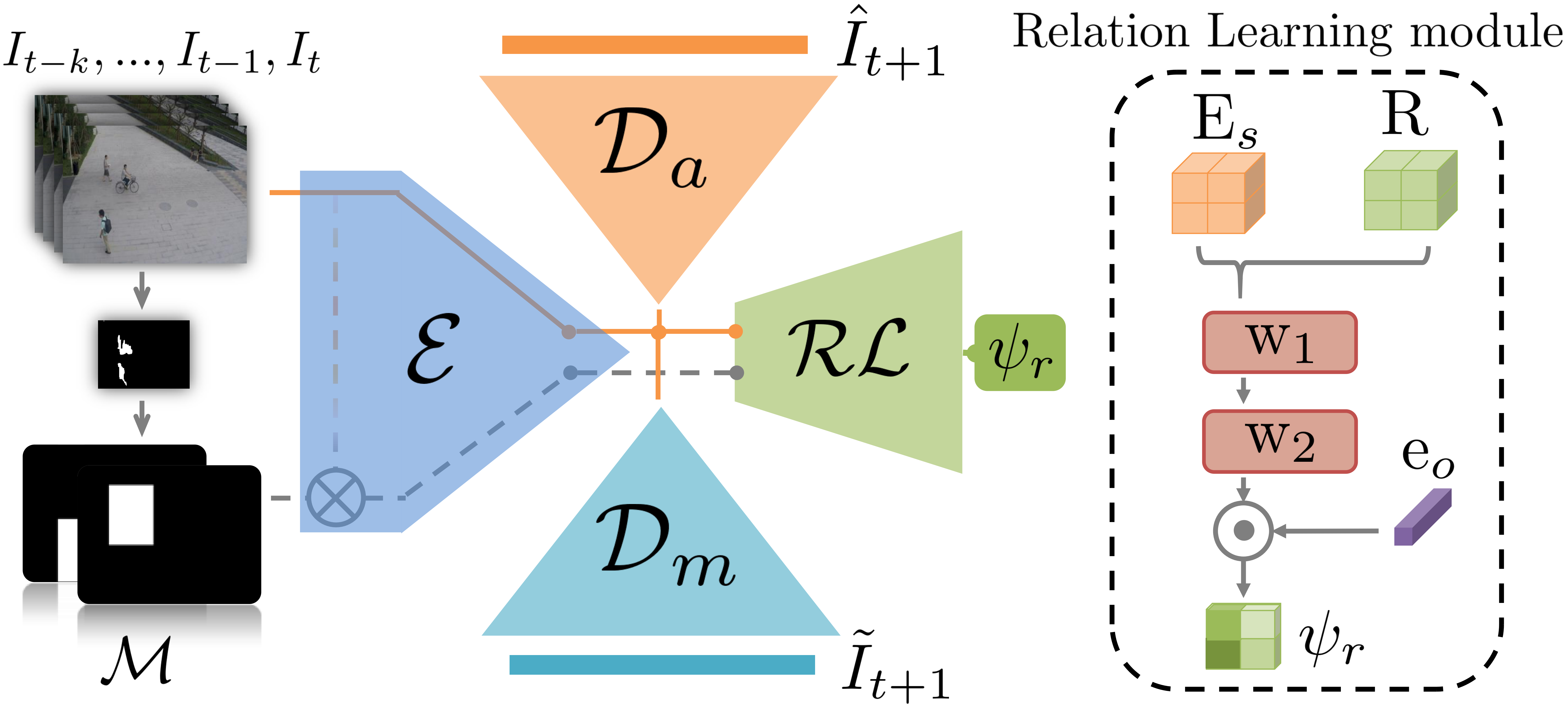

⊗: Element-wise multiplication ⊙: Dot production $\mathcal{S}_{\mathrm{rl}}$: Relation score

$\tilde{I}$: Warped future frame $\mathcal{E}$: Encoder $\mathcal{M}$: Region mask set

$\hat{I}$: Predicted future frame $\mathcal{D}$: Decoder $\mathrm{E}_s$: Scene Embedding

$I$: Ground-truth frame w: Conv filter $\psi_r$: Relation score map

$\mathcal{RL}$: Relation Learning module e$_o$: Object concensus embedding