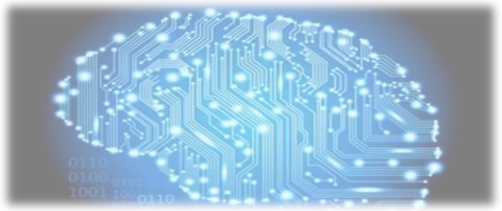
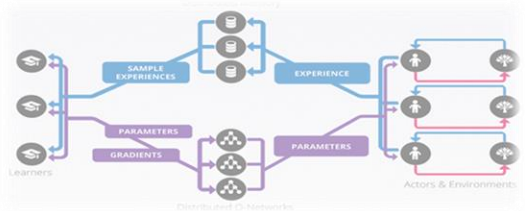



ResNeXt 구현 및 실험 결과 보고

AutoML 연구실
석사과정 문아성
20.09.22

Introduction

Aggregated Residual Transformations for Deep Neural Networks

2017 CVPR Conference “Aggregated Residual Transformations for Deep Neural Networks” 라는 연구 결과 발표

2016 년 말 arXiv로 공개되어, 2016년 ILSVRC 대회에서 2등을 수상

“Cardinality” 개념 도입

이전 논문인 ResNet 과 더 정확한 비교를 위해

FLOPs complexity를 유지한 상태로 소개

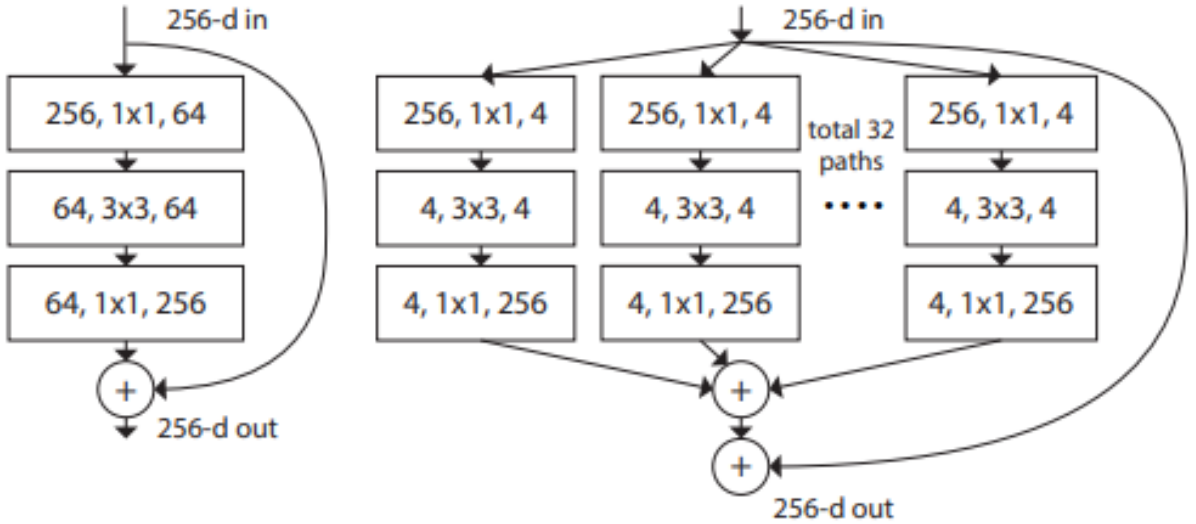


Figure 1. **Left**: A block of ResNet [14]. **Right**: A block of ResNeXt with cardinality = 32, with roughly the same complexity. A layer is shown as (# in channels, filter size, # out channels).

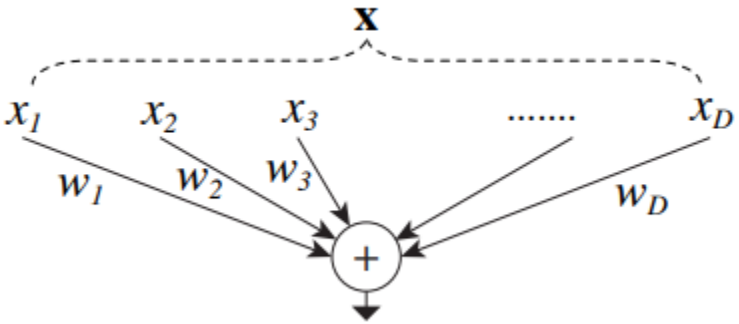
Aggregated Residual Transformations for Deep Neural Networks (1)

Aggregated Transformations

(i) *Splitting*: the vector x is sliced as a low-dimensional embedding.
it is a single-dimension subspace x_i.

(ii) *Transforming*: the low-dimensional representation is transformed. => $w_i x_i$

(iii) *Aggregating*: the transformations in all embeddings are aggregated by $\sum_{i=1}^{D} w_i x_i$



A simple neuron that performs inner product.

$$\mathcal{F}(\mathbf{x}) = \sum_{i=1}^{C} \mathcal{T}_i(\mathbf{x})$$

C (cardinality) : size of the set of transformations to be aggregated

Aggregted Residual Transformations : $\mathbf{y} = \mathbf{x} + \sum_{i=1}^{C} \mathcal{T}_i(\mathbf{x})$

Aggregated Residual Transformations for Deep Neural Networks (2)

Cardinality

CNN 에서 하나의 Block 안의 transformation 개수, 혹은 group의 개수로 정의

모델을 더 Wider 또는 Deeper 하게 만드는 것 보다 정확도 면에서 더 높게 나올 수 있다는 관점을 제시

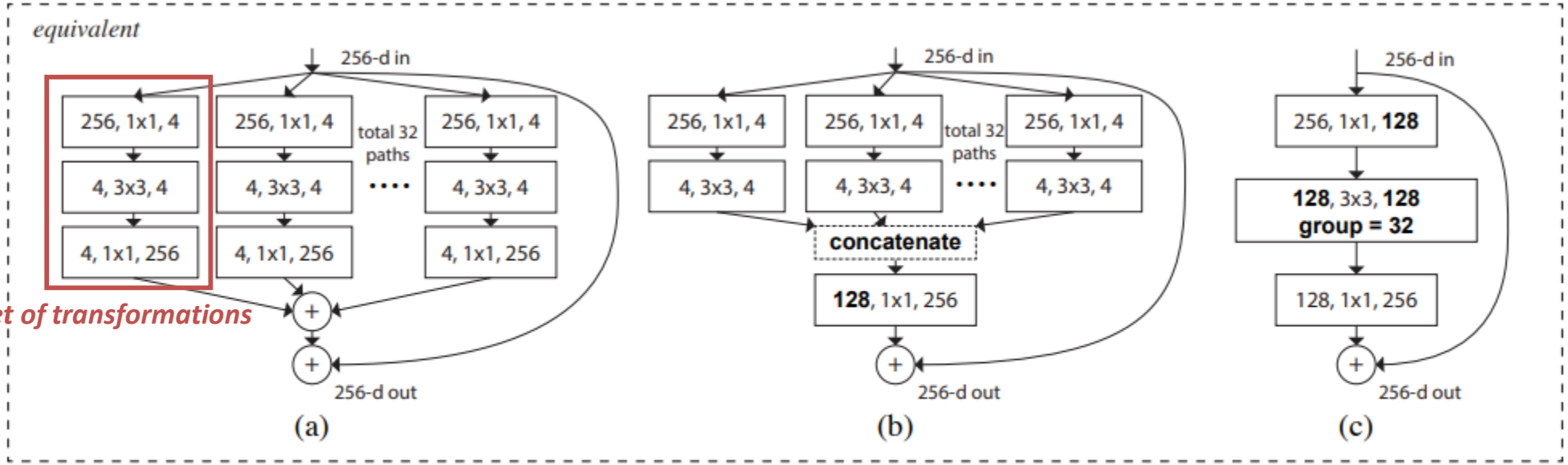


Figure 3. Equivalent building blocks of ResNeXt. **(a)**: Aggregated residual transformations, the same as Fig. 1 right. **(b)**: A block equivalent to (a), implemented as early concatenation. **(c)**: A block equivalent to (a,b), implemented as grouped convolutions [24]. Notations in **bold** text highlight the reformulation changes. A layer is denoted as (# input channels, filter size, # output channels).

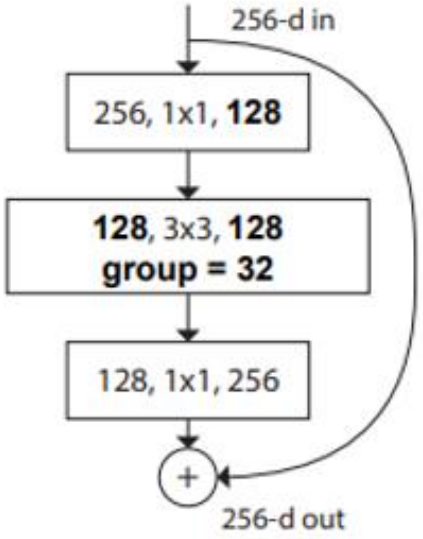
Grouped Convolution

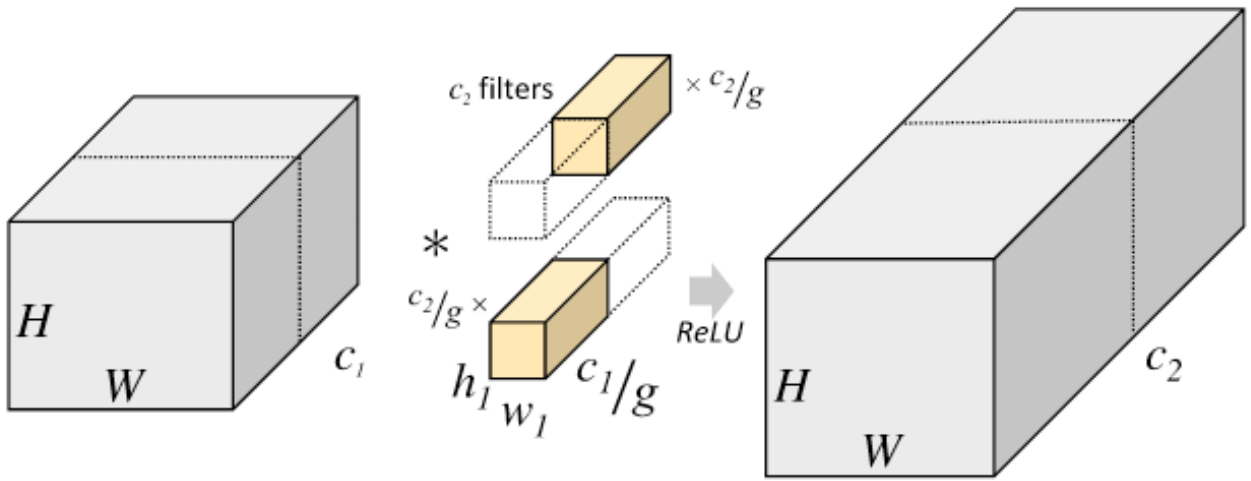
Input 의 channel들을 여러 개의 group으로 나누어 독립적으로 Convolution을 수행

Ex) Input Channel : 128 group : 32 Output Channel : 128 일 때,

128개 채널을 32개의 그룹으로 나누어 Convolution을 수행하며, 각 그룹은 4개의 채널을 갖는다.

그 다음, 그룹에서 나온 출력을 이어붙여 (concatenate) 다시 128 채널을 가진다.





g : 나눌 그룹의 수

Aggregated Residual Transformations for Deep Neural Networks (3)

Model Capacity

ResNet-50 FLOPs : $[1 * 1 * 256 * 64] + [3 * 3 * 64 * 64] + [1 * 1 * 64 * 256] = 70k$

ResNeXt FLOPs : $C * [(1 * 1 * 256 * d) + (3 * 3 * d * d) + (1 * 1 * d * 256)]$ 에서 $(C, d) = (32, 4)$, $\fallingdotseq$ $70k$

stage	output	ResNet-50	**ResNeXt-50 (32×4d)**
conv1	112×112	7×7, 64, stride 2	7×7, 64, stride 2
conv2	56×56	3×3 max pool, stride 2	3×3 max pool, stride 2
		[1×1, 64 3×3, 64 1×1, 256] ×3	[1×1, 128 3×3, 128, C=32 1×1, 256] ×3
conv3	28×28	[1×1, 128 3×3, 128 1×1, 512] ×4	[1×1, 256 3×3, 256, C=32 1×1, 512] ×4
conv4	14×14	[1×1, 256 3×3, 256 1×1, 1024] ×6	[1×1, 512 3×3, 512, C=32 1×1, 1024] ×6
conv5	7×7	[1×1, 512 3×3, 512 1×1, 2048] ×3	[1×1, 1024 3×3, 1024, C=32 1×1, 2048] ×3
	1×1	global average pool 1000-d fc, softmax	global average pool 1000-d fc, softmax
# params.		**25.5**×10^6	**25.0**×10^6
FLOPs		**4.1**×10^9	**4.2**×10^9

즉, 동일한 complexity 에서 더 좋은 성능을 얻음

	setting	5K-way classification		1K-way classification	
		top-1	top-5	top-1	top-5
ResNet-50	1 × 64d	45.5	19.4	27.1	8.2
ResNeXt-50	32 × 4d	**42.3**	**16.8**	**24.4**	**6.6**
ResNet-101	1 × 64d	42.4	16.9	24.2	6.8
ResNeXt-101	32 × 4d	**40.1**	**15.1**	**22.2**	**5.7**

Table 6. Error (%) on **ImageNet-5K**. The models are trained on ImageNet-5K and tested *on the ImageNet-1K val set*, treated as a 5K-way classification task or a 1K-way classification task at test time. ResNeXt and its ResNet counterpart have similar complexity. The error is evaluated on the single crop of 224×224 pixels.

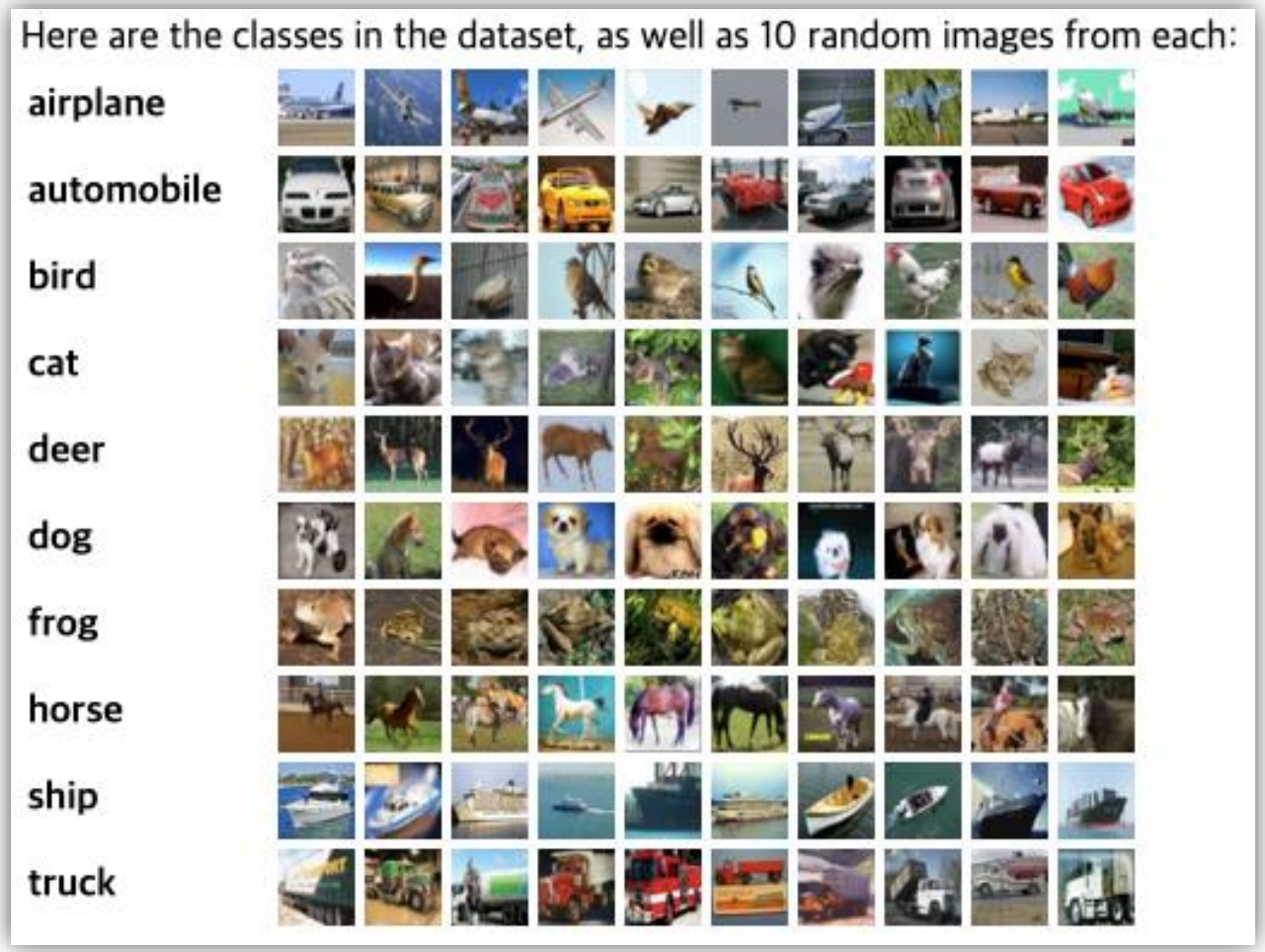
DATASET CIFAR-10

32x32사이즈의 RGB 이미지와

10가지의 레이블로 구성 $(32, 32, 3),\ (10)$

INPUT DATA : 50000

TEST DATA : 10000



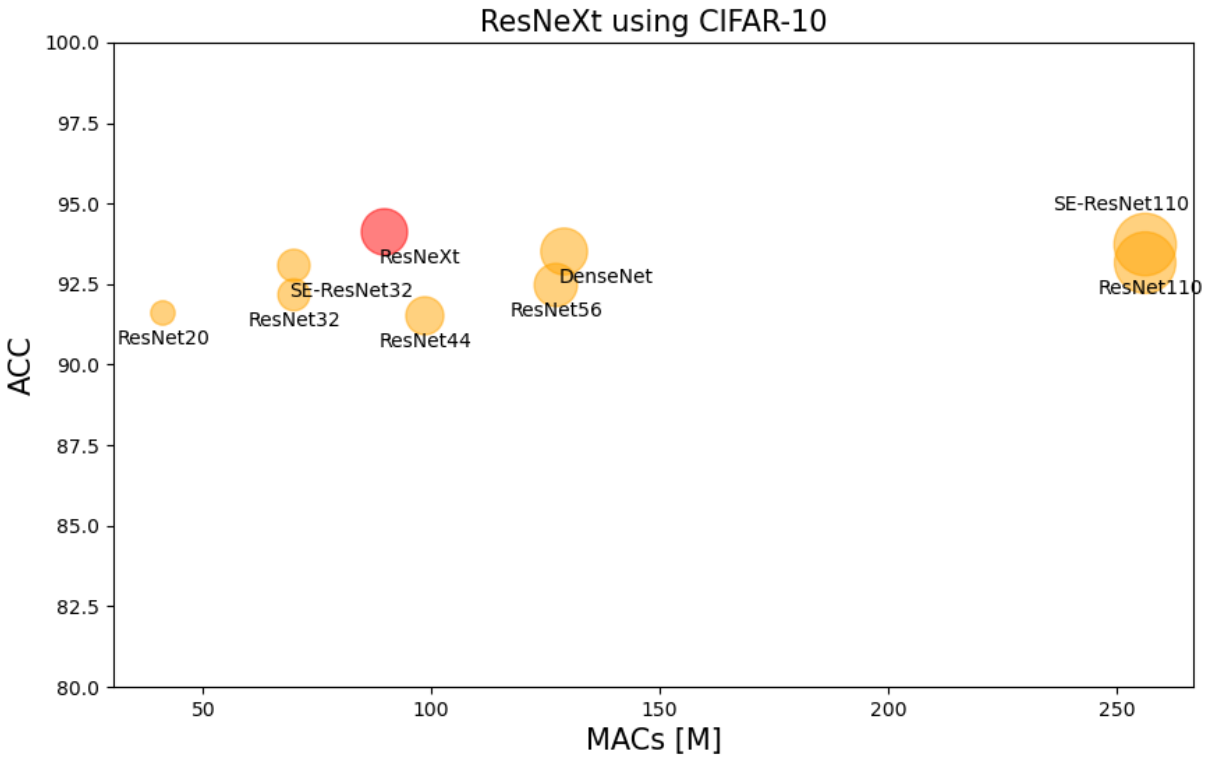
실험 결과

정밀도(Precision): $\frac{TP}{TP+FP}$

재현율(Recall): $\frac{TP}{TP+FN}$

정확도(Accuracy): $\frac{TP+TN}{TP+FP+FN+TN}$

Name	# params	precision	recall	accuracy
ResNet-20	0.27M	91.69%	91.62%	**91.63%**
ResNet-32	0.46M	92.21%	92.18%	**92.19%**
ResNet-44	0.66M	91.57%	91.52%	**91.53%**
ResNet-56	0.85M	92.46%	92.48%	**92.48%**
ResNet-110	1.7M	93.19%	93.17%	**93.17%**
DenseNet	1.78M	93.76%	93.77%	**94.22%**
ResNeXt-29(2 x 32d)	3.75M	93.71%	93.32%	**94.11%**



Which data fits ResNeXt model?

independent한 블럭 구조에서 channel을 representation 함으로써, feature를 더 다양하게 조합할 수 있음

group 구조가 아닌 모델에 비해, group으로 나눠서 독립적인 저차원으로 블럭 구조가 구성되어 있어 표현력에 제한

➡ 제한된 범위 안에서 다양한 feature를 원하는 상황일 때, ResNeXt 모델이 적합하다고 판단