

딥러닝을 이용한 위성 영상의 의미 정보 추출 기법 동향 분석 위성 영상에서 자동으로 정보를 검출하는 기법 분석

저자 (Authors)	김재선, 허용석
출처 (Source)	정보와 통신 열린강좌 35(2(별책8호)) , 2018.11, 3-9(7 pages) KICS Information & Communication Magazine - Open Lecture Series 35(2(별책8호)) , 2018.11, 3-9(7 pages)
발행처 (Publisher)	한국통신학회 Korea Institute Of Communication Sciences
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07615406
APA Style	김재선, 허용석 (2018). 딥러닝을 이용한 위성 영상의 의미 정보 추출 기법 동향 분석. 정보와 통신 열린강좌, 35(2(별책8호)), 3-9
이용정보 (Accessed)	아주대학교 202.30.20.*** 2020/03/16 12:41 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.



딥러닝을 이용한 위성 영상의 의미 정보 추출 기법 동향 분석

위성 영상에서 자동으로 정보를 검출하는 기법 분석

김재선, 허용석
아주대학교 전자공학과

1. 서론

컴퓨터 비전 (computer vision)은 영상으로부터 그 안에 있는 다양한 정보를 파악하는 것을 목표로 하는 연구 분야이다. 일반적으로 사람은 영상으로부터 정보 파악을 매우 잘한다. 가령, 어린아이에게도 고양이를 촬영한 영상을 한 장 보여 주면 그 영상 안에 있는 대상이 고양이라는 사실을 부지불식간에 인식한다. 사람에게는 이처럼 쉬운 일을 컴퓨터가 잘하게 하기란 쉽지가 않다. 그 이유는 컴퓨터가 입력으로 받는 영상 정보는 영상의 밝기 값에 해당하는 단순한 숫자의 나열이기 때문이다. 사람도 숫자의 나열만을 보고 의미를 파악하기란 쉽지가 않듯이, 이러한 작업을 컴퓨터가 자동으로 처리하게 하는 알고리즘을 만들기는 쉽지가 않은 일이다. 이러한 이유로 영상으로부터 의미를 파악하는 연구는 컴퓨터 비전 연구 중에서 가장 어려운 일 중의 하나로 알려져 왔다.

이러한 의미 파악을 위한 대표적인 컴퓨터 비전 연구 분야는 영상 분류(classification) 및 관심 물체 검출 (object detection) [1], 의미론적 영상 분할 (semantic segmentation) [2] 등이 있다. 영상 분류는 영상 전체에 대해서 미리 정의된 몇 개의 클래스(class 또는 label)들 중의 하나를 결정하는 것을 목표로 하고 있고, 물체 검출은 관심 물체에 대해서 영상 분류뿐만 아니라 물체의 위치(localization) 정보까지도 사각형 경계(bounding box) 형태로 예측하는 기법이다. 의미론적 영상 분할은 영상을 이루는 각 픽셀에 대해서 미리 정의된 클래스들 중의 하나를 결정하는 연구이다.

이러한 연구들은 최근 대용량의 데이터 저장 기술, CPU 및 GPU등의 하드웨어 기술의 발전, 그리고 인공지능 (artificial intelligence) 알고리즘 및 딥러닝 (deep learning)과 같은 머신러닝 (machine learning) 알고리즘 연구의 발전과 맞물려서 좋은 결과를 내는 연구들이 많이 발표되고 있어서 사람들의 관심이 점점 더 커지고 있다.

한편, 소형 위성 및 항공 영상 기술의 발전으로 위성 및 항공 영상 관련 산업 또한 전 세계적으로 주목을 받고 있다. 기존의 군 위주의 위성 영상 산업이 최근에는 민간 기업들 중심으로 위성 영상 정보와 인공지능 알고리즘을 결합하여 새로운 정보를 제공하는 서비스 산업으로 다양하게 확대되어 가고 있다. Google, Microsoft, Facebook, Uber 등 글로벌 대기업뿐만 아니라 Orbital Insight와 Cape Analytics등 다양한 스타트업 회사들이 최근 몇 년 동안 급성장하며 항공 및 위성 영상을 이용하여 새로운 서비스를 앞세우며 시장을 창출하고 있다[3][4][5].

따라서, 본고에서는 이처럼 급격한 발전을 이루고 있는 인공지능 기술의 하나인 딥러닝을 이용하여 위성 영상으로부터 다양한 의미를 추출하는 기법들에 대해서 소개하고자 한다.

최근 위성 산업에 많은 민간 기업들이 관심을 가지면서, 딥러닝을 이용한 위성 영상 분석 알고리즘 연구 또한 증가하고 있다.

II. 위성 영상에서의 의미 추출 기법

위성 영상은 일반적인 이미지에 비해 더 구조적이고 균일한 데이터를 포함하고 있기 때문에 강력한 정보 원천이라 할 수 있다. 그럼에도 불구하고 컴퓨터 비전 분야에서 대부분의 연구는 일반적인 영상에 대해서 이뤄져왔고, 딥러닝을 이용하여 상당한 발전을 이루었지만 위성 영상에 대한 관심과 연구는 상대적으로 적었다.

최근, DeepGlobe는 최근 위성 영상에 대한 연구를 활성화하기 위해 DeepGlobe 위성 영상 이해 챌린지(Satellite Image Understanding Challenge)[6]를 2018년부터 개최하기 시작했다. 이 챌린지에는 도로 추출(Road Extraction), 건물 검출(Building Detection), 토지 피복 분류(Land Cover Classification)의 위성 영상 이해 Task 세 가지와 이를 딥러닝으로 학습하기 위한 각각의 데이터셋(dataset)을 제공한다.

우선, 도로 추출(Road Extraction) Task는 위성 영상의 각 픽셀에 대해 도로인지 아닌지 분류

하는 이진 분류 문제이다. 위성 영상에서 도로를 추출하는 것은 지도로 만들어지지 않은 지역을 비율에 맞춰 지도로 만들 수 있도록 하고, 재난이 발생했을 때 지도를 신속하게 수정하여 구호 활동을 돕거나 세계의 도시 개발을 보다 상세하게 추적하는 것에 응용할 수 있는 연구 분야이다[6].

위성 영상에서의 딥러닝을 이용한 도로 추출에는 고려해야 할 특징이 세 가지 있다. 첫 번째로, 위성 영상이 고해상도 영상이기 때문에 합성곱 신경망은 넓은 수용 영역(Receptive Field)을 가져야 한다. 두 번째로는, 때때로 위성 영상에서 도로가 얇거나, 복잡, 작게 존재할 수 있기 때문에 작은 영역에 대해서도 세부 정보를 보존해야 한다. 세 번째로는, 도로는 자연스러운 연결성과 긴 범위를 갖고 있기 때문에 이러한 특성을 필수적으로 고려해야 한다. L. Zhou et al. 이 제안한 D-LinkNet[7]은 이와 같은 특징들을 모두 고려하여 개발되었고, 도로 추출 Task에서 가장 우수한 성능을 보였다.

DeepGlobe 챌린지는 도로 추출, 건물 검출, 토지 피복 분류의 세 가지 위성 영상 분석 알고리즘에 대해서 데이터셋을 제공하고 성능을 평가한다.

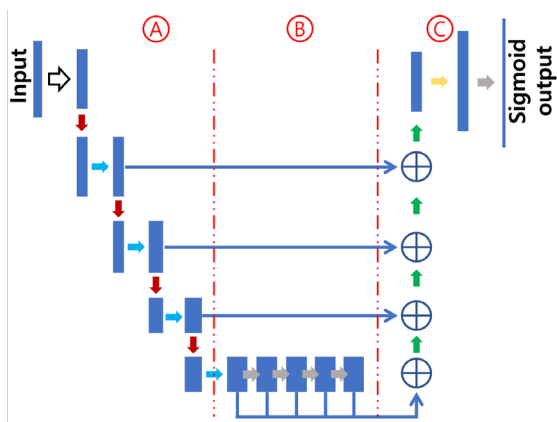


그림 1. D-LinkNet[7] 구조

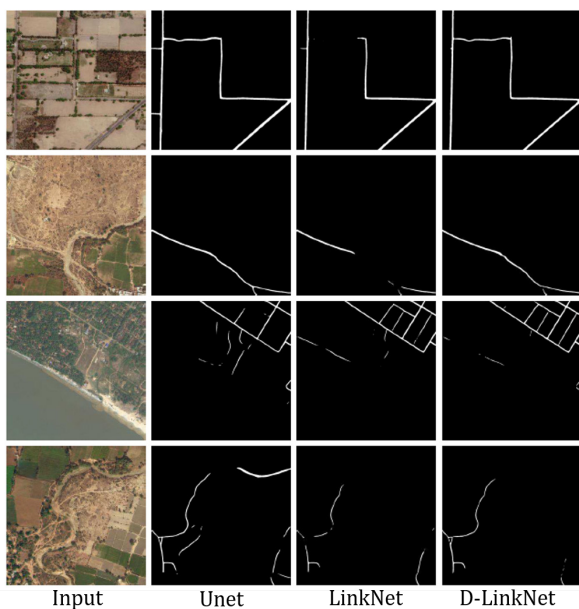


그림 2. 도로 추출 결과 비교

D-LinkNet은 미리 학습된 LinkNet[8]을 기본 골자로 하고, 중간 층에 Dilated 합성곱 층을 추가하여 구성되었다. Dilated 합성곱은 합성곱 신경망의 수용 영역을 넓히는 데 자주 쓰이는 유용한 방법이다.

〈그림 1〉은 [7]에서 제안하는 D-LinkNet의 자세한 구조로, 크게 A, B, C 부분으로 나뉜다. A 부분은 Residual block[9]으로 구성된 ResNet34를 엔코더로 사용한 것이다. 엔코더는 영상을 입력으로 받아서 딥러닝 학습에 사용하기에 용이한 특징 지도(Feature Map) 형태로 추출하는 기능을 하는 부분을 의미한다. Residual block은 합성곱 신경망 알고리즘에서 가장 널리 사용되는 합성곱의 기본 구성 단위로, 합성곱의 결과와 입력의 합을 출력한다. B 부분은 Dilated 합성곱 층으로, 세부 공간 정보(Spatial Information)를 보존하기 위해 효율적으로 수용 영역을 넓히는 데 기여한다. C 부분은 추출한 특징점을 바탕으로 입력 영상의 크기와 같은 결과를 생성하는 디코더이다. 이는 일반적으로 엔코더의 출력이 입력 영상의 크기보다 작은 해상도의 형태로 출력되는 경우가 많기 때문에, 디코더에서는 특징 지도의 해상도를 원래 입력 이미지의 크기로 키워주는 역할을 한다. [7]에서 참고한 LinkNet의 경우는 단순히 엔코더인 A 부분과 디코더인 C부분으로만 구성되어있다.

〈그림 2〉는 DeepGlobe 도로 추출 Validation 데이터셋에 대한 Unet, LinkNet, D-LinkNet의 결과 영상 예시이다. 전체적으로 D-LinkNet은 Unet과 LinkNet 보다 좋은 성능을 보인다.

한편, 건물 검출(Building Detection) Task는 위성 영상의 각 픽셀에 대해 건물인지 아닌지 분류하는 이진 분류 문제이다. 위성 영상에서의 건물 검출은 마케팅, 도시 운영, 인구 추정 등의 다양한 응용에 활용될 수 있기 때문에 잠재 가치가 있는 연구 분야라고 할 수 있다.

최근 합성곱 신경망을 이용한 위성 영상에서의 건물 검출 알고리즘들이 좋은 성능과 함께 제시되었지만, 건물 검출에 있어 중요한 두 가지 특징을 고려하지 않았다. 첫 번째로, 건물 크기의 다양성이다. 실제로 DeepGlobe 데이터셋을 보면 건물의 크기가 다양하게 분포하는데, 크기가 큰 건물과 작은 건물은 시각적인 외관이 서로 매우 다르다. 기존의 알고리즘들은 건물의 크기에 상관없이 모든 건물을 하나의 클래스에 할당했다. 두 번째로는, 거시적인 맥락(context) 정보의 활용이다. 건물 주변의 거시적인 맥락 정보는 때때로 유용하다. 특히, 도로와 건물은 함께 존재하기 때문에 도로에 대한 정보는 매우 유용하다. 이러한 특징들을 고려하여 R. Hamaguchi와 S. Hikosaka는 건물의 크기에 따른 Small, Medium, Large세 가지 분류의 건물 검출과 도로를 검출하는 합성곱 신경망 알고리즘을 제안하였다[11]. 건물 검출 Task에서 가장 우수한 성능을 보

다른 데이터셋에 대하여 미리 학습된 합성곱 신경망을 이용하는 것은 학습 시간을 단축시키면서 성능을 끌어올리는 기법 중의 하나이다.

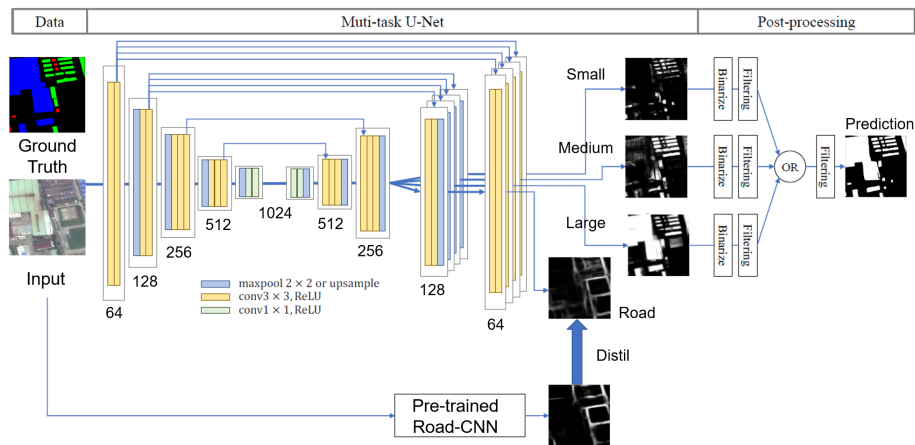


그림 3. [11]의 건물 검출 모델 구조

인 이 합성곱 신경망은 Unet[10]을 기반으로 세 가지 분류의 건물을 각각 검출하고, 후처리 과정을 통해 검출한 모든 건물을 하나의 클래스로 통합한다.

〈그림 3〉은 R. Hamaguchi, S. Hikosaka가 제안한 Unet 기반의 건물 검출 합성곱 신경망 모델의 구조이다. 엔코더 부분에서 추출한 특징 지도를 바탕으로 Small, Medium, Large 크기의 건물과 도로 검출을 하는 4 개의 디코더로 나뉘어져 각각을 검출하도록 학습한다. 후처리 작업을 통해 세 가지 크기의 건물 검출 결과를 통합하여 참값과 비교하고, 도로 검출 결과는 미리 학습된 도로 검출 모델의 결과와 비교한다.

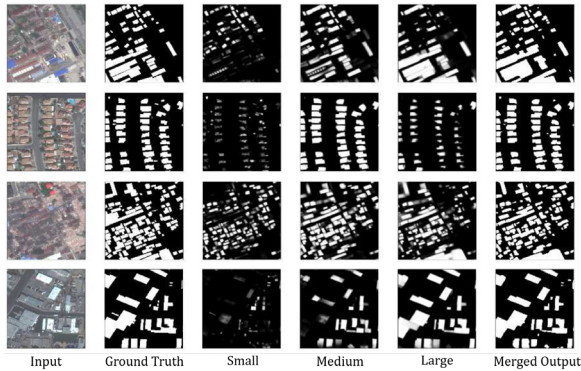


그림 4. [11]의 건물 검출 결과

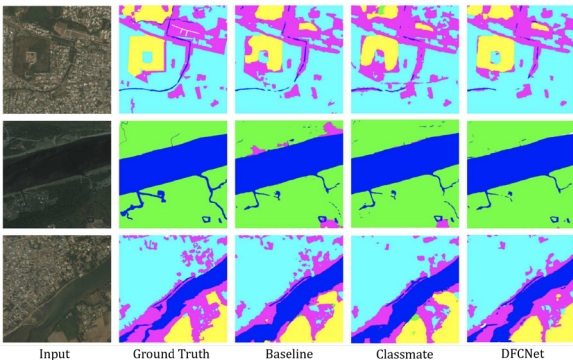


그림 5. 토지 피복 분류 결과 비교

한 기술이다.

합성곱 신경망을 이용한 의미론적 영상 분할의 성능을 높이는 방법에는 크게 세 가지가 있다. 첫 번째로는 엔코더-디코더 구조이다. 엔코더-디코더 구조는 현재 많은 의미론적 영상 분할 모델에서 쓰이고 있다. 엔코더를 통해 특징 지도를 추출하고, 디코더를 통해 입력 영상과 같은 크기의 결과 영상을 만들어낸다. 두 번째로는 특징 결합(Feature Fusion)이다. 입력 영상 전체 영역에서 작은 영역까지의 정보를 충분히 이용하고 표현하기 위해, 특징 결합을 사용한다. 세 번째로는 공간 정보(Spatial Information) 보강이다. 공간 정보는 의미론적 영상 분할을 수행하는 합성곱 신경망 모델의 성능을 높일 수 있다.

이러한 방법들을 사용하여 Chao Tian et al.이 제안한 DFCNet[13]은 DeepGlobe 토지 피복 분류 챌린지에서 가장 우수한 성능을 보였다. DFCNet 은 DeepGlobe 데이터셋에서 도로는 도심, 논밭, 방목장에 나타나는 경향이 있고, 더욱 선명한 도로는 주로 도심에서 나타나는 경향이

〈그림 4〉는 DeepGlobe 건물 검출 Test 데이터셋에 대한 Res-Unet 모델 [11]의 결과 영상 예시이다. 첫 번째 열은 입력 영상, 두 번째 열은 참값, 세 번째부터 다섯 번째 열은 제안하는 각각 모델이 예측한 Small, Medium, Large 건물, 여섯 번째 열은 세 가지 분류의 건물을 하나의 클래스로 통합한 결과 영상이다. 참값과 비교했을 때, 대체로 좋은 결과를 낸 것을 알 수 있다.

또한, 토지 피복 분류(Land Cover Classification) Task는 위성 영상의 각 픽셀에 대해 미리 정의된 복수개의 클래스 중에서 하나의 클래스를 할당하는 의미론적 영상 분할(Semantic Segmentation) 문제이다. 위성 영상을 이용해서 토지 피복을 자동으로 분류하는 기술은 지속 가능한 개발, 자율적 농업, 도시 계획에 중요

엔코더-디코더 구조는 합성곱 신경망을 이용한 의미론적 영상 분할 알고리즘 중에서 가장 널리 쓰이는 구조이다.

있는 특징을 이용하여 학습하였다. 또한 규칙적으로 많은 특징 결합을 반복하여 성능을 높였다. <그림 5>는 DeepGlobe 토지 피복 분류 Test 데이터셋에 대한 각 모델의 결과영상 예시이다. DFCNet의 결과가 가장 참값(Ground Truth)에 가까운 것을 알 수 있다.

한편, 대부분의 위성 영상의 특징은 위에서 아래로 내려다 본 시점에서 영상을 취득하기 때문에 단순히 거리에서 영상을 취득하는 경우와는 매우 다른 특징을 가지고 있다. 딥러닝 학습을 잘하기 위해서는 정답이 있는 다량의 데이터셋이 필요한데 위성 영상은 이러한 특성 때문에 다량의 데이터셋을 확보하기가 어려워 딥러닝의 성능을 최대로 발휘하기가 쉽지 않다. Y. Liu et al.[14]은 이러한 위성 영상의 단점을 극복하기

다량의 데이터셋을 확보하는 것이 딥러닝 학습의 핵심이라 할 수 있다. 딥러닝 알고리즘의 성능을 높이기 위해 다양한 데이터셋에서 추출한 정보를 융합하여 사용하기도 한다.

위하여, 영상의 컬러 정보뿐만 아니라 LiDAR와 같은 3차원 깊이 센서의 정보도 융합하여 보다 향상된 의미론적 영상 분할 기법을 제안하였다[14]. <그림 6>은 [14]에서 제안하는 네트워크의 개요를 보여주고 있다. 위성 영상에 대해서는 기존의 의미론적 영상 분할 기법 [2]을 적용하고, LiDAR 데이터에 대해서는 사람이 고안해낸 특징점 (hand-crafted features)를 사용하여 컬러 영상 특징점 정보와 LiDAR 영상 특징점 정보를 융합하여 최종적인 분류에 적용하였다. 또한 정확한 물체의 경계영역 정보를 위해서 CRF (Conditional Random Field)와 같은 확률적인 유추 방법론도 적용하였다.

<그림 7>은 컬러 위성 영상만을 이용한 의미론적 영상 분할 (FCN-8s) 결과와 LiDAR센서 데이터만을 이용한 결과, 그리고 두 가지 센서 정보들을 융합하여 사용했을 때의 결과를 비교하고 있다. 기대했던 것과 같이 두 가지 센서 정보를 융합한 결과가 각각만을 사용했을 때보다 정확한 결과를 내는 것을 확인할 수 있다.

한편, Azimi et al.[15]은 위성 영상에 대해서 차선과 같이 매우 얇고 가는 물체를 검출하는 기법을 제안하였다. 차선의 검출은 자율 주행과 같이 최근 이슈가 되고 있는 기술에 있어서 매우 중요한 기법 중의 하나이다. 일반 거리 영상에 대해서 차선을 검출하는 기법은 많이 제안되어 왔으나, 위성 영상에 대해서 차선을 검출하는 기법은 거의 없는 실정이다. 위성 영상에 대해서 차선 검출이 어려운 이유는 여러 가지가 있지만 매우 작고 가는 형태를 가지고 있고, 그림자나 노이즈, 또는 배경이 복잡한 영상에 대해서는 잘못 판단할 여지가 많고, 다른 물체들에 의해서 가리어짐이 있는 경우가 흔하기 때문이다.

이러한 문제에 대해서 보다 정확도를 향상시키기 위하여 [15]에서는 <그림 8>과 같이 인코더와

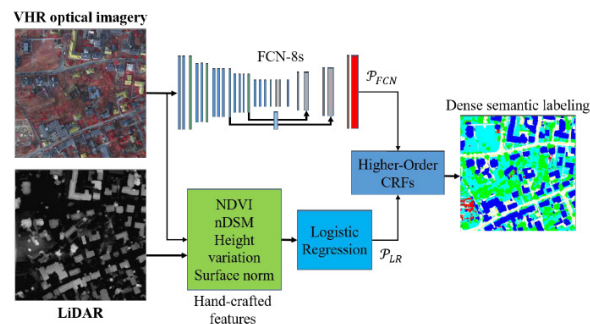


그림 6. [14]에서 제안한 구조

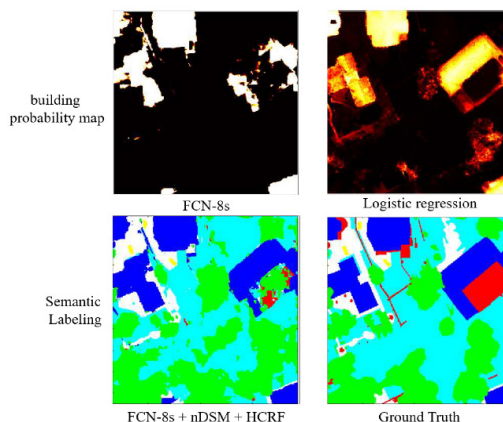


그림 7. [14]에서 제안한 센서 융합 방법 비교

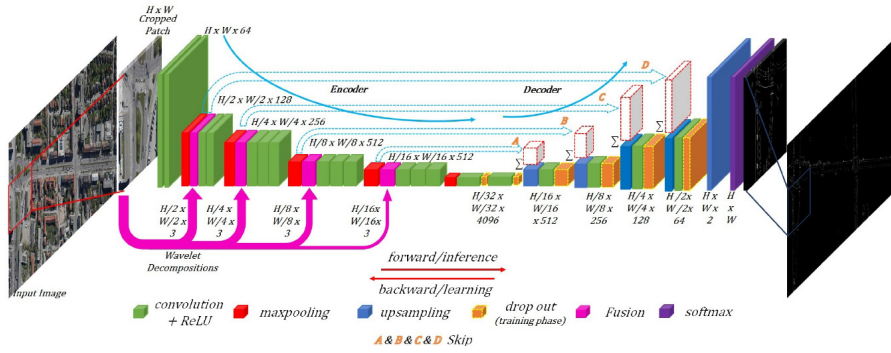


그림 8. [15]에서 제안한 구조

디코더 기반의 네트워크를 제안하였다. 기존의 네트워크와는 다르게 엔코더에서 각 층의 입력부의 특징점에 그 특징점에 해당하는 DWT(Discrete Wavelet Transform)의 결과도 같이 삽입하여 사용함으로써 보다 효율적으로 가늘거나 작은 형태의 물체를 검출하는데 최적화되도록 설계되었다.

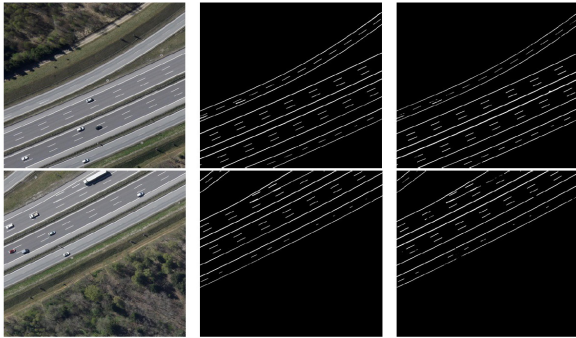


그림 9. [15]에서 제안하는 차선 검출 결과

다. <그림 9>는 [15]에서 제안하는 방법의 결과를 보여주고 있다. <그림 9>의 첫번째 열은 입력 이미지를, 두번째 열은 참값에 해당하는 이미지이고, 세번째 열은 [15]에서 제안하는 방법에 대한 결과이다. <그림 9>에서 관찰할 수 있듯이 [15]의 방법은 매우 정확한 차선 검출 결과를 내는 것을 확인할 수 있다.

III. 결론

본고에서는 딥러닝을 이용하여 위성영상에서 다양한 의미 정보를 추출하는 기법들에 대해서 살펴보았다. 위성 영상은 보다 넓은 지역을 위에서 아래로 취득한다는 특징 덕분에 도로 추출, 건물 검출, 토지 피복 분류, 차선 검출 등과 같은 거시적인 정보를 취득하기에 적합하다. 그러나, 딥러닝을 위해서는 방대한 데이터 셋의 취득이 필수적인데, 일반적인 거리 영상과는 달리 위성 영상은 참값이 있는 데이터셋을 만들기가 쉽지 않은 단점도 존재한다. 이러한 단점을 극복하기 위해서는 위성 영상에 특화된 딥러닝 구조가 필요하다. 본고에서는 이러한 위성 영상에 특화된 다양한 딥러닝 구조들에 대해서 살펴보았다.

위성 영상은 참값이 있는 데이터셋을 대량으로 확보하기 쉽지 않기 때문에 딥러닝을 이용한 위성 영상 분석 알고리즘을 개발하기 위해서는 위성 영상에 특화된 네트워크 구조가 필요하다.

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (2018-0-01424).

참고 문헌

- [1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE International Conference on Computer Vision (ICCV), 2015.
- [2] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
- [3] <https://www.technologyreview.com/the-download/611701/facebook-is-creating-an-internet-satellite/>.
- [4] <https://orbitalinsight.com/>.
- [5] <https://capeanalytics.com/>.
- [6] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar. "Deepglobe 2018: A challenge to parse the earth through satellite images," arXiv preprint arXiv:1805.06561, 2018.
- [7] L. Zhou, C. Zhang, M. Wu, "D-LinkNet: LinkNet With Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction," The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, USA, June 2018.
- [8] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," arXiv preprint arXiv:1707.03718, 2017.
- [9] K. He, X. Zhang, S. Ren, and J. Sun "Deep residual learning for image recognition," IEEE conference on computer vision and pattern recognition, pages 770-778, 2016.
- [10] O. Ronneberger, P. Fischer, and T. Brox, "Unet: Convolutional networks for biomedical image segmentation," International Conference on Medical image computing and computer-assisted intervention, pages 234-241. Springer, 2015.
- [11] R. Hamaguchi, S. Hikosaka, "Building Detection from Satellite Imagery using Ensemble of Size-specific Detectors," IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, USA, June 2018.
- [12] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition," CoRR, 2014.
- [13] C. Tian, C. Li, J. Shi, "Dense Fusion Classmate Network for Land Cover Classification," IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, USA, June 2018.
- [14] Y. Liu, S. Piramanayagam, S.T. Monteiro, E. Saber, "Dense Semantic Labeling of Veri-High-Resolution Aerial Imagery and LiDAR with Fully-Convolutional Neural Networks and Higher-Order CRFs," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [15] S. M. Azimi, P. Fischer, M. Korner, P. Reinartz, "Aerial LaneNet: Lane Marking Semantic Segmentation in Aerial Imagery using Wavelet-Enhanced Cost-sensitive Symmetric Fully convolutional Neural Network," CoRR, 2018.