Journal of KIIT. Vol. 23, No. 5, pp. 1-12, May 31, 2025. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2025.23.5.01

학습 영상 생성 강화를 이용한 악천후 영상의 시인성 개선

이세완*, 이성학**

Image Visibility Enhancement under Bad Weather with Intensified Generative Module for Train Set

Se-Wan Lee*, Sung-Hak Lee**

이 논문은 2024학년도 경북대학교 연구년 교수 연구비에 의하여 연구되었음

요 약

Image-to-Image translation은 이미지를 입력으로 받아 새로운 이미지로 변환하는 기술이다. 그중 딥러닝을 이용한 이미지 변환 학습은 과적합 현상을 방지하기 위해 많은 학습 데이터가 필요하므로, 본 연구에서는 우천 상황 등 악천후 영상의 개선 학습을 위해 CycleGAN(Cycle-Consistent Aversarial Network) 모델과 CNN(Convolutional Neural Network) 모델을 통해 가상의 물방울 이미지를 생성 및 선별하여 학습 데이터를 효율적으로 확보하는 방법과 이를 활용한 시인성 개선 방법을 제안한다. 선별한 이미지를 기존 학습 데이터 셋에 추가하여 증식된 데이터 셋으로 학습하면 기존 학습에 비해 향상된 물방울 제거 능력을 확인할 수 있다. 또한 톤 매핑이 적용된 타겟 이미지로 모듈을 추가적으로 학습하여 2단계의 개선 모듈을 적용하였고, 결과적으로 물방울 제거뿐만 아니라 디테일 손실 및 명암비를 개선하는 방법을 제안한다.

Abstract

Image-to-image translation is a technique that takes an image as input and transforms it into a new image. Since deep learning-based image translation requires a large amount of training data to prevent overfitting, this study proposes a method to efficiently secure training data by generating and selecting fake water-droplet images using Cycle-Consistent Adversarial Network (CycleGAN) and Convolutional Neural Network (CNN) models for the image enhancement under bad weather conditions. Additionally, we present an approach to improve image visibility using the augmented data. By adding the selected images to the existing dataset and training with the augmented dataset, improved water-droplet removal performance can be observed compared to the original training. This paper proposes an additional training step with tone-mapped target images. This approach not only enhances water-droplet removal but also improves detail preservation and contrast ratio.

Keywords

image deep learning, CycleGAN, CNN, data augmentation, water-droplet removal, tone mapping

* 경북대학교 IT대학 전자공학부 학부과정
- ORCID: https://orcid.org/0009-0007-1244-0168
** 경북대학교 IT대학 전자공학부 교수(교신저자)
- ORCID: https://orcid.org/0000-0002-1030-381X

· Received: Feb. 25, 2025, Revised: Apr. 01, 2025, Accepted: Apr. 04, 2025
· Corresponding Author: Sung-Hak Lee
School of Electronics Engineering, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu, 41566, Korea
Tel.: +82-53-950-7216, Email: shak2@ee.knu.ac.kr

Ⅰ. 서 론

인공지능 기술의 발달로 컴퓨터 비전 기반의 자율주행차 개발이 활발해지고 있다. 자율주행 기술은 차량이 주변 환경을 스스로 인식하고 주행할 수 있는, 미래 교통 시스템을 혁신적으로 변화시킬 수 있는 기술이지만, 우천 시 전방 카메라 렌즈에 맺히는 물방울은 영상의 선명도를 크게 저하시키며, 보행자, 차량, 도로 표지판과 같은 객체를 명확히 인식하는 데 어려움을 초래한다. 이러한 가시성 문제는 객체 인식 성능에 부정적인 영향을 미쳐 주행 중 사고 위험을 증가시킬 수 있으며, 이를 개선하기 위한 기술 개발은 자율주행의 안전성과 신뢰성을 높이는 데 필수적이다[1].

딥러닝을 이용한 Image-to-Image Translation은 특정 영역의 영상을 목표 영역의 영상으로 변환하는 기술이다. 특히 GAN(Generative Adversarial Network)은 경쟁적 학습을 통해 양질의 이미지를 생성할 수 있어 Image-to-Image Translation에 적극적으로 사용된다[2][3]. GAN 기반의 대표적인 모델 Pix2pix는 학습을 위해 입력 이미지와 대응 이미지가 쌍을 이루는 데이터 셋이 필요하고 많은 양의 데이터를 필요로 하기 때문에 활용에 한계가 있다. 이를 극복하기 위해 CycleGAN(Cycle-Consistent Aversarial Network)이 도입되었다[4].

물방울 제거를 위한 영상 학습을 언페어(Unpair) 데이터 셋으로 학습을 진행하는 CycleGAN은 페어(Pair) 데이터 셋으로 학습을 진행하는 인공지능 모델에 비해 데이터 셋의 확보가 용이하다는 장점이 있다[4]. 그러나 그림 1(b)에서처럼 언페어 데이터 셋으로 학습을 진행한 CycleGAN은 변환 결과 영상에서 영상의 색이나 밝기에 왜곡이 생기고 충분한 물방울 제거 효과가 나타나지 않는 단점을 보인다. 이는 페어 영상 셋이 아닌 언페어 영상 셋을 학습 데이터 셋으로 사용하기 때문에 요구되는 특징을 지정하여 변환하기 어려워지고 상대적으로 물방울 제거에 있어 한계가 발생된다. 반면에 그림 1(c)에서처럼 페어 데이터 셋으로 학습한 CycleGAN은 변환 결과 영상에서 물방울 특징을 잘 검출하여 영상의 색이나 밝기에 언페어 학습과 비교하여 왜곡 없이 변환하는 것을 확인할 수 있다.

(a) (b) (c)

그림 1. 물방울 입력 및 CycleGAN을 이용한 제거 결과
(a) 물방울 원본 영상, (b) Unpaired CycleGAN 물방울 제거 결과, (c) Paired CycleGAN 물방울 제거 결과
Fig. 1. Water droplet input and results of water droplet removal using CycleGAN (a) Original water droplet image, (b) Result of water droplet removal using unpaired CycleGAN, (c) Result of water droplet removal using paired CycleGAN

페어 데이터 셋으로 CycleGAN 학습을 할 때, 데이터 셋이 충분하지 않은 경우 부족한 학습 데이터에 지나치게 최적화되어 테스트 데이터에 일반화하지 못하는 과적합 현상이 발생할 수 있기 때문에 충분한 페어 데이터 셋 확보가 필요하다[5]. 또한 데이터 셋을 보충하면 모델이 더 다양한 패턴을 학습할 수 있기 때문에, 물방울의 다양한 특성을 인식하고 효과적으로 제거하여 시인성을 개선할 수 있다. 따라서 본 연구에서는 우선 데이터 셋 보충을 위해 가상의 물방울 생성 모듈을 학습하였고, 분류 모델인 CNN(Convolutional Neural Network)을 사용하여 선별된 양질의 데이터 셋만이 기존 데이터 셋에 추가되어 학습을 위한 증식된 데이터 셋을 구성하게 된다.

물방울 제거 학습을 하기에 앞서, CycleGAN 모델의 예측 값과 실제 값의 차이를 계산하는 손실(Loss) 함수 중 L1 Loss 관련 함수들의 입력과 타겟 부분을 수정하여 물방울 영역에 초점을 맞춘 학습이 이루어지도록 하였다. L1 Loss가 수정된 CycleGAN 모델은 물방울을 중점으로 학습이 더욱 가속화되어 기존 학습보다 epoch가 증가함에 따라 Loss가 더 빠르게 감소하고, 물방울이 없는 배경에서의 왜곡이 개선된다[6].

본 논문에서는 CycleGAN 모델을 기반으로 하여 학습한 2가지 모듈을 적용함으로써 물방울이 제거된 영상을 복원하여 시인성을 개선하는 방법을 제안한다. 첫 번째 모듈은 앞서 언급한 증식된 데이터 셋을 L1 Loss가 수정된 CycleGAN 모델로 학습하여 효율적인 물방울 제거에 초점이 맞춰져있다.

두 번째 모듈은 동일하게 증식된 데이터 셋을 활용하지만 타겟 이미지에 톤 매핑을 적용하여, 추가적인 물방울 제거와 톤 매핑에 초점이 맞춰져있다. 두 가지 모듈을 순차적으로 적용한 결과 영상은 단순히 기존 페어 데이터 셋으로 학습한 모듈의 결과 영상과 비교했을 때 영상의 물방울 부분을 더 잘 제거함과 동시에, 디테일 손실과 명암비가 개선되는 것을 확인할 수 있다.

II. 관련 연구

2.1 딥러닝 변환 학습

Image-to-image translation 학습은 입력 이미지와 출력 이미지 간의 특징적 변환 매핑을 학습하는 것을 목표로 하는 컴퓨터 비전의 한 분야이다. 이 기술은 정렬된 이미지 쌍으로 구성된 페어 데이터 셋을 이용하여 수행되는데, 효과적인 학습을 위해 충분한 페어 데이터 셋을 구할 수 어려운 경우가 많다. 이를 극복하기 위해서 CycleGAN 기법이 도입되었다. CycleGAN은 생성자(Generator) 2개, 판별자(Discriminator) 2개로 구성되어 있다[4]. 생성자 신경망은 ResNet 구조이며, 판별자 신경망은 PatchGAN 구조 기반이다. ResNet은 입력 값과 출력 값 간의 차이를 학습하는 방식인 잔차 학습(Residual learning)을 사용하는 구조로, 이미지 생성에서 원본 이미지와 생성 이미지 간의 차이를 학습하는 데 적합하며, 스킵 연결(Skip connections)을 사용하여 깊은 신경망에서도 기울기 손실 문제를 완화한다[7]. PatchGAN은 전체 영역이 아닌 특정 크기의 패치(Patch) 단위로 생성자가 생성한 이미지의 진위 여부를 판단한다.

입력된 이미지는 평가를 위해 판별자로 이동하거나 다른 도메인으로 변환되어 다른 판별자가 평가하고 다시 변환되어 돌아온다. 이러한 과정을 또 다른 네트워크에서 반대로 진행하도록 설계되어 있다[8]. 식 (1)은 GAN에서 일반적으로 사용되는 적대적 손실(Adversarial loss)을 나타낸다.

$$L_{GAN}(G,D_Y,X,Y) = E_{y\sim p_{data}(y)}[\log D_Y(y)] + E_{x\sim p_{data}(x)}[\log(1-D_Y(G(x)))] \quad (1)$$

여기서 G는 $X\rightarrow Y$ 변환을 위한 매핑 함수이다. D_Y는 Y 도메인의 판별기이며 E는 기댓값(Expected value)을 의미한다. 또한 $y\sim p_{data}(y)$는 real 데이터의 분포, $x\sim p_{data}(x)$는 fake 데이터의 분포이다. 적대적 손실의 첫 번째 항은 도메인 Y의 각 real data y를 판별자에 넣었을 때 나오는 결과를 log 취했을 때 얻는 기댓값이다. 두 번째 항은 도메인 X의 각 fake data x를 생성자에 넣었을 때 나오는 결과를 판별자에 넣은 결과를 $\log(1-$결과$)$를 취했을 때 얻는 기댓값이다.

식 (2)는 기존의 GAN과 차별화된 CycleGAN의 대표적인 특징인 순환 일관성 손실(Cycle consistency loss)을 나타낸다.

$$L_{cyc}(G,F) = E_{x\sim p_{data}(x)}[\|F(G(x))-xvert|_1] + E_{y\sim p_{data}(y)}[\|G(F(y))-yvert|_1] \quad (2)$$

여기서 F는 $Y\rightarrow X$ 변환을 위한 매핑 함수이며 D_X는 X 도메인의 판별기이다. 순환 일관성 손실은 적대적 손실의 매핑 함수 G와 F를 거쳐서 순환을 할 때 영상 변환 사이클이 원래 영상으로 되돌릴 수 있도록 한다. 첫 번째 항은 도메인 X의 각 영상 x에 대해 매핑 함수 G에 대입한 영상 $G(x)$를 다시 역방향으로 매핑 함수 F에 대입하여 원래 영상 x와의 차이를 절댓값으로 나타내었을 때 얻는 기댓값이다. 두 번째 항은 도메인 Y의 각 영상 y에 대해 매핑 함수 F에 대입한 영상 $F(y)$를 다시 역방향으로 매핑 함수 G에 대입하여 원래 영상 y와의 차이를 절댓값으로 나타내었을 때 얻는 기댓값이다.

식 (3)은 적대적 손실과 순환 일관성 손실을 결합한 CycleGAN의 최종적인 Loss를 나타낸다.

$$L(G,F,D_X,D_Y) = L_{GAN}(G,D_Y,X,Y) + L_{GAN}(F,D_X,Y,X) + \lambda L_{cyc}(G,F) \quad (3)$$

여기서 λ는 순환 일관성 손실에 대한 가중치를 나

타낸다. 매핑 함수 G에 대한 적대적 손실과 매핑 함수 F에 대한 적대적 손실, 그리고 순환 일관성 손실을 가중치 λ를 통해 결합한 최종적인 CycleGAN Loss를 나타내었다.

또한, image-to-image-translation의 경우, 입력과 출력 간의 색상 구성을 유지하도록 매핑하기 위해 Identity Loss가 추가적으로 도입되었다. 식 (4)는 Identity Loss를 나타낸다.

$$L_{identity}(G,F) = E_{y\sim p_{data}(y)}[\|G(y)-yvert|_1] + E_{x\sim p_{data}(x)}[\|F(x)-xvert|_1] \quad (4)$$

첫 번째 항은 매핑 함수 G가 도메인 Y의 실제 데이터 y를 입력으로 받을 때 출력 $G(y)$와 원래 입력 y 간의 차이를 측정한다. 이 항의 목적은 매핑 함수 G가 타겟 도메인의 실제 이미지를 그대로 출력하도록 유도하여, 입력 이미지의 색상과 특성이 변하지 않도록 보장하는 것이다. 두 번째 항은 매핑 함수 F가 도메인 X의 실제 데이터 x를 입력으로 받을 때 출력 $F(x)$와 원래 입력 x간의 차이를 측정한다. 이 항의 목적도 첫 번째 항과 유사하다.

2.2 영상 분류 학습

CNN 학습 모델은 영상을 분석하기 위해 패턴을 찾는데 유용한 알고리즘이다[9]. CNN은 이미지 등의 시각 데이터를 처리하는데 특화된 신경망 구조로서, 1980년대 인공신경망으로 처음 제안된 이후 2012년 AlexNet 모델은 딥러닝 학습을 통해 대규모 이미지 분류에서 탁월한 성능을 발휘함을 입증하며 딥러닝 산업에 큰 영향력을 주었다[10]. 따라서 현재까지도 관련된 논문이 지속적으로 인용되어 연구되고 있으며, 최근에는 기존의 활성화 함수 시그모이드(Sigmoid)나 하이퍼볼릭탄젠트(Tanh)가 아닌 렐루(ReLU)를 채택하여 기울기 소멸 문제를 해결하고 학습 속도를 가속화하고 있다[11].

CNN은 특징 추출을 위한 합성곱 층(Convolutional layer)과 풀링 층(Pooling layer)으로 구성된다. CNN에 입력되는 데이터는 높이, 너비, 채널의 3차원 텐서(Tensor)로 구성되어 있고 행렬(Matrix) 형태로 나타난다. 합성곱 층에서는 입력 데이터와 필터의 합성곱 연산을 채널마다 수행한다. 이때 영상의 채널 개수만큼 필터가 존재하기 때문에 합성곱의 결과는 채널, 즉 필터 수만큼 나오게 된다. 각 채널의 결과를 모두 더해서 하나의 출력을 얻게 되는데, 이를 특징 맵(Feature map)이라고 한다. 이때 필터의 간격은 스트라이드(Stride)를 지정함으로써 변경할 수 있다. 합성곱 과정에서 영상의 크기는 점점 작아지게 되는데, 이는 특정 픽셀에 대한 정보가 사라짐을 의미한다. 이러한 문제를 해결하기 위해 패딩(Padding)을 사용한다. 패딩은 입력 영상의 가장자리에 일정한 크기의 테두리를 추가하는 것인데, 0으로 채워진 픽셀을 사용하는 제로 패딩(Zero padding) 방식이 보편적이다. 패딩을 적용함으로써 얻을 수 있는 이점은, 입력 영상과 출력 영상의 크기를 비슷하게 조정하여 영상의 가장자리에 해당하는 정보를 잘 보존할 수 있다는 것이다.

풀링 층에서는 합성곱 층의 출력, 즉 특징 맵을 입력으로 받아 다운샘플링(Downsampling)하여 특징 맵의 크기를 줄이는 풀링 연산이 이루어진다. 풀링 연산은 합성곱 연산과 다르게 학습해야 할 가중치가 없으며, 연산 후에 채널 수가 변하지 않는다.

풀링 층의 결과 데이터는 완전 연결층(Fully connected layer)으로 전달되어야 한다. 하지만 풀링 층을 거친 특징 맵은 여전히 3차원 형태를 가지고 있다. 따라서 이를 1차원 벡터로 변환하는 평탄화(Flatten) 과정을 거쳐서 완전 연결층에 입력으로 제공한다. 완전 연결층에서는 모든 입력 뉴런과 출력 뉴런이 연결되고 입력 데이터의 모든 특징을 종합적으로 고려하여 복잡한 패턴 및 관계를 학습하게 된다.

완전 연결층의 결과 데이터는 출력층으로 전달되어 소프트맥스(Softmax) 함수 또는 시그모이드(Sigmoid) 함수와 같은 활성화 함수를 적용하게 된다.

III. 제안 방법

그림 2는 본 논문에서 제안하는 효과적인 물방울 제거를 위한 알고리즘을 학습 데이터 생성과 시인성 개선으로 나누어 구성한 흐름도를 나타낸다.

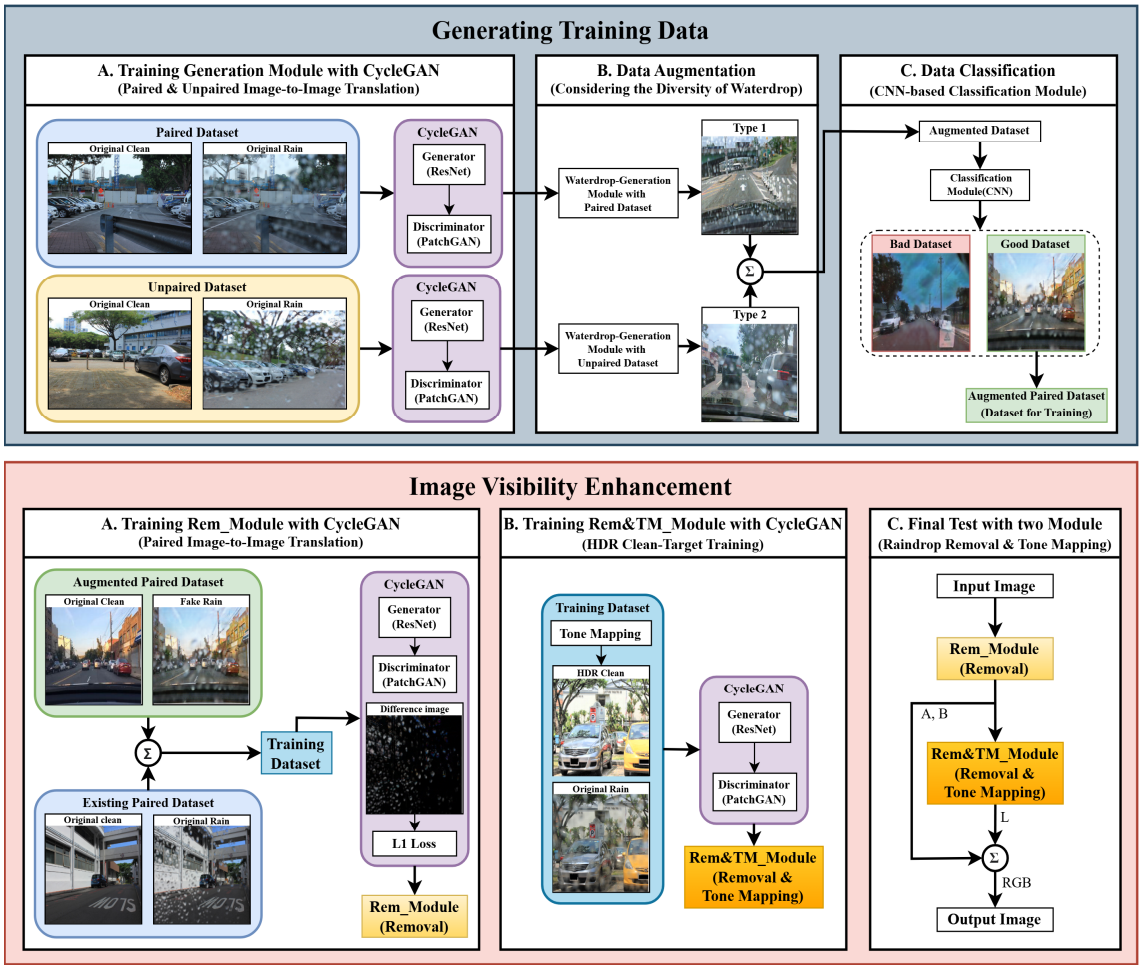


그림 2. 제안된 알고리즘의 흐름도
Fig. 2. Flowchart of the proposed method

흐름도는 A부터 F까지 순서대로 총 6개의 단계로 구성되어 있다. A는 물방울 가상 생성 모듈 학습, B, C는 데이터 증식을 효율적으로 하기 위한 방법, D는 물방울 제거 모듈(Rem_module) 학습, E, F는 물방울 제거 및 톤 매핑 모듈(Rem&TM_module) 학습에 대한 단계이다.

3.1 학습 데이터 생성

3.1.1 CycleGAN 기반 물방울 가상 생성 모듈 학습

앞서 설명했듯이, 페어 데이터 셋으로 CycleGAN 학습을 하기 위해서는 충분한 페어 데이터 셋 확보가 중요하다. 본 학습에서의 페어 데이터 셋은 물방울이 없는 깨끗한 영상과, 동일 배경에서 물방울이 맺혀있는 물방울 영상으로 이루어져 있다. 그러나 동일한 배경에서 페어 데이터 셋을 확보하는 것은 현실적으로 어렵다. 따라서 본 연구에서는 학습을 위한 기존의 부족한 페어 데이터 셋을 가상의 데이터 셋으로 보충하기 위해서, 흐름도의 첫 번째 블록에서 가상의 물방울 생성 모듈을 학습하였다. 기존에 보유한 800개의 페어 데이터 셋과 400개의 언페어 데이터 셋을 각각 CycleGAN에 입력으로 제공하여 서로 다른 두 가지 물방울 생성 학습을 하였다. 이때 학습 epoch는 동일하게 300으로 설정하였다.

3.1.2 데이터 증식

두 번째 블록에서, 이전 단계에서 학습한 2가지 생성 모듈로 물방울이 없는 깨끗한 영상에 가상의 물방울을 생성하였다. 그림 3에서 그 결과를 나타내었다.

그림 3(a)는 800개의 페어 데이터 셋으로 학습한 모듈의 가상 물방울 생성 결과로, 물방울이 작고 뚜렷하며 밀도가 높은 것을 확인할 수 있다. 반면에 그림 3(b)는 400개의 언페어 영상 셋으로 학습한 모듈의 가상 물방울 생성 결과로, 물방울이 크고 투명하며 밀도가 낮은 것을 확인할 수 있다.

그림 3에서 볼 수 있듯이, 페어 데이터 셋의 경우 입력과 GT(Ground Truth)가 1:1로 매칭된 지도 학습에 해당되어 비지도 학습인 언페어 데이터 셋의 경우보다 화질이 우세하다. 그럼에도 물방울의 여러가지 패턴들을 학습하면, 실제로 다양하게 맺히는 물방울을 더욱 효과적으로 제거하여 시인성을 개선할 수 있기 때문에 두 모듈을 모두 사용해서 데이터 증식을 하였다.

(a) (b)

그림 3. 2가지 생성 모듈의 가상 물방울 생성 결과
(a) 800개의 페어 데이터 셋으로 학습한 모듈의 결과,
(b) 400개의 언페어 데이터 셋으로 학습한 모듈의 결과
Fig. 3. Results of generating fake water droplet (a) Result of module trained with 800 paired dataset, (b) Result of module trained with 400 unpaired dataset

3.1.3 데이터 분류 모델 학습

세 번째 블록에서, 우리는 증식된 데이터 중 좋은 데이터를 분류하기 위해 CNN 분류 모델을 학습하여 사용했다. 그림 4는 두 가지 가상 물방울 생성 결과를 보여준다.

(a) (b)

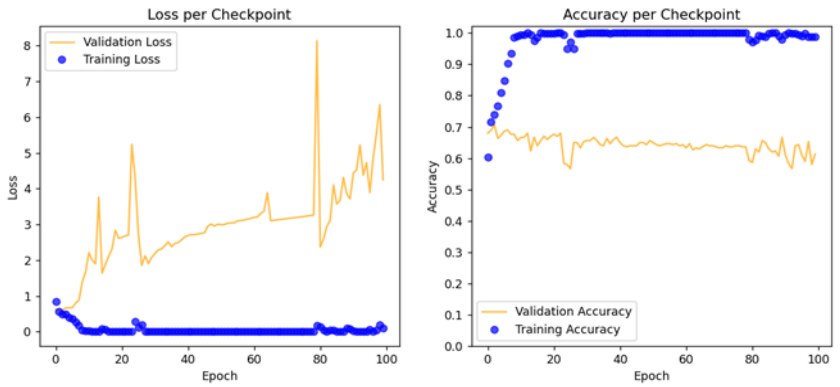
그림 4. 가상의 물방울 생성 결과 (a) 물방울이 잘 생성된 영상, (b) 물방울이 잘 생성되지 않은 영상
Fig. 4. Results of generating fake water droplet (a) Well-formed image, (b) Bad-formed image

대부분의 가상 물방울은 그림 4(a)와 같이 잘 형성되지만, 일부는 그림 4(b)와 같이 제대로 형성되지 않는다. 이러한 데이터를 하나씩 수동으로 분류하는 작업은 매우 번거롭기 때문에 CNN 분류 모델을 학습하여 효율적으로 양질의 데이터를 선별하였다.

본 연구에서 사용된 CNN 모델은 3개의 합성곱층과 풀링층으로 구성되어 있으며, 각 합성곱층에는 3×3 크기의 필터를 사용하고 패딩을 1로 설정하여 입력 영상의 크기를 유지하였다. 활성화 함수로는 ReLU를 적용하였으며, 2×2 크기의 최대 풀링 연산을 수행하여 특징 맵의 크기를 절반으로 축소하였다. 이후에 특징 맵은 Flatten 과정을 거쳐서 1차원 벡터로 변환되며, 완전 연결층으로 전달되어 최종적인 분류를 하게된다. 학습은 600개의 페어 데이터에 대해서 총 100 epoch 동안 진행되었으며, 손실 함수로는 교차 엔트로피 손실 함수(Cross entropy loss), 기울기 손실 최적화 기법으로는 Adam 옵티마이저를 사용하였다[12].

그림 5는 데이터 증식을 적용하지 않고 CNN 모델을 학습한 결과를 보여준다.

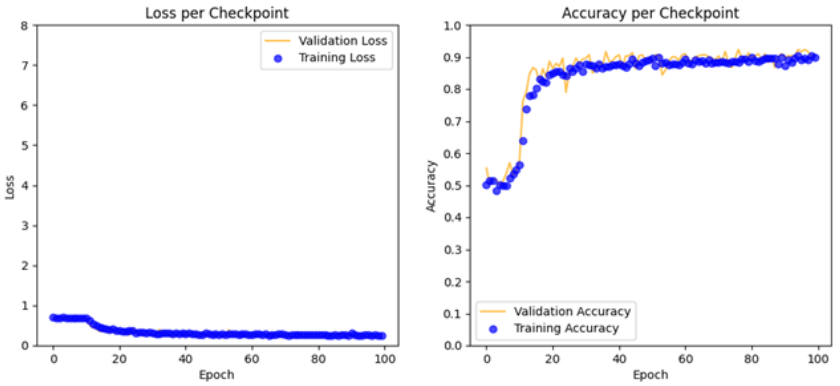
그림 5(a)에서 볼 수 있듯이, 학습 손실(Training loss)은 0에 가까워지는 반면, 검증 손실(Validation loss)은 약 8.5까지 증가한다. 마찬가지로, 그림 5(b)에서 확인할 수 있듯이, 학습 정확도(Training accuracy)는 100%에 근접하지만, 검증 정확도(Validation accuracy)는 55%에서 70% 사이에서 변동한다. 반면 그림 6은 데이터 증식 기법인 RandomFlip과 RandomCrop을 적용하여 CNN 모델을 학습한 결과를 보여준다.



(a) (b)

그림 5. 데이터 증식을 적용하지 않은 CNN 모델 학습 결과 (a) 손실, (b) 정확도

Fig. 5. Results of the CNN model training without data augmentation (a) Loss, (b) Accuracy



(a) (b)

그림 6. 데이터 증식을 적용한 CNN 모델 학습 결과 (a) 손실, (b) 정확도

Fig. 6. Results of the CNN model training with data augmentation (a) Loss, (b) Accuracy

그림 6(a)에서 볼 수 있듯이, 훈련 손실은 0에 근접하며, 검증 손실도 훈련 손실을 따라 0에 가까워진다. 또한 그림 6(b)에서 볼 수 있듯이, 초기 CNN 모델에 비해 훈련 정확도는 다소 낮아 90%에 근접하지만, 검증 정확도도 훈련 정확도를 따라 90%에 가까워진다. 따라서 본 연구에서는 데이터 증식을 적용하여 CNN 모델을 학습하였다. 학습된 CNN 모델로 가상의 물방울 영상 4,000장에 대해서 잘 형성된 가상의 물방울 영상과 잘못 형성된 영상을 분류했다. 표 1은 분류 결과에 대해 분석한 것을 나타내었다.

표 1에서 알 수 있듯이 잘 생성된 가상의 물방울 영상은 4,000장 중 1,200장이다. 이중 800 페어 데이터 셋으로 학습한 모듈의 결과가 685장으로, 400 페어 데이터 셋으로 학습한 모듈의 결과인 515장 보다 조금 더 비중이 큰 것을 알 수 있다. 1,200장의 잘 생성된 가상의 물방울 영상들만 깨끗한 원본과 쌍을 이루어 최종 물방울 제거 모듈의 학습에 사용되었다.

표 1. 분류 결과에 대한 분석

Table 1. Analysis of classification results

	800 paired dataset	400 unpaired dataset	Total dataset
Well formed	685	515	1,200
Bad formed	1,315	1,485	2,800
Total	2,000	2,000	4,000

3.2 시인성 개선

3.2.1 CycleGAN 기반 물방울 제거 모듈 학습

네 번째 블록에서, Rem_module의 물방울 부분 집중 학습을 위해 L1 Loss의 입력과 타겟 부분을 수정했다. 물방울 영상과 깨끗한 영상의 차이를 구한 차영상을 추가로 로드하였고, 이를 정규화(Normalization)하여 L1 Loss의 입력과 타겟 부분에 곱해줬다. 결과적으로 L1 Loss를 적용하는 Cycle Consistency Loss, Identity Loss를 계산할 때 물방울 부분을 집중적으로 학습할 수 있게 되었다.

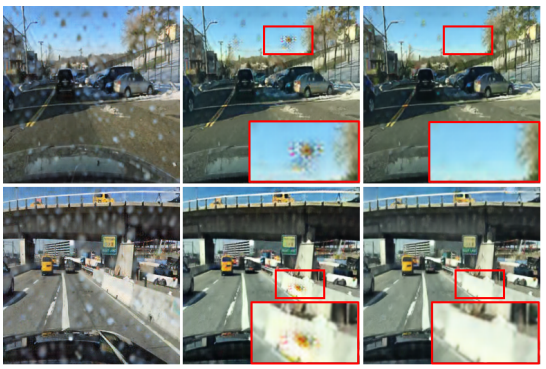
식 (5)와 식 (6)은 각각 수정된 Cycle Consistency Loss, Identity Loss 수식을 나타낸다.

$$L_{cyc}(G,F) = (E_{x\sim p_{data}(x)}[\|M^*(F(G(x))-x)\|_1] + E_{y\sim p_{data}(y)}[\|M^*(G(F(y))-y)\|_1])*100 \quad (5)$$

$$L_{identity}(G,F) = (E_{y\sim p_{data}(y)}[\|M^*(G(y)-y)\|_1] + E_{x\sim p_{data}(x)}[\|M^*(F(x)-x)\|_1])*100 \quad (6)$$

M은 정규화된 차영상으로, 0~1 값을 가지는 물방울 부분을 집중적으로 타겟팅하는 강조 마스크(Attention mask) 역할을 한다. 또한 강조 마스크를 곱했을 때, 손실이 너무 작아지는 것을 방지하기 위해 100을 곱해서 스케일을 조정했다.

그림 7은 L1 Loss 수정 여부에 따른 비교를 보여준다. 그림 7(a)는 L1 Loss를 수정하지 않은 모듈의 물방울 제거 결과, 그림 7(b)는 L1 Loss를 수정한 모듈의 물방울 제거 결과이다. 표시된 부분을 보면 L1 loss를 수정한 모듈이 Loss를 수정하지 않은 모듈에 비해서 물방울 제거시 원본 영상에서 배경을 왜곡하지 않고 잘 보존하는 모습을 확인할 수 있다. 또한 그림 7의 두 번째 사진에서 표시된 부분을 보면 Loss를 수정한 모듈이 Loss를 수정하지 않은 모듈에 비해서 물방울 제거 시 중앙 분리대의 경계 정보가 뚜렷하게 잘 보존하는 것을 확인할 수 있다.



(a) (b) (c)

그림 7. L1 Loss 수정 여부에 따른 비교 (a) 물방울 원본 영상, (b) L1 Loss를 수정하지 않은 모듈의 물방울 제거 결과, (c) L1 Loss를 수정한 모듈의 물방울 제거 결과
Fig. 7. Comparison based on L1 Loss modification (a) Original water droplet image, (b) Results of water droplet removal using module without modifying the L1 Loss, (c) Results of water droplet removal using module with modifying the L1 Loss

따라서 L1 Loss를 수정하고 앞서 기존 데이터 셋에 증식된 데이터 셋을 더한 2,000개의 페어 데이터 셋을 CycleGAN 입력으로 제공하여 Rem_module 학습을 하였다. 이때 학습 epoch는 300으로 설정하였다.

3.2.2 CycleGAN 기반톤 매핑 & 물방울 제거 모듈 학습

다섯 번째 블록에서는 Rem&TM_module 학습을 진행하였다. 이 모듈은 Rem_module 학습에 사용된 데이터 셋과 동일한 데이터 셋을 사용하였으나, 깨끗한 영상 셋에 톤 매핑을 적용하여 새로운 clean target 영상 셋으로 대체한 후 학습을 진행하였다. 톤 매핑은 Retinex 이론과 다중 스케일 CLAHE(Contrast Limited Adaptive Histogram Equalization)를 결합한 기법을 사용하였다[13]. 그림 8(a)는 톤 매핑을 적용하지 않은 기존의 깨끗한 영상, 그림 8(b)는 톤 매핑을 적용한 clean target 영상이다.

본 논문에서 다루는 우천 시 상황은 대개 저조도 환경이므로, 이러한 방식으로 학습된 모듈은 어두운 영역의 디테일을 보존하고 대비를 향상시켜 객체 검출 성능을 개선할 수 있다.

(a) (b)

그림 8. 톤 매핑 적용 결과 (a) 기존의 깨끗한 영상, (b) 톤 매핑을 적용한 clean target 영상

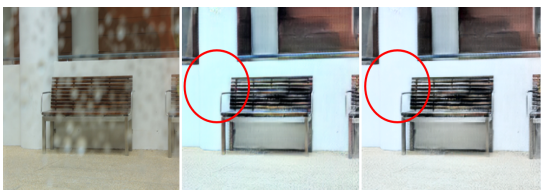
Fig. 8. Results of tone mapping (a) Original clean image, (b) Clean target image with tone mapping

Rem&TM_module은 Rem_module과 동일하게 2,000개의 페어 데이터 셋을 CycleGAN에 입력으로 제공하였으며, 학습 epoch는 300으로 설정하였다.

3.2.3 두 가지 모듈을 사용한 최종 테스트

마지막 블록에서는 학습한 두 가지 모듈을 순차적으로 적용하였다. 먼저 Rem_module을 적용하여 물방울을 제거한 후, 결과 영상을 LAB 색공간으로 변환하고 색상 정보 채널인 A, B 채널을 따로 보존한다. 그 후 Rem&TM_module을 적용한 결과에서 밝기 채널인 L(Lightness) 채널만 추출하여, 보존된 A, B 채널과 합성하여 최종 결과 영상을 생성한다.

그림 9와 그림 10은 A, B 채널 보존 여부에 따른 비교를 나타낸다.



(a) (b) (c)

그림 9. A, B 채널 보존 여부에 따른 비교 (a) 물방울 원본 영상, (b) A, B 채널을 보존하지 않고 Rem&TM_module을 적용한 결과, (c) A, B 채널을 보존하고 Rem&TM_module을 적용한 결과

Fig. 9. Comparison based on A and B channels preservation (a) Original water droplet image, (b) Result of applying Rem&TM_module without preserving A and B channels, (c) Result of applying Rem&TM_module with preserving A and B channels

(a) (b) (c)

그림 10. A, B 채널 보존 여부에 따른 비교 (a) 물방울 원본 영상, (b) A, B 채널을 보존하지 않고 Rem&TM_module을 적용한 결과, (c) A, B 채널을 보존하고 Rem&TM_module을 적용한 결과

Fig. 10. Comparison based on A and B channels preservation (a) Original water droplet image, (b) Result of applying Rem&TM_module without preserving A and B channels, (c) Result of applying Rem&TM_module with preserving A and B channels

그림 9(a)와 그림 10(a)는 A, B 채널을 보존하지 않고 Rem&TM_module을 적용한 결과, 그림 9(b)와 그림 10(b)는 A, B 채널을 보존하고 Rem&TM_module을 적용한 결과이다. 그림 9의 표시된 부분에서 확인할 수 있듯이 A, B 채널을 보존하지 않고 Rem&TM_module을 적용했을 때, 기존에 흰색이었던 벽이 색이 차가운 색감을 띄며 푸른색 계열로 나타난다. 이는 색 공간 변환 과정에서 색 균형(Color Balance)이 맞지 않아 색이 부자연스럽거나 왜곡되는 현상인 화이트 밸런스 오차(White Balance Error)가 발생한 것이다. 반면 A, B 채널을 보존하고 Rem&TM_module을 적용하면 화이트 밸런스 오차가 개선된 것을 확인할 수 있다,

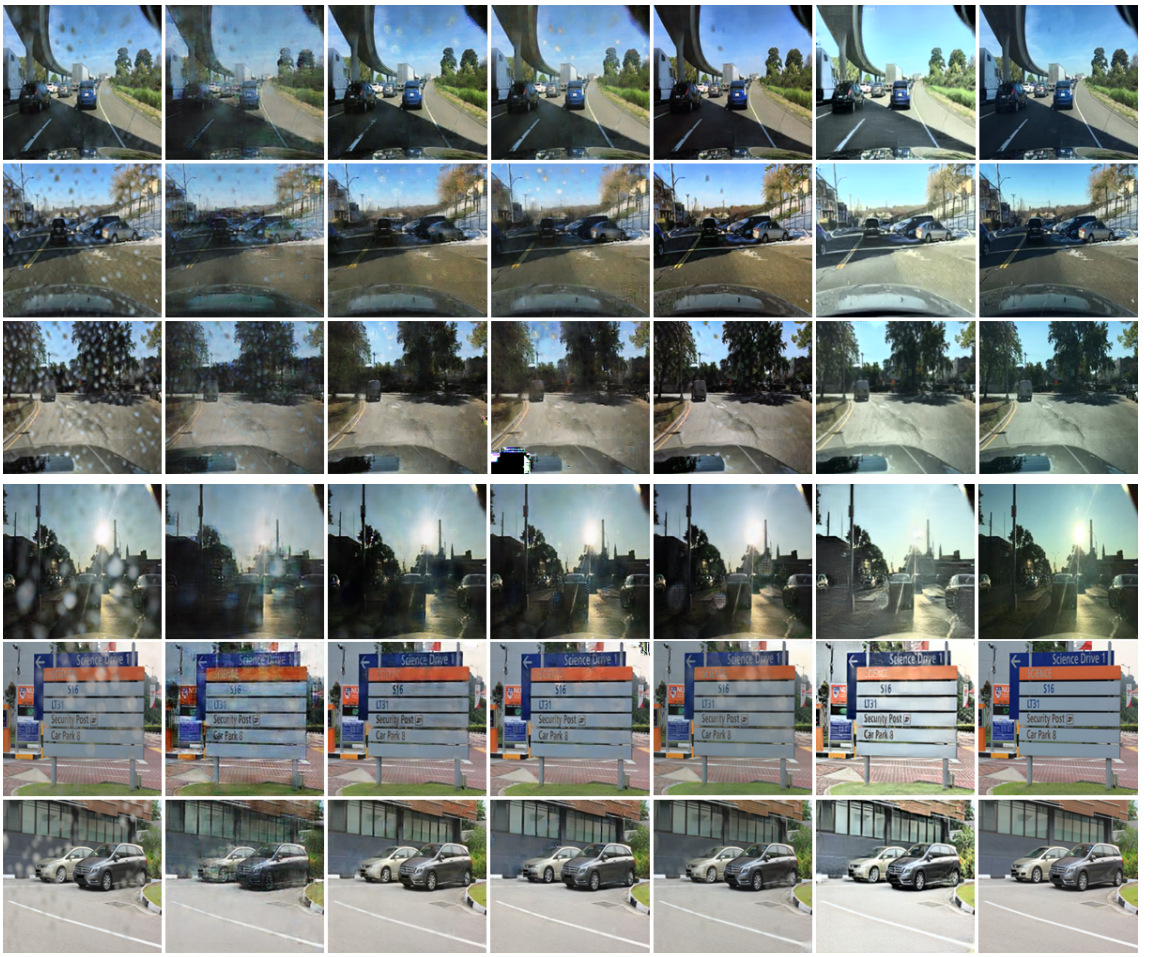
또한 그림 10에서 표시된 부분을 보면 A, B 채널을 보존하지 않고 Rem&TM_module을 적용했을 때, 경계 영역에서 색이 번지는 현상이 발생하는 것을 확인할 수 있다. 이는 색 공간 변환 과정에서 색상 정보 채널인 A, B 채널이 손실되면서 원본 색상의 명확한 구분이 어려워지고, 결과적으로 색이 부자연스럽게 퍼지는 현상(Color Bleeding)이 나타나는 것이다. 반면 A, B 채널을 보존하고 Rem&TM_module을 적용하면 색상 정보가 유지되어 보다 선명한 경계가 나타나며, 색 번짐 현상이 개선된 것을 확인할 수 있다.

Ⅳ. 실험 결과

제안 방법의 물방울 제거 성능 검증을 위해 기존의 rwm(region-weighted method)[14], Pix2pix, HAN et al.[15], CycleGAN의 방법들과 비교 평가 실험을 진행하였다.

그림 11은 입력 영상 그림 11(a)에 대한 제안 방법과 기존 방법을 적용한 비교 영상들이다. 그림 11(b), 그림 11(c), 그림 11(d), 그림 11(e)는 기존 방법들로, 각각 rwm, Pix2pix, HAN et al., CycleGAN의 결과 영상이다. 그림 11(f)는 제안된 방법의 결과 영상이다.

제안 방법은 기존 방법에 비해 물방울을 더 효과적으로 제거했을 뿐만 아니라 명암비(Contrast Ratio, CR)을 개선하여 어두운 영역의 디테일이 향상되는 것을 확인할 수 있다. 특히 그림 11의 첫 번째와 두 번째 비교 영상에서는, 물방울 원본 영상과 기존 방법의 결과에서 그늘 위에 존재하는 차량과 차선이 뚜렷하게 인식되지 않는다. 반면 제안된 방법의 결과에서는 물방울을 효과적으로 제거함과 동시에 명암비를 개선하여, 그늘 위의 차량과 차선이 더욱 명확하게 인식된다.



(a) (b) (c) (d) (e) (f) (g)

그림 11. 기존 방법 및 제안 방법 결과 비교
(a) 물방울 원본 영상, (b) Rwm, (c) pix2pix, (d) HAN et al., (e) CycleGAN, (f) 제안된 방법, (g) 타겟 영상
Fig. 11. Comparisons of the results using the existing methods and the proposed method
(a) Original water droplet images, (b) Rwm, (c) pix2pix, (d) HAN et al., (e) CycleGAN, (f) Proposed method, (g) Target images

또한 그림 11의 네 번째 비교 영상에서, 물방울 원본 영상 및 기존 방법의 결과에서는 역광과 저조도 환경의 영향으로 인해 영상의 4대 차량이 뚜렷하게 인식되지 않는다. 반면 제안된 방법은 물방울을 효과적으로 제거함과 동시에 명암비를 개선하여, 기존 방법들 보다 4대 차량이 더욱 명확하게 인식된다. 표 2는 기존 방법 및 제안 방법의 처리 시간을 나타내었다.

표 2에서 알 수 있듯이 제안 방법은 2가지 모듈을 거치기 때문에 기존 CycleGAN에 비해서 처리 시간이 길다.

표 2. 기존 방법 및 제안 방법의 처리 시간
Table 2. Process time of existing methods and proposed method.

	Rwm	Pix2pix	Han et al.	CycleGAN	Proposed
Process time(sec)	0.111	0.192	0.201	0.067	0.106

V. 결 론

본 논문에서는 우천 시 자율주행 차량의 카메라 및 렌즈 상에 맺힌 물방울을 효과적으로 제거하는 새로운 접근 방식을 제안하였다. 기존의 데이터 셋이 부족한 문제를 해결하기 위해, 먼저 CycleGAN 딥러닝 모델을 기반으로 생성 모듈을 학습한 후 물방울 영상을 생성하였다. 생성된 영상들은 CNN 기반의 분류 모델을 통해 평가되었으며, 품질이 높은 영상들만 선별하여 기존 데이터 셋에 추가함으로써 데이터 증식을 수행하였다. 이후 증식된 데이터를 활용하여 물방울 제거 네트워크를 단계적으로 학습하였다. Rem_module은 기본적인 물방울 제거를 수행하며, Rem&TM_module은 톤 매핑이 적용된 깨끗한 영상을 타겟으로 하여 물방울 제거 및 톤 매핑을 수행하는 역할을 한다. 두 모듈을 순차적으로 적용하는 과정에서, Rem_module을 적용한 영상에서 A, B 채널을 보존하고, Rem&TM_module에서 얻은 결과로부터 L 채널을 추출하여 합성함으로써 최종적인 물방울 복원 영상을 생성하였다. 실험 결과, 본 연구에서 제안한 방식이 기존 방법들과 비교하여 더 자연스럽고 왜곡이 적은 복원 영상을 생성함을 확인하였다. 향후 연구에서는 더욱 다양한 환경에서의 물방울 제거 성능을 개선하기 위해 데이터 셋을 확장하고, 실시간 처리를 고려한 경량화된 네트워크 설계를 진행할 예정이다.

References

[1] H.-J. Kwon and S.-H. Lee, "Raindrop-removal image translation using target-mask network with attention module", Mathematics, Vol. 11, No. 15, 3318, Jul. 2023. https://doi.org/10.3390/math11153318.

[2] I. Goodfellow, et al., "Generative adversarial nets", NIPS'14: Proc. of the 28th International Conference on Neural Information Processing Systems, Montreal, Canada, Vol. 2, pp. 2672-2680, Dec. 2014.

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks", Proc. of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, pp. 1125-1134, Jul. 2017. https://doi.org/10.1109/CVPR.2017.632.

[4] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks", Proc. of the IEEE international conference on computer vision, Venice, Italy, pp. 2223-2232, Oct. 2017. https://doi.org/10.1109/ICCV.2017.244.

[5] X. Ying, "An overview of overfitting and its solutions", Journal of physics: Conference series, Vol. 1168, No. 2, 022022, Mar. 2019. https://doi.org/10.1088/1742-6596/1168/2/022022.

[6] K. Janocha and W. M. Czarnecki, "On loss functions for deep neural networks in classification", Vol. 25, pp. 49-59, 2016. https://doi.org/10.4467/20838476SI.16.004.6185.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", Proc. of the IEEE conference on computer vision and

pattern recognition, Las Vegas, NV, USA, pp. 770-778, Jun. 2016. https://doi.org/10.1109/cvpr.2016.90.

[8] D. Kinga and J. B. Adam, "A method for stochastic optimization", International conference on learning representations (ICLR), San Diego, California, May 2015. https://doi.org/10.48550/arXiv.1412.6980.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", Proc. of the IEEE, Vol. 86, No. 11, pp. 2278-2324, Nov. 1998. https://doi.org/10.1109/5.726791.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", Adv Neural Inf Process Syst, Vol. 60, No. 6, pp. 84-90, May 2017. https://doi.org/10.1145/3065386.

[11] C. Y. Jang and T. Y. Kim, "Hand feature enhancement and user decision making for CNN hand gesture recognition algorithm", Journal of the Institute of Electronics and Information Engineers, Vol. 57, No. 2, pp. 60-70, Feb. 2020. https://doi.org/10.5573/ieie.2020.57.2.60.

[12] S. Ruder, "An overview of gradient descent optimization algorithms", arXiv preprint arXiv:1609.04747, Sep. 2016. https://doi.org/10.48550/arXiv.1609.04747.

[13] Y.-J. Kim, D.-M. Son, and S.-H. Lee, "Retinex Jointed Multiscale CLAHE Model for HDR Image Tone Compression", Mathematics, Vol. 12, No. 10, 1541, May 2024. https://doi.org/10.3390/math12101541.

[14] S.-W. Jung, H.-J. Kwon, and S.-H. Lee, "Enhanced Tone Mapping Using Regional Fused GAN Training with a Gamma-Shift Dataset", Applied Sciences, Vol. 11, No. 16, 7754, Aug. 2021. https://doi.org/10.3390/app11167754.

[15] Y.-K. Han, S.-W. Jung, H.-J. Kwon, and S.-H. Lee, "Rainwater-Removal Image Conversion Learning with Training Pair Augmentation", Entropy, Vol. 25, No. 1, 118, Jan. 2023. https://doi.org/10.3390/e25010118.

저자소개

이 세 완 (Se-Wan Lee)

2020년 2월 ~ 현재 : 경북대학교 IT대학 전자공학부 학부과정
관심분야 : Deep Learning, 컴퓨터 비전, Object Detection, 신호처리

이 성 학 (Sung-Hak Lee)

1997년 2월 : 경북대학교 전자공학과(공학사)
1999년 2월 : 경북대학교 전자공학과(공학석사)
1999년 2월 ~ 2004년 6월 : LG 전자 영상제품연구소 선임연구원
2008년 2월 : 경북대학교 전자공학과(공학박사)
2009년 8월 ~ 2017년 7월 : 경북대학교 IT대학 전자공학부 연구초빙교수
2018년 3월 ~ 현재 : 경북대학교 IT대학 전자공학부 교수
관심분야 : Color Image Processing, Color Appearance Model, Color Management, HDR 영상처리, 영상융합, 인공지능영상처리