



华东师范大学

East China Normal University, ECNU



第十四章 智能图像分割



1

分割概念基础

2

现有的智能分割的典型算法

3

分类算法实例分析

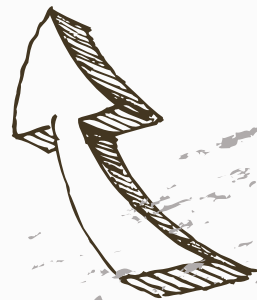
4

分割技术发展中的主要问题

5

分割技术应用实例

PART 01



分割概念基础



分割概念基础

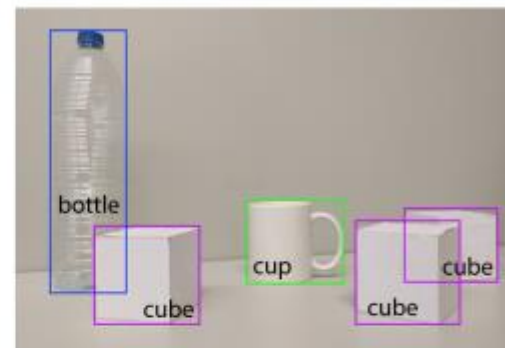


语义分割：将每个像素分类为属于对象类的过程。

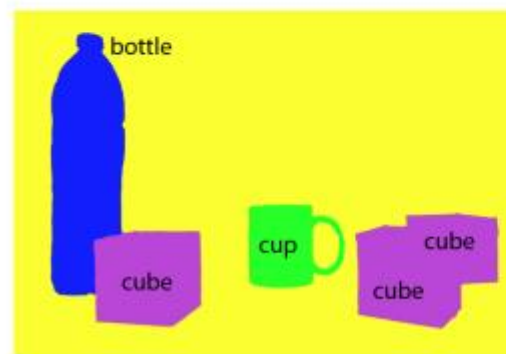
不同于分类任务，图像语义分割所追求的是输入一张图片，输出也是一张图片，学习像素到像素的映射。



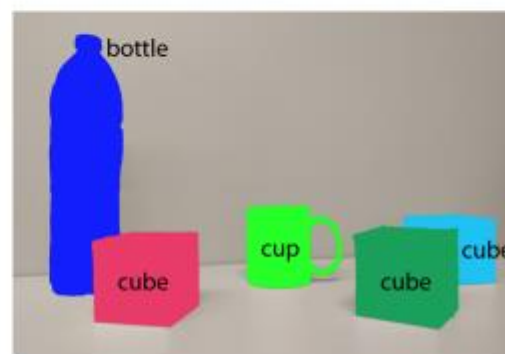
(a) Image classification



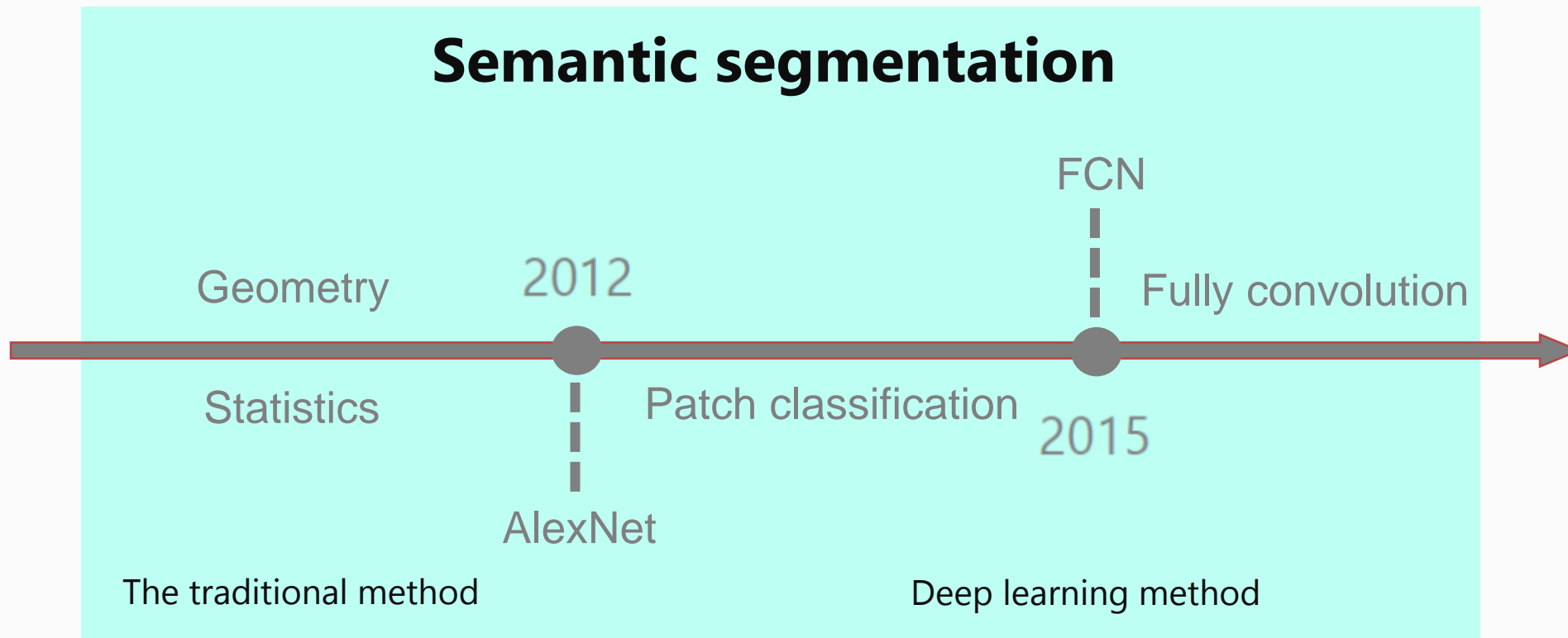
(b) Object localization



(c) Semantic segmentation



(d) Instance segmentation





分割概念基础



语义分割中有许多数据集，简单介绍以下四种：

Pascal VOC |

通常采用PASCAL VOC 2012，最开始有1464 张具有标注信息的训练图片，2014 年增加到10582张训练图片。主要涉及了日常生活中常见的物体，包括汽车，狗，船等20个分类

Microsoft COCO |

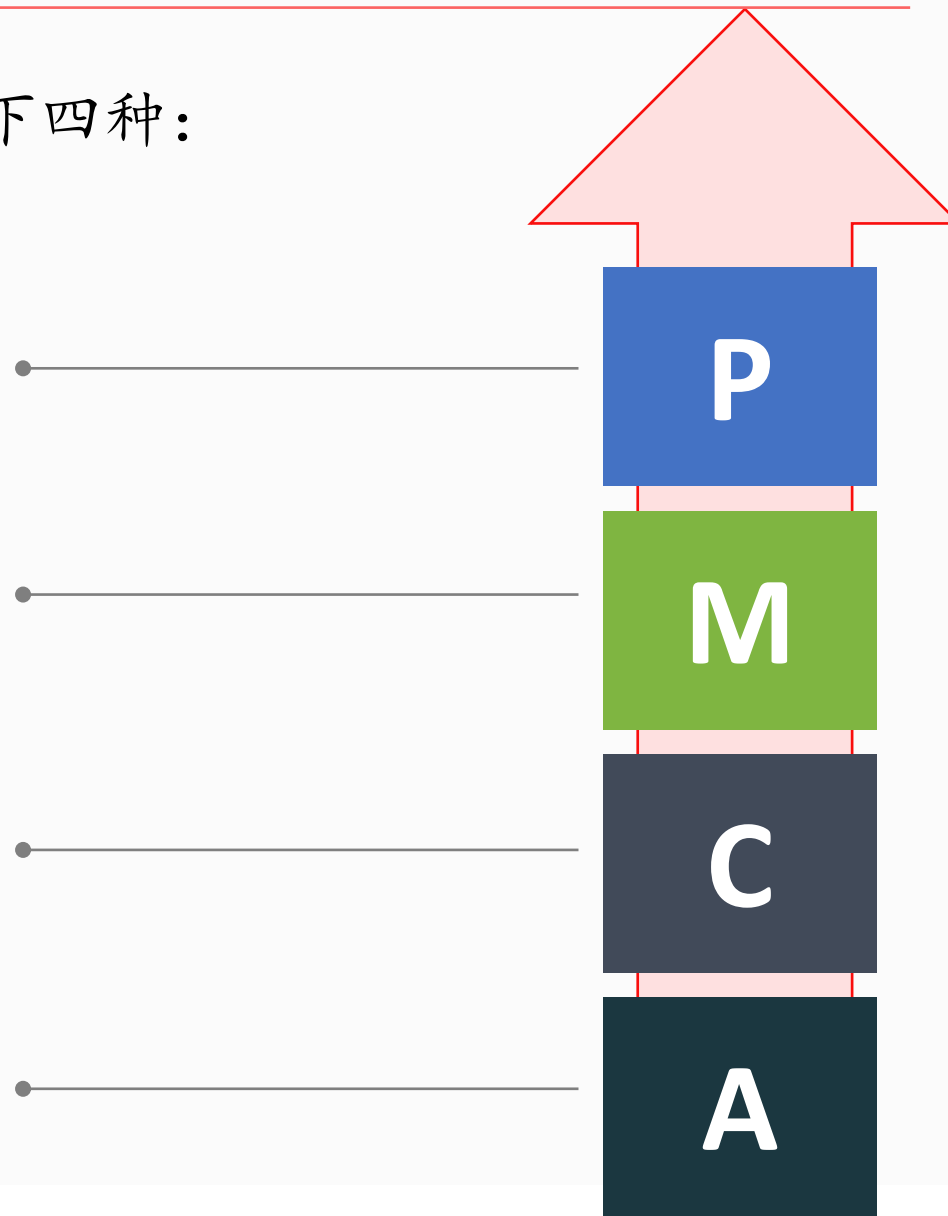
一共有80个类别。这个数据集主要用于**实例级别的分割**（Instance-level Segmentation）以及图片描述Image Caption）。

Cityscapes |

适用于**汽车自动驾驶**的训练数据集，包括19种都市街道场景：road、side-walk、building、wal、fence、pole、traficlight、traffic sign、vegetation、terain、sky、person、rider、car、truck、bus、train、motorcycle 和 bicycle。该数据库中用于训练和校验的精细标注的图片数量为3475，同时也包含了 2 万张粗糙的标记图片。

ADE20K |

包括室内和室外场景共计150类。拥有20000张训练数据，2000张验证数据，以及若干张测试数据。





分割概念基础



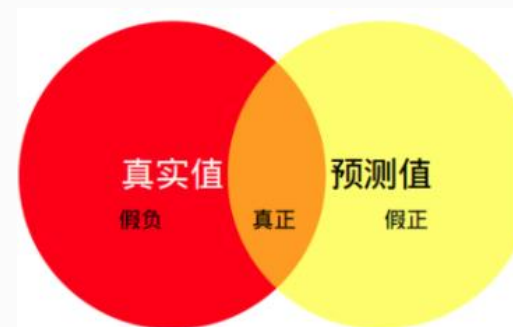
语义分割中常见的评价指标有**平均像素准确率 (mPA)**和**平均交并比 (mIoU)**

平均像素准确率是分别计算每个类别分类正确的像素数占有所有预测为该类别像素数的比例，即精确率，然后累加求平均

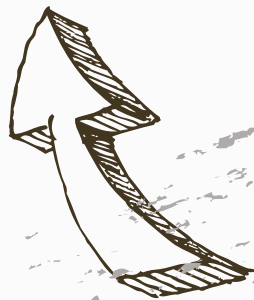
$$MPA = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}}$$

平均交并比是对每一类预测的结果和真实值的交集与并集的比值求和平均的结果

$$MIoU = \frac{1}{n+1} \sum_{i=0}^n \frac{p_{ii}}{\sum_{j=0}^n p_{ij} + \sum_{j=0}^n p_{ji} - p_{ii}}$$



PART 02



现有的智能分割的典型算法



现有的智能分割的典型算法



- **AlextNet**
- **FCN**
- **SegNet**
- **U-Net**
- **DeepLab 系列**
- **Attention 机制**



现有的智能分割的典型算法



- 最初，图像块分类(patch classification)是常用的深度学习方法，即利用每个像素周围的图像块分别将各像素分成对应的类别。
- 其中，使用图像块的主要原因是分类网络通常具有全连接层，其输入需为固定大小的图像块。



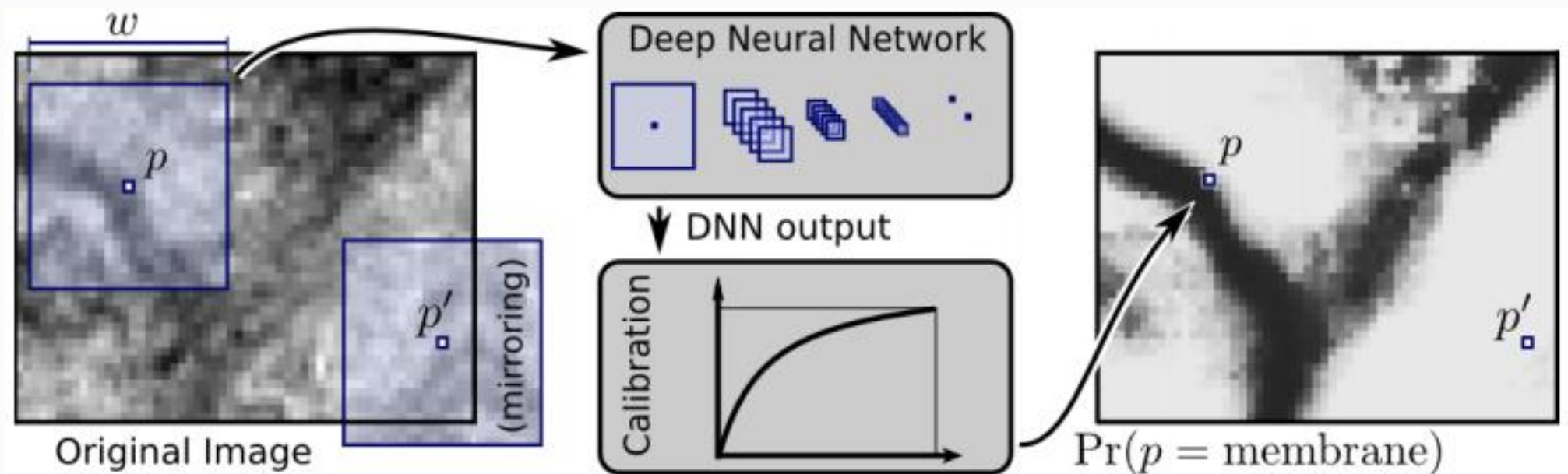
(a) AlexNet



现有的智能分割的典型算法



基于AlexNet采用**Patch-classification**方法和EM算法进行语义分割：





现有的智能分割的典型算法



基于AlexNet采用**Patch-classification**方法和EM算法进行语义分割：

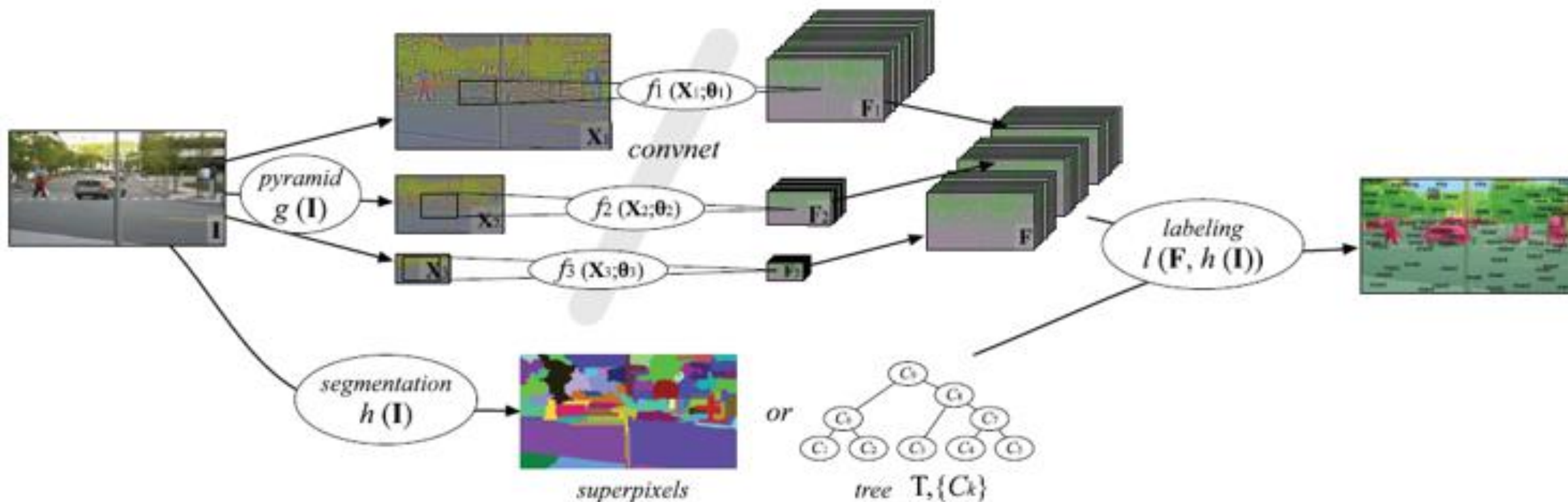
Layer	Type	Maps and neurons	Kernel size
0	input	1 map of 95x95 neurons	
1	convolutional	48 maps of 92x92 neurons	4x4
2	max pooling	48 maps of 46x46 neurons	2x2
3	convolutional	48 maps of 42x42 neurons	5x5
4	max pooling	48 maps of 21x21 neurons	2x2
5	convolutional	48 maps of 18x18 neurons	4x4
6	max pooling	48 maps of 9x9 neurons	2x2
7	convolutional	48 maps of 6x6 neurons	4x4
8	max pooling	48 maps of 3x3 neurons	2x2
9	fully connected	200 neurons	1x1
10	fully connected	2 neurons	1x1



现有的智能分割的典型算法



基于AlexNet采用**Patch-classification**方法和CRF算法进行语义分割：

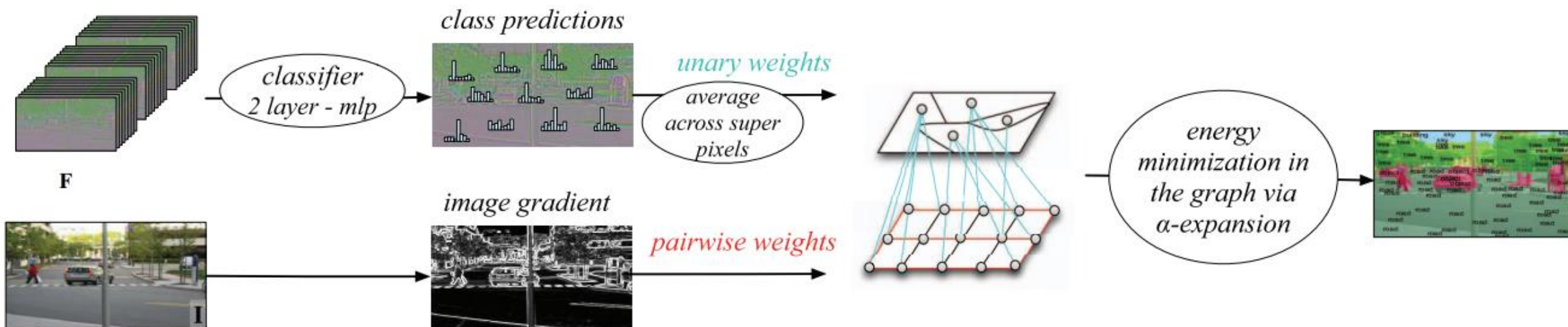




现有的智能分割的典型算法



基于AlexNet采用**Patch-classification**方法和CRF算法进行语义分割：

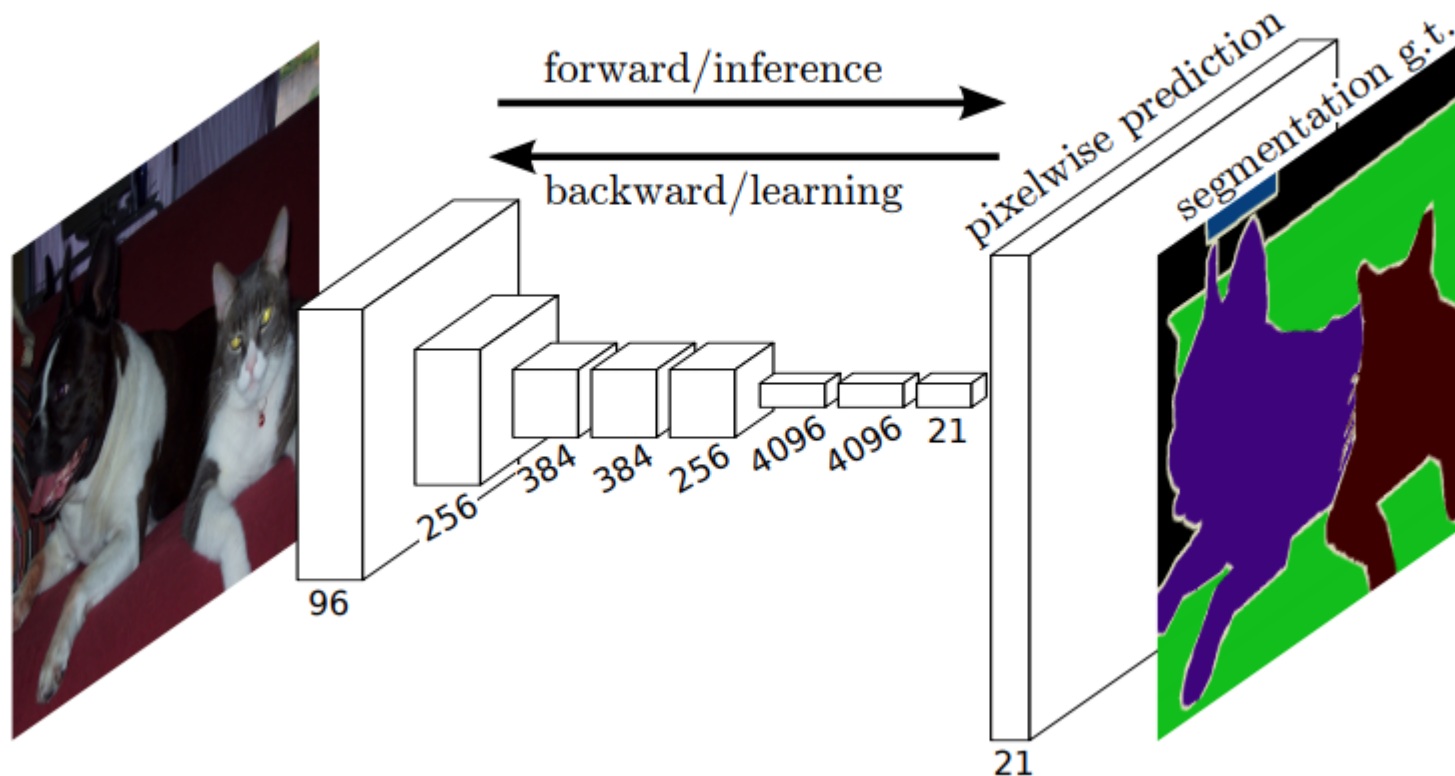




现有的智能分割的典型算法



2015，全卷积网络（FCN）横空出世，FCN将网络全连接层用卷积取代，因此使任意图像大小的输入都变成可能，而且速度比Patch classification方法快很多。

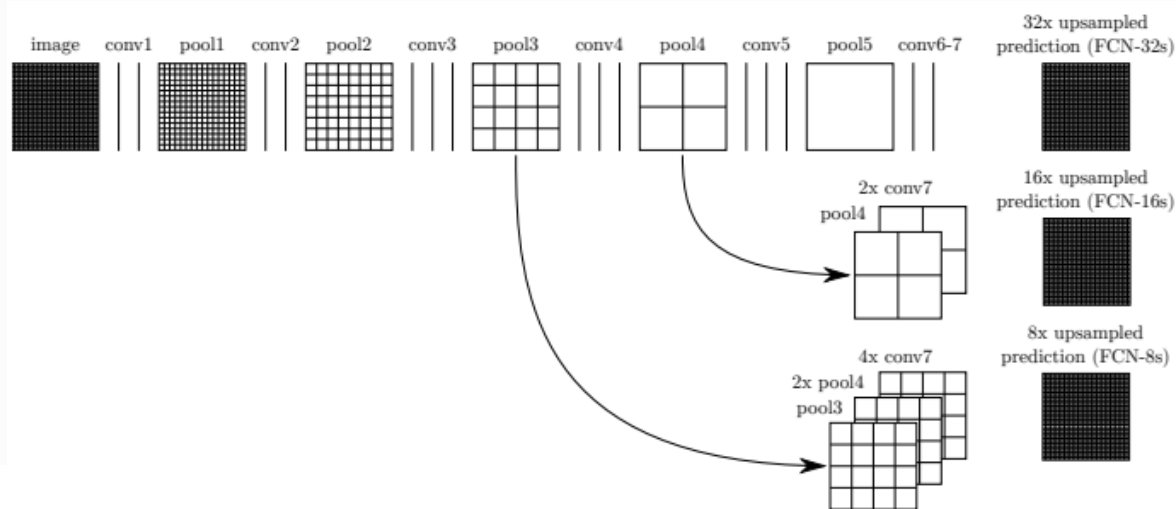




现有的智能分割的典型算法



- 首先，将一幅 RGB 图像输入到卷积神经网络后，经过多次卷积及池化过程得到一系列的特征图；
- 然后，利用反卷积层对最后一个卷积层得到的特征图进行上采样，使得上采样后特征图与原图像的大小一样，从而实现对特征图上的每个像素值进行预测的同时保留其在原图像中的空间位置信息；
- 最后对上采样特征图进行逐像素分类，逐个像素计算 softmax 分类损失。





FCN的主要贡献

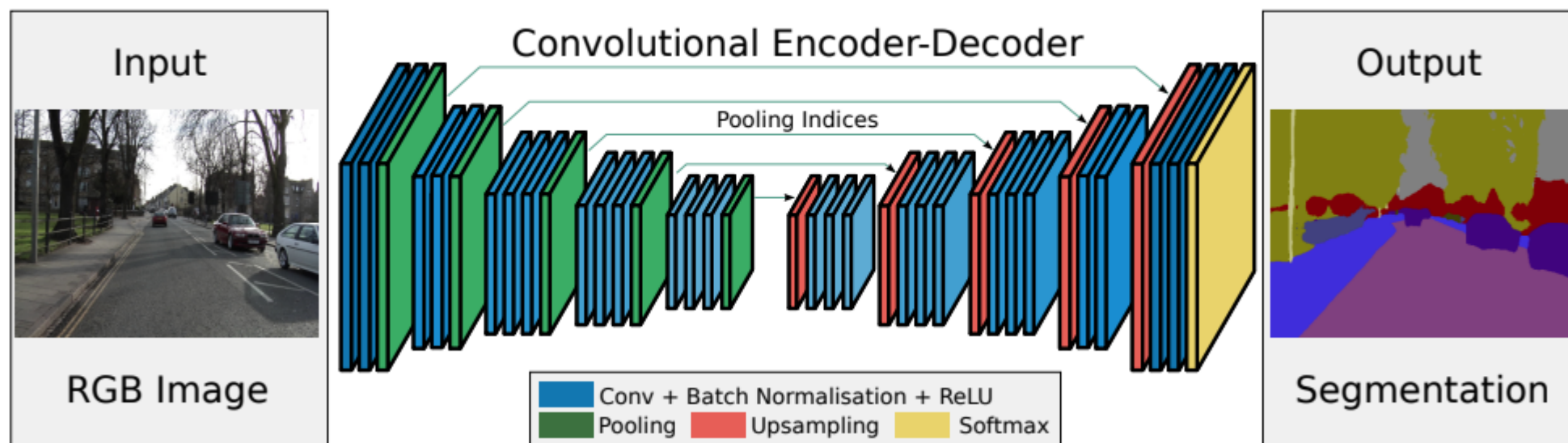
- 不含全连接层(fc)的全卷积(fully conv)网络。可适应任意尺寸输入。
- 增大数据尺寸的反卷积(deconv)层。能够输出精细的结果。
- 结合不同深度层结果的跳级(skip)结构。同时确保鲁棒性和精确性。

FCN的缺点同样明显

- 得到的结果不够精细，虽然8倍上采样比32倍上采样效果好，但是上采样的结果仍然比较模糊和平滑，对于图像中的细节部分不敏感。
- 对于图像中各个像素的分类没有充分考虑到像素与像素之间的关系，缺乏空间一致性。

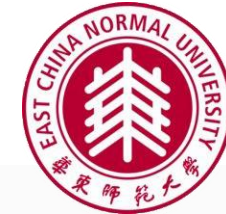


SegNet

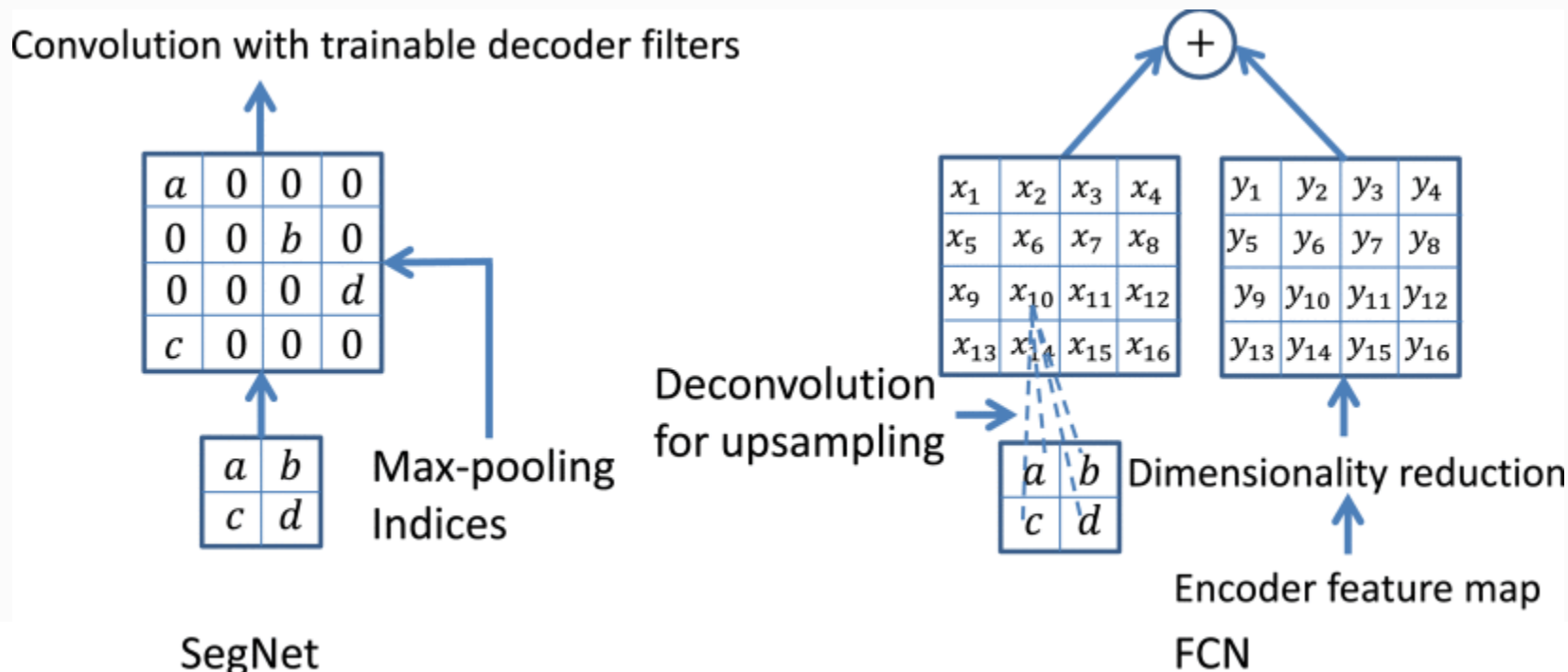




现有的智能分割的典型算法



- SegNet首次使用具有相同阶段数的Encoder-decoder结构，为后人将feature map还原为原来的大小提供了新的思路。
- 同时，SegNet 网络采用了 pooling indices 来保存图像的轮廓信息，在降低了参数数量的情况下增加了精度。

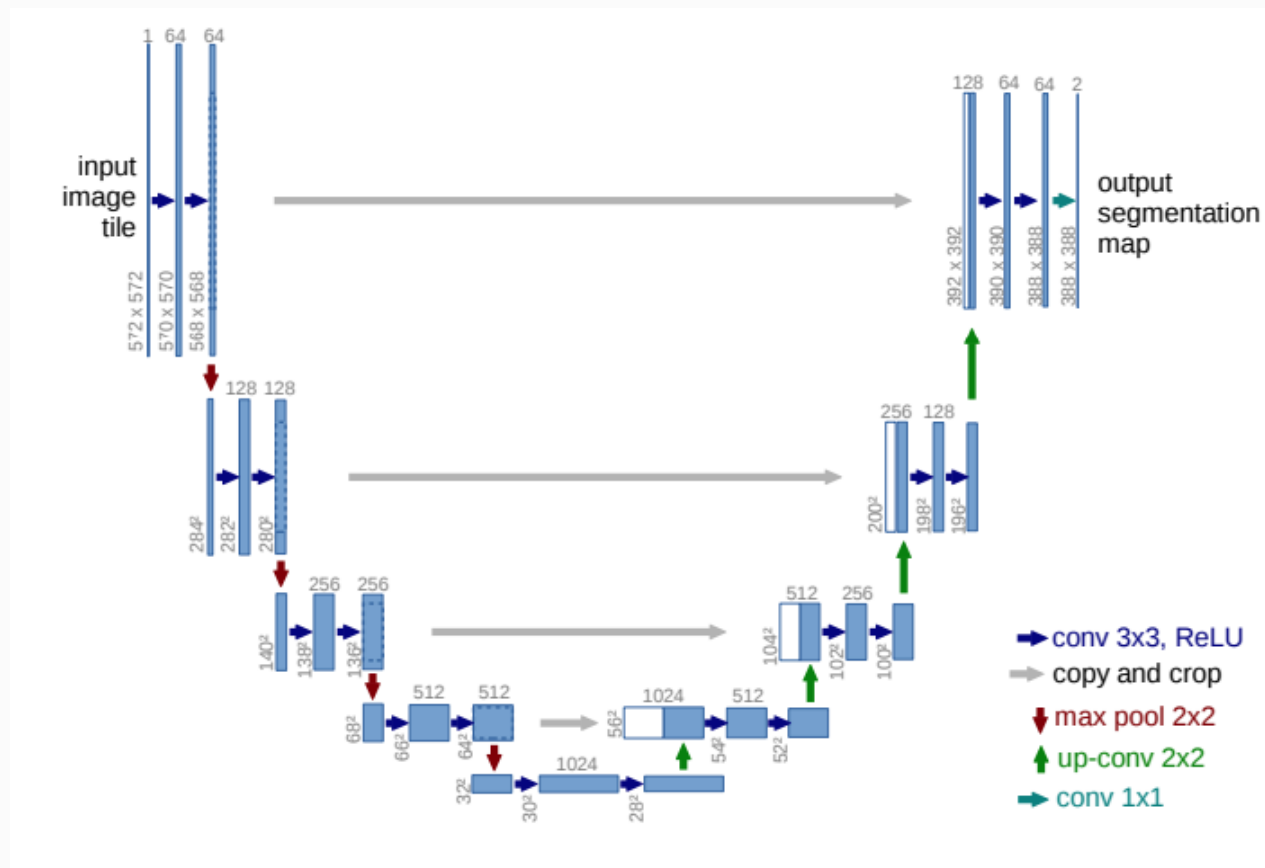




现有的智能分割的典型算法



为了解决医疗图像语义分割任务中数据量不足的情况，U-Net在2015年被提出



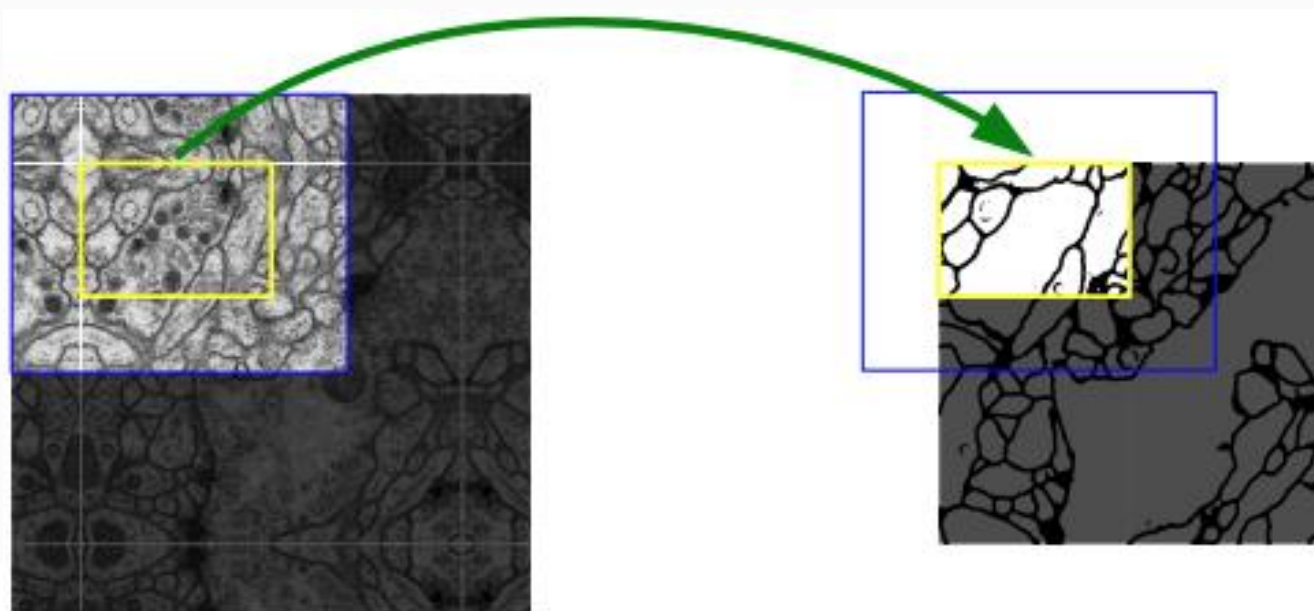
Ronneberger, Olaf, Fischer, Philipp, Brox, Thomas. "U-Net: Convolutional Networks for Biomedical Image Segmentation." (2015).



现有的智能分割的典型算法



- U-Net 基于 FCN，UNet 共进行了 4 次上采样，并在同一个 stage 使用了 **skip connection**，这样就保证了：最后恢复出来的特征图融合了更多的 low-level 的 feature，也使得不同 scale 的 feature 得到了融合，从而可以进行多尺度预测和 Deep-Supervision。4 次上采样也使得分割图恢复边缘等信息更加精细。
- 同时，采用了 Overlap-tile strategy，镜像拷贝。





DeepLab是Google团队一系列semantic image segmentation的paper, 截止2018年, 共4篇。

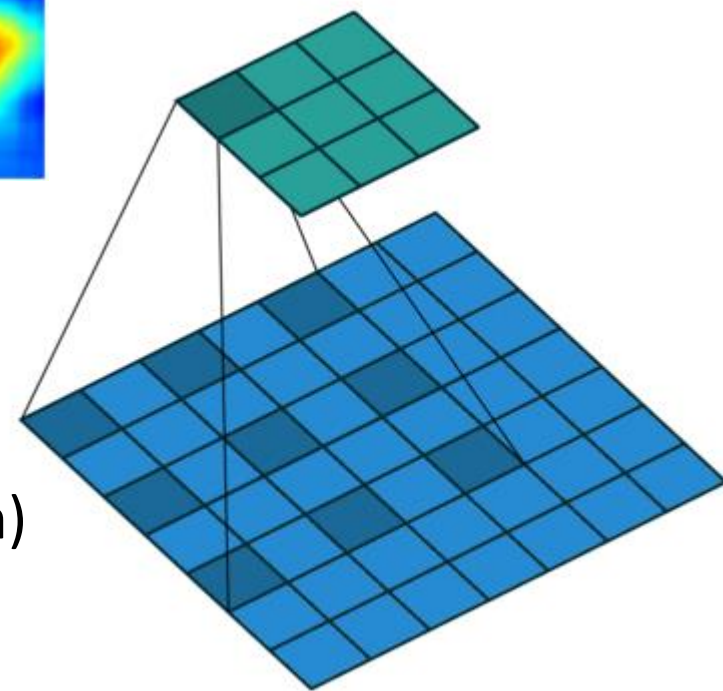
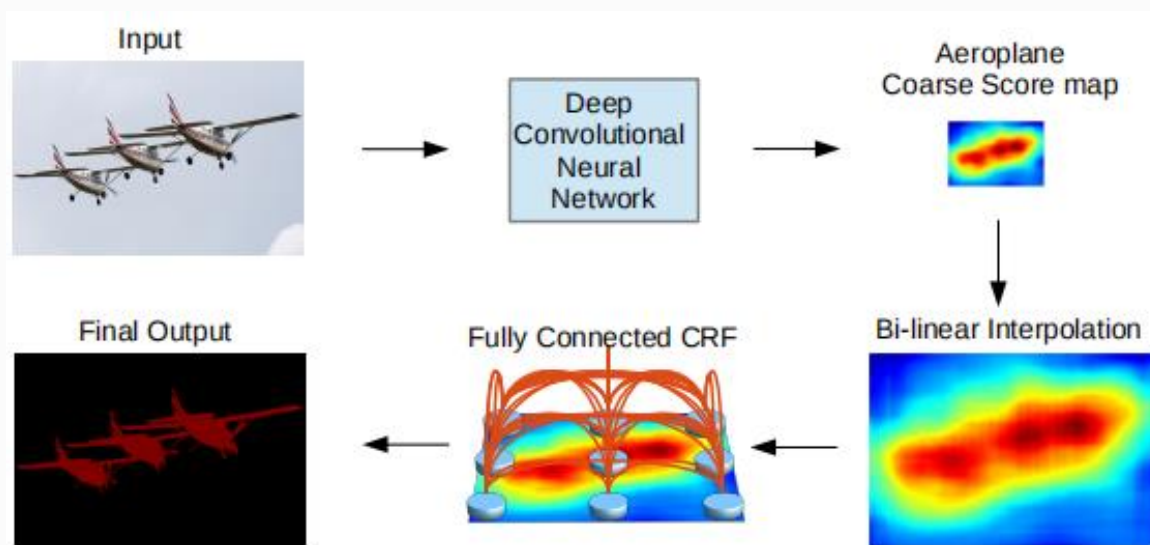
- ◆ Deeplabv1(2015)
- ◆ Deeplabv2(2016)
- ◆ Deeplabv3(2017)
- ◆ Deeplabv3-plus(2018)



现有的智能分割的典型算法



Deep labv1

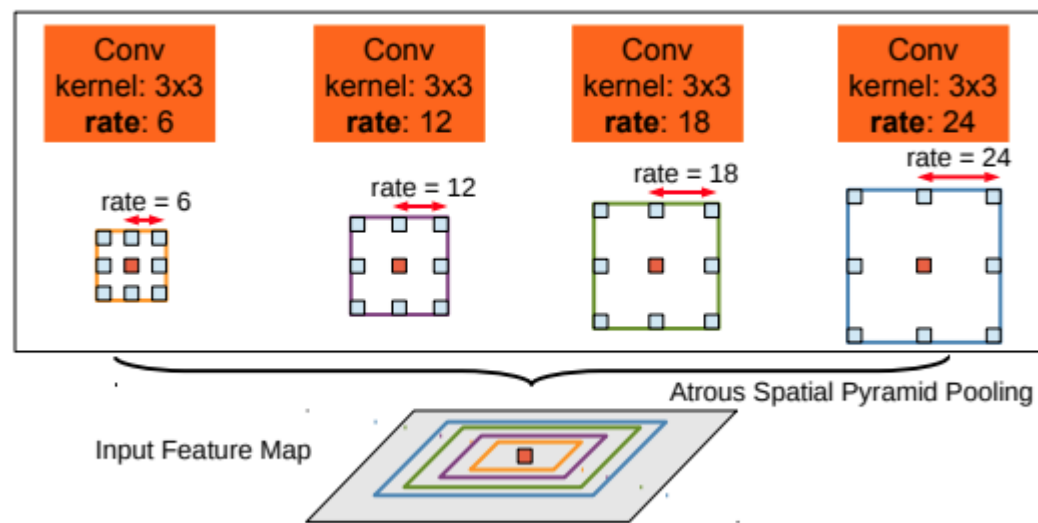


- VGG16的全连接层转为卷积
- 最后的两个最大池化层去掉了下采样
- 后续卷积层的卷积核改为了空洞卷积(atrous convolution)
- 在ImageNet上预训练的VGG16权重上做finetune



DeepLabv2

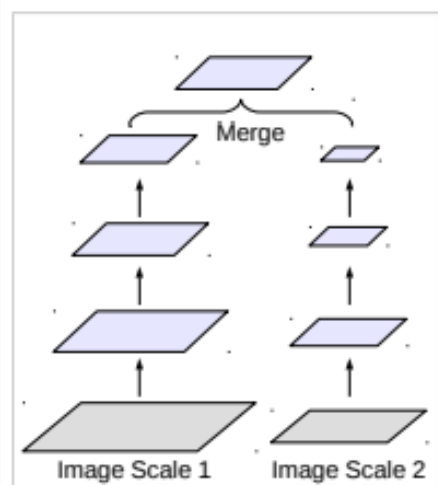
针对物体的多尺度问题，提出ASPP(atrous spatial pyramid pooling)模块；在卷积之前以多种采样率在给定的特征层上进行重采样；使用多条平行的有不同采样率的空洞卷积层。



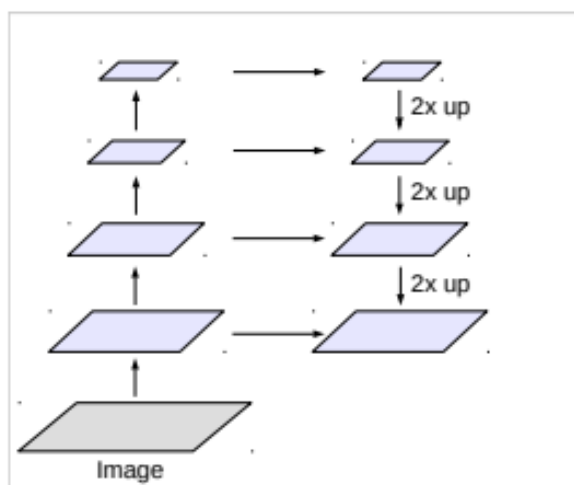


DeepLabv3

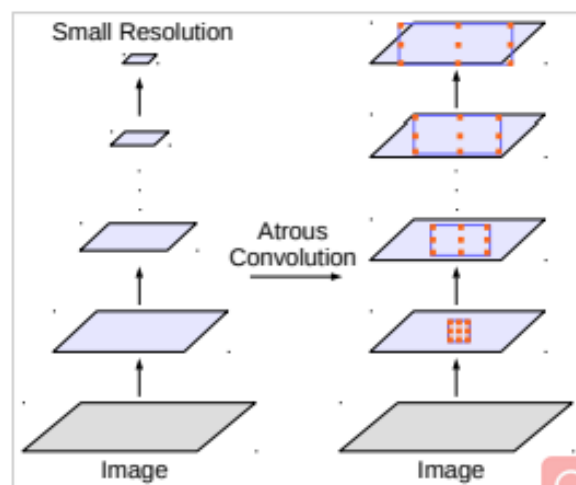
作者对比了几种常见的捕获multi-scale context的方法，对于原本的ASPP进行了改进



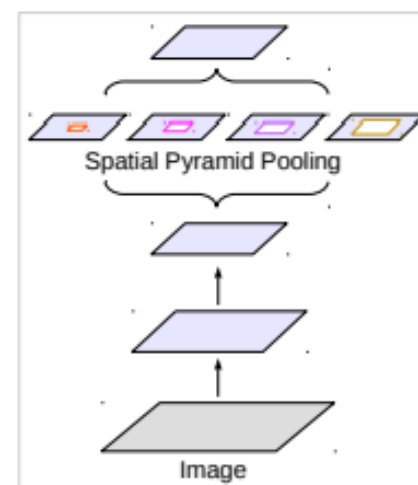
(a) Image Pyramid



(b) Encoder-Decoder



(c) Deeper w. Atrous Convolution



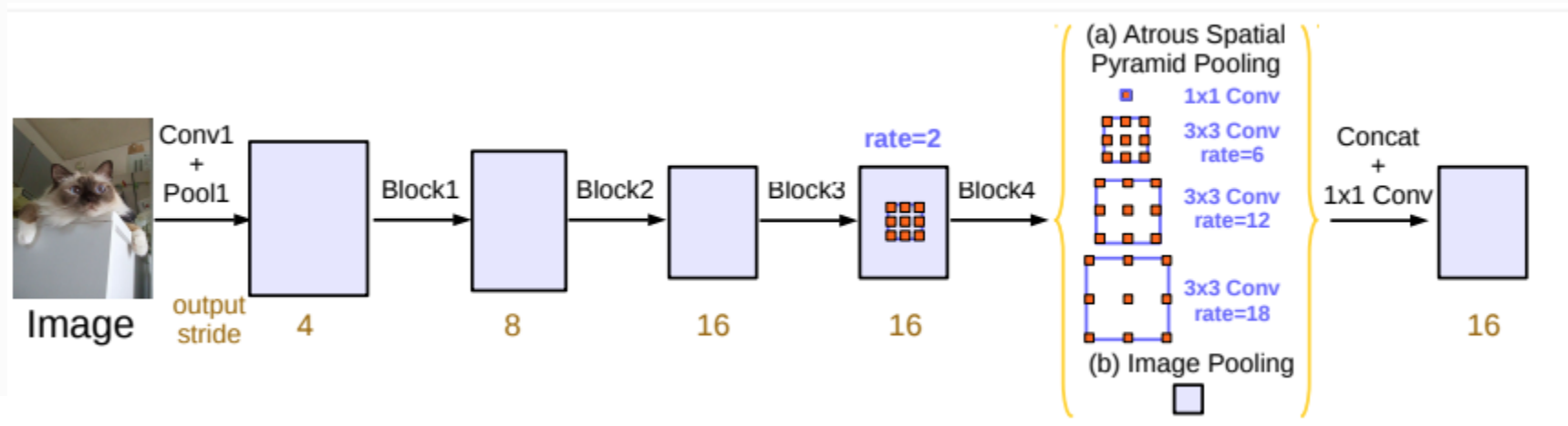
(d) Spatial Pyramid Pooling



DeepLabv3

新的ASPP模块包括：

- 一个 1×1 卷积和3个 3×3 的空洞卷积(采样率为(6,12,18))，每个卷积核都有256个且都有BN层
- 包含图像级特征image-level features(即全局平均池化Global Average Pooling)



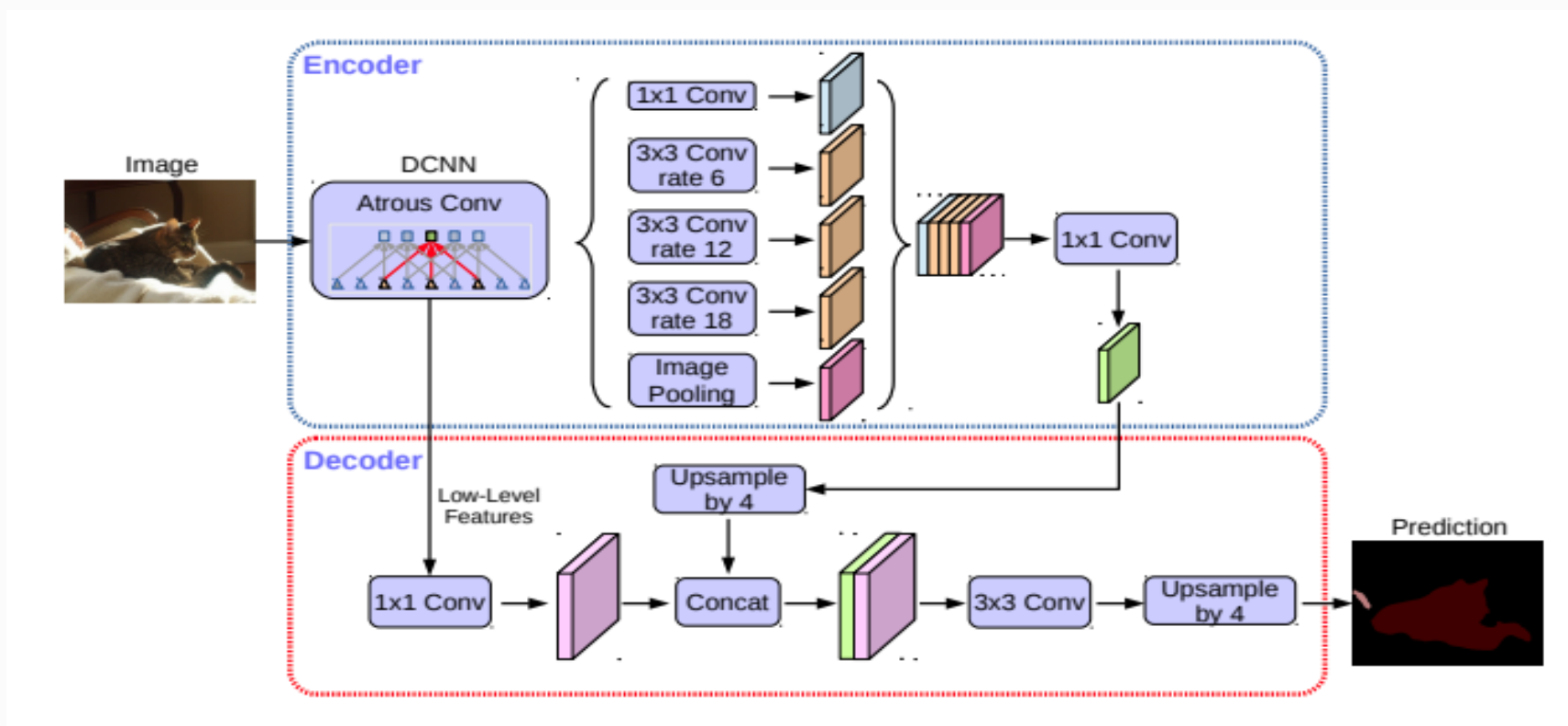


现有的智能分割的典型算法



DeepLabv3+

原DeepLabv3当作encoder，添加decoder得到新的模型（DeepLabv3+）



Liang-Chieh Chen, George Papandreou, Florian Schroff, Hartwig Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation" . (2018)

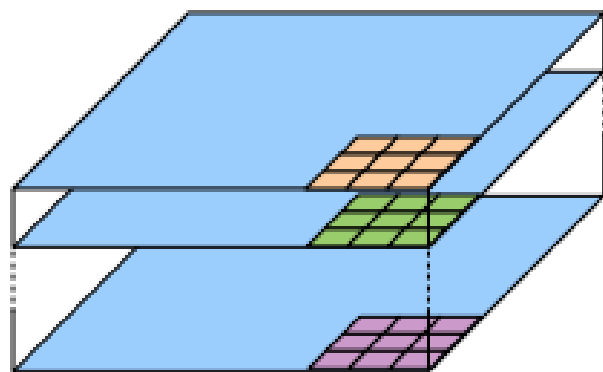


DeepLabv3+

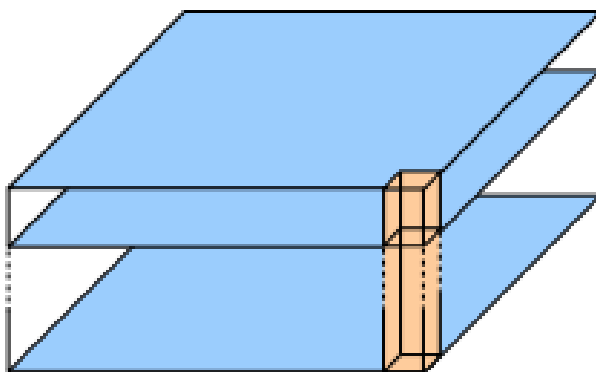
Deeplabv3+的创新点:

一是设计基于Deeplabv3的decode module,

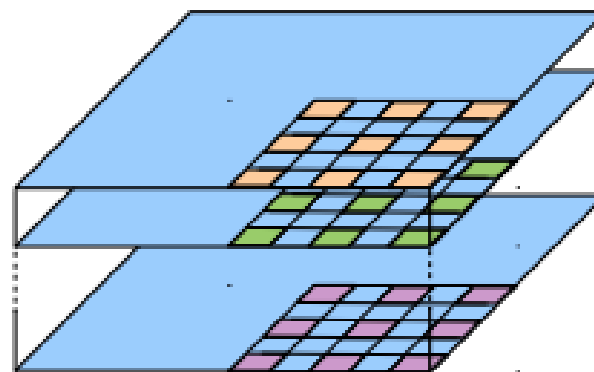
二是用modify xception作为backbone, 采用深度可分离卷积减少参数量和计算量。



(a) Depthwise conv.



(b) Pointwise conv.



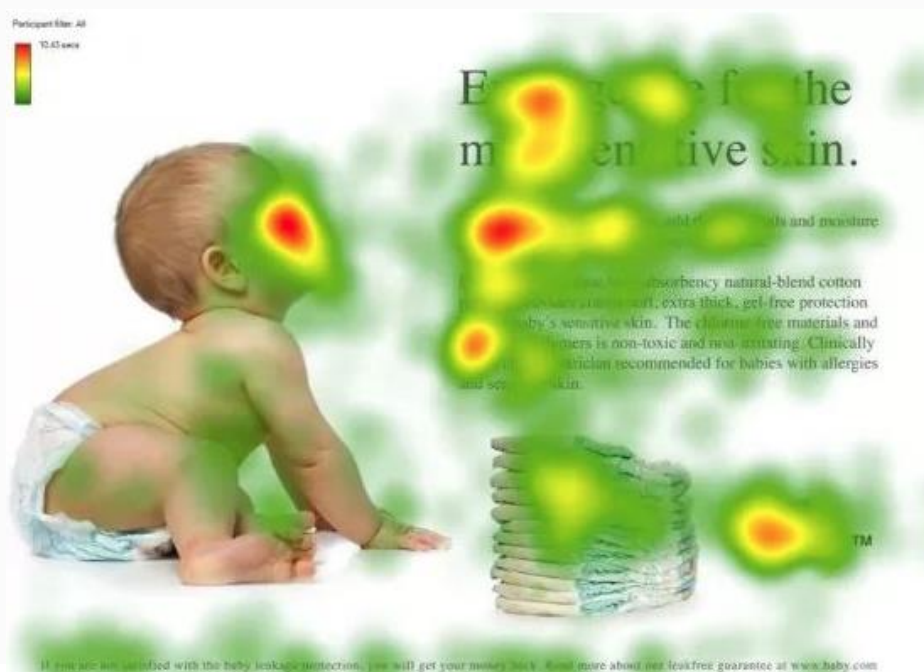
(c) Atrous depthwise conv.



现有的智能分割的典型算法



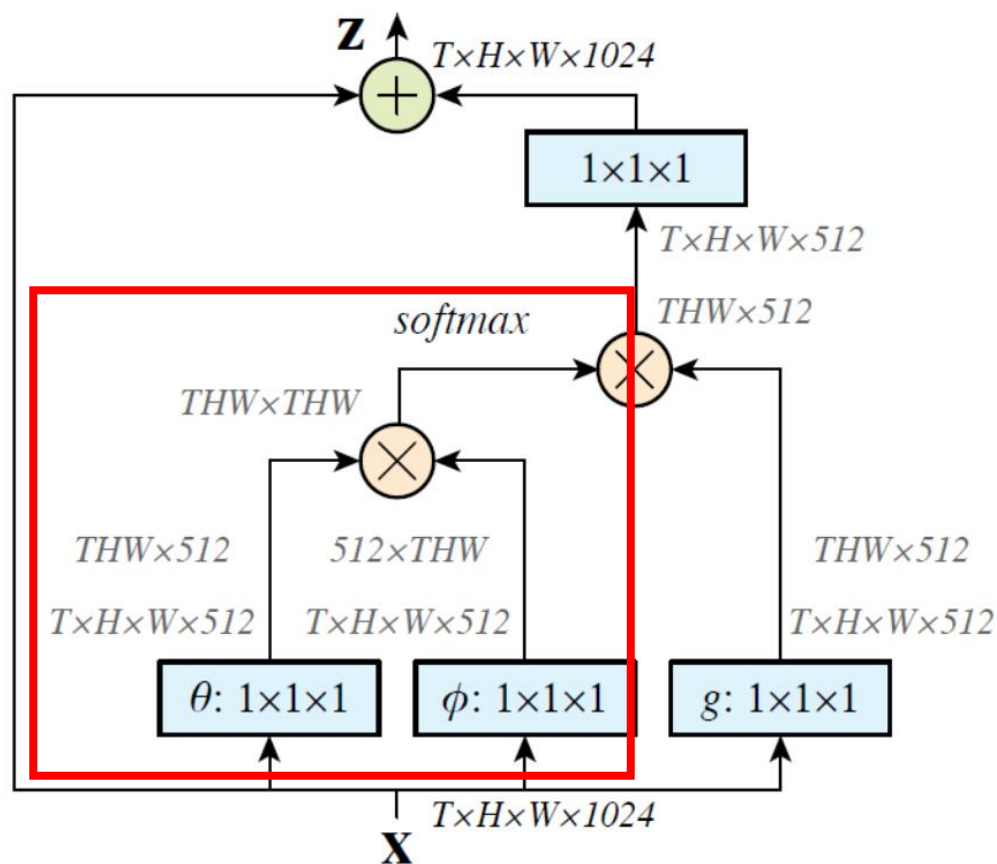
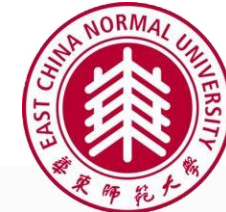
Attention 机制继在 NLP 领域取得主导地位之后，近几年在 CV 领域也开始独领风骚。率先将之引入的是 Kaiming He 组的 Nonlocal。此后层出不穷的文章，引发了一波研究attention机制的热潮。仅2018年，在语义分割领域就有多篇高影响力文章出炉，如 PSANet, DANet, OCNNet, CCNet等



- 深度学习中的注意力机制从本质上讲和人类的**选择性视觉注意力机制**类似，
- 核心目标也是从众多信息中选择出对当前任务目标更关键的信息。



现有的智能分割的典型算法



$$\text{Eq1: } y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j)$$

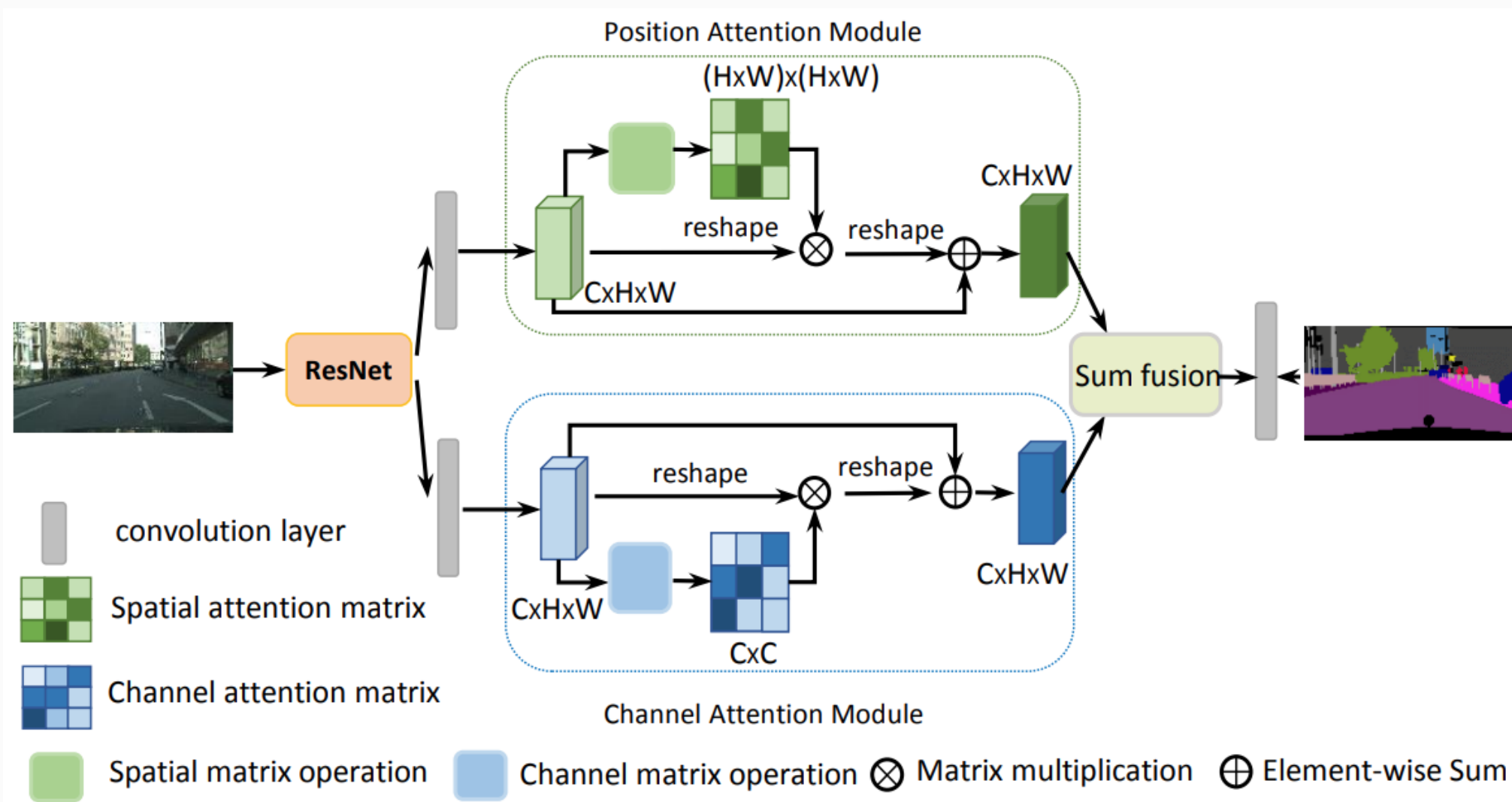
$$\text{Eq6: } z_i = W_z y_i + x_i$$

对于计算两点特征关系的函数 f 有多种选择

1. Gaussian, $f(x_i, x_j) = \exp\{x_i^T x_j\}$
2. Embedding Gaussian, $f(x_i, x_j) = \exp\{\theta(x_i)^T \phi(x_j)\}$
3. Dot Product, $f(x_i, x_j) = \theta(x_i)^T \phi(x_j)$
4. Concatenation, $f(x_i, x_j) = \text{ReLU}(w^T [\theta(x_i)^T, \phi(x_j)])$



现有的智能分割的典型算法





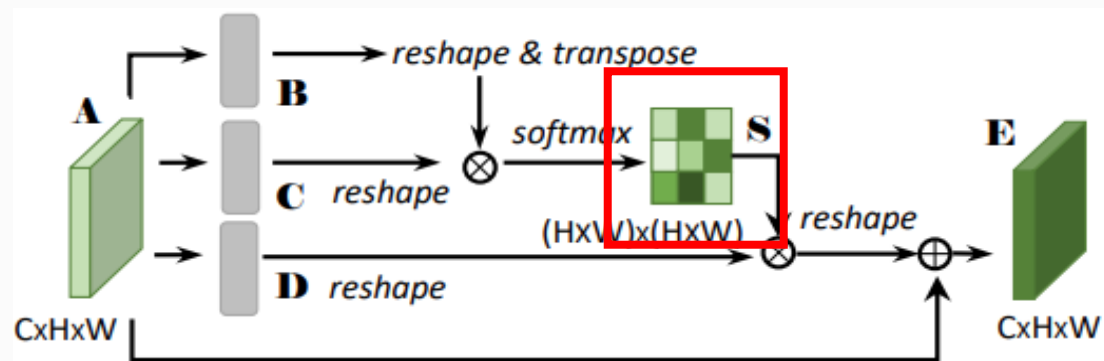
现有的智能分割的典型算法



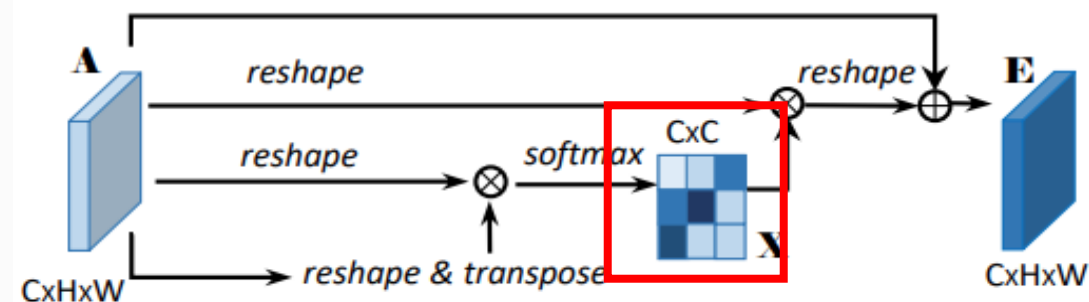
Dual Attention Networks加入两种类型的attention module。其中，

- **position attention module**选择性地通过所有位置的加权求和聚集每个位置的特征。
- **channel attention module**通过所有channle的特征map中的特征选择性地强调某个特征图。

最后将两种attention module的output求和得到最后的特征表达。



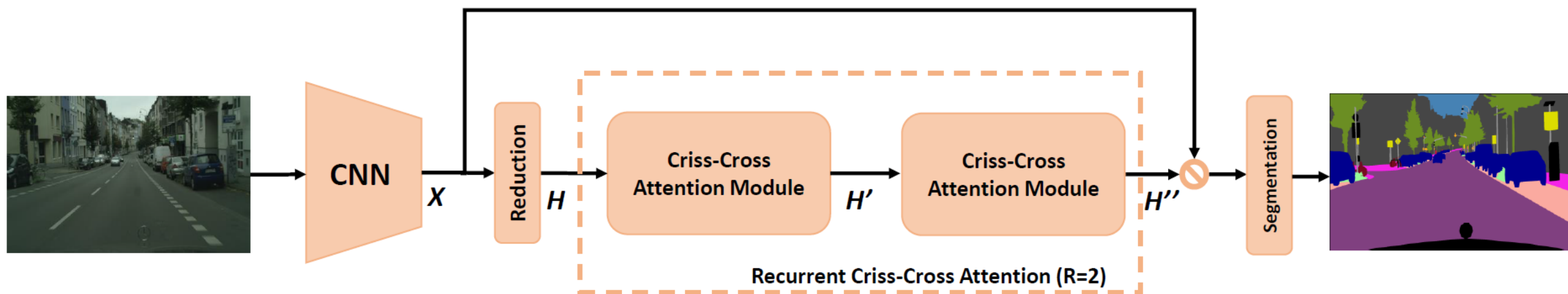
A. Position attention module



B. Channel attention module

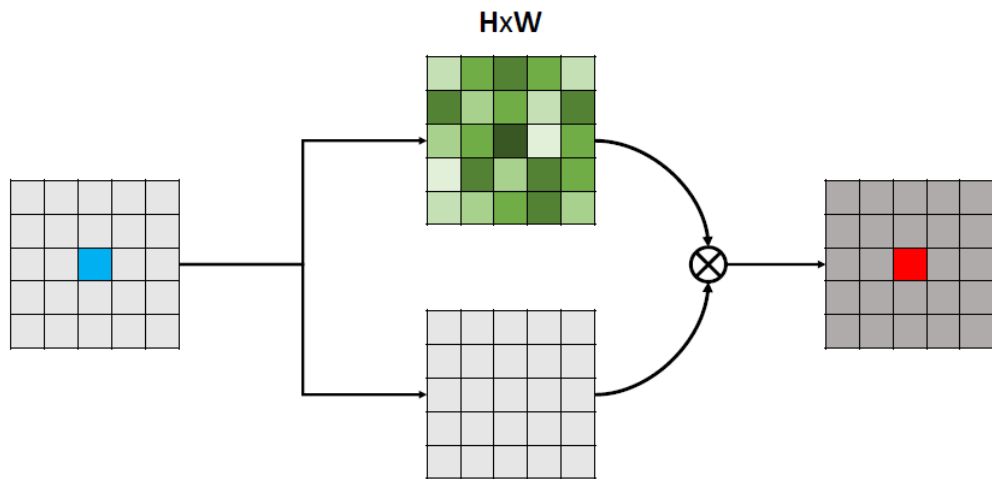


现有的智能分割的典型算法

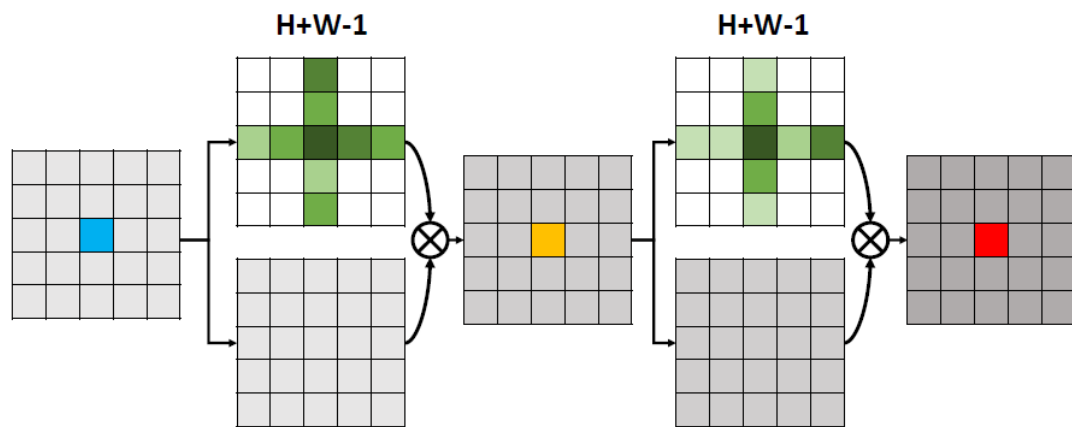




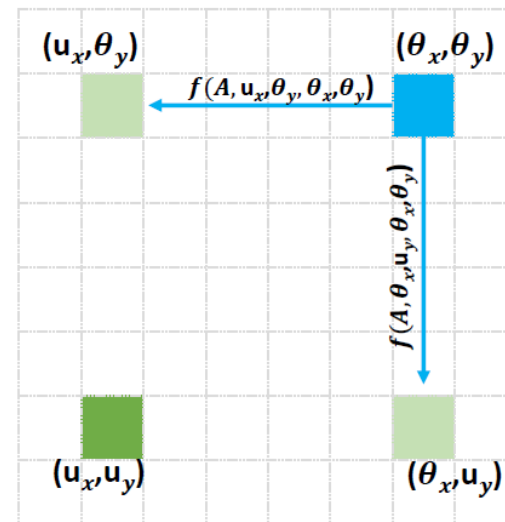
现有的智能分割的典型算法



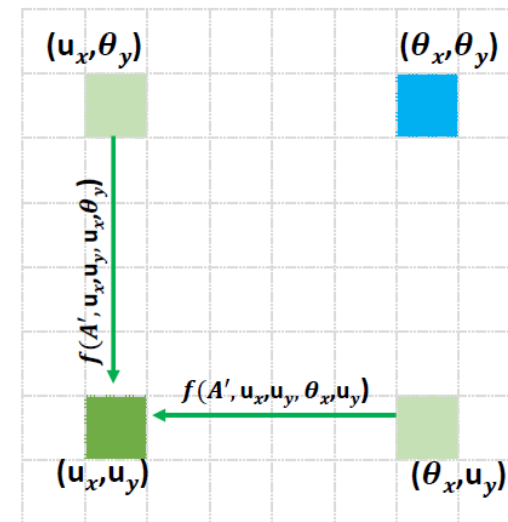
(a) Non-local block



(b) Criss-Cross Attention block



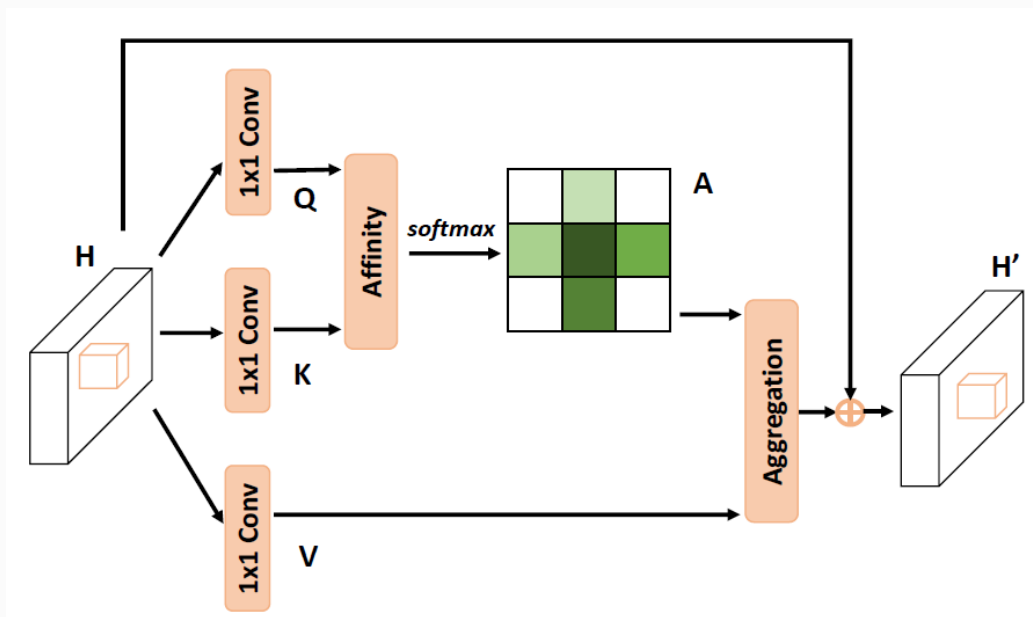
Loop 1



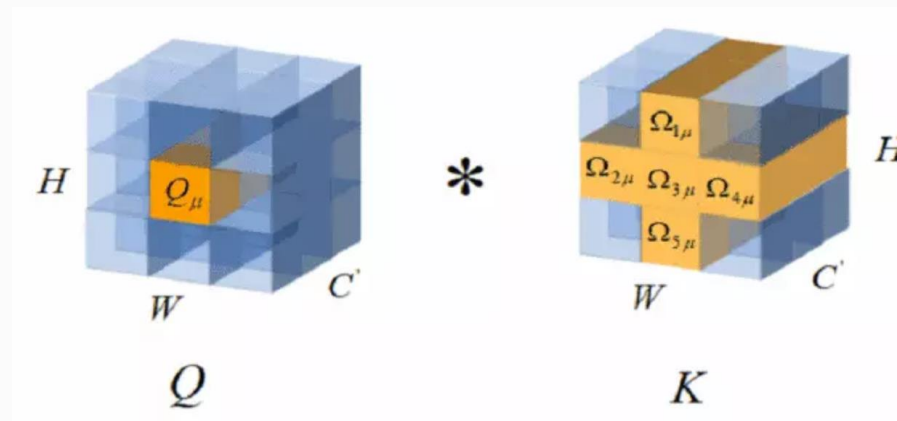
Loop 2



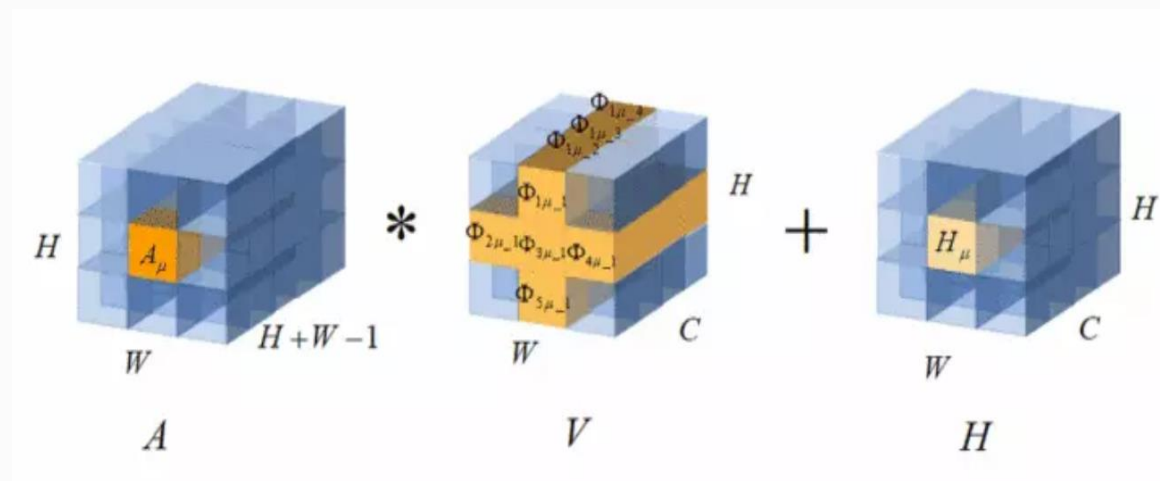
现有的智能分割的典型算法



Affinity操作



Aggregation 操作



PART 03



分类算法实例分析



分类算法实例分析



- 监督_PSPNet
- 弱监督_MDC
- 无监督

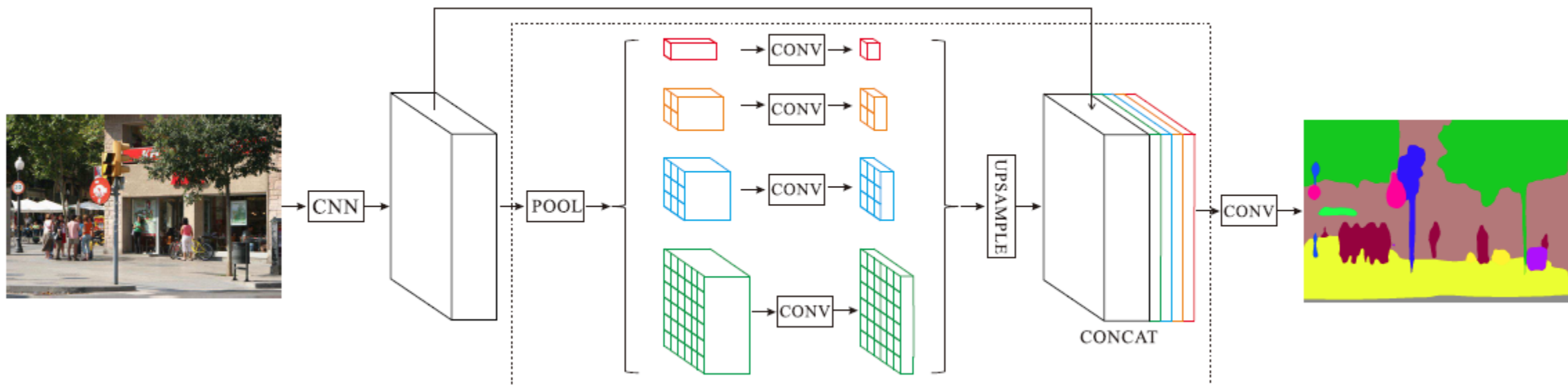


语义分割分为以下几个类型：

- ◆ **有监督语义分割**：进行语义分割模型训练时，数据集是有标签的。
- ◆ **弱监督/半监督语义分割**：在弱监督语义分割中，采用并不是精细标注的标签，而是粗糙标注的标签。
- ◆ **无监督/自监督语义分割**：无监督的语义分割不使用标签，而是通过算法迭代，不断地给具有相似语义的像素分配相同的标签。
- **自监督学习**通常分为两个步骤：无监督的预训练和对目标任务的微调。预训练步骤只需要数据本身，而无需数据的标注结果。它需要设计一个预训练任务，利用从数据本身到的的监督信号来训练。



有监督的语义分割网络PSPNet



Zhao, H. , Shi, J. , Qi, X. , Wang, X. , & Jia, J. . (2017). Pyramid Scene Parsing Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

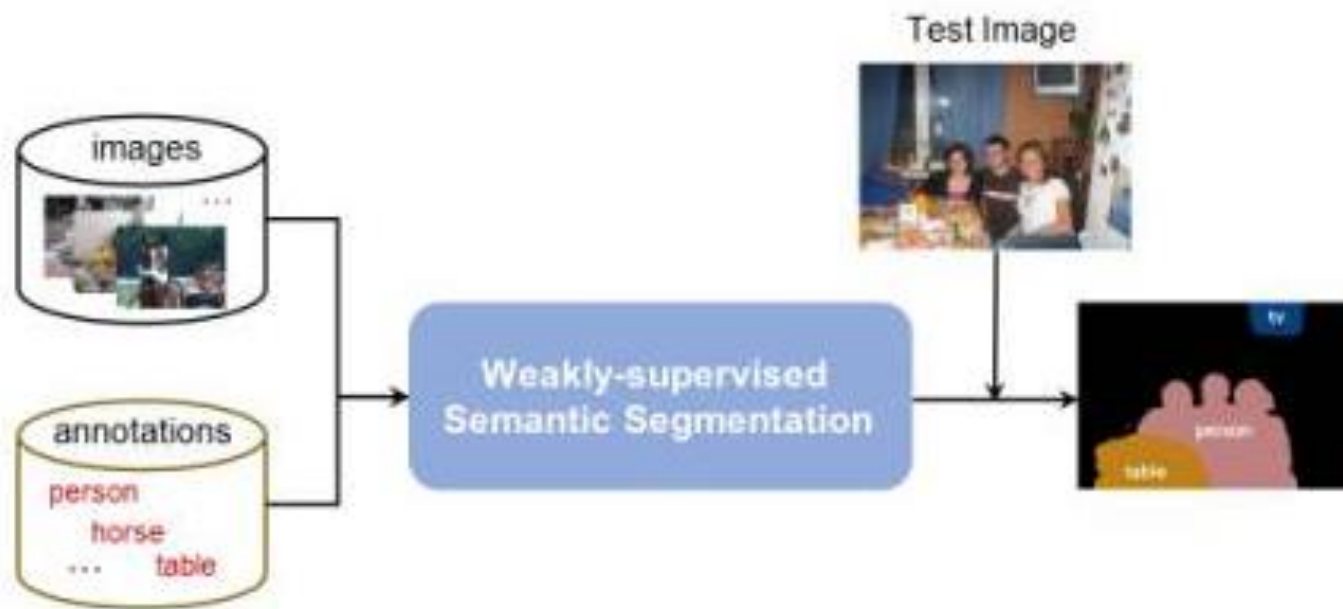


分类算法实例分析



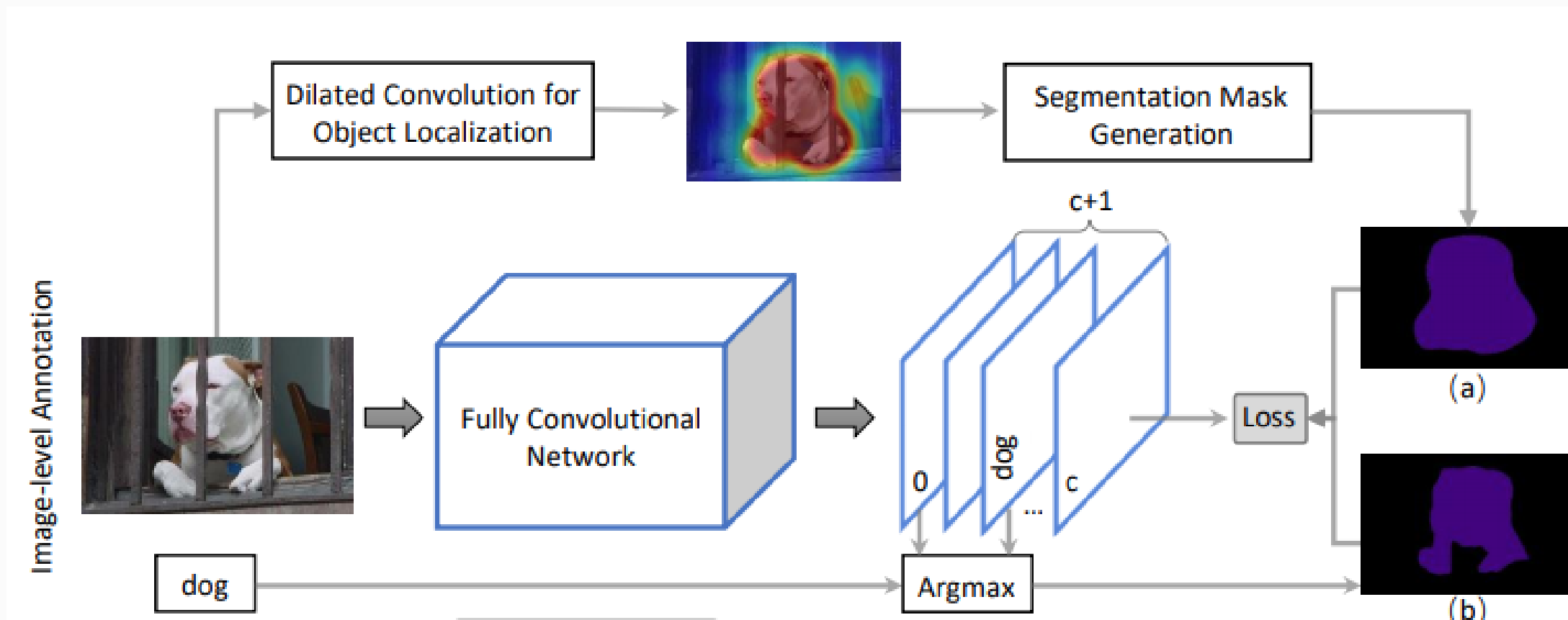
“弱监督学习”是指仅利用 image-level 的标注信息，通过弱监督方式学出一个模型，该模型可以预测出图像的语义分割结果。

- 这类弱监督学习的关键问题是：如何去构建 image-level 的标签语义和像素点的关联，推断出图像所对应的 segmentation mask，从而利用全卷积神经网络去学习分割模型。



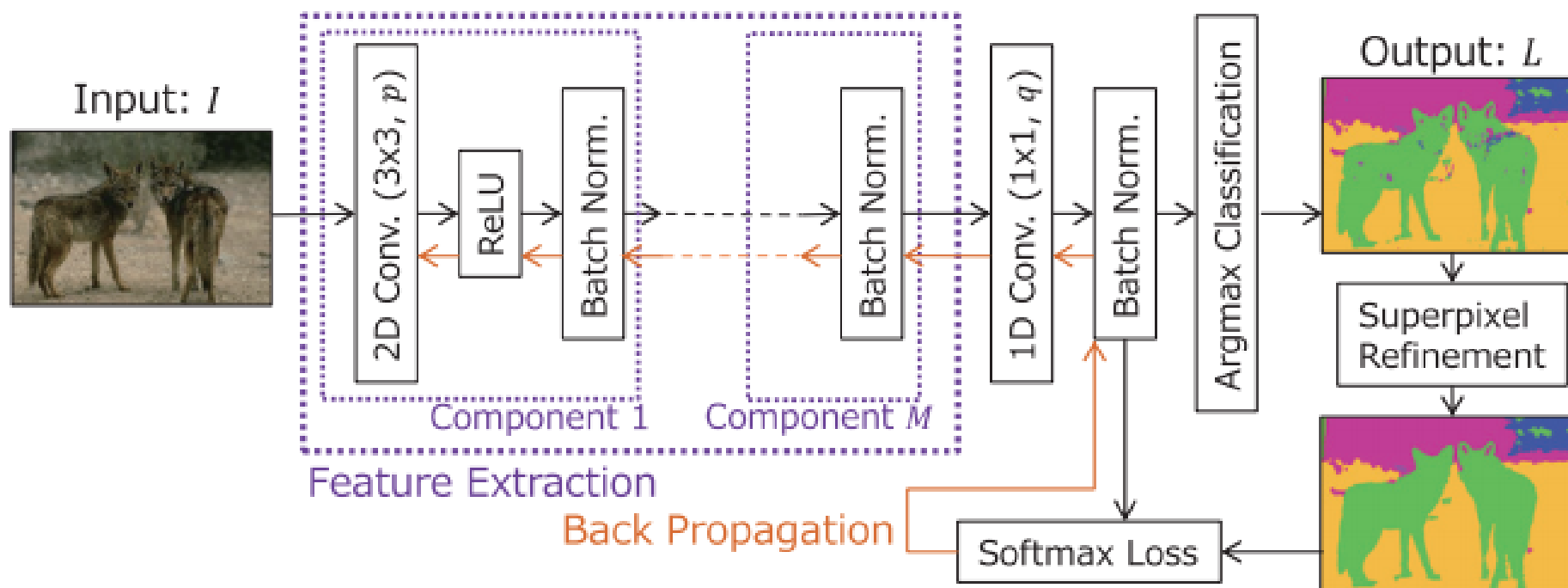


弱监督的语义分割网络MDC





无监督的语义分割网络



A. Kanezaki, "Unsupervised Image Segmentation by Backpropagation," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, 2018, pp. 1543-1547, doi: 10.1109/ICASSP.2018.8462533.



分类算法实例分析



算法：无监督图像分割 Unsupervised image segmentation

输入： $\mathcal{I} = \{i_n \in \mathbb{R}^3\}_{n=1}^N$ 输入RGB图片

输出： $\mathcal{J} = \{j_n \in \mathbb{Z}\}_{n=1}^N$ 输出语义分割的结果图片

$Net.parameter = Init(Xaiver)$ 初始化神经网络，保持每层的方差与均值

$\{\mathcal{S}_k\}_{k=1}^K = PreSeg(\{i_n\}_{n=1}^N)$ 对图片进行初步聚类

for $t = 1$ **to** T 迭代T次

$\{x_n\}_{n=1}^N = Net(\{i_n\}_{n=1}^N)$ 使用卷积网络得到特征图

$\{c_n\}_{n=1}^N = \{\arg \max x_n\}_{n=1}^N$ 根据特征图，取数值最大者为对应像素的标签

for $k = 1$ **to** K 对于经典语义分割聚类结果

$j_{max} = \arg \max |j_n|_{n \in \mathcal{S}_k}$ 统计每个聚类中，出现次数最多的类别

$c'_n = c_{max}$ **for** $n \in \mathcal{S}_k$ 将这个聚类中的所有像素，都记录为这个类别

$\mathcal{L} = Softmax(x_n, c'_n)$ 计算损失函数 (softmax竟然有中文名：归一化指数)

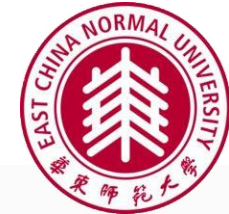
$Net.SDG(\mathcal{L})$ 使用随机梯度下降更新参数



(a) 网络推理过程

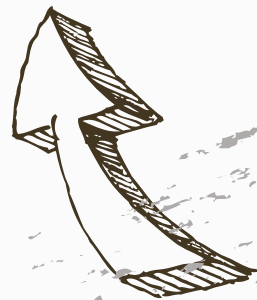


分类算法实例分析



- 先使用经典的机器学习算法，为输入图片进行**预分类**：调整算法参数，为语义信息明显相同的小区域分配相同的语义标签。由于具有相同语义的像素通常存在于一张图片中的连续区域。
- 然后使用深度学习结合自动编码器结构，对输入图片进行分类，分类的目标是：使输出的语义分割结果，尽可能地符合**预分类的结果**。训练到收敛。
- 最后，深度学习的语义分割结果，会在符合**预分类结果**基础上，将具备相同语义信息的小区块进行合并，得到大区块。

PART 04



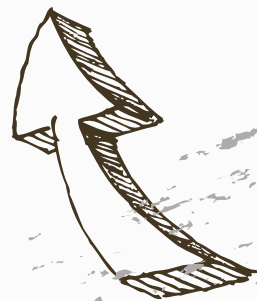
分割技术发展中的主要问题



语义分割中存在的几个难点：

- **数据问题：**分割不像分类等任务，只需要标注一个类别就可以拿来使用，分割需要精确的像素级标注。
- **计算资源问题：**如果要想得到较高的精度就需要使用深层次的网络。同时，分割预测了每一个像素，这就要求Feature map的分辨率尽可能的高，这都说明了计算资源的问题。
- **小目标分割：**查看Cityscape结果可以看出，很多算法的像道路、建筑物等类别，分割精度很高，能达到98%，而对于细小的类别，像行人、交通标志等等，由于其目标占比太小，而无法精确的定位轮廓，造成精度较低；

PART 05



分割技术应用实例



分割技术应用实例



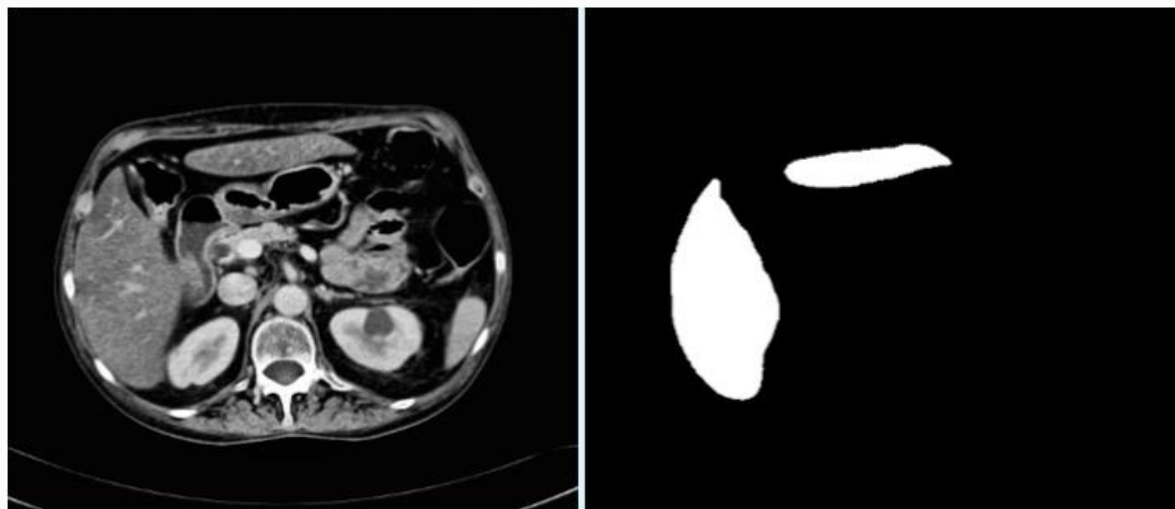
语义分割是计算机视觉中的基本任务，在语义分割中我们需要将视觉输入分为不同的语义可解释类别。目前语义分割的应用领域主要有地理信息系统，无人车驾驶，医疗影像分析，机器人、商业等领域。



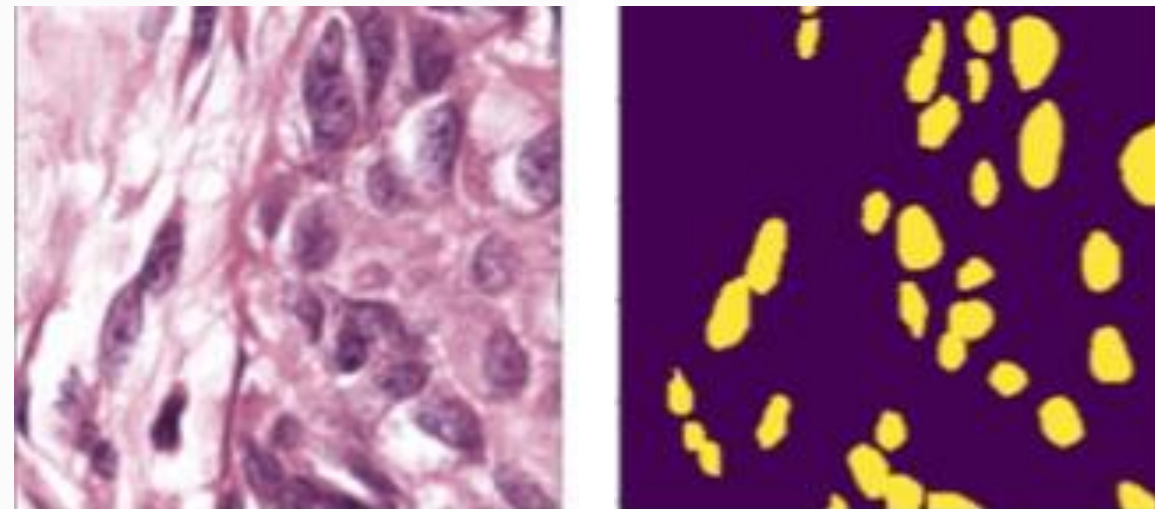
(a) 无人驾驶汽车



分割技术应用实例



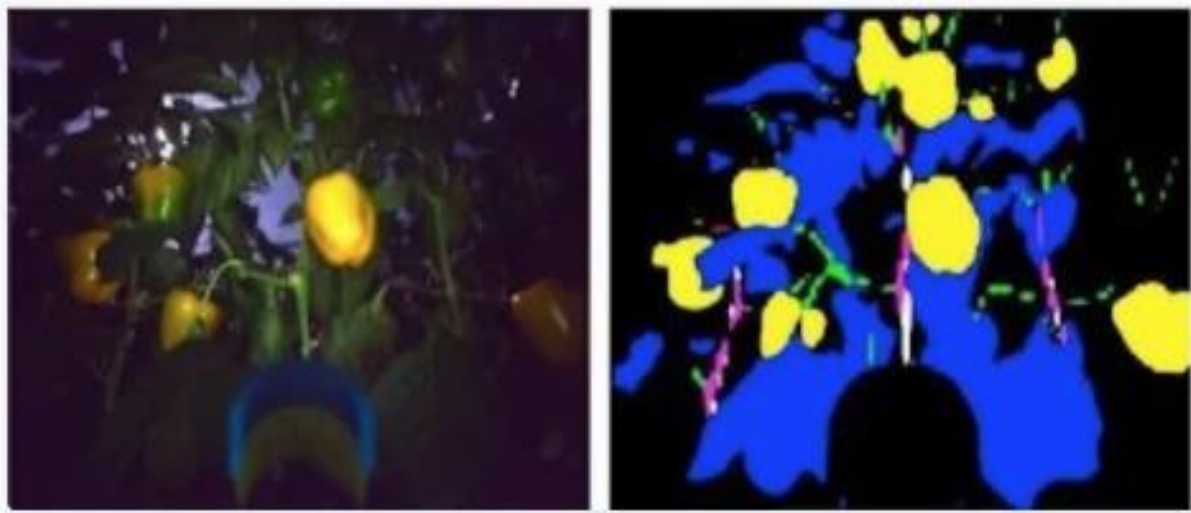
(a) 腹部CT提取肝脏



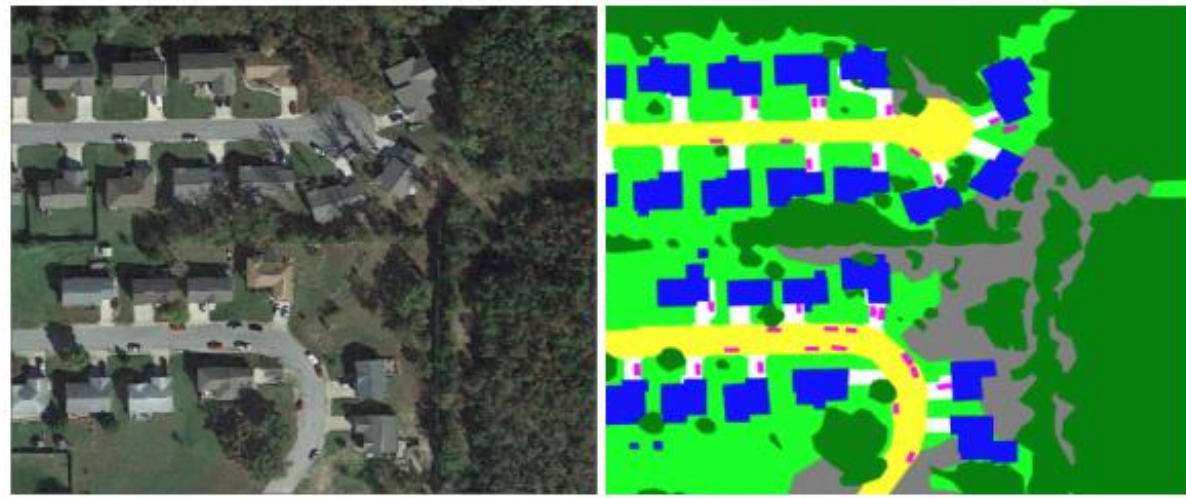
(b) H&E 染色细胞核分割



分割技术应用实例



(a) 精准农业_农田中的作物提取



(b) 遥感图像地物分离



分割技术应用实例



(a) 面部分割



(b) 时尚——分类服装



Assignment14

- 查阅文献，完成本章内容的《计算机视觉课程报告》
- 完成《计算机视觉实验报告-综合版》中的实践《智能图像分割》



THANK YOU