

Abstract

automatically learn the image-dependent structural constraint and sequence-dependent temporal context by using a multi-stage sequential refinement

1. a 2D pose module extracting the image-dependent pose representations
2. a 3D pose recurrent module regressing 3D poses
3. a feature adaption module serving as bridge between 1 and 2 to enable the representation transformation from 2D to 3D domain.

1 Introduction

To resolve the challenges share with 2D and the 3D pose rescovery from monocular imagery, a preferable way is to investigate how to simultaneously **enforce 2D spatial relationship, 3D geometry constraint and temporal consistency within one single model**.

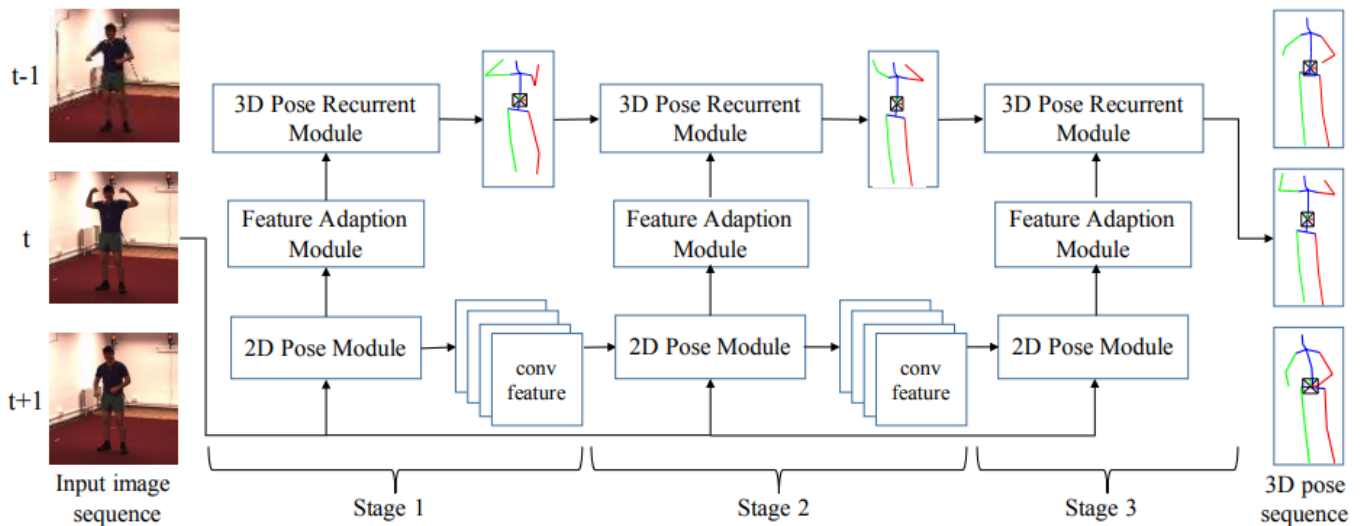
Inspired by the pose machine and CPM, our RPSM proposes a multi-stage training to capture long-range dependencies among multiple body-parts for 3d pose prediction.

Also include several stage, in each stage, RPSM is composed by **a 2D pose module, a feature adaption module and a 3D pose recurrent module**.

Main Contribution:

- learns to recurrently integrate rich spatial and temporal long-range dependencies using a multi-stage sequential refinement, 而不使用人为定义的一些限制
- 使用了recurrent network models

2 Recurrent 3D Pose Sequence Machines



2.1 Multi-stage Optimization

$$\begin{aligned}
 f_{2D}^{t,k} &= \Psi_p(I_t, f_{2D}^{t,k-1}; W_p), \\
 f_{3D}^{t,k} &= \Psi_a(f_{2D}^{t,k}; W_a), \\
 S_t^k &= \Psi_r(f_{3D}^{t,k}, H_{t-1}^k, S_t^{k-1}; W_r),
 \end{aligned}$$

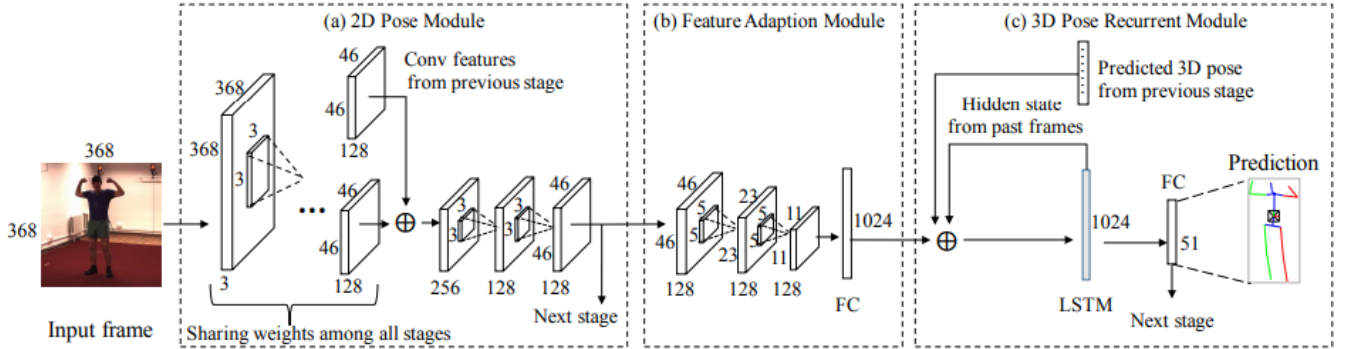
其中 W_p, W_a, W_r 是 Ψ_p, Ψ_a, Ψ_r 的参数, I_t 是 t 步的图片输入, H_{t-1}^k 是 3D pose recurrent module 上一个隐藏层输出, S_t^{k-1} 是上一层预测的 3D pose 结果。

2.2 2D pose module

Divide 2D pose module into two parts:

- the shared convolution layers
- the specialized pose-aware convolutio layers in each stage.

如下图 2D pose module 部分所示:



2.3 Feature Adaption Module

如上图中 (b) 部分所示, feature adaption module consists of two convolutional layers and one fully connected layer.

2.4 3D Pose Recurrent Module

如图 (c)所示, the adapted features, the previous hidden states and the previous 3D pose predictions are concatenated together as the current input of 3D poses Recurrent module.