



Bài 4: Mô hình DeepLab

AI Research Team - Sun* AI Research

Nội dung chính

01

Lý thuyết mô hình

02

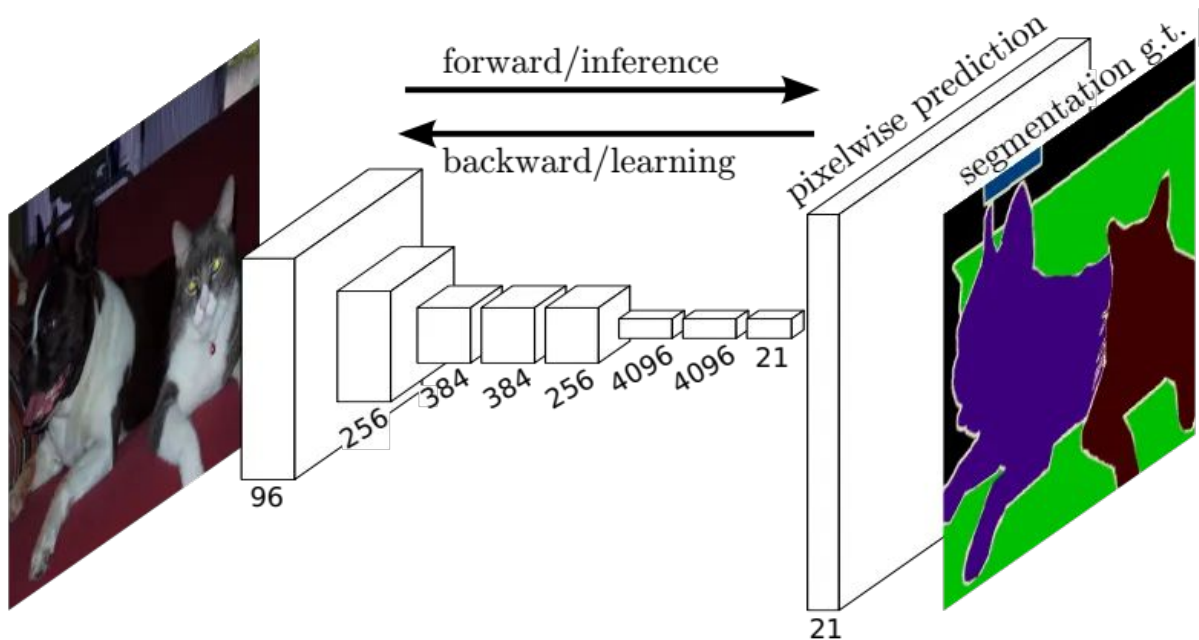
Thực hành với
PyTorch



01

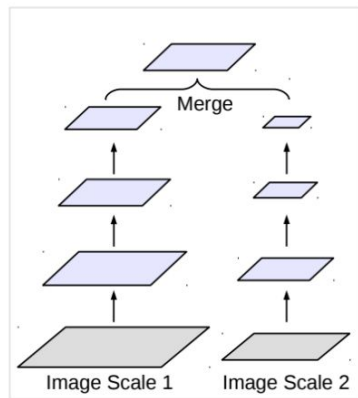
Lý thuyết mô hình DeepLab

Nhắc lại Deep Learning cho bài toán Semantic Segmentation

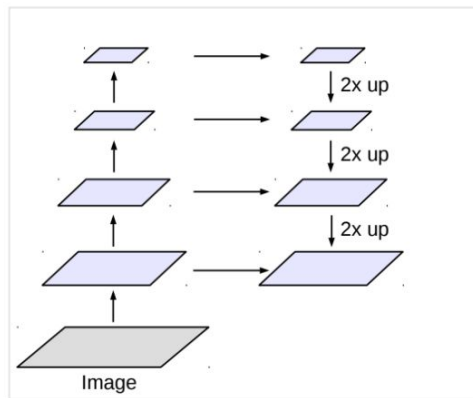


Nguồn: <https://towardsdatascience.com/review-fcn-semantic-segmentation-eb8c9b50d2d1>

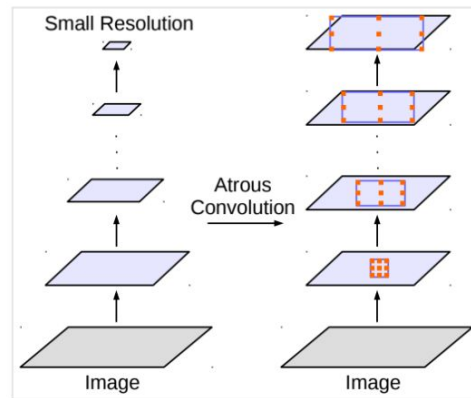
Nhắc lại Deep Learning cho bài toán Semantic Segmentation



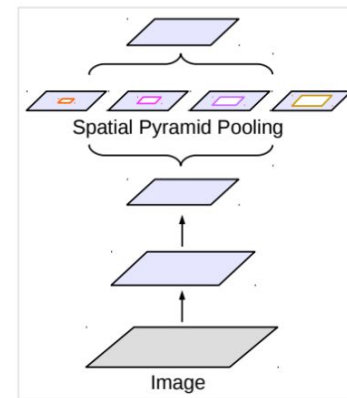
(a) Image Pyramid



(b) Encoder-Decoder



(c) Deeper w. Atrous Convolution



(d) Spatial Pyramid Pooling

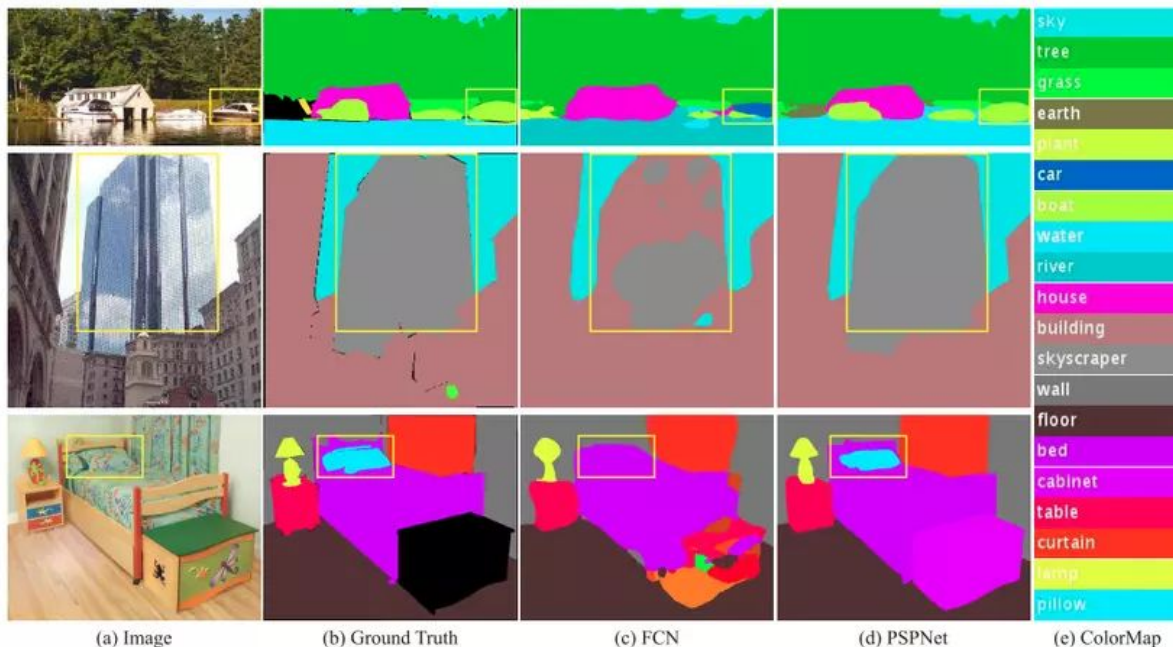
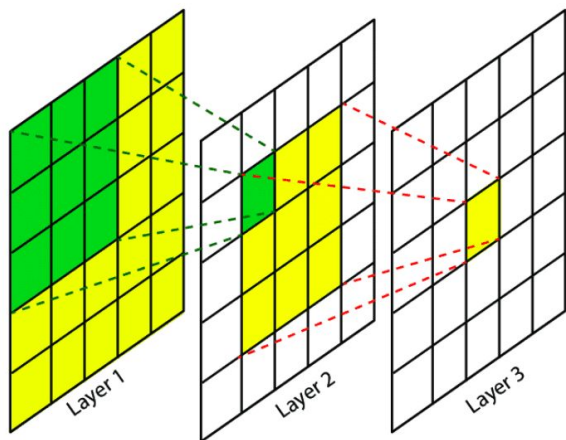
Figure 2. Alternative architectures to capture multi-scale context.

Nguồn: Paper DeepLabv3 (<https://arxiv.org/pdf/1706.05587v3.pdf>)

Nhắc lại vấn đề về Receptive Field

Các mô hình Deep Learning tìm cách mở rộng Receptive Field?

- PSPNet dùng PPM để tăng RF
- DeepLab?



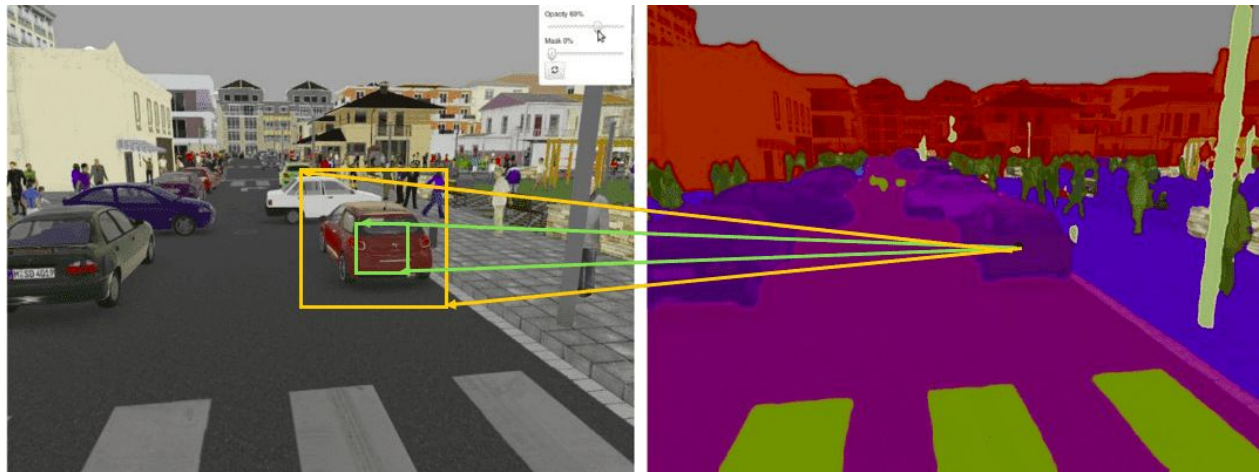
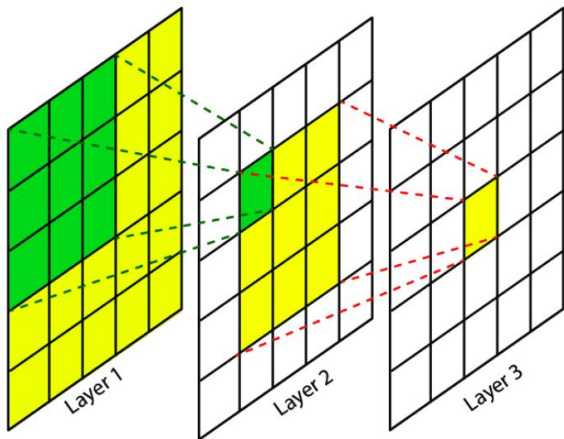
Nguồn: [Understanding the receptive field of deep convolutional networks](#) | AI Summer ([theaisummer.com](#))

Nguồn: Paper Pyramid Scene Parsing Network

Nhắc lại vấn đề về Receptive Field

Các mô hình Deep Learning tìm cách mở rộng Receptive Field?

- PSPNet dùng PPM để tăng RF
- DeepLab?

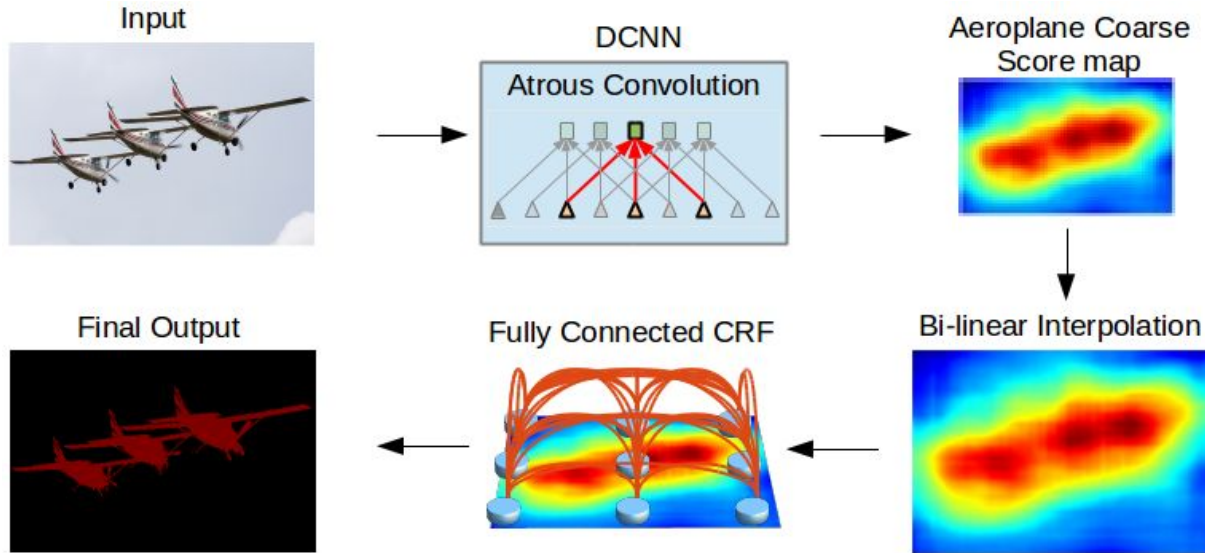


Nguồn:

<https://developer.nvidia.com/blog/image-segmentation-using-digits-5/>

Nguồn: [Understanding the receptive field of deep convolutional networks](https://theaisummer.com/understanding-the-receptive-field-of-deep-convolutional-networks/) | AI Summer (theaisummer.com)

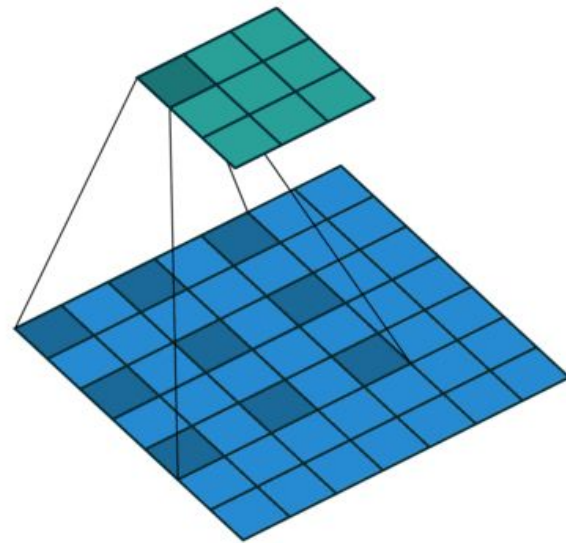
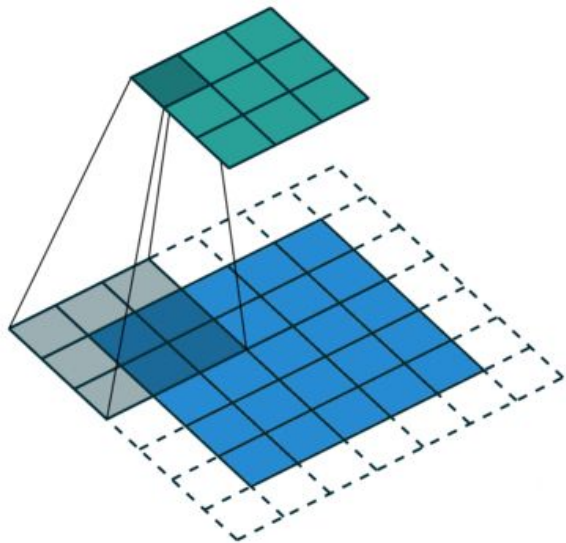
DeepLab



- Về cơ bản giống với kiến trúc Fully Convolutional Network (FCN)
- Sử dụng Atrous Convolution (Dilated Convolution) để mở rộng Receptive Field
- Sử dụng thêm thành phần Fully Connected Conditional Random Field để làm mịn kết quả phân vùng

Nguồn Paper DeepLab: (<https://arxiv.org/pdf/1606.00915.pdf>)

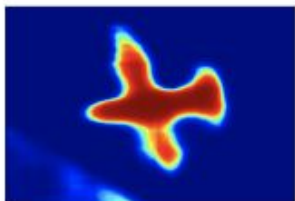
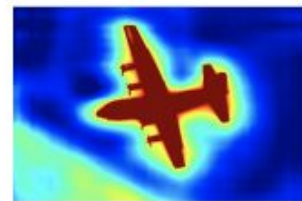
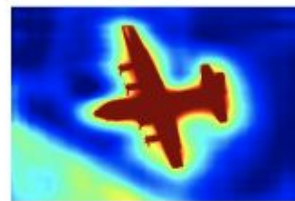
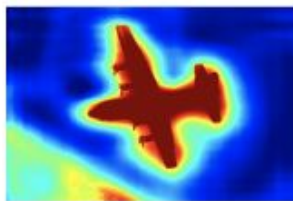
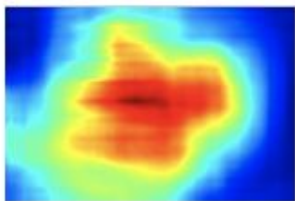
Dilated Convolution



- Dilated conv giống với phép conv thông thường
- Dilated conv mở rộng kích thước kernel bằng cách thêm các số 0 (theo chỉ số rate)
- Từ đó làm tăng kích thước quét được của input, mở rộng Receptive Field cũng như không làm tăng số lượng tham số

Nguồn: <https://towardsdatascience.com/review-dilated-convolution-semantic-segmentation-9d5a5bd768f5>

Conditional Random Field



Image/G.T.

DCNN output

CRF Iteration 1

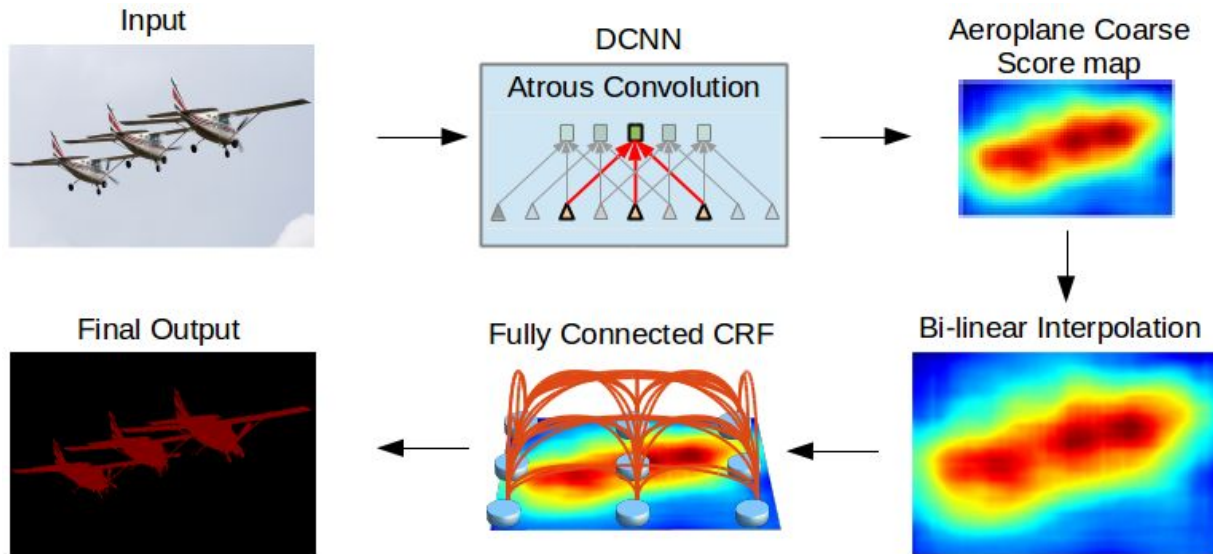
CRF Iteration 2

CRF Iteration 10

- Feature map đầu ra từ mạng backbone được phóng to bằng thuật toán nội suy
- Sử dụng CRF để làm mượt kết quả phân vùng

Nguồn: Paper DeepLab (<https://arxiv.org/pdf/1606.00915.pdf>)

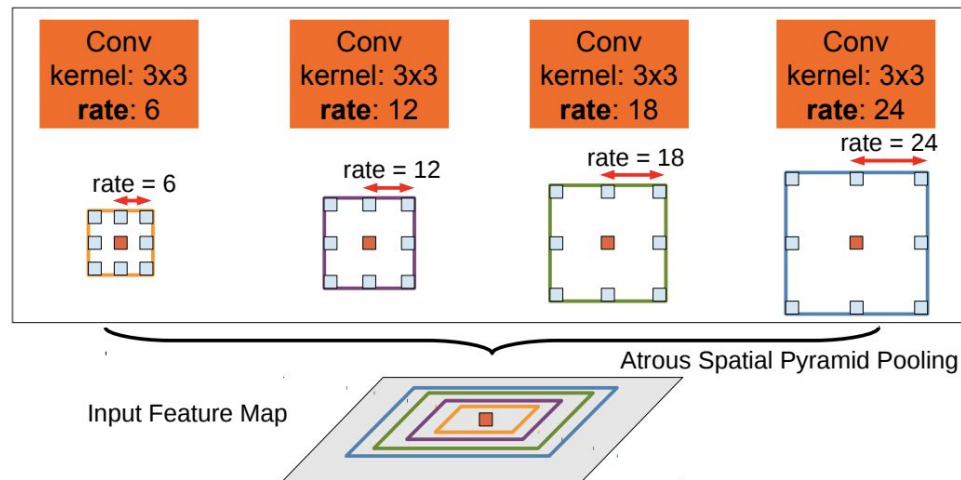
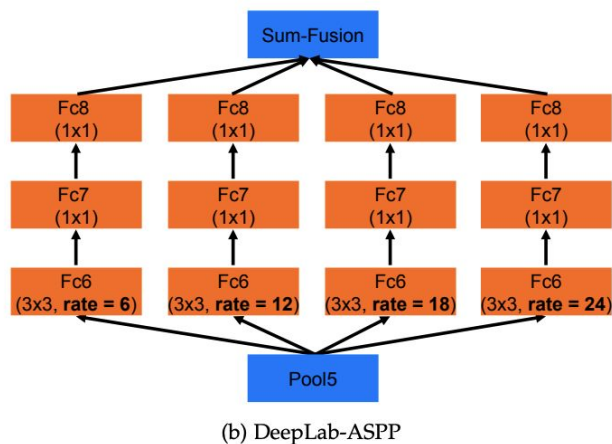
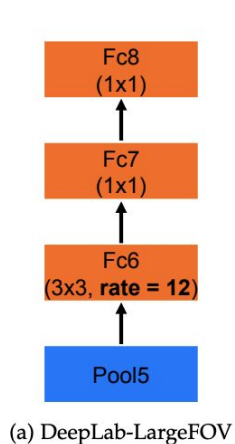
DeepLabv1 & v2



- Về cơ bản giống với kiến trúc Fully Convolutional Network (FCN)
- Sử dụng Atrous Convolution (Dilated Convolution) để mở rộng Receptive Field
- Sử dụng thêm thành phần Fully Connected Conditional Random Field để làm mịn kết quả phân vùng

Nguồn: Paper DeepLab (<https://arxiv.org/pdf/1606.00915.pdf>)

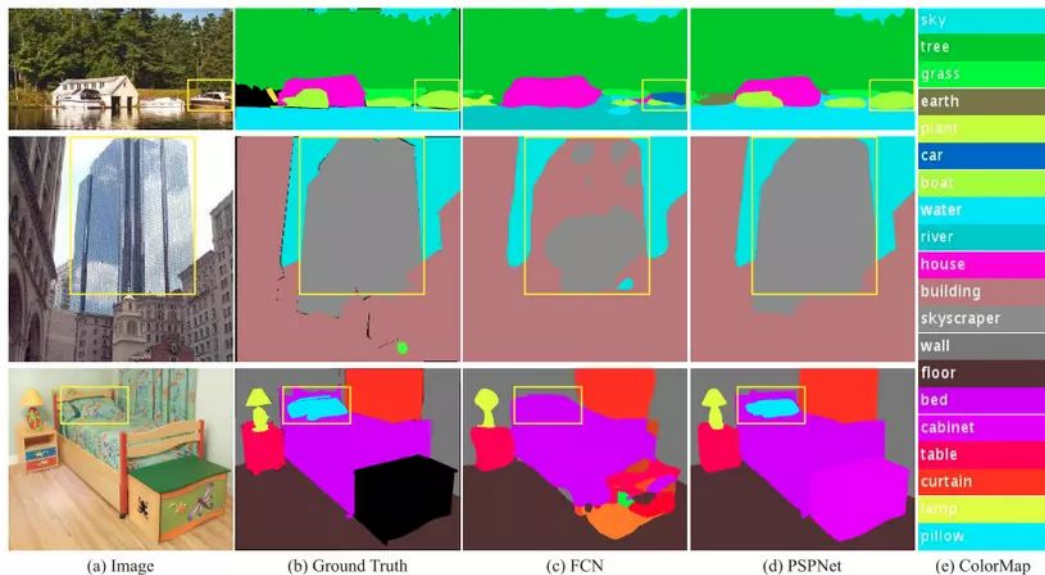
DeepLabV1 & V2



- DeepLabV2 sử dụng thêm Atrous Spatial Pyramid Pooling (ASPP)
- Atrous Spatial Pyramid Pooling (ASPP) giống với ý tưởng Pyramid Pooling Module, gồm nhiều dilated conv với các rate khác nhau để bắt được đặc trưng đa dạng hơn

Nguồn: Paper DeepLab (<https://arxiv.org/pdf/1606.00915.pdf>)

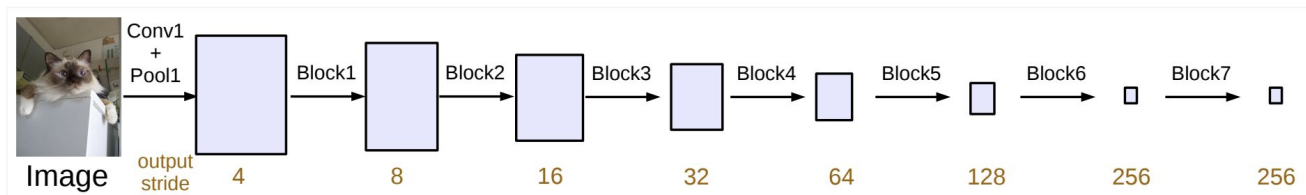
DeepLabV3



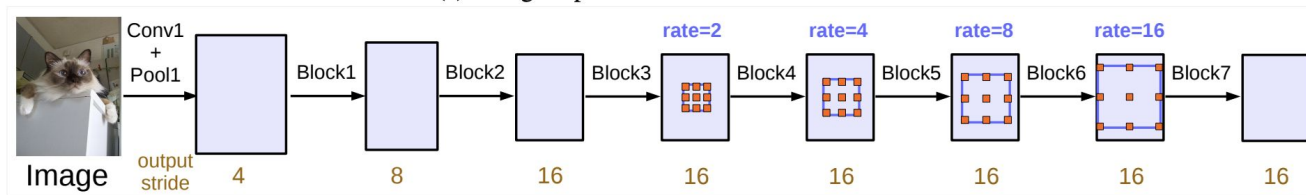
- Để có kết quả phân vùng tốt?
 - Tăng Receptive Field để mô hình nắm được thông tin đặc trưng ngữ nghĩa
 - Duy trì được thông tin đặc trưng cục bộ (local feature)

Nguồn: Paper Pyramid Scene Parsing Network

DeepLabV3



(a) Going deeper without atrous convolution.



(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output_stride = 16$.

Figure 3. Cascaded modules without and with atrous convolution.

- Yêu cầu bài toán: Phân loại chính xác tới từng pixel trong ảnh
- Khi feature map ở layer cuối trong mạng quá bé, gặp vấn đề khi phải phóng to trở về kích thước gốc
- Loại bỏ Fully Connected CRF

Nguồn: Paper DeepLabv3 (<https://arxiv.org/pdf/1706.05587v3.pdf>)

DeepLabV3

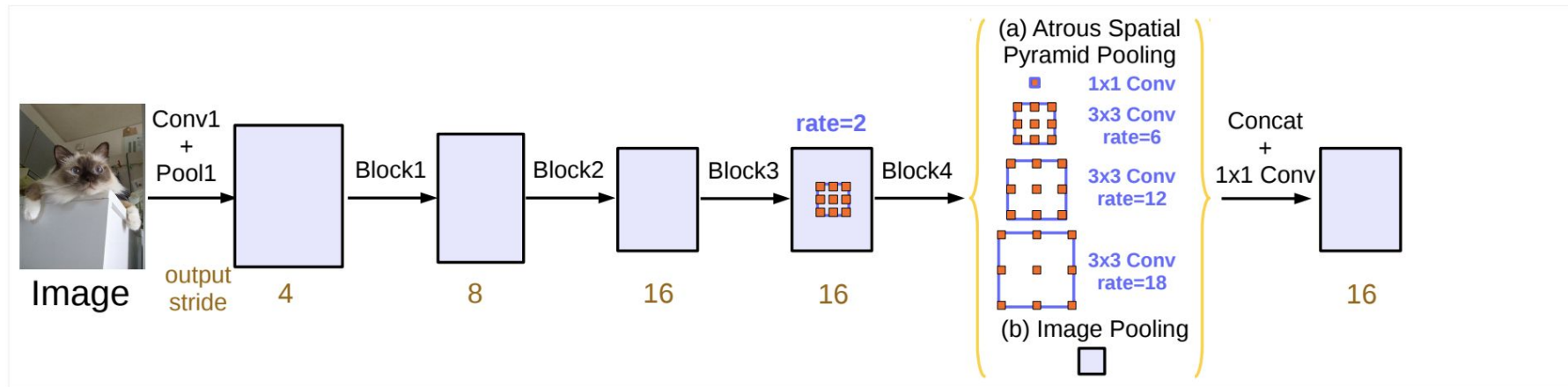


Figure 5. Parallel modules with atrous convolution (ASPP), augmented with image-level features.

- Vẫn sử dụng ý tưởng ASPP để đa dạng đặc trưng
- Sử dụng thêm kỹ thuật Multi Grid

Nguồn: Paper DeepLabv3 (<https://arxiv.org/pdf/1706.05587v3.pdf>)



02

**Thực hành
với PyTorch**