We gratefully acknowledge the constructive comments from the anonymous reviews and the editor to our submission "Unsupervised Progressive Learning for Large-scale Neuronal Population Reconstruction - TMI-2019-1442". We are glad to see that the merit of our work is appreciated. We carefully revised our manuscript under the guidance provided by reviewers. Here, we would like to address the concerns of the reviewers and the editor point by point.

[Comments from Editor]

All reviewers have expressed their interests in this work, but also have provided extensive comments that may eventually improve the quality of the papers. In particular, one of the reviewers pointed out the overlap of the MICCAI paper from the authors' group and provided new qualitative results. The authors are advised to address reviewers' comments carefully in the revision.

[Response] We made substantial revision according to the comments of the anonymous reviewers and editors. First, we rewrote the whole section of Section III-B "Ultra-scale Neuronal Population Reconstruction" with more detailed description and figures to explain our UltraNPR algorithm. Secondly, we abbreviate Section III-A "PLNPR for Robust Neuron Reconstruction" to reduce overlap with our MICCAI paper. Thirdly, a series of additional experiments were conducted to evaluate our UltraNPR algorithm and compare with state-of-the-art large-scale reconstruction algorithms including UltraTracer [Peng et al 2017] and MEIT [Wang et al 2018]. The added experiments demonstrate the effectiveness and robustness of our UltraNPR method.

[Comments from Reviewer #1]

[Q1-1] The main idea of the PLNPR is valuable, especially in terms of not requiring manual annotation. However, the main idea of PLNPR seems closer to being an adaptive enhancement of the input image to the tracing module, rather than a deep learning framework that uses unsupervised learning for the task of reconstruction. For this reason, the presentation of the idea is misleading. The "Unsupervised Progressive Learning" in the title gives an implication that the deep learning network is the driving force of the reconstruction and the training of the deep learning network is unsupervised. The deep learning network is technically trained with labels, which are generated by a conventional tracing method. The focus should be more on how the iteratively improving enhancement of the input image to the tracing module improves the quality of the reconstruction.

[R1-1] Thank you for your comments. We agree with you that the key concept of this paper is the iterative enhancement instead of unsupervised training. To avoid misunderstanding, we modified our title to "Neuronal population reconstruction from ultra-scale optical microscopy images via progressive learning". In order to evaluate the effectiveness of our PLNPR method on enhancing noisy and low-contrast images, we test seven neuron tracing methods on the same images enhanced by our method. The performance comparisons are shown in Fig. 12. For any of the seven neuron tracing methods, the reconstruction results from our enhanced images are much better than the results performed on the raw images under the same parameter settings.

[Q1-2] Some paragraphs should be abridged: avoid any unnecessary/ambiguous expressions: e.g. "it can be seen that…"

[R1-2] Thanks for pointing this out. We carefully revise our expression to make it more concise.

[Q1-3] It's difficult to see the UltraNPR algorithm as something other than a block-by-block reconstruction scheme with some slightly adaptable features. If there is any novel coherent principle behind it, it would be helpful to include a diagram or an algorithm that encapsulates the idea.

[R1-3] Yes, our UltraNPR follows a general block-by-block reconstruction scheme. The diagram of our UltraNPR method is shown in Fig. 2. The novelty in our UltraNPR is the block-propagation strategy by setting multiple neurite tips as pseudo-somas for dense neuronal population reconstruction, which is different from existing large-scale tracing methods that are designed for single-neuron tracing. Integrating PLNPR with an efficient block-wise tracing and fusion strategy, our UltraNPR is able to reconstruct dense neuronal populations from an ultra-scale OM images of a mouse brain slice.

We added comparison with two state-of-the-art large-scale neuron reconstruction methods, including UltraTracer [Peng et al 2017] and MEIT [Wang et al 2018] in Fig. 14 of our revised manuscript. Given an ultra-scale OM image, UltraTracer and MEIT fail to reconstruct individual neurons and trace a complete neuronal population in the image. In comparison, thanks to the signal enhancement by our PLNPR and block propagation strategy designed for dense neurons, our UltraNPR is more robust to reconstruct a more complete neuronal population with more accurate neurite assignment.

[Q1-4] The evaluation using the BigNeuron dataset should be revised. For testing on a different dataset, it is not fair to use a deep-learning-based method that is already trained on a previous dataset (VISoR-40). Furthermore, the method by Li et al. uses a APP2 as a main tracing method. If the progressive learning aspect of the currently proposed method is to be highlighted, it would be helpful to also compare using the PLNPR model using APP2.

[R1-4] Thank you for your comments. Following your suggestion, we used our PLNRP framework to train a new model on the BigNeuron dataset with random initialization, instead of using the pretrained model on VISoR-40. We also trained another model with our PLNPR framework using APP2 as the base tracer. The results are reported in Table II. "PLNPR-APP2" uses APP2 as the base tracer to generate pseudo labels for network training and "PLNPR- NGPST" uses NGPST as the base tracer for progressive learning. Without pretraining the segmentation network on the VISoR-40 dataset, our method still achieves higher overall performance than previous methods.

In our previous submission, we finetune the model that is pretrained on VISoR-40 with the BigNeuron dataset for only one iteration and the results showed the effectiveness of our PLNPR method. While starting from random initialization on BigNeuron dataset, we performed the progressive learning for four iterations in this revised version. Compared with the results on BigNeuron reported in our previous submission, we achieved better results with random initialization but with more iterations.

Compare with [Li et al, 2017], both our PLNPR-APP2 and PLNPR-NGPST achieve comparable performance. However, our method does not require any manual annotations for network training. In addition, our PLNPR can integrate many base tracers and many deep neural segmentation networks and to improve the reconstruction results. We tested four base tracers,

including APP1, APP2, MOST, and NGPST, and three widely-used deep segmentation networks, including 3D HRNet, 3D DSN, and 3D U-Net in our PLNPR framework. The results of multiple variants at different iterations are shown in Fig. 9. It clearly shows that our PLNPR progressively improves all the tested base tracer.

[Comments from Reviewer #2]

[Q2-1] This paper presented a framework to reconstruct grouped neuronal population from large-scaled 3D optical microscopic images. The framework consists of two components: (1) A self-supervised fully convolutional CNN is trained using the ground-truth reconstructions produced iteratively by a base neuron tracer to enhance the neuron images; (2) The neuronal structures are reconstructed block by block and then merged to complete neuron models using the proposed population reconstruction algorithm. To evaluate the proposed methods on neuron populations, authors compared the proposed methods on the newly released dataset VISoR-40. Authors also benchmarked on single neurons released by BigNeuron as well as show-cased the neuron population reconstructed on a mouse brain image.

Pros:

1. The proposed framework is sound and practical.

2. Most of the manuscript is well-written and easy to follow.

3. The related works are comprehensive and helpful to understand the context.

4. The visual inspections look promising.

5. This paper released a new dataset which could trigger future studies in this domain.

[R2-1] We thank the reviewer for finding the work of interest, and appreciate the insightful points that have been raised. We will address them point by point in the following.

[Q2-2] Section III-B-2 is hard to follow for the readers. I would suggest adding a visual depiction to demonstrate the algorithm. Authors should also highlight how the proposed block-by-block tracing differ from the methods used in the mentioned methods, such as Ultra-Tracer?

[R2-2] We re-wrote Section III-B to explain our approach in more details with illustrations. UltraTracer and MEIT have the capability of neuronal population reconstruction from the large-scale image. However, for the local regions with low signal-noise-ratio, they fail to separate individual neurons and trace complete dendrites in a dense neuron population, as shown in Fig. 14. In comparison, thanks to the signal enhancement by our deep network and block propagation designed for dense neurites, our UltraNPR is more robust to reconstruct a more complete neuronal population from the low-quality image while individual neurons are continuously and smoothly traced.

[Q2-3] It is unclear why the authors selected different quantitative criteria in Table 1 and Table2. The metrics used in Table1 are voxel-wise / node-wise metrics which do not support the claims regarding the neuron fiber identifications nor the topology. However, at the same time, to assign fibers to the right somas is one of the major claimed advantages of the proposed framework. I would suggest the authors consider adding metrics to show to what extent the fiber could be assigned to the right soma. Without such numbers presented, it is hard to judge whether the automatically reconstructed models could be used for further analysis.

[R2-3] Thank you for your suggestion. We add more detailed description of the four metrics used in Table 1 in Section IV-A-2. These metrics follow the definition in NGPST [Quan2015]. The four metrics are first computed on each individual neuronal tree according to the manually labeled skeleton, and then averaged in a neuronal population weighted by the total length of the neuronal processes of each neuron. Therefore, these metrics are voxel-wise / node-wise metrics on individual neurons to reflect the correct assignment to somas. The reconstructed individual neurons of different methods from VISoR-40 test images are visualized in Fig. 11 of the revised manuscript and Fig. 3 of the supplementary file. Separated neurons are shown in different colors.

In addition, we add these four metrics (precision, recall, F-Score and Jaccard) of different methods on BigNeuron dataset in Table 2. However, neither the implementation of the learning-based tracing method nor the reconstruction results of [Li et al. 2017] is not available. In order to compare with [Li et al. 2017], we only compare three evaluation metrics reported in [Li et al. 2017] on the same test data. The three metrics include the entire structure average (ESA), different structure average (DSA) and percentage of different structures (PDS).

[Q2-4] Most of the compared traditional methods are highly threshold dependent. It is not clear whether the compared methods were tuned as expected. It is questionable if the low recalls showed in Table 1 were caused by under optimized thresholds. Authors please justify.

[R2-4] We have tuned the parameters for more than 10 times for each method and each test image block to find the best results. In Section IV-A-6, we add the description that "The parameters of these tracing methods are manually adjusted for each image block to get the best performance for fair comparison."

[Q2-5] It remains unclear how much the CNN learning framework really helped in Table1 as the block-by-block tracing methods seem not to be identical to NPGST [38] (please advise if it is). The alpha value is also chosen to be 0.1 rather at the F1 score peak. Authors should add the quantitative results with alpha=0 to see the performance gap with and without the self-supervised learning.

[R2-5] Thanks for your elaborate comments. In order to evaluate the effectiveness of our progressive learning framework on image enhancement, we test existing seven neuron tracing methods on the same images enhanced by our method, and compare with their reconstruction results without using self-supervised learning. Fig. 12 in our revised manuscript illustrates the improvement from four quantitative metrics. For any of the seven neuron tracing methods, the reconstruction results from our enhanced images are much better than the results performed on the raw images under the same parameter settings.

With regard to the enhancement parameter $alpha$, we report the four reconstruction metrics in Fig.9 (c) under different values (from 0 to 1). We finally choose alpha=0.1 (F-Score=0.850) rather than alpha=0.3 (F-Score=0.863) because that despite similar F-Scores, the precision is higher at alpha=0.1 (Precision=0.978) than the precision at alpha=0.3 (Precision=0.948). This is also because that there is a large portion of subtle neurites in the reconstruction results that do not affect the main neuron structure too much but will lead to

sensitive precision. After consulting experts majored in neurobiology, we choose alpha=0.1 to balance the F-Score and precision metrics.

[Q2-6] It is intuitive that DL based image enhancement could improve neuron tracing performance. It has been shown in previous works though most of them were supervised learning frameworks. Though authors showed that the image enhancement could be applied to the other base-tracing methods in Fig 7, the same comparison are not shown in Table1. It would be nice to see if the proposed tracing and merging algorithms would make a difference given the same enhanced images.

[R2-6] Thanks for your suggestion. We added a group of experiments to test seven tracers on the same enhanced images by the segmentation network trained with pseudo labels from NGPST and show the comparison with the base tracer in Fig.12 in our revised manuscript. For any of the seven neuron tracing methods, the reconstruction results using our enhanced images are much better than the results performed on the raw images under the same parameter settings. The results demonstrate that our DNN-enhanced images lead to better reconstruction results for other base tracers.

[Q2-7] BigNeuron dataset contains many small subsets which are significantly different. It might help to include the summaries of the chosen subsets in this manuscript.

[R2-7] Thanks for pointing this out. We add more description of the test data from BigNeuron that we used in Section IV-B in our revised manuscript. "Following [Li et al 2017], we select the same 68 images that are from a variety of species to evaluate the performance of dense neurite reconstruction. Manual reconstruction by experts is associated with each image. 51 images are used for network training in [Li et al 2017] and the remaining 17 images are used for evaluation. Note that we do not use the manual annotations in our PLNPR in training the deep neural network."

[Q2-8] One of the most relevant and interesting comparisons to see would be the previous block-by-block tracing methods such as UltraTracer regarding the tracing performance, memory peak, speed, etc rather than the other image processing tracers.

[R2-8] Thank you for your suggestion. We add the comparison between our UltraNPR and two large-scale neuron tracing methods, UltraTracer and MEIT in Section IV-C. The comparison results are shown in Fig. 14. Given a large-scale OM image, UltraTracer and MEIT fail to reconstruct separated individual neurons and trace a complete neuronal population in the image. In comparison, thanks to the signal enhancement by our PLNPR and block propagation strategy designed for dense neuronal population, our UltraNPR is more robust to reconstruct a more complete neuronal population form the challenging image. The peak memory is 22.4GB, 54.6GB, and 3.58GB for UltraTracer, MEIT, and our method, respectively. The tracing time is 42 minutes, 811 minutes, and 153 minutes for UltraTracer, MEIT, and our method, respectively. Our method requires much less computer memory than other methods and much less tracing time than MEIT. The results demonstrate the robustness, effectiveness, and low cost of our method.

[Q2-9] Without numbers against the ground-truth on maybe even a subset of the dataset, it is not clear if the mouse brain slice experiment provides additional information. Please justify how were the neurons proven to be successfully traced as claimed?

[R2-9] The two state-of-the-art methods, UltraTracer and MEIT, employ block-by-block framework to trace neurons from large-scale images. Since the ground-truth reconstructions for the BigNeuron dataset are available, we compare our PLNPR with UltraTracer and MEIT in Table 2 and Fig. 13. The results show that our method outperforms the two large-scale tracing methods. For dense neuron population reconstruction from ultra-scale images, since it is infeasible to manually annotate the dense neuronal population in an ultra-scale image, we qualitatively compare our UltraNPR with UltraTracer and MEIT in Section IV-C. Based on the visible somas and neurites in the raw images, the reconstructed neuron population using our UltraNPR is visually better (more complete and distinguishable for individual neurons) than other methods.

[Comments from Reviewer #3]

[Q3-1] In this paper, the authors propose an unsupervised progressive learning method for neuron segmentation in optical microscopy images. The main contribution of this work is introducing a novel iterative deep neural network training framework for segmentation without user's supervision. For this, a conventional neuron tracing algorithm generates pseudo labels, which is used to improve the accuracy of the segmentation network. The authors also extended the proposed method to the integrated workflow to reconstruct large-scale neuron populations in the microscopy dataset. The efficacy and performance of the proposed method are demonstrated using two neuron datasets. This paper deals with a research topic that will be of interest to many readers in the biology field. The proposed method seems useful because generating training labels is time-consuming and labor-intensive in biological datasets. However, the method is mainly based on heuristics, so rigorous validation/justification of the method is recommended. The exposition and writing could be improved as well. The following questions/comments should be addressed in the revised paper to be accepted to IEEE TMI.

[R3-1] We thank the reviewer for the agreement on the value of our work. We also appreciate the inspiring discussion points raised by the reviewer. We will address the concerns one by one in the following.

[Q3-2] The proposed method relies on a stand-alone neuron tracing method to automatically generate pseudo labels. Then, the generated pseudo labels are used to train the neural network in a supervised way. The main assumption behind the proposed method is that the iterative execution of the above two processes makes the training converge to an optimal solution. However, there is no guarantee that the proposed training process improves the accuracy of the segmentation network, i.e., closer to the ground truth (the proposed loss function only measures how the output of the segmentation network is close to the neuron tracing results from imperfect input volume). In-depth discussions about the convergence of the proposed method would be helpful.

[R3-2] Yes, the main assumption is that the iterative execution of the neuron tracing and deep segmentation network makes the training converge to an optimal solution. Unfortunately, there is no theoretical guarantee on the convergence to the optimal solution. However, our experiments on a series of combination of different tracing methods and segmentation network

verify this assumption. Without losing too much precision, our PLNPR is not affected by the noises in the pseudo-labels too much, and is able to progressively improve the recall and the overall reconstruction performance.

As shown in Fig. 4 in our revised manuscript, our PLNPR method progressively improves the segmentation results and reconstruction results. We also evaluate the reconstruction results of our PLNPR with four neuron tracing methods at eight iterations. The results are shown in Fig.9 (a) and Fig. 10. Our progressive learning strategy effectively facilitates conventional tracing methods to reconstruct more complete neuronal populations. In addition, the performance improvement gets stable after five iterations of the progressive learning for all the tested tracing methods. Moreover, we test seven neuron tracing methods on the images enhanced by our method, and the performance comparisons are shown in Fig. 12. For any of the seven tracing methods, the reconstruction results using our enhanced images are much better than the results performed on the raw images. These results show that our method is effective and can greatly improve the neuron reconstruction performance with progressive learning.

[Q3-3] The fusion algorithm, described in Section III B, seems heuristic as well. The method relies on simple discarding and merging techniques based on the spatial proximity between neurites with empirically chosen voxel distance thresholds. The description of the algorithm is also confusing and difficult to understand. I assume there will be many topological errors during merging, but the performance of the fusion algorithm was not assessed qualitatively.
[R3-3] Thank you for your comments. We agree with you that our UltraNPR is a heuristic approach in the commonly-used block-by-block scheme. In order to trace multiple neurons and avoid topological errors, we design a block-propagation strategy by setting the terminal tips from reconstructed blocks as pseudo somas for reconstructing neurons from unreconstructed blocks. The fusion strategy is also designed to eliminate topological errors by checking the structure consistency in the overlapped regions of two neighboring blocks.

To illustrate the effectiveness of our UltraNPR, we compare the reconstruction results with two state-of-the-art large-scale neuron tracing methods, UltraTracer [Peng et al. 2017] and MEIT [Wang et al 2018] in Fig. 14 in our revised manuscript. UltraTracer and MEIT perform poorly on reconstructing separated neurons and a complete neuronal population in the large-scale image. We also tested the combination of our block propagation and fusion strategy with NPGST [Quan et al. 2015] on the raw image blocks. Without image enhancement by our PLNPR, it fails to reconstruct subtle neurites for most neurons. In comparison, thanks to the signal enhancement by our PLNPR and block propagation strategy designed for dense neurons, our UltraNPR achieves a more complete neuronal population from the challenging image.

[Q3-4] In general, the paper did not describe the methods and algorithms well with sufficient details. For example, there is no clear description of the network architecture/layer sizes/parameters/training methods/algorithms, etc. There is no formal definition of loss functions used in the proposed network either. Without such information, it is difficult to reproduce the proposed method.

[R3-4] Thank you for pointing this out. We re-organized our manuscript to make the whole pipeline clearer. We add Fig. 2 as illustration of the algorithm overview in the revised manuscript. We also add more detailed explanation of our UltraNPR and illustrations in Section III-B. For the implementation details of the segmentation network, because our PLNPR is a generic framework that can integrate any traditional neuron tracers and deep 3D segmentation networks, we follow the same parameter setting, and loss functions in our PLNPR with each neuron tracer or segmentation network. Due to the page limitation, we did not list all the details that can be found in their original paper. We briefly describe the training hyper-parameters in Section IV-A.

[Q3-5] All the performance scores (e.g., F-score, precision, recall, etc) from the test dataset are single numbers. I assume there are many individual neurites or data volumes (e.g., there are 40 volumes in VISoR-40 set and each contains multiple neurites), so I want to see how the performance is consistent over individual results rather than the averaged scores. You may want to use a boxplot (or square plot) to visualize those.

[R3-5] Thank you for your comments. We add boxplot visualization for the eight tracing methods on the VISoR-40 test image blocks in Fig. 4 of the supplementary file. For each tracing method, we show the distribution of four metrics (precision, recall, F-Score, Jaccard) on the seven test blocks, including the minimum, the maximum, the median, the average, and the first and third quartiles of each metric. Compared with other methods, our PLNPR achieves the highest overall performance and more consistent results over individual blocks. Due to the space limit, we do not show this figure in our paper.

[Q3-6] I have concerns that the difference between this submission and the authors' previous MICCAI paper is marginal. There are some additional figures and texts in this submission, but the main differences are introducing the UltraNPR algorithm and adding one more tracing result (TReMAP), which may not be sufficient for the journal extension (e.g., the main experimental results in Table I and II are almost identical except the new addition of TReMAP result). I suggest adding new qualitative results of the UltraNPR algorithm.

[R3-6] Thanks for your constructive suggestion. We re-wrote much of our manuscript to make our contribution more clear. First, we rewrote the whole section Section III-B "Ultra-scale Neuronal Population Reconstruction" with more detailed explanation and figures to introduce our UltraNPR algorithm. Our block propagation and neurite fusion approach efficiently reconstruct large-scale neuron populations, which is not supported in our MICCAI paper. Secondly, we abbreviate the section of "PLNPR" to reduce overlap with our MICCAI paper. Third, a series of experiments were conducted to evaluate our UltraNPR algorithm and compare with state-of-the-art large-scale reconstruction algorithms including UltraTracer [Peng et al 2017] and MEIT [Wang et al. 2018].