论文解析：《K-Net:面向统一图像分割》

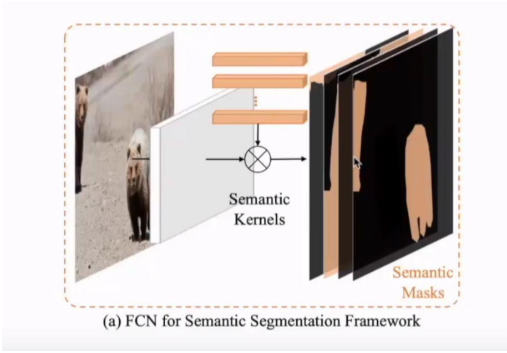
概括

这个名为 K-Net 的框架，通过一组可学习的内核一致地分割实例和语义类别，其中每个内核负生成 mask。

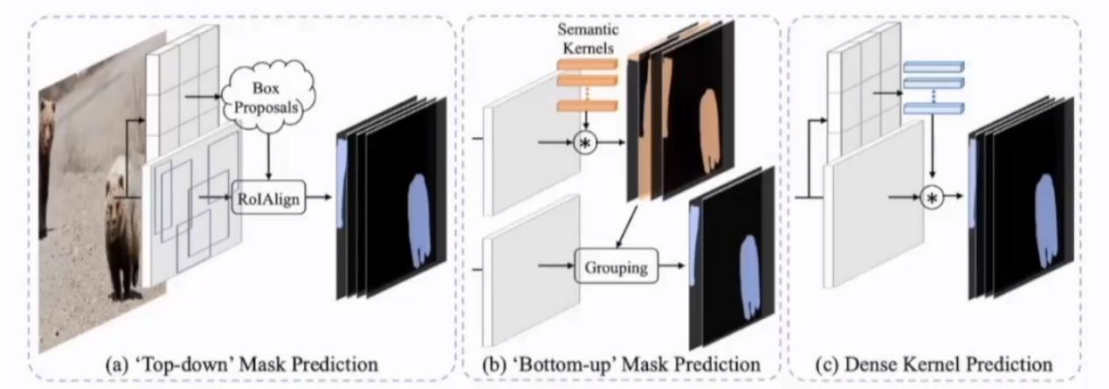
先行知识

图像分割目标：将图像中不同的具有相似性、一致性的像素聚集在一起
语义分割：将每一个像素映射到一个语义类别（每个 pixel 代表一个语义类别）
实例分割：将每一个像素映射到一个 instance ID （同一个 object 上）
全境分割：将每一个像素映射到一个 instance ID 或语义类别{存在 stuff（不可数区域）}

语义分割框架：CNN 生成图像表征--卷积核与特征进行卷积得到 mask，即先检测后分割的框架



实例分割与全境分割框架较复杂

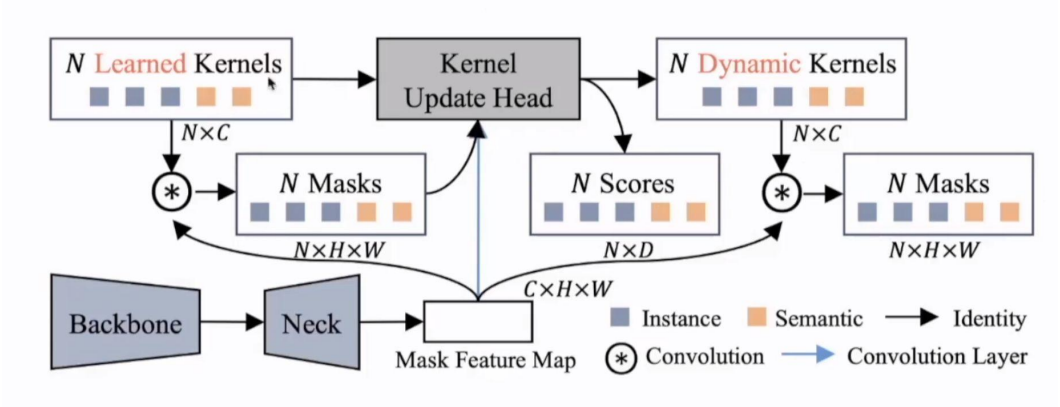


语义分割：每一个语义核（semantic kernel）生成一个 mask，对应一个语义类别
--→试图寻找‘instance kernels’以解决实例分割
--→将 semantic kernels 与 instance kernel 结合以解决全境分割
寻找 instance kernels 的难点：semantic kernels 特点易得，instance kernels 需要动态分配，试图构造‘dynamic kernels’

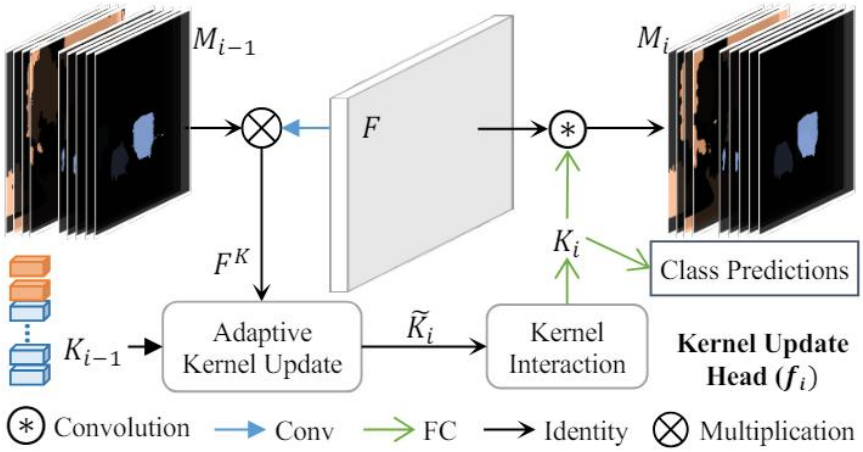
K-Net

初始框架

Backbone 和 neck 提取 mask 特征图（Mask Feature Map）--→可学习核（learned kernels）与特征图卷积得到新的 mask（N masks）--→将 learned kernels、 N masks、Mask Feature Map 作为输入，产生 dynamic kernels，同时对 N masks 分类得到 N scores--→判别特性不断加强



Kernel Update Head 的结构



聚合每组的特征：

Kernel 和特征图响应产生 mask 的部分（mask 是对 pixel 是否属于自己 group 的分配/预测）

$$F^K = \sum_u^H \sum_v^W M_{i-1}(u,v) \cdot F(u,v), F^K \in R^{B \times N \times C},$$

自适应内核更新：

噪声。为了减少组特征中噪声的不利影响，我们设计了一种自适应内核更新策略。具体来说，我们首先在F^K和K_{i-1}as之间进行元素明智的乘法

$$F^G = \phi_1(F^K) \otimes \phi_2(K_{i-1}), F^G \in R^{B\times N\times C}, \tag{4}$$

其中ϕ_1和ϕ_2是线性变换。然后头部学习两个门，G^F和G^K，分别将F^K和K_{i-1}的贡献适应于更新后的内核K，~。其公式为

$$\begin{aligned} G^K &= \sigma(\psi_1(F^G)), G^F = \sigma(\psi_2(F^G)), \\ \tilde{K} &= G^F \otimes \psi_3(F^K) + G^K \otimes \psi_4(K_{i-1}), \end{aligned} \tag{5}$$

式中ψ_n，n = 1，…，4为不同的全连通层(FC)，依次为LayerNorm (LN)，σ为Sigmoid函数。然后在核交互中使用K ~。

动态调整对于新 kernel 的贡献程度*(这一部分没有看明白)*

新的 K-Net

生成 Kernel Update Head--→得到新的 dynamic kernels 、N masks --→堆叠 Kernel Update Head --→不断改良 dynamic kernels 和 N masks

如图所示，其中 dynamic kernels 包含 semantic kernels 和 instance kernels