

Supplementary Materials of SACoD: Sensor Algorithm Co-Design Towards Efficient CNN-powered Intelligent PhlatCam

Yonggan Fu¹, Yang Zhang², Yue Wang¹, Zhihan Lu¹, Vivek Boominathan¹,
Ashok Veeraraghavan¹, Yingyan Lin¹
¹Rice University ²MIT-IBM Watson AI Lab

1. Details of the experiment settings

1.1. Search and train on CIFAR-10/100

The search space. For the experiments on CIFAR-10/100, we adopt the same search space as FBNet [5] except each group’s stride settings, i.e., a set of blocks with the same number of output channels, for adapting to the image resolution of CIFAR-10/100. In particular, we follow the strides settings for MobileNetV2 on CIFAR-10/100 as described in [4], which is [1, 1, 2, 2, 1, 2, 1] for all the seven groups.

Search settings. Following [5], we apply the Gumbel Softmax [3] on each architecture parameter option’s contributing weights to the supernet, where the initial temperature is 3 and decays by 0.92 at the end of each epoch. Specifically, we search for 50 epochs with a batch size of 64, and update the supernet weights on half of the training dataset using an SGD optimizer with a momentum of 0.9, and an initial learning rate of 0.025 with cosine decay, and update the network architecture parameters on the other half of the training dataset using an Adam optimizer with a momentum of 0.9, and a fixed learning rate of 3e-4.

Training settings. For training the derived network architectures from scratch, we adopt an SGD optimizer with a momentum of 0.9, and an initial learning rate of 0.1 with cosine decay for 600 epochs with a batch size of 96.

1.2. Search and train on the segmentation and image translation tasks

Segmentation on Cityscapes. We follow the same search space and the exact search/training settings in [1], except that the optical layer is co-searched with the network in our SACoD. In particular, Table 1 of the main text adopts input images with a resolution of 2048×1024 to calculate the mIOU and FLOPs.

Image translation on horse2zebra and zebra2horse. We follow the same search space and the exact

search/training settings in [2], except that the optical layer is co-searched with the network in our SACoD.

2. Visualizations of the optimal masks for different tasks

To further support the analysis of the specificity of SACoD generated masks in Section 3.4 of the main text, we visualize the microscope images of the fabricated masks searched by SACoD on the CIFAR-10 and FlatCam Face datasets in Fig. 1. We can see that the optimal masks for each task look very different in visualization, motivating the necessity of task-specific mask designs to achieve decent task performance. This aligns with both the analysis in Section 3.4 and the experimental ablation studies in Section 4.6 of the main text.

3. More visualization results of the image translation tasks

We provide more visualization results of the image translation tasks in Fig. 2 in addition to Figure 6 of the main text. We can see that again the Gabor-mask baseline suffers greatly from color shift and distortion especially on the horse2zebra task, and the proposed SACoD provides better or competitive visualization effects with high-contrast textures compared with the Co-train baseline, while achieving a 92.6% reduction in FLOPs.

References

- [1] Wuyang Chen, Xinyu Gong, Xianming Liu, Qian Zhang, Yuan Li, and Zhangyang Wang. Fasterseg: Searching for faster real-time semantic segmentation. [arXiv preprint arXiv:1912.10917](#), 2019. 1
- [2] Yonggan Fu, Wuyang Chen, Haotao Wang, Haoran Li, Yingyan Lin, and Zhangyang Wang. Autogan-distiller: Searching to compress generative adversarial networks. [arXiv preprint arXiv:2006.08198](#), 2020. 1

