Accelerating Realization of Effective Capacity in Lightweight Vision Models via Self-Competitive Distillation

Ahmet Arda Dalyanci, Computer Science

Mentors: Prof. Huan Liu – Weidong Zhang, Doctoral Student

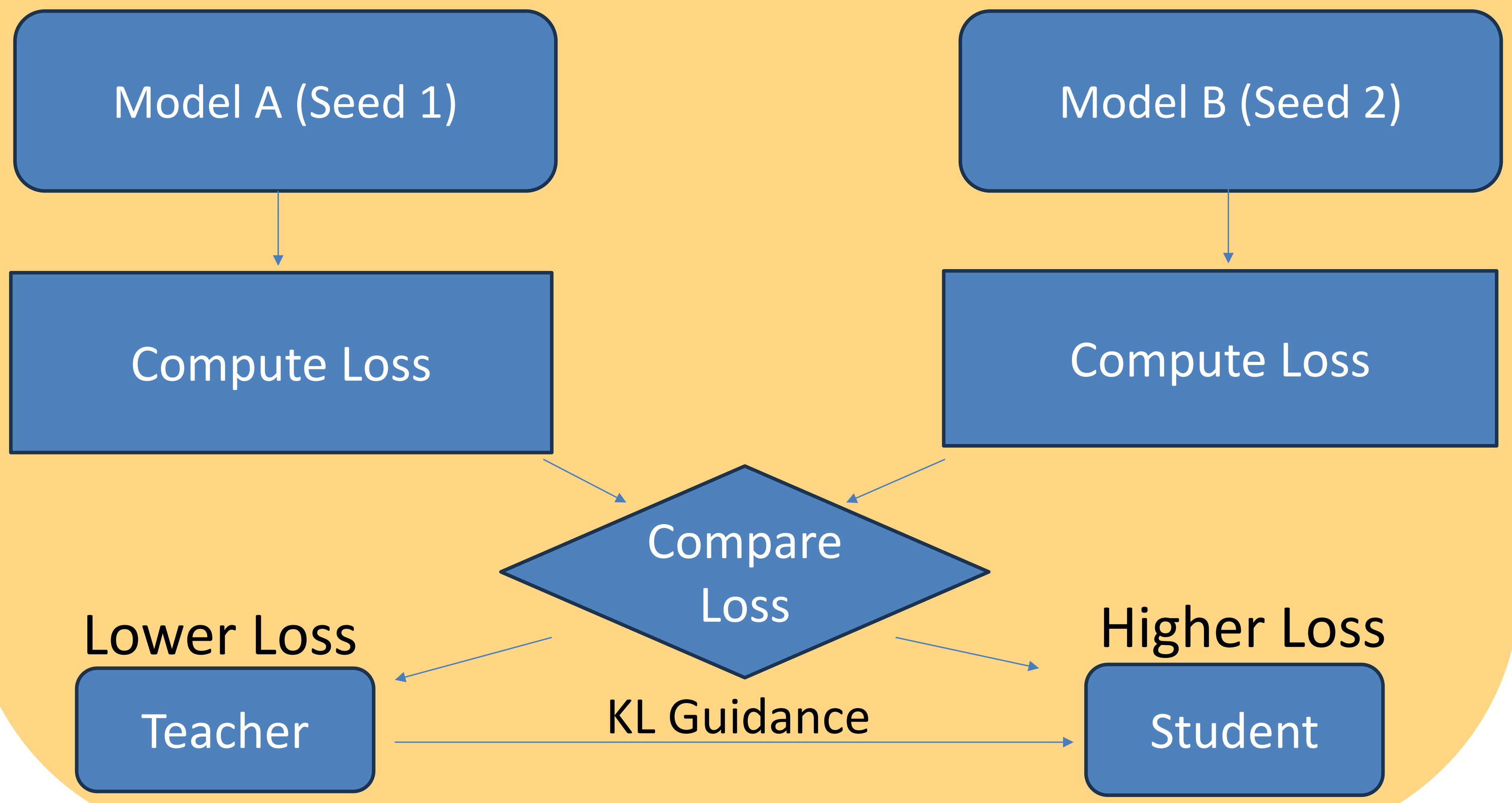
School of Computing and Augmented Intelligence (SCAI)

Motivation & Problem

- **Mobile Constraints:** Edge AI models must operate under strict parameter and memory limits (~2.5M params).
- **The Bottleneck:** Standard training fails to fully realize the model's effective capacity.
- **Key Question:** Can we improve performance without increasing model size or inference cost?

Our Approach: Self-Competitive Distillation (SCD)

- Two identical models train in parallel from different seeds.
- At each step, the lower-loss branch becomes the **Teacher**.
- The other branch (Student) learns using cross-entropy + KL-divergence.
- **Zero Inference Overhead:** The weaker branch is discarded after training.



Experimental Setup

- **Models:** EfficientNet, ConvMixer, MobileNet, ConvNeXt.
- **Datasets:** CIFAR-10, Stanford Dogs, HAM10000, MIT Indoor-67.
- **Training:** Baseline vs. SCD (100 & 200 epochs).

Key Findings

- SCD consistently improves test accuracy and **xScore** across all datasets.
- SCD leverages divergence between training trajectories to improve performance.
- SCD provides larger gains for higher-capacity models, while improving all models.

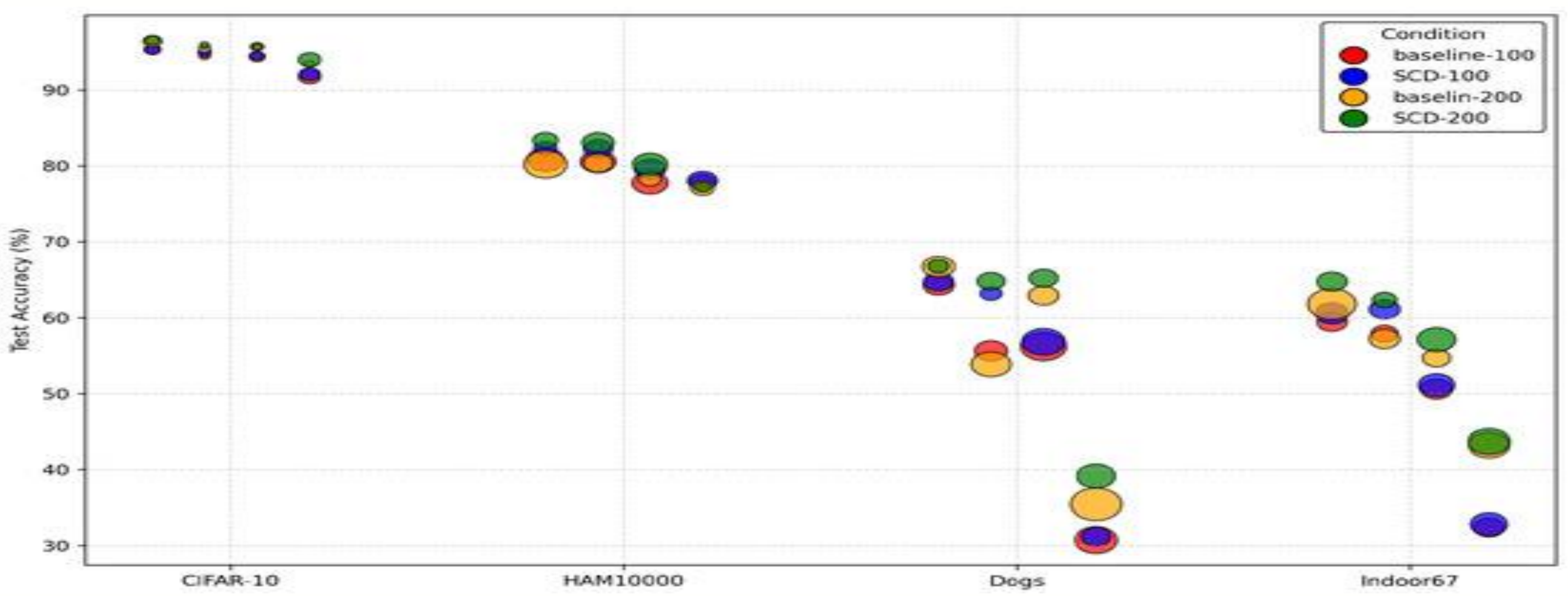


Figure 1: Test accuracy comparison across four datasets at 100 and 200 epochs.

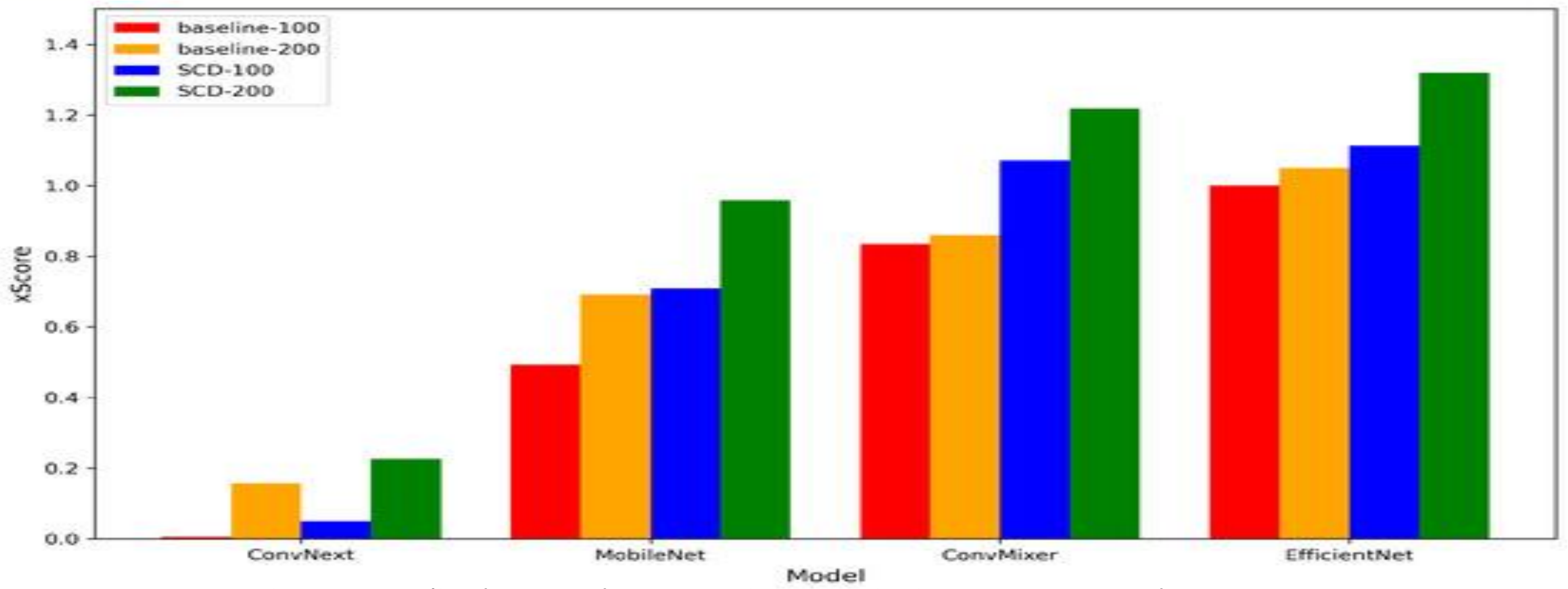


Figure 2: Realized xScore demonstrating capacity improvements under SCD.

SCD vs. DML

- **DML:** Symmetric and bidirectional. Promotes consensus. Provides stable support for lower-capacity models.
- **SCD:** Asymmetric and performance-driven. Encourages competition. Benefits higher-capacity models.
- **Key Insight:** SCD provides larger gains for stronger models.

Conclusion

- Training dynamics are as important as architecture.
- SCD improves effective capacity with zero additional inference cost.

Key Takeaways

- No increase in model size.
- Zero additional inference overhead.
- Improved effective capacity under fixed budgets.
- Stronger gains in higher-capacity models.

Future Work

- Study seed diversity.
- Extend to detection and segmentation.

Acknowledgements: Supported by FURI (Spring 2026). Special thanks to Weidong Zhang and Prof. Huan Liu.

FURI

Ira A. Fulton Schools of Engineering
Arizona State University