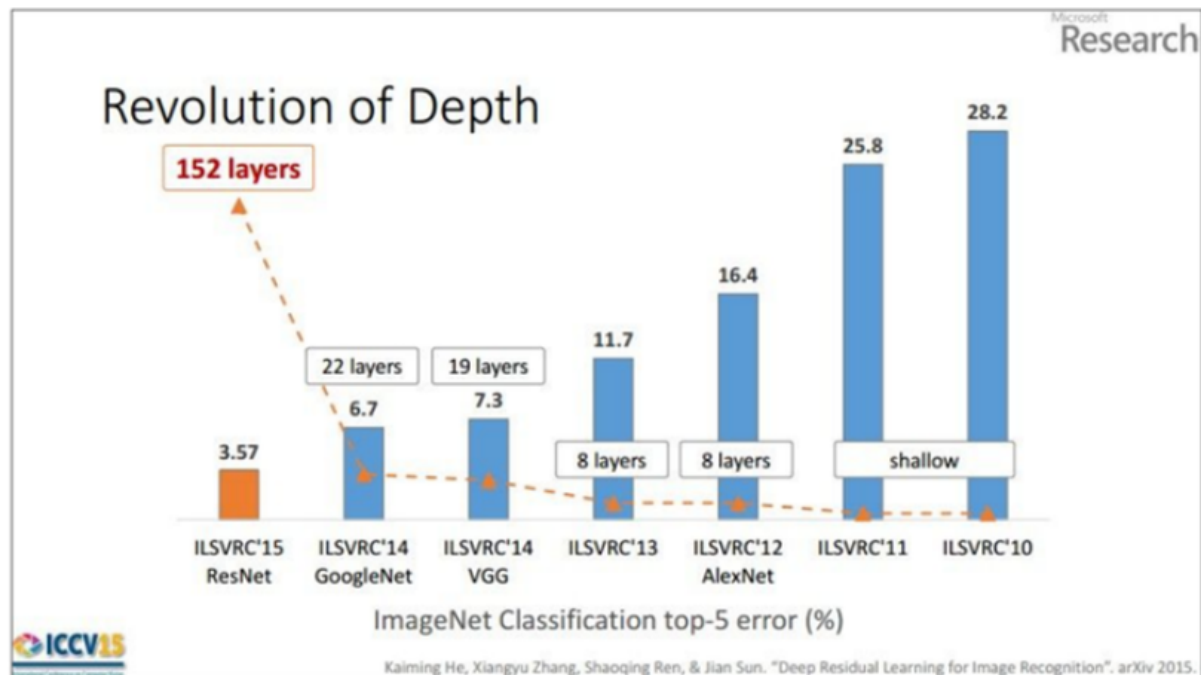


VGGNet

ImageNet Large Scale Visual Recognition Challenge(ILSVRC)에서 14년도 준우승을 한 모델이다. 여기서 말하는 VGGNet은 16개 또는 19개의 층으로 구성된 모델을 의미한다. (VGG16,VGG19)

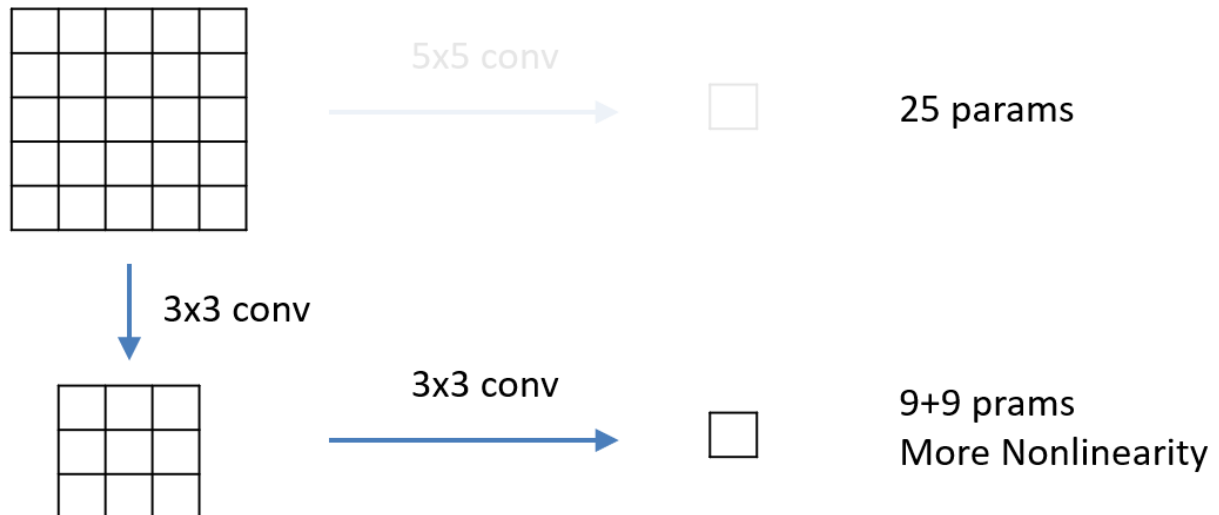


네트워크가 깊어질수록 성능이 좋아졌음을 위 그림을 통해 확인할 수 있다.

VGGNet은 사용하기 쉬운 구조와 좋은 성능 덕분에 그 대회에서 우승을 거둔 조금 더 복잡한 형태의 GoogLeNet보다 더 인기를 얻었다.

VGGNet의 구조

VGGNet의 original 논문의 개요에서 밝히고 있듯이 이 연구의 핵심은 네트워크의 깊이를 깊게 만드는 것이 성능에 어떤 영향을 미치는지를 확인하고자 한 것이다. VGG 연구팀은 깊이의 영향만을 최대한 확인하고자 컨볼루션 필터커널의 사이즈는 가장 작은 3 x 3으로 고정했다.



filter의 크기가 크게 좋을까? 작은게 좋을까?

filter의 크기가 크면 더 많은 local feature를 나타낼 수 있으므로 장점이있는데 반면 paramater가 너무 많아지고, 작은 filter로 반복하면 비선형성을 더 증가시킨다.

따라서 작은크기의 filter가 더 유용하다.

INPUT: [224x224x3] memory: 224*224*3=150K params: 0 (not counting biases)

CONV3-64: [224x224x64] memory: 224*224*64=3.2M params: (3*3*3)*64 = 1,728

CONV3-64: [224x224x64] memory: 224*224*64=3.2M params: (3*3*64)*64 = 36,864

POOL2: [112x112x64] memory: 112*112*64=800K params: 0

CONV3-128: [112x112x128] memory: 112*112*128=1.6M params: (3*3*64)*128 = 73,728

CONV3-128: [112x112x128] memory: 112*112*128=1.6M params: (3*3*128)*128 = 147,456

POOL2: [56x56x128] memory: 56*56*128=400K params: 0

CONV3-256: [56x56x256] memory: 56*56*256=800K params: (3*3*128)*256 = 294,912 C

ONV3-256: [56x56x256] memory: 56*56*256=800K params: (3*3*256)*256 = 589,824 CO

NV3-256: [56x56x256] memory: 56*56*256=800K params: (3*3*256)*256 = 589,824

POOL2: [28x28x256] memory: 28*28*256=200K params: 0

CONV3-512: [28x28x512] memory: 28*28*512=400K params: (3*3*256)*512 = 1,179,648 C

ONV3-512: [28x28x512] memory: 28*28*512=400K params: (3*3*512)*512 = 2,359,296 CO

NV3-512: [28x28x512] memory: 28*28*512=400K params: (3*3*512)*512 = 2,359,296

POOL2: [14x14x512] memory: 14*14*512=100K params: 0

CONV3-512: [14x14x512] memory: 14*14*512=100K params: (3*3*512)*512 = 2,359,296 C

ONV3-512: [14x14x512] memory: 14*14*512=100K params: (3*3*512)*512 = 2,359,296 CO

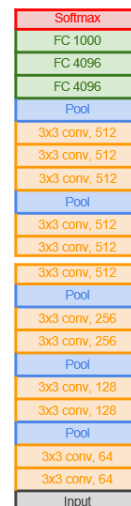
NV3-512: [14x14x512] memory: 14*14*512=100K params: (3*3*512)*512 = 2,359,296

POOL2: [7x7x512] memory: 7*7*512=25K params: 0

FC: [1x1x4096] memory: 4096 params: 7*7*512*4096 = 102,760,448 F

C: [1x1x4096] memory: 4096 params: 4096*4096 = 16,777,216

FC: [1x1x1000] memory: 1000 params: 4096*1000 = 4,096,000



VGG16

TOTAL memory: 24M * 4 bytes ~ 96MB / image (only forward! ~*2 for bwd)

TOTAL params: 138M parameters