

# 딥러닝 기반 물체 감지 모델을 활용한 손 글씨 감지 기법

곽철영, 박세웅 (서울대학교)

cykwak@netlab.snu.ac.kr, sbahk@snu.ac.kr

## I. 서론

손 글씨 인식은 다양한 종류의 어플리케이션에 활용될 수 있다. 손 글씨 인식을 위해서는, 이미지 속 손 글씨의 위치를 정확히 파악할 필요가 있다. 최근, 딥러닝 기반의 이미지 속 물체를 빠르고 정확하게 감지할 수 있는 기법이 많이 제안되었다. 이미지 속 손 글씨를 감지 할 대상으로 간주한다면, 마찬가지로 딥러닝 기반의 물체 감지 모델을 통해 이미지 속 손 글씨의 위치를 파악할 수 있다. 본 논문에서는 물체 감지 모델을 통해 손 글씨 인식을 하는 방법을 제안한다. 모델 학습에 필요한 데이터의 레이블을 가공하는 방법과, 최종적으로 학습된 모델의 감지 결과를 제시한다.

## II. 본론

### 1. 물체 감지 모델

최근 이미지 속 물체 감지를 위한 딥러닝 기반의 물체 감지 모델이 제안되고 있다. 최근 제안된 SSD(Single Shot Multibox Detector) [1]는 다양한 크기의 anchor box를 활용하여 이미지 속 물체를 빠르고 정확하게 검출할 수 있는 물체 감지 모델이다. 본 논문에서는 SSD를 손 글씨 검출을 위한 모델로 사용하였다.

### 2. IAM 손 글씨 데이터

물체 감지 모델을 학습시키기 위해서는, 손 글씨의 레이블이 있는 데이터가 필요하다. IAM 손 글씨 데이터는 [2], 다양한 종류의 사람이 작성한 손 글씨 데이터를 제공하고 있다.

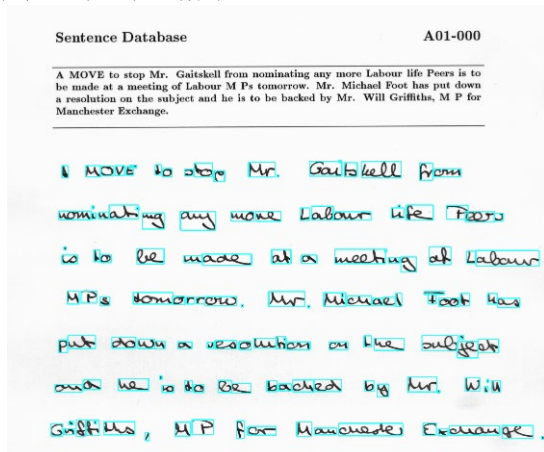


그림 1. IAM 손 글씨 데이터 예시

위 그림과 같이, IAM 손 글씨 데이터 속에는 손 글씨가 있는 곳에 일정 단위로 직사각형 모양의 상자와 상자의 좌표를 레이블로 제공한다. 하지만 그림 2에서 볼 수 있듯이 한 단어임에도 여러 개의 상자로 구성되어 있는 경우가 있다. 이러한 데이터를 학습 데이터로 사용할 경우 모델이 감지할 대상을 하나의 단어임에도 한 개로 인식하지 않고 여러 개로 인식하는 형태의 학습이 일어날 수 있다. 이를 방지하기 위해서는, 학습 데이터에 들어있는 레이블을 단어 단위의 레이블로 재구성 할 필요가 있다. 따라서, 단어에 들어있는 각 데이터를 단어를 포함하는 가장 작은 직사각형 모양의 새로운 상자를 만든 후 그 상자의 좌표를 새로운 레이블로 사용하였다. 레이블 가공 방법에 대한 예시는 그림 2와 같다.

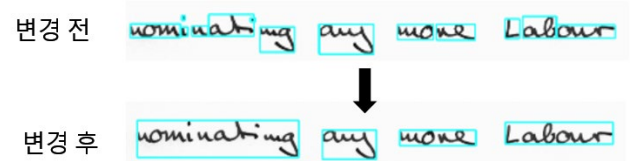


그림 2. 학습 데이터 레이블 가공 예시

최종적으로 class는 손 글씨가 있는 경우(text)와 없는 경우(Background)로 구분하였다.

### 3. 실험 결과

IAM 손 글씨 데이터 1,372 개를 학습 데이터로 사용하였다. SSD의 Backbone 네트워크는 ImageNet으로 사전에 학습된 값을 사용하였고, 나머지 모델의 가중치(weight)는 IAM 손 글씨 데이터를 이용하여 학습하였다. 최종적으로 학습된 모델을 사용한 검출 결과의 예시는 아래 그림과 같다.

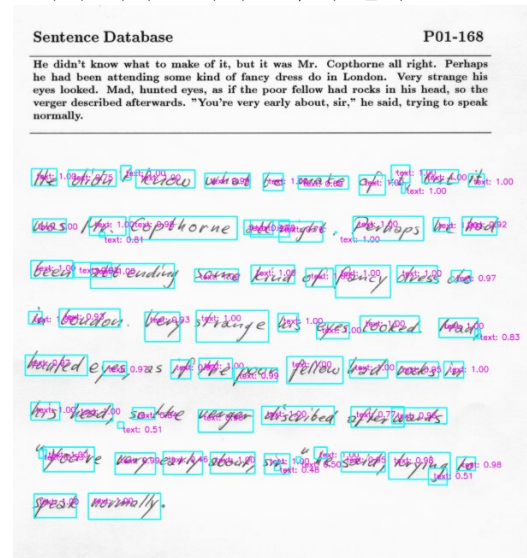


그림 3. 검출 결과 예시

학습 데이터로 사용한 정답 레이블과 유사하게, 단어 단위의 손 글씨를 검출 하는 방식으로 동작하는 것을 확인할 수 있었다.

## III. 결론

본 논문에서는 딥러닝 기반 물체 감지 모델을 통해, 이미지 속 손 글씨 감지 방법을 제안하였다. 학습에 필요한 IAM 손 글씨 데이터의 레이블 가공방법을 제안하고, 학습된 모델의 검출 성능을 확인하였다.

## ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2017R1E1A1A01074358).

## 참고 문헌

- [1] Liu, Wei, et al., "SSD: Single shot multibox detector," in *Proc. ECCV*, 2016.
- [2] U. Marti and H. Bunke, "The IAM-database: An English Sentence Database for Off-line Handwriting Recognition," *Int. J. Document Analysis and Recognition*, vol. 5, pp. 39-46, 2002.