Decoupling Features in Hierarchical Propagation for Video Object Segmentation

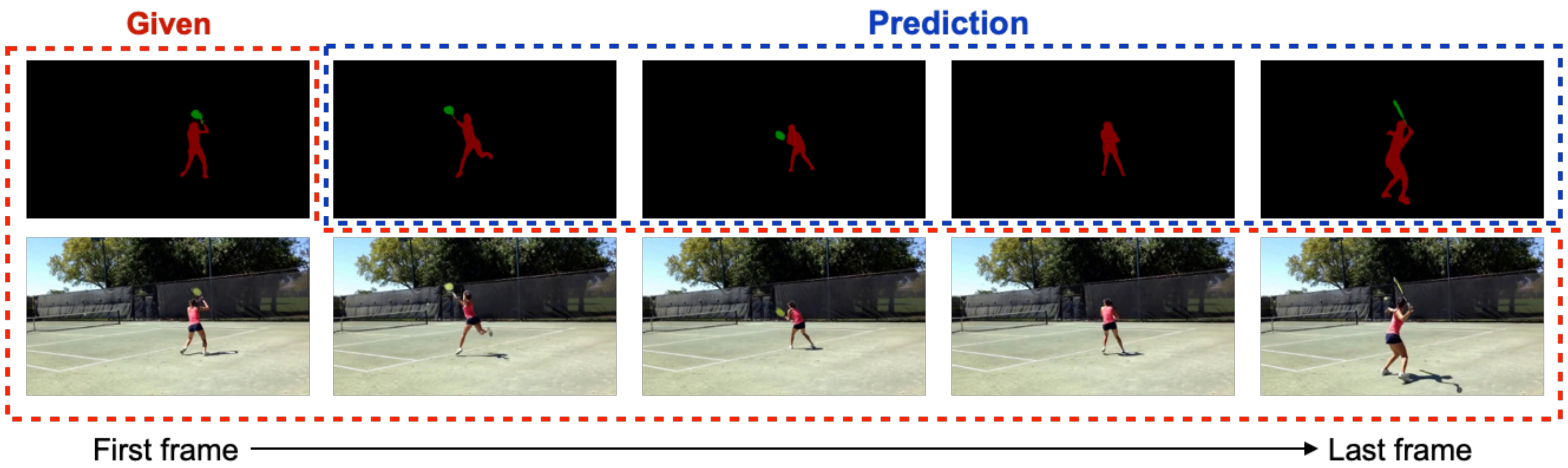
Zongxin Yang, Yi Yang

浙江大學
ZHEJIANG UNIVERSITY

Baidu Research

Task

Semi-supervised Video Object Segmentation (VOS)



Single-object results

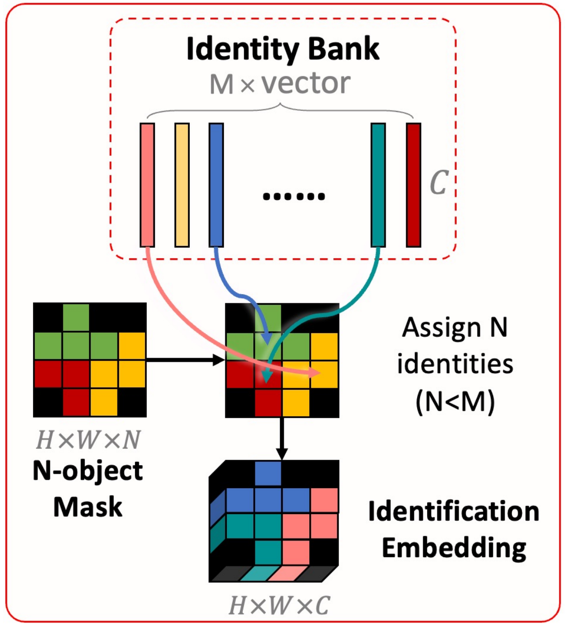
Multi-object (panoptic) results

ZHEJIANG UNIVERSITY

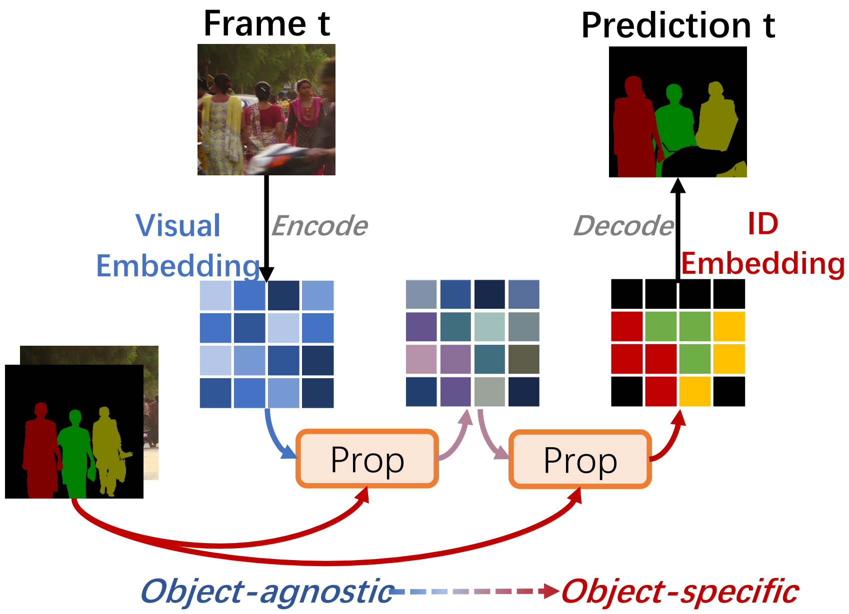
Baidu Research

Revisit Hierarchical Propagation for VOS

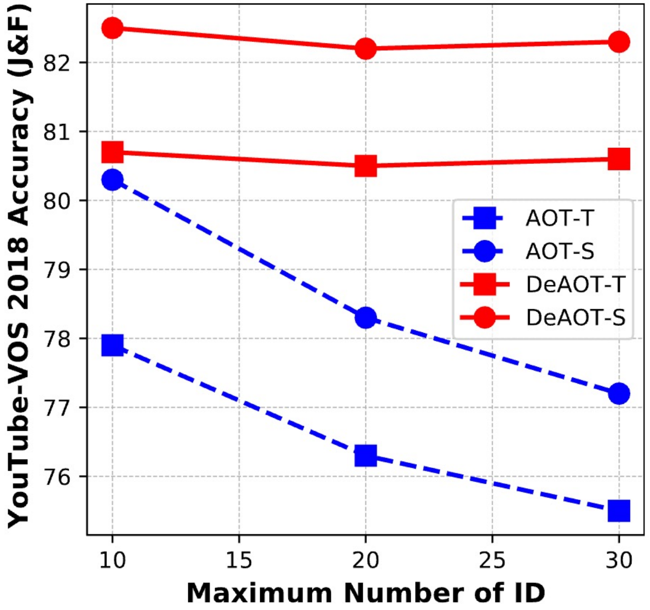
Absorbing the ID information leads to the oblivion of visual information



IDentification (ID) embedding



AOT[1]-like propagation



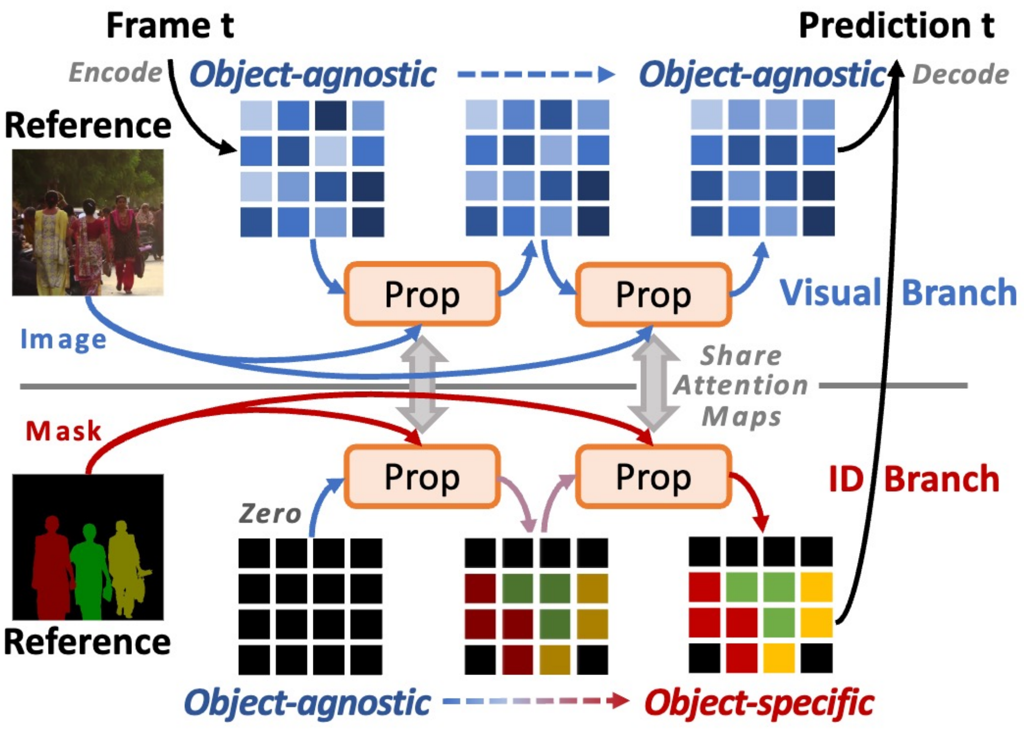
In AOT, more ID information, worse accuracy

[1] Yang, Zongxin, Yunchao Wei, and Yi Yang. "Associating objects with transformers for video object segmentation." NeurIPS 2021

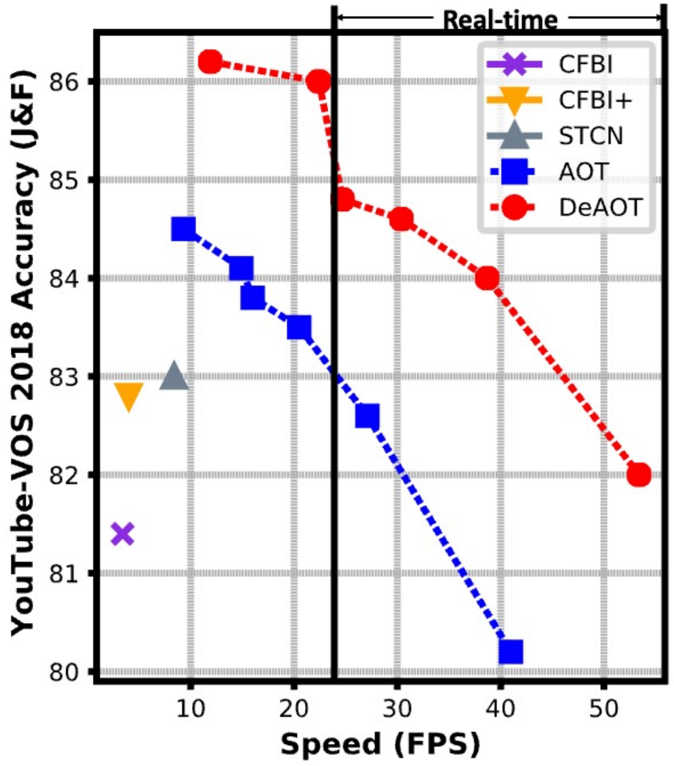
浙江大學 ZHEJIANG UNIVERSITY Baidu Research

Our Solution: Decoupling Features

Decouple object-agnostic and object-specific informations



Decoupling Features in Hierarchical Propagation (DeAOT)



DeAOT variants achieve superior accuracy and efficiency

浙江大学 ZHEJIANG UNIVERSITY

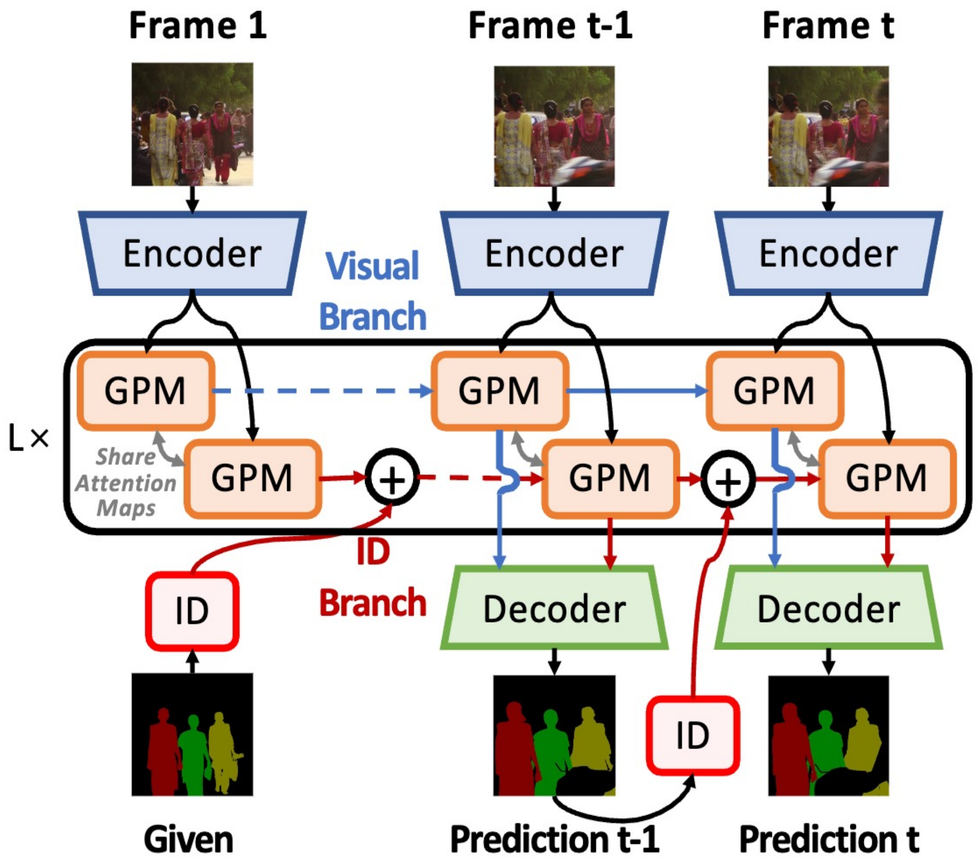
Baidu Research

Decoupling Features in Hierarchical Propagation

Overview

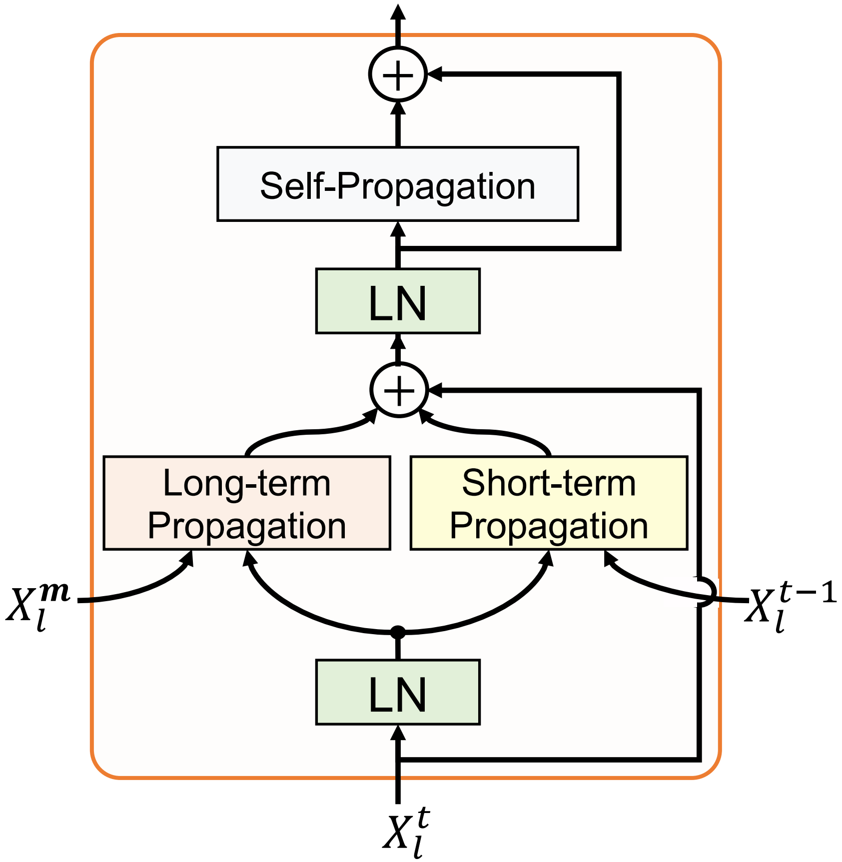
Our DeAOT decouples the propagation of visual embedding and ID embedding in two branches, i.e., **Visual Branch** and **ID Branch**.

The efficient propagation module, Gated Propagation Module (**GPM**) , shares attention maps between two branches.

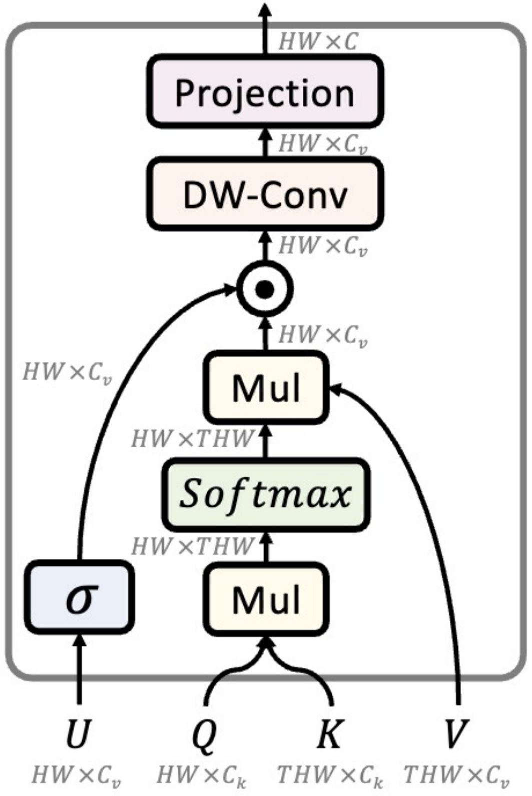


Gated Propagation Module

For efficient hierarchical propagation



Gated Propagation Module



Gated Propagation Function

	Robustness	Computation
Multi-head attention	Good	Heavy
Single-head attention	Limited	Light
Gated propagation	Good	Light

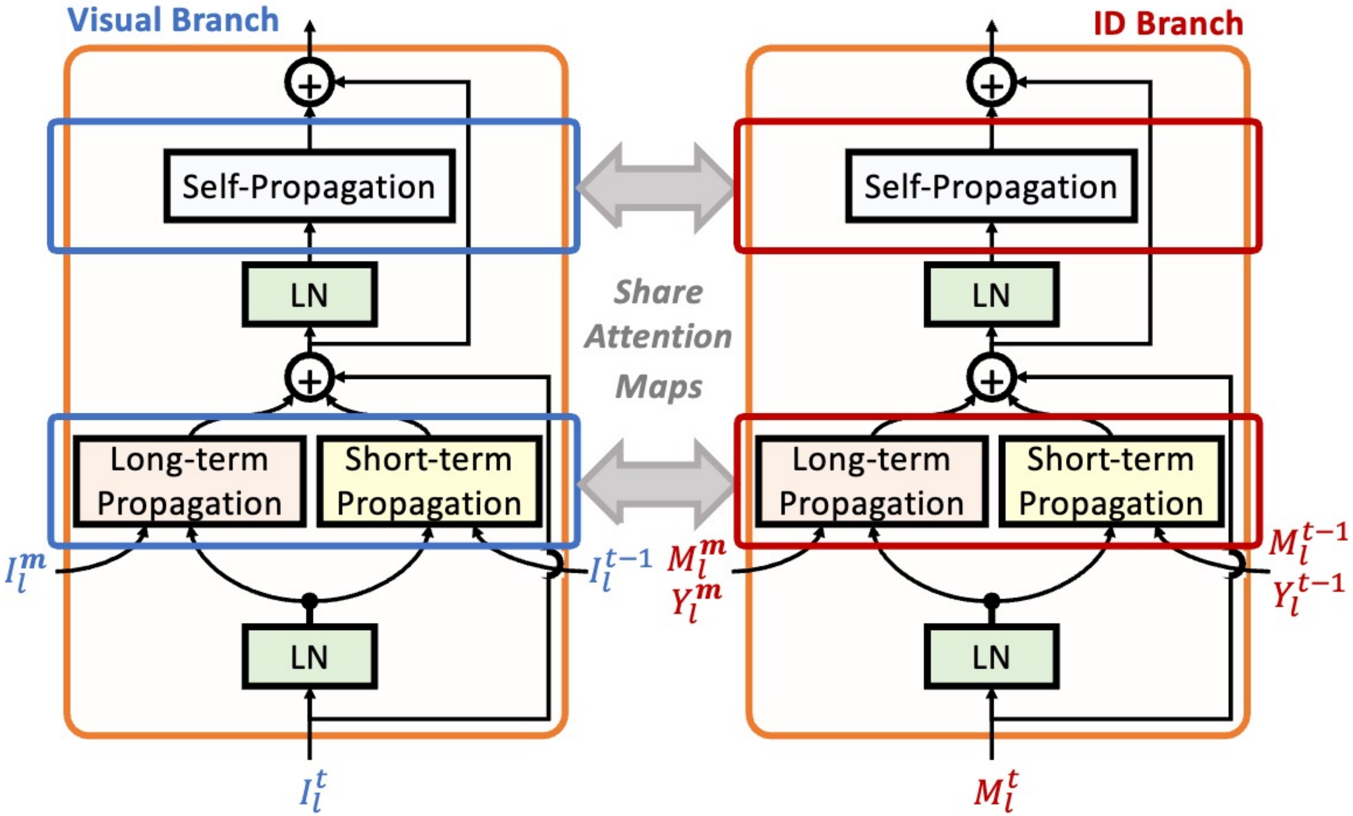
Gated propagation improves single-head attention by light-weight gated process and depth-wise convolution (DW-Conv)

ZHEJIANG UNIVERSITY 浙江大學

Baidu Research

Dual-branch Propagation

For decoupling visual and identification embeddings



- **Visual Branch**: calculate attention maps, propagate visual embedding
- **ID Branch**: reuse the attention maps from Visual Branch, propagate ID embedding

ZHEJIANG UNIVERSITY

Baidu Research

Results: Multi-object benchmarks

Compare DeAOT variants with SOTA methods

YouTube-VOS (large-scale) DAVIS 2017 (small-scale)

	YouTube-VOS 2018 Val					YouTube-VOS 2019 Val						DAVIS-17 Val			DAVIS-17 Test			
Method	Avg	$\mathcal{J}_S$	$\mathcal{F}_S$	$\mathcal{J}_U$	$\mathcal{F}_U$	Avg	$\mathcal{J}_S$	$\mathcal{F}_S$	$\mathcal{J}_U$	$\mathcal{F}_U$	fps	Avg	$\mathcal{J}$	$\mathcal{F}$	Avg	$\mathcal{J}$	$\mathcal{F}$	fps
KMN[ECCV20] [43]	81.4	81.4	85.6	75.3	83.3	-	-	-	-	-	-	82.8	80.0	85.6	77.2	74.1	80.3	-
CFBI[ECCV20] [62]	81.4	81.1	85.8	75.3	83.4	81.0	80.6	85.1	75.2	83.0	3.4	81.9	79.3	84.5	76.6	73.0	80.1	2.9
SST[CVPR21] [17]	81.7	81.2	-	76.0	-	81.8	80.9	-	76.6	-	-	82.5	79.9	85.1	-	-	-	-
HMMN[ICCV21] [44]	82.6	82.1	87.0	76.8	84.6	82.5	81.7	86.1	77.3	85.0	-	84.7	81.9	87.5	78.6	74.7	82.5	$3.4^{\ddagger}$
CFBI+[TPAMI21] [64]	82.8	81.8	86.6	77.1	85.6	82.6	81.7	86.2	77.1	85.2	4.0	82.9	80.1	85.7	78.0	74.4	81.6	3.4
STCN[NeurIPS21] [11]	83.0	81.9	86.5	77.9	85.7	82.7	81.1	85.4	78.2	85.9	$8.4^{\star}$	85.4	82.2	88.6	76.1	72.7	79.6	$19.5^{\star}$
RPCM[AAAI22] [58]	84.0	83.1	87.7	78.5	86.7	83.9	82.6	86.9	79.1	87.1	-	83.7	81.3	86.0	79.2	75.8	82.6	-
AOT-T [63]	80.2	80.1	84.5	74.0	82.2	79.7	79.6	83.8	73.7	81.8	41.0	79.9	77.4	82.3	72.0	68.3	75.7	51.4
DeAOT-T	**82.0**	**81.6**	**86.3**	**75.8**	**84.2**	**82.0**	**81.2**	**85.6**	**76.4**	**84.7**	**53.4**	**80.5**	**77.7**	**83.3**	**73.7**	**70.0**	**77.3**	**63.5**
AOT-S [63]	82.6	82.0	86.7	76.6	85.0	82.2	81.3	85.9	76.6	84.9	27.1	**81.3**	**78.7**	**83.9**	73.9	70.3	77.5	40.0
DeAOT-S	**84.0**	**83.3**	**88.3**	**77.9**	**86.6**	**83.8**	**82.8**	**87.5**	**78.1**	**86.8**	**38.7**	80.8	77.8	83.8	**75.4**	**71.9**	**79.0**	**49.2**
AOT-B [63]	83.5	82.6	87.5	77.7	86.0	83.3	82.4	87.1	77.8	86.0	20.5	**82.5**	**79.7**	**85.2**	75.5	71.6	79.3	29.6
DeAOT-B	**84.6**	**83.9**	**88.9**	**78.5**	**87.0**	**84.6**	**83.5**	**88.3**	**79.1**	**87.5**	**30.4**	82.2	79.2	85.1	**76.2**	**72.5**	**79.9**	**40.9**
AOT-L [63]	83.8	82.9	87.9	77.7	86.5	83.7	82.8	87.5	78.0	**86.7**	16.0	83.8	**81.1**	86.4	**78.3**	**74.3**	**82.3**	18.7
DeAOT-L	**84.8**	**84.2**	**89.4**	**78.6**	**87.0**	**84.7**	**83.8**	**88.8**	**79.0**	**87.2**	**24.7**	**84.1**	81.0	**87.1**	77.9	74.1	81.7	**28.5**
R50-AOT-L [63]	84.1	83.7	88.5	78.1	86.1	84.1	83.5	88.1	**78.4**	86.3	14.9	84.9	**82.3**	87.5	79.6	75.9	83.3	18.0
R50-DeAOT-L	**86.0**	**84.9**	**89.9**	**80.4**	**88.7**	**85.9**	**84.6**	**89.4**	**80.8**	**88.9**	**22.4**	**85.2**	82.2	**88.2**	**80.7**	**76.9**	**84.5**	**27.0**
SwinB-AOT-L [63]	84.5	84.3	89.3	77.9	86.4	84.5	84.0	88.8	78.4	86.7	9.3	85.4	82.4	88.4	81.2	77.3	85.1	12.1
SwinB-DeAOT-L	**86.2**	**85.6**	**90.6**	**80.0**	**88.4**	**86.1**	**85.3**	**90.2**	**80.4**	**88.6**	**11.9**	**86.2**	**83.1**	**89.2**	**82.8**	**78.9**	**86.7**	**15.4**

more

GPM Number

less

DeAOT-L: state-of-the-art

DeAOT-T: real-time

ZHEJIANG UNIVERSITY

Baidu Research

Results: Single-object benchmarks

Compare DeAOT variants with SOTA methods

DAVIS 2016:
Video Object Segmentation

VOT 2020:
Visual Object Tracking

	DAVIS 2016				VOT 2020	
Method	Avg	$\mathcal{J}$	$\mathcal{F}$	fps	EAO	EAO^{RT}
CFBI+ [64]	89.9	88.7	91.1	5.9	-	-
RPCM [58]	90.6	87.1	94.0	5.8	-	-
HMMN [44]	90.8	89.6	92.0	10.0	-	-
STCN [11]	91.6	90.8	92.5	27.2*	-	-
AlphaRef [59]	-	-	-	-	0.482	0.486
RPT [33]	-	-	-	-	0.530	0.290
MixFormer-L [14]	-	-	-	-	0.555	-
AOT-T [63]	86.8	86.1	87.4	51.4	0.435	0.433
DeAOT-T	**88.9**	**87.8**	**89.9**	**63.5**	**0.472**	**0.463**
AOT-S [63]	**89.4**	**88.6**	90.2	40.0	0.512	0.499
DeAOT-S	89.3	87.6	**90.9**	**49.2**	**0.593**	**0.559**
AOT-B [63]	89.9	88.7	91.1	29.6	0.541	0.533
DeAOT-B	**91.0**	**89.4**	**92.5**	**40.9**	**0.571**	**0.542**
AOT-L [63]	90.4	89.6	91.1	18.7	0.574	0.560
DeAOT-L	**92.0**	**90.3**	**93.7**	**28.5**	**0.591**	**0.554**
R50-AOT-L [63]	91.1	90.1	92.1	18.0	0.569	0.540
R50-DeAOT-L	**92.3**	**90.5**	**94.0**	**27.0**	**0.613**	**0.571**
SwinB-AOT-L [63]	92.0	90.7	93.3	12.1	0.586	0.523
SwinB-DeAOT-L	**92.9**	**91.1**	**94.7**	**15.4**	**0.622**	**0.559**

more

GPM
Number

less

DeAOT-L:
state-of-the-art

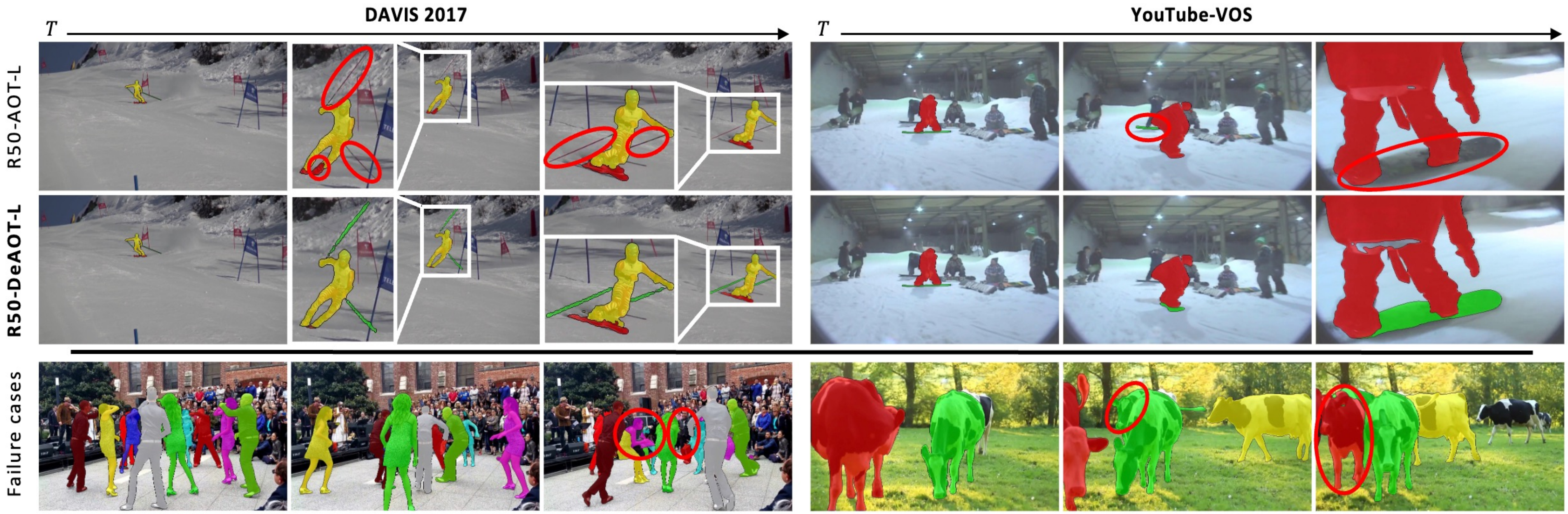
DeAOT-T:
real-time

ZHEJIANG UNIVERSITY 浙江大学

Baidu Research

Results: Qualitative Results

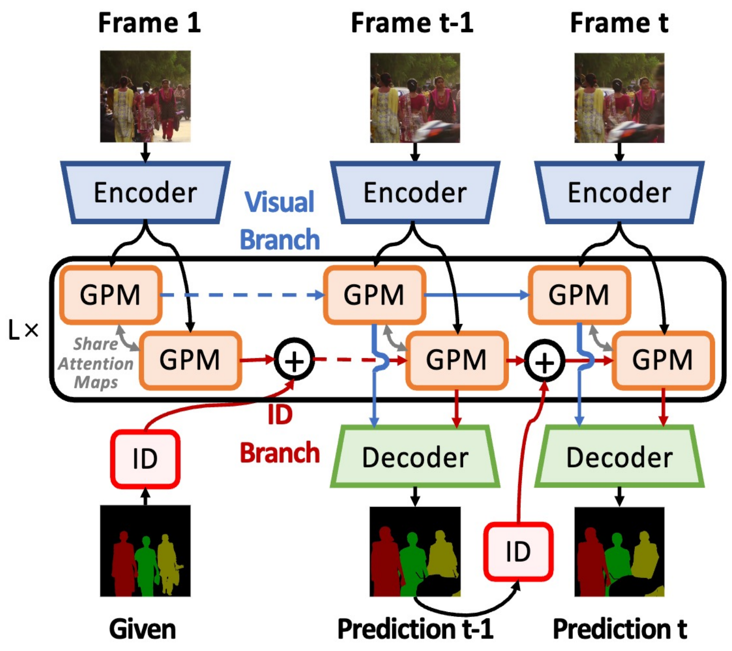
Compare DeAOT variants with SOTA methods



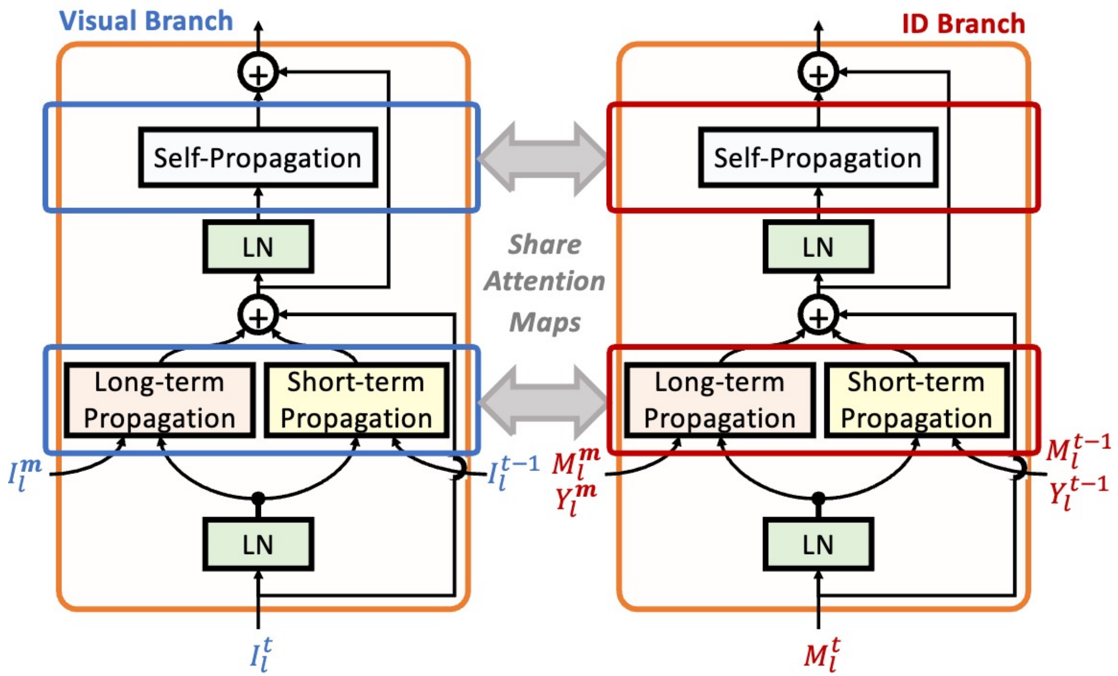
浙江大學
ZHEJIANG UNIVERSITY
Baidu Research

Decoupling Features in Hierarchical Propagation

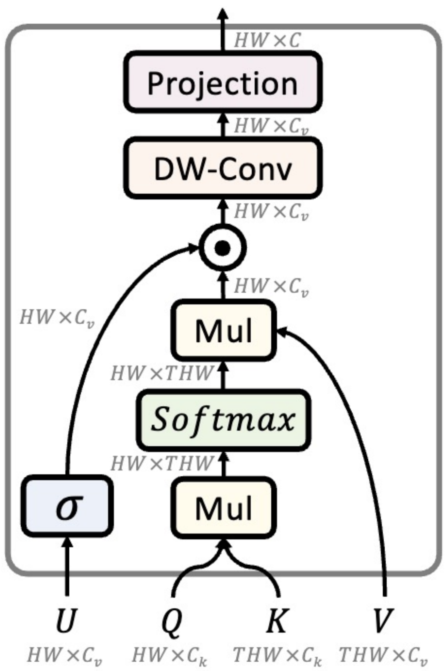
Overview



Dual-branch Propagation



Gated Propagation



ZHEJIANG UNIVERSITY

Baidu Research