Estimating the Gait Deviation Index of Children with Cerebral Palsy

Vincent Salpietro, Jon Stingel, and Cara Welker

CS 230 Final Poster Presentation, Fall 2018

Introduction

Cerebral palsy (CP) is a motor disorder that causes impaired musculoskeletal development, and affects approximately 3 out of every 1000 children born [1]. There are a range of treatment options from physical therapy to surgery, and quantitative measures calculated from gait analysis such as the Gait Deviation Index (GDI) can be used to help inform treatment decisions. Historically, marker-based motion capture has been used to collect gait analysis data, but these setups are expensive (tens of thousands of dollars), and require specialized spaces and personnel. Compared to the current standard of one or two visits max to the gait lab for each child, if GDI could be calculated with 2D video, this would allow patients (or their parents) to collect gait data at home, and allow doctors to continuously monitor their progress.

Motivation

Can we use deep learning to predict the gait deviation index of children with cerebral palsy from 2D videos?

Dataset

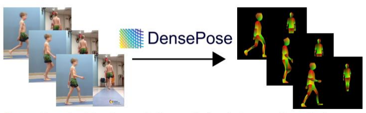

Figure: Sample video frames before and after being run through the DensePose network [2]

Our dataset was provided by Gillette Children's Hospital and consists of $\approx$ 500 videos of children with CP walking, along with paired GDI scores calculated from motion capture for each child from that visit. We augmented this data by splitting each video into segments of 124 frames, giving us $\approx$ 6500 videos. These videos were initially run through the DensePose network, which associates points on a 2D image to a surface-based representation of the human body.

Model

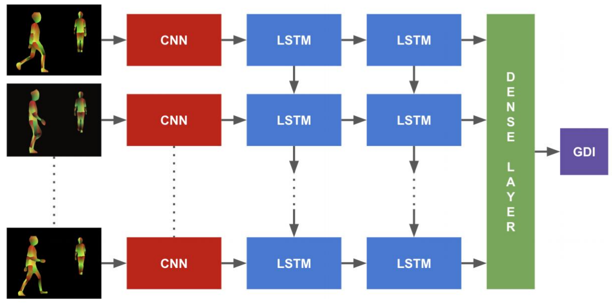

Figure: Model architecture where densepose video frames are fed into a CNN to get embeddings, which are passed through two LSTM layers to predict the output GDI (adapted from [3])

Methods

We started by feeding individual DensePose video frames into a small convolutional network with a classifier predicting individual image GDI in one of seventeen bins. We then used transfer learning on Alexnet with various numbers of frozen layers to get a deeper network. We decided to add LSTM layers to account for the time dependance of our video frames. We gathered feature vectors from the output of Alexnet, and fed those into the LSTM layers to get a single GDI score for an entire video's worth of DensePose image frames.

Results

Model	Val Acc	Correlation
CNN Baseline	19.8%	0.331
Pretrained Alexnet	25.8%	
Finetuned Alexnet	28.5%	0.202
Alexnet + LSTM	**33.3%**	**0.568**

Table: Summary statistics comparing different models run on the DensePose dataset and showing improvement from the baseline CNN to the addition of LSTM

Discussion

Given the nature of our GDI data (i.e. temporal, clustered near middle scores), predicting the output using a purely CNN model was difficult. To train the CNN, it was required that we associate the patients' GDI with each individual image versus associating a single GDI with a series of images. This removed the temporal aspect from the data, which we believe contributed to our low validation accuracies. In order to combat this, we implemented a sequential LSTM network taking as inputs the features pulled out from the CNN. Due to the non-representative training of the CNN, along with the lack of data spread across bins, the LSTM still suffered from low validation accuracies (though its correlation value of 0.568 is not vastly different than the correlation of 0.74 in a state-of-the-art network given the same task)[4]. Training an end-to-end CNN into LSTM network with more data may solve a lot of these issues and result in higher accuracies.

Contact Information

- Github: https://github.com/cgwelker/clinical_parameters_gait
- Email: cgwelker@stanford.edu, stingjp@stanford.edu, vsal@stanford.edu

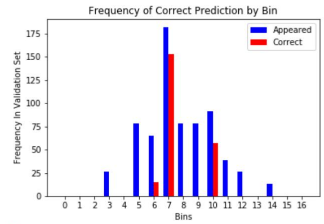

Figure: Frequency of target appearance compared to correct prediction of LSTM model on validation set.

Future Work

Future work on the model focuses on more personalized training of the model on the provided videos, as well as focus on improving the model accuracy through reduction of bias:

- Consider an end-to-end approach
- Gather more videos
- Preprocess/augment the data with additional information
- Retrain a full AlexNet on images with sequence model
- Increase the size of the network

References

[1] Arneson et al.
Prevalence of cerebral palsy: Autism and developmental disabilities monitoring networke.
2004 Disability and Health Journal, pages 45–48, 2009.

[2] Guler et al.
Densepose: Dense human pose estimation in the wild.
arXiv, 2018.

[3] Donahue et al.
Long-term recurrent convolutional networks for visual recognition and description.
arXiv, 2016.

[4] Kidzinski et al.
Automatic diagnostics of gait pathologies using a mobile phone.
unpublished work.

Acknowledgements

Thanks to Gillette Children's Hospital for providing the dataset, Lukasz Kidzinsksi for running the videos through DensePose and for project mentorship, and the CS230 teaching team for the guidance provided on this project, especially Aarti Bagul.