nature portfolio

# Peer Review File

# Pushing the limits of fluorescence imaging with a restoration neural network aggregating large-view statistics

Corresponding Author: Professor Peng Xi

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

**This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.**

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In "Pushing the limits of fluorescence imaging with a restoration neural network aggregating large-view statistics", Hou et al develop a neural network architecture optimized to enable significantly larger image areas to be assessed as a single entity. They claim that this model is particularly good for restoration of fluorescence microscopy data due to its ability to consider both local and global statistical affects in its reconstructions.

Their model, called LargePNet, is presented with a range of minor variations and applied to a range of fluorescence imaging modalities and compared with a number of current best-in-class approaches. They have deposited the image data gathered for this paper in an open repository to assist with future research for others.

Overall this is an extremely nice paper, well presented and seems to represent a significant step forward for processing both single images and volumes or time series data sets. I it is suitable for publication with corrections.

Major issue:

I feel that the paper very much bypasses any drawbacks to their model over existing approaches. The only mention of any downsides at all in the 3 sentences at the end of the discussion. A rigorous comparison with existing techniques requires a more balanced review of the strengths and weaknesses compared to the existing approaches. The authors mention that it doesn't perform better for small ROI, can we have some data on when it is better and when it performs similarly. They also say that it is slower than Unet but requires similar amounts of computation, the only data regarding this appears to be in fig 1, where LargePNet has significantly fewer multiply-accumulate operations than Unet but is still significantly slower.

Minor issues:
Although mostly well written there are a number of minor issues with

the text:

1)
Page 5:
"Especially, as our modification in LFFE, we also employ instance normalization in HFFE to replace the commonly used batch normalization..."

Do the authors mean "Especially, like our modification in the LFFE,"?

2)
"And we found that instance normalization is beneficial for training LargePNet"

Delete the And "We found that...."

3)
"...architecture focuses on a region with less in a 128-pixel window"

Would be better phrased as "architecture focuses on a region in a 128-pixel, or less, window"

4)
Page 6
"... especially in the heavy, noisy conditions."
Noise is more usually referred to a "high noise" rather than "heavy". This phraseology is repeated several times through the paper. EG fig2 with light noise and heavy noise, should be low noise and high noise and extended data fig 2. I am sure there are other places too.

5)
Page 8
"...and we found that harsher conditions would lead to collapsed performance of DFCAN and UniFMIR"

By context this implies that they mean lower intensity, but it is not clear what the "Harsher conditions" refers too.

6) "...taking only ~1/4 and 1/20 of the time cost by DFCAN and UniFMIR", should be "...of the time cost of DFCAN..."

7)
Fig 2, box and whisker plots are great, but what do the boxes and whisker represent?

8)
Page 10
"...with fewer computational cost ..."
It should be "lower computational cost"

9)

"We devise LargeP-GAN using LargePNet as the generator and modifying the CNN architecture in ERSGAN48 series to a multiscale version as the discriminator (Supplementary Note 3)"

I think the word series should be deleted.

10) page 15 "Besides assisting fluorescence imaging in the widefield system, the trained LargePNet models can also be utilized to achieve sustained multicolor super-resolution imaging that can be hard to

achieve in the hardware dimension in the point-detecting confocal system"

This needs to be rephrased as I struggle to see how you get super resolution in hardware with a point-detecting confocal system.

11) "With the clear images after restoration, we have observed that a mitochondrial end was inlaid in the ER and microtubule network at most of the 1-hour recording duration,"

I don't know what you mean by "inlaid in" and do you mean "for most of the..."?

12) page 17 - "...such as label-free and electron microscopes that image biological specimens with rich identity structures over a large view."

I don't understand the phrase "rich identity structures"

13)
Extended data figure 4, Low and heavy computational costs would be better as low and high.

14)
Page 24, "It usually takes 800-1,200 512×512 image pairs to well feed LargePNet"

Do you mean "to train LargePNet"?

15)
"COS-7 cells were firstly pre-fixed" should be "first pre-fixed"

16) page 25 " after slightly blurring the GT image with a Gaussian kernel with 1.0 sigma, and train DL models to transform the degenerated image to the GT"

Is the 1.0 in pixels, or some other unit? The rest of this paragraph also references further blurring, again not specifying what unit the sigma refers to.

17) "Three commercially available fixed slices were used for imaging, including a mouse kidney section (FluoCells Prepared Slide #3, Invitrogen, F24630), a homemade stem of the Nerlum sample."

Only two of these appear to be commercial available.

18)
page 26
"in 35 mm class bottom" should be "glass bottom" - repeated elsewhere as well.

19)
"After washing with 1× PBS for 3 times" - should be "After washing with 1× PBS, 3 times" - the additional "for" is also in several other places

20)
"The secondary antibody was labeled as previously described."
Reference needed, also they don't say anything about the DNA probes used. Are they commercial, described in the not included reference or something else?

21) "Airy-NovaSD (Bejing, China) spinning- disk confocal microscopy to capture"
Should be "microscope"

22) "In previous assessments, 128×128 patches with an average photon count of 25-600 are selected in the calculation of image metrics, and the calculation method was also described"

Is the average photon count per patch or per pixel, it is not clear.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
Article summary:

In this paper, the authors discuss a new ML method for image restoration. This method leverages scale separation by synthesizing the outputs of small, deep convolutions and large, shallow convolutions. They argue that the latter element—through a much larger effective receptive field—is better suited to image restoration tasks by taking global statistics into account. They demonstrate this method to produce a 0.5-2dB increase in SNR compared to canonical method (U-Net), as well as a significant 4-20x increase in computational efficiency. They argue that the increased processing speed makes this method better suited to large timeseries datasets.

Overall comments:

Overall, this manuscript offers an interesting discussion of effective receptive fields and their effect(s) on the accuracy of network predictions for image restoration. The authors test a new architecture—LargePNet—based on the synthesis of low- and high-frequency feature extraction steps that act on large (512x512) patches. The authors demonstrate that LargePNet offers substantial increases in computational efficiency and processing time compared to current methods such as the commonly used UNet architecture. This is an important consideration for users! The manuscript would be strengthened by a discussion (or expanded discussion) of how these efficiency gains are achieved.

However, while there are some instances and some metrics for which LargePNet outperforms current methods, there is not substantial evidence that this new architecture offers significant general performance gains. Moreover, any apparent disparities in performance appear to be dependent on the biological structure present in the images. This manuscript would be strengthened by a more quantitative statistical presentation of comparative performance gains for LargePNet relative to current methods. Moreover, a discussion of why LargePNet performs better/worse when restoring the images of different biological structures would be of interest.

With that said, it is more clear that LargePNet outperforms the current methods when restoring higher-dimensional imaging data such as time series and volumes. The manuscript would be strengthened by discussing potential reasons for this improvement.

The microscopy methods are generally underreported. Please make sure for all image data collected the objective specifications, spatial and temporal sampling (ex. Pixel size and time interval), laser lines, detectors, and post-processing algorithims (ex. SIM reconstruction) are reported.

General Comments:

The authors make repeated claims on the effect/shortfalls of small effective receptive field of architectures such as U-Net but only cite a single analysis of this fact. In one contrasting example, 10.1117/1.JMI.11.5.054004 suggests that larger true receptive fields (TRF, and by extension, the ERF) are not always correlated with superior network performance in the context of segmentation. This is not to claim that this is also true for image restoration tasks, but a deeper source of support for this claim would strengthen the manuscript. Moreover, this paper also makes the point that different sized structures benefit from differently sized TRF, which could plausibly also apply to restoration tasks.

Similarly, the authors claim “inconsistenc[ies] of the patch size during training, evaluation, and final application” are often overlooked. It would be good to add some discussion or citations regarding the perils of these inconsistencies.

The content in Figure S1 does not correlate with the figure caption. The authors argue that the histograms from the 4 ROIs of the BioSR image are “identical”. However, it is clear that these histograms—particularly comparing ROI1 and ROI2—are distinct. Moreover, from the image, it seems as though an ROI in the dimmer part of the cell (presumably around the nucleus) would yield yet another distinct histogram. Please address these discrepancies.

Moreover, the motivation derived from these data is not clear. Even if small ROIs contain ~similar intensity information, and thus, similar statistics, then isn’t this an argument for small patch sizes? If any one patch is similar to another, then there would appear to be no need to consider more information. Furthermore, if, as it seems here, it is plausible that biological images in fact contain regions of distinct spatial patterns, isn’t that also an argument for small patches, as one would want the network to learn these features independently? It would seem that using large patches could obfuscate some of these

more granular details. A more thorough explanation of these topics is warranted.

While measurements of PSNR, NRSRE, MS-SSIM are good, it would be good to also include a measure of downstream accuracy such as segmentation quality. Image restoration for the sake of restoration is of minimal value in terms of image analysis. Moreover, it could be good to identify regions that contain potentially significant hallucinations, if they exist. In particular, in Fig S2, there is a potential hallucinated connection in all model predictions that does not appear to exist in the GT (just west and north of the middle of the image).

I am curious how the computational efficiency increase(s) relative to current methods such as DFCAN scales with input image size (see Figure 1f).

Across the manuscript, stating whether the gains in these metrics are statistically meaningful compared to current methods would strength conclusions.

A discussion of choice of metrics and their relation to resolving biological structures would be helpful.

As a general comment, many of the arrows used to highlight regions or line profiles do not sufficiently or precisely identify the ROI in question.

Figure-specific Questions and Comments:

Figure 1

The connection between the displayed ERF's in b and c is not obvious. In particular, the LargePNet ERF in b appears significantly broader than that in c. Please explain why this is the case.

Figure 2

From the plots in a, the results appear significant for some structures, but maybe not others. A discussion of why this might be would be useful. In other words, what sorts of structures is LargePNet better at recognizing than others?

What do the white arrows and curves represent in c? The curves, in particular, do not appear to correlate with structures in the image. Moreover, the LargePNet output appears to miss the fiber between the arrows from ~10 o'clock to 5, while DFCAN-Stitching and UniFMIR do not to the same degree, at least qualitatively.

In e, the array pointing to "signal loss" in DFCAN-Stitch could just as well be used in the LargePNet prediction. A quantitative assessment demonstrating that the prediction in LargePNet is improved relative to DFCAN-Stitch would better support this assertion.

In g, while LargePNet appears to perform qualitatively better than DFCAN and UniFMIR, its prediction also appears far off from the GT SIM data. A comparison between the GT and the prediction would strengthen this figure.

Figure 3

For these predictions, were 512x512 blocks also used as in the other examples?

In a, it is not obvious that LargePNet performs much better than UNetRCAN across the board. An argument could be made for LargePNet vs SwinIR, but not vs UNetRCAN, at least by eye.

For b-e, adding zoomed-in ROIS corresponding to the regions from which these curves are taken would make it easier to correlate the intensity profiles with the biological structure. The arrows seem to refer to such a small ROI that it is difficult to see the structures being highlighted.

In g, it is once again difficult to assess whether LargeP-GAN is actually performing significantly better than UNet-GAN. Other than FID, there is not an appreciable difference in many/all of the metrics. Commenting on why FID is used and what this signifies would also strengthen this point.

Panel h is difficult to interpret. Perhaps zoomed-in ROIs would highlight differences between the methods. Otherwise, as large images, it is very hard to see observe much difference.

Figure 4

For training the timeseries model, were depth-wise convolutions used? In other words, clarifying whether the temporal information is incorporated across timepoints would be helpful.

Carrying over the white ROI box in all panels of a-b would better communicate the selection in c-d.

This figure—outside the MT case in e—does appear to show that LargePNet outperforms other current methods. However, see the comment about quantifying the significance of this performance increase. Moreover, considering that this example

appears to demonstrate the most significant increase in performance for LargePNet compared to current methods across the manuscript, a discussion of why this network seems to perform comparatively better in higher dimensions would be interesting.

For the line profiles in j, the highlighted feature (the small bump before the large peak) does not seem significant enough to highlight. In the microscopy images, it does not appear to correspond to any meaningful structure and could potentially be noise. I would suggest removing these profiles and only display the ROIs.

Figure 5

Without GT as a comparison, it is very difficult to judge whether the recovered structures exist in the raw data. As a suggestions, the same experiment could be performed of imaging at low SNR every 30 seconds, but interleave a high SNR image every couple of hours. This could then showcase the same point being made currently while also allowing for a more direct ground truth comparison in the same data set.

The sentence “LargePNet has effectively removed the massive noise in the raw recordings, enabling fine segmentation of the microtubules structures.” could be toned down. Moreover, unless the segmentations are provided, it is not necessarily fair to claim that the segmentation is enabled.

Extended data:

Fig 1

The images are very small in e. It is very difficult to assess model performance.

Fig 2

I would consider reordering the rows/columns in d. Putting the GT next to the outputs makes comparison easier. I would also suggest moving the image labels off the images themselves so that the entire image can be seen.

Fig 3

See comments about statistical significance of performance gains in a.

See comments about arrows identifying regions in b-e.

Specifying which profiles correspond to which images—i.e. does f correspond to the region in b or c?—would reduce ambiguity in comparing the line profiles to the ROIS.

Fig 4

The table of comparisons is the most useful part of this figure.

Fig 5

It is difficult to parse any meaningful differences between the images in a-b at this scale.

The similarity between UNet-GAN and LargeP-GAN in c is hard to ignore.

I am not sure that it is robust to assert the existence of a 3rd peak in either the GT or the LargeP-GAN output in d.

Once again, the metrics presented in e-f do not seem to suggest a significant difference between these methods, when taken together. However, the potentially significant increase in PSNR from LargeP-GAN is noted in f. Once again, commenting on the difference in performance across the structures would be welcome. Moreover, I wonder why f is included only in the extended data and not the main figure.

Fig 6

See comments about statistical significant and disparity in comparative performance across different structures.

The difference between the middle peaks in the LargePNet prediction and GT should be discussed if they are going to be highlighted.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)
The repository does not appear to contain the pre-trained models required to test inference using the provided Jupyter notebooks. I can thus not comment on the efficacy of the code in terms of reproducing the reported results.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In the revised version of "Pushing the limits of fluorescence imaging with a restoration neural network aggregating large-view statistics", Hou et al have significantly improved on their original submission.

They responded to my original points and have produced a clearer and more complete paper suitable for publication.

I do feel that the 70 page response to reviewers was totally excessive and the responses could have been more concise and still adequately covered the ground.

Additional minor issues in added text:

Introduction

"The small patch truncated many contextual structures and imposes the network to learn restoration with limited information."

You must impose on something, so this could be reworded but probably better to just replace imposes with forces.

"However, the application image typically span large views with much richer contextural structure information, which shall be important restoration clues but remain largely unexplored"

Either needs to be "image typically spans" or "images typically span" depending on if they are referring to a single application image or the whole set of images. "Which shall be" would be better phrases as "which are"

"Moreover, the difference of the structure statistics in the large view and small patch", The critical factor is not the difference in the large view and small patch results but the differences between them.

Discussion

"It should be noted that despite LargePNet significantly improves the restoration performance of previous models, "

Should the "significantly improving the"

"Besides fluorescence microscopy, we believe that LargePNet can also improve the DL restoration performance in images similar to fluorescence images, such as label-free and electron microscope images."

If the authors wish to claim that their model CAN improve other modalities they should show that. More realistically they should say may or could improve.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The authors have done a fantastic and thorough job of addressing the points laid out in the previous review, and we therefore feel it is sufficiently ready for publication.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

# Response to Reviewers' Comments

We very much appreciate the critical reading of our manuscript and valuable suggestions from the reviewers. We have carefully reviewed the comments and have revised the manuscript accordingly. To ensure clarity and facilitate the review process, all significant changes have been highlighted in red. The responses to the comments are listed one by one as follows (in blue):

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

In "Pushing the limits of fluorescence imaging with a restoration neural network aggregating large-view statistics", Hou et al develop a neural network architecture optimized to enable significantly larger image areas to be assessed as a single entity. They claim that this model is particularly good for restoration of fluorescence microscopy data due to its ability to consider both local and global statistical affects in its reconstructions.
Their model, called LargePNet, is presented with a range of minor variations and applied to a range of fluorescence imaging modalities and compared with a number of current best-in-class approaches. They have deposited the image data gathered for this paper in an open repository to assist with future research for others.
Overall this is an extremely nice paper, well presented and seems to represent a significant step forward for processing both single images and volumes or time series data sets. I it is suitable for publication with corrections.
Response:

Thanks for your very kind summary and the positive attitude of our manuscript. Your advice below has helped us to improve the quality of the paper significantly, which we greatly appreciate.

Major issue:
I feel that the paper very much bypasses any drawbacks to their model over existing approaches. The only mention of any downsides at all in the 3 sentences at the end of the discussion. A rigorous comparison with existing techniques requires a more balanced review of the strengths and weaknesses compared to the existing approaches. The authors mention that it doesn't perform better for small ROI, can we have some data on when it is better and when it performs similarly. They also say that it is slower than Unet but requires similar amounts of computation, the only data regarding this appears to be in fig 1, where LargePNet has significantly fewer multiply-accumulate operations than Unet but is still significantly slower.
Response:

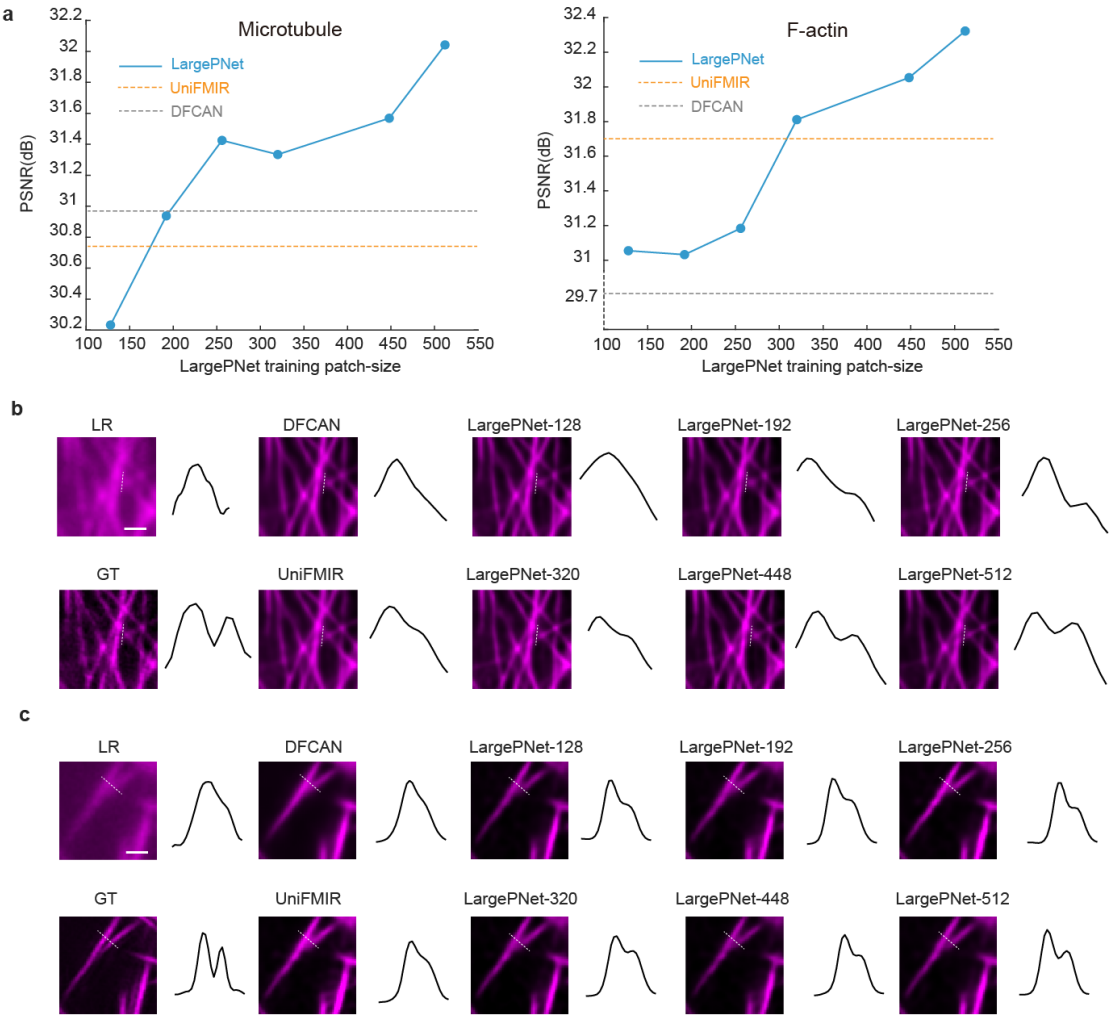
Thanks for your feedback and suggestions. In response to the specific points raised, we have now expanded our discussion and added data to provide a more balanced and rigorous comparison of LargePNet's strengths and limitations relative to existing approaches. Below, we address the key concerns in detail:

**(1) Drawbacks of LargePNet**

We appreciate the reviewer's insightful suggestion regarding the need for a more explicit discussion of the model's limitations. Here, since the architecture of LargePNet is devised to train with large images/patches, when only small patches are available, this architecture may not perform better than state-of-the-art small-patch networks.

We illustrate this issue using microtubule and F-actin structures from the BioSR dataset, which are two complex structures. We test the performance of LargePNet when

forcing it to be trained with patch sizes of 128, 192, 256, 320, 448, and 512 pixels. The results are shown below, which are also added as Supplementary Fig. 19 in the revised manuscript. As shown, when ROI is smaller than 300 pixels, LargePNet may not outperform or perform similarly with previous state-of-the-art small-patch networks in this task (DFCAN and UniFMIR). In contrast, when ROI>300 pixels are available, LargePNet evidently outperforms previous methods. Moreover, the best performance of LargePNet can be obtained when the full-size image (with ROIs of 512 pixels) is used directly to train LargePNet.




**Supplementary Fig. 19. Investigation of the LargePNet performance when only small ROIs are available for training. a.** The PSNR metrics of microtubule and F-actin structures in BioSR datasets. **b.** Instance inference results of the microtubule structures. **c.** Instance inference results of the F-actin structures. Scale bars: 0.5 μm.

Action:

We have added data demonstrating that LargePNet may not outperform or perform similarly when only small ROIs are available. The results are added as Supplementary Fig. 19 in the revised supplementary information. This issue is discussed in more detail in the Discussion section of the revised manuscript.

## (2) Processing speed

We thank the reviewer for raising this important point. We shall explain this issue more clearly to avoid confusion from readers. Although LargePNet has fewer multiply-accumulate operations than Unet, it is still evidently slower. This is due to the fact that the current off-the-shelf deep learning tools (such as Pytorch) are more well optimized for models like Unet that stack very deep 3×3 convolutions. While for the major operation in LargePNet, that is the reparametrized-ultra-large-kernel-convolution (RepLKConv), the optimization in Pytorch is not sufficiently optimized as a conventional CNN architecture. This issue was also demonstrated and discussed in a previous literature [*Ding, X. et al. Scaling Up Your Kernels to 31×31: Revisiting Large Kernel Design in CNNs. In IEEE Conference on Computer Vision and Pattern Recognition (2022)*]. Despite this, LargePNet is still much more efficient than some advanced networks (such as DFCAN, UniFMIR) while providing a better performance, as shown in the processing time table in Figs. 2-4. These networks typically provide significantly better performance than UNet in restoration tasks.

Action:

We have explained in more detail with citation of a relevant analysis in the Discussion section of the revised manuscript:

"*Secondly, while possessing a computational cost lower than some advanced models, LargePNet's processing speed is still evidently lower than the commonly used UNet, despite the total computation amounts being equivalent. This is due to the fact that current off-the-shelf deep-learning tools (such as Pytorch) are more well-optimized for conventional CNNs that stack very deep 3×3 convolutions compared with RepLKConv*[34] *that occupies major computational cost in LargePNet. Despite this, LargePNet is still much more efficient than some advanced networks (such as DFCAN and UniFMIR), while providing better performance. Optimization of the RepLKConv efficiency can further benefit the deployment of LargePNet.*"

Minor issues:

Although mostly well written there are a number of minor issues with the text:

Response:

We are very grateful that you carefully checked the presentations over the manuscript. We have seriously revised the manuscript based on your feedback.

1)

Page 5:

"Especially, as our modification in LFFE, we also employ instance normalization in HFFE to replace the commonly used batch normalization..."

Do the authors mean "Especially, like our modification in the LFFE,"?

Response:

Thanks for your feedback. Yes, we mean "like our modification in the LFFE". We have modified this sentence on Page 5 according to your advice.

2)

"And we found that instance normalization is beneficial for training

LargePNet"
Delete the And "We found that...."
Response:
Thanks for your feedback. We have corrected this error on Page 5.

3)
"...architecture focuses on a region with less in a 128-pixel window"
Would be better phrased as "architecture focuses on a region in a 128-pixel, or less, window"
Response:
Thanks for your good suggestion. We have modified this sentence on Page 5 according to your advice.

4)
Page 6
"... especially in the heavy, noisy conditions." Noise is more usually referred to a "high noise" rather than "heavy". This phraseology is repeated several times through the paper. EG fig2 with light noise and heavy noise, should be low noise and high noise and extended data fig 2. I am sure there are other places too.
Response:
Thanks for your careful review of the manuscript. We have modified the descriptions to “low-noise” and “high-noise”, and checked other parts of the paper to revise similar errors.

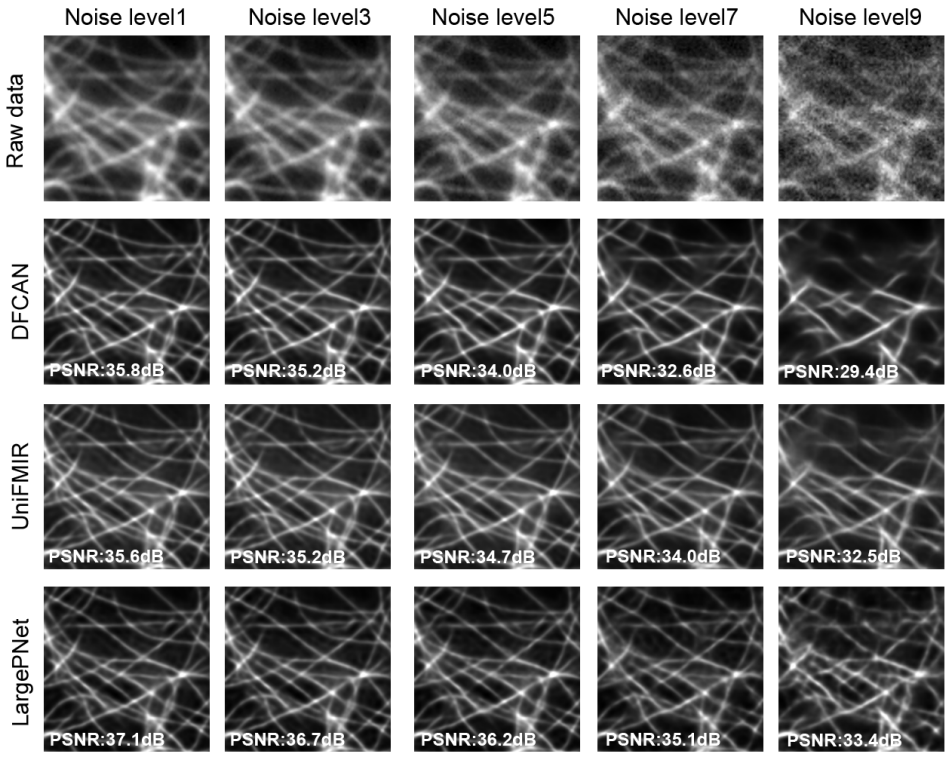
5)
Page 8
"...and we found that harsher conditions would lead to collapsed performance of DFCAN and UniFMIR"
By context this implies that they mean lower intensity, but it is not clear what the "Harsher conditions" refers too.
Response:
Thanks for your good suggestion. We aim to denote “lower intensity” here. We have modified this word on Page 8 for a better description according to your advice.

6) "...taking only ~1/4 and 1/20 of the time cost by DFCAN and
UniFMIR", should be "...of the time cost of DFCAN..."
Response:
Thanks for your feedback. We have modified this word on Page 8 according to your advice.

7)
Fig 2, box and whisker plots are great, but what do the boxes and whisker represent?
Response:
Thanks for your question. In this paper, we employ the box chart to display the

distribution of the restoration metrics. The lower and upper boundaries of the box correspond to the first quartile (Q1, the value at the 25th percentile when the data is sorted in ascending order) and the third quartile (Q3, the value at the 75th percentile) of the data, respectively. The median value and the average value are marked by the centerline and diamond point in the box, respectively. The interquartile range (IQR) is calculated as IQR=Q3−Q1. The lower and upper whisker endpoints are presented as Q1−1.5×IQR and Q3+1.5×IQR.

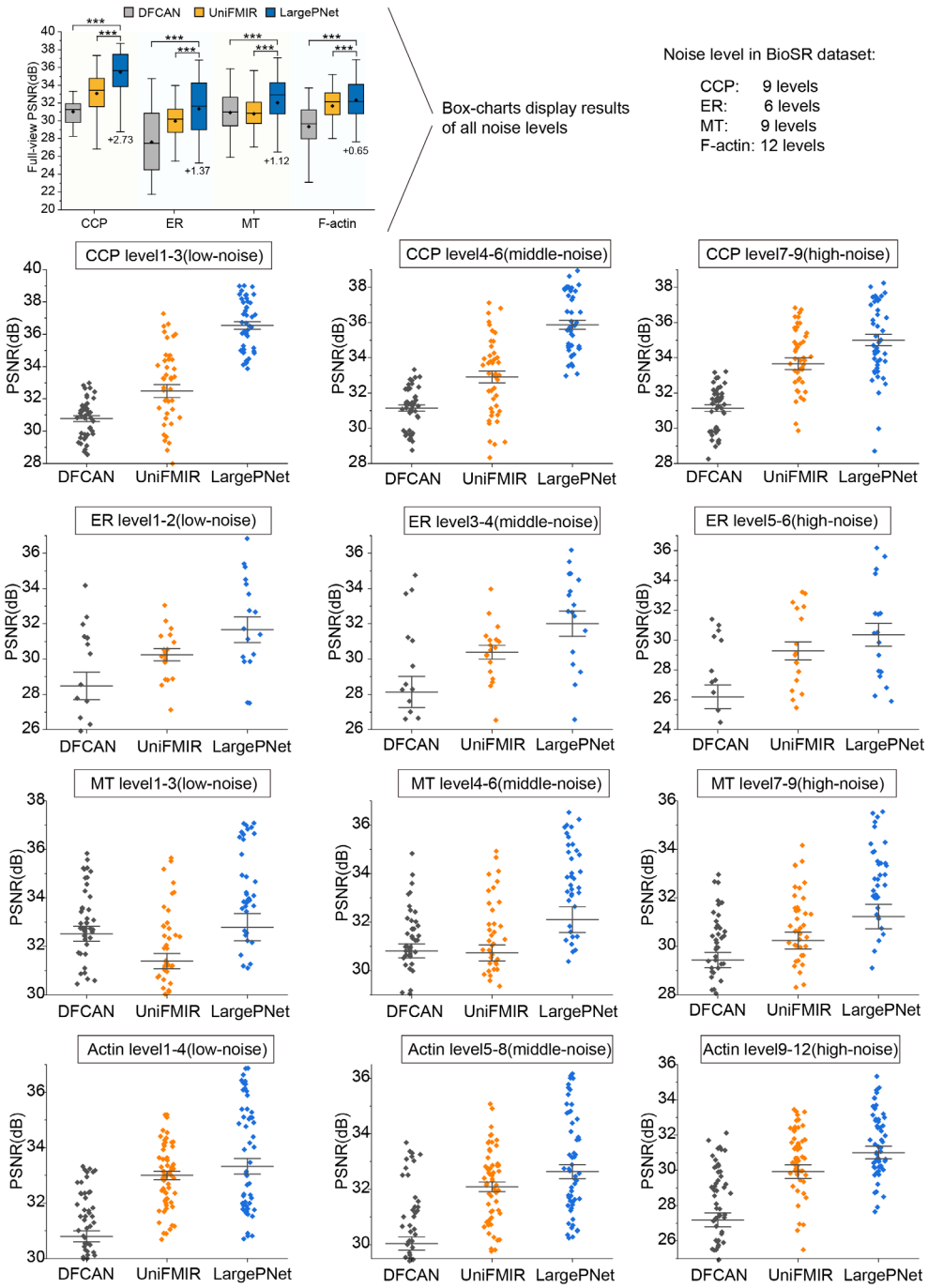
The box and whisker plot represents the deviation of the restoration metrics of different FOVs across a wide range of noise levels (e.g., F-actin in BioSR datasets is collected with 12 noise levels for a single FOV). All model performance will sensitively decrease with the increase in the noise level. This deviation causes the long whisker and the overlap of the boxes of different models in the box chart display. An example is shown in the Response Fig. 1-1 below.




Response Fig. 1-1. All models show evident performance degradation as raw image noise levels increase. This deviation causes the long whisker and the overlap of the boxed of different models in the box chart display.

If we display the results at relatively separated noise level groups, the error bar would not be that great, as shown in Response Fig. 1-2 below.

We appreciate you for pointing out this question, which may also be arisen by readers. We have provided a detailed explanation to this issue.




Response Fig. 1-2. Statistical comparisons in the BioSR super-resolution dataset with separated noise levels. Each point displays the metric of an image. The error bar displays the average value and standard error (SE).

Action:

We have added a description about what the box represents and the origin of the great whisker in the box chart in the Methods section of the revised manuscript. Response Fig. 1-2 is made as the new Supplementary Fig. 6 in the revised Supplementary Information to help understand this issue. The Methods section "Box-chart display" is written as:

"*In this paper, we employ the box chart to display the distribution of the restoration*

*metrics. The lower and upper boundaries of the box correspond to the first quartile (Q1, the value at the 25th percentile when the data is sorted in ascending order) and the third quartile (Q3, the value at the 75th percentile) of the data, respectively. The median value and the average value are marked by the centerline and diamond point in the box, respectively. The interquartile range (IQR) is calculated as IQR=Q3−Q1. The lower and upper whisker endpoints is presented as Q1−1.5×IQR and Q3+1.5×IQR.*

*As presented in this paper, the boxes and whiskers are typically relatively long. This is because restored image datasets usually acquire different levels of SNR for a single FOV (e.g., the F-actin dataset in BioSR includes 12 noise levels). The performance of all models degrades significantly as the SNR of the original images decreases, which results in the elongated boxes and whiskers. In addition, although box plots effectively present the fluctuations of restoration metrics across samples with varying SNRs, they may obscure the performance differences between different models. Statistical charts stratified by different noise levels could help to better understand the performance of various models under different initial SNR conditions of original images. Some illustrative examples are provided in Supplementary Figs. 6, 9.*"

8)
Page 10
"...with fewer computational cost ..."
It should be "lower computational cost"
Response:
Thanks for your feedback. We have modified this word on Page 10 according to your advice.

9)
"We devise LargeP-GAN using LargePNet as the generator and modifying the CNN architecture in ERSGAN48 series to a multiscale version as the discriminator (Supplementary Note 3)"
I think the word series should be deleted.
Response:
Thanks for your feedback. We have deleted the word "series" on Page 10 according to your advice.

10) page 15 "Besides assisting fluorescence imaging in the widefield system, the trained LargePNet models can also be utilized to achieve sustained multicolor super-resolution imaging that can be hard to achieve in the hardware dimension in the point-detecting confocal system"
This needs to be rephrased as I struggle to see how you get super resolution in hardware with a point-detecting confocal system.
Response:
Thanks for pointing this out. We apologize for any confusion caused by our initial description. We may get used to the terminology in some commercial microscopes like "super-resolution confocal system", which actually denotes the STED microscope. We

have revised the original description “in the point-detecting confocal system” to “in the point-detecting STED system” on Page 15 to avoid confusion.

11) "With the clear images after restoration, we have observed that a mitochondrial end was inlaid in the ER and microtubule network at most of the 1-hour recording duration," I don't know what you mean by "inlaid in" and do you mean "for most of the..."?
Response:

Thanks for pointing this out. By "inlaid in" we meant that the mitochondrial tip was persistently apposed to and maintained a contact site with the ER and the microtubule network. We agree that "inlaid in" is not the best wording, and we also intended to say "for most of the 1-hour recording" here. A clearer revised sentence is, "*With the restored images, we observed that one mitochondrial end maintained persistent contact with ER tubules and the microtubule network for most of the 1-hour recording.*"
Action:

We have revised this sentence on Page 10 in the revised manuscript to avoid confusion.

12) page 17 - "...such as label-free and electron microscopes that image biological specimens with rich identity structures over a large view."
I don't understand the phrase "rich identity structures"
Response:

Thanks for pointing this out. We aim to denote that this method can be potentially generalized to other images that may share similar characteristics with fluorescence images, such as label-free and electron microscope images. We have revised this sentence to “*Besides fluorescence microscopy, we believe that LargePNet can also improve the DL restoration performance in images similar to fluorescence images, such as label-free and electron microscope images*” to avoid possible confusion.
Action:

We have revised this sentence on Page 17 in the revised manuscript to avoid confusion.

13)
Extended data figure 4, Low and heavy computational costs would be better as low and high.
Response:

Thanks for pointing this out. We have revised the word “heavy” to “high” in Extended Data Fig. 4 in the revised manuscript.

14)
Page 24, "It usually takes 800-1,200 512×512 image pairs to well feed LargePNet"
Do you mean "to train LargePNet"?
Response:

Thanks for your feedback. We do mean “train LagePNet”. We have revised this

word on Page 25 to avoid ambiguity.

15)
"COS-7 cells were firstly pre-fixed" should be "first pre-fixed"
Response:
Thanks for your feedback. We have corrected this error on Page 17 in the revised manuscript.

16) page 25 " after slightly blurring the GT image with a Gaussian kernel with 1.0 sigma, and train DL models to transform the degenerated image to the GT"
Is the 1.0 in pixels, or some other unit? The rest of this paragraph also references further blurring, again not specifying what unit the sigma refers to.
Response:
Thanks for pointing this out. We shall make it clearer. The blurring uses a Gaussian kernel with a function: $g = \exp[-(x^2+y^2)/2\sigma^2]$. The "sigma" we denoted shall be the value of $\sigma$, and the unit should be pixels. We have added a relevant description in this part.

Action:
This part has been revised more precisely as: "*We employed Gaussian blurry kernel: $g = \exp[-(x^2+y^2)/2\sigma^2]$ to simulate the blurring. In the first condition, we add Poisson-Gaussian mixed noise at different levels to the GT image, after slightly blurring the GT image with a Gaussian kernel with a σ value of 1.0 (pixel), and train DL models to transform the degenerated image to the GT. In the second condition, we blur the GT image with a Gaussian kernel with a σ value of 2.0-4.0 (pixel) with moderate Poisson-Gaussian mixed noise, ...*"

17) "Three commercially available fixed slices were used for imaging, including a mouse kidney section (FluoCells Prepared Slide #3, Invitrogen, F24630), a homemade stem of the Nerlum sample."
Only two of these appear to be commercial available.
Response:
Thanks for your feedback. We have corrected this error.
Action:
We have corrected this error on Page 28 in the revised manuscript as: "*Three commercially available fixed slices were used for imaging, including a mouse kidney section (FluoCells Prepared Slide #3, Invitrogen, F24630), an onion epidermal cell slice (Sagaoptics, China), and an nerlum indicum stem slice (Sagaoptics, China)*".

18)
page 26
"in 35 mm class bottom" should be "glass bottom" - repeated elsewhere as well.
Response:

Thanks for pointing out this error. We have corrected this error and checked other parts throughout the manuscript.

19)
"After washing with 1× PBS for 3 times" - should be "After washing with 1× PBS, 3 times" - the additional "for" is also in several other places
Response:
Thanks for pointing out this error. We have modified "After washing with 1× PBS for 3 times" to "After washing with 1× PBS, 3 times" throughout the manuscript.

20)
"The secondary antibody was labeled as previously described." Reference needed, also they don't say anything about the DNA probes used. Are they commercial, described in the not included reference or something else?
Response:
Thank you for pointing this out. We apologize for the lack of detail in the original manuscript. Specifically, the secondary antibodies were conjugated to 5′-thiol-modified DNA strands via thiol–maleimide chemistry following a previously established protocol [*Schnitzbauer et al., Super-resolution microscopy with DNA-PAINT, 2017, Nature Protocols*; reference added]. The DNA docking strand used for antibody conjugation had the sequence 5′-TTGATCTACAT-3′ (5′-thiol modified). The complementary imager strand was labeled with Cy3b and had the sequence 5′-TATCTAGATC-3′. Cy3b-labeled DNA imager strands were synthesized commercially (Thermo Fisher Scientific), whereas the antibody–DNA conjugation was performed in-house according to the published protocol.
Action:
We have added citation [*Schnitzbauer et al., Super-resolution microscopy with DNA-PAINT, 2017, Nature Protocols*] to this protocol.

21) "Airy-NovaSD (Bejing, China) spinning- disk confocal microscopy to capture" Should be "microscope"
Response:
Thanks for your feedback. We have corrected this error and checked other parts of the paper to avoid similar errors.

22) "In previous assessments, 128×128 patches with an average photon count of 25-600 are selected in the calculation of image metrics, and the calculation method was also described"
Is the average photon count per patch or per pixel, it is not clear.
Response:
Thanks for your feedback. We are sorry for this ambiguity. The average photon count is calculated per patch based on the established protocol from the BioSR dataset paper [*Qiao et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy.* 2021, *Nature Methods*]. According to your

feedback, we have added additional explanation in this sentence to avoid possible confusion.

Action:

We have modified this description to: "*In previous assessments, it calculated the average photon count per patch (128×128-size). The patches with an average photon count of 25-600 were selected in the calculation of image metrics, and the calculation method was also described*[16]."

Reviewer #2 (Remarks to the Author):

Article summary:

In this paper, the authors discuss a new ML method for image restoration. This method leverages scale separation by synthesizing the outputs of small, deep convolutions and large, shallow convolutions. They argue that the latter element—through a much larger effective receptive field—is better suited to image restoration tasks by taking global statistics into account. They demonstrate this method to produce a 0.5-2dB increase in SNR compared to canonical method (U-Net), as well as a significant 4-20x increase in computational efficiency. They argue that the increased processing speed makes this method better suited to large timeseries datasets.

Response:

Thanks for your kind summary of our manuscript. Your constructive suggestions below help us to improve the quality of the manuscript a lot, which we deeply appreciate.

Regarding the performance comparison, we would like to gently clarify a minor point in the summary to avoid potential confusion. The 0.5–2 dB PSNR improvement and 4–20× computational efficiency gain referenced in our study are not compared specifically to U-Net, but rather to state-of-the-art convolution-based and Transformer-based networks—such as the DFCAN and UniFMIR in the high-profile BioSR single-image-super-resolution task.

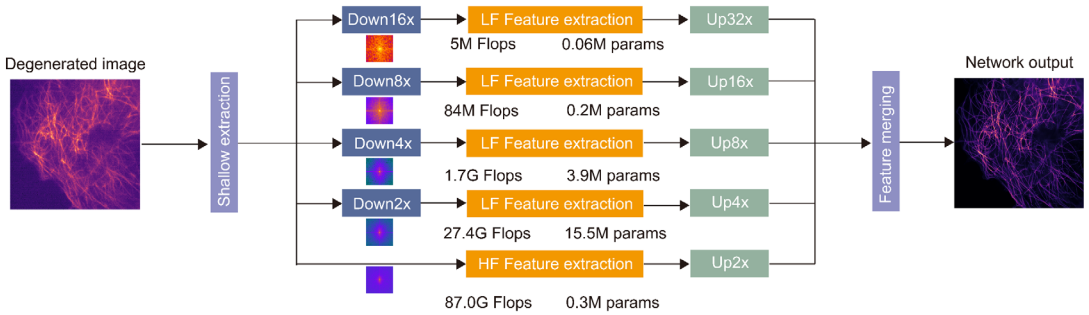
We have carefully addressed all suggestions below and believe the revisions significantly strengthen the paper. Thank you again for your thorough review.

Overall comments:

Overall, this manuscript offers an interesting discussion of effective receptive fields and their effect(s) on the accuracy of network predictions for image restoration. The authors test a new architecture—LargePNet—based on the synthesis of low- and high-frequency feature extraction steps that act on large (512x512) patches. The authors demonstrate that LargePNet offers substantial increases in computational efficiency and processing time compared to current methods such as the commonly used UNet architecture. This is an important consideration for users! The manuscript would be strengthened by a discussion (or expanded discussion) of how these efficiency gains are achieved.

Response:

Thanks for your kind summary and constructive suggestion to improve our manuscript. The efficiency gain originates from our scale separation design and avoidance of the redundancy caused by over-deep network design. Let’s recall the LargePNet design:




Response Fig. 2-1. LargePNet architecture design

In LargePNet design, the branch that processes the full resolution feature (HF Feature extraction) is composed of the reparametrized-large-kernel-convolution (RepLKConv). We only need a total of 4 RepLKConv layers (composed of 4 LKConv layers and 19 small-kernel-conv reparameterization layers) to accomplish a satisfactory ERF for large-patch (512x512) training. This branch occupies the largest computational amount in LargeNet. Since it only contains 23 trainable layers in total, it is much more efficient than some advanced networks with very deep layers. For example, DFCAN has 83 trainable layers, UniFMIR has 108 trainable layers. In these models, attention operations are repeatedly applied to achieve good performance, further increasing their computational burden.

Despite RepLKConv offering an efficient way to leverage the global information and preserve the high-frequency features, its shallow depth also has drawbacks. Notably, it lacks sufficient nonlinear representation capability, which is required in the image restoration task. The nonlinearity is essential to remove the random noise and other artifacts in images, meaning that the network needs deep layers to gradually transform the degenerated images. Therefore, in the LargePNet design, we compensate for this drawback of RepLKConv by adding several LF Feature extraction branches. These LF branches process features with multiple down-sampling rates, where the averaged local pixels aid the removal of noise/artifacts, and simultaneously reduce the computational cost. These branches are processed by deep networks, which can provide sufficient nonlinear capability to remove noise/artifacts.

Besides the efficiency in architecture, the training view (512 pixels or larger) of LargePNet matches the view that is used in practical applications (512 pixels or larger). As a result, LargePNet can directly infer the full view image with uncompromised performance. While for other small-patch networks, it may need to infer overlapped small patches and then stitch them to a final full view, to ensure the detail quality (as shown in Fig. 2). This further highlights the efficiency improvement of LargePNet in practical application.

In summary, LargePNet's efficiency gains arise from using deep networks to low-resolution branches (where it is the most effective and cheapest in computational burden) and using the efficient shallow RepLKConv for high-resolution branches (which builds a large ERF and aggregates global information). This scale-adaptive division and rationale choice of network depth achieved these efficiency gains. Its large-view training and inference consistency further improve its efficiency in practical application (without the need for stitching small views).

Action:

We have added a relevant discussion about how these efficiency gains are achieved in LargePNet. The above discussion has been added in Supplementary Note 3 in the revised Supplementary Information.

However, while there are some instances and some metrics for which LargePNet outperforms current methods, there is not substantial evidence that this new architecture offers significant general performance gains. Moreover, any apparent disparities in performance appear to be dependent on the biological structure present in the images. This manuscript would be strengthened by a more quantitative statistical presentation of comparative performance gains for LargePNet relative to current methods. Moreover, a discussion of why LargePNet performs better/worse when restoring the images of different biological structures would be of interest.

Response:

Thanks for your insightful comment, which helps improve our manuscript. Since this comment contains several issues, we answer the comment in a point-by-point manner:

**(1) General performance gains**

We may first explain that this paper focuses on the fluorescence imaging field. The goal of using fluorescence excitation is to mark a single biological structure of interest. This is different from medical imaging methods (e.g., CT, MRI), where the captured image contains many structures without specificity. Therefore, the images in the fluorescence image dataset are all marked with a single biological structure. For example, the high-profile open-source BioSR dataset only contains 4 structures: CCP, ER, MT, and F-actin. Our experiments already cover all four structures in the datasets, which demonstrate the general ability in the fluorescence image field. The features of these structures are usually considered sufficient for generality in all fluorescence imaging structures.

**(2) Statistical significance of LargePNet's improvement in different structures**

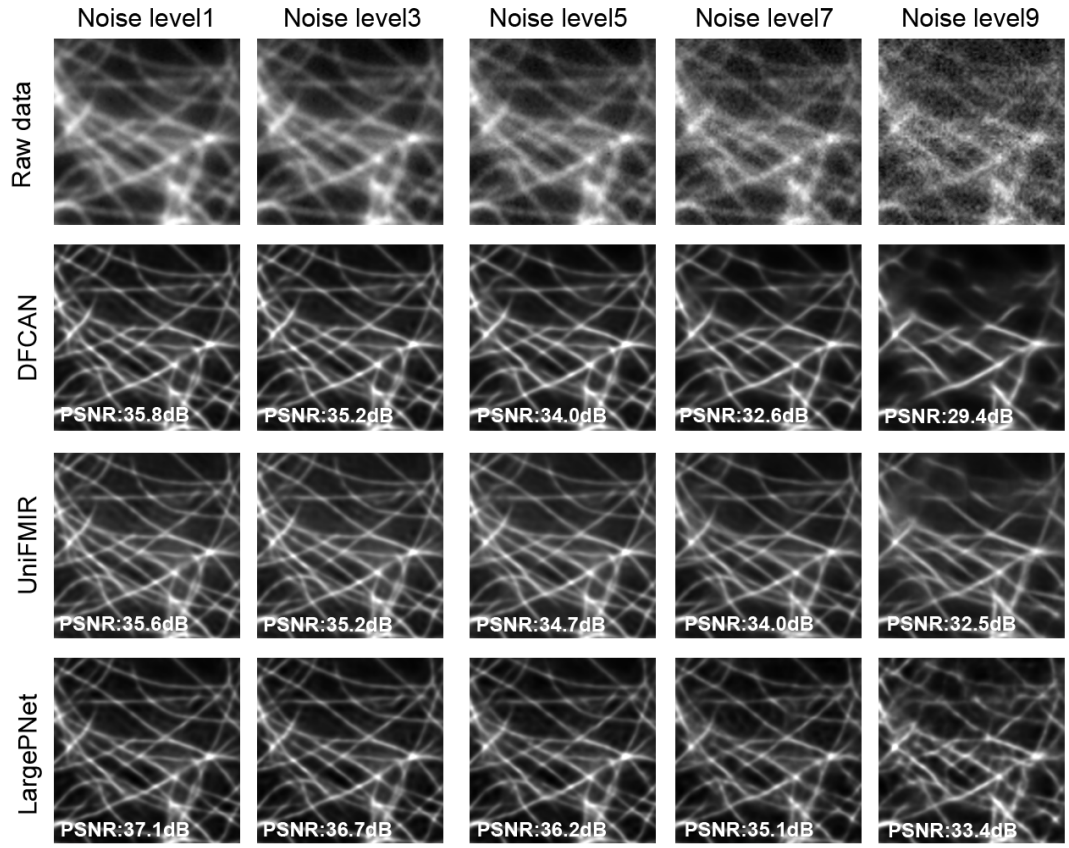
We are sorry that previously we ignored this issue. We have performed the significance test on all presented box charts in the paper to check the significance of LargePNet's improvement in quality metrics through t-test (we implement the test using the *ttest* function in MATLAB). The results show that LargePNet outperforms other models in all tasks and all structures with very high significance ($p < 0.001$, ***) in the PSNR and NRMSE metrics. The significance level of the LargePNet improvement is not that sensitive to different structures. As an example, here we put the significance test results of all four structures in the BioSR datasets. Other results can be seen in our following response to your Figure-specific Questions and Comments. The source data are all uploaded in the .xlsx format to ensure reproducibility.

Response Table 2-1. Significance test of the PSNR metrics in the BioSR super-resolution dataset

| Significance of LargePNet improvement through t-test | | | | | |
|---|---|---|---|---|---|
| Model | Metric | CCPs | ER | Microtubules | F-Actin |

| | | | | | |
|---|---|---|---|---|---|
| Versus DFCAN | PSNR (dB) | $p$ = 1.50e-67(***) | $p$ = 4.22e-15(***) | $p$ = 3.82e-71(***) | $p$ = 1.01e-33(***) |
| Versus UniFMIR | PSNR (dB) | $p$ = 1.49e-26(***) | $p$ = 6.41e-5(***) | $p$ = 8.54e-85(***) | $p$ = 2.23e-5(***) |

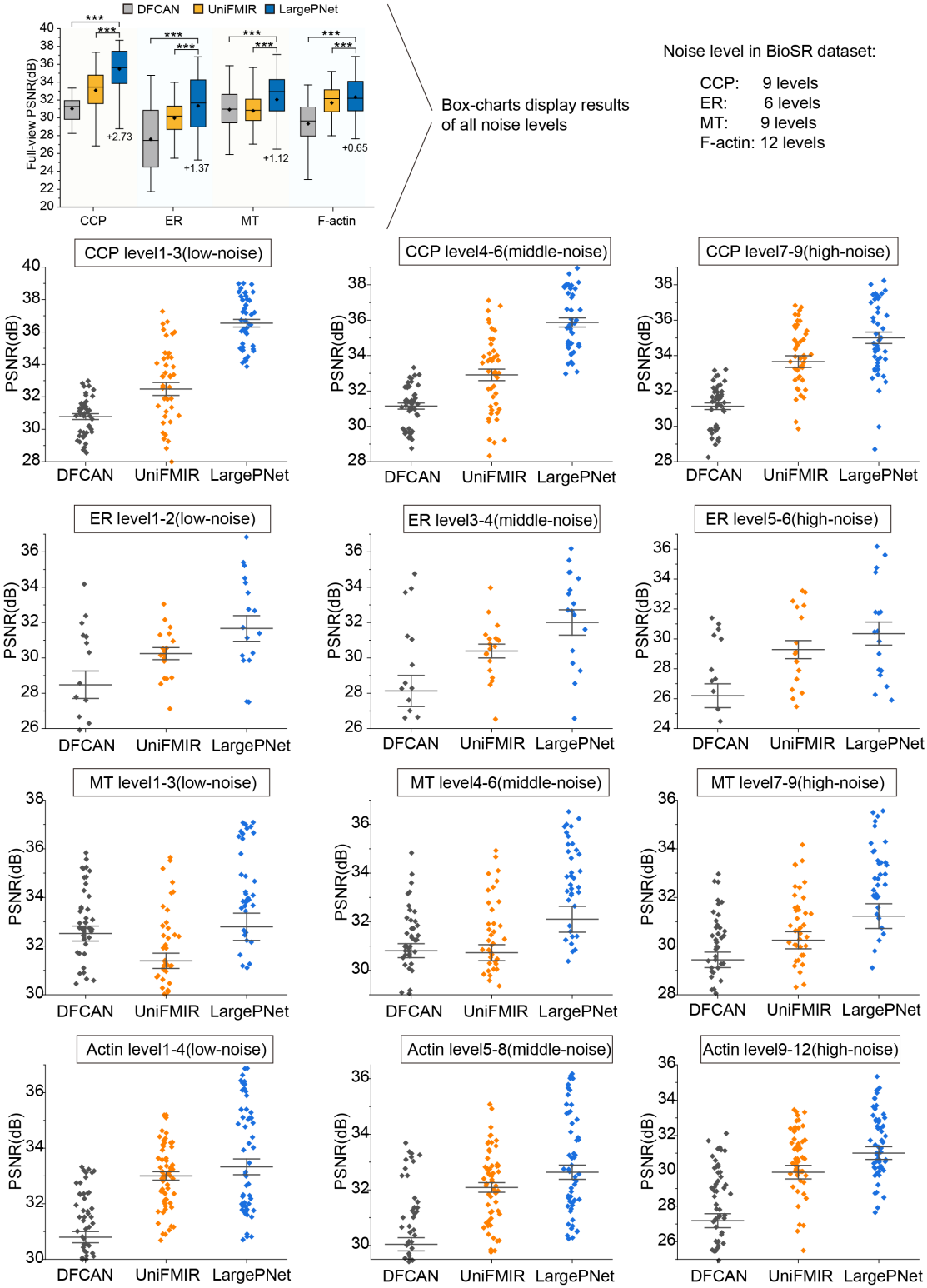
We may explain that the box chart display method may leave a superficial impression that the improvement is not significant. This is because the box displays the results with all the noise levels or degeneration levels. In the datasets such as BioSR, the same FOV is collected with multiple noise levels (e.g., F-actin in BioSR datasets is collected with 12 noise levels for a single FOV). All model performance will sensitively decrease with the increase in the noise level. An example is shown in Response Fig. 2-2 below. The deviation of the restored PSNR across different noise levels (~6 dB) is more evident than the model difference (~1.5 dB). As a result, there is a large overlap of the box displays of different models, and the whiskers of all box charts are relatively long. This may make the improvement of LargePNet seem insignificant in the box chart display.




Response Fig. 2-2. All models show evident performance degradation with the raw image noise levels. This deviation causes the long whisker and the overlap of the boxes of different models in the box chart display. This obscures the improvement of LargePNet over other models.

Although the box chart provides a good presentation of the quality metrics across all raw image noise levels, it fails to present the restoration metric's connection with the noise levels. If we display the results of some adjacent noise levels separately, as shown

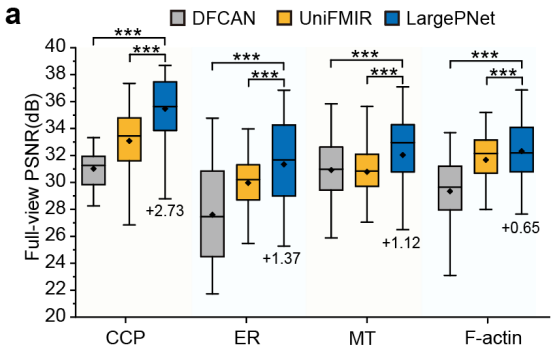
in Response Fig. 2-3 below, or visually following Response Fig. 2-2, you can see that the improvement of LargePNet is indeed appreciable. Usually, when the average value in this box chart (denoted by the diamond point) is perceivably higher, it indicates that the global improvement is very significant.




Response Fig. 2-3. Statistical comparisons in the BioSR super-resolution dataset with relatively separated noise levels. Each point displays the metric of an image. The error

bar displays the average value and standard error (SE).

Following your advice, we have strengthened the paper by marking the quantitative statistical significance result in all the comparative box charts. The exact improvement value by LargePNet compared with the second-best model is also marked in the box chart for clearer presentation. An example is shown below. You can find other results in our following response to your Figure-specific comments and questions.




The optimized box chart display in Fig. 2a

We very much appreciate that the reviewer proposed this constructive suggestion. This has greatly helped strengthen our manuscript. We hope that through our statement of statistical significance in this paper, the researchers in this field can also follow and set this to avoid possible confusion in the box chart display results.

Action:

We added significance test to all the statistical comparisons. The results are all marked to the box chart display with star numbers. The exact improvement value of LargePNet compared with the second-best model is also marked for clearer presentation.

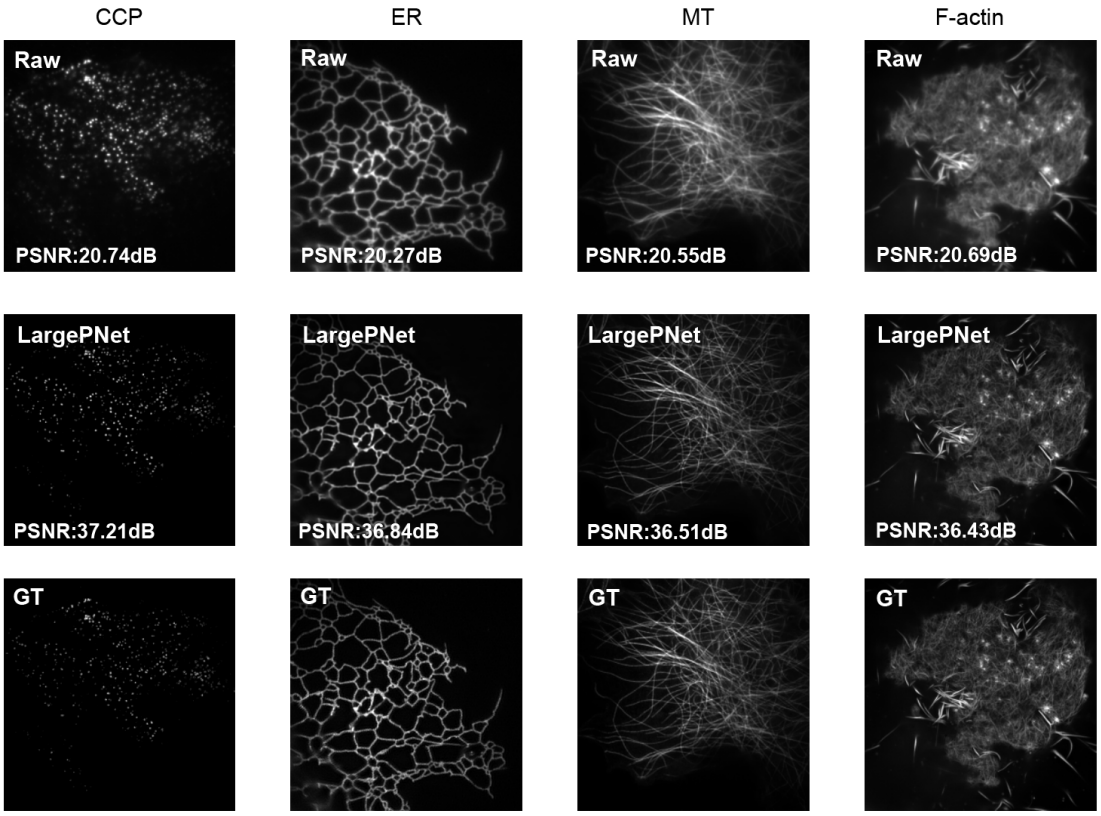
The potential drawbacks of box chart display for obscuring the significance are discussed in the Methods section in the revised manuscript as follows:

"*As presented in this paper, the boxes and whiskers are typically relatively long. This is because restored image datasets usually acquire different levels of SNR for a single FOV (e.g., the F-actin dataset in BioSR includes 12 noise levels). The performance of all models degrades significantly as the SNR of the original images decreases, which results in the elongated boxes and whiskers. In addition, although box plots effectively present the fluctuations of restoration metrics across samples with varying SNRs, they may obscure the performance differences between different models. Statistical charts stratified by different noise levels could help to better understand the performance of various models under different initial SNR conditions of original images. Some illustrative examples are provided in Supplementary Figs. 6, 9.*".

**(3) Why LargePNet performs better/worse when restoring the images of different biological structures**

Thanks for your constructive suggestion. Firstly, it should be noted that the most important factor affecting the restoration performance in fluorescence images, as demonstrated in the above discussion, is the SNR of the raw image rather than its

structure. This is because in the restoration task, the original image of the fluorescence image and the GT image already have roughly similar contours, and the network only learns how to remove degradation (e.g., noise and blur) from the raw image. The degradation process itself has little correlation with the structure. This may be different from the situation where there are significant differences between different structures in fields such as object detection or segmentation. In addition, even if the parameters, such as laser power, are the same when collecting data from different structures, it is difficult to ensure that their SNR is the same because the excitation efficiency of different fluorescent dyes is quite different. This makes it difficult for us to strictly separate the effects of degradation and structure in analysis. We provide an example restoration result in the BioSR single-image super-resolution dataset, where we select images of different structures with comparable raw PSNR (Response Fig. 2-4). It can be seen that the LargePNet restored PSNR is relatively close in different structures. The structure-induced PSNR difference (~0.8 dB) is evidently lower than the difference induced by noise level in the raw images (~6 dB, as previously shown in Response Fig. 2-2).



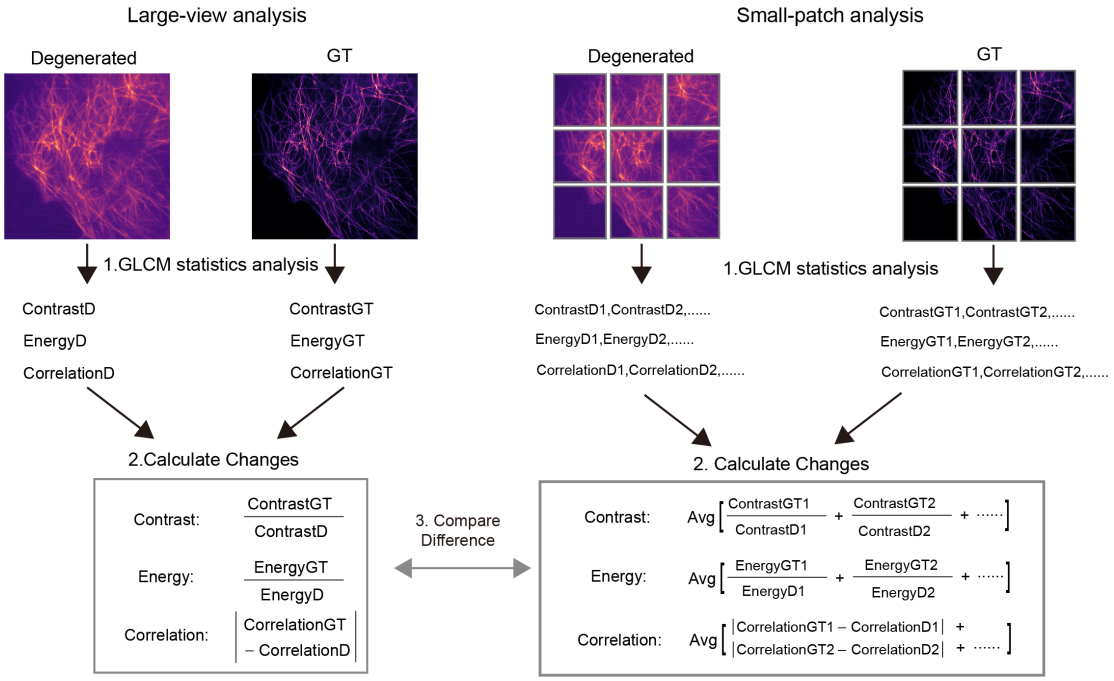

Response Fig. 2-4. Inspection of the model performance in restoring different structures when the raw degenerated images have close PSNR values.

Nevertheless, we appreciate the reviewer's question on which biological structure benefits most from LargePNet compared to other models. Analyzing this issue through quantitative evidence will make the manuscript more comprehensive and compelling. In addition, it should be pointed out that LargePNet has shown significant improvement in the PSNR metric in all structures in statistical significance ($p < 0.001$, ***). Here,

we use the increment of PSNR value to evaluate the degree of LargePNet's advantage.

The main difference between LargePNet and other networks lies in the aggregation of the large-view information. Therefore, we examine when LargePNet performs much better by analyzing the differences in structural statistical information between large and small view images. We found that this indicator can explain most of our experimental results well: when the structural statistical information of the small-view deviates greatly from that of the large-view, LargePNet will achieve greater improvement in the restoration performance. When the structural statistical information of the small-view is closer to that of the large-view, the performance improvement of LargePNet will be limited. The analysis method is as follows:

We employ the Gray-Level Co-occurrence Matrix (GLCM) to analyze the structural statistics information in an image. The statistics of GLCM measure the texture information and structural connection in an image. Here we employ the commonly used Contrast, Energy, and Correlation statistics metrics of GLCM, which can be easily obtained via the MATLAB inner function *graycomatrix* and *graycoprops*. It is worth noting that in image restoration, due to the roughly similar contours between the raw and GT images, the statistics of the structure itself may not be suitable to explain the performance of LargePNet. How the structure is disrupted in the view of large and small images (i.e., the difference between the raw and GT images) is a more reasonable explanatory factor. Based on this rationale, we develop the calculation process for explaining the structure-dependent advantages of LargePNet using GLCM, as shown in the Response Fig. 2-5.




Response Fig. 2-5. Calculation process for explaining the structure-dependent advantages of LargePNet.

The first step is to analyze the GLCM statistical metrics of all images that belong to a certain structure before and after degradation. This analysis is conducted separately in large-view images and small-view images. The second step is to calculate the

changes in GLCM statistical metrics before and after degradation. For GLCM's Contrast and Energy metrics, which are obtained by addition or subtraction, we focus on their fold differences. For the Correlation metric, we consider the absolute value difference in their values. This analysis was also conducted separately in the large-view images and the small-view images. The third step is to compare the degree of change in GLCM statistics obtained from the large-view and those obtained from the small-view. By comparing the PSNR increments in various experiments, we found that typically, the larger the difference in GLCM statistics changes between the small and large images, the greater the PSNR increment obtained by LargePNet.

Based on this method, we analyze the difference in the structural advantage of LargePNet throughout the tasks that appear to show structural differences in the paper:

- BioSR single-image super-resolution task (Fig. 2).

Analysis results of this dataset are shown in Response Table 2-2. It can be seen that LargePNet has a relatively lower PSNR gain in the F-actin compared to other structures. The GLCM statistics show that there are more significant differences in statistics between the large and small views of other structures: the differences of the GLCM Correlation metric between the large-small views of CCP, ER, and MT all exceeded 0.03; the difference in Energy statistics of CCP reached 2.8-fold, and MT reached 1.5-fold. The large-small view differences in F-actin on these statistical indicators are not as significant as in other structures. Mapping these statistical indicators onto the image, F-actin is a very complex structure, and its local regions may already contain rich information about the structure and degradation process. Therefore, the structural information lost by small-patch truncation may not be as much as other structural losses. This could be the reason why LargePNet provides a lower PSNR gain in F-actin compared to other structures.

Response Table 2-2. Explaining structure-dependent advantage of LargePNet in the BioSR single-image-super-resolution task using the GLCM analysis method.

| LargePNet PSNR gain compared with the best small-patch model | | | | |
|---|---|---|---|---|
| Structure | CCP | ER | MT | F-actin |
| PSNR(dB) gain | **+2.73dB** | **+1.37dB** | **+1.12dB** | +0.65dB |
| GLCM Contrast statistics difference analysis | | | | |
| Large-view | 1.54-fold | 3.28-fold | 1.76-fold | 2.08-fold |
| Small-view | 1.55-fold | 3.29-fold | 2.17-fold | 2.47-fold |
| Large-small Diff | 1.0-fold | 1.0-fold | 1.2-fold | 1.2-fold |
| GLCM Energy statistics difference analysis | | | | |
| Large-view | 0.185-fold | 0.168-fold | 0.280-fold | 0.352-fold |
| Small-view | 0.065-fold | 0.135-fold | 0.189-fold | 0.319-fold |
| Large-small Diff | **2.8-fold** | 1.2-fold | **1.5-fold** | 1.1-fold |
| GLCM Correlation statistics difference analysis | | | | |
| Large-view | 0.0061 | 0.2351 | 0.0031 | 0.0334 |
| Small-view | 0.0379 | 0.2710 | 0.0363 | 0.0555 |

| | | | | |
|---|---|---|---|---|
| Large-small Diff | **0.0318** | **0.0359** | **0.0332** | 0.0221 |

- STED denoising task (Fig. 3).

Analysis results of this dataset are shown in Response Table 2-3. It can be seen that LargePNet has achieved greater advantages over other structures in the Mito IMM structure. By examining the GLCM statistic, it can be found that Mito IMM has a difference of up to 3.17-fold between the large and small views in the Energy metric; the difference between the large and small views in the Correlation metric is as high as 0.0698. These significant differences in GLCM statistical indicators reflect that it lost more complete structural information about Mito IMM and inter-regional connections with small patches. This may explain why LargePNet achieves greater advantages in Mito IMM compared to other structures.

Response Table 2-3. Explaining structure-dependent advantage of LargePNet in the STED denoising task using the GLCM analysis method.

| LargeP-TISR PSNR gain compared with the best small-patch model | | | |
|---|---|---|---|
| Structure | MT | Mito IMM | ER |
| PSNR(dB) gain | +0.25dB | **+0.56dB** | +0.36dB |
| GLCM Contrast statistics difference analysis | | | |
| Large-view | 7.15-fold | 5.79-fold | 12.97-fold |
| Small-view | 6.28-fold | 5.82-fold | 8.07-fold |
| Large-small Diff | 1.14-fold | 1.00-fold | 1.61-fold |
| GLCM Energy statistics difference analysis | | | |
| Large-view | 5.23-fold | 7.58-fold | 1.84-fold |
| Small-view | 4.76-fold | 2.39-fold | 1.17-fold |
| Large-small Diff | 1.10-fold | **3.17-fold** | **1.57-fold** |
| GLCM Correlation statistics difference analysis | | | |
| Large-view | 0.1018 | 0.0497 | 0.8172 |
| Small-view | 0.1530 | 0.1195 | 0.8134 |
| Large-small Diff | **0.0512** | **0.0698** | 0.0038 |

- BioTISR video super-resolution (Fig. 4)

Analysis results of this dataset are shown in Response Table 2-4. It can be seen that LargePNet has achieved greater advantages over other structures in the Mito OMM structure compared with MT. By examining the GLCM statistic, it can be found that the Mito IMM structure has a much more evident large-small view difference compared with the MT structure. Its large-small view difference in the Contrast metric is 6.85-fold, in the Energy metric is 4.16-fold. This means that in small-patch cropping, the Mito OMM structure loses more structural contextual information. It may explain why LargePNet achieves greater advantages in Mito OMM compared to MT.

Response Table 2-4. Explaining structure-dependent advantage of LargePNet in the BioTISR video super-resolution task using the GLCM analysis method.

| LargeP-TISR PSNR gain compared with the best small-patch model |
|---|

| Structure | MT | Mito OMM |
|---|---|---|
| PSNR(dB) gain | +0.50dB | **+1.18dB** |
| GLCM Contrast statistics difference analysis | | |
| Large-view | 4.29-fold | 2.22-fold |
| Small-view | 4.37-fold | 3.09-fold |
| Large-small Diff | 1.02-fold | **6.85-fold** |
| GLCM Energy statistics difference analysis | | |
| Large-view | 2.16-fold | 3.28-fold |
| Small-view | 3.10-fold | 1.27-fold |
| Large-small Diff | 1.43-fold | **4.16-fold** |
| GLCM Correlation statistics difference analysis | | |
| Large-view | 0.1652 | 0.0062 |
| Small-view | 0.1286 | 0.0225 |
| Large-small Diff | 0.0366 | 0.0395 |

- STED deblurring dataset. (Extended Data Fig. 5)

Analysis results of this dataset are shown in Response Table 2-5. It can be seen that LargePNet has achieved greater advantages over other structures in the Mito IMM structure compared with MT. Inspecting the GLCM statistics, it can be found that the Mito IMM shows an evidently higher large and small views difference (2.02-fold in Energy metric, 0.0652 in Correlation metric) compared with MT (1.47-fold in Energy metric, 0.0184 in Correlation metric). Mapping this indicator to the practical image, this task aims to remove the blurring degeneration in STED images. For MT, as a relatively simple structure, small-patch cropping may not disturb the contextual structural information as the complex Mito IMM structure does. While for the Mito IMM structure, which is relatively complex, the advantage of LargeP-GAN's large-view statistics aggregation will exert, and cause a significant performance difference.

Response Table 2-5. Explaining structure-dependent advantage of LargePNet in the STED deblurring task using the GLCM analysis method.

| LargeP-GAN PSNR gain compared with the best small-patch model | | |
|---|---|---|
| Structure | MT | Mito IMM |
| PSNR(dB) gain | +0.42dB | **+1.16dB** |
| GLCM Contrast statistics difference analysis | | |
| Large-view | 4.80-fold | 1.02-fold |
| Small-view | 3.27-fold | 1.41-fold |
| Large-small Diff | 1.47-fold | 1.38-fold |
| GLCM Energy statistics difference analysis | | |
| Large-view | 0.085-fold | 0.155-fold |
| Small-view | 0.058-fold | 0.313-fold |
| Large-small Diff | 1.47-fold | **2.02-fold** |
| GLCM Correlation statistics difference analysis | | |
| Large-view | 0.0177 | 0.0961 |

| | | |
|---|---|---|
| Small-view | 0.0361 | 0.0309 |
| Large-small Diff | 0.0184 | **0.0652** |

Through the above analysis, we can see that GLCM analysis of the difference in information between small and large view images can qualitatively explain why LargePNet has different performance advantages on different structures. We appreciate you bringing up this point and helping to improve our manuscript.

Action:

We added discussions about why LargePNet performs better/worse when restoring the images of different biological structures in the Discussion section. Supplementary Note 6 and Supplementary Fig. 18 are added to illustrate the calculation methods and analysis of the results, which are listed in Supplementary Tables 11-14.

With that said, it is more clear that LargePNet outperforms the current methods when restoring higher-dimensional imaging data such as time series and volumes. The manuscript would be strengthened by discussing potential reasons for this improvement.

Response:

Thanks for your insightful question. The timeseries (the open-source BioTISR dataset) and volumes data (self-collected SD-confocal dataset) were not collected with multiple noise levels as in single-image datasets. In time series acquisition, the samples are constantly moving, making it difficult to capture videos with multiple levels of signal-to-noise ratio matching in strictly the same FOV. For Fig. 4g-k, we expect to examine the network's ability to remove background noise present in 3D fluorescence imaging instead of removing random noise under low light or short exposure conditions. Therefore, volume data were not collected with multiple signal-to-noise ratios.

As a result, the box charts displaying the timeseries and volumes data seem more compact than the single-image data. This may make it superficially clearer that LargePNet outperforms the current methods when restoring higher-dimensional data. Please see the statistical presentation of the data with separated noise levels in the response to the above comment and Response Fig. 2-3.

The microscopy methods are generally underreported. Please make sure for all image data collected the objective specifications, spatial and temporal sampling (ex. Pixel size and time interval), laser lines, detectors, and post-processing algorithims (ex. SIM reconstruction) are reported.

Response:

Thanks for your constructive suggestion to improve our manuscript. We added relevant Supplementary Tables summarizing this information in our manuscript, as shown below:

Response Table 2-6. The microscopy methods in the collected datasets

| **Self-collected imaging dataset** | | | | | | |
|---|---|---|---|---|---|---|
| **Dataset** | **Structure** | **Imaging setup** | **Sampling** | **Ex (nm)** | **Objective** | **Detectors** |

| | | | | | | |
|---|---|---|---|---|---|---|
| STED (Fig. 3, Ex Fig. 5) | Microtubule | Leica Stellaris 8 | 25-55nm | 610 nm | 100×/1.40 HC PL Apo Oil STED | Leica HyD hybrid detector |
| | Mitochondria | | 25-55nm | 610 nm | | |
| | ER | | 25-55nm | 610 nm | | |
| SMLM (Fig.3) | Microtubule | NovaSD | 15 nm | 488 nm | CFI SR Apochromat TIRF 100X oil, NA 1.49 | sCMOS |
| SD-confocal (Fig. 4) | Nerlum | NovaSD | 108 nm lateral 800 nm z step | 610 nm | CFI Plan Fluor 20×/0. 5 NA | sCMOS |
| | Onion | | 108 nm lateral 800 nm z step | 610 nm | | |
| | Mouse WGA | | 325 nm lateral 800 nm z step | 488 nm | CFI Plan Fluor 10×/0.25 NA | sCMOS |
| | Mouse Phallodin | | 325 nm lateral 800 nm z step | 561 nm | | |
| **Open-sourced imaging dataset** | | | | | | |
| **Dataset** | **Structure** | **Imaging setup** | **Sampling** | **Ex (nm)** | **Objective** | **Detectors** |
| BioSR SIM (Fig. 2, Ex. Fig.6) | CCP | GI-SIM or TIRF-SIM | 31.3 nm | 488 nm | APON100XH OTIRF 1.7 NA, Olympus | sCMOS |
| | ER | | 31.3 nm | 488 nm | | |
| | Microtubule | | 31.3 nm | 488 nm | | |
| | F-actin | | 31.3 nm | 488 nm | | |
| BioTISR SIM (Fig. 4) | Microtubule | Multi-SIM | 30.6 nm spatial 0.25 s temporal | 488 nm | CFI SR Apochromat TIRF 100X oil, NA 1.49 | sCMOS |
| | Mitochondria | | 30.6 nm spatial 0.5 s temporal | 488 nm | | |

Note: the self-collected SMLM data require reconstruction, which is performed using the commercial Huygens Localizer software. The open-source BioSR and BioTISR SIM data already contained the reconstructed SIM data. The reconstruction is performed through the home-written SIM reconstruction algorithms by the dataset establishers. Please find details in the data link and their papers.

Action:

We add the Supplementary Table 16 summarizing the microscopy methods.

General Comments:

The authors make repeated claims on the effect/shortfalls of small effective receptive

field of architectures such as U-Net but only cite a single analysis of this fact. In one contrasting example, 10.1117/1.JMI.11.5.054004 suggests that larger true receptive fields (TRF, and by extension, the ERF) are not always correlated with superior network performance in the context of segmentation. This is not to claim that this is also true for image restoration tasks, but a deeper source of support for this claim would strengthen the manuscript. Moreover, this paper also makes the point that different sized structures benefit from differently sized TRF, which could plausibly also apply to restoration tasks.

Response:

Thanks for your insightful comment, which helps improve the description of our manuscript.

We are sorry that our repeated claims may have misled you. These descriptions should be refined to avoid ambiguity. Our essential claim is that "discarding large-view statistics" but not exact "small ERF" is the shortfalls of the small-patch network. Our "large-view" here means a view ≥ 512*512 pixels, which is not completely identical to "large ERF".

Large ERF is one essential component to aggregate large-view information, because a small ERF means that the model does not take a large view into account. Another essential component is the stable training capability on large images. In this sense, previous models like UNet cannot fulfill large-view information aggregation due to: (1) too-small ERF to cover the large-size view and (2) instability of training on large-size images (as demonstrated in Ex. Data Fig. 2). The difference of LargePNet over them is not just large-ERF, but the specially designed model architecture (RepLKConv, scale separation, and instance normalization) for stable training on large-size images.

We appreciate you very much for providing a wonderful literature [10.1117/1.JMI.11.5.054004] for us to learn. We carefully studied it, and definitely agree with your idea and also the results in 10.1117/1.JMI.11.5.054004, that a larger ERF does not always mean a better performance. Since LargePNet has a different training view-size over small-patch models, this conclusion could be made more rigorously as: "Within a fixed-size patch, a larger ERF does not always mean a better performance". This phenomenon can also be observed in our LargePNet model.

The ERF of LargePNet can be tuned by modifying the kernel size of the RepLKConv operations in LargePNet. A larger RepLKConv kernel size will lead to a larger ERF, as demonstrated in [*Ding, X. et al. Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs. CVPR (2022)*]. We test the performance of LargePNet with different ERFs in the BioSR datasets; the results are shown in the Response Table 2-7 below.

Response Table 2-7. PSNR(dB) metrics of LargePNet with different ERFs on the BioSR dataset.

| Structure | LargePNet-11 | LargePNet-13 | LargePNet-15 | LargePNet-17 | LargePNet-19 |
|---|---|---|---|---|---|
| CCP | 35.5593 | 35.6468 | **35.8077** | 35.7443 | 35.781 |
| ER | 31.4821 | 30.6697 | 31.3432 | **31.7895** | 31.4765 |
| MT | 31.7426 | 31.8876 | 32.0415 | **32.2077** | 31.1861 |

| | | | | | |
|---|---|---|---|---|---|
| Actin | 31.7963 | 31.4313 | **32.3224** | 31.5853 | 30.6305 |

Note: LargePNet-k represents LargePNet with k×k-size RepLKConv. A larger k value leads to a larger ERF.

From the comparative results, it can be seen that: (1) the performance varies with the ERFs and (2) a larger ERF does not always bring better performance. In addition, the sizes of different structures in the fluorescence microscopy dataset are relatively close. Therefore, their respective optimal ERF changes will not be too significant (LargePNet-15 or 17 achieved the best results). This may be different from the segmentation of organs in medical images that have evidently different sizes. We believe this table effectively conveys to the readers that a larger ERF is not always equate to better performance. Thank you for providing us with constructive feedback, which helps improve our manuscript.

Action:

We have discussed the ERF issue in the Discussion section in the revised manuscript. The literature [10.1117/1.JMI.11.5.054004, *Loos V, Pardasani R, Awasthi N. Demystifying the effect of receptive field size in U-Net models for medical image segmentation. J Med Imaging*] has been cited as an analysis evidence. The results in the above table are added as a Supplementary Table as experimental evidence. The content is written as follow:

"*Although we have established a large enough ERF in LargePNet to aggregate large-view information, it should be noted that an even larger ERF is not necessarily better, as shown in* ***Supplementary Table 10****. The optimal ERF may be related to factors such as structure size and other relevant parameters*[54]*.*"

Similarly, the authors claim "inconsistenc[ies] of the patch size during training, evaluation, and final application" are often overlooked. It would be good to add some discussion or citations regarding the perils of these inconsistencies.

Response:

Thanks for your constructive suggestion. We shall discuss this issue and cite some literature to clarify this issue. From our perspective, the perils of these inconsistencies may have two aspects:

(1) While practical fluorescence microscopy images typically span large resolutions (≥512 pixels), most previous models are trained exclusively on small local patches (e.g., 128≥ pixels). Small-patch truncates much contextual structural information, so in the training, the network only learns how to use limited structural contexts to restore the image. However, in practical applications, large images contain richer contextual information, which has not been fully utilized for small-patch networks. Compounding this issue, the potential benefits of direct full-view training for improved global information aggregation remain largely unexplored. Taking more pixels and richer contextual structural information into consideration when doing restoration carries more potential for better restoration results. This view can be supported by a theoretical literature [*Patch Complexity, Finite Pixel Correlations and*

*Optimal Denoising, ECCV, 2012*].

(2) The inconsistency of feature statistics distribution in training patch-size and final application patch-size would degrade the model performance. This phenomenon has been observed and discussed in [*Scale Calibrated Training: Improving Generalization of Deep Networks via Scale-Specific Normalization. CVPR, 2020*; and *Improving Image Restoration by Revisiting Global Information Aggregation. ECCV, 2022*]. In the results presented in our paper, this phenomenon can also be observed. In Fig. 2 and Supplementary Fig. 4, it can be seen that DFCAN-Full provides better quality metrics (PSNR, NRMSE) than DFCAN-Stitching, but DFCAN-Stitching better recovers the detailed structures. This also illustrates that the train-test patch-size inconsistency would degrade the model performance.

Action:

We have added discussion and citations regarding the perils of train-test patch-size inconsistencies.

The content in Figure S1 does not correlate with the figure caption. The authors argue that the histograms from the 4 ROIs of the BioSR image are "identical". However, it is clear that these histograms—particularly comparing ROI1 and ROI2—are distinct. Moreover, from the image, it seems as though an ROI in the dimmer part of the cell (presumably around the nucleus) would yield yet another distinct histogram. Please address these discrepancies.

Response:

Thanks for your insightful question. We apologize that the usage of "identical" is ambiguous here. This description is not precise. We aim to illustrate that fluorescence images share a relatively higher intra-ROI similarity compared with natural images. This statement is made in a comparative manner and does not necessarily mean that the various ROIs in fluorescence images are absolutely similar. More precisely and directly, we shall describe that the histogram of each ROI in fluorescence images shows lower distinction compared with the histogram in natural images. The histograms displayed here are only for leaving a direct visual impression of this issue. In the right subgraphs, we also rigorously calculate the gray-level co-occurrence matrix (GLCM) energy and average gradient to quantify the intra-ROI similarity of the BioSR and DIV2K image. This quantitatively demonstrates that fluorescence images have higher intra-ROI similarity compared with natural images. Thus, leveraging large-view statistics is easier in fluorescence images relative to natural images.

Action:

We have revised the imprecise wording "identical" to "share a relatively higher intra-ROI similarity compared with natural images". Moreover, the graph is marked to avoid possible misunderstanding.

Moreover, the motivation derived from these data is not clear. Even if small ROIs contain ~similar intensity information, and thus, similar statistics, then isn't this an argument for small patch sizes? If any one patch is similar to another, then there would appear to be no need to consider more information. Furthermore, if, as it seems here, it

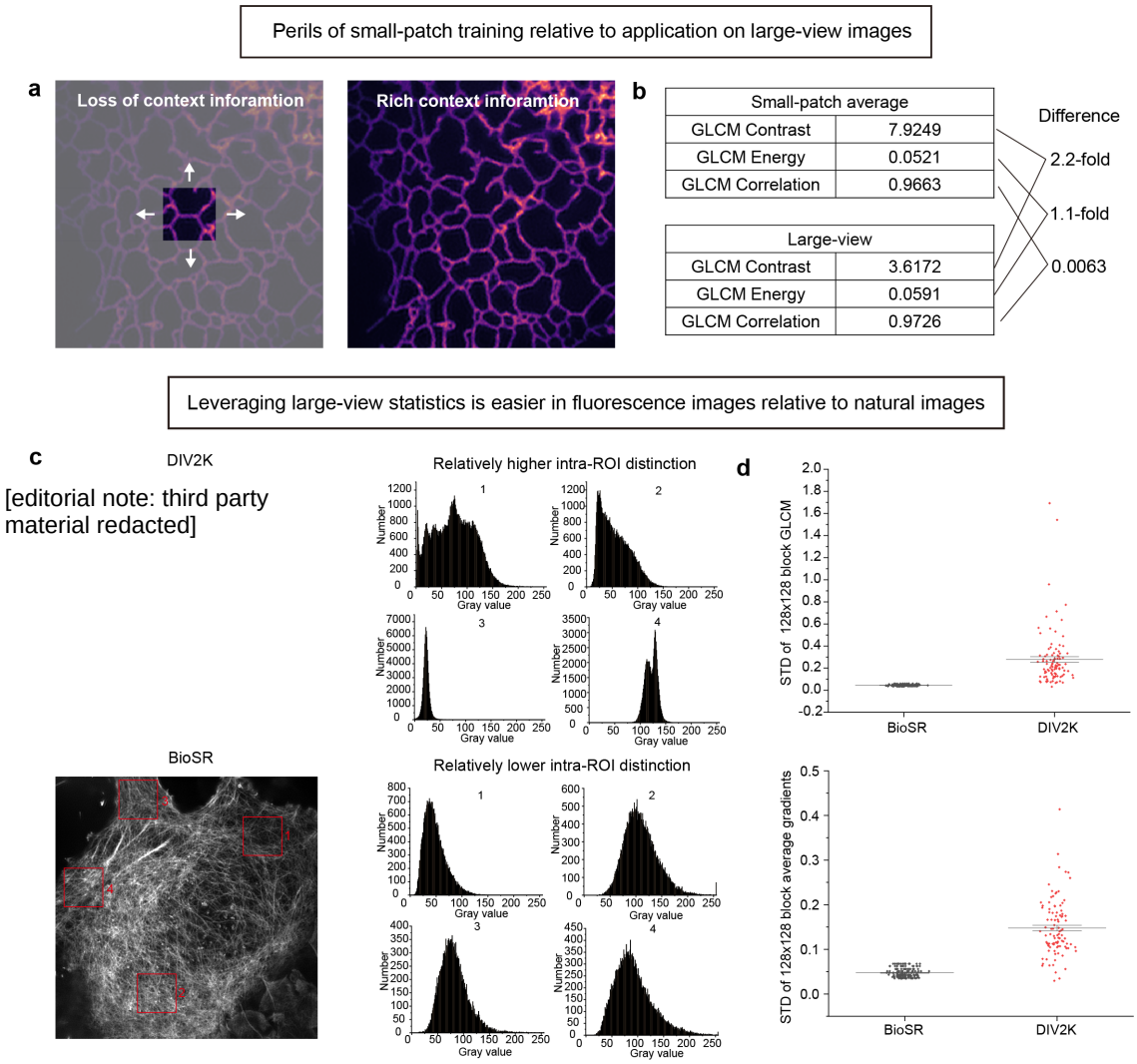
is plausible that biological images in fact contain regions of distinct spatial patterns, isn't that also an argument for small patches, as one would want the network to learn these features independently? It would seem that using large patches could obfuscate some of these more granular details. A more thorough explanation of these topics is warranted.

Response:

Thanks for pointing out the drawback of this figure. We apologize that this figure did not truly convey our motivation. We agree with you that the similarity/dissimilarity of ROIs is not sufficient proof to judge whether a large patch or a small patch is better. In fact, the displayed content regarding the similarity/dissimilarity of ROIs only shows the reason why we believe it is much easier to construct a generic large-view restoration network on fluorescent images compared to natural images. Before this content, we should first introduce our motivation for using large-view training instead of small-patch training. The perils of small-patch training lie in: (1) small-patch training loses the rich context structural information as an important clue for restoration; and (2) the structural statistics are different in the training-stage small patches and the application-stage large-view image. Now we have revised the figure as follows to better clarify our motivations:

In the newly added subgraphs Fig. S1a and b, we illustrate the perils of small-patch training through an illustrative example. Firstly, the small-patch truncates much contextual structural information, so in the training, the network only learns how to use limited structural contexts to restore the image. However, in practical applications, large images contain richer contextual information, which has not been fully utilized for small-patch networks (Fig. S1a). The phenomenon that leverage large-view context information for better restoration can even be traced back even before the era of deep-learning. For example, in the powerful BM3D denoising, which find some similar blocks, stack them, and perform collaborative filtering to remove noise/artifacts. BM3D used to be one of the most powerful denoising algorithm. When training large-view images, deep neural network can seek more complex contextual information than the simple similarity in BM3D to assist in restoration.

In addition, comparing the statistics of structural information of the entire view and small patches of the same FOV, it can be found that there are differences between them (Fig. S1b). This can also lead to performance degradation when the small-patch-trained network is directly applied to large-view images. This phenomenon was also pointed out in the literature [*Improving Image Restoration by Revisiting Global Information Aggregation. ECCV, 2022*]. However, the literature also points out that the feasibility of directly training large-view and aggregating large-view information in natural images is relatively low. Because it requires a complex network to aggregate useful large-view information from distinct patches (i.e. establishing large-view connections with complex Attention mechanism). This will bring huge computational burden. This is also why we specifically explain that fluorescence images are more amenable to direct large-view training compared with natural images (Fig. S1 c, d). This represents a key finding and innovation of this paper.




**Supplementary Fig. 1. Motivations of developing LargePNet. a.** Small-patch training forces the network to learn restoration with much truncated contextual structural information, while large-view image provides more contextual clue to benefit restoration. **b.** The inconsistency of feature statistics distribution in training small-patch and final application on large-view image would degrade the model performance. The contrast, energy, and correlation of the gray-level co-occurrence matrix (GLCM) are used to quantify the feature statistics distribution. **c.** A representative image from the DIV2K and BioSR dataset, and the histogram from four selected blocks. **d.** Statistical comparisons of intra-ROI similarity in fluorescence and natural image dataset. We divide a full-size image into several 128×128 patches without overlapping. Then we calculated the GLCM energy and average gradient metrics of each 128×128 patch that characterizes the structural features. A spot in the graph represents the standard deviation of the two metrics from a group of 128×128 patches generated from a full-size image. The results in c and d show that fluorescence image has a relatively lower intra-ROI distinction compared with natural images. Thus, leveraging large-view statistics is easier in fluorescence images relative to natural images.

Continuing the discussion above, we point out that fluorescence images have more significant similarities and connections between ROIs compared to natural images (Fig.

S1 c, d). In practice, this makes the network easier to learn how to produce a better restoration result by leveraging large-view information. The LargePNet is such an architecture that can achieve better performance by training with a large patch size while maintaining good computational efficiency.

We agree that the similarity or dissimilarity between the ROIs is not sufficient to motivate the use of large-view training instead of small-patch training. Instead this comparison is our motivation that it is easier to train large-view images in fluorescent images than in natural images (Fig. S1c, d). We are very sorry for the misunderstanding caused by the previous presentation. The loss of contextual structural information and the difference in structural statistics between small images and large images are the motivations for using large images for training. The two issues have been visualized in the newly added Fig. S1 a, b.

Moreover, we may explain that for image-restoration task, using large patches would not obfuscate some of these more granular details. Unlike target detection or semantic segmentation (where different patch-sizes may help isolate and recognize object with different sizes), the objective of fluorescence image restoration is to preserve true biological details while removing degradation (e.g., noise, blur). The noisy input and GT images already share similar structural contours, and the restoration network's learning target is to invert the degradation process (e.g., denoise, deblur) rather than modify or reshape the underlying biological structure. Thus, large-view do not "obfuscate" details; instead, they provide the necessary spatial context and regional reference to ensure that the network does not lose true details as degradation or retain degradation as false details. This can also be supported by our experiment results throughout the paper.

Thanks again for your insightful comments, which have helped us clarify the motivation for developing LargePNet in fluorescence image restoration.

Action:

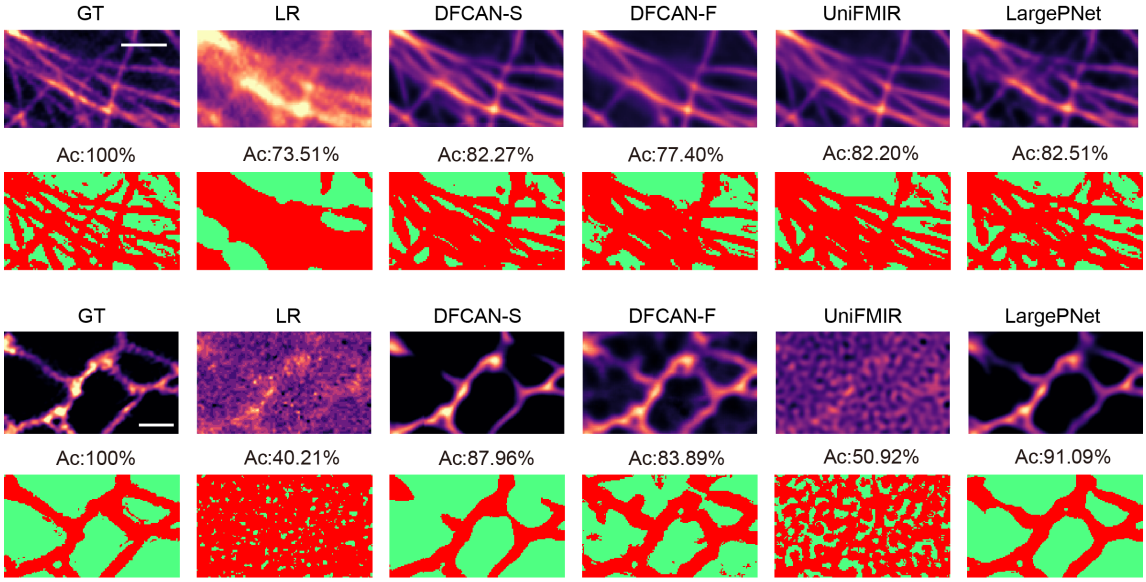
We have revised Supplementary Fig. 1 and its caption to clarify our motivation for developing LargePNet.

While measurements of PSNR, NRSRE, MS-SSIM are good, it would be good to also include a measure of downstream accuracy such as segmentation quality. Image restoration for the sake of restoration is of minimal value in terms of image analysis. Moreover, it could be good to identify regions that contain potentially significant hallucinations, if they exist. In particular, in Fig S2, there is a potential hallucinated connection in all model predictions that does not appear to exist in the GT (just west and north of the middle of the image).

Response:

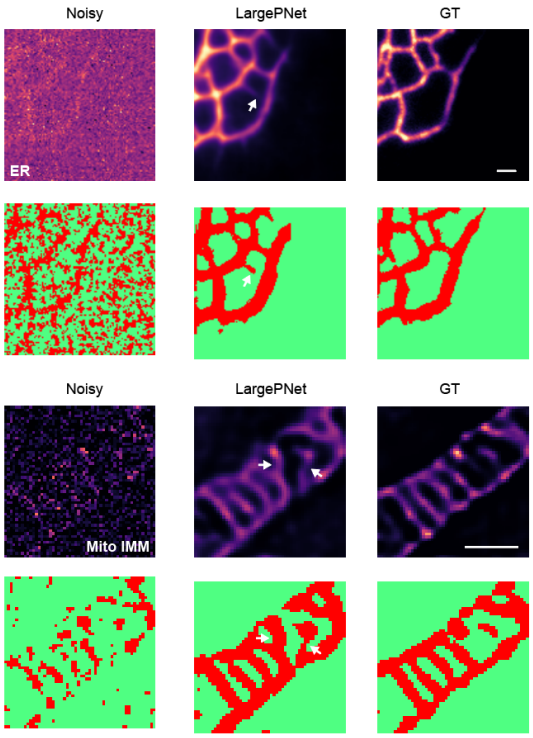
Thanks for your insightful suggestion, which helps improve the quality of our manuscript. Higher quality metrics like PSNR, NRMSE, MS-SSIM usually accompanies higher segmentation quality. We employ the commonly used TWS segmentation tool to analyze the results. We regard the GT segmentation results as a label and assess the segmentation quality of different restoration results. An example of segmentation results of the BioSR dataset is shown below (Response Fig. 2-6). It can

be found that in some very-low-intensity cases such as the ER results shown below, LargePNet shows extraordinary performance and lead to a much-gained segmentation performance. You can find more results in the revised manuscript and supplementary information.




Response Fig. 2-6. Segmentation accuracy of the single-image super-resolution results. Scale bars: 1 μm.

We greatly appreciate the reviewer's suggestion to identify the areas where the model has illusions. When the noise in the raw image is too-strong, the restoration of the model may experience hallucinations. We stated this issue in the discussion section of the paper and provided some additional examples of segmentation results including the case in Fig. S2 as follows:




Response Fig. 2-7. When the noise is too strong, there may be hallucinations in

LargePNet restoration. The white arrows denote the example hallucination region. Scale bars: 0.5 $\mu$m.

Action:

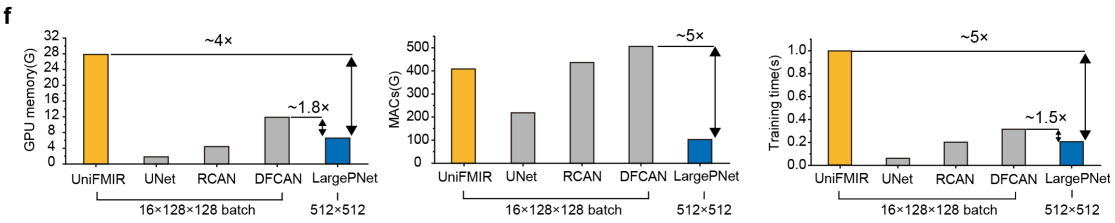
Segmentation results are added as an additional assessment of model restoration performance. The results are made as Supplementary Figs. 5, 8, 11, 14, 17 and Extended Data Fig. 5 in the revised manuscript and supplementary information.

The issue of possible hallucinations in the model when the noise is too strong has been declared in the discussion section, and Supplementary Fig. 17 is provided as examples.

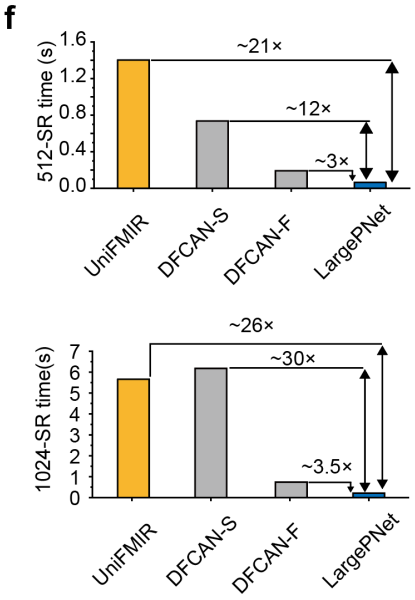
I am curious how the computational efficiency increase(s) relative to current methods such as DFCAN scales with input image size (see Figure 1f).

Response:

Thanks for your good question. We are sorry that our previous presentation of Figure 1f is not clear, which may leave misunderstanding. Figure 1f displays the training efficiency of processing an equivalent volume. For LargePNet, we display its efficiency in processing a 512×512-size image, while for other models devised for training small-patches, we display their efficiencies in processing a 16×128×128 batch. Previously, this information is only in the figure caption, which seems indirect. Now we have marked this information in Figure 1f directly to avoid possible misunderstanding. We speculate that you want to ask the how the computational efficiency of LargePNet increases relative to current models with input image size at the inference stage. This result can be found in Figure 2f (the figure is also refined to illustrate this issue more clearly).




The revised Figure 1f




The revised Figure 2f

Compared with previous small-patch CNN architecture, if it adopts the full-size inference mode, LargePNet's fold improvement in efficiency would slightly increase with increasing image sizes. If small-patch CNN adopts the small-size inference and stitching mode, LargePNet's fold improvement in efficiency would significantly increase with increasing image sizes. In contrast to Swin Transformer-derived models (UniFMIR and SwinIR), the fold improvement in efficiency of LargePNet would slightly increase as image sizes increase. Detailed analysis using the example result in Figure 2f is as follows:

- LargePNet versus DFCAN-F (full-size inference mode):

Theoretically, the computational efficiency improvement of LargePNet over DFCAN-Full would be relatively stable with input image size. This is due to that both two are CNN architecture, and the convolutional operation complexity would increase proportionally with the square of image size. In practice, memory footprint also acts as a key factor: LargePNet has a lower memory occupation than DFCAN-Full, thus imposing a smaller memory burden as input size increases, which leads to a slight rise in its fold efficiency improvement. In the provided results in Figure 2f, LargePNet achieves ~3-fold efficiency gains in 512-size (0.065 sec versus 0.193 sec) inference and ~3.5-fold in 1024-size inference (0.214 sec over 0.736 sec).

- LargePNet versus DFCAN-S (small-size inference and stitching mode)

For DFCAN, when it infers large-size image using the stitching mode, If there is no overlap between the small images of inference, then its computational complexity will also increase proportionally with the square of the image size. However, this can lead to serious edge artifacts. If the inference strategy of overlapping small graphs is adopted, its computational complexity will further increase significantly on this basis. For example, if a 50% overlap strategy is adopted to infer a 512-size image, DFCAN needs to infer 49 128-size images; To infer an image with a size of 1024, DFCAN needs to infer 225 128-size images. This has evidently exceeded the square growth. Moreover, additional cache pressure, parallelism loss, and data transmission will further reduce its

actual speed. As shown in Figure 2f, LargePNet achieves ~12-fold efficiency gain in 512-size and ~30-fold efficiency gain in 1024-size over DFCAN-stitching.

- LargePNet versus Swin Transformer derived model (UniFMIR and SwinIR)

UniFMIR is derived from Swin Transformer. When inferring full-size image, its shifted window attention mechanism ensures computational efficiency, and allows adjacent 8×8-size windows to overlap and transmit context, avoiding boundary blocks. As a result, it also follows a square growth with the input image size in the inference stage, just like a CNN adopts stitching mode but without overlap. Similarly, the memory cost also influences the practical speed. LargePNet has a lower memory occupation than UniFMIR, thus imposing a smaller memory burden as input size increases, which leads to a slight rise in its fold efficiency improvement. Moreover, the multi-head self-attention operation in UniFMIR/SwinIR causes very heavy computational burden. LargePNet has a significantly higher efficiency than UniFMIR whatever the input size. As shown in Figure 2f, LargePNet achieves ~21-fold efficiency gains in the 512-size input and 26-fold in1024-size input over UniFMIR.

The fold improvement in efficiency was discussed above. In addition, it is worth noting that as the input size increases, the total time savings brought by LargePNet will also greatly increase. This also constitutes an important consideration for practical deployment.

We appreciate you for pointing out this question, which helps improve our manuscript.

Action:

Figure 1f is marked with more information to avoid possible misunderstanding. Figure 2f is refined to illustrate this issue more clearly.

The computational efficiency of LargePNet increases relative to current models with input image size has been discussed in Supplementary Note 3 in the revised Supplementary Information.

Across the manuscript, stating whether the gains in these metrics are statistically meaningful compared to current methods would strength conclusions.

Response:

Thanks for your constructive suggestion. This issue should be clarified to strengthen the conclusions.

We have performed statistical significance t-test of all the experiments performed in this manuscript. The results show that in all experiments LargePNet shows very significant improvements ($p < 0.001$, ***) in PSNR and NRMSE metrics. The results are shown in your Figure-specific question below. Moreover, we have uploaded all source data of the PSNR, MS-SSIM and NRMSE metrics in the .xlsx format to ensure the reproducibility of the t-test results.

As our response to your Overall Comment, in this field the box chart display of quality metrics may leave a superficial impression that the improvement is insignificant. This is due to that in all fluorescence-image-restoration experiments the same FOV with multiple noise levels (for example, BioSR dataset has 12 noise levels for a single F-actin FOV) are used to train and test the model performance. While the raw noise level

is an important factor that influences the model performance. All restored quality metrics would decay as the increase of the noise level in the raw image. Therefore, the metrics box chart of different models would have a wide overlap region. However, the difference in the average value, as shown in our test results, is very significant. You can also back-check the results and explanations in our response to your Overall Comment and Response Figs. 2-2, 2-3.

We are aware that randomness in some other fields may influence the significance of the statistical comparison. While for fluorescence-image-restoration as a low-level vision task, the whisker (deviation) in the box chart is largely originates from different noise levels in the sample rather than the sample randomness. Another factor that may influence the t-test result is the sample numbers. In each fluorescence dataset typically at least 24 samples are used to test the model performance, ensuring the significance of LargePNet improvement.

We very much appreciate that the reviewer proposed this constructive suggestion. This has greatly helped strengthen our manuscript. We hope that through our statement of statistical significance in this paper, the researchers in this field can also follow and set this to avoid possible confusion on the box chart display results.

Action:

We have added statistical significance tests of all the experiments performed in the manuscript. The source data are uploaded in the .xlsx format to ensure the reproducibility of the results.

A discussion of choice of metrics and their relation to resolving biological structures would be helpful.

Response:

Thanks for your constructive suggestion, which helps to further clarify the rationality of our metric selection and its relevance to biological structure analysis. We chose the combination of pixel-precision metrics (PSNR, NRMSE) and structural similarity metric (MS-SSIM) based on the dual requirements of fluorescence image restoration for quantitative intensity fidelity and biological structure preservation. Their complementary roles effectively address the multi-dimensional evaluation needs in biological imaging scenarios.

Pixel-precision metric (PSNR and NRMSE) quantifies pixel-level grayscale fidelity that ensures signal discriminability of biological structures. Loss of the completeness of the biological structure or distortion in the recovered gray value compared to the ground-truth image will be directly reflected in PSNR and NRMSE metrics. A higher pixel-precision metric also provides a more reliable grayscale basis for subsequent structural identification such as segmentation. The pixel-precision metric shall be the premier standard for evaluating the restoration quality in scientific imaging field like fluorescence imaging that requires high-fidelity.

SSIM focuses on similarity of two images in brightness (gray value average) and contrast (gray value standard deviation) over a region instead of the pixel-precision difference. MS-SSIM is a multi-scale extension of SSIM that incorporates calculations

across multiple scales, addressing the limitations of SSIM's single-scale evaluation. However, since MS-SSIM considers the similarity of different image scales, it is insensitive to the integrity of small subtle biological structures such as mitochondrial cristae shown in Fig. 3. It mainly serves as an auxiliary evaluation standard for pixel-precision metrics to better reflect the structural distortions.

Besides the three metrics, we added the FID (Fréchet Inception Distance) metric [*GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. arXiv (2018).*] to assist the assessment of the SMLM virtual sampling task. This task differs slightly from other tasks in removing noise or blurring targets, as it involves filling in insufficiently sampled single-molecule emission points. From the gaining of the samplings in SMLM, we can more confidently assess the existence of a microtubule filament. For example, in the under-sampled SMLM image we can hardly identify the existence of a complete microtubule filament, but in a well-sampled SMLM image the contour of microtubule becomes clear. While pixel-precision or local similarity metrics like PSNR, NRMSE, and MS-SSIM can measure the differences, they fail to provide the semantic level comparison, which may also be an important aspect that reflects the restoration quality of the SMLM images. The core idea of FID is to measure the distance between the perception of the distribution of GT images and the restored images on the pre-trained Inception-v3 model. It can evaluate semantic quality information such as whether the restored microtubules are as complete as the GT SMLM images. Therefore, here we use FID value as an additional evaluation metrics for assessing the quality of the restored SMLM images.

We very much appreciate the reviewer for pointing this out. We have added relative discussion in the manuscript to clearly illustrate this issue.

Action:

We have added a discussion of the choice of metrics and their relation to resolving biological structures in the Methods section of the revised manuscript.

As a general comment, many of the arrows used to highlight regions or line profiles do not sufficiently or precisely identify the ROI in question.

Response:

Thank you for pointing out. Throughout the paper, the arrows are used to denote a structure or mark an intensity profile with two arrowheads. The latter usage may not be precise and ambiguous. Therefore, we have refined the usage of arrows throughout the paper. Now we only use an arrow to denote a structure to get attention, and use a dashed line to mark an intensity profile line.

Detailed modifications can be found in our following answers to Figure-specific questions and comments.

Action:

In the revised manuscript, we uniformly use dashed line to mark an intensity profile for more precise identification. For modification details please see our response to Figure-specific Questions and Comments below.

Figure-specific Questions and Comments:

Figure 1

The connection between the displayed ERF's in b and c is not obvious. In particular, the LargePNet ERF in b appears significantly broader than that in c. Please explain why this is the case.

Response:

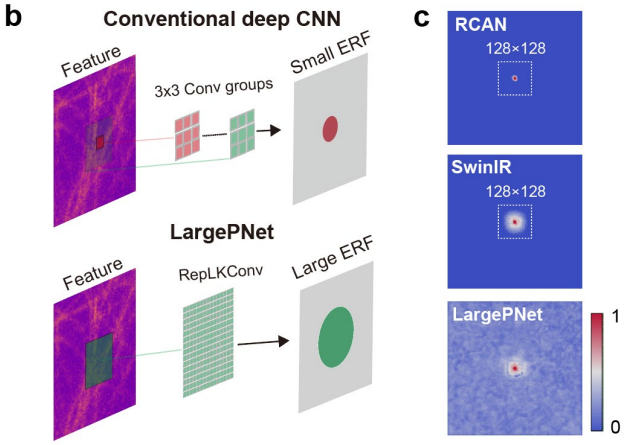
Thanks for your constructive feedback. We are sorry that our previous presentation may have caused this confusion. The ERF map in b actually displays a smaller region, and the ERF map in c displays a large region (512*512-size). Essentially, the ERF in c shall be broader than b.

Moreover, the ERF map in b is illustrative, aiming to show that the RepLKConv operation provides a much larger ERF than stacking deep 3x3 convolutions. While the ERF map in c is obtained from real data, comparing the ERF of LargePNet with advanced convolution and Transformer-based models. The reason for putting an illustrative map in b is that the analysis of the ERF map requires model training and analysis of the inference stage on real data.

We appreciate your feedback that points out this confusing issue. We have modified b to avoid this misunderstanding, making readers clearly aware of its illustrative feature.

Action:

We have modified Fig. 1b to avoid ambiguity.




The modified Fig. 1b

Figure 2

From the plots in a, the results appear significant for some structures, but maybe not others. A discussion of why this might be would be useful. In other words, what sorts of structures is LargePNet better at recognizing than others?

Response:

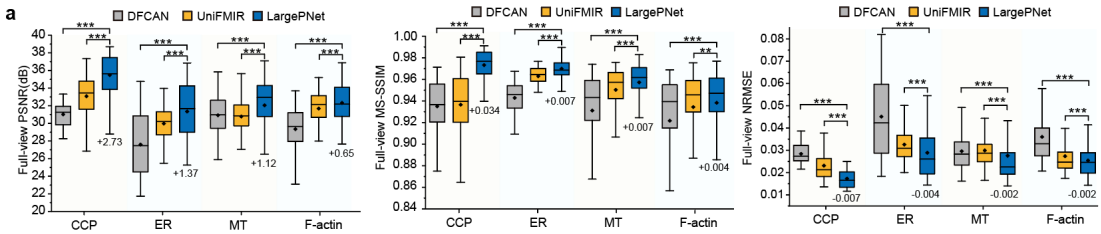
Thanks for your good suggestion. We have performed t-test to examine the significance of LargePNet improvement over previous methods. The results show that the improvement significance $p<0.001$(***) holds for all the structures. The raw data

are uploaded in the .xlsx format to ensure reproducibility.

Response Table 2-8. Performance and significance tests on the BioSR dataset results.

| Model | Metric | CCPs | ER | Microtubules | F-Actin |
|---|---|---|---|---|---|
| DFCAN | PSNR (dB) | 31.0205±1.7055 | 27.6028±4.8866 | 30.9200±3.2214 | 29.3412±3.5111 |
| | MS-SSIM | 0.9350±0.0313 | 0.9429±0.0167 | 0.9311±0.0577 | 0.9217±0.0488 |
| | NRMSE | 0.0284±0.0058 | 0.0451±0.0280 | 0.0296±0.0116 | 0.0360±0.0151 |
| UniFMIR | PSNR (dB) | 33.0813±3.6723 | 29.9711±2.9006 | 30.7877±2.7212 | 31.6725±2.3337 |
| | MS-SSIM | 0.9365±0.0403 | 0.9630±0.0118 | 0.9504±0.0290 | 0.9344±0.0313 |
| | NRMSE | 0.0231±0.0107 | 0.0326±0.0117 | 0.0299±0.0089 | 0.0273±0.0074 |
| LargePNet | PSNR (dB) | 35.8077±2.3331 | 31.3432±4.5586 | 32.0415±4.8193 | 32.3224±3.3113 |
| | MS-SSIM | 0.9708±0.0181 | 0.9699±0.0122 | 0.9574±0.0286 | 0.9383±0.0290 |
| | NRMSE | 0.0166±0.0044 | 0.0289±0.0169 | 0.0276±0.0160 | 0.0254±0.0089 |
| Significance of LargePNet improvement through t-test | | | | | |
| Model | Metric | CCPs | ER | Microtubules | F-Actin |
| Versus DFCAN | PSNR (dB) | *p*=4.35e-67(***) | *p*=1.49e-15(***) | *p*=3.02e-7(***) | *p*=1.01e-33(***) |
| | MS-SSIM | *p*=5.38e-47(***) | *p*=5.80e-20(***) | *p*=4.06e-29(***) | *p*=6.74e-15(***) |
| | NRMSE | *p*=4.1e-125(***) | *p*=3.1e-150(***) | *p*=7.46e-57(***) | *p*=2.3e-141(***) |
| Versus UniFMIR | PSNR (dB) | *p*=1.49e-26(***) | *p*=1.97e-6(***) | *p*=2.09e-8(***) | *p*=2.23e-5(***) |
| | MS-SSIM | *p*=3.56e-31(***) | *p*=1.46e-10(***) | *p*=1.54e-25(***) | *p*=0.0044(**) |
| | NRMSE | *p*=2.97e-22(***) | *p*=8.25e-27(***) | *p*=8.50e-12(***) | *p*=1.83e-11(***) |

As discussed in our response to your Overall Comment (Response Figs. 2-3, 2-4), the lack of clarity in the Box chart comparison stems from its inclusion of all raw noise levels in the results. In fact, the improvement of LargePNet is statistically significant. To address this, we also annotated the test results directly on the graph, and also marked the PSNR, MS-SSIM improvement, and NRMSE reduction of LargePNet compared to the second-best model:




The revised Fig. 2a

Regarding the structural differences, we recall the GLCM analysis result here. For the detailed calculation method, please see our response to your Overall Comment.

Recall: Response Table 2-2. Explaining structure-dependent advantage of LargePNet in the BioSR single-image-super-resolution task using the GLCM analysis method.

| LargePNet PSNR gain compared with the best small-patch model | | | | |
|---|---|---|---|---|
| Structure | CCP | ER | MT | F-actin |
| PSNR(dB) gain | **+2.73dB** | **+1.37dB** | **+1.12dB** | +0.65dB |
| GLCM Contrast statistics difference analysis | | | | |
| Large-view | 1.54-fold | 3.28-fold | 1.76-fold | 2.08-fold |
| Small-view | 1.55-fold | 3.29-fold | 2.17-fold | 2.47-fold |
| Large-small Diff | 1.0-fold | 1.0-fold | 1.2-fold | 1.2-fold |
| GLCM Energy statistics difference analysis | | | | |
| Large-view | 0.185-fold | 0.168-fold | 0.280-fold | 0.352-fold |
| Small-view | 0.065-fold | 0.135-fold | 0.189-fold | 0.319-fold |
| Large-small Diff | **2.8-fold** | 1.2-fold | **1.5-fold** | 1.1-fold |
| GLCM Correlation statistics difference analysis | | | | |
| Large-view | 0.0061 | 0.2351 | 0.0031 | 0.0334 |
| Small-view | 0.0379 | 0.2710 | 0.0363 | 0.0555 |
| Large-small Diff | **0.0318** | **0.0359** | **0.0332** | 0.0221 |

It can be seen that LargePNet has a relatively lower PSNR gain in the F-actin compared to other structures. The GLCM statistics show that there are more significant differences in statistics between the large and small views of other structures. The differences of the GLCM Correlation metric between the large-small views of CCP, ER, and MT all exceeded 0.03. The difference in Energy statistics of CCP reached 2.8-fold, and MT reached 1.5-fold. The large-small view differences in F-actin on these statistical indicators are not as significant as in other structures. Mapping these statistical indicators onto the image, F-actin is a very complex structure, and its local regions may already contain rich information about the structure and degradation process. Therefore, the structural information lost by small-patch truncation may not be as much as other structural losses. This could be the reason why LargePNet provides a lower PSNR gain in F-actin compared to other structures.

Action:

Significance tests have been performed and marked on these box charts.

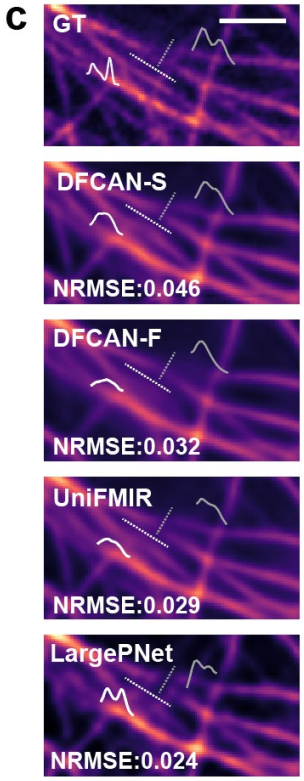
LargePNet's gain in PSNR/MS-SSIM and reduction in NRMSE value compared with the second-best model are also annotated.

What do the white arrows and curves represent in c? The curves, in particular, do not appear to correlate with structures in the image. Moreover, the LargePNet output appears to miss the fiber between the arrows from ~10 o'clock to 5, while DFCAN-Stitching and UniFMIR do not to the same degree, at least qualitatively.

Response:

Thanks for your insightful question. The arrows in c aim to mark a line between the two arrowheads, and the curve is the intensity profile along the line between the two arrowheads. It may be more appropriate to be modified to a dashed line to indicate the intensity profile to be investigated.

The fiber between the arrows from ~10 o'clock to 5 actually exists in LargePNet recovery results. We have shown the intensity profile below (gray-colored). The reason that the fiber may be dimmer in LargePNet is that it resolves the two adjacent fibers. While DFCAN and UniFMIR did not clearly separate the two fibers at this position and superimposed their intensity, making it seem bright. We also calculate and mark the NRMSE value of this magnified region, to quantitatively compare the quality of different results. We are grateful that the reviewer points out, which helps us refine the manuscript.




The revised Figure 2c

Action:

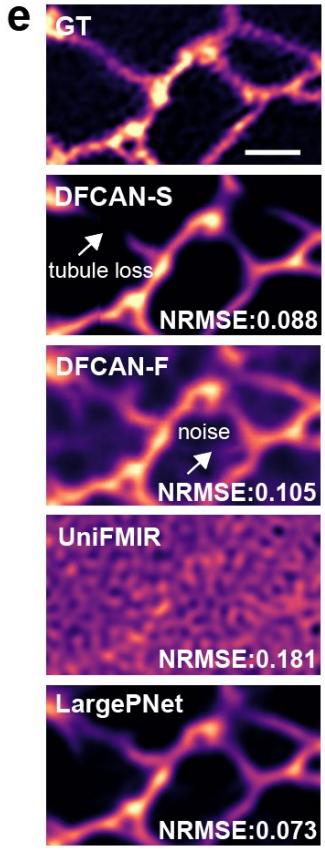
We refine Fig. 2c. The arrows have been modified to dashed lines, to indicate the sampling line of the intensity profile more clearly. An additional gray dashed line and intensity profile is added to mark the fiber from ~10 o'clock to 5. NMRSE value is

marked in the graph to quantitively compare the quality of different results.

In e, the array pointing to “signal loss” in DFCAN-Stitch could just as well be used in the LargePNet prediction. A quantitative assessment demonstrating that the prediction in LargePNet is improved relative to DFCAN-Stitch would better support this assertion.

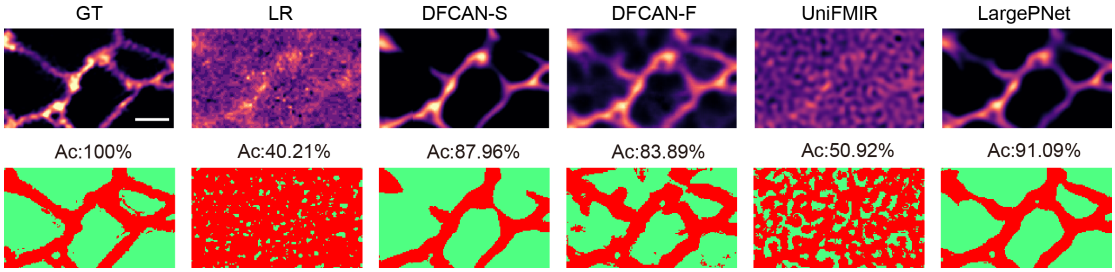
Response:

Thanks for your insightful suggestion. The arrow aims to denote that DFCAN-stich loses this ER tubule structure. The word “signal loss” shall be modified to “tubule loss” for a more straightforward description. We have calculated the NRMSE value as a quantitative assessment demonstrating that the prediction in LargePNet is improved relative to DFCAN-Stitch, as shown below:




The revised Figure 2e

Moreover, we have seriously taken your advice in the General Comments, to add a segmentation accuracy comparison to strengthen the conclusion:




Recall: Response Fig. 2-6. Segmentation accuracy of the single-image super-resolution results. Scale bars: 1 μm.

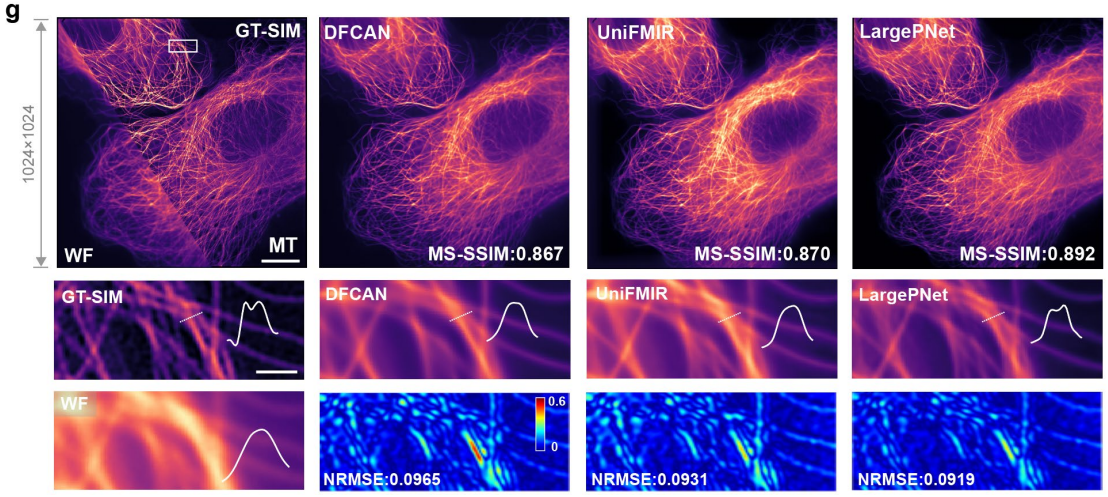
Action:

We refine Fig. 2e. The "signal loss" wording has been revised to "tubule loss" for a more precise description. The NRMSE value has been calculated and marked. Segmentation results have been added in Supplementary Fig. 5 to strengthen the conclusion.

In g, while LargePNet appears to perform qualitatively better than DFCAN and UniFMIR, its prediction also appears far off from the GT SIM data. A comparison between the GT and the prediction would strengthen this figure.

Response:

Thanks for your insightful suggestion. LargePNet precisely predicts the structural information (the microtubule filaments) and is consistent with GT SIM data. The key difference lies in the retention of background content, which is more pronounced in our result compared to the cleaner GT SIM data. We have added an error map in this figure and marked the NRMSE value to show the comparison between GT more clearly.




The revised Fig. 2g

Action:

We refine Fig. 2g. The error map between each prediction and the GT is added. The NRMSE value is also calculated and marked.

Figure 3

For these predictions, were 512x512 blocks also used as in the other examples?

Response:

Thanks for your insightful question. The answer is yes for all convolutional models. For LargePNet, since it is directly trained with a large patch, we directly predict the full-view image. For UNetRCAN, despite being trained with small patches, full-view prediction is also adopted, since we found that it produces higher quality metric values (PSNR, MS-SSIM, and NRMSE) than the stitching method in this task. This is the same with DFCAN's behavior in the BioSR task, as shown in Fig. 2 and Supplementary Fig. 4d. Since objective quality metrics should be a more important factor in restoration

tasks, we display UNetRCAN performance at its best state. In addition, as we have already shown the comparison of the advantages and disadvantages of traditional CNN direct full-view inference and small-patch-stitching in Fig. 2, in order to avoid excessive redundancy in the article, we present the results of direct full-view inference in all the following contents.

For SwinIR, since it is a Transformer-based architecture, it can only take the same input-size (128-size) information as in the training stage. When inferring large-view image, it gradually shifts its attention window to fulfill the inference of the whole image.

In a, it is not obvious that LargePNet performs much better than UNetRCAN across the board. An argument could be made for LargePNet vs SwinIR, but not vs UNetRCAN, at least by eye.

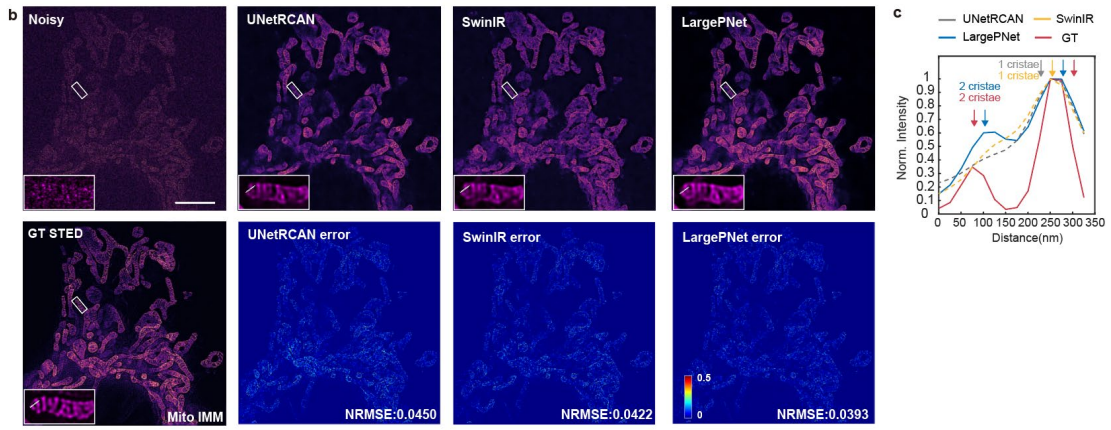
Response:

Thanks for your insightful question. We have adopted significance analysis. The results are shown below. The source data are also uploaded in the xlsx format to ensure the reproducibility of the results.

Response Table 2-9. Performance and significance tests on the STED denoising dataset results.

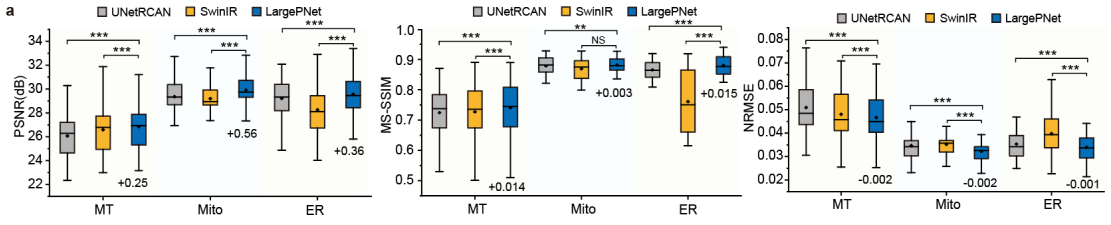
| Model | Metric | Microtubules | Mito | ER |
|---|---|---|---|---|
| UNetRCAN | PSNR (dB) | 26.0681±2.7485 | 29.3609±2.3268 | 29.2095±2.3496 |
| | MS-SSIM | 0.7241±0.1283 | 0.8785±0.0517 | 0.8655±0.0443 |
| | NRMSE | 0.0509±0.0173 | 0.0346±0.0099 | 0.0353±0.0100 |
| SwinIR | PSNR (dB) | 26.5923±2.9001 | 29.1935±2.0527 | 28.2601±3.4403 |
| | MS-SSIM | 0.7269±0.1412 | 0.8687±0.0511 | 0.7609±0.1490 |
| | NRMSE | 0.0481±0.0173 | 0.0352±0.0078 | 0.0398±0.0138 |
| LargePNet | PSNR (dB) | 26.8437±2.7108 | 29.9172±1.6248 | 29.5671±2.4895 |
| | MS-SSIM | 0.7408±0.1307 | 0.8815±0.0366 | 0.8800±0.0570 |
| | NRMSE | 0.0466±0.0155 | 0.0322±0.0063 | 0.0339±0.0090 |
| Significance of LargePNet improvement through t-test | | | | |
| Model | Metric | Microtubules | Mito | ER |
| Versus UNetRCAN | PSNR (dB) | $p$=1.09e-17(***) | $p$=7.50e-6(***) | $p$=1.56e-4(***) |
| | MS-SSIM | $p$=2.03e-18(***) | $p$=0.1754(NS) | $p$=6.19e-8(***) |
| | NRMSE | $p$=5.21e-17(***) | $p$=1.98e-4(***) | $p$=1.53e-4(***) |
| Versus SwinIR | PSNR (dB) | $p$=1.92e-13(***) | $p$=9.46e-8(***) | $p$=4.78e-13(***) |
| | MS-SSIM | $p$=2.60e-6(***) | $p$=0.0013(**) | $p$=3.77e-5(***) |
| | NRMSE | $p$=1.09e-12(***) | $p$=6.39e-10(***) | $p$=1.63e-6(***) |

From the test results, it can be seen that LargePNet outperforms UNetRCAN in all the quality metrics (PSNR, MS-SSIM and NRMSE) with very high significance (***) for MT and ER structures. For the Mito structure, the improvement of LargePNet over UNetRCAN on the PSNR and NRMSE metrics is in very high significance (***). While for the Mito's MS-SSIM metric, LargePNet (0.8815) exceeds UNetRCAN (0.8785) in the average value, but the improvement is not significant. This is due to that the Mito inner membranes are very subtle, tiny structures. While MS-SSIM calculates the SSIM of the predicted image with ground-truth at full-resolution and several down-sampling resolutions. In these down-sampled images the difference of Mito inner membrane structures may not be evident since it is averaged in local pixels. Therefore, in the MS-SSIM metric of this Mito IMM structure LargePNet does not provide significant improvement over UNetRCAN. This also highlights that multiple metrics should all be considered for quality evaluation. For this denoising task, the pixel-level precision calculation (PSNR, NRMSE) shall more precisely reflect the quality. In the Mito inner membrane results shown in Fig. 3b, c, it can be observed that LargePNet better retain the subtle mitochondrial cristae structure, demonstrating improvement in visual quality.

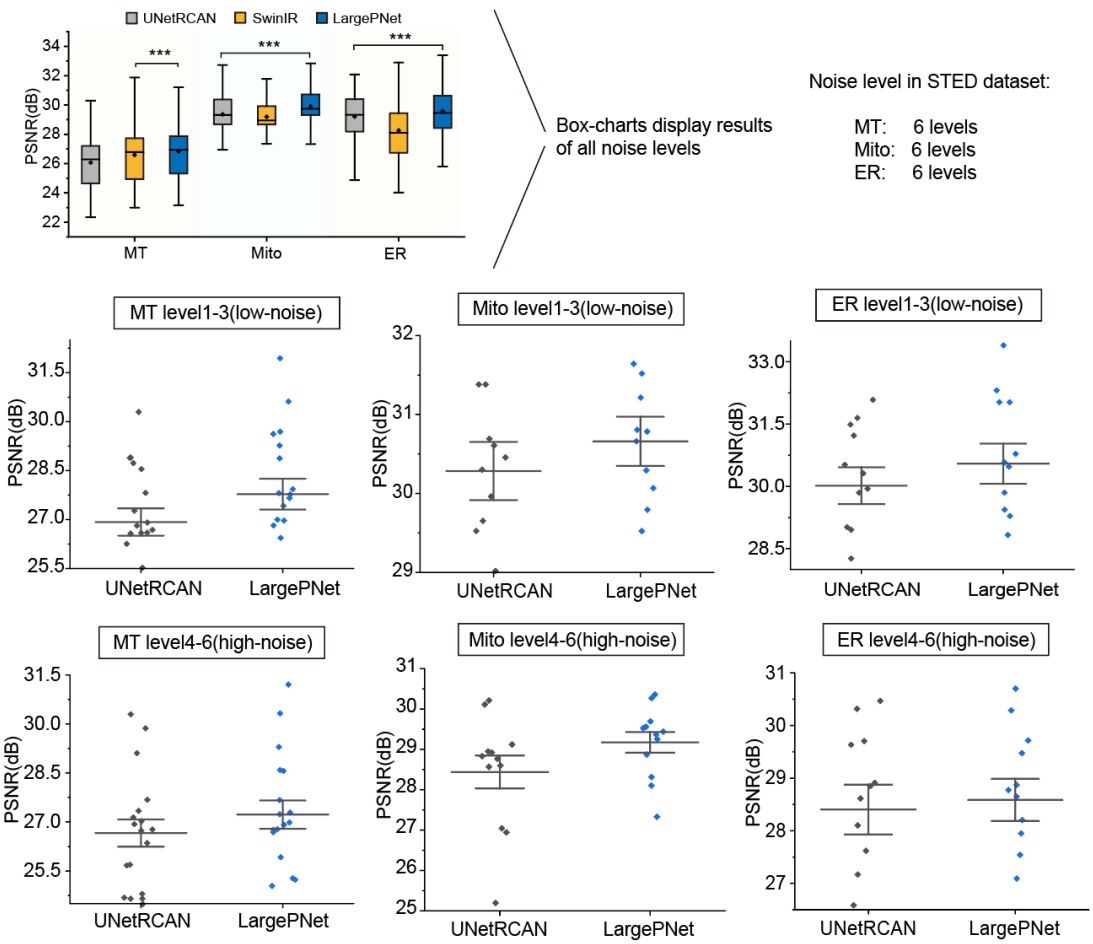



The revised Fig. 3b, c

We also mark the testing results in the main figure to directly show the statistical significance. LargePNet's gain in PSNR/MS-SSIM and reduction in NRMSE value compared with the second-best model are also annotated for clearer presentation. As that done in the BioSR dataset, we also display the data distribution at relatively separated noise levels to compensate for the drawback box chart




The modified Fig. 3a




Response Fig. 2-8. Statistical comparisons in the STED denoising dataset with separated noise levels. Each point displays the metric of an image. The error bar displays the average value and standard error (SE).

Action:

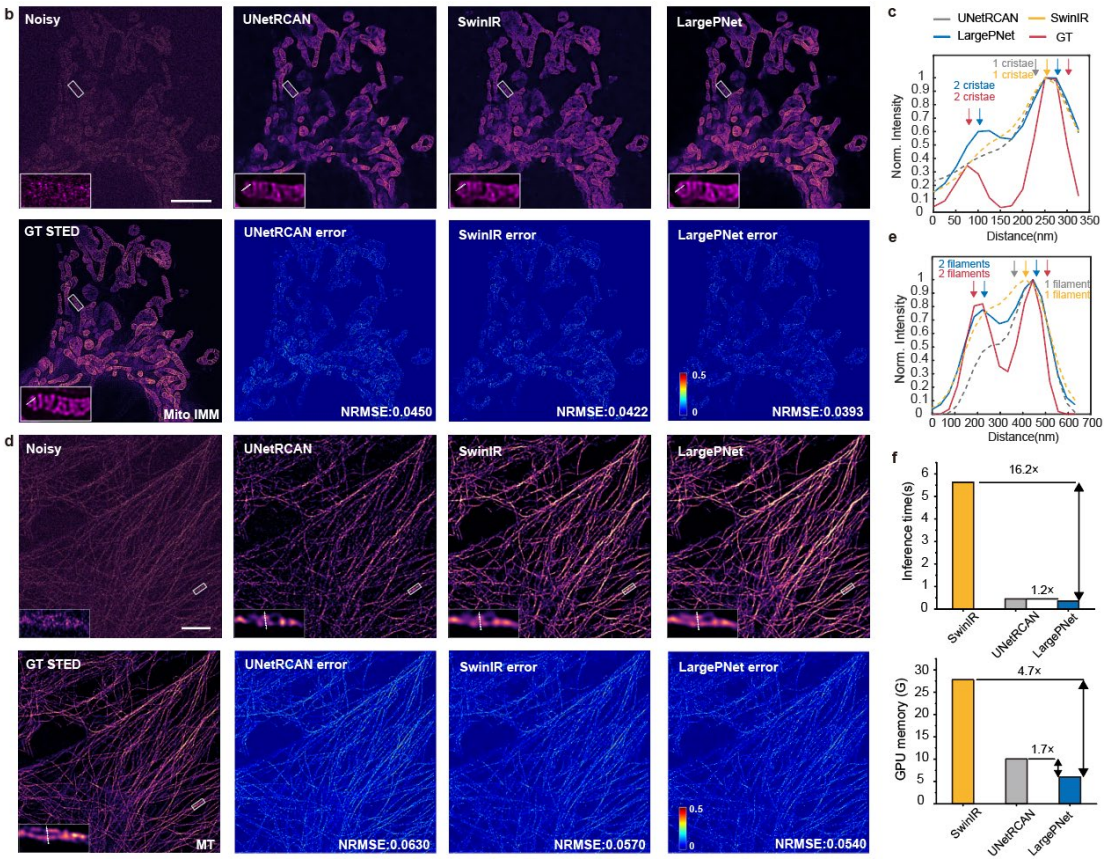
Significance tests have been performed and marked on these box charts. LargePNet's gain in PSNR/MS-SSIM and reduction in NRMSE value compared with the second-best model are also annotated for clearer presentation.

The content in Response Fig. 2-8 is made as Supplementary Fig. 9 in the revised Supplementary Information to assist understanding.

For b-e, adding zoomed-in ROIS corresponding to the regions from which these curves are taken would make it easier to correlate the intensity profiles with the biological structure. The arrows seem to refer to such a small ROI that it is difficult to see the structures being highlighted.

Response:

Thanks for your kind suggestion, which helps improve the quality of our manuscript. We have modified this figure to make it easier to correlate the intensity profiles with the biological structure.




The revised Fig. 3b-e

Action:

Magnified regions have been shown to locate the ROIs and intensity profiles more precisely.

In g, it is once again difficult to assess whether LargeP-GAN is actually performing significantly better than UNet-GAN. Other than FID, there is not an appreciable difference in many/all of the metrics. Commenting on why FID is used and what this signifies would also strengthen this point.

Panel h is difficult to interpret. Perhaps zoomed-in ROIs would highlight differences between the methods. Otherwise, as large images, it is very hard to see observe much difference.

Response:

Thanks for your constructive suggestion.

- Significance test

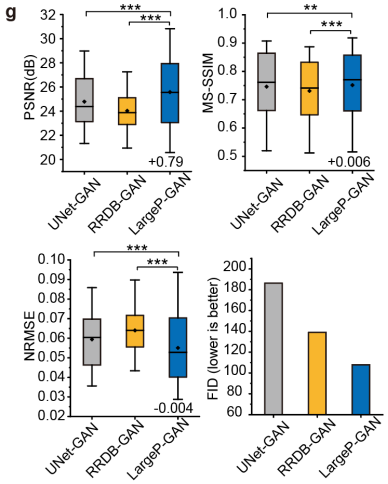
We perform significance test of this data (Response Table 2-10). It can be seen that the improvement of LargeP-GAN over the other two structures are very significant. The source data are also uploaded in the .xlsx format to ensure the reproducibility of the results.

Response Table 2-10. Significance tests on the SMLM restoration dataset results.

| Model | Metric | Microtubules |
|---|---|---|
| UNet-GAN | PSNR (dB) | 24.7897±3.2976 |

| | | |
|---|---|---|
| | MS-SSIM | 0.7461±0.1587 |
| | NRMSE | 0.0594±0.0200 |
| | FID | 186.3 |
| RRDB-GAN | PSNR (dB) | 24.0293±2.4831 |
| | MS-SSIM | 0.7310±0.1490 |
| | NRMSE | 0.0640±0.0167 |
| | FID | 139.0 |
| LargeP-GAN | PSNR (dB) | 25.5779±3.6021 |
| | MS-SSIM | 0.7517±0.1618 |
| | NRMSE | 0.0551±0.0246 |
| | FID | 107.8 |
| Significance of LargePNet improvement through t-test | | |
| Model | Metric | Microtubules |
| Versus UNet-GAN | PSNR (dB) | *p*=2.38e-13(***) |
| | MS-SSIM | *p*=0.0021(**) |
| | NRMSE | *p*=2.68e-12(***) |
| Versus RRDB-GAN | PSNR (dB) | *p*=1.02e-19(***) |
| | MS-SSIM | *p*=9.13e-23(***) |
| | NRMSE | *p*=2.41e-22(***) |

The overlap of the box chart mainly originates from the fact that the same FOV with multiple-samplings are considered. All model performance will decay with decreasing samplings in the raw SMLM images. We also marked the significance test result in the box charts to strengthen the comparisons. LargePNet's gain in PSNR/MS-SSIM and reduction in NRMSE value compared with the second-best model are also annotated for clearer presentation.



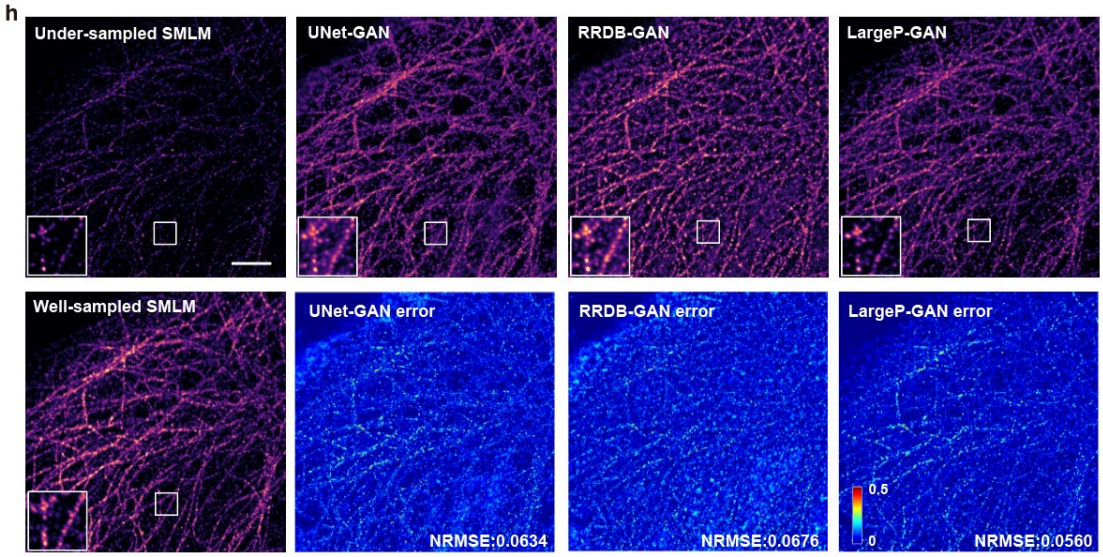

The revised Fig. 3g

- Explanation of FID

FID (Fréchet Inception Distance) is a semantic quality metric [*GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. arXiv (2018).*]. Here it is used as an additional assessment of the visual quality of GAN-based SMLM restoration task. This task differs slightly from other tasks in removing noise or blurring targets, as it involves filling in insufficiently sampled single-molecule emission points. From the gaining of the samplings in SMLM, we can more confidently assess the existence of a microtubule filament. For example, it the under-sampled SMLM image we can hardly assess the existence of a complete microtubule filament, but in a well-sampled SMLM image the counter of microtubule becomes clear. While pixel-precision or local similarity metrics like PSNR, NRMSE, and MS-SSIM can measure the differences, they fail to provide the semantic level comparison, which may also be an important aspect that reflects the restoration quality of the SMLM images. The core idea of FID is to measure the distance between the perception of the distribution of GT images and the restored images on the pre-trained Inception-v3 model. It can evaluate semantic quality information such as whether the restored microtubules are as complete as the GT SMLM images. Therefore, here we use FID value as an additional evaluation metrics for assessing the quality of the restored SMLM images. We very much appreciate the reviewer for pointing this out. We have added relative discussion in the manuscript to clearly illustrate this issue.

- Zoomed-in ROIs

Thanks for pointing this out. We have added zoomed in ROI in this figure to assist the assessment of the image quality. From the magnified region it can be seen that LargeP-GAN provides more precise virtual sampling of the under-sampled SMLM, while the UNet-GAN and RRDB-GAN would amplify some noise signals in the background. This is also consistent with the quantitative metric comparisons.




The revised Fig. 3h

Action:

Significance tests have been performed and marked on these box charts. LargePNet's gain in PSNR/MS-SSIM and reduction in NRMSE value compared with

the second-best model are also annotated for clearer presentation.

In the Methods section, we detailed described the rationale of choosing FID metric as an assistant metric for assessing the SMLM image quality. The literature [*GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. arXiv, 2018*] that proposes this metric is also cited.

We added a zoomed-in ROI in Fig. 3h to facilitate comparison of different methods.

Figure 4

For training the timeseries model, were depth-wise convolutions used? In other words, clarifying whether the temporal information is incorporated across timepoints would be helpful.

Response:

Thanks for your insightful suggestion. In the developed LargeP-TISR architecture, the temporal information is incorporated across time points. It is achieved through the usage of deep-wise convolution in an optical flow reconstruction network (OFRNet), which is proposed in a pioneering study [*Deep Video Super-Resolution using HR Optical Flow Estimation. IEEE Transactions on Image Processing PP, 1-1 (2020).*]. It leverages deep-wise convolution and channel shuffle operation to extract temporal information. The output of OFRNet will expand the channels of the input frame (for example, when the input video frame number is 7, the OFRNet will output feature maps with 25 channles). The output of OFRNet is subsequently input to our LargePNet for super-resolution, and the output channel number of LargePNet is the same with the raw input video frame number. While besides the boundary frames, LargeP-TISR would only preserve the result of the central frame, which leverages adjacent temporal information at best. Then it gradually shifts frame by frame to accomplish the super-resolution of the full video. This processing protocol is the same with the pioneer works [*Qiao, C. et al. A neural network for long-term super-resolution imaging of live cells with reliable confidence quantification. Nature Biotechnology (2025).*]. The major innovation in LargeP-TISR is that it adopts the LargePNet for super-resolution with a large-view training paradigm. We appreciate the reviewer for pointing this out, which makes the manuscript clearer.

Action:

In the main text, the way that LargeP-TISR leverages temporal information is described clearly:

"*We devise LargeP-TISR that can leverage temporal information for video super-resolution (Supplementary Note 4). It incorporates an optical flow reconstruction network[49] (OFRNet), which leverages deep-wise convolution and channel shuffle to analyze temporal information. The output from OFRNet is consecutively sent to LargePNet for super-resolution.*"

In Supplementary Note 4, the processing procedure (as described in our response) of LargeP-TISR is clearly illustrated.

Carrying over the white ROI box in all panels of a-b would better communicate the selection in c-d.

Response:

Thanks for your good suggestion, which helps improve our manuscript. We have refined this figure as suggested.

Action:

We have carried over the white ROI box in all panels of Fig. 4a-b.

This figure—outside the MT case in e—does appear to show that LargePNet outperforms other current methods. However, see the comment about quantifying the significance of this performance increase. Moreover, considering that this example appears to demonstrate the most significant increase in performance for LargePNet compared to current methods across the manuscript, a discussion of why this network seems to perform comparatively better in higher dimensions would be interesting.

Response:

Thanks for your insightful suggestion, which helps improve our manuscript.

We perform significance test, and the results are shown in Response Tables 2-11 and 2-12 below. It shows that the improvement of LargePNet is significant in all cases. The source data are also uploaded in the xlsx format to ensure the reproducibility of the results.

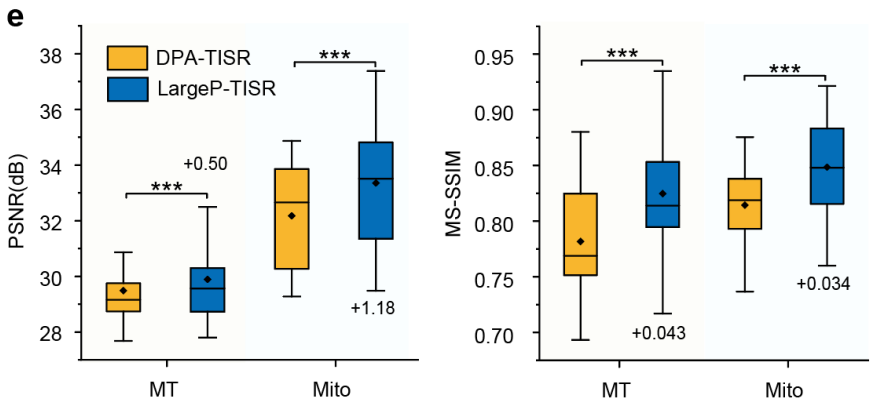
Response Table 2-11. Significance test results of the BioTISR video super-resolution task

| Significance of LargeP-TISR improvement through t-test | | | |
|---|---|---|---|
| Model | Metric | Microtubule | Mitochondrion |
| Versus DPA-TISR | PSNR (dB) | $p$=5.91e-72(***) | $p$=5.18e-51(***) |
| | MS-SSIM | $p$=1.25e-73(***) | $p$=4.30e-58(***) |

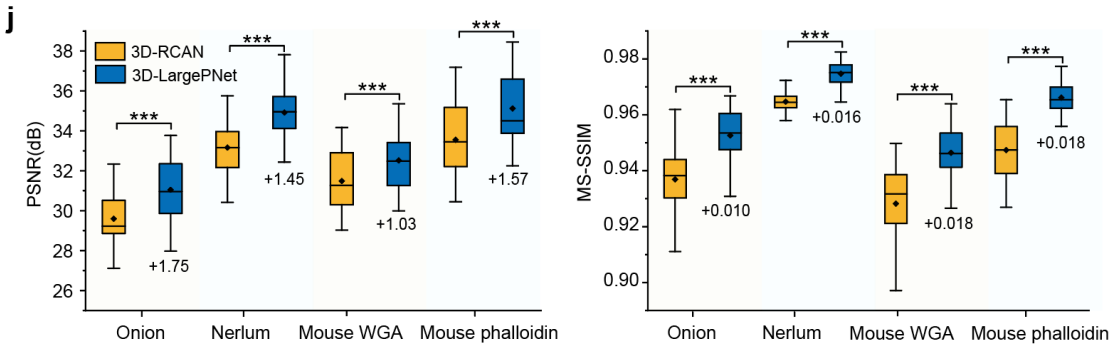
Response Table 2-12. Significance test results of the volumetric background removal task

| Significance of 3D-LargePNet improvement through t-test | | | | | |
|---|---|---|---|---|---|
| Model | Metric | Nerlum indicum stem | Onion epidermal cell | Mouse kidney WGA | Mouse kidney phalloidin |
| Versus 3D-RCAN | PSNR (dB) | $p$=8.94e-19(***) | $p$=1.11e-14(***) | $p$=1.35e-8(***) | $p$=5.87e-9(***) |
| | MS-SSIM | $p$=3.21e-20(***) | $p$=3.77e-18(***) | $p$=8.36e-8(***) | $p$=1.64e-9(***) |

We have also marked the significance level directly in the box charts. LargePNet's gain in PSNR/MS-SSIM value are also annotated for clearer presentation.




The revised Fig. 4e




The revised Fig. 4j

Moreover, it should be noted that the timeseries (the open-source BioTISR dataset) and volumes data (self-collected SD-confocal dataset) were not collected with multiple noise levels as that in single-image conditions. In time series acquisition, the samples are constantly moving, it is difficult to capture images with multiple levels of signal-to-noise ratio matching in strictly the same FOV. For Fig. 4g-k, we expect to examine the network's ability to remove background noise present in 3D fluorescence imaging instead of removing random noise under low light or short exposure conditions. Therefore, volume data were not collected with multiple signal-to-noise ratios.

As a result, the box charts displaying the timeseries and volumes data seem more compact than the single-image data. This may make it superficially clearer that LargePNet outperforms the current methods when restoring higher-dimensional data. Please see the statistical presentation of the data with separated noise levels in the response to the above comment and Response Fig. 2-3.

Action

Significance tests have been performed and marked on these box charts. LargePNet's gain in PSNR/MS-SSIM value is also annotated for clearer presentation.

For the line profiles in i, the highlighted feature (the small bump before the large peak) does not seem significant enough to highlight. In the microscopy images, it does not appear to correspond to any meaningful structure and could potentially be noise. I would suggest removing these profiles and only display the ROIs.

Response:

Thanks for your insightful suggestion. We have removed these components in this figure to avoid such a misunderstanding.

Action:

Then intensity profiles have been removed to avoid possible confusion.

Figure 5

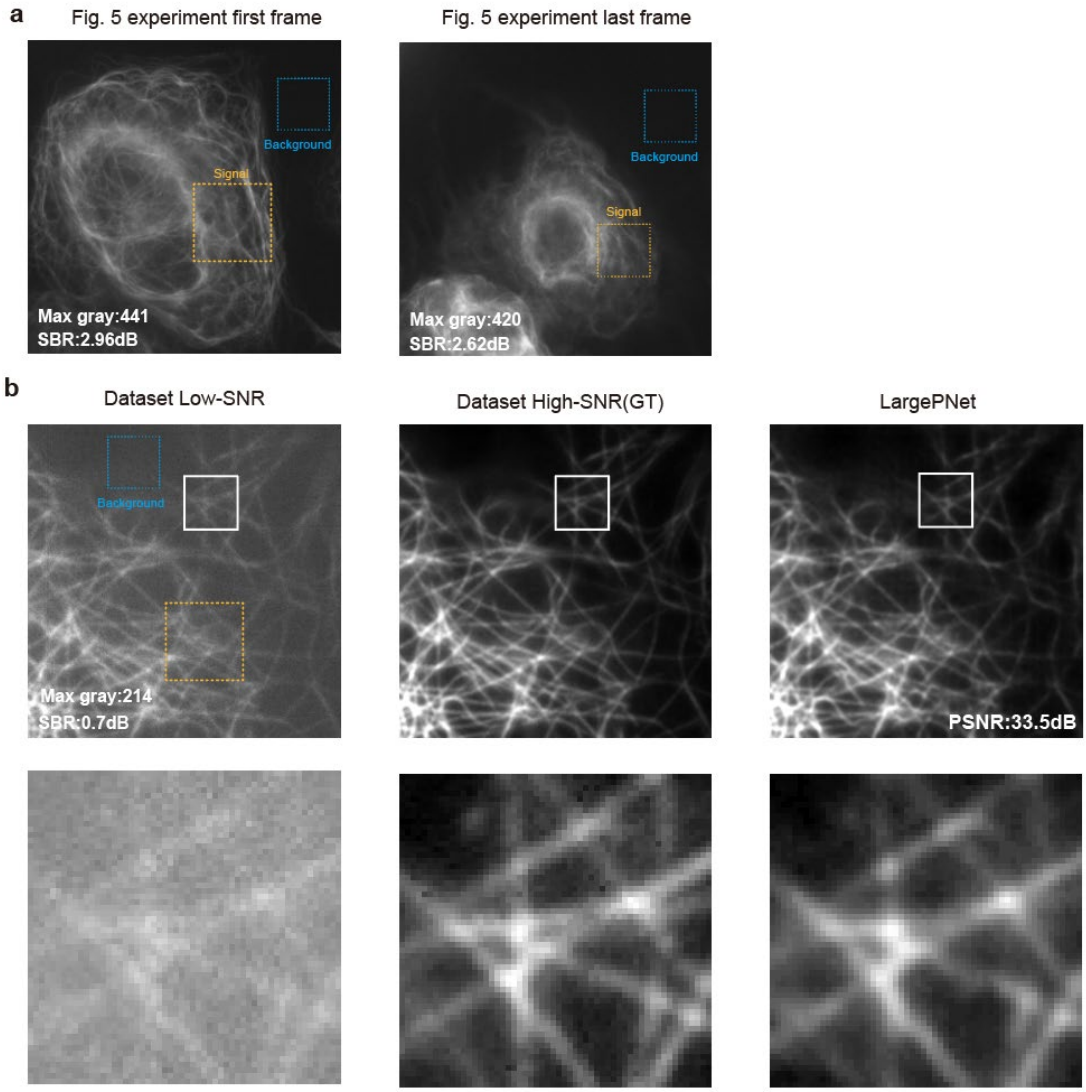
Without GT as a comparison, it is very difficult to judge whether the recovered structures exist in the raw data. As a suggestions, the same experiment could be performed of imaging at low SNR every 30 seconds, but interleave a high SNR image every couple of hours. This could then showcase the same point being made currently while also allowing for a more direct ground truth comparison in the same data set.

Response:

Thanks for your constructive suggestion.

In this field, researchers typically demonstrate the application of their models under low light (low toxicity) on live cells after validating their model performance. As shown earlier, restoration datasets typically collect raw images with multiple noise levels for restoration. Usually, as long as the model has good recovery ability at a certain SNR raw image in the dataset, it can be considered applicable to live cells with the equivalent SNR. This can be seen in some representative previous works [*Qiao, C. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. Nature Methods* ***18****, 194-202 (2021)*; *Qu, L. et al. Self-inspired learning for denoising live-cell super-resolution microscopy. Nature Methods 21, 1895-1908 (2024).*].

We agree that we may need to add some validation experiments to illustrate this point. Response Fig. 2-9 below shows an example from our test dataset. This image has a higher noise level compared with the experiment data in Fig. 5. It can be seen that our model has a very good restoration effect (PSNR>30dB) at an even higher noise levels than Fig. 5. According to the convention of research papers in this field, it can be considered that this also ensures the robustness of the model performance in Fig. 5. We have already employed a conservative condition in Fig. 5 compared to LargePNet's capability on the test dataset to ensure the reliability of the recovery results.

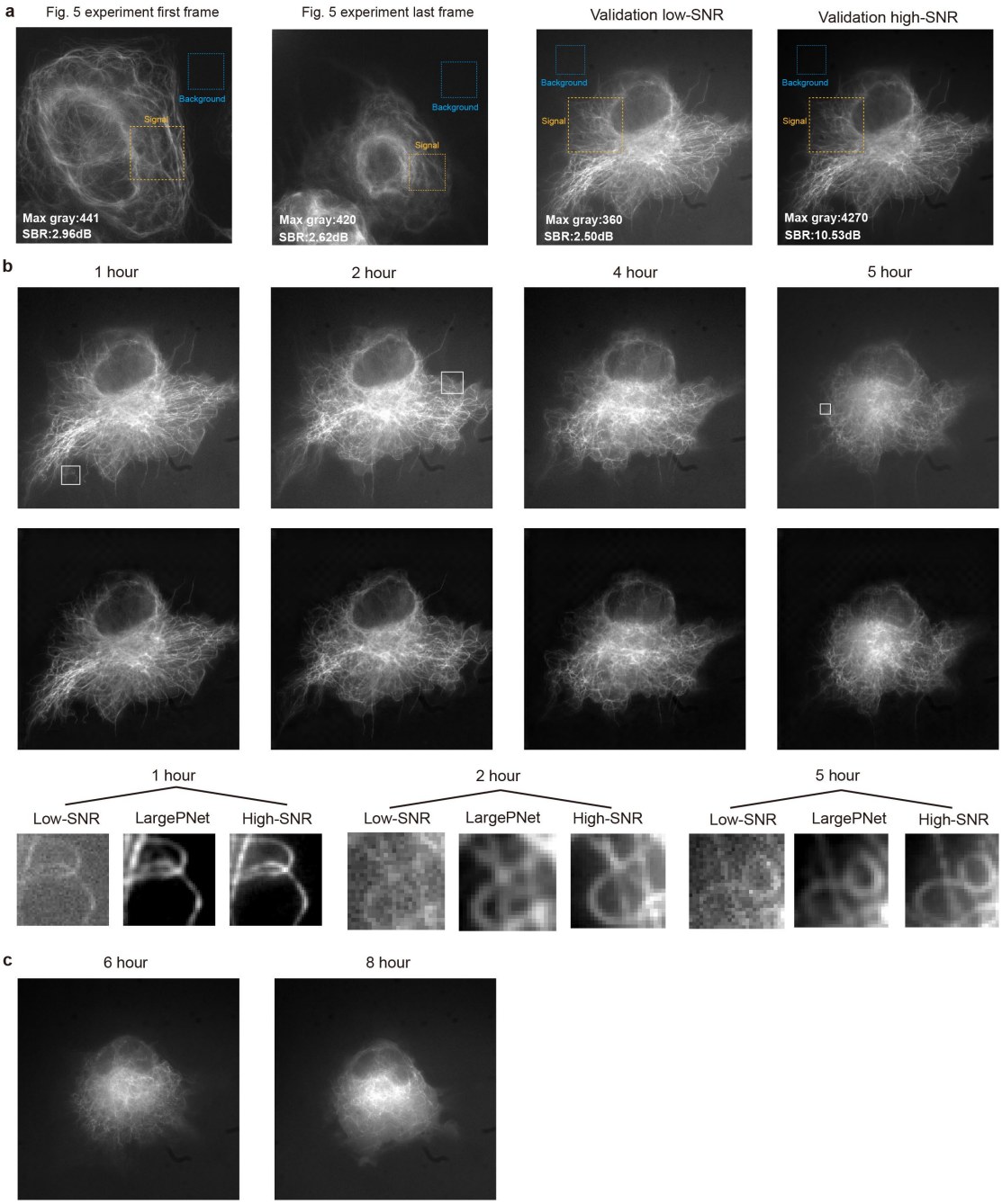



Response Fig. 2-9. Validation of the model recovery capability in Fig. 5 by inspecting the model performance on the test dataset with a lower signal intensity. a. The image in the experiment in Fig. 5. The orange box is used as signal sampling region, the blue box is used as the background sampling region. b. A test image in the BioSR dataset. The magnified boxed regions are shown below.

We also tried your suggestion to capture low light images (0.2% laser intensity) every 30 seconds and high SNR images (20% laser intensity) every hour. It can be seen that the SBR of the raw image in this experiment is comparable or lower than that of Fig. 5.

We conducted a validation of the restoration effect of LargePNet with high SNR control. It can be seen that LargePNet's noise reduction in the enlarged complex areas are satisfactory. Additionally, it should be noted that due to time delay, microtubules in high-SNR images may have shifted relative to the low-SNR and LargePNet recovered ones. Microtubules are light sensitive structures. Due to the use of a 20 times higher laser power to capture high-SNR images, a significant negative impact on the activity of microtubules is observed. The similar issue of phototoxicity was also mentioned in

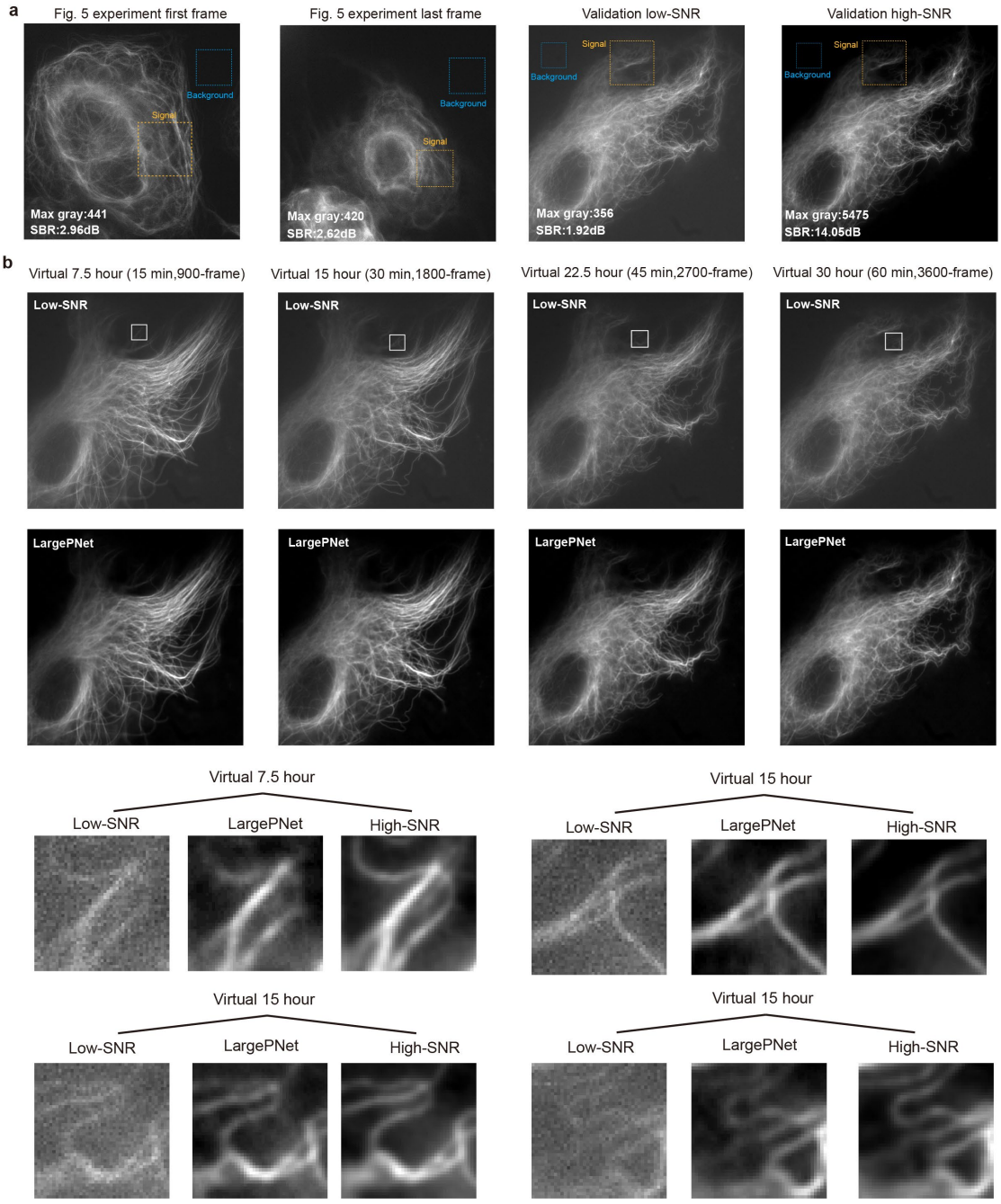
[*Editorial. Artifacts of light. Nat. Methods 10, 1135 (2013).*]. The result was that after 6 hours, although there was no obvious photobleaching, the microtubules underwent significant wrinkling and were no longer suitable for further observation. We performed the same experiment at least 3 times and observed similar results. This also emphasizes the important role of using deep learning assisted continuous low light observation to reduce phototoxicity.




Response Fig. 2-10. Validation of LargePNet restored images with high-SNR reference. a. The raw image quality of the experiment in Fig. 5 and the raw image quality of the validation experiment. b. Validation results within 1-5 hours. c. Cell states at 6 and 8 hours.

We speculate that this is also the reason why previous studies also did not use this verification method, because introducing strong light does significantly affect the morphology and activity of organelles, making it difficult to observe their changes in their natural state.

In addition, we also tried another verification method. We have reduced the time period from 30 hours to 1 hour but maintained the same light dosage. We take one low-SNR image every 1 second and one high-SNR GT image every 3 minutes. This validation experiment ensures the same total amount of light dose as the 30-hour experiment in Fig. 5. Due to the relatively short recording time (1 hour), the morphology of the cell may not change quickly after being stimulated by high-power laser to capture high-SNR image, and it may still maintain their original shape for us to check the model's noise reduction ability. In the validation image, we selected an area with a lower signal-to-noise ratio than the Fig. 5 experiment (Response Fig. 2-11a). If LargePNet can effectively extract signals at this noise level, it can also be considered that the noise reduction ability of LargePNet in Fig. 5 is sufficient. As shown in the results, LargePNet effectively extracts signals from noise and can be validated in high-SNR images (Response Fig. 2-11b).




Response Fig. 2-11. Validation of LargePNet restored images with high-SNR reference. The validation experiment is performed in an equivalent photon dosage method. (low-SNR image is captured every 1 second within an hour, maintaining the same frame number and light dosage as the 30-hour experiment in Fig. 5). a. The raw image quality of the experiment in Fig. 5 and the raw image quality of the validation experiment. b. Validation results within the virtual 30 hours.

From all the results above, it can be seen that we have already adopted a conservative lighting strategy in Fig. 5. The denoising capability of LargePNet can meet an even higher noise level. We hope that the above experiment is sufficient for you to verify the applicability of the model under Fig. 5 conditions. If you have any other

questions, please ask us. We appreciate this insightful question, which helps improve our manuscript.

Action:

The validation experiment under equivalent raw image SNR is added as Supplementary Figure 15 in the revised manuscript.

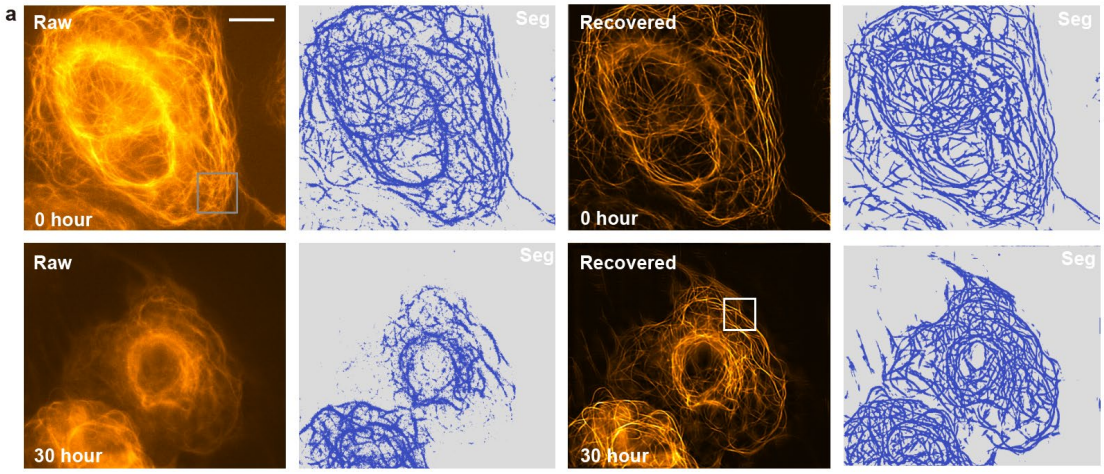
The sentence "LargePNet has effectively removed the massive noise in the raw recordings, enabling fine segmentation of the microtubules structures." could be toned down. Moreover, unless the segmentations are provided, it is not necessarily fair to claim that the segmentation is enabled.

Response:

Thanks for your informative feedback. We are sorry that the previous presentation of Figure 5 may not be that straightforward. We actually provided the segmentation results on the right side of each image. But we did not label this as the segmentation results, which may confuse readers. Now, in the revised version, we have added the word mark for clearer presentation. Moreover, we agree with you that this claim can be toned down.

Action:

The word "Seg" has been added to the segmentation results of each graph.




The revised Fig. 5a, where the segmentation results are marked with the word "Seg"

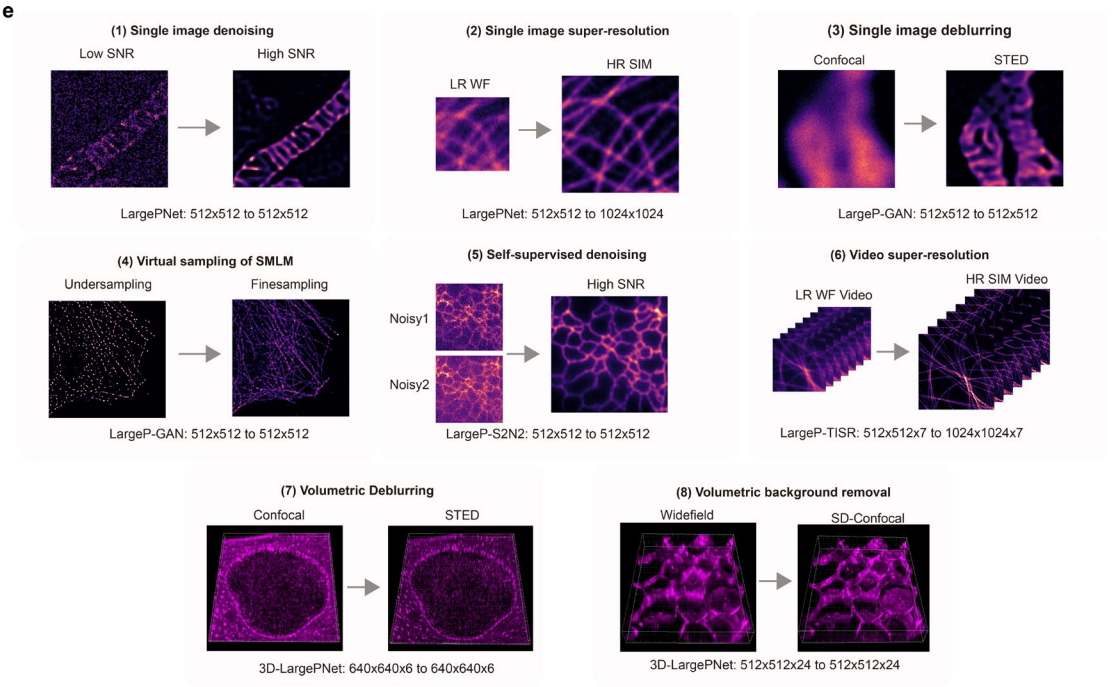
Extended data:

Fig 1

The images are very small in e. It is very difficult to assess model performance.

Response:

Thanks for your informative feedback. The subgraph e is illustrative, aiming to show the training data pairs in each restoration tasks, but not the assessment of model performance. We appreciate your feedback that the too-small image in e may hinder the readers. We now revise it to show enlarged areas for better visual perception.




The revised Extended Data Fig. 1e

Action

We have refined Extended Data Fig. 1e to strengthen its visibility.

Fig 2

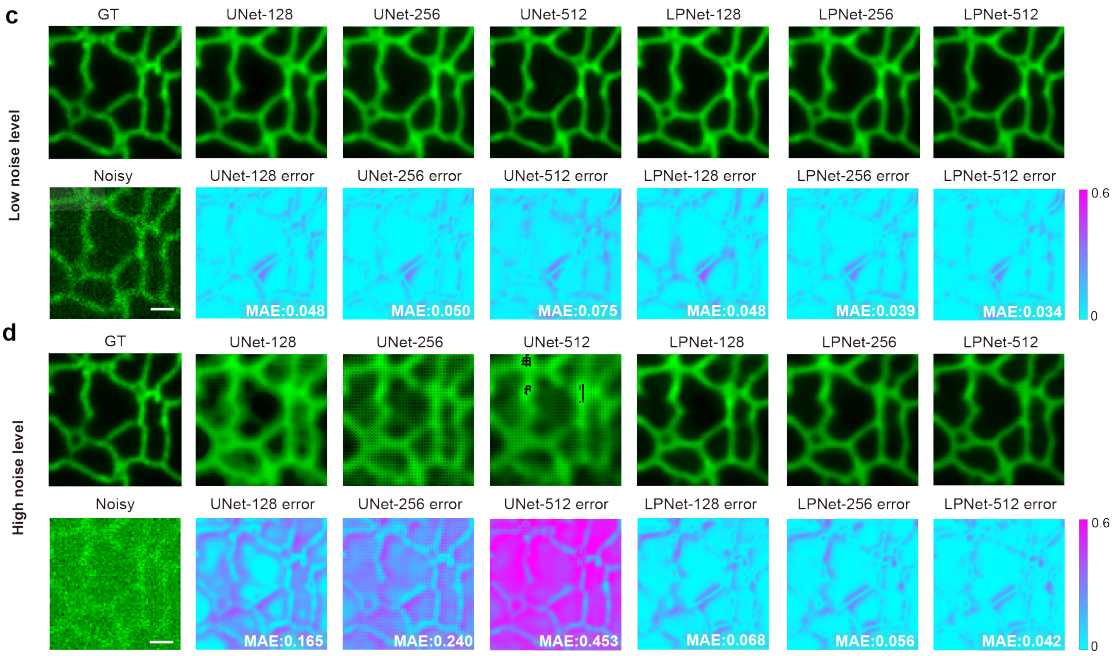
I would consider reordering the rows/columns in d. Putting the GT next to the outputs makes comparison easier. I would also suggest moving the image labels off the images themselves so that the entire image can be seen.

Response:

Thanks for your insightful suggestion. We have modified this figure according to your suggestion to make better visual quality.

Action:

We have revised Extended Data Fig. 2 according to your advice.




The revised Extended Data Fig. 2

Fig 3

See comments about statistical significance of performance gains in a.

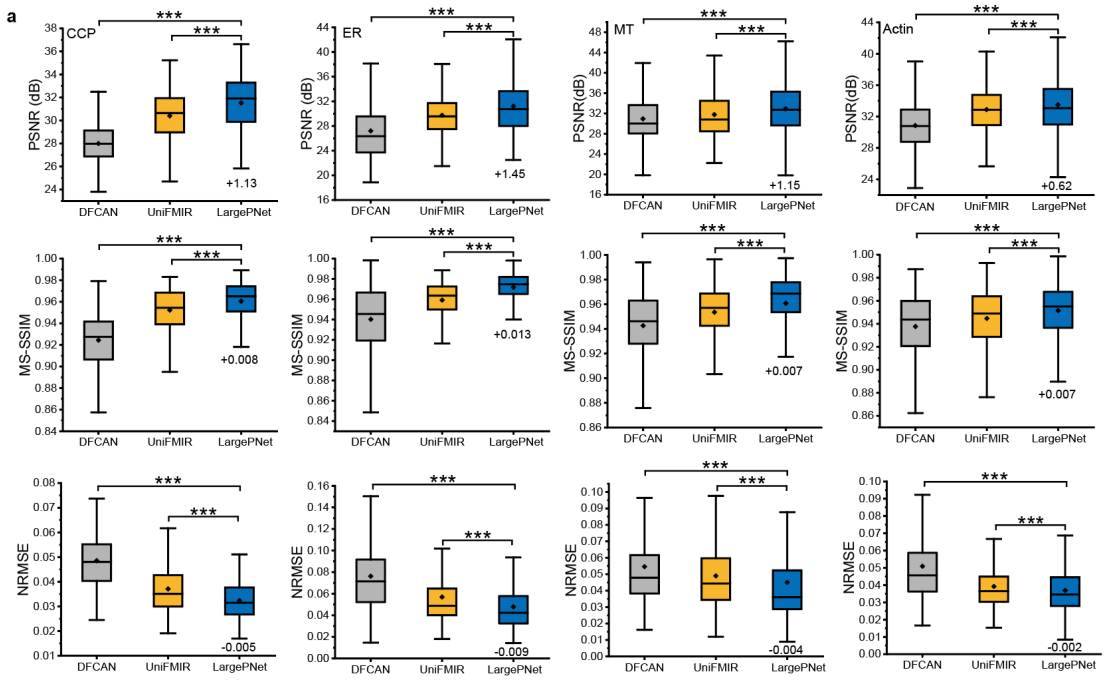
Response:

Thanks for your good suggestion. We have performed significance test of this statistical comparison, the results are shown below:

Response Table 2-13. Significance test results on Extended Data Fig. 3

| Significance of LargePNet improvement through t-test | | | | | |
|---|---|---|---|---|---|
| Model | Metric | CCPs | ER | Microtubules | F-Actin |
| Versus DFCAN | PSNR (dB) | $p$ = 1.49e-140(***) | $p$ = 3.69e-167(***) | $p$ = 1.17e-115(***) | $p$ = 4.57e-235(***) |
| | MS-SSIM | $p$ = 2.33e-91(***) | $p$ = 8.67e-129(***) | $p$ = 8.39e-215(***) | $p$ < 2.2e-308(***) |
| | NRMSE | $p$ = 4.05e-125(***) | $p$ = 3.05e-150(***) | $p$ = 7.46e-57(***) | $p$ = 1.22e-189(***) |
| Versus UniFMIR | PSNR (dB) | $p$ = 3.26e-24(***) | $p$ = 2.22e-33(***) | $p$ = 1.13e-37(***) | $p$ = 1.37e-24(***) |
| | MS-SSIM | $p$ = 1.25e-13(***) | $p$ = 5.36e-68(***) | $p$ = 4.30e-99(***) | $p$ = 5.02e-266(***) |
| | NRMSE | $p$ = 2.97e-22(***) | $p$ = 8.25e-27(***) | $p$ = 8.50e-12(***) | $p$ = 1.90e-17(***) |

From the results it can be seen that the improvement is significant in statistics. We also mark the test results directly in the figure.




The revised Extended Data Fig. 3a

Action:

Statistical significance test is performed on all the box charts. The results are marked in the figure.

See comments about arrows identifying regions in b-e.

Specifying which profiles correspond to which images—i.e. does f correspond to the region in b or c?—would reduce ambiguity in comparing the line profiles to the ROIS.

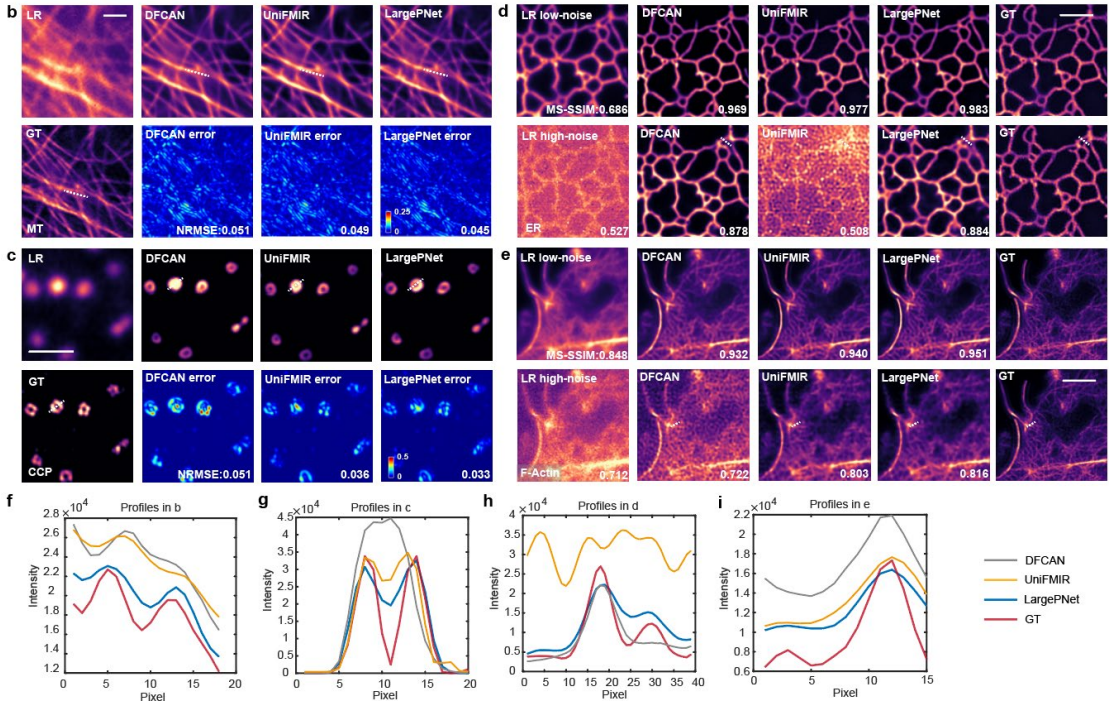
Response:

Thanks for pointing this out. The arrows in this figure aim to mark a line between the two arrowheads, and the curve is the intensity profile along the line between the two arrowheads. It may be more appropriate to be modified to a dashed line to indicate the intensity profile to be investigated.

Moreover, we have marked the identified profiles in f-i to reduce the ambiguity in comparing the line profiles.

Action:

We have refined Extended Data Fig. 3b-i following the suggestions.




The revised Extended Data Fig. 3b-i.

Fig 4

The table of comparisons is the most useful part of this figure.

Response:

Thanks for your feedback. We have removed the other parts in this figure, which shall be a common sense in this field.

Action:

We have only reserved the comparison table in the revised Extended Data Fig. 4.

Fig 5

It is difficult to parse any meaningful differences between the images in a-b at this scale.

Response:

Thank you for pointing this out. In previous version we displayed the whole image in a-b, and displayed some magnified region in c-d. We agree that for the STED deblurring task displayed in this figure, comparisons at this scale are hard to see any meaningful differences. Therefore, we modified this figure and show an ROI region between the size of the original a-b and c-d, to display the structures at a balanced view and also keep the figure compact. You can also see the modifications in our following response to your other comments.

Action:

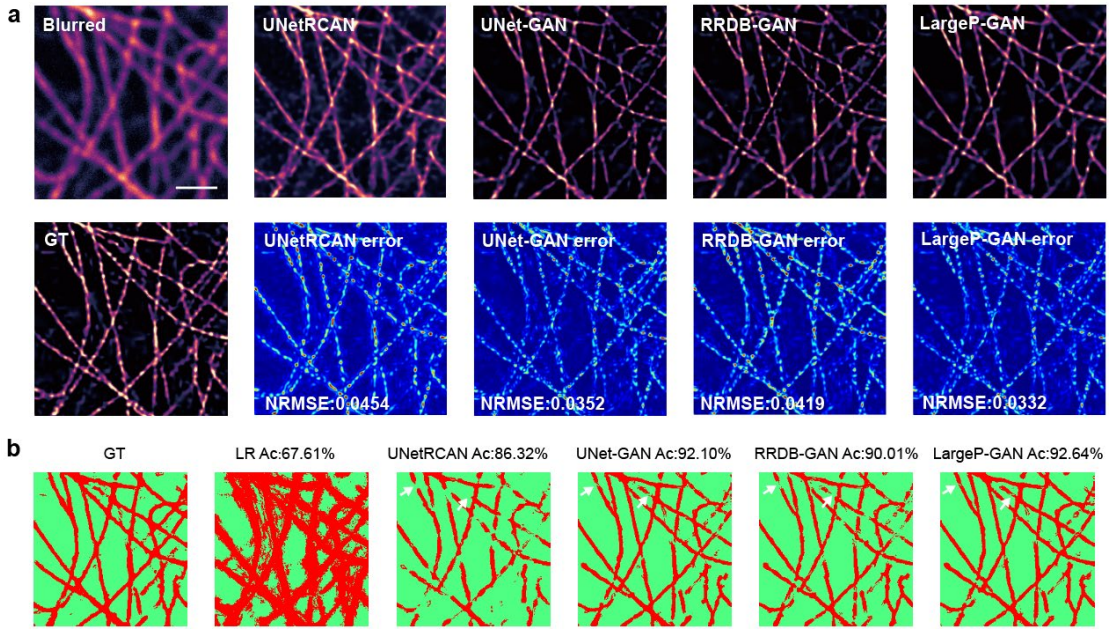
We have adjusted the displayed view size in Extended Data Fig. 5.

The similarity between UNet-GAN and LargeP-GAN in c is hard to ignore.

Response:

Thank you for pointing this out. We acknowledge that in the morphology the difference of UNet-GAN and LargeP-GAN in the microtubule network is relatively

small. This is due to that the blurred image is already very similar with the GT image, making all models learn this feature relatively easier. While LargeP-GAN (NRMSE:0.0332) still shows higher pixel-precision fidelity compared with UNet-GAN (NRMSE:0.0352), as shown in the newly added error map in a and also the statistics in subfigure d. We have also seriously taken your advice to add segmentations of this result to enrich the comparison, as shown in the newly added subfigure b. The higher gray-value fidelity helps the LargeP-GAN result to achieve a higher segmentation precision. The white arrows in b denote some microtubules that LargeP-GAN can help better identify.




The revised Extended Data Fig. 5a, b

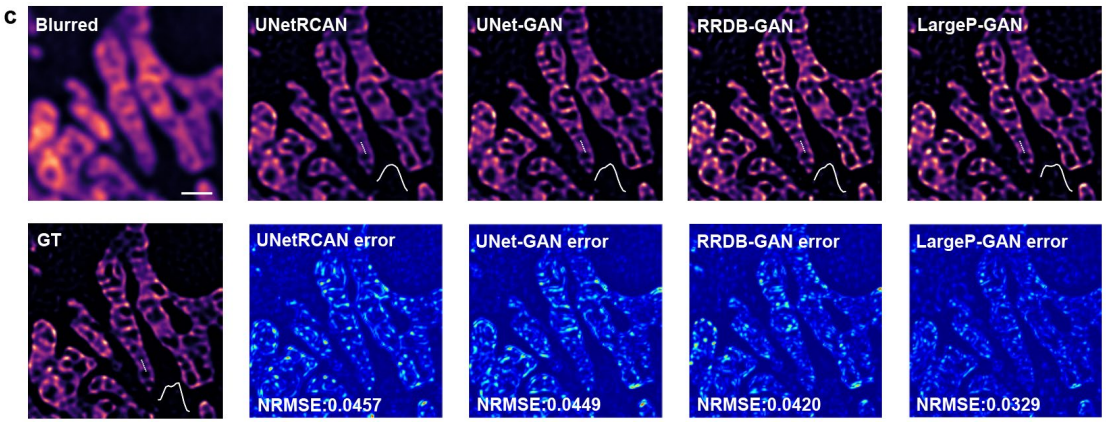
Action:

We have adjusted the displayed view size in Extended Data Fig. 5.

I am not sure that it is robust to assert the existence of a 3rd peak in either the GT or the LargeP-GAN output in d.

Response:

We appreciate you for pointing this out, since the original intensity profiles displaying a too-long sampling line would make it hard to clearly see the existence of the 3rd peak. Since the 2nd and 3rd peaks are the major difference between LaregP-GAN/GT with other prediction results, we shall focus on this more compared with the 1st peak. Therefore, we have modified this figure and only display the original 2nd and 3rd peaks, which are clear in GT and the LargeP-GAN results while not for the others, as shown below:




The revised Extended Data Fig. 5c

Action:

We have refined this subfigure to clearly show the existence of the existence of the peak in GT and LargeP-GAN results.

Once again, the metrics presented in e-f do not seem to suggest a significant difference between these methods, when taken together. However, the potentially significant increase in PSNR from LargeP-GAN is noted in f. Once again, commenting on the difference in performance across the structures would be welcome. Moreover, I wonder why f is included only in the extended data and not the main figure.

Response:

Thanks for your insightful question, which helps strengthen our manuscript. Since this comment contains several points, below we respond in a point-by-point manner.

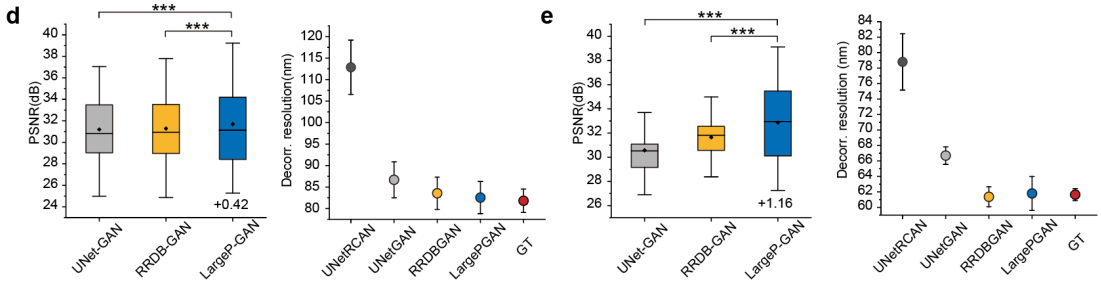
(1) Significance test

We have performed significance test of the box chart in this figure, the result is shown in Response Table 2-14 below. It can be seen that the improvement of LargePNet is statistically significant. The source data are also uploaded to ensure the reproducibility. Multiple noise and blurring levels are the reason for the long whisker in the box chart display, which may obscure the disparity in different models.

Response Table 2-14. Significance test results on Extended Data Fig. 5

| Model | Metric | Microtubules | Mito |
|---|---|---|---|
| Versus UNet-GAN | PSNR (dB) | $p$ = 3.27e-4(***) | $p$ = 1.06e-12(***) |
| Versus RRDB-GAN | PSNR (dB) | $p$ = 3.83e-4(***) | $p$ = 1.52e-6(***) |

We also marked the test results directly on the graph, and also annotated the PSNR improvement of LargePNet compared to the second-best model in the figure:




The revised Extended Data Fig. 5d, e

Action:

Significance test is performed on the results. The PSNR improvement of LargePNet over the second-best model is annotated in the figure for clearer presentation.

(2) Commenting on difference in performance across the structures

Analysis results of this dataset are shown in Response Table 2-5, which is recalled below. It can be seen that LargePNet has achieved greater advantages over other structures in the Mito IMM structure compared with MT. Inspecting the GLCM statistics, it can be found that the Mito IMM shows an evidently higher large and small views difference (2.02-fold in Energy metric, 0.0652 in Correlation metric) compared with MT (1.47-fold in Energy metric, 0.0184 in Correlation metric). Mapping this indicator to the practical image, this task aims to remove the blurring degeneration in STED images. For MT as a relatively simple structure, small-patch cropping may not disturb the contextual structural information as the complex Mito IMM structure does. While for the Mito IMM structure, which is relatively complex, the advantage of LargeP-GAN's large-view statistics aggregation will exert, and cause a significant performance difference.

Recall: Response Table 2-5. Explaining structure-dependent advantage of LargePNet in the STED deblurring task using the GLCM analysis method.

| LargeP-GAN PSNR gain compared with the best small-patch model | | |
|---|---|---|
| Structure | MT | Mito IMM |
| PSNR(dB) gain | +0.42dB | **+1.16dB** |
| GLCM Contrast statistics difference analysis | | |
| Large-view | 4.80-fold | 1.02-fold |
| Small-view | 3.27-fold | 1.41-fold |
| Large-small Diff | 1.47-fold | 1.38-fold |
| GLCM Energy statistics difference analysis | | |
| Large-view | 0.085-fold | 0.155-fold |
| Small-view | 0.058-fold | 0.313-fold |
| Large-small Diff | 1.47-fold | **2.02-fold** |
| GLCM Correlation statistics difference analysis | | |
| Large-view | 0.0177 | 0.0961 |
| Small-view | 0.0361 | 0.0309 |
| Large-small Diff | 0.0184 | **0.0652** |

Action:

Structural disparity is discussed in Supplementary Note 6 and Supplementary Table 14 in the revised manuscript.

(3) Why f is included only in the extended data and not the main figure

The statistics in f is essentially for this figure specifically, which is essentially a different restoration task from the main figure. The main figure (Fig. 3) displays the STED denoising task, aiming to remove the random noise in STED image. While this figure (Extended Fig. 5) displays the STED deblurring task, aiming to transform Confocal or low-resolution STED images to a high-resolution one. Since the STED denoising could be more commonly used, we place it in the main figure.

Fig 6

See comments about statistical significant and disparity in comparative performance across different structures.

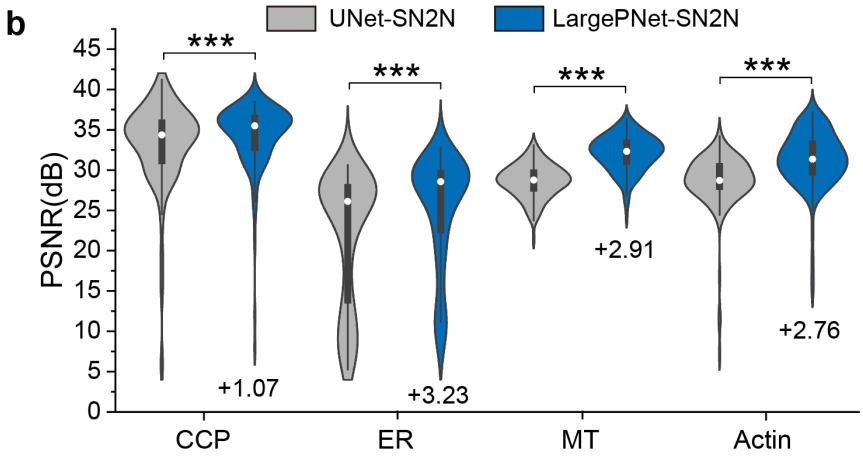
Response:

Thanks for your constructive suggestion. We have performed significance analysis of the violin chart result, showing that in all cases the improvement of LargeP-SN2N over UNet-SN2N is very significant ($p < 0.001$, ***). The source data are uploaded with the manuscript to ensure the reproducibility.

Response Table 2-15. Significance test on the results in Extended Data Fig. 6.

| Significance of LargePNet improvement through t-test | | | | | |
|---|---|---|---|---|---|
| Model | Metric | CCPs | ER | Microtubules | F-Actin |
| Versus UNet-SN2N | PSNR (dB) | $p$ = 1.63e-6(***) | $p$ = 6.53e-21(***) | $p$ = 4.02e-29(***) | $p$ = 5.11e-13(***) |

We have also marked the test results and the PSNR improvement of LargePNet in the violin chart.




The revised Extended Data Fig. 6

In this experiment, LargePNet achieved significant improvement in PSNR across

all structures. The PSNR of CCP structure is slightly weaker compared to other structures. It should be pointed out that this is a self-supervised denoising experiment. The model will learn to distinguish structure and noise information from noisy images without GT information supervision. Therefore, unlike the GLCM analysis of previous supervised restoration tasks, we no longer calculate the degree of change in GLCM statistics between noisy images and GT images. We only examine the GLCM statistical characteristics of noisy images from the views of small and large images. The results are shown in Response Table 2-16.

Response Table 2-16. Examining structural difference in GLCM statistics.

| Structure | CCP | ER | MT | F-actin |
|---|---|---|---|---|
| PSNR(dB) gain | **+1.07dB** | **+3.23dB** | **+2.91dB** | **+2.76dB** |
| GLCM Contrast statistics of the noisy image | | | | |
| Large-view | 9.8727 | 23.5682 | 14.1993 | 13.3843 |
| Small-view | 3.1173 | 7.6815 | 4.2809 | 4.8178 |
| Large-small Diff | **3.2-fold** | **3.1-fold** | **3.3-fold** | **2.8-fold** |
| GLCM Energy statistics of the noisy image | | | | |
| Large-view | 0.156 | 0.0408 | 0.0235 | 0.0152 |
| Small-view | 0.0678 | 0.0142 | 0.0116 | 0.0099 |
| Large-small Diff | **2.3-fold** | **2.9-fold** | **2.0-fold** | **1.5-fold** |
| GLCM Correlation statistics difference analysis | | | | |
| Large-view | 0.8518 | 0.6909 | 0.9452 | 0.9081 |
| Small-view | 0.7426 | 0.5396 | 0.8616 | 0.8618 |
| Large-small Diff | **0.1092** | **0.1513** | **0.0836** | **0.0463** |

From the results, it can be seen that there are significant differences in GLCM statistics between small and large views on all structures. This explains why LargePNet has achieved significant improvement in PSNR across all structures. But the difference between the large and small views of CCP is not weaker than that in other structures, which may not be enough to explain why the increment in CCP structure is relatively small. We speculate that this structure may be too simple in this task (self-supervised wide-field CCP denoising only need to resolve the spherical shape of the CCP, while the CCP super-resolution task in Fig. 2 needs to resolve its circular structure). The UNet-SN2N model has already achieved very good restoration results (PSNR>33dB), so the improvement of LargePNet on this basis is slightly lower.

Action:

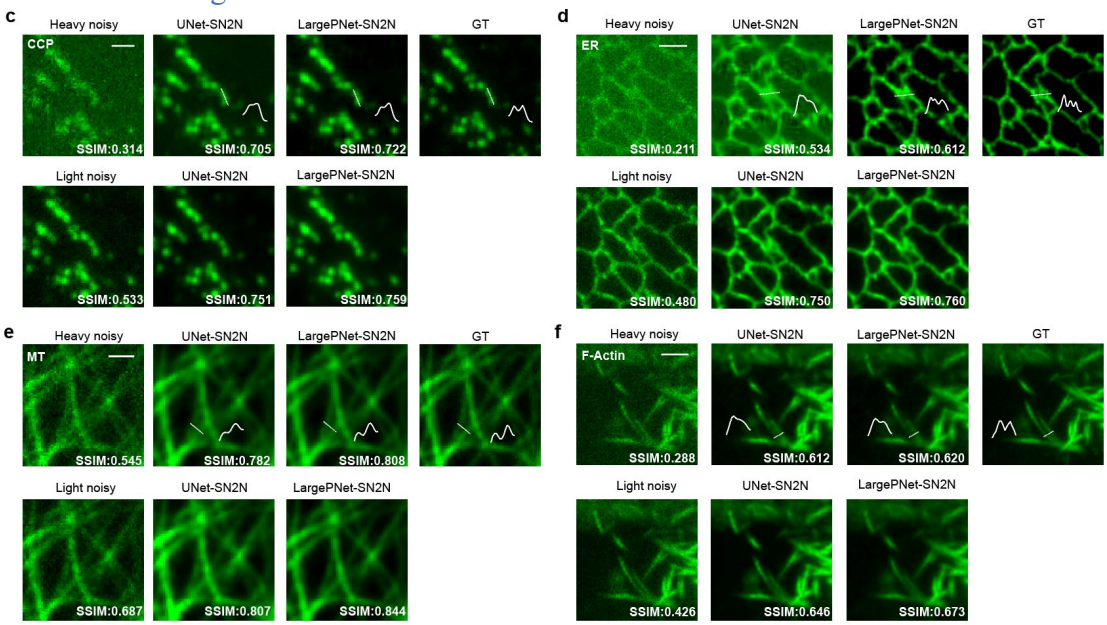
Statistical significance test is performed and marked on this violin chart result.

The difference between the middle peaks in the LargePNet prediction and GT should be discussed if they are going to be highlighted.

Response:

Thanks for your kind feedback. This figure demonstrates that the advantages of LargePNet can be generalized to the self-supervised denoising task. In this task the model would learn to transform noisy images to a clean one without knowing the

ground-truth clean image. Since the raw image is down-sampled to two sub-images and put for training, the model may not be aware of the precise position information in the original scale, different from the supervised task. This may cause the more evident shifts of peaks in the prediction image compared with the supervised tasks. We apologize for the misunderstanding caused by the previous arrangement. We are not aiming to emphasize these profiles, they are just some additional data displays. An objective quantitative metric like PSNR is the gold standard for evaluating models. We have rearranged the graph in consistency with the displays in other figures to avoid misunderstandings:




Extended Data Fig. 6c-f

Action:

Extended Data Fig. 6c-f have been refined in the revised manuscript.

Reviewer #3 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Response:

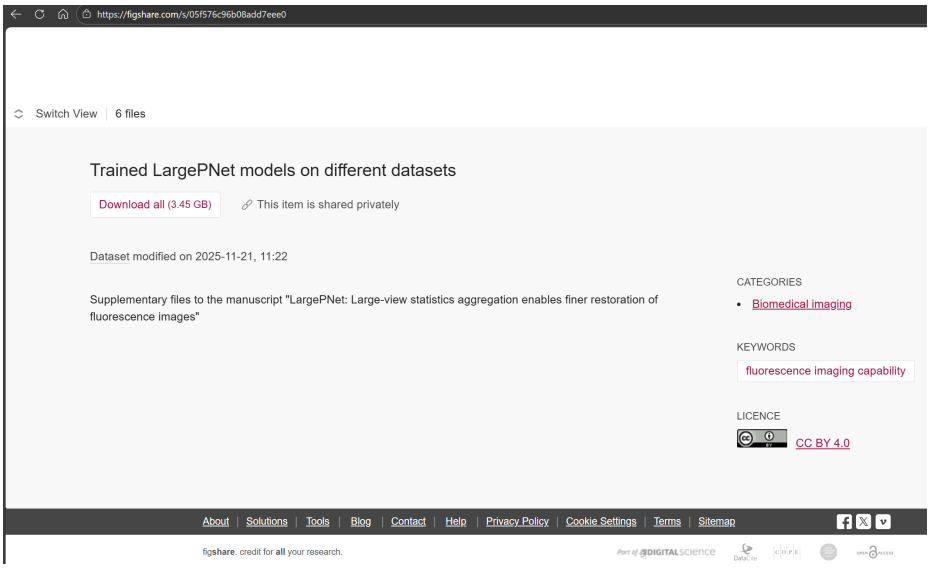
We would like to extend our sincere gratitude to you for dedicating your valuable time and expertise to the peer-review of our manuscript. We are particularly grateful for your participation in the collaborative review model initiated by Nature Communications. This helps improve our manuscript in its breadth, depth, and solidness.

Reviewer #3 (Remarks on code availability):

The repository does not appear to contain the pre-trained models required to test inference using the provided Jupyter notebooks. I can thus not comment on the efficacy of the code in terms of reproducing the reported results.

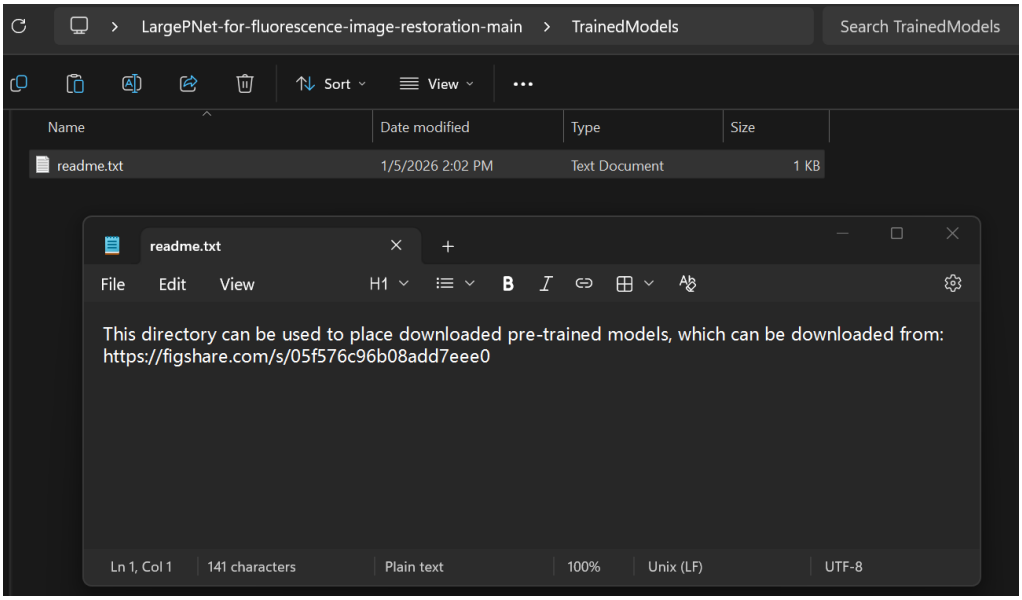
Response:

Thanks for your feedback. Due to the file size constraint (all models 3.45G), the trained models are uploaded in a figshare link for reproduction: https://figshare.com/s/05f576c96b08add7eee0. As shown below:




A screen-shot from the trained model link: https://figshare.com/s/05f576c96b08add7eee0

In previous version, we only state this link in the Data availability section in the manuscript, which may be inconvenient for users to find. Now we have added this link information to the "TrainedModels" folder of the code file to remind users. In the revised code file, you can find a 'Readme.txt' file, which now states this link, as shown below:




A screen-shot of the added readme.txt file in the TrainedModels folder

Action:

In addition to the statement in the Data Availability section, the figshare link of the trained models is also stated in the code repository and the uploaded source code file.

# Response to Reviewers' Comments

We very much appreciate the critical reading of our manuscript and valuable suggestions from the reviewers. We have carefully reviewed the comments and have revised the manuscript accordingly. To ensure clarity and facilitate the review process, all significant changes have been highlighted in red. The responses to the comments are listed one by one as follows (in blue):

Reviewer #1 (Remarks to the Author):

In the revised version of "Pushing the limits of fluorescence imaging with a restoration neural network aggregating large-view statistics", Hou et al have significantly improved on their original submission.
They responded to my original points and have produced a clearer and more complete paper suitable for publication.
I do feel that the 70 page response to reviewers was totally excessive and the responses could have been more concise and still adequately covered the ground.
Response:

We would like to express our sincere gratitude to you for the constructive comments and valuable suggestions, which have greatly helped us improve the quality of the manuscript.

Additional minor issues in added text:
Introduction
"The small patch truncated many contextual structures and imposes the network to learn restoration with limited information."
You must impose on something, so this could be reworded but probably better to just replace imposes with forces.
Response:

Thanks for your constructive suggestion, we have revised “imposes” to “forces” for better presentation.

"However, the application image typically span large views with much richer contextural structure information, which shall be important restoration clues but remain largely unexplored"
Either needs to be "image typically spans" or "images typically span" depending on if they are referring to a single application image or the whole set of images. "Which shall be" would be better phrases as "which are"
Response:

Thanks for your constructive suggestion. We aim to refer to a single application image here. We have revised “span” to “spans”. We also revise “shall be” to “which are” for better presentation.

"Moreover, the difference of the structure statistics in the large view and small patch",
The critical factor is not the difference in the large view and small patch results but the differences between them.
Response:

Thanks for pointing this out. We have revised the original sentence to "Moreover, the differences of the structure statistics between the large view and small patch",

Discussion

"It should be noted that despite LargePNet significantly improves the restoration performance of previous models, "
Should the "significantly improving the"
Response:
Thanks for your constructive suggestion, we have modified “significantly improves” to “greatly improving” in the revised manuscript.

"Besides fluorescence microscopy, we believe that LargePNet can also improve the DL restoration performance in images similar to fluorescence images, such as label-free and electron microscope images."
If the authors wish to claim that their model CAN improve other modalities they should show that. More realistically they should say may or could improve.
Response:
Thanks for your constructive suggestion, we have revised “can” to “could” for better presentation.

Reviewer #2 (Remarks to the Author):

The authors have done a fantastic and thorough job of addressing the points laid out in the previous review, and we therefore feel it is sufficiently ready for publication.
Response:
We would like to extend our deepest gratitude for your positive evaluation and recognition of our revised manuscript. We will continue to maintain the rigor and quality of the manuscript to ensure it meets the publication standards of the journal.

Reviewer #3 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.
Response:
We sincerely thank you for participating in the peer review process of our manuscript. We fully support the Nature Communications initiative to facilitate training in peer review and provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Additional revisions

In addition to the reviewer's comments, we also made detailed revisions to the article according to the requirements outlined in the editor's checklist document. We conducted an overall inspection of the article for grammatical and spelling errors and made the necessary corrections.