http://bit.ly/dlcv2018

#DLUPC

Day 2 Lecture 3

Semantic Segmentation

Míriam Bellver
miriam.bellver@bsc.edu

PhD Candidate
Barcelona Supercomputing Center

Acknowledgements

Amaia Salvador
amaia.salvador@upc.edu

PhD Candidate
Universitat Politècnica de Catalunya

[DLCV 2016]

Verónica Vilaplana
veronica.vilaplana@upc.edu

Associate Professor
Universitat Politècnica de Catalunya

[DLCV 2017]

Segmentation

Segmentation

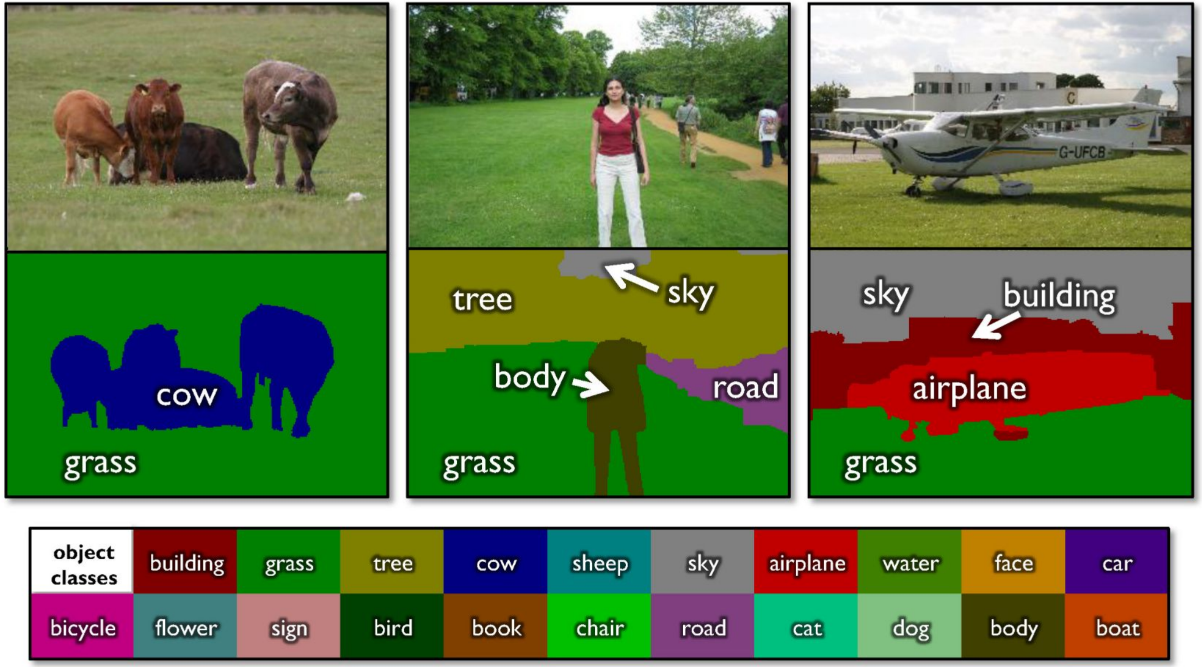
Define the accurate boundaries of all objects in an image

Semantic Segmentation

Label every pixel!

Don't differentiate instances (cows)

Classic computer vision problem



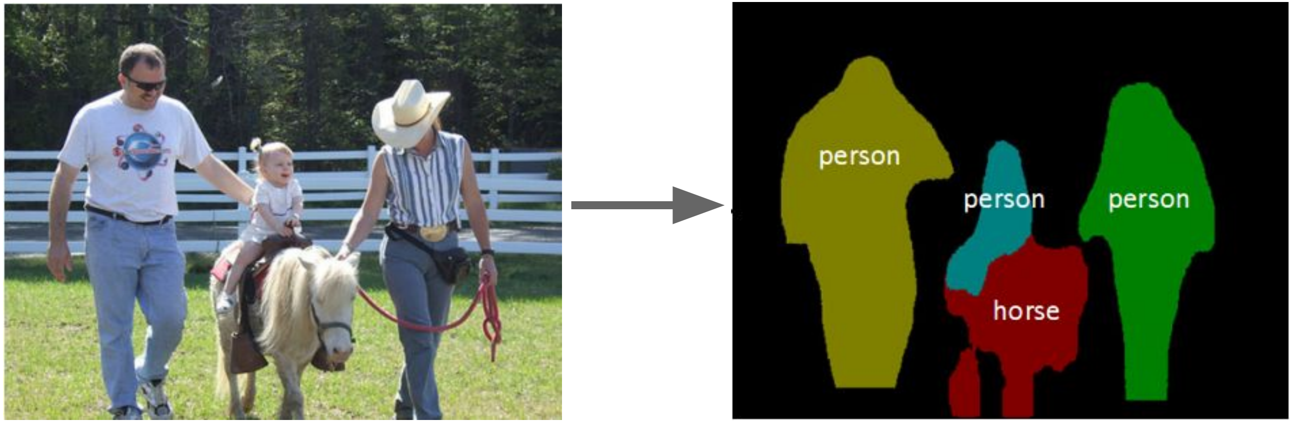
Slide Credit: [CS231n]

Instance Segmentation

Detect instances, give category, label pixels

“simultaneous detection and segmentation” (SDS)

Label are class-aware and instance-aware



Slide Credit: CS231n

Outline

Segmentation Datasets

Semantic Segmentation Methods

- Deconvolution (or transposed convolution)
- Dilated Convolution
- Skip Connections

Segmentation: Datasets

Pascal Visual Object Classes

- 20 categories
- +10,000 images
- Semantic segmentation GT
- Instance segmentation GT

Pascal Context

- Real indoor & outdoor scenes
- 540 categories
- +10,000 images
- Dense annotations
- Semantic segmentation GT
- Objects + stuff

Segmentation: Datasets

ADE20K

- Real general scenes
- +150 categories
- +22,000 images
- Semantic segmentation GT
- Instance + parts segmentation GT
- Objects and stuff

COCO Common Objects in Context

- Real indoor & outdoor scenes
- 80 categories
- +300,000 images
- 2M instances
- Partial annotations
- Semantic segmentation GT
- Instance segmentation GT
- Objects, but no stuff

Segmentation: Datasets

[CityScapes]

- Real driving scenes
- 30 categories
- +25,000 images
- 20,000 partial annotations
- 5,000 dense annotations
- Semantic segmentation GT
- Instance segmentation GT
- Depth, GPS and other metadata
- Objects and stuff

[Mapillary Vistas Dataset]

- Real driving scenes
- 100 categories
- 25,000 images
- Semantic segmentation GT
- Instance + parts segmentation GT
- Objects and stuff

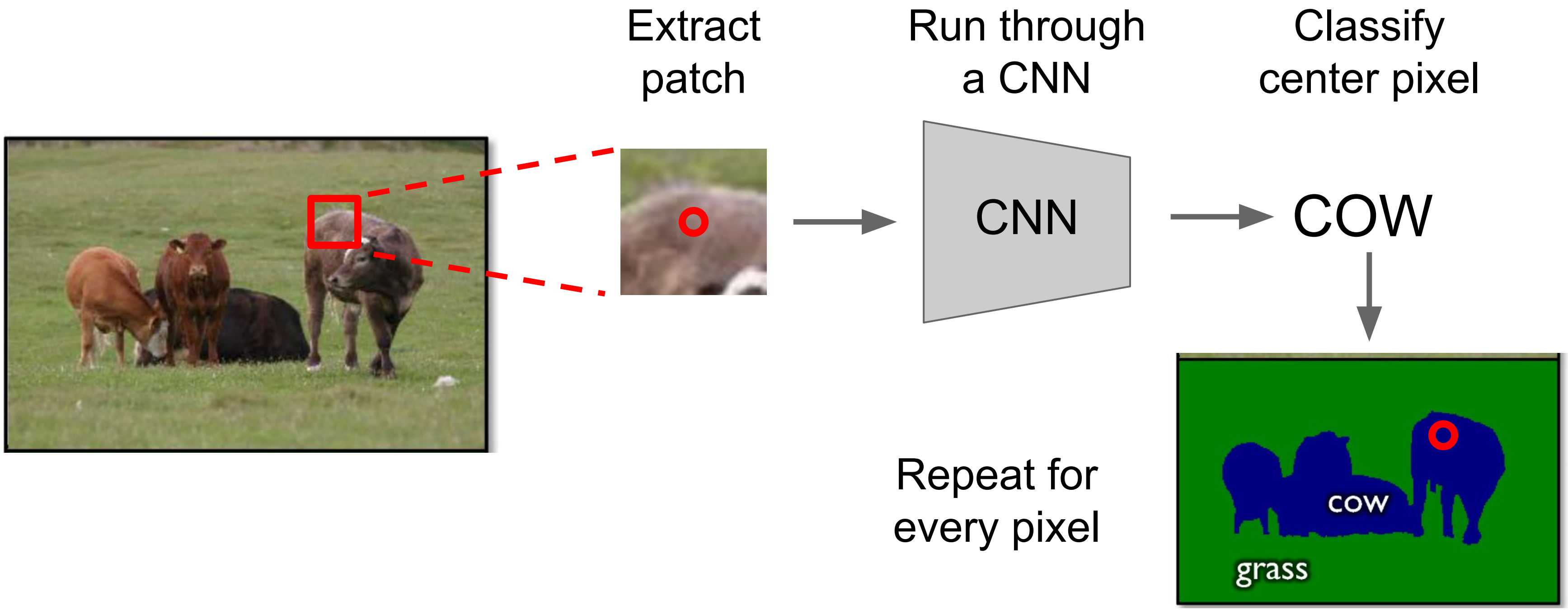
Outline

Segmentation Datasets

Semantic Segmentation Methods

- Deconvolution (or transposed convolution)
- Dilated Convolution
- Skip Connections

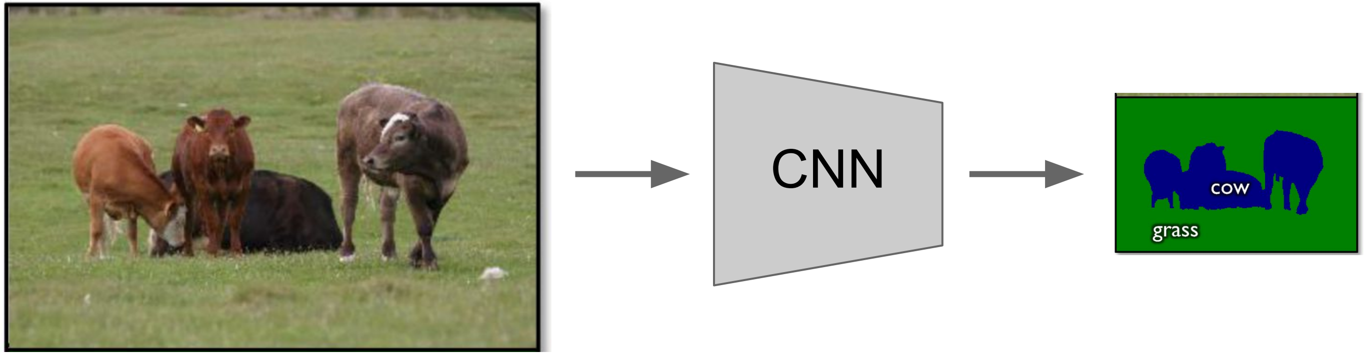
From Classification to Segmentation



Slide Credit: [CS231n]

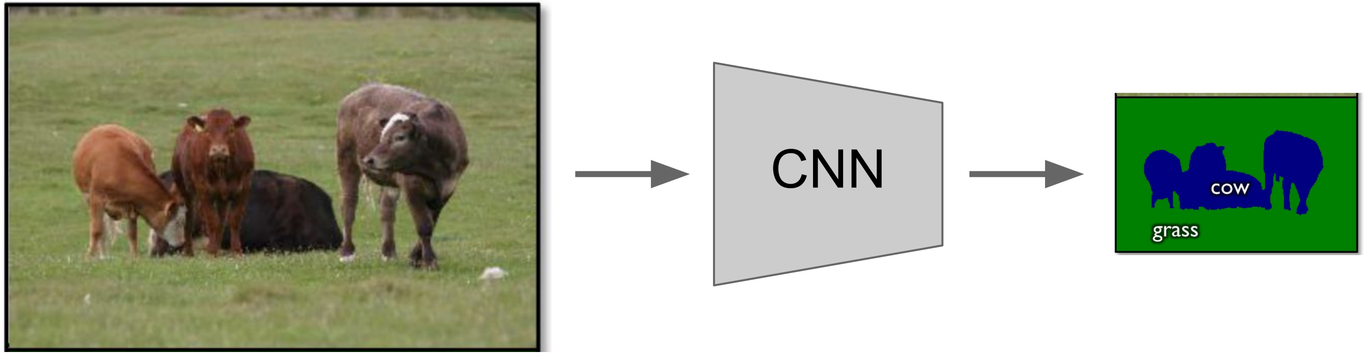
From Classification to Segmentation

Run “fully convolutional” network
to get all pixels at once



Slide Credit: [CS231n]

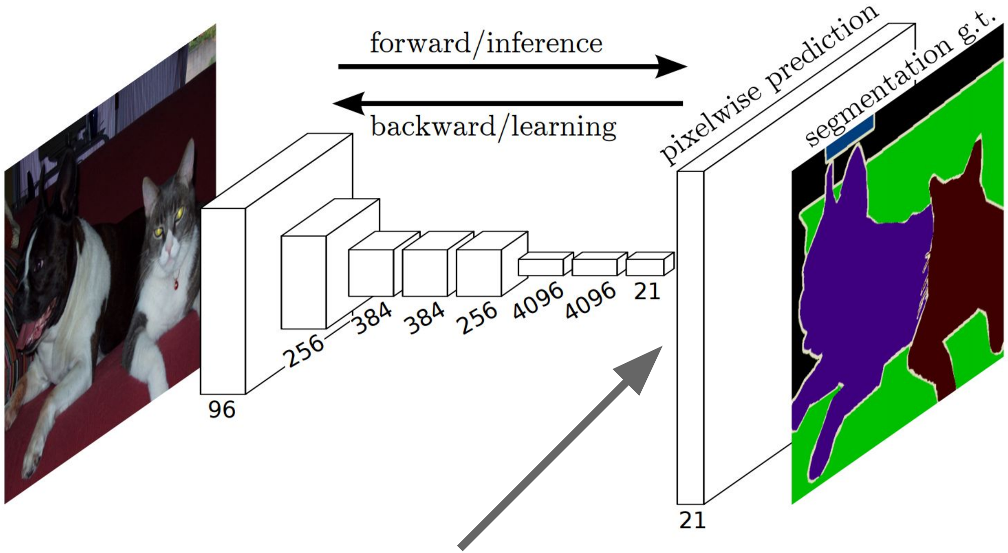
Semantic Segmentation



Problem 1:

Smaller output due to pooling

Learnable upsampling

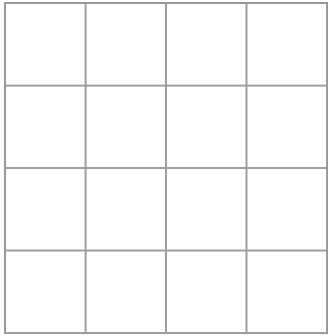


Long et al. Fully Convolutional Networks for Semantic Segmentation. CVPR 2015

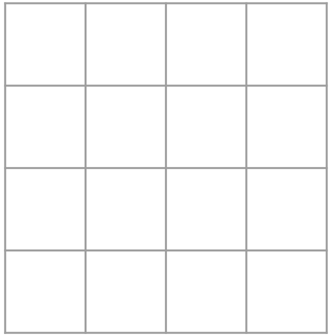
Slide Credit: CS231n

Reminder: Convolutional Layer

Typical 3 x 3 convolution, stride 1 pad 1



Input: 4 x 4

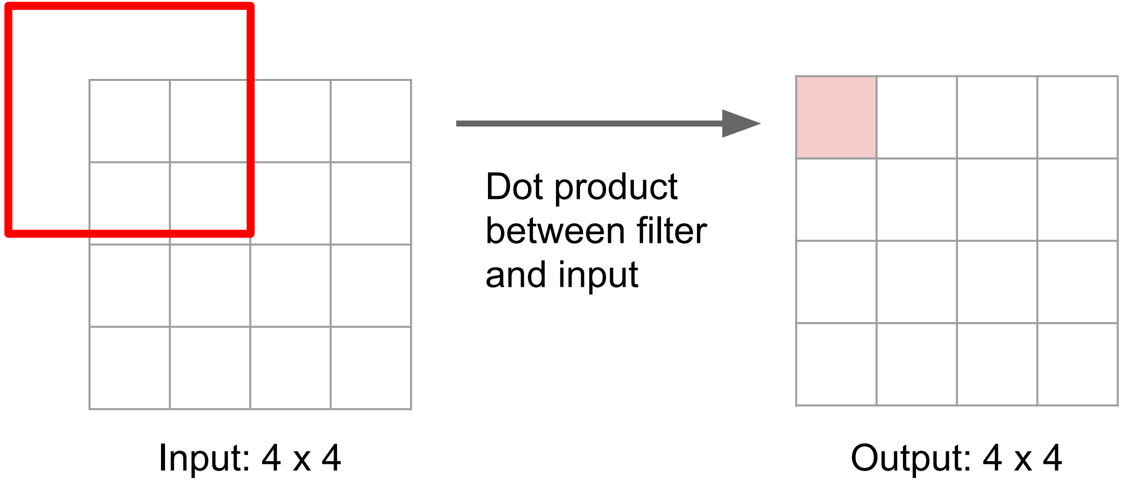


Output: 4 x 4

Slide Credit: [CS231n]

Reminder: Convolutional Layer

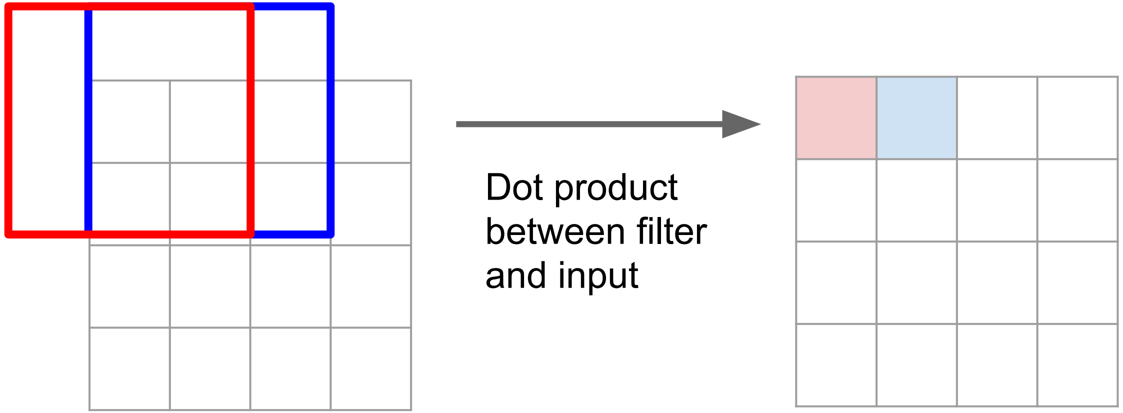
Typical 3 x 3 convolution, stride 1 pad 1



Slide Credit: [CS231n]

Reminder: Convolutional Layer

Typical 3 x 3 convolution, stride 1 pad 1



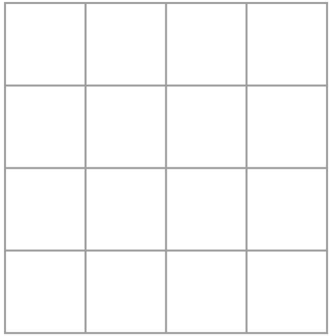
Input: 4 x 4

Output: 4 x 4

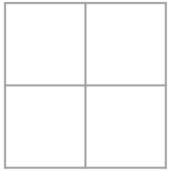
Slide Credit: CS231n

Reminder: Convolutional Layer

Typical 3 x 3 convolution, **stride 2** pad 1



Input: 4 x 4

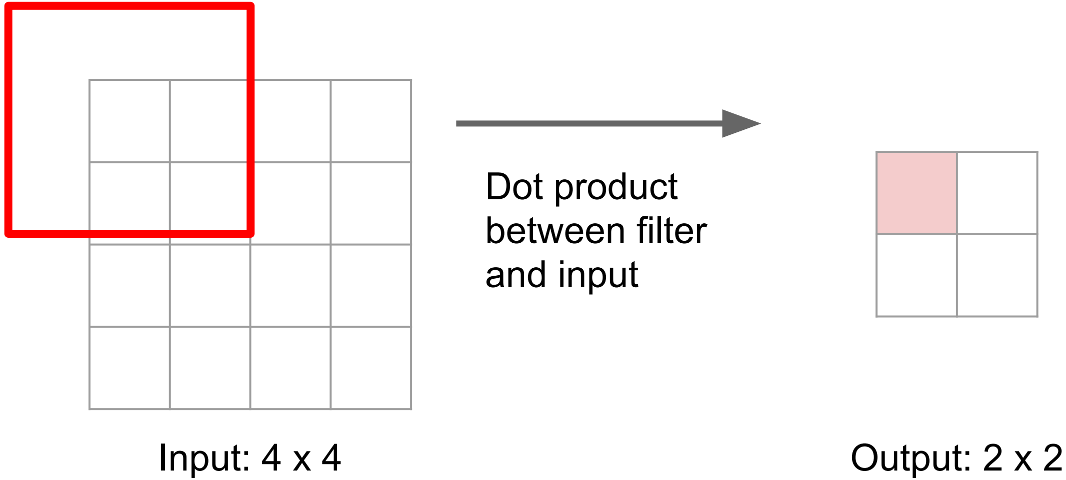


Output: 2 x 2

Slide Credit: CS231n

Reminder: Convolutional Layer

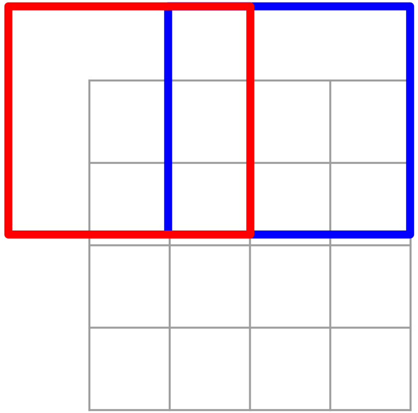
Typical 3 x 3 convolution, stride 2 pad 1



Slide Credit: [CS231n]

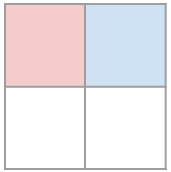
Reminder: Convolutional Layer

Typical 3 x 3 convolution, stride 2 pad 1



Input: 4 x 4

Dot product between filter and input

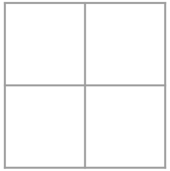


Output: 2 x 2

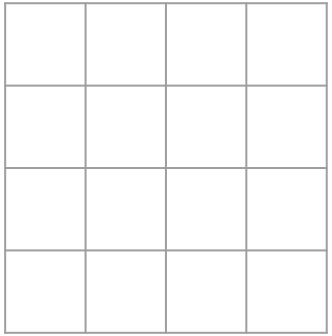
Slide Credit: CS231n

Learnable Upsample: Transposed Convolution

3 x 3 “deconvolution”, stride 2 pad 1



Input: 2 x 2

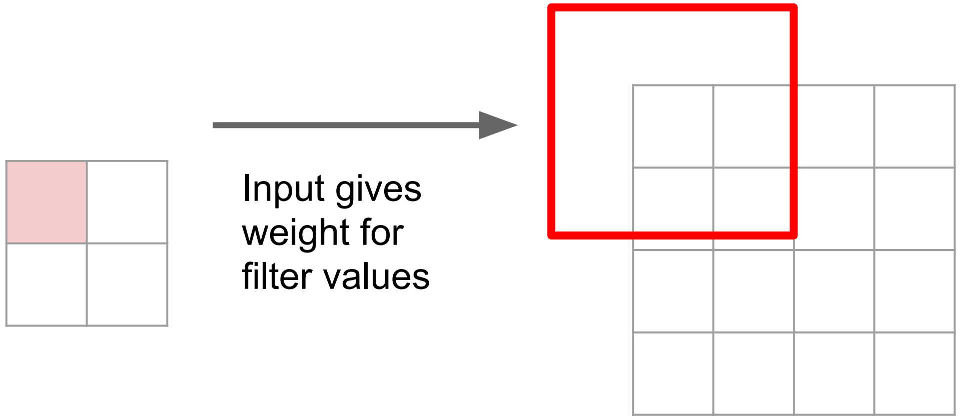


Output: 4 x 4

Slide Credit: [CS231n]

Learnable Upsample: Transposed Convolution

3 x 3 “deconvolution”, stride 2 pad 1

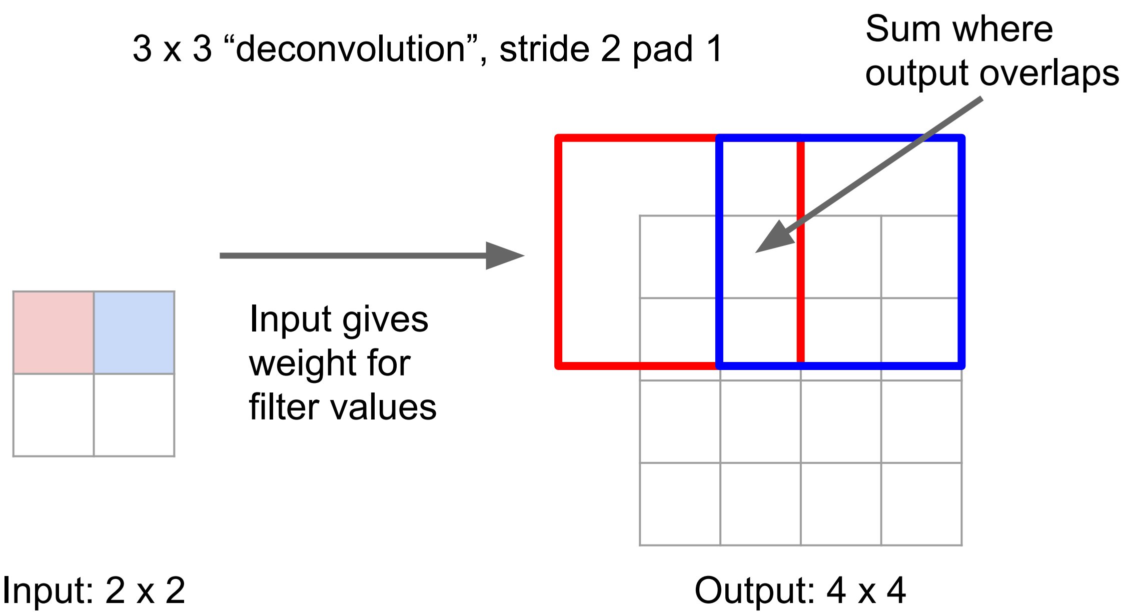


Input: 2 x 2

Output: 4 x 4

Slide Credit: CS231n

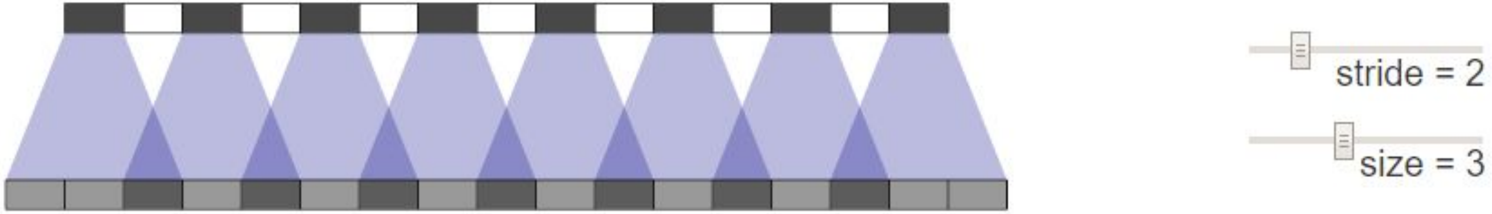
Learnable Upsample: Transposed Convolution

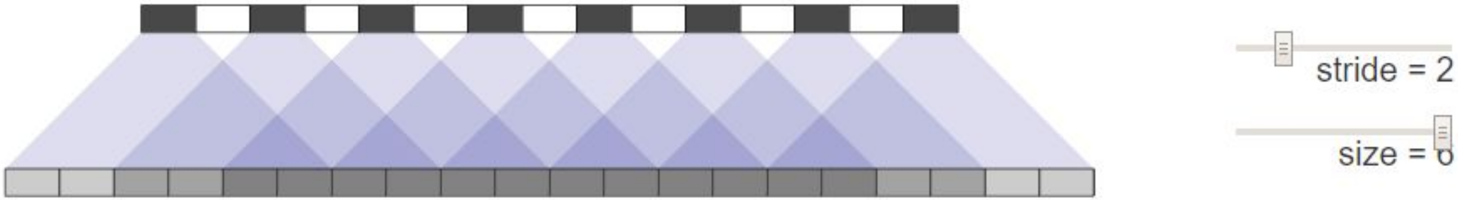


Slide Credit: [CS231n]

Learnable Upsample: Transposed Convolution

Warning: Checkerboard effect when kernel size is not divisible by the stride

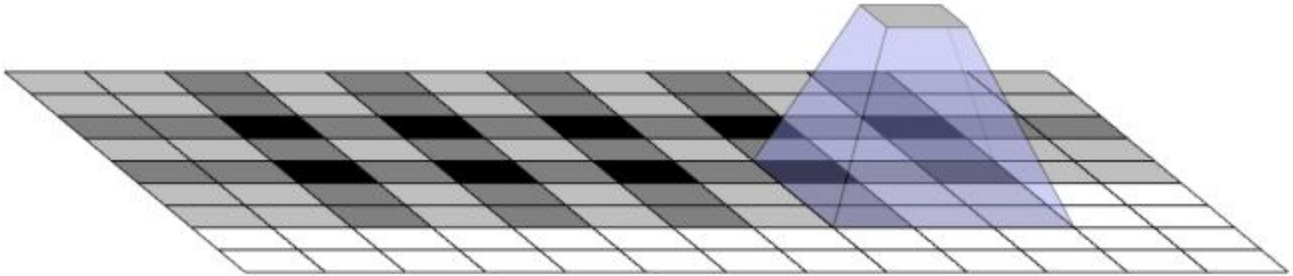




Source: distill.pub

Learnable Upsample: Transposed Convolution

Warning: Checkerboard effect when kernel size is not divisible by the stride

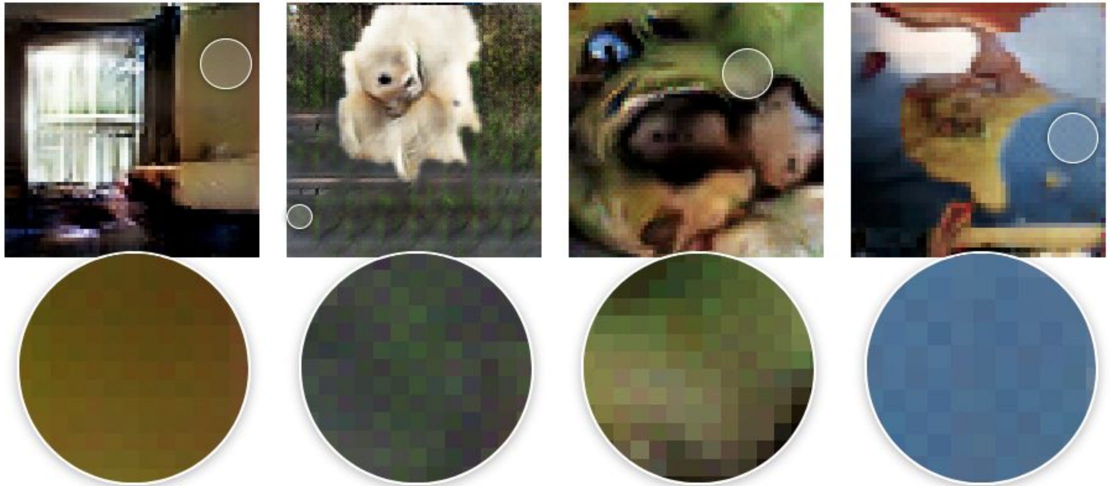


stride = 2, kernel_size = 3

Source: distill.pub

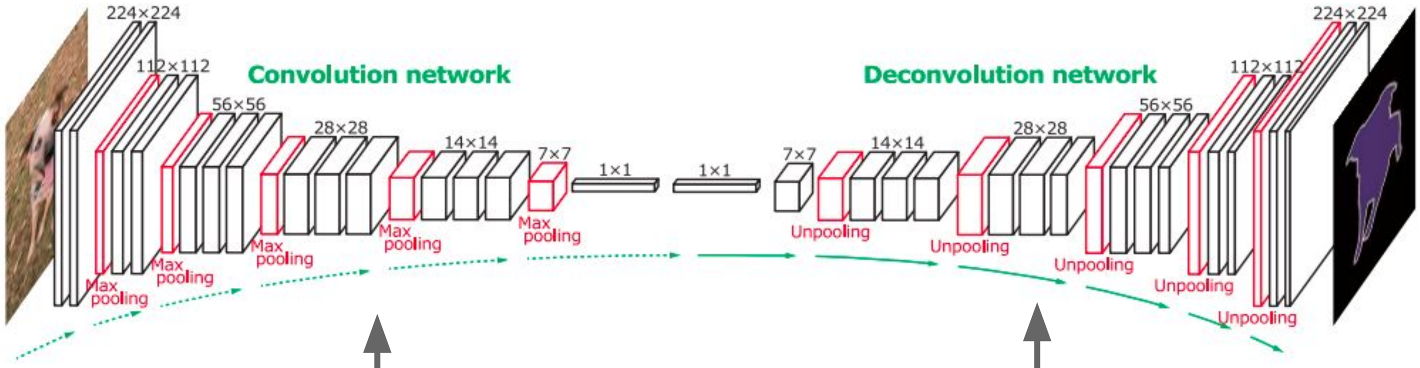
Learnable Upsample: Transposed Convolution

Warning: Checkerboard effect in images generated by neural networks



Radford, et al., 2015 [1] Salimans et al., 2016 [2] Donahue, et al., 2016 [3] Dumoulin, et al., 2016 [4]

Learnable Upsample: Transposed Convolution



“Regular” VGG

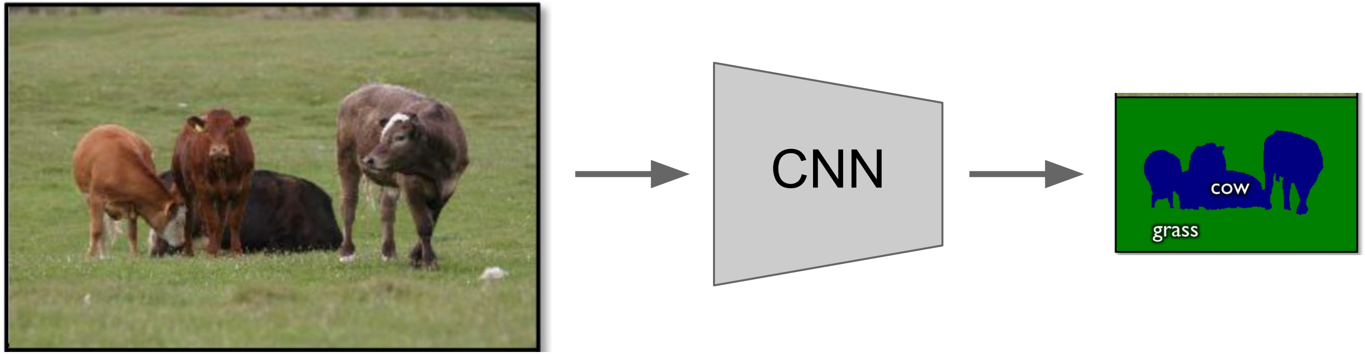
“Upside down” VGG

Noh et al. [Learning Deconvolution Network for Semantic Segmentation](). ICCV 2015

Slide Credit: [CS231n]()

Semantic Segmentation

Problem 2:

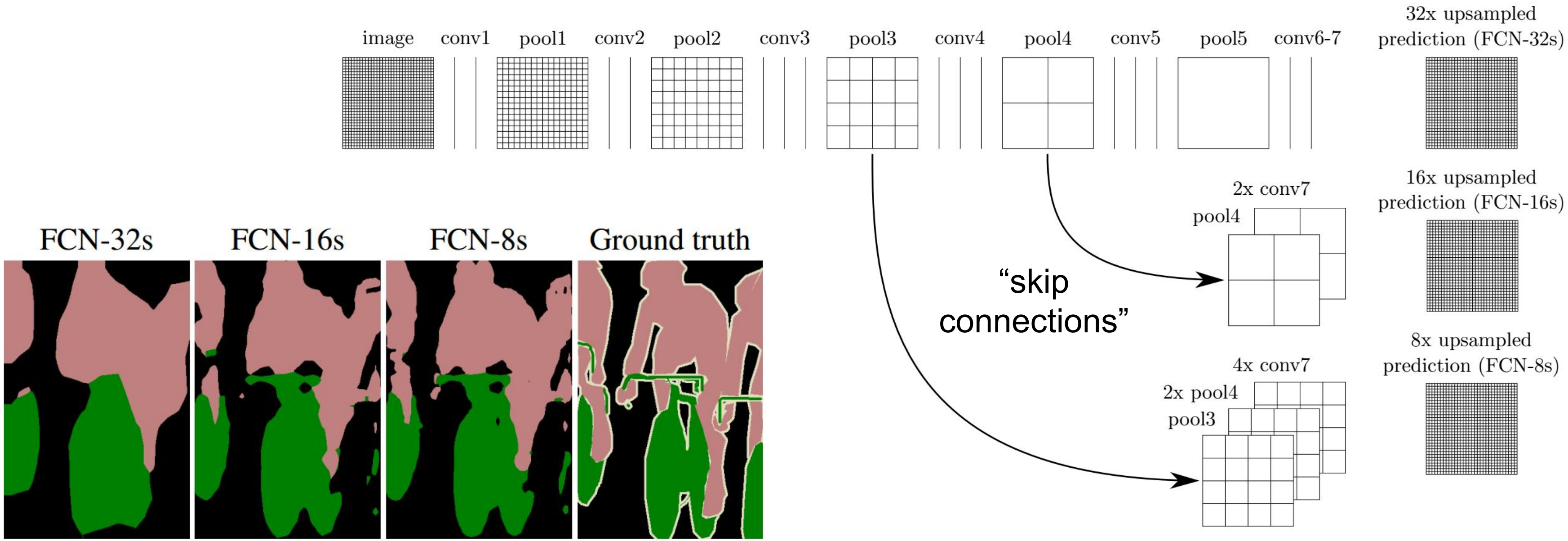


Coarse output

High-level features (e.g. conv5 layer) from a pretrained classification network are the input for the segmentation branch

Skip Connections

Recovering low level features from early layers



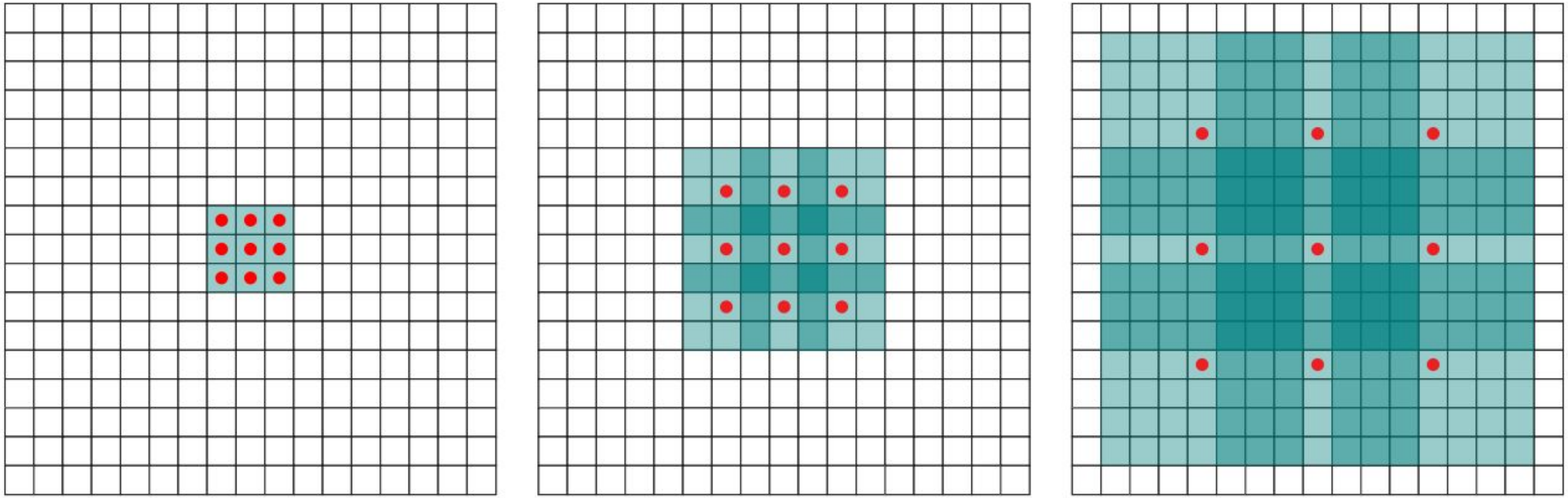
Skip connections = Better results

Long et al. [Fully Convolutional Networks for Semantic Segmentation.](#) CVPR 2015

Slide Credit: [CS231n](#)

Dilated Convolutions

Structural change in convolutional layers for dense prediction problems (e.g. image segmentation)

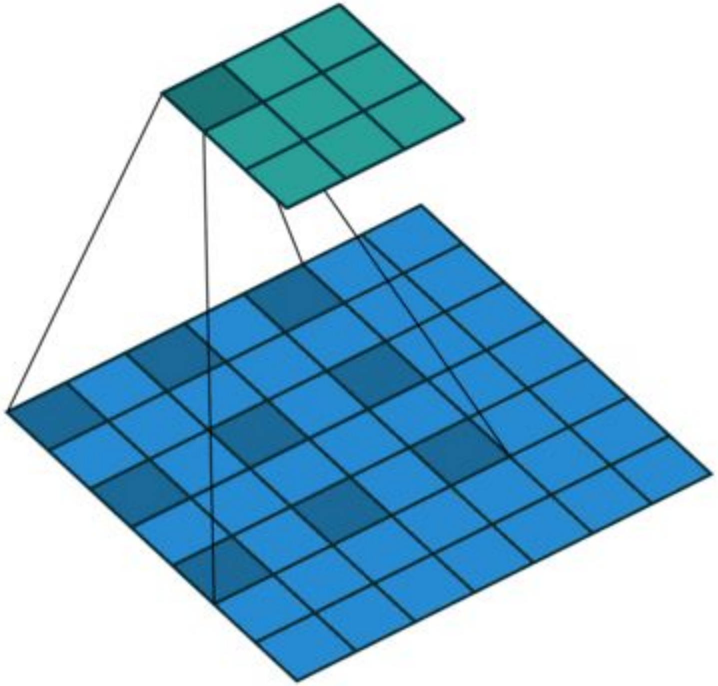


(a) (b) (c)

- The receptive field grows exponentially as you add more layers → more context information in deeper layers wrt regular convolutions
- Number of parameters increases linearly as you add more layers

Yu & Koltun. Multi-Scale Context Aggregation by Dilated Convolutions. ICLR 2016

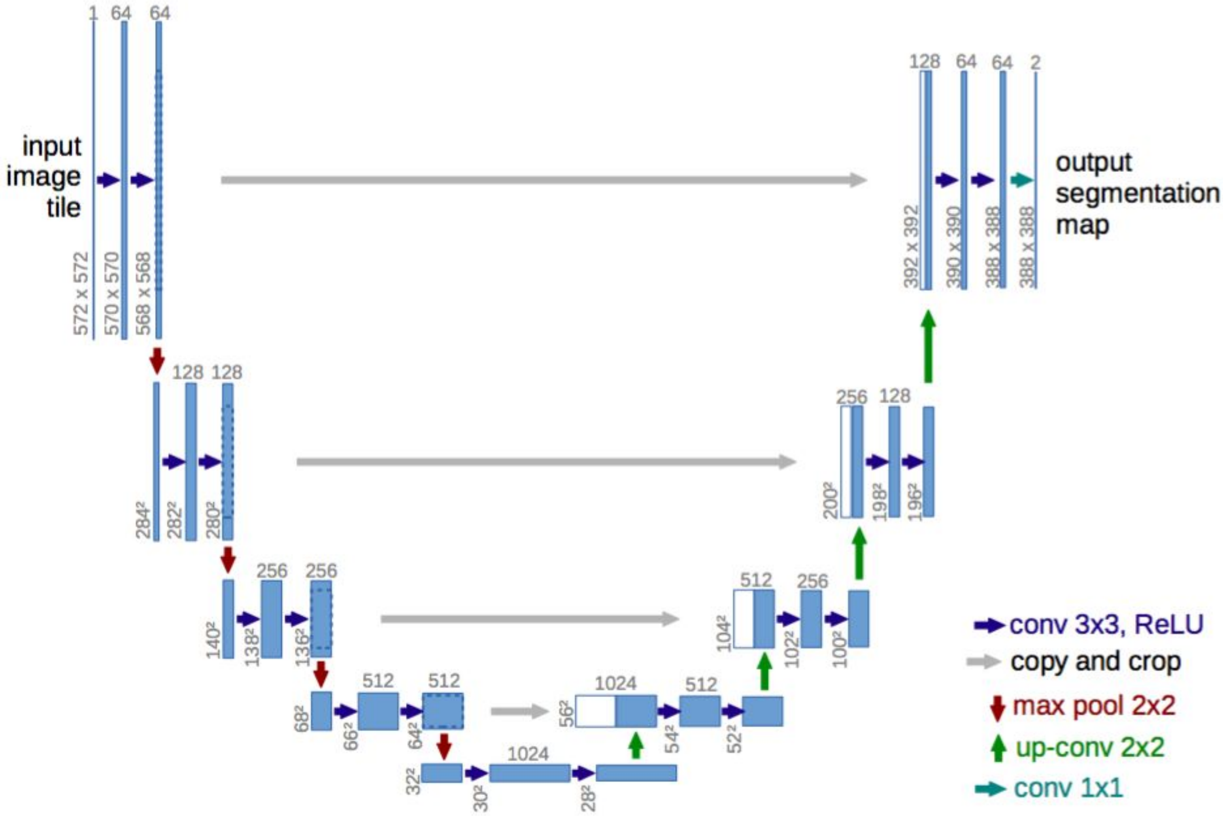
Dilated Convolutions



Source: https://github.com/vdumoulin/conv_arithmetic

State-of-the-art models

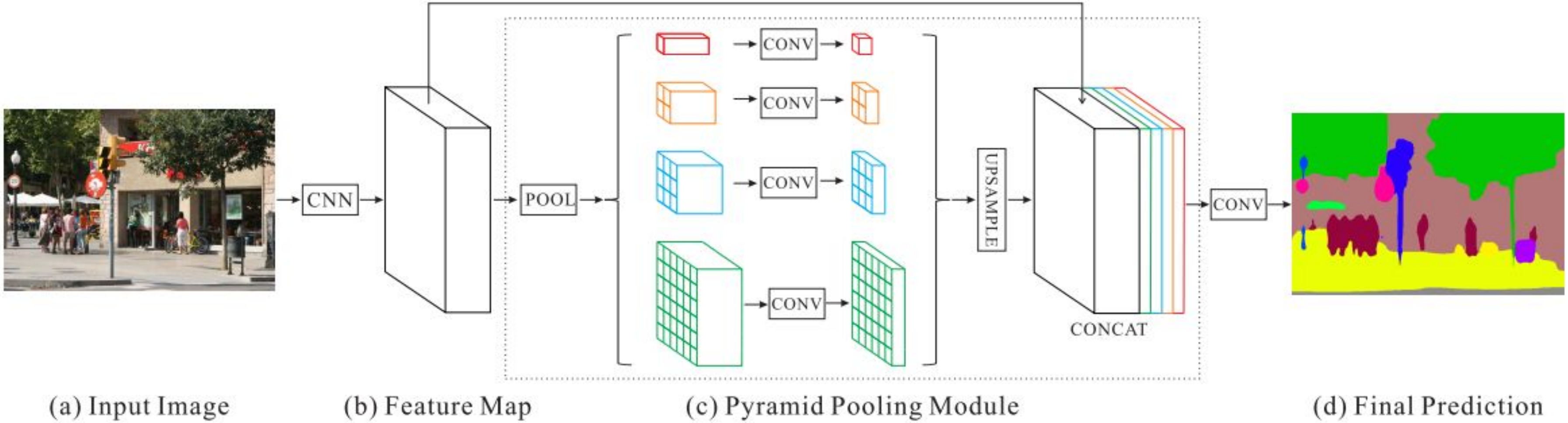
- U-Net
  - Deconvolutions
  - skip connections



Ronneberger et al. U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015

State-of-the-art models

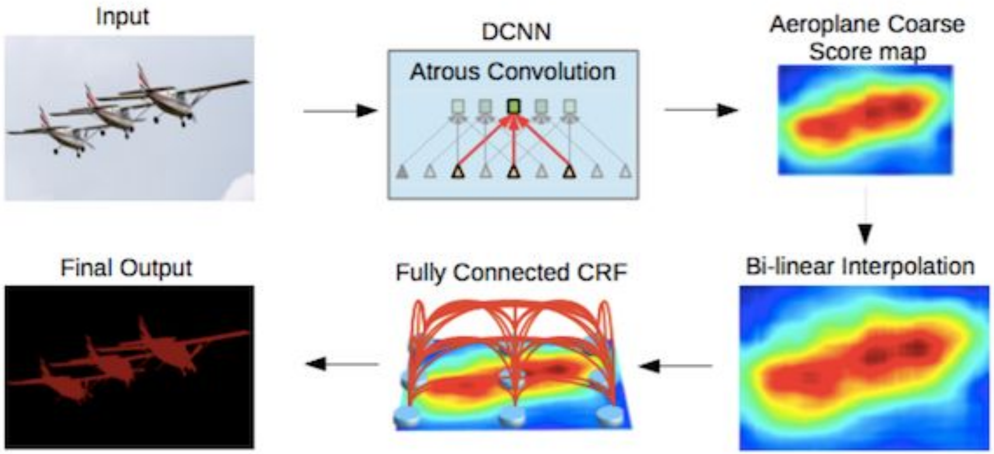
- PSPNet (dilated convolutions + pyramid pooling)



Zhao et al. [Pyramid Scene Parsing Network](). CVPR 2017

State-of-the-art models

- DeepLab v2 (dilated convolutions + CRF)



- DeepLab v3 (added pyramid pooling. Removed CRF)

Chen et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. TPAMI 2017
Chen et al. Rethinking Atrous Convolution for Semantic Image Segmentation. TPAMI 2017

Summary

Segmentation Datasets

Semantic Segmentation Methods

- Deconvolution (or transposed convolution)
- Dilated Convolution
- Skip Connections

Questions?