DeepLab V3 paper review

윤태우

1. Introduce

Image segmantation에서 Deep Convolutional Neural Network(DCNN)은 pooling, convolution striding을 통해 해상도를 줄여왔습니다.

이는 이미지의 특성맵을 잘 잡아냈으나, 그 과정에서 변형되는 이미지는 세부적인 예측을 방해했습니다.

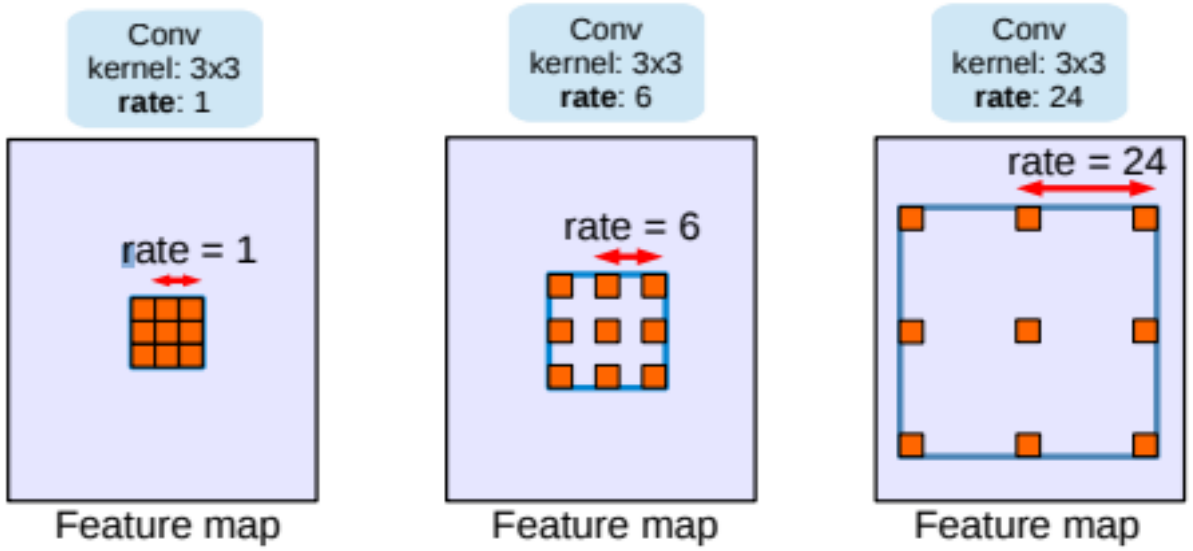
이 문제점을 극복하기 위해 이 논문에서는 atrous convolution을 시행합니다.

(atrous convolution은 dilated convolution으로 불리기도 합니다.)

2. DeepLab V3의 특징

❖ Atrous Convolution

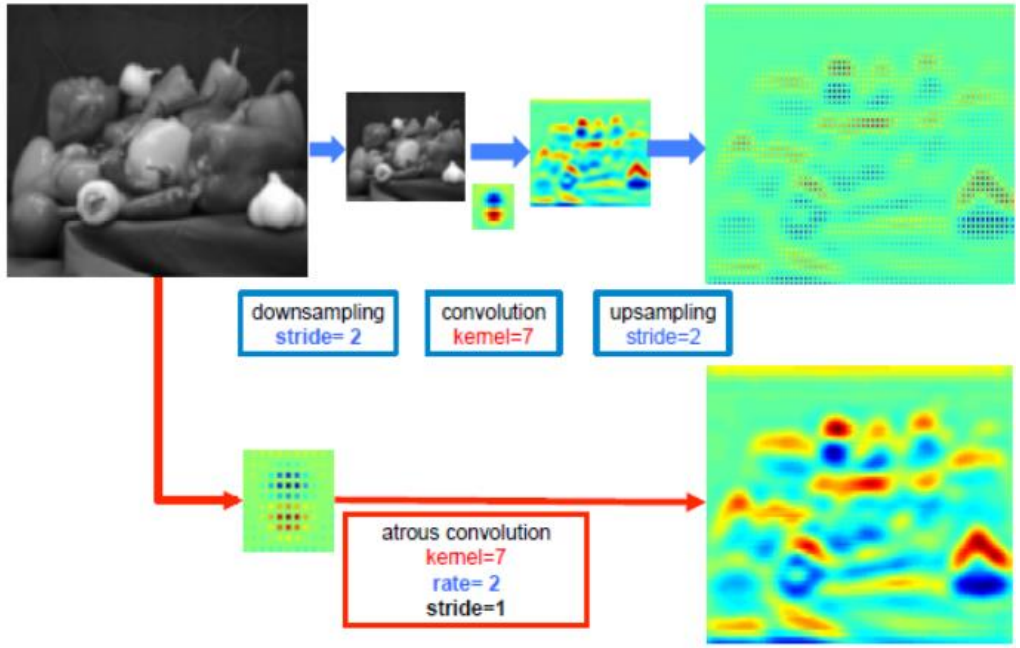
Atrous convolution의 atrous는 '구멍이 나있는' 이란 프랑스의 어원에서 유래했습니다. 이 convolution은 이름 그대로 구멍이 나 있는듯한 구조를 가집니다. 설정한 비율만큼 feature map의 값들을 떨어뜨려서 나타냅니다. 즉, kernel 사이 사이를 띄우는 것이죠.



2. DeepLab V3의 특징

❖ Atrous Convolution

Atrous convolution를 통해 kernel 사이를 띄우는 이유는 같은 computational cost로 더 큰 *receptive field를 나타낼 수 있기 때문입니다.



→ 일반적인 convolution

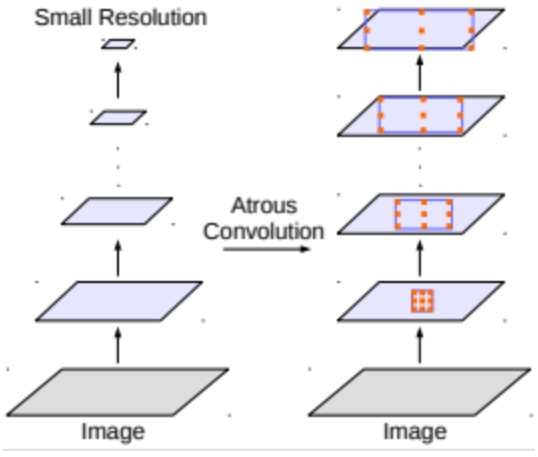
→ atrous convolution

Receptive field : 값이 존재하는 범위.
ex) 3X3 = 9, 5X5 = 25

2. DeepLab V3의 특징

❖ Atrous Convolution

기존의 convolution은 detail한 특성 보다는 대상의 존재를 파악하는데에 집중했습니다.
즉, detail information 보다는 global information에 포커싱을 두었습니다.
Image segmentation에서는 픽셀단위의 조밀한 예측을 하기 때문에 detail information에 포커싱 해야 더 좋은 결과를 낼 수 있습니다.
이러한 이유 때문에 pooling layer를 없애고 atrous convolution을 통해 receptive field를 확장시키는 것입니다.

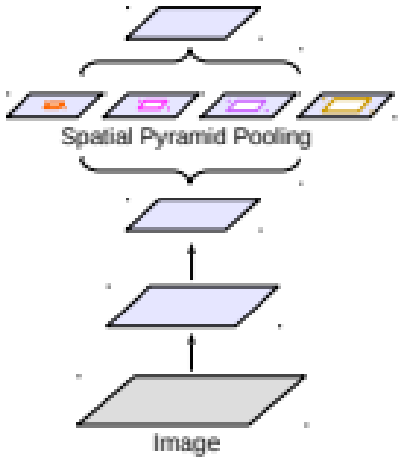


2. DeepLab V3의 특징

❖ Spatial Pyramid Pooling(SPP)

Spatial Pyramid Pooling(SPP)는 convolution layer를 거쳐 생성된 feature map들을 인풋으로 받고 여러 영역으로 나눕니다.

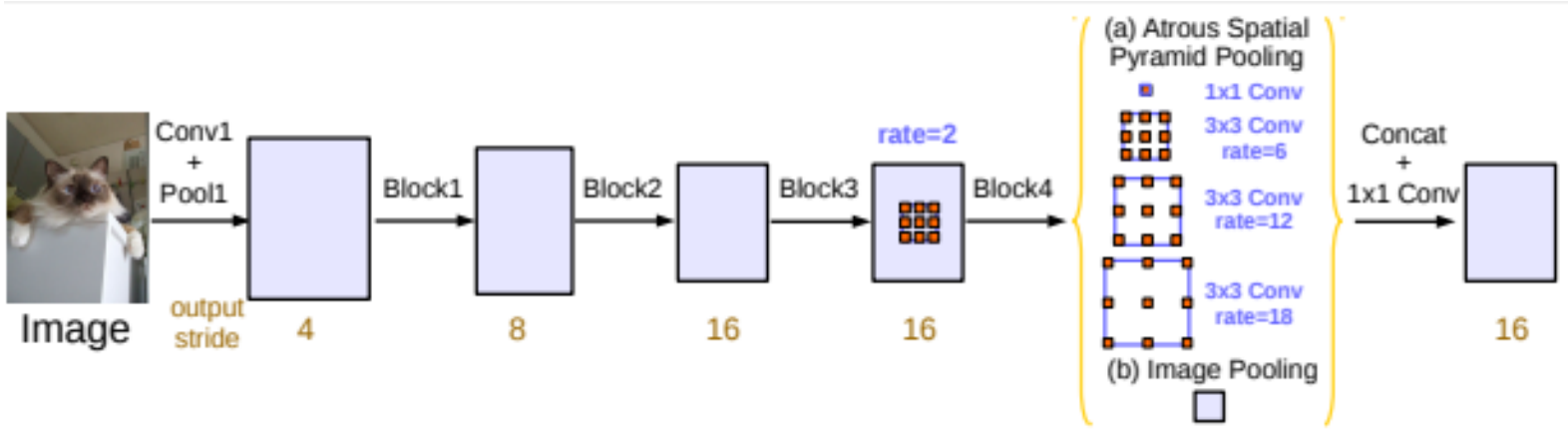
4X4, 2X2, 1X1로 나눈다면 각각이 하나의 피라미드가 되어 총 3개의 피라미드가 있는것입니다. 이 피라미드의 한칸을 bin이라 하고, 4X4의 경우 총 16 bin, 2X2는 4 bin 입니다.



2. DeepLab V3의 특징

❖ Atrous Spatial Pyramid Pooling(ASPP)

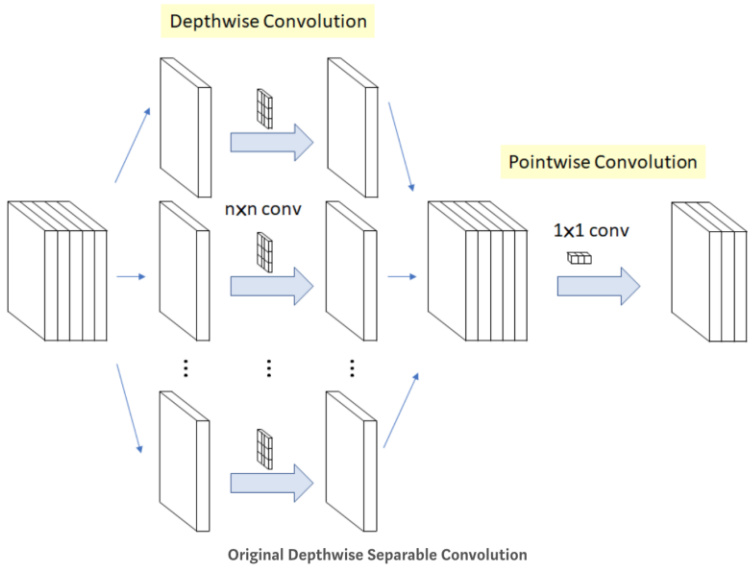
ASPP는 Spatial Pyramid Pooing에 atrous convolution을 더한것입니다.
기존 SPP의 convolution을 atrous convolution으로 대체한 것이죠.
여러 Grid scale에서 pooing이 진행된 값들끼리 concatenate 합니다.



2. DeepLab V3의 특징

❖Depthwise Separable Convolution

Depthwise Separable Convolution은 연산량과 파라미터 수를 줄여 학습 속도를 증가시키기 위하여 사용합니다. 명칭 그대로를 풀어서 말하자면 depthwise : 깊이별로 separable : 나누어서 convolution한다는 말입니다.



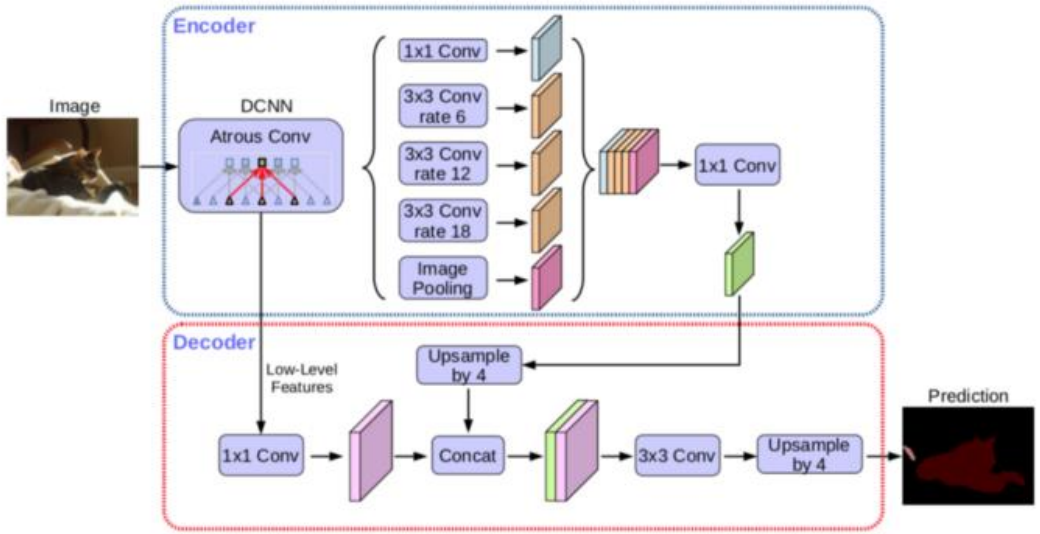
2. DeepLab V3의 구조

❖ 인코더-디코더

DeepLabV3 모델은 인코더와 디코더가 나누어져 있습니다. 인코더는 2개의 모델을 사용하는것을 제시합니다.

인코더 a는 ResNet을 backbone으로 사용합니다.

인코더 b는 Xception을 backbone으로 사용합니다.



2. DeepLab V3의 구조

❖ 인코더-디코더

Low level feature에 1X1 convolution을 하는 이유는 인코더의 결과물과 채널을 줄이기 위해서입니다.