



[논문리뷰] Residual Attention Network [2017]

Introduction

✨ 해당 논문은 Attention Mechanism을 사용한 **Residual Attention Network**를 설명합니다.

해당 방식은 End-to-End 방식으로 SOTA를 달성한 **Feed Forward Net**과 통합하는 것이 가능합니다.

Residual Attention Network은 attention-aware features를 생성하는 **Stacking Attention Module**로 만들어졌습니다.

☀ **Importantly**, scaled up to hundreds of layers까지 쉽게 학습하기 위해 **Deep Residual Attention**

Network를 Train하기 위해서는 **Attention Residual**을 학습해야 합니다.

Attention은 focused location을 선택하여 serve할 뿐만 아니라, location내에 다른 Object의 representations을 향상시킵니다.

최근 **Network**와 **Attention Mechanism**이 집중되면서, “**Very Deep**” Structure에서 mixed attention mechanism을 사용한 **Residual Attention Network**를 제안합니다.

Residual Attention Network는 attention-aware features를 생성하는 **Multiple Attention Module**로 구성되어 있습니다.

Residual Attention Network Advantage

- 다양한 형태의 **Attention Module**이 extensively 포착하므로, [Fig1] **Attention Module**은 일관된 성능을 향상으로 이어집니다.
- SOTA를 달성한 Model과 end-to-end training fashion의 방식으로 통합할 수 있습니다.

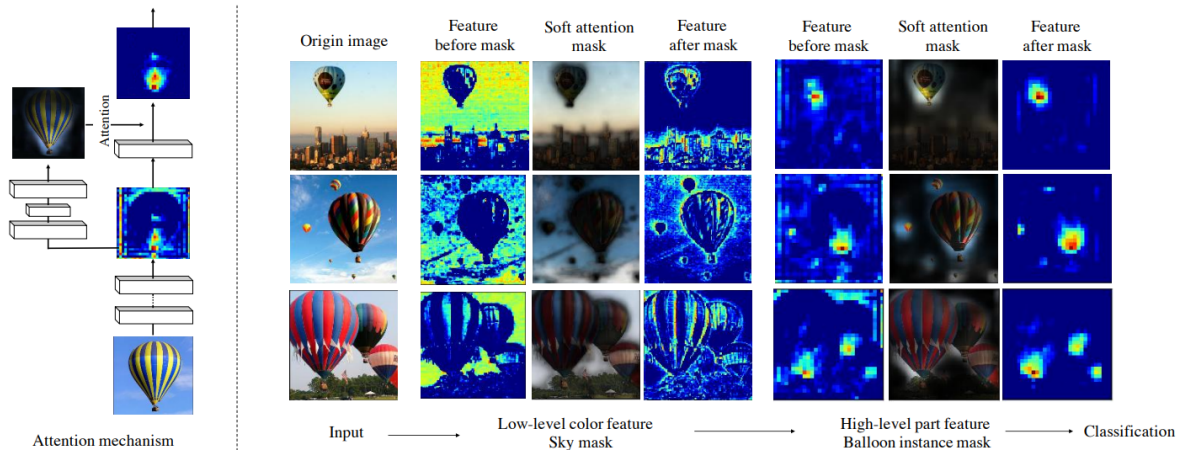


Figure 1: **Left:** an example shows the interaction between features and attention masks. **Right:** example images illustrating that different features have different corresponding attention masks in our network. The sky mask diminishes low-level background blue color features. The balloon instance mask highlights high-level balloon bottom part features.

3가지 방법으로 **Residual Attention Module**에 기여하여 **Advantage**를 얻었습니다.

1. **Residual Attention Network**는 **Stacking Multiple Attention Module**로 구성되어 있습니다.

stacked structure구조는 **mixed attention mechanism**으로 구성되어 있습니다. 그러므로, **Different types of attention**들이 **different attention module**로부터 **capture**됩니다.

2. Attention Module을 여러 층 쌓는 방식은 명백히 성능을 저하시킵니다. 수백만 개의 **Layer**를 최적화하기 위해서, **Deep Residual Attention Network**를 제안합니다.
3. Bottom-up top-down feedforward attention은 성공적으로 **human pose estimation**과 **image segmentation**에 성공적으로 적용되었습니다. 우리는 **Attention Module**의 부분에 **feature**에 **soft weights**를 추가합니다.

Residual Attention Network

Residual Attention Network는 **Stacking Multiple Attention module**에 의해 생성되었습니다.

각각의 **Attention Module**은 두 가지로 나뉘어집니다.

- **trunk branch**는 **feature processing**을 하며, **SOTA**의 **network structures**에서 활용가능합니다.
- **mask branch**는 Output feature를 부드럽게 하는 **Mask(x)**를 학습하기 위해서 **bottom-up top-down structure**를 사용합니다.

✨ Bottom-up top-down structure는 매우 빠르게, attention process를 mimics 합니다.

In Attention Modules, the attention mask can not only serve as a feature selector during forward inference, but also as a gradient update filter during back propagation. In the soft mask branch, the gradient of mask for input feature is:

$$\frac{\partial M(x, \theta) T(x, \phi)}{\partial \phi} = M(x, \theta) \frac{\partial T(x, \phi)}{\partial \phi} \quad (2)$$

Attention Residual Learning

일반적인 stacking Attention Module은 성능을 저하시킵니다.

- 0~1 사이의 범위 내에서 dot production을 반복적으로 하는 것은 deep layer의 feature의 value를 degrade합니다.
- Softmask는 잠재적으로 trunk branch의 good property를 잃습니다.

그러므로, Output H of Attention Module을 수정합니다.

attention. Thus we modify output H of Attention Module as

$$H_{i,c}(x) = (1 + M_{i,c}(x)) * F_{i,c}(x) \quad (3)$$

$M(x)$ ranges from $[0, 1]$, with $M(x)$ approximating 0, $H(x)$ will approximate original features $F(x)$. We call this method attention residual learning.



ResNet의 Residual Function은 $H_{i,c}(x) = x + F_{i,c}(x)$ 로 공식화되며, Residual Function은 $F_{i,c}(x)$ 로 근사합니다.

Formulation내에서, $F_{i,c}(x)$ 는 Deep Convolution Network에서 생성한 feature를 의미합니다.

핵심은 $M(x)$ 의 Mask branch에 놓여 있습니다.

feature selection을 함으로써, enhance good feature를 하며, trunk features from부터 noise를 suppress합니다.

추가로, Stacking Attention Module은 incremental nature에 의하여, residual learning이 백업됩니다.

Attention Residual Learning은 original feature로부터 good properties를 가지고 있을 뿐만 아니라, soft mask branch를 분기를 하기도 하며, mask branch's feature selection ability를 weaken하게도 합니다.



Attention Residual Learning을 사용함으로써, Residual Attention Network의 깊이가 증가할수록 성능이 향상됩니다.

Soft Mask Branch

In Convolutional Neural Network에서는, bottom-up 그리고 top-down 두 구조로 전개됩니다.

Max pooling은 receptive field를 빠르게 증가시킵니다.

낮은 해상도에 도달한 이후에, Symmetrical top-down architecture가 input feature를 각각의 position에 guide함으로써, Global Information를 확장합니다.

Attention Module

✨ 정리하면, Attention Module은 Soft Mask Branch와 Trunk Branch, 두 갈래로 나뉩니다.

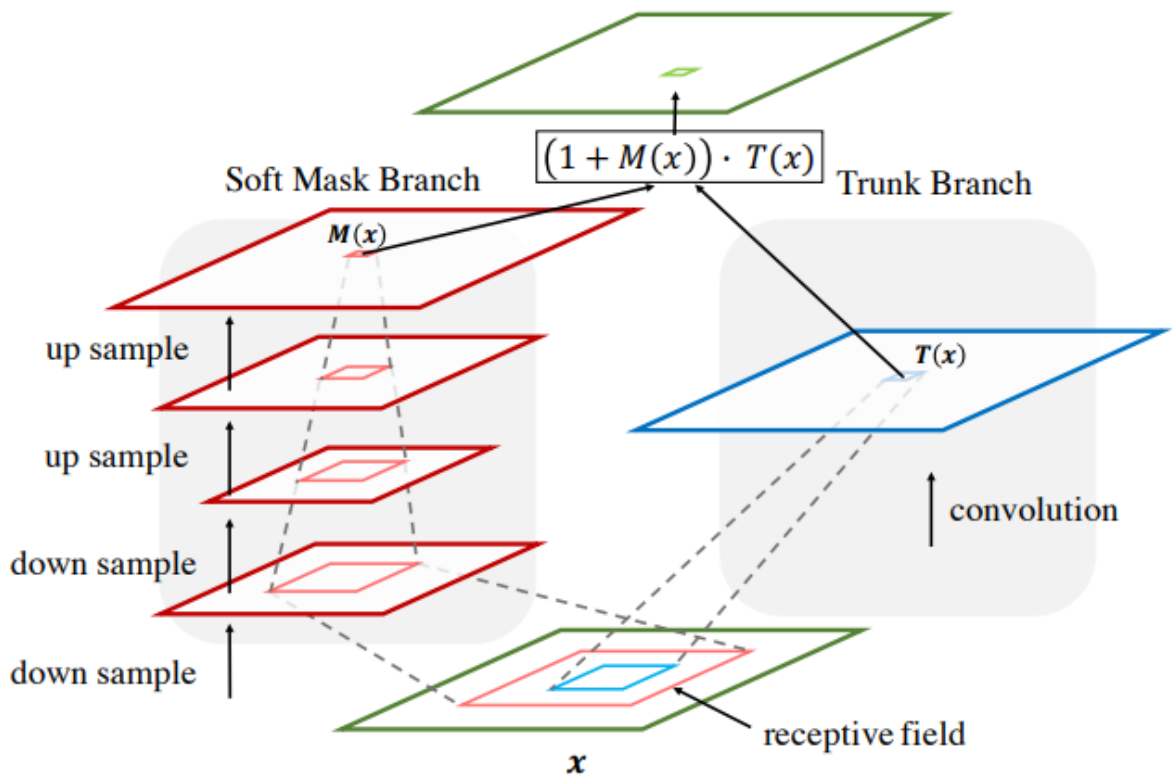


Figure 3: The receptive field comparison between mask branch and trunk branch.

(1) Trunck Branch

- **Trunck Branch**는 입력값으로부터 특징을 추출합니다.
Trunck Branch는 pre-activation resnet이나 resnext와 같은 최신 Model을 사용하여, 특징을 추출합니다. 출력값으로 $T(x)$ 를 출력합니다.

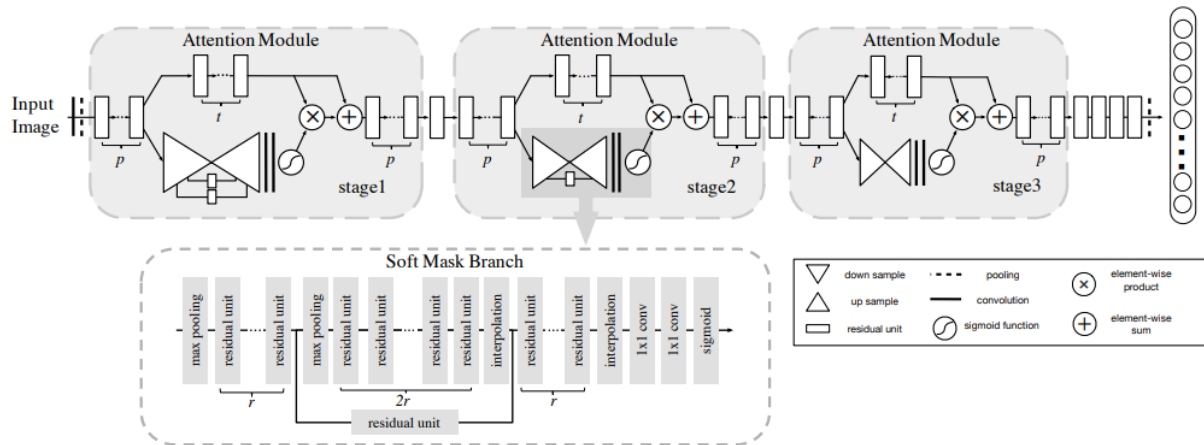
(2) Soft Mask Branch

- **Soft Mask Branch**는 0~1범위의 Mask를 생성합니다. $M(x)$ 로 표기합니다.
입력값을 최소 해상도까지 down-sample하고, up-sample하여 입력값까지 해상도를 키웁니다.
- Upsamle과정에서 선형 보간법(linear interpolation)을 사용한 후 softmax를 활용한 과정이후,
Input값에서 중요한 특징만이 살아남게 됩니다. Max-Pooling으로 가장 큰 값을 이용함으로써,
상대적으로 영향이 적은 픽셀들은 제거되는 것입니다.
그리고 크기를 키 trunk branck와 출력값의 크기와 같아지도록 합니다.



즉, Mask Branch를 활용하여, down-sample로 중요한 Pixel을 추출하고, Up-sample를 활용하여, 입력값에 중요한 Pixel의 위치를 얻는 것입니다.

그리고, Mask는 trunk Branch의 Output에 곱해져, 출력값을 정제합니다.
(논문에서는 feature selection이라고 부릅니다.)



Attention Module은 하나의 특징만을 capture하기때문에, 여러 Module을 쌓아서 활용해야 합니다.

Comparison & Conclusion

Network	params $\times 10^6$	FLOPs $\times 10^9$	Test Size	Top-1 err. (%)	Top-5 err. (%)
ResNet-152 [10]	60.2	11.3	224×224	22.16	6.16
Attention-56	31.9	6.3	224×224	21.76	5.9
ResNeXt-101 [36]	44.5	7.8	224×224	21.2	5.6
AttentionNeXt-56	31.9	6.3	224×224	21.2	5.6
Inception-ResNet-v1 [32]	-	-	299×299	21.3	5.5
AttentionInception-56	31.9	6.3	299×299	20.36	5.29
ResNet-200 [11]	64.7	15.0	320×320	20.1	4.8
Inception-ResNet-v2	-	-	299×299	19.9	4.9
Attention-92	51.3	10.4	320×320	19.5	4.8

Multiple Attention Module인 Residual Attention Network를 제안합니다.

두 가지에서 Benefit이 존재합니다.


1. Different Attention Modules은 different types of attention이 feature를 learning하도록 되어있습니다.

2. 각각의 **Attention Module**에서는 **bottom-up top-down feedforward**구조를 사용하며,
top-down Attention으로 이점이 옵니다.

추가로, **Deep Residual Attention Network**에서 **training**하는 것이 가능하다는 것입니다.

Residual Attention Network for Image Classification


In this work, we propose "Residual Attention Network", a convolutional neural network using attention mechanism which can incorporate with state-of-art feed forward network

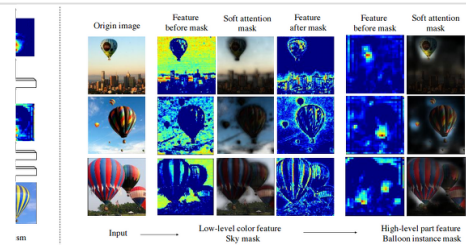
 <https://arxiv.org/abs/1704.06904>



[논문 읽기] Residual Attention Network(2017) 리뷰

안녕하세요! 이번에 소개할 논문은 Residual Attention Network 입니다. Residual Attention Network는 자연어 처리에서 활발하게 이용하고 있는 Attention에서 영감을 받아 탄생한 모델입니다. 실제

 <https://deep-learning-study.tistory.com/536>




Residual Attention Network - 2017 | DataCrew

여길봐 여길! :wave: Feature 생성에 도움을 주는 Attention!

Residual과 만난 Attention, Residual Attention Network Review.

2015년은 명실공히 ResNet의 해였습니다. 그만큼 영향력이 상당해

 <http://datacrew.tech/residual-attention-network-2017/>

