Vision Based Prediction of ICU Mobility Care Activities Using Recurrent Neural Networks

Gabriel Bianconi*[1] Rishab Mehra*[1] Serena Yeung[1] Francesca Salipur[1] Jeffrey Jopling[2] Lance Downing[2]
Albert Haque[1] Alexandre Alahi[1,3] Brandi Campbell[4] Kayla Deru[4] William Beninati[4] Arnold Milstein[2] Li Fei-Fei[1]

[1] Stanford AI Lab [2] Stanford Medicine
[3] École Polytechnique Fédérale de Lausanne
[4] Intermountain Healthcare

INTRODUCTION

Intensive Care Units are amongst the highest density areas of patient care activities in hospitals, yet documentation and understanding of these activities remains sub-optimal.

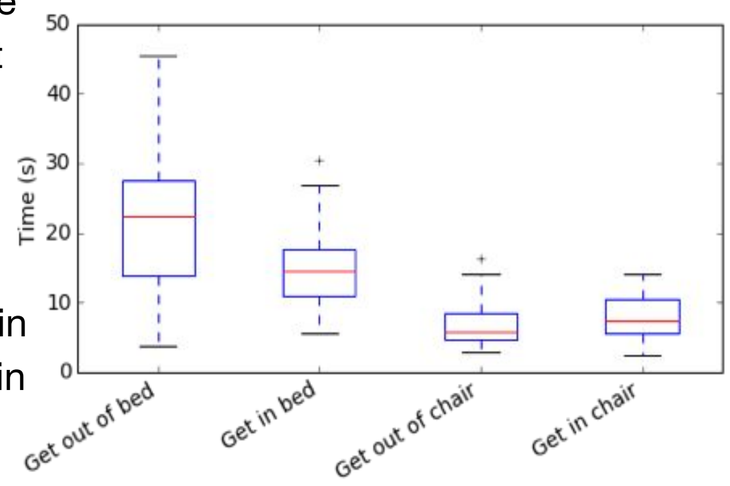
Understanding the occurrence of these activities at a per-patient level is important for ensuring adherence to protocols, and for studying the correlation between the adherence, and patient outcomes.

DATA

We set up a simulation room with one depth sensor in an ICU ward room at LDS Hospital, Salt Lake City, Utah.

We collected a dataset of nurse-led simulations for 4 patient mobility activities: getting out of bed, getting in bed, getting out of chair and getting in chair.

In total we collected 45 minutes of data, leading to a total of 9,755 frames of depth data.



METHOD

Data Split
We split our simulation data by consecutive time into 21 minutes (4,617 frames) for training, 8 minutes (1,763 frames) for validation and 18 minutes (3,375 frames) for testing. We combined the training and validation set for the final model.

Finetuned Resnet-18
We fine-tuned a Resnet-18 pre-trained on Imagenet to perform one vs. all single frame classification for each of the activities.

Resnet-18 + LSTM
Instead of fine tuning the Resnet-18, we extracted features from its last layer, and inputted them into an LSTM, a type of Recurrent Neural Network.

We inputted features of 16 frame sequences into the LSTM with the supervision being the activity label at the final frame. We trained the LSTM using a binary cross entropy loss and the Adam optimizer.

We experimented with using one vs all LSTM classifiers for each of the activities, and a single multiclass LSTM classifier. The one vs. all LSTM classifiers outperformed the multiclass LSTM.

Results

Activity	CNN	LSTM-CNN
Getting out of bed	62.1	**72.2**
Getting in bed	24.2	**53.5**
Getting out of chair	8.3	**44.8**
Getting in chair	9.2	**39.4**
Mean AP	25.5	**52.5**

As seen in the correct predictions on the right, the model is able to successfully distinguish between getting in and out of bed using temporal context. However, as seen in the incorrect predictions, the model occasionally confuses the two, particularly when the patient struggles, and takes a long time to get out of bed.

Overall, we show that our temporal modeling achieves effective frame-level prediction and produces meaningful activity timelines.