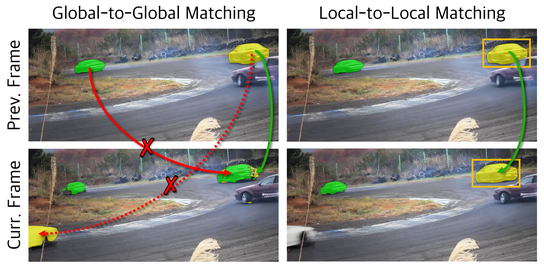
Efficient Regional Memory Network for Video Object Segmentation

Haozhe Xie, Hongxun Yao, Shangchen Zhou, Shengping Zhang, Wenxiu Sun

CVPR

Project Page
https://git.io/RMNet

Motivation



Two Typical Feature Matching Errors

- The target object in the current frame matches to the wrong object in the past frame *(solid red line)*
- The target object in the past frame matches to the wrong object in the current frame *(dotted red line)*

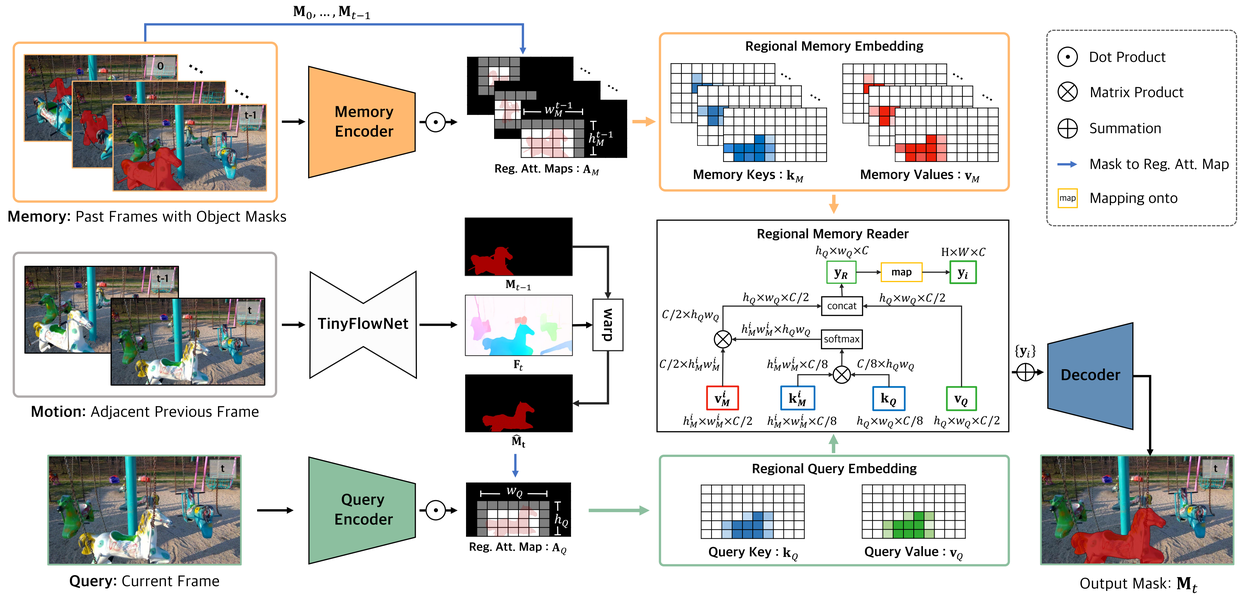
Observation

- The target objects appear ONLY in small regions in each frame

Solution

- Perform local-to-local feature matching in regions containing target objects

The Proposed Method: RMNet



Quantitative Evaluation

DAVIS 2017 val set

Method	$\mathcal{J}$ Mean	$\mathcal{F}$ Mean	Avg.
PReMVOS	0.739	0.817	0.778
STM	0.792	0.843	0.818
EGMN	0.800	0.859	0.829
CFBI	0.791	0.846	0.819
RMNet	**0.810**	**0.860**	**0.835**

DAVIS 2017 test-dev set

Method	$\mathcal{J}$ Mean	$\mathcal{F}$ Mean	Avg.
STM	0.680	0.740	0.710
CFBI	0.711	**0.785**	0.748
RMNet	**0.719**	0.781	**0.750**

YouTube-VOS val set (2018 version)

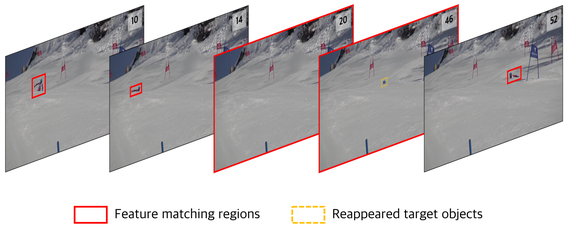
Method	$\mathcal{J}$ Mean (Seen)	$\mathcal{F}$ Mean (Seen)	$\mathcal{J}$ Mean (Unseen)	$\mathcal{F}$ Mean (Unseen)	Avg.
STM	0.797	0.842	0.728	0.809	0.794
EGMN	0.807	0.851	0.740	0.809	0.802
CFBI	0.811	**0.858**	0.753	**0.834**	0.814
RMNet	**0.821**	0.857	**0.757**	0.824	**0.815**

Qualitative Evaluation

Contribution

- We propose Regional Memory Network (RMNet) for semi-supervised VOS, which memorizes and tracks the regions containing target objects. RMNet effectively alleviates the ambiguity of similar objects.
- We present Regional Memory Reader that performs local-to-local matching between object regions in the past and current frames, which reduces the computational complexity.
- Experimental results on the DAVIS and YouTube-VOS datasets indicate that the proposed RMNet outperforms the state-of-the-art methods with much faster running speed.

Fallback to Global Matching for Occlusion



Object Size vs. Image Size