



DeepLab V1

: Semantic Segmentation with deep
convolutional nets and fully connected CRFs

16기 김연선



Problem

기존 Deep Convolutional Neural Network를
Semantic Segmentation에 적용했을 때 발생하는 2가지 문제

1) Single downsampling

- Max pooling layer & Downsampling & Stride 의 반복으로 Signal resolution reduction 발생

1) Spatial insensitivity (Spatial invariance)

- 기존의 DCNN의 경우 pixel 단위의 detail한 localization이 필요한 semantic segmentation에서는 공간적 정확도가 떨어짐



DeepLab V1 제안

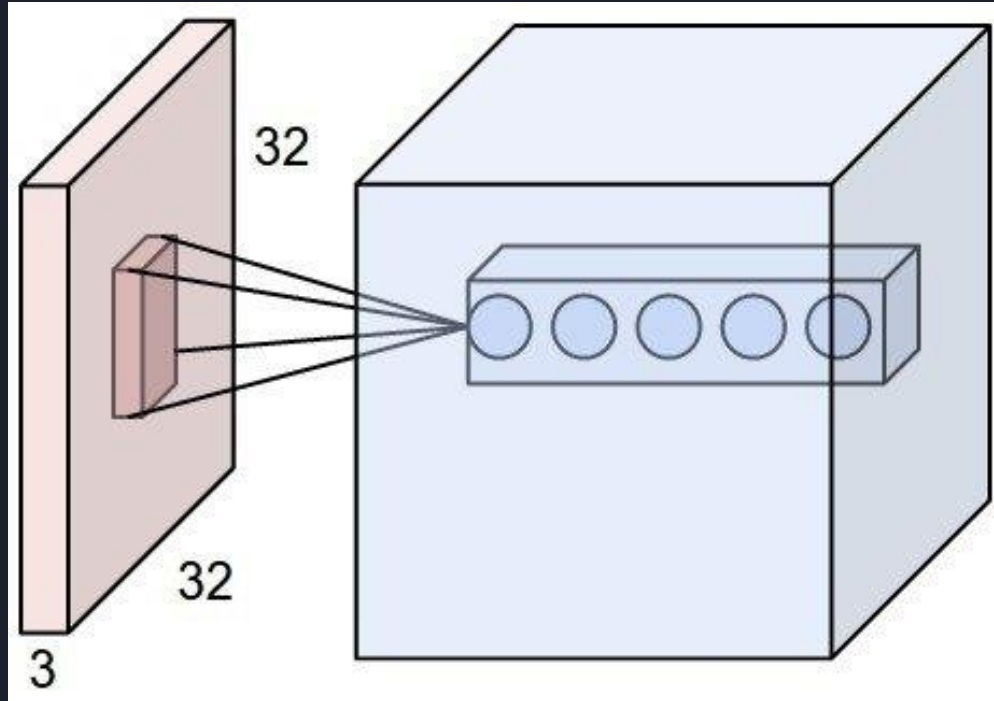
1. Atrous algorithm(hole algorithm) 적용

- 공간적 정보 손실을 줄이기 위해 convolution 사이에 빈 공간(hole)을 채워넣고 수행하는 Atrous Convolution 제안, 더 넓은 scale 확인 가능.

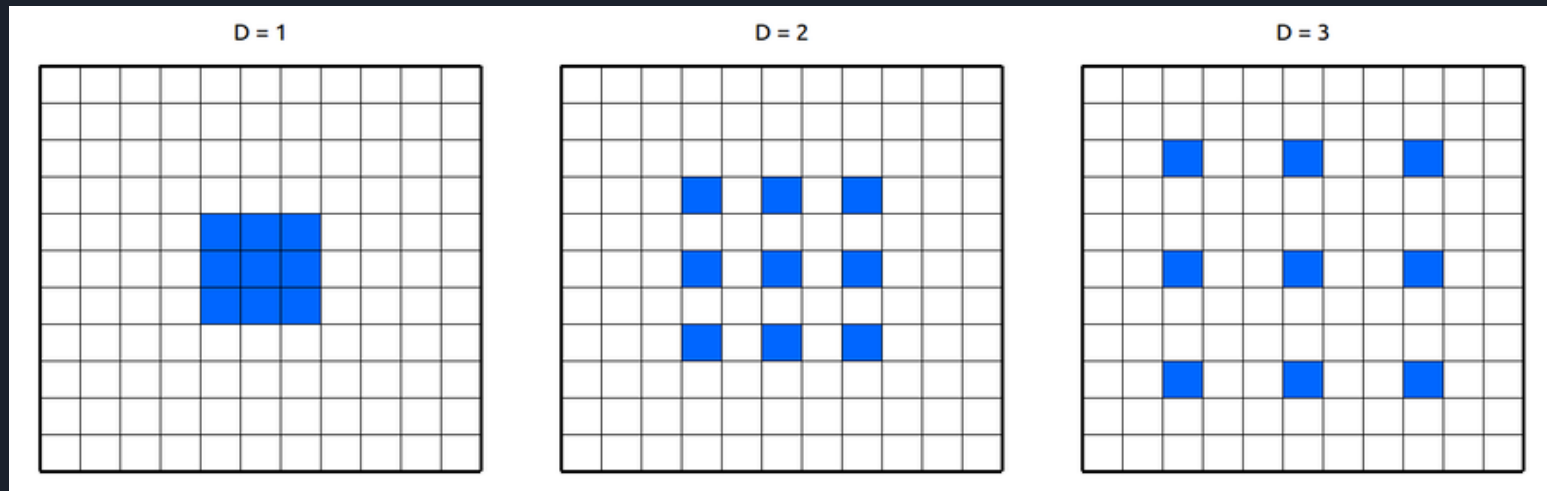
2. DCNN-based pixel-level classifier와 Fully Connected CRF 결합

- Conditional Random Field를 Convolutional network에 적용해 픽셀 단위 예측의 정확도 향상
- Segmentation의 localization 에 기여
- pixel- level classification(semantic segmentation)을 해결하기 위한 graphical model

Receptive Field



Atrous Convolution

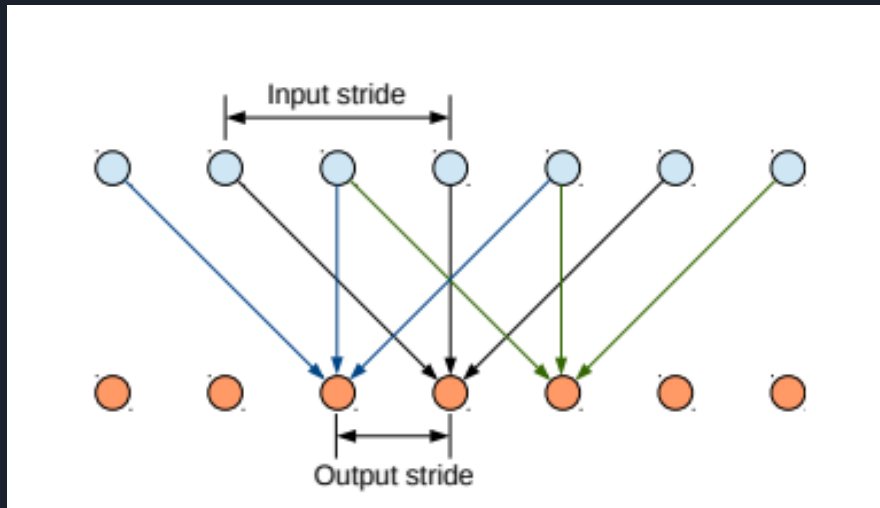


일반적인 Convolution 사이에 공간(hole)을 넣은 구조.

Dilated convolution: receptive field가 D 값에 따라 다른 Convolution

Convolutional Neural Network Dense Image Labeling

- VGG-16 를 dense feature extractor로 사용



- 기존 DCNN: receptive field가 크기 때문에, 병목 현상이 발생
 - 네트워크의 receptive field size 조절하는 'hole algorithm' 적용
- 1) VGG-16의 FC 계층 \rightarrow Convolution 으로 변환
 - 2) 마지막 2개의 max-pooling layer 제거, hole algorithm 도입하여 길이 증가
- 필터의 크기는 그대로 유지, input stride 2px or 4px 사용
 - CNN에 비해 연산량 Down, 효율적으로 계산



Detailed Boundary Recovery: Fully Connected Conditional Random Fields and Multi-Scale Prediction

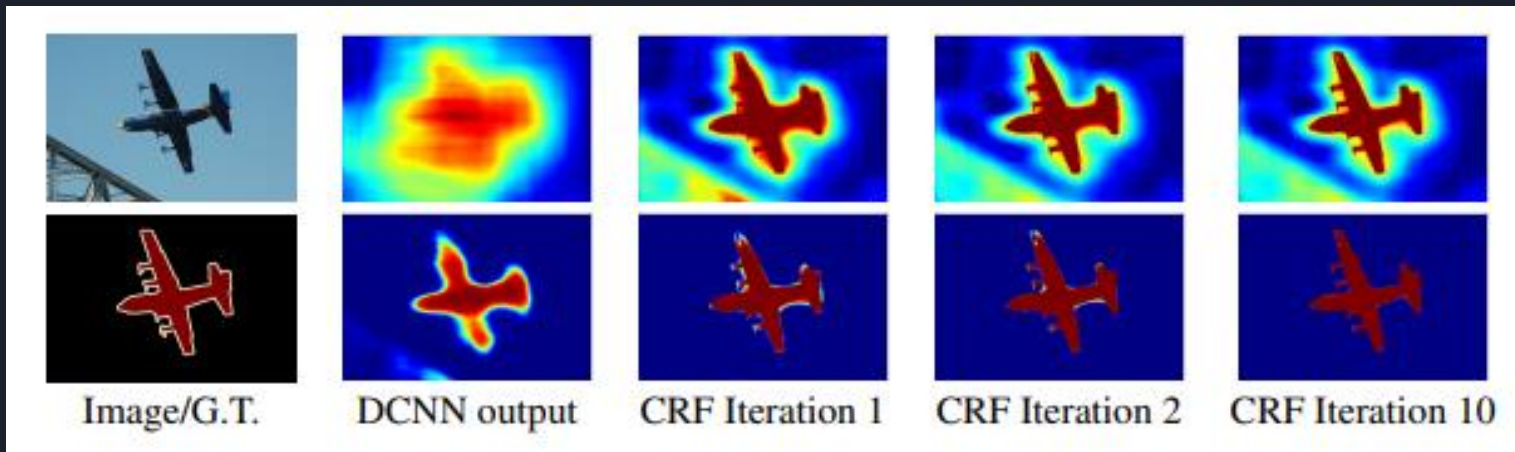
[Localization Challenge]

- DCNN Score map은 presence와 대략적 위치 예측에는 좋으나, 객체의 정확한 윤곽을 pin-pointing 하기에는 적합하지 않음
- Convolutional Network 를 활용한 Classification accuracy와 localization accuracy 사이의 trade-off 관계 有

[Localization Challenge 해결을 위한 접근법]

- 1) convolutional network의 여러 layer 정보를 활용하여 더 나은 추정
- 2) Super-pixel Representation 사용하여 localization task를 low-level segmentation method로 위임

Detailed Boundary Recovery: Fully Connected Conditional Random Fields and Multi-Scale Prediction



일반적으로 CRF는 noise가 있는 segmentation maps를 원활하게 하기 위해 사용되어 왔으며, 인접한 픽셀에 동일한 레이블에 할당하는 것을 선호함.

그에 비해 DCNN은 기존 short-range CRF를 Segmentation에 적용할 때 발생하는 feature map의 크기 감소로 인한 detail 정보 취득의 어려움을 해결하고자, Fully Connected CRFs 모델을 시스템에 결합함.

Detailed Boundary Recovery: Fully Connected Conditional Random Fields and Multi-Scale Prediction

Fully-Connected CRFs Energy Function

$$E(\mathbf{x}) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j) \leftarrow \text{Fully connected model}$$

Diagram illustrating the components of the energy function $E(\mathbf{x})$:

- $\theta_i(x_i)$ is derived from **From DCNN label probabilities**.
- $\theta_{ij}(x_i, x_j)$ is derived from a **Gaussian, pairwise** model, which is further defined by:
$$w_1 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2}\right) + w_2 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2}\right)$$
 - The first term is influenced by **Differences in position and intensity**.
 - The second term is influenced by **Just position**.

x : pixel의 label

I : Image

p_i, p_j : 픽셀의 위치

I_i, I_j : pixel color intensity

Unary term : DCNN을 통해 계산된 픽셀 i 에서의 label 할당 확률. CNN연산을 통해 얻은 수 있음

Pairwise term : 픽셀간 detail한 예측에 활용. pixel 값의 유사도와 위치 유사도를 함께 고려.

- 1) 첫 번째 커널: 비슷한 위치 비슷한 색상을 갖는 픽셀들에 대하여 비슷한 label이 붙을 수 있도록 함
- 2) 두 번째 커널: 픽셀의 근접도에 따라 smooth 수준을 결정함.

Multi Scale Prediction

- Boundary localization 정확도 향상을 위한 Multi-Scale Prediction 시도.
- input image & 첫 4개의 max-pooling layer의 output을
마지막 layer의 feature map에 연결된 two-layer MLP와
연결.
- MLP를 추가할 경우 localization 성능이 향상 되나 크진 않음

Method	mean IOU (%)
DeepLab	59.80
DeepLab-CRF	63.74
DeepLab-MSc	61.30
DeepLab-MSc-CRF	65.21
DeepLab-7x7	64.38
DeepLab-CRF-7x7	67.64
DeepLab-LargeFOV	62.25
DeepLab-CRF-LargeFOV	67.64
DeepLab-MSc-LargeFOV	64.21
DeepLab-MSc-CRF-LargeFOV	68.70

(a)

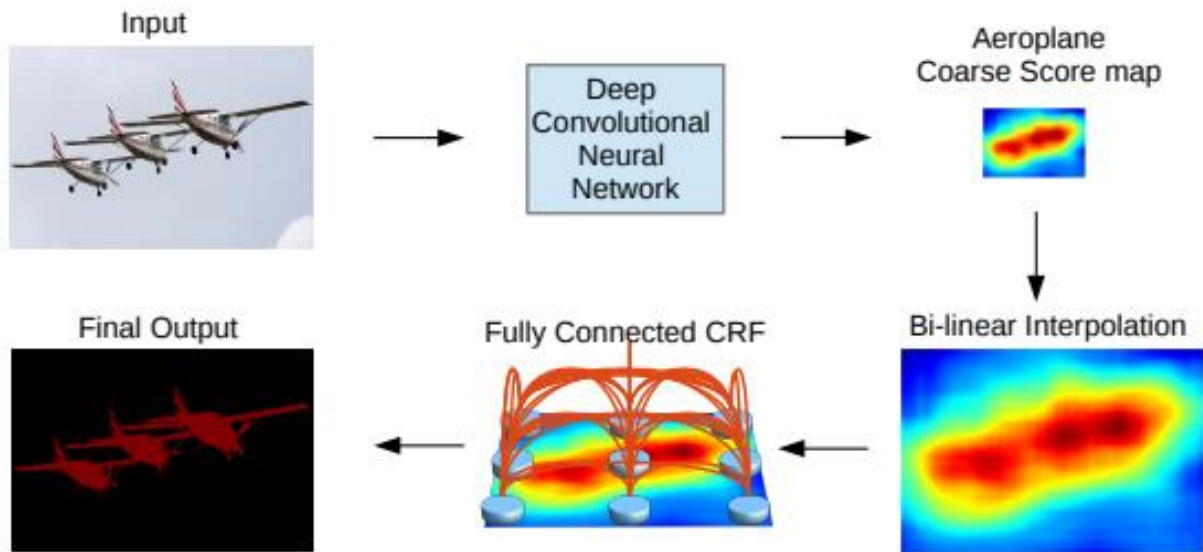
Method	mean IOU (%)
MSRA-CFM	61.8
FCN-8s	62.2
TTI-Zoomout-16	64.4
DeepLab-CRF	66.4
DeepLab-MSc-CRF	67.1
DeepLab-CRF-7x7	70.3
DeepLab-CRF-LargeFOV	70.3
DeepLab-MSc-CRF-LargeFOV	71.6

(b)

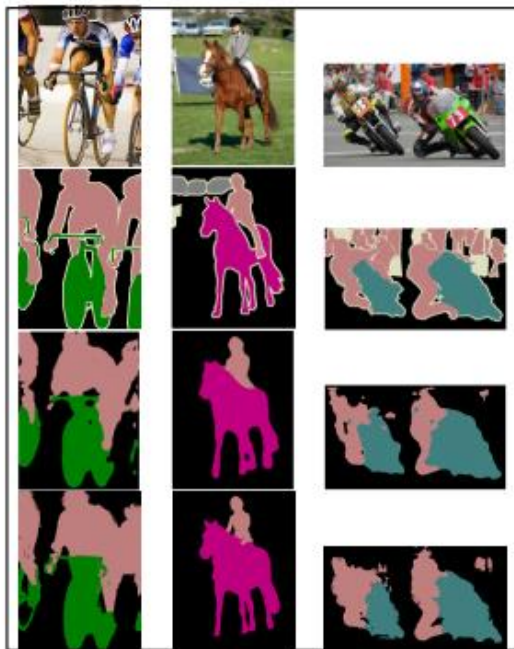
(A) PASCAL VOC 2012 valid set 모델 성능

(B) PASCAL VOC 2012 test set 모델 성능

DeepLab V1 요약



Experimental Evaluation



(a) FCN-8s vs. DeepLab-CRF



(b) TTI-Zoomout-16 vs. DeepLab-CRF



Experimental Evaluation

Field of View

- Atrous Algorithm을 통해 stride를 조절, FOV를 임의로 제어할 수 있게 함
- 해상도 저하 문제 해결, 연산 시간 절감

Multi-Scale features

- Field of view 의 효과에 Multi-Scale features 추가.
- Fully Connected layer의 kernel size와 Atrous Algorithm에 사용된 stride 값에 Multi-Scale features 추가 시 모델 성능 1.5% 향상



DeepLab-CRF test

- DeepLab-CRF-7X7
 - kernel size 7 x 7, stride 4
 - 67.64% 성능 (느림)
- DeepLab-CRF-4X4
 - FOV input stride 8, 성능 향상
- DeepLab-CRF-LargeFOV
 - kernel size 3X3, stride 12, 마지막 two-layer filter size 4096 -> 1024 로 축소.
 - DeepLab-CRF-7X7 성능과 일치, 속도 향상.



Conclusion

Deep Convolutional Neural Network

+

Fully connected conditional random fields idea

=

Semantically accurate predictions and detailed segmentation maps

감사합니다

