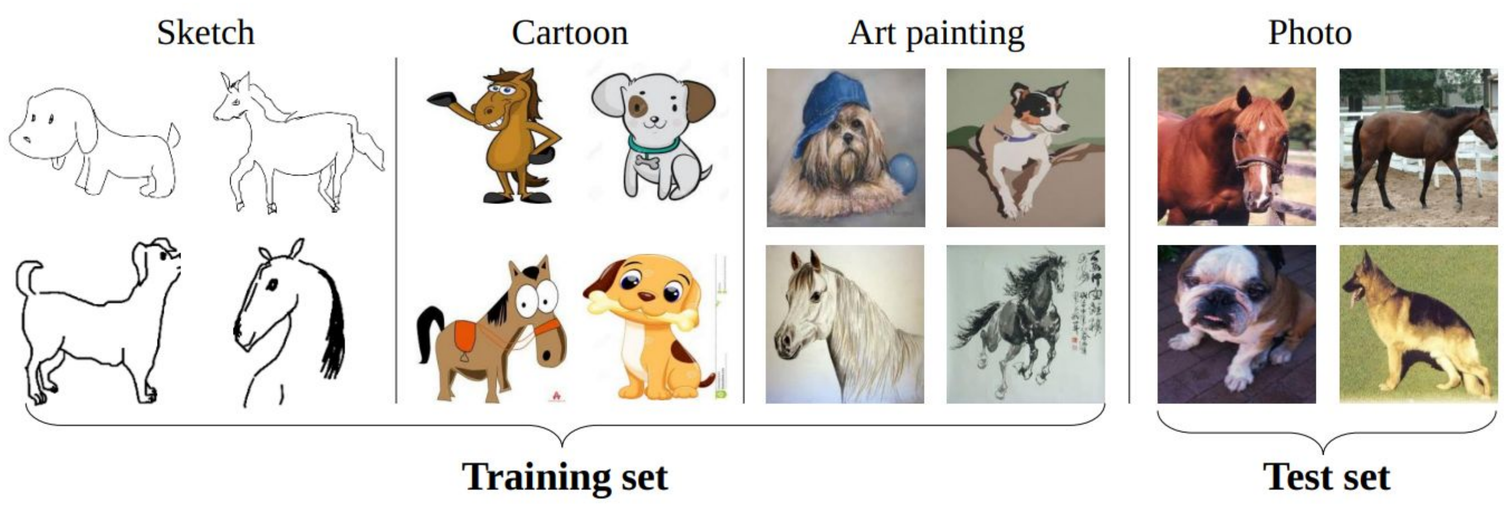
Domain Generalization with Nuclear Norm Regularization

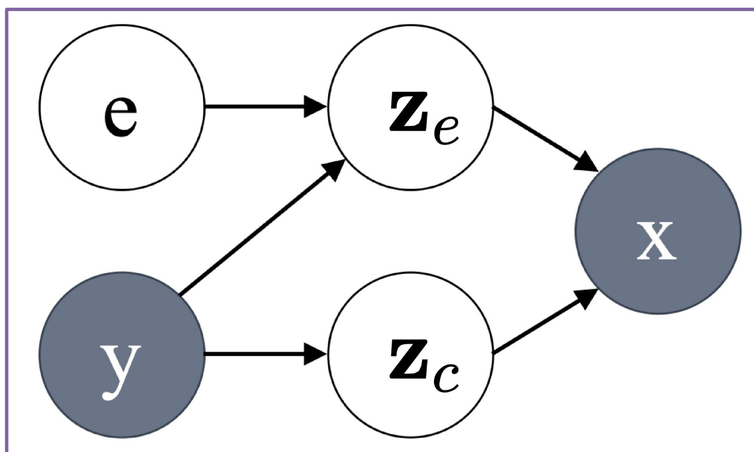
Zhenmei Shi*, Yifei Ming*, Ying Fan*, Frederic Sala, Yingyu Liang

Motivation



Domain Generalization

Source "Generalizing to Unseen Domains: A Survey on Domain Generalization." (2022)



Data Distribution

z_e: environmental feature, z_c: invariant feature

Inspired by "The risks of invariant risk minimization." (2021)

Problem Setup

- Test on **unseen** domains:

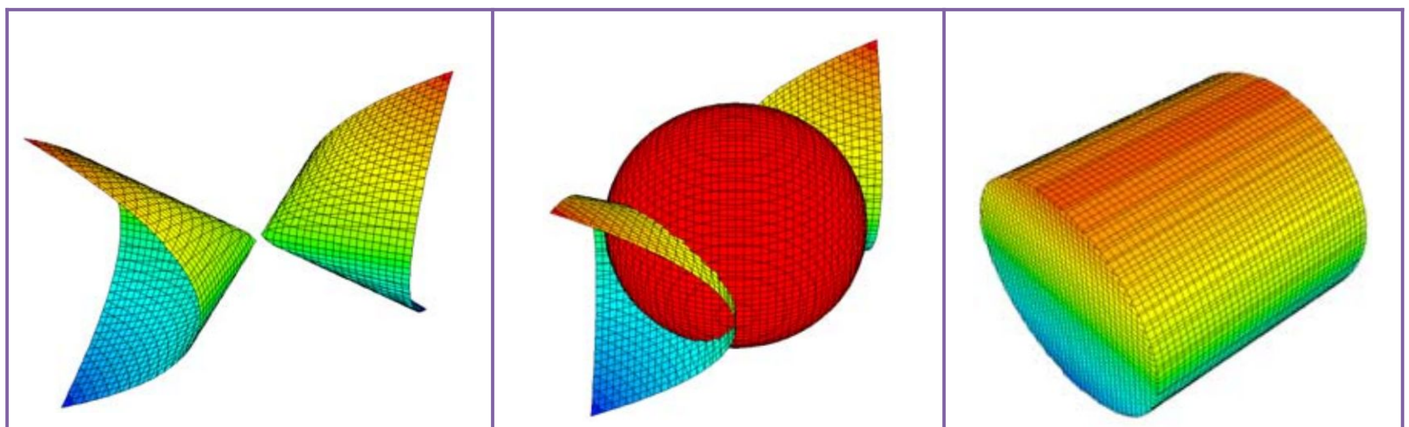
$$\mathcal{L}(\hat{f}, \Phi) = \mathbf{E}_{(\mathbf{x},y)\sim\mathcal{D}_{\text{ood}}}\left[\ell(\hat{f}(\Phi(\mathbf{x})), y)\right]$$

- Matrix Rank regularization (ERM-rank):

$$\min_{\hat{f},\Phi} \mathcal{L}(\hat{f}, \Phi) + \lambda \text{rank}(\Phi(\mathbf{X}))$$

- Convex envelope - **nuclear norm:**

$$\min_{\hat{f},\Phi} \mathcal{L}(\hat{f}, \Phi) + \lambda \|\Phi(\mathbf{X})\|_*$$



(a) rank 1 (b) nuclear norm ball (c) convex hull

Manifold of constant penalty for a symmetric 2×2 matrix X = [x, y; y, z]

Experiments

Model

ResNet-50, pre-trained on ImageNet.

Dataset

Colored MNIST, Rotated MNIST, PACS, VLCS, Office-Home, Terra Incognita, and DomainNet.

Evaluation & Methods

Model selection criterion:
Training-domain validation set.

Batch-wise nuclear norm regularization.

Results

- DomainNet: ours beating all SOTA on all domains;
- DomainBed: ours achieving the best compared to all baselines.

Algorithm	clip	info	paint	quick	real	sketch	Average
ERM [41]	58.1 ± 0.3	18.8 ± 0.3	46.7 ± 0.3	12.2 ± 0.4	59.6 ± 0.1	49.8 ± 0.4	40.9
IRM [1]	48.5 ± 2.8	15.0 ± 1.5	38.3 ± 4.3	10.9 ± 0.5	48.2 ± 5.2	42.3 ± 3.1	33.9
GroupDRO [37]	47.2 ± 0.5	17.5 ± 0.4	33.8 ± 0.5	9.3 ± 0.3	51.6 ± 0.4	40.1 ± 0.6	33.3
Mixup [45]	55.7 ± 0.3	18.5 ± 0.5	44.3 ± 0.5	12.5 ± 0.4	55.8 ± 0.3	48.2 ± 0.5	39.2
MLDG [23]	59.1 ± 0.2	19.1 ± 0.3	45.8 ± 0.7	13.4 ± 0.3	59.6 ± 0.2	50.2 ± 0.4	41.2
CORAL [39]	59.2 ± 0.1	19.7 ± 0.2	46.6 ± 0.3	13.4 ± 0.4	59.8 ± 0.2	50.1 ± 0.6	41.5
MMD [25]	32.1 ± 13.3	11.0 ± 4.6	26.8 ± 11.3	8.7 ± 2.1	32.7 ± 13.8	28.9 ± 11.9	23.4
DANN [12]	53.1 ± 0.2	18.3 ± 0.1	44.2 ± 0.7	11.8 ± 0.1	55.5 ± 0.4	46.8 ± 0.6	38.3
CDANN [27]	54.6 ± 0.4	17.3 ± 0.1	43.7 ± 0.9	12.1 ± 0.7	56.2 ± 0.4	45.9 ± 0.5	38.3
MTL [6]	57.9 ± 0.5	18.5 ± 0.4	46.0 ± 0.1	12.5 ± 0.1	59.5 ± 0.3	49.2 ± 0.1	40.6
SagNet [29]	57.7 ± 0.3	19.0 ± 0.2	45.3 ± 0.3	12.7 ± 0.5	58.1 ± 0.5	48.8 ± 0.2	40.3
ARM [49]	49.7 ± 0.3	16.3 ± 0.5	40.9 ± 1.1	9.4 ± 0.1	53.4 ± 0.4	43.5 ± 0.4	35.5
VREx [22]	47.3 ± 3.5	16.0 ± 1.5	35.8 ± 4.6	10.9 ± 0.3	49.6 ± 4.9	42.0 ± 3.0	33.6
RSC [18]	55.0 ± 1.2	18.3 ± 0.5	44.4 ± 0.6	12.2 ± 0.2	55.7 ± 0.7	47.8 ± 0.9	38.9
AND-mask [30]	52.3 ± 0.8	16.6 ± 0.3	41.6 ± 1.1	11.3 ± 0.1	55.8 ± 0.4	45.4 ± 0.9	37.2
SelfReg [19]	58.5 ± 0.1	20.7 ± 0.1	47.3 ± 0.3	13.1 ± 0.3	58.2 ± 0.2	51.1 ± 0.3	41.5
Fishr [33]	58.2 ± 0.5	20.2 ± 0.2	47.7 ± 0.3	12.7 ± 0.2	60.3 ± 0.2	50.8 ± 0.1	41.7
ERM-NU (ours)	**60.9** ± 0.0	**21.1** ± 0.2	**49.9** ± 0.3	**13.7** ± 0.2	**62.5** ± 0.2	**52.5** ± 0.4	**43.4**

Results on DomainNet

Algorithm	CMNIST	RMNIST	VLCS	PACS	OfficeHome	TerraInc	DomainNet	Average
ERM	51.5 ± 0.1	98.0 ± 0.0	77.5 ± 0.4	85.5 ± 0.2	66.5 ± 0.3	46.1 ± 1.8	40.9 ± 0.1	66.6
IRM	52.0 ± 0.1	97.7 ± 0.1	78.5 ± 0.5	83.5 ± 0.8	64.3 ± 2.2	47.6 ± 0.8	33.9 ± 2.8	65.4
GroupDRO	52.1 ± 0.0	98.0 ± 0.0	76.7 ± 0.6	84.4 ± 0.8	66.0 ± 0.7	43.2 ± 1.1	33.3 ± 0.2	64.8
Mixup	52.1 ± 0.2	98.0 ± 0.1	77.4 ± 0.6	84.6 ± 0.6	68.1 ± 0.3	47.9 ± 0.8	39.2 ± 0.1	66.7
MLDG	51.5 ± 0.1	97.9 ± 0.0	77.2 ± 0.4	84.9 ± 1.0	66.8 ± 0.6	47.7 ± 0.9	41.2 ± 0.1	66.7
CORAL	51.5 ± 0.1	98.0 ± 0.1	**78.8** ± 0.6	86.2 ± 0.3	**68.7** ± 0.3	47.6 ± 1.0	41.5 ± 0.1	67.5
MMD	51.5 ± 0.2	97.9 ± 0.0	77.5 ± 0.9	84.6 ± 0.5	66.3 ± 0.1	42.2 ± 1.6	23.4 ± 9.5	63.3
DANN	51.5 ± 0.3	97.8 ± 0.1	78.6 ± 0.4	83.6 ± 0.4	65.9 ± 0.6	46.7 ± 0.5	38.3 ± 0.1	66.1
CDANN	51.7 ± 0.1	97.9 ± 0.1	77.5 ± 0.1	82.6 ± 0.9	65.8 ± 1.3	45.8 ± 1.6	38.3 ± 0.3	65.6
MTL	51.4 ± 0.1	97.9 ± 0.0	77.2 ± 0.4	84.6 ± 0.5	66.4 ± 0.5	45.6 ± 1.2	40.6 ± 0.1	66.2
SagNet	51.7 ± 0.0	98.0 ± 0.0	77.8 ± 0.5	**86.3** ± 0.2	68.1 ± 0.1	48.6 ± 1.0	40.3 ± 0.1	67.2
ARM	**56.2** ± 0.2	**98.2** ± 0.1	77.6 ± 0.3	85.1 ± 0.4	64.8 ± 0.3	45.5 ± 0.3	35.5 ± 0.2	66.1
VREx	51.8 ± 0.1	97.9 ± 0.1	78.3 ± 0.2	84.9 ± 0.6	66.4 ± 0.6	46.4 ± 0.6	33.6 ± 2.9	65.6
RSC	51.7 ± 0.2	97.6 ± 0.1	77.1 ± 0.5	85.2 ± 0.9	65.5 ± 0.9	46.6 ± 1.0	38.9 ± 0.5	66.1
AND-mask	51.3 ± 0.2	97.6 ± 0.1	78.1 ± 0.9	84.4 ± 0.9	65.6 ± 0.4	44.6 ± 0.3	37.2 ± 0.6	65.5
SelfReg	52.1 ± 0.2	98.0 ± 0.1	77.8 ± 0.9	85.6 ± 0.4	67.9 ± 0.7	47.0 ± 0.3	41.5 ± 0.2	67.1
Fishr	52.0 ± 0.2	97.8 ± 0.0	77.8 ± 0.1	85.5 ± 0.4	67.8 ± 0.1	47.4 ± 1.6	41.7 ± 0.0	67.1
ERM-NU (ours)	51.8 ± 0.2	98.0 ± 0.1	77.8 ± 0.7	85.6 ± 0.1	68.1 ± 0.1	**49.6** ± 0.6	**43.4** ± 0.1	**67.8**

Results on DomainBed

Theoretical Analysis

Proposition (Informal)

Under our linear data and linear model setting (refer to Section 3), the *optimal* solution for the **ERM-rank** on the ID tasks has 100% OOD test accuracy, while the *optimal* solution for the **ERM-weight-decay** on the ID tasks may have OOD test accuracy much worse than the random guessing.

Take-Home Message

Nuclear norm regularization works better than ERM or ERM-weight-decay.

Key Intuition: ERM-weight-decay encodes all features correlated with labels, even weak correlation (spurious features), while ERM-rank only encodes features having a strong correlation with labels (intrinsic features).