

# 18\_4\_cnn\_architectures

☰ Tags

CNN 아키텍처

[Alexnet](#)

[VGGnet](#)

[GoogleNet](#)

[ResNet](#)

[MobileNet](#)

[EfficientNet](#)

[이 후의 CNN 아키텍처들](#)

[정리](#)

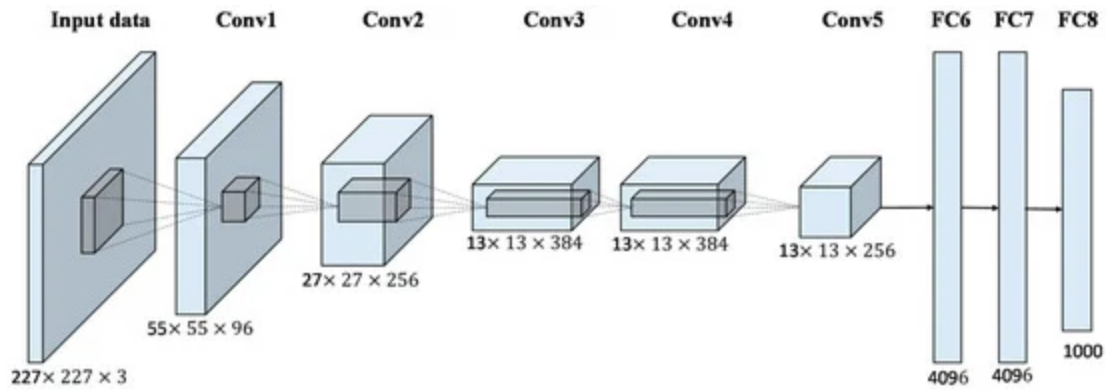
## CNN 아키텍처

지금까지 간단한 CNN 모델을 구축하고 학습시켜 보았습니다. 그런데 이미지넷과 같이 수백만 장의 이미지가 있고, 예측해야하는 클래스의 개수가 1000개라면 어떻게 할까요? CNN 모델을 단순히 더 깊게, 더 넓게 쌓아주면 될까요? 그럴 경우 vanishing gradient 현상이 일어나진 않을까요? 크기는 작지만 성능은 뛰어난 모델을 만들 수 는 없을까요?

이러한 고민들을 담아서 수많은 아키텍처들이 제안되었습니다. 모두 살펴볼 수는 없으니, 핵심적인 모델들과 주요 아이디어만 간단히 살펴보고 넘어가겠습니다.

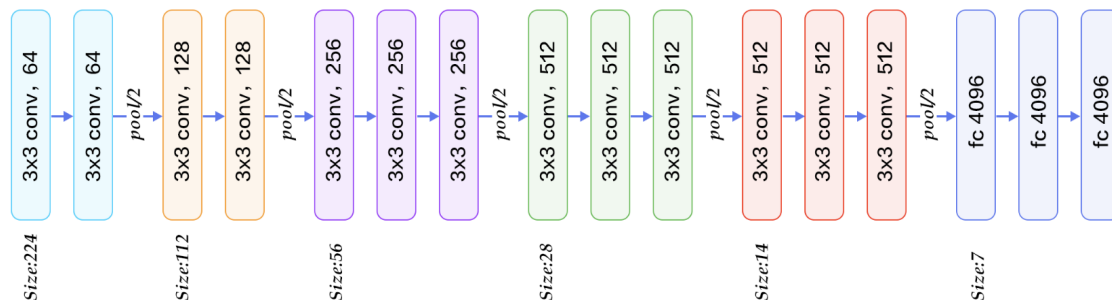
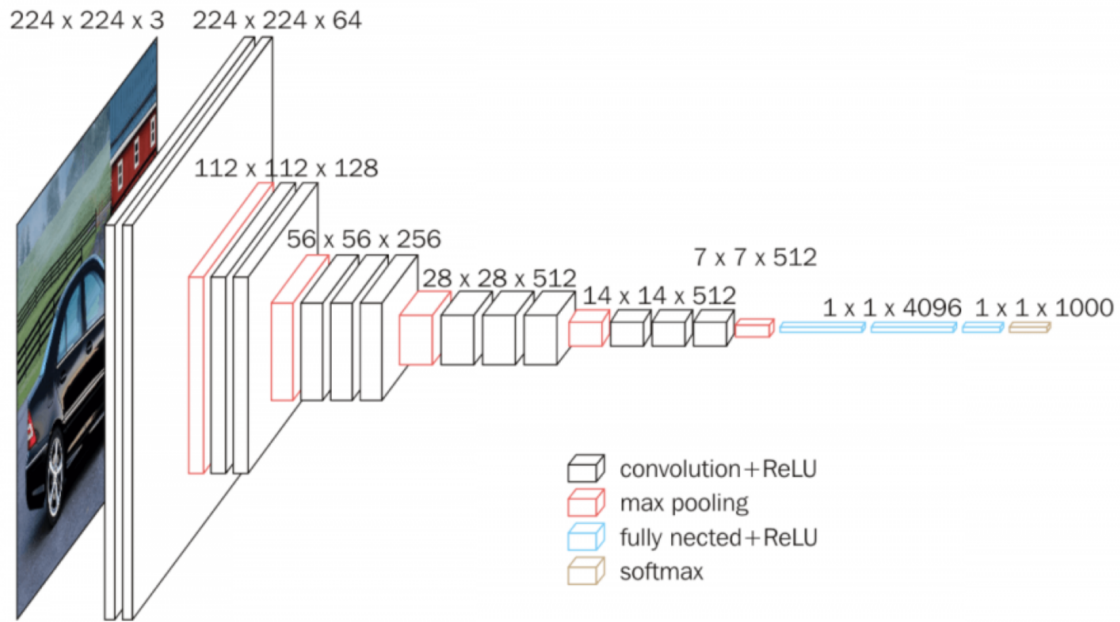
### Alexnet

제프리 힌튼 교수님 연구실에서 제안한 모델입니다. 이미지넷 챌린지에 처음 등장하여 분류 테스트의 성능을 대폭 향상시키면서 컴퓨터 비전 분야에 딥러닝 열풍을 몰고 왔습니다. 지금 시점에서 보면 굉장히 단순한 아키텍처입니다.



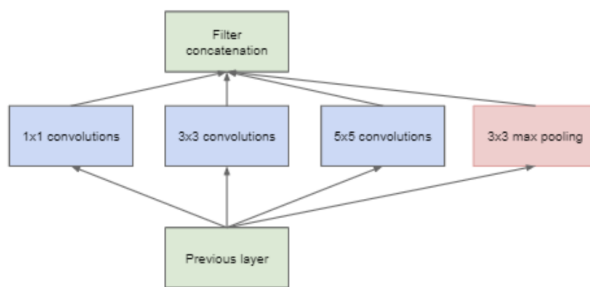
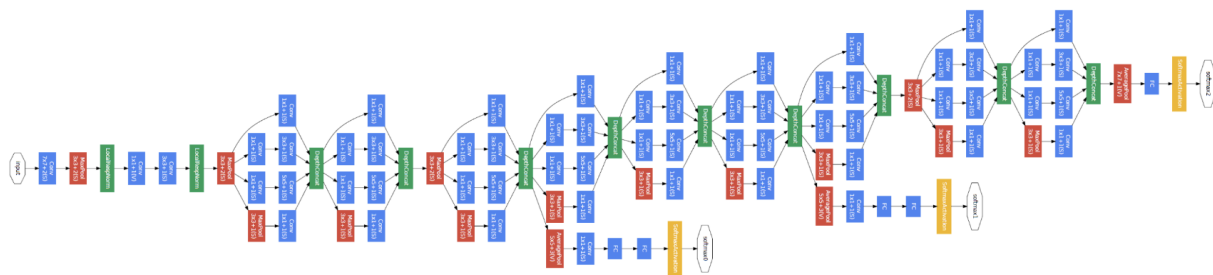
227x227 크기의 이미지를 입력으로 받아서 컨볼루션 레이어를 통과하면서 정보를 추상화 하고, Fully Connected Layer를 두번 통과하여 최종적으로 1000개의 클래스에 대해서 각각에 속할 확률값을 리턴해줍니다. 가장 기본적인 신경망 구조입니다.

## VGGnet

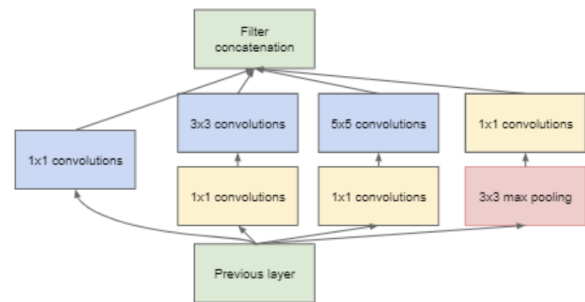


VGG는 영국 옥스포드 대학교의 컴퓨터 비전 연구실에서 제안한 모델입니다. (연구실 이름이 VGG입니다.) Alexnet 이후에 연구자들은 CNN 신경망을 더 깊게 쌓는 것에 초점을 두었습니다. VGGNet은 3x3 크기의 컨볼루션 필터를 반복해서 쌓아주면 깊은 신경망을 효과적으로 학습시킬 수 있다는 것을 발견하고 제안하였습니다. 아래 다이어그램을 보면 모든 컨볼루션 필터들의 크기가 3x3인 것을 확인할 수 있습니다.

## GoogleNet



(a) Inception module, naïve version



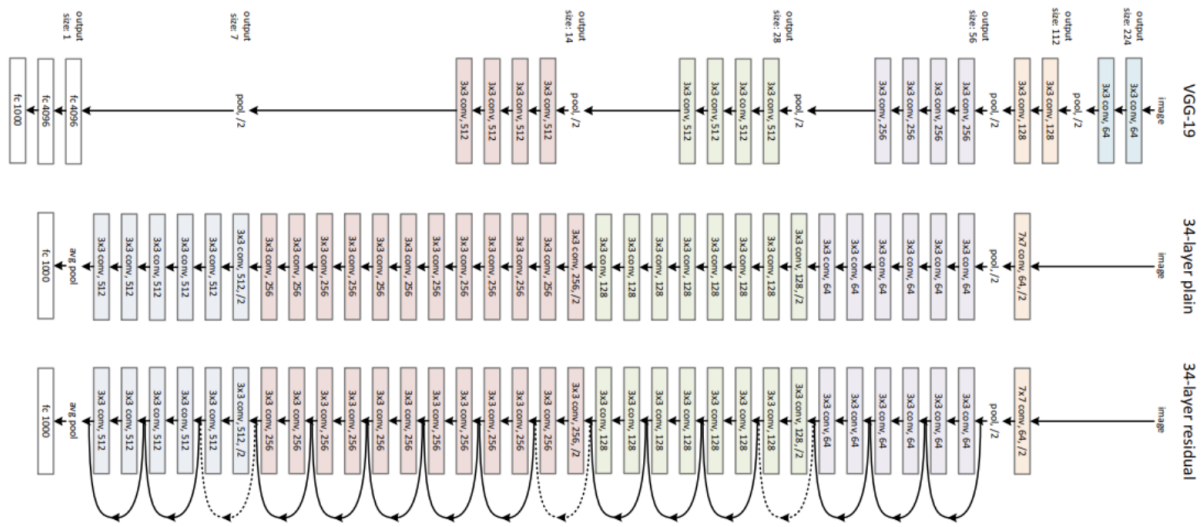
(b) Inception module with dimension reductions

구글에서 제안한 아키텍처입니다. 여기서는 서로 다른 크기의 컨볼루션 필터를 동시에 적용한 다음, 하나로 합쳐서 피쳐맵을 뽑는 Inception Module이라는 것을 제안합니다. 그리고 피쳐맵의 차원을 효과적으로 줄여줄 수 있는 1x1 컨볼루션이라는 기법을 제안했습니다.

첫번째 다이어그램에서 중간 중간 보이는 노란색 셀은 vanishing gradient를 방지하기 위해서 신경망의 중간 중간에서 로스를 계산하고 거기서 역전파를 시켜주는 기법을 의미합니다.

아키텍처가 엄청 요란하지만 그렇게 효과적이지 못했고, 1x1 컨볼루션으로 피쳐맵의 차원을 줄이는 기법만 많이 사용됩니다.

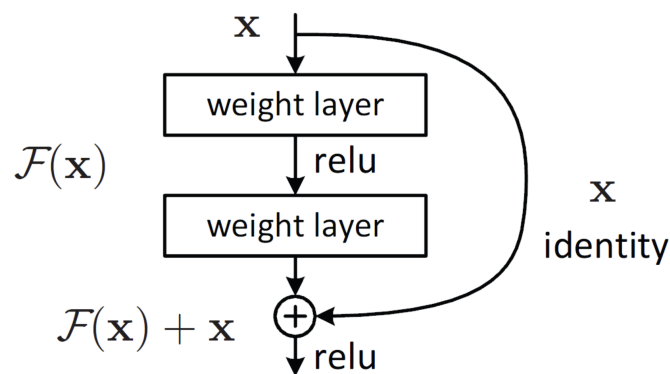
## ResNet



ResNet-34 architecture

구 MS, 현 메타 소속의 연구자, Kaming He가 제안한 아키텍처입니다. 컴퓨터 공학 역사상 가장 많은 인용을 받은 페이퍼 중에 하나인 Residual Network입니다. 이전까지 등장한 모든 아키텍처들은 컨볼루션 레이어를 20층 이상 쌓지 못했습니다. VGG 같은 경우 19층이 가장 큰 모델이었으며, 그 이상 쌓으면 vanishing gradient 때문에 학습이 잘 이루어지지 않았습니다.

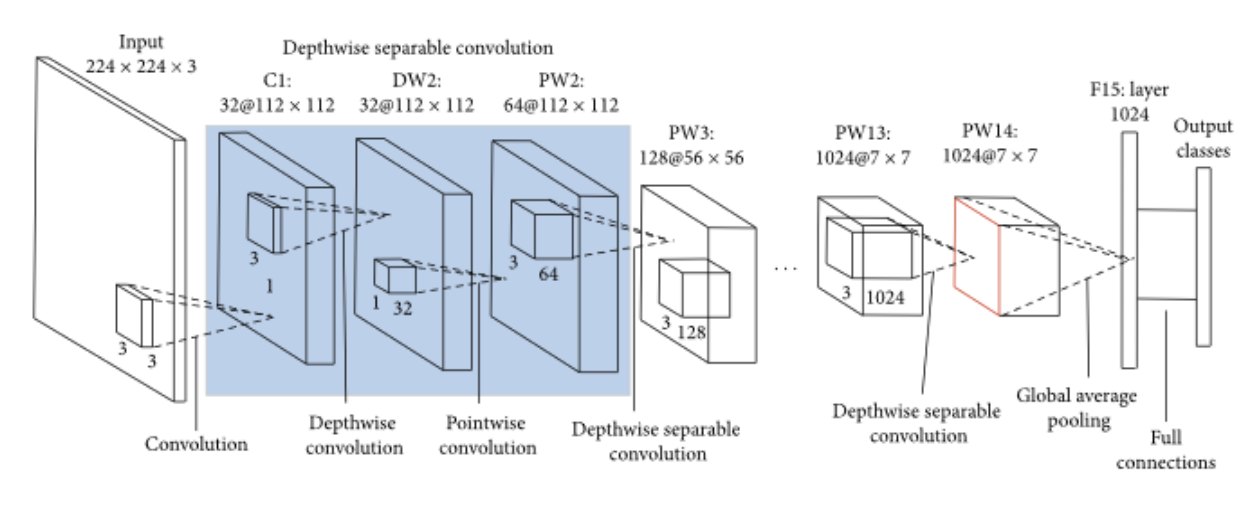
Residual Network는 skip connection이라는 개념을 사용해서 무려 151층의 CNN 모델을 학습시키는데 성공합니다. 그리고 당시 존재하던 거의 모든 컴퓨터 비전 벤치마크를 갱신합니다. 이후에 제안되는 거의 대부분의 딥러닝 모델들이 ResNet에서 제안한 skip connection 아이디어를 사용하며, 트랜스포머 모델에도 skip connection이 사용됩니다.



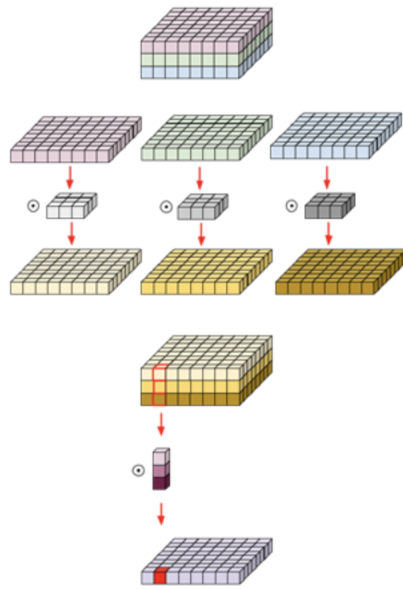
skip connection 아이디어를 간단히 설명하면, “깊은 신경망이 적어도 얇은 신경망 만큼의 성능은 낼 수 있도록 보장해주자”입니다. 깊은 신경망을 단순히 신경망을 깊게 쌓은 모델이 아니라, 얇은 신경망의 각 층 사이사이에 추가적인 층을 끼워넣어주는 관점으로 접근합니다. 그리고 새롭게 끼워넣은 레이어를 통과하지 않고 스킵한 값을 새로 추가한 값과 더해주는 skip connection 개념을 제안합니다. 이렇게 하면 원래의 얇은 신경망의 성능이 보존되면서, 레이어 추가로 인한 성능 향상 효과를 얻을 수 있습니다.

Resnet은 뛰어난 성능으로 현재도 많이 사용되는 아키텍처입니다. 그리고 skip connection 아이디어는 이후에 등장하는 거의 모든 신경망 구조에 사용됩니다. 여담으로 ResNet을 집필한 Kaming He는 이후에 메타로 이직하여 딥러닝을 이용한 컴퓨터 비전 분야에 눈부신 기여들을 합니다.

## MobileNet



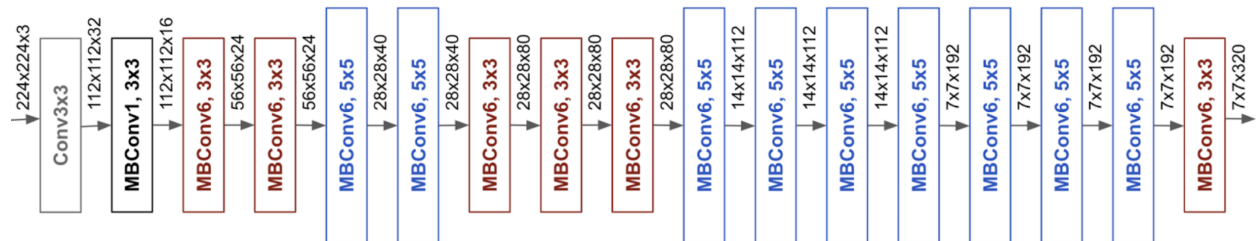
ResNet 이후로 신경망을 크고 깊게 쌓는 것 보다는 성능 저하는 최소화 시키면서 더 얇고 가벼운 구조가 많이 등장했습니다. 경량화 된 신경망 모델을 모바일 디바이스에서 구동시키는 것이 목적이었습니다. 그 중 대표적인 것이 구글에서 제안한 MobileNet입니다.



Depth-wise Separable Convolution

mobilenet에서 제안한 핵심적인 아이디어는 depthwise separable convolution입니다. 보통 CNN의 경우 모든 채널에 대해서 일괄적으로 컨볼루션 연산을 수행합니다. 반면에 depthwise separable convolution은 각 채널을 분리하여 따로따로 컨볼루션 연산을 수행하고, 이를 하나로 합쳐주는 방식을 제안합니다. 이렇게하면 획기적으로 모델의 파라미터 수를 줄이면서도, 뛰어난 성능을 얻을 수 있습니다.

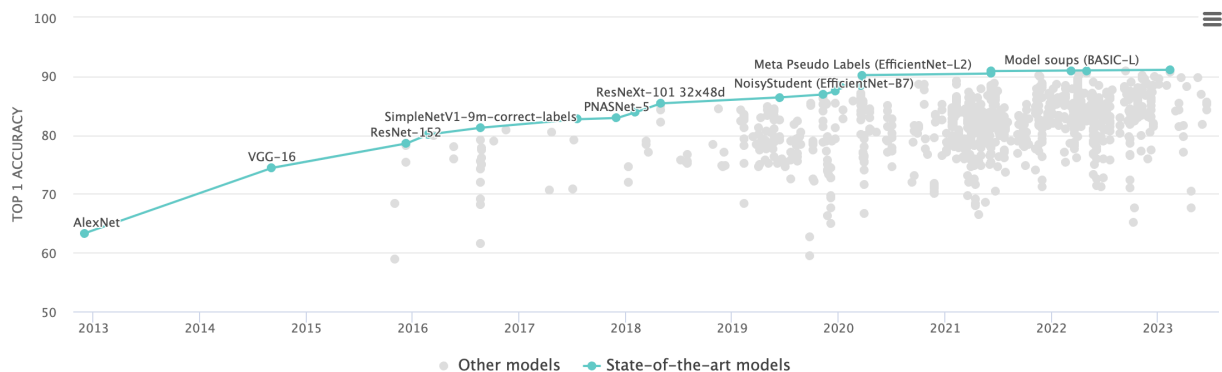
## EfficientNet



이후에 트렌드는 강화학습을 이용해서 최적의 CNN 아키텍처를 찾아내는 것으로 넘어갔습니다. 이런 연구 분야를 autoML 분야 중에서도 Network Architecture Search라고 부릅니다. 이는 조합 가능한 신경망 블록들을 재료로 주고, 반복해서 여러 조합으로 모델을 만들어 학습을 시킨 다음, 가장 뛰어난 성능을 내는 신경망을 최종 선택하는 기법입니다.

NAS를 이용하여 찾아낸 신경망 구조 중 대표적인 것이 구글에서 발표한 EfficientNet입니다. MBConv라고 표시된 블록은 Mobile Inverted Bottleneck Convolution의 약자로, mobilenet에서 제안한 depthwise separable convolution과 resnet에서 제안한 residual 개념을 결합시킨 블록입니다. 가볍고 성능이 뛰어나기 때문에 많이 사용됩니다.

## 이 후의 CNN 아키텍처들



2021년도에 EfficientNet 발표 이후 꾸준히 새로운 CNN 아키텍처들이 발표되고 있지만, 그다지 주목을 받지 못하였습니다. 대신에 Transformer를 컴퓨터 비전 분야에 적용한 Vision Transformer가 선풍적인 인기를 끌면서, CNN이 연구자들의 관심에서 멀어지게 됩니다. 그래도 여전히 컴퓨터 비전 분야에서는 방대한 데이터를 미리 학습시켜 놓은 CNN 모델들을 사용하는 경우가 많으며, 활용도가 높은 모델입니다.

## 정리

이번 챕터에서는 CNN 아키텍처들의 변천사를 간단히 살펴보았습니다. 하나하나 자세히 들여다 보면 새로운 아이디어가 제안되어 기존에 막혀있었던 문제가 해결되는 과정이 상당히 흥미롭습니다. 관심이 있으신 분들은 ResNet 논문을 한번 읽어보시는 걸 추천드리면서 마치겠습니다.