ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks

Qilong Wang[1], Banggu Wu[1], Pengfei Zhu[1], Peihua Li[2], Wangmeng Zuo[3], Qinghua Hu[1]

[1]Tianjin University, [2]Dalian University of Technology, [3]Harbin Institute of Technology

Motivation

- Attention is widely used to improve deep CNNs, such as SENet (CVPR18), CBAM (ECCV18), GSoP (CVPR19), etc.
- Most existing attention modules achieve better performance, but inevitably increase model complexity.
- *Question : Can one learn effective channel attention in a more efficient way?*

Analysis and Findings

Methods	Attention	#.Param.	Top-1	Top-5
Vanilla	N/A	0	75.20	92.25
SE	$\sigma(f_{\{W_1,W_2\}}(y))$	$2\times C^2/r$	76.71	93.38
SE-Var1	$\sigma(y)$	0	76.00	92.90
SE-Var2	$\sigma(w \odot y)$	C	77.07	93.31
SE-Var3	$\sigma(Wy)$	C^2	77.42	93.64
SE-GC1	$\sigma(GC_{16}(y))$	$C^2/16$	76.95	93.47
SE-GC2	$\sigma(GC_{C/16}(y))$	$16\times C$	76.98	93.31
SE-GC3	$\sigma(GC_{C/8}(y))$	$8\times C$	76.96	93.38
ECA-NS	$\sigma(\omega)$with Eq.(7)	$k\times C$	77.35	93.61
ECA(Ours)	$\sigma(C1D_k(y))$	$k=3$	**77.43**	**93.65**

I: Avoiding Dimensionality Reduction (DR)

$$\mathbf{W}_{var2} = \begin{bmatrix} w^{1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w^{C,C} \end{bmatrix} \quad \mathbf{W}_{var3} = \begin{bmatrix} w^{1,1} & \cdots & w^{1,C} \\ \vdots & \ddots & \vdots \\ w^{1,C} & \cdots & w^{C,C} \end{bmatrix}$$

- SE-Var2 > SE: Avoiding dimensionality reduction is more important than consideration of nonlinear channel dependencies.

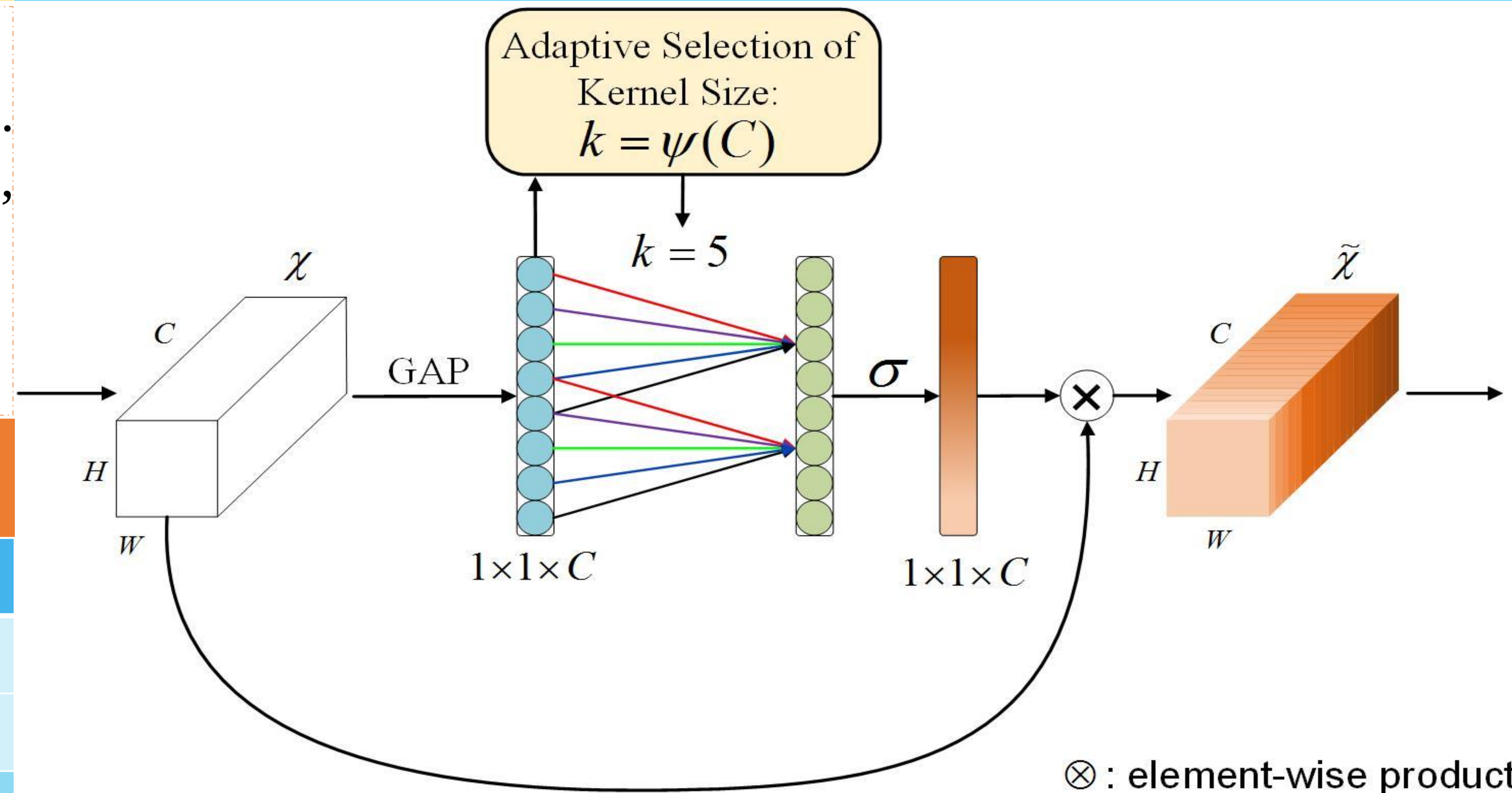
II: Cross-Channel Interaction (CCI) is helpful.

- SEVar-3 > SE-Var2: Cross-channel interaction is beneficial to learn channel attention, but leads to high model complexity.

$$\mathbf{W}_{GC} = \begin{bmatrix} \mathbf{W}_G^1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{W}_G^G \end{bmatrix}$$

- Group Conv. (GC) is not effective to capture cross-channel interaction.
- Reason: SE-GC completely discards dependences among different groups.

ECA Module



Our goal: No DR & Effective CCI in efficient way

ECA-NS: $$\omega_i = \sigma\left(\sum_{j=1}^{k} w_i^j y_i^j\right), y_i^j \in \Omega_i^k \qquad \text{Eq.(7)}$$

ECA: $$\omega_i = \sigma\left(\sum_{j=1}^{k} w^j y_i^j\right), y_i^j \in \Omega_i^k \Rightarrow \omega = \sigma(C1D_k(y))$$

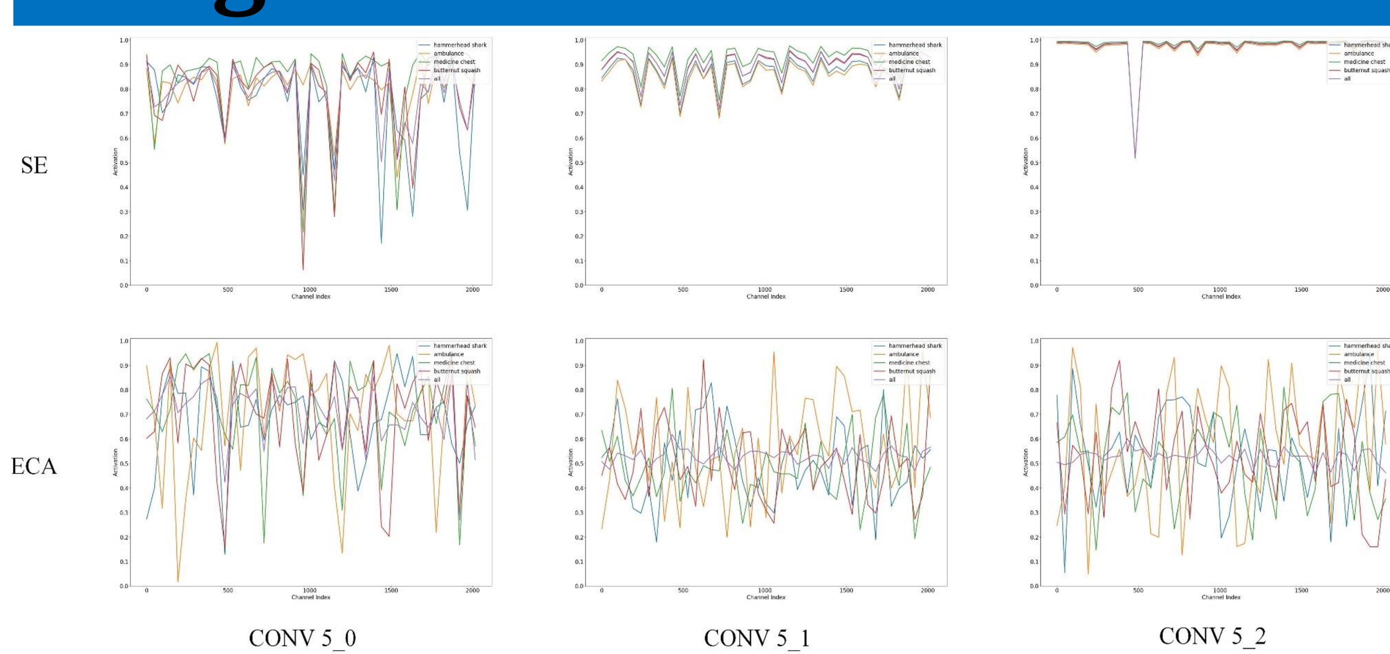
where Ω_i^k indicates the set of k adjacent channels of y_i.

- Our ECA can avoid complete independence among different groups, achieving promising performance with much lower model complexity.

How to adaptively compute kernel size k?

We introduce a nonlinear guideline: $C=\phi(k)=2^{(\gamma * k+b)} \Rightarrow k=\psi(C)=\left|\frac{\log_2(C)}{\gamma}+\frac{b}{\gamma}\right|_{odd}$

Weights Visualization



The weights learned by the proposed ECA module for different categories in the high-level semantics are significantly more discriminative than the SE module.

Experiments on ImageNet-1K

Method	Backbone	#.Param.	FLOPs	Top-1	Top-5
ResNet	ResNet-18	11.148M	1.699G	70.40	89.45
SENet		11.231M	1.700G	70.59	89.78
ECA-Net (Ours)		11.148M	1.700G	70.78	89.92
ResNet	ResNet-34	20.788M	3.427G	73.31	91.40
SENet		20.938M	3.428G	73.87	91.65
CBAM		20.943M	3.428G	74.01	91.76
ECA-Net (Ours)		20.788M	3.428G	74.21	91.83
ResNet	ResNet-50	24.37M	3.86G	75.20	92.52
SENet		26.77M	3.87G	76.71	93.38
CBAM		26.77M	3.87G	77.34	93.69
GSoP-Net1		28.05M	6.18G	77.68	93.98
ECA-Net (Ours)		24.37M	3.86G	77.48	93.68
ResNet	ResNet-101	42.49M	7.34G	76.83	93.48
SENet		47.01M	7.35G	77.62	93.93
CBAM		47.01M	7.35G	78.49	94.31
ECA-Net (Ours)		42.49M	7.35G	78.65	94.34
MobileNetV2	MobileNetV2	3.34M	319.4M	71.64	90.20
SENet		3.40M	320.1M	72.42	90.67
ECA-Net (Ours)		3.34M	319.9M	72.56	90.81

Experiments on MS-COCO

Method	Detector	#.Param.	GFLOPs	AP	Gains
ResNet-101	Faster R-CNN	60.52M	283.14	38.7	-
+ SE block		65.24M	283.33	39.6	↑0.9
+ ECA(Ours)		60.52M	283.32	40.3	↑1.6
ResNet-101	Mask R-CNN	63.17M	351.65	39.4	-
+ SE block		67.89M	351.84	40.7	↑1.3
+ ECA(Ours)		63.17M	351.83	41.3	↑1.9
ResNet-101	RetinaNet	56.74M	315.39	37.7	-
+ SE block		61.45M	315.58	38.7	↑1.0
+ ECA(Ours)		56.74M	315.57	39.1	↑1.4