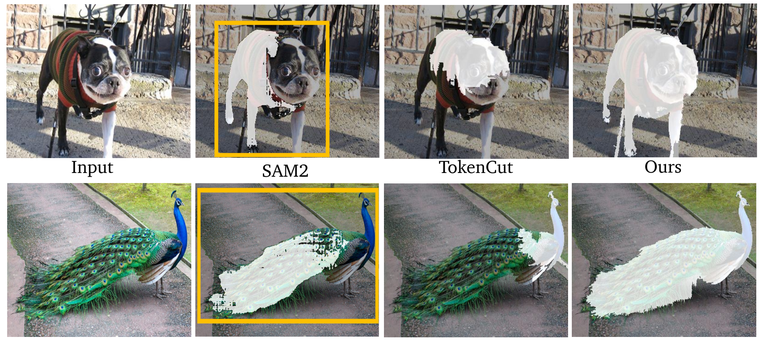
Wholly Unsupervised! Segmenting Objects by Contrast and Context

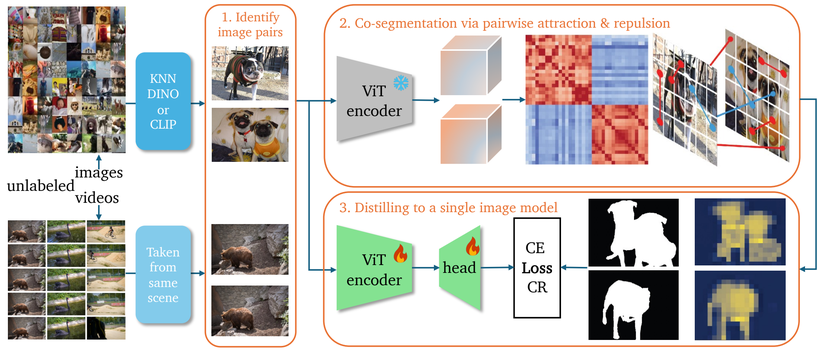
Fei Pan* Yixing Wang* Sangryul Jeon Stella X. Yu

NEURAL INFORMATION PROCESSING SYSTEMS

Unsupervised Whole Object Segmentation Remains Challenging

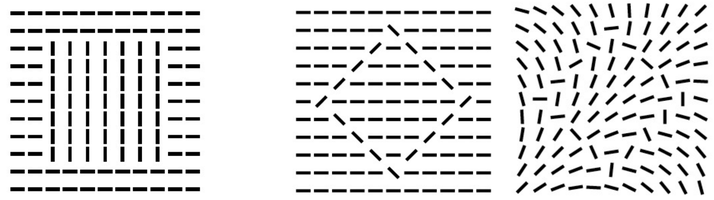


Our Method: Co-segmentation on Image pairs + Distillation



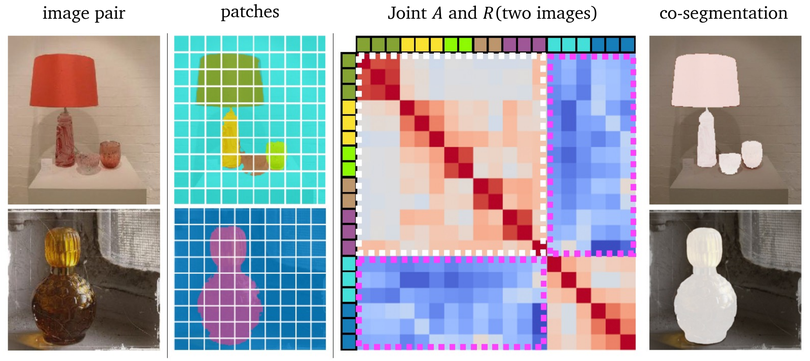
1. CCG Outperforms SAM2 on Saliency Detection

Objects Emerge via Self-similarity and Contextual Contrast



Homogenous parts cluster naturally. *The enemy of my enemy is my friend.*

Co-segmentation Provides Richer Figure-ground Relationship

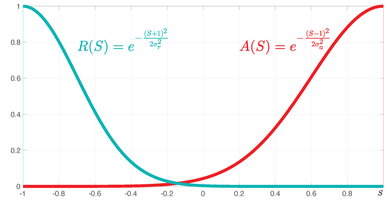


Graph Cut with Attraction and Repulsion

$S = \frac{\langle F_i, F_j \rangle}{||F_i|| \cdot ||F_j||}$: patch similarity

$A(S) = e^{-\frac{(S-1)^2}{2\sigma_a^2}}$: pairwise attraction

$R(S) = e^{-\frac{(S+1)^2}{2\sigma_r^2}}$: pairwise repulsion



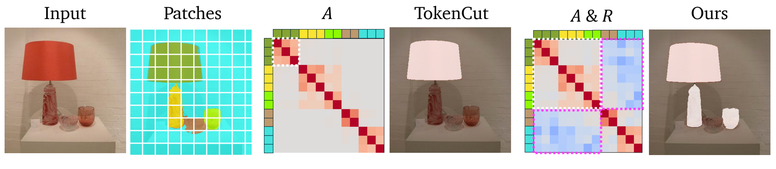
$$\max \quad \xi_{AR}(p) = \sum_{t=1}^{2} \frac{p_t^T W p_t}{p_t^T D p_t}$$

p_t is the binary indicator for group t

$W = A - R + D_R, \quad D = D_A + D_R$ $\quad D_A/D_R$: Diagonal degree matrix of A/R

$\Longrightarrow$ Find the largest eigenvector of $D^{-1}W$.

Single Image Segmentation with Attraction and Repulsion



CCG Outperforms Unsupervised Methods across Four Benchmarks

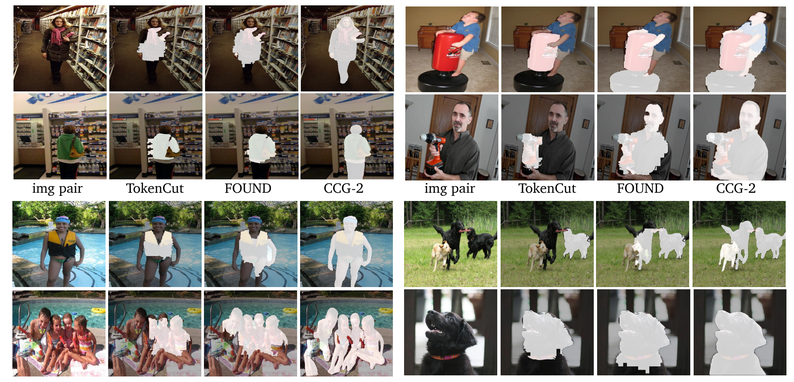
Saliency	train?	ViT	ECSSD			DUTS			DUT-OMRON		
			maxF$_\beta$	IoU	Acc.	maxF$_\beta$	IoU	Acc.	maxF$_\beta$	IoU	Acc.
FUIS	×		–	71.3	91.5	–	52.8	89.3	–	50.9	88.3
LOST	×	S/16	75.8	65.4	89.5	61.1	51.8	87.1	47.3	41.0	79.7
DSS	×	–	–	**73.3**	–	–	51.4	–	–	**56.7**	–
TokenCut	×	S/16	80.3	71.2	91.8	67.2	57.6	90.3	60.0	53.3	88.0
CCG-1	×	S/16	82.7	72.8	93.1	**69.5**	60.2	92.8	62.6	55.3	**90.7**
			+2.4	+0.6	+1.3	+2.3	+2.6	+2.5	+2.6	+2.0	+2.7
CCG-2	×	S/16	**83.1**	73.2	**94.7**	69.3	**60.5**	**93.2**	**63.3**	56.4	90.6
			+2.8	+2.0	+2.9	+2.1	+2.9	+2.9	+3.3	+3.1	+2.6
SelfMask	✓	S/8	–	78.1	94.4	–	62.6	92.3	–	58.2	90.1
FOUND	✓	S/8	**95.5**	80.7	94.9	71.5	64.5	93.8	66.3	57.8	91.2
PEEKABOO	✓	S/8	95.3	79.8	94.6	**86.0**	64.3	93.9	**80.4**	57.5	91.5
HEAP	✓	S/8	93.0	81.1	94.5	75.7	64.4	94.0	69.0	59.6	92.0
CCG-1	✓	S/8	94.1	83.6	95.2	78.0	65.9	**94.6**	70.7	60.8	93.5
			+1.1	+2.5	+0.7	+2.3	+1.5	+0.6	+1.7	+1.2	+1.5
CCG-2	✓	S/8	94.5	**83.9**	**95.8**	78.2	**66.5**	94.4	71.2	**61.3**	**93.8**
			+1.5	+2.8	+1.3	+2.5	+2.1	+0.4	+2.2	+1.7	+1.8

Unsupervised Nuclei Segmentation	accuracy	mIoU	F_1 score
UNSEG	43.6	41.4	48.2
CCG-1	58.3 (+14.7)	54.5 (+13.1)	57.9 (+9.7)
CCG-2	61.1 (+17.5)	56.9 (+15.5)	58.6 (+10.4)

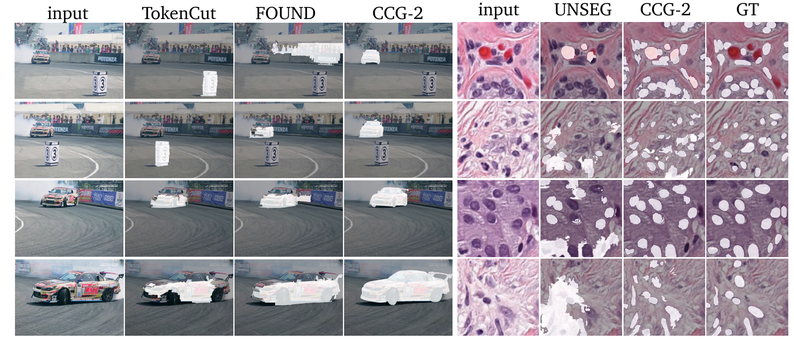
Object Discovery	train?	ViT	VOC07	VOC12	COCO20K
DINO-seg	×	S/16	45.8	46.2	42.0
LOST	×	S/16	61.9	64.0	50.7
DSS	×	S/16	62.7	66.4	52.2
TokenCut	×	S/16	68.8	72.1	58.8
CCG-1	×	S/16	71.4	**73.8**	60.3
			+2.6	+1.7	+1.5
CCG-2	×	S/16	**72.3**	73.7	**61.7**
			+3.5	+1.6	+2.9
SelfMask	✓	S/8	72.3	75.3	62.7
FOUND	✓	S/8	72.5	76.1	62.9
PEEKABOO	✓	S/8	72.7	75.9	64.0
HEAP	✓	S/8	73.2	77.1	63.4
CCG-1	✓	S/8	76.4	79.8	65.6
			+3.2	+2.7	+2.2
CCG-2	✓	S/8	**77.7**	**80.8**	**66.2**
			+4.5	+3.7	+2.8

VOS	train?	use flow?	DAVIS	FBMS	SegTV2
TokenCut	×	×	64.3	60.2	59.6
CCG-1	×	×	66.4	62.5	61.2
			+2.1	+2.3	+1.6
CCG-2	×	×	**67.9**	**64.1**	**62.1**
			+3.6	+3.9	+2.5
CIS	✓	✓	71.5	63.6	62.0
CMC	✓	✓	**75.4**	66.8	62.6
AMD	✓	×	45.7	28.7	42.9
VideoCutLER	✓	×	68.4	64.6	62.5
CCG-1	✓	×	71.8	66.4	64.5
			+3.4	+1.8	+2.0
CCG-2	✓	×	72.4	**67.9**	**66.1**
			+4.0	+3.3	+3.6

2. CCG Surpasses Baselines by 3-5% on Saliency Detection



3. CCG Achieves a 4% Performance Gain on Object Discovery



4. CCG Generalizes well to Out-of-distribution Cell Images