ARIA
APPLIED RESEARCH IN ACTION

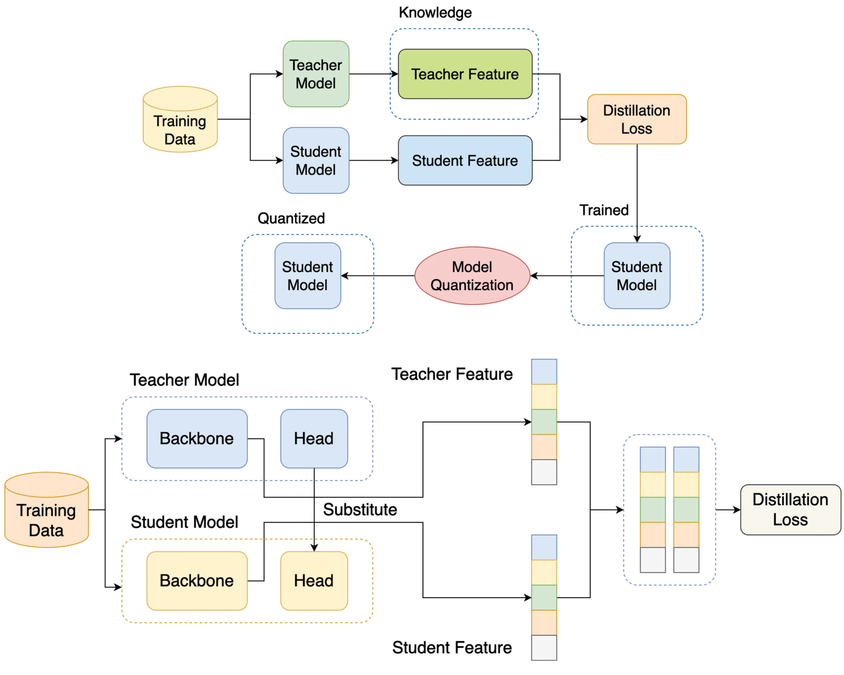
Model Compression for Unmanned Aerial Vehicle's Companion Computer

Accuracy & Latency Optimization and Tradeoff for Segmentation Models

Yuxuan Zhang

Nandita Vijaykumar
ACADEMIC SUPERVISOR

Sunil Jacob
INDUSTRY SUPERVISOR



PROJECT SUMMARY

Constrained by limited computational resources, lightweight models are widely used in edge devices, like unmanned aerial vehicle's companion computers. However, these lightweight models are often not as powerful as large models, thus leading to insufficient performance. This performance gap can be mitigated through knowledge distillation, a technique that transfers knowledge from a teacher model (typically a heavyweight model) to a smaller student model. However, most proposed knowledge distillation approaches concentrate on image classification, which can not be directly transmitted to dense prediction tasks like semantic segmentation. To improve lightweight models' performance for dense prediction tasks, this paper introduces EfficientKD, an innovative and efficient knowledge distillation approach tailored for semantic segmentation tasks. EfficientKD consists of two pivotal components: feature alignment knowledge distillation, which streamlines the knowledge transfer process from the teacher to the student model, and reused decode head, which further enhances the performance of semantic segmentation. Comprehensive experiments on the ADE20K dataset substantiate both the effectiveness and efficiency of our proposed method.

SOTI

Computer Science
UNIVERSITY OF TORONTO
Master of Science in Applied Computing