Last time: Repurposing pre-trained networks

- Fully supervised training:
  - Train on large dataset, test on the same task

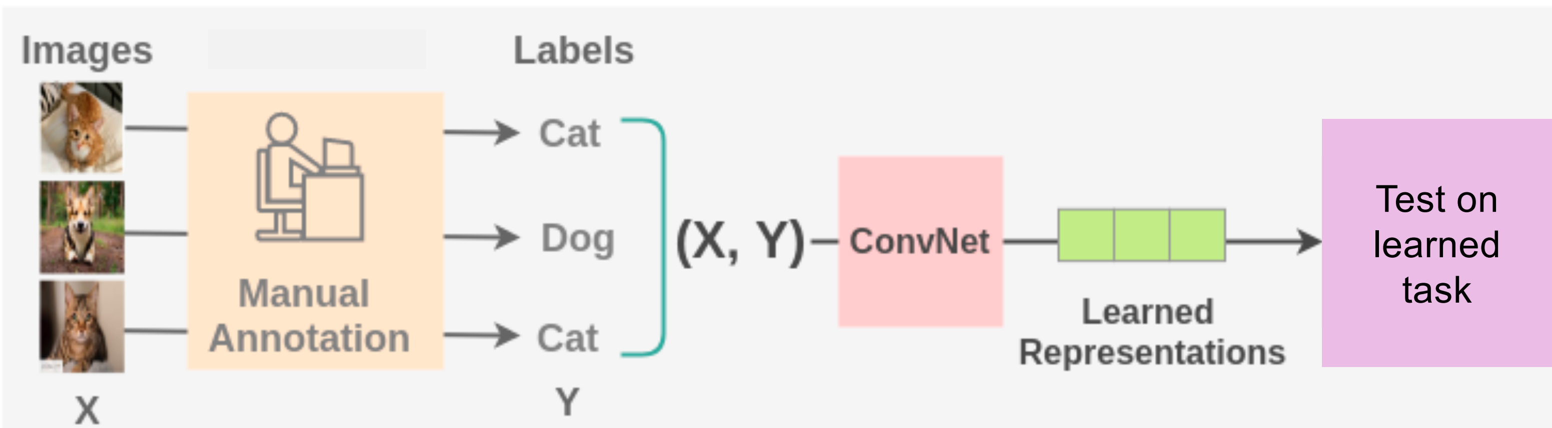


Figure source

Last time: Repurposing pre-trained networks

- Transfer learning:
  - *Pre-train* on large dataset
  - *Fine-tune* for a different task on a smaller dataset

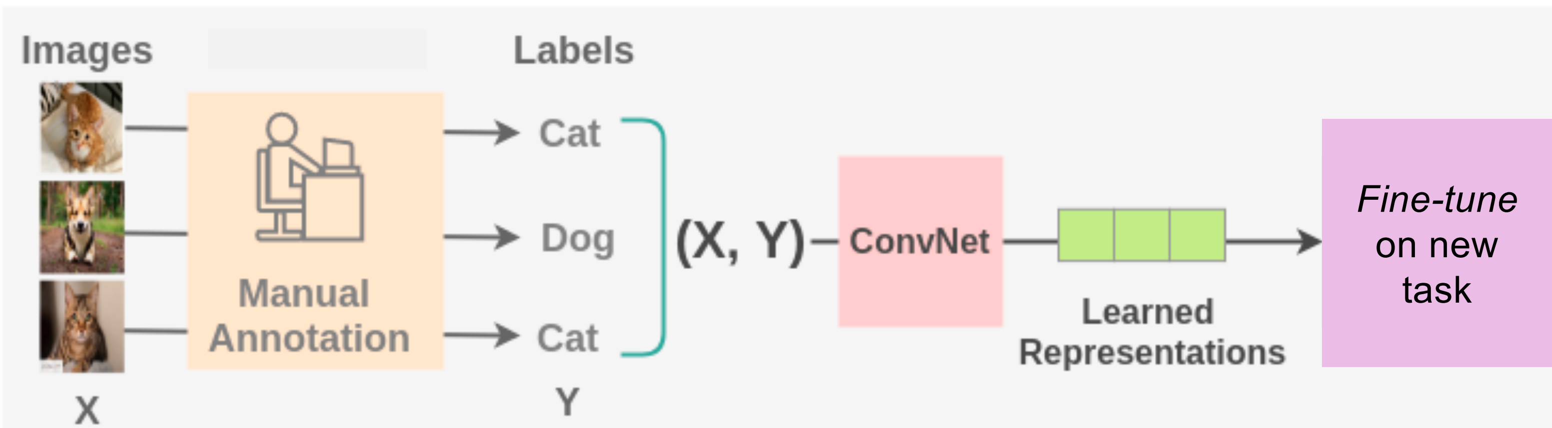


Figure source

Last time: Repurposing pre-trained networks

- Self-supervised pre-training:
  - *Pre-train* on large dataset *without labels* using a *pretext task* objective
  - *Fine-tune* for a different task on a smaller dataset (with labels)

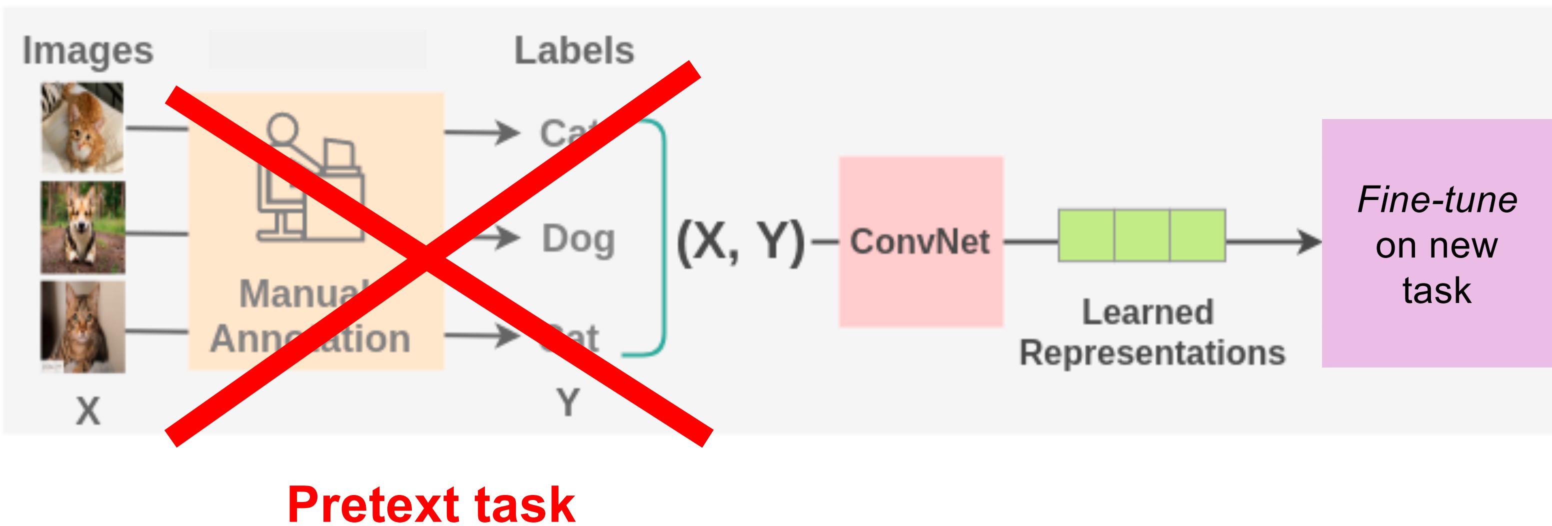
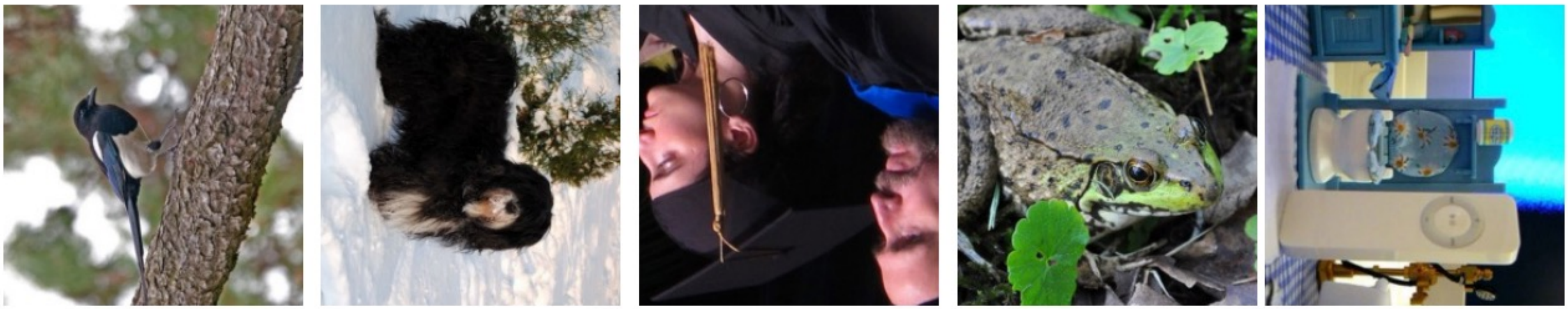


Figure source

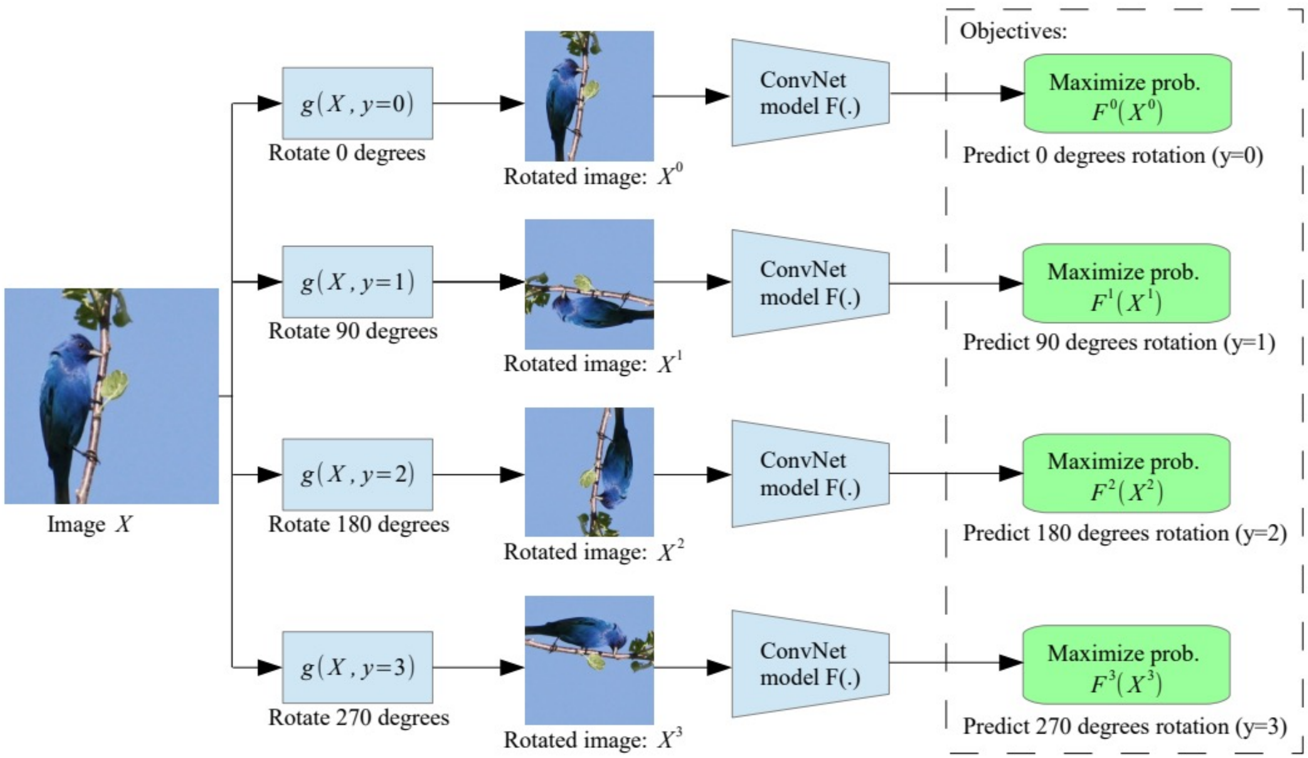
Pretext task: Rotation prediction (MP3)

- Can you recognize the image rotation (0, 90, 180, 270 degrees)?



S. Gidaris, P. Singh, and N. Komodakis. Unsupervised representation learning by predicting image rotations. ICLR 2018

Pretext task: Rotation prediction (MP3)



During training, feed in all four rotated versions of an image in the same mini-batch

S. Gidaris, P. Singh, and N. Komodakis. [Unsupervised representation learning by predicting image rotations.](#) ICLR 2018

CNNs for dense prediction

Image classification

Semantic segmentation

Semantic segmentation

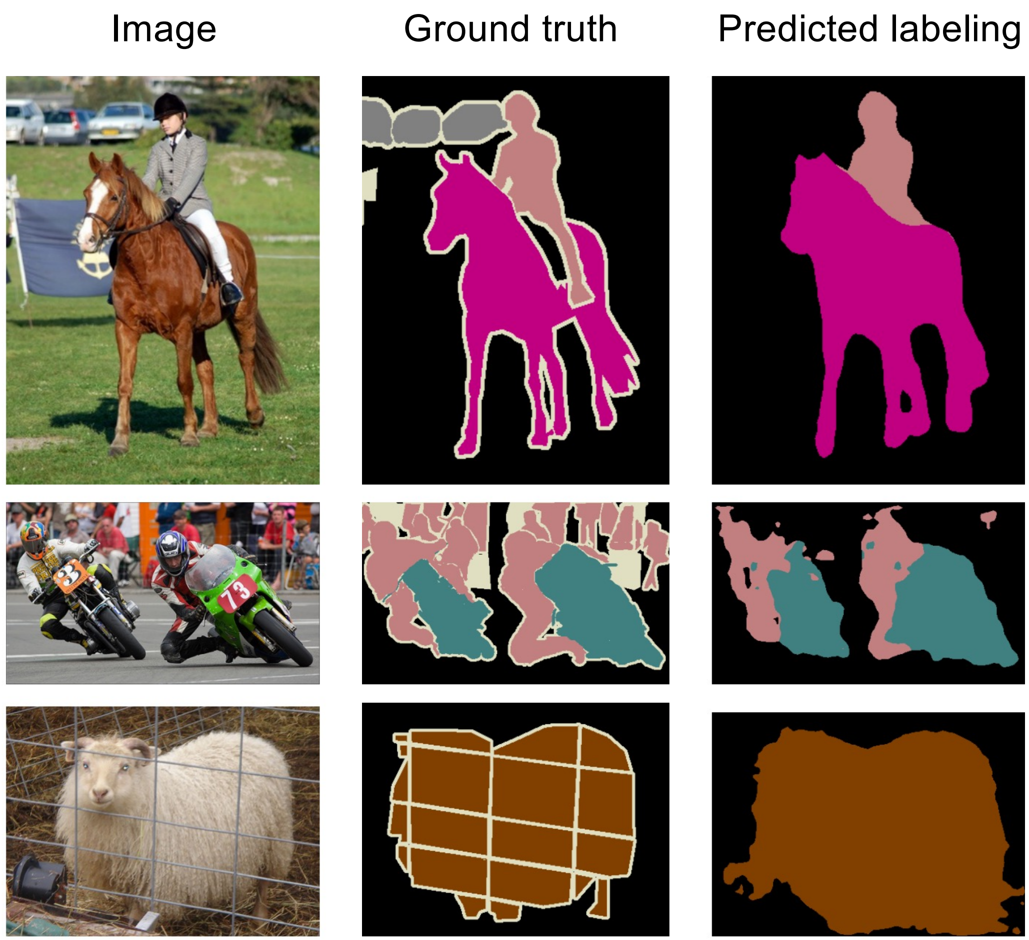
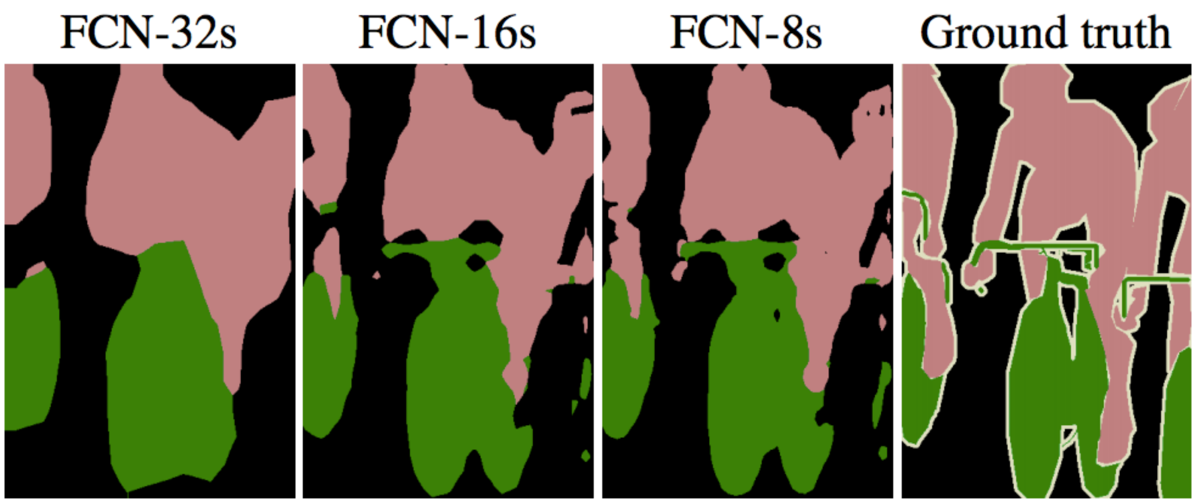


Figure source

Semantic segmentation

- Example evaluation:



Comparison on a subset of PASCAL 2011 validation data

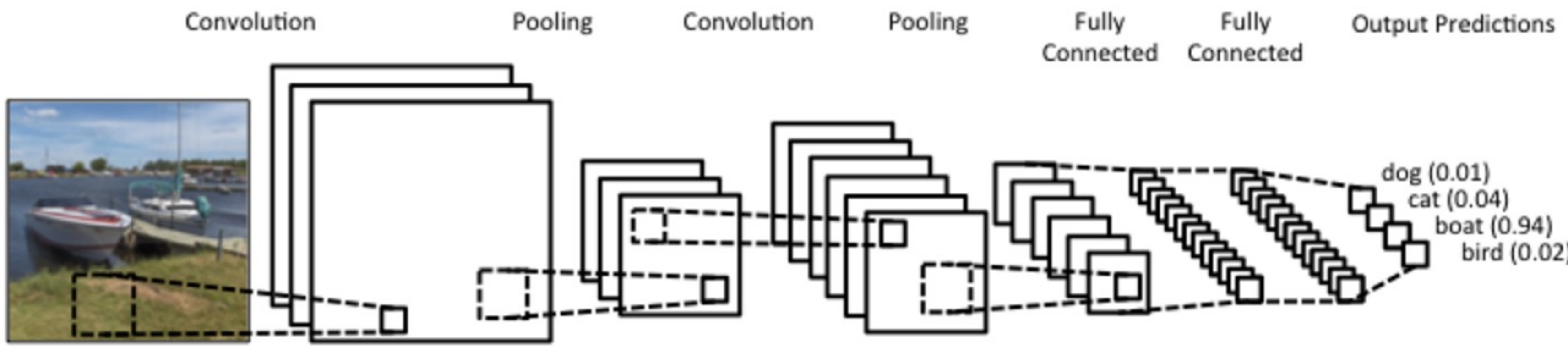
	pixel acc.	mean acc.	mean IU
FCN-32s-fixed	83.0	59.7	45.4
FCN-32s	89.1	73.3	59.4
FCN-16s	90.0	75.7	62.4
FCN-8s	**90.3**	**75.9**	**62.7**

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

Outline

- Dense prediction architectures
- Feature map upsampling
- U-Net
- A tour of dense prediction problems

Convolutional architectures so far

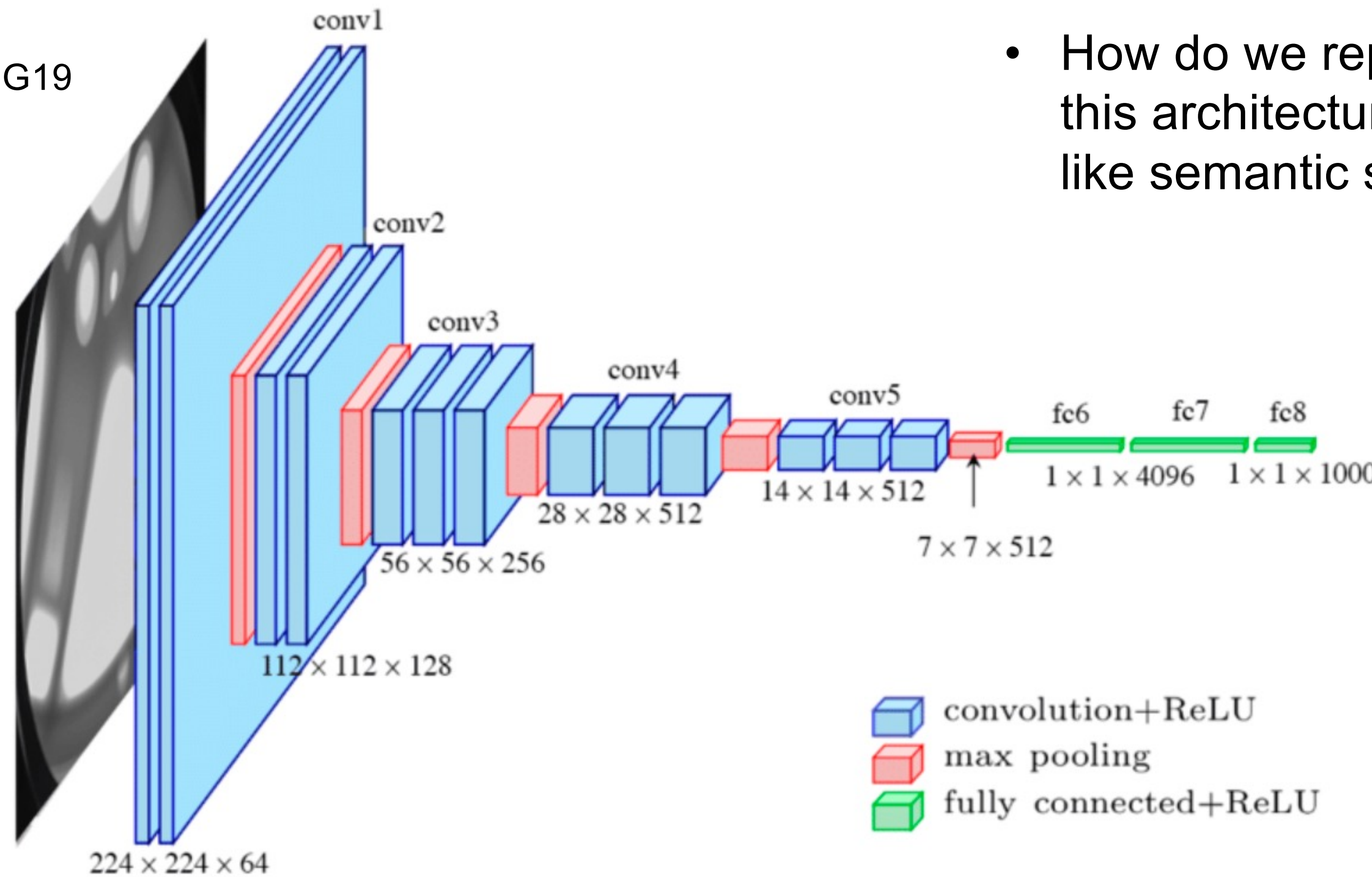


Stem/
backbone

Head

Convolutional architectures so far

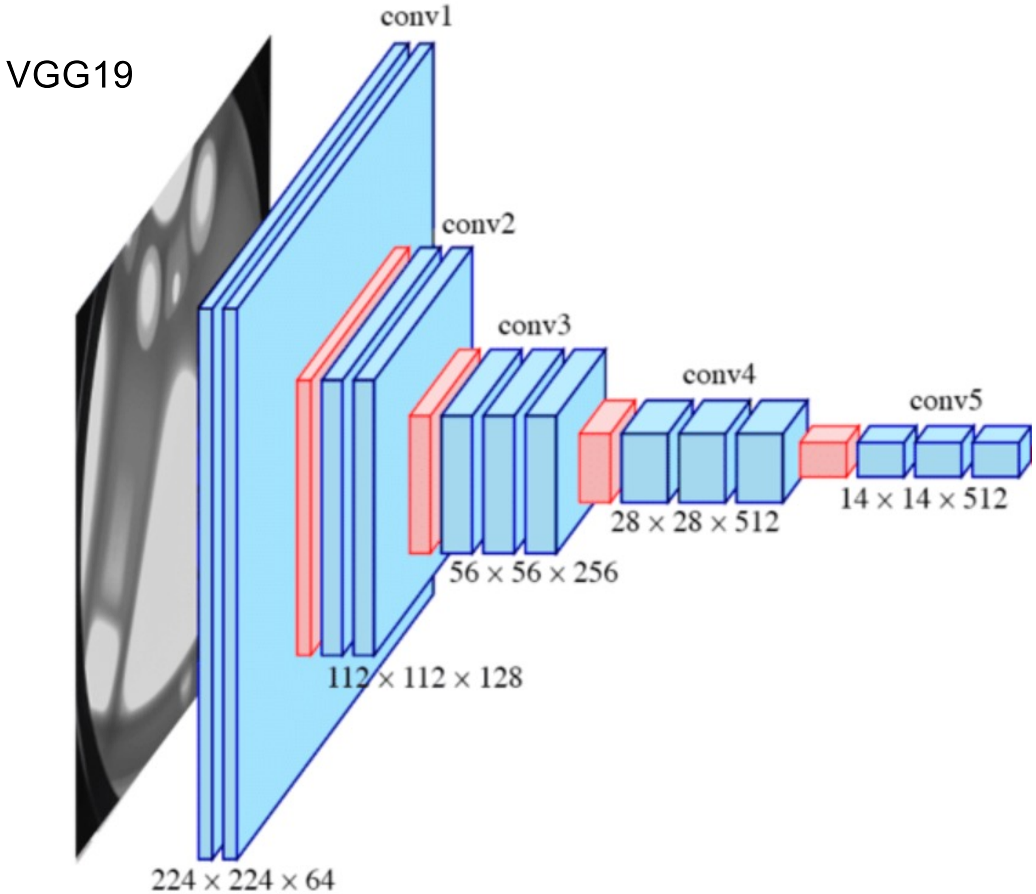
VGG19



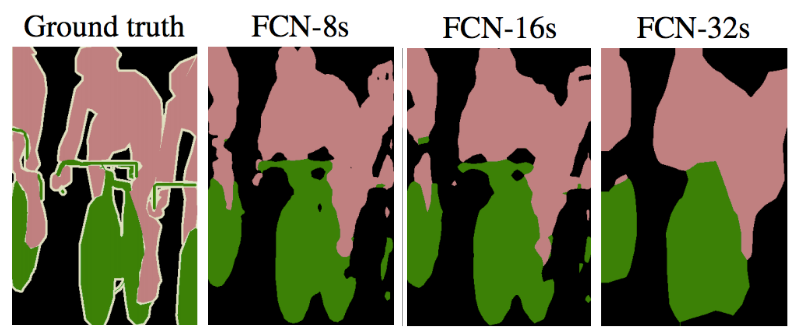
- How do we repurpose this architecture for other tasks, like semantic segmentation?

Image source

Convolutional architectures so far



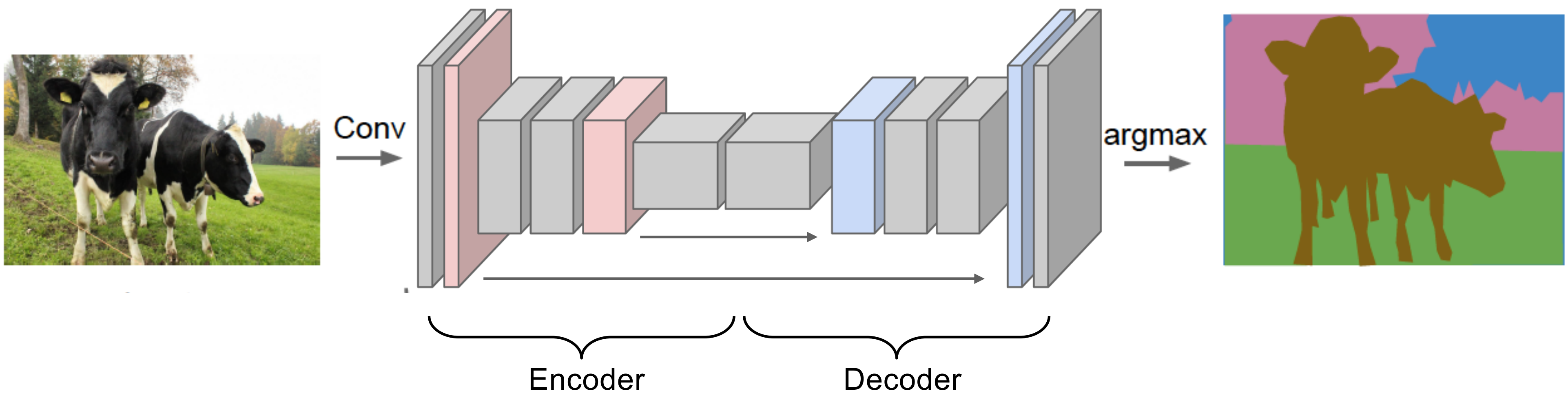
- How do we repurpose this architecture for other tasks, like semantic segmentation?
- Retain the stem, build some new layers on top of the output feature maps: e.g., 1x1 conv, elementwise softmax
- How to deal with the loss of spatial resolution?



[Image source](#)

Dense prediction architecture

- Idea 1: need to upsample the intermediate feature maps to produce full-resolution predictions
- Idea 2: need to fuse encoder and decoder feature maps to avoid loss of spatial information

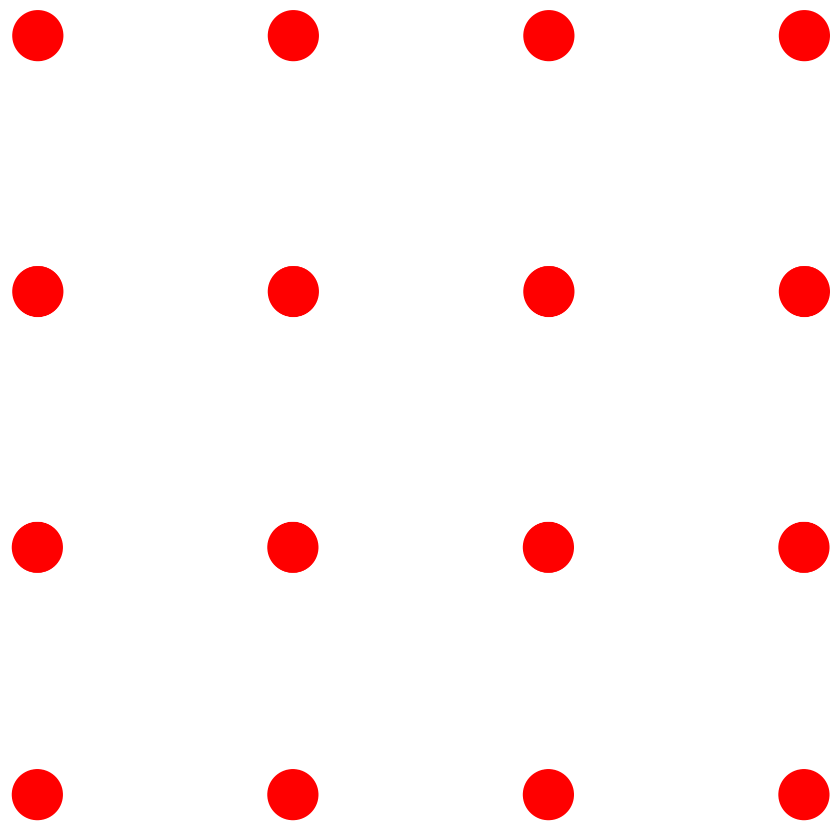


Source: [Stanford CS231n]

Outline

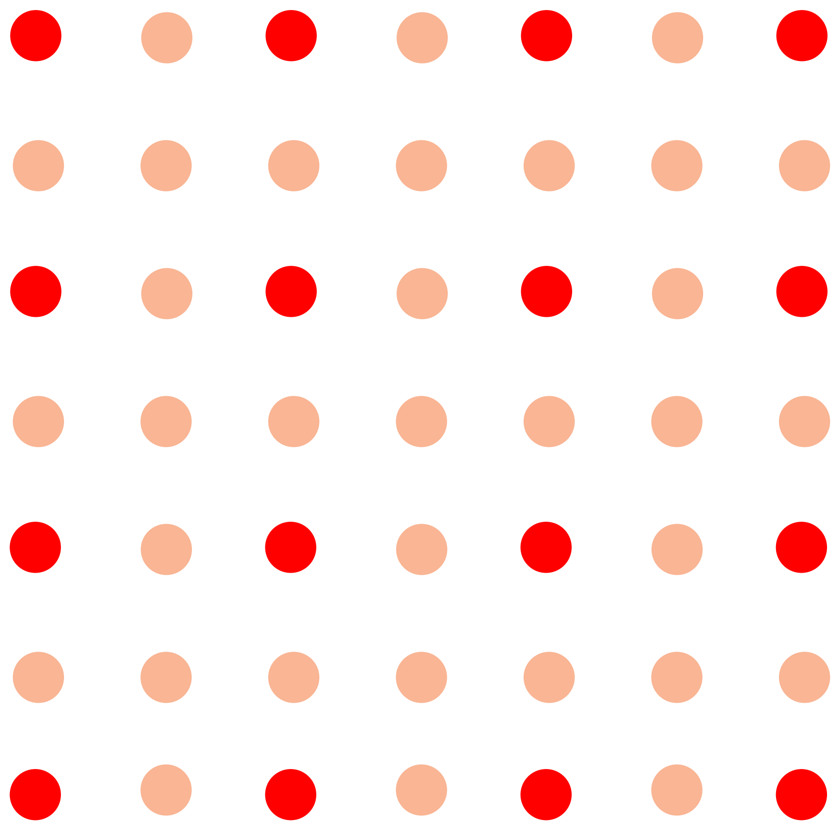
- Dense prediction architectures
- Feature map upsampling

How to upsample a feature map by a factor of 2?



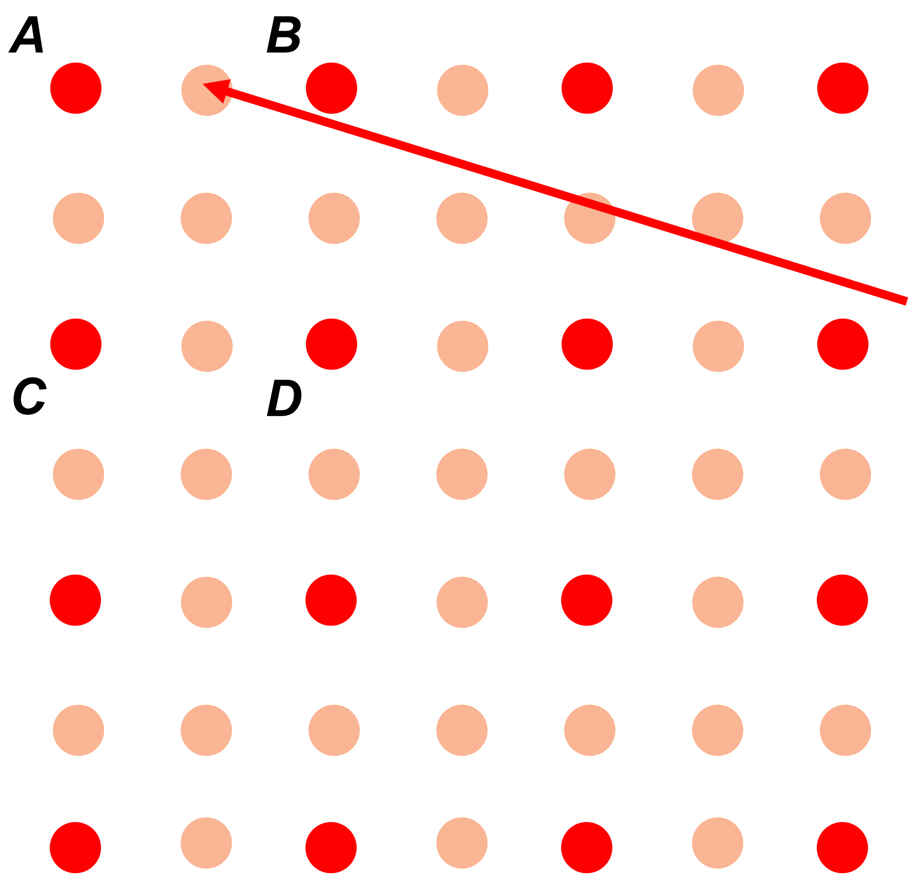
Step 1: increase the resolution of the feature grid

How to upsample a feature map by a factor of 2?



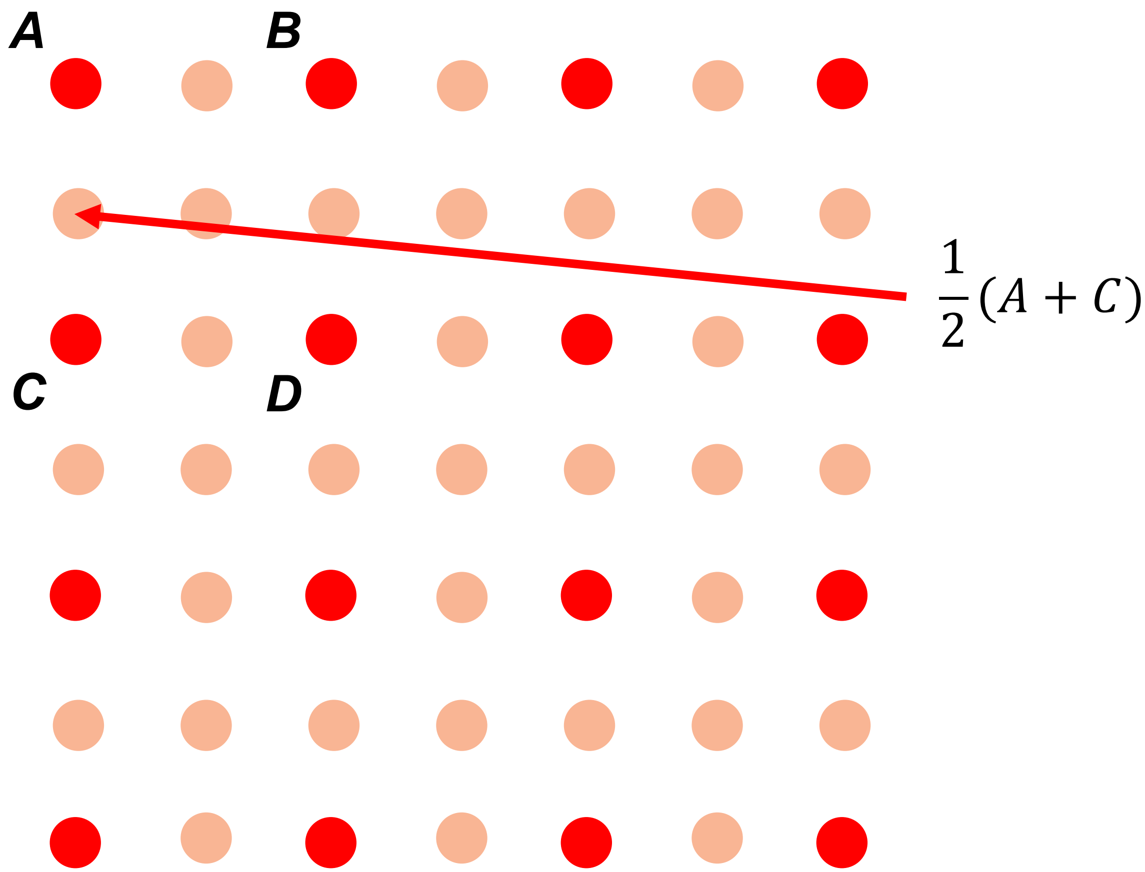
Step 2: *interpolate* to get the missing values

How to upsample a feature map by a factor of 2?

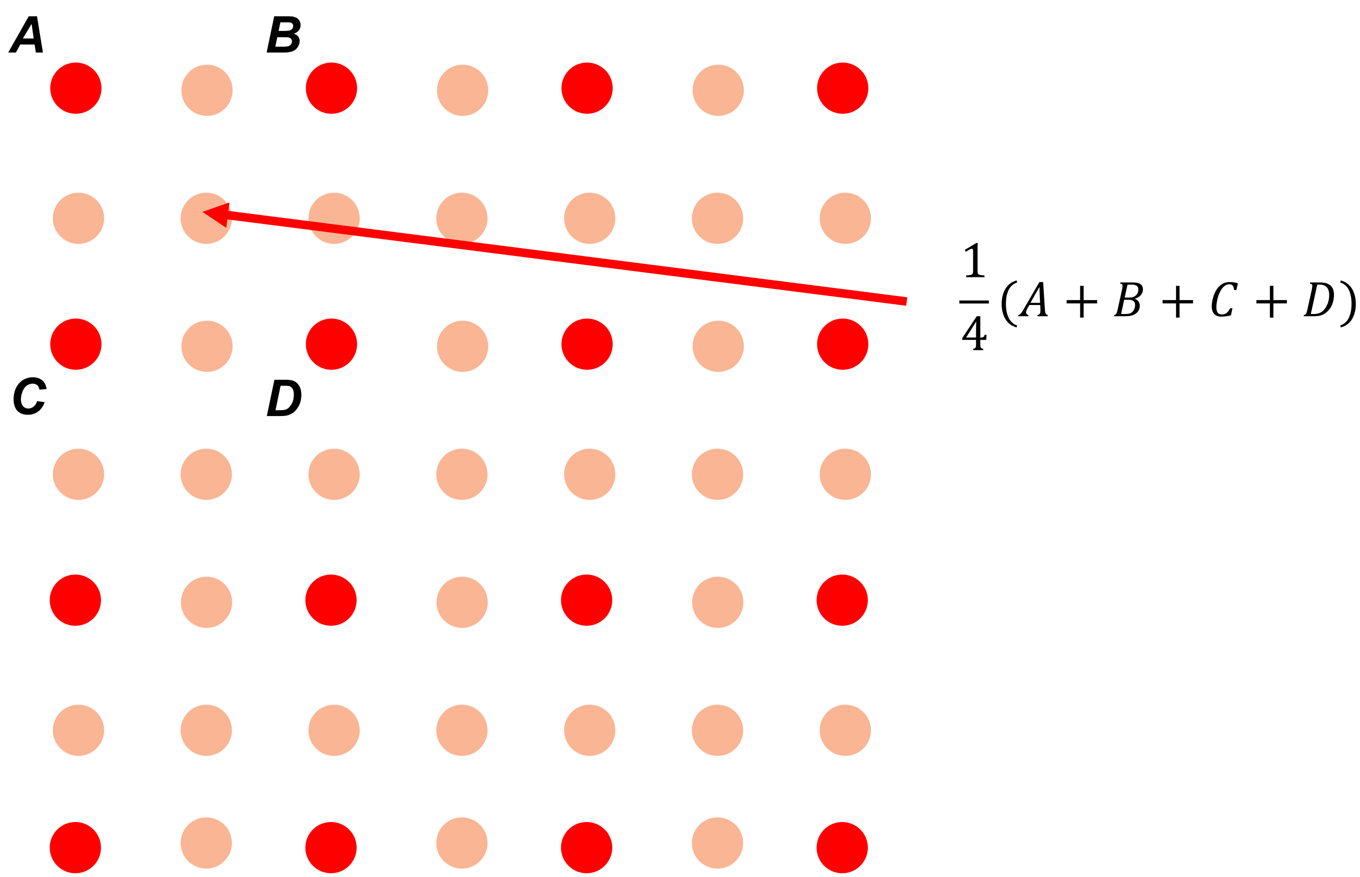


$$\frac{1}{2}(A+B)$$

How to upsample a feature map by a factor of 2?

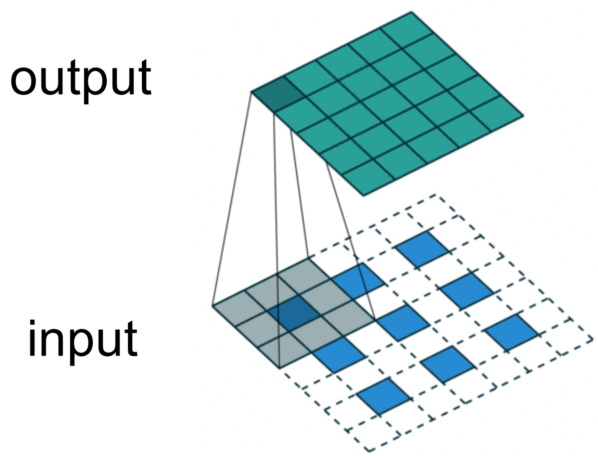


How to upsample a feature map by a factor of 2?



Upsampling by filtering

- For 2x upsampling, dilate the input by inserting rows and columns of zeros between adjacent entries, convolve with upsampling filter



Upsampling filter:

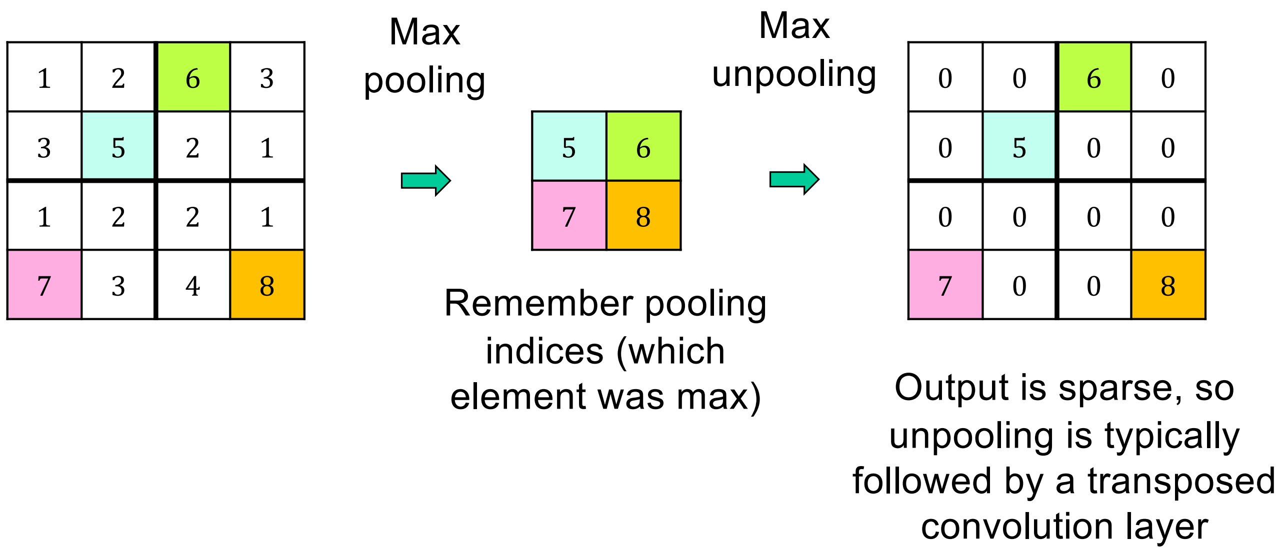
$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
$\frac{1}{2}$	1	$\frac{1}{2}$
$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

A learned upsampling filter can be used as well

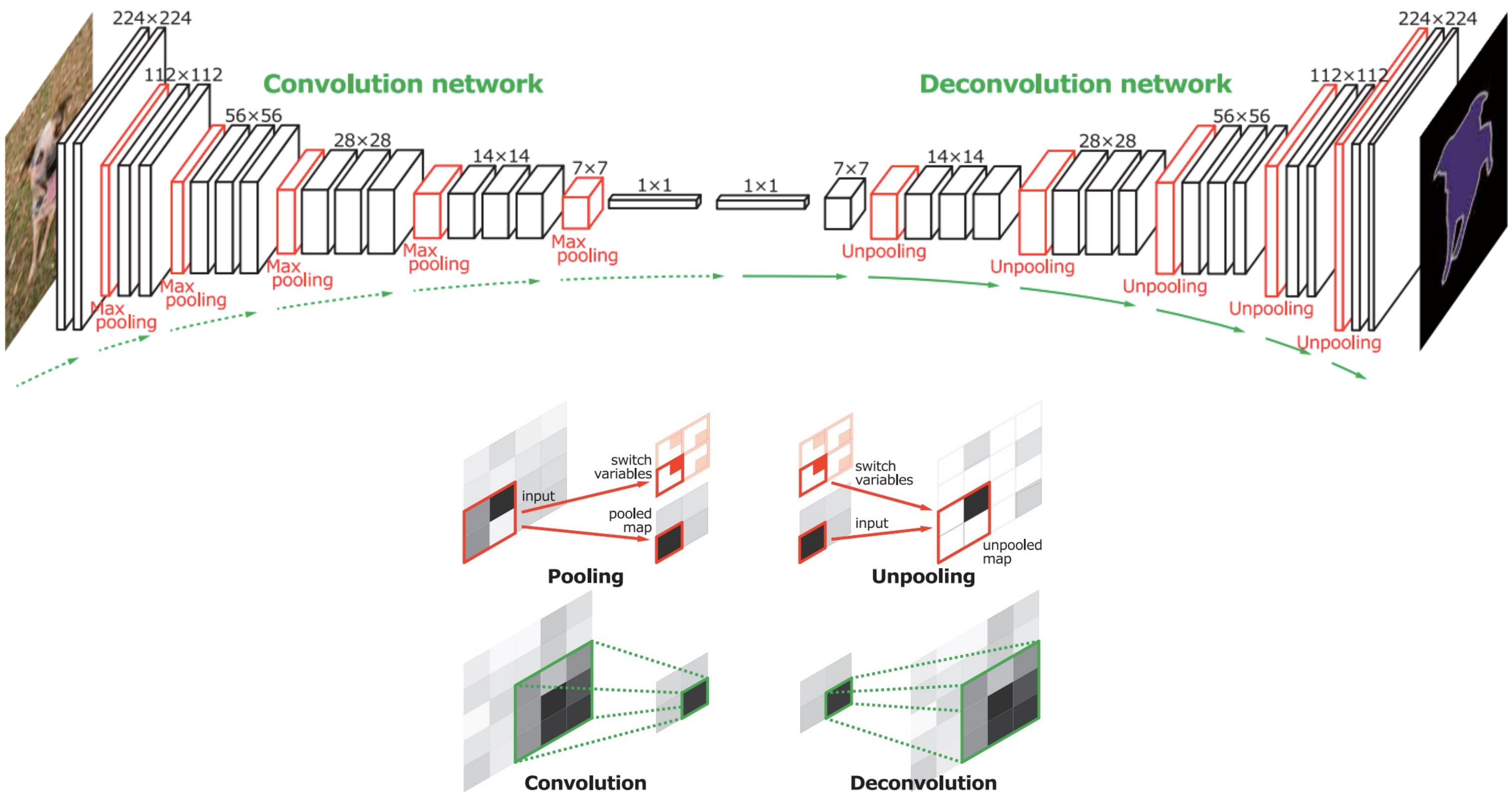
V. Dumoulin and F. Visin, A guide to convolution arithmetic for deep learning, arXiv 2018

Upsampling by max unpooling

- “Reversing” a max pooling operation:

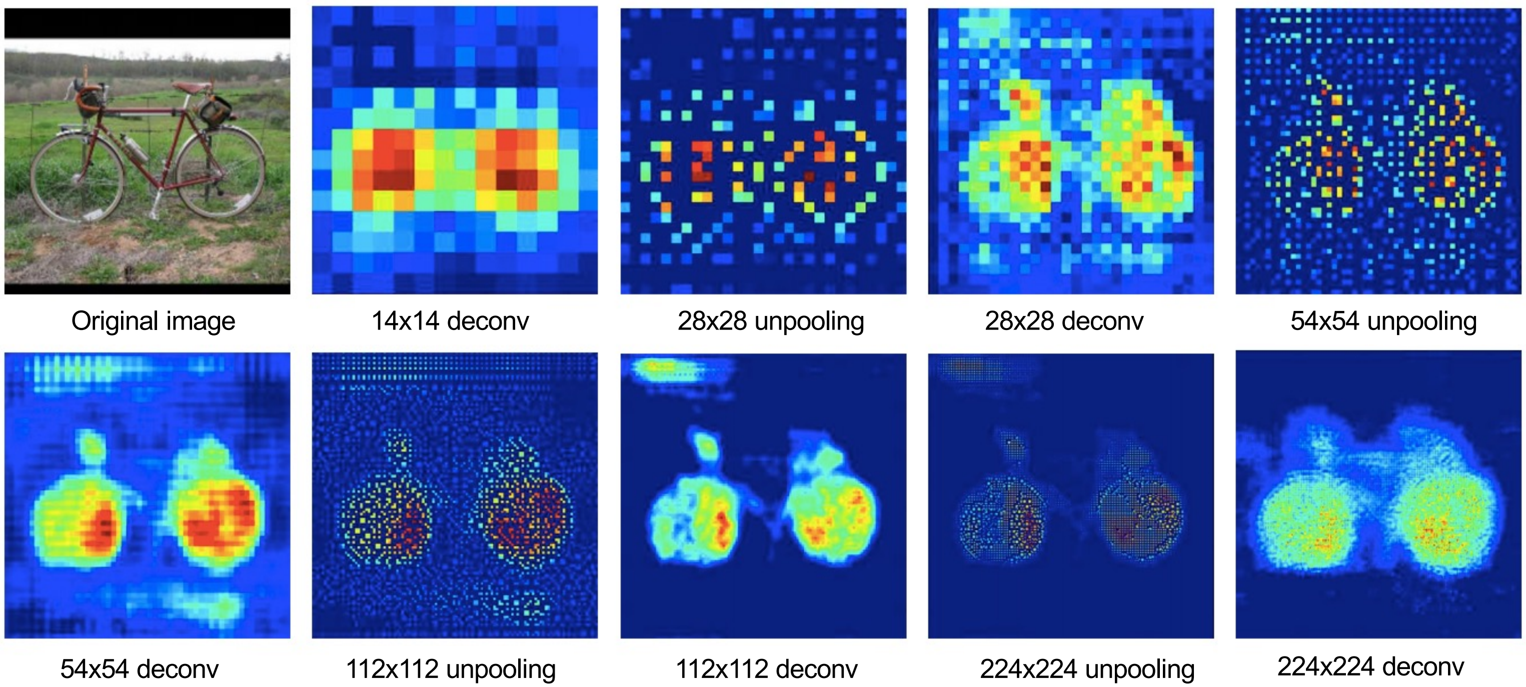


Early dense prediction architecture: DeconvNet



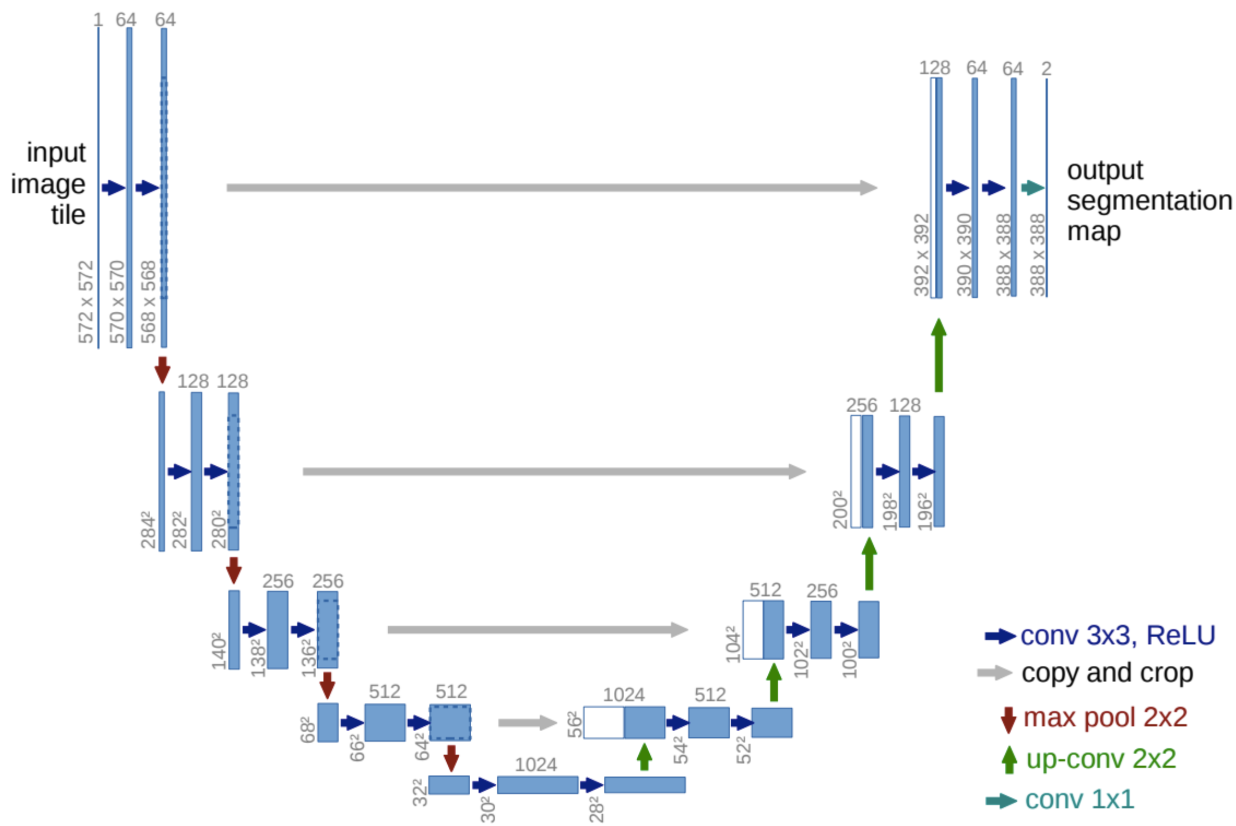
H. Noh et al. [Learning Deconvolution Network for Semantic Segmentation](). ICCV 2015

Early dense prediction architecture: DeconvNet



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

U-Net



O. Ronneberger et al. [U-Net: Convolutional Networks for Biomedical Image Segmentation](). MICCAI 2015

Example U-Net result

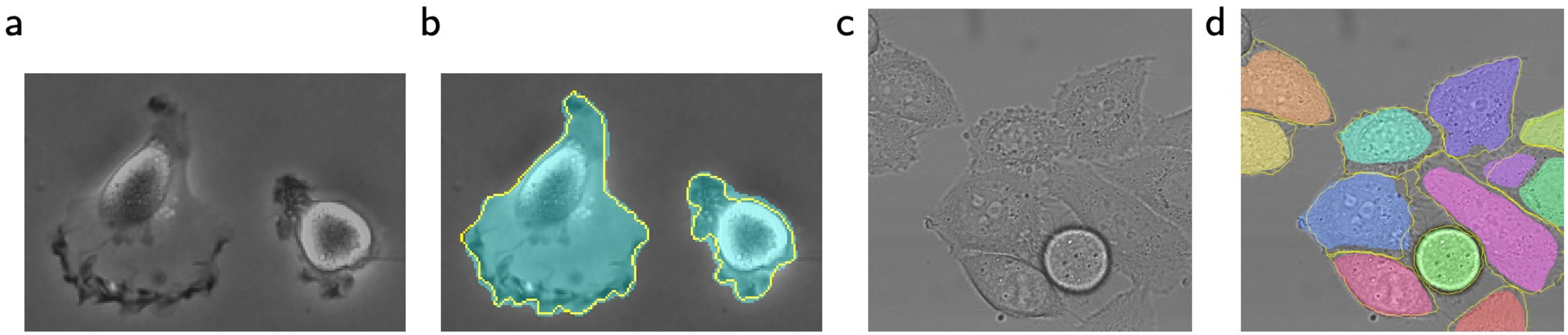


Fig. 4. Result on the ISBI cell tracking challenge. (**a**) part of an input image of the "PhC-U373" data set. (**b**) Segmentation result (cyan mask) with manual ground truth (yellow border) (**c**) input image of the "DIC-HeLa" data set. (**d**) Segmentation result (random colored masks) with manual ground truth (yellow border).

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

A more typical representation of U-Net

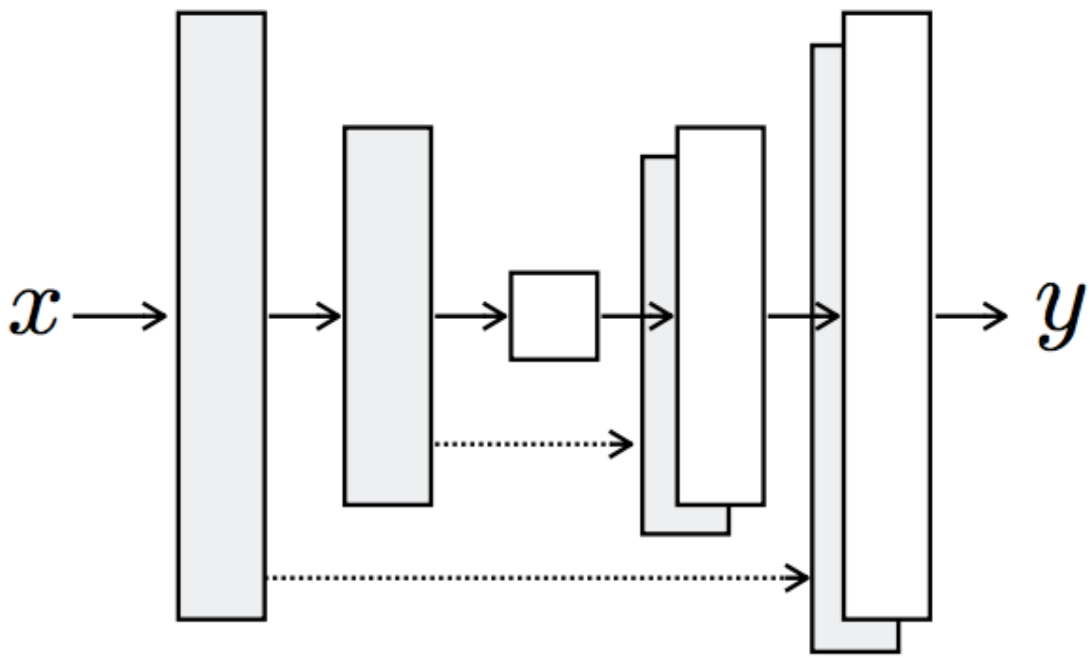
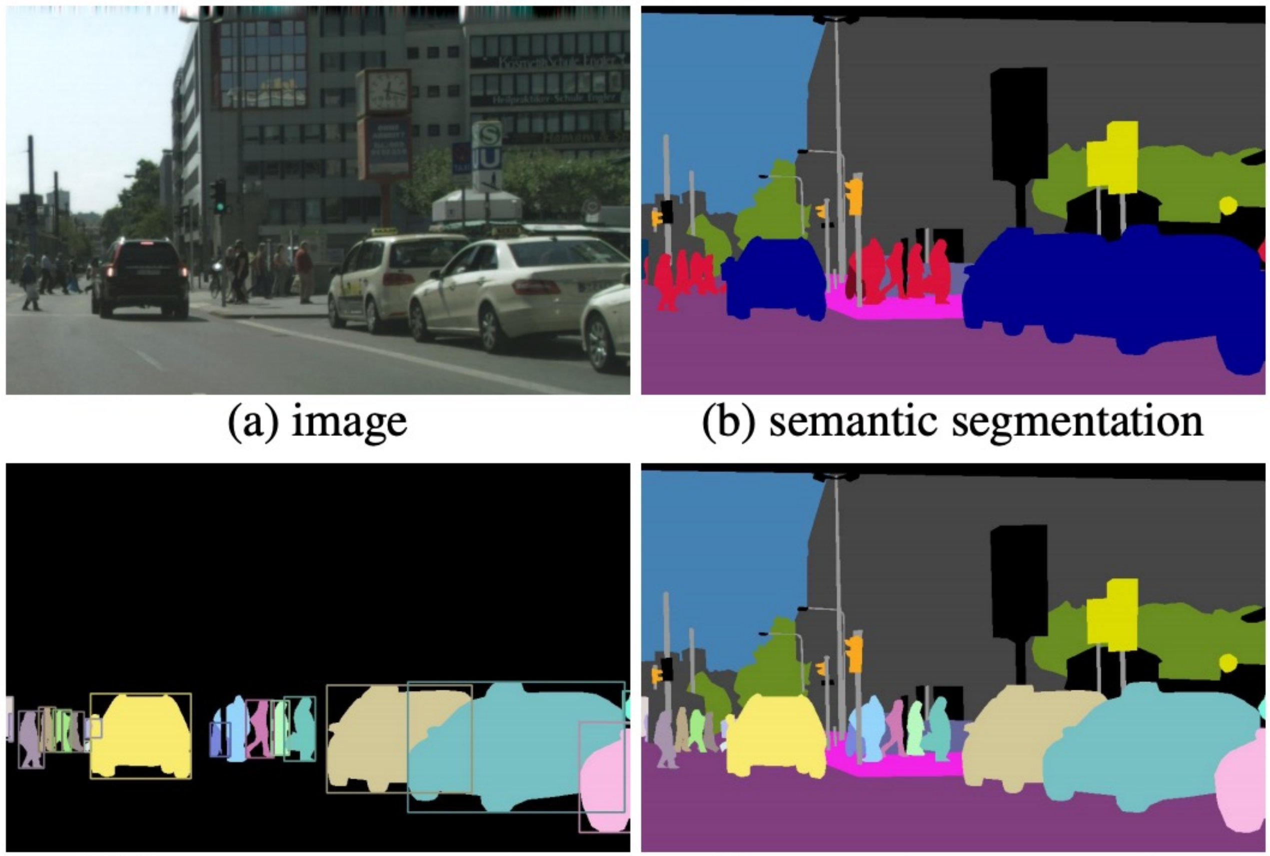


Figure source

Outline

- Dense prediction architectures
- Feature map upsampling
- U-Net
- A tour of dense prediction problems

Instance and panoptic segmentation



(a) image (b) semantic segmentation

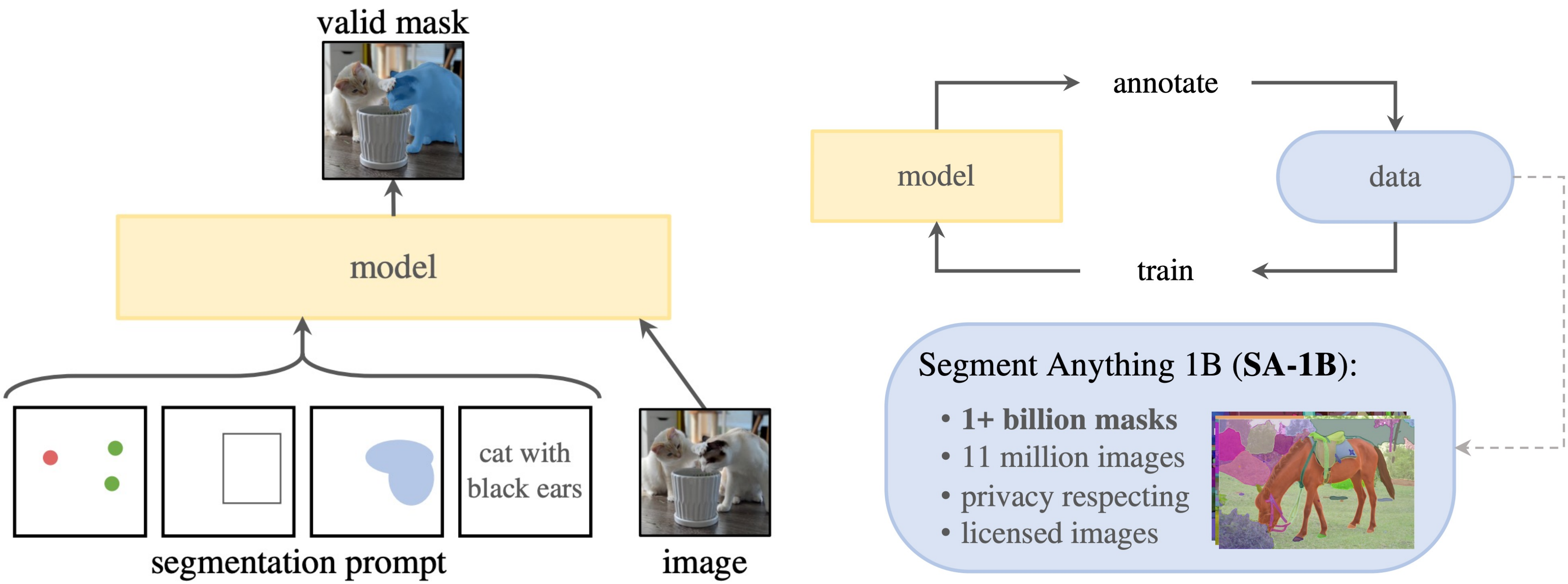
(c) instance segmentation (d) panoptic segmentation

A. Kirillov et al. [Panoptic segmentation](). CVPR 2019

Amodal segmentation

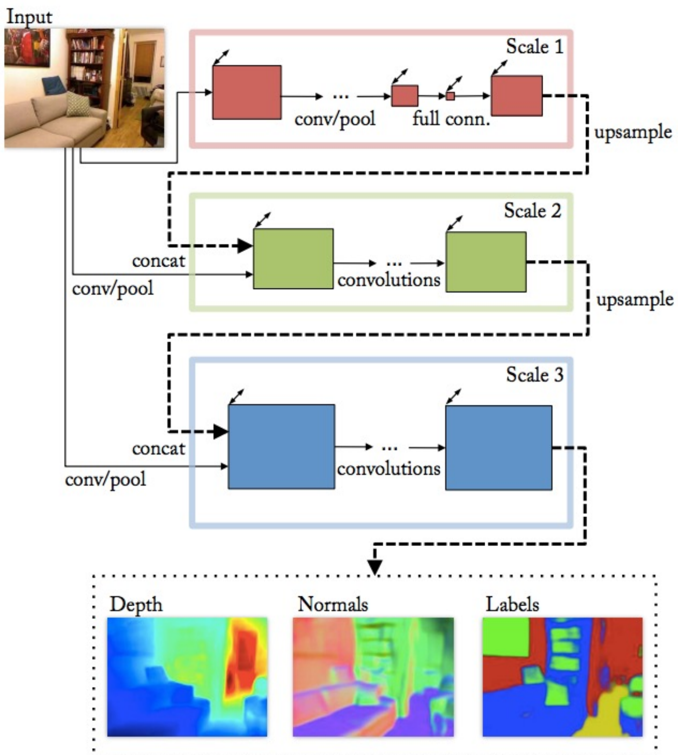
K. Li and J. Malik. [Amodal instance segmentation.](#) ECCV 2016

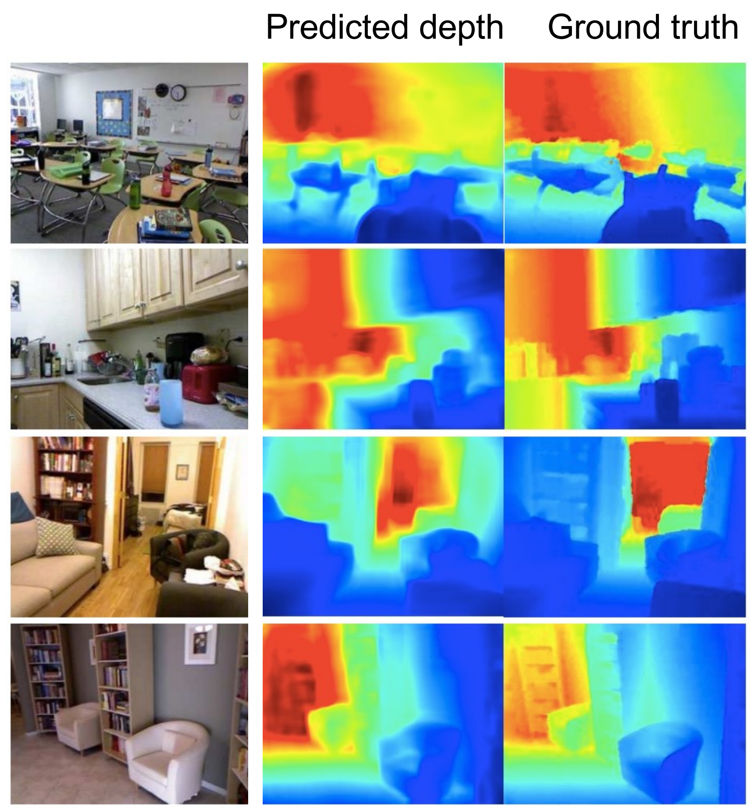
Promptable segmentation



A. Kirilllov et al. [Segment anything](). ICCV 2023

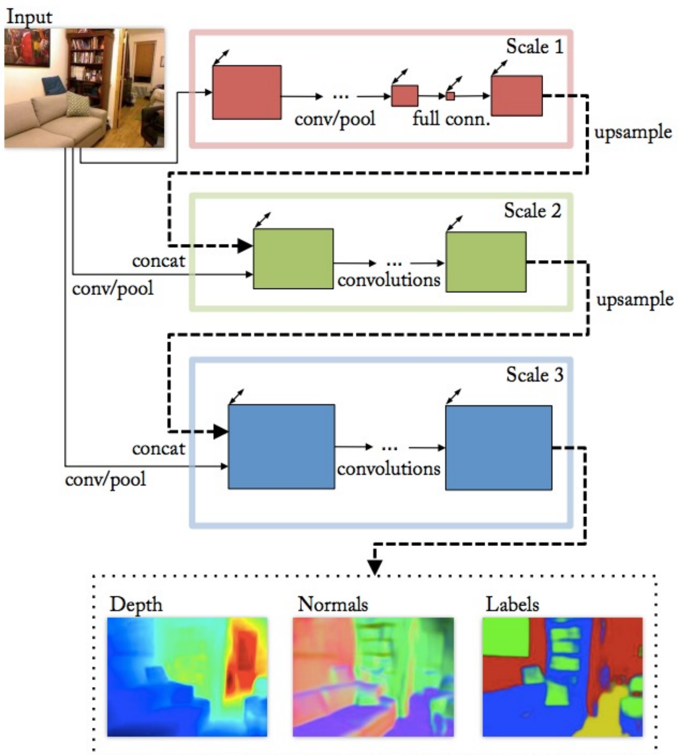
Depth and normal estimation



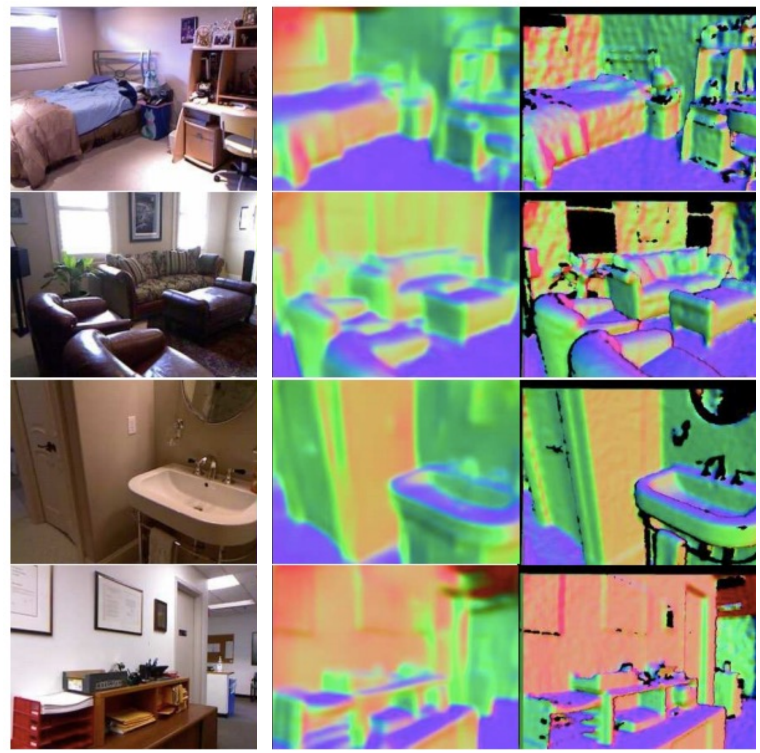


D. Eigen and R. Fergus, [Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture](), ICCV 2015

Depth and normal estimation

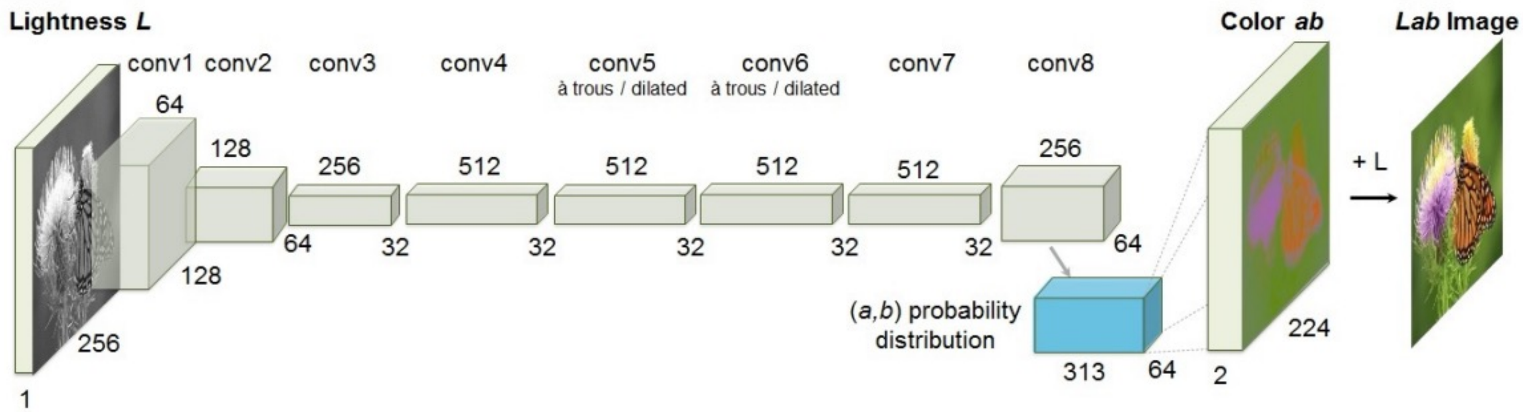


Predicted normals Ground truth



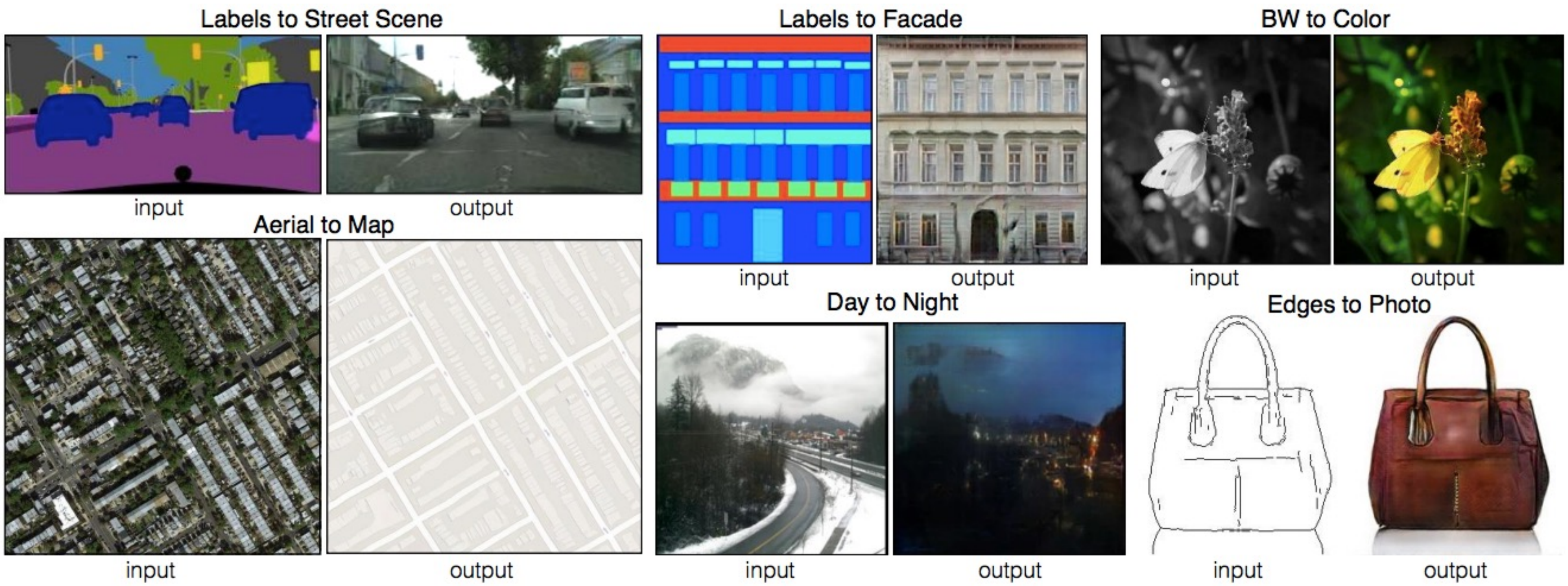
D. Eigen and R. Fergus, [Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture](), ICCV 2015

Colorization



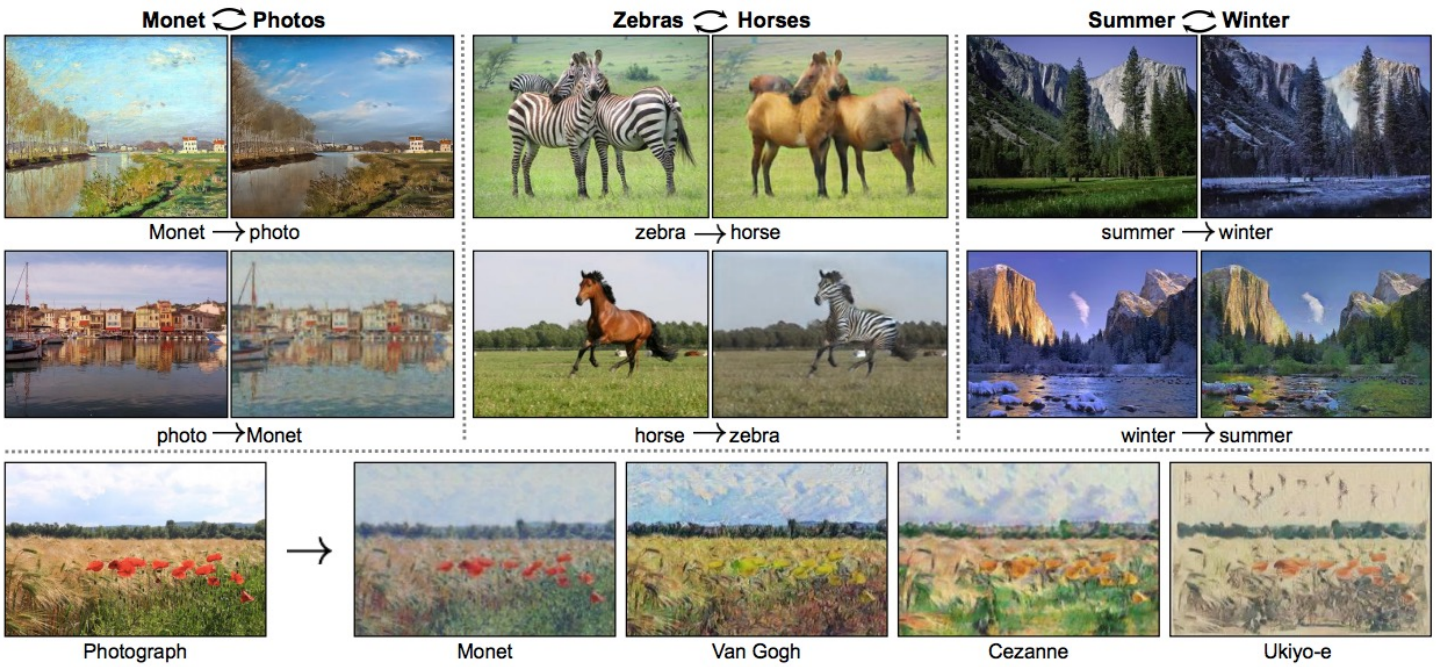
R. Zhang, P. Isola, and A. Efros, Colorful Image Colorization, ECCV 2016

Image-to-image translation (paired)



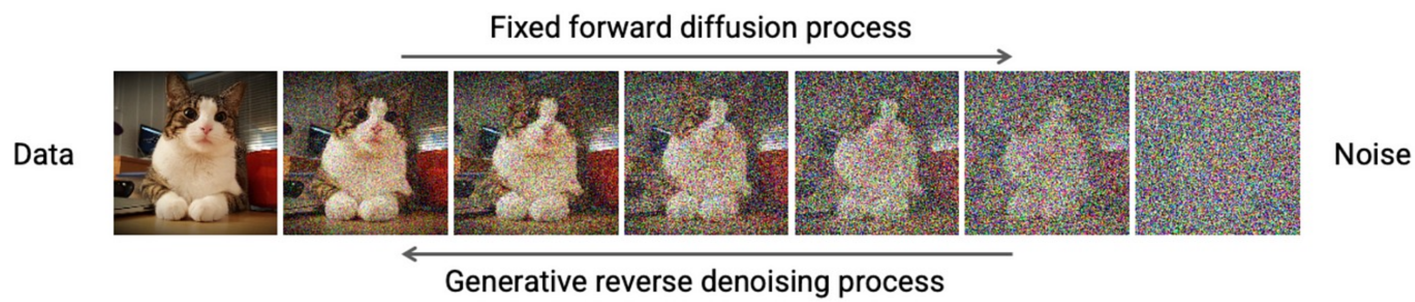
P. Isola, J.-Y. Zhu, T. Zhou, A. Efros, [Image-to-Image Translation with Conditional Adversarial Networks](), CVPR 2017

Image-to-image translation (unpaired)



J.-Y. Zhu, T. Park, P. Isola, A. Efros, [Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks](), ICCV 2017

Denoising diffusion probabilistic models



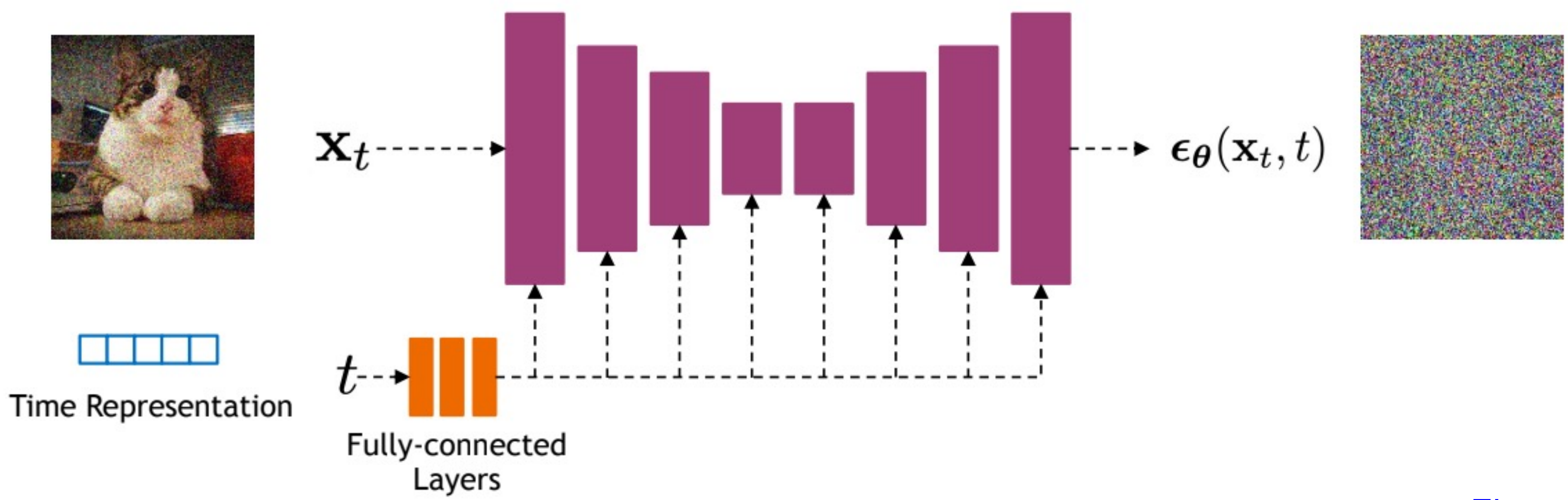


Figure source