

CNN을 활용한 음식 분류 결과 보고

YBIGTA 24기 임세은

단순한 CNN 모델 2개 (vanillaCNN1, vanillaCNN2)와 ImageNet으로 pretrain된 VGG19 모델을 사용해 사진에 따라 음식을 분류해주는 모델을 구축했습니다. 총 4838개의 train data, 955개의 validation을 가지고 학습을 진행했으며, 분류 카테고리는 20개입니다. 실험은 3개의 모델 모두 epoch 50, batch size 32, learning rate 1e-4의 동일한 설정으로 진행되었습니다.

1. 실험 결과

Model	Epoch	Validation Score
VanillaCNN1	45	0.4094
VanillaCNN2	50	0.3822
VGG19	45	0.7874

*Validation Score는 4번째 자리까지 반올림 한 결과입니다.

2. 결과 분석

VGG19의 성능이 가장 우수했고, 나머지 2개의 VanillaCNN 모델의 경우 유사한 성능을 보였으나 VanillaCNN2가 더 좋지 않은 결과를 보였습니다.

1) VGG 19

VGG19는 ImageNet 데이터를 사용해 사전 훈련된 모델입니다. 이는 곧 모델이 사전 훈련을 통해 다양한 이미지와 상황을 학습했음을 의미하며, 곧 기본적인 CNN 모델보다 일반화 능력이 상대적으로 높을 수 있음을 의미합니다. 더불어, 사전 훈련을 통해 학습된 가중치를 사용하기에 훈련 초기부터 복잡한 특징 인식이 가능했다는 것 역시 하나의 요인이 된다고 생각합니다.

2) VanillaCNN2

VanillaCNN1과 VanillaCNN2는 fc layer (head)외에는 완전히 유사한 구조를 보이고 있습니다. 이론적으로 생각해보면 단순히 하나의 linear layer만 사용한 VanillaCNN1보다 여러 개의 Linear layer, ReLU, Dropout 등을 이어 붙인 VanillaCNN2의 성능이 더 높을 것으로 예견됩니다. 그럼에도 불구하고 VanillaCNN2가 더 좋지 않은 성능을 보인 이유는 학습 속도에 있다고 생각합니다. VanillaCNN2는 VanillaCNN1에 비해 더 복잡한 head를 사용하고 있기 때문에 데이터 표현의 학습 범위가 확장된다는 장점이 있으나, 더 많은 계산을 요구한다는 단점 역시 있습니다. 훈련 과정에서 loss나 val score가 거의 유사하게 나왔음을 고려할 때, epoch 수가 더 많았다면, 즉, VanillaCNN2에게 학습의 기회가 조금 더 있었다면 VanillaCNN1보다 더 좋은 성능을 보였을 것이

라고 생각합니다.

3. Pytorch 사용 의의

지난 3주차 내용과 비교해봤을 때, Pytorch 사용 의의는 풍부한 리소스와 편리함에 있다고 생각합니다. 일일이 조건과 상황을 고려해 numpy로 수식을 구현하지 않고도 코드 한 줄로 딥러닝에 필요한 연산을 적용할 수 있고, 사전 학습된 모델이나 다양한 라이브러리를 통해 직접 구현한 모델과의 비교 분석 역시 가능하기 때문입니다.