Video Salient Object Detection via Cross-frame Cellular Automata

Jingfan Guo, Tongwei Ren, Lei Huang, Xingyu Liu, Ming-Ming Cheng, Gangshan Wu

Background

Video salient object detection (VidSOD) aims to detect the most attractive objects for human beings in a given video.

Compared to image salient object detection (ImgSOD), VidSOD faces some distinct challenges:

- **Object motion** usually plays an important role in VidSOD, which is ignored in ImgSOD
- Object appearances may change over time in videos, which makes it difficult to obtain **coherent salient objects** over frames

Method

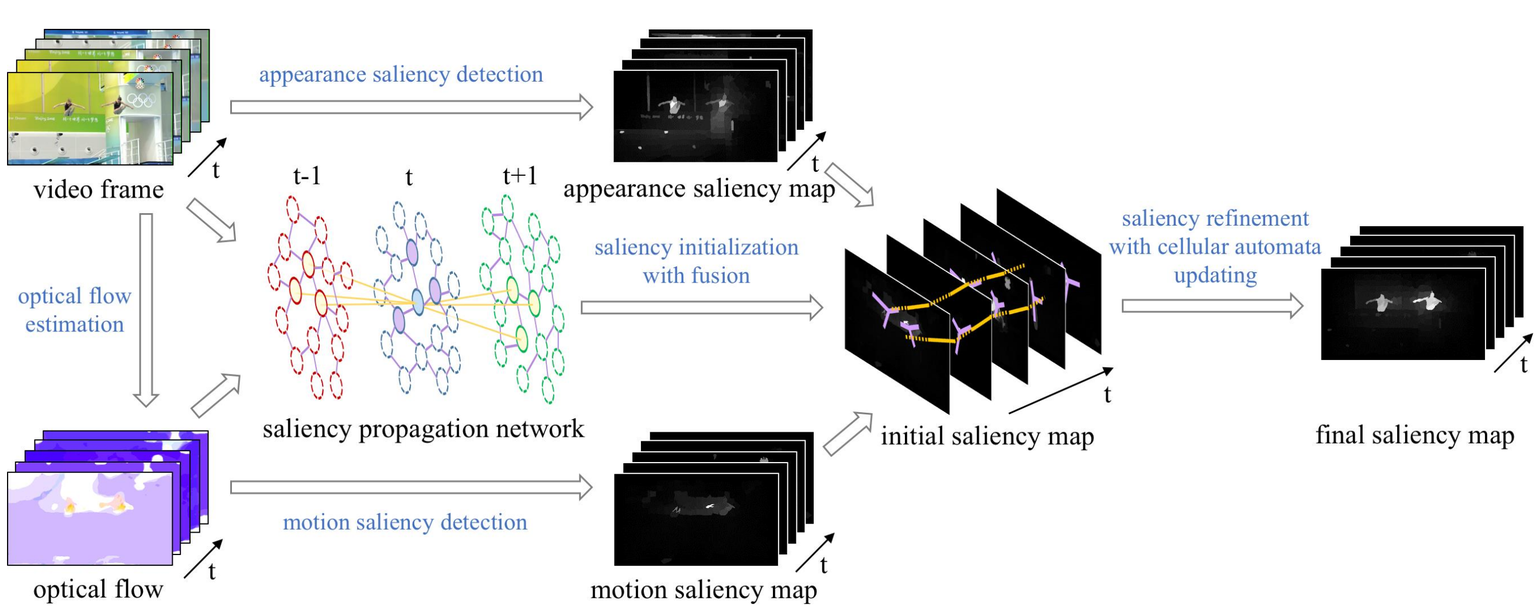
Basic idea

Given a video, we construct a graph model called saliency propagation network based on spatio-temporal connectivity, fuse appearance saliency and motion saliency to initiate the propagation network, and iteratively update the network by saliency propagation between neighboring nodes

Contribution

- We firstly apply cellular automata in VidSOD by constructing a cross-frame saliency propagation network
- We utilize both appearance and motion features followed by entropy-based adaptive fusion in saliency initialization
- We construct a VidSOD dataset NoMot, which consist of 10 videos without obvious camera motion or object motion

Framework

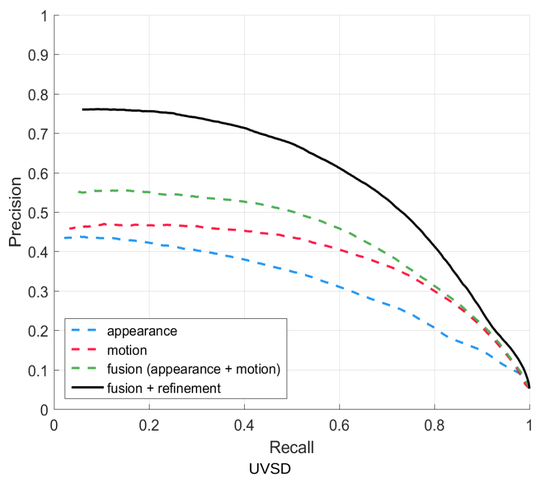


- **Pre-processing:** super-pixel segmentation and optical flow estimation
- **Propagation network construction:** connect adjacent super-pixels in the same frame and connect matched ones in two adjacent frames
- **Saliency map initialization:** detect appearance saliency and motion saliency by color contrast and geodesic distance of motion histograms, respectively, and fuse them with entropy-based adaptive fusion strategy
- **Saliency refinement:** iteratively update the propagation network by cellular automata updating

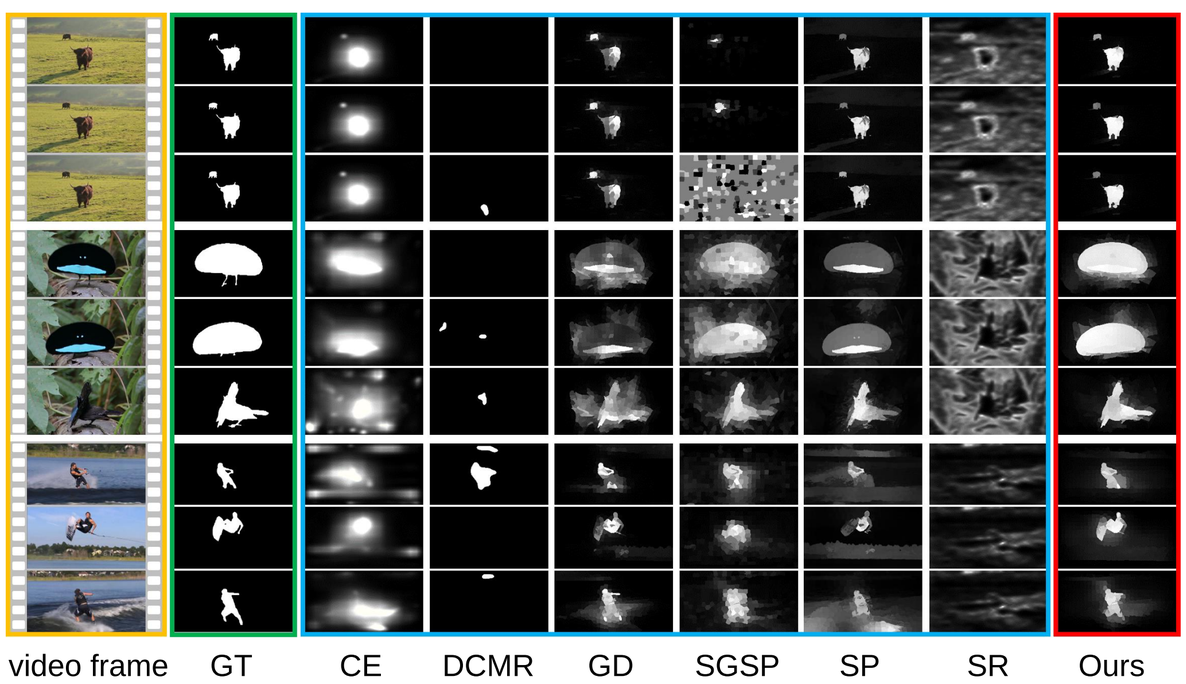
Experiments

Ablation study

We evaluate the performance of each step of our method on UVSD dataset, and show that all steps of our method are effective using PR curves

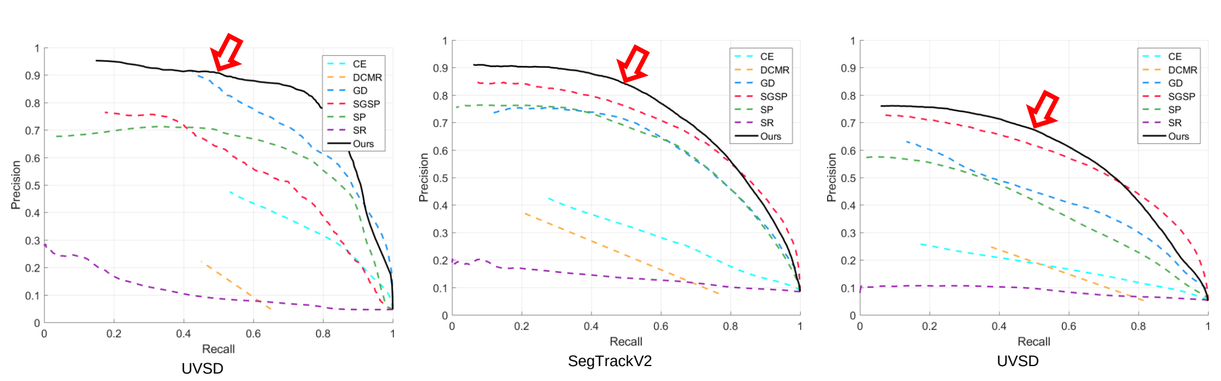


Result examples



Comparison with state-of-the-art methods

	NoMot		SegTrackV2		UVSD		average	
	F_β^ω	Time(s)	F_β^ω	Time(s)	F_β^ω	Time(s)	F_β^ω	Time(s)
CE (Matlab)	0.1652	3.9397	0.1826	3.4351	0.1237	2.9538	0.1572	3.4429
DCMR (C++)	0.1363	0.0465	0.147	0.0413	0.148	0.0426	0.1438	0.0435
GD (Matlab)	0.3243	7.2564	**0.3557**	8.7907	0.2338	11.2767	0.3046	9.1079
SGSP (Matlab)	0.1994	6.9027	0.3045	11.2483	0.2021	10.6294	0.2353	9.5935
SP (Matlab)	0.27	7.5342	0.2662	11.6843	0.1771	11.0561	0.2358	10.0915
SR (Matlab)	0.0549	0.1811	0.0883	0.1576	0.0662	0.1524	0.0698	0.1637
Ours (Matlab)	**0.4832**	7.1523	0.3434	9.4792	**0.2631**	9.1834	**0.3632**	8.605



- Our method outperforms other methods on all the data sets except that GD performs better on SegTrackV2
- The efficiency of our method is similar to those of SGSP and GD since we all rely on LDOF to compute optical flow, which occupies most of the computation cost

Contact to: rentw@nju.edu.cn

MAGUS