



A Short Tutorial on Domain Generalization

Christos Diou & Aristotelis Ballas

Department of Informatics and Telematics
Harokopio University of Athens, Greece

Outline

Part 1: Introduction to DG

Part 2: Overview of representative DG methods

Part 3: DG datasets and benchmarks

Part 4: DG for biomedical signals: The BioDG benchmark

Part 5: Conclusions

Part 1: Introduction to the Domain Generalization problem



(A) **Cow: 0.99**, Pasture: 0.99, Grass: 0.99, No Person: 0.98, Mammal: 0.98



(B) No Person: 0.99, Water: 0.98, Beach: 0.97, Outdoors: 0.97, Seashore: 0.97



(C) No Person: 0.97, **Mammal: 0.96**, Water: 0.94, Beach: 0.94, Two: 0.94



Predicted: **wolf**
True: **wolf**



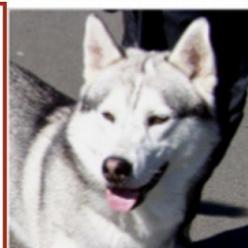
Predicted: **husky**
True: **husky**



Predicted: **wolf**
True: **wolf**



Predicted: **wolf**
True: **husky**



Predicted: **husky**
True: **husky**



Predicted: **wolf**
True: **wolf**

[source](#)

Background (1)

Supervised statistical machine learning methods aim at learning a function

$$\hat{y} = f(x; \theta)$$

which provides an estimate/probabilities \hat{y} of the true value of the label(s)/target(s) y , given a previously unseen data sample x

In our previous example, $x =$



$$\hat{y} = \begin{array}{l} \text{(A) Cow: 0.99, Pasture:} \\ 0.99, \text{ Grass: 0.99, No Person:} \\ 0.98, \text{ Mammal: 0.98} \end{array}$$

Background (2)

Let P_{XY} on $\mathcal{X} \times \mathcal{Y}$ be the data distribution of the training data. A fundamental assumption of most ML methods is that the unseen (test) data follow the same distribution (i.i.d. assumption).

However, in our example:



$$P_{XY}^{(1)} \neq P_{XY}^{(2)}$$



(1): Cows + grass

(2): Cows + beach

Domains

Let's call each $P_{XY}^{(i)}$, $i = 1, \dots, K$ a domain

Example domains:

- Cows with grass
- Cows with beach
- Wolves with snow
- Wolves with gravel

Domains

Let's call each $P_{XY}^{(i)}$, $i = 1, \dots, K$ a domain

Example domains:

- Cows with grass
- Cows with beach
- Wolves with snow
- Wolves with gravel

Problem 1: How can we train a model to recognize Cows with grass and beach?

Solution 1: Just train a model with pictures from both domains

Domains

Let's call each $P_{XY}^{(i)}$, $i = 1, \dots, K$ a domain

Example domains:

- Cows with grass
 - Cows with beach
 - Wolves with snow
 - Wolves with gravel
- Dataset distribution can be device-specific
 - X-rays of device A
 - X-rays of device B
 - Dataset distribution may vary across individuals
 - Blood sample from person 1
 - Blood sample from person 2
 - Privacy / legal / technical restrictions

Problem 2: In many real-world scenarios it is difficult to have access to large volumes of data from all domains during training

Transfer learning

(for the same labels)

Given a model trained on a **source domain**: $P_{XY}^{(S)}$

Train a model on the **target domain**: $P_{XY}^{(T)}$ by performing a (hopefully small) number of training iterations

Requirements:

- Large numbers of labeled data on source domain
- Availability of labeled data on target domain



Domain adaptation

Train a model on the **source domain(s)**: $P_{XY}^{(S)}$

Use unlabeled data (or very small number of labeled data) to adapt the model to the **target domain**, $P_{XY}^{(T)}$

Requirements:

- Unlabelled data (or small number of labeled data) from the target domain

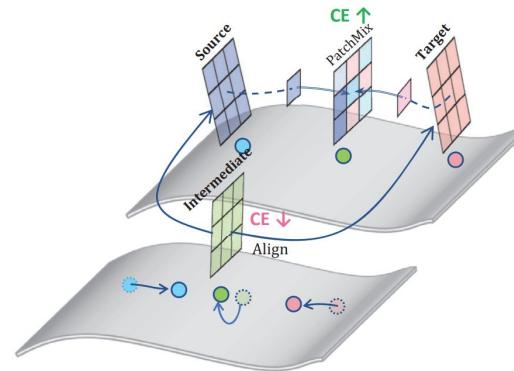


Image from: Zhu, J. et al. "Patch-Mix Transformer for Unsupervised Domain Adaptation: A Game Perspective.", CVPR 2023

Domain Generalization (1)

In this scenario we don't have access to **any** data from the target domain during training.

Our goal is to train a model using data from the source domain(s) that are effective in the unseen target domain(s)

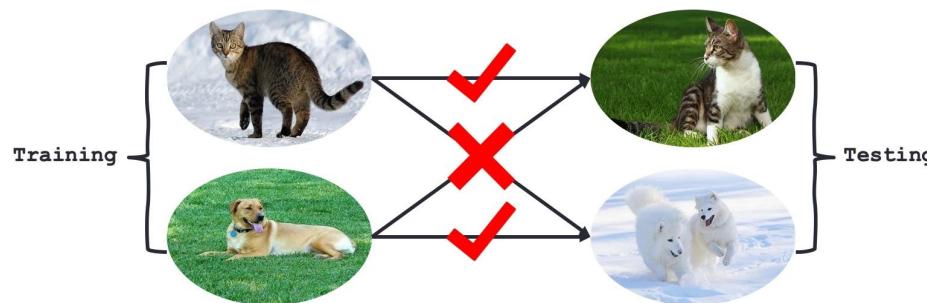
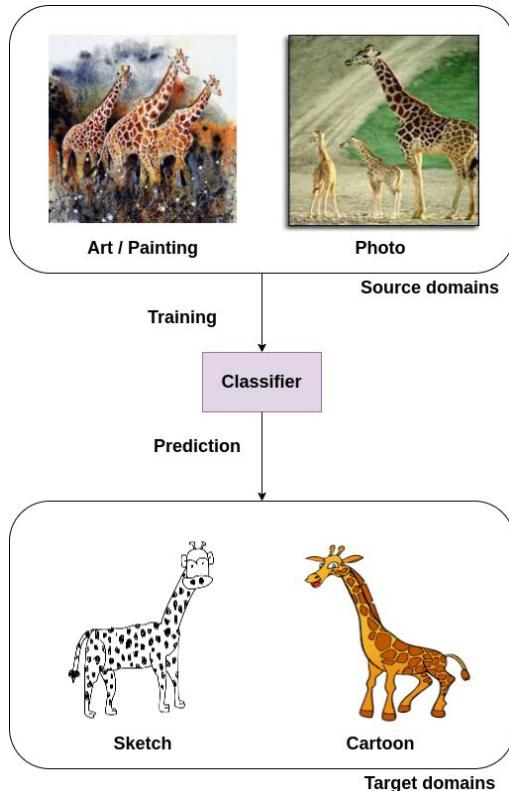


Image: Zhang et al., NICO++: Towards Better Benchmarking for Domain Generalization
<https://github.com/xxgege/NICO-plus>

Domain Generalization (2)



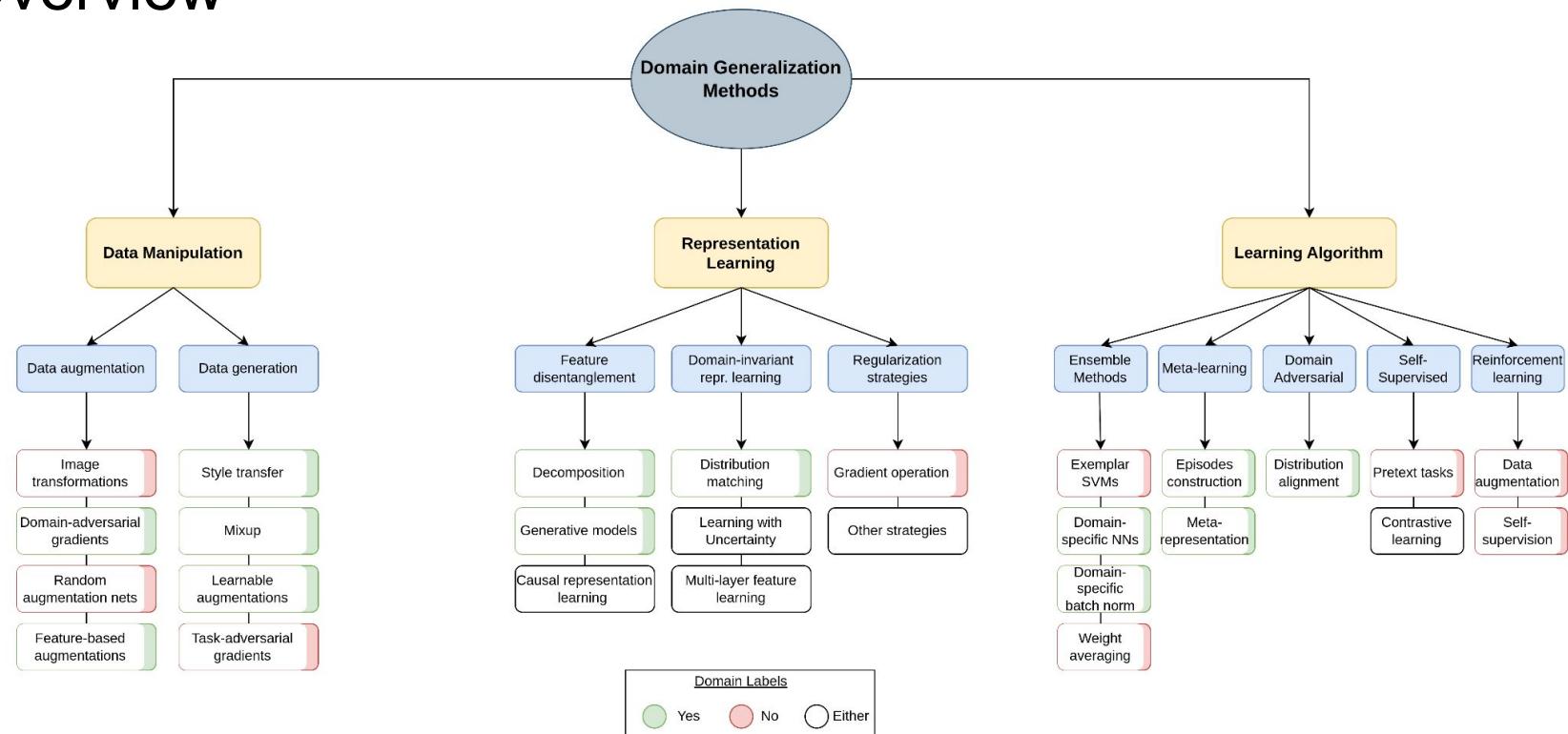
Observations:

- Some features are common (invariant) across domains
- These may have a different appearance, depending on the domain

Goal: Learn the invariant features, not spurious domain features

- **Multi-source DG:** Multiple training domains, access to domain labels (we know the domain of each training sample)
- **Single-source DG:** Single training domain and/or no-access to domain labels

Overview



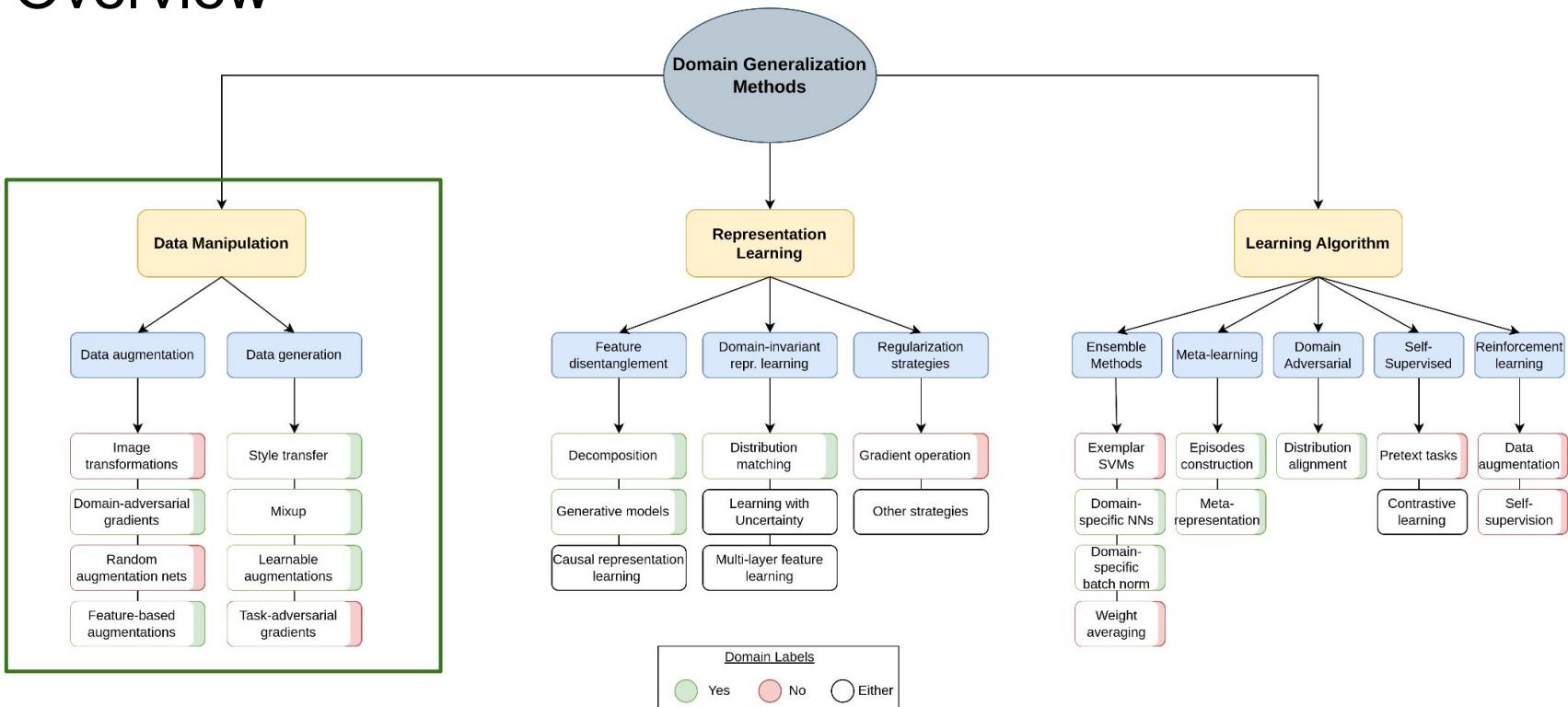
Domain Labels

● Yes ● No ● Either

This taxonomy is a variation / extension of the ones on Wang, J. et al. "Generalizing to unseen domains: A survey on domain generalization." IEEE Transactions on Knowledge and Data Engineering (2022), and Zhou, Kaiyang, et al. "Domain generalization: A survey." IEEE Transactions on Pattern Analysis and Machine Intelligence (2022).

Part 2: Overview of representative DG methods

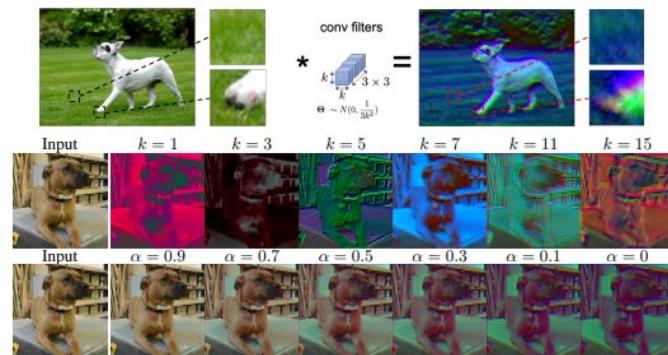
Overview



This taxonomy is a variation / extension of the ones on Wang, J. et al. "Generalizing to unseen domains: A survey on domain generalization." IEEE Transactions on Knowledge and Data Engineering (2022), and Zhou, Kaiyang, et al. "Domain generalization: A survey." IEEE Transactions on Pattern Analysis and Machine Intelligence (2022).

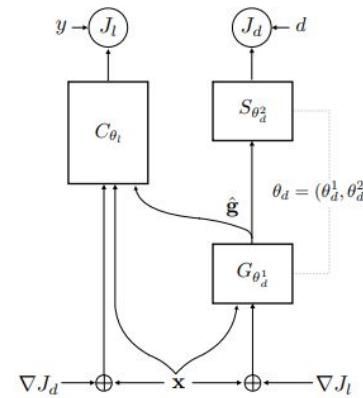
Data Augmentation

RandConv: Random Convolutions



RandConv uses multi-scale random-convolutions to generate images with random texture while maintaining global shapes

CrossGrad: Cross-Gradient

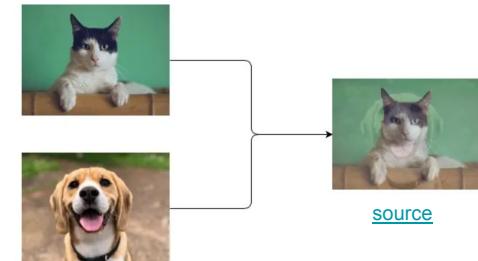


Adversarially augment data via gradient training to generate data that share the same label y but different domain label d .

Data Generation

Mixup: Improve model robustness by training the network on convex combinations of pairs of examples and their labels.

$$\begin{aligned}\tilde{x} &= \lambda x_i + (1 - \lambda)x_j, \\ \tilde{y} &= \lambda y_i + (1 - \lambda)y_j, \\ \text{where } \lambda &\in [0, 1]\end{aligned}$$



MixStyle: Mix the styles of training instances in each mini-batch to increase the domain diversity of the source domains

$$\begin{aligned}MixStyle(x) &= \gamma_{mix} \frac{x - \mu(x)}{\sigma(x)} + \beta_{mix} \\ \gamma_{mix} &= \lambda \sigma(x) + (1 - \lambda) \sigma(\tilde{x}), \\ \beta_{mix} &= \lambda \mu(x) + (1 - \lambda) \mu(\tilde{x}), \\ \text{where } \lambda &\sim B(a, a)\end{aligned}$$

$$x = [x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6]$$

$$\tilde{x} = [x_5 \ x_6 \ x_4 \ x_3 \ x_1 \ x_2]$$

(a) Shuffling batch w/ domain label

$$x = [x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6]$$

$$\tilde{x} = [x_6 \ x_1 \ x_5 \ x_3 \ x_2 \ x_4]$$

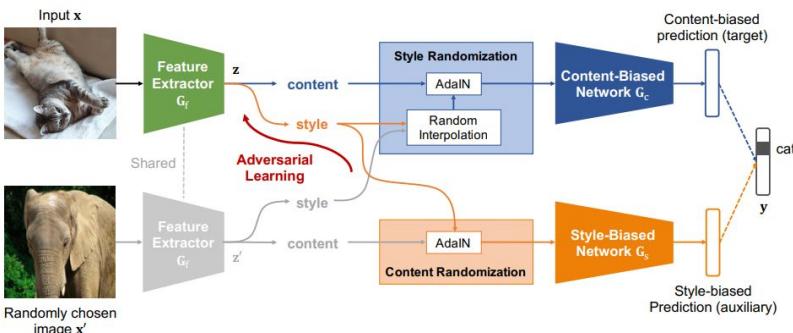
(b) Shuffling batch w/ random shuffle

Zhang et al. "Mixup: Beyond empirical risk minimization". arXiv 2017

Zhou et al. "Domain generalization with mixstyle". ICLR 2021

Feature Disentanglement

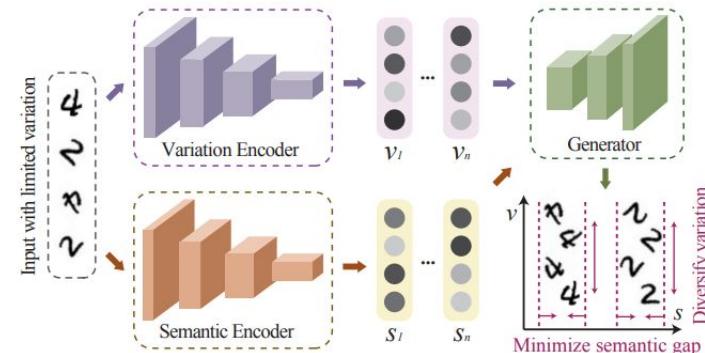
SagNet: Style Agnostic Networks



SagNets disentangle style encodings from class categories to prevent style biased predictions and focus more on the contents.

Nam et al. "Reducing domain gap by reducing style bias". CVPR 2021.

DDG: Disentanglement-constrained DG

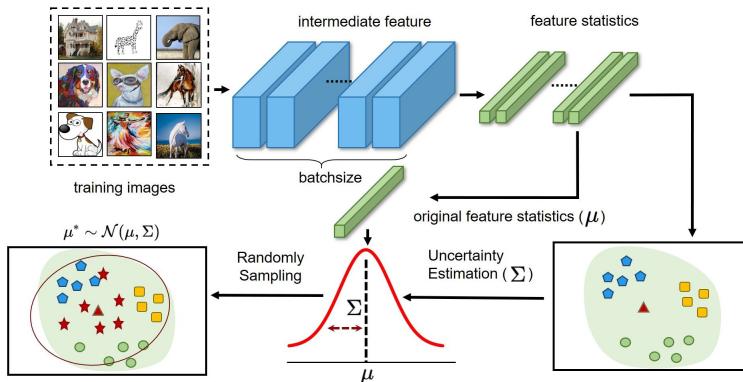


DDG seeks to minimize the semantic difference of the generated samples from the same class while diversifying the variation across source domains.

Zhang et al. "Towards Principled Disentanglement for Domain Generalization". CVPR 2022

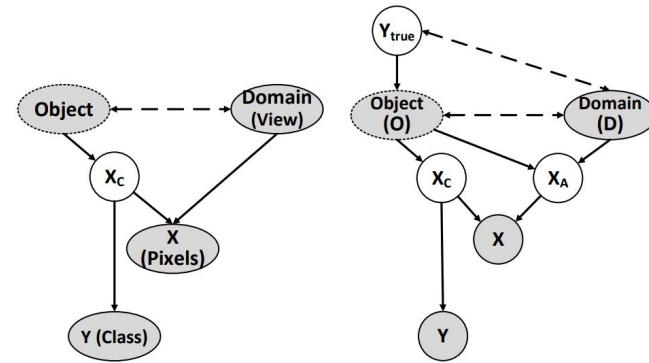
Domain Invariant Representation Learning

DSU: Domain Shifts with Uncertainty



Under the assumption that the feature statistics of each source data distribution follows a multivariate Gaussian distribution, DSU models the uncertainty of domain shifts with synthesized feature statistics during training.

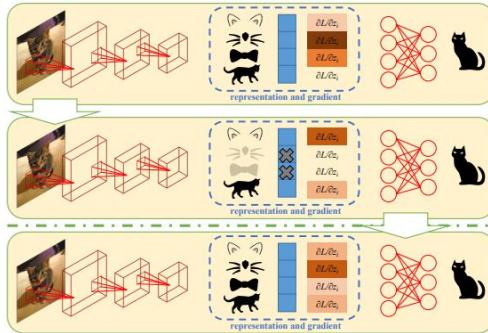
MatchDG: Build Representations via Causal Matching



MatchDG utilizes contrastive learning to build a representation such that inputs sharing the same causal features are closer to one another.

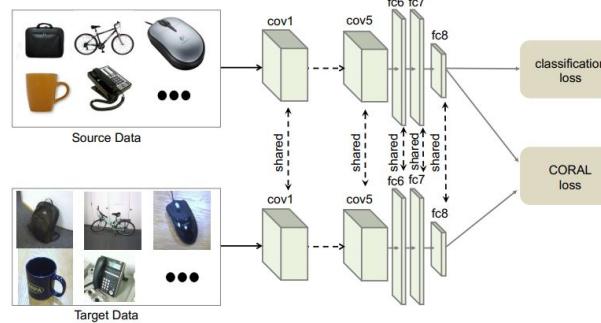
Regularization Strategies

RSC: Representation Self-Challenging



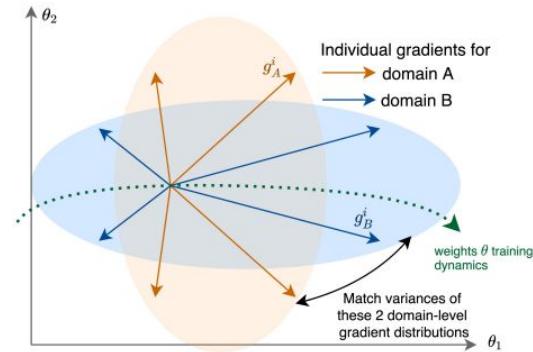
RSC mutes the feature representations associated with the highest gradient, such that the network is forced to predict the labels through other features

CORAL: Correlation Alignment



CORAL focuses on minimizing the distance between the second-order statistics (covariances) between domains.

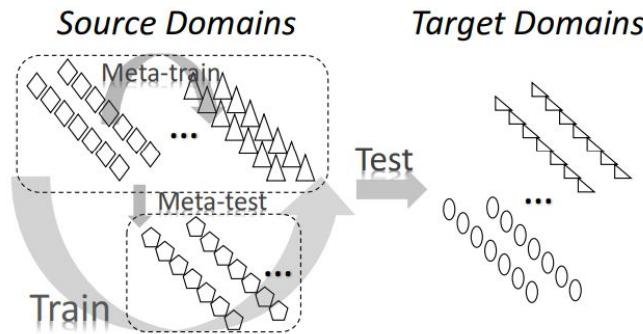
FISHR: Regularization with Fisher Information



FISHR introduces a regularization term that matches the domain-level variances of gradients across training domains

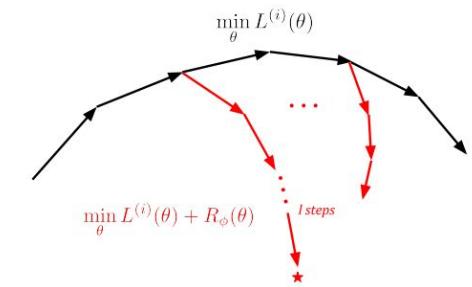
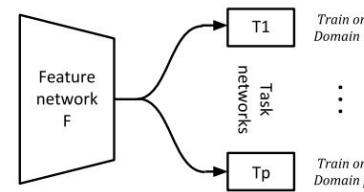
Meta-Learning

MLDG: Meta-Learning for DG



In each mini-batch the model is evaluated on meta-test domains. This meta-test evaluation simulates testing on new domains with different statistics, thus improving robustness.

MetaReg: Meta-Learning for regularization



The goal is to learn a regularizer model from multiple domain-specific networks in order to update the parameters of the base classifier.

Ensemble Methods

MIRO: Mutual Information Regularization with Oracle

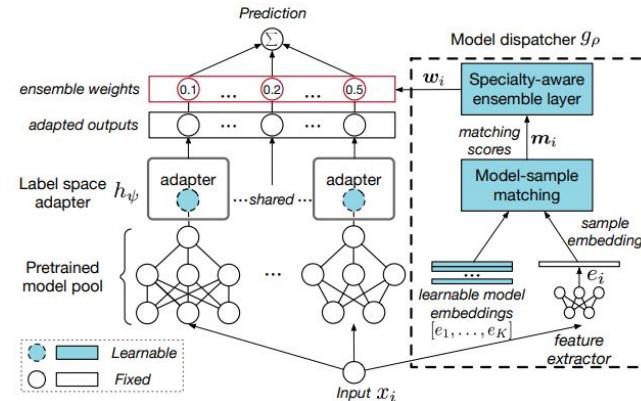
Algorithm 1: Mutual Information Regularization with Oracle (MIRO)

```
Input: feature extractor  $f$ , classifier  $g$ , mean encoder  $\mu$ , variance encoder  $\Sigma$ , regularization coefficient  $\lambda$ , batch size  $N$ .  
Init: initialize  $f$  to pre-trained feature extractor  $f^0$ .  
Output: learned feature extractor  $f$  and learned classifier  $g$ .  
for sampled mini-batch  $(\mathbf{x}, \mathbf{y})$  do  
     $\mathbf{z}_f = f(\mathbf{x})$   
     $\mathbf{z}_{f^0} = f^0(\mathbf{x})$   
     $\mathcal{L} = \frac{1}{N} \sum_i^N [\text{CrossEntropy}(g(z_f^i), y^i) + \lambda (\log |\Sigma(z_f^i)| + \|z_{f^0}^i - \mu(z_f^i)\|_{\Sigma(z_f^i)^{-1}}^2)]$   
    update  $f, g, \mu, \Sigma$  to minimize  $\mathcal{L}$   
end
```

In MIRO, a large pre-trained model is used as an “Oracle” model and is leveraged to guide a classifier by minimizing the Mutual Information between their representations.

Cha et al. "Domain generalization by mutual-information regularization with pre-trained models." ECCV 2022.

SIMPLE: Specialized Model-Sample Matching

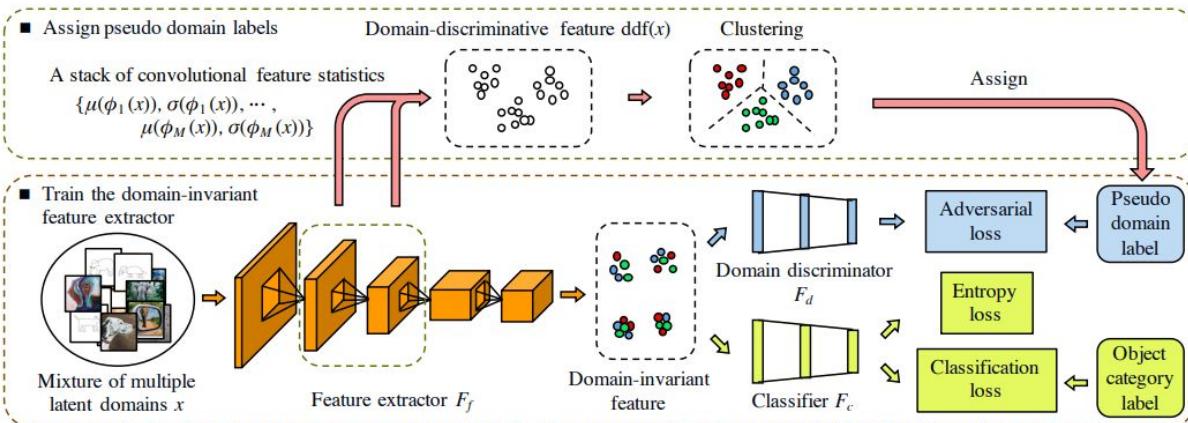


SIMPLE utilizes a pool of fixed pretrained models and selects the best subset for each downstream task/dataset.

Li et al. "SIMPLE: Specialized Model-Sample Matching for Domain Generalization." ICLR 2022.

Domain Adversarial Learning

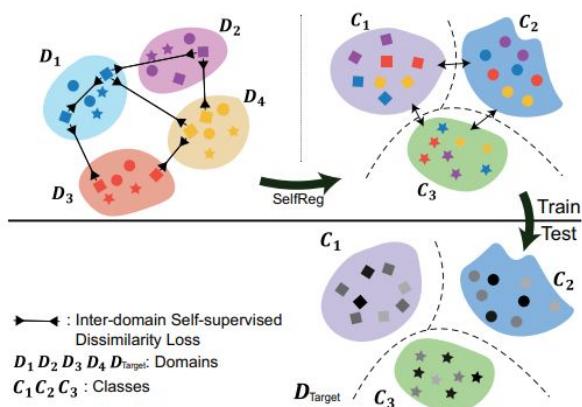
MMLD: Mixture of Multiple Latent Domains



MMLD iteratively assigns pseudo domain labels by clustering domain discriminative features extracted from lower layers of the feature extractor, and trains the domain-invariant feature extractor via adversarial learning.

Self-Supervised Learning

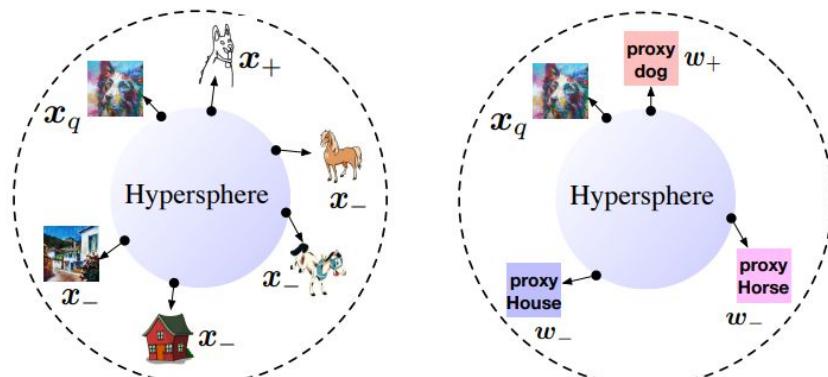
SelfReg: Self-Supervised Contrastive Regularization



SelfReg maps the latent representations of same-same class samples close together to learn domain-invariant properties.

Kim et al. "Selfreg: Self-supervised contrastive regularization for domain generalization". ICCV 2021.

PCL: Proxy-Based Contrastive Learning



The PCL paper claims that aligning positive sample-to-sample pairs hinders the model generalization and propose aligning proxy-to-sample representations.

Yao et al. "Pcl: Proxy-based contrastive learning for domain generalization". CVPR 2022.

Part 3: Datasets and Benchmarks

Benchmarking DG Algorithms

When evaluating a model in a DG setting, there are 3 things to consider:

- 1) Which dataset should I use?
- 2) Is there a codebase I can use or should I just write my own scripts?
- 3) How do I evaluate my model?

Although there is no ‘gold standard’, there are some widely adopted settings in the community.

DG Applications & Datasets

There are several applications and datasets which can be used for evaluating DG algorithms. To name a few:

Object Recognition

- PACS
- VLCS
- OfficeHome
- DomainNet
- WILDS
- TerraInc
- NICO++



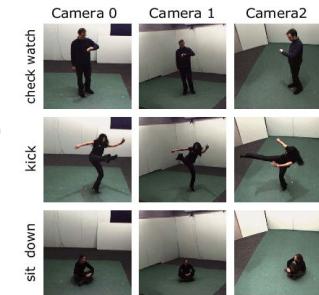
PACS

Action Recognition

- IXMAS
- UCF-HMDB

Sentiment Classification

- Amazon Reviews



IXMAS

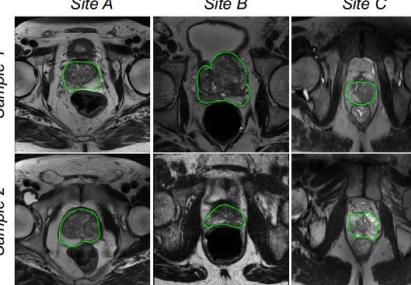
Medical Imaging

- Multi-site Prostate MRI Seg
- Chest X-rays

Reinforcement Learning

- Coinrun
- OpenAI Procgen

MRI Seg

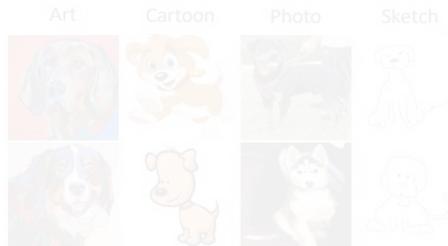


DG Applications & Datasets

There are several applications and datasets which can be used for evaluating DG algorithms. To name a few:

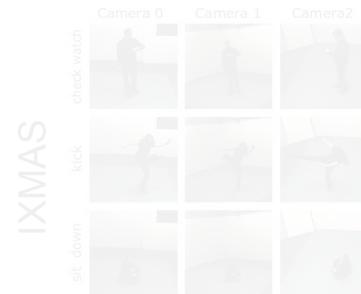
Object Recognition

- PACS
- VLCS
- OfficeHome
- DomainNet
- WILDS
- TerraInc
- NICO++



Action Recognition

- ICMAS
- Amazon Reviews



MRI Seg



Medical Imaging

- Multi-Site Prostate MRI Seg

Reinforcement Learning

- Coinrun
- OpenAI Procgen

DG Benchmarks & Codebases (1)

Lately, the **DomainBed** benchmark for domain generalization image classification has been generally recognized and used by researchers.

The currently available datasets are:

- RotatedMNIST
- ColoredMNIST
- VLCS
- PACS
- Office-Home
- TerraIncognita
- DomainNet
- A SVIRO
- WILDS FMoW
- WILDS Camelyon17
- Spawrious

Algorithm	CMNIST	RMNIST	VLCS	PACS	Office-Home	TerraInc	DomainNet	Avg
ERM	34.2 ± 1.2	98.0 ± 0.0	76.8 ± 1.0	83.3 ± 0.6	67.3 ± 0.3	46.2 ± 0.2	40.8 ± 0.2	63.8
IRM	36.3 ± 0.4	97.7 ± 0.1	77.2 ± 0.3	82.9 ± 0.6	66.7 ± 0.7	44.0 ± 0.7	35.3 ± 1.5	62.9
DRO	32.2 ± 3.7	97.9 ± 0.1	77.5 ± 0.1	83.1 ± 0.6	67.1 ± 0.3	42.5 ± 0.2	32.8 ± 0.2	61.8
Mixup	31.2 ± 2.1	98.1 ± 0.1	78.6 ± 0.2	83.7 ± 0.9	68.2 ± 0.3	46.1 ± 1.6	39.4 ± 0.3	63.6
MLDG	36.9 ± 0.2	98.0 ± 0.1	77.1 ± 0.6	82.4 ± 0.7	67.6 ± 0.3	45.8 ± 1.2	42.1 ± 0.1	64.2
CORAL	29.9 ± 2.5	98.1 ± 0.1	77.0 ± 0.5	83.6 ± 0.6	68.6 ± 0.2	48.1 ± 1.3	41.9 ± 0.2	63.9
MMD	42.6 ± 3.0	98.1 ± 0.1	76.7 ± 0.9	82.8 ± 0.3	67.1 ± 0.5	46.3 ± 0.5	39.3 ± 0.9	64.7
DANN	29.0 ± 7.7	89.1 ± 5.5	77.7 ± 0.3	84.0 ± 0.5	65.5 ± 0.1	45.7 ± 0.8	37.5 ± 0.2	61.2
C-DANN	31.1 ± 8.5	96.3 ± 1.0	74.0 ± 1.0	81.7 ± 1.4	64.7 ± 0.4	40.6 ± 1.8	38.7 ± 0.2	61.1

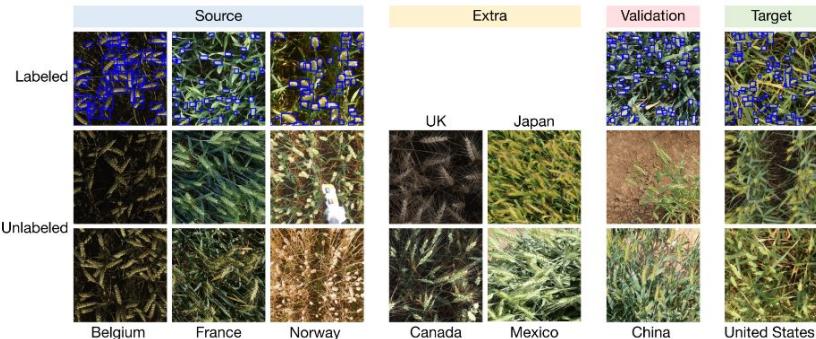
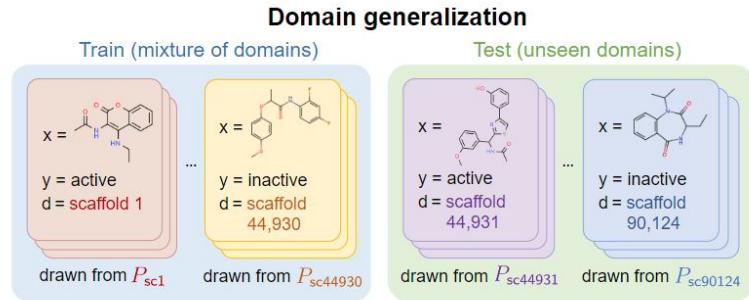
Gulrajani et al. "In search of lost domain generalization." arXiv 2020.
Code: <https://github.com/facebookresearch/DomainBed>

DG Benchmarks & Codebases (2)

WILDS is a collection of benchmark datasets that represent distribution shifts faced in the wild.

WILDS provides the following datasets:

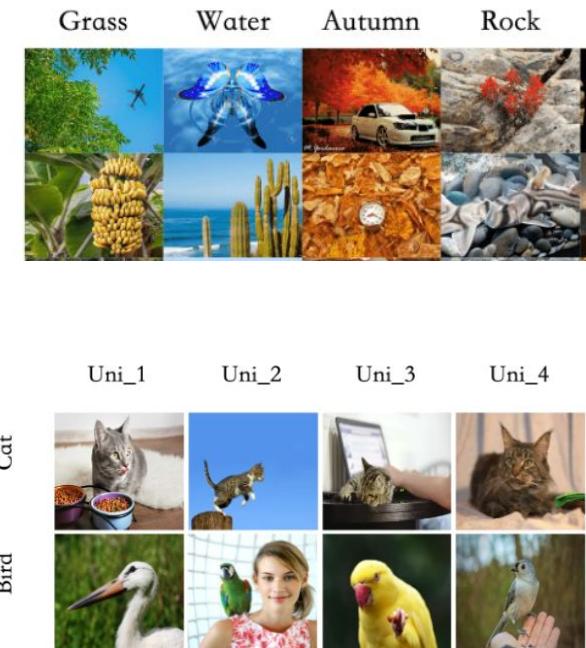
- iWildCAM
 - Camelyon17
 - VLCS
 - RxRx1
 - OGB-MolPCBA
 - GlobalWheat
 - DomainNet
 - CivilComments
 - FMoW
 - PovertyMap
 - Amazon
 - Py150



Koh et al. "Wilds: A benchmark of in-the-wild distribution shifts." PMLR 2021.
 Code: <https://github.com/p-lambda/wilds/>

DG Benchmarks & Codebases (3)

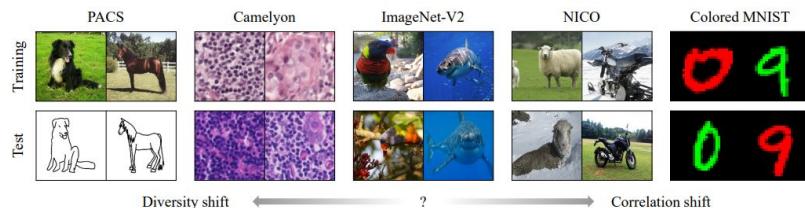
NICO++ is specifically designed for OOD (Out-of-Distribution) image classification. It simulates a real world setting that the testing distribution may induce arbitrary shifting from the training distribution



DG Benchmarks & Codebases (4)

Building on DomainBed, additional benchmarks have been proposed.

OoD-Bench adopts similar datasets to DomainBed and splits the benchmark into two dimensions, focusing on data suffering from either *diversity* or *correlation* shift.



Ye et al. "Ood-bench: Quantifying and understanding two dimensions of out-of-distribution generalization." CVPR 2022.

Code: <https://github.com/xxgege/NICO-plus>

DeepDG is proposed as a simplified version of DomainBed with additional features.

Specifically it:

- Avoids huge hyperparameter tuning
- Provides a more friendly interface and
- Has better customization

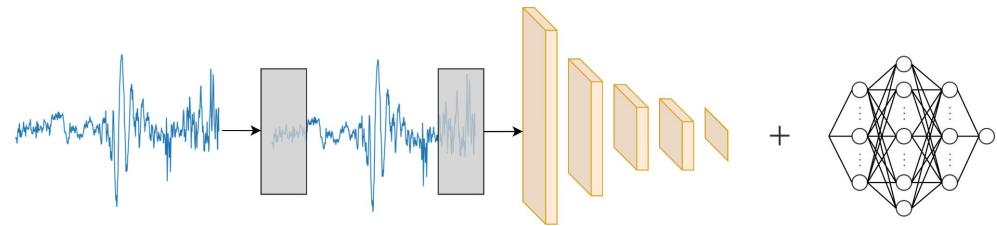
J. Wang et al., "Generalizing to Unseen Domains: A Survey on Domain Generalization". IEEE TKDE 2022

Code: <https://github.com/jindongwang/transferlearning>

Part 4: Application to biomedical signals and the BioDG benchmark

Domain Generalization in Biosignal Classification

Up until now, we have presented an abundance of methods / datasets / benchmarks for the DG problem in images.



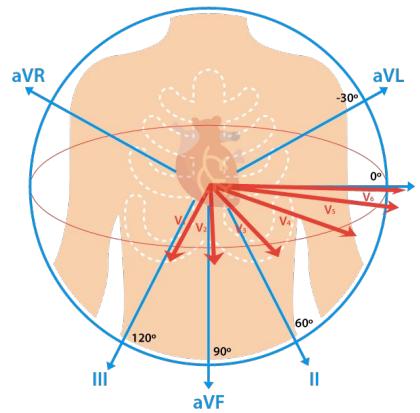
*The DG problem is somewhat under-represented in other settings such as **Biosignal Classification**.*

Distributional shifts can occur due to:

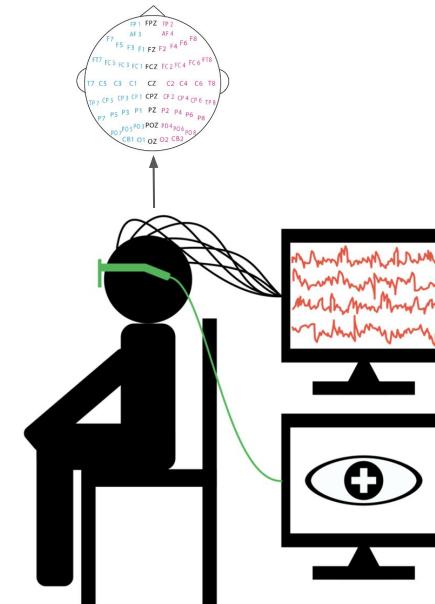
- Different hospital equipment
- Different or evolving hospital procedures
- Different testing environments
- Different physiology of patients

Biosignal Classification in General

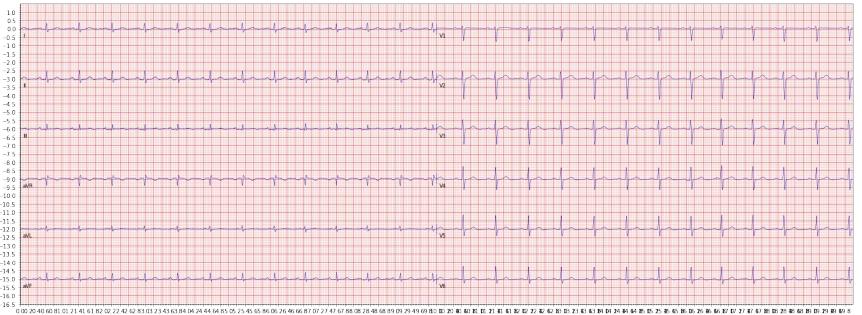
We consider domain generalization for the **12-lead ECG** and **62-channel EEG** classification task and aim to improve the ability of a model $M\theta : X \rightarrow Y$, to detect domain-invariant features of a class.



12-Lead ECG. [source](#)



62-channel EEG. [source](#)



Typical 12-lead electrocardiogram.



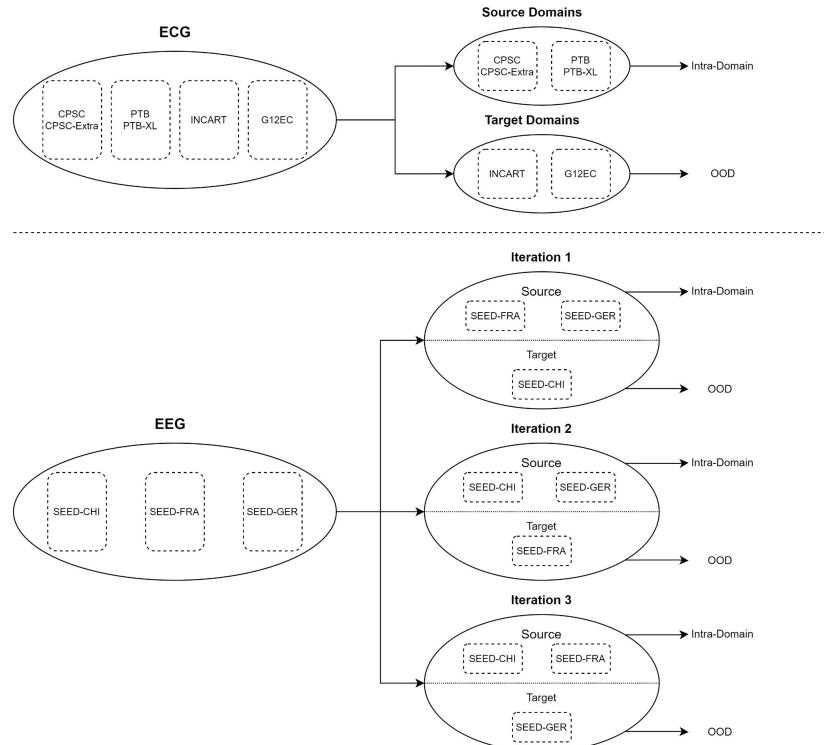
Spike and wave discharges monitored EEG

The BioDG Benchmark (1)

We introduce a novel DG open-source evaluation benchmark, namely **BioDG** for the biosignal classification setting.

Specifically, BioDG:

- Proposes a DG setup for ECG and EEG classification, using datasets from PhysioNet and SEED
- Adapts state-of-the-art image DG algorithms to 1D signal classification and
- Provides researchers with the means to pour further research in biomedical DG.



The BioDG Benchmark (2)

For the DG setup we used publicly available datasets from the PhysioNet (ECG) and SEED (EEG) repositories.

ECG

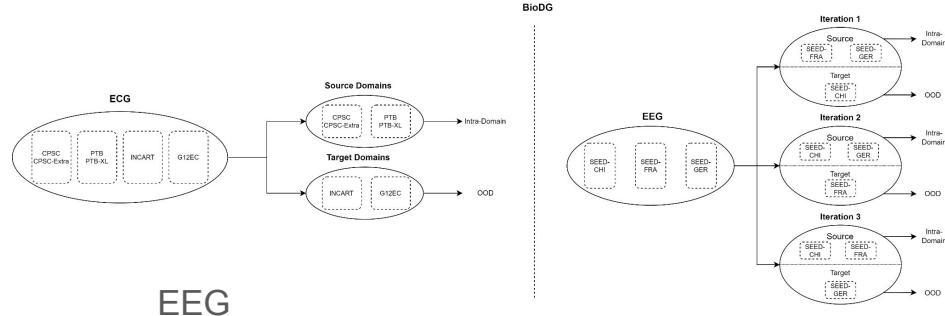
We used datasets from 4 different sources:

- CPSC, CPSC Extra
- G12EC
- INCART
- PTB, PTB-XL

The ECG signals are multi-class and multi-label (24 labels).

Task: Disease prediction

Alday et al. "Classification of 12-lead ecgs: the physionet/computing in cardiology challenge 2020." *Physiological measurement* 2020.



EEG

We used the datasets from 3 different populations:

- SEED-CHI
- SEED-FRA
- SEED-GER

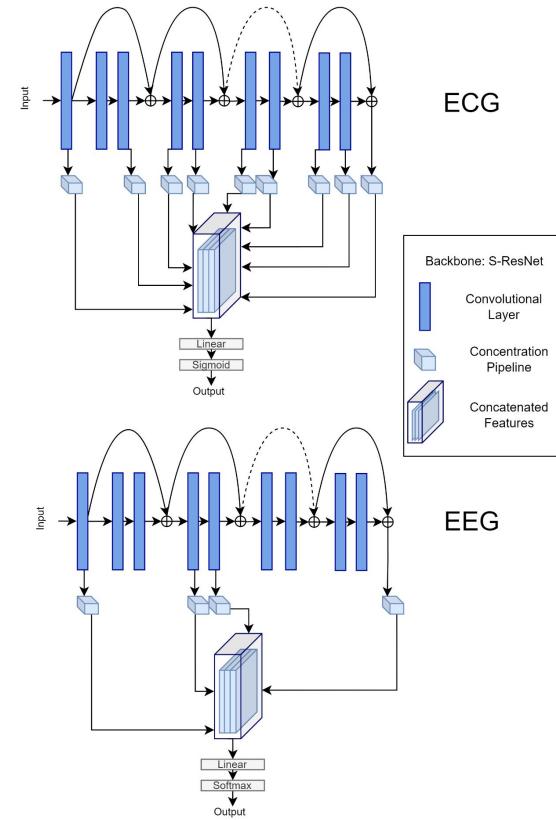
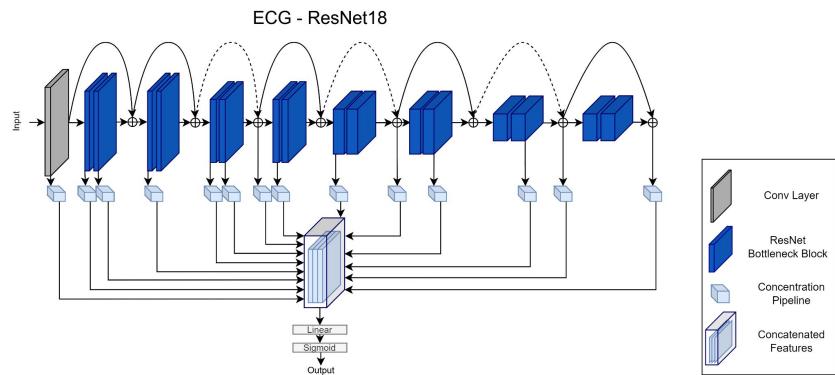
The EEG signals are multi-class (3 labels) problem.

Task: Emotion classification

Zheng et al, "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks". IEEE TAE 2015.

The BioDG Benchmark (3)

We also propose an alternative architecture to tackle DG in 1D biosignal classification, which utilizes compressed intermediate features of a backbone CNN model.



The BioDG Benchmark (4)

BioDG provides evaluation metrics for both
Intra-Distribution and Out-of-Distribution data splits.

ECG Classification Results

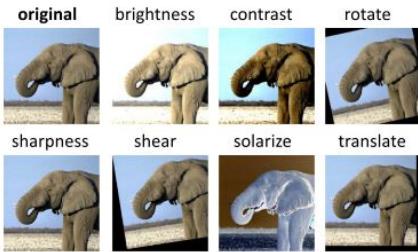
Backbone: S-ResNet Diagnosis	Intra-Distribution						OOD					
	ERM	IRM	MMD	RSC	CORAL	BioDG	ERM	IRM	MMD	RSC	CORAL	BioDG
1st degree AV block	16.14	0	7.69	15.00	10.67	61.88	7.30	0	5.84	0.51	7.38	62.24
Atrial fibrillation	55.84	18.18	23.98	36.99	24.81	87.05	32.58	12.16	13.28	12.16	13.22	62.51
Atrial flutter	19.05	0	0	13.33	0	50.00	0	0	0	0	0	10.70
Bradycardia	0	0	0	5.26	7.69	0	0	0	0	0	0	0
Complete right bundle branch block	73.38	16.16	60.41	72.58	64.19	82.46	60.15	11.50	50.90	47.75	52.64	64.55
Incomplete right bundle branch block	11.01	0	13.05	4.86	22.19	43.03	14.56	0	20.09	10.85	25.65	30.27
Left anterior fascicular block	54.67	0	50.83	52.8	53.37	70.29	31.03	0	21.01	17.16	20.84	33.29
Left axis deviation	58.10	31.15	58.34	56.26	60.56	66.87	44.98	17.79	37.75	36.09	39.16	50.42
Left bundle branch block	80.49	0	0	78.33	45.49	85.48	62.44	0	0	0	40	74.04
Low QRS voltages	0	0	0	0	0	0	0	0	0	0	0	0
Non-specific intraventricular conduction disorder	0	0	0	4.55	0	2.34	0	0	0	0	0	0
Pacing rhythm	58.62	0	37.84	43.90	48.15	78.46	0	0	0	0	0	0
Premature ventricular contractions	0	0	0	0	0	0	0	0	0	0	0	0
Prolonged PR interval	0	0	0	0	0	15.00	0	0	0	0	0	0
Prolonged QT interval	0	0	0	0	0	0	0	0	0	0	0	0
Q wave abnormal	0	0	0	0	0	11.11	0	0	0	0	0	0
Right axis deviation	5.48	0	23.27	9.52	27.92	24.53	0	0	18.58	21.54	17.35	2.84
Sinus arrhythmia	2.17	0	0	0	0	0	1.94	0	0	0	0	10.85
Sinus bradycardia	4.49	0	0	9.03	0	36.70	31.56	31.56	31.56	31.56	31.56	47.33
Sinus rhythm	84.60	80.94	80.94	82.56	80.94	91.44	38.89	0	12.54	0.62	15.17	82.68
Sinus tachycardia	27.86	0	10.37	20	15.54	79.82	1.35	0	9.98	0	8.07	1.92
Supraventricular premature beats	9.27	0	10.07	11.28	8.17	3.68	0	0	10.02	0	6.26	5.84
T wave abnormal	18.57	15.68	28.62	13.31	26.14	33.89	8.11	38.40	37.22	0	30.35	4.53
T wave inversion	0	0	0	0	0	0	0	0	0	0	0	0
# of Predicted Classes	16	5	12	17	15	18	12	5	12	9	13	15
Macro-F1 (Predicted)	41.46	9.01	38.00	17.03	37.82	53.91	31.15	7.43	32.43	14.92	31.70	36.44
Macro-F1 (All)	31.10	6.75	28.50	12.77	28.37	40.44	19.47	4.64	20.27	9.33	19.81	22.77

EEG Classification Results

Method	Intra		
	FRA-GER	CHI-GER	CHI-FRA
ERM	70.94 \pm 2.51	73.63 \pm 2.08	76.10 \pm 2.27
IRM	70.75 \pm 2.81	71.42 \pm 2.98	73.90 \pm 3.15
CORAL	71.42 \pm 1.67	73.72 \pm 1.91	74.96 \pm 2.59
MMD	71.23 \pm 2.23	74.34 \pm 1.87	76.26 \pm 1.95
RSC	70.15 \pm 2.43	72.01 \pm 2.37	76.15 \pm 2.08
BioDG	77.59 \pm 2.23	76.11 \pm 1.25	75.61 \pm 1.15
Method	OOD		
	CHI	FRA	GER
ERM	54.60 \pm 3.44	45.83 \pm 3.53	63.75 \pm 4.01
IRM	53.70 \pm 1.42	43.82 \pm 0.75	66.87 \pm 3.49
CORAL	53.18 \pm 3.29	45.00 \pm 2.51	67.04 \pm 3.27
MMD	54.17 \pm 2.68	45.80 \pm 3.46	66.04 \pm 3.39
RSC	53.86 \pm 2.55	44.37 \pm 2.99	58.79 \pm 3.93
BioDG	57.09 \pm 0.44	51.74 \pm 0.35	68.04 \pm 0.24

Data Augmentation

Typical Image Transformations



Random Convolutions

Z. Xu, D. Liu, J. Yang, C. Raffel, and M. Niethammer, "Robust and generalizable visual representation learning via random convolutions," in Proc. Int. Conf. Learn. Representations, 2021

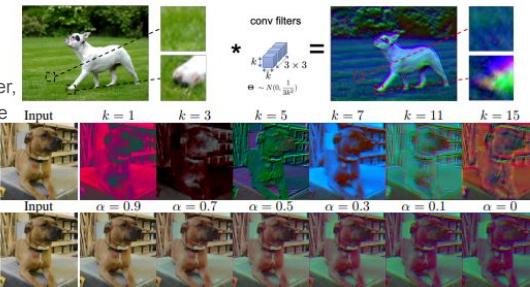
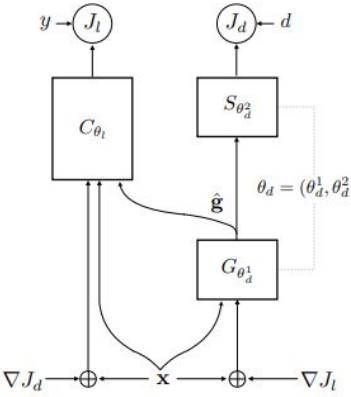


Figure 1: **Top:** Illustration that RandConv randomize local texture but preserve shapes in the image. **Middle:** First column is the input image of size 224^2 ; following columns are convolutions results using random filters of different sizes k . **Bottom:** Mixing results between an image and one of its random convolution results with different mixing coefficients α .

Adversarial data augmentation



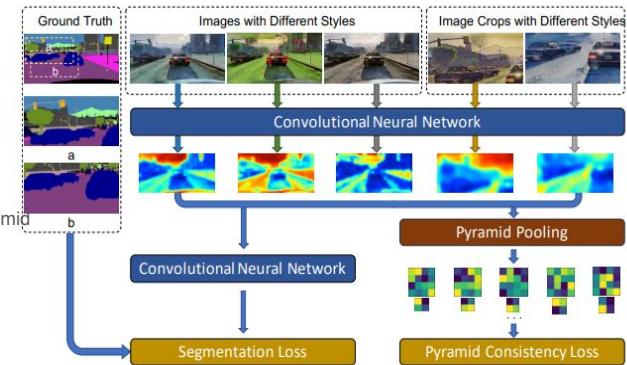
CrossGrad:
Adversarially
augment data via
gradient training

The idea is to
generate data that
share the same
label y but different
domain label d .

Shankar et al. Generalizing across Domains via Cross-Gradient Training. ICLR 2018 (CrossGrad)

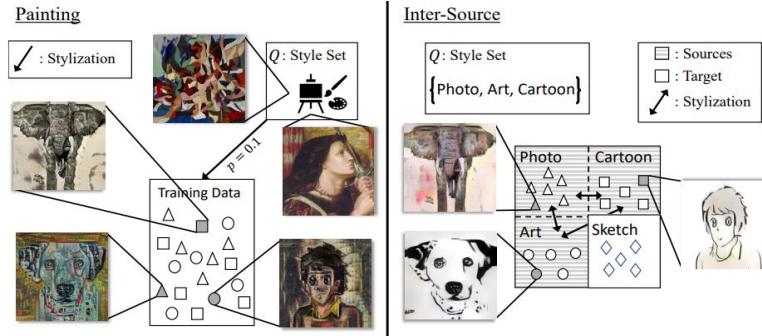
Domain Randomization

Yue et al. Domain Randomization and Pyramid Consistency: Simulation-to-Real Generalization without Accessing Target Domain Data. ICCV, 2019.



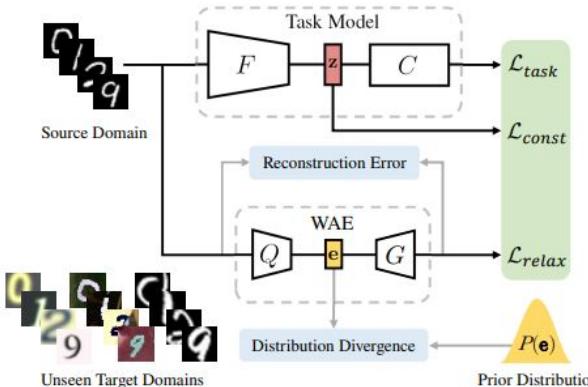
Data Generation

Image Stylization



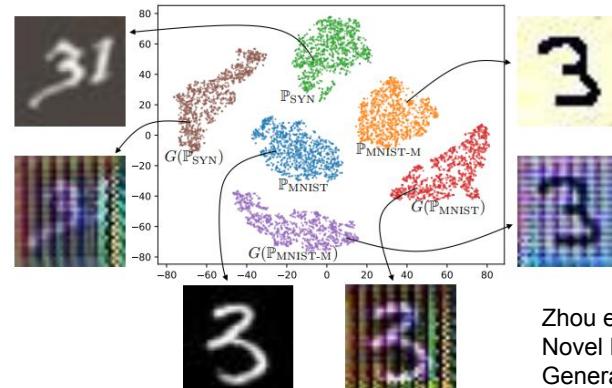
Somavarapu et al. Frustratingly Simple Domain Generalization via Image Stylization. 2020.

Image Generation with VAE



Qiao et al. Learning to Learn Single Domain Generalization. CVPR 2020.

Conditional GAN

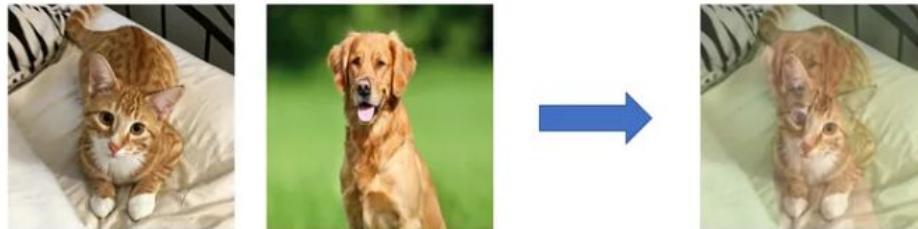


Zhou et al. Learning to Generate Novel Domains for Domain Generalization. ECCV 2020

MIXUP Methods

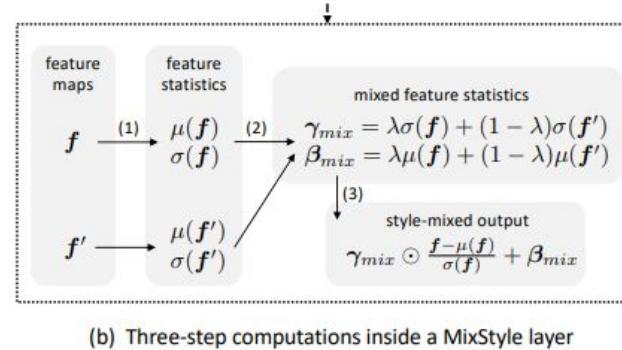
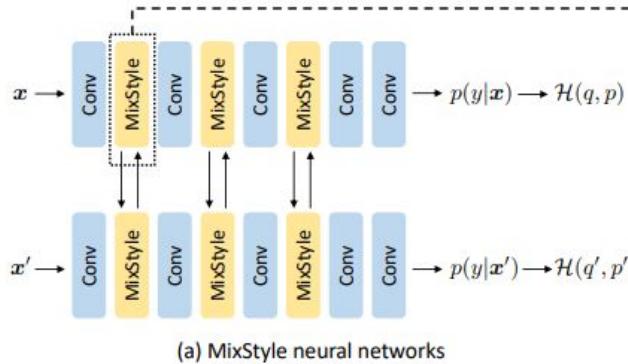
MIXUP

Photo taken from: [link](#)



Zhang, Hongyi, et al. "mixup: Beyond empirical risk minimization." *arXiv preprint arXiv:1710.09412* (2017).

MixStyle

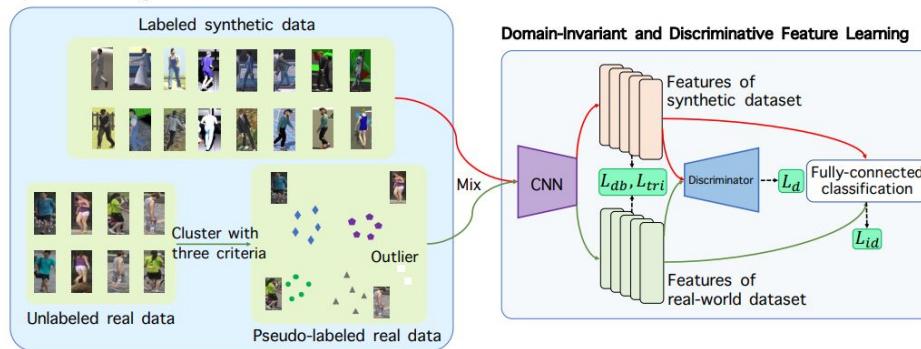


Zhou et al. Domain generalization with mixstyle. ICLR 2021.

Mixup Methods

DomainMix

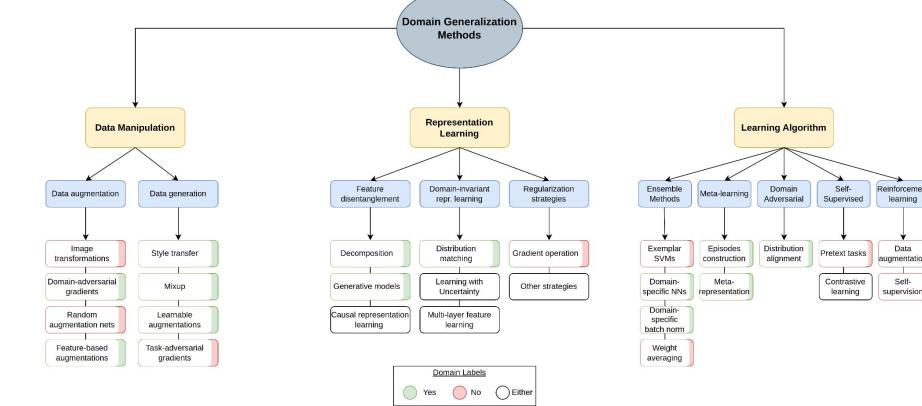
Dynamic Training Dataset Generation



Wang et al. DomainMix: Learning Generalizable Person Re-Identification Without Human Annotations. 2020.

Conclusions

Recap



Future directions (Christos)

An open problem

Applies to several application areas (many of them still unexplored)

Several subproblems emerge

- E.g., federated DG

Links to model explainability, model testing / assessment of robustness, bias etc.

References

- [1] A. A. Deshmukh, Y. Lei, S. Sharma, U. Dogan, J. W. Cutler, and C. Scott, “A generalization error bound for multi-class domain generalization,” *arXiv preprint arXiv:1905.10392*, 2019.
- [2] E. A. P. Alday *et al.*, “Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020,” *Physiol. Meas.*, vol. 41, no. 12, p. 124003, Dec. 2020, doi: 10.1088/1361-6579/abc960.
- [3] B. Sun and K. Saenko, “Deep CORAL: Correlation Alignment for Deep Domain Adaptation,” in *Computer Vision – ECCV 2016 Workshops*, G. Hua and H. Jégou, Eds., in Lecture Notes in Computer Science, vol. 9915. Cham: Springer International Publishing, 2016, pp. 443–450. doi: 10.1007/978-3-319-49409-8_35.
- [4] J. Cha, K. Lee, S. Park, and S. Chun, “Domain Generalization by Mutual-Information Regularization with Pre-trained Models,” in *Computer Vision – ECCV 2022*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds., in Lecture Notes in Computer Science, vol. 13683. Cham: Springer Nature Switzerland, 2022, pp. 440–457. doi: 10.1007/978-3-031-20050-2_26.
- [5] F. M. Carlucci, A. D’Innocente, S. Bucci, B. Caputo, and T. Tommasi, “Domain Generalization by Solving Jigsaw Puzzles,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA: IEEE, Jun. 2019, pp. 2224–2233. doi: 10.1109/CVPR.2019.00233.
- [6] T. Matsuura and T. Harada, “Domain generalization using a mixture of multiple latent domains,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, pp. 11749–11756.
- [7] D. Mahajan, S. Tople, and A. Sharma, “Domain Generalization using Causal Matching,” in *Proceedings of the 38th International Conference on Machine Learning*, PMLR, Jul. 2021, pp. 7313–7324. Accessed: Jul. 16, 2023. [Online]. Available: <https://proceedings.mlr.press/v139/mahajan21b.html>
- [8] K. Zhou, Y. Yang, Y. Qiao, and T. Xiang, “Domain Generalization with MixStyle,” in *International Conference on Learning Representations*, 2021. [Online]. Available: <https://openreview.net/forum?id=6xHJ37MVxxp>
- [9] X. Yue, Y. Zhang, S. Zhao, A. Sangiovanni-Vincentelli, K. Keutzer, and B. Gong, “Domain Randomization and Pyramid Consistency: Simulation-to-Real Generalization Without Accessing Target Domain Data,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South): IEEE, Oct. 2019, pp. 2100–2110. doi: 10.1109/ICCV.2019.00219.
- [10] W. Wang, “DomainMix: Learning Generalizable Person Re-Identification Without Human Annotations”.

References

- [11] A. Rame, C. Dancette, and M. Cord, "Fishr: Invariant Gradient Variances for Out-of-Distribution Generalization," in *Proceedings of the 39th International Conference on Machine Learning*, PMLR, Jun. 2022, pp. 18347–18377. Accessed: Jul. 16, 2023. [Online]. Available: <https://proceedings.mlr.press/v162/rame22a.html>
- [12] N. Somavarapu, C.-Y. Ma, and Z. Kira, "Frustratingly simple domain generalization via image stylization," *arXiv preprint arXiv:2006.11207*, 2020.
- [13] S. Shankar, V. Piratla, S. Chakrabarti, S. Chaudhuri, P. Jyothi, and S. Sarawagi, "Generalizing Across Domains via Cross-Gradient Training," in *International Conference on Learning Representations*, 2018. [Online]. Available: <https://openreview.net/forum?id=r1Dx7fbCW>
- [14] J. Wang *et al.*, "Generalizing to Unseen Domains: A Survey on Domain Generalization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 8, pp. 8052–8072, Aug. 2023, doi: 10.1109/TKDE.2022.3178128.
- [15] I. Gulrajani and D. Lopez-Paz, "In search of lost domain generalization," *arXiv preprint arXiv:2007.01434*, 2020.
- [16] W.-L. Zheng and B.-L. Lu, "Investigating Critical Frequency Bands and Channels for EEG-based Emotion Recognition with Deep Neural Networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015, doi: 10.1109/TAMD.2015.2431497.
- [17] D. Li, Y. Yang, Y.-Z. Song, and T. Hospedales, "Learning to generalize: Meta-learning for domain generalization," in *Proceedings of the AAAI conference on artificial intelligence*, 2018.
- [18] K. Zhou, Y. Yang, T. Hospedales, and T. Xiang, "Learning to Generate Novel Domains for Domain Generalization," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., in *Lecture Notes in Computer Science*, vol. 12361. Cham: Springer International Publishing, 2020, pp. 561–578. doi: 10.1007/978-3-030-58517-4_33.
- [19] F. Qiao, L. Zhao, and X. Peng, "Learning to Learn Single Domain Generalization," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA: IEEE, Jun. 2020, pp. 12553–12562. doi: 10.1109/CVPR42600.2020.01257.
- [20] Y. Balaji, S. Sankaranarayanan, and R. Chellappa, "MetaReg: Towards Domain Generalization using Meta-Regularization," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., Curran Associates, Inc., 2018. [Online]. Available: <https://proceedings.neurips.cc/paper/2018/file/647bba344396e7c8170902bcf2e15551-Paper.pdf>

References

- [21] X. Zhang, Y. He, R. Xu, H. Yu, Z. Shen, and P. Cui, “NICO++: Towards Better Benchmarking for Domain Generalization.” 2023. [Online]. Available: <https://openreview.net/forum?id=q08xeIw1HA1>
- [22] N. Ye *et al.*, “OoD-Bench: Quantifying and Understanding Two Dimensions of Out-of-Distribution Generalization,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA: IEEE, Jun. 2022, pp. 7937–7948. doi: 10.1109/CVPR52688.2022.00779.
- [23] X. Yao *et al.*, “PCL: Proxy-based Contrastive Learning for Domain Generalization,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA: IEEE, Jun. 2022, pp. 7087–7097. doi: 10.1109/CVPR52688.2022.00696.
- [24] G. Zhang, H. Zhao, Y. Yu, and P. Poupart, “Quantifying and Improving Transferability in Domain Generalization,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2021, pp. 10957–10970. Accessed: Jul. 16, 2023. [Online]. Available: <https://proceedings.neurips.cc/paper/2021/hash/5adaacd4531b78ff8b5cedfe3f4d5212-Abstract.html>
- [25] S. Beery, G. Van Horn, and P. Perona, “Recognition in Terra Incognita,” in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., in Lecture Notes in Computer Science, vol. 11220. Cham: Springer International Publishing, 2018, pp. 472–489. doi: 10.1007/978-3-030-01270-0_28.
- [26] Z. Xu, D. Liu, J. Yang, C. Raffel, and M. Niethammer, “Robust and Generalizable Visual Representation Learning via Random Convolutions.” arXiv, May 03, 2021. doi: 10.48550/arXiv.2007.13003.
- [27] Z. Wu, X. Wang, D. Lin, D. Lischinski, D. Cohen-Or, and H. Huang, “SAGNet: structure-aware generative network for 3D-shape modeling,” *ACM Trans. Graph.*, vol. 38, no. 4, p. 91:1-91:14, Jul. 2019, doi: 10.1145/3306346.3322956.
- [28] Z. Huang, H. Wang, E. P. Xing, and D. Huang, “Self-challenging Improves Cross-Domain Generalization,” in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, pp. 124–140. doi: 10.1007/978-3-030-58536-5_8.
- [29] D. Kim, Y. Yoo, S. Park, J. Kim, and J. Lee, “SelfReg: Self-supervised Contrastive Regularization for Domain Generalization,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada: IEEE, Oct. 2021, pp. 9599–9608. doi: 10.1109/ICCV48922.2021.00948.
- [30] Z. Li, K. Ren, X. JIANG, Y. Shen, H. Zhang, and D. Li, “SIMPLE: Specialized Model-Sample Matching for Domain Generalization,” in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=BqrPeZ_e5P
- [31] H. Ye, C. Xie, T. Cai, R. Li, Z. Li, and L. Wang, “Towards a Theoretical Framework of Out-of-Distribution Generalization,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2021, pp. 23519–23531. Accessed: Jul. 16, 2023. [Online]. Available: <https://proceedings.neurips.cc/paper/2021/hash/c5c1cb0bebd56ae38817b251ad72bedb-Abstract.html>

References

- [32] H. Zhang, Y.-F. Zhang, W. Liu, A. Weller, B. Scholkopf, and E. P. Xing, “Towards Principled Disentanglement for Domain Generalization,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA: IEEE, Jun. 2022, pp. 8014–8024. doi: 10.1109/CVPR52688.2022.00786.
- [33] X. Li, Y. Dai, Y. Ge, J. Liu, Y. Shan, and L. DUAN, “Uncertainty Modeling for Out-of-Distribution Generalization,” in *International Conference on Learning Representations*, 2022.
- [34] P. W. Koh *et al.*, “WILDS: A Benchmark of in-the-Wild Distribution Shifts.” arXiv, Jul. 16, 2021. Accessed: Jul. 16, 2023. [Online]. Available: <http://arxiv.org/abs/2012.07421>

Appendix: A little bit of theory (Christos)

Notes:

DA theory / bounds

From DA to DG