JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

SCConv: Spatial and Channel Reconstruction Convolution for Feature Redundancy

Jiafeng Li[1] , Ying Wen[1*], Lianghua He[2]

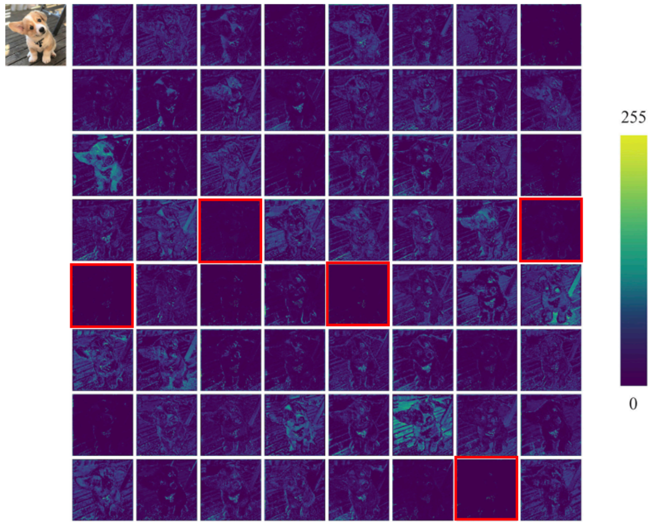
[1]School of Communication and Electronic Engineering, East China Normal University, China

[2]Department of Computer Science and Technology, Tongji University, China

Tag: TUE-PM-194

Overview

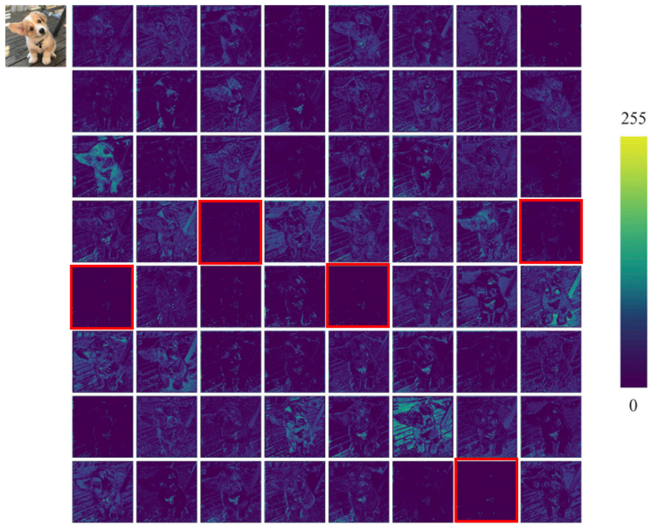
- In CNN networks, deep features tend to be non-compact, partly due to convolutional layers extracting redundant features, as shown in Figure, which often waste tremendous storage and computational resources.
- We propose a two-step procedure to exploit spatial and channel redundancy of intermediate features for CNN compression, termed as SCConv (Spatial and Channel reconstruction Convolution).



The visualization of feature maps within a layer of a ResNet model trained on the ImageNet dataset for classification.

Overview

- In CNN networks, deep features tend to be non-compact, partly due to convolutional layers extracting redundant features, as shown in Figure, which often waste tremendous storage and computational resources.
- We propose a two-step procedure to exploit spatial and channel redundancy of intermediate features for CNN compression, termed as SCConv (Spatial and Channel reconstruction Convolution).



The visualization of feature maps within a layer of a ResNet model trained on the ImageNet dataset for classification.

Overview

- The proposed SCConv module tackles feature redundancy in both spatial and channel dimensions through the integration of two units: the spatial reconstruction unit (SRU) and the channel reconstruction unit (CRU).
- SRU employs a ***Separate***-and-***Reconstruct*** operation, which effectively separates redundant features and reconstructs them to suppress spatial redundancy and enhance feature representation.
- CRU utilizes a ***Split***-***Transform***-and-***Fuse*** strategy to further diminish the redundancy in channel dimension while saving computational costs and storage requirements.

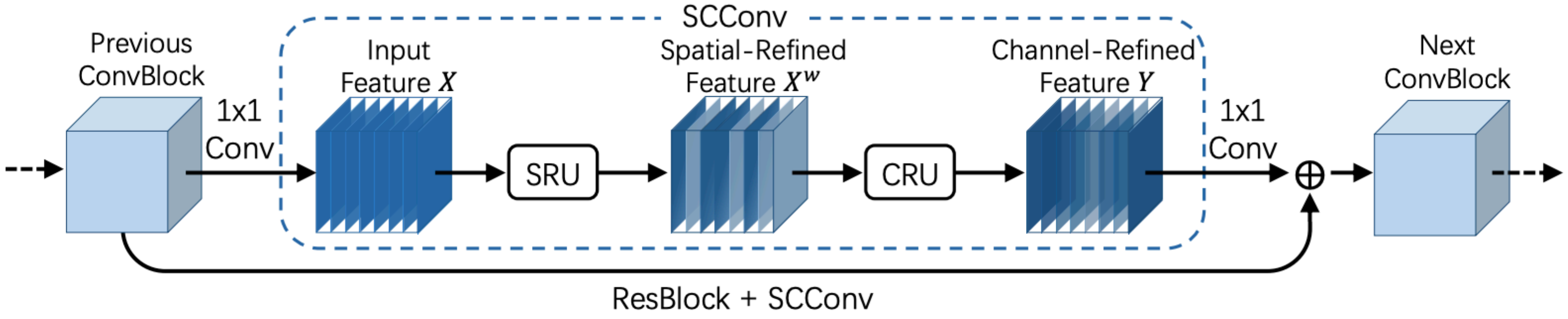


Figure 1. The architecture of SCConv integrated with Spatial Reconstruction Unit (SRU) and Channel Reconstruction Unit (CRU). This figure shows the exact position of our SCConv module within a ResBlock.

Overview

- The proposed SCConv module tackles feature redundancy in both spatial and channel dimensions through the integration of two units: the spatial reconstruction unit (SRU) and the channel reconstruction unit (CRU).
- SRU employs a ***Separate***-and-***Reconstruct*** operation, which effectively separates redundant features and reconstructs them to suppress spatial redundancy and enhance feature representation.
- CRU utilizes a ***Split***-***Transform***-and-***Fuse*** strategy to further diminish the redundancy in channel dimension while saving computational costs and storage requirements.

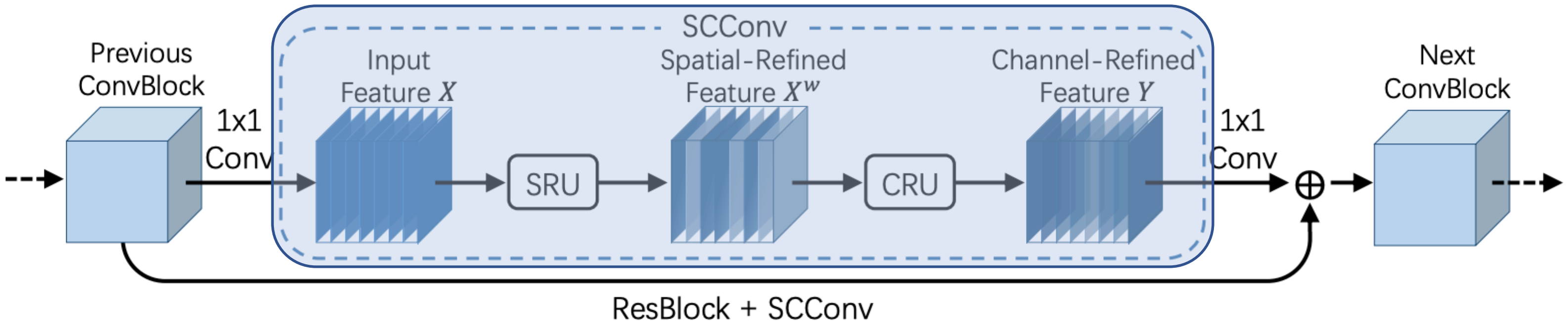


Figure 1. The architecture of SCConv integrated with Spatial Reconstruction Unit (SRU) and Channel Reconstruction Unit (CRU). This figure shows the exact position of our SCConv module within a ResBlock.

Overview

- The proposed module can be seamlessly embedded into various architectures without any additional modifications, augmenting feature extraction and representation capabilities.
- The proposed method not only cuts down on the number of model parameters and FLOPs, but also achieves an improved trade-off between efficiency and performance.

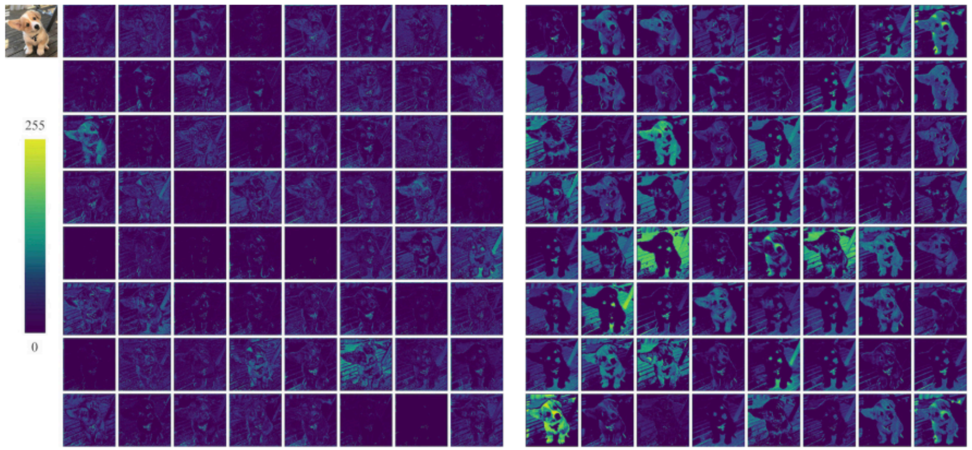


Figure 5. Left: Features from the first-stage of original ResNet50, Right: Features from the first-stage of SRU-embedded ResNet50.

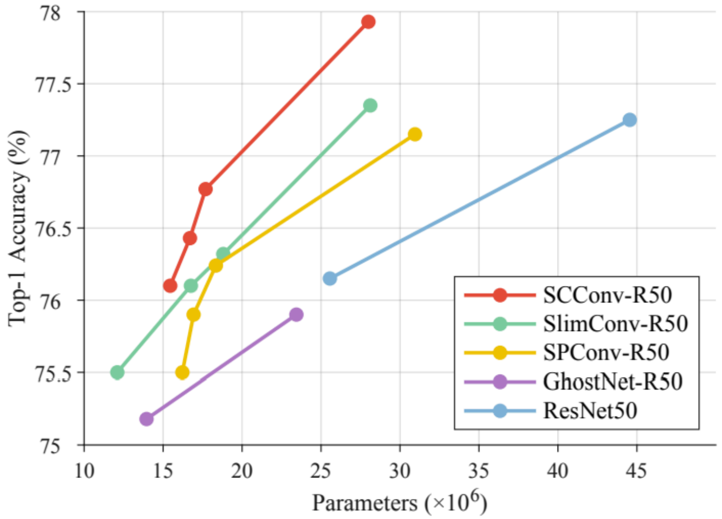


Figure 6. Top1-Accuracy v.s. FLOPs for ResNet50 on ImageNet.

SCConv: Spatial and Channel Reconstruction Convolution for Feature Redundancy

More Details

Motivation

- How to reduce the feature redundancy and improve the efficiency of the neural network?
  - Existing methods either focus on reducing the redundancy in channel dimension or in spatial dimension
  - Various model compression strategies have limitations and inefficiency in dealing with inherent redundancy.
- This paper aims at reducing the inherent redundancy in both dimension and achieving smaller model sizes, improved computational efficiency and with significant promotion of accuracy.

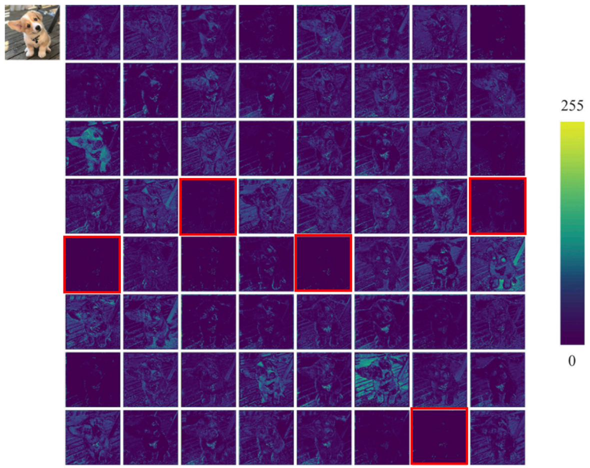


Figure. The visualization of feature maps within a layer of a ResNet model trained on the ImageNet dataset for classification

Method

- We propose an efficient convolution module called SCConv that consists of two units: spatial reconstruction unit (SRU) and channel reconstruction unit (CRU).
  - SRU and CRU are arranged in a sequential manner to reduce redundancy, improve feature representation, and facilitate more efficient feature learning.

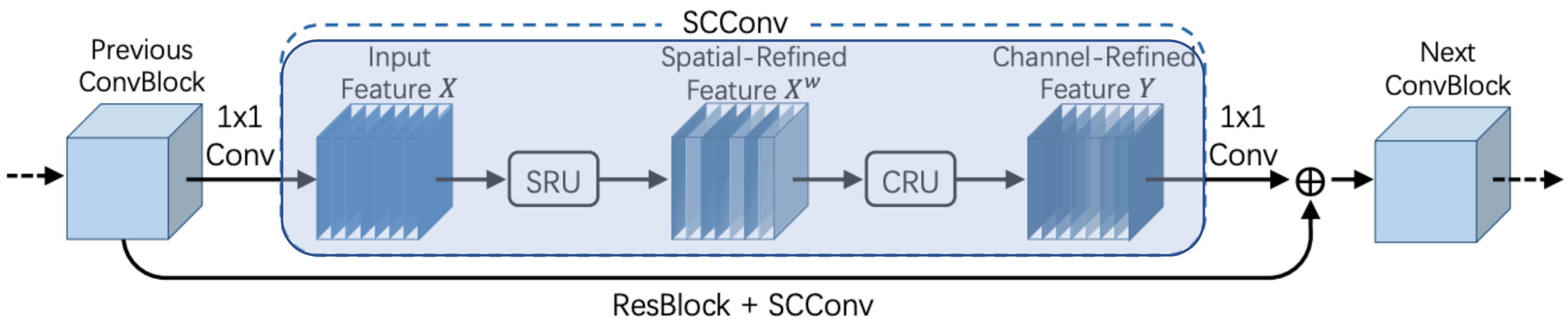


Figure 1. The architecture of SCConv integrated with Spatial Reconstruction Unit (SRU) and Channel Reconstruction Unit (CRU). This figure shows the exact position of our SCConv module within a ResBlock.

Method

- The proposed Spatial Reconstruction Unit (SRU):
  - Utilizes a *Separate*-and-*Reconstruct* operation to suppress the redundant features in spatial dimension

1. *Separate* operation separate those informative feature maps from less informative ones corresponding to the spatial content, assessed by the GN.
2. *Reconstruct* operation reconstruct those with rich information, while compressing the less informative ones to suppress redundant features in the spatial dimension to enhance the representation ability.

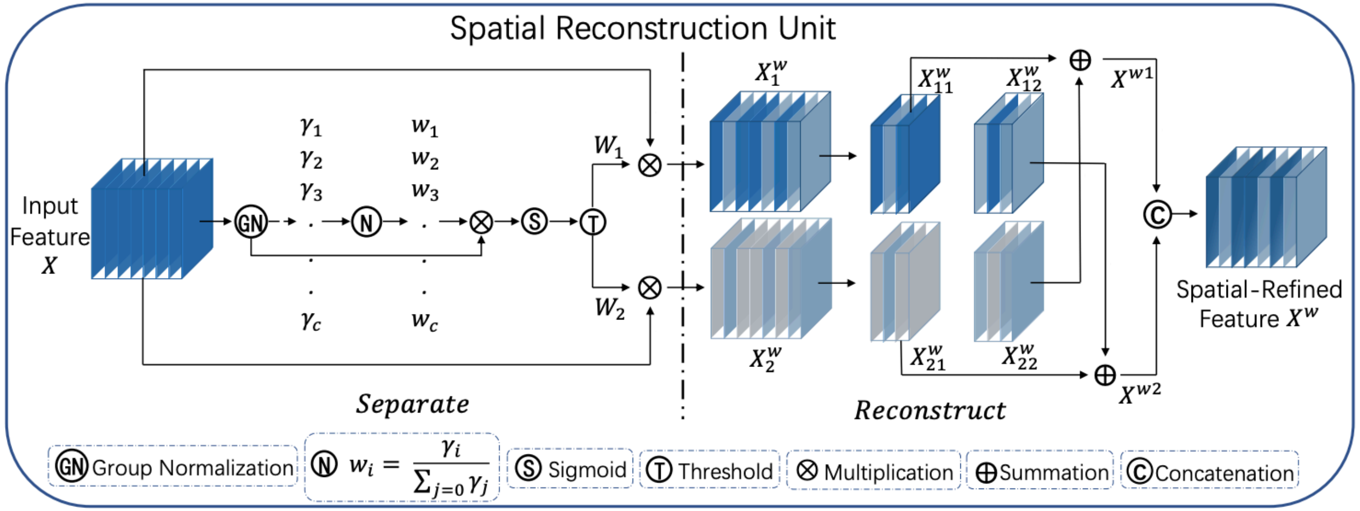


Figure 2. The architecture of Spatial Reconstruction Unit.

Method

- To further exploit channel redundancy, we introduce Channel Reconstruction Unit (CRU) :
  - Utilizes a *Split*-*Transform*-and-*Fuse* strategy.

1. *Split* operation aims to split and squeeze the channels of feature maps for its computing efficiency.
2. *Transform* operation processes two parts through different transformation, one part to extract rich representative features while another part redundant features to decrease inherent redundant computing.
3. *Fuse* operation aims to adaptively merge the two transformed features in a channel-wise manner.

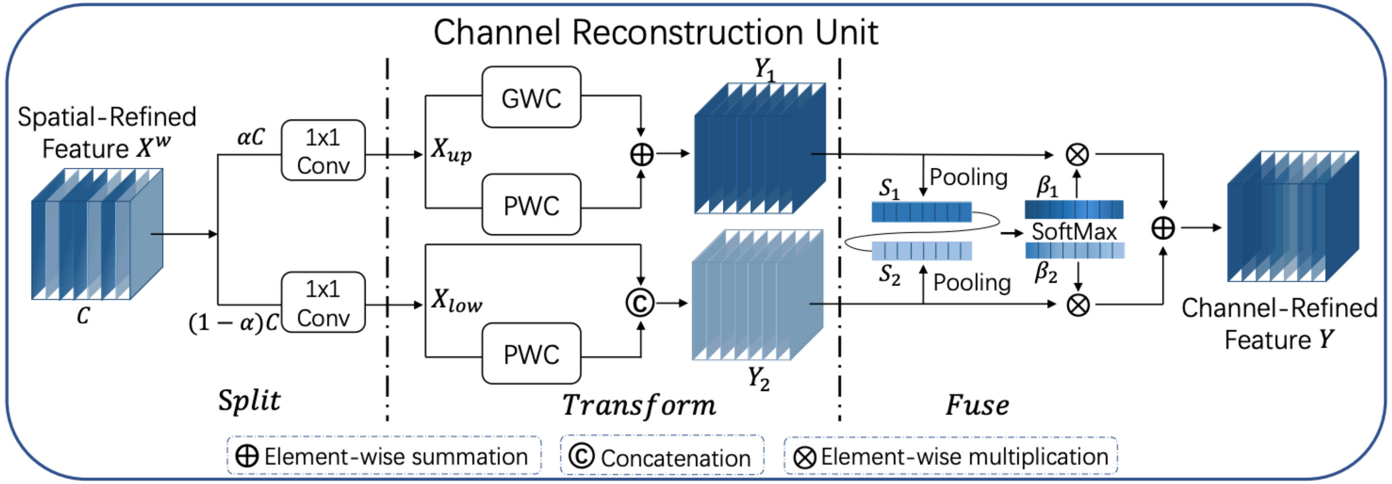
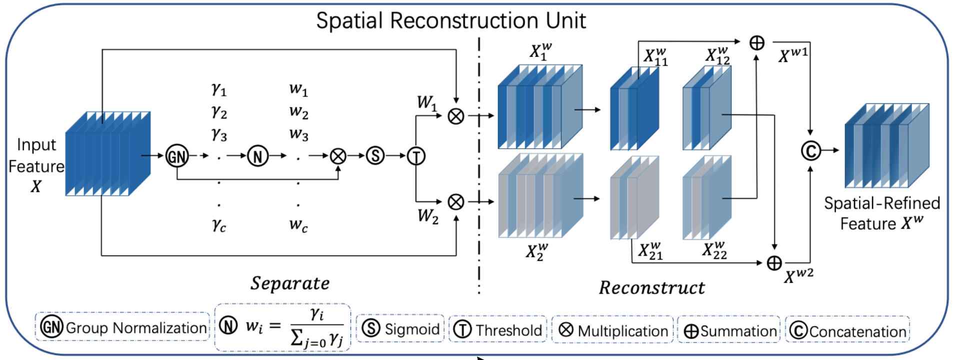
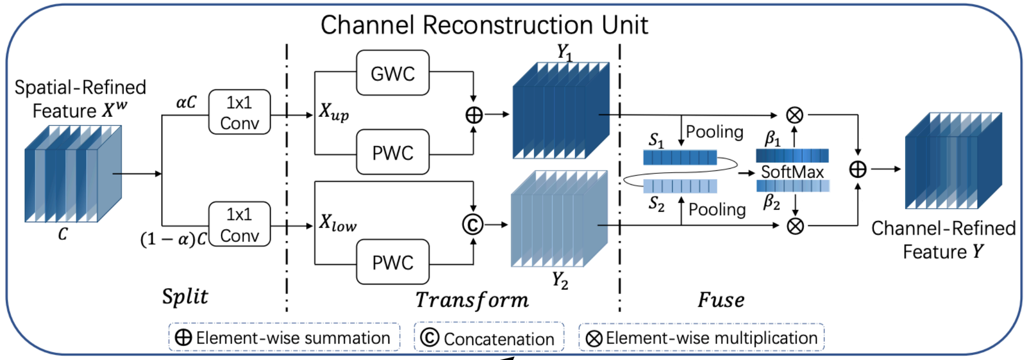


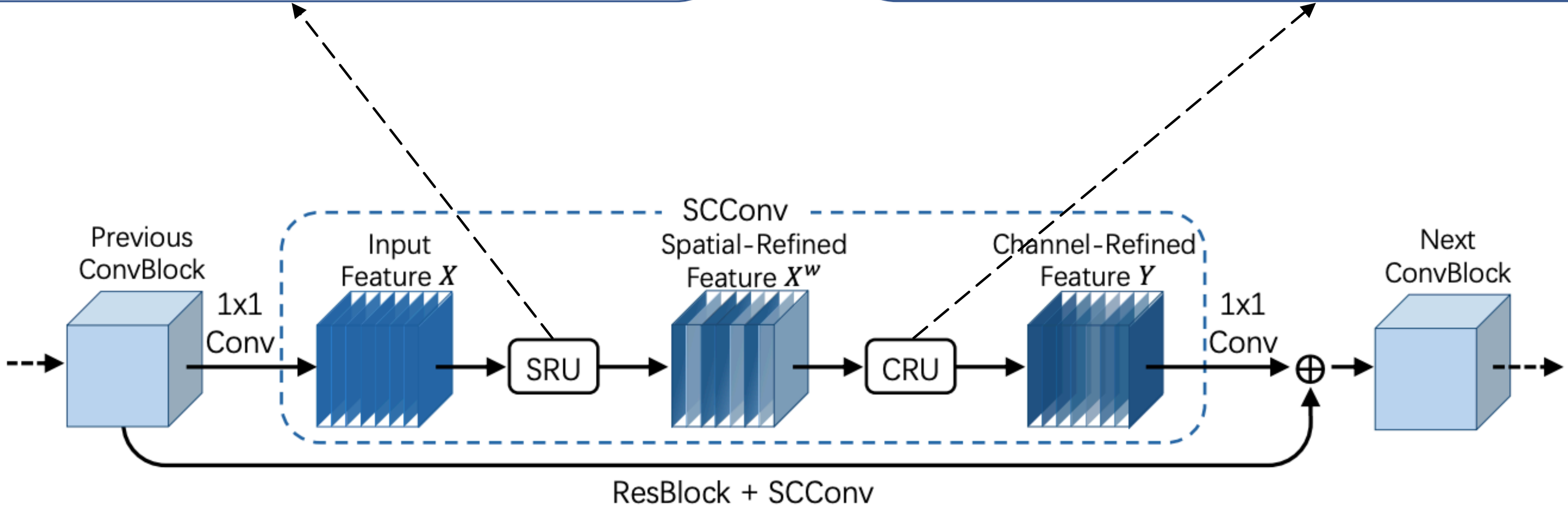
Figure 3. The architecture of Channel Reconstruction Unit.

Method

➢ Integration of SRU and CRU in a sequential manner







Experiments

- Comparison with related approaches for Image Classification on CIFAR
  - Ablation studies find that the spatial-first order (S+C) and the optimal split ratio α = 1/2 settings can achieve a better trade-off between performance and efficiency.

Table 1. Experimental results with different combination methods of SRU and CRU on CIFAR-100 dataset. (S and C is short for SRU and CRU respectively)

Description	Params	FLOPs	Top-1ACC(%)
ResNet50	23.71M	1.30G	78.60
ResNet50 + S	23.53M	1.30G	79.59
ResNet50 + C	14.74M	843.81M	79.21
ResNet50 + C & S	14.74M	843.81M	79.26
ResNet50 + C + S	14.74M	843.81M	79.54
ResNet50 + S + C	14.74M	843.81M	79.89

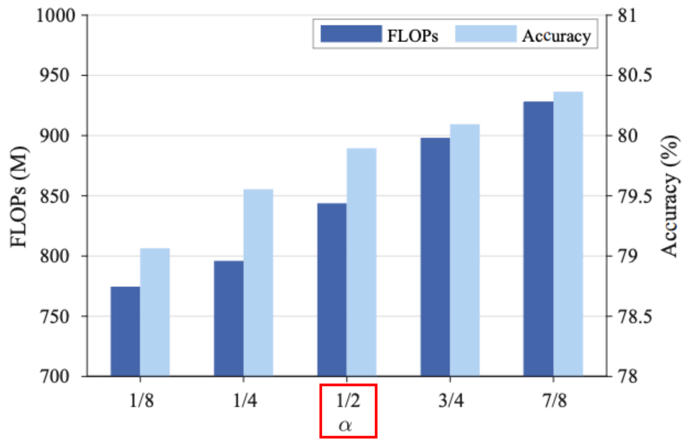


Figure 4. The trade-off between FLOPs and Accuracy on CIFAR-100 with different split ratios α in SCConv-embedded ResNet50.

Experiments

- Comparison with related approaches for Image Classification on CIFAR
  - Ablation studies find that the spatial-first order (S+C) and the optimal split ratio α = 1/2 settings can achieve a better trade-off between performance and efficiency.

Table 2. Comparison of SOTA methods for common CNN architectures over Top-1 accuracy, the number of parameters and FLOPs on CIFAR-10 and CIFAR-100 dataset.

Network Architecture	FLOPs	Params	CIFAR-10 ACC(%)	CIFAR-100 ACC(%)
ResNet56 (R56)	126.84M	0.86M	93.27	71.50
OctConv-R56 (α=1/2)	126.84M	0.62M	93.11	70.93
SPConv-R56 (α=1/2)	87.95M	0.58M	93.49	71.51
GhostNet-R56 (s=2)	**68.35M**	**0.45M**	92.35	70.42
SlimConv-R56 (k=1)	88.82M	0.60M	93.51	71.71
TiedConv-R56 (b=2)	94.99M	0.54M	92.87	71.10
SCConv-R56 (Ours)	79.55M	0.52M	**94.12**	**72.56**
ResNet50 (R50)	1.30G	23.52M	95.09	78.60
OctConv-R50 (α=1/2)	727.26M	23.52M	95.25	78.91
GhostNet-R50 (s=2)	**632.45M**	**14.56M**	94.30	77.54
SlimConv-R50 (k=4/3)	853.50M	14.83M	95.05	78.85
SPConv-R50 (α=1/2)	972.31M	16.14M	95.32	79.23
SlimConv-R50 (k=1)	942.38M	16.78M	95.35	79.26
TiedConv-R50 (b=2)	998.78M	15.03M	95.44	79.52
SCConv-R50 (Ours)	831.18M	14.69M	**95.92**	**79.89**
ResNeXt-29 (RX29)	4.55G	28.27M	95.68	81.54
GhostNet-RX29 (s=2)	3.60G	**22.22M**	95.14	80.23
SPConv-RX29 (α=1/2)	4.79G	31.17M	96.03	81.76
SlimConv-RX29 (k=4/3)	4.38G	28.24M	95.85	81.93
TiedConv-RX29 (b=2)	5.96G	24.67M	95.66	81.32
SCConv-RX29 (Ours)	**3.57G**	22.31M	**96.20**	**82.56**
WideResNet-28 (WRN28)	5.96G	36.55M	95.21	79.40
GhostNet-WRN28 (s=2)	3.98G	22.12M	95.25	79.27
SPConv-WRN28 (α=1/2)	4.16G	24.20M	95.37	80.25
SlimConv-WRN28 (k=1)	4.25G	25.00M	95.43	79.52
TiedConv-WRN28 (b=2)	4.54G	22.03M	95.48	78.61
SCConv-WRN28 (Ours)	**3.75G**	**21.12M**	**95.64**	**80.83**
DenseNet-121 (D121)	898.23M	7.05M	95.09	79.43
GhostNet-D121 (s=2)	**517.36M**	**5.04M**	93.96	78.51
SPConv-D121 (α=1/2)	641.54M	5.69M	95.15	79.64
SlimConv-D121 (k=1)	670.21M	5.97M	94.63	78.90
TiedConv-D121 (b=2)	695.68M	5.45M	95.23	79.73
SCConv-D121 (Ours)	594.34M	5.45M	**95.37**	**80.24**

Table 1. Experimental results with different combination methods of SRU and CRU on CIFAR-100 dataset. (S and C is short for SRU and CRU respectively)

Description	Params	FLOPs	Top-1ACC(%)
ResNet50	23.71M	1.30G	78.60
ResNet50 + S	23.53M	1.30G	79.59
ResNet50 + C	14.74M	843.81M	79.21
ResNet50 + C & S	14.74M	843.81M	79.26
ResNet50 + C + S	14.74M	843.81M	79.54
ResNet50 + S + C	14.74M	843.81M	79.89

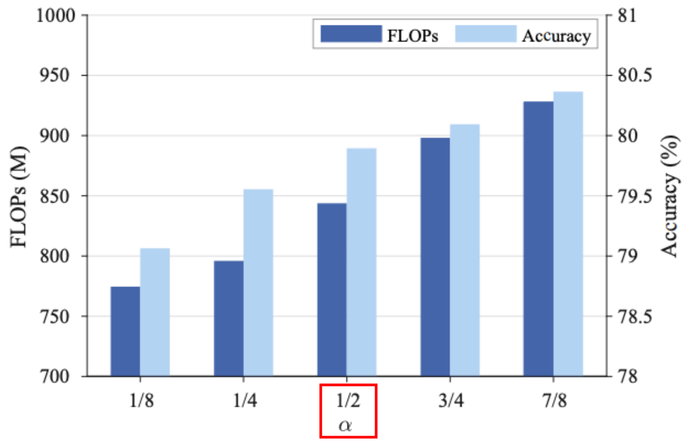


Figure 4. The trade-off between FLOPs and Accuracy on CIFAR-100 with different split ratios α in SCConv-embedded ResNet50.

Experiments

➢ Comparison with related approaches for Image Classification on ImageNet and Object Detection on COCO

Table 3. Image classification results on ImageNet-1K dataset.

Network Architecture	FLOPs(G)	Params(M)	Top-1 (%)
ResNet50 (R50) (Baseline)	4.09	25.56	76.15
Versatile-R50 (NIPS2018)	**1.80**	18.7	75.50
GhostNet-R50_s=2 (CVPR2020)	2.15	13.95	75.18
SlimConv-R50_k=8/3 (TIP2021)	1.88	**12.10**	75.32
SPConv-R50_α=1/2 (IJCAI2020)	2.97	18.34	76.26
OctConv-R50_α=1/2 (CVPR2020)	2.40	25.56	76.34
SlimConv-R50_k=4/3 (TIP2021)	2.65	16.76	76.12
PfLayer-R50_max (ICLR2022)	2.90	18.00	76.15
SlimConv-R50_k=1 (TIP2021)	3.00	18.81	76.32
TiedConv-R50_b=2 (AAAI2021)	3.19	17.07	76.04
SCConv-R50_α=1/2	2.70	16.78	76.41
SCConv-R50_α=3/4	2.87	17.69	**76.79**
ResNet_101(R101)(Baseline)	7.83	44.55	77.25
SCConv-R101	**4.90**	**28.00**	**77.93**

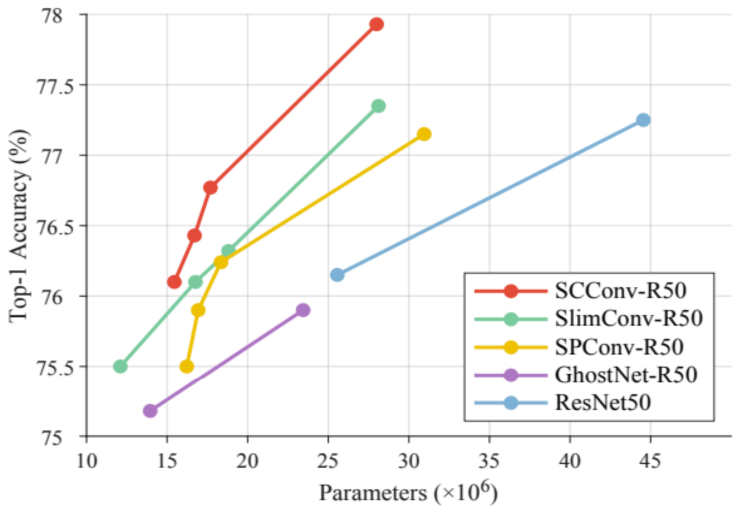


Figure 6. Top1-Accuracy v.s. FLOPs for ResNet50 on ImageNet.

Table 4. Object detection experiments on the PASCAL VOC 2007 and 2012 dataset.

Backbone	Params(M)/FLOPs(G)	AP@.5	AP@.75	mAP@[.5,.95]
ResNet50(R50)	25.56/63.09	77.89	55.31	52.26
SPConv-R50	19.76/49.23	78.05	55.47	52.48
SlimConv-R50	18.81/47.12	77.96	55.38	52.42
SCConv-R50	**16.78/41.36**	**78.68**	**56.26**	**53.16**
ResNet101(R101)	44.55/121.3	79.23	56.31	53.32
SCConv-R101	**27.90/75.26**	**80.36**	**57.05**	**54.12**

Table 5. Object detection results on MS COCO val2017.

Backbone	Params(M)/FLOPs(G)	AP@.5	AP@.75	mAP@[.5, .95]
ResNet50(R50)	25.56/63.09	54.2	37.4	35.2
SPConv-R50	19.76/49.23	54.5	37.6	35.3
SlimConv-R50	18.81/47.12	54.0	37.1	35.0
SCConv-R50	**16.78/41.36**	**55.1**	**38.2**	**35.6**

Conclusion

- This work introduces SCConv module, serving as a practical solution for addressing feature redundancy in the convolutional layers across the spatial and channel dimensions.

- Concretely, our approach adopts a two-step procedure where SRU and CRU are sequentially applied to reduce redundancy and enhance feature representation. This plug-and-play module provides a generic alternative to standard convolution, facilitating an improved trade-off between efficiency and performance.

- Experimental results demonstrate that the proposed method can substantially save the computing load yet obtain promising model performance on challenging tasks.

- We anticipate that the proposed method will inspire further research in the development of more efficient architectural designs.

Thanks