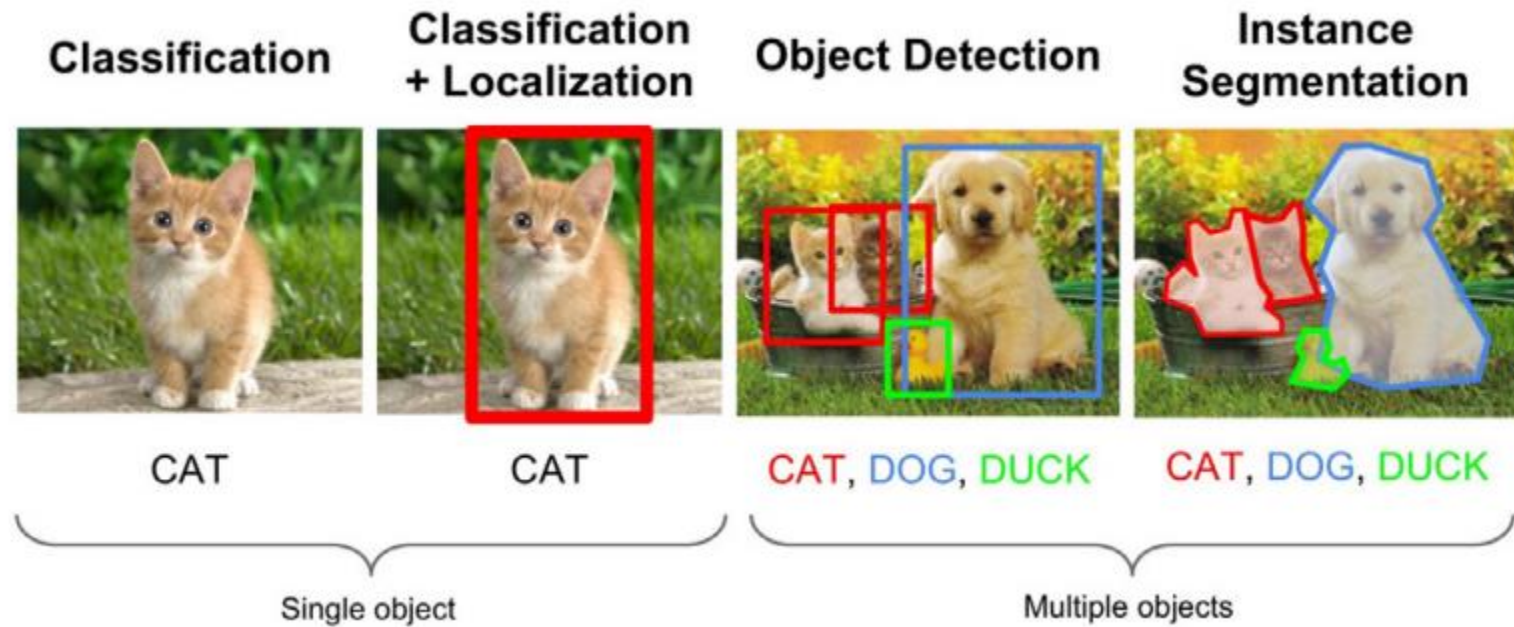


CNN

(AlexNet, VGGNet, GoogleNet, ResNet,
SENet, NASNet, AmoebaNet)

서영민



객체 검출에 사용되는 CNN은 주로 이미지 분류에서 사용한 CNN과 동일한 구조를 사용

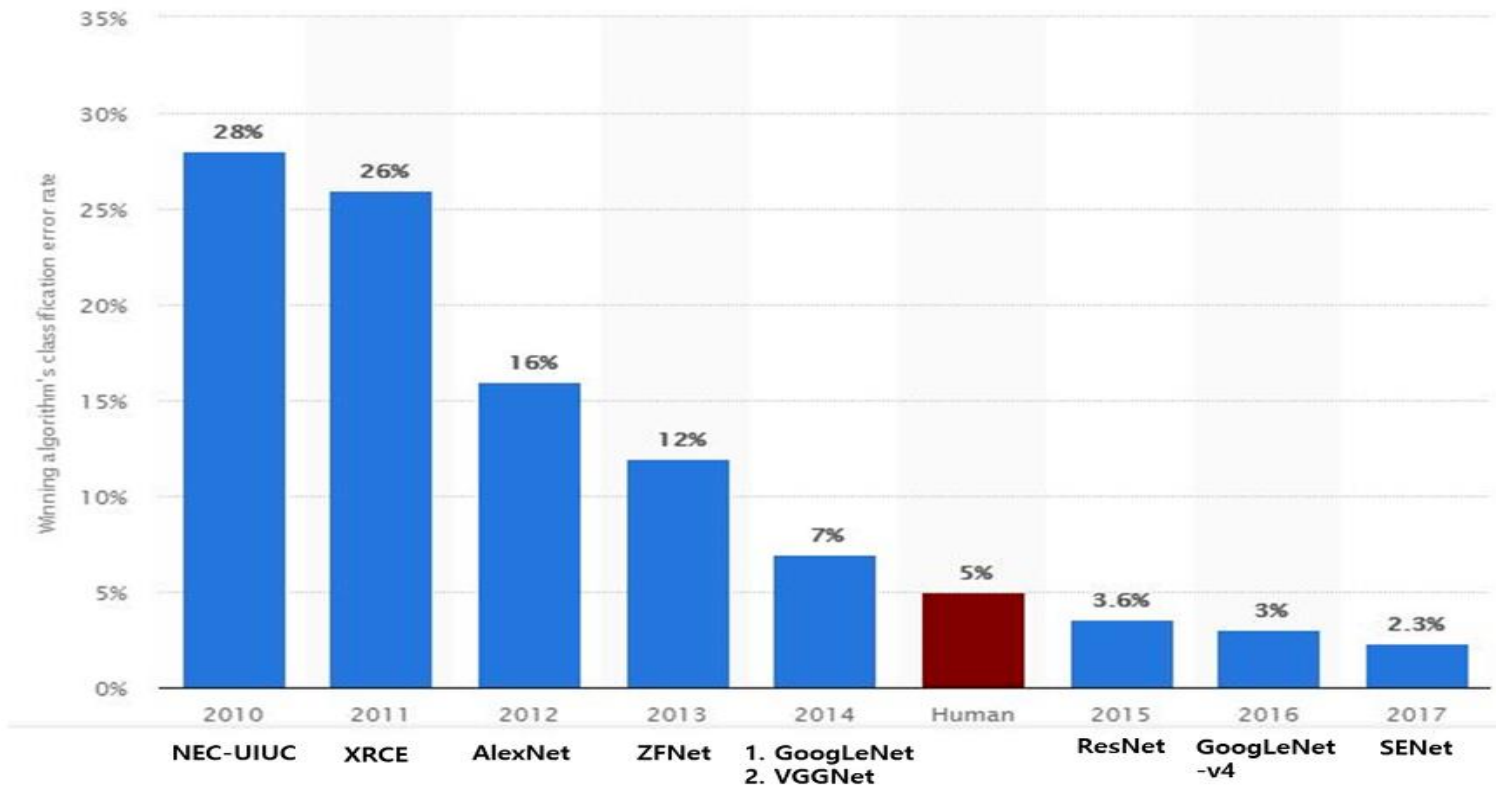


CNN의 성능이 객체 검출 성능에 많은 영향을 줌

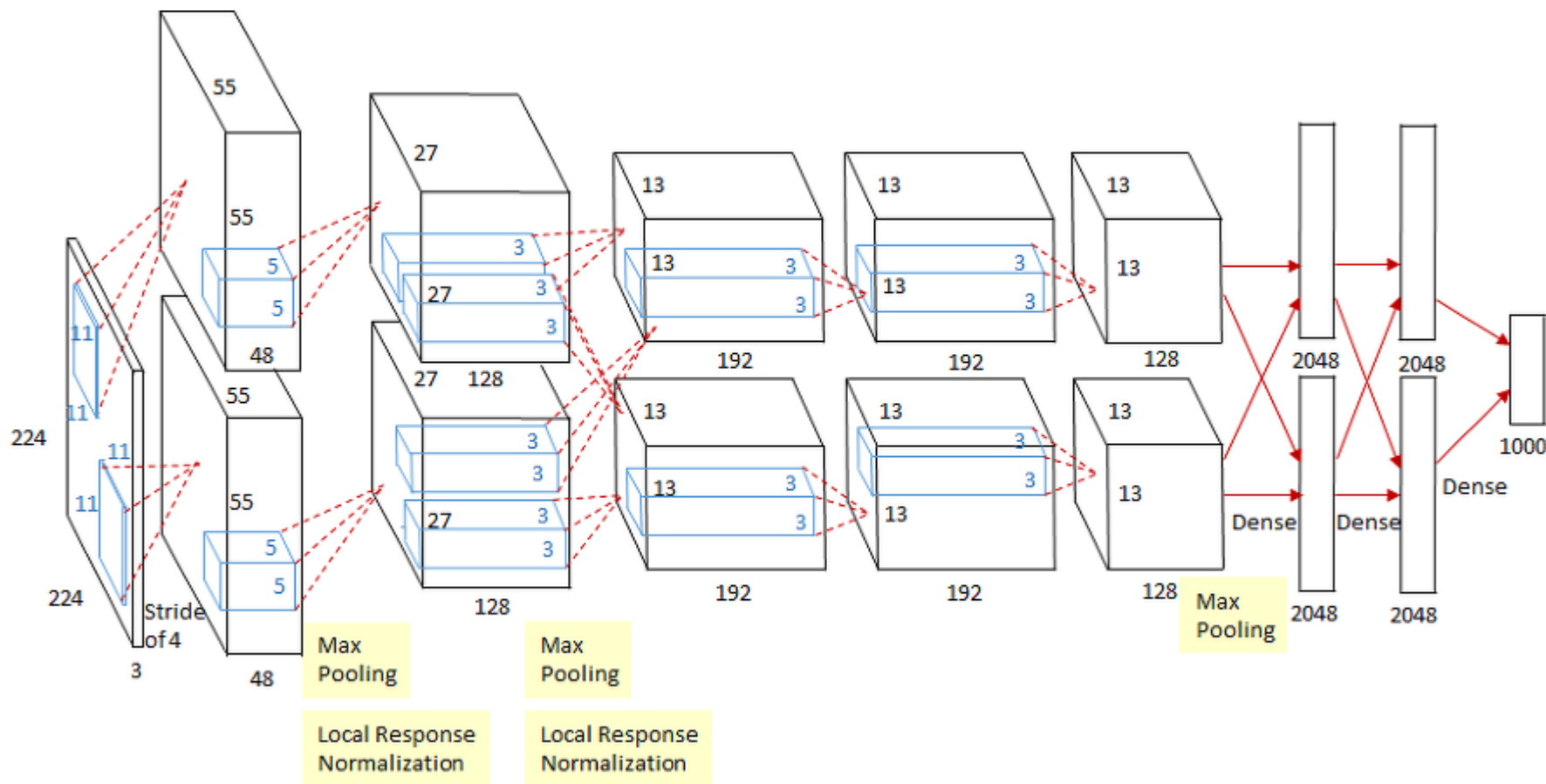
ILSVRC

ImageNet Large Scale Visual Recognition Challenge ~17

대용량의 이미지셋에 대해 이미지 인식 알고리즘의 성능을 평가하는 대회



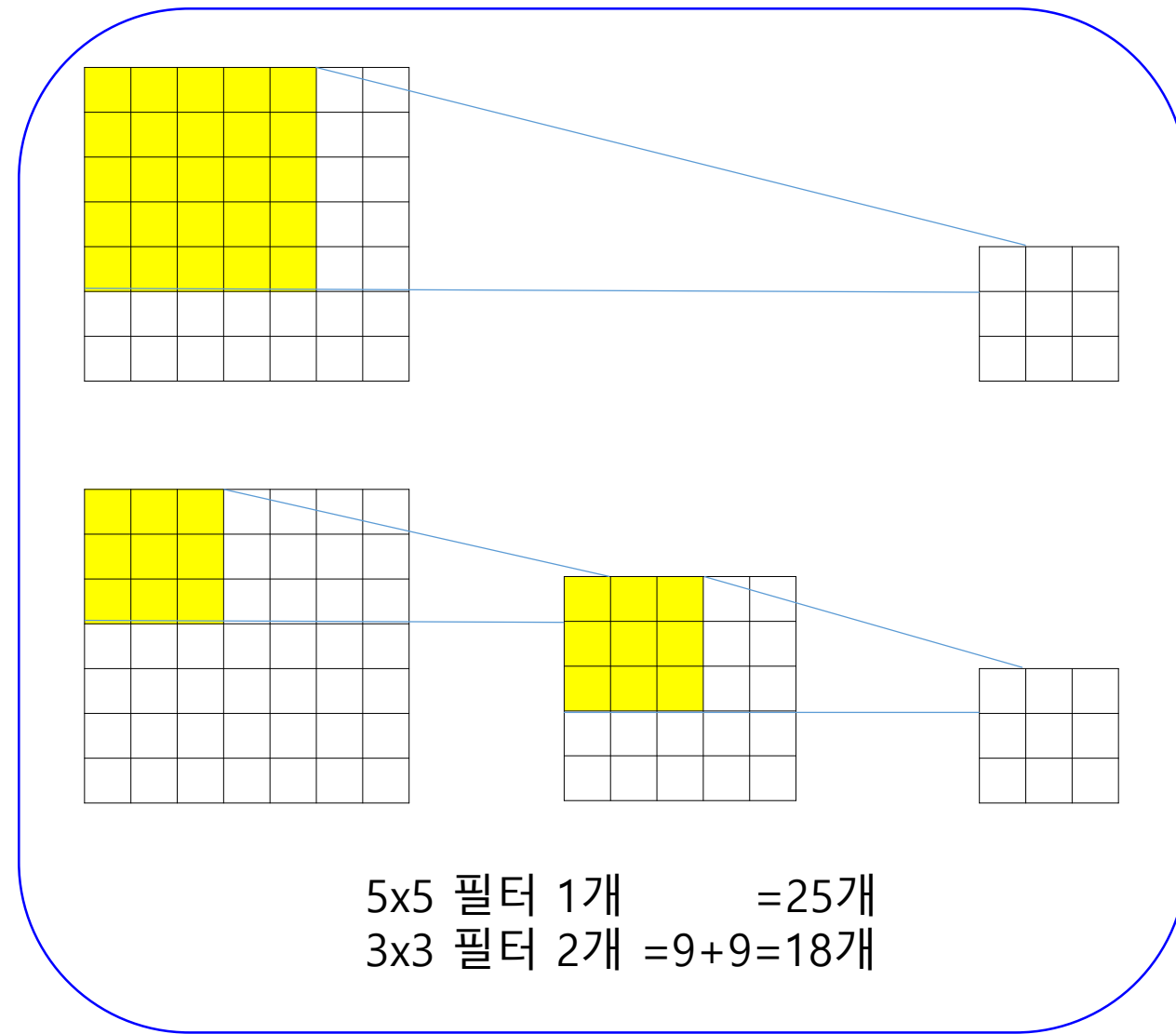
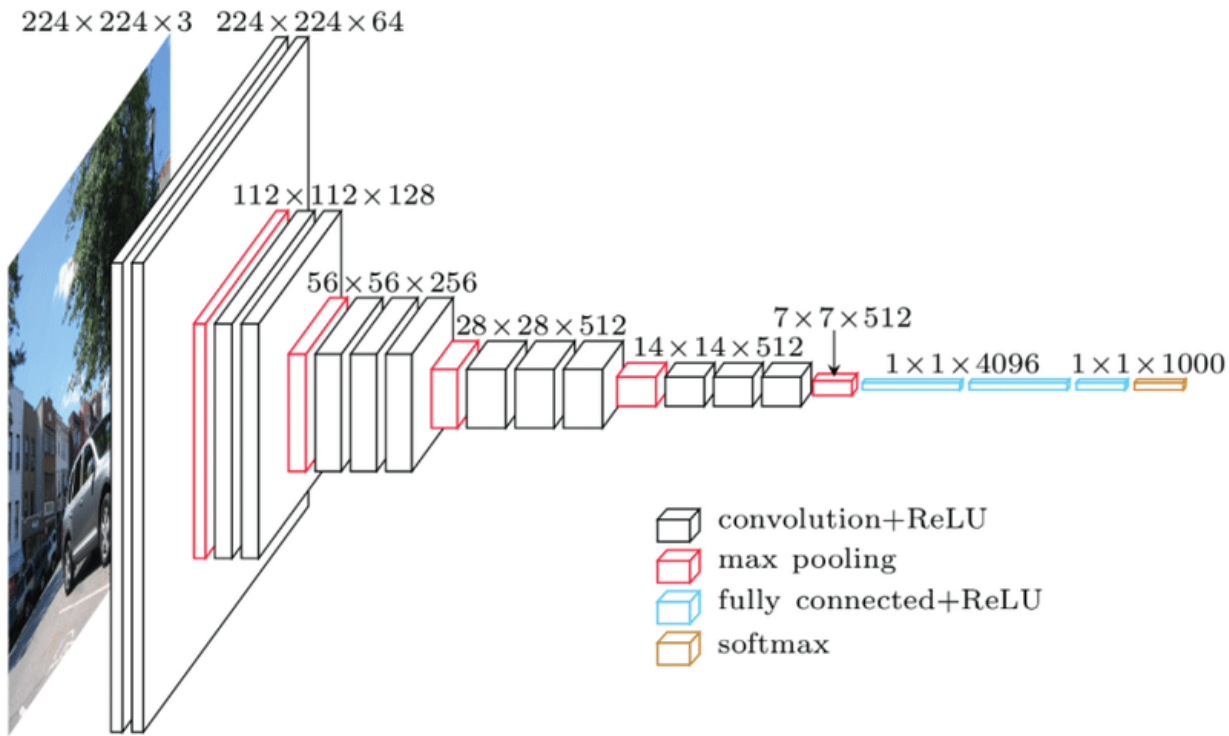
AlexNet



[제안내용]

1. 이미지 분류를 위해 **딥러닝을 사용**하였으며 좋은 성능을 낼 수 있음을 보여줌
2. **데이터 증강**이라는 테크닉을 사용
3. **드롭아웃** 기법 적용
4. 컨볼루션+풀링+FC레이어 구조 → **기본적인 CNN 구조로 영향을 끼침**

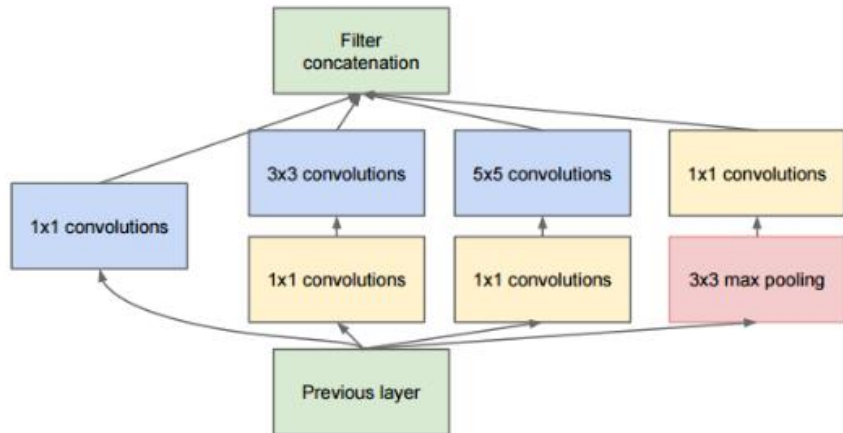
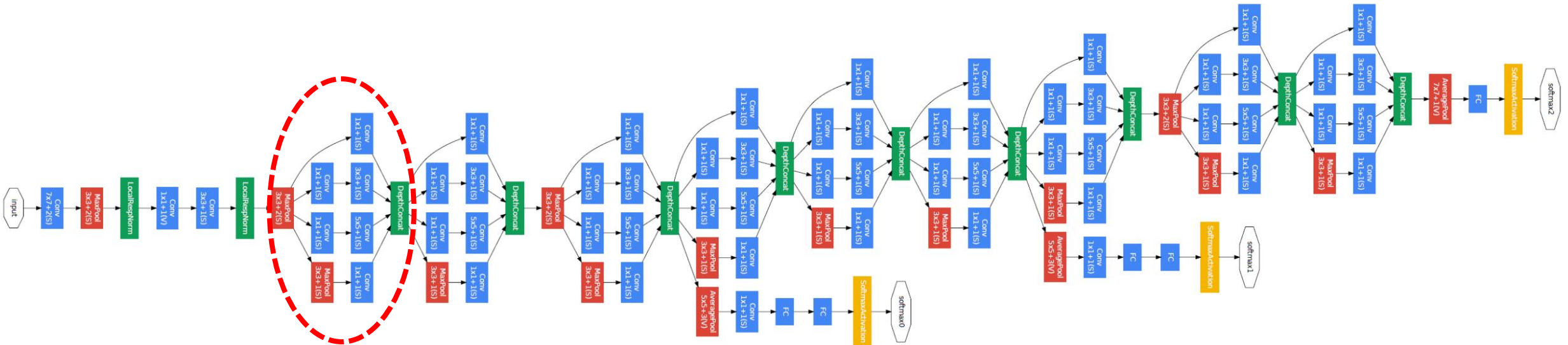
VGGNet



[제안내용]

1. **ReLU**를 사용하여 비선형성을 높임
2. **3x3 컨볼루션을 연속적으로 사용한다**
 - 5x5 크기 레이어의 수용 범위와 유사하면서 파라미터수는 더 적게 가져갈 수 있다.
 - 단순한구조, 성능우수
3. 파라미터수가 너무 많다.

GoogleNet



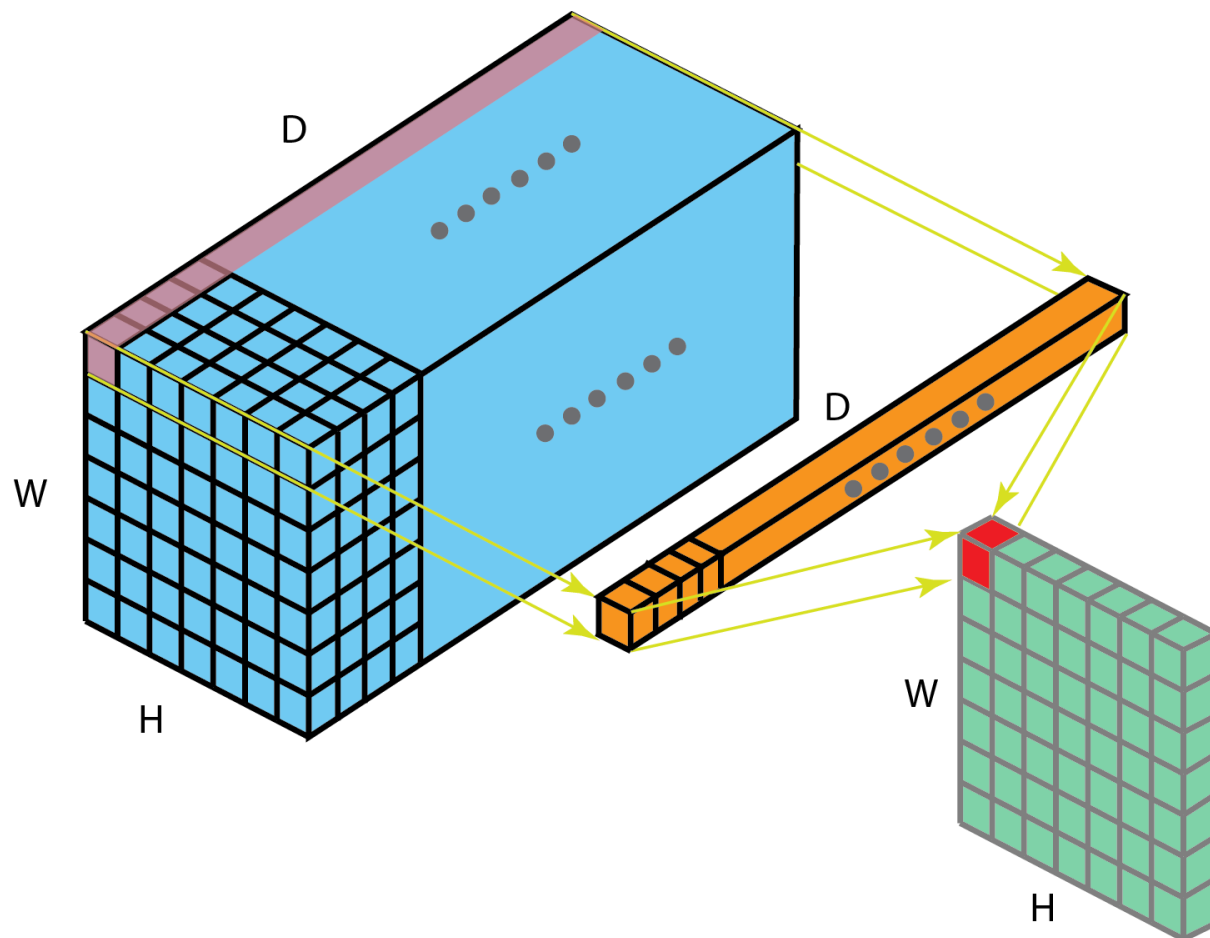
Inception module with dimension reductions

일반적인 컨볼루션 : W,H은 줄이고, C는 늘어남

“연산량을 줄이면서 같은 효과를 어떻게 낼까?”

C은 1x1 컨볼루션으로, W,H는 Max Pooling으로 조절하자

GoogleNet



VGG16 param: 4,096,000

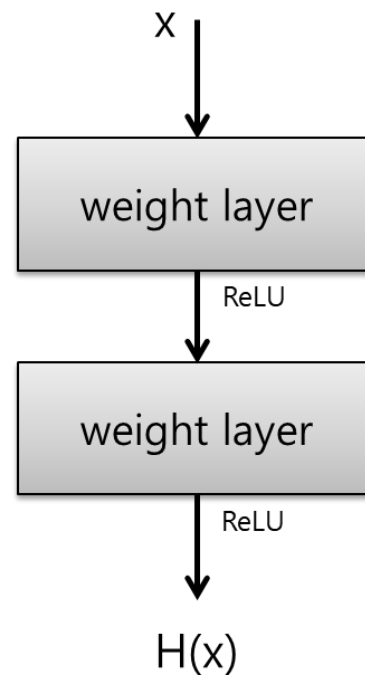
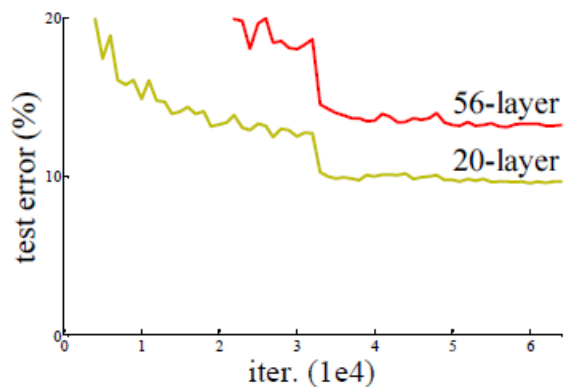
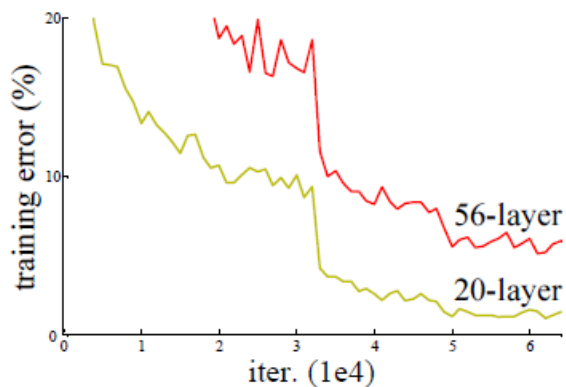
GoogleNet param: 1,388,000

ResNet

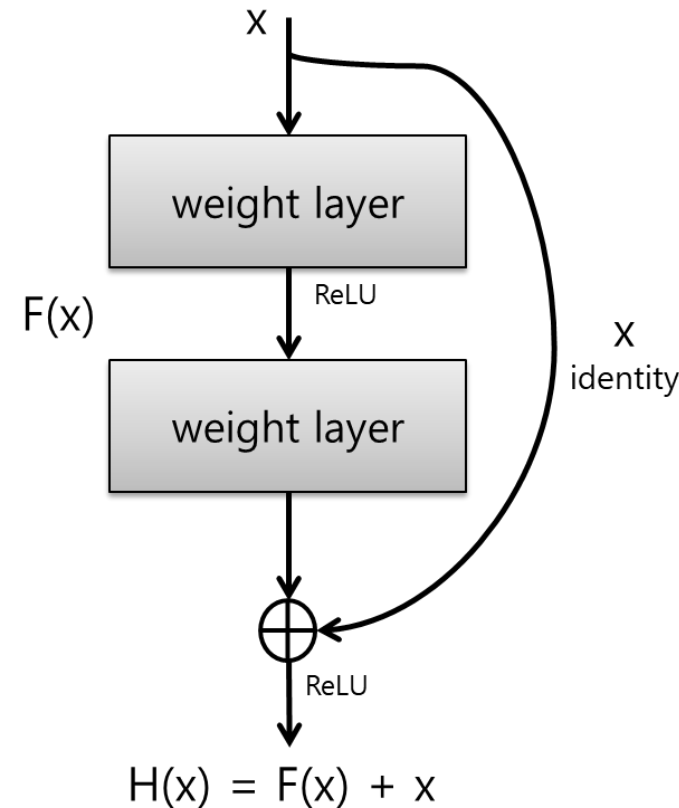
ResNet-18, 34, 50, 101, **152**, 1002 layer

인간인식 에러 : 5%

ResNet 에러 : 3.57%



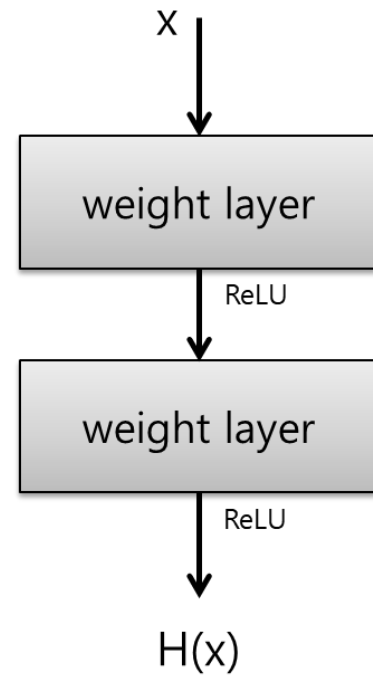
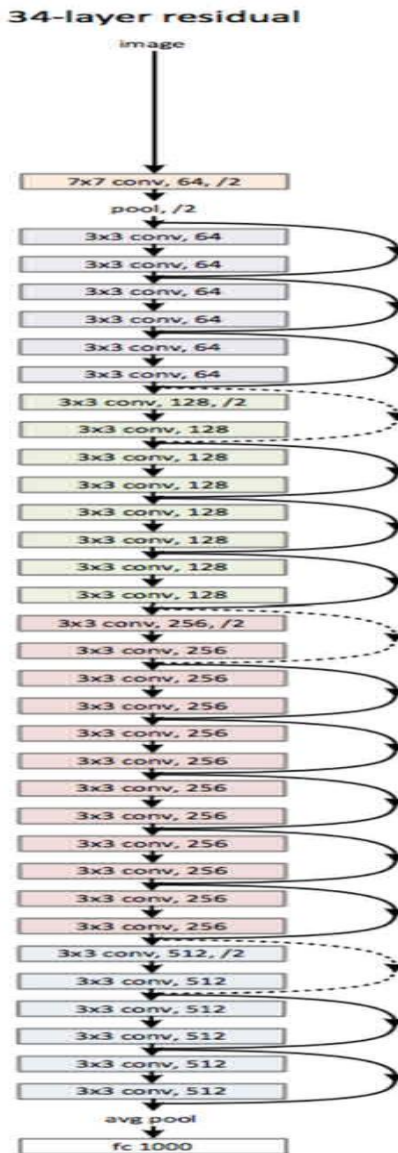
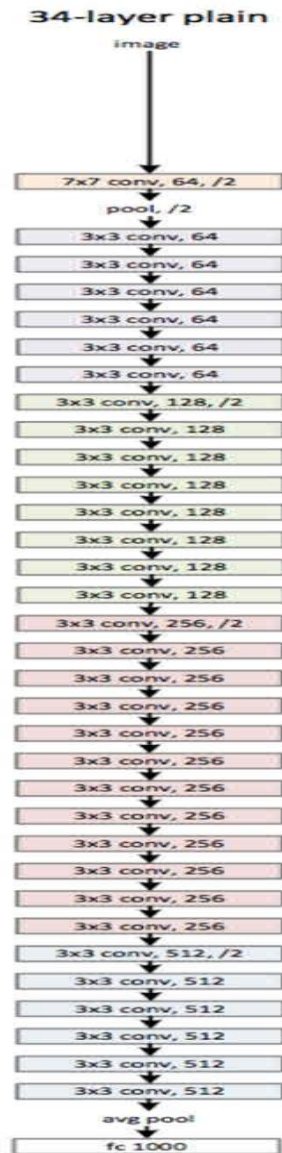
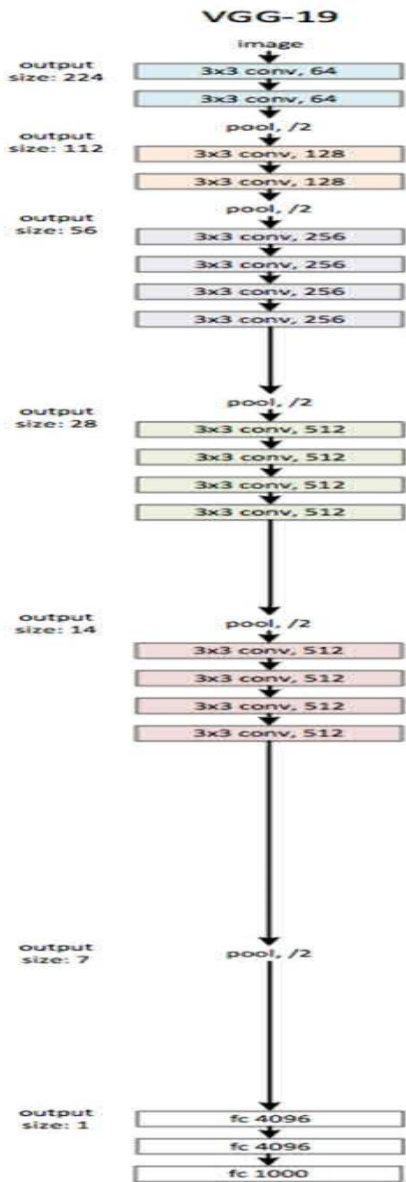
기존 방식



Residual block

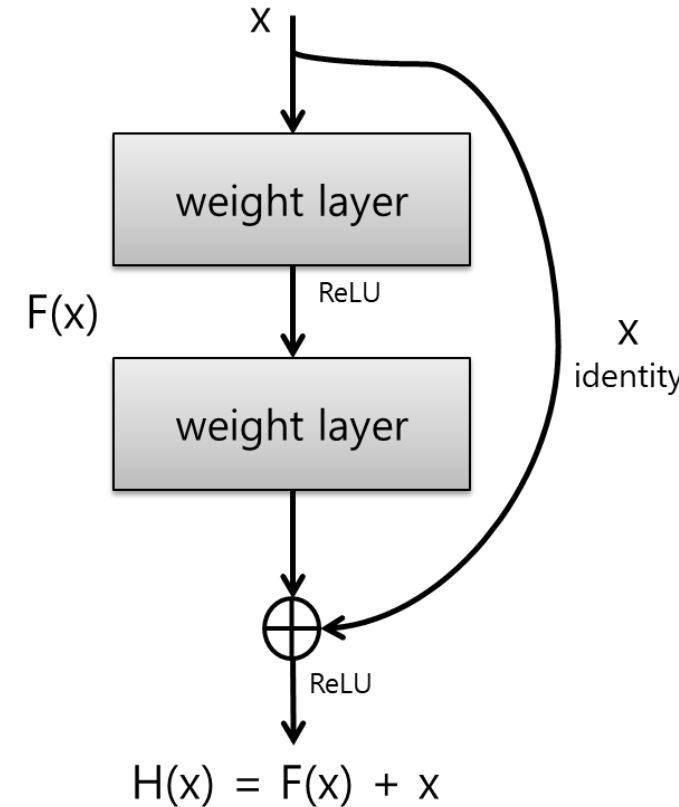
"네트워크는 구조가 깊고 넓을수록 성능이 좋다" → 파라미터가 많아짐 → 무조건 깊게 해도 잘 안나옴
"새로운 방법....?!"
"Skip connection"

ResNet



기존 방식

$H(x) - y = 0$
목표 : $H(x) = y$



Residual block

$H(x) = F(x) + x$
목표 : $F(x) = H(x) - x = 0$
 $H(x) = x$

SENet

Squeeze-and-Excitation Block

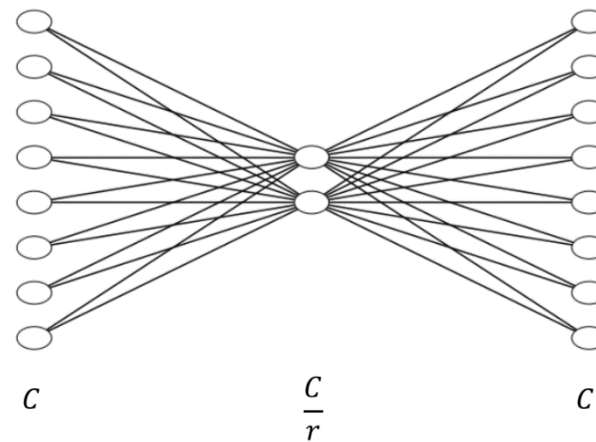
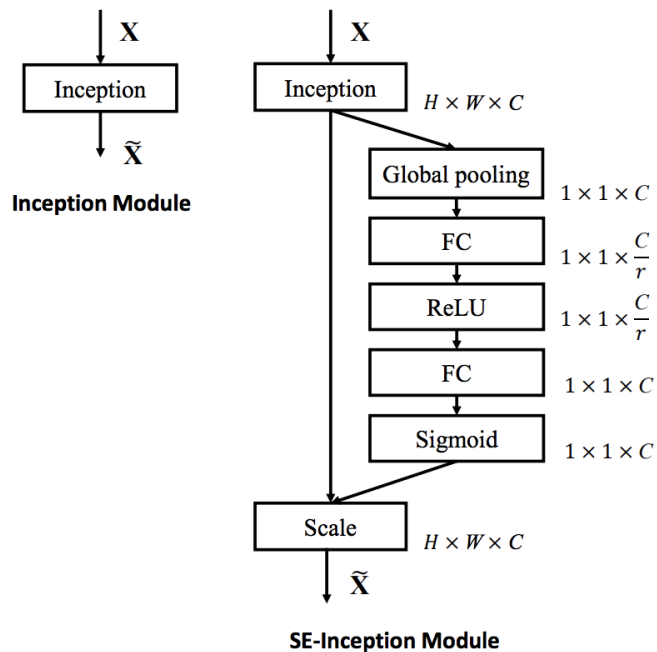
기존 GoogleNet(Inception)이나 ResNet에 바로 적용 가능

피쳐맵의 차원을 Global Average Pooling으로 축소(Squeeze)

➡ 각 피쳐맵의 정보를 요약

FC레이어를 통해 어떤 피쳐에 가중치를 둘지 정하고
Sigmoid로 비선형성을 높임(Excitation: 자극)

➡ 각 피쳐맵의 중요도를 정함

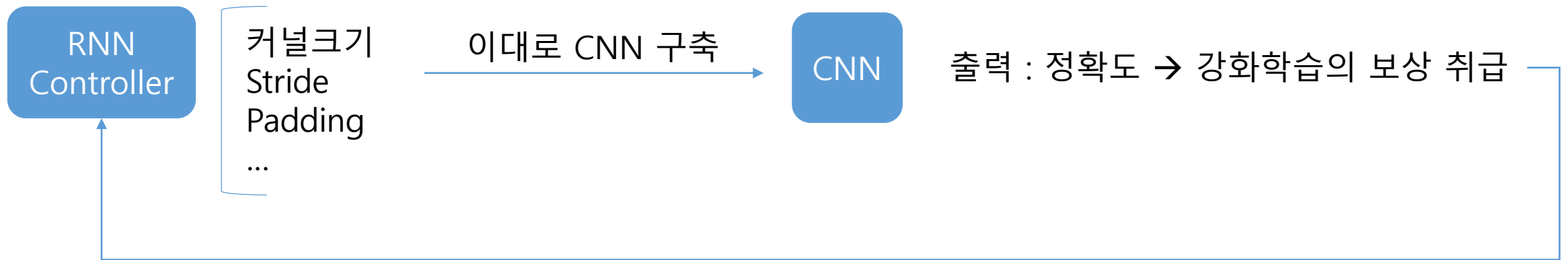


NASNet

AutoML 기반

- 실험, 경험에서 나온 layer 구조
- Hyperparameter : 학습에 큰 영향을 주는 요인(LearningRate, BatchSize..)

➡ 학습을 통해 획득



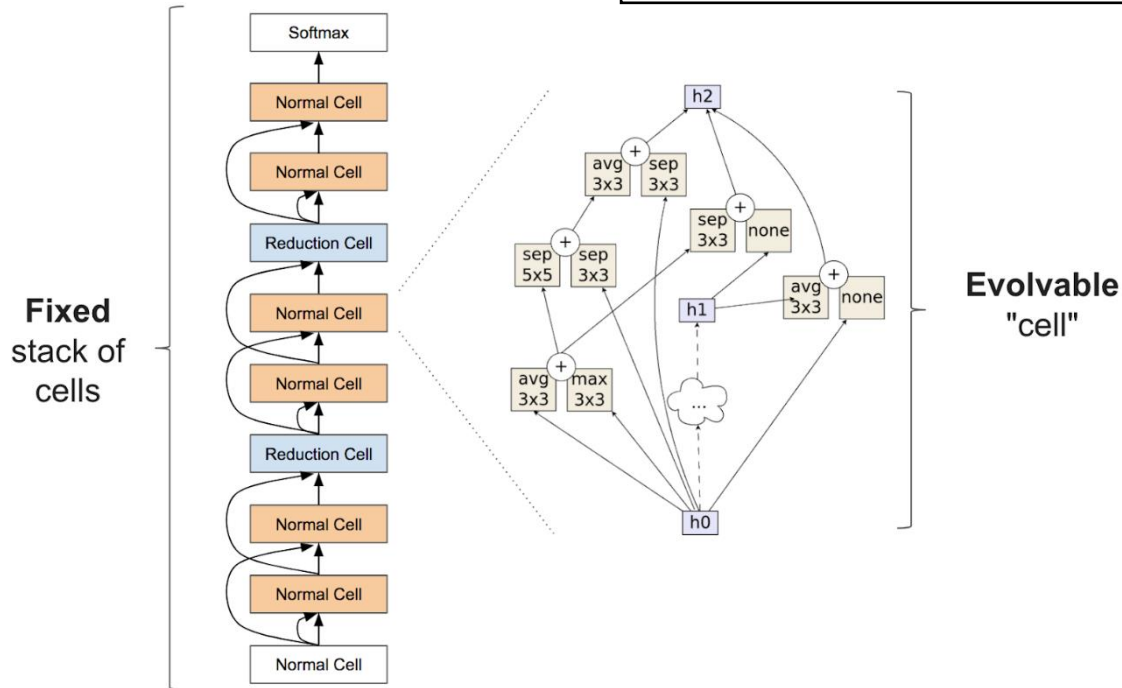
CIFAR-10 : 5만장으로 10개 분류 → 800대의 GPU로 28일

➡ (개선)
Cell 이라는 단위로 레이어 조합에 필요한 경우의 수를 줄임 ➡ 500대의 GPU로 4일

AmoebaNet

진화알고리즘(Evolutionary Algorithm) : 필요한 요소만 진화되고, 불필요한것은 퇴화

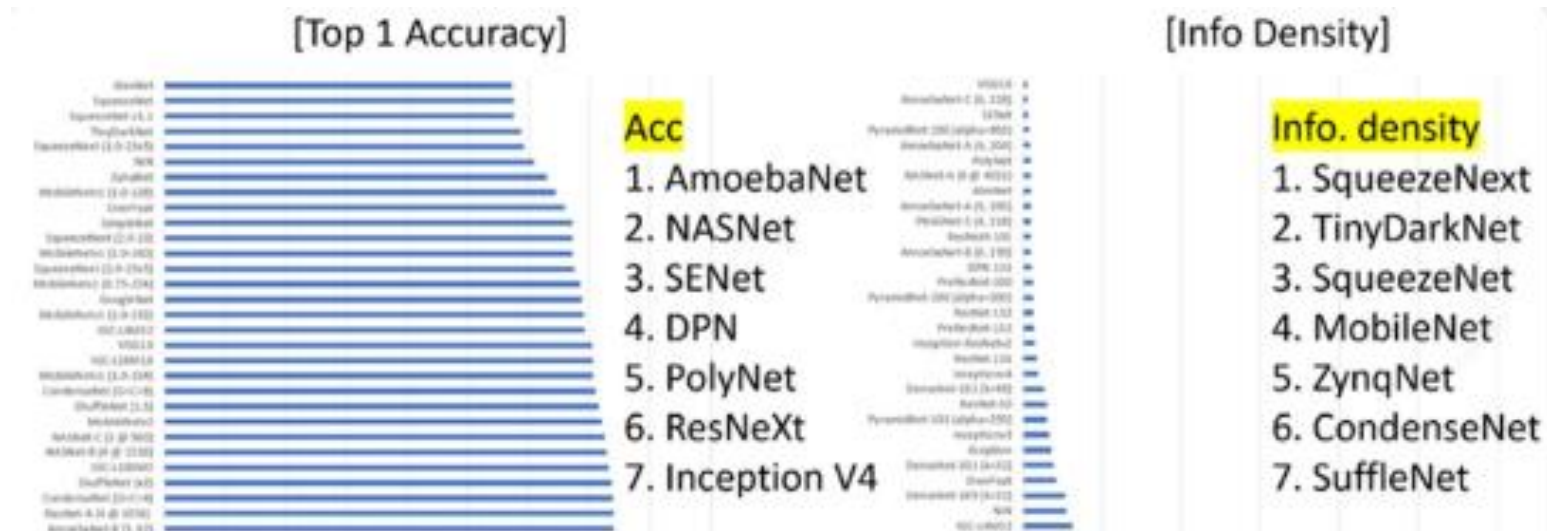
→ 천개의 간단한 모델집합에서 무작위로 한 쌍을 선택, 높은 정확도를 가진 신경망이 선택되고, 복제 모델을 만들어 변이 시킨 뒤, 다시 천개의 모델 집합에 추가



연결 해제
레이어 사이 skip connection
학습 속도 변경
...

반복할 경우 모델의 집합 개수는 그대로지만, 우수한 신경망들로만 이루어져 있다.

END



동일한 데이터셋 기반 실험(18)
Accuracy(좌), 연산량(우)

Reference

AlexNet : A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097-1105, 2012.

VGGNet : K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

GoogleNet : C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In CVPR, 2015.

ResNet : K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In IEEE Conference on Computer Vision and Pattern Recognition, 2016.

SENet : J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. arXiv preprint arXiv:1709.01507, 2017.

NASNet : B. Zoph, V. Vasudevan, J. Shlens and Q. V. Le. Learning Transferable Architectures for Scalable Image Recognition. In CVPR, 2018.

AmoebaNet : E. Real, A. Aggarwal, Y. Huang, and Q. V Le. Regularized evolution for image classifier architecture search. arXiv preprint arXiv:1802.01548, 2018.