

# 英語論文#5

2023/10/30

M1

建元 了

# 論文の概要

- タイトル
  - **Semi-supervised bidirectional alignment for Remote Sensing cross-domain scene classification**
- 執筆者
  - Wei Huang, Yilei Shi, Zhitong Xiong, Qi Wang, Xiao Xiang Zhu
- 掲載
  - ISPRS Journal of Photogrammetry and Remote Sensing cross-domain
- 選択理由
  - 最近のリモートセンシングに関するSSLと分類問題を扱う論文として調査

# 背景

- リモートセンシング（RS）画像シーン分類は、土地利用や都市計画など、広範な応用の可能性からより注目されている
- 利用可能なRS画像が増加するにつれ、ラベルがつけられていないデータをどう活用するかが課題となっている

# 導入

- RS画像シーン分類のほとんどの方法は、教師ありの深層学習モデルに基づいたラベル付きのデータに依存している
- ほかの半教師あり学習を扱うモデルでは、ラベルなしデータの自己学習によってアノテーションへの大きな依存を軽減する手法が試みられている

# 導入

- 異なるデータセット間ではドメインが大幅にシフトするためにシーン分類に適応するのは困難
- 教師なしドメイン適応(UDA)と半教師ありドメイン適応(SSDA)を用いて、ドメインのシフトを軽減する

# ドメイン適応(domain Adaptation)

- ドメイン適応 (domain adaptation)

- 学習時の訓練データ (Source domain) と推論時のデータ (Target domain) でドメインが異なる場合、一般的に悪くなる
- このドメインが変化する場合でも推論をうまく行えることを目指す

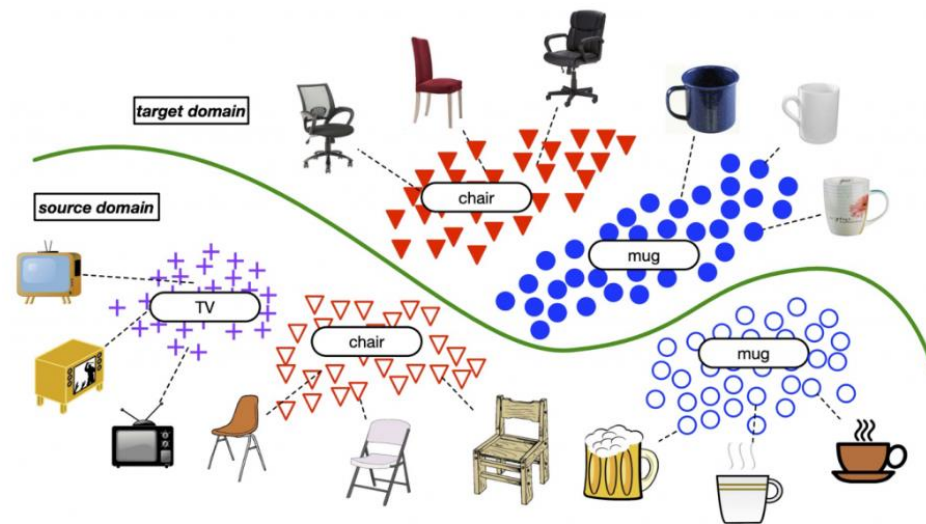


図 4 ドメイン適応のシナリオの概要 ([Cao 18] より引用)。

# 提案手法

- **bidirectional sample-class alignment(BSCA)**

- ラベル付けされたソースドメインS

$$D_s = \{(\mathbf{x}^s, y^s)\}_{i=1}^{N_s}$$

- 限定されたラベルがつけられたターゲットドメインT

$$D_l = \{(\mathbf{x}^l, y^l)\}_{i=1}^{N_l},$$

- 完全にラベルないターゲットドメインU

$$D_u = \{(\mathbf{x}^u)\}_{i=1}^{N_u},$$

# shared network architecture

- 画像  $x$  から特徴抽出器  $\mathcal{E}$  によって特徴マップ  $m$  を出力

$$\mathbf{m} = \mathcal{E}(\mathbf{x}).$$

- $m$  からGAP層を通してグローバル特徴ベクトル  $f$  を計算

$$\mathbf{f} = \mathcal{G}(\mathbf{m}) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W m_{ij}.$$

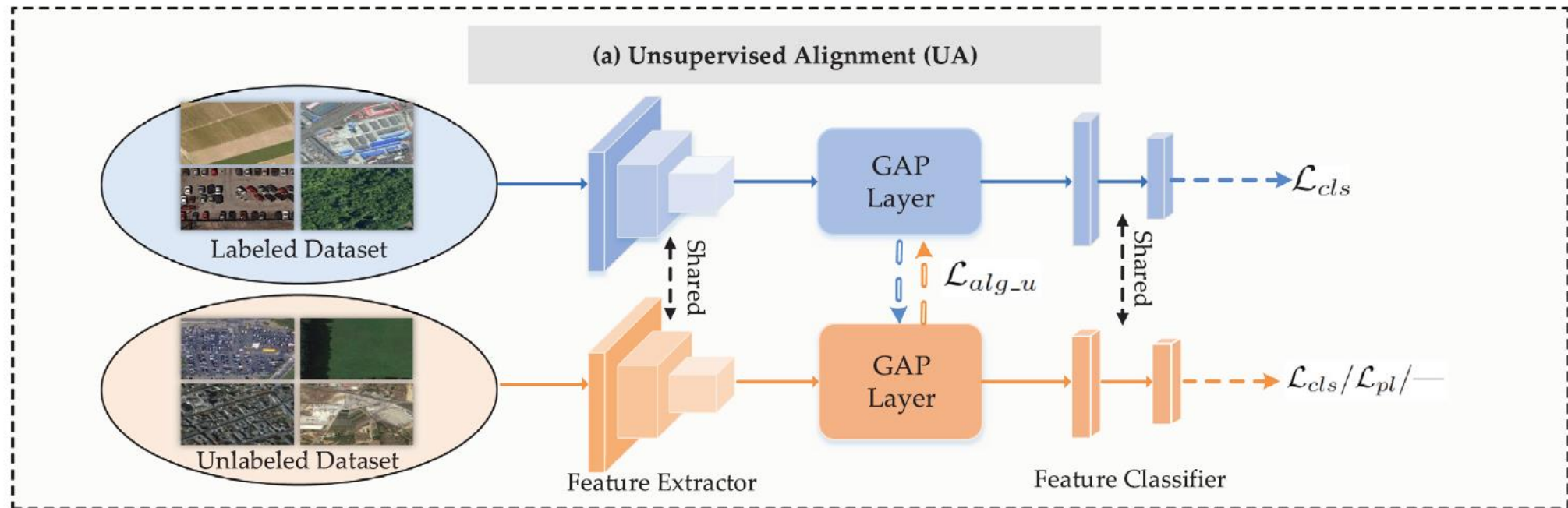
- $f$  を分類器に入力し、クラスごとの予測ベクトル  $p$  を取得

$$\begin{cases} \mathbf{f} = \frac{\mathbf{f}}{\|\mathbf{f}\|}, \\ \mathbf{p} = \sigma(C(\mathbf{f})/T), \end{cases}$$



# Bidirectional sample-class alignment

- **Unsupervised Alignment(UA)**



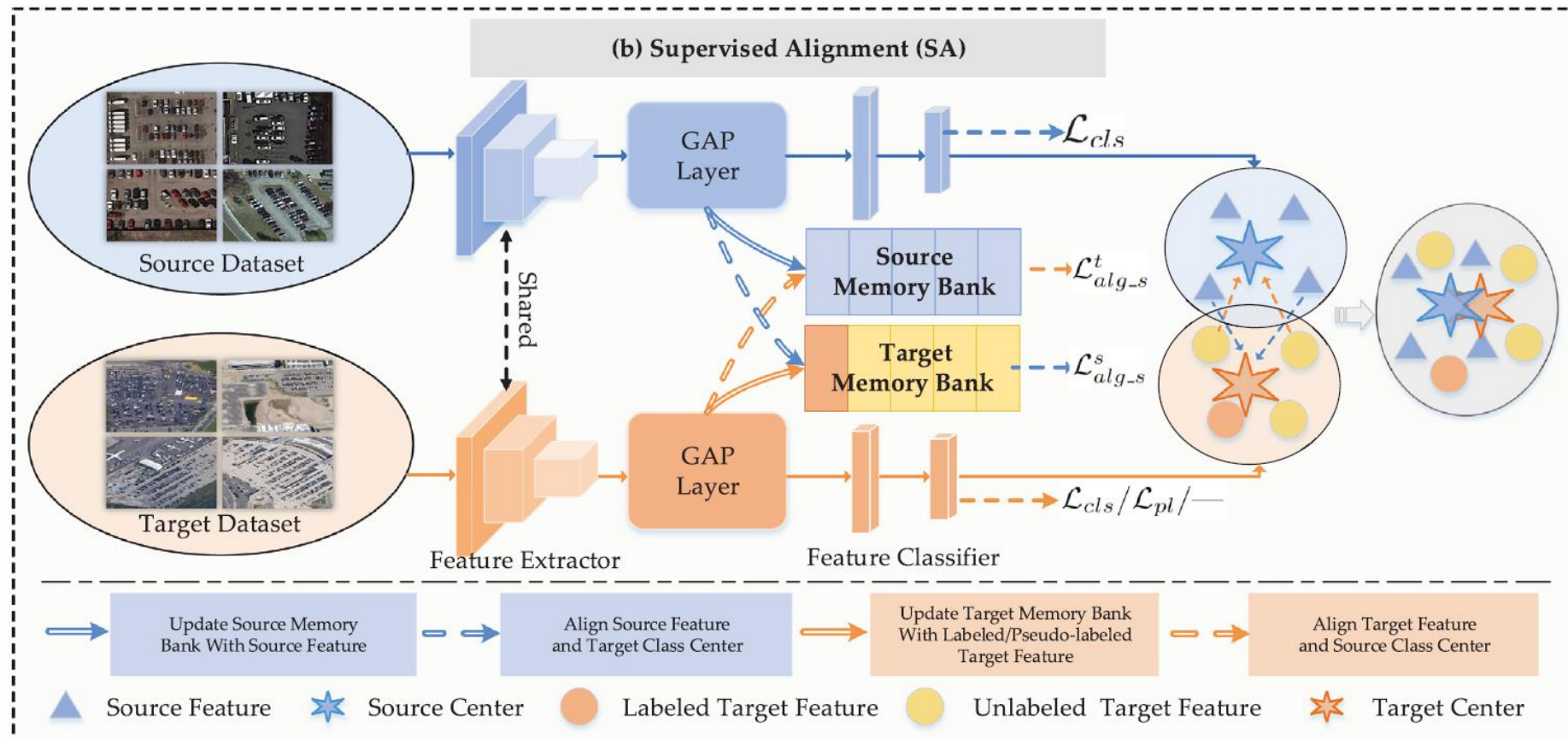
# Unsupervised Alignment(UA)

- ラベル付きサンプルとラベルなしサンプルから抽出された特徴との間の教師なしドメイン間アライメントに注目
- カーネルヒルベルト空間(RKBS)におけるMMDの距離を 2 つの分布の平均値の類似性の評価に使用

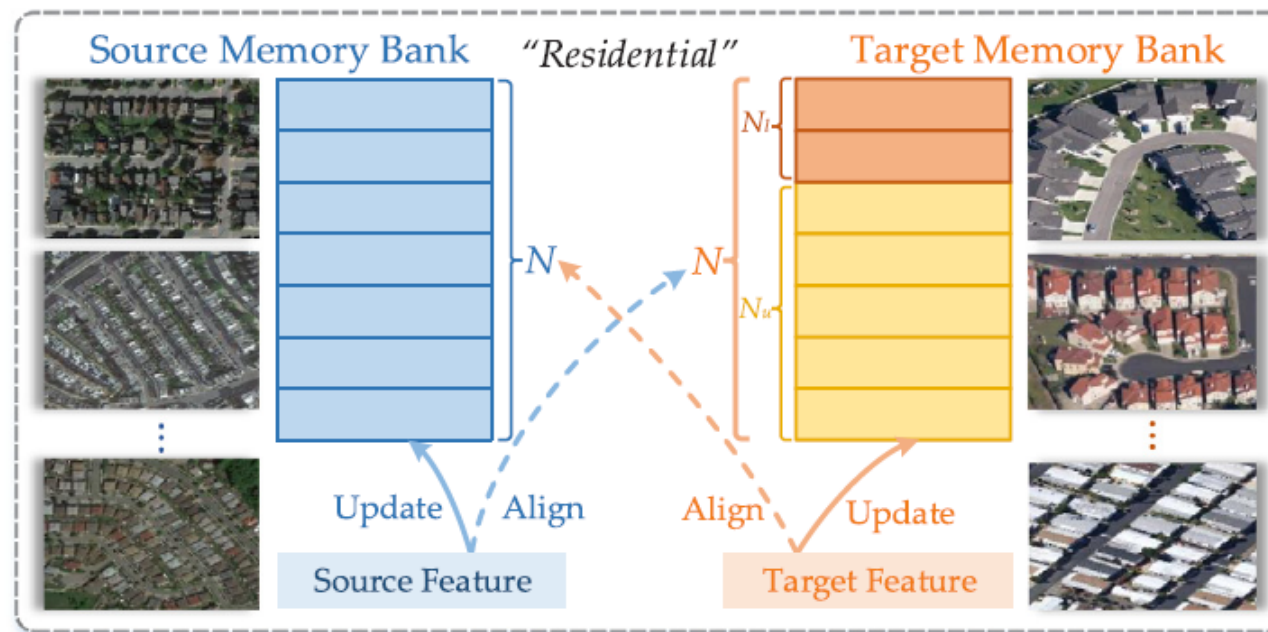
$$\mathcal{L}_{alg\_u} = d_{MMD}^2 = \left\| \frac{1}{N_s + N_l} \sum_{i=1}^{N_s + N_l} \phi(\mathbf{f}_i^{sl}) - \frac{1}{N_u} \sum_{i=1}^{N_u} \phi(\mathbf{f}_i^u) \right\|_{\mathcal{H}}^2,$$

# Bidirectional sample-class alignment

- **Supervised Alignmet(SA)**



# Supervised Alignment(SA)



**Fig. 2.** The illustration of SA for RS cross-domain scene classification, with the class “residential” as an example. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

# Supervised Alignmet(SA)

## 1. 特徴とラベルのペアを生成

- $[f_s, y_s]$ と $[f_l, y_l]$ のペアを生成
- $f_u$ は確率的予測ベクトルから最大のクラスを疑似ラベルとして採用し、 $[f_u, p_u, y_u]$ のペアを生成

$$\hat{y}^u = \operatorname{argmax}_{i=\{1,\dots,K\}} p_i^u,$$

# Supervised Alignment(SA)

## 2. 特徴とラベルのペアを使用してメモリバンクの更新

---

**Algorithm 1** Updating Memory Banks

---

**Input:** Feature-label pairs:  $\{[f^s, y^s]\}$ ,  $\{[f^l, y^l]\}$ , and  $\{[f^u, p^u, \hat{y}^u]\}$ , the size of a mini-batch  $M$ , the entropy threshold  $\tau$ , memory banks:  $\mathcal{M}^s$ ,  $\mathcal{M}^l$ , and  $\mathcal{M}^u$

**Output:** updated memory banks:  $\mathcal{M}^s$ ,  $\mathcal{M}^l$ , and  $\mathcal{M}^u$

```
1: while  $i < M$  do
2:   Update:  $\mathcal{M}^s[y_i^s, 1:N-1] \leftarrow \mathcal{M}^s[y_i^s, 2:N]$ 
3:   Update:  $\mathcal{M}^s[y_i^s, N] \leftarrow f_i^s$ 
4: end while
5: while  $i < M$  do
6:   Update:  $\mathcal{M}^l[y_i^l, 1:N_l-1] \leftarrow \mathcal{M}^l[y_i^l, 2:N_l]$ 
7:   Update:  $\mathcal{M}^l[y_i^l, N_l] \leftarrow f_i^l$ 
8: end while
9: while  $i < M$  do
10:  if  $E_p < \tau$  then
11:    Update:  $\mathcal{M}^u[\hat{y}_i^u, 1:N_u-1] \leftarrow \mathcal{M}^u[\hat{y}_i^u, 2:N_u]$ 
12:    Update:  $\mathcal{M}^u[\hat{y}_i^u, N_u] \leftarrow f_i^u$ 
13:  end if
14: end while
15: return  $\mathcal{M}^s$ ,  $\mathcal{M}^l$ , and  $\mathcal{M}^u$ 
```

---

# Supervised Alignmet(SA)

- 閾値を超えない疑似ラベルのサンプルはラベルのないターゲットドメインの重みを改善するようにネットワークをトレーニングすることもできる

$$\begin{cases} E_p = - \sum_{i=1}^K \mathbf{p}_i^u \log \mathbf{p}_i^u, \\ \mathcal{L}_{pl} = - \sum_{\mathbf{x}^u \in D_u} 1\{E_p < \tau\} \cdot \hat{y}^u \log \mathbf{p}_{\hat{y}^u}^u, \end{cases}$$

# Supervised Alignmet(SA)

## 3. クロスドメインのサンプルクラスの調整

- ソースからターゲット、ターゲットからソースへの両方で、特徴とクラスのクロスドメイン調整を行う

$$\begin{cases} \mathbf{c}_k^s = \frac{1}{N} \sum_{i=1}^N \mathcal{M}_{ki}^s, \\ \mathbf{c}_k^t = \frac{1}{N} \left( \sum_{i=1}^{N_l} \mathcal{M}_{ki}^l + \sum_{i=1}^{N_u} \mathcal{M}_{ki}^u \right). \end{cases}$$



# Supervised Alignment(SA)

- 全体のSA損失を教師ありアライメント損失、ターゲット特徴アライメント損失から計算

$$\begin{cases} \mathcal{L}_{alg\_s}^s = \frac{1}{M} \sum_{i=1}^M (\mathbf{f}_i^s - \mathbf{c}_{y_i^s}^t), \\ \mathcal{L}_{alg\_s}^t = \frac{1}{M} \left[ \sum_{i=1}^{M^l} (\mathbf{f}_i^l - \mathbf{c}_{y_i^l}^t) + \sum_{i=1}^{M^u} (\mathbf{f}_i^u - \mathbf{c}_{\hat{y}_i^u}^s) \right], \quad \hat{y}_i^u > \tau, \\ \mathcal{L}_{alg\_s} = \mathcal{L}_{alg\_s}^s + \mathcal{L}_{alg\_s}^t. \end{cases}$$

# Overall loss

- BSCAベースのSSDAの全体的な損失は、教師あり損失、半教師あり損失、教師なしアライメント損失、教師ありクロスドメインアライメント損失の合計

$$\mathcal{L} = \mathcal{L}_{cls} + \mathcal{L}_{pl} + \alpha \mathcal{L}_{alg_u} + \beta \mathcal{L}_{alg_s},$$

# 評価

- 4つのRS用ベンチマークで検証



# 評価

**Table 1**

Characteristics of RS image scene classification sets.

	NWPU-RESISC45	AID	RSSCN7	WHU-RS19
Years	2017	2016	2015	2010
Resolution(m)	0.2–30	0.5–8	–	0.5
Image size	$256 \times 256$	$600 \times 600$	$400 \times 400$	$600 \times 600$
Class 1: <i>Farmland</i>	700	370	400	50
Class 2: <i>Forest</i>	700	250	400	53
Class 3: <i>Industrial</i>	700	390	400	53
Class 4: <i>Meadow</i>	700	280	400	61
Class 5: <i>Parking</i>	700	390	400	50
Class 6: <i>Residential</i>	700	410	400	54
Class 7: <i>River</i>	700	410	400	56

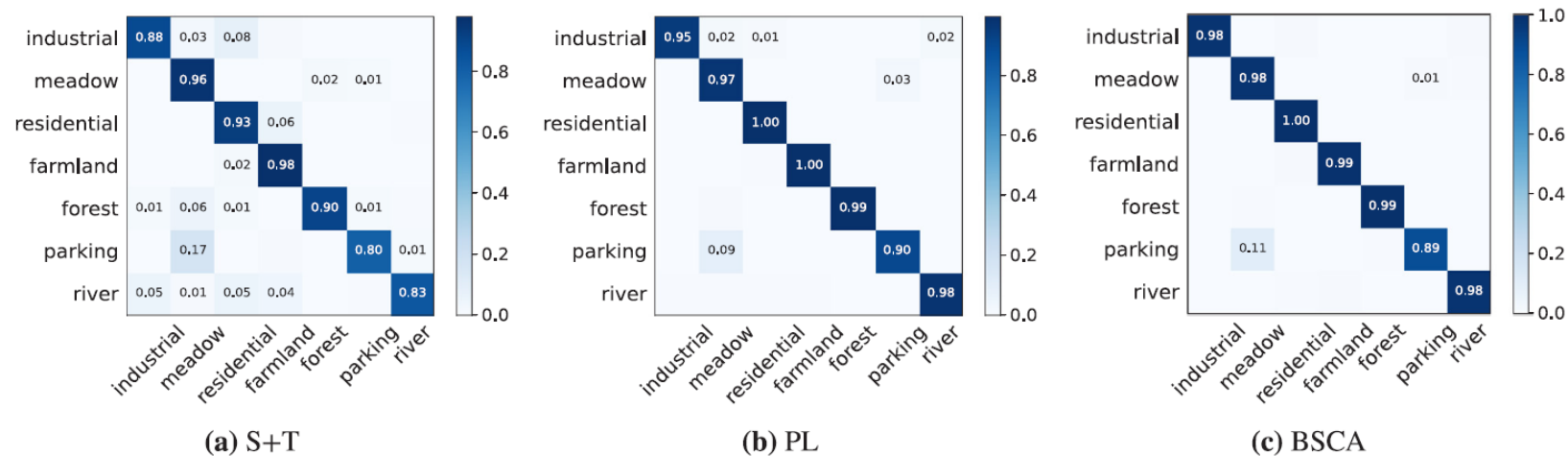
# 評価

- ablation study

**Table 2**

Classification accuracy (%) of ablation study under the 3-shot setting with **ResNet34** as the CNN backbone.

$\mathcal{L}_{cls}$	$\mathcal{L}_{pl}$	$\mathcal{L}_{alg_s}$	$\mathcal{L}_{alg_u}$	A→N	R→N	W→N	N→A	R→A	W→A	N→R	A→R	W→R	N→W	A→W	R→W	Mean
✓				90.7	77.4	84.3	93.3	86.0	94.2	75.9	72.5	72.9	89.9	98.3	92.5	85.7
✓	✓			94.6	86.5	94.2	95.9	94.7	97.2	78.9	79.0	81.6	97.6	97.9	93.4	91.0 (+5.3)
✓	✓	✓		91.9	<b>89.8</b>	93.2	96.7	96.0	96.7	80.5	79.9	82.3	<b>97.9</b>	99.0	96.9	91.7 (+6.0)
✓	✓		✓	96.6	89.0	94.5	96.8	96.8	<b>97.6</b>	81.6	81.9	<b>84.1</b>	<b>97.9</b>	99.0	96.2	92.7 (+7.0)
✓	✓	✓	✓	<b>96.8</b>	89.7	<b>94.7</b>	<b>97.2</b>	<b>97.2</b>	97.5	<b>84.1</b>	<b>82.6</b>	84.0	<b>97.9</b>	<b>99.7</b>	<b>98.3</b>	<b>93.3 (+7.6)</b>



**Fig. 4.** The confusion matrices of the methods of S+T, PL, and the proposed BSCA in the N→A adaptation scenario, with ResNet34 as the CNN backbone.

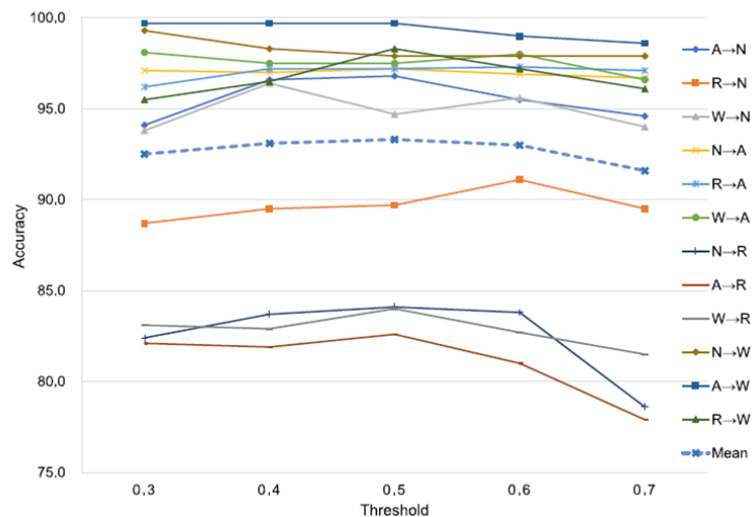
# 評価

- ハイパーパラメータ調整

Table 3

Classification accuracy (%) of threshold tuning experiments under the 3-shot setting with **ResNet34** as the CNN backbone.

Threshold	A→N	R→N	W→N	N→A	R→A	W→A	N→R	A→R	W→R	N→W	A→W	R→W	Mean
0.3	94.1	88.7	93.8	97.1	96.2	<b>98.1</b>	82.4	82.1	83.1	<b>99.3</b>	<b>99.7</b>	95.5	92.5
0.4	96.6	89.5	<b>96.4</b>	97.0	97.2	97.5	83.7	81.9	82.9	98.3	<b>99.7</b>	96.5	93.1
0.5	<b>96.8</b>	89.7	94.7	<b>97.2</b>	97.2	97.5	<b>84.1</b>	<b>82.6</b>	<b>84.0</b>	97.9	<b>99.7</b>	<b>98.3</b>	<b>93.3</b>
0.6	95.5	<b>91.1</b>	95.6	96.9	<b>97.3</b>	98.0	83.8	81.0	82.7	97.9	99.0	97.2	93.0
0.7	94.6	89.5	94.0	96.7	97.1	96.6	78.6	77.9	81.5	97.9	98.6	96.1	91.6

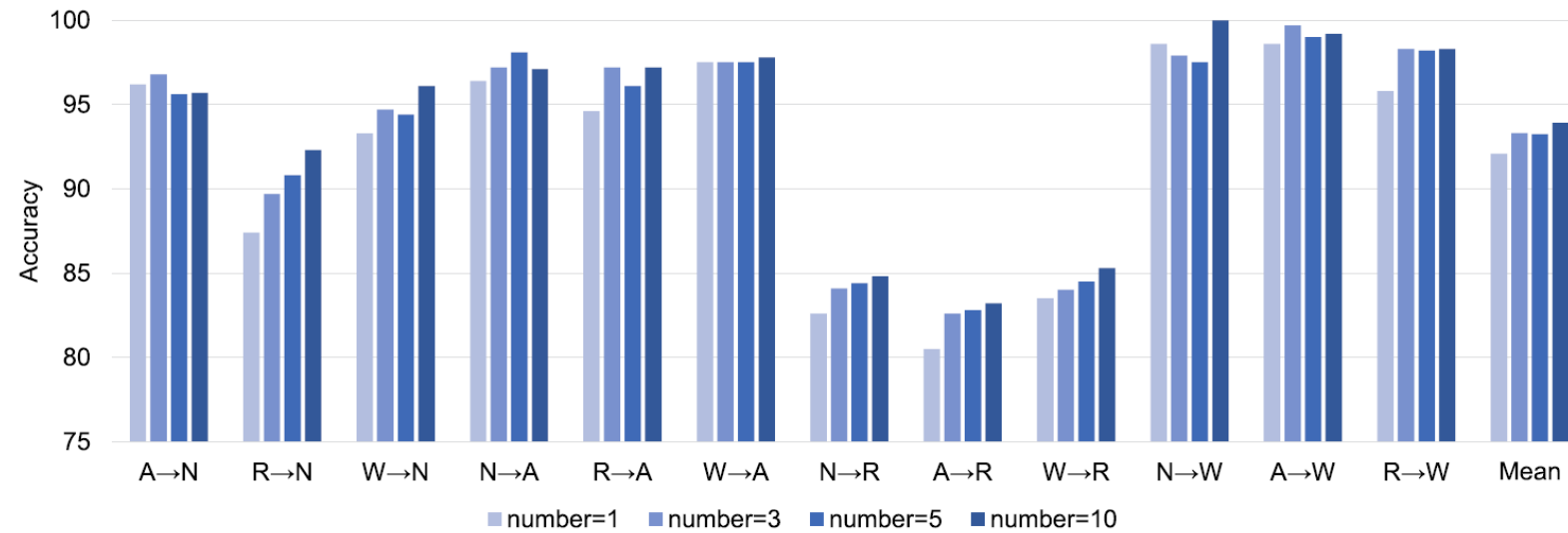


# 評価

**Table 4**

Classification accuracy (%) of different numbers of labeled samples per class with **ResNet34** as the CNN backbone.

Shot number	A→N	R→N	W→N	N→A	R→A	W→A	N→R	A→R	W→R	N→W	A→W	R→W	Mean
1	96.2	87.4	93.3	96.4	94.6	97.5	82.6	80.5	83.5	98.6	98.6	95.8	92.1
3	96.8	89.7	94.7	97.2	97.2	97.5	84.1	82.6	84.0	97.9	99.7	98.3	93.3
5	95.6	90.8	94.4	98.1	96.1	97.5	84.4	82.8	84.5	97.5	99	98.2	93.2
10	95.7	92.3	96.1	97.1	97.2	97.8	84.8	83.2	85.3	100.0	99.2	98.3	93.9



**Fig. 6.** Classification accuracy with different numbers of labeled target samples per class.



# 評価

- 比較実験（ベースライン）

**Table 5**

Classification accuracy (%) of comparison experiments on the 12 adaptation scenarios from the RS-SSDA datasets under the 3-shot setting with **VGG16** as the CNN backbone. The best results are in **bold**.

Type	Method	A→N	R→N	W→N	N→A	R→A	W→A	N→R	A→R	W→R	N→W	A→W	R→W	Mean
SSL	FixMatch (Xiong et al., 2021)		77.9			96.3			67.5			93.4		83.8
UDA	ECB-FAM (Ma et al., 2021b)	90.5	74.6	80.6	93.9	79.5	92.3	69.7	71.0	68.2	94.1	97.7	87.6	83.3
SSDA	S+T	92.6	83.6	88.0	94.1	88.9	93.8	72.9	72.4	76.9	96.9	98.3	92.4	87.6
	ADDA (Wang et al., 2018)	81.8	69.2	73.9	90.5	82.4	77.9	61.8	64.8	66.7	86.1	89.6	80.9	77.1
	RevGrad (Lasloun et al., 2021)	88.2	81.7	90.0	93.9	88.9	89.8	73.8	71.9	77.5	96.9	98.3	93.4	87.0
	SSDAN (Lasloun et al., 2021)	94.6	84.7	89.5	<b>96.8</b>	94.2	95.5	<b>79.6</b>	80.6	81.8	98.3	<b>99.3</b>	96.9	91.0
	ENT (Grandvalet et al., 2005)	86.5	72.0	86.9	92.2	86.0	91.7	60.4	68.0	73.0	80.2	98.6	89.9	82.1
	CDAC (Li et al., 2021)	85.6	74.6	71.7	89.9	81.8	79.5	65.0	70.2	69.8	81.2	89.2	73.6	77.7
	<b>Our BSCA</b>	<b>96.5</b>	<b>94.2</b>	<b>95.9</b>	96.5	<b>94.5</b>	<b>96.5</b>	78.0	<b>82.8</b>	<b>84.7</b>	<b>98.6</b>	99.0	<b>99.3</b>	<b>93.0</b>



# 評価

- 比較実験（ベースライン）

**Table 6**

Classification accuracy (%) of comparison experiments on the 12 adaptation scenarios from the RS-SSDA datasets under the 3-shot setting with **ResNet34** as the CNN backbone. The best results are in **bold**.

Type	Method	A→N	R→N	W→N	N→A	R→A	W→A	N→R	A→R	W→R	N→W	A→W	R→W	Mean
SSL	FixMatch (Xiong et al., 2021)		81.3			95.0			63.5			97.6		84.4
UDA	ECB-FAM (Ma et al., 2021b)	88.5	74.5	83.2	92.5	77.5	92.8	66.9	68.9	65.4	91.7	97.6	79.9	81.6
SSDA	S+T	89.9	76.4	84.4	91.7	84.5	91.6	71.4	68.4	69.4	92.0	97.6	91.0	84.0
	ADDA (Wang et al., 2018)	87.3	78.6	83.1	77.2	76.5	90.7	64.7	72.4	65.0	89.9	94.8	85.8	80.5
	RevGrad (Lasloum et al., 2021)	90.2	80.2	86.9	93.5	86.8	91.7	76.7	73.6	74.1	95.5	98.3	88.5	86.3
	SSDAN (Lasloum et al., 2021)	95.5	86.6	93.5	96.9	95.0	97.3	80.2	80.2	83.6	97.3	99.3	96.2	91.8
	ENT (Grandvalet et al., 2005)	79.5	67.7	77.2	91.2	91.2	87.5	66.6	60.5	62.6	89.9	91.0	83.7	85.4
	CDAC (Li et al., 2021)	85.4	80.6	75.2	83.8	90.0	88.5	65.8	65.5	72.7	86.1	80.6	92.0	79.8
	<b>Our BSCA</b>	<b>96.8</b>	<b>89.7</b>	<b>94.7</b>	<b>97.2</b>	<b>97.2</b>	<b>97.5</b>	<b>84.1</b>	<b>82.6</b>	<b>84.0</b>	<b>97.9</b>	<b>99.7</b>	<b>98.3</b>	<b>93.3</b>

# 評価

- 比較実験（ベースライン）

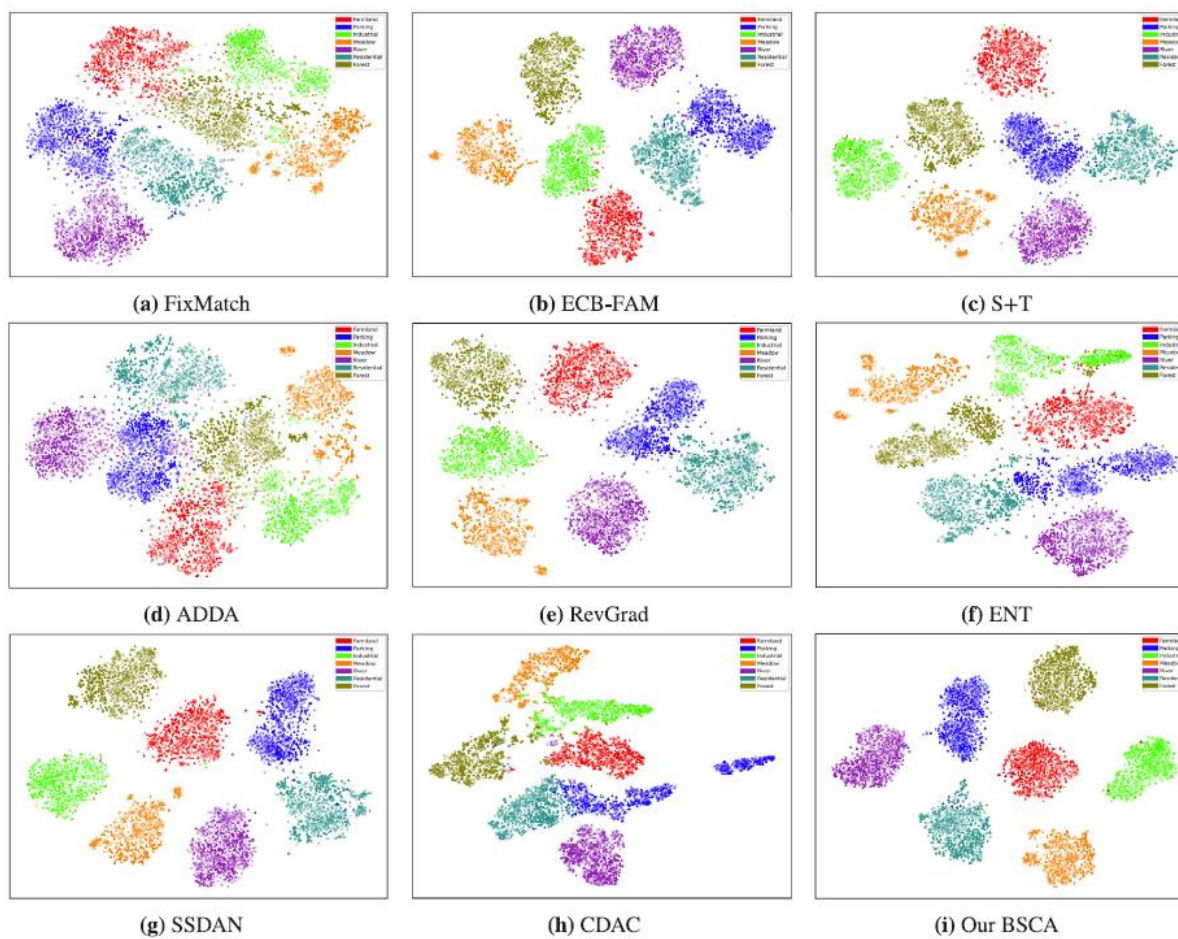
**Table 7**

Classification accuracy (%) of comparison experiments on the 12 adaptation scenarios from the RS-SSDA datasets under the 3-shot setting with **EfficientNet\_B0** as the CNN backbone. The best results are in **bold**.

Type	Method	A→N	R→N	W→N	N→A	R→A	W→A	N→R	A→R	W→R	N→W	A→W	R→W	Mean
SSL	FixMatch (Xiong et al., 2021)		72.2			93.2			63.0			96.9		81.3
UDA	ECB-FAM (Ma et al., 2021b)	91.7	76.8	83.0	91.8	80.5	90.8	68.6	69.5	62.4	92.4	98.6	84.7	82.6
SSDA	S+T	91.1	81.5	86.0	92.6	84.2	91.4	71.4	71.9	66.2	93.1	97.6	90.4	84.8
	ADDA (Wang et al., 2018)	90.2	77.7	84.3	89.1	79.8	90.8	69.8	69.9	69.3	94.1	96.6	83.3	82.9
	RevGrad (Lasloum et al., 2021)	90.9	78.4	84.5	92.7	84.4	91.5	72.8	69.6	72.0	94.1	<b>99.0</b>	89.2	84.9
	SSDAN (Lasloum et al., 2021)	90.6	79.6	83.7	90.8	86.0	89.6	73.1	73.8	72.0	96.5	<b>99.0</b>	87.2	85.2
	ENT (Grandvalet et al., 2005)	88.3	78.2	84.8	91.5	83.2	90.5	70.7	68.2	71.9	96.5	97.9	91.0	84.4
	CDAC (Li et al., 2021)	87.0	84.7	86.0	88.5	78.6	84.8	68.5	66.3	68.5	88.9	92.7	81.2	81.3
	<b>Our BSCA</b>	<b>96.1</b>	<b>88.3</b>	<b>93.8</b>	<b>96.5</b>	<b>95.8</b>	<b>96.4</b>	<b>77.0</b>	<b>78.9</b>	<b>83.0</b>	<b>98.3</b>	98.6	<b>98.6</b>	<b>91.8</b>

# 評価

- 最先端の手法と比較



# 評価

- バックボーン

**Table 8**

Comparison of BSCA in parameters, computing complexity, and 12-adaptation-scenario mean accuracy based on different CNN models.

Model	Params (M) ↓	MACs (G) ↓	Mean accuracy (%) ↑
VGG16	14.98	15.35	93.0
ResNet34	21.55	3.68	<b>93.3</b>
EfficientNet_B0	<b>4.67</b>	<b>0.41</b>	91.8

# 結論

- RSにおいて最新のSSL/UDA/SSDA手法と比較して、BSCAはクロストメインのクラス内特徴表現により、ラベルのないターゲットドメインで優れた分類精度を実現