

안개영상의 의미론적 분할 및 안개제거를 위한 심층 멀티태스크 네트워크

송태용[†], 장현성^{**}, 하남구^{***}, 연운모^{****}, 권구용^{*****}, 손광훈^{*****}

Deep Multi-task Network for Simultaneous Hazy Image Semantic Segmentation and Dehazing

Taeyong Song[†], Hyunsung Jang^{**}, Namkoo Ha^{***}, Yoonmo Yeon^{****},
Kuyong Kwon^{*****}, Kwanghoon Sohn^{*****}

ABSTRACT

Image semantic segmentation and dehazing are key tasks in the computer vision. In recent years, researches in both tasks have achieved substantial improvements in performance with the development of Convolutional Neural Network (CNN). However, most of the previous works for semantic segmentation assume the images are captured in clear weather and show degraded performance under hazy images with low contrast and faded color. Meanwhile, dehazing aims to recover clear image given observed hazy image, which is an ill-posed problem and can be alleviated with additional information about the image. In this work, we propose a deep multi-task network for simultaneous semantic segmentation and dehazing. The proposed network takes single haze image as input and predicts dense semantic segmentation map and clear image. The visual information getting refined during the dehazing process can help the recognition task of semantic segmentation. On the other hand, semantic features obtained during the semantic segmentation process can provide cues for color priors for objects, which can help dehazing process. Experimental results demonstrate the effectiveness of the proposed multi-task approach, showing improved performance compared to the separate networks.

Key words: Semantic Segmentation, Dehazing, Deep Learning, Multi-task Learning

1. 서 론

컴퓨터 비전의 영상 인식은 자율주행 시스템 등을 포함한 다양한 응용 분야에서 핵심적인 역할을 수행

하는 연구 분야이다 [1, 26]. 최근 하드웨어 성능의 향상과 다량의 데이터 구축에 따른 컨벌루션 신경망 (Convolutional Neural Network, CNN)의 발달과 함께 영상 분류(Classification) [2], 탐지(Detection) [3,

※ Corresponding Author : Kwanghoon Sohn, Address: (120-749) Digital Image Media Laboratory, C129, The 3rd Engineering Building, Yonsei University 50 Yonsei-ro, Seodaemun-Gu, Seoul 120-749, Korea, TEL : +82-2-2123-2879, FAX : +82-2-2123-7712, E-mail : khsohn@yonsei.ac.kr

Receipt date : Aug. 12, 2019, Revision date : Sep. 3, 2019
Approval date : Sep. 5, 2019

[†] School of Electrical Electronic Engineering, Yonsei University, E-mail: sty37@yonsei.ac.kr,

^{**} EO/IR R&D Lab., LIG Nex1 Co., Ltd., 16911, Yongin-city, South Korea, E-mail: hyunsung.jang@gmail.com

^{***} EO/IR R&D Lab., LIG Nex1 Co., Ltd., 16911, Yongin-city, South Korea

E-mail: hanamkoo369a@lignex1.com

^{****} EO/IR R&D Lab., LIG Nex1 Co., Ltd., 16911, Yongin-city, South Korea

E-mail: yoonmo.yeon@lignex1.com

^{*****} EO/IR R&D Lab., LIG Nex1 Co., Ltd., 16911, Yongin-city, South Korea

E-mail: kuyong.kwon@lignex1.com

^{*****} School of Electrical Electronic Engineering, Yonsei University

27] 등 영상 인식의 다양한 세부 분야는 큰 성능 향상을 이루어 냈다. 이 중, 의미론적 분할(Semantic segmentation)은 영상의 각 픽셀이 어떠한 물체에 속하는지 예측 하는 기술이다. 다른 기술들과 마찬가지로, 의미론적 분할 기술 또한 CNN의 발전과 더불어 큰 성능 향상을 이루어 냈다 [9, 10, 11, 12, 18, 28]. 그러나 이러한 기존 기술은 입력 영상이 열화 없이 촬영된 (Fig. 1(a)) 것을 가정하고 안개 등 날씨의 영향으로 인한 손상이 있는 경우(Fig. 1(d))를 별도로 가정하지 않으며, 따라서 영상의 왜곡 또는 손상이 있을 경우 큰 성능 저하를 보인다 (Fig. 1(e)).

안개는 공기 중의 수증기, 먼지 등이 빛을 산란시켜 발생하는 자연 현상으로, 일반적으로 식 (1)과 같이 모델링 된다.

$$I(x) = J(x)t(x) + A(1-t(x)) \quad (1)$$

여기서 $I(x)$, $J(x)$, $t(x)$ 는 각각 픽셀 x 에서의 안개 영상, 안개 없는(Clean) 영상, 그리고 투과맵(Transmission Map)을 뜻하며, A 는 산란된 빛인 Airlight를 뜻한다. 투과맵은 영상의 깊이 정보 d 와 빛의 산란 계수인 β 를 통해 $t(x) = e^{-\beta d(x)}$ 와 같이 정의된다. 안개가 있는 환경에서 촬영된 영상은 대비(Contrast)가 저하되고 색상이 희미해지는 등 영상의 세부 정보가 부족해, 시각적으로 좋지 않을 뿐만 아니라 해당

영상을 입력으로 사용해 이루어지는 영상 인식 기술의 성능을 저하시킨다. 이와 같은 이유로 안개제거 기술 또한 컴퓨터 비전에서 활발히 연구되어 온 기술이다. 단일 영상 안개제거 기술은 하나의 입력 영상으로부터 Clean 영상을 얻는 것을 목표로 한다. 이는 단일 영상에 대해 다수의 Clean 영상의 추정이 가능한 ill-posed 문제이다. 단일 영상의 안개제거를 위한 기존의 연구들은 이러한 문제를 해결하기 위해 여러 가정을 사용하거나 [4, 5] 인공적으로 생성된 안개 영상을 이용해 심층 네트워크를 학습하는 방식으로 [6, 7, 8] 안개제거를 수행한다. 이러한 방식들은 사물의 의미론적 정보와는 상관없이 지역적인 픽셀(Pixel) 또는 작은 패치(Patch)의 색상에만 의존해 안개제거를 수행하여 물체 색상을 잘못 복원하는 등의 문제점이 있다.

본 논문에서는 의미론적 분할과 안개제거가 함께 학습 되었을 때 서로 도움을 줄 수 있다는 가정에 착안하여, 하나의 단일 네트워크로 두 가지를 동시에 수행하는 멀티태스크(Multi-task, 다중 작업) 심층 네트워크를 제안한다. 문헌 조사에 따르면 본 연구는 의미론적 분할과 안개제거를 동시에 수행한 최초의 연구이다. 제안한 네트워크를 학습하기 위해 인공적으로 합성된(Synthetic) 다량의 안개 영상을 이용하였으며, 결과적으로 네트워크는 다양한 상황의 안개

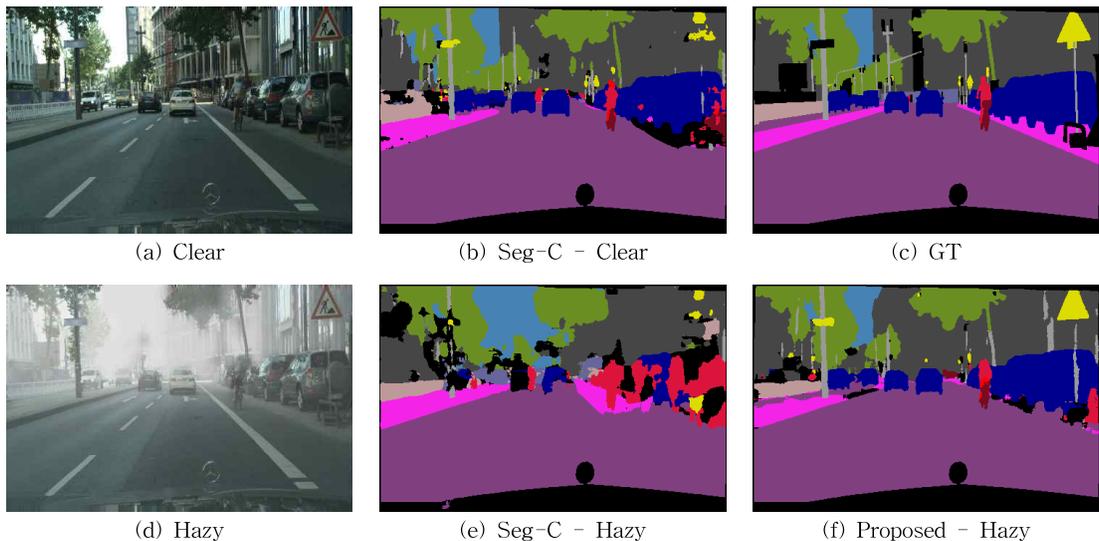


Fig. 1. Challenge of semantic segmentation with hazy images. (a) Clean image, (b) Segmentation result of clean image, (c) Ground truth segmentation image, (d) Synthesized hazy image, (e) Segmentation result of hazy image, (f) Segmentation result of hazy image with proposed method.

영상에 효율적으로 적용될 수 있다. 제안한 네트워크는 상호 보완적인 두 작업을 동시에 학습함으로써 가중치 공유(Weight sharing)를 통해 개선된 성능을 보인다. 또한, 각각의 작업을 위해 별도의 네트워크를 사용하는 것, 혹은 두 작업을 순서대로 처리하는 것보다 메모리와 계산량의 측면에서 효율적이다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문과 관련된 영상의 의미론적 분할, 안개제거, 그리고 멀티태스크 학습의 연구에 관해 설명한다. 3장에서는 제안하는 네트워크 구조와 학습 방법, 그리고 목적 함수에 대해 설명하며, 4장에서는 네트워크를 학습하기 위한 구성, 데이터셋과 실험 결과를 제시한다. 마지막으로 5장에서는 본 논문에서 제안한 방법의 결론을 맺는다.

2. 관련 연구

2.1 의미론적 분할(Semantic segmentation)

영상의 의미론적 분할은 영상에서 같은 클래스(Class)의 물체에 해당하는 픽셀끼리 나누는 작업으로, 자율주행 등 산업의 각종 분야의 자동화를 위해 필수적인 연구인 영상인식의 한 분야이다. 최근 GPU 등 하드웨어의 발달과 다양한 데이터 셋에 기반을 둔 딥 러닝 방식들이 FCN [9] 을 시작으로 개발되어 왔다. 이후, SegNet [10], UNet [11] 등 인코더-디코더 기반의 구조를 사용해 low-level과 high-level Feature를 모두 추출하여 융합할 수 있는 연구들이 제안되었다. 또한 PSP-Net [12]은 PPM (Pyramid Pooling Module)을 이용하여 다양한 스케일에서의 Feature를 추출하였다. 기존의 이와 같은 연구들은 일반적으로 준수한 성능을 보이지만, 안개 등에 의한 왜곡을 별도로 고려하지 않아 안개에 의해 대비가 저하되고 색상정보가 손실된 영상에 대해서는 낮은 성능을 보인다.

이러한 문제점을 완화하여 안개 상황 속에서도 영상의 의미론적 분할에서 좋은 성능을 얻기 위한 연구가 최근 제안되었다 [13, 14]. 이들 연구는 의미론적 분할에 대한 Ground-truth가 있는 데이터셋의 영상에 안개를 합성하고, 분할을 위한 네트워크를 합성된 영상으로 학습하는 방식으로 안개 영상에서 분할 네트워크가 좋은 성능을 보일 수 있도록 한다. 그러나 위 연구들에서는 안개제거의 과정에서 얻어질 수 있

는 Feature 수준에서의 영상의 세부(Detail) 정보를 이용하지 않고, 안개 영상이라는 새 도메인(Domain)에서의 개선된 분할 결과를 얻는 것에만 집중하였다.

2.2 안개제거(Haze removal)

영상에서 안개를 제거하기 위해 과거 Dark Channel Prior [4]나 Color Attenuation Prior [5] 등 다양한 가정들이 사용되었다. 이러한 방법들은 식 (1)을 기반으로 투과맵 $t(x)$ 와 Airlight A 를 추정하여 식 (1)의 역계산으로 Clean 영상을 추정한다. 이러한 방식은 사용한 가정이 맞지 않는 경우 큰 성능 저하를 보인다. 이후 개발된 딥 러닝 기반 연구들도 식 (1)을 기반으로 인공적으로 생성된 안개 영상을 통해 투과맵과 Airlight를 추정하는 방식이 추가 되어왔다. Dehaze Net [6]은 BReLU 활성화함수와 Maxout [15]을 사용하여 투과맵을 추정하였으며, MSCNN [7]은 Coarse 스케일과 Fine 스케일 네트워크를 이용하여 투과맵을 추정하였다. 위 방식들의 네트워크는 물체의 구조적 정보 또는 의미론적 정보를 고려하지 않고 픽셀의 색상 등 지역적인 특징만으로 투과맵을 추정해 물체의 색상 등에 따라 실제 투과맵과 다른 값을 추정하는 등의 문제가 있으며, Airlight는 경험적 가정 [3]에 의해 추정한다는 한계점이 있다. 최근 DCPDN [8]에서는 투과맵과 더불어 Airlight 또한 다량의 데이터를 이용한 학습을 통해 추정하는 방식을 제안하였다. 이는 경험적 가정을 통해 Airlight를 추정하는 기존 방식의 단점을 완화하였으나, 안개 제거 과정에 영상의 의미론적 정보는 고려하지 않았다.

안개제거에 의미론적 정보의 사용을 고려한 기존 연구로는 [16]의 방식이 있다. 이는 동영상의 안개제거를 목표로 하며, 사전에 학습된(Pre-trained) 분할 네트워크를 [17] 사용하여 안개 동영상의 각 프레임(Frame)으로부터 의미론적 분할 영상을 생성한다. 이후 분할 결과를 안개제거 네트워크의 중간 Feature와 Concatenate 시킨 뒤 최종적으로 투과맵을 추정한다. 해당 연구는 의미론적 정보가 투과맵 추정에 도움이 될 수 있음을 보였지만, 안개에 의해 저하되는 의미론적 분할 결과의 정확도는 고려하지 않았다. 또한, 해당 방식은 영상에서의 의미론적 정보를 투과맵을 추정하기 위한 보조 역할로만 수행하여 두 작업을 동시에 처리함으로써 얻을 수 있는 상호 보완성은 고려하지 못했다.

2.3 멀티태스크 학습(Multi-task learning)

멀티태스크 학습은 여러 개의 다양한 작업을 함께 학습함으로써 단일 학습 대비 두 작업 모두에서 향상된 결과를 얻는 것을 목표로 한다. 멀티태스크 학습에는 두 작업이 물리적, 수학적으로 관련이 있는 경우도 있지만 [18, 19], 비교적 약한 연관성을 지닌 경우도 있다 [20, 21]. Song의 논문 [18]에서는 스테레오 매칭과 안개의 밀도가 $t(x) = e^{-\beta d(x)}$ 라는 관계 통해 연결되고 거리에 대한 서로 다른 단서를 갖고 있으며, 단서가 상호보완적이라는 점에 착안하여 두 작업을 함께 학습하였다. Qi의 논문 [19]은 장면의 깊이와 연직성분(Normal)이 상호 도출될 수 있다는 점에 착안하였다. Li의 연구 [20]는 블러제거(Deblur)와 초해상도(Super-resolution)을 블러의 측면에서 공통성이 있다고 보아 함께 학습하는 방안을 제안하였다. Pan의 연구 [21]에서는 블러제거와 Scene flow를 동시에 추정하며, 이는 움직임이라는 요소와 공통적으로 연관이 있음에 기반을 둔 연구이다. 이 두 논문 [20, 21]은 두 결과물이 일련의 수식으로 도출되지는 않지만, 서로 관련이 있다는 점에서 멀티태스크 학습을 제안하였다. 위 논문들은 공통적으로 서로 다른 두 개 이상의 작업이 서로 연관성이 있으며, 따라서 네트워크가 두 작업을 수행하기 위해 학습되는 과정에서 파라미터 공유(Parameter Sharing)을 통해 향상된 성능을 낼 수 있음을 제시하였다. 본 연구에서는 비교적 서로 다른 두 작업을 동시에 수행하는 네트워크에서의 파라미터 공유를 통한 성능 향상을 제시하였다.

3. 제안한 방법

본 논문에서는 단일 안개 영상(I)을 네트워크의 입력으로 하여 영상의 의미론적 분할 결과(\hat{M})와 안개제거 영상(\hat{J})을 출력으로써 얻는 것을 목표로 한다. 이를 위해 두 가지 작업을 동시에 수행하는 멀티태스크 네트워크를 설계하였다. 이는 안개 과정에서 점차 개선되는 영상의 특징들(Feature)이 영상의 의미론적 분할에 도움을 줄 수 있으며, 반대로 영상 내 물체들의 의미론적 정보가 안개제거의 수행에 있어서 각 물체의 본래 색상을 복원하는 과정에 단서를 주어 상호보완성을 지닐 수 있음에 착안하였다.

3.1 네트워크 구조

본 논문에서 제안한 네트워크에 대한 구조는 Fig. 2에 제시되었다. 제안한 네트워크는 기존의 멀티태스크 논문들 [19, 21]에서 일반적으로 사용되는 인코더-디코더(Encoder-Decoder) 구조를 사용하였다. 안개 영상을 입력으로 받는 하나의 인코더 네트워크와 해당 인코더 feature를 입력으로 하는 두 개의 디코더 네트워크로 구성된다. 두 개의 디코더는 분할 영상과 안개제거 영상을 각각 출력한다. 안개제거의 경우, 투과맵과 Airlight를 추정 한 뒤 식 (1)의 역계산을 수행하는 기존 방법들 [4, 5, 6, 7]과는 달리 네트워크의 출력으로써 바로 안개제거 영상을 얻는다. 인코더는 VGG-19 [22] 구조를 기반으로 하여 3×3 컨벌루션(Convolution), 배치 정규화(Batch Normalization), Leaky ReLU의 기본 단위와 최대값 풀링(Max Pooling)을 갖는 블록들로 이루어져 있다. 각 블록의 끝에서 이루어지는 최대값 풀링을 통해

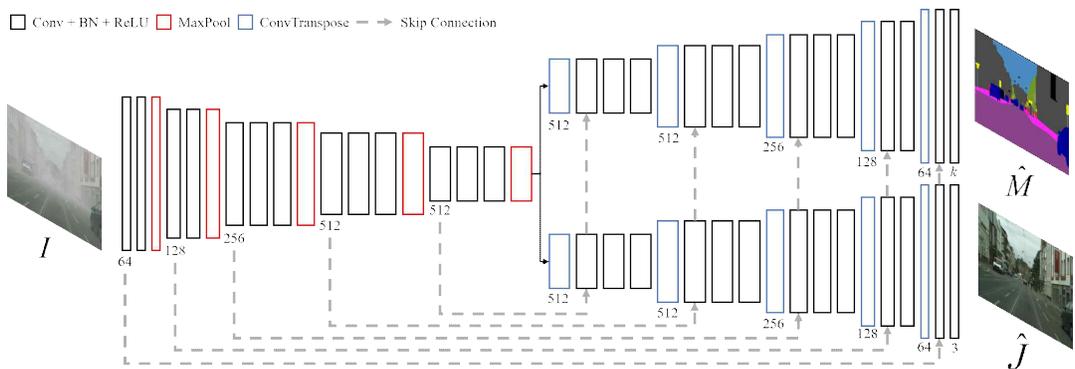


Fig. 2. Overall architecture of the proposed multi-task network for semantic segmentation and dehazing.

Feature의 사이즈는 절반으로 줄어 들고, 다음 블록의 첫 컨볼루션에서 Feature의 채널(Channel)의 수는 두 배가 된다. 인코더는 총 5개의 블록을 포함하며, 인코더의 마지막 Feature는 입력 영상 크기의 $1/32$ 크기를 갖는다.

네트워크 후반부의 디코더를 구성하는 각 블록은 이전 Feature를 4×4 컨볼루션 전치(Convolution Transpose)를 이용해 해상도를 2배로 늘린 뒤 그에 해당하는 해상도를 갖는 인코더 Feature를 Skip connection을 통해 합친다. 이후 3×3 컨볼루션, 배치 정규화, Leaky ReLU를 반복해 Feature를 처리한다. 두 개의 디코더는 다섯 개의 블록을 포함해 기본적으로 같은 구조를 갖으며, 입력 영상과 같은 크기의 출력을 각각 생성하지만 다음과 같은 차이가 있다. 분할을 위한 디코더는 물체의 클래스(Class)의 개수인 k 개의 채널을 갖는 Feature에 각 픽셀에서 채널 축으로 Softmax 함수를 적용한 것을 출력으로 생성한다. 반면, 안개제거를 위한 디코더는 출력이 안개를 제거한 RGB 영상이므로 3개의 채널을 갖고, 영상의 Dynamic range를 제한하기 위해 마지막에 Tanh 함수를 적용한다.

3.2 네트워크 학습 데이터 및 비용함수

제안한 네트워크는 의미론적 분할과 안개제거 모두 다량의 데이터를 사용한 지도학습(Supervised Learning)을 통해 학습된다. 따라서 학습을 위해 입력 안개 영상과 그에 해당하는 Clean 영상, 영상 분할에 대한 Ground-truth 쌍의 학습 데이터 구성이 필요하다. 이를 위해 Ground-truth 분할 영상이 제공되는 Clean 영상에 인공적인 안개를 합성한다. 안개 영상 합성을 위해서 일반적으로 사용되는 안개 영상의 수학적 모델(식 (1))을 사용하였다.

본 논문에서는 Cityscapes 데이터셋 [23]에서 파생된 Foggy Cityscapes-DBF 데이터셋 [14]을 사용하여 학습 데이터를 구성하였다. 해당 데이터셋은 Cityscapes에서 제공되는 잡음이 있는(Noisy) 깊이 영상을 RGB 영상과 의미론적 분할에 대한 Ground-truth인 M_{gt} 를 사용한 Dual-reference Cross-Bilateral Filtering (DBF)을 사용하여 개선 한 뒤 여러 값의 산란계수 β 를 사용해 투과맵을 생성한다. DBF 처리를 통해 테두리가 물체와 잘 정렬된 투과맵을 계산하여 자연스러운 안개를 합성할 수 있다. 본 논

문에서는 Foggy Cityscapes-DBF에서 제공되는 투과맵과 (0.7,1.0) 범위의 Airlight A 를 사용해 식 (1)을 통해 안개 영상을 생성하여 결과적으로 학습 데이터 $\{I, J_{gt}, M_{gt}\}$ 를 구성한다.

이후 네트워크 학습을 위한 비용함수를 정의한다. 본 연구에서 정의하는 비용함수는 의미론적 분할과 안개 제거를 위한 두 비용함수의 선형 조합이다(식 (2)).

$$L_{total} = L_{seg} + \lambda L_{dehaze} \quad (2)$$

여기서 λ 는 두 목적함수간의 균형을 맞추기 위한 파라미터이다. 의미론적 분할에 대한 학습을 위해 기존 의미론적 분할 연구 [9, 10, 11, 12]에서 일반적으로 사용되는 픽셀 별 Cross-entropy Loss를 추정값 \hat{M} 과 Ground-truth M_{gt} 사이에 적용하였다(식 (3)).

$$L_{seg} = - \sum M_{gt} \log(\hat{M}) + (1 - M_{gt}) \log(1 - \hat{M}) \quad (3)$$

안개제거를 위해서는 L_1 비용함수를 식 (4)와 같이 정의하여 안개제거 영상과 Ground-truth Clean 영상 사이에 적용하였다.

$$L_{dehaze} = \| \hat{J} - J_{gt} \|_1 \quad (4)$$

L_1 비용함수는 L_2 비용함수에 비해 결과 영상의 Blurring이 덜하고 세부정보가 보존되는 방향으로 학습되도록 한다.

4. 실험 결과

본 장에서는 제안한 방법의 효율성을 보이기 위한 실험 결과를 제시한다. 먼저 실험 환경에 대해 설명하고, 각 작업을 별개로, 혹은 순차적으로 처리하는 것보다 제안한 멀티태스크 방식을 통해 향상된 결과를 얻을 수 있음을 보인다. 마지막으로, 안개제거와 영상의 의미론적 분할을 다루는 기존 방법들과의 성능을 비교한다.

4.1 실험 환경 및 구현 세부사항

본 연구에서는 학습을 위해 3.8GHz CPU와 24GB RAM, NVIDIA Titan X GPU가 장착된 PC를 이용하였으며, 딥 러닝 라이브러리는 Tensorflow [24]를 사용하였다. 학습과 평가 시에 영상은 320×480 크기로 조정하였다. 학습 시 배치(Batch)의 크기는 2로

Table 1. Performance of semantic segmentation with different configurations

Network	Seg-C		Seg-H	Proposed
Test Image	Clean	Hazy	Hazy	Hazy
mIoU	0.4730	0.2986	0.3419	0.3842

설정하였으며, Adam optimizer [25]를 사용하였다. 학습률(Learning Rate)은 초기 10^{-4} 로 설정하여 매 10 epoch마다 0.1을 곱하여 총 30 Epoch동안 학습하였으며, 식 (2)의 λ 는 1로 설정하였다. 네트워크 학습에는 Cityscapes 데이터셋 [23]의 학습 세트인 2,993장의 Clean 영상과 Mask를 GT로 하는 안개영상을 Foggy Cityscapes-DBF [14]의 방식으로 생성하여 네트워크의 입력으로 사용하였다. 또한 물체의 클래스(Class)수인 k 는 Cityscapes 데이터셋에 맞도록 20으로 설정하였다. Cityscapes 데이터셋은 Test Set에 대한 Ground-truth 분할 영상이 제공되지 않아 입력 안개 영상을 생성하거나 성능 평가를 할 수 없는 문제점이 있어, 성능평가는 500장의 Clean 영상으로 이루어진 Validation Set으로부터 생성된 1,500장의 안개 영상으로 진행하였다. 영상의 의미론적 분할을 위한 측정 방식으로는 mIoU (mean Intersection over Union)을 사용하였고, 안개제거의 성능 측정은 PSNR (Peak Signal to Noise Ratio)과 SSIM (Structure Similarity)을 사용하였다.

4.2 멀티태스크 네트워크의 효율성

제안한 멀티태스크 네트워크의 효율성을 보이기 위하여 안개 영상을 입력으로 하는 의미론적 분할 네트워크, 혹은 안개제거 네트워크를 단독으로 학습

한 경우와 결과를 비교하였다. 분할 네트워크를 Clean 영상과 안개영상으로 각각 학습하였으며(Seg-C, Seg-H), 안개제거 네트워크(Dehazing)와 멀티태스크 네트워크(Proposed)는 안개영상만으로 학습하였다.

영상의 의미론적 분할에 대한 결과는 Table 1과 Fig. 3에 주어졌다. Seg-C 네트워크의 Clean 영상에 대한 성능을 기준으로 잡아 다른 결과들과 비교하였다. Clean 영상에 학습한 네트워크(Seg-C)로 안개영상에 대한 분할을 수행 할 경우 mIoU가 0.4730에서 0.2986으로 현저히 저하되는 것을 확인할 수 있다. 이러한 성능 저하는 네트워크를 안개 영상으로 학습시키는 것(Seg-H)으로 일부 개선시킬 수 있다 (mIoU = 0.3419). 제안한 멀티태스크 네트워크는 안개제거 과정에서 얻어지는 영상의 세부 정보를 이용하여 더욱 정확한 의미론적 분할을 수행하였다 (mIoU = 0.3842).

안개제거에 대한 성능 비교는 Fig. 4에 주어졌다. 단일 네트워크만 학습한 경우 PSNR = 27.66, SSIM = 0.9363의 성능을 보였으며, 제안한 멀티태스크 네트워크는 분할 과정에서 얻어지는 물체의 의미론적 정보를 추가적으로 사용하여 PSNR = 28.31, SSIM = 0.9483로 향상된 성능을 보인다. 안개제거 단일 네트워크 또한 전반적으로 우수한 성능을 보이지만 식물의 색상을 열게 만들거나 (Saturated) 건물의 색상을 잘 복원하지 못하는 경우가 관찰되었다. 그러나 의미론적 분할과 함께 학습된 멀티태스크 네트워크는 해당 물체들의 색상을 더욱 정확히 복원하였다.

4.3 기존 방법과의 비교

본 절에서는 영상의 의미론적 분할과 안개제거에

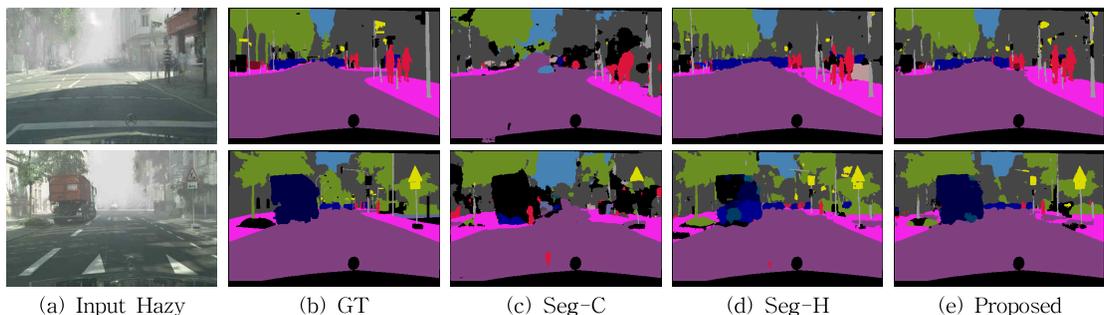


Fig. 3. Semantic segmentation on hazy images with different networks, (a) Input hazy image, (b) Ground-truth segmentation map (c)~(e) Semantic segmentation results with (c) Seg-C, (d) Seg-H and (e) Proposed network.

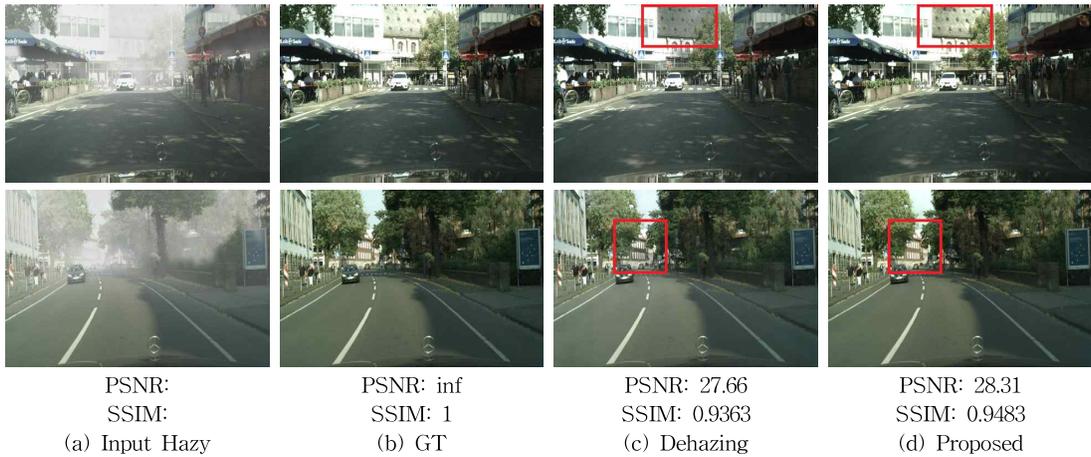


Fig. 4. Dehazing results of dehazing network and proposed multi-task network, Objective metrics are averaged over all images, (a) Hazy image, (b) Ground-truth clean image, (c) Dehazing result of single network and (d) Dehazing result of proposed multi-task network,

관한 기존 연구들의 결과와 제안한 멀티태스크 네트워크의 결과를 비교한다. 의미론적 분할과 안개제거를 동시에 수행한 기존 연구가 부재하므로 각각을 수행한 연구의 결과들과 비교를 진행하였다. 의미론적 분할에 대해서는, 안개를 고려하지 않은 의미론적 분할 네트워크가 안개 상황에서 성능이 크게 저하되는 것을 4.2절에서 보였으므로, 안개를 고려하여 분할 네트워크를 학습한 기존의 연구 SFSU [13]와 비교한다. 안개제거는 CNN 기반 단일 영상을 안개 제거 방식들인 DehazeNet [6], MSCNN [7], DCPDN [8]과 성능을 비교한다. 비교한 방법들의 소스코드와 네트워크 파라미터는 모두 저자가 공개한 것을 사용

하였다.

의미론적 분할에 대한 비교 결과는 Fig. 5에 제시되었다. 기존 연구인 SFSU [13]는 mIoU = 0.3717의 성능을 보였으며, 제안한 멀티태스크 네트워크는 mIoU = 0.3842의 성능으로 기존 연구에 비해 향상된 성능을 보였다. SFSU는 전반적으로 부드러운 결과를 내지만 반대로 작은 구조를 잘 포착하지 못하며, 특히 먼 거리의 길은 안개가 존재하는 영역에 대해 그러한 현상이 도드라진다. 반면, 제안한 방식은 먼 거리의 물체들에 대해서도 세밀하고 정확한 분할 결과를 제공한다.

안개제거에 대한 비교 결과는 Fig. 6에 제시되었

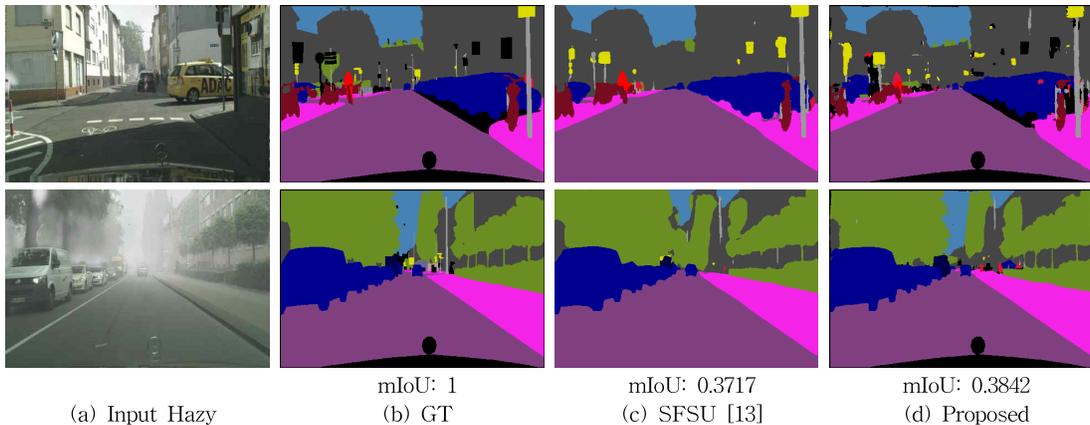


Fig. 5. Semantic segmentation results on hazy image. (a) Input hazy image, (b) Ground-truth, (c) Result of SFSU [13], and (d) Result of proposed method,

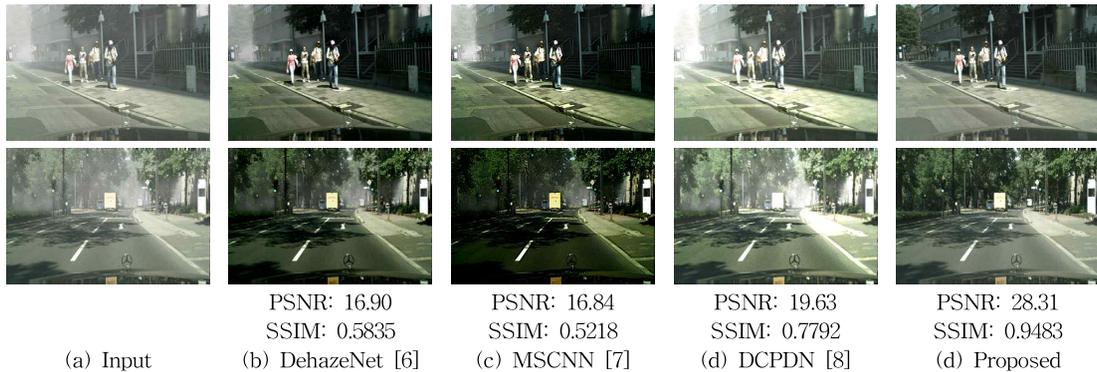


Fig. 6. Dehazing results of previous method and the proposed method. Objective metrics are averaged over all test images. (a) Input hazy image, (b)~(d) Dehazing results of (b) DehazeNet, (c) MSCNN, (d) DCPDN, and (e) Proposed network.

다. 기존의 방법인 DehazeNet [6]과 MSCNN [7] 모두 전반적으로 안개제거 영상을 과하게 어둡게 만들며, 영상의 상당 부분에서 안개를 성공적으로 제거하지 못한다. 반면, DCPDN [8] 영상을 전반적으로 과하게 밝게 만들며, 안개 또한 성공적으로 제거하지 못하는 모습을 보였다. 정량적 평가로는 DehazeNet은 PSNR = 16.90, SSIM = 0.5835, MSCNN은 PSNR = 16.84, SSIM = 0.5218, 그리고 DCPDN은 PSNR = 22.63, SSIM = 0.7792로 낮은 성능을 보인다. 반면, 제안한 멀티태스크 네트워크는 과한 안개로 영상의 정보가 완전히 손실된 부분을 제외하고는 정답 Clean 영상과 가까운 안개제거 영상을 정확하게 복원하며, 정량적 측면에서도 PSNR = 28.31, SSIM = 0.9483으로 월등히 우수한 결과를 보인다.

5. 결 론

본 논문에서는 컴퓨터 비전의 주요 분야인 영상의 의미론적 분할 기술이 안개가 낀 환경에서 획득된 영상에서 성능이 저하되는 문제점과, 장면의 의미론적 정보 없이 이루어지는 기존 안개제거 연구의 한계점에 대해 시사하였다. 이러한 문제점을 해결하기 위해 본 논문에서는 영상의 의미론적 분할과 안개제거를 동시에 수행하는 딥 러닝 기반 멀티태스크 학습 프레임워크를 제안하였다. 이는 안개제거 과정에서의 영상 세부정보가 의미론적 분할에 도움을 주고, 의미론적 분할의 물체에 대한 정보가 안개제거의 색상복원에 도움을 줄 수 있다는 것을 가정한다. 본 연구의 결과로 각 작업을 따로 학습했을 때보다 동시에

학습했을 때 성능이 향상되는 것을 확인하였으며, 기존의 각 분야 연구 결과보다 향상된 성능을 내는 것을 확인하였다. 이와 같은 프레임워크는 향후 다른 영상인식과 영상개선 분야에 접목하여 확장이 가능하다.

REFERENCE

- [1] C. Oh, B. Kim, and K. Sohn, "Automatic Illumination Invariant Road Detection with Stereo Vision," *Proceedings of the IEEE Conference on Industrial Electronics and Applications*, pp. 889-893, 2012.
- [2] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.
- [3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks," *Advances in Neural Information Processing Systems*, pp. 91-99, 2015.
- [4] K. He, J. Sun, and X. Tang, "Single Image Haze Removal Using Dark Channel Prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 12, pp. 2341-2353, 2011.
- [5] Q. Zhu, J. Mai, and L. Shao, "A Fast Single

- Image Haze Removal Algorithm Using Color Attenuation Prior," *IEEE Transactions on Image Processing*, Vol. 24, No. 11, pp. 3522–3533, 2015.
- [6] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An End-to-end System for Single Image Haze Removal," *IEEE Transactions on Image Processing*, Vol. 25, No. 11, pp. 5187–5198, 2016.
- [7] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.H. Yang, "Single Image Dehazing via Multi-scale Convolutional Neural Networks," *Proceeding of European Conference on Computer Vision*, pp. 156–169, 2016.
- [8] H. Zhang and V.M. Patel, "Densely Connected Pyramid Dehazing Network," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3194–3203, 2018.
- [9] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.
- [10] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A Deep Convolutional Encoder-decoder Architecture for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 12, pp. 2481–2495, 2017.
- [11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional Networks for Biomedical Image Segmentation," *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 234–241, 2015.
- [12] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6230–6239, 2016.
- [13] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic Foggy Scene Understanding with Synthetic Data," *International Journal of Computer Vision*, 2018, Vol. 126, No. 9, pp. 973–992.
- [14] D. Dai, C. Sakaridis, S. Hecker, and L. Van Gool, "Curriculum Model Adaptation with Synthetic and Real Data for Semantic Foggy Scene Understanding," *International Journal of Computer Vision*, pp. 1–23, 2019.
- [15] I. Goodfellow, D. Warde-Farley, M. Mirza, A. C. Courville, and Y. Bengio, "Maxout Networks," *Proceedings of the 30th International Conference on Machine Learning*, pp. 1319–1327, 2013.
- [16] W. Ren, J. Zhang, X. Xu, L. Ma, X. Cao, G. Meng, et al., "Deep Video Dehazing with Semantic Segmentation," *IEEE Transactions on Image Processing*, Vol. 28, No. 4, pp. 1895–1908, 2018.
- [17] G. Lin, A. Milan, C. Shen, and I. Reid, "Refine-net: Multipath Refinement Networks with Identity Mappings for High-resolution Semantic Segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1925–1934, 2017.
- [18] T. Song, Y. Kim, C. Oh, and K. Sohn, "Deep Network for Simultaneous Stereo Matching and Dehazing," *British Machine Vision Conference*, 2018.
- [19] X. Qi, R. Liao, Z. Liu, R. Urtasun, and J. Jia, "Geonet: Geometric Neural Network for Joint Depth and Surface Normal Estimation," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 283–291, 2018.
- [20] X. Zhang, H. Dong, Z. Hu, W. Lai, F. Wang, and M. Yang, "Gated Fusion Network for Joint Image Deblurring and Super-resolution," *British Machine Vision Conference*, 2018.
- [21] L. Pan, Y. Dai, M. Liu, and F. Porikli, "Simultaneous Stereo Video Deblurring and Scene Flow Estimation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4382–4391, 2017.
- [22] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-scale

- Image Recognition," *Proceeding of International Conference on Learning Representations*, 2015.
- [23] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, et al., "The Cityscapes Dataset for Semantic Urban Scene Understanding," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3213-3223, 2016.
- [24] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean et al., "Tensorflow: A System for Large-scale Machine Learning," *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pp. 265-283, 2016.
- [25] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *Proceeding of International Conference on Learning Representations*, 2015.
- [26] B. Lee and S. Kwon, "Automatic Classification System of Tablets with Various Colors and Shapes," *Journal of Korea Multimedia Society*, Vol. 21, No. 6, pp. 659-666, 2018.
- [27] Y. Jeong, I. Ansari, J. Shim, and J. Lee, "A Car Plate Area Detection System Using Deep Convolution Neural Network," *Journal of Korea Multimedia Society*, Vol. 20, No. 8, pp. 1166-1174, 2017.
- [28] H. Lee, J. Sa, H. Shin, Y. Chung, D. Park, and H. Kim, "Separation of Occluding Pigs using Deep Learning-based Image Processing Techniques," *Journal of Korea Multimedia Society*, Vol. 22, No. 2, pp 136-145, 2019.



송 태 옹

2017년 연세대학교 전기전자공학과 학사
2017년~현재 연세대학교 전기전자공학과 석/박 통합과정
관심분야: 영상처리, 영상개선, 탐지/추적



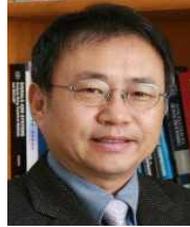
권 구 옹

2014년 한국 항공대학교 학사
2014년~현재 LIG넥스원, 선임연구원
관심분야: 영상처리, 임베디드 시스템



장 현 성

2007년 부산대학교 컴퓨터공학과 학사
2009년 부산대학교 컴퓨터공학과 석사
2011년~현재 LIG넥스원, 선임연구원



손 광 훈

1983년 연세대학교 전자공학과 학사
1985년 미국 미네소타대학교 전자공학과 석사
1991년 미국 노스캐롤라이나 주립대학교 전기·컴퓨터공학과 박사

관심분야 : 딥러닝, 실시간 시스템, 영상처리

1988년~1992년 North Carolina A&T State University, Image Proc. & Computer Vision Lab., Research Associate

1992년~1993년 한국전자통신연구소 위성통신기술연구단, 선임연구원

1994년 Georgetown University MRI Center, Post Doctoral Fellow

1995년~현재 연세대학교 전기전자공학부 교수

2002년~2003년 Nanyang Technological University, School of Computer Engineering, Visiting Professor

2009년~2010년 국방과학연구소 전문위원

관심분야: 3차원 영상 신호 처리, 컴퓨터 비전



하 남 구

2008년 경북대학교 컴퓨터공학과 박사
2009년~현재 LIG넥스원, 수석연구원

관심분야: 영상처리, 딥러닝, 감시정찰 시스템



연 윤 모

2006년 경북대학교 전자전기컴퓨터공학부 학사
2009년 경북대학교 전자전기컴퓨터공학부 석사
2009년~현재 LIG넥스원, 선임연구원

관심분야 : 영상처리, 임베디드 시스템, 실시간 시스템