Covariance-aware Feature Alignment with Pre-computed Source Statistics for Test-time Adaptation to Multiple Image Corruptions (TA.PA.2)

Kazuki Adachi, Shin'ya Yamaguchi, and Atsutoshi Kumagai
NTT Computer and Data Science Laboratories

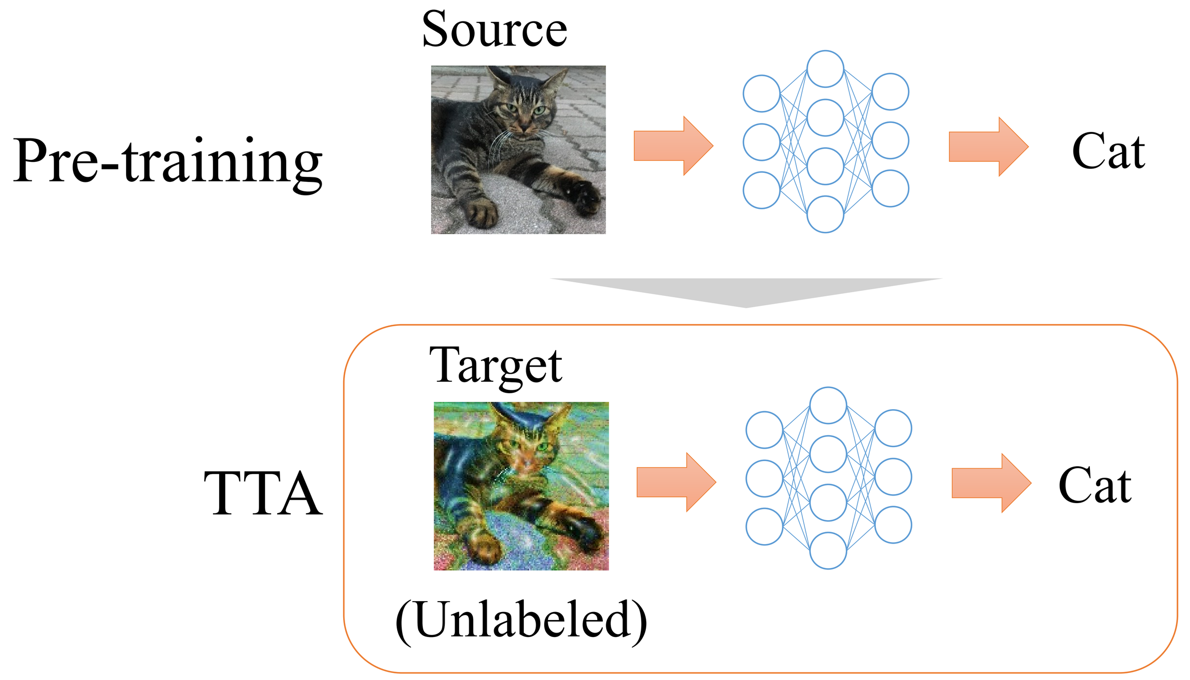
Paper and poster
(IEEE SIGPORT)

Introduction

- Models degrade accuracy when distribution shift occurs
  → Need to adapt models to the target domain
- Fine-tuning: additional annotation for target data required☹️
- Domain adaptation: simultaneous access to source and target data required
  - Obtaining target data in advance can be difficult☹️

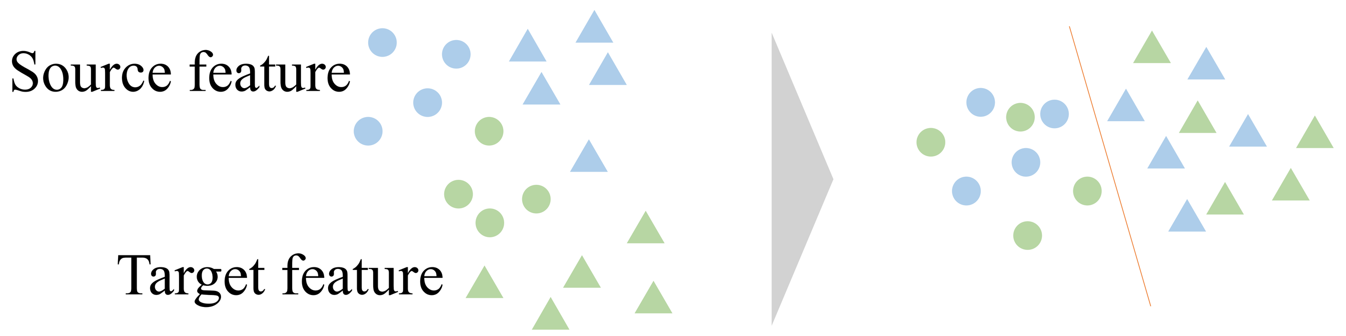
Test-time Adaptation (TTA)

- Adapts a pre-trained model to the target domain with unlabeled target data
- Does not access source data



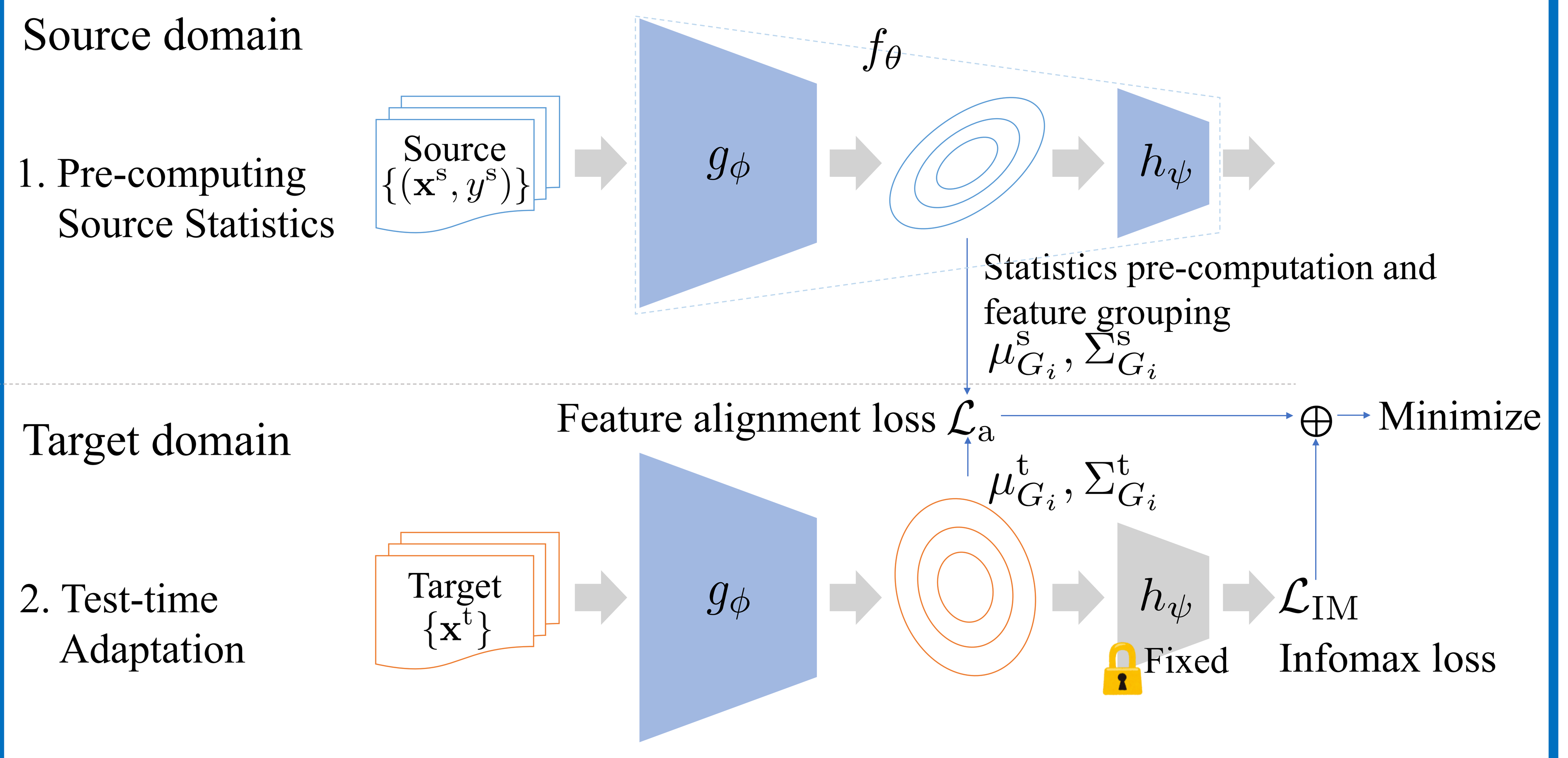
Proposed Method

- **Key Idea**
  - Feature alignment is important in domain adaptation
    → Can we improve TTA by feature alignment?



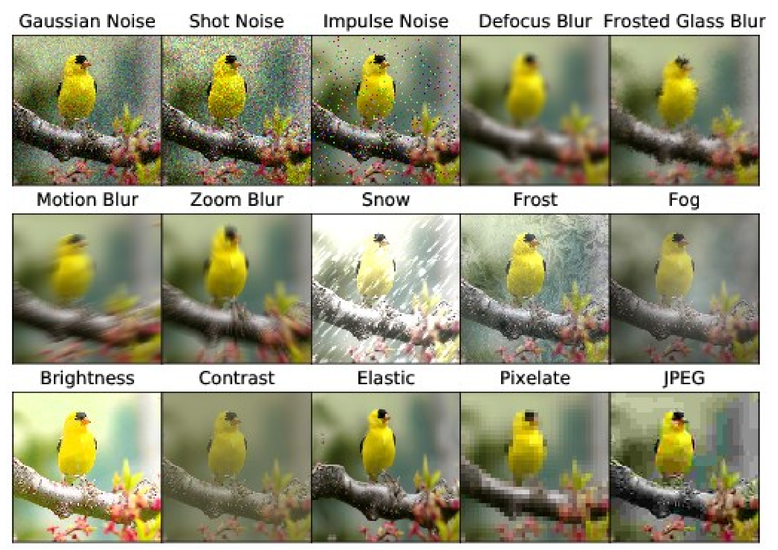
- **Covariance-aware Feature alignment (CAFe)**
  - Aligns source and target features with pre-computed source statistics
    → Source dataset itself is not required during TTA😀
  - Considering correlations between feature dimensions
    → accurate feature alignment😀

- **Feature alignment**
  1. Pre-computes the source statistics (mean and covariance)
  2. Feature-grouping: makes groups of feature dimensions correlated with each other by spectral clustering
  3. Aligns target batch statistics to the source one with KL-divergence



Experiment

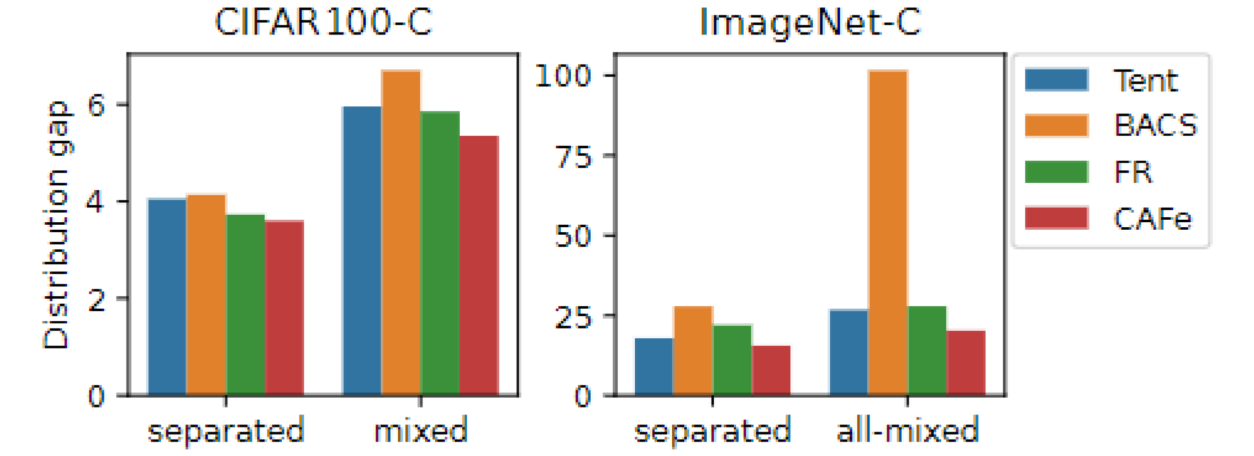
- Benchmark TTA performance under image corruption
- **Dataset**
  - Source: CIFAR10/100, ImageNet
  - Target: CIFAR10/100-C, ImageNet-C



- **Type of distribution shift**
  - Separated: a single type and strength of corruption
  - Severity-mixed: strengths are mixed
  - All-mixed: types and strengths are mixed (harder)
- **Result**
  - CAFe outperforms other TTA baselines, especially in all-mixed case
  - CAFe reduces distribution gap remains after TTA by feature alignment

Test accuracy after TTA

Method	CIFAR10-C Separated	CIFAR10-C Mixed	CIFAR100-C Separated	CIFAR100-C Mixed	ImageNet-C Separated	ImageNet-C Severity-mixed	ImageNet-C All-mixed
Source	63.75	$63.46_{\pm 0.61}$	34.24	$34.16_{\pm 0.20}$	39.14	$39.43_{\pm 0.00}$	$39.16_{\pm 0.01}$
AdaBN [4]	$80.26_{\pm 0.30}$	$67.62_{\pm 0.13}$	$51.10_{\pm 0.25}$	$38.52_{\pm 0.27}$	$50.28_{\pm 0.02}$	$48.00_{\pm 0.17}$	$39.85_{\pm 0.18}$
T3A [9]	$66.02_{\pm 0.02}$	$63.92_{\pm 0.42}$	$36.05_{\pm 0.07}$	$34.10_{\pm 0.49}$	$39.05_{\pm 0.01}$	$39.28_{\pm 0.03}$	$37.46_{\pm 0.09}$
Tent [7]	$80.86_{\pm 0.06}$	$68.59_{\pm 0.30}$	$52.09_{\pm 0.07}$	$38.95_{\pm 0.65}$	$58.97_{\pm 0.03}$	$57.15_{\pm 0.05}$	$44.44_{\pm 0.22}$
BACS [8]	$81.51_{\pm 0.02}$	$68.69_{\pm 0.09}$	$\mathbf{53.00_{\pm 0.12}}$	$39.65_{\pm 0.32}$	$57.01_{\pm 0.19}$	$55.05_{\pm 0.29}$	$33.07_{\pm 1.38}$
FR [21]	$80.71_{\pm 0.40}$	$68.31_{\pm 0.64}$	$51.50_{\pm 0.03}$	$39.44_{\pm 0.32}$	$53.54_{\pm 0.01}$	$50.38_{\pm 0.20}$	$40.52_{\pm 0.16}$
Infomax [23]	$81.40_{\pm 0.02}$	$69.01_{\pm 0.50}$	$52.48_{\pm 0.02}$	$39.78_{\pm 0.36}$	$60.20_{\pm 0.05}$	$57.52_{\pm 0.23}$	$46.52_{\pm 0.08}$
CAFe (w/o infomax)	$81.11_{\pm 0.02}$	$69.02_{\pm 0.62}$	$51.83_{\pm 0.02}$	$38.71_{\pm 0.11}$	$57.35_{\pm 0.02}$	$54.43_{\pm 0.14}$	$43.83_{\pm 0.16}$
CAFe (dimwise)	$81.40_{\pm 0.02}$	$69.10_{\pm 0.38}$	$52.48_{\pm 0.02}$	$39.83_{\pm 0.24}$	$60.29_{\pm 0.08}$	$58.60_{\pm 0.36}$	$47.19_{\pm 0.24}$
CAFe	$\mathbf{81.66_{\pm 0.01}}$	$\mathbf{70.06_{\pm 0.25}}$	$52.79_{\pm 0.02}$	$\mathbf{40.01_{\pm 0.36}}$	$\mathbf{60.77_{\pm 0.09}}$	$\mathbf{59.04_{\pm 0.22}}$	$\mathbf{48.55_{\pm 0.26}}$



Distribution gap (Fréchet distance) between source and target distributions remaining after TTA