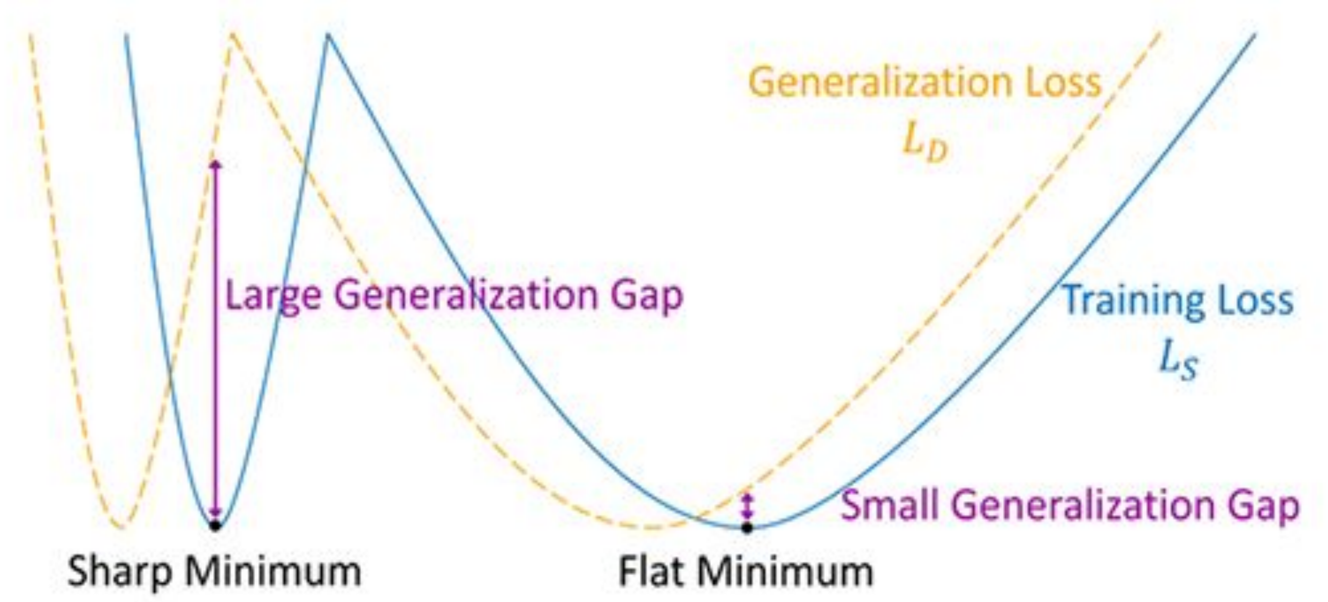
A Closer Look at Smoothness in Domain Adversarial Training

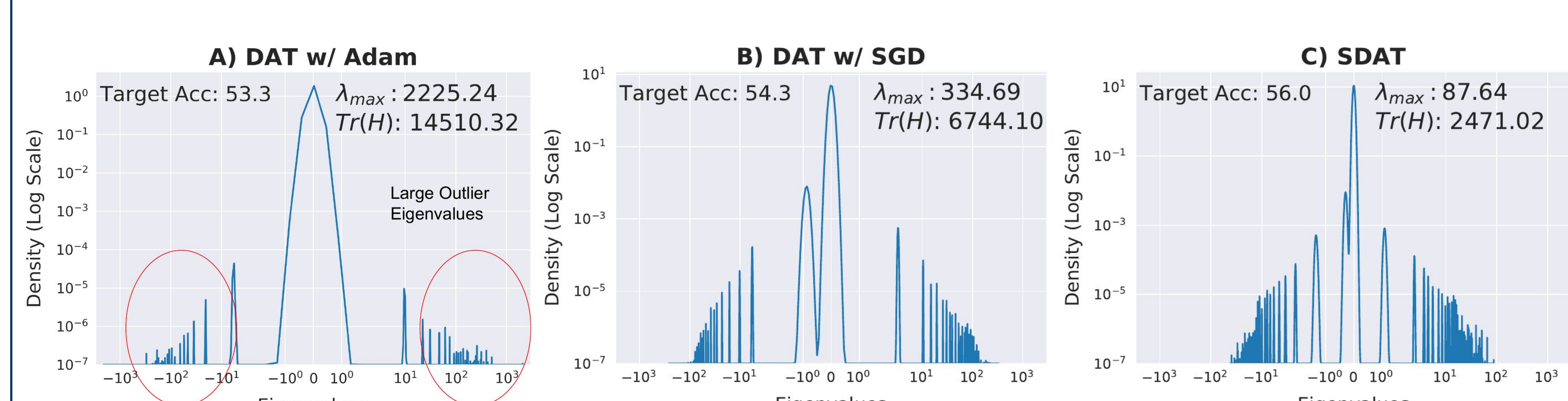
Harsh Rangwani* Sumukh K Aithal* Mayank Mishra Arihant Jain R. Venkatesh Babu
Video Analytics Lab, Indian Institute of Science

Motivation



- As compared to sharp minima, convergence to a flatter (smooth) region within a loss landscape improves generalization in supervised learning.
- Our work analyses the effect of smoothness in loss landscape for domain adversarial training (DAT), which performs domain adaptation of classifier from source to target domain by learning domain invariant representations.
- The DAT objective is a combination of task loss (i.e. classification etc.) and adversarial loss.

Analysis of Task Loss

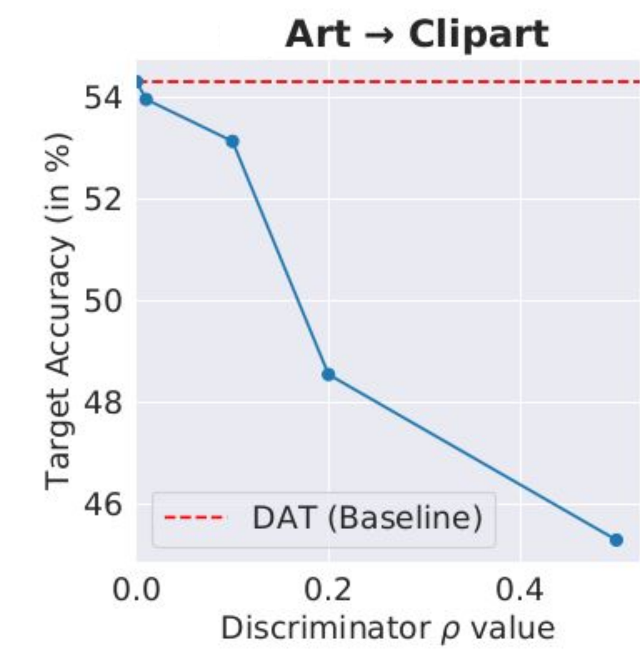


- Eigen spectral density of Hessian of the task loss above analyzes the smoothness of the task loss.
- Low λ_{max} and low Tr(H) as we go from left to right indicates convergence to smoother region for various methods in loss landscape, which is correlated with improved accuracy on target.

Task Loss Smoothness ↑ Performance ↑

Analysis of Adversarial Loss

Theorem 2 (Remark): We show that for gradient lipschitz functions the discriminator is suboptimal between source and target domain, with smooth version of adversarial loss .

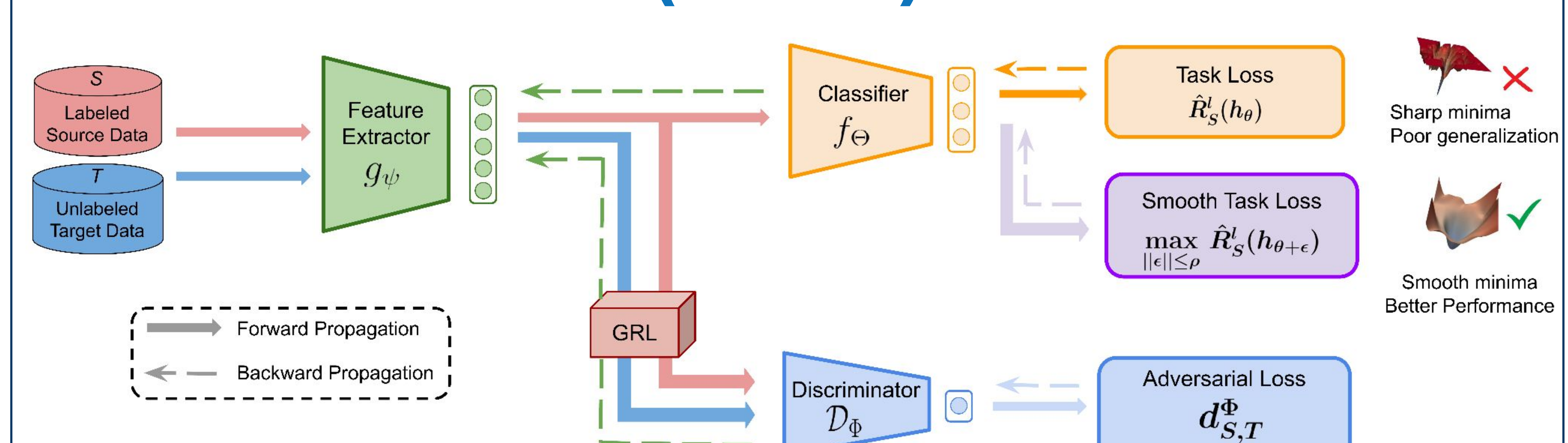


Empirical: As the smoothness increases (ρ), the target accuracy decreases (Fig. on right) indicating that smoothing adversarial loss leads to suboptimal generalization.

Adversarial Smoothness ↑

Performance ↓

Smooth Domain Adversarial Training (SDAT)



SDAT introduces smooth task loss inspired by Sharpness Aware Minimization (SAM) objective:

$$\min_\theta \max_\Phi \mathbb{E}_{x\sim P_S}[l(h_\theta(x), y(x))] + d^\Phi_{S,T}$$

Why Smooth Domain Adversarial Training?

- SDAT consistently improves the performance of standard DAT methods (eg. CDAN, DANN, MCC)
- Requires only a few lines of code change for integration.

Experiments

Integrating SDAT with DAT methods improves their performance significantly.

Dataset	Method	Backbone	DAT	SDAT
Office-Home	CDAN+MCC	ResNet-50	71.3	**72.2**
Office-Home	CDAN+MCC	ViT-B/16	82.2	**84.3**
VisDA-2017	CDAN+MCC	ResNet-50	83.6	**84.3**
VisDA-2017	CDAN+MCC	ViT-B/16	87.7	**89.8**
DomainNet	CDAN	ResNet-50	40.3	**42.1**

SDAT improves over recent SOTA methods with less compute and training resources.

	TVT (2021)	CDTrans (2022)	SDAT
Additional modules	✓	✓	✗
Memory requirement	∼35GB	∼26.3GB	<12GB
Pretraining	ImageNet-21k	ImageNet-1k	ImageNet-1k
Accuracy (Office-Home)	83.6	80.5	**84.3**
Accuracy (VisDA-2017)	83.2	88.4	**89.8**

Analysis of SDAT (Office-Home)

SDAT outperforms other smoothing techniques proposed for ERM, because it selectively smoothens the task loss.

Method	Ar→Cl	Cl→Pr	Rw→Cl	Pr→Cl	Avg
DAT	54.3	69.5	60.1	55.3	59.2
VAT	54.6	70.7	60.8	54.4	60.1 (+0.9)
SWAD	54.6	71.0	60.9	55.2	60.4 (+1.2)
LS	53.6	71.6	59.9	53.4	59.6 (+0.4)
SAM	54.9	70.9	59.2	53.9	59.7 (+0.5)
SDAT	**56.0**	**73.2**	**61.4**	**55.9**	**61.6** (+2.4)

Only smoothing Task Loss (i.e. SDAT) leads to the best performance on the target domain.

Smooth Task	Smooth Adv	Accuracy
✗	✗	54.3
✗	✓	51.0
✓	✗	**55.7**
✓	✓	54.9

Acknowledgement

This work was supported in part by SERB-STAR Project (Project:STR/2020/000128), Govt. of India. Harsh Rangwani is supported by PMRF fellowship.

Paper: https://arxiv.org/abs/2206.08213
Code: https://github.com/val-iisc/SDAT