光电技术及应用

DOI：10.16818/j.issn1001－5868.2022022301

基于偶然不确定性的单人姿态估计模型测试时增强算法

李 杰[1,2,3]，亓 波[1,2,3]，张建林[2,3]*

（1. 中国科学院光束控制重点实验室，成都 610209；2. 中国科学院光电技术研究所，成都 610209；
3. 中国科学院大学，北京 100049）

摘 要： 针对现有单人姿态估计网络结果缺乏可靠性评估和鲁棒性保障等问题，提出了一种基于偶然不确定性的测试时增强方法。该方法首先利用随机并行的数据增强和模型推理得到多样化输出，随后通过计算该输出的偶然不确定性得到其可靠性评估，最后根据可靠性将该输出及其不确定性进行加权融合以得到更准确鲁棒的结果及其评估。在 MPII 数据集上的实验表明，该算法可即插即用地应用于任意现有单人姿态估计网络，从而得到更精确鲁棒的结果及其不确定性评估。

关键词： 单人姿态估计；关键点检测；偶然不确定性；测试时增强

中图分类号： TP183 **文章编号：** 1001－5868(2022)05－0968－06

A Testing-Time-Augmentation Algorithm for Single Human Pose Estimation Based on Aleatoric Uncertainty

LI Jie[1,2,3], QI Bo[1,2,3], ZHANG Jianlin[2,3]

（1. Key Lab. of Optical Beam Control Engineering, Chinese Academy of Sciences, Chengdu 610209, CHN;
2. Institute of Optics and Electronics Technology, Chinese Academy of Sciences, Chengdu 610209, CHN;
3. University of Chinese Academy of Sciences, Beijing 100049, CHN）

Abstract: Aiming at the problem of lacking reliability evaluation and robustness guarantee in existing single-person pose estimation networks' results, a testing-time-augmentation (TTA) algorithm based on aleatoric uncertainty was proposed. In this TTA algorithm, diverse outputs were firstly obtained by stochastic parallel data augmentation and model inference. Then, the reliability evaluations of those outputs are acquired by calculating their aleatoric uncertainty. Finally, those outputs and their uncertainty were fused according to the reliabilities to obtain a more accurate and robust result as well as its evaluation. Experiments on the MPII dataset show that this algorithm can be applied to any existing single-person pose estimation network in a plug-and-play manner, leading to a more precise and robust result with its uncertainty evaluation.

Key words: single human pose estimation; key points detection; aleatoric uncertainty; testing-time-augmentation

0 引言

人体姿态估计即针对输入的人体图像进行关节点定位和识别，其中单人姿态估计是该领域的核心和基础。该任务的难点主要在于光照视角等场景变换，尺度不一、类别多样的姿态差异以及目标之间的遮挡干扰等问题，但随着大数据、GPU 计算和深度学习技术的引入，人体姿态估计得到了迅速发展[1]。

其中单人姿态估计模型的典型代表网络有 Hourglass[2]，SimpleBaseline[3] 和 HRNet[4] 等，但是由于深度学习往往基于大数据进行黑盒训练，缺乏可解释性和可靠性评估，这对其实际应用造成了

收稿日期：2022－02－23.

*通信作者：张建林 E-mail：jlin@ioe.ac.cn

较大的困扰，尤其对于面向实际部署的轻量化模型，如 Lite-HRNet[5]，MobileNet[6]，ShuffleNet[7] 等，更是存在精度和可靠性未知的问题。

为了解决这些问题，本文提出了一种基于偶然不确定性的单人姿态估计测试时增强算法，其流程可分为以下三个阶段：基于随机并行的数据增强和模型推理阶段、基于不确定性的可靠性评估阶段和基于可靠性加权的结果融合阶段。该算法通过随机并行的数据增强和模型推理得到模型的多样化输出，同时通过评估该输出的偶然不确定性，进而依据可靠性加权得到更优的输出和鲁棒性评估。

1　相关工作

1.1　单人姿态估计网络

单人姿态估计（或关键点检测）是人体姿态估计领域的基础和核心，该任务在相对可控的环境下可以将研究重点放在如何针对性地设计出优秀网络模型上。而在这之后，优秀的单人姿态估计网络稍加改进也可以延伸适用于多人姿态估计或者三维人体姿态估计。

常用单人姿态估计经典网络通常将人体姿态估计问题当作关键点检测问题，其通过神经网络输出不同种类的热力图分布来响应关键点位置。按模型大小分类，前者有 Hourglass[2]，SimpleBaseline[3] 和 HRNet[4] 等。而后者代表则有 Lite-HRNet[5] 和 EEffPose[8] 等。此外，通用模型如 ResNet[9]，MobileNet[6] 和 ShuffleNet[7] 等通过嫁接热力图输出头也可用于姿态估计。

1.2　数据增强

数据增强是指通过对输入图像进行一系列变换，如旋转、缩放、平移以及明暗色彩乃至对比度变换，以期扩充数据和引入强烈先验，从而使得网络得到更多样的输入或输出[10]。

该技术应用在网络训练阶段时，通常是被作为一种常用的数据预处理策略，以使得网络得到旋转、尺度或者视角不变性同时防止过拟合。但由于各领域的数据增强策略往往仅适用于和依赖于特定数据集和任务领域，为了得到一种自适应的数据增强策略，自动数据增强算法（Auto-Augment）被提出以自动迭代寻找最优的数据增强策略的超参数[11]。然而由于人体姿态估计领域人体目标较为固定，所以从 Hourglass[2] 开始就使用较统一沿用的数据增强策略，如随机水平翻转、随机缩放和旋转等。

当数据增强应用在测试阶段时，即为测试时增强（Test-Time-Augmentation，TTA），它是一种可以得到更准确鲁棒输出的模型集成策略，广泛应用在从图像识别到人体姿态估计等各种深度学习任务之中，例如图像识别任务中 Alexnet[12] 和 ResNet[9] 都应用了“十次随机裁剪”测试时增强策略，而在人体姿态估计领域，SimpleBaseline[3] 和 HRNet[4] 则应用了“水平翻转测试”和“六尺度测试”等测试时增强策略。

1.3　不确定性

作为机器学习中的重要概念，不确定性可分为偶然不确定性（Aleatoric Uncertainty）和认知不确定性（Epistemic Uncertainty），其中偶然不确定性来源于随机性，是由固有随机导致的实验结果的可变性，其往往不可消除；与此相反，认知不确定性是指由于缺乏知识或数据而导致的不确定性，同时这种由无知引起的不确定性原则上可以通过添加额外信息而消减[13]。在姿态估计领域中，人们对不确定性研究得较少。但在同为像素级任务的医疗图像分割领域，有工作[14] 为了得到准确结果和可靠性评估，提出基于偶然不确定性的测试时增强。

2　本文算法

为进一步得到更准确鲁棒的关键点检测结果及其不确定性评估，本文提出一种基于偶然不确定性的单人姿态估计测试时增强算法。该算法在充分利用 GPU 并行能力而不显著增加运行延迟的同时，可得到更精确鲁棒的关键点检测结果及其热力图不确定性输出。同时，作为一种测试时增强算法，该算法无需重新训练或超参数微调，可适应于任意已有单人姿态估计模型，即可做到真正无痛即插即用。

该算法作为一种全新测试时增强算法，具有以下创新：（1）与传统测试时增强相比，该算法中基于随机并行的数据增强和模型推理，不会延迟测试推理的速度；（2）与常见不确定性相比，该算法提出了一种独特的基于均方根误差的偶然不确定性，具有简洁质朴的合理性；（3）与普通均值融合相比，该算法依据可靠性进行加权融合，具有更强的可解释性。

依据以上创新点表述，该算法总体流程如图 1 所示，具体可分为三个阶段：基于随机并行的数据增强和模型推理阶段、基于不确定性的可靠性评估阶段和基于可靠性加权的结果融合阶段。各阶段的技术细节将在接下来的章节分别介绍。

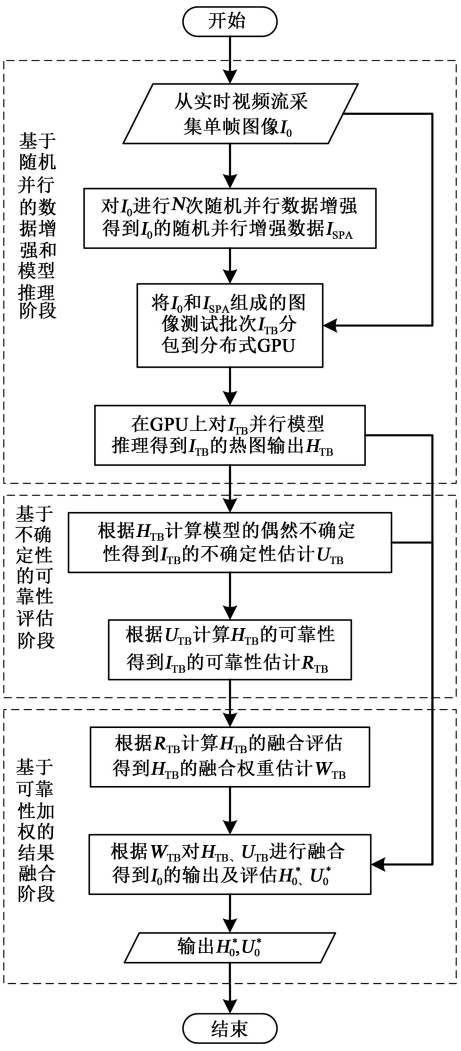


图 1 本文算法的总体流程图

2.1 基于随机并行的数据增强和模型推理阶段

在此阶段中,由于单人姿态估计测试流程中,每次只能从视频流中采集单帧图像,从而制约了系统充分利用 GPU 并行计算的能力。同时结合对更准确鲁棒热力图结果及其不确定性评估的需求,构建了该单帧图像的数据增强以对模型进行偶然不确定性估计。即本文提出的基于随机并行的数据增强和模型推理,以期随机并行地构建如图 1 所示的单帧图像 I_0 的测试批次 I_{TB}(Image of Test Batch)及其热力图输出 H_{TB}(Heatmap of Test Batch)。

如图 1 所示,该阶段随机并行分别体现在"随机数据增强"和"并行模型推理"两部分。具体地讲,如图 2 所示,首先对采集到的单帧图像 I_0 经过 N 次随机并行数据增强 T_n 得到 N 张随机并行增强图像 I_{SPA}(Image of Stochastic Parallel Augment),如式(1)的集合所示;之后再将 I_{SPA} 与 I_0 组成测试批次图像 I_{TB},如式(2)的集合合并所示;最后用多卡 GPU 进行分布式并行推理从而得到该测试批次热力图输出 H_{TB},如式(3)所示。

$$I_{SPA}=\{I_n \mid I_n=T_n(I_0),n\in[1,N]\} \quad (1)$$

$$I_{TB}=I_{SPA}\cup I_0 \quad (2)$$

$$H_{TB}=\{H_n \mid H_n=\mathrm{Net}(I_n),I_n\in I_{TB}\} \quad (3)$$

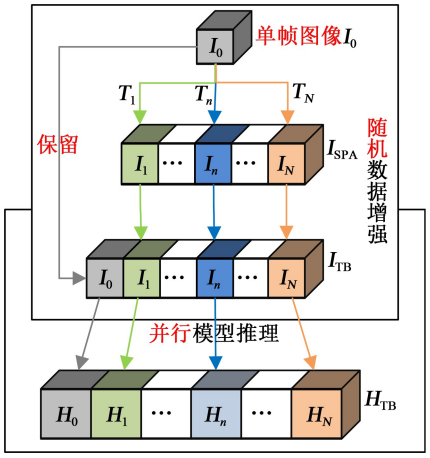


图 2 基于随机并行的数据增强和模型推理阶段示意图

该阶段核心在于式(1)中的 T 和 N。一般测试时,增强往往需要设计特定的数据增强策略和超参数,而本文的随机数据增强则直接沿用该模型训练时所采用的数据增强策略 T,从而无需针对模型进行特定的策略设计和寻优。因为经由 T 数据增强策略下训练得到的模型 Net 在测试时也应该能对 T 策略下的增强数据进行准确推理。这既符合该网络训练阶段拟合数据的概率空间,又降低针对每个模型具体参数的寻优操作;至于 N,其值越大算法效果越好越稳定,但过大时将增加计算成本。本文为了充分利用 4 核 FPGA/4 卡 GPU 分布式终端,令 $N=7$ 得到共 8 张增强数据,即每核/每卡各负责两张,从而保证不会显著增加推理延迟。

2.2 基于不确定性的可靠性评估阶段

由于深度学习模型参数繁杂,需要训练大量有标签的数据,而面对无标签的测试数据时则不能有效验证。因此,设计了如下基于不确定性的可靠性评估阶段,主要对深度学习"黑盒子"模型输出的热力图 H_{TB} 估计其偶然不确定性 U_{TB}(Uncertainty of Test Batch)和可靠性 R_{TB}(Reliability of Test Batch),进而在实际应用中提供可解释评估。

$$U_{TB}=\left\{U_i \mid U_i=\sqrt{\frac{1}{N+1}\sum_{j=0}^{N}(H_i-H_j)^2},\ H_i \text{ 和 } H_j\in H_{TB}\right\} \quad (4)$$

$$R_{\mathrm{TB}}=\left\{R_i \mid R_i=-U_i+\sum_{j=0}^{N}U_j, U_i \text{ 和 } U_j \in U_{\mathrm{TB}}\right\} \tag{5}$$

如图 1 所示，该阶段可分为“偶然不确定性评估”和“可靠性估计”两部分。具体地讲，如图 3 所示，先由上阶段输出 H_{TB} 定义并计算的不确定性估计 U_{TB}，进而利用 U_{TB} 定义和计算其可靠性 R_{TB}。其中，H_{TB} 如式(3)所示是一个具有 $N+1$ 张关键点响应热力图的集合。而偶然不确定性 U_{TB} 则由式(4)所定义，其中 H_i 和 H_j 是来自于集合 H_{TB} 的元素，即某热力图 H_i 的偶然不确定性 U_i 为假设以其自身 H_i 为真实标签时全体热力图 H_{TB} 所对应的均方根误差。该值反映了该输出热力图 H_i 在其他热力图验证下的偶然不确定性。为将不确定性取反，本文定义可靠性集合如式(5)所示。其中，U_i 和 U_j 都是来自于不确定性集合 U_{TB} 的元素，某不确定性 U_i 对应的可靠性 R_i 定义为除自身外所有不确定性的总和。值得说明的是，这种可靠性无任何实际物理意义。

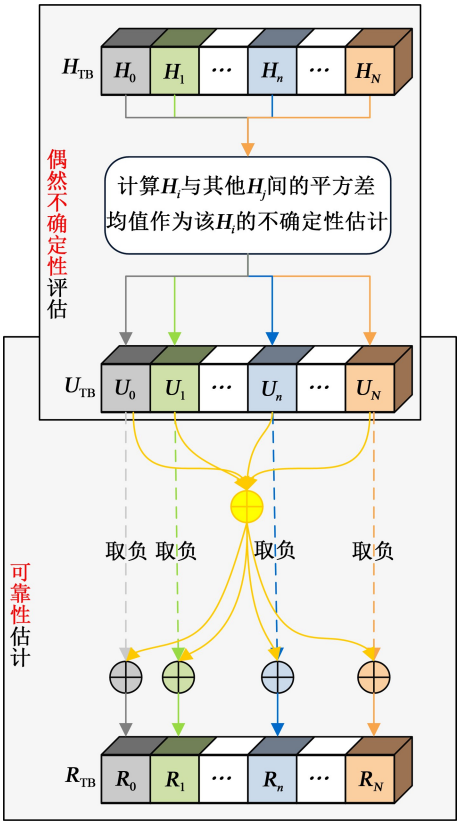


图 3　基于不确定性的可靠性评估阶段示意图

2.3　基于可靠性加权的结果融合阶段

为了得到更准确鲁棒的结果及其可解释性评估，本文设计了此基于可靠性加权的结果融合阶段。该阶段将利用并行批次输出的可靠性估计 R_{TB} 进行权重评估以得到相应的归一化权重 W_{TB}(Weight of Test Batch)，从而加权融合其结果 H_{TB} 及其不确定性 U_{TB} 和可靠性 R_{TB}。

$$W=\left\{W_i=\frac{\exp(R_i)}{\sum_j \exp(R_j)}, R_i \text{ 和 } R_j \in R_{\mathrm{TB}}\right\} \tag{6}$$

$$H^*=\sum_i W_i \times H_i, \text{有 } W_i \in W_{\mathrm{TB}}, H_i \in H_{\mathrm{TB}} \tag{7}$$

$$U^*=\sum_i W_i \times U_i, \text{有 } W_i \in W_{\mathrm{TB}}, U_i \in U_{\mathrm{TB}} \tag{8}$$

如图 1 所示该阶段分为“融合权重估计”和“加权融合输出与评估”两部分。具体如图 4 所示：对可靠性评估 R_{TB} 进行 Softmax 归一化计算得到该测试批次的融合权重 W_{TB}，具体计算细节如式(6)所示；而后利用该权重对该测试批次输出 H_{TB} 和评估 U_{TB} 分别进行扩展元素相乘的加权融合，从而得到原始单帧图像 I_0 的最终热力图输出 H_0^* 和不确定性估计 U_0^*，具体如式(7)和(8)所示。这一阶段最终得到了原始输入的更为准确鲁棒的输出，又得到了具有可解释性的不确定性估计输出。显而易见且可证明，经过可靠性加权的不确定性 U_0^* 会小于 U_{TB} 的元素。

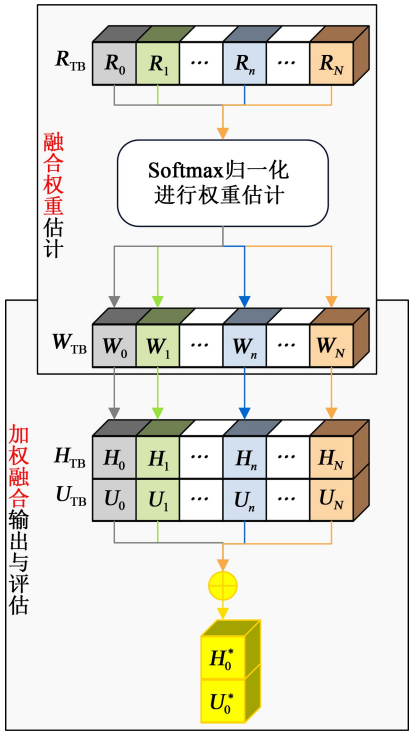


图 4　基于可靠性加权的结果融合阶段示意图

3 实验结果与分析

3.1 实验环境

虽然本文所提算法无需额外训练，但仍需依靠GPU进行并行模型推理。所以将该算法应用在4核K7325T FPGA和4卡NVIDIA Jetson TX2的分布式终端推理平台上，所使用的框架为PyTorch和MMPose开源系统。为控制变量，所有对照结果均来源于该开源库，因此可能与源论文结果有微小出入。

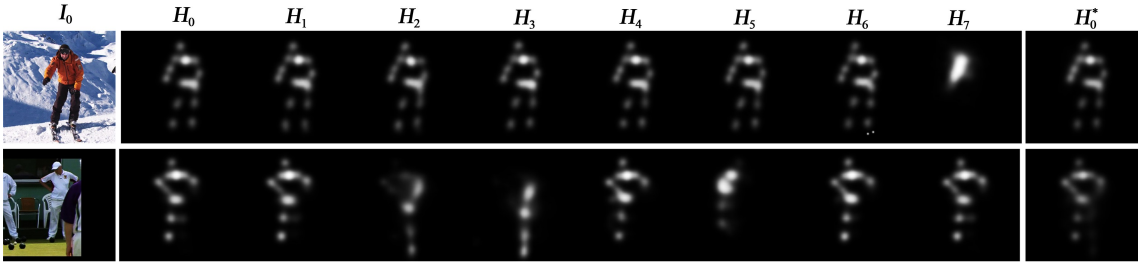
3.2 数据集和指标介绍

本文选取MPII单人数据集作为实验分析基准，其是最常用的人体姿态估计权威基准之一，尤其是在二维单人姿态估计领域。该数据集有专门整理筛选的单人姿态估计的RGB图像及其16个关键点位置类别标注，同时该图像人体姿态丰富多样且尽可能保证在图像中心，非常适合于调试和检验单人姿态估计模型或算法。该单人数据集分别有约28,000/11,000张人体姿态训练集和验证集实例，且包含了不同的光照、视角、姿态和遮挡场景。

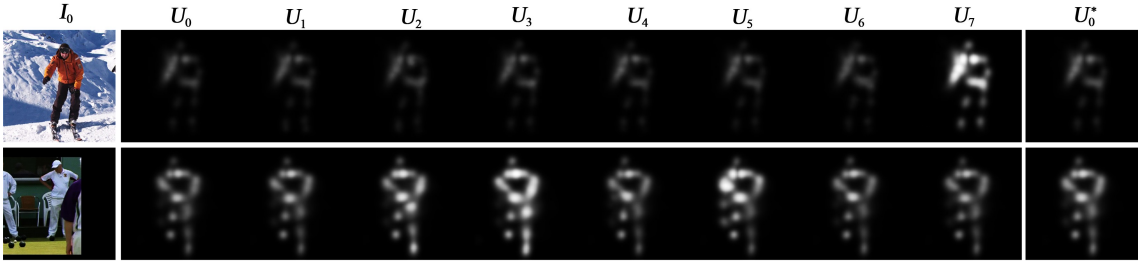
该数据集常用指标为PCK，即正确分类关键点的比例(Percentage of Correct Keypoints)。该指标通过参照长度和设定阈值来判断关键点是否检测正确，如常用的$PCK_h@0.5$和$PCK_h@0.1$则是以头部长度h作为参照长度，同时分别设定阈值α为0.5和0.1，即当预测关键点与标签位置距离小于α_h时即为正确。

3.3 结果可视化分析

图5展示了部分样例的热力图输出及其不确定性评估。其中I_0为输入图像，$H_0 \sim H_7/U_0 \sim U_7$为测试增强的热力图及其不确定输出，而$H_0^*/U_0^*$为最终输出及评估。可以看到，最终热力图输出能够更准确抗干扰，且其不确定性更低，从而达到了本算法的预期效果。



(a) 热力图输出结果示例



(b) 不确定性输出结果示例

图5 部分样例的热力图输出及其不确定性评估

3.4 结果对比

为了证明本文算法的有效性，同时展示其即插即用能力，本文在MPII单人数据集上对各种经典单人姿态估计模型进行了实验。为了消除随机干扰，各进行8次随机实验并报告了其指标统计量的均值和标准差。

由表1所示的结果可知：对于大模型，本算法的$PCK_h@0.5/PCK_h@0.1$提升了0.4～0.7/0.9～1.6；而对于小模型，本算法则的$PCK_h@0.5$和$PCK_h@0.1$提升了0.4～1.3/0.8～2.2；同时各指标的较小标准差(括号内)显示本算法足够有效且受随机影响较小。利用分布式并行计算，观察到本算法推理延迟没有显著增加。

综上所示，在MPII数据集上，本文算法可即插即用地适用于各种网络模型，即在不降低模型推理速度的同时，单人姿态估计模型都可以得到性能增益，同时输出如图5所示的之前所不具备的不确定性评估。

表 1　本文算法应用于各模型前后在 MPII 数据集的结果对比

模型类别	模型名称	模型大小	输入大小	原始结果		测试增强后结果	
				PCK_h@0.5	PCK_h@0.1	PCK_h@0.5	PCK_h@0.1
小模型	LiteHRNet-18	1.10 M	256×256	85.902	25.993	86.326(±0.045)	27.032(±0.073)
	LiteHRNet-30	1.76 M	256×256	86.916	27.067	87.310(±0.051)	27.902(±0.075)
	ShuffleNetv1	6.94 M	256×256	82.345	19.477	83.437(±0.063)	21.720(±0.129)
	ShuffleNetv2	7.55 M	256×256	82.766	20.549	84.104(±0.054)	22.550(±0.102)
	MobileNetv2	9.57 M	256×256	85.436	23.456	86.351(±0.064)	25.377(±0.126)
大模型	HRNet_w32	28.54 M	256×256	90.016	33.417	90.590(±0.052)	34.391(±0.042)
	HRNet_w48	63.60 M	256×256	90.062	33.739	90.640(±0.047)	34.728(±0.077)
	ResNet_50	34.00 M	256×256	88.225	28.569	88.909(±0.049)	30.200(±0.084)
	ResNet_101	52.99 M	256×256	88.754	28.977	89.268(±0.041)	30.518(±0.129)
	ResNet_152	68.64 M	256×256	88.925	30.273	89.598(±0.028)	31.791(±0.086)
	Hourglass_52	94.85 M	256×256	88.873	31.673	89.353(±0.043)	32.659(±0.055)
	Hourglass_52	94.85 M	384×384	89.391	36.641	89.802(±0.050)	37.498(±0.057)

4　总结

本文提出了一种基于偶然不确定性的单人姿态估计测试时增强算法，该算法通过随机并行的数据增强和模型推理得到多样化结果输出，而后通过定义输出之间的偶然不确定性从而评估其可靠性，接着通过可靠性加权融合得到最后带有不确定性和可靠性估计的热力图输出，最后对关键点热力图进行解码得到关键点坐标信息。MPII 上实验表明，该算法可即插即用地适用于任意单人姿态估计模型，以得到准确鲁棒的结果及其不确定性评估。

参考文献：

[1] Chen Y, Tian Y, He M. Monocular human pose estimation: A survey of deep learning-based methods[J]. Computer Vision and Image Understanding, 2020, 192: 102897.

[2] Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation[C]// Proc. of European Conf. on Computer Vision (ECCV), 2016: 483-499.

[3] Xiao B, Wu H, Wei Y. Simple baselines for human pose estimation and tracking[C]// Proc. of European Conf. on Computer Vision (ECCV), 2018: 466-481.

[4] Sun K, Xiao B, Liu D, et al. Deep high-resolution representation learning for human pose estimation[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2019: 5693-5703.

[5] Yu C, Xiao B, Gao C, et al. Lite-HRNet: A lightweight high-resolution network[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2021: 10440-10450.

[6] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]// Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2018: 4510-4520.

[7] Ma N, Zhang X, Zheng H T, et al. ShuffleNet V2: Practical guidelines for efficient CNN architecture design[C]// Proc. of the European Conf. on Computer Vision (ECCV), 2018: 116-131.

[8] Li J, Wang Z, Qi B, et al. Meme: A mutually enhanced modeling method for efficient and effective human pose estimation[J]. Sensors, 2022, 22(2): 632.

[9] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2016: 770-778.

[10] Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning[J]. J. of Big Data, 2019, 6(1): 1-48.

[11] Cubuk E D, Zoph B, Mane D, et al. AutoAugment: Learning augmentation strategies from data[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2019: 113-123.

[12] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Adv. in Neural Information Processing Systems, 2012, 25(2): 84-90.

[13] HüLlermeier E, Waegeman W. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods[J]. Machine Learning, 2021, 110(3): 457-506.

[14] Wang G, Li W, Aertsen M, et al. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks[J]. Neurocomputing, 2019, 338: 34-45.

作者简介：

李　杰(1995—)，男，河北省邯郸市人，博士研究生，主要研究方向为人体姿态估计/关键点检测、目标检测与跟踪；

张建林(1976—)，男，四川省乐山市人，博士，研究员，主要研究方向为计算机视觉、深度学习、图像分析与理解技术。