사물인터넷융복합논문지 Vol. 9, No. 1, pp. 97-104, 2023 https://doi.org/10.20465/KIOTS.2023.9.1.097

ESRGAN과 Semantic Soft Segmentation을 이용한 객체 분할

윤동식[1], 곽노윤[2*]
[1]고려대학교 전기전자공학과 석사과정, [2]백석대학교 컴퓨터공학부 교수

Object Segmentation Using ESRGAN and Semantic Soft Segmentation

Dongsik Yoon[1], Noyoon Kwak[2*]
[1]Graduate Student, Department of Electronics and Electrical Engineering, Korea University
[2]Professor, Division of Computer Engineering, Baekseok University

요 약 본 논문은 ESRGAN(Enhanced Super Resolution GAN)과 SSS(Semantic Soft Segmentation)을 이용한 객체 분할에 관한 것이다. 본 논문의 연구진이 앞서 제안한 Mask R-CNN과 SSS를 이용한 객체 분할 방법의 분할 성능은 전반적으로 양호하지만 객체의 크기가 상대적으로 작은 경우 분할 성능이 저조해지는 문제점이 있었다. 본 논문은 이러한 문제점을 해소하기 위한 것이다. 제안된 방법은 Mask R-CNN을 통해 검출된 객체의 크기가 일정 기준치 이하인 경우, ESRGAN을 통해 초해상화를 수행한 후, SSS을 수행함으로써 소형 객체의 분할 성능을 개선하고자 한다. 제안된 방법에 따르면, 기존의 방법에 비해 크기가 작은 객체의 분할 특성을 좀 더 효과적으로 개선할 수 있음을 확인할 수 있었다.

주제어 : ESRGAN, Mask R-CNN, Semantic Soft Segmentation, 초해상화, 영상 분할

Abstract This paper is related to object segmentation using ESRGAN(Enhanced Super Resolution GAN) and SSS(Semantic Soft Segmentation). The segmentation performance of the object segmentation method using Mask R-CNN and SSS proposed by the research team in this paper is generally good, but the segmentation performance is poor when the size of the objects is relatively small. This paper is to solve these problems. The proposed method aims to improve segmentation performance of small objects by performing super-resolution through ESRGAN and then performing SSS when the size of an object detected through Mask R-CNN is below a certain threshold. According to the proposed method, it was confirmed that the segmentation characteristics of small-sized objects can be improved more effectively than the previous method.

Key Words : ESRGAN, Mask R-CNN, Semantic Soft Segmentation, Super-resolution, Image Segmentation

본 논문은 2022년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 연구과제(2021RIS-004)로 수행되었음.
*교신저자 : 곽노윤(nykwak@bu.ac.kr)
접수일 2022년 12월 4일 수정일 2023년 1월 5일 심사완료일 2023년 1월 8일

1. 서론

널리 알려진 Mask R-CNN[1]은 객체 검출(Object Detection)과 인스턴스 분할(Instance Segmentation) 두 측면에서 유용한 모델로 의료분야나 모바일·임베디드 환경, 데이터셋 생성 등 인스턴스 분할을 활용할 수 있는 여러 분야에서 사용된다. 기존의 Mask R-CNN은 세밀하지 못한 마스크 데이터셋과 훈련과정의 오차로 인해 객체를 검출하는 성능은 우수하지만 객체를 분할하는 성능은 기대에 미치지 못하는 경우가 빈도 높게 발생한다. 이러한 문제점을 보완하기 위해 본 논문의 연구진은 Mask R-CNN[1]과 SSS(Semantic Soft Segmentation)[2]를 사용한 객체 분할 방법[3]을 앞서 제안한 바 있다. 여기서 채택하고 있는 SSS[2]는 ResNet[4]을 활용하여 추출한 특징 벡터 이미지와 Spectral Matting[5]에 새로운 항을 추가하여 좀 더 부드러운 의미론적 분할을 수행하는 것이 특징이다. 본 연구진이 앞서 제안한 객체 분할 방법[3]은 Mask R-CNN으로 미진하게 분할된 객체를 SSS을 사용하여 분할 결과를 보정한 것이다. 하지만 이 방법은 Mask R-CNN을 통해 검출된 객체의 크기가 일정 이하인 경우에 SSS의 성능이 미진해 오히려 분할 성능이 이전보다 저하되는 문제점이 있었다.

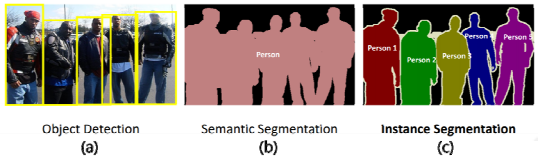
본 논문은 이러한 문제를 해결하기 위해 Mask R-CNN을 통해 검출한 객체의 크기가 일정 이하인 경우 ESRGAN(Enhanced Super Resolution GAN)[6]을 사용하여 초해상화(super-resolution)를 수행한 후, SSS을 수행함으로써 상대적으로 크기가 작은 객체의 분할 성능을 개선함에 목적이 있다. 최종적으로 Mask R-CNN을 통해 검출한 객체의 크기가 일정 이하인 경우에도 우수한 분할 성능을 제공하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 살펴보고, 3장에선 제안된 객체 분할 방법에 대해 설명한다. 4장에서는 시뮬레이션 결과 및 고찰을 서술하고 마지막으로 5장에서는 결론을 맺는다.

2. 관련 연구

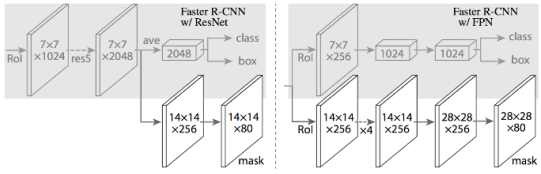
2.1 Mask R-CNN을 이용한 객체 검출 및 분할

딥러닝 기술의 급속한 발전에 기인해 그간 객체 검출과 관련된 주목할만한 연구들[1, 7, 8, 9, 10, 11, 12]이 다수 진행되었다. 본 논문에서 객체 검출(object detection) 및 인스턴스 분할(instance segmentation)을 위해 채택한 Mask R-CNN[1]은 Faster R-CNN[13]에서 가능한 경계박스 단위의 분할과 FCN(Fully Convolutional Network)[14]에서 가능한 의미론적 분할(semantic segmentation)을 결합한 것으로, 경계박스 단위로 전경 객체를 분할하는 모델이다. Faster R-CNN이 객체 검출만을 위한 모델이라면, Mask R-CNN은 이를 확장하여 검출된 경계박스 내 마스크를 학습하는 모델이다. 이 모델은 객체를 분할하기 위해 물체를 분류하는 브랜치와 경계박스를 회귀(regression)하는 브랜치에 추가로 객체 마스크를 예측하는 브랜치를 부가한 것이다. 즉 Mask R-CNN은 Faster R-CNN에 각 픽셀이 객체에 해당하는지 아닌지를 예측하는 추가적인 마스크 브랜치를 삽입해 경계박스 내 각각의 픽셀이 그 객체의 일부인지를 판별하는 이진 마스크를 구한다.



[Fig. 1] (a) Object detection in Faster R-CNN, (b) Semantic segmentation of FCN, and (c) Object segmentation in Mask R-CNN[1, 13, 14].

통상, Mask R-CNN에서는 정확한 픽셀 위치가 필요하기 때문에 양성형 보간(bilinear interpolation)을 사용한 RoIAlign을 추가적으로 채택함으로써 오정합 현상을 방지하고 각 특징맵과 원본 이미지의 해당 영역이 더 잘 정합되도록 한다.



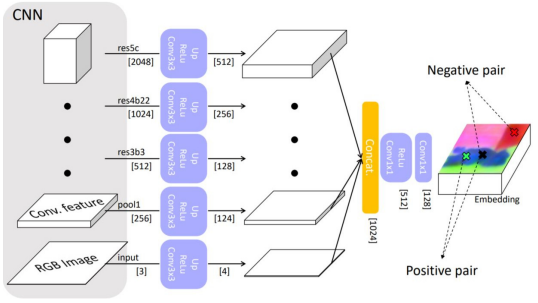
[Fig. 2] The architecture composed by adding mask branch to Faster R-CNN(ResNet/FPN)

2.2 Semantic Soft Segmentation

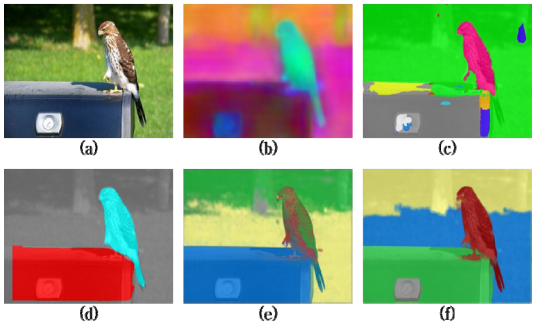
다수의 객체 분할 방법들[3, 15, 16, 17]에서 의미론적 분할을 위해 채택한 SSS는 [Fig. 3]과 같이 DeepLab ResNet-101 기반[18]의 네트워크를 사용해 128차원의 특징 벡터를 추출한다. ResNet-101의 'input', 'pool1', 'res3b3', 'res4b22', 'res5c'층에서 출력된 특징맵들을

대상으로 3×3 합성곱을 취한 후, 양선형 보간법을 사용하여 원본 이미지의 해상도로 업샘플링하고 이를 모두 결합하여 ReLU에 통과시킨다. 이때 중간 특징의 차원은 {3, 256, 512, 1024, 2048} 차원인데, 이를 {4, 124, 128, 256, 512} 차원으로 각각 압축하여 총 1024의 차원이 되게 한다. 그 후 두 번의 1×1 합성곱 층을 통해 1024의 특징 차원을 512차원, 128차원으로 단계적으로 감축한다[2]. 이 네트워크의 훈련에는 COCO Stuff 데이터셋이 사용된다.

그래프 이론 기반의 라플라시안 행렬(Laplacian matrix)을 사용하는 Spectral Matting[5]는 작은 패치들의 저수준 통계 특성(low-level statistics)만을 다루기 때문에 객체를 판별하는 능력이 제한되는 단점이 있다. 이러한 단점을 보완하기 위해 SSS은 라플라시안 행렬에 의미론적 특징을 혼합하고 장면 객체(scene objects)와 같은 고수준 개념(high-level concepts)을 추출해 이미지를 폭넓은 시점에서 판별하게 한다.



[Fig. 3] Feature extraction network architecture of SSS



[Fig. 4] (a) Original Image, (b) Feature vector extracted from 128 dimension using PCA, (c) Segmentation result using PSPNet, (d) Segmentation result using Mask R-CNN, (e) Segmentation result using Spectral Matting, (f) Segmentation result using SSS[1, 3, 5, 20]

SSS는 색상 기반의 원거리 상호작용을 하는 저수준 유사도항을 비국부 색상 유사도(nonlocal color affinity)로 정의하였다. 이는 과분할된 이미지를 기반으로 가이디드 샘플링(guided sampling)을 진행하고 이미지 크기의 20% 이내에 해당하는 반지름을 가진 2,500개의 슈퍼 픽셀[19]을 생성하여 각 슈퍼 픽셀들 간의 유사도를 측정한다. 이 방법은 특징을 대표하기 충분한 각 슈퍼 픽셀들을 하나의 샘플로 사용하며, 큰 반경을 사용하기 때문에 근접하지 않은 영역끼리의 연결도 가능하게 하는 장점이 있다. 추가적인 정보 없이 비국부 색상 유사도로 확장된 공간상에서 상호작용을 통한 이미지 분할을 진행할 시, 색상이 유사한 다른 객체들을 하나로 합치는 문제가 발생할 수 있다. SSS는 의미론적으로 유사한 영역들이 하나의 덩어리로 분할되도록 세분화하기 위해 의미론적 유사도항을 추가하였다. 이는 같은 장면 객체들의 픽셀들은 서로 같이 분류되도록 도와주면서 서로 다른 객체끼리는 분리되도록 도와준다. 이상의 설명을 정리하자면, SSS는 DeepLab ResNet-101 기반의 네트워크를 사용하여 나온 128차원의 벡터를 PCA을 통해 3차원으로 줄이고, 기존의 Spectral Matting 연산 시에 사용하는 라플라시안 행렬에 새로 만든 비국부 색상 유사도와 의미론적 유사도를 추가해 의미론적 분할을 진행하여 분할 성능을 제고할 수 있다.

2.3 ESRGAN(Enhanced Super Resolution GAN)

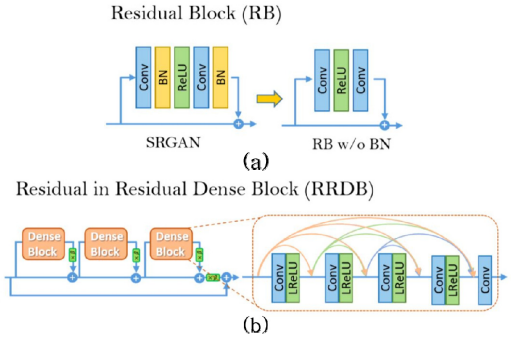
기존의 SRGAN(Super Resolution GAN)[21]은 한 장의 이미지만으로 초해상화를 수행해 사실적인 질감을 생성한다. 하지만 부분적으로 시각적 거부감이 초래되는 바, ESRGAN[6]은 세 가지 핵심 요소를 추가해 화질을 개선한다.

2.3.1 Residual in Residual Dense Block(RRDB)

본 논문에서 이미지 초해상화를 위해 채택한 기존의 ESRGAN은 화질을 향상시키기 위해 SRGAN의 생성자 구조에서 다음과 같은 두 가지 보완 사항을 포함하고 있다. 하나는 모든 BN(Batch Normalization)층을 삭제한 것이고, 다른 하나는 [Fig. 5]와 같이 제안된 RRDB (Residual-in-Residual Dense Block)을 사용하는 것이다. BN층은 배치의 평균과 분산을 사용해 특징들을 정규화하여 훈련하고, 테스트 시엔 전체 데이터셋의 평균과 분산을 사용하기 때문에 훈련과 테스트 데이터셋의 통계가 크게 다른 경우, 시각적 거부감을 야기하고 일반

화하는 성능을 저하시키는 경향이 있다. ESRGAN은 BN 층을 제거함으로써 훈련을 안정화시키면서 일반화 성능을 높이고 계산 복잡도를 줄인 것이다.

ESRGAN의 고차원 네트워크 구조는 SRGAN의 구조를 유지하면서 [Fig. 5(b)]와 같은 RRDB라는 새로운 기본 블록을 사용한다. 더 많은 층과 연결이 있는 경우 성능이 향상된다는 Residual Networks[22]를 기반으로 제안한 RRDB는 기존의 Residual Block보다 더 깊고 복잡한 구조를 가진다.



[Fig. 5] (a) The structure in which the BN layer is deleted from the residual block of SRGAN, (b) Dense block structure of the proposed RRDB

이로 인해 ESRGAN은 RRDB 기반의 매우 깊은 네트워크를 학습시킬 수 있도록 두 가지 추가적인 기술을 사용한다. 그 중 하나는 Residual Scaling으로, 학습의 불안정을 방지하기 위해 연산된 블록을 주경로(main path)와 결합하기 전에 0에서 1사이의 상수를 곱해서 블록을 축소하는 것이다. 또 다른 하나는 Residual 네트워크의 구조가 시작 파라미터의 분산이 작은 경우에 학습이 더 잘되는 것을 이용해 작은 초기값을 사용하는 것이다.

2.3.2 상대 판별자(Relativistic Discriminator)

ESRGAN은 생성자의 구조를 향상 시키는 것뿐만 아니라 Relativistic GAN[23]을 기반으로 판별자도 강화한다. 기존의 SRGAN의 표준 판별자는 입력 이미지 x가 얼마나 진짜 같은지에 대한 가능성을 판별하는 것이었다면, ESRGAN의 판별자는 진짜 이미지 x_r이 생성한 가짜 이미지 x_f보다 얼마나 더 진짜 같은지를 예측하는 상대적 평균 판별자(relativistic average discriminator)를 사용한다. SRGAN의 표준 판별자는 $D(x)=\sigma(C(x))$의 식으로 나타낼 수 있다. 이때 σ는 시그모이드 함수이고 $C(x)$는 아직 변환되지 않은 판별자의 출력을 의미한다. ESRGAN에서 사용한 상대 평균 판별자는 식 (1)과 같다. 여기서 $\mathbb{E}_{x_f}[C(x_f)]$는 미니 배치에 있는 모든 가짜 데이터의 평균을 의미한다. 이를 사용한 판별자의 손실함수는 식 (2)와 같으며, 생성자의 손실함수는 식 (3)과 같이 판별자의 손실함수와 대칭적인 형태를 보인다.

$$D_{Ra}(x_r,x_f)=\sigma(C(x_r)-\mathbb{E}_{x_f}[C(x_f)]) \quad (1)$$

$$L_D^{Ra}=-\mathbb{E}_{x_r}[\log(D_{Ra}(x_r,x_f))]-\mathbb{E}_{x_f}[\log(1-D_{Ra}(x_f,x_r))] \quad (2)$$

$$L_G^{Ra}=-\mathbb{E}_{x_r}[\log(1-D_{Ra}(x_r,x_f))]-\mathbb{E}_{x_f}[\log(D_{Ra}(x_f,x_r))] \quad (3)$$

식 (2)와 식 (3)에서 $x_f=G(x_i)$이며 x_i는 입력 LR(Low Resolution) 이미지를 의미한다. 식 (3)에서와 같이 생성자의 손실함수는 x_r와 x_f 모두를 포함하기 때문에 SRGAN과 달리 생성된 이미지와 원본 이미지 모두 경쟁적 학습에 영향을 미치는 것을 볼 수 있다.

2.3.3 지각 손실(perceptual loss)

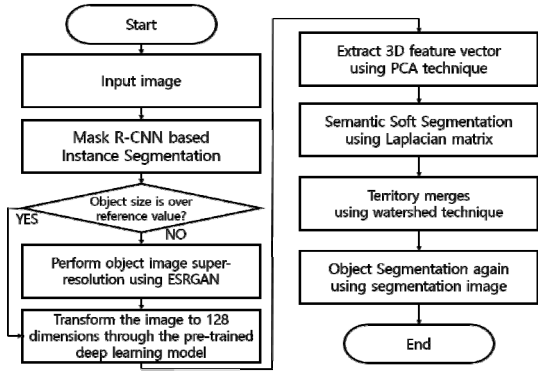
ESRGAN은 좀 더 효과적인 지각 손실을 이용하는데, 보통의 지각 손실은 사전 학습된 네트워크의 활성화 층을 통과한 특징을 최소화하는 것으로 정의한다. 이와 달리 ESRGAN에서는 활성화 층을 통과하기 전의 특징을 사용하는 것으로 두 가지 문제를 개선한다. 네트워크가 매우 깊은 경우 활성화되는 특징들이 매우 드물게 나타나기 때문에 좋지 못한 결과를 초래하는 문제와 활성화 층을 통과한 특징을 사용하는 경우 원본 이미지와 비교하여 재구성된 밝기가 일관되지 않은 문제이다. 결론적으로 ESRGAN의 생성자 손실 함수는 식 (4)과 같이 정의할 수 있다.

$$L_G=L_{percep}+\lambda L_G^{Ra}+\eta L_1 \quad (4)$$

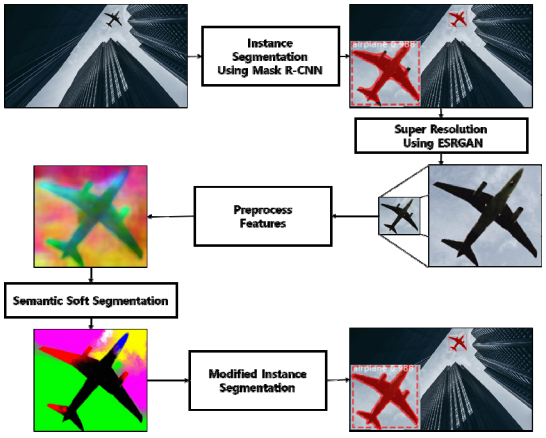
여기서, $L_1=\mathbb{E}_{x_i}\|G(x_i)-y\|$은 콘텐트 손실로, 생성된 이미지 $G(x_i)$와 원본 이미지 y의 차이이고, L_{percep}은 생성 이미지와 원본 이미지의 네트워크를 통과시킨 손실이다. λ와 η는 다른 손실과의 균형을 맞추기 위한 계수이다.

3. 제안된 객체 분할 방법

본 논문은 Mask R-CNN으로 객체를 검출하고 SSS을 통해 의미론적 분할을 수행한다. [Fig. 6]은 제안된 방법의 순서도를 나타낸 것이고, [Fig. 7]은 제안된 방법의 개념 블록도를 나타낸 것이다.



[Fig. 6] Flowchart of the proposed method



[Fig. 7] Concept block diagram of the proposed method

이때 Mask R-CNN을 통해 검출된 객체의 크기가 일정 기준치 이하라면 ESRGAN을 활용해 해상도를 높인 이미지를 생성한 후, SSS을 수행하고, 그렇지 않은 경우에는 곧바로 SSS을 통해 의미론적 분할을 수행한다. 이후 검출 객체와 분할 결과를 대응시켜 최종적으로 객체를 재분할한다. 결과적으로 Mask R-CNN으로 검출한 객체의 크기가 작은 경우에 미진했던 SSS의 분할 특성을 개선해 분할 성능을 향상시키는 효과가 있다.

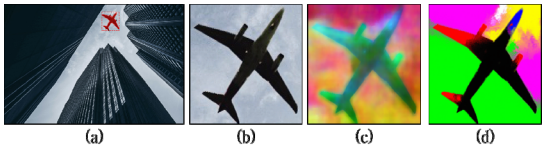
3.1 객체 검출 및 초해상화

3.1.1 객체 검출 및 분할

Mask R-CNN 모델을 활용하여 입력 이미지에서 객체를 검출하고 분할한다. Mask R-CNN 모델은 [Fig. 8(a)]와 같이 각 객체의 경계박스 정보와 객체 분할된 마스크 정보, 객체 종류(class_ids) 및 정확도를 반환한다.

3.1.2 초해상화

Mask R-CNN을 활용하여 검출된 객체 중에서 크기가 일정 기준 이하인 경우, ESRGAN을 사용해 [Fig. 8(b)]와 같이 소형 객체 이미지를 대상으로 4배 초해상화를 수행한다. 그렇지 않은 경우, 해당 객체는 초해상화를 수행하지 않는다.



[Fig. 8] (a) Image with object detection completed, (b) Super-resolution of objects based on ESRGAN, (c) Image extracted feature vector using PCA, (d) Image with SSS applied

3.2 SSS(Semantic Soft Segmentation)

3.2.1 특징 벡터 추출

ResNet-101 기반의 사전 훈련된 네트워크를 사용하여 128차원의 벡터를 추출한다. 그 후 PCA을 통해 [Fig. 8(c)]와 같이 3차원의 특징 벡터를 추출한다.

3.2.2 Semantic Soft Segmentation 수행

PCA를 통해 3차원으로 만든 특징 벡터와 원본 이미지를 사용해 [Fig. 8(d)]와 같이 SSS을 수행한다.

3.3 워터쉐드를 이용한 영역 분할

워터쉐드 분할(watershed segmentation)을 통해 의미론적 분할 결과의 보정 및 영역 병합을 수행하기 위해서는 우선, [Fig. 9(a)] 및 [Fig. 9(b)]과 같이 각각 흰색으로 나타낸 확실한 배경 영역 및 확실한 전경 영역에 각각의 배경 마커와 전경 마커를 부여한다. 이때 기존의 Mask R-CNN의 마스크와 SSS의 분할 결과를 픽셀 단위로 대조하여 전경 객체에 해당하는 분할 영역의 색상을 구한다. 예컨대, [Fig. 8(a)]의 마스크 영역과 [Fig. 8(d)]를 대조하여 전경 객체에 해당하는 색상인 빨강, 검

정, 파랑 색상을 얻는다. 이후 SSS의 분할 결과를 기반으로 완벽하게 의미론적으로 구분된 픽셀만을 선별해 앞서 구한 분할 영역 색상에 해당하는 픽셀은 전경으로, 분할 영역 색상에 해당하지 않는 픽셀은 배경으로 설정한다. 특징 벡터인 [Fig. 8(c)]와 마커로 설정한 [Fig. 9(c)]를 참조하여 [Fig. 9(d)]와 같은 워터쉐드 분할을 수행한다.

[Fig. 9] (a) Definite background area(white), (b) Definite foreground area(white), (c) Images with markers set on the foreground and background for which the semantic segmentation result is certain, (d) Image segmented by watershed with Fig 8(c) and Fig 9(c)

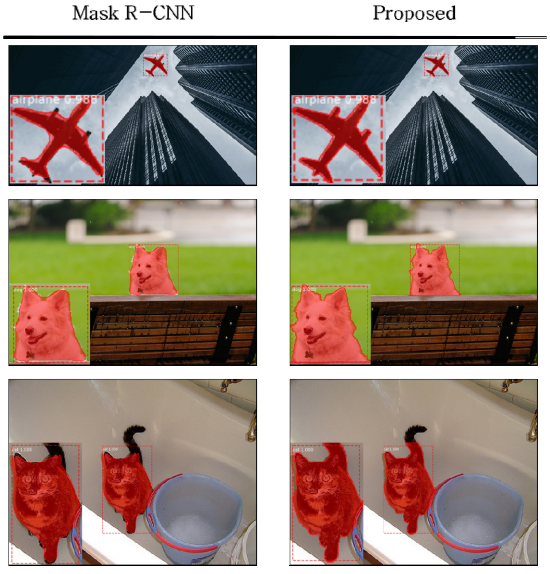
3.4 객체 분할

3.3절의 워터쉐드 기법으로 분할된 여러 마커들 중 의미론적 분할 결과를 기반으로 검출된 객체에 해당하는 마커들을 찾는다. 이때 각각의 마커가 전경 객체의 분할 영역 색상을 포함하고 있는 비율을 계산하여 전경 객체에 해당하는 마커를 찾는다. 그 후 객체에 해당되는 마커들을 Mask R-CNN의 마스크와 교체하여 객체 분할을 완료함으로써 오분할 억제할 수 있다.

4. 시뮬레이션 결과 및 고찰

본 논문에서는 Jupyter Notebook과 NVIDIA GTX 1070Ti GPU 환경에서 사전 학습된 Mask R-CNN, SSS 및 ESRGAN 모델을 사용하여 컴퓨터 시뮬레이션을 수행하였다.

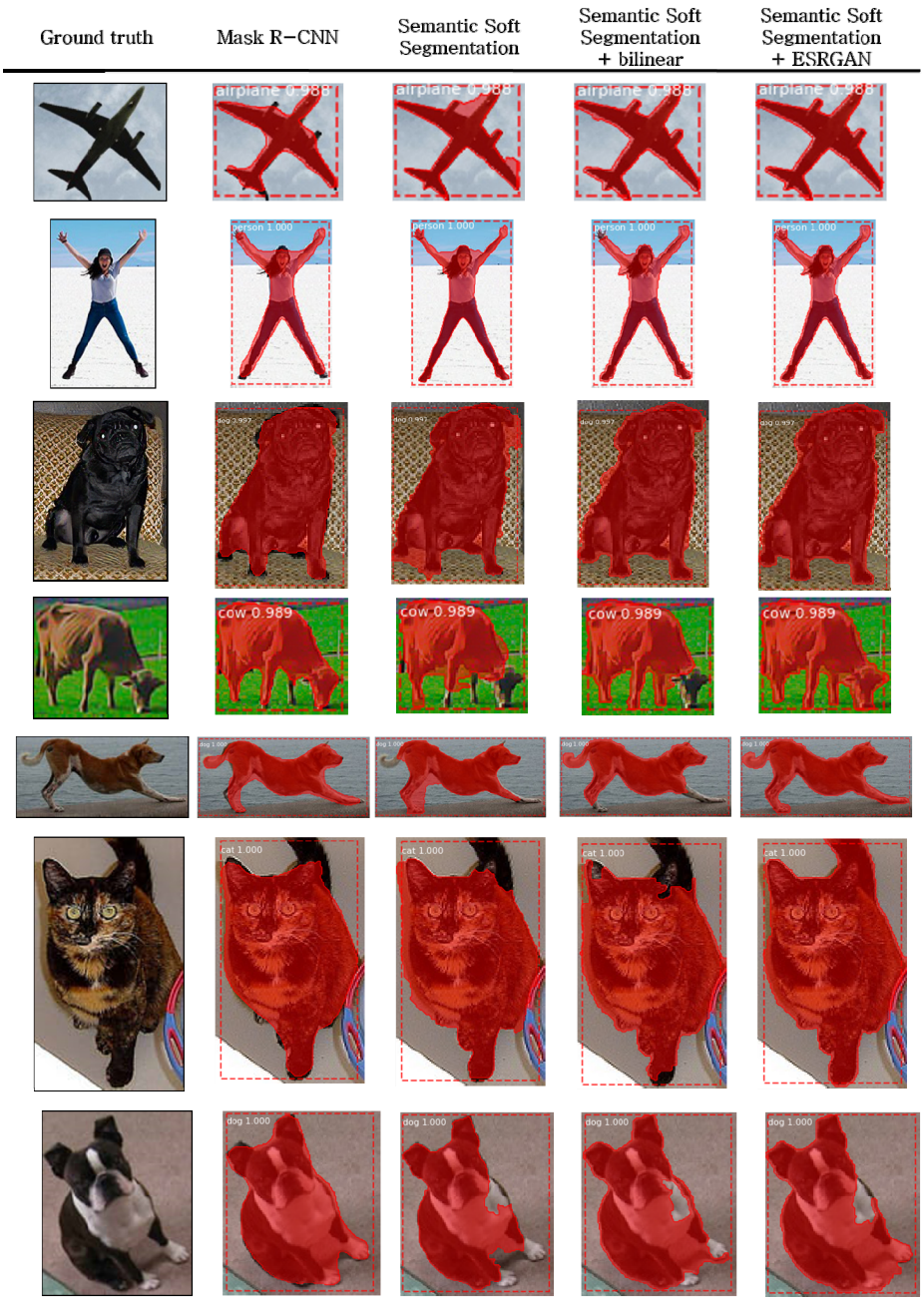
[Fig. 10]은 Mask R-CNN과 제안된 방법의 분할 결과를 비교한 것으로, 전경과 배경 구분이 뚜렷함에도 불구하고 Mask R-CNN은 객체 분할 결과가 만족스럽지 못한 반면, 제안된 분할 방법은 상대적으로 우수한 객체 분할 성능을 제공함을 시각적으로 확인할 수 있었다. [Fig. 11]은 제안된 객체 분할 방법의 각종 테스트 영상들에 대한 시뮬레이션 결과를 나타낸 것이다. [Fig. 11]에서 볼 수 있듯이 검출된 객체의 크기가 작은 경우, 초해상화를 수행한 후, 의미론적 분할을 진행하는 것이 대체적으로 그 성능이 우수함을 확인할 수 있다. 특히, 양선형 보간법보다 ESRGAN을 사용한 경우의 의미론적 분할 결과가 더 우수한 것을 확인할 수 있다. 하지만 드물게 [Fig. 11]의 마지막 강아지 사진과 같이 전경 객체와 배경의 색상이 유사하여 SSS의 성능이 객체의 크기와 상관없이 미진한 경우, 제안된 객체 분할 결과는 Mask R-CNN의 분할 결과보다 나빠지는 경우가 발생하기도 한다.



[Fig. 10] Simulation results using Mask R-CNN and the proposed method

5. 결론

제안된 방법은 Mask R-CNN과 SSS을 이용한 객체 분할 시 검출된 객체의 크기가 작은 경우, ESRGAN을 이용한 초해상화를 통해해 SSS의 의미론적 분할 성능을 개선한 것이다. 제안된 방법은 검출된 객체의 크기가 작더라도 우수한 분할 성능을 제공함을 확인할 수 있었다. 하지만 Mask R-CNN의 분할 결과를 보정하는 일련의 과정을 거치기 때문에 기존의 객체 검출보다 시간이 오래 걸리는 단점이 있다. 따라서 제안된 방법을 직접 실시간 처리에 사용하기보다는 실시간처리 모델에서 사용하는 학습 데이터셋을 만들거나 고품질의 영상 편집 및 객체 인식 등에 활용 가능할 것으로 기대된다. 예컨대, 객체 분류 모델을 학습시키기 위해선 사람 손으로 일일이 레이블링을 수행하여 COCO와 같은 데이터셋들을 생성하는데, 제안된 방법을 활용해 수작업을 최소화할 수 있도록 돕는 반자동 레이블러를 개발한다면, 데이터셋 생성 과정에서 비용과 시간을 절감할 뿐만 아니라 좀 더 정밀한 데이터셋을 통해 학습 모델의 성능 향상을 도모할 수 있을 것이다.



[Fig. 11] Simulation results using multiple object segmentation methods

REFERENCES

[1] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," Proceedings of IEEE International Conference on Computer Vision, pp.2980-2988. 2017.

[2] Y. Aksoy, T.-H. Oh, S. Paris, M. Pollefeys, and W. Matusik, "Semantic Soft Segmentation," ACM Trans. Graph., 2018.

[3] D. Yoon and N. Kwak, "Object Segmentation Using MaskR-CNN and Semantic Soft Segmentation," Proceedings of 2020 Winter Conference of Korean Society of Communications and Communications, pp.872-873, Feb. 2020.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Proceedings of Conference on Computer Vision and Pattern Recognition, pp.770-778, 2016.

[5] A. Levin, A. Rav-Acha, and D. Lischinski, "Spectral Matting," IEEE Trans. Pattern Anal. Mach. Intell., Vol.30, No.10, pp.1699-1712, Oct. 2008.

[6] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. -C Loy, Y. Qiao, and X. Tang, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," Proceedings of European Conference on Computer Vision, 2018.

[7] Z. Li, C. Peng, G. Yu, X. Zhang, Y. Deng, and J. Sun, "Light-head R-CNN: In Defense of Two-stage Object Detector," arXiv preprint arXiv:1711.07264, 2017.

[8] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into High Quality Object Detection," Proceedings of Conference on Computer Vision and Pattern Recognition, 2018.

[9] P. Purkait, C. Zhao, and C. Zach, "SPP-Net: Deep Absolute Pose Regression with Synthetic Views," arXiv preprint arXiv:1712.03452, 2017.

[10] Z. Li and F. Zhou, "FSSD: Feature Fusion Single Shot Multibox Detector," arXiv preprint arXiv:1712.00960, 2017.

[11] H. Law and J. Deng, "CornerNet: Detecting Objects as Paired Keypoints," Proceedings of European Conference on Computer Vision, 2018.

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-time Object Detection," Proceedings of Conference on Computer Vision and Pattern Recognition, 2016.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks," Proceedings of Conference on Neural Information Processing Systems, 2015.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," Proceedings of Conference on Computer Vision and Pattern Recognition, 2015.

[15] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A Deep Neural Network Architecture for Real-time Semantic Segmentation," arXiv preprint arXiv:1606.02147, 2016.

[16] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation," 2017 IEEE Visual Communications and Image Processing, 2017.

[17] E. Romera, J. M. Álvarez, L. M. Bergasa, and R. Arroyo, "ERFNet: Efficient Residual Factorized Convnet for Real-time Semantic Segmentation," IEEE Transactions on Intelligent Transportation Systems, Vol.19, pp.263-272, 2017.

[18] L. -C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," IEEE Trans. Pattern Anal. Mach. Intell., Vol.40, 2017.

[19] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," IEEE Trans. Pattern Anal. Mach. Intell., Vol.34, No.11, pp.2274-2281, 2012.

[20] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," Proceedings of Conference on Computer Vision and Pattern Recognition, 2017.

[21] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic Single Image Super-resolution Using a Generative Adversarial Network," Proceedings of Conference on Computer Vision and Pattern Recognition, pp.4681-4690, 2017.

[22] B. Lim, S. Son, H. Kim, S. Nah, and K. Lee, "Enhanced Deep Residual Networks for Single Image Super-resolution," Proceedings of Conference on Computer Vision and Pattern Recognition, 2017.

[23] A. Jolicoeur-Martineau, "The Relativistic Discriminator: a Key Element Missing from Standard GAN," arXiv preprint arXiv:1807.00734, 2018.

윤 동 식(Dongsik Yoon) [준회원]

- 2021년 2월 : 백석대학교 ICT학부 (공학사)
- 2023년 2월 : 고려대학교 대학원 전기전자공학과 (공학석사)

〈관심분야〉

딥러닝 기반 영상처리 및 컴퓨터비전, 딥러닝 기반 객체 분할, 초해상화 및 인페이팅, 스타일 트랜스퍼

곽 노 윤(Noyoon Kwak) [종신회원]

- 1994년 2월 : 한국항공대학교 항공전자공학과 (공학사)
- 1996년 2월 : 한국항공대학교 대학원 항공전자공학과 (공학석사)
- 2000년 2월 : 한국항공대학교 대학원 항공전자공학과 (공학박사)
- 2000년 3월 ~ 현재 : 백석대학교 컴퓨터공학부 교수

〈관심분야〉

딥러닝 기반 영상처리 및 컴퓨터비전, 얼굴 및 시선 인식, 객체 트래킹, 3D 재구성, 제스처 기반 UI/UX, 인공지능