석사학위논문

도메인 간 특징 분리 및 합성 기반의 내재적 생성 데이터 증강 기술을 활용한 Rainy 이미지 분할 향상

2024년

한 성 대 학 교 대 학 원

A I 응 용 학 과

A I 응 용 학 전 공

용 윤 정

석 사 학 위 논 문
지도교수 오희석

도메인 간 특징 분리 및 합성 기반의 내재적 생성 데이터 증강 기술을 활용한 Rainy 이미지 분할 향상

Built-in Generative Data Augmentation via Cross-Domain Feature Disentanglement and Synthesis for Enhancing Rainy Image Segmentation

2024년 6월 일

한 성 대 학 교 대 학 원

A I 응 용 학 과

A I 응 용 학 전 공

용 윤 정

석 사 학 위 논 문
지도교수 오희석

도메인 간 특징 분리 및 합성 기반의 내재적 생성 데이터 증강 기술을 활용한 Rainy 이미지 분할 향상

Built-in Generative Data Augmentation via Cross-Domain Feature Disentanglement and Synthesis for Enhancing Rainy Image Segmentation

위 논문을 공학 석사학위 논문으로 제출함

2024년 6월 일

한 성 대 학 교 대 학 원

A I 응 용 학 과

A I 응 용 학 전 공

용 윤 정

용윤정의 공학 석사학위 논문을 인준함

2024년 6월 일

심사위원장 김 명 선 (인)

심 사 위 원 오 희 석 (인)

심 사 위 원 이 웅 희 (인)

국 문 초 록

도메인 간 특징 분리 및 합성 기반의 내재적 생성 데이터 증강 기술을 활용한 Rainy 이미지 분할 향상

한 성 대 학 교 대 학 원
A I 응 용 학 과
A I 응 용 학 전 공
용 윤 정

비는 일반적인 기상 조건 중 하나로, 물체 감지 및 분할과 같은 자율 주행 시스템에서 화질 저하의 주요 원인이 되어 악천후 조건에서의 자율 주행을 어렵게 만든다. 현재 맑은 날씨 조건에서의 시맨틱 세그멘테이션 모델은 큰 발전을 이루었으나, 이미지 품질이 저하되는 악천후 상황에서는 성능이 급격히 떨어진다. 비 오는 상황에서도 세그멘테이션 성능을 유지하기 위해, 비 오는 영상과 깨끗한 영상이 쌍으로 존재하는 데이터셋을 사용하여 세그멘테이션 모델을 파인튜닝하려 했으나, 세그멘테이션 GT가 함께 존재하지 않는다는 문제가 발생했다. 또한 지금까지의 연구는 비 오는 영상을 디레이닝한 후 다른 세그멘테이션 모델로 추론하는 방식이었으나, 성능이 좋지 않았다.

기존의 비 오는 영상과 원본 영상이 쌍으로 존재하는 데이터셋은 주행 영상이 아닌 다른 도메인의 영상으로 구성되어 있어 세그멘테이션 GT가 없고,

세그멘테이션 데이터셋은 비 오는 영상을 제공하지 않는다. 이를 해결하기 위해 본 연구에서는 비 오는 영상을 생성할 수 있으며, 생성된 비 오는 영상에서 세그멘테이션이 잘 이루어지도록 해야 한다고 주장한다. DSANet은 배경 특징과 비 특징을 분리하고 이를 다시 합성하여 배경을 더 잘 보존하고 다양한 비 효과 적용을 용이하게 함으로써 비 오는 영상을 증강할 수 있다.

본 연구에서는 DSANet을 활용한 rainy image augmentation을 통해 자율주행 영상을 비 오는 영상으로 생성하고, 이를 사용해 세그멘테이션 모델을 파인튜닝하였다. 그 결과, 기존 세그멘테이션 모델에 비해 우리 모델의 성능이 훨씬 우수했으며, 기존의 디레이닝 모델을 통해 비를 제거한 후 사전 학습된 세그멘테이션 모델에 적용한 다른 모델들에 비해 탁월한 성능을 보였다.

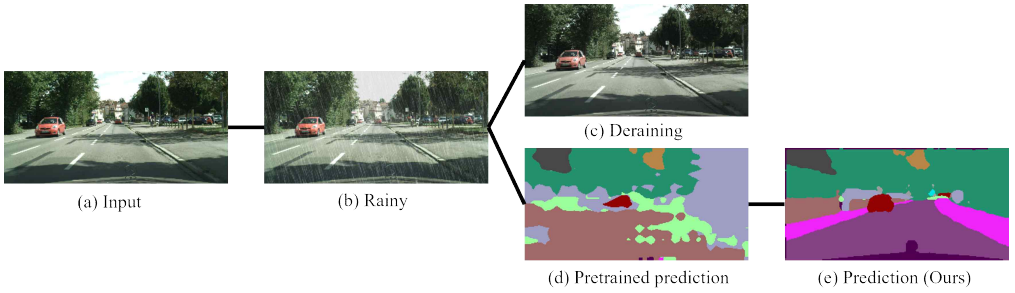
【주요어】 어그멘테이터, 세그멘테이션, 자율주행, 자율주행자동차

목 차

표 목 차

그 림 목 차

제 1 장 서 론



[그림 1-1] DSANet과 파인튜닝된 세그멘테이션 모델을 통해 도출된 모든 기여. (a) Cityscapes의 깨끗한 영상 (b) Cityscapes의 비 오는 영상 (c) DSANet을 활용하여 비 오는 영상의 비를 제거한 영상 (d) 사전 학습된 DeepLabV3 모델에 비 오는 영상을 입력한 결과 (e) 파인 튜닝된 모델에 비 오는 영상을 입력한 결과

비는 일반적인 기상 조건 중 하나로, 물체 감지[4] 및 분할[5]과 같은 자율 주행 시스템[1, 2, 3]과 관련된 분야에서 화질 저하의 부정적인 주요 요인이 되며 결과적으로 악천후 조건에서의 자율 주행은 어려운 과제이다. 현재 맑은 날씨 조건의 주행 상황에서의 시맨틱 세그멘테이션 모델은 상당히 큰 발전을 이루었으나 악천후와 같이 이미지 품질 저하[9]시키는 경우에는 성능이 급격히 떨어진다. 시맨틱 및 인스턴스 세그멘테이션과 같은 자율 주행[5, 6, 7, 8]에 중요한 기계 인식 작업에서 비 오는 상황에도 세그멘테이션 성능이 저하되지 않도록 DSANet과 세그멘테이션 파인 튜닝 모델을 제안한다. 하지만 비 오는 주행 상황일 때 세그멘테이션 성능을 높이기 위해서는 비 오는 주행 영상이 입력으로 사용해야 하기 때문에 필요하다. 대표적인 합성 비 데이터셋[11, 12, 13, 14] 경우에는 주행 영상이 아니라 다른 도메인의 영상으로 구성되어 있으며 비 합성 이미지로써 사실적인 비의 형태나 다양성을 실제와 같이 반영하기 어렵다. 또한 (SPAdata)실제의 비 장면에서 사진 촬영을 하거나 인터넷으로 수집할 경우도 있으나 비용이 많이 발생하여 이는 큰 제

약 사항으로 작용한다[10].이에 따라, 비 오는 주행 영상의 부족으로 인하여 우리는 실제 빗줄기와 유사하게 비 오는 영상을 생성할 수 있는 방법을 모색해야 한다. 초기 연구에서는 주로 실사 렌더링 기법[14, 17, 18]을 사용하여 빗줄기를 생성하고 생성한 비와 깨끗한 원본 영상과 결합하여 비 데이터셋을 구축하는데 중점을 두었다. 그러나 비의 각도, 위치, 깊이, 강도, 밀도, 길이, 폭 등이 다양하므로 실제 비의 복잡한 분포를 반영하기 어렵다. [16, 17]에서는 선명한 영상과 빗줄기 간의 더 나은 융합 매커니즘을 연구하였으나 빗줄기의 일부 파라미터를 수동으로 설정하여 합성해야 하므로 매우 복잡한 실제 빗줄기의 형태를 정확하게 반영하지 못한다. 빗줄기 자체보다 깨끗한 원본 영상과 비 오는 영상 간의 격차를 완화하는 것에만 주목하기 때문이다. 따라서, DSANet은 배경 콘텐츠와 비 콘텐츠를 분리하여 서로 다른 비 콘텐츠와 배경 콘텐츠를 합성하여 새로운 비 오는 영상을 만들 수 있을 뿐만 아니라 실제 비 오는 영상과 유사하게 학습하기 위하여 특징단에 직접적으로 손실을 부여하여 각각의 콘텐츠에 더욱 집중하도록 한다. 또한 생성된 영상을 다시 분리 및 합성 시뮬레이션을 진행하여 원본 영상을 생성하도록 유도하며 특징 간의 분리와 합성을 더 강력하게 도와준다. 그림 1-1에서 DSANet은 새로운 비 오는 영상을 생성할 수 있는 data augmentator이자 deraining 영상 생성 역할까지 책임지고 있다.

최근 시맨틱 세그멘테이션 네트워크가 자주 등장하고 있지만, 비 오는 영상과 원본 영상의 쌍의 부족이나 품질보다 대부분 비 제거에 초점을 두고 있다. SAM[15]와 VLTSeg[33] 같이 성능 좋은 기존 모델들도 비 오는 장면에서 시맨틱 세그멘테이션에 대해서는 기대에 미치지 못하며 지금까지는 우천 환경에서 특화된 시맨틱 세그멘테이션 네트워크가 많이 없다. 따라서 데이터의 부족이나 빗줄기의 다양성과 같은 요인을 해결해야 한다. [14]은 MultiScaleSE 브록과 비대칭 스킵을 이용하여 다양한 입력 영상 스케일에 잘 적응하도록 연구를 진행하였다. 일반적으로 현재까지는 다른 모델과 결합하지 못하고 모델 종속적이며 심지어 augmentator로써의 역할도 하지 못하고 있어 이 둘을 해결하기 위해 우리 모델을 제안한다. 본 연구는 DSANet을 통해 생성된 비 오는 주행 영상을 입력으로 DeepLabv3와 SegFormer를 파인 튜닝하

였다.

우리 기술의 주요 기여 내용을 요약해보면 다음과 같다:

- 비 오는 영상의 의미론적 분할 성능을 향상시키기 위해 모델 agnostic하게 데이터 증강을 적용하고자 한다.
- 비 오는 요소와 배경 요소를 분리하기 위해 특징 도메인에서 기능적 분리를 수행한다. 즉, 특징 도메인의 분리를 활용하여 비 오는 영상 데이터 증강기를 구축한다.
- 실험 결과 비 오는 장면 분할에 대해 경쟁력 있는 성과를 도출하였으며 state-of-the-art를 달성하는 수준은 아니지만, 우리 기술은 다른 세그멘테이션 작업에 대해서도 어느 정도의 universal한 성능을 보인다.

제 2 장 연구 배경

제 1 절 Synthetic rain image generation (Learnable data Augmentator)

기존의 비 데이터셋은 실제 비 영상과 배경 영상의 쌍을 사용할 수 없으므로 주로 실사 렌더링 기법[14]을 활용하여 빗줄기를 생성한 후, 이를 깨끗한 원본 영상과 혼합하여 합성된 비 데이터셋을 구축하는 것에 초점을 맞추었다. Garg와 Nayar는 비의 모양과 이미징 과정을 분석하고, 실사 렌더링 기법[18]으로 빗줄기 데이터베이스를 합성했다. 실제 SIRR 연구에서 자주 사용되는 합성 비 데이터셋에는 Rain100L/Rain100H[12], Rain800[14], Rain 1200[13], Rain 1400[11]이 있다. 하지만 사실적인 영상들로 이루어져 있지 않으며 합성 영상이라는 한계에 그쳤다. 또한 우리 논문에서 제시하는 주행 영상에 대한 데이터셋은 부족하며 다른 도메인의 영상만 존재할 뿐 아니라 사실적이지 않다는 단점이 있다. Wang 등은 합성된 비 데이터셋의 현실성이 부족하고 실제 비 이미지에 대한 공공 벤치마크가 부족하여 비 제거 방법의 평가가 어렵다는 문제점이 있었다. 비 데이터셋과 원본 영상이 쌍으로 된 데이터셋의 부족 문제로 인해, 날씨에 의해 손상된 영상을 생성하기 위한 방법으로 image-to-image 변환 분야가 크게 주목받고 있다. 피자티 등[19]은 빗방울이나 흙, 먼지와 같은 오클루전으로부터 장면을 분리하여 가려진 모델의 물리적 파라미터를 회귀하는 적대적 파이프라인을 활용하며 사실적인 변환을 생성할 수 있는 방법을 제안하였다. 이러한 생성 방법은 주로 인간의 주관적인 가정에 의존하며, 모델 파라미터를 설정해야 하는 경우가 많아 합성된 비의 다양성을 제한할 수 있다. 따라서 실제로 촬영하거나 수집한 비 영상에 정해진 비 제거 과정을 적용하여 깨끗한 이미지를 생성함으로써 다양한 자연 비 장면을 담고 있는 SPA-data[6] 모델은 대규모 비 데이터셋을 구축했다. 또한 [12]에서와 같이 비의 방향 및 비의 밀도와 같은 비의 물리적 특성을

연구하기 위한 일부 연구가 수행되었지만 비의 밀도를 세 가지 수준 중 하나로 분류하는 것은 매우 어려운 경우가 많으며 이러한 설정은 특정 유형의 비인 경우에만 사용할 수 있으며 영상 자체가 어두워질 경우 사용하기 힘들다. 또한, 비 영상 자체를 촬영하는 비용도 비싸기 때문에 다양한 형태의 비 영상을 포착하기는 어렵다. Adobe Photoshop의 경우도 특정 유형의 비에만 사용할 수 있거나 배경이 어두워져 디테일이 손실될 수 있다.

DSANet에서는 대규모 실제 비 데이터셋을 구축하기 위해 배경과 비 특징을 채널에 따라 분리하여 서로 다른 비 특징과 결합하여 Adversarial loss, Feature Consistency loss 및 MAE loss를 사용하여 실제 비 데이터셋과 유사하게 학습할 수 있도록 한다. 이에 따라 렌더링 방법으로는 달성하기 어려운 자연스러운 비 패턴을 만들기 유리하며 특정한 경우에만 한정되어 생성하지 않도록 다양한 비의 형태를 생성하도록 유도한다. 기존의 렌더링 기법과 딥러닝 기법을 활용한 모델들은 대부분 특정 데이터셋에서만 최적화되어 있고 다른 유형의 비나 다양한 환경에서 촬영된 이미지에 대해서는 성능이 저하될 수 있으며 비 제거 과정에서 비가 아닌 영상의 일부분을 잘못 제거하거나, 원본 이미지의 디테일을 손상시키기 쉽다. [20]에서도 실제 데이터와 합성 데이터의 배경이 크게 다를 경우, deraining결과가 좋지 못하며 흐린 배경 데이터셋으로 훈련된 경우에는 비 오는 영상과 비 오는 영상과 배경이 비슷해야만 가장 좋은 성능을 얻는 단점이 있어 도메인 간의 격차를 해결하지 못하였다. 하지만 DSANet은 이와 달리 영상 간에 사칙 연산으로 해결하는 것이 아니라 feature단에서 서로 내포하고 있는 특징들을 분리하여 학습하고 feature단에서 MSE loss를 통하여 학습하므로 이와 같은 단점들을 개선하였다. 비 오는 영상을 생성하는 연구들을 따라서, 주행 영상에 비 콘텐츠를 합성하여 다양한 유형의 비 오는 주행 영상을 생성할 수 있는 rainy data augmentator를 제안한다.

제 2 절 Deraining 기술

최근 연구 동향을 고려할 때, 시맨틱 세그멘테이션은 비 영상 데이터 생성

분야에서 흥미로운 발전을 보이고 있는 것으로 보인다. 특히, 이러한 분야에서는 'deraining' 기술이 주목받고 있으며, 이는 비가 내리는 영상에서 비의 영향을 제거하는 기술을 의미한다. 대표적인 딥러닝 기반 deraining 알고리즘 기법들에 대해 알아보고자 한다. 딥러닝 방법은 빗물 제거에서 탁월한 성능을 보인다. 예를 들어, RESCAN은 상황 정보를 최대한 활용하기 위해 컨볼루션 및 순환 신경망 기반 방법을 소개한다. DetailNet은 이전 기반 심층 세부 네트워크를 사용하여 음의 잔여 정보로 빗줄기를 추정한다. ID-CGAN은 빗방울과 그 이웃 배경에 주의를 기울이기 위한 attention-based GAN을 제안했다. PreNet은 입력 및 중간 레이어를 점진적으로 처리하기 위해 얕은 잔여 네트워크(shallow residual network)가 반복적으로 실행된다. 최근 deraining 방법은 다양한 크기와 방향의 빗줄기를 활용하기 위해 multi-scale learning을 통합하기 시작했다. 예를 들어, MSPFN은 빗줄기 정보의 미세 융합을 지도하기 위해 multi-scale pyramid architecture를 활용하며 feature map을 얻기 위해 attention mechanism을 적용한 네트워크이다. Syn2Real의 작업은 합성 이미지와 레이블이 지정되지 않은 실제 비오는 이미지를 사용하여 네트워크 성능을 높일 수 있는 Gaussian 프로세스 기반 준지도 학습(semi-supervised learning)을 개발했다. [21]에서는 준지도 학습 부분과 지식 증류(knowlege distilation) 부분을결합하여 네트워크를 제한하도록 rain direction regularizer를 설계했다. unpaired deraining 방법의 경우, CycleGan은 unpaired image-to-image 변환 문제를 해결하고 deraining 프로세스를 이루기 위해 널리 사용되었다. 최근 작업[22, 23, 24, 25]은 모두 개선된 CycleGAN 아키텍처와 제한된 전이 학습을 사용하여 비가 오는 이미지 도메인과 비가 내리지 않는 이미지 도메인을 공동으로 사용합니다. 앞에서 언급했듯이 이러한 방법은 도메인 지식이 비대칭이기 때문에 주기 일관성 제약 조건(cycle-consistency constraints)을 사용하여 두 도메인 간의 정확한 변환을 학습하기는 어렵다. 그 대신, 우리의 접근 방식은 이러한 한계를 극복하고 adversarial train과 disentangling 기술을 통합하여 짝이 없는 환경에서 derainig 견고성을 더욱 향상시킬 수 있다.

제 3 절 Semantic Segmentation

비 오는 환경에서의 semantic segmentation은 자율 주행, 감시 시스템 및 기타 실외 컴퓨터 비전 응용 분야에서 중요한 과제이다. Semantic segmentation은 이미지 노이즈 제고, 이미지 디블러링, 이미지 디레이닝과 같은 저수준 비전 작업에 유용하게 사용되어 왔다. 비가 내리는 동안 영상 품질이 저하되고, 물체 경계가 모호해지며, 배경 잡음이 증가하여 semantic segmentation의 정확도를 떨어뜨릴 수 있다. 따라서, 높은 수준의 비 오는 환경에서의 semantic segmentation의 난이도가 높아진다. 예를들어, 자율 주행 시스템에서 중요한 문제 중 하나는 비 오는 날의 도로 주행 segmentation이다. 또한, 비 오는 날씨에서의 주행 영상은 부정확한 segmentation 결과로 인해 세그멘테이션 알고리즘이 경계를 정확하게 파악하기 어려워진다. 우리 기술은 비 오는 환경에서도 Semantic Segmentation이 잘되게 하기 위하여 비 오는 영상을 입력으로써 파인 튜닝하는 구조로 설계하였다. 먼저, 완전 컨볼류션 네트워크(FCN) 기반 모델은 다양한 semantic segmentation 벤치마크에서 상당한 성능 개선을 입증해왔다. FCN[26]은 엔드 투 엔드 방식으로 픽셀 대 픽셀 분류를 수행하여 이미지의 semantic segmentation을 가능하게 한다. Dilated Convolution과 Atrous Spatial Pyramid Pooling (ASPP)은 수용 영역을 확대하여 더 많은 문맥 정보를 활용하는 기술로, DeepLab 시리즈에서 주로 사용된다[27, 28]. FCN 이후의 연구는 수용 영역 확대, 문맥 정보 세분화, 경계 정보 도입, 주의 메커니즘 설계, AutoML 기술 사용 등 다양한 측면에서 FCN을 개선해왔다[29-32]. 이러한 접근법은 시맨틱 세그멘테이션 성능을 크게 향상시켰지만, 비 오는 환경에서는 여전히 한계가 존재합니다. 비전에서 가장 널리 사용되는 문제 중 하나인Semantic Segmentation은 일관된 개체 클래스에 속하는 이미지 부분을 함께 클러스터링(Clustering)하는 작업이다. 이미지의 각 픽셀은 범주에 따라 분류되므로 픽셀 수준 예측의 한 형태이다. 우리는 주로 PSPNet과 HRNet인 Semantic Segmentation을 위한 두 가지 방법을 소개한다. 특히 Semantic Segmentation을 위한 우수한 프레임워크인 PSPNet은 다른 region based context의 feature를 완전히 활용하는 새로운

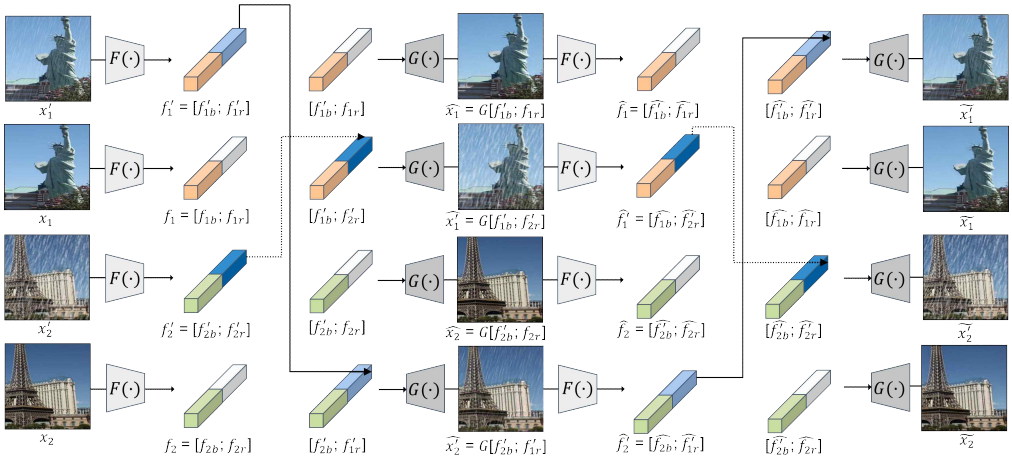
PPM(Pyramid Pooling Module) 모듈을 제안한다. 고품질의 장면 파싱 결과는 글로벌 사전 표현의 도움으로 생성된다. HRNet의 경우 고해상도 표현 네트워크라는 새로운 분할 네트워크를 개발했다. 이 네트워크는 일반적인 CNN이 한 개의 가지(branch)를 갖고 있는 것과 달리 입력 크기의 활성화 지도에 해당하는 가지들도 갖으면서 각 가지의 활성화 지도가 서로 정보를 교환하는 CNN이다. 기존 딥러닝 방법은 특징 추출 및 압축 과정에서 해상도가 낮아져 객체에 대한 특징이 손실되어 분할 정확도가 떨어지는 문제가 있다. 따라서 영상분할 분야에서 높은 정확도를 나타내고 있는 DeepLabv3 모델을 사용하였다. DeepLabv3 는 기존 ResNet 구조에Atrous convolution을 활용한 것이고 DeepLabv3는 Depthwise separable convolution과 Atrous convolution을 결합한 Atrous separable convolution을 제안했다. SegFormer는 Transformer 기반의 인코더를 사용하여 이미지의 장거리 의존성을 효과적으로 캡처하고, 다중 레벨의 특징을 결합하여 다양한 스케일의 정보를 통합한다.

그러나 현재까지의 세그멘테이션 네트워크들은 비 오는 영상을 입력으로 사용할 경우 모두 성능이 저조했으며, 비 오는 영상에서도 더 나은 결과를 얻기 위해서는 네트워크를 재학습시켜야만 했다. 본 연구에서는 DeepLabv3와 SegFormer 네트워크를 파인 튜닝하여 기존의 파인 튜닝하지 않은 모델과 성능을 비교한 결과, 비 오는 영상을 통해 파인 튜닝한 경우 훨씬 더 뛰어난 성능을 확인할 수 있었다. 이에 따라 우리는 DeepLabv3를 파인 튜닝하여 비 오는 영상을 효과적으로 분할할 수 있도록 설계하였다. 본 논문에서 제안하는 파인 튜닝된 시맨틱 세그멘테이션 모델은 어떤 시맨틱 세그멘테이션 모델을 사용하여 파인 튜닝하더라도, 기존 모델에 비해 비 오는 이미지에서 현저히 향상된 성능을 보입니다. 또한, Ablation study의 그림 4-1에서와 같이 DSANet을 통해 end-to-end 학습도 진행하였으며 더 깨끗한 비 오는 영상을 생성하여 비를 제외한 다른 노이즈는 제거한 상태로 더욱 세그멘테이션 성능을 극대화한 결과도 볼 수 있다.

제 3 장 DSANet (Disentangling and Synthesizing Augmentator Network)

제 1 절 Framework Architecture

자율 주행 분야에서 중요한 문제 중 하나는 빗줄기 영상과 원본 영상 간의 짝지어진 영상을 제공하는 데이터셋이 부족하며 학습에 사용 가능한 이미지 수는 현저히 부족하다. 이로 인해 비 오는 환경에서도 효과적인 세그멘테이션 모델을 학습하는 것이 어렵다.



[그림 3-1] DSANet: Disentangling and Synthesizing Augmentator Network (DSANet)

본 연구는 자율 주행 시스템의 실제 비 오는 상황에서도 효과적인 세그멘테이션을 달성하는 것이 본 연구의 목적이다. 세그멘테이션 모델을 학습하기 이전에 본 연구에서는 대량의 빗줄기 영상 생성이 필수적이다. 이에 따라 다양한 형태의 빗줄기를 포함하는 영상 확장을 위해 빗줄기 영상 데이터 확장을 위해 데이터 증강기 DSANet를 제안한다.

구체적으로 DSANet의 전반적인 구조는 분리와 합성 과정을 모두 포함한다. $F(\cdot)$는 배경 특징과 비 특징을 분리하는 과정을 수행하고 $G(\cdot)$는 배경 특징과 비 특징을 합성하는 과정을 수행한다. 제안된 DSANet은 서로 다른 쌍의 영상 x_1', x_1와 x_2', x_2이 $F(\cdot)$를 통과하게 되어 추출된 특징 f_1'의 채널 앞부분은 배경 특징, 뒷부분은 빗줄기 성분 특징으로써 분리하고 서로 다른 영상의 배경 특징과 비 특징을 concat한다. 즉, 그림에서 x_1'의 배경 특징인 f_{1b}'과 x_2'의 빗줄기 특징인 f_{2r}'을 concat하고 x_2'의 배경 특징인 f_{2b}'과 x_1'의 빗줄기 특징인 f_{1r}'을 concat한다. 그리고 $G(\cdot)$를 통해 x_1'의 배경에 x_2의 비를 포함하는 $\widehat{x_1'}$, x_2'의 배경에 x_1'의 비를 포함하는 $\widehat{x_2'}$를 얻게 된다. 비 내리지 않는 영상 x_1을 입력 영상에 포함시켜 f_{1r}와 같이 빗줄기가 존재하지 않는 특징을 얻어 다시 한번 $\widehat{x_1}$, $\widehat{x_1'}$, $\widehat{x_2}$, $\widehat{x_2'}$를 $F(\cdot)$의 입력 영상으로 넣어 이 과정을 한 번 더 반복하게 된다. 또 한 번의 얽힘과 풀림 현상을 거친 후 마지막에 입력 영상과 유사한 영상이 출력되어 MSE, L_1, Adversarial loss를 통해 특징과 영상에 의미를 부여한다. 이를 통해 배경과 비 특징 간의 분리 관계를 더 강력하게 구축할 수 있으며 배경은 일관성 있게 유지하며 서로 다른 비 특징 공간 사이에서의 변환만 가능해진다. 따라서 비 내리지 않는 원본 특징에 새로운 빗줄기 특징을 포함시키는 특징 전파에 큰 이점이 된다. 또한 배경 특징과 비 특징을 추출하고 분리함으로써 분리한 빗줄기 성분 특징을 기반으로 새로운 비 모양을 생성할 수 있으며 이 과정을 반복하여 다양한 비의 형태를 포함한 비 오는 영상을 생성 가능하며 이로써 DSANet의 일반화 능력을 향상시킬 수 있다.

결과적으로 DSANet는 그림에서의 $\widehat{x_1'}$와 $\widehat{x_2'}$와 같이 원본 영상에 다른 형태의 빗줄기를 추가함으로써 새로운 비 오는 영상을 생성할 수 있고 실제 환경에서 더 잘 일반화되도록 도와 효과적으로 기상 현상을 시뮬레이션하고 실제 비 오는 상황을 재현하는 데 도움이 된다. 이로써 DSANet은 기존의 영상을 모의실험하고 예측하는 데에 적용할 수 있는 유용한 도구가 될 수 있다.

제 2 절 Training Objective

이 섹션에서는 손실 함수와 목적 함수에 대해 자세히 설명한다. 중간에 생성되는 특징 및 출력에 대해 feature loss, background loss, adversarial loss 및 reconstruction loss를 적용한다.

1) Feature Loss

비 오는 영상 x_1'과 그의 원본 영상 x_1은 동일한 배경 콘텐츠를 포함하고 있으며 x_1'이 $F(\cdot)$를 통하여 추출된 특징 f_{1b}'은 x_1이 $F(\cdot)$를 통하여 추출된 특징 f_{1b}과 서로 특징들 간의 일관성을 가지도록 하기 위해 이 손실을 사용한다. 또한 이 손실은 두 배경 특징 간의 각 특징의 절대 차이를 계산하여 $F(\cdot)$가 원본 영상의 배경 콘텐츠만을 올바르게 재현하도록 유도한다. $G(\cdot)$로부터 재구성된 $\widehat{x_1}$, $\widehat{x_1'}$ 영상은 원본 영상인 x_1', x_1와 동일한 배경 콘텐츠로부터 유래한 영상이므로 $\widehat{x_1}$, $\widehat{x_1'}$ 영상이 $F(\cdot)$를 통과하면 나오는 배경 성분 특징 f_{1b}'과 $\widehat{f_{1b}'}$도 동일한 배경 콘텐츠로부터 도출된 특징으로써 MSE 손실을 사용하면 배경 성분 특징들이 일관성을 유지하도록 하며 원본 영상의 배경 특징 성분 간의 차이를 최소화하는데 도움이 된다. 아래는 x_1'와 x_1 영상에 관련된 수식이며 x_2'와 x_2영상도 포함되어 있다.

$$L_{mse}^{bf} = \|f'_{1b} - f_{1b}\|_2^2 + \|\widehat{f'_{1b}} - f_{1b}\|_2^2 + \|\widehat{f_{1b}} - f_{1b}\|_2^2 + \|\widehat{f_{1b}} - \widehat{f'_{1b}}\|_2^2 + \|f'_{2b} - f_{2b}\|_2^2 + \|\widehat{f'_{2b}} - f_{2b}\|_2^2 + \|\widehat{f_{2b}} - f_{2b}\|_2^2 + \|\widehat{f_{2b}} - \widehat{f'_{2b}}\|_2^2 \quad (1)$$

또한, 원본 영상으로부터 도출된 배경 성분 특징이 비 오는 영상으로부터 도출된 배경 성분 특징보다 더 깨끗한 배경 성분을 담고 있을 가능성이 높다. 따라서, 배경 성분 특징은 배경만을, 비 성분 특징은 배경을 제외한 부분만을 닮아가도록 학습시켜 확실히 분리되게 학습시킬 수 있으며 $F(\cdot)$를 통과한 배경 성분 특징은 원본 영상으로부터 $F(\cdot)$를 통과하여 추출된 특징을 따라가도록 학습시키기 위해 MSE 손실을 사용한다.

2) Background Loss

원본 영상에 접근하지 않으면서 배경 특징과 비 특징의 분리를 강화하기 전체 구조의 중간 과정에서 생성되는 $\widehat{x_1}$와 $\widehat{x_2}$영상에 평균 절대 오차 손실(L_1)을 적용하여 깨끗한 원본 영상의 배경 성분이 유지될 수 있도록 도와주며 분리와 합성 과정에서 중요한 역할을 수행한다.

$$L_{B1} = E_{\mathrm{B}1}[\| \chi_1 - \widehat{\chi_1} \|_1]$$
$$L_{B2} = E_{\mathrm{B}2}[\| \chi_2 - \widehat{\chi_2} \|_1] \quad (2)$$

3) Adversarial Loss

원본 영상에 접근하지 않으면서 배경 특징과 비 특징의 분리를 강화하기 위해, 분해한 뒤 x_1 및 x_2의 배경과 비 성분이 없는 비 특징 부분을 합성해 생성한 $\widehat{x_1}$ 과 사실적인 원본 영상을 구별하는 원본 영상(비 안 오는 영상) 판별자 D_x를 사용한다. 그리고 생성된 영상을 토대로 다시 이 과정을 반복하여 $\widehat{x_1'}$ 및 $\widehat{x_2'}$의 배경과 비 성분이 있는 비 특징 부분을 합성해 생성한 $\widetilde{x_1'}$ 및 $\widetilde{x_2'}$(비 오는 영상)과 사실적인 입력 영상(비 오는 영상)을 구별하는 입력 영상(비 오는 영상) 판별자 $D_{x'}$을 사용한다.

한편, $G(\cdot)$는 입력 영상과 주어진 조건에 따라 배경 성분 특징과 비 성분 특징을 합성하므로 비 오는 영상과 비 안 오는 영상 모두를 생성 가능하다. 보다 사실적인 비 안 오는 원본 영상 또는 비 오는 영상을 생성하여 판별기를 속이려고 시도한다.

사실적인 배경 영상을 생성하는 $G(\cdot)$에 대해서는 중간 과정에 생성되는 비 안 오는 영상과 마지막에 출력되는 비 오는 영상 모두에게 적대적 손실을 부여한다. 따라서, $G(\cdot)$는 주어진 배경 특징과 비 특징을 조합하여 비 오는 영상과 복원된 원본 영상 모두 생성할 수 있다.

먼저, $F(\cdot)$를 통과하여 추출된 특징 f_i는 해당 i번째 영상의 배경 특징과 비 특징을 나타낸다. 해당 i번째 배경과 j번째 비 특징이 결합 되어 $S(f_i; f_j)$로 합성된다.

$$f_i = [f_{ib}; f_{ir}] \quad (3)$$

$$S(f_i; f_j) = [f_{ir}; f_{jr}] \quad (4)$$

여기서 x_1', x_1 영상이 $F(\cdot)$를 통하여 추출된 x_1'의 배경 특징과 x_1의 비 특징을 결합하여 $G(\cdot)$는 f_{1b}'에 대한 배경 특징과 f_{1r}에 대한 비 특징을 합성한 후에 비 안 오는 영상 $\hat{x_1}$을 생성하게 된다. 생성된 $\hat{x_1}$와 x_1간의 적대적 손실을 적용한다.

$$\begin{aligned} L_{adv}^{B}(D_B, G, S, F) = & E_{\chi_1}[\log D_B(\chi_1)] \\ & + E_{\chi, \chi_1}[\log(1 - D_B(G(S(F(\chi'_1), F(\chi_1)))))] \end{aligned} \quad (5)$$

x_2', x_2에도 이와 동일하게 진행하여 $F(\cdot)$를 통하여 추출된 x_2'의 배경 특징과 x_2의 비 특징을 결합하여 $G(\cdot)$는 f_{2b}'에 대한 배경 특징과 f_{2r}에 대한 비 특징을 합성한 후에 비 안 오는 영상 $\hat{x_2}$을 생성하게 된다. 생성된 $\hat{x_2}$와 원본 영상 x_2간의 적대적 손실을 적용한다. 따라서, 전체적인 비 안 오는 원본 영상에 대한 adversarial loss는 아래와 같다.

$$\begin{aligned} L_{adv}^{B}(D_B, G, S, F) = & E_{\chi_1}[\log D_B(\chi_1)] \\ & + E_{\chi, \chi'_1}[\log(1 - D_B(G(S(F(\chi'_1), F(\chi_1)))))] \\ & + E_{\chi_1}[\log D_B(\chi_2)] \\ & + E_{\chi_2, \chi'_2}[\log(1 - D_B(G(S(F(\chi'_2), F(\chi_2)))))] \end{aligned} \quad (6)$$

또한, 생성된 비 오는 영상에 대한 절대적 손실을 제안하며 $\hat{x_1}$, $\hat{x_1}'$, $\hat{x_2}$, $\hat{x_2}'$이 $F(\cdot)$를 통과하여 분리된 특징끼리 합한 $[\hat{f_{1b}}'; \hat{f_{1r}}']$ 및 $[\hat{f_{2b}}'; \hat{f_{2r}}']$ 이 $G(\cdot)$를 통하여 비 오는 영상을 생성하게 된다. 생성된 비 오는 영상이 입력 영상에 비해 더 사실적으로 보이게 하기 위하여 절대적 손실을 사용하였으며 이로써 분리된 특징에서 배경 특징은 더욱 더 배경 특징에 집중하고 비 특징은 비 관련 부분에만 더욱 집중을 할 수 있도록 학습이 되고 배경 및 비 특징이

서로 잘 얽히게 하여 비 오는 영상을 생성할 수 있도록 도와준다.

따라서, 전체 생성된 합성 영상에 대한 절대적 손실을 아래와 같이 정의한다.

$$\begin{aligned} L_{adv}^{B}(D_B, G, S, F) = & E_{\chi_1}[\log D_S(\chi'_1)] \\ & + E_{\hat{\chi_1}, \hat{\chi'_1}}[\log(1 - D_S(G(S(F(\hat{\chi_1}), F(\hat{\chi'_1})))))] \\ & + E_{\chi_1}[\log D_S(\chi'_2)] \\ & + E_{\hat{\chi_2}, \hat{\chi'_2}}[\log(1 - D_S(G(S(F(\hat{\chi_2}), F(\hat{\chi'_2})))))]] \end{aligned} \quad (7)$$

생성된 합성 영상과 실제 이미지에 대해 비 오는 영상 및 비 안 오는 원본 영상 판별기를 적용하여 적대적 손실을 생성한다.

4) Reconstruction Loss

Feature Consistency Loss와 Adversarial Loss 외에도 원본 영상에 대응하는 생성된 비 오는 영상과 비 안 오는 영상의 경우, 분리 네트워크 $F(\cdot)$와 합성 네트워크 $G(\cdot)$의 훈련을 감독하기 위해 평균 절대 오차(MAE) 손실을 활용한다. 이는 다음과 같이 정의된다.

$$\begin{aligned} L_{\text{rec}}^{B1'} &= E_{\text{B1}'}\left[\| \text{x}_1' - \widetilde{\text{x}_1'} \|_1 \right] \\ L_{\text{rec}}^{B1} &= E_{\text{B1}}\left[\| \text{x}_1 - \widetilde{\text{x}_1} \|_1 \right] \\ L_{\text{rec}}^{B2'} &= E_{\text{B2}'}\left[\| \text{x}_2' - \widetilde{\text{x}_2'} \|_1 \right] \\ L_{\text{rec}}^{B2} &= E_{\text{B2}}\left[\| \text{x}_2 - \widetilde{\text{x}_2} \|_1 \right] \end{aligned} \quad (8)$$

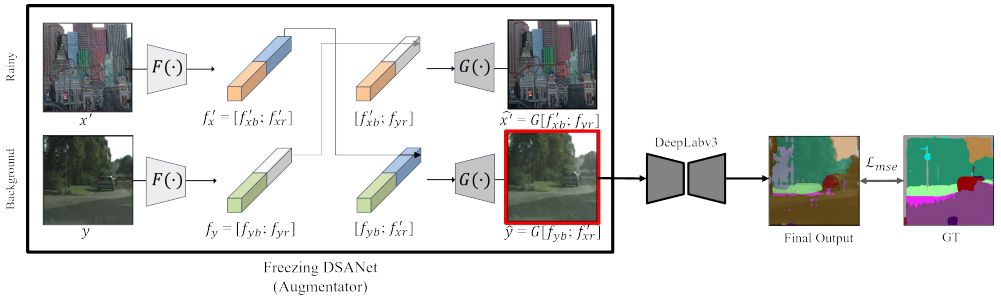
5) Full Objective

전체 목적 함수는 다음과 같이 Feature Consistency Loss와 Adversarial Loss 및 MAE Loss를 포함한다.

$$L(F,G,D_B,D_S,S) = \lambda_{mse} L_{\text{mse}}^{\text{bf}} + \lambda_{\text{B}}(\text{L}_{\text{B1}} + \text{L}_{\text{B2}}) + \lambda_{\text{adv}}(\text{L}_{\text{adv}}^{\text{B}} + \text{L}_{\text{adv}}^{\text{S}}) + \lambda_{\text{rec}}(\text{L}_{\text{rec}}^{\text{B1}'} + \text{L}_{\text{rec}}^{\text{B1}} + \text{L}_{\text{rec}}^{\text{B2}'} + \text{L}_{\text{rec}}^{\text{B2}'}) \quad (9)$$

전체 목적 함수를 부여함으로써 배경 및 비 특징을 확실히 분리할 수 있으며 비 오는 영상 데이터셋의 부족을 완화시키기 위해 실제와 같은 새로운 비 오는 영상을 생성할 수 있게 된다. 우리의 프레임워크에서 각 분리 및 합성 네트워크는 배경과 비 콘텐츠 정보를 보존하면서 서로 다른 콘텐츠 정보와 합쳐 새로운 영상 생성에 무리 없이 출력의 영역을 보장하기 위해 적어도 하나의 비지도 적대적 손실과 하나의 Feature Consistency Loss로 제약을 받으며, 결과적으로 새로운 비를 포함하는 영상을 생성할 수 있게 분리와 합성을 역할을 효과적이게 해내도록 프레임워크가 안내한다.

제 3 절 Generating Rainy Scenes and Fine-tuning of Segmentation Models for Improved Performance



[그림 3-2] DeepLabV3를 활용한 파인 튜닝 세그멘테이션 모델 구조

비 오는 상황에서 자율 주행 시스템의 세그멘테이션 성능을 향상시키기 위해서는 대부분의 경우 비 오는 이미지와 비 안 오는 이미지가 쌍으로 제공되지 않아 부족하다. 이에 따라, 우리는 배경과 비 특징을 분리하고 새로운 비오는 장면을 합성하기 위해 $F(\cdot)$와 $G(\cdot)$를 활용하여 DSANet 네트워크 구조를 설계하였다. 이와 같이, 쌍으로 된 데이터셋이 없기 때문에, 우리는 Rain100H와 같은 자율주행과는 다른 도메인의 데이터셋을 사용하여

DSANet을 학습하고, 이후 세그멘테이션 모델을 파인 튜닝하였다. 그림 3-2과 같이, 초기 단계에서는 고정된 DSANet을 사용하여 추론을 수행한다. 비 오는 영상인 x'을 $F(\cdot)$에 입력하여 배경 특징과 비 특징을 분리하고, 주행 영상인 y도 $F(\cdot)$에 입력하여 배경 특징과 비 특징을 분리한다. 그리고 비 오는 특징인 f_{xr}'과 주행 영상의 배경 특징인 f_{yb}을 연결하여 $G(\cdot)$를 통해 합성을 수행한다. 이로써, 비오는 주행 영상인 $\hat{y}$을 생성할 수 있으며, 이를 통해 쌍으로 된 주행 영상을 대량으로 생성할 수 있다. 그리고 이제 생성된 비 오는 주행 영상, 즉 $\hat{y}$을 사용하여 세그멘테이션 파인 튜닝을 진행한다.

DeepLabv3모델을 파인 튜닝하여 세그멘테이션 모델로 활용하였다. 일반적으로 Deeplabv3 모델은 주행 영상을 입력으로 받아들여 해당 영상의 출력과 세그멘테이션 Ground Truth(GT) 사이의 교차 엔트로피 손실을 통해 학습된다. 그러나 우리는 비 오는 주행 영상을 직접 입력으로 사용하여 모델을 파인 튜닝했다. 이로써 비 오는 영상에 대한 세그멘테이션이 효과적으로 학습될 수 있게 되었다. 또한, 주어진 주행 영상이 비 오는 상황에서 세그멘테이션을 위한 효과적인 학습을 위해 우리는 DeepLabv3모델을 파인 튜닝하는 과정에서 몇 가지 추가적인 전략을 채택했다. 우선, 비 오는 주행 영상을 사용하기 위해 데이터 전처리 단계에서 강수량에 따른 픽셀 라벨링을 수행하고, 이를 통해 학습 데이터셋을 보강하였다.

기존에 학습된 DeepLabv3 모델을 적용한 결과와 파인 튜닝 후의 성능을 비교해본 결과, 파인 튜닝을 적용한 경우에 테스트 성능이 현저히 향상되는 것을 확인할 수 있다. 그리고 비 오는 영상을 사용하여 학습한 모델은 비 오지 않는 이미지만을 사용하여 학습한 모델보다 더 강력한 도메인 일반화 능력을 갖추게 되며 다양한 환경에서도 안정적으로 작동할 수 있는 능력이 향상된다. 비 오는 환경에서는 노이즈와 불규칙성이 더 많이 발생할 수 있으나 파인 튜닝을 통해 비 오는 영상을 사용하여 학습한 모델은 이러한 노이즈와 불규칙성에 대응할 수 있도록 능력을 키워준다. 이로써 우리의 제안은 자율주행 시스템이 비 오는 환경에서도 안정적으로 작동하며, 주변 환경을 정확하게 인식할 수 있도록 도움을 줄 수 있는 유망한 접근법임을 입증하였다. 따라서, 파인 튜닝을 통해 비 오는 영상을 사용하여 학습한 DeepLabv3 모델은

기존에 비 오지 않는 영상으로 학습된 모델에 비해 우수한 성능을 보이게 된다.

제 4 장 실 험

제 1 절 Datasets and Experimental Setting

- Cityscapes. 본 논문에서는 Cityscapes[1] 데이터셋을 이용하여 세그멘테이션 네트워크에 대한 실험을 진행하였다. Cityscapes 이미지는 5,000장 가량의 주행 이미지로 이루어져 있으며, 본 논문에서는 비 오는 주행 이미지를 다량으로 생성하기 위해 다른 비 오는 이미지로 구성된 데이터셋으로부터 비 영역을 추출하여 Cityscapes 데이터셋에 적용한다. 이렇게 함으로써 다량의 비가 내리는 주행 이미지를 얻어낼 수 있었으며, 따라서 원활하게 segmentation 실험을 진행할 수 있었다.
- RainHQ. Cityscapes 데이터셋이 비가 내리지 않는 실제 주행 이미지였던 것과 달리, RainHQ 데이터셋은 비 영역이 존재하는 데이터셋이다. 1,000장 가량으로 이루어져 있으며, 비가 내리는 이미지와 내리지 않는 이미지를 모두 제공한다. 본 논문에서는 이 RainHQ 데이터셋으로부터 비 영역을 추출하여 Cityscape 데이터셋에 적용할 수 있게끔 하고 있다. 이렇게 여러 데이터셋을 혼합하여 사용함으로써 각각의 장점을 더욱 살릴 수 있었다.
- KITTI. KITTI 데이터셋은 15,000장 가량의 주행 이미지로 이루어진 데이터셋이며, 본 논문에서는 테스트를 할 때 이용하였다. Cityscape와 같은 주행 이미지 데이터셋을 통해 모델의 파인 튜닝이 잘 되어 있었는지 확인할 수 있었다.

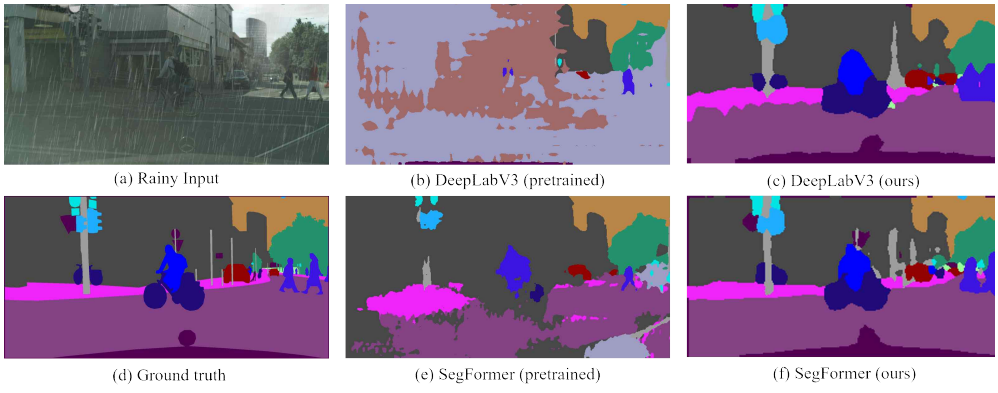
제 2 절 Evaluation Metric

세그멘테이션 결과를 평가하기 위해 mIoU와 mAP를 사용하며, 두 지표 모두 값이 높을수록 우수한 성능을 보인다. 또한, 그림 3-1에서 DSANet을 이용하여 비를 제거한 영상 $\widehat{x_1}$, $\widehat{x_2}$ 를 생성할 수 있다. DSANet의 비 제거

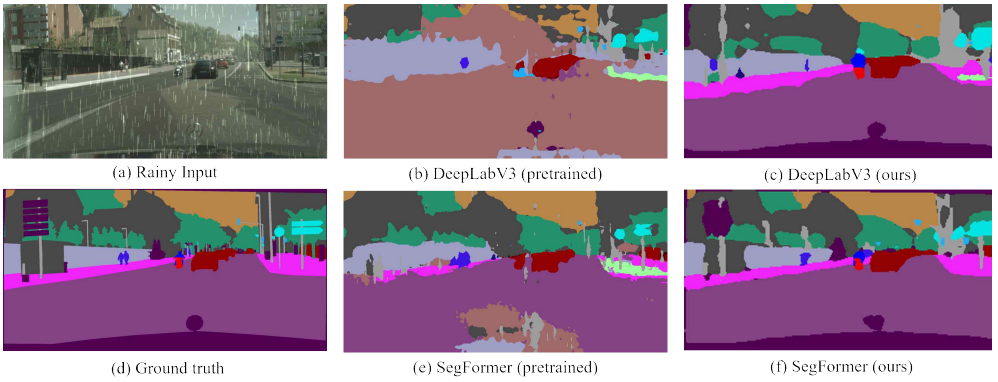
효과를 평가하기 위해 두가지 영상 품질 평가 방법을 사용하여 확인한다. PSNR(피크 신호 대 잡음 비율)과 SSIM(구조적 유사성 지수)을 사용한다. 이 두 평가 방법에서 값이 클수록 더 좋은 품질을 나타낸다. PSNR은 영상의 잡음 수준을 측정하는데 사용되며, 값이 높을수록 더 낮은 잡음 수준을 의미한다. 반면 SSIM은 영상의 구조적 유사성을 측정하는데 사용되며, 값이 1에 가까울수록 원본과 유사한 품질을 나타낸다.

제 3 절 Segmentation Experiments on Generated Rainy Images

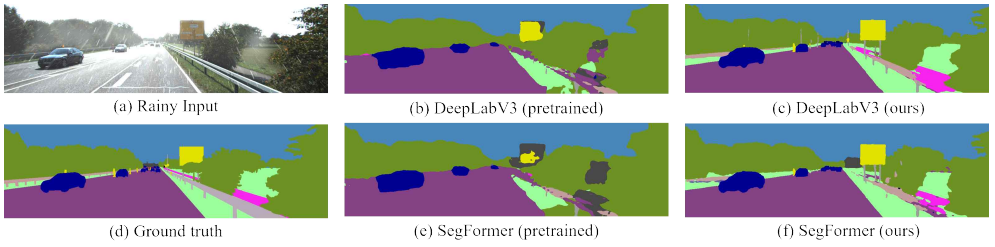
그림 4-3에서는 DSANet을 활용하여 Cityscape 및 KITTI 데이터셋으로부터 새로운 비가 내리는 환경을 시뮬레이션한 영상의 세그멘테이션 시각적 결과가 제시된다. DSANet은 다양한 비 오는 영상에 대한 세그멘테이션을 향상시키기 위해 새로 생성된 영상을 기반으로 DeepLabV3와 SegFormer를 파인튜닝하였다. 이로써 생성된 새로운 비 오는 데이터셋은 표 4-1에 정량적으로 평가되었다. Cityscape 및 KITTI 데이터셋은 각각 훈련 데이터셋과 테스트 데이터셋을 9:1 비율로 분할하여 학습하였다.



[그림 4-1] Visualization of segmentation results on Cityscapes



[그림 4-2] Visualization of segmentation results on Cityscapes



[그림 4-3] Visualization of segmentation results on KITTI

우리는 두 가지 세그멘테이션 모델을 각각 세분화하여 총 네 가지 기준 모델로 평가하였다. 기존에 미리 학습된 모델을 사용하여 평가한 결과, 만족스러운 세그멘테이션 결과를 얻지 못했다. 이는 비 영상으로 학습된 모델이 아니므로 기존에 학습된 모델이 세그멘테이션 결과의 의미적 정보를 손상시킬 수 있다는 것을 시사한다. 따라서, 우리는 생성한 비 오는 영상을 활용하여 기존 모델을 파인 튜닝하였으며, 이 과정에서 비 오는 영상에 대한 세그멘테이션 성능을 확인하였다.

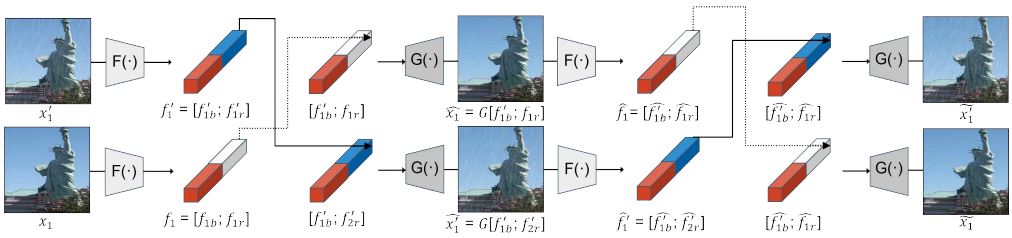
[표 4-1] Comparison results of difference segmentation models in terms of mIoU(%) and mAP(%) on Generated rainy images by DSANet

Method	Cityscapes		KITTI	
	mIoU (%)	mAP (%)	mIoU (%)	mAP (%)
DeepLabV3(Input-1)	54.48	61.53	72.08	76.18
DeepLabV3(End-to-End)	60.44	65.50	**77.51**	**79.35**
DeepLabV3(Ours)	57.54	64.33	74.19	78.28
SegFormer(Input-1)	59.03	66.11	61.60	66.94
SegFormer(End-to-End)	**66.74**	**71.27**	70.32	74.95
SegFormer(Ours)	63.25	68.60	66.86	71.63

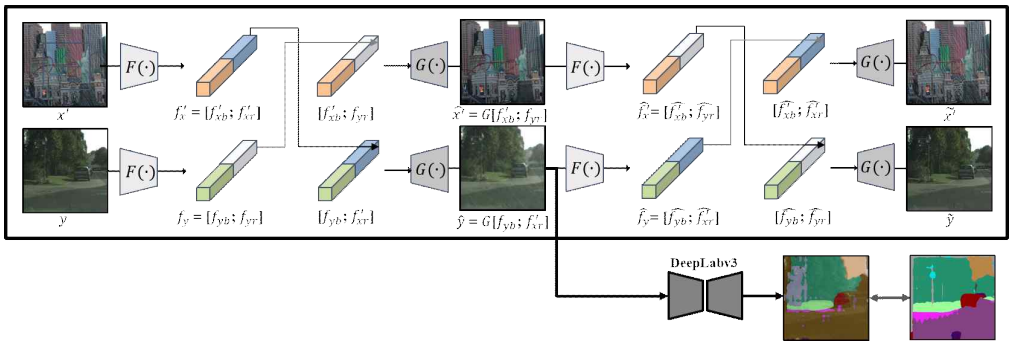
[표 4-2] The mIoUcomparison was conducted for each of the 19 categories in our dataset using both the DeepLabV3 and SegFormer models.

Method	road	sidewalk	build	wall	fence	pole	light	sign	veget°	terrain	sky	person	rider	car	truck	bus	train°	motor	bicycle	mIoU (%)
DeepLabV3	63.9	24.2	39.9	19.2	3.7	0.1	0	0.2	27.1	0	6.6	0.1	0	29.1	17.4	13.7	15.1	0	0	13.7
DeepLabV3 (Ours)	90.9	74.1	79.7	73.1	64.1	15.8	22.6	29.6	78.2	49.1	67.4	47.1	41.5	82.6	69.2	61	64.1	46.7	37.3	57.5
SegFormer	64.1	24.4	40.4	17.7	8.1	0.3	0	0.2	22.8	3.6	10.9	6.7	0.1	30.2	20.3	15.4	17.9	0.1	0.3	14.9
SegFormer (Ours)	**91.0**	**75.9**	**80.2**	**75.2**	**68.2**	**25.4**	**34.2**	**37.6**	**78.3**	**54.4**	**71.9**	**51.8**	**49.8**	**83.5**	**79.3**	**71.7**	**76.9**	**51.2**	**45.5**	**63.2**

결과적으로, 우리의 파인 튜닝 모델은 기존의 미리 학습된 모델에 비해 훨씬 향상된 결과를 보였다. 이는 DSANet으로 새로운 비 오는 영상을 대량 생성한 후 해당 영상을 활용하여 파인 튜닝한 모델이 성능적으로 효과적임을 입증한다. 더불어, DSANet을 통해 파인 튜닝된 세그멘테이션 모델은 어떠한 세그멘테이션 모델을 사용하더라도 향상된 성능을 제공함을 확인하였다.



[그림 4-4] A network trained with a single input without exchanging different rain in DSANet(Input-1)



[그림 4-5] End-to-End Fine-Tuning Framework for Segmentation Model Training

각 클래스에 대한 개별적인 영향을 더 자세히 분석하기 위해, 표 4-2에 mIoU를 정리하였다. 총 19개의 클래스가 포함되어 있으며, 우리가 제안한 파인 튜닝 모델이 모든 클래스에서 가장 높은 점수를 기록했음이 분명하다. 전반적으로, SegFormer를 우리의 방식으로 파인 튜닝한 모델이 성능이 가장 우수하지만, 기존에 미리 학습된 DeepLabV3과 우리의 방식대로 파인 튜닝한 DeepLabV3를 비교해보면 우리의 방식이 모든 클래스에서 기존 모델에 비해 상당히 효과적임을 확인할 수 있다.

제 4 절 Ablation Study

기존 이 섹션에서는 Cityscapes 및 KITTI 데이터셋을 활용해 새로 생성된

비 오는 영상에 대한 세그멘테이션 연구가 수행되어 Augmentation 방식의 유/무 및 Training Strategy의 효과를 평가한다.

1) Effectiveness of the presence of Augmentation

표 4-3에서 Input-1은 DSANet에서 그림 8과 같이 입력 영상 x'_1, x_1을 활용하여 분리와 합성 네트워크를 거쳐 학습된다. 이로써 우리의 contribution인 rainy image augmentator로써의 역할은 할 수 없으며 합성 영상을 입력으로 사용하여 학습하여 생성된 비 오는 영상은 처음 입력 영상과 유사한 영상이며 비 제거 영상을 생성하는데 그친다. 표 4-3에서의 Input-1과 우리 방식의 파인튜닝과 비교하여도 Input-1에서 작지만 성능 하락이 존재하며 이로써 우리 방식의 DSANet을 사용한 세그멘테이션 방법이 성능이 더 좋으며 그림 8에서 우리는 하나의 도메인에서 분리와 합성을 진행하므로 세그멘테이션과 같은 주행 영상 도메인과 같이 서로 다른 도메인의 경우 세그멘테이션 성능이 감소하게 된다. DSANet에서는 서로 다른 도메인과의 배경 및 비 성분을 분리 및 합성하여 Input-1에 비해 다양한 도메인에 이점을 보인다.

[표 4-3] The ablation study for different models'presence or absence of augmentation techniques and training strategies for the generated rainy images

Method	Cityscapes		KITTI	
	mIoU (%)	mAP (%)	mIoU (%)	mAP (%)
DeepLabV3(Input-1)	54.48	61.53	72.08	76.18
DeepLabV3(End-to-End)	60.44	65.50	**77.51**	**79.35**
DeepLabV3(Ours)	57.54	64.33	74.19	78.28
SegFormer(Input-1)	59.03	66.11	61.60	66.94
SegFormer(End-to-End)	**66.74**	**71.27**	70.32	74.95
SegFormer(Ours)	63.25	68.60	66.86	71.63

[표 4-4] PSNR and SSIM comparison on Test100, Test1200, Rain100L and Rain100H

Method	Test100		Test1200		Rain100L		Rain100H	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
GMM	23.55	0.832	28.74	0.878	29.02	0.862	15.13	0.447
RESCAN	25.00	0.835	30.51	0.882	31.52	0.890	26.18	0.827
RGN	26.56	0.866	32.42	0.941	32.74	0.898	25.25	0.841
JORDER	26.95	0.884	32.89	0.913	33.21	0.916	26.54	0.835
Rainy	20.67	0.779	25.23	0.771	25.48	0.808	12.26	0.359
DDN	21.19	0.806	25.86	0.791	26.12	0.836	12.57	0.418
Syn2Real	19.28	0.799	23.53	0.719	23.76	0.828	14.83	0.341
MOSS	21.83	0.825	26.65	0.814	26.91	0.855	16.49	0.436
MPRNet	**30.27**	**0.897**	32.91	0.916	33.57	**0.934**	28.52	0.872
EffDerain	21.06	0.785	25.70	0.786	25.96	0.814	15.32	0.462
PreNet	24.81	0.851	31.36	0.911	34.82	0.852	27.96	0.844
MSPFN	27.50	0.876	32.39	0.916	29.32	0.875	24.52	0.646
Ours	29.98	**0.897**	**32.94**	**0.948**	**34.89**	0.924	**28.73**	**0.874**

2) Effectiveness of Training Strategy

현재 우리의 접근 방식은 현재 DSANet을 훈련한 후에, Cityscapes 및 KITTI 데이터셋을 활용하여 DSANet을 통해 비가 오는 영상을 생성하고, 이를 세그멘테이션 파인 튜닝에 활용하는 것이다. 그러나 일반적으로 End-to-End 학습 방식은 일반화 능력을 더 효과적으로 갖추므로 입력과 출력 간의 관계를 보다 강화하기 위해 DSANet과 세그멘테이션 파인 튜닝 모델을 한 번에 학습하는 방식으로 접근한다. 여기서 x'은 RainHQ의 비가 오지 않는 영상을 의미하고, y는 Cityscapes를 나타낸다.

3) Comparison on Synthetic Rainy Dataset

표 4-4에서는 합성 비 오는 데이터셋에 대한 정량적 비교를 시행했다. 이는 DSANet이 비 오는 이미지 생성을 담당하는 Rainy image augmentator로써 탁월한 비 제거 능력을 보여주기 때문에 부가적으로 가장 뛰어난 PSNR과 SSIM을 보여준다. 특히 Rain100H 데이터셋은 가장 어려운 경우이지만, DSANet이 여전히 표 4-4에서 우수한 PSNR과 SSIM을 기록했으며 그림 1-1에서 시각적 결과로써 비 제거가 잘된 것을 확인할 수 있다.

따라서, 표 4-4의 정량적 평가 결과를 통해 우리의 기술이 특징 간의 분리를 효과적으로 달성했음을 확인할 수 있으며, 더불어 비 제거 효과에도 이점을 지녔음을 알 수 있다. 이로써 우리의 기술은 Rainy image augmentator, Deraining, 그리고 세그멘테이션까지 가능한 종합적이고 다재다능한 기술로 평가될 수 있다.

제 5 장 결 론

논문에서는 비가 내리는 기상 환경에서도 원활한 segmentation이 이루어질 수 있도록 하였다. 이를 위해 비가 오는 자동차 주행 이미지가 많이 필요하였으며, disentangling을 통해 비가 오는 이미지로부터 비와 배경을 분리한 뒤 주행 이미지에 이를 적용하여 비 오는 주행 이미지의 augmentation을 구현하였다. 또, 이렇게 생성된 비 오는 주행 이미지들을 이용해 segmentation 모델을 fine tuning 및 규격화 하여 다양한 segmentation 모델을 본 논문의 시스템에 적용하게끔 할 수 있게 하였다. 따라서 다양한 비 오는 주행 이미지를 통해 학습 및 원활한 segmentation을 하여 자율주행 자동차의 비 오는 상황에서의 성능을 향상시킬 수 있게 하였다.

다만, 본 논문에서 구현한 시스템은 2D 이미지만을 기반으로 작동하게끔 되어 있어, 향후 3D 이미지에 대해서도 시스템이 작동하게 하여 더 큰 성능 향상을 구현할 예정이다.

참 고 문 헌

1. 국외문헌

M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In IEEE CVPR, 2016. 2, 3, 5.

S. Yogamani, C. Hughes, J. Horgan, G. Sistu, P. Varley, D.O. Dea, M. Uricar, S. Milz, M. Simon, K.Amende, et al. Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving. arXiv preprint arXiv:1905.01489, 2019. 2.

F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. arXiv preprint arXiv:1805.04687, 2018. 2, 3, 5.

S. Li, I. Araujo, W. Ren, Z. Wang, E. Tokuda, R. Junior, R. Cesar-Junior, J. Zhang, X. Guo, and X. Cao. Single image deraining: A comprehensive benchmark analysis. In IEEE Conf. Comput. Vis. Pattern Recog., pages 3838-3847, 2019. 1, 2.

C.H. Bahnsen and T.B. Moeslund. Rain removal in traffic surveillance: Does it. matter? IEEE Trans. Intell. Transp. Syst., 20(8):2802-2819, 2018. 1.

M. Treml, J.A. Medina, T. Unterthiner, R. Durgesh, F. Friedmann, P. Schuberth, A. Mayr, M. Heusel, M. Hofmarcher, M. Widrich, et al. Speeding up semantic segmentation for autonomous driving. In MLITS, NIPS Workshop, volume 2, 2016.

H. Blum, P.E. Sarlin, J. Nieto, R. Siegwart, and C. Cadena. Fishyscapes:

A. benchmark for safe semantic segmentation in autonomous driving. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pages 0－0, 2019. 1.

A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollar. Panoptic segmentation. In IEEE CVPR, 2019. 1.

S. Zang, M. Ding, D. Smith, P. Tyler, T. Rakotoarivelo, and M.A. Kaafar. The impact of adverse weather conditions on autonomous vehicles: how rain, snow, fog, and hail affect the performance of a self-driving car. IEEE Vehicular Technology Magazine, 14(2):103－111, June 2019. 1.

T. Wang, X. Yang, K. Xu, S. Chen, Q. Zhang, and R.W. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12270－12279, 2019.

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, pages 3855－3863, 2017.

W. Yang, R.T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan. Deep joint rain detection and removal from a single image. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1357－1366, 2017.

H. Zhang and V.M.Patel. Density-aware single image de-raining using a multi-stream dense network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 695－704, 2018.

H.E. Zhang, V. Sindagi, V.M. Patel, Image De-Raining Using a Conditional. Generative Adversarial Network, IEEE Trans. Circuits Syst. Video Technol. 30 (11) (2020) 3943－3956.

A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T.

Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollar, and R. Girshick, "Segment anything," arXiv:2304.02643, 2023.

X. Hu, C. Fu, L. Zhu, and P. Heng. Depth-attentional features for single-image. rain removal. In IEEE Conf. Comput. Vis. Pattern Recog., pages 8022-8031, 2019. 1, 2, 3, 5, 6.

S. Halder, J. Lalonde, and R. Charette. Physics-based rendering for improving. robustness to rain. In Int. Conf. Comput. Vis., pages 10203-10212, 2019. 2, 5, 6.

K. Garg and S.K. Nayar. Photorealistic rendering of rain streaks. ACM Transactions on Graphics (TOG), 25(3):996-1002, 2006. 2, 6.

S.S. Halder, J.F. Lalonde, and R.Charette. Physics-based rendering for improving robustness to rain. In Proceedings of the IEEE International Conference on Computer Vision, pages 10203-10212, 2019. 2, 7.

Pizzati, Fabio, P. Cerri, and R. Charette. "Model-based occlusion disentanglement for image-to-image translation." Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XX 16. Springer International Publishing, 2020.

Ye, Yuntong, et al. "Closing the loop: Joint rain generation and removal via disentangled image translation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.

Cordts, Marius, et al. "The cityscapes dataset for semantic urban scene understanding." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

K. Han and X. Xiang. Decomposed cyclegan for single image deraining with unpaired data. In IEEE International Conference on Acoustics, Speech and Signal Processing, pages 1828-1832, 2020.

X. Jin, Z. Chen, J. Lin, Z. Chen, and W.Zhou. Unsupervised single

image deraining with selfsupervised constraints. In IEEE International Conference on Image Processing, pages 2761－2765, 2019.

Y. Wei, Z. Zhang, Y. Wang, M. Xu, Y. Yang, S. Yan, and M. Wang. Deraincyclegan: Rain attentive cyclegan for single image deraining and rainmaking. IEEE Transactions on Image Processing, 30:4788－4801, 2021.

H.Zhu, Xi Peng, Joey Tianyi Zhou, Songfan Yang, Vijay Chanderasekh, Liyuan Li, and Joo-Hwee Lim. Singe image rain removal with unpaired information: A differentiable programming perspective. In Proceedings of the AAAI Conference on Artificial Intelligence, pages 9332－9339, 2019.

Long, Jonathan, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

Chen, L. Chieh, et al. "Semantic image segmentation with deep convolutional nets and fully connected crfs." arXiv preprint arXiv:1412.7062 (2014).

Chen, L. Chieh, et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." IEEE transactions on pattern analysis and machine intelligence 40.4 (2017): 834-848.

Shaw, Albert, et al. "Squeezenas: Fast neural architecture search for faster semantic segmentation." Proceedings of the IEEE/CVF international conference on computer vision workshops. 2019.

Chen, Wuyang, et al. "Fasterseg: Searching for faster real-time semantic segmentation." arXiv preprint arXiv:1912.10917 (2019).

Li, Yanwei, et al. "Learning dynamic routing for semantic segmentation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

Liu, Chenxi, et al. "Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

Hümmer, Christoph, et al. "VLTSeg: Simple Transfer of CLIP-Based Vision-Language Representationsfor Domain Generalized Semantic Segmentation." arXiv preprint arXiv:2312.02021 (2023).

ABSTRACT

Built-in Generative Data Augmentation via Cross-Domain Feature Disentanglement and Synthesis for Enhancing Rainy Image Segmentation

Yong, Yun-Jeong

Major in Applied Artificial Intelligence

Dept. of Applied Artificial Intelligence

The Graduate School

Hansung University

Rain is a common weather condition that significantly degrades image quality, making autonomous driving challenging under adverse weather conditions. While semantic segmentation models have made significant advancements under clear weather conditions, their performance drastically drops when image quality deteriorates in rainy conditions. To maintain segmentation performance in rainy conditions, we attempted to fine-tune a segmentation model using a dataset comprising paired images of rainy and clear conditions. However, we encountered the issue of missing segmentation ground truth (GT) for such paired

images. Additionally, previous studies have approached this problem by deraining the images first and then using another segmentation model for inference, but this method has shown suboptimal performance.

Existing datasets containing paired rainy and clear images are typically from non-driving domains and lack segmentation GT, while segmentation datasets do not include rainy images. To overcome this challenge, our study proposes generating synthetic rainy images that still enable effective segmentation. We argue that it is essential to generate rainy images where segmentation can be accurately performed.

DSANet can augment rainy images by disentangling and synthesizing background and rain features, thereby better preserving the background and facilitating the application of various rain effects. In this study, we used DSANet as a rainy image augmentator to generate rainy images from autonomous driving videos and fine-tuned the segmentation model using these images. As a result, our model significantly outperformed existing segmentation models and demonstrated superior performance compared to other models that applied a pretrained segmentation model after deraining the images using conventional deraining models.

【Key words】 Augmentator, Disentangling, Segmentation, Synthesizing, Self-driving Car