**RM: 精**

|  |  |
| --- | --- |
| 记录项 | 记录内容 |
| 文献名称 | FFB6D: A Full Flow Bidirectional Fusion Network for 6D Pose Estimation |
| 文献作者 | Yisheng He， Haibin Huang， Haoqiang Fan， Qifeng Chen  Jian Sun |
| 来源信息 | Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 3003-3013. (CVPR2021 Oral) |
| 关键词 | Feature Fusion, Farthest Point Sampling |
| 针对性问题 | 对于以前的方法，分别从RGB和点云中提取特征，将他们融合起来做位姿估计，但是由于外观特征和几何特征是分别提取的，两个网络无法共享信息，所以使精度受到限制。 |
| 主要方法 | **DenseFusion**是分别使用CNN和点云特征提取网络提取特征，再将他们融合进行位姿估计，由于外观特征和几何特征是分开提取的，所以限制了特征的表达能力。这篇文章基于此，在两个特征提取网络之间添加通信桥梁（双向融合模块），使得两个分支的信息能够互补。 |
| 主要贡献 | 1. 一种全流双向融合网络，从单个场景的RGBD中学习，可以推广到3D目标检测。 2. 一种简单但是有效的3D关键点选择算法，利用了纹理和几何信息。 3. 在YCB-Video、LineMod和Occulusion上展示了最好的效果。 |
| 1. The paper proposes a Full Flow Bidirectional fusion network (FFB6D) for 6D pose estimation from a single RGBD image. 2. The FFB6D network combines appearance and geometry information for representation learning and output representation selection. 3. Bidirectional fusion modules are built in the full flow of the two networks for representation learning, where fusion is applied to each encoding and decoding layer. 4. A simple but effective 3D keypoints selection algorithm is designed for output representation selection, which considers the texture and geometry information of objects. 5. Experimental results show that the proposed method outperforms the state-of-the-art by large margins on several benchmarks. 6. The code and video of the proposed method are available at https://github.com/ethnhe/FFB6D.git. |
| 核心算法记录（精读） | 1. Full Flow Bidirectional Fusion Network 全流双向融合网络        1. **主要是两个模块**，分别是像素到点（pixel-to-point）和点到像素两个模块。像素到点融合模块指的是将RGB特征融合到点云特征中。 2. 像素到点融合模块的具体方法：对于每个point，在XYZ Map（使用相机的内参将每个像素的深度提升到对应的3D点，获得与RGB对齐的XYZ Map）中找到他的Kr2p最邻近点，并从RGB特征图中收集外观特征，通过最大池化和共享MLP处理这些特征以获得最重要的外观特征。最后通过一个共享的MLP融合外观特征和几个特征从而获得融合的点云特征（the fused point feature）。点到像素的融合与之类似，最终获得融合的像素特征。 3. **密集的RGBD特征**（Dense RGBD feature embedding）：通过相机内参矩阵将点投影到平面获得他们之间的对应关系，根据关系。我们得到成对的外观和几何特征并将他们连接到一起获得密集的RGBD特征，输入语义分割模块和3D关键点检测模块。 4. **基于3D关键点的6D位姿估计（3D Keypoint-based 6D Pose Estimation）：**改进PVN3D的3D关键点选择算法，从而充分利用纹理和几何信息，具体是检测场景中每个对象选择的3D关键点，然后利用最小二乘算法来恢复物体位姿。   **每个物体的关键点检测（Per-object 3D keypoint detection）：**在PVN3D的基础上添加语义分割模块，用以区分不同对象的实例、并使用关键点投票来恢复3D关键点，最终得到每个对象的3D关键点。  **关键点选择（Keypoint selection）：**以前的采样常试用最远点采样（Farthest Point Sampling ，FPS），但是该算法只考虑了欧几里得距离，因此得到的点可能在非显著区域，例如没有纹理的平面，这些点很难检测到，最终会影响位姿估计。所以本文提出了一种能够充分利用物体纹理信息和几何信息的算法，SIFT-FPS（改进SIFT，使用SIFT检测纹理图像中与众不同的2D点，将他们提升到3D，然后应用FPS选择前N个关键点），该算法选择的关键点均匀分布在物体表面，并且纹理鲜明，易于检测。 |
| 主要实验平台 |  |
| 实验数据集 | 制作的人工标注的数据集，包含分割、6D位姿  地址：labelfusion.csail.mit.edu |
| 性能指标 | IOU |
| 性能比较对象 |  |
| 主要实验及目的（简记） | **概览：**  方法概览：更具体地说，给定 RGBD 图像作为输入，我们利用 CNN 从 RGB 图像中提取外观特征，并利用点云网络从点云中提取几何特征。在两个网络的特征提取流程中，点到像素和像素到点融合模块被添加到每一层中作为通信桥梁。这样，两个分支可以利用对方额外的外观（几何）信息来促进自己的表示学习。然后将提取的逐点特征输入实例语义分割和 3D 关键点检测模块，以获得场景中每个对象的 3D 关键点。最后，应用最小二乘拟合算法来恢复 6D 姿态参数。  1. 全流双向融合网络  给定对齐的 RGBD 图像，我们首先使用相机内在矩阵将深度图像提升到点云。然后应用 CNN 和点云网络 (PCN) 分别从 RGB 图像和点云中提取特征。当信息流经两个网络时，添加点对像素和像素对点融合模块进行双向通信。通过这种方式，每个分支都可以利用来自另一个分支的本地和全局信息来促进其表示学习。  **从图像特征到点云特征的像素到点融合。**这些模块将从 CNN 提取的外观信息共享给 PCN。一种简单的方法是从 RGB 特征图生成全局特征，然后将其连接到每个点云特征。然而，由于大多数像素都是背景，并且场景中存在多个对象，因此全局压缩 RGB 特征图会丢失大量详细信息，并损害后续的姿态估计模块。相反，我们引入了一种新颖的像素到点特征融合模块。由于给定的 RGBD 图像对齐良好，我们可以使用 3D 点云作为桥梁来连接逐像素和逐点特征。更具体地说，我们使用相机内在矩阵将每个像素的深度提升到其对应的 3D 点，并获得与 RGB 图对齐的 XYZ 图。如图2（b）左侧所示，对于每个点特征及其3D点坐标，我们在XYZ图中找到其Kr2p最近点，并从RGB特征图中收集其相应的外观特征。然后，我们按照[51]使用最大轮询来集成这些相邻的外观特征，并应用共享的多层感知器（MLP）将其压缩到与点云特征相同的通道大小：    其中 Fri 是 RGB 特征的第 i 个最接近的像素，Fr2p 是集成的像素。然后，我们将集成的外观特征 Fr2p 与点特征 Fpoint 连接起来，并使用共享 MLP 来获得融合的点特征：    其中 ⊕ 是连接运算。  值得一提的是，在外观特征编码的流程中，当网络越深时，RGB特征图的高度和宽度就会变小。因此，我们需要维护一个对应的XYZ图，以便特征的每个像素都能找到其3D坐标。由于减小的特征图尺寸是通过卷积核以步长扫描原始特征图而生成的，因此内核的中心成为特征图的新坐标。一种简单的方法是应用相同大小的内核来计算其中 XYZ 的平均值，以生成像素的新 XYZ 坐标。然后通过使用与 CNN 相同步长的平均核扫描 XYZ 图来获得相应的 XYZ 图。然而，由于前景物体和背景之间的边界上的深度变化显着，均值运算会产生噪声点。相反，更好的解决方案是将 XYZ 图的大小调整为与最近插值算法中的特征图相同的大小。  **从点云特征到图像特征的点到像素融合。**这些模块搭建了将从 PCN 获得的几何信息传输到 CNN 的桥梁。该过程如图2(b)右侧所示。与像素到点融合模块相同，我们密集地融合特征，而不是简单地将全局点特征连接到每个像素。具体来说，对于具有 XYZ 坐标的特征的每个像素，我们从点云中找到其 Kp2r 最近点并收集相应的点特征。我们将点特征压缩到与 RGB 特征相同的通道大小，然后使用最大池化来整合它们。然后将集成的点特征与相应的颜色特征连接起来，并通过共享 MLP 进行映射以生成融合的特征：    其中Fpj表示第j个最近的点特征，Fp2r表示积分点特征，⊕表示连接操作。  密集RGBD特征嵌入通过所提出的全流融合网络，我们从CNN分支获得密集外观嵌入，从PCN分支获得密集几何特征。然后，我们通过使用相机固有矩阵将每个点投影到图像平面来找到它们之间的对应关系。根据对应关系，我们获得成对的外观和几何特征，并将它们连接在一起以形成提取的密集 RGBD 特征。然后将这些特征输入到实例语义分割模块和 3D 关键点检测模块中，以在下一步中进行对象姿态估计。  2.基于 3D 关键点的 6D 姿态估计  最近，He 等人的 PVN3D [17] 工作。为使用 3D 关键点估计物体姿态开辟了新的机会。在这项工作中，我们遵循他们的 3D 关键点公式，但进一步改进了 3D 关键点选择算法，以充分利用对象的纹理和几何信息。具体来说，我们首先检测场景中每个对象选择的 3D 关键点，然后利用最小二乘拟合算法来恢复姿势参数。  每个对象 3D 关键点检测 到目前为止，我们已经获得了密集的 RGBD 嵌入。然后，我们遵循 PVN3D [17]，通过添加实例语义分割模块来区分不同的对象实例和关键点投票模块来恢复 3D 关键点，获得每个对象的 3D 关键点。实例语义分割模块由语义分割模块和中心点投票模块组成，前者预测每点语义标签，后者学习每点到对象中心的偏移量以区分不同实例。对于每个对象实例，关键点投票模块学习到所选关键点的逐点偏移，这些关键点以 MeanShift [10] 聚类方式投票支持 3D 关键点。  关键点选择 先前的工作[47, 17]使用最远点采样（FPS）算法从目标对象表面选择关键点。具体来说，他们维护一个由对象表面上的随机点初始化的关键点集，并迭代地添加距离该集中最远的其他点，直到获得 N 个点。通过这种方式，选定的关键点分布在物体表面并稳定后续的姿态估计过程[47, 17]。然而，由于该算法仅考虑欧氏距离，因此所选点可能出现在非显着区域，例如没有明显纹理的平面。这些点很难检测到，估计姿态的准确性也会降低。为了充分利用对象的纹理和几何信息，我们提出了一种简单但有效的 3D 关键点选择算法，称为 SIFT-FPS。具体来说，我们使用 SIFT [40] 算法来检测纹理图像中独特的 2D 关键点，然后将它们提升到 3D。然后应用FPS算法来选择其中的前N个关键点。这样，所选的关键点不仅在物体表面分布均匀，而且纹理特征鲜明，易于检测。  最小二乘拟合 给定对象坐标系{pi}N i=1 中选定的3D 关键点，以及相机坐标系{p\* i }N i=1 中相应的3D 关键点。最小二乘拟合[1]算法通过最小化平方损失来计算位姿参数R和T： |
| 对文献的个人思考 | 1. 融合模块有提到如何提取重要特征，本文使用的是最大池化（后面写的是最大轮询）和MLP的方式，我觉得这会是一个改进的点，也许把注意力机制弄进去会更好？ 2. 这篇论文的作者和PVN3D的是同一个，PVN3D使用DenseFusion和hough投票来搞网络。所以后续可以暂时跳过以上两篇论文。 |
| 扩展阅读 | 1. PVN3D 2. Kr2p算法 3. 最远点采样FPS |