Appendix

In this supplementary material, we provide details on the pathologists' assessment and more qualitative results on PanNuke [1], Lizard [2], and EndoNuke [3].

Appendix A. Qualitative Assessment by Pathologists

Fig. 1 shows the instructions for the qualitative assessment survey conducted by pathologists. The following text describes the open-response questions asked in the survey and the pathologists' responses.

Q1. What were the key considerations when evaluating image quality?

- Whether the image represent a tissue morphology that could exist in reality.
- Whether nuclei membranes, nucleoli, and chromatin can be distinguished.
- Whether the boundaries between cells can be seen, including resolution, noise, and color contrast.
- Good match to actual histologic structures
- Diagnostic potential of the tissue.

Q2. What were the key considerations when evaluating label quality?

- Errors seem to occur when connective tissue cells are in between inflammation, but evaluation of this is limited due to difficulty in accurately determining GT for these cells.
- Differentiation of epithelial and inflammatory cell.
- Eosinophil matching, differentiation of (1) neutrophil and karyorrhexis, (2) plasma and stromal cell, (3) lymphocyte and degenerated epithelial cell.
- Whether to differentiate between lymphocytes and connective tissue, which are relatively difficult to distinguish.
- Identify the location and type of cell.

Q3. What could be improved in the synthetic images?

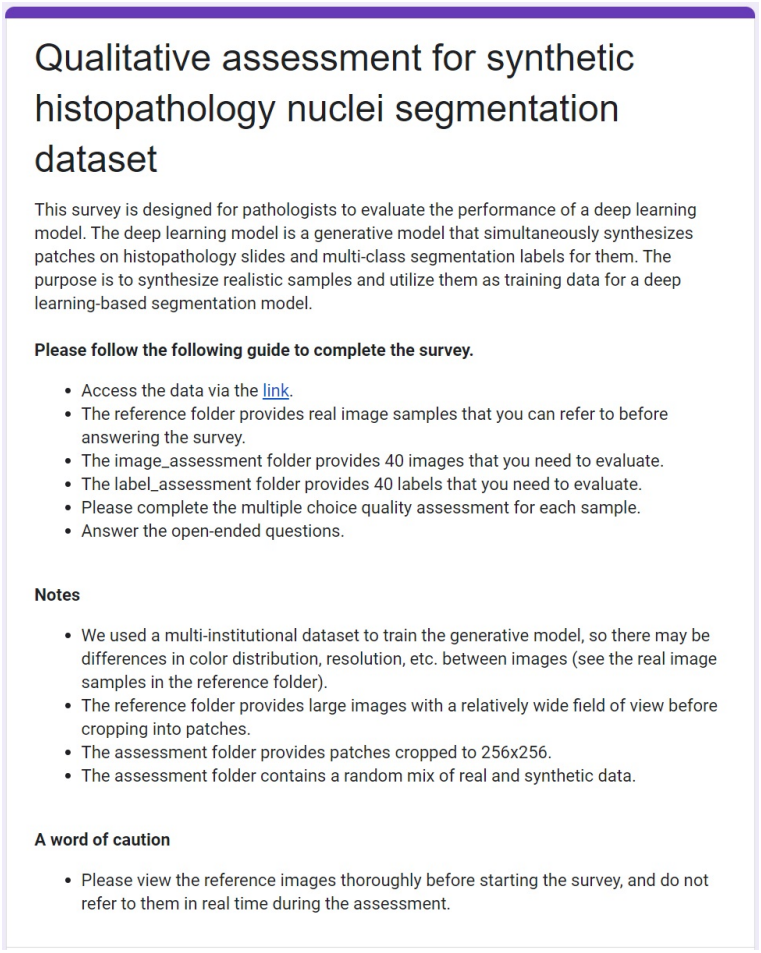
- Reproduction of polarity loss and disorientation depending on the actual malignancy of the cell.
- Sharpness (resolution), finer differentiation of nuclei.
- H&E stain is too intense or too light in some areas.
- Some blurry or fragmented images.

Q4. What could be improved in the synthetic labels?

- Error occurs when connective tissue is between inflammation.
- Reliable differentiation of epithelial cells.
- Differentiation of (1) neutrophil and karyorrhexis, (2) lymphocytes and fibroblasts in connective tissue, and (3) connective tissue and lymphocyte.

Q5. Please feel free to provide any additional comments.

- Since there were only three possible choices to the question, I graded more conservatively, but overall, the synthetic image quality and label performance were good.



Qualitative assessment for synthetic histopathology nuclei segmentation dataset

This survey is designed for pathologists to evaluate the performance of a deep learning model. The deep learning model is a generative model that simultaneously synthesizes patches on histopathology slides and multi-class segmentation labels for them. The purpose is to synthesize realistic samples and utilize them as training data for a deep learning-based segmentation model.

Please follow the following guide to complete the survey.

- Access the data via the link.
- The reference folder provides real image samples that you can refer to before answering the survey.
- The image_assessment folder provides 40 images that you need to evaluate.
- The label_assessment folder provides 40 labels that you need to evaluate.
- Please complete the multiple choice quality assessment for each sample.
- Answer the open-ended questions.

Notes

- We used a multi-institutional dataset to train the generative model, so there may be differences in color distribution, resolution, etc. between images (see the real image samples in the reference folder).
- The reference folder provides large images with a relatively wide field of view before cropping into patches.
- The assessment folder provides patches cropped to 256x256.
- The assessment folder contains a random mix of real and synthetic data.

A word of caution

- Please view the reference images thoroughly before starting the survey, and do not refer to them in real time during the assessment.

Fig. 1. Introduction to the qualitative assessment of synthetic data survey.

Appendix B. Color Quality Evaluation

To evaluate the color quality, we present several synthetic and real images for each dataset as depicted in Figs. 2-4. Our method consistently mimics the color distribution of the real data. Moreover, Fig. 2 shows that our method generates images considering the color features that vary depending on the tissue type. SDM-generated samples exhibit unrealistic but diverse colors, resulting in a high Inception Score (IS).

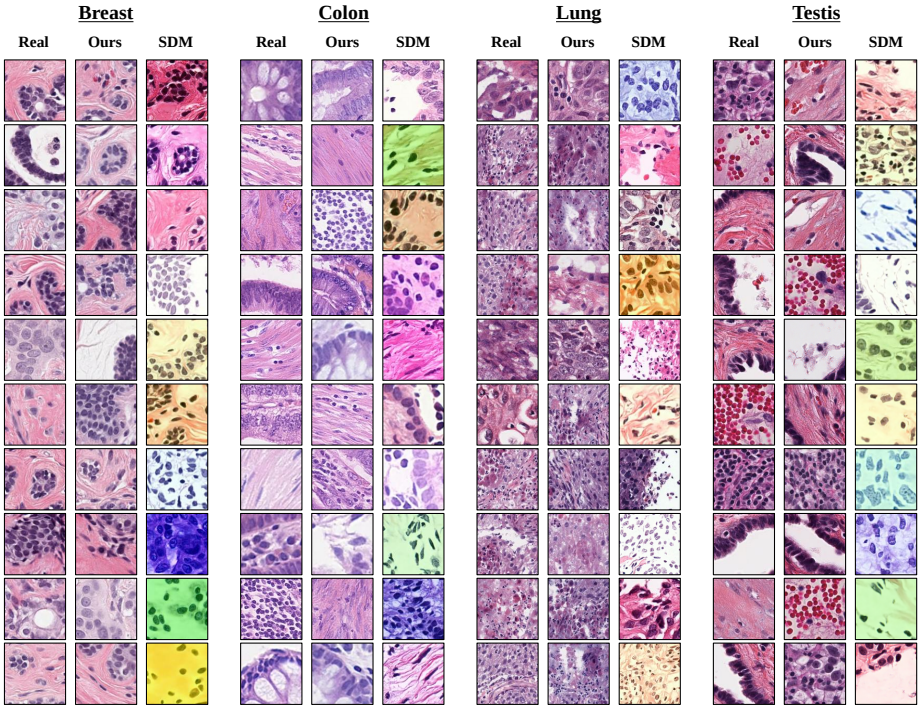


Fig. 2. Qualitative result on PanNuke.

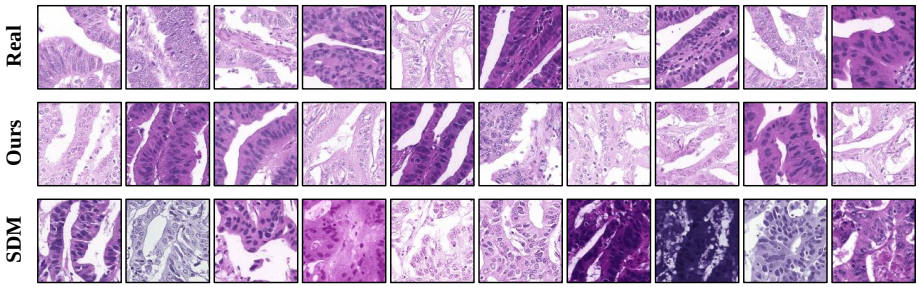


Fig. 3. Qualitative result on Lizard.

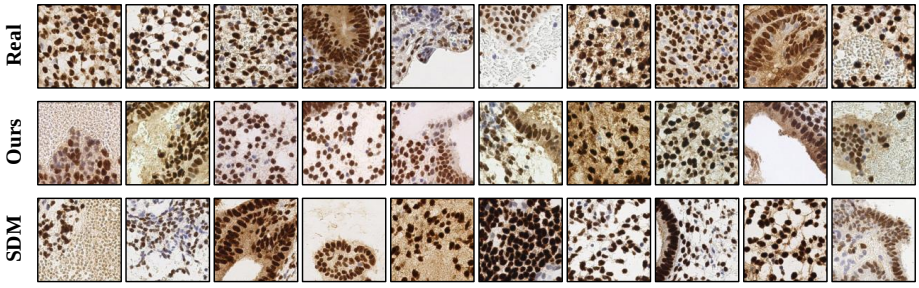


Fig. 4. Qualitative result on EndoNuke.

References

1. Gamper, J., Koohbanani, N.A., Benes, K., Graham, S., Jahanifar, M., Khurram, S.A., Azam, A., Hewitt, K., Rajpoot, N.: Pannuke dataset extension, insights and baselines. arXiv preprint arXiv:2003.10778 (2020)
2. Graham, S., Jahanifar, M., Azam, A., Nimir, M., Tsang, Y.W., Dodd, K., Hero, E., Sahota, H., Tank, A., Benes, K., et al.: Lizard: A large-scale dataset for colonic nuclear instance segmentation and classification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 684–693 (2021)
3. Naumov, A., Ushakov, E., Ivanov, A., Midiber, K., Khovanskaya, T., Konyukova, A., Vishnyakova, P., Nora, S., Mikhaleva, L., Fatkhudinov, T., et al.: Endonuke: Nuclei detection dataset for estrogen and progesterone stained ihc endometrium scans. Data **7**(6), 75 (2022)