

MAML을 이용한 마스크 착용 여부 검사 시스템

윤영식[○]

동국대학교 컴퓨터공학과
yys3606@gmail.com

Mask wear inspection system using MAML

Young-Sik Yun[○]

Dept. of Computer Science and Engineering, Dongguk Univ., Seoul

요 약

COVID-19로 인해서 현재는 건물 입구마다 마스크 착용 여부를 확인하는 기계들이 설치되어 있다. 하지만, 이러한 기계 내부에 들어가는 마스크 탐지 기술들은 딥러닝을 이용했지만, 최소 6개월이라는 오랜 시간이 걸려서 상용화가 되었다. 그 이유 중 하나가 기존의 연구가 되지 않고 있어서 데이터가 부족하다고 생각해서 적은 데이터로 학습을 진행할 수 있는 메타 러닝을 이용하였고, 분류 별 1,020개의 데이터를 이용해서 5000번을 반복 학습시킨 결과 88%의 정확도가 나오는 것을 확인할 수 있었다.

1. Introduction

COVID-19 이전에는 마스크 착용은 의무가 아닌 선택 사항이었지만, 전 세계적으로 COVID-19 사태가 심각해지면서 마스크 착용이 의무화되었다. 건물 내부의 마스크 착용이 의무화되면서 건물을 입장할 때 마스크 착용 여부를 확인하는 기계들이 입구에 설치되었고 최근에 학교에도 건물 입구마다 마스크 착용 여부와 온도를 측정하는 기계를 설치하여서 건물에 출입하는 사람들이 마스크를 착용하였는지 확인한다. 이러한 기계들로 입구에서 사람들이 확인하지 않아도 되므로 인력과 비용을 절감하게 되었다. 하지만, COVID-19가 시작된 2019년도 12월 비하면 입구에서 마스크 착용 여부를 확인하는 기계들은 나와서 상용화되기 시작한 시점인 2020년도 중반은 시간이 오래 걸린 것을 알 수 있었다. 이러한 이유 중 하나는 COVID-19 사태가 발생하기 전까지는 마스크 착용 여부를 확인하는 것에 관한 연구를 많이 진행하지 않았기 때문이라 생각한다. 실제로 RISS에서 마스크 착용 여부에 관한 연구를 찾아본 결과 많은 연구가 2020~2021에 이루어졌다. 이전까지 많은 연구가 이뤄지지 않은 상황에서 데이터를 만들거나 수집하는 것에 대한 시간이 오래 소모된 것이라고 생각이 든다. 또한, 기존의 딥러닝 모델을 학습하기 위해서는 많은 데이터 세트가 필요하다. [1]

이러한 이유로 본 논문에서는 적은 데이터로도 학습을 진행할 수 있는 메타러닝(meta-learning)을 이용해서 마스크를 올바르게 착용했는지, 잘못 착용했는지, 마스크를 미착용했는지를 구분한다. COVID-19와 같은 갑작스러운 상황에서 적은 데이터로 학습을 진행하려고 할 때 메타러닝을 이용하면 이전의 딥러닝보다 빠른 기간에 학습을 진행할 수 있고 제품에 적용할 수 있어서 제품의 상용화도 빨리 될 수 있다고 생각한다.

2. Meta Learning

2.1. Meta Learning이란

기존의 Deep Learning은 점점 그 분류 오류가 줄어들고 정확성이 증가하고 있다. 하지만, Deep Learning의 정확성을 증가시키기 위해서는 현재 많은 양의 데이터와 높은 수준의 컴퓨터 성능을 요구한다. 또한, 의료와 관련된 사진, 적게 사용하는 언어의 번역, 로보틱스, 추천 등과 같은 데이터가 충분하지 않거나 많은 데이터를 얻을 수 없는 환경에서는 기존의 데이터 의존도가 높은 Deep Learning 학습 방법은 좋은 성능을 발휘하는 것이 어렵다. 이러한 환경에서의 기존 Deep Learning의 단점을 극복하기 위해서 제시된 것이 메타러닝(meta learning)이다. 메타러닝(meata learning)은 학습하는 방법을 학습한다(Learning to learn)는 개념으로 기존의 많은 데이터로부터 일일이 학습을 진행하던 Deep Learning과 다르게 task(episode)를 통해서 얻은 이전의 경험, 지식을 통해 공통된 지식을 추출하고 이렇게 얻은 지식으로부터 새로운 task에 대해서 빨리 학습을 빠르게 진행하면서 적은 데이터로부터 효율적인 학습을 가능하게 한다. [2]

2.2. Few-shot Learning

few-shot classification은 기존에 image classification에서 pre-trained된 model을 사용하는 방식인 supervised

learning에 meta learning을 적용한 예로 앞서서도 설명을 했듯이 적은 양의 사진들을 이용해서 대상의 클래스를 분류하는 학습을 말한다. few-shot classification은 N-way, K-shot classification이라고도 하는데 이때, N은 class의 수를 의미하고 K는 각 클래스에 대한 데이터의 개수를 의미한다. [3]

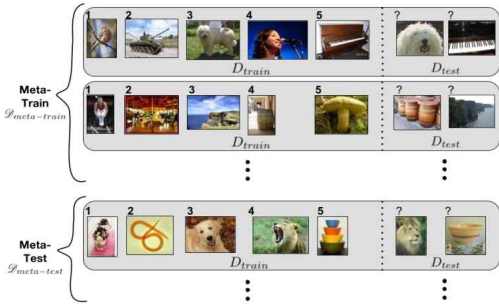


그림 1. few-shot classification 예시

위 그림에서 각각의 행을 task라고 한다. Meta-Train의 첫 번째 task를 보면 D_{train} 에 5개의 번호로 이루어진 class가 있고 각각의 class에는 한 장의 사진으로 구성되어 있다. 이렇게 구성된 task들이 계속 나열된 것을 확인할 수 있는데 이때 이를 5 classes, 1 shot classification이라고 한다. 또한, 각각의 task는 learning을 위한 D_{train} 과 training과 testing을 위한 D_{test} 로 구성되어 있는데 D_{train} 은 support set S라고 하고 D_{test} 는 prediction set B라고 해서 $D=\langle S, B \rangle$ 로 표현한다. 이렇게 구성된 task들을 이용해서 Meta-train 단계에서 training을 진행하고 Meta-test 단계에서 support set S를 통해서 prediction set을 예측하는 방식으로 test를 진행한다.

2.3. 관련 연구

관련된 연구로는 메타 러닝을 이용한 연구는 아니지만, Haar-Cascade를 이용해서 얼굴을 감지한 뒤에 마스크 객체의 점들을 찾아서 점들의 위치에 따라서 마스크 착용 여부를 확인하는 방법이나 [1] 마스크를 검출하는데 학습 과정에서 기울기 소실 문제를 완화하면서 이미지를 분류하는 ResNet를 이용하거나 [4] YOLO-v3 모델을 사용해 빠른 시간 내에 객체를 인식하여 마스크를 검출하는 연구도 있었다. [5] 이러한 연구들은 학습 데이터 세트수집을 10,000장 이상 하거나 YOLO-v3을 이용한 경우에는 학습하는 과정에서 1번의 학습 동안 사용하는 사진이 64장이 사용되었다.

본 논문에서는 메타 러닝을 이용해서 적은 데이터 세트를 이용해서 마스크 착용 여부를 확인하는 시스템을 제안하였다.

3. MAML를 이용한 마스크 착용 여부 검사 시스템

본 논문에서는 먼저, 수집한 고해상도의 데이터를 저해상도로 바꾸는 이미지 전처리 과정을 수행한 다음에 메타러닝 중에서도 최적화 기반(optimization based) 접근방식 중에 MAML (Model-Agnostic Meta-Learning)을 이용해서 마스크 착용 여부를 판별할 것이다.



그림 2. 시스템의 전체 구조

3.1. Data Pre-Processing

수집한 데이터 세트의 해상도가 높아서 전체적인 학습에 걸리는 시간이 많으므로 해상도를 낮추는 작업을 진행하였다. 수집한 데이터를 먼저 흑백 사진을 바꾸고 데이터의 size를 (128, 128) 조절하여서 최종적으로 모든 데이터를 (128, 128, 1)에 저해상도로 저장해서 다음 단계에서 MAML에 입력으로 사용한다.

3.2. MAML (Model-Agnostic Meta-Learning)

MAML은 gradient descent를 통해서 학습하는 어떤 model에도 적용이 가능한 optimization algorithm이다. optimal parameter ϕ 의 가까운 위치에서부터 출발하면 효율적인 학습이 가능하므로 학습을 반복하면서 각각의 optimal parameter 값을 갱신한다. [6]

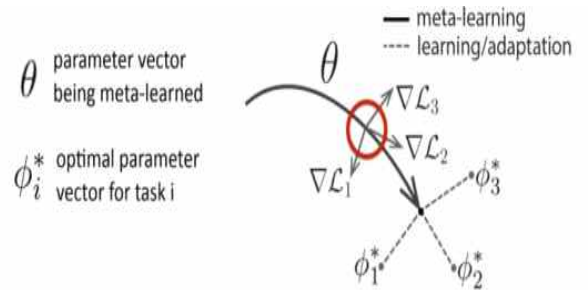


그림 3. MAML 학습

MAML은 2중 반복문 구조로 학습이 진행된다. 먼저, task를 구성한 다음에 내부 반복문에서는 각각의 task를 통해서 optimal parameter인 ϕ_i^* 을 갱신하고 갱신된 각각의 ϕ_i^* 을 통해서 외부 반복문에서 parameter vector

인 θ 을 갱신하는 과정을 반복하면서 layer의 weight를 수정해 나간다. [6]

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β : step size hyperparameters

- 1: randomly initialize θ
- 2: **while** not done **do**
- 3: Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
- 4: **for all** \mathcal{T}_i **do**
- 5: Evaluate $\nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$ with respect to K examples
- 6: Compute adapted parameters with gradient descent: $\theta'_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$
- 7: **end for**
- 8: Update $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'_i})$
- 9: **end while**

그림 4. MAML 학습 알고리즘

본 논문에서는 3x3 filter로 구성된 4개의 convolution layer와 1개의 fully connected dense layer로 구성된 CNN을 MAML Model에 적용하였고 parameter를 갱신하는데 사용하는 손실 함수는 tensorflow에서 지원하는 softmax cross entropy 함수를 사용하였다.

4. Experiments

4.1. 실험 환경 및 데이터

표 1과 같은 환경에서 실험을 진행하였다.

표 1. 컴퓨터 실험 환경

CPU	Intel(R) Core(TM) i5-9400F 2.90 GHz
RAM	16GB

학습의 사용할 데이터 세트는 <https://github.com/cabani/MaskedFace-Net>에서 수집했다. 마스크를 착용하지 않은 사진, 마스크를 잘못 착용한 사진, 마스크를 올바르게 착용한 사진으로 분류하여 수집하였고, 각각의 분류마다 1820개의 사진을 수집하였다. 이 중에서 각각 1020장의 사진을 Meta-train에 사용할 task를 만드는 데 사용하였고, 각각 680장의 사진은 Meta-test에 사용할 task를 만드는 데 사용하였다. Meta-train에서 사용하는 task는 1024*1024*3의 크기의 사진을 Pre-Processing을 진행한 사진으로 분류한 폴더에서 각각 1장씩을 선택해서 3 way - 1 shot으로 Meta-train에 대한 D_{train} 과 D_{test} 을 구성하였고, 학습률은 0.4로 총 5000번의 학습을 진행하였다.

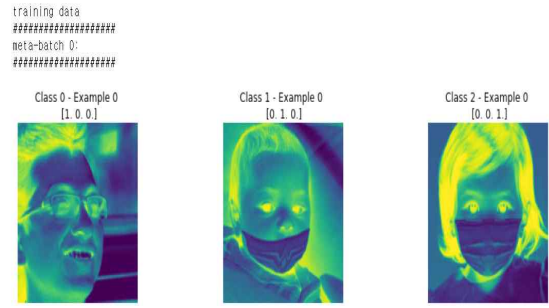


그림 5. Meta-train D_{train}

4.2. 마스크 착용 여부 실험 결과

먼저, Meta-train을 수행하면서 정확도를 확인했을 때 결과는 task가 3 way - 1 shot이어서 적은 반복 횟수에 비해서 높은 정확도가 나온 것을 확인할 수 있었다. 처음 학습을 진행했을 때는 40%의 정확도가 나왔는데 학습을 1000번과 2000번을 반복했을 때에는 약 94%의 정확도가 나왔고 3000번 때에는 85%의 정확도로 떨어졌지만 4000, 5000번째에 92%의 정확도로 다시 올라온 것을 확인할 수 있었다.

표 2. Meta-train 반복 횟수에 대한 정확도

iteration	0	1000	2000	3000	4000	5000
Accuracy (%)	40	94	94	85	92	92

초기에 정확도가 증가하는 방향으로 학습이 진행되다가 중간에 3000번을 반복했을 때는 정확도가 떨어졌지만, 그 뒤로는 92%의 정확도를 유지하였다.

Meta-test도 Meta-train과 똑같이 각각의 640장의 사진에서 1장씩을 선택해서 3 way - 1 shot의 task를 생성해서 500번을 반복해서 진행했다. 100개의 task에 대해서 test를 진행했을 때는 약 89% 정확도가, 200개에 대해서도 약 89%의 정확도가 300개에 대해서도 약 89%에 정확도가 400, 500개에 대해서는 약 88%의 정확도가 나온 것을 확인할 수 있었고, 전체적인 평균 정확도가 88%인 것을 확인할 수 있었다.

표 3. Meta-test 반복 횟수에 대한 정확도

iteration	100	200	300	400	500
Accuracy (%)	89	89	89	88	88

많은 사진에 대해서 반복하면서 정확도는 조금 떨어졌지만 88~89% 사이에 test 정확도를 유지하는 것을 확인할 수 있었고 학습된 데이터에 대한 수치가 과적합이 나지 않았다고 할 수 있다고 생각한다.

5. Conclusion

현재는 딥러닝을 이용한 마스크 착용 여부를 확인하는 다양한 방법들이 있었지만, 이전까지는 연구가 많이 되고 있던 분야가 아니었고, 우리 사회에 필요가 있는 시스템도 아니었기 때문에 충분한 데이터 세트가 없는 상황이었다. 거기에 기존의 딥러닝은 정확도를 상승시키기 위해서 많은 데이터와 학습을 많이 반복해야 하므로 시스템을 구축하는데 많은 시간이 걸린다. 실제로 지금 같이 건물 입구 마스크 착용 여부를 확인하는 기계들은 코로나 발생 시점인 2019년 12월에 비하면 긴 시간이 지나고 상용화되었다.

하지만 이를 메타 러닝을 이용해서 학습을 진행하면 비교적 이른 시간에 시스템을 구축할 수 있다. 실제로 조사한 기존의 딥러닝에 방법들은 마스크 착용 사진을 10,000 이상을 사용하였지만, 이번 실험에서는 1,020장씩 3개의 그룹의 사진만을 이용하였고, 1개의 task에는 총 6장의 사진으로 D_{train} 과 D_{test} 을 구성하여서 진행하였다.

또한, 비록 결과의 정확도가 88%여서 비교적 부정확하다고 할 수 있지만, 학습을 5,000번만을 반복한 거에 비하면 반복 횟수 대비 높은 효율의 정확도가 나왔다고 할 수 있다.

COVID-19라는 갑작스러운 상황이 앞으로도 없을 거라는 보장은 할 수 없다. 그런 상황 속에서 마스크 착용 여부와 같이 작게는 2진 분류 문제 크게는 여러 label을 분류하는 문제에서 데이터가 별로 존재하지 않을 때 메타 러닝을 활용하면 짧은 기한 내에 좋은 시스템을 구축할 수 있을 것이다.

현재 본 논문은 메타 러닝의 학습적인 부분을 중심으로 작성되었기 때문에 이를 확장해서 영상 속에서 마스크 착용 여부를 확인하는 추가적인 연구가 더 필요하다.

REFERENCES

[1] Adnane C, Karim H, Halim B, & Mahmoud M. MaskedFace-Net - A dataset of correctly/incorrectly masked face images in the context of COVID-19. (2021).

<https://www.sciencedirect.com/science/article/pii/S2352648320300362?via%3Dihub>

[2] “CS 330: Deep Multi-Task and Meta Learning” (JUN 10, 2021) Stanford, <https://cs330.stanford.edu/>

[3] Sachin Ravi, Hugo Larochelle. Optimization as Model for Few-Shot Learning. ICLR 2017 conference submission. (2017)

[4] Nam, C., Nam, E., & Jang, K.-S. 딥러닝을 이용한 마스크 착용 여부 검사 시스템. 한국정보통신학회논문지, 25(1), 44 - 49. (2021).

[5] Lee, M. J., Kim, Y. M., & Lim, Y. M. 딥러닝을 활용한 마스크 착용 얼굴 체온 측정 시스템. 한국멀티미디어학회논문지, 24(2), 208 - 214. (2021).

[6] Chelsea Finn, Piter Abbeel, Sergey Levine. Model-agnostic meta-learning for fast adaption of deep networks. ICML 17. 70, 1126 - 1135. (2017)