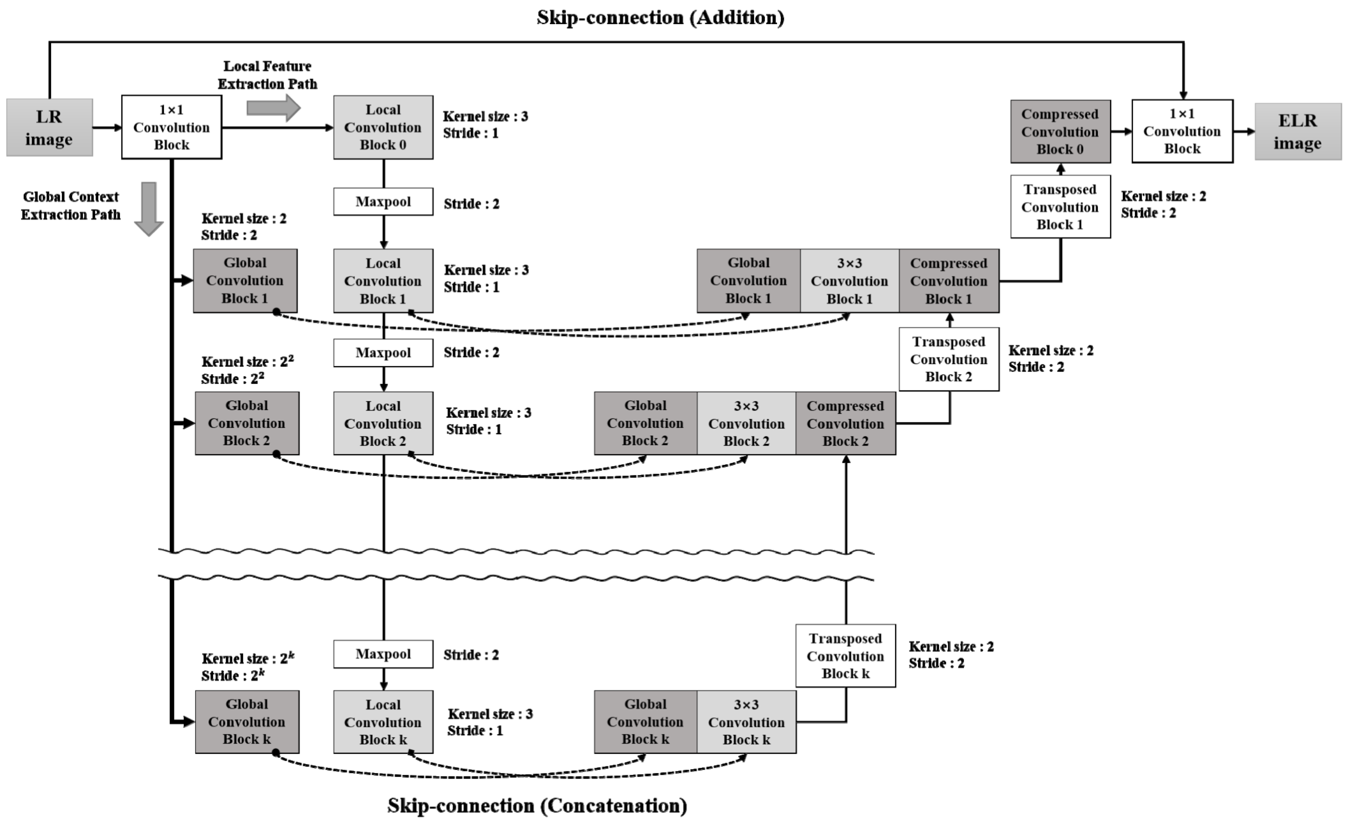

FIGURE 2. Overall structure of proposed image enhancement network.

B. STRUCTURE OF IMAGE-ENHANCEMENT NETWORK

Fig. 2 depicts the entire structure of the proposed IEN. The IEN uses several types of convolutional blocks, including convolution layers and the leaky-ReLU activation function. The structure of the IEN is inspired by U-net [42], which is composed of a CNN-based encoder–decoder architecture and skip-connection. With skip-connection of the extracted features, the IEN can reconstruct images by maintaining information across multiple scales. In the encoding part, transposed convolution layers are employed instead of simple up-sampling to generate enhanced images with a more detailed appearance.

Additionally, we employ convolutional blocks for extracting the global context information of input images. The blocks are composed of a convolution layer with non-overlapping kernels of various sizes to extract and compress image features at various scales. The sizes of the kernel and stride are doubled, starting from 2 to the size of the input image. By using these blocks, we expect that the proposed IEN can extract object-related features efficiently from LR images at various scales.

When an LR image is input into the IEN, it passes through 1×1 convolution layers once to ensure non-linearity. Then, the images are transmitted in two ways: local and global way. In the local way, image features are extracted and compressed with convolutional blocks with small kernels. In the global way, global image features are extracted using the convolutional blocks for global contexts with various receptive fields. The locally and globally extracted features are concatenated in the reconstruction process according to scale. Finally, the output of the network is added to the input image with a residual scaling of 0.1, which is known to be effective for stable learning [25].

C. TRAINING IMAGE ENHANCEMENT NETWORK

The IEN is trained in the second stage shown in Fig. 1(a). The task of the IEN is to reconstruct high resolution images from very low resolution images, which is an ill-posed problem. Therefore, to obtain the desired enhancement results, it is important to design an appropriate loss function that is suited for the ultimate goal of the given problem. Because our ultimate goal is to increase recognition accuracy, we need to train the IEN to reconstruct high resolution images by focusing on object-related information. To this end, we propose to combine four different types of loss signals. Based on the typical reconstruction loss for super resolution, we add three loss signals that can play secondary roles in inducing the learning to generate enhanced images.

1) RECONSTRUCTION LOSS

The conventional pixel-wise MSE between two images is the typical reconstruction loss used for SR. Consider a mini-batch training set composed of N pairs of HR and LR images, such as $\{(I_i^{HR}, I_i^{LR})\}_{i=1,\ldots,N}$. For a given i-th LR image I_i^{LR}, the IEN generates an enhanced image I_i^{ELR}. Then, the pixel-wise MSE between I_i^{ELR} and I_i^{HR} can be written as follows:

$$Loss_{reconstruction} = \frac{1}{N}\sum_{i=1}^{N} ||I_i^{HR} - I_i^{ELR}||_F^2, \tag{1}$$

where $||\cdot||_F$ denotes the Frobenius norm of a matrix. All images used in this study are three-channel RGB images in the form of a matrix.

Although pixel-wise MSE is a well-known and efficient loss function for super resolution, it is inadequate for reconstructing the semantic information that is essential for object recognition. Especially in the extremely low resolution case, in which most of the detail information is absent, additional guiding signals can boost the learning along the desired direction. To achieve this, we combine three additional loss functions.

2) PERCEPTUAL LOSS

The perceptual loss function has been developed to recover texture information [43]–[45], and it has been used for super resolution [27]–[33]. By using the perceptual loss function, it is expected to increase the perceptual fidelity of the reconstructed image.

For given I_i^{HR} and I_i^{ELR}, perceptual features are extracted from the intermediate layer of the object recognition network, and they are denoted as $\mathbf{f}(I_i^{HR})$ and $\mathbf{f}(I_i^{ELR})$, respectively. Then, the perceptual loss is calculated using the Euclidean distance between two feature vectors as

$$Loss_{perceptual} = \frac{1}{N}\sum_{i=1}^{N} ||\mathbf{f}(I_i^{HR}) - \mathbf{f}(I_i^{ELR})||_2^2. \tag{2}$$

Note that we extract the feature vectors $\mathbf{f}(I_i^{HR})$ and $\mathbf{f}(I_i^{ELR})$ from the frozen parts of the object recognition network, which are fixed in training stage 1. Since the fixed part has pre-trained weights by using large data set, we can expect to obtain more generalized features from the fixed layers rather than from the fine-tuned parts to the given specific dataset.

3) CLASSIFICATION LOSS

Classification loss provides the IEN with conceptual information that can be used to reconstruct distinguishable feature information for classifying objects. For a given LR image I_i^{LR}, the classification loss for the corresponding output I_i^{ELR} is defined as the cross-entropy loss of the object recognition network that takes I_i^{ELR} as its input. When the output of the object recognition network for the given input I_i^{ELR} is obtained as the one-hot vector $\mathbf{y}(I_i^{ELR})$, its cross-entropy loss can be written as follows:

$$Loss_{classification} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} \{c_j(I_i^{ELR}) \ln \mathrm{y}_j(I_i^{ELR})\}, \tag{3}$$

where $\mathrm{y}_j(I_i^{ELR})$ is the j-th element of the output vector $\mathbf{y}(I_i^{ELR})$, and $c_j(I_i^{ELR})$ is a binary target value that becomes 1 only when I_i^{ELR} is in the j-th class.

Use of the classification loss can be considered similar to the task-driven learning scheme used in [30]. Although the classification loss cannot provide explicit information about the true output values of the IEN, it can play a secondary role of guiding the proposed method toward the ultimate goal.

4) EDGE LOSS

While perceptual loss and classification loss are somewhat indirect signals from the object recognition network, edge loss is designed to represent the direct loss measured in the IEN output. Based on the supposition that the edge of an image is important for object recognition, we define a measure to account for the differences between the edges in HR images and those in the IEN output images.

To compare the edges of I_i^{HR} and I_i^{ELR}, we first design a simple edge-extraction operator $ED(I)$ by using the Sobel operator [24], which generates the edge images of a given input I. Then, we compute the edge loss based on the difference between $ED(I_i^{HR})$ and $ED(I_i^{ELR})$, as defined in the following equation:

$$Loss_{edge} = \frac{1}{N}\sum_{i=1}^{N} |(ED(I_i^{HR}) - ED(I_i^{ELR})) * K|. \tag{4}$$

Note that we additionally apply a smoothing convolution with a 3×3 filter $K = \{k_{ij}\}(k_{ij} = 1 \text{for all } i, j)$ to average out the one-pixel difference. Using this edge loss, we expect the IEN to learn by focusing on the edges that may represent object shapes.

5) TOTAL LOSS FOR IMAGE-ENHANCEMENT NETWORK

The entire loss function for the proposed IEN can then be expressed as follows:

$$\begin{aligned} Loss_{IEN} = {} & \alpha_1 Loss_{reconsturction} + \alpha_2 Loss_{perceptual} \\ & + \alpha_3 Loss_{classification} + \alpha_4 Loss_{edge}, \end{aligned} \tag{5}$$

where $\alpha_1, \alpha_2, \alpha_3$, and α_4 are user-defined hyperparameters in the range [0, 1]. In this work, we determine the hyperparameters empirically by considering the range of actual values of each loss function, as well as the tradeoff between image quality and perceptual index [34].

While the ultimate goal of the proposed method is to improve object recognition performance, the primary goal of IEN learning is to minimize reconstruction error. This is based on the assumption that if IEN can generate ELR images close to HR, the recognition accuracy of the object recognition network will be increased accordingly. Since perceptual and edge loss also depend on the difference between HR and ELR, they induce the IEN learning to generate images close to HR, especially focusing on the texture and edges of the image. Though classification loss does not depend directly on the difference between HR and ELR, it can be predicted that the loss will decrease when ELR images are close to HR, because the object recognition network is only trained for HR images. Based on these considerations, we can expect that the proposed combination of four different losses can improve the performance of IEN in the view of image quality as well as recognition accuracy.

Once the IEN is trained with the proposed losses, the output of IEN can be used to retrain the object recognition network in the third stage of the learning process.

D. TRAINING STRATEGIES FOR OBJECT RECOGNITION NETWORK

The object recognition network is trained twice: preliminary training using HR images in stage 1 and secondary training using HR and ELR images in stage 3. As shown in Fig. 1(a), in training stage 1, a pre-trained object recognition network is imported and fine-tuned using HR images from a desired dataset. In this process, early network layers are frozen, and only the later network layers are fine-tuned. The outputs from the intermediate frozen block are used to calculate the perceptual loss, and the outputs of the network are used to calculate the classification loss. After the IEN is trained, we retrain the object recognition network with ELR images without freezing any layer, as shown in training stage 3 in Fig. 1(a). After the second round of learning, the object recognition network can recognize low resolution and high resolution objects.

The reason why we trained object recognition network in two stages is to fully utilize feature extraction ability of pre-trained network and to generate the classification loss signal effectively. If the network were to be fine-tuned without freezing in the first stage, the features extracted in its intermediate layer would contain some bias toward the training data, which is undesirable for estimating the perceptual loss. In addition, if the object recognition network were to be trained using both HR and LR images in the first stage, the IEN would stop reducing the classification loss in the early steps of learning because early ELR images that are rather similar to the LR images can be well classified from the start of learning. Therefore, we trained the object recognition network with only HR images in first stage to generate a greater number of well-interpretable images that are similar to HR images.

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SETUP

To verify the performance of the proposed method, CIFAR-10, CIFAR-100 [46], and Downsampled ImageNet [47] were employed as the benchmark datasets. CIFAR-10 and CIFAR-100 comprise 50,000 training images and 10,000 test images in 10 and 100 object classes, respectively. Downsampled ImageNet contains 1.28M+ training images and 50,000 test images in 1000 object classes. The resolution of the original images in the CIFAR datasets is 32 $\times$ 32 pixels, and we composed an HR image set by using these images. To generate the LR image set, we downsampled the HR images to 8 $\times$ 8 pixels and upsampled them to 32 $\times$ 32 pixels with the bilinear method, as shown in Fig. 3. Instead of using the original ImageNet images with resolutions higher than 224 $\times$ 224 pixels, we used the ImageNet 32 $\times$ 32 pixel subset included in Downsampled ImageNet to use the same settings as those used for the CIFAR datasets. All images were resized to 224 $\times$ 224 pixels with the bilinear method when inputting them to the object recognition network, because the object recognition network has pretrained weights for 224 $\times$ 224 pixel image inputs.

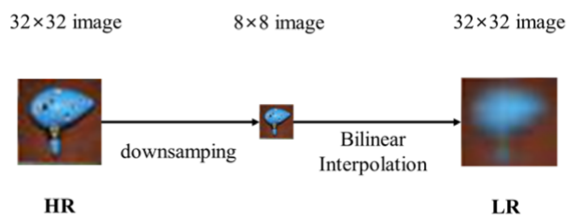

FIGURE 3. Generating LR images from HR images.

To implement the IEN, each convolutional block in local way was composed of three pairs of convolution layers with a 3 $\times$ 3 filter and the leaky-ReLU activation function. The convolutional blocks in global way are composed of a pair of convolution layers without overlap and the leaky-ReLU activation function. The blocks for concatenated features used the same output channels as the other convolutional blocks. To upscale the features by a factor of 2, we transposed the convolution layers with the leaky-ReLU activation function. The 1 $\times$ 1 convolution layer for input images employs the leaky-ReLU activation function, while the 1 $\times$ 1 convolution layer for the network output employs the identity activation function. All convolutional blocks except the 1 $\times$ 1 block for network output have 16 channels. The negative slope of the leaky-ReLU activation function was set to 0.2. The depth (k) of IEN was set to 5 for treating the 32 $\times$ 32 pixel images.

Among the loss functions, the edge loss was obtained through convolutional operation of the bi-directional Sobel operator [48]. Weight values of the loss functions were set empirically; the weight value of the perceptual, classification, and edge loss functions was set to 0.01 and that of the reconstruction loss function was set to 1. Moreover, all loss functions except reconstruction loss were applied after two epochs from the beginning to secure stability, as suggested in [30]. We chose the Adam optimizer for training the IEN, because it achieved lower loss value than SGD in the experiment.

For the object recognition networks, the ResNet-152 model [11] trained using the ILSVRC dataset was imported. In the first learning stage, the model was fine-tuned using HR images from each dataset. Note that the layers prior to the first residual block ('conv2_x' in paper) in the network were frozen and later parts of network were tuned. We used the SGD optimizer for training the object recognition network, which has been used for the pre-training of the imported

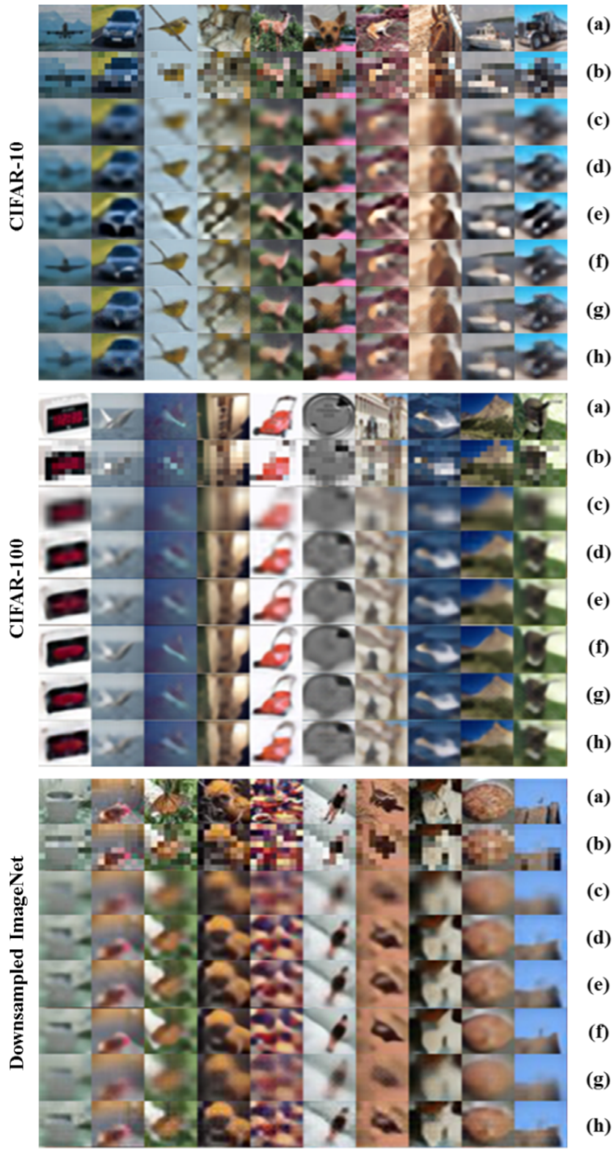

FIGURE 4. Sample images and images generated for benchmark datasets. (a) HR images, (b) downsampled HR images, (c) LR images, (d) ELR images obtained with SRCNN, (e) ELR images obtained with VDSR, (f) ELR images obtained with DBPN, (g) ELR images obtained with SRGAN, and (h) ELR images obtained with proposed IEN.

network. In primary training stage, the object recognition network was trained with early stopping based on the classification loss and the desired accuracy for the training data. The performance of the proposed method was evaluated in terms of image enhancement quality and object recognition performance.

B. IMAGE ENHANCEMENT PERFORMANCE

Before evaluating the recognition performance of the entire proposed method, we investigated the image-reconstruction performance of the proposed IEN relative to that of four popular SR models: SRCNN [23], VDSR [24], DBPN [26], and SRGAN [30]. SRCNN and VDSR are pioneering deep network models in the field of SR, and DBPN won first place in track 1 of the CVPR NTIRE 2018 challenge on single-image SR. SRGAN is a representative model that uses perceptual loss from a discriminator network to increase the perceptual fidelity. All models were adopted from open-source repositories and retrained for each benchmark dataset. In the training of SRGAN, we used the VGG22 model, which resulted in superior values of PSNR and SSIM index than those of the VGG54 model.

Table 1 lists the results of quantitative evaluation of image quality with PSNR and SSIM index, and Fig. 4 presents

TABLE 1. Image quality on benchmark dataset.

Input images	CIFAR-10		CIFAR-100		Downsampled ImageNet	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
LR (baseline)	29.136	0.430	29.897	0.460	28.897	0.382
ELR by SRCNN	29.604	0.597	29.661	0.599	29.437	0.545
ELR by VDSR	29.020	0.581	29.788	0.615	29.328	0.543
ELR by DBPN	30.117	0.651	30.231	0.656	29.797	0.596
ELR by SRGAN	29.824	0.616	29.925	0.621	28.880	0.390
ELR by Proposed IEN	29.830	0.619	29.755	0.610	29.401	0.550

TABLE 2. Recognition performance on benchmark dataset according to image quality.

Input images	CIFAR-10		CIFAR-100		Downsampled ImageNet	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
HR	96.47	99.96	86.55	97.96	47.58	71.97
LR	21.93	75.28	3.68	10.63	0.16	0.95
ELR by SRCNN	30.32	82.24	12.65	29.30	2.16	5.99
ELR by VDSR	36.14	83.14	12.07	28.59	2.03	5.84
ELR by DBPN	46.79	89.63	23.72	44.52	3,76	8.94
ELR by SRGAN	54.16	92.52	26.32	50.56	0.21	1.26
ELR by proposed IEN	60.51	95.27	28.33	54.53	7.31	17.15

TABLE 3. Image quality and recognition performance according to structure and losses.

Methods	CIFAR-10				CIFAR-100				Downsampled ImageNet			
	PSNR		SSIM		PSNR		SSIM		PSNR		SSIM	
LR image (baseline)	29.136	-	0.430	-	29.887	-	0.460	-	28.897		0.382	-
ELR image w/o GC (R)	29.828	+0.692	0.621	+0.191	29.889	+0.002	0.624	+0.164	29.538	+0.641	0.563	+0.181
ELR image with GC (R)	29.831	+0.003	0.620	-0.001	29.903	+0.014	0.629	+0.005	29.549	+0.011	0.567	+0.004
ELR image with GC (R+E)	29.847	+0.016	0.622	+0.002	29.894	-0.009	0.629	0	29.560	+0.011	0.570	+0.003
ELR image with GC (R+E+P)	29.898	+0.051	0.629	+0.007	29.963	+0.069	0.634	+0.005	29.550	-0.010	0.569	-0.001
ELR image with GC (R+E+P+C)	29.830	-0.068	0.619	-0.010	29.755	-0.208	0.610	-0.024	29.400	-0.150	0.550	-0.019
	Top-1 Acc. (%)		Top-5 Acc. (%)		Top-1 Acc. (%)		Top-5 Acc. (%)		Top-1 Acc. (%)		Top-5 Acc. (%)	
LR image (baseline)	21.93	-	75.28	-	3.68	-	10.63	-	0.16	-	0.95	-
ELR image w/o GC (R)	33.89	+11.96	81.71	+6.43	15.07	+11.39	34.42	+23.79	2.78	+2.62	7.20	+6.25
ELR image with GC (R)	33.89	0	82.60	+0.89	15.95	+0.88	34.97	+0.55	2.82	+0.04	7.46	+0.26
ELR image with GC (R+E)	34.82	+0.93	82.51	-0.09	16.92	+0.97	36.12	+1.15	2.75	-0.07	7.01	-0.45
ELR image with GC (R+E+P)	36.45	+1.63	83.16	+0.65	17.81	+0.89	38.02	+1.90	2.98	+0.23	7.46	+0.45
ELR image with GC (R+E+P+C)	60.51	+24.06	84.47	+1.31	28.33	+10.52	54.53	+16.51	7.31	+4.33	17.15	+9.69

GC: Global Convolutional Block, R: Reconstruction Loss, E: Edge Loss, P: Perceptual Loss, and C: Classification Loss

LR and HR image samples from three benchmark datasets, as well as the ELR images obtained using the SR methods and the proposed IEN. As shown in Table 1, DBPN has the best PSNR and SSIM index for all datasets. Though the proposed IEN exhibits slightly inferior performance, it is superior to SRCNN and competitive against VDSR and SRGAN, which have considerably more complex structures (Note that the proposed IEN has only 166,239 training parameters, while SRCNN, VDSR, DBPN, and the SRGAN generator have 263,075, 668,227, 10,426,358, and 1,549,443 parameters, respectively).

In addition to the pixel-based image quality, we checked the image-enhancement quality from the viewpoint of object recognition. To this end, we compared the accuracy of object recognition on the various ELR images, as summarized in Table 2. To check only the effect of image enhancement, we used the object recognition network trained using only HR images in the first training stage of the proposed method. The first two rows of Table 2 indicate drastic decreases in recognition accuracy for the LR input images, which implies the difficulty of the given task. The poor performance can be understood from sample LR images in Fig. 4, in which large amounts of perceptual information are absent. Despite the difficulty, the SR methods improve the recognition accuracy to some extent, as can be seen from Table 2. In particular, the proposed IEN achieves the best recognition performance on all datasets, and the performance of SRGAN is superior to that of DBPN on the CIFAR-10 and CIFAR-100 datasets.

From the results, we can confirm that higher PSNR and SSIM do not necessarily lead to improved recognition performance, and the reconstruction error in the sense of MSE loss cannot be the best solution for the specific task of object recognition. For the same reason, the use of perceptual loss and other additional losses seems to positively influence the performance of SRGAN and the proposed IEN. Additionally, the performance degradation of SRGAN on Downsampled ImageNet may be ascribed to the adversarial training effects of GAN models, in which a well-trained discriminator can interfere with the learning process of a generator.

Although the results in Table 1 and 2 demonstrate the efficiency of the proposed IEN model, we investigated the effect of the proposed loss functions and the global convolutional blocks more thoroughly. Starting from the IEN model without the global convolutional blocks trained with only the MSE reconstruction loss, we added the global convolutional blocks and other losses one-by-one to evaluate the changes in PSNR, SSIM, and recognition accuracy.

The first two rows of Table 3 indicate that even the basic IEN model without the global convolutional blocks trained with only the MSE loss can improve image quality as well

TABLE 4. Object recogniton performance on CIFAR-10 and CIFAR-100 dataset.

Top-1 accuracy (%)		CIFAR-10			CIFAR-100		
		HR	LR	average	HR	LR	average
Results in [20]		86.22*	81.23**	83.73	64.98*	58.48**	61.73
ResNet-152, Trained by HR images		95.01	45.16	70.09	84.96	20.57	52.77
ResNet-152, Trained by LR images		48.11	81.54	64.83	50.21	26.53	38.37
ResNet-152, Trained by HR and LR images		90.19	82.45	86.32	82.15	23.28	52.72
Proposed Method	**ResNet-152, Re-trained by ELR images**	90.90	82.43	86.67	67.69	59.52	63.61
	ResNet-152, Re-trained by HR and ELR images	**97.45**	83.14	90.30	**85.16**	59.67	**72.42**
	ResNet-152, Re-trained by HR, ELR and LR images	96.49	**84.26**	**90.38**	82.67	**61.27**	71.97

* Copied from best results of Model I
** Copied from best results of Model IV

TABLE 5. Object recognition on downsampled imagenet dataset.

Accuracy (%)		HR		LR		average	
		Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
ResNet-152, Trained by HR images		47.78	72.09	0.96	3.56	24.37	37.83
ResNet-152, Trained by LR images		1.41	4.08	11.43	26.25	6.42	15.17
ResNet-152, Trained by HR and LR images		39.84	65.01	16.85	35.22	28.35	50.12
Proposed Method	**ResNet-152, Re-trained by ELR images**	29.58	51.03	**24.82**	**45.39**	27.20	48.21
	ResNet-152, Re-trained by HR and ELR images	**48.48**	**72.74**	23.05	42.89	**35.77**	**57.82**
	ResNet-152, Re-trained by HR, ELR and LR images	48.10	72.28	22.98	43.23	35.54	57.76

as recognition accuracy, and the addition of the global convolutional blocks improves the performance further in most cases. The addition of edge loss and perceptual loss causes marginal performance degradation in a few cases, but they have positive effects on average. In the overall sense, edge loss is more effective for enhancing image quality, and perceptual loss is more effective for increasing recognition accuracy. Finally, owing to introduction of the classification loss, PSNR and SSIM index decrease, but the recognition accuracy improves remarkably. Based on these results, we confirm that the proposed loss functions work cooperatively to achieve the ultimate goal.

C. OBJECT RECOGNITION PERFORMANCE

Though we confirm that the proposed IEN can achieve remarkable improvement in recognition accuracy as shown in Table 2 and 3, it is still unsatisfactory. To further increase the accuracy, it is essential to conduct the training stage 3 for the object recognition network by using ELR images. Tables 4 and 5 summarize the evaluation results on the object recognition performance of the entire proposed model obtained through training stage 3. To verify the performance of the proposed method, we have presented the performances of ResNet-152, which were obtained by using various training protocols. The results obtained by using HR and LR images simultaneously can be considered as a 'mixed learning' protocol in [21]. In addition, we have depicted the performance reported in [20] in table 4, which is the only study reporting the results obtained using 8×8 pixel images. For the proposed method, we tried three different compositions of datasets for training in the third stage.

As shown in the first row of the tables, [20] achieved remarkably high accuracy for LR images, albeit at the expense of accuracy for HR images. The low performance of the model in [20] on HR images may be ascribed to the relatively simple underlying network model that was specifically designed for recognizing LR images. On the contrary, the conventional object recognition networks achieve high recognition accuracy for HR images but significantly low recognition accuracy for LR images. Although the performance for LR images can be improved by using LR training images, it still has similar problems to the case that using HR images. The mixed learning protocol that uses HR and LR images has reasonable performance for both test sets. In the proposed method, we have utilized the conventional well-trained object recognition network to ensure that the method performs competently on both HR and LR images. In addition, we confirmed that remarkable performance improvement can be achieved by implementing the proposed third stage of learning. When trained with the ELR images alone, the proposed method exhibits a marginal decrease in accuracy for HR images, but this decrease can be compensated by training the method with HR and ELR images simultaneously. Moreover, the simultaneous use of LR and HR images can further increase the accuracy on LR images. The tradeoff between the accuracy on HR images and LR images can be treated with an appropriate combination of three training sets (HR, ELR, and LR). To the best of our knowledge, this is the first work dealing with the 8×8 Downsampled ImageNet.

V. CONCLUSION

This paper proposes a collaborative training system comprising an IEN and object recognition network for recognizing very low resolution objects. Using the training signals originating from the object recognition network, the IEN can generate images with improved quality in terms of appearance and perception. The proposed IEN employs considerably fewer parameters than conventional SR networks, but it can efficiently reconstruct high resolution information that is essential for object recognition. This purpose-driven reconstruction is achieved with appropriately designed loss functions that actively use of the object recognition networks. The object recognition network, which has been imported from a well-trained conventional model, can generate good loss signals for the IEN. In addition, through retraining using the outputs of the IEN, the recognition ability of the object recognition network can be extended to very low resolution objects. Consequently, the proposed systematic collaboration between two deep networks can serve as an efficient solution for to the task of very low resolution object recognition. Even though we have focused on the object recognition problem, the proposed framework can be applied to other low resolution problems, such as faces and letters, which will be done in future studies.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, Sep. 2015, pp. 1–14.

[3] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[5] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.

[6] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. Int. Conf. Learn. Represent. Workshop*, Feb. 2016, pp. 4278–4284.

[7] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[8] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Lille, France, Jul. 2015, pp. 448–456.

[9] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, Haifa, Israel, Jun. 2010, pp. 807–814.

[10] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jun. 2014.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.

[12] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[13] *Large Scale Visual Recognition Challenge 2014*. Accessed: May 20, 2019. [Online]. Available: http://www.image-net.org/challenges/LSVRC/2014/

[14] W. W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 327–340, Jan. 2012.

[15] M.-C. Yang, C.-P. Wei, Y.-R. Yeh, and Y.-C. F. Wang, "Recognition at a long distance: Very low resolution face recognition and hallucination," in *Proc. Int. Conf. Biometric.*, May 2015, pp. 237–242.

[16] Z. Wang, Z. Miao, Q. J. Wu, Y. Wan, and Z. Tang, "Low-resolution face recognition: A review," *Vis. Comput.*, vol. 30, no. 4, pp. 359–386, Apr. 2014.

[17] E. Zangeneh, M. Rahmati, and Y. Mohsenzadeh, "Low resolution face recognition using a two-branch deep convolutional neural network architecture," Jun. 2017, *arXiv:1706.06247*. [Online]. Available: https://arxiv.org/abs/1706.06247

[18] M. Wang and W. Deng, "Deep face recognition: A survey," Apr. 2018, *arXiv:1804.06655*. [Online]. Available: https://arxiv.org/abs/1804.06655

[19] S. P. Mudunuri, S. Sanyal, and S. Biswas, "GenLR-net: Deep framework for very low resolution face and object recognition with generalization to unseen categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit, Workshop*, Jun. 2018, pp. 602–611.

[20] 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.*, Jun. 2016, pp. 4792–4800.

[21] X. Peng, J. Hoffman, S. X. Yu, and K. Saenko, "Fine-to-coarse knowledge transfer for low-res image classification," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 3683–3687.

[22] D. Cai, K. Chen, Y. Qian, and J. Kämäräinen, "Convolutional low-resolution fine-grained classification," *Pattern Recognit. Lett.*, vol. 119, pp. 166–171, Mar. 2019.

[23] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[24] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1646–1654.

[25] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop*, Jul. 2017, pp. 136–144.

[26] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1664–1673.

[27] J. Bruna, P. Sprechmann, and Y. LeCun, "Super-resolution with deep convolutional sufficient statistics," in *Proc. Int. Conf. Learn. Represent.*, 2016.

[28] J. Johnson, A. Alahi, and L. Fei-Fei, ''Perceptual losses for real-time style transfer and super-resolution,'' in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.
[29] M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch, ''EnhanceNet: Single image super-resolution through automated texture synthesis,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2017, pp. 4501–4510.
[30] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, ''Photo-realistic single image super-resolution using a generative adversarial network,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4681–4690.
[31] M. Cheon, J.-H. Kim, J.-H. Choi, and J.-S. Lee, ''Generative adversarial network-based image super-resolution using perceptual content losses,'' in *Proc. Eur. Conf. Comput. Vis. Workshop*, Sep. 2018, pp. 51–62.
[32] B. Wu, H. Duan, Z. Liu, and G. Sun, ''SRPGAN: Perceptual generative adversarial network for single image super resolution,'' Dec. 2017, *arXiv:1712.05927*. [Online]. Available: https://arxiv.org/abs/1712.05927
[33] O. Voynov, A. Artemov, V. Egiazarian, A. Notchenko, G. Bobrovskikh, D. Zorin, and E. Burnaev, ''Perceptual deep depth super-resolution,'' Dec. 2018, *arXiv:1812.09874*. [Online]. Available: https://arxiv.org/abs/1812.09874
[34] M. Haris, G. Shakhnarovich, and N. Ukita, ''Task-driven super resolution: Object detection in low-resolution images,'' May 2018, *arXiv:1803.11316*. [Online]. Available: https://arxiv.org/abs/1803.11316
[35] K. Nasrollahi and T. B. Moeslund, ''Super-resolution: A comprehensive survey,'' *Mach. Vis. Appl.*, vol. 25, no. 6, pp. 1423–1468, Aug. 2014.
[36] A. Hore and D. Ziou, ''Image quality metrics: PSNR vs. SSIM,'' in *Proc. Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
[37] 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.
[38] H. R. Sheikh and A. C. Bovik, ''Image information and visual quality,'' *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
[39] W. Xue, X. Mou, L. Zhang, and X. Feng, ''Perceptual fidelity aware mean squared error,'' in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 705–712.
[40] M. Mathieu, C. Couprie, and Y. LeCun, ''Deep multi-scale video prediction beyond mean square error,'' in *Proc. Int. Conf. Learn. Represent.*, Nov. 2016.
[41] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, ''SSD: Single shot multibox detector,'' in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2016, pp. 21–37.
[42] O. Ronneberger, P. Fischer, and T. Brox, ''U-net: Convolutional networks for biomedical image segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Nov. 2015, pp. 234–241.
[43] A. Mahendran and A. Vedaldi, ''Understanding deep image representations by inverting them,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5188–5196.
[44] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, ''Understanding neural networks through deep visualization,'' in *Proc. Int. Conf. Mach. Learn.*, Jun. 2015.
[45] L. Gatys, A. S. Ecker, and M. Bethge, ''Texture synthesis using convolutional neural networks,'' in *Proc. Neural Inf. Process. Syst.*, 2015, pp. 262–270.
[46] A. Krizhevsky, ''Learning multiple layers of features from tiny images,'' Tech. Rep., 2009.
[47] P. Chrabaszcz, I. Loshchilov, and F. Hutter, ''A downsampled variant of imagenet as an alternative to the cifar datasets,'' Jul. 2017, *arXiv:1707.08819*. [Online]. Available: https://arxiv.org/abs/1707.08819
[48] S. Gupta and S. G. Mazumdar, ''Sobel edge detection algorithm,'' *Int. J. Comput. Sci. Manage. Res.*, vol. 2, no. 2, pp. 1578–1583, Feb. 2013.

JEONGIN SEO received the B.S., M.S., and Ph.D. degrees in computer science from Kyungpook National University, Daegu, South Korea, in 2009, 2013, and 2019, respectively, where he is currently a Postdoctoral Researcher. His current research interests include machine learning, artificial intelligence, and computer vision.

HYEYOUNG PARK received the B.S. degree *(summa cum laude)* and the M.S. and Ph.D. degrees in computer science from Yonsei University, Seoul, South Korea, in 1994, 1996, and 2000, respectively. She was a Member of Research Staff with the Brain Science Institute, RIKEN, Japan, from 2000 to 2004. She is currently a Professor with the School of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea. Her current research interests include computational learning theory, machine learning theory and their application to various fields, such as pattern recognition, image processing, and data mining.