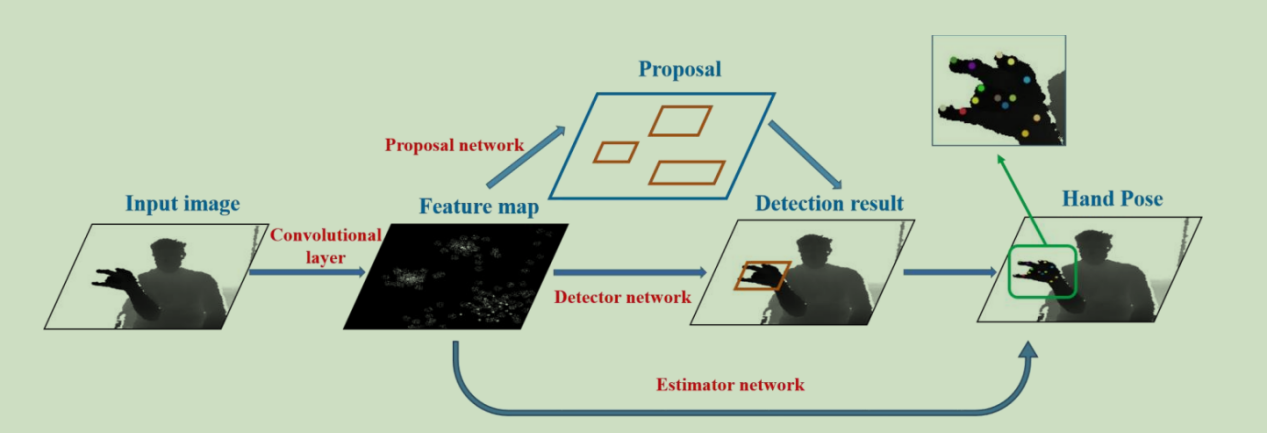
**Deep Learning for Integrated Hand Detection and Pose Estimation**

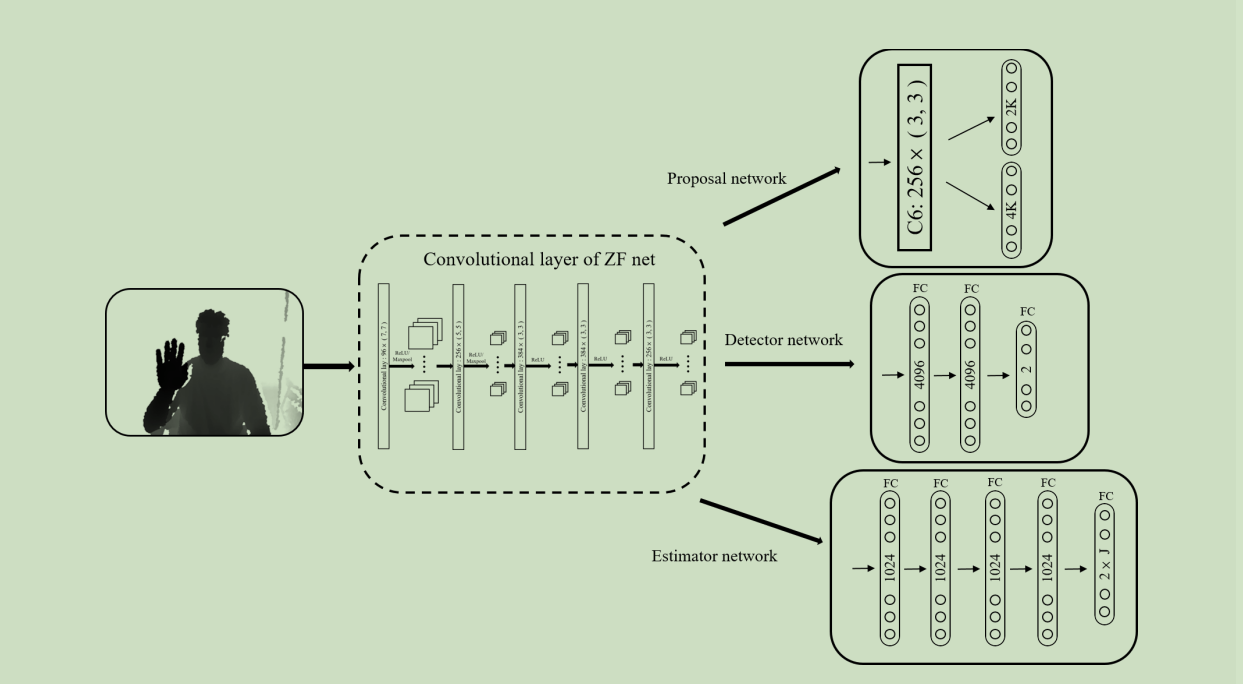


系统框架总览

该系统包含三个**卷积神经网络**(proposal, detector and estimator network)框架和一个**优化阶段**(optimization stage)。

**系统流程：**

首先,对原始深度图像做卷积运算得到一个特征图像（feature map）;然后,proposal network利用特征图像生成一系列手部的候选对象，detector network对生成的候选对象进行运算得到最有可能是手得对象，并在由原始图像得到的特征图像上标记出；接着, estimator network利用框选出的手部对象，回归出每个关节的坐标；最后，对这些回归出的位置进行优化。下图为神经网络框架:



**实验结果：**

准确率: 90%

误差: ±8.92 pixels.

速率: 74毫秒每帧(13.5FPS)

**Depth-based 3D Hand Pose Tracking**

**本文所做的主要贡献**

第一个贡献是一种新的检测跟踪策略，通过考虑使用RNN的预测历史结果，将基于CNN的单帧检测方法扩展到多帧跟踪方法。第二个贡献是使用RNN模拟3D模型对输入数据的拟合。

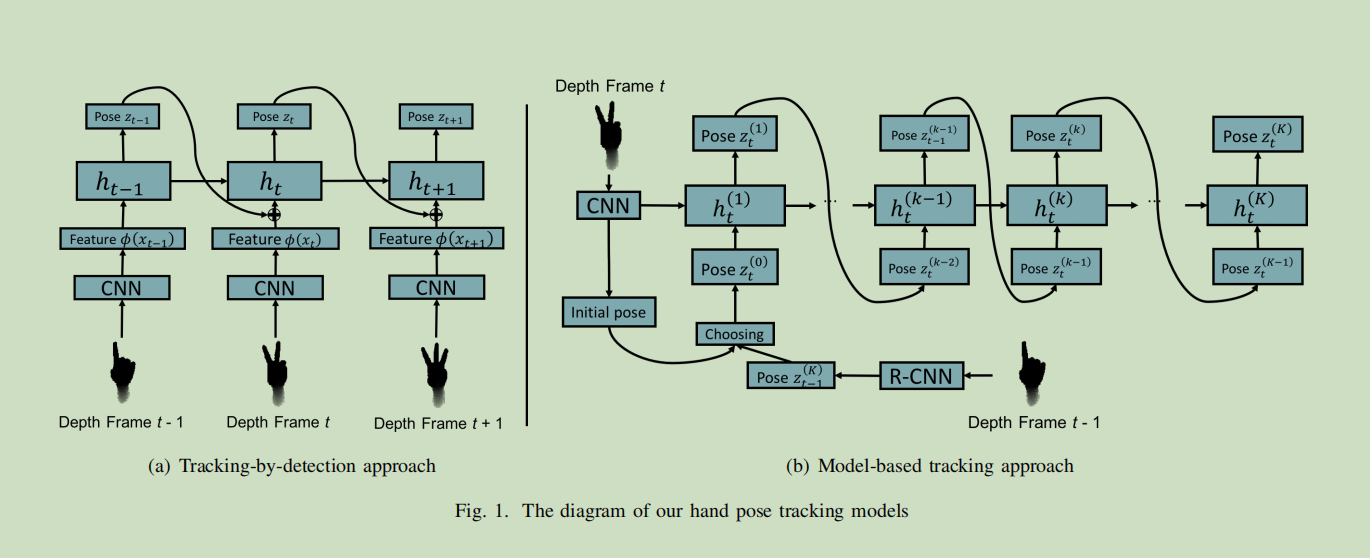
**本文提到的其他方法**

**基于模型生成的手部跟踪**(the model-base generative tracking)和**基于判别式手势检测**(the discriminative hand pose detection)。各自的缺点如下:

基于模型生成的手部跟踪: 当检测器在发生阻塞时导致模型初始化失败，跟踪框架无法从初始错误中恢复。

基于判别式手势检测（自己添加的）：鲁棒性不强，难以处理自遮挡的情况。

而本文则是结合上述的两种方法创建一种更强健的监测策略。



手姿势跟踪模型框架图

**框架说明：**

第一个部分是RNN输出根据CNN生成的图像的视觉特征和RNN的记忆状态（记忆的先前的帧特征）计算出的手姿态预测(如上图(a)所示)。换句话说，RNN起到记忆来自先前帧的手姿势预测的作用。所以，基于CNN的手部姿势估计的结果现在受到先前帧而不是当前输入帧的影响。最后，预测可以选择性地强调输入中的关注区域，然后在下一帧中将其馈送到卷积网络中。上图（ａ）所示，生成每一帧的特征都需要结合前一帧的特征。

第二个部分是一种新的基于生成模型的跟踪，其中RNN中的每个时间步骤可被视为传统的基于模型的跟踪中的有效近似模型拟合步骤。从训练数据中学习相似度函数并将其集成到RNN的学习中。CNN用于从当前深度帧中提取特征，并在必要时产生良好的初始手势（如图1（b）所示）。上图（ｂ）中，深度图像中的当前帧和上一帧同时作为网络的输入，然后网络利用两副图的特征回归出当前帧中的位置。

**实验结果：**

准确率: 89-99%

误差: 20~25mm