Scale-Prior Deformable Convolution for Exemplar-Guided Class-Agnostic Counting

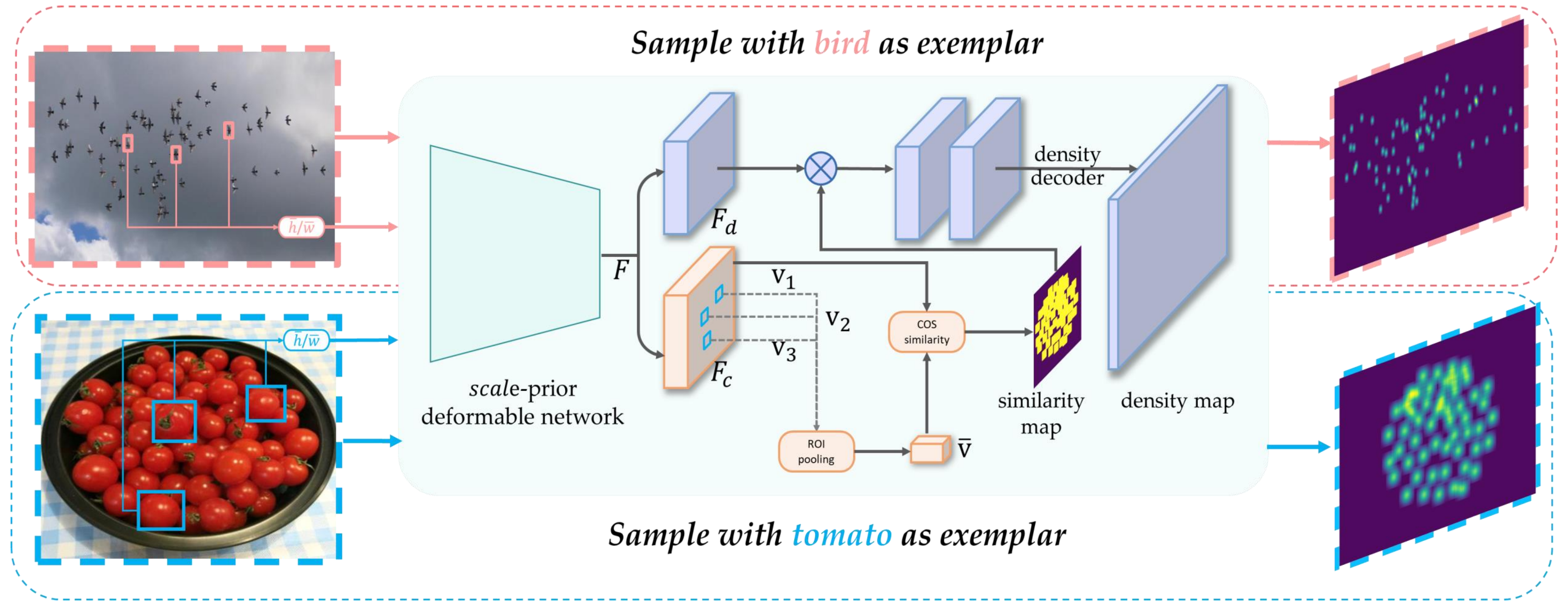
Wei Lin[1], Kunlin Yang[2], Xinzhu Ma[3], Junyu Gao[4], Lingbo Liu[5], Shinan Liu[2], Jun Hou[2], Shuai Yi[2], Antoni B. Chan[1]

[1]City University of Hong Kong, [2]SenseTime Group Limited, [3]The University of Sydney, [4]Northwestern Polytechnical University, [5]The Hong Kong Polytechnic University

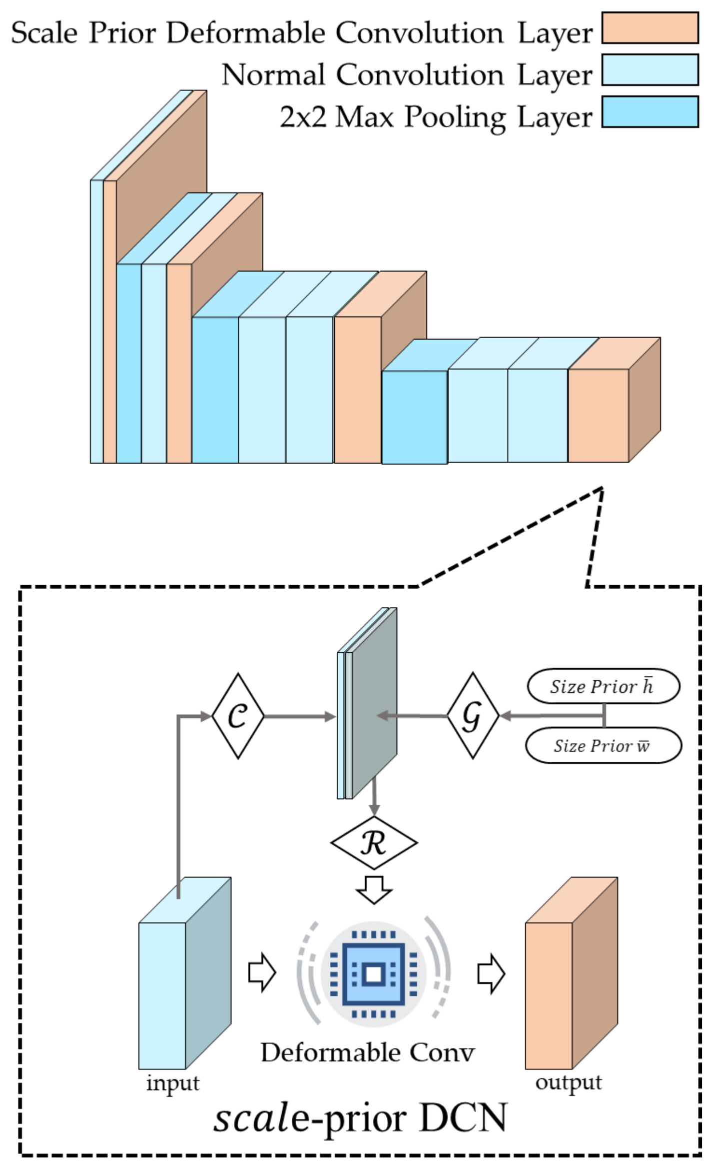
香港城市大學 City University of Hong Kong 電腦科學系 Department of Computer Science

BMVC 2022

Scale-Prior Deformable Convolution Network



- Previous works focus on designing self-similarity matching rules between exemplars and query images;
- SPDCN is developed to better extract exemplar-related features;



scale-prior backbone

non-linear $\mathcal{C}$	non-linear $\mathcal{G}$
Input size: $n \times h \times w$	*Input size:* 2
Conv_3x3$(n,64)$	Linear$(2,64)$
ReLU	ReLU
Conv_3x3$(64,32)$	Linear$(64,32)$
Output d_c: $32 \times h \times w$	*expand to* d_g: $32 \times h \times w$
non-linear $\mathcal{R}$	
Concatenate $\mathcal{C}$ and $\mathcal{G}$: $64 \times h \times w$	
Conv_3x3$(64,32)$	
ReLU	
Conv_3x3$(32,18)$	
Output size: $18 \times h \times w$	

- The offsets in vanilla deformable convolution only comes from local embedding $\mathcal{C}$;
- In SPDCN, the offsets is transformed from the combination of local embedding $\mathcal{C}$ and global embedding $\mathcal{G}$. A non-linear module $\mathcal{R}$ is used to fuse them.
- The global embedding is the average height and width of given exemplars.

$$d_g = \mathcal{G}(\bar{h}, \bar{w}), \quad \bar{h} = \sum_{e_i \in E_I} \frac{h_{e_i}}{|E_I|}, \bar{w} = \sum_{e_i \in E_I} \frac{w_{e_i}}{|E_I|}$$

Scale-Sensitive Generalized Loss

$$\mathcal{L}_{\mathbf{C}} = \min_{\mathbf{P}} \langle \mathbf{C}, \mathbf{P} \rangle - \varepsilon H(\mathbf{P}) + \tau \|\mathbf{P}\mathbf{1}_m - \mathbf{a}\|_2^2 + \tau \|\mathbf{P}^\top \mathbf{1}_n - \mathbf{b}\|_1$$

$$\mathbf{C}_{ij} = \|\hat{x}_i - \hat{y}_j\|_2, \quad [\hat{x}_i \quad \hat{y}_j] = \begin{bmatrix} 1/s_h & 0 \\ 0 & 1/s_w \end{bmatrix} [x_i \quad y_j]$$

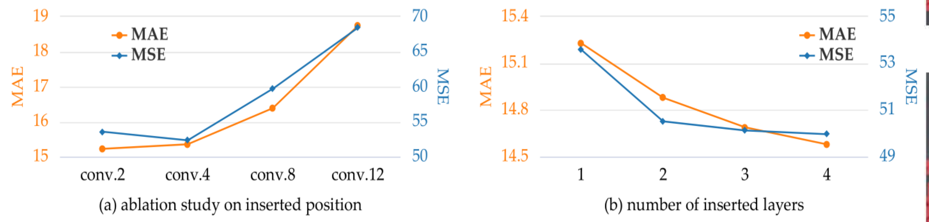
$$\left.\begin{matrix} s_h = \mathcal{S}(\bar{h}) \\ s_w = \mathcal{S}(\bar{w}) \end{matrix}\right\} \mathcal{S}(k) = \frac{\alpha}{1 + \exp(-(k-\mu)/\sigma)} + \beta$$

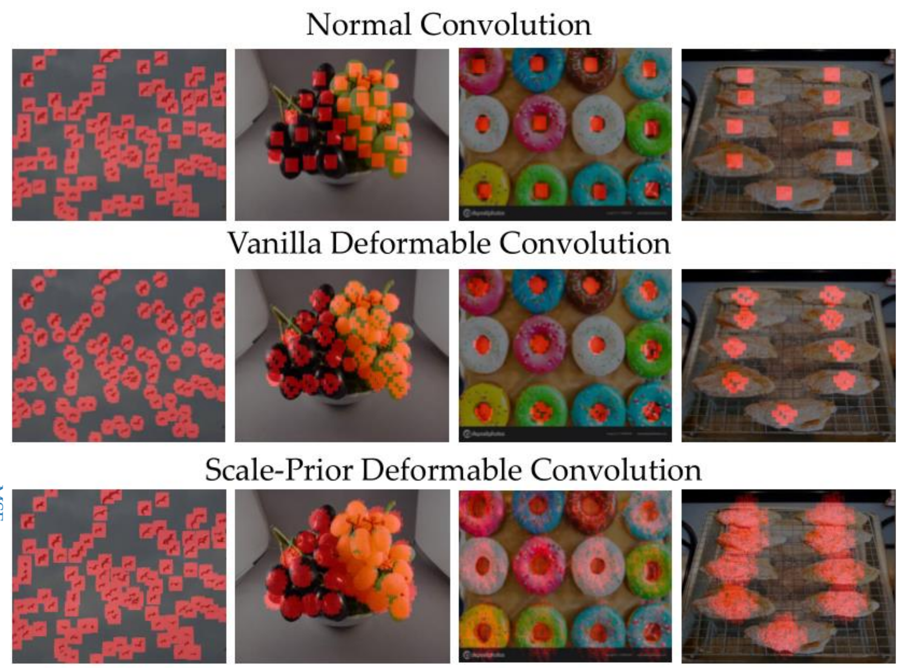
Effect of Scale-sensitive Loss (MAE/MSE)

VGG-19	L2 loss		Generalized loss	
	vanilla	scale-sensetive	vanilla	scale-sensetive
w/o scale-prior	23.67/72.81	22.85/70.90	21.60/71.83	21.23/70.75
w/ scale-prior	15.89/52.24	15.45/50.18	15.55/51.00	**14.59/49.97**

Effect And Visualization of Scale Prior

embedding global	embedding local	validation set MAE	validation set MSE	test set MAE	test set MSE
		21.24	71.16	19.41	128.26
	✓	19.99	69.49	17.58	119.69
✓		16.63	53.30	14.54	104.69
✓	✓	**14.59**	**49.97**	**13.51**	**96.80**





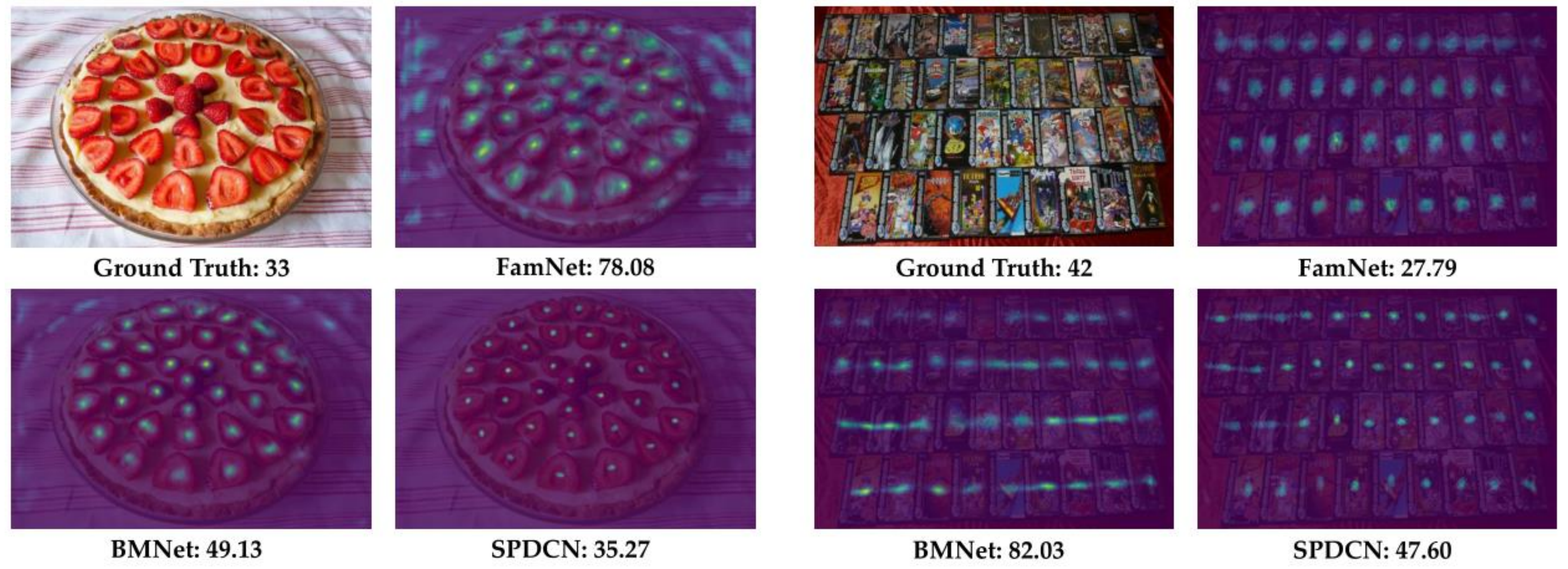
Adaption to CARPK Dataset

method	w/o fine-tuning FamNet	w/o fine-tuning BMNet	w/o fine-tuning SPDCN	w/ fine-tuning FamNet	w/ fine-tuning BMNet	w/ fine-tuning SPDCN
MAE	28.84	**17.30**	18.15	18.19	**9.66**	10.07
MSE	44.47	21.89	**21.61**	33.66	14.84	**14.12**

Comparison with State-of-the-arts on FSC-147

Methods		Validation Set MAE	Validation Set MSE	Test Set MAE	Test Set MSE
FR FSD [13]	ICCV'19	45.45	112.53	41.64	141.04
FSOD FSD [5]	CVPR'20	36.36	115.00	32.53	140.65
MAML [6]	PRML'17	25.54	79.44	24.90	112.68
GMN [16]	ACCV'18	29.66	89.81	26.52	124.57
FamNet [23]	CVPR'21	23.75	69.07	22.08	99.54
VCN [22]	CVPR'22	19.38	60.15	18.17	95.60
BMNet [26]	CVPR'22	15.74	58.53	14.62	**91.83**
SPDCN (ours)		15.55	51.00	14.48	100.01
SPDCN† (ours)		**14.59**	**49.97**	**13.51**	96.80

Visualization of Recent Methods and Ours



Robustness to Scale Variation And Rotation

The scale prior is only used to adjust the receptive field of network. The matching process and density estimation are based on semantic information instead of scale information.