Unifying Synergies between Self-supervised Learning and Dynamic Computation

Tarun Krishna[1,3]
tarun.krishna2@mail.dcu.ie
Ayush K. Rai[1,2]
ayush.rai3@mail.dcu.ie
Alexandru Drimbarean[4]
Alexandru.Drimbarean@xperi.com
Eric Arazo[1,3,5]
eric.arazo@insight-centre.org
Paul Albert[1,3,5]
paul.albert@insight-centre.org
Alan F. Smeaton[1,2]
alan.smeaton@dcu.ie
Kevin McGuinness[1,3]
kevin.mcguinness@dcu.ie
Noel E. O'Connor[1,3]
noel.oconnor@dcu.ie

[1] Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland
[2] School of Computing, Dublin City University, Dublin, Ireland
[3] School of Electronic Engineering, Dublin City University, Dublin, Ireland
[4] Xperi Corporation, Galway, Ireland
[5] Equal Contribution

1 Data Augmentation

We follow the augmentation strategy as outlined in VICReg [1] which remains a standard augmentation strategy in almost every SSL scenario. The following augmentations were applied sequentially:

- Random cropping with an area uniformly sampled with size ratio between 0.2 to 1.0, followed by resizing to size: (224×224) (ImageNet-100), (96×96) (STL-10) and (32×32) (CIFAR-10/100).
- Random horizontal flip with $p = 0.5$[1].
- Color jittering of brightness, contrast, saturation and hue, with $p = 0.8$, with ColorJitter params as (0.4, 0.4, 0.2, 0.1).
- Grayscale with $p = 0.2$.
- Solarization with $p = 0.1$.

[1] p denotes probability.

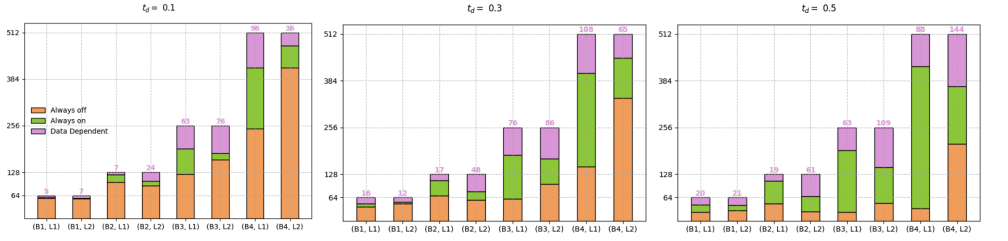

Figure 1: Learned channel distribution for CIFAR-100 with varying t_d.

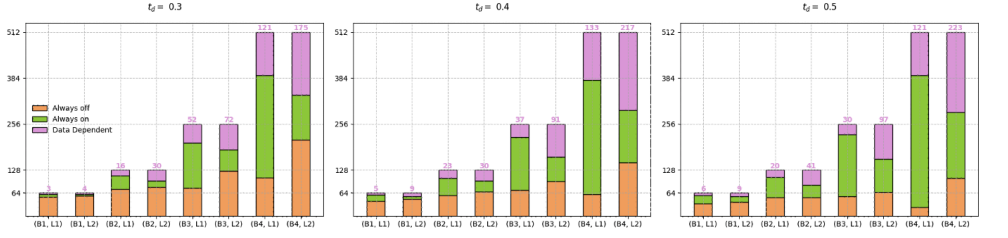

Figure 2: Learned channel distribution for ImageNet-100 with varying t_d

2 Learned Policy: Qualitative assessment

Figure 1 and 2 depicts the learned channel distribution through our gating module. In order to make fair comparison with Knowledge distillation method specifically SEED [2] and SimReg [3], student networks were sampled from R18's subspace following policy learned by our gated network Figure 1 & 2. We combine "data dependent" + "always on" channels to get the architecture of student for uniform comparison.

References

[1] Adrien Bardes, Jean Ponce, and Yann Lecun. VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning. Technical report.

[2] Zhiyuan Fang, Jianfeng Wang, Lijuan Wang, Lei Zhang, Yezhou Yang, and Zicheng Liu. Seed: Self-supervised distillation for visual representation. *arXiv preprint arXiv:2101.04731*, 2021.

[3] KL Navaneet, Soroush Abbasi Koohpayegani, Ajinkya Tejankar, and Hamed Pirsiavash. Simreg: Regression as a simple yet effective tool for self-supervised knowledge distillation. 2021.