SemiSANet： シャムネットワークを用いた半教師付き高解像度リモートセンシング画像変化検出モデル

------------------------------

呉江江$^{1}$ , 陳昊$^{1,2,}$ \*.

------------------------------

シャムネットワークを用いたモデル

------------------------------

National University of Defense Technology, College of Electronic Science and Technology, Changsha 410073, China; sunchengzhe@nudt.edu.cn (C.S.); wujiangjiang@nudt.edu.cn (J.W.); duchun@nudt.edu.cn (C.D.) Key Laboratory of Natural Resources Monitoring and Supervision in Southern Hilly Region, Ministry of Natural Resources, Changsha 410083, China Correspondence: hchen@nudt.edu.cn

------------------------------

要旨：変化検出（CD）はリモートセンシングの重要なアプリケーションの一つであり、災害評価、土地利用検出、都市スプロール追跡において重要な役割を果たしている。現在、高精度な完全教師付き手法がCDタスクの主な手法である。しかし、これらの手法では、二時間画像とその変化マップからなる大量のラベル付きデータが必要である。さらに、変化マップの作成には多大な労力と時間がかかる。この限界に対処するため、本論文では、一貫性正則化と強補強に基づく簡単な半教師付き変化検出法を提案する。まず、グラフ注目機構を持つシャム入れ子UNet(SANet)を構築し、少量のラベル付きデータで事前訓練する。次に、ラベル無しデータを事前学習済みSANetに投入し、信頼度閾値フィルタを用いて信頼度の高い擬似ラベルを得る。同時に、ラベルなしデータに対して強い増強処理を行うことで、歪んだ画像を生成する。モデルは、歪んだ画像のCD結果が対応する擬似ラベルと一致するように学習される。つの高解像度リモートセンシングデータセットを用いて広範な実験を行った。その結果、我々の手法が、ラベルが不十分な状況下での変化検出の性能を効果的に改善できることが実証された。我々の手法は、最先端の手法と比較して、IoUを25%以上向上させることができる。

------------------------------

キーワード：変化検出、リモートセンシング、半教師付き学習、データ補強、一貫性正則化

------------------------------

リモートセンシング(RS)画像の変化検出(CD)タスクは、異なる時刻に取得された同一地域のリモートセンシング画像に生じる意味的な変化を識別することを目的としている[1]。Sentinel、WorldView、GeoEye、ZY-3、GFなどの地球観測プログラムの増加により、様々なセンサーが、変化検出タスクのために、解像度の異なる多数のリモートセンシング画像を提供している。同時に、ディープラーニング(DL)技術の発展により、最近の変化検出手法は、対象地域の変化情報を迅速に取得できるようになった。これまでのところ、リモートセンシング画像変化検出技術は、生態系モニタリング[2]、土地被覆と土地利用マッピング[3]、被害評価[4]、都市拡大モニタリング[5]の分野で広く利用されている。リモートセンシング画像変化検出技術に関する研究は、数十年にわたってリモートセンシングコミュニティで行われてきた。従来の変化検出手法は、(1)画像演算ベース、(2)画像変換ベース、(3)ポスト分類ベースの3つのクラスに分けられます。画像演算に基づく方法は、まず、複数の時間画像のピクセル値を比較することで差分画像を生成し、しきい値を使用してピクセルを変化したクラスと変化していないクラスに分類します。主成分分析（PCA）のような画像変換に基づく方法は、画像のスペクトルの組み合わせを特定の特徴空間に変換することにより、変化したピクセルを識別する。ポスト分類ベースの手法は、まず2つの時間画像を独立に分類する。その後、分類結果を直接比較することで、変化検出結果を得ることができる[7]。現在、主な変化検出アルゴリズムは、ディープラーニングに基づく完全教師あり手法であり、高精度に到達するためには大量のデータを必要とする。リモートセンシング画像CDタスクでは、データは二時間RS画像であり、そのラベルは人工的にラベル付けされた変化画素である。二枚の画像のラベルを作成する場合、まず変化した部分を目で確認し、変化した部分と変化していない部分を255と0といった異なる画素値で塗りつぶす必要がある。この制約により、完全教師ありの手法をリモートセンシング画像の変化検出分野に広く適用することは困難である。通常、この限界を克服する方法は3つある。1つ目は、回転、シフト、切り抜き、反転などのデータ補強によってラベル付きデータ量を増やす方法である。この方法は、完全教師ありCDタスクの精度を向上させることができるが、拡大画像は元画像のバリエーションであるため、未見のデータセットではうまく機能しない可能性がある。もう一つは弱教師付き学習[8]であり、画像レベルのラベル[9]、バウンディングボックス[10]、落書き[11]によってデータに粗い注釈を付ける。一般に、粗くラベル付けされた情報は、ニューラルネットワークの教師あり学習をさらに導くための擬似ラベルを生成するのに役立つ。弱教師付き手法は、完全教師付き手法に比べて、労力と時間を節約できる。しかし、高解像度のリモートセンシング画像における複雑な領域CDタスクの場合、変化を人為的に識別し、粗くラベル付けするには、依然として多くの労力がかかる。第三の方法は、半教師付き学習（SSL）[12]である。前の2つの方法とは異なり、半教師付き学習は、少数のラベル付きデータと大量のラベルなしデータを使ってモデルを学習することを目的としている。半教師あり学習は、ラベル付けにかかる労力と時間を大幅に削減すると同時に、ラベル付けされていないデータの特徴を利用してオーバーフィッティングを回避することができる。リモートセンシングの分野では、画像の数は膨大であるが、ラベルを作成するのは困難である。リモートセンシング画像処理の性能を向上させるために、ラベルなしデータを利用するのは自然なことである。半教師あり学習は、生成的敵対ネットワーク（GAN）[13]、オートエンコーダ[14]、グラフニューラルネットワーク[15]などによって実現できる。リモートセンシングの分野では、半教師付き学習は、山火事の燃料マッピング[16,17]、ハイパースペクトル画像の分類[18]、セマンティックセグメンテーション[19]などに用いられている。しかし、半教師付きリモートセンシング画像変化検出を扱った研究はごく少数しかない[20]。例えば、Pengら[20]は、Generative Adversarial Network (GAN)に基づく半教師付きCD法を提案している。ラベル付けされたデータとラベル付けされていないデータとの間の検出結果の特徴分布の整合性を強化するために2つの識別器を構成することにより、大量のラベル付けされていないデータを利用することでモデルの性能を向上させている。この手法は半教師付き変化検出を実現することに成功しているが、GANの学習が安定せず、ハイパーパラメータの調整が難しいため、実際に適用するにはやや複雑である。さらに、ラベル付きデータが不十分な場合の変化検出の性能は改善できる。この限界に対処するため、本論文ではシンプルで効果的な半教師付き変化検出法を検討する。ラベル付けされていない画像に対して、強力な補強の有無に関わらず一貫した変化検出結果が得られるようにモデルを学習させるだけで、本手法は変化検出モデルの反干渉ロバスト性を向上させることができる。これにより、ラベル付けされた画像が不十分な場合でも高い変化検出精度が得られる。同時に、より良い変化検出性能を達成するために、グラフ注目メカニズム（GAT）を用いてモデルを改良する。本論文の主な貢献は以下の通りである：

------------------------------

出版社注：MDPIは中立の立場を保ちます。

------------------------------

における管轄権の主張に関して、MDPIは中立を保つ。

------------------------------

における管轄権の主張に関しては中立を保ちます。

------------------------------

この論文はオープンアクセス論文です

------------------------------

クリエイティブ・コモンズの条件

------------------------------

表示(CC BY)ライセンス (https:／／)

------------------------------

擬似ラベルと一貫性正則化に基づく新しい半教師付きRS画像変化検出フレームワークを提案する。まず、ラベル付きデータを用いる。

------------------------------

を採用し、ピクセルレベルの閾値フィルタの助けを借りて、ラベル付けされていないRS画像に対する高信頼度の擬似ラベルを生成する。そして、歪んだRS画像の変化検出結果を擬似ラベルと一致させることで、本モデルの性能と頑健性を向上させる。シャムユネットに基づく新しいRS画像変化検出ネットワークが提案され、グラフアテンションメカニズムの追加により、モデルは潜在特徴間の長距離依存性を抽出する能力を持つ。2つの高解像度リモートセンシングデータセットを用いた広範な実験結果から、本モデルが、ラベル数が不十分な場合の変化検出結果を大幅に改善し、最先端の変化検出手法を凌駕することが実証される。この半教師付きCD法の最適なハイパーパラメータ値の範囲は、広範なアブレーション実験によって与えられる。

------------------------------

本節では、CDのためのDLフレームワーク、注意メカニズム、一貫性正則化に基づく半教師付き手法を簡単に説明する。

------------------------------

を用いてネットワークを学習し、ピクセルレベルの閾値フィルタの助けを借りて、ラベル付けされていないRS画像に対する高信頼度の擬似ラベルを生成する。そして、歪んだRS画像の変化検出結果を擬似ラベルと一致させることで、本モデルの性能と頑健性を向上させる。シャムユネットに基づく新しいRS画像変化検出ネットワークが提案され、グラフアテンションメカニズムの追加により、モデルは潜在特徴間の長距離依存性を抽出する能力を持つ。2つの高解像度リモートセンシングデータセットを用いた広範な実験結果から、本モデルが、ラベル数が不十分な場合の変化検出結果を大幅に改善し、最先端の変化検出手法を凌駕することが実証される。この半教師付きCD法の最適なハイパーパラメータ値の範囲は、広範なアブレーション実験によって与えられる。

------------------------------

本節では、CDのためのDLフレームワーク、注意メカニズム、一貫性正則化に基づく半教師付き手法を簡単に説明する。

------------------------------

近年、ビッグデータ時代の到来とコンピューティングパワーの急速な発展に伴い、ディープラーニング（DL）アルゴリズムが研究のホットスポットとなっている[21]。ディープラーニングは、画像[22]、音声[23]、テキスト[24]を理解し、それらから意味情報を抽出するのに役立つ、複数のレベルの表現と抽象化を学習することができる。従来のアルゴリズムよりもはるかに強力なモデリング能力を持ち、画像処理、言語認識、自然言語処理などの人工知能分野で多くの成功を収めている。リモートセンシング画像処理の分野でも、ディープラーニングは優れた性能を示しており[25]、画像レジストレーション[26]、相互重ね合わせ[27]、道路抽出[28]、画像-地図変換[29,30]、画像セグメンテーション[31]などの問題で広く利用されている。ディープラーニングはまた、リモートセンシング画像の変化検出問題に対する効果的なソリューションでもある[32]。リモートセンシングのビッグデータの恩恵を受け、完全教師あり学習に基づく多くのCDアルゴリズムが提案されている。CDタスクのためのDLネットワークのほとんどは、畳み込みニューラルネットワーク（CNN）に基づいている。その中でも、完全畳み込みネットワークに基づくUNetは最もポピュラーであり、多くの拡張とともにCDタスクのための標準的なCNNアーキテクチャの一つとなっている[32]。UNetは対称的なエンコーダ・デコーダ構造であり、ダウンサンプリング部では特徴量を抽出するために文脈情報を取り込み、アップサンプリング部では画像を再構成して最終的な変化検出結果を出力する[33]。エンコーダとデコーダの間にスキップ接続を追加することで、UNetは深い意味情報と浅い空間情報をよりよく統合し、CDの精度を向上させることができる。CDタスクの入力が両時間のRS画像であるという特徴から、Daudtら[34]はUNetにシャム構造を組み合わせ、エンコーダ部を共有パラメータを持つデュアルチャネル構造とした。この研究では、まず2つの時間帯のリモートセンシング画像をそれぞれダウンサンプリング処理し、次に抽出された特徴を差分または接続によって統合し、デコーダの入力とする。この改良により、画像間の差異を比較するためのネットワークがより良くガイドされる。Shaoら[35]は、異なるサイズの画像を調整するためにUNetのエンコーダ部を修正することで、異なるサイズの衛星画像とUAV画像の変化検出を実現した。Pengら[36]は、より多くの中間（ネストされた）畳み込みブロックを追加し、UNetの接続をスキップすることで、ネストされたUNet（UNet++）構造をCDタスクに導入し、接続された特徴がより高い意味的類似性を持つようにし、エンコーダとデコーダの間の意味的ギャップをなくした。Fangら[37]は、前の2つの研究を組み合わせ、ネストされたUNetに基づくシャムネットワークを提案し、ニューラルネットワークにおける深部局在情報の損失を緩和した。

------------------------------

注意メカニズムは人間の視覚システム[38]に由来し、画像処理の分野で研究のホットスポットとなっている。これは、ネットワークが異なる特徴をより効果的に融合することを可能にし、画像処理問題の精度向上に大きな効果をもたらす。Zhangら[6]は、ネットワークのデコーダ部分にチャネル注意機構と空間注意機構を導入し、ネットワークがアップサンプリング処理中に情報をより効果的に利用できるようにした。チャネル注意メカニズムは、多層パーセプトロン（MLP）ネットワークを通じて、CD結果に対する異なるチャネルの寄与を学習し、寄与に応じて各チャネルに割り当てられる注意を決定する。空間的注意メカニズムは、7×7の畳み込みブロックを通して異なる位置の寄与を学習する。異なる位置と時間における個々の画素間の時空間相関を完全に利用するために、Chenら[39]は時空間注意メカニズムを提案した。時空間注意メカニズムは自己注意メカニズム[40]に触発されたもので、3つの異なる畳み込み層を利用し、入力特徴テンソルからクエリー、キー、バリューテンソルを計算する。出力ベクトルは、計算によってこれら3つのベクトルをマッピングすることで得られる。出力ベクトルを得る過程では、画素間の空間的・時間的依存関係を十分に利用し、照明に不変で、ミスレジストレーションに頑健な特徴を得ることができる。グラフアテンション（GAT）機構[41]は、強力な長距離特徴集約能力を持つグラフベースのアテンション手法である。近年、リモートセンシング分野におけるグラフネットワーク手法の応用において、いくつかの関連する成果がある[31,42]。31]では、ZiらがGATと自己アテンションに基づく画像セグメンテーション手法を提案し、エンコーダ部で抽出された特徴量をグラフのノードとして扱い、GATを用いてこれらの特徴量を融合する。グラフの長距離情報集約能力により、特徴量間の関係性を掘り起こし、活用することができる。

------------------------------

2.3. 一貫性正則化に基づく半教師付き手法

------------------------------

半教師付き学習(Semi-supervised learning, SSL)は、完全教師付き学習と教師なし学習の中間に位置するディープラーニングの一分野である。SSLは、大量のラベルなしデータと少量のラベル付きデータを組み合わせて、特定の学習タスクを達成することを目的としている[12]。一貫性正則化に基づく半教師付き学習法は、SSLの一種であり、近年人気がある。一貫性正則化の原則の主な仮定は、入力またはモデルに摂動が加えられる前と後で、モデルの出力が一貫していなければならないということである。一貫性正則化は以下の式で測定できる：

------------------------------

ここでDはメトリック関数であり、クロスエントロピー損失、二乗平均損失などがある。p$\_{model}$はモデル、θはモデルパラメータである。半教師付きアプローチでは、オーグメンテーション法は画像を歪める。一貫性正則化の原理は、異なる歪みを持つ同じ画像に対してモデルが同じ予測を行うことを強制し、モデルの頑健性と汎化性を高める。一貫性正則化の原理に基づいて、Laineら[43]はΠモデルを提案した。学習プロセスの各エポックにおいて、同じラベルなしデータが2回前方に伝播され、ネットワークに入る前に確率論的増大を経て、2つの異なる予測をもたらす。Πモデルは、これらの予測ペアの差を縮めるように訓練することで、モデルのロバスト性を向上させる。改良されたモデルである時間的アンサンブルも[43]で提案されている。テンポラル・アンサンブリングでは、ラベル付けされていないデータは各エポックで1回だけ前方に伝播され、前のエポックから得られた予測値は指数移動平均（EMA）後の出力擬似ラベルとして使用される。Tarvainenら[44]は、指数移動平均のアイデアをモデル自体に適用したMean Teacherモデルを提案した。具体的には、バックプロパゲーションアルゴリズムによってパラメータが更新される生徒モデルと、以前のエポックにおける生徒モデルのパラメータの指数移動平均によってパラメータが更新される教師モデルがある。Wangら[19]は、Mean Teacherを用いることで、半教師付きリモートセンシング画像セマンティックセグメンテーションに成功している。半教師付き学習ネットワークの性能をさらに向上させるために、Berthelotら[45]はエントロピー最小化の原理を組み合わせてMixMatchを提案した。最小化エントロピーの原理とは、ラベル付けされていないデータに対して、モデルが低エントロピーの予測を行うことを意味する。擬似ラベルは最小化エントロピーの原理を応用したものである。MixMatchは、入力されたラベルなしデータの予測値を平均化しシャープにすることで、擬似ラベルを作成する。MixMatchは従来の半教師付き手法よりも優れた結果を達成する。摂動をよりドラスティックにするために、Xieら[46]は入力データを処理するために、画像の形や色に劇的な変化を引き起こす可能性のある強いオーグメンテーションを導入し、UDAモデルを提案した。擬似ラベルは、フリップやシフ トなどの弱いオーグメンテーションを行った後の入力データの 予測値をシャープにすることで得られた。次に、同じデータに対して強い補強を行い、学習用ネットワークに投入した。UDAモデルにより、MixMatchを超える性能が達成された。Sohnら[47]は、UDAモデルに基づいて擬似ラベルを強化するFixMatchを提案した。FixMatchは予測をシャープにする代わりに、閾値を超える予測を選択することで擬似ラベルを作成し、モデルをミスリードする偽の擬似ラベルを効果的に防ぐ。一貫性正則化に基づく上記の半教師付き手法は、画像分類領域において満足のいく精度を達成することができた。しかし、画像分類に使われる画像は、オブジェクトの数が少ない。例えば、1枚の画像に含まれるのは、猫1匹、リンゴ数個などである。対照的に、リモートセンシング画像は、はるかに豊富な意味情報を持っている。つまり、リモートセンシング画像に摂動を与える際に、利用可能な意味情報が破壊されやすく、CDの結果が悪くなる可能性がある。半教師付き学習ベースのCDタスクを実装するために、どのように一貫性正則化を効果的に使用するかは、調査する価値のある問題である。

------------------------------

このセクションでは、まず提案するセミSANetのアーキテクチャを説明する。次にSANetの詳細を示す。最後に、SANetの教師あり部分と教師なし部分の損失関数を定義する。我々の提案する半教師付きCD法のアーキテクチャを図1に示す。これは教師あり部分と教師なし部分の2つから構成される。教師ありパートでは、ラベル付き二時 間リモートセンシング画像を弱補強後にSANetに入力し、生成された予測値 Ŷ$\_{l}$ を二値ラベルY$\_{l}$ で損失関数計算を行い、教師あり損失L s を求める。 教師なしパートでは、ラベルなしデータをモデルに入力し、擬似ラベル Ŷ$\_{pl}$ と予測値↪Lu76↩ u を求める。まず、ラベルなしデータに対して、ランダムな反転やシフトを含む弱補強を行い、X w を得る。その後、X w は2つの処理の流れに従う。最初の処理ストリーム（図1の緑矢印）では、X w がモデルに入力され、2値画像 ↪Lu\_1 w が得られる。Ŷ w は次に、擬似ラベル Ŷ$\_{pl}$ を得るために、強い補強と信頼度閾値フィルタにかけられる。 もう1つの処理の流れ（図1の黄色矢印）では、X w はまず X s を得るために強い補強にかけられ、次に予測値 Ŷ u を得るためにモデルに投入される。擬似ラベルŶ$\_{pl}$ と予測値 Ŷ u のピクセル毎の対応を保証するために、ここで使われる2つの強い補強は同一であることに注意。教師なし損失 L u は、擬似ラベルと予測値 ↪Lu\_176 u を用いて計算される。最終的な損失 L$\_{CD}$ は、教師あり損失 L s と教師なし損失 L u の重み付き和によって得られる。L$\_{CD}$ を用いてバックプロパゲートすることで、ネットワークパラメータを更新できる。

------------------------------

図1. semiSANet法のフローチャート。ラベル付けされたデータは、弱補強後に直接SANet を通過し、完全教師あり損失L s を得る。 ラベルなしデータは、弱補強（Ŷ w を得るため）と信頼度閾値フィルタを経 てSANetに供給され、擬似ラベルŶ$\_{pl}$ を得、同時に強 補強を用いて歪んだ画像を生成し、SANetによってCD結果 Ŷ u を得る。教師なし損失 L u は Ŷ u と Ŷ$\_{pl}$ によって計算される。

------------------------------

我々の半教師付き処理は、以下の式でより簡潔に示すことができる：

------------------------------

θとP$\_{SANet}$はそれぞれSANetのパラメータと予測値を表す。Aug wとAug sは、それぞれ弱い補強と強い補強を表す。CTFは信頼閾値フィルタ、D$\_{l}$とD uはそれぞれ教師あり部分と教師なし部分に使われる損失関数を表す。

------------------------------

3.1.1. semiSANetにおけるデータ補強

------------------------------

semiSANetで使用されるデータ補強は、弱い補強と強い補強に分けられる。弱いオーグメンテーションには、ランダムな反転とシフティングがあり、 データセットを拡張するために使用される。本研究で使用される強オーグメントは、Cubuk[48]によるRandaugmentという研究を改良したものである。データ拡張にRandaugment戦略を用いながら、リモートセンシング画像の特性に応じて追加と削除を行い、最終的に合計15の強力な拡張手法を得た。表1に、本研究で使用した強補強手法とその強度範囲、および具体的な説明を示す。

------------------------------

表1. 強力な補強方法とその程度と説明

------------------------------

画像の明るさを制御する

------------------------------

画像のカラーバランスを調整する

------------------------------

各色チャンネルのビット数を減らす

------------------------------

画像を水平軸に沿って切断する

------------------------------

縦軸方向に画像を剪断する

------------------------------

しきい値以上のすべてのピクセル値を反転します。

------------------------------

図2は、リモートセンシング画像に様々な強補強法を適用した結果を示している。リモートセンシング画像に強オーグメンテーションを用いる場合、表1に示す手法（切り出しは常に選択される）からランダムに4つのオーグメンテーションを選択し、強オーグメンテーション手法を形成し、その効果を図3に示す。

------------------------------

図2. リモートセンシング画像で使用される強オーグメンテーション。

------------------------------

図3. リモートセンシング画像にランダムに適用される複数の強力な補強方法。

------------------------------

各補強の次数はランダムであり、次数の範囲は表1の通りである。ネットワークを学習する際、強い補強が十分にランダムであることを保証するために、同じエポックにおける異なる二時期のリモートセンシング画像のペア、および異なるエポックにおける同じ画像のペアに対して、異なる強い補強を使用する。しかし、擬似ラベルŶ$\_{pl}$ と教師なし予測値 Ŷ u の画素が互いに対応することを保証するために、同じエポックで同じペアの二時 間リモートセンシング画像とその擬似ラベルに対して、同じ強補強を用いるべきであることに注意する必要がある。強い補強は2つの目的を果たす。1つは、色補強によってリモートセンシング画像間の色ギャップを増減させることである。もう一つは、形状補強によって特徴を変形させることである。一貫性正則化は、モデルがこれらの変化を無視し、オブジェクトの変化に焦点を当てることを要求し、意味理解能力とモデルの頑健性を向上させる。

------------------------------

信頼度閾値フィルタはFixMatch[47]から参照され、閾値τを用いて信頼度の低い擬似ラベルをフィルタリングすることを目的としている。変化検出タスクの特性上、擬似ラベルをピクセルレベルでフィルタリングする。図4に示すように、閾値τより小さいピクセルはマスクされる。擬似ラベルŶ$\_{pl}$ と予測値 Ŷ u を用いて教師なし損失 L u を計算するとき、マスクされた画素は計算に関与しない。信頼度閾値τ以上のピクセルのみが教師なし損失L uの計算に関与し、誤ったピクセルレベルの擬似ラベルによるモデルの学習不良を防ぐことができる。

------------------------------

図4. 赤い部分はτより小さく、損失計算に関与しないためマスクされている。

------------------------------

SANetは、semiSANetでCDに使用されるエンド・ツー・エンドのエンコーダ・デコーダネットワークである。SANetはUNetのバックボーンを持ち、エンコーダとデコーダの間を密にスキップ接続することで、高解像度の特徴量ときめ細かな局在情報を保持し、マルチスケール情報を効率的に抽出することができる。UNetとは異なり、CDタスクの特性に適応するため、SANetのエンコーダ部分としてシャムネットワークを用いる。図5に示すように、ビット時間画像はシャムネットワークの2つのFCN分岐にそれぞれ入力される。畳み込みカーネルのサイズによって制限されるため、CNNは特徴の長距離依存性を融合することができない。我々は、完全な畳み込みネットワークから抽出された特徴を用いてグラフを構築する。特徴量は{$^{~}$ h$\_{1}$ ,$^{~}$ h$\_{2}$ ,$^{~}$ h$\_{3}$ , - - ,$^{~}$ h$\_{N}$ }で表されるノードとして使用し、Nは特徴量の数を表す。ノードiとノードjの間のエッジの重み、及び$^{~}$ h$\_{i}$ と$^{~}$ h$\_{j}$ の間の注目係数をa$\_{ij}$ とする。グラフ注意メカニズムにより、図5bに示すように、隣接ノードの情報を$^{~}$ h$\_{i}$に融合させることができる。まず、注目係数a$\_{ij}$を以下のように計算する：

------------------------------

ここで、～aとWは学習可能なパラメータであり、ニューラルネットワークの学習により自動的な注意配分を実現できる。｜｜ は連結演算を表す。

------------------------------図5.SANetの説明図。(a）SANetのバックボーン。(b）多頭グラフ注意ユニットと単頭グラフ注意ユニットを含むGATブロック。(c）注意ユニット。

------------------------------

この過程で、ノードiは隣接ノードの特徴を融合する。N$\_{i}$はiを含むiの近傍ノードのインデックスの和であり、$^{~}$ h$\_{i}^{0}$は以下のように計算される：

------------------------------

図5bに示すように、モデルで使用する最初のGAT層はマルチヘッド注意を持つGATであり、注意メカニズムの汎化能力を向上させることができる。互いに独立したK組の注意層を用い、それらの結果を連結する。多頭注意メカニズムを用いた$^{~}$ h$\_{i}^{0}$の計算式は以下の通りである： k

------------------------------

特徴量の数はグラフのサイズを直接決定し、さらにGPUメモリによって制限されるパラメータ数を決定するため、グラフ注目機構をエンコーダの最後に追加し、抽出された最も深い意味特徴を用いてグラフを構成するだけである。また、エンコーダで抽出された特徴情報と空間コンテキスト情報を、二時期のリモートセンシング画像の同じ位置に対応させるため、GATネットワークを含むシャムネットワークの両チャンネルで重みを共有する。さらに、高レベル特徴量と低レベル特徴量の間の意味的ギャップを抑制するために、エンコーダとデコーダの間の密なスキップ接続にチャネルアテンション機構を追加した[20]で言及された密なアテンション接続を採用する。チャネルアテンションメカニズムは、深い意味情報と浅い空間情報を接続する際に、異なるチャネルに重みを割り当てることができ、深い情報と浅い情報の間の意味ギャップを抑制することができる。

------------------------------

変化しない画素数が変化した画素数よりはるかに多いCDタスクでは、損失関数としてクロスエントロピー損失L ceとダイス損失L$\_{dice}$の組み合わせを用いる。ラベルをY、CDの結果をŶ = ŷ$\_{k}$ 、k = 1, 2, . . ここで、Hは画像の高さ、Wは画像の幅、ŷ$\_{k}$ は画像のk番目の画素の画素値である。cは0または1とし、ラベルのk番目の画素の変化の有無を表す。クロスエントロピー損失とサイコロ損失は以下のように計算される：

------------------------------

Ŷ$\_{pl}$を擬似ラベルとすると、教師あり損失L sと教師なし損失L uは以下のように定義される：

------------------------------

半教師付き手法全体の最終的な損失は、L s と L u の重み付き和によって得られる。重み付け係数をλとすると、最終損失は以下のように計算される：

------------------------------

明らかに、λは教師なし損失のモデルへの影響度を決定し、λ = 0のとき、モデルは完全教師ありモデルに縮退する。

------------------------------

この方法の有効性を検証するために、我々は2つのリモートセンシング画像CDデータセットを選択した：LEVIRデータセットとWHUデータセットであり、どちらもVHRデータセットであり、多数の画像を含んでいる。

------------------------------

LEVIRデータセット(https:／／justchenhao.github.io／LEVIR／, accessed on 3 May 2022)は、Google EarthのVHR画像637枚からなる大規模なリモートセンシング建物変化検出データセットである。LEVIRデータセットの二重時間画像は、米国テキサス州の複数の都市にある20の異なる地域から取得され、2002年から2018年にかけて撮影されたものである。このデータセットは、主に建物における土地利用の変化に焦点を当てており、コテージハウス、高層アパート、小さなガレージ、大きな倉庫など、様々な種類の建物をカバーしている。画像には季節や照明による擬似的な変化が多く、建物による変化も形や大きさが異なるため、非常に難易度の高いデータセットとなっている。GPUメモリの制限から、画像ペアを256×256ピクセルに分割し

------------------------------

に分割し、ランダムに学習セットとテストセットに分割する。トレーニングセットには7120枚の画像が含まれ、テストセットには1024枚の画像が含まれる。WHUデータセット（https:／／study.rsgis.whu.edu.cn／pages／download／building\_dataset.html、2022年5月3日アクセス）には、2012年と2016年にニュージーランドのクライストチャーチの同じ地域で撮影された2つのリモートセンシング画像とその変化マップが含まれている。各画像のサイズは32,507 × 15,345ピクセル、解像度は0.075 m/ピクセルで、主な対象は建物です。GPUメモリの制限により、LEVIRデータセットと同様に、256×256ピクセルで分割し、2000の画像ペアからなるトレーニングセットと996の画像ペアからなるテストセットを得た。WHUデータセットは、変化オブジェクトのばらつきが大きく、極めて異質な分布をしているため、難しいデータセットである。

------------------------------

様々な手法のCD効果を効果的に評価するため、評価指標としてF1スコア、総合精度（OA）、カッパ係数（Kappa）、交差-結合（IoU）を用い、これらは以下のように定義される：

------------------------------

TPを真陽性数、FPを偽陽性数、TNを真陰性数、FNを偽陰性数とすると、P、R、PREは以下のように計算できる：

------------------------------

提案手法はPytorchフレームワークで実装され、CPUはIntel Core i7-10700(2.9GHz、8コア、16GB RAM)、GPUはNVIDIA GTX 3090(24GBRAM)を搭載している。AdamWオプティマイザを採用し、学習率は1×10$^{-3}$、重み減衰は1×10$^{-2}$である。提案する半教師付き手法は、LEVIRデータセットとWHUデータセットの両方で2000回の反復学習を行った。ラベル付きデータの割合をkとし、バッチサイズはラベル付きデータに対して4、ラベルなしデータに対して$\_{k}^{4}$ - 1とする。さらにハイパーパラメータλを8、信頼度閾値τを0.95とした。λとτの最適範囲については4.5.2節と4.5.3節で説明することに留意されたい。SANetでは、各コンボリューションモジュールのコンボリューションカーネルの数を{8, 16, 32, 64, 128}とし、コンボリューションカーネルのサイズを3×3とする。そこで、LEVIR2000データセットと呼ばれる2000ペアをランダムに選択し、実験に使用する。訓練セットの他の画像ペアは、教師なしデータ量が半教師付きCDタスクの精度に与える影響を検証する実験に使用されることは注目に値する。本手法の有効性を証明するため、ラベル付き画像の割合として2.5%を選択した。しかし、アプリケーションでは、より良いパフォーマンスを得るために、より高いラベル付き画像の割合を選択することを提案する。したがって、より多くのラベル付き画像の影響を示すために、実験では他の3つの割合も高く設定した。最後に、LEVIR2000データセットとWHUデータセットの両方で、ラベル付きデータの割合を{2.5%、5%、10%、20%}に設定した。

------------------------------

本手法の有効性を検証するため、提案手法を最先端の完全教師ありCDおよび半教師ありCD手法と比較する：

------------------------------

FC-Conc[34]とFC-Diff[34]である：FC-ConcとFC-DiffはCDタスクのベースライン手法であり、その中でもFC-ConcとFC-Diffはより効果的であるため、本稿では比較手法として引用する。どちらもUNetとシャムネットワークの組み合わせに基づいている。FC-Concは連結を用いるのに対し、FC-Diffは差分を用いるという違いがある。SNUNet-ECAM [37]．SNUNet-ECAMはネステッドUNetとシャムネットワークを組み合わせたもので、強力な特徴抽出と意味情報融合機能を持つ。s4GAN[49]：s4GANは半教師付きセマンティックセグメンテーションに用いられる。この半教師ありアプローチは、半教師ありデータから予測されるセグメンテーションマップと真のセグメンテーションマップの差を最小化するために特徴マッチング損失を設定し、モデルの有効性を向上させるために生成器と識別器のバランスをとるために自己学習損失を使用することで、教師なしデータを使用して実装される。SemiCDNet [20]： SemiCDNetは生成的敵対ネットワークを用いた半教師付きCD手法である。ラベル付けされたデータとラベル付けされていないデータ間のセグメンテーションとエントロピーグラフの特徴分布の一貫性を高めるために、2つの識別器を使用することにより、半教師付きCDの精度を向上させる。

------------------------------

公平性を保つため、比較モデルの教師ありモデルはラベル付きデータのみを用いて学習される。

------------------------------

LEVIR2000データセットにおける提案手法とベースラインの定量結果を表2に示す。ラベル付きデータの異なるサンプリングレート設定に基づき、F1スコア、Kappa、IoUの3つの評価指標を計算し、まとめた。表2から、LEVIR2000データセットにおける我々の半教師付きモデルは、完全教師あり手法や他の半教師付き手法よりも有意に優れていることがわかる。その中でも、同じく半教師付き手法であるs4GANとsemiCDNetは、ラベル付きデータのサンプリング率が2.5%の場合、完全教師付き手法と比較して、この実験ではモデル性能があまり向上していない。これには2つの理由が考えられる。一方では、我々のサンプリング率は原論文[20]のサンプリング率よりも低く、データセットの分布を完全に学習することが難しい。一方、元の論文[20]では、変化したピクセルを含むリモートセンシング画像のみを使用している。ご存知のように、GANは識別器を扱うために生成器を訓練することを目的としている。GANに基づく半教師付き変化検出法では、識別器は、どれがラベルで、どれがモデルによって出力された予測であるかを決定する必要がある。変化していない画像を使うということは、真っ黒な画像をラベルとして使うということである。これは、生成器が黒い画像を出力することで識別器を扱うという結果につながる可能性がある。これはGANに基づく半教師付き変化検出の限界となりうる。我々の実験では、変化なしも一種のラベルであると考え、変化なし画像を用いている。このため、GANの性能は学習の後半で低下する。しかし、我々の手法は未変更画像を使用することの影響を克服することができる。さらに、学習データを選択する際に、変更画像と未変更画像を区別しない方が実用上便利であると考える。表2から分かるように、我々の手法は全ての指標において比較手法を大きく上回っている。本手法の有効性と限界については、4.5.1節で詳述する。

------------------------------

表2.LEVIR2000データセットでの実験結果。

------------------------------

この比較をより直感的に示すために、サンプリング率10％における各モデルの変化検出マップを図6に示す。

------------------------------

図6.

LEVIR2000データセットにおける、学習に10%のラベル付き画像を用いた、異なるアプローチによるCDマップの視覚的比較。(a) 画像T$\_{1}$ .(b) 画像T$\_{2}$ . (c) グラウンドトゥルース。(d) Fc-Diff. (e) Fc-Conc. (f) SNUNet-ECAM. (g) s4GAN. (h) semiCDNet. (i) semiSANet.

------------------------------

図6の最初の2行は、誤検出エラーに対する我々の半教師付き手法の効果的な解を示している。比較法では、画像中の建物の基礎を完成した建物と判断しているが、我々のsemiSANetではこのような擬似的な変化を識別している。図6の3行目と4行目は、semiSANetが検出ミスを克服できることを示している。変化している建物の色が周囲の特徴の色と類似している場合、他の手法はこれを背景と判断する、すなわち変化が生じていないと判断する。さらに、我々のsemiSANetは、このような変化する物体の検出に成功している。誤検出と検出ミスの減少は、モデルのロバスト性を向上させるために使用する強力な補強方法によるものです。図6の5行目と6行目は、semiSANetが、変化するオブジェクトのペアの輪郭をよりよく復元でき、ボイドやぼかしを生成しないことを示している。

------------------------------

WHUデータセットにおける提案手法とベースラインの定量結果を表3に示す。ラベル付きデータの異なるサンプリングレート設定に基づき、F1スコア、OA、Kappa、IoUの3つの評価指標を計算し、まとめた。表3からわかるように、サンプリング率が20%以下の場合、完全教師ありの手法の結果は、ラベル付きデータ量が増加しても増加傾向を示さない。さらに、F1スコアは常に59%以下、Kappaは54%以下、IoUは42%以下である。これは、WHUデータセットのシーン分布がLEVIRデータセットよりも複雑で、色の変化による擬似変化が多いことによる。ラベル付けされたデータが不十分な場合、モデルはデータセットの全体的な分布を学習しない。そのため、ラベル付きデータのサンプリング率が2.5%から20%の間は、モデルがうまく機能せず、ジリ貧になる。

------------------------------

表3. WHUデータセットでの実験結果。

------------------------------

半教師付き手法s4GAN、semiCDNet、semiSANetは、WHUデータセットでも全く異なるパフォーマンスを示す。ラベル付きデータのサンプリング率が20%の場合、つまり400ペアのラベル付き画像を使用した場合、semiCDNetは教師あり手法に比べてF1スコア、Kappa、IoUが約15%向上する。しかし、F1スコアは60.25%、Kappaは56.22%、IoUは43.12%と、まだあまり満足のいく結果ではない。これら2つの半教師付き手法の結果が芳しくない主な理由は2つある。1つ目は、WHUデータセットが複雑なシーン、多様な変化オブジェクト、多くの擬似変化を含む難しいものであること。第二に、我々は未変化のリモートセンシング画像ペアを一種のラベル付きデータとして使用するだけでなく、元の論文[20]よりも低いサンプリングレートを使用していることである。セクション4.4.1で分析したように、この理由はGANの劣化につながる可能性がある。我々のsemiSANetはWHUデータセットで最もよく機能し、モデル性能を大幅に改善した。これはsemiCDNetと同じ傾向を示し、すなわち、完全教師あり手法のようにジッタリングするのではなく、ラベル付き画像数が増えるとモデルの性能が向上する。semiSANetは全てのサンプリング率で良好に機能し、F1スコアは78%以上、Kappaは75%以上、IoUは64%以上である。サンプリング率が20%の場合、semiSANetはF1スコア87.86%、Kappa86.64%、IoU78.34%と優れた性能を達成することができる。我々の半教師付きアプローチによるモデル性能の向上をより直感的に見るために、サンプリング率10%における各モデルの変化検出結果を図7に示す。

------------------------------

図7. 図7.WHUデータセットにおける10%のラベル付き画像を用いたCDマップの視覚的比較。(a) 画像 T$\_{1}$ .(b) 画像 T$\_{2}$ . (c) グラウンドトゥルース。(d) Fc-Diff. (e) Fc-Conc. (f) SNUNet-ECAM. (g) s4GAN. (h) semiCDNet. (i) semiSANet.

------------------------------

図7から、比較手法には多くの誤検出があり、これらの誤検出は主にリモートセンシング画像ペアの擬似変化に起因していることがわかる。

LEVIR2000データセットと比較すると、WHUデータセットの変化オブジェクトは、より多様な形状、サイズ、色を持っている。変化した家屋の多くは、土地や建物の基礎のように、地図上で明確な立体感を持たない。これらが、モデルに対する混乱の原因である。しかし、我々の半教師付き手法は、ラベル数が少ない場合でも良い結果を得ることができる。これは、強力な補強と一貫性の正則化を用いることでモデルのロバスト性が強化され、モデルがこれらの色や形状に邪魔されにくくなるためである。

------------------------------

CDタスクの有効性に対するモデルの各コンポーネントの効果をより良く決定するために、我々は、半教師付き手法、半教師付き損失重みλ、信頼閾値τ、ラベル無しデータの量、グラフ注意モジュールがモデルに与える影響を探索するためのアブレーション実験を設計した。

------------------------------

4.5.1.半教師付き手法の効果

------------------------------

半教師付きアルゴリズムの効果を実証するため、完全教師ありアルゴリズムSANetを用いた実験と、半教師付きアルゴリズムsemiSANetを用いた実験を、各ラベル付きデータのサンプリング比率で行った。LEVIR2000での実験結果を表4に示す。

------------------------------

表4. LEVIR2000データセットでの実験結果。

------------------------------

サンプリング率が2.5%の場合、すなわち50枚のラベル付き画像のみを用いた場合、我々のモデルはF1スコアを53.99%から78.75%に、Kappaを52.38%から77.90%に、IoUを36.96%から64.94%に、いずれも24%以上向上させた。サンプリング率が5%の場合、すなわち100枚のラベル付き画像を使用した場合、我々のモデルはF1スコアを65.48%から80.95%に、Kappaを64.19%から80.19%に、IoUを48.68%から68.00%に改善し、いずれも15%以上の改善を示した。サンプリング率が10％の場合、すなわち200枚のラベル付き画像を使用した場合、半教師付き手法はF1スコアを6.6％、Kappaを6.86％、IoUを9.5％向上させる。サンプリング率が20％、すなわちラベル付き画像数が400の場合、半教師付き手法はF1スコアを5.6％、Kappaを5.84％、IoUを8.36％向上させることができる。実験結果は、我々の半教師付き手法の有効性を示しており、LEVIR2000データセットにおけるモデル性能の向上は、ラベル付きデータのみを用いた場合と比較して顕著である。さらに、ラベル付きデータの量が少ないほど、その改善は顕著である。これは、教師なしデータ変化検出のモデルの一貫性を強化するために強力な補強を用いることで、モデルの頑健性と汎化性が大幅に改善されるためである。ラベル付けされたデータが少ない場合でも、ラベル付けされていないデータを用いて、色変化干渉や形状変化干渉などの干渉に対する耐性を高めることで、モデルのロバスト性と汎化性を向上させることができ、ラベル付けされたデータが少ない場合の変化検出性能を大幅に向上させることができる。この改善をより視覚的に見るために、サンプリング率2.5%と5%における完全教師あり結果と半教師あり結果を図8に示す。図8から、ラベル付き画像が50枚しかない場合、完全教師ありの結果は非常に悪く、変化したオブジェクトを検出することはほとんど不可能であることがわかる。半教師付き手法では、さらに1950枚のラベルなし画像を追加することで、検出効果が大幅に向上し、すべての変化オブジェクトの検出に成功している。ラベル付き画像が100枚しかない場合、完全教師ありSANetはいくつかの変化オブジェクトを検出するが、検出結果は非常に粗い。さらに1900枚のラベル無し画像を用いて半教師付き法を用いると、全てのオブジェクトが検出されるだけでなく、検出結果もより正確で輪郭が明瞭になる。

------------------------------

図8. LEVIR2000データセットにおけるCDマップのSANetとsemiSANetの視覚的比較。(a) 画像 T$\_{1}$ . (b) 画像 T$\_{2}$ . (c) Ground Truth.(d) 50枚のラベル付き画像を用いたSANet。(e) 50枚のラベル付き画像によるsemiSANet。(f) 100個のラベル付き画像によるSANet。(g) 100個のラベル付き画像を持つsemiSANet。

------------------------------

WHUデータセットに対してSANetとsemiSANetの比較実験を行い、その結果を表5に示す。

半教師付きアルゴリズムを使用することで，各サンプリングレートにおいてモデルの性能が大幅に向上していることがわかる．F1スコアの改善は20%以上であり、KappaとIoUの改善はともに25%以上である。サンプリング率が20%のとき、semiSANetの結果は優れている。F1スコア、Kappa、IoUの改善率はいずれも36%以上である。

------------------------------

表5. WHUデータセットの実験結果。

------------------------------

この改善をより直感的に見るために、サンプリング率が2.5%と5%の場合の完全教師あり、半教師ありの結果を図9に示す。図9から、完全教師ありアルゴリズムのみを使用した場合、結果に多くの誤検出があることがわかる。これは、WHUデータセットでは、建物の周囲に道路など建物と色が似ている地理的オブジェクトが多数存在するためです。半教師ありアルゴリズムは、このような誤検出を非常にうまく排除することができる。これは、画像に強い補強を用いると、異なる地理的オブジェクト間の色のギャップが、時に拡大され、時に最小化されるからである。一貫性の正則化により、オブジェクトの色が似ている場合と大きく異なる場合の両方で、モデルは同じ変化検出結果を生成する。最終的には、モデルは色のギャップよりも、オブジェクトの意味的なギャップに焦点を当てるようになる。------------------------------図9. WHUデータセットに対するCDマップのSANet法とsemiSANet法の視覚的比較。(a) 画像T$\_{1}$ . (b) 画像T$\_{2}$ . (c) グラウンドトゥルース。(d) 50枚のラベル付き画像を用いたSANet。(e) 50個のラベル付き画像を持つsemiSANet。(f) 100個のラベル付き画像を持つSANet。(g) 100枚のラベル付き画像によるsemiSANet．

------------------------------

4.5.2. 教師なし損失重みλの効果

------------------------------

λの大きさはラベルなし画像がどの程度モデル効果に影響するかを決定する。semiSANetでλの最適な値の範囲を見つけるために、信頼閾値τ=0.95、λ={ 0, 0.05, 0.5, 1, 2, 4, 8, 12, 16, 32, 64 }を設定し、LEVIR2000データセットとWHUデータセットで実験を行った。実験結果を図10に示す。図10aから、LEVIR2000データセットにおいてλが増加すると、変化検出の効果が急激に上昇し、その後緩やかに減少することがわかる。我々の実験では、λ∈[ 0, 2 ]が立ち上がりの速い段階であり、この段階でλの値を大きくすることで、モデル効果を大幅に改善することができる。λ ∈ [ 4, 16 ]のとき、モデル効果は満足のいくものであり、F1スコアは約85%で安定し、λ = 8のとき最高値85.32%に達する。図10bから、WHUデータセットでも同じ傾向が見られる。λ∈[ 0, 4 ]が急上昇の段階である。その後、F1スコアはゆっくりと増加し、λ = 16で84.19%の最大値を得る。これは、教師なし損失L uが教師あり損失L sよりはるかに小さいため、λの値が大きくても、教師あり損失L sのモデルへの誘導効果をL uが破壊するほどではないからであろう。しかし、λの値が小さすぎると、モデル効果への寄与が弱くなる。また、区間[ 4, 20 ]内の値を取ることが最良の結果を与える可能性がある。

------------------------------

図10. (a)LEVIR2000データセット。(b) WHUデータセット。

------------------------------

4.5.3. 信頼閾値τの効果

------------------------------

τの値は、擬似ラベルの何画素が損失を計算するために利用されるかを決定し、モデル効果に重要な影響を与える。semiSANet上でτの最適な値域を見つけるために、λ = 8とτ = { 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99 }を設定し、LEVIR2000データセットとWHUデータセットで実験を行った。実験結果を図11に示す。図11aから、LEVIR2000データセットではモデルの性能はあまり変わらない。F1スコアは、τ∈[ 0.5, 0.95 ]のとき84.94%から85.39%の間で変動し、τ=0.99のときだけF1スコアが84.39%に低下した。これは、表4に示すように、このモデルがLEVIR2000データセットで事前によく訓練されているため、信頼度閾値τが小さな値であっても、擬似ラベルをうまく取得できるためと考えられる。しかし、τが大きすぎると、モデルの学習に有用な擬似ラベルが代わりにフィルタリングされてしまい、モデルの性能が低下してしまう。

------------------------------

図11. (a)LEVIR2000データセット。(b) WHUデータセット。

------------------------------

図11bから、WHUデータセットではτの増加とともにF1スコアが増加し、その後減少する傾向を示していることがわかる。

これは、WHUデータセットに対する事前学習モデルの有効性が低いため、信頼度閾値τでフィルタリングしなければならない擬似ラベルに多くの偽ラベルが存在するためである。 τ∈[ 0.5, 0.8 ]のとき、τが大きくなるにつれてF1スコアは79.36%から83.35%に徐々に増加し、間違った擬似ラベルが徐々にフィルタリングされることがわかる。τ∈[ 0.8, 0.95 ]のとき、モデルはより安定した段階にあり、F1スコアは83.19%から83.53%の間でジッタリングしている。しかし、τ = 0.99のとき、F1スコアは82.8%に低下する。この実験から、信頼性の閾値τは、事前学習効果の良し悪しを参考にして決めるべきであることがわかる。一般的に、τ∈[ 0.8, 0.95 ]のとき、モデルはより良い結果を達成する。

------------------------------

4.5.4. ラベルなしデータ量の効果

------------------------------

ラベルなしデータ量の増加が半教師付き性能の向上に寄与するかどうかを検証するため、LEVIRデータセットとWHUデータセットで実験を行った。前回はLEVIRデータセットから2000組の画像のみを学習セットとして使用したが、今回の実験では、より大量の教師なしデータの効果を検証するため、LEVIRデータセットから最大4000枚の画像を学習セットとして使用した。LEVIRデータセットの学習セット数を｛200, 400, 600, 1000, 2000, 3000, 4000｝とし、ラベル付きデータ量を200とした。WHUデータセットの学習セット数は｛200, 400, 600, 1000, 1500, 2000｝とし、ラベル付きデータ量は同じく200である。実験結果を図12に示す。全体として、教師なしデータ量が増加するにつれて、モデルの変化検出効果は増加する傾向にある。図12aのLEVIRの結果から、この上昇過程にもジッタがあり、教師なしデータ量の増加すべてがモデル性能を向上させているわけではない。さらに、学習セット数が400しかない場合、半教師付き手法はすでにモデル性能が大きく向上している。これは、ラベル付きデータとして200組の画像がすでにLEVIRデータセットの全体的な分布をよく表しており、より高い変化検出精度を得るためにモデルのロバスト性と汎化性を向上させるには、さらに200組のラベルなし画像を追加するだけでよいためと考えられる。図12bのWHUデータセットの結果も同じ傾向を示している。モデル性能は、学習セット数が400のときに大きく改善し、その後F1スコアは徐々に改善するが、データ量の制限により、データ量が2000を超えた後のWHUデータセットの状況を観察することはできない。

------------------------------

図12. データ量の効果。(a) LEVIRデータセット。(b) WHUデータセット。

------------------------------

特徴相関を計算し、長距離情報を融合する能力をモデルに持たせるため、SANetのエンコーダの最後にGATモジュールを追加した。GATモジュールの有用性を検証するため、LEVIR2000データセットとWHUデータセットのラベル付き画像とGoogleデータセットのサンプリング率を10％にしたアブレーション実験を行い、その結果を図13に示す。この結果は、GATモジュールがモデルの性能を実際に改善できることを示している。サンプリング率10％のLEVIR2000データセットにGATを追加すると、F1スコア、Kappa、IoUがそれぞれ0.88％、0.9％、1.33％向上した。サンプリング率10%のWHUデータセットにGATを追加すると、F1スコア、Kappa、IoUがそれぞれ0.96%、1.13%、1.4%向上した。

------------------------------

提案する半教師付き変化検出手法は、ラベル付けされた画像が不十分な場合でも、効果的な変化検出結果を得ることができる。

まず、提案手法はシンプルである。FixMatch[47]に触発され、本研究では、強補強と信頼閾値フィルタのみを用いることで、半教師付き変化検出を実現する。平均教師(mean-teacher)[19]やGAN[20,49]を用いた半教師付きリモートセンシング画像処理手法と比較して、本手法は実装が容易である。さらに、広範な実験により、本作業が従来の作業よりも効果的であることが証明された。少量のラベル付きデータで、本作品は最新の教師あり手法や半教師あり手法よりも優れた変化検出結果を達成する。しかし、本作品には主に2つの限界がある。第一に、ラベルなしデータの割合は GPU メモリによって制限される。セクション4.3で述べたように、ラベル付きデータのバッチサイズは少なくとも1でなければならず、ラベルなしデータのバッチサイズは$\_{k}^{1}$ - 1となる。ラベルなしデータの割合が増えると、入力データのバッチサイズが大きくなり、より多くのGPUメモリを必要とする。第二に、強力な補強は低・中解像度リモートセンシング画像の有用な情報を破壊する可能性がある。本研究は高解像度リモートセンシング画像の変化検出を目的としており、強力な補強は本研究において最も重要な技術である。しかし、より粗く、乱れやすい低・中解像度画像に関しては、強 補正はその意味情報を破壊する可能性がある。

------------------------------

本論文では、シャムUNetと一貫性正則化に基づく半教師付きリモートセンシング画像変化検出法（semiSANet）を提案する。semiSANetでは、抽出された深い意味特徴間の情報を融合するために、グラフ注目メカニズムを利用し、長距離特徴情報の融合を可能にする。非ラベルデータを利用するために、我々は一貫性正則化の原理を用いる、すなわち、歪曲前と歪曲後の画像に対してモデルが同じ変化検出結果を出すようにする。具体的には、ラベル付けされていないデータは、まず事前に訓練されたモデルと信頼度閾値フィルタに入力され、高い信頼度を持つ擬似ラベルを得る。一方、強い補強を用いて歪んだ画像を取得し、歪んだ画像のCD結果をモデルによって取得する。CD結果と擬似ラベルの差を小さくすることで、強オーグメンテーションによる干渉を緩和し、モデルのロバスト性を向上させる。2つの高解像度リモートセンシング画像データセットを用いた実験結果は、我々の手法が、わずかな量のラベル付きデータで性能を大幅に向上させることができ、最先端の手法を凌駕することを示している。ラベル付き画像が2.5%しかない場合、我々の手法は最先端の手法と比較してIoUを25%以上向上させることができる。今後は、本研究の限界を解決することに注力する。具体的には、第一に、GPUメモリの要件を制御するために、ラベル付きデータとラベルなしデータのバッチサイズは、ラベル付きデータの割合によって変化するのではなく、固定であるべきである。第二に、強い増強範囲が異なる解像度のリモートセンシング画像の変化検出にどのような影響を与えるかを研究する。

------------------------------

著者貢献 概念化、C.S.およびH.C.、方法論、C.S.およびH.C.、ソフトウェア、C.S.、検証、C.S.、C.D.およびH.C.、データ管理、H.C.、執筆・原案作成、C.S.、執筆・査読・編集、C.D.およびJ.W.、監修、H.C.、プロジェクト管理、H.C.。

------------------------------

Data Availability Statement： 該当なし。

------------------------------

利益相反： 著者らは利益相反がないことを宣言する。

------------------------------