Supplementary material

Photoacoustic microscopy with sparse data by convolutional neural networks

Jiasheng Zhou, Da He, Xiaoyu Shang, Zhendong Guo, Sung-Liang Chen, Jiajia Luo

Study of training and inference time of the CNN models

In our experiment, we trained 40000 iterations with a batch size of 8 for Residual U-Net and our network and batch size of 6 for EDSR. During training, the intermediate models were evaluated every 200 iterations using the validation set. The intermediate model with the best validation performance was selected and used as the final model in this work.

The actual training time after 40000 iterations of the three CNNs using the leaf vein dataset is listed in Table S1. The values were calculated using the platform with Intel Xeon Silver 4110 CPU and a single Nvidia 2080Ti GPU. Note that the validation time is included during the training.

Table S1: Training time of the three CNN methods (mins/40K iterations).

Method	EDSR	Residual U-Net	Ours
Training time for 2× scaling	1350.00	214.83	437.97
Training time for 4× scaling	480.97	246.87	403.20

The average inference/evaluation time of the three CNNs using the leaf vein test set is listed in Table S2. The values were also calculated using the platform with Intel Xeon Silver 4110 CPU and a single Nvidia 2080Ti GPU.

Table S2: Inference time of the three CNN methods (s/image).

Method	EDSR	Residual U-Net	Ours
Inference time for 2× scaling	0.2022	0.0806	0.1078
Inference time for 4× scaling	0.0963	0.0829	0.1065

According to the above two tables, Residual U-Net is the fastest for both training and inference. The proposed method is slower than Residual U-Net for training time and comparable for inference time. EDSR requires the longest training time and takes much more time for the 2× scaling case than the 4× scaling case, while the other two CNN methods take comparable time for the 2× and 4× scaling cases.

Study using a quantitative perceptual metric

As indicated in Section 3.3 and Fig. 6 in the main text, there are some limits for PSNR and SSIM metrics as they give high scores to the over-smoothed images. Despite the perceptual comparison without quantitative metrics mainly described in Section 3.3 and Fig. 6 in the main text, some quantitative metrics were proposed before for perceptual evaluation. Therefore, a quantitative perceptual metric was used for comparison, and the results are shown in Table S3. Specifically, we applied Fréchet Inception Distance (FID) [S 1] to measure the perceptual quality of the results obtained with different loss functions (i.e., MSE, MAE, and the perceptual loss) for the ablation study. Smaller FID values indicate higher quality.

Table S3: Comparison among the models trained with different loss functions on 4× leaf vein test set.

Loss function	PSNR (dB)	SSIM	FID
MSE	23.4113	0.6930	222.8991
MAE	**23.8995**	**0.7512**	173.0127
Perceptual loss	23.1760	0.7159	**43.6614**

According to Table S3, although the model trained with MAE achieves the highest mean PSNR and mean SSIM values, the model trained with perceptual loss shows the best performance with respect to the FID metric, which is consistent with the qualitative results in Fig. 6. Therefore, although not a gold standard, FID can be used for perceptual evaluation to some extent.

To our understanding, it is a challenge for the research community about how to guarantee that "more qualitatively similar images represent a more accurate recreation of the ground truth", which may be out of the scope of this study. Instead, we followed the convention in previous studies [S 2][S 3] to consider the trade-off between PSNR (or SSIM) and the visual performance in our demonstrations in the main text.

Study of asymmetric low-sampling PAM data

In the main text, the same downsampling rates along the two scanning axes were studied and presented. We further conducted experiments using the leaf vein dataset to study "asymmetric" low-sampling PAM data (i.e., different downsampling rates along the two scanning axes). By changing the scaling arguments in the "Upconv" module in the proposed CNN, our method can be applied to the case of asymmetric low-sampling PAM data.

Without loss of generality, two asymmetric low-sampling operations were applied in our experiments. First, we used the low-sampling rates of 1/2 along the vertical direction (or y direction) and of 1/4 along the horizontal direction (or x direction) of the PAM image. This case is denoted as "2×—4×". On the other hand, the case of "4×—2×" was also studied. The sparse sampling density

of the two cases is thus 1/8. Other training settings for these two new cases were kept the same, as described in the main text. After training, the statistical performance on the test set is included in Table S4. One representative result is shown in Fig. S1 for qualitative evaluation.

Table S4: Statistical results for experiments with different low-sampling PAM data.

Scaling	Sampling density	Bicubic		Our method	
		PSNR (dB)	SSIM	PSNR (dB)	SSIM
2×—2×	1/4	23.4936±1.8718	0.7721±0.0457	**26.1431**±1.7022	**0.8183**±0.0599
2×—4×	1/8	21.7742±2.0937	0.6547±0.0575	**24.1429**±1.6319	**0.7870**±0.0718
4×—2×	1/8	22.0679±2.0713	0.6756±0.0541	**24.1399**±2.0457	**0.7913**±0.0501
4×—4×	1/16	19.9941±1.9204	0.5773±0.0683	**23.1760**±1.9290	**0.7159**±0.0602

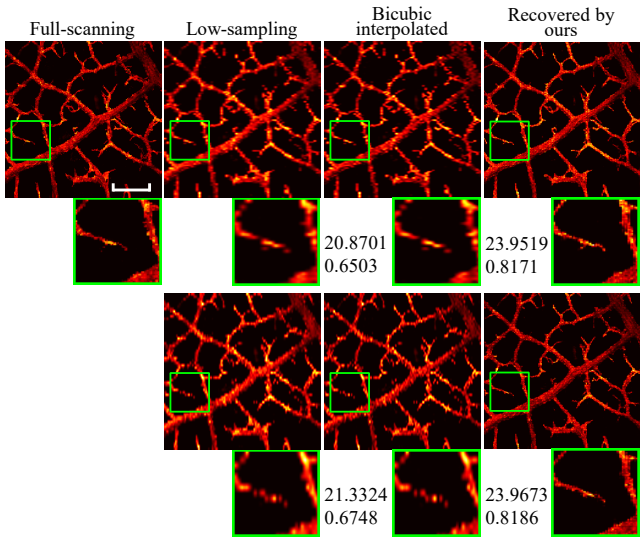

Fig. S1: Results of the leaf vein experiment with asymmetric low-sampling PAM data. The first row (except the image of "full-scanning") indicates the case of "2×—4×", and the second row corresponds to the case of "4×—2×." The numbers below the images indicate the PSNR (dB) and SSIM values (by comparing the entire image with the corresponding ground truth). The sample comes from a magnolia leaf. Scale bar: 500 μm. All images, excluding zoom images, share the same scale bar.

According to Table S4 and Fig. S1, the proposed method can still well recover the low-sampling PAM images in these two asymmetric cases. From Table S4, the mean PSNR and SSIM values of the two asymmetric cases (sampling density of 1/8) are between those of the symmetric cases (sampling density of 1/4 and 1/16), which is reasonable. Besides, the improvement in PSNR and SSIM of the asymmetric cases keeps the same trend as in the main text (i.e., more improvements for the case of lower sparse sampling density by our method

compared with bicubic interpolation). In Fig. S1, issues like blurring, over smoothing, and discontinuity are not observed with our method. Therefore, our method is suitable for the recovery of asymmetric low-sampling sparse PAM data as well.

References

[S1] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, Gans trained by a two time-scale update rule converge to a local nash equilibrium, in: Advances in neural information processing systems, 2017, pp. 6626–6637.

[S2] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4681–4690. doi:10.1109/cvpr.2017.19.

[S3] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: European conference on computer vision, Springer, 2016, pp. 694–711. doi:10.1007/978-3-319-46475-6_43.