# Recover to Predict: Progressive Retrospective Learning for Variable-Length Trajectory Prediction

## Supplementary Material

In this supplementary file, we provide additional details and results to demonstrate the benefits of the proposed framework further. The contents include the following appendices:

- Appendix for RSTS (Section 7)
- Appendix for Loss Functions (Section 8)
- Appendix for Evaluation Metrics (Section 9)
- Appendix for Qualitative Evaluations. (Section 10)
- Appendix for Interpretability Analysis. (Section 11)

## 7. Appendix for RSTS

The proposed Rolling-Start Training Strategy (RSTS), described in Section 3.5, improves data efficiency by incorporating incomplete observations during training. Fig. 6 illustrates the applications of RSTS on the Argoverse 2 dataset, with a standard observation horizon of $T_o = 50$ and a prediction horizon of $T_f = 60$.

When $T_v = 50$, which corresponds to the standard observation horizon, a standard sample pair ([1,50], [51,110]) can be segmented into observation windows {[41,50], [31,50], [21,50], [11,50], [1,50]}. These observation windows are then encoded to train retrospective units $\{\Phi^4, \Phi^3, \Phi^2, \Phi^1\}$, with the encoded feature of the standard-length observation window [1,50] being used to train the decoder.

Then, the start point is shifted to $T_v = 40$, generating a sample pair ([1,40],[41,100]). This sample pair is segmented into observation windows {[31, 40], [21, 40], [11, 40], [1, 40]}. These observation windows are encoded to train respective units $\{\Phi^4, \Phi^3, \Phi^2\}$. The encoded feature of the incomplete observation window [1, 40] is distilled by unit $\Phi^1$ to match the standard observation length, which is then used to train the decoder.

Subsequently, the start point is shifted to $T_v = 30$, producing a sample pair ([1, 30],[31,90]). This sample pair is segmented into observation windows {[21, 30], [11, 30], [1, 30]}. These observation windows are encoded to train respective units $\{\Phi^4, \Phi^3\}$. The encoded feature of the incomplete observation window [1,30] is sequentially distilled by units $\Phi^2$ and $\Phi^1$ to match the standard observation length, which is used to train the encoder.

Finally, the start point is shifted to $T_v = 20$, yielding a sample pair ([1,20],[21,80]). This sample pair is segmented into observation windows {[11, 20], [1, 20]}. The two observation windows are encoded to train unit $\Phi^4$, with the encoded feature of the incomplete observation window [1, 20] being sequentially distilled by units $\Phi^3$, $\Phi^2$, and $\Phi^1$ to match the standard observation length, which is used to train the decoder.

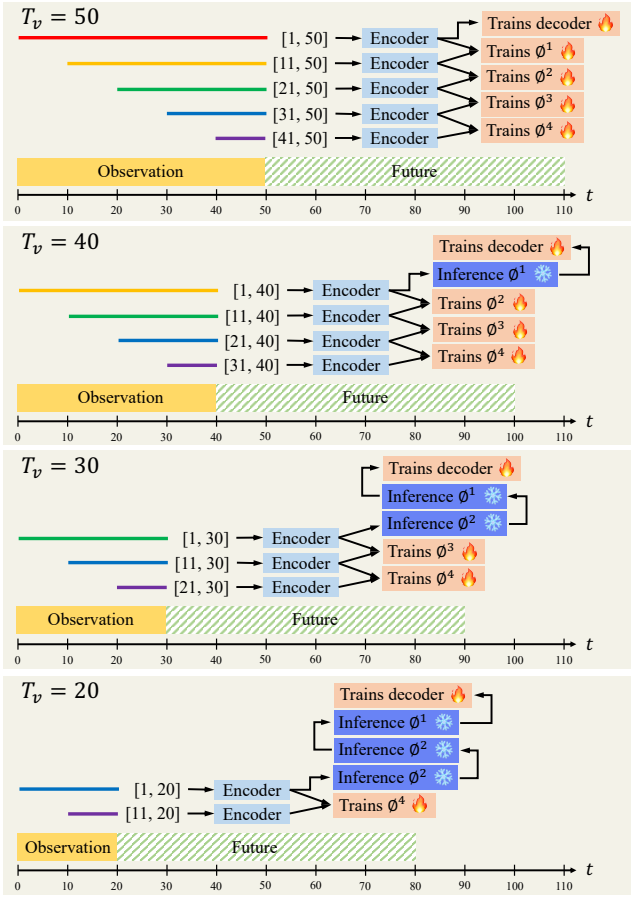
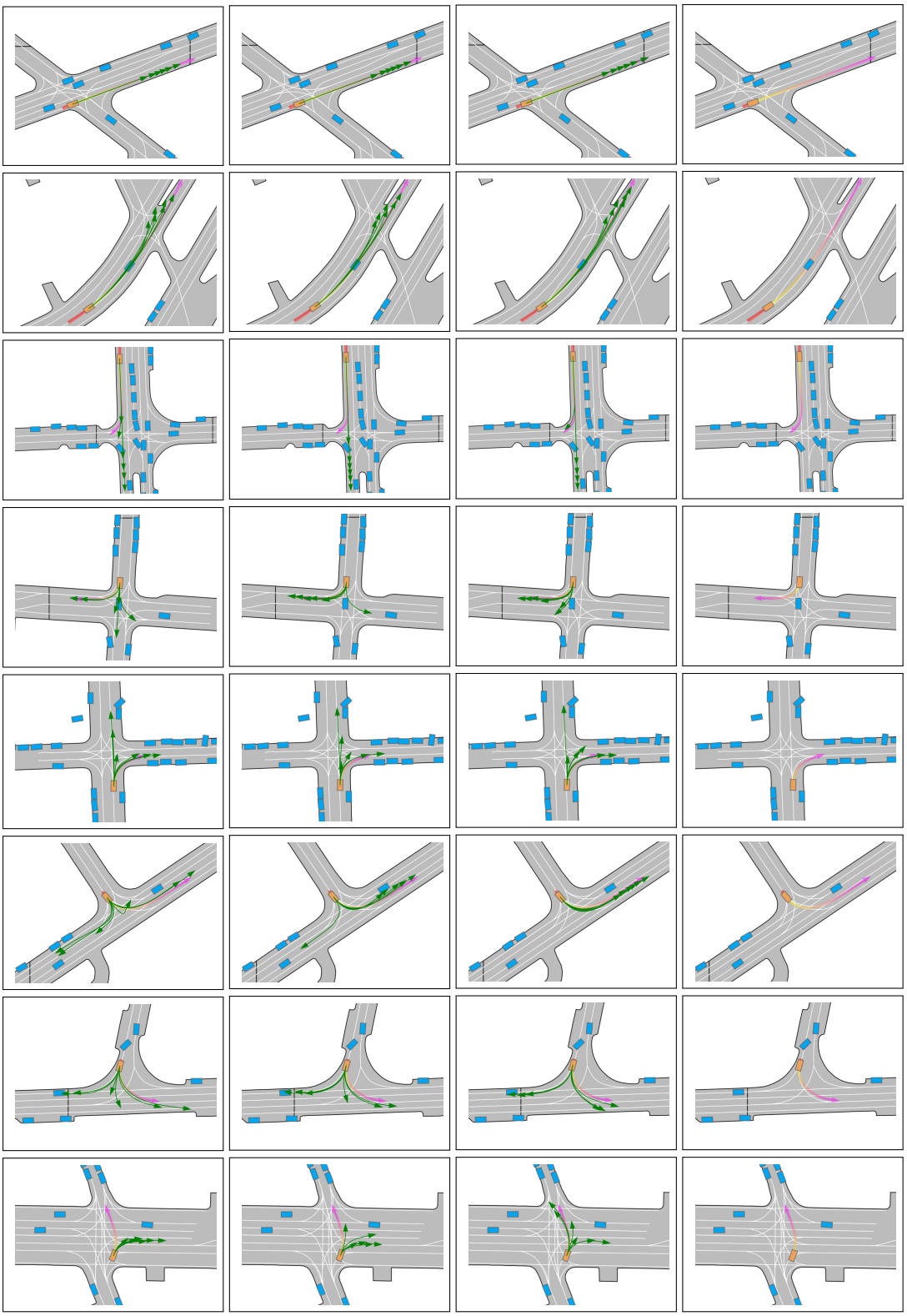



Figure 6. Illustration of the RSTS on the Argoverse 2 dataset with a standard observation horizon of $T_o = 50$ and a prediction horizon of $T_f = 60$. As the prediction start point shifts from 50 to 40, 30, and 20, additional training samples are generated to train the retrospective units and the decoder.

In summary, RSTS generates {4,3,2,1} samples to train the retrospective units $\{\Phi^4, \Phi^3, \Phi^2, \Phi^1\}$, respectively, and 4 samples to train the decoder, using a standard training sequence.

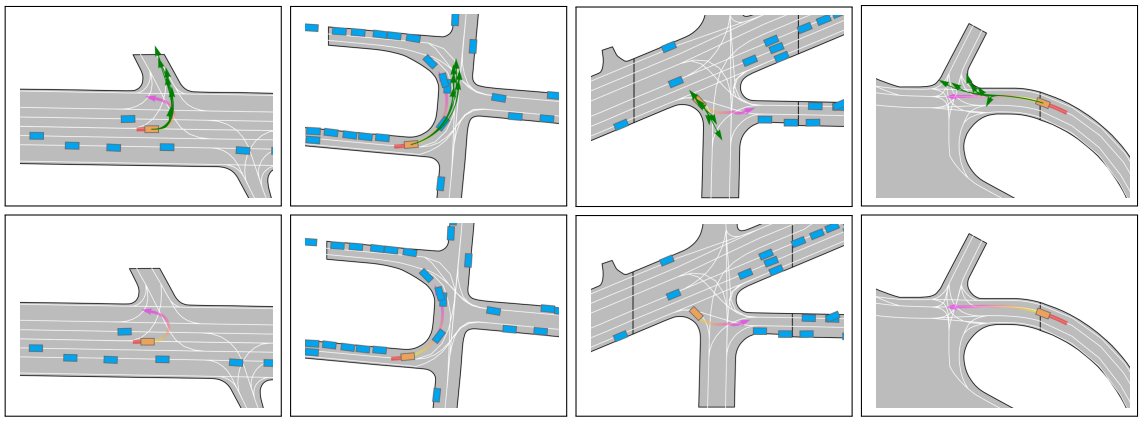
## 8. Appendix for Loss Functions

A smooth-L1 loss and a cross-entropy loss are employed to train the decoder and RPM, as introduced in Section 3.6. The ground-truth future trajectories, predicted future trajectories, and their probability are represented by $\mathbf{Y} \in \mathbb{R}^{N_a \times T_f \times 2}$, $\tilde{\mathbf{Y}} \in \mathbb{R}^{N_a \times K \times T_f \times 2}$, and $\mathbf{P} \in \mathbb{R}^{N_a \times K}$, where $N_a$, $K$, $T_f$, and 2 represents the number of predicted agents,



(a) DeMo-IT (b) DeMo-CLLS (c) DeMo-PRF (Our) (d) GT

Figure 7. More qualitative results on the Argoverse 2 validation set. Incomplete observations, predicted trajectories, and ground truth trajectories are shown in yellow, green, and pink, respectively. The absence of an observation trajectory indicates that the vehicle is stationary. Our predictions align more closely with the ground truth compared to other methods.



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3 (d) Scenario 4

Figure 8. Failure cases of DeMo-PRF on the Argoverse 2 validation set. The first and second rows visualize the predicted trajectories and ground-truth trajectories, respectively. Incomplete observations, predicted trajectories, and ground truth trajectories are shown in yellow, green, and pink, respectively. The absence of an observation trajectory indicates that the vehicle is stationary.

the number of predicted modes, the prediction horizon, and the coordinate dimensions, respectively. These variables are used to compute the smooth-L1 loss and cross-entropy loss.

**Smooth-L1 regression loss.** The smooth-L1 regression loss is computed using the ground-truth future trajectories $\mathbf{Y}$ and predicted future trajectories $\tilde{\mathbf{Y}}$ as follows:

$$\mathcal{L}_{\text{reg}} = \frac{1}{N_a T_f} \sum_{i=1}^{N_a} \sum_{t=1}^{T_f} \text{SmoothL1}\big(\tilde{\mathbf{Y}}_{i,k_i^\star,t} - \mathbf{Y}_{i,t}\big)$$

where $k_i^\star$ denotes the index of the best predicted mode for agent $i$.

**Cross-entropy classification loss.** For probability score classification, the index $k_i^\star$, corresponding to the mode with the smallest ADE of agent $i$, is used as the ground-truth class label. Then, using the predicted probability $\mathbf{P}$ and the ground-truth class label, the classification loss is calculated as:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N_a} \sum_{i=1}^{N_a} \log \mathbf{P}_{i,k_i^\star}. \tag{13}$$

The overall training loss for the decoder and RPM is the sum of the regression and classification terms.

## 9. Appendix for Evaluation Metrics

We adopt commonly used metrics, namely $\text{mADE}_K$, $\text{mFDE}_K$, $\text{b-mFDE}_K$, and $\text{MR}_K$, to evaluate PRF, as described in Section 4.1. The ground-truth future trajectories $\mathbf{Y}$, predicted future trajectories $\tilde{\mathbf{Y}}$, and their associated probabilities $\mathbf{P}$ are used to compute these metrics. Specifically, for each agent $i$ and mode $k$, the ADE and FDE are defined as follows:

$$\begin{aligned} \text{ADE}_{i,k} &= \frac{1}{T_f} \sum_{t=1}^{T_f} \left\| \tilde{\mathbf{Y}}_{i,k,t} - \mathbf{Y}_{i,t} \right\|_2, \\ \text{FDE}_{i,k} &= \left\| \tilde{\mathbf{Y}}_{i,k,T_f} - \mathbf{Y}_{i,T_f} \right\|_2. \end{aligned} \tag{14}$$

Then, $\text{mADE}_K$ and $\text{mFDE}_K$ are calculated as the average minimum ADE and FDE over $K$ modes, respectively:

$$\begin{aligned} \text{mADE}_K &= \frac{1}{N_a} \sum_{i=1}^{N_a} \min_{1 \le k \le K} \text{ADE}_{i,k}, \\ \text{mFDE}_K &= \frac{1}{N_a} \sum_{i=1}^{N_a} \min_{1 \le k \le K} \text{FDE}_{i,k}. \end{aligned} \tag{15}$$

The $\text{b-mFDE}_K$ metric augments $\text{mFDE}_K$ with a Brier-style penalty based on the probability of the best predicted mode:

$$\text{b-mFDE}_K = \frac{1}{N_a} \sum_{i=1}^{N_a} \left[ \text{FDE}_{i,k_i^\star} + \big(1 - P_{i,k_i^\star}\big)^2 \right]. \tag{16}$$

The $\text{MR}_K$ metric measures the fraction of agents for which even the best of the $K$ predicted trajectories deviates from the ground truth by more than a threshold $\delta = 2.0$ meters at the final time step:

$$\text{MR}_K = \frac{1}{N_a} \sum_{i=1}^{N_a} \mathbf{1}\Big(\text{FDE}_{i,k_i^\star} > \delta\Big), \tag{17}$$

where $\mathbf{1}(\cdot)$ is the indicator function that returns 1 if the condition is true and 0 otherwise. These metrics can be extended to the entire dataset by averaging over the total number of predicted agents across all scenes.

## 10. Appendix for Qualitative Evaluations

Additional qualitative results, complementing those presented in Fig. 4, are shown in Fig. 7. All results are predicted from very short observation horizons of only 10 timesteps. In some scenarios, the absence of an observation trajectory indicates that the vehicle is stationary during the observation window. These qualitative results, spanning

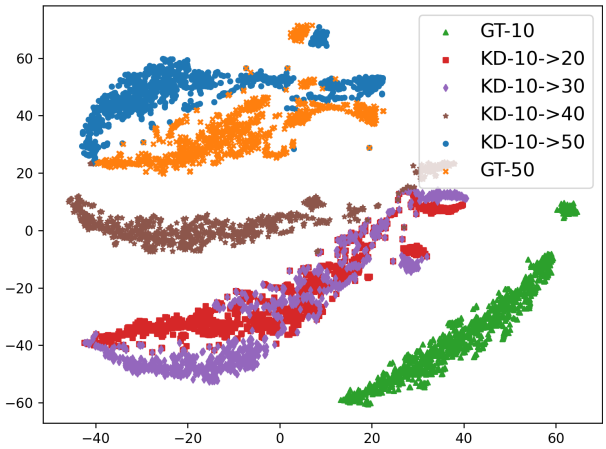



Figure 9. t-SNE visualization of features distilled by the progressive strategy. Green and orange points represent features extracted from trajectories with observation lengths of 10 and 50 timesteps, respectively. Red, purple, brown, and blue points correspond to features distilled from 10-step observations to those with 20, 30, 40, and 50 steps, which gradually shift from the manifold of 10-step observations toward that of 50-step observations.

various driving scenarios, further highlight the state-of-the-art performance of the proposed PRF.

**Failure case.** Fig. 8 illustrates four failure cases when using very short observation horizons of only 10 timesteps. Fig. 8a and Fig. 8b depict failure cases in U-turn scenarios. Fig. 8c shows a failure case in a compound turn scenario, while Fig. 8d presents a failure case in the on-ramp merging scenario. These scenarios present long-tail problems for trajectory prediction, even with complete observation lengths. With such short and incomplete observations, the proposed PRF initially tracks the ground-truth motion but eventually deviates as the maneuver becomes more complex. To improve predictions in these scenarios, future work could focus on enhancing the modeling of interactions among multiple agents and incorporating additional high-level context, such as traffic signals and right-of-way rules, as structural constraints on the predicted trajectories.

## 11. Appendix for Interpretability Analysis

Additional interpretability analysis, complementary to Fig. 5, is presented in Fig. 9. This figure visualizes the t-SNE of features extracted from observation lengths of 10 and 50 timesteps, as well as features distilled from 10-step observations to those with 20, 30, 40, and 50 timesteps. The visualization shows that, as progressive distillation proceeds, features distilled from trajectories with an observation length of 10 timesteps gradually converge toward the features obtained from 50-step observations. This demonstrates that decomposing direct distillation into a sequence of progressive distillation steps reduces the difficulty of the distillation process and effectively distills representations from short trajectories into those of complete trajectories.