한국컴퓨터그래픽스학회
Korea Computer Graphics Society Vol. 25, No. 5, P. 1~9

적대적생성신경망을 이용한 연안 파랑 비디오 영상에서의 빗방울 제거 및 배경 정보 복원

허동°[1] 김재일[2] 김진아*[3]

경북대학교°[1,2], 한국해양과학기술원*[3]

herick@knu.ac.kr°[1], jaeilkim@knu.ac.kr[2], jakim@kiost.ac.kr*[3]

Raindrop Removal and Background Information Recovery in Coastal Wave Video Imagery using Generative Adversarial Networks

Dong Huh°1 Jaeil Kim[2] Jinah Kim*[3]

Kyungpook National University°[1,2], Korea Institute of Ocean Science and Technology*[3]

요 약

본 논문에서는 강우시 빗방울로 인해 왜곡된 연안 파랑 비디오 영상에서 빗방울 제거와 제거된 영역에 대한 배경 정보를 복원하기 위한 적대적생성신경망을 이용한 영상 강화 방법을 제안하고자 한다. 영상 변환에 널리 사용되는 *Pix2Pix* 네트워크와 현재 단일 이미지에 대한 빗방울 제거에 좋은 성능을 보여주고 있는 *Attentive GAN*을 실험 대상 모델로 구현하고, 빗방울 제거를 위한 공개 데이터 셋을 이용하여 두 모델을 학습한 후 빗방울 왜곡 연안 파랑 영상의 빗방울 제거 및 배경 정보 복원 성능을 평가하였다. 연안 파랑 비디오에 영상에 대한 빗방울 왜곡 보정 성능을 향상시키기 위해 실제 연안에서 빗방울 유무가 짝을 이룬 데이터 셋을 직접 획득한 후 사전 학습된 모델에 대하여 전이 학습에 사용하여 빗방울 왜곡 보정에 대한 성능 향상을 확인하였다. 모델의 성능은 빗방울 왜곡 영상으로부터 파랑 정보 복원 성능을 최대 신호 대 잡음비와 구조적 유사도를 이용하여 평가하였으며, 전이 학습을 통해 파인 튜닝된 *Pix2Pix* 모델이 연안 파랑 비디오 영상의 빗방울 왜곡에 대한 가장 우수한 복원 성능을 보였다.

Abstract

In this paper, we propose a video enhancement method using generative adversarial networks to remove raindrops and restore the background information on the removed region in the coastal wave video imagery distorted by raindrops during rainfall. Two experimental models are implemented: *Pix2Pix* network widely used for image-to-image translation and *Attentive GAN*, which is currently performing well for raindrop removal on a single images. The models are trained with a public dataset of paired natural images with and without raindrops and the trained models are evaluated their performance of raindrop removal and background information recovery of rainwater distortion of coastal wave video imagery. In order to improve the performance, we have acquired paired video dataset with and without raindrops at the real coast and conducted transfer learning to the pre-trained models with those new dataset. The performance of fine-tuned models is improved by comparing the results from pre-trained models. The performance is evaluated using the peak signal-to-noise ratio and structural similarity index and the fine-tuned *Pix2Pix* network by transfer learning shows the best performance to reconstruct distorted coastal wave video imagery by raindrops.

키워드: 연안 파랑 비디오 영상, 빗방울 제거, 배경 정보 복원, 적대적생성신경망, 영상 강화

Keywords: Coastal wave video imagery, Raindrop removal, Background information recovery, Generative adversarial networks, Video enhancement

*corresponding author: Jinah Kim/Korea Institute of Ocean Science and Technology(jakim@kiost.ac.kr)

Received : 2019.10.04./ Review completed : 1st 2019.11.04. / Accepted : 2019.11.07.
DOI : 10.15701/kcgs.2019.25.5.1
ISSN : 1975-7883(Print)/2383-529X(Online)

1. 서론

연안/해안은 바다와 육지가 맞닿아 밀접한 상호작용이 발생하는 지역/해역으로 해변, 갯벌, 만, 사구 등 다양한 형태를 가지며 항만, 어항, 관광지, 해양 스포츠 등으로 이용도가 높은 동시에 태풍, 해일, 고파랑, 해수면 상승, 해변 침식, 이안류 등의 연안 재해에 매우 취약한 지역이기도 하다. 특히 해안 개발, 폭풍 강도 및 내습 빈도 증가, 기후 변화 등 해양 환경 변화로 연안 재해가 더욱 빈번하게 발생하고 있으며, 이로 인한 인명 및 재산 피해가 갈수록 증가하고 있다. 이러한 연안 재해 저감 및 대응을 위해 연안에서의 파랑 현상에 대한 이해와 예측이 필수적이나, 연안 지역에서의 파랑은 기존 센서를 이용한 정점 관측이 용이하지 않고, 공간적 파랑 변형의 높은 비선형성으로 인해 현상의 정확한 이해에 한계가 있다. 이는 월파, 이안류, 침수범람, 연안침식, 연안 구조물 안정성 등 파랑 기인 연안 재해 대응 및 저감에 어려움으로 이어진다.

이에 CCTV 비디오, 저궤도 위성 등 광학 영상 및 레이다 등 원격탐사 기술과 영상처리 기법 적용을 통한 2·3차원 영상기반 연안 파랑 관측와 모델링 기술이 활발하게 연구되고 있다 [1]. 특히 컴퓨팅 파워 및 딥러닝 기술이 급속도로 발전하면서 대용량 비디오 영상의 이해와 모델링 연구가 용이해짐에 따라 대용량의 실해역 비디오 영상을 통한 연안에서의 파랑 이해를 위한 융합 연구가 국내외에서 활발히 이루어지고 있다 [2, 3, 4].

이와 같은 딥러닝 기술을 적용하여 연안에서 파랑 거동 학습을 통해 정확히 파랑을 관측하고 예측하기 위해서는 충분한 학습 비디오 영상이 필요하며, 특히 파랑 거동에 높은 비선형성을 보이는 국지적 악기상에 의한 재해성 파랑 발생시에는 보다 장기간에 다양한 형태의 파랑 거동이 기록된 충분한 학습 데이터가 필수적이다. 그런데 연안 재해가 주로 발생하는 악기상시에는 강우가 동반되는 경우가 대부분이므로, 카메라 렌즈에 묻은 빗방울과 함께 촬영된 파랑 비디오 영상이 저장된다. 이를 딥러닝 모델의 학습 데이터로 사용하기 위해서는 모든 비디오 영상에 대해 빗방울 제거 및 제거된 부분의 배경 파랑 정보의 복원이 필요하다.

이에 본 연구에서는 적대적생성모델(Generative Adversarial Network, GAN)[5]을 이용하여 현재 단일 이미지에 대한 빗방울 제거에 우수한 성능을 보이는 Attentive GAN[6]과 Pix2Pix[7] 모델을 구현하고, Qian 등이 공개한 빗방울 유무로 짝을 이룬 학습 이미지 데이터 1,119건을 (이후 *Raindrop 1119*로 지칭) 이용하여 사전 학습(pre-trained)을 실시하고, 강우시 연안 파랑 비디오 영상을 적용하여 성능을 살펴본다.

또한 강우시 연안 파랑 비디오 영상에 대한 빗방울 제거와 배경 파랑 정보 복원의 성능을 높이기 위해 실해역인 강릉 안목 해변에서 빗방울 유무로 짝을 이룬 연안 파랑 비디오 영상을 직접 획득하고 이를 이용하여 사전 학습을 마친 두 모델에 대한 전이 학습(transfer learning)[8]을 통해 성능 평가를 실시한다. 성능 평가는 빗방울 제거 후 배경 정보 복원 영상의 화질 손실정보를 측정하는 최대 신호 대 잡음비(Peak Signal-to-Noise Ratio, PSNR)과 복원 영상의 구조적 유사 지수를 측정하는 구조적 유사도(Structural Similarity Index, SSIM) 값의 계산을 통해 실험 케이스별 성능을 비교·평가하고자 한다.

2. 관련 연구

단일 이미지 및 비디오 영상에서 빗방울 제거에 대한 연구는 크게 빗방울의 수학적 표현에 대한 사전 정의를 통한 필터링 기법과 빗방울 제거에서 나아가 배경 정보의 복원을 위한 빗방울 유무에 따른 짝을 이루는 데이터를 이용한 학습 기반 심층신경망을 이용한 방법이 있다.

Luo[8] 등은 사전 정의한 빗방울, 빗줄기, 빗물의 다양한 형태와 크기에 대한 수학적 표현을 통한 필터링 기법으로 빗방울을 제거하였으며, Kim[9] 등은 이에 추가로 카메라 렌즈 표면과 거리 차이 등에 의한 효과를 함께 고려하여 다양한 빗방울 정보들을 사전에 수학적으로 모델링했으며, You[10] 등은 단일 빗방울에 대한 시간 변화를 함께 고려하여 수학적으로 모델링하여 이미지 내에서 빗방울 제거를 위한 필터링을 적용하는 영상처리 기법을 제안하였다. 하지만 이와 같은 방법들은 강우 시 발생하는 빗방울에 대한 무수한 형태와 시간에 따른 변형을 사전에 모두 정의하는데 한계가 있을수 있으며, 특히 빗방울 제거 후 제거된 영역에 대한 배경 정보의 복원은 어렵다.

Fu[11] 등과 Shen[12] 등은 빗방울 검출을 통해 이를 제거하고, 제거된 영역의 배경 정보를 복원하기 위한 딥러닝 방법을 적용하였고, 작은 크기로 흩어진 빗방울에 대한 제거 및 배경 정보 복원 성능은 양호했으나 빗방울이 크고 밀집되어 있는 경우 복원된 부분이 흐리게 표현되는 등 배경 정보 복원 성능은 좋지 않았다. 이를 개선하기 위해 Qian[6] 등은 *Raindrop 1119* 공개 데이터셋 (Figure 1(a))을 구축하고 Attentive GAN이라는 적대적생성신경망을 이용하여 빗방울이 제거된 후 배경 정보로 복원되어야 하는 영역에 시각적으로 주의(attention)를 주어 검출된 빗방울 영역마다 복원될 영역을 배경 정보로 복원함으로써 기존 연구 대비 우수한 성능을 보여주었다. 그러나 학습 데이터의 대부분이 육상에서 획득한 이미지로 구성되어 있어 연안 파랑 비디오 영상 적용시 빗방울 제거 및 배경 정보 복원에 대한 성능은 좋지 않았다.

3. 연구 방법

본 연구에서는 악기상시 촬영된 연안 파랑 비디오 영상에 대한 빗방울 제거와 배경 파랑 정보의 복원을 위하여, *Pix2Pix* 및 *Attentive GAN* 모델을 구현하고, 빗방울 왜곡 영상을 포함하는 공개 데이터 셋인 *Raindrop 1119*으로 학습하여 파랑 비디오 영상의 화질 개선 성능을 확인한다.

또한 위처럼 공개 데이터로 사전 학습된 두 모델을 강릉 해변에서 획득한 연안 비디오 영 상을 이 용해 재학습 시 키는 전이 학습을수행하고, 이에 따른 왜곡 복원 성능 향상 수준을 평가한다.

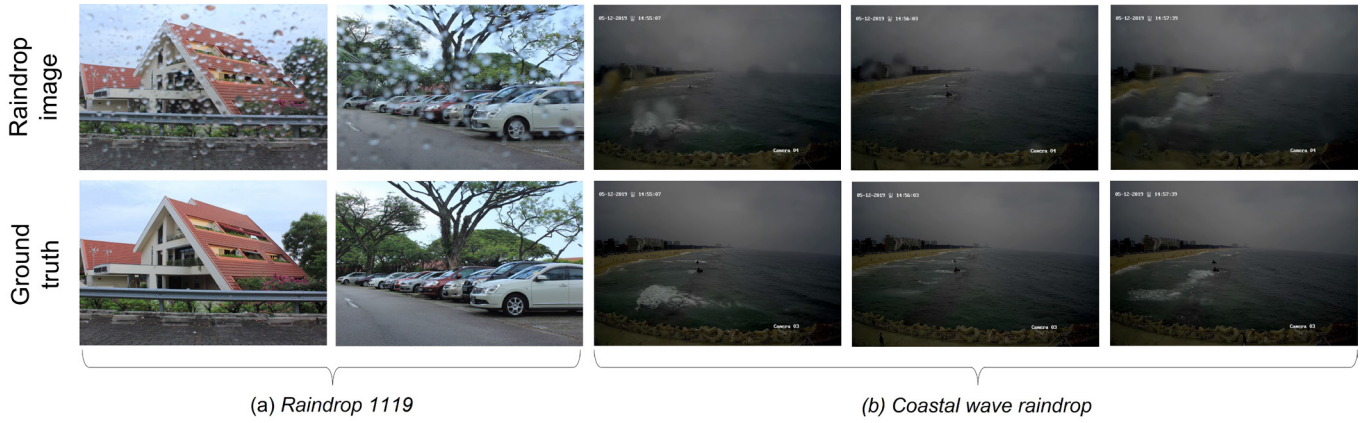


Figure 1: Samples of (a) *raindrop 1119* and (b) *coastal wave raindrop* datasets. Top: The images degraded with raindrops. Bottom: The corresponding ground-truth images.

3.1 학습 데이터 구축

강릉 안목 해변에서 빗방울로 왜곡된 비디오 영 상과 그에 대응하는 깨끗한 영상을 2개의 CCTV를 서로 가로로 부착하여, 2 대가동시에 촬영하는 방법으로 빗방울 유무를 포함하는 짝 비디오 영 상 데이터를 획득하였다 (이후 *Coastal wave raindrop*이라 지칭). 획득한 비디오영상의 시·공간 해상도는 30 FPS, 1920 × 1080이다. 악기상 시 빗방울로 인한 CCTV 렌즈 왜곡 패턴을재현하기 위해서, 두 대의 CCTV 카메라에 20 × 15cm 크기의 투명 유리판을 각각 렌즈 앞에 부착한 뒤, 한 대의 카메라에만 분무기를사용하여 빗방울 패턴을 모사하였다. 다른 카메라는 동일 시간,동일유리판이 부착된 상태로 깨끗한 비왜곡 영상도동시 획득하였다. 카메라 2대의위치 차이로 인하여, 영상에서 위치와 각도의차이가 미세하게 발생하였다. 하지만 모든 평가 영상 전체에 차이가 일정하고, 지도적 학습을이용하는 인공신경망을 대상으로연구를 수행하기 때문에 데이터 셋의 보정은 수행하지 않았다.

수집된 영상은 1분 이내의 간격으로 총 13개의 다양한 빗방울 및 시간에 따른 빗방울 변형 영상을 포함하고 있으며, 약 10여분 동안 빗방울의 시·공간적 형태 변화를 담고 있다. 13가지의 비디오 영상에서 학습 및 검증에 짝을 이룬 17,002개 프레임을 추출하였고, 그 중 일부 샘플 이미지를 Figure 1(b)에서 보여주고 있다.

3.2 빗방울 왜곡 보정을 통한 영상 강화를 위한 *Pix2Pix* 모델 구현 및 실험

Figure 2(a)와 같이 Pix2Pix 모델은 영상 전이(image transfer)를 위한 대표적 적대적생성모델 형태 중 하나로 입력 영상(x)로부터 목표 영상(y)를 생성하는 생성자(Generator, G) 네트워크와 생성 영상과 목표 영상을 구분하는 판별자(Discriminator, D) 네트워크로 구성된다.

빗방울 왜곡으로 인한 영상 강화를 위한 *Pix2Pix* 모델에서 x는 빗방울에 의해 저해된 영상으로 생성자에 입력되며, 생성자는 입력으로부터 빗방울 왜곡이 보정된 영상 $G(x)$를 생성한다. 생성자는 $G(x)$를 최대한 목표 영상 y에 가깝게 만들면서, 동시에 판별자가 $G(x)$와 y를 구분하지 못하도록 $G(x)$를 생성하는 학습을 한다. 판별자는 $G(x)$와 y를 가짜 영상과 진짜 영상으로 구분하도록 학습된다. 이러한 반복과정을 통해 생성자는 판별자와 서로 적대적으로 학습을 하며 판별자를 속이기 위해서 빗방울 왜곡 영상과 목표 영상을 구분하는 주요 특징들을 대상으로 빗방울 왜곡이 보정된 영상으로 복원하게 된다.

생성자 G는 인코터(encoder)와 디코더(decoder)로 구성되며, 인코더는 입력 영상으로부터 목표 영상 복원을 위한 주요 특징을 합성곱 블럭(convolution block)을 연속으로 쌓은 신경망 형태로 학습하고, 디코더는 압축된 영상 특징 정보를 다시 본래 해상도 영상으로 복원하면서 빗방울 왜곡이 없는 목표 영상(y)과 가깝게 사상한다. 합성곱 블럭은 합성곱 레이어와 batch normalization, ReLU(Rectified Linear Unit) 활성화 함수로 연결된다. 비왜곡 영상 복원 과정 중 인코더에서의 영상 압축 과정에 의해 지역적인 영상 상세 특징들을 손실하게 되는데, 이러한 정보 손실을 최소화하기 위해서 인코더의 각 합성곱 블럭의 결과 벡터를 디코더의 합성곱 블럭의 입력으로 추가하는 skip connection[13]을 이용한다. 생성자의 손실함수는 수식 (1)과 같이 첫 항은 목표 영상과 생성 영상 간의 차이를 나타내며, 마지막 두 항은 판별자를 속이기 위한 판별 손실 함수에 해당한다.

$$L_g = \mathbb{E}_{x,y}[\|y - G(x)\|_1] + \mathbb{E}_{x,y}[log(D(x,y))] + \mathbb{E}_{x,y}[log(1 - D(x, G(x)))] \quad (1)$$

판별자 D는 목표 영상(y)과 생성 영상($G(x)$)을 실제 영상과 가짜 영상으로 구분하는 합성곱 신경망으로써, 합성곱 블럭과 max pooling 레이어를 번갈아 여러 층 쌓아서 판별을 위한 영상 특징을 학습한다. 판별자는 x와 y 혹은 x와 $G(x)$를 짝으로 묶어서 입력받아서, 입력 영상 x를 구분을 위한 조건으로써 사용한다.

영상의 크기가 크고 영역에 따른 특징이 다양하기 때문에 영상을 전체가 아닌 작은 패치(patch)로 분리하여 실제와 생성 영상

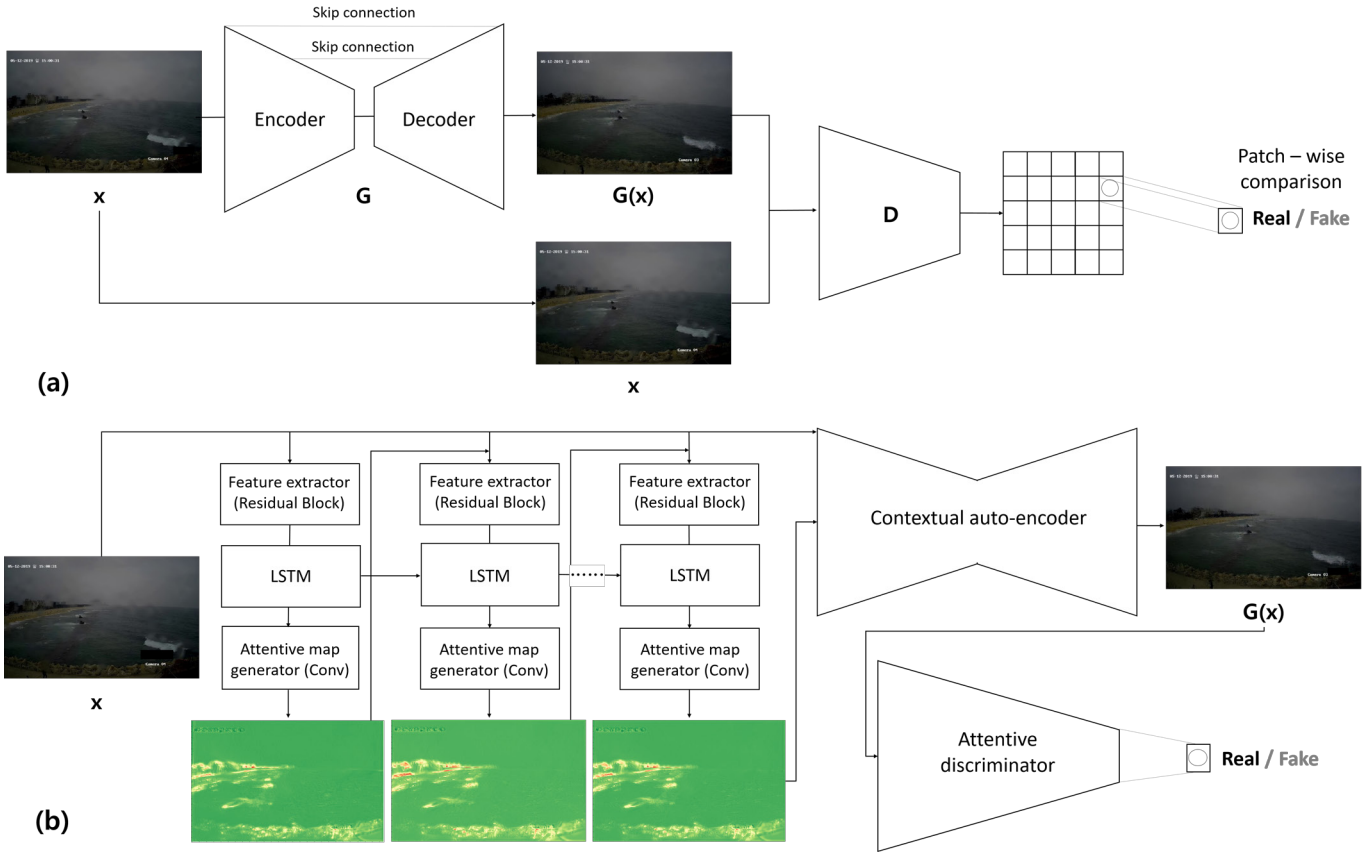


Figure 2: The architecture of implemented (a) *Pix2Pix* network and (b) *Attentive GAN*.

패치로 구분하는 PatchGAN 판별자를 사용하며, 판별자의 손실 함수는 수식 (2)와 같다.

$$L_d = \mathbb{E}_{x,y}[1 - log(D(x,y))] + \mathbb{E}_{x,y}[log(D(x,G(x)))] \quad (2)$$

빗물 왜곡 보정을 위해 구현한 *Pix2Pix* 모델의 생성자는 $720 \times 480 \times 3$ 크기의 영상을 입력으로 받아 인코더에 해당하는 4번 합성곱 블록을 포함하고, 디코더는 인코더 출력을 받아서 3번의 전치 합성곱(transpose convolution)과 합성곱 블록으로 구현한다. 전치 합성곱 레이어를 사용함으로써 생성 영상은 입력 영상과 동일한 크기를 갖게 된다.

판별자는 영상을 60×60의 패치로 구분하여, 각 패치에 대한 판별을 수행한다. 학습 중 배치 크기는 4로 설정하였으며, Adam 알고리즘(learning rate=0.001, β1=0.9, β2=0.999)을 최적화를 위해 사용하였다.

3.3 빗방울 왜곡 보정을 통한 영상 강화를 위한 *Attentive GAN* 모델 구현 및 실험

Qian 등이 제안한 *Attentive GAN*은 *Pix2Pix*와 달리 빗방울 영역을 결정하는 주의 집중(attentive) 순환 신경망(recurrent network)과 빗방울 영역을 중심으로 비왜곡 영상을 복원하는 문맥적 오토인코더(contextual auto-encoder), 그리고 복원 영상과 실제 영상을 빗방울 영역을 중심으로 구분하는 주의 집중 판별자로 구성된다 (Figure 2(b)). 주의 집중 순환 신경망과 문맥적 오토인코더는 비왜곡 영상 생성자에 포함되어 동시에 역전파(back-propagation)를 통해 학습된다[6].

주의 집중 순환 신경망은 빗방울로 왜곡된 영상에서 빗방울 영역을 검출하기 위해 순환 신경망의 출력 영상이 빗방울이 맺힌 영역을 표시하는 빗방울 영역의 마스크(mask)과 동일하도록 학습한다. 순환 신경망으로써 출력을 자신의 입력으로 다시 사용하면서 은닉(hidden) 특징 벡터를 함께 전달하는 반복적인 입·출력 과정을 가지며 이를 통해 빗방울 영역을 점차 명확하게 결정한다. 주의 집중 순환 신경망이 최종 생성하는 주의 집중 맵(attentive map)은 입력 영상(x)에서 빗방울 영역을 1로 표현하고, 배경을 0으로 표현한다. 주의 집중 순환 신경망의 학습을 위해서 *Pix2Pix*와는 달리 *Attentive GAN*은 입력 영상의 빗방울 영역을 나타내는 빗방울 영역 마스크가 추가로 요구된다. 빗방울 영역 마스크는 짝으로 이루어진 빗방울 왜곡 영상과 비왜곡 영상 차이를 계산한 뒤, 차이값에 대한 역치를 이용해 이진화하여 생성한다. 주의 집중 맵 생성을 위한 손실 함수는 수식 (3)과 같다. N은 순환 신경망의 총 순환 횟수를 의미하며, 메모리 사용량을 고려하여 4회로 구현한다. t는 순환 횟수이고, A_t는 주위 집중 순환 신경망이 생성한 주의 집중 맵이다. M은 빗방울 영역 마스크를 의미한다. θ는 매 순환 시 발생하는 손실에 대한 가중치로써 0.8

Figure 3: Results of comparing pre-trained *Pix2Pix* and pre-trained *Attentive GAN* using *raindrop 1119* dataset.

로 지정한다.

$$L_{att} = \sum_{t=1}^{N} \theta^{N-t} \mathbb{E}_{A_t,M}[(A_t - M)^2] \tag{3}$$

문맥적 오토인코더는 주의 집중 맵과 빗방울 왜곡 영상(x)을 동시에 입력으로 받은 후 주의 집중 맵에 나타난 빗방울 영역을 중심으로 빗방울에 의해 가려진 배경을 복원하는 역할을 수행한다. 주의 집중 맵을 통해 문맥적 오토인코더가 명시적으로 입력 영상마다 다른 왜곡 영역에 대한 참조를 할 수 있기 때문에 왜곡 영상에서 빗방울의 영역이 명확하게 나타날 경우 *Pix2Pix*에 비해 보다 보다 정확하게 배경을 복원하게 될 것이다. 문맥적 오토인코더는 *Pix2Pix*의 생성자 신경망과 동일하게 합성곱 블록 및 전치 합성곱 레이어를 사용한 인코더와 디코더로 구성되고, skip connection으로 연결된다. 문맥적 오토인코더는 디코더에서 크기를 늘릴 때마다 해당 크기에 대한 영상 복원 손실을 계산하는 다중 크기 손실 함수(multi-scale loss)와 사전 학습된 VGG 신경망을 사용하여 목표 영상과 생성 영상의 VGG 신경망 특징 벡터가 유사하게 만드는 지각 손실 함수(perceptual loss), 그리고 판별자에 대응하기 위한 GAN 손실 함수를 통해 학습한다(수식 4). 수식 4에서 S는 다중 크기로 나눈 횟수이며, λ_i는 각 크기에 해당하는 손실 함수의 가중치이다. $G_i(x)$는 각 크기에 대해 문맥적 오토인코더의 생성 영상이다. *Attentive GAN* 제안 논문[6]과 동일하게 S는 3단계(1/16, 1/4, 1:1 크기)로 지정하였고, 작은 크기부터 λ_i를 0.6, 0.8, 1.0으로 구현한다. GAN 손실 함수에는 가중치 (λ_g)로 10^{-2}을 적용한다.

$$\begin{aligned} L_{context} = \sum_{i=1}^{S} \lambda_i \mathbb{E}_{x,y}[(G_i(x) - y)^2] + \\ \mathbb{E}_{x,y}[(VGG(G(x)) - VGG(y))^2] + \\ \lambda_g \mathbb{E}_{x,y}[log(1 - D(G(x)))] \end{aligned} \tag{4}$$

주의 집중 판별자는 생성 영상과 실제 역상을 판별하는 생성자에 적대적 역할을 담당하며, 문맥적 오토인코더는 판별자를 속이기 위한 생성 영상에서 실제 영상과 구분되는 특징을 없애기 위한 학습을 한다. 주의 집중 판별자는 *Pix2Pix*의 판별자와 달리 판별자 스스로 주의 집중 맵을 생성하여 해당 영역에 대한 가중치로써 사용함으로써 빗방울 영역에 대한 판별의 초점을 맞춘다. 주의 집중 판별자는 *Pix2Pix*와 같이 판별을 위한 손실 함수와 더불어 판별자에서 생성하는 주의 집중 맵이 주의 집중 순환 신경망의 결과와 유사하도록 지도하는 손실 함수를 사용한다. 해당 손실 함수는 수식 5와 같다. D_{map}은 판별자가 생성한 주의 집중 맵이며, 생성자가 생성한 G(x)로부터는 x에 대한 주의 집중 맵 A_N과 유사하게 만들고, 비왜곡 영상 y를 판별할 때는 D_{map}이 0이 되도록 손실 함수를 구현한다.

$$\begin{aligned} L_{map} = \mathbb{E}_{G(x),A_N}((D_{map}(G(x)) - A_N)^2) + \\ \mathbb{E}_y((D_{map}(y) - 0)^2) \end{aligned} \tag{5}$$

Attentive GAN 학습은 *Pix2Pix*와 동일한 구성으로 학습 중 배치 크기는 4로 설정하였으며, Adam 알고리즘(learning

rate=0.001, β1=0.9, β2=0.999)을 최적화를 위해 사용하였다.

4. 결과

*Pix2Pix*와 *Attentive GAN* 모델의 빗방울 왜곡 복원 성능을 평가하기 위해 최대 신호 대 잡음비(Peak Signal-to-Noise Ratio, PSNR)와 구조적 유사도(Structural Similarity, SSIM)을 사용한다. PSNR은 영상 화질의 손실 정보를 정량화하는 측정 방법 중하나로써 왜곡 보정 영상과 정답 영상 간 차이값의 평균을 로그스케일로 계산한다. PSNR은 수식 6으로 연산된다. 수식 6에서 s는 영상 데이터 타입의 최대값과 최소값의 차이를 나타내며, 본연구에서는 255이다. MSE는 비왜곡 고 화질 영상과 왜곡 복원영상 간 평균 제곱 오차이다.

$$PSNR = 10log\frac{\mathrm{s}^2}{MSE} \tag{6}$$

SSIM은 비왜곡 고 화질 영상과 복원 영 상 사 이의 구조적 유사지수를 수식 7로 측정한다. SSIM은 고화질 영상 x와 복원영상 y에 대해 두 영상 간 밝기 유사도($l(x, y)$), 대조(contrast) 유사도 ($c(x, y)$), 구조 유사도 ($s(x, y)$)를 각 영상의 화소 평균 (μ)과 화소표준 편차(σ), 그리고 영상 간 공 분산(σ_{xy})을 이 용해 0과 1사이로정량화 한다. 동일 영상 간 SSIM은 1이다. 수식 7에서 c_1, c_2, c_3는 수식의 안정성을 위 해 사용한다.

$$\begin{aligned} l(x,y) &= \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}, \\ c(x,y) &= \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}, \\ s(x,y) &= \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}, \\ SSIM(x,y) &= l(x,y)c(x,y)s(x,y) \end{aligned} \tag{7}$$

빗방울에 대한 왜곡이 본래 배경과 유사하게 복원될수록 영상 화질의 손상이 적어지고 정답 영상과 동일한 형태를 갖게됨으로써 PSNR과 SSIM이 모두 향상된다. 수집한 안목 해변 영상에서 빗방울 왜곡 영상과 비왜곡 영상 간 평균 PSNR과 SSIM은 각각 25.01, 0.70이었다.

먼저, *Pix2Pix* 및 *Attentive GAN* 모델을 *Raindrop 1119* 데이터로 학습을 한 후, 안목 해변 파랑 비디오 영상에 대한 빗방울 제거 및 배경 정보 복원 성능을 평가하였다. *Pix2Pix* 모델은 PSNR에서 26.11, SSIM에서 0.71, *Attentive GAN* 모델은 26.04, 0.71의 PSNR과 SSIM이 계산되었다(Table 1(a)와 (b)). Figure 4 (c)와 (d)에서 두 모델을 통해 얻은 복원 영상 중 4가지 결과에 대한 샘플 이미지를 보여주고 있다.

Pix2Pix 모델이 PSNR에서 약간 높은 성능을 기록하였으며, 두 모델의 정량적 평가에서 유의한 차이를 확인하지 못하였다. 그리고 *Pix2Pix* 모델은 공개 데이터에 포함되지 않는 패턴인 CCTV 시간 기록 글자 등에서 왜곡된 결과를 보여주었다(Figure 4 (c)).

Table 1: PSNR and SSIM for the pre-trained and fine-tunned *Pix2Pix* and *Attentive GAN*.

Experiments	**PSNR**	**SSIM**
(a) Pre-trained *Pix2Pix*	26.11	0.71
(b) Pre-trained *Attentive GAN*	26.04	0.71
(c) Fine-tuned *Pix2Pix*	30.29	0.78
(d) Fine-tuned *Attentive GAN*	26.71	0.69

Attentive GAN 모델은 *Pix2Pix*와 달리 글씨에 왜곡은 없었으나, 생성 영상이 *Pix2Pix*에 비해 흐릿하였으며, 이에 따라 PSNR에서 낮은 수치를 보인 것으로 보인다. 왜곡 영상과 비왜곡 영상 간 PSNR과 SSIM과 비교하였을 때, 두 모델이 약간의 보정을 하였으나 큰 성능 향상을 보이지 못했음을 확인하였다. 이는 실제 연안 비디오가 공개 데이터의 배경과 달리 연안의 광범위한 영역을 배경으로 하고 있으며, 파랑 고유의 특징들을 공개 데이터에서는 찾을 수 없기 때문이다.

안목 해변에서 획득한 영상 샘플로 사전 학습된 *Pix2Pix* 및 *Attentive GAN* 모델을 전이 학습시킴으로서 모델 파라미터를 파인 튜닝한 뒤, 빗방울 왜곡 보정에 대한 성능을 평가하였다. 본 평가에서는 13개 빗방울 패턴에서 12개 패턴에 대한 영상을 학습 데이터로 사용하고, 다른 1개 패턴을 평가하는 LOCV 교차검증 (leave-one out cross-validation)을 수행하였다.

*Pix2Pix*는 PSNR과 SSIM이 각각 30.29, 0.78이었으며, *Attentive GAN*은 26.71, 0.69로 계산되었다(Table 1(c)와 (d)). 파인 튜닝에 의한 *Pix2Pix*의 성능 향상은 PSNR에서 16%, SSIM에서 9% 정도였으나, *Attentive GAN*은 PSNR에서 약간의 성능 향상을 보인 반면 SSIM에서는 도리어 성능 하락이 있었다.

Figure 4 (e)와 (f)에서 두 모델을 통해 얻은 복원 영상 중 4가지 결과에 대한 샘플 이미지를 보여주고 있다. 생성 영상에서도 *Attentive GAN* 모델은 물결 모양의 왜곡을 함께 생성하였고, 파랑 영역에서 빗방울에 의한 왜곡 복원이 사전 학습 모델에 비해 제대로 이루어지지 않았다. 이러한 결과는 연안 CCTV 영상에서 빗방울 영역이 공개 데이터에 비해 선명하지 않기 때문에 *Attentive GAN*에서 요구하는 빗방울 영역 마스크를 제대로 생성하기 어렵고, *Attentive GAN* 모델의 파라미터 수가 *Pix2Pix* 모델에 비해 월등히 많아서 쉽게 과적합될 수 있기 때문인 것으로 보인다.

또한 파랑 모양 복원에 있어서 빗방울에 의해 파랑이 많이 왜곡될수록 파랑 복원의 정확성이 하락하는 결과를 확인하였다(Figure 4의 두번째 열). 반면에 *Pix2Pix* 모델은 파인 튜닝을 통해 PSNR과 SSIM이 월등히 향상되었으며, 실제 연안 데이터와 전이 학습을 이용한 딥러닝 모델 학습이 빗방울 왜곡 보정에 유효함을 확인하였다.

5. 결론 및 향후 연구

본 논문은 연안 파랑 비디오 영상에서 자주 발생하는 빗방울에 의한 왜곡을 제거하고 배경 파랑 정보를 복원하기 위하여 현재 영상 변환에 널리 사용되는 *Pix2Pix* 네트워크와 현재 단일 이미지에 대한 빗방울 제거에 좋은 성능을 보여주고 있는 *Attentive GAN*을 구현·평가하였다. *Raindrop 1119* 공개 데이터로 두 모델을 사전 학습하고, 강릉 안목 해변에서 직접 촬영한 강우 시 파랑 비디오 영상을 이용하여 전이 학습한 모델에 대해 빗방울 왜곡 복원 성능을 평가하였다.

*Attentive GAN*은 전이 학습에서 성능 향상을 확인하지 못하였으나, *Pix2Pix* 모델은 전이 학습을 통해 연안 비디오에서 빗방울 왜곡 제거에 유효한 결과를 보였다. 이는 연안 CCTV 비디오는 공개 데이터에 비해 넓은 영역을 배경으로 하고 있으며, 지역에 따른 고유 파랑 패턴을 학습할 필요가 있음을 의미한다.

연안 파랑 비디오 영상의 왜곡 보정을 위해서는 실제 CCTV 연안 비디오의 획득이 필수적이지만, 강우 중 발생하는 다양한 빗방울 패턴을 포함하면서 정확하게 대응되는 비왜곡 파랑 영상의 획득은 매우 어렵다. 이에 따라 빗방울 유무가 포함된 짝 영상을 사용하는 지도 학습 기반 방식에는 한계가 있으므로, 향후 서로 다른 도메인의 영상 변환을 지원하는 비지도 학습(unsupervised learning) 기반 모델을 통해 보다 높은 성능 향상이 필요할 것으로 보여진다.

감사의 글

본 연구는 한국해양과학기술원(PE99742, 'AI기반 파랑기인 연안재해 모델링 플랫폼 및 해무 예측기술 개발')의 연구비 지원, 연구재단(NRF-2017R1A2B4010108)의 데이터 공유 및 국가초고성능컴퓨팅센터로부터 초고성능컴퓨팅 자원과 기술지원을 받아 수행된 연구성과(KSC-2019-CRE-0100)입니다. 연구수행을 위한 지원에 감사드립니다.

References

[1] R. Holman and M. C. Haller, "Remote sensing of the nearshore," *Annual Review of Marine Science*, vol. 5, pp. 95–113, 2013.

[2] T. Kim, J. Kim, and J. Kim, "Hydrodynamic scene separation from video imagery of ocean wave using autoencoder," *Journal of The Korea Computer Graphics Society*, vol. 25, no. 4, pp. 9–16, 2019.

[3] J. Kim, J. Kim, and S. Shin, "Wave celerity estimation using unsupervised image registration from video imagery," *Journal of KIISE*, vol. 42, no. 10, pp. 1207–1221, 2019.

[4] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. S. Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown, *et al.*, "Tackling climate change with machine learning," *arXiv preprint arXiv:1906.05433*, 2019.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[6] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2482–2491.

[7] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[8] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *Proceedings of ICML workshop on unsupervised and transfer learning*, 2012, pp. 17–36.

[9] J.-H. Kim, J.-Y. Sim, and C.-S. Kim, "Video deraining and desnowing using temporal correlation and low-rank matrix completion," *IEEE Transactions on Image Processing*, vol. 24, no. 9, pp. 2658–2670, 2015.

[10] S. You, R. T. Tan, R. Kawakami, Y. Mukaigawa, and K. Ikeuchi, "Adherent raindrop modeling, detectionand removal in video," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 9, pp. 1721–1733, 2015.

[11] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3855–3863.

[12] L. Shen, Z. Yue, Q. Chen, F. Feng, and J. Ma, "Deep joint rain and haze removal from a single image," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 2821–2826.

[13] L. Wang, B. Yin, A. Guo, H. Ma, and J. Cao, "Skip-connection convolutional neural network for still image crowd counting," *Applied Intelligence*, pp. 1–12, 2018.

[14] Y. Luo, Y. Xu, and H. Ji, "Removing rain from a single image via discriminative sparse coding," in *Proceedings of the*

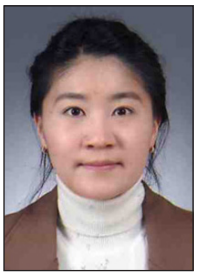
IEEE International Conference on Computer Vision, 2015, pp. 3397–3405.

[15] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

〈저자소개〉

허 동
- 2018년 ~ 현재 경북대학교 인공지능학과 석사과정
- 관심분야: 딥러닝, 인공지능, 영상복원 등
- https://orcid.org/0000-0002-3695-8320

김 재 일
- 2014년 ~ 2016년 삼성전자 책임연구원
- 2016년 ~ 2018년 University of North Carolina at Chapel Hill 박사후연구원
- 2018년 ~ 현재 경북대학교 IT대학 컴퓨터학부 조교수
- 관심분야: 영상 처리, 시계열 데이터 모델링, 예외 검출, 기계학습 등
- https://orcid.org/0000-0002-9799-1773



김 진 아
- 2015년 한국과학기술원 전산학과 박사
- 2005년~현재 한국해양과학기술원 연안재해재난연구센터 책임연구원
- 관심분야: 해양 및 기상자료의 데이터 사이언스(Data Science)
- https://orcid.org/0000-0002-8110-6047

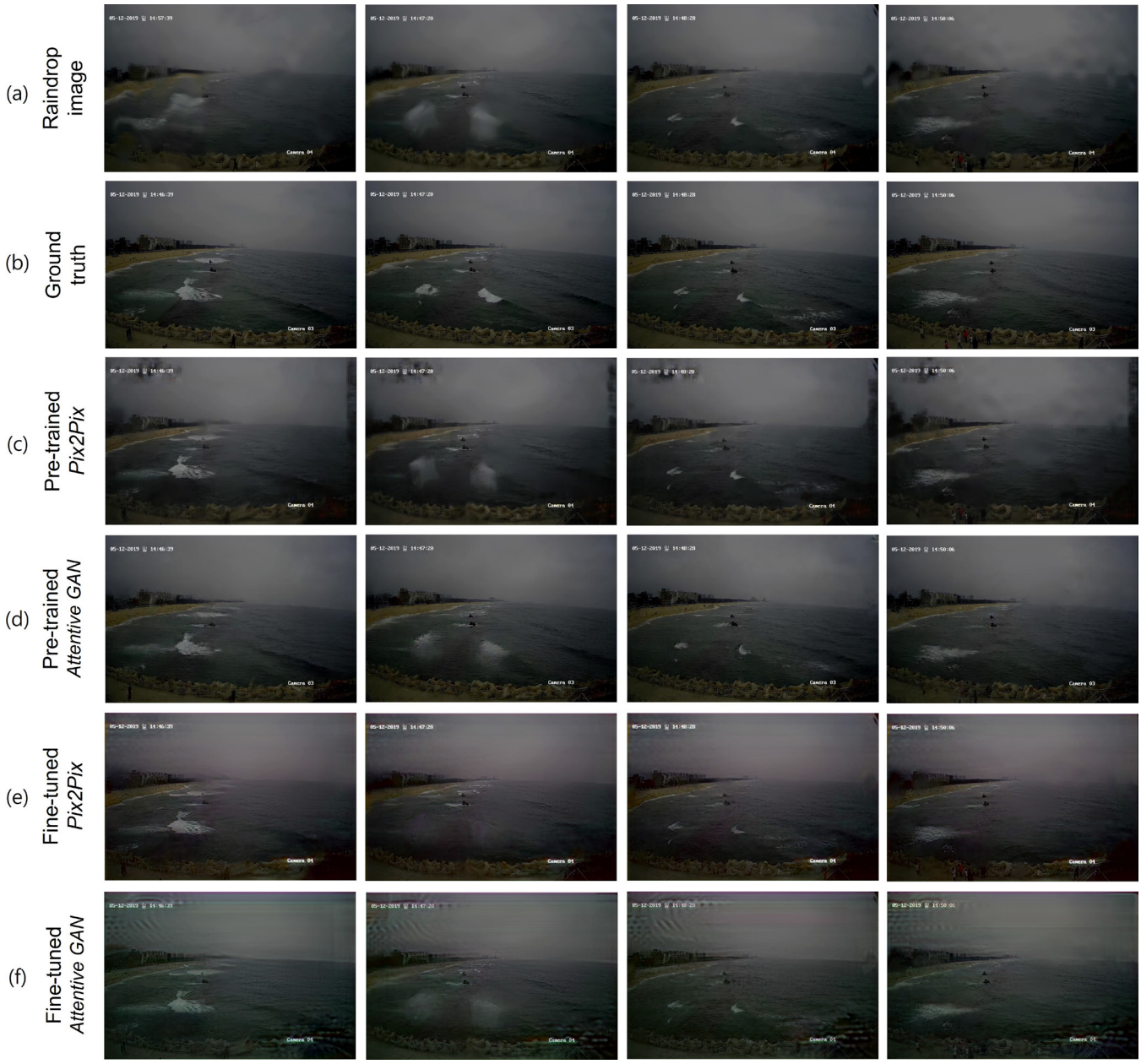


Figure 4: Four samples of results to compare fine-tuned *Pix2Pix* and fine-tuned *Attentive GAN* using *coastal wave raindrop* dataset. From top to bottom: raindrop image (input), ground truth, result of pre-trained *Pix2Pix*, pre-trained*Attentive GAN*, fine-tuned *Pix2Pix*, and fine-tuned *Attentive GAN* for raindrop removal and background information recovery.