

# 다양한 배경을 가지는 영상에서의 딥러닝 기반 신분증 검출

이지운, 박재혁, 황인창, 강현수\*

충북대학교

dlwldns2@gmail.com, hyuk3880@naver.com, inchang9507@gmail.com, hskang@cbnu.ac.kr\*

## Deep Learning based ID card detection in images with various backgrounds

Lee Ji Woon, Park Jae Hyuk, Hwang In Chang, Kang Hyun Soo\*

Chungbuk National Univ.

### 요약

본 논문은 영상 내에서 신분증 검출을 위하여 U-net 기반 세가지 모델을 제시한다. 기존의 딥러닝을 사용하지 않고 경계선을 검출하는 방식과 이미지 분할을 목적으로 제안된 Fully-Convolutional Network 기반 모델인 U-Net을 활용하여 CNN based U-net, VGG16을 인코더로 사용하는 U-net, Resnet50을 인코더로 사용하는 U-net을 각각 구성, 학습하여 다양한 배경을 가지는 영상에서 신분증을 검출하였다. 그 결과 Resnet50을 인코더로 사용한 U-net 성능이 가장 우수하였으며, 이는 다양한 영상에서 신분증 경계선 검출 및 신분증을 활용한 인터넷, 모바일 작업, 또한 그와 관련된 신분증 위변조 보안 문제 등의 전처리과정으로 유용하게 사용될 수 있다.

### I. 서론

U-Net[1]는 생물학 분야에서 이미지 분할(Image Segmentation)을 목적으로 제안된 End-to-End 방식의 Fully-Convolutional Network 기반 모델이다. 네트워크 구성의 형태('U')로 인하여 U-Net이라는 이름이 붙여졌다. U-net은 입력 이미지의 특징을 추출할 수 있도록 채널 수를 늘리면서 차원을 축소해나가는 인코더(Encoder) 부분과 저차원으로 인코딩된 정보를 이용하여 채널 수를 줄이고 차원을 늘려서 고차원의 이미지를 복원하는 디코더(Decoder) 부분으로 구성된다. 본 논문에서는 인코더 부분에 일반적인 CNN(Convolution Neural Network) 모델, VGG 16[2] 모델, Resnet50[3] 모델을 사용하여 각각의 성능을 비교, 분석하였다.

본 연구 결과는 딥러닝을 사용하지 않고 물체를 찾는 기존 방식과 비교하여 딥러닝을 사용하여 물체를 찾는 방식의 결과를 제시하고 향후 신분증 관련 인터넷, 모바일 작업에 도움이 되는 것을 목표로 한다.

### II. 본론

영상 분할(Image Segmentation)이란 영상을 여러 개의 픽셀 집합으로 분류하는 것을 말한다. 영상 분할은 특히 영상에서 물체와 경계를 찾는 데 사용되며, 딥러닝을 이용한 영상 분할 알고리즘 중 본 논문에서는 영상 분할을 위하여 U-net을 사용하여 학습을 진행하였다.

그림 1은 U-net의 구조를 나타낸 것이다. U-net은 오토인코더(autoencoder)와 같은 인코더-디코더 기반 모델에 속한다. 그림 1에서 중앙을 기준으로 왼쪽은 특징 추출을 위해 채널 수를 늘리고 차원을 축소해나가는 인코더 부분이며, 오른쪽은 인코딩된 정보를 이용하여 고차원의 이미지를 복원하는 디코더 부분이다. 본 논문에서는 인코더 부분에 CNN, VGG16, Resnet50 모델을 사용하여 학습을 진행하였으며, 디코더 부분에 디콘볼루션을 사용하여 이미지를 복원하였다. 또한 인코딩 단계의 각 레이어에서 얻은 특징을 디코딩 단계의 각 레이어에 합치는(concatenation) 방법을 사용하였으며, 이 직접 연결을 스킵 연결(skip connection)이라고

한다.

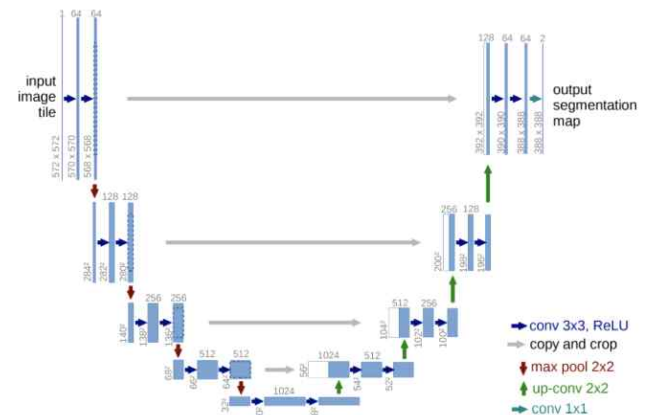


그림 1 U-net 구조

본 논문에서는 정확도를 높이기 위하여 Binary Cross Entropy가 아닌 Focal Loss[4]를 사용하였다. Focal Loss는 Cross Entropy의 클래스 불균형 문제를 다루기 위한 개선된 버전이라고 말할 수 있으며, 어렵거나 쉽게 오분류되는 케이스에 대하여 더 큰 가중치를 주는 방법을 사용한다. 따라서 전경(background) 대비 배경(background)의 객체가 많이 나오는 영상 분할에 효과적으로 사용될 수 있다. Focal Loss는 Cross Entropy Loss에 조절 항  $(1 - p_t)^\gamma$ 을 추가한 것이다.  $\gamma$ 는 조절 가능한 focusing parameter로 0 이상의 값으로 세팅하며, 다음과 같이 Focal Loss를 정의한다.

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t) \quad (1)$$

그림2는  $\gamma$ 을 0~5 사이의 값으로 조절하였을 때의 함수 형태이다.

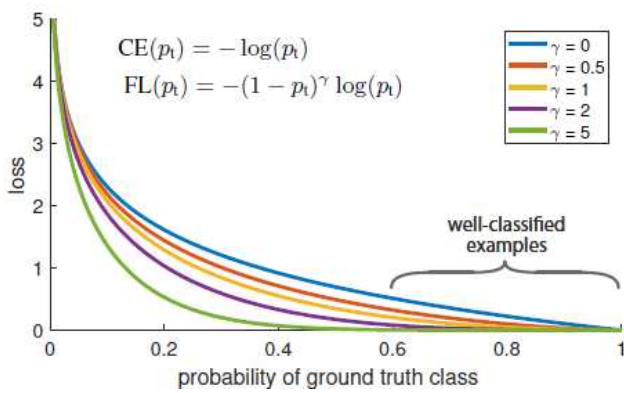


그림 2 Focal Loss

본 논문에서는 데이터셋으로 신분증이 포함된 사진을 이용하여 직접 라벨 작업을 통하여 255와 0값만을 가지는 흑백 이미지를 생성하였다. 신분증에 해당되는 영역은 255값을 가지는 하얀색으로 변환되며, 신분증이 아닌 배경에 해당되는 영역은 0값을 가지는 검은색 바탕으로 변환된다. 그림3은 원본 이미지와 원본 이미지를 바탕으로 생성된 라벨 이미지이다. 이렇게 생성된 6,000개의 데이터셋을 사용하여 실험을 진행하였으며, 환경 설정은 아래 표 1과 같다.

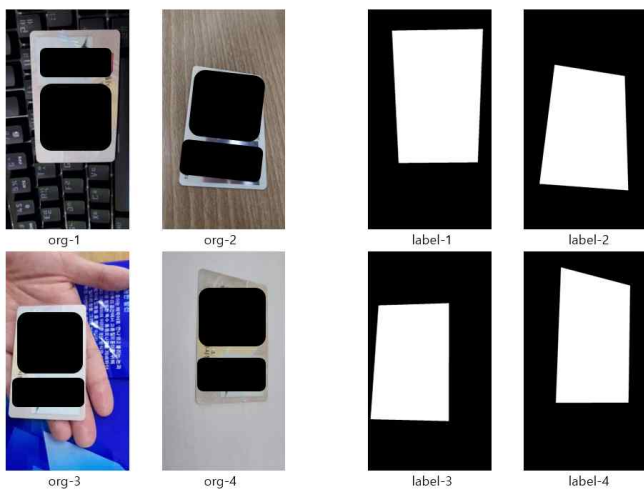


그림 3 원본 이미지 및 라벨 이미지

표 1 학습 환경 설정

테스트 환경	
Optimizer	ADAM
dataset	train:6000 validation:1000
loss	BinaryFocalLoss(gamma=2)
하드웨어 환경	
그래픽카드	RTX3090
CUDA	11.x
tensorflow	2.6.0

위의 데이터와 환경으로 학습을 한 결과, Resnet50을 backbone으로 사

용한 모델이 성능이 가장 좋게 나왔으며 다음으로 VGG16, U-net을 차지하였다. 각 모델별 검증데이터 세트 loss와 accuracy는 다음 표2와 같으며, 이 때 accuracy는 각 픽셀값의 예측 정확도이다.

표 2 학습 결과

	validation loss	validation accuracy
U-net	0.0527	0.9785
VGG16	0.0080	0.9854
RESNET50	0.0048	0.9905

이를 바탕으로 딥러닝을 사용하지 않고 Gaussian과 Canny 등의 알고리즘을 이용하여 이미지 내에서 신분증의 경계선을 검출하는 방식과 딥러닝을 사용하여 이미지 내에서 신분증을 검출하는 방식을 비교하였을 때, 검은색 배경 등 신분증과 배경이 명확히 구분되는 배경에서는 기존 방식도 검출하지만, 배경이 조금만 다양해질 시 딥러닝을 사용하는 방식이 성능이 상당히 우수하였다. 그 결과 비교는 다음 그림 4와 같다.

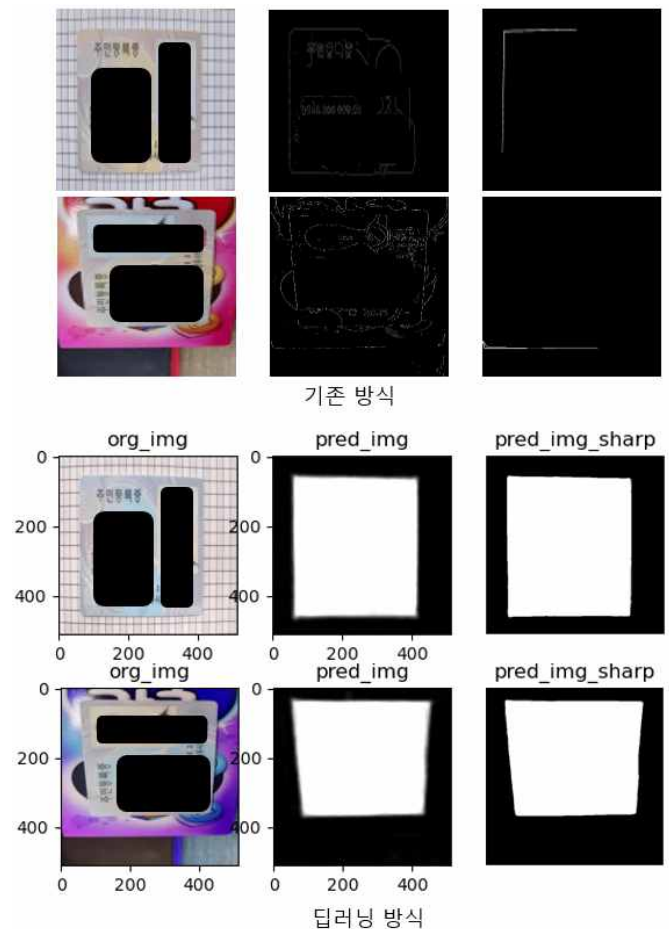


그림 4 기존 방식과 딥러닝 방식 결과

### III. 결론

본 논문에서는 U-Network기반 Segmentation Network를 활용한 이미지 내에서의 객체 검출 기법을 바탕으로 다양한 배경을 가지는 사진에서 신분증을 검출하는 네트워크를 구성하였으며, 각각 기본 U-Net, VGG16 기반 U-net, Resnet50 기반 U-net을 활용하여 실험을 진행하였으며,

Resnet50 기반 U-net이 성능이 가장 높게 나타나는 것을 확인하였으며, 딥러닝을 사용하지 않는 기존 방식을 이용하여 신분증을 검출하는 방식보다 성능이 좋은 것을 확인하였다.

인터넷, 모바일 관련 카드 및 신분증을 활용한 작업이 증가함에 따라 본 방식을 기반으로 이미지 내에서 신분증을 검출한다면, 그 이후 보안 등의 신분증 관련 작업에 다양하게 활용될 것이다.

## ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2022-2020-0-01462) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

## 참 고 문 헌

- [1] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, May 2015.
- [2] Karen Simonyan, Andrew Zisserman, Very Deep Convolutional Networks For Large-Scale Image Recognition, April 2015
- [3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Microsoft Research, Deep Residual Learning for Image Recognition, December 2015.
- [4] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollar, Facebook AI Research(FAIR), Focal Loss for Dense Object Detection, February 2018.