

视听觉信息理解

--视觉部分

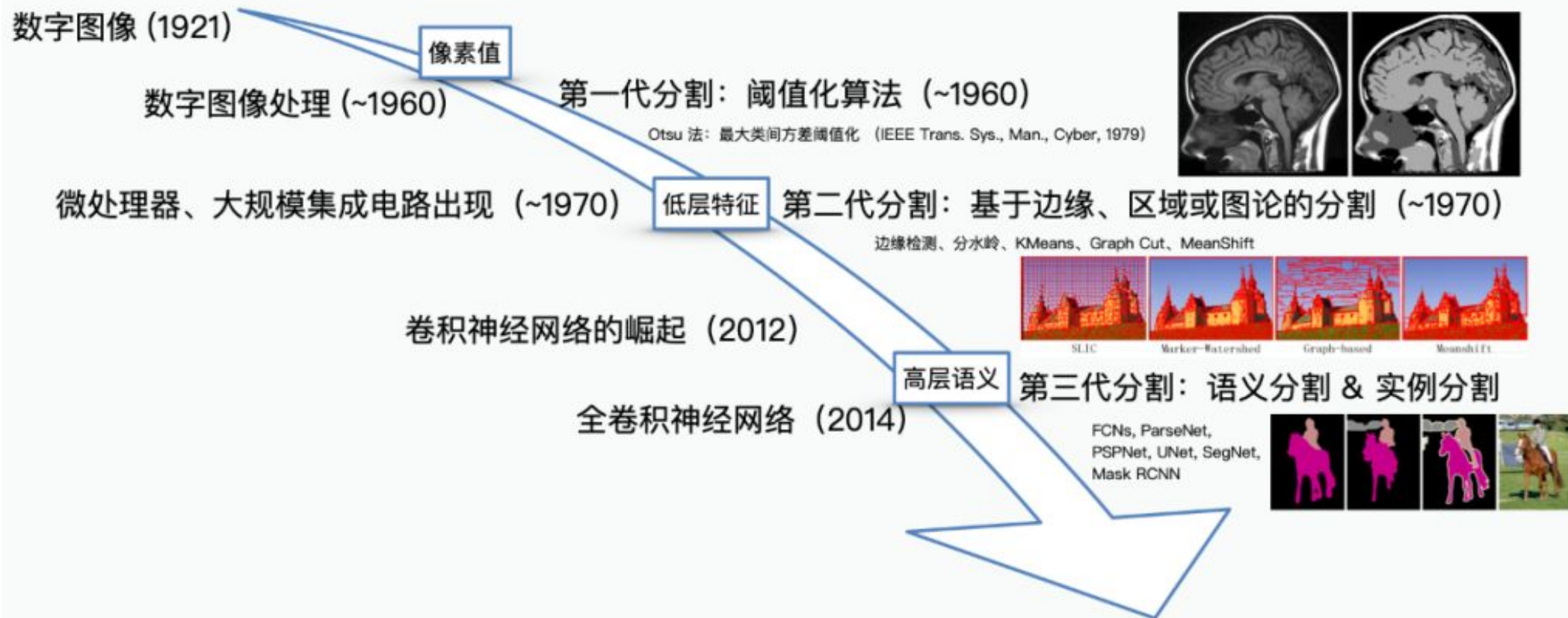
Segmentation

任冬伟

内容

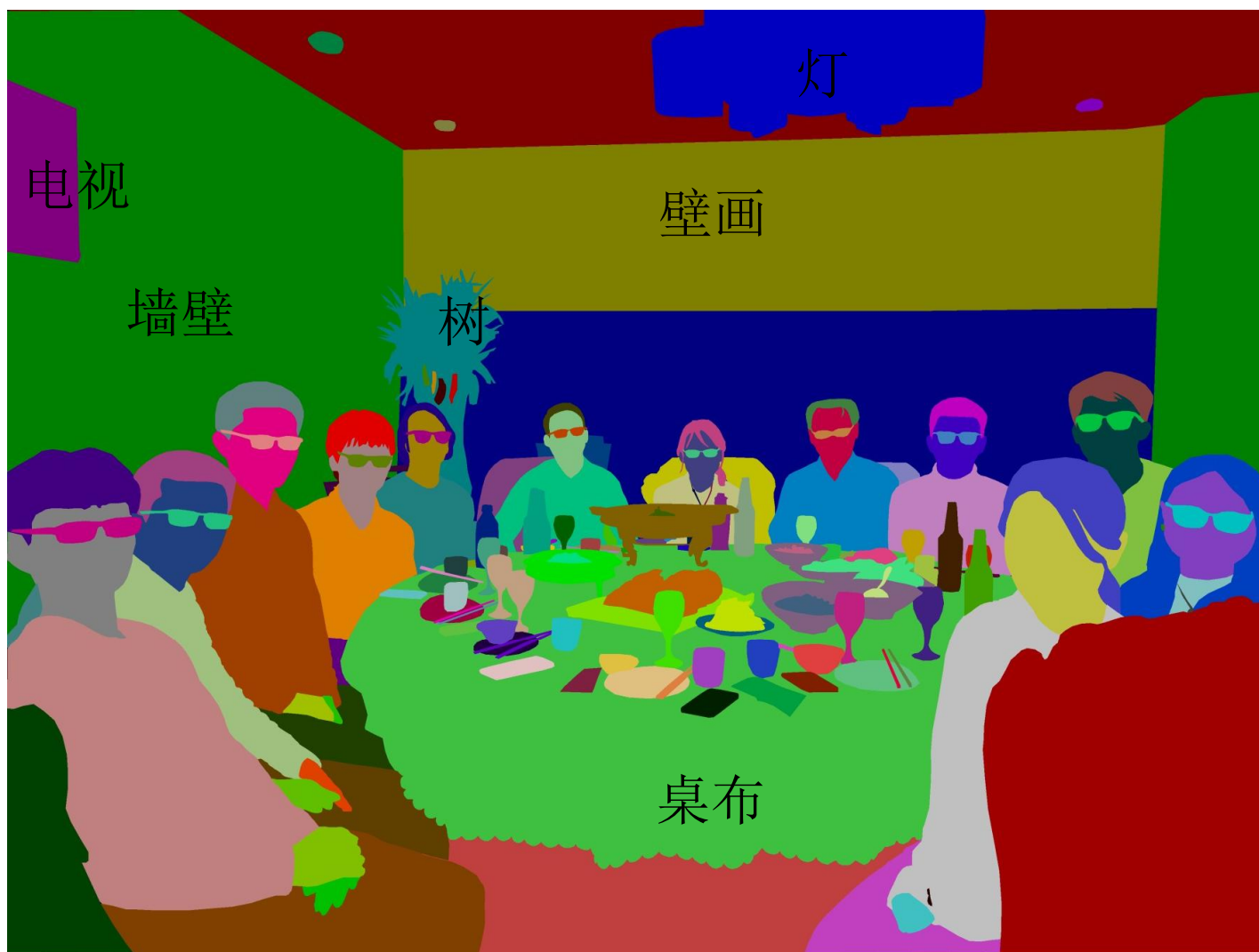
- 什么是视觉分割
- 语义分割网络
 - 使用特征融合得到语义信息的分割网络
 - 全卷积分割网络(DeepLab系列)
- 语义分割的应用

视觉分割技术发展历程





小组生日聚餐合照



可能会关注到整个场景，这是一个餐厅包间



更会注意到其中的人

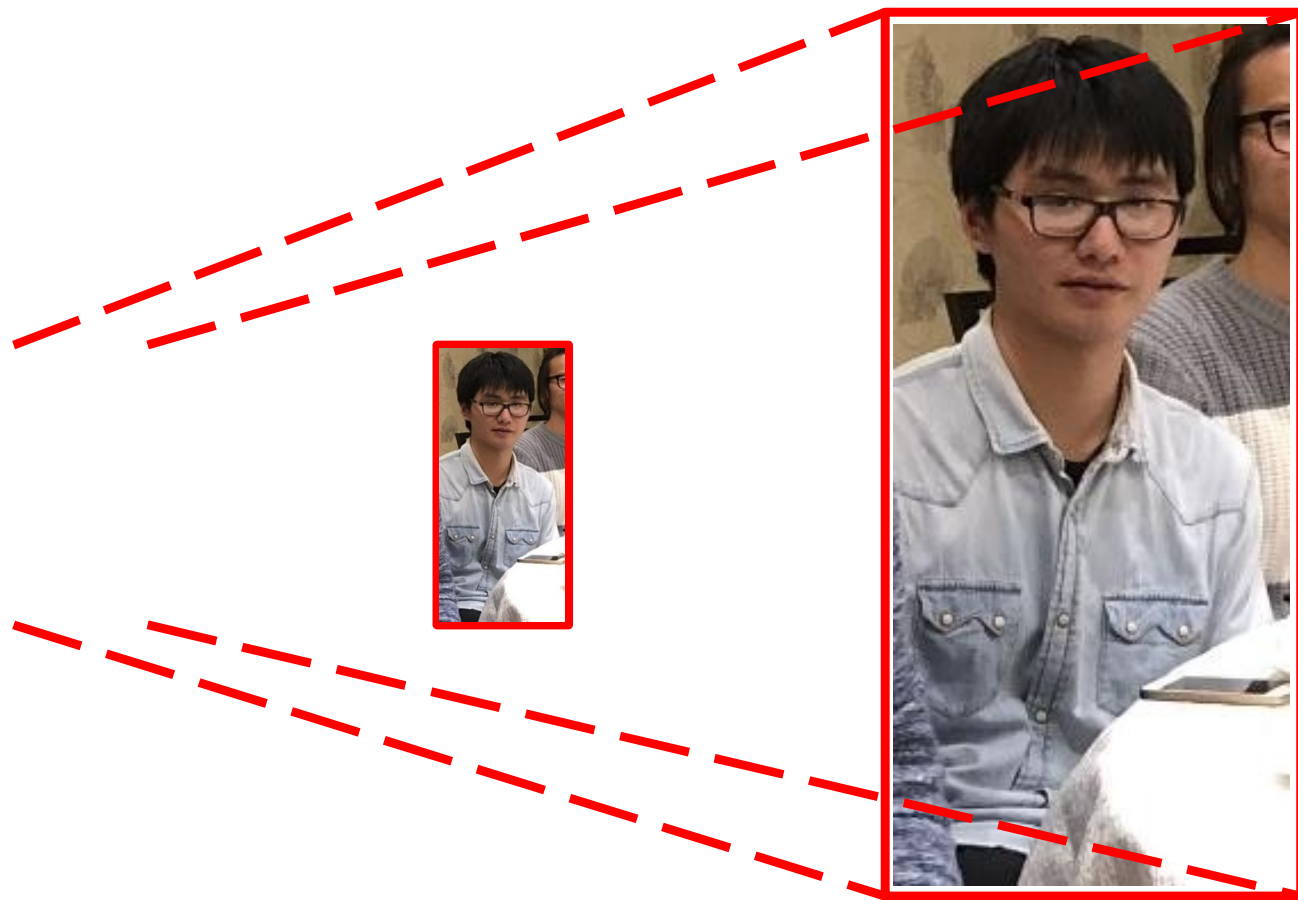


比如，我们会注意到寿星同学



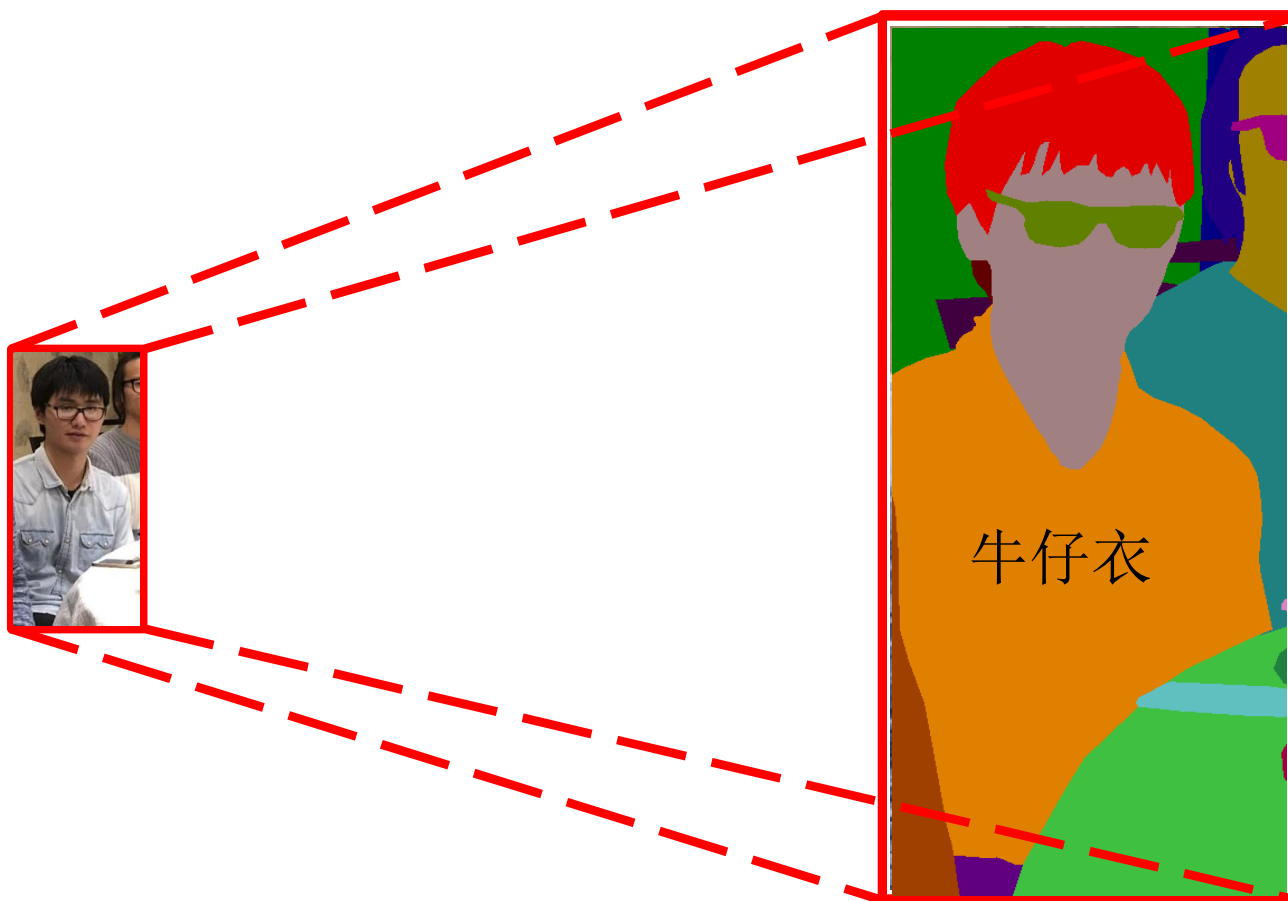
聚焦该寿星同学

视觉分割



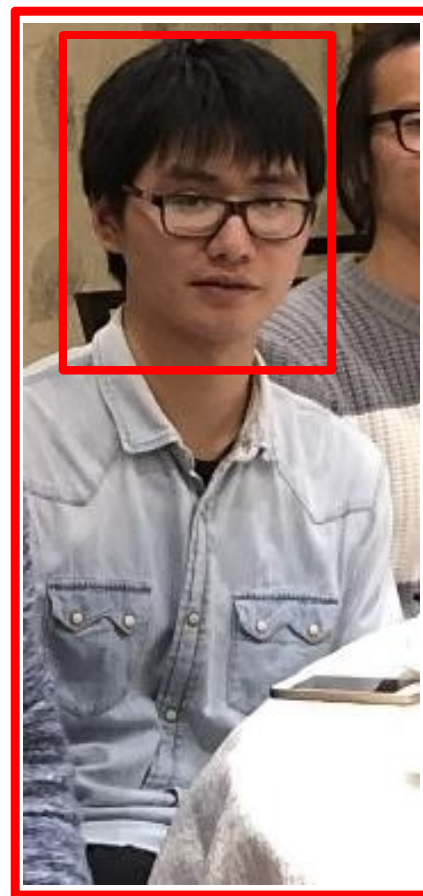
聚焦该寿星同学

视觉分割



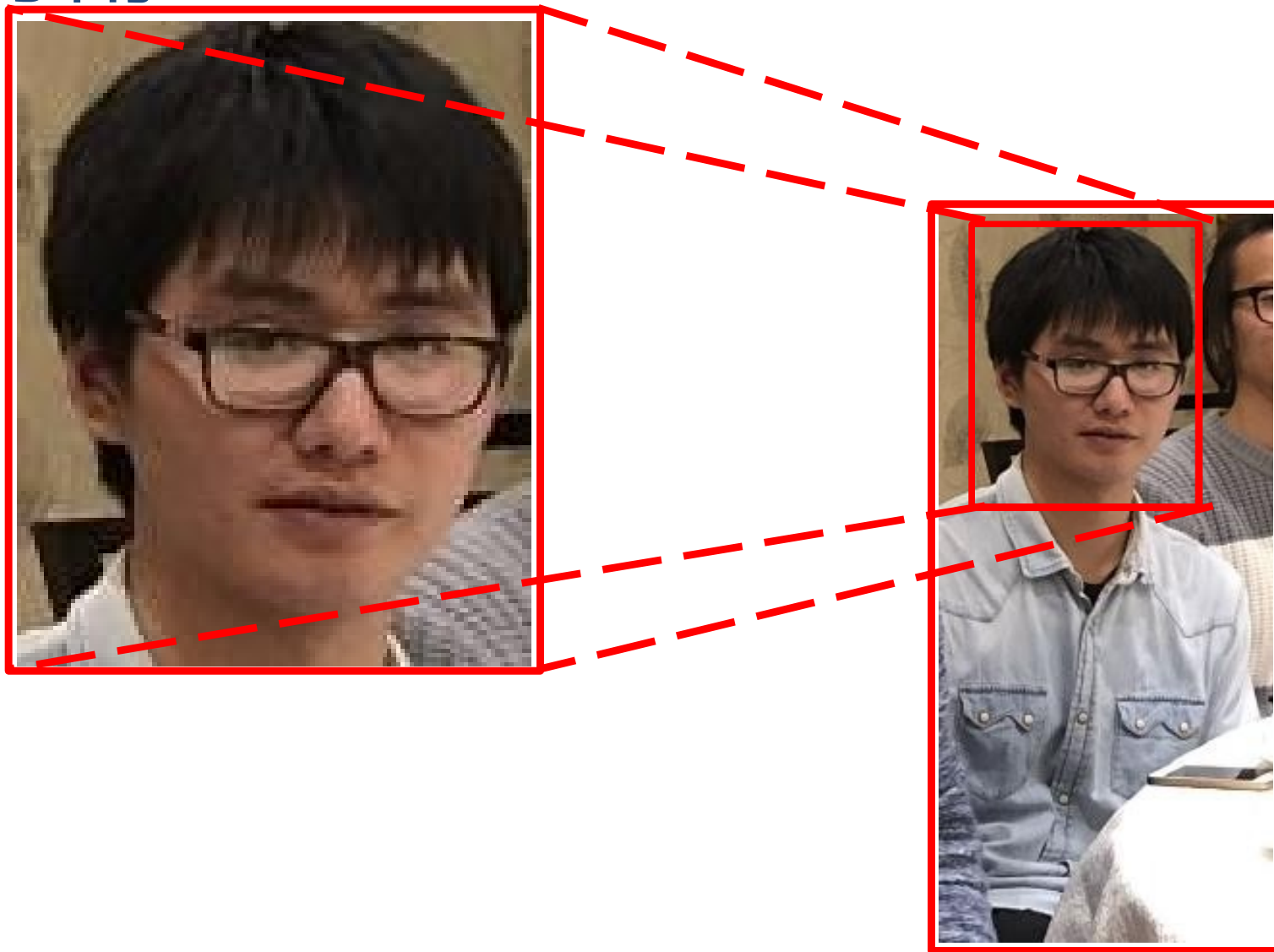
关注到寿星的服饰(淡蓝色牛仔衣)、
状态 (坐在餐桌旁) 等

视觉分割



进一步，会关注到寿星的脸部特征

视觉分割



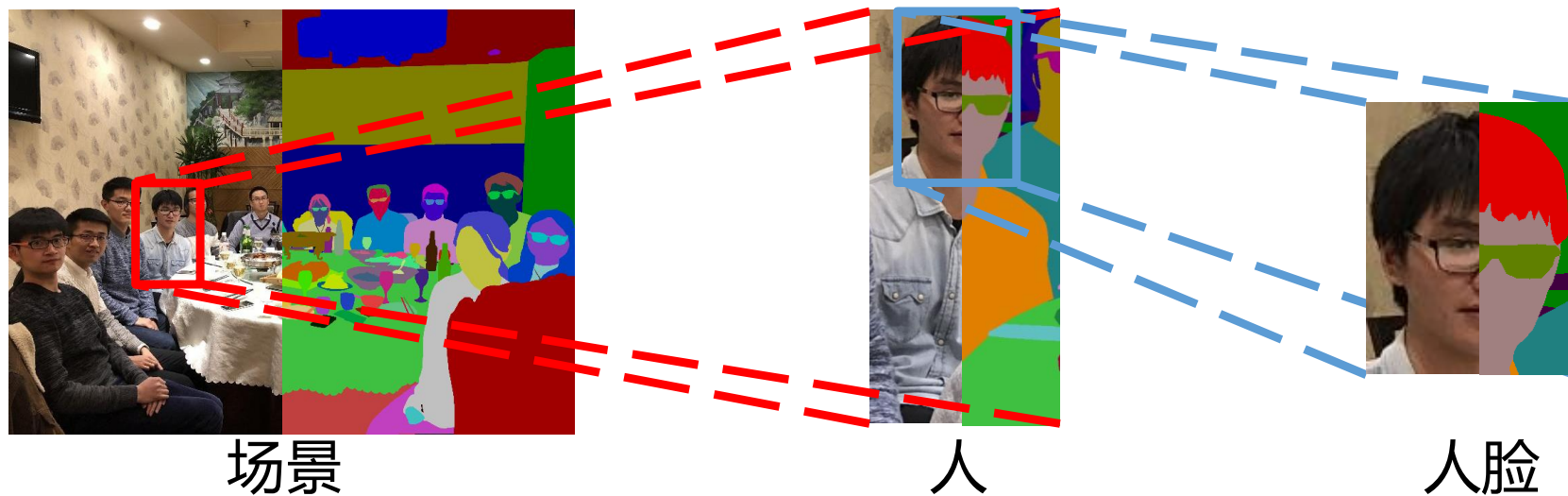
进一步，会关注到寿星的脸部特征

视觉分割



寿星是短发、戴眼镜

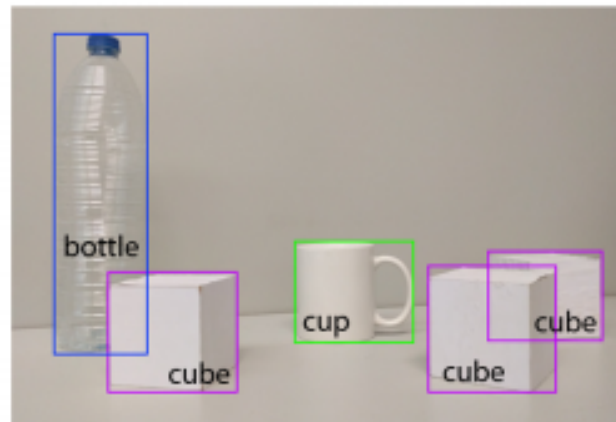
视觉分割



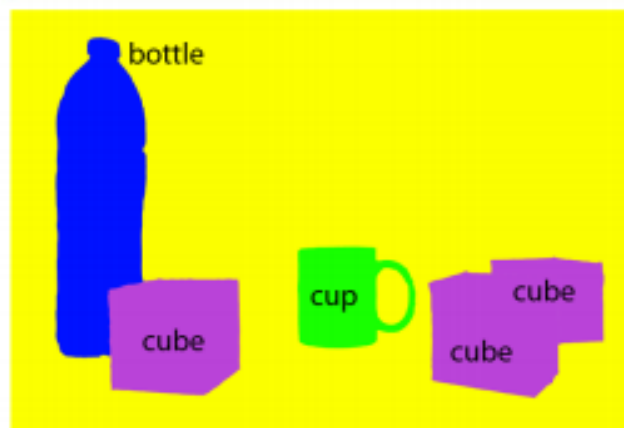
视觉分割



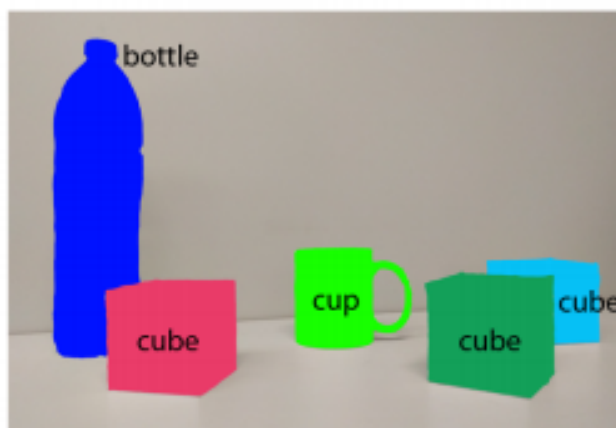
分类



检测



语义分割



实例分割

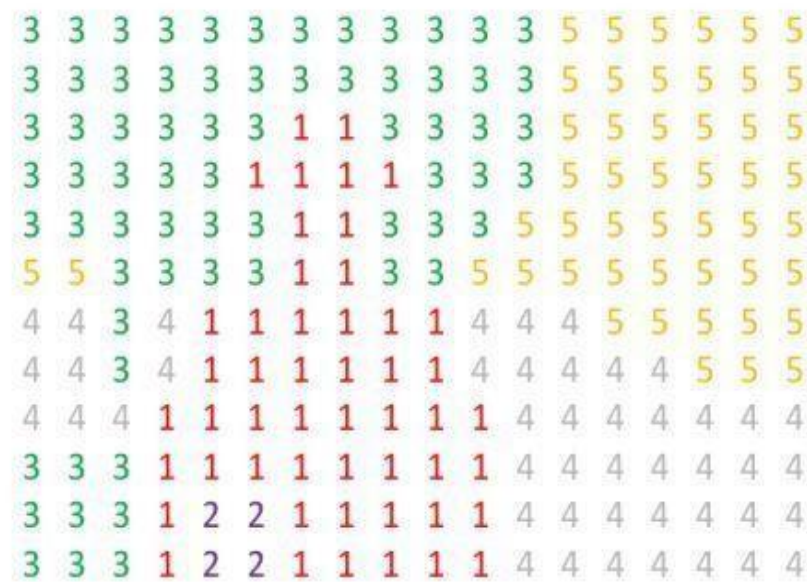
语义分割



Input



- 1: Person
- 2: Purse
- 3: Plants/Grass
- 4: Sidewalk
- 5: Building/Structures



Semantic Labels

语义分割



0: Background/Unknown

1: Person

2: Purse

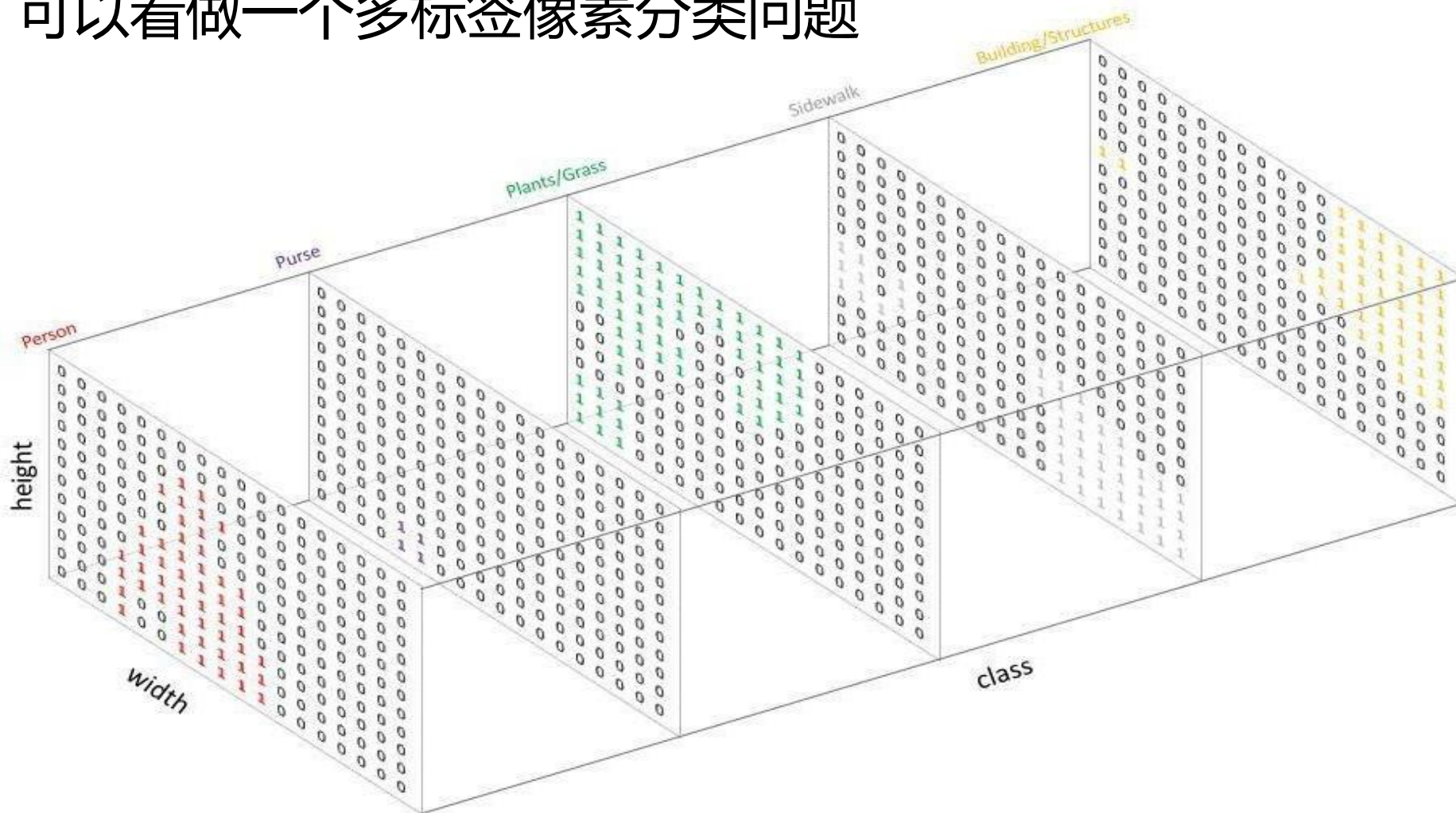
3: Plants/Grass

4: Sidewalk

5: Building/Structures

语义分割

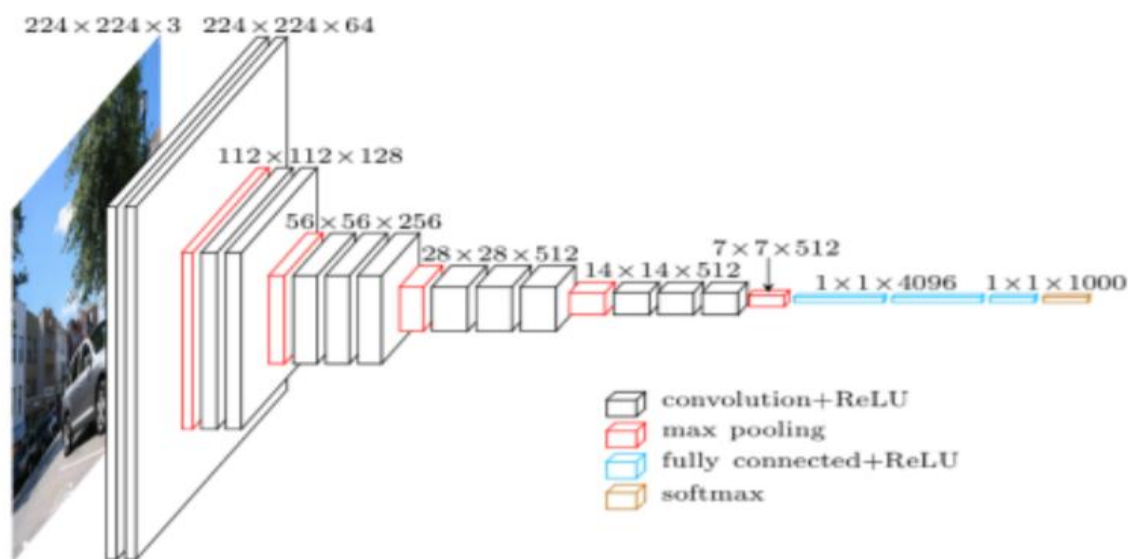
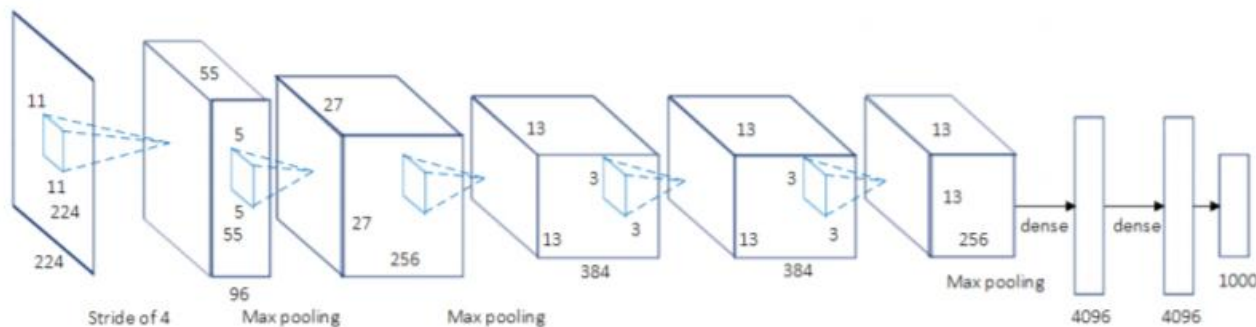
可以看做一个多标签像素分类问题



内容

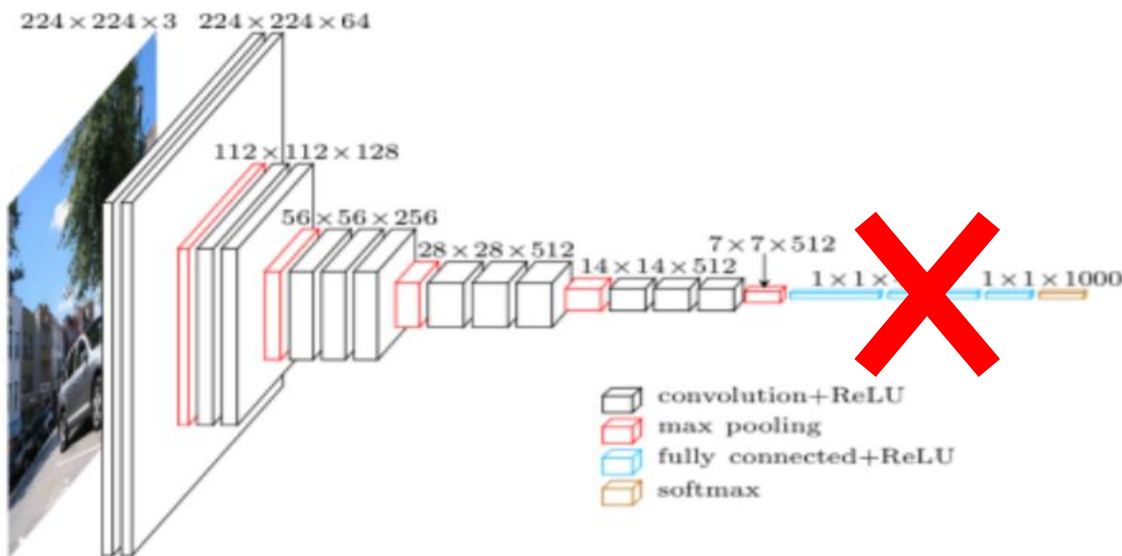
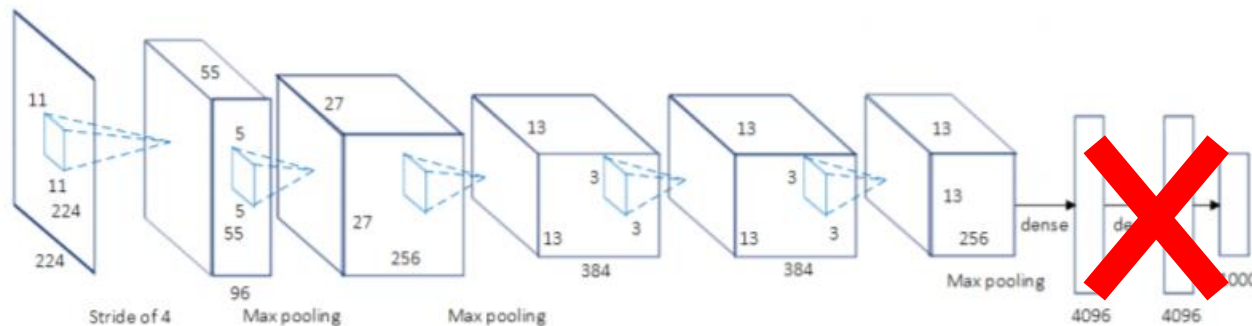
- 什么是视觉分割
- 语义分割网络
 - 使用特征融合得到语义信息的分割网络
 - 全卷积分割网络(DeepLab系列)
- 语义分割的应用

基于深度学习提取特征的方法



Segmentation?

基于深度学习提取特征的方法



Segmentation?

基于深度学习提取特征的方法

使用CNN进行语义分割的问题：

- 深层特征包含语义信息丰富，更适合做分类任务，而不适合做dense prediction的分割任务
- 多尺度目标的存在会增加语义分割任务中类别分类的困难

基于深度学习提取特征的方法

- 深层特征包含语义信息丰富，更适合做分类任务，而不适用dense prediction的分割任务

Encoder-Decoder
UNet



深层与浅层特征融合

- 分割对象尺度不同，多尺度问题增加分割困难

Atrous Spatial Pyramid Pooling
(Deeplab)

Pyramid Pooling Module
(PSPNet)



多尺度池化，融合多尺度的特征信息和上下文信息

语义分割网络

使用特征融合得到语义信息丰富且分辨率高的特征的分割网络

- 典型架构

- ✓ FCN
- ✓ SegNet
- ✓ UNet

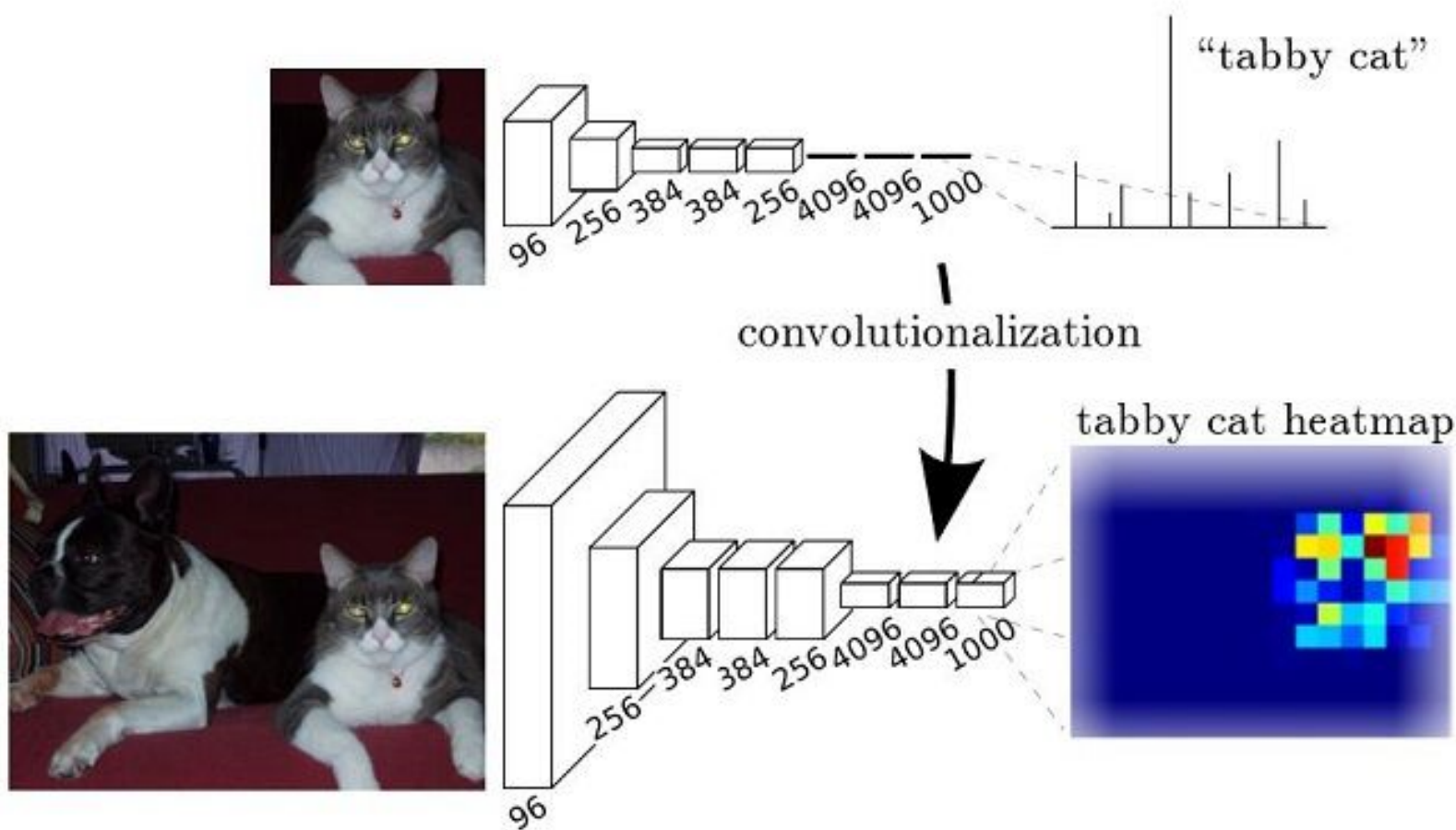
使用全卷积网络以增大感受野并结合上下文信息的分割网络

- 典型架构

- ✓ DeepLab V1
- ✓ DeepLab V2
- ✓ DeepLab V3
- ✓ DeepLab V3+

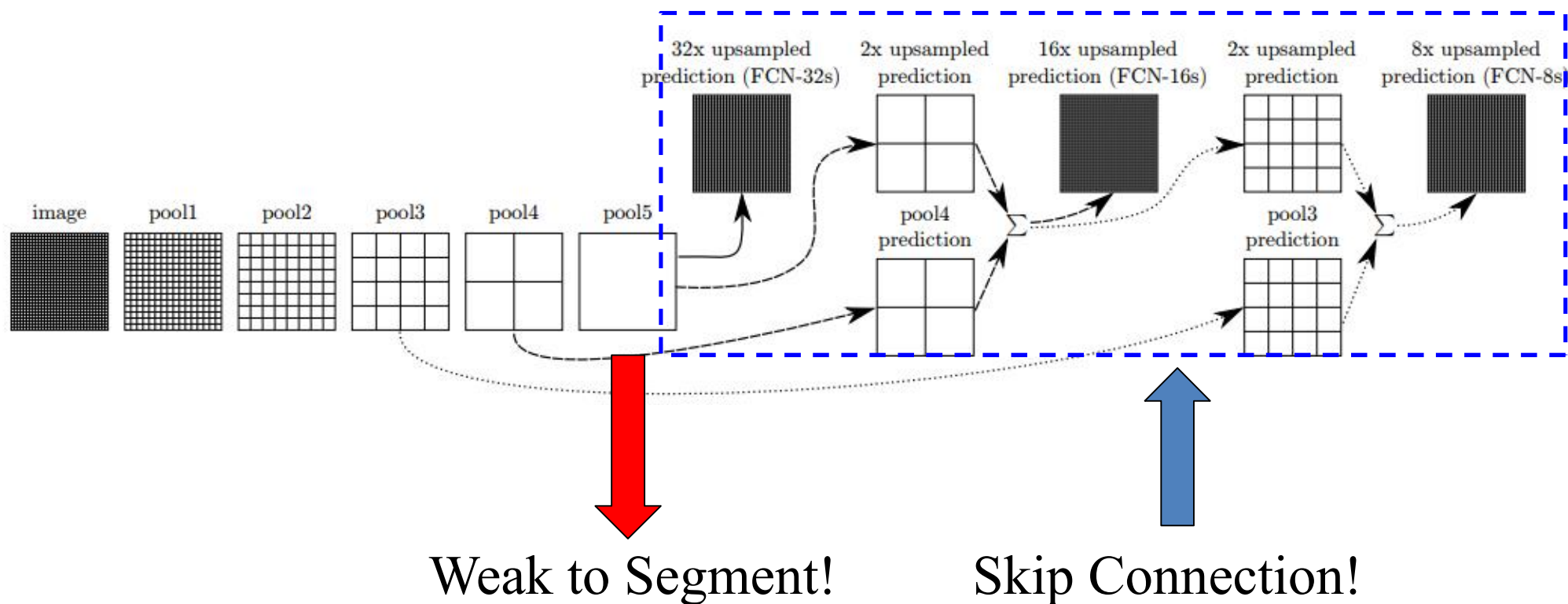
使用特征融合得到语义信息的分割网络

分类网络中的语义特征



使用特征融合得到语义信息的分割网络

FCN (Fully Convolutional Networks)



语义分割网络

使用特征融合得到语义信息丰富且分辨率高的特征的分割网络

- 典型架构

- ✓ FCN
- ✓ SegNet
- ✓ UNet

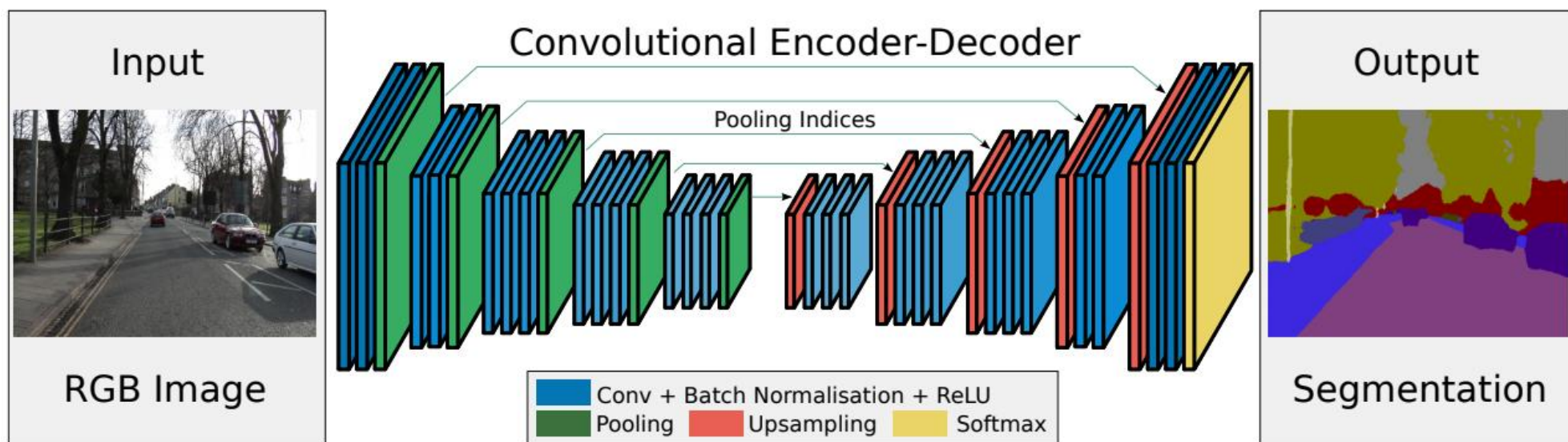
使用全卷积网络以增大感受野并结合上下文信息的分割网络

- 典型架构

- ✓ DeepLab V1
- ✓ DeepLab V2
- ✓ DeepLab V3
- ✓ DeepLab V3+

使用特征融合得到语义信息的分割网络

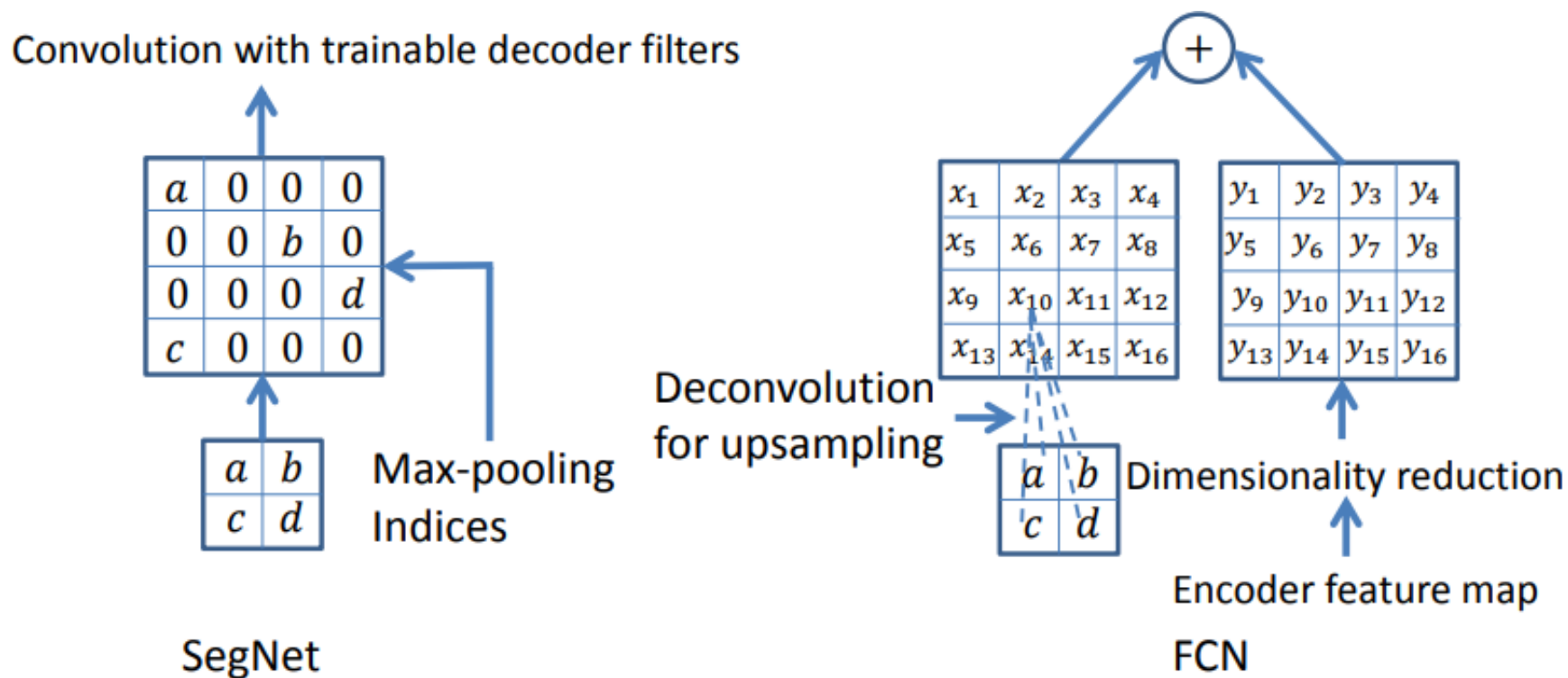
SegNet



Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.

使用特征融合得到语义信息的分割网络

SegNet



语义分割网络

使用特征融合得到语义信息丰富且分辨率高的特征的分割网络

- 典型架构

- ✓ FCN
- ✓ SegNet
- ✓ UNet

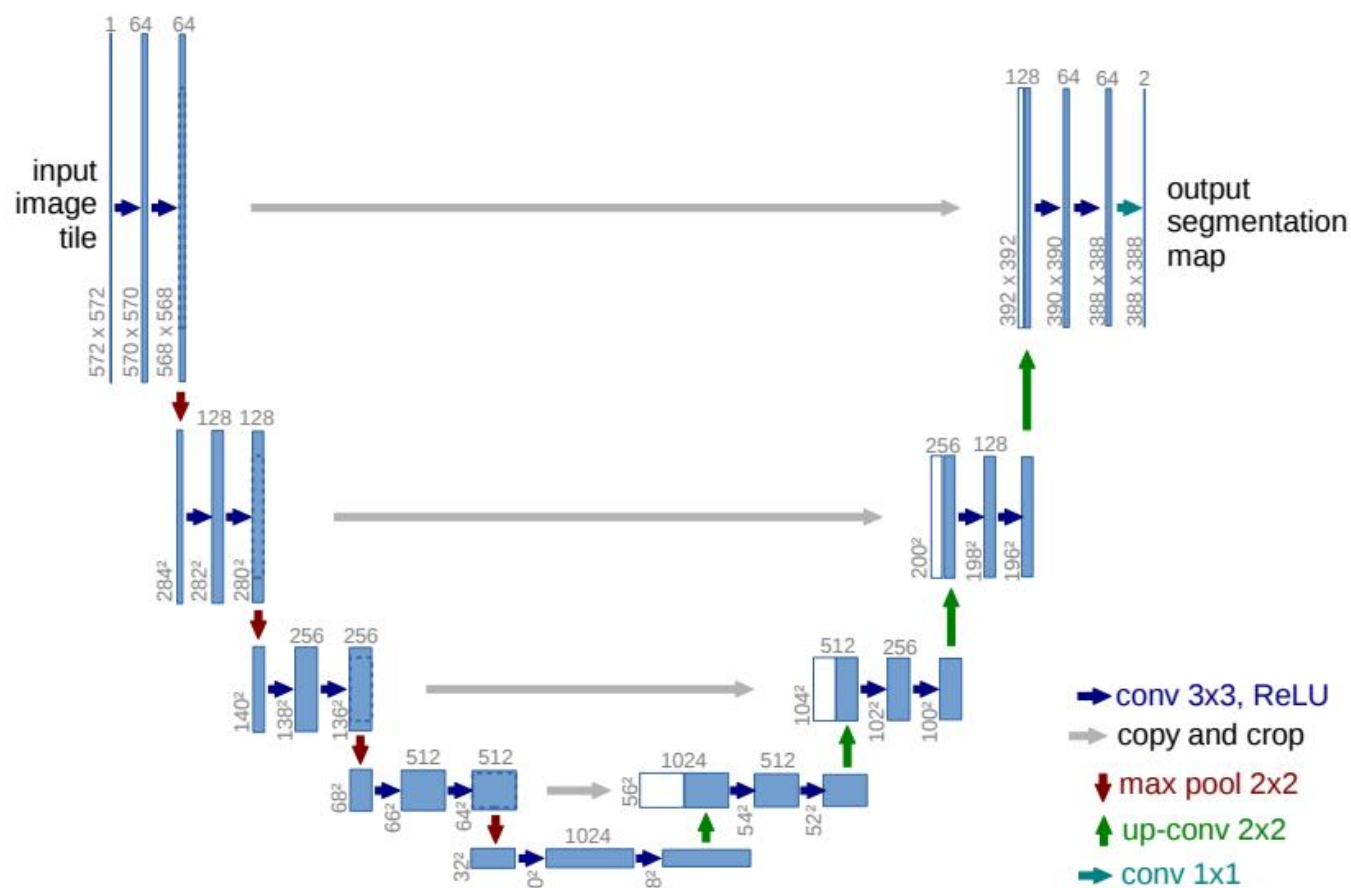
使用全卷积网络以增大感受野并结合上下文信息的分割网络

- 典型架构

- ✓ DeepLab V1
- ✓ DeepLab V2
- ✓ DeepLab V3
- ✓ DeepLab V3+

使用特征融合得到语义信息的分割网络

UNet



Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

内容

- 什么是视觉分割
- 语义分割网络
 - 使用特征融合得到语义信息的分割网络
 - 全卷积分割网络(DeepLab系列)
- 语义分割的应用

语义分割网络

使用特征融合得到语义信息丰富且分辨率高的特征的分割网络

- 典型架构

- ✓ FCN
- ✓ SegNet
- ✓ UNet

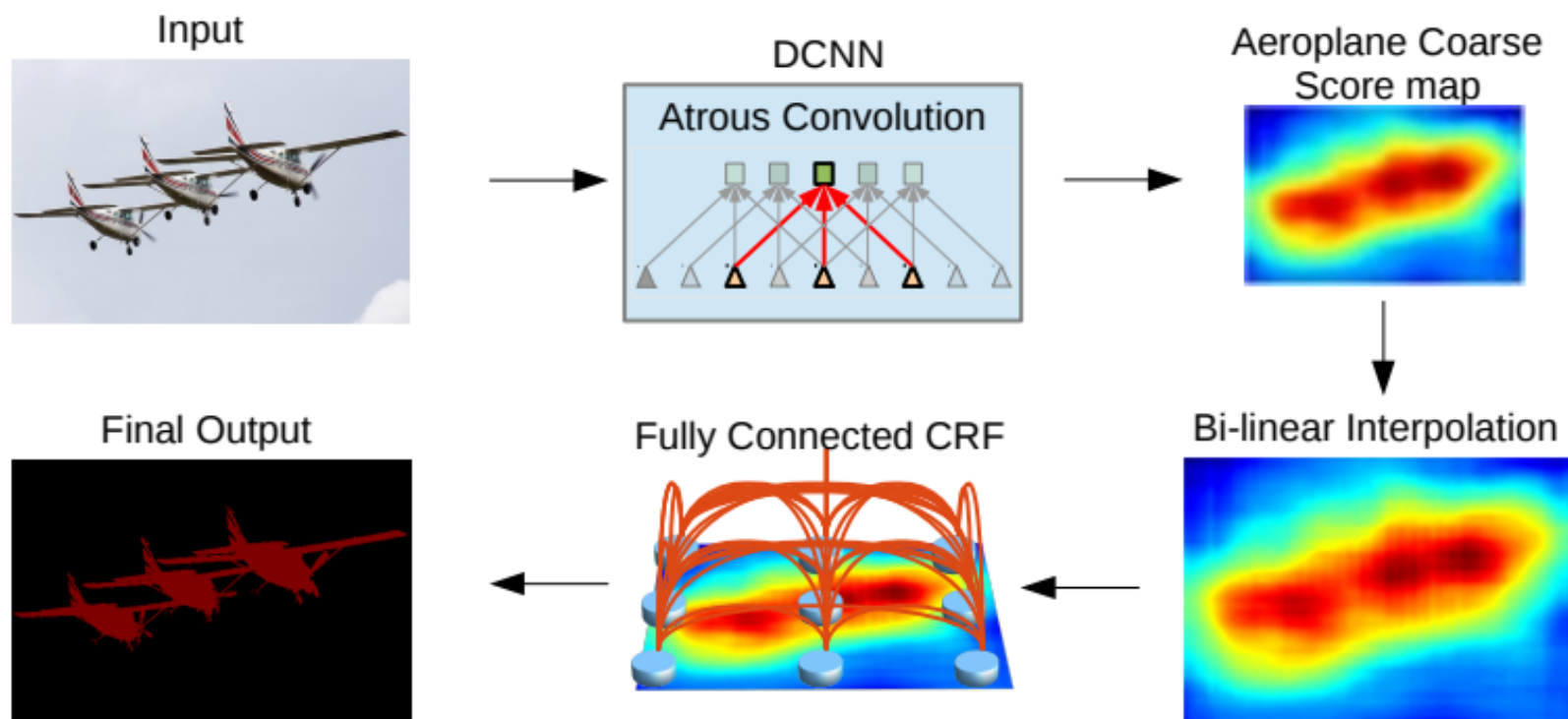
使用全卷积网络以增大感受野并结合上下文信息的分割网络

- 典型架构

- ✓ DeepLab V1
- ✓ DeepLab V2
- ✓ DeepLab V3
- ✓ DeepLab V3+

全卷积网络增大感受野

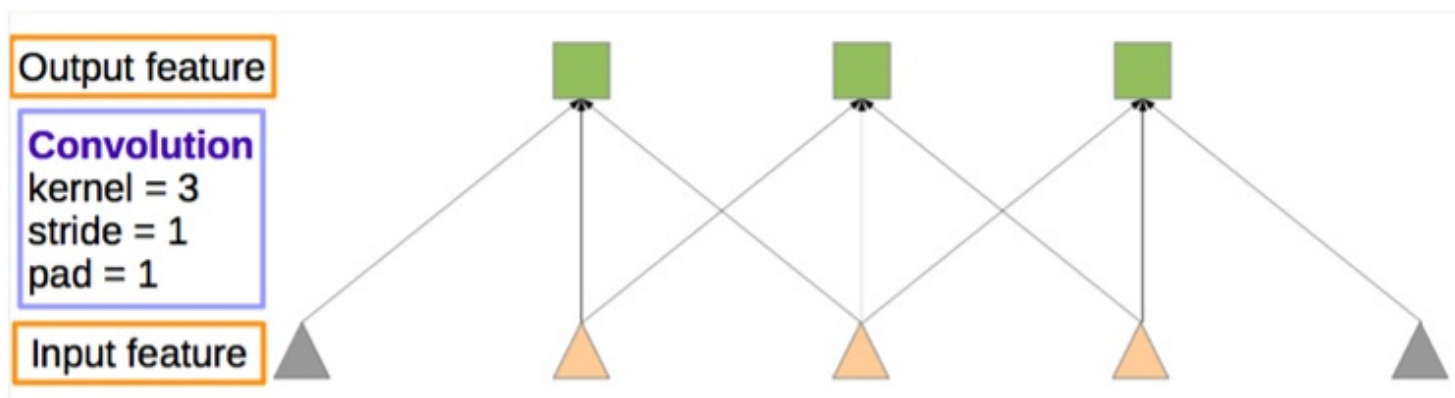
DeepLab v1



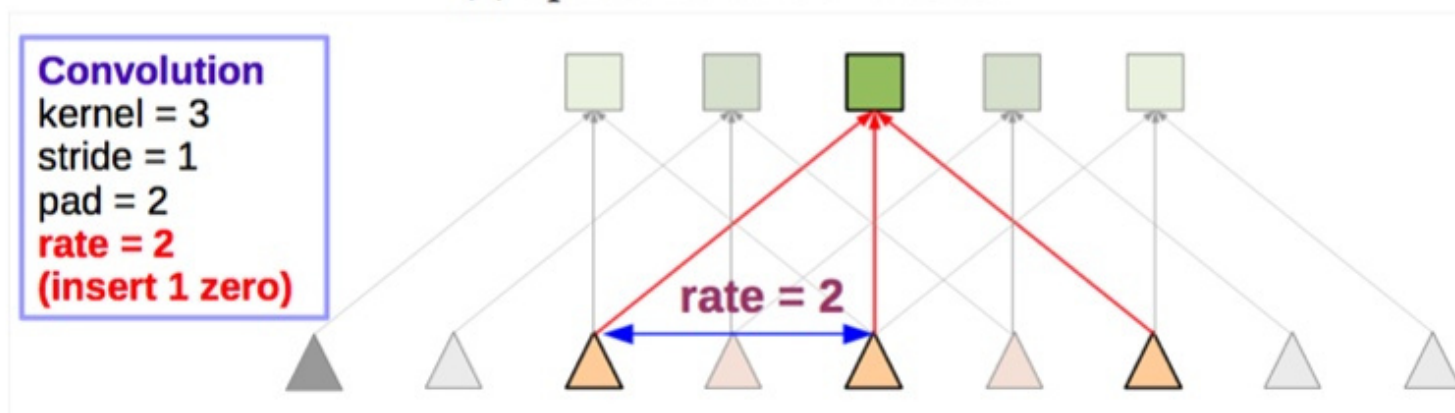
Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv preprint arXiv:1412.7062, 2014.

DeepLab v1

空洞卷积



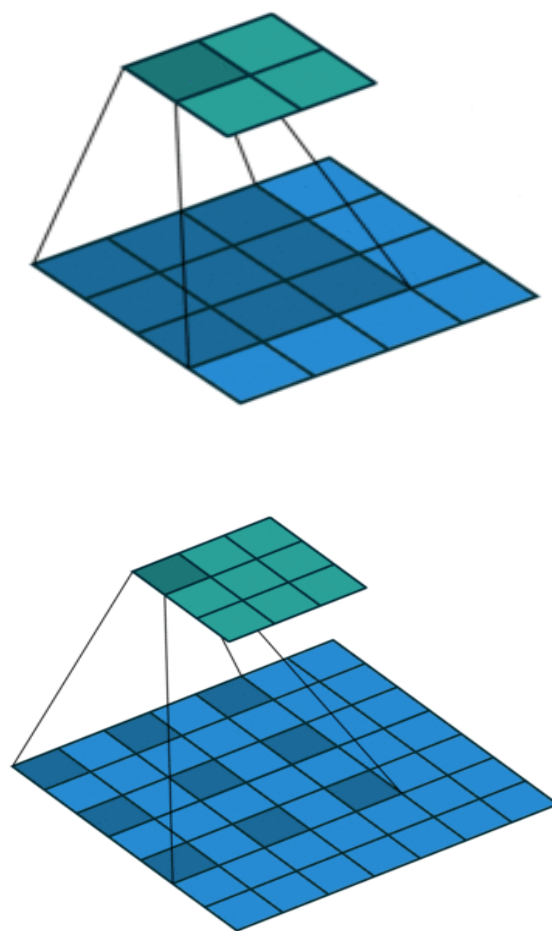
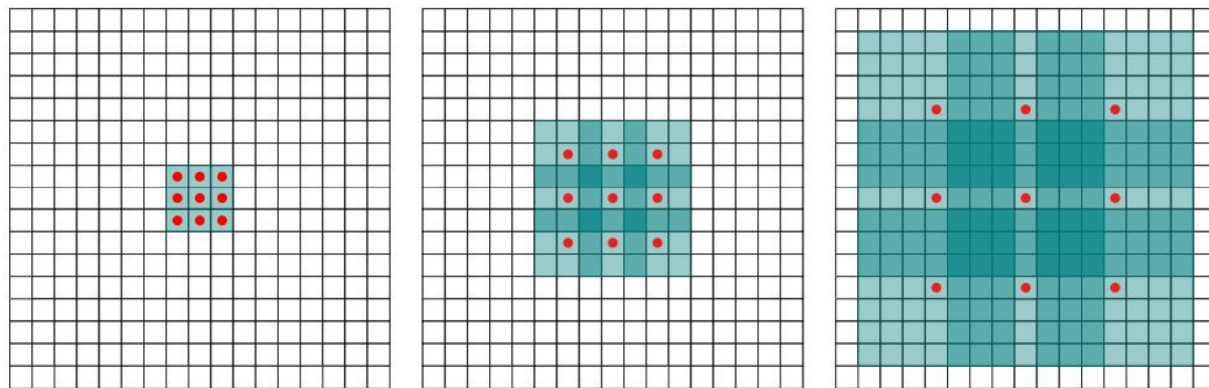
(a) Sparse feature extraction



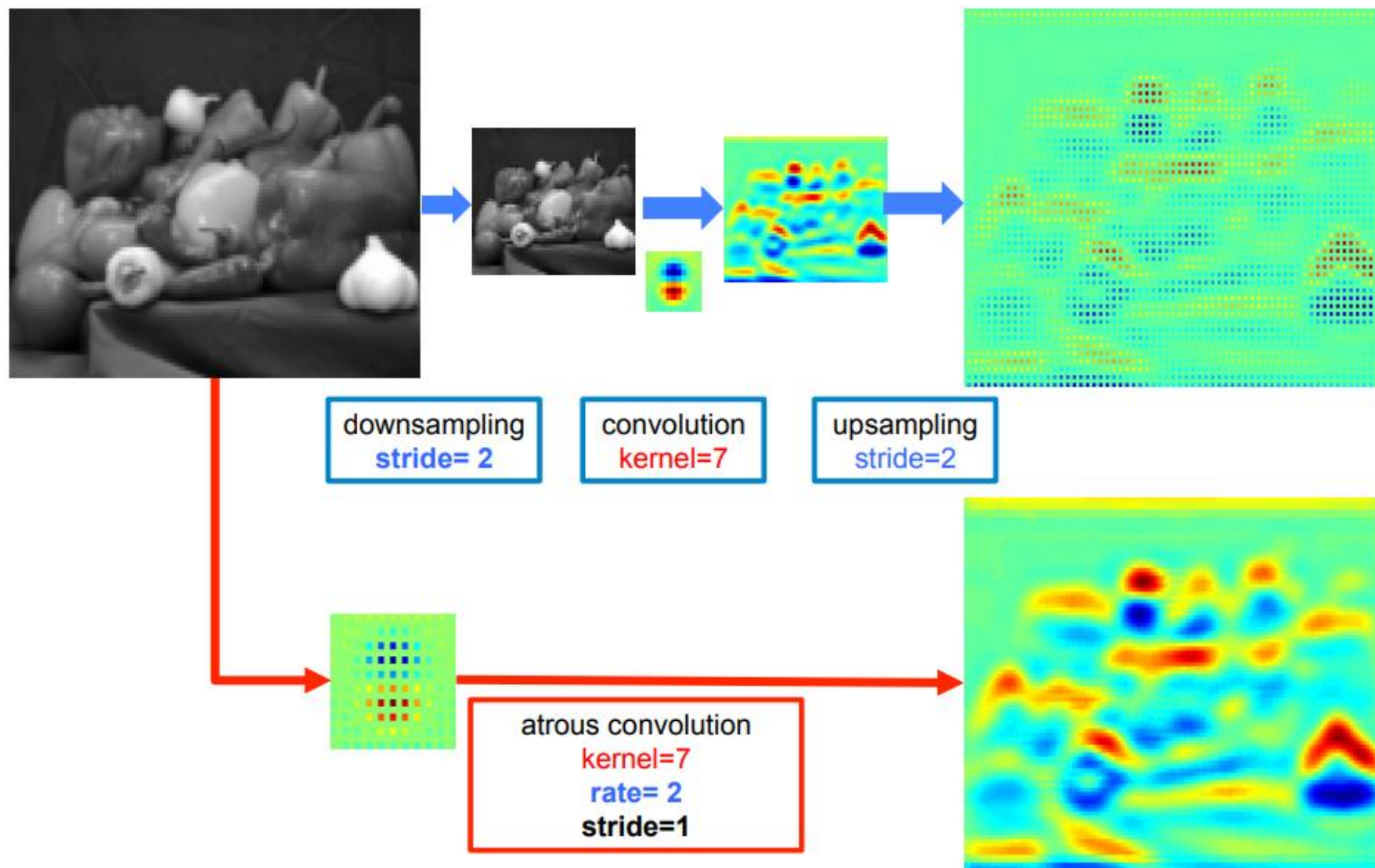
(b) Dense feature extraction

DeepLab v1

空洞卷积



空洞卷积



语义分割网络

使用特征融合得到语义信息丰富且分辨率高的特征的分割网络

- 典型架构

- ✓ FCN
- ✓ SegNet
- ✓ UNet

使用全卷积网络以增大感受野并结合上下文信息的分割网络

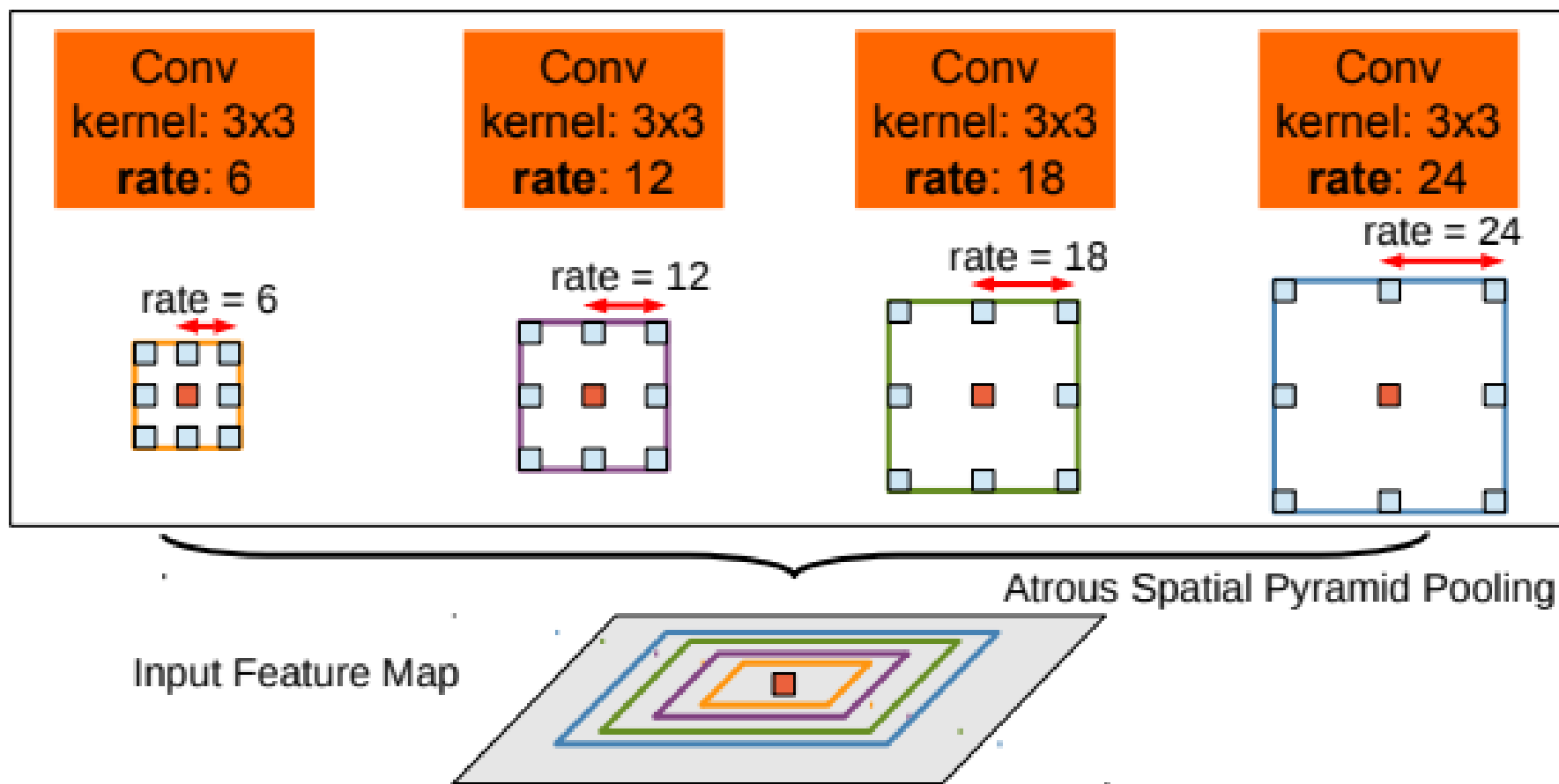
- 典型架构

- ✓ DeepLab V1
- ✓ DeepLab V2
- ✓ DeepLab V3
- ✓ DeepLab V3+

全卷积网络增大感受野

DeepLab v2

多尺度的空洞卷积 ASPP



语义分割网络

使用特征融合得到语义信息丰富且分辨率高的特征的分割网络

- 典型架构

- ✓ FCN
- ✓ SegNet
- ✓ UNet

使用全卷积网络以增大感受野并结合上下文信息的分割网络

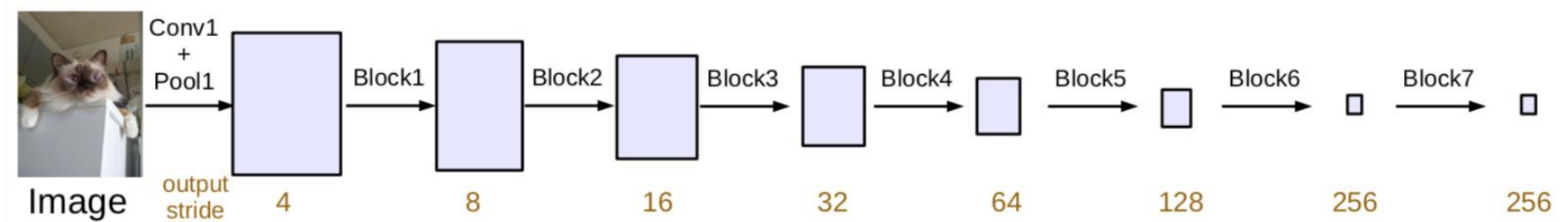
- 典型架构

- ✓ DeepLab V1
- ✓ DeepLab V2
- ✓ DeepLab V3
- ✓ DeepLab V3+

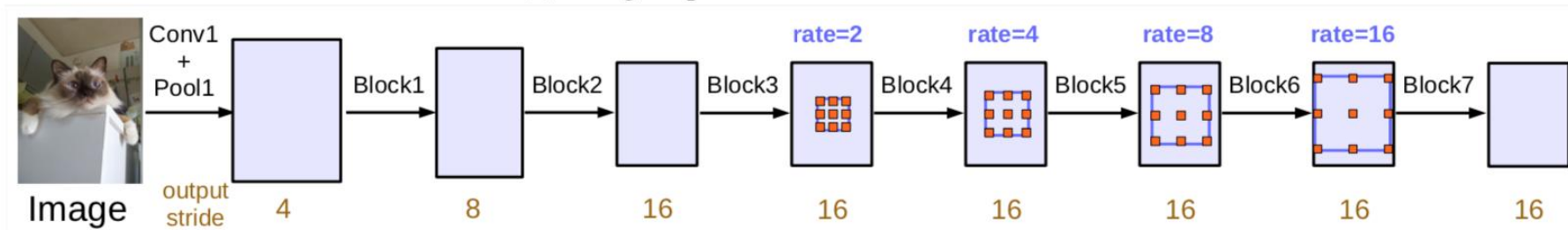
全卷积网络增大感受野

DeepLab v3

串行结构



(a) Going deeper without atrous convolution.

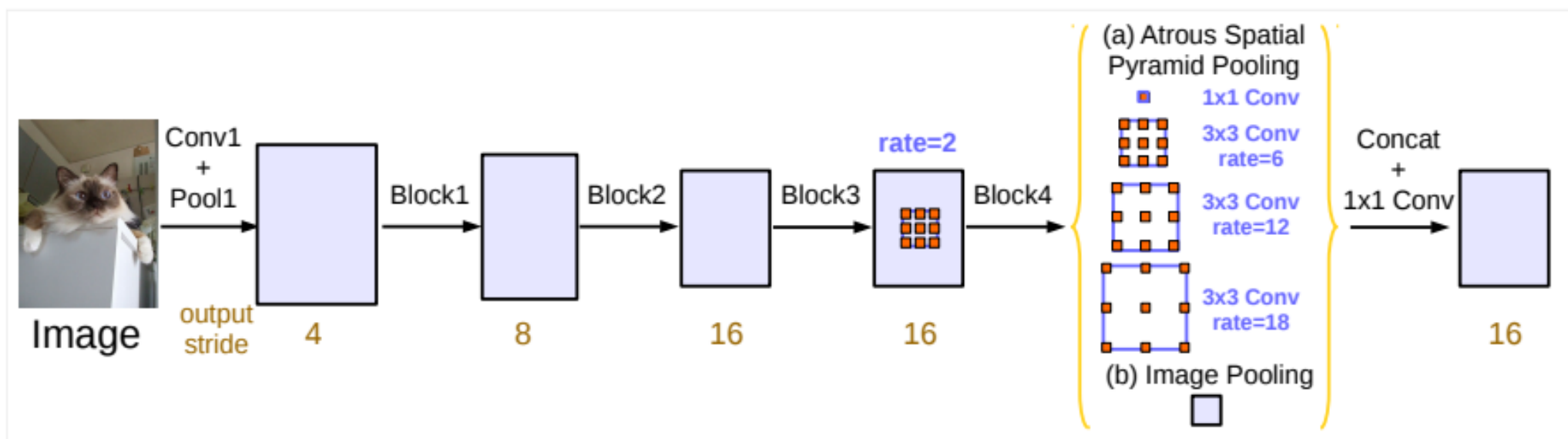


(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output_stride = 16$.

全卷积网络增大感受野

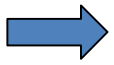
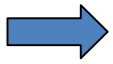
DeepLab v3

并行结构



全卷积网络增大感受野

DeepLab v3

- Deeplab v3设计多比例的空洞卷积级联来捕获多尺度背景；
- 改进ASPP模块，用于探索多尺度卷积特征，将全局背景基于图像层次进行编码获得特征；  串行结构
- 条件随机场（CRF）被去除，模型简洁易懂。  并行结构

语义分割网络

使用特征融合得到语义信息丰富且分辨率高的特征的分割网络

- 典型架构

- ✓ FCN
- ✓ SegNet
- ✓ UNet

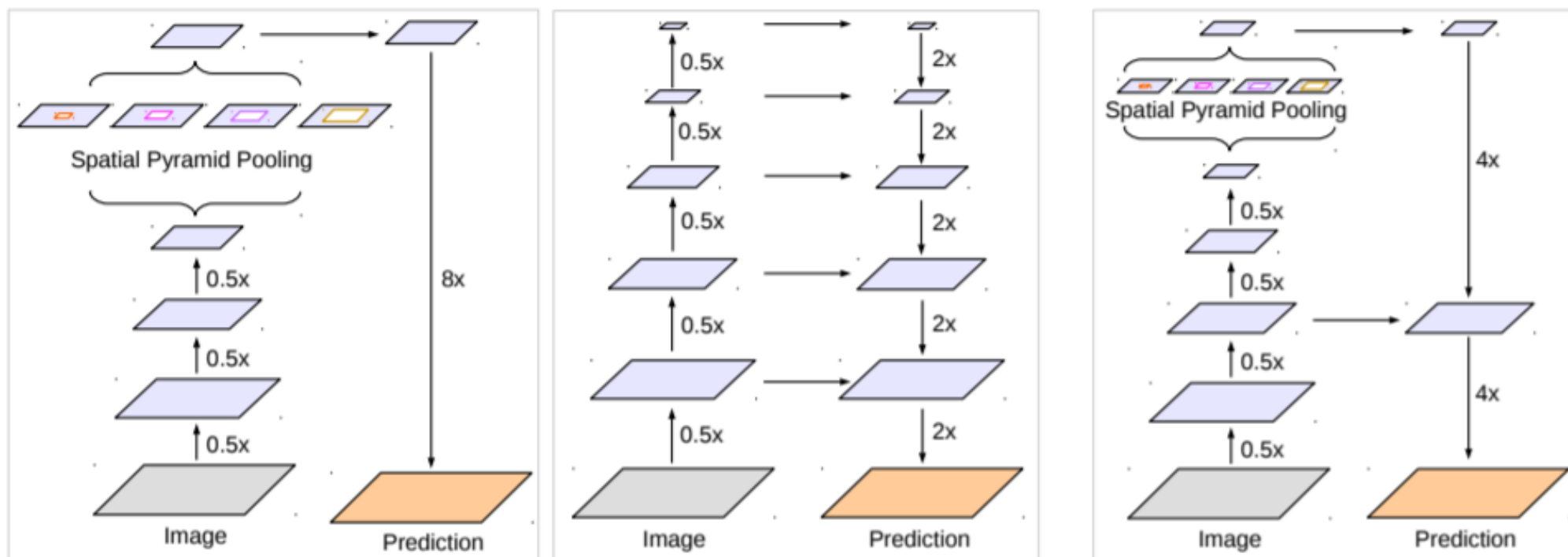
使用全卷积网络以增大感受野并结合上下文信息的分割网络

- 典型架构

- ✓ DeepLab V1
- ✓ DeepLab V2
- ✓ DeepLab V3
- ✓ DeepLab V3+

全卷积网络增大感受野

DeepLab v3+



(a) v3的纵式结构；(b)常见的编码-解码结构；(c) v3+提出的encode-decode结构

全卷积网络增大感受野

DeepLab v3+

- 将原DeepLabv3当作Encoder，添加Decoder得到新的模型（DeepLabv3+）；
- 把Depthwise separable convolution应用到Atrous Spatial Pyramid Pooling和Decoder中。

DeepLab小结

DeepLab v1: 空洞卷积 + DenseCRF

✓不丢失信息情况下增大感受野，考虑标签之间的概率关系

DeepLab v2: 空洞卷积多尺度特征ASPP

✓不同尺度物体的特征提取

DeepLab v3: 增加了多尺度串行、增强ASPP、去掉CRF

DeepLab v3+: 结合low-level feature修复分割边界

内容

- 什么是视觉分割
- 语义分割网络
 - 使用特征融合得到语义信息的分割网络
 - 全卷积分割网络(DeepLab系列)
- 语义分割的应用

语义分割的应用

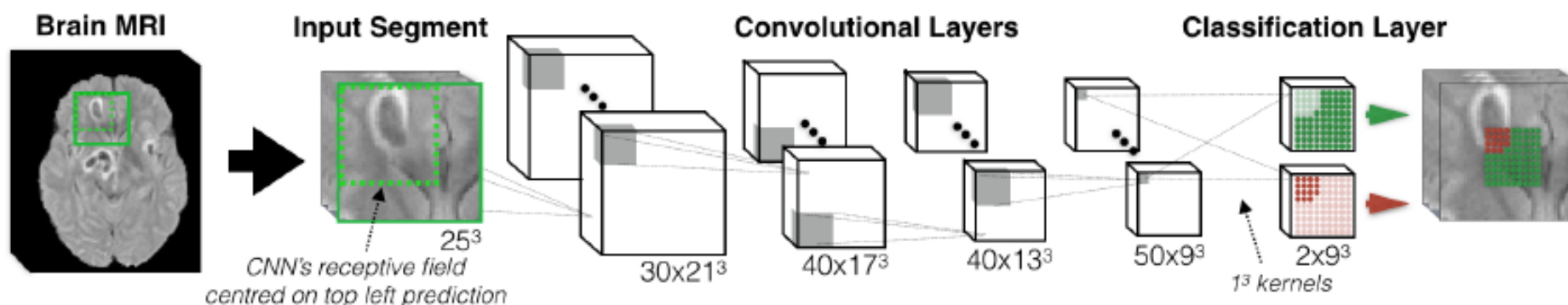
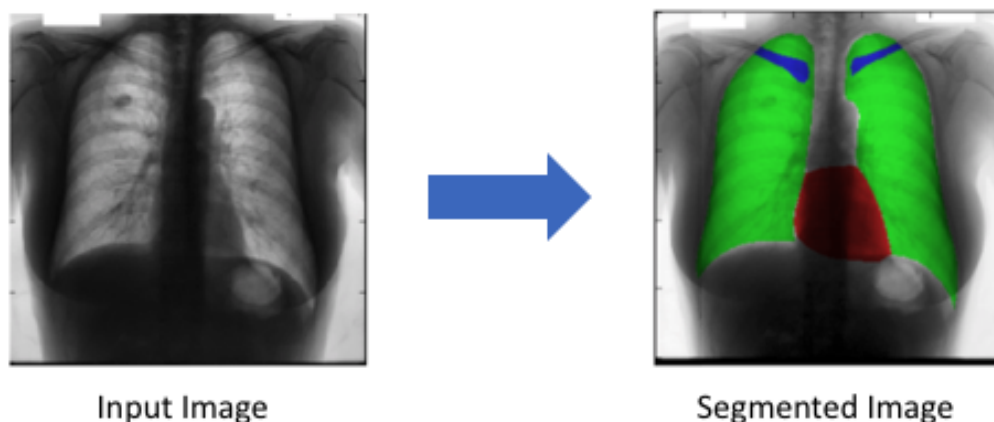
➤ 医学图像分割

➤ 自动驾驶

.....

语义分割的应用

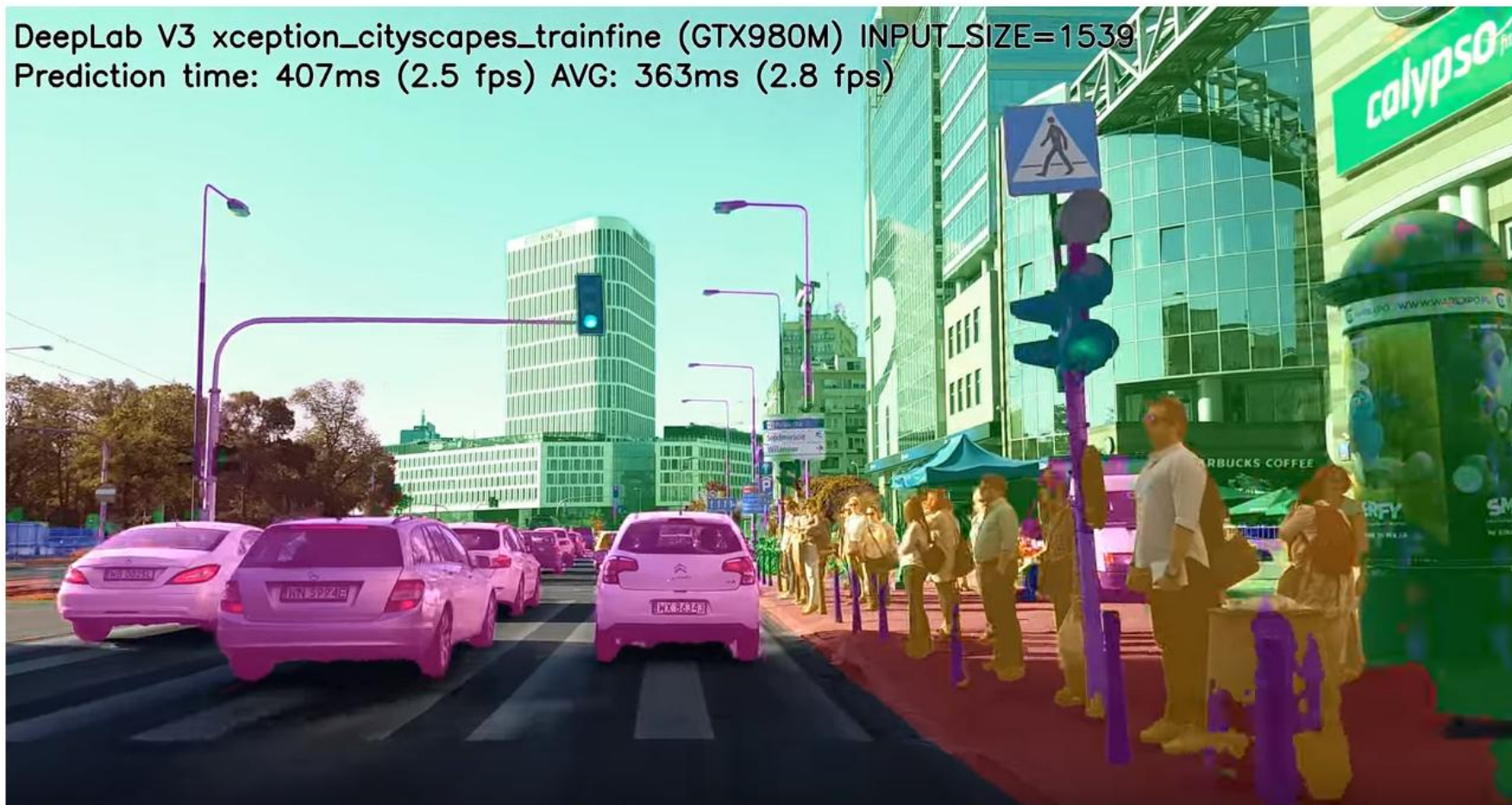
医学图像分割



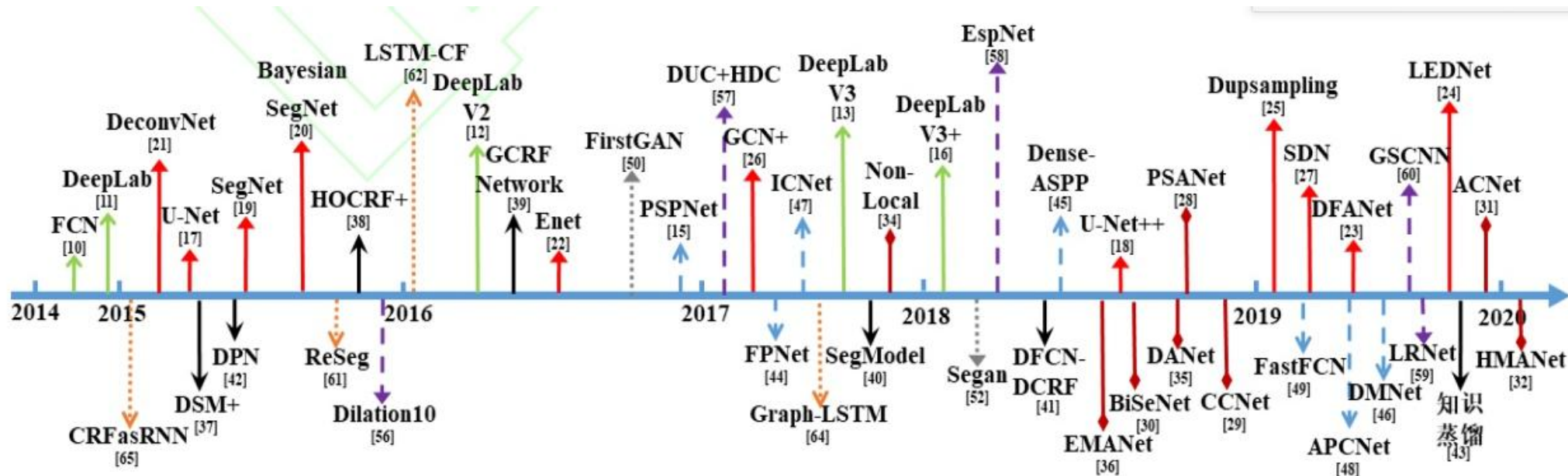
语义分割的应用

自动驾驶

DeepLab V3 xception_cityscapes_trainfine (GTX980M) INPUT_SIZE=1539
Prediction time: 407ms (2.5 fps) AVG: 363ms (2.8 fps)



更多内容

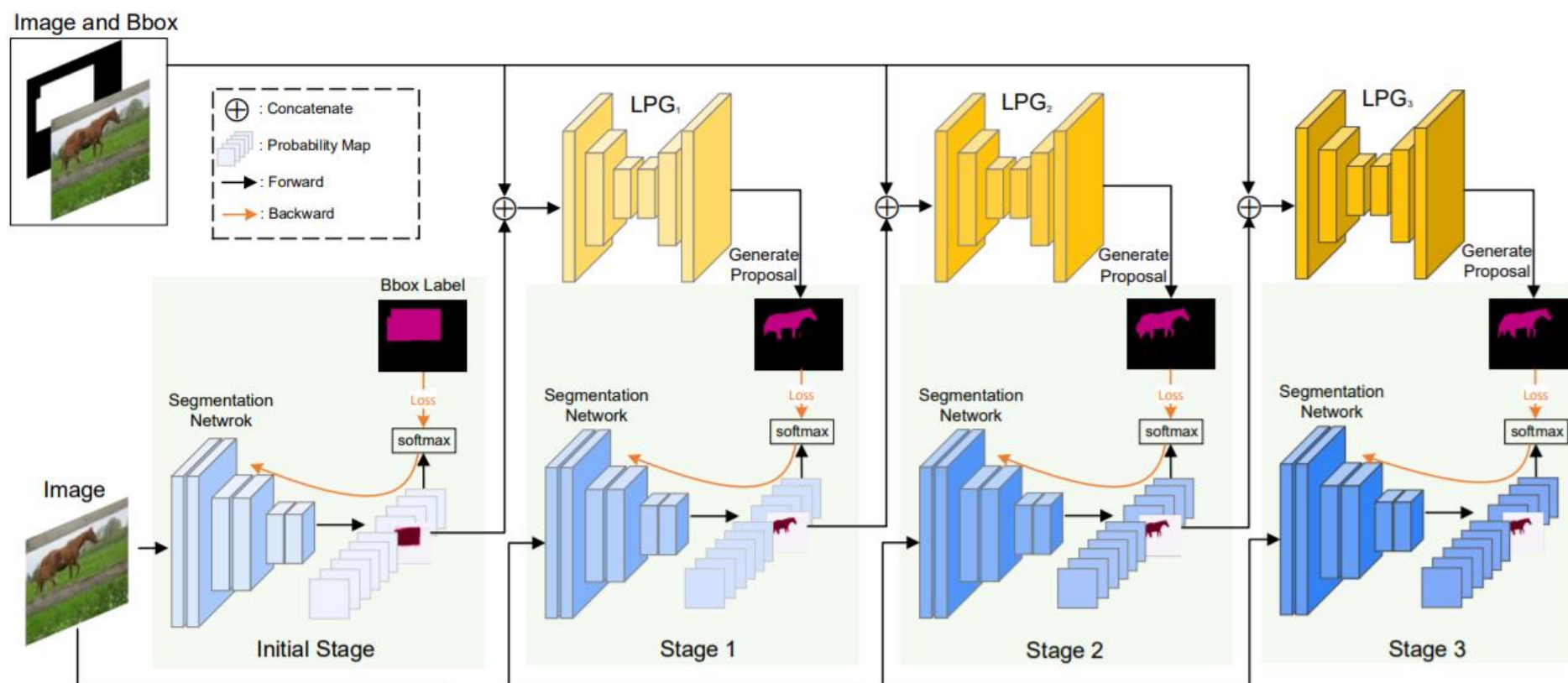


弱监督语义分割



弱监督语义分割

只给定标注框，训练语义分割网络



Learning Class-Agnostic Pseudo Mask Generation for Box-Supervised Semantic Segmentation, arxiv 2021.

谢谢