

摘要

近些年来，深度卷积神经网络研究获得了飞速进展，使得计算机视觉中的诸多领域都取得了巨大进步，如目标检测、图像分割、图像去噪、图像分类等。人体姿态估计也是计算机视觉中一个很重要的领域，是很多上层落地应用感知层面的技术基础。它旨在将一张输入图片中的人体的每一个关键点位置预测出来，即预测图片中人体姿态。以终为始，人们希望所设计的深度神经网络可以像人一样，准确地识别出图像中的人体姿态。近期的研究发现，现有的即使是最先进的神经网络都不具有平移等变性。其具体表现为输入图片产生微小的偏移，模型的预测结果会大相径庭。这不符合我们所期望的网络应有的能力，这一点在现有的人体姿态估计网络中也不例外。另外，这一性质的欠缺在业界应用广泛的自顶向下方法中尤为致命：后接的人体姿态估计网络往往由于前置人体检测器输出的人体边界框的偏移而导致精度下降，视觉效果表现为关键点的抖动和跳变，这是人们所难以接受的。

然而，目前现有的方法里，仅有对图像分类网络和图像重建网络的平移等变性研究。而人体姿态估计领域的平移等变性问题并未有研究者涉足，目前即使是最先进的人体姿态估计网络也不具有平移等变性。

基于上述的问题与挑战，本文希望探索一种能够为人体姿态估计网络引入平移等变性的通用方法。这种方法将适用于各种主流的人体姿态估计网络架构中。同时本文希望所引入的平移等变人体姿态估计网络能够提升识别准确率，应用在实际的工作流程中也能够表现出抗输入边界框抖动的性质。

首先，本文在人体姿态估计任务的关键点编解码方法上进行了研究。本文首先对当前主流关键点位置表示方法和后处理方法进行描述并进行误差分析；并针对所发现的问题提出了基于热力图的关键点无偏表示方法和后处理方法，以确保关键点位置信息在编码、解码、量化三个主步骤上的无偏性，不会引入额外的系统误差。

然后，本文在深度卷积神经网络的组件的平移等变性上也进行了研究。本文提出的平移等变人体姿态估计网络结构针对性地解决了网络填充、采样，以及网络非对称性导致的平移等变性的损失。实验表明，使用本文提出的网络结构，不仅能够获得优于基线方法一个数量级的平移等变性，同时可以带来关键点检测准确性的提升。同时，应用在实际的工作流程中，网络也表现出了更强的抗扰动能力，缓解了由于前置检测器输出边界框偏移导致的精度损失，表现出优秀的学术价值和实际应用价值。

关键词：深度学习，人体姿态估计，平移等变性