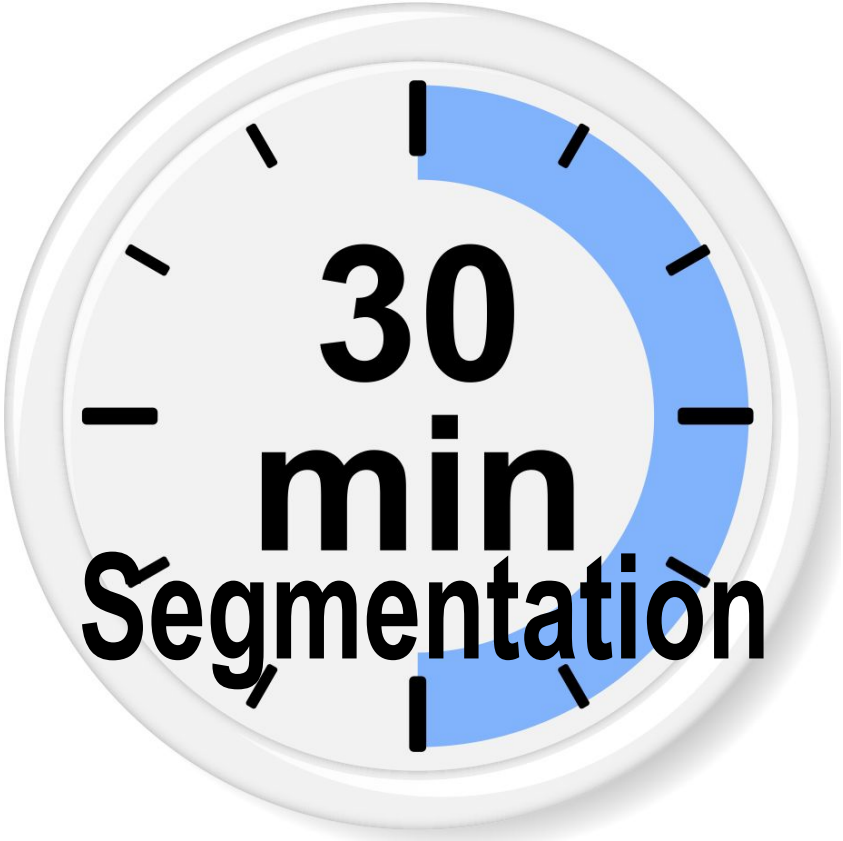
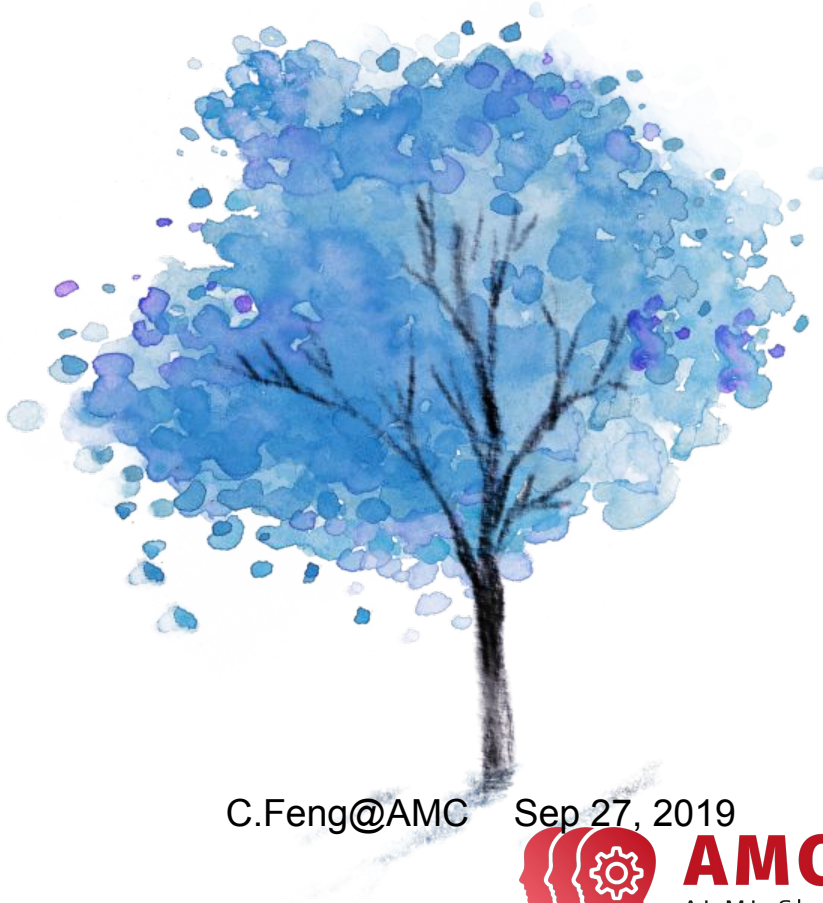




C.Feng@AMC Sep 27, 2019

AMC
AI-ML.Club

0
预备知识

AMC
AI-ML.Club

计算机视觉

(Computer Vision, CV)

图像分类 (Image Classification)

物体检测 (Object Recognition)

语义分割 (Semantic Segmentation)

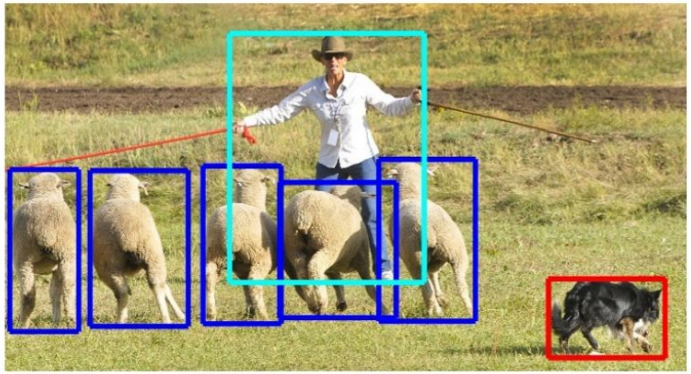
视频分析 (Video Analysis)

人脸支付

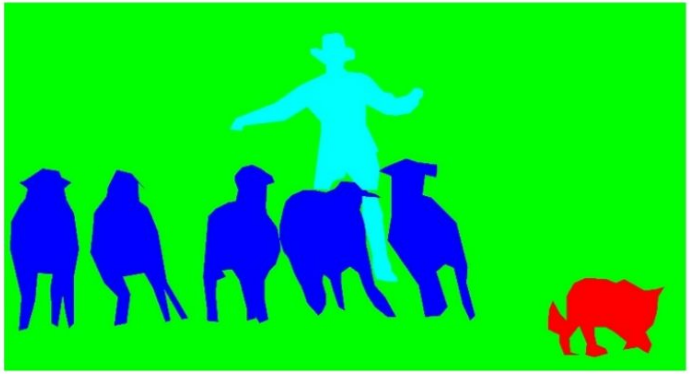
自动驾驶

医疗影像
辅助诊断

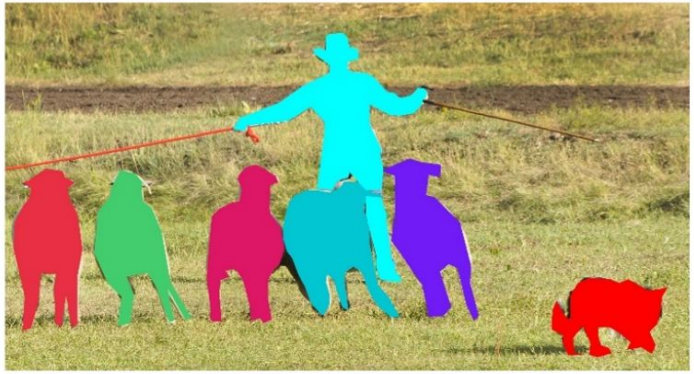
(a) Image classification



(b) Object localization



(c) Semantic segmentation

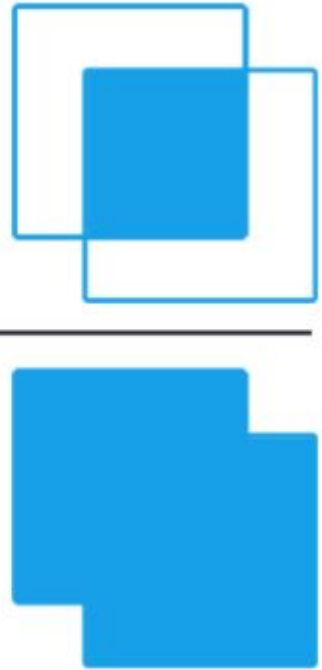


(d) This work

常用评价指标(IoU/mIoU/PA)

- ★ IoU: 交并比
- ★ mIoU: 基于类进行计算的IoU就是将每一类的IoU计算之后累加，再进行平均，得到的就是基于全局的评价
- ★ PA(Pixcal-accuracy, 像素精度): 从字面上理解就可以知道，PA是指预测正确的像素占总像素的比例

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$



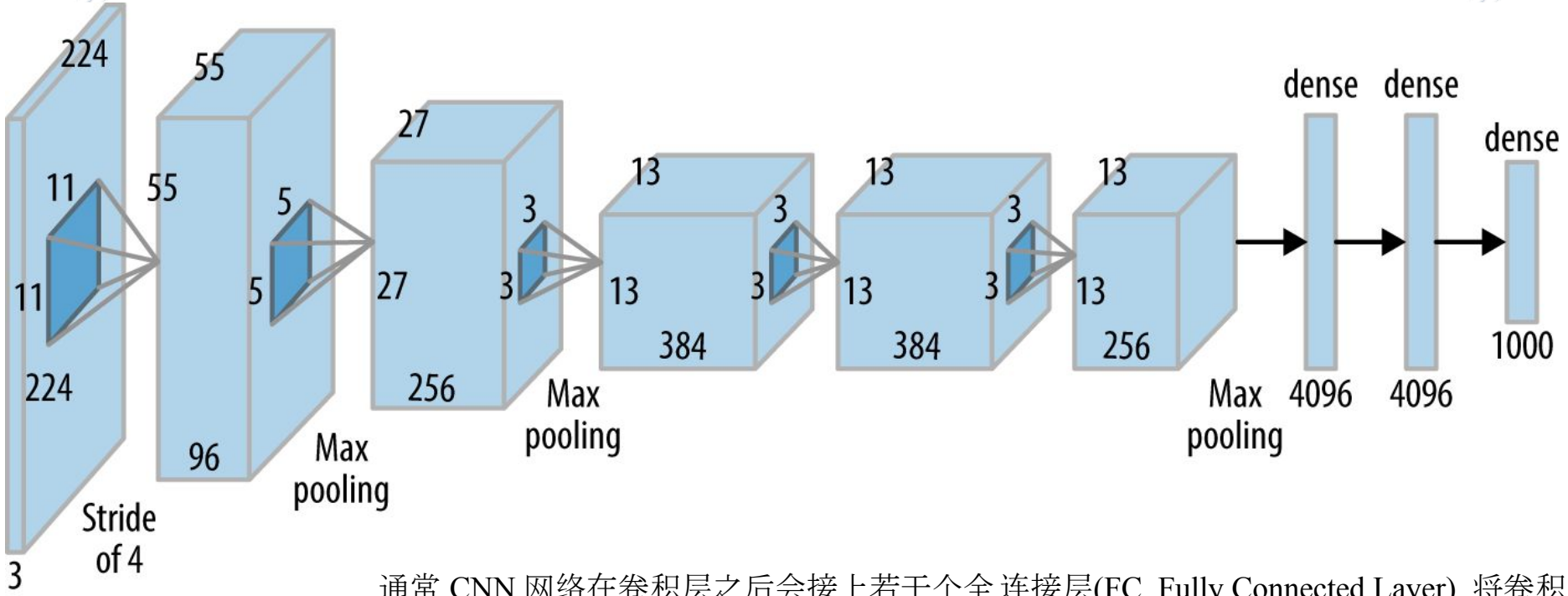
1
Why FCN
[a.k.a. Why NOT AlexNet]

为什么分类网络无法用于分割任务

AMC
AI-ML.Club

Krizhevsky, Alex et al. “ImageNet Classification with Deep Convolutional Neural Networks.” Commun. ACM 60 (**2012**): 84-90.

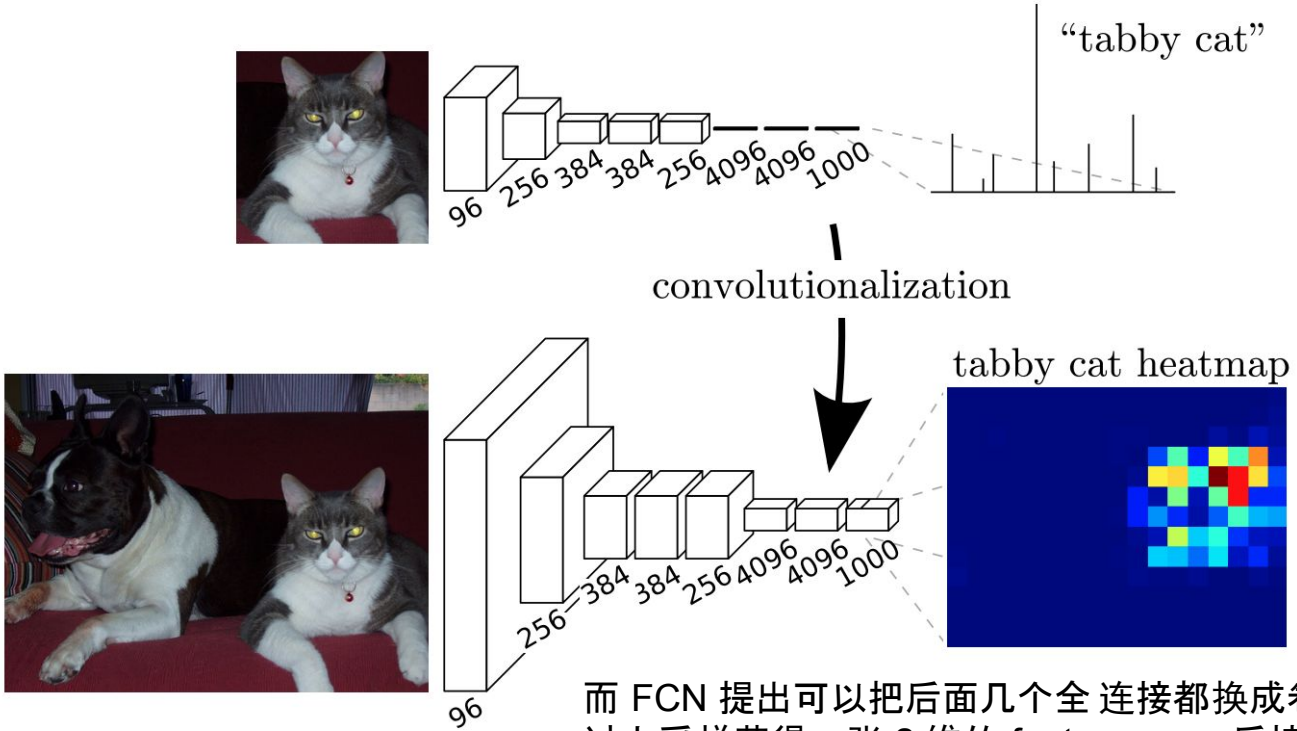
回顾AlexNet-like分类网络



通常 CNN 网络在卷积层之后会接上若干个全连接层(FC, Fully Connected Layer), 将卷积层产生的特征图 (feature map) 映射成一个固定长度的特征向量, 经过 softmax 后就可以获得类别概率信息.但是这个概率信息是 1 维的, 即只能标识整个图片的类别, 不能标识每个像素点的类别, 所以这种全连接方法不适用于图像分割.

Shelhamer, Evan et al. “Fully Convolutional Networks for Semantic Segmentation.” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (**2014**): 3431-3440.

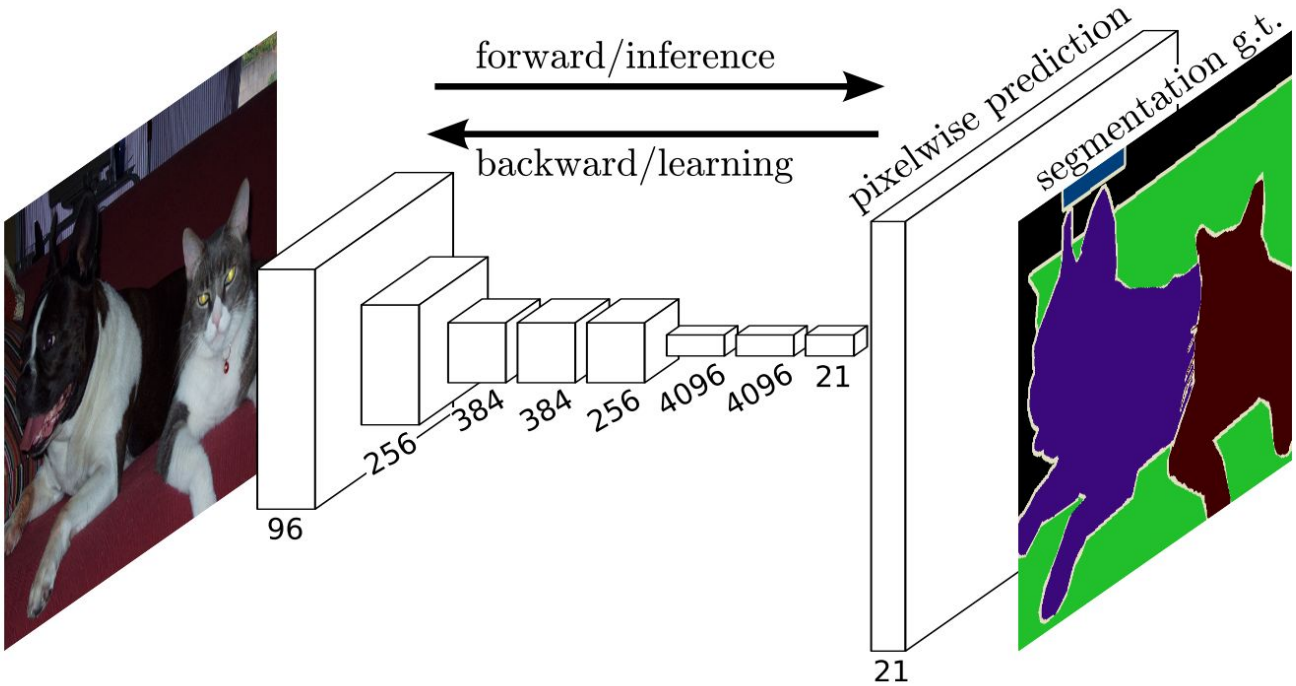
语义分割开山之作 FCN



而 FCN 提出可以把后面几个全连接都换成卷积.这样就可以经过上采样获得一张 2 维的 feature map, 后接 softmax获得每个像素点的分类信息, 从而解决了语义级别的图像分割问题.

Shelhamer, Evan et al. “Fully Convolutional Networks for Semantic Segmentation.” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (**2014**): 3431-3440.

语义分割开山之作 FCN

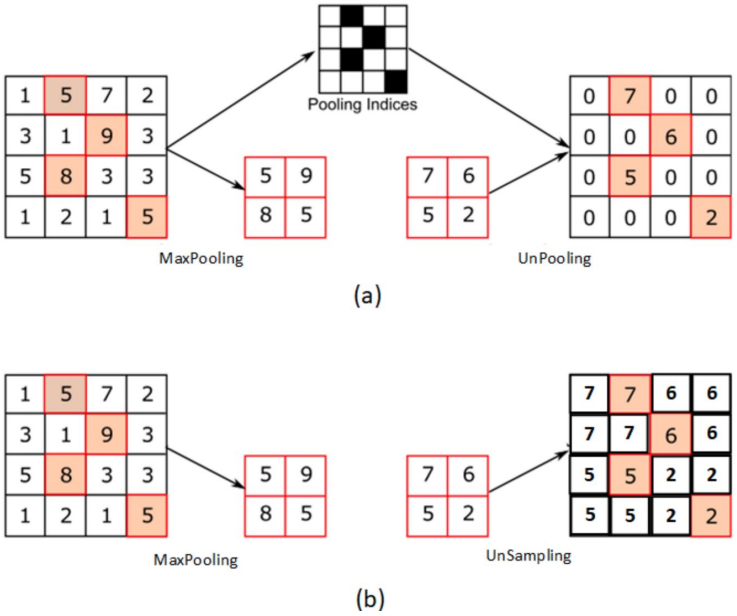


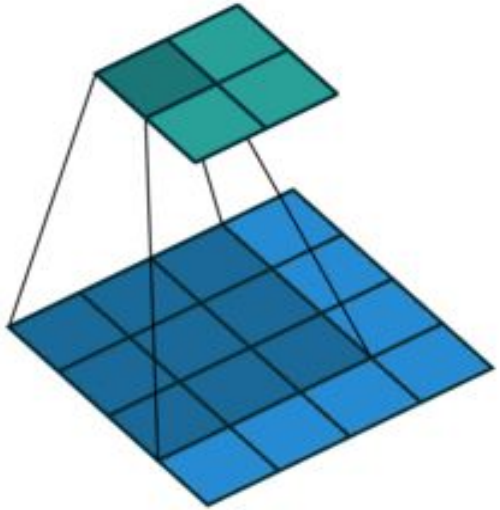
FCN在分割过程中能够恢复像素所属的类别, 极大地推动了语义分割领域的发展.然而该领域仍然存在两个问题:

1. 图像经过池化操作后, 特征图的**分辨率不断降低**, 部分像素的空间位置信息丢失;
2. 分割过程未能有效地**考虑上下文(image context)信息(图像上下文/标签上下文)**, 无法充分利用丰富的空间位置信息, 导致局部特征和全局特征的利用率失衡.

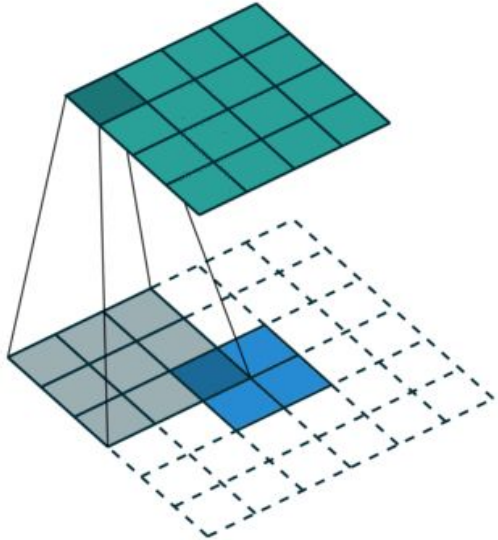
FCN未能有效地解决这两个问题, 致使分割结果粗糙、分割边界不连续.

扩展: 上池化/上采样/转置卷积





Vanilla
Convolution



Transposed
Convolution

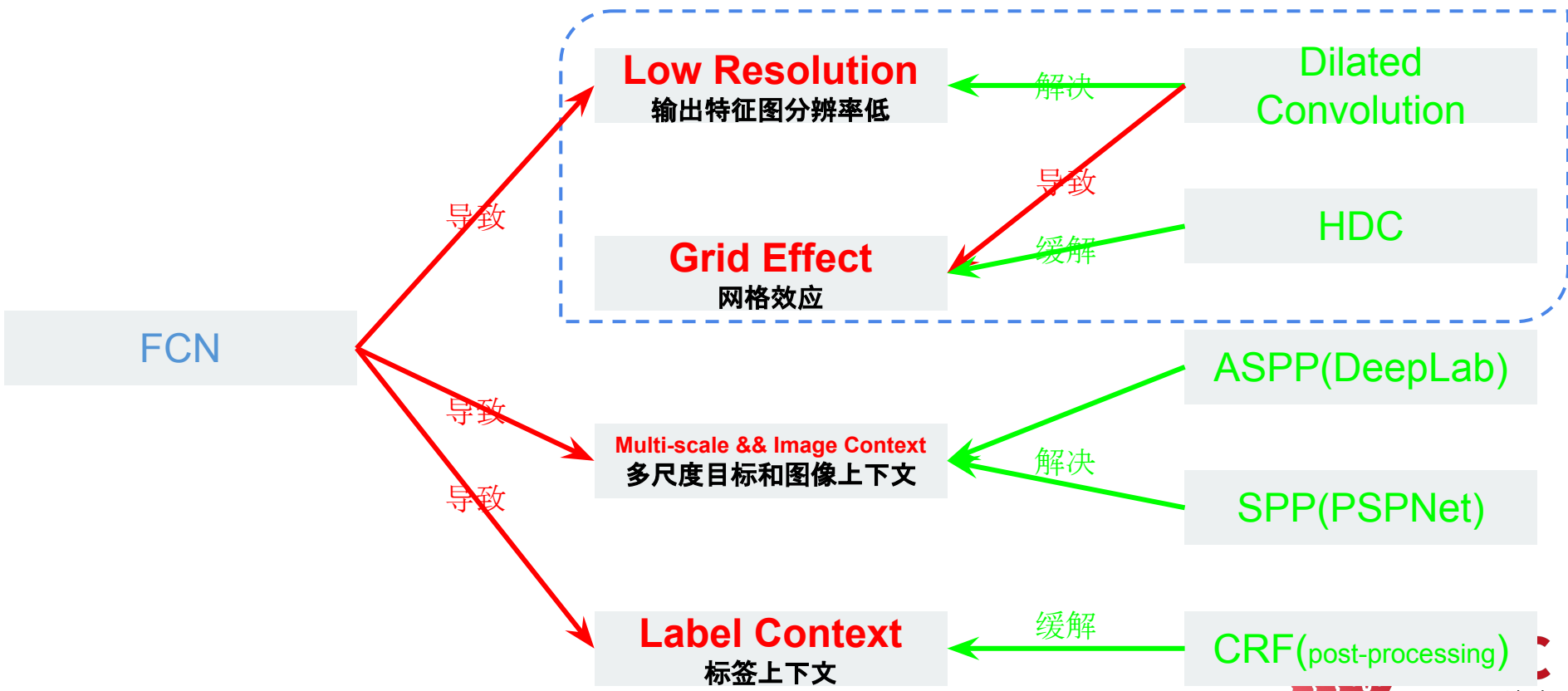
AMC
AI-ML.Club

2
Domain Issues && Solutions

图像分割领域的问题和他们的解决方案

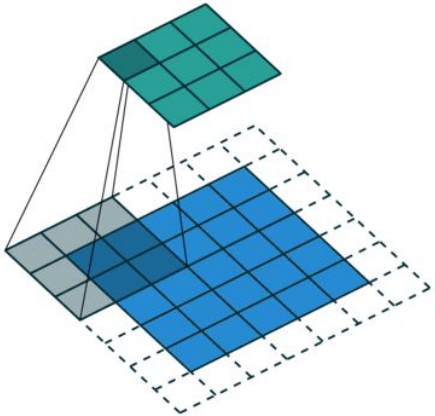
AMC
AI-ML.Club

领域问题和解决方案(Q&A)

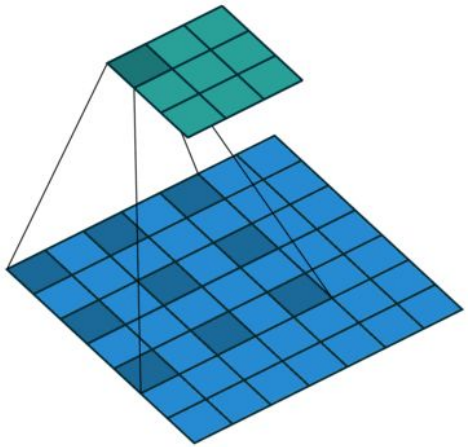


Yu, Fisher, and Vladlen Koltun. "Multi-scale context aggregation by dilated convolutions." 2015 The International Conference on Learning Representations (ICLR) (**2015**)

让分辨率不再下降: Dilated Convolution



Vanilla Convolution



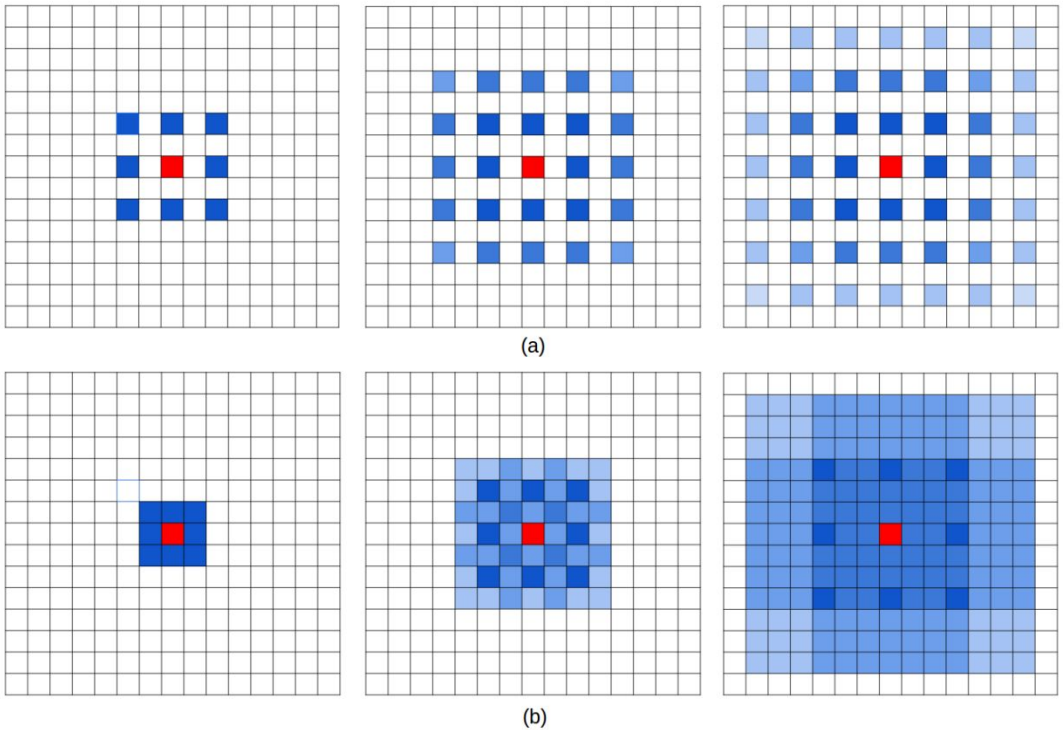
Dilated Convolution

然而 Deep CNN 对于其他任务还有一些致命性的缺陷. 较为著名的是 up-sampling 和 pooling layer 的设计. 这个在 Hinton 的演讲里也一直提到过. 主要问题有:

1. Up-sampling / pooling layer (e.g. bilinear interpolation) is deterministic. (a.k.a. not learnable)
2. 内部数据结构丢失; 空间层级化信息丢失.
3. 小物体信息无法重建 (假设有四个 pooling layer 则任何小于 2 4 = 16 pixel 的物体信息将理论上无法重建.)

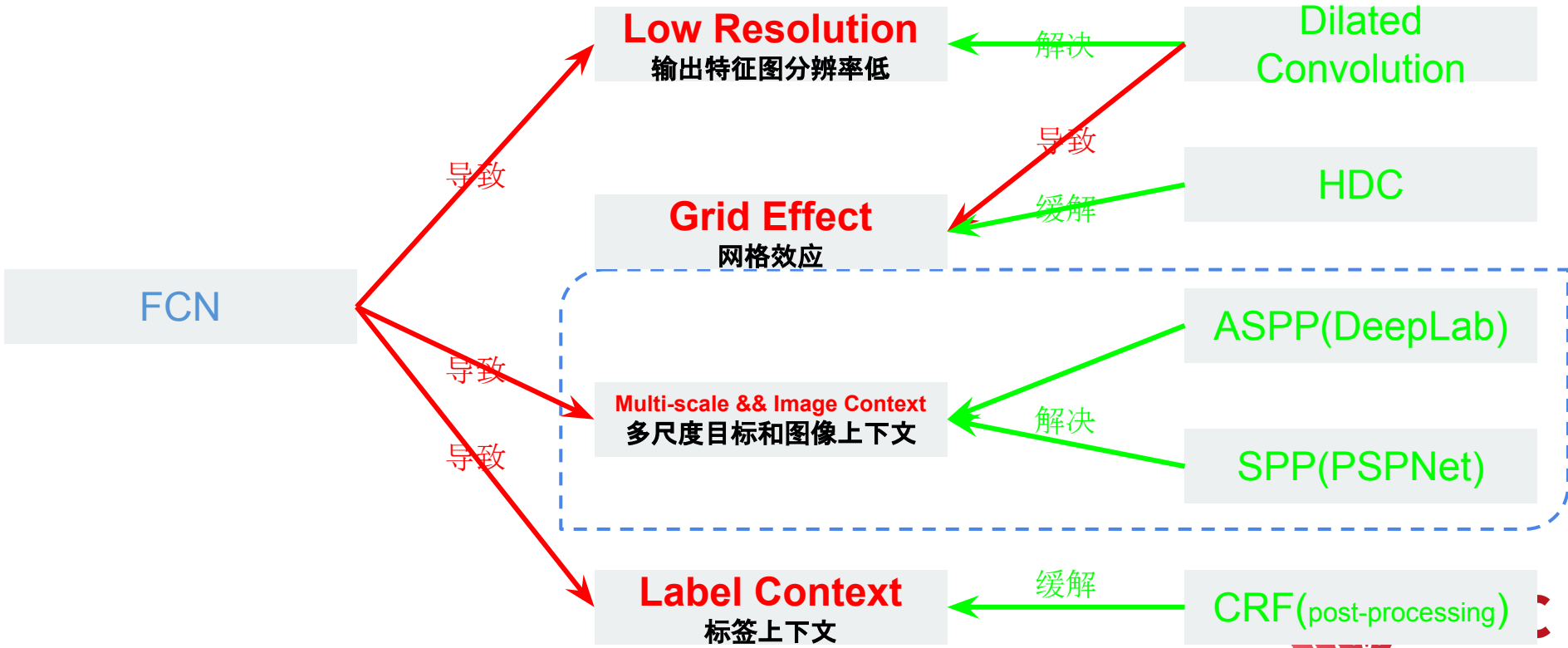
在这样问题的存在下, 语义分割问题一直处在瓶颈期无法再明显提高精度, 而 dilated convolution 的设计就良好的避免了这些问题.

扩展: The Gridding Effect and Hybrid Dilated Convolution



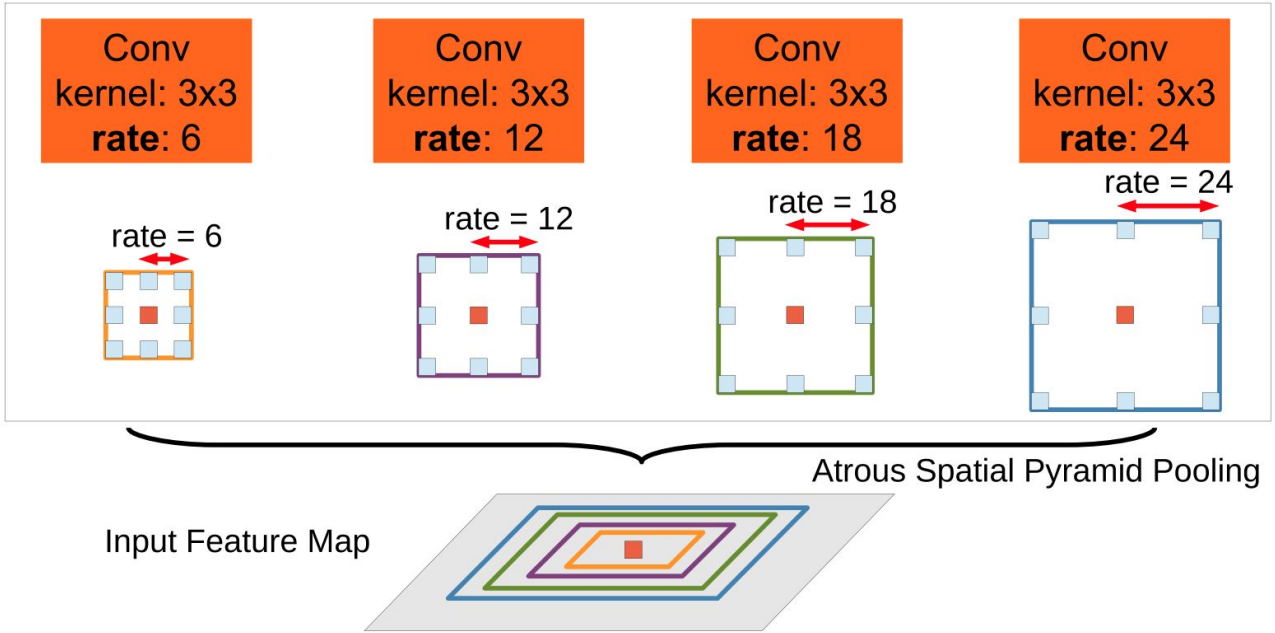
假设我们仅仅多次叠加 dilation rate 2 的 3 x 3 kernel 的话，我们会发现我们的 kernel 并不连续，也就是并不是所有的 pixel 都用来计算了。因此这里将信息看做 checker-board 的方式会损失信息的连续性，这对 pixel-level dense prediction 的任务来说是致命的。对于这个问题，图森组的文章中设计了一个称之为 HDC 的设计结构，并通过对比实验指出一个良好设计的 dilated convolution 网络能够有效避免 gridding effect。

领域问题和解决方案(Q&A)



Chen, Liang-Chieh et al. “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs.” IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (**2016**): 834-848.

使用Atrous Spatial Pyramid Pooling(ASPP)进行多尺度分割

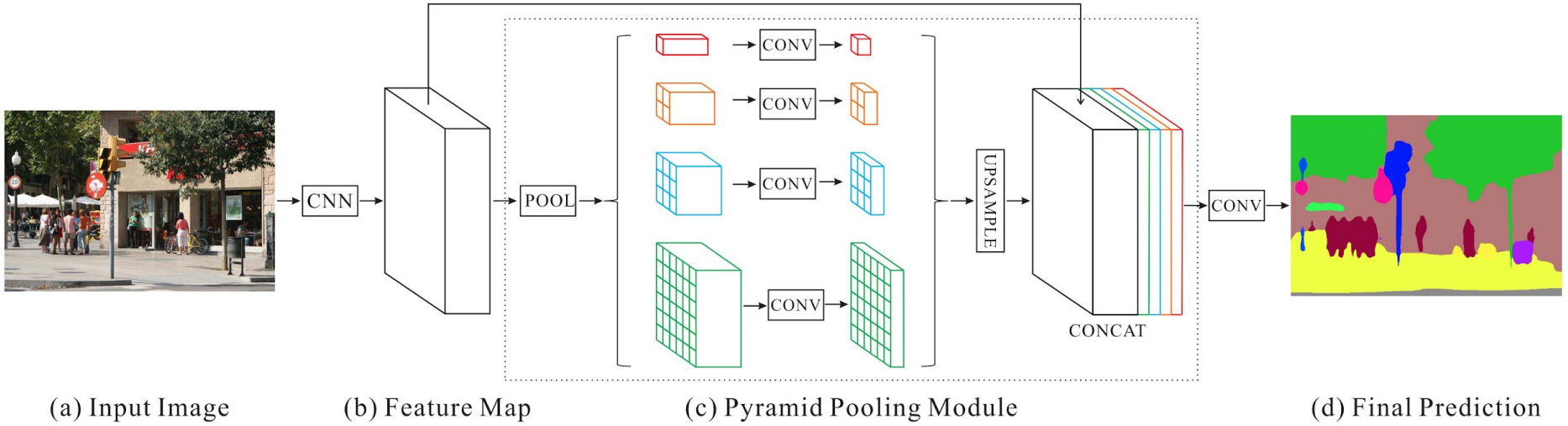


然仅仅 (在一个卷积分支网络下) 使用 dilated convolution 去抓取多尺度物体是一个不正 统的方法. 比方说, 我们用一个 HDC 的方法来获取一个大/近车辆的信息, 然而对于一个小/远车辆的信息都不再受用. 假设我们再去用小 dilated convolution 的方法重新获取小车辆的信息, 则这么做非常的冗余。

DeepLab在多尺度上为分割对象进行带洞空间金字塔池化 (ASPP),其在特征顶部映射图使用了四中不同采样率的空洞卷积. 这在一定程度上解决了多尺度 (Multi-scale)问题。

Zhao, Hengshuang et al. “Pyramid Scene Parsing Network.” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (**2016**): 6230-6239.

层次化全局语义信息的融合 Pyramid Pooling Module

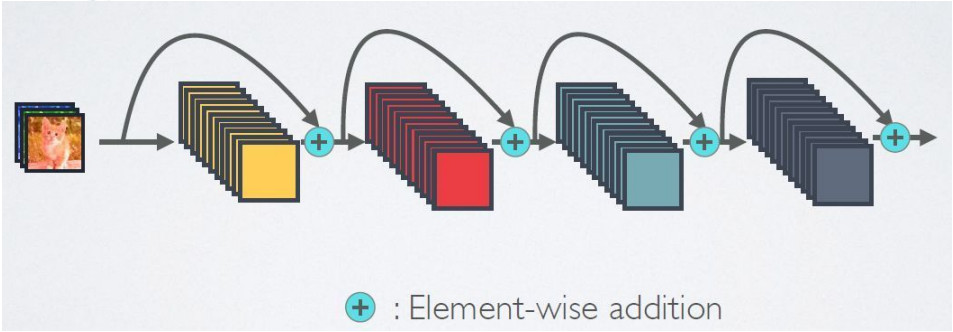


作者认为现有模型由于没有引入足 够的上下文信息及不同感受野下的全局信息而存在分割出 现错误的情景，于是提出的金字塔池化模 块 (pyramid pooling module) 能够聚合不同区域的上下文信息 ,从而提高 获取全局信息的能力。

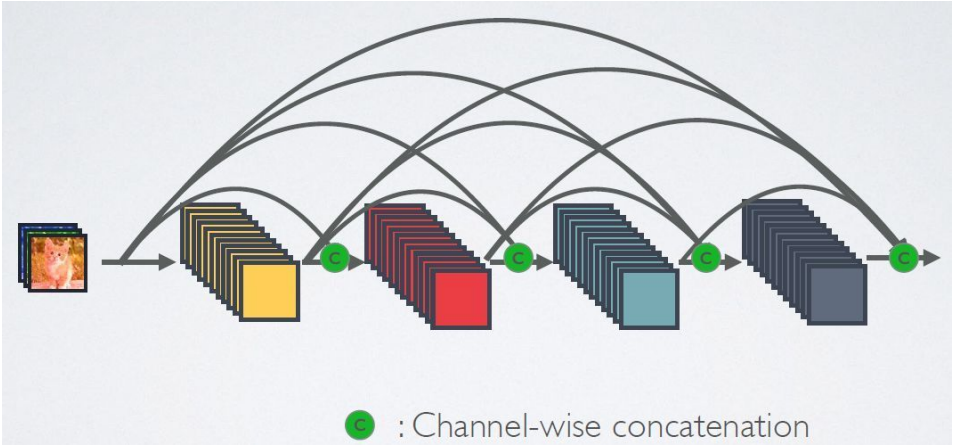
本文提出的网 络结构简单来说就是将DeepLab ASPP之前的Feature Map进行四种尺度的pooling之后，将5种feature map concat到一起经过卷积最后进行prediction的过程。

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 (**2015**): 234–241. Crossref. Web.

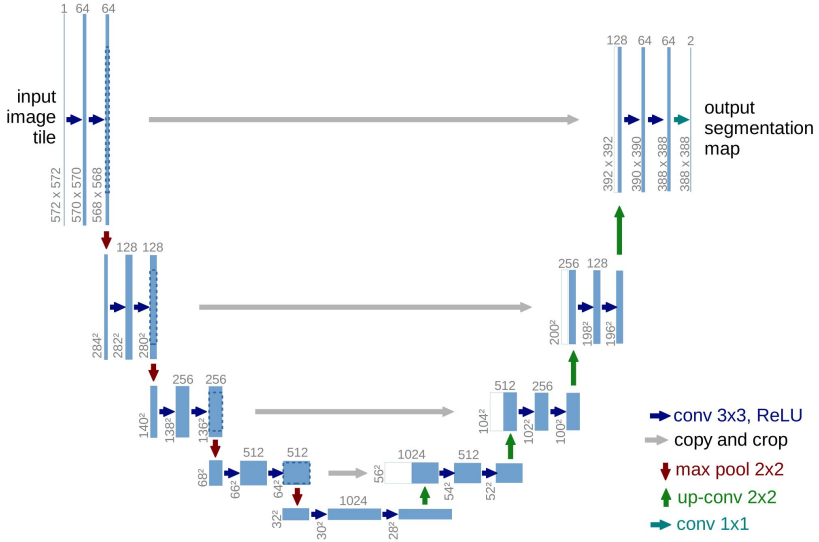
扩展: 另类特征融合方式 U-Net



ResNet网络的短路连接机制(其中+代表的是元素级相加操作)



DenseNet网络的密集连接机制(其中c代表的是channel级连接操作)

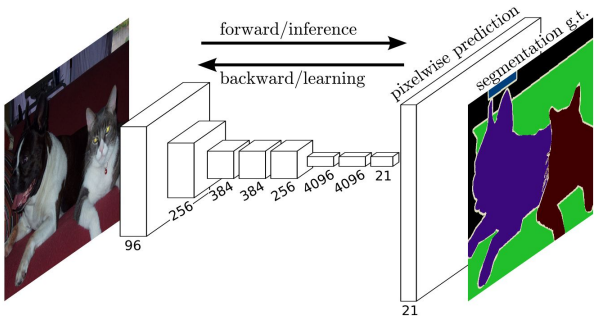


在**相加**的方式下,feature map 的维度没有变化, 但每个维度都包含了更多特征, 对于普通的分类任务这种不需要从 feature map 复原到原始分辨率的任务来说, 这是一个高效的选择.

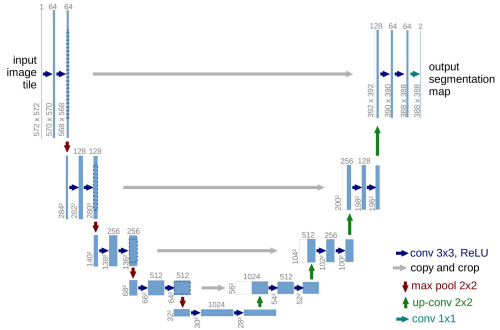
而**拼接**则保留了更多的维度/位置信息, 这使得后面的 layer 可以在浅层特征与深层特征自由选择, 这对语义分割任务来说更有优势.

Wu, Huikai et al. “FastFCN: Rethinking Dilated Convolution in the Backbone for Semantic Segmentation.” ArXiv abs/1903.11816 (2019): n. pag.

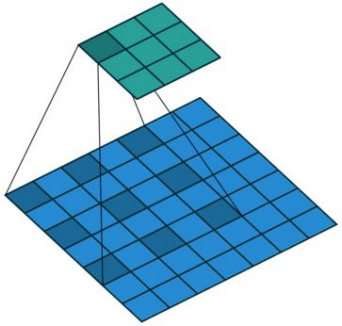
扩展: 可视化不同形态的主干网络



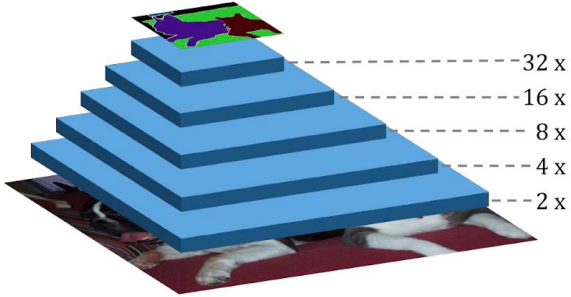
(a) FCN [22]

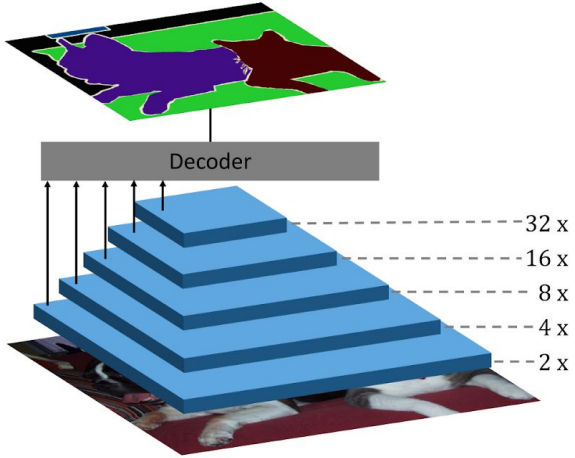


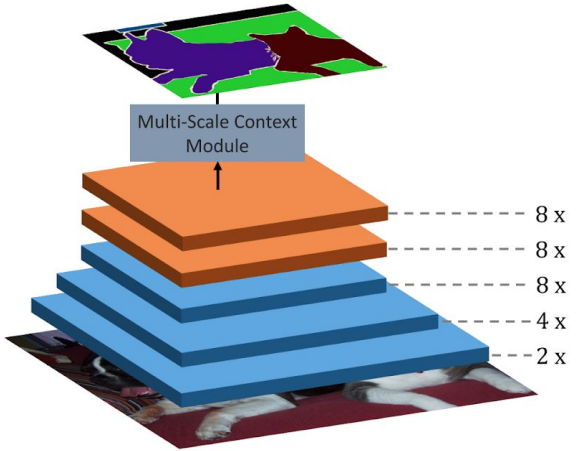
(b) EncoderDecoder



(c) DilatedFCN





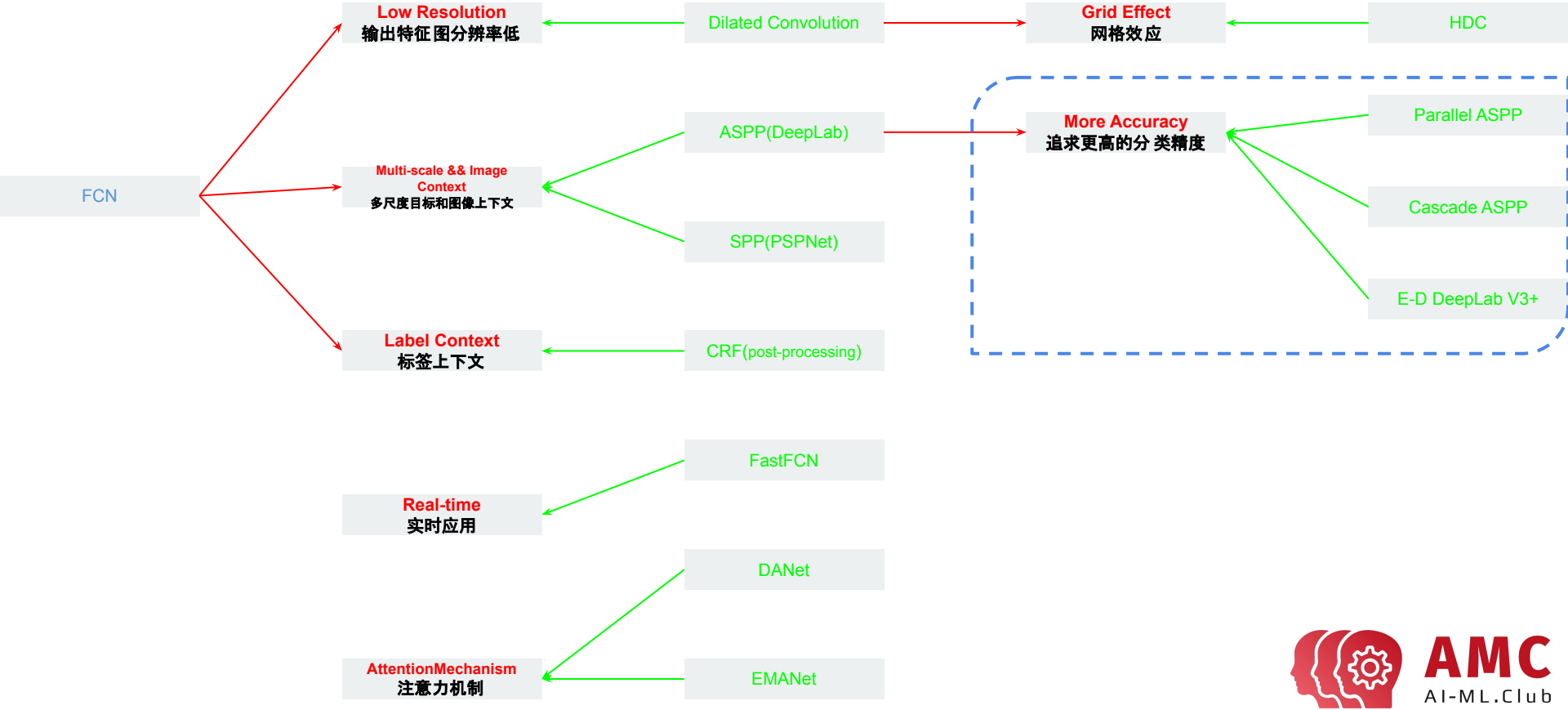


3
More Issues && Solutions

更多的问题和探索方向

AMC
AI-ML.Club

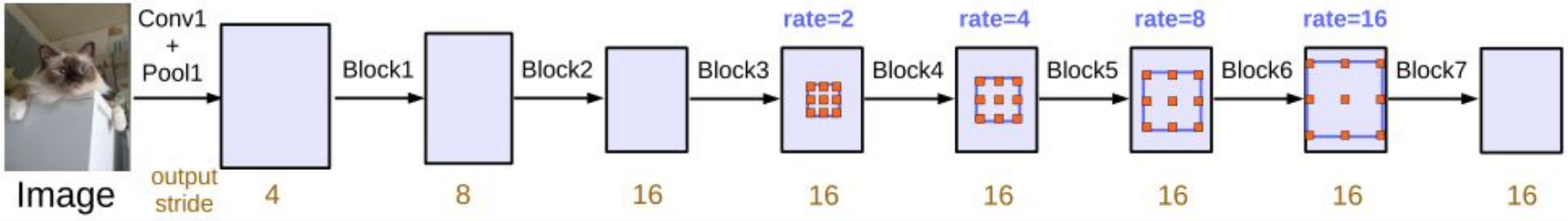
更多领域问题和解决方案(Q&A)

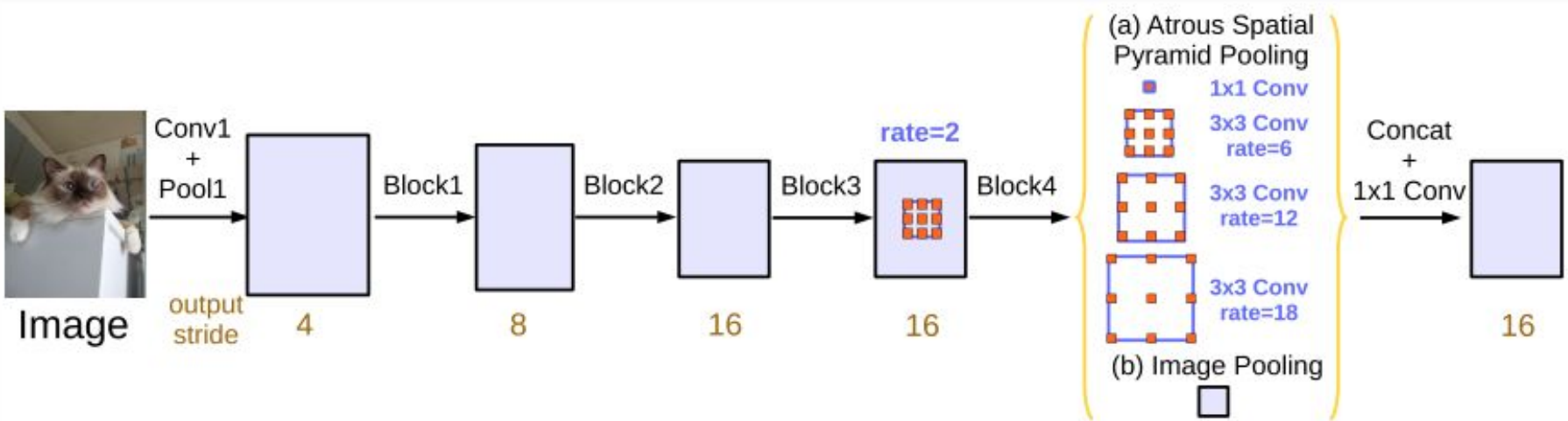


AMC
AI-ML.Club

Chen, Liang-Chieh et al. “Rethinking Atrous Convolution for Semantic Image Segmentation.” ArXiv abs/1706.05587 (**2017**): n. pag.

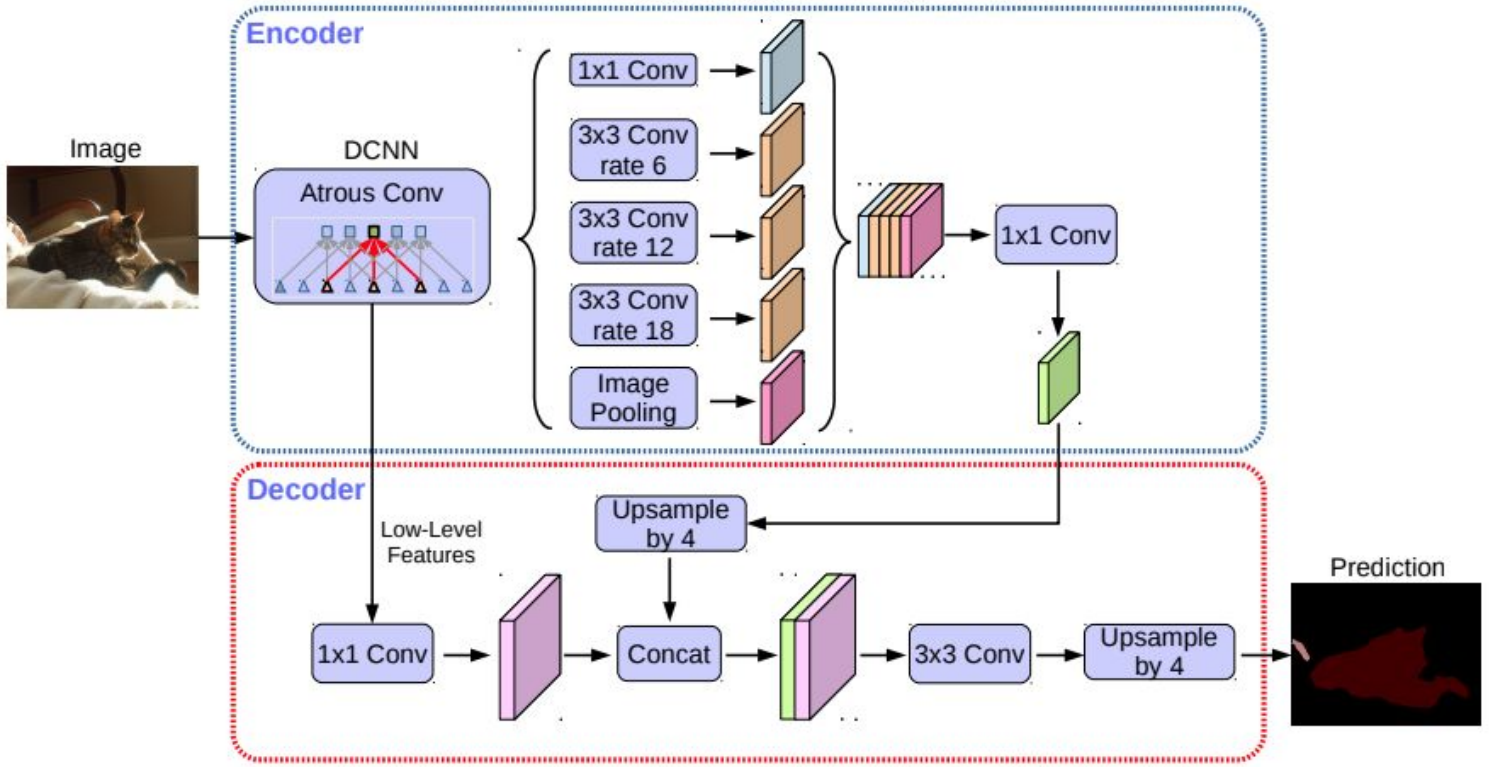
追求更高效的多尺度上下文: DeepLab V3



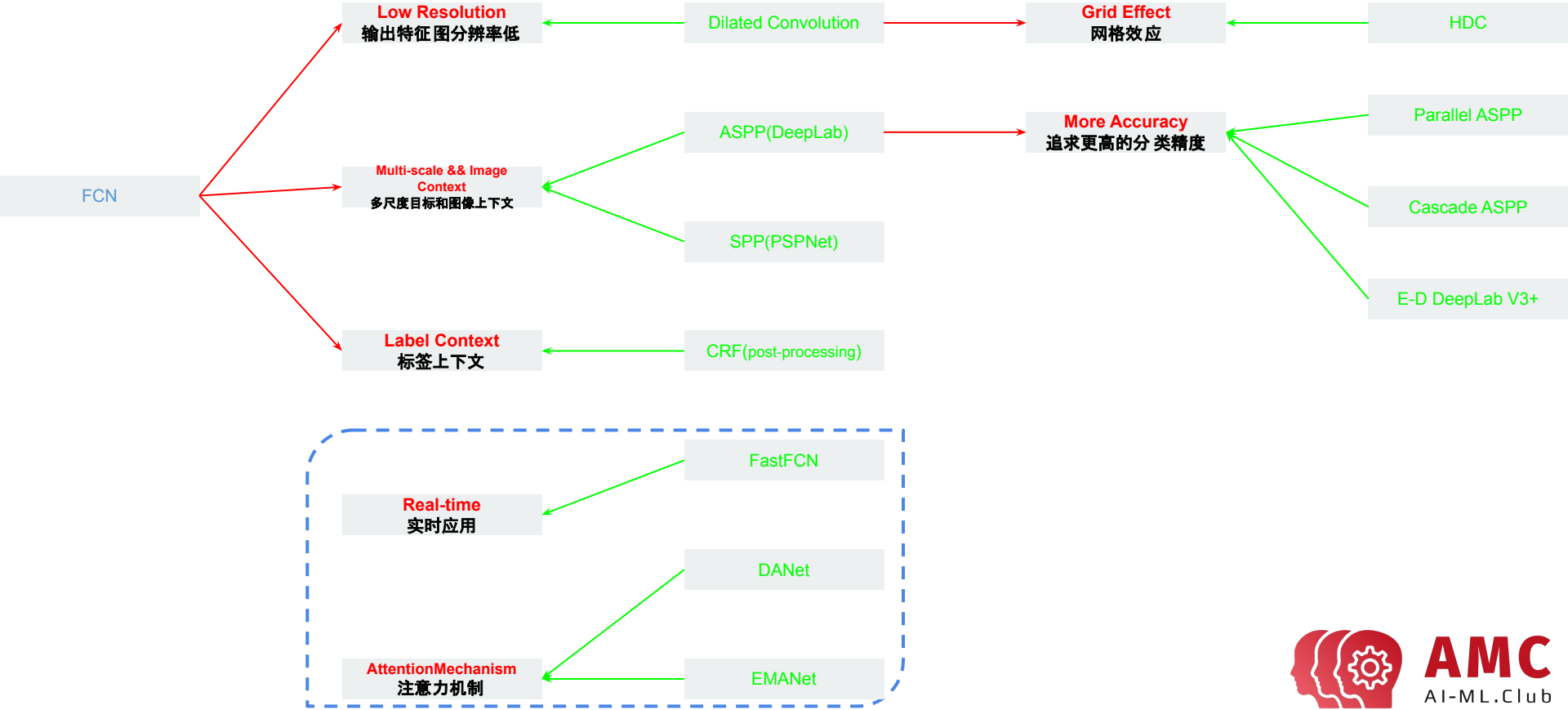


Chen, Liang-Chieh et al. “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation.” ECCV (**2018**).

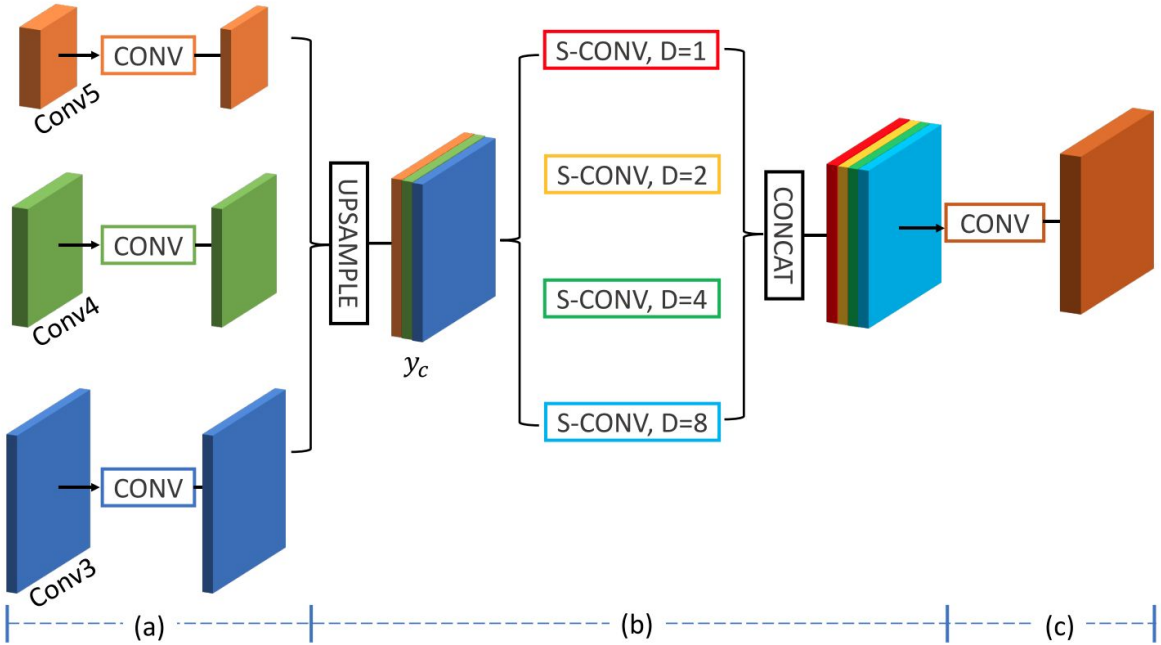
基于编解码结构的高效多尺度上下文模块: DeepLab V3+



更多领域问题和解决方案(Q&A)



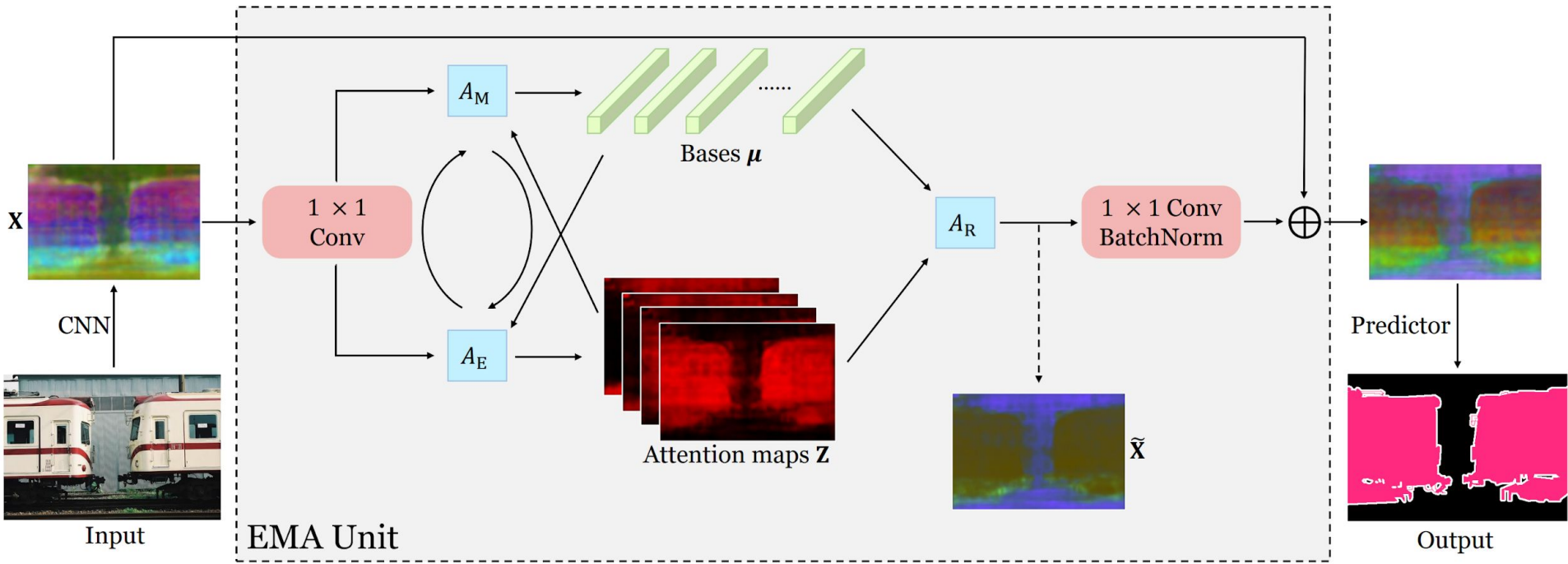
空洞卷积如何做到实时和轻量化: Joined Pyramid Upsampling (FastFCN)



现代语义分割方法通常会在主干中部署空洞卷积来提取高分辨率特征图, 这带来了极大的计算复杂度和内存占用. 为了取代耗时又耗内存的空洞卷积, 研究者提出了新的联合上采样模块 (Joint Pyramid Upsampling, JPU) 将提取高分辨率特征图的任务化为了联合上采样问题, 它在不影响性能的情况下, 将计算复杂度降低了三倍多. 实验表明,JPU在多个数据集上达到SOTA同时运行速度快了三倍.

Fu, Jun et al. “Dual Attention Network for Scene Segmentation.” *2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019)
Li, Xia et al. “Expectation-Maximization Attention Networks for Semantic Segmentation.” ICCV (2019)

将注意力机制应用于语义分割中: DANet/EMANet



AMC
AI-ML.Club

End.

AMC
AI-ML.Club