Domain Generalization via Model-agnostic Learning of Semantic Features

Qi Dou, Daniel C. Castro, Konstantinos Kamnitsas, Ben Glocker

Biomedical Image Analysis Group, Department of Computing, Imperial College London, UK

Motivation

Problem: Generalization capability to **unseen domains** is crucial for machine learning models when deploying to real-world conditions.

- *wide success of deep neural networks in visual understanding tasks*
- *large variety in image acquisition conditions*
- *domain shift leads to performance degradation*

Goal:

Train a model using **multi-domain source data**, such that it can **directly generalize** to new domains without need of retraining.

Related Work

Model-dependent approaches:

- *adversarial feature alignment for domain invariance* [Li et al. ECCV 2018]
- *decompose networks parameters to domain-specific/invariant* [Khosla ECCV 2012]
- *data augmentation based methods* [Shankar et al. ICLR 2018; Volpi et al. NeurIPS 2018]
- *multi-task or self-supervised signals* [Ghifary et al. ICCV 2015; Carlucci et al. CVPR 2019]

Model-agnostic approaches via gradient-based meta-learning:

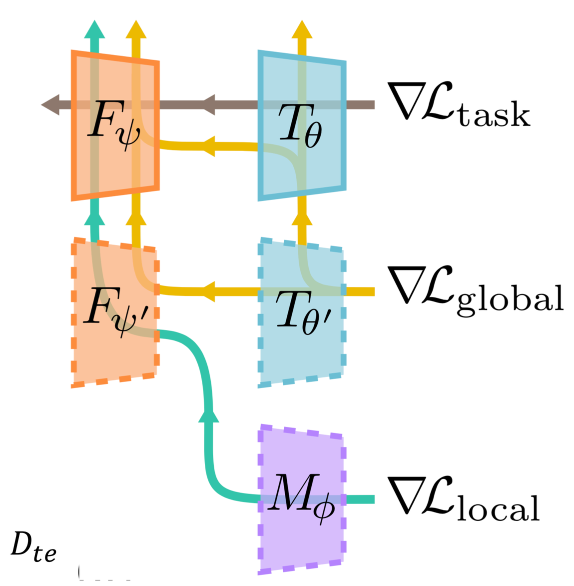
- *MLDG: directly applying episodic training paradigm* [Li et al. AAAI 2018]
- *MetaReg: meta-learning of weights regularization term* [Balaji et al. NeurIPS 2018]
- *Epi-FCR: episodic training with alternative model updates* [Li et al. ICCV 2019]

Method

MASF: model-agnostic learning of semantic features

- **Episodic training paradigm**

A **model-agnostic learning** paradigm with gradients based meta-learning procedures to expose model optimization to **simulated domain shift**



available domains: $D = \{D_1, D_2, \ldots, D_K\}$

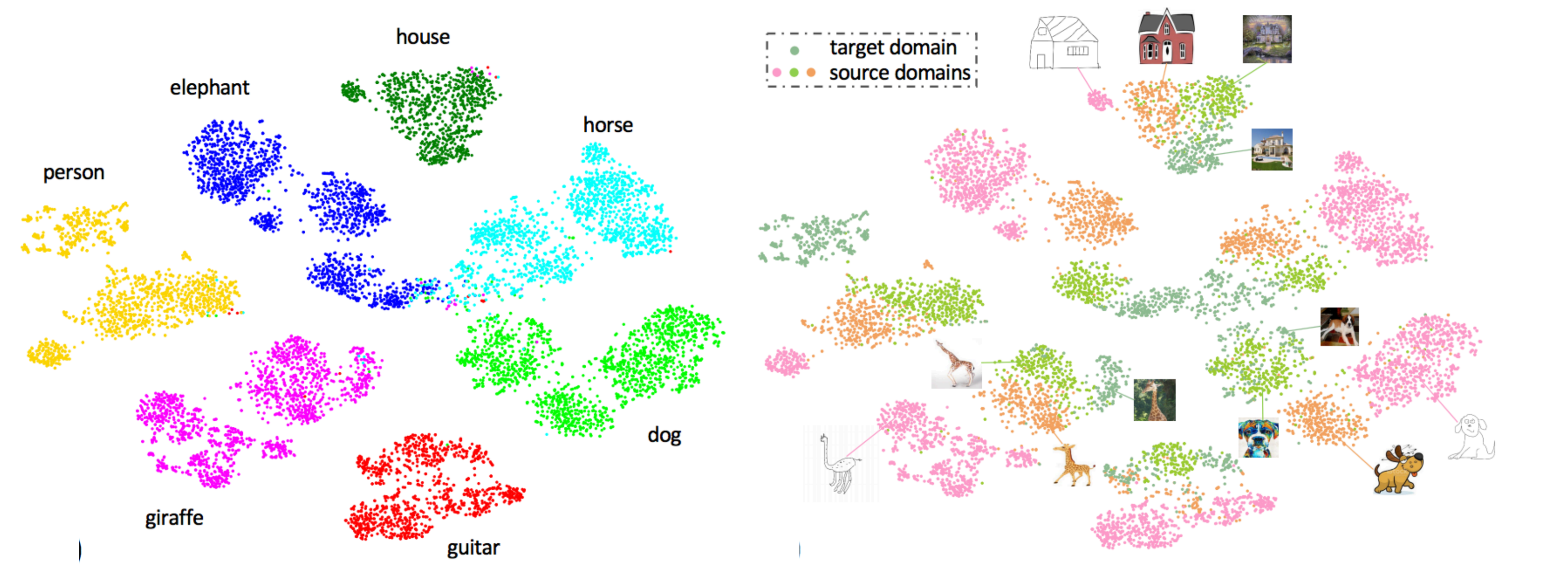
neural network is composed of feature extractor F_ψ and task classifier T_θ

In each iteration: split into meta-train D_{tr} and meta-test D_{te}

update the parameters one or more steps with gradient descent:

$$(\psi', \theta') = (\psi, \theta) - \alpha \nabla_{\psi,\theta} \mathcal{L}_{\text{task}}(\mathcal{D}_{\text{tr}}; \psi, \theta)$$

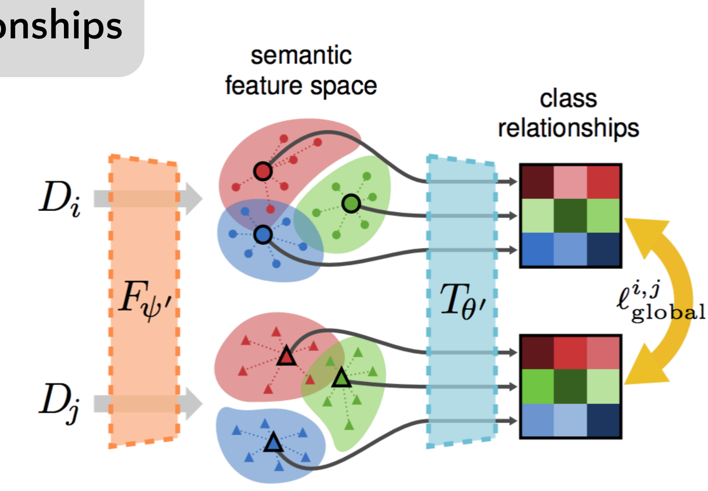
with the temporarily updated parameters, apply meta-learning step, to enforce certain properties to be exhibited on held-out meta-test domain, for globally and locally regularizing semantic features.



Method

- **Global class alignment**

Align a derived **soft confusion matrix** to preserve general knowledge about **inter-class relationships**



In each domain, compute class-specific mean feature vector:

$$\bar{\mathbf{z}}_c^{(k)} = \frac{1}{N_k^{(c)}} \sum_{n: y_n^{(k)} = c} F_{\psi'}(\mathbf{x}_n^{(k)}) \approx \mathbb{E}_{D_k}[F_{\psi'}(\mathbf{x}) \mid y = c]$$

compute soft label distribution:

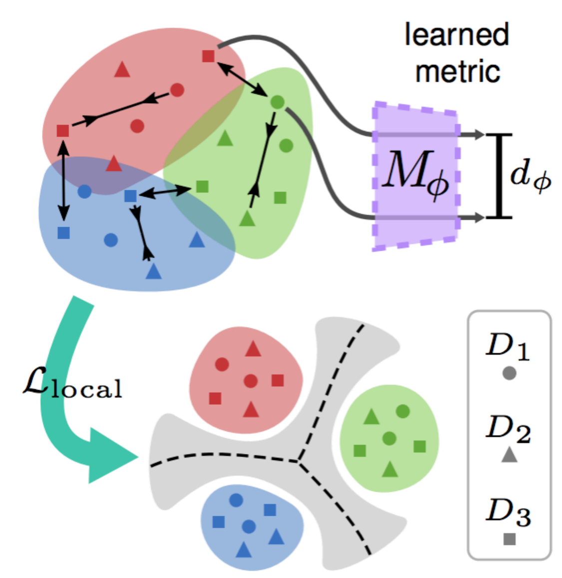
$$\mathbf{s}_c^{(k)} = \text{softmax}(T_{\theta'}(\bar{\mathbf{z}}_c^{(k)})/\tau)$$

with $(D_i, D_j) \in D_{tr} \times D_{te}$, regularize consistency of inter-class alignment:

$$\ell_{\text{global}}(D_i, D_j; \psi', \theta') = \frac{1}{C} \sum_{c=1}^{C} \frac{1}{2} [D_{\text{KL}}(\mathbf{s}_c^{(i)} \,\|\, \mathbf{s}_c^{(j)}) + D_{\text{KL}}(\mathbf{s}_c^{(j)} \,\|\, \mathbf{s}_c^{(i)})]$$

- **Local sample clustering**

We promote domain independent class-specific **cohesion and separation** of sample features with a **metric-learning** component



metric-learning with an embedding network, obtain a learnable distance function:

$$d_\phi(\mathbf{z}_n, \mathbf{z}_m) = \|\mathbf{e}_n - \mathbf{e}_m\|_2 = \|M_\phi(\mathbf{z}_n) - M_\phi(\mathbf{z}_m)\|_2$$

regularize with contrastive loss, separately collapse each group of same-class exemplars to a distinct single point:

$$\ell_{\text{con}}^{n,m} = \begin{cases} d_\phi(\mathbf{z}_n, \mathbf{z}_m)^2, & \text{if } y_n = y_m \\ (\max\{0, \xi - d_\phi(\mathbf{z}_n, \mathbf{z}_m)\})^2, & \text{if } y_n \neq y_m \end{cases}$$

Or use triplet loss, make pairs of samples from the same class closer than pairs from different classes, by a certain margin ξ

$$\ell_{\text{tri}}^{a,p,n} = \max\{0, \; d_\phi(\mathbf{z}_a, \mathbf{z}_p)^2 - d_\phi(\mathbf{z}_a, \mathbf{z}_n)^2 + \xi\}$$

Experiments

- **Image recognition benchmark of VLCS**

❖ *5 classes: bird, car, chair, dog, person*
❖ *4 domains corresponding to 4 datasets: VOC2007, LabelMe, Caltech, SUN09*
❖ *input is pre-extracted DeCAF features (4096-dimensional vector)*

Table 1: Domain generalization results on VLCS dataset with object recognition accuracy (%).

Source	Target	D-MTAE [12]	CIDDG [30]	CCSA [34]	DBADG [25]	MMD-AAE [28]	MLDG [26]	Epi-FCR [27]	JiGen [3]	DeepAll (Baseline)	MASF (Ours)
L,C,S	V	63.90	64.38	67.10	69.99	67.70	67.7	67.1	70.62	68.67±0.09	69.14±0.19
V,C,S	L	60.13	63.06	62.10	63.49	62.60	61.3	64.3	60.90	63.10±0.11	64.90±0.08
V,L,S	C	89.05	88.83	92.30	93.63	94.40	94.4	94.1	96.93	92.86±0.13	94.78±0.16
V,L,C	S	61.33	62.10	59.10	61.32	64.40	65.9	65.9	64.30	64.11±0.17	67.64±0.12
Average		68.60	69.59	70.15	72.11	72.28	72.3	72.9	73.19	72.19	74.11

- **Image recognition benchmark of PACS**

❖ *7 classes: dog, elephant, giraffe, guitar, house, horse, person*
❖ *4 domains with distinct styles, i.e., art painting, cartoon, photo, sketch*
❖ $F_\psi \circ T_\theta$ *uses AlexNet pretrained on ImageNet,* M_Φ *uses FC with triplet loss*

Table 2: Domain generalization results on PACS dataset with recognition accuracy (%) using AlexNet.

Source	Target	D-MTAE [12]	CIDDG [30]	DBADG [25]	MLDG [26]	Epi-FCR [27]	MetaReg [1]	JiGen [3]	DeepAll (Baseline)	MASF (Ours)
C,P,S	Art painting	60.27	62.70	62.86	66.23	64.7	69.82	67.63	67.60±0.21	70.35±0.33
A,P,S	Cartoon	58.65	69.73	66.97	66.88	72.3	70.35	71.71	68.87±0.22	72.46±0.19
A,C,S	Photo	91.12	78.65	89.50	88.00	86.1	91.07	89.00	89.20±0.24	90.68±0.12
A,C,P	Sketch	47.68	64.45	57.51	58.96	65.0	59.26	65.18	61.13±0.30	67.33±0.12
Average		64.48	68.88	69.21	70.01	72.0	72.62	73.38	71.70	75.21

Experiments

- **Analytical experiments**

❖ ablation of key components of episodic training, global loss and local loss

Table 3: Ablation study on key components of our method with the PACS dataset (accuracy, %).

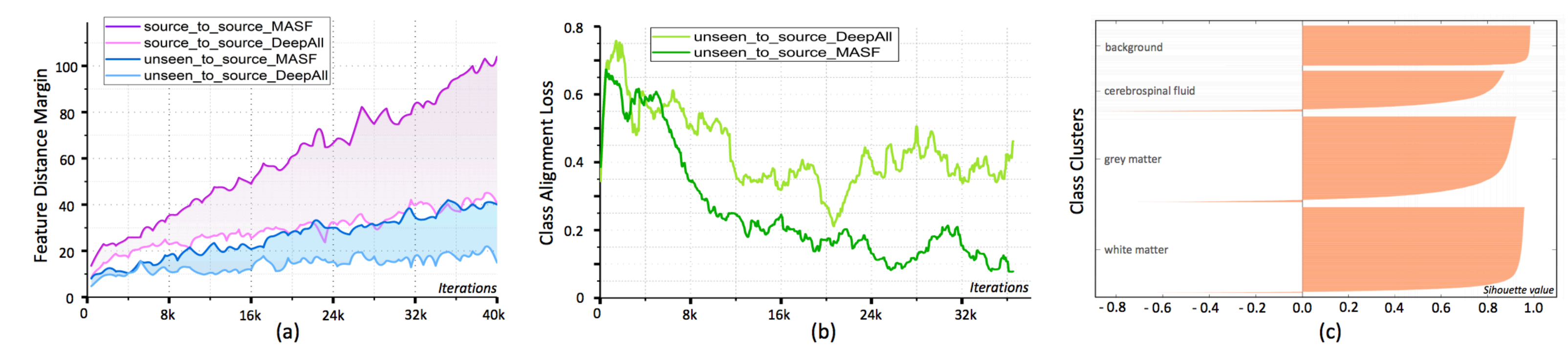
Episodic	$\mathcal{L}_{\text{global}}$	$\mathcal{L}_{\text{local}}$	Art	Cartoon	Photo	Sketch	Average
-	-	-	67.60±0.21	68.87±0.22	89.20±0.24	61.13±0.30	71.70
✓	-	-	69.19±0.10	70.66±0.37	90.36±0.18	59.89±0.26	72.52
-	✓	-	69.43±0.29	70.22±0.21	90.64±0.15	60.11±0.17	72.60
-	-	✓	69.50±0.15	70.25±0.13	90.12±0.12	63.02±0.12	73.22
-	✓	✓	69.48±0.20	71.15±0.16	90.16±0.15	64.73±0.34	73.88
✓	✓	-	69.94±0.15	72.16±0.28	90.10±0.12	63.54±0.13	73.93
✓	-	✓	69.50±0.20	71.44±0.34	90.16±0.15	64.97±0.28	74.02
✓	✓	✓	70.35±0.33	72.46±0.19	90.68±0.12	67.33±0.12	75.21

❖ scalability to deep models

Table 4: PACS results with deep residual network architectures (accuracy, %).

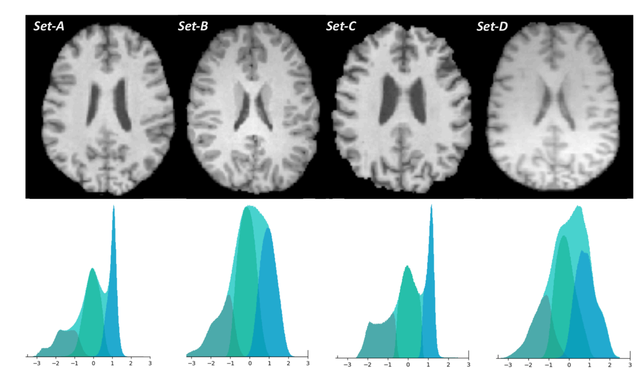
Source	Target	ResNet-18		ResNet-50	
		DeepAll	MASF (ours)	DeepAll	MASF (ours)
C,P,S	Art-painting	77.38 ± 0.15	80.29 ± 0.18	81.41 ± 0.16	82.89 ± 0.16
A,P,S	Cartoon	75.65 ± 0.11	77.17 ± 0.08	78.61 ± 0.17	80.49 ± 0.21
A,C,S	Photo	94.25 ± 0.09	94.99 ± 0.09	94.83 ± 0.06	95.01 ± 0.10
A,C,P	Sketch	69.64 ± 0.25	71.69 ± 0.22	69.69 ± 0.11	72.29 ± 0.15

❖ observation of learning procedure



- **Application to a real-world medical task**

❖ pixel-wise segmentation of brain tissue in MRI
❖ *data acquisition differences in scanners, imaging protocols, etc.*
❖ *3 brain tissues: white matter, gray matter and cerebrospinal fluid*
❖ *4 domains corresponding to 4 different clinical sites*



Test \ Train	Set-A	Set-B	Set-C	Set-D	DeepAll	MASF
Set-A	90.62	88.91	88.81	85.03	89.09	89.82
Set-B	85.03	94.22	81.38	88.31	90.41	91.71
Set-C	93.14	92.80	95.40	88.68	94.30	94.50
Set-D	76.32	88.39	73.50	94.29	88.62	89.51

Conclusion

We have presented a new approach to domain generalization of predictive models by incorporating global and local constraints for learning semantic feature spaces. The better generalization capability is demonstrated by new state-of-the-art results on popular benchmarks and a dense classification task for medical images.

References

[1] Da Li, Jianshu Zhang, Yongxin Yang, Cong Liu, Yi-Zhe Song, and Timothy M. Hospedales. *Episodic training for domain generalization*. ICCV, 2019.
[2] Yogesh Balaji, Swami Sankaranarayanan, and Rama Chellappa. *MetaReg: Towards domain generalization using meta-regularization*. NeurIPS, 2018.
[3] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M. Hospedales. *Learning to generalize: Meta-learning for domain generalization*. AAAI, 2018.

Acknowledgement

project mira • ERC StG 2017
Imperial College London
BioMedIA Imperial College London

Code & Poster