Low-shot visual object counting

Matej Kristan

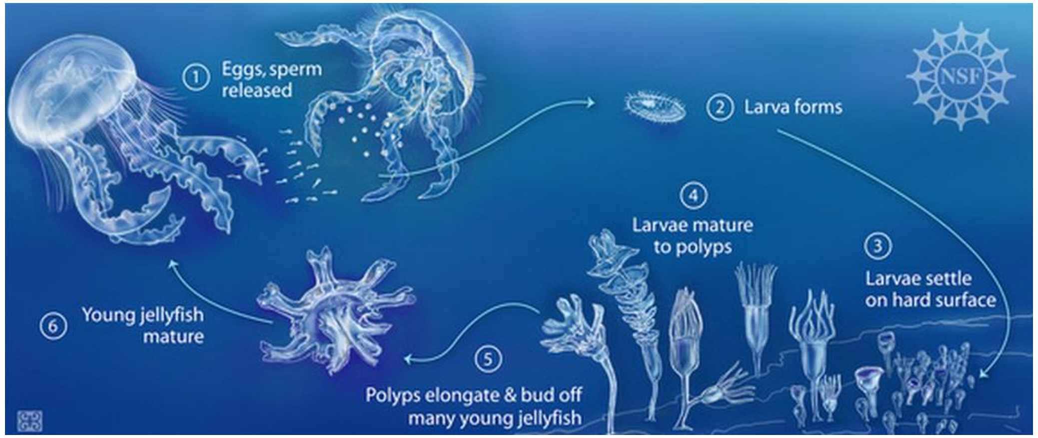
Visual Cognitive Systems Laboratory
Faculty of computer and information science
University of Ljubljana, Slovenia

Data Science, FRI, November 2023

How I got into counting

- ~2016 an astrophysicist-turned-marine-biologist asked me for help

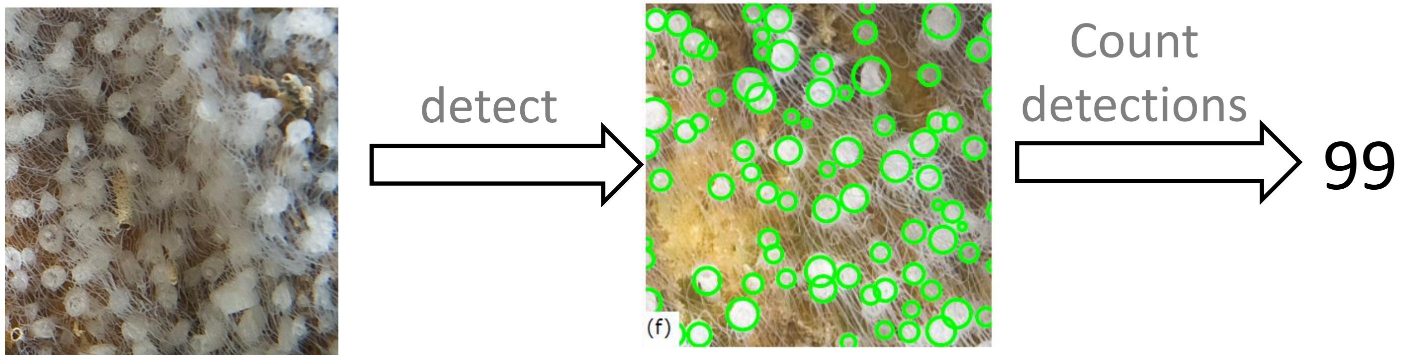
How many polyps in the image?



State-of-the art approaches in counting

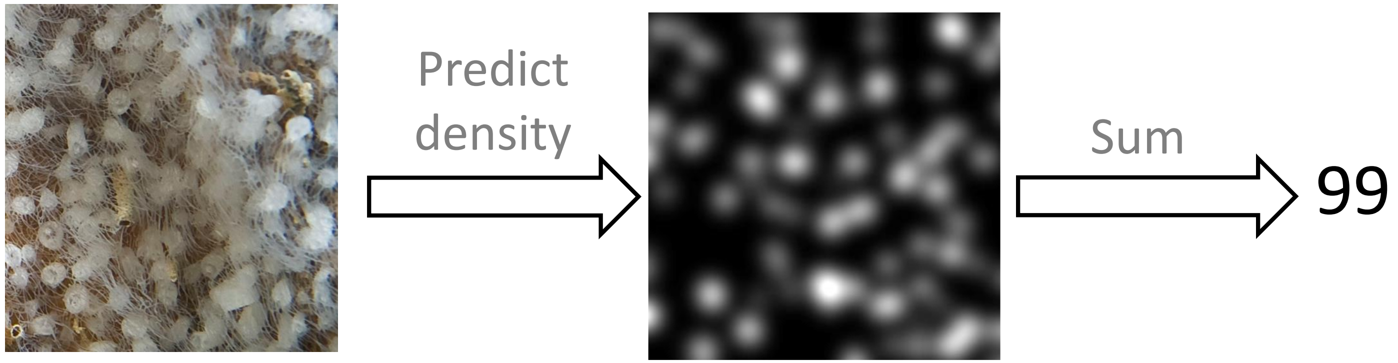
Counting by detection

Lin et al. SMCA2001; Zhao et al. CVPR2003; Ge et al. CVPR2009; Leibe et al. CVPR2005; Idrees et al. CVPR2013



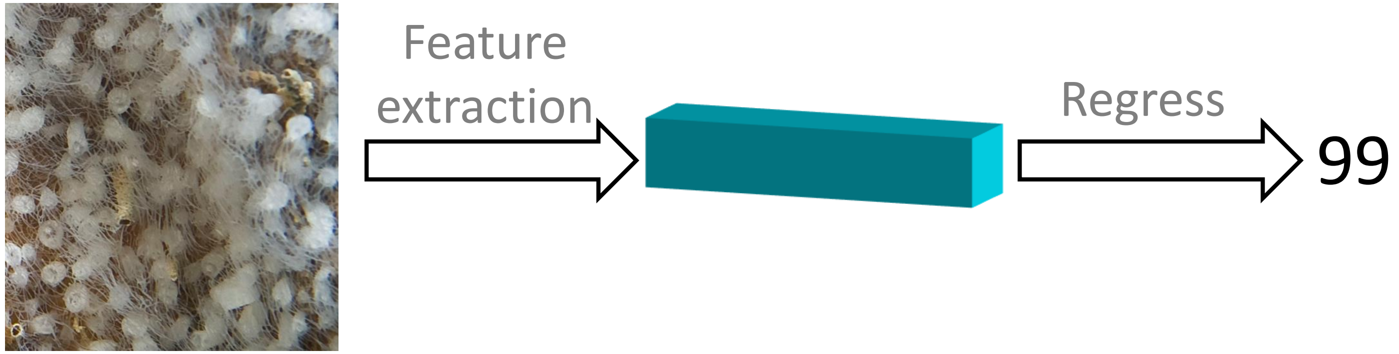
Counting by density est.

Arteta et al. ECCV2014; Pham et al. ICCV2015; Zhang et al. CVPR2015; Zhang et al. CVPR2016; Zeng et al. ICIP2017; Sindagi et al. AVSS2017; Cao et al. ECCV2018; Ranjan et al. ECCV2018; Ma et al. ICCV2019; Zhang et al. ICCV2019; Jiang et al. CVPR2019; Wan et al. ICCV2019; Liu et al. CVPR2019; Wan et al. Neurips2020; Wan et al. CVPR2021; Cheng et al. CVPR2022



Counting by regression

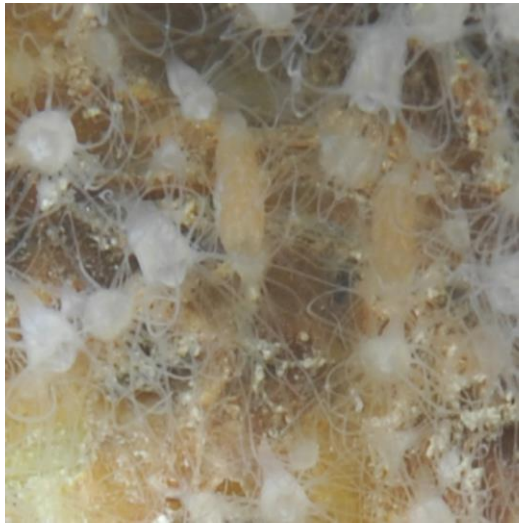
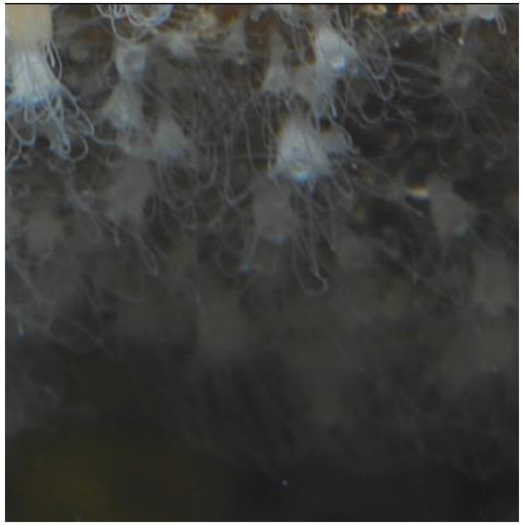
Chan et al. CVPR2008; Ryan et al. DIC2009; Kong et al. ICPR2006; Chen et al. BMVC2012



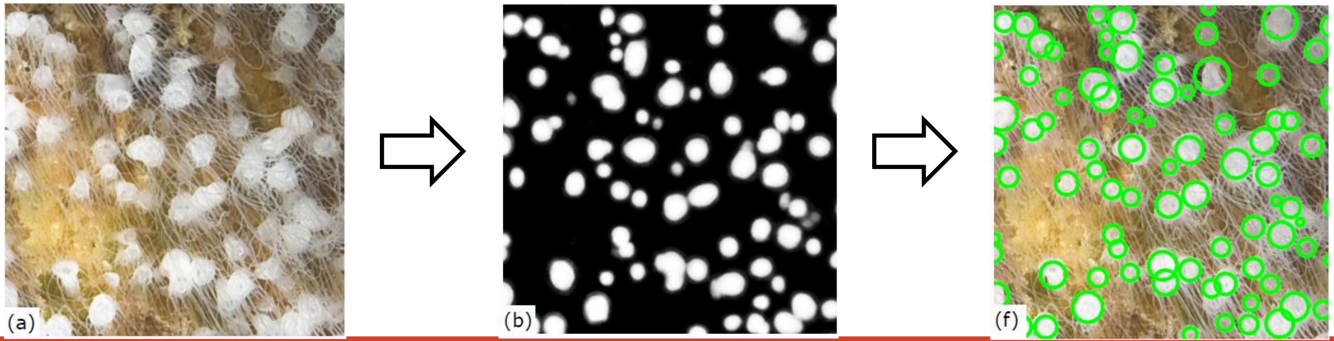
The challenges of polyp counting

- Requirement: visualized locations and sizes
- High size variability, high density, bluring, nonconstant contrast, ...



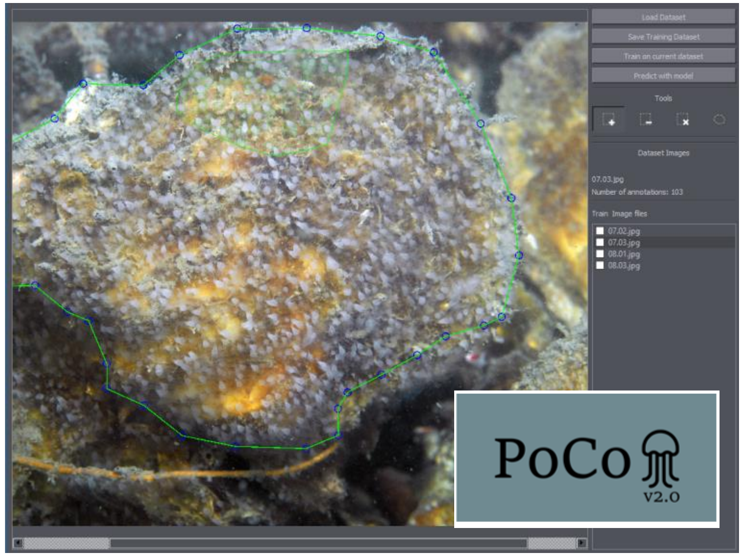
- Our approach (v2): Segment and fit circles into the mask



Detection-by-segmentation counting

- Above human-level performance, PoCo2 released to biology community

Method	Ratio	Rel. err.	AR
PoCo2 [4,64]	**0.99 ± 0.02**	**0.01 ± 0.02**	**0.94 ± 0.01**
RetinaNet	0.92 ± 0.05	0.08 ± 0.05	0.89 ± 0.04



Zavrtanik, Vodopivec, Kristan, *A segmentation-based approach for polyp counting in the wild*, Engineering Applications of Artificial Intelligence, Elsevier, 2020

Adapted for industrial surface inspection

Defect counting by segmentation

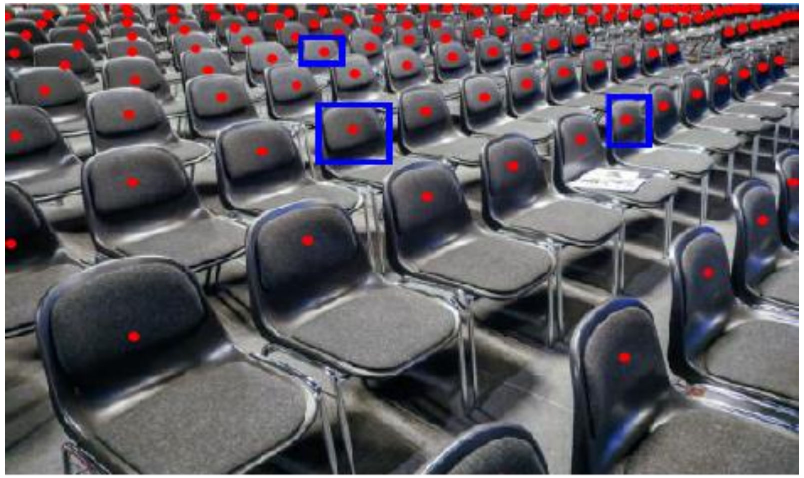
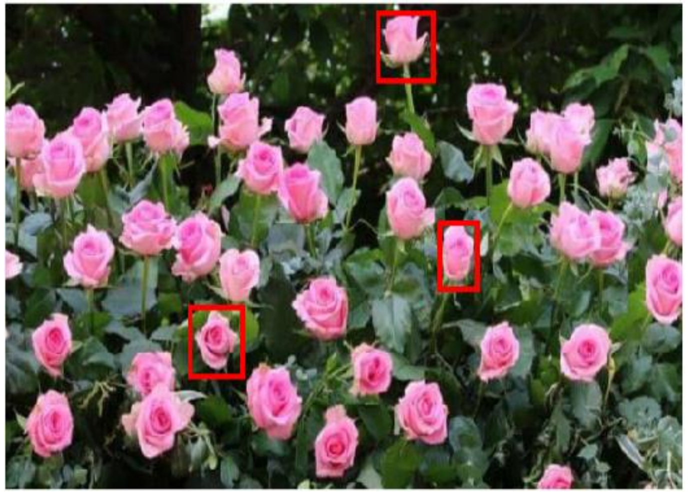
Low-shot counting – the FSC147 benchark

- The FSC147[1] benchmark with *few-shot* *(& 0-shot)* challenge

  [1]Ranjan, et al. "Learning to count everything." *CVPR* 2021

- Few-shot counting:

  "Given a few exemplars, count all objects of the same class in the image."

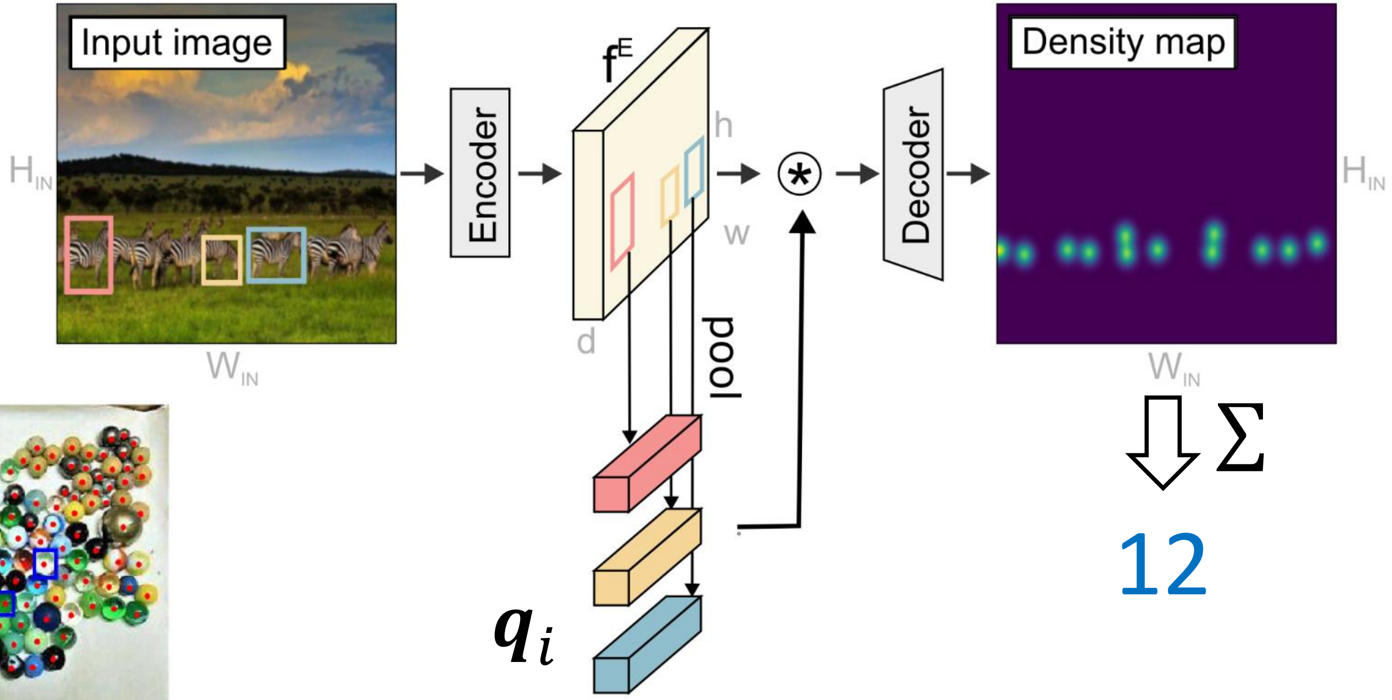




- Zero-shot counting: *"Count the majority class objects in the image."*

Few-shot counting: the standard pipeline[1]

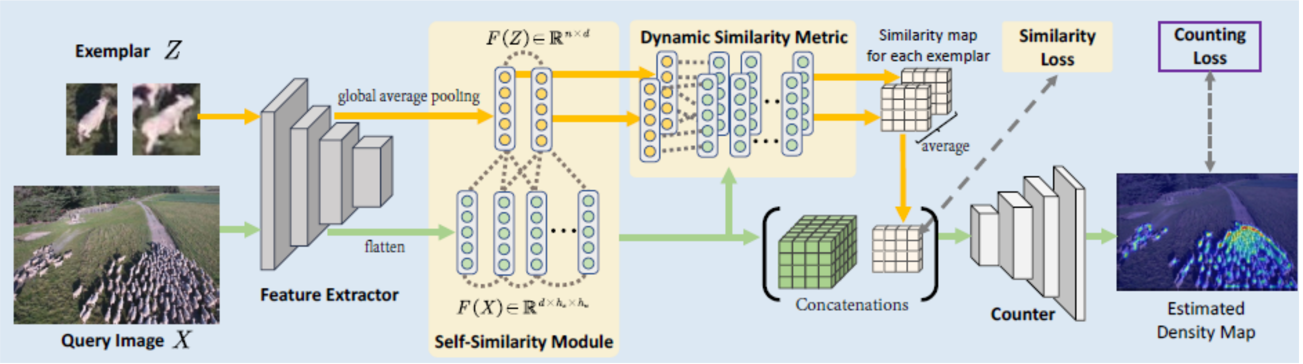
- Extract features of the entire image ($\boldsymbol{f}^E$)
- Pool features for each exemplar into queries (object prototypes) ($\boldsymbol{q}_i$)
- Correlate the queries with the extracted features and decode into density
- Sum the density map

- Challenge: *The prototypes should be able to detect all instances well*

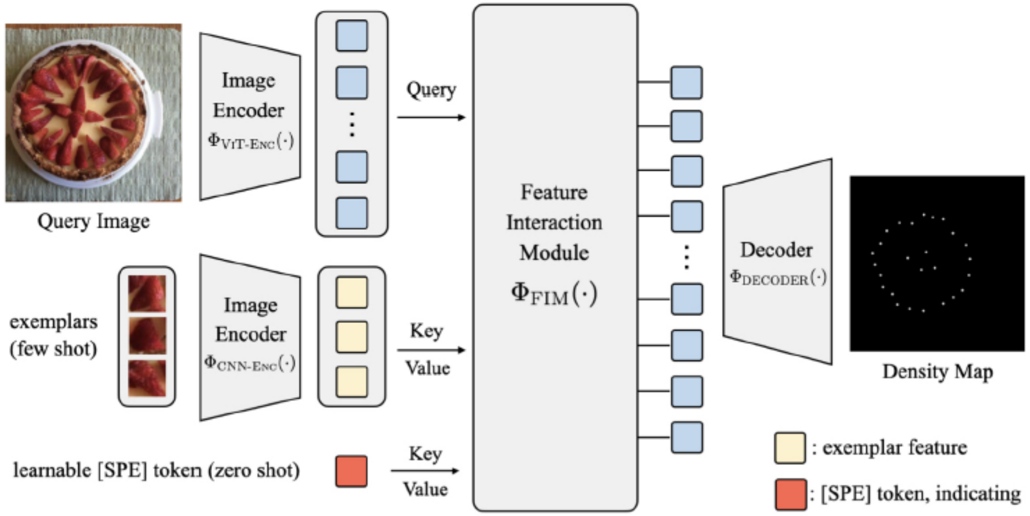


[1]Ranjan, et al. "Learning to count everything." *CVPR* 2021

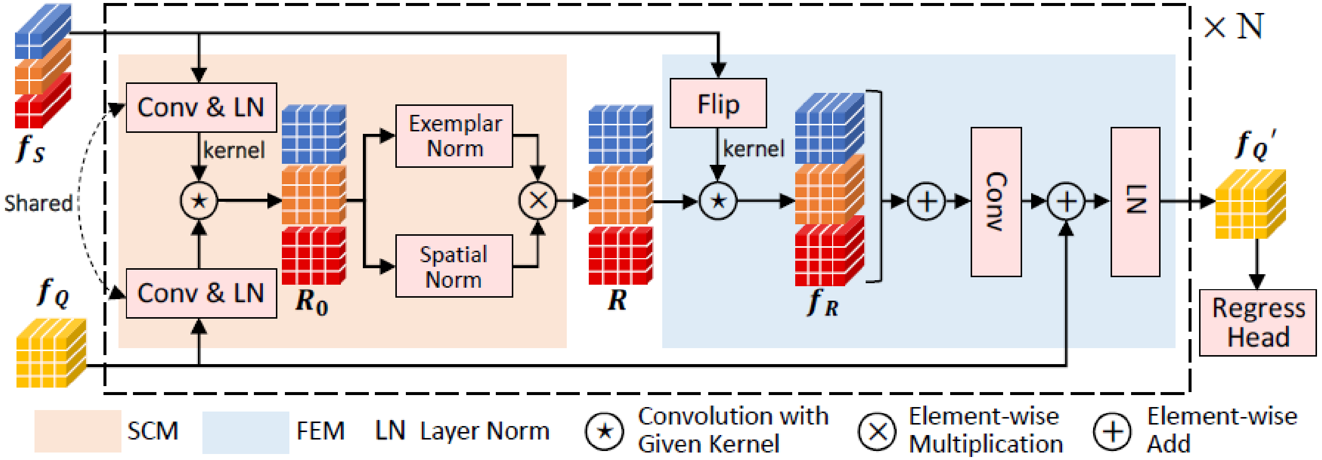
Related works



Metric learning [Shi et al., CVPR2022]



Transformer [Chang et al., BMVC2022]



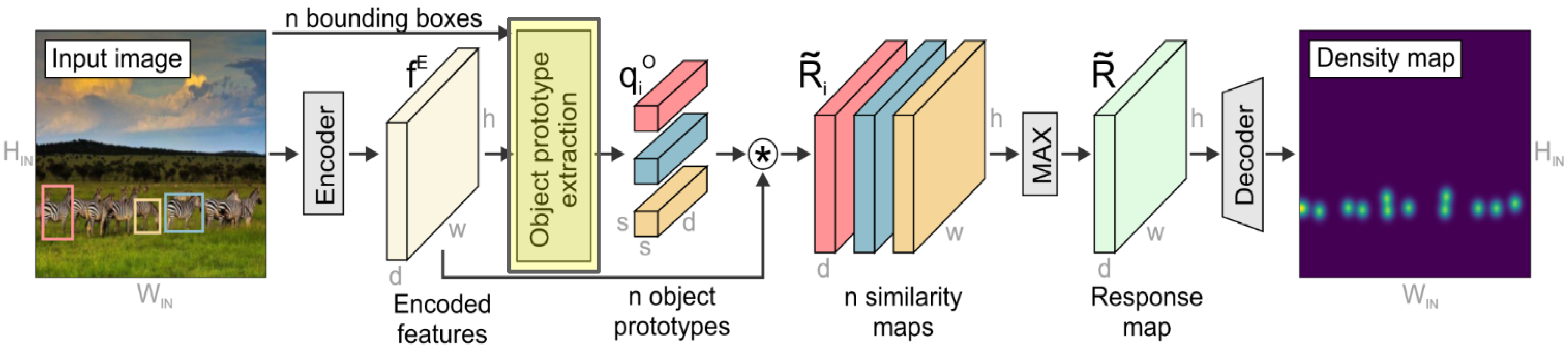
Feature enhancing [You et al., WACV2023]

Observations:

- Prototype constructed by pooling to a fixed-size correlation filter
- Shape information is lost
- Shape should guide filter construction

Our approach: LOCA

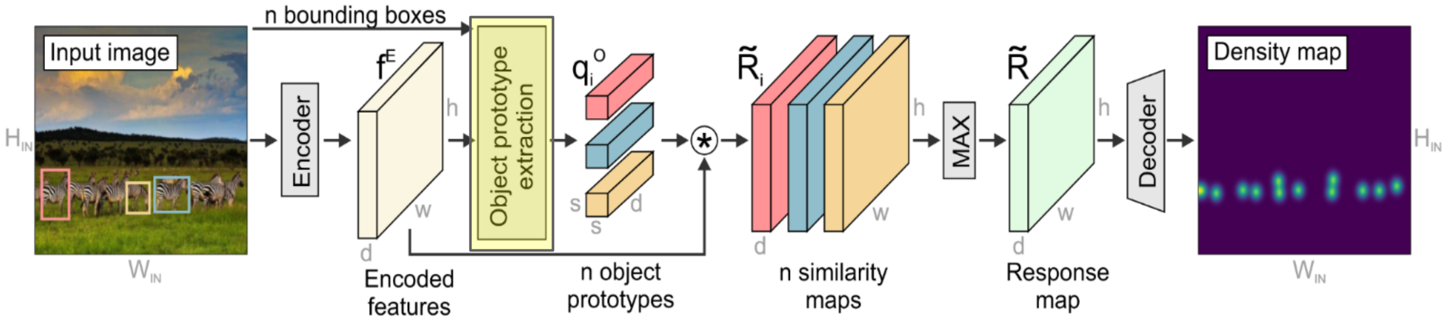
- Low-shot Object Counting network with iterative prototype Adaptation
- Explicitly addresses scale information
- Propose object prototype extraction (OPE) module, which is modulated by the shape information for improved correlation filter construction

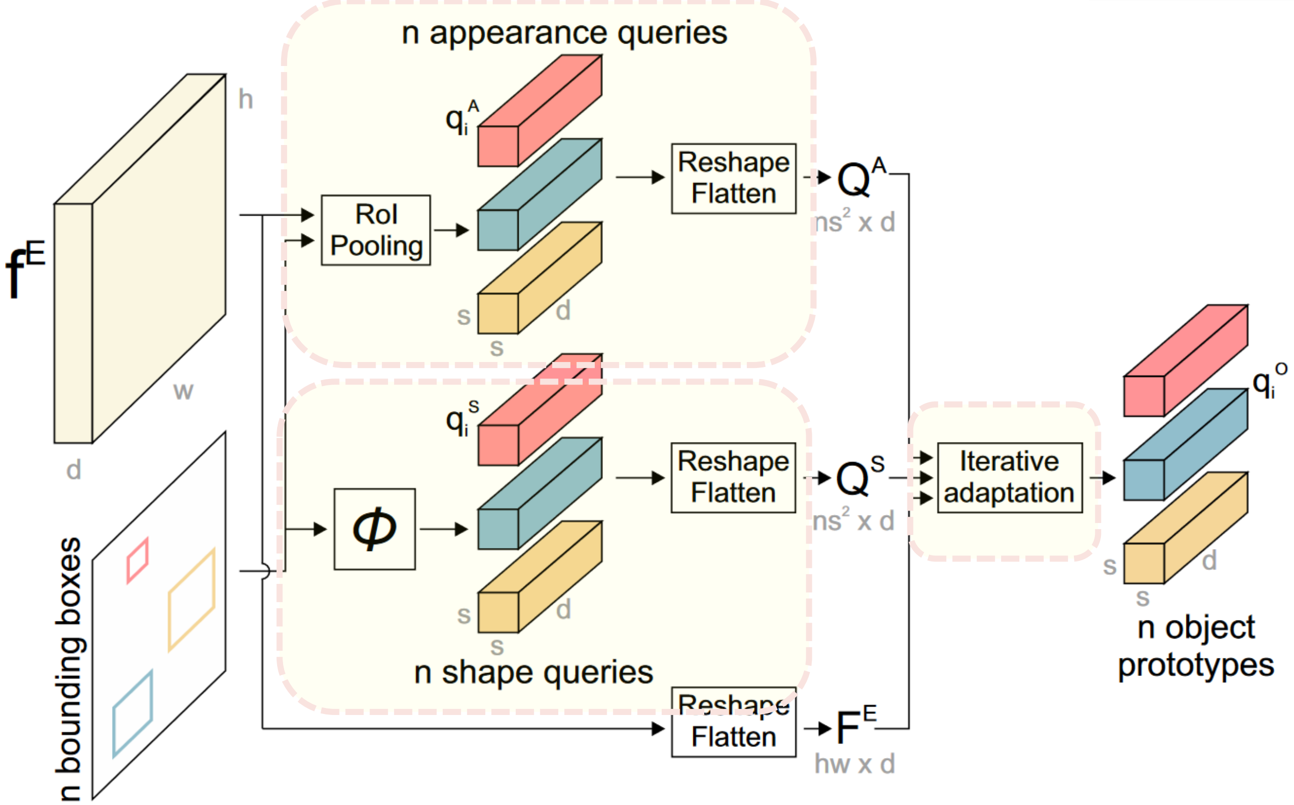


Đukić, Zavrtanik, Lukežič, Kristan, A Low-Shot Object Counting Network With Iterative Prototype Adaptation, ICCV2023

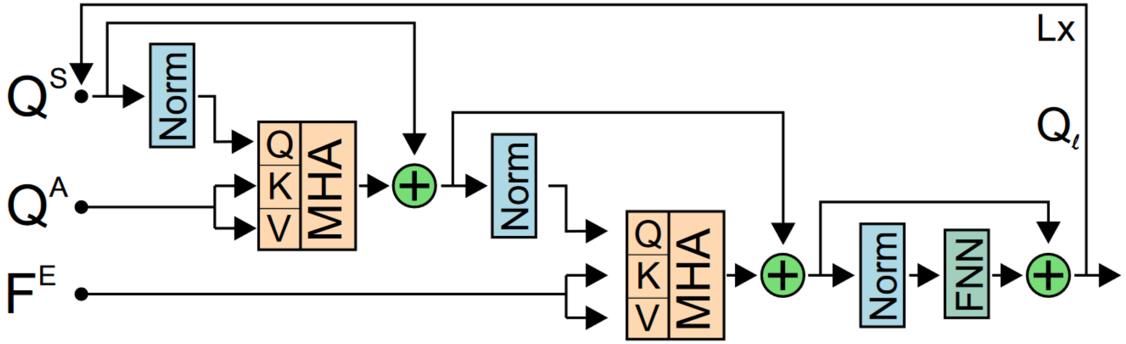
Object prototype extraction (OPE)

- Shape and appearance queries extracted separately





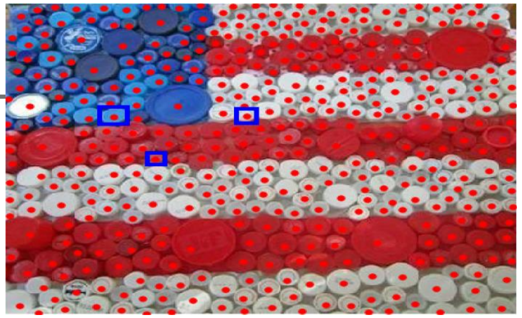
OPE: Iterative adaptation



Shape query network

$$[W_i, H_i] \rightarrow \boxed{\Phi} \rightarrow Q^s$$

LOCA: Experimental results on FSC147[1]

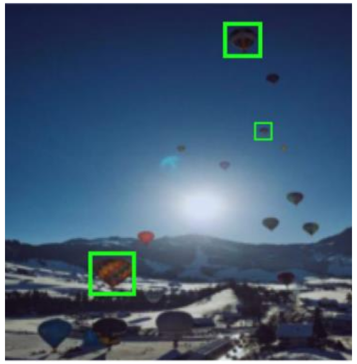


- 147 object categories (89 in training, and 29 in test set)
- 6000 images (3659 for training)
- On average 56 objects per image (between 7 and 3731)
- 3 objects annotated with a bounding box in each image

[1]Ranjan, et al. "Learning to count everything." *CVPR* 2021

Few-shot performance

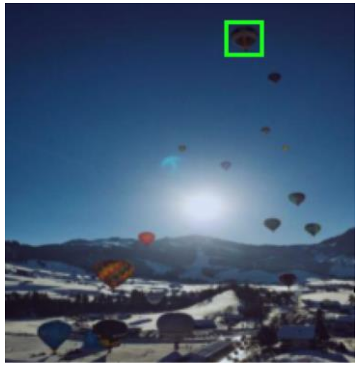
Three-shot setup



Method	Validation set MAE	Validation set RMSE	Test set MAE	Test set RMSE
GMN [ACCV2018]	29.66	89.81	26.52	124.57
MAML [PMLR2017]	25.54	79.44	24.90	112.68
FamNet [CVPR2021]	23.75	69.07	22.08	99.54
CFOCNet [WACV2021]	21.19	61.41	22.10	112.71
BMNet+ [CVPR2022]	15.74	58.53	14.62	91.83
SAFECount [WACV2023]	15.28③	47.20②	14.32③	85.54②
CounTR [BMVC2022]	13.13②	49.83③	11.95②	91.23③
LOCA (ours)	10.24①	32.56①	10.79①	56.97①

RMSE is improved by 33.4%

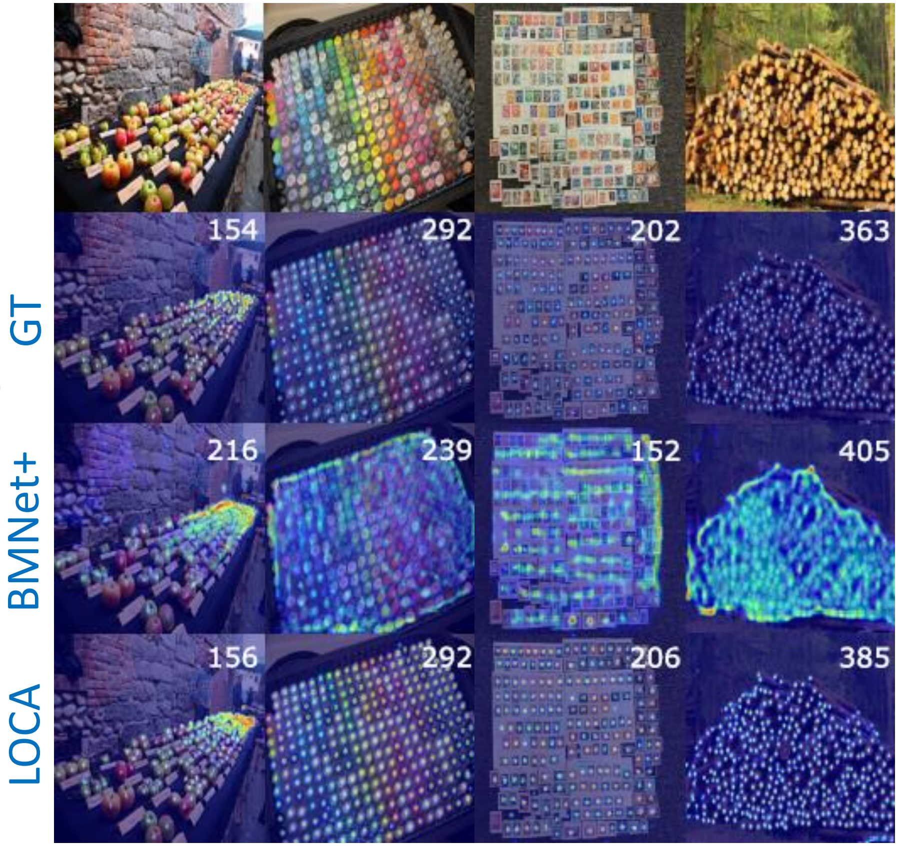
One-shot setup



Method	Validation set MAE	Validation set RMSE	Test set MAE	Test set RMSE
GMN [ACCV2018]	29.66	89.81	26.52	124.57
CFOCNet [WACV2021]	27.82	71.99	28.60	123.96
FamNet [CVPR2021]	26.55	77.01	26.76	110.95
BMNet+ [CVPR2022]	17.89	61.12	16.89	96.65③
LaoNet [Arxiv2021]	17.11③	56.81③	15.78③	97.15
CounTR [BMVC2022]	13.15②	49.72②	12.06①	90.01②
LOCA (ours)	11.36①	38.04①	12.53②	75.32①

RMSE is improved by 16.3%

Three-shot performance

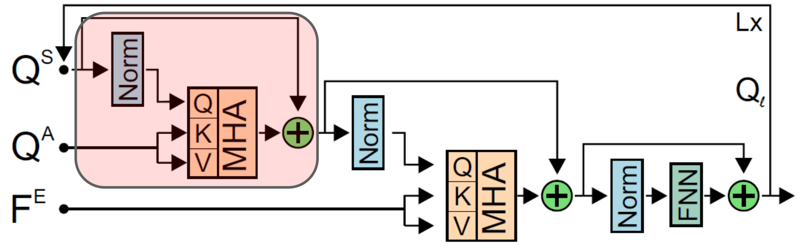


LOCA delivers a more accurate density

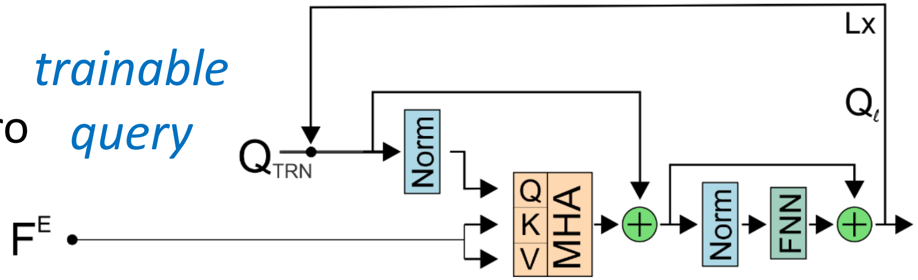
Zero-shot (vs Few-shot) performance

- Appearance/shape queries replaced by a trainable prototype

OPE

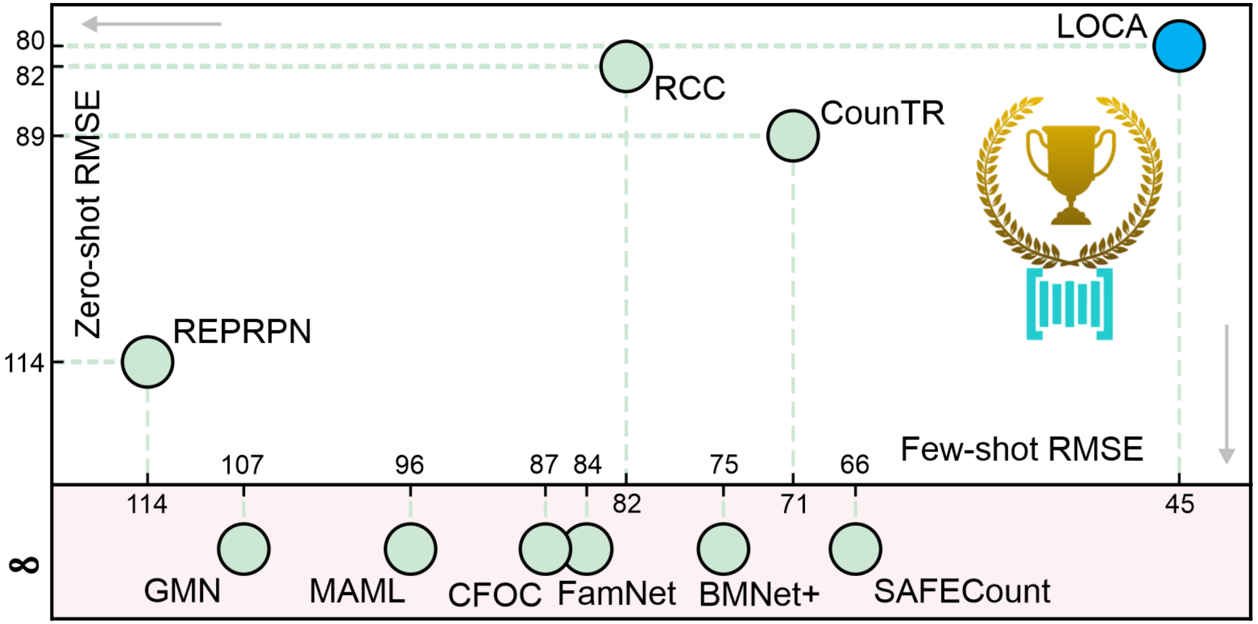


OPE_{zero} *trainable query*



Method	Validation set		Test set	
	MAE	RMSE	MAE	RMSE
RepRPN-C [23]	29.24	98.11	26.66	129.11
RCC [11]	17.49③	58.81②	17.12③	104.53②
CounTR [16]	17.40①	70.33③	14.12①	108.01③
LOCA (ours)	17.43②	54.96①	16.22②	103.96①

LOCA is the top overall counter

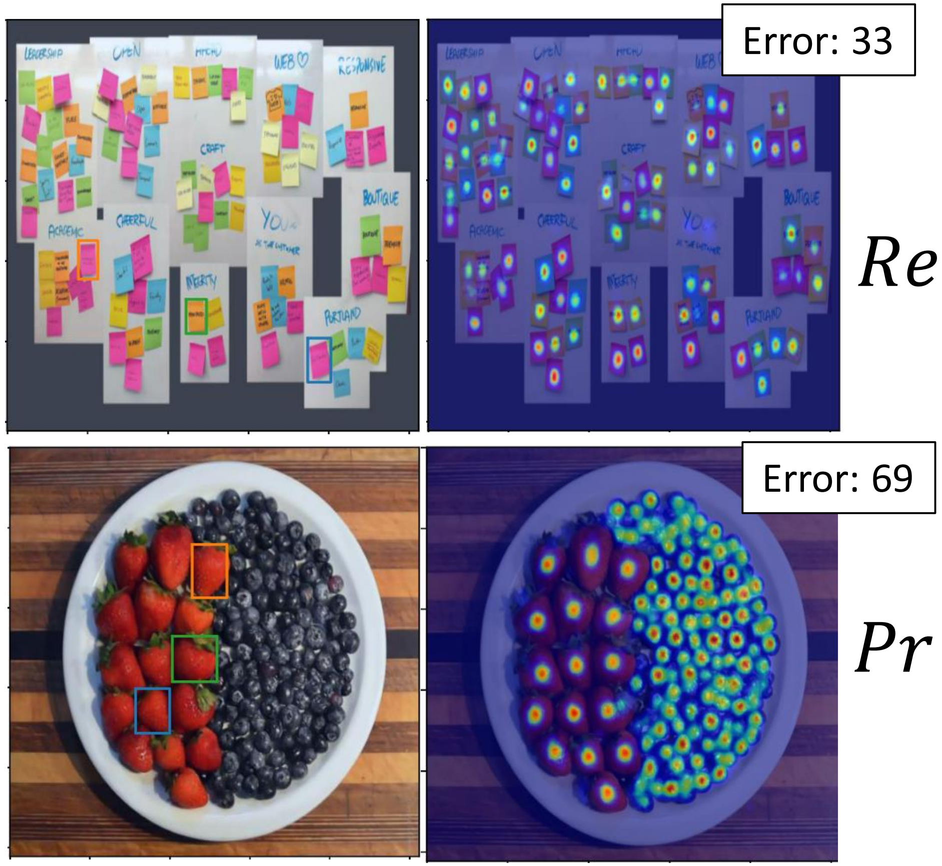


→ On par with the state-of-the-art

Đukić, et al., A Low-Shot Object Counting Network With Iterative Prototype Adaptation, ICCV2023

Issues with SOTA low-shot (LS) counters

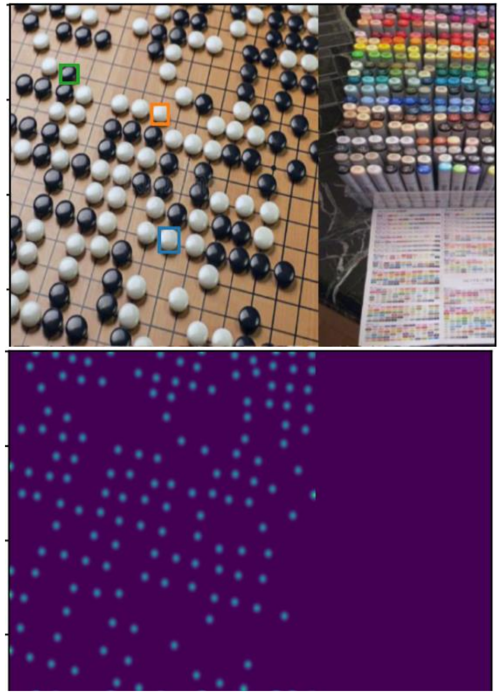
- Large within-class diversity requires general features to maximize recall
- Application to multi-class images often leads to reduced Precision

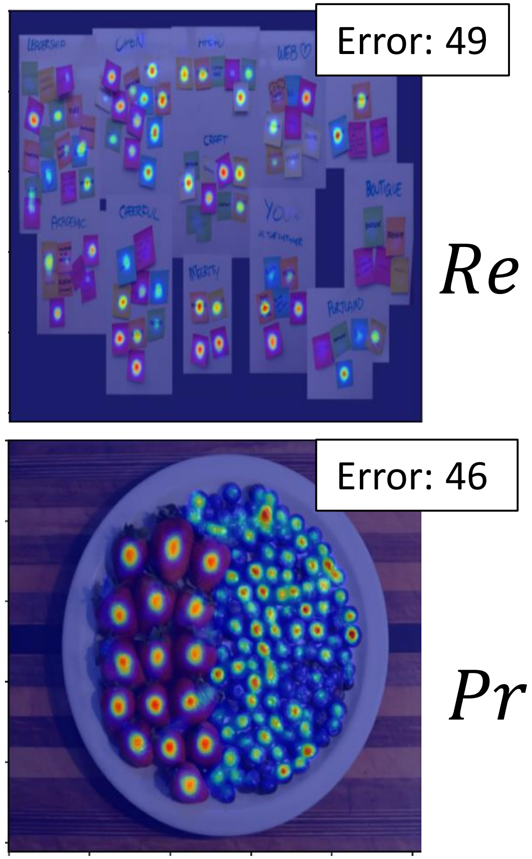


$Re \uparrow$

$Pr \downarrow$

"Solution": Augment training set images with mixed-class examples





$Re \downarrow$

$Pr \uparrow$

A novel DAVE counter

- Slides ommitted from publix since the paper’s under review...

Pelhan, Lukežič, Zavrtanik, Kristan, DAVE – A Detect-and-Verify Paradigm for Low-Shot Counting, (submitted) 2023

Density-based counting performance

3-shot counting

Method	Validation set		Test set	
	MAE	RMSE	MAE	RMSE
GMN [20]	29.66	89.81	26.52	124.57
MAML [10]	25.54	79.44	24.90	112.68
FamNet [27]	23.75	69.07	22.08	99.54
CFOCNet [38]	21.19	61.41	22.10	112.71
BMNet+ [29]	15.74	58.53	14.62	91.83
VCN [25]	19.38	60.15	18.17	95.60
SAFECount [39]	15.28	47.20 (3)	14.32	85.54 (3)
CounTR [16]	13.13 (3)	49.83	11.95 (3)	91.23
LOCA [6]	10.24 (2)	32.56 (2)	10.79 (2)	56.97 (2)
DAVE	8.91 (1)	28.08 (1)	8.66 (1)	32.36 (1)

Prompt-based counting

Method	Validation Set		Test Set	
	MAE	RMSE	MAE	RMSE
ZeroClip [37]	26.93	88.63	22.09	115.17
CLIP-Count [15]	18.79 (3)	61.18 (2)	17.78 (3)	106.62 (2)
CounTX [1]	17.70 (2)	63.61 (3)	15.73 (2)	106.88 (3)
DAVE_{prm}	15.48 (1)	52.57 (1)	14.90 (1)	103.42 (1)

0-shot counting

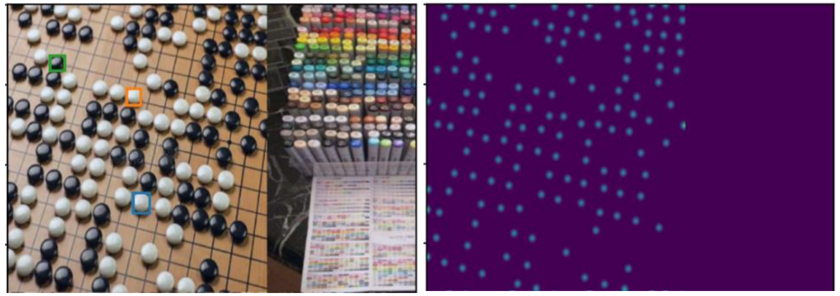
Method	Validation Set		Test Set	
	MAE	RMSE	MAE	RMSE
RepRPN-C [26]	29.24	98.11	26.66	129.11
RCC [13]	17.49	58.81 (3)	17.12	104.53 (3)
CounTR [16]	17.40 (2)	70.33	14.12 (1)	108.01
LOCA [6]	17.43 (3)	54.96 (2)	16.22 (3)	103.96 (2)
$\text{DAVE}_{\text{0-shot}}$	15.54 (1)	52.67 (1)	15.14 (2)	103.49 (1)

- Outperforms 3-shot sota (substantial RMSE reduction) [LOCA ICCV2023]
- Outperforms prompt-based sota [ZeroClip CVPR2023, CounTX BMVC2023]
- Slightly outperforms or performs on par with 0-shot sota [LOCA ICCV2023]

False response activation reduction

Performance on multi-class images

- Multi-class test set FSCD147$_{mul}$
- Retrained LOCA with multi-class images:



→ LOCA$_{mul}$

	FSCD147			FSCD147$_{mul}$	
	MAE(↓)	RMSE(↓)	AP50(↑)	MAE(↓)	RMSE(↓)
LOCA [6]	10.79②	56.97②	-	21.28	43.67
LOCA$_{mul}$ [6]	12.63	78.95	-	13.25②	22.57②
CounTR [16]	11.95③	91.23③	-	14.56③	27.41③
DAVE	8.66①	32.36①	61.08①	3.05①	4.94①

- LOCA$_{mul}$ improves over LOCA on multi-class, but not on the original dataset
- Margin between DAVE and top-performer further increases on FSCD147mul *(~40% FSCD vs ~80% FSDC$_{mul}$)*

Few-shot *detection* performance analysis

Method	Validation Set AP↑	Validation Set AP50↑	Test Set AP↑	Test Set AP50↑
FSDetView-PB [35]	-	-	13.41	32.99
FSDetView-RR [35]	-	-	17.21	33.70
AttRPN-RR [9]	-	-	18.53	35.87
AttRPN-PB [9]	-	-	20.97③	37.19③
C-DETR [22]	17.27②	41.90②	22.66②	50.57②
DAVE	24.20①	61.08①	26.81①	62.82①
$\mathrm{DAVE}_{0\text{-shot}}$	16.31②	46.87①	18.55②	50.08②

(medals are vs CDETR)

- DAVE outperforms CDETR[1] by ~20% (AP & AP50)
  - Better Pr / Re : (1), (2)
  - Better in high-density regions: (3), (4)
  - Better learns what to count (5)
- Even 0-shot DAVE performs on par with CDETR (!)

(5)

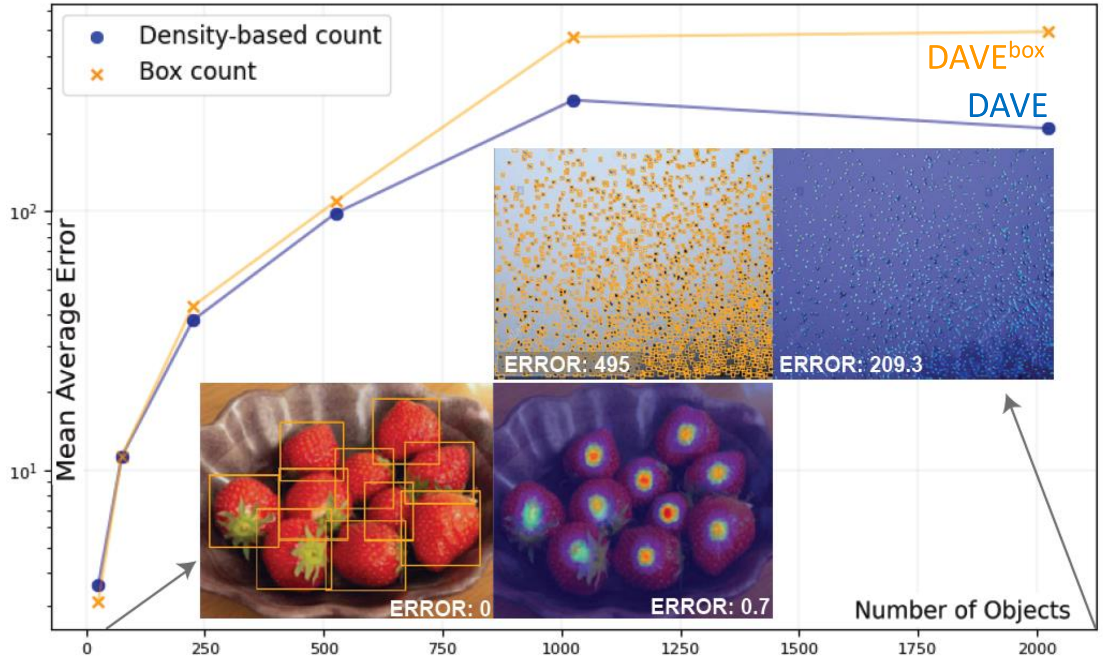
[1]Nguyen et al, ECCV2022

Counting-by-detection performance

- Estimate the counts by the number of detections: *DAVE*^{box}

Few-shot setup

	Validation Set		Test Set	
Method	MAE	RMSE	MAE	RMSE
FSDetView-RR [35]	-	-	37.83	146.56
FSDetView-PB [35]	-	-	37.54	147.07
AttRPN-RR [9]	-	-	32.70	141.07③
AttRPN-PB [9]	-	-	32.42③	141.55
C-DETR [22]	20.38②	82.45②	16.79②	123.56②
DAVE^{box}	9.75①	40.30①	10.45①	74.51①
LOCA [6]	10.24②	32.56②	10.79②	56.97②
DAVE	8.91①	28.08①	8.66①	32.36①



- DAVE outperforms detection sota by ~40% MAE/RMSE
- Not only detection-based, also density-based estimates of LOCA
- DAVE^{box} lags behind DAVE in high-density scenes with small objects

Conclusion

- Overviewed our recent work on counting

PoCo!

Zavrtanik, Vodopivec, Kristan, A segmentation-based approach for polyp counting in the wild, Engineering Applications of Artificial Intelligence, Elsevier, 2020

LOCA!

Đukić, Zavrtanik, Lukežič, Kristan, A Low-Shot Object Counting Network With Iterative Prototype Adaptation, ICCV2023

DAVE!

Pelhan, Lukežič, Zavrtanik, Kristan, DAVE – A Detect-and-Verify Paradigm for Low-Shot Counting, (submitted)

- Several directions for future research:
  - Further increase detection capabilities on high-density regions.
  - Explore possibility of (automatic) exemplar re-selection & interactive counting.
  - Connection to retrieval, tracking, etc.

Thanks

The PoCO/LOCA/DAVE core team

Jer Pelhan

Nikola Đukić

Vitjan Zavrtanik

Alan Lukežič

Matej Kristan

Slovenian
Research
Agency