



[논문리뷰] SENet [2018]

Introduction

✨ 해당 논문에서는 Channel RelationShip에 집중한, 모델인 SE Block을 소개합니다.

SEBlock은 채널 사이에서, 모델들이 독립적으로 반응함으로써, adaptively하게 channel-wise feature들이 재조정합니다.



여러 Dataset에서 일반화된 SENet은, SEBlock을 Stack하여 모델을 만드는 것이 가능합니다.

각 Conv2d Filter들은 Image 또는 feature map의 receptive field에 있는 정보들의 조합인 local을 학습합니다.

SEBlock은 Network가 feature들을 재조정하는 가능하도록 합니다. 해당 방식은 Global Information으로부터 선택적으로 강조되는 feature들을 학습하고, 덜 유용한 것들은 억제합니다.

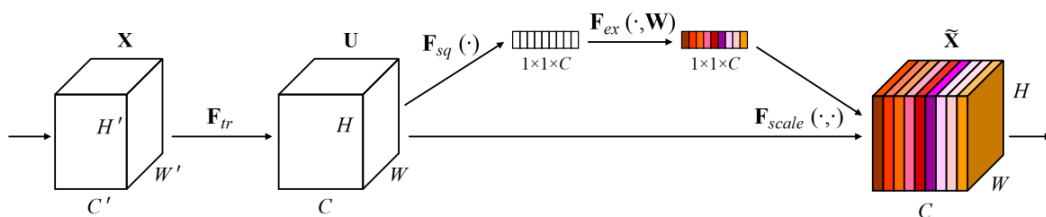


Fig. 1. A Squeeze-and-Excitation block.

Feature U가 squeeze operation을 첫째로 통과할 때, spatial dimensions의 여러 feature Map의 집합으로부터, channel descriptor를 생성합니다.

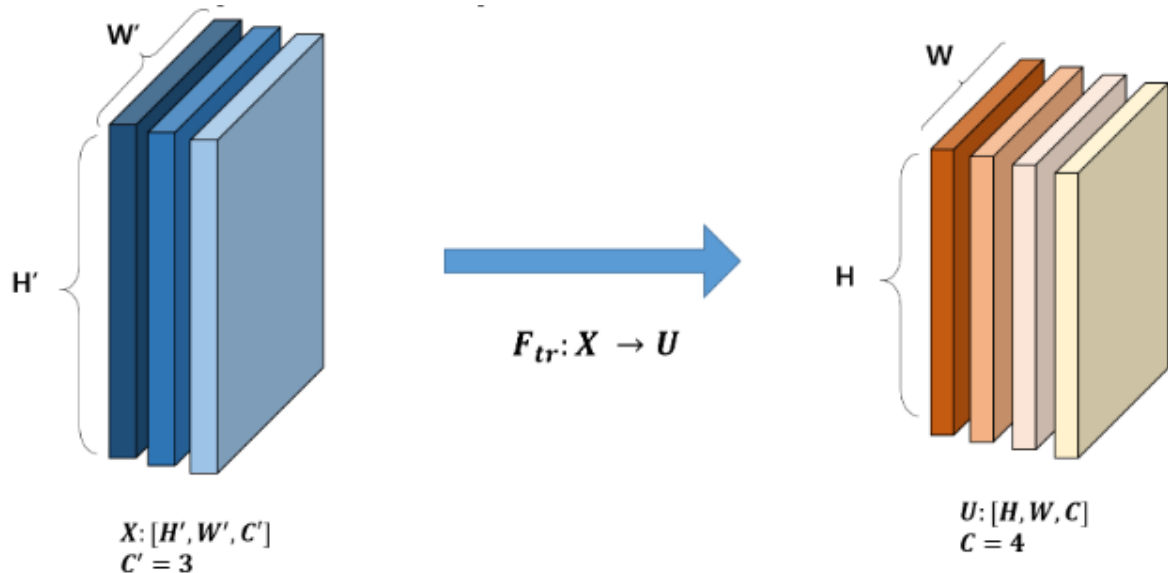
Squeeze-and-Excitation Blocks

✨ Reference의 LINK를 참고하시면, 보다 상세히 알 수 있습니다.

SENet을 이루는 것은 SE block이라고 할 수 있다. SE block은 $[H, W, C]$ 차원을 가지는 입력 X 를 Feature maps 출력인 U 로 변경하는 transformation $F : X \rightarrow U$ 가 있을 때 적

용하는 것이 가능하다.

출력 U에 대해 각 채널 별 정보를 뽑아내, 그 정보를 바탕으로, 스칼라 가중치를 계산해 해당 채널에 곱해주는 것이다.



Squeeze : Global Information Embedding

각각의 **local receptive field** 그리고 **transformation output U**는 **contextual information**을 밖으로 추출하는 것이 불가능하다.

이러한 문제를 해결하기 위해 모든 픽셀의 정보를 고려하는 **channel descriptor**를 제안합니다.

Channel-wise statistics를 생성하기 위해 **global average pooling**을 사용함으로써, 성취하게 됩니다.

즉, 채널 정보의 전체를 활용하여, 채널을 대표하는 Z값을 만든다는 것이다.

출력 transformation U는 전체 이미지에 대한 통계적 표현으로 해석되어집니다.

$$z_c = \mathbf{F}_{sq}(\mathbf{u}_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j).$$

Excitation : Adaptive Recalibration

Squeeze operation 내에서 집계된 정보를 활용하기 위해서, **fully capture channel-wise**의 종속성을 고려해야합니다.

이를 달성하기 위해서, 주의사항으로 두가지 있는데,

- Must be flexible - 채널 간의 복잡한(nonlinear) 관계를 찾을 수 있어야 합니다.
- one-hot activation에 초점을 두는 것이 아니라, 여러 채널들을 골라서 강조할 수 있어야 합니다.

$$\mathbf{s} = \mathbf{F}_{ex}(\mathbf{z}, \mathbf{W}) = \sigma(g(\mathbf{z}, \mathbf{W})) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z})), \quad (3)$$

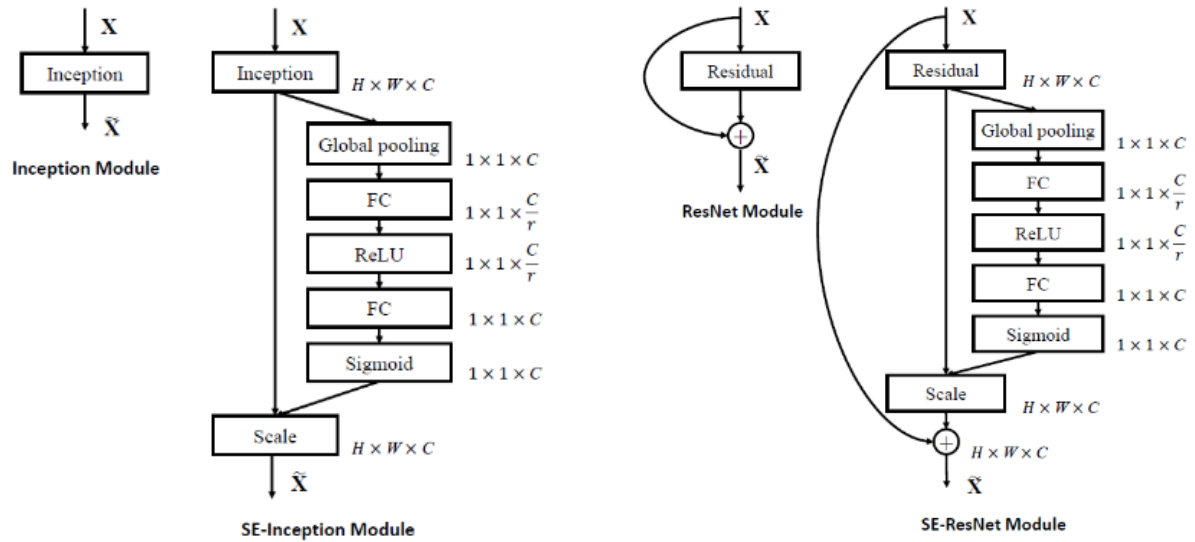
where δ refers to the ReLU [63] function, $\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $\mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$. To limit model complexity and aid generalisation, we parameterise the gating mechanism by forming a bottleneck with two fully-connected (FC) layers around the non-linearity, i.e. a dimensionality-reduction layer with reduction ratio r (this parameter choice is discussed in Section 6.1), a ReLU and then a dimensionality-increasing layer returning to the channel dimension of the transformation output \mathbf{U} . The final output of the block is obtained by rescaling \mathbf{U} with the activations \mathbf{s} :

$$\tilde{\mathbf{x}}_c = \mathbf{F}_{scale}(\mathbf{u}_c, s_c) = s_c \mathbf{u}_c, \quad (4)$$

Advantages

1. 유연합니다.

- SE Block은 한 Convolution layer에서만, 적용 가능한 것이 아니라, 모든 이미지에 적용하는 것이 가능합니다.



2. 추가적인 계산량이 적습니다.

- 기존 CNN에 SE Block을 추가하였을 때, 늘어나는 파라미터는 거의 없습니다.

Model Architecture and Experiment

TABLE 1
(Left) ResNet-50 [13]. (Middle) SE-ResNet-50. (Right) SE-ResNeXt-50 with a $32 \times 4d$ template. The shapes and operations with specific parameter settings of a residual building block are listed inside the brackets and the number of stacked blocks in a stage is presented outside. The inner brackets following by *fc* indicates the output dimension of the two fully connected layers in an SE module.

Output size	ResNet-50	SE-ResNet-50	SE-ResNeXt-50 ($32 \times 4d$)
112 × 112		conv, 7 × 7, 64, stride 2	
56 × 56		max pool, 3 × 3, stride 2	
	$\begin{bmatrix} \text{conv}, 1 \times 1, 64 \\ \text{conv}, 3 \times 3, 64 \\ \text{conv}, 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} \text{conv}, 1 \times 1, 64 \\ \text{conv}, 3 \times 3, 64 \\ \text{conv}, 1 \times 1, 256 \\ \text{fc}, [16, 256] \end{bmatrix} \times 3$	$\begin{bmatrix} \text{conv}, 1 \times 1, 128 \\ \text{conv}, 3 \times 3, 128 \\ \text{conv}, 1 \times 1, 256 \\ \text{fc}, [16, 256] \end{bmatrix} \times 3$
28 × 28	$\begin{bmatrix} \text{conv}, 1 \times 1, 128 \\ \text{conv}, 3 \times 3, 128 \\ \text{conv}, 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} \text{conv}, 1 \times 1, 128 \\ \text{conv}, 3 \times 3, 128 \\ \text{conv}, 1 \times 1, 512 \\ \text{fc}, [32, 512] \end{bmatrix} \times 4$	$\begin{bmatrix} \text{conv}, 1 \times 1, 256 \\ \text{conv}, 3 \times 3, 256 \\ \text{conv}, 1 \times 1, 512 \\ \text{fc}, [32, 512] \end{bmatrix} \times 4$
14 × 14	$\begin{bmatrix} \text{conv}, 1 \times 1, 256 \\ \text{conv}, 3 \times 3, 256 \\ \text{conv}, 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} \text{conv}, 1 \times 1, 256 \\ \text{conv}, 3 \times 3, 256 \\ \text{conv}, 1 \times 1, 1024 \\ \text{fc}, [64, 1024] \end{bmatrix} \times 6$	$\begin{bmatrix} \text{conv}, 1 \times 1, 512 \\ \text{conv}, 3 \times 3, 512 \\ \text{conv}, 1 \times 1, 1024 \\ \text{fc}, [64, 1024] \end{bmatrix} \times 6$
7 × 7	$\begin{bmatrix} \text{conv}, 1 \times 1, 512 \\ \text{conv}, 3 \times 3, 512 \\ \text{conv}, 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} \text{conv}, 1 \times 1, 512 \\ \text{conv}, 3 \times 3, 512 \\ \text{conv}, 1 \times 1, 2048 \\ \text{fc}, [128, 2048] \end{bmatrix} \times 3$	$\begin{bmatrix} \text{conv}, 1 \times 1, 1024 \\ \text{conv}, 3 \times 3, 1024 \\ \text{conv}, 1 \times 1, 2048 \\ \text{fc}, [128, 2048] \end{bmatrix} \times 3$
1 × 1	global average pool, 1000-d <i>fc</i> , softmax		

TABLE 16
Effect of Squeeze operator on ImageNet (error rates %).

	top-1 err.	top-5 err.	GFLOPs	Params
ResNet-50	23.30	6.55	3.86	25.6M
NoSqueeze	22.93	6.39	4.27	28.1M
SE	22.28	6.03	3.87	28.1M

Conclusion


SE block은, **dynamic channel-wise feature recalibration**을 가능하게 함으로써, Network의 표현력을 상당히 개선시키는 것이 가능합니다.

게다가 SE Block은 **model channel-wise feature dependencies**를 조절하기 위해 이전의 모델에서 능력이 없던 것들에 대해 약간 성능을 발휘하게 합니다.

Reference

Squeeze-and-Excitation Networks

The central building block of convolutional neural networks (CNNs) is the convolution operator, which enables networks to construct informative features by fusing both spatial and

 <https://arxiv.org/abs/1709.01507>



Squeeze-and-Excitation Networks

Squeeze-Excitation-Network 이 글은 Squeeze-and-Excitation Networks 논문을 읽고 요약/정리한 것이다. SENet은 VGGNet, ResNet, DenseNet에 이어 2017 ILSVRC 에서 1등을 한 CNN 구조

☹ <https://www.iiiiii.tistory.com/entry/SqueezeandExcitation-Networks>

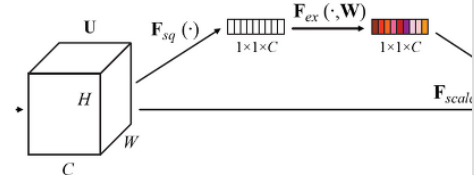


Figure 1: A Squeeze-and-Excitation block.