Supplementary Material for Budget-aware Dynamic Spatially Adaptive Inference

Georgios Zampokas[1,2]

Christos-Savvas Bouganis[1]

Dimitrios Tzovaras[2]

[1] Imperial College London
London, UK

[2] Information Technologies Institute
Centre for Research and Technology
Hellas,
Thessaloniki, Greece

B Ablation Studies

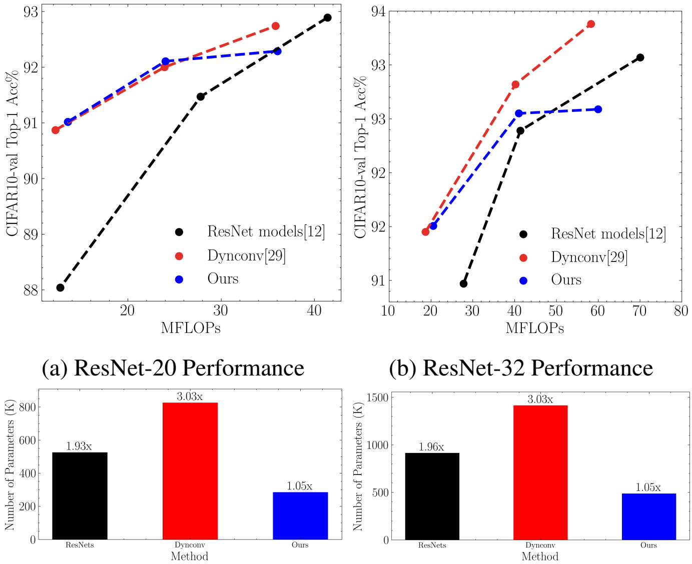
In this section, a series of ablation studies are conducted, to investigate the impact of various components and hyperparameters of our method. The ablation experiments are performed on ResNet models and follow a consistent setting. Models are are trained from scratch for image classification on the CIFAR10 [1] dataset. Training follows the schedule of [2] but optimizes for three target budgets of $\mathcal{B} = \{30\%, 50\%, 70\%\}$, unless stated otherwise (range of budgets ablation). To alleviate noise in results and extract statistically significant conclusions, also given the relatively smaller size of CIFAR10 dataset, results are obtained by performing training 10 times and keeping mean scores for all metrics.

Model Depth. The multi-target method outperforms the multiple ResNet models at most cases, in both performance to cost and memory trade-offs. When compared to Dynconv [2], we observe that our method achieves comparable performance, particularly at lower budgets, while a growing discrepancy is observer at higher budgets, especially in ResNet-32. However, the memory cost of parameters of the proposed method is significantly lower, with a mere 5% increase to base model parameter count.

Target Awareness. We examine the impact of our target aware loss weighting scheme and also compare l1 and l2 norms for the calculation of w^{b_i} weights in Figure 2a where the use of l2-norm outperforms l1-norm and no awareness weighting across all budgets. The evolution of loss weights during training (Figure 2b), shows that training focuses mostly on improving performance for the highest target budget after the halfway mark.

Knowledge Distillation. We conduct an ablation study over potential teachers between the base model (100%) and spatially adaptive models trained with Dynconv [2] to facilitate knowledge distillation. Figure 2c demonstrates that using the base model as teacher for all budgets yields the best performance. Furthermore, Figure 2d presents performance for multiple values of the KD temperature T while using the base model as the teacher, with $T = 10$ being the most optimal.

Range of budgets Finally, the performance of our method over multiple ranges $r = 20\%, 40\%, 60\%, 70\%$ of target budgets centered around 50% is examined in Figure 2e, 2f,



(a) ResNet-20 Performance (b) ResNet-32 Performance

(c) ResNet-20 Parameters (d) ResNet-32 Parameters

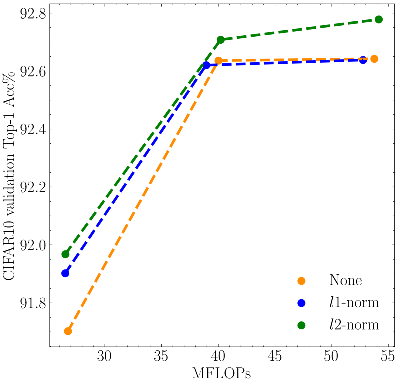
Figure 1: Model depth ablation: Results of multi-target inference with ResNet models on CIFAR10 dataset.

for both ResNet-20 and ResNet-32. As expected, smaller range leads to better performance, whereas widening the target budget support range negatively impacts performance.

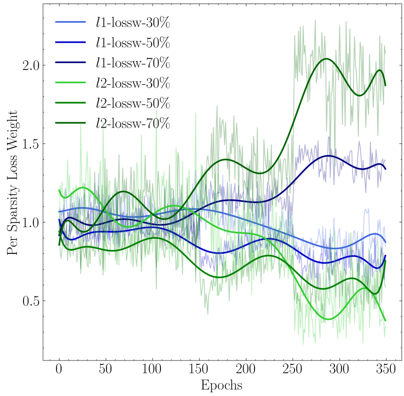
References

[1] Alex Krizhevsky. Learning multiple layers of features from tiny images. *University of Toronto*, 05 2012.

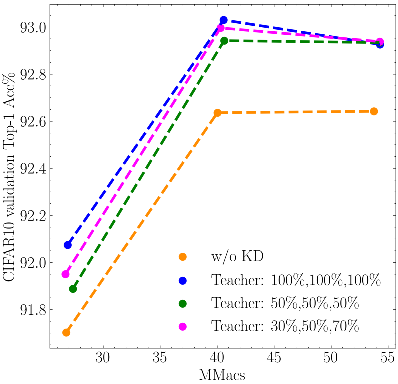
[2] Thomas Verelst and Tinne Tuytelaars. Dynamic convolutions: Exploiting spatial sparsity for faster inference. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2020. doi: 10.1109/cvpr42600.2020.00239. URL https://doi.org/10.1109%2Fcvpr42600.2020.00239.



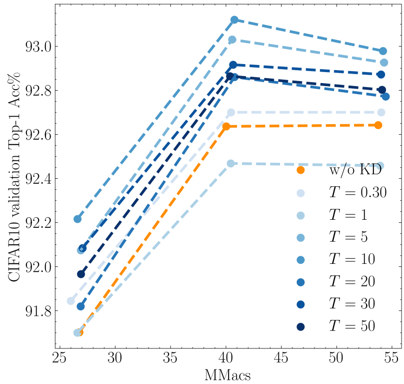
(a) Target Awareness norm



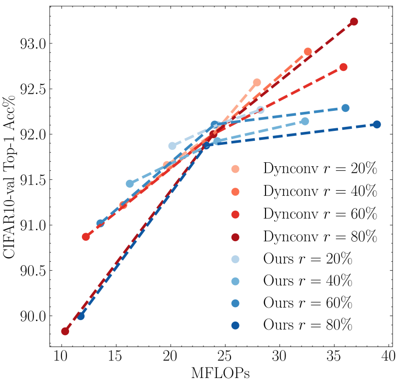
(b) Loss weights w^{b_i} during training



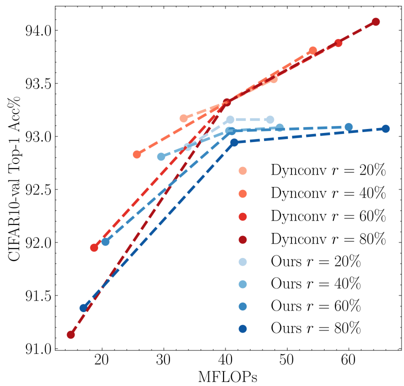
(c) KD Teacher Ablation



(d) KD Temperature T Ablation



(e) ResNet-20 Range Ablation



(f) ResNet-32 Range Ablation

Figure 2: Ablation studies for target awareness weighting, knowledge distillation setting and range of supported budgets.