

GSN

1 方法 (Methodology)(Proposed Approach)

我们提出了一种图析出网络 (Graph Separate Network), 用于学习两幅图片之间包含特征点坐标、描述子和邻域特征的联系 (correspondences)。采用可学习的配对点析出网络相比于固定的特征匹配策略能够发现匹配点之间的一些隐藏信息, 传统的相似度评价指标 (如余弦相似度、特征点距离等) 只能挖掘特征向量之间的显式关系, 一些相似度不高的真实匹配会因此被忽略。但在匹配种子点预选阶段仍然采用这些不可学习策略来过滤错误匹配, 去除明显错误的匹配点, 也减小了网络复杂度, 并且很大部分匹配点之间的联系是很微弱的, 加入学习过程是毫无意义的。

1.1 注意力策略

图像拼接过程中除了聚合关键点位置信息和视觉描述子外, 邻域特征也可用来增强匹配点的背景特征。图注意力网络 (Attention Graph Network) 对于这类非规则结构数据有很好的处理能力, 通过聚合局部或全局信息来捕捉节点之间的联系, 学习视觉描述子之间的拓扑关系。介绍一种典型的带权重的注意力聚合方式。[?]

对于两个特征空间 $\mathbf{X} \in \mathbb{R}^{m \times d}$, $\mathbf{Y} \in \mathbb{R}^{n \times d}$, m 和 n 为特征向量个数, d 表示特征维度, 带权重的注意力机制可以表示为,

$$\mathbf{X}' = \text{Att}(\mathbf{X}, \mathbf{Y}, \omega) = \mathbf{X} + \text{MLP}(\mathbf{X} \parallel \mathbf{m}) \quad (1.1)$$

这里, $\mathbf{m} = \text{softmax}(\frac{\mathbf{QK}^T}{\sqrt{d_k}})\omega\mathbf{V}$ [?], 如果是图内注意力 $\mathbf{Q} = \mathbf{X}, \mathbf{K} = \mathbf{X}, \mathbf{V} = \mathbf{X}$, 如果是图间注意力 $\mathbf{Q} = \mathbf{X}, \mathbf{K} = \mathbf{Y}, \mathbf{V} = \mathbf{Y}$ 。

权重向量 ω 用于调整被聚合元素的重要程度。在本文采用一种可学习的评价网络层来学习注意力权重 ω , 并采用析出策略来拒绝明显错误的匹配点。引入上下文标准化 (Context Normalization) [?] 对每个图像特征进行 normalization 和不同匹配对应关系的 normalization, 这使得特征分布能够编码场景几何结构和相机运动信息, 嵌入上下文信息到训练层的每个匹配关系输入。(画出一个图来说明)

$$\omega_s = \text{Cov}(\text{CN}(\mathbf{X} \parallel \mathbf{Y})) \quad (1.2)$$

其中 \parallel 表示将两幅图片的匹配点节点按照特征维度进行连接, CN 表示上下文正则化, Conv 表示卷积操作。

在注意力权重分配策略中, 常常是对于有效信息进行权重奖励, 忽略或惩罚无效信息。这里采用析出策略舍弃异常离群值, 将粗匹配种子点在学习过程中发现的异常匹配点进行舍弃。(!!!)

$$\mathbf{X}' = \text{Att}(\mathbf{X}, \mathbf{Y}, \omega_s) \quad (1.3)$$

这里 $\mathbf{X}' \in \mathbb{R}^{s \times d}$, $s < \{m, n\}$ (!!!)

1.2 邻域特征聚合

图像特征点