Domain-Inspired Sharpness-Aware Minimization Under Domain Shifts

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

上海人工智能实验室 Shanghai Artificial Intelligence Laboratory

Ruipeng Zhang[1,2], Ziqing Fan[1,2], Jiangchao Yao[1,2], Ya Zhang[1,2], Yanfeng Wang[1,2]

[1] Cooperative Medianet Innovation Center, Shanghai Jiao Tong University, [2] Shanghai AI Laboratory

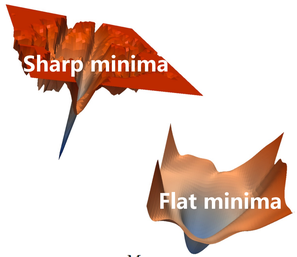
Github Repo

ICLR International Conference On Learning Representations 2024

Introduction

Sharpness

$$s(w, D) \triangleq \max_{\|\epsilon\|_2 \le \rho} \mathbb{E}_{\xi \in D}[\mathcal{L}(w+\epsilon;\xi) - \mathcal{L}(w;\xi)].$$

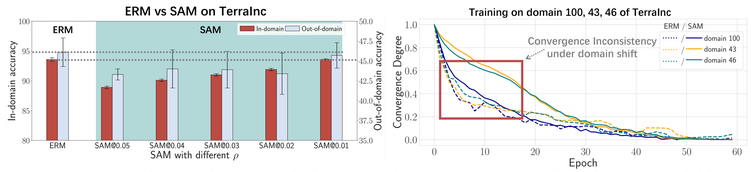


Sharpness-Aware Minimization (SAM)

$$\min_w \max_{\|\epsilon\|_2 \le \rho} \mathbb{E}_{\xi \in D}[\mathcal{L}(w+\epsilon;\xi)].$$

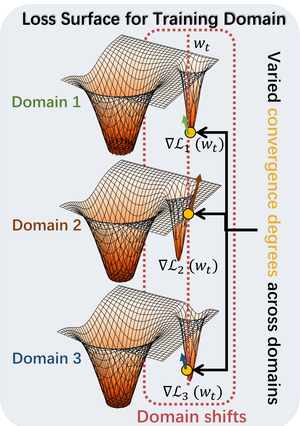
SAM under domain shifts

$$\min_w \mathbb{E}_{\xi \in \mathcal{S}}[\mathcal{L}_{SAM}(w;\xi)] = \min_w \max_{\|\epsilon\|_2 \le \rho} \mathbb{E}_{\xi \in \mathcal{S}}[\mathcal{L}(w+\epsilon;\xi)] \overset{?}{\approx} \min_w \max_{\|\epsilon\|_2 \le \rho} \sum_{i=1}^{M} \alpha_i \mathcal{L}_i(w+\epsilon).$$



SAM' s degradation of the training process under domain shifts

Motivation



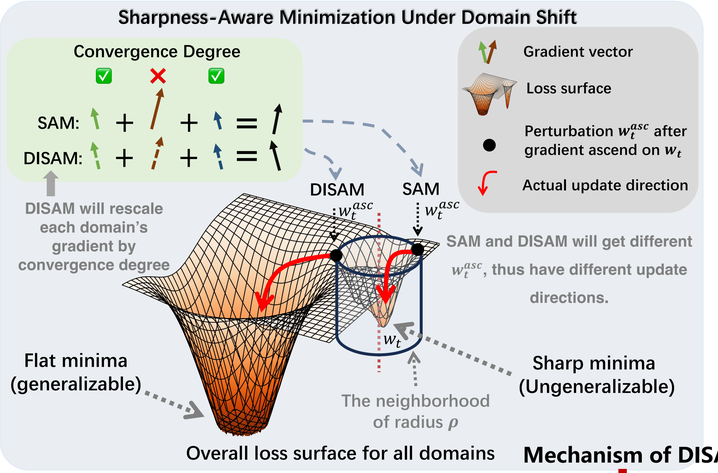
The **inherent heterogeneity among domains** can **considerably distort their sharpness estimation**.

SAM misdirects the perturbation direction towards the domain with the largest gradient vectors.

The direction should **focus more on domains with higher convergence degree** while being **mild to domains with lower convergence degree.**

New objective: $\min_w \mathbb{E}_{\xi \in \mathcal{S}}[\mathcal{L}_{DISAM}(w;\xi)] \triangleq \min_w \max_{\|\epsilon\|_2 \le \rho} \left[\sum_{i=1}^{M} \alpha_i \mathcal{L}_i(w+\epsilon) - \lambda \mathrm{Var}\{\mathcal{L}_i(\hat{w}+\epsilon)\}_{i=1}^{M}\right]$

Method: DISAM



Mechanism of DISAM

$$\nabla \mathcal{L}_{DISAM} = \sum_{i=1}^{M}\left(\alpha_i - \frac{2\lambda}{M}\left(\mathcal{L}_i(w) - \frac{1}{M}\sum_{j=1}^{M}\mathcal{L}_j(w)\right)\right)\nabla \mathcal{L}_i(w) = \nabla \mathcal{L}_{SAM} - \underbrace{\sum_{i=1}^{M}\frac{2\lambda}{M}\left(\mathcal{L}_i(w) - \frac{1}{M}\sum_{j=1}^{M}\mathcal{L}_j(w)\right)\nabla \mathcal{L}_i(w)}_{\text{Adaptive adjustment: increase weights for smaller losses, reduce for larger ones.}}$$

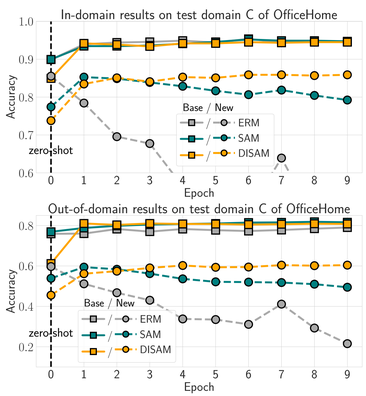
Main Results

DG benchmarks with ResNet50

Algorithm	PACS	VLCS	OfficeHome	TerraInc	DomainNet	Avg.
In-domain results						
ERM	96.6 ± 0.2	84.6 ± 0.4	84.2 ± 0.3	93.6 ± 0.3	67.1 ± 1.6	85.2
SAM	97.3 ± 0.1	84.8 ± 0.3	85.8 ± 0.2	88.9 ± 0.2	68.5 ± 0.1	85.1
Domain-Inspired	97.8 ± 0.1	84.4 ± 0.3	86.3 ± 0.2	94.8 ± 0.2	70.2 ± 0.1	86.7
GSAM	97.8 ± 0.2	83.9 ± 0.2	85.9 ± 0.2	92.1 ± 0.2	69.1 ± 0.1	85.8
Domain-Inspired	97.9 ± 0.1	**85.1** ± 0.4	86.2 ± 0.2	94.8 ± 0.3	70.0 ± 0.1	86.8
SAGM	97.6 ± 0.1	84.6 ± 0.3	86.1 ± 0.2	92.0 ± 0.2	69.2 ± 0.1	85.9
Domain-Inspired	**97.9** ± 0.1	85.0 ± 0.2	**86.5** ± 0.3	**94.9** ± 0.2	**70.5** ± 0.1	**87.0**
Out-of-domain results						
ERM	85.5 ± 0.2	77.3 ± 0.4	66.5 ± 0.3	46.1 ± 1.8	43.8 ± 0.1	63.9
CORAL (SOTA)	86.2 ± 0.3	78.8 ± 0.3	68.7 ± 0.3	47.6 ± 1.0	41.5 ± 0.1	64.5
SAM	85.8 ± 0.2	79.4 ± 0.1	69.6 ± 0.1	43.3 ± 0.7	44.3 ± 0.0	64.5
Domain-Inspired	87.3 ± 0.2	80.1 ± 0.5	70.7 ± 0.2	47.9 ± 0.8	45.8 ± 0.2	66.4
GSAM	85.9 ± 0.1	79.1 ± 0.2	69.3 ± 0.0	47.0 ± 0.8	44.6 ± 0.2	65.1
Domain-Inspired	87.2 ± 0.3	80.0 ± 0.3	70.8 ± 0.3	**50.6** ± 1.2	45.6 ± 0.1	66.8
SAGM	86.6 ± 0.2	80.0 ± 0.3	70.1 ± 0.2	48.8 ± 0.9	45.0 ± 0.2	66.1
Domain-Inspired	**87.5** ± 0.3	**80.7** ± 0.2	**71.0** ± 0.2	50.0 ± 1.2	**46.0** ± 0.1	**67.0**
+ CORAL	88.4 ± 0.3	81.2 ± 0.4	71.7 ± 0.2	51.7 ± 0.3	46.3 ± 0.2	67.9

DG benchmarks with CLIP-ViT

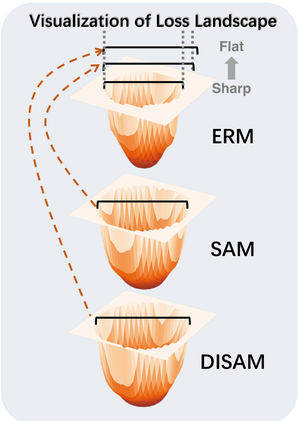
Algorithm	PACS	VLCS	OfficeHome	TerraInc	DomainNet	Avg.
Zero-shot	96.2	81.7	82.0	33.4	57.5	70.2
CoOp	96.8	81.2	84.2	44.9	59.9	73.4
+ SAM	97.1 ± 0.1	81.3 ± 0.8	84.6 ± 0.2	47.7 ± 1.3	60.3 ± 0.2	74.2
+ DISAM	**97.2** ± 0.1	81.8 ± 0.4	84.8 ± 0.2	49.5 ± 1.2	60.6 ± 0.2	74.8
ERM	96.1 ± 0.5	83.0 ± 0.2	83.3 ± 0.3	60.9 ± 0.2	59.9 ± 0.1	76.7
CLIPOOD[†]	97.3 ± 0.1	85.0 ± 0.4	87.0 ± 0.2	60.4 ± 0.7	63.5 ± 0.1	78.6
CLIPOOD*[2]	96.6 ± 0.4	84.1 ± 0.3	86.1 ± 0.2	59.7 ± 0.8	63.1 ± 0.1	77.9
+ SAM	96.9 ± 0.2	84.3 ± 0.6	84.4 ± 0.4	60.0 ± 1.4	58.6 ± 0.2	76.9
+ DISAM	97.1 ± 0.1	**85.6** ± 0.2	**86.6** ± 0.0	**61.1** ± 0.7	**63.6** ± 0.1	**78.8**

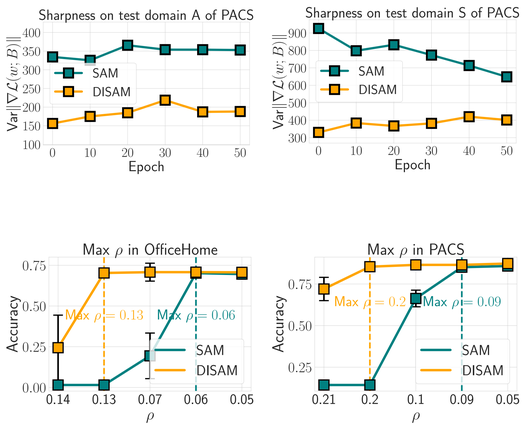


Open class generalization with CLIP

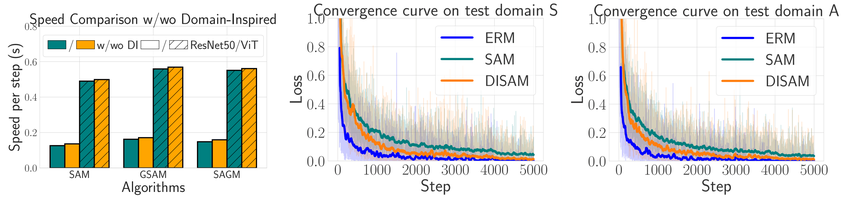
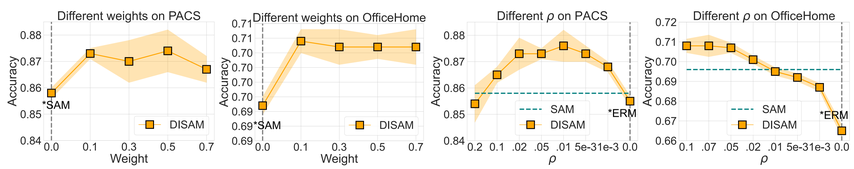
Visualization and Analysis

Enhanced Generalization with DISAM





Computation Cost & Convergence Comparison



Sensitivity of hyperparameters in DISAM