핸즈온 머신러닝

Chapter 14.10: 시맨틱 분할

Presentation: **Jeiyoon Park**

6th Generation, TAVE

Semantic Segmentation

1. 시맨틱 분할 (Semantic segmentation)

- 각 픽셀은 픽셀이 속한 객체의 클래스로 분류됨 (e.g. 자동차, 보행자, 건물)
- 클래스가 같은 건물은 구분되지 않음
- 일반적인 CNN을 통과할 때 점진적으로 위치정보를 잃음
(1 이상의 stride를 사용하는 layer들 때문)

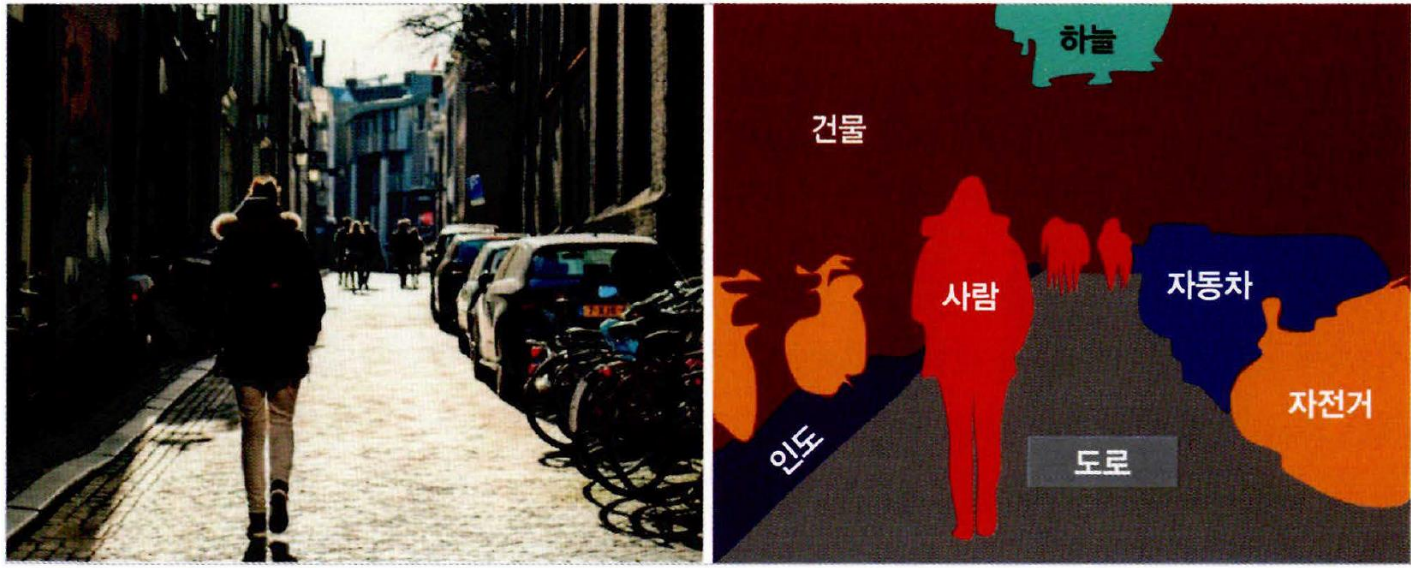


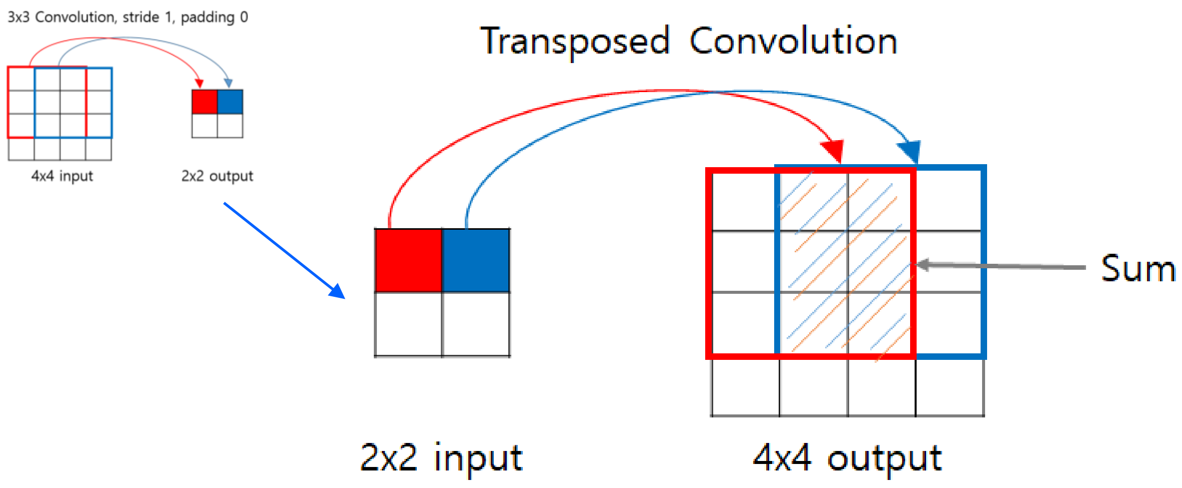
그림 14-26 시맨틱 분할

Semantic Segmentation

2. Fully Convolutional Networks for Semantic Segmentation *(Long and Shelhamer et al., CVPR 2015)*

(1) Pretrain된 CNN을 FCN으로 변환함. 전체 stride 합은 32. 이는 처음 이미지 보다 마지막 이미지가 32배 작은 feature map으로 출력되는 것을 의미함

(2) 해상도를 32배 올리는 Upsampling layer (transposed convolution layer)를 하나 추가함



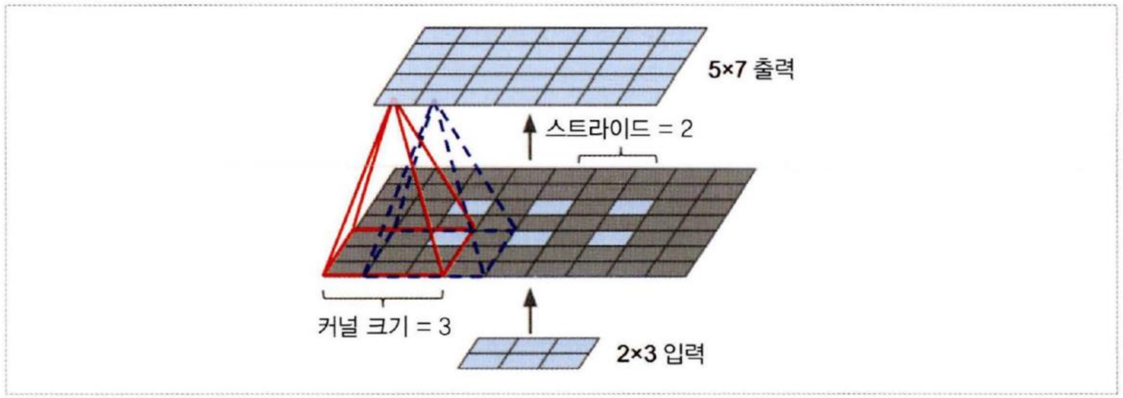


그림 14-27 전치 합성곱 층을 사용한 업샘플링

Semantic Segmentation

2. Fully Convolutional Networks for Semantic Segmentation *(Long and Shelhamer et al., CVPR 2015)*

(3) 여전히 떨어지는 정확도를 개선하기 위해 Skip architecture를 추가함
(e.g. 2배 upsampling & 아래층 output을 더해서 해상도를 두 배로 키움).
(4) 이를 초해상도(super-resolution) 이라고 함.

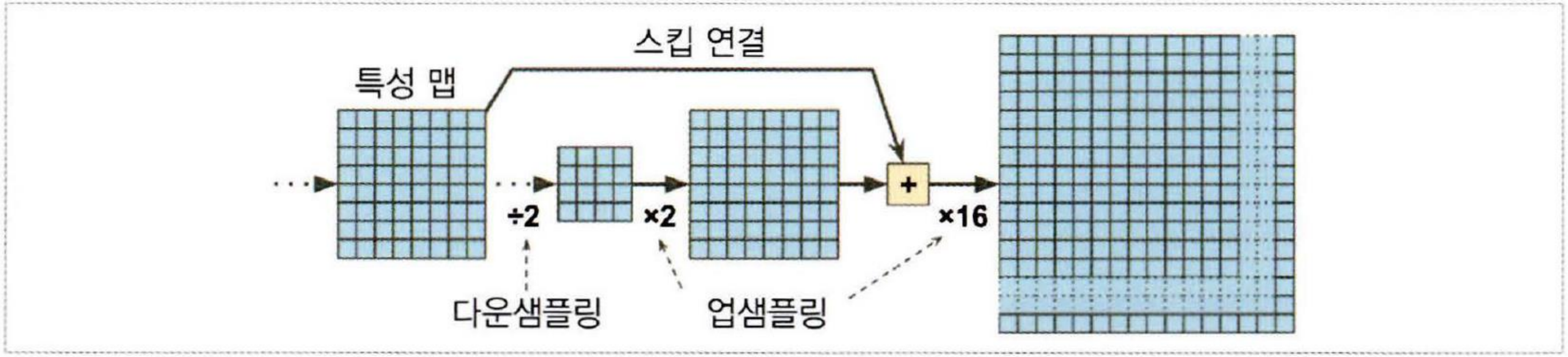


그림 14-28 아래쪽 층에서 공간 정보를 복원하는 스킵 연결 층

Semantic Segmentation

2. Fully Convolutional Networks for Semantic Segmentation *(Long and Shelhamer et al., CVPR 2015)*

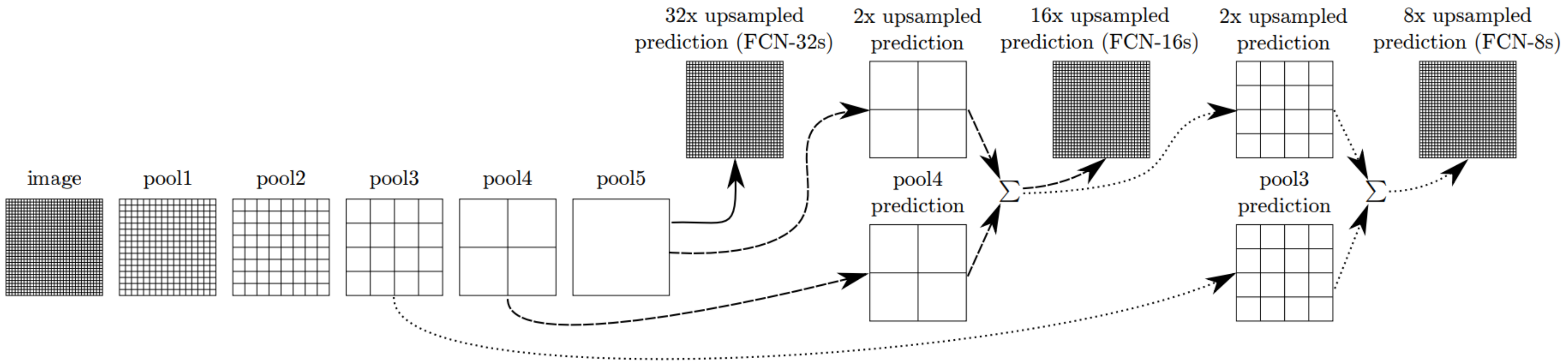


Figure 3. Our DAG nets learn to combine coarse, high layer information with fine, low layer information. Layers are shown as grids that reveal relative spatial coarseness. Only pooling and prediction layers are shown; intermediate convolution layers (including our converted fully connected layers) are omitted. Solid line (FCN-32s): Our single-stream net, described in Section 4.1, upsamples stride 32 predictions back to pixels in a single step. Dashed line (FCN-16s): Combining predictions from both the final layer and the `pool4` layer, at stride 16, lets our net predict finer details, while retaining high-level semantic information. Dotted line (FCN-8s): Additional predictions from `pool3`, at stride 8, provide further precision.

Semantic Segmentation

2. Fully Convolutional Networks for Semantic Segmentation *(Long and Shelhamer et al., CVPR 2015)*

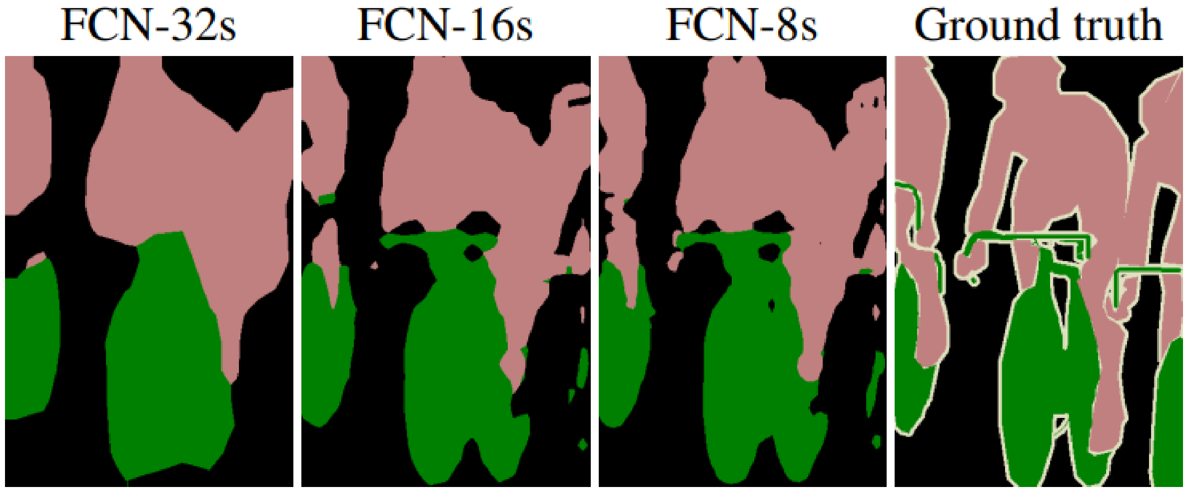


Figure 4. Refining fully convolutional nets by fusing information from layers with different strides improves segmentation detail. The first three images show the output from our 32, 16, and 8 pixel stride nets (see Figure 3).

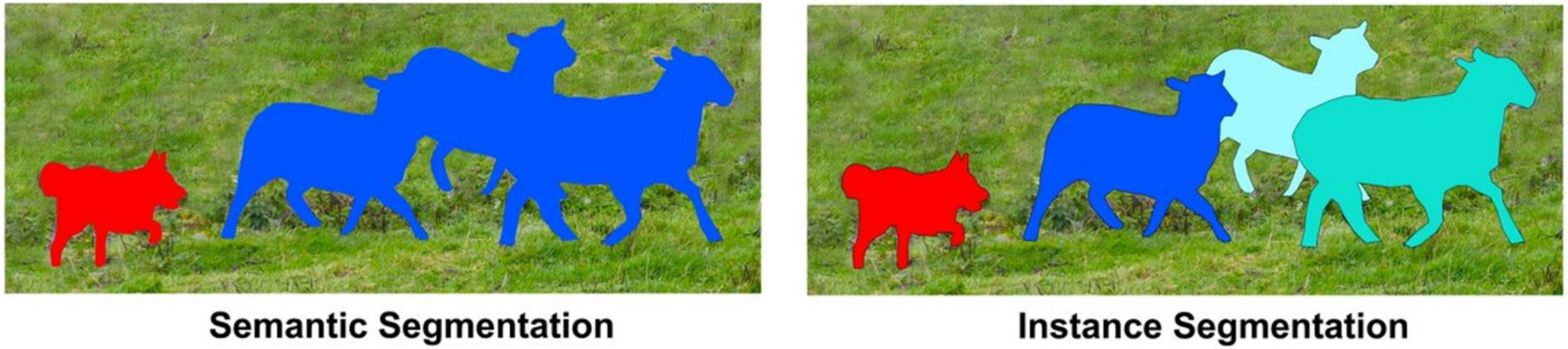
Table 2. Comparison of skip FCNs on a subset of PASCAL VOC2011 validation[7]. Learning is end-to-end, except for FCN-32s-fixed, where only the last layer is fine-tuned. Note that FCN-32s is FCN-VGG16, renamed to highlight stride.

	pixel acc.	mean acc.	mean IU	f.w. IU
FCN-32s-fixed	83.0	59.7	45.4	72.0
FCN-32s	89.1	73.3	59.4	81.4
FCN-16s	90.0	75.7	62.4	83.0
FCN-8s	**90.3**	**75.9**	**62.7**	**83.2**

Semantic Segmentation

3. 인스턴스 분할 (Instance segmentation)

- 시맨틱 분할과 비슷하지만 동일한 클래스 물체를 하나의 덩어리로 인식하는게 아니라 각 물체를 구분하여 표시함



- ~~Optional~~ Must-read thing: [*Mask R-CNN (He et al., ICCV 2017)*]

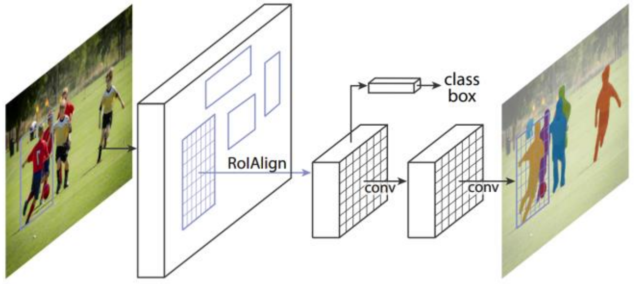


Figure 1. The **Mask R-CNN** framework for instance segmentation.

Thank you

🧑‍🎓: https://jeiyoon.github.io/

☁: https://www.instagram.com/cloudwantsasnack/

TAVE
TECHNOLOGY WAVE