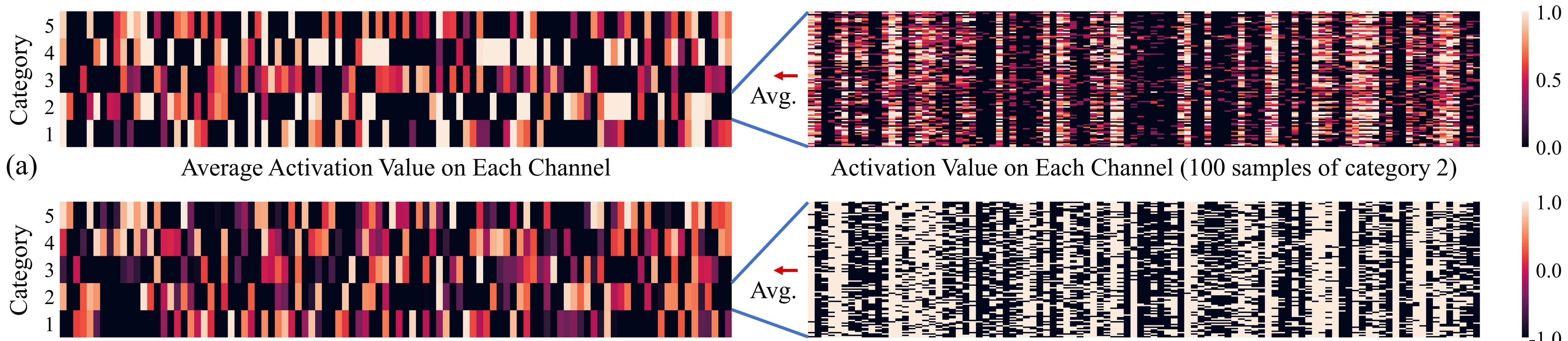
Dual-Perspective Activation: Efficient Channel Denoising via Joint Forward-Backward Criterion for Artificial Neural Networks

Tian Qiu, Chenchao Gao, Zunlei Feng, Jie Lei, Bingde Hu, Xingen Wang, Yi Gao, Mingli Song
https://github.com/horrible-dong/DPA
ZHEJIANG UNIVERSITY
NEURAL INFORMATION PROCESSING SYSTEMS

Highlights

- This work observed the limited ability of existing activation mechanisms in ANNs to suppress irrelevant features for pure sparsity and verified the negative impact of noise from irrelevant channels on the network's final decision.
- A novel end-to-end trainable mechanism called Dual-Perspective Activation (DPA) is proposed, which efficiently identifies irrelevant channels and applies channel denoising by incorporating criteria established and updated online from both forward and backward propagation perspectives while preserving activation responses from relevant channels.
- Extensive experiments have been conducted to assess the effectiveness and generalization of the proposed DPA mechanism. DPA is parameter-free and fast and achieves remarkable performance compared to existing activation counterparts across various mainstream ANN architectures and datasets, as well as multiple tasks and domains.

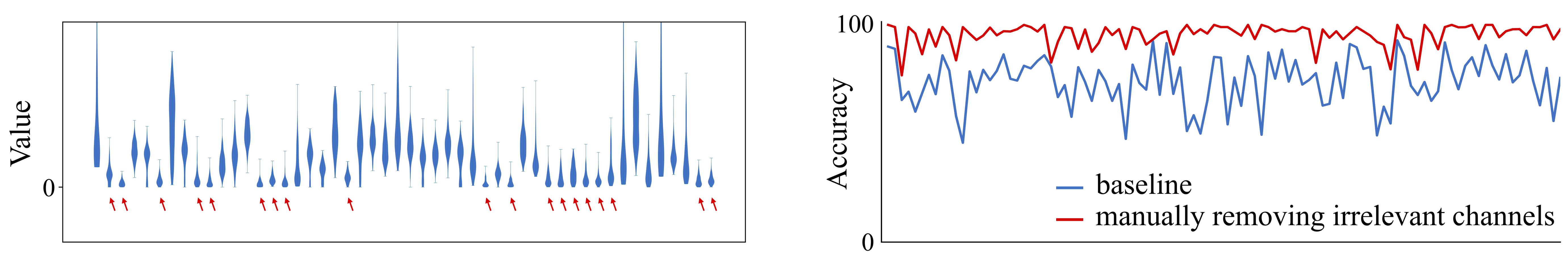
Observations



(a) Average Activation Value on Each Channel — Activation Value on Each Channel (100 samples of category 2)

(b) Average Activation Gradient on Each Channel — Activation Gradient on Each Channel (100 samples of category 2)

Observation 1: Whether from the perspective of (a) forward propagation or (b) backward propagation, intra-class consistency and inter-class differences can be observed. Each category is only highly correlated with sparse and specific channels, indicating that a significant proportion of channels are redundant and ideally should not generate any responses. Additionally, the judgments from these two perspectives in opposite directions share an overlap that points to some common channels.



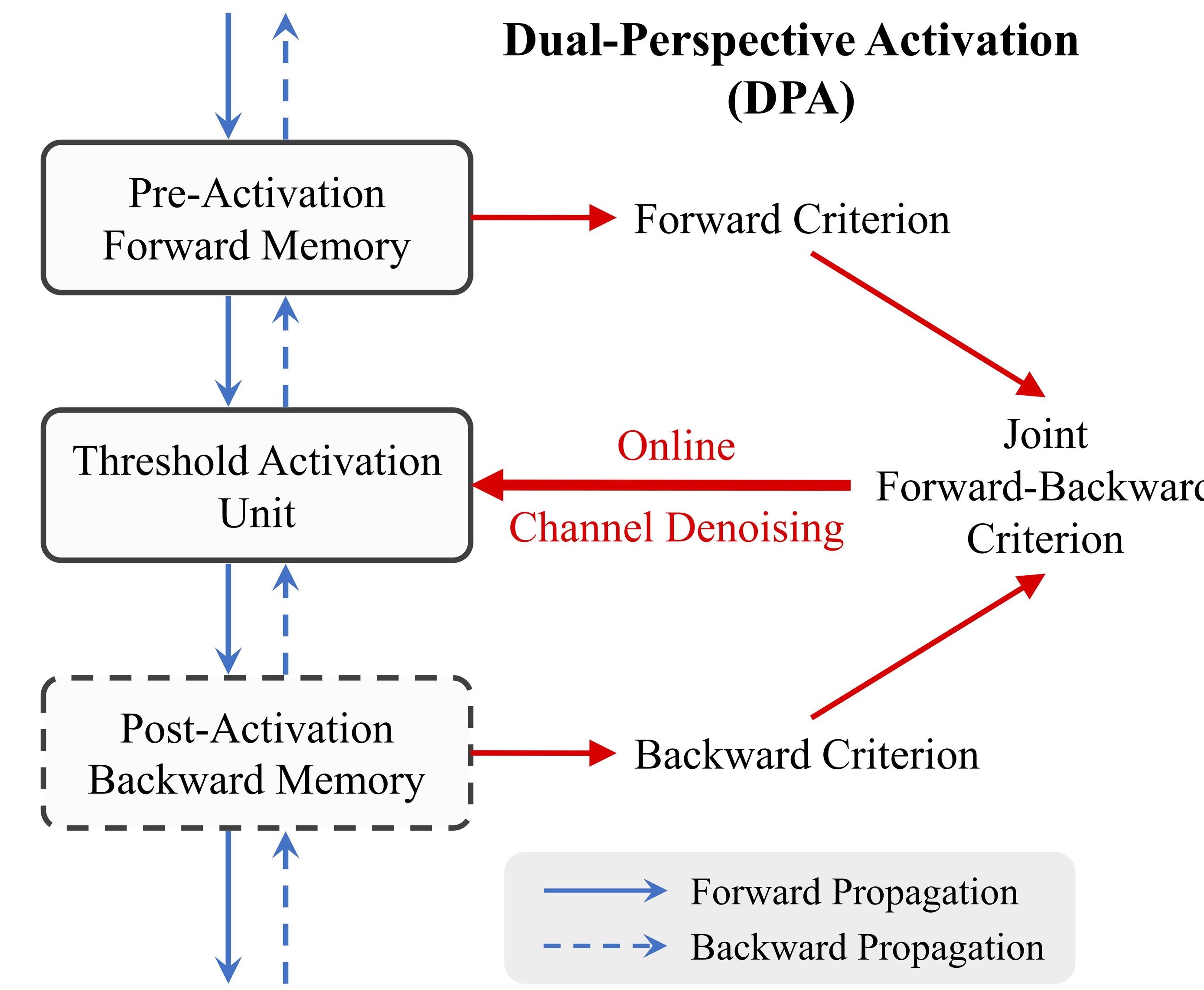
(a) Channel (b) Category

Observation 2: (a) Potential irrelevant channels are indicated by the red arrows, and it is evident that the current activation mechanism, such as ReLU, does not entirely suppress irrelevant channels, as a considerable number of responses still persist in these channels. (b) After *manually* removing the irrelevant channels for each category, there is a substantial improvement in the *training* accuracy, suggesting that the responses from irrelevant channels adversely affect the network's final decision.

Method

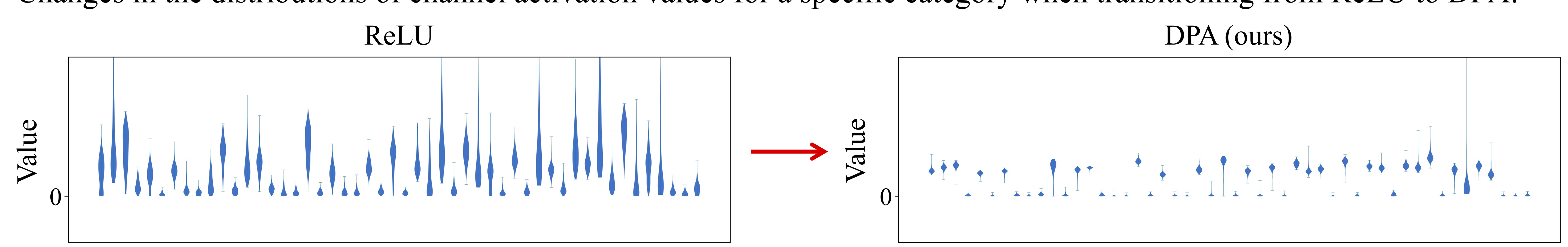
The proposed method consists of Pre-Activation Forward Memory (PreA-FM), Threshold Activation Unit (TAU), and Post-Activation Backward Memory (PostA-BM).

- TAU processes the input signal of the network.
- PreA-FM tracks the historical response value of each category before TAU in real time.
- PostA-BM tracks the historical response gradient of each category after TAU in real time.
- By utilizing the memories from both PreA-FM and PostA-BM, a joint forward-backward criterion is established and updated online to measure the relevance of channels, subsequently filtering out irrelevant channels.
- Under the guidance of this real-time updated criterion, channel denoising (regularization) is performed to suppress responses from irrelevant channels during training.



Experiments

Changes in the distributions of channel activation values for a specific category when transitioning from ReLU to DPA:



Top-1 Acc / %		Softplus	ELU	SELU	SiLU	ReLU	GELU	GDN	DPA
CIFAR-10	ViT-Tiny	84.3	82.0	79.4	85.5	89.9	89.2	81.8	**91.3**
	CaiT-XXS	82.5	80.7	78.4	86.6	89.4	88.7	80.0	**91.4**
	PVT-Tiny	90.6	89.3	85.4	92.5	93.0	92.8	82.8	**93.8**
	TNT-Small	88.3	85.4	83.7	90.5	90.8	91.1	85.1	**92.4**
CIFAR-100	ViT-Tiny	62.4	60.0	57.5	65.5	65.7	65.4	59.4	**70.5**
	CaiT-XXS	60.4	59.3	55.8	63.9	65.8	65.5	56.2	**68.5**
	PVT-Tiny	69.5	69.3	65.7	70.2	70.9	70.6	64.4	**75.3**
	TNT-Small	65.2	63.8	60.9	65.1	65.4	64.4	62.5	**72.0**
ImageNet-1K	ViT-Tiny	70.0	64.2	63.1	66.9	70.9	70.4	65.2	**72.2**
	CaiT-XXS	70.3	68.1	66.7	73.2	74.0	73.6	66.1	**75.0**
	PVT-Tiny	71.5	69.2	68.5	72.8	73.7	73.5	66.5	**75.2**
	TNT-Small	72.0	70.7	70.3	71.5	73.4	73.3	68.2	**77.8**

Top-1 Acc / %		Softplus	ELU	SELU	SiLU	ReLU	GELU	GDN	DPA
CIFAR-10	AlexNet	85.6	86.1	85.7	86.0	86.0	85.8	85.4	**86.4**
	VGG-11	91.3	92.0	91.5	91.9	**92.2**	91.9	91.1	**92.2**
	MobileNet	87.4	87.7	87.2	**87.8**	87.4	87.4	87.0	**87.8**
	ResNet-18	94.6	94.7	94.6	**95.1**	95.0	94.9	94.0	**95.1**
CIFAR-100	AlexNet	57.6	58.4	58.1	58.1	57.2	57.4	56.8	**58.5**
	VGG-11	69.6	69.9	69.7	69.9	70.2	70.0	70.1	**70.3**
	MobileNet	65.4	65.5	65.6	65.2	66.0	65.4	64.8	**67.2**
	ResNet-18	75.5	75.7	75.6	76.1	75.7	75.6	74.3	**76.8**
ImageNet-1K	AlexNet	56.1	56.3	56.1	56.4	56.5	56.4	55.6	**57.5**
	VGG-11	68.4	68.2	67.8	69.0	69.0	69.1	68.1	**69.7**
	MobileNet	67.2	66.7	67.1	67.4	68.1	68.2	66.3	**68.9**
	ResNet-18	69.3	69.4	68.9	69.7	69.7	69.4	68.3	**70.3**