Grounding Stylistic Domain Generalization with Quantitative Domain Shift Measures and Synthetic Scene Images

Yiran Luo* Joshua Feinglass* Tejas Gokhale^ Kuan-Cheng Lee* Chitta Baral* Yezhou Yang*

* Arizona State University
^ UMBC

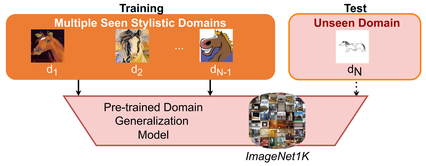
ARXIV

GITHUB

CVPR JUNE 17-21, 2024 SEATTLE, WA

Concept

Current methodology for Image Stylistic Domain Generalization (DG) often relies on vast weakly supervised pre-training data to reduce qualitative distances amongst stylistic domains.



We observe such methods possess risks for **inflated DG benchmarking performance based on ungrounded spurious correlation:**

- Pre-dominant **bias of photo-realistic style** in commonly used pre-training data, e.g. ImageNet1K, when benchmarking[1,2].
- Lack of **quantitative understandings of shifts between specific domains** and hence difficulty in extracting domain-invariant features[3,4].

Contributions

- **The Measures - IDD and ICV** to quantify inter- and intra-domain shifts in visual features.
- **The Precursor Dataset - SuperMarioDomains (SMD)%** of unique and diverse synthetic style video game graphics dissimilar to that of ImageNet1K.
- **The DG Method - SMOS** which uses SMD to obtain a precursor model that grounds the training of DG. Achieves top performance on multiple benchmarks esp. when targeting abstract-styled domains.

References and Acknowledgements

[1] Deng, et al. ImageNet: A large-scale Hierarchical Image Database. CVPR 2009.
[2] Li, et al. Simple: Specialized Model-sample Matching for Domain Generalization. ICLR 2022.
[3] Cha, et al. Domain Generalization by Mutual-information Regularization with Pre-trained Models. ECCV 2022.
[4] Gulrajani and Lopez-Paz. In Search of Lost Domain Generalization. arXiv:2007.01434 2020.

% The release of SMD will abide *Nintendo Game Content Guidelines for Online Video & Image Sharing Platforms* (Updated 10/24/2023).

The authors acknowledge Research Computing at Arizona State University for providing HPC resources and support for this work. This work was supported by NSF RI grants #1750082 and #2132724.

Methodology

The Measures

Inter-Domain Dissimilarity (IDD)

$$\mathrm{IDD}(\mathcal{D}_1, \mathcal{D}_2) = D_{\mathrm{JS}}(P_{\mathcal{D}_1}, P_{\mathcal{D}_2})$$

D_{JS} - Jenson-Shannon Divergence

$P_{\mathcal{D}_1}$ - Feature distribution of images in domain $\mathcal{D}_1$

Intra-Class Variation (ICV) of domain $\mathcal{D}$

$$\mathrm{ICV}(\mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n} D_{\mathrm{JS}}(P_i, Q_i)$$

n - Number of classes in domain $\mathcal{D}$

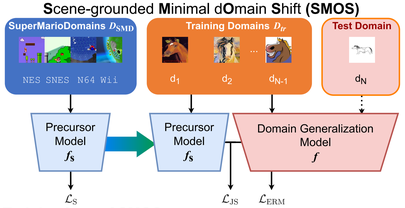
P_i, Q_i - Feature distributions of two randomly split equal halves of the images belonging to the same class i

The Precursor Dataset

SuperMarioDomains (SMD) of 82K video game scenes

The DG Method

Scene-grounded Minimal dOmain Shift (SMOS)



Training loss of SMOS:

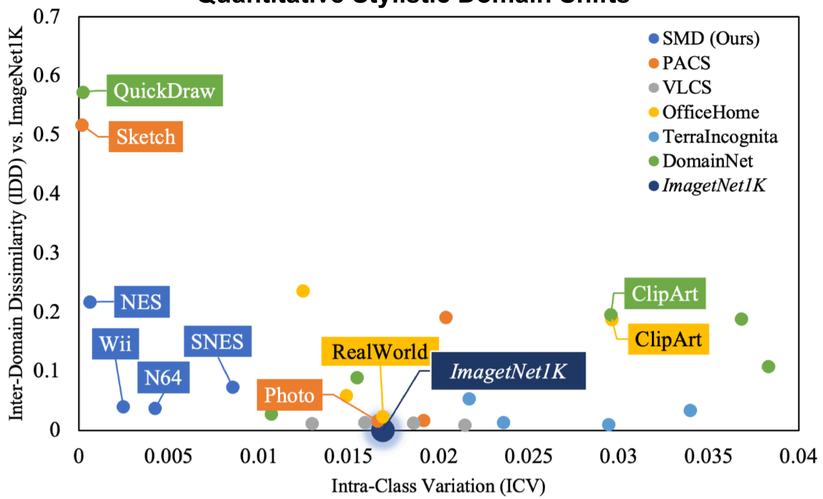
$$\mathcal{L}_S = \frac{1}{N_{\mathrm{SMD}}} \sum_{x_j, y_j \in D_{\mathrm{SMD}}} \mathcal{L}_{\mathrm{CrossEnt}}(f_S(x_j), y_j)$$

$$\mathcal{L}_{\mathrm{ERM}} = \frac{1}{N-1} \sum_{x_i, y_i \in D_{tr}} \mathcal{L}_{\mathrm{CrossEnt}}(f(x_i), y_i)$$

$$\mathcal{L}_{\mathrm{JS}} = D_{\mathrm{JS}}(f_S(x_i), f(x_i))$$

$$\mathcal{L}_{\mathrm{SMOS}} = \mathcal{L}_S + \mathcal{L}_{\mathrm{ERM}} + \lambda \mathcal{L}_{\mathrm{JS}}$$

Quantitative Stylistic Domain Shifts



- Domains of similar styles appear close to each other.
- Abstract-style domains are placed high vertically.
- ImageNet1K's IDD against itself is 0.
- The domains of SMD are dissimilar from all stylistic domains of involved DG benchmarks as well as ImageNet1K.

Domain Generalization Results

Method	PACS	VLCS	OfficeHome	TerraIncognita	DomainNet	Avg.
Mixstyle	85.2 ±0.3	77.9 ±0.5	60.4 ±0.3	44.0 ±0.7	34.0 ±0.1	60.3
GroupDRO	84.4 ±0.8	76.7 ±0.6	66.0 ±0.7	43.2 ±1.1	33.3 ±0.2	60.7
IRM	83.5 ±0.8	78.5 ±0.5	64.3 ±2.2	47.6 ±0.8	33.9 ±2.8	61.6
ARM	85.1 ±0.4	77.6 ±0.3	64.8 ±0.3	45.5 ±0.3	35.5 ±0.2	61.7
VREx	84.9 ±0.6	78.3 ±0.2	66.4 ±0.6	46.4 ±0.6	33.6 ±2.9	61.9
CDANN	82.6 ±0.9	77.5 ±0.1	65.8 ±1.3	45.8 ±1.6	38.3 ±0.3	62.0
DANN	83.6 ±0.4	78.6 ±0.4	65.9 ±0.6	46.7 ±0.5	38.3 ±0.1	62.6
RSC	85.2 ±0.9	77.1 ±0.5	65.5 ±0.9	46.6 ±1.0	38.9 ±0.5	62.7
MTL	84.6 ±0.5	77.2 ±0.4	66.4 ±0.5	45.6 ±1.2	40.6 ±0.1	62.9
Mixup	84.6 ±0.6	77.4 ±0.6	68.1 ±0.3	47.9 ±0.8	39.2 ±0.1	63.4
MLDG	84.9 ±1.0	77.2 ±0.4	66.8 ±0.6	47.7 ±0.9	41.2 ±0.1	63.6
ERM	84.2 ±0.1	77.3 ±0.1	67.6 ±0.2	47.8 ±0.6	44.0 ±0.1	64.2
SagNet	86.3 ±0.2	77.8 ±0.5	68.1 ±0.1	48.6 ±1.0	40.3 ±0.1	64.2
CORAL	86.2 ±0.3	78.8 ±0.6	68.7 ±0.3	47.6 ±1.0	41.5 ±0.1	64.5
CCFP	86.6 ±0.2	78.9 ±0.3	68.9 ±0.1	48.6 ±0.4	41.7 ±0.0	64.8
MIRO	85.4 ±0.4	79.0 ±0.0	70.5 ±0.4	50.4 ±1.1	44.3 ±0.2	65.9
SMOS⁻(ours)	88.7 ±0.2	79.7 ±0.1	71.0 ±0.0	**55.5** ±0.8	44.5 ±0.0	67.9
SMOS⁺(ours)	**89.4** ±0.3	**79.8** ±0.1	**71.6** ±0.1	55.4 ±0.4	**45.3** ±0.0	**68.3**

Method	PACS				OfficeHome			DomainNet			
	ACS → Photo	PAS → Cartoon	PAC → Sketch	Avg.	ACP → RealWorld	RAP → Clipart	Avg.	CSIPQ → Real	RSIPQ → Clipart	CRSIP → Quickdraw	Avg.
ERM	97.0±0.1	79.8±1.0	74.7±0.7	84.2	77.4±0.2	53.2±1.2	67.6	60.9±0.1	50.8±0.2	10.5±0.1	44.0
MIRO	97.3±0.2	80.5±0.5	75.9±1.0	85.4	**81.1**±0.4	53.8±0.8	70.5	63.4±0.2	54.2±0.8	11.3±0.0	44.3
SMOS⁻(ours)	**98.1**±0.0	**84.8**±0.2	81.3±0.4	88.7	80.7±0.3	57.5±0.3	71.0	63.9±0.3	60.3±0.1	13.9±0.0	44.5
SMOS⁺(ours)	**98.1**±0.1	**84.8**±0.3	**83.2**±0.6	**89.4**	80.8±0.4	**58.6**±0.4	**71.6**	**64.0**±0.2	**62.3**±0.0	**14.7**±0.0	**45.3**

- All are using the ResNet50 backbone, average of 3 trials.
- **SMOS⁻** - Precursor f_S w/o pre-training; **SMOS⁺** - f_S pre-trained w/ ImageNet1K.

Method	Precursor Train. Data $\mathcal{D}_S$	Benchmark Avg. PACS	Benchmark Avg. OfficeHome
ERM	-	84.2	67.6
MIRO	-	85.4	70.5
SMOS⁻	DomainNet	87.7	65.2
	Places365	87.4	66.5
	SMD (ours)	88.7	71.0
SMOS⁺	DomainNet	87.9	70.1
	Places365	87.5	66.5
	SMD (ours)	**89.4**	**71.6**

Ablation study of using replacement precursor data instead of SMD

Pair-wise IDDs when testing on the Sketch domain of PACS by multiple DG methods