**Squeeze-and-Excitation Networks 리뷰**

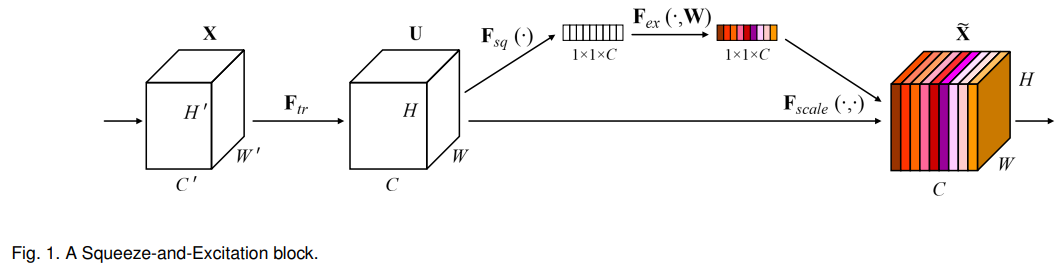
1. **Abstract**

CNN의 central building block은 convolution operator로 networks가 각 layer의 local receptive fields 내에서 spatial과 channel-wise information을 융합함으로 informative features를 구성할 수 있게 한다. 이전 연구는 이 relationship을 조사해 feature hierarchy 전체의 spatial encoding을 향상시켜 CNN의 representational power를 강화하고자 하였다.

대신 이 논문에서는 channel relationship에 초점을 두고 “Squeeze-and-Excitation” (SE) block이라 칭하는 새로운 architectural unit을 제안한다. 이것은 channels 간의 interdependencies를 명시적으로 modelling하여 channel-wise feature responses를 적응적으로 recalibrates한다.

1. **INTRODUCTION**

이 논문에서 우리는 network design의 channels 간의 relationship 측면을 조사한다. 그리고 convolutional features간의 상호의존성을 명시적으로 modelling해 network에 의해 생성된 representations의 품질을 향상시키는 목표와 함께 Squeeze-and-Excitation (SE) block이라 칭하는 새로운 architectural unit을 소개한다. 끝에는 network가 feature recalibration을 수행할 수 있게 하는 매커니즘을 제안한다. 이를 통해 informative features를 선택적으로 강조하고 덜 유용한 것을 억제해서 global information을 사용해 학습할 수 있다.



SE building block의 구조는 다음과 같다. 주어지는 transformation Ftr은 input X를 feature maps U로 mapping한다. 우리는 feature recalibration을 수행하기 위해 해당하는 SE block을 구성할 수 있다. Features U는 먼저 squeeze operation을 통과한다. 이것은 spatial dimensions 전체에서 feature maps을 aggregating해 channel descriptor을 생성한다. 이 descriptor의 기능은 channel-wise feature responses의 global distribution의 embedding을 생성하여 network의 global receptive field의 information이 모든 layers에서 사용될 수 있도록 한다(#오역 가능성 있음). 이 aggregation은 다음으로 embedding을 input으로 취하고 per-channel modulation weights의 collection을 만드는 간단한 self-gating mechanism의 형태를 취하는 excitation operation을 통과한다. 이 weights는 feature maps U에 적용되어 network의 subsequent layers로 직접적으로 공급될 수 있는 SE block의 output을 생성한다.

이것은 SE blocks의 collection을 간단히 stacking 함으로 SE network (SENet)를 구성할 수 있다. 더 나아가 이 SE blocks은 network architecture의 depths의 range에서 original block에 대한 drop-in replacement로 사용될 수 있다(#오역 가능성 있음). Building block의 template은 일반적이지만 그것이 다른 depths에서 수행하는 역할은 network 전체에서 다르다. earlier layers에서 그것은 class-agnostic manner로 informative features를 자극하고 shared low-level representations을 강화한다. Later layer에서 SE blocks은 점점 specialized된다. 그리고 다른 inputs에 대해 highly class-specific manner로 응답한다. 결과적으로 SE blocks에 의해 수행되는 feature recalibration의 benefits은 network를 통해 누적될 수 있다.

1. **RELATED WORK**

~

1. **SQUEEZE-AND-EXCITATION BLOCKS**

Squeeze-and-Excitation block은 input X를 feature maps U로 mapping 하는 transformation Ftr 위에서 지어질 수 있는 computational unit이다. Ftr의 수식은 다음과 같다.

텍스트, 손목시계, 시계이(가) 표시된 사진

자동 생성된 설명

vc는 c번째 filter의 parameters, X는 input, u는 output이다(#아마도). Bias는 생략되었다.

우리는 convolutional features의 learning이 명시적인 modelling channel interdependencies에 의해 향상될 것을 기대한다. 따라서 network는 subsequent transformations에서 활용될 수 있는 informative features에 대한 민감도를 향상시킬 수 있다. 결과적으로 우리는 global information에 대한 access를 제공하고 next transformation로 제공하기 전 squeeze and excitation의 두 단계로 filter responses를 recalibrate하고자 한다.

* 1. **Squeeze: Global Information Embedding**

Channel dependencies를 이용함의 문제를 해결하기 위해 먼저 output features의 각 channel의 signal을 고려했다. 각 learned filters는 local receptive field와 operates되고 결과적으로 각 transformation output U의 unit은 이 region의 outside에서 contextual information을 이용할 수 없다.

이 문제를 축소하기 위해 우리는 global spatial information을 channel descriptor로 squeeze할 것을 제안한다. 이것은 channel-wise statistics를 생성하기 위해 global average pooling를 사용함으로 달성된다. 공식적으로 statistic z는 spatial dimensions HxW를 통해 U를 shrinking 함에 의해 생성된다. c번째 z는 다음 수식으로 계산된다.

텍스트, 시계이(가) 표시된 사진

자동 생성된 설명

Transformation U의 output는 전체 image에 대해 설명되는 통계의 local descriptors의 collection으로 해석될 수 있다. 이러한 information을 이용하는 것은 이전 feature engineering work에서 널리 퍼져 있었다. 우리는 가장 간단한 aggregation technique인 global average pooling을 선택하였다. 여기서 더 정교한 strategies를 사용할 수 있었다.

* 1. **Excitation: Adaptive Recalibration**

Squeeze operation의 aggregated된 information을 사용하려면 우리는 channel-wise dependencies를 fully capture하는 것을 목표로 하는 second operation으로 이어진다. 이 목적을 위해 function은 두 기준을 충족해야 한다. 먼저 유연해야 한다(특히 그것은 channels 간의 nonlinear interaction을 학습할 수 있어야한다). 그리고 두번째로 non-mutually-exclusive relationship을 학습할 수 있어야 한다. Multiple channels이 강조될 수 있음을 보장하고 싶기 때문이다(one-hot activation 대신에). 이 두 기준을 충족하기 위해 우리는 sigmoid activation이 있는 간단한 gating mechanism을 선택했다. 수식은 다음과 같다.



δ 는 ReLU function이다. Model complexity를 제한하고 generalization을 돕기 위해 우리는 non-linearity 주위에 두개의 fully-connected (FC) layers로 bottleneck을 형성해 gating mechanism을 parameterize한다. 예를 들어 reduction ratio r의 dimensionality-reduction layer(#잘 모르겠음). ReLU 그 다음 dimensionality-increasing layer는 transformation output U의 channel dimension으로 returning한다. Block의 final output은 activations s로 U를 rescaling하여 얻는다. 수식은 다음과 같다.



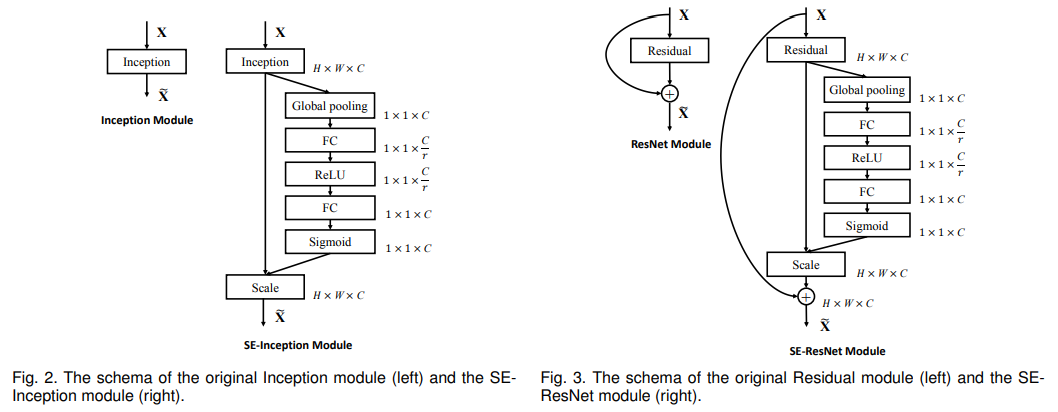
이는 scalar sc 그리고 feature map uc 간의 channel-wise multiplication이다.

Excitation operator은 input specific descriptor z를 channel weights의 set으로 maps한다. 이와 관련하여 SE blocks은 본질적으로 input에 conditioned된 dynamics를 도입한다. 이는 convolutional filters가 응답하는 local receptive field에 제한되지 않는 relationships의 channels에 대한 self attention function으로 간주될 수 있다.

* 1. **Instantlations**

SE block은 각 convolution을 뒤따르는 non-linearity 이후에 삽입으로 VGGNet 같은 standard architectures로 integrated 될 수 있다. 더 나아가 SE block의 유연성은 이것이 standard convolutions을 넘어 직접 transformations에 적용될 수 있음을 의미한다.

먼저 우리는 Inception networks를 위한 SE blocks의 구성을 고려한다. 먼저 transformation Ftr을 전체 Inception module로 간주한다(Fig.2). 그리고 architecture의 각 module을 변환하여 SE-Inception network를 얻는다(#오역 가능성 있음). 또한 SE block은 직접 residual networks와 함께 사용될 수 있다(Fig. 3). 여기서 SE block transformation Ftr은 residual module의 non-identity branch로 간주된다. Squeeze와 Excitation은 둘 다 identity branch와 summation 전에 작동한다.



SE block의 유연한 특성의 한 결과는 그것이 architectures로 통합될 수 있는 몇가지 가능한 방법이 있다는 것이다.

1. **MODEL AND COMPUTAIONAL COMPLEXITY**

실용적으로 SE block design을 사용하기 위해 그것은 향상된 성능과 model complexity 증가 간의 좋은 절충안을 제안해야 한다. 각 SE block은 squeeze phase에서 global average pooling operation을 사용하고 excitation phase에서 두개의 작은 FC layers를 사용한 뒤 저렴한 channel-wise scaling operation을 사용한다.

우리는 SE block에 의해 초래되는 작은 additional computational cost가 model performance의 기여에 의해 정당화된다고 믿는다.

다음으로 SE block에 의해 additional parameters를 고려한다. 이 additional parameters는 오직 두개의 gating mechanism의 FC layers의 결과다. 따라서 전체 network capacity의 작은 fraction를 구성한다. 구체적으로 FC layers의 weight parameters에 의해 도입된 전체 수는 다음과 같다.

텍스트, 손목시계이(가) 표시된 사진

자동 생성된 설명

r은 reduction ratio, S는 stage의 수(stage는 common spatial dimension의 feature maps에서 operating하는 block의 collection를 나타낸다), Cs는 output channels의 dimension 그리고 Ns는 stage s에서 repeated blocks의 수를 나타낸다. 이 parameters의 대부분은 가장 많은 수의 channels에서 excitation operation이 수행되는 network의 final stage에서 온다. 하지만 우리는 비교적 비싼 SE blocks의 final stage를 relative parameter increase를 ~4%로 줄이게 performance에 작은 비용으로 제거될 수 있음을 발견했다. 이는 parameter usage가 key consideration일 때 유용함을 입증할 수도 있다.

1. EXPERIMENTS

~

1. ABLATION STUDY

~

1. ROLE OF SE BLOCKS

~

1. CONCLUSION

이 논문에서 우리는 dynaic channel-wise feature recalibration을 수행할 수 있게 하여 network의 representational power를 향상하도록 설계된 architectural unit인 SE block을 제안했다. 연구의 넓은 범위는 multiple dataset과 tasks에서 state-of-the-art performance를 도달하는 SENets의 효과를 보여준다. 추가적으로 SE blocks은 이전 architectures가 channel-wise feature dependencies를 적절히 model할 수 없음을 보여준다. 우리는 이 통찰이 strong discriminative features가 요구되는 다른 tasks에 유용함을 입증할 수 있기를 바란다. 마지막으로 SE blocks에 의한 feature importance values produced는 model compression을 위한 network pruning 같은 tasks에 사용될 수도 있다.