

補足資料 - 正確なガイダンスなしに学習する：低解像度の履歴ラベルから大規模高解像度土地被覆マップを更新する

この補足では、提案するフレームワークとデータセットの構成について詳細に説明する。さらに実験結果も示す。これら3つの部分を順次示す。

A. パラフォーマーの詳細

提案するParaformerでは、ロバストな特徴抽出器。パラレルはダウンサンプリングフリーのCNNブランチとTransformerブランチをハイブリッドする。CNNとTransformer分岐の構造をより明確に示すために、[図S1](#)と[図S2](#)にCNNとTransformer分岐の基本ユニットを示す。

このセクションでは、CNN分岐の基本ユニットを詳細に説明することに重点を置く。[図S1](#)に示す解像度保存（RP）ブロックは、我々の先行研究で最初に提案されたものである：L2HNet¹。ここでは、 $\mathbf{I}^{(b)}$ 、 $\mathbf{M}^{(b)}$ 、 $\mathbf{F}^{(b)}$ を使って、 b 番目のブロックの入力、中間、融合特徴マップを表す。具体的には、最初のブロックの入力特徴マップは、4つの入力層を持つ3×3の畳み込み入力層によって生成される。

チャンネル（すなわち、画像のR-G-B-NIRバンド）と C_I 出力チャンネルを置く。 $(1)N \in C_I \times H_I \times W_I \times N$ はバッチサイズを表し、 C_I H_I W_I はそれぞれマップのチャンネル、高さ、幅を表す。演算記号については、1ストライド（ $n \times n$ ）の畳み込み層を表し、 C_I 入力チャンネルと C_I

出力チャンネルを $M^{n \times n}$ （ $n = 3, 5$ の場合はパディングあり）とする。さらに、整流線形ユニット（ReLU）関数を持つバツ

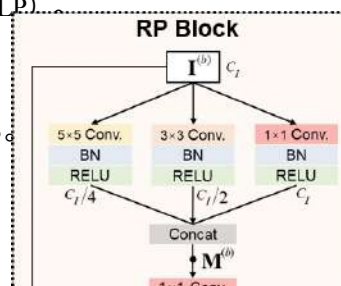
$\mathbf{M}^{(b)}$ の次元を $C_I (1 + 1/2 + 1/4)$ から C_I に変更し、ブロックを軽量化した。さらに、浅い特徴を維持し、残差学習を有効にするために、 $\mathbf{I}^{(b)}$ から $\mathbf{F}^{(b)}$ へのショートカット接続が採用されている。その結果、最終的な $\mathbf{F}^{(b)}$ は次のように記述できる：

$$\mathbf{F}^{(b)} = \text{bn}(\mathbf{M}^{(b)} * W_{C_I(1+1/2+1/4), C_I}^{1 \times 1}) + \mathbf{I}^{(b)}. \quad (\text{S2})$$

式(S1)-(S2)より、 $\mathbf{F}^{(b)}$ は $\mathbf{I}^{(b)}$ と同じサイズ、チャンネル、解像度を持つマルチスケールフュージョン特徴マップである。この構造に基づいて、RPブロックは、オーバーダウンサンプリングによる特徴解像度の低下を適切に防ぐために、マルチスケールフュージョン属性と残差学習能力を同期的に結合する。さらに、複数のRPブロックの特徴フュージョンの後、ソフトマックス関数と1×1コンボリュショナルレイヤー $W^{1 \times 1}$ で構成される分類器によって、予測値と対応するCPマップが生成される。

はバックボーン全体で維持されるバーであり、 L は土地被覆カテゴリの数によって決定される出力チャンネルである。さらに、Transformer分岐の基本ユニットを [図S2](#)に示す。

層知覚（MLP）



チ正規化層は、単に $bn()$ と表記される。

* は畳み込み演算子を表す。これに基づいて、 $\mathbf{I}^{(b)}$ から $\mathbf{M}^{(b)}$ へのマルチスケール特徴融合プロセスは、次のように記述できる：

$$\mathbf{M}^{(b)} = \text{concat} \left(\begin{array}{l} bn(\mathbf{I}^{(b)} * W_{C_1}^{1 \times 1}), \\ bn(\mathbf{I}^{(b)} * W_{C_3}^{3 \times 3}), \\ bn(\mathbf{I}^{(b)} * W_{C_5}^{5 \times 5}) \end{array} \right). \quad (S1)$$

式(S1)に示されるように、マルチカーネルのカーネル数 \sqrt{b} は、1,000,000,000,000,000である。スケールの畳み込み層を $\omega = \{ \frac{1}{2^{(1-n)}} \}_{n=1,3,5}$ とすると、これはカーネルサイズに反比例する。

その後マルチスケールレイヤーを連結するために、1畳み込みレイヤーを採用する。

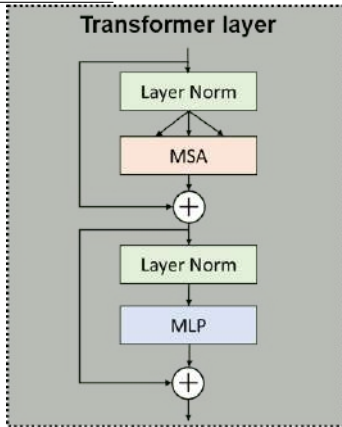
図S1.RPブロックの説明図。入力マップ $\mathbf{I}^{(b)}$ は、 1×1 、 3×3 、 5×5 のサイズを持つ3つの畳み込み層によってサンプリングされ、各層の畳み込みカーネルは、オーバーダウンサンプリングによる特徴分解能の低下を防ぐために ω の割合に設定される。

¹<https://doi.org/10.1016/j.isprsjprs.2022.08.008>

マイクロソフト社⁽²⁾によってグループ化され、1メートルの解像度の画像と、30メートルの解像度の土地被覆プロダクトが学習データとして含まれており、評価のために1メートルの解像度の地上基準も含まれている。図S3は、チェサピーク湾データセットの位置、デジタル標高モデル（DEM）、タイル数、データサンプルを示している。より詳細には、データソースを以下のように示す：

1. HRリモートセンシング画像は、1mの解像度を持つ。
米国農務省の全米農業画像プログラム（NAIP）。画像は赤、緑、青、近赤外線（NIR）の4つのバンドを含む。
2. 30メートルの解像度を持つ過去の大まかな土地被覆プロダクトは、National.

²<https://lila.science/datasets/chesapeake-landcover>



図S2. Transformerレイヤーの説明図。このレイヤーはレイヤー正規化(Layer Norm)、多頭自己注意(MSA)、多層知覚(MLP)を含む。

B. 調査地域と使用データの詳細

この節では、2つの大規模データセットの詳細を示す。図S3と図S4は、チェサピーク湾とポーランドのデータセットの位置、カバー率、データサンプルを示す。表S1と表S2は、LRラベルとHRグラントトゥールズ間の土地被覆クラス統一関係を示す。

チェサピーク湾のデータセットチェサピーク湾は、サスケハナ川の北の源流から大西洋の出口までの全長約320kmのアメリカ最大の河口である。チェサピーク湾の流域は、周辺の流域の約16万 km^2 をカバーしている。流域には、ニューヨーク州、ペンシルベニア州、デラウェア州、メリーランド州、バージニア州、ウェストバージニア州のアメリカの6つの州が含まれる。チェサピーク湾流域には、豊富な生態系と多様な植物相を持つ様々な地形があり、大規模な高解像度（HR）土地被覆マップ作成に課題をもたらしている。チェサピーク湾のデータセットは、

米国地質調査所（USGS）の土地被覆データベース。NLCDデータは16の土地被覆タイプを含み、提案するParaformerフレームワークの学習過程でラベルとして利用される。

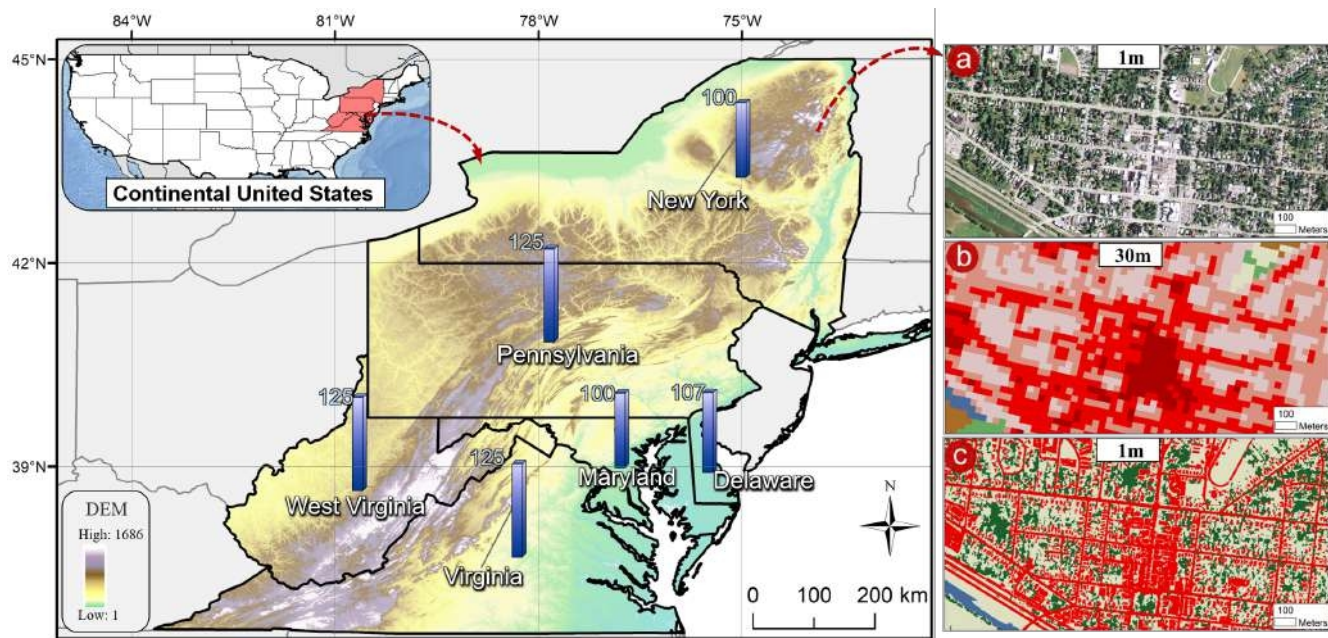
3. 1m 解像度の HR 地上参照データは、Chesapeake Bay Conservancy Land Cover (CCLC) プロジェクトから入手した。CCLC データは、1mのNAIP画像と6種類の土地被覆を含むLi-DARデータに基づいて解釈された。本データベースでは、CCLCデータは定量的・定性的評価のための基準としてのみ使用され、フレームワークのトレーニングや最適化プロセスには関与していない。

ポーランドのデータセットポーランド共和国は、中央ヨーロッパ平原を縦断し、北はバルト海から南はスデーテン山脈とカルパティア山脈まで広がっている。地形的には、平坦で長い海と丘陵の多い山岳地帯があり、ポーランドの景観は多様な地形、河川システム、生態系によって特徴づけられる。ポーランドのデータセットには、Po-morskie, Łódzkie, Lubuskie, Dolnoslaskieなどの14の州が含まれている。ポーランドのデータセットには、0.25メートルの解像度の画像、3種類の10メートルの解像度の土地被覆プロダクト、および30メートルの解像度の土地被覆プロダクトが含まれており、さまざまな組み合わせで学習データのペアを構成する。図S4は、ポーランドのデータセットの位置、DEM、タイル数、データサンプルを示したものである。より詳細には、データソースを以下のように示す：

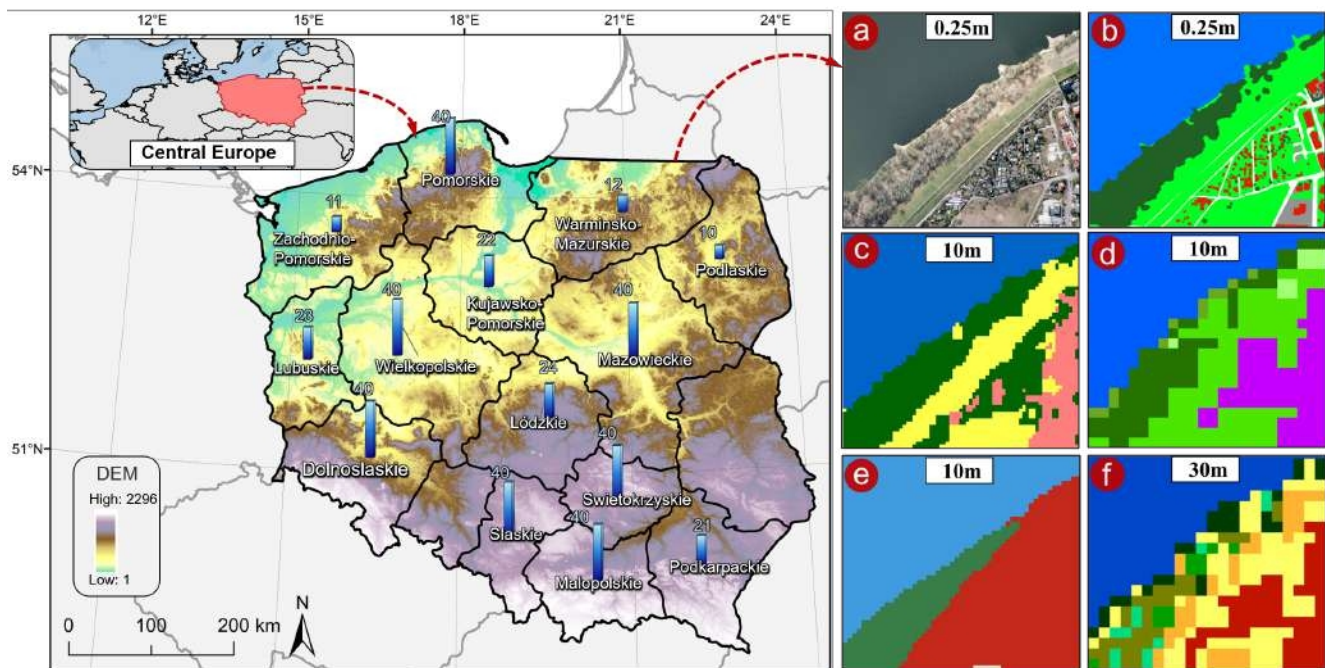
1. 0.25mと0.5mの解像度のHRリモートセンシング画像は、Land-Cover.aiデータセットから

収集されたもので、画像ソースはLand Parcel Identification System (LPIS)で使用されている公共測地系リソースである。画像には赤、緑、青の3つのバンドが含まれている。

2. 10m 分解能のラフヒストリカルラベルデータは、(1) 清華大学が提供する FROM GLC10、(2) 欧州宇宙機関（ESA）が提供する ESA WorldCover v100、(3) ESRI Inc.と IO Inc.が提供する ESRI 10m 全球土地被覆（略称 ESRI GLC10）の3種類の全球土地被覆プロダクトから収集した。30m解像度のラベル付きデータは、中国科学院（CAS）が提供する30m全球土地被覆プロダクトGLC FCS30から収集した。
3. HR 地上参照は、東京大学が提供する OpenEarthMapデータセットから取得した。地上参照は、0.25mと0.5mの解像度のLPIS画像に基づいて解釈され、5つの土地被覆タイプを含んでいる。



図S3.チェサピーク湾のデータセットはアメリカの6つの州をカバーし、(a) NAIPの1m画像、(b) 30mNLCDラベル、(c) 1mグランドトゥルースのデータソースを含む。青い列はタイル数を示す。



図S4.(a)0.25m画像、(b)0.25m地上絵、(c) 10m FROM GLC10、(d)10m ESA GLC10、(e)10m ESRI GLC10、(f)30m GLC FCS30。青い列はタイル数を示す。

C. 補足実験結果

Paraformerの性能を総合的に実証するために、以下のような補足的な実験結果を順次示す：

チェサピーク湾データセットの視覚的結果： 図

S5-S7は、Paraformerと4つの典型的な手法との大規模な比較と小規模な比較を示している。これらの視覚的な結果から、ParaformerはHR画像から正確なHR土地被覆マップを更新することができる。TransUNetは

名称	NLCD	CCLC	対象クラス
所属USGS	(米国)	チェサピーク・コンサーバ	
ンシー (米国)			
分解能	30メートル	1メートル	
クラス	開発されたオープンスペース	道路	
	開発途上低位	不毛の	ビルトアップ
	開発途上中位	地	
	開発途上高位		
	落葉樹林		
	照葉樹林	混	樹冠
	交林	木質湿地	
	帯		
	やせた土地	低木	
	／低木	草地 牧草	低植生
	地／ハー	耕作作物	
	草本性湿地		
	オープンウォーター	水	水注
: USGS=米国地質調査所;			

表S1.LRラベル（NLCD）とHR地上真理との間の土地被覆クラス統一関係。最初の列はLRラベルの凡例を示す。最後の列は、精度評価の対象となるクラスと、視覚結果に示されたその色を示す。

名称	GLC10より	ESRI_GLC10	ESA_GLC10	GLC_FCS30	オープンアースマップ	対象クラス
所属	THU、中国	米国Esri&IO	ESA、ヨーロッパ	CAS、中国	日本、東京	
決議	10メートル	10メートル	10メートル	10メートル	0.25/0.5メートル	樹冠
クラス	森林	樹木	樹木	落葉広葉樹林		
				常緑針葉樹林		
				混交林		
	シュラブランド	スクラブ／	シュラブランド	オーチャード	R ゲルランド	低植生
	草地 農	低木	草地 農	まばらな低木林		
	地	草地作物		草原		
				草本カバー		
				灌漑農地	A 農地	
				雨水耕作地		
					ビルディング	
	不浸透面積	建築面積	ビルトアップ		不浸透 広告	溜まっ
	水域	水	オープンウォ	水域	開発された空間	た水
注: THU=清華大学、ESRI=ESRI Inc.、IO=IO Inc.、ESA=欧州宇宙機関、CAS=中国科学院、UTokyo=東京大学。						

表S2.4種類のLRラベルとHR地上真理との間の土地被覆クラス統一関係。1-4列はLRラベルの凡例を示す。最後の列は、精度評価の対象となるクラスと、視覚結果に示されたその色を示す。

都市パターンを過小評価するが、既成市街地を過小評価する。UNetは典型的なCNNベースのエンコーダー・デコーダーフレームワークであり、LRラベルと一致する大まかな結果を示す。L2HNetは、LRラベルからHR土地被覆の結果を更新する最先端の手法であ

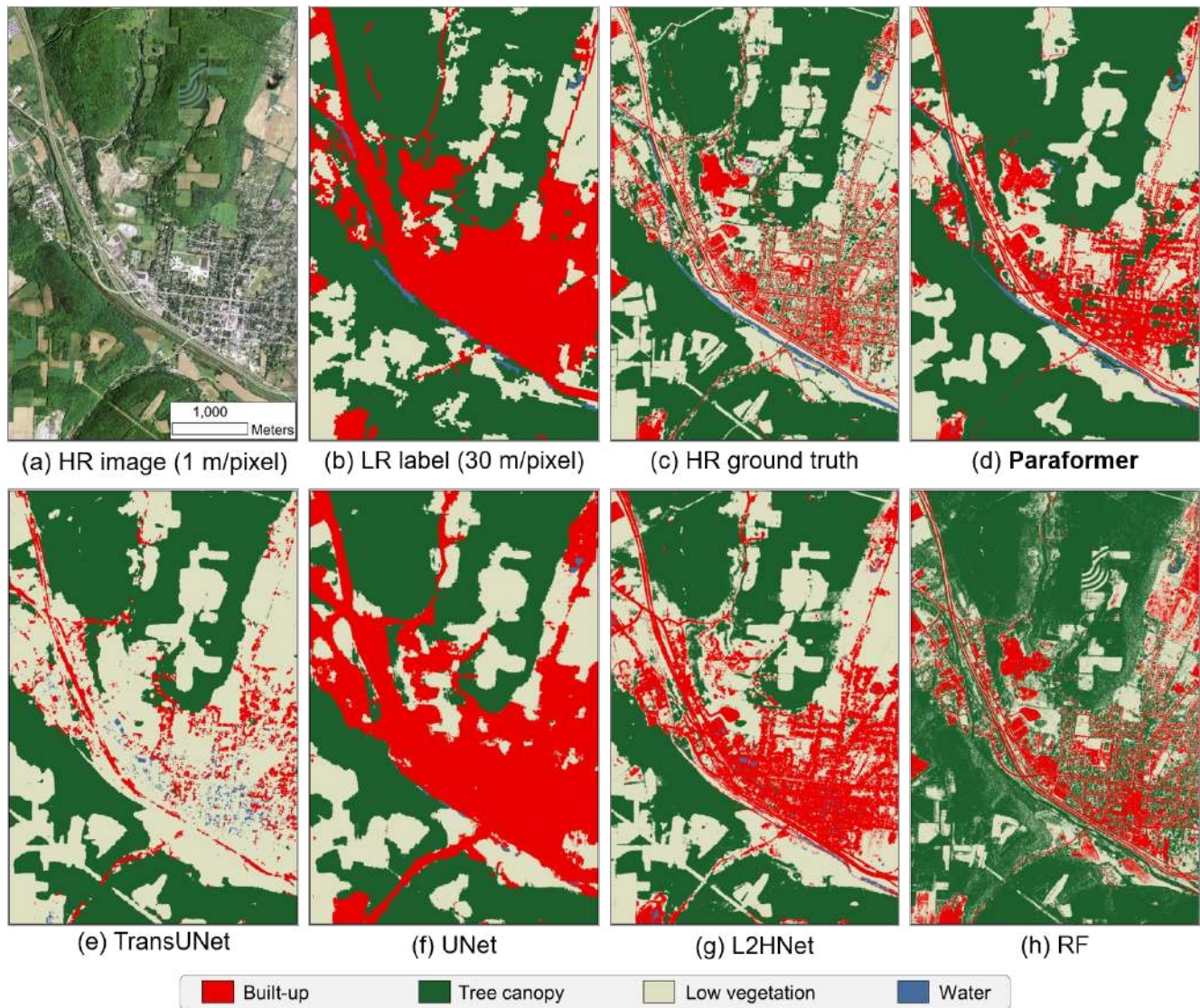
り、土地オブジェクトの正確なエッジを示すが、結果にはまだ誤った断片がある。RFは画素間学習法であり、最も細かいエッジを持つが、文脈情報の学習に欠けるため、全体として不十分な結果となる（水と低植生を過小評価する）。

ポーランドデータセットの視覚的結果：図S8-S11は

Paraformerと、異なるLR土地被覆ラベルで学習させた他の3つの代表的な手法との視覚的比較を示す。視覚的な結果から、Paraformerは、LRの土地被覆ラベルから明確な土地被覆パターンを絞り込むことができる。

異なるタイプのLR土地被覆ラベル。実証パッチに含まれるいくつかのクラスが保持されていなくても、Paraformerは局所的および大域的なコンテキストを共同で捕捉し、HR画像と一致するHR結果を生成することができる。

さらなる議論このパートでは、中国における損失変動と補足的な大規模実験の詳細を示す。図S12は、 \mathcal{L}_{cc} と \mathcal{L}_{mce} のフレームワーク学習時の損失関数である。この2つの学習損失は、チェサピーク湾データセットの6つの状態において、安定して減少している。このことは、不正確なLRラベルからの学習において、擬似ラベル支援学習（PLAT）モジュールが有効であることを示している。Paraformerの適用可能性をさらに議論するために、我々は中国武漢市全域で大規模な実験を行った。

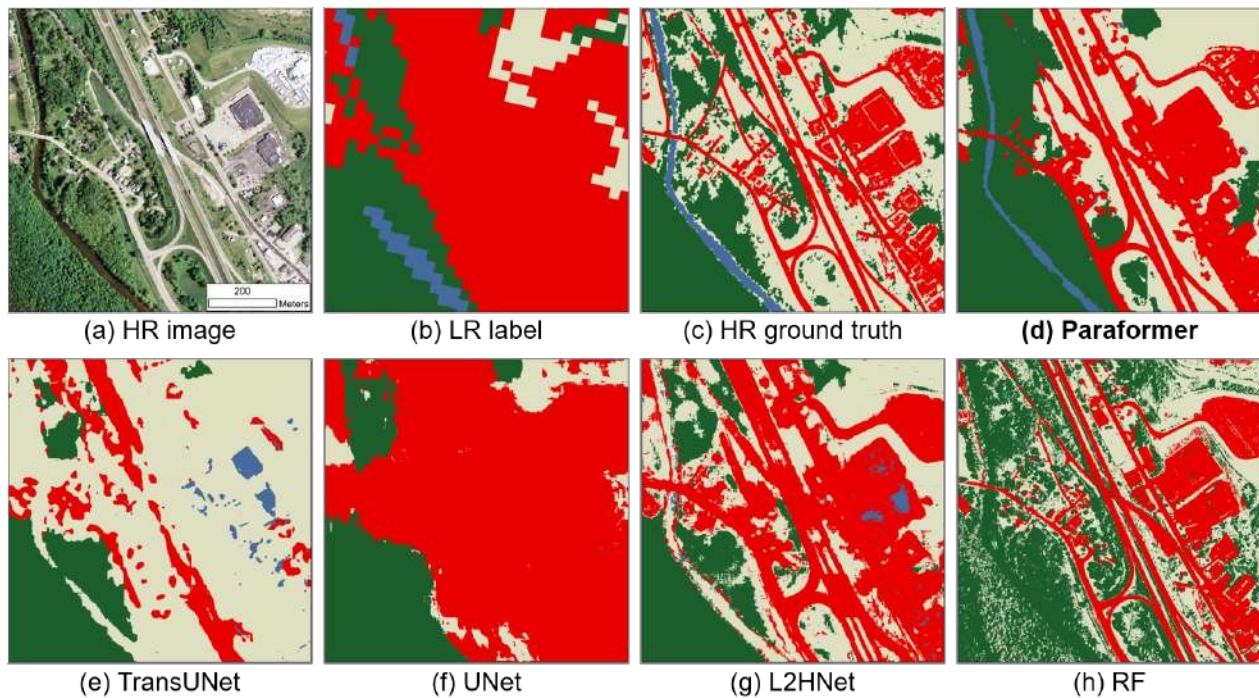


図S5. つの統一クラスを持つチェサピーク湾データセットにおける、トレーニングデータのデモと、Paraformerと他の代表的な手法の視覚的比較。(a) HR画像。(b) LRラベル。(c) HR グラウンドトゥルース。(d) Paraformerによる土地被覆マッピング結果。(e-h) 4つの代表的な手法による土地被覆マッピング結果。

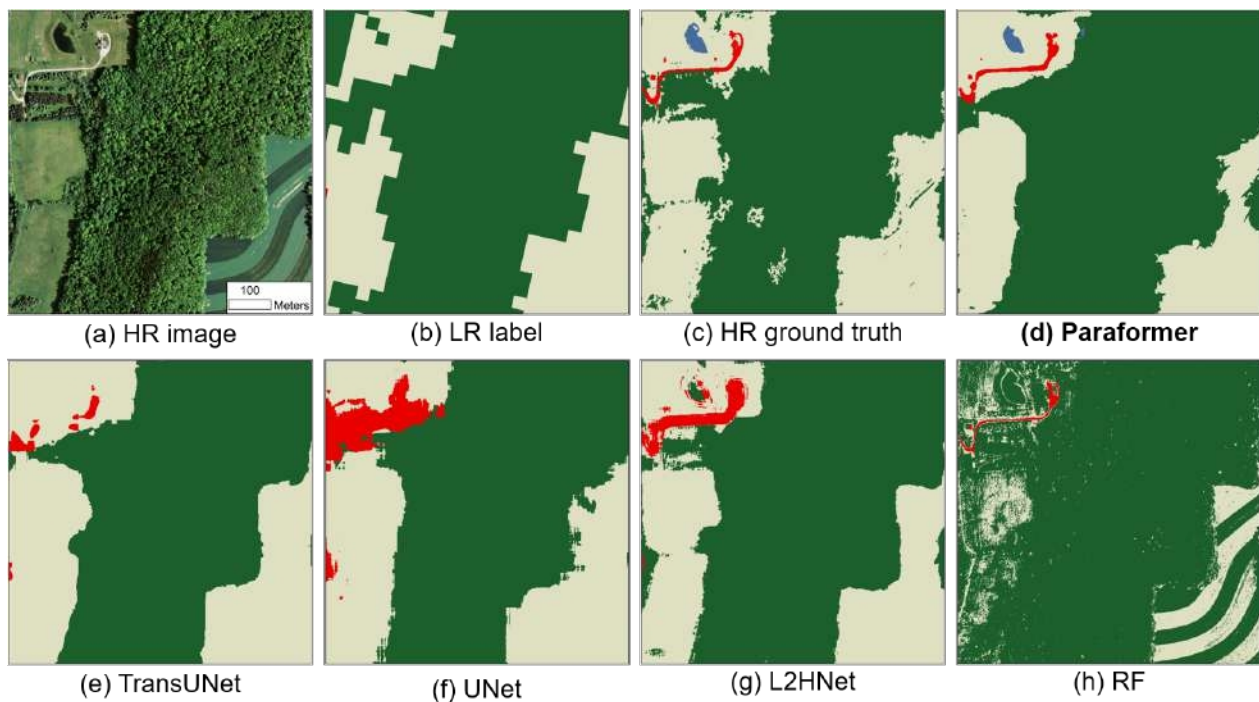
SinoLC-1³（すなわち、中国初の1-m土地被覆地図）に関する我々の以前の研究に基づき、3つの10-m土地被覆プロダクト（ESA GLC10、Esri GLC10、FROM GLC10）の交差結果を、1-m Google Earth画像のLRトレーニング・ラベルとみなす。図S13(a)に示すように、1mGoogle Earth画像は、土地の詳細が明瞭である。図S13 (b-d) は、3種類の10-m土地被覆プロダクトを示している。図S13 (e)に示すオリジナルの1-m SinoLC-1と比較すると、Paraformerは図S13 (f)に示すように、より正確な都市パターンを精製することができる。武漢市全体では、SinoLC-1の総合精度（OA

）は72.40%と報告されている。提案するParaformerの更新結果は74.98%に達し、2.58%向上している。

³<https://doi.org/10.5194/essd-15-4749-2023>



図S6.訓練データのサンプルAと、4つの統一クラスを持つチェサピーク湾データセットにおけるParaformerと他の代表的な手法の視覚的比較。(a) HR画像。(b) LRラベル。(c) HR グラウンドトゥルース。(d) Paraformerによる土地被覆マッピング結果。(e-h) 4つの代表的な手法による土地被覆マッピング結果。



図S7.訓練データのサンプルBと、4つの統一クラスを持つチェサピーク湾データセットにおけるParaformerと他の代表的な手法の視覚的比較。(a) HR画像。(b) LRラベル。(c) HR グラウンドトゥルース。(d) Paraformerによる土地被覆マッピング結果。(e-h) 4つの代表的な手法による土地被覆マッピング結果。

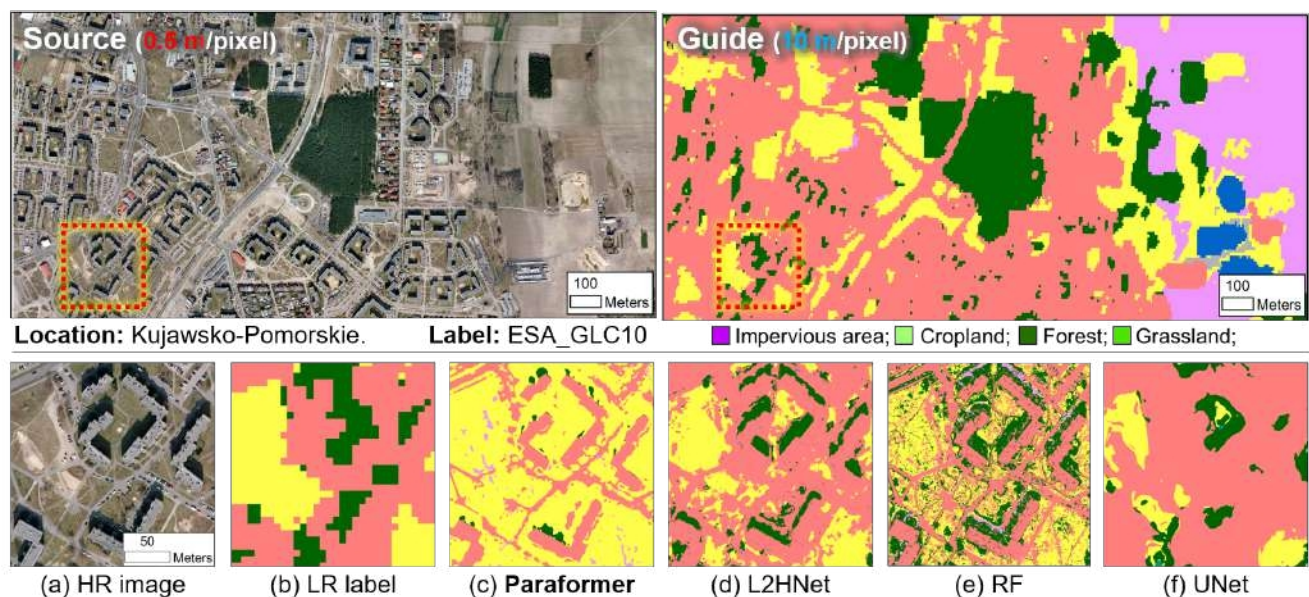


図8.10mのESA GLC10トレーニングラベルを用いたポーランドデータセットの視覚的結果。(a) 0.5mの画像、(b) E S A GLC10からサンプリングした10mのラベル。(c) Paraformerの結果。(d) L2HNetの結果。(e) RFの結果。(f) UNetの結果。

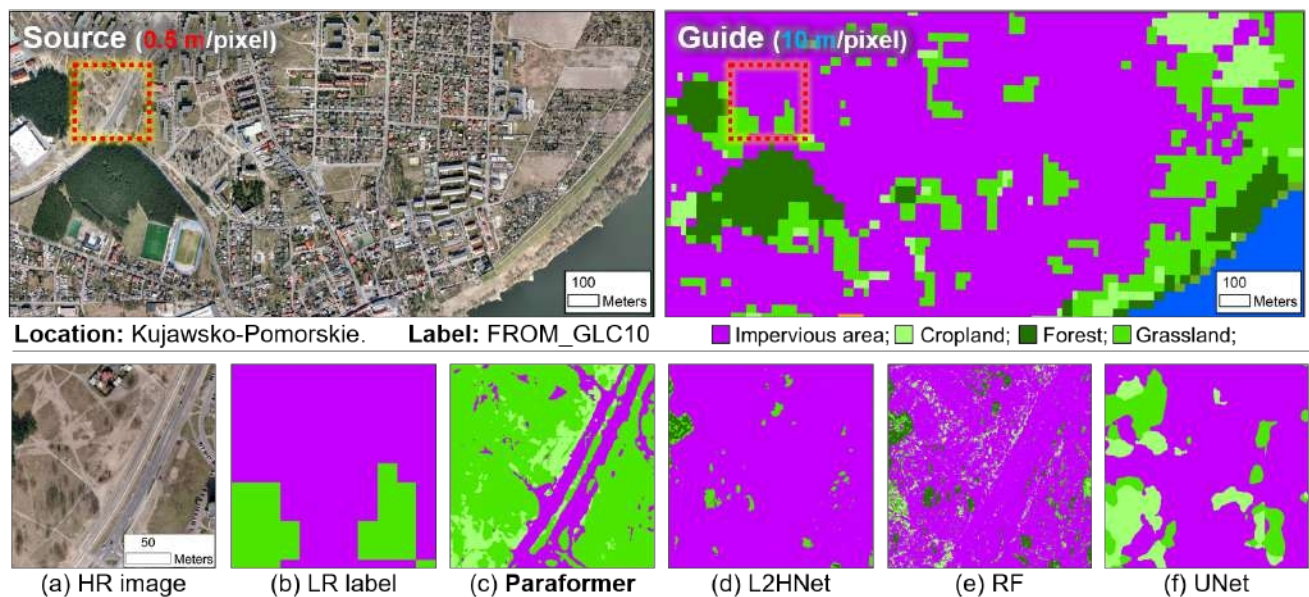
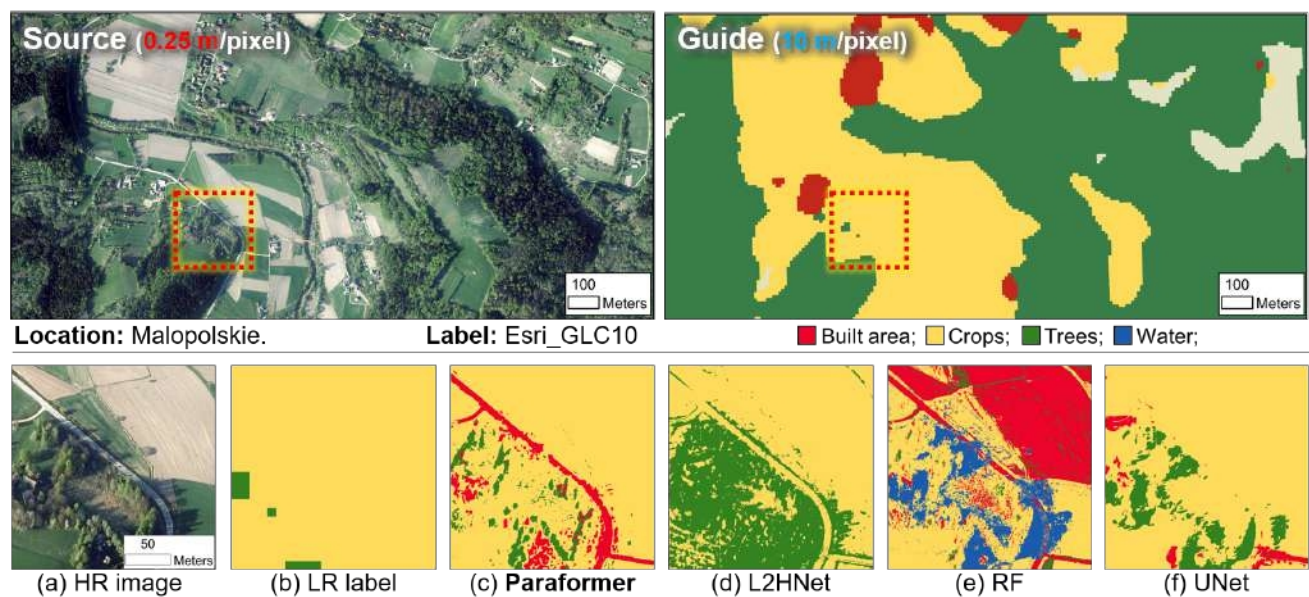
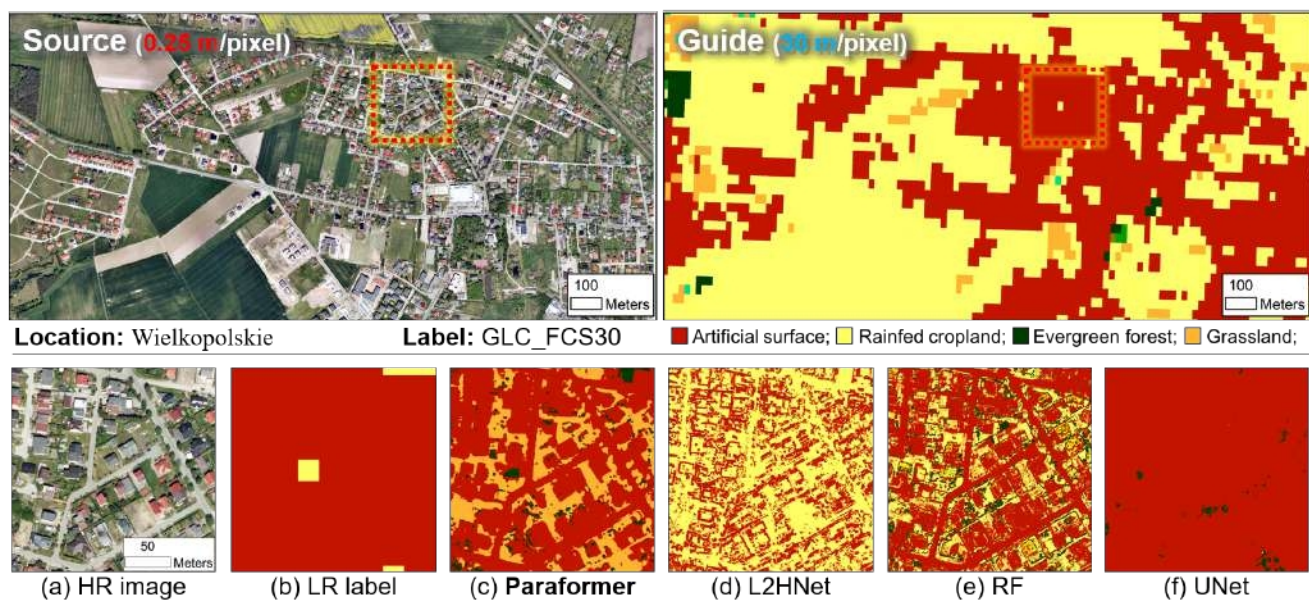


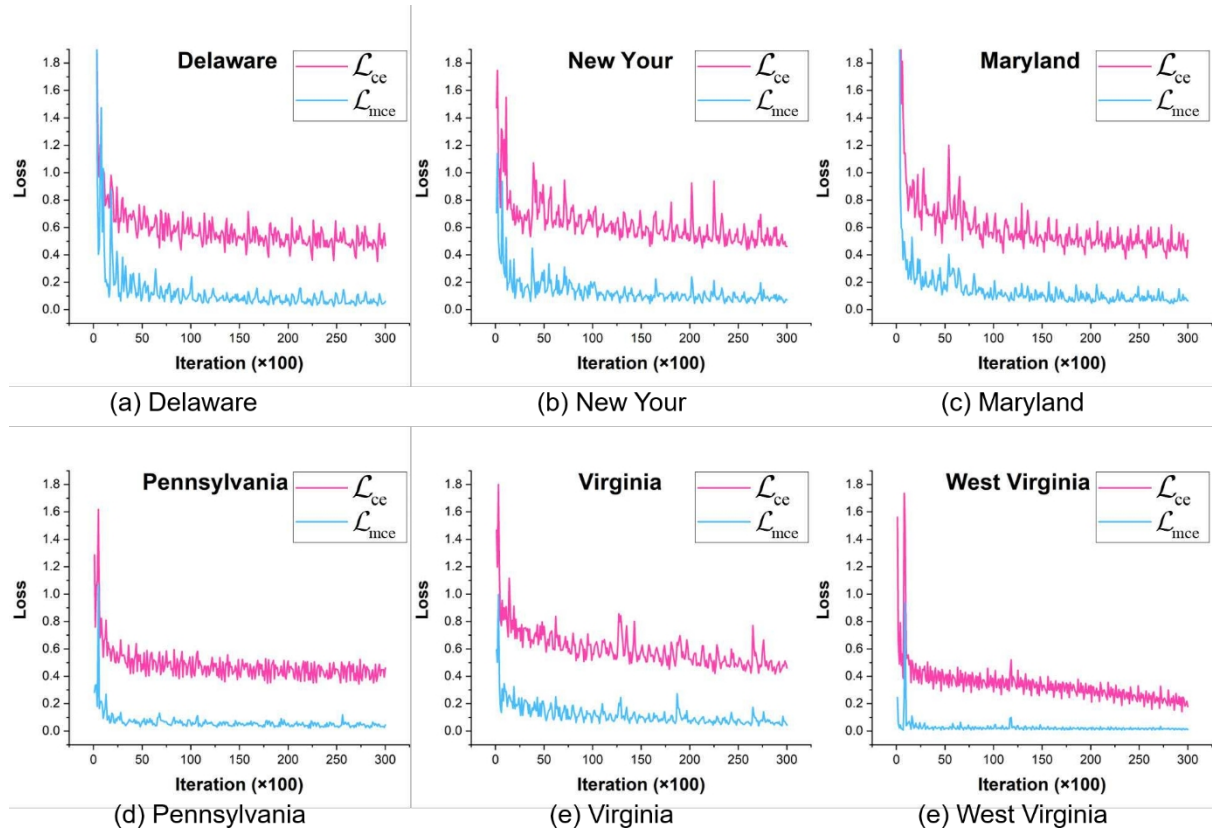
図9.10mのFROM GLC10トレーニングラベルを用いたポーランドデータセットの視覚的結果。(a) 0.5mの画像、(b) FROM GLC10からサンプリングした10mのラベル。(c) Paraformerの結果。(d) L2HNetの結果。(e) RFの結果。(f) UNetの結果。



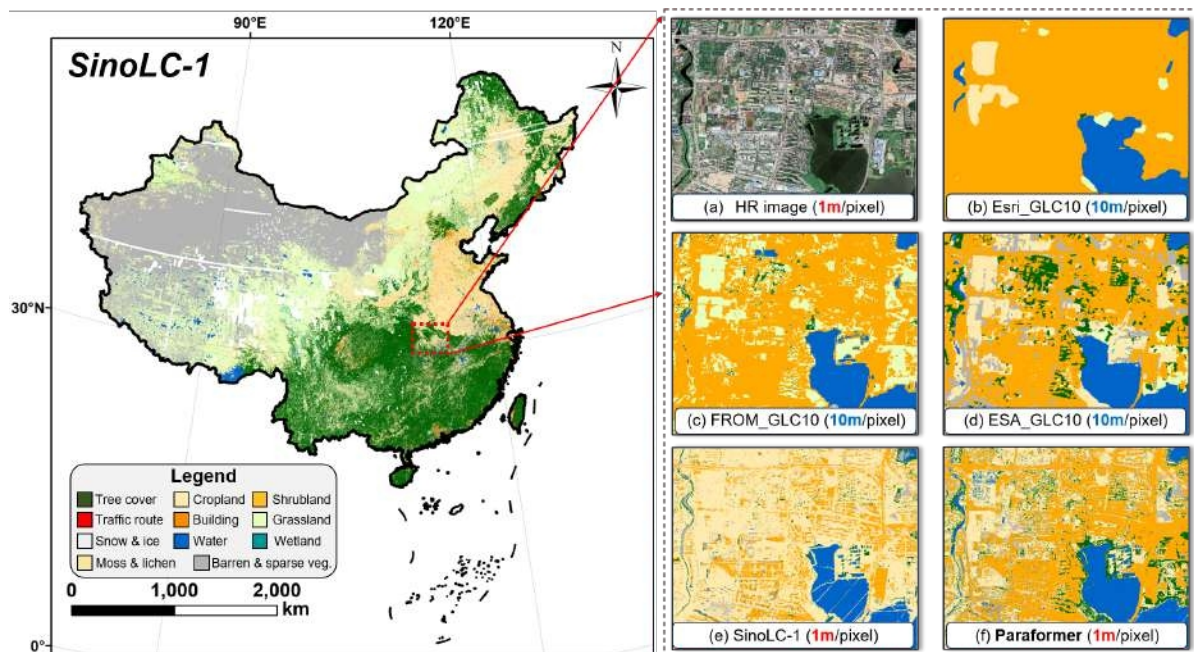
図S10.10mのEsri GLC10トレーニングラベルを用いたポーランドデータセットの視覚的結果。(a) 0.25mの画像、(b) Esri GLC10からサンプリングした10mのラベル。(c) Paraformerの結果。(d) L2HNetの結果。(e) RFの結果。(f) UNetの結果。



図S11.30mのGLC FCS30トレーニングラベルを用いたポーランドデータセットの視覚的結果。(a) 0.5mの画像、(b) GLC FCS30からサンプリングした10mのラベル。(c) Paraformerの結果。(d) L2HNetの結果。(e) RFの結果。(f) UNetの結果。



図S12.フレームワーク学習中の損失関数 \mathcal{L}_{cc} と \mathcal{L}_{mce} のデモ。(a)～(e)はチェサピーク湾データセットの6つの状態における学習過程を示す。



図S13.SinoLC-1データセットの補足実験のデモンストレーション。視覚的な結果は中国武漢市からサンプリングされたものである。Paraformerを使って武漢市全域の1-m土地被覆図を更新し、74.98%のOAを報告している。