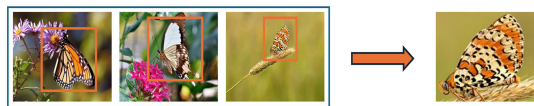


Background

Practical applications of the research direction:

- Collaborative thumbnail: Generate a representative cover or thumbnail for photo albums, e-commerce product groups, news image collections, etc.
- Image search: When a user inputs a query image, the system automatically searches for related images.



(a) Image search

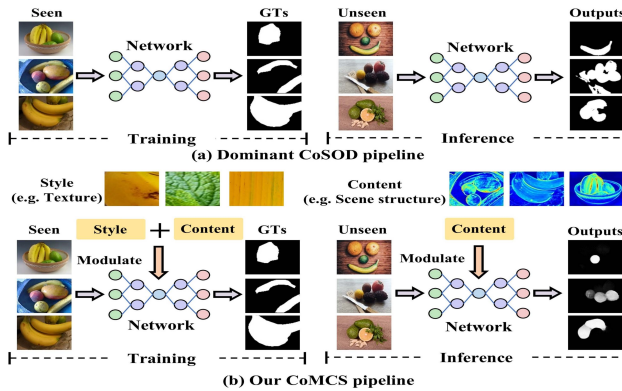


(b) Collaborative thumbnail



Co-salient object detection

Motivation



Simultaneously modulate both the **style** and the **content** information

Experiment

Methods	CoCA				CoSOD3k				CoSal2015			
	$S_\alpha \uparrow$	$F_\beta^{max} \uparrow$	$E_\phi^{max} \uparrow$	$MAE \downarrow$	$S_\alpha \uparrow$	$F_\beta^{max} \uparrow$	$E_\phi^{max} \uparrow$	$MAE \downarrow$	$S_\alpha \uparrow$	$F_\beta^{max} \uparrow$	$E_\phi^{max} \uparrow$	$MAE \downarrow$
GICD(ECCV2020)	0.658	0.513	0.715	0.126	0.797	0.770	0.848	0.079	0.844	0.844	0.887	0.071
DeepACG(CVPR2021)	0.688	0.552	0.771	0.102	0.792	0.756	0.838	0.089	0.854	0.842	0.892	0.064
CADC(ICCV2021)	0.681	0.548	0.744	0.132	0.801	0.840	0.859	0.096	0.866	0.862	0.906	0.064
GCoNet(CVPR2021)	0.673	0.544	0.760	0.105	0.802	0.777	0.860	0.071	0.845	0.847	0.887	0.068
DCFm(CVPR2022)	0.710	0.598	0.783	0.085	0.810	0.805	0.874	0.067	0.838	0.856	0.892	0.067
CoEGNet(TPAMI2022)	0.612	0.493	0.717	0.106	0.762	0.736	0.825	0.092	0.836	0.832	0.882	0.077
UFO(TMM2023)	0.697	0.571	0.782	0.095	0.819	0.797	0.874	0.073	0.860	0.865	0.906	0.064
TCNet(TCSVT2023)	0.685	0.552	0.777	0.101	0.832	0.804	0.878	0.068	0.870	0.864	0.911	0.054
CADC+(PAMI2023)	0.708	0.609	0.791	0.107	0.823	0.808	0.876	0.070	0.875	0.889	0.922	0.060
CoRP(TPAMI2023)	0.703	0.575	0.741	0.110	0.820	0.794	0.864	0.075	0.859	0.864	0.896	0.060
MCC(LAAI2023)	0.714	0.590	0.796	0.103	0.856	0.837	0.903	0.061	0.890	0.891	0.927	0.051
GCoNet+(PAMI2023)	0.738	0.637	0.812	0.081	0.843	0.834	0.901	0.062	0.881	0.891	0.924	0.056
UniTR(TMM2024)	0.708	0.574	0.789	0.089	0.853	0.834	0.900	0.072	0.860	0.857	0.906	0.063
ICSM(TIM2024)	0.725	0.626	0.806	0.099	0.853	0.843	0.890	0.061	0.890	0.893	0.924	0.052
IPPO(VRIH2024)	0.747	0.644	0.810	0.084	0.836	0.819	0.889	0.065	0.877	0.889	0.922	0.055
ASCoD(ICASSP2025)	0.738	0.640	0.816	0.083	0.830	0.819	0.900	0.063	0.860	0.867	0.907	0.061
E2HCoD(ICASSP2025)	0.729	0.626	0.805	0.081	0.837	0.820	0.886	0.071	0.880	0.877	0.912	0.055
CoMCS	0.747	0.649	0.821	0.084	0.857	0.843	0.907	0.058	0.890	0.896	0.929	0.052

Table 1. Quantitative comparison between our CoMCS and other state-of-the-art methods on three regular benchmark datasets. Red and blue fonts indicate best and second-best performance, respectively.

Limitation

Similar object interference



(a) Lack of sufficient discrimination

Unknown category interference



(c) Lack of sufficient generalization

How to solve?

Method

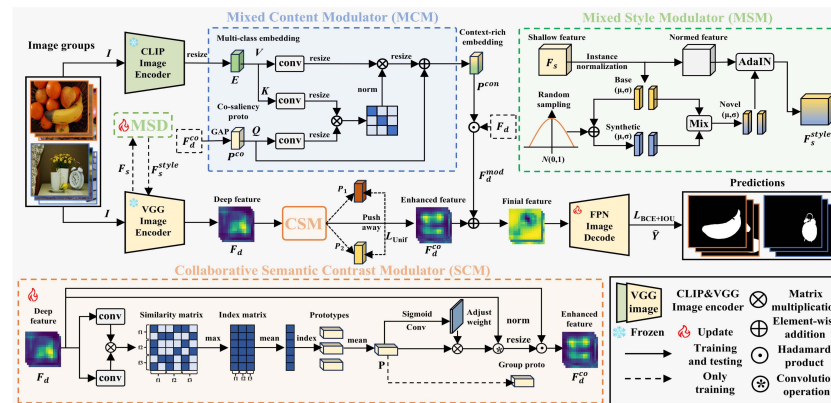
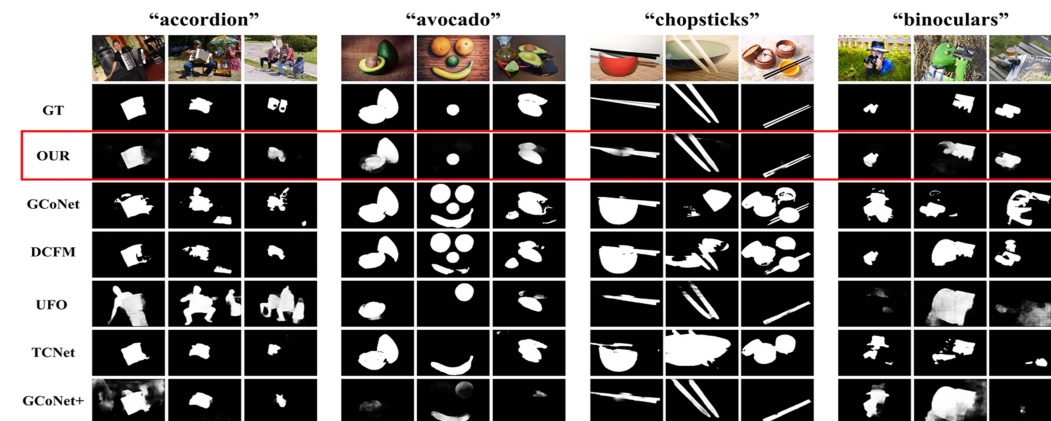


Figure 2. Pipeline of our CoMCS. The whole network consists of three main modules: MCM, MSM and SCM. Among them, MCM captures domain-invariant structural information by establishing collaborative object prototypes and multi-class semantic embeddings; MSM increases the style diversity of the training domain by simulating multiple styles; SCM is designed to capture cooperative features and reduce the interfering noise contained in them. Specifically, MSM is only used during training.



In the cooperative image groups containing numerous interfering objects unrelated to the co-objects, our CoMCS effectively identifies the true co-objects while eliminating these interferences. For example, in the **avocado** group, the CoMCS precisely identifies and excludes on cooperative objects such as **oranges and bananas**, which have highly similar appearances to the co-objects.