ProX: A REVERSED ONCE-FOR-ALL NETWORK TRAINING PARADIGM FOR EFFICIENT EDGE MODELS TRAINING IN MEDICAL IMAGING

Shin Wei Lim, Chee Seng Chan, Erma Rahayu Mohd Faizal, Kok Howg Ewe
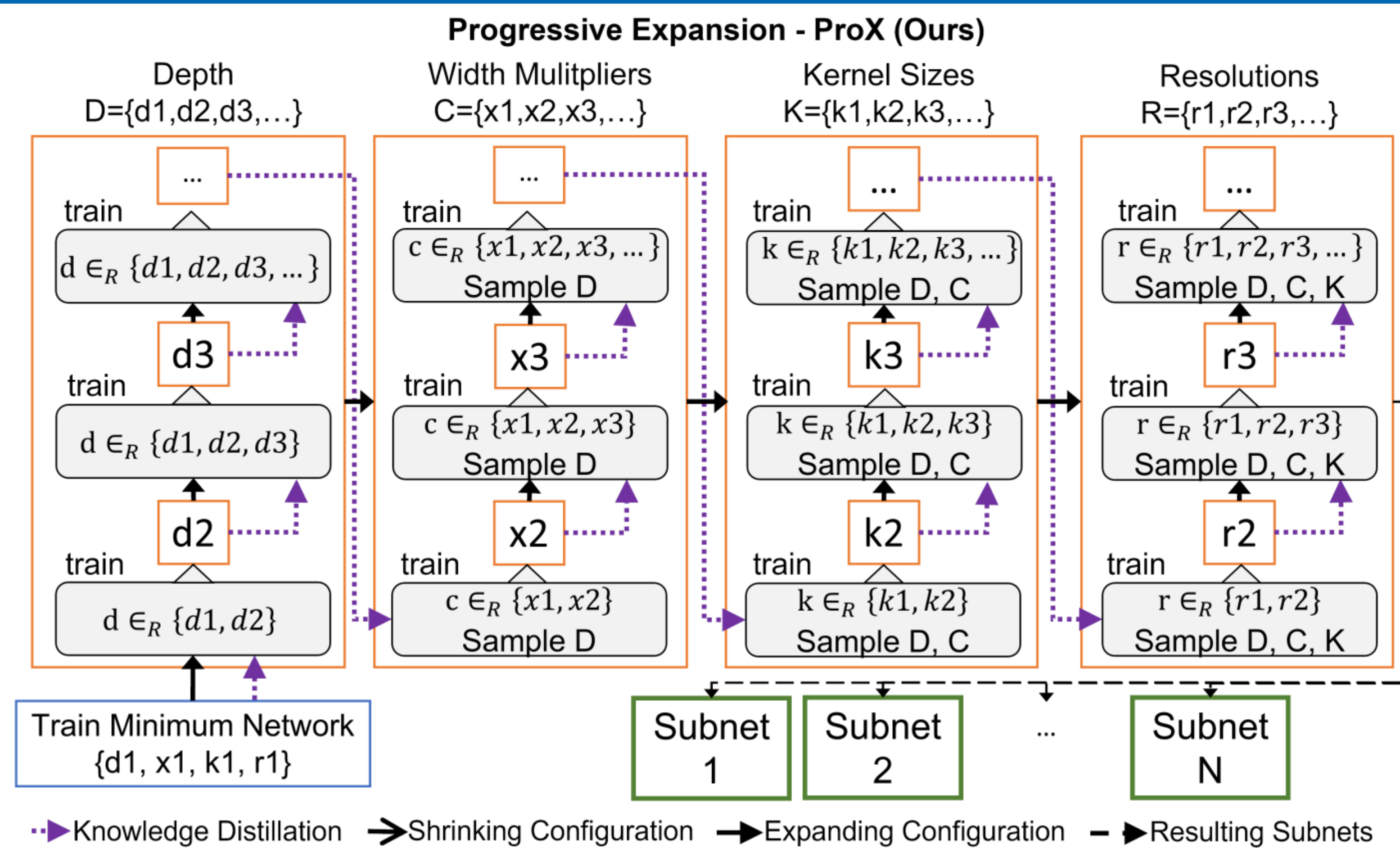
Introduction

- Digital inequality is an ongoing problem in the under-privileged community, i.e. hard to access medical imaging service.

- Power and cost efficient edge models can solve this by popularizing the medical imaging inference service access at a lower cost.

- Designing edge models that can fit well on different computing platforms with different compute capabilities is time and resource consuming.

- How can we automate the design of the edge network with a shorter time and better resource usage?

Previous Solution

- The Once-for-All (OFA) network directly hosts a large number of subnetworks without any retraining by utilising the Progressive Shrinking (*PS*) [1] training method.

- However, the *PS* method have several drawbacks:

  - Not resource friendly;

  - Requires longer training time

Proposed Method – A Reversed Paradigm

Progressive Expansion - ProX (Ours)

- Knowledge Distillation → Shrinking Configuration → Expanding Configuration ⇢ Resulting Subnets
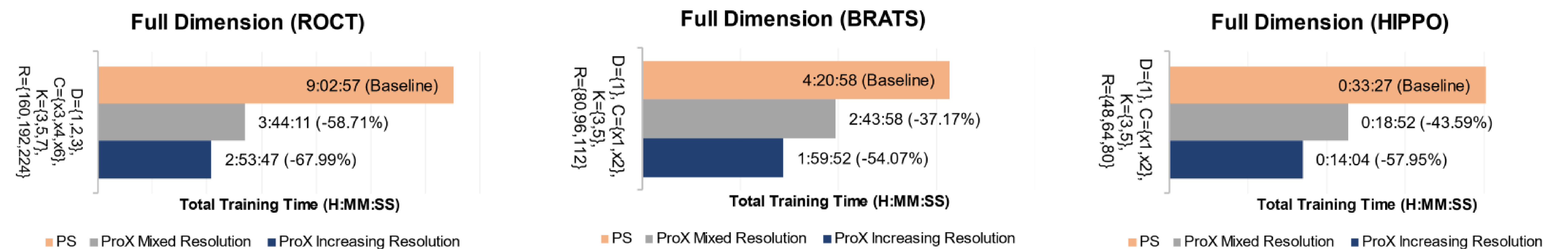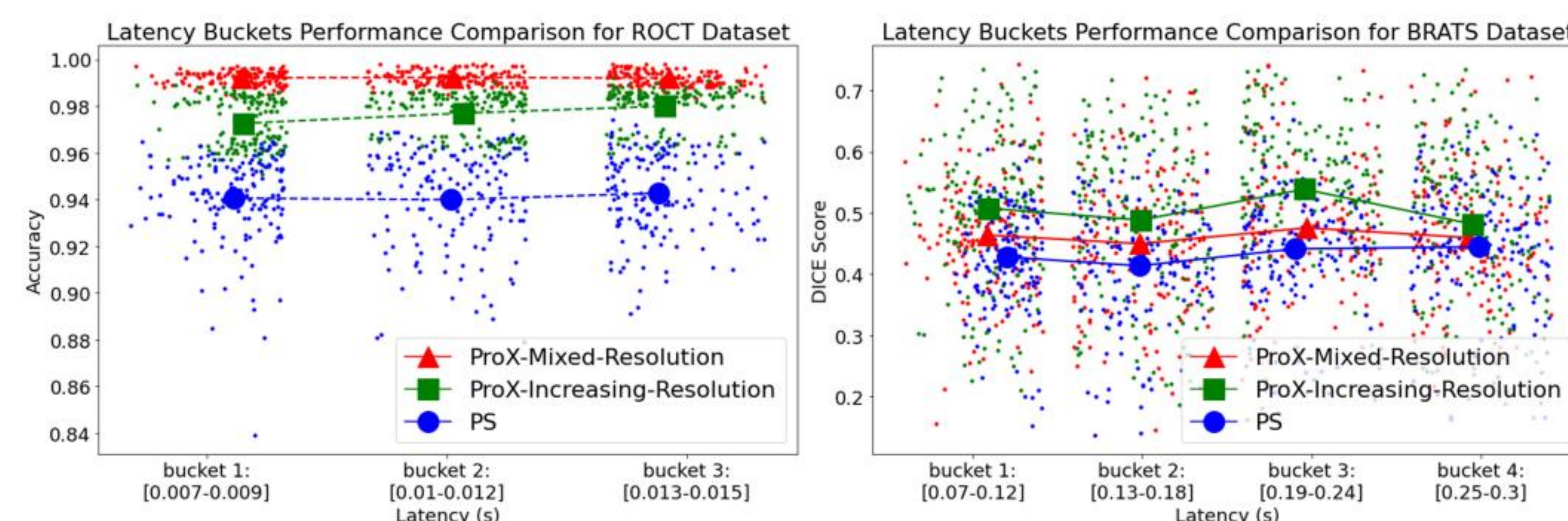
Experiment Results

- Tested on ROCT (classification – MobileNet), BRATS and HIPPO (3D segmentation – 3D-UNet) medical datasets.

- Generally, *ProX* is able to:

  a. **Monitor and inspect overparameterization earlier, avoiding training redundant layers;**
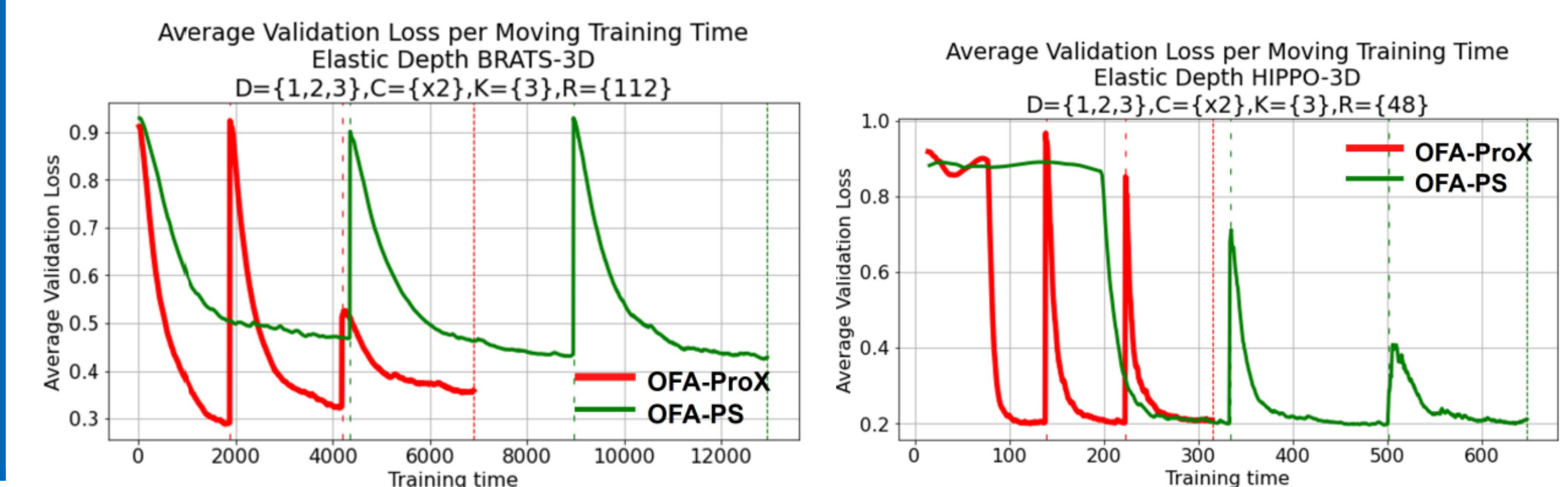
b. **Reduce OFA network training time up to 68%;**

c. **Produce higher scoring subnetworks.**

Conclusion

- By using the expanding paradigm, *ProX* method can train the OFA network faster, allowing better resource usage monitoring while producing the subnetworks with better quality (higher scores).

UNIVERSITY OF MALAYA   intel.

[1] H. Cai, et. al., "Once-for-all: Train one network and specialize it for efficient deployment," in ICLR, 2020.