A. Additional Trajectory Visualizations across Methods

Additional examples in the style of Figure 2 of performance comparisons across methods on specific trajectory sequences can be seen in Figures 4-10.

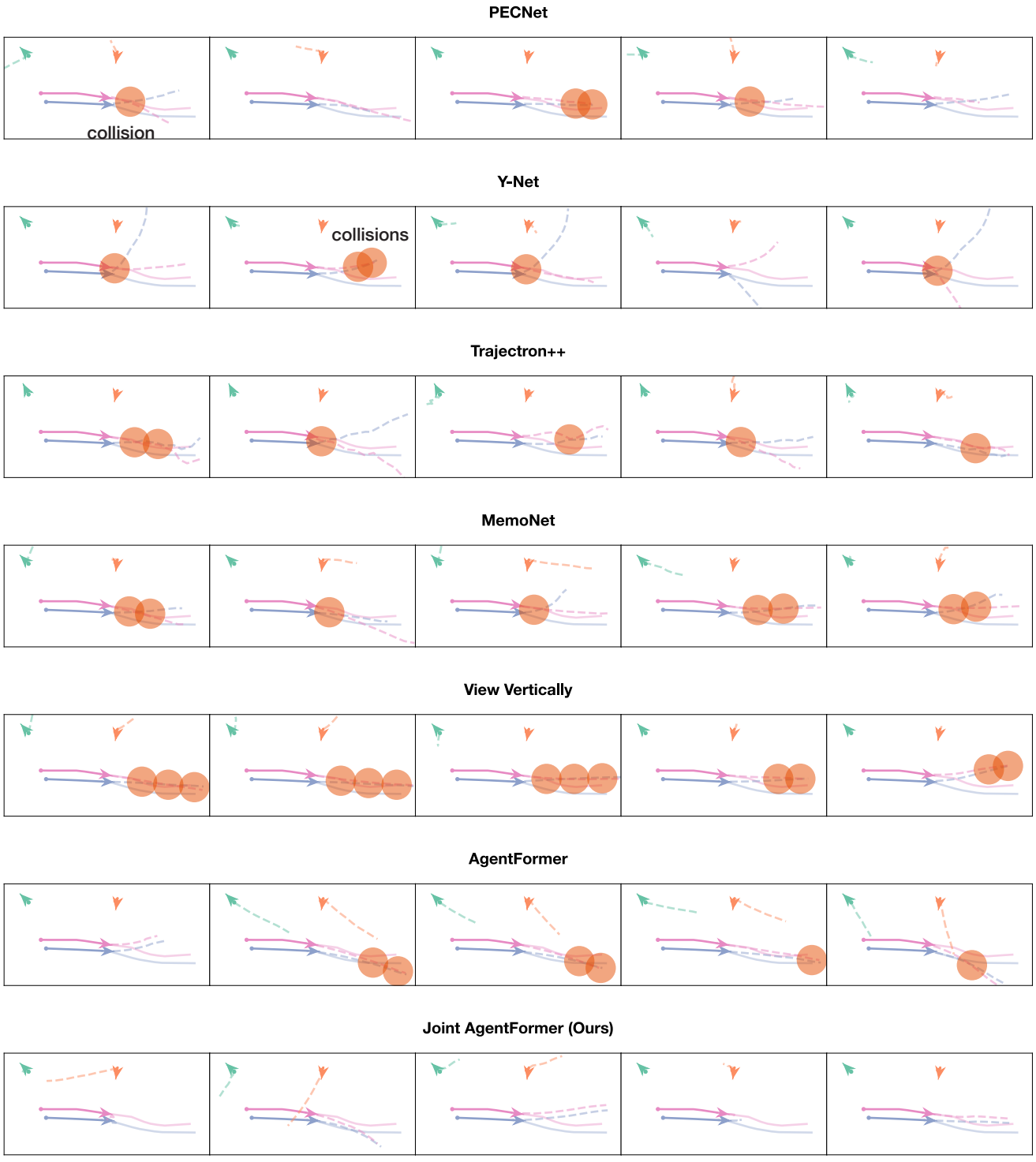

Figure 4: Our method exhibits improved collision-avoidance over baselines. While other methods (first 6 rows), including AgentFormer, predicts trajectories with collisions (denoted by orange circles), our method (last row) predicts trajectories with no collisions.

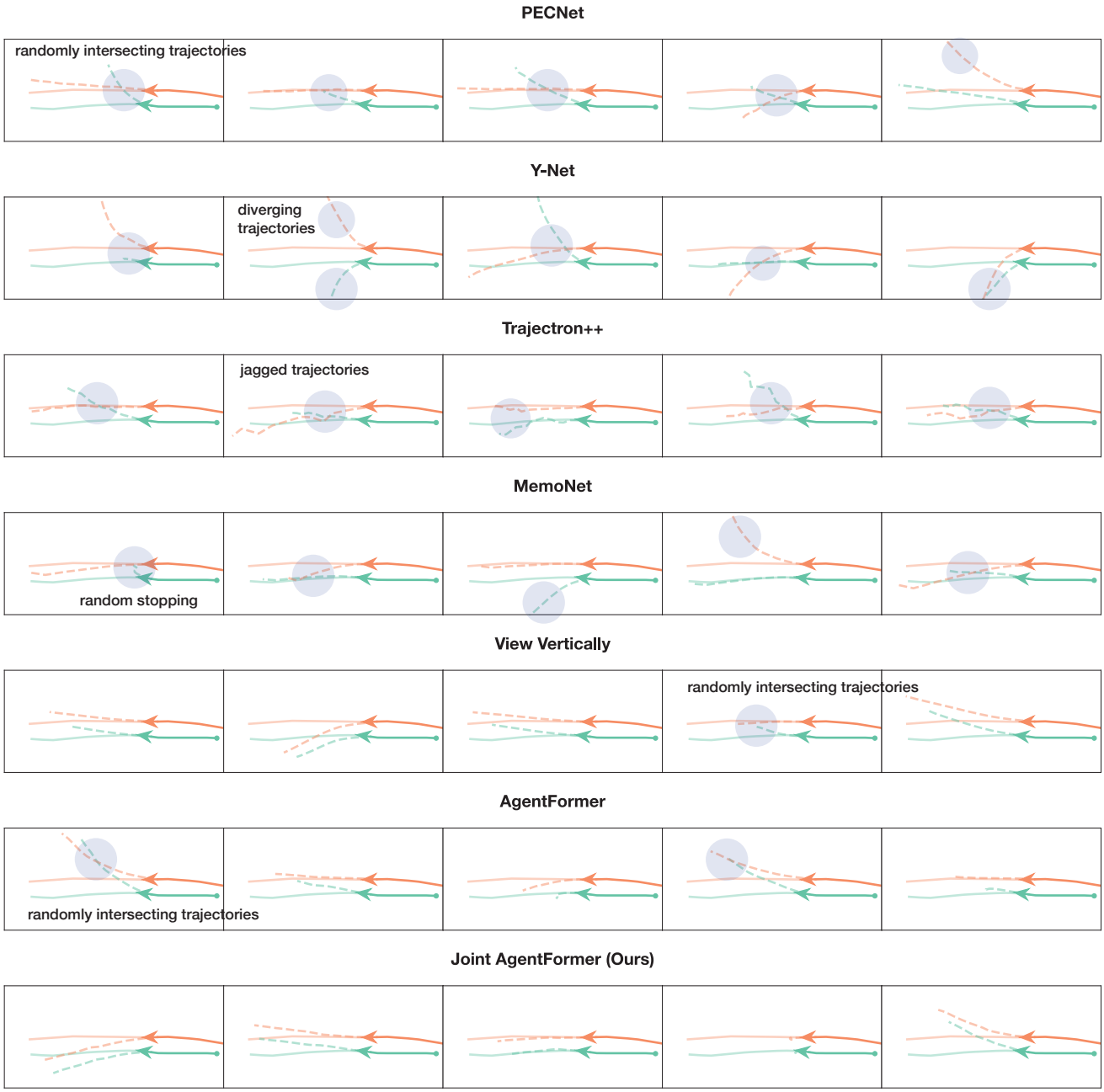

Figure 5: Our method exhibits more natural grouping behavior over baselines. While other methods (first rows) predict unnatural trajectories such as diverging or intersecting trajectories (denoted by blue circles) for agents who are obviously grouped, our method (last row), optimized for joint metrics, predicts reasonable grouping behavior while still maintaining diversity of predictions.

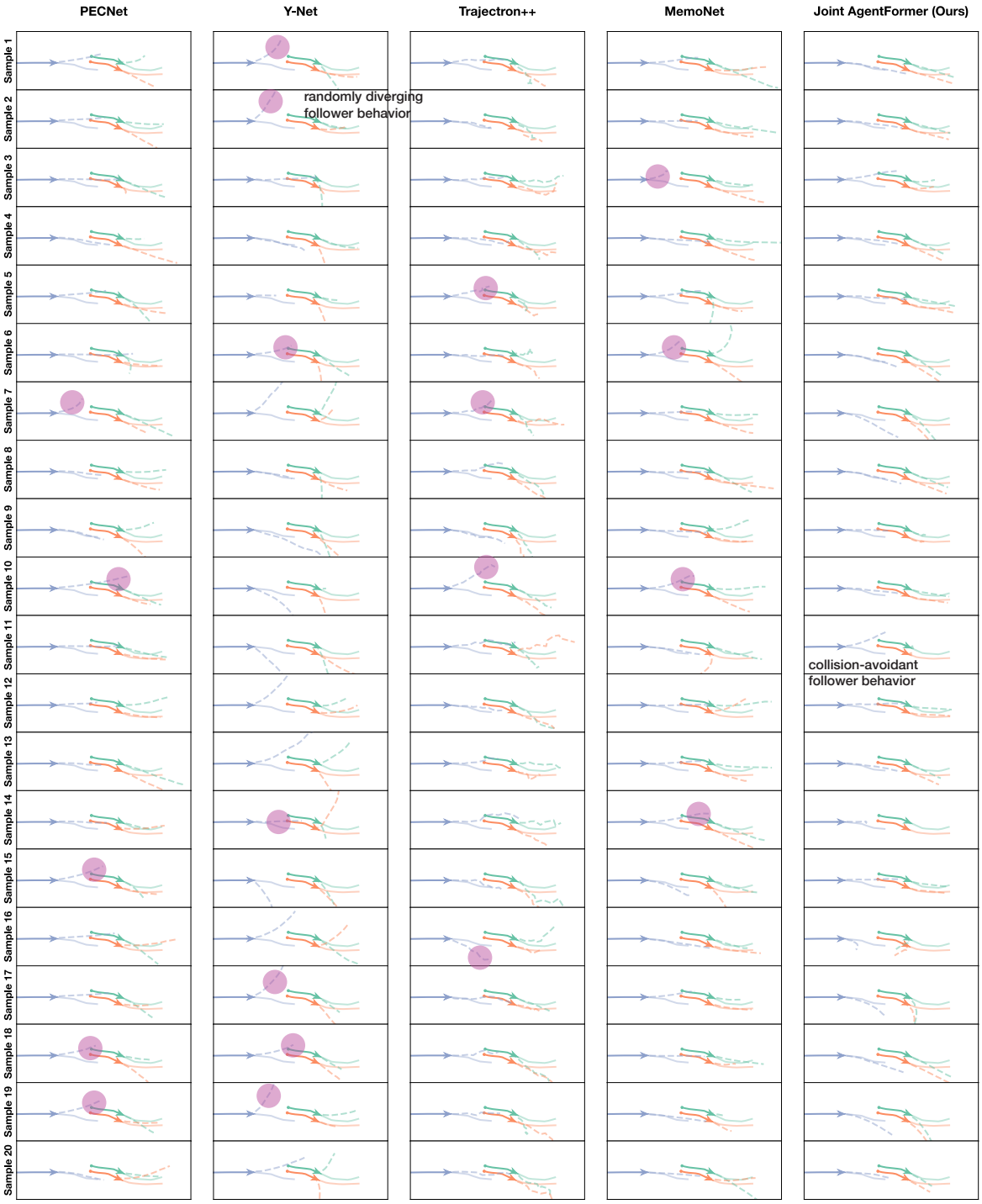

Figure 6: Our method exhibits more natural leader-follower behavior over baselines. Other methods (first four columns) predict unnatural trajectories such as sudden direction changes (denoted by purple circles) for the blue agent, who is clearly following along the natural path defined by the orange and green agent. On the other hand, our method (last column) predicts natural leader-follower behavior for the blue agent, while still maintaining visible diversity of predictions. Also of interest is our method's superiority with respect to predicting collision-avoidance as well as natural grouping behavior of the orange and green agents, but we leave it unmarked in this visual to leave the reader's attention upon leader-follower behavior.

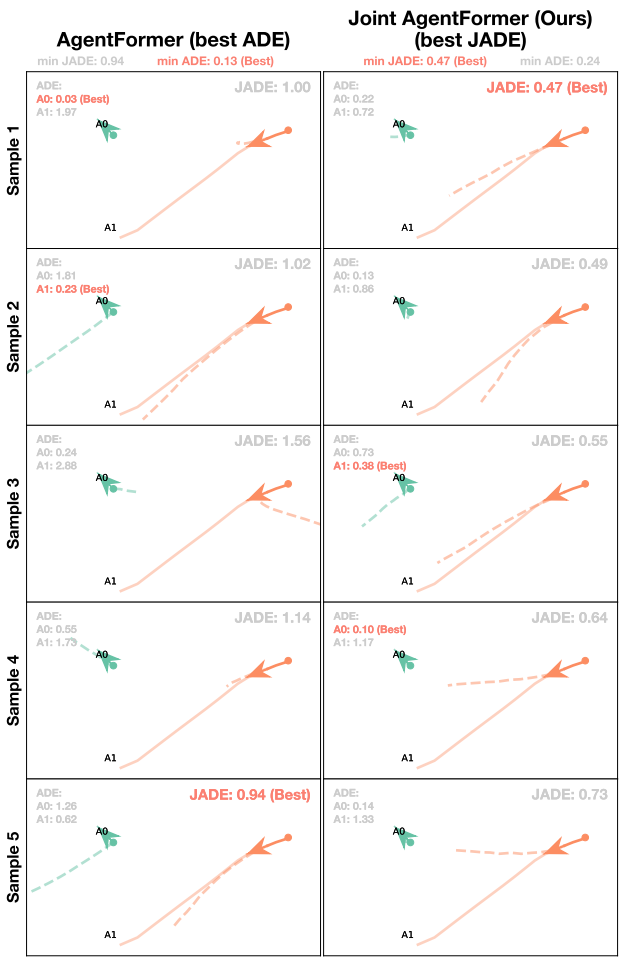

Figure 7: AgentFormer vs. Our Method. Of particular interest is qualitative comparison between AgentFormer and our method, as our method was built on AgentFormer. In our qualitative observations, AgentFormer heavily represents the future mode in which a static pedestrian suddenly begins moving at regular speed. This static-to-moving mode is not very likely; static-to-moving pedestrians account for only $1887/34161 \approx 5.5\%$ of pedestrians in the ETH / UCY datasets. The reason AgentFormer does this is because under marginal loss, over-representing unlikely modes is not penalized. Marginal optimization favors exploration of unlikely modes over exploitation of the most probable modes, because marginal evaluation assumes that any of the 20^N possible ways to mix-and-match 20 trajectory predictions for N agents are valid. Thus, marginal evaluation favors prediction diversity at the expense of realism. On the other hand, under joint loss optimization, our method is allowed not 20^N but rather only 20 "tries" to generated an accurate joint prediction. Not having the luxury to explore, our method reduces the representation of the unlikely static-to-moving mode. Thus, even though AgentFormer outperforms our method with respect to *ADE*, our method, which outperforms AgentFormer with respect to *JADE*, produces a more realistic set of predictions.

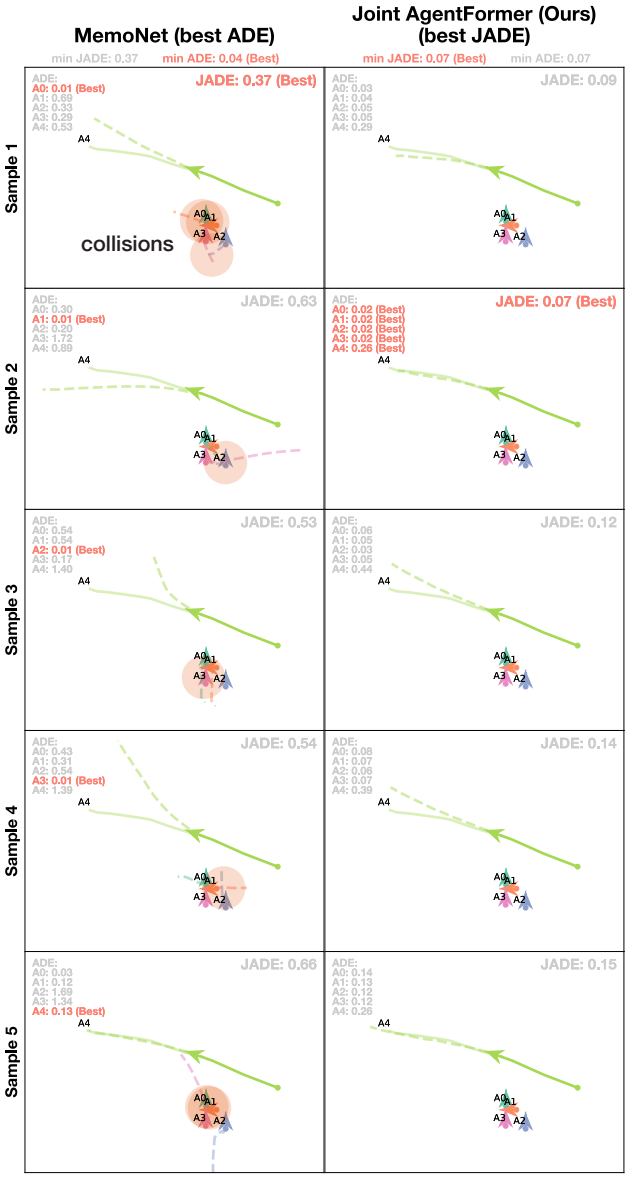

Figure 8: Our method outperforms MemoNet, the SOTA with respect to *XDE*, at producing collision-free joint predictions. Although MemoNet [54] is SOTA with respect to *ADE*/ *FDE*on ETH / UCY Average, it appears to possess minimal social awareness, producing predictions full of collisions (denoted by red circles). Furthermore, in this Figure we also observe that MemoNet produces unnatural and inconsistent trajectories. For example, the group of static pedestrians in the lower-right corner is expected to maintain its static position, or perhaps all begin moving as a group. However, MemoNet occasionally predicts them moving off in different directions. Thus, even though MemoNet outperforms our method at *ADE*; our method, which outperforms with respet to *JXDE*, generates more natural and consistent predictions.

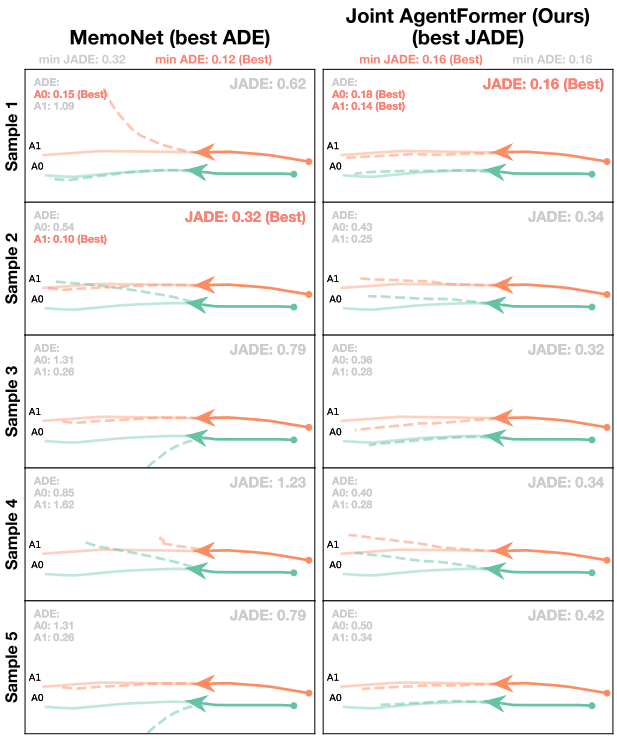

Figure 9: Our method outperforms MemoNet, the SOTA with respect to *XDE*, at producing natural and consistent joint predictions. Although MemoNet [54] is SOTA with respect to *ADE*/ *FDE*on ETH / UCY Average, it appears to possess minimal social awareness, producing predictions with clearly grouped agents going off in different directions. Thus, even though MemoNet outperforms our method at *ADE*; our method, which outperforms with respet to *JXDE*, generates more natural and consistent predictions.

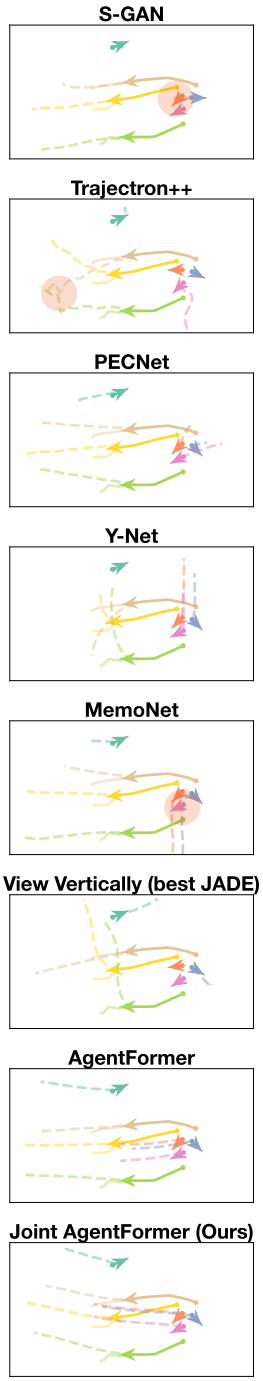

Figure 10: Worst-case Analysis. For the sequence shown in this figure, we select the max *JADE*example for each method. Even the worst-*JADE*prediction of our method produces natural-looking trajectories, while other methods produce collisions (S-GAN, Trajectron++, MemoNet) or inconsistent and random predictions (Y-Net, View Vertically).