

# Two-Stage Deep Learning Based Algorithm for Cosmetic Object Recognition

Jongmin Kim · Daeho Seo<sup>†</sup>

Dageom, CO. LTD

## 화장품 물체 인식을 위한 Two-Stage 딥러닝 기반 알고리즘

김종민 · 서대호<sup>†</sup>

다겸 주식회사

With the recent surge in YouTube usage, there has been a proliferation of user-generated videos where individuals evaluate cosmetics. Consequently, many companies are increasingly utilizing evaluation videos for their product marketing and market research. However, a notable drawback is the manual classification of these product review videos incurring significant costs and time. Therefore, this paper proposes a deep learning-based cosmetics search algorithm to automate this task. The algorithm consists of two networks: One for detecting candidates in images using shape features such as circles, rectangles, etc and Another for filtering and categorizing these candidates. The reason for choosing a Two-Stage architecture over One-Stage is that, in videos containing background scenes, it is more robust to first detect cosmetic candidates before classifying them as specific objects. Although Two-Stage structures are generally known to outperform One-Stage structures in terms of model architecture, this study opts for Two-Stage to address issues related to the acquisition of training and validation data that arise when using One-Stage. Acquiring data for the algorithm that detects cosmetic candidates based on shape and the algorithm that classifies candidates into specific objects is cost-effective, ensuring the overall robustness of the algorithm.

**Keywords :** Computer Vision, Deep Learning, Object Detection, Classification

### 1. 서 론

온라인, 오프라인 구분 없이 제품을 구매하는 고객은 제품을 언제, 어디서, 어떻게 등 다양한 요소를 고려한다. 이는 구매자가 물건을 구매했을 때 올바른 구매로 인식되기 위함이다. 그래서 구매자는 구매 물건의 신뢰할만한 정보를 얻기 위해 실제로 물건을 직접 사용해본 구매자들의 의견을 참고한다[9, 10, 16].

최근에는 인터넷 및 모바일 기술의 발달로 과거 오프라

인 매장이거나 홈쇼핑에서만 물건을 구매할 수 있었던 시기와 다르게 아마존닷컴이나 쿠팡과 같은 온라인에서 물건을 구매할 수 있게 되었다. 특히 1995년에 시작한 아마존닷컴은 온라인에 등록된 판매 제품의 평가 1억 개를 판매 전략으로 활용하여 큰 성공을 거두었다[11]. 또한 83%의 고객이 호텔 예약을 하기 전에 호텔 리뷰를 참고한다고 하였으며 시장조사전문기관 엠브레인 트렌드모니터(2015)에서도 73.8%의 응답자가 광고보다 평가 글 또는 영상을 더 신뢰한다고 하였다.

유튜브에는 다양한 콘텐츠의 영상들이 있어 유튜브 정보를 찾는 빈도가 증가하고 있다[3]. 동영상 플랫폼인 유튜브는 온라인 개인방송 서비스 시장에서 가장 큰 비중을 차지하여 많은 기업들의 마케팅 용도로 적극 활용된다.

Received 24 November 2023; Finally Revised 13 December 2023;  
Accepted 13 December 2023

<sup>†</sup> Corresponding Author : seo\_daeho@naver.com

2020년 12월 글로벌 리뷰 플랫폼(global review platform)은 소비자의 제품 평가 동영상은 사진이나 일반 텍스트보다 더 많은 정보를 신뢰성있게 전달할 수 있다는 이유로 소비자가 제품 구매 전 동영상을 많이 참고한다고 밝혔다. 특히 패션 및 뷰티 제품의 경우 조회수가 1억 건이 넘는 것도 존재한다.

회사마다 판매되는 제품이 다양할 뿐만 아니라 시간 흐름에 따라 판매되는 제품이 변경되고 그로 인해 유튜브에 등록된 다양한 물체의 평가 영상의 개수가 역동적으로 변한다. 그 결과 현재 수작업으로 이루어지는 특정 제품의 평가 영상 검색에 많은 노동과 시간이 소요된다. 따라서 본 논문은 물체 검색 작업의 자동화를 가능케 하는 화장품 인식 알고리즘을 제안한다.

화장품 인식 알고리즘은 Two-Stage로 구성된다. 첫 번째는 모양 특성 기반으로 영상 내의 화장품 후보군을 검출한다. 두 번째는 각 화장품 후보 영역을 분류 모델에 입력하여 판매되는 화장품의 여부를 판단한다. 알고리즘을 Two-Stage로 구성한 이유는 One-Stage 모델 사용시 발생하는 학습 데이터 구축 문제 때문이다. 유튜브에 등록된 화장품 평가 영상을 살펴보면 촬영 장소, 평가자, 조명 등 다양한 조건에서 촬영된다. 또한 각 제품의 평가 영상이 거의 존재하지 않거나 없는 경우가 많아 One-Stage 모델의 학습 데이터 구축을 어렵게 만든다. 따라서 본 논문은 모양 특성기반의 후보군 검출 모델과 화장품 분류 모델로 구성된 Two-Stage 모델을 제안한다. 모양 특성을 활용한 후보군 모델 데이터는 유튜브 화장품 평가 영상으로 구축되며, 화장품 분류 모델의 데이터는 실제 제품의 촬영으로 구축된다. 일반적으로 클래스(class) 정보만 존재하는 분류 모델 데이터 구축 비용이 검출 모델 데이터 구축 비용보다 상대적으로 낮다.

본 논문의 구성은 다음과 같다. 제2장에서는 본 연구와 관련된 선행 연구를 살펴본다. 제3장에서는 제안한 모델을 효율적으로 학습하기 위해 전이 학습 기법을 적용 및 분석한다. 제4장에서는 제안된 화장품 인식 알고리즘의 성능을 검증하고 마지막 5장에서는 본 논문의 내용을 정리하고 향후 연구 방향을 제시한다.

## 2. 이론적 배경

본 연구는 화장품 인식 알고리즘의 검출 네트워크와 분류 네트워크에 YOLO(You Only Look Once)[2, 13, 14]와 EfficientNet[17, 18]을 사용하였다. YOLO는 One-Stage 네트워크로서 객체를 빠르게 검출할 수 있으며 영상을 한번에 바라보는 형식으로 구성되어 있어 거짓 양성(False Positive)이 발생할 확률이 상대적으로 적다. 또한 분류 모

델에는 ResNet[6], MobileNet[8, 15], EfficientNet 등의 다양한 모델이 있지만 그 중 EfficientNet은 네트워크 구조의 효율성에 초점을 두어 파라미터 크기 대비 높은 성능을 갖는다. 주된 이유는 NAS(Network Architecture Search) [19] 기반으로 구성된 Baseline 네트워크와 많은 실험으로 얻은 모델의 깊이, 너비 그리고 입력 영상 크기 간 관계식 때문이다.

### 2.1 YOLO(You Only Look Once)

검출 네트워크로 YOLO가 사용되었다. YOLO는 One-Stage 검출기로 빠른 속도와 상대적으로 높은 정확도를 가지고 있다. YOLO는 이미지로부터 N by M 격자 모양을 띤 특징 맵을 추출하고 각 격자 위치로부터 물체 유무, 물체 종류 그리고 영상에 투영된 물체 크기를 동시에 예측한다. 이는 기존 two stage detector보다 짧은 시간으로 물체를 검출할 수 있도록 한다.

YOLO 네트워크는 크게 24개의 컨볼루션 레이어(Convolution Layer)로 구성되어 있다. 앞 20개의 네트워크 초기 파라미터는 ImageNet 데이터 셋[4]으로 훈련된 파라미터(parameter)로 설정되고 뒤 4개의 컨볼루션 모듈과 완전 연결 계층(Fully Connected Layer) 파라미터는 본 연구에서 구축한 화장품 데이터 셋의 훈련으로 설정된다.

### 2.2 ResNet

ResNet은 네트워크의 깊이 증가로 발생하는 기울기 소실(gradient vanishing) 문제를 잔차 맵핑(residual mapping)으로 해결한다. 잔차 맵핑은 앞 단 레이어(layer)로 기울기를 추가로 전달하는 구조로서 기울기 소실 문제를 방지한다. 기존에는 네트워크의 깊이가 커질수록 분류 성능이 떨어지는 모습을 보였지만 잔차 맵핑을 갖는 ResNet은 모델의 깊이와 분류 성능이 비례 관계를 갖는다. 이에 2015년 ILSVRC[9]에서 ResNet-152 모델로 top-5 test error 3.57%로 우승한다.

### 2.3 MobileNet

MobileNet은 경량 네트워크로 모바일 및 에지 기기에서 효율적인 실시간 이미지 분류 및 객체 감지를 위해 제안되었다. MobileNet은 공간과 채널을 분리하여 컨볼루션 필터(convolution filter)를 적용한다. 공간은 'depthwise separable convolution filter' 그리고 채널은 'pointwise convolution filter'을 적용하며 이 두 연산은 연속적으로 수행된다. 필터(filter) 크기가 3인 경우 연산량이 8~9배나 줄어든다.

## 2.4 EfficientNet

EfficientNet은 최적의 모델 깊이, 너비 그리고 입력 이미지 크기로 설정되어 다른 유사한 크기의 모델보다 높은 분류 성능을 가진다. 깊이, 너비 그리고 이미지 크기 3가지 요소의 균형을 맞추어 조절할 수 있는 Compound Scaling 방법과 이 방법에 적합한 네트워크를 설계하였다.

총 8개 형태의 EfficientNet ‘B0~B7’이 있으며 단계가 올라갈수록 모델의 크기가 크고 성능도 함께 올라간다. 각 단계 별 유사 크기를 갖는 다른 네트워크 간 성능 비교에서 더 높은 수치를 갖는다.

## 2.5 전이 학습(Transfer Learning)

Transfer Learning은 기존 과제 데이터 셋(Source Task Dataset)으로 학습된 파라미터를 목표 과제 데이터 셋(Target Task Dataset)에 맞게 변형시키는 기법이다. 그래서 기존 과제 데이터 셋과 타겟 과제 데이터 셋의 유사도나 데이터 양 등에 따라 네트워크의 성능이 달라질 수 있다. Azizpour et al.[1]은 기존 과제 데이터 셋의 특성에 따라 목표 과제의 성능이 크게는 10% 차이가 나는 것을 실험 결과로 제시하였다.

또한 전이 학습은 네트워크 학습 속도를 단축시키는 효과가 있다. Ying et al.[19]의 연구 결과에 따르면 동일 조건에서 랜덤으로 모델을 초기화하여 학습을 시작하는 것보다 사전 학습된 모델의 파라미터 값으로 초기화하면 더 빠르게 높은 정확도로 도달할 수 있으며 크게는 2~3배의 학습 시간을 단축될 수 있다.

## 3. 연구 방법

본 연구에서 제안한 방법은 <Figure 1>처럼 검출 모델과 분류 모델로 구성된 Two-Stage 기반 화장품 인식 알고리즘이다. 첫 번째는 직사각형, 원 등과 같이 모양 특

성을 이용하여 영상에서 화장품 후보군을 검출한다. 그리고 분류 모델로 하여금 원(original) 영상에서 잘라낸 각 후보 영역을 입력으로 목표(target) 화장품 여부를 판단하게 한다.

### 3.1 YOLO을 활용한 화장품 후보군 검출

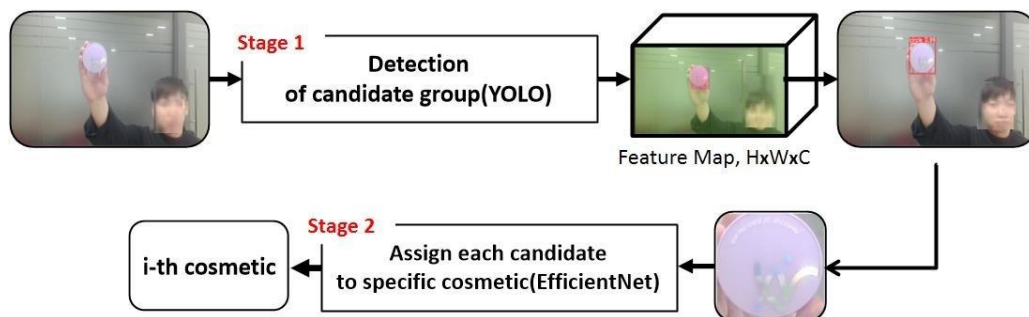
후보군 검출 네트워크로 YOLO를 사용하였다. YOLO는 높이(height), 너비(width) 그리고 채널(channel) 모양을 띤 특징 맵(feature map)을 추출한다. <Figure 1>의 경우 YOLO에서 추출된 특징 맵의 높이와 너비는 각각 4와 8이며, 화장품이 포함된 특징 그리드가 빨간색으로 표시되었다. 이는 물체가 존재하는 영역이 주변 영역보다 활성화 정도가 크다는 것을 나타내며 결과적으로 해당 위치에서 관심 물체의 크기, 클래스 정보가 내포되어 있다.

### 3.2 데이터 문제 및 방안 제시

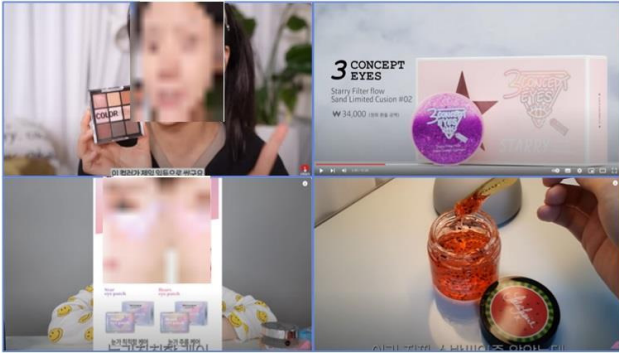
네트워크 성능은 데이터의 양과 질에 의존하므로 물체 특성 및 환경(background) 정의가 중요하다. 제안된 알고리즘은 <Figure 2>처럼 다양한 환경 속에 있는 제품을 검출한다. 하지만 해당 제품에 대한 영상이 거의 존재하지 않아 데이터 구축에 어려움이 있다. 만약 적은 양의 데이터로 네트워크로 학습시키는 경우 과적합(overfitting)이 발생하여 성능이 저하된다.

따라서 본 연구는 One-Stage 대신 Two-Stage 기반의 알고리즘을 제안하였고 첫 번째 단계인 YOLO가 화장품 후보군을 검출한다. <Figure 3>처럼 모든 화장품 물체를 사각형이나 원 등으로 재정의하였고 그 결과 YOLO가 모양 특성을 이용하여 화장품 후보군을 검출한다.

본 논문에서는 서로 다른 30개의 화장품이 직사각형(rectangle), 원(circle), 주둥이 병(snout bottle), 붓(brush)으로 재분류되기 때문에 유튜브에서 화장품 영상을 무작위로 추출하여 4가지 종류의 YOLO 모델 학습 및 검증 데이터를 구축하였다.



<Figure 1> Cosmetics Recognition Algorithm Flow Chart



<Figure 2> Background of Cosmetics Detection



<Figure 3> Shape Based Cosmetic Classification

### 3.3 EfficientNet의 구조 변경 및 학습 방법론

화장품 후보군을 특정 물체로 할당하는 알고리즘은 EfficientNet이다. EfficientNet의 입력 데이터는 <Figure 3>과 같이 YOLO가 검출한 후보 영역이다. EfficientNet의 효율적 학습 및 일반화를 위해 파라미터 초기화, 학습 파라미터 비율 등 해당 값을 그리드 서치(grid search) 방법으로 결정하였다.

모델을 학습시킬 때 모델 일반화와 빠른 학습을 위해 ImageNet과 같은 대용량 데이터로 학습된 파라미터를 모델 초기화에 많이 사용한다. 본 연구 또한 ImageNet으로 훈련된 파라미터를 EfficientNet 파라미터 초기화 값으로 사용하였다. 더불어 특징 추출 부의 전반 레이어를 고정시키고 그 외 파라미터만 훈련시켜 모델의 일반화 성능을 한 층 더 끌어올렸다. <Table 1>는 초기화 방법을 ImageNet으로 학습된 파라미터, kaiming initializer 기법[7] 그리고 xavier initializer 기법[5]을 사용한 경우를 비교한 것이다. 초기화 방법에 따라 성능 차이가 약 47%로 크게 나는 이유는 훈련 데이터 셋이 각 물체 당 약 1000장 정도로 적기 때문이다.

모델 성능 평가 척도는 mAP(mean Average Precision)을

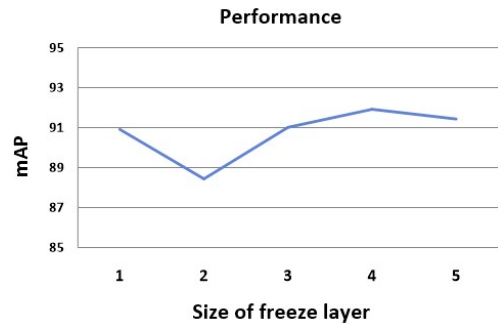
사용하였다. mAP에 대한 것은 실험 결과 부분에 서술하였다.

모델 일반화를 위해 ImageNet으로 학습된 파라미터를 모델 초기화에 적용한 경우 네트워크 앞 단부터 파라미터를 고정시킨다. 본 논문의 경우 <Figure 4>처럼 5가지 경우에 대해 각 모델을 훈련시켜 최적의 고정 파라미터 크기를 정하였다.

<Table 1> Performance of two Model Parameter Initialization Methods

Performance(mean Average Precision)	
xavier initializer	40.403
kaiming initializer	43.910
ImageNet	91.917

<Figure 4>의 세로 항목은 모델의 성능 수치를 나타내며 가로 항목의 각 크기는 파라미터를 고정시킨 MBCConv (mobile inverted bottleneck convolution)의 모듈 개수를 의미한다. 따라서 본 논문은 물체 인식 성능이 가장 우수했던 4개의 MBCConv 모듈 파라미터를 고정시켜 모델 학습을 진행하였다. 참고로 MBCConv는 EfficientNet의 특징 추출부를 구성하는 모듈이다.



<Figure 4> Model Performance by Fixed Parameter Size

## 4. 실험 결과

### 4.1 실험 데이터 구축

실험에 사용된 하드웨어 구성은 13세대 인텔 i7-13700K 3.40 GHz, NVIDIA GeForce RTX 3060, 파이토치 프레임워크(pytorch framework)[12]이다. 그리고 소프트웨어 구성은 Python-3.10.13, Pytorch-1.13.1, cudatoolkit-11.6, opencv-4.6.0, pillow-10.0.1이다.

검출 모델과 분류 모델로 구성된 화장품 인식 알고리즘 훈련을 위해 두 종류의 데이터를 구축하였다. 첫 번째 검출 모델 데이터는 총 500장으로 300장은 훈련용 200장은

검증용이다. 데이터는 유튜브 영상에서 무작위하게 추출되었다. 분류 모델 데이터는 총 54000장으로 훈련용 33000장 검증용 21000장이다. 데이터는 30종류의 화장품을 촬영하여 얻었다.

#### 4.2 물체 검출 모델의 성능 평가

물체 검출 모델의 성능은 mAP(mean Average Precision)로 결정된다. mAP는 정밀도(precision)와 재현율(recall)로 계산되며, 정밀도와 재현율은 아래 수식처럼 TP(True Positive), FP(False Positive), FN(False Negative), TN(True Negative)으로 계산된다.

재현율과 정밀도는 반비례 관계를 지니고 있으므로 두 수치가 함께 고려된 AP(Average Precision)를 물체 종류마다 계산하고 더불어 각 물체 종류의 AP를 평균 내어 모델의 성능 지표인 mAP를 계산한다.

$$precision = \frac{TP}{TP+FP}, recall = \frac{TP}{TP+FN}$$

#### 4.3 분류 모델 비교 실험

분류 모델은 3가지 모델(ResNet-18, MobileNet-V3-Large, EfficientNet)의 비교 실험으로 결정된다. 3가지 모델은 <Table 2>와 같은 파라미터 크기를 가진다.

이 3종류 모델 또한 파라미터 초기화를 ImageNet으로 학습된 파라미터로 설정하였으며 <Table 3>과 같이 각 모델을 구성하는 하부 모듈 단위로 일부 파라미터를 고정시키는 기법도 적용하였다. <Table 3>을 보면 모든 경우에 대해 모듈 3개의 파라미터를 고정시킨 EfficientNet이 가장 우수하였다. <Table 3>의 's'는 모듈 개수를 나타낸다. 각 모델은 서로 다른 모듈로 구성되기 때문에 각 모델의 고정 파라미터 개수를 같게 설정하는 것은 불가능하다. 따라서 파라미터 개수 대신 각 모델을 구성하는 모듈의 개수를 기준으로 비교 실험을 수행하였다.

분석 결과로 <Table 3>을 살펴보면 모델 중 EfficientNet의 성능이 91.917%로 가장 우수하였다. 따라서 본 연구는 EfficientNet을 Two-Stage 화장품 인식 알고리즘으로 채택하였다.

<Table 2> Number of Parameters for Each Model

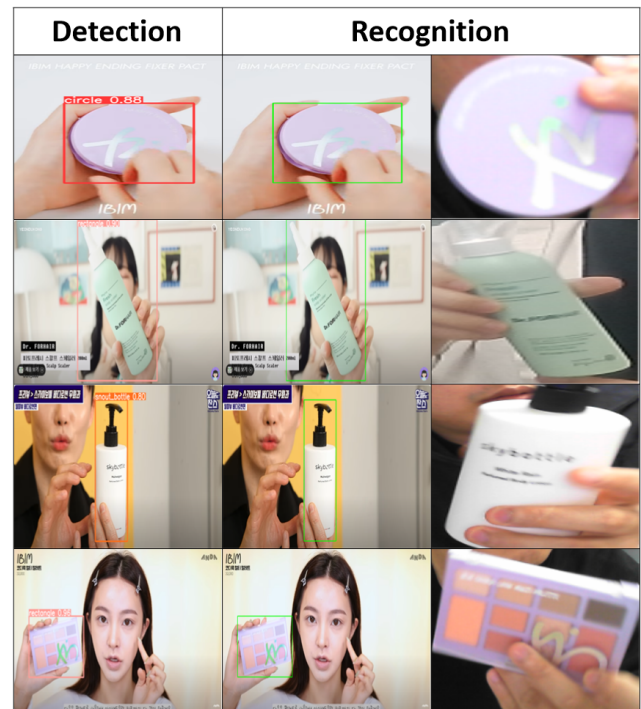
Model	Parameter Size
ResNet-18	9.1M
MobileNet-V3-Large	11.6M
EfficientNet	5.4M

<Table 3> Performance by Model and Fixed Parameters

mAP(%)	s=1	s=2	s=3	s=4
ResNet-18	81.39	87.89	85.97	82.67
MobileNet-V3-Large	87.32	87.69	59.83	87.73
EfficientNet	90.91	88.43	91.01	91.91

#### 4.4 정성적 결과

<Figure 5>는 실제 유튜브 영상에 제안한 알고리즘으로 얻은 결과이다. 왼쪽 첫 번째 열은 YOLO에서 얻은 결과를 나타낸 것이고 두세 번째 열은 EfficientNet이 특정 화장품으로 할당한 결과를 나타낸 것이다. <Figure 5>는 제안한 알고리즘이 유튜브에 등록된 실제 제품 평가 영상에 적용한 결과를 나타내며, 이는 다양한 환경에서 제안한 모델이 화장품을 잘 검출할 수 있음을 의미한다.



<Figure 5> Qualitative Results from YouTube Videos

## 5. 결론

딥러닝 모델 성능에서 데이터의 양과 질은 성능에 많은 영향을 끼친다. 하지만 특정 물체에 대한 유튜브 평가 영상과 같이 데이터가 충분한 양의 데이터를 구축하기가 매우 어려운 경우는 매우 많다. 따라서 본 연구는 상위 수준인 모양으로 물체를 재분류하였다. 그 결과로 YOLO를 이



용하여 화장품 평가 영상으로부터 물체 후보군을 검출하고 해당 후보 영역을 분류 모델인 EfficientNet에 입력하여 특정 화장품으로 할당시켰다. 그리고 정성 및 정량적 결과를 통해 제안한 알고리즘이 효과가 있음을 알 수 있다. 다만 관심 화장품 목록이 주기적으로 변경될 수 있다는 점을 고려하면 제안한 알고리즘의 학습 시간이 길다는 점이 단점이 있다. 따라서 추후 연구에 소량 학습 기법(few-shot learning) 등을 적용하는 연구가 필요하다.

## References

- [1] Azizpour, H., Razavian, A.S., Sullivan, J., Maki, A., and Carlsson, S., Factors of transferability for a generic convnet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, Vol. 38, No. 9, pp. 1790-1802.
- [2] Bochkovskiy, A., Wang, C.Y., and Liao, H.Y.M., Yolov4: Optimal speed and accuracy of object detection, arXiv, 2015, arXiv:2004.10934.
- [3] Chosunbiz, [https://biz.chosun.com/site/data/html\\_dir/2021/04/19/2021041901823.html](https://biz.chosun.com/site/data/html_dir/2021/04/19/2021041901823.html)
- [4] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., and Fei-Fei, L., , Imagenet: A large-scale hierarchical image database, *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, pp. 20-25.
- [5] Glorot, X. and Bengio, Y., Understanding the difficulty of training deep feedforward neural networks, *Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings*, Sardinia, Italy, 2010, pp. 249-256.
- [6] He, K., Zhang, X., Ren, S., and Sun, J., Deep Residual Learning for Image Recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 770-778.
- [7] He, K., Zhang, X., Ren, S., and Sun, J., Delving Deep Into Rectifiers: Surpassing Human-level Performance on Imagenet Classification, *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, 2015, pp. 1026-1034.
- [8] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H., Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv, 2017, arXiv:1704.04861.
- [9] IMAGENET, <https://image-net.org/challenges/LSVRC/2015/>.
- [10] Khan, M. L., Social Media Engagement: What Motivates User Participation and Consumption on YouTube?, *Computers in Human Behavior*, 2017, Vol. 66, pp. 236-247.
- [11] Kim, J.H., Byeon, H.S., and Lee, S.H., Enhancement of User Understanding and Service Value Using Online Reviews, *The Journal of Information Systems*, 2011, Vol. 20, No. 2, pp. 21-36.
- [12] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S., Pytorch: An Imperative Style, High-performance Deep Learning Library, *Advances in Neural Information Processing Systems*, Vancouver, Canada, 2019, pp. 8024-8035.
- [13] Redmon, J. and Farhadi, A., Yolov3: An Incremental Improvement, arXiv, 2018, arXiv:1804.02767.
- [14] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., You only look once: Unified, real-time object detection, 2015, arXiv:1506.02640.
- [15] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L., Mobilenetv2: Inverted residuals and linear bottlenecks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, UT, USA, pp. 4510-4520.
- [16] Seok, B. and Kim, H., Effects of Motivation to use YouTube Beauty Channels as One-Person Media on Channel Evaluation, Product Attitude, and Purchasing Intention, *Korean Association of AD & PR*, 2019, Vol. 21, No. 4, pp. 168-198.
- [17] Tan, M. and Le, Q., Efficientnet: Rethinking model scaling for convolutional neural networks, *International Conference on Machine Learning*, Long Beach, CA, USA, 2019, pp. 6105-6114.
- [18] Tan, M. and Le, Q., Efficientnetv2: Smaller Models and Faster Training, *International Conference on Machine Learning*, Virtual, 2021, pp. 10096-10106.
- [19] Ying, C., Klein, A., Real, E., Christiansen, E., Murphy, K., and Hutter, F., Nas-bench-101: Towards Reproducible Neural Architecture Search, *International Conference on Machine Learning*, CA, USA, 2019, pp. 7105-7114.

## ORCID

Jong Min Kim | <http://orcid.org/0000-0002-1927-6520>  
 Dae Ho Seo | <http://orcid.org/0000-0003-2963-4819>