Computer Science and Application 计算机科学与应用, 2024, 14(5), 126-136
Published Online May 2024 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2024.145121

Hans汉斯

面向嵌入式平台的红外人体姿态估计系统

侯雨琪*，白　玉

沈阳航空航天大学电子信息工程学院，辽宁 沈阳

收稿日期：2024年4月23日；录用日期：2024年5月22日；发布日期：2024年5月29日

摘　要

针对在低照度环境下的人体姿态估计精度下降严重，且模型参数量大导致部署在嵌入式设备时效率低的问题。本文设计了一种基于红外摄像头的Jetson Xavier NX平台轻量化人体姿态估计系统，提出基于HRNet的人体姿态估计方法。首先，引入残差模块并在可见域和红外域之间进行迁移学习；其次，提出结合通道剪枝模块的HRNet，消除分支中的冗余，以低开销进行规模感知特征融合；最后，利用TensorRT方法优化深度学习模型并部署到Jetson Xavier NX嵌入式平台。实验结果表明，改进后模型更符合对嵌入式设备的实时性需求，参数量相比原模型减少46%，与同样规模相比具有更高的检测精度，模型的mAP保持在74.2%以上，经过TensorRT加速优化后系统检测速度可达33 fps。

关键词

人体姿态估计，HRNet，剪枝，迁移学习

Infrared Human Pose Estimation for Embedded System

Yuqi Hou*, Yu Bai

School of Electronic Information Engineering, Shenyang Aerospace University, Shenyang Liaoning

Received: Apr. 23rd, 2024; accepted: May 22nd, 2024; published: May 29th, 2024

Abstract

In response to the significant degradation in human pose estimation accuracy under low-illumination conditions, and the challenges posed by the large number of model parameters leading to low efficiency when deployed on embedded devices, this paper introduces a lightweight human pose estimation system based on the Jetson Xavier NX platform utilizing infrared cameras. We propose a

*通讯作者。

文章引用: 侯雨琪, 白玉. 面向嵌入式平台的红外人体姿态估计系统[J]. 计算机科学与应用, 2024, 14(5): 126-136.
DOI: 10.12677/csa.2024.145121

novel human pose estimation method, which is founded on HRNet. Initially, we opted to introduce a residual module and perform transfer learning between the visible and infrared domains. Given the absence of large, this paper introduces HRNet combined with a channel pruning module, which eliminates redundancy within the branches, enabling scalable feature fusion with low overhead. Subsequently, we utilize the output keypoint heatmaps for simple action classification. Finally, the deep learning model is optimized using TensorRT methods to enhance inference speed and deploy it on the Jetson Xavier NX embedded platform. Experimental results demonstrate that the improved model has a 46% reduction in parameters compared to the original model, offering higher detection accuracy when compared to models of similar size. The model's mAP remains above 74.2%, and after acceleration optimization, the detection speed reaches 33 fps.

Keywords

Human Pose Estimation, HRNet, Pruning, Transfer Learning

Copyright © 2024 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

由于在人体姿态估计[1]中，往往需要高分辨率的表示以保证估计的准确性，所以目前的人体姿态估计算法绝大部分都是基于可见光的。2014 年 Jain 等人[2]提出 DeepPose，首次将深度学习中的卷积神经网络引入人体关键点检测。2016 年从 CPM [3]开始，CNN 已经可以把关键点特征表示以及关键点的空间位置信息建模进去。同年还出现了一个非常重要的 MSCOCO 数据集。2018 年 Xiao 等人[4]提出 Simple Baselines，使用转置卷积提高特征图的分辨率代替之前常用的线性插值方法，引起了对于如何得到高分辨率预测热图的思考。2019 年 SUN [5]提出了高分辨率网络 HRNet。考虑到在现实情况下，经常会有在低照度[6]条件下进行检测的需求；另外人体姿态估计是一项既需要准确性又需要效率的任务，特别是当它遇到资源有限的设备上的实时应用时，模型的参数量对于模型部署效率有很大影响。

基于以上问题，本文提出一种基于改进 HRNet 的人体姿态估计方法。结合一种无损通道剪枝模块在不显著影响准确率的情况下大幅减少参数的数量。并使用迁移学习方法克服红外图像数据较少的问题，本文用 Microsoft COCO 2017 数据集的灰度图像对网络进行预训练，再在改进的网络前加入一层瓶颈结构进一步提取特征，后用 UCH-Thermal-Pose 数据库在热域对其进行微调，从而训练得到优化后的模型。最后使用 TensorRT 对模型进行 Jetson Xavier NX 平台优化加速，采用上述系统实现基于嵌入式平台的红外人体姿态估计。

2. 人体姿态估计的关键算法

2.1. 基于迁移学习的人体姿态估计

由于目前能够收集到的用于红外图像的人体关键点检测的数据集通常很小，不能满足深度网络的训练要求，数据不足必然会遇到模型过拟合的问题，因此本文采用基于迁移学习的微调方法将提出的人体姿势估计方法从可见域扩展到热域。迁移学习预训练－微调[7] [8]的学习范式通常首先在大型数据集上对模型进行预训练，以获取更广泛的通用知识。本文利用灰度图像和红外图像数据格式都是单通道图像的

特点为红外图像的人体关键点检测任务提供有效的解决方案，具体流程如图 1。

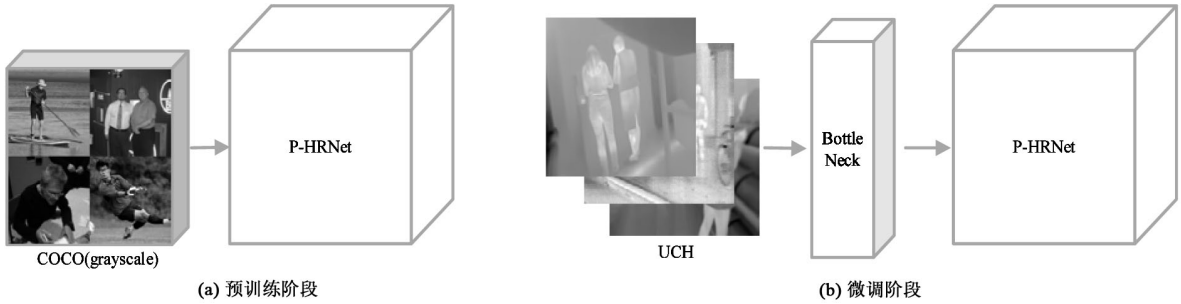


Figure 1. Transfer learning flowchart
图 1. 迁移学习流程图

为了使训练样本更接近红外图像的目标域，本文将 COCO 数据集中的图像进行标准灰度化处理，转换为和红外图像数据格式相同的灰度图像用来预训练 P-HRNet，获得可用的预训练权值作为对该方法进行微调的基础。考虑到红外图像图像对比度低，细节信息不丰富等特点，本文又在 P-HRNet 网络前引入了一个残差模块，先降维聚焦关键信息，再升维恢复分辨率，在不改变特征层大小的情况下对红外图像进行进一步的特征提取。接着采用迁移学习微调的方法，使用 UCH-Thermal-Pose 数据集对加入了瓶颈结构的 P-HRNet 网络整体进行微调，得到最终基于红外图像的人体姿态估计模型。

2.2. P-HRNet 结构

由于红外图像往往分辨率较低，细节信息有限，这要求检测模型可以维持图像分辨率，降低由于低分辨率图像信息表征能力弱带来的检测难度。而且红外图像中的目标轮廓和关键特征可能不如可见光图像清晰，针对这些挑战本文选择了在提取人体姿态估计的多尺度特征方面表现出卓越的能力的 HRNet 作为基础网络。HRNet 的高分辨率处理能力可以最大限度地提取和利用有限的信息，从而在姿态估计中提供更高的准确性。而且考虑到后期的嵌入式部署，本文需要一个高效稳定的人体姿态估计网络，HRNet 的并行子网络结构可以在不增加过多计算负担的情况下实现高性能的姿态估计且具有很好的鲁棒性。

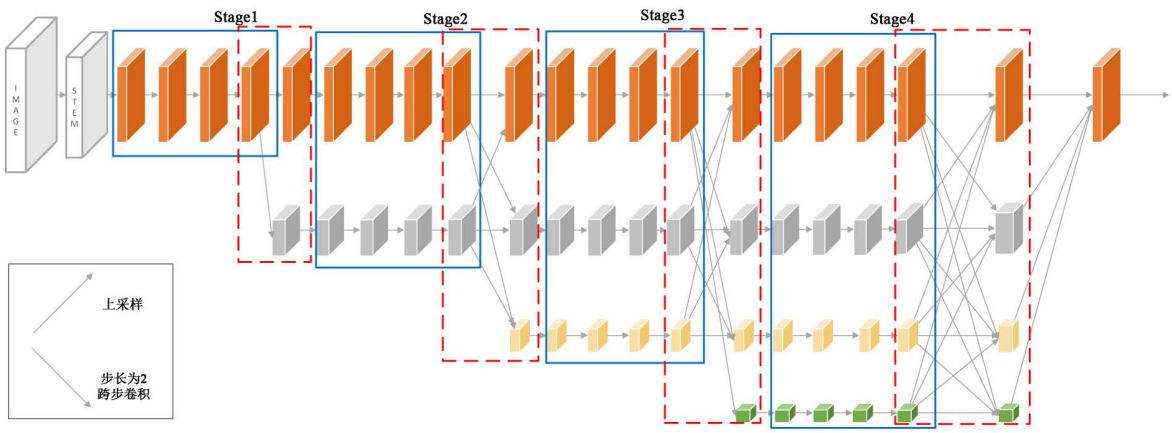


Figure 2. The overall structure of P-HRNet
图 2. P-HRNet 网络整体结构

如图 2 所示，P-HRNet 网络一共生成四级的多尺度特征融合结构。首先将 256 × 192 × 1 的原始图像输入到由两个跨步的 3 × 3 卷积组成的 Stem 结构，将分辨率降低到 1/4。接着图片将通过一系列改进的 Stage 结构，在图中蓝色框和红色框两部分组成一个 Stage。蓝色框部分是每个 Stage 的基本结构，由多个 branch 组成，Stage1 蓝色框使用的是 P-Bottle Neck，其余 Stage 的蓝色框使用 P-Basic Block。红色框部分是每个 Stage 的过渡结构，有 Transition 层和 Fuse 层。在 Stage1 的时候通过重复堆叠 4 次 P-Bottle Neck 提高各层之间梯度的传播能力，使特征图的通道数增多，接着 Transition 将图片通过并行两个卷积核大小为 3 × 3 的卷积层得到两个不同的尺度分支，即下采样 4 倍的尺度以及下采样 8 倍的尺度。后几个 Stage 结构类似，以 Stage3 为例进行说明。对于每个尺度分支都首先通过 4 个 P-Basic Block，然后融合不同尺度上的信息，对于下采样 4 倍分支的输出，它是分别将下采样 4 倍分支的输出、下采样 8 倍分支的输出上采样 2 倍、下采样 16 倍分支的输出上采样 4 倍进行相加最后通过 ReLU 得到下采样 4 倍分支的融合输出。如此经过 4 个 Stage 以后便输出 64 × 48 × 17 的特征图，即 17 个身体部位关键点的 heatmap。

2.3. 无损通道剪枝模块原理

对于人体姿态估计任务来说，随着模型深度的增加，网络的收敛性和模型规模大小也受到影响。因此，如何在不牺牲准确率的同时削减模型的规模是一个亟待解决的问题。ResRep [9]是一种通道剪枝方法。它将 CNN 重参数化为两部分，分别用于保持模型性能[10]和修剪。它通过减少卷积层的宽度来精简 CNN。受其启发，本文提出了一种不会显著影响精度的剪枝方法，将结构重参数化思想的剪枝方式融入人体姿态估计算法中。该模型在保证精度前提下，参数量大幅减少，模型复杂度更低。

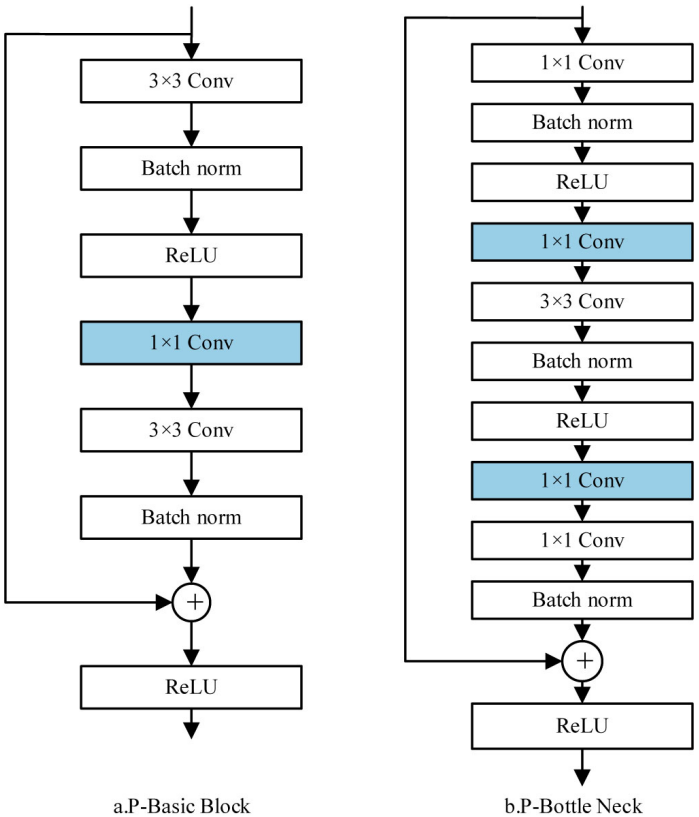


Figure 3. Residual module structure combined with pruning
图 3. 结合剪枝的残差模块结构

如图 3 所示，为了在预测精度和计算成本之间取得平衡，本文提出 P-Bottle Neck 和 P-Basic Block 来替换 Stage 结构中的所有残差块，即对 Stage 结构中的 Bottle Neck 和 Basic Block 进行剪枝。P-Bottle Neck 将修剪 Bottle Neck 的第一层(1 × 1)和第二层(3 × 3)卷积层，在其前两个卷积层经过 BN 以后各加一个 1 × 1 的卷积，这样减少了第一层第二层的输出通道和第二层第三层的输入通道，而对第一层的输入和第三层的输出没有影响，所以不影响残差。P-Basic Block 则是融合通道剪枝的 Basic Block，在其第一个卷积层(3 × 3)经过 BN 以后加上一个 1 × 1 的卷积，剪枝思路和 Bottle Neck 同理。

此操作意在将原 CNN 等价拆分成负责保持性能的部分和负责修剪的部分，前者通过初始化 1 × 1 卷积参数为单位矩阵，不改变原目标函数、保持精度不降低，后者通过去掉某些通道，取得更高压缩率、更少精度损失。然后将原始卷积和 1 × 1 的卷积线性组合成一个卷积，并且删除通道成一个更小的模型。具体剪枝过程如图 4 所示。

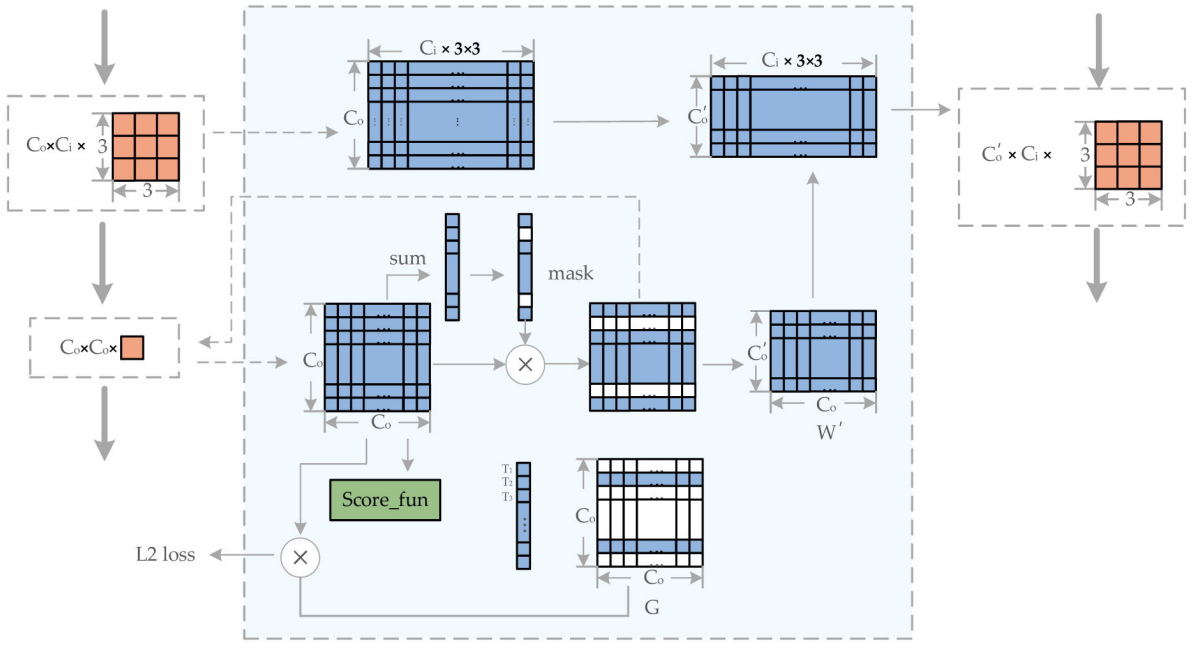


Figure 4. Schematic diagram of pruning process
图 4. 剪枝过程原理图

修剪 Basic Block 中 3 × 3 卷积层时，输入通道是 C_i，输出通道是 C_o，加入的 1 × 1 的卷积输入通道和输出通道都为 C_o，对于卷积核大小为 1 × 1 的 2D 卷积，它的参数 $F_{(1\times1)} \in R^{C_o\times C_o\times1\times1}$ 可以转化为一个二维的矩阵 $W \in R^{C_o\times C_o}$，W 的行和列分别对应卷积的输入通道和输出通道，将 W 中某一行置零相当于对卷积对应的输入通道进行剪枝。为了确定 W 中每个通道的重要性，本文通过估计将某一参数(w)置 0 后对损失的影响来评价该参数的重要性。网络的输入为 x，标签为 y，损失函数为 L。表示网络参数的集合。通常，网络的参数都处于 0 附近的小范围内，因此某一参数 w 置 0 后对损失 L 的变化可以有效地近似为损失 L 对 w 的一阶泰勒级数展开。计算方式如下：

$$L(x,y,\Theta_{w\leftarrow0})-L(x,y,\Theta)=\frac{\partial L(x,y,\Theta)}{\partial w}=T(x,y,w) \tag{1}$$

其中 T 表示参数的重要性分数，T 越大，将该参数置 0 后对损失的影响越大。即该参数越重要。W 中输入通道 p 的重要性 $T_p \in R^{C_o}$ 可以表示为

$$T_p = \left| \sum_{q=0}^{C_o} T\left(x, y, W_{p,q}\right) \right| \tag{2}$$

用 Ψ 表示网络中所有剪枝模块的 T_p 的值集合。因为本文使用的网络与传统剪枝方法针对的分类网络相比，参数量较小。对所有参数都添加正则化会导致模型欠拟合。所以不同于对所有权重都添加正则化的传统剪枝方法，本文只对分数最小的前 Q 个参数添加正则损失，而其他的参数不添加正则损失。则剪枝模块的参数在第 t 步更新方式如下：

$$W(t+1) \leftarrow W(t) - \alpha\left(Z(t+1) - \eta G(t) W(t)\right) \tag{3}$$

$$G_{p,:}(t) = \begin{cases} 1 & \text{当} T_p(t) < \Psi\text{中第Q大的值} \\ 0 & \text{其他} \end{cases} \tag{4}$$

其中 α 表示学习率，η 是普通权重的衰减系数，$Z(t+1)$ 是根据优化器和损失函数计算的梯度项，$G \in R^{C_o \times C_o}$ 为掩码矩阵，用于选择是否对权重添加正则化损失。对于权重 $W(t)$ 中重要性高的通道，掩码矩阵 G 上的对应位置为 0，在更新时不考虑正则化损失。对于 $W(t)$ 中重要性低的权重，掩码矩阵 G 上的对应位置为 1，在更新时通过正则化损失使其逐渐降低为 0。在训练阶段，如果 W(t)中的某一个输入通道的参数之和小于阈值 τ，则将 mask 中的对应通道置零。

$$mask_p = \begin{cases} 0 & \text{当} \left| \sum_{q=0}^{C_o} W_{p,q} \right| < \tau \\ 1 & \text{其他} \end{cases} \tag{5}$$

训练结束后，mask 为 0 的位置对应 W 中需要被剪枝的通道。$W' \in R^{C_o' \times C_o}$ 表示 W 被剪枝后的结果。W' 对应的是输入通道为 C_o'，输出通道为 C_o，卷积核大小为 1 × 1 卷积的权重 $F_{P_{(1\times1)}} \in R^{\left(C_O' \times C_O \times 1 \times 1\right)}$。在测试阶段，K × K 的卷积和 1 × 1 的卷积共同转化成一个 K × K 的卷积，转换过程公式如下：

$$F'_{(K\times K)} = F_{(K\times K)} \otimes TRANS\left(F_{P_{(1\times1)}}\right) \tag{6}$$

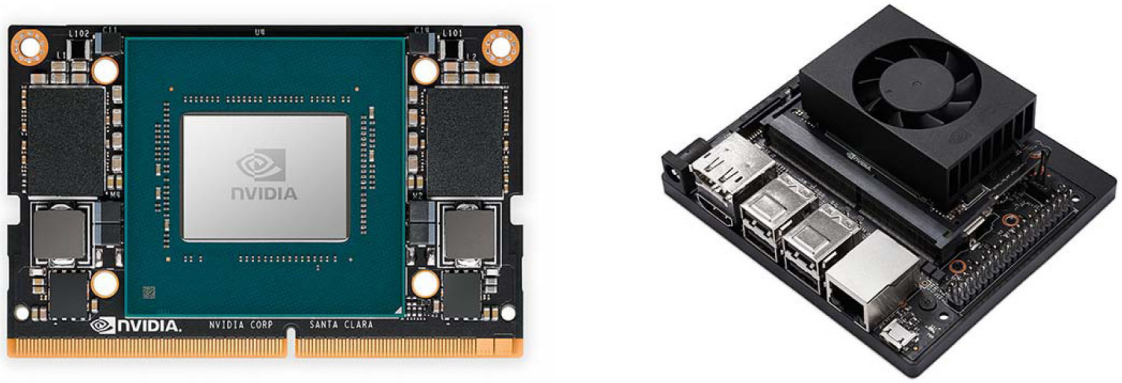
其中 ⊗ 表示卷积操作。因为 $C_o < C_o'$，所以变换后的 $F'_{(K\times K)} \in R^{\left(C_O' \times C_O \times K \times K\right)}$ 的参数小于 $F_{(K\times K)}$，从而减少了推理阶段的计算量。

3. 基于嵌入式的人体姿态估计系统

3.1. 系统组成及工作原理

针对在人体姿态估计时嵌入式应用成本高、检测速度慢难以满足实际应用需求等难点问题，本文设计了基于嵌入式 Jetson Xavier NX 平台的人体姿态估计系统，由 Jetson Xavier NX 开发板、显示屏、热红外摄像头以及鼠标键盘组成。如图 5 所示，Jetson Xavier NX 是一款小巧而强大的人工智能开发平台[11]。尽管体积小巧，其计算性能却极为出色，在 15W 的功耗下，它能够达到每秒 14 万亿次的计算速度。这款开发板规模仅为 103 mm × 90.5 mm × 34 mm，并且支持多网络的并行计算技术。此外，它还提供了 HDMI 和 DP 的视频输出接口，USB 方面，它支持 USB 3.1 (4 端口)、以及 USB 2.0 Micro-B (1 端口)。

基于嵌入式 Jetson Xavier NX 平台的人体姿态估计系统首先利用红外摄像头传入的视频流作为输入，将视频流经过 GStreamer 框架处理，通过逐帧抽取图像作为实时检测系统的输入，采用多线程的设计策略增强嵌入式系统的实时性。以 Jetson Xavier NX 作为高性能嵌入式计算平台执行 TRT 推理引擎得到检测结果，并将检测结果通过屏幕进行可视化显示，实现在功耗低、体积小的嵌入式平台完成实时的基于红外视频的人体站坐躺行为检测。



（a）Jetson Xavier NX 模组　　（b）Jetson Xavier NX 开发套件

Figure 5. Jetson Xavier NX embedded platform
图 5. Jetson XavierNX 嵌入式平台

3.2. TensorRT 优化及模型部署

在进行模型推理时，网络的不同层之间频繁调用函数会消耗很多时间[12]。我们通过采用 TensorRT 对人体姿态估计模型进行优化，可以提高大约 15 倍[13]的网络吞吐量，同时将功耗降低 55% [14]。本文的人体姿态估计模型中的所有张量以 32 位浮点数的形式表示，为了在确保人体关键节点检测精度的前提下，进一步提高处理能力，本文将模型通过 TensorRT 的量化方案，转换为 FP16 的精度。此外，进行模型结构的优化也是提升效率的重要手段，包括采用垂直和水平融合两种方式[15]。这样的优化操作可以缩减网络的深度和宽度，同时减少了在姿态估计模型中函数调用的数量，进而提升推理速度。

针对嵌入式设备对检测实时性的需求，本文采用了偏向底层语言的 C++编程来加载 TensorRT 模型，加速实施人体姿态预估模型的部署。在对 Pytorch 训练出的模型启用 TensorRT 的推理加速前，先将 Pytorch 保存的模型文件转化为 ONNX 模型，流程如图 6 所示：

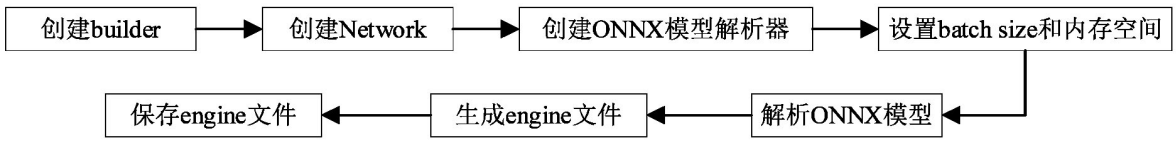


Figure 6. ONNX-TensorRT model conversion flowchart
图 6. ONNX-TensorRT 模型转换流程图

在 TensorRT 解析了人体姿态预估模型后，就可以将模型和对应的层进行匹配。首先，创建 engine 类作为构建器，定义网络，然后创建 ONNX 的解析器并进行解析。接下来，配置网络构建参数，让 TensorRT 可以进行对应的模型构建和优化。有了网络定义和 builder 配置，就使用 builder 创建引擎。然后将模型序列化并保存，最后创建推理运行时 runtime，并反序列化模型以获得推理 engine，随后执行推理和后处理操作，在 Jetson Xavier NX 上编译项目，对性能进行优化。

4. 实验结果与分析

4.1. 数据集及评价指标

4.1.1. 数据集

COCO2017 [16]数据集包含 80 类图像和大约 25 万个人类实例。整个数据集划分如下：训练集 train2017

包含 57000 张图像和 150 万个人类实例；测试集 Test-dev2017 包含 20000 个图像；验证集 val2017 包含 5000 张图像。

UCH-Thermal-Pose [17]数据集由 A 和 B 两组组成，Set-A 是通过收集不同公共来源的热图像并对其进行标注而构建的。由 800 张图像组成，包括室内和室外环境。其中 600 张图像用于训练集，其余 200 张用作验证集。B 组使用由 FLIR FC-690 S 热像仪捕获的 104 幅热图像组成，分辨率为 640 × 480，共有 3 个不同的相机角度。数据集的每张图像均使用 LabelMe 工具手动标注人体关键点。

4.1.2. 评价指标

关键点检测需要同时检测人体目标和定位人体关键点的坐标，这是一项检测和关键点估计同时进行的任务。本姿态估计模型采用对象关键点相似度(Object Keypoint Similarity, OKS)来计算真实关节点与预测关节点之间的相似度。一个人的真实关节点和预测关节点分别定义为：

$$[x_1, y_1, v_1, \cdots, x_N, y_N, v_N] \tag{7}$$

$$[\hat{x}_1, \hat{y}_1, \hat{v}_1, \cdots, \hat{x}_N, \hat{y}_N, \hat{v}_N] \tag{8}$$

其中 N 表示图像上关键点的个数。x_i 和 y_i 是关键点的坐标。v_i 为真实值的是否可见的标记。可以计算出每个人体实例关节点的真实值与检测到的关键点之间的欧几里得距离：

$$d_i = \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2} \tag{9}$$

OKS 的公式如下：

$$OKS = \frac{\sum_i \exp\left(-d_i^2 / 2s^2 k_i^2\right) \delta(v_i > 0)}{\sum_i \delta(v_i > 0)} \tag{10}$$

其中，k_i 越大代表这个人体实例关节点标注难度越高。s 为对象的因子，是第 i 个人体关键点的归一化因子，是通过计算数据集中所有真实值标准差而来的，可以反映出此人体实例关节点标注时候的标准差。本文性能指标遵循 mAP，即 OKS 以 0.05 为间隔从 0.5 取到 0.95，得到的所有的 AP 结果取平均。AP^{50} 和 AP^{75} 是在 0.50 和 0.75 阈值下预测数据的评价平均精度。AP^M 和 AP^L 分别表示中等尺度和大尺度下的 AP 值，计算方法与计算 mAP 类似，只是在计算每个尺度下的 AP 时使用了不同的框尺寸和长宽比。

4.2. 实验结果

本文使用 Microsoft COCO 2017 数据集对模型进行预训练。将图像转换为灰度，设置输入大小为 256 × 192，每个图像将通过一系列的数据增强操作，包括随机旋转([−30˚, 30˚])，随机缩放([0.75, 1.5])，随机平移([−40, 40])以及随即翻转的操作。总训练回合数为 300 次，学习率 LR 初始设置为 2e-3；在第 200 次衰变到 1e-4；在第 260 次衰变到 1e-5。在训练中设置 Adam 优化器计算热图损失。并将两者的损失权值分别设为 1 和 1e-3，以平衡热图损失与分组损失的关系。在剪枝过程中随着权值衰减和剪枝率的增加，剪枝规模越大，但是随着迭代次数的增加，精度也会出现波动现象。当权值衰减为 0.01，剪枝率为 0.3 时，精度曲线变化趋势比较平稳，在 300 个 epoch 时，实现了达到大约 46%的剪枝规模的同时，仅有 0.6%的精度下降。实验所使用的软硬件如表 1 所示，部分测试集检测结果可视化如图 7。

Table 1. Hardware and software used in the experiment
表 1. 实验使用的软硬件

软硬件平台	训练服务器
操作系统	Windows 10

续表

深度学习框架	Py Torch 1.11.0
CPU	Intel (R) Core (TM) i5-1035G1 CPU @1.7 GHz 2.19 GHz
GPU	RTX 3090
CUDA	10.1
cuDNN	8.4

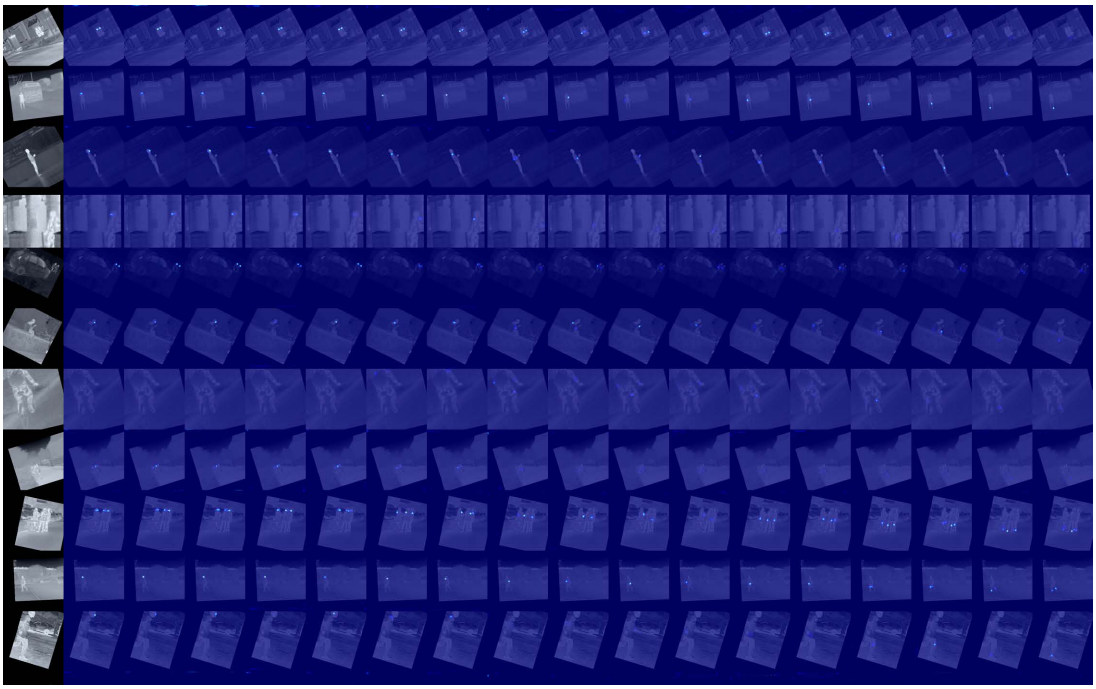


Figure 7. Prediction heatmap results on the UCH dataset
图 7. UCH 数据集上预测热图结果

COCO 数据集上与其他经典轻量级 HRNet 算法精度对比如表 2，可以看出本文的模型在精度上的显著优势。模型剪枝前后在 COCO 数据集上数据对比如表 3，可以看出本文模型经过剪枝后虽平均精确率较原模型略有下降，但可以说优于同类型轻量级模型，且推理速度有大幅提升，参数量和模型体积甚至几乎仅有原模型的一半。

Table 2. Comparison of P-HRNet with other classic lightweight HRNet algorithms on the COCO dataset
表 2. COCO 数据集上 P-HRNet 与其他经典轻量级 HRNet 算法对比

网络模型	mAP	AP^{50}	AP^{75}	AP^{M}	AP^{L}
Dite-HRNet-30	70.6	90.8	78.2	67.4	76.1
Lite-HRNet-30	69.7	90.7	77.5	66.9	75.0
Lite-HRNet-18	66.9	89.4	74.4	64.0	72.2
P-HRNet	75.3	90.6	80.6	68.7	78.4

Table 3. Comparison of model data before and after pruning on the COCO dataset
表 3. 剪枝前后模型在 COCO 数据集上对比

网络模型	mAP	AR	GFLOPs	参数量/M	模型体积/M
HRNet	75.9	78.9	7.2	28.6	116.6
P-HRNet	75.3	73.9	4.5	15.5	63.8

预训练后使用来自 UCH-Thermal-Pose Set-A 的 600 张注释图像对改进后的模型进行微调。学习率 LR 设置为 1e-3，对模型进行 50 个 epoch 的迭代训练。改进前后模型在 UCH-Thermal-Pose set-A 上的 200 张红外图像组成的验证子集上进行评估平均精度如表 4。本文还将改进模型和其他优秀的人体姿态估计网络在相同的红外图像测试集上进行了对比，结果如表 5。不难看出经过改进的模型在红外数据集上的表现优于原模型，与其他优秀算法对比优势明显，证明本文的可见域与热域迁移学习的策略不失为针对红外图像的关键点检测领域一种可靠的解决方案。

Table 4. Comparison of model accuracy on the UCH dataset before and after transfer learning
表 4. 迁移学习前后模型在 UCH 数据集上的精度对比

网络模型	mAP	AP^{50}	AP^{75}	AP^{M}	AP^{L}
HRNet	62.8	89.8	73.8	60.1	68.9
P-HRNet	74.2	90.8	79.0	66.1	76.6

Table 5. Comparison with other excellent algorithms on the UCH dataset
表 5. UCH 数据集上与其他优秀算法对比

方法	主干网络	参数量/M	GFLOPs	mAP
Hourglass	Hourglass	277.8	206.9	56.6
PersonLab	ResNet-152	68.7	405.5	66.5
HRNet	HRNet-W32	28.5	38.9	64.1
HigherHRNet	HRNet-W32	28.6	47.9	63.6
改进 HRNet	改进 HRNet-W32	15.0	20.3	74.2

4.3. Jetson Xavier NX 上的结果与分析

在 Jetson Xavier NX 平台进行模型优化效果实验，环境配置如下：Ubuntu 20.04 操作系统、CUDA 11.4、cuDNN 8.4、Jetpack 5.0.2、TensorRT 8.4.1.5、OpencV4.5.4，成功部署后的检测效果如图 8。

Figure 8. Detection performance on the Jetson Xavier NX
图 8. Jetson Xavier NX 平台检测效果

经过 TensorRT 优加速后，网络的单帧平均推理速度最高达到 33 fps，优化后推理速度较原模型推理速度提高 61%，在低功耗嵌入式平台的推理速度基本满足了嵌入式实时人体姿态估计的要求。

5. 结束语

本文设计了面向嵌入式平台的红外人体姿态估计系统。针对嵌入式平台人体姿态估计网络的实时效果，本文通过引入结合结构重参数化方法的剪枝模块，改进 HRNet 的网络结构，在对精度牺牲较小的情

况下大幅减小模型体积。又针对目前人体姿态估计算法几乎都是基于可见光，低照度环境下的姿态识别效果不佳的情况，提出使用迁移学习方法先用 COCO 数据集的灰度图像训练，然后用加入了瓶颈结构的 P-HRNet 在 UCH-Thermal-Pose 数据库上进行热域的微调，实现了针对红外场景的人体姿态估计。接着进一步采用 GStreamer 视频流处理框架结合 TensorRT 对模型进行了优化和部署，使推理速度能够基本满足实时性需求。未来研究可以收集更丰富场景下的红外人体图像进行标注，提高模型精度并完善丰富嵌入式人体姿态估计系统功能。

基金项目

辽宁省教育厅项目(LJKZ0174)。

参考文献

[1] 冯晓月, 宋杰. 二维人体姿态估计研究进展[J]. 计算机科学, 2020, 47(11): 128-136.

[2] Toshev, A. and Szegedy, C. (2014) Deeppose: Human Pose Estimation via Deep Neural Networks. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 1653-1660. https://doi.org/10.1109/CVPR.2014.214

[3] Wei, S.E., Ramakrishna, V., Kanade, T., *et al*. (2016) Convolutional Pose Machines. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, USA, 27-30 June 2016, 4724-4732. https://doi.org/10.1109/CVPR.2016.511

[4] Xiao, B., Wu, H. and Wei, Y. (2018) Simple Baselines for Human Pose Estimation and Tracking. *European Conference on Computer Vision*, Munich, Germany, 8-14 September 2018, 466-481. https://doi.org/10.1007/978-3-030-01231-1_29

[5] Sun, K., Xiao, B., Liu, D., *et al*. (2019) Deep High-Resolution Representation Learning for Human Pose Estimation. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, USA, 15-20 June 2019, 5693-5703. https://doi.org/10.1109/CVPR.2019.00584

[6] 臧影. 低照度下人体姿态估计及行为识别研究[D]: [博士学位论文]. 北京: 中国科学院大学(中国科学院沈阳计算技术研究所), 2022.

[7] Weiss, K., Khoshgoftaar, T.M., Wang, D.D., *et al*. (2016) A Survey of Transfer Learning. *Journal of Big Data*, **3**, Article No. 9. https://doi.org/10.1186/s40537-016-0043-6

[8] Zhuang, F., *et al*. (2021) A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, **109**, 43-76. https://doi.org/10.1109/JPROC.2020.3004555

[9] Ding, X., Zhang, X., Han, J., *et al*. (2021) Diverse Branch Block: Building a Convolution as an Inception-Like Unit. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 10886-10895. https://doi.org/10.1109/CVPR46437.2021.01074

[10] Ding, X., Zhou, X., Guo, Y., *et al*. (2019) Global Sparse Momentum SGD for Pruning Very Deep Neural Networks. *Advances in Neural Information Processing Systems*, **32**, 1-13.

[11] 唐乾琛. 英伟达公司发布全球最小边缘 AI 超级计算模块[J]. 科技中国, 2019(12): 108.

[12] 张宇昂, 李琦, 薛芳芳, 等. 基于 Jetson TX2 的路面裂缝检测系统设计[J]. 公路, 2023, 68(12): 337-344.

[13] 周立君, 刘宇, 白璐, 等. 使用 TensorRT 进行深度学习推理[J]. 应用光学, 2020, 41(2): 337-341.

[14] Jeong, E.J., Kim, J. and Ha, S. (2022) TensorRT-Based Framework and Optimization Methodology for Deep Learning Inference on Jetson Boards. *ACM Transactions on Embedded Computing Systems* (*TECS*), **21**, 1-26.

[15] Song, Z. and Shui, K. (2019) Research on the Acceleration Effect of Tensorrt in Deep Learning. *Scientific Journal of Intelligent Systems Research*, **1**, 45-50.

[16] Lin, T.Y., Maire, M., Belongie, S., *et al*. (2014) Microsoft Coco: Common Objects in Context. *European Conference on Computer Vision*. Switzerland, Zurich, 6-12 September 2004, 740-755. https://doi.org/10.1007/978-3-319-10602-1_48

[17] Smith, J., Loncomilla, P. and Ruiz-Del-Solar, J. (2023) Human Pose Estimation Using Thermal Images. *IEEE Access*, **11**, 35352-35370. https://doi.org/10.1109/ACCESS.2023.3264714