_____

heum _____

_____

_____

_____ ✒

SEnet

○ Channel 간의 inter-dependency를 modeling   ↗채널간 출력값
 → Channel-wise feature response를 adaptive 하게
   recalibration 해주는 Unit을 만들려고.
⇒ Squeeze-and-Excitation (SE) block을 제안.

○ SE block을 위한 조금의 Cost만 지원해준다면
기존의 Architecture 성능화 크게 향상 효과 기대

Fig.1 A Squeeze and Excitation block

Squeeze	top-1 err.	top-5 err.
Max	22.57	6.09
Avg	**22.28**	**6.03**

Table 1.

Excitation	top-1 err.	top-5 err.
ReLU	23.47	6.98
Tanh	23.00	6.38
Sigmoid	**22.28**	**6.03**

Table 2.

○ CNN에서는 conv layer에 `non-linear activation`과 `downsampling operator`
를 사용함으로서   ↗계층적 패턴   ex) maxpooling..
  o Hierarchical pattern 의 capture
  o Global theoretical receptive field
위 목적을 달성하는 Powerful image representation을 찾는것.

network 에서 생성된 representation을 향상시키기 위하여
Squeeze-and-excitation (SE) block을 제안.
(축소)   (확대)
(개념화) (정밀화)

Q Squeeze   GMaxP < GAverP   High   width
— Spatial dimension (HxW)에 걸친 feature map을 aggregation 함으로서
  channel descriptor를 생성.   (융합)

— 이 descriptor은 channel-wise feature response 의 global distribution
  을 임베딩하여, 네트워크의 모든 layer에서 global field에서 (information)
  얻을수있는 information을 사용할수 있도록 해준다.

값, 단어들의 features를
표로 Vector로 표현

Q Excitation   ReLU < Tanh < sigmoid
— 임베딩을 input으로 하여 Per-channel modulation weight 를 생성하는
  simple self-gating 메커니즘의 형태   ↳FC를 하나 넣는다는 말인가요

— 이 weight 들은 feature map U에 적용되며, 이렇게 생성된 SE block의 출력은
  후속 layer에 바로 사용될수있다.

○ 또한 SE block은 다양한 depth의 네트워크 architecture에서
  original block의 대체로도 사용가능
                                    ⇒ flexable

in earlier layer
→ class에 구애받지 않는 형식으로 유용한 feature 들을 자극
→ Shared low-level representation을 강화
in later layer
→ 점점 더 specialised 되며
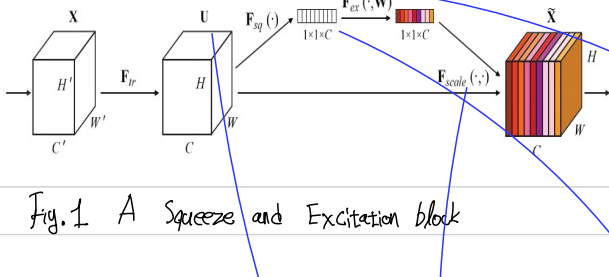→ class에 특화된 방식으로 다른 input에 응답

↳ class2 = image

→ SE block의 구조는 간단하면서도 성능을 효과적으로 향상가능
→ 기존 architecture에서 SE block 에 대응되는 component만 교체 가능
→ 계산적으로도 가볍기 때문에 `model complexity` 와 `computational cost`를
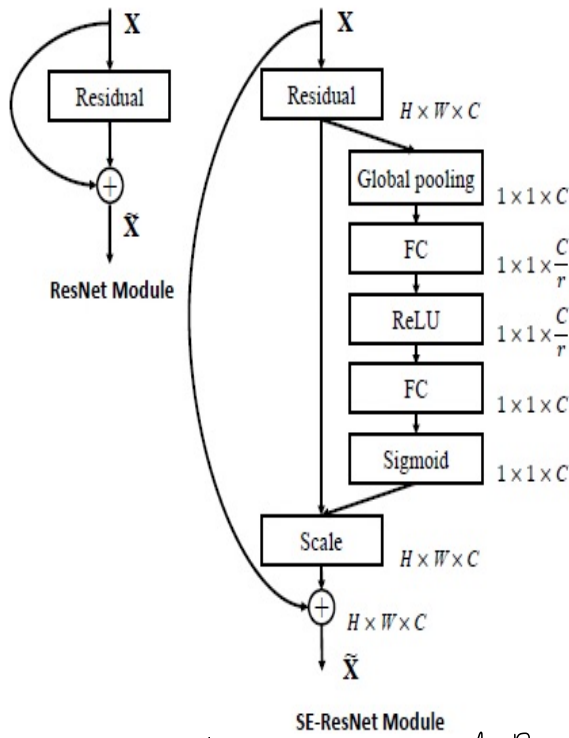 약간만 증가시킬 뿐이다.

○ 특정 task나 dataset 에 제한되지 않는다고 제시함

Fig 3. The schema of the original Residual module (left) and the SE-ResNet module (right)

Ratio r	top-1 err.	top-5 err.	Params
2	22.29	6.00	45.7M
4	22.25	6.09	35.7M
8	22.26	5.99	30.7M
16	22.28	6.03	28.1M
32	22.72	6.20	26.9M
original	23.30	6.55	25.6M

적절 ✓ (16)

Table 10.

○ representational Power 네트워크의 표현력

○ Attention and gating mechanisms

→ Attention은 이용 가능한 computational resource를, signal 중 가장 유익한 요소에 할당하는 성향을 갖게끔 유도할 수 있다.

ex) Sequence learning
ex) localization and understanding in image
ex) Image captioning
ex) Lip reading.

→ 이러한 응용 연구들에선 보통의 기법들 간의 적용을 위하, 고차원의 abstraction을 나타내는 1개 이상의 layer로 이루어진 하나의 operator로 작동 (구성)

SE block 에서는 lightweight gating 메커니즘로 구성

⇒ computationally efficient 한 channel-wise relationship을 모델링, 네트워크의 representational power를 향상시키는 것에 중점을 둠.

○ hyper-parameter
— mini-batch size =1024
— Initial learning rate = 0.6
— 30 epoch → 0.1 lr
— He initialization (weight initialization)
— Train for 100 epoch.
— reduction ratio $r = 16$

final stage (Paper..)
↕
○ 상대적으로 later layer에 있는 channel 이 대부분의 parameter를 차지함

○ final stage, later layer에 SE block 을 없앨 경우
⌄4%의 parameter 증가율로 0.1% 미만의 작은 손실만 피해볼수 있다.

ex) final stage의 channel이 512라 할때 이 stage 엔 SE block 사용 안함.
<u>final stage -1</u> 의 SE block만 사용.

Q SE block 자체가 depth 를 증가시킨다는 점
→ 매우 효율적인 계산 방식으로 동작
→ 기본 architecture 의 depth가 늘어남에 따라 return 이 감소하는 지점에서도 좋은 결과를 얻을수 있다.
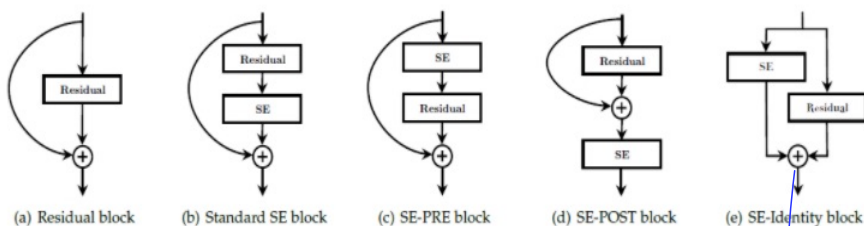
ex) Resnet-50 < SE-Resnet-50 < Resnet-101
7.48%   6.62%   6.52%

25M Param   25M+2.5 Param

10%를 cost 더 줌   100 layer에 비슷한 효과를 발휘있다라 함. 카더라 통신에!

Stage	top-1 err.	top-5 err.	GFLOPs	Params
ResNet-50	23.30	6.55	3.86	25.6M
SE_Stage_2	23.03	6.48	3.86	25.6M
SE_Stage_3	23.04	6.32	3.86	25.7M
SE_Stage_4	22.68	6.22	3.86	26.4M
SE_All	22.28	6.03	3.87	28.1M

Table 13.

(a) Residual block (b) Standard SE block (c) SE-PRE block (d) SE-POST block (e) SE-Identity block

○ How to join SE block?
→ SE-POST block을 제외한 나머지는 비슷한 성능을 보임

table 5. SE block integration designs explored in the ablation study
(통합 설계)

Design	top-1 err.	top-5 err.
SE	22.28	6.03
SE-PRE	22.23	6.00
SE-POST	22.78	6.35
SE-Identity	22.20	6.15

table 4. Effect of different SE block integration strategies
with ResNet-50 on ImageNet (error rates)

○ 아무래도 resnet의 Paper에 영향을
넣지 않았나.. 라는 생각.

○ SE-PRE block이 조금이라도 더 좋아보이긴 하는데..
이왕 쓸거 standard SE block 보다 SE-PRE block을
한번 써볼까?
(뭐.. 거의 차이가 없긴한데..)

○ branch aggregation 이전에 적용해야 효과가
있다는 사실 만 알고 가자. (지점 접계)ㄱㄱ

즉, squeeze 와 Excitation 이라는 연산 작업을 통하여
정보의 압축, 재 조정을 효율적으로 가능.
적은 cost (연산량) 으로도 성능향상 가능.
flexible (유연) 하게 기존 네트워크 (architecture)에도 사용가능.

궁금증//

하위(초기) layer 에선 general한 feature를 학습,
상위(후기) layer 에서 고차원의 feature을 학습 한다 했는데

conv1D에서 어떤 filter로 general 과 고차원의 feature를 구분하지?
비슷한가?