

# 딥러닝 기반 집-나무-사람 검사 분석 모델의 개발

조승계, 조건우, 김영욱  
알고리마

[sjcho@algorima.io](mailto:sjcho@algorima.io), [geonwoocho@algorima.io](mailto:geonwoocho@algorima.io), [ywkim@algorima.io](mailto:ywkim@algorima.io)

## Development of Deep Learning-Based House-Tree-Person Test Analysis Model

Seung-Je Cho, Geon-Woo Cho, Young-wook Kim  
Algorima

### 요 약

심리학에서 사람의 심리 상태를 알아보기 위해 사용되는 검사 방법 중, 집-나무-사람 검사(HTP Test)는 피실험자가 그린 집, 나무, 사람을 포함하는 그림을 사용하여 피실험자의 심리를 분석하는 투영 검사법이다. 본 논문에서는 딥러닝 모델을 이용해 HTP Test에 사용되는 그림을 분석하는 시스템을 제안하며, 성능 평가를 통해 심리학에서의 딥러닝 모델 적용 가능성을 확인한다. 또한 그림 데이터 분석에 적합한 사전 훈련 모델을 개발하기 위해, ImageNet과 스케치 데이터셋으로 사전 훈련하여 성능을 비교한다. 본 논문에서 제안하는 시스템은 크게 감정 분석을 위한 이미지 객체 추출부, 추출된 객체로 피실험자의 감정을 분류하는 감정 분류부로 구성되어 있다. 객체 추출과 이미지 분류 모두 CNN(Convolution Neural Network) 기반의 딥러닝 모델을 사용하며, 이미지 분류 모델은 서로 다른 데이터셋으로 모델을 사전 훈련한 후, 훈련 데이터셋으로 전이 학습하여 모델의 성능을 비교한다. 그림 심리 분석을 위한 HTP test 스케치 데이터셋은, HTP Test와 동일하게 피실험자가 3개 클래스의 집, 나무, 사람의 그림을 그려 자체 수집하였다.

### 1. 서론

딥러닝 기술은 의료, 금융, 보안, 화학 등과 같은 여러 산업에서 사용되면서 그 성능을 입증해오고 있다. 하지만, 심리학에서 딥러닝을 적용하는 것은 어려움이 있는데, 이는 심리학의 학문적 특성 때문이기도 하다. 딥러닝은 사람이 어떠한 행동을 했을 때 그 사람의 다음 행동을 예측하는 것에 능하지만, 왜 그러한 행동을 취했는지에 대해서는 설명하지 못하는 것에 비해, 심리학은 예측보다는 사람의 심리 구조나 과정에 대해 이해하는 것을 더 중요하게 여기기 때문에 [1] 심리학에 딥러닝이 적용되기에 어려움이 있어 다른 분야들에 비해 관련 연구가 활발하지 않다.

그럼에도 불구하고, 심리학에 딥러닝 기술을 접목하려는 시도가 있으며 일례로, 집과 나무와 사람 그림을 그린 후 이를 분석하여 피실험자의 심리를 파악하는 검사법인 집-나무-사람 검사[2](House-Tree-Person test: HTP test)에 딥러닝 모델을 사용하는 연구들이 있다. [3][4]

본 논문에서는 HTP test에서 요구되는 집, 나무, 사

람을 그린 그림에서 피실험자의 감정이 가장 뚜렷하게 나타나는 사람의 얼굴을 객체로 추출하는 객체 인식 과정을 거치고, 추출된 객체를 통해 감정을 분류하여 피실험자의 심리 상태를 판단하는 시스템을 제안한다. 감정은 기쁨과 행복, 분노와 화남, 우울과 슬픔의 3가지 클래스로 분류된다.

또한 감정 분류 모델은 ImageNet[5], TU-Berlin 데이터셋[6], Quick Draw 데이터셋[7]의 3가지의 데이터셋을 사용하여 사전 훈련된 모델을 backbone으로 사용하여 감정 분류 성능을 비교 실험한다.

### 2. 선행 연구

#### 2.1 객체 인식 모델

이미지 분류를 수행하는 딥러닝 모델이 발전함에 따라 객체 인식 모델 또한 꾸준히 발전하고 있다.

객체 인식 모델은 CNN 기반의 모델이며 크게 1-Stage Detector, 2-Stage Detector로 구분된다.

1-Stage Detector의 대표적인 모델인 YOLO[8]는 클래스 분류(Class Classification), 지역 제안(Region Proposal)을 동시에 수행하기 때문에 2-Stage Detector

모델에 비해 수행 속도가 빠르지만, 상대적으로 성능이 좋지 못하다. 이를 개선하기 위해 꾸준한 연구가 이루어지고 있으며, 현재 YOLO v5[9] 모델이 개발되어 계속해서 발전을 이루고 있다.

### 2.2 이미지 분류 모델

CNN 은 1989 년 LeCun 에 의해 처음 소개되었다.[10] 이후 컴퓨터 비전에서 이미지 분류를 위해 CNN 을 기반으로 모델을 개발하였고, 현재 ImageNet 데이터셋의 이미지 분류를 수행하는 모델은 대부분 CNN 모델을 기반으로 연구되었으며, VGG16[11], ResNet[12], EfficientNet[13]와 같은 CNN 기반의 모델이 State-of-the-art 를 달성하며 그 성능을 계속해서 입증해왔다.

이 중 ResNet 은 Residual Block 을 사용하여 모델의 성능을 개선한 모델이며[12], 현재까지도 ResNet 모델을 기반으로 성능을 높이기 위한 연구가 진행되고 있으며[14], 모델의 크기에 따라 ResNet-18 부터 ResNet-200 까지 다양한 모델이 있다.

### 2.3 스케치 데이터셋

스케치는 선사 시대부터 시각적인 것을 그리기 위해 사용되었고, 아마 모든 사람이 가장 쉽게 사용할 수 있는 렌더링 기법일 것이다. [6] 대부분의 사람은 사물을 스케치 할 때, 사물의 특징을 강조하여 그리기 때문에 알고리즘은 실제 스케치 데이터셋을 통해 학습하여야 하고, 이를 위해 스케치 데이터셋이 구축되었다.

본 논문에서는 스케치 데이터셋을 사용하여 사전 학습한 모델과, 실제로 존재하는 이미지들로 구성된 ImageNet 으로 사전 훈련된 모델의 성능을 비교하기 위해 TU-Berlin 데이터셋, Quick-Draw 데이터셋의 스케치 데이터셋을 사용하였다.



(그림 1) 왼쪽부터 ImageNet, TU-Berlin 데이터셋, Quick Draw 데이터셋이다.

Quick Draw 데이터셋은 TU-Berlin 데이터셋에 비해 노이즈가 많다[15]. 또한 TU-Berlin Dataset 은 250 개의 클래스가 있으며 각 class 당 80 장의 데이터로 이루어져 있고, Quick Draw 데이터셋은 345 개의 클래스가 있으며 약 5 천만장의 데이터로 구성되어 있다.

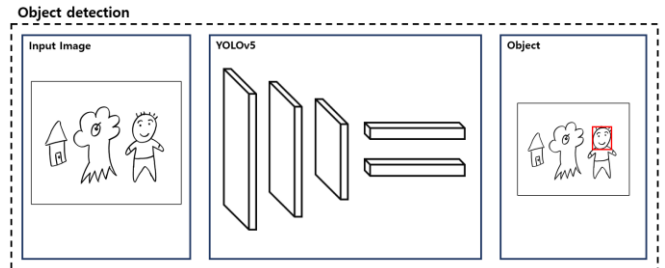
### 3. 전체 시스템 구성

본 논문에서, 우리는 피실험자가 HTP test 를 수행한 결과를, 딥러닝 모델만을 사용하여 피실험자의 심리 상태를 분류하는 시스템을 제안한다. 피실험자가 그린 집, 나무, 사람 그림은 딥러닝 모델의 입력으로 사용된다.

우리가 제안하는 시스템의 구성은 크게 2 단계이며 각 단계에서는 CNN 기반의 모델이 사용된다. 첫 번째 단계인 객체 추출부에서는 객체 인식 모델을 이용해, 입력된 그림에서 사람의 얼굴 객체를 추출하고, 두 번째 단계인 감정 분류부에서는, 첫 번째 단계에서 추출한 객체를 이미지 분류 모델을 사용해 심리 상태를 분류한다.

### 3.1 객체 인식 및 추출 (객체 추출부)

피실험자의 심리 상태를 파악하기 위해 우리는 첫 번째로 그림에서 심리 상태를 분류하기 위한 정보를 가지고 있는 객체를 추출한다. 이를 위해 그림 2 와 같은 객체 인식 모델을 사용하는데, 객체가 어떤 class 인지 분류하는 단계와 어느 위치에 객체가 위치하고 있는지에 대한 바운딩 박스(bounding box)의 좌표를 찾는 지역 제안 단계를 동시에 수행하는 1-Stage Detector 모델인 YOLOv5 모델[9]을 사용하였으며, ImageNet 으로 사전 훈련된 모델을 전이 학습 하여 모델을 개발하였다.



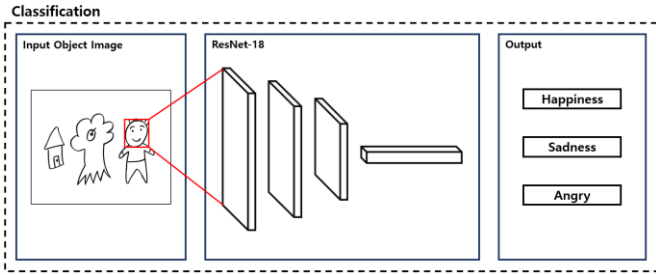
(그림 2) 객체 추출부. 이미지에서 사람의 얼굴 객체가 추출된다.

### 3.2 이미지 분류 모델 (감정 분류부)

일상생활에서 볼 수 있는 실제 이미지를 분류하기 위한 모델을 개발할 때 ImageNet 으로 사전 훈련된 모델을 가져온 후, 전이 학습을 통해 해결하고자 하는 문제에 적합한 모델을 만든다.[16][17][10] 하지만, 본 논문에서는 전이 학습을 위해 사용하는 데이터는 피실험자가 HTP test 를 진행하며 그린 스케치 데이터이기 때문에 스케치 데이터셋을 사용하여 사전 훈련 후 전이 학습한 모델과 실제 사진 데이터셋으로 사전 훈련 후 전이 학습한 모델과 비교한다.

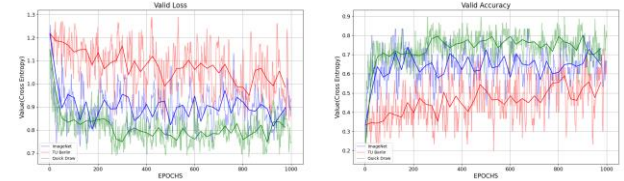
스케치 데이터셋은 TU-Berlin 데이터셋[6]과 Quick-Draw 데이터셋[7]으로 사전 훈련 후 전이학습 하여 스케치 데이터셋의 크기와 특성에 따른 성능 차이를 비교하며, 실제 사진 데이터셋은 ImageNet 을 사용하여 사전 훈련 후 전이 학습하여 3 개의 모델을 비교 실험한다.

우리는 객체 인식 모델로 추출된 객체를 통해 피실험자의 심리 상태를 파악하기 위해 그림 3 과 같이 이미지 분류 과제에서 베이스라인 모델로 자주 사용되는 CNN 기반의 이미지 분류 모델인 ResNet-18 모델[16]을 사용한다.



(그림 3) 감정 분류부. 객체 추출부에서 추출된 객체를 ResNet-18 모델을 통해 감정을 분류한다.

= 1e-6, weight decay = 1e-6, epoch = 1000 으로 설정하였다.

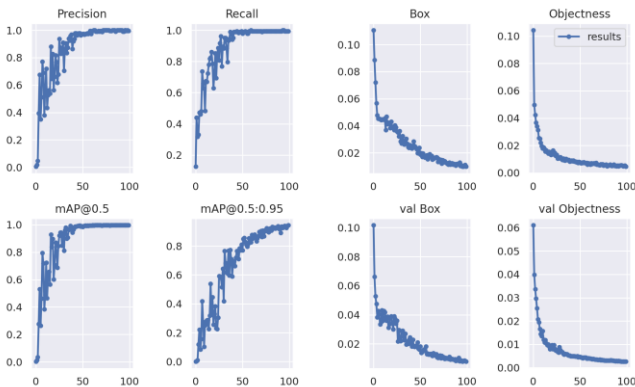


(그림 5) 감정 분류 모델 훈련 시, 테스트 데이터셋의 Loss, Accuracy 그래프

## 4. 실험

### 4.1 객체 인식 모델 구현

객체 인식 모델 훈련을 위해 ImageNet 으로 사전 훈련된 YOLOv5 모델을 HTP test 스케치 데이터셋으로 훈련하였고, HTP test 스케치 데이터셋은 이미지의 크기가 가로, 세로 각각 2000 픽셀 이상이며 224 장의 데이터로 구성되어 있다. 이 중 80%는 훈련 데이터셋 (182 장), 나머지 20%는 테스트 데이터셋(42 장)으로 분류하여 전이 학습을 하였다. 데이터셋의 이미지 크기가 2000 이상인 점을 고려하여, 전이 학습 시 이미지 크기는 2000, 에포크는 100 으로 지정하였다. 하나의 이미지에 여러 객체가 탐지될 경우, confidence 값이 가장 높은 객체만을 선택하여 추출한다. 모델의 훈련 결과는 그림 4 와 같다.



(그림 4) YOLOv5 모델 훈련 결과. (좌) Bounding Box, Object Classification 의 Loss 그래프. (우) Precision, Recall, mAP 그래프

### 4.2 감정 분류 모델 구현

감정 분류를 위해 ImageNet 사전 훈련 모델, TU-Berlin 사전 훈련 모델, Quick Draw 사전 훈련 모델을 각각 전이 학습하여 비교 실험하였다.

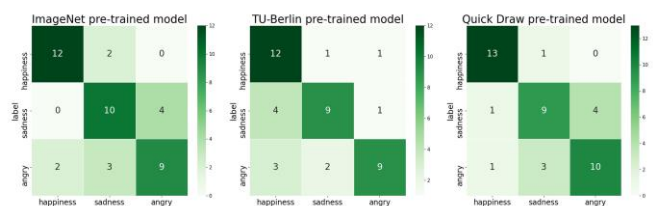
이미지 분류 모델은, HTP test 스케치 데이터셋에 YOLOv5 모델을 사용하여 추출된 객체들을 사용하여 전이 학습하였다. 이미지 분류 모델 훈련을 위해 HTP test 스케치 데이터셋의 사람 얼굴 그림에서 감정이 뚜렷하지 않은(무표정) 데이터는 제외한 뒤 이미지 분류 모델을 위한 데이터셋을 새롭게 구축하였다. 이미지 분류 모델을 위한 데이터셋의 80%는 훈련 데이터셋, 나머지 20%는 테스트 데이터셋으로 분류하였고 image size = 224, optimizer = Adam, scheduler = CosineAnnealingLR, learning rate = 1e-3, min learning rate

3 가지 사전 훈련 모델을 비교 실험한 결과, 그림 5 에서 볼 수 있듯, TU-Berlin 데이터셋으로 사전 학습한 모델을 사용한 경우의 정확도는 0.73 으로 가장 낮았으며, 다음으로 ImageNet 으로 사전 학습한 모델의 정확도가 0.83, Quick Draw 데이터셋으로 학습한 모델의 정확도는 0.89 로 가장 성능이 높았다.

TU-Berlin 데이터셋이 스케치 데이터셋임에도 불구하고 성능이 낮은 이유는 TU-Berlin 데이터셋은 250 개의 클래스로 구성되며 각 클래스 당 80 장의 데이터로 구성되어 있어 데이터셋의 크기가 다른 데이터셋에 비해 작기 때문이다. 그 결과 다른 데이터셋에 비해 사전 학습이 불안정하게 이루어지기 때문에 전이 학습 시 가장 낮은 성능을 보인다. 그에 비해 ImageNet 과 Quick Draw 데이터셋은 사전 학습에 사용되는 데이터셋의 크기가 크기 때문에 안정적으로 사전 학습이 가능하며, 실험 결과와 같이 ImageNet 사전 훈련 모델에 비해 스케치 데이터셋인 Quick Draw 데이터셋으로 사전 학습한 모델을 사용한 경우가 가장 높은 성능을 보였다.

### 4.3 결과

그림 심리 분석 모델의 실험 결과는 그림 6 과 같다. 두 개의 비교 실험 모델 중, 객체 인식 모델은 Yolo v5 모델로 동일하며 감정 분류 모델은 각각 이미지넷 사전 훈련 모델, TU-Berlin 사전 훈련 모델, Quick-Draw 사전 훈련 모델을 비교 실험하였다.



(그림 6) 그림 심리 분석 모델 실험 결과(감정 분류부 모델 비교)

그림 6 에서 볼 수 있듯 TU-Berlin 사전 훈련 모델의 정확도가 0.714 로 가장 낮으며, ImageNet 사전 훈련 모델은 0.738, Quick Draw 사전 훈련 모델의 정확도가 0.761 로 가장 높다.

이는 사전 훈련에 사용한 데이터셋의 크기가 클 경우, 스케치 데이터셋을 사용하여 모델을 사전 훈련하는 것이 그림 데이터를 다루기에 적합하다는 것을 알 수 있으며, 그림 심리 분석에 본 논문에서 제안하는

시스템이 적용 가능함을 알 수 있다.

## 5. 결론

그림 심리 검사 중 하나인 HTP test 에 딥러닝 모델만을 사용하여 심리를 예측함으로써, 심리학에서의 딥러닝 모델 적용 가능성을 확인하였고, 이를 위해 HTP test 데이터셋을 직접 구축하였다. 또한 그림 데이터에서 심리 분석에 필요한 정보를 가지고 있는 객체를 추출하고, 추출된 객체를 통해 피실험자의 심리를 예측하는 2 단계의 과정을 거치도록 전체 모델을 구성하여 그림 심리 분석에 적합한 모델을 제안하였다.

본 논문에서는 ImageNet 으로 사전 훈련된 모델보다, 많은 양의 스케치 데이터로 사전 훈련된 모델을 사용하는 것이 그림 데이터 분석에 있어 더 적합함을 확인하였다. 또한 우리는 ImageNet 으로 사전 학습된 객체 인식 모델을 사용하였고, 심리 분석을 위한 객체는 사람의 얼굴만을 추출하여 훈련하였다. 하지만, 객체 인식 모델 또한 스케치 데이터셋으로 사전 훈련된 모델을 사용하고, 나무와 집, 배경에서도 정보를 추출할 수 있다면 성능이 더욱 향상될 것으로 보인다.

## 참고문헌

- [1] 김청택. "빅데이터를 이용한 심리학 연구 방법." 한국심리학회지: 일반, Vol. 38, No. 4, pp. 519-54, 2019.
- [2] J.N. Buck, "The HTP test", Journal of Clinical Psychology, No. 4, pp. 151-159, 1948.
- [3] 김수경, and 유건아. "손그림 자동분석에 기반한 퍼지 추론 심리진단 응용의 개발." 한국디지털콘텐츠학회 논문지, Vol. 22, No. 3, pp. 519-525, 2021.
- [4] Kim, T., et al. "Application of Deep Learning in Art Therapy." International Journal of Machine Learning and Computing 11.6, 2021.
- [5] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, pp. 248-255, 2009.
- [6] Eitz, Mathias, James Hays, and Marc Alexa. "How do humans sketch objects?." ACM Transactions on graphics (TOG), Vol. 31, No. 4, pp. 1-10, 2012.
- [7] Ha, David, and Douglas Eck. "A neural representation of sketch drawings." arXiv preprint arXiv:1704.03477, 2017.
- [8] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779-788.
- [9] Glenn Jocher. Yolov5 in pytorch. <https://github.com/ultralytics/yolov5>, 06 2020. (undefined 28/4/2021 10:16).
- [10] LeCun, Yann, et al. "Backpropagation applied to handwritten zip code recognition." Neural computation 1.4, pp. 541-551, 1989.
- [11] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
- [12] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, pp. 770-778.
- [13] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International Conference on Machine Learning. PMLR, 2019.
- [14] Bello, Irwan, et al. "Revisiting resnets: Improved training and scaling strategies." arXiv preprint arXiv:2103.07579, 2021.
- [15] Xu, Peng, et al. "Sketchmate: Deep hashing for million-scale human sketch retrieval." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, pp. 8090-8098.
- [16] Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." IEEE Transactions on knowledge and data engineering 22.10, pp.1345-1359, 2009.
- [17] Chu, Brian, et al. "Best practices for fine-tuning visual classifiers to new domains." European conference on computer vision. Springer, Cham, 2016, pp. 435-442.