Supplementary Material: Dynamic Neural Relational Inference

Colin Graber Alexander Schwing
University of Illinois at Urbana-Champaign
{cgraber2, aschwing}@illinois.edu

This supplementary material contains the following information: first, we describe implementation details, including description of experimental setup and additional model information. Next, we provide extra visualizations of results for the motion capture data. Finally, we visualize extra instances from the basketball data. For both visualizations we also provide additional quantitative assessment.

1. Implementation Details

We first begin by describing details common to all experiments. All models were trained using the Adam optimizer. To train the decoder for the static NRI model, it was provided with ground-truth features for the first 10 time steps. Beyond the 10-th step, model predictions were used as input. This setup follows work by Kipf *et al*. [19], who report that this form of training leads to more accurate results. We were able to confirm this independently. As mentioned in Sec. 3 of the main paper, this strategy led to poor performance for dNRI though. Instead we found it to be beneficial if the decoder is always provided with ground-truth inputs during training. To match the training setup of dNRI, we also provided the FCGraph, SingleLSTM, and JointLSTM models with ground-truth inputs for all steps during training.

After every epoch, losses on the validation data were calculated by every model. We report results on the model having the largest conditional data likelihood on the validation data. All training data were normalized to lie between -1 and 1 using the formula $x' = 2\frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} - 1$, where x_{max} is the maximum value for an individual coordinate of all the training features and x_{min} is the minimum value for an individual coordinate for all the training features. Location and velocity are normalized independently of each other. The same normalization formula was also used on the validation/test data using the values of x_{min} and x_{max} from the training data. All reported errors are computed on this normalized data. Every model was trained 5 times using different random initializations. The error plots display mean errors, and the shaded areas represent standard deviations. All models are implemented using PyTorch version 1.2. The code implementing the models and running the training/evaluation procedures is available at https://github.com/cgraber/cvpr_dNRI.

Synthetic Physics Simulation Experiments: Each trajectory consists of 50 simulated time steps of motion. Train, validation, and test sets each consisted of 100 trajectories. All models were trained for 400 epochs with a batch size of 16. The starting learning rate was 0.0005, and the learning rate was decayed by 0.5 after 200 epochs.

Motion Capture Experiments: All models were trained for 600 epochs with a batch size of 8. The starting learning rate was 0.0005, and the learning rate was decayed by 0.5 after 300 epochs.

For subject #35, we use the same data splits as Kipf *et al*. [19]. This consists of the 23 trials labeled as 'Walk,' with 12 trials used for training, 4 trials used for validation, and 7 trials used for testing. For training/validation data, each trial was split into sequences of length 50. For test data, each trial was split into sequences of length 100.

For subject #118, the first 30 trials (those labeled as 'Jump') were utilized, with the first 14 used for training, the next 7 used for validation, and the remaining 9 used for testing. For this subject, we left the validation and test sequences whole. For training, we extracted sequences of length 50 from the training trials; during every epoch, a random starting point within each 50-step sequence was selected, and the following 50 steps were extracted as a single sequence. Velocities for this data were approximated for each time step by using the difference between the current location and the location for the next time step.

Basketball Experiments: All models were trained for 100 epochs with a batch size of 128. A learning rate of 0.0005 was used for the entirety of training. The train/validation/test data consisted of 79,456/27,690/13,845 trajectories, respectively.

Traffic Trajectory Experiments: All models were trained for 400 epochs with a batch size of 8. The starting learning rate was 0.0005, and the learning rate was decayed by 0.5 after 200 epochs.

2. Additional Motion Capture Visualizations

We present additional visualizations for Motion Capture subjects: Figs. 1, 3, 5, 7, and 9 display location predictions for subject #35, Figs. 2, 4, 6, 8, and 10 display sample dNRI relation predictions for subject #35, Figs. 11, 13, and 15 display location predictions for subject #118, and Figs. 12, 14, and 16 display sample dNRI relation predictions for subject #118. All models illustrated here use 4 predicted relation types, with one of them being hard-coded as 'no relation.' The red, solid skeleton represents the ground-truth joint locations, and the blue, dotted skeleton represents model predictions. Below each frame, we display the mean squared error of the unnormalized 3D position for that frame's predictions.

For the figures displaying sample edge predictions, we also illustrate a subset of the predicted relations connected to the left heel, right heel, left wrist, and right wrist. The provided relations are those that change across these three frames and are chosen to highlight the dynamic predicted relations. The image in the bottom row depicts all predicted relations between the 31 joints. Each color represents a different relation type, with green representing the hard-coded 'no relation' type. For subject #118, each displayed frame is predicted 20 time steps after the most recent ground-truth was provided. We observe that, across both datasets, the dNRI predictions are more accurate than the other approaches in most cases. Additionally, dNRI predicts different relation types during different "phases" of activity. This is most apparent for subject #118, where the dominant relation type changes from blue to red when the jump begins.

3. Additional Basketball Visualizations

We present additional sample trajectories in Figs. 17 – **??** and relation predictions in Figs. 20 – **??** on 20 test trajectories from the basketball dataset. In every case, the first 40 frames are provided to the models (transparent), and the models are tasked with predicting the final 9 frames (solid). Below each prediction, we display the mean squared error of the unnormalized 2D position for the prediction at the final time step. Relation predictions are displayed for frames 3 and 45, representing predictions near the beginning and near the end of the trajectories. The pink lines represent relations which are predicted to be present in a given frame, but which are not predicted at other time steps. In every case, the dNRI predictions have lower error than those made by the other models. In many cases, dNRI predicts more relations which change over time than the other models – this allows the dNRI decoder to better adapt to the changes in entity states through selecting different edge models, giving it more flexibility than the other decoders.

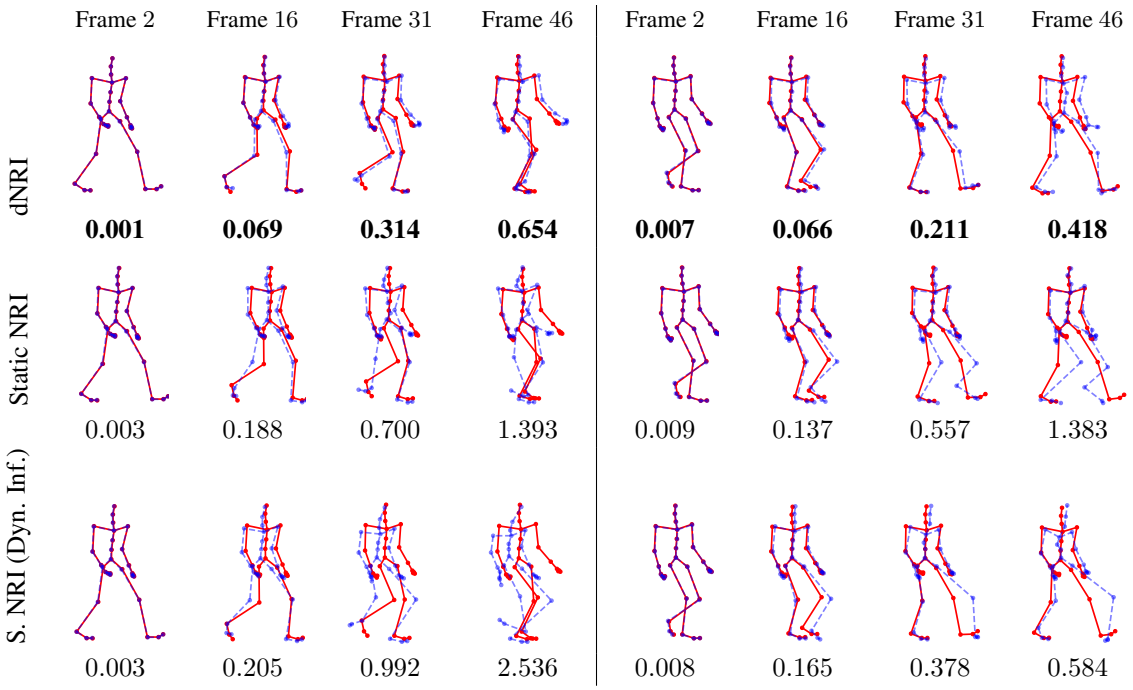

Figure 1: Predicted motion for additional test trajectories #1 (left) and #2 (right) for subject #35. Displayed numbers are mean squared error of the unnormalized 3D positions for that frame's predictions.

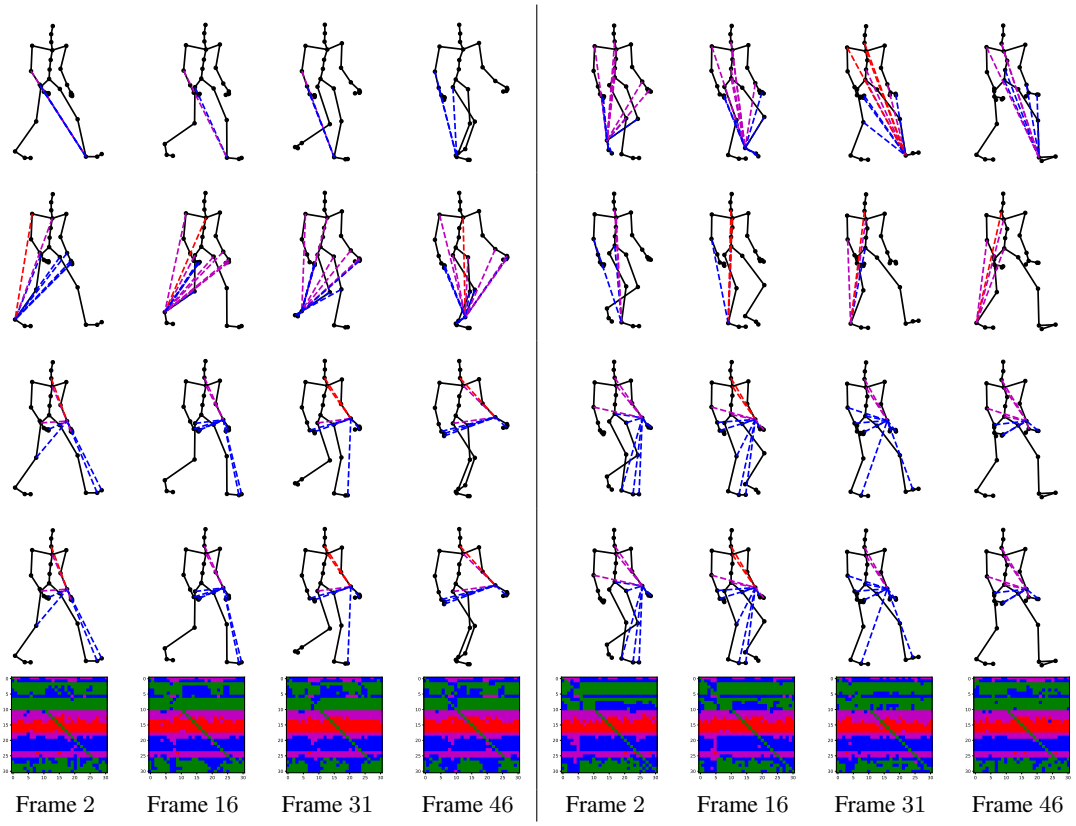

Figure 2: Predicted relations for additional test trajectories #1 (left) and #2 (right) for subject #35.

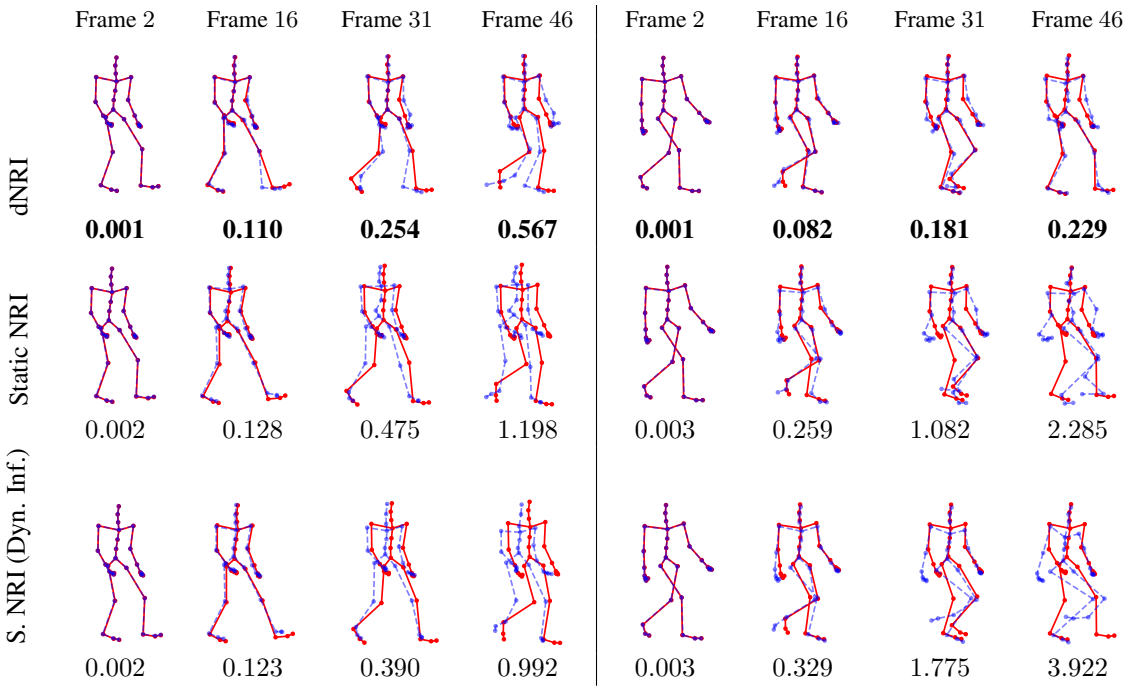

Figure 3: Predicted motion for additional test trajectories #3 (left) and #4 (right) for subject #35. Displayed numbers are mean squared error of the unnormalized 3D positions for that frame's predictions.

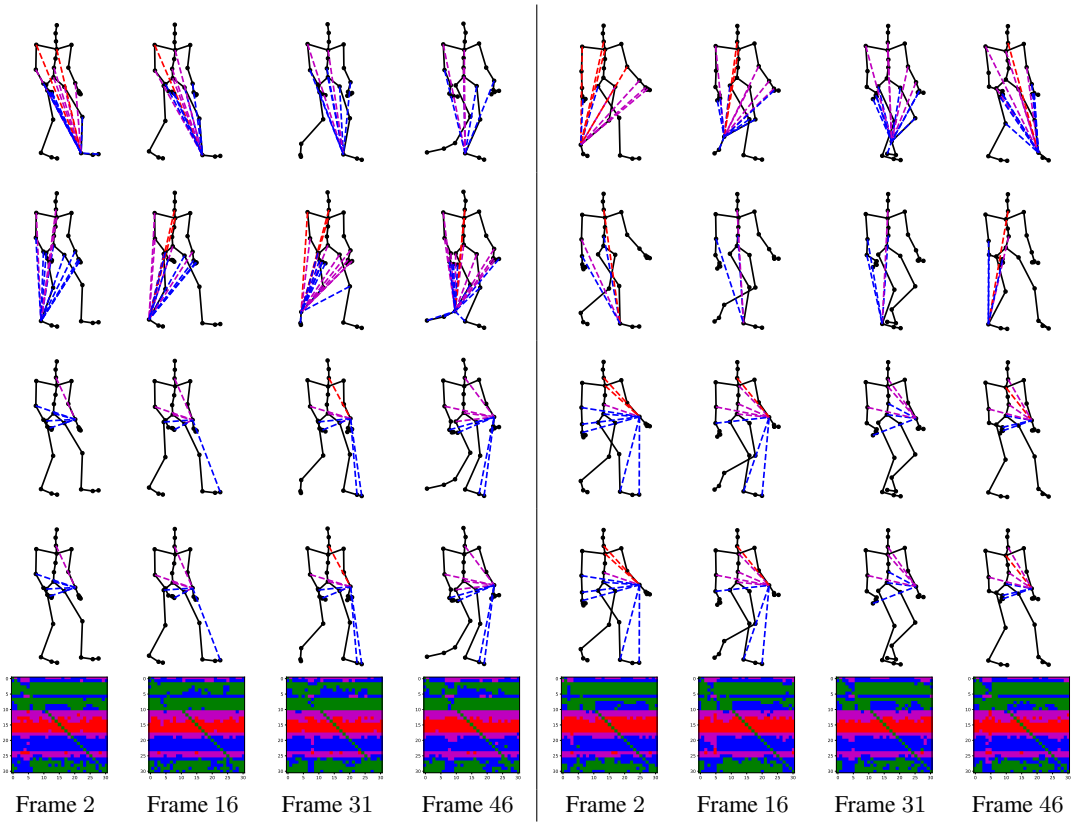

Figure 4: Predicted relations for additional test trajectories #3 (left) and #4 (right) for subject #35.

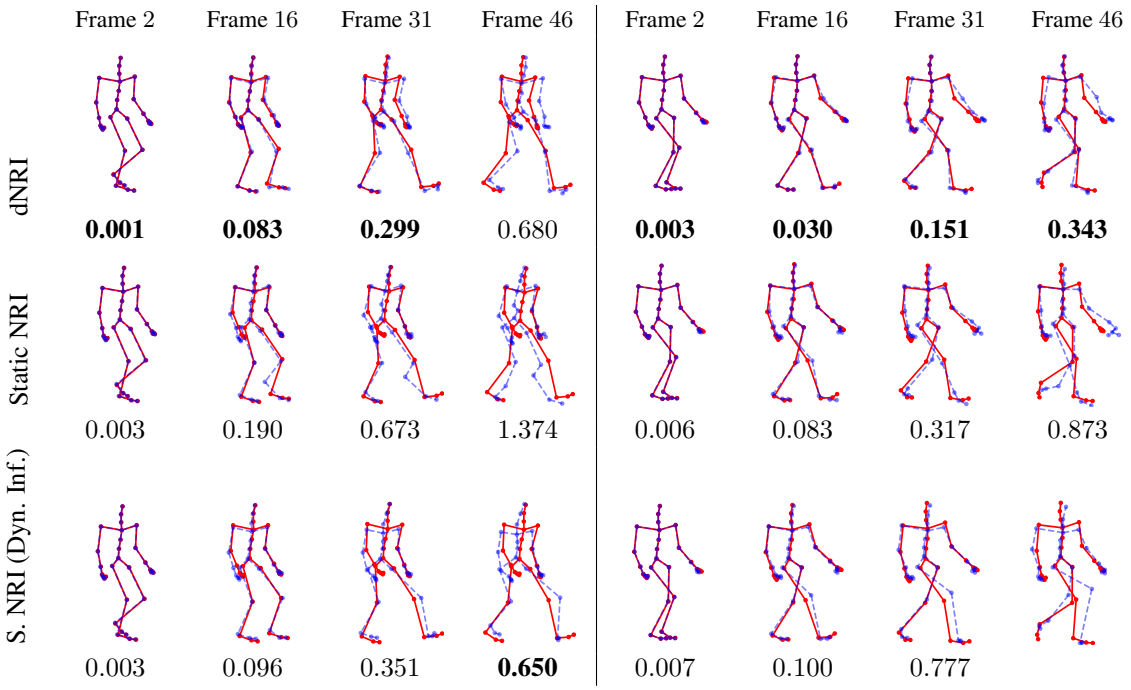

Figure 5: Predicted motion for additional test trajectories #5 (left) and #6 (right) for subject #35. Displayed numbers are mean squared error of the unnormalized 3D positions for that frame's predictions.

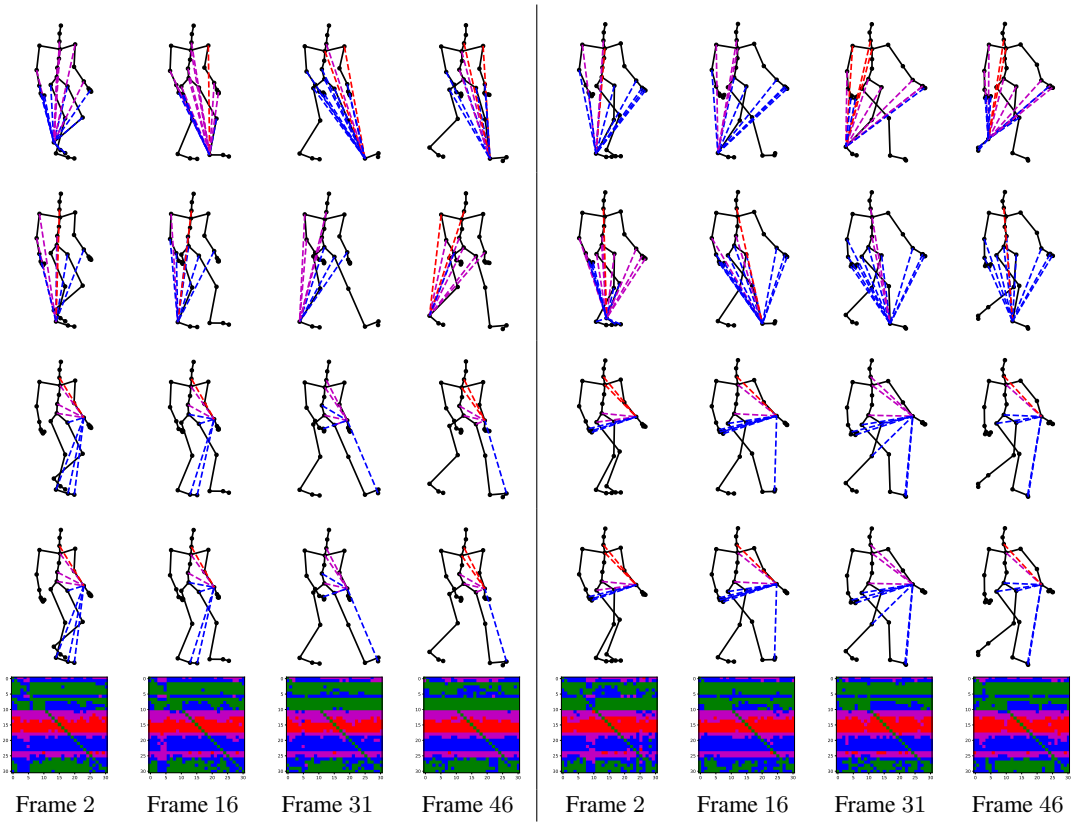

Figure 6: Predicted relations for additional test trajectories #5 (left) and #6 (right) for subject #35.

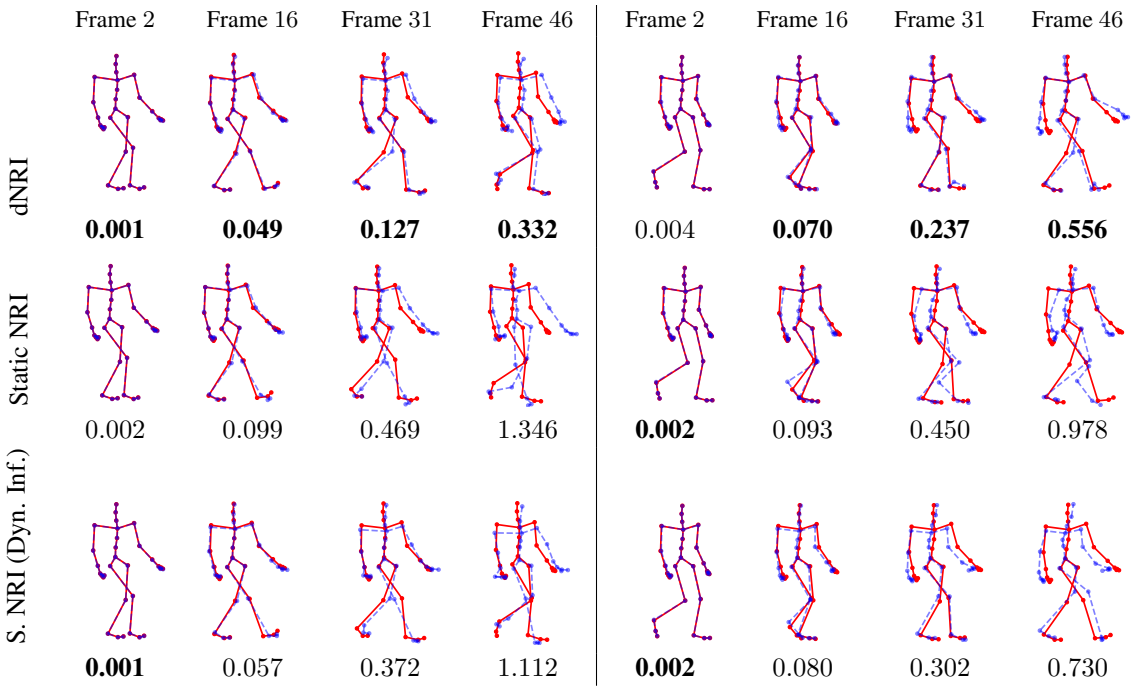

Figure 7: Predicted motion for additional test trajectories #7 (left) and #8 (right) for subject #35. Displayed numbers are mean squared error of the unnormalized 3D positions for that frame's predictions.

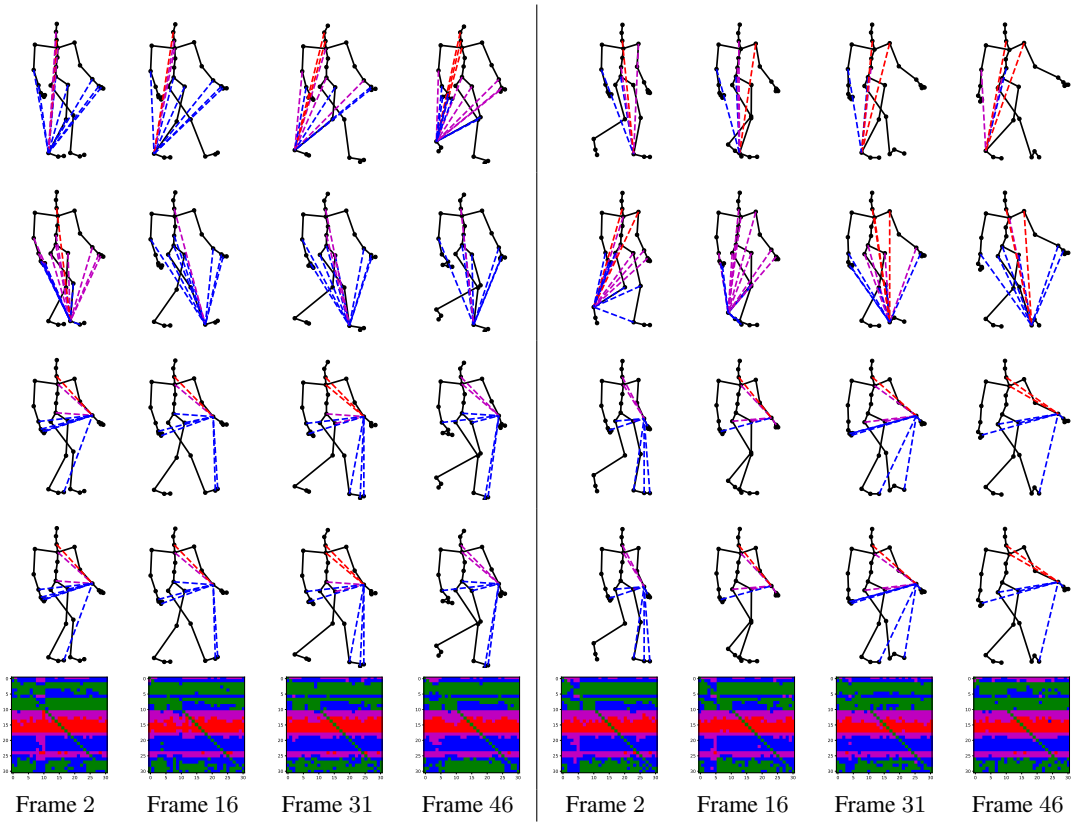

Figure 8: Predicted relations for additional test trajectories #7 (left) and #8 (right) for subject #35.

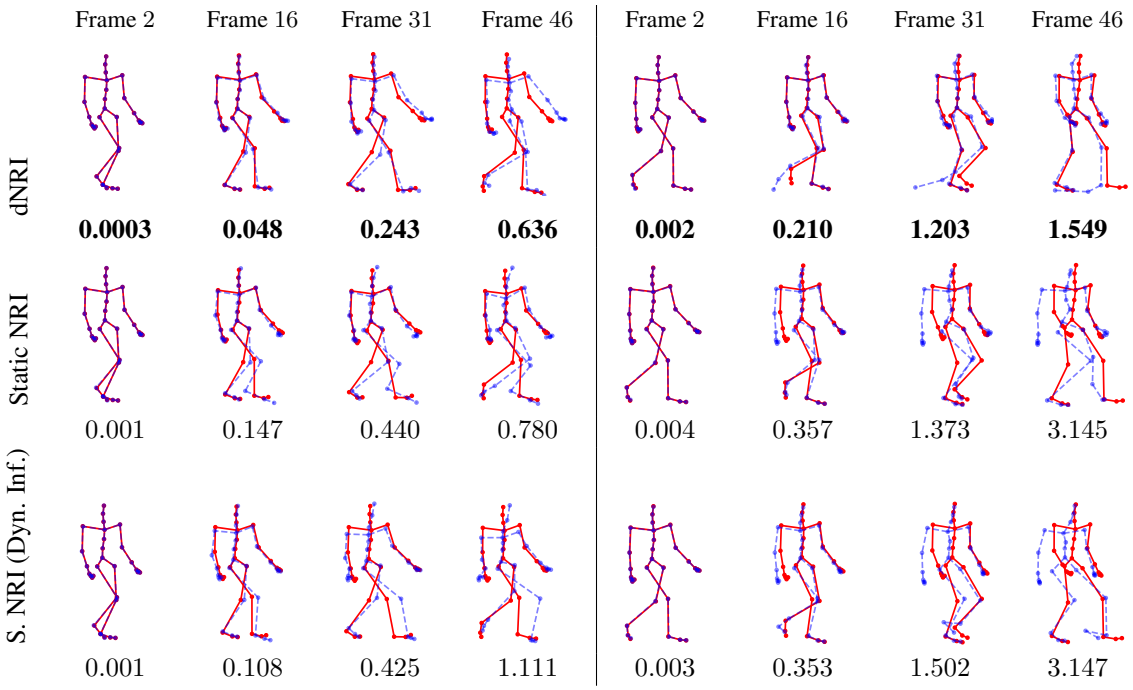

Figure 9: Predicted motion for additional test trajectories #9 (left) and #10 (right) for subject #35. Displayed numbers are mean squared error of the unnormalized 3D positions for that frame's predictions.

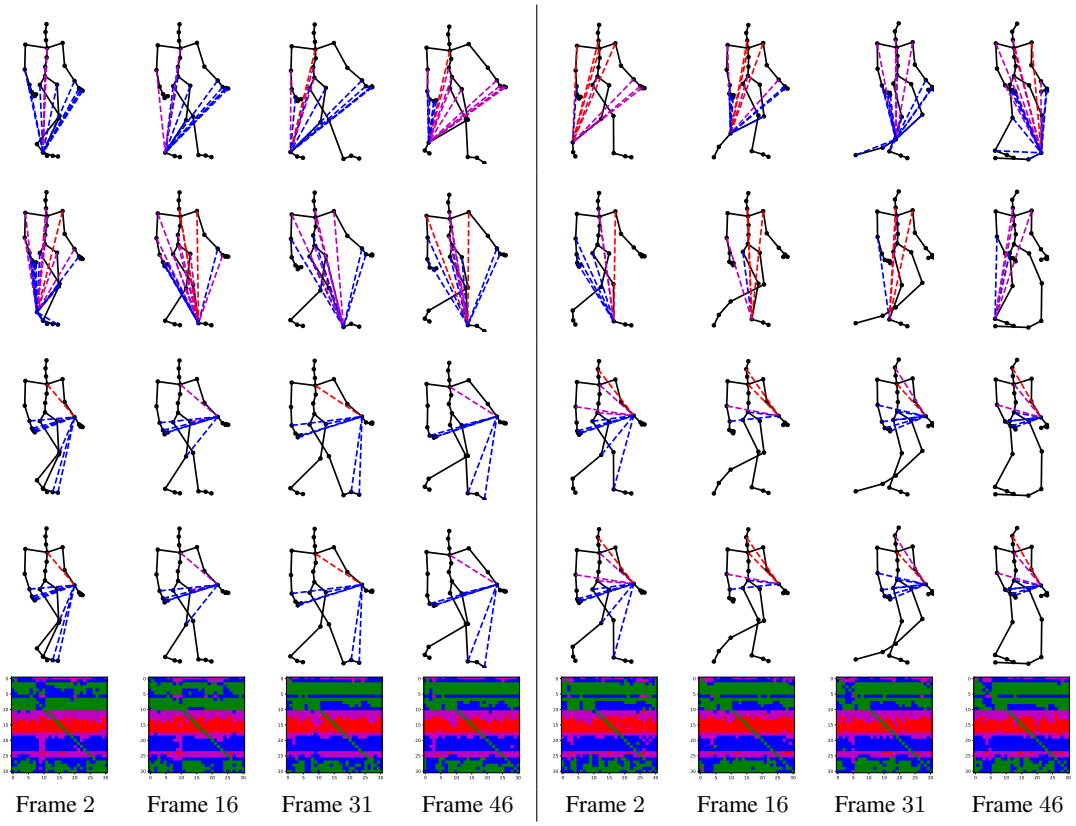

Figure 10: Predicted relations for additional test trajectories #9 (left) and #10 (right) for subject #35.

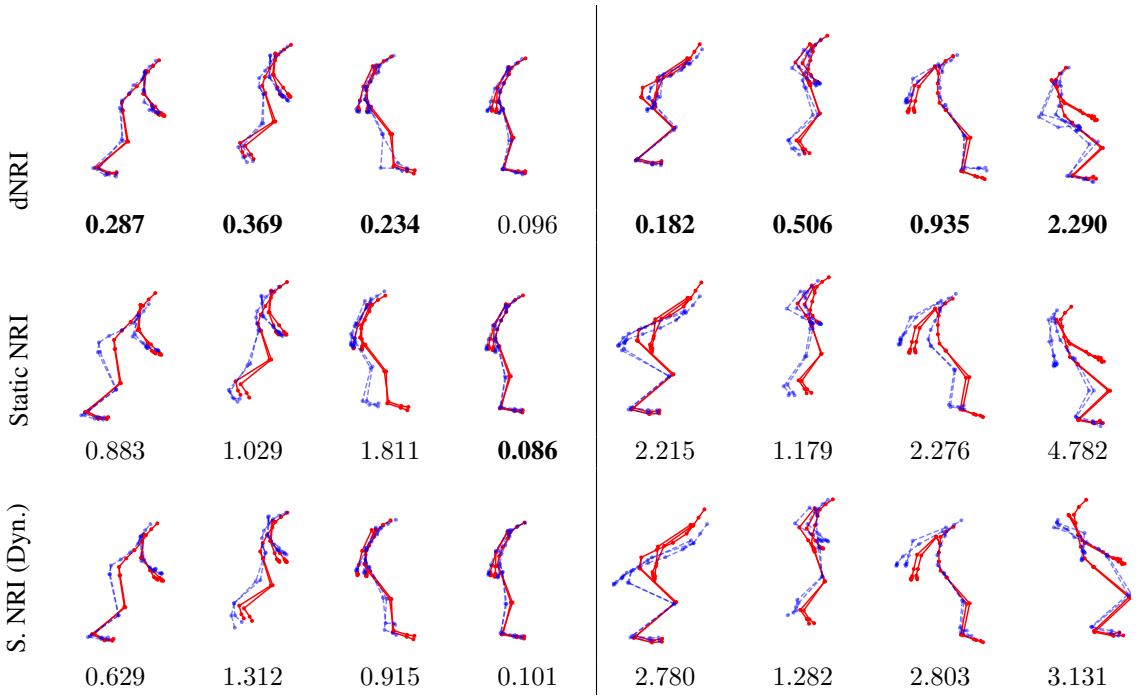

Figure 11: Predicted motion for additional test trajectories #1 (left) and #2 (right) for subject #118. Displayed numbers are mean squared error of the unnormalized 3D positions for that frame's predictions.

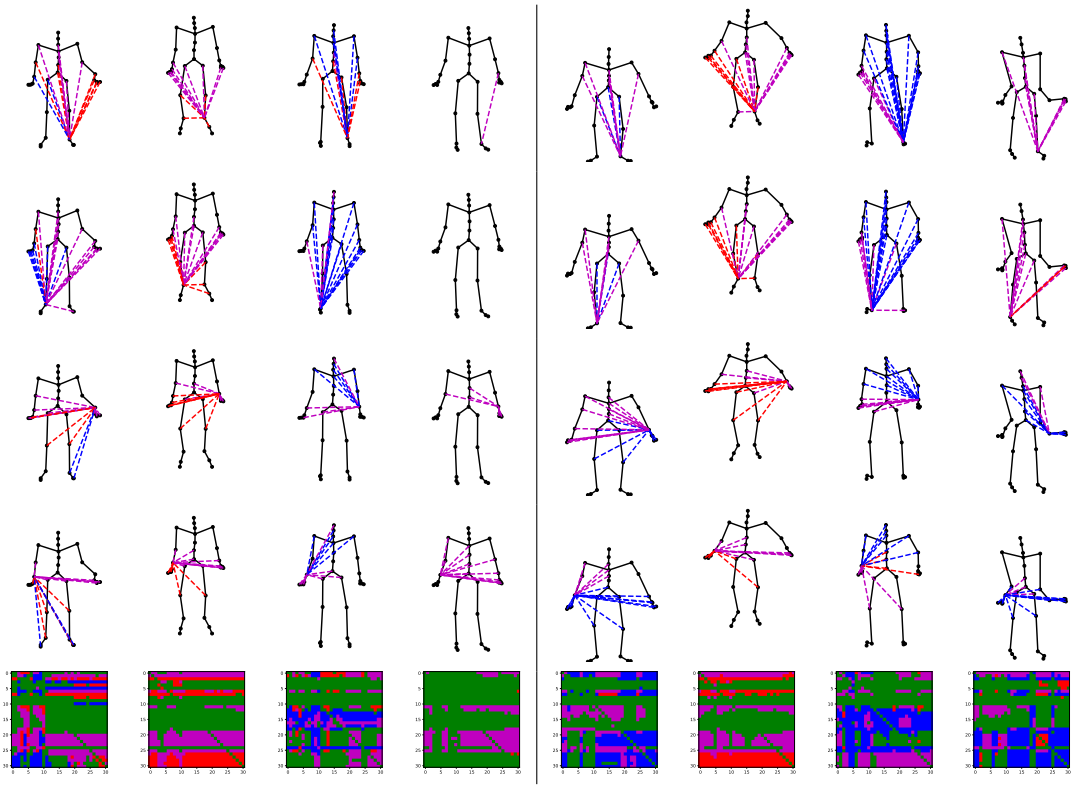

Figure 12: Predicted relations for additional test trajectories #1 (left) and #2 (right) for subject #118.

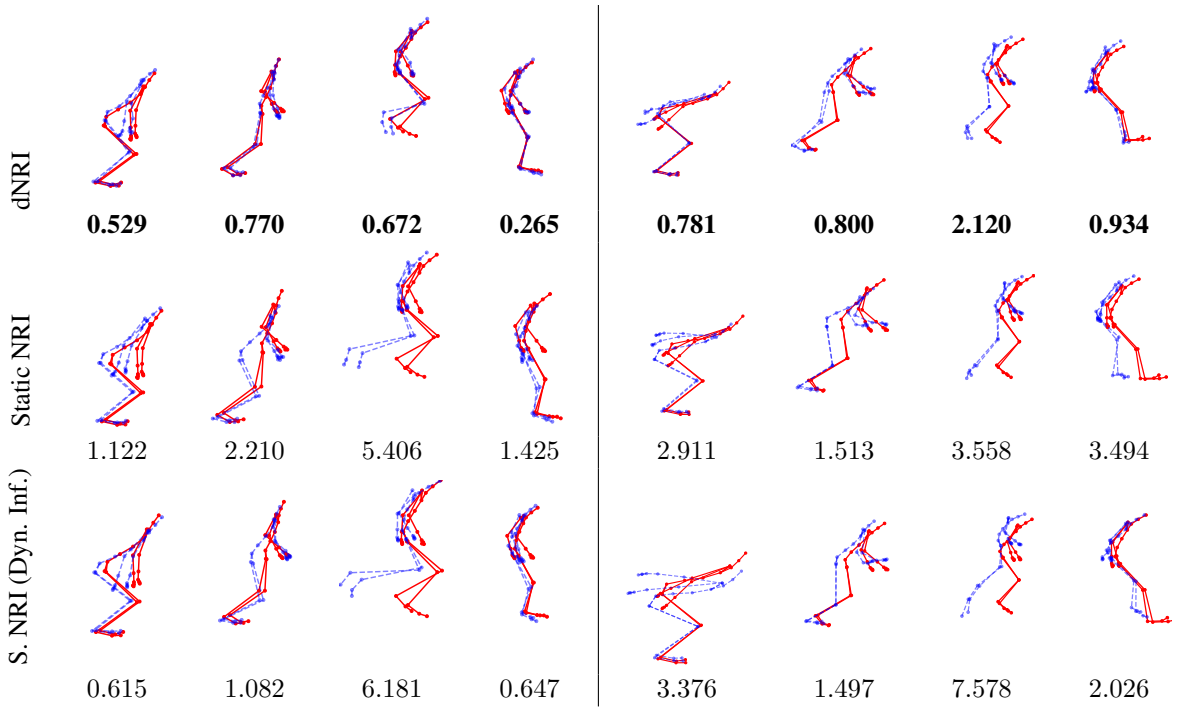

Figure 13: Predicted motion for additional test trajectories #3 (left) and #4 (right) for subject #118. Displayed numbers are mean squared error of the unnormalized 3D positions for that frame's predictions.

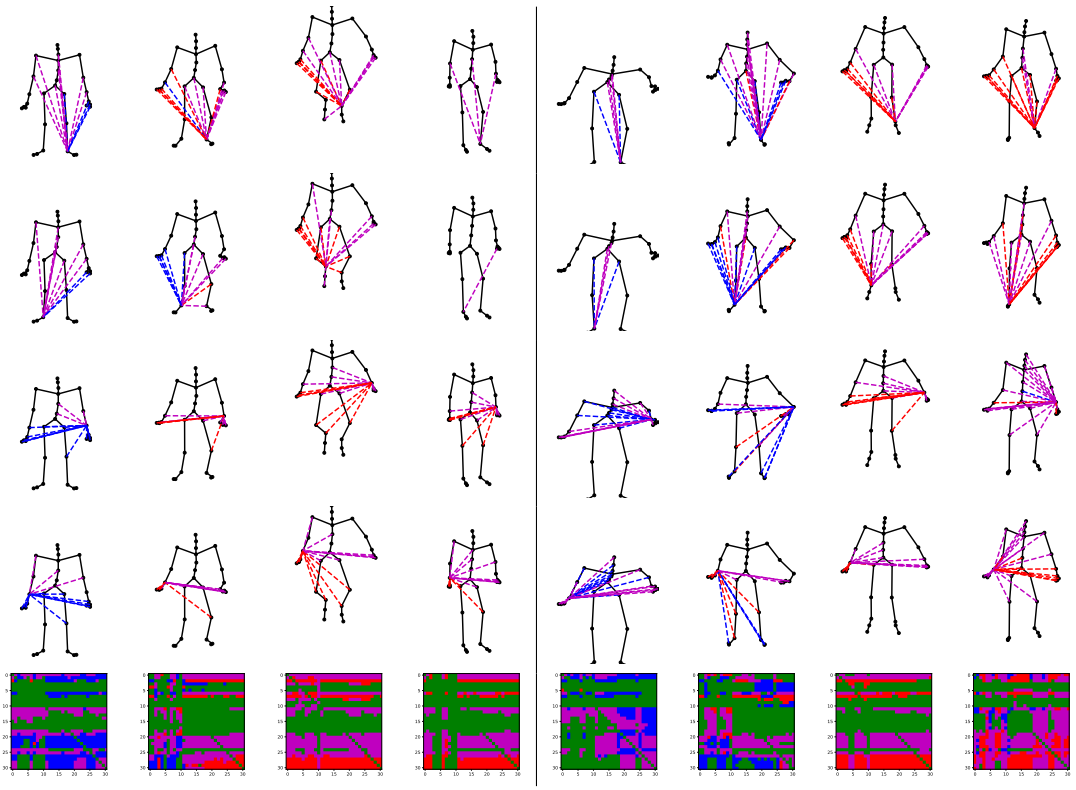

Figure 14: Predicted relations for additional test trajectories #3 (left) and #4 (right) for subject #118.

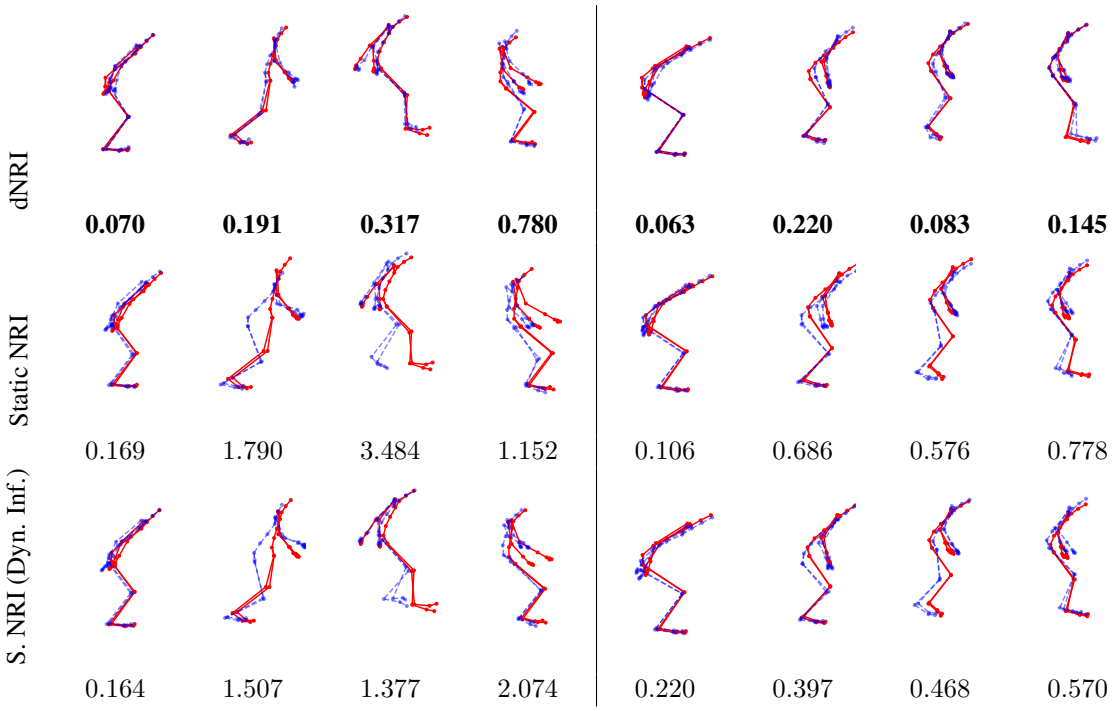

Figure 15: Predicted motion for additional test trajectories #5 (left) and #6 (right) for subject #118. Displayed numbers are mean squared error of the unnormalized 3D positions for that frame's predictions.

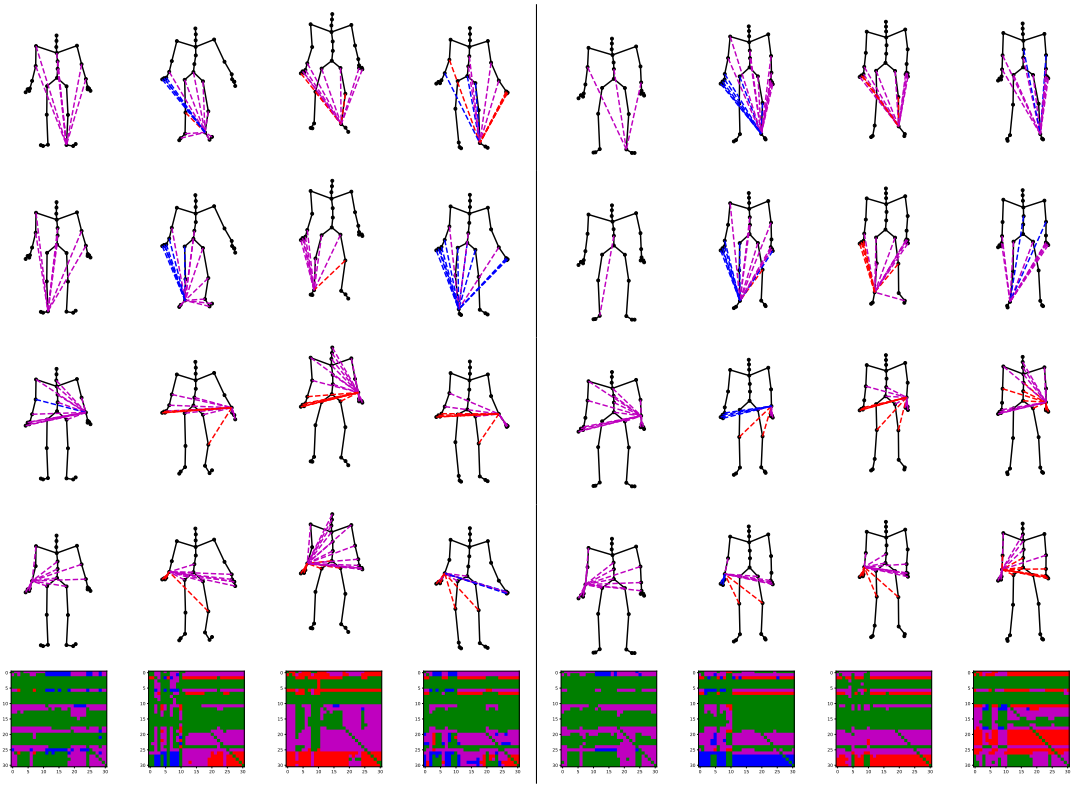

Figure 16: Predicted relations for additional test trajectories #5 (left) and #6 (right) for subject #118.

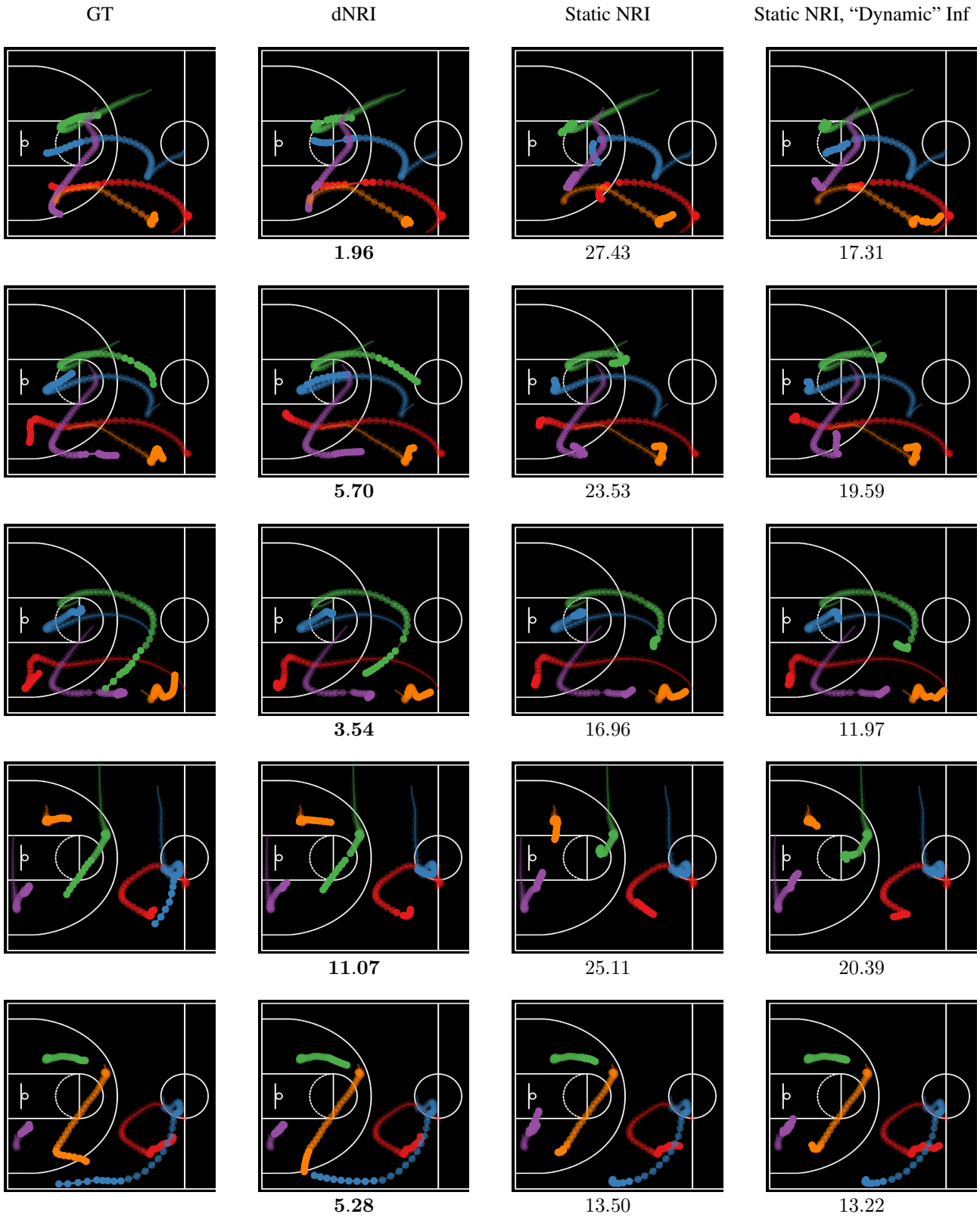

Figure 17: Predicted trajectories for test trajectories #1-#5 of basketball dataset. Displayed numbers are mean squared error of the unnormalized 2D position for each model’s prediction for the final time step.

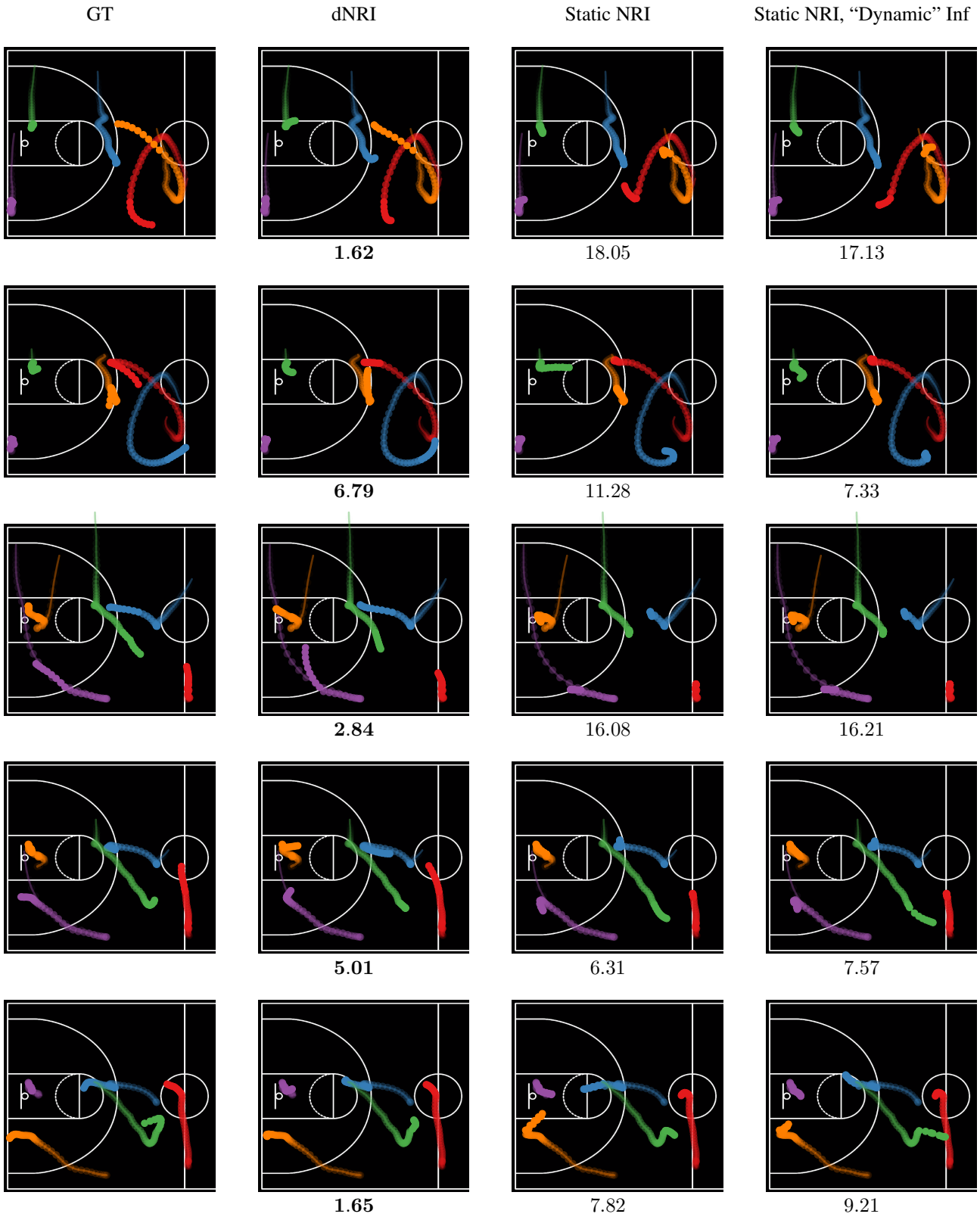

Figure 18: Predicted trajectories for test trajectories #6-#10 of basketball dataset. Displayed numbers are mean squared error of the unnormalized 2D position for each model’s prediction for the final time step.

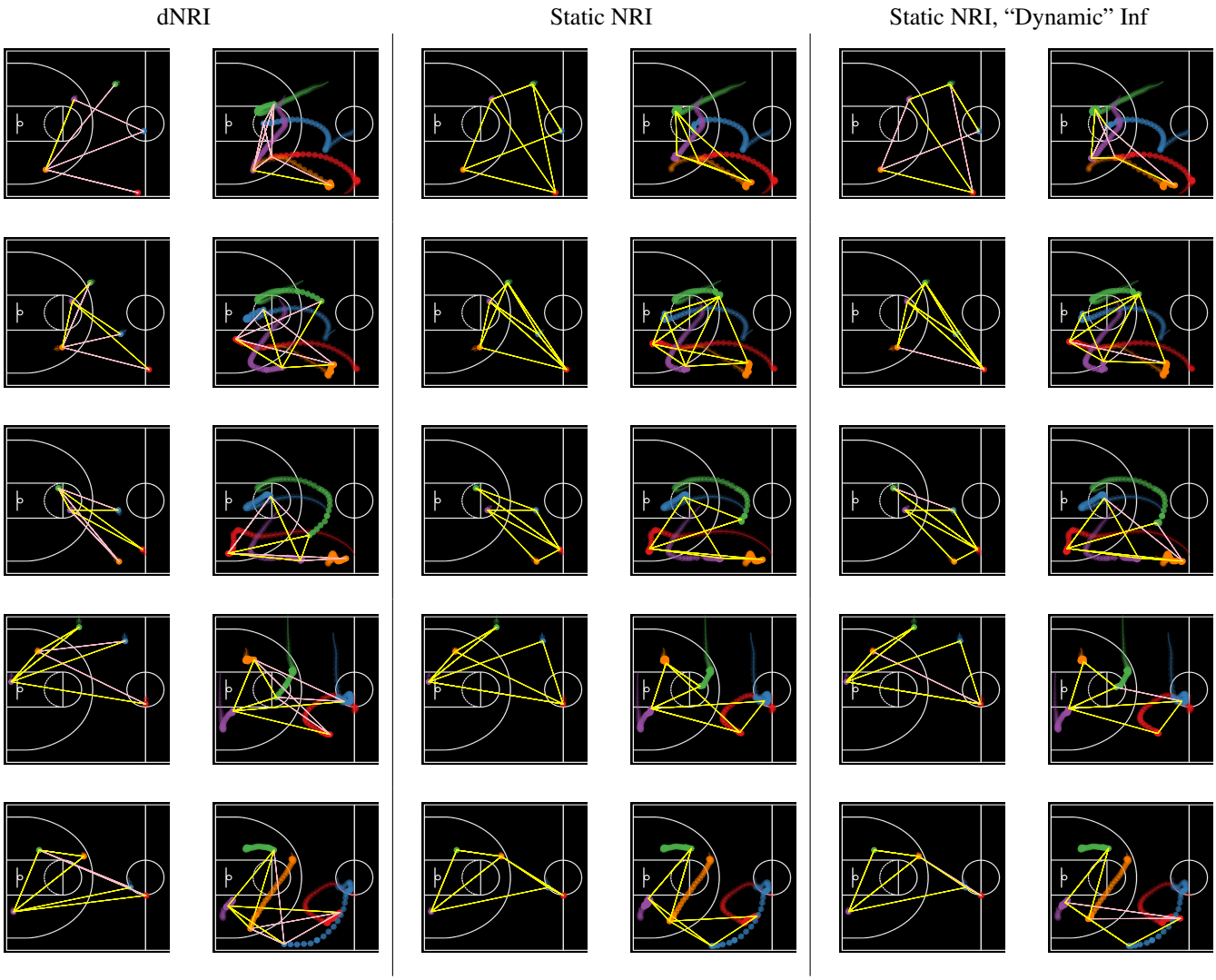

Figure 19: Predicted relations for test trajectories #1-#5 of basketball dataset.

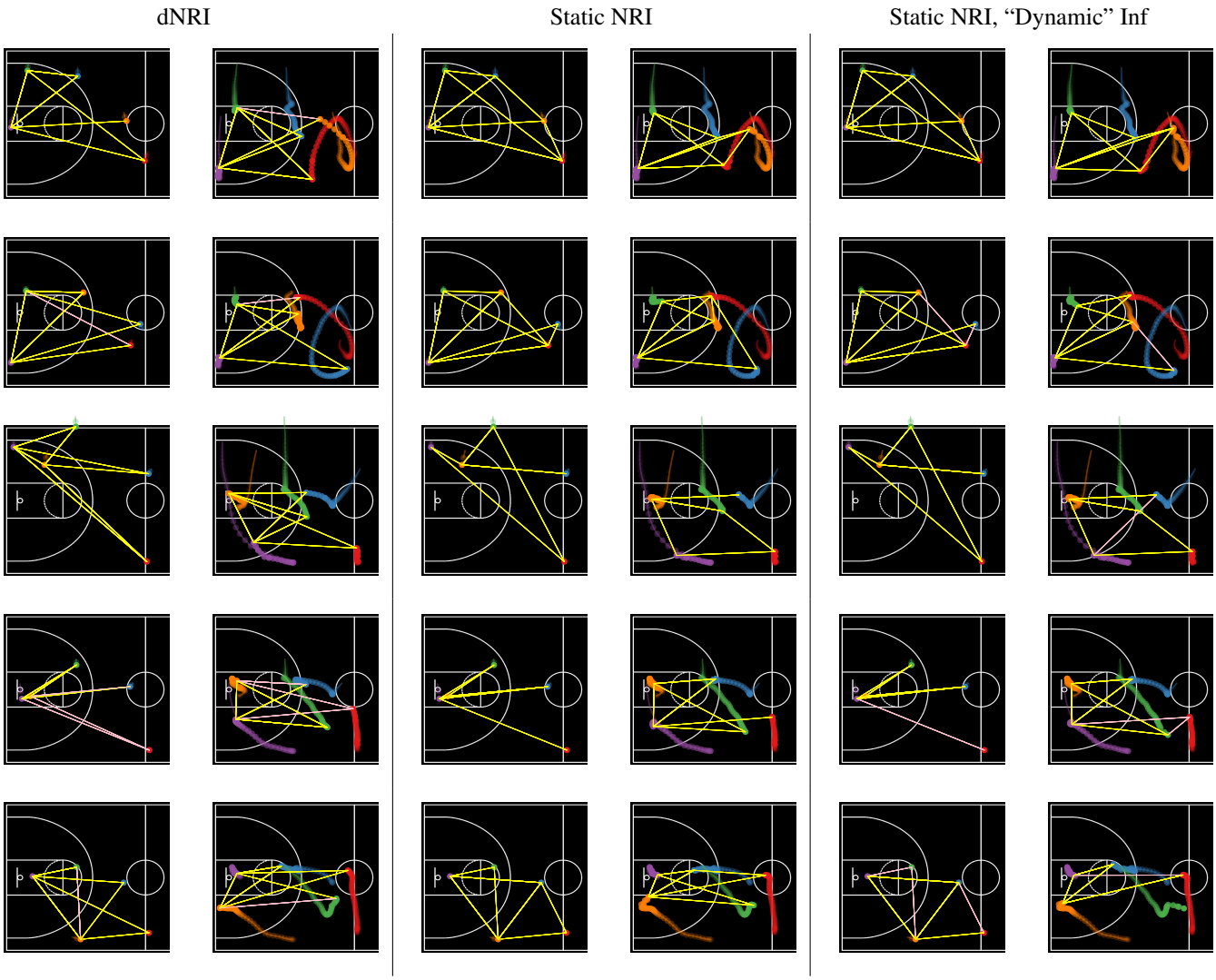

Figure 20: Predicted relations for test trajectories #6-#10 of basketball dataset.