

이미지 분류를 위한 기술 조사 및 성능 비교

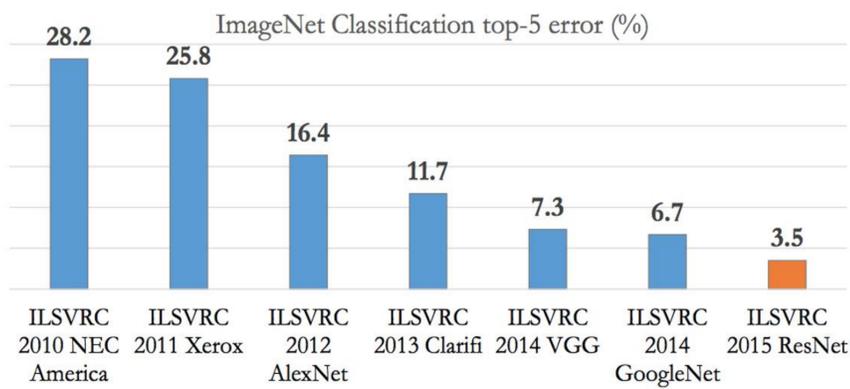
이름 정찬호

지도교수 조현석

연구배경

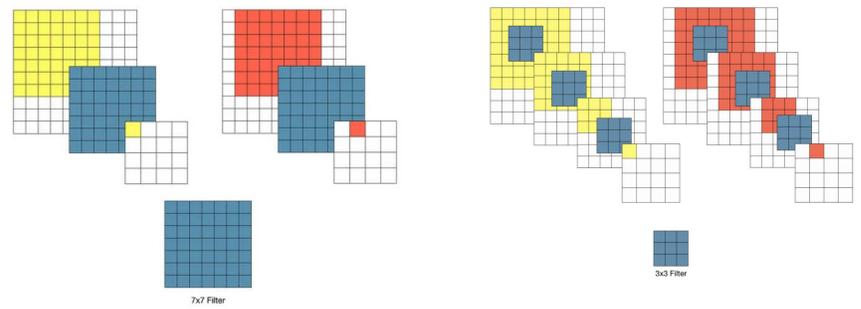
이미지 분류 모델은 대량의 이미지들을 활용하여 여러 분야에서 기술적인 혁신을 이뤄내고 있다. 본 연구에서는, 과거부터 현재 까지 이미지 분류 모델 논문들을 분석하여 사용된 알고리즘의 발전 동향을 살펴보고 정리한다. 추가로 각 모델의 실제 성능을 확인하여, 이미지 분류 모델의 발전 과정을 알아본다.

연구 관련 이미지 분류 경진 대회



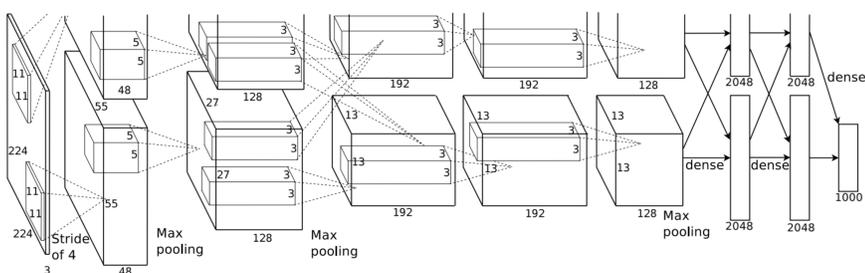
위 표는 2010년에 시작된 ILSVRC (ImageNet Large Scale Visual Recognition Challenge) 대회에서 매년 우승을 차지한 모델의 top-5 error 지표를 나타내고 있다. ILSVRC 경진대회는 참가자들이 대회에서 주어진 데이터셋을 사용하여 본인들이 준비한 모델을 학습시키고, 테스트 데이터셋에서 최고의 성능을 보이는 모델을 제출하는 방식으로 진행된다. Top-5 error는 모델이 테스트 데이터 셋에서 가장 높은 확률로 예측한 5개의 클래스 중 실제 클래스가 포함되어 있지 않은 이미지의 비율을 나타낸다. 매년 최신 모델들이 경쟁하며, 대회에서 우승한 모델들은 이미지 분류 분야의 발전에 큰 기여를 하였다. 특히 2012년 이후로는 deep learning 모델의 초석인 Alexnet 아키텍처의 등장으로 top-5 error rate가 현저히 낮아지는 시기였으며 이후부터 deep learning 기반 모델 아키텍처들이 계속 발전해 왔다. 따라서 Alexnet 부터 어떠한 방식으로 아키텍처가 구성되어 왔는지, 그리고 어떠한 방향으로 아키텍처가 발전해 왔는지 기술 조사 및 성능비교를 수행 하고자 한다.

위 그림은 2012년 발표된 AlexNet CNN 아키텍처의 형태를 보여준다. Alexnet 에 도입된 CNN(Convolutional Neural Network) 모델 방식은 1990년대부터 이미지 인식 및 패턴 인식 작업을 위해 개발된 모델이다. AlexNet에서는 이 모델의 성능 향상을 위해 여러 기법을 사용하였고, 이 중 두가지를 설명하고자 한다. 첫번째는 ReLU 활성화 함수를 사용한 것이고 두번째는 Multiple Gpu Training을 이용한 것이다. 활성화 함수란 인공 신경망의 각 뉴런에서 받아온 입력신호를 적절한 형태로 변환시켜 출력하는 함수이다. 적절한 형태는 인공 신경망 학습에서 비선형적인 구조를 이용하여 출력결과를 연속성 있게 하여 신경망의 표현성을 높이는 것을 의미한다. 대표적으로는 시그모이드 함수가 있다. 이 활성화 함수를 사용하여 error-rate를 현저하게 낮출 수 있게 되었다. Multiple GPU는 이전부터 CNN 모델은 구성되었으나 대량학습의 어려움이 있었던 부분을 해결한 것이다. GPU 2대를 병렬로 사용함으로써 연산량을 급격히 줄일 수 있게 되었다.



위 그림은 2014년 발표된 VGGNet 의 가장 주요한 특징인 3x3filter를 사용한 부분을 보여준다. VGGNet 모델에서는 매우 작은 컨볼루션 필터를 사용함으로써 AlexNet보다 컨볼루션 네트워크의 깊이를 더 확장함과 동시에 연산의 양 또한 줄일 수 있게 되었다. 왼쪽 그림은 7x7의 필터로 convolution을 수행한 결과이고 오른쪽 그림은 3X3의 필터로 convolution을 수행한 결과이다. 오른쪽 그림과 같이 3X3의 필터로 3번의 convolution을 수행하면 출력 특징 맵(4x4)의 픽셀당 Receptive Field 는 $3 \times 3 \times 3 = 27$ 의 크기를 가질 수 있지만, 왼쪽과 같은 큰 필터는 $7 \times 7 = 49$ 라는 보다 큰 크기의 Receptive Field를 가지게 된다. 이러한 작은 필터를 통해 overfitting 문제를 해결함과 동시에 깊은 신경망을 이용함으로써 모델 특징의 식별성이 증가하는 장점까지 얻을 수 있다. 이를 통해 2년 사이에도 기술의 발전이 계속 되었음을 알 수 있다.

기술조사 연구진행과정



결론

앞서 소개한 두 개의 deep learning CNN 아키텍처 이외에도 2014년에 나온 GoogleNet, 2015년에 나온 ResNet 등등 기술의 발전은 계속되어 왔다. 이번 연구를 통해 이미지 분류 기술에 있어서 어떠한 동향으로 발전이 되어왔는지 파악할 수 있는 기회가 되었고, 미처 포스터에서 설명하지 못한 이후의 학습 모델에 대해서는 자기주도연구 보고서에서 더 다루도록 한다