https://doi.org/10.5392/JKCA.2021.21.03.023

의미론적 영상 분할의 정확도 향상을 위한 에지 정보 기반 후처리 방법

Post-processing Algorithm Based on Edge Information to Improve the Accuracy of Semantic Image Segmentation

김정환, 김선혁, 김주희, 최형일
숭실대학교 미디어학과

Jung-Hwan Kim(96junghwan@naver.com), Seon-Hyeok Kim(k_sh153@naver.com), Joo-heui Kim(createtime2@naver.com), Hyung-Il Choi(hic@ssu.ac.kr)

요약

컴퓨터 비전 분야의 의미론적 영상 분할(Semantic Image Segmentation) 기술은 이미지를 픽셀 단위로 분할 하여 클래스를 나누는 기술이다. 이 기술도 기계 학습을 이용한 방법으로 성능이 빠르게 향상되는 중이며, 픽셀 단위의 정보를 활용할 수 있는 높은 활용성이 주목받는 기술이다. 그러나 이 기술은 초기부터 최근까지도 계속 '세밀하지 못한 분할'에 대한 문제가 제기되어 왔다. 이 문제는 레이블 맵의 크기를 계속 늘리면서 발생한 문제이기 때문에, 자세한 에지 정보가 있는 원본 영상의 에지 맵을 이용해 레이블 맵을 수정하여 개선할 수 있을 것으로 예상할 수 있었다. 따라서 본 논문은 기존 방법대로 학습 기반의 의미론적 영상 분할을 유지하되, 그 결과인 레이블 맵을 원본 영상의 에지 맵 기반으로 수정하는 후처리 알고리즘을 제안한다. 기존의 방법에 알고리즘의 적용 한 뒤 전후의 정확도를 비교했을 때 평균적으로 약 1.74% 픽셀 정확도와 1.35%의 IoU(Intersection of Union) 정확도가 향상되었으며, 결과를 분석했을 때 성공적으로 본래 목표한 세밀한 분할 기능을 개선했음을 보였다.

■ 중심어 : | 컴퓨터비전 | 머신러닝 | 딥러닝 | 영상처리 | 의미론적 분할 |

Abstract

Semantic image segmentation technology in the field of computer vision is a technology that classifies an image by dividing it into pixels. This technique is also rapidly improving performance using a machine learning method, and a high possibility of utilizing information in units of pixels is drawing attention. However, this technology has been raised from the early days until recently for 'lack of detailed segmentation' problem. Since this problem was caused by increasing the size of the label map, it was expected that the label map could be improved by using the edge map of the original image with detailed edge information. Therefore, in this paper, we propose a post-processing algorithm that maintains semantic image segmentation based on learning, but modifies the resulting label map based on the edge map of the original image. After applying the algorithm to the existing method, when comparing similar applications before and after, approximately 1.74% pixels and 1.35% IoU (Intersection of Union) were applied, and when analyzing the results, the precise targeting fine segmentation function was improved.

■ keyword : | Computer Vision | Machine Learning | Deep Learning | Image Processing | Semantic Segmentation |

접수일자 : 2020년 11월 19일
수정일자 : 2021년 01월 20일
심사완료일 : 2021년 01월 20일
교신저자 : 최형일, e-mail : hic@ssu.ac.kr

I. 서 론

1. 연구 배경 및 목적

AI가 급부상하면서 IT 분야에 국한되지 않고 현재 모든 기술 분야에서 빠르게 기존 알고리즘을 대체하거나 병행함으로써 성능을 향상하고 있다. 이에 따라 기존에는 구현이 완벽하지 않거나 불가능했던 AR(Augmented Reality), VR(Virtual Reality)과 같은 기술과 서비스들이 점차 구현되고 계속 보완되는 추세이다. 컴퓨터 비전 분야에서도 발전되고 있는 머신러닝 기법을 이용해 기존 알고리즘과 병행하거나 완전히 대체하면서 날이 갈수록 뛰어난 효율과 성능 향상을 보인다. 컴퓨터비전의 의미론적 영상 분할은 [그림 1]과 같이 이미지를 픽셀 단위로 분할 하여 벽, 침대, 이불과 같은 클래스로 나누는 기술로, 각 픽셀에 레이블 번호를 할당하여 레이블 맵을 결과로 얻는 구조이다. 이 기술은 픽셀 단위의 정보를 활용할 수 있는 높은 활용성이 주목을 받고 있으며, 자율주행 자동차, AR, 의료 등 이미지를 이용하는 모든 분야에서 더 세밀한 자료로써 활용될 수 있다.

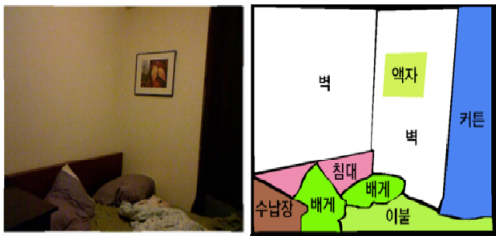


그림 1. 의미론적 영상 분할 예시

기존 의미론적 영상 분할 방법들은 [그림 1]과 같이 에지 부분의 영상 분할이 제대로 수행되지 않는 '세밀하지 못한 분할' 문제가 계속 제기되었다. 딥러닝을 본격적으로 활용한 초기의 의미론적 영상 분할 연구[1]에서 나온 문제점이 성능을 상승시켜도 계속 나타난 것이다. 실제 의미론적 영상 분할이 된 이미지를 살펴보면, 초창기의 영상 분할과 최근의 성능이 좋았던 방법들[1-7]에서 모두 에지 부분의 분할이 정확하지 않은 현상을 발견할 수 있었다. 이 문제점은 의미론적 영상 분할의 결과물인 작은 레이블 맵을 원본 영상과 같은 크기까지 늘리면서 발생한 문제로 볼 수 있다. 따라서 이 문제점을 해결하기 위해 자세한 에지 정보가 있는 에지 맵을 기반으로 레이블 맵을 수정할 수 있을 것이 예상된다. 학습 기반의 의미론적 영상 분할을 위해 에지 맵을 참고한 다른 사례가 있었는데, 해당 연구에서는 후처리 알고리즘으로 에지 맵을 활용한 것이 아니라 학습 단계에서 에지 맵을 활용했었다[8]. 이에 따라 본 논문에서는 이미 의미론적 영상 분할이 된 이미지를 대상으로, 에지 맵을 참고해 레이블 맵을 수행하여 분할 정확도를 향상하는 후처리 알고리즘 방법을 제안한다.

II. 관련 연구

1. 의미론적 영상 분할을 위한 CNN(Convolutional Neural Network)

의미론적 영상 분할(Semantic Image Segmentation)은 컴퓨터비전 분야의 기술 중 하나로, 이미지를 픽셀 단위로 분할하여 클래스를 나누는 기술이다. 예를 들어 영상 안에서 책상이나 벽, 사람 등과 같은 의미를 지니는 단위인 '클래스'로 분할하는데, 이를 픽셀 단위로 수행한다. 전체적으로는 원본 영상을 입력받아 같은 크기의 레이블 맵을 출력하는 구조이다. 최근에는 의미론적 영상 분할을 위해 CNN(Convolutional Neural Network)[9]이나 RNN(Recurrent Neural Network)[10]등을 활용한 학습 네트워크를 설계해 학습하는 방법이 주로 쓰이고 있다. 기존 DNN을 사용해 이미지를 학습시키면 이미지 고유의 2차원 공간 정보가 사라지는 문제에 대한 해결책으로 떠오른 학습 방법이다.

CNN이 기존 DNN과 다른 부분은 입력 이미지의 공간 정보를 유지하기 위해 영상 처리 기법인 convolution을 이용한 계층과 이미지의 크기를 줄이는 pooling 계층을 삽입한 점이다. Convolution에 사용되는 필터는 구성된 가중치(Weight)에 따라 에지를 검출하거나 이미지에서 고주파나 저주파 성분을 제거할 수 있으며 특정 패턴을 지닌 부분을 검출해낼 수 있다. 그 크기는 3x3, 5x5, 11x11 등으로 다양하고 크기가 클수록 넓은 범위에서 필터에 해당하는 패턴을 추출

할 수 있다. CNN은 필터의 가중치를 학습시켜 특정 패턴을 검출해낼 수 있도록 설계되었고 학습시킨 필터에 입력 이미지를 convolution 연산했을 때, 학습된 패턴과 유사할수록 높은 가중치를 가진 특징 맵(Feature Map)이 출력된다. 이후 반복해서 여러 필터를 거친 값들이 완전 연결 계층(Fully-Connected Layer)에서 종합되어 최종적으로 결과인 클래스가 출력되는 구조이다. 실제로 AlexNet은 convolution과 pooling 계층을 여러 번 반복하여 총 8계층으로 신경망을 구성하여 풍부한 특징을 학습할 수 있도록 하였다.

2. 의미론적 영상 분할 연구 동향

의미론적 영상 분할을 수행하기 위해서 CNN을 그대로 사용하면 문제가 발생한다. 보통 CNN 후반부에는 완전 연결 계층이 존재하여 모든 데이터가 연결되어 종합되기 때문에 2차원 이미지의 위치 정보가 최종적으로 사라지기 때문이다. [1]에서는 CNN 상에서 완전 연결 계층에 도달하기 전 얻어진 정보에 이미 분류가 가능할 정도의 충분한 특징 패턴이 있고, 그 위치에 대한 정보도 지금까지 convolution 및 pooling만을 거쳤기 때문에 유지하고 있다는 점에 집중하였다. 그래서 마지막 계층인 완전 연결 계층을 없애고 전부 convolution 및 pooling 계층으로 구성한 FCN(Fully Convolutional Networks)를 발표하였다. CNN의 특성상 convolution과 pooling을 거치게 되면 특징 맵의 크기가 줄어들게 된다. 원본 이미지의 크기와 같은 크기로 픽셀 단위의 세밀한 영상 분할을 하려면 줄어든 특징 맵의 결과를 다시 키우는 과정을 거쳐야 하는데, 그 과정을 Up-scale 등으로 부른다.

간단한 Up-scale 방법은 양선형 보간법(Bilinear Interpolation)을 수행해 크기를 늘리는 방법이 있으나, 그 방법만을 사용한다면 원본을 복원하기엔 세밀함이 떨어진다. 따라서 FCN[1]에서는 세밀한 정보를 보강하기 위해 각 convolution 계층별로 남아있는 pooling이 수행되기 이전에 조금 더 큰 크기의 중간 결과(특징 맵)를 참고하여 원본의 세밀한 부분을 살려 정교하게 예측하고자 하였다.

이후 원본 크기로 복원한 영상 분할의 세밀함이 부족하다는 문제와 부족한 인식 성능 문제가 제기되면서, 그 이후 해마다 성능을 향상한 방법을 제안하는 논문들이 다수 발표되었다. [2]는 데이터 셋에 포함돼있는 깊이 맵의 정보를 같이 학습하여 두 학습의 결과를 합친 방법을 사용했고 [3]은 Deeplab과 ResNet[11]의 101 계층 버전을 조합하여 학습하였다. [4]는 pooling 시에 핵심적인 부분의 매핑 정보를 저장하고 활용하여 세밀한 복원을 하도록 구성한 신경망을 구성했고, [5]는 잔여 학습을 활용한 Encoder-Decoder 구조를 설계하는 등의 연구들이 진행되었다. 이같이 최근 발표된 학습을 통한 의미론적 영상 분할 분야의 연구는 주로 ResNet과 같은 신경망을 기반으로 하거나 새로운 신경망 구조로 대체하는 연구, 또는 영상 분할 맵을 세밀하게 복원하는 방법[12]을 제안하는 추세이다.

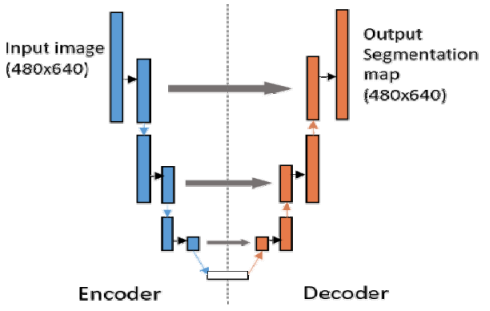


그림 2. 의미론적 영상 분할 전체 흐름도 예시

본 논문에서는 최근까지 성능을 향상한 구조인 Deeplab[6][7]을 본 논문에서 전처리 과정으로써 활용하였다. Deeplab은 핵심적으로 Atrous Convolution, ASPP(Atrous Spatial Pyramid Pooling), Encoder-Decoder 구조 등을 이용해 성능을 향상 시켜왔다. Atrous convolution이란 convolution 시에 필터 일부분만 사용하고 나머지는 0으로 채워 연산하는 방법으로, 학습시킬 필터의 가중치 개수가 줄어드는 효과를 얻을 수 있다. 그로 인해 기존에 연산량 때문에 적용하지 못했던 큰 크기의 필터를 사용할 수 있는 이점을 얻었다. ASPP는 Atrous convolution을 활용해 여러 크기의 필터를 연산하고 이를 다시 하나의 특징 맵으로 합쳐주는 방법이며, 더 넓은 범위의 특징을 연산의 증가 없이 검출할 수 있게 되었다[6]. Encoder - Decoder 구조는 Encoder 부분에 CNN을 배치하여

중간에서 최종적 특징 맵을 추출한다. Decoder 부분에는 Encoder와 대칭이 되는 신경망을 배치하고 대칭되는 각 convolution 계층에 남아있던 특징 맵을 참고로 하여 Up-Scale을 수행해서 세밀한 영상 분할에 초점을 두었고 pooling 계층을 일부 삭제해서 특징 맵이 갈수록 줄어드는 부분을 줄였다[7]. 이러한 요소들로 인해 Deeplab은 현재 우수한 의미론적 영상 분할 성능을 보여주고 있다.

[그림 2]는 최근 기계 학습을 이용한 의미론적 영상 분할 구조에서 가장 보편적으로 사용되고 있는 Encoder-Decoder 구조를 간략하게 그림으로 표현한 것이다. 그림의 왼쪽 부분인 Encoder에서는 이미지가 입력되면 convolution 계층에서 특징 맵과 연산이 수행되고, 그 후 Pooling 계층에서 이미지의 크기가 작아진다. 이 과정을 반복하여 [그림 2]의 중앙 하단에 있는 부분에서 각 클래스 필터 별로 다르게 분할된 최종 특징 맵을 얻는다. 이 최종 특징 맵을 하나로 합치게 되면 작은 크기의 초기 영상 분할 맵, 즉 레이블 맵이 생성된다. 따라서 [그림 2]의 중앙 하단 부분은 가장 작은 크기의 레이블 맵을 의미한다. 그 이후 그림의 오른쪽 부분인 Decoder에서는 작은 레이블 맵을 원본 영상 크기만큼 확대한다. 이 과정에서 앞서 언급했듯이 다양한 방법을 사용한 성능 향상 연구가 진행되고 있다. 학습 시에는 각 convolution 계층에서 사용되는 특징 맵을 학습하는 구조이다.

Ⅲ. 제안하는 방법

의미론적 영상 분할 기술에서는 에지 부분에서 레이블의 분할이 정확하지 않아 세세한 분할 정확도가 부족했으며 이 문제는 초기 방법인 FCN[1]부터 현재까지 계속 지적되어오고 있는 문제점이다. 물론 최근까지도 의미론적 영상 분할의 정확도는 상승하고 있으나 아직 정확도가 완벽에 가깝지 않으며, 에지를 기준으로 레이블이 잘못 예측이 되는 경우가 있다.

에지 주변의 세밀하지 못한 분할의 원인으로는 레이블 맵의 크기를 늘리는 과정에서 발생한 문제로 볼 수 있다. 이 문제를 해결하기 위해 누락이 된 세밀한 정보를 활용할 수 있을 것이고, 영상의 세밀한 정보 중 하나인 에지 맵을 이용해 잘못 분할된 영역을 보정 할 수 있을 것이다. 따라서 본 논문에서는 의미론적 영상 분할이 된 레이블 맵을 대상으로 에지 맵 기반 알고리즘으로 레이블 값을 수정하여, 정확도와 세밀한 분할 기능을 강화하는 알고리즘을 제안한다.

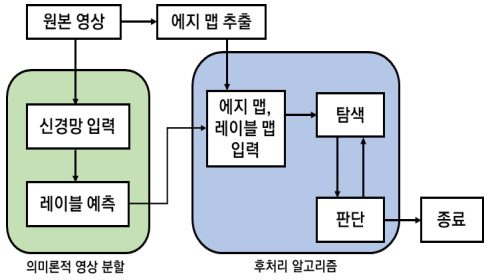


그림 3. 제안하는 알고리즘을 적용한 의미론적 영상 분할의 전체 흐름도

1. 제안하는 방법의 개요

본 논문은 의미론적 영상 분할의 결과물인 레이블 맵에 대해, 원본 영상의 에지 맵을 기반으로 하여 레이블을 수정을 후처리 알고리즘을 제안한다. 알고리즘을 적용했을 때를 가정하여 전체적인 과정을 설계한다면 [그림 4]와 같이 표현할 수 있다.

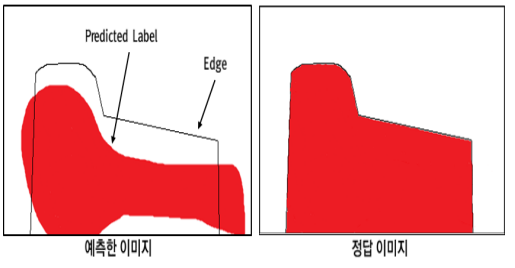


그림 4. 세밀하지 못한 분할의 예시

원본 영상을 학습시킨 신경망에 입력하여 결과물인 레이블 맵을 얻고, 원본 영상에 캐니 에지 검출기를 이용해 에지 맵을 추출한다. 그 후 에지 맵과 레이블 맵이 후처리 알고리즘에 입력되고 후처리 알고리즘 수행을 시작한다. 후처리 알고리즘은 탐색-판단의 과정으로 나눌 수 있는데, 독립적으로 수행되는 것이 아니라 서로

반복하며 수행된다. 오직 에지 픽셀에서만 탐색을 시작하고, 한 에지 픽셀은 8방향으로 전진하며 레이블 값을 수집한다. 탐색이 끝나면 곧바로 이어지는 판단 과정에서는 탐색 때 저장한 레이블 값과 여러 정보를 바탕으로 현재 레이블을 유지할 것인지, 변경할 것인지 판단한다. 한 판단이 끝나면 해당 방향은 종료되어 다음 방향으로 탐색을 시작하고, 8방향 모두 탐색과 판단을 끝마치면 다음 에지 픽셀로 이동해 8방향의 탐색-판단 과정을 다시 반복한다. 최종적으로 모든 에지 픽셀에 대해 수행되면 알고리즘은 종료된다.

알고리즘의 목표는 에지 부분의 레이블을 수정하여 소폭의 정확도를 향상하는 것이다. 학습 기반의 의미론적 영상 분할은 점점 그 정확도가 상승하는 중이기 때문에, 알고리즘은 기본적으로 확실한 정보에 대해서만 레이블을 수정하고, 안정적인 성능을 보여야 한다. 예를 들어 알고리즘을 수행하기 전의 정확도가 70%일 경우와 90%인 경우가 있을 때, 알고리즘은 안정적으로 두 경우 모두 비슷한 소폭의 정확도 상승과 에지 부분의 레이블 향상이 있도록 설계하였다.

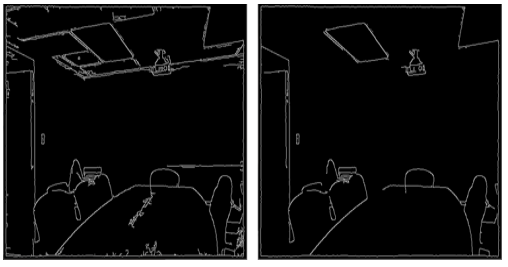


그림 5. 에지 맵 추출 예시 (좌측이 낮은 Threshold, 우측이 높은 Threshold)

2. 전처리 과정

알고리즘을 수행하기에 앞서, 에지 맵과 의미론적 영상 분할된 레이블 맵이 필요하다. 레이블 맵은 앞서 2장에서 설명한 것과 같이 얻었다고 가정하고, 에지 맵을 원본 영상에서 추출해야 한다. 에지를 검출하는데 사용한 방법은 Canny 에지 검출기를 사용했으며, 이 과정에서 2개의 에지 맵을 생성한다. 하나는 낮게 Threshold 값을 설정하여 에지 검출을 민감하고 세밀한 부분까지 검출하도록 했고, 다른 하나는 높게 Threshold 값을 설정하여 전자와 비교해 굵고 확실한 에지만을 검출하도록 하였다. 이를 통해 얻은 에지 맵을 탐색-판단 알고리즘에 입력하여 사용한다. 그 결과 얻어진 에지 맵은 [그림 5]와 같이 서로 차이를 보인다.

3. 탐색-판단 알고리즘

제안하는 알고리즘은 전처리 과정에서 얻어진 에지 맵과 레이블 맵이 입력되어 수행된다. 에지 부분의 레이블 값 수정이 주요 목표이기 때문에, 강한 에지 픽셀부터 8방향 탐색 후 판단을 한다. [그림 6]과 같은 알고리즘으로 구성되어 있다. 탐색의 시작을 강한 에지 픽셀로부터 한 이유는 강인하고 확실한 에지를 대상으로 레이블 값을 수정하기 위해서이며, 탐색 종료 조건을 다음 에지 픽셀까지이지만 이는 민감한 에지에도 탐색이 종료될 수 있게 설정하였다.

```
For p = 1, ··· , Edges          # 강한 에지 픽셀의 수 만큼 반복
  For d = 1, ··· , 8                    # 8방향 탐색을 위해 8회 반복

    탐색)    해당 방향으로 1픽셀 씩 이동하며 레이블 값 저장
             (1) 레이블 값이 변한 경우, 해당 좌표를 저장한 후 다시 탐색
             (2) 에지 픽셀에 도달 시, 총 distance 저장 후 탐색 종료

    판단)    if (Label_Change == True) :
               (1) if ( a*3 < b AND b/distance >= DR ) :
                     a의 레이블을 전부 b의 레이블로 변경
               (2) if ( a==c AND b <= 5 ) :
                     b 레이블을 a와 c의 레이블로 변경
```

그림 6. 제안하는 알고리즘

[그림 7]과 같이 탐색 과정에서는 레이블 값, 레이블 값 변화 지점, 탐색 거리 등을 기록한다. 그 종료 조건은 전진 탐색 중 에지 픽셀에 도달했을 경우이다. 판단 과정에서는 탐색 과정에서 얻어진 정보를 바탕으로 레이블 값 수정 여부를 판단한다. 레이블 변화가 없는 경우에는 해당 방향이 정상적으로 분할된 레이블로 판단하고, 레이블 변화가 있는 경우에만 판단 과정의 (1), (2)을 수행한다.

판단 과정에서의 a는 시작 지점부터 레이블 첫 변화 지점까지의 거리, b는 첫 레이블 변화 지점부터 다음 레이블 변화 지점까지의 거리, c는 두 번째 레이블 변화 지점부터 세 번째 레이블 변화 지점까지의 거리를

정의하였다. distance는 해당 방향의 총 탐색 거리이며, DR은 Dominative Rate를 의미하고 해당 방향에서 레이블이 지배적인지 판단하는 비율로, 0.8을 설정하여 총 탐색 거리에서 한 레이블이 80% 이상을 차지하면 지배적으로 판단하도록 설정하였다.

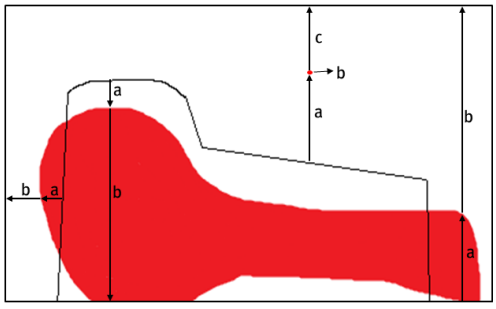


그림 7. 알고리즘 수행 예시

판단 (1)의 a*3 〈 b 조건은 레이블을 보수적으로 수정하기 위한 조건으로, 수정할 대상인 a가 b의 1/3을 넘지 않을 조건에만 수정하도록 하였다. b/distance 〉= DR 조건은 해당 방향의 전체 탐색 거리에서 b의 레이블이 지배적인지 판단한다. 이에 따라 판단의 (1)에서는 [그림 7]과 같은 형태의 잘못 분류된 레이블을 수정하게 된다. 판단 (2)에서는 노이즈, 얼룩과 같이 잘못 분류된 레이블을 수정한다. a와 c가 같은 레이블이고, 그 사이의 b가 5 이하인 비교적 작은 부분일 경우, a, b, c가 전부 같은 레이블로 되도록 변경한다. [그림 7]의 예시로 나타낼 수 있다.

제안하는 알고리즘은 이러한 과정을 거쳐 모든 강한 에지 픽셀에 대해 레이블을 8방향으로 수정한다. 조건을 전부 만족할 때에만 레이블이 수정되도록 설계했기 때문에, 보수적이지만 안정적인 성능을 발휘하도록 하였다.

Ⅳ. 실험 및 결과

1. 실험 환경

실제 학습과 실험에 사용된 컴퓨터의 CPU는 AMD Ryzen 5 1600 3.2GHz, 메모리는 16GB, GPU는 NVIDIA GTX 1060 6GB, OS는 Windows 10 Pro 64-bit 환경을 사용하였고 Python 3.7, CUDA 10.1 버전 등을 활용하였다. 모든 학습에 사용된 신경망은 ResNet[11]의 50 Layer 구조를 사용한 Deeplab[6][7]의 v3+ 버전을 사용했고 Tensorflow로 구현한 소스코드를 참조하였다. Hyper-Parameter 설정은 이미지 crop size를 256x256으로 설정하였고, 손실 함수로는 Cross Entropy, 활성화 함수를 ReLU, Learning rate는 0.0001에 Adam 기법을 적용해 학습하였다. 제안하는 알고리즘은 위에 설명한 신경망과 파라미터와는 관계가 없기에 다른 형태로 구성해도 상관없다.

2. 정확도 측정 방법

본 논문에서 픽셀 정확도는 예측한 레이블 맵을 정답 레이블 맵과 비교하여 (정답 픽셀 개수 / 전체 픽셀 개수)로 한 이미지에서 레이블을 올바르게 분류한 확률을 계산하였다. 그리고 의미론적 영상 분할 기술에서 보편적으로 사용되는 정확도 측정 방법인 IoU(Intersection of Union)를 사용하였다. IoU는 클래스 별로 정확도를 측정하며, 수식(1)과 같다. A는 이미지에서 해당 클래스의 정답 값 영역, B는 이미지에서 해당 클래스의 예측값 영역을 의미하고 정답 값 영역과 예측값 영역의 교집합/합집합을 수식화한 것이다.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (1)$$

3. NYU v2 데이터셋 실험

우선 NYU Depth v2 데이터 셋[13]을 이용한 실험을 수행하였다. 데이터들은 실내 공간 이미지 1,448장으로 이루어져 있고 전부 480x640 크기이며, 같은 크기로 의미론적 영상 분할이 된 레이블 맵이 포함되어 있다. 추가로 깊이 정보가 있으나 본 실험에서는 사용하지 않았다.

표 1. NYU v2 데이터 셋 실험, 평균 정확도 비교

항목	알고리즘 적용 전	알고리즘 적용 후
픽셀 정확도	67.71%	69.45%
IoU 정확도	44.43%	45.78%

[표 1]은 위 실험 환경에 명시한 대로 Deeplab v3+ 신경망을 학습한 후 NYU Depth v2 데이터 셋에 의미론적 영상 분할을 수행하고, 그 결과인 1,448장의 이미지(레이블 맵)를 대상으로 알고리즘을 수행한 결과이다. 알고리즘 적용 전과 적용 후를 비교했으며 평균적으로 적용 후에 픽셀 정확도가 약 1.74%p, 평균 IoU는 약 1.35%p만큼 정확도가 상승하였다.

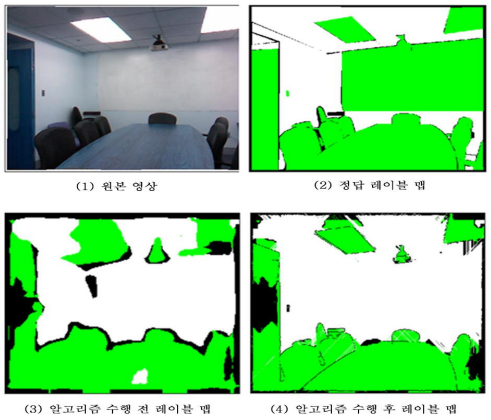


그림 8. NYU v2 데이터 셋 실험 : 알고리즘 수행 전/후의 실제 예시

[그림 8]은 실제 데이터 셋에 알고리즘을 적용 전과 적용 후의 예시이다. 예시를 보았을 때, 비어있는 레이블 부분과 얼룩처럼 잘못 예측된 레이블이 성공적으로 수정되었고, 다른 레이블도 에지를 따라서 수정이 된 것을 볼 수 있다.

표 2. NYU v2 데이터 셋 실험 : 알고리즘 적용 후 정확도 상승 비율 분석

기준 픽셀 정확도	전체 항목 수	정확도 상승 항목 수	정확도 상승 비율	정확도 상승 평균
51% ~ 60%	252	198	79%	1.16%
61% ~ 70%	513	457	89%	1.77%
71% ~ 80%	452	419	93%	2.11%
81% ~ 90%	130	125	96%	2.40%
91% ~ 100%	2	2	100%	1.95%
종합	1448	1255	87%	1.74%

[표 2]는 [표 1]을 자세히 분석한 표로, 알고리즘 적용 전의 픽셀 정확도를 기준으로 정렬해 비교한 것이다. 전체 데이터 셋에서 픽셀 정확도가 상승한 이미지는 1448장 중 1255장이고, 이를 백분율로 환산하면 약 86.67%이다. 정확도 상승 비율에 집중해서 보면, 기준 픽셀 정확도가 높을수록 정확도가 비례적으로 안정적인 성능을 보였음을 알 수 있다.

4. CamVid 데이터셋 실험

추가로 CamVid 데이터 셋[14]을 이용한 실험을 수행하였다. 데이터들은 도로 영상이 대부분인 실외 공간 이미지 701장으로 이루어져 있고 전부 960x720 크기이며, 같은 크기로 의미론적 영상 분할이 된 레이블 맵이 포함되어 있다.

[표.3]은 위 실험 환경에 명시한 대로 Deeplab v3+ 신경망을 학습한 후 CamVid 데이터 셋에 의미론적 영상 분할을 수행하고, 그 결과인 701장의 레이블 맵을 대상으로 알고리즘을 수행한 결과이다. 알고리즘 적용 전과 적용 후를 비교했으며 평균적으로 적용 후에 픽셀 정확도가 약 0.73%p, 평균 IoU는 약 0.49%p만큼 정확도가 상승하였다.

표 3. CamVid 데이터 셋 실험 : 알고리즘 적용 전/후 평균 정확도 비교

항목	알고리즘 적용 전	알고리즘 적용 후
픽셀 정확도	91.96%	92.69%
IoU 정확도	70.85%	71.34%

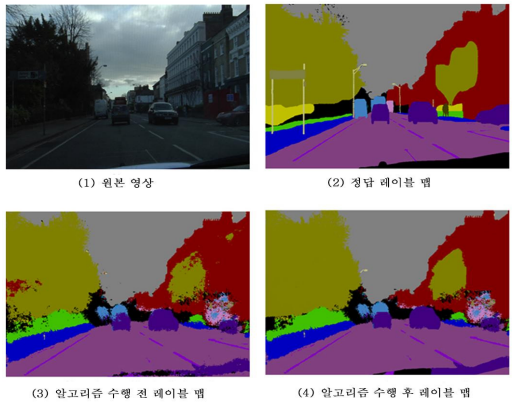


그림 9. CamVid 데이터 셋 실험 : 알고리즘 수행 전/후의 실제 예시

[그림 9]는 CamVid 데이터 셋에 알고리즘을 수행한 실제 예시이다. 특이사항으로, CamVid 데이터 셋의 실험에서는 정확도가 하락하는 경우가 없어서 안정적

인 성능을 보였다.

5. 실험 결과 비교

NYU v2 데이터 셋과 CamVid 데이터 셋에 대해 수행한 실험을 비교해보면, NYU v2 데이터 셋 실험의 경우가 비교적 정확도 상승 비율이 높고, 안정적인 성능을 보였다. 그 첫 번째 이유로는 NYU v2 데이터 셋은 실내 공간 이미지이며, 사람이나 나무, 잔디와 같이 정확한 에지 측정이 힘든 실외 도로 영상보다 비교적 에지가 뚜렷하게 검출되었음을 들 수 있다. 두 번째 이유로 CamVid 데이터 셋은 영상의 깊이 자체가 도로 영상이기 때문에 실제 멀리 있는 물체는 에지로 잘 표현되지도 않고 해상도가 낮아서 픽셀의 누락이 일어났음을 들 수 있다. 따라서 두 실험의 결과를 비교했을 때, 제안하는 알고리즘은 에지 검출이 제대로 되지 않고, 동시에 영상의 깊이가 깊은 실외 도로 영상에서는 비교적 레이블 맵 향상이 적었지만 에지가 뚜렷하게 나타나는 실내 공간 영상에서는 눈에 띄는 성능 향상이 있었다. 두 실험 모두 정확도 하락 비율이 매우 낮거나 없어서 안정성 측면에서 좋은 성능을 보였다.

V. 결론

NYU v2 데이터 셋 기준으로 약 1.74%p의 평균 픽셀 정확도 상승과 1.35%p의 IoU 정확도 상승을 보였다. 알고리즘 수행 후 정확도 비교 데이터를 분석한 결과, 기본적으로 후처리 전에 정확도가 높을수록 알고리즘이 안정적으로 작동하였다.

추가 실험에서는 CamVid 데이터 셋을 대상으로 알고리즘을 수행했으며, 같은 신경망 조건으로 설정하였다. 실험 결과로는 약 0.73%p의 픽셀 정확도 상승과 약 0.49%p의 IoU 정확도 상승을 보였다. 앞선 실험보다는 낮은 향상을 보였는데, 그 이유로 NYU v2 데이터 셋보다 CamVid 데이터 셋이 확실한 에지 검출이 힘들었고 도로 영상의 특성상 영상의 깊이가 깊어서 근본적으로 충분한 특징이 표현되지 않은 것이 정확도 상승이 비교적 낮은 이유로 예상된다. 하지만 두 실험 모두 정확도가 하락하는 비율은 낮거나 없어서 안정성 측면에서 좋은 성능을 보였다.

[표 2]에 보이는 것처럼, 기반 정확도가 높을수록 제안하는 알고리즘이 안정적인 성능을 보이기 때문에 제안하는 알고리즘은 후에도 준수한 성능을 낼 수 있을 것으로 보인다. 제안하는 방법은 원본 영상의 에지 맵에 영향을 크게 받기 때문에 추후 에지 검출 최적화 기법에 관한 연구 및 파라미터 결정 방법에 관한 연구가 더 진행된다면 더 안정적이고 좋은 성능을 낼 수 있을 것이다.

참 고 문 헌

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

[2] Jing Liu, Yuhang Wang, Yong Li, Jun Fu, Jiangyun Li, and Hanqing Lu, "Collaborative deconvolutional neural networks for joint depth estimation and semantic segmentation," IEEE transactions on neural networks and learning systems, Vol.29, No.11, pp.5655-5666, 2018.

[3] Carlos H. Perdiguero, C. R. Cabrera, J. Roberto, and L. Sastre, "In pixels we trust: From Pixel Labeling to Object Localization and Scene Categorization," 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS), 2018.

[4] Yanhua Cheng, Rui Cai, Zhiwei Li, Xin Zhao, and Kaiqi Huang, "Locality-sensitive deconvolution networks with gated fusion for rgb-d indoor semantic segmentation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[5] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid, "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation," Proceedings of the IEEE conference on computer vision and

pattern recognition, 2017.
[6] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE transactions on pattern analysis and machine intelligence, Vol.40, No.4, pp.834-848, 2017.
[7] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," Proceedings of the European conference on computer vision (ECCV), 2018.
[8] D. Marmanisac, K. Schindlerb, J. D. Wegnerb, S. Gallianib, M. Datcua, and U. Stillac, "Classification with an edge: Improving semantic image segmentation with boundary detection," ISPRS Journal of Photogrammetry and Remote Sensing, Vol.135, pp.158-172, 2018.
[9] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom, *A Convolutional Neural Network for Modelling Sentences*, Department of Computer Science University of Oxford, 2014.
[10] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals, "Recurrent Neural Network Regularization," Under review as a conference paper at ICLR, 2015.
[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
[12] Jianbo Jiao, Yunchao Wei, Zequn Jie, Honghui Shi, Rynson Lau, and Thomas S. Huang, "Geometry-Aware Distillation for Indoor Semantic Segmentation," Computer Vision and Pattern Recognition (CVPR), 2019.
[13] Nathan Silberman, Derek HoiemPushmeet and KohliRob Fergus, "Indoor segmentation and support inference from rgbd images," European Conference on Computer Vision, Springer, Berlin, Heidelberg, 2012.
[14] Gabriel J. Brostow, Jamie Shotton, Julien Fauqueur, and Roberto Cipolla, "Segmentation and recognition using structure from motion point clouds," European conference on computer vision, Springer, Berlin, Heidelberg, 2008.

저 자 소 개

김 정 환(Jung-Hwan Kim) 정회원

- 2017년 8월 : 숭실대학교 평생교육원 컴퓨터공학과(공학사)
- 2020년 2월 : 숭실대학교 미디어학과(석사)
- 2020년 ~ 현재 : ㈜셀빅 사원

〈관심분야〉 : 컴퓨터비전, 기계 학습, 실내 공간 인식 등

김 선 혁(Seon-Hyeok Kim) 준회원

- 2016년 8월 : 평생교육원 컴퓨터공학과 (공학사)
- 2016년 9월 ~ 현재 : 숭실대학교 미디어학과 석사과정

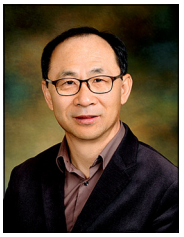
〈관심분야〉 : 컴퓨터비전, 기계 학습, 영상처리 등

김 주 희(Joo-heui Kim) 정회원

- 1997년 2월 : 국립안성산업대학교 전자계산학과(공학사)
- 2005년 2월 : 국립한경대학교 멀티미디어종보통식(석사)
- 2021년 1월 : 숭실대학교 미디어학과 콘텐츠공학 박사과정

〈관심분야〉 : 인터랙티브 콘텐츠, 멀티미디어 콘텐츠, AR 콘텐츠 기획 등

최 형 일(Hyung-Il Choi) **정회원**



- 1979년 : 연세대학교 전자공학과 공학사
- 1983년 : 미시간대학교 전기전산학과 공학석사
- 1987년 : 미시간대학 전기전산학과 공학박사
- 1989년 ~ 1999년 : 숭실대학교 컴퓨터학부 교수
- 2000년 ~ 현재 : 숭실대학교 미디어학과 교수

〈관심분야〉 : 컴퓨터 비전, 패턴인식, 증강현실 등