Boyan Gao[1], Henry Gouk[1],

Yongxin Yang[1], Timothy Hospedales[1,2]

Loss Function Learning for Domain Generalization by Implicit Gradient

[1]School of Informatics, University of Edinburgh, United Kingdom

[2]Samsung AI Centre, Cambridge, United Kingdom

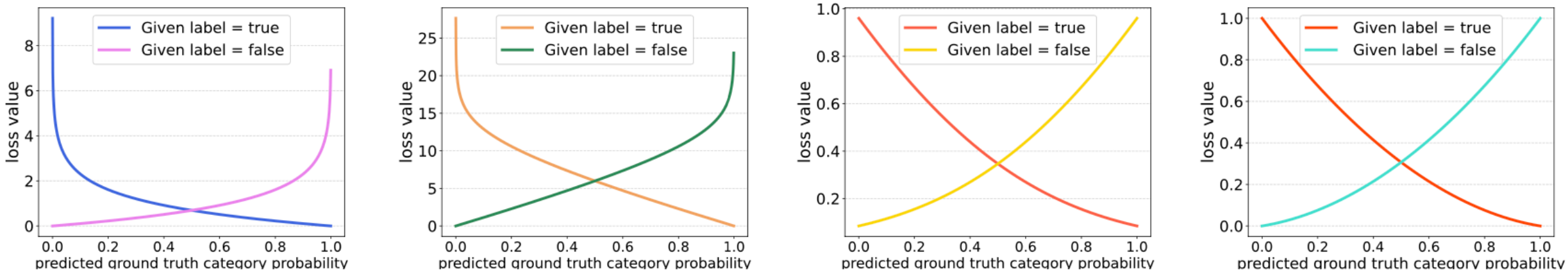
ICML
International Conference On Machine Learning

THE UNIVERSITY of EDINBURGH
informatics

ICML | 2022
Thirty-ninth International Conference on Machine Learning

THE UNIVERSITY of EDINBURGH
informatics

Motivation

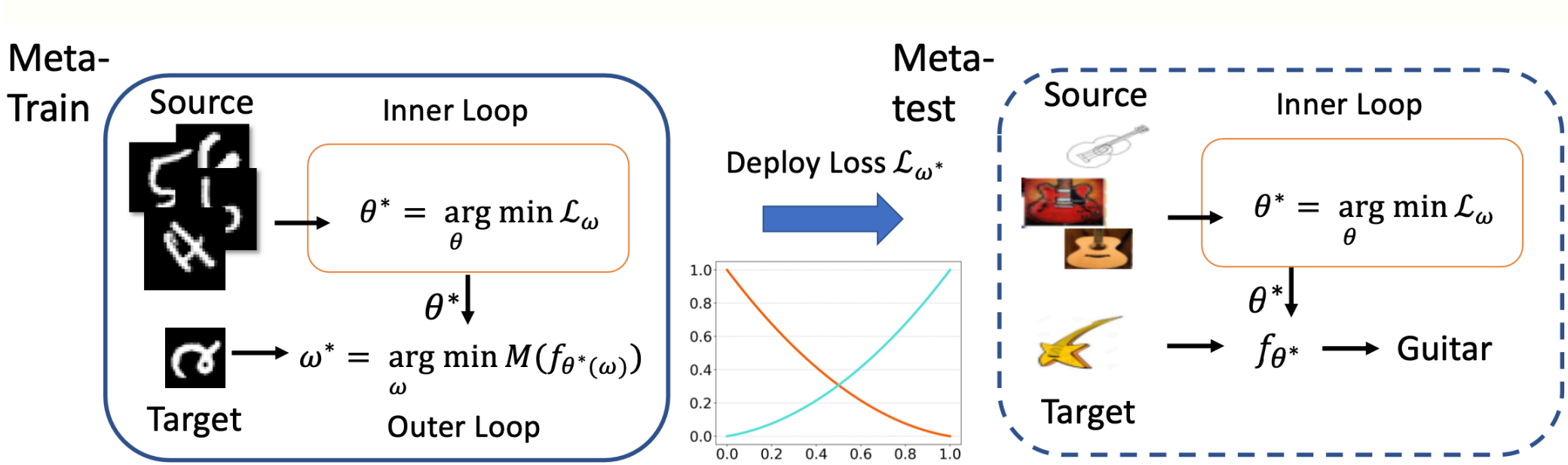


Comparison of loss function (from left to right: CE, SCE, FOCAL, ITL (our learned loss function))

- Loss Function Study for DG:
  - We provide the first study on the significance of supervised loss function choice in DG
- Loss Function Learning for DG:
  - We demonstrate the first efficient solution to loss-learning based on meta-gradients computed b the Implicit Function Theorem

THE UNIVERSITY of EDINBURGH
informatics

Method



- General idea Loss Function Learning:
  - Meta-train: The loss function is learned through the proposed algorithm.
  - Meta-test: The learned loss function is deployed on the novel DG tasks.

THE UNIVERSITY of EDINBURGH
informatics

Loss Function Learning

- Optimization framework
  - Bilevel optimization

$$\omega^* = \arg\min_\omega \hat{\mathbb{E}}_{\vec{x},y \in D_i^{(t)}} \left[\mathcal{M}(f_{\theta_i^*(\omega)}(\vec{x}), y) \right]$$
$$s.t.\ \theta_i^*(\omega) = \arg\min_\theta \hat{\mathbb{E}}_{\vec{x},y \in D_i^{(s)}} \left[\mathcal{L}_\omega(f_\theta(\vec{x}), y) \right]$$

- Simulating the source-target pipeline
  - Outer loop: Hold-out (Target) domain with Implicit Gradient
  - Inner loop: Source domain with SGD

- Loss parameterization
  - Flexible enough to include interesting losses
  - Simple enough to generalize

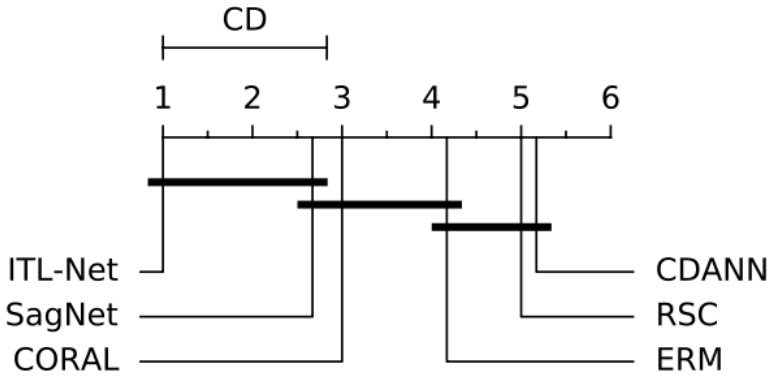
$$\mathcal{L}_\omega(\vec{\hat{y}}, \vec{y}) = \frac{1}{C} \sum_{i=1}^{C} \ell_\omega(\vec{\hat{y}}_i, \vec{y}_i)$$
$$\ell_\omega(\vec{\hat{y}}_i, \vec{y}_i) = \sum_{n=0}^{\beta} \frac{1}{n!} \nabla^n \ell([\omega_0, \omega_1])^T ([\vec{\hat{y}}_i, \vec{y}_i] - [\omega_0, \omega_1])^n$$

THE UNIVERSITY of EDINBURGH
informatics

Experiments

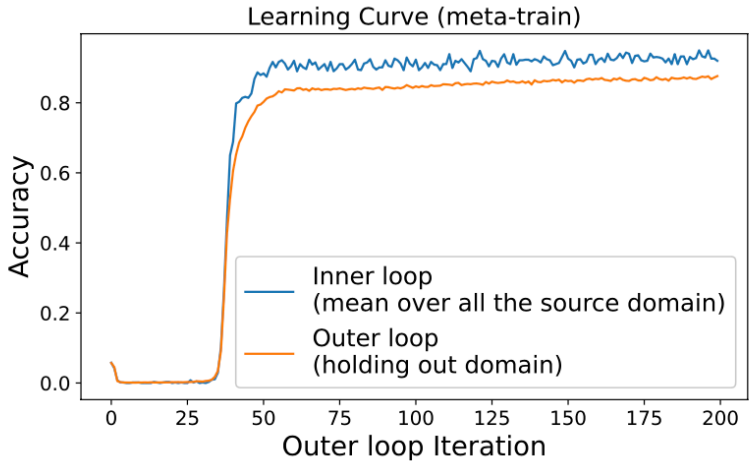
- Standard PACS setting with ResNet-18

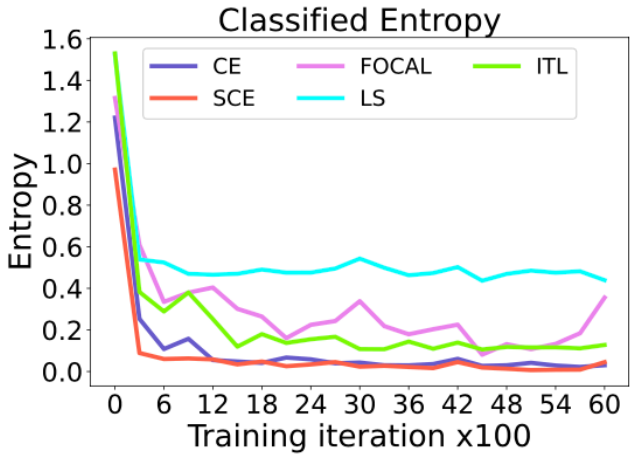
- DomainBed with critical different diagram

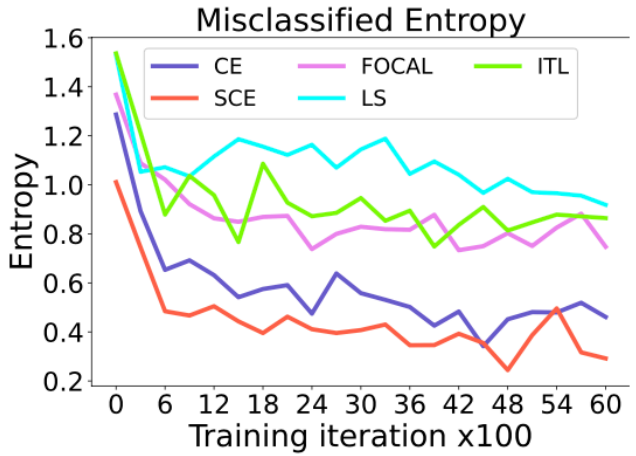


THE UNIVERSITY of EDINBURGH
informatics

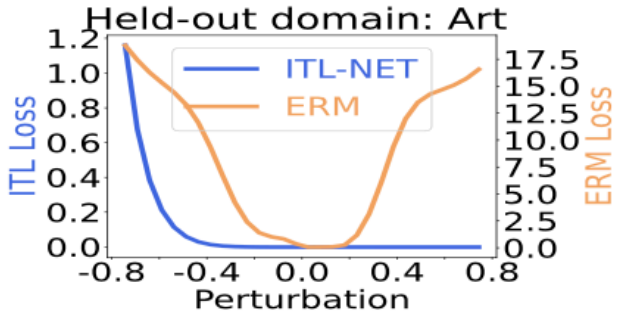
Loss Analysis

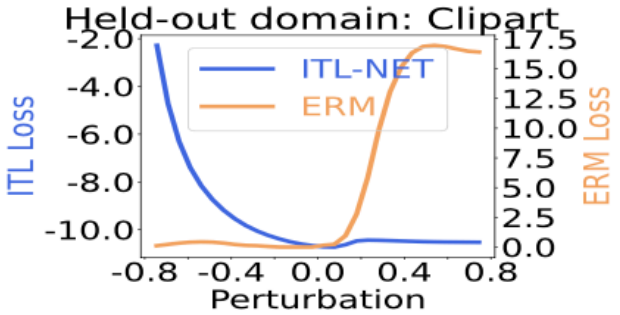






- Convergence Analysis
- In/Cross distribution Analysis
- Repeatability Analysis
- Entropy Analysis
- Loss Landscape Analysis





Thank you

THE UNIVERSITY of EDINBURGH
informatics

ICML | 2022

Thirty-ninth International
Conference on Machine Learning

ICML

International Conference
On Machine Learning