DeepLapV3+ Model

Introduction

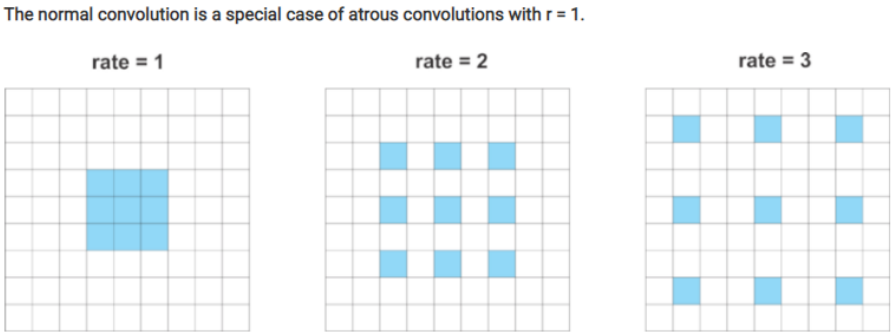
DeepLap V3+ cùng với các phiên bản trước của nó được tạo ra và phát triển bởi Google. Model này đem lại hiệu quả cao trong việc Segmentation, gán nhãn cho mỗi pixel trong một bức ảnh hoặc video. Từ khi được giới thiệu lần đầu vào năm 2016, model này đã có nhiều cải tiến qua các phiên bản:

- **DeepLab-V1**: Sử dụng ***atrous convolution (1)*** để kiểm soát độ phân giải của những đặc trưng thu được khi qua qua CNN.
- **DeepLab-V2**: Sử dụng ***atrous spatial pyramid pooling (ASPP) (2)*** để giải quyết các vấn đề mắc phải đối với các đối tượng có tỉ lệ khác nhau và cải thiện độ chính xác của model.
- **DeepLab-V3**: Thêm vào ASPP image-level feature và áp dụng batch normalization để dễ dàng training hơn.
- **DeepLab-V3+**: Mở rộng DeepLabv3 với decoder module đơn giản nhưng hiệu quả mới nhằm cải thiện kết quả segmentation.

Related Technique

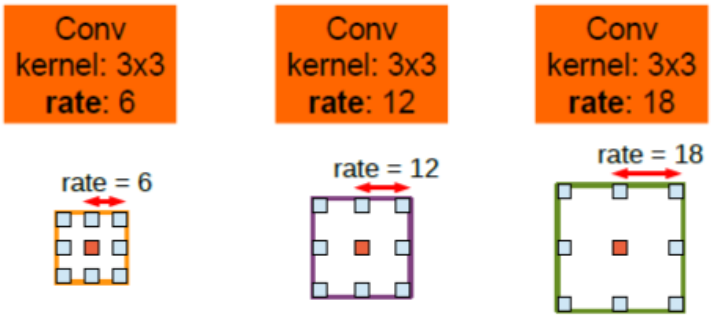
Atrous convolution (1)

Giới thiệu tham số rate được mô tả như sau:



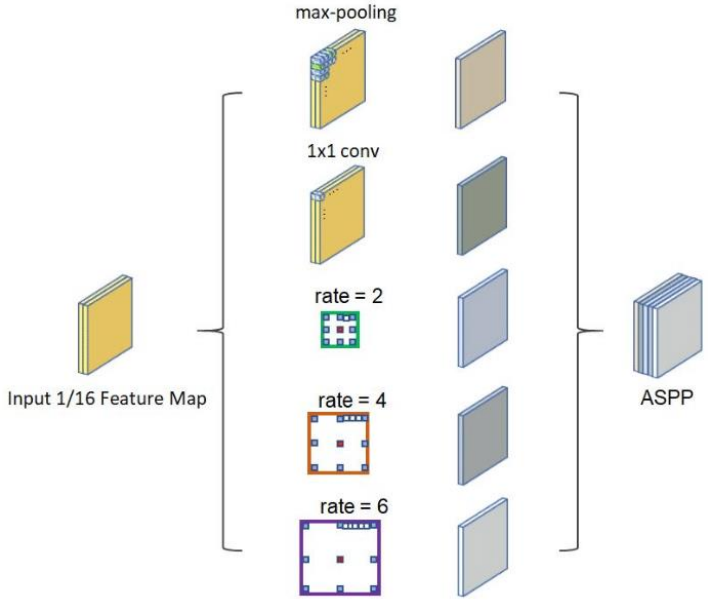
Việc lấy feature của Convolution network có tham số rate như thế này giúp những đối tượng có tỉ lệ nhỏ vẫn không bị mất đi. Vì đặc trưng, thông số của các object nhỏ không tập trung lại 1 chỗ và có thể mất đi qua nhiều lần pooling, thay vào đó nó được rải rác theo hệ số rate, nên dù pooling nhiều lần vẫn có thể giữ được đặc trưng cơ bản.

Dựa vào thực nghiệm cho thấy, Atrous convolution hoạt động hiệu quả với rate là 6, 12, 18.



Autrous spatial pyramid pooling (ASPP) (2)

ASPP bao gồm Max-pooling, 1x1 convolution và 3 atrous convolution với tỉ lệ tương ứng, ở ví dụ dưới là 2, 4, 6. Trong thực tế, deeplabv3+ sử dụng rate = [6, 12, 18] với output-ratio=16 và rate = [12, 24, 36] với output-ratio=8. ASSP tổng hợp các đặc trưng của đầu vào thông qua các layer kể trên, với các rate khác nhau, các layer này đảm bảo có thể lấy được nhiều đặc trưng nhất có thể, đặc biệt là các object nhỏ



ASPP đi qua 1x1 convolution để cho ra kết quả của phần encoding trong DeepLab model

Batch normalization (3)

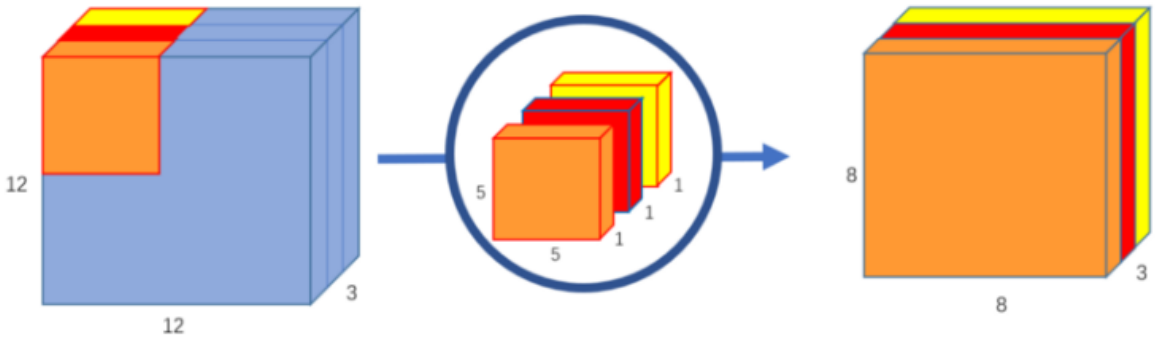
Chuẩn hóa dữ liệu về zero mean và unit-variance: giải quyết vấn đề non zero mean và high variance. Non zero mean là hiện tượng dữ liệu không phân bố quanh giá trị 0, kết hợp với high variance khiến dữ liệu có thành phần rất lớn hoặc rất nhỏ. Khi đi qua các activation có vùng bão hòa thì nhiều giá trị ra sẽ rơi vào vùng này dẫn tới gradient = 0 tại vùng bão hòa, kết quả train ko được cải thiện.

Depthwise Separable Convolutions

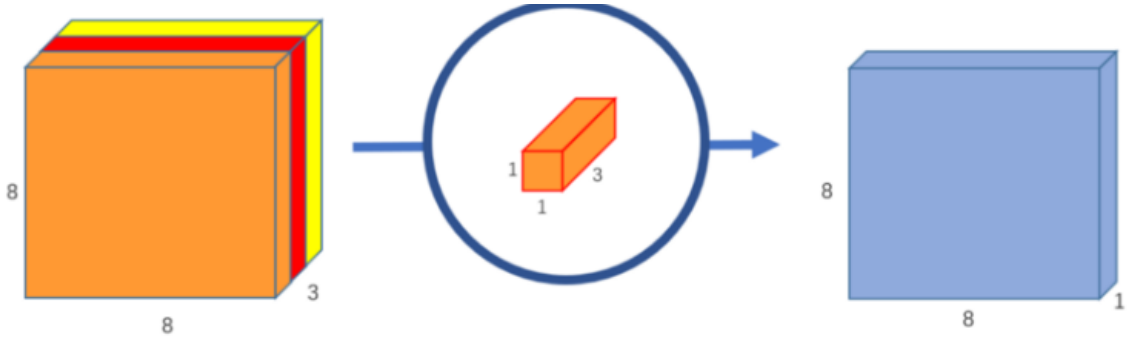
Depthwise Convolutions là một kỹ thuật thực hiện Convolution với số lượng tính toán ít hơn thông thường. Bằng cách chia convolution thành 2 bước:

- Depthwise convolution (1)
- Pointwise convolution (2)

Ví dụ, ta có bức ảnh 12x12x3 và ta muốn dùng kernel 5x5 với input này để tạo ra output 8x8x1. Thông thường, vì ảnh có 3 chanels nên ta có 3 kernels 5x5 cho mỗi input channel. Tuy nhiên khi chia ra 2 steps, ta chỉ cần 1 kernel 5x5 để tạo thành 8x8x3 sau đó dùng kernel 1x1x3 để đưa về 8x8x1.

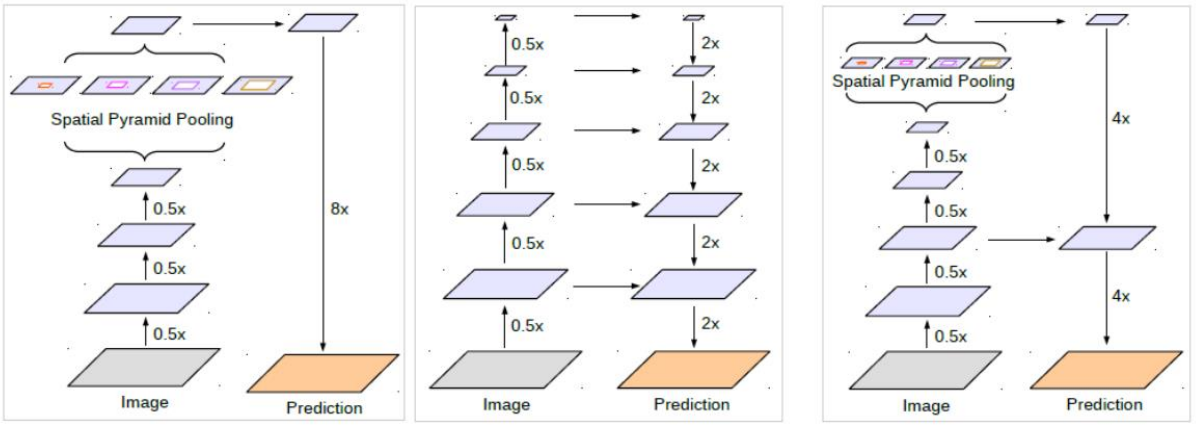


(1)



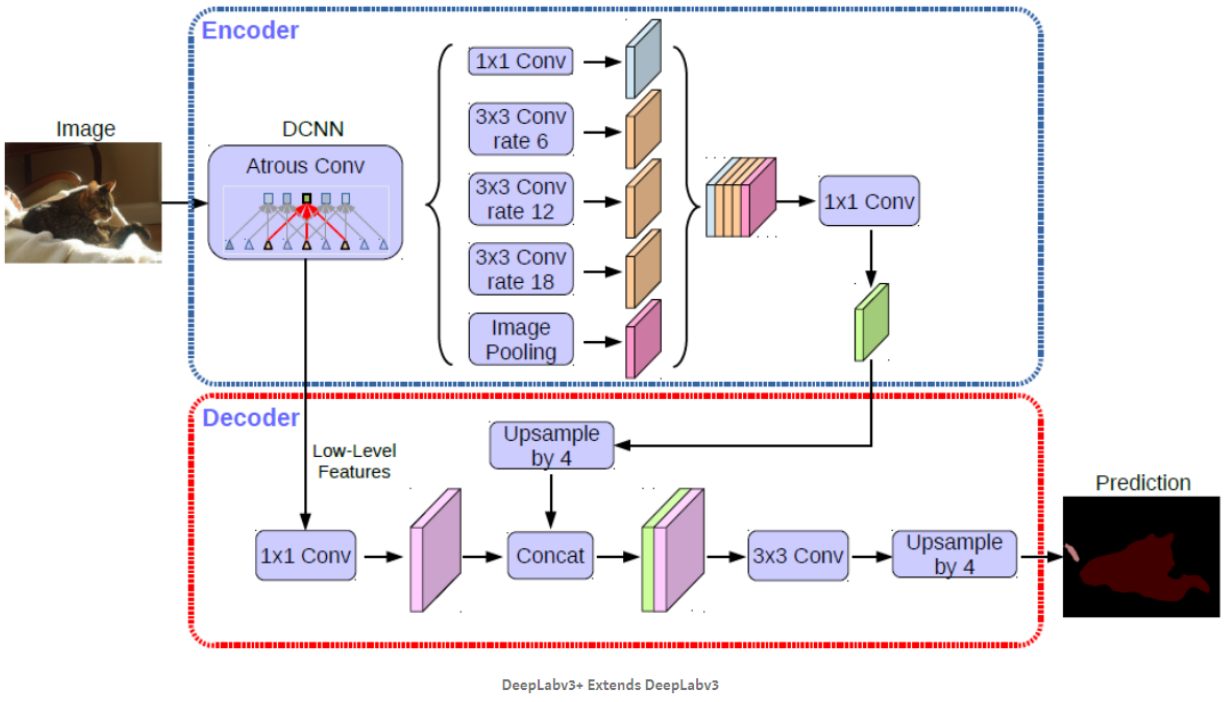
(2)

Architecture



(a) Spatial Pyramid Pooling (b) Encoder-Decoder (c) Encoder-Decoder with Atrous Conv

DeepLabv3+ có sự kết hợp giữa Spatial Pyramid Pooling và Encoder-Decoder(U-net). Được mô tả cụ thể như sau:

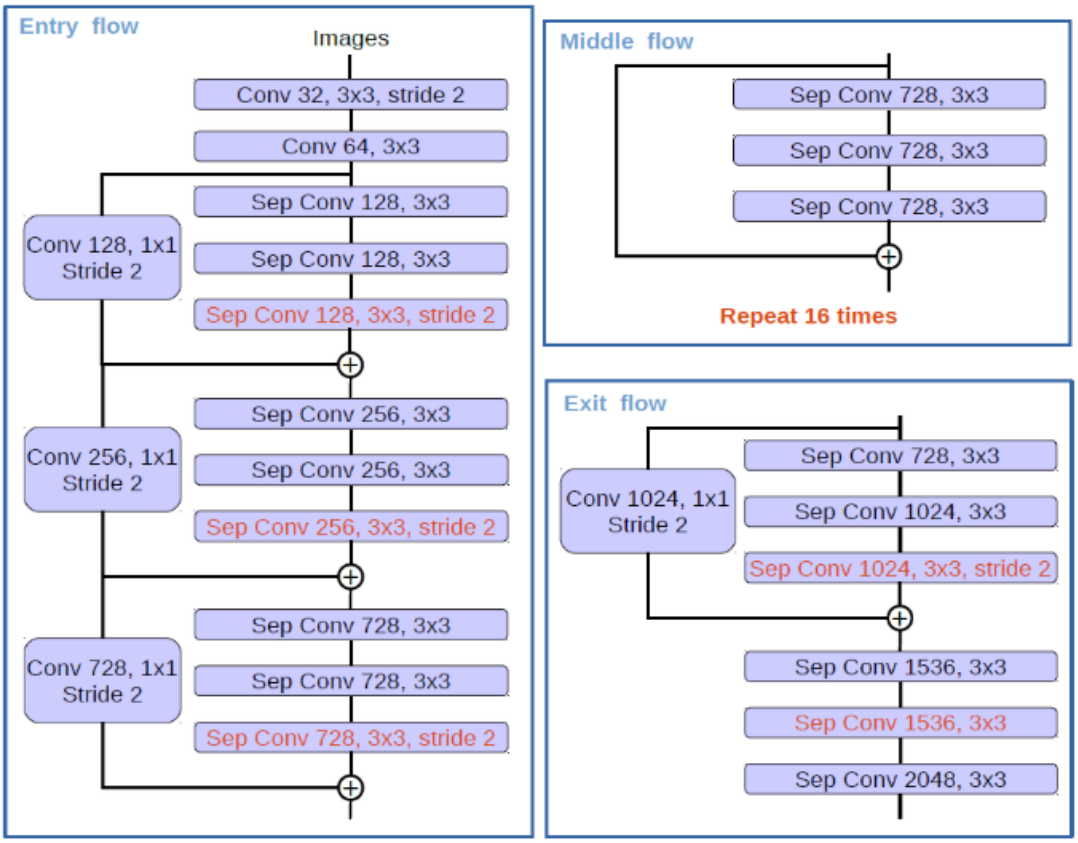


DeepLabv3+ Extends DeepLabv3

DCNN là phần Modified Aligned XCeption được mô tả ở phần dưới. Sau khi có được output từ CDNN, ta thực hiện lấy các đặc trưng bằng các layer conv với kernel và rate khác nhau cùng với image pooling, sau đó dùng kernel 1x1 cho conv layer để trích xuất các feature trước khi thực hiện upsampling. Kq này được kết hợp với 1 phần tương ứng của encoder để tiếp tục upsampling và cho ra kq qua activation thích hợp

Modified Aligned XCeption as Encoder

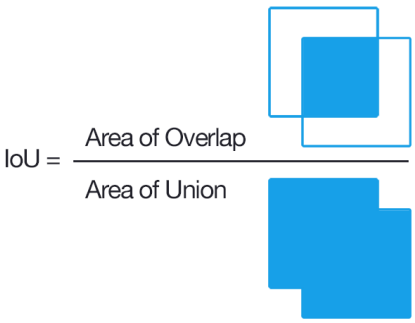
- Sử dụng nhiều layers hơn
- Tất cả max-pooling layers được thay bằng depthwise separable convolution với stride = 2
- Sau mỗi depthwise convolution, dùng batch normalization và ReLU



Modified Aligned Xception

Tham số để đánh giá kết quả của model: mIOU (mean Intersection Over Union)

IOU (Intersection Over Union) của 1 đối tượng là một số liệu dùng để đo lường độ chính xác khi segment được tính dựa vào công thức dưới:



Với area được tính bằng số pixel.

mIOU là trung bình IOU các classes có trong ảnh

C:\Program Files\CMake\bin