Attentive Fine-Grained Structured Sparsity for Image Restoration
– *Supplementary Document* –

Junghun Oh[1] Heewon Kim[1] Seungjun Nah[1,3] Cheeun Hong[1] Jonghyun Choi[4] Kyoung Mu Lee[1,2]

[1]Dept. of ECE, ASRI, [2]IPAI, Seoul National University [3]NVIDIA [4]Yonsei University

[1]{dh6dh, ghimhw, cheeun914, kyoungmu}@snu.ac.kr, [3]seungjun.nah@gmail.com, [4]jc@yonsei.ac.kr

S1. Overview

We elaborate on the validity of our proposed pruning method in detail. To summarize, we have following items:

- In Section S2, we validate the necessity of the regularization loss annealing during training.
- In Section S3, we show that our Search framework is also effective at various N:M configurations.
- In Section S4, we provide further experimental results on the adaptive inference scheme on image deblurring task.
- Finally, in Section S5, we provide the implementation details for reproducibility.

S2. Effectiveness of Scheduling λ_{reg}

Our pruning framework prunes a pretrained network by jointly optimizing the task loss and the computational amount regularization loss until the pruned model meets the target computational budget. As discussed in Section 3.5, a large constant weight for λ_{reg} leads to immature pruning. The induced performance drop is significant even after the following fine-tuning process. On the other hand, if λ_{reg} is smaller, the model fails to meet the target budget. To overcome this dilemma, we adaptively change λ_{reg} by monitoring the progress of the pruning rates in the last few epochs (Eq. (7)).

In addition to the quantitative validation in Table 3 in the main manuscript, we contrast the effect of proper λ_{reg} scheduling by loss curve as the learning progresses in Figure S1. From a constant λ_{reg} without the scheduling, the model rapidly deteriorates with the spiking task loss. In contrast, our scheduling on λ_{reg} prunes the model in a moderate speed so that the optimization becomes easier even with a relatively less number of fine-tuning epochs. These results imply the effectiveness of our λ_{reg} scheduling.

S3. Generalization to Various M Values

In Tables 1 and 2 of the main manuscript, we presented the N:M pruning results when $M = 32$ to elaborate the effectiveness of our SLS framework in an extreme pruning ratio up to roughly 93.75%. We show that our SLS generalizes to other configurations by quantitative comparisons with various M values.

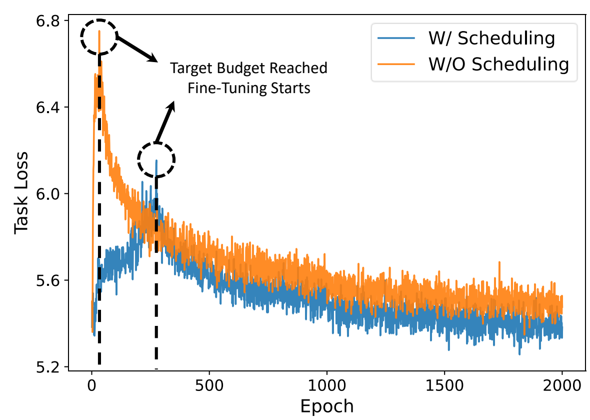

Figure S1. Task loss curves with (blue curve) and without (orange curve) λ_{reg} scheduling. The highlighted parts indicate the moment when the target budget ($\frac{1}{8}$ of the original computational costs) is reached in each case. In the case of 'W/O Scheduling', we set λ_{reg} as the final value of λ_{reg} in 'W/ Scheduling'.

Tables S1 and S2 show the pruning results on image deblurring and super-resolution models for $M \in \{8, 16, 32\}$, respectively. For all methods, the restoration performance tend to degrade when M is smaller. Nevertheless, SLS consistently outperforms others on both tasks and for all M. Empirically, SLS is generalized well on various M.

S4. Adaptive Inference on Image Deblurring

In Section 4.7 and Figure 7 in the main manuscript, we showed that our adaptive inference exhibits an improved trade-off between PSNR and GMACs for super-resolution task. We further validate the adaptive inference is also effective in image deblurring, as shown in Figure S2. Specifically, we trained an auxiliary network to predict the MSE error of the deblurred patches from 3 DMPHN models pruned by SLS and identity mapping (no processing). Using the es-

Table S1. Image deblurring performance comparisons on GOPRO dataset [39].

Model	Method	GMACs	Num. Param.	PSNR$_\uparrow$ / SSIM$_\uparrow$ / LPIPS$_\downarrow$
UNet	Unpruned	458.04	6.79M	29.46 / 0.8837 / 0.1686
	Filter pruning	115.42	1.70M	28.79 / 0.8692 / 0.1893
	One-shot (8:32)	117.24	1.70M	29.19 / 0.8771 / 0.1795
	SR-STE (8:32)	117.24	1.70M	28.85 / 0.8691 / 0.1860
	SLS (Ours)	116.64	1.55M	**29.37 / 0.8811 / 0.1740**
	One-shot (4:16)	117.24	1.70M	29.17 / 0.8765 / 0.1801
	SR-STE (4:16)	117.24	1.70M	28.81 / 0.8684 / 0.1876
	SLS (Ours)	116.31	1.51M	**29.31 / 0.8800 / 0.1751**
	One-shot (2:8)	117.24	1.70M	29.11 / 0.8741 / 0.1812
	SR-STE (2:8)	117.24	1.70M	28.75 / 0.8675 / 0.1898
	SLS (Ours)	116.18	1.49M	**29.20 / 0.8767 / 0.1794**

Table S2. Image super-resolution performace (PSNR$_\uparrow$) comparisons on benchmark datasets with the scaling factor of 4.

Model	Method	GMACs	Num. Param.	Set14 / B100 / Urban100
RFDN	Unpruned	39.86	828.75K	28.52 / 27.51 / 25.91
	Filter pruning	10.44	214.77K	16.50 / 17.32 / 15.52
	One-shot (8:32)	10.10	210.00K	28.46 / 27.47 / 25.73
	SR-STE (8:32)	10.10	210.00K	28.33 / 27.41 / 25.59
	SLS (Ours)	10.05	240.17K	**28.50 / 27.50 / 25.84**
	One-shot (4:16)	10.10	210.00K	28.44 / 27.46 / 25.73
	SR-STE (4:16)	10.10	210.00K	28.35 / 27.42 / 25.60
	SLS (Ours)	10.03	250.66K	**28.50 / 27.49 / 25.82**
	One-shot (2:8)	10.10	210.00K	28.43 / 27.46 / 25.71
	SR-STE (2:8)	10.10	210.00K	28.40 / 27.42 / 25.63
	SLS (Ours)	10.02	260.58K	**28.48 / 27.49 / 25.82**

timated MSE and the expected computational costs for the input patch, the adaptive inference path is determined by the Equation (8). For instance, a relatively sharp input patch can be fed to output by identity mapping or a model with high pruning ratio (*i.e.* 93.75%) and a severely blurry patch could be processed by the model with a higher accuracy. The results show that our adaptive inference scheme can find the better trade-off between efficiency and the deblurring performance. In all experimental results on the adaptive inference for both super-resolution and deblurring tasks, we report the computational costs containing the additional overhead from overlapping.

S5. Additional Implementation Details

To show the effectiveness of our SLS framework, we performed additional experiments on state-of-the-art image deblurring and super-resolution networks. For image deblurring, we used a residual UNet [40], SRN [51], and DMPHN [59]. The residual UNet used in [40] is a light-weight model in terms of GMACs while SRN [51] and DMPHN [59] are the relatively heavier models. Following [40], we used a modified version of DMPHN by removing the multi-patch hierarchy for higher baseline accuracy. All models for deblurring were trained and tested on GOPRO dataset [39]. After pretraining each model for 2,000 epochs, we learn to prune the model for 2,000 additional epochs.

For image super-resolution, we used EDSR [28], CARN [2], and RFDN [31] models. EDSR is a heavy architecture in terms of GMACs, CARN being a mid-weight, and RFDN [31] is a light-weight architecture. Specifically, we used EDSR-baseline which has 16 ResBlocks. The super-resolution models were trained on DIV2K dataset [1] and evaluated on 3 benchmark datasets, Set14 [58], B100 [35], and Urban100 [18]. The super-resolution models were pretrained for 300 epochs, followed by the pruning process for 300 additional epochs.

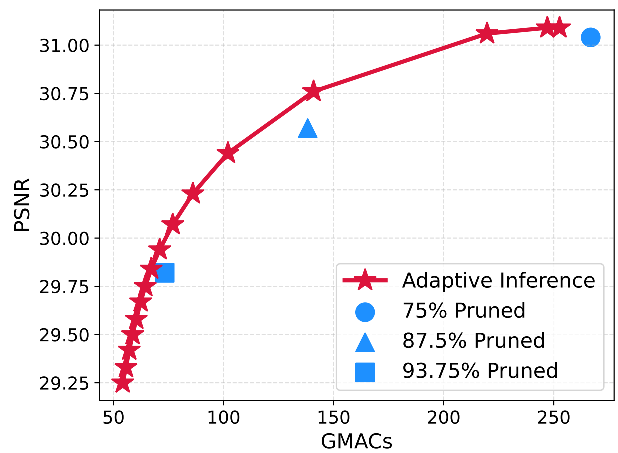

Figure S2. Results of adaptive inference for image deblurring on GOPRO dataset. We use the identity function and 3 DMPHN models pruned by SLS with different target budgets.

When applying SR-STE [63], the model was trained from scratch for 4,000 epochs for deblurring, 600 epochs for super-resolution, respectively, for fair comparison. We prune all convolutional layers with input channel dimension divisible by M.

For adaptive inference, the auxiliary convolutional network for MSE estimation consists of 0.15 M parameters and requires 3.45 GMACs for an HD image (1,280×720). For image deblurring, we crop the test images in GOPRO dataset into several patches with the size of 246×266 and overlap 3 and 5 pixels in vertical and horizontal directions, respectively. For image super-resolution, we crop the test images on Urban100 dataset into several patched with the size of 50×50 with the stride of 48. The restored patches are combined to the original image by removing the overlapping areas. We select the value of β by a uniform sampling in the range of [0,10].

License of the Used Assets

- GOPRO dataset [39] is a publicly available dataset released under CC BY 4.0 license.
- DIV2K dataset [1] is made available for academic research purposes.
- B100 dataset [35] is made available by Computer Vision Group, UC Berkeley

- Urban100 dataset [18] is made available at https://github.com/jbhuang0604/SelfExSR

Limitations

We study the layer-wise sparsity search on the general N:M configurations for extremely efficient image restoration networks. However, the latest Ampere-generation NVIDIA GPUs only support the acceleration of 2:4 sparsity pattern, limiting the acceleration of the efficient networks searched by SLS. The efficient execution of our pruned models is our future work.