Domain Generalization via Model-Agnostic Learning of Semantic Features

Qi Dou **Daniel C. Castro** **Konstantinos Kamnitsas** **Ben Glocker**
Biomedical Image Analysis Group, Imperial College London, UK
{qi.dou,dc315,kk2412,b.glocker}@imperial.ac.uk

(NeurIPS 2019)

2021.07.16
임진혁

Index

임진혁<eeplearning@units.ac.kr>

Introduction: what is Domain Generalization?

Deep Learning's remarkable success
BUT it's based on the assumption that training and test data are sampled **from the same distribution**

In Real Word Application : this assumption is often violated
(Train/Test => different distribution => Test sampled from **Unseen Domain)**

Domain shift로 인해 실제 industrial 적용 시에는 performance degradation인 경우가 흔함

임진혁<eeplearning@units.ac.kr>

Introduction: what is Domain Generalization?

Domain shift로 인해 실제 industrial 적용 시에는 performance degradation인 경우가 흔함

=>Generalization capability to unseen domains is crucial when deploying to real-world conditions.

=> Domain Generalization은 Unseen Domain Performance Degradation를 해결하고자 하는 연구분야

=> Data Distribution이 바뀌는 Domain Drifts에 강건한 모델 학습

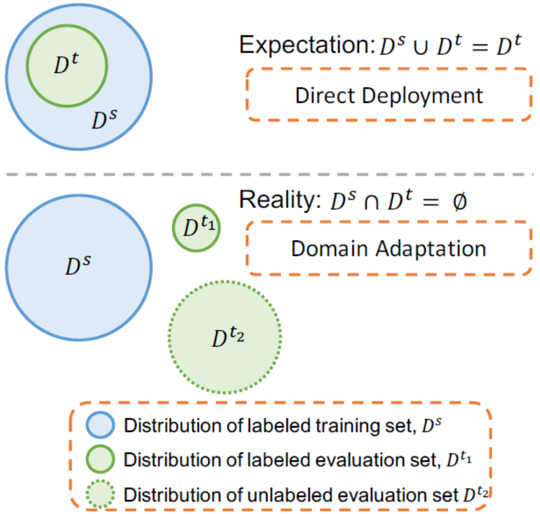


Figure 1. Domain adaptation in the true data space: Expectation vs. Reality.

임진혁<eeplearning@units.ac.kr>

Domain Generalization

- assumes "The existence of domain-variant patterns" in the inputs
- learn domain-invariant feature representation (Generalizable features)
    - (in CV: "capture general semantic features for object recognition"
- train on multi-domain source data for generalizing to new target domain without retraining
- perform well across seen and unseen domains
- information or explicit knowledge about the new domains is not available (real-world applications)

임진혁<eeplearning@units.ac.kr>

How to learn domain-invariant feature representation(Generalizable features)
which is discriminative for the specific task but insensitive to changes of domain-specific statistics

- regularization of the feature space
  Domain generalization via invariant feature representation *(ICML 2013)*

- adversarial feature alignment via maximum mean discrepancy
  Domain generalization with adversarial feature learning (CVPR 2018)

- Using episodic learning(Meta Learning)
  - splitting the available source domains into meta-train and meta-test at each iteration (to simulate domain shift)
  - (1) Task error : ***Learning to generalize: Meta-learning for domain generalization.(aaai 2018)***
  - (2) Classifier regularizer : ***MetaReg: Towards domain generalization using meta-regularization (NIPS2018)***
  - (3) Feature Critic Module : ***Unsupervised domain adaptation with residual transfer networks (NIPS2016)***
  - Etc) ***Episodic Training for Domain Generalization (arXiv 2019)***

임진혁<eeplearning@unist.ac.kr>

How to learn domain-invariant feature representation(Generalizable features)

MASF(**M**odel-**A**gnostic Learning of **S**emantic **F**eature)

[1] training a model on multi-domain source
- directly generalize to **target domains with unknown statistics(목표)**

[2] expose the optimization to domain shift
- gradient-based meta-train and meta-test procedures

[3] two complementary losses
- explicitly regularize the semantic structure of the feature space

[4] a derived soft confusion matrix
- to preserve general knowledge about inter-class relationships

[5] Locally
- domain- independent class-specific
- cohesion and separation of
- sample features with
- a metric-learning component

임진혁<eeplearning@unist.ac.kr>

How to learn domain-invariant feature representation(Generalizable features)

MASF(**M**odel-**A**gnostic Learning of **S**emantic **F**eature)

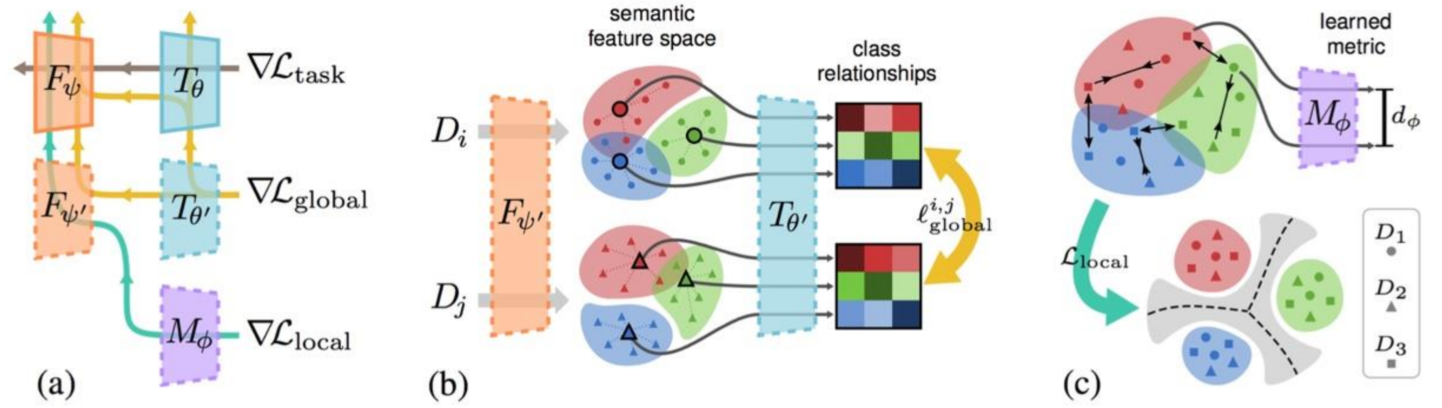


Figure 1: An overview of the proposed model-agnostic learning of semantic features (MASF): (a) episodic training under simulated domain shift, with gradient flows indicated; (b) global alignment of class relationships; (c) local sample clustering, towards cohesion and separation. F_ψ and T_θ are the feature extractor and the task net, $F_{\psi'}$ and $T_{\theta'}$ are their updated versions by inner gradient descent on the task loss $\mathcal{L}_{\text{task}}$, the M_ϕ is a metric embedding net, and D_k denotes different source domains.

MASF(Model-Agnostic Learning of Semantic Feature)

Algorithm 1 Model-agnostic learning of semantic features for domain generalization

Input: Source training domains $\mathcal{D} = \{D_k\}_{k=1}^{K}$; hyperparameters $\alpha, \eta, \gamma, \beta_1, \beta_2 > 0$

Output: Feature extractor F_ψ, task network T_θ, embedding network M_ϕ

1: **repeat**

2: Randomly split source domains $\mathcal{D}$ into disjoint meta-train $\mathcal{D}_{\text{tr}}$ and meta-test $\mathcal{D}_{\text{te}}$

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

4: Compute global class alignment loss:

$\mathcal{L}_{\text{global}} \leftarrow \frac{1}{|\mathcal{D}_{\text{tr}}|} \sum_{D_i \in \mathcal{D}_{\text{tr}}} \frac{1}{|\mathcal{D}_{\text{te}}|} \sum_{D_j \in \mathcal{D}_{\text{te}}} \ell_{\text{global}}(D_i, D_j; \psi', \theta')$ // Section 3.2

5: Compute local sample clustering loss:

$\mathcal{L}_{\text{local}}(\mathcal{D}; \psi', \phi) \leftarrow \mathbb{E}_{\mathcal{D}}[\ell_{\text{con}}^{n,m}]$ or $\mathbb{E}_{\mathcal{D}}[\ell_{\text{tri}}^{a,p,n}]$ // Section 3.3

6: $\mathcal{L}_{\text{meta}} \leftarrow \beta_1 \mathcal{L}_{\text{global}} + \beta_2 \mathcal{L}_{\text{local}}$

7: $(\psi, \theta) \leftarrow (\psi, \theta) - \eta \nabla_{\psi,\theta}(\mathcal{L}_{\text{task}} + \mathcal{L}_{\text{meta}})$

8: $\phi \leftarrow \phi - \gamma \nabla_\phi \mathcal{L}_{\text{local}}$

9: **until** convergence

임진혁<eeplearning@unist.ac.kr>

MASF(Model-Agnostic Learning of Semantic Feature)

Algorithm 1 Model-agnostic learning of semantic features for domain generalization

Input: Source training domains $\mathcal{D} = \{D_k\}_{k=1}^{K}$; hyperparameters $\alpha, \eta, \gamma, \beta_1, \beta_2 > 0$
Output: Feature extractor F_ψ, task network T_θ, embedding network M_ϕ

repeat
  Randomly split source domains $\mathcal{D}$ into disjoint meta-train $\mathcal{D}_{\text{tr}}$ and meta-test $\mathcal{D}_{\text{te}}$
  $(\psi', \theta') \leftarrow \boxed{(\psi, \theta)} - \alpha \nabla_{\psi,\theta} \mathcal{L}_{\text{task}}(\mathcal{D}_{\text{tr}}; \psi, \theta)$
  Compute global class alignment loss:
  $\mathcal{L}_{\text{global}} \leftarrow \frac{1}{|\mathcal{D}_{\text{tr}}|} \sum_{D_i \in \mathcal{D}_{\text{tr}}} \frac{1}{|\mathcal{D}_{\text{te}}|} \sum_{D_j \in \mathcal{D}_{\text{te}}} \ell_{\text{global}}(D_i, D_j; \psi', \theta')$ // Section 3.2
  Compute local sample clustering loss:
  $\mathcal{L}_{\text{local}}(\mathcal{D}; \psi', \boxed{\phi}) \leftarrow \mathbb{E}_{\mathcal{D}}[\ell_{\text{con}}^{n,m}]$ or $\mathbb{E}_{\mathcal{D}}[\ell_{\text{tri}}^{a,p,n}]$ // Section 3.3
  $\mathcal{L}_{\text{meta}} \leftarrow \beta_1 \mathcal{L}_{\text{global}} + \beta_2 \mathcal{L}_{\text{local}}$
  $\boxed{(\psi, \theta)} \leftarrow \boxed{(\psi, \theta)} - \eta \nabla_{\psi,\theta}(\mathcal{L}_{\text{task}} + \mathcal{L}_{\text{meta}})$
  $\boxed{\phi} \leftarrow \boxed{\phi} - \gamma \nabla_\phi \mathcal{L}_{\text{local}}$
until convergence

MASF:

Model is composed of Feature Extractor : X -> Z

Task Network : Z -> R^c

- Z: feature space (low dimensional than x)
- C: number of classes in Y

$$F_\psi : \mathcal{X} \rightarrow \mathcal{Z},$$

$$T_\theta : \mathcal{Z} \rightarrow \mathbb{R}^C$$

$$p(y \mid \mathbf{x}; \psi, \theta) = \hat{\mathbf{y}} = \operatorname{softmax}(T_\theta\left(F_\psi(\mathbf{x})\right))$$

임진혁<eeplearning@unist.ac.kr>

MASF:

Just optimized for predictions to task (Task Loss)

Just ψ, θ are optimized about task-specific Loss(Loss(task))

$$\ell_{\text{task}}(y, \hat{\mathbf{y}}) = -\sum_c \mathbf{1}[y = c] \log \hat{y}_c$$

⇒ Of course, it makes the model to be excellent predictor for training domain, and produce highly discriminative features Z

⇒ But it can't prevent the model to be overfitted to source domain & suffering from degradation o unseen test domain

MASF: How to capture general semantic features?

"***Episodic training scheme***" :

To expose the model optimization to distribution mismatch

⇒ *The model is trained on a sequence of simulated episode with domain shift*

: At each iteration, available domains D are randomly split into sets of meta-train and meta-test domains.

: be trained to semantically perform well on held-out test domain after being optimized with one or more steps of gradient descent with train domains.

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

임진혁<eeplearning@unist.ac.kr>

MASF: How to capture general semantic features?

Algorithm 1 Model-agnostic learning of semantic features for domain generalization

Input: Source training domains $\mathcal{D} = \{D_k\}_{k=1}^{K}$; hyperparameters $\alpha, \eta, \gamma, \beta_1, \beta_2 > 0$
Output: Feature extractor F_ψ, task network T_θ, embedding network M_ϕ

1: **repeat**
2: Randomly split source domains $\mathcal{D}$ into disjoint meta-train $\mathcal{D}_{\text{tr}}$ and meta-test $\mathcal{D}_{\text{te}}$
3: $\boxed{(\psi', \theta') \leftarrow (\psi, \theta) - \alpha \nabla_{\psi,\theta} \mathcal{L}_{\text{task}}(\mathcal{D}_{\text{tr}}; \psi, \theta)}$
4: Compute global class alignment loss:
$\mathcal{L}_{\text{global}} \leftarrow \frac{1}{|\mathcal{D}_{\text{tr}}|} \sum_{D_i \in \mathcal{D}_{\text{tr}}} \frac{1}{|\mathcal{D}_{\text{te}}|} \sum_{D_j \in \mathcal{D}_{\text{te}}} \ell_{\text{global}}(D_i, D_j; \psi', \theta')$ // Section 3.2
5: Compute local sample clustering loss:
$\mathcal{L}_{\text{local}}(\mathcal{D}; \psi', \phi) \leftarrow \mathbb{E}_{\mathcal{D}}[\ell_{\text{con}}^{n,m}]$ or $\mathbb{E}_{\mathcal{D}}[\ell_{\text{tri}}^{a,p,n}]$ // Section 3.3
6: $\mathcal{L}_{\text{meta}} \leftarrow \beta_1 \mathcal{L}_{\text{global}} + \beta_2 \mathcal{L}_{\text{local}}$
7: $(\psi, \theta) \leftarrow (\psi, \theta) - \eta \nabla_{\psi,\theta}(\mathcal{L}_{\text{task}} + \mathcal{L}_{\text{meta}})$
8: $\phi \leftarrow \phi - \gamma \nabla_\phi \mathcal{L}_{\text{local}}$
9: **until** convergence

임진혁<eeplearning@unist.ac.kr>

Proposed Method

MASF: How to capture general semantic features?

Algorithm 1 Model-agnostic learning of semantic features for domain generalization

Input: Source training domains $\mathcal{D} = \{D_k\}_{k=1}^{K}$; hyperparameters $\alpha, \eta, \gamma, \beta_1, \beta_2 > 0$

Output: Feature extractor F_ψ, task network T_θ, embedding network M_ϕ

1: **repeat**

2: Randomly split source domains $\mathcal{D}$ into disjoint meta-train $\mathcal{D}_{tr}$ and meta-test $\mathcal{D}_{te}$

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

4: Compute global class alignment loss:
$\mathcal{L}_{global} \leftarrow \frac{1}{|\mathcal{D}_{tr}|} \sum_{D_i \in \mathcal{D}_{tr}} \frac{1}{|\mathcal{D}_{te}|} \sum_{D_j \in \mathcal{D}_{te}} \ell_{global}(D_i, D_j; \psi', \theta')$ // Section 3.2

5: Compute local sample clustering loss:
$\mathcal{L}_{local}(\mathcal{D}; \psi', \phi) \leftarrow \mathbb{E}_{\mathcal{D}}[\ell_{con}^{n,m}]$ or $\mathbb{E}_{\mathcal{D}}[\ell_{tri}^{a,p,n}]$ // Section 3.3

6: $\mathcal{L}_{meta} \leftarrow \beta_1 \mathcal{L}_{global} + \beta_2 \mathcal{L}_{local}$

7: $(\psi, \theta) \leftarrow (\psi, \theta) - \eta \nabla_{\psi,\theta} (\mathcal{L}_{task} + \mathcal{L}_{meta})$

8: $\phi \leftarrow \phi - \gamma \nabla_\phi \mathcal{L}_{local}$

9: **until** convergence

To make the objective function quantifying these properties(Learning general semantic features),
Design Loss(meta) and this LOSS is computed based on the updated parameters (Loss(task))

: Semantic Meta Objective Loss(meta) composed of Loss(global) and Loss(local) with weighting coefficients

임진혁<eeplearning@unist.ac.kr>

MASF: **How** to capture general semantic features?

Semantic Meta Objective Loss(meta)

Enforce the feature space to encode "semantically relevant" properties

- inter-class relationship from different domains

- ***Features should be cluster by class labels regardless of domains***

Loss(global)

- Transferable knowledge

(class relationships information preserved from unseen domain feature space)

Loss(local)

- Robust semantic features

(locally cluster(class) regardless of the domain)

(class boundary information with respect to samples) (class-specific cohesion and separation)

임진혁<eeplearning@unist.ac.kr>

Semantic Meta Objective Loss(meta)

Loss(global)

- **Inspired by Knowledge Distillation**
- **Class Ambiguities: form of per-class soft label**
- **[The class feature of each domain] is mean feature vectors so it represent current 'concept' of each class**

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

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

\`

- **This form of class relationship information represent "inter-class relationships" of particular domain**
- **And we want this information same between train domain and test domain**

임진혁<eeplearning@unist.ac.kr>

2 Proposed Method

Semantic Meta Objective Loss(meta)

Loss(global)

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

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

`

- **we want this information same between train domain and test domain**

minimising their symmetrized Kullback–Leibler (KL) divergence, averaged over all C classes:

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

임진혁<eeplearning@unist.ac.kr>

Semantic Meta Objective Loss(meta)

Loss(local)

- **Robust semantic features for class-specific clusters reducing their overlap (For the problem of ambiguous decision boundaries) : complementary to global structure**
- **+ still be sensitive to unseen domain shift**
- **=> samples in same class lie close to each other and away form different class**

between feature vectors (rather than between raw inputs):

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

- **Use regularization but not to Loss(task) or Loss(global) worry for hurting performance about domain shift**
- **Use Metric-Learning to extracted features Z with embedding network M**
- **Sample pairs randomly drawn from all source domains**

Proposed Method

Semantic Meta Objective Loss(meta)

Loss(local)

between feature vectors (rather than between raw inputs):

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

- **Use Metric-Learning to extracted features Z with embedding network M**
- **Sample pairs randomly drawn from all source domains**

Contrastive Loss

- **In case of mild domain shift**

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

- Attracting samples of the same class and replling samples of different class

Triplet Loss

- **In case of extreme domain shift**

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

- Given ‘anchor’ a, ‘positive’ p, ‘negative’ n to make samples from same closer than different

임진혁<eeplearning@unist.ac.kr>

Proposed Method

Semantic Meta Objective Loss(meta)

Enforce the feature space to encode “semantically relevant” properties

inter-class relationship from different domains

Features should be cluster by class labels regardless of domains

Algorithm 1 Model-agnostic learning of semantic features for domain generalization

Input: Source training domains $\mathcal{D} = \{D_k\}_{k=1}^{K}$; hyperparameters $\alpha, \eta, \gamma, \beta_1, \beta_2 > 0$
Output: Feature extractor F_ψ, task network T_θ, embedding network M_ϕ

1: **repeat**
2: Randomly split source domains $\mathcal{D}$ into disjoint meta-train $\mathcal{D}_{\mathrm{tr}}$ and meta-test $\mathcal{D}_{\mathrm{te}}$
3: $(\psi', \theta') \leftarrow \boxed{(\psi, \theta)} - \alpha \nabla_{\psi,\theta} \mathcal{L}_{\mathrm{task}}(\mathcal{D}_{\mathrm{tr}}; \psi, \theta)$
4: Compute global class alignment loss:
$\mathcal{L}_{\mathrm{global}} \leftarrow \frac{1}{|\mathcal{D}_{\mathrm{tr}}|} \sum_{D_i \in \mathcal{D}_{\mathrm{tr}}} \frac{1}{|\mathcal{D}_{\mathrm{te}}|} \sum_{D_j \in \mathcal{D}_{\mathrm{te}}} \ell_{\mathrm{global}}(D_i, D_j; \psi', \theta')$ // Section 3.2
5: Compute local sample clustering loss:
$\mathcal{L}_{\mathrm{local}}(\mathcal{D}; \psi', \boxed{\phi}) \leftarrow \mathbb{E}_{\mathcal{D}}[\ell_{\mathrm{con}}^{n,m}]$ or $\mathbb{E}_{\mathcal{D}}[\ell_{\mathrm{tri}}^{a,p,n}]$ // Section 3.3
6: $\mathcal{L}_{\mathrm{meta}} \leftarrow \beta_1 \mathcal{L}_{\mathrm{global}} + \beta_2 \mathcal{L}_{\mathrm{local}}$
7: $\boxed{(\psi, \theta)} \leftarrow \boxed{(\psi, \theta)} - \eta \nabla_{\psi,\theta}(\mathcal{L}_{\mathrm{task}} + \mathcal{L}_{\mathrm{meta}})$
8: $\boxed{\phi} \leftarrow \boxed{\phi} - \gamma \nabla_\phi \mathcal{L}_{\mathrm{local}}$
9: **until** convergence

Semantic Meta Objective Loss(meta) composed of Loss(global) and Loss(local) with weighting coefficients

임진혁<eeplearning@unist.ac.kr>

2 Proposed Method

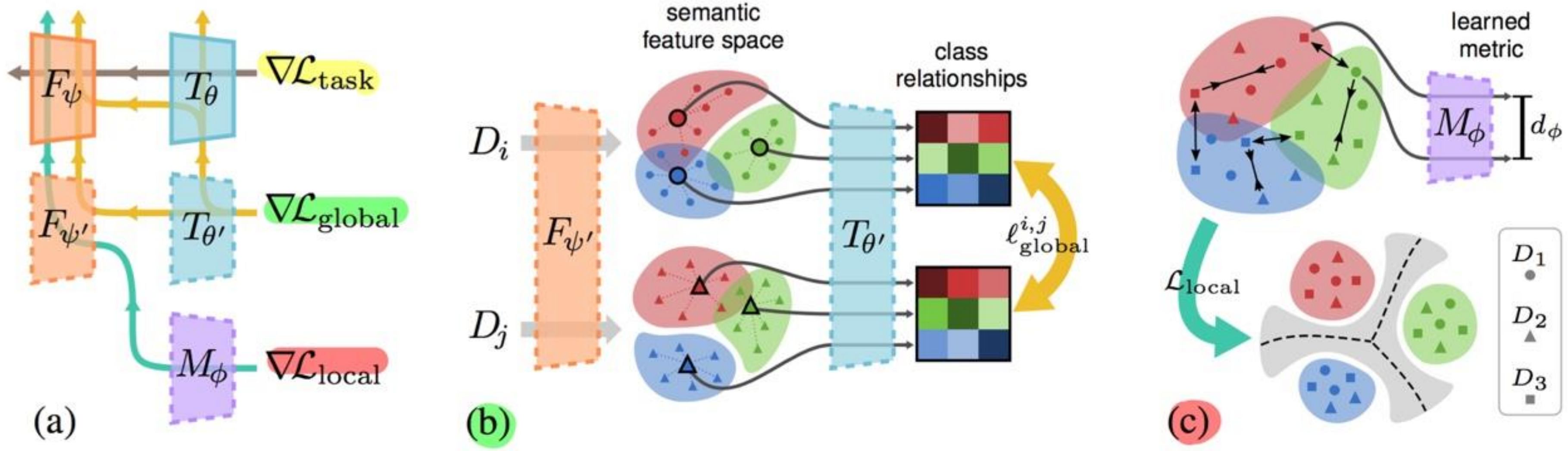


Figure 1: An overview of the proposed model-agnostic learning of semantic features (MASF): (a) episodic training under simulated domain shift, with gradient flows indicated; (b) global alignment of class relationships; (c) local sample clustering, towards cohesion and separation. F_ψ and T_θ are the feature extractor and the task net, $F_{\psi'}$ and $T_{\theta'}$ are their updated versions by inner gradient descent on the task loss $\mathcal{L}_{\text{task}}$, the M_ϕ is a metric embedding net, and D_k denotes different source domains.

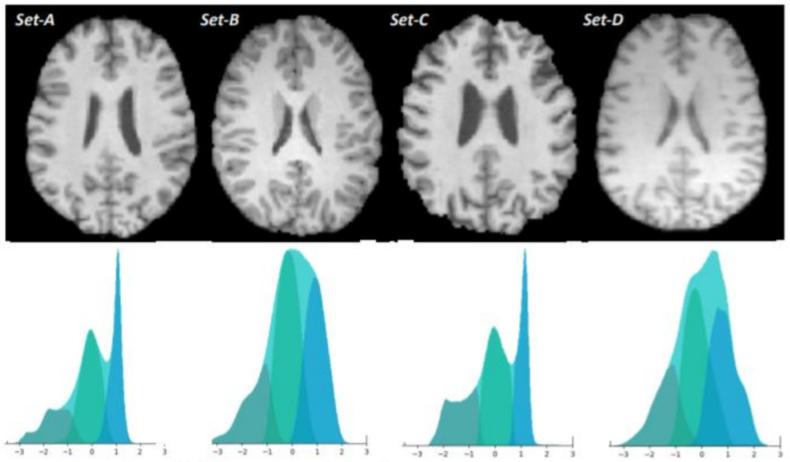
Experiment

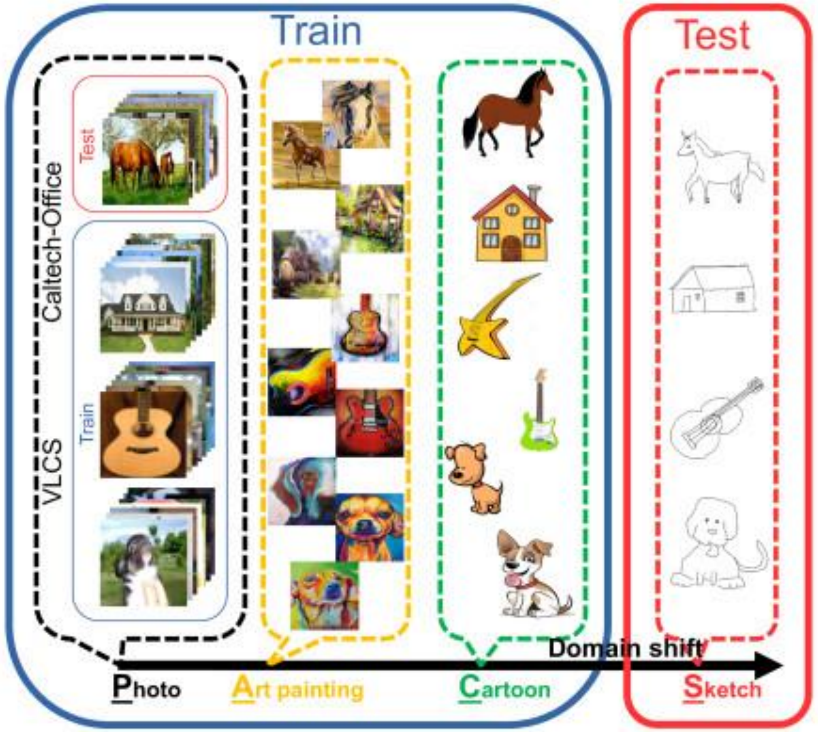
- Leave-one-domain-out cross-validation
- 3 Independent run and average it

(1) VLCS Dataset

(2) PACS Dataset

(3) Tissue Segmentation in Multi-site Brain MRI





Experiment

(1) VLCS Dataset

- The classic domain generalization benchmark for image classification
- Includes images from **four datasets**
- The multi-class object recognition task includes **five classes**: bird, car, chair, dog and person
- Randomly dividing each domain into 70% training and 30% test

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

- **Baseline:** merging all source domains and training

임진혁<eeplearning@unist.ac.kr>

(2) PACS Dataset

- Recent benchmark with more severe distribution shift between domains
- Four domains: art painting, cartoon, photo, sketch,
- Seven classes: dog, elephant, giraffe, guitar, house, horse, person
- leave-one-domain-out cross-validation
- adopt an AlexNet pre-trained on ImageNet
- The multi-class object recognition task includes **five classes**: bird, car, chair, dog and person

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

임진혁<eeplearning@unist.ac.kr>

Experiment

(2) PACS Dataset

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

- [1, 26, 27] exposing the training procedure to domain shift benefits model generalization to unseen domains.
- **MASF** considers the semantic structure
- regarding both **global** class alignment and **local** sample clustering
- Across all domains, **MASF** increases average accuracy by 3.51% over the baseline
- **Note** when the unseen domain is **sketch**
  which has a distinct style and requires more general knowledge about semantic concepts

임진혁<eeplearning@unist.ac.kr>

(2) PACS Dataset: Ablation study

For what?

- to investigate 2 key points

- 1) the contribution of each component to our method's performance
- 2) how the semantic feature space is influenced by our proposed meta losses

How?

- test all possible combinations of including the key component

- 1) episodic meta-learning simulating domain shift
- 2) global class alignment loss
- 3) local sample clustering 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

(2) PACS Dataset: t-SNE visualization

- to analzye the feature space learned with our proposed model and the DeepAll baseline

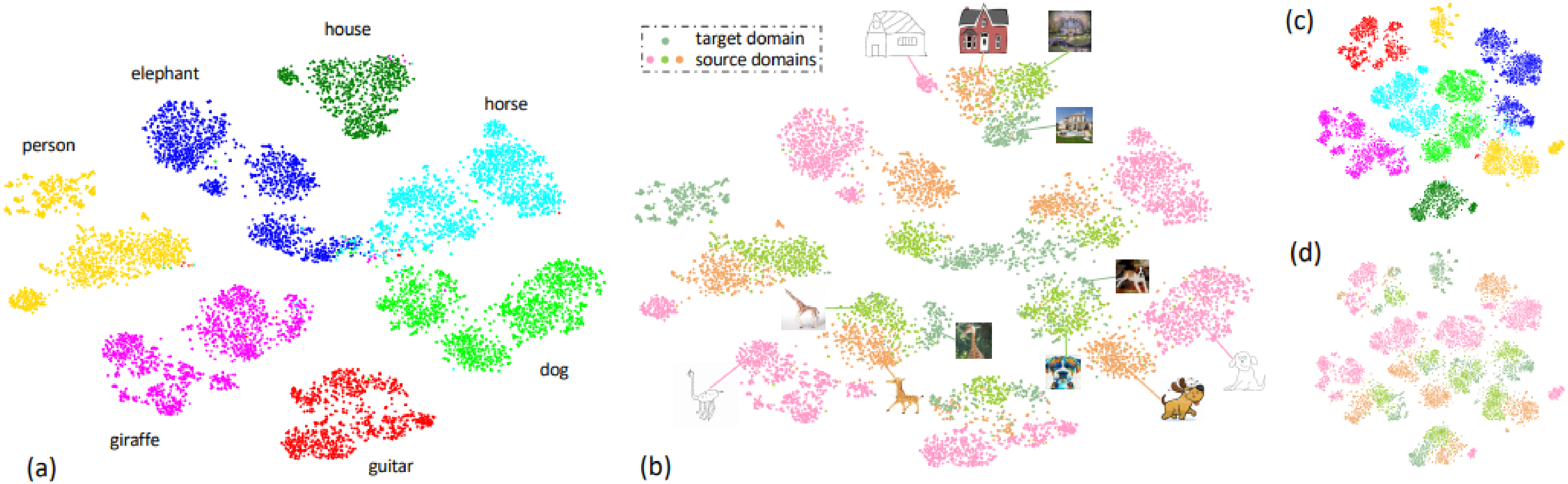
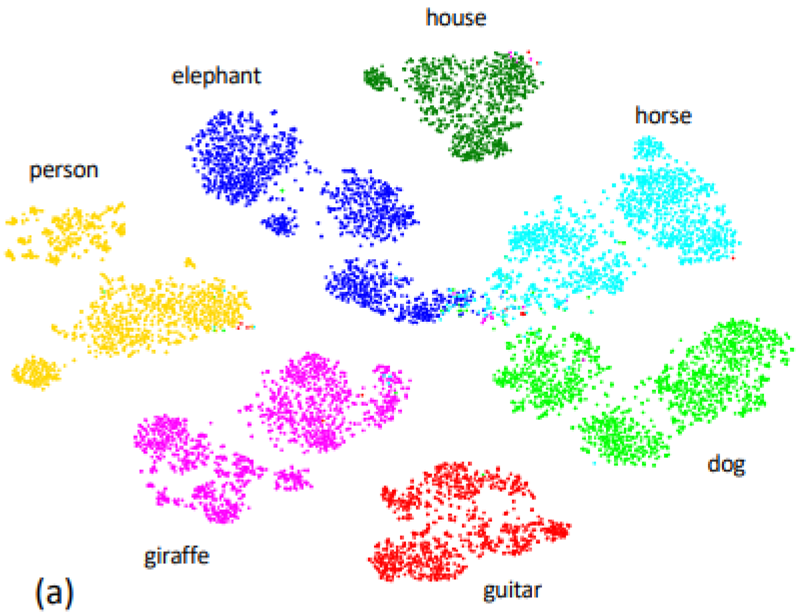
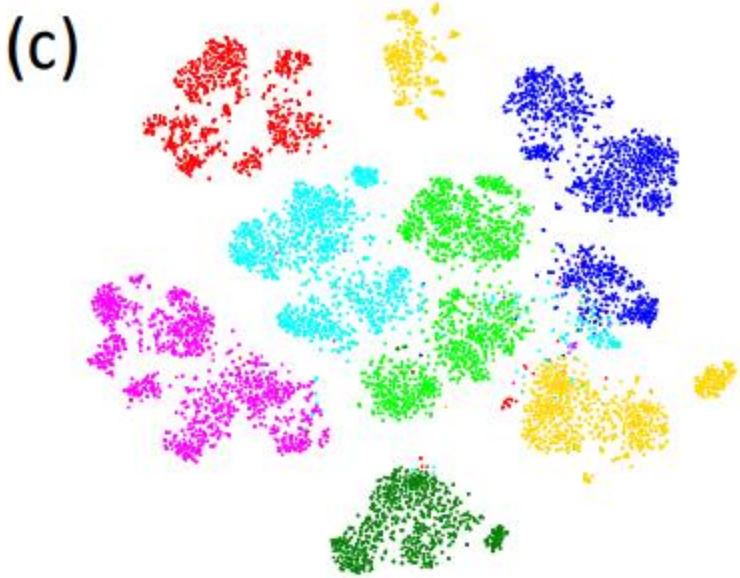


Figure 2: The t-SNE visualization of extracted features from F_ψ, using our proposed (a-b) MASF and the (c-d) DeepAll model on PACS dataset. In (a) and (c), the different colors indicate different classes; correspondingly in (b) and (d), the different colors indicate different domains.

(2) PACS Dataset: t-SNE visualization



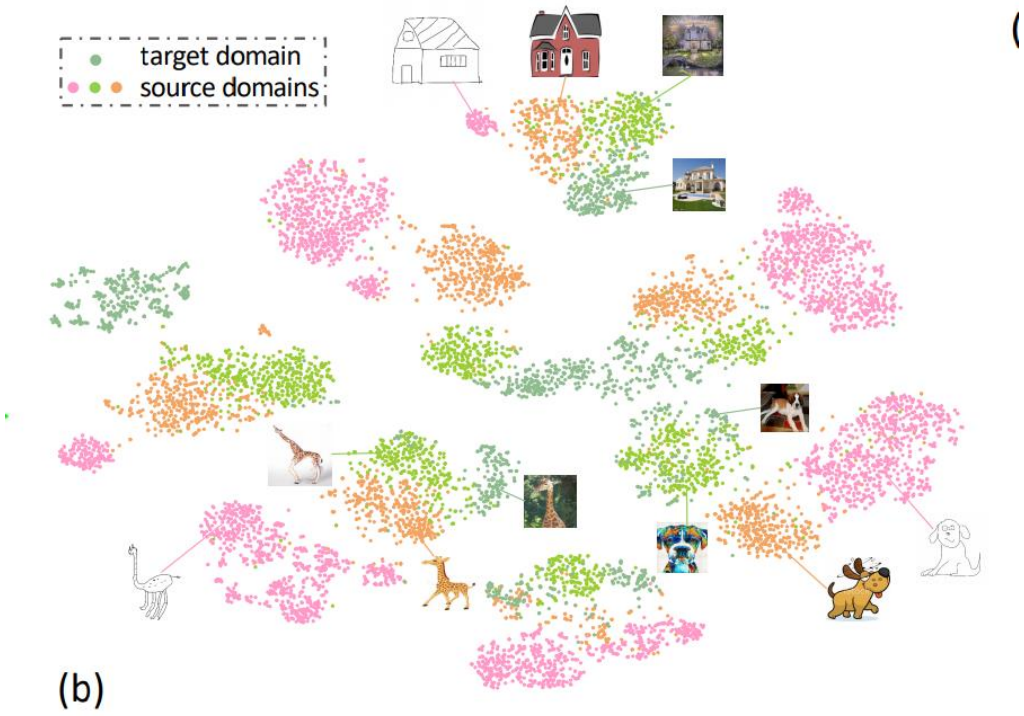


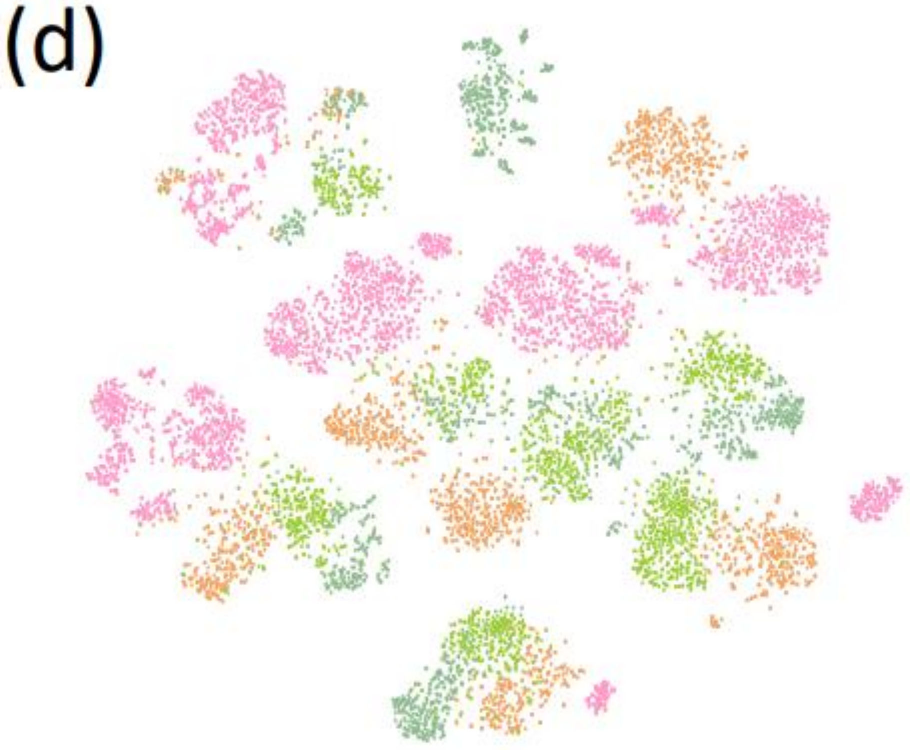
(a) vs (c)

Different color => Different class

임진혁<eeplearning@unist.ac.kr>

(2) PACS Dataset: t-SNE visualization





(b) vs (d)
Different color => Different domain

- MASF model yields a better separation of classes
- note that the sketch domain is further apart from art painting and cartoon

Deeper architectures

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

- This suggests our proposed algorithm is also beneficial for domain generalization with **deeper** feature extractors.

Experiment

(3) Tissue Segmentation in Multi-site Brain MRI

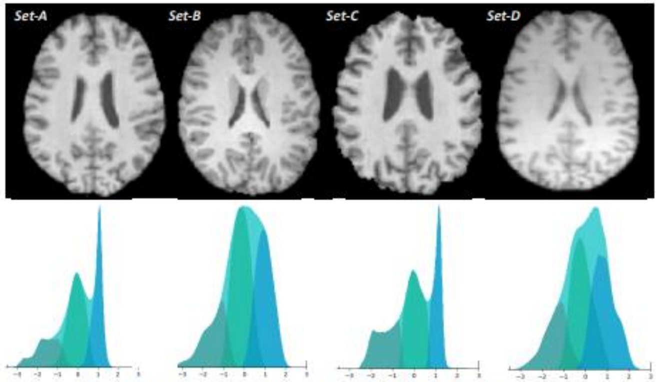


Figure 4: Different brain MRI datasets with example images and intensity histograms.

Table 5: Evaluation of brain tissue segmentation (Dice coefficient, %) in different settings: *columns 1–4:* train model on single source domain, test on all domains; *columns 5–6:* train on three source domains, test on remaining domain.

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

- Real-world medical imaging task of brain tissue segmentation (MRI)
- Data was acquired from four clinical centers (denoted as Set-A/B/C/D).
- Domain shift occurs due to differences in scanners
- Randomly split each domain to 80% for training and 20% for testing in experimental settings.

4 Conclusion

- domain generalization

- by incorporating global and local constraints
- for learning semantic feature spaces.

임진혁<eeplearning@unist.ac.kr>

캄사...