

Deep Residual Network를 이용한 신발 상품 분류

컴퓨터공학과
2015104221 정준현

개 요

1. 서론

1-1. 연구배경

1-2. 연구목표

2. 기존 연구

2-1. 기존 연구 1

2-2. 기존 연구 2

2-3. 기존 연구의 문제점

3. 프로젝트

3-1. 기존 연구와 차이점 및 해결방안

3-2. 프로젝트 내용

4. 진행 일정

5. 결론

참고문헌

1. 서론

1-1. 연구배경

SNS와 인터넷에 게시되는 신발 사진은 대부분 판매용이 아닌 이상 그 신발의 브랜드, 정확한 상품명등 신발에 대한 자세한 정보를 기재하지 않은 채 게시된다. 이로 인해, 이 신발을 구매하고 싶거나 이 신발에 대해서 자세한 정보를 찾아보고 싶어하는 요구가 있을 때 이에 대한 객관적인 솔루션이 존재하지 않아 자세한 정보를 사용자에게 제공할 수 없었다. 본 연구에서는 이러한 요구에 대해서 Deep Learning의 대표적인 이미지 인식 인공지능 경망 모델인 Residual Network(ResNet)을 이용하여 해결하고자 한다. 즉, 어떠한 신발 사진에 대해서 이 신발의 상품명을 ResNet을 통해 예측하고 이 신발과 유사한 다른 신발의 상품명 또한 알려주는 소프트웨어를 개발하고자 한다.

1-2. 연구목표

본질적인 목표는 어떤 신발 사진에 대해서 이 신발의 정확한 상품명과, 이 신발과 매우 유사한 신발의 상품명 10개 정도를 사용자에게 알려주는 것이다.

이러한 본질적인 목표를 달성하기 위해서, 먼저 ResNet의 신발 상품명 분류 학습을 위해서 적절한 학습 데이터를 구해야한다. 학습 데이터를 확보한 후, 학습을 어느정도 진행 시키고 나서 여러 각도에서 촬영한 특정 신발 이미지를 가지고 테스트를 진행하여 정확한 상품명에 출력되는 지, 유사한 신발의 상품명에 출력되는지를 확인하여 만약 결과가 좋지 않다면 학습의 parameter를 조정해가며 다시 학습하고, 테스트 하는 과정을 반복할 것이다.

2. 기존 연구

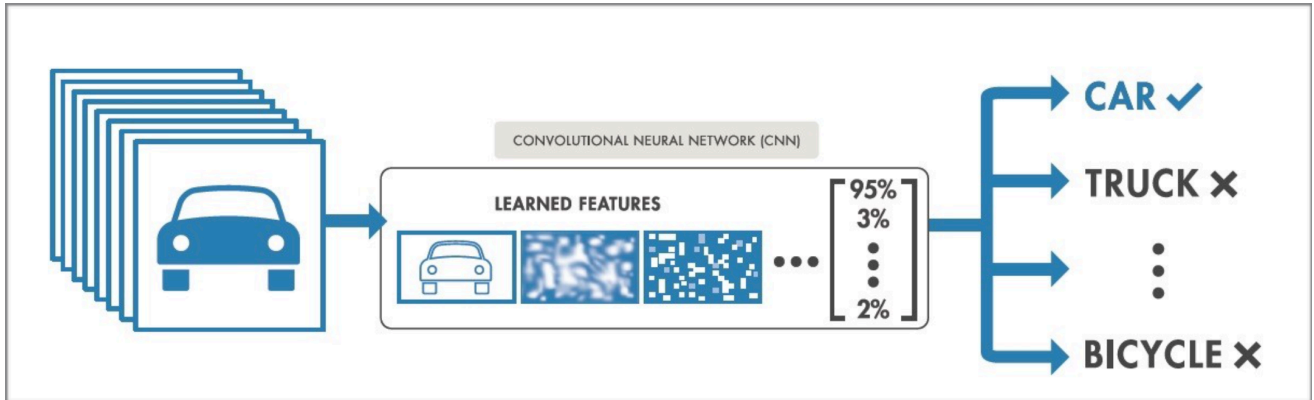
2-1. 기존 연구 1

Deep Learning의 대표적인 인공지능 경망 모델 중 하나인 ‘Convolutional Neural Network(CNN)’를 통해 image classification 을 할 수 있다. <그림 1>는 자동차, 트럭, 자전거 등의 이미지 데이터에 대해서 이를 CNN architecture를 통해 학습 시킨 후, 어떤 이미지를 architecture에 넣었을 때 이 이미지가 자동차, 트럭, 자전거 혹은 다른 어떤 것인지 알려주는 image classification의 예이다.

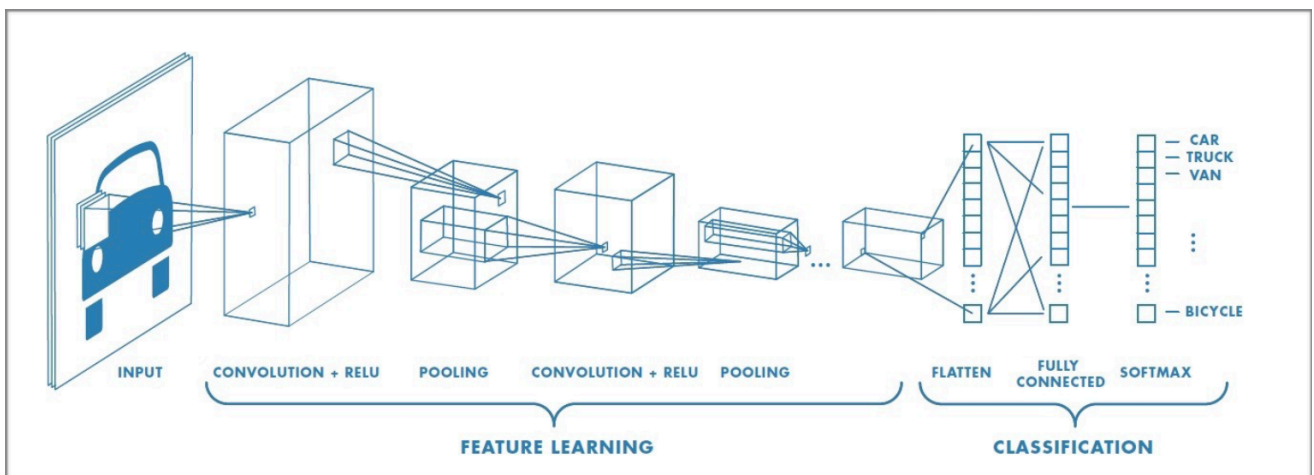
일반적인 CNN의 구조는 <그림 2>와 같다. 첫번째로, matrix 형태의 weight를 가지고 convolution을 진행하고, 이를 활성화 함수인 Relu를 거치게 한 후 Max pooling을 통해 이미지의 dimension을 줄인다.

이 과정은 이미지의 대표적인 특징을 추출하기 위한 단계이다. 이러한 이미지의 대표적인 특징을 추출하는 일련의 작업들을 반복적으로 진행한후, 1차원 행렬로 flatten하여 이를 fully connected layer (Multi Layer Perceptron)의 input으로 넣어준 후, 최종적으로 차, 트럭, 밴, 자전거등 각각의 Class에 대한 확률을 계산하게 된다. Fully connected layer을 통해 2차원 영상 정보로부터 receptive field와 강한 신호 선택의 특성을 살려,

topology 변화에 강인한 인식 능력을 갖게 된다. 이를 통해 어떠한 Image에 대해서 정확하게 Classification할 수 있게 되는 것이다.



<그림 1. CNN architecture를 이용한 Image Classification의 예>



<그림 2. Image Classification CNN의 일반적인 구조>

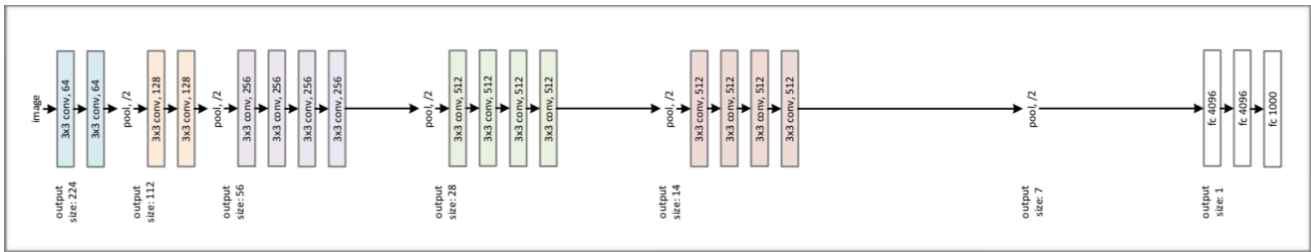
2-2. 기존 연구 2

2012년에 AlexNet의 영향에 따라, Convolution layer 개수를 높여 network를 더욱더 deep하게 구성한 ‘VGG16 model’이 2014년에 발표되었다. 이는 16개의 convolution layer로 이루어져 있으며, 다수의 max pooling layer와 마지막에 3개의 connected layers로 이루어져 있다.

VGG Model의 주요 특징은 3x3 filter를 처음으로 도입했다는 점이다. AlexNet은 11x11의 filter를 사용했던 반면 VGG model은 3x3 filter를 사용한 이유는 반복되는 패턴을 더 쉽게 인식할 수 있게 되고, 학습 가중치 파라미터들은 줄어들게 되므로 사용하였다.

그 결과로, 2014년 ImageNet Challenge에서 7.3의 error rate를 보였다고 한다. 이는 AlexNet의 ImageNet에서의

error rate를 2배 이상 줄인 결과이다.

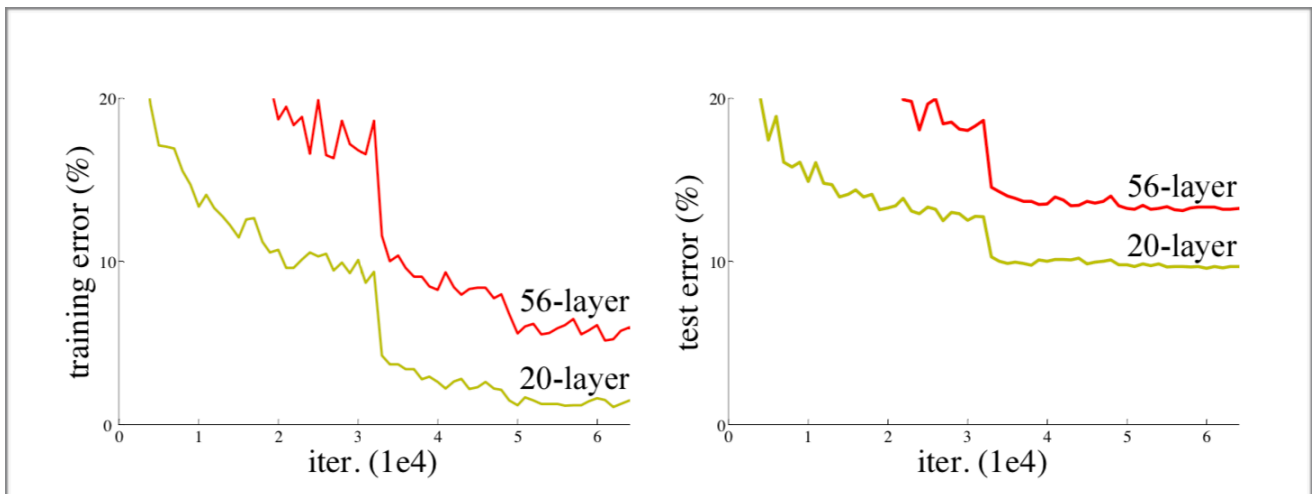


<그림 3. VGG 16 Model의 구조>

2-3. 기존 연구의 문제점

AlexNet, VGG Model, Inception modules 등의 뛰어난 Performance를 바탕으로, Convolutional neural network model의 트렌드가 convolution layer의 개수를 더 많이 뒤서 network를 깊게 두는 것이 되었다. 그러나, 네트워크를 깊게하면 깊게 할 수록, error rate 는 더욱 더 올라가는 현상을 보였다. 이는 오버피팅때문이 아닌, 깊은 network의 hyper parameter를 조정하고, 학습을 최적화하는 작업이 매우 힘들기 때문이다.

<그림 4>에서 볼 수 있듯, 56개의 layer로 구성된 plain CNN model은 20개의 layer로 구성된 plain CNN model보다 training error와 test error에서 모두 높은 것을 볼 수 있다.



<그림 4. 20개의 layer와 50개의 layer로 구성된 Network의 CIFAR-10 성능 측정 결과>

3. 프로젝트

3-1. 기존 연구와 차이점 및 해결방안

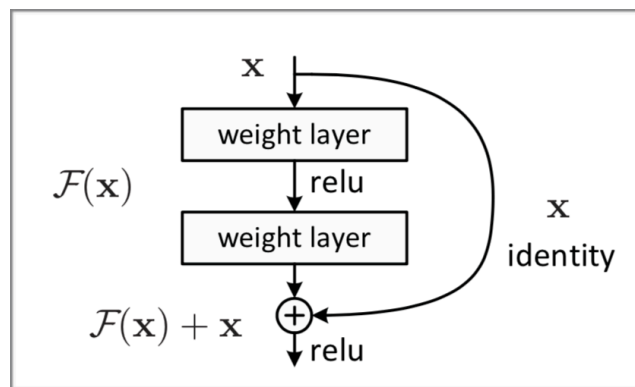
기존 연구는 network의 layer를 깊게 쌓으면 쌓을 수록 Vanishing gradient등의 이유로 성능이 낮아졌으므로 이

를 해결하기 위해 ‘Residual learning’이라는 Concept를 본 프로젝트에서는 이용하려고 한다.

<그림 5>에서 , Residual learning이란 한개 혹은 여러개의 convolution layer의 결과로 나온 $F(x)$ 와 original input X 와의 connection을 만들어 주는 것을 의미한다. 즉, Residual Learning을 통해 input X 로 부터 나오는 정보는 대부분 유지하고, 조금씩의 변화만 줄 수 있게 된다.

결과적으로, Input image로부터 나오는 다양한 패턴들을 deep한 layer를 통해 학습시킬 수 있게 된다. 추가적으로 학습을 해야되는 파라미터도 없고, computational cost 또한 증가하지 않으면서 이미지 classification 성능을 높일 수 있는 모델이다.

따라서 본 프로젝트에서는 이러한 Residual Block으로 구성된 ResNet을 이용하여 Network를 깊게 두어 신발 상품명을 classification하는 모델을 구축하고자 한다.



<그림 5. Residual block의 구조>

3-2. 프로젝트 내용

프로젝트의 진행 내용은 크게 3가지로 될 것이다. 첫 번째, 학습을 위한 다양한 각도에서 찍은 신발 이미지 데이터 수집 및 확보. 두번째, ResNet을 이용하여 Classification Model 구축. 세번째, 학습진행 및 테스트를 통한 hyper parameter(layer 수, learning rate 등) 조절.

첫 번째, 학습을 위한 신발 이미지 데이터 수집 및 확보는 인터넷 쇼핑몰에서 나와있는 신발 상품 이미지들을 크롤링하여 진행하고자 한다.

두번째, ResNet을 구성하는 데 있어서 classify해야하는 class가 본 프로젝트에서는 굉장히 많다. 예를 들어 브랜드 별로 신발이 약 400개씩 있다고 가정하면 브랜드가 20개만 있어도 classify 해야 할 class가 8000개가 넘는다. 따라서 이러한 large scale classification 문제를 해결하기 위한 ResNet 모델을 구축해야한다.

세번째, 학습을 진행하고 테스트를 해보면서 다양한 각도의 신발 이미지를 model의 input으로 주어 정확하게 신발의 상품명을 알려주는 지 확인 할 것이다. 만약 결과의 정확도가 높게 나오지 않으면, 학습 파라미터들을 조정해가면서 정확도가 높게 나올 때 까지 학습 반복적으로 진행하고자 한다.

4. 진행 일정

진행 주차	담당	내용
4주차	정준현	주제 확정 및 관련 논문 및 자료 조사
5주차~6주차	정준현	모델 설계, 데이터 수집 및 논문 초안 작성
7주차~10주차	정준현	설계 검증 및 초기 모델 구현
11주차~14주차	정준현	데이터 학습 및 테스트 데이터에 대해 결과 검증
15주차~16주차	정준현	어플리케이션 프로토타입 준비, 발표 준비

5. 결론

신발 이미지에 대해서 이 신발의 상품명, 브랜드, 이 상품과 유사한 다른 브랜드의 상품들을 추천해주는 소프트웨어를 구현할 것이다. 이 소프트웨어를 통해 사용자들은 원하는 신발에 대한 자세한 정보들을 쉽게 얻어낼 수 있게 될 것이다.

참 고 문 헌

- (1) Karen Simonyan, Andrew Zisserman, 「Very Deep Convolutional Networks for Large-Scale Image Recognition」, 2014.
- (2) Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jean Sun, 「Deep Residual Learning for Image Recognition」, 2015.