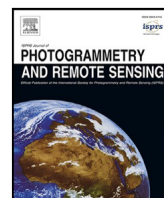




ScienceDirectで利用可能なコンテンツリスト

ISPRS写真測量リモートセンシングジャーナル

ジャーナル・ホームページ: www.elsevier.com/locate/isprsjprs



リモートセンシングのための半教師付き双方向アライメント シーン分類

ウェイ・ファン^a 施益蕾^b 王齊^a 王齊^c 朱曉祥^{a,*}

^a ミュンヘン工科大学地球観測データ科学講座、ミュンヘン、80333、ドイツ

^b ミュンヘン工科大学リモートセンシング技術講座、ミュンヘン、80333、ドイツ

^c 西北理工大学人工知能・光学・電子学院 (iOPEN)、西安市、710072

ア・ラ・ティ・シー・イ・イン・フ・ア・ビー・システムズ・トラック・ア・シー・ティー

キーワード

リモートセンシング

半教師付きドメイン適応 クロス

ドメイン分類 双方向サンプルク

ラスアライメント

リモートセンシング (RS) 画像のシーン分類は、その幅広い応用の可能性から、ますます注目を集めている。従来の完全教師ありアプローチでは、人手でラベル付けされた大量のデータが必要であった。より多くのリモートセンシング画像が利用可能になるにつれ、これらのラベル付けされていないデータをいかに活用するかが喫緊の課題となっている。半教師あり学習は、少数のラベル付きデータを用いて多数のラベルなしデータの自己学習を導くものであり、直感的な戦略である。しかし、異なるデータセット間のドメインシフトが大きいため、データセット横断的（ドメイン横断的）なシーン分類に適用することは困難である。このため、完全にラベル付けされたRSシーンデータセット（ソースドメイン）から、限定ラベル付けされたRSシーンデータセット（ターゲットドメイン）へ、ドメインシフトを低減し、さらに知識を転送することができる半教師付きドメイン適応（SSDA）が、実現可能な解決策となる。本論文では、RSのクロスドメインシーン分類のための双方向サンプルクラスアライメント（BSCA）と呼ばれるSSDA手法を提案する。BSCAは教師なしアライメント (unsupervised alignment: UA) と教師ありアライメント (supervised alignment: SA) の2つのアライメント戦略から構成される。UAは、クラスラベルを必要とせず、ドメイン間の最大平均不一致の距離を縮めることに集中する。対照的に、SAは、ソースサンプルから関連ターゲットクラス中心へ、ターゲットサンプルから関連ソースクラス中心へ、それぞれのクラスを意識した分布整合を達成することを目的とする。提案手法の有効性を検証するため、広く利用されている4つのRSシーン分類データセットを用いて構築したRS-SSDAベンチマークに対して、広範なアブレーション、比較、可視化実験を行った。実験の結果、いくつかの最新手法と比較して、我々のBSCAは、コンパクトな特徴表現と低エントロピーな分類境界により、優れたクロスドメイン分類性能を達成することが示された。我々のコードは<https://github.com/hw2hwei/BSCA>。

1. はじめに

リモートセンシング (RS) 画像シーン分類 (Wang et al., 2020; Cheng et al., 2017; Bai et al., 2022; Ma et al., 2021a; Xiong et al., 2022) は、土地利用 (Ma et al., 2017) や都市計画 (

Rottensteiner et al., 2012; Qiu et al., 2019) などの幅広い実世界への応用のため、注目が高まっている。しかし、RS画像シーン分類のほとんどの最先端手法は、完全教師ありディープラーニングモデルに基づいており、多数の手動ラベル付けデータに大きく

依存している。対照的に、より多くのラベル付けされていないRS画像が蓄積され、研究者がアクセスできるようになっている。このような状況において、ラベル付けされていないRS画像をどのように有効活用するかが喫緊の課題となっている。

RS画像のシーン分類に関するいくつかの半教師付き学習（SSL）作品（[Han et al, 2018](#); [Miao et al, 2022](#)）が試みられている。

には、ラベル付けされていないデータで自己学習することで、アノテーションへの依存度を下げることができる。しかし、既存のデータセットが持つ大規模なラベル付きデータを直接利用することはできない。具体的には、RSシーン分類のデータセット間には、異なる時間、様々なスケール、異なる撮影機材、固定されていない撮影アングルなど、多くのドメインシフトをもたらす要因が存在する。実際には、これらの要因が混在することで、ドメインシフトはさらに複雑になる。一方、RSのクロスドメインシーン分類のために、いくつかの非スーパービジョン領域適応（UDA）手法が検討されている（[Ma et al., 2021b](#); [Zheng et al., 2022b](#); [Yu et al., 2022](#); [Ahmed et al.](#)

* 筆者。

メールアドレス: w2wei.huang@tum.de (W. Huang), yilei.shi@tum.de (Y. Shi), zhitong.xiong@tum.de (Z. Xiong), crabwq@nwpu.edu.cn (Q. Wang), xiaoxiang.zhu@tum.de (X.X. Zhu)。

<https://doi.org/10.1016/j.isprsjprs.2022.11.013>

2022年7月13日受領; 2022年10月18日改訂版受領; 2022年11月18日受理

2022年12月5日オンライン公開

0924-2716/© 2022 The Author(s).国際写真測量リモートセンシング学会（ISPRS）を代表してElsevier B.V.が発行。これはCC BYライセンス

(<http://creativecommons.org/licenses/by/4.0/>)に基づくオープンアクセス記事です。

を共有するクラス空間内で使用することができる。しかし、UDAの一般化は、ラベル付けされたターゲットデータの不足により制限される。ターゲットデータは、ソースドメインとターゲットドメインの間の橋渡しとして、ドメインシフトを低減するために不可欠である。この観点から、SSLとUDAの組み合わせとして、半教師付きドメイン適応 (SSDA) は、少数のラベル付きターゲットデータのガイダンスの下で、アノテーションが豊富なソースドメインから僅かにラベル付けされたターゲットドメインに知識を転送することができるため、実現可能なソリューションとなる。しかし、UDAと同様、SSDAもドメインシフトに悩まされる。

特徴量のアライメントは、経験的に主要なズレを減少させる効果的な方法であり、その結果、対象ドメインの分類性能を高めることができる。特徴量の観点から、現在のSSDAの特徴量アライメントは、サンプルアライメント(Motiiian et al., 2017)とドメインアライメント(Kim and Kim, 2020)の2種類に大別できる。サンプルアライメントはドメイン間のクラス内インスタンスのアライメントを目的とし、ドメインアライメントは2つのドメインの大域的な特徴分布をアライメントしようとするものである。その中間戦略として、クラスアライメントは、ノイズの多いサンプルに対する感度が低いため、サンプルアライメントよりも頑健であり、焦点化操作のため、ドメインアライメントよりもきめ細かい。正確で安定したクラスアライメントを行うためには、学習段階で複数のサンプルからクラス中心を計算する必要がある。残念ながら、各ミニバッチにおいて同一クラス内のサンプル数は非常に限られており、クラスごとの中心を正確かつ頑健に計算することや、さらなるクラス整列を行う妨げとなっている。

高い計算負荷なしに柔軟で頑健なクラスレベル特徴アライメントを実現するために、我々はRSクロスドメイン情景分類のための双方向サンプルクラスアライメント (BSCA) と呼ばれる新しい全体的SSDAフレームワークを考案した:

(1) ラベル付けされたサンプル (ソースサンプルとラベル付けされたターゲットサンプルを含む) から抽出された特徴量と、ラベル付けされていないターゲットサンプルから抽出された特徴量の間の大域的な分布整合を実現するために、教師なし整列 (UA) モジュールが導入されている。これは、ソースドメインと未ラベルターゲットドメイン間だけでなく、ラベルターゲットドメインと未ラベルターゲットドメイン間の最大平均不一致 (MMD) (Gretton et al;

中心との特徴量アライメントを実現するために、独自の教師ありアライメント(SA)モジュールを提案する。ここで、クラス中心は、ソースサンプルとターゲットサンプルから抽出された特徴量を個別に格納できるメモリバンク機構によって計算される。SAでは、ソースメモリバンクとターゲットメモリバンクという2つの分離されたメモリバンクが存在し、2種類のクラスセンターに対応する。(a)同じクラスのソースメモリバンクを更新するために使用され、(b)ターゲットメモリバンクから計算される同じクラスのターゲットクラス中心に整列される。同様に、ターゲットサンプルから抽出された特徴量は、ターゲットメモリバンクの更新に使用され、対応するソースクラス中心に整列される。従って、我々のSAは、ソースサンプルからターゲット中心への特徴アライメントと、ターゲットサンプルからソース中心への特徴アライメントの両方、すなわち双方向アライメントを実現することができる。また、メモリバンクを利用することで、過去のサンプルの豊富なクラス対応特徴量を保存し、少ない計算資源でリアルタイムにサンプルクラス特徴量のアライメントを行うことができる。言い換えれば、BSCAのSAは、ミニバッチの制限を超えて、豊富なクラス認識特徴量からロバストなクラス中心を計算し、オンライン方式でクラス単位の特徴量アライメントをサポートすることができる。

提案するBSCAの有効性を客観的に評価するために、広く使われている4つのRS画像シーン分類データセットの共通クラスを用いてRS-SSDAベンチマークを構築し、このベンチマーク上で我々の手法をいくつかの最先端のSSL、UDA、SSDA手法と比較する。実験結果は、我々のBSCAが以下の性能を上回ることを示している。

(2) 本論文では、サンプルと対応するクロスドメインクラス

また、ベンチマークにおける12の適応シナリオの中で最高の平均分類性能を達成した。全体として、我々の貢献は以下のように要約できる：

- ・本研究では、教師ありアライメント(SA)と教師なしアライメント(UA)の2つのモジュールから構成される、RS-SSDAのための双方向サンプルクラスアライメント(BSCA)手法を提案する。教師ありアライメント(SA)は元サンプルから元クラス中心へ、ラベル付け/擬似ラベル付けされた元サンプルから元クラス中心への双方向の特徴量アライメントを実現する。対照的に、UAは、ラベル付けされたサンプルとラベル付けされていないサンプルから抽出された特徴量間の大域的なアライメントに焦点を当て、それらのクラス情報を必要としない。
- ・提案手法の有効性を検証するため、広く利用されている4つのRS画像シーン分類データセットから7つの共通クラスを含むRS-SSDAベンチマークデータセットを収集し、それに基づく広範な実験を実施する。
- ・いくつかの最新のSSL/UDA/SSDA手法と比較して、我々の提案するBSCAは、上記のRS-SSDAベンチマークにおいて最良の平均分類結果を得ており、我々の手法の有効性を実証している。

2. 関連作品

2.1. 半教師付きドメイン適応

SSDAは、SSLとUDAの組み合わせと見ることができる、最近台頭してきた研究分野である。我々の知る限り、SSDA手法には、敵対的訓練に基づくSSDA (Tzengら、2017; Jiangら、2020)、エントロピー最適化に基づくSSDA (Grandvaletら、2005; Saitoら、2019; Li and Hospedales、2020)、特徴アライメントに基づくSSDA (Motiianら、2017; Kim and Kim、2020; Liら、2021)の3種類がある。具体的には、Tzengら (2017)は、SSDA問題とディープモデルの汎化に対処するためのシャムアーキテクチャを提案した。このアーキテクチャは、ラベル付けされたソースサンプルとターゲットサンプルの特徴を扱うために、分類と対照的意味整列 (CCSA) 損失を導入することによって、識別埋め込み部分空間を学習するが、ラベル付けされていないターゲットサンプルの使用法は無視する。Saitoら (2019)は、SSDAモデルを敵対的に最適化する新しい最小エントロピー(MME)アプローチによって、この分野に注目した。MMEは、ラベル付けされていないターゲットデータの条件付きエントロピーを、特徴分類器に対して交互に最大化し、特徴

エンコーダに対して最小化することができる。Jiangら (2020)は、ドメインギャップを越えて敵対的な例を導くための一般的な双方向敵対的訓練(BiAT)を考案した。LiとHospedales(2020)は、計算が容易で、マルチソース教師なしドメイン適応(MSDA)とSSDAの両方に有効なオンライン最短パスメタ学習フレームワークを提案した。KimとKim(2020)は、ドメイン内の不一致を低減することで特徴量を揃えることを目的とし、主にattraction、perturbation、explorationの3つのスキームからなるSSDAフレームワークを提示した。Liら(2021)は、領域間適応と領域内適応の両方を実現するために、領域間適応クラスタリング(CDAC: cross-domain adaptive clustering)と呼ばれるアプローチを考案している。

2.2. RSシーン分類のドメイン適応

近年、RS画像シーン分類のドメイン適応 (DA) に注目する研究が増えている (Song et al., 2019; Lu et al., 2020; Zheng et al., 2021; Zhang et al., 2020; Zhu et al., 2021; Zheng et al., 2022a; Lasloun et al.) Songら(2019)では、CNNモデルに追加された新しい部分空間整列層がRS画像シーン分類のDAのために提案され、特徴部分空間におけるソースドメインとターゲットドメインを整列させる。

を提案した。Zhangら(2020)では、RS画像シーン分類のための相関部分空間動的分布整列法が提案され、これは、ソースドメインデータを無関係な部分空間にマッピングすることを回避しようとする部分空間相関最大化(SCM)と、クロスドメイン分布の不一致を低減することを目的とする動的統計分布整列(DSDA)から構成される。注意に基づくマルチスケール残差適応ネットワーク (AMRAN) がクロスシーン分類タスクのために提案された (Zhu et al., 2021)。AMRANでは、マージナル分布と条件付き分布の両方が考慮され、ロバストな特徴と完全な情報を抽出するためにマルチスケール注意メカニズムが用いられた。

Zhengら(2022a)では、RS応用のための単一ソース複数ターゲット領域適応タスクが探求され、2段階適応ネットワーク(TSAN)と名付けられた新しいアルゴリズムが提示された: (1)敵対的学習アプローチを利用して、ソース・ドメインと混合マルチターゲットのタスクの両方において、クロスドメイン分布の不一致を低減する。複数のドメインの残差を動的に調整する学習。

ドメインラベルは、複数のターゲットドメインの固有特徴を学習するために使用される。Laslounら (2021) では、セクション 2.1 で述べた MME アルゴリズム (Saito et al., 2019) を用いて、我々の研究は、主に RS 画像におけるシーン分類に集中している。

単一ソース単一ターゲットの SSDA 設定において、単一ソースドメインから単一ターゲットドメインへの伝達可能な知識を学習し、マルチソース/マルチターゲット研究にも役立つドメイン不変の特徴表現を得ることを意図している。

2.3. メモリーバンク

メモリーバンクモジュールは、ニューラルネットワークの外部に余分な情報を保存することができるため、広く応用されている。

辞書を参照することができる。例えば、最も明白な情報を外部記憶モジュールに書き込むことを学習できる、記憶増強型時間的双方向学習ネットワークが、次のような目的で提案された。

Yuan et al. (2019) では、人間の行動認識を行った。Lvら(2021)では、リアルタイムフレーム構築のためのプロトタイプとして、通常のダイナミクスを符号化する双方向プロトタイプユニットが提示された。対照学習の観点から、教師なし視覚表現学習のために、MoCo(momen-tum contrast)の移動平均キューエンコーダを持つメモリーバンクに基づいて双方向辞書が構築された(He et al., 2020)。Alonsoら(2021)では、半教師付きセマンティックセグメンテーションのために、セグメンテーションネットワークを強化するために、メモリーバンク機構を対比学習と組み合わせ、

3.2. 共有ネットワーク・アーキテクチャ

いくつかの既存の SSDA 文献 (Chen et al., 2019; Saito et al., 2019) に従い、本稿では RS クロスドメインシーン分類に CNN モデルを用いる。これは 3 つのコンポーネントから構成される: CNN ベースの特徴抽出器 \mathcal{E} 、グローバル平均プーリング (GAP) 層 \square 、特徴分類器 \mathcal{C} 。これらはすべてソースドメインとターゲットドメインの間で共有される。

まず、特徴マップ $\mathbf{m} \in \mathbb{R}^{H \times W \times C}$ 、ここで H 、 W 、 C は空間的なマップである。高さ、空間幅、特徴次元はそれぞれ、特徴抽出器 \mathcal{E} によって画像 $\hookrightarrow \mathbf{L1_1D431}$ から抽出される。

$$\mathbf{m} = \mathcal{E}(\hookrightarrow \mathbf{L1_1D431}). \quad (1)$$

そして、グローバル特徴ベクトル $\mathbf{f} \in \mathbb{R}^C$ 、GAP 層 \square (Lin et al., 2013) を介して \mathbf{m} から次のように計算される。

$$\mathbf{f} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \mathbf{m}_{ij}. \quad (2)$$

最後に、 $\hookrightarrow \mathbf{L1_1D41F}$ は特徴分類器 \mathcal{C} に入力され、クラスごとの確率的予測ベクトル $\mathbf{p} \in \mathbb{R}^K$ が得られる。

$$\mathbf{p} = \frac{\mathbf{C}(\mathbf{f})}{T}. \quad (3)$$

ここで最初の式は l_2 の正規化を指す (Ranjan et al.) 項 $\hookrightarrow \mathbf{L1_1D70E}$ はソフトマックス関数であり、予測ベクトルのすべての要素の合計を 1 に正規化できる; これは $\hookrightarrow \mathbf{L1_1D429}$ の要素値が対応するクラスの確率と見なせることを意味する。 $\hookrightarrow \mathbf{L1_1D44}$ は温度のハイパーパラメータである。

SSDA 文献 (Saito et al., 2019; Li et al., 2021) における一般的な手法に従い、ソースドメイン $Q \hookrightarrow \mathbf{L160}$ とラベル付きターゲットドメイン $Q \hookrightarrow \mathbf{L1459}$ の両方からの全てのラベル付きサンプルが、以下のような分類損失 (標準的なクロスエントロピー損失) を介してネットワークの学習に使用される。

$$\mathcal{G}_{CS} = \mathbb{E}_{\{\mathbf{x} \hookrightarrow \mathbf{L1_1D466}\} \in Q_S \cup Q_T} \mathcal{L}(\mathbf{p}). \quad (4)$$

領域内クラス横断サンプルに類似したピクセルレベルの特徴表現を得たが、ラベル付けされたデータからの特徴ベクトルのみで更新し、ラベル付けされていないデータの利用は無視した。

3.3 双方向サンプルクラスアライメント

BSCA は、UA と SA から構成され、図1 に示すように、教師なし・教師ありの特徴アライメントを実現することを目的としている。

3.3.1. 教師なしアライメント

図1 の "unsupervised alignment" の部分に示されているように、UA

3. 方法論

本節では、まず普遍的なSSDAのいくつかの表記法を紹介し、次に共有ネットワークアーキテクチャを示し、最後に提案するBSCAをその学習手順の概要とともに詳述する。BSCAに基づくSSDAアーキテクチャのワークフローを図1に示す。

3.1. 表記

SSDAの設定では、3つのサブドメインがある：完全にラベル付けされたソース・ドメインS、限定ラベル付けされたターゲット・ドメインf、そしてラベル付けされていない対象領域f₁。そのサンプル集合を $Q_s = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N_s}$ とする。ここで、 \mathbf{x}_i, y_i, N はそれぞれ

それぞれ、画像、そのラベル、その領域のサンプル数である。ここで、 $\phi_{L_{1D441}}$ は $\phi_{L_{1D59}}$ よりもはるかに小さい。ソースラベルとターゲットラベル、 y^s と y^t は同じクラス空間 $\{1, \dots, K\}$ を持つ。クラスの。

は主にラベル付けされたサンプルから抽出された特徴とラベル付けされていないサンプルから抽出された特徴の間の教師なしドメイン間アライメントに集中している。再現カーネル・ヒルベルト空間 (RKBS) (Borgwardt et al., 2006) におけるMMD距離 (Gretton et al., 2012a; Pan et al., 2010) を用いて、これら2つの分布の平均値の類似性を次のように評価する。

$$Gal \, g_{MD}^2 = \frac{1}{N+1} \sum_{i=1}^{N+1} \phi(\mathbf{f}_{L_{1D460}}) \phi(\mathbf{f}_{L_{1D462}})^T, \quad (5)$$

ここで、 $\mathbf{f}_{L_{1D460}}$ は式(1)~(3)で抽出された大域的特徴ベクトルを表す。

(2) ソースドメインとラベル付けされたターゲットドメインの画像から、そして $\mathbf{u}_{L_{1D462}}$ はラベル付けされていないターゲット領域の画像から抽出されたグローバル特徴ベクトルを表す。関数 $\phi_{L_{1D719}}$ は、 \mathbf{u} と \mathbf{v} を RKBS に投影できるガウス放射基底関数 (RBF) カーネルである。特徴表現の能力を向上させるために、MMDの多重カーネル変形、すなわちMK-MMD (Gretton et al.

ここで特筆すべきは、ソースデータとターゲットデータ間のアライメントだけでなく、UAは以下のアライメントに焦点を当てていることである。

を含む) すべてのラベル付きドメイン間の特徴分布。ソース・ドメインとラベル付けされたターゲット・ドメイン)とラベル付けされていないターゲット・ドメイン(ソース・ドメインとラベル付けされた各ドメインがあるターゲットデータは

教師あり訓練によって識別可能な特徴表現を提供し、ラベル付けされていないデータから抽出された特徴のよりロバストなアンカーとして、ドメインレベルでより包括的な特徴分布を提供するために、完全にラベル付けされたソースデータと連携することができる。

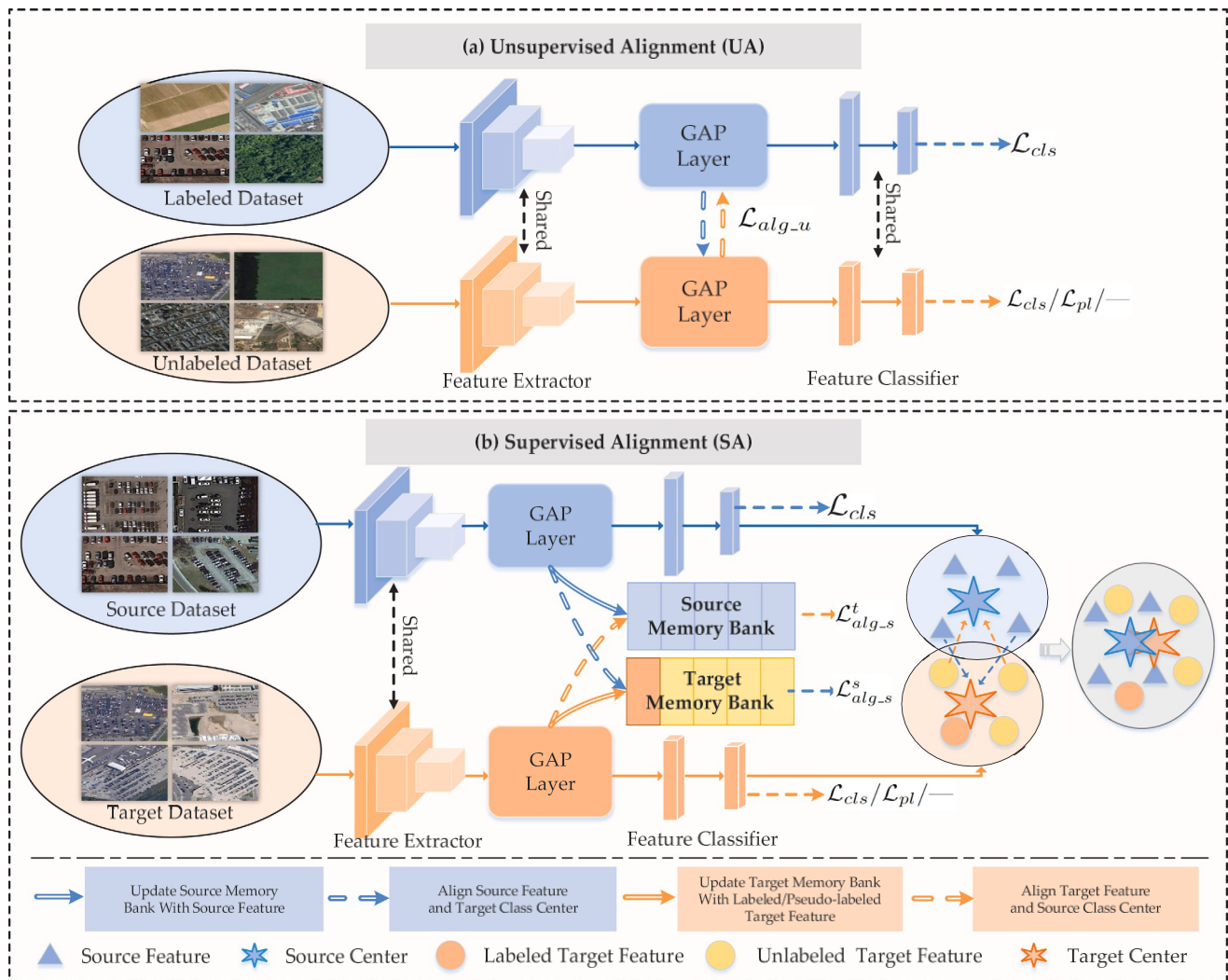


図1. 提案するRS-SSDA用BSCAのワークフローは以下の2つの部分から構成される。

(b) クロスドメインのクラス認識特徴量をクラスラベル/擬似ラベルの要求と整合させるSA。

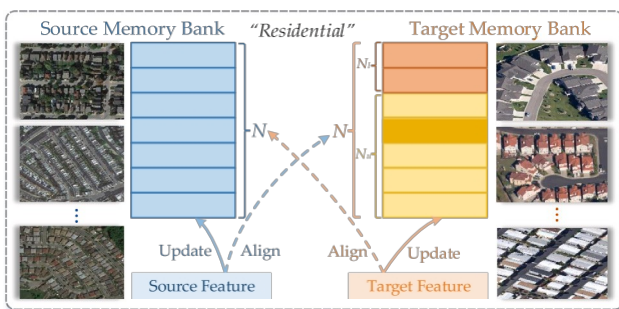


図2. RSのクロスドメインシーン分類のためのSAの説明図。(この図の凡例における色の言及と解釈については、読者はこの記事のウェブ版を参照されたい)

3.3.2. 監督付きアライメント

図1(b)に示すように、SAは、メモリバンク機構を介して、サンプルからクロスドメインクラス中心への特徴量アライメントを実現することを目的とする。SAでは、ソースメモリバンクとターゲットメモリバンクが存在し、それぞれ n^s 、 $n^t \in k^{K \times N \times C}$

、ソースサンプルとターゲットサンプルから抽出された特徴を格納する。ここで $\hookrightarrow Lu_1D43E$ はデータセットのクラス番号、 $\hookrightarrow Lu_1D436$ は特徴量のチャンネル次元、そして

Lu_1D441 は、 n^s および n^t の各クラスの項目番号である。ターゲット・メモリ・バンク n^t は、ラベル付きターゲット・メモリ・バンク $n^t \in k^{K \times N_u \times w}$ と、ラベルなしターゲット・メモリ・バンク $n^u \in k^{K \times N_u \times w}$ である。これらのメモリバンクはゼロで初期化される。

メモリバンク・ベースのSAのパイプラインを図2に示す。図2 図中、青い四角形はソース画像から抽出された特徴量、オレンジ色の四角形はラベル付けされたターゲット画像から抽出された特徴量、黄色の四角形は信頼度の高いラベル付けされていないターゲット画像から抽出された特徴量を表す。メモリバンクに基づくSAは、サンプルアライメントとクラスアライメントの両方の利点を兼ね備えており、サンプルは柔軟な最適化を可能にし、クラスセンターは安定でロバストなアンカーを提供する。具体的には、メモリバンクに格納された複数の特徴量からロバストなクラス中心を計算することができ、計算量の多い特徴抽出処理を必要としない。一方、ミニバッチのサンプルから抽出された特徴量は、関連するクロスドメインのクラス中心にリアルタイムでアライメントできる。一方、ミニバッチのサンプルから抽出された特徴量は、リアルタイムで関連するクロスドメインクラス中心にアライメントすることができる。

SAの詳細なプロセスは、以下の3つのステップに分けられる：

(a) **特徴とラベルのペアの生成**

は、各反復で特徴ラベル対 $[f, y]$ となる。式(1)-(2)を用いて、画像 x, s, x^l, x^u は \mathcal{E} と \square の組み合わせに入力される。

そして、対応するグローバル特徴ベクトル \mathbf{f} 、 \mathbf{s} 、 \mathbf{f} が出力される。これら3つのベクトルのうち、 \mathbf{f}^k は直接、関連するラベル $\hookrightarrow L_{160}$ と \mathbf{f} と組み合わせられ、特徴とラベルのペア $[\mathbf{s}, \mathbf{y}_s]$ と $[\mathbf{f}^t, \mathbf{y}_f]$ を形成することができる。関連ラベルを持たない \mathbf{f} については、その最大活性クラスが擬似ラベル $\hookrightarrow L_{1D466}$ として用いられる。

$$y = \underset{y \in \{1, \dots, N\}}{\operatorname{argmax}} P(\mathbf{f} | y) \quad (6)$$

ここで、 w は式(3)によって u から計算される。項 \mathbf{f} 、 $\hookrightarrow L_{1D462}$ \mathbf{p} 、 u 、 $y \hookrightarrow Mn_{302}$ は擬似特徴ラベルペア $[u, y \hookrightarrow Mn_{302}]$ を形成する。

(b) 特徴量とラベルのペアでメモリバンクを更新する。その後

メモリバンクを更新するための合理的な更新戦略を設計することが重要である。メモリバンクはサンプルクラスのアライメントにアンカーを提供するため、SSDAの性能を決定する。 n^s 、 n^f 。

の干渉を減少させる。

n^u 、信頼度の高いラベルで更新される。

\mathbf{f} 、その擬似クラス確率が閾値 τ より高い。は

更新戦略をアルゴリズム1に示す。このうち、 $n[\hookrightarrow L_{1D466}, 1: \check{u} \boxtimes 1]$ は、クラス $\hookrightarrow L_{1D466}$ の1から $\check{u} \boxtimes 1$ までのアイテムを、対応するメモリバンク n に格納する。

BSCAの場合、メモリバンクは反復ごとに更新される：

(1) 遅れをとるクラスセンターではなく、リアルタイムのクラスリアルタイムで画像から抽出された特徴と一致する。

(2) メモリバンクのサポート可能なアイテムを生成する。

誤って擬似標識されたターゲット・サンプルは置き換えられるからだ。

を反復する。メモリーバンクのもうひとつの利点は、私たちのクラスのサンプル数に依存しないBSCA。技術的には、BSCAは少なくとも1つのラベル付きソース・サンプルと1つのラベル付きクラスごとのターゲット・サンプルは、その関連クラス中心を計算する。 $\hookrightarrow L_{1D462}$ は以下のように与えられる。

アルゴリズム1 メモリバンクの更新

入力: 特徴ラベルのペア: $\{[\mathbf{f}^s, \mathbf{y}^s]\}$, $\{[\mathbf{f}^f, \hookrightarrow L_{1D463}]\}$, $\{[\mathbf{f}^u, \mathbf{p}^u, \mathbf{y}^u]\}$, ミニバッチのサイズ T , エントロピーの閾値 θ , メモリバンク: n^s , n^f , n^u

出力: 更新されたメモリーバンク: n^s 、 n^f 、 n^u

1: while $i < M$ do
2: 更新: $n^s[y^s[1:N-1]]_s \leftarrow n^s[y^s[2:N]]$

3: 更新: $n^s[y^s[i], N] \leftarrow f_i$
4: end while
5: while $i < M$ do
6: 更新: $n^f[y^f[1:N-1]] \leftarrow n^f[y^f[2:N]]$
7: 更新: $n^f[y^f[i], N] \leftarrow f_i$
8: end while
9: while $i < M$ do

を計算する必要があります。アライメントを行う前に、ソースとターゲットのクラス中心を計算する必要があります。 \mathbf{c} , \mathbf{s} , \mathbf{f} $\in \mathbb{R}^{K \times C}$ と表記します。 $\hookrightarrow L_{1D8}$ 番目のクラス・センターは、関連するメモリ・バンク内のそのクラスの全アイテムの平均であり、次式で定式化される。

$$\mathbf{c}_k = \frac{1}{N} \sum_{i=1}^N \mathbf{n}_{ki}^s, \quad \mathbf{c}_k^t = \frac{1}{N} \left(\sum_{i=1}^t \mathbf{n}_{ki}^f + \sum_{i=1}^{N-t} \mathbf{n}_{ki}^u \right) \quad (8)$$

ここで、トレーニングの初期には、メモリバンクを更新するのに十分な特徴量がないため、いくつかの項目が初期ゼロの状態のままであることを述べておく。これらの干渉を防ぐため、この段階ではこれらのゼロの項目はクラス中心に計算されない。ソースクラス中心とターゲットクラス中心は、アラインメント（位置合わせ）

のためのアンカーとして使用される。特徴量 \mathbf{f} , \mathbf{f} のメンター。教師ありアライメント損失

G_{alg_s} は、ソース特徴アライメント損失とターゲット特徴アライメント損失 G^t の2方向から構成される。

$$G_{alg_s} = \frac{1}{M} \sum_{i=1}^M (\mathbf{f}_i^s - \mathbf{y}_i^s)^2, \quad G^t = \frac{1}{M} \left[\sum_{i=1}^M (\mathbf{f}_i^f - \mathbf{y}_i^f)^2 + \sum_{i=1}^{M_u} (\mathbf{f}_i^u - \mathbf{y}_i^u)^2 \right], \quad \text{if } y > \tau \quad (9)$$

$$||G_{alg_s}|| = G_{alg_s}^s + G_{alg_s}^t.$$

ここで、特徴とラベルのペア $[\mathbf{f}_s, \mathbf{y}_s]$, $[\mathbf{f}_f, \mathbf{y}_f]$, $[\mathbf{f}_u, \mathbf{y}_u]$

ステップ(1)を参照。Liら(2021)に従い、 \mathbf{f}^u をアライメントに用いる、は、RandAugment技法 (Cubuk et al., 2020) により同じターゲット画像の拡張部分から抽出された対応部分で置き換えられる。ミニバッチ内では、 M^u は、ラベル付けされていない特徴量の数である。

擬似ラベル \mathbf{f}^u が τ より大きく、 \mathbf{f}^u が \mathbf{c}_s に最も近い。平均二乗誤差(MSE)を測定値として使用する。BSCAでは、クラス中心は特徴のアンカーとしてのみ機能し、対応する生画像への勾配バックプロパゲーションは行わない。その結果、このような特徴クラス・アライメントは、処理中にほとんど計算資源を消費しません。

10: もし $E_{1D45D} < \tau$ なら
11: 更新: $n^u[y^u[1:y-1]] \leftarrow n^u[y^u[2:u]]$
12: 更新: $n^u[y^u[u], u] \leftarrow f^u$
13: エンド・イフ
14: end while

15: n^s 、 n^f 、 n を返す。

一方、信頼度の高いラベルなしサンプルの擬似ラベルは、以下のように、ラベルなしターゲット領域の重みを改善するためにネットワークを訓練するために使用することもできる。

$$|_{Ep} = \sum_K p_{ul},$$

$$G = \sum_{i=1}^i 1_{\{E < \tau\}} \cdot y_u^{log p u}, \quad (7)$$

ここで、 $1_{\{E < \tau\}}$ は、 E_p のエン트로ピー和の値がエン트로ピー閾値 τ より小さいときに値が1となり、そうでないときに値が0となる指標関数である。

(c) クロスドメインサンプルクラスアライメントの作成。最後に、特徴量とクラス中心のクロスドメインアライメントを行います。

トレーニングの段階。

3.3.3. BSCAの全体的な損失とトレーニングの手順

BSCAベースのSSDAモデルの総合損失は、教師あり分類損失 G_{cls} 、擬似ラベル分類損失 G_N 、教師なし特徴アライメント損失 G_{\square} 、および教師ありクロスドメイン特徴クラスアライメント損失 G_{\square} の組み合わせであり、次式で定式化される。

$$G = G_{cls} + \alpha G_{alg_l} + \beta G_{alg_l} + G_{\square}, \quad (10)$$

ここで α と β は G_a と G_{g_s} の重みである。

本手法のワークフローを明確に理解するために、イテレーションを行う。

BSCAの学習手順を以下に示す。各反復において、3つのミニバッチ画像 (x_s) がそれぞれ Q_1 からランダムにサンプリングされ、共有特徴抽出器と特徴検出器に送られる。したがって、大域的特徴 (f_s, l, f, u) とクラスごとの予測 (p, l, s, p, l, l) の3つのミニバッチが、共有特徴抽出器と特徴分類器に送られる。 p が得られる。このうち p と l は計算に使用される。

クロスエン트로ピー損失 G_{cls} の高信頼性部分を使用。擬似ラベル分類損失 G_N の計算のため f^u と p^l と f^l と p^s を使用される。

アライメント損失 G_{\square} は、 f^s と f^l と p^s と p^l の連結は、以下のように使用される。

を計算し、アライメント損失 G_{a_s} を求める。次に、上記の損失から式(10)により総合損失を求め、これを

勾配バックプロパゲーションとモデル最適化。最後に、 u はソースメモリバンクの更新に、 s はラベル付けされたターゲットメモリバンクの更新に、 f^u の高信頼部分はラベル付けされていないターゲットメモリバンクの更新に使用される。



図3. 3.4つのRSシーンデータセットに共通する7つのクラスの例。様々なスケールや角度など、異なるデータセット間でいくつかのドメインシフトがあることが観察された。

4. 実験

本節では、ベンチマークデータセットや実装の詳細など、いくつかの実験設定を紹介する。次に、提案するBSCAの境界を探るために、2種類のハイパーパラメータチューニング実験を行う。続いて、以下の実験を行う。

提案された方法について、いくつかの最新技術と比較した。SSL/UDA/SSDA法を用いた。次に、アブレーション試験が行われる。

BSCAの各要素の個別効果を探るためである。最後にBSCAと他の手法の特徴分布の可視化は直感的な比較のために用意されている。

4.1. 実験設定

ベンチマークデータセット提案手法を包括的に評価するために、広く使われている4つのRS画像シーン分類データセットから7つの共有クラスを持つRS-SSDAベンチマークを収集・構築する：RSSCN7 (Zou et al., 2015)、NWPU-RESISC45、AID、WHU-RS19。

各データセットは、ソース・ドメインとターゲット・ドメインの両方の役割を果たすことができるため、全部で12の適応シナリオがある。各データセットについて、すべての画像は、クラス・ラベルを持つ。

表1

RS画像シーン分類セットの特徴。				
	NWPU-RESISC45	エイド	RSSCN7	WHU-RS19
年	2017	2016	2015	2010
解像度(m)	0.2-30	0.5-8	-	0.5
画像サイズ	256 × 256	600 × 600	400 × 400	600 × 600
クラス1: 農地	700	370	400	50
クラス2: 森林	700	250	400	53
クラス3: 工業用	700	390	400	53
クラス4: 草地	700	280	400	61
クラス5: 駐車場	700	390	400	50
クラス6: レジデンス	700	410	400	54
クラス7: 川	700	410	400	56

をソース・ドメインとし、各クラスのランダムな N 画像のみをラベル付き (ラベル付きターゲット・ドメイン) とし、残りの画像はすべて unlabeled (ラベルなしターゲット・ドメイン)。ラベル付けされていない対象データはすべて教師なしモデル学習に使用される。このうち、1クラスあたり6サンプルが検証セットとして使用され、残りのデータは全てテストセットとして使用される。これらのデータのリリース年、キャプチャ解像度、元画像サイズ、クラス名、サンプル数を表1に、いくつかの例を図3に示す。以下では、NWPU-RESISC45、AID、RSSCN7、WHU-RS19 をそれ

He et al.、A、R、W と略す。

実装の詳細。VGG16 (Simonyan と Zisserman、2014)、ResNet34 (Heら、2016)、EfficientNet_B0 (TanとLe、2019) の3つの広く使われている先進的なCNNバックボーンに基づく実験を採用し、これらはすべてImageNet (Dengら、2009) で事前訓練されている。これらの特徴マップは、対応するグローバル

(2)のように GAP 層で特徴ベクトルを生成し、最後の完全連結層を(3)の特徴分類器 C で置き換える。ソースとターゲットのメモリバンクのアイテム数は32であり、ラベル付きターゲットメモリバンクのそれは $2 \times NI$ であり、したがってラベルなしターゲットメモリバンクのそれは残りの $(32 - 2 \times NI)$ である。式(7)と式(9)において、エントロピー閾値 α は、アブレーション、比較、可視化実験におけるターゲット領域として、4つのデータセットすべてについて0.5に設定される。式(10)において、教師なしアライメント損失重み α と教師ありアライメント損失 β は、それぞれ10と0.025に設定される。上記のBSCA特有のハイパーパラメータに加え、公平な比較のため、一般的な学習設定は全て本稿で用いた他のSSDA手法と同じにしてある。具体的には、モメンタムを0.9とした確率的勾配降下法 (SGD) を最適化器として使用し、モデルの学習を行う。学習率は0.01で初期化され、0.005の重み減衰で減少する。すべての手法のモデルは、12の適応シナリオすべてにおいて2000反復学習され、50反復ごとに検証セットによって検証される。学習段階において、検証結果が最も良いものは保存され、学習後にテストセットでテストされる。

表2

ResNet34をCNNバックボーンに用いた3ショット設定でのアブレーション研究の分類精度 (%)。

Gcs	Gpl	Galg _s	Galg _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	平均
✓				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	89.8	93.2	96.7	96.0	96.7	80.5	79.9	82.3	97.9	99.0	96.9	91.7 (+6.0)
✓	✓		✓	96.6	89.0	94.5	96.8	96.8	97.6	81.6	81.9	84.1	97.9	99.0	96.2	92.7 (+7.0)
✓	✓	✓	✓	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 (+7.6)

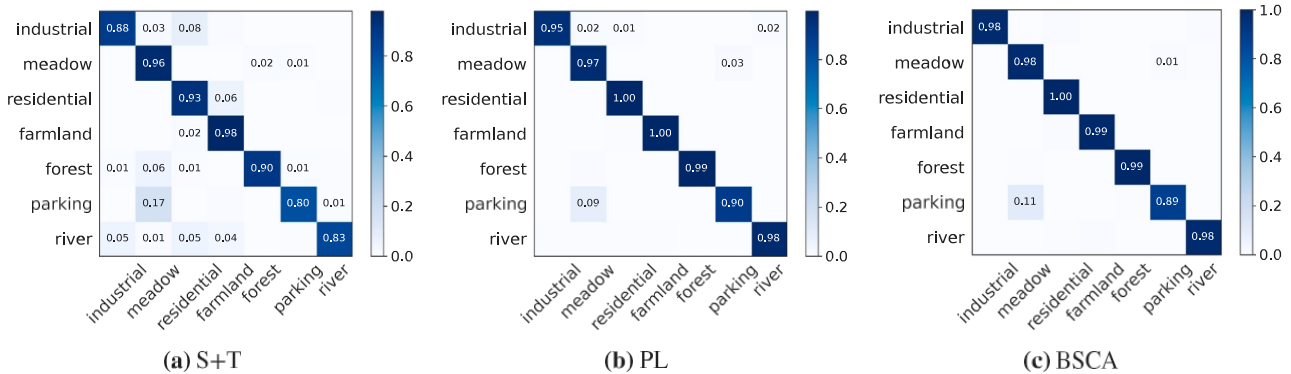


図4. ResNet34をCNNバックボーンとして、N→A適応シナリオにおけるS+T、PL、および提案BSCAの手法の混同行列。

実験は、PyTorch 1.9.1¹(Paszke et al., 2017)に基づいて、2つの GeForce RTX 2080Ti GPU上で実施された。

4.2. アブレーション研究

提案するBSCAの目的関数は、ラベル付きデータに対する監視分類損失 $G_{\square\square}$ 、ラベルなしデータに対する擬似ラベル分類損失 $G_{N\square}$ 、ラベルなしデータに対する監視分類損失 $G_{\square\square}$ の4つから構成される。教師なしクロスドメイン特徴量間の教師ありアライメント損失 $G_{\square\square\square}$ 、クラス認識/擬似クラス認識特徴量間の教師ありサンプルクラスアライメント損失 $G_{\square\square\square}$ 。知識伝達に対するそれぞれの効果を検証するため、本小節ではCNNバックボーンとしてResNet34を用い、12の適応シナリオすべてについてBSCAの検証実験を行う。実験結果を表2に示す。

Gcsのみを用いた場合、12の適応シナリオすべてにおいて平均85.7%の分類精度が得られた。埋め込み後

Gpaは、ラベル付けされていないターゲットデータのインボリューションにより、分類境界のエントロピーを減少させ、ソースドメインの過剰な影響を減少させることができるため、平均精度は91.7%に劇的に増加する。さらに、GaとGlgとGguの両方が、Gg+Glに基づく更なる性能向上に異なる程度寄与する。Gc+Glの91.0%に対してGalを加えるだけで0.7%の利得を得ることができるが、これはおそらく誤って擬似ラベル付けされたターゲットデータによって制限される。幸いなことに、GalgとGasを組み合わせると、Galsの個々の改善率はさらに0.6%上昇する。3%となり、Gcのベースラインに対して7.6%、Gc+Gsに対して2.3%の優位性がある。アブレーションの結果は、提案するBSCAの各コンポーネントの有効性を示している。

擬似ラベル学習は、分類性能を向上させる一方で、Q_1D462のラベルにノイズを与えることになる。しかし、切除結果

特徴アライメントの段階で、ターゲット領域のクラス中心は、BSCAにおけるこれらのノイズ擬似ラベルの悪影響を受けないことが明らかになった。幸いなことに、この問題を緩和する5つの要因がある：

- (1) ソース・ドメインとターゲット・ドメインは同じクラス空間を共有しているため、ソース・ドメインで学習したモデルをターゲット・ドメインでもある程度使用できる。

¹ <https://pytorch.org/>.

(3)ラベル付けされたターゲットサンプルから抽出された特徴量は、ターゲットメモリバンク内の一定数の項目を占め、特に初期段階では純粋なサンプルを提供する。(5)BSCAのUAモジュールは、ソースドメインとラベル付けされていないターゲットドメインの間のクラス非関連ドメインレベルのアライメントを達成し、後者の分布をより安定させ、ノイズ擬似ラベルの影響を少なくすることができる。

$S+T(G_c)$ 、 $PL(G_c + G_{lc})$ 、および完全なBSCAの混同行列を図4に示す。より多くの成分

を追加した場合、特に"residential"のクラスにおいて、より明確な混同行列が得られ、このことは我々のアルゴリズム設計の有効性をさらに検証するものである。

4.3. ハイパーパラメーター調整実験

BSCAの性能境界をさらに探るために、エントロピー閾値チューニングとショット数（クラスごとのラベル付きサンプル数）チューニングの2種類のハイパーパラメータチューニング実験を行う。

0.3は低いため、比較的少ない数の擬似ラベル付けされたターゲットサンプルを学習に使用することができ、モデルのアンダーフィッティングを引き起こす。全体として、図5の傾向から、0.5が最高の平均分類性能を達成することが観察された。以下の実験では、エントロピー閾値はデフォルトで0.5に設定されている。

一方、ショット数チューニングの実験結果を表4と図6に示す。一般に、基本性能（1ショット）が95%以下の場合、ショット数の増加とともに分類精度が向上する。ショット数が3から10に増加した場合、平均精度は0.7%増加し、性能向上は限定的であることがわかる。実際の応用の観点から、我々は3ショット設定での実験に焦点を当てるが、これはいくつかの古典的なSSDA研究（Saito et al.）

表3

ResNet34をCNNバックボーンに用いた3ショット設定での閾値チューニング実験の分類精度 (%)。

しきい値	A→N	R→N	W→N	N→A	R→A	W→A	N→R	A→R	W→R	N→W	A→W	R→W	平均
0.3	94.1	88.7	93.8	97.1	96.2	98.1	82.4	82.1	83.1	99.3	99.7	95.5	92.5
0.4	96.6	89.5	96.4	97.0	97.2	97.5	83.7	81.9	82.9	98.3	99.7	96.5	93.1
0.5	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
0.6	95.5	91.1	95.6	96.9	97.3	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

表4

ResNet34をCNNバックボーンとして、クラスごとにラベル付けされたサンプル数を変えた場合の分類精度 (%)。

ショット番号	A→N	R→N	W→N	N→A	R→A	W→A	N→R	A→R	W→R	N→W	A→W	R→W	平均
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

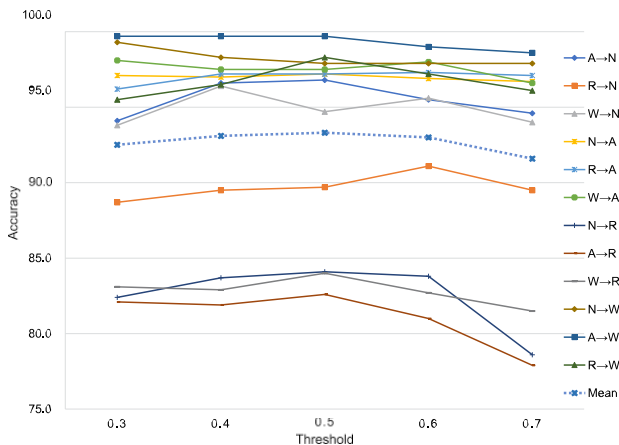


図5. ResNet34に基づく提案BSCAのエントロピー閾値を変えた場合の分類精度。

4.4. 比較実験

提案するRS-SSDAベンチマークの12種類の適応シナリオに基づいて、RSのクロスドメインシーン分類における我々のBSCAの有効性を公正かつ包括的に評価するために、3つの一般的なCNNバックボーンを用いて、我々の手法といくつかの最先端のUDA、SSL、SSDAアルゴリズムとの比較実験を行う。

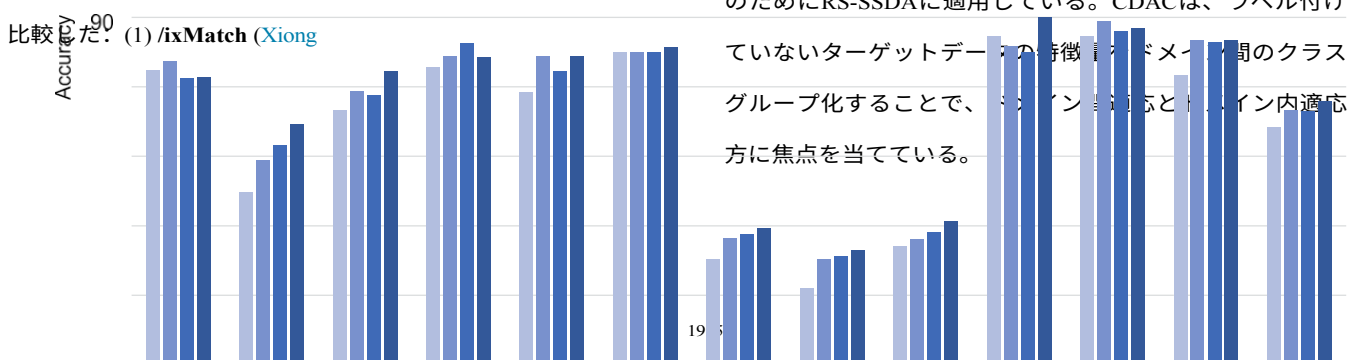
4.4.1. ベースライン

RSのクロスドメインシーン分類におけるSSL/UDAに対するSSDAの優位性を検証するために、我々のBSCAを以下のものと比較した。(1) **ixMatch** (Xiong

et al., 2021; Sohn et al., 2020) は、古典的なSSL手法の1つであり、

半教師付きRSシーン分類に適用され、この分野で最先端の結果の1つを達成している。より包括的な評価のため、我々のBSCAを、本研究で再現されたSSDA手法と比較した：

(3) **S+T**は、ソースデータとラベル付きターゲットデータのみを学習に使用する基本的な手法である；(4) **ADDA** (Wang et al., 2018; Tzeng et al., 2017)は、我々の知る限りRS-SSDAに関する最初の研究である。これは、2つの個別抽出器と共有分類器に基づいて、ソースデータとターゲットデータをadversariallyに訓練する；(5) **RevGrad** (Lasloum et al., 2021; Lu et al., 2019)は、DANN (Ganin et al., 2016; Saito et al., 2019)の手法から開発された、RSクロスドメインシーン分類に一般的に適用された。(6) **SSDAN** (Lasloum et al., 2021)は、MME (Saito et al., 2019)の手法から派生した、RS-SSDAのための現在の最先端アプローチである。(7) **ENT** (Grandvalet et al., 2005; Saito et al., 2019)は、ユニバーサルSSDA分野で広く適用されている古典的な半教師付き学習法 (Grandvalet et al., 2005)である。これは、より信頼性の高い分類のために、分類境界のエントロピーを減少させることを目的としている；そして、(8) **CDAC** (Li et al., 2021)は、コンピュータビジョンにおける最先端のSSDA手法の一つであり、本稿では比較のためにRS-SSDAに適用している。CDACは、ラベル付けされていないターゲットデータの特徴量とドメイン間のクラスにグループ化することで、クロスドメインとドメイン内適応の両方に焦点を当てている。



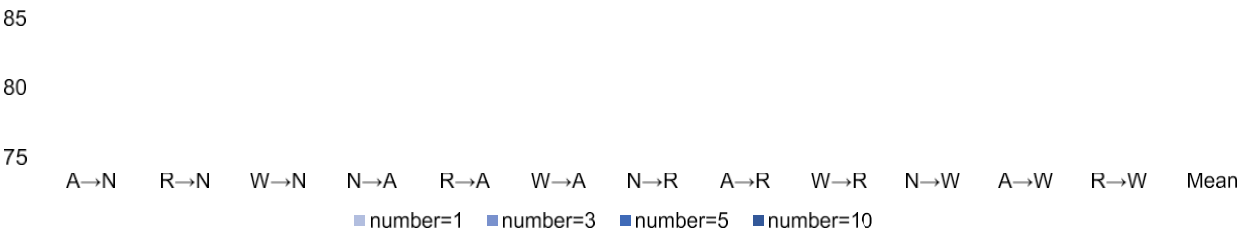


図 6. クラスごとにラベル付けされたターゲットサンプルの数を変えた場合の分類精度。

表5

VGG16をCNNバックボーンとして、3ショット設定でRS-SSDAデータセットの12の適応シナリオの比較実験の分類精度（％）。最も良い結果は**太字**である。

タイプ	方法	A→N	R→N	W→N	N→A	R→A	W→A	N→R	A→R	W→R	N→W	A→W	R→W	平均
SSL	フィックスマッチ (Xiongら、2021年)		77.9			96.3			67.5			93.4		83.8
UDA	ECB-FAM (マーら、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.)	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
	レブグラッド (ラスロームら、2021年)	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 (Lasloum et al., 2021)	94.6	84.7	89.5	96.8	94.2	95.5	79.6	80.6	81.8	98.3	99.3	96.9	91.0
	耳鼻咽喉科 (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 (李ら、2021年)	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
BSCAについて		96.5	94.2	95.9	96.5	94.5	96.5	78.0	82.8	84.7	98.6	99.0	99.3	93.0

表6

ResNet34をCNNバックボーンとして、RS-SSDAデータセットの12の適応シナリオについて、3ショット設定での比較実験の分類精度（％）。最も良い結果は**太字**である。

タイプ	方法	A→N	R→N	W→N	N→A	R→A	W→A	N→R	A→R	W→R	N→W	A→W	R→W	平均
SSL	フィックスマッチ (Xiongら、2021年)		81.3			95.0			63.5			97.6		84.4
UDA	ECB-FAM (マーら、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.)	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
	レブグラッド (ラスロームら、2021年)	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
	耳鼻咽喉科 (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 (李ら、2021年)	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
BSCAについて		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

表7

EfficientNet_B0をCNNバックボーンとして、RS-SSDAデータセットの12の適応シナリオについて、3ショット設定での比較実験の分類精度（％）。最も良い結果は**太字**である。

タイプ	方法	A→N	R→N	W→N	N→A	R→A	W→A	N→R	A→R	W→R	N→W	A→W	R→W	平均
SSL	フィックスマッチ (Xiongら、2021年)		72.2			93.2			63.0			96.9		81.3
UDA	ECB-FAM (マーら、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.)	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
	レブグラッド (ラスロームら、2021年)	90.9	78.4	84.5	92.7	84.4	91.5	72.8	69.6	72.0	94.1	99.0	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	99.0	87.2	85.2
	耳鼻咽喉科 (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 (李ら、2021年)	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
BSCAについて		96.1	88.3	93.8	96.5	95.8	96.4	77.0	78.9	83.0	98.3	98.6	98.6	91.8

ここでは、公式コード（S+T & ENT & SSDAN、²ADDA³RevGrad、⁴および CDAC⁵）、CNNモデルに対するこれらの手法の頑健性を、広く使われている3つのCNNバックボーンで評価する：(1)複雑な視覚表現のための古典的なディープCNNアーキテクチャであるVGG16 (Simonyan and Zisserman, 2014)、(2)スキップ残差接続を利用して勾配消失の問題を回避するResNet34 (He et al., 2016)、(3)神経アーキテクチャ探索手法によって設計され、効率的にモデルパラメータを削減するEfficientNet_B0 (Tan and Le, 2019)。

4.4.2. 最先端の手法との比較

VGG16、ResNet34、EfficientNet_B0の比較結果をそれぞれ表5、6、7に報告する。全体として、提案BSCAは、RS-SSDAベンチマークデータセットの12の適応シナリオの平均分類精度に関して最高の性能を達成した。BSCAは、VGG16、ResNet34、EfficientNet_B0において、それぞれ93.0%、93.3%、91.8%の平均分類精度を達成し、SSDANの下位最適手法に対して、2.0%、1.5%、6.6%の性能優位性を持つ。

² https://github.com/VisionLearningGroup/SSDA_MME.

³ <https://github.com/corenel/pytorch-adda>.

⁴ <https://github.com/fungtion/DANN>.

⁵ <https://github.com/lijichang/CVPR2021-SSDA>.

示し、VGG16はそれにわずかに及ばず、EfficientNet_B0がそれに続く。ここで、EfficientNet_B0が最悪の性能を示したことは特筆に値するが、これは訓練に十分なラベル付きデータがない場合に、その縮小メカニズムによって制限される可能性がある。RSシーンの画像には、地面に多数の物体が存在する。そのため、RSシーン画像に含まれる有用な情報が、知識伝達の過程で破棄される可能性がある。さらに、これらのCNNバックボーンをより包括的に評価するため、表8に示すように、パラメータ、計算複雑度（MACs（multiply-accumulates）により評価）、平均分類精度の3つの観点から、我々のBSCAの比較を行った。この表から、EfficientNet_B0はパラメータと計算量が最小であり、その結果、性能が相対的に低いことがわかる。一方、ResNet34は、効率的な残差接続と最大のモデルパラメータの恩恵を受けて、中間の計算複雑度で最高の性能を達成している。

手法の種類別では、SSL手法であるFixMatchは、UDAやSSDAと比較して劇的な差で最も悪い結果となり、RSシーン画像のターゲットドメイン分類には、半教師付き学習戦略よりもドメイン適応戦略の方が有効であることが検証された。ECB-FAMのUDA法では、その性能はソースドメインに影響される。例えば、VGG16上のNWPU-RESISC45の同じターゲットドメインでは、その精度は90.5%（AIDがソースドメイン）から74.6%（RSSCN7がソースドメイン）に低下し、その差は次のようになる。

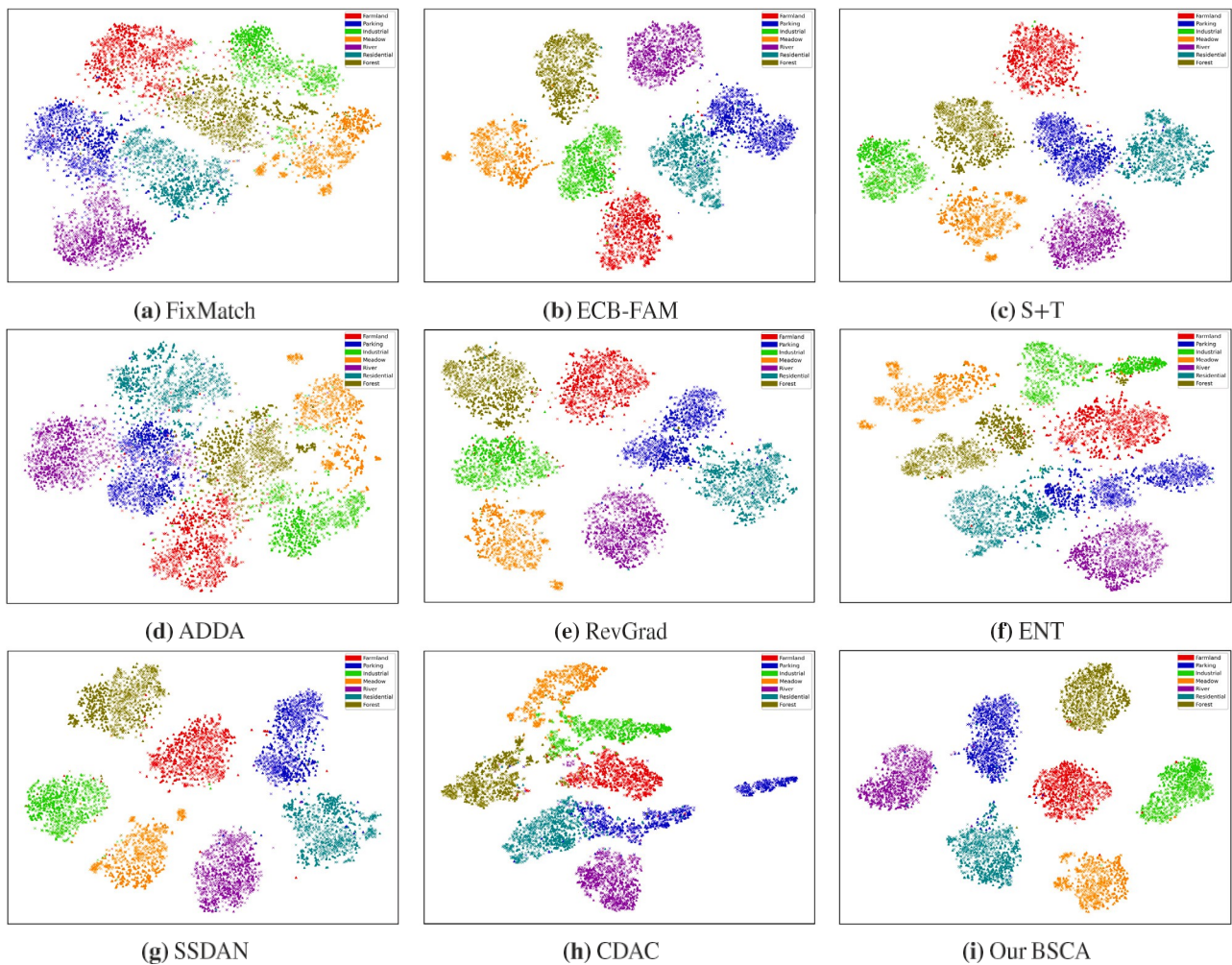


図7. ResNet34をCNNバックボーンとして、N→A適応シナリオにおける提案BSCAの手法と他の比較手法の特徴分布のt-SNE可視化。記号"x"と"•"はそれぞれソースサンプルとターゲットサンプルから抽出された特徴を表す。

表8

異なるCNNモデルに基づくBSCAのパラメータ、計算複雑度、12適応シナリオの平均精度の比較。

モデル	Params (M) ↓	MAC (G) ↓	平均精度 (%)
VGG16	14.98	15.35	93.0
ResNet34	21.55	3.68	93.3
EfficientNet_B0	4.67	0.41	91.8

15.9%。対照的に、SSLとUDAの組み合わせとして、実験のSSDAメソッドのほとんどは、ソースドメインの影響を受けにくく、より優れた安定した性能を有する。これは、SSDAがUDAとSSLの利点を組み合わせることで、よりロバストになったためと考えられる。

SSDA手法の観点から見ると、我々のBSCAはテストされたすべてのSSDA手法の中で最も優れた安定した性能を達成している。S+Tの基本手法と比較すると、SSDANはResNet34とVGG16に対して十分な改善を示している。しかし、SSDANは

EfficientNet_B0ではわずかな改善しか得られず、これは異なるCNNバックボーンに直面したときの不安定性を示している。ADDAは、おそらくそのユニークな分離されたソース抽出器とターゲット抽出器のため、性能が低く、これはドメイン間のモデル共有の重要性を明らかにする。RevGradは性能に対して何のメリットも示さず、ENTは一般的にクロスドメイン分類に対して有害でさえある。RS-SSDAにおいてCDACの性能が悪いのは、おそらくRSシーンのクラス間でオブジェクトと特徴の重複が多いため、ラベル付けされていない特徴類似度に基づくクラスタリングがうまくいかないためであろう。

この現象は、セクション4.2で説明したCDACの特徴分布から直感的に理解できる。

特定の適応シナリオの観点から、これらの実験結果は、WHUがターゲット・ドメインとなるいくつかのシナリオを除き、我々のBSCAがほとんどのシナリオで最良の知識伝達を達成できることを示している。その理由は、サンプル数が少ないため、クロスドメインアライメントの有効性とさらなる分類性能が制限されるためと考えられる。

一般的に、実験結果は、SSL、UDA、および他の最先端のSSDA手法と比較して、様々なRS適応シナリオとCNNバックボーンに直面した場合の提案BSCAの有効性、優位性、および安定性を実証している。

4.5. 特徴分布の可視化

提案するBSCAの特徴アライメント効果を直感的に示すために、図7ではt-SNE手法 (Van der Maaten and Hinton, 2008) を用いて、ResNet34に基づくN→Aの適応シナリオにおける、我々のBSCAと他の比較手法の特徴分布を図示している。

FixMatchのSSL法は、ターゲット特徴量を大まかに集約するが、ソース特徴量の利用を無視するため、クロスドメインの意味的ギャップが明らかになる。ECB-FAMのUDA法はクロスドメインの不一致を減少させるが、"Residential"や"Industrial"のような一部のクラスではクラス境界が明確ではない。SSDA法では、ベースラインであるS+T法と比較して、RevGradはほとんど意味のずれを示さない。

また、ENTのエントロピー最小化演算は、ドメイン間のクラス内特徴を分離させるという負の役割を果たすが、これは分類結果と一致する。SSDANは、クロスドメイン内クラス分布に干渉することなく、クラスを互いに遠ざけることでうまく機能する。CDACはRS-SSDAにおいて悪い結果を示すが、これはおそらく、2つの画像に含まれる物体と特徴の類似性に基づく、ラベルのない対象特徴の適応的クラスタリング手法による影響であろう。しかし、RS-SSDAでは、建物や道路など、異なるRSシーンクラス間で重複している物体や特徴が多く存在する。図7(h)に示すように、"Industrial"と"Residential"の特徴量は大きく混在しており、教師なし特徴量クラスタリングにより、両者の特徴量分布が不当に近くなっていることがわかる。これに対して、図7(i)は、我々のBSCAがクラス間の識別性を向上させるだけでなく、クラス内の領域間距離、すなわちクラス内分散を減少させることができ、少数の誤った擬似ラベル対象サンプルがあるにもかかわらず、最高の知識伝達性能を達成できることを明らかにする。全体として、この可視化は、我々のBSCAが、クロスドメイン特徴アライメントを介して、ソースドメインからターゲットドメインへの知識移転を達成できることを示している。

5. 結論

本論文では、リモートセンシング(RS)のクロスドメインシーン分類のための半教師付きドメイン適応(SSDA)を目的として、ソースドメインとターゲットドメイン間の意味的不一致を低減する双方向サンプルクラスアライメント(BSCA)手法を設計する。BSCAは、教師なしアライメント(unsupervised alignment: UA)と教師ありアライメント(supervised alignment: SA)の2種類のアライメント機構から構成される。まず、UAモジュールは、教師無しで、ソースドメインと未ラベルターゲットドメインの間だけでなく、ラベルターゲットドメインと未ラベルターゲットドメインの間のドメイン間特徴アライメントに集中し、それぞれのMMD距離を減少させることで達成される。次に、SAモジュールは、以下のサブステップによって、クラスレベルでのドメインシフトをさらに低減することを目的とする：メモリバンクから2つのドメインのクラス単位の中心を計算する；ソースサンプルから関連するターゲットクラスの中心へ、および、ターゲットサンプルから関連するソースクラスの中心への、サンプル-クラス特徴アライメントを行う；サンプルから抽出された特徴を2つのドメインのクラス単位のメモリバンクに格納する。いくつかの最新のSSL/UDA/SSDA手法と比較した結果、我々のBSCA

は、クラス内特徴表現と低エントロピーの分類境界により、ラベルのないターゲットドメインにおいて優れた分類性能を達成した。この結果は、提案手法の有効性と頑健性を示している。

将来的には、その結果生じるパフォーマンスのボトルネックを考慮する。

我々のBSCAでは、誤った擬似ラベル付きターゲットサンプルから、カリキュラム学習のような動的閾値技術によって擬似ラベルの品質を向上させることに焦点を当てる。

利益相反宣言

著者らは、本論文で報告された研究に影響を及ぼすと思われるような、競合する金銭的利益関係や個人的関係はないことを宣言する。

謝辞

本研究は、ドイツ研究財団 (DFG GZ: ZH 498/18-1、プロジェクト番号: 519016653)、欧州連合 (EU) の研究・イノベーションプログラム「ホライゾン2020」の下での欧州研究会議 (ERC) (助成金契約番号: [ERC-2016- StG-714087], 頭字語: So2Sat)、ヘルムホルツ優秀教授枠を通じたヘルムホルツ協会 (Helmholtz Association) による共同支援を受けている。[ERC-2016- StG-714087]、頭字語: So2Sat)、ヘルムホルツ協会によるヘルムホルツ・エクセレント・プロフェッサーシップ「地球観測におけるデータ科学-都市研究のためのビッグデータ融合」(助成金番号: W2-W3-100)、ドイツ連邦教育省によるもの。

AI4EO-地球観測のための人工知能」(助成金番号: 01DD20001)、およびドイツ連邦経済・気候行動省による国際的な未来AIラボ「AI4EO-地球観測のための人工知能」(助成金番号: 01DD20001)の枠組みによるものです: 推論、不確実性、倫理、そしてその先へ」(助成金番号: 01DD20001)、およびドイツ連邦経済・気候行動省による「ナショナル・センター・オブ・エクセレンスML4Earth」(助成金番号: 50EE2201C)の枠組み。

参考文献

- Ahmed, N., Saha, S., Shahzad, M., Fraz, M.M., Zhu, X.X., 2021. 衛星画像における森林マッピングのための漸進的非スーパービジョン深層伝達学習。In: Proceedings of the IEEE/CVF International Conference on Computer Vision.
- Alonso, I., Sabater, A., Ferstl, D., Montesano, L., Murillo, A.C., 2021. クラスワイズ・メモリ・バンクからのピクセルレベルのコントラスト学習による半教師付き意味分割。In: IEEE/CVF International Conference on Computer Vision.
- Bai, L., Liu, Q., Li, C., Ye, Z., Hui, M., Jia, X., 2022. オクターブ畳み込みによるマルチスケール特徴融合共分散ネットワークを用いたリモートセンシング画像シーン分類。IEEE Trans.Geosci.Remote Sens.
- Borgwardt, K.M., Gretton, A., Rasch, M.J., Kriegel, H.P., Schölkopf, B., Smola, A.J., 2006. カーネル最大平均不一致による構造化生物学的データの統合。Bioinformatics 22, e49-e57.
- Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C.F., Huang, J.B., 2019. A closer look at few-shot classification. arXiv preprint arXiv:1904.04232.
- Cheng, G., Han, J., Lu, X., 2017. Remote sensing image scene classification: Benchmark and state of the art. Proc. IEEE 105, 1865-1883.
- Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V., 2020. Randaugment: 探索空間を縮小した実用的な自動データ補強。In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: 大規模階層型画像データベース. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp.248-255.
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V., 2016. Domain-adversarial training of neural networks. J. Mach. Learn. Res. 17, 2096-2030.
- Grandvalet, Y., Bengio, Y., et al., 2005. エントロピー最小化による半教師付き学習. cap 367, 281-296.
- Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A., 2012a. A kernel two-sample test. J. Mach. Learn. Res. 13, 723-773.
- Gretton, A., Sejdinovic, D., Strathmann, H., Balakrishnan, S., Pontil, M., Fukumizu, K., Sriperumbudur, B.K., 2012b. 大規模2標本検定における最適カーネル選択. In: 神経情報処理システムの進歩. Citeseer, pp.1205-1213.
- Han, W., Feng, R., Wang, L., Cheng, Y., 2018. A semi-supervised generative framework with deep learning features for high resolution remote sensing image scene classification. ISPRS J. Photogramm. 145, 23-43.
- He, K., Fan, H., Wu, Y., Xie, S., Girshick, R., 2020. (1)学習する視覚表現(2)学習する視覚表現(3)学習する視覚表現. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- Jiang, P., Wu, A., Han, Y., Shao, Y., Qi, M., Li, B., 2020. 半教師付き領域適応のための双方向敵対的学習. IJCAI: IJCAI. pp.
- Kim, T., Kim, C., 2020. 惹きつけ、振動させ、探索する: このような学習は、学習者の学習意欲を向上させるだけでなく、学習者の学習意欲を向上させ、学習者の学習意欲を向上させる。In: European Conference on Computer Vision. Springer, pp.591-607.

- ン分類のためのマルチソース半教師付き領域適応ネットワーク。リモートセンシング。
- Li, D., Hospedales, T., 2020. 多ソース・半教師付きドメイン適応のためのオンラインメタ学習。In: European Conference on Computer Vision. Springer, pp. 382-403.
- Li, J., Li, G., Shi, Y., Yu, Y., 2021. 半教師付きドメイン適応のためのクロスドメインアダプティブクラスタリング。In: IEEE/CVF Conference of the Computer Vision and Pattern Recognition.
- Lin, M., Chen, Q., Yan, S., 2013. ネットワークの中のネットワーク。Comput.Sci.
- Long, M., Cao, Y., Wang, J., Jordan, M., 2015. Learning transferable features with deep adaptation networks. In: International Conference on Machine Learning. PMLR, pp. 97-105.
- Lu, X., Gong, T., Zheng, X., 2019. Multisource compensation network for remote sensing cross-domain scene classification. IEEE Trans. Geosci. 58, 2504-2515.
- Lu, X., Gong, T., Zheng, X., 2020. リモートセンシングのクロスドメインシーン分類のためのマルチソース補償ネットワーク。IEEE Trans. Geosci. 58, 2504-2515. <http://dx.doi.org/10.1109/TGRS.2019.2951779>.
- Lv, H., Chen, C., Cui, Z., Xu, C., Li, Y., Yang, J., 2021. メタプロトタイプネットワークを用いた動画中の正常ダイナミクスの学習。In: このような場合、「閾値」を設定することで、「閾値」が設定された「閾値」よりも高い閾値が設定された「閾値」よりも低い閾値が設定されることになる。

- Ma, L., Li, M., Ma, X., Cheng, L., Du, P., Liu, Y., 2017.教師ありオブジェクトベースの土地被覆画像分類のレビュー。ISPRS J. Photogramm.130, 277-293.
- Ma, C., Sha, D., Mu, X., 2021b.リモートセンシングシーン 分類のためのエラー補正境界と特徴適応メトリックを用いた教師なし敵対的領域適応。Remote Sens.
- Ma, A., Wan, Y., Zhong, Y., Wang, J., Zhang, L., 2021a.SceneNet: Remote sensing scene classification deep learning network using multi-objective neural evolution architecture search.ISPRS J. Photogramm.172, 171-188.
- Miao, W., Geng, J., Jiang, W., 2022.表現一貫性シャムネットワークを用いた半教師付きリモートセンシング画像シーン分類。IEEE Trans.Geosci.Remote Sens.
- Motiani, S., Piccirilli, M., Adjero, D.A., Doretto, G., 2017.Unified deep supervised domain adaptation and generalization.In : Proceedings of the IEEE International Conference on Computer Vision.
- Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q., 2010.Domain adaptation via transfer component analysis.IEEE Trans.22, 199-210.
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017.Pytorch における自動分化。
- Qiu, C., Mou, L., Schmitt, M., Zhu, X.X., 2019.リカレント残差 ネットワークを用いた多季節センチネル-2 画像からの局所気候ゾーンベースの都市土地被覆分類。ISPRS J. Photogramm.154, 151-162.
- Ranjan, R., Castillo, C.D., Chellappa, R., 2017.L2-constrained softmax loss for discriminative face verification. arXiv.
- Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., Breitkopf, U., 2012.The isprs benchmark on urban object classification and 3d building reconstruction.ISPRS Ann.Photogram. リモートセンシング. Spat.1-3 (2012), 293-298, Nr. 1 I.
- 齋藤和彦、キム、D.、Sclaroff, S.、Darrell, T.、Saenko, K., 2019.最小エントロピーによる半教師付きドメイン適応。In : Proceedings of the IEEE/CVF International Conference on Computer Vision.
- Simonyan, K., Zisserman, A., 2014.Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L., 2020.Fixmatch : 一貫性と信頼性を持つ半教師付き学習の簡略化。Adv. Neural Inf.Process.Syst.33, 596-608.
- Song, S., Yu, H., Miao, Z., Zhang, Q., Lin, Y., Wang, S., 2019.畳み込みニューラルネットワークベースのリモートセンシングシーン分類のためのドメイン適応。IEEE Geosci.Remote Sens.Lett.16, 1324-1328.
- Tan, M., Le, Q., 2019.Efficientnet: Rethinking model scaling for convolutional neural networks.In : International Conference on Machine Learning.PMLR, pp.Tzeng, E., Hoffman, J., Saenko, K., Darrell, T., 2017.Adversarial discriminative domain adaptation.In : コンピュータビジョンとパターンに関するIEEE会議の議事録 7167-7176ページ。
- Van der Maaten, L., Hinton, G., 2008.t-sneを使ったデータの可視化。J. Mach.Learn. Res.9。
- Wang, R., Collins, L.M., Bradbury, K., Malof, J.M., 2018.Semisupervised adversarial discriminative domain adaptation, with application to remote sensing data.In : IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium.IEEE, pp.3611-3614.
- Wang, Q., Huang, W., Xiong, Z., Li, X., 2020.シーンをより近くに見る: リモートセンシング画像シーン分類のためのマルチスケール表現学習。IEEE Trans.Neural Netw.Learn.Syst.
- Wurm, M., Stark, T., Zhu, X.X., Weigand, M., Taubenböck, H., 2019.Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks.ISPRS J. Photogramm.150, 59-69.
- Xiong, Y., Xu, K., Dou, Y., Zhao, Y., Gao, Z., 2021.Wrmach: 数ショットのリモートセンシングシーン 分類のための重み付き核ノルム正則化によるfixmatchの改善。IEEE Trans.Geosci.60, 1-14.
- Xiong, Z., Zhang, F., Wang, Y., Shi, Y., Zhu, X.X., 2022.Earthnets: 地球観測におけるaiのエンパワメント. arXiv preprint arXiv:2210.04936.
- Yu, T., Lin, J., Mou, L., Hua, Y., Zhu, X., Wang, Z.J., 2022.Scida: 単一ラベルから複数ラベルへの航空画像からの自己補正統合領域適応。IEEE Trans.Geosci. リモートセンシング。
- Yuan, Y., Wang, D., Wang, Q., 2019.行動認識のための記憶増強時間動的学習。In : Proceedings of the AAAI Conference on Artificial Intelligence.
- Zhang, J., Liu, J., Pan, B., Shi, Z., 2020.リモートセンシング画像シーン分類のための 相関部分空間動的分布整列に基づく領域適応。IEEE Trans.Geosci.58, 7920-7930.
- Zheng, X., Gong, T., Li, X., Lu, X., 2021. 小規模データセットからのマルチタスク学習による一般化されたシーン分類。IEEE Trans.Geosci.Remote Sens.
- Zheng, J., Wu, W., Yuan, S., Zhao, Y., Li, W., Zhang, L., Dong, R., Fu, H., 2022a.A two-stage adaptation network (tsan) for remote sensing scene classification in single-source-mixed-multiple-target domain adaptation (s m² t da) scenarios.IEEE Trans.Geosci. リモートセンシング。
- Zheng, Z., Zhong, Y., Su, Y., 2022b.リモートセンシングクロスシーン分類のためのタスク別分類器フレームワークによるドメイン適応。IEEE Trans.Geosci.Remote Sens.
- Zhu, S., Du, B., Zhang, L., Li, X., 2021.Attention-based multiscale residual adaptation network for cross-scene classification.IEEE Trans.Geosci.60, 1-15.Zou, Q., Ni, L., Zhang, T., Wang, Q., 2015.Deep Learning based feature selection for remote sensing scene classification.IEEE Geosci.Remote Sens.Lett.12, 2321-2325.