Response to Reviewer 3

The authors would like to express sincere appreciations to the reviewer for the thorough and constructive comments. We present a detailed response to the individual points raised by the reviewer. In this reply, we indicate the revised parts with blue color when directly quote them from the revised manuscript.

> **Concern # 0:** Compared to the previous submission, the authors have made substantial revisions based on the reviewers' comments.

Response: Many thanks for your positive and valuable comments that really improve the quality of our manuscript. We next present a point-by-point response as follows.

> **Concern # 1:** In the first-round review, the reviewer points out the novelty of this work is weak. Unfortunately, in the reply, the author does not defense the innovation of the work, but mentions the results of the method again, as well as the newly added comparison method.

Response: We thank the reviewer for the constructive comments. The novelty of this work is that, to the best of our knowledge, we are the first to develop semi-supervised image-to-image translation framework on SAR-optical image matching task. Besides, our framework indeed combines the benefits of Pix2pix (correct gray value assignment) and CycleGAN (good edge-preserving), and avoids the disadvantage of the two methods (i.e., Pix2pix obtains blur results; CycleGAN assigns wrong gray value to some features). Meanwhile, our framework is simple yet effective.

In the revised manuscript, we introduce more contributions compared to Pix2pix and CycleGAN and highlight that our framework is simple yet effective. We excerpt the major revisions as follows.

In Section-I. Introduction,

> "...In this work, we propose a simple yet effective framework for semi-supervised image-to-image translation, which merges two well-known supervised and unsupervised image-to-image translation methods, i.e., Pix2pix [1] and CycleGAN [2]. Our framework combines the benefits of Pix2pix (correct gray value assignment) and CycleGAN (good edge-preserving), and avoids the disadvantage of the two methods (i.e., Pix2pix obtains blur results; CycleGAN assigns wrong gray value to some features)..."

In Abstract,

> "...To this end, we combine the benefits of both supervised and unsupervised well-known image-to-image translation methods, i.e., Pix2pix and CycleGAN, and propose a simple yet effective semi-supervised image-to-image translation framework..."

In Section-I. Introduction,

> "...In this work, we propose a simple yet effective framework for semi-supervised image-to-image translation, which merges two well-known supervised and unsupervised image-to-image translation methods, i.e., Pix2pix [1] and CycleGAN [2]..."

In Section-IV. Conclusion,

> "...In this work, we proposed a simple yet effective semi-supervised image-to-image translation framework integrated by Pix2pix and CycleGAN...."

Concern # 2: It may be unfair for TCR which is chose as a comparison method. After all, TCR directly performs cross-domain translation, while the proposed method converts the cross-domain translation into the co-domain translation. The methods are completely different in nature.

Response: We thank the reviewer for the valuable comments. We agree with the reviewer that our framework has co-domain losses (cycle consistency loss), which TCR hasn't. Hence, we think this is also an advantage of our framework compared TCR.

In addition, SAR-optical image matching results in this work are based on both SAR-SAR and optical-optical matchings, i.e., TCR is trained twice for a certain experiment to obtain SAR-to-optical and optical-to-SAR translation models. In contrast, our framework is trained once a time to get the two translation models. Hence, we think the comparison between our framework and TCR is fair in our manuscript.

In the revised manuscript, we highlight the fairness of the comparison. We excerpt the major revisions as follows.

In Section-III-D. Comparison of Image Matching,

> "...Note that because SAR-optical image matching results in this work are based on both SAR-SAR and optical-optical matchings, Pix2pix and TCR are trained twice to obtain SAR-to-optical and optical-to-SAR translations...."

Concern # 3: Finally, this is not a huge problem, but the reviewer still wants to bring the authors to their attention. About the abbreviation (SPC-GAN) to the name of the proposed method, i.e., Semi-supervised-Pix2pix-Cycle-GAN. Although it is difficult to give a more appropriate name, I still think the current name is inappropriate. From the point of view of word formation, "Semi-supervised" denotes the learning strategy, that is to say, it is neither fully supervised nor fully unsupervised, but both which is the combination of a supervised method Pix2pix and an unsupervised method "CycleGAN". The hyphen between SPC and GAN tends to confuse readers into thinking that this is a new GAN method, namely SPC GAN, which is obviously not what the authors expect.

Response: We thank the reviewer for the constructive comments. We remove all the "SPC-GAN" from the revised manuscript. Please refer to our revised manuscript (the highlighted version) for more details.

Concern # 4: "number of qualied matchings" in the text of section III.D, it is written as NOQMs, while in Fig.7 and Fig.8, it is NQMs.

The use of parentheses is confusing between line10 to line 13 in page 2. Please recheck and correct them.

Response: We thank the reviewer for the constructive comments. We correct all the "NQMs" to "NOQMs" in the revised manuscript, and we correct the parentheses in page 2. We excerpt the major revisions as follows.

In Section II. Method,

> "...($\mathcal{A}_{\mathcal{X}} = \{\mathcal{A}_x^k\}_{k=1}^K \in \mathcal{X}$, $\mathcal{A}_{\mathcal{Y}} = \{\mathcal{A}_y^k\}_{k=1}^K \in \mathcal{Y}$) while given unaligned M SAR ($\mathcal{U}_{\mathcal{X}} = \{\mathcal{U}_x^i\}_{i=1}^M \in \mathcal{X}$) and N optical image pairs ($\mathcal{U}_{\mathcal{Y}} = \{\mathcal{U}_y^j\}_{j=1}^N \in \mathcal{Y}$). Hence, in each training iteration of the *semi-supervised version*, a pair of aligned and unaligned data are fed into the unsupervised and supervised modules, respectively..."

In Figures 7 and 8,

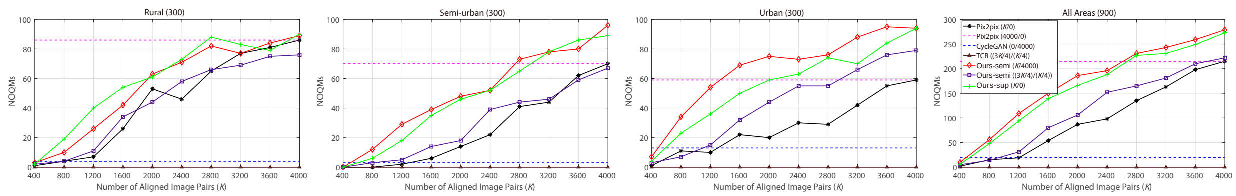

Figure 7: NOQMs for each method on 900 test image pairs as the amount of aligned training data varies.

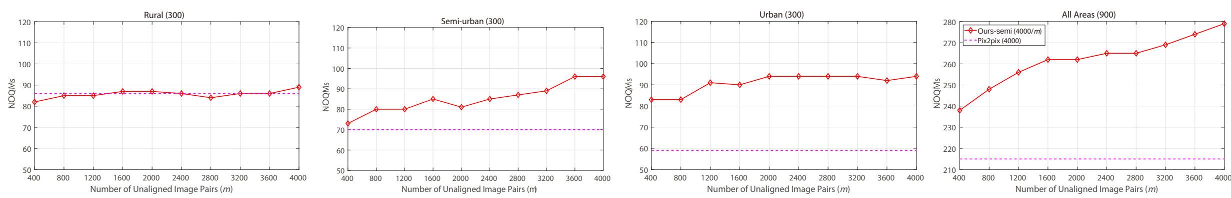

Figure 8: NOQMs for our framework on 900 test image pairs as the amounts of unaligned training data varies.

References

[1] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proc. CVPR*, pages 5967–5976, 2017.

[2] J. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. ICCV*, pages 2242–2251, 2017.