Remote Kinematic Analysis for Mobility Scooter Riders Leveraging Edge AI

Thanh-Dat Nguyen, Chenrui Zhang, Melvin Gitbumrungsin, Amar Raheja, Tingting Chen

California State Polytechnic University, Pomona
datnguyen1@cpp.edu, chenruizhang@cpp.edu, melving@cpp.edu, raheja@cpp.edu, tingtingchen@cpp.edu

Abstract

Current kinematic analysis for patients with upper or lower extremity challenges is usually performed indoors at the clinics, which may not always be accessible for all patients. On the other hand, mobility scooter is a popular assistive tool used by people with mobility disabilities. In this study, we introduce a remote kinematic analysis system for mobility scooter riders to use in their local communities. In order to train the human pose estimation model for the kinematic analysis application, we have collected our own mobility scooter riding video dataset which captures riders' upper-body movements. The ground truth data is labeled by the collaborating clinicians. The evaluation results show high system accuracy both in the keypoints prediction and in the downstream kinematic analysis, compared with the general-purpose pose models. Our efficiency test results on NVIDIA Jetson Orin Nano also validate the feasibility of running the system in real-time on edge devices.

Introduction

Kinematic Analysis is often used as an outcome measure to evaluate the performance of patients with upper or lower extremity challenges after an injury or with certain diseases such as Parkinson's (An 1984). Current kinematic analysis is usually performed indoors at the clinics based on motions captured by optoelectronic camera systems (Ponsiglione et al. 2022). However, such evaluation at the clinics cannot always be accessible for patients because of various reasons such as difficulty of transportation and the COVID-19 pandemic.

In this study, we aim to explore the feasibility of performing real-time kinematic analysis by leveraging the deep learning models on devices attached to mobility scooters, a popular assistive mobility tool, for patients to use at their communities. When riding mobility scooters, patients constantly have upper-body movements showing their muscle and joint abilities, such as the voluntary posture sways on an uneven surface and the arm extension movements when reaching a door opener. We leverage these opportunities to perform seamless kinematic analysis in a portable fashion. This work is also meaningful in providing a system to help monitor mobility scooter riders' safety when their upper extremity symptoms may possibly progress.

As used by most of the kinematic analysis carried out in the clinics, we leverage cameras to capture patients' motion patterns and design a system to perform remote analysis based on the video frames. There are several challenges in designing such a system: 1) As mobility scooters are often used in the outdoor setting, there are various backgrounds and environmental conditions that are new to traditional kinematic analysis. It requires the use of cutting-edge deep learning models such as those for human pose estimation (Zheng et al. 2023) that have been trained and tested with real-world scenarios. 2) Validating the effectiveness and accuracy of kinematic analysis results poses another challenge for this application. General-purpose deep learning models may not serve the needs of clinicians in understanding the motions of patients. Human experts with domain knowledge should be included in the loop of system design and development. 3) To enable remote kinematic analysis and protect the patients' privacy, the model inference should be deployed on local devices installed on the mobility scooter. It requires high efficiency of the system on resource constrained computing platforms.

To address the aforementioned challenges, our remote kinematic analysis system for mobility scooter riders has the following features: a) We generate our own dataset by capturing real-world mobility scooter riding videos from physical therapy patients. This tailored data is used to train human pose estimation models, ensuring relevance and accuracy; b) Clinicians are actively involved in both the data collection and annotation processes. Their expertise is leveraged to create accurate upper-body keypoints ground truth, which is essential for precise kinematic analysis; c) The system is implemented and deployed on NVIDIA Jetson Orin Nano using TensorRT. This ensures efficient processing on edge devices, enabling real-time analysis without compromising performance.

Our remote kinematic analysis system design focuses on practicality, clinical relevance, and efficiency. This paper covers some initial results to show the method's feasibility and we plan to further validate various kinematic analysis measurements generated from the system, by comparing them with the traditional IMU-based solutions (Stanev et al. 2021).

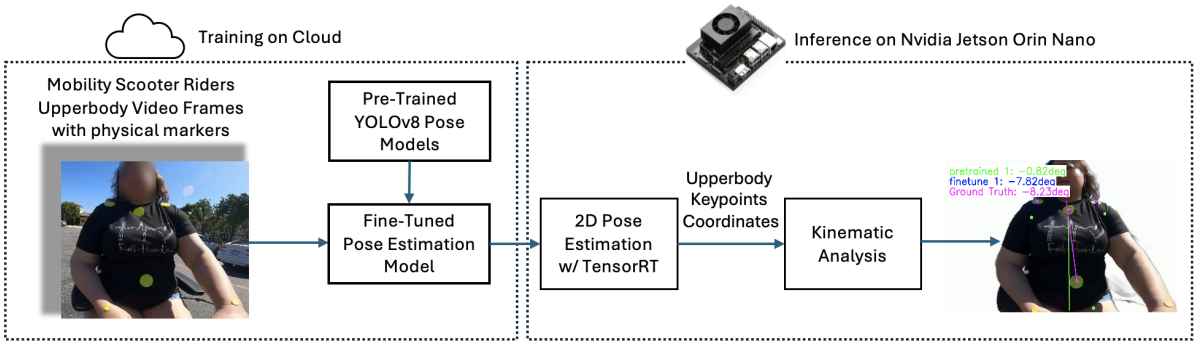


Figure 1: Pipeline of Remote Kinematic Analysis System based on Human Pose Estimation

System Overview and Methods

System Pipeline

To analyze the kinematic patterns of mobility scooter riders in a seamless fashion, our system relies on the patients' movement data collected from cameras, consistent with such systems in the clinical setting (Mikami, Shiraishi, and Kamo 2022). In the future, a full version of the system will perform various kinematic analysis measurements. In this paper, we demonstrate one function, i.e., trunk lateral flexion analysis, to show the system workflow and the feasibility of methods.

We recognize and emphasize the importance of patients' privacy, so most computing tasks are completed on edge devices locally with the patients, except the final calculation results. In this section, we present the system pipeline first and then describe each component in greater details.

As shown in Figure 1, our system takes video frames of the mobility scooter riders' upper-body movements as input. We collect our new dataset of mobility scooter driving videos in which physical markers are placed on riders' upper body. The new dataset is used to fine-tune a pre-trained YOLOv8 human pose estimation model (Jocher, Chaurasia, and Qiu 2023). The model training is carried out in the cloud environment.

In the inference stage, the fine-tuned model is converted with TensorRT (Corporation 2024) for better efficiency and deployed on edge device NVIDIA Jetson Orin Nano (NVIDIA 2024). The fine-tuned human pose estimation model for mobility scooter riding outputs upper-body keypoints in 2D coordinates. The extracted upper-body keypoints of riders are used to perform the downstream kinematic analysis. The kinematic analysis results from the edge device can be sent to the clinicians via secure network communications, to enable real-time remote monitoring.

Riders Upperbody Keypoints Extraction

In order to perform kinematic analysis of mobility scooter riders remotely, we first extract their upper-body keypoints from the video frames by leveraging the widely applied human pose estimation models. In particular, we apply two specific models for accuracy and efficiency comparisons: YOLOv8x-pose-p6 and YOLOv8n-pose (Jocher, Chaurasia, and Qiu 2023). The benchmark tests show higher accuracy for YOLOv8x-pose-p6 and higher inference efficiency for YOLOv8n-pose. Although YOLOv8 detects 17 keypoints from the human body, our mobility scooter rider kinematic analysis system only uses 9 of them located on the upper body: neck, left shoulder, right shoulder, left elbow, right elbow, left wrist, right wrist, left hip, and right hip. YOLOv8 pose models output 2D coordinates of the 9 keypoints for each input video frame with 30 frames per second.

The general-purpose human pose estimation models serve as a solid baseline for our remote kinematic analysis. However, for the mobility scooter riding postures, we need to fine-tune the pose estimation models with the new domain-specific keypoint ground truth data.

Keypoint Ground Truth

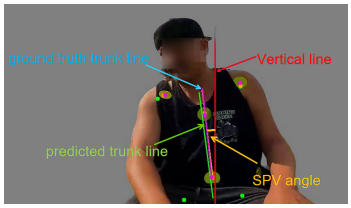
To establish a reliable ground truth for keypoints used for kinematic analysis, we collaborate with clinicians in the Kinesiology Department and leverage their domain expertise. Before collecting the mobility scooter riding videos, clinicians place round stickers with a 2-inch diameter on the patient's upper body, as illustrated in Figure 2a. These markers are strategically positioned on specific anatomical landmarks such as joints and key body points. The placement of these markers is meticulously carried out by trained professionals to ensure they correspond accurately to the anatomical features of interest. Each marker's position is extracted using YOLOv8x object detection model to create a comprehensive set of labeled keypoints, which constitutes the ground truth for our system.

This carefully curated ground truth data is used for fine-tuning and testing the pose estimation models and our subsequent kinematic analysis. Please note that the markers are only placed on patients when collecting training data, but not needed in the pose estimation model inference or kinematic analysis.

Kinematic Analysis

From the keypoints predicted by the fine-tuned pose estimation models, we perform kinematic analysis. The initial results are specifically focused on trunk lateral flexion to provide insights on the patient's ability to perform upper-body movement.

(a) Placing Markers



(b) SVP angle illustration

Figure 2: Illustration of the process of placing markers on patients' upperbody keypoints and the trunk lateral flexion angle calculation.

The trunk lateral flexion angle, or SPV angle, is defined as the angle between the vertical axis and the axis of lateral flexion, which is commonly used in physical therapy and biomechanics to assess the range of motion (Mikami, Shiraishi, and Kamo 2022). As depicted in Figure 2b, the trunk lateral flexion angle is measured using specific anatomical landmarks: the belly button (the center of mass in a sitting position) and the top of the sternum.

From the human body keypoints defined in the COCO-Pose Dataset (Lin et al. 2015), we use the coordinates of the left and right shoulders and left and right hips (the green dots in Figure 2b) to calculate the trunk lateral flexion angle. In particular, the midpoint (averaging the x and y coordinates respectively) of two shoulder points is mapped to the top of the sternum, and the midpoint of two hip points is estimated as the center of mass, as the belly button landmarks are not directly available in COCO-Pose. The SPV angle (in degrees) is calculated simply as $\theta = \frac{180}{\pi} \arctan \frac{y_t - y_b}{x_t - x_b}$ where (x_t, y_t), (x_b, y_b) are the calculated x,y coordinates for the top of sternum and the belly button respectively. We can also see from Figure 2b that the keypoints coordinates as the results of pose estimation may be different than the ground truth (red dots) annotated by the physical markers. Consequently, the calculated SPV angle value may deviate with some errors as well.

Preliminary Evaluation and Results

Our human pose-based remote kinematic analysis prototype system is developed in Python with libraries PyTorch with torchvision (Ansel et al. 2024), and OpenCV(Bradski 2000) among others, and deployed on the NVIDIA Jetson Orin Nano Developer Kit 8GB module, providing 1024-core GPU and 32 Tensor Cores, which can be placed in the basket of the mobility scooter. Using the real-world mobility scooter riding data, we carry out both accuracy and efficiency tests. The models are re-trained on Delta system with NVIDIA A100 GPUs with 40GB HBM2 RAM at National Center for Supercomputing Applications (Gateway 2024).

Data Collection

We collect mobility scooter riding data from 11 patients with different medical conditions including stroke, neuropathy, brain injury, and Arthritis. Participants are instructed to complete various driving tasks on a Drive Medical Phoenix LT 4 Wheel Mobility Scooter. We mount an IMX219 120° HD camera on the mobility scooter handle, facing the rider to capture the video frames of their upper extremity motions. In total, from 15 video clips, we collected 388,435 video frames, all of which contain physical markers on patients' upper-bodies to generate the ground truth. Image preprocessing steps include background removal (Kim et al. 2022), noise reduction, and normalization to ensure the images were optimally suited for training the pose estimation models. In the dataset, 85% of the frames are used for training, and 15% are used for testing. Our data collection and experiments have been approved by the university's Institutional Review Board (CPP-IRB 22-88).

Model	OKS
YOLOv8x-pose-p6	0.941
YOLOv8n-pose	0.639
Our fine-tuned YOLOv8x-pose-p6	**0.994**
Our fine-tuned YOLOv8n-pose	**0.985**

Table 1: Average OKS for different models with $\sigma = 0.05$

Performance Evaluation

To evaluate our fine-tuned pose estimation models and the kinematic analysis performance, we perform two sets of experiments, one focusing on testing the keypoints prediction accuracy in the mobility scooter driving scenarios and the other on the trunk lateral flexion angle accuracy.

Human Pose Keypoints Prediction Accuracy The metrics we use to evaluate the human pose keypoints prediction accuracy include Object Keypoint Similarity (OKS) (Lin et al. 2015), Percentage of Correct Keypoints (PCK) (Andriluka et al. 2014) (predicted keypoints are under a distance threshold with the ground truth), and Area Under the Curve (AUC) (Fawcett 2006) for PCK when varying the threshold.

Table 1 shows the average OKS values of the two shoulder points across all tested video frames on different pose estimation models. The keypoint standard deviation σ is set to 0.05 in the test. Higher OKS indicates that the predicted keypoints are closer to the ground truth keypoints, considering the object's scale and keypoint visibility. We observe that our fine-tuned pose estimation models have higher OKS compared to the pre-trained YOLOv8 models.

We also measure the PCK values of our fine-tuned pose estimation models when varying the thresholds from 0.1 to 0.3, as shown in Figure 6. The thresholds represent the fraction of the distance to the length of the object (human body) being analyzed. Compared with the original YOLOv8 models, our fine-tuned models exhibit notably higher PCK values across all thresholds, indicating improved keypoint prediction accuracy. Figure 6 also includes the overall AUC values of PCK. The AUC for our models is calculated over the range of thresholds from 0.1 to 0.3 and then normalized to a scale from 0 to 1 as standard. As shown, our fine-tuned models outperform the original YOLOv8 models, demonstrating superior performance in keypoint detection accuracy.

Trunk Lateral Flexion Angles Evaluation To evaluate the

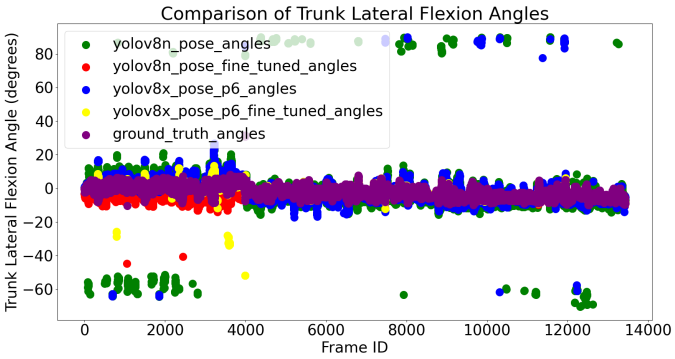


Figure 3: Comparison of Trunk Lateral Flexion Angles

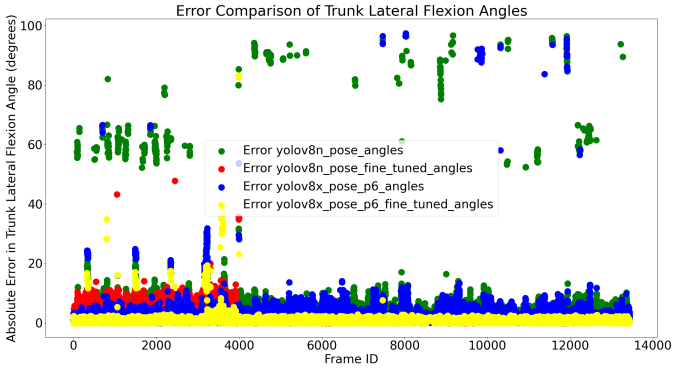


Figure 4: Absolute Error in Trunk Lateral Flexion Angle

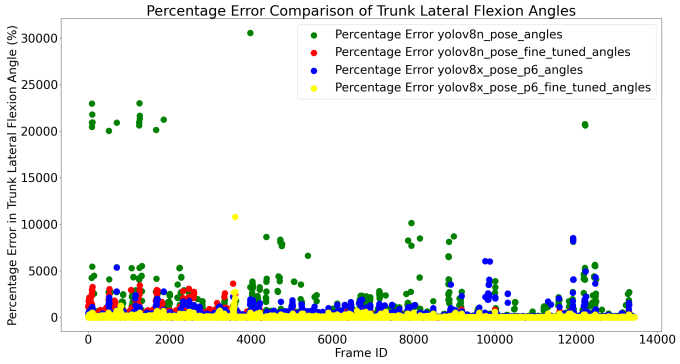


Figure 5: Percentage Error Comparison of Trunk Lateral Flexion Angles

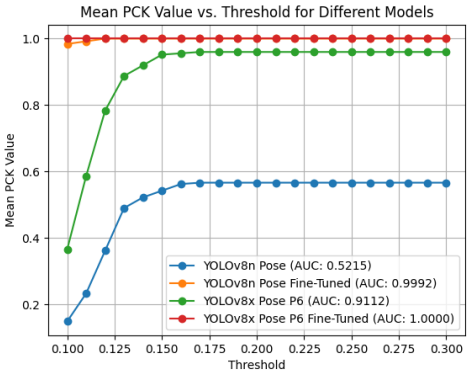


Figure 6: PCK Curve of keypoints predicted on different pose models with thresholds ranging from 0.1 to 0.3

trunk lateral flexion angle calculation accuracy, we use the angle calculated using the physical markers on the drivers' upper body as the ground truth.

The evaluation of the angle results is based on 13,433 frames. Figure 3, 4 and 5 respectively depict the calculated angles in degrees by the models and the ground truth for all frames, absolute deviation from the ground truth, and the percentage error of different models. They consistently show that for the downstream kinematic analysis, our fine-tuned models have superior accuracy performance compared to the pre-trained YOLOv8 models. It indicates the importance of including domain-specific ground truth data with the human experts' intervention.

Efficiency Test

We apply human pose estimation models pre-trained YOLOv8n-pose and our fine-tuned based on YOLOv8n-pose to generate keypoints coordinates for different efficiency needs. We converted and imported the aforementioned models to TensorRT (Corporation 2024) for better inference efficiency performance.

Model	fps	time (ms)
YOLOv8n-pose w/ TensorRT	20	30.4
YOLOv8n-pose w/o TensorRT	16	78.5
Our fine-tuned model w/ TensorRT	**31**	**15.1**
Our fine-tuned model w/o TensorRT	**21**	**50.2**

Table 2: Prototype System Efficiency Test Results on NVIDIA Jetson Orin Nano.

In Table 2, the time column refers to the inference time for processing a single frame using the YOLOv8n-pose models with and without TensorRT on Jetson platforms. The results show that TensorRT significantly enhances performance, with the YOLOv8n-pose model processing frames at 20 fps and 30.4 ms per frame, compared to 16 fps and 78.5 ms without TensorRT. By fine-tuning the pre-trained model with high precision and focusing on fewer important keypoints, the fine-tuned model benefits even more from TensorRT, achieving 31 fps and 15.1 ms per frame, while the same model without TensorRT processes frames at 21 fps and 50.2 ms. These results illustrate that TensorRT optimization greatly reduces inference time and improves frame rate, making it a crucial tool for the efficient deployment of pose estimation models on edge devices.

Acknowledgments

This work is supported partly by grant NSF CNS #2318671.

This work used the Delta system at the National Center for Supercomputing Applications through allocation [allocation number] from the Advanced Cyberinfrastructure

Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

We thank our collaborators Dr. Mai Jara, Joshua Rogers and Michihito Ichihara from Department of Kinesiology and Health Promotion at Cal Poly Pomona for their efforts and supports in mobility scooter riding video data collection and annotation in this project.

References

An, K.-N. 1984. Kinematic analysis of human movement. *Annals of biomedical engineering*, 12: 585–597.

Andriluka, M.; Pishchulin, L.; Gehler, P.; and Schiele, B. 2014. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ansel; et al. 2024. PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. In *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*. ACM.

Bradski, G. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Corporation, N. 2024. NVIDIA TensorRT. https://developer.nvidia.com/tensorrt.

Fawcett, T. 2006. Introduction to ROC analysis. *Pattern Recognition Letters*, 27: 861–874.

Gateway, D. S. 2024. Delta Science Gateway Documentation. https://gateway.delta.ncsa.illinois.edu/wiki.

Jocher, G.; Chaurasia, A.; and Qiu, J. 2023. Ultralytics YOLO. https://github.com/ultralytics/ultralytics.

Kim, T.; Kim, K.; Lee, J.; Cha, D.; Lee, J.; and Kim, D. 2022. Revisiting Image Pyramid Structure for High Resolution Salient Object Detection. In *Proceedings of the Asian Conference on Computer Vision*, 108–124.

Lin, T.-Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C. L.; and Dollár, P. 2015. Microsoft COCO: Common Objects in Context. arXiv:1405.0312.

Mikami, K.; Shiraishi, M.; and Kamo, T. 2022. Effect of subjective vertical perception on lateral flexion posture of patients with Parkinson's disease. *Scientific Reports*, 12(1): 1532.

NVIDIA. 2024. Jetson Orin Nano Datasheet. https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-orin/.

Ponsiglione, A. M.; Ricciardi, C.; Amato, F.; Cesarelli, M.; Cesarelli, G.; and D'Addio, G. 2022. Statistical Analysis and Kinematic Assessment of Upper Limb Reaching Task in Parkinson's Disease. *Sensors*, 22: 1708.

Stanev, D.; Filip, K.; Bitzas, D.; Zouras, S.; Giarmatzis, G.; Tsaopoulos, D.; and Moustakas, K. 2021. Real-Time Musculoskeletal Kinematics and Dynamics Analysis Using Marker- and IMU-Based Solutions in Rehabilitation. *Sensors*, 21(5).

Zheng, C.; Wu, W.; Chen, C.; Yang, T.; Zhu, S.; Shen, J.; Kehtarnavaz, N.; and Shah, M. 2023. Deep learning-based human pose estimation: A survey. *ACM Computing Surveys*, 56(1): 1–37.