

빗줄기 제거를 위한 구조 인식 잔차 네트워크

김주찬*, 손창환**

Structure-Aware Residual Network for Rain Streaks Removal

Ju-Chan Kim*, Chang-Hwan Son**

이 논문은 2020년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임 (P0012769, 2020년 산업혁신인재성장지원사업). 또한 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2020R1A2C1010405)

요 약

본 논문에서는 입력 빗줄기 영상에서 빗줄기를 제거하고 텍스처의 선명도를 강화할 수 있는 빗줄기 제거 기법을 제안한다. 최근 영상 복원 분야에서 큰 주목을 받고 있는 심층 합성곱 신경망 기법을 빗줄기 제거 분야에 적용할지라도 평탄 영역에서의 빗줄기 잔존 문제와 텍스처 선명도 저하 문제는 여전히 현안으로 남아 있다. 따라서 본 논문에서는 입력 빗줄기 영상에서 빗줄기를 효과적으로 제거하고 텍스처의 선명도를 강화할 수 있는 복원 기법을 제안하고자 한다. 특히 입력 빗줄기 영상으로부터 원본 영상의 라플라시안 필터링 결과인 에지 맵을 추정함으로써, 평탄 영역에서 빗줄기를 제거하고 디테일 영역에서 선명도를 강화할 수 있는 영상 구조에 적응적인 구조 인식 잔차 네트워크를 소개하고자 한다. 또한 실험 평가를 통해, 제안한 구조 인식 잔차 네트워크가 소기의 목적을 성취할 수 있으며 기존의 최첨단 기법들보다 정량적 평가에서 더 우수한 성능을 달성할 수 있음을 보이고자 한다.

Abstract

In this paper, rain streaks removal is proposed to remove rain streaks and to enhance the sharpness of textures from input rain images. Even though deep convolutional neural networks, which have recently attracted great attention in the field of image restoration, are applied to rain streaks removal, residual rain streaks on the flat areas and deterioration of texture sharpness in the texture areas remain current issues. Therefore, this paper proposes a restoration method that can effectively remove rain streaks and enhance texture's sharpness from input rain images. In particular, by estimating the edge map, which is the result of Laplacian filtering applied to the original image, the structure-aware residual network that is able to adaptive to image structures is introduced to remove the rain streaks in the flat areas and enhance the sharpness in the detail areas. Moreover, through experimental evaluation, it is confirmed that the proposed structure-aware residual network can achieve the desired purpose and obtain better performance in quantitative evaluation than the existing state-of-the-art methods.

Keywords

rain removal, residual networks, filtering, generative adversarial networks, convolutional neural networks, transfer learning

* 군산대학교 소프트웨어융합공학과
- ORCID: <https://orcid.org/0000-0002-8700-3583>
** 군산대학교 소프트웨어융합공학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0001-7077-3074>

· Received: Sep. 15, 2020, Revised: Oct. 13, 2020, Accepted: Oct. 16, 2020
· Corresponding Author: Chang-Hwan Son
Department of Software Convergence Engineering, Republic of Korea,
Tel.: 82+63-469-8915, Email: cson@kunsan.ac.kr

1. 서 론

최근 심층 학습(Deep learning) 기반의 컴퓨터 비전 기술은 자율 주행 시스템의 성능을 비약적으로 발전시켰다. 하지만 객체 검출이나 이미지캡셔닝과 같은 컴퓨터 비전 기술은 청명하고 쾌적한 날씨 또는 조도가 충분한 환경에서 촬영된 영상을 실험 대상으로 삼고 있다. 폭우, 폭설, 안개와 같은 악천후 환경에서 촬영된 영상은 빗줄기 패턴, 잡음, 헤이즈와 같은 결함을 수반한다. 이로 인해 최종 컴퓨터 비전의 성능은 저하될 수밖에 없다. 예를 들어, 폭우 환경에서 촬영한 영상은 빗줄기 패턴을 포함한다. 이러한 빗줄기 패턴은 객체의 에지의 모양을 변형시키기 때문에 유용한 특징을 추출하는 데 방해요소로 작용할 수 있다[1]. 따라서 다양한 기상환경이나 조도 환경을 반영한 컴퓨터 비전 기술을 개발할 필요성이 있다. 이 논문에서는 폭우 환경에서 촬영된 영상의 화질을 개선하기 위한 빗줄기 제거 기법을 소개하고자 한다. 빗줄기 제거 기법(Rain Removal)이란 폭우 환경에서 촬영된 입력 영상에서 빗줄기 패턴을 제거해서 결함이 없는 선명한 영상을 복원하는 과정을 말한다.

빗줄기 제거 기법은 크게 비디오 기반과 단일 영상 기반으로 구분될 수 있다. 하지만 본 연구에서는 단일 영상 기반의 빗줄기 제거 기법만 다루고자 한다. 단일 영상 기반의 빗줄기 제거 기법은 영상분해(Image decomposition) 접근 방법을 사용한다. 즉, 입력 빗줄기 영상은 빗줄기가 없는 우리가 찾고자 하는 원본 영상인 장면 레이어(Scene layer)와 빗줄기만 포함한 빗줄기 레이어(Rain layer)의 합으로 모델링될 수 있다. 만약 빗줄기 레이어를 추정할 수 있다면, 입력 빗줄기 영상에서 빗줄기 레이어를 빼줌으로써 빗줄기가 없는 원본 영상인 장면 레이어를 획득할 수 있게 된다. 빗줄기 레이어를 추정하기 위해서는 빗줄기 에지의 모양이나 방향에 대한 사전 확률(Prior Probability)을 모델링할 필요가 있다. 이를 위해, 스파스 코딩(Sparse coding)[2]-[4], 가우시안 혼합 모델(Gaussian mixture model)[5], 사전학습(Dictionary learning)[6]과 같은 기계학습 기법들이 개발되었다. 물론 영상 분해 접근 방식외에도, 수작업 특징 추출 기반[7][8] 또는 심층 학습 기반으로

빗줄기 영역을 검출한 후[9], 빗줄기 패턴을 제거하거나 인페인팅(Inpainting) 기법으로 빗줄기 영역의 픽셀 값을 이웃 픽셀로 채우는 방식도 있다.

최근 빗줄기 제거 기법은 심층 합성곱 신경망(DCNN, Deep Convolutional Neural Networks)[10][11], 생성적 적대 신경망(Generative Adversarial Networks: GAN)[12]-[14]과 같은 심층 학습 기법을 사용해서 입력 빗줄기 영상과 원본 영상간의 매핑 관계를 학습하는 형태를 취하고 있다. 심층 합성곱 신경망은 합성곱 레이어(Conv, Convolution layer), 배치 정규화(BN, Batch Normalization), 정류선형유닛(ReLU, Rectified Linear Unit) 등과 같은 레이어를 연속적으로 깊이 쌓은 후, 입력 빗줄기 영상과 출력 원본 영상 간의 관계를 모델링한다. 그리고 내리막경사법(Gradient descent methods)과 같은 최적화 기법을 통해 각 레이어의 파라미터를 추정하여 입력 영상과 출력 영상 간의 매핑 관계를 추정한다. 깊은 레이어를 학습함으로써, 저수준의 특징 추출뿐만 아니라 고수준의 특징까지도 뽑아낼 수 있다. 이 결과로, 빗줄기가 잘 제거된 고품질의 영상을 제공할 수 있으며 종래의 저수준 특징을 묘사하기 위한 사전확률 모델링 기반의 기법들보다 더 탁월한 성능을 낼 수 있다.

생성적 적대 신경망은 생성망과 적대망이라는 두 종류의 네트워크를 학습한다. 생성망은 입력 빗줄기 영상에서 빗줄기를 제거해서 원본 영상과 유사한 가짜 영상을 만드는 역할을 한다. 반면 적대망은 생성망이 만든 가짜 영상과 원본 영상을 판별하는 역할을 한다. 생성망은 적대망이 가짜 영상을 구별할 수 없도록 속이려고 학습이 되고 반면 적대망은 생성망의 가짜 영상을 잡아 내도록 학습이 진행된다. 이는 생성자와 대립자의 적대적인 관계로 이해할 수 있으며 최소최대학습(Minmax learning)을 통해 최적의 해에 수렴하게 된다. 즉, 내시 평형(Nash equilibrium)에 도달하게 된다. 생성적 적대 신경망은 심층 합성곱 신경망처럼 입력 영상에서 출력 영상을 직접적으로 매핑하는 판별 모델(Discriminative model)과는 달리, 출력 영상의 확률 분포를 추정하는 생성 모델(Generative model)에 해당한다. 이러한 생성 모델은 현재 음성, 텍스트와 같은 다양한 응용도메인에 적용되고 있고 특히 영상 생성 분야에도

탁월한 성능을 거두고 있다. 대표적인 기법에는 Pix2Pix[12], CycleGAN[13] 등이 있다.

기존의 심층 합성곱 신경망이나 생성적 적대 신경망 기반의 빗줄기 제거 기법들은 화질 측면에서 상당한 개선을 이룬 것은 사실이다. 하지만 여전히 텍스처나 에지와 같은 영역에서 선명도가 떨어지거나 빗줄기가 완전히 제거되지 않는 문제점이 현안으로 남아 있다. 따라서 본 논문에서는 이러한 문제점을 개선하기 위해, 새로운 "구조 인식 잔차 네트워크"(SARN, Structure-Aware Residual Network)을 제시하고자 한다. 제안한 네트워크는 입력 빗줄기 영상에서 원본 영상에서 라플라시안(Laplacian) 필터를 통과시킨 에지 맵을 추정하는 라플라시안 서브네트워크(LS, Laplacian Subnetwork)와 빗줄기 영상에서 잔차 영상을 복원하는 구조 인식 잔차 복원 서브네트워크(SARRS, Structure-Aware Residual Reconstruction Subnetwork)로 구성된다. 특히 라플라시안 서브네트워크를 통해 원본 영상의 에지를 추정함으로써, 구조 인식 잔차 복원 서브네트워크가 빗줄기 영상에서 원본 영상의 에지와 평탄 영역을 정보를 활용할 수 있게 된다. 이로 인해, 평탄 영역에서 빗줄기 패턴을 제거하고 디테일 영역에서는 에지의 선명도를 보존할 수 있는 효과를 얻을 수 있다. 즉, 영상 구조(Image structure)에 적응적으로 학습할 수 있는 심층 합성곱 신경망을 구현할 수 있다.

이 논문의 구성은 다음과 같다. 먼저 제 2장에서는 빗줄기 제거를 위한 기존의 다양한 심층 학습 모델을 소개한 후, 제 3장에서는 제안한 구조 인식 잔차 네트워크 기반의 빗줄기 제거 기법에 대해 자세히 설명하고자 한다. 제 4장에서는 제안한 기법의 성능을 평가 및 분석하고 마지막 제 5장에서 결론을 맺고자 한다.

II. 관련 연구

2.1 직접 매핑 네트워크

직접 매핑 네트워크(Direct mapping networks)는 심층 합성곱 신경망을 학습할 때, 입력 빗줄기 영상을 출력 원본 영상으로 매핑하는 방식을 말한다. 이는 심층 합성곱 신경망의 출력 값의 범위가 원본

영상의 픽셀 값의 범위로 매핑 된다는 것을 말한다. 일반적인 신경망 학습 방식에 해당한다. 직접 매핑 네트워크를 수식으로 표현하면 아래와 같다.

$$L = \sum_i \|h(X_i) - Y_i\|_F^2 \quad (1)$$

여기서 X_i 와 Y_i 는 i 번째 훈련 영상인 빗줄기 영상과 원본 영상을 가리킨다. h 는 심층 합성곱 신경망을 의미하고 L 은 총 손실(Loss)이다. 심층 합성곱 신경망은 L 의 값이 작아지도록 신경망의 파라미터가 갱신된다. 다시 말하면, h 를 통과한 추정치, 즉 빗줄기가 제거된 영상인 $h(X_i)$ 가 원본 영상인 Y_i 와 같아지도록 학습이 된다.

2.2 잔차 네트워크

직접 매핑 네트워크와는 달리, 잔차 네트워크(Residual networks)는 심층 합성곱 신경망을 학습할 때, 입력 빗줄기 영상을 원본 영상에서 빗줄기 영상을 뺀 잔차 영상(Residual image)으로 매핑하는 방식을 말한다[10,11]. 기존의 직접 매핑 네트워크는 출력 영상의 값의 범위가 상대적으로 넓게 분포하기 때문에 신경망의 정확도가 떨어질 수 있다. 잔차 네트워크는 이러한 직접 매핑 네트워크가 지닌 문제를 해결하기 위해, 최종 신경망의 출력 값의 범위를 좁혀서 학습을 진행한다. 출력 값의 범위를 좁히기 위해, 원본 영상을 그대로 사용하는 것이 아니라 원본 영상에서 빗줄기 영상을 뺀 잔차 영상을 사용한다. 잔차 네트워크의 수식은 다음과 같이 표현된다.

$$L = \sum_i \|h(X_i) - (Y_i - X_i)\|_F^2 \quad (2)$$

학습 단계에서는 손실 L 을 최소화하기 위해 신경망의 출력 값인 $h(X_i)$ 와 잔차 영상인 $Y_i - X_i$ 를 같아지도록 학습을 한다. 테스트 단계에서는 빗줄기가 제거된 영상을 얻기 위해, 신경망을 통과한 잔차 영상 $h(X_i)$ 에 입력 빗줄기 영상 X_i 를 더해줌으로써 최종 결과 영상을 만들어 낸다. 잔차 네트워크는 출력 값의 분포를 좁혀줌으로써, 최종 신경망이 추

정해야 될 값을 좀 더 수월하게 찾을 수 있도록 해준다. 이를 통해 학습의 정확도를 높이는 효과를 유도할 수 있다.

2.3 디테일 네트워크

디테일 네트워크(Detail networks)는 잔차 네트워크의 변종으로서, 신경망의 입력으로 빗줄기 영상을 그대로 사용하는 것이 아니라 빗줄기 영상의 디테일 레이어를 사용한다[11]. 출력 영상은 잔차 네트워크와 동일하게 원본 영상에서 빗줄기 영상을 뺀 잔차 영상을 사용한다. 즉, 디테일 네트워크는 빗줄기 영상의 디테일 레이어와 잔차 영상을 학습 영상으로 사용한다. 디테일 네트워크에 관한 수식은 다음과 같이 표현된다.

$$X_i = X_{D,i} + X_{B,i} \quad (3)$$

$$L = \sum_i \| h(X_{D,i}) - (Y_i - X_i) \|_F^2 \quad (4)$$

여기서 $X_{D,i}$ 와 $X_{B,i}$ 는 입력 빗줄기 영상 X_i 의 디테일 레이어와 베이스 레이어를 각각 의미한다. 여기서 베이스 레이어 $X_{B,i}$ 는 입력 빗줄기 영상 X_i 에 스무딩 필터를 적용해서 만든 영상으로써, 밝기의 변화가 거의 없는 평탄 영역으로 구성된다. 반면에 디테일 레이어 $X_{D,i}$ 는 X_i 에서 $X_{B,i}$ 를 뺀 영상으로써, 객체의 에지, 윤곽선, 그리고 빗줄기 패턴을 포함한 디테일 성분에 해당한다.

수식 (4)에서 보듯이, 디테일 네트워크는 신경망의 입력으로 디테일 레이어 $X_{D,i}$ 를 받아서 출력으로 잔차 영상 $Y_i - X_i$ 를 내보낸다. 잔차 네트워크와의 차이점은 신경망의 입력으로 디테일 레이어를 사용한다는 점이다. 잔차 네트워크는 입력과 출력 영상이 모두 배경이 제거된 디테일 성분으로 구성된다는 것을 알 수 있다. 또한 디테일 레이어는 대부분의 값이 '0'으로 채워지기 때문에 스파스(Sparse)한 특징을 지니게 된다. 이렇게 배경이 제거된 입출력 영상의 특성과 스파스한 특징은 신경망의 성능을 한 층 더 개선할 수 있는 것으로 보고되고 있다.

III. 제안한 구조 인식 잔차 네트워크 기반의 빗줄기 제거

3.1 제안한 접근 방법

본 논문에서는 기존의 심층 합성곱 신경망 기반의 빗줄기 제거 기법에서 나타나는 디테일 정보 소실과 평탄 영역에서의 빗줄기 잔존 문제를 동시에 해결할 수 있는 "구조 인식 잔차 네트워크"를 제안하고자 한다. 기존의 잔차 및 디테일 네트워크는 출력 영상으로 잔차 영상을 활용함으로써, 출력 값의 범위를 좁게 만들어 복원 영상의 화질을 개선할 수 있었다. 하지만 여전히 디테일 선명도 저하와 빗줄기 잔존 문제에서 벗어날 수 없다. 왜냐하면 빗줄기 패턴으로 인해 객체가 지닌 디테일 정보가 손실된 상태에서, 평탄 영역에 원치 않는 빗줄기 패턴으로 인해 원본 영상의 디테일을 선명하게 복구하고 평탄 영역에서 빗줄기를 제거하기가 쉽지 않기 때문이다. 이러한 문제를 해결하기 위해, 제안한 기법에서는 라플라시안 서브네트워크를 새롭게 도입하여 객체의 디테일 정보와 평탄 영역의 정보를 추정하여, 추후 영상 복원을 담당하는 구조 인식 잔차 복원 서브네트워크가 영상 구조에 적응적으로 학습될 수 있는 네트워크를 설계하고자 한다.

따라서 제안한 구조 인식 잔차 네트워크는 두 종류의 서브네트워크로 구성된다. 첫 번째 서브네트워크는 "라플라시안 네트워크"로써, 입력 빗줄기 영상으로부터 원본 영상에 라플라시안 필터를 통과시킨 에지 맵을 추정한다. 두 번째 서브네트워크는 "구조 인식 잔차 복원 네트워크"로써, 입력 빗줄기 영상과 추정된 에지 맵을 입력으로 받아서 최종 빗줄기가 제거된 영상을 복원한다. 특히 라플라시안 서브네트워크를 통해 원본 영상의 에지를 추정함으로써, 구조 인식 잔차 복원 서브네트워크가 빗줄기 영상에서 원본 영상의 에지와 평탄 영역의 정보를 활용할 수 있게 된다. 이로 인해, 평탄 영역에서 빗줄기 패턴을 제거하고 디테일 영역에서는 에지의 선명도를 보존할 수 있는 효과를 유도할 수 있다. 다시 말하면, 영상 구조(Image structure)에 적응적인 구조 인식 잔차 네트워크를 실현할 수 있다.

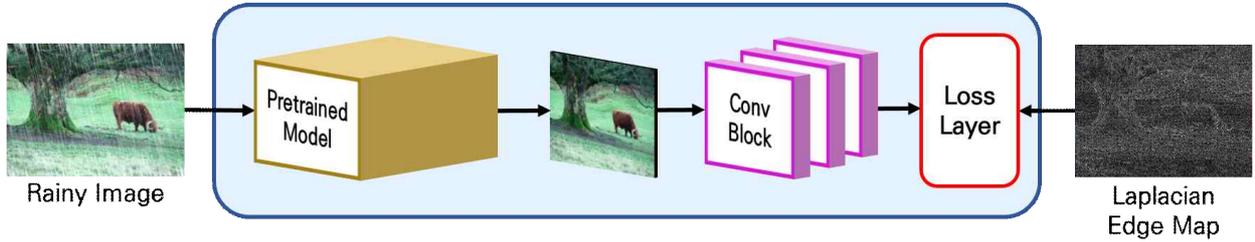


그림 1. 제안한 라플라시안 서브네트워크 아키텍처
Fig 1. Architecture of the proposed Laplacian subnetwork

3.2 학습 데이터 수집

제안한 라플라시안 서브네트워크와 구조 인식 잔차 복원 서브네트워크를 학습하기 위해서는 총 3종류의 훈련 영상 셋이 필요하다. 즉, 빗줄기 영상, 빗줄기가 없는 원본 영상, 그리고 원본 영상에 라플라시안 필터를 적용한 에지 영상이다. 원본 영상은 객체 검출 분야에 많이 사용되는 PASCAL VOC (PASCAL Visual Object Classes)[15] 데이터 셋을 사용했다. 그리고 라플라시안 필터를 PASCAL VOC의 원본 영상에 각각 적용하여 에지 영상을 생성하였다. 빗줄기 영상은 원본 영상에 빗줄기 패턴을 추가함으로써, 생성이 가능하다. 빗줄기 영상 생성 과정은 다음과 같이 모델링 된다[5,6].

$$X_i = Y_i + S_i \quad (5)$$

여기서 X_i , Y_i , S_i 는 i 번째 빗줄기 영상, 빗줄기가 없는 원본 영상, 그리고 빗줄기 패턴만 포함한 영상을 가리킨다. 수식 (5)에서 보듯이, 최종 빗줄기 영상은 원본 영상과 빗줄기 패턴만 포함한 영상을 더함으로써 모델링된다. 여기서 빗줄기 패턴만 포함한 영상 S_i 는 랜덤 잡음을 생성한 후, 모션 블러링 (Motion Blurring)을 적용하여 빗줄기 패턴을 만들어 낼 수 있다. 이 과정은 포토샵과 같은 영상 편집 툴을 사용할 수 있고 아니면 프로그래밍으로도 구현 가능하다.

본 연구에서는 학습될 네트워크가 영상 스케일에 강인한 특성을 보유하기 위해, 수집된 훈련 영상들에 데이터 확충(Data augmentation) 과정을 수행하였다. 다시 말하면, 수집된 훈련 영상 각각에 대해

1.0, 0.9, 0.8, 0.7, 0.6배의 비율로 영상의 크기를 가변하면서 데이터 확충 작업을 완료하였다.

일반적으로 신경망 학습에 사용될 훈련 영상을 수집된 영상 그 자체가 아닌, 영상에서 랜덤하게 추출된 패치들을 사용한다. 그러한 이유는 원본 영상의 차원이 너무 크기 때문에 학습 속도와 메모리 용량 문제가 발생하기 때문이다. 빗줄기 제거 기법에서 사용되는 패치의 크기는 대략 64×64 픽셀 정도에 해당한다.

3.3 라플라시안 서브네트워크 학습

그림 1은 제안한 라플라시안 서브네트워크의 아키텍처를 보여주고 있다. 라플라시안 필터를 통과한 에지 맵을 빗줄기 영상에서 바로 추정할 수 있지만, 제안한 기법에서는 전이 학습(Transfer learning)을 도입하였다. 그러한 이유는 빗줄기가 제거된 초기 영상에서 에지 맵을 찾는 것이 파라미터의 초기화를 쉽게 설정할 수 있고 학습 효과를 높일 수 있기 때문이다. 실제로 라플라시안 필터는 3×3 과 같은 하나의 필터로 구현된다. 이는 빗줄기가 제거된 초기 영상이 주어진 경우, 소수의 합성곱 신경망으로 추정이 가능하다는 뜻이다.

제안한 라플라시안 서브네트워크는 그림 1에서 보듯이, 사전 학습된 모델(Pretrained model)에서 손실 레이어를 제거한 뒤, (합성곱, 배치 정규화) 또는 (합성곱, 배치 정규화, 정류선형유닛)으로 구성된 합성곱 블록(Convolution Block: Conv Block) 3개를 연속해서 쌓은 구조로 되어 있다. 그리고 맨 마지막에는 손실 레이어가 달려 있다. 손실 레이어는 추정된 에지 맵과 원본 라플라시안 에지 맵의 오차를 계산

한다. 본 연구에서는 사용된 오차의 종류는 평균제곱오차(Mean squared errors)이다.

사전 학습된 모델은 ResNet 구조 형태를 취하고 있으며 총 26개의 합성곱 블록과 12개의 스킵 연결로 구현되었다. 참고로 사전 학습된 모델은 라플라시안 서브네트워크가 학습이 진행되는 동안 그 파라미터의 값이 변경된다. 즉 사전 학습된 모델의 초기 영상은 추후 공정한 평가를 위해 영상 복원을 위해 사용되지 않는다. 사전 학습된 모델은 오직 라플라시안 에지 맵을 효과적으로 추정하기 위해서만 사용되었다는 뜻이다.

3.4 구조 인식 잔차 복원 서브네트워크 학습

그림 2는 제안한 구조 인식 잔차 복원 서브네트워크의 아키텍처를 보여주고 있다. 제안한 구조 인식 잔차 복원 네트워크는 ResNet처럼 합성곱 블록을 연속해서 쌓고 스킵 연결로 구성된 형태를 취하고 있다. 하지만 기존의 잔차 네트워크와는 달리, 제안한 구조 인식 잔차 복원 서브네트워크는 두 종류의 입력을 사용한다. 즉 라플라시안 서브네트워크에서 추정한 라플라시안 에지 맵과 입력 빗줄기 영상을 함께 사용한다. 기존의 잔차 네트워크는 빗줄기 영상만을 입력으로 사용하기 때문에 학습하는 동안 어떤 부분이 객체의 디테일 영역인지 어떤 부분이 빗줄기가 없는 평탄한 영역인지를 자기 스스로 학습하기가 어렵다. 따라서 제안한 구조 인식 잔

차 복원 서브네트워크에서는 이러한 문제를 해소하기 위해, 라플라시안 서브네트워크에서 제공한 에지 맵을 추가적인 정보로 활용한다. 이를 통해, 평탄 영역에서 빗줄기 패턴을 제거하고 디테일 영역에서는 에지의 선명도를 강화할 수 있는 학습 효과를 유도할 수 있다. 즉, 영상 구조에 적응적인 구조 인식 네트워크를 달성할 수 있다. 제안한 기법은 최근에 개발되고 있는 주의집중모델(Attention Models) [16,17]의 한 종류로 볼 수 있다. 즉, 선명도 강화를 위해 입력 영상에서 어떤 부분을 주의 깊게 봐야 되는지를 학습하기 위해, 라플라시안 서브네트워크를 도입해서 그 목적을 효과적으로 달성한 경우라 볼 수 있다.

3.5 빗줄기 제거 과정

앞서 말한 두 종류의 서브네트워크가 학습이 완료된 후, 입력 빗줄기 영상에서 빗줄기 패턴을 제거하는 과정은 그림 3과 같다. 먼저, 빗줄기 영상을 라플라시안 서브네트워크에 통과시켜 라플라시안 에지 맵을 추정한 뒤, 입력 빗줄기 영상과 라플라시안 에지 맵을 스택으로 쌓은 후, 구조 인식 잔차 복원 서브네트워크에 통과시킨다. 이때 구조 인식 잔차 복원 서브네트워크의 출력은 잔차 영상이다. 따라서 최종 빗줄기가 제거된 복원 영상을 생성하기 위해서는 입력 빗줄기 영상과 잔차 영상을 더해줘야 된다.

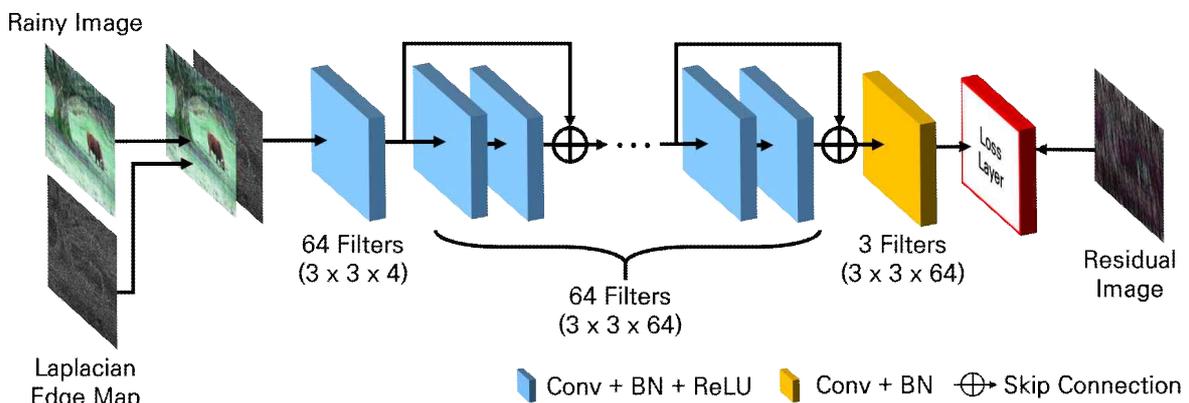


그림 2. 제안한 구조 인식 잔차 복원 서브네트워크 아키텍처
 Fig. 2. Architecture of the proposed structure-aware residual reconstruction subnetwork

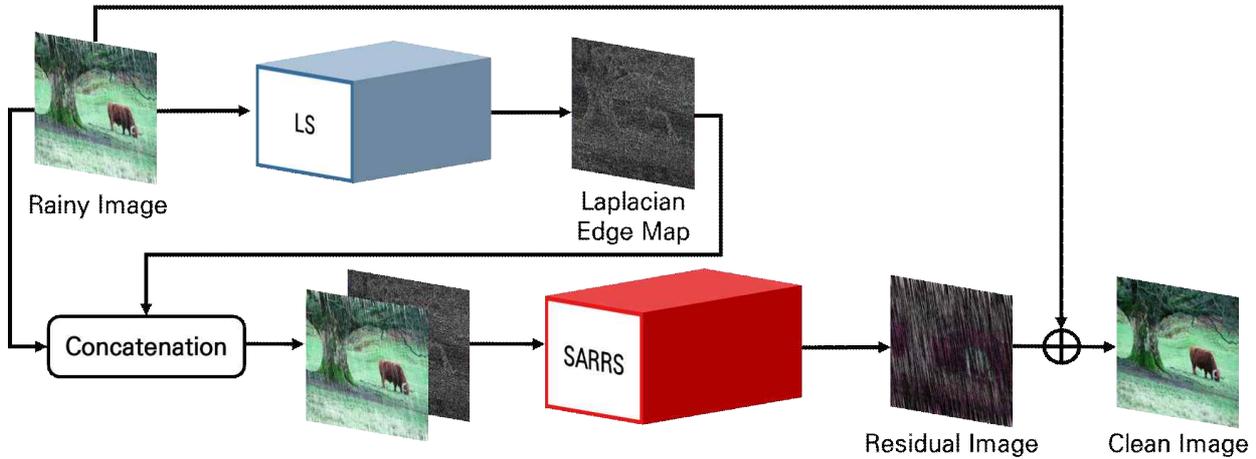


그림 3. 제안한 빗줄기 제거 과정
Fig. 3. Process of the proposed rain streaks removal

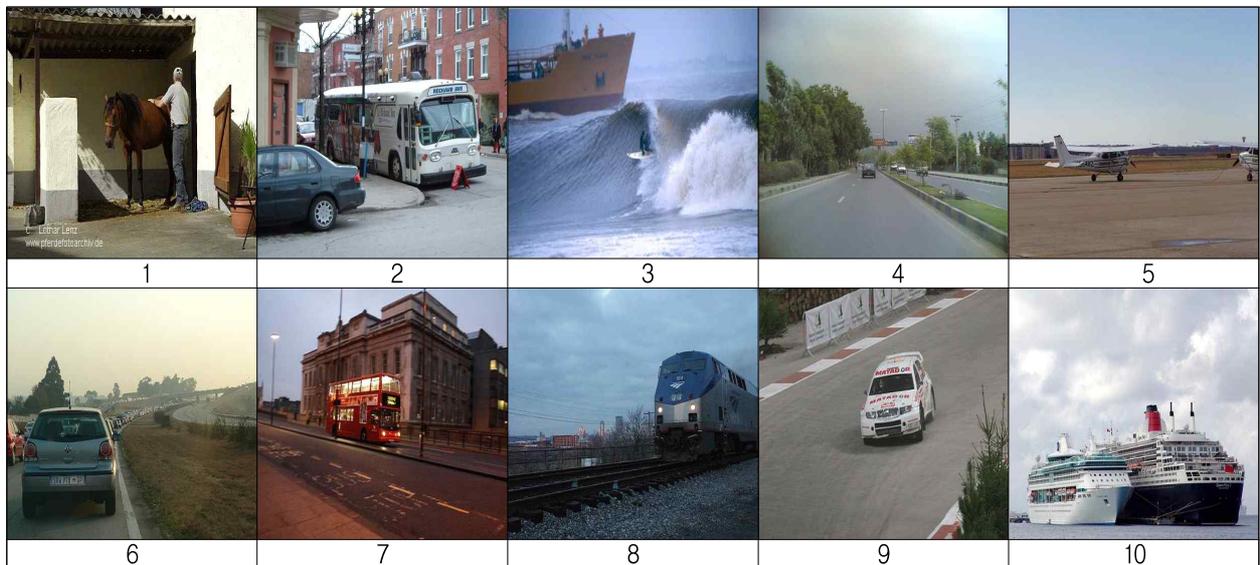


그림 4. 정량적 평가에 사용된 테스트 영상 예시
Fig. 4. Examples of test images used for quantitative evaluation

IV. 실험 및 결과

4.1 실험 환경

정량적 성능 평가를 위해, 본 연구에서는 Pascal VOC2007 오픈 데이터셋을 사용했다. Pascal VOC2007 영상중에서 총 2,000장을 무작위로 선택했다. 그리고 훈련 영상과 테스트 영상 집합의 비율을 3:1로 설정했다. 그림 4는 테스트 영상으로써, 총 500장으로 구성된 테스트 집합의 일부 영상에 해당한다. 그림에서 보듯이 테스트 영상이 다양한 영상 구조, 예

를 들어 라인, 곡선 등을 포함하고 있는 것을 볼 수 있다. 그리고 테스트 영상을 구분하기 위해 영상 하단에 인덱스 번호를 붙여 놓았다. 비교 대상으로 최첨단 성능을 보유한 잔차 네트워크[10], 생성적 적대 신경망의 한 종류인 Pix2Pix[12], 디테일 네트워크[11] 기법을 선택했다.

공평한 실험을 위해 기존의 잔차 네트워크, 디테일 네트워크, 그리고 제안의 방법에서 사용된 신경망의 최적화 기법은 모두 확률적 경사 하강법(Stochastic Gradient Descent)을 사용했다. 또한 배치 크기, 학습률, 총 에폭의 수도 동일하게 설정했다.

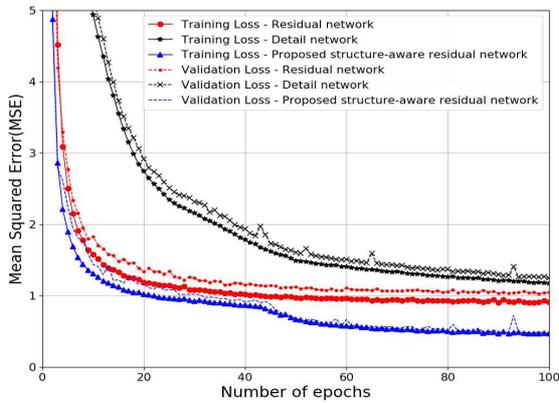


그림 5. 에폭의 수에 따른 기존 방법들과 제안한 방법의 총 손실 변화 비교

Fig. 5. Comparison of the total loss change between the existing methods and the proposed method, according to the number of epochs

표 1. 테스트 집합에 대한 정량적인 화질 평가

Table 1. Quantitative image quality evaluation for test set

Methods	PSNR	SSIM
Residual Network [10]	33.8225	0.9247
Detail Network [11]	31.1713	0.9130
Pix2Pix [12]	29.5182	0.9136
Proposed Structure-Aware Residual Network	34.1123	0.9372

즉, 배치 크기는 64, 총 에폭(Epochs)은 100회이고 학습률(Learning rate)은 10^{-5} 에서 시작해서 10^{-6} 까지 가변하도록 설정되었다. 그리고 제안한 합성곱 블록에 사용된 필터의 크기는 3×3, 필터 개수는 64개로 설정되었다.

제안한 방법과 기존 방법의 성능을 평가하기 위해, 영상 복원 분야에서 널리 사용되는 최대 신호 대 잡음비(PSNR, Peak Signal-to-Noise Ratio)와 구조적 유사성(SSIM, Structural Similarity) [18]을 평가적으로 사용하였다. 최대 신호 대 잡음비는 원본 영상과 복원 영상 간의 픽셀 값의 오차가 얼마나 나는지를 판별하는 척도이며 반면에 구조 유사성은 두 영상 간의 라인, 곡선, 패턴의 모양이 얼마나 유사한지를 반영할 수 있는 척도이다. 두 척도 모두 수치가 높을수록 복원 영상과 원본 영상이 유사하다는 것을 의미한다. 제안한 기법은 MatConvNet 매트랩 딥 러닝 툴로 구현되었고 2080Ti GPU 2개로 구성된 워크스테이션에서 실행되었다.

4.2 실험 결과

그림 5는 기존의 잔차 네트워크, 디테일 네트워크, 그리고 제안한 구조 인식 잔차 네트워크 기법에 대한 총 손실 추이를 보여준 그래프이다. 그림에서 보듯이, 제안한 구조 인식 잔차 네트워크 기법이 동일한 에폭 수에 대해 훈련 집합(Training loss)과 검증 집합(Validation loss)에 대해 총 손실을 가장 빠르게 감소시키는 것을 볼 수 있다. 이는 라플라시안 서브네트워크를 통해 추정된 라플라시안 에지 맵을 활용하면 총 오차를 더 빨리 감소할 수 있음을 의미한다. 또한 구조 인식 잔차 복원 서브네트워크에 입력 빗줄기 영상에서 객체의 에지가 어디에 존재하는지에 관한 정보를 제공함으로써, 영상 구조에 적응적인 학습 효과를 얻을 수 있음을 보여 준다.

그림 6-9는 기존의 잔차 네트워크, Pix2Pix, 디테일 네트워크, 그리고 제안한 기법을 적용한 결과 영상을 보여주고 있다. 먼저 그림 6의 빨간색 박스 영역에서 보면, 기존의 잔차 네트워크, Pix2Pix, 디테일 네트워크는 벽면과 같은 평탄한 영역에서 빗줄기의 잔재가 남아 있는 것을 볼 수 있다. 반면 제안한 구조 인식 잔차 네트워크는 평탄 영역의 빗줄기 잘 제거한 것을 볼 수 있고 또한 텍스처의 선명도나 형태의 복원 정도가 기존의 방법보다 더 나은 것을 확인할 수 있다. 그림 7에서는 제안한 기법이 선박 하단부의 라인과 텍스처를 가장 잘 표현한 것을 볼 수 있다. 또한 빗줄기가 말끔히 잘 제거된 것을 볼 수 있다. 그림 8과 9에서도 나무의 텍스처, 버스 외벽 및 타이어의 선명도를 개선하고 빗줄기 잔재 문제를 잘 해결한 것을 볼 수 있다.

그림 10은 라플라시안 서브네트워크로 추정된 에지 맵을 보여 주고 있다. 원본 라플라시안 에지 맵과 비교했을 때 거의 유사한 것을 볼 수 있다. 즉, 라플라시안 서브네트워크는 입력 빗줄기 영상으로부터 원본 영상의 에지 영역과 평탄 영역을 구분할 수 있는 에지 맵을 생성할 수 있음을 확인해 준다.

표 1은 테스트 집합에 대한 최대 신호 대 잡음비와 구조적 유사성에 대한 정량적 평가 결과를 보여 주고 있다. 정량적 평가에 사용된 영상은 총 500장이다. 표 1에 기입된 값은 총 500장 영상에 대한 PSNR과 SSIM의 평균치를 나타낸다.

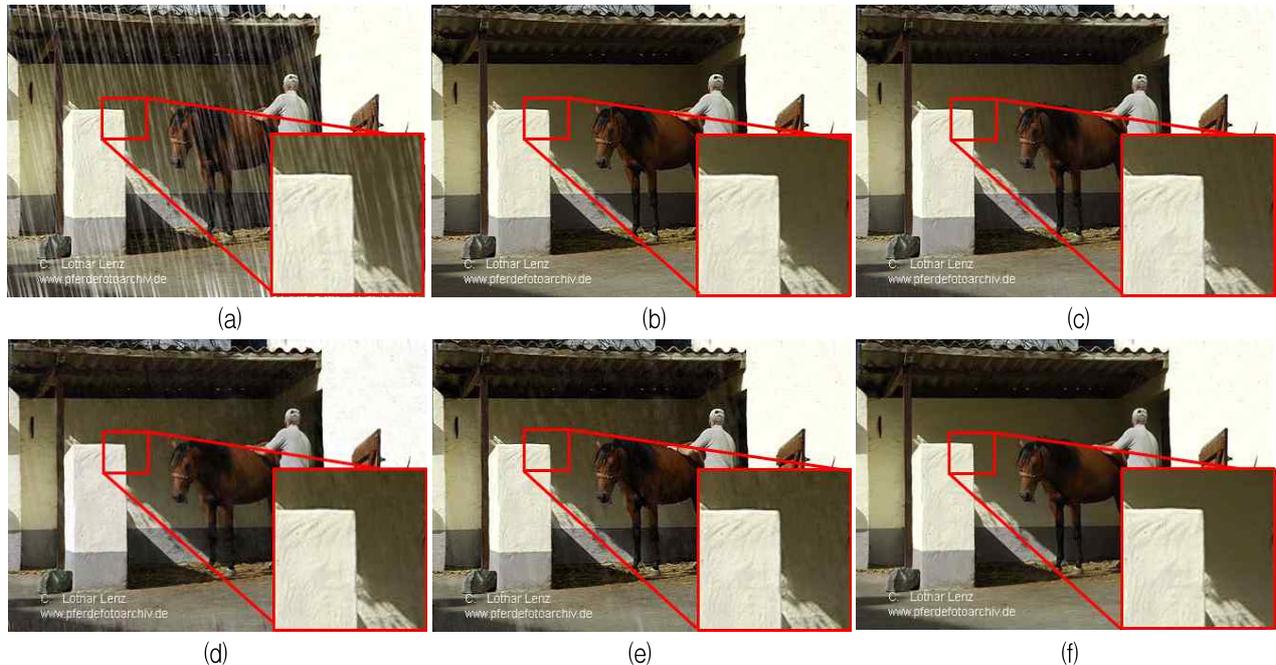


그림 6. 결과 영상; (a) 빗줄기 영상, (b) 원본 영상, (c) 잔차 네트워크[10], (d) 디테일 네트워크[11], (e) Pix2Pix[12], (f) 제안한 기법

Fig. 6. Resulting images; (a) Rain image, (b) Original image, (c) Residual network[10], (d) Detail network[11], (e) Pix2Pix[12], and (f) Proposed method

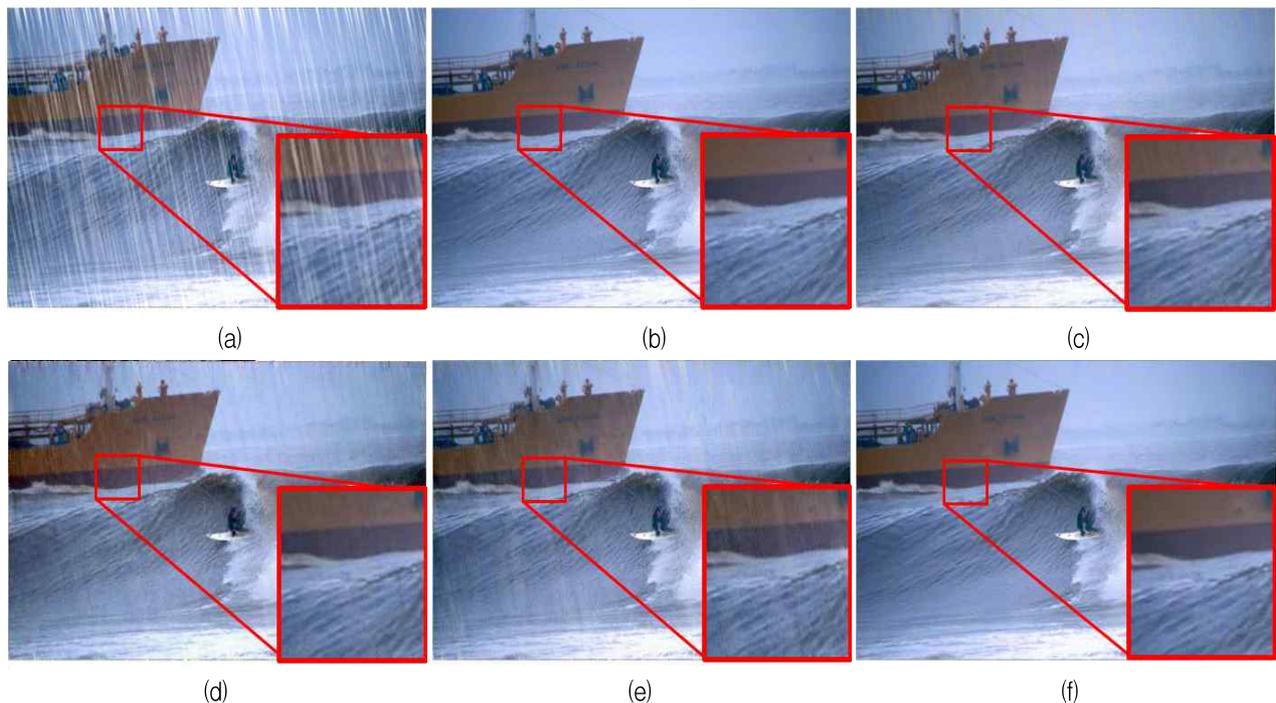


그림 7. 결과 영상; (a) 빗줄기 영상, (b) 원본 영상, (c) 잔차 네트워크[10], (d) 디테일 네트워크[11], (e) Pix2Pix[12], (f) 제안한 기법

Fig. 7. Resulting images; (a) Rain image, (b) Original image, (c) Residual network[10], (d) Detail network[11], (e) Pix2Pix[12], and (f) Proposed method

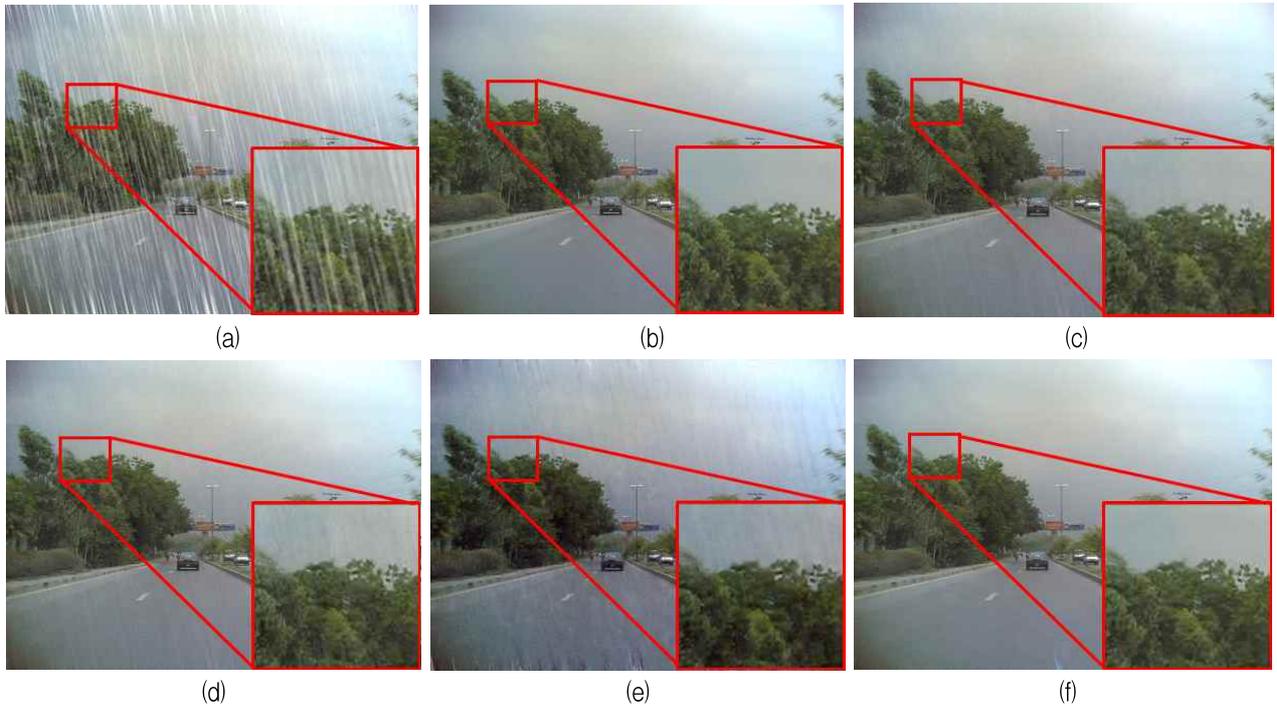


그림 8. 결과 영상; (a) 빗줄기 영상, (b) 원본 영상, (c) 잔차 네트워크[10], (d) 디테일 네트워크[11], (e) Pix2Pix[12], (f) 제안한 기법

Fig. 8. Resulting images: (a) Rain image, (b) Original image, (c) Residual network[10], (d) Detail network[11], (e) Pix2Pix[12], and (f) Proposed method

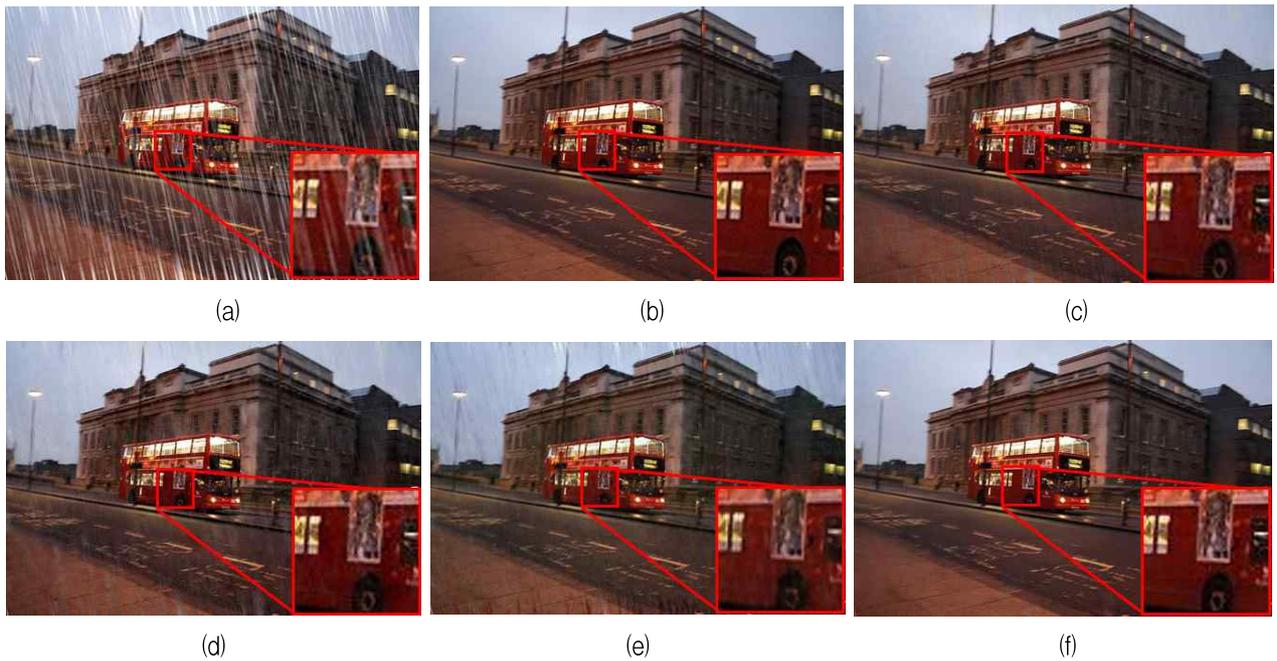


그림 9. 결과 영상; (a) 빗줄기 영상, (b) 원본 영상, (c) 잔차 네트워크[10], (d) 디테일 네트워크[11], (e) Pix2Pix[12], (f) 제안한 기법

Fig. 9. Resulting images: (a) Rain image, (b) Original image, (c) Residual network[10], (d) Detail network[11], (e) Pix2Pix[12], and (f) Proposed method

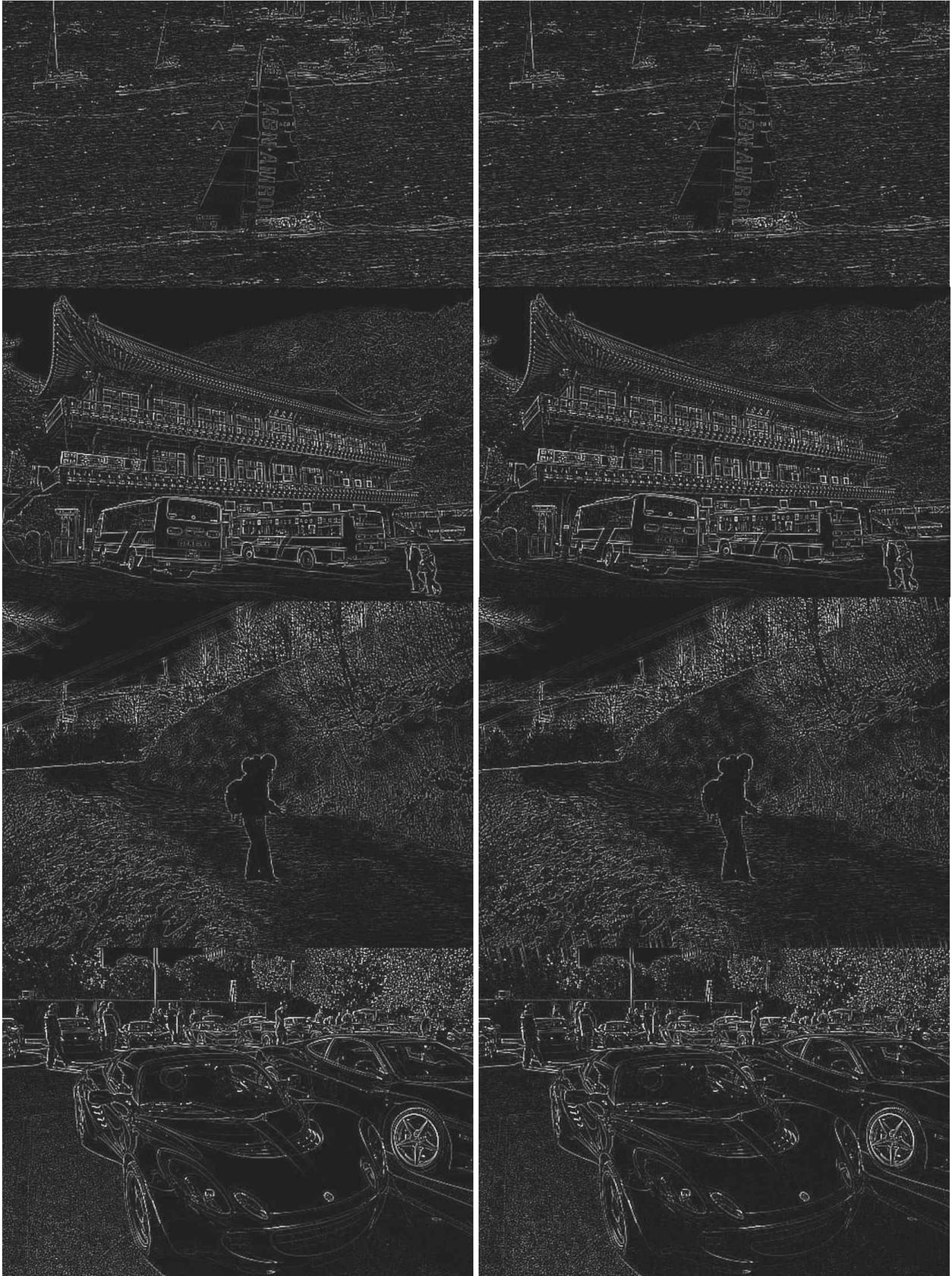


그림 10. 원본 라플라시안 에지 맵 (좌측 열), 추정된 라플라시안 에지 맵(우측 열)

Fig. 10. Original Laplacian edge maps (left column), estimated Laplacian edge map (right column)

표 2. 그림 4의 10장의 테스트 영상에 대한 정량적인 화질 평가
 Table 2. Quantitative image quality evaluation for 10 test images in Fig. 4

Test Images		1	2	3	4	5	6	7	8	9	10	AVG.
Methods												
Residual Network [10]	PSNR	34.196	31.010	33.497	33.955	35.360	31.636	32.023	32.671	29.888	32.455	32.454
	SSIM	0.9473	0.8995	0.9521	0.8697	0.9560	0.9177	0.8814	0.8819	0.8629	0.7129	0.8881
Detail Network [11]	PSNR	31.281	29.183	31.710	31.887	32.553	29.699	29.808	31.490	28.924	30.459	30.699
	SSIM	0.9162	0.8737	0.9435	0.8153	0.9338	0.8957	0.8474	0.8422	0.8520	0.6648	0.8585
Pix2Pix [12]	PSNR	29.315	28.597	29.845	23.894	30.694	29.020	29.271	30.515	28.027	27.676	28.685
	SSIM	0.8395	0.8829	0.9207	0.7604	0.9357	0.8974	0.7946	0.8596	0.8584	0.6280	0.8377
Proposed Method	PSNR	34.597	31.235	35.067	34.113	36.869	32.334	32.305	32.978	30.213	32.818	33.253
	SSIM	0.9562	0.9081	0.9627	0.8867	0.9683	0.9295	0.8962	0.8924	0.8666	0.7234	0.8990

표 2는 그림 4에 제시된 테스트 영상 샘플 각각에 대한 PSNR과 SSIM 결과치를 보여준다. 표 1에서 보듯이 제안한 기법이 기존의 방법보다 PSNR 평가에서 약 0.29dB 더 개선된 결과를 얻었으며, SSIM 평가에서도 약 0.0125 더 개선된 점수를 획득하였다. 이는 제안한 기법이 기존의 방법에 비해, 원본 영상과의 오차가 더 작으면서도 라인이나 텍스처와 같은 영상구조는 더 가깝게 표현했음을 의미한다.

제안한 구조 인식 잔차 네트워크가 기존의 잔차 네트워크와 디테일 네트워크보다 우수한 성능을 달성할 수 있었던 이유는 영상 구조에 적응적인, 즉 입력 빗줄기 영상에서 원본 에지의 형상과 평탄 영역에 대한 정보를 활용함으로써, 라플라시안 에지 맵 기반의 주의집중 네트워크를 구현했기 때문이다. 이로 인해, 평탄 영역에서 빗줄기 잔재를 말끔히 해결하고 디테일 영역에서는 에지의 선명도를 강화할 수 있는 학습 효과를 유도할 수 있었다. 그 결과로써, 정량적 평가 부분에서 더 우수한 성능을 얻게 되었고 시각적 측면에서도 화질의 선명도를 더 개선할 수 있었다.

V. 결 론

본 논문에서는 텍스처의 선명도를 강화하면서도 평탄 영역의 빗줄기 잔재 문제를 해소할 수 있는 구조 인식 잔차 네트워크를 새롭게 제안하였다. 특히 두 종류의 서브네트워크인 라플라시안 서브네트워크와 구조 인식 잔차 복원 서브네트워크를 새로

게 소개하였다. 라플라시안 서브네트워크를 통해 입력 빗줄기 영상으로부터 원본 영상의 텍스처의 모양과 평탄 영역에 대한 위치 정보를 추정하여 활용함으로써, 영상 구조에 따른 적응적인 네트워크 학습을 유도할 수 있었다. 실험 결과로 통해, 제안한 구조 인식 잔차 네트워크가 평탄 영역의 빗줄기를 제거하고 텍스처의 복원 능력도 개선할 수 있음을 확인하였다. 뿐만 아니라 최첨단 성능을 보유한 기존의 빗줄기 제거 기법보다 SSIM 및 PSNR 평가에서도 더 우수한 성능을 달성할 수 있었다.

제안한 기법은 빗줄기 영상에서 빗줄기 패턴만을 제거하는 기법만 개발한 상태이다. 하지만 폭우 환경에서는 빗줄기 패턴뿐만 아니라 헤이즈, 블러링 [19] 효과까지 동반한다. 따라서 향후 계획으로 빗줄기 영상에서 빗줄기 패턴과 헤이즈, 블러링을 동시에 제거하는 보다 고도화된 빗줄기 제거 기법을 개발하고자 한다. 더 나아가 폭우, 폭설 환경에서 눈이나 빗줄기를 동시에 제거 가능한 일반화된 기법도 개발할 계획이다.

References

- [1] Y. Choi and C.-H. Son, "Rain removal via deep convolutional neural networks considering orientation and strength of rain streaks", The Journal of Korean Institute of Information Technology, Vol. 17, No. 1, pp. 85-98, Jan. 2019.
- [2] D.-Y. Chen, C.-C. Chen, and L.-W. Kang, "Visual depth guided color image rain streaks removal

- using sparse coding", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 24, No. 8, pp. 1430-1455, Aug. 2014.
- [3] Y. Luo, Y. Xu, and H. Ji, "Removing rain from a single image via discriminative sparse coding", in Proc. IEEE International Conference on Computer Vision, Santiago, Dec. 2015, pp. 3397-3405.
- [4] C.-H. Son and X.-P. Zhang, "Rain removal via shrinkage of sparse codes and learned rain dictionary" in Proc. IEEE International Conference on Multimedia and Expo Workshop, Seattle, Jul. 2016.
- [5] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Single image rain streak decomposition using layer priors", IEEE Transactions on Image Processing, Vol. 26, No. 8, pp. 3874-3885, Aug. 2017.
- [6] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, "Automatic single-image-based rain streaks removal via image decomposition", IEEE Transactions on Image Processing, Vol. 21, No. 4, pp. 1742-1755, Apr. 2012.
- [7] J.-H. Kim, C. Lee, J.-Y. Sim, and C.-S. Kim, "Single-image deraining using an adaptive nonlocal means filter", in Proc. IEEE International Conference on Image Processing, Melbourne, VIC, Sept. 2013, pp. 914-917.
- [8] S.-C. Pei, Y.-T. Tsai, and C.-Y. Lee, "Removing rain and snow in a single image using saturation and visibility features", in Proc. IEEE International Conference on Multimedia and Expo Workshops, Chengdu, July 2014, pp. 14-18.
- [9] C.-H. Son and X.-P. Zhang, "Rain detection and removal via shrinkage-based sparse coding and learned rain dictionary", Journal of Imaging Science and Technology, Vol. 64, No. 3, pp. 30501-1-30501-17(17), May 2020.
- [10] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: residual learning of deep CNN for image denoising", IEEE Transactions on Image Processing, Vol. 26, No. 7, pp. 3142-3155, Jul. 2017.
- [11] X. Fu, J. Huang, D. Zeng, Y. Huang, and X. Ding, J. Paisley, "Removing rain from single images via a deep detail network," in Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Honolulu, USA, July 2016, pp. 1-9.
- [12] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks", in Proc. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 1125-1134, July 2017.
- [13] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks", in Proc. IEEE International Conference on Computer Vision, Venice, Italy, pp. 2223-2232, Oct. 2017.
- [14] S.-W. Jung, H.-J. Kwon, Y.-C. Kim, S.-H. Ahn, and S.-H. Lee, "Image translation method based on GAN for surveillance under dim surround", The Journal of Korean Institute of Information Technology, Vol. 18, No. 8, pp. 9-17, Aug. 2020.
- [15] M. Everingham, L. Van Gool, C. K. I Williams, J. Winn, and A. Zisserman "The PASCAL visual object classes (VOC) challenge", International Journal of Computer Vision, Vol. 88, pp. 303-338, 2010.
- [16] Y.-J. Yu, C.-H. Son, and D. H. Lee, "Apple leaf disease identification through region-of-interest-aware deep convolutional neural network", Journal of Imaging Science and Technology, Vol. 64, No. 2, pp. 20507-1-20507-10, Jan. 2020.
- [17] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer", arXiv:1612.03928v3, Feb. 2017.
- [18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", IEEE Transactions on Image Processing, Vol. 13, No. 4,

pp. 600-612, Apr. 2004.

- [19] K.-M. Lee and C.-H. Lin, "Video stabilization algorithm of shaking image using deep learning", The Journal of the Institute of Internet, Broadcasting and Communication, Vol. 19, No. 1, pp. 145-152, Feb. 2019.

저자소개

김 주 찬 (Ju-Chan Kim)



2017년 3월 ~ 현재 : 군산대학교
소프트웨어융합공학과 학사과정
관심분야 : 컴퓨터 비전, 영상처리,
딥 러닝, 프로그래밍

손 창 환 (Chang-Hwan Son)



2002년 2월 : 경북대학교
전자전기공학부(공학사)
2004년 2월 : 경북대학교
전자공학과(공학석사)
2008년 8월 : 경북대학교
전자공학과(공학박사)
2017년 4월 ~ 현재 : 군산대학교

소프트웨어융합공학과 조교수
관심분야 : 컴퓨터 비전, 영상처리, 기계학습, 딥 러닝