# FastFCN

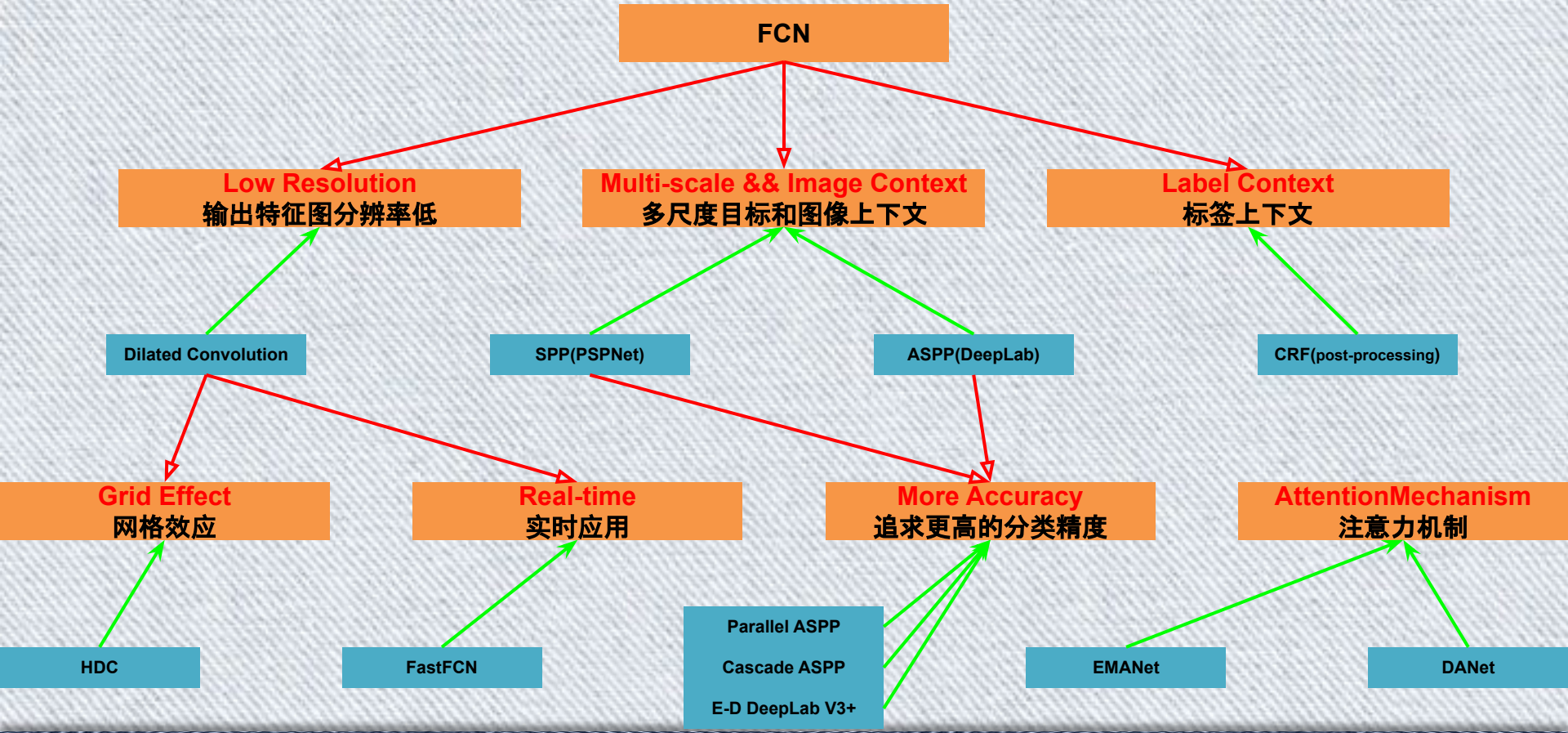
C.Feng@AMC Oct 01, 2019

Wu, H., Zhang, J., Huang, K., Liang, K., & Yu, Y. (2019).

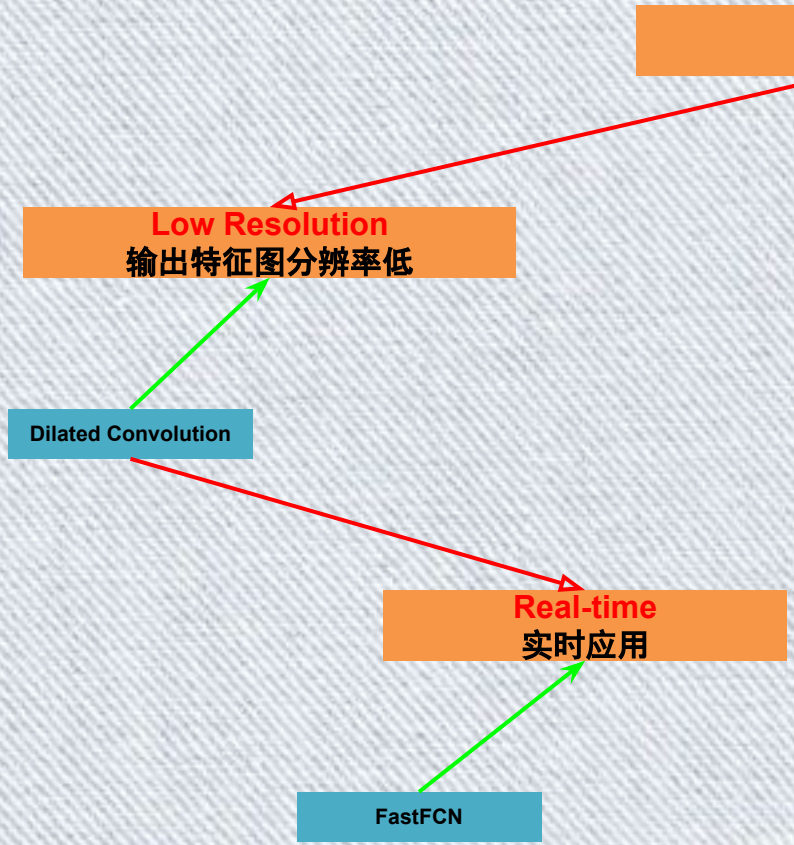
FastFCN: Rethinking Dilated Convolution in the Backbone for Semantic Segmentation.

ArXiv, abs/1903.11816.

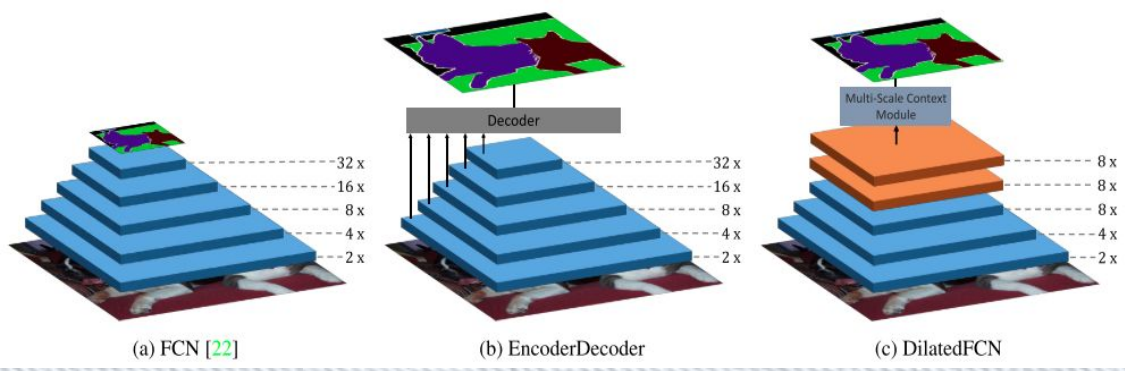
# 分割领域多种问题和其解决方案(Q&A)



# 空洞卷积计算实时性解决方案(Q&A)



# Why




DeepLab 方法移除FCN最后两层下采样操作并引入扩张卷积来保持特征图感受野不变,后跟一个多尺度的语义模块从而得到最终效果, 如图(c) 所示, 其中扩张卷积在保持最终特征图的分辨率方面作用明显, 大大提升了编解码语义分割方法的分割精度, 然而扩展卷积大大增加了计算复杂度和内存占用, 限制了其在实时问题上的应用.

min^h()2H //

# How

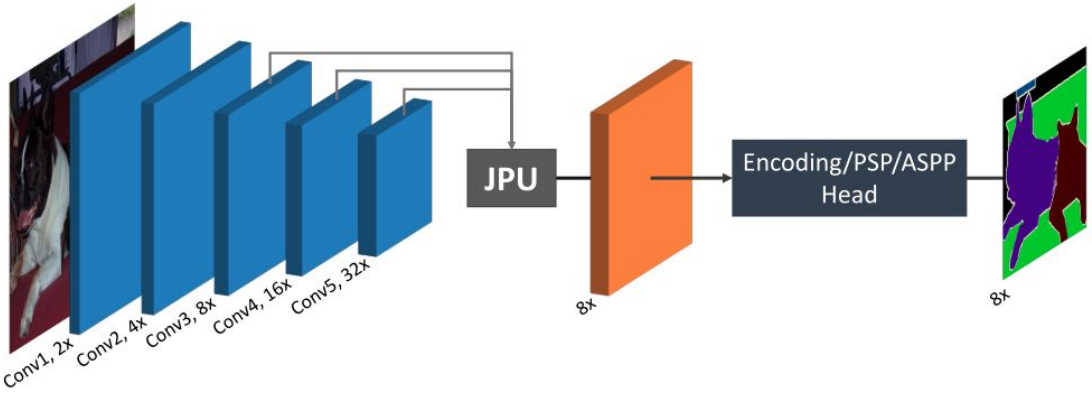
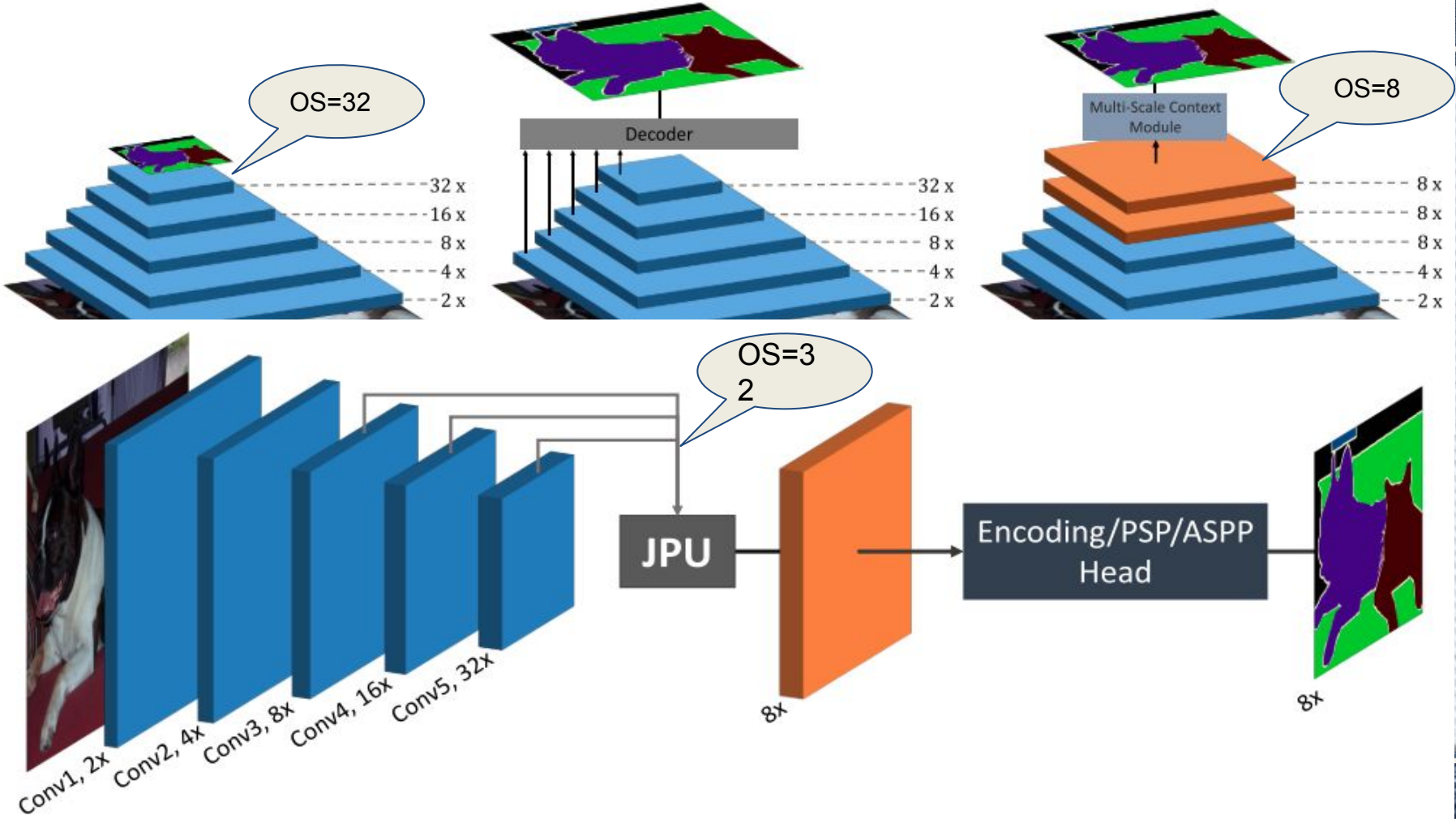
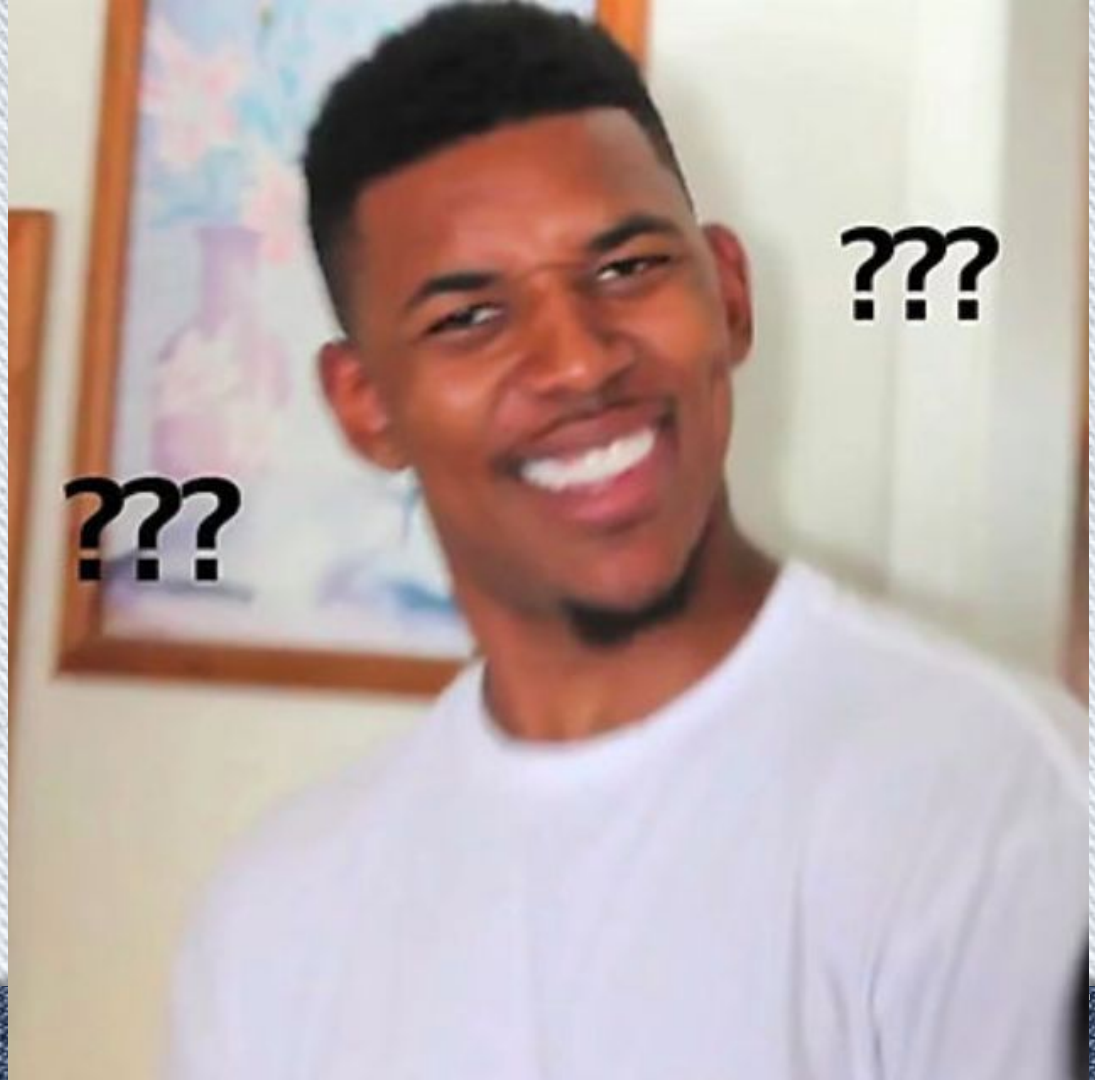



Figure 2: **Framework Overview of Our Method.** Our method employs the same backbone as the original FCN. After the backbone, a novel upsampling module named Joint Pyramid Upsampling (JPU) is proposed, which takes the last three feature maps as the inputs and generates a high-resolution feature map. A multi-scale/global context module is then employed to produce the final label map. Best viewed in color.


OS=32
32 x
16 x
8 x
4 x
2 x
Decoder
32 x
16 x
8 x
4 x
2 x
Multi-Scale Context Module
OS=8
8 x
8 x
8 x
4 x
2 x
OS=32
Conv1, 2x
Conv2, 4x
Conv3, 8x
Conv4, 16x
Conv5, 32x
JPU
8x
Encoding/PSP/ASPP Head
8x



# What is Joint Upsampling

**联合上采样 Joint Upsampling** 给定一个低分辨率的目标图像和一个高分辨率的指导图像，联合上采样旨在从指导图像中转换出结构和细节来生成高分辨率的目标图像。一般而言，低分辨率的目标图像 $y_l$ 由低分辨率的指导图像 $x_l$ 通过变换 $f(\cdot)$ 生成，也就是说 $y_l = f(x_l)$。给定 $x_l$ 和 $y_l$，我们需要获得变换 $\hat{f}(\cdot)$ 来趋近 $f(\cdot)$，其中 $\hat{f}(\cdot)$ 的计算复杂度远低于 $f(\cdot)$。举例而言，如果 $f(\cdot)$ 是多层感知机 MLP，那么 $\hat{f}(\cdot)$ 可以被简化为线性变换。高分辨率的目标图像 $y_h$ 就通过 $\hat{f}(\cdot)$ 作用在高分辨率的指导图像 $x_h$ 来获得。严格来说，给定 $x_l$, $y_l$ 和 $x_h$，联合上采样定义如式 7:

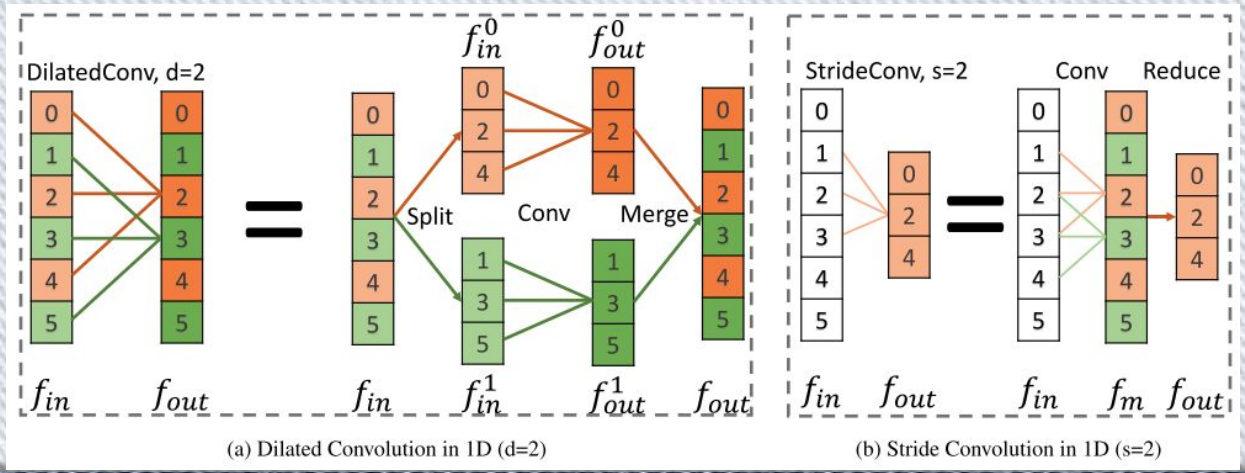
$$y_h = \hat{f}(x_h), \text{where} \hat{f}(\cdot) = \underset{\hat{h}(\cdot)\in\mathcal{H}}{\operatorname{argmax}} \|y_l - h(x_l)\| \tag{7}$$

其中 H 是所有可能的变换函数的集合，$\|\cdot\|$ 是预定义的距离度量。

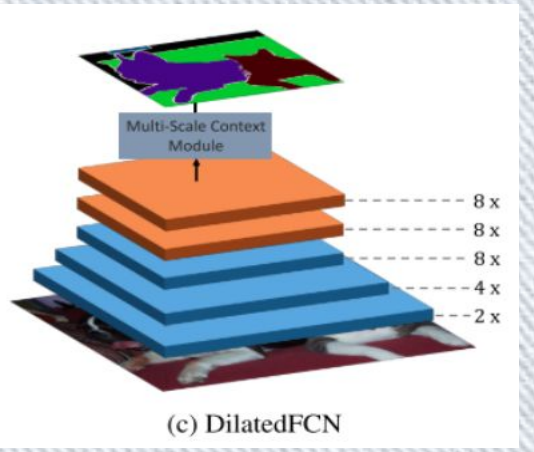
# The Essence of Convolution

如图(a) 所示,DilatedConv = Split(S) + Conv(Cr) + Merge(M). 其中 Split 就是将输入按照序号奇偶不同分为两列,Conv 就是一般的卷积,Merge 就是交错融合.

如图(b) 所示,StrideConv = Conv(Cr) + Reduce(R). 首先对输入fin 进行一次卷积得到中间特征fm, 再删除基数位置数值得到最后的结果.

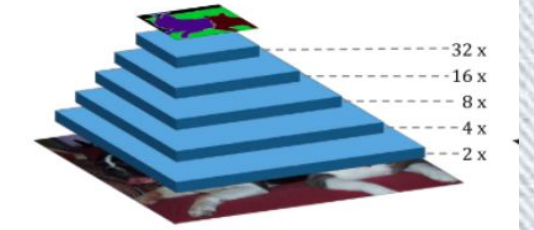


# Analysis of Output



(c) DilatedFCN

基于Dilated Convolution的Backbone



(a) FCN [22]

基于Stride Convolution的Backbone

$$
\begin{aligned}
y_d &= x \to C_r \to \underbrace{C_d \to \cdots\cdots \to C_d}_{n} \\
&= x \to C_r \\
&\to \underbrace{SC_rM \to \cdots\cdots \to SC_rM}_{n} \\
&= x \to C_r \to S \to \underbrace{C_r \to \cdots\cdots \to C_r}_{n} \\
&\qquad \to M \\
&= y_m \to S \to C_r^n \to M \\
&= \left\{y_m^0, y_m^1\right\} \to C_r^n \to M
\end{aligned}
$$

$$
\begin{aligned}
y_s &= x \to C_s \to \underbrace{C_r \to \cdots\cdots \to C_r}_{n} \\
&= x \to C_r \to R \to \underbrace{C_r \to \cdots\cdots \to C_r}_{n} \\
&= y_m \to R \to C_r^n \\
&= y_m^0 \to C_r^n
\end{aligned}
$$

其中的,Cr, Cd 和Cs 分别表示的是常规卷积、扩展卷积与步幅卷积,S,M 和R 表示的分离、聚合和减去操作, 其中的Cnr 表示n 层常规卷积操作. 通过观察上面的两个公式可以发现不论是步幅卷积还是扩展卷积都可以通过常规卷积得到.

之所以比较二者的不同是因为二者展示的是一个优化的问题, 出x,ys,利用二者之间的联系可以得到yd.

然后作者用CNN 模型来近似这个优化过程, 也就出现了JPU(Joined Pyramid Upsampling).

$$
\begin{aligned}
&y = \left\{y_m^0, y_m^1\right\} \to \hat{h} \to M \\
&\text{where } \hat{h} = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \left|\left| y_s - h(y_m^0) \right|\right|, \\
&y_m = x \to C_r
\end{aligned}
$$

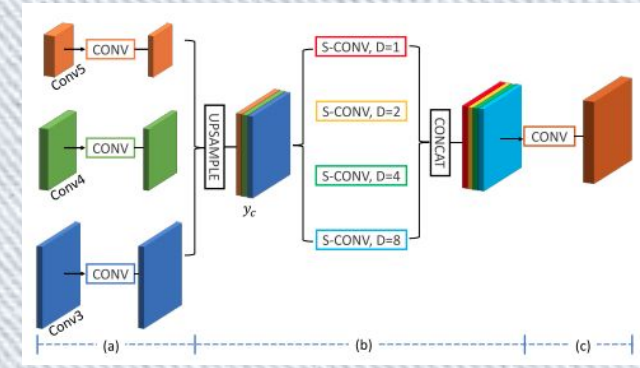
# Joined Pyramid Upsampling (JPU) Module

如上述公式阐明, 联合上采样过程在梯度下降过程中的收敛会花费很多时间, 于是本文通过设计CNN 模块来解决这个问题. 具体的, 首先在给定x 的条件下生成ym, 然后y0m和y1m的特征聚合在一起学习一个映射$\hat{h}$, 最后通过卷积模块得到最终的预测y,由此设计的JPU 模块如图所示.

首先每一个输入特征图通过一个普通卷积模块(阶段a): 通过x 生成ym, 将fm 转换为具有缩减尺寸的嵌入空间, 有助于融合和降低计算复杂度.

接着所有生成的特征图进行上采样和融合生成yc(阶段b):yc 通过多尺度语义模块, 其中空洞率为1 的平行连接用于捕获y0m和ym 其他部分的关系(y0m和y1m之间的关系), 其他空洞率的平行连接用于学习将y0m转化为ys 的映射$\hat{h}$, 映射关系如下图. 所以JPU 可以在多层特征图上提取多尺度文本信息, 与ASPP(在单层上提取多尺度信息) 不同.

阶段对提取到的特征进行一次Conv, 最后得到分割图(阶段c).



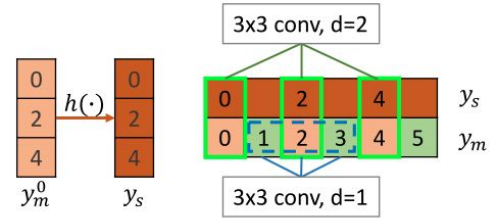




Figure 5: The convolution with dilation rate 1 focuses on $y_m^0$ and the rest part of $y_m$, and the convolution with dilation rate 2 aims at $y_m^0$ and $y_s$. Best viewed in color.

# Back 2

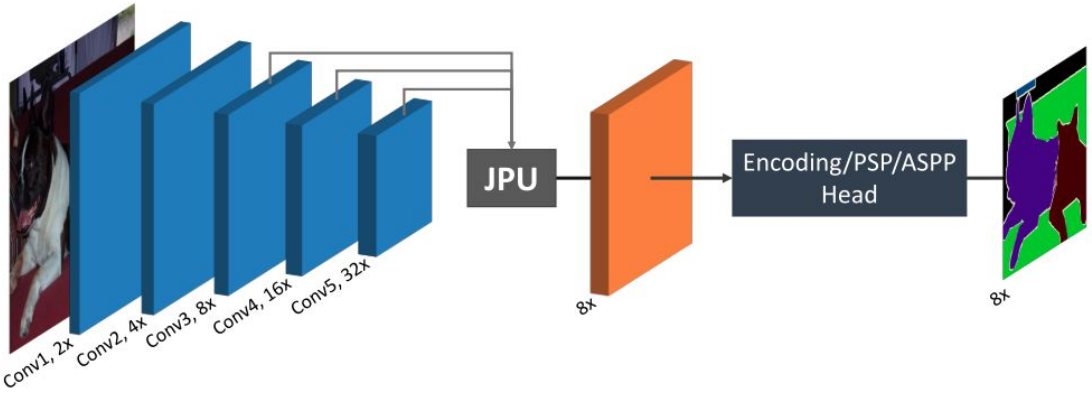



Figure 2: **Framework Overview of Our Method.** Our method employs the same backbone as the original FCN. After the backbone, a novel upsampling module named Joint Pyramid Upsampling (JPU) is proposed, which takes the last three feature maps as the inputs and generates a high-resolution feature map. A multi-scale/global context module is then employed to produce the final label map. Best viewed in color.

# 脑洞

莱特兄弟-鸟-飞机

郎之万-海豚-声纳

斯帕拉捷-蝙蝠-雷达

仿生学-控制论

利奥那多·达·芬奇-鸟-飞行器