Physical Parameters for Image Learning in Pediatrics

Examining PSF Blur and RGB Channel Weights for Learning Infant Poses Estimation

Sebi Gutierrez
BME 548L: Machine Learning and Imaging
Fall 2020

Abstract

Pose estimation and tracking, or the tracking of key anatomical landmarks in video or images, offers insight and are keys to the early detection of certain congenital and developmental neuromuscular disorders, to include Cerebral Palsy (CP) and Autism Spectrum Disorder (ASD). For these, there is no one test used as a diagnostic criteria and longitudinal data through ages 2 to 3 of life is usually used for diagnosis. State of the art pose estimation model for clinical applications used here quantitatively measure the characteristic and features of general and fidgety movements typical of these ages but are usually used only controlled, clinical environments. Absence of these can indicate the presence of early forms of neuromotor dysfunction. The fine-tuned domain-adapted infant pose (FiDIP) model, explored here, begins to explore the use of pose-estimation for infant and pediatric images of everyday life. Here, we set off to look at two physicals parameters of RGB channels weights and PSF blur kernels as a means to explore accuracy and adaptability to further non-clinical, non-ideal settings and what modulating these can tell us about important details of images. We find that the full image is ultimately the best at locating key anatomic landmarks, but that RGB channel and blurring have varying and significant effects on this process. This result may arise from certain aspects of the images important to these models, to include native lighting, angle, background.

Introduction

Cerebral Palsy (CP) is one of the most common physical and movement disabilities in children in highly developed countries, with a prevalence of cases 2.1 per 1000 births. In about 80% of these cases, an exact causal mechanism is not understood for the development of CP, so corrective interventions treating the disease etiology are not well. This is also the case with movements displayed of increased risk indicators for the development of autism spectrum disorders (ASD), which affects. Most early interventions for the treatment of pediatric CP and ASD target the physical symptoms of the disease, to include physical occupational, and speech therapy, or corrective surgery if indicated. These interventions focus on improving the quality of live and reducing the impact on the body the patient experiences, with studies showing moderate to good improvements in motor and coordinative functions with early intervention. Therefore, it is critical that screening and diagnostics be done with a great accuracy and as early as possible.

The earliest signs of CP and ASD in most cases can be seen in infancy, when newborn infants learn and develop motor skills by efforts of spontaneous movement that progress into more controlled, directed movement with time. Thus, a diagnosis of CP can usually be made in the 12-24-month range of late infancy, but can be made as early as 4 months in some cases if certain physical signs are present. The current standard of care (SOC) and most common monitoring and diagnostic methods include observational movement exams and assessments, such as Prechtl's General Movement Assessment (GMA) and the Test of Infant Motor Performance (TIMP) for infants under 5 months and Alberta Infant Motor Scale (AIMS) and Developmental Assessment of Young Children (DAYC) older infants. Each of these assessments has proven successful with high inter-rater reliability but requires a qualified and trained physical therapists in person under ideal clinical settings to perform the exam. Consequently, issues of applicability have arisen when using these exams, with some calling for its expanded use outside the clinical setting to allow for more continuity of monitoring and applicability in more remote settings where access to providers is more restricted.

Advances of the recent decades in imaging analysis technologies have allowed for the development of marker-less, motion capture techniques that are able to extract quantitative movement parameters through processing of key anatomical landmark identification. These powerful analysis tool historically have been used in athletics and military performance applications to track, record and model the movements of people or objects through space. They have also found use in

certain medical settings, like those of rehabilitation, particularly in the area of injury recovery and evaluation. In instance, it's applied by researchers here at the Duke Krzyzewski Human Performance Laboratory (K-Lab) for study of musculoskeletal injury and rehab amongst college athletes. This technology is now starting to be applied in other areas of clinical medicine, as the one explored here, for automation of movement assessments of infants, used to predict and stratify infants' relative risks of developing neuromotor or neurocognitive disease. From normal, RGB color images and video, infant anatomic landmarks can be tracked over time and compiled to form infant stick-figure-like assemblies representing the trajectories and movements of the different body parts through the use of complex, neural network based models, such as those seen in the image below by Marchi et al using OpenPose, an open source computer vision software by Cao et al.

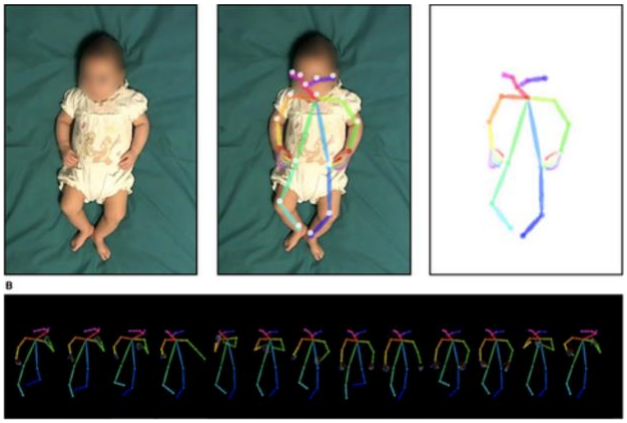

Figure 1. 13 sample pose estimations of infant from RGB video (Marchi et al)

The difficulty in the expansion to out-of-clinical setting of such modalities, however, lies in the image/video point of view and recording quality. In studies such as Marchi et al, efforts were made to stabilize handheld shake from camera motion by fixing axis origin at the neck identified key point, but these processing steps can only be used for visualizations when the subject is supine, relatively centered in the frame. If the goal is to expand the applicability of such pose estimation models for use outside of the clinical setting, say at home, daycare, or any location where a child might locomote or display a variety of spontaneous to volitional movements, then the robustness of these models must be validated for these settings. This has been studied in the work done Huang et al upon whose work this study aims to build. Huang et al, through the use of internet-acquired images of videos, have demonstrated the application of their developed fine-tuned domain-adapted infant pose (FiDIP) estimation model.

Here, the authors seek to explore two physical parameters and how these may affect pose estimation accuracy. The first of these is the use of different weights for each of the three red, green and blue color channels in the jpg color images. Such a step helps elucidate on the important of certain colors in images be it in contrast between objects or similarities of colors between images. A sample image from the dataset used in this paper is seen below, as its original image in the top left and 3 distinctive color channels. This can also be used to gain insight into artificial image coloring, should it be necessary, for applications when images might be black and white or infrared, when monitoring infant's movements at night during sleep.

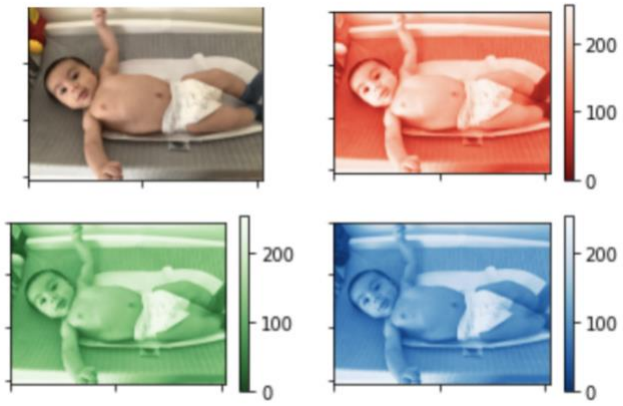

Figure 2. Sample infant image and it's respective RGB channel weights

The second physical parameter explored is the use of a blurring filter, the one used here a Gaussian blurring filter. A Gaussian kernel acts as filter, effectively blurring through a convolution, an image at the level of the aperture. This blurring simulates the application in lower resolution before downsampling to enter the pose model as well as the possibility of out of focus images and videos passed into the model. This information can inform applications in areas where ideal imaging may be difficult such as throughout the home or daycare, where imagine capture systems may be fixed far away from the subject of interest. An example of the application of this Gaussian kernel is seen in the figure below on the following page.

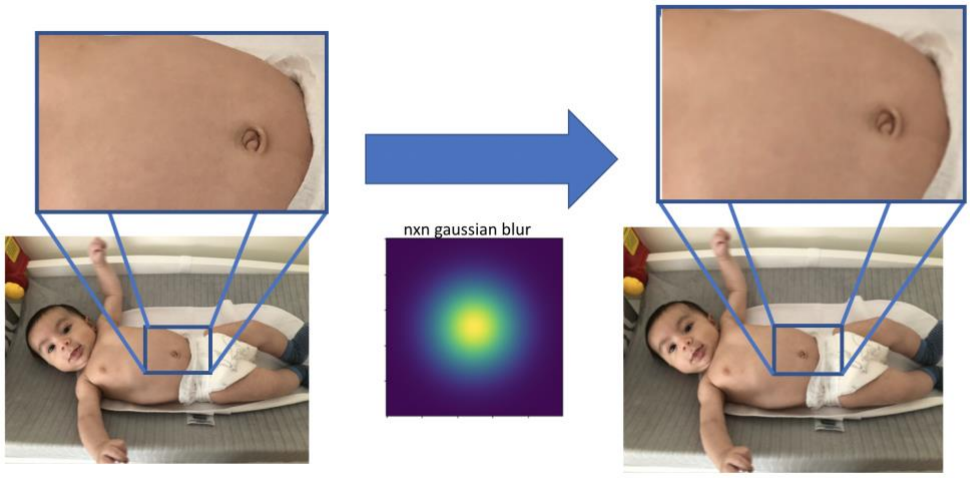

Figure 3. Sample infant image and blurred convolving using a Gaussian blur kernel

Related Work

Prior work in this area of investigation is quite varied in its images, techniques and primary outcomes. Primarily, there is a difficulty in obtaining images of infants, as many sets of images used in these instances are collected as part of clinical trial work, and bound by The Health Insurance Portability and Accountability Act of 1996 (HIPAA) if national, or other patient safety and information acts worldwide. Huang et al, the dataset used for this project, was manually collected from real images of infants from across the web, harnessing media repositories and search engines like Google Images and YouTube. In this application, however this turns out to be advantages, as the images collected through these means offer a larger variety of posse, backgrounds and lighting conditions.

Others work seeks to improve the automation of pose estimation and bridge the gap between computer vision and clinical diagnostic capability. Marchi et al explore the agreement between automated risk stratification and independently, human-reviewed pose models and images, showing that continuous, skeleton pose models do poses enough clinical data to allow for diagnosis, compared to standard RBG video.

Still others take a different approach to infant and movement monitoring, employing wireless accelerometer system and comparing their continuous information stream to video media interpreted by physical therapy professionals with high (>70%) accuracy, in Singh et al. Joshi et al explored the use of ballistographic data for continuous movement monitoring in infants using thin film sensors placed underneath the subject, again being compared to annotated video as standards of reference. Although only a small fraction of state of the art work is described here, no research seems to make headway in exploring physical parameter layers to pose estimation architectures or its implications for use in other settings.

Methods

The data used for this study was acquired from Augmented Cognition Laboratory at Northeastern University's School of Engineering in Boston, MA. As downloaded, it consisted a pre-split training and validation set of infant images of various sizes and web sources as was described in the prior section. The sizes and type are further explored in the table below.

	Training	Testing
Image Size	Real: Varying (~500x500 to 2000x2000) Computer: 480x640	Varying (~500x500 to 2000x2000)
Origin	YouTube, Google Images, Computer Generated	YouTube, Google Images
Set Size	904 images	100 images

Figure 4. Quantitative descriptors of infant image dataset, by subset

The images were also accompanied by .json files for each, the training and testing set, with annotated ground truth coordinates for the infant adapted pose network. The model employed for this project was a modified application of the fine-tuned domain-

adapted infant pose (FiDIP), also developed by Augmented Cognition Laboratory at Northeastern University. This model is an extension of PoseResNet, which is the primary composed of a ResNet-50 architecture with added deconvolution layers in the latter stages. The model, when downloaded, already is pretrained on images from COCO adult pose dataset, but for every iteration is re-"fine trained" from the infant training set. The model employs an Adam optimize with learning rate set to 0.001, but uses different batch sizes and epochs for both training and testing due to their differences in size. For fine-tuning stage, in this application for this project the training step, there were 80 epochs and 85 images in a batch. The model, accepts in images of all size, and includes a resizing feature as its first layer, resizing all image to maintain aspect ratio and padding to an input size of 384 by 288. The output of the model after running on the infant validation data is: a .json with the infant's key anatomic features' as predicated by the model with a pose loss score for each infant as the mean square error (MSE) between the predated mapping and the ground truth for each point i:

$$L_P = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

.

Individual epoch time in calculation, loss and accuracy, as well as their running averages, were also presented as logger messages when running the testing scripts. A subset of the testing images area also output, in their final 348x288 form, with heatmap locations of each of the 16 anatomic key points, as can be seen in the figure below.
The model was access through its github repository and downloaded to python directly. Color and blur processing of images was done prior to initialization of the model, given the innate complexity of the multiple files associated with the model. The model was run 5 times, making minimal changes to the original code: Once as is, once grayscale image, once allowing for RGB color weights, once with the variable blur kernel and once with both the weights and blur kernel. The Gaussian blur kernel shape was based on the native resolution of the image after padding to maintain aspect ratio, with initiated to a standard deviation that was trainable. When run as a grayscale images, with and without training of weights, images were converted to grayscale using the using the luminance method as described here:

$$Grayscale = 0.21 * R + 0.72 * G + 0.07 * B$$

Figure 5 Example visual, heatmap output of the Fine-tuned domain-adapted infant pose (FiDIP)

Results

The results of training and testing rounds for each of the five runs are summarized in the table below as accuracy presented for both the last epoch to be run and the running average across all epochs. They were collected from the logger display built into the file.

	Accuracy (MSE)
Images	(Final Epoch, Avg)
As Is	0.939 (0.917)
Gray	0.733 (0.712)
RGB	0.747 (0.741)
Blur	0.846 (0.849)
Blur + RBG	0.819 (0.820)

Figure 6 Accuracy Outcomes as MSE from FiDIP run

As can be seen from cumulative averages, the best preforming run of the model was a vanilla run, making no modifications to the image set of the model. This scored in the 91.7% accuracy range, close to the original developer's published max

accuracy of 92.5%, reassuring of a successful implementation of the model in this code.
When examining the physical parameters after being run through the model, they are shown below.
The results of the physical layer's kernel blur for the two, one the right the kernel from the blur alone and the left from the run grayscale run with colors channels. It is interesting to note that both seem to have converted to a larger Gaussian kernel, with a very small covariance compared to the seeded kernel from above. Since the blur kernels in application acts as a low-pass filter, allowing only lower frequencies to pass through when convolved, the maximization of these implies the importance of higher frequencies in the mode. For the purposes of pose estimation, this manifests the importance as edges in the spatial domain, with sharper variations producing sections in the high frequencies of the spectrum.

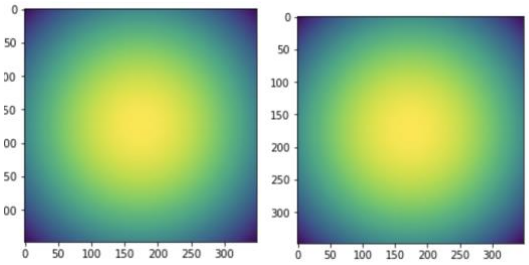

Figure 7 Blur kernels from model

Discussion

Although the results of the implementation of the physical models did not necessary improve model performance, as was hoped, they do still offer insight into the kinds of data valued by pose estimations models built on similar architectures. The final blur kernel suggest that as implemented, the model benefits from more resolved images before being downsampled. This would limit its implementation in imaging systems where the subject would be smaller than a few hundred pixels square. The variation in backgrounds and poses of the subject, sometimes not centered, and variety of lighting conditions did not produce optimization either of the scalar weights for RGB, with them appearing to have only slight variation from initial settings. Since in both applications, this layer appears to have only negatively impacted accuracy, a simple approach like channel weighing may be insufficient and more robust methods may be needed to address the instance of color.

Other limitation on this project with room for further study were imposed by the limited size of the data set, as generally datasets containing thousands of images are used to train. Of note regarding the dataset was also the underrepresentation of infants of color, drawing to question the applicability and accuracy for implementation in products or software for wider use. AI and health technology have a history of implicit bias against certain historically marginalized populations at all levels of application, to include biometric security software used by governmental organizations.

Future work may also look into the application of FiDIP like models for video, where video tracking, as oppose to frame by frame estimation as is done in single images, offers an extra dimensionally to the data perhaps opening an avenue still where physical layers may play a role in developing the future of specialized pose estimations models for clinical application.

References

Ei Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. 7291–7299. https://openaccess.thecvf.com/content_cvpr_2017/html/Cao_Realtime_Multi-Person_2D_CVPR_2017_paper.html

Einspieler, C., Bos, A. F., Libertus, M. E., & Marschik, P. B. (2016). The General Movement Assessment Helps Us to Identify Preterm Infants at Risk for Cognitive Dysfunction. Frontiers in Psychology, 7. https://doi.org/10.3389/fpsyg.2016.00406

Ferrari, F., Cioni, G., Einspieler, C., Roversi, M. F., Bos, A. F., Paolicelli, P. B., Ranzi, A., & Prechtl, H. F. R. (2002). Cramped Synchronized General Movements in Preterm Infants as an Early Marker for Cerebral Palsy. Archives of Pediatrics & Adolescent Medicine, 156(5), 460. https://doi.org/10.1001/archpedi.156.5.460

Grunewaldt, K. H., Fjørtoft, T., Løhaugen, G. C. C., Evensen, K. a. I., Brubakk, A.-M., & Skranes, J. (2011). Lack of Fidgety Movements at 15 Weeks Post-Term Relates to Cerebral Palsy and Adverse Cognitive Outcome in Preterm Born Children at 10 Years of Age. Pediatric Research, 70(5), 322–322. https://doi.org/10.1038/pr.2011.547

Hadders-Algra, M. (1993). General movements in early infancy: What do they tell us about the nervous system? Early Human Development, 34(1), 29–37. https://doi.org/10.1016/0378-3782(93)90038-V

Huang, X., Fu, N., & Ostadabbas, S. (2020). Infant Pose Learning with Small Data. ArXiv:2010.06100 [Cs]. http://arxiv.org/abs/2010.06100

Joshi, R., Bierling, B. L., Long, X., Weijers, J., Feijs, L., Pul, C. V., & Andriessen, P. (2018). A Ballistographic Approach for Continuous and Non-Obtrusive Monitoring of Movement in Neonates. IEEE Journal of Translational Engineering in Health and Medicine, 6, 1–10. https://doi.org/10.1109/JTEHM.2018.2875703

Marchi, V., Hakala, A., Knight, A., D'Acunto, F., Scattoni, M. L., Guzzetta, A., & Vanhatalo, S. (2019). Automated pose estimation captures key aspects of General Movements at eight to 17 weeks from conventional videos. Acta Paediatrica, 108(10), 1817–1824. https://doi.org/10.1111/apa.14781

Støen, R., Songstad, N. T., Silberg, I. E., Fjørtoft, T., Jensenius, A. R., & Adde, L. (2017). Computer-based video analysis identifies infants with absence of fidgety movements. Pediatric Research, 82(4), 665–670. https://doi.org/10.1038/pr.2017.121

Zuk, L. (2017). Fidgety movements, cerebral palsy, and cognitive ability. Developmental Medicine & Child Neurology, 59(6), 568–568. https://doi.org/10.1111/dmcn.13410