

# Description and recall of the object using processes with scene change using LSTM

○Masaki Yano, Tadashi Matsuo, Nobutaka Shimada (Ritsumeikan Univ.)

## 1. Motivation

- There are many kinds of objects which we usually use.
- It is very hard to develop softwares which make the robots manipulate any objects.



We aim that a robot learns a human's action automatically.

### Our Study:

We propose a method that describes and recalls co-occurrence of a human action and a scene change when a person uses an object.

### This Poster:

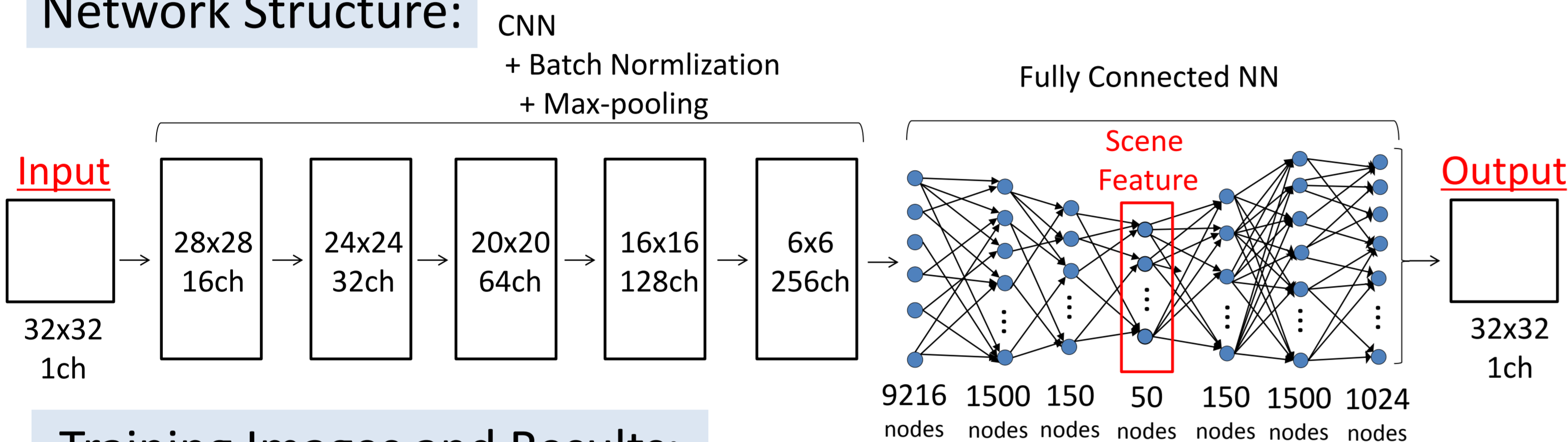
We describe and recall them when a person uses a chair.

## 2. Describing the Scene Feature by Sparse Auto-Encoder

We use Sparse Auto-Encoder[1] for describing the scene depth image as the low dimensional vector.

[1] Tadashi Matsuo, Nobutaka Shimada "Construction of Latent Descriptor Space of Hand-Object Interaction", The 22nd Joint Workshop on Frontiers of Computer Vision (FCV2016): pp. 117-122

### Network Structure:



### Training Images and Results:

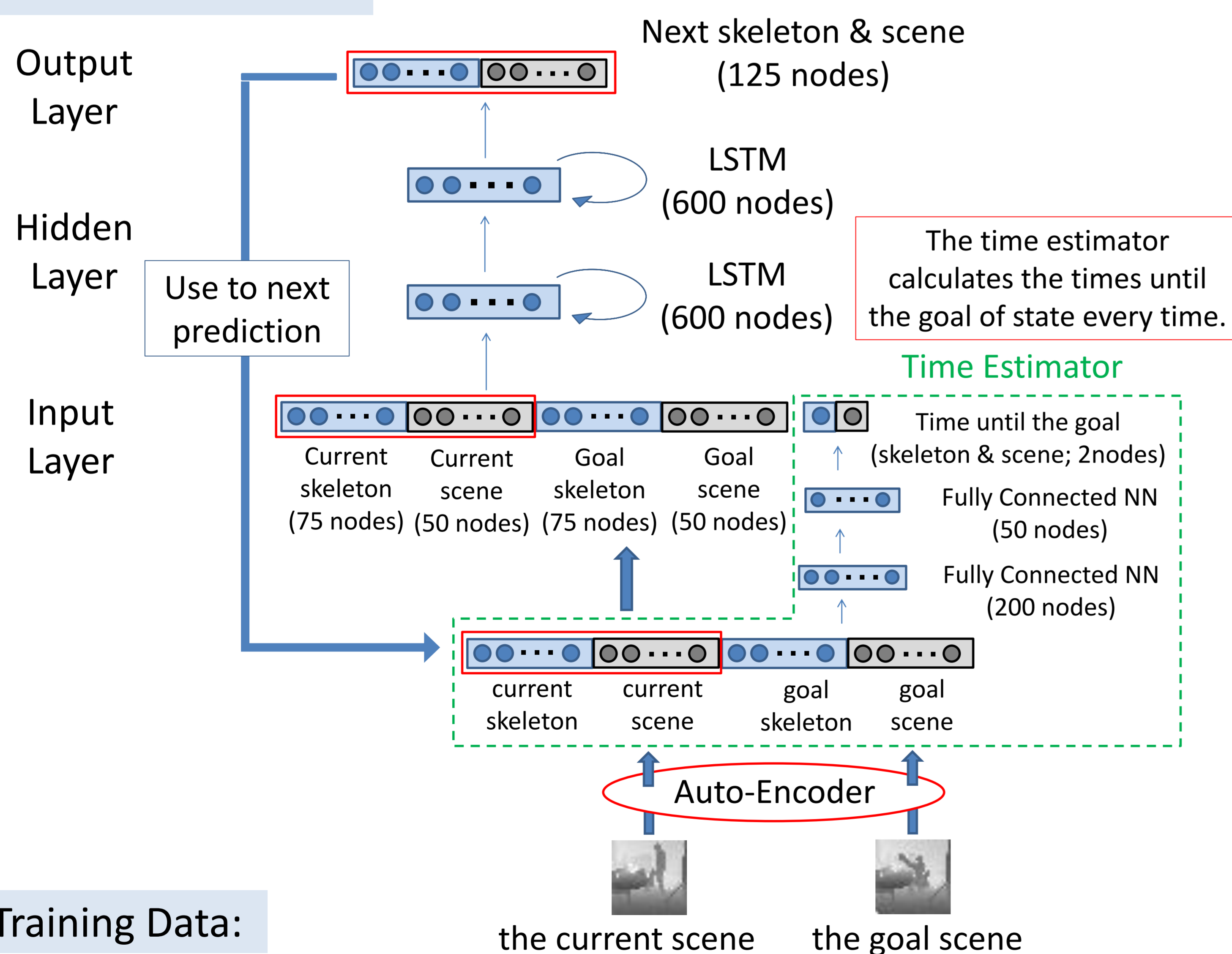


## 3. Describing the Human Action and the Scene Change

We train a LSTM(Long Short-Term Memory) model from the human posture features and the scene features.

➡ The model can recall the human action and scene change which corresponds to the current state.

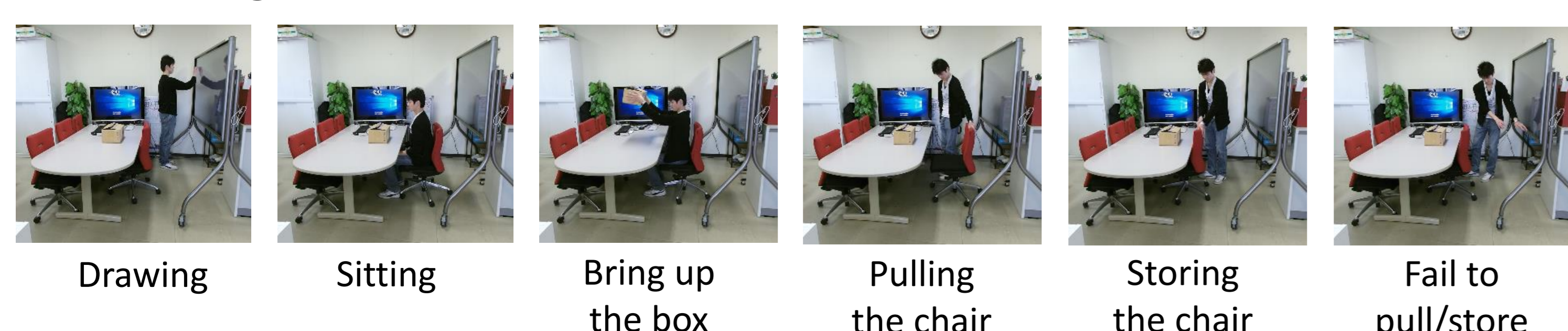
### Network Structure:



### Training Data:

the current scene the goal scene

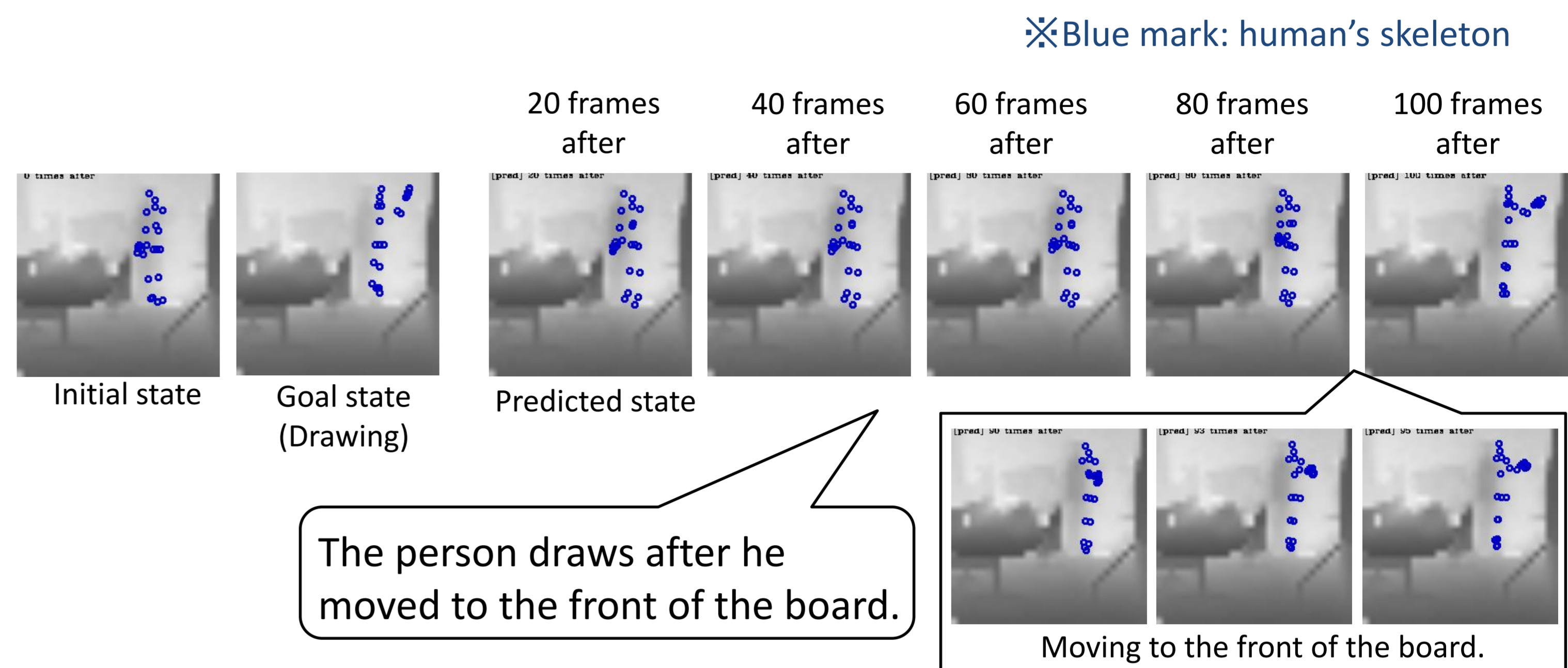
The training data includes the six kinds of human action.



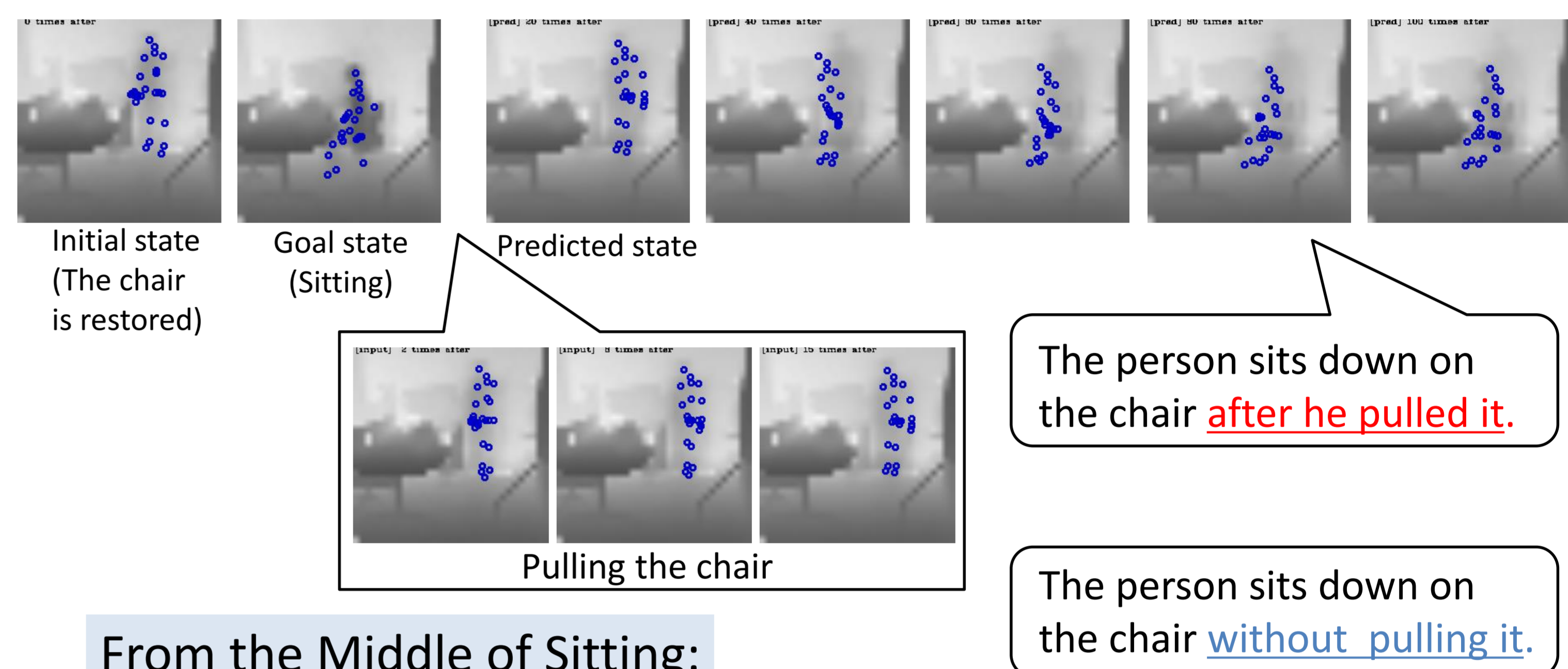
## 4. Recalling the Human Action and the Scene Change

### Results of Recall:

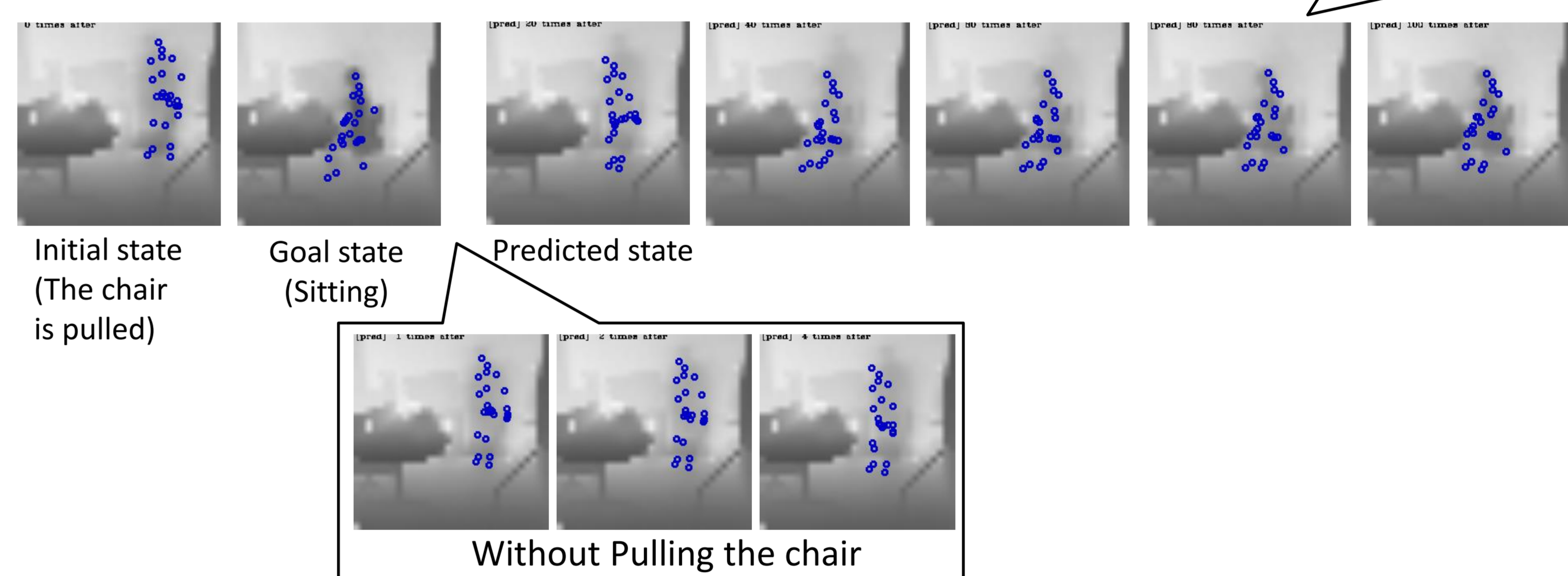
Results of the recall from the initial state until the goal state.



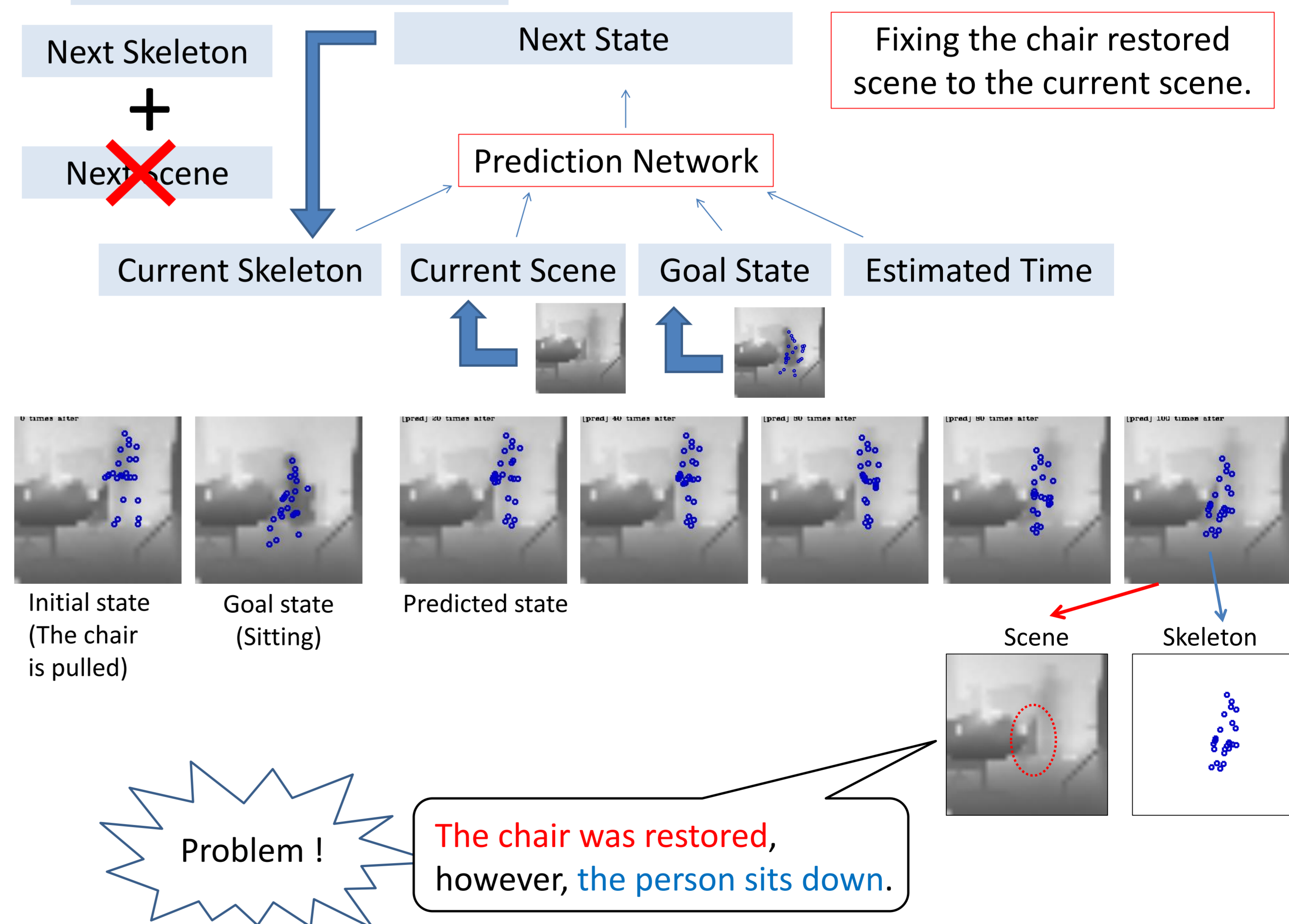
### Another Goal State:



### From the Middle of Sitting:



### Fail to Pull the Chair:



## 5. Conclusion and Future Work

- We built the LSTM model and trained it with the human skeletons and the scene features.
- In the experiment, the model recalled them the same order as actual.
- We are extending to allow our proposed method to work even if the human fails to change the scene state.

## Acknowledgement

This work was supported by JSPS KAKENHI Grant Numbers JP24500224, JP15H02764.