



Article

Towards Edge-Precise Cloud and Shadow Detection on the GaoFen-1 Dataset: A Visual, Comprehensive Investigation

Libin Jiao ¹, Mocun Zheng ¹, Ping Tang ² and Zheng Zhang ^{2,*}

¹ Department of Computer Science and Technology, School of Mechanical Electronic and Information Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China

² Aerospace Information Research Institute (AIR), Chinese Academy of Sciences (CAS), Beijing 100101, China

* Correspondence: zhangzheng@aircas.ac.cn

Abstract: Remote sensing images are usually contaminated by opaque cloud and shadow regions when acquired, and therefore cloud and shadow detection arises as one of the essential prerequisites for restoration and prediction of the objects of interest underneath, which are required by further processing and analysis. Cutting-edge, learning-based segmentation techniques, given a well-labeled, sufficient sample set, are significantly developed for such a detection issue and can already achieve region-accurate or even pixel-precise performance. However, it may possibly be problematic to attempt to apply the sophisticated segmentation techniques to label-free datasets in a straightforward way, more specifically, to the remote sensing data generated by the Chinese domestic satellite GaoFen-1. We wish to partially address such a segmentation problem from a practical perspective rather than in a conceptual way. This can be performed by considering a hypothesis that a segmentor, which is sufficiently trained on the well-labeled samples of common bands drawn from a source dataset, can be directly applicable to the custom, band-consistent test cases from a target set. Such a band-consistent hypothesis allows us to present a straightforward solution to the GaoFen-1 segmentation problem by treating the well-labeled Landsat 8 Operational Land Imager dataset as the source and by selecting the fourth, the third, and the second bands, also known as the false-color bands, to construct the band-consistent samples and cases. Furthermore, we attempt to achieve edge-refined segmentation performance on the GaoFen-1 dataset by adopting our prior Refined UNet and v4. We finally verify the effectiveness of the band-consistent hypothesis and the edge-refined approaches by performing a relatively comprehensive investigation, including visual comparisons, ablation experiments regarding bilateral manipulations, explorations of critical hyperparameters within our implementation of the conditional random field, and time consumption in practice. The experiments and corresponding results show that the hypothesis of selecting the false-color bands is effective for cloud and shadow segmentation on the GaoFen-1 data, and that edge-refined segmentation performance of our Refined UNet and v4 can be also achieved.

Keywords: cloud and shadow detection; object edge refinement; end-to-end inference of UNet and CRF; GaoFen-1 dataset customization



Citation: Jiao, L.; Zheng, M.; Tang, P.; Zhang, Z. Towards Edge-Precise Cloud and Shadow Detection on the GaoFen-1 Dataset: A Visual, Comprehensive Investigation. *Remote Sens.* **2023**, *15*, 906. <https://doi.org/10.3390/rs15040906>

Academic Editor: Farid Melgani

Received: 22 December 2022

Revised: 26 January 2023

Accepted: 1 February 2023

Published: 6 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Spectral remote sensing images are widely used in terrestrial remote sensing applications but are usually contaminated by clouds and shadows when acquired [1,2]. In particular, opaque cloud and shadow regions fatally override the presence of the objects of interest, giving rise to low-usable remote sensing products for further processing and analysis. Restoration and prediction of such objects underneath cloud and shadow regions, therefore, arise from the requirement of the usage of cloud- and shadow-free remote sensing images, such that perceiving and locating cloud and shadow regions becomes an essential prerequisite for such further manipulations of remote sensing images [1,3,4]. Given sufficient, typical, well-labeled samples together with cloud and shadow labels, current

learning-based segmentation techniques are significantly developed for such a detection issue and can already achieve region-accurate or even pixel-precise performance, in which the segmentors based on convolutional neural networks (CNNs) are representatives and are thus exploited, such as UNet, SegNet, and CNN variants [1,5–7]. In addition, the exploration of cloud and shadow detection using cutting-edge, learning-based techniques is still proceeding.

A demanding requirement of training samples together with at least region-accurate cloud and shadow labels, however, arises from the popular usage of such learning-based segmentation methods. In practice, the segmentors for the Landsat 8 Operational Land Imager (L8) dataset [3] can be obtained due to the relatively region-accurate labels derived from the quality-assessment (QA) bands, whereas the currently available approaches to generating the approximate labels for the label-free Chinese domestic satellite GaoFen-1 (GF1) dataset are threshold segmentation [8] and manual labeling [9], making it practically hard to apply the learning-based segmentation techniques to the GF1 datasets in a straightforward way. This requires the solutions to the label-free learning problems or the label-transferring techniques between the source and the target sets.

Naturally, it is our study goal in this work to find a way of reusing the L8 labels in order to address the GF1 edge-precise segmentation problem of cloud and shadow detection from a practical perspective rather than in a conceptual way. This can be partially performed by considering a hypothesis that a segmentor, which is sufficiently trained on the well-labeled samples of common bands drawn from a source dataset, can be directly applicable to the custom, band-consistent test cases from a target set. More specifically, given the source and the target datasets, we first create two band-consistent input datasets by selecting the common bands and then train the segmentor on the set derived from the source. The inference for the test cases drawn from the target can be directly performed because of the band-consistent inputs. Such a band-consistent hypothesis allows us to present a straightforward solution to the GF1 segmentation problem by treating the well-labeled L8 dataset as the source and by selecting the fourth, the third, and the second bands, also known as the false-color bands, to construct the band-consistent samples and cases. In particular, common bands 4, 3, and 2 of both the L8 and GF1 data are adopted, not only because of the band consistency of inputs but also the sufficiently visible cloud and shadow regions and edges. Then the initial cloud and shadow labels for the GF1 data can be given using such a solution. In addition, we attempt to satisfy the requirement of edge-precise segmentation by applying our prior Refined UNet and v4 to the GF1 segmentation, giving rise to edge-refined segmentation proposals for GF1 inputs. We finally verify the effectiveness of the band-consistent hypothesis and the edge-refined approaches by performing a relatively comprehensive investigation, including visual comparisons, ablation experiments regarding bilateral manipulations, explorations of critical hyperparameters within our implementation of the conditional random field (CRF), and time consumption in practice. Some typical segmentation results, as illustrated in Figure 1, show that edge-refined labels of cloud and shadow regions are generated both by Refined UNet and by v4, in comparison with UNet giving region-accurate labels. In general, the main contributions of our investigation are as follows.

- Label-transferring solution to the GF1 segmentation problem: we present a label-transferring solution to the GF1 segmentation problem by considering that the segmentor can be directly applicable to the custom, band-consistent test cases from the GF1 data if it is sufficiently trained on the well-labeled samples of common bands drawn from the L8 dataset.
- Deployment of the Refined UNets on the custom GF1 datasets: we achieve edge-refined segmentation performance on the custom GF1 data by adopting our Refined UNet and v4.
- Comprehensive investigation into the edge-precise GF1 segmentation solution: we confirm our edge-refined achievements on the GF1 data by performing a relatively visual, comprehensive investigation, including visual comparisons, ablation exper-

iments regarding bilateral manipulations, explorations of critical hyperparameters within our CRF, and time consumption in practice.

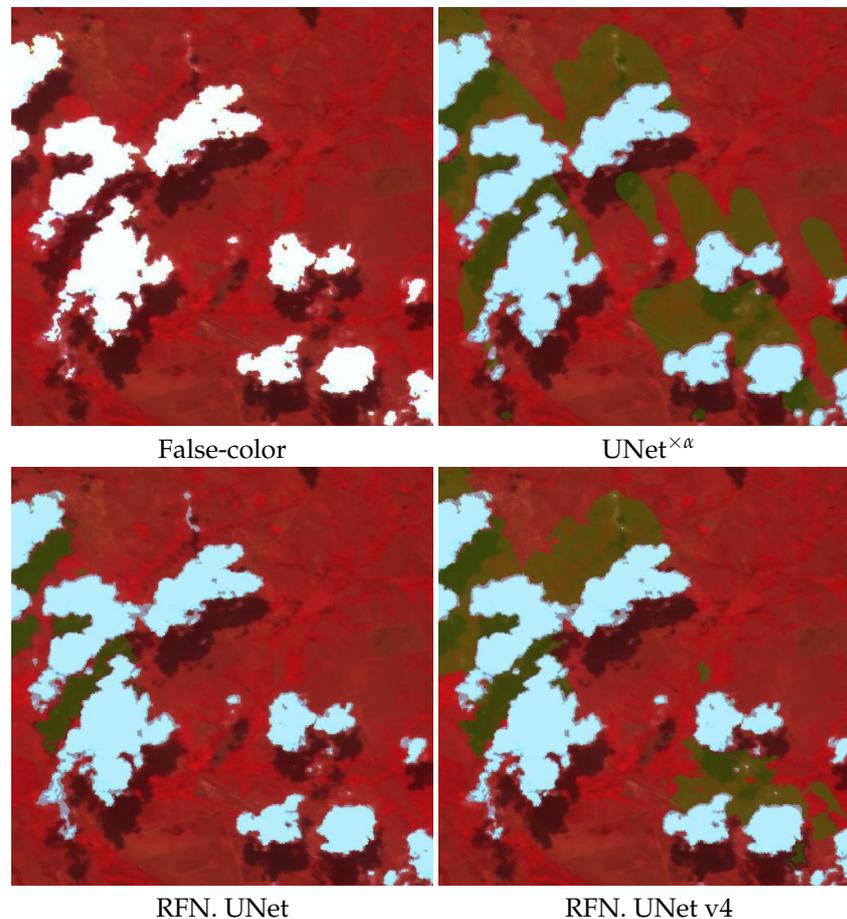


Figure 1. Illustration of a typical segmentation patch. The typical patch is drawn from GF1WV1.16m.2014267223114.21LTD.FAGUO1.SR.tiff, the false-color image (bands 4, 3, and 2) and the segmentation proposals of which, yielded, respectively, by UNet^α, by our Refined UNet, and by v4, are presented. It can be seen that the edge-refined labels of cloud and shadow regions are generated both by Refined UNet and by v4, in comparison with UNet^α giving region-accurate labels.

The rest of the paper is organized as follows. Section 2 investigates some relevant research. Section 3 revisits the implementation of our Refined UNets and the customization of the band-consistent datasets. Section 4 presents the experiments, the results, and the corresponding findings. Section 5 concludes this paper.

2. Related Work

It is reported that fundamental and advanced learning-based segmentation methods are gradually introduced to the remote sensing applications of pixel-wise classification. The fully convolutional network (FCN) [10] is known as a convolutional neural network-based segmentor initiating the neural image segmentation, while U-Net [5] is widely used in multiple visual-related applications because of its lightweight yet effective architecture. Various advances are proposed for the purpose of improving the segmentation efficacy, including SegNet [6], Bayesian SegNet [11], RefineNet [12], PSPNet [13], FastFCN [14], and the DeepLab series [7,15–17]. Typical solutions focus (1) on considering the scales of receptive fields [5,18–21], (2) on reusing features captured by previous layers [5,6,11–13], and (3) on introducing gradient discrepancies [22]. Such solutions definitely improve the segmentation performance in terms of global scores, such as accuracy and mIoU, even though they focus more on reaching the accurate performance of semantic regions instead

of pixel-precise segmentation. Alternatively, boundary-aware techniques particularly consider the segmentation performance surrounding edges (1) by introducing the dedicated block perceiving edges [23–26], (2) by introducing a particular term added to the objective function [27], and (3) by improving segmentation performance in an iterative way [28–30]. Such boundary-aware solutions focus more on the iterative improvement of the resolutions of the segmentation proposals and on the class assignments of the newborn pixels classified by the dedicated block, and attempt to improve the performance by introducing the images of the original resolutions. Naturally, they achieve a substantially edge-precise improvement but thereby introduce the critical dependencies of the pretrained modules, ignore insignificant objects, or have extra computational cost, and they also suffer from the absence of the edge features of the original inputs and from building explicit edge-sensitive formulations. Standalone or plug-in edge-sensitive blocks, for example, CRFs [7,31–35], are also thoroughly investigated to globally optimize refinement by building probabilistic graphical models with contrast-sensitive bilateral terms. Edge-preserving filters, such as the guided filter [36], are investigated to yield edge refinement. CRFasRNN [37] and the learnable guided filter [38] initiate producing segmentation proposals together with online edge refinement by presenting innovative network architectures, yet the computational cost or the property of multiple pixel classification can be further improved. Such advances significantly contribute to natural image segmentation and provide a prototypical framework for object segmentation of remote sensing images.

In addition to the aforementioned pipelines or end-to-end models, we wish to achieve end-to-end, edge-precise cloud and shadow segmentation by presenting multiple variants of Refined UNet [2,39–41], which adopt a pipeline or end-to-end models comprising a trainable UNet backbone and a plug-in CRF module. Such methods are reported to achieve edge-refined segmentation for cloud and shadow detection, in terms of our comprehensive experiments on our L8 dataset. We further initiate an attempt at the deployment of our Refined UNet on the target GF1 dataset, given the well-labeled L8 dataset as the source.

3. Instantiating Refined UNets on the Custom GF1 Dataset

We begin our efficacy investigation of the GF1 segmentation by introducing the model deployment of our Refined UNets. We first revisit the forward procedure of Refined UNets and then introduce the band selection of the given GF1 dataset.

3.1. Inference of Refined UNets Revisited

We first briefly revisit the formulations of the inference procedure of our Refined UNets, including Refined UNet and v4. Given a multi-spectral image \mathcal{I} comprising HW pixels $\{I_i\}$, where H and W represent the height and the width of the image, $I_i = [I_{i1}, I_{i2}, \dots, I_{ic}]^T$, and $i = 1, \dots, HW$, the target of the pixel-wise classification is to find a potential label assignment for each pixel, each label proposal of which is from the candidate target set \mathcal{T} comprising N label targets $\{t_n\}$. As presented in our previous work, we first give a coarse-grained segmentation proposal of the regions of interest using the UNet backbone, and then the spatial performance of the regions within the proposal is refined by the subsequent CRF in either a global or a local way. Specifically, the logits \mathbf{O} of the inference of UNet f is denoted by

$$\mathbf{O} = f(\mathcal{I}; \theta) \quad (1)$$

where θ is the set of trainable parameters, either pretrained or updated by the gradient descent optimization. The logits \mathbf{O} can be regarded as the unnormalized probabilities of the label assignment, in terms of the objective function, and therefore the actual probabilities of the most likely label assignment are given using the softmax function, in the form

$$P_0(x_i = t_n) = \frac{1}{Z_i} \exp(o_{in}) \quad (2)$$

where the normalization term Z_i is given by

$$Z_i = \sum_n \exp(o_{in}). \quad (3)$$

We then have the coarse-grained prediction of the label assignment of pixel i , given by

$$x_i^* = \arg \max_{t_n} P_0(x_i = t_n). \quad (4)$$

Spatial edge-refinement is formulated by a CRF $(\mathbf{X}, \mathcal{I})$ characterized by a Gibbs distribution in the form

$$P(\mathbf{X}|\mathcal{I}) = \frac{1}{Z(\mathcal{I})} \exp(-E(\mathbf{x}|\mathcal{I})) \quad (5)$$

and the corresponding Gibbs energy within is given by

$$E(\mathbf{x}) = \sum_i \psi_u(x_i) + \sum_i \sum_{j>i} \psi_p(x_i, x_j) \quad (6)$$

where the unary potential $\psi_u(x_i)$ of the probable label t_n is given by

$$\psi_{u,n}(x_i) = -\log(P_{GT} \cdot P_0(x_i = t_n)). \quad (7)$$

and the pairwise potentials ψ_p of the probable label t_n have the form

$$\psi_{p,n}(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^K w^{(m)} k^{(m)}(f_i, f_j). \quad (8)$$

Considering the gain of saving computational cost, together with the concatenation of the coarse-grained segmentation and edge refinement to form an end-to-end segmentation solution, we choose to evaluate the approximate probabilities $Q(\mathbf{X})$ instead of the exact probabilities $P(\mathbf{X})$ by using the mean-field approximation method to minimize the KL-divergence $D(Q||P)$ such that we will have the maximum a posteriori (MAP) estimation of the label assignment, in the form

$$x_i^{**} = \arg \max_{t_n} Q(x_i = t_n). \quad (9)$$

Specifically, $Q(\mathbf{X})$ is initialized by

$$Q(x_i) = \frac{1}{Z_i} \exp\{-\phi_u(x_i)\} \quad (10)$$

where the normalization term Z_i is given by

$$Z_i = \sum_n \exp\{-\phi_{un}(x_i)\}. \quad (11)$$

Then, $Q(\mathbf{X})$ is given using an iterative update equation, in the form

$$Q(x_i) \leftarrow \frac{1}{Z_i} \exp\left(-\psi_u(x_i) - \mu \sum_{m=1}^K w^{(m)} \sum_{j \neq i} k^{(m)}(f_i, f_j) Q(x_j)\right). \quad (12)$$

Finally, the refinement of the label assignment is given by (9) as discussed earlier.

Noting that we have discussed such formulations in detail earlier and wish to focus on the discussion of the GF1 data, it is then recommended to refer to our prior reports [2,39] for more formulation details and to refer to our open-access implementation of Refined UNet v4, available at <https://github.com/92xianshen/refined-unet-v4> (accessed on 24 January 2023). In addition, the end-to-end Refined UNet v4 performs edge-refined

segmentation on each proposal tile of the input images, while the pipeline of Refined UNet applies global edge refinement to the segmentation proposal of full resolution.

3.2. Details Regarding Model Deployment

The L8 data has general, region-accurate QA bands indicating the numerical possibility of the predefined classes, whereas the GF1 data only has four spectral bands of B, G, R, and NIR. The custom training and inference on the GF1 dataset are thus currently problematic due to the lack of such labels. More generally, we wish to partially address the GF1 segmentation problem by considering a hypothesis that a segmentor which is sufficiently trained on the well-labeled samples of common bands drawn from a source dataset can be directly applicable to the custom, band-consistent test cases from a target set. This can be specifically performed by restricting our attention to the common bands shared both by the well-labeled L8 dataset and by the label-free GF1 set: training samples comprising common bands, together with well-defined labels, can be drawn from the L8 set and the corresponding QA references, and then a well-trained segmentor can be directly applied to the band-consistent test cases from the GF1 set. In particular, the fourth, third, and second bands (4, 3, and 2), also known as the false-color bands, of both the L8 and the GF1 data are selected to form the band-consistent inputs, the region-accurate labels are derived from the QA references, the training can be run on the custom L8 set together with the QA references, and the inference can subsequently be performed on such a dataset of the common bands due to inputs of band consistency and well-training. Our implementations are built in part using the TensorFlow framework [42] and in part using PyDenseCRF (<https://github.com/lucasb-eyer/pydensecrf> (accessed on 7 December 2022)) and in part using our prior v4 implementation available at <https://github.com/92xianshen/refined-unet-v4> (accessed on 24 January 2023).

4. Experiments and Discussions

We finally investigate the inference performance of the Refined UNet instances by conducting comprehensive experiments, including visual comparisons, ablation experiments regarding bilateral manipulations, and explorations of critical hyperparameters within CRF.

4.1. Experiment Setup

We briefly introduce the experiment setup regarding our GF1 edge-refined segmentation. The source L8 dataset is inherited from our prior research [2,39] while the test cases from the target GF1 dataset are:

- GF1WV4.16m.2020271224917.20LQQ.FAGUO1.SR.tiff,
- GF1WV1.16m.2014127221306.21KVB.FAGUO1.SR.tiff,
- GF1WV1.16m.2014164221646.21KVB.FAGUO1.SR.tiff,
- GF1WV1.16m.2014267223114.21LTD.FAGUO1.SR.tiff,
- GF1WV2.16m.2017241224210.21KVB.FAGUO1.SR.tiff,
- GF1WV2.16m.2020262223057.20LQQ.FAGUO1.SR.tiff,
- GF1WV4.16m.2014128223630.21KVB.FAGUO1.SR.tiff,
- GF1WV4.16m.2017352224828.21KXB.FAGUO1.SR.tiff,
- GF1WV4.16m.2019241223346.21KXB.FAGUO1.SR.tiff,
- GF1WV4.16m.2020261220853.21KZB.FAGUO1.SR.tiff.

Data information is given by splitting up the file name by the dots. In particular, the third fraction, formatted as YYYYDDHMMSS, indicates Julian date, hours, minutes, and seconds of acquisition, and the fourth fraction represents the geocoordinate of the Military Grid Reference System (MGRS). We would like to use an identifier *date.grid* to represent a particular test case. Please kindly refer to the relevant research [8] for more details regarding the GF1 satellite as we wish to restrict our attention to the segmentation problem. In addition, the experiment implementations are derived from our prior Refined UNet and

v4, in which the secondary hyperparameters are also inherited from our prior research [2,39] while the principal and the critical hyperparameters are given in the later discussions.

4.2. Visual Efficacy of Edge-Refining Techniques

We first report the inference performance of our Refined UNet instances deployed on the custom GF1 datasets in terms of the visual results. Please note that our Refined UNets are customized to facilitate edge-refined segmentation, in which the edges of the region labels are spatially aligned to those of the regions of interest, and consequently the region shapes of the labels are sufficiently analogous to those of the regions of interest as well. We therefore first evaluate the edge-refined performance from both the global and the local perspectives. Please note that in the context of our evaluations, the visual results of the full images are regarded as global evaluations whereas the zoom-in patches are local evaluations. Global results from the backbone, Refined UNet, and v4 are presented in Figures 2 and 3, in which *False-color* represents the false-color illustrations, $UNet^{\times\alpha}$ represents the UNet backbone, *RFN. UNet* represents our Refined UNet, and *RFN. UNet v4* represents our Refined UNet v4. As illustrated, it can be seen that considerable cloud and shadow regions are generally detected and correctly labeled, and the boundaries between the valid (cloud, shadow, and background pixels) and invalid (fill values) regions are sufficiently precise. In addition, the tile gaps between patches are effectively neutralized by Refined UNet, attributed to its global processing. In particular, the UNet backbone is trained using dynamically adaptive weights and it can capture more shadow regions compared with the original UNet, in terms of our previous report. These results demonstrate that our Refined UNets can achieve cloud and shadow detection compared with the UNet variant, attributed to the significant segmentation effectiveness of the involved models given only region-accurate training samples.

We further zoom in on typical patches to observe the performance of edge refinement, illustrated in Figures 4 and 5. Fortunately, significant edge-refined achievements can be seen in such illustrations: the backbone can give generally coarse-grained labels for the regions of interest and thus can still achieve region-accurate segmentation, whereas it can be also seen that the edges of the label regions are not spatially aligned to those of the regions of interest and therefore the label edges are not sufficiently precise. Instead, the edges of the cloud labels, given by our Refined UNet instances, are spatially aligned to those of the potential cloud regions, and accordingly, these segmentation proposals are regarded as edge-precise results. Such achievements of cloud detection suggest that our Refined UNet instances can achieve edge-refined segmentation for cloud regions and can be attributed to the CRF module embedded in our Refined UNet instances: the segmentation proposals are significantly refined by the CRF modules attached to the UNet backbone after the UNet backbone yields the coarse-grained location of cloud and shadow regions. This explanation is also supported by the visual differences between the proposals given by the UNet backbone and our Refined UNet instances as the segmentation proposals of Refined UNets are degenerated to those of the UNet backbone when the CRF module is disabled, as discussed later. Unfortunately, it is observed that edge-refined segmentation is partially achieved in shadow detection: some typical shadow regions are labeled and are refined, but some are missing; some labels grow arbitrarily toward the edges of irrelevant objects. However, shadow detection is reported as a challenging, tricky task, and such growth attempts to perform edge alignment and thus is consistent with our edge-precise purpose. We would like to further explore a solution to edge-precise shadow detection. Consequently, we conclude that our Refined UNet instances can successfully perform the inference of edge-refined segmentation for cloud and shadow regions of the custom GF1 data, in terms of the aforementioned, comprehensive comparisons.

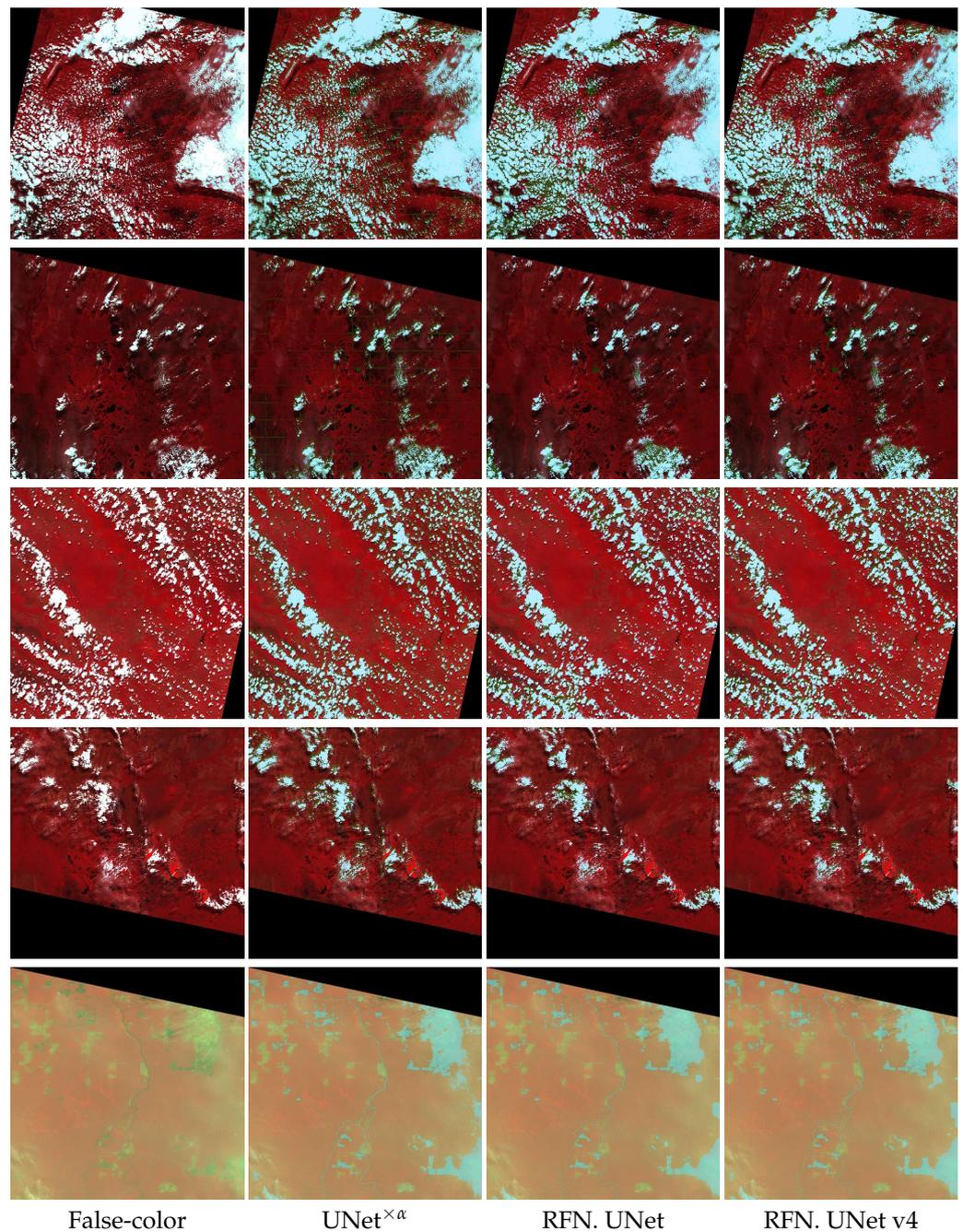


Figure 2. Illustrations of segmentation proposals from a global perspective (1). Presented are the results of 2014127221306.21KVB, 2014164221646.21KVB, 2014267223114.21LTD, 2017241224210.21KVB, and 2020262223057.20LQQ, respectively, including the false-color images (bands 4, 3, and 2) and the segmentation proposals of full resolution, yielded, respectively, by UNet^α, by our Refined UNet, and by v4. It can be seen that considerable cloud and shadow regions are generally detected and correctly labeled, confirming the effectiveness of the involved methods for cloud and shadow detection.

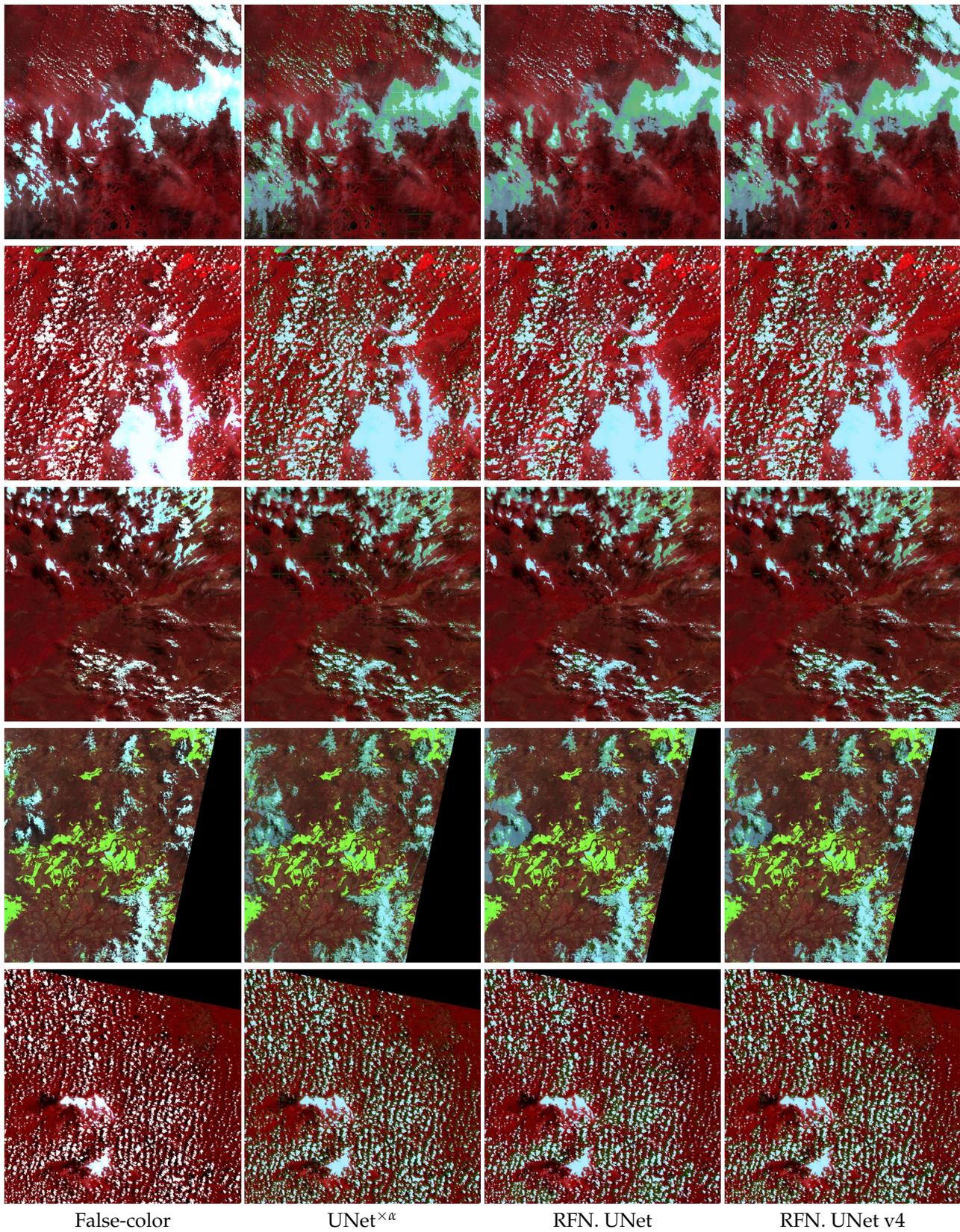


Figure 3. Illustrations of segmentation proposals from a global perspective (2). Presented are the results of 2014128223630.21KVB, 2017352224828.21KXB, 2019241223346.21KXB, 2020261220853.21KZB, and 2020271224917.20LQQ, respectively.

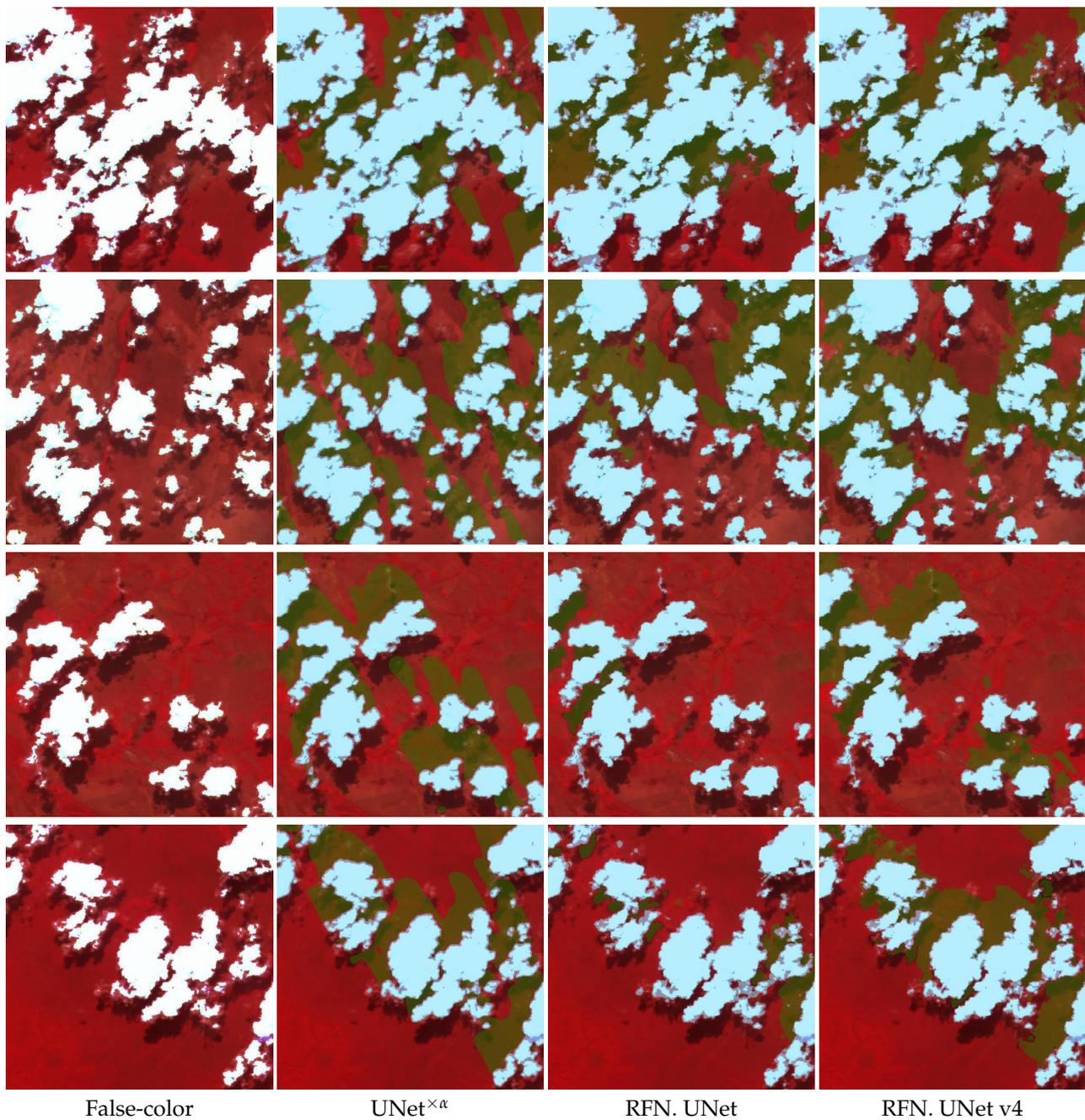


Figure 4. Illustrations of segmentation proposals from a local perspective (1). The typical patches are drawn from GF1WV1.16m.2014267223114.21LTD.FAGUO1.SR.tiff, the false-color images (bands 4, 3, and 2) and the segmentation proposals of which, yielded respectively by UNet^α, our Refined UNet, and v4, are presented. It can be seen that the edge-refined labels of cloud and shadow regions are generated by both Refined UNet and v4, in comparison with UNet^α giving region-accurate labels. Such illustrations further support the edge-refined efficacy of our sophisticated methods.

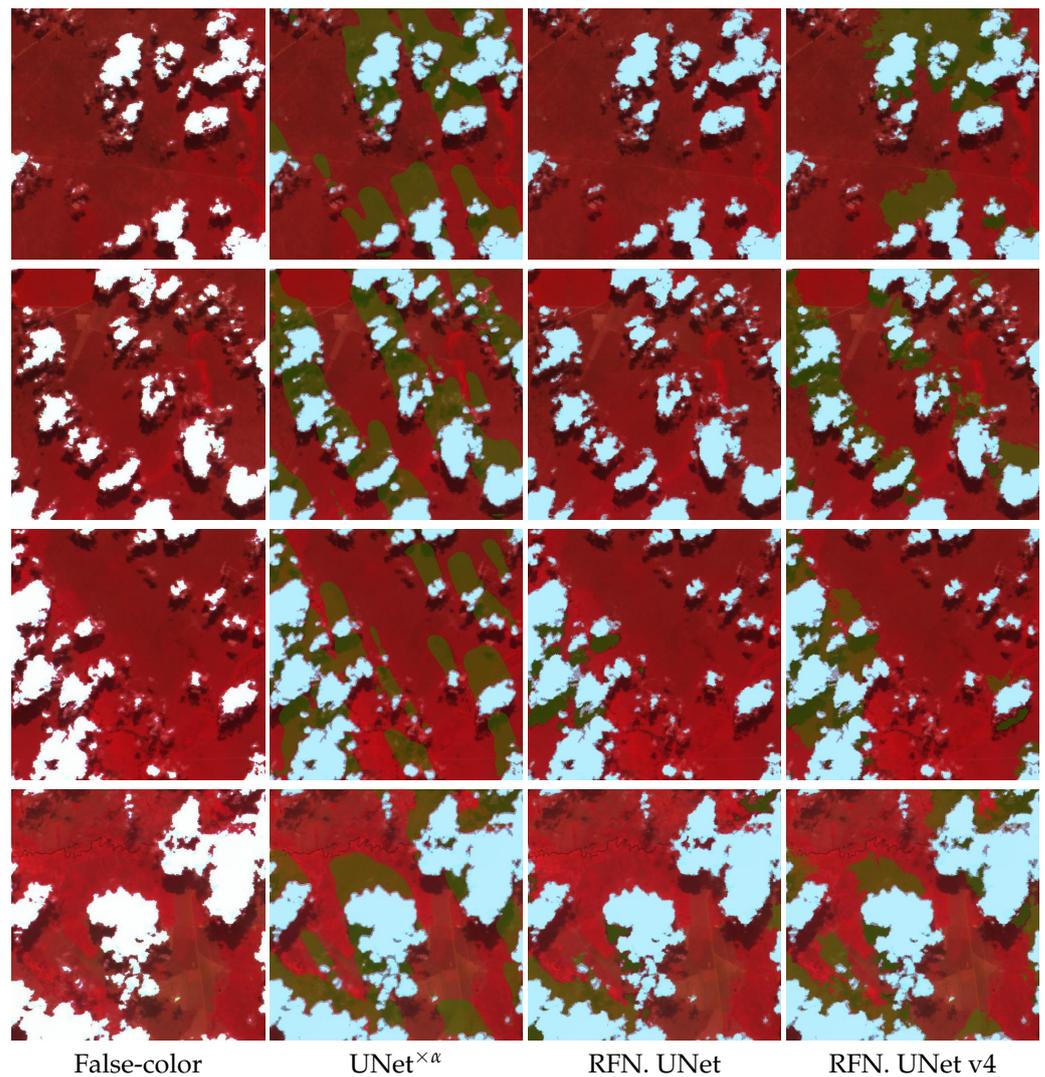


Figure 5. Illustrations of segmentation proposals from a local perspective (2).

4.3. Ablation Study Regarding Edge Refinement

We further explore the specific effects of each critical module, including the CRF module and the bilateral message-passing manipulation within. In general, the UNet backbone can provide coarse-grained locations of cloud and shadow regions, while the CRF module can perform edge refinement, given the unary potentials supplied by the coarse-grained locations. We currently verify such a hypothesis using the ablation experiment, in which the CRF module and the bilateral message-passing manipulation are disabled in turn. Corresponding results are illustrated in Figures 6 and 7, in which *Backbone* represents that the CRF module is disabled and *W/o bi-step* represents that the bilateral manipulation is disabled. It also can be seen that in the proposals of v4 without the bilateral manipulation, the region edges are not sufficiently tight and sharp, compared with those of v4; we suggest that the Refined UNet v4 without the bilateral manipulation degenerates to the UNet backbone in terms of the visual comparisons. We therefore conclude that the bilateral message-passing manipulation plays a critical role in the context of edge refinement.

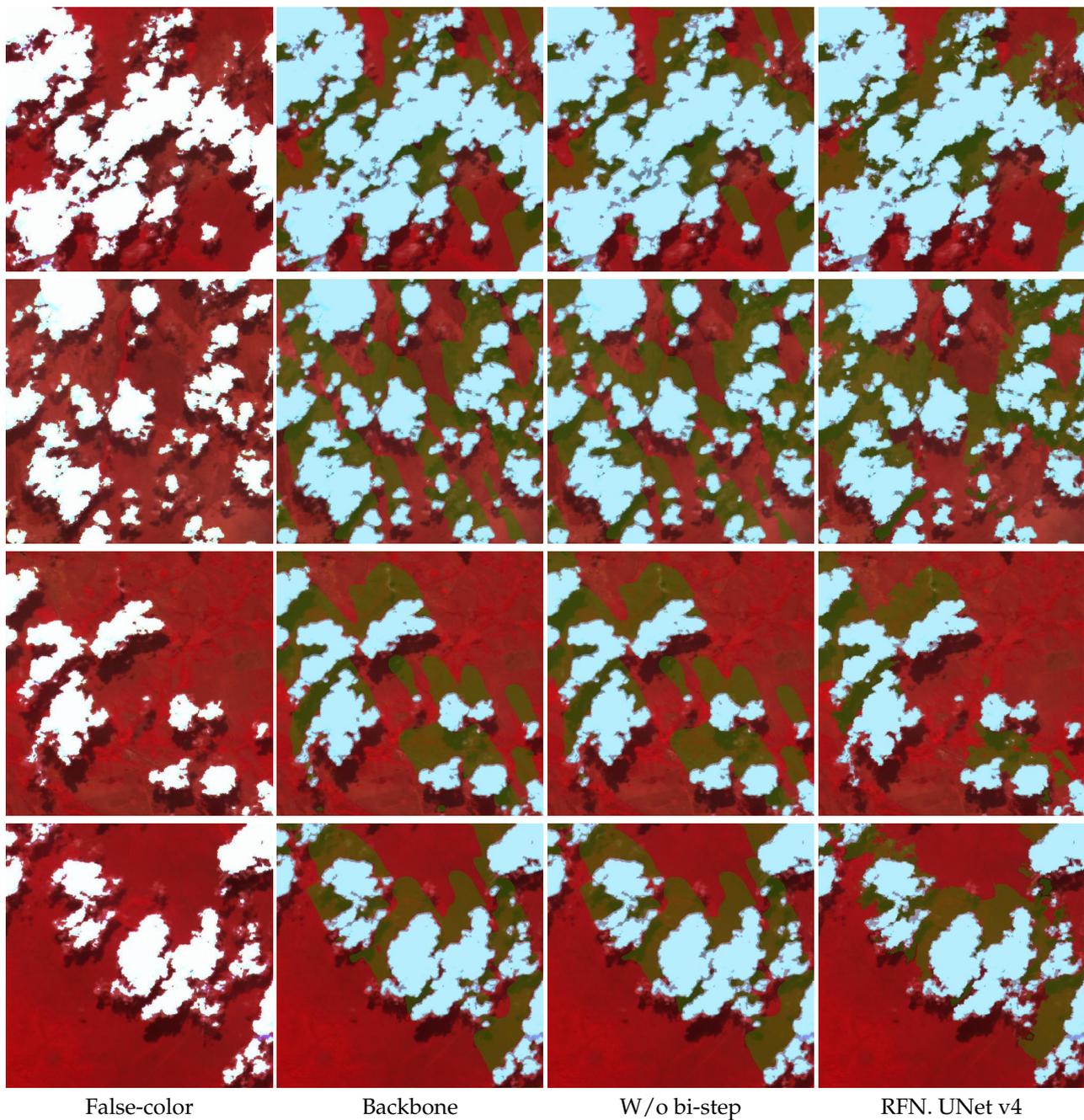


Figure 6. Illustrations of the effects of the critical modules (1). The typical patches are drawn from GF1WV1.16m.2014267223114.21LTD.FAGUO1.SR.tiff. It is observed that the proposals given by the backbone and the full model without the bilateral manipulation are visually close but those given by the full model are significantly refined on the edges. Such results confirm that the bilateral manipulation within the CRF module is critical to edge refinement.

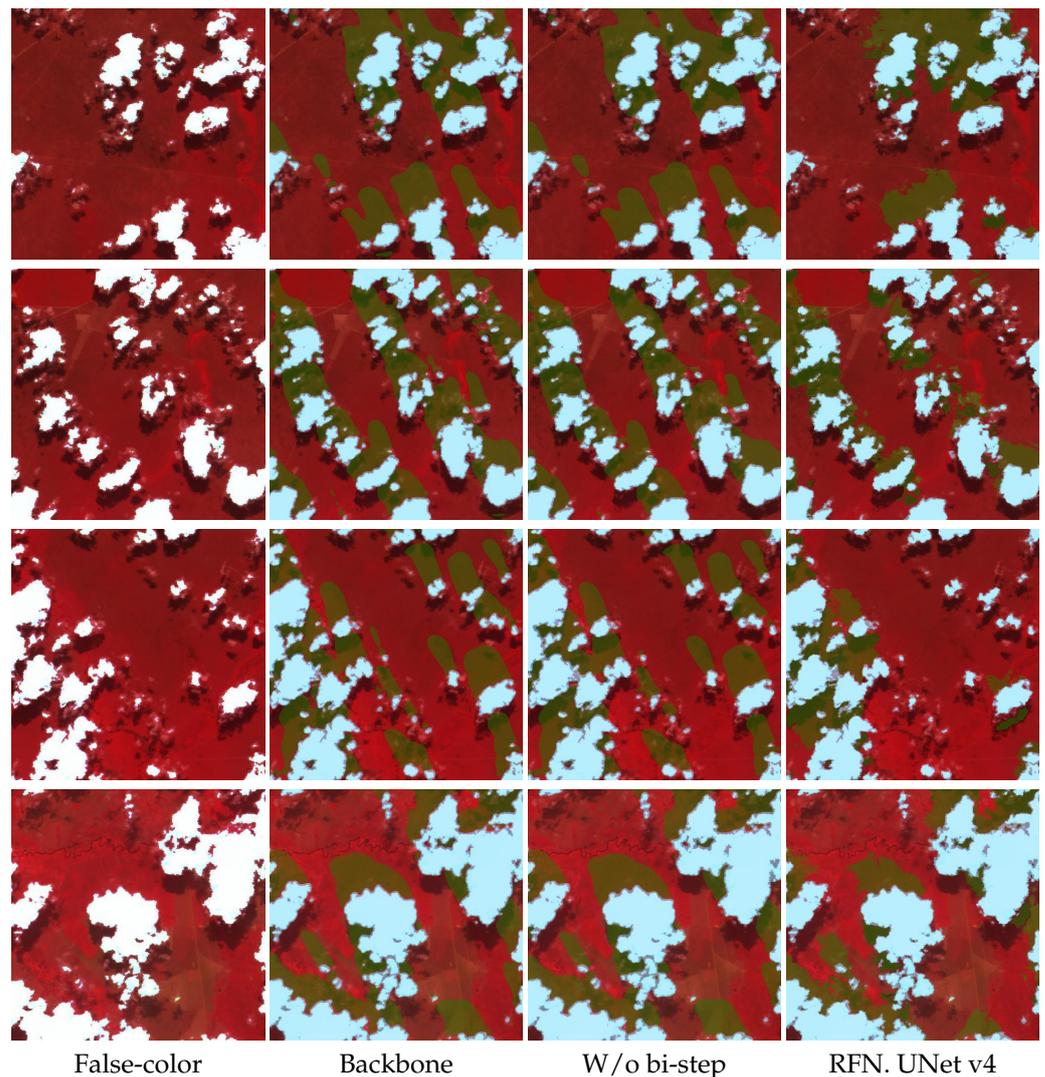


Figure 7. Illustrations of the effects of the critical modules (2).

4.4. Visual Effect of Critical Hyperparameters

We then investigate the hyperparameter effect on the edge-precise segmentation performance by observing the segmentation proposals yielded by the models with different hyperparameters. Naturally, θ_α and θ_β are sufficiently critical to the edge-refined segmentation performance in terms of our previous report: the CRF module with sufficiently small θ_β gives proposals with tight and sharp edges, while θ_α has not as significant effect as θ_β does. Such a hypothesis is confirmed in terms of our experiments, illustrated in Figure 8 and 9: the edges of label regions are tighter, sharper, and sufficiently close to those of the regions of interest, given the CRF module with a small θ_β of 0.03125. By contrast, the effect of θ_α is not so significant as θ_β is, but the visual difference is still observed when θ_α drops to 10. We attribute such results to the nature of θ_α and θ_β introduced by the edge-preserving filters and the fully-connected CRF: θ_β controls the sensitivity of the filter with respect to the color intensities while θ_α affects the range of passing messages, therefore the value of θ_β has visually significant effects in our case while the results are not sensitive with θ_β because of its relatively large value.



Figure 8. Illustrations of the effects of the critical hyperparameters (1). The typical patches are drawn from GF1W1V1.16m.2014267223114.21LTD.FAGUO1.SR.tiff. θ_α and θ_β are consistently reported as critical hyperparameters with respect to the extent of edge refinement, and θ_β is relatively significant in comparison with θ_α . The illustrations presented above show the differences governed by such multiple hyperparameters, in which θ_β can significantly govern the sensitivity of refining the edges of regions of interest whereas θ_α plays a secondary role. Such findings are consistent with our previous reports.

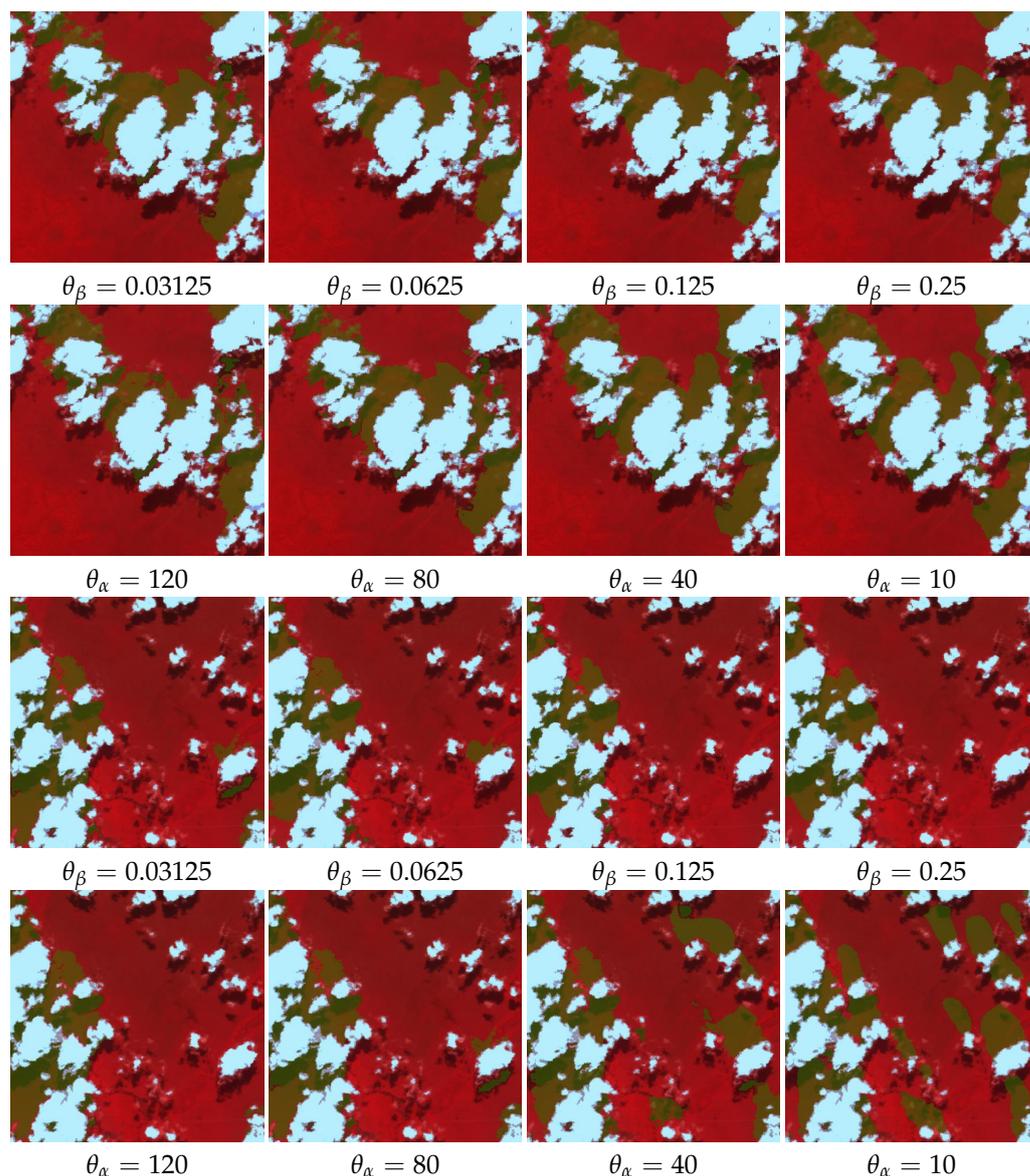


Figure 9. Illustrations of the effects of the critical hyperparameters (2).

4.5. Time Consumption in Practice

We finally report the practical efficiency of our Refined UNets in terms of time consumption. The actual time costs of Refined UNet and v4 are 254.11 ± 6.84 s and 276.12 ± 11.31 s, respectively, which suggests that such methods are computationally expensive in comparison with their lightweight UNet backbone. In addition, we also see that the efficiency is dramatically related to the critical hyperparameters within the CRF module. The actual time costs of the inference of a particular case with respect to θ_α s of 120, 80, 40, and 10 are 274.77 s, 281.11 s, 312.64 s, and 1410.53 s, respectively. The differences in such time costs are possibly attributed to the sparsity of bilateral sampling. However, the actual time costs of the inference of a particular case with respect to θ_β s of 0.03125, 0.0625, 0.125, and 0.25 are 281.11 s, 265.26 s, 256.29 s, and 254.84 s, respectively. Such differences are not sufficiently significant in comparison with those of θ_α even if θ_β plays a critical role in affecting visual performance. We conclude that activating edge refinement should be determined in terms of usage, considering the trade-off between performance and time consumption.

5. Conclusions

In this paper, we wish to partially address the cloud and shadow segmentation problem of the label-free GF1 dataset by considering the hypothesis that the segmentor can be applicable to the target dataset after being trained on well-labeled, band-consistent source samples. We accordingly present a straightforward solution by treating the L8 set as the source and by selecting the fourth, the third, and the second bands to construct band-consistent samples. We then apply our prior Refined UNet and v4 to such a solution for the purpose of satisfying the requirement of edge-precise cloud and shadow segmentation, making it possible to give rise to edge-refined segmentation proposals. We finally verify the effectiveness of such a band-consistent hypothesis and edge-refined solution by performing a relatively comprehensive investigation, including visual comparisons, ablation experiments regarding bilateral manipulations, explorations of critical hyperparameters within CRF, and time consumption in practice. Our main findings are as follows. Specifically, the visual comparisons show that our Refined UNet and v4, sufficiently trained on the well-labeled, band-consistent L8 samples, are able to perform edge-refined cloud and shadow segmentation on the custom GF1 data. The bilateral message-passing step is confirmed to be critical to the edge-refined performance in terms of the ablation experiments. θ_β plays a more important role in governing the edge-refined performance whereas the model with sufficiently large θ_β is not sensitive to edges, in terms of the hyperparameter tests. The instances of Refined UNets are computationally expensive even if they can provide a significantly edge-refined segmentation; we therefore have to make a trade-off between the quantitative performance and time consumption of the inference. We would like to proceed with improving the segmentation performance for the shadow regions in the future as they are a tricky part in the context of cloud and shadow detection.

Author Contributions: Conceptualization, L.J. and P.T.; Funding acquisition, P.T. and Z.Z.; Methodology, L.J. and M.Z.; Supervision, P.T. and Z.Z.; Writing—original draft, L.J.; Writing—review & editing, P.T. and Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the National Key Research and Development Program of China grant number 2019YFE0197800, in part by Engineering Research Center of Integration and Application of Digital Learning Technology, Ministry of Education grant number 1221022, in part by the Foreign Expert Programs of Ministry of Science and Technology of China grant number DL2021123002L, and in part by the Fundamental Research Funds for the Central Universities of China grant number 2022XJJD02.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to appreciate the open-source implementation of the fully-connected conditional random field *PyDenseCRF* (<https://github.com/lucasb-eyer/pydensecrf> (accessed on 7 December 2022)) as it is a part of Refined UNet. The authors also would like to appreciate Changmiao Hu and Lianzhi Huo as they provide test data used in the experiments. Other implementations are built using the TensorFlow framework (<https://tensorflow.google.cn/> (accessed on 7 December 2022)).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chai, D.; Newsam, S.; Zhang, H.K.; Qiu, Y.; Huang, J. Cloud and cloud shadow detection in Landsat imagery based on deep convolutional neural networks. *Remote Sens. Environ.* **2019**, *225*, 307–316. [CrossRef]
2. Jiao, L.; Huo, L.; Hu, C.; Tang, P.; Zhang, Z. Refined UNet V4: End-to-End Patch-Wise Network for Cloud and Shadow Segmentation with Bilateral Grid. *Remote Sens.* **2022**, *14*, 358. [CrossRef]
3. Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* **2016**, *185*, 46–56. [CrossRef] [PubMed]
4. Lin, D.; Xu, G.; Wang, X.; Wang, Y.; Sun, X.; Fu, K. A remote sensing image dataset for cloud removal. *arXiv* **2019**, arXiv:1901.00600.
5. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241.

6. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)] [[PubMed](#)]
7. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)] [[PubMed](#)]
8. Li, Z.; Shen, H.; Li, H.; Xia, G.; Gamba, P.; Zhang, L. Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery. *Remote Sens. Environ.* **2017**, *191*, 342–358. [[CrossRef](#)]
9. Khoshboresh-Masouleh, M.; Shah-Hosseini, R. A deep learning method for near-real-time cloud and cloud shadow segmentation from gaofen-1 images. *Comput. Intell. Neurosci.* **2020**, *2020*, 8811630. [[CrossRef](#)] [[PubMed](#)]
10. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 640–651. [[CrossRef](#)] [[PubMed](#)]
11. Kendall, A.; Badrinarayanan, V.; Cipolla, R. Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding. In Proceedings of the British Machine Vision Conference (BMVC), London, UK, 4–7 September 2017; pp. 57.1–57.12.
12. Lin, G.; Milan, A.; Shen, C.; Reid, I. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1925–1934.
13. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
14. Wu, H.; Zhang, J.; Huang, K.; Liang, K.; Yu, Y. FastFCN: Rethinking dilated convolution in the backbone for semantic segmentation. *arXiv* **2019**, arXiv:1903.11816.
15. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
16. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
17. Wang, P.; Chen, P.; Yuan, Y.; Liu, D.; Huang, Z.; Hou, X.; Cottrell, G. Understanding convolution for semantic segmentation. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1451–1460.
18. Chen, L.C.; Yang, Y.; Wang, J.; Xu, W.; Yuille, A.L. Attention to scale: Scale-aware semantic image segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 3640–3649.
19. Farabet, C.; Couprie, C.; Najman, L.; LeCun, Y. Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1915–1929. [[CrossRef](#)] [[PubMed](#)]
20. Mostajabi, M.; Yadollahpour, P.; Shakhnarovich, G. Feedforward semantic segmentation with zoom-out features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3376–3385.
21. Zhang, H.; Dana, K.; Shi, J.; Zhang, Z.; Wang, X.; Tyagi, A.; Agrawal, A. Context encoding for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 7151–7160.
22. Zhang, H.; Patel, V.M. Densely connected pyramid dehazing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 3194–3203.
23. Kirillov, A.; Wu, Y.; He, K.; Girshick, R. Pointrend: Image segmentation as rendering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 9799–9808.
24. Yuan, Y.; Xie, J.; Chen, X.; Wang, J. Segfix: Model-agnostic boundary refinement for segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 489–506.
25. Zhou, P.; Price, B.; Cohen, S.; Wilensky, G.; Davis, L.S. Deepstrip: High-resolution boundary refinement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10558–10567.
26. Tang, C.; Chen, H.; Li, X.; Li, J.; Zhang, Z.; Hu, X. Look closer to segment better: Boundary patch refinement for instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 13926–13935.
27. Borse, S.; Wang, Y.; Zhang, Y.; Porikli, F. Inverseform: A loss function for structured boundary-aware segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 5901–5911.
28. Li, K.; Hariharan, B.; Malik, J. Iterative instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 3659–3667.
29. Cheng, H.K.; Chung, J.; Tai, Y.W.; Tang, C.K. Cascadepsp: Toward class-agnostic and very high-resolution segmentation via global and local refinement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 8890–8899.
30. Huynh, C.; Tran, A.T.; Luu, K.; Hoai, M. Progressive semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 16755–16764.

31. Lin, G.; Shen, C.; Van Den Hengel, A.; Reid, I. Efficient piecewise training of deep structured models for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 3194–3203.
32. Krähenbühl, P.; Koltun, V. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. In Proceedings of the Advances in Neural Information Processing Systems, Granada, Spain, 12–14 December 2011; pp. 109–117.
33. Krähenbühl, P.; Koltun, V. Parameter learning and convergent inference for dense random fields. In Proceedings of the International Conference on Machine Learning (ICML), Atlanta, GA, USA, 17–19 June 2013; pp. 513–521.
34. Liu, Z.; Li, X.; Luo, P.; Loy, C.C.; Tang, X. Semantic image segmentation via deep parsing network. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 11–18 December 2015; pp. 1377–1385.
35. He, X.; Zemel, R.S.; Carreira-Perpinán, M.A. Multiscale conditional random fields for image labeling. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, 27 June–2 July 2004; pp. II–695–II–702.
36. He, K.; Sun, J.; Tang, X. Guided Image Filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1397–1409. [[CrossRef](#)] [[PubMed](#)]
37. Zheng, S.; Jayasumana, S.; Romera-Paredes, B.; Vineet, V.; Su, Z.; Du, D.; Huang, C.; Torr, P.H. Conditional random fields as recurrent neural networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 11–18 December 2015; pp. 1529–1537.
38. Wu, H.; Zheng, S.; Zhang, J.; Huang, K. Fast end-to-end trainable guided filter. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 1838–1847.
39. Jiao, L.; Huo, L.; Hu, C.; Tang, P. Refined UNet: UNet-based refinement network for cloud and shadow precise segmentation. *Remote Sens.* **2020**, *12*, 2001. [[CrossRef](#)]
40. Jiao, L.; Huo, L.; Hu, C.; Tang, P. Refined UNet V2: End-to-End Patch-Wise Network for Noise-Free Cloud and Shadow Segmentation. *Remote Sens.* **2020**, *12*, 3530. [[CrossRef](#)]
41. Jiao, L.; Huo, L.; Hu, C.; Tang, P. Refined UNet v3: Efficient end-to-end patch-wise network for cloud and shadow segmentation with multi-channel spectral features. *Neural Netw.* **2021**, *143*, 767–782. [[CrossRef](#)] [[PubMed](#)]
42. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: [tensorflow.org](https://www.tensorflow.org) (accessed on 27 October 2020).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.