

## 차량 번호판 화질 개선을 위한 GAN 기반의 초해상도 기법

Deep Super-resolution Method via Generative Adversarial Networks for License Plate Image Enhancement

---

저자 (Authors)	최준명, 강동중 Junmyung Choi, Dongjoong Kang
출처 (Source)	<a href="#">제어로봇시스템학회 논문지 23(8)</a> , 2017.8, 635-643(9 pages) <a href="#">Journal of Institute of Control, Robotics and Systems 23(8)</a> , 2017.8, 635-643(9 pages)
발행처 (Publisher)	<a href="#">제어로봇시스템학회</a> Institute of Control, Robotics and Systems
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07221369">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07221369</a>
APA Style	최준명, 강동중 (2017). 차량 번호판 화질 개선을 위한 GAN 기반의 초해상도 기법. 제어로봇시스템학회 논문지, 23(8), 635-643
이용정보 (Accessed)	국민대학교 113.198.***.81 2021/08/30 18:55 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

# 차량 번호판 화질 개선을 위한 GAN 기반의 초해상도 기법

## Deep Super-resolution Method via Generative Adversarial Networks for License Plate Image Enhancement

최 준 명, 강 동 중\*

(Junmyung Choi<sup>1</sup> and Dongjoong Kang<sup>1,\*</sup>)

<sup>1</sup>School of Mechanical Engineering, Pusan National University

**Abstract:** Convolutional neural networks and generative adversarial neural networks have recently shown outstanding performance in single-image super-resolution. In this paper, we propose a deep super-resolution method based on generative adversarial networks to reconstruct a high-resolution license plate image from a low-resolution license plate image to improve the accuracy of low-resolution license plate recognition. To achieve this, we create a super-resolution model using deep residual blocks. In addition, our model uses the benefits of adversarial loss, pixel-wise loss, and perceptual loss. Adversarial loss plays a role in enabling the discriminator network to distinguish between the high-resolution image and the fake high-resolution image generated by the generator network well, while it facilitates the generator network to generate a realistic high-resolution image with the goal of fooling the discriminator network. Pixel-wise loss and perceptual loss help the generator network to reconstruct a better high-resolution image. We compare the proposed method and other state-of-the-art methods with the performance of the license plate image recognizer. The experimental results demonstrate that our proposed method performs well by showing that the accuracy of the low-resolution image is improved from 38.45 to 73.65% by the proposed method.

**Keywords:** super resolution, generative adversarial network, deep learning, image recognition, license plate

### I. 서론

CCTV나 차량의 블랙박스 등 같은 영상 획득 장치를 이용하여 얻은 영상은 경찰조사나 과학수사 등에서 중요한 역할을 한다. 하지만 영상 획득 장치를 통해 얻은 영상은 카메라와 대상 물체 사이의 거리와 촬영 각도, 빛, 노이즈와 같은 외부 영향이나 영상 획득 장치의 기술적 한계로 인해 영상 열화가 발생한다. 그로 인해 영상에서 차량의 번호판을 인식하는데 어려움을 겪는 사례가 빈번히 발생한다[1].

과거에도 초해상도 기법을 통한 저해상도 영상 복원에 관한 연구가 이루어져 왔다[4-13]. 그러나 자동차 번호판은 영상에서 작은 부분을 차지하고 열화된 영상으로 인해 이를 온전히 복원하기에는 기존의 특징 기반 영상처리나 딥러닝 기반 방법(그림 1(a))으로는 한계가 있다. 따라서 본 논문에서는 새로운 초해상도 기법을 통해 저해상도 흑백 자동차 번호판을 고해상도 영상으로 복원하여

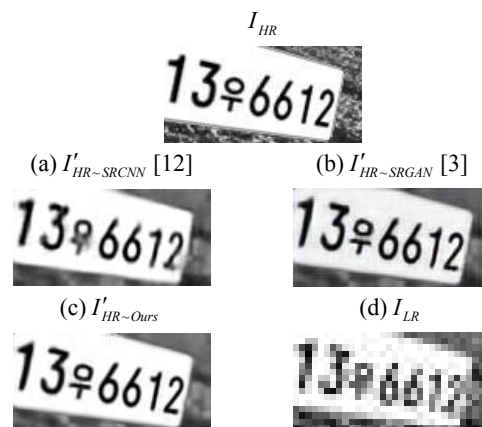


그림 1. 원본과 복원된 영상 샘플. (a) SRCNN[12], (b) SRGAN[3], (c) 제안한 방법, (d)  $16 \times 32$  입력 데이터 (복원 대상)

Fig. 1. Illustration of reconstruction results. (a) SRCNN, (b) SRGAN, (c) proposed method, (d)  $16 \times 32$  input

\* Corresponding Author

Manuscript received May 2, 2017 / revised June 13, 2017 / accepted June 27, 2017

최준명, 강동중: 부산대학교 기계공학부

(jmchoi@pusan.ac.kr/djkang@pusan.ac.kr)

※ 본 연구는 2017년도 미래창조과학부의 재원으로 한국연구재단 (2016R1A2B4007608)과 SW융합기술고도화사업(S0602-17-1001)의 지원을 받아 수행된 연구임.

※ 이 논문은 2017 제 32회 제어·로봇·시스템학회 학술대회에 초안이 발표되었음.

Copyright© ICROS 2017

자동차 번호판 인식 성능을 개선하는 방법을 제안한다. 제안한 방법은 생성적 적대 신경망(Generative Adversarial Networks, GAN)과 레지듀얼 신경망(Residual network 이하 Resnet, [2])을 사용하여 저해상도 영상을 복원한다. 생성적 적대 신경망은 최근 연구되고 있는 비지도 학습(unsupervised learning) 방식으로써 영상의 생성과 변환에 관한 많은 연구가 진행되고 있으며, 영상 화질 개선에도 좋은 성능을 보인다[3,31]. Resnet 기반의 모델은 그 이전

의 VGG16[21] 등에 비교하여 빠른 학습 속도, 높은 정밀도를 가지며, 더 복잡한 비선형성을 수용할 수 있도록 깊은 학습이 가능하며 일반적인 컨볼루션 신경망(Convolutional Neural Networks 이하 CNN)으로 구성된 모델보다 더 높은 정확도를 얻을 수 있다[2].

본 논문의 구성은 다음과 같다. II장에서 제안하는 기법과 관련한 연구에 대해 기술한다. III장에서는 제안하는 기법의 상세 학습과정, 생성적 적대 신경망의 구조와 학습에 사용된 손실 함수에 관해 기술한다. IV장에서는 학습 데이터 생성 과정, 번호판 인식기 신경망 구조와 학습 파라미터 설정에 관해 기술한다. 그리고 V장에서는 기존의 방식으로 복원된 영상 결과와 제안하는 방법을 이용한 영상 복원 결과를 바탕으로 번호판 인식기의 인식 정확도를 비교 분석한다. 끝으로 VI장에서 결론을 기술한다.

## II. 관련 연구

### 1. 초해상도 기법

초해상도 방법에 관한 연구는 오래전부터 지속해서 연구된 분야이다. Nasrollashi, Moeslund[4]와 Yang et al. [5]는 다양한 방식의 초해상도 방법에 대한 내용을 기술하였다. 초해상도 방법 예로는 전통적인 예측 기반의 방법(bilinear, bicubic, Lanczos, [6])을 포함하여, 다수의 저해상도 영상들로부터 고해상도 영상을 복원하는 다중 프레임 기반 방법[7,8], edge를 이용한 방법[9], 통계학적 기법을 이용한 방식[10] 등이 있다. 또한 최근에는 딥러닝 기반의 컨볼루션 신경망(Convolutional Neural Networks 이하 CNN)을 이용한 방식[3,11-13] 등이 있다. 이처럼 근래에는 다양한 학습 기반의 방식이 영상 개선을 위해 적용되었고, 그 중에서 딥러닝 기반 초해상도 기법이 뛰어난 성능을 내고 있다[12,13]. 기존의 초해상도 방법은 학습 과정에서 번호판 영상의 특징 추출 과정이 반드시 선행되어야 한다. 따라서 사용자가 임의로 정의하는 특징 추출 과정에 따라 성능이 좌우되는 한계점이 있다. 그러나 딥러닝 기반의 학습 방식은 그 이전의 학습 방식에서 요구되던 특징 추출 과정[14,32]이 필요하지 않기 때문에 일반화 성능이 높다. 제안하는 방식 또한 특징 추출 과정없이 학습 과정이 이루어지며, 학습 과정에서 열화 영상이 포함된 문자와 숫자에 대한 정보가 필요 없다. 제안하는 방식은 기존의 딥러닝 기반의 방식보다 더 깊은 신경망으로 구성된 생성적 적대 신경망을 이용한 방식으로 다양한 손실 함수를 통해 학습한다.

### 2. 생성적 적대 신경망

생성적 적대 신경망은 비지도 학습의 일종으로 Goodfellow et al. [15]가 처음 제안한 네트워크이다. 그 후 딥 컨볼루션 생성적 적대 신경망(deep convolutional generative adversarial network)[15]이 영상 생성에 뛰어난 성능을 보여줌으로써 생성적 적대 신경망 성능 개선에 관한 연구뿐만 아니라 영상의 변환, 생성, 복원 등과 같은 많은 분야에 관한 응용 연구가 활발히 수행되고 있다[3,16-18].

특히 초해상도 생성적 적대 신경망을 이용한 저해상도 영상 복원(그림 1(b))은 기존의 CNN 기반의 모델보다 더 좋은 성능을 보여줄 뿐만 아니라 선명한 영상을 생성할 수 있음을 보여준다[3].

본 논문에서 활용한 초해상도 생성적 적대 신경망이란 Ledig et al. [3]이 처음 제안한 방법이며, 생성적 적대 신경망을 초해상도 기법에 적용한 것이다. 초해상도 생성적 적대 신경망은 생성자와 구별자라는 두 개의 신경망이 서로 대립하면서 각자의 성능을 점차 개선해가며 학습하는 방식이다. 먼저 생성자 신경망은 저해상도 영상( $I_{LR}$ )을 고해상도 영상( $I_{HR}$ )과 비슷하게 생성한다. 구별자 신경망은 생성자 신경망을 통해 생성된 고해상도 영상과 실제 고해상도 영상을 분류하도록 학습한다. 학습이 진행됨에 따라 구별자 신경망은 실제 고해상도 영상과 생성자 신경망을 통해 생성된 고해상도 영상을 잘 분류하게 된다. 그러나 학습 과정에서 생성자 신경망이 고해상도와 비슷한 영상을 생성하도록 학습함으로써 구별자 신경망은 생성된 고해상도 영상과 실제 고해상도 영상의 구분이 점점 모호해진다. 이처럼 서로 다른 목적을 가진 두 신경망이 서로 경쟁적인 구도로 학습을 진행하고 학습이 진행될수록 생성자 신경망은 구별자 신경망이 잘 구별할 수 없을 정도로 실제 고해상도 영상과 비슷한 영상을 생성하는 원리이다.

## III. 제안하는 초해상도 생성적 적대 신경망 기법

### 1. 상세 학습 과정

제안하는 기법의 학습개략도는 그림 2와 같다. 생성자 신경망( $G_{\theta_G}(x_{LR}), x_{LR} \sim I_{LR}$ )의 입력데이터는 저해상도 영상( $I_{LR}$ )이며, 복원된 영상( $I'_{HR} = G_{\theta_G}(I_{LR})$ )를 출력한다. 구별자 신경망( $D_{\theta_D}(x_{HR}), x_{HR} \sim I_{HR}, I'_{HR}$ )은 고해상도 영상( $I_{HR}$ )과 복원된 영상( $I'_{HR}$ )을 입력으로 받아 확률 분포의 평균값을 출력한다. 기존의 초해상도 생성적 적대 신경망과는 달리 복원된 영상( $I'_{HR}$ )과 고해상도 영상( $I_{HR}$ )은 내용 손실 함수뿐만 아니라 픽셀 단위 손실 함수의 변수로써도 이용된다. 즉, 복원된 영상( $I'_{HR}$ ), 고해상도 영상( $I_{HR}$ ), 구별자 신경망을 통해 출력된 값( $D_{\theta_D}(I'_{HR})$ )은 생성자 신경망 손실 함수( $L_G$ )의 구성요소가 되며,  $D_{\theta_D}(I_{HR})$ 와  $D_{\theta_D}(I'_{HR})$ 는 구별자 신경망의 손실 함수( $L_D$ )의 구성요소가 된다. 손실 함수는 III.3절에서 상세히 설명한다. 여기서  $\theta_G, \theta_D$ 는 각각 생성자 신경망과 구별자 신경망의 파라미터이며,  $I_{HR}, I'_{HR} \in R^{64 \times 128 \times 3}, I_{LR} \in R^{16 \times 32 \times 3}$ 이다.

학습은  $\theta_G, \theta_D$ 를 최적화하는 과정이다. 학습 과정에서 생성자 신경망( $G_{\theta_G}(x)$ )을 통해 복원된 영상( $I'_{HR}$ )이 고해상도 영상( $I_{HR}$ )과 유사해지도록 생성자 신경망의 파라미터

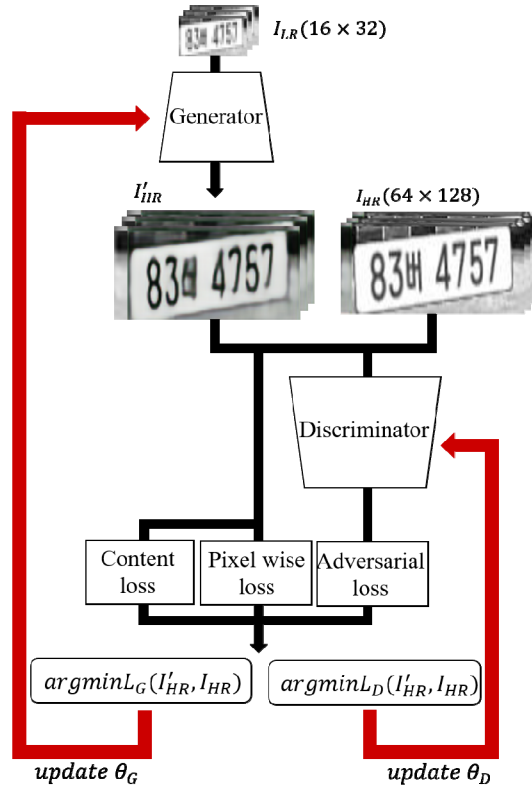


그림 2. 초해상도 생성적 적대 신경망 학습 개략도.

Fig. 2. Training procedure diagram of Super-resolution generative adversarial network.

$\theta_G$ 를 반복적으로 갱신한다. 최종적으로  $L_G$ 가 최소값을 가지는 최적 파라미터  $\hat{\theta}_G$  학습하도록 진행되며, 식 (1)과 같이 정의한다.

$$\hat{\theta}_G = \arg \min_{\theta_G} L_G(I'_{HR}, I_{HR}) \quad (1)$$

이때 구별자 신경망  $D_{\theta_D}(x)$ 은 복원된 영상( $I'_{HR}$ )과 고해

상도 영상( $I_{HR}$ )을 잘 구별할 수 있도록 하는 최적 구별자 신경망의 파라미터( $\hat{\theta}_D$ )를 찾기 위한 과정으로 학습이 진행된다. 즉, 구별자 파라미터  $\theta_D$ 를 반복적으로 갱신하여  $L_D$ 가 최소값을 가지는 최적 파라미터  $\hat{\theta}_D$ 를 학습한다. 그 식은 식 (2)과 같이 정의한다.

$$\hat{\theta}_D = \arg \min_{\theta_D} L_D(I'_{HR}, I_{HR}) \quad (2)$$

생성자 신경망 파라미터  $\theta_G$ 의 최적화 과정과 구별자 신경망 파라미터  $\theta_D$ 의 최적화 과정은 순차적으로 진행된다. 학습 목표는 고해상도 영상과 유사한 복원된 영상을 생성하는 것이므로  $L_G$ 이 최소값으로 수렴하게 되면 학습 종료한다.

## 2. 초해상도 생성적 적대 신경망 구조

본 논문에서 제안한 생성적 적대 신경망의 생성자 신경망과 구별자 신경망의 구조는 그림 3과 같다. 먼저 생성자 신경망은 Resnet 기반의 레지듀얼 블록(Residual block, [2])과, 업샘플링 층(Upsampling layer), 컨볼루션 층(Convolution layer)으로 구성되어 있으며, 학습 안정화를 위해 batch normalization(BN, [19])과 활성화 함수는 Rectified linear unit(ReLu, [20])를 이용하였다. 생성자 신경망은 저해상도 영상을 입력으로 받아 레지듀얼 블록을 통해 저해상도 영상 중의 번호판에 해당하는 부분의 특징을 추출하고 업샘플링 층에서 크기 4배 확대한다. 그 후 컨볼루션 층을 통하여 업샘플링 과정에서 확대된 영상의 화질을 개선한다. 구별자 신경망은 VGG 신경망[21] 기반의 컨볼루션 층, 서브샘플링 층(Subsampling layer), 완전 연결 층(Fully connected layer)으로 구성되어 있다. 또 VGG 신경망을 일부 수정하여 서브샘플링 과정은 컨볼루션의 거리(stride)를 2로 하여 서브샘플링하였고, Radford et al. [15]이 제안한 것과 같이 활성화 함수는 음수 기울기 0.2의

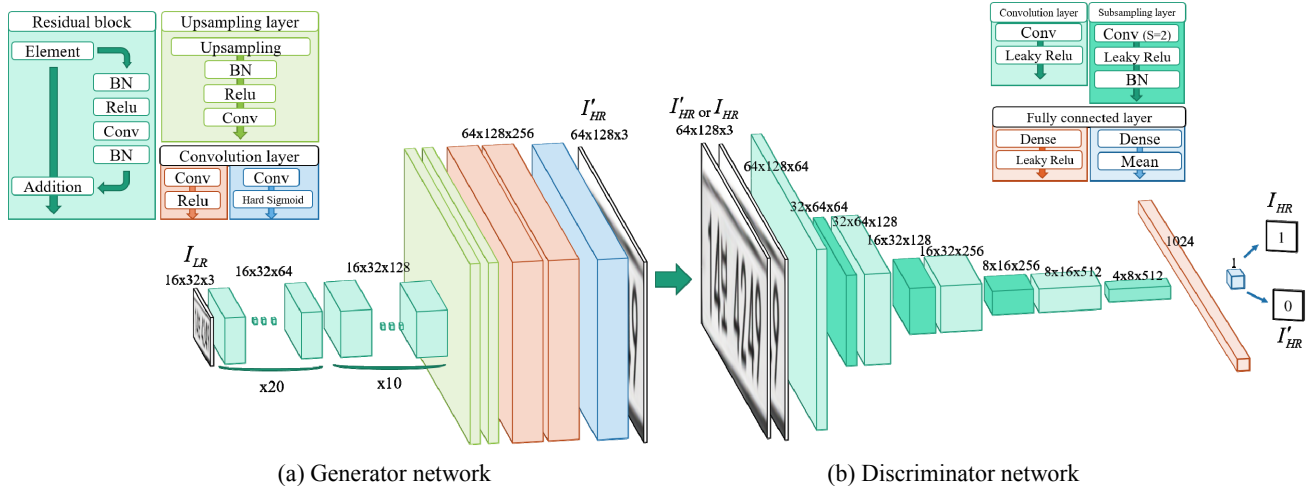


그림 3. 제안하는 생성적 적대 신경망의 생성자와 구별자 신경망 구조.

Fig. 3. Generator and discriminator network architecture of the proposed method.

leaky rectified linear unit(Leaky Relu, [22])을 이용하였다. 구별자 신경망은 복원된 영상과 고해상도 영상을 입력으로 받아 컨볼루션 층에서 복원된 영상과 고해상도 영상의 특징을 추출하는 역할을 하며, 완전 연결 층을 통해 분류를 한다. 이때 고해상도 영상의 분류 값은 1에 가까운 값을 복원된 영상의 분류 값은 0에 가까운 값을 출력하도록 학습하여 분류한다. 이때 모든 커널(kernel)의 크기는  $3 \times 3$ 이다.

### 3. 손실 함수(Loss function)

제안한 방법의 학습에 사용되는 손실 함수는 크게 생성자 신경망 손실 함수와 구별자 신경망 손실 함수로 정의할 수 있다.

먼저 생성 신경망 손실 함수는 식 (3)과 같이 정의한다.

$$L_{G_{\theta_G}} = 10^{-1} \times L_{Adversarial_G} + 10^3 \times L_{pixelwise} + 10 \times L_{Content} \quad (3)$$

생성자 손실 함수는 적대 손실 함수( $L_{Adversarial_G}$ )와 MSE와 같은 픽셀 단위 손실 함수( $L_{pixelwise}$ )[12]와 내용 손실 함수( $L_{Content}$ ) [3,23]으로 구성된다. 각 손실 함수의 가중치는 실험을 통한 경험에 근거하여 결정했다.

구별자 신경망 손실 함수는 식 (4)로 정의한다.

$$L_{D_{\theta_D}} = -\frac{1}{N} \sum_{n=1}^N (\log D_{\theta_D}(I_{HR}^n)) - \frac{1}{N} \sum_{n=1}^N (\log(1 - D_{\theta_D}(I_{HR}'^n))) \quad (4)$$

여기서,  $I_{HR}^n, n=1, \dots, N$  은  $n$ 번째 고해상도 영상이며,  $I_{HR}'^n, n=1, \dots, N$  은  $n$ 번째 복원된 영상이다. 구별자 신경망( $D_{\theta_D}(x)$ )은 최적화 과정에서  $D_{\theta_D}(I_{HR})$ 이 1이 되도록 학습하는 동시에  $D_{\theta_D}(I_{HR}')$ 이 0이 되도록 학습한다.

#### 3.1. 적대 손실 함수(Adversarial loss function)

적대 손실 함수( $L_{Adversarial_{loss_G}}$ )는 식 (5)와 같이 정의한다.

$$L_{Adversarial_{loss_G}} = -\frac{1}{N} \sum_{n=1}^N (\log D_{\theta_D}(I_{HR}'^n)) \quad (5)$$

적대 손실 함수의 주된 목적은 구별자 신경망이 복원된 영상( $I_{HR}'$ )과 고해상도 영상( $I_{HR}$ )의 구별이 모호해지도록 학습하는 것이다. 즉, 최적화 과정에서 구별자 신경망은 식 (4)의  $D_{\theta_D}(I_{HR})$ 가 0이 되도록 학습하는데 반해, 생성자 신경망은 식 (5)의  $D_{\theta_D}(I_{HR}')$ 가 1이 되도록 학습하여 구별자 신경망이 복원된 영상과 고해상도 영상을 분류하는 것을 어렵게 한다.

#### 3.2. 픽셀 단위 손실 함수(Pixel-wise loss function)

픽셀 단위 손실 함수( $L_{pixelwise_{loss}}$ )는 MSE와 같은 초해상도 기법에서 일반적으로 사용되는 손실 함수이다[12]. 픽셀 단위 손실 함수는 식 (6)과 같다.

$$L_{pixelwise_{loss}} = \frac{1}{N} \sum_{n=1}^N \left( \frac{1}{WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{HR-x,y}^n - I_{HR-x,y}'^n)^2 \right) \quad (6)$$

여기서,  $W$ 와  $H$ 는 각각 저해상도 영상의 너비와 높이

이며,  $x$ 와  $y$ 는 너비와 높이의 인덱스이다.

### 3.3. 내용 손실 함수(Content loss function)

내용 손실 함수는 학습에 있어 중요한 요소 중 하나이다. 내용 손실 함수는 픽셀 단위 손실 함수가 가진 단점을 보완하고자 사용한다. 예를 들어, 영상에 미세한 변화가 있을 때, 육안으로는 같은 영상으로 보이지만 픽셀 단위 손실 함수를 사용했을 경우에는 완전히 다른 영상으로 취급한다. 또한 픽셀 단위 손실 함수는 영상의 모든 부분의 오차를 측정하기 때문에 영상 안에서 자동차 번호판에만 집중하기 어렵다. 따라서 픽셀 단위 손실 함수가 가지는 단점을 보완하기 위해 위치와 상관없이 영상의 유사성을 고려하는 동시에 자동차 번호판의 특징 정보를 가진 내용 손실 함수를 추가한다. 그 식은 (7)과 같이 정의한다.

$$L_{Content_{loss}} = \frac{1}{N} \sum_{n=1}^N \left( \frac{1}{C_{i,j} W_{i,j} H_{i,j}} \sum_{z=1}^{C_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I_{HR-x,y}) - \phi_{i,j}(I_{HR-x,y}'^n))^2 \right) \quad (7)$$

여기서  $\phi_{i,j}$ 는 ImageNet 데이터[24]로 기 학습(pretrained)된 VGG16 신경망[21]의 중간 층 출력 함수다. 이때,  $i, j$ 는 VGG16 신경망의 블록과 층 수를 나타내며  $W_{i,j}, H_{i,j}, C_{i,j}$ 는 VGG16 신경망의  $i, j$ 층 특징맵(feature map, [25])의 너비, 높이와 채널을 나타낸다. 즉, 특정 층( $i$ 번째 블록의  $j$ 번째 층)의 특징맵 손실 함수를 계산함으로써 두 영상이 가진 번호판의 위치와 상관 없이 번호판의 유사성만 고려하여 학습을 할 수 있다. 본 논문에서는 실험을 통해 최종적으로 VGG16 신경망의  $i=2, j=2$  층을 이용하였다.

## IV. 실험

딥러닝 학습을 위한 충분한 양의 실제 자동차 번호판을 구하는 것은 매우 어렵기 때문에 본 논문에서는 학습 과정에서 가상으로 생성한 200,000장의 영상 데이터를 사용하였다. 테스트 과정에서는 실제 데이터를 수집하여 테스트하여 학습된 모델의 실효성을 검증한다. 학습에 사용된 저해상도 영상( $I_{LR}$ )은 고해상도 영상( $I_{HR}$ ) 표본을  $r$  복원 배율로 축소하여 얻는다. 즉, 고해상도 영상( $I_{HR}$ )의 크기가  $W \times H \times C$ 일 때, 저해상도 영상( $I_{LR}$ )은  $rW \times rH \times C$ 가 된다[3]. 본 논문에서는  $r$  복원배율(0.25)을 사용하였다. 고해상도 영상( $I_{HR}$ )은 학습 과정에서만 필요하며, 테스트 과정에서는 필요하지 않다.

#### 1. 학습 데이터 생성

생성된 가상 데이터는 다양한 각도, 빛, 노이즈에 대응할 수 있도록 생성한다. 생성 방법은 먼저 국가법령정보센터에서 제공하는 번호판의 숫자와 문자 샘플 영상을 이용하여 무작위로 6개의 숫자와 1개의 문자를 이용해 가상의 번호판 영상을 생성한다. 그 후 다양한 각도, 크기와 주변 밝기에 대응하기 위해 호모그래피(Homo-

graphy) 변환과 밝기 변화를 적용한다. 생성된 가상의 번호판 영상은 SUN database[26]로 만든 64×128 크기의 배경 영상에 합성된다. 저해상도 영상은 지역 보간법(Area interpolation)을 이용하여 16×32 크기의 영상을 획득한다. 그 후 표준편차가 0.1인 가우시안 노이즈를 저해상도 영상에 추가한다. 노이즈는 복원된 영상의 노이즈 제거 효과를 위해 추가한다[27].

## 2. 딥러닝 기반 자동차 번호판 인식기

본 논문에서 딥러닝 기반 자동차 번호판 인식기의 인식률을 통해 제안한 기법의 실효성을 검증한다. 자동차 번호판 인식기는 기존의 딥러닝 기반 자동차 번호판 인식기[28]를 수정하였다. 기존의 딥러닝 기반 자동차 번호판 인식기의 경우 영상에서 자동차 번호판을 검출한 후 번호판의 숫자와 문자를 인식한다. 그러나 본 논문에서 사용된 인식기의 목적은 복원된 영상에서 번호판의 숫자와 문자를 인식하여 제안한 방법의 실효성을 평가하기 위함이다. 따라서 실제 영상에서의 번호판 영역의 검출 과정을 생략한 모델을 구성한다. 따라서 인식기의 입력 데이터는 추출된 자동차 번호판이며 16×32 크기의 영상이다. 인식기는 레지듀얼 보틀넥 신경망(Residual bottleneck network)을 이용하여 학습 모델(그림 4)을 구성하였다. 그림 3에서 s는 컨볼루션의 거리를 뜻하고, n1과 n2는 각 층에서 컨볼루션의 깊이를 뜻한다. 학습 데이터는 250,000장의 가상 번호판 영상으로 학습하였다. 이때, 학습 최적화를 위한 최적화 함수는 Adam( $\beta=0.9$ , [29])을 이용하였다. 초기 학습률(learning rate)은 0.001로 하였고, 1 epoch(300k iteration)마다 절반으로 감소시켰다. 학습 시 배치(batch)의 크기는 16으로 정하고, 총 5 epoch 반복 학습하였다.

## 3. 초해상도 기법의 파라미터 설정

생성자와 구별자 신경망 학습에 필요한 하이퍼파라미터(hyperparameter)는 반복 학습을 통한 경험에 의해 결정하였다. 초해상도 기법의 학습 최적화를 위한 최적화 함수로 Adam( $\beta=0.9$ )을 이용하였다. 초기 학습률(learning

rate)은 0.002로 하였고, 3000회 반복할 때마다 절반으로 감소시켰다. 학습 시 배치(batch)의 크기는 16으로 하였고, 50k 반복 학습하였다.

## V. 실험 결과

제안한 학습 방법의 실효성 검증을 위해 학습에 이용되지 않은 가상의 자동차 번호판 영상과 실제 자동차 번호판 영상을 이용하여 기존의 딥러닝 기반 초해상도 기법과 본 연구에서 제안한 방법의 정량적 및 정성적 성능을 비교 분석한다.

### 1. 손실 함수에 따른 학습 결과 비교

본 논문에서는 다양한 손실 함수에 따른 초해상도 기법의 성능을 비교한 후 가장 좋은 결과를 보인 손실 함수를 선택하였다. 테스트는 학습에 사용되지 않은 4,000장의 저해상도 영상을 이용하였다. 손실 함수에 따른 제안한 기법의 양적 결과는 표 1과 같으며, 그림 5는 임의의 영상에 대한 정성적 비교 실험 결과를 나타낸다. 그림 5에서  $I'_{HR-MSE}$ 는 픽셀 단위 손실 함수만 최적화하는 방법으로 학습한 결과이다.  $I'_{HR-VGG2.2}$  [12]는 VGG 신경망의  $\phi_{2.2}$ 을 이용하여 깊은 층에서 영상이 가진 특징맵들에 대한 내용 손실 함수만 최적화하는 방법으로 학습한 결과이다[28]. 표 1은 PSNR(Peak signal-to-noise ratio)[30]과 4.2장에서 설명한 자동차 번호판 인식기 모델의 정확도로써 기존의 딥러닝 기반 초해상도 기법과 제안한 방법의 양적 결과를 비교하였다. 이때 정확도는 번호판 숫자에 대한 정확도, 문자에 대한 정확도, 그리고 전체 정확도(숫자와 문자 정확도 평균)로 표시하였다. 실험 결과 픽셀 단위 손실 함수와 내용 손실 함수 개별적으로 사용하는 것보다 제안한 방법이 PSNR과 자동차 번호판 인식기의 인식 정확도에서 가장 좋은 결과를 얻었다.

### 2. 기존의 초해상도 기법과 비교 결과

기존의 초해상도 기법인 바이큐빅 보간법, SRCNN[12], SRResnet[3]과 제안한 방법의 성능을 비교하였다. Ledig et al. [3]에 의하면 초해상도 생성적 적대 신경망(SRGAN)보

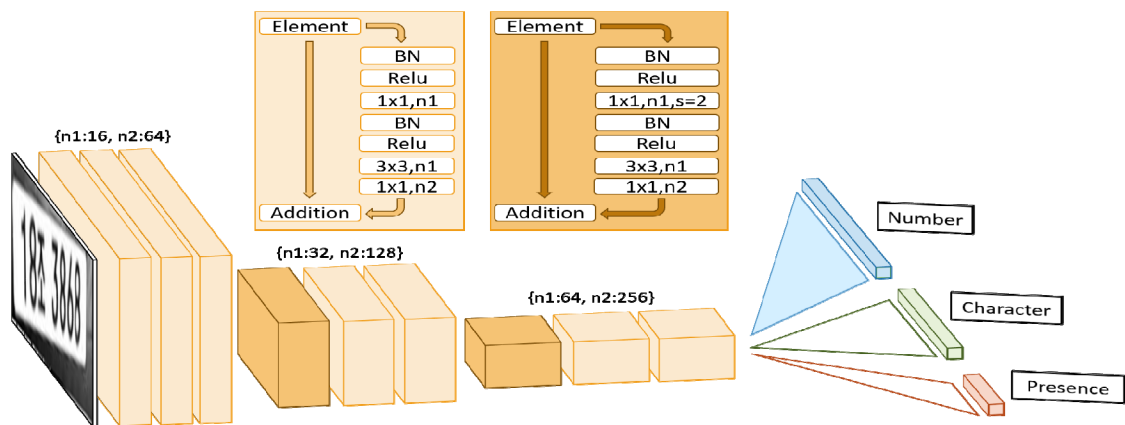


그림 4. 자동차 번호판 인식기 신경망의 구조. 1×1, 3×3은 컨볼루션의 커널의 크기를 의미.

Fig. 4. Architecture of license plate recognition network. 1×1, 3×3 represent kernel size of convolution.



표 1. 가상 번호판 영상을 이용한 바이큐빅 보간법, 픽셀 단위 손실 함수, 내용 손실 함수, 제안한 방법의 양적 결과 비교.

Table 1. Comparison of quantitative results of bicubic interpolation, pixel-wise loss, content loss, proposed method using virtual license plate.

	$I_{HR}$	$I'_{HR \sim Bicubic}$	$I'_{HR \sim MSE}$	$I'_{HR \sim VGG2,2} [12]$	$I'_{HR \sim Ours}$
PSNR [dB]	$\infty$	17.5829	21.7745	21.1365	<b>23.7035</b>
Digital Accuracy [%]	98.7	45.1	79.2	83.7	<b>89.3</b>
Character Accuracy [%]	96.6	31.8	43.5	51.1	<b>58</b>
Total Accuracy [%]	97.65	38.45	61.35	67.4	<b>73.65</b>



그림 5. 손실 함수 요소에 따른 초해상도 기법 결과 비교. 왼쪽에서부터 원본,  $16 \times 32$  저해상도 영상, 바이큐빅 보간법, 픽셀 단위 손실 함수, 내용 손실 함수, 제안한 방법에 따른 결과.

Fig. 5. Comparison of the reconstruction results with respect to different elements of loss function. From the leftmost, original,  $16 \times 32$  input, bicubic interpolation, pixel-wise loss, content loss, proposed method.

표 2. 가상 번호판 영상을 이용한 기존의 초해상도 기법과 제안한 방법의 양적 결과 비교.

Table 2. Comparison of quantitative results of state-of-art method and proposed method using virtual license plate.

	$I_{HR}$	$I'_{HR \sim Bicubic}$	$I'_{HR \sim SRCNN} [12]$	$I'_{HR \sim SRResnet} [3]$	$I'_{HR \sim Ours}$
PSNR [dB]	$\infty$	17.5829	21.1426	23.5083	<b>23.7035</b>
Digital Accuracy [%]	98.7	45.1	63.6	88.5	<b>89.3</b>
Character Accuracy [%]	96.6	31.8	27.6	<b>58.7</b>	<b>58</b>
Total Accuracy [%]	97.65	38.45	91.2	73.6	<b>73.65</b>

다 Resnet을 이용한 초해상도 기법인 SRResnet을 이용하였을 때 PSNR가 더 높고 실제 자동차 인식기의 성능 또한 좋았기 때문에 SRResnet을 비교하였다. 그림 6은 임의의 영상에 대한 비교 실험 결과를 나타낸다. 여기서,  $I'_{HR \sim SRCNN}$ 은 딥러닝 기반 초해상도 기법 중 일반적으로 쓰이는 CNN기반 초해상도 복원 결과이며,  $I'_{HR \sim SRResnet}$ 은 Resnet을 이용하여 초해상도 복원한 결과이다. 표 2는 각 기법의 PSNR과 인식 정확도에 대한 정량적 결과이다. SRresnet을 사용하여 복원한 경우 문자 정확도에서만 58.7%로 가장 높게 나왔지만, 제안한 방법으로 복원하였을 때 가장 높은 PSNR과 평균 정확도가 가장 많이 향상되었음을 확인할 수 있다.

마지막으로 제안한 방법과 자동차 인식기의 실효성 검증 위해 실제 자동차 번호판 영상 250장을 이용해 테스트하였다. 검증 단계에서 실제 자동차 번호판 영상에서 번호판 부분(그림 7의 빨간 박스)만 추출하여 검증에 사용하였다. 그림 7과 표 3은 실제 번호판을 가지고 검증한 정성적 및 정량적 결과이다. 이를 통해 제안한 방

법은 가상 데이터로 학습하여도 실제 번호판 영상 복원 대한 실효성이 있음을 알 수 있다.

## VI. 결론

본 논문에서는 저해상도 자동차 번호판 영상의 인식률을 높이기 위해 생성적 적대 신경망을 이용한 초해상도 기법을 제안하였다. 제안한 모델은 Ledig et al. [3]이 제안한 초해상도 생성적 적대 신경망과 달리 내용 손실 함수에 픽셀 단위 손실 함수를 추가하고, 자동차 번호판 영상에 최적화되도록 모델을 구성하였다. 생성자 신경망은 Resnet 기반 신경망으로 구성하였고, 구별자 신경망은 VGG 기반의 신경망으로 구성하였다. 또한 학습 과정에서 한쪽 신경망의 성능에 치우쳐 지지 않고 균형이 이루어지도록 실험을 통한 경험에 근거하여 손실 함수 모델을 구성하였다. 모델 학습은 가상으로 생성한 번호판 데이터를 이용하였다. 실험 결과, 가상 번호판 영상을 이용했을 경우 번호판 인식기의 저해상도 열화 영상 인식률이 38.45%에서 제안한 방법을 이용한 경우 73.65%로 향

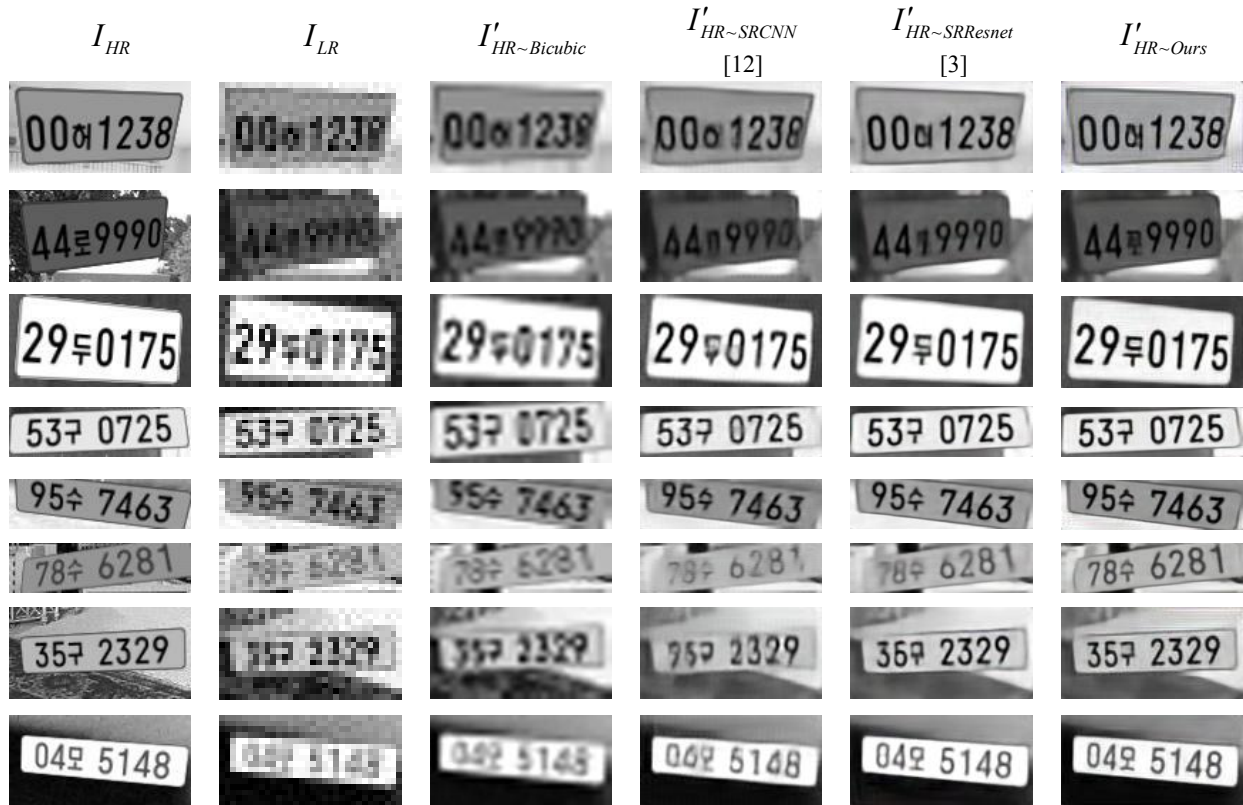


그림 6. 기존의 초해상도 기법 및 제안한 방법을 이용한 저해상도 영상 복원 결과 비교. 왼쪽부터 원본,  $16 \times 32$  저해상도 영상(복원 대상), 바이큐빅 보간법, SRCNN, SRGAN, 제안한 방법.

Fig. 6. Comparison of the reconstruction results by using state-of-the-art super-resolution method and proposed method. From the leftmost, original,  $16 \times 32$  input, bicubic interpolation, SRCNN, SRResnet, proposed method.



그림 7. 실제 저해상도 자동차 번호판을 이용한 복원 결과 샘플.

Fig. 7. Examples of reconstruction results using actual low-resolution license plate.

표 3. 실제 저해상도 영상의 번호판 복원 결과.

Table 3. Quantitative results of reconstructed image using actual low-resolution license plate.

	$I'_{HR \sim Bicubic}$	$I'_{HR \sim Ours}$
Digital Accuracy [%]	54.7	<b>97.3</b>
Chracter Accuracy [%]	50.9	<b>79.3</b>
Total Accuracy [%]	52.8	<b>88.3</b>

상되었다. 또한, 실제 번호판 영상을 이용했을 경우 저해상도 열화 영상 인식률이 52.8%에서 88.3%로 향상되었음을 확인하였다. 위 결과를 바탕으로 가상의 데이터만으로 학습을 했음에도 충분히 실제 영상에 적용 가능하며, 저해상도 자동차 번호판 영상 복원에 있어서 기존의 딥러닝 기반 초해상도 기법을 사용했을 때보다 제안한 방법의 성능이 더 뛰어남을 보였다.

그러나 문자 인식 정확도는 제안하는 방법의 경우 58%로 숫자 인식 정확도(89.3%)에 비해 상대적으로 낮은 결과를 얻었다. 그 이유는 저해상도 영상이 가진 정보의 부족으로 인해 생기는 한계로 사료된다. 예를 들어, 저해상도 영상에서 문자를 형성하는 픽셀의 화소수가 매우 작다. 때문에 “o”, “口”, “日” 과 같이 비슷한 외형의 문자를 구별할 만한 충분한 정보를 가지고 있기 않기 때문



이다. 따라서 복원 과정에서도 저해상도 영상에서 서로 비슷한 외형을 가진 픽셀의 경우 뚜렷한 복원에 실패하였다. 이로 인해 숫자보다 상대적으로 작은 픽셀의 화소수를 가진 문자의 경우 낮은 인식 정확도를 보였다. 저해상도 영상에서 서로 비슷한 외형을 띤 문자를 구분하는 문제를 해결한다면 문자 인식의 정확도를 향상시킬 수 있을 것으로 사료된다.

## REFERENCES

- [1] P. Comelli, P. Ferragina, M. Notturmo Granieri, and F. Stabile, "Optical recognition of motor vehicle license plates," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 4, pp. 790-799, 1995.
- [2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.
- [3] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, and Z. Wang, "Photo-realistic single image super-resolution using a generative adversarial network," *arXiv preprint arXiv:1609.04802*, 2016.
- [4] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," *arXiv preprint arXiv:1511.05440*, 2015.
- [5] C. Yang, C. Ma, and M. Yang, "Single-image super-resolution: A benchmark," *European Conference on Computer Vision*, pp. 372-386. Springer International Publishing, 2014.
- [6] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231-239, 1991.
- [7] K. V. Suresh, G. M. Kumar, and A. N. Rajagopalan, "Superresolution of license plates in real traffic videos," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 321-331, 2007.
- [8] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327-1344, 2004.
- [9] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 2, pp. 12, 2011.
- [10] K. I. Kim and Y. H. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 32, no. 6, pp. 1127-1133, 2010.
- [11] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 370-378, 2015.
- [12] C. Dong, C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," *European Conference on Computer Vision*, pp. 184-199, Springer International Publishing, 2014.
- [13] J. W. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1646-1654, 2016.
- [14] J. H. Jo and D. J. Kang, "An ensemble classifier based method to select optimal image features for license plate recognition," *The Transactions of the Korean Institute of Electrical Engineers (in Korean)*, vol. 65, no. 1, pp. 142-149, 2016.
- [15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, pp. 2672-2680, 2014.
- [16] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
- [17] T. S. Kim, M. S. Cha, H. S. Kim, J. W. Lee, and J. W. Kim, "Learning to discover cross-domain relations with generative adversarial networks," *arXiv preprint arXiv:1703.05192*, 2017.
- [18] D. Berthelot, T. Schumm, and L. Metz, "BEGAN: boundary equilibrium generative adversarial networks," *arXiv preprint arXiv:1703.10717*, 2017.
- [19] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [20] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 807-814, 2010.
- [21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [22] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," *Proc. ICML*, vol. 30, no. 1, 2013.
- [23] J. Johnson, A. Alahi, and F. Li, "Perceptual losses for real-time style transfer and super-resolution," *European Conference on Computer Vision*, pp. 694-711. Springer International Publishing, 2016.
- [24] 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, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211-252, 2015.
- [25] M. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," *European Conference on Computer Vision*, pp. 818-833. Springer International Publishing, 2014.
- [26] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba, "Sun database: Large-scale scene recognition from abbey to zoo," *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, pp. 3485-3492, IEEE, 2010.
- [27] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, pp. 3371-3408, 11. Dec, 2010.
- [28] H. H. Kim, J. K. Park, J. H. Oh, and D. J. Kang, "Multi-task convolutional neural network system for license plate recognition," *International Journal of Control, Automation and Systems*, Accepted.
- [29] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [30] H. Quan and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics letters* 44, no. 13, pp. 800-801, 2008.
- [31] Y. K. Park, J. K. Park, H. I. On, and D. J. Kang, "Convolutional Neural Network-based System for Vehicle Front-Side Detection," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 21, no. 11, pp. 1008-1016, Nov. 2015.
- [32] S. M. Kang, K. M. Huh, and Y. B. Joo, "Development of

Sleepy Status Monitoring System using the Histogram and Edge Information of Eyes,” *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 22, no. 5, pp. 361-366, 2016

- [33] J. M. Choi and D. J. Kang, “Super-resolution for license plate using generative adversarial networks,” *Proc. of 2017 32<sup>nd</sup> ICROS Annual Conference (in Korean)*, Gangwon-do Sokcho, Korea, May 2017.



#### 최 준 명

2016년 부산대학교 조선해양공학과 졸업.  
2017년~ 현재 부산대학교 대학원 기계공학부 제어자동화시스템 전공 석사과정 재학중. 관심분야는 딥러닝, 머신러닝.

#### 강 동 중

제어·로봇·시스템학회 논문지, 제 17 권 제 1 호 참조.