

基于改进 FastPose 的轻量型人体姿态估计算法



祝施冉^{1,*}, 万擎², 贾世杰³

¹大连交通大学自动化与电气工程学院, 辽宁大连 116028

²天津轨道交通运营集团有限公司, 天津 300380

³大连交通大学计算机与通信工程学院, 辽宁大连 116028

摘要: 人体姿态估计网络模型性能逐渐提高, 过深的网络结构带来的是更为庞大的参数量和复杂计算量。针对以上问题, 提出了 FastPose-Lite 轻量型人体姿态估计网络, 该网络由 GSE-ResNet 特征提取网络、上采样 DUC 模块以及 CBAM 模块组成。其中构成 GSE-ResNet 特征提取网络的 GBNK 基础模块是由 Ghost 模块和 SE 模块组成, 一方面, 为减少参数量和计算量提出利用 Ghost 模块代替传统卷积模块; 另一方面, 为保证网络模型性能不变, 引入 SE 注意力机制模块。在空间和通道两方面为增强了网络模型对特征信息的处理能力, 将 CBAM 模块引入到上采样 DUC 模块之间, 减少了上采样过程中带来的损失。在 COCO 数据集上的实验结果表明, 本文提出的 FastPose-Lite 相对于 FastPose 网络模型的参数量和计算量分别减少了 51.4% 和 50.8%; 与 SHN、CPN 和 SimpleBaseline 等常见的热门网络模型相比, FastPose-Lite 网络模型不仅参数量和计算量更少, 而且预测精度更高。

关键词: 人体姿态估计; FastPose; Ghost 模块; 注意力机制; 轻量化

DOI: [10.57237/j.cst.2023.02.003](https://doi.org/10.57237/j.cst.2023.02.003)

Lightweight Human Pose Estimation Algorithm Based on Improved FastPose

Zhu Shiran^{1,*}, Wan Qing², Jia Shijie³

¹School of Automation and Electrical Engineering, Dalian Jiaotong University, Dalian 116028, China

²Tianjin Rail Transit Group Corporation, Tianjin 300380, China

³School of Computer and Communication Engineering, Dalian Jiaotong University, Dalian 116028, China

Abstract: The performance of human pose estimation network model is gradually improved, and the over-deep network structure brings a large number of parameters and complex calculation. To solve these problems, the FastPose-Lite lightweight human pose estimation network is proposed, which is composed of GSE-ResNet feature extraction network, up-sampling DUC module and CBAM module. The basic GBNK module of GSE-ResNet feature extraction network is composed of Ghost module and SE module. On the one hand, Ghost module is proposed to replace the traditional convolutional module in order to reduce the number of parameters and calculation. On the other hand, in order to keep the performance of the network model unchanged, SE attention mechanism module is introduced. In order to enhance the processing ability of the network model to the feature information in both spatial and channel aspects, CBAM module is

*通信作者: 祝施冉, 793182834@qq.com, 1013825469@qq.com

收稿日期: 2023-03-01; 接受日期: 2023-04-06; 在线出版日期: 2023-04-07

<http://www.computscitech.com>

introduced into the up-sampling DUC module to reduce the loss in the up-sampling process. The experimental results on the COCO dataset show that the proposed FastPose-Lite reduce the number of parameters and calculation by 51.4% and 50.8% respectively compared with the FastPose network model. Compared with the common popular network models such as SHN, CPN, and SimpleBaseline, the FastPose-Lite network model not only has fewer parameters and computations, but also has higher prediction accuracy.

Keywords: Human Pose Estimation; FastPose; Ghost Module; Attention Mechanism; Lightweight

1 引言

人体姿态估计是从图像或视频中检测出人的骨骼关节点。其优点是不易受外观因素影响，在监控摄像头、虚拟现实等领域有着不错的应用前景。传统方法使用人工设定的 HOG [1]和 SHIFT [2]来提取特征，不能充分利用图像的特征信息，导致算法容易受到视角、遮挡等客观因素的干扰。

深度学习的不断发展，带来了可从图像中提取丰富语义信息的卷积神经网络（CNN）[3]，使得人体姿态估计领域有了不错的发展。CPM [4]首次使用热图对骨骼关节点建模。堆叠式沙漏网络（SHN）[5]通过重复的下采样再上采样获得更强的语义信息。随着检测技术的不断发展，现阶段主要是针对多人的姿态估计，分为自上向下和自下向上两种形式。

自上向下首先利用目标检测器检测出图像中的人，再输入单人姿态估计网络中检测关键点，最后映射到原图中。自上向下预测的结果更为准确，但会受到检测器性能的影响。RMPE [6]是这样的思路，后期提出的 Alpha Pose [7]更新了用于快速和精细定位的对称积分关键点回归（SIKR）和用于消除冗余人工检测的参数姿态非极大抑制（P-NMS），同时提出了 FastPose 姿态估计网络。自下向上先检测出图像中所有人体的骨骼关节点，再通过匹配连接得到最终姿态。这种方法虽然速度快实时性高，但当目标位置较近时，易出现错误链接。其中最经典的算法是 Open Pose [8]，通过 CPM 确定所有人在图像中的关键点，并根据部分亲和场（PAF）将关节点与躯干拼接生成人体姿态。

人体姿态估计算法性能的不不断提升，伴随着其参数数量和计算量的增加，而与之匹配的硬件设备在实际应用中很少，因此需要考虑适当减少模型的参数量，降低计算量的同时，不改变性能和对关键点的预测精度。MobileNet [9]是由 Google 提出的轻量型的网络，通过使用深度可分离卷积替换传统的卷积层来降低计算量。MobileNetV2 [10]中引入了线性瓶颈的倒置残差

结构。MobileNetV3 [11]在 V1、V2 的基础上，引入了 SE 注意力模块调整通道权重，使用 H-Swish 函数，减少运算量，提升网络精度。另一方面，设备的运算能力有限，为防止信息超载，可以通过引入注意力模块来提升轻量型网络的效率，合理分配计算资源。以 SE 注意力机制模块为核心的压缩激励网络 SENet [12]通过全局池化构建通道间的关系来增强网络模型的表征能力，达到轻量化的目的。

基于上述研究，针对 FastPose 网络模型参数量多、计算量大的问题，本文采用轻量级的 Ghost 模块取代骨干网络中的传统卷积模块，以减少冗余的参数量和计算量，并引入 SE 注意力模块，保证模型精度不发生改变。同时在上采样阶段引入 CBAM 注意力机制模块，依次从空间和通道两方面增强网络模型对特征信息的处理能力，有助于在上采样过程整合特征信息，使其达到与 FastPose 网络模型对关键点的预测精度相同的水平。

2 本文算法

2.1 网络结构

本文采用自上向下的人体姿态估计算法，其网络架构图如图 1 所示，先用 YOLOV4 [13]人体检测器对图像中的每个人进行检测，再对子图像中的人体进行关键点预测。为减少模型参数量和计算量，本文基于 FastPose 网络提出了轻量型人体姿态估计网络 FastPose-Lite，并设计了轻量型的 GSE-ResNet 特征提取骨干网络，以替换 FastPose 中的 ResNet [14]骨干网络，在密集上采样卷积模块中，引入 CBAM [15]注意力模块，依次从通道与空间两个维度上增加注意力，确保其能够集成到 DUC 模块中，最后以一个 1×1 的卷积层生成热图，实现人体姿态估计。

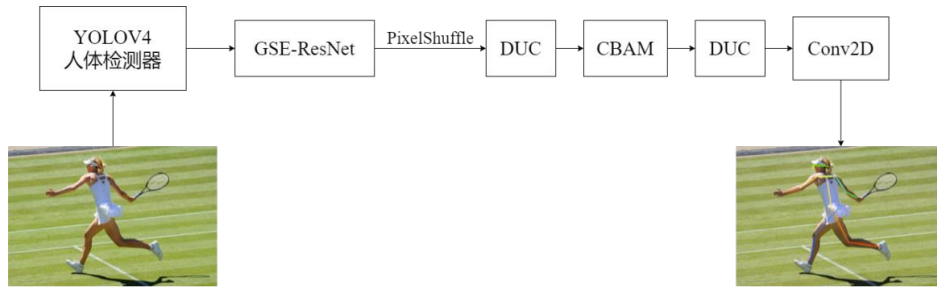


图 1 网络架构图

2.1.1 GSE-ResNet 特征提取网络

参考 ResNet50 特征提取网络，提出了轻量型的 GSE-ResNet 特征提取网络，具体结构如图 2 所示：

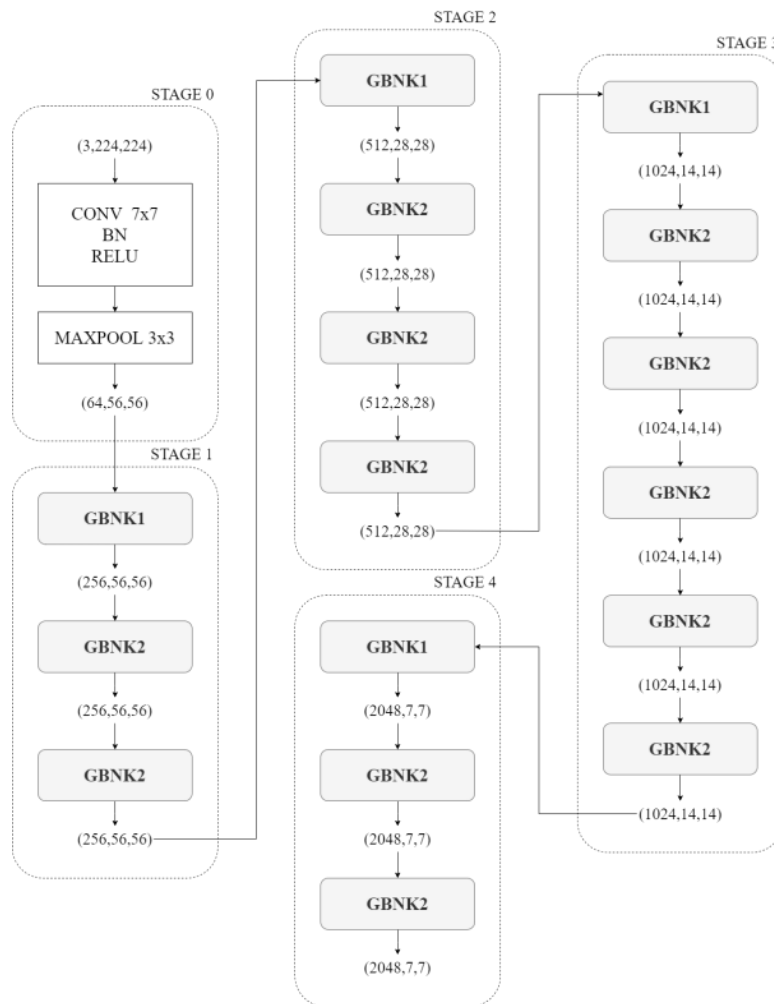


图 2 GSE-ResNet 网络结构

GSE-ResNet 网络继承了 ResNet 网络原有的模型架构，包含 Stage0、Stage1、Stage2、Stage3 和 Stage4 共五个阶段。具体处理过程如下：

Stage0 阶段为预处理阶段，将输入特征的维度变

为 (56, 56)，通道数变为 64；

Stage1 阶段是以 Stage0 阶段输出的特征图作为输入，依次通过 3 个 GBNK 模块处理特征图；

Stage2、Stage3、Stage4 阶段按照同上一阶段一样

的步骤分别依次通过4、6、3个GBNK模块处理特征图。

接下来将详细介绍组成GSE-ResNet特征提取网络的模块结构。

(1) Ghost 模块

Ghost [16]模块相较于传统卷积模块，以相对较少的计算量，输出与传统卷积模块同等数量且相似的特征图。如图3所示，该模块的具体处理过程如下：

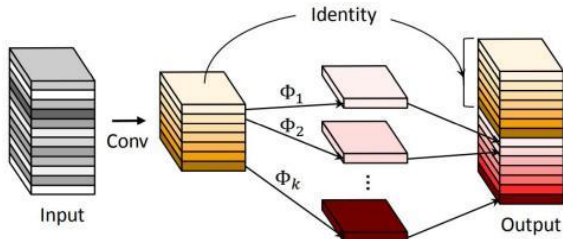


图3 Ghost 模块

首先，对输入特征图采用 1×1 卷积操作生成基础特征图；

然后，将步骤一得到的输出采用 3×3 的分组卷积线性运算生成Ghost特征图；

最后，将前两步的输出融合，得到与传统卷积操作过程相似的特征图。

Ghost模块与传统卷积的参数比 r_P 和计算量比 r_G ，如公式(1)和(2)所示：

$$r_P = \frac{n \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot d \cdot d} \approx \frac{s \cdot c}{s+c-1} \approx \frac{sc}{c} = s \quad (1)$$

$$r_G = \frac{n \cdot h' \cdot w' \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot h' \cdot w' \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot h' \cdot w' \cdot d \cdot d} \approx \frac{s \cdot c}{s+c-1} \approx s \quad (2)$$

其中， c 和 n 分别为输入和输出特征图的通道数；

h' 、 w' 分别为输出特征图的高和宽；

k 为传统卷积核大小， d 为Ghost模块中线性运算卷积核大小，且 $d \approx k$ ；

s 为输出特征图中每个基础特征图对应的冗余特征的个数，且 $s \ll c$ ；

分子是传统卷积，分母是Ghost模块。相较于传统卷积模块Ghost模块的参数量和计算量都更低，因而可以改善骨干网络的参数量优化计算量。

(2) GBNK 模块

GBNK1和GBNK2是组成GSE-ResNet特征提取网络的基础模块，其由轻量型Ghost模块和SE注意力模块共同组成，具体结构如图4所示，其中(a)为GBNK1，(b)为GBNK2。

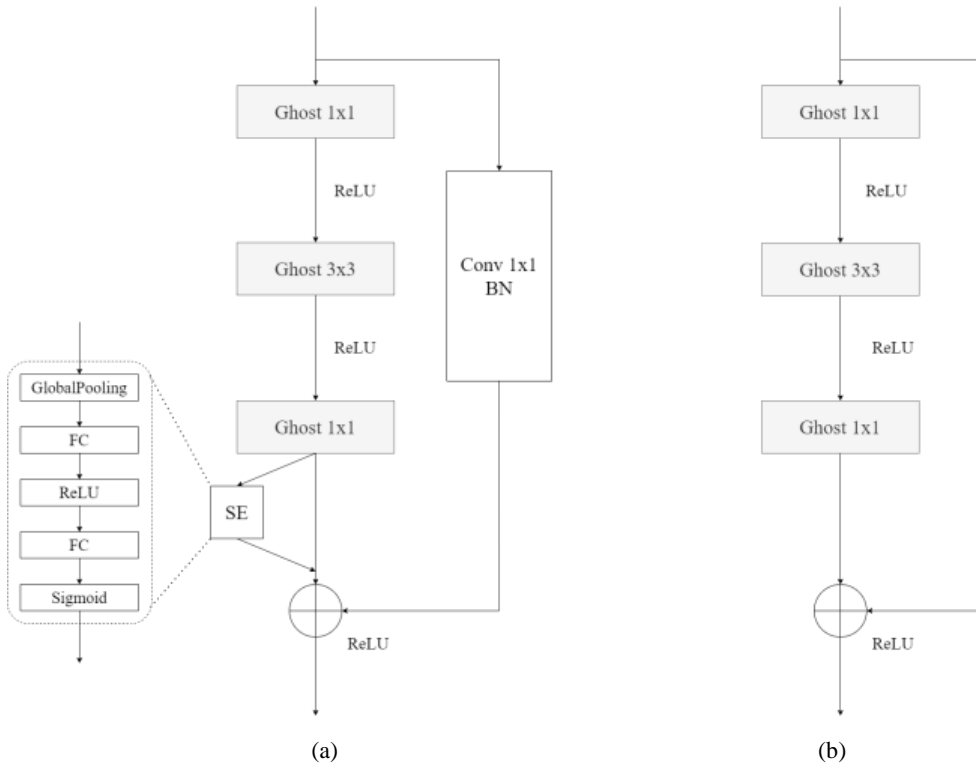


图4 GBNK 模块

GBNK 模块参考的是构成 ResNet50 网络结构中的瓶颈模块 Bottleneck。首先,采用 Ghost 模块替换瓶颈模块中的传统卷积模块,以压缩的方式减少深度网络的参数量,同时需要重新考虑层间系数,以保证网络模型的特征提取能力。其次,在 GBNK1 模块中引入了用于获取通道特征信息的 SE 注意力模块,SE 注意力模块通过压缩和激励的方式构建通道间的关系,并对通道特征信息进行加权处理,以增强网络的表征能力,从而做到适当减少模型的参数量和计算量的同时不改变模型性能和对关键点的预测精度。设计 GBNK 模块时,保留了瓶颈模块的整体结构,缩减中间特征图的通道数更利于构建网络。下面分别给出 GBNK1 和 GBNK2 的参数量 P 和计算量 G:

$$P_{GBNK1} = P_{G1} + P_{G3} + P_{G1} + P_{SE} \quad (3)$$

$$G_{GBNK1} = G_{G1} + G_{G3} + G_{G1} + G_{SE} \quad (4)$$

$$P_{GBNK2} = P_{G1} + P_{G3} + P_{G1} \quad (5)$$

$$G_{GBNK2} = G_{G1} + G_{G3} + G_{G1} \quad (6)$$

其中, G1 表示的是卷积核为 1 的 Ghost 模块, G3 表示的是卷积核为 3 的 Ghost 模块, SE 表示的是 SE 注意力机制模块。

2.1.2 DUC 模块

传统的网络模型上采样阶段会选择双线性插值,输出结果的分辨率与输入相同,但由于双线性插值无法学习,因此会丢失部分重要信息。而密集上采样卷积(DUC)[17]会生成像素级的预测,成功解码并获取到在双线性插值上采样时丢失的更重要的信息。具体过程如图 5 所示:

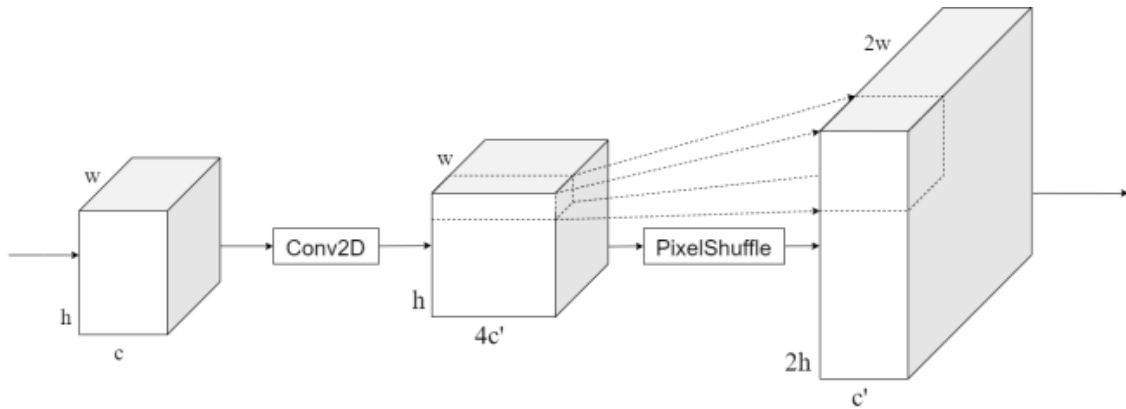


图 5 DUC 模块

输入维度为 $h \times w \times c$ 的特征图,首先通过一个 3×3 的卷积层和 BN 层,得到维度为 $h \times w \times 4c'$ 的特征图,然后进行像素混洗(Pixel Shuffle)操作将其重构为 $2h \times 2w \times c'$ 的特征图。

2.1.3 CBAM 注意力模块

卷积注意力机制模块(CBAM)可用于前馈卷积

神经网络,其结构如图 6 所示。本文将 CBAM 模块添加到两个 DUC 模块之间,依次从通道和空间两个维度判断特征图,输出含有注意力权重的特征图并进行自适应优化,以加强模型对特征信息的精准定位,有助于在上采样过程整合特征信息,提升预测精度。最终在降低模型复杂程度的同时又能保证性能,达到了轻量化的目的。

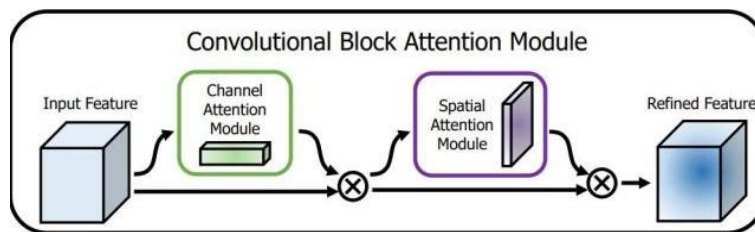


图 6 CBAM 注意力机制模块结构图

2.2 损失函数

本文所选损失函数为均方误差损失函数 (MSELoss)，反映的是预测量与被预测量之间差异程度，公式如式 3 所示：

$$Loss(x, y) = \frac{1}{n} \sum_{i=1}^N (x_i - y_i)^2 \quad (7)$$

其中， x 为真实热图 (Ground Truth Heatmap)， y 为预测热图 (Predicted Heatmap)。每个关键点的真实值是以该关键点为峰值位置的 2D 高斯函数，以每个关键点的真实位置为中心，标准差为 1 像素，生成真实值热图。于是定义损失函数为真实值生成的热图与预测生成的热图之间的均方差。

3 实验结果与分析

3.1 数据集

COCO [18] 数据集是专门用来实验研究的数据集，如目标检测、人体姿态估计、语义分割等。它包含了 17 个标注的人体骨骼关节点，标注顺序如图 7 所示。一共 118000 个数据集用于训练，5000 个数据集用于验证，41000 个数据集用于测试。本文的网络模型在 COCO 2017 train 上训练，在 COCO 2017 test 上与最先进的模型进行对比。

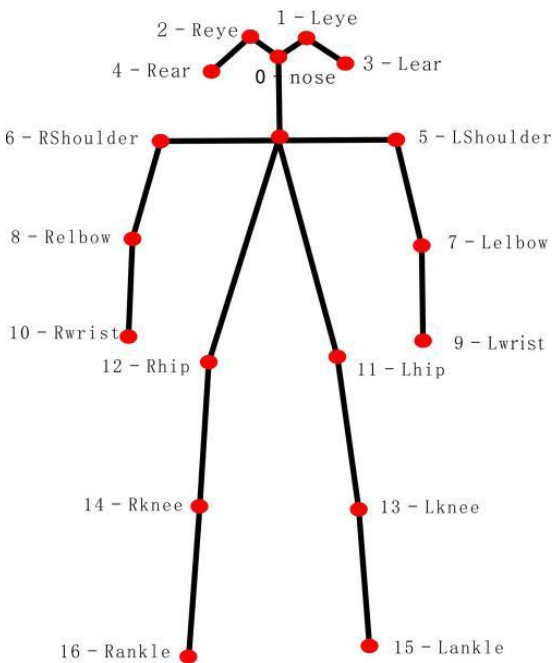


图 7 COCO 数据集关键点标注位置

3.2 实验设置

本文开发平台为 Windows10 操作系统，编程语言为 Python3.8，编程环境为 Cuda10.1，GPU 为 RTX 3090，在深度学习框架 Pytorch1.12.0 下完成模型设计和训练。优化器为 Adam，batchsize 为 32，训练时初始学习率为 0.01，在训练到 280 个 epoch 时，学习率由 0.01 衰减至 0.0001 结束训练，共训练 280 个 epoch。

3.3 评价指标

精度方面本文采用 MS COCO 官方的目标关键点相似度 OKS 测量方式。其实现方法如式(8)所示，OKS 定义了不同关键点之间的相似性，其中， d_i 表示真实值与预测的关键点间的欧几里得距离， s 表示目标比重， v_i 表示真实关键点的可见度， k_i 表示控制衰减的常数。

$$OKS = \frac{\sum_i \exp(-d_i^2 / 2s^2 k_i^2) \delta(v_i > 0)}{\sum_i \delta(v_i > 0)} \quad (8)$$

定义 mAP 为 OKS = 0.50, 0.55, ..., 0.95 时，对应 10 个阈值所预测的关键点精度的平均值。

轻量化方面本文以参数量 (params) 和计算量 (FLOPs) 来描述，参数量指的是算法在空间维度的复杂度，参数量越大在硬件层面对计算机内存的消耗也越大；计算量指的是算法在时间维度的复杂度，网络模型执行时间的长短。

3.4 实验结果分析

为验证本文提出的人体姿态估计网络模型的效能，选择在 COCO 数据集上同常见的热门网络模型分别从精度和轻量化两方面进行对比实验，实验结果如表 1 所示。根据实验结果可以看出，本文提出的 FastPose-Lite 人体姿态估计网络与常见的热门网络模型在 COCO 数据集上相比，能够以较少的参数量和计算量，得到了一个不错的精度，mAP 可以达到 70.7%。

相较于 FastPose 网络，平均预测精度基本接近；轻量化方面，整体网络模型的参数量减少了 51.4%，计算量降低了 50.8%。同 SHN、CPN [19] 以及 SimpleBaseline [20] 这些热门的非轻量型网络模型相比，精度方面分别提升 3.8%、1.6% 以及 0.5%；轻量化方面参数量分别减少 43%、47% 以及 57.9%，计算量分别降低 79.7%、53.2% 以及 67.4%。相比于 HRNet [21] 网络模型，虽然平均精度均值略低，但在轻量化方面，参

数量降低了 49.8%，计算量减少了 81.9%，在 OKS=0.5 时预测的平均精度高出 HRNet 网络模型 1.2 个百分点。

实验结果表明，本文提出的 FastPose-Lite 网络模型与常见的热门网络模型在 COCO 数据集上相比，能

够以较少的参数量和较低的计算量，得到一个不错的精度，在达到与原网络模型性能相当的同时，参数量和计算量都减少了超过 50 个百分点，可以得出，本文提出的网络实现了对 FastPose 网络模型的轻量化。

表 1 COCO 数据集不同方法结果对比

Method	Input size	#Params/ 10^6	GFLOPs	mAP/%	AP ⁵⁰ /%	AP ⁷⁵ /%	AP ^M /%	AP ^L /%
Hourglass	256×192	25.1	14.3	66.9	-	-	-	-
CPN	256×192	27.0	6.2	68.6	-	-	-	-
SimpleBaseline	256×192	34.0	8.9	70.2	90.9	78.3	67.1	75.9
HRNet	384×288	28.5	16.0	73.4	89.5	80.7	70.2	80.1
FastPose	256×192	29.4	5.9	71.8	91.9	80.3	72.8	74.2
FastPose-Lite	256×192	14.3	2.9	70.7	90.7	80.4	72.0	73.9

3.4.1 可视化研究及分析

为直观地验证本文提出的人体姿态估计网络的检测效果，选择在 COCO 验证集上进行了可视化验证，随机选择有遮挡和无遮挡的图片，检测效果如图 8 所示。

可以看出当目标人体的骨骼关节点受到遮挡或交叉重合时，本文提出的 FastPose-Lite 网络模型，依然能够准确地检测出图像中人体的姿态。图(a)中两个人都是背面且无遮挡的情况，模型能够检测出两个人各自骨骼关键点的位置；图(b)中的人物身前有物体遮挡，

且手臂与身体存在不同程度的交叉重合的情况，模型同样能够检测出每个人骨骼关键点的位置；图(c)和图(d)通过对比可以看出当两个人靠的过近时，本文的算法仍然能准确识别出每个人的骨骼关键点位置，并成功将不同人体的姿态区分开。

可视化结果表明，本文提出的 FastPose-Lite 模型，即使在有遮挡或人物间有交叉重叠等因素干扰的情况下，仍然能够准确的检测出每个人骨骼关键点的位置，生成正确的人体姿态图，说明本文提出的网络模型的泛化能力和抗干扰能力不错。



图 8 可视化效果图

3.4.2 消融实验

本文选择在 COCO 数据集上对所提出的 FastPose-Lite 人体姿态估计网络的几个模块进行消融实验。定义方法 1 是原始网络模型 FastPose；方法 2 将 ResNet 网络替换为 GSE-ResNet 网络，即将 ResNet 中的 Bottleneck 模块替换成 GBNK 模块；方法 3 将 ResNet 网络替换成 MobileNetV3 网络；方法 4 是本文提出的包含 GBNK 模块以及 CBAM 模块的 FastPose-Lite 网络。

消融实验结果如表 2 所示，比较方法 1 与方法 2 的结果，不难看出，本文提出的 GSE-ResNet 特征提取网络，与 ResNet 特征提取网络相比，参数量减少了 59.9%，计算量降低了 49.2%，做到了对原模型特征提取网络的轻量化；比较方法 2 和方法 3 的实验结果，可以看出，本文提出的轻量型 GSE-ResNet 特征提取网

络同当下较为热门的轻量型特征提取网络 MobileNetV3 相比，虽然 GSE-ResNet 的参数量和计算量略低于 MobileNetV3，但检测准确率更高，mAP 值高出 MobileNetV3 网络 2.5 个百分点，符合轻量型网络的特点；对比方法 2 和方法 4，在加入了 CBAM 模块后，使得整个模型的 mAP 达到 70.7%，证实了在两个 DUC 模块之间加入 CBAM 注意力机制模块，可以达到增强网络模型对特征信息的处理能力，减少上采样过程带来的损失，在针对关键点预测的准确度提升方面，带来了积极的影响作用。

实验结果表明，本文提出的基于 GBNK 模块的 GSE-ResNet 特征提取网络在加入 CBAM 模块后对模型预测精度有一定提升的同时，实现了对原网络模型轻量化的目的。

表 2 消融实验结果

Method	GBNK	MobileNetV3	CBAM	#Params/10 ⁶	GFLOPs	mAP/%
FastPose（方法 1）				29.4	5.9	71.8
+GSE-ResNet（方法 2）	√			11.8	3.0	70.2
+MobileNetV3（方法 3）		√		11.3	2.6	68.5
FastPose-Lite（方法 4）	√		√	14.3	2.9	70.7

4 结论

本文以改进 FastPose 网络模型的参数量和计算量为目的，提出了 FastPose-Lite 轻量型人体姿态估计网络。在特征提取网络部分构建了基于 GSE-ResNet 的骨干网络，同时在 DUC 模块间引入了高效的 CBAM 注意力模块。在 COCO 数据集上进行实验证明，FastPose-Lite 网络能够以较少的参数量和较低的计算量达到与主流姿态估计算法相当的性能和预测准确率。今后将考虑如何将模型应用于实际场景中，进一步提升实时检测效率。

参考文献

[1] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C] // 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee, 2005, 1: 886-893.

[2] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International journal of computer vision, 2004,

60 (2): 91-110.

[3] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition [J]. Neural computation, 1989, 1 (4): 541-551.

[4] Wei S E, Ramakrishna V, Kanade T, et al. Convolutional pose machines [C] // Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2016: 4724-4732.

[5] Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation [C] // European conference on computer vision. Springer, Cham, 2016: 483-499.

[6] Fang H S, Xie S, Tai Y W, et al. RMPE: Regional multi-person pose estimation [C] // Proceedings of the IEEE international conference on computer vision. 2017: 2334-2343.

[7] Fang H S, Li J, Tang H, et al. AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.

[8] Cao Z, Simon T, Wei S E, et al. Realtime multi-person 2d pose estimation using part affinity fields [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7291-7299.

- [9] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [J]. arXiv preprint arXiv: 1704.04861, 2017.
- [10] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.
- [11] Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3 [C] // Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1314-1324.
- [12] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
- [13] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov: Optimal speed and accuracy of object detection [J]. arXiv preprint arXiv: 2004.10934, 2020.
- [14] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [15] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module [C] // Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [16] Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations [C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1580-1589.
- [17] Wang P, Chen P, Yuan Y, et al. Understanding convolution for semantic segmentation [C] // 2018 IEEE winter conference on applications of computer vision (WACV). Ieee, 2018: 1451-1460.
- [18] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context [C] // European conference on computer vision. Springer, Cham, 2014: 740-755.
- [19] Chen Y, Wang Z, Peng Y, et al. Cascaded pyramid network for multi-person pose estimation [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7103-7112.
- [20] Xiao B, Wu H, Wei Y. Simple baselines for human pose estimation and tracking [C] // Proceedings of the European conference on computer vision (ECCV). 2018: 466-481.
- [21] Sun K, Xiao B, Liu D, et al. Deep high-resolution representation learning for human pose estimation [C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 5693-5703.