Infant-Pose: Apply OpenPose and Infant Keypoints Dataset to Evaluate Infant Posture*

Xu Cao and Zening Chen

New York University
{xc2057,zc2256}@nyu.edu

Abstract. In this report, we will briefly discuss fundamental ideas from Openpose and try to extend their model and related methods into automatic rehabilitation evaluation research. The discussion section will cover some of the ideas for deploy our system in the future. Our implementation is available at https://github.com/IrohXu/Infant-Pose-pytorch

Keywords: Infant Posture · Openpose · Key-point.

1 Introduction

OpenPose is a Real-time multi-person 2D pose estimation based on deep learning, which is a key component in enabling machines to have an understanding of people in images and videos. It placed first in the COCO 2016 keypoints challenge.

In this report, we will explore the application of Openpose on clinical medicine: automatically detecting infant posture for General Movements Assessment (GMs). This is a very significant research topic in children's rehabilitation medicine. GMs is a video evaluation method by using Gestalt vision to observe and interpret the infant's overall movement performance from a short video[4]. The evaluation results of GMs have been proven to have a strong correlation with Cerebral Palsy[3]. Although GMs are widely used in clinical practice, the evaluation process of GMs is very cumbersome and needs the participation of professionals.

In order to explore whether deep learning can be used to build automated GMs models, we have made some preliminary attempts. In this paper, we try to extract the posture of the baby in the video through deep learning models, thereby reducing the dimension of the input data and at the same time protecting the privacy of the patient. We labeled a key-points dataset for infants and finetune pre-trained Openpose model on this new labeled infant key-point dataset.

The next section will briefly introduce the main skeleton of Openpose. Section 3 will cover the introduction for COCO dataset and our infant key-points dataset. Then, Section 4 presents and discusses the experimental result of our research. In section 5 and the final section, we will discuss the usage of infant keypoints in the future and summarize the whole report.

* Supported by Shenzhen Children's' Hospital and Shenzhen Baoan Maternal and Child Health Hospital. Code is available at https://github.com/IrohXu/Infant-Pose-pytorch

2 Openpose

Openpose is one of the most efficient bottom-up methods for multi-person pose estimation, which is proposed by Cao et al.[2] It firstly apply Part Affinity Fields (PAFs), a set of vectors that encode the location and orientation of limbs for human pose analysis. In this section, we will introduce and discuss the main ideas of Openpose and also introduce the confidence maps and PAFs that used in Openpose.

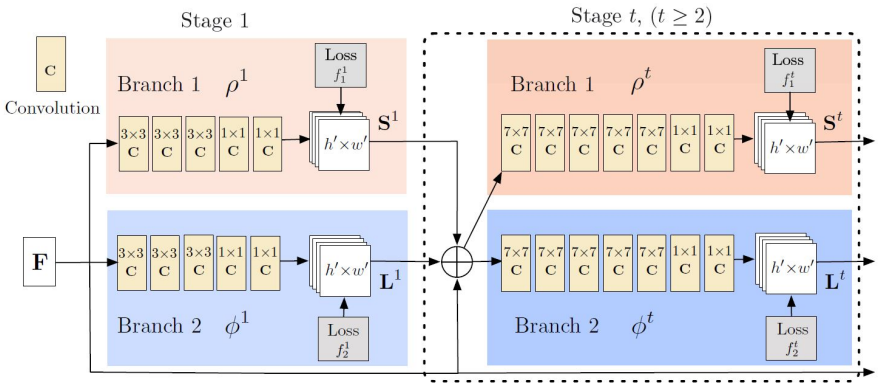


Fig. 1: Openpose

Fig.1 illustrates the overall framework of Openpose. The input of Openpose is an image of size w × h, while the output is the locations of corresponding key-points for each person of the input image. Firstly, the input image will pass the first 10 layers of VGG-19 and obtain extracted feature map F for further analysis. Then, Openpose applies a two-branch multi-stage CNN to separately compute and predict confidence maps S^t and PAFs L^t. After each stage, the predicted feature maps of the two branches will be concatenated with VGG-19 feature map F.

As figure 1 shows, for each stage of Openpose, there are two branches. We can use two functions to define the operation for each branch:

$$S^t = \begin{cases} \rho^1(F) & t = 1 \\ \rho^t(F, S^{t-1}, L^{t-1}) & t >= 2 \end{cases} \quad L^t = \begin{cases} \phi^1(F) & t = 1 \\ \phi^t(F, S^{t-1}, L^{t-1}) & t >= 2 \end{cases} \tag{1}$$

,where ρ^t and ϕ^t are the CNNs for the inference at stage t. $S^t \in R^{w \times h}$ and $L^t \in R^{w \times h \times 2}$ are detection confidence maps and part affinity fields for stage t separately. During the training, we are using L_2 loss between the each stages' estimated predictions and the ground truth confidence maps and PAFs. The intermediate supervision at each stage addresses the vanishing gradient problem by replenishing the gradient periodically[6]. The overall loss function is the combination of each L_2 loss for confidence maps and PAFs at different stages.

2.1 Confidence Maps for Part Detection

Confidence map is a 2D representation of the belief that a particular body part occurs at each pixel location. We can generate the ground truth confidence map from the labeled key-points location easily. For our self-labeled dataset, each input image only contains one infant, we can generate the confidence map S_j^* below:

$$S_j^*(p) = exp(-\frac{\|p - x_j\|_2^2}{\sigma^2}) \tag{2}$$

,where σ controls the spread of the peak. Since there is only one person in our Infant-Pose dataset, we do not need to aggregate the confidence values via a max operator applied by Openpose.

2.2 Part Affinity Fields for Part Association

Part Affinity Fields is a component that can preserve both location and orientation information across the region of support of limbs for keypoints detection. The part affinity is a 2D vector field for each limb. Let $x_{j_1,k}$ and $x_{j_2,k}$ be the groundtruth positions of body parts j_1 and j_2 from the limb c for person k in the image. If a points p lies on the limb, the value at $L_{c,k}^*(p)$ is a unit vector that points from j_1 to j_2. For all other points, the vector is zero-valued. The below function is the definition of the ground truth part affinity vector field $L_{c,k}^*$ at image point p.

$$L_{c,k}^*(p) = \begin{cases} v \;\; if \;\; p \; on \; limb \; c,k \\ 0 \;\; otherwise \end{cases} \tag{3}$$

Here, $v = (x_{j_2,k} - x_{j_1,k})/\|x_{j_2,k} - x_{j_1,k}\|_2$ is the unit vector in the direction of the limb.

3 Dataset

3.1 CoCo Dataset

COCO is a large image dataset designed for object detection, segmentation, person keypoints detection, stuff segmentation, and so on.[1] This dataset is aiming to advance the state-of-the-art in object recognition. And this is achieved by gathering images of complex everyday scenes containing common objects in their natural context. Objects are labelled using per-instance segmentations to aid in precise object localization[5].

In this project, We used COCO's label 'people key points'. As shown in the above images, each person in coco images has 17 points, and each point has three data, the first two is the x value, y value. The third value indicates the state of the point, 0: not exist, 1:node is labelled but covered by other things, 2 labelled and able to be seen.

Fig. 2: A sample of the CoCo dataset image and its labels.

3.2 Infant Key-point Dataset

This dataset is a small 2D Infant Pose Dataset collected from Shenzhen Children's hospital and Shenzhen Baoan Maternal and Child Health Hospital(due to the clinical data has laws on privacy, we can not put a sample here). As this dataset is without labels, we label it manually using labelme. In detail, the bbox, which describe the area where the infant appears, is indicated and the key points of the infant body are marked just like what the COCO dataset did. After that, the labelme JSON file is transferred to COCO style annotations which can be info can be read by COCO API and able to use COCO API to evaluate the result.

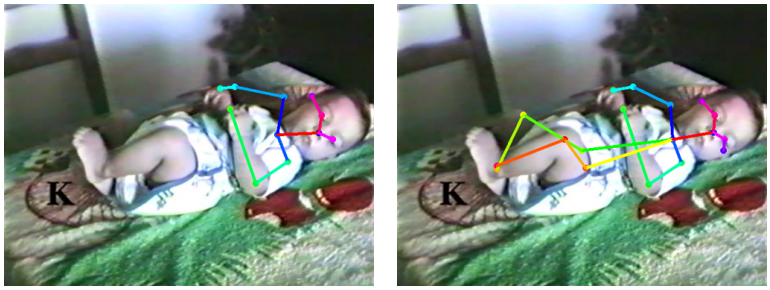
This dataset is used to fine-tune the model to check if the paediatricians can use OpenPose to evaluate an infant's movement.

4 Experiment

After finishing the dataset part, the basic framework of Openpose is implemented by PyTorch. The code is now available at GitHub. Then, because of the time and resource limit, we select a well-trained model from the authors to save training time. Our pre-trained model is trained by a lab server with 4 NVIDIA 1080Ti, and it achieves similar performance as the Openpose paper showed. After we pre-train Openpose on COCO 2014 dataset with 5 epochs, the Infant Key-point Dataset is used to finetune the model. In this part, an SGD optimizer with momentum is used to optimize models and the learning rate is set to 0.01. The models generally converge within 50 epochs. The precision and recall are computed on the validation set. All models are trained on one NVIDIA RTX 2070 GPU.

Table 1: Validation results on original Openpose and Openpose after finetune.

Framework	Average Precision(%)	Average Recall(%)
Openpose (original)	8.30	13.1
Openpose (finetune on Infant-Pose)	73.6	82.7



(a) (b)

Fig. 3: Comparison

As shown in the image and table, the original Openpose(a) can not fit the key points of infant well because the limbs and postures of infants are quite different from adults on the CoCo dataset. After fine-tuning on our self-labelled dataset, the performance improves a lot. See the result table calculated using CoCo API, the precision is increase from 8.3 to 73.6 and recall increase from 13.1 to 82.7.

5 Discussion

In this report, we introduce the initial research plan for Infant-Pose and implement the framework on a self-labeled dataset. However, their is still some gaps for deploying the framework to clinical application. In this section, we will discuss the defect by using Openpose and offer future direction for achieving automatic general movements assessment.

5.1 Defect of using Openpose

Though Openpose's architecture achieves good performance in Infant-Pose dataset, it will still cause some problems. One of the most significant problems is that it lose the 3D keypoints of information. For 2D pictures, we can still obtain 3D features according to the size of the object and parallax. If we want to solve this problem, a possible method is to label the depth of each keypoint in the image and use this information to adjust confidence maps and PAFs.

5.2 Automatic General Movements Assessment(GMs)

In our experiment, we prove that we can use the deep learning method to transfer infant images or videos to a set of keypoints. However, how to use these keypoints in clinical medicine? In the future, we hope to start a clinical experiment to research the equivalence between using original videos and using keypoints video in general movements assessment. If the experiment result shows that they have a strong correlation, we will then use the keypoints set to predict cerebral palsy.

5.3 Relation with Federated Learning

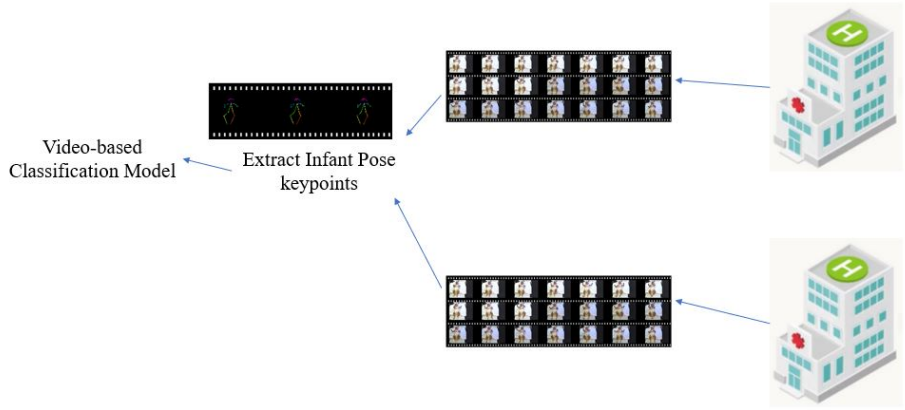


Fig. 4: The deployment goal in the future

As we know, the general idea of federated learning is "Bringing the code to the data, instead of the data to the code". It can solve the problem of protecting the privacy, ownership, and locality of the data during deep learning training and can also achieve remotely shared learning. Infant-Pose can become a useful tool similar to federated Learning in the future. As Fig. 4 shows, Infant-Pose can become a useful tool similar to the effect of federated Learning in the future. The deployment goal of Infant-Pose is to help hospitals using this model to transfer the original videos to keypoints videos and use the new data to train or aggregate a video-based classification model without disclose patients' privacy.

6 Conclusion

In this paper, we propose an Openpose-based method to predict key-points for infant and achieve good performance in experiment. It proves that we can use COCO Key-points Challenge Dataset to pretrain corresponding models and then use transfer learning to fine-tune parameters using infant key-points datasets. In the future, we plan to implement Openpose or its improved version in a larger infant key-points dataset and train an end-to-end model for evaluating infant movement.

References

1. Coco - common objects in context: https://github.com/cocodataset/cocoapi

2. Cao, Z., Hidalgo, G., Simon, T., Wei, S.E., Sheikh, Y.: Openpose: realtime multi-person 2d pose estimation using part affinity fields. IEEE transactions on pattern analysis and machine intelligence **43**(1), 172–186 (2019)
3. Einspieler, C., Bos, A.F., Krieber-Tomantschger, M., Alvarado, E., Barbosa, V.M., Bertoncelli, N., Burger, M., Chorna, O., Del Secco, S., DeRegnier, R.A., et al.: Cerebral palsy: early markers of clinical phenotype and functional outcome. Journal of clinical medicine **8**(10), 1616 (2019)
4. Einspieler, C., Prechtl, H.F.: Prechtl's assessment of general movements: a diagnostic tool for the functional assessment of the young nervous system. Mental retardation and developmental disabilities research reviews **11**(1), 61–67 (2005)
5. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
6. Wei, S.E., Ramakrishna, V., Kanade, T., Sheikh, Y.: Convolutional pose machines. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 4724–4732 (2016)