ICONIP 2024

31st International Conference on Neural Information Processing

December 2–6, 2024 · Auckland, New Zealand iconip2024.org

Transferring Teacher's Invariance to Student Through Data Augmentation Optimization

Tamotsu Kurioka (Science Tokyo)

Teppei Suzuki (Denso IT Laboratory)

Rei Kawakami (Science Tokyo)

Ikuro Sato (Science Tokyo/Denso IT Laboratory)

ICONIP 2024
31st International Conference on Neural Information Processing
December 2–6, 2024 · Auckland, New Zealand iconip2024.org

Sponsors: g.tec NEW ZEALAND 100% PURE NEW ZEALAND AUT KNOWLEDGE ENGINEERING & DISCOVERY RESEARCH INNOVATION Springer

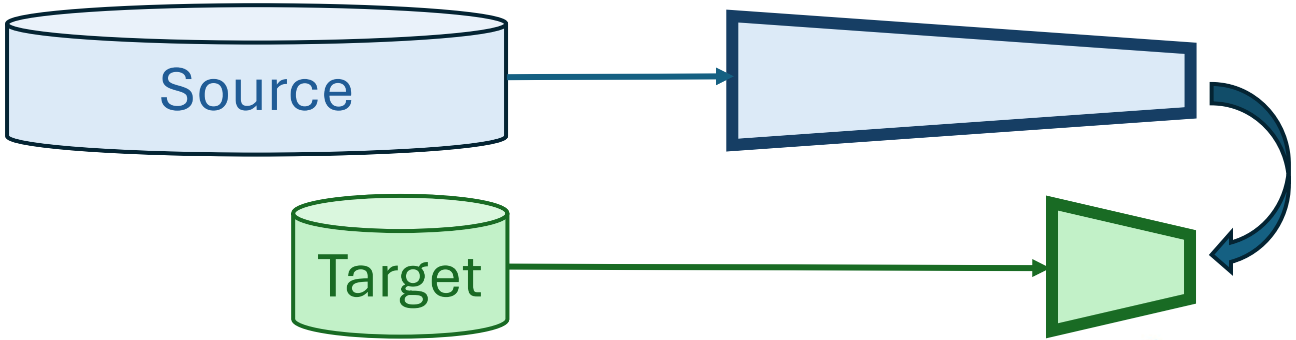
Supported by: AUT UNIVERSITY UNIVERSITY OF AUCKLAND Waipapa Taumata Rau NEW ZEALAND NTU Nottingham Trent University Asia Pacific Neural Network Society

Transfer Learning

- **Transfer learning** is a machine learning technique where a model developed for a particular task is reused as the starting point for a model on a second task.

Advantage

- Enables leveraging large datasets (e.g., ImageNet) to improve performance on smaller datasets.



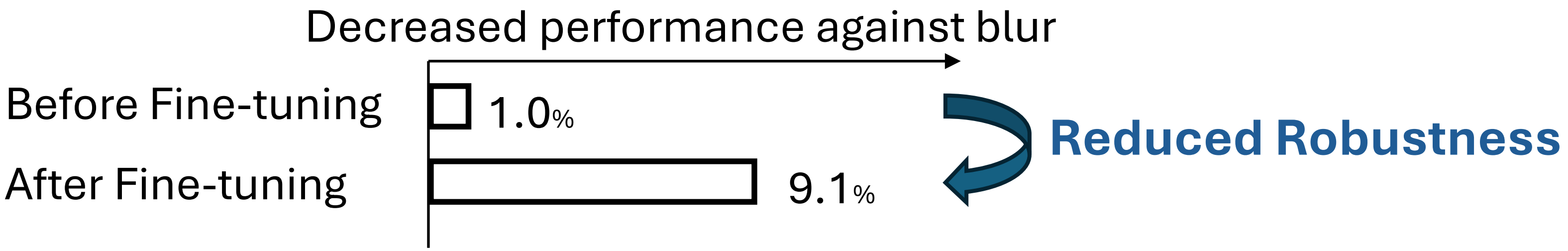
Teach something to the target model as a teacher model

The Challenge of Robustness in Transfer Learning

- **Robustness** refers to a model's ability to maintain performance despite variations in input data, such as noise or distortions.

Issue

- Traditional transfer learning methods often fail to preserve the robustness of pre-trained models when adapting to new tasks.[Yamada et al., 2022]



Decreased performance against blur

Before Fine-tuning 1.0%

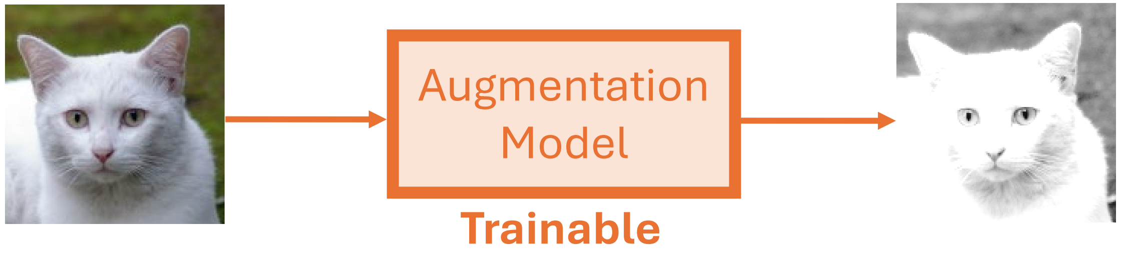
After Fine-tuning 9.1%

Reduced Robustness

Y. Yamada and M. Otani. Does Robustness on ImageNet Transfer to Downstream Tasks?. In *Conference on Computer Vision and Pattern Recognition (CVPR).* pp. 9205-9214, 2022

Gaining Robustness through Data Augmentation

- Data augmentation enhances the robustness of models by artificially expanding training datasets through transformations like rotations, flips, and color adjustments.
  - In recent years, methods have emerged to perform **data augmentation with parameters that can be optimized** using the error back propagation method.



Issue

- Conventional data augmentation is not specifically designed to maintain the robustness acquired by teacher models when transferring knowledge to the target model.

Introducing TransInv

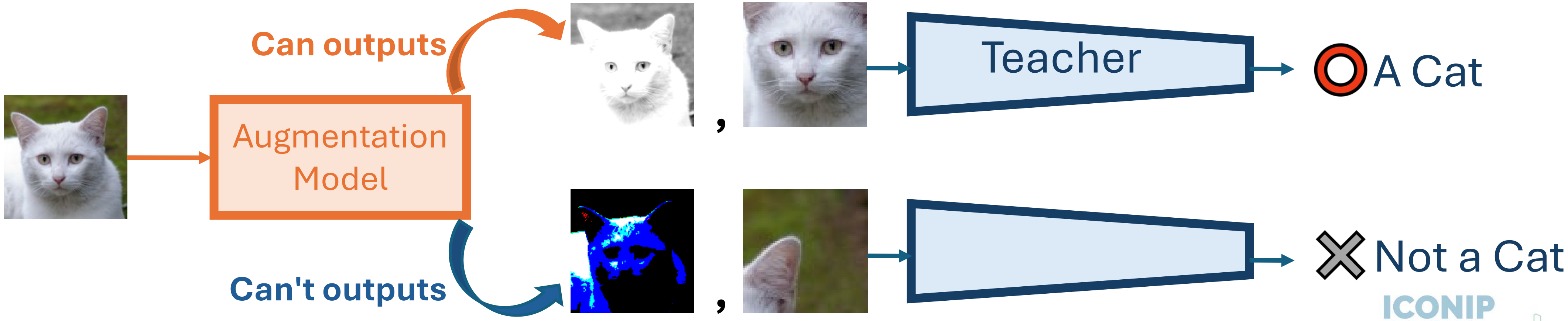
- **TransInv** is a framework that jointly optimizes two models: a data augmentation model and a target model.
- It leverages the knowledge of a teacher model to enhance the robustness of the target model against intra-class variations.

Research Objectives

- To demonstrate that TransInv can maintain robustness during the fine-tuning process.

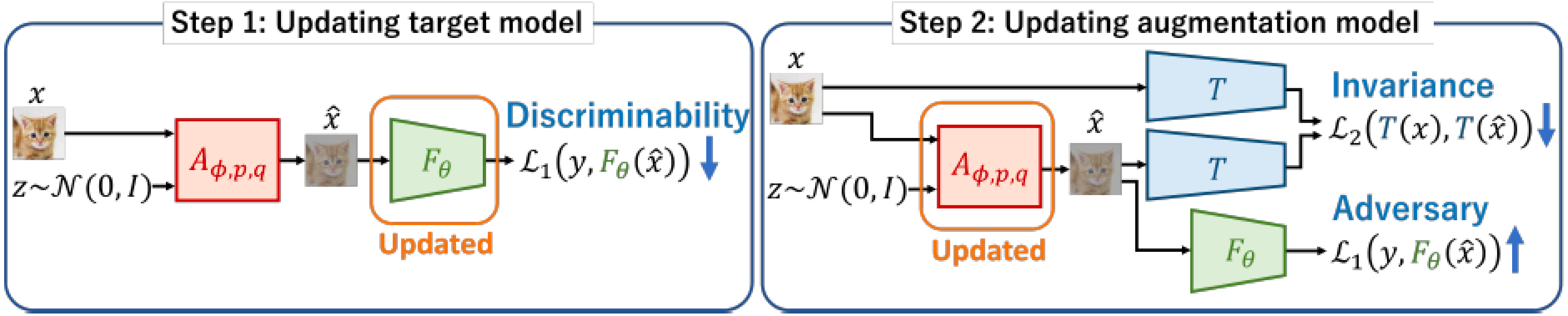
Methodology Overview

- TransInv employs a min-max optimization framework that simultaneously trains the data augmentation model and the target model.
- The data augmentation model generates adversarial samples that maintain the invariance of the teacher model.



Joint Optimization

- The min-max objective allows the data augmentation model to adaptively learn which transformations are most effective based on the teacher model's robustness.
- This process ensures that the augmented data falls within a range where the teacher model exhibits high invariance.



Step 1: Updating target model
x
$z\sim\mathcal{N}(0,I)$
$A_{\phi,p,q}$
$\hat{x}$
F_θ
Updated
Discriminability
$\mathcal{L}_1(y, F_\theta(\hat{x}))$ ↓
Step 2: Updating augmentation model
x
$z\sim\mathcal{N}(0,I)$
$A_{\phi,p,q}$
Updated
$\hat{x}$
T
T
F_θ
Invariance
$\mathcal{L}_2(T(x), T(\hat{x}))$ ↓
Adversary
$\mathcal{L}_1(y, F_\theta(\hat{x}))$ ↑

Experimental Setup

Objective

- To evaluate the performance of TransInv in enhancing the robustness of target models using a teacher model.

Model Architecture

- We utilized WideResNet-40-2 for both teacher and target models.

Experimental Setup

Teacher model

- 15 different teacher models with different invariants.
  - Teacher models were trained on the corrupted image dataset **CIFAR-10-C** and its uncorrupted version **CIFAR-10**.
  - Each teacher model was trained on specific types of corruptions. (Gaussian Noise, Contrast, Elastic Transform, Defocus Blur, etc ...)

Target model

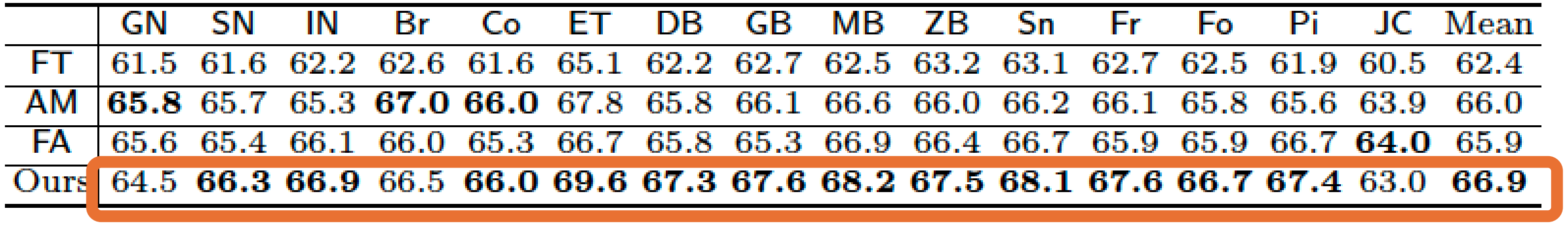
- Fine-tuning with the parameters of each teacher model as initial values.
  - Target model were trained on clean **CIFAR-100.**

Augmentation model

- 4 kinds of augmentation.
  - Contrast, Geometric transform, Gaussian blur and Gaussian noise.

Experimental Result : Accuracy

- The following table compares the accuracy of each target model on the CIFAR-100 test data set.
- The table also includes results for FIne-tuning (FT) and Fine-tuning with AugMix (AM, one of the data augmentation optimization methods), and Fine-tuning with FixedAug (FA, where the data augmentation is the same as the corruption learned by the teacher model).

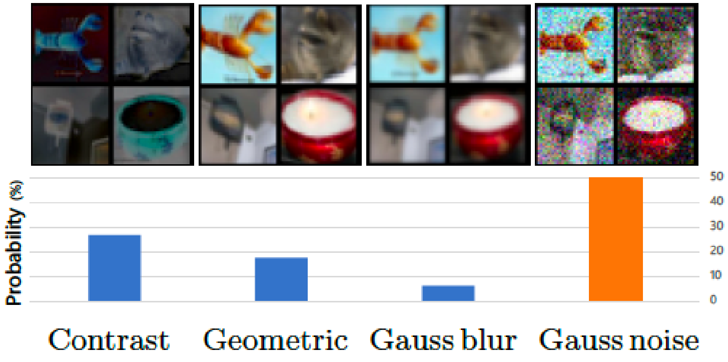


	GN	SN	IN	Br	Co	ET	DB	GB	MB	ZB	Sn	Fr	Fo	Pi	JC	Mean
FT	61.5	61.6	62.2	62.6	61.6	65.1	62.2	62.7	62.5	63.2	63.1	62.7	62.5	61.9	60.5	62.4
AM	**65.8**	65.7	65.3	**67.0**	**66.0**	67.8	65.8	66.1	66.6	66.0	66.2	66.1	65.8	65.6	63.9	66.0
FA	65.6	65.4	66.1	66.0	65.3	66.7	65.8	65.3	66.9	66.4	66.7	65.9	65.9	66.7	**64.0**	65.9
Ours	64.5	**66.3**	**66.9**	66.5	**66.0**	**69.6**	**67.3**	**67.6**	**68.2**	**67.5**	**68.1**	**67.6**	**66.7**	**67.4**	63.0	**66.9**

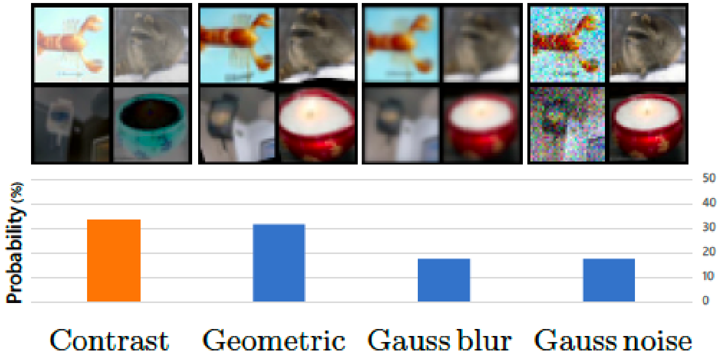
TransInv achieves the best in 12/15 cases !

Experimental Result : Data Augmentations Obtained

- The following shows the final data extension obtained when using a certain teacher model.



(a) Shot Noise teacher model



(b) Fog teacher model

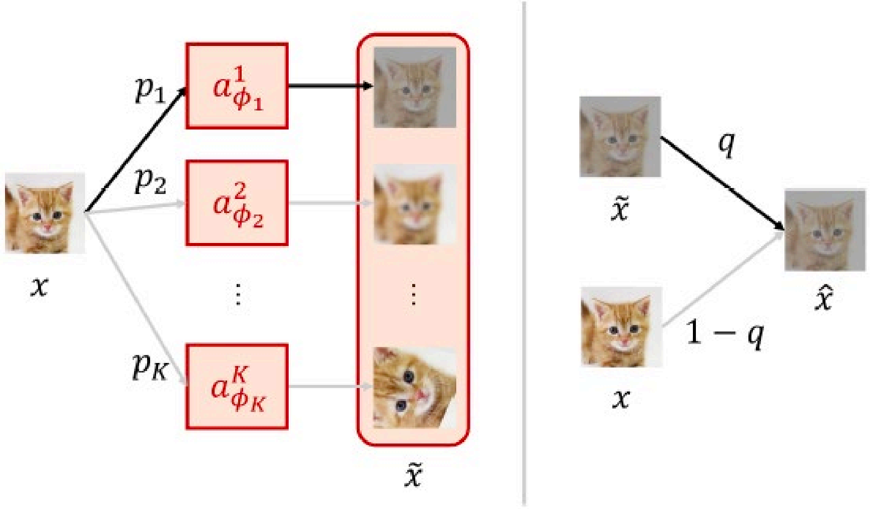
- Data augmentations obtained close to those corruptions where the teacher model exhibits invariance. This suggests that TransInv may have allowed the target model to acquire some of the invariant properties of the teacher model.

Summary

- We proposed a framework, TransInv, which enhances the invariance of the target model by performing data augmentation reflecting the invariant properties of the teacher model.
  - The method of joint optimization of the data augmentation model and the target model was used.
  - Experiments suggest that the data augmentations obtained by the proposed method may reflect the invariant properties of the teacher, and that the invariance of the target model trained with the augmentations may be extended.

Appendix : Data Augmentation Model

- Utilizes K data augmentation primitives (e.g., contrast shifting, blurring, rotation) to modify original image samples.
- Each primitive is defined by a neural network with learnable parameters and a specific deformation function.


p_1
p_2
p_K
$a^1_{\phi_1}$
$a^2_{\phi_2}$
$a^K_{\phi_K}$
x
$\tilde{x}$
q
$1-q$
$\tilde{x}$
x
$\hat{x}$