Chapter 08. 좋은 딥러닝 구조를 찾아내는 딥러닝 (Neural Architecture Search)

자동 모델 구조 최적화

NAS Methods

Reinforcement Learning (RL)

Evolutionary Algorithm (EA)

Differentiable Architecture Search (DARTS)

다양한 종류의 NAS 방법을 소개하고자 한다.

NAS

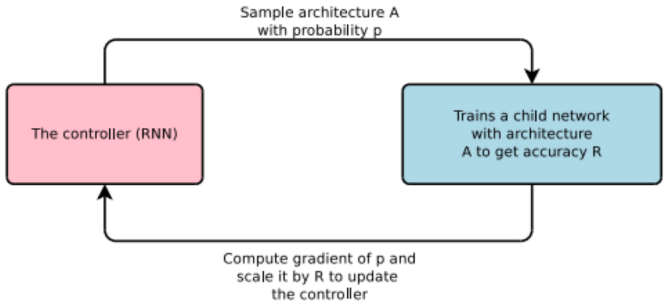
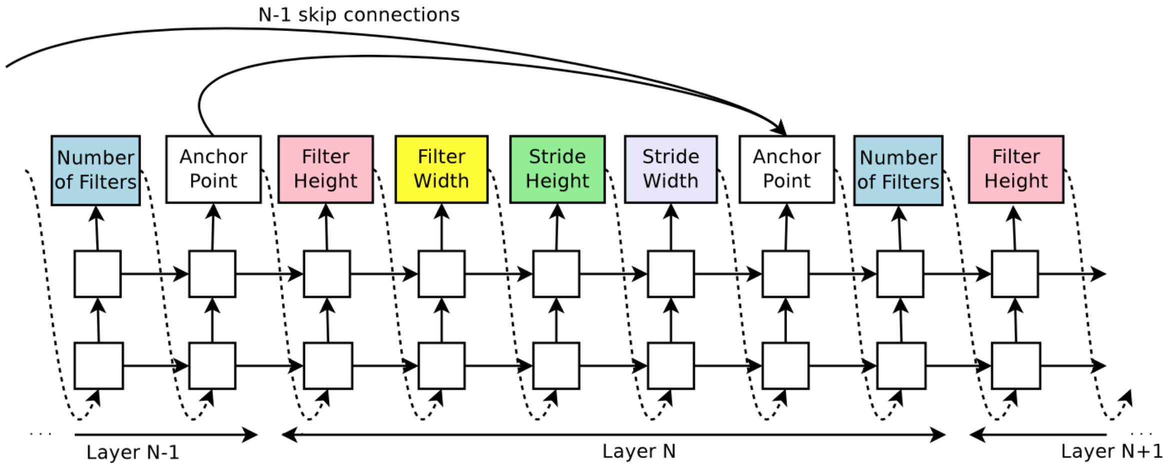


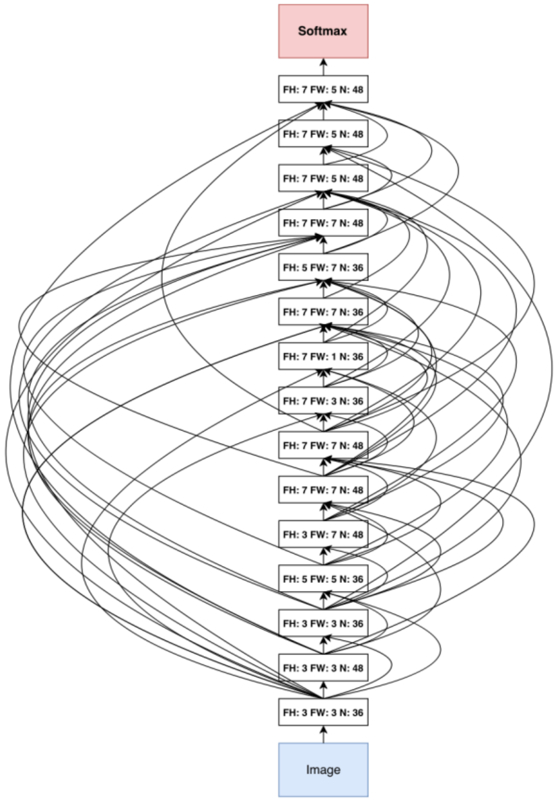
Figure 1: An overview of Neural Architecture Search.



Google Brain에서 제안한 첫 NAS 논문. RNN Controller를 이용한 RL 방법이다.

https://openreview.net/pdf?id=r1Ue8Hcxg

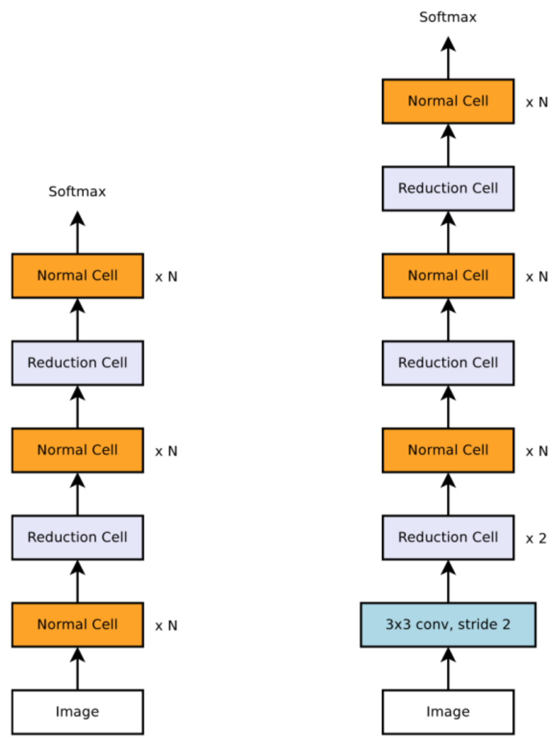
NAS



NAS로 도출한 Optimum Image Classifier. GPU Time으로 약 1000day가 들어갔다.

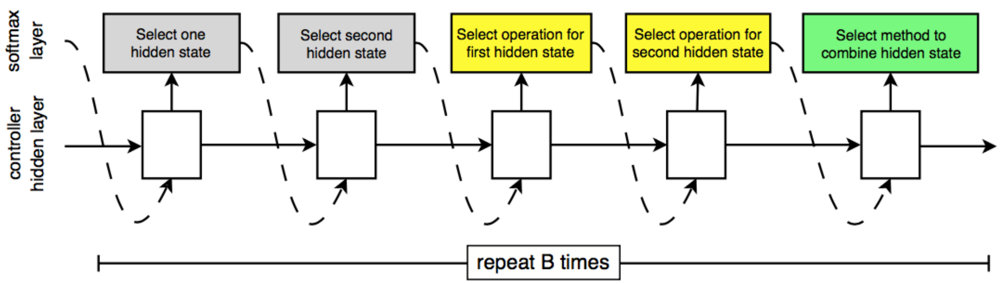
https://openreview.net/pdf?id=r1Ue8Hcxg

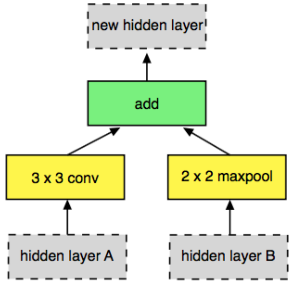
NASNet



CIFAR10 Architecture

ImageNet Architecture



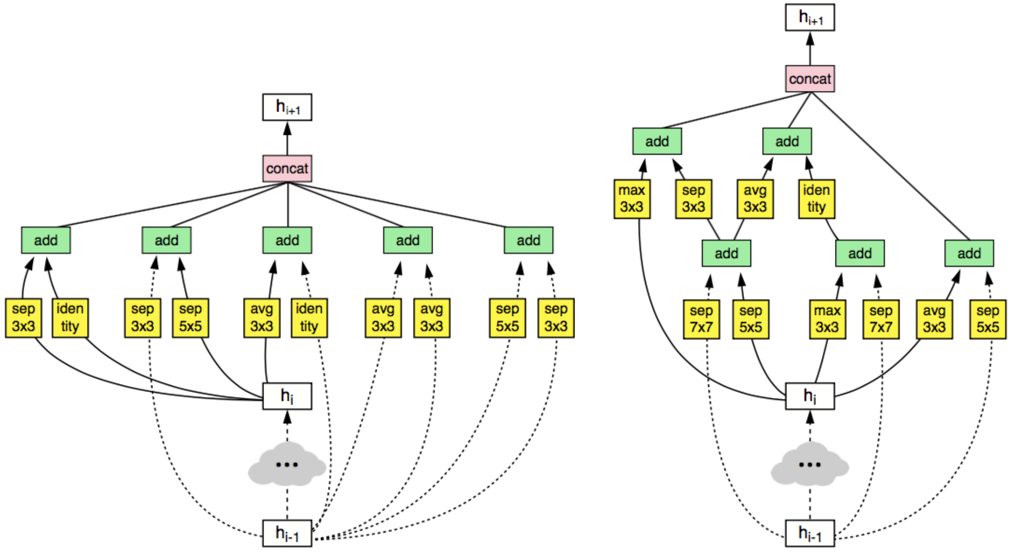


- identity
- 1x7 then 7x1 convolution
- 3x3 average pooling
- 5x5 max pooling
- 1x1 convolution
- 3x3 depthwise-separable conv
- 7x7 depthwise-separable conv
- 1x3 then 3x1 convolution
- 3x3 dilated convolution
- 3x3 max pooling
- 7x7 max pooling
- 3x3 convolution
- 5x5 depthwise-seperable conv

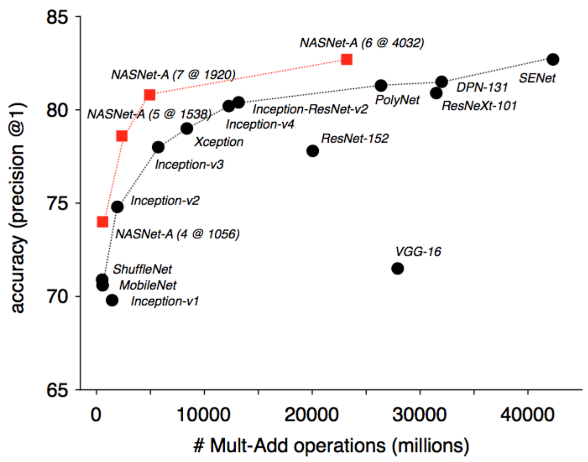
Normal Cell과 Reduction Cell로 나누고, 자주 쓰이는 Operation을 후보로 두었다.

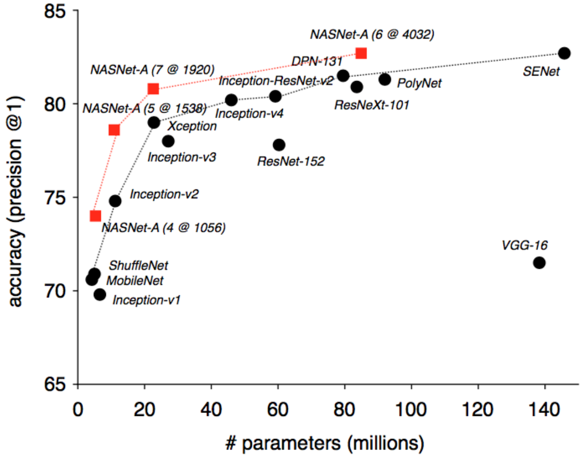
https://arxiv.org/pdf/1707.07012.pdf

NASNet



Normal Cell Reduction Cell

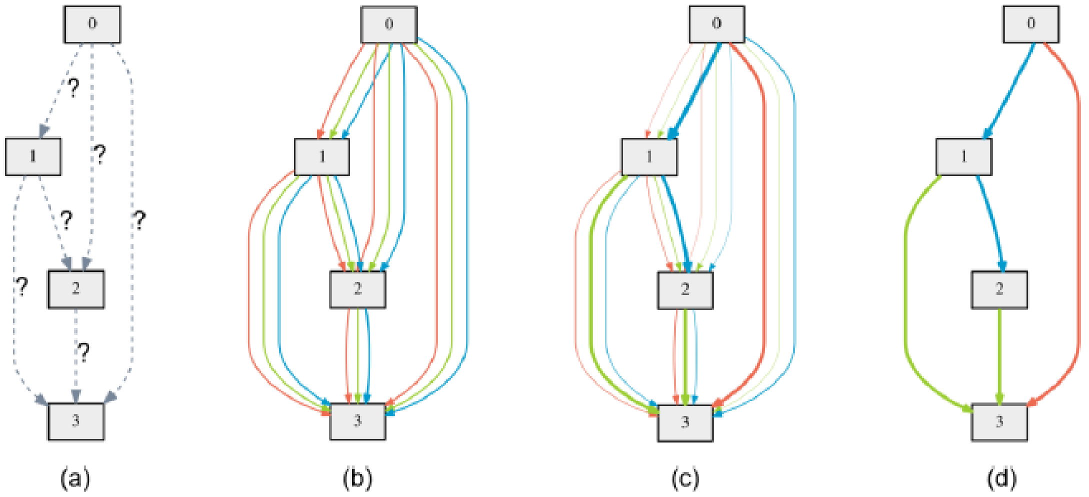




NASNet으로 찾은 구조로, ImageNet에서 SOTA의 성능을 보여주었다.

https://arxiv.org/pdf/1707.07012.pdf

DARTS



$$\min_{\alpha} \quad \mathcal{L}_{val}(w^*(\alpha), \alpha)$$
$$\text{s.t.} \quad w^*(\alpha) = \text{argmin}_w \; \mathcal{L}_{train}(w, \alpha)$$

Mixed Operation의 가중치를 이용해, Child network의 선택이 미분 가능하게 한다.

https://arxiv.org/abs/1806.09055

DARTS

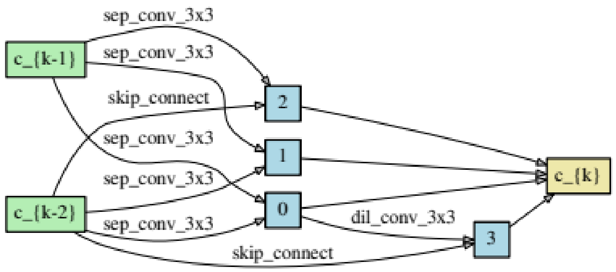


Figure 4: Normal cell learned on CIFAR-10.

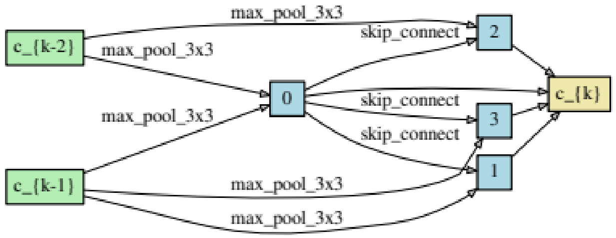


Figure 5: Reduction cell learned on CIFAR-10.

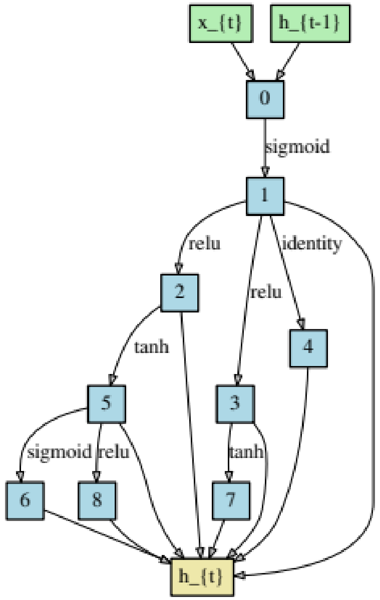


Figure 6: Recurrent cell learned on PTB.

DARTS에 의해 학습된 Cell Architectures

https://arxiv.org/abs/1806.09055

DARTS

Architecture	Test Error (%)	Params (M)	Search Cost (GPU days)	Search Method
DenseNet-BC (Huang et al., 2017)	3.46	25.6	–	manual
NASNet-A + cutout (Zoph et al., 2017)	2.65	3.3	1800	RL
NASNet-A + cutout (Zoph et al., 2017)†	2.83	3.1	3150	RL
AmoebaNet-A + cutout (Real et al., 2018)	3.34 ± 0.06	3.2	3150	evolution
AmoebaNet-A + cutout (Real et al., 2018)†	3.12	3.1	3150	evolution
AmoebaNet-B + cutout (Real et al., 2018)	2.55 ± 0.05	2.8	3150	evolution
Hierarchical Evo (Liu et al., 2017b)	3.75 ± 0.12	15.7	300	evolution
PNAS (Liu et al., 2017a)	3.41 ± 0.09	3.2	225	SMBO
ENAS + cutout (Pham et al., 2018b)	2.89	4.6	0.5	RL
Random + cutout	3.49	3.1	–	–
DARTS (first order) + cutout	2.94	2.9	1.5	gradient-based
DARTS (second order) + cutout	2.83 ± 0.06	3.4	4	gradient-based

GPU Hour 대비 매우 좋은 성능이 나타남을 알 수 있다.

https://arxiv.org/abs/1806.09055