**Hands Deep in Deep Learning for Hand Pose Estimation**

**本文贡献:**

1. 实验结果表明，该方法能够学习到手部姿态的先验模型，并将其无缝地集成到网络中，提高了预测精度。这会导致网络出现一个不寻常的“瓶颈”，即一层的神经元比最后一层少。
2. 与之前的工作一样，我们使用细化阶段来独立地改进每个关节的位置估计。由于这是一个回归问题，因此在这一阶段应谨慎使用空间池和子采样。为了解决这个问题，我们使用多个以关节的初始估计为中心的输入区域，对于较小的输入区域使用非常小的集合区域，对于较大的输入区域使用较大的集合区域。较小的区域提供准确性，较大的区域提供上下文信息。

**本文提到的其他方法:**

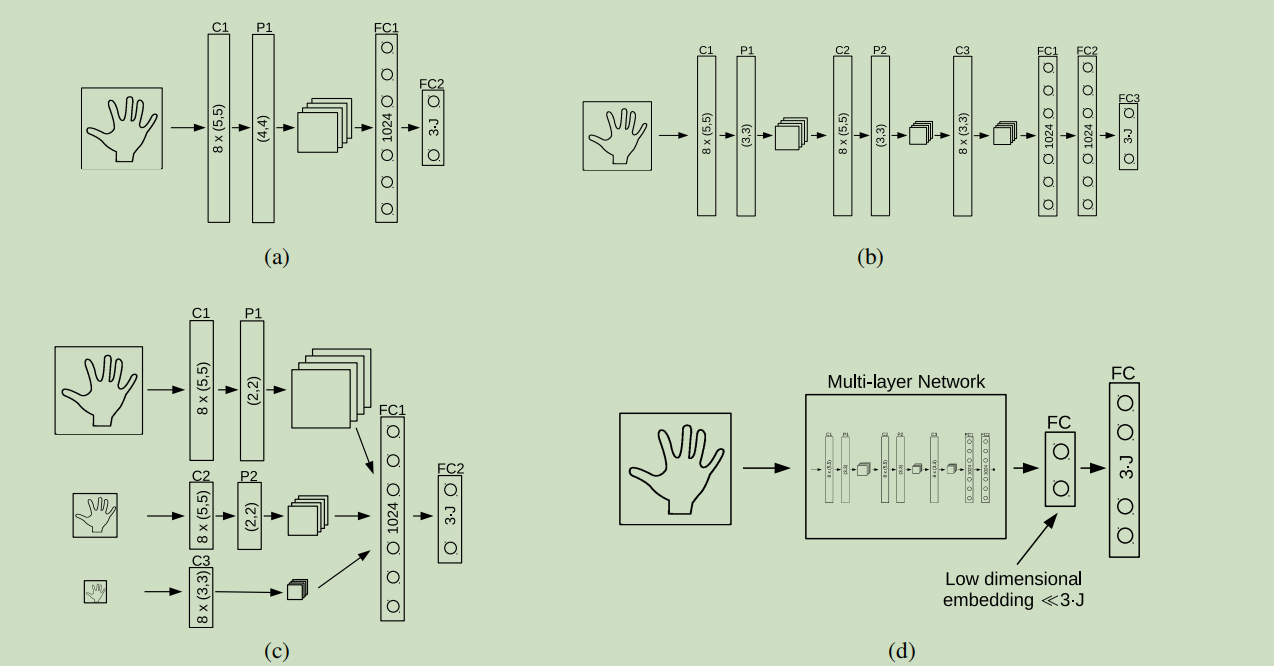
1. 第一种方法是基于生成的、基于模型的跟踪方法.
2. 第二种方法是辨别法，旨在直接从RGB或RGB-D图像预测关节的位置。如随机森林，CNN。

**本文方法描述:**

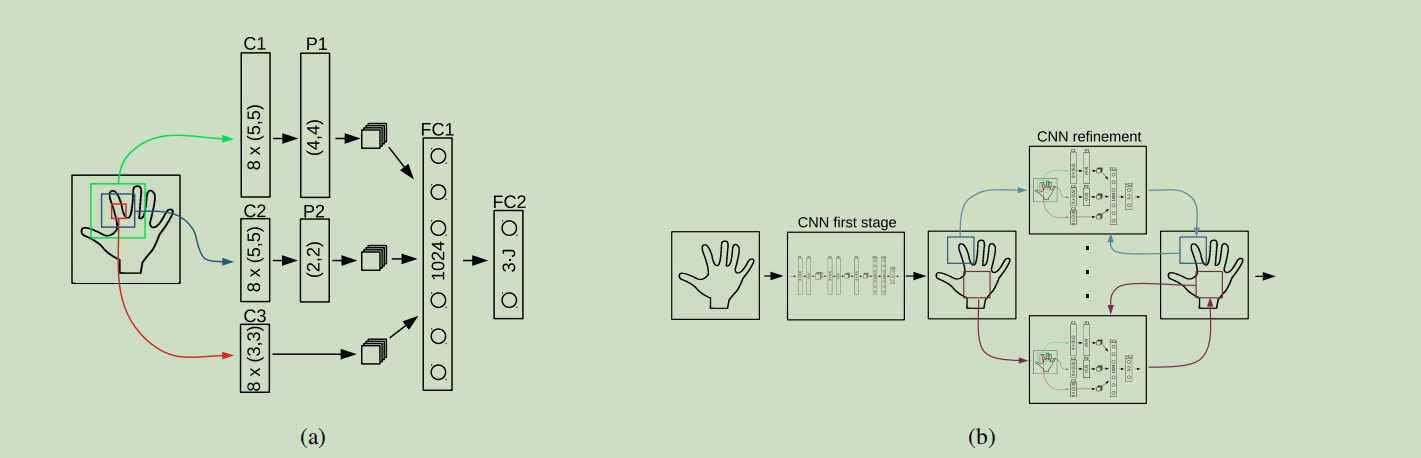
本文中的手部姿态回归框架由两部分组成:**手部检测器**和**姿态预测器**。**手部检测器**为一个简单的2D手检测器，我们用它来获得手的粗略边界框作为基于CNN的姿势预测器的输入。**姿态预测器**由两个阶段组成**:**对于第一阶段，用于同时预测所有关节的位置，该阶段可以预测较低维空间中的姿势；第二阶段，它独立于第一阶段的预测来细化关节的位置。

**手部检测器.**本文首先使用一种简单的方法来估计一个包含手的粗略的3Ｄ边界框，这种方法类似于假设手是离相机最近的对象：我们从深度图像中提取一个以该对象的质心为中心的固定大小的立方体，并将其大小调整为128×128,并将深度值标准化为[-1，1]。对于深度不可用的点（例如，相机未捕获到的点或深度比立方体的背面更深的点）指定其深度为1。为了保持从手到相机的不同距离不变性，这种标准化对CNN很重要。

**姿态预测器。**我们首先考虑了两个标准的CNN架构。第一个如下图(a)所示，是简单的浅层网络，由单个卷积层、最大池层和单个完全连接的隐藏层组成。第二个架构如下图(b)所示，它是一个更深但仍然通用的网络，具有三个卷积层、最大池化层和两个完全连接的隐藏层。所有层都使用校正线性单元激活函数(Rectified Linear Unit )。此外，我们评估了一种多尺度方法。这种方法的动机是使用多个尺度可以帮助捕获上下文信息。它使用输入图像的几个缩小版本作为网络的输入，如下图(C)图所示。我们的结果将表明，多尺度方法比深层架构表现更好，深层架构比浅层架构表现更好。 但是，我们在接下来的两节中描述的贡献带来了更多的改进。到目前为止，我们只考虑直接预测关节的三维位置。然而，考虑到手部的物理约束，不同的三维关节位置之间存在着很强的相关性，并且之前的工作已经表明，低维嵌入足以参数化手部的三维姿态。因此，我们可以在低维空间中预测姿态参数，而不是直接预测三维关节位置。由于这加强了手部姿态的约束，可以预期它提高了预测的可靠性，这将得到我们实验的证实。如图1d所示，我们通过在最后一层引入“瓶颈”来实现在网络结构中的姿势优先。这个瓶颈是一个神经元数量少于全姿态表示所需的3·j的层，它迫使网络学习训练数据的低维表示，从而实现手的物理约束。类似于[28]，我们依赖于线性嵌入。嵌入由瓶颈层实现，嵌入到位姿空间的重构作为单独的隐藏层添加在瓶颈层之上。通过设置重建层的权值来计算三维关节空间的反投影。因此，生成的网络直接计算完整姿势。通过对手姿态数据进行主成分分析，利用主成分对重构权值进行初始化，然后利用反向传播对整个网络进行训练。使用这种方法，我们训练前一节描述的网络。

****

第二阶段:优化阶段.

****