英語論文#7

2024/06/19
M2
建元　了

論文の概要

- タイトル
  - **Learning without Exact Guidance: Updating Large-scale High-resolution Land**
- 執筆者
  - Zhuohong Li, Wei He, Jiepan Li, Fangxiao Lu, Hongyan Zhang
- 掲載
  - CVPR2024
- 選択理由
  - 解像度と衛星に関する最新論文

背景

- 土地被覆データは、自然や人間の活動によって頻繁に景観が変化するため、継続的に更新が必要

- HR(高分解能)画像のマッピングはCNNが大半だが、より広範囲で様々な地形への適応には限界があり、局所的で詳細な表示には不向きである

導入

- HR画像は不足しているが、LR（低分解能）の土地被覆データは数十年分ある
- HR画像とLRラベルでは学習ペアの不一致で教師あり学習では不可

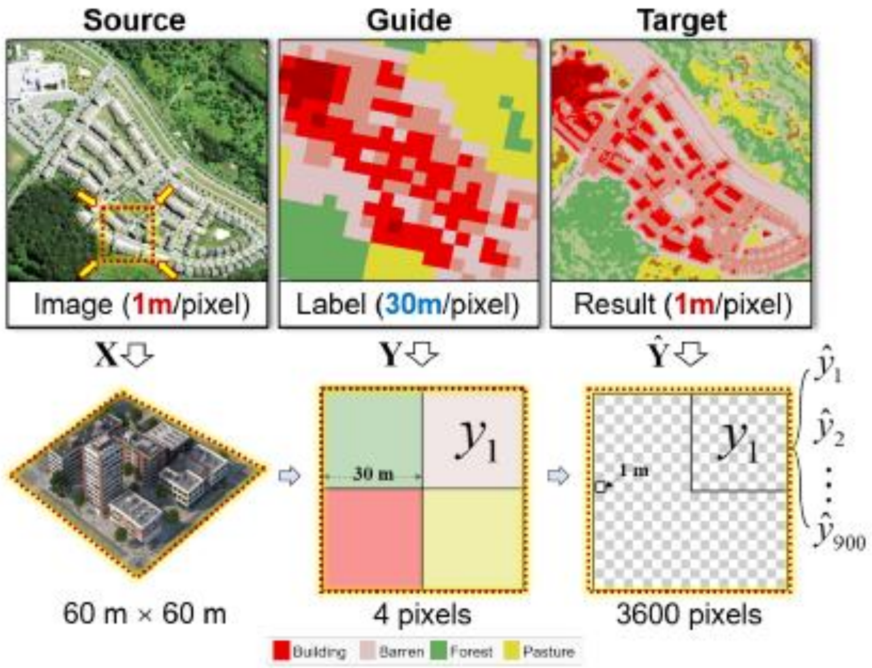

Figure 1. Illustration of resolution mismatched issue in using the HR remote-sensing image (**Source**) and LR historical labels (**Guide**) to generate HR land-cover results (**Target**).

導入

- HRラベルを使用しないEnd-to-Endフレームワークとして Parafomerを提案
- 疑似ラベル支援学習(PLAT)を採用し、LRラベルから信頼性の高い特徴を得る

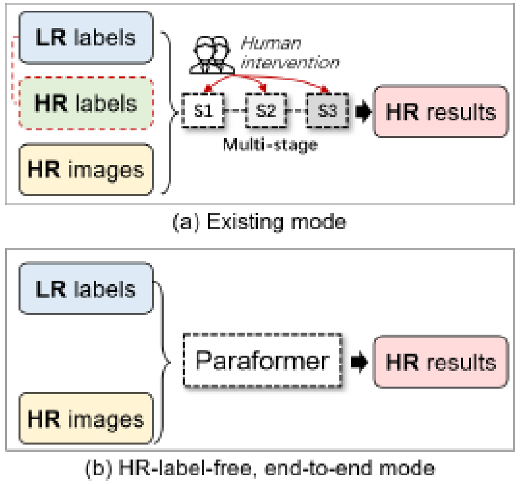

Figure 2. Two modes of large-scale HR land-cover mapping with LR labels. (a) Existing modes either reply on partial HR labels or require non-end-to-end training with human interventions. (b) **Paraformer** aims to form a mode that is HR-label-free and end-to-end trainable.

提案手法

- **Parafomer**

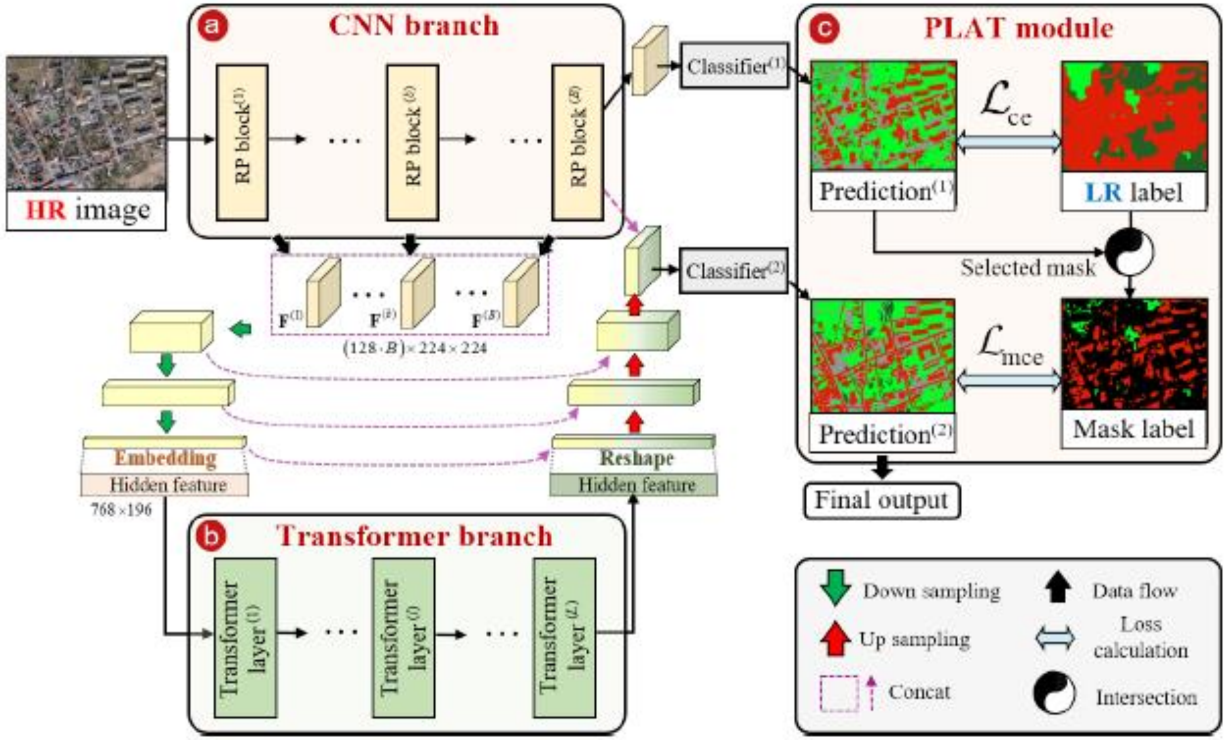

Figure 3. Overall workflow of Paraformer. The framework only takes the HR images and LR labels as training input and includes three components: (a) CNN-based resolution-preserving branch, (b) Transformer-based global-modeling branch, and (c) Pseudo-Label-Assisted Training (PLAT) module.

提案手法

- HR画像とLRラベルでの不一致例

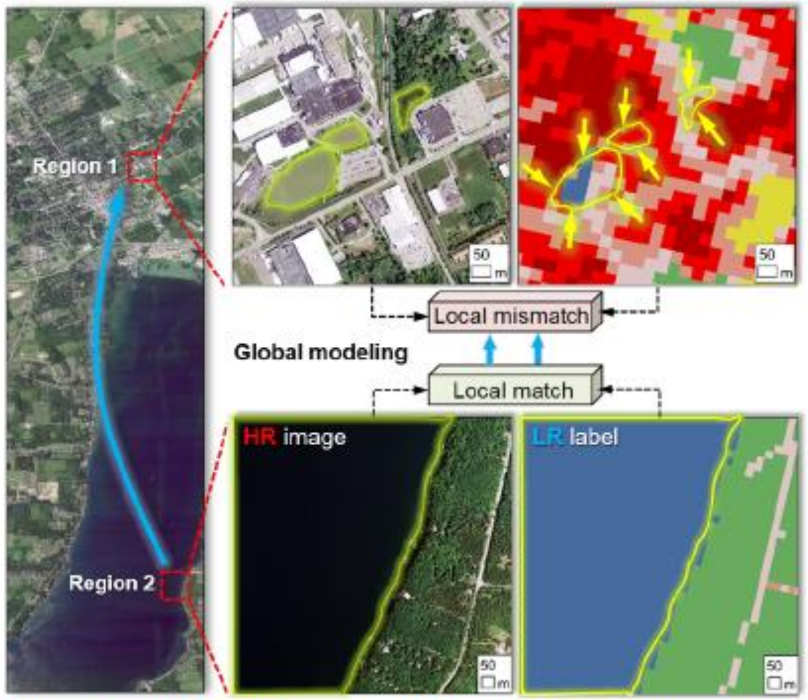

Figure 4. Example of the local mismatch/match in two regions. The edge of water is marked with yellow boundaries. Region 1 shows dispersed lakes around urban areas with unmatched annotation. Region 2 shows a large-scale river with matched annotation.

提案手法

- PLAT
  - HRで表現される予測値$\mathbf{Y}$と、$\hat{\mathbf{Y}}'$として表現されているLRラベルとのクロスエントロピーを計算

$$\mathcal{L}_{\mathrm{ce}}(\mathbf{Y}, \hat{\mathbf{Y}}') = \frac{\sum_{i=0}^{W} \sum_{j=0}^{H} \left[\sum_{l=1}^{L} y_{ij}^{(l)} \log(\hat{y}_{ij}'^{(l)}) \right]}{H \times W}. \quad (1)$$

提案手法

- フレームワークの最終出力は$\hat{\mathbf{Y}}''$として表現

$$\mathcal{L}_{\mathrm{mce}}(\mathbf{M}\cdot\mathbf{Y},\hat{\mathbf{Y}}'') = \frac{\sum_{i=0}^{W}\sum_{j=0}^{H}\left[\sum_{l=1}^{L} y_{ij}^{(l)} m_{ij} \log(\hat{y}_{ij}^{\prime\prime(l)})\right]}{\mathrm{Sum}(\mathbf{M}(i,j)=1)}. \tag{2}$$

$$m_{ij} = \begin{cases} 1 | \, Y_{ij} = Y'_{ij} \\ 0 | \, Y_{ij} \neq Y'_{ij}. \end{cases} \tag{3}$$

提案手法

- 弱教師あり学習PLAT
  - ２つのブランチの損失を組み合わせる

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{ce}} + \mathcal{L}_{\mathrm{mce}}. \tag{4}$$

データセット

- The Chesapeake Bay dataset
  - アメリカの河口周辺から取得した732のタイルから構成
  - size 6000 × 7500
  - HR画像(1m/pixel)はNAIPの4band
  - LRラベル(30m/p)はNLCDの16class
  - HR Ground truths(1m/p)はCCLCから

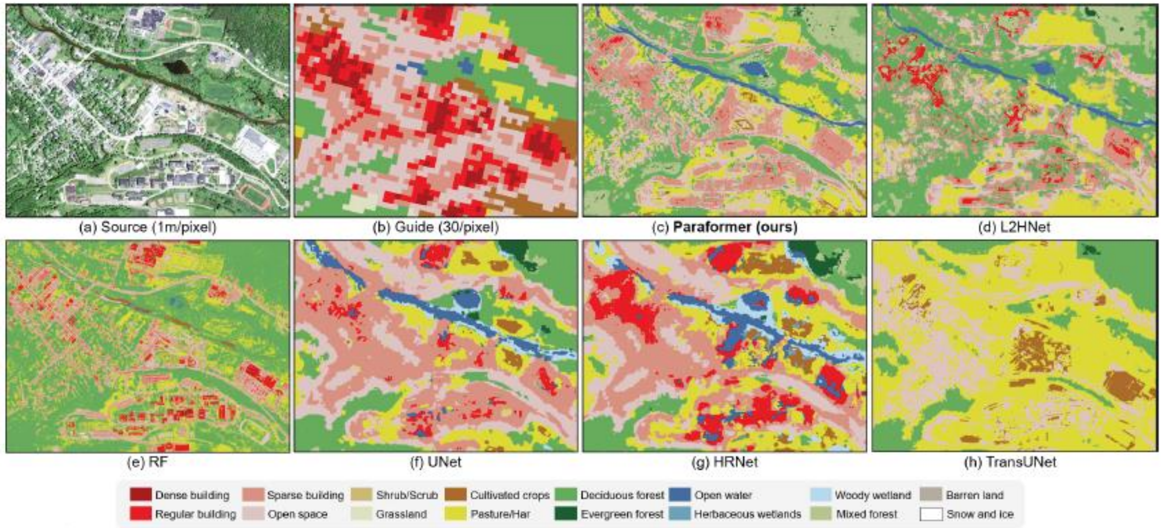

Figure 5. Demonstration of the training data and visual comparisons of the **Paraformer** and other typical methods on the Chesapeake Bay dataset with 16 classes. (a) HR image. (b) LR label. (c) land-cover mapping result of Parafomer. (d–h) land-cover mapping results of five typical methods.

データセット

- The Poland dataset
  - ポーランドの14の州から取得した403のタイル
  - size 1024 × 1024
  - HR画像(0.25m or 0.5m/p)はLandCover.aiから取得し、3band
  - LRラベルは30mと10mから構成
  - HR Ground truthsはOpenEarthMapの7class

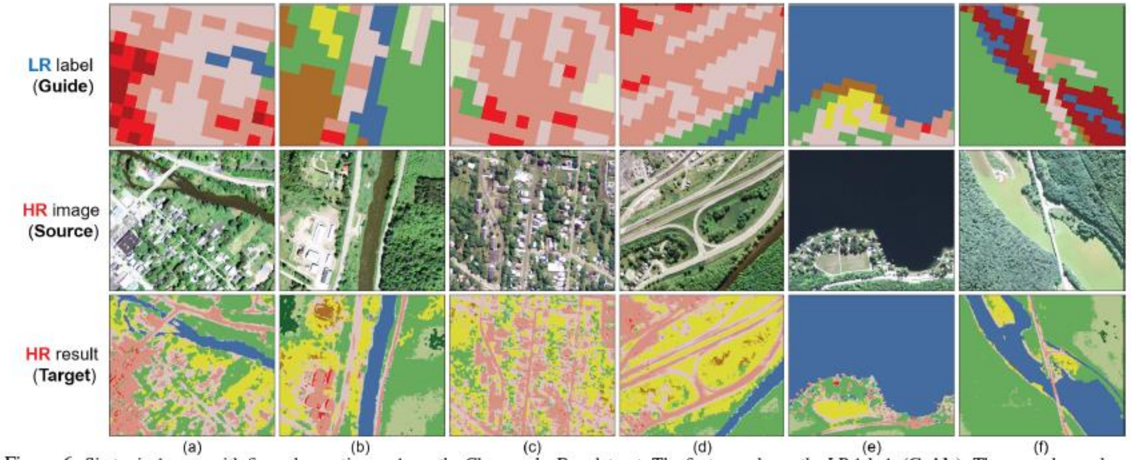

Figure 6. Six typical areas with finer observation scale on the Chesapeake Bay dataset. The first row shows the LR labels (**Guide**). The second row shows the HR images (**Source**). Third row shows the HR results (**Target**) produced by **Paraformer**.

評価

- 全ての手法でLRラベルのみ学習して行う
- Parafomerはパッチサイズ224×224で、オプティマイザーはAdamWで学習
- 学習率0.01で8epoch毎に10%減少
- 出力結果と正解ラベルはmIoUで計算

結果

- The Chesapeake Bay dataset

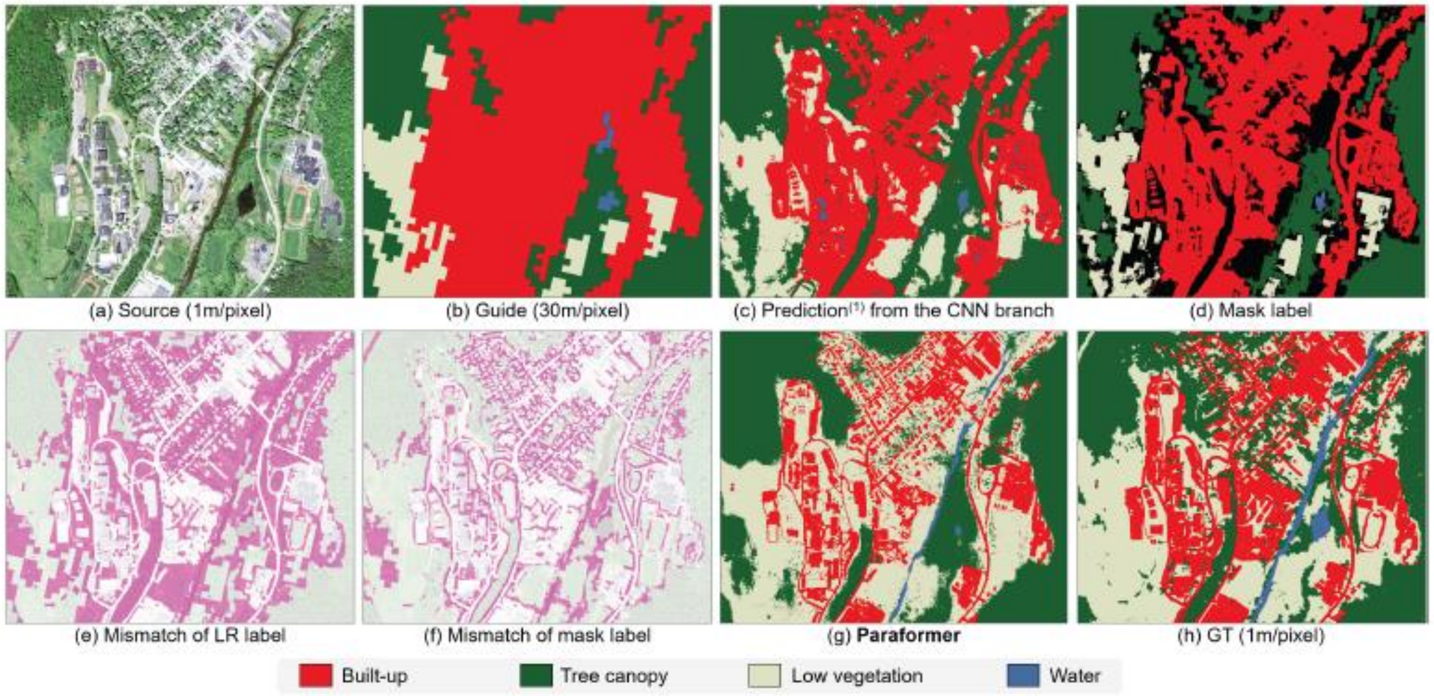

Figure 8. Example of training data and different outputs of Paraformer sampled from the Chesapeake Bay dataset with four unified classes. (a) HR images. (b) LR labels. (c) the primal prediction from the CNN branch. (d) Mask label, as the intersection parts of (b) and (c). The **black areas** are set to void without supervised information. (e–f) the incorrect samples (with pink color) of LR label and mask label. (g) the final results of Paraformer. (h) HR ground truth.

結果

- The Chesapeake Bay dataset

Resolution gap	Method	mIoU (%) of six states in the Chesapeake Bay watershed						
		Delaware	New York	Maryland	Pennsylvania	Virginia	West Virginia	Average
30×	**Paraformer**	**65.57**	**71.43**	**70.20**	**60.04**	68.01	52.62	**64.65**
	L2HNet [28]	61.77	68.12	65.24	58.52	**69.39**	**55.43**	63.08
	TransUNet [10]	53.15	60.53	60.42	51.08	66.21	47.52	56.49
	ConViT [18]	55.26	60.71	61.58	53.94	59.80	49.11	56.73
	CoAtNet [16]	56.89	62.83	61.25	53.57	65.67	51.34	58.59
	MobileViT[34]	58.03	61.32	61.84	55.53	57.04	48.64	57.07
	EfficientViT[5]	53.72	61.28	59.48	51.38	57.34	48.76	55.33
	UNetFormer[48]	58.85	65.11	61.34	59.10	60.84	47.20	58.74
	DC-Swin[47]	59.65	65.99	58.60	58.06	64.11	48.15	59.09
	UNet [38]	54.16	58.79	56.42	53.21	57.34	46.11	54.34
	HRNet [45]	52.11	56.21	50.76	50.03	57.48	45.42	52.00
	LinkNet [8]	58.27	62.05	52.96	52.11	48.71	48.93	53.84
	SkipFCN [26]	60.97	64.83	59.44	55.37	64.72	54.66	60.00
	SSDA [43]	57.91	61.54	54.85	51.71	57.71	47.15	55.15
	RF [7]	59.35	55.03	55.26	51.07	52.29	54.36	54.56

Table 1. The quantitative comparison of the Paraformer and other methods on six states of the Chesapeake Bay watershed. All methods were trained with the 1-m images and 30-m labels. The mIoU (%) of different methods was calculated between their results and the 1-m ground truth.

結果

- The Poland dataset

Max gap	LR label	Paraformer (ours)	L2HNet [28]	TransUNet [10]	ConViT [18]	MobileViT [34]	DC-Swin [47]	HRNet [45]	SkipFCN [26]	RF [7]
		mIoU (%) of different methods								
40×	FROM_GLC10 [9]	**56.57**	50.15	38.44	39.36	41.03	43.56	43.66	27.14	21.48
	ESA_GLC10 [44]	**55.19**	52.13	35.58	36.09	38.42	40.05	49.81	28.34	26.97
	Esri_GLC10 [22]	**55.07**	50.78	37.79	38.78	38.50	39.91	46.65	28.18	19.36
120×	GLC_FCS30 [56]	**49.39**	43.62	26.20	29.16	29.57	30.14	41.46	23.67	17.02

Table 2. The quantitative comparison on the Poland dataset. The mIoU (%) of the Paraformer and other methods that trained with three types of 10-m labels (i.e., FROM_GLC10, ESA_GLC10, and Esri_GLC10) and one type of 30-m label (i.e., GLC_FCS30) are demonstrated.

アブレーション実験

- Parafomerの機能評価

Ablation method	mIoU (%) of six states in the Chesapeake Bay watershed								
	Delaware	New Your	Maryland	Pennsylvania	Virginia	West Virginia	Average	Params	FLOPs
Paraformer	**65.57**	**71.43**	**70.20**	**60.04**	**68.01**	**52.62**	**64.65**	109.4M	141.3G
Sole CNN branch	59.57	67.87	64.30	53.86	65.26	50.01	60.15	4.5M	56.1G
Sole Transformer branch	53.15	60.53	60.42	51.08	66.22	47.52	56.49	96.9M	83.3G
Hybrid without PLAT	62.69	70.39	67.15	58.33	67.47	50.83	62.81	109.4M	141.3G

Table 3. The ablation results of the Paraformer on six states of the Chesapeake Bay watershed. The sole CNN branch, sole Transformer branch, and Hybrid without PLAT aim to investigate the contribution of the CNN branch, Transformer branch, and PLAT module, respectively.

アブレーション実験

- 各ブランチでのコンテキストの可視化

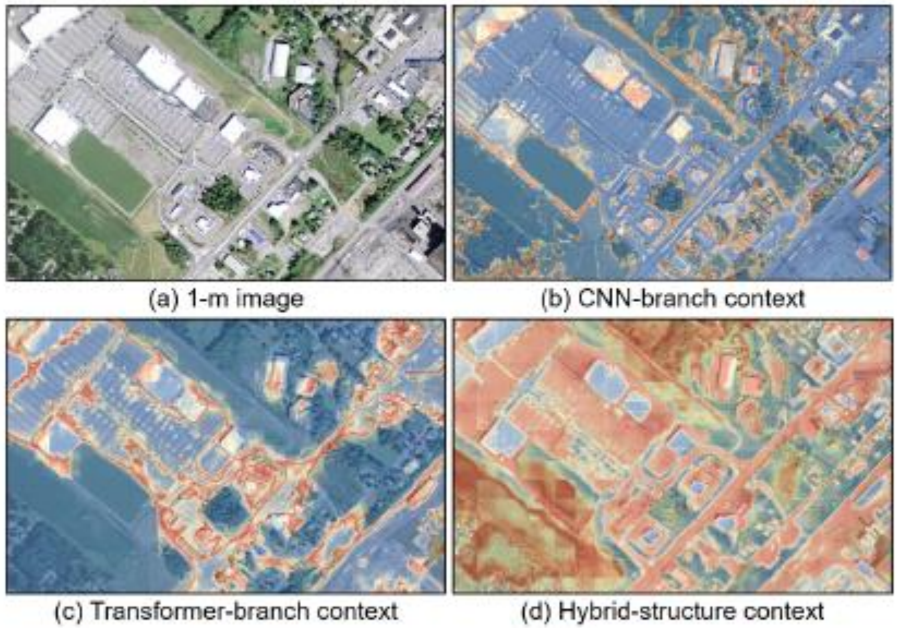

Figure 9. Demonstration of the extracted contexts from the ablation methods. (a) the original HR image. (b) the contexts extracted by the sole CNN branch. (c) the contexts extracted by the sole Transformer branch . (d) the contexts extracted by the CNN-Transformer hybrid backbone.

結論

- HR土地被覆マップの作成に、LRラベルを安定的に利用できる
- アブレーションより、並列CNN-Transfomer構造とPLATが有効
- 中間結果とブランチの可視化よりコンテキストを示せた