

# 딥러닝 보조 레이블링 작업 시스템

이완해, 정형구  
 서울시립대학교 컴퓨터과학부  
 e-mail : {train96, hjung}@uos.ac.kr

## DALT: Deep Learning Assisted Labeling Task

Wanhae Lee, Hyunggu Jung  
 Department of Computer Science and Engineering, University of Seoul

### 1. 서론

지도학습을 사용하여 머신러닝 모델을 개발할 경우 학습을 위한 레이블링 데이터가 필수적으로 필요하다. 일반적으로 하나의 모델을 완성하기 위해 필요한 데이터의 개수가 몇 개인지는 알려져 있지 않지만, ILSVRC<sup>1)</sup> 2012에서 최고 성능을 보여준 것으로 알려진 AlexNet의 경우 학습을 위해 1500만개 이상의 레이블 된 데이터를 이용한 것으로 알려져 있다 [1].

하지만 이러한 레이블 된 데이터를 얻기 위한 레이블링 과정은 현재 순수하게 사람의 능력에 의존하고 있다. 대표적으로 앞서 예시로 든 AlexNet에서 데이터 셋으로 활용한 ImageNet의 경우 사진을 레이블링하기 위하여 크라우드소싱기법을 사용한 것으로 알려져 있다 [2].

본 논문에서는 이러한 레이블링 과정을 개선시키기 위해 실시간으로 레이블링 된 데이터를 기반으로 학습을 진행하고 학습된 모델을 통해 바로 레이블 되어야할 데이터의 추천 선택지를 보여주는 시스템인 DALT를 제안한다. 본 시스템은 레이블링 작업자의 의사결정에 걸리는 시간을 최소화하고 나아가 레이블링 작업의 효율성을 높여줄 것으로 기대된다.

### 2. 시스템 설계 및 구현

본 시스템은 Docker를 활용한 웹 서비스 환경으로 구성되어 있다. 그림 1에 제시된 바와 같이 다섯 개 부분으로 구성된 본 시스템을 통해 사용자는 웹서버에 접속하여 레이블링 작업을 진행하며 그 결과는 실시간으로 데이터베이스에 저장된다.

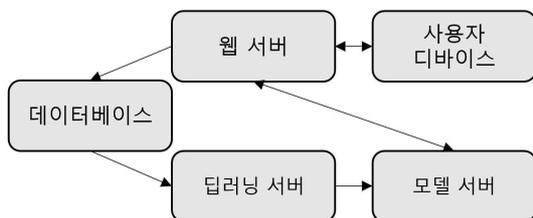


그림 1. DALT 시스템 개요도

이때 딥러닝 서버는 데이터베이스에서 주기적으로 학습할 데이터를 가져와 학습을 진행하게 된다. 추가적으로, 데이터베이스에는 원본 데이터 뿐만이 아닌 Auto Encoder [3] 기법을 사용하여 추출해낸 압축된 특징 데이터가 미리 저장되어 있어 원본 데이터를 학습하는 것보다 더 빠르고 효율적인 실시간 학습을 가능하게 한다. 학습된 모델은 주기적으로 모델서버에 업로드되며 사용자는 모델서버의 예측결과가 옳은지 옳지 않은지만 판단을 해주면 된다.

### 3. 결론 및 향후 연구

지금까지 DALT에 대해서 살펴보았다. 본 시스템은 실시간으로 학습을 진행하며 앞으로 진행되어야 할 레이블링 작업을 보조해 주며, 비지도 선행학습을 통하여 실시간 컴퓨터자원 사용량을 최소화 하였다. 그러나 이것이 실제 레이블링 작업에 도움이 되는지는 아직 검증되지 않았으므로, 후속 연구로는 사용자가 레이블링 작업하는데 있어 본 시스템이 작업효율성에 얼마나 유의미한 영향을 끼쳤는지 평가해보는 것이 될 것이다.

### ACKNOWLEDGEMENT

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단 -현장맞춤형 이공계 인재양성 지원사업의 지원을 받아 수행된 연구임(NRF-2017H1D8A1030582).

### 참고문헌

[1] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.  
 [2] Fei-Fei, Li. "ImageNet: crowdsourcing, benchmarking & other cool things." CMU VASC Seminar. Vol. 16. 2010.  
 [3] Hinton, Geoffrey E., and Richard S. Zemel. "Autoencoders, minimum description length and Helmholtz free energy." Advances in neural information processing systems. 1994.

1) ImageNet Large Scale Visual Recognition Challenge