# Supplementary Material of Future Moment Assessment for Action Query

In this supplementary material, we provide some qualitative examples of starting moment assessment.

Figure 1 are sampled from Breakfast Dataset. The first three frame images (from left to right) of each row are sampled from an observed sequence. The white mask indicates that the future frames are unseen. The query action is shown in the last frame image of each row. Given the observation and query action, we generate 100 predictions using VQNet and DR-VQNet. We show examples with query actions at different time-horizons. Below we describe the details of each sample.

(a) In this sample, the subject is preparing cereals. The query action 'Stir Cereals' occurs at 6s in the future. The prediction of DQNet is 3.5s. The mean values of 100 predictions of DR-VQNet and VQNet are 5.2s and 38.2s, respectively.

(b) This sample contains an activity of making tea. The starting moment of the query action 'Stir Tea' is 12s. The prediction of DQNet is 4.2s. The mean values of 100 predictions of DR-VQNet and VQNet are 18.4s and 43.7s, respectively.

(c) The activity of this sample is preparing fruit salad. The query action 'Take Bowl' occurs at 62s after the observation. The prediction of DQNet is 21.1s. The mean values of 100 predictions of DR-VQNet and VQNet are 59.1s and 53.9s, respectively.

(d) In this sample, the subject is frying egg. The query action 'Put Egg' occurs at 69s after the observation. The prediction of DQNet is 21.5s. The mean values of 100 predictions of DR-VQNet and VQNet are 55.9s and 50.9s, respectively.

(e)The activity of this sample is preparing fruit salad. The starting moment of the query action 'Stir Fruit' is 174s. The prediction of DQNet is 107.2s. The mean values of 100 predictions of DR-VQNet and VQNet are 149.3s and 72.8s, respectively.

(f) In this sample, the subject is making pancake. The query action is 'Take Plate'. It occurs at 282s in the future. The prediction of DQNet is 393.1s. The mean values of 100 predictions of DR-VQNet and VQNet are 289.8s and 191.6s, respectively.

(g) This sample is a failure case where the predictions of the three methods are distant from the ground-truth value. This is due to that the query action (pour oil) has appeared once in the observation. This makes the three methods anticipate that query action will occur shortly after the observation.

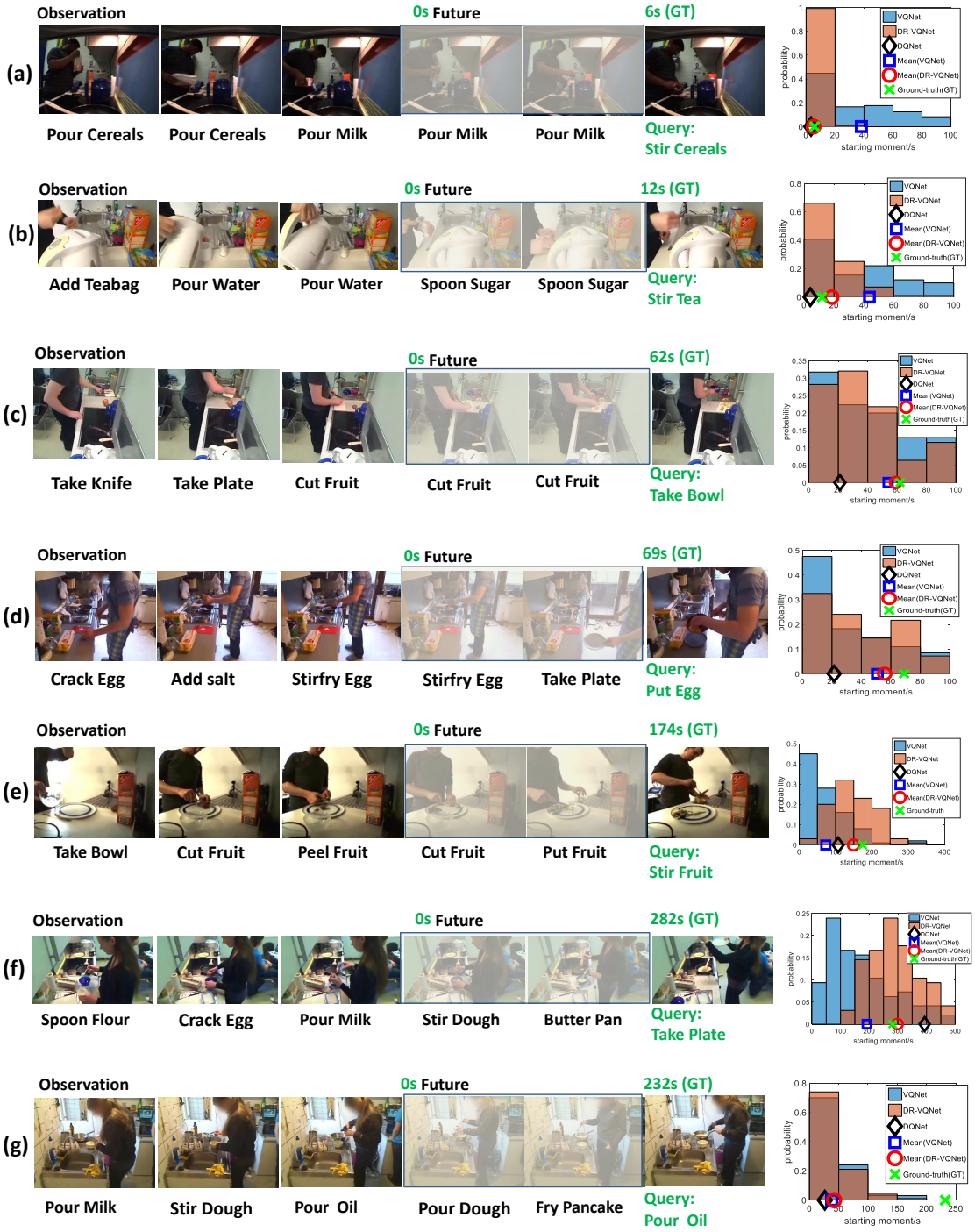


Figure 1: Visualizations of starting moment predictions (last column) generated by DQNet, VQNet and DR-VQNet on the Breakfast Dataset. The blue and red histograms illustrate the distributions of the 100 starting moment predictions generated by VQNet and DR-VQNet, respectively. The black ♢ denotes the starting moment predictions generated by DQNet. The blue □ denotes the mean of the starting moment predictions generated by VQNet. The red ◯ denotes the mean of the starting moment predictions generated by DR-VQNet. The green × denotes the ground-truth starting moment of the query action.