

Ansh Mujral
amujral@ucsd.edu

Investigating CoTTA: Validating Real-Time Neural Network Adaptations

Ifunanya Okoroma
iokoroma@ucsd.edu

Keenan Serrao
kserrao@ucsd.edu

Nicholas Swetlin
nswetlin@ucsd.edu

Mentor: Jun-Kun Wang
jkw005@ucsd.edu

Motivation

Machine learning (ML) systems can suffer from **test-data distribution shifts**.

Established “solution”: **Test-time adaptation with pseudo-labeling (TTAPL)**.

We need **better variations of TTAPL** that: produce **more accurate models, generalize to different models**.

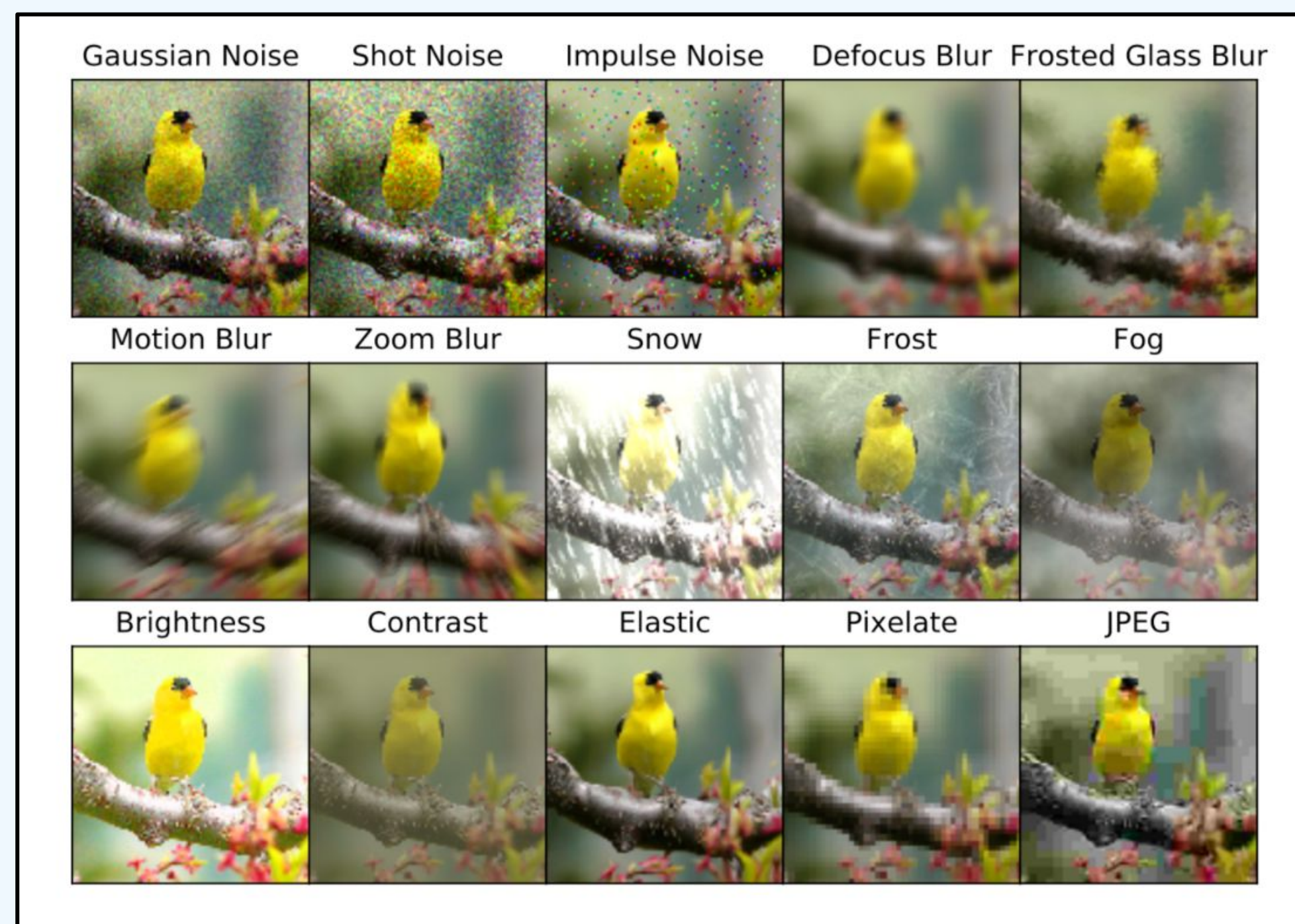


Figure 1: Types of Distribution Shifts

Background

Goal of TTAPL: adapt model weights θ to better suit a distribution-shifted dataset of unlabeled test images, X :

1. Generate pseudo-labels (\hat{y}) for each test image in X .
2. Adapt model weights θ from pseudo-labels (\hat{y}).

$$\begin{aligned} \text{for each image } x \text{ in } X: \\ \hat{y} &= f_{\theta}(x) \\ \theta &\leftarrow \theta - L(x, \hat{y}) \end{aligned}$$

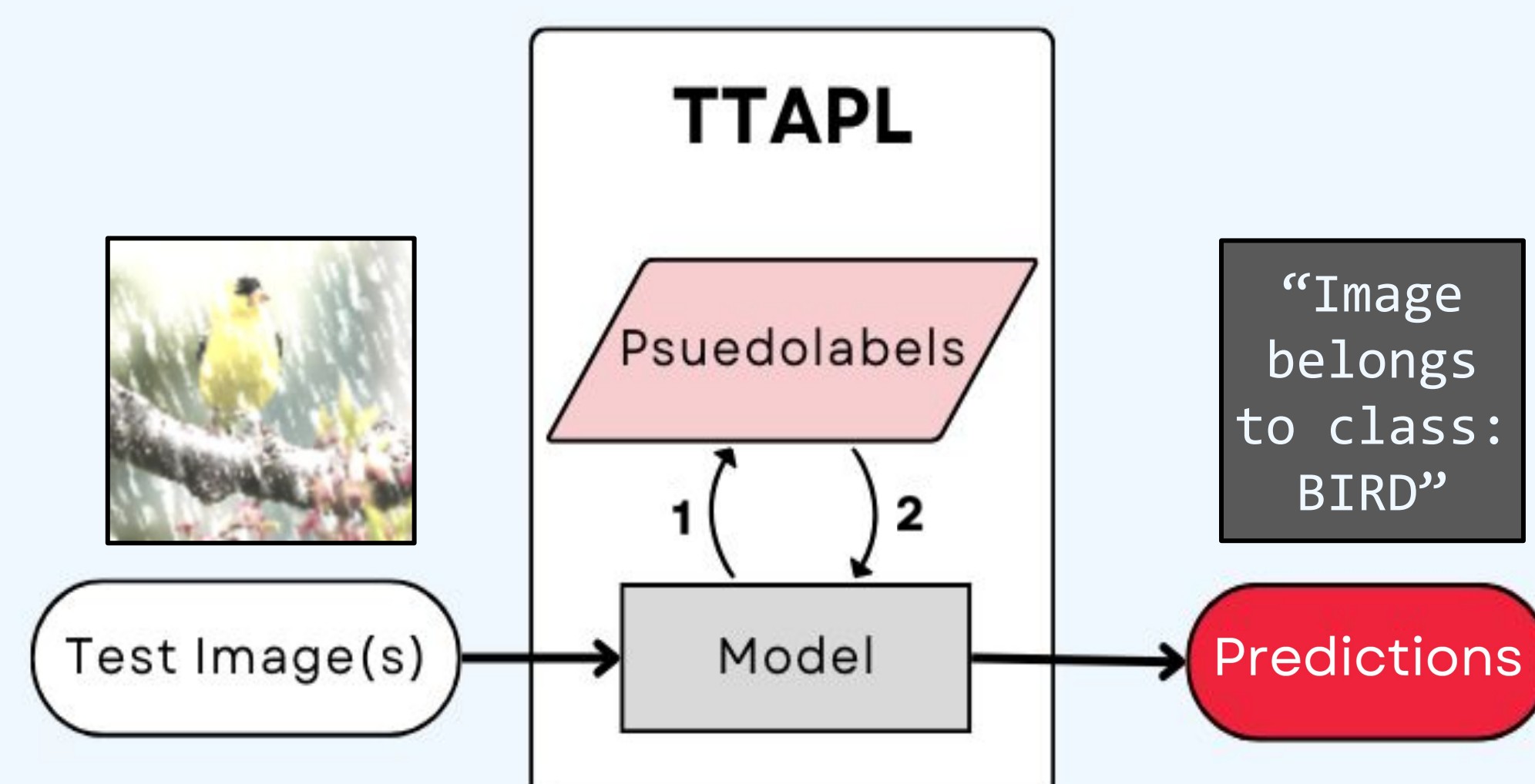


Figure 2: TTAPL General Algorithm

Methods

TTAPL Choice:

CoTTA – one type of **continual TTAPL** that shows significant promise in existing literature.

Consistency Loss – Part of CoTTA that determines how much a pseudo label should affect model weights.

$$\mathcal{L}_{\theta_t}(x_t^T) = - \sum_c y_{tc}^T \log \hat{y}_{tc}^T$$

Figure 4: Example of a Consistency Loss (Cross-Entropy Loss)

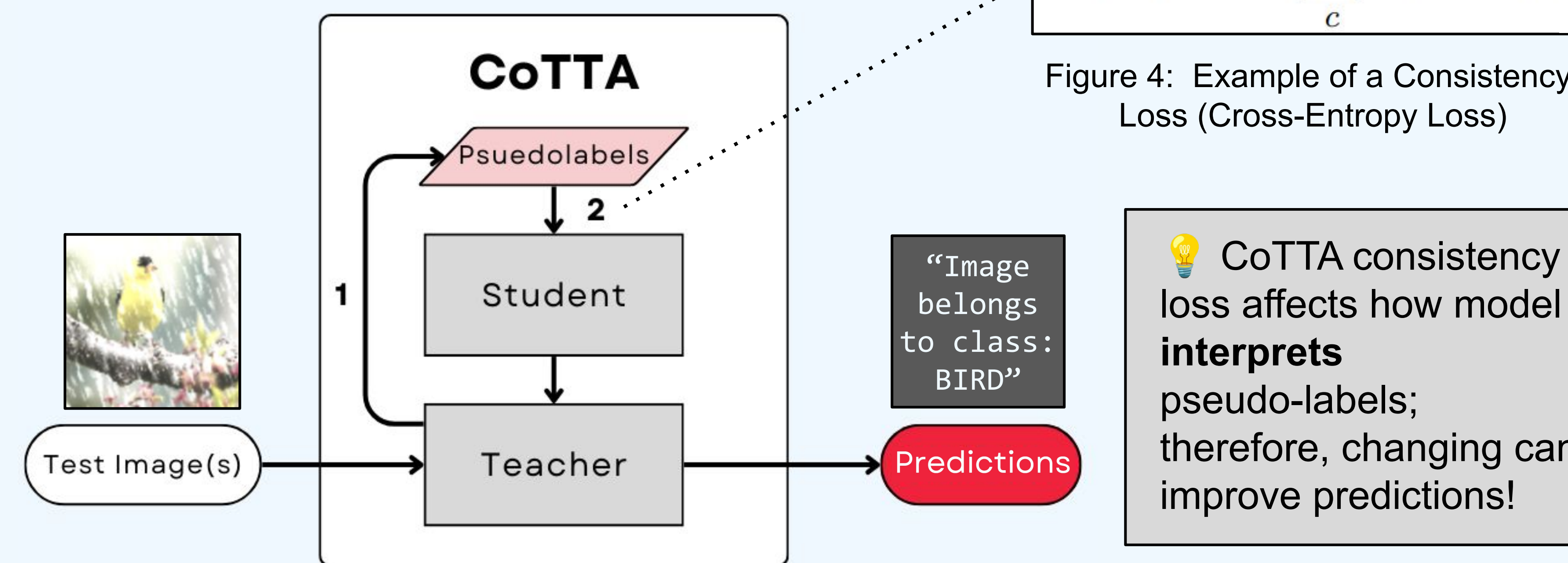


Figure 3: CoTTA Algorithm

Procedure:

1. **Test CoTTA on seven common models** – each pretrained and with a different model structure.
2. For **each model**:
 - Adapt four times – each time with different consistency loss.
3. For **each combination of model, consistency loss**:
 - Perform **CIFAR10-to-CIFAR10C**, an established computer-vision experiment for TTAPL.
 - Evaluate model quality with accuracy, precision, recall, F1 score.

For each model:

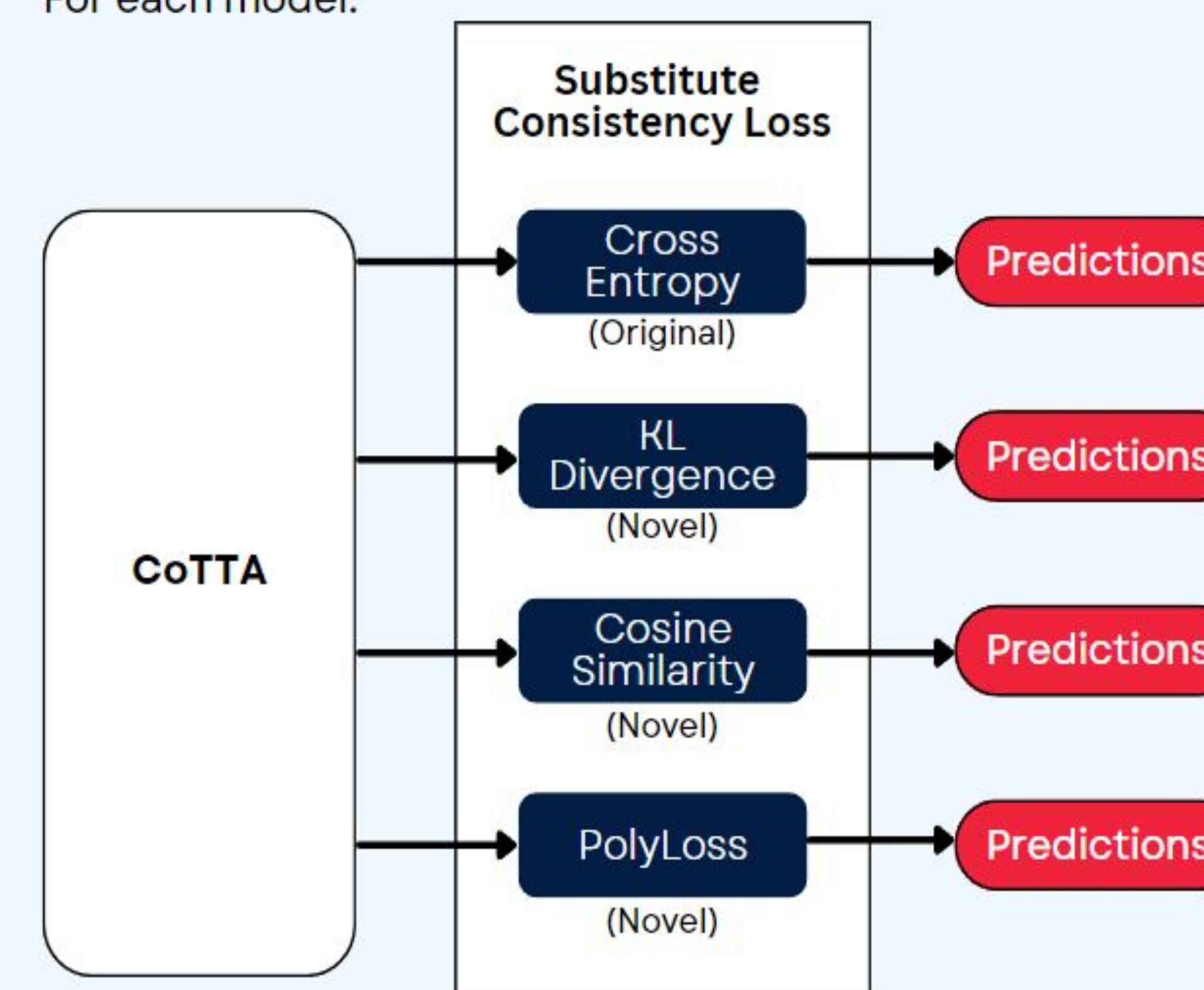


Figure 5: Experiment flowchart for a single model.

Results

CoTTA Consistency Loss	Model Type						
		I	II	III	IV	V	VI
	CE (Original)	8.4	11.7	5.2	26.5	6.4	-2.5
	Poly (Ours)	9.0	11.2	5.2	25.7	6.6	-7.4
	KL (Ours)	8.3	11.7	5.2	26.6	6.4	-2.6
	Cosine (Ours)	7.4	11.2	4.2	25.8	5.6	-5.6

Table 1: Accuracy Increase (%) per Model-Loss Combination. The highest value(s) per model (column) are shaded red.

Better Predictions: higher increases in accuracy, precision, recall, F1 score for our variations on select models compared to original CoTTA.

Model Type Matters: size and architecture of model still highly influence quality of CoTTA.

In Practice: when using CoTTA to adapt your model, consider the model being adapted; incorporating different consistency loss likely leads to better prediction quality.

Future work:

- Determine CoTTA efficacy on larger ML models (larger than 40 neural network layers).
- Test performance on higher resolution data.

Contributions

AM: Implemented two models, analyzed result metrics, contributed to website.
IO: Implemented two models, contributed to poster and paper.
KS: Restructured original CoTTA repository to run on different architectures, contributed to website.
NS: Implemented two models, contributed to poster and paper.

References

Wang, Gool Dai, Fink. 2022. “Continual Test-Time Adaptation.” *Computer Vision and Pattern Recognition (CVPR) 2022*.