Appendix

α	MLT17		MLT19	
	Avg	Last	Avg	Last
1	75.4	66.5	71.5	64.4
5	76.3	68.2	71.9	64.0
10	76.8	68.1	72.1	62.7
15	**78.4**	**69.8**	**73.8**	**65.5**
20	77.5	67.4	72.8	64.3

Table 1. Ablation study on domain loss weights (α).

# Router	MLT17		MLT19	
	Avg	Last	Avg	Last
1	**78.4**	**69.8**	**73.8**	**65.5**
2	77.1	68.3	72.5	64.2
3	76.8	68.1	72.1	62.7

Table 2. Ablation study on the number of DM-Router modules.

A. Analyses on the Domain Loss Weights

The first is the α in manuscript Equ.3, which adjusts the contribution of language domain identification and text recognition. Tab. 1 shows the results on several enumerated α. The accuracy experiences a firstly increased and then decreased procedure, where the best accuracy is reached when α equals to 15. Text recognition loss plays a dominant role in MRN. This is reasonable as text recognition is the main task while it accumulates loss at the character level. This also indicates that the classification of language domains also contributes to joint optimization.

B. Analyses on the Number of DM-Router

In DM-Router module, one might guess that stacking the internal language dependence exploration parts several times may drive a better dependence utilization. We also empirically validate this and Tab.2 gives the results. As seen, the results answer that performing DM-Router module once is sufficient to explore the language dependence. It in turn proves the rationality of the DM-Router structure.

C. visualizes on the DM-Router

In Fig. 1, we visualize three examples, each containing a different language. DM-Router gives considerable well language identification, and each language all gives its recognition.

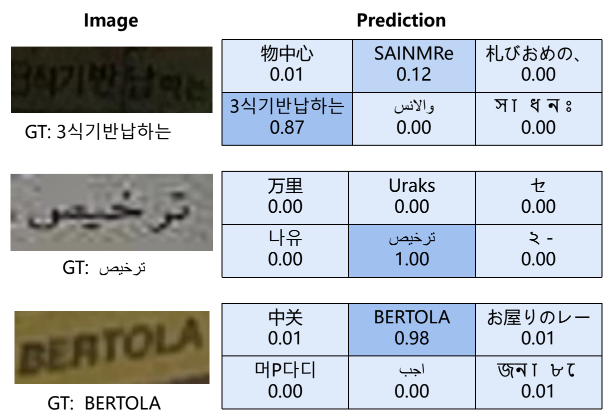

Figure 1. Illustrative examples with their predictions and domain scores.

D. Accuracy of Chinese Data at Each Task

To analyze the degree of forgetting the same language by different incremental methods, we show the accuracy of Chinese data tested on MLT17 and MLT19 for different combinations of text recognition and incremental learning methods.MLT17 and MLT19 show the same trend. Incremental learning methods (except MRN and DER) are completely unable to maintain the memory of the old language due to the rehearsal-imbalance. DER memorability depends on a larger memory budget, as the memory budget decreases, DER is unable to maintain the memorability of the old language. MRN maintains a stable performance advantage and is less dependent on the rehearsal memory budget.

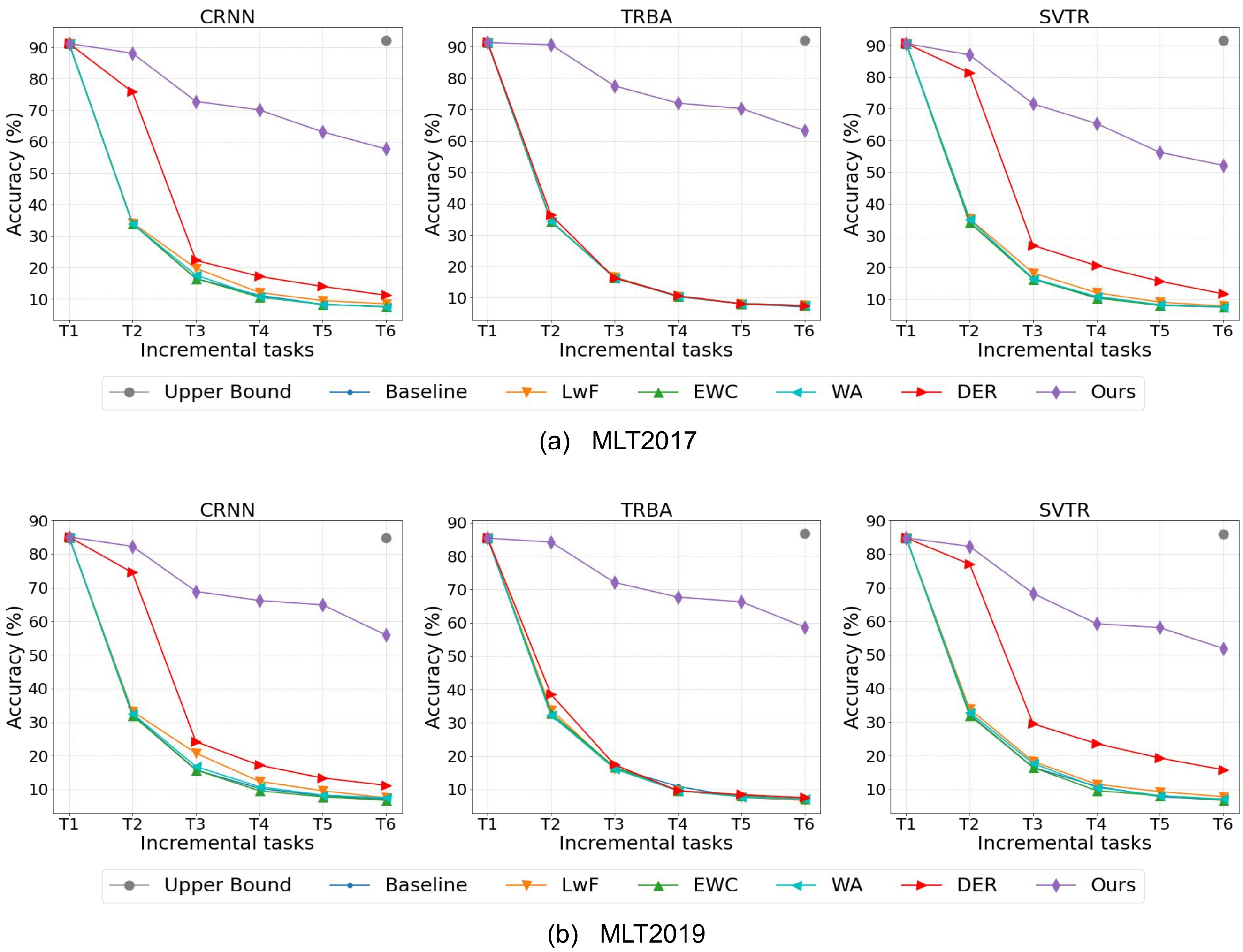

Figure 2. The Accuracy (%) of Chinese data on MLT17 and MLT19 tested against different combinations of text recognition and incremental learning methods.