ME5413 Report Homework 1: Perception

Zhao Yimin Matric Number: A0285282X

1 Task 1: Single Object Tracking

1.1 Template Matching Principle

An advanced template matching approach named timing-base dynamic template matching is developed. this approach enhances the original template matching technique by incorporating dynamic search areas, preprocessing for robustness, and contour analysis for adaptability. These improvements collectively aim to increase the accuracy and robustness of object tracking in video sequences, particularly in challenging conditions with variations in object appearance, size, and movement. The following part summarizes the principle and efforts made to improve the original template matching method:

Dynamic Search Region

The search region is adjusted through the calculation of a direction vector based on the location of detection bounding boxes in previous two frames. This vector contributes to predicting location of search region in the current frame based on the following equation:

$$\vec{r}_{\text{cur}} = \max\left(0, \left\lfloor \vec{r}_{\text{prev}} + \vec{d} \cdot s \right\rfloor\right) \tag{1}$$

where $\vec{r}_{\text{cur}}$ is the search region location of the current frame, and $\vec{r}_{\text{prev}}$ is that of the previous frame. A scale factor s is multiplied with the direction vector $\vec{d}$ mentioned above, to adjust effects that direction vector applied on the prediction of search region movement. Furthermore, the search area's size is dynamically adjusted based on the module of that vector, allowing large area for fast movements and small one for slow movements.

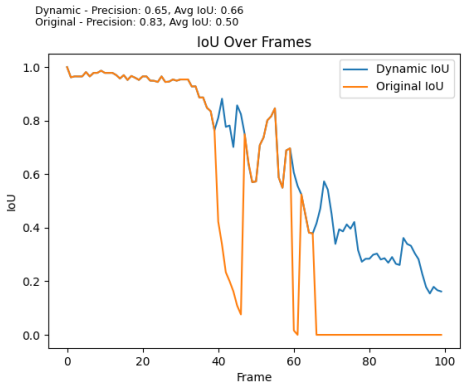
Contour Detection and Bounding Box Cropping

After finding a match, the code pre-processes the matched region by converting it to grayscale, applying Gaussian blur for noise reduction, and then equalizing the histogram to enhance contrast. Then, Canny edge detection is performed to identify edges, where the edges closest to the detection frame are used to form a contour to enable cropping of the bounding box. This step ensures that the template adapts to changes in the object's size, improving tracking accuracy.

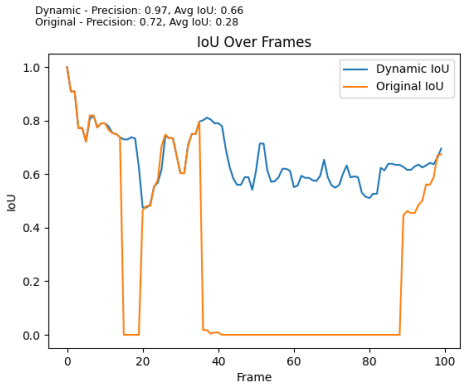
1.2 Template Matching Evaluation

To evaluate the effectiveness of the proposed novel template matching method, the Intersection of Union (IoU) diagram of all frames in each video is plotted as Fig. 1. It is very obvious that the new method captures object in more frames, especially for video 2 and video 5.

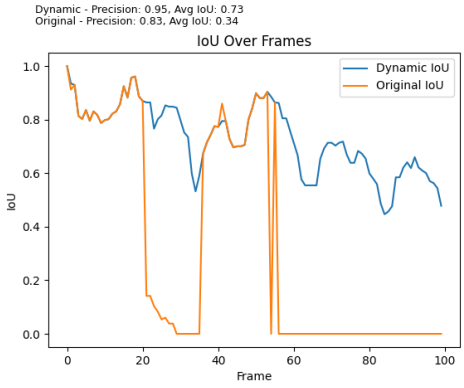
By referring the results videos, it is found that the correct search region setting can avoid high-score but non-targeted object in the whole frame being identified.



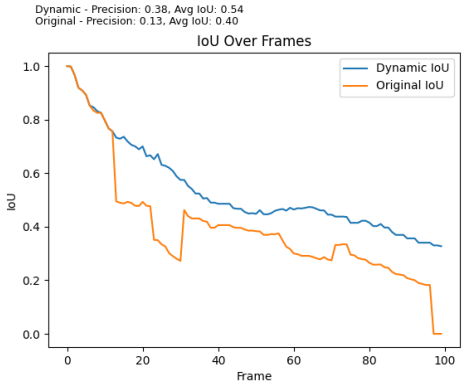
(a) Video 1



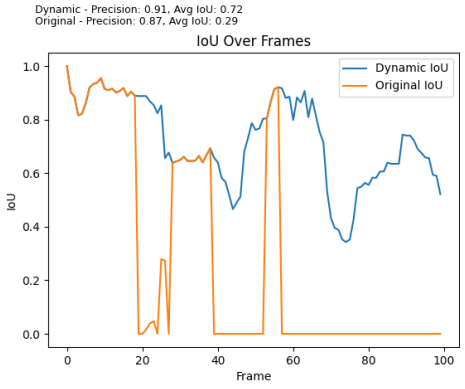
(b) Video 2



(c) Video 3



(d) Video 4



(e) Video 5

Figure 1: Comparison of IoU variation in all videos between original template matching and customized timing-base dynamic template matching IoU diagram of all frames in each video.

To further evaluate the propose method, the metrics comparison table is illustrate by Tab. 1. All data from the proposed one are better than the original template matching except of precision of sequence 1, and that of sequence 4 performs not well with only 0.38. As shown in results videos, Fig. 1a and Fig. 1d, the ground truth area becomes small. Hence, the IoU

deceases and ultimately results in low precision. This indirectly reflects the edge detection and contour cropping are not effective.

Sequence	Original			Proposed Dynamic		
	Precision	Success	Time (s)	Precision	Success	Time (s)
Seq 1	**0.83**	0.50	4.38	0.65	**0.66**	**0.67**
Seq 2	0.72	0.28	0.99	**0.97**	**0.66**	**0.28**
Seq 3	0.83	0.34	2.57	**0.95**	**0.73**	**0.53**
Seq 4	0.13	0.40	2.15	**0.38**	**0.54**	**0.37**
Seq 5	0.87	0.29	2.36	**0.91**	**0.72**	**0.46**

Table 1: Comparison of precision and success (average IoU) between original template matching and self-designed timing-base dynamic template matching

However, the low processing duration presents applying template matching within the correct search region can significantly reduce computing resources and enhance efficiency. Moreover, high precision and high success indicate, the strategy of predicting search region based on target movement in previous frames, performs very well.

1.3 Kalman Filter Evaluation

The ground truth data is used as the measurement of the Kalman Filter, so there is no metrics evaluation for this section. Refer to the red bounding boxes in the mp4 videos, it is obvious that their movement becomes soother.

2 Task 2: Multi Object Prediction

For every kind of prediction model, the average data of the previous 1 second is utilized as input. Some invalid data in agent 15 is found and removed. The final visualization is shown in Fig. 2, which indicates agent 22 has no reference value. In that case, the following metrics calculation ignores this agent. In terms of other results, the trajectories are reasonable and accurate when there are little yaw angles of ground truth.

Model Category	ADE			FDE		
	Future 1s	Future 2s	Future 3s	Future 1s	Future 2s	Future 3s
CV	0.3263	0.8509	1.6778	0.6736	2.0884	4.5059
CA	**0.2608**	**0.6250**	**1.1756**	**0.5260**	**1.4643**	**3.0446**
CTRV	0.6615	1.4362	2.4333	1.2722	3.0808	5.6501

Table 2: Mean ADE and mean FDE of 8 agents comparisons between Constant Velocity Model, Constant Acceleration Model, and Constant Turn Rate and Velocity(CTRV) Model. The results of future 1 second, future 2 seconds, and future 3 seconds are all presented.

The Fig. 3 prominently presents those result. The agent 15 and agent 4, which are the two vehicles driving in turning scenario, have the highest error curves. The situation might be caused by the strategy of predictions using data. Another one is only utilize the driving

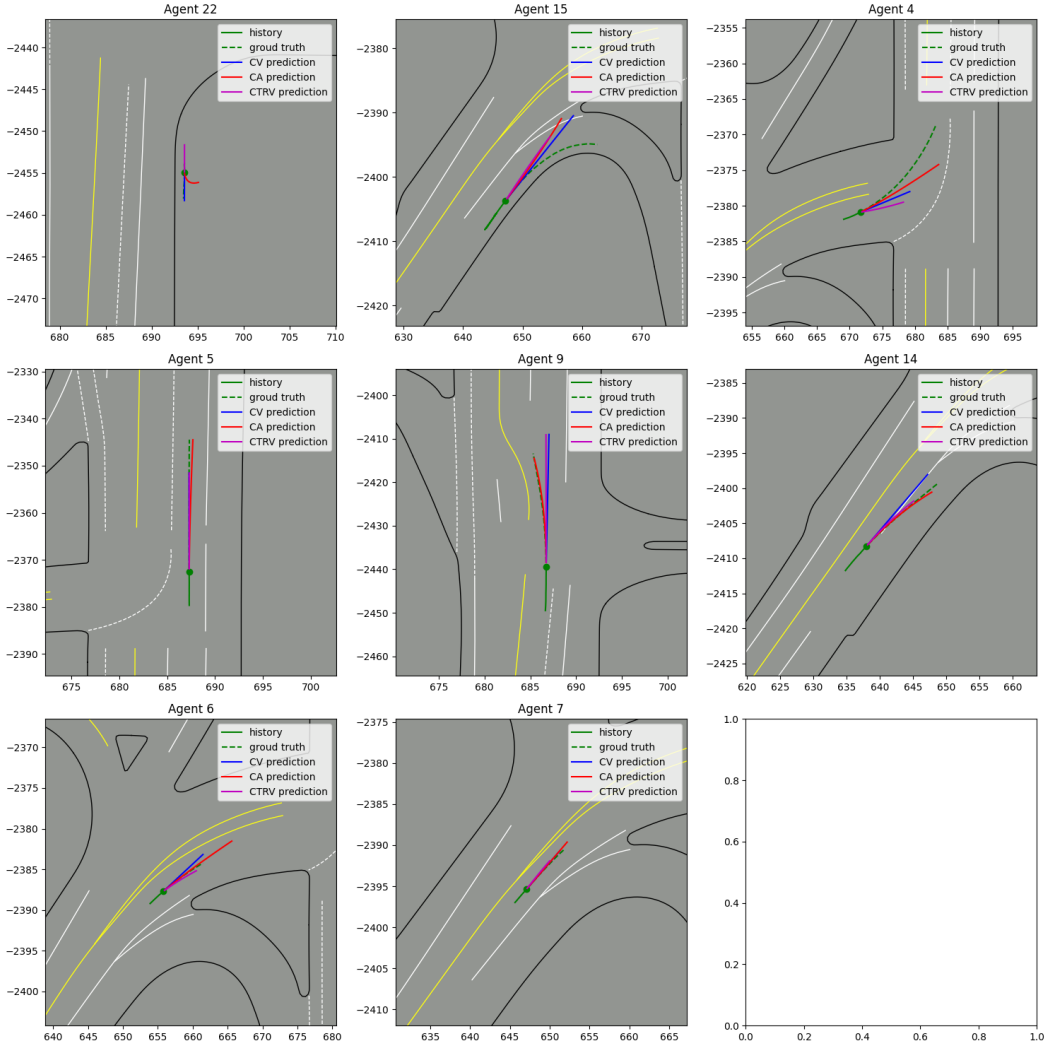


Figure 2: The 8 agents trajectories visualizations of prediction in three models, history, and ground truth.

data at the current position, which could be more accurate and effective. However, an invalid current data might cause serious consequence. Hence, the strategy of using current driving data is not applied.

Furthermore, the Tab. 2 reflects the metrics comparison results of mean ADE and mean FDE between three different models. Constant Acceleration Model is the best performing model. It excels in dynamic adaptation to accelerating or decelerating objects, closely aligning with real-world movement patterns, especially in urban traffic where stop-and-go scenarios are prevalent. For Constant Velocity Model, it is too simple to predict complex scenario such as u-turns and turns. It seems like Constant Turn Rate and Velocity(CTRV) Model is suitable to handle that complex scenario, however, heading direction is sometimes different from that of the velocity. Force majeure external factors such as sensor error, wheel-slip may cause this problem, which would cause unpredictable consequences.

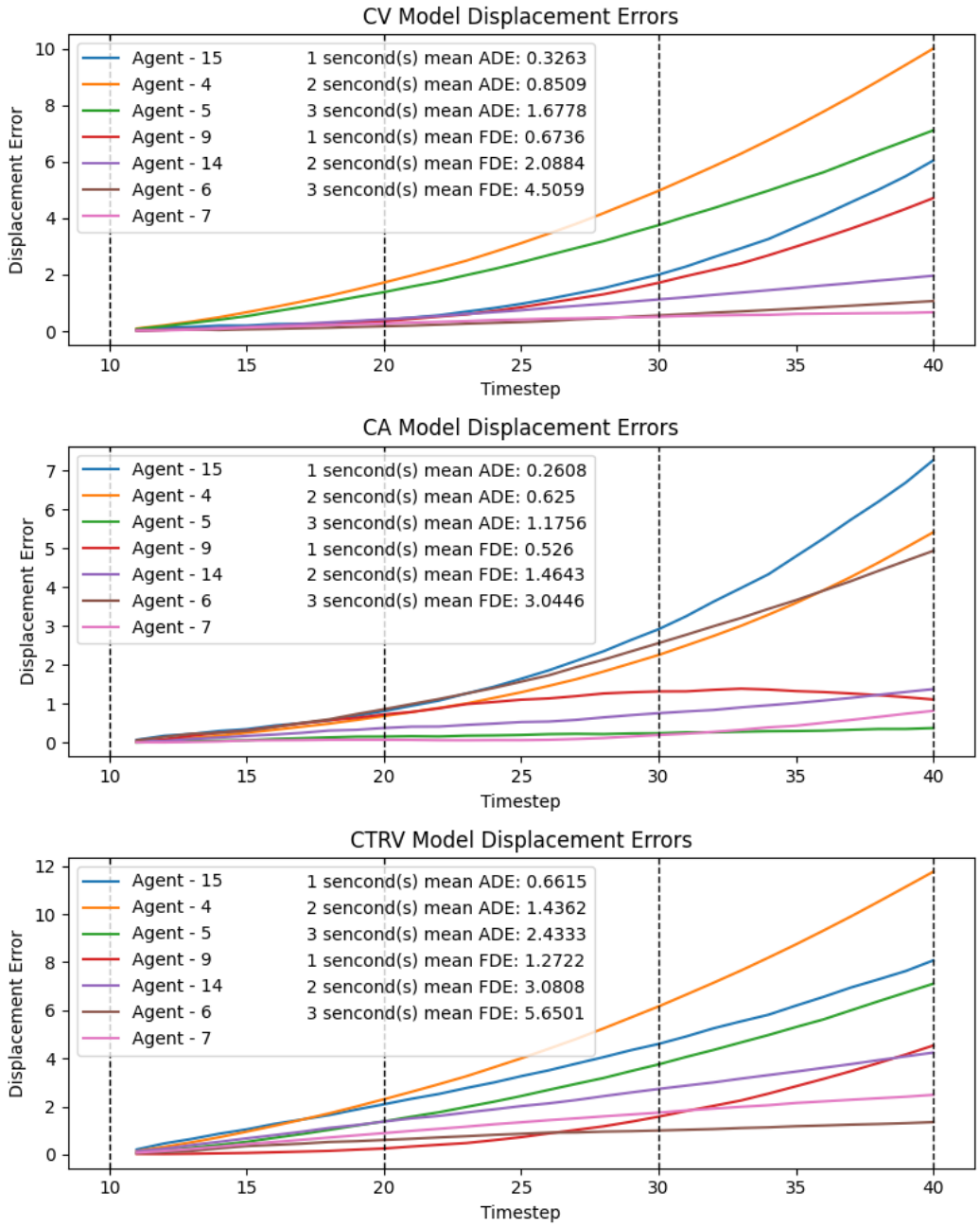


Figure 3: For three models, the 8 agents metrics visualizations of displacement errors between trajectories of predictions and ground truth.

3 Bonus Task: Single Object Tracking in ROS

3.1 Experimental Platform Construction

The most significant part in the structure of ROS is communication between ros nodes. After building a catkin workspace, ros core should be initiated first. Then a 'detector' node is turned on to publish the student matrix number, the ground truth and the detected target

through ros topics. It can also receive the raw image data by subscribing the topics sent by the provided ros bag.

3.2 Queue Deployment

Another ROS feature is the data transmission is parallel. It is unpredictable that which image from video is input first. As the proposed template matching method is relevant to time series, a queue is required to organized store all input images. To match each frame with ground truth data, a frame number counter is introduced. The counter will be added 1 after attaching the ground truth to the target image, so that it can be used as the index for selecting the ground truth in the next round.

3.3 Ros Bag Recording

After initiating every ros nodes, open a new terminal and input 'rosbag record' followed by the names of required topics. Then play the provided ros bag and halt the recording at the end of the play.