International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 323-328 www.ijaem.net ISSN: 2395-5252

Pose Estimation on Humans

[1]Puneeth P, [2]Bharath V,[3] Arjun M,[4] Aishwarya R shetty
[1] Assistant Professor,[2 3 4] Engineering Students,Department of Information Science and Engineering
Maharaja institute of Technology Mysore,India

Submitted: 01-08-2021 Revised: 07-08-2021 Accepted: 10-08-2021

ABSTRACT: Single-person human pose estimation facilitates markerless movement analysis in sports, as well as in clinical applications. Still, state-of-the-art models for human pose estimation generally do not meet the requirements of real-life applications. The proliferation of deep learning techniques has resulted in the development of many advanced approaches. However, with the progresses in the field, more complex and inefficient models have also been introduced, which have caused tremendous increases in computational demands. To cope with these complexity and inefficiency challenges, we propose a novel convolutional neural network architecture,called EfficientPose, which exploits recently proposed EfficientNets in order to deliver efficient and scalable single-person pose estimation. EfficientPose is a family of models harnessing an effective multi-scale feature extractor and computationally efficient detection blocks using mobile inverted bottleneck convolutions, while at the same time ensuring that the precision of the poseconfigurations is still improved. Due to its low complexity and efficiency, EfficientPose enables real-world applications on edge devices by limiting the memory footprint and computational cost. The results from our experiments, using the challenging MPII single-person benchmark, show that the proposed EfficientPose models substantially outperform the widely-used OpenPose model both in terms of accuracy and computational efficiency. In particular, our top-performing model achieves state-of-the-art accuracy on single-person MPII, with low-complexity ConvNets.

KEYWORDS: Human pose estimation · Effeicient pose · Model scalability · High precision · Computational efficiency · Openly available

I. INTRODUCTION

Single-person human pose estimation (HPE) refers to the computer vision task of localizing human skeletal keypoints of a person from an image or video frames. Single person HPE has many real-world applications,ranging from outdoor activity recognition and computer animation to clinical assessments of motor repertoire and skill practice among professional athletes.The proliferation of deep convolutional neural networks (ConvNets) has advanced HPE and further widen its application areas. ConvNet based HPE with its increasingly complex network structures,combined with transfer learning, is a very challenging task. However, the availability of high performing ImageNet backbones,together with large tailor-made datasets, such as MPII for 2D pose estimation,has facilitated the development of new improved methods to address the challenge.

The OpenPose network [6] (OpenPose for short) hasbeen one of the most applied HPE methods in real-world applications. It is also the first open-source real-time system for HPE. OpenPose was originally devel-oped for multi-person HPE, but has in recent years beenfrequently applied to various single-person applicationswithin clinical research and sport sciences [15, 32, 34].The main drawback with OpenPose is that the levelof detail in keypoint estimates is limited due to itslow-resolution outputs. This makes OpenPose less suit-able for precision-demanding applications, such as elitesports and medical assessments, which all depend onhigh degree of precision in the assessment of movementkinematics. Moreover, by spending 160 billion floating-point operations (GFLOPs) per inference, OpenPoseis considered highly inefficient.

In this paper, it stresses the lack of publicly available methods for single person HPE that are both computationally efficient and effective in terms of estimation precision. To this end,we exploit recent advances in ConvNets and propose an improved approach called EfficientPose. Our main idea is to modify OpenPose into a family of scalable ConvNets for high-precision and computationally efficient single-person pose estimation from 2D images. To assess the performance of our approach, this performs two separate comparative studies.First,we evaluate the Efficient Pose model by comparing it against the original OpenPose model on single-person HPE. Second, we compare it against the current state-of-the-art single-person HPE methods on the official MPII challenge, focusing on accuracy as a function of the number of parameters. The proposed Efficient-Pose models aim to elicit high computational efficiency, while bridging the gap in availability of high-precision HPE networks.

II. LITERATURE SURVEY

We find that A new beauty of strategies is proposed to motive at 2D-to-3-D picture Reconstruction that is primarily based in reality mostly on the extensively tremendous method of analyzing from examples. One method that is proposed is primarily based absolutely totally on reading a thing mapping from nearby photo attributes to scene-intensity. The specific approach is based totally on globally estimating the complete depth state of affairs of a query immediately from a repository of Depth+Position pairs the use of nearest neighbor based totally regression.It objectively validates the polygon mesh common trendy customary performance in opposition to modern day algorithms.While the nearby technique changed into outperformed through manner of specific algorithms, it's far quite rapid as it's miles,basically,based completely clearly genuinely on desk look up.However, the world wide method completed better than the contemporary day algorithms in phrases of cumulative regular common normal performance all through datasets and finding out strategies,and has finished so at a Position. Anaglyph pictures produced through the usage of the algorithms bring about a relaxed 3-D revel in how ever are not clearly void of distortions. Clearly, there may be room for improvement within the future. With the constantlygrowing quantity of 3-d facts online and with the swiftly developing computing in thecloud, the proposed framework seems a promising alternative to operator-assisted 2D-to-3Dimage.

A study is conducted on processing of images based on the radically different approachof learning from examples. method proposed is based on learning a point mapping fromlocal image attributes to scene-depth. The other method is based on reconstructionestimating the entire depth field of a query directly from a repository of image + depthpairs using nearest neighbor-based regression. In some papers they have objectivelyvalidated their algorithms' performance against state-of-the-art algorithms. While thelocalmethod was outperformed by otheralgorithms.

III. EXPERIMENTATION.(

Figure 1 and Figure 2 depict the architectures of Open-Pose and EfficientPose, respectively. As can be observed in these two figures, although being based on OpenPose,the EfficientPose architecture is different from the OpenPose architecture in several aspects, including 1)both high and low-resolution input images, 2) scalableEfficientNet backbones, 3) cross-resolution features, 4)and 5) scalable Mobile DenseNet detection blocks infewer detection passes, and 6) bilinear upscaling. For amorethorough ImageNet (step 2a and 2b in Figure 2). High-level semantic information is obtained from the high-resolution image using the initial three blocks of a high-scale EfficientNet with $\varphi \in [2, 7]$ (see Equation 1), outputting Cfeature maps (2a in Figure 2). Low-level local information is extracted from the low-resolution image by thefirst two blocks of a lower-scale EfficientNet-backbone(2b in Figure 2) in the range $\varphi \in [0, 3]$. Table 1 provides an overview of the composition of Efficient Netbackbones, from low-scale B0 to high-scale B7.The first block of EfficientNets utilizes the MBConvs shown inFigure 3a and 3b, whereas the second and third blocks comprise the MB Conv layers in Figure 3cand3d.

The features generated by the low-level and high-level EfficientNet backbones are concatenated to yieldcross-resolution features (step 3 in Figure 2). This en-ables the EfficientPose architecture to selectively emphasize important local factors from the image of interest and the overall structures that guide high-quality pose estimation. In this way, we enable an alternative simultaneous handling of different features at multiple abstraction levels.

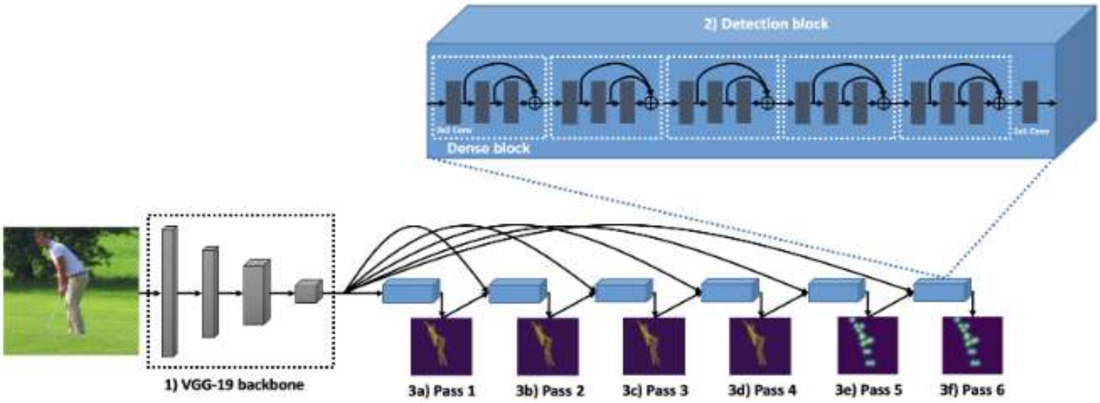

Fig. 1 OpenPose architecture utilizing 1) VGG-19 feature extractor, and 2) 4+2 passes of detection blocks performing 4+2 passes of estimating part affinity fields (3a-d) and confidence maps (3e and 3f)

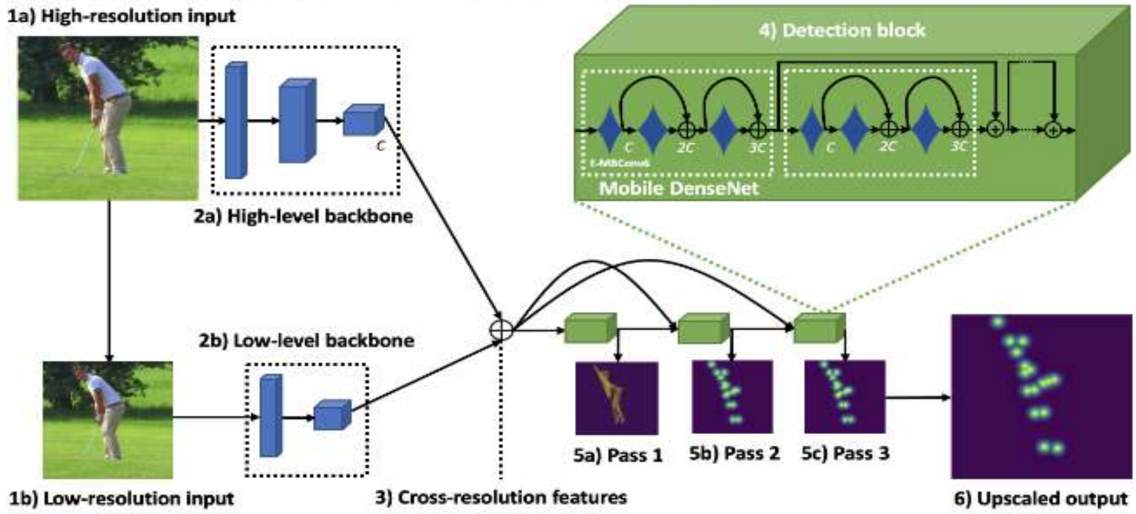

Fig. 2 Proposed architecture comprising 1a) high-resolution and 1b) low-resolution inputs, 2a) high-level and 2b) low-level EfficientNet backbones combined into 3) cross-resolution features, 4) Mobile DenseNet detection blocks, 1+2 passes for estimation of part affinity fields (5a) and confidence maps (5b and 5c), and 6) bilinear upscaling

Block	B0	B1	B2	B3	B4	B5	B7
1	Conv(3×3,32,2) BN Swish			Conv(3×3,40,2) BN Swish	Conv(3×3,48,2) BN Swish		Conv(3×3,64,2) BN Swish
	MBConv1 (3×3,16,1)			MBConv1 (3×3,24,1)			MBConv1 (3×3,32,1)
	–	MBConv1* (3×3,16,1)		MBConv1* (3×3,24,1)		MBConv1* (3×3,24,1) ×2	MBConv1* (3×3,32,1) ×3
2	MBConv6 (3×3,24,2)			MBConv6 (3×3,32,2)		MBConv6 (3×3,40,2)	MBConv6 (3×3,48,2)
	MBConv6* (3×3,24,1)	MBConv6* (3×3,24,1) ×2		MBConv6* (3×3,32,1) ×2	MBConv6* (3×3,32,1) ×3	MBConv6* (3×3,40,1) ×4	MBConv6* (3×3,48,1) ×6

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 8 Aug 2021, pp: 323-328 www.ijaem.net ISSN: 2395-5252

3	MBConv6 (5×5,40,2)		MBConv6 (5×5,48,2)		MBConv6 (5×5,56,2)	MBConv6 (5×5,64,2)	MBConv6 (5×5,80,2)
	MBConv6* (5×5,40,1)	MBConv6* (5×5,40,1) ×2	MBConv6* (5×5,48,1) ×2		MBConv6 * (5×5,56,1) ×3	MBConv6* (5×5,64,1) ×4	MBConv6* (5×5,80,1) ×6
I	224×224	240×240	260×260	300×300	380×380	456×456	600×600
C	40		48		56	64	80
αφ	1.2^0=1.0	1.2^1=1.2	1.2^2=1.4	1.2^3=1.7	1.2^4=2.1	1.2^5=2.5	1.2^7=3.6

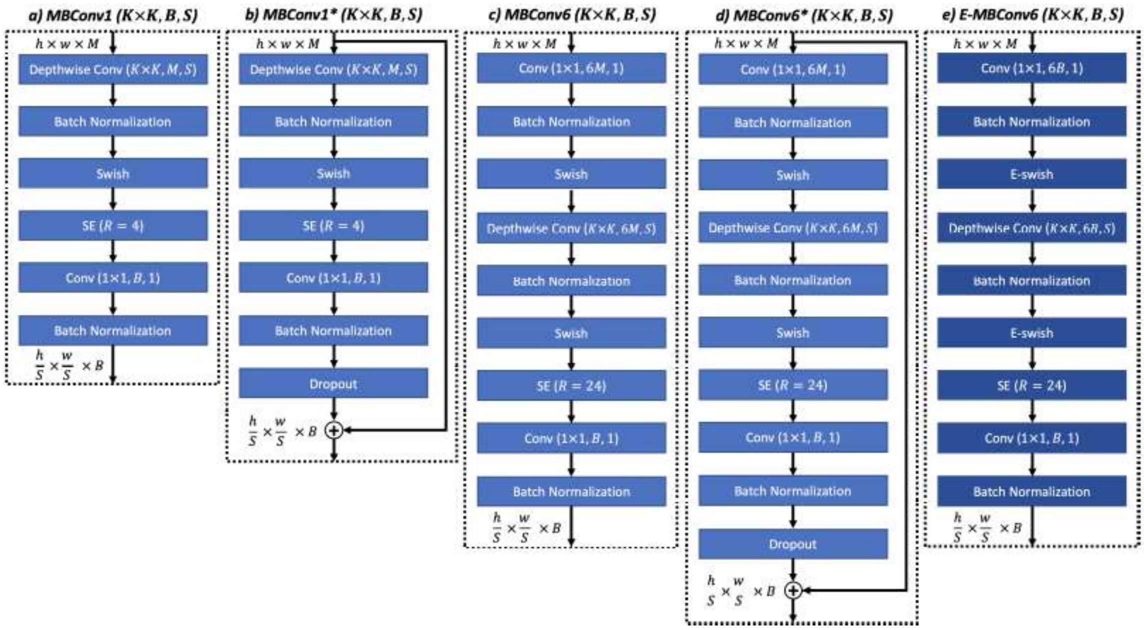

Fig. 3 The composition of MBConvs. From left: a-d) MBConv(K K, B, S) in EfficientNets performsdepthwiseconvolutionwithfiltersize KKandstrideS,andoutputsBfeaturemaps.M BConv*(bandd)extendsregularMBConvsb yincludingdropoutlayerandskipconnection.e)E-MBConv6(KK, B, S)in Mobile DenseNets adjusts MBConv6 with E-swishactivation and number of feature maps in expansion phase as 6B. All MBConvs take as input M feature maps with spatialheightandwidthofhandw,respectively .RisthereductionratioofSE

Fromtheextractedfeatures,thedesired key pointsarelocalizedthroughaniterativedetectionproces s,whereeachdetectionpassperformssupervisedpredictionofoutput maps. Each detection pass comprises a detectionblockandasingle1 × 1convolutionforoutputprediction.Thedetectionblocksacrossalldetection passeselicitthesamebasicarchitecture,comprising MobileDenseNets(seestep4inFigure2).DatafromMobileDenseNetsareforwardedtosubsequentlayersof thedetectionblockusing residualconnections. The MobileDenseNetisinspiredbyDenseNetssupportingreuseoffeatures,avoidingredundantlayers,andMBConvwithSE,thusenablinglowmemoryfootprint.Inouradaptationof the MBConv operation (E-MBConv6(K × K, B, S) inFigure3e),weconsistentlyutilizethehighestperformingcombinationfrom, i.e., a kernelsize(K×K)of5×5andanexpansionratioof6. Wealsoavoiddown sampling(i.e.,S=1)andscalethewidthofMobileDenseNetsbyoutputtingnumberofchannelsrelativetothe high level backbone (B = C).WemodifytheoriginalMBConv6operationbyincorporatingE-swishasactivationfunctionwithβvalueof 1.25 . This has a tendency to accelerate progressionduring training compared to the regular Swish activation. We also adjust the first 1 × 1 convolution togenerate a number of feature maps relative to the out-put feature maps B rather than the input channels M. This reduces the memory consumption and computational latency since B ≤ M , with C ≤ M≤ 3C. Witheach Mobile DenseNet consisting of three consecutiveE-MBConv6 operations, the module outputs 3C featuremaps.

EfficientPoseperformsdetectionintworoundsFirst ,theoverallposeofthepersonis anticipated through a single pass of skeleton estimation. This aims to facilitate the detection of feasibleposes and to avoid confusion in case of several personsbeing present in an image. Skeleton estimation is per-formed utilizing part affinity fields as proposed in [7].Followingskeletonestimation,twodetectionpassesareperformed to estimate heat maps for key points of interest.Theformeroftheseactsasacoarsedetector(5bin Figure 2), whereas the latter (5c in Figure 2) refineslocalizationtoyieldmoreaccurateoutputs. Note that in OpenPose, the heatmaps of the finaldetection pass are constrained to a low spatial resolution,whichareincapableofachievingtheamountofdetails that are normally inherent in the high-resolutioninput. Toimprovethislimitationof Open Pose,a series of three transposed convolutions performing bilinear up sampling are added for 8× up scaling of thelow-resolution heat maps (step 6 in Figure 1). Thus, weprojectthelow-resolutionoutputontoaspaceofhigherresolution in order to allow an increased level of detail.To achieve the proper level of interpolation while operating efficiently, each transposed convolution increasesthemapsizebyafactorof2,usingastride of kurnel.

IV. CONCLUSION

This project provides an efficient way for pose estimation on humans. There is main characteristicsofclassificationarespeedandaccuracy. Hencethereisworkingondevelopment of automatic, efficient, fast and accurate system which is used for different pose estimation on humans. Work can be extended for development of hybrid algorithms & neural networks in order to increase the recognition rate of final classification process. Further needed to compute number of poses on humans andobjects.

REFERENCES

[1]. Wei-ZhiNie,Min-JieRen,An-AnLiu*,ZhendongMao,JieNie*"M-GCN:Multi- Branch Graph Convolution Network for 2d Image-Based On 3d Model Retrieval" Journal Of Latex Class Files, Vol. 14, No. 8, August 2015.

[2]. Song Dai, Xiang Bai, Senior Member, IEEE, Zhichao Zhou, Xhaoxiang Zhang, Senior Member, IEEE, Qi Tian and Longin Jan Latecki, "GIFT: Towords Scalable 3d Shape Retrieval" IEEE Transaction on Multimedia, Vol. 19, No. 6, June2017.

[3]. Pengfei Xu, Hongbo Fu, Youyi Zheng, Karan Singh, Hui Huang, Chiew-Lan Tai, "Model-Guided 3d Sketching" IEEE Transactions on Visualization and Computer Graphics.

[4]. Shubham Tulsiani, Tinghui Zhou, Alexei A. Efros, Jitendra Malik University of California, Berkeley, "Multi-view Supervision for Single-view Reconstruction via Differenciable Ray Consistency"Available:https://shubhtuls.github.io/drc/

[5]. M Kavitha* and E Kannan, "2D to 3D Conversion Using Key Frame Extraction" Indian Journal of Science and Technology, Vol. 9(28), July 2016.

[6]. LiJiang,ShaoShuaiShi,XiaojuanQiandJiayaJia,"GAL:GeometricAdversarial Loss for Single-View 3d-Object Reconstruction" in ECCV 2018 Available: https://link.springer.com/conference/eccv

[7]. Zhenzhong Kuang, Jun Yu, Jianping Fan,

Min Tan, "Deep Point Convolutional Approach for 3D Model Retrieval" in 978-1-5386-1737-3/18© 2018IEEE.

[8]. CaitlinT.Yeo,MD,*AndrewMacDonald,MD, †TamasUngi,MD,PhD,†Andras Lasso, PhD,Diederick Jalink, MD, FRCSC, * Boris Zevin, MD, PhD, FRCSC, * Gabor Fichtinger, PhD,† and Sulaiman Nanji, MD, PhD, FRCSC "Utility of 3D Reconstruction of 2D Liver Computed Tomography/ Magnetic Resonance Images As a Surgical Planning Tool for Residents in Liver Resection Surgery" Journal of Surgical Education & 2017 Association of Program Directors inSurgery.

[9]. TakahikoFuruya1,RyutarouOhbuchi1"Deep AggregationofLocal3DGeometric Features for 3D Model Retrieval" JSPS Grant-in-Aid for Young Scientists (B) #16K16055 and JSPS Grants-in-Aid for Scientific Research (C)#26330133.