Institute for Infocomm Research
I²R

ROBUST DOMAIN-FREE DOMAIN GENERALIZATION WITH CLASS-AWARE ALIGNMENT

Zhang Wenyu, Mohamed Ragab, Ramon Sagarna

Machine Intellection Department,
Institute for Infocomm Research

Nanyang Technological University

IEEE ICASSP 2021, 6-11 June

Outline

Contents:

CREATING GROWTH, ENHANCING LIVES

Motivation

- Deep neural network performance is often reliant on the assumption that train and test sample distributions are the same
- Data collection is often resource-constrained, and out-of-distribution samples can be present at test-time
    - Practical application examples: new road conditions for self-driving car, new operating conditions for machinery, new users of device
- Model robustness is needed to avoid compromising the accuracy of trained models at deployment in the wild

CREATING GROWTH, ENHANCING LIVES

Motivation: Domain Generalization

- Domain: a data generation regime
- Domain generalization aims to learn a robust model from source/training domains that can directly generalize to new target/testing domains
  - No target samples used at training (difference from domain adaptation)

Source domains

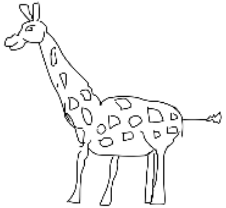
Painting

Cartoon

Photo

Source domain labels

Target domain



(Sketch)

Unknown target domain label

Motivation: Domain-Free Domain Generalization

- Domain-free: no domain labels i.e. unable to group source samples by domain labels during training.
- Domain labels may be unavailable in practice, and dataset labels cannot replace domain labels when samples of a dataset are drawn from a mixture of domains

Source domains

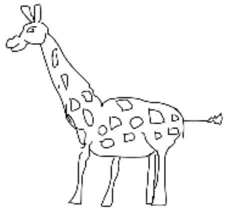
~~Painting~~

~~Cartoon~~

~~Photo~~

Unknown source domain labels

Target domain



(Sketch)

Unknown target domain label

Problem Setup and Notations

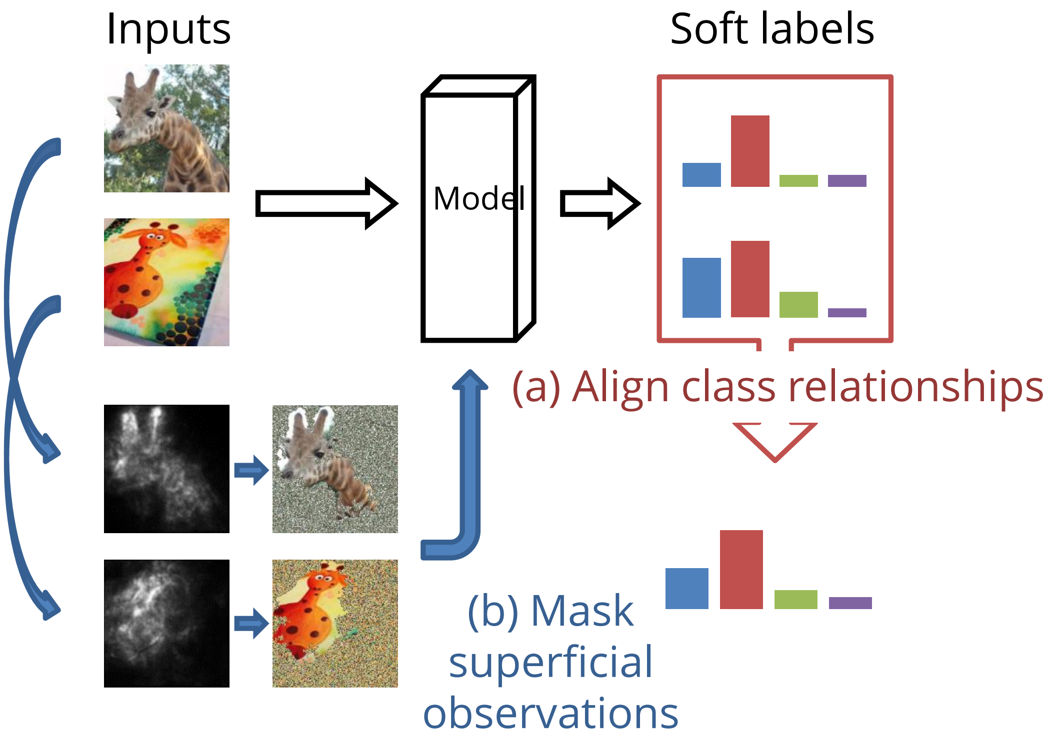
Setup for classification tasks:

- For each domain D, samples $\left(x_i^{(D)}, y_i^{(D)}\right)$ are drawn from a fixed distribution $\left(X^{(D)}, Y^{(D)}\right) \sim P^{(D)}$, y_i is one-hot vector of true class label in C classes
- For model f parameterized by θ, soft labels or vector of predicted class probabilities $p_i = softmax\left(f(x_i; \theta)\right)$

Goal:

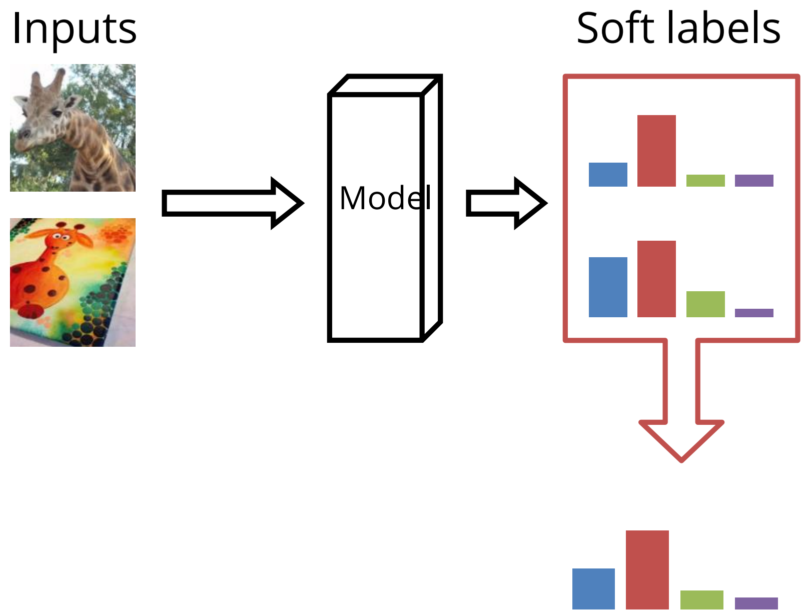
- Learn a robust model from source domains that can generalize to new unseen target domains, without source domain labels

Proposed Method: DFDG



CREATING GROWTH, ENHANCING LIVES

Proposed Method: DFDG



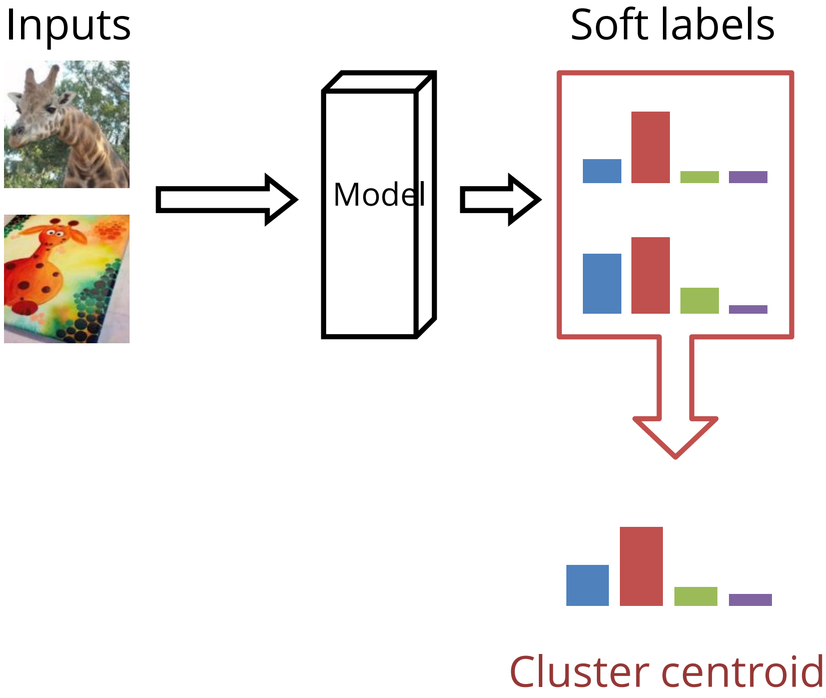
Objective function for batch B:

$$L = \ell_{ce} + \alpha \ell_{align}$$

- Samples from the same class to have similar class relationships regardless of domains.

CREATING GROWTH, ENHANCING LIVES

Proposed Method: DFDG



Objective function for batch B:

$$L = \ell_{ce} + \alpha \ell_{align}$$

Cross entropy loss

$$\ell_{ce} = -\frac{1}{|B|}\sum_{i\in B} y_i^T \log(p_i)$$

Class relationship alignment loss:

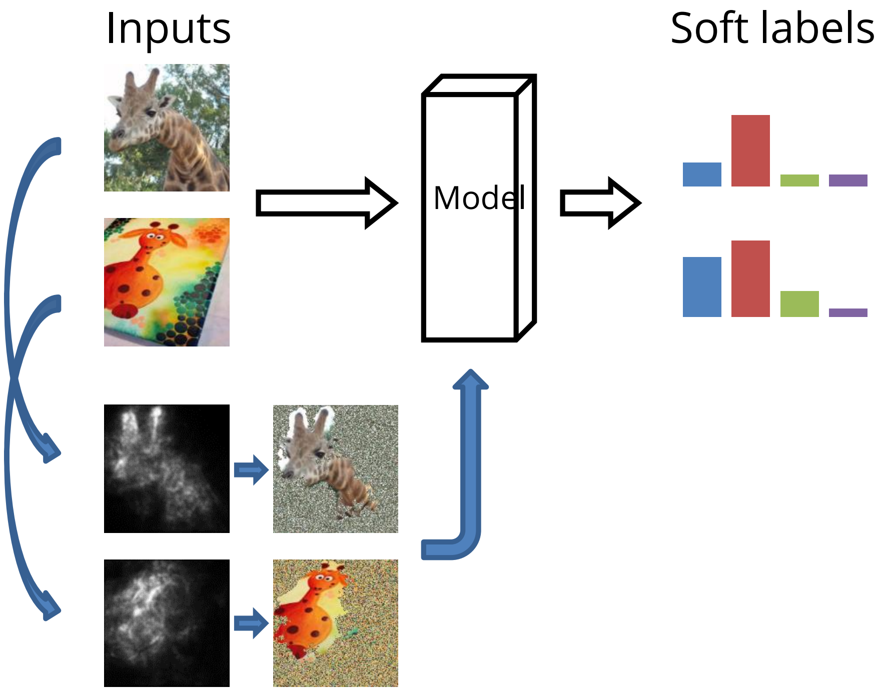
$$\ell_{align} = \sum_{c=1}^{C} \frac{1}{|B(c)|} \sum_{i\in B(c)} \left\| p^{(i)} - \mu(c) \right\|_2^2$$

where

$$B(c) = \{i | i \in B, [1, \cdots, C] y_i = c\}$$

$$\mu(c) = \frac{1}{|B(c)|} \sum_{i\in B(c)} p^{(i)}$$

Proposed Method: DFDG

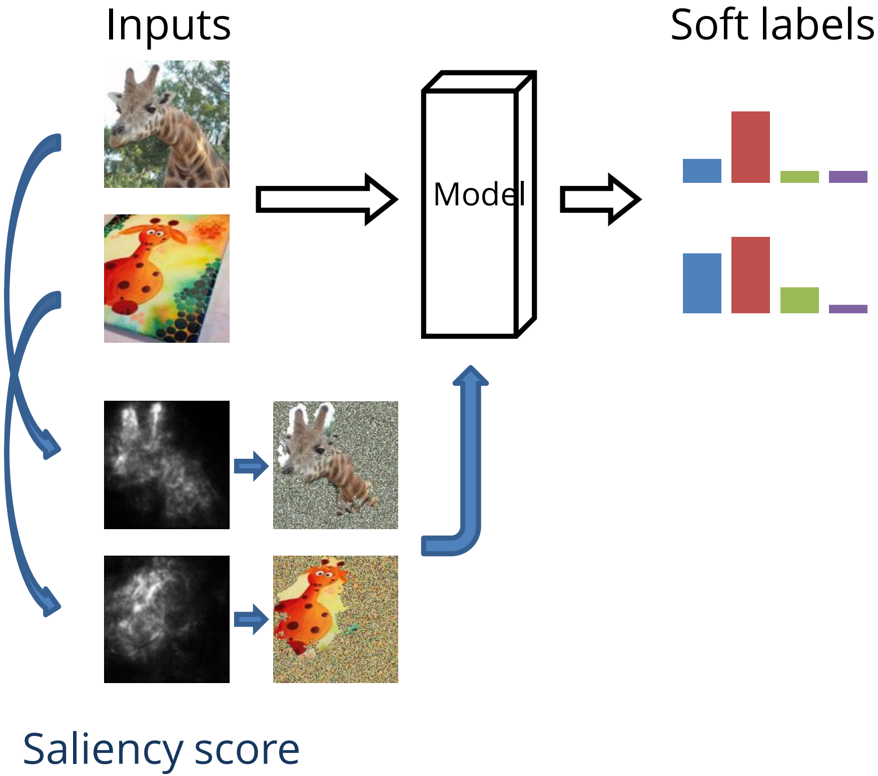


Mask superficial observations

- Superficial observations (e.g. backgrounds, styles in images) can lead to overfitting to training data
- Perturb inputs so that trained model is more robust to variations in superficial observations

CREATING GROWTH, ENHANCING LIVES

Proposed Method: DFDG



Mask superficial observations

1. Rank pixels by SmoothGrad saliency score

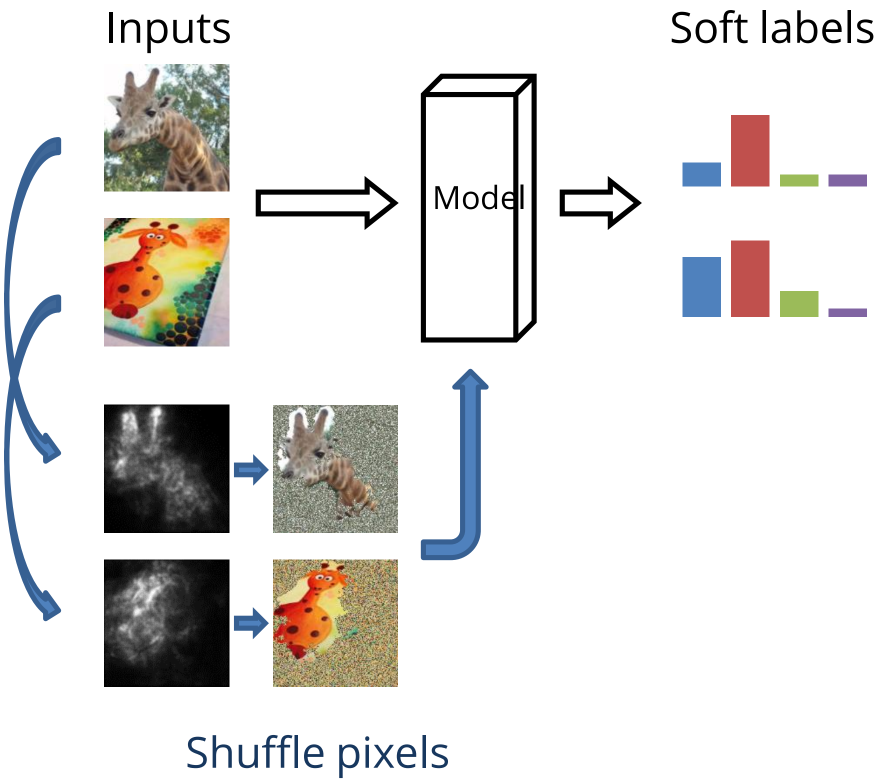
Vanilla saliency score:

$$g(x, c) = \left(\frac{\partial f(x, \theta)[c]}{\partial x}\right)^2$$

SmoothGrad saliency score:

Averages n replicates of $g(x, c)$ where Gaussian noise is added to x in each replicate

Proposed Method: DFDG



Mask superficial observations

1. Rank pixels by SmoothGrad saliency score
2. Sample $q \sim Unif(0, qMax)$
3. Pixels with saliency score below the q^{th} percentile are shuffled

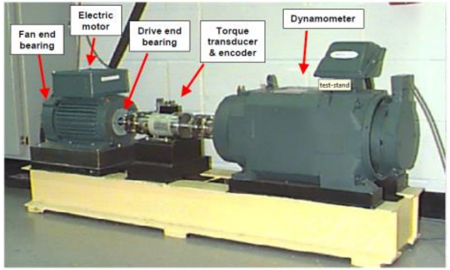
In each batch, augment $m\%$ of samples.

($m = 50, qMax = 70$ in experiments)

Experiments: Datasets

Bearings (vibration sensor signals)

- 10 fault classes
- 8 operating conditions: 4 loading torques x 2 bearing locations



HHAR (motion sensor signals)

- 6 human activity classes
- 9 users

PACS (images)

- 7 classes
- 4 art styles

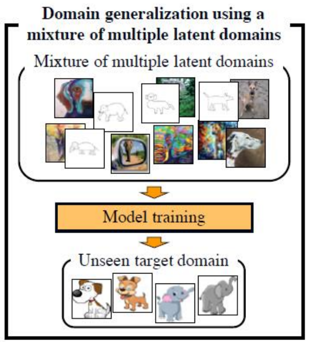
Experiments: Competing Methods

TrainAll

- Baseline with cross entropy loss

MMLD

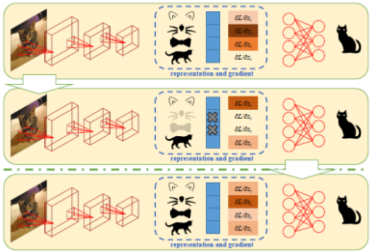
- Estimates source domain labels using convolutional feature statistics, and aligns the domains with a discriminator network



Matsuura, Toshihiko and Harada, Tatsuya. "Domain Generalization Using a Mixture of Multiple Latent Domains." AAAI (2020).

RSC

- Zeros out penultimate-layer feature representations associated with the highest gradient in the final classification layer



Huang, Zeyi, Haohan Wang, E. Xing and Dong Huang. "Self-Challenging Improves Cross-Domain Generalization." *ECCV (2020)*

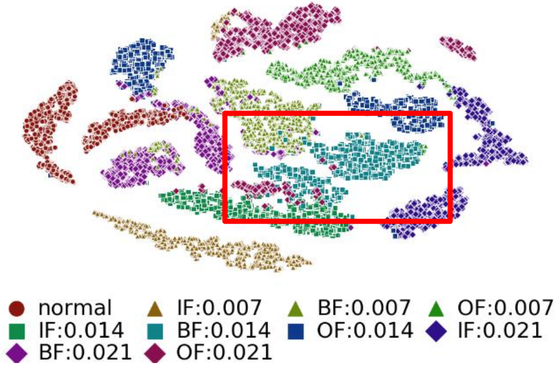
Experiments: Generalization Performance

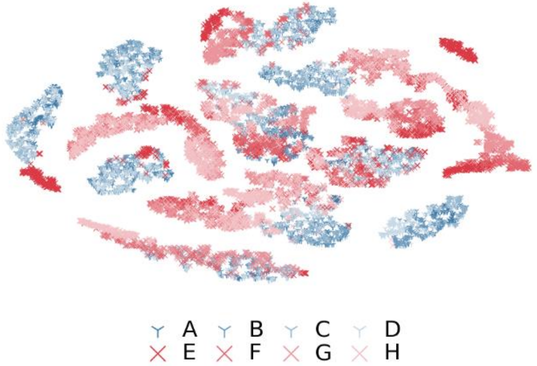
- Bearings

Target	Accuracy (%)				
	TrainAll	RSC	MMLD	DFDG	DFDG+RSC
A	52.33	68.43	65.70	66.60	68.23
B	90.50	91.10	95.37	89.30	90.90
C	92.90	97.60	89.33	89.67	92.47
D	74.97	77.73	65.87	77.73	77.97
E	70.73	70.33	64.20	73.40	74.77
F	88.53	86.40	79.03	91.50	90.23
G	87.20	90.20	86.83	91.40	94.30
H	85.20	90.50	75.53	85.30	88.37
Avg	80.30	84.04	77.73	83.13	**84.65**

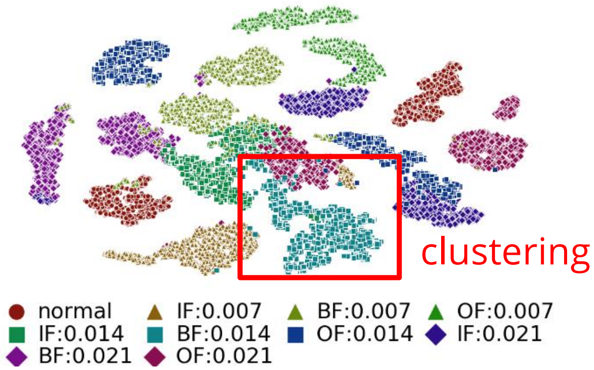
Experiments: Generalization Performance

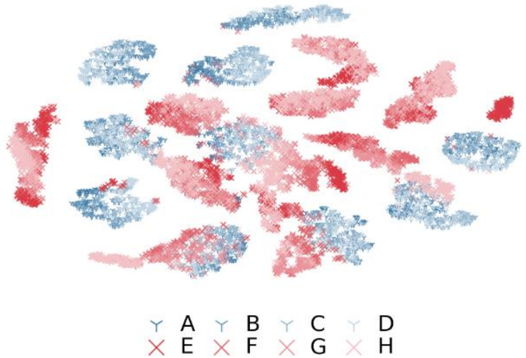
TrainAll





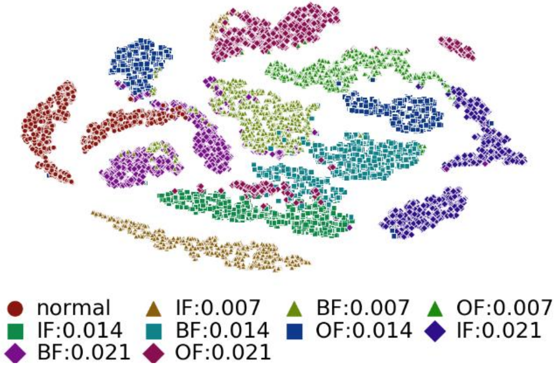
DFDG

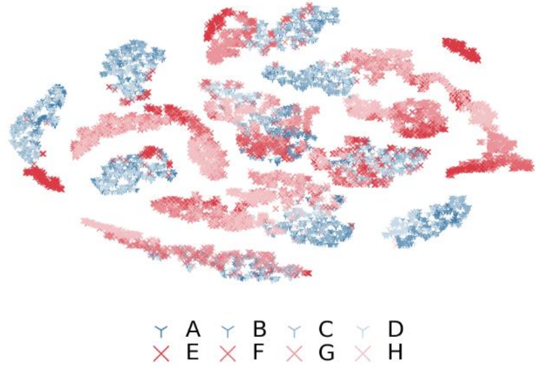




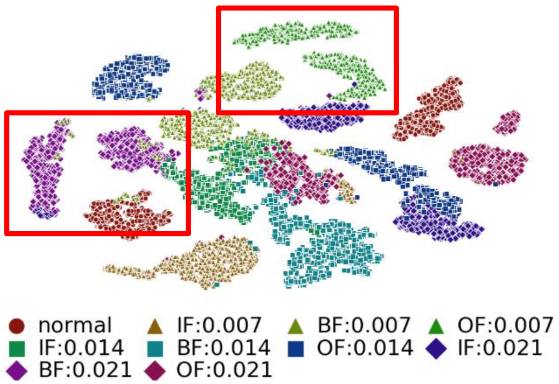
Experiments: Generalization Performance

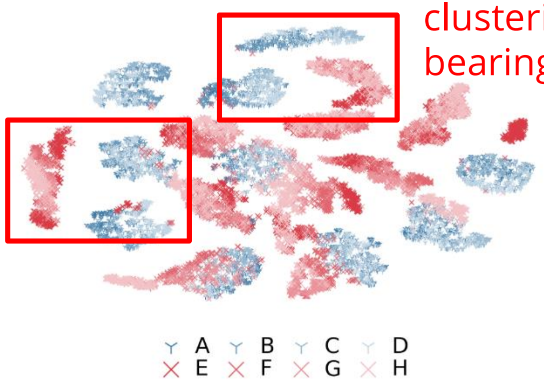
TrainAll





DFDG





clustering by bearing location

Experiments: Generalization Performance

- HHAR

Target	Accuracy (%)			
	TrainAll	RSC	MMLD	DFDG
A	43.27	41.28	45.82	36.76
B	48.91	44.54	60.43	70.91
C	49.15	47.68	46.73	54.63
D	45.73	52,23	49.93	62.60
E	46.59	44.80	52.88	66.40
F	41.98	43.02	46.95	69.28
G	30.36	29.99	43.90	60.64
H	54.84	57.75	52.38	47.30
I	40.86	40.59	47.19	55.07
Avg	45.15	45.23	49.58	**58.18**

Experiments: Generalization Performance

- PACS

Target	Accuracy (%)			
	TrainAll	RSC	MMLD	DFDG
Art	78.73	80.37	78.99	79.23
Cartoon	74.30	76.84	77.06	75.84
Photo	94.55	94.99	95.41	95.45
Sketch	76.19	74.40	62.56	77.87
Avg	80.94	81.65	78.51	**82.10**

CREATING GROWTH, ENHANCING LIVES

CREATING GROWTH, ENHANCING LIVES

Experiments: Ablation Studies

- Bearings

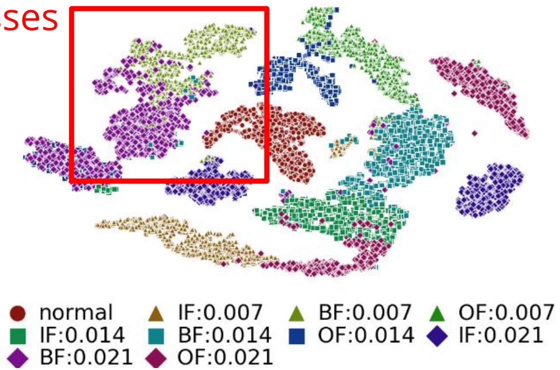
Saliency Map	Avg Accuracy (%)
Vanilla gradient	80.77
SmoothGrad	83.42

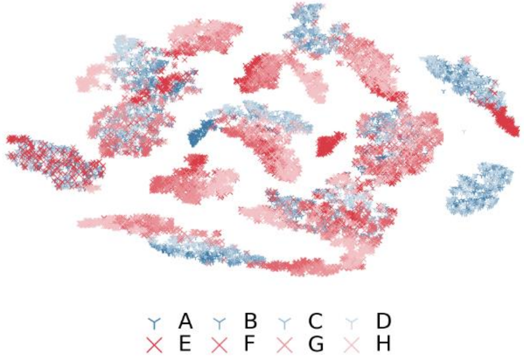
Regularization	Masking		Avg Accuracy (%)
α	m	qMax	
-	-	-	79.75
0.01	-	-	80.21
0.1	-	-	79.97
-	33	30	83.21
-	50	70	83.42

Experiments: Ablation Studies

Less separation between classes

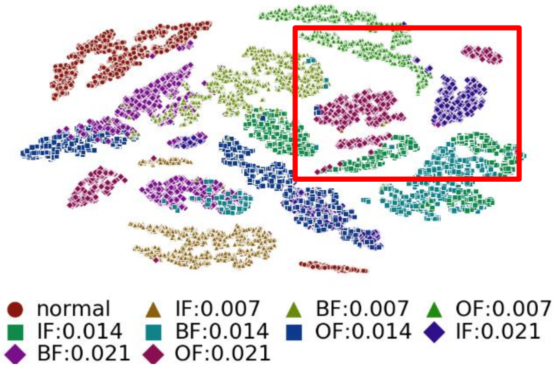
DFDG w/o mask

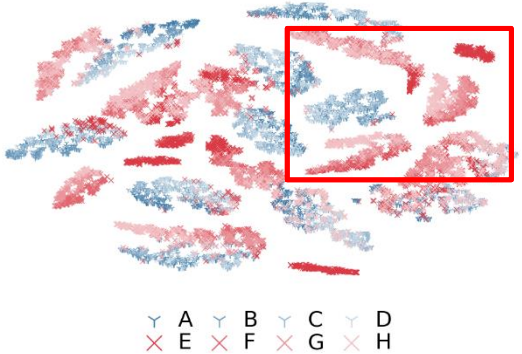




Domains may be individually clustered

DFDG w/o align





Summary

Proposed DFDG method

- Aligns class relationships of source samples
- Masks superficial observations from source samples

☑ Model-agnostic

☑ Does not require domain labels for training

☑ Attains better performance over baseline and competing methods for both time series sensor and image classification tasks

  ☑ Complements existing method (RSC) to achieve best performance on Bearings dataset

CREATING GROWTH, ENHANCING LIVES

Selected References

Mancini, Massimiliano, S. R. Bulò, B. Caputo and E. Ricci. “Best Sources Forward: Domain Generalization through Source-Specific Nets.” *25th IEEE International Conference on Image Processing (ICIP)* (2018): 1353-1357.

T. S. Talagala, R. J. Hyndman, and G. Athanasopoulos, “Meta-learning how to forecast time series,” *Tech. Rep., Monash University, Dept. of Econometrics and Business Statistics* (2018).

H. Li, S. J. Pan, S. Wang, and A. C. Kot, “Domain generalization with adversarial feature learning,” *ICCV* (2018).

Dou, Qi, Daniel Coelho de Castro, K. Kamnitsas and B. Glocker. “Domain Generalization via Model-Agnostic Learning of Semantic Features.” *NeurIPS* (2019).

D. Li, J. Zhang, Y. Yang, C. Liu, Y-Z. Song, and T. M.Hospedales, “Episodic training for domain generalization,” *ICCV* (2019).

CREATING GROWTH, ENHANCING LIVES

Selected References

Huang, Zeyi, Haohan Wang, E. Xing and Dong Huang. "Self-Challenging Improves Cross-Domain Generalization." *ECCV (2020).*

Matsuura, Toshihiko and Harada, Tatsuya. "Domain Generalization Using a Mixture of Multiple Latent Domains." AAAI (2020).

D. Smilkov, N. Thorat, B. Kim, F. Viegas, and M. Wattenberg, "Smoothgrad: removing noise by adding noise," *Workshop on Visualization for Deep Learning, ICML* (2017).

Smith, W. and R. Randall. "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study." *Mechanical Systems and Signal Processing* 64 (2015): 100-131.

Garrett Wilson, Janardhan Rao Doppa, and Diane J. Cook, "Multi-source deep domain adaptation with weak supervision for time-series sensor data," *KDD* (2020).

CREATING GROWTH, ENHANCING LIVES

Institute for Infocomm Research

I²R

THANK YOU

www.a-star.edu.sg