# Knowledge Distillation for high-end compute intensive DL process to run on a low-end system

Computer Vision (CS6350)

TPA-14

1. Knowledge Distillation

Knowledge distillation is the technique of transferring knowledge from a large model or a set of models (teacher network) to a single smaller model (student). In neural networks, knowledge can be any of the following: 1) learned weights and bias, 2) logits as a source of teacher's knowledge, or 3) different types of activations and neurons or parameters of the teacher network.
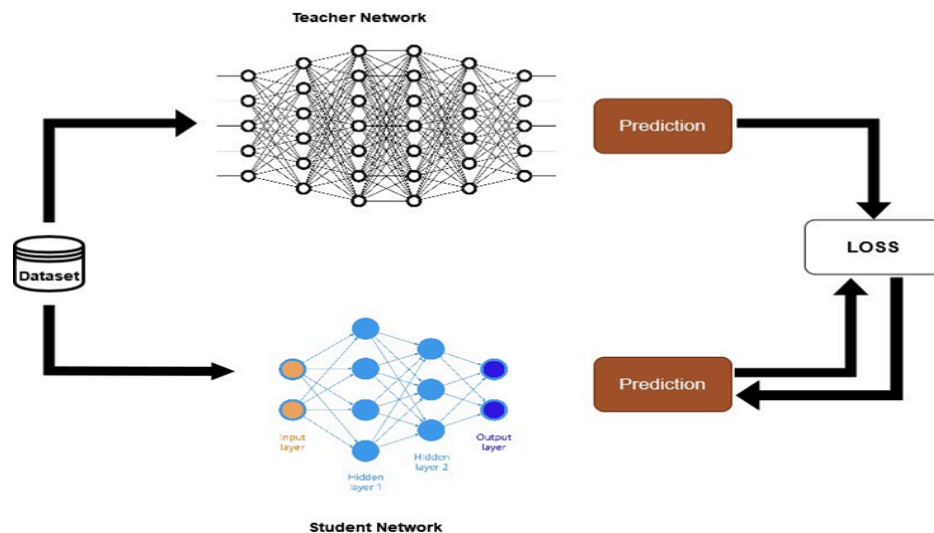
2. Problem Statement

In this project, you are supposed to select one (preferably) Computer Vision task(s): classification, object detection, segmentation, or depth estimation, and find the state-of-the-art performing trained model (Example: YOLO for object detection or HRNet for semantic segmentation, ResNet101 for classification).

Then you need to design and model a student network to perform the same task with comparable accuracy as the teacher network. Students network will be a smaller network with fewer parameters, such as MobileNet, VGG 16, AlexNet, or any custom network.

Input and Output for the Teacher Network
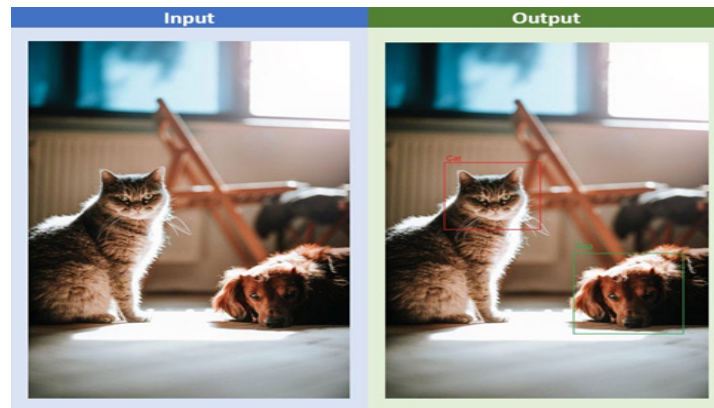
For Object Detection

Image classification

Segmentation

Input and Output for the Student Network

The student model should be able to perform the same task as the teacher network with comparable accuracy.

Dataset

It is recommended to use the same or a subset of the dataset used to train the teacher network.

Reference

- Gou, J., Yu, B., Maybank, S.J. *et al.* Knowledge Distillation: A Survey. *Int J Comput Vis* **129**, 1789–1819 (2021). https://doi.org/10.1007/s11263-021-01453-z
- Li, Zhihui, Pengfei Xu, Xiaojun Chang, Luyao Yang, Yuanyuan Zhang, L. Yao and Xiaojiang Chen. "When Object Detection Meets Knowledge Distillation: A Survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45 (2023): 10555-10579.
- Base Paper: Z. Zheng *et al.*, "Localization Distillation for Object Detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 10070-10083, Aug. 2023, doi: 10.1109/TPAMI.2023.3248583.
- Q. Li, S. Jin and J. Yan, "Mimicking Very Efficient Network for Object Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 7341-7349, doi: 10.1109/CVPR.2017.776.
- Guobin Chen, Wongun Choi, Xiang Yu, Tony Han, and Manmohan Chandraker. 2017. Learning efficient object detection models with knowledge distillation. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 742–751.
- Sample Implementation:

- - https://www.analyticsvidhya.com/blog/2022/01/knowledge-distillation-theory-and-end-to-end-case-study/
  - https://towardsdatascience.com/data-distillation-for-object-detection-92a89fe5d996
- ***Saurabh Kumar Jain, Sukhendu Das***; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2023, pp. 4579-4588
- For more papers on Knowledge Distillation refer: https://github.com/LutingWang/awesome-knowledge-distillation-for-object-detection