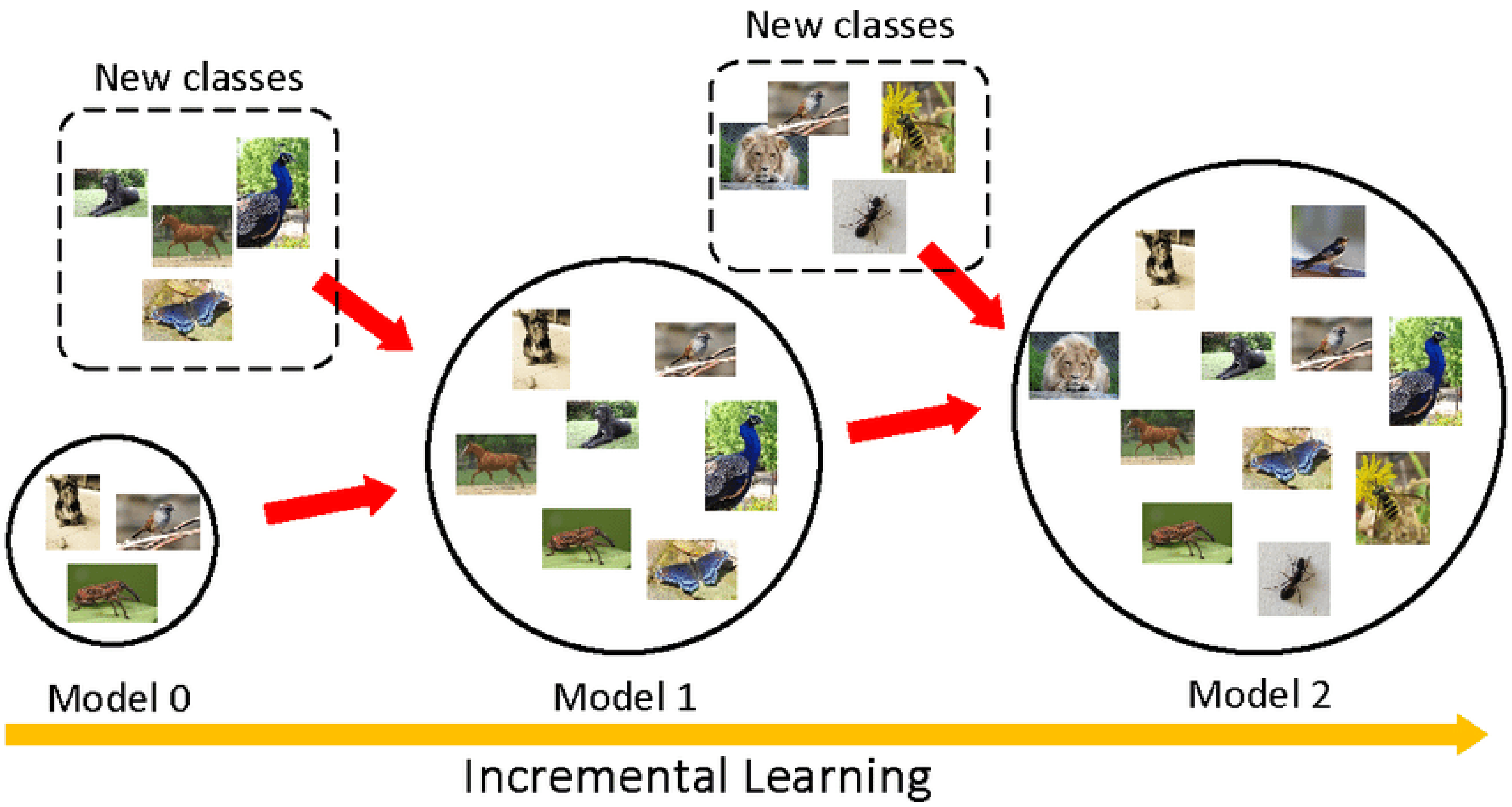
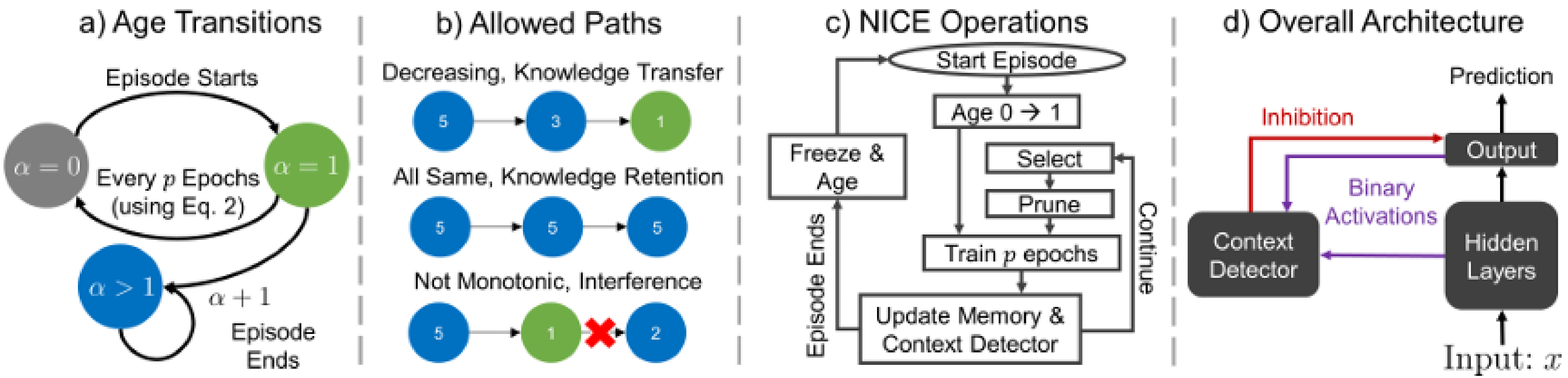
NICE: Neurogenesis Inspired Contextual Encoding for Replay-free Class Incremental Learning

A20586593 Seonghwan Lim

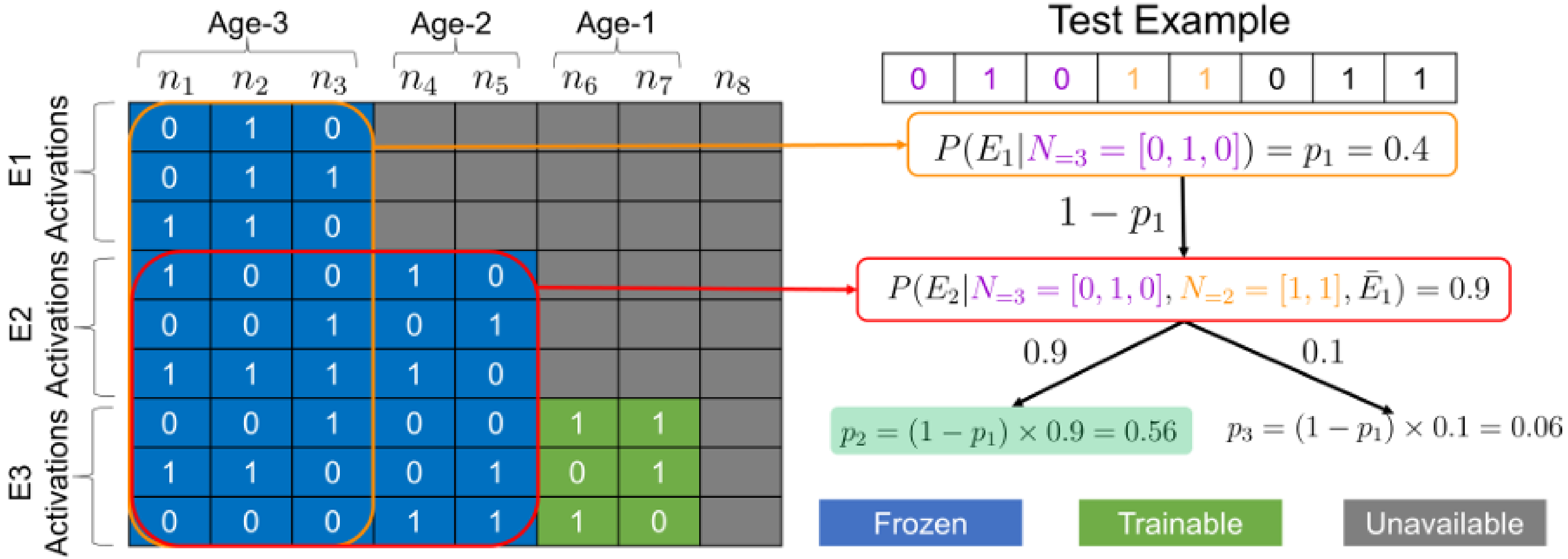
Class Incremental Learning



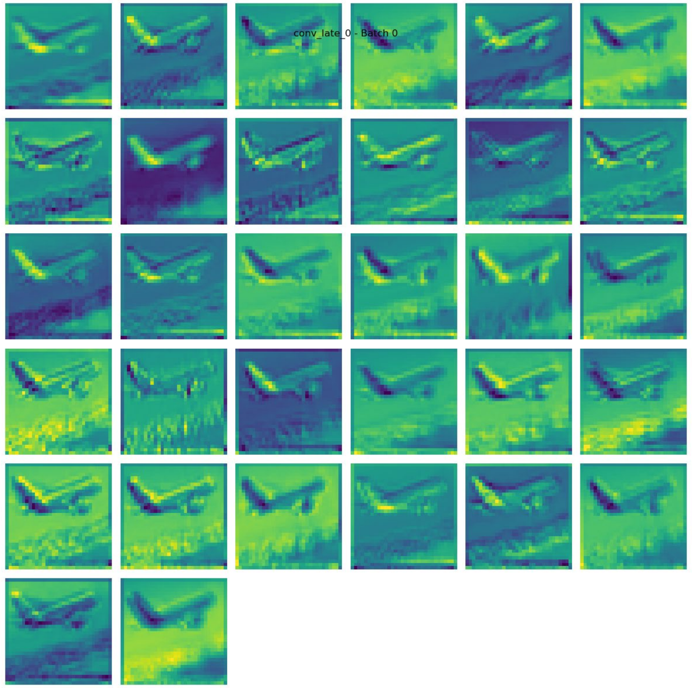
NICE: Neurogenesis Inspired Contextual Encoding for Replay-free Class Incremental Learning

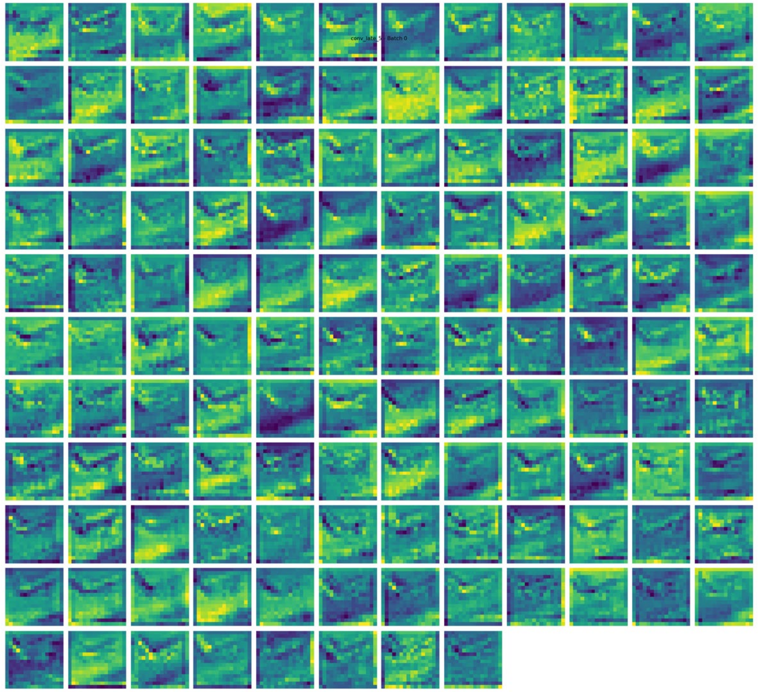


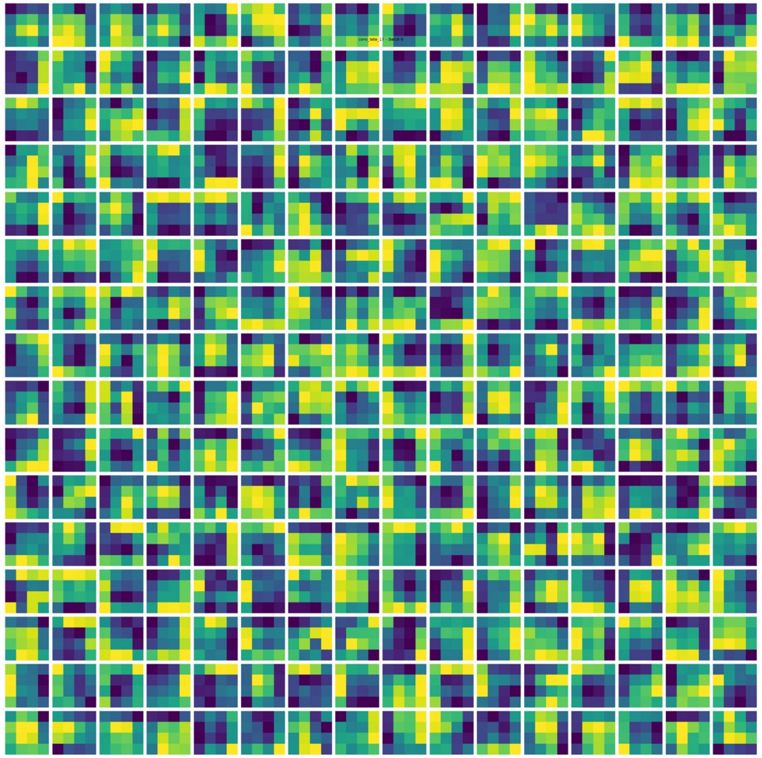
NICE: Neurogenesis Inspired Contextual Encoding for Replay-free Class Incremental Learning



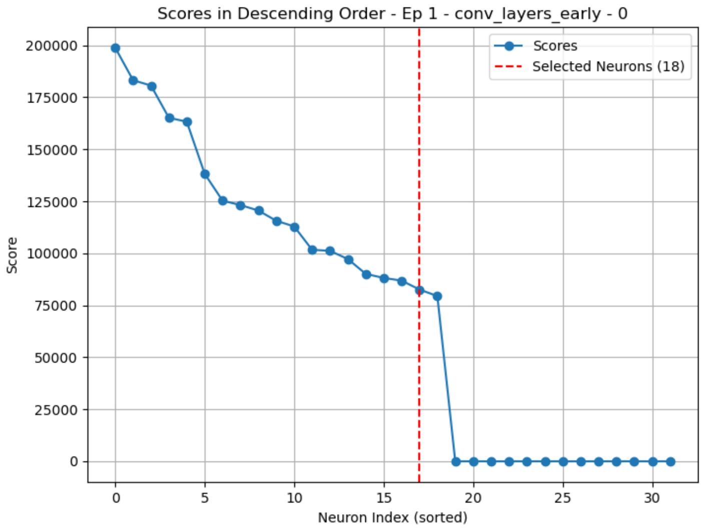
NICE: Neurogenesis Inspired Contextual Encoding for Replay-free Class Incremental Learning

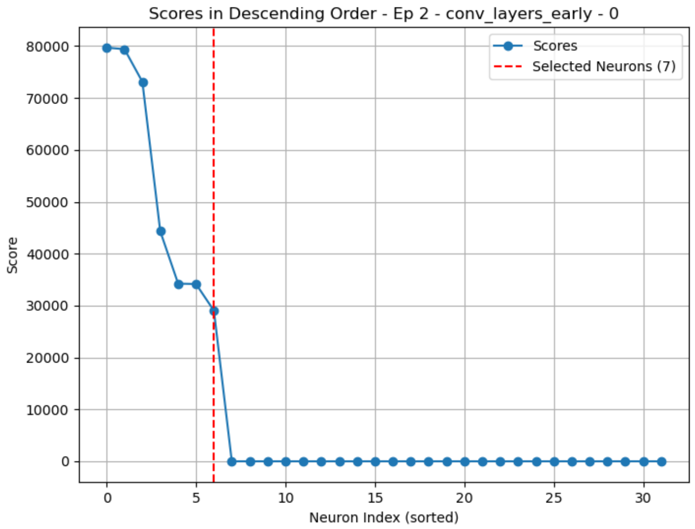


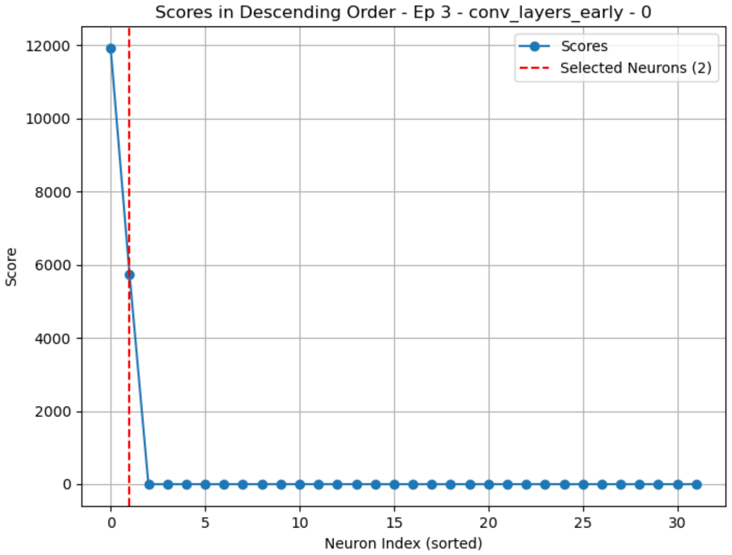




NICE: Neurogenesis Inspired Contextual Encoding for Replay-free Class Incremental Learning



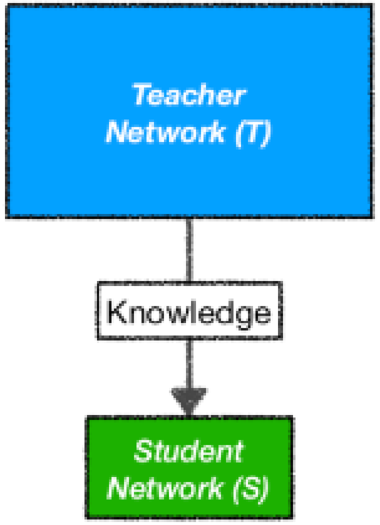




Knowledge Distillation

Introduction

Knowledge Distillation



1. **Teacher Network (T)**
   - cumbersome model
   ex) ensemble / a large generalized model
   - (pros) excellent performance
   - (cons) computationally expansive
   - can not be deployed when limited environments

2. **Student Network (S)**
   - small model
   - suitable for deployment
   - (pros) fast inference
   - (cons) lower performance than T

Knowledge Distillation

Key Loss Formula for Knowledge Distillation

$$L_{\text{total}} = \alpha \cdot L_{\text{CE}} + (1 - \alpha) \cdot L_{\text{KD}}$$

Where:

- L_{CE}: Cross-Entropy Loss (difference between student model outputs and ground truth labels).
- L_{KD}: Knowledge Distillation Loss (difference between teacher and student output distributions).
- α: Balancing weight for L_{CE} and L_{KD}.
- T: Temperature for softening probability distributions.

Knowledge Distillation

Knowledge Distillation Loss (KD Loss)

$$L_{\text{KD}} = T^2 \cdot \text{KL}\big(\text{Softmax}(\frac{z_t}{T}) \,||\, \text{Softmax}(\frac{z_s}{T})\big)$$

Where:

- z_t: Teacher model logits.
- z_s: Student model logits.
- **KL Divergence:** Measures the difference between two probability distributions.

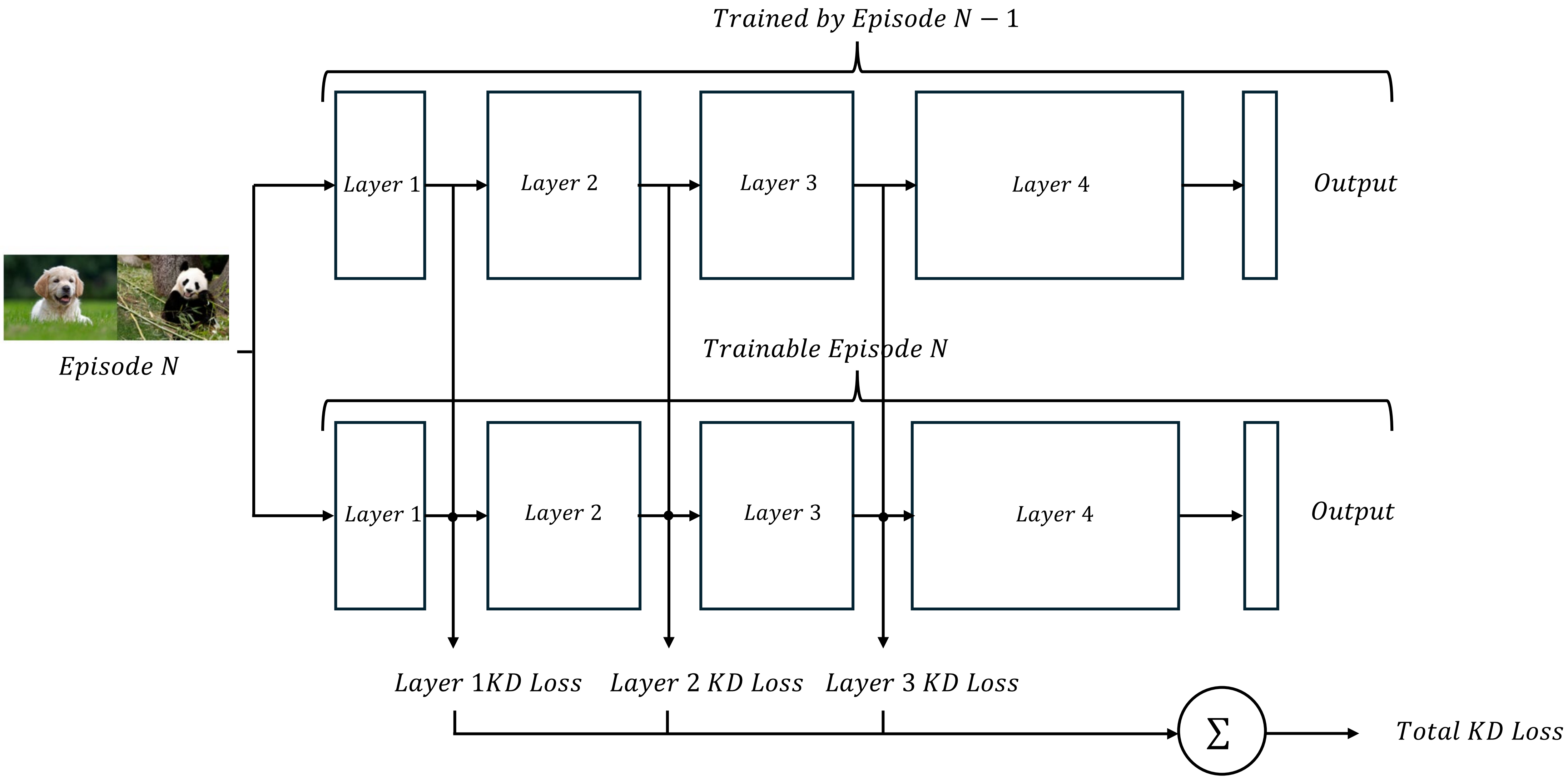
KL Divergence Formula

$$\text{KL}(P \,||\, Q) = \sum_i P(i) \cdot \log\left(\frac{P(i)}{Q(i)}\right)$$

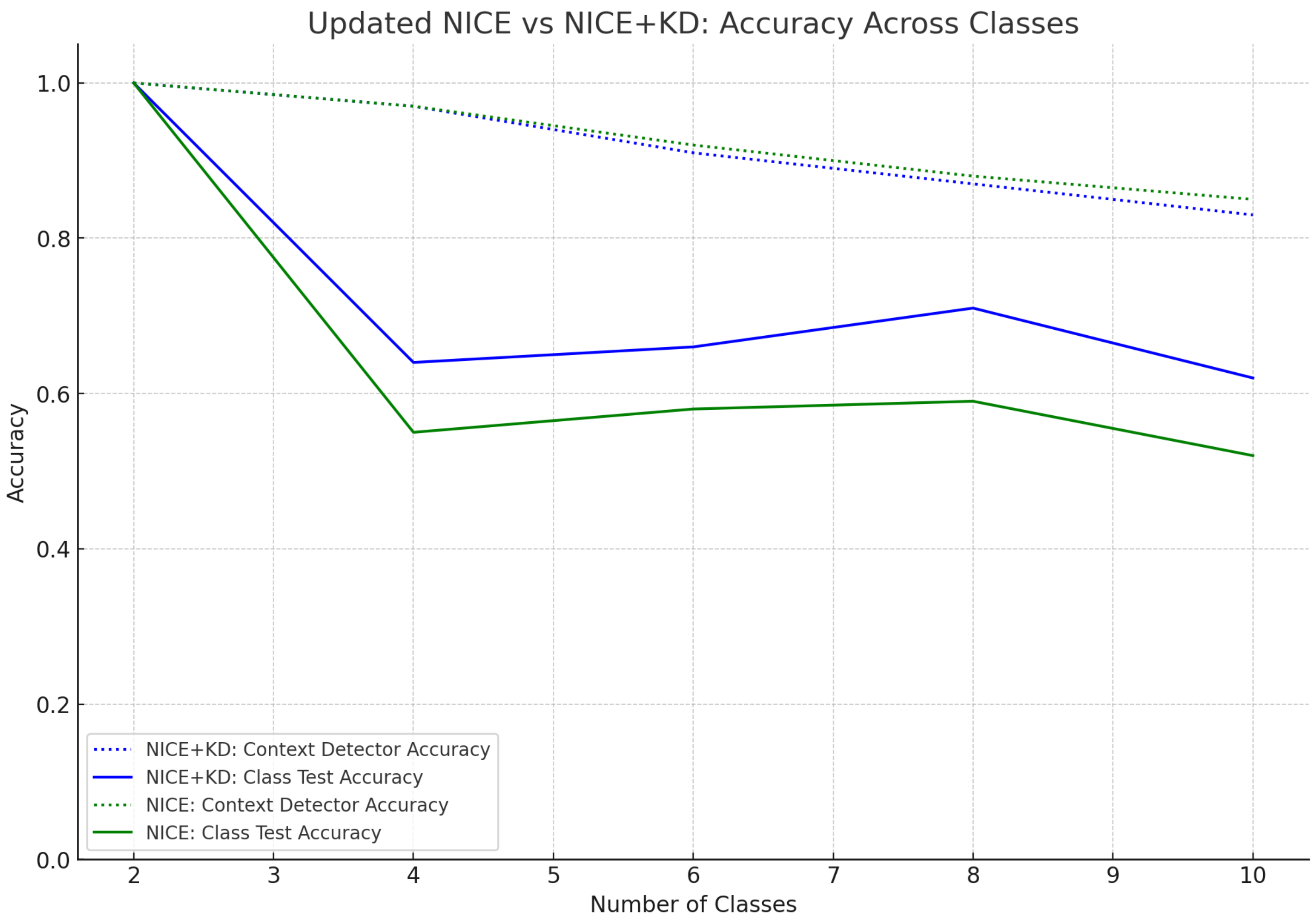
Where:

- $P(i)$: The true probability distribution (e.g., teacher's softened output).
- $Q(i)$: The approximate probability distribution (e.g., student's softened output).

Knowledge Distillation

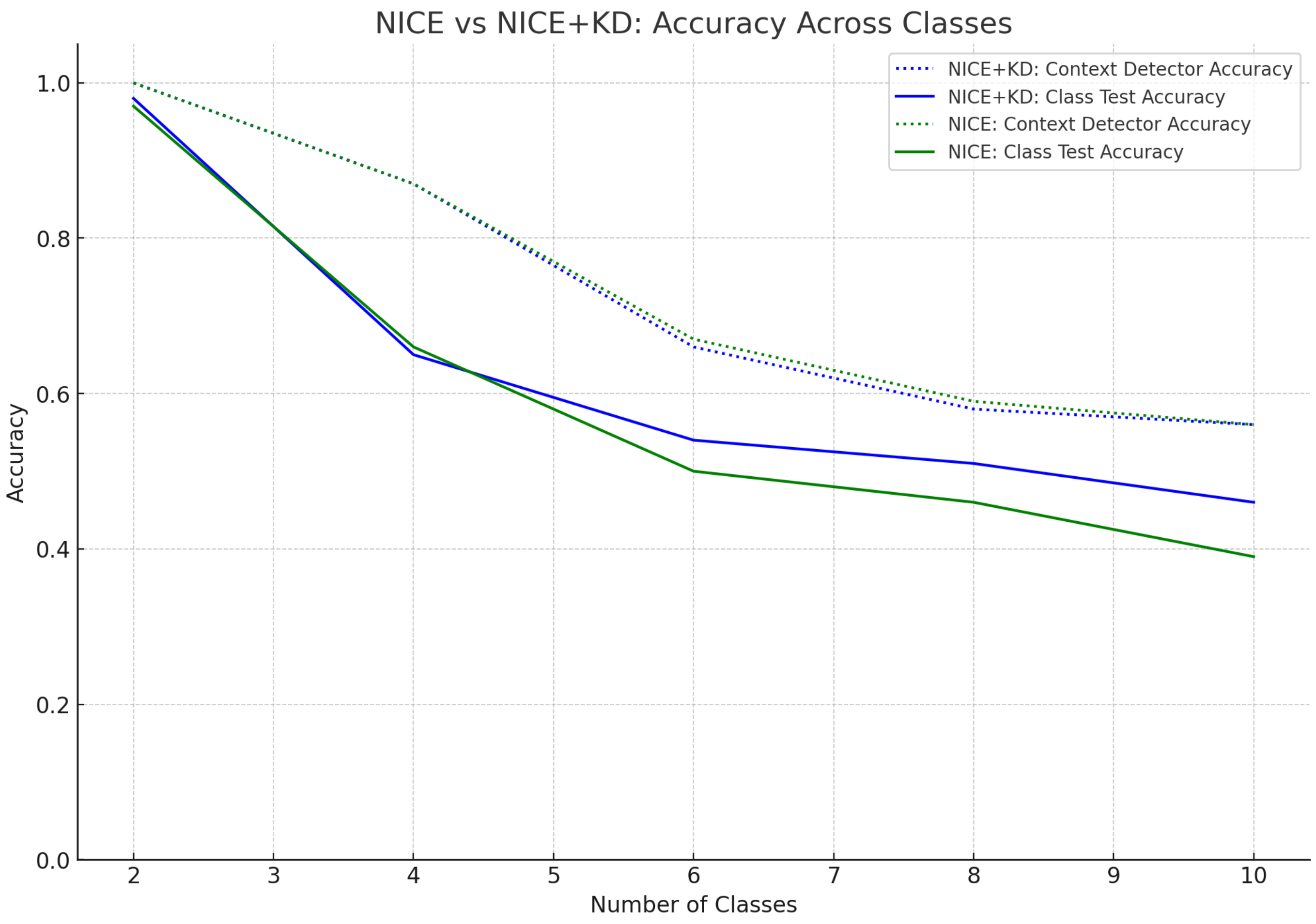


Knowledge Distillation



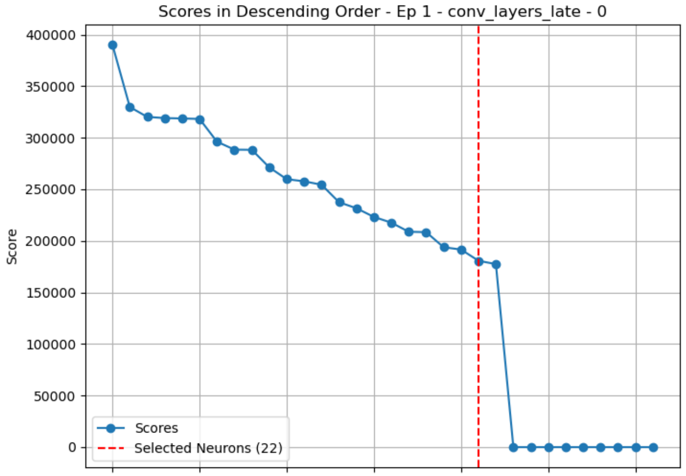
Model: CNN_MNIST
Data: MNIST

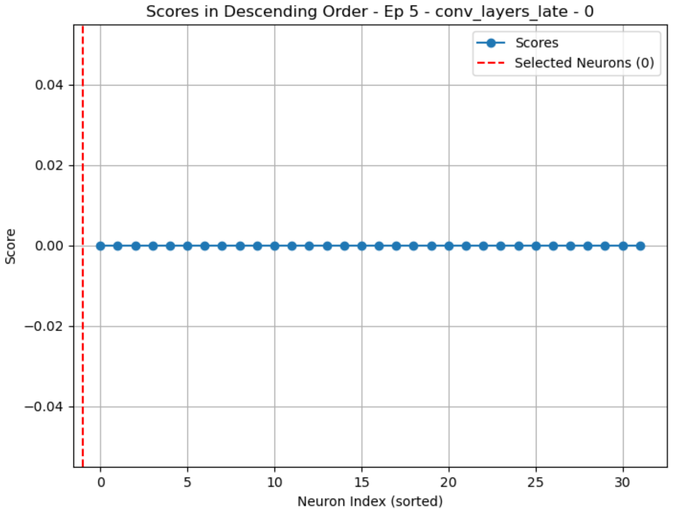
Knowledge Distillation

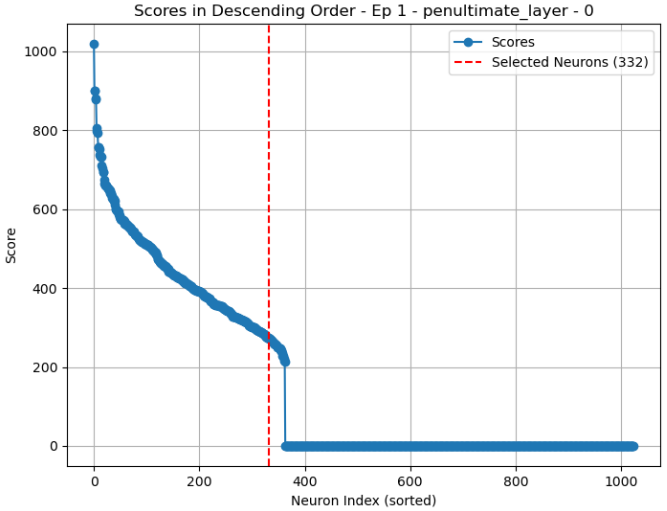


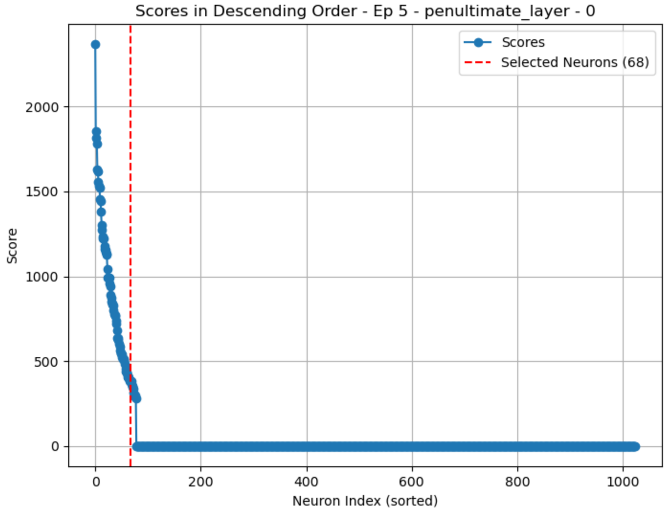
Model: VGG11 SLIM
Data: CIFAR10

Future Works

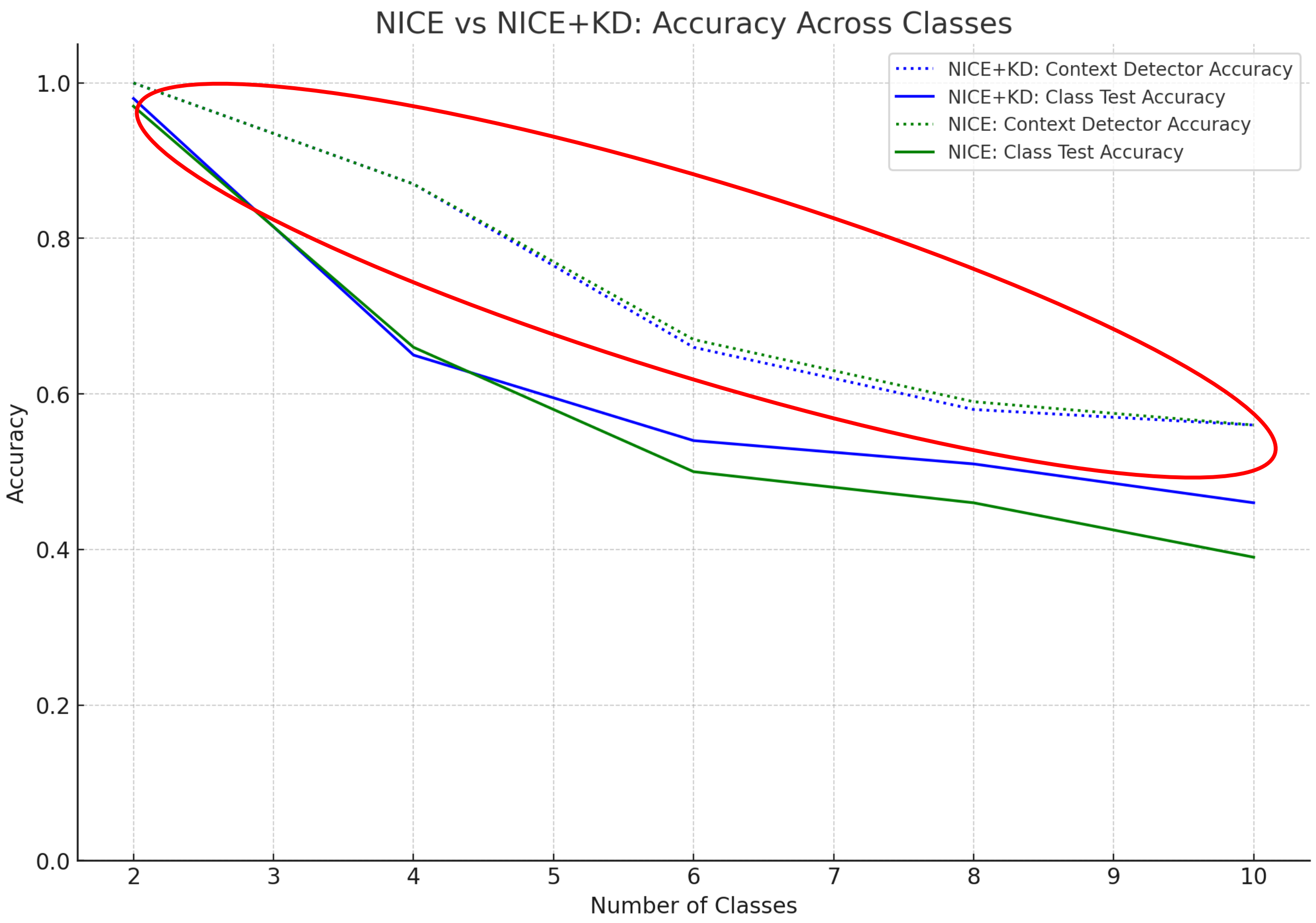








Future Works



Question

Thank You