REAL-TIME INVERSE KINEMATICS
FROM MOTION CAPTURE WITH NEURAL NETWORKS

[1]Taras Rodzin, [2]Anna Korol and [2]Valeriya Gritsenko

[1]Ukrainian Catholic University, Lviv, Ukraine; [2]WVU School of Medicine, Rockefeller Neuroscience Institute, Morgantown, West Virginia, USA

INTRODUCTION

Human motion analysis is imperative for Computer Vision, Robotics, Prosthetics, and Orthotics; primarily, the **inverse kinematics** (IK) problem lies in the general task of computing the optimal rigid body joint configurations, a.k.a. Euler angles, based on the desired endpoint positions. Usually, Euler angles are used to evaluate and quantify neurological movement deficits during specific activities and become an unbias test in contrast to Fugl-Meyer Assessment, Action Research Arm Test, Wolf Motor Function Test, etc., performed by physical therapists. Besides the decrease in model efficiency of classic numeric IK solutions with increasing degrees of freedom, the subject's anthropometric data also introduces variability. However, **Deep Learning** (DL) may demonstrate a high-quality performance using motion capture (MoCap) markers located on the rigid body during movement. Therefore, we hypothesize that the DL model will produce a real-time generalized IK solution for the wrist using only MoCap markers of the arm. This project aims to extend Zabava's work on designing accurate and fast neural network models and optimizing them using structural and data manipulation techniques. Our results demonstrate two MLP (Figures 1, 2) and one LSTM (Figure 3) models with efficient results both in prediction accuracy and in time performance.

MODELS

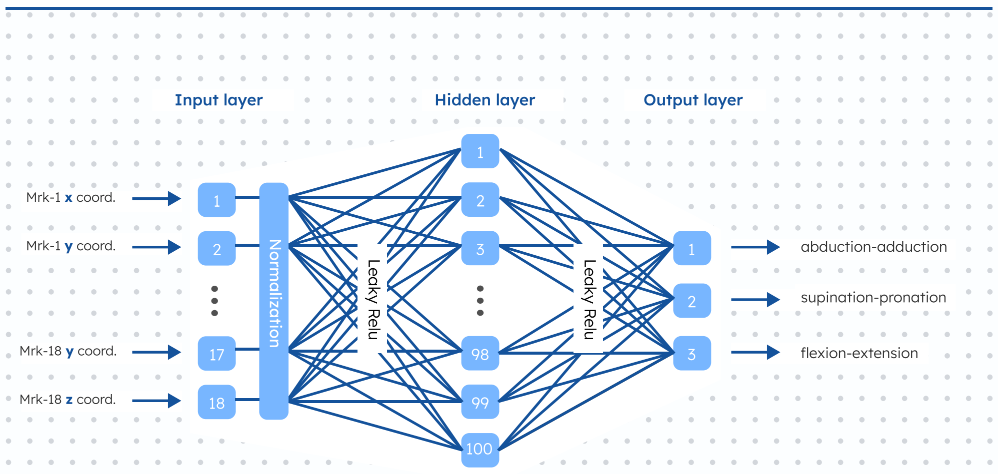


Figure 1: The two-hidden-layer MLP consists of an input layer of size 90: the coordinates of five ms are concatenated in one row and the output will correspond to the angle on the fifth ms; MLP has a normalization layer, the first hidden layer of size 60, the second one of size 24 and the output layer of size 3. The model uses Leaky ReLU activation function between each layer. This architecture allows the model to be noise invariant, although have worse results on the data with no noise.

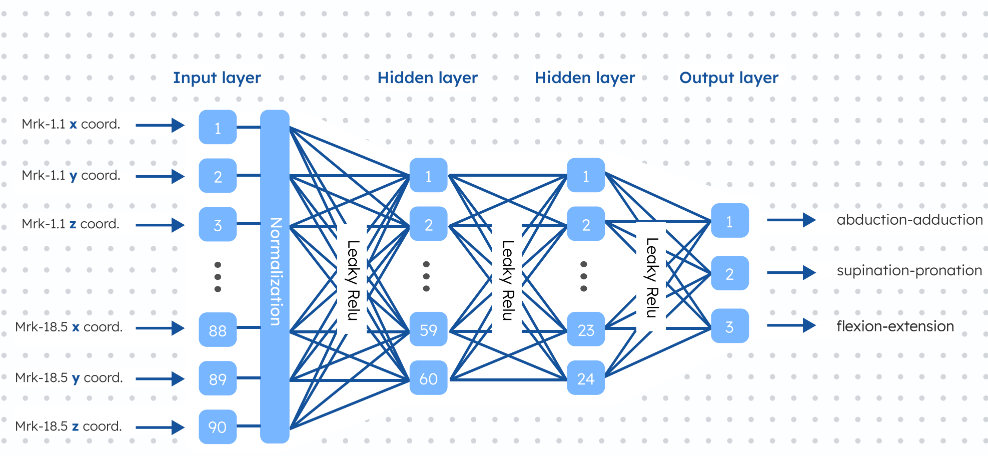


Figure 2: The two-hidden-layer MLP consists of an input layer of size 90: the coordinates of five ms are concatenated in one row and the output will correspond to the angle on the fifth ms; MLPhas a normalization layer, the first hidden layer of size 60, the second one of size 24 and the output layer of size 3. The model uses Leaky ReLU activation function between each layer. This architecture allows the model to be noise invariant, although have worse results on the data with no noise.

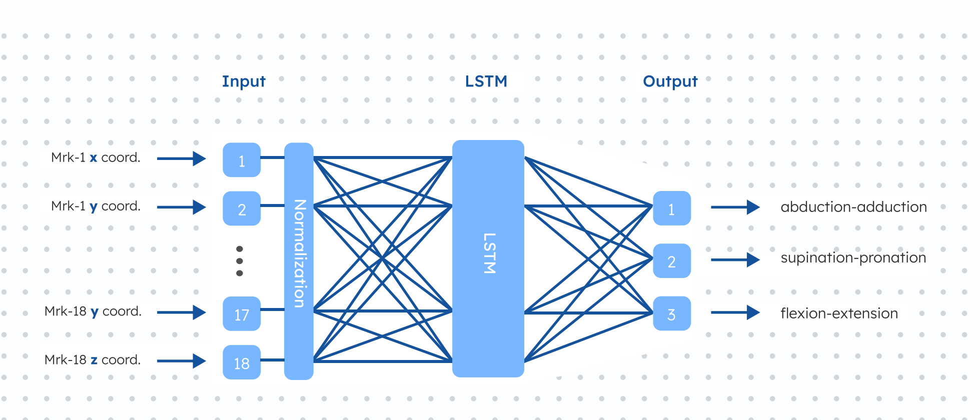


Figure 3: The Long Short Term Memory (LSTM) model consists of an input layer of size 18, hidden layer of size 36 and an output layer of size 3. Each hidden neuron represents a memory cell and hidden state units for saving the results of the previous timesteps and using this history to predict the output. In detail, the cell state is responsible for a long-term memory and the hidden state is responsible for transfer of the results from timestep t-1 to the timestep t, establishing a recurrent dependance. The marker data is preprocessed as 3D array, where the first dimension is the number of samples (dataset size), the second one – the number of timesteps (n=1) and the third one – the number of features (n=18). This architecture demosnstrates the lowest RMSE results among all three models, despite losing in performance time to both MLP's.

METHODS

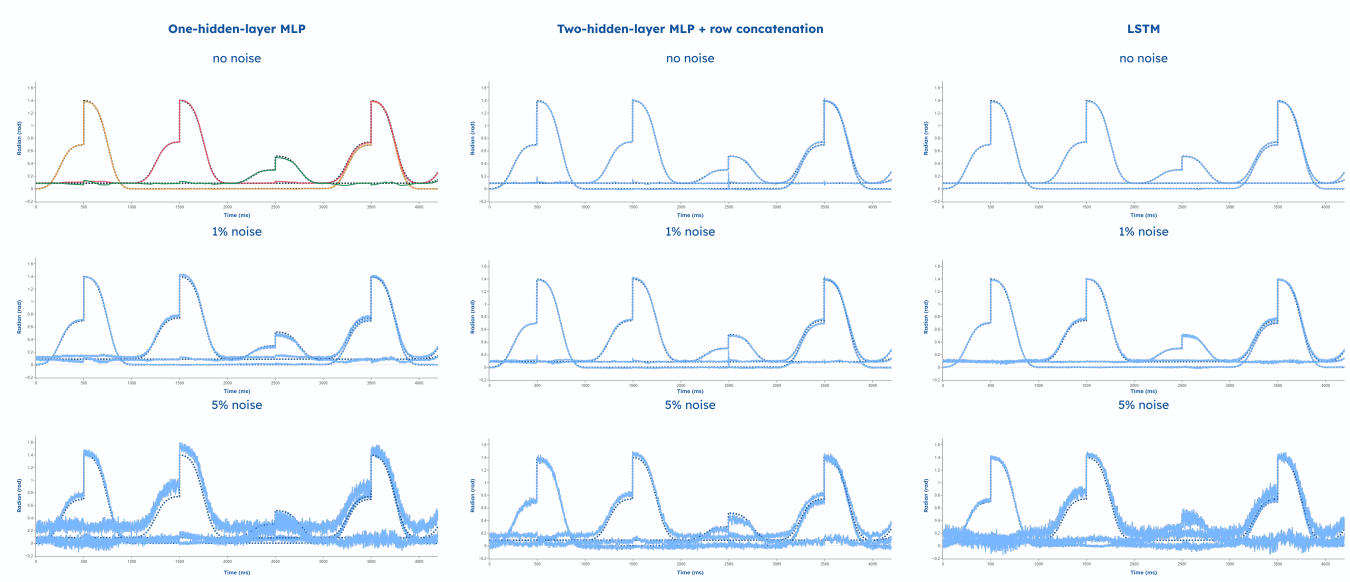
Using the PyTorch library, the model architectures (i.e., MLP on Figures 1,2 and LSTM on Figure 3) with optimal hyperparameters, optimization and regularization functions were determined. The input data consist of coordinates of MoCap markers placed on the humerus, radial, and metacarpal joints. The output layer (n=3) returns three degrees of freedom for the wrist: abduction-adduction, supination-pronation, and flexion-extension. Twenty-one-second-length movement of an arm is used to train each model; the results were validated with 1% and 5% of noise.

All three models are trained on 20000 epochs with a learning rate of 0.01. The batches are not used while training. Instead, the whole training dataset (n=16800) is used each time. The Adam optimizer is used to update models' paramters after each epoch.

Results

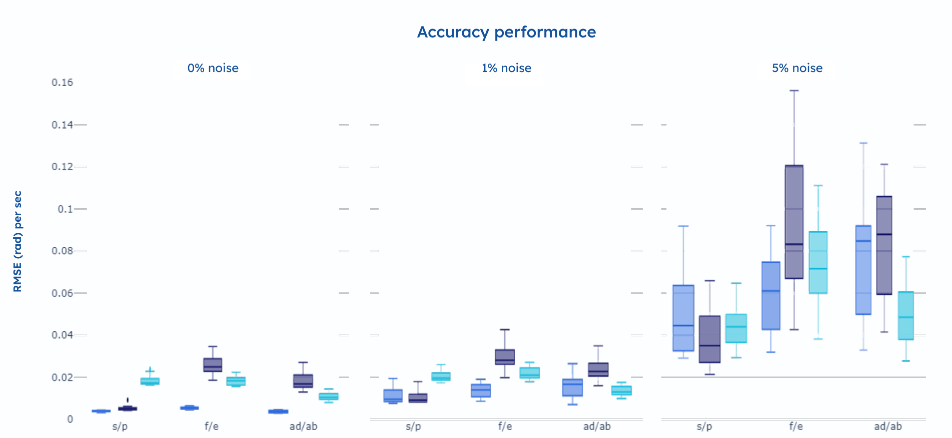
Overall, LSTM model with recurrent history dependence performs with an average RMSE error of 0.004 rad in 0.001 sec per one second (per n=1000 observations) of rigid body movement. Because of the history-dependent input, the two-hidden-layer MLP shows more accurate results with the increase of noise, outperforming LSTM on 5% noise data. The accuracy performance of LSTM and one-hidden-layer MLP models is more sensitive to the noise. The LSTM results demonstrate the lowest RMSE: twice better on data with no noise and with 1% noise, although its time performance starts increasing with decreasing signal/noise ratio.

Despite the accuracy drop, one-hidden-layer MLP shows the lowest time performance out of all three models.



Predicted data
Real data
Supination-pronation
Abduction-adduction
Flexion-extension

Figure 4: An arm movement data sample of 4500 observations (ms) from the total 21000-length datasets is plotted. The plots demonstrate overlapping between real (dotted) vs. predicted (solid) angles in time with 0%, 1% and 5% of noise.



LSTM
One-hidden-layer MLP
Two-hidden-layer MLP

Figure 5: The error distribution per one sec of rigid body movement for each model (0%, 1% and 5% noise) during twenty train-test iterations. One-hidden-layer MLP is the most noise-sensitive model; the two-layer-mlp shows relatively worse performance on a data with no noise, but efficiently generalizes 1% and 5% noise datasets. The LSTM model demonstrates the the lowest error with small influence of noise onto its variation.

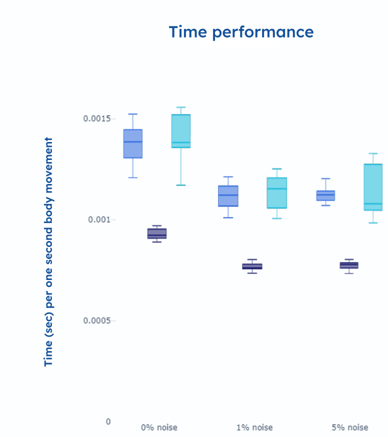


Figure 6: Time performance distribution per one sec of rigid body movement of three models after twenty iterations. One-hidden-layer MLP has the quickest performance, while LSTM and two-hidden-layer MLP stay on the same level.

Conclusion

This work encompasses two goals: (1) advancing Zabava's models and (2) constructing architectures based on predicted observations. By introducing LSTM, we aimed to balance both time dependence and accuracy, and future steps would be to explore other recurrent neural networks, e.g., gated recurrent unit model, to increase the time performance. Moreover, the IK problem can be splitted into stages and represented as the combination of several models with different architectures and hyperparameters.

We succeeded in applying data pre-processing methodology, i.e., smoothing, increasing model generalizability both in one-hidden-layer MLP and in LSTM models on the noisy data. The findings show robust results that can be applied on the real-world recordings.

Despite high model performance, the trade-off between time execution and predicted accuracy has to be considered when applied to the specific applications (see **Results** section).

The research has been sponsored by the **West Virginia University Center for Foundational Neuroscience Research & Education** in conjunction with the **West Virginia University Department of Neuroscience and Rockefeller Neuroscience Institute.**