Robust Thick Cloud Removal for Multitemporal Remote Sensing Images Using Coupled Tensor Factorization

Jie Lin[1],

Ting-Zhu Huang[1], Xi-Le Zhao[1], Yong Chen[2], Qiang Zhang[3], Qiangqiang Yuan[3]

1. University of Electronic Science and Technology of China
2. Jiangxi Normal University
3. Wuhan University

CSIAM 2022

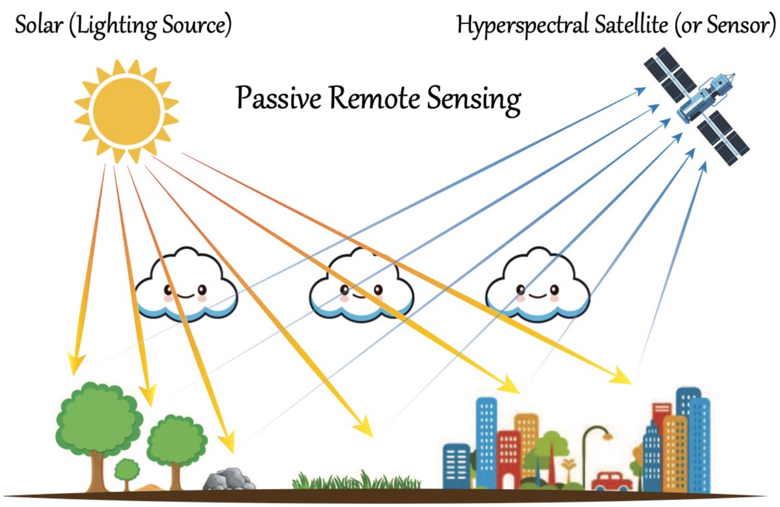
Outline

- Background
- Methodology
- Experiment
- Conclusion

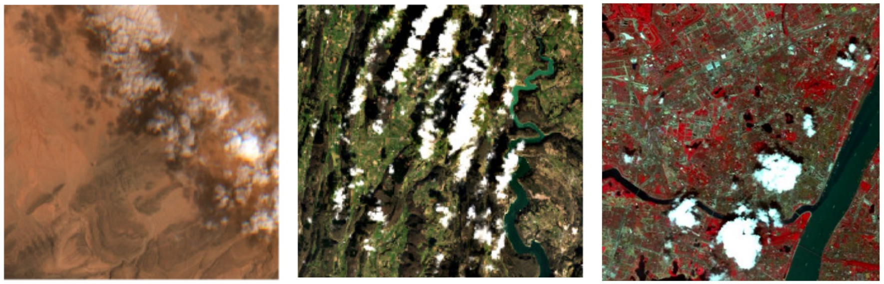
Background

According to the research [1], the **cloud** covers approximately **35%** of **the earth's surface** in anytime.

➢Imaging Process[2]



➢Observed RS Images



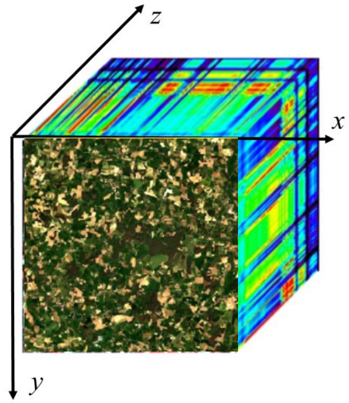
Sentinel-2 Landsat-8 GF 1

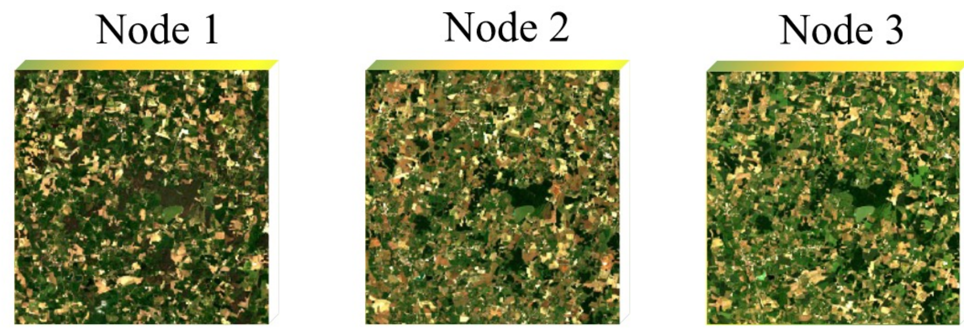
Thick cloud greatly reduce data usability! → **Multitemporal images Thick Cloud Removal**

[1] Junchang Ju, David P. Roy, "The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally", *Remote Sensing of Environment*, 2008,

[2] D. Hong *et al.*, "Interpretable Hyperspectral Artificial Intelligence: When nonconvex modeling meets hyperspectral remote sensing", in *IEEE Geoscience and Remote Sensing Magazine*, 2021.

Background





➢Existing Thick Cloud Removal Methods

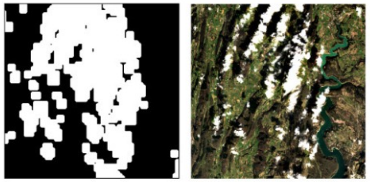
Auxiliary information

- **Spatial** information-based methods
- **Spectral** information-based methods
- **Temporal** information-based methods
- **Hybrid**-based methods
  - Can handle the large thick cloud contamination

➢ Existing Thick Cloud Removal Methods

Given mask?
- Non-Blind
- Blind

- Can not achieve a more ideal cloud removal **when the mask is inaccurate**



■ TVLRSDC model [3]

$$\min_{X,S,N} \frac{1}{2}\|N\|_F^2 + \lambda_1 \|X\|_* + \lambda_2 \|S\|_1, \quad s.\ t.\ \ Y = X + S + N.$$

- Low-rankness is not strong.
- Discard all given mask information

[3] Yong Chen, Wei He, Naoto Yokoya, Ting-Zhu Huang, Blind cloud and cloud shadow removal of multitemporal images based on total variation regularized low-rank sparsity decomposition, *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019.

Outline

- Background
- Methodology
- Experiment
- Conclusion

Methodology

➢ Question

- Is there any **latent relationships** **between the multitemporal RS images**, which can be exploited to finely reconstruct themultitemporal information?

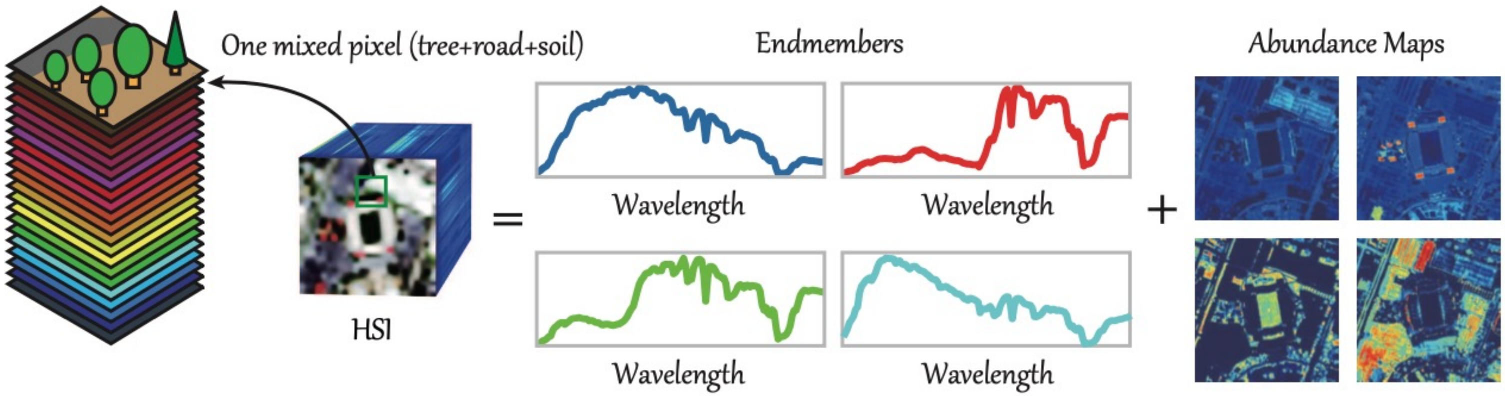
- How to make a **balance** **between the** **nonblind** **methods and the** **blind** **methods** to achieve the reasonable use of themasks that comes with RS imagery products?

➢ **Aim:** Thick cloud removal for remote sensing image

Hybrid-based + Semi-Blind Method

Methodology

➢Inspiration [2]

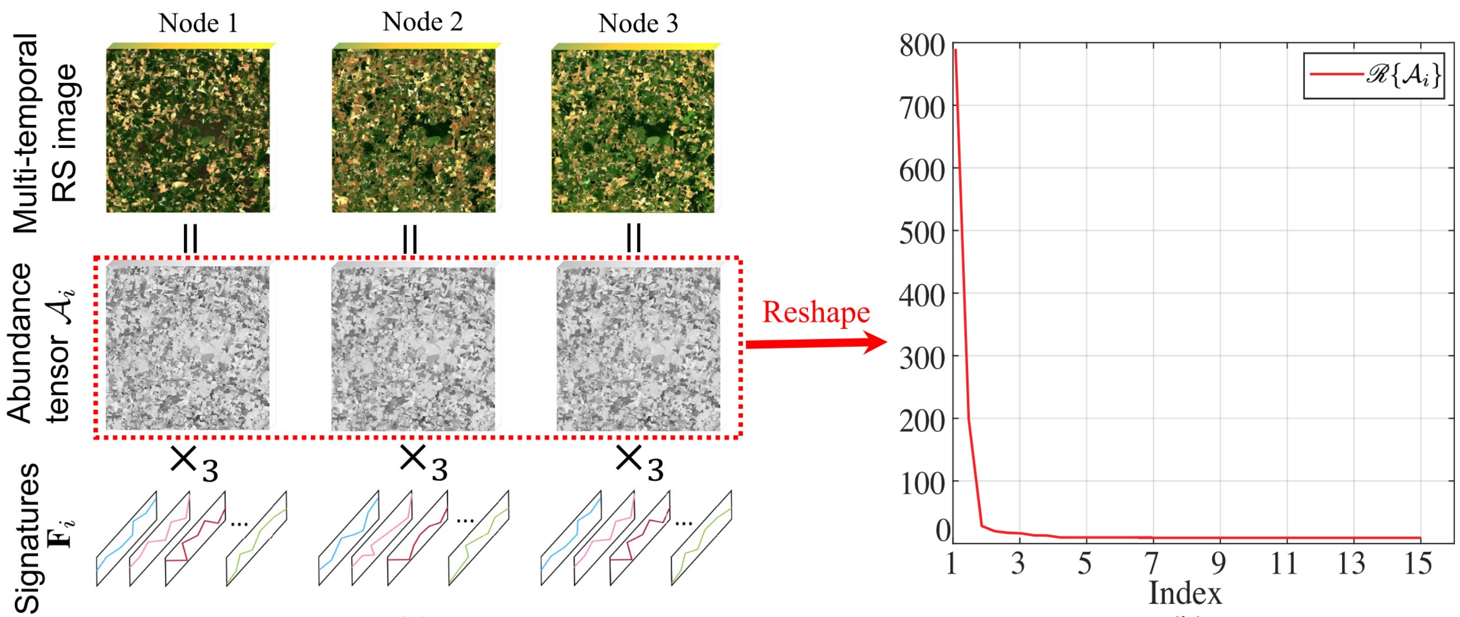


Inspired by unmixing , as the distribution of surface material is constant over a period and the same material shows different spectral signatures at different time nodes, **the multi-temporal images in the same scene share the same abundances.**

[2] D. Hong *et al.*, "Interpretable Hyperspectral Artificial Intelligence: When nonconvex modeling meets hyperspectral remote sensing", in *IEEE Geoscience and Remote Sensing Magazine*, 2021.

Methodology

➢Key Observation



Then, we use a coupled tensor factorization to explore this relationship, which decomposes the image at each time node into an abundance tensor that implies material distribution and orthogonal endmembers. **There is a strong similarity between abundance tensors over a period.**

Methodology

Proposed Method

- Degradation model

$$\mathcal{Y} = \mathcal{M} \odot \mathcal{X} + \mathcal{C}$$

- Decomposition model

$$\mathcal{X}_i = \mathcal{A}_i \times_3 \mathbf{F}_i,$$

- Proposed Model

$$\min_{\mathcal{X},\mathcal{C},\mathcal{A}_i,\mathbf{F}_i} \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \odot \mathcal{X} - \mathcal{C}\|_F^2 + \beta\|\mathcal{C}\|_0 + \alpha \mathrm{Rank}(\mathbf{A})$$
$$\text{s.t. } \mathcal{X}_i = \mathcal{A}_i \times_3 \mathbf{F}_i, \quad \mathbf{F}_i^T \mathbf{F}_i = \mathbf{I}$$

$$\boxed{\begin{aligned}&\min_{\mathcal{X},\mathcal{C},\mathcal{A}_i,\mathbf{F}_i} \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \odot \mathcal{X} - \mathcal{C}\|_F^2 + \beta\|\mathcal{C}\|_1 + \alpha\|\mathbf{A}\|_*\\ &\text{s.t. } \mathcal{X}_i = \mathcal{A}_i \times_3 \mathbf{F}_i, \quad \mathbf{F}_i^T \mathbf{F}_i = \mathbf{I}.\end{aligned}}$$

Developed ALM algorithm

Algorithm 2 ALM Algorithm for Cloud/Shadow Removal

Input: Target RS images $\mathcal{Y}$, regularization parameters α and β, and penalty parameters γ and ρ.

Initialize: $\mathcal{X} = \mathcal{Y}$, $\mathcal{C} = \mathcal{P}_i = \mathcal{O}$, and $\mathbf{W} = \mathbf{Q} = \mathbf{0}$.
while not converged **do**
  Update $\{\mathbf{F}_i^{k+1}\}$ by (5);
  Update $\{\mathcal{A}_i^{k+1}\}$ by (6);
  Update $\mathbf{W}^{k+1}$ by (7);
  Update $\mathcal{C}^{k+1}$ by (8);
  Update $\mathcal{X}^{k+1}$ by (11);
  Update $\{\mathcal{P}_i^{k+1}\}$ and $\mathbf{Q}^{k+1}$ by (12);
  Refine mask $\mathcal{M}$ by Algorithm 1;
  Check the convergence condition:

$$\|\mathcal{X}^{k+1} - \mathcal{X}^k\|_F^2 / \|\mathcal{X}^k\|_F^2 \leq 10^{-4}.$$

end while

Output: Reconstructed RS images $\mathcal{X}$.

Methodology

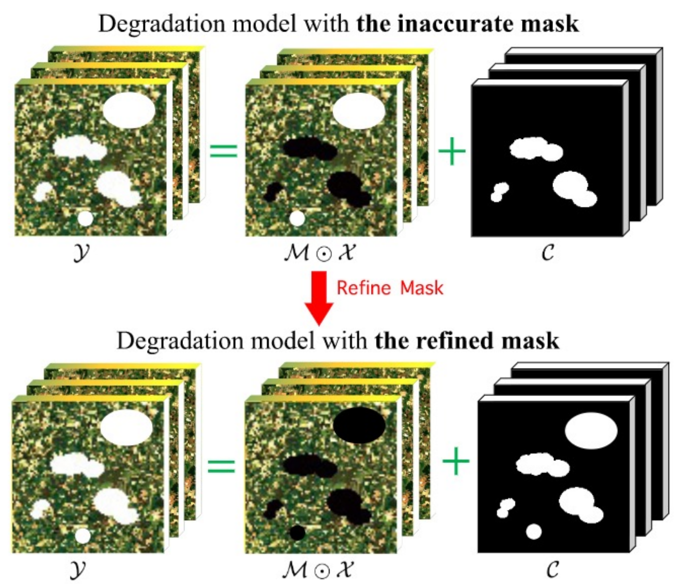
➢Mask Refinement

Algorithm 1 Adaptive Threshold Algorithm for Mask Refinement

Input: Error component $\mathcal{E}$, given mask $\mathcal{M}^0$, and corresponding cloud/shadow index set Ω^0.

Initialize: $\Omega = \Omega^0$, $\mathcal{M} = \mathcal{M}^0$.
for $i = 1{:}t$ **do**
  for $p_1 = 1{:}m$ **do**
    for $p_2 = 1{:}n$ **do**
      Compute $a = \text{mean}[\mathcal{E}(p_1, p_2, (i\text{-}1)b{+}1{:}ib)]$;
      Compute
        $\tau = \min\{|\text{mean}[(\mathcal{E})_{\Omega^0}(p_1, p_2, (i\text{-}1)b{+}1{:}ib)]|\}$;
      $\Omega = \Omega \cup (p_1, p_2, (i\text{-}1)b{+}1{:}ib)$, if $|a| > \tau$;
    end for
  end for
end for
Let $(\mathcal{M})_\Omega = 0$;

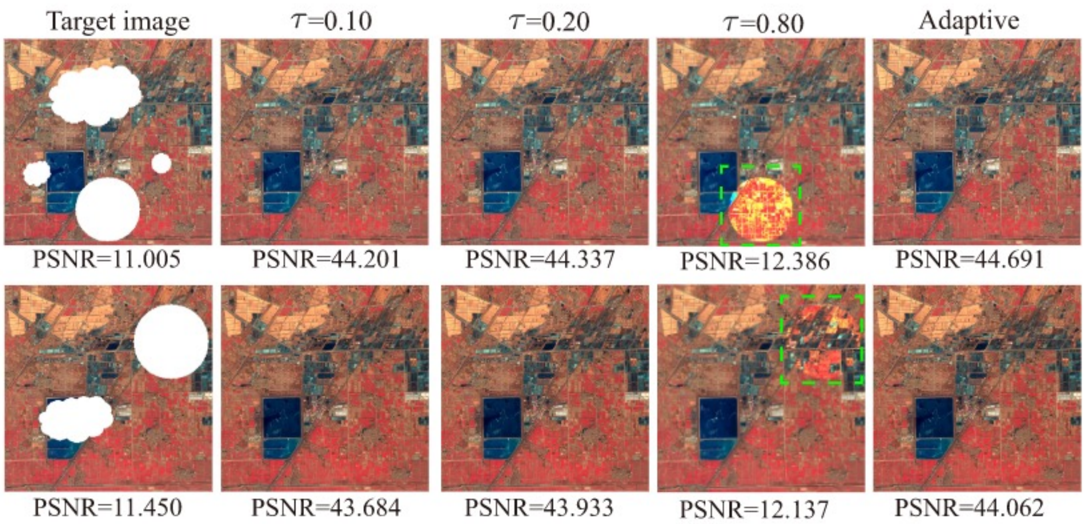
Output: Refined mask $\mathcal{M}$.

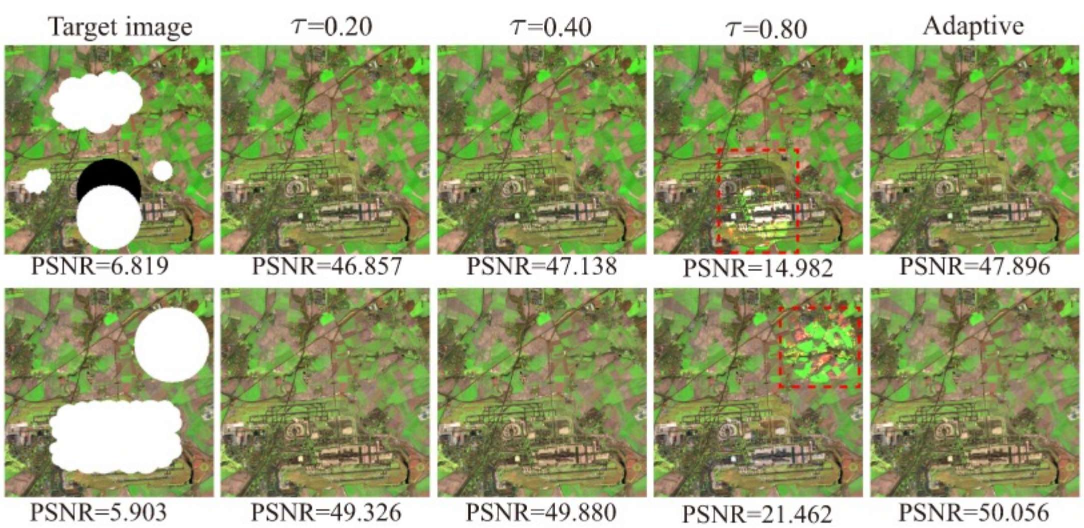


We **embed the cloud/shadow detection (Algorithm 1) in each iteration of Algorithm 2** to refine the mask. The refined mask will help to introduce true information from observed images for multi-temporal feature learning.

Methodology

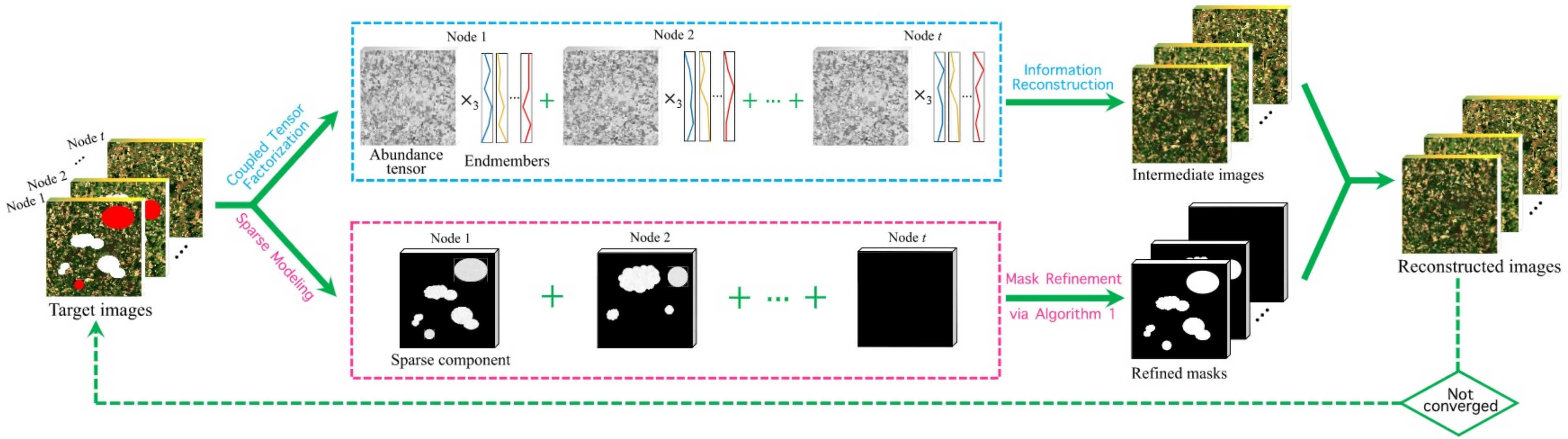
➢Discussion





Methodology

➢Flowchart



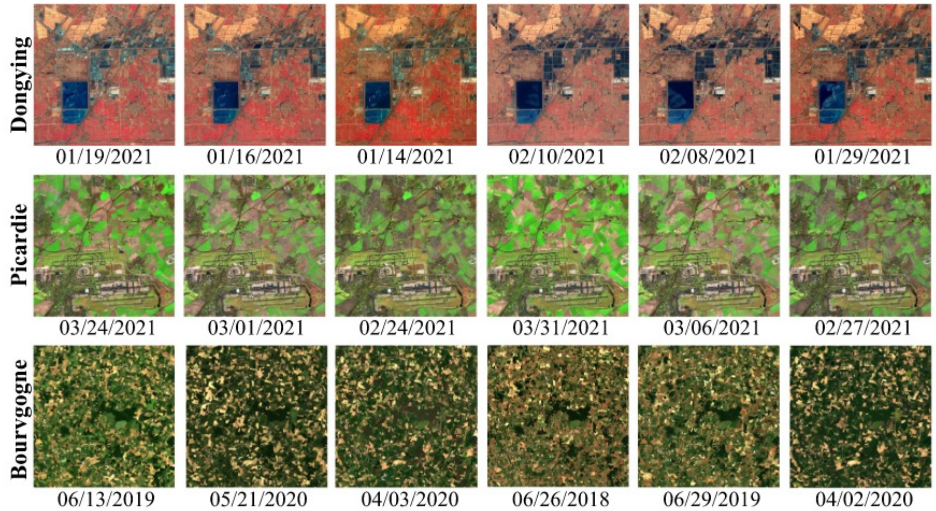
Outline

- Background
- Methodology
- Experiment
- Conclusion

➢ Simulated Experiment

■ Dataset

1) *Dongying*[1]*:* This dataset is taken over Dongying, China, by Sentinel-2, and each time node contains four spectral bands (B2, B3, B4, and B8) with 10-m spatial resolution. The subimages of size $500 \times 500 \times 4$ of six time nodes are used in experiments.
2) *Picardie*[1]*:* This dataset is taken over Picardie, France, by Sentinel-2, and each time node contains six spectral bands (B5, B6, B7, B8A, B11, and B12) with 20-m spatial resolution. The subimages of size $1000 \times 1000 \times 6$ of six time nodes are used in experiments.
3) *Bourgogne*[2]*:* This dataset is taken over Bourgogne, France, by Landsat-8, and each time node contains seven spectral bands (B1, B2, B3, B4, B5, B6, and B7) with 30-m spatial resolution. The subimages of size $400 \times 400 \times 7$ of six time nodes are used in experiments.

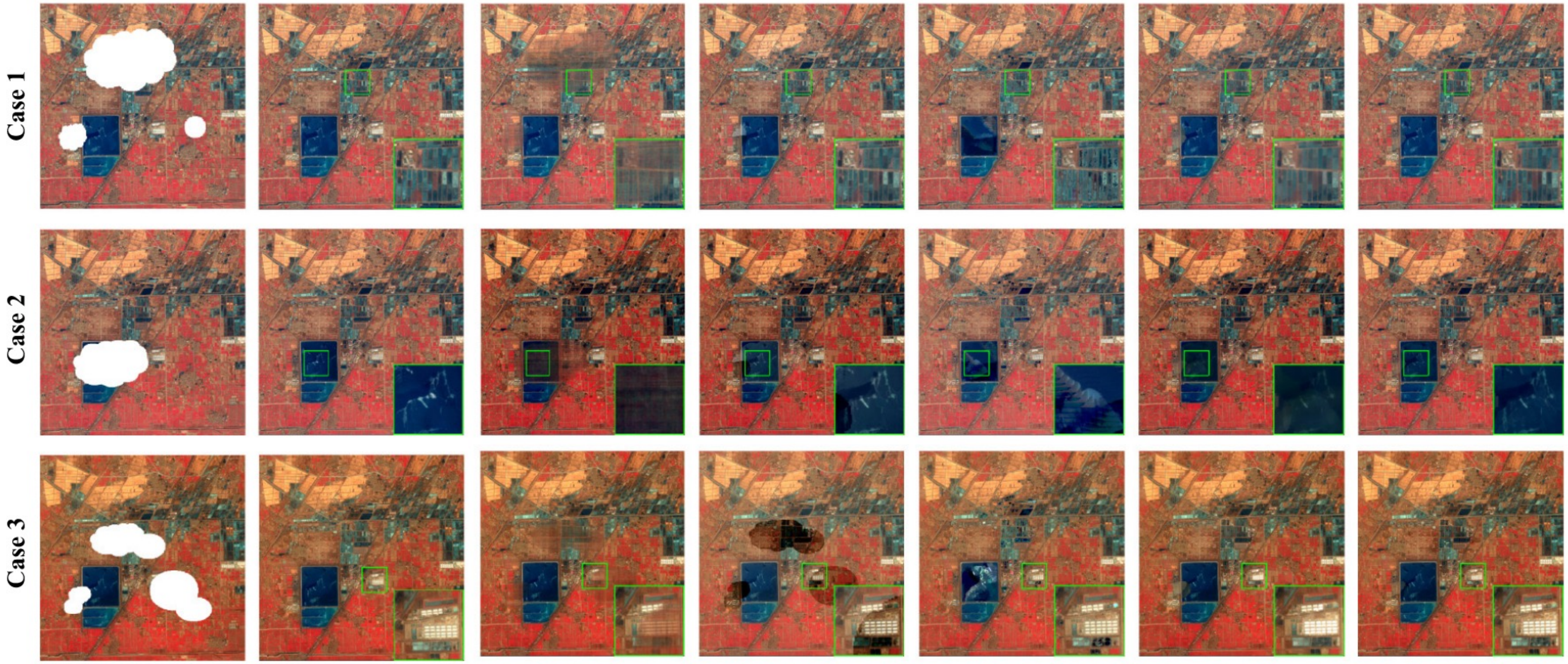


Experiment

➢Simulated Experiment

- Accurate Mask

		Dongying					
Case	Index	Target	HaLRTC	ALM-IPG	WLR	TVLRSDC	RTCR
Case 1	PSNR	12.802	36.913	41.761	39.097	43.131	**46.536**
	SSIM	0.8686	0.9702	0.9931	0.9837	0.9927	**0.9963**
	CC	0.3315	0.9776	0.9962	0.9842	0.9953	**0.9974**
Case 2	PSNR	15.138	38.761	39.226	38.710	42.237	**45.788**
	SSIM	0.9334	0.9856	0.9934	0.9829	0.9935	**0.9964**
	CC	0.1198	0.9741	0.9876	0.9821	0.9920	**0.9975**
Case 3	PSNR	14.579	40.754	31.726	37.197	38.221	**45.656**
	SSIM	0.8937	0.9861	0.9738	0.9818	0.9926	**0.9963**
	CC	0.3445	0.9898	0.9426	0.9636	0.9865	**0.9961**
Time (min)		—	**3.104**	6.876	4.962	8.210	3.763



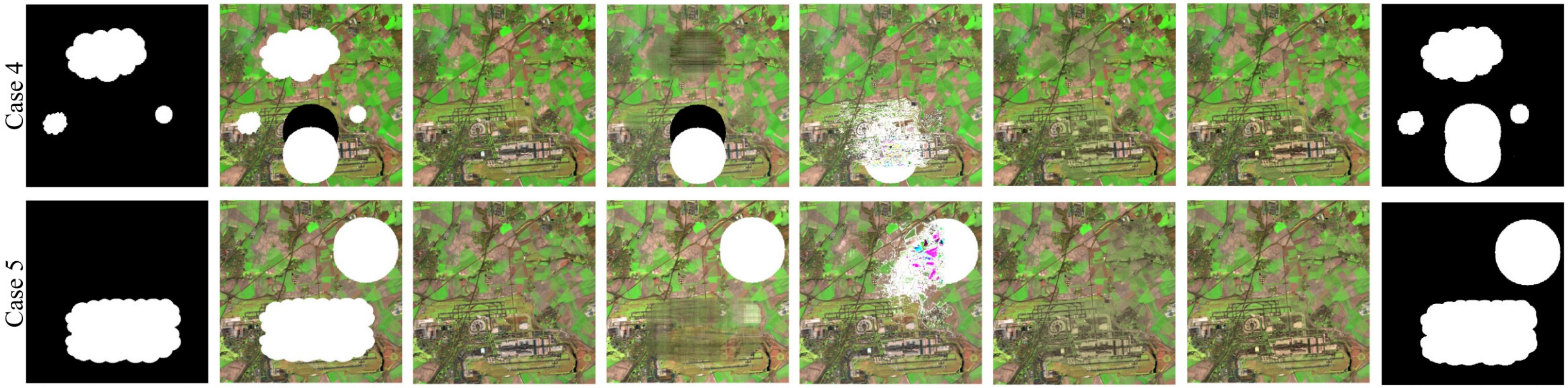
(a) Target (b) Cloud-free (c) HaLRTC (d) ALM-IPG (e) WLR (f) TVLRSDC (g) RTCR

Experiment

➢ Simulated Experiment

■ Inaccurate Mask

Dataset	Case	Index	Target	HaLRTC	WLR	TVLRSDC	RTCR
Picardie	Case 4	PSNR	6.819	11.879	13.673	45.757	**47.896**
		SSIM	0.7055	0.8530	0.7757	0.9912	**0.9952**
		CC	0.0988	0.1398	0.1607	0.9559	**0.9838**
	Case 5	PSNR	5.903	10.573	10.822	46.398	**50.052**
		SSIM	0.6530	0.8413	0.7405	0.9912	**0.9971**
		CC	0.0649	0.0068	0.0205	0.9597	**0.9847**
	Time (min)		—	**1.047**	4.971	4.542	1.508

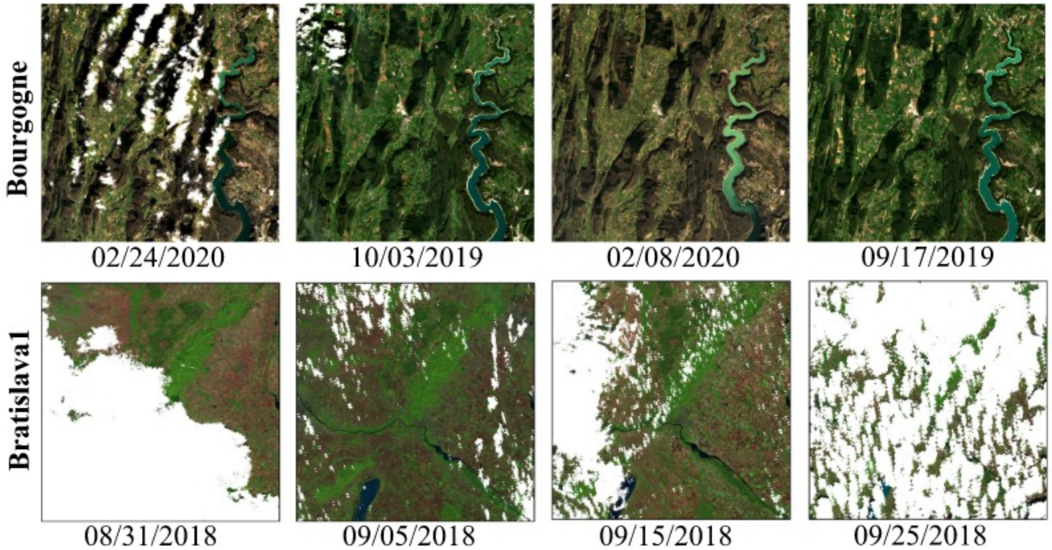


(a) Given mask (b) Target (c) Cloud-free (d) HaLRTC (e) WLR (f) TVLRSDC (g) RTCR (h) Final mask

➢Real Experiment

- Dataset

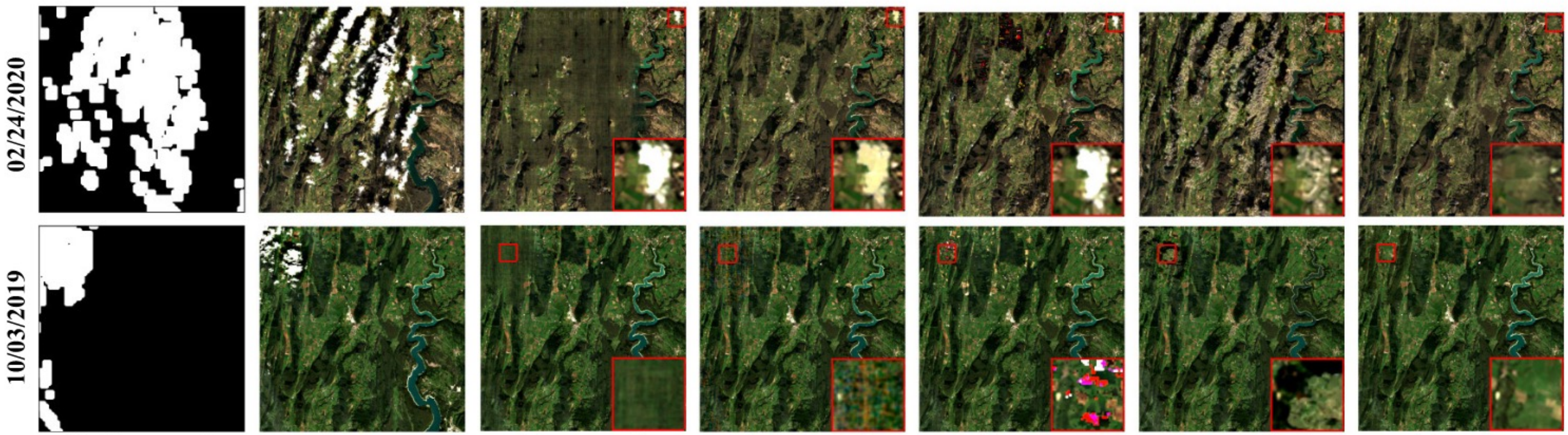
1) *Bourgogne*[2]*:* This dataset is taken by Landsat-8, and each time node contains seven spectral bands with 30-m spatial resolution. The subimages of size $600 \times 600 \times 7$ of four time nodes are used in experiments.
2) *Bratislava*[1]*:* This dataset is taken over Bratislava, Slovakia, by Sentinel-2, and each time node contains six spectral bands with 20-m spatial resolution. The full images of size $5490 \times 5490 \times 6$ of four time nodes are used in experiments.



Experiment

➢Real Experiment

- Inaccurate Mask

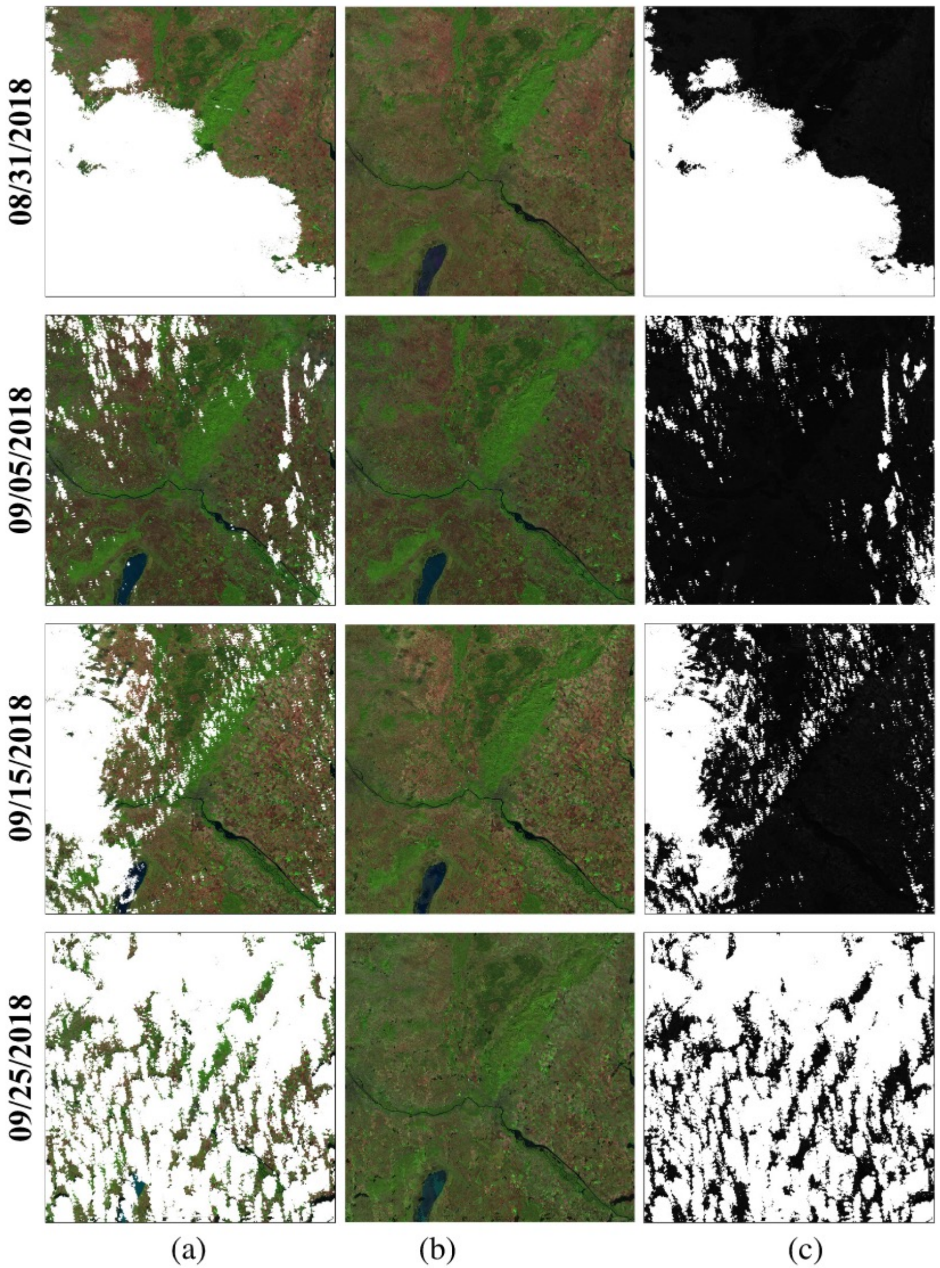


(a) Given mask (b) Target (c) HaLRTC (d) ALM-IPG (e) WLR (f) TVLRSDC (g) RTCR

Experiment

➢Real Experiment

- Large Scene



Outline

- Background
- Methodology
- Experiment
- Conclusion

Conclusion

➢New Perspective:

$$\min_{\mathcal{X},\mathcal{C},\mathcal{A}_i,\mathbf{F}_i} \frac{1}{2}\|\mathcal{Y}-\mathcal{M}\odot\mathcal{X}-\mathcal{C}\|_F^2+\beta\|\mathcal{C}\|_0+\alpha\mathrm{Rank}(\mathbf{A})$$
$$\text{s.t. } \mathcal{X}_i=\mathcal{A}_i\times_3\mathbf{F}_i, \quad \mathbf{F}_i^T\mathbf{F}_i=\mathbf{I}$$

➢Semi-blind Decloud

A balance between the non-blind and the blind.

Conclusion

Thanks!

Jie Lin

University of Electronic Science and Technology of China

Homepage: *https://jielin96.github.io*