

本科生《计算机视觉》

基于深度学习的视觉理解与生成

第四节 分割

黄雷

人工智能研究院

huangleiAI@buaa.edu.cn

2023年10月19日

主要内容

- 深度学习基础
 - 神经网络及反向传播算法
 - 卷积神经网络中的视觉表示思想
- 视觉理解任务
 - 目标检测
 - 分割
- 视觉生成
 - 深度生成模型
 - 图像翻译任务详解
- 深度神经网络训练技巧

Segmentation



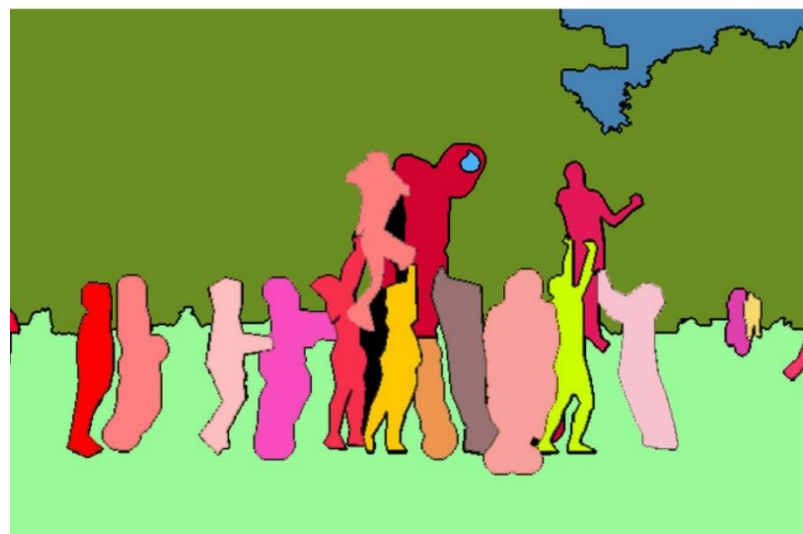
Semantic Segmentation



Instance Segmentation



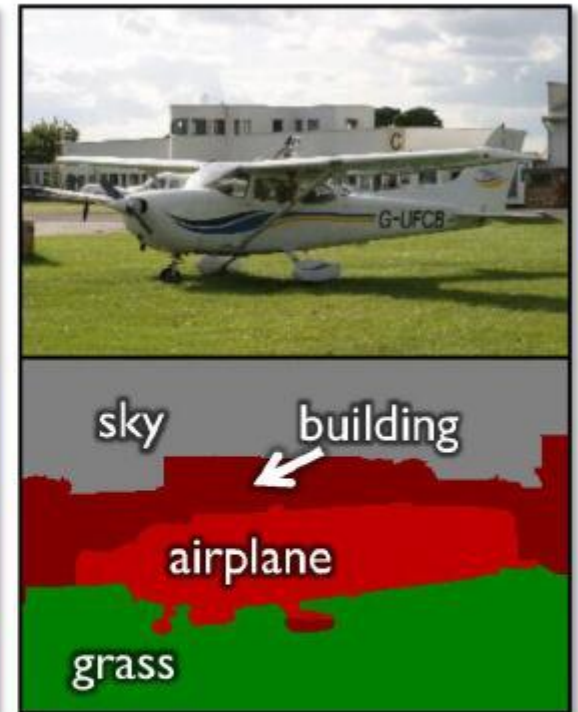
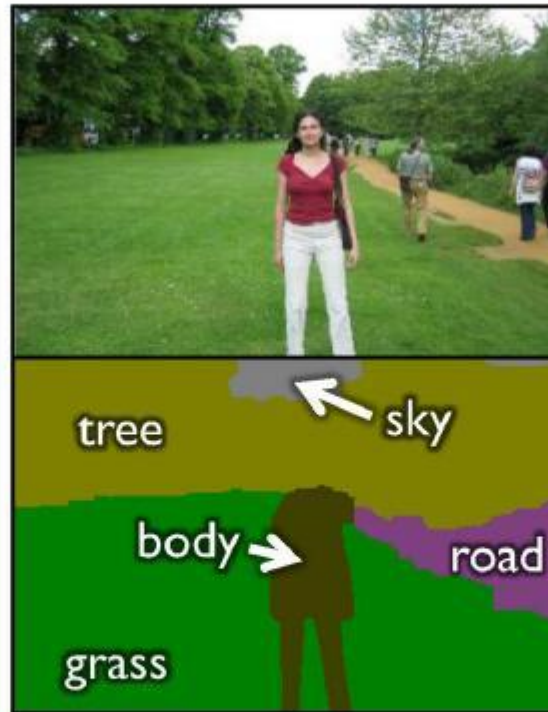
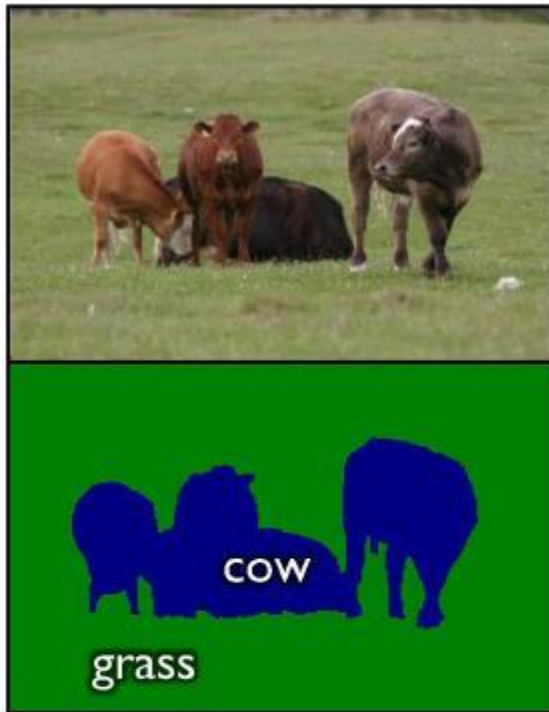
Panoptic Segmentation



outline

- Semantic Segmentation
- Instance Segmentation
- Panoptic Segmentation
- Prompt based Segmentation
- Image Translation---Dense prediction

Semantic Segmentation



| | | | | | | | | | | |
|----------------|----------|-------|------|------|-------|------|----------|-------|------|------|
| object classes | building | grass | tree | cow | sheep | sky | airplane | water | face | car |
| bicycle | flower | sign | bird | book | chair | road | cat | dog | body | boat |

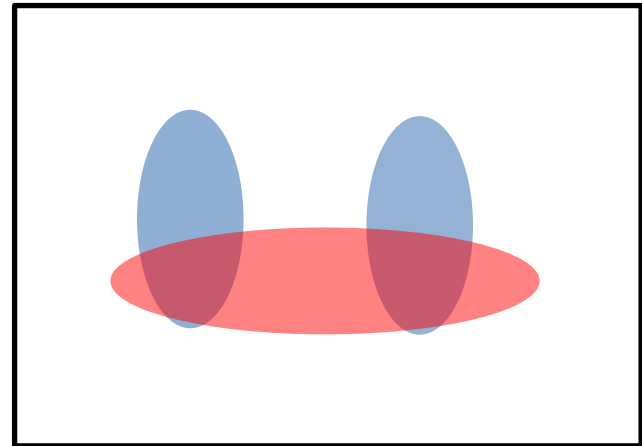
The Task



- person
- grass
- trees
- motorbike
- road

Evaluation metric

- Pixel classification!
- Accuracy?
 - Heavily unbalanced
 - Common classes are over-emphasized
- *Intersection over Union*
 - Average across classes and images
- Per-class accuracy
 - Compute accuracy for every class and then average



Things vs Stuff

THINGS

- Person, cat, horse, etc
- Constrained shape
- Individual instances with separate identity
- May need to look at objects



STUFF

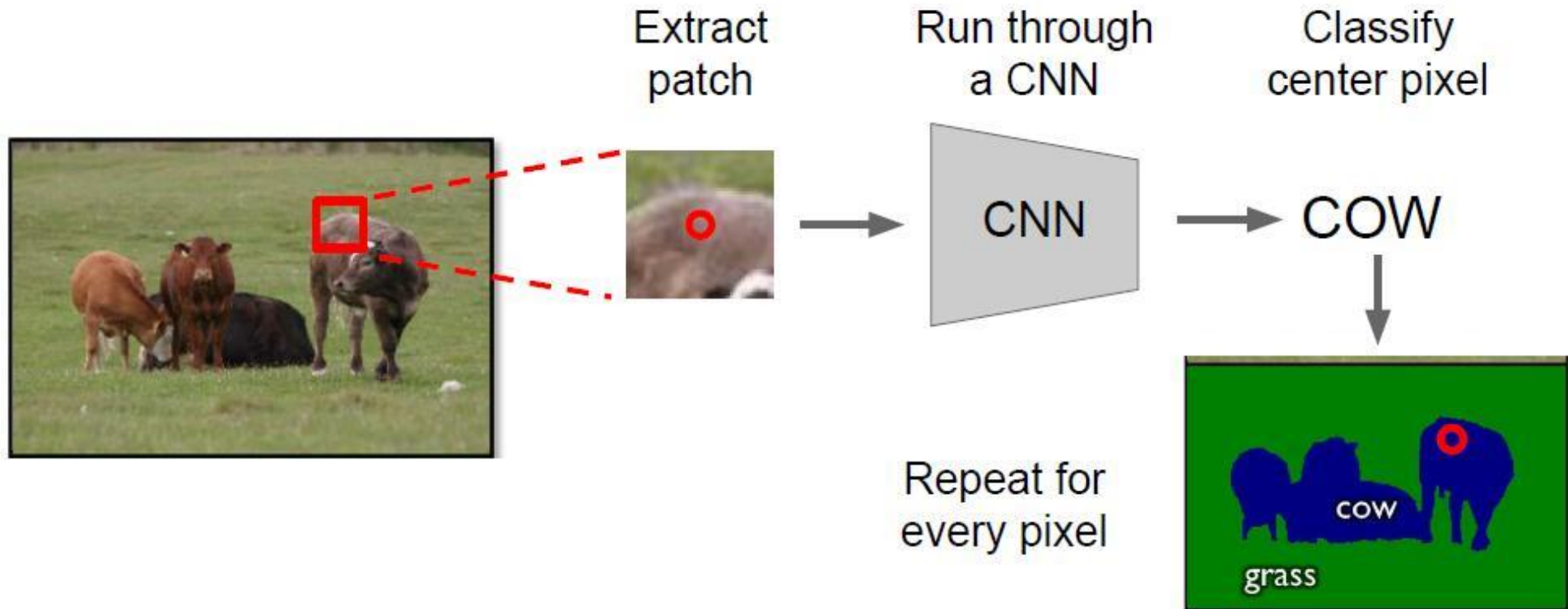
- Road, grass, sky etc
- Amorphous, no shape
- No notion of instances
- Can be done at pixel level
- “texture”



Challenges in data collection

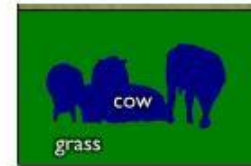
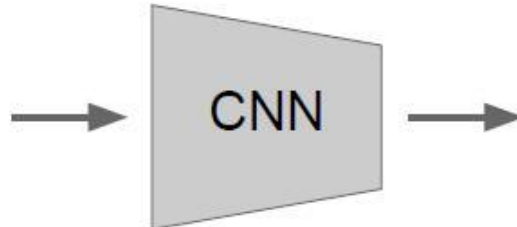
- Precise localization is hard to annotate
- Annotating every pixel leads to heavy tails
- Common solution: annotate few classes (often things), mark rest as “Other”
- Common datasets: PASCAL VOC 2012 (~1500 images, 20 categories), COCO (~100k images, 20 categories)

Simple Routine



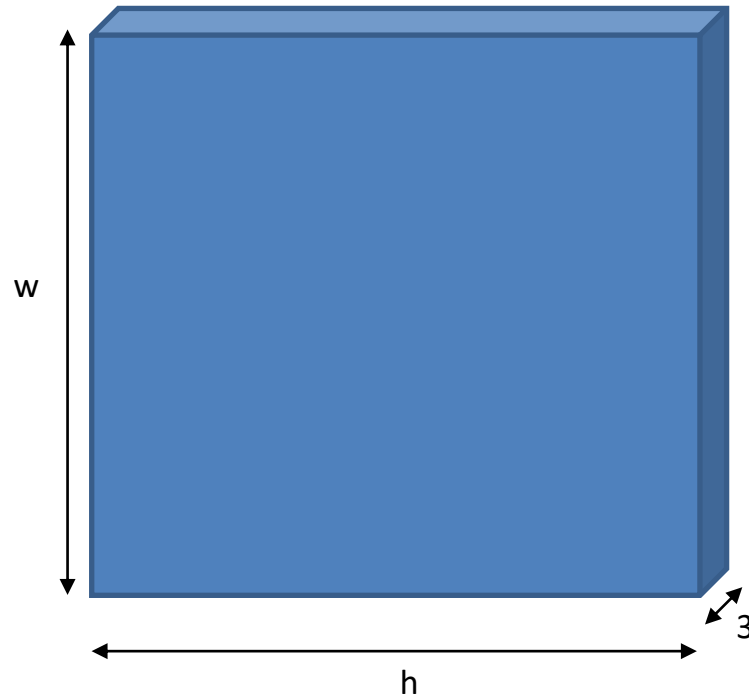
Another Routine

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

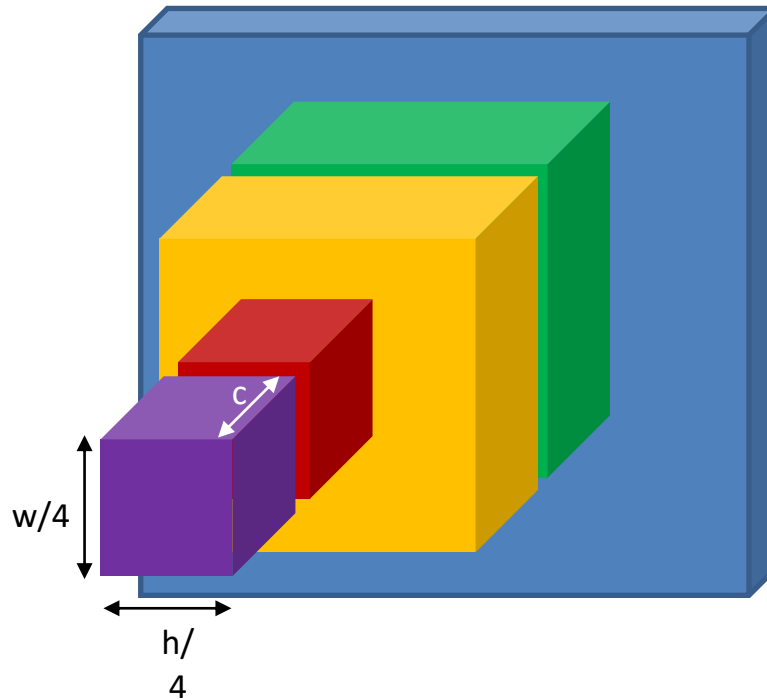


Smaller output
due to pooling

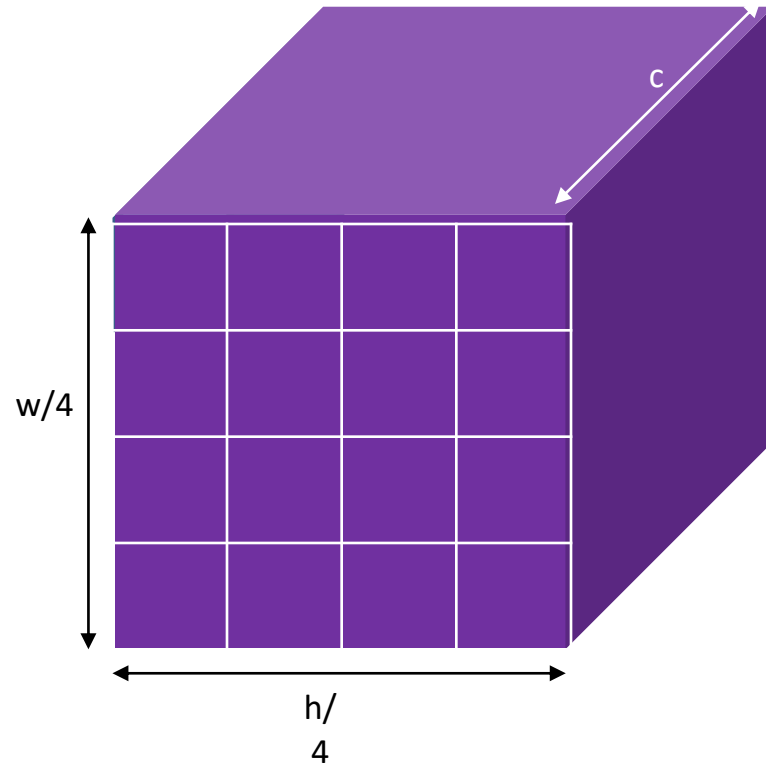
Semantic segmentation using convolutional networks



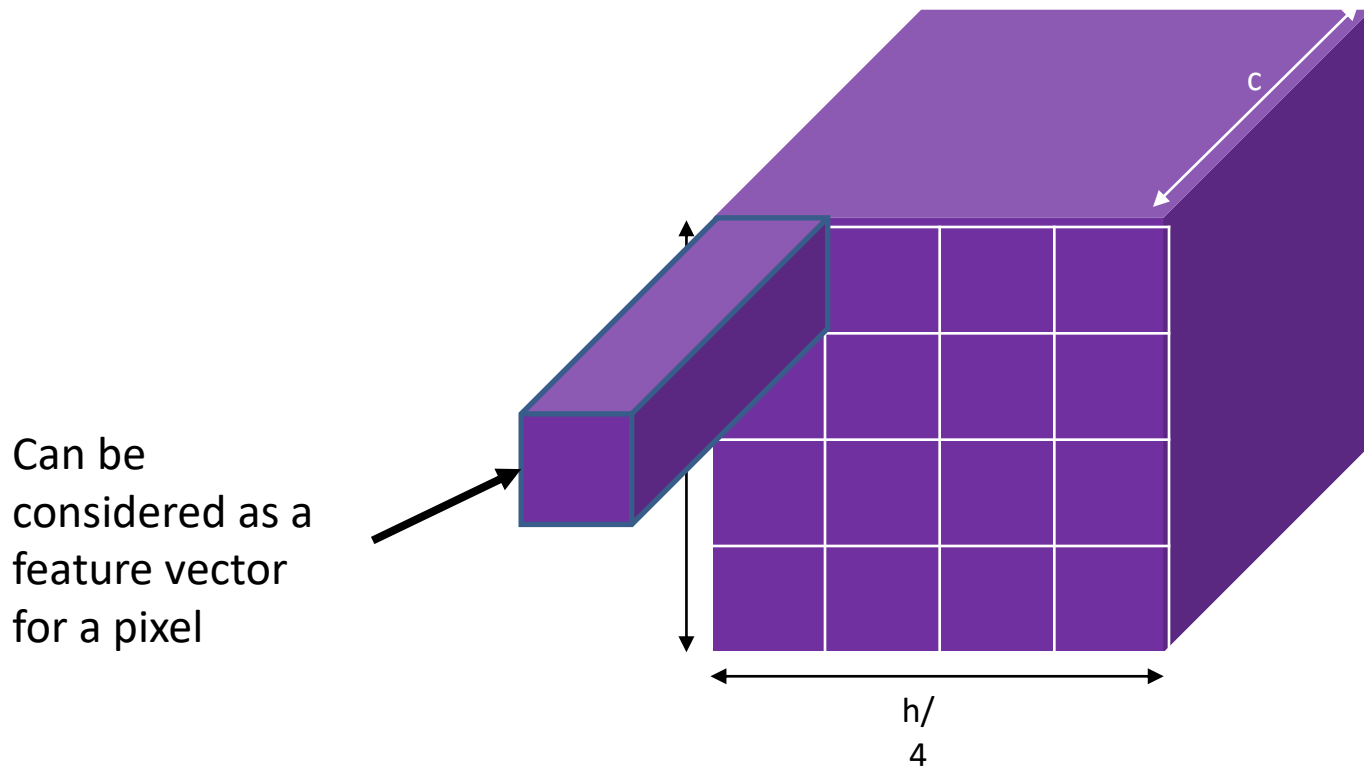
Semantic segmentation using convolutional networks



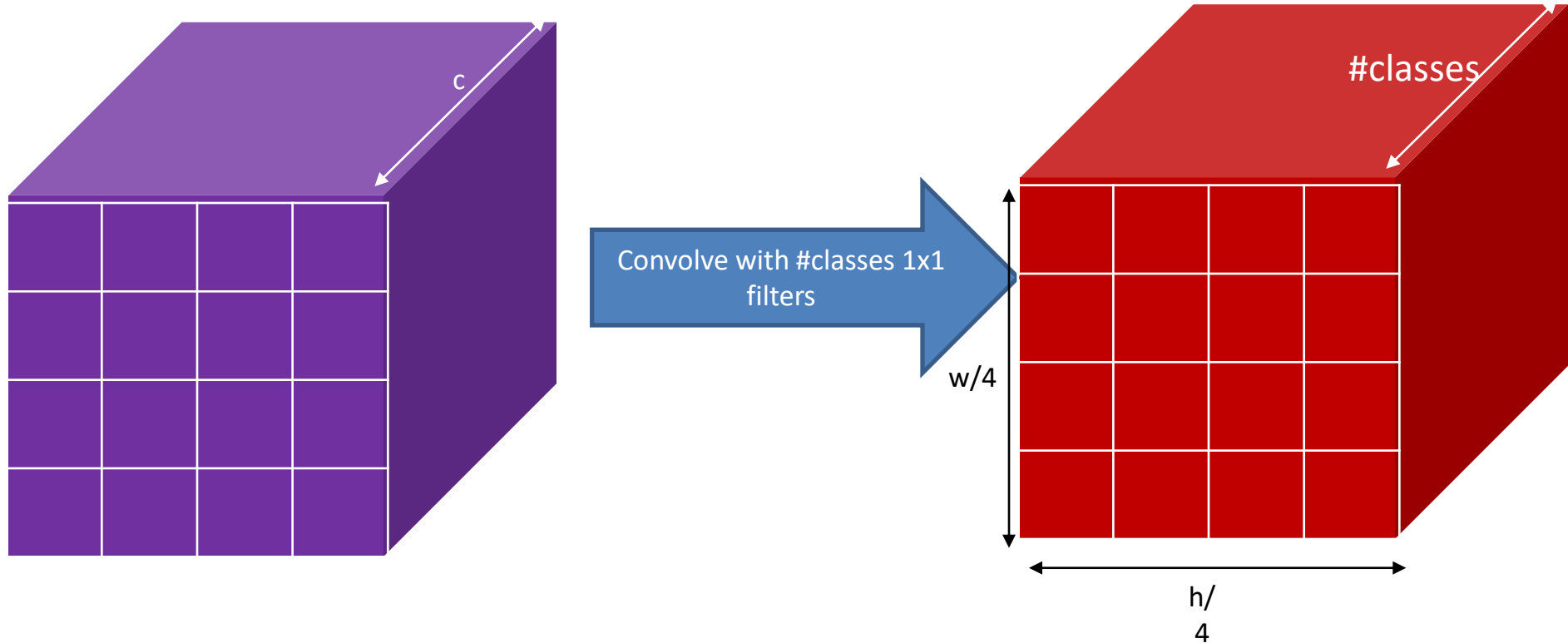
Semantic segmentation using convolutional networks



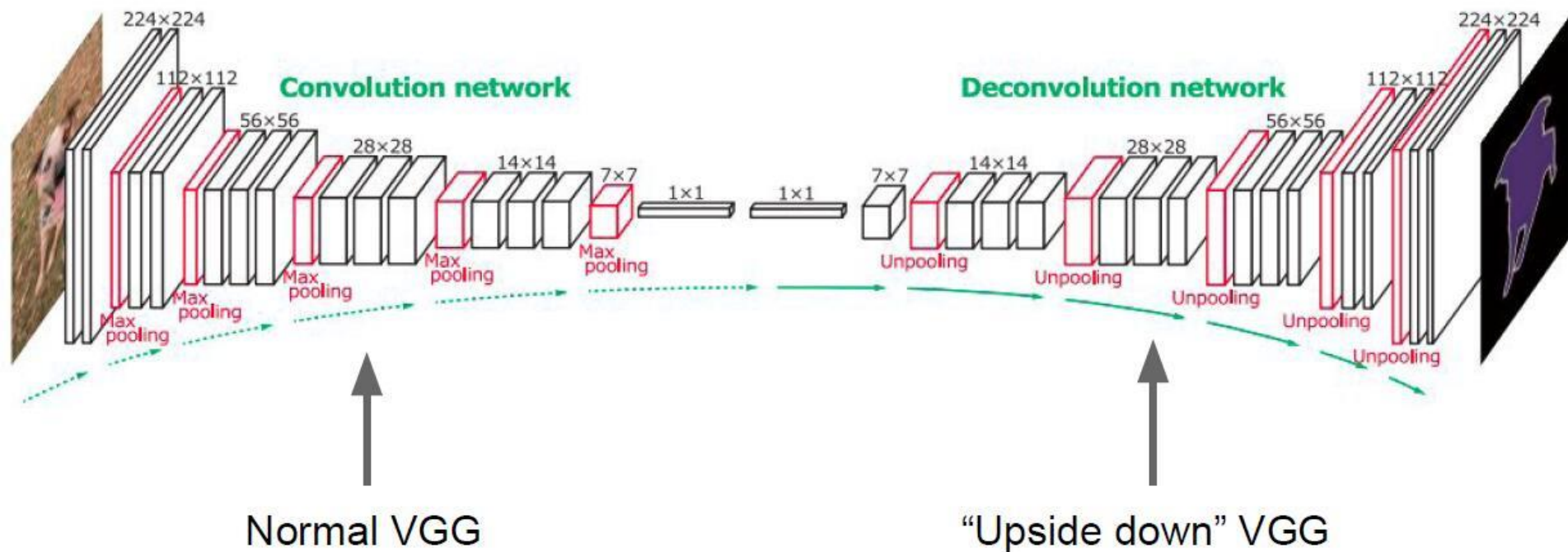
Semantic segmentation using convolutional networks



Semantic segmentation using convolutional networks

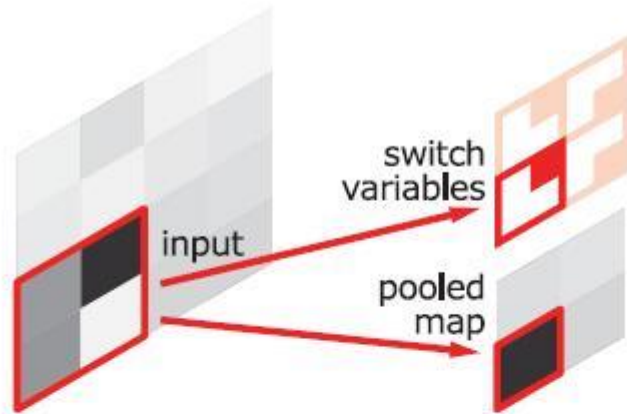


Semantic Segmentation: Upsampling

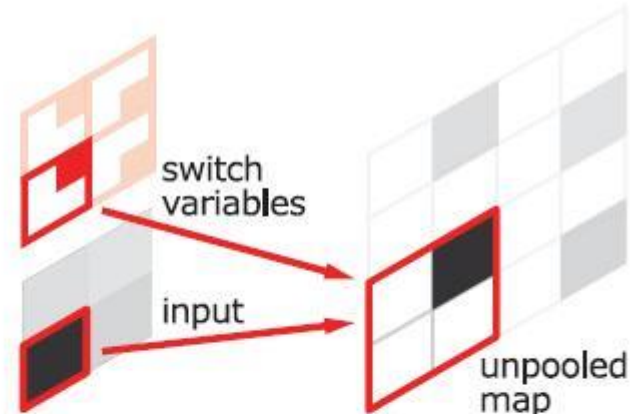


Noh et al Learning Deconvolution Network for Semantic Segmentation, ICCV 2015

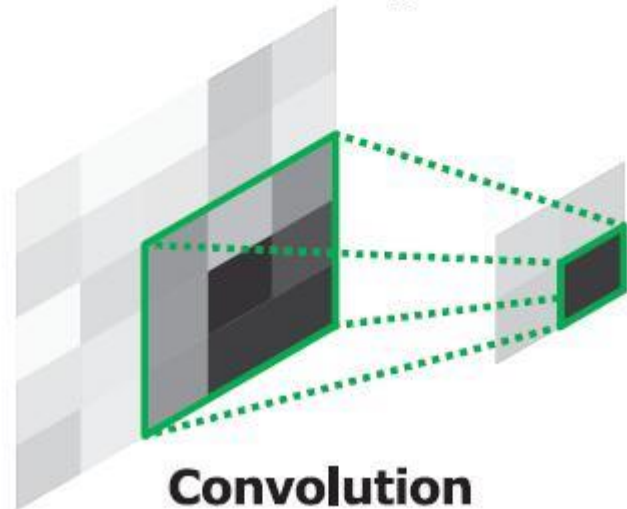
Semantic Segmentation: Upsampling



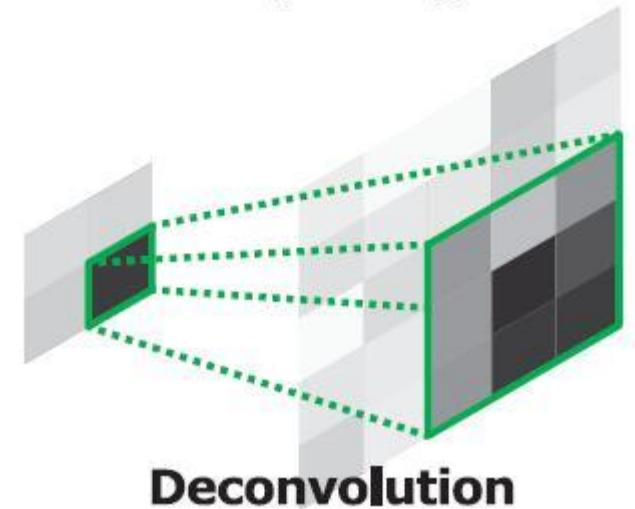
Pooling



Unpooling



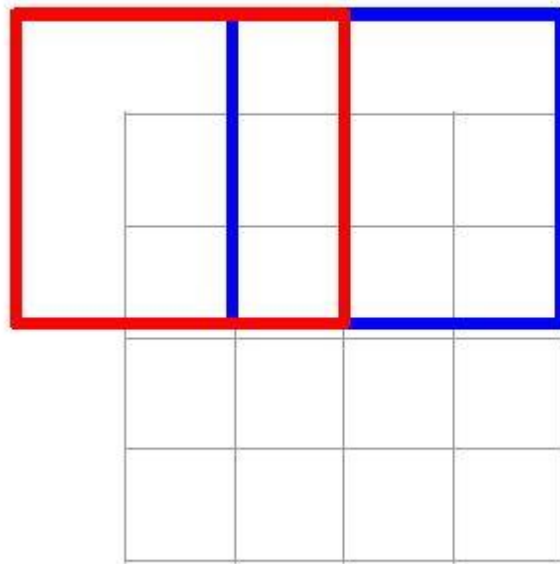
Convolution



Deconvolution

Semantic Segmentation: Upsampling

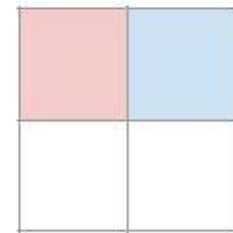
Typical 3 x 3 convolution, stride 2 pad 1



Input: 4 x 4



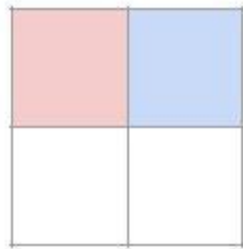
Dot product
between filter
and input



Output: 2 x 2

Semantic Segmentation: Upsampling

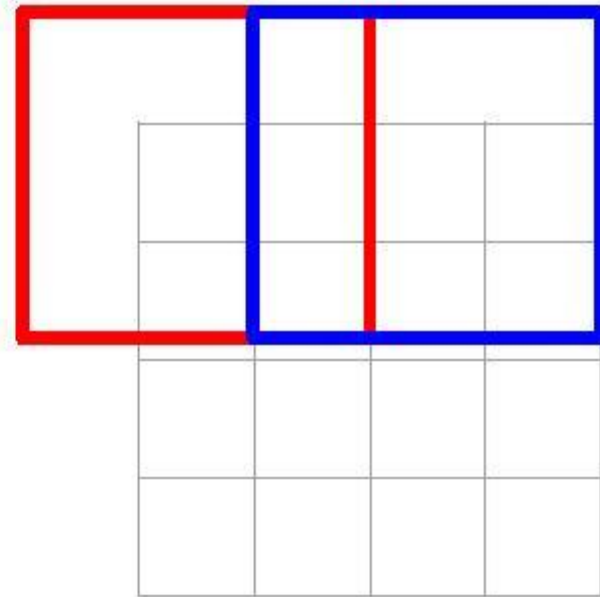
3 x 3 “deconvolution”, stride 2 pad 1



Input: 2 x 2



Input gives
weight for
filter

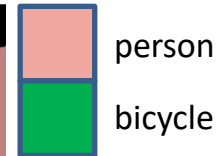
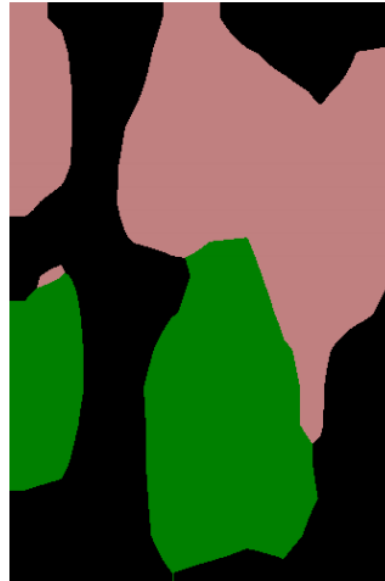


Output: 4 x 4

Semantic segmentation using convolutional networks

- Pass image through convolution and subsampling layers
- Final convolution with #classes outputs
- Get scores for *subsampled* image
- Upsample back to original size

Semantic segmentation using convolutional networks



outline

- Semantic Segmentation
- Instance Segmentation
- Panoptic Segmentation
- Prompt based Segmentation
- Image Translation---Dense prediction

Instance Segmentation



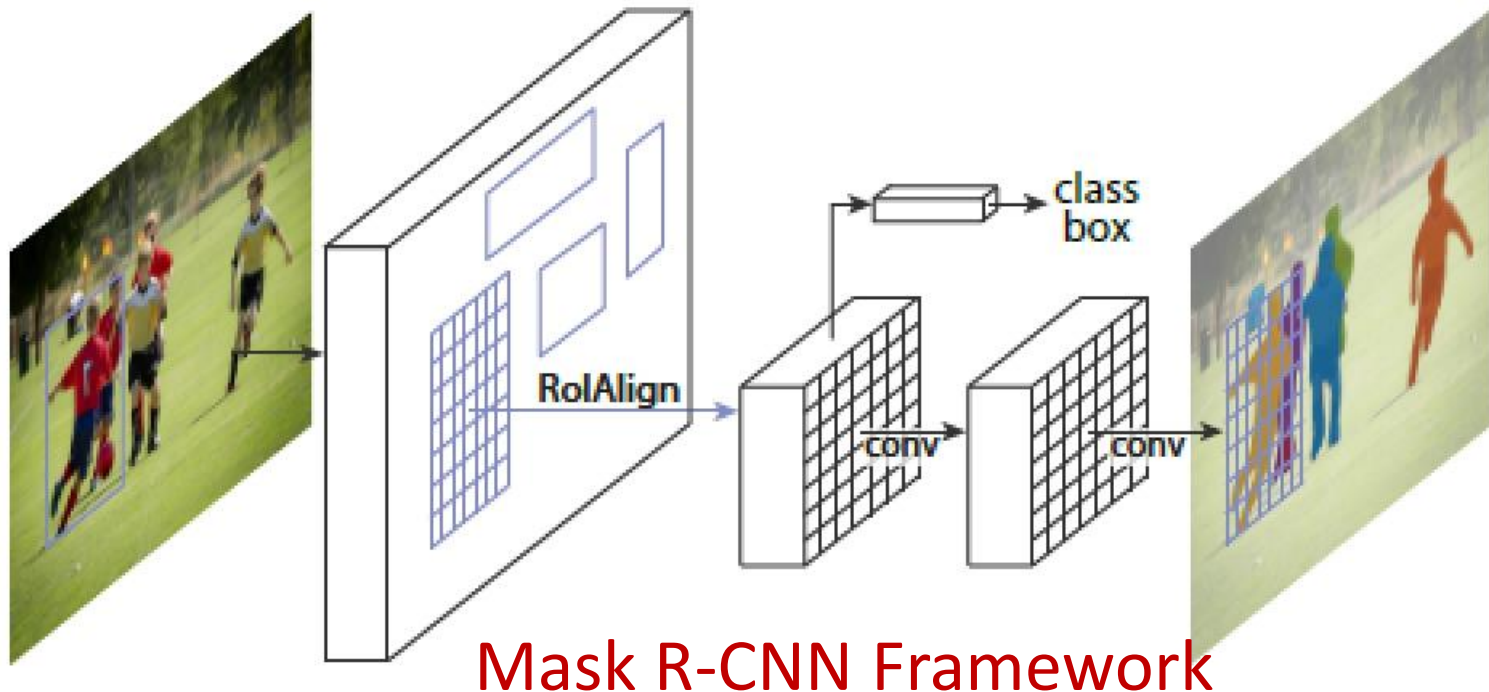
Instance Segmentation



Top-Down



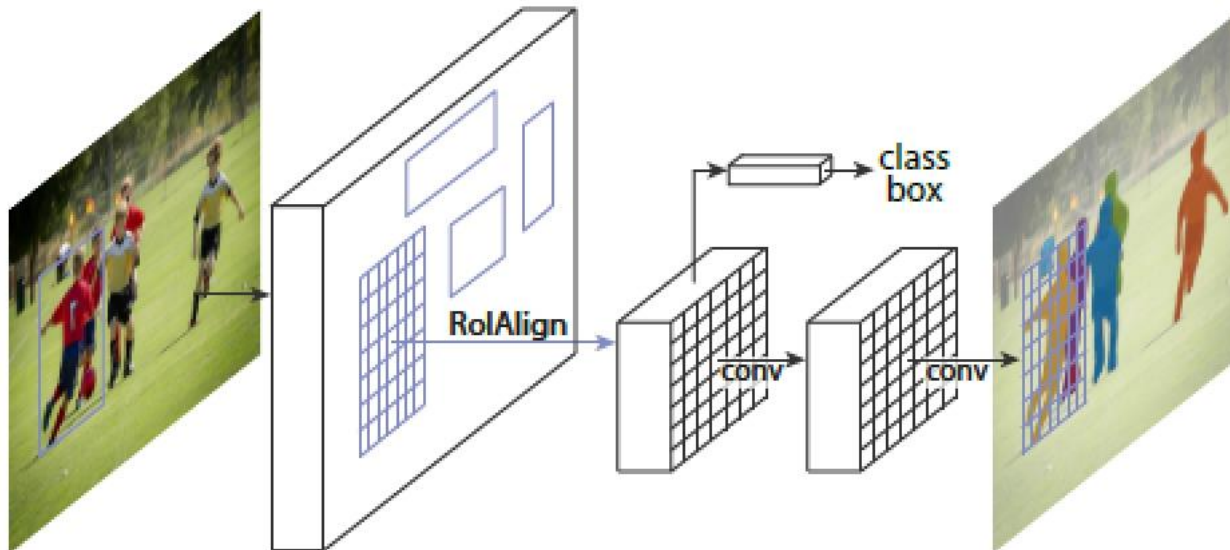
Object Detection+ mask prediction



Top-Down

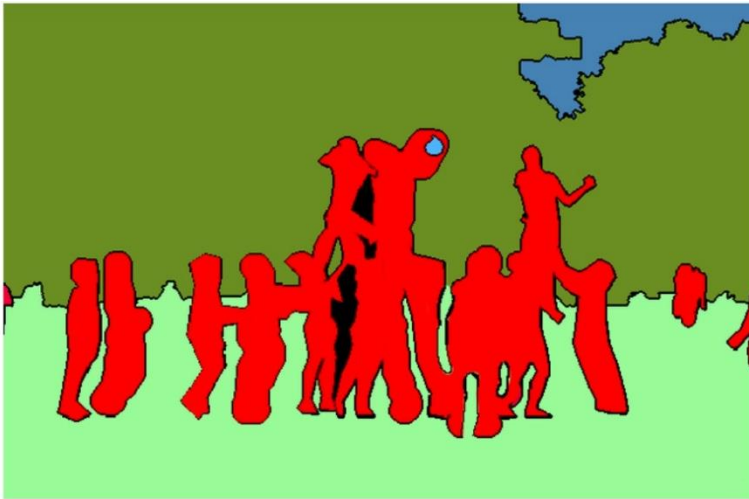


Region Proposal + Semantic segmentation



Bottom Up

- Semantic segmentation + Group



Instance Segmentation

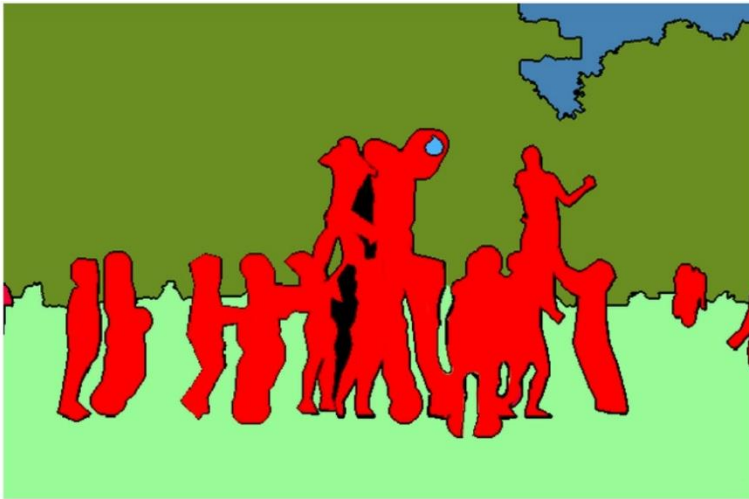


outline

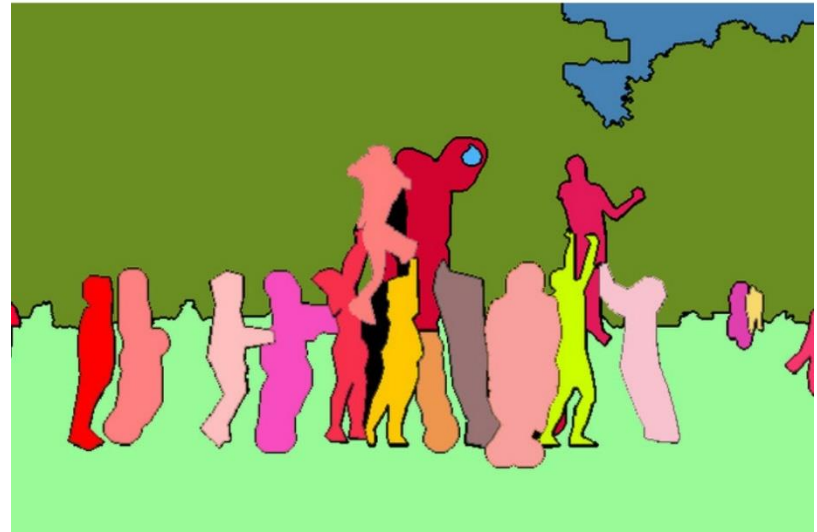
- Semantic Segmentation
- Instance Segmentation
- **Panoptic Segmentation**
- Prompt based Segmentation
- Image Translation---Dense prediction

Panoptic Segmentation

- Semantic segmentation + Group



Panoptic Segmentation

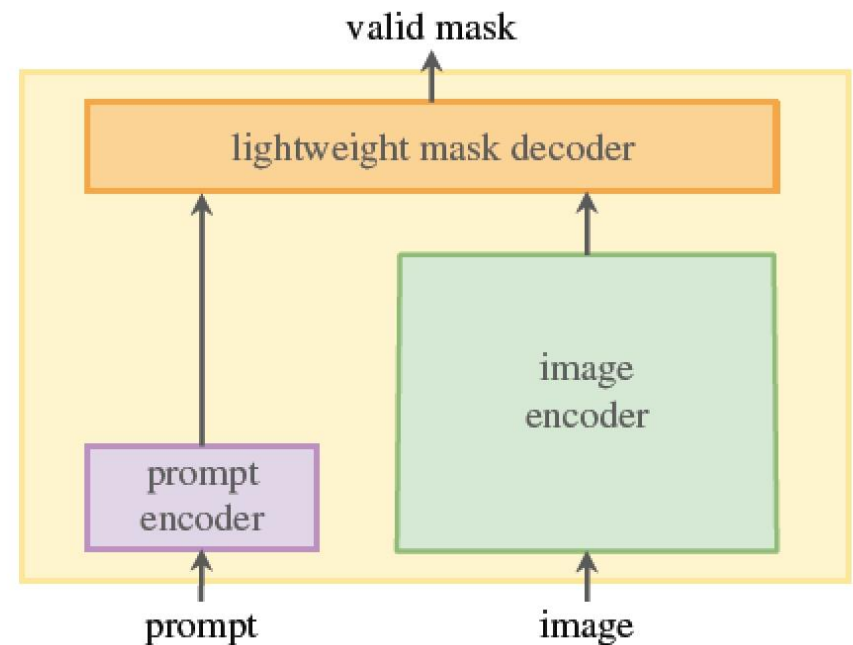
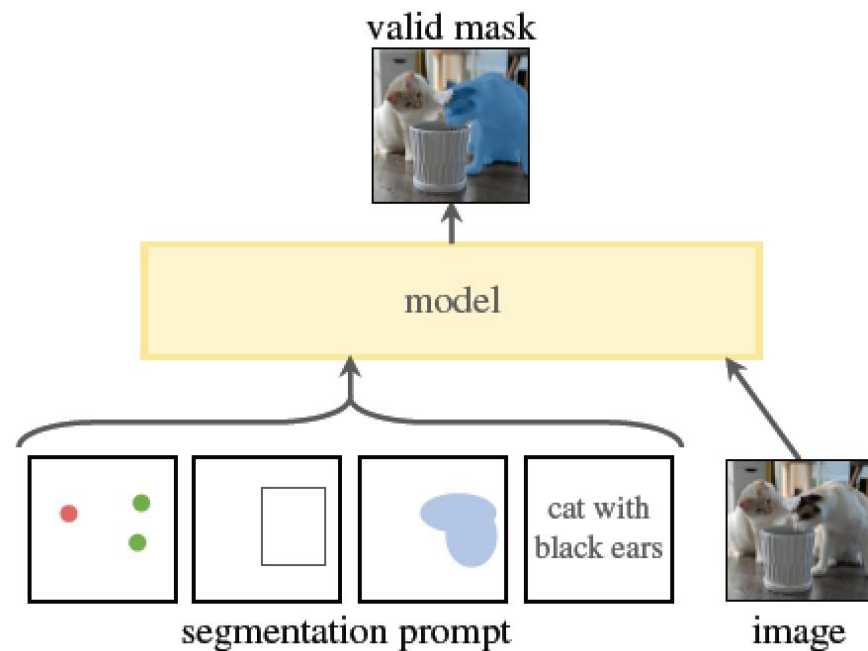


outline

- Semantic Segmentation
- Instance Segmentation
- Panoptic Segmentation
- Prompt based Segmentation
- Image Translation---Dense prediction

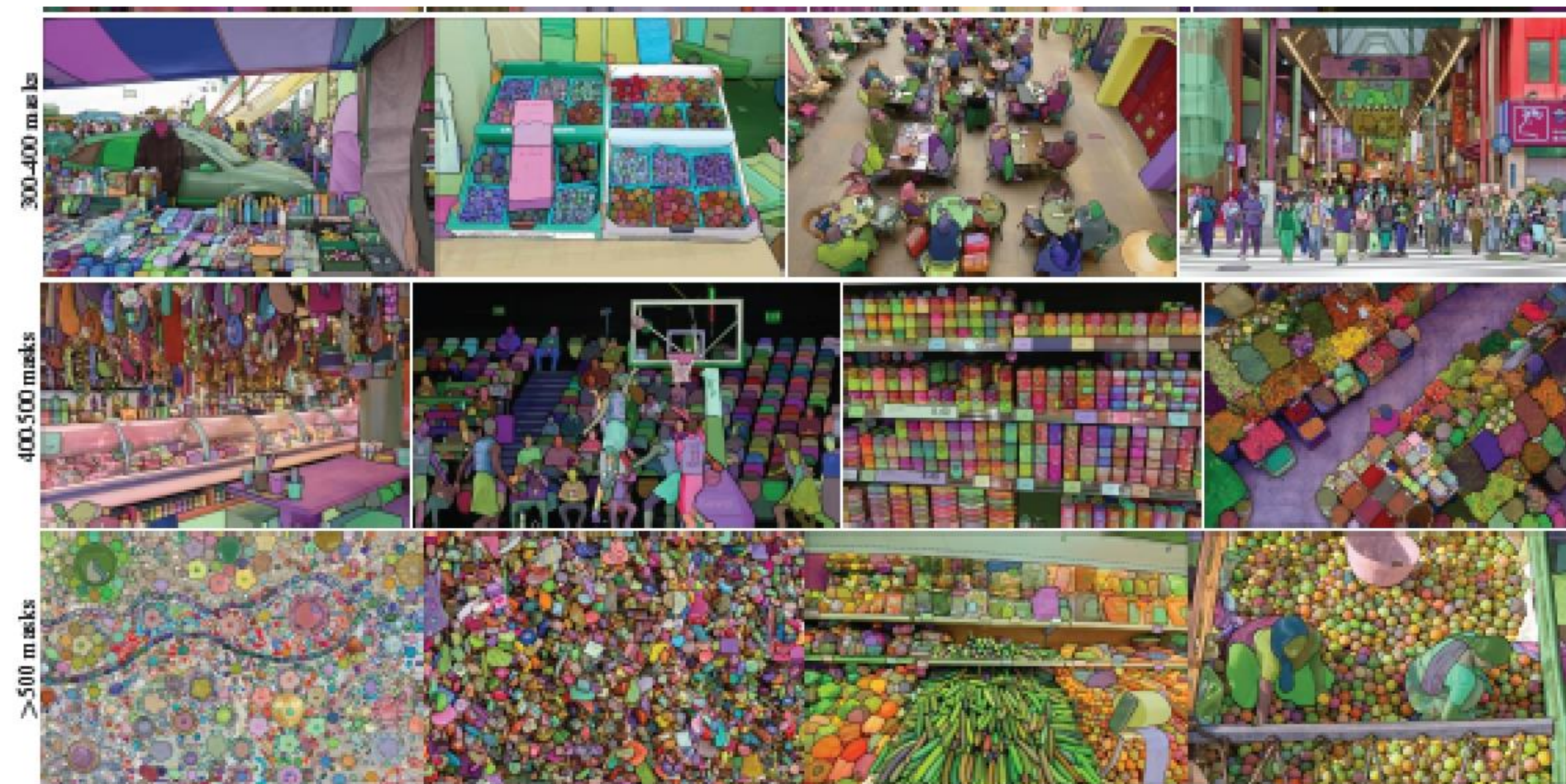
Prompt-based segmentation

Segment Anything Model(SAM)



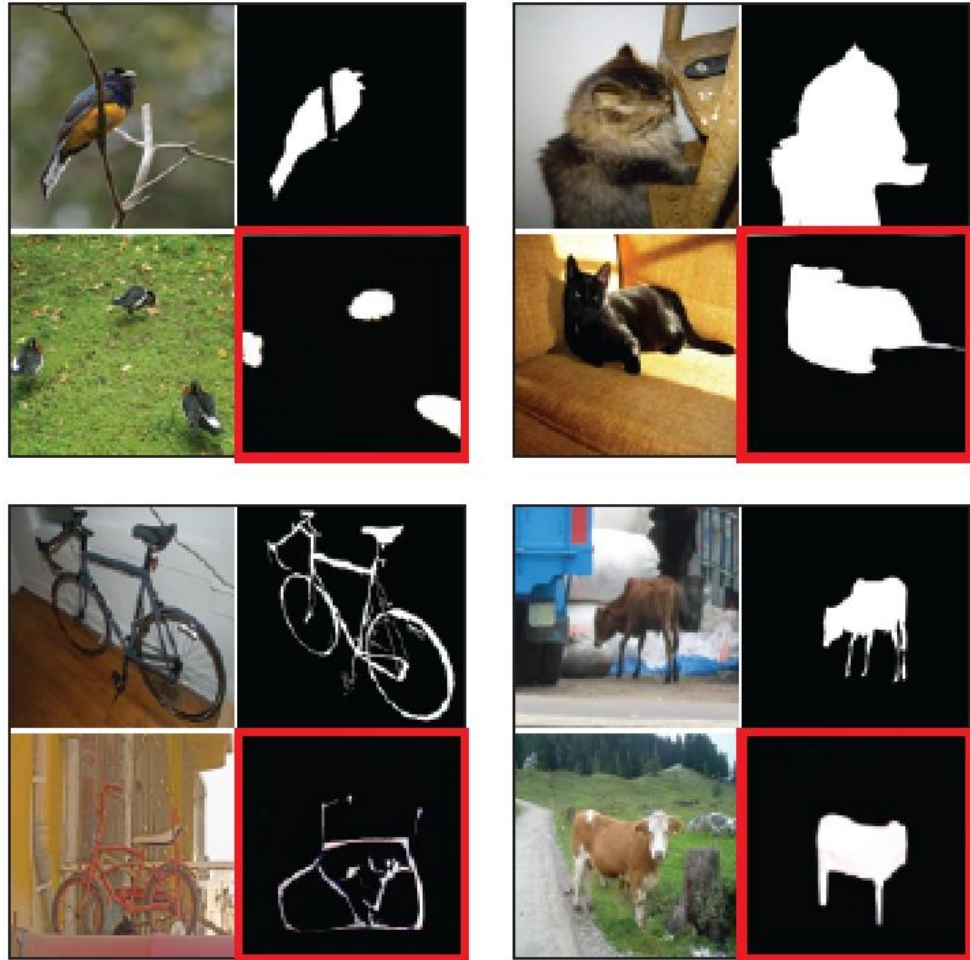
Prompt-based segmentation

Segment Anything Model(SAM)

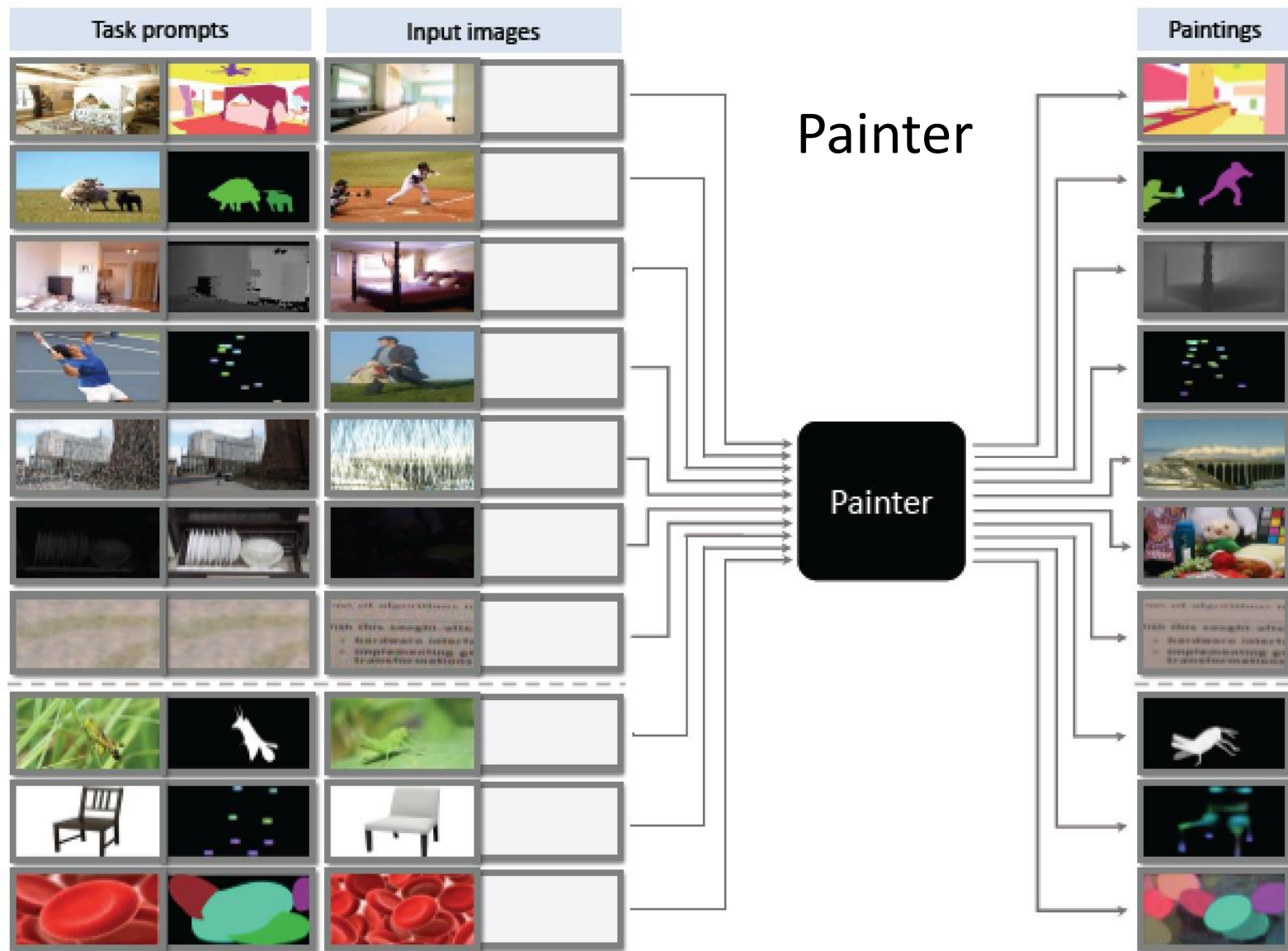


Prompt-based segmentation

Vision Prompter



Prompt-based segmentation



outline

- Semantic Segmentation
- Instance Segmentation
- Panoptic Segmentation
- Prompt based Segmentation
- Image Translation---Dense prediction

Image-to-image translation problems

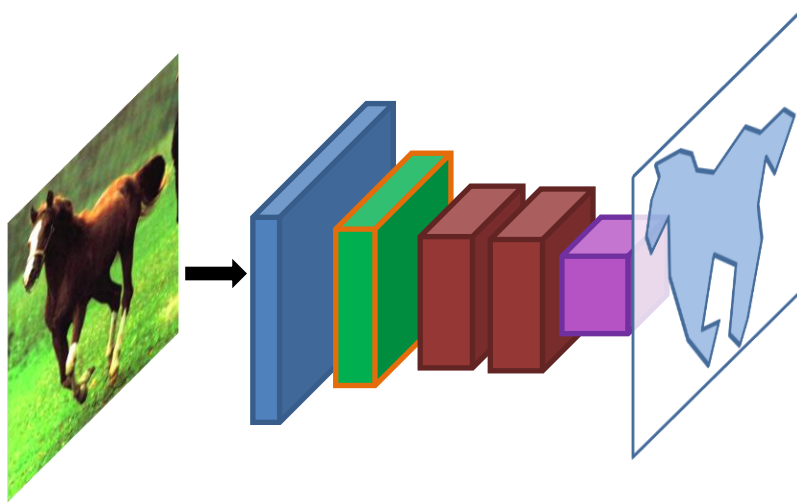
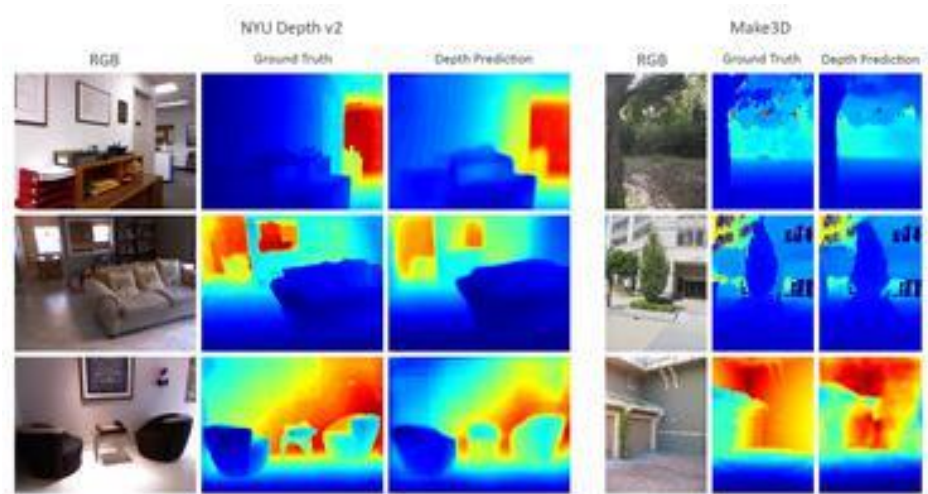


Image-to-image translation problems

- Segmentation
- Optical flow estimation
- Depth estimation
- Normal estimation
- Boundary detection
- ...



谢谢！

