

# 효율적인 온-디바이스 머신러닝을 위한 이미지 분석 신경망 설계 및 모델 생성 방법

(A Method to Design Image Analysis Neural Network and  
Create Model for Efficient On-device Machine Learning)

이 원 호, 정 철 우, 천 승 만\*  
(재)경북IT융합산업기술원

(Wonho Lee, Cheol Woo Jung, Seung-Man Chun)  
(Gyeongbuk Institute of IT Convergence Industry Technology)

Abstract : In this paper, we propose an on-device machine learning method that can perform machine learning in the Android mobile environment. Recently, as the demand for mobile devices such as smartphones and embedded devices increases, the demand for on-device machine learning to operate machine learning in low-performance devices is also increasing. Therefore, in this paper, we study the process for performing machine learning in a mobile environment with low performance and low cost, rather than server computer oriented machine learning as in the past, and how to create a model file for a mobile environment using the PyTorch mobile. In addition, using the model file generated according to the proposed method, a human skin picture is input to a smartphone which installed Android operating system. Thereafter, the analysis result generated by smartphone is compared with the analysis result on the server computer to evaluate the accuracy. As a result, when using independently builded on-device machine learning, the classification accuracy was 93.9% on average compared to the existing server computer oriented analysis results.

Keywords : On-device, Machine learning, Android, PyTorch, Mobile

## I. 서 론

딥러닝이나 머신러닝과 같은 인공지능에 대한 관심이 급증함에 따라 그 시장도 함께 커지고 있다 [1], [2]. 인공지능 분야는 기본적으로 방대한 데이터와 빠른 처리 속도를 요구하기 때문에 대다수의 연구는 서버급 고사양의 컴퓨터를 통해서만 이루어지고 있다. 그러나 최근 인공지능 기술이 연구자들 뿐만 아니라 대중에게도 보급됨에 따라 강제되는 고사양 컴퓨터가 인공지능 기술에 대한 접근성을 낮추며, 진입장벽을 세우는 결과를 초래하고 있다.

본 논문에서는 이러한 기술의 접근에 대한 어려움을 완화하기 위해 안드로이드 운영체제가 설치된 스마트폰에서 머신러닝을 동작시킬 수 있는 방법을

\*Corresponding Author(smchun@gitc.or.kr)

이원호, 정철우, 천승만: (재)경북IT융합산업기술원

※ 본 연구는 2020년도 중소벤처기업부의 기술개발사업 지원에 의한 연구임[S3012967]

소개한다. 또한 실제 사람의 피부 사진을 입력 받아 모공을 분석하며, 이를 서버에서 생성한 모델로 분석한 결과와 비교해 정확도 및 모바일 모델의 실용성을 평가한다.

## II. 관련 연구

### 1. 신경망 선택 및 구성

일반적인 컴퓨터와 같은 성능을 보장할 수 없는 모바일 환경에서 머신러닝이 수행되어야 하므로 학습 네트워크는 빠르면서 높은 정확도를 보장할 수 있어야 한다. 이러한 이유로 본 논문에서는 U-Net을 이용해 신경망을 구성하였으며 그 모습은 그림 1과 같다 [3]. U-Net은 CNN의 일종으로 수축 단계 및 팽창 단계를 거쳐 입력된 데이터의 특징을 규명하는데 유리한 네트워크이다. 또한 U-net은 이미 검증이 완료된 데이터에 대해 추가적인 검증을 수행하지 않도록 End-to-End 학습 방식을 사용하기 때문에 처리 속도 및 정확도 면에서 우수하다.

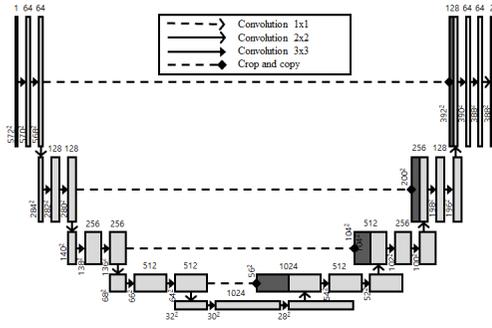


그림 1. 제안된 U-Net 신경망의 구조 및 동작 방식 다이어그램

Fig. 1. Proposed U-Net neural network structure and operation diagram

## 2. 모바일 모델 생성

기존의 인공지능 기술 적용은 서버급 컴퓨터에서 이뤄졌기 때문에 Windows 및 Linux 등의 운영체제에서 학습을 진행시키기에 최적화되어 있다. 그러나 최근 모바일 기기에 대한 수요가 강세를 보이고 이러한 이유로 모바일 기기와 같이 비교적 성능이 낮은 장치에서 인공지능 기술을 도입하고자 하는 수요도 증가하고 있다.

모바일 기기는 운영체제 및 하드웨어의 설계가 일반적인 서버 등의 PC와는 다르며, 하드웨어의 성능 또한 비교적 낮기 때문에 기존의 인공지능 모델을 모바일 기기에 그대로 도입하는 것은 불가능하다. 그러나 2019년 하반기에 Facebook에서 PyTorch mobile을 발표하게 되면서 안드로이드 모바일 기기를 위한 모델 생성 가능성이 열리게 되었다 [4]. 표 1의 Algorithm 1은 PyTorch mobile을 통해 안드로이드 기기에 탑재할 수 있는 모델을 만드는 방법을 보여주는 예시이다. 또한 Algorithm 2는 이렇게 생성된 모델 파일을 Java 기반 안드로이드 어플리케이션에 탑재하는 코드 예시를 나타낸다. 단, 본 논문에서는 모바일 모델에 중점을 맞추기 때문에 서버 모델에 대한 내용은 다루지 않는다.

## III. 실험 및 평가

### 1. 실험 환경

본 논문에서는 U-Net을 기반으로 만들어진 모바일용 모델의 성능을 평가하기 위해 동일한 데이터를 기존의 서버용 모델 및 모바일용 모델에 입력하여 도출되는 분석 결과를 비교한다. 이 때, 모델을 만들기

표 1. 모바일 모델을 만들기 위한 소스 코드 예시  
Table 1. Source code example to create a mobile model

Algorithm 1. create_model_mobile(model)	
1.	torch.save(net.state_dict(), model)
2.	dst_model = torch.load(model)
3.	net = UNet()
4.	net.load_state_dict(dst_model)
5.	net.eval()
6.	script_model = torch.jit.script(net)
7.	script_model._save_for_lite_interpreter('./script_model')
8.	optim_model = optimize_for_mobile('./script_model')
9.	optimize_model._save_for_lite_interpreter('./optim_model')
Algorithm 2. load_model_mobile(model)	
1.	return LiteModuleLoader.load(model)

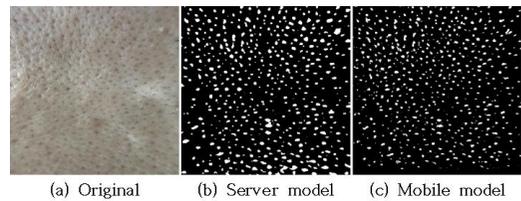


그림 2. 원본 사진 및 기기별 모공 분석 결과  
Fig. 2. Original photo and pore analysis by each device

위해 사용한 데이터는 660장으로 구성된 사람의 피부 사진이며, 결과적으로 생성되는 모델은 새로운 피부 사진을 입력 받으면 모공의 위치 및 크기를 분석 및 반환하게 된다. 표 2는 이렇게 구성된 신경망에서 사용된 하이퍼 파라미터를 나타낸다. 또한, 본 논문에서 제시하는 신경망은 End-to-End 학습의 효율 및 처리 속도 향상을 위해 U-Net 신경망의 활성화 함수 일부를 LeakyReLU로 변경하여 사용한다.

### 2. 실험 평가

그림 2는 각각 원본 사진, 그리고 서버 및 모바일 모델을 통해 모공을 분석한 사진을 보여준다. 이를 위해 임의로 선택된 총 16장의 사진을 이용해 서버 및 모바일에서 분석된 결과를 각각 비교했을 때, 평균 93.9%의 분석 정확도를 보여주었으며, 표 3은 해당 비교 결과의 상세를 나타낸다.

표 2. 모델 생성에 이용된 신경망의 하이퍼 파라미터

Table 2. Hyper-parameters of the neural network used in the experiment

Kernel size	Stride	Max pool	Activator	Input nodes
3 <sup>3</sup>	1	-	ReLU	64 <sup>3</sup> * 1
3 <sup>3</sup>	2	3 <sup>3</sup>	-	64 <sup>3</sup> * 48
3 <sup>3</sup>	1	-	ReLU	32 <sup>3</sup> * 48
3 <sup>3</sup>	2	3 <sup>3</sup>	-	32 <sup>3</sup> * 96
3 <sup>3</sup>	1	-	ReLU	16 <sup>3</sup> * 96
3 <sup>3</sup>	2	3 <sup>3</sup>	-	16 <sup>3</sup> * 192
3 <sup>3</sup>	1	-	ReLU	8 <sup>3</sup> * 192
2 <sup>3</sup>	2	-	LeakyReLU	8 <sup>3</sup> * 384
3 <sup>3</sup>	1	-	ReLU	16 <sup>3</sup> * 384
2 <sup>3</sup>	2	-	LeakyReLU	16 <sup>3</sup> * 192
3 <sup>3</sup>	1	-	ReLU	32 <sup>3</sup> * 192
2 <sup>3</sup>	2	-	LeakyReLU	32 <sup>3</sup> * 96
3 <sup>3</sup>	1	-	ReLU	64 <sup>3</sup> * 96
1 <sup>3</sup>	1	-	SoftMAX	64 <sup>3</sup> * 48

표 3. 서버 및 모바일 머신러닝 결과 비교

Table 3. Comparison of server and mobile machine learning results

개수	평균	표준 편차	분산	최소	최대
16	0.939	0.075	0.006	0.853	0.992

#### IV. 결론 및 향후 연구

앞으로의 인공지능은 조금 더 낮고 넓게 소비자 지향적인 환경을 주시해야 한다. 즉, 모바일 기기를 통해 일상에서도 함께 해야 한다는 수요를 충족하는 기술이 되어야 한다. 그러므로 본 논문에서는 효율적인 이미지 분석을 위해 U-Net을 이용하여 구성되는 신경망 및 이를 모바일 환경에서 사용할 수 있도록 하는 모델 생성 방법 등에 대해 소개하였다. 결과적으로, 모바일 환경에 탑재되는 모델일지라도 기존의 서버 환경에서 이용된 모델과 비교할 때 성능상의 감쇄는 그리 크지 않았다.

본 연구의 향후 연구는 다음과 같다. 첫 번째는 입력되는 이미지의 전처리 방법이다. Blur나 Sharpen과 같은 필터링 기술을 이용하여 입력된 사진을 분석 전에 가공한다. 결과적으로 인공지능 입장에서는 분석친화적인 값을 제공 받을 수 있으며, 처리 속도 또한 향상될 것이다. 두 번째는 이미지의 부분 밝기 조정 방법이다. 밝기는 빛의 여러 가지 복합적 특성에 의해 결정되므로 단 한 장의 사진에서도 여러 결과가 혼재할 가능성을 만든다. 그러므로 밝기를 균등하게 보정하게 되면 이 역시 모델의 성능 증대를 기대할 수 있다.

#### References

- [1] Z. Shao, S. Yuan, Y. Wang, and J. Xu, (2021). "Evolutions and trends of artificial intelligence (AI): research, output, influence and competition". Library Hi Tech, 2021.
- [2] 박정규, 전홍석, 노삼혁, "모바일 로봇의 네비게이션을 위한 빠른 경로 생성 알고리즘", 대한임베디드공학회논문지, Vol. 9, No. 2, page. 101, 2014.
- [3] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241, October, 2015.
- [4] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan and S. Chintala, et al, "Pytorch: An imperative style, high-performance deep learning library", Advances in neural information processing systems, Vol. 32, pp. 8026-8037, 2019.