



Medical Boundary Diffusion Model for Skin Lesion Segmentation

Jiacheng Wang^{1,2}, Jing Yang¹, Qichao Zhou^{2(✉)}, and Liansheng Wang^{1(✉)}

¹ Department of Computer Science at School of Informatics, Xiamen University,
Xiamen 361005, China

{jiachengw, jingy77}@stu.xmu.edu.cn, lswang@xmu.edu.cn

² Manteia Technologies Co., Ltd., Xiamen 361005, China
zhouqc@manteiatech.com

Abstract. Skin lesion segmentation in dermoscopy images has seen recent success due to advancements in multi-scale boundary attention and feature-enhanced modules. However, existing methods that rely on end-to-end learning paradigms, which directly input images and output segmentation maps, often struggle with extremely hard boundaries, such as those found in lesions of particularly small or large sizes. This limitation arises because the receptive field and local context extraction capabilities of any finite model are inevitably limited, and the acquisition of additional expert-labeled data required for larger models is costly. Motivated by the impressive advances of diffusion models that regard image synthesis as a parameterized chain process, we introduce a novel approach that formulates skin lesion segmentation as a boundary evolution process to thoroughly investigate the boundary knowledge. Specifically, we propose the Medical Boundary Diffusion Model (MB-Diff), which starts with a randomly sampled Gaussian noise, and the boundary evolves within finite times to obtain a clear segmentation map. First, we propose an efficient multi-scale image guidance module to constrain the boundary evolution, which makes the evolution direction suit our desired lesions. Second, we propose an evolution uncertainty-based fusion strategy to refine the evolution results and yield more precise lesion boundaries. We evaluate the performance of our model on two popular skin lesion segmentation datasets and compare our model to the latest CNN and transformer models. Our results demonstrate that our model outperforms existing methods in all metrics and achieves superior performance on extremely challenging skin lesions. The proposed approach has the potential to significantly enhance the accuracy and reliability of skin lesion segmentation, providing critical information for diagnosis and treatment. All resources will be publicly available at <https://github.com/jcwang123/MBDiff>.

Keywords: Skin lesion segmentation · Diffusion model

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43901-8_41.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14223, pp. 427–436, 2023.
https://doi.org/10.1007/978-3-031-43901-8_41

1 Introduction

Segmentation of skin lesions from dermoscopy images is a critical task in disease diagnosis and treatment planning of skin cancers [17]. Manual lesion segmentation is time-consuming and prone to inter- and intra-observer variability. To improve the efficiency and accuracy of clinical workflows, numerous automated skin lesion segmentation models have been developed over the years [1, 2, 7, 10, 18, 19, 21]. These models have focused on enhancing feature representations using various techniques such as multi-scale feature fusion [10], attention mechanisms [1, 7], self-attention mechanisms [18, 19], and boundary-aware attention [2, 18, 19], resulting in significant improvements in skin lesion segmentation performance. Despite these advances, the segmentation of skin lesions with ambiguous boundaries, particularly at extremely challenging scales, remains a bottleneck issue that needs to be addressed. In such cases, even state-of-the-art segmentation models struggle to achieve accurate and consistent results.

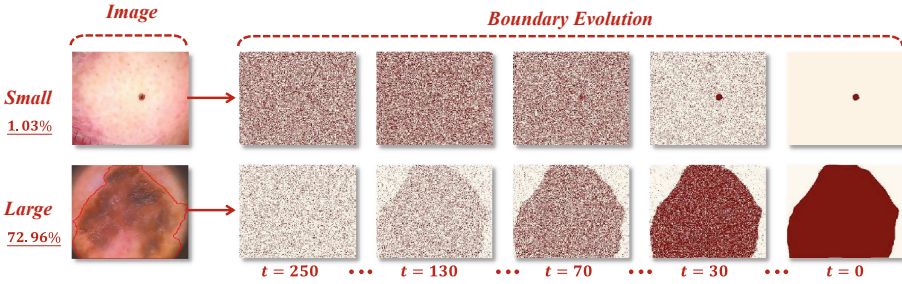


Fig. 1. The boundary evolution process. It could be seen that various lesions can be accurately segmented by splitting the segmentation into sequential timesteps (t), named as **boundary evolution** in this work.

Two representative boundaries are visualized in Fig. 1, where one extremely small lesion and one particularly large lesion are presented. The small one covers 1.03% in the image space and the large one covers 72.96%. As studied prior, solving the segmentation problems of such two types of lesions have different strategies. (1) For the small lesions, translating the features at a lower depth to the convolutional layers at a higher depth can avoid losing local contexts [10]. (2) For the large lesions, enlarging the receptive field by dilated convolution [1], and even global attention [18] can capture the long-range dependencies to improve the boundary decision. Besides the challenge of how to yield stable representations for various scales, multi-scale lesions will cause training fluctuation, that is, small lesions usually lead to large Dice loss. Feeding more boundary-aware supervision can reduce these negative effects to some degree [2, 19]. The latest transformer, Xbound-Former, comprehensively addresses the multi-scale boundary problem through cross-scale boundary learning and exactly reaches higher performance on whatever small or large lesions.

However, current models for skin lesion segmentation are still struggling with extremely challenging cases, which are often encountered in clinical practice. While some approaches aim to optimize the model architecture by incorporating local and global contexts and multi-task supervision, and others seek to improve performance by collecting more labeled data and building larger models, both strategies are costly and can be limited by the inherent complexity of skin lesion boundaries. Therefore, we propose a novel approach that shifts the focus from merely segmenting lesion boundaries to predicting their evolution. Our approach is inspired by recent advances in image synthesis achieved by diffusion probabilistic models [6, 9, 14, 15], which generate synthetic samples from a randomly sampled Gaussian distribution in a series of finite steps. We adapt this process to model the evolution of skin lesion boundaries as a parameterized chain process, starting from Gaussian noise and progressing through a series of denoising steps to yield a clear segmentation map with well-defined lesion boundaries. By predicting the next step in the chain process rather than the final segmentation map, our approach enables the more accurate segmentation of challenging lesions than previous models. We illustrate the process of boundary evolution in Fig. 1, where each row corresponds to a different step in the evolution process, culminating in a clear segmentation map with well-defined boundaries.

In this paper, we propose a **Medical Boundary Diffusion model (MB-Diff)** to improve the skin lesion segmentation, particularly in cases where the lesion boundaries are ambiguous and have extremely large or small sizes. The MB-Diff model follows the basic design of the plain diffusion model, using a sequential denoising process to generate the lesion mask. However, it also includes two key innovations: Firstly, we have developed an efficient multi-scale image guidance module, which uses a pretrained transformer encoder to extract multi-scale features from prior images. These features are then fused with the evolution features to constrain the direction of evolution. Secondly, we have implemented an evolution uncertainty-based fusion strategy, which takes into account the uncertainty of different initializations to refine the evolution results and obtain more precise lesion boundaries. We evaluate our model on two popular skin lesion segmentation datasets, ISIC-2016 and PH² datasets, and find that it performs significantly better than existing models.

2 Method

The key objective of MB-Diff is to improve the representation of ambiguous boundaries by learning boundary evolution through a cascaded series of steps, rather than a single step. In this section, we present the details of our cascaded boundary evolution learning process and the parameterized architecture of the evolution process. We also introduce our evolution-based uncertainty estimation and boundary ensemble techniques, which have significant potential for enhancing the precision and reliability of the evolved boundaries.

2.1 Boundary Evolution Process

We adopt a step-by-step denoising process to model boundary evolution in MB-Diff, drawing inspiration from recent diffusion probabilistic models (DPMs). Specifically, given the image and boundary mask distributions as $(\mathcal{X}, \mathcal{Y})$, assuming that the evolution consists of T steps in total, the boundary at T -th step (y_T) is the randomly initialized noise and the boundary at 0-th (y_0) step denotes the accurate result. We formulate the boundary evolution process as follows:

$$y_0 \sim p_\theta(y_0|x) := \int p_\theta(y_{0:T}|x) dy_{1:T} := p(y_T) \prod_{t=1}^T p_\theta(y_{t-1}|y_t, x), \quad (1)$$

where $p(y_T) = \mathcal{N}(y_T; 0, \mathbf{I})$ is the initialized Gaussian distribution and $p_\theta(y_{t-1}|y_t)$ is each learnable evolution step, formulated as the Gaussian transition, denoted as:

$$p_\theta(y_{t-1}|y_t) := \mathcal{N}(y_{t-1}; \mu_\theta(y_t, x, t), \Sigma_\theta(y_t, x, t)). \quad (2)$$

Note that the prediction function takes the input image as a condition, enabling the evolving boundary to fit the corresponding lesion accurately. By modeling boundary evolution as a step-by-step denoising process, MB-Diff can effectively capture the complex structures of skin lesions with ambiguous boundaries, leading to superior performance in lesion segmentation.

To optimize the model parameters θ , we use the evolution target as an approximation of the posterior at each evolution step. Given the segmentation label y as y_0 , the label is gradually added by a Gaussian noise as:

$$q(y_{1:T}|y_0) := \prod_{t=1}^T q(y_t|y_{t-1}) := \prod_{t=1}^T \mathcal{N}(y_t; \sqrt{1 - \beta_t}y_{t-1}, \beta_t\mathbf{I}), \quad (3)$$

where $\{\beta_t\}_{t=1}^T$ is a set of constants ranging from 0 to 1. After that, we compute the posterior $q(y_{t-1}|y_t, y_0)$ using Bayes' rule. The MSE loss function is utilized to measure the distance between the predicted mean and covariance of the Gaussian transition distribution and the evolution target $q(y_{t-1}|y_t, y_0)$.

2.2 Parameterized Architecture with Image Prior

The proposed model is a parameterized chain process that predicts the μ_{t-1}^* and Σ_{t-1}^* at each evolution step t under the prior conditions of the image x and the prior evolution y_t^* . To capture the deep semantics of these conditions and perform efficient fusion, we adopt a basic U-Net [16] architecture inspired by the plain DPM and introduce novel designs for condition fusion, that is the efficient multi-scale image guidance module.

The architecture consists of a multi-level convolutional encoder and a symmetric decoder with short connection layers between them. To incorporate the variable t into the model, we first embed it into the latent space. Then, the prior evolution y_t^* is added to the latent t before each convolution. At the bottleneck layer, we fuse the evolution features with the image guidance to constrain the evolution and ensure that the final boundary suits the conditional image.

To achieve this, priors train a segmentation model concurrently with the evolution model and use an attention-based parser to translate the image features

in the segmentation branch into the evolution branch [22]. Since the segmentation model is trained much faster than the evolution model, we adopt a pretrained pyramid vision transformer (PVT) [20] as the image feature extractor to obtain the multi-scale image features. Let $\{f_l\}_{l=1}^4$ denote the extracted features at four levels, with a 2x, 4x, 8x, 16x, smaller size of the original input. Each feature at the three lower levels is resized to match the scale of f_4 using Adaptive Averaging Pooling layers. After that, the four features are concatenated and fed into a full-connection layer to map the image feature space into the evolution space. We then perform a simple yet effective addition of the mapped image feature and the encoded prior evolution feature, similar to the fusion of time embeddings, to avoid redundant computation.

2.3 Evolution Uncertainty

Similar to typical evolutionary algorithms, the final results of boundary evolution are heavily influenced by the initialized population. As a stochastic chain process, the boundary evolution process may result in different endpoints due to the random Gaussian samples at each evolution step. This difference is particularly evident when dealing with larger ambiguity in boundary regions. The reason is that the image features in such ambiguous regions may not provide discriminative guidance for the evolution, resulting in significant variations in different evolution times. Instead of reducing the differences, we surprisingly find that these differences can represent segmentation uncertainty. Based on the evolution-based uncertainty estimation, the segmentation results become more accurate and trustworthy in practice [4, 5, 12].

Uncertainty Estimation: To estimate uncertainty, the model parameters θ are fixed, and the evolution starts with a randomly sampled Gaussian noise $y_T^* \sim \mathcal{N}(0, \mathbf{I})$. Let $\{y_T^{*,i}\}_{i=1}^n$ denote a total of n initializations. Once the evolution is complete, the obtained $\{\mu^{*,i}\}_{i=1}^n, \{\sum^{*,i}\}_{i=1}^n$ are used to the sample final lesion maps as: $y^{*,i} = \mu^{*,i} + \exp(\frac{1}{2} \sum^{*,i}) \mathcal{N}(0, \mathbf{I})$. Unlike traditional segmentation models that typically scale the prediction into the range of 0 to 1, the evolved maps generated by MB-Diff have unfixed distributions due to random sampling. Since the final result is primarily determined by the mean value μ , and the predicted \sum has a limited range [6], we calculate the uncertainty as:

$$\delta = \sqrt{\frac{1}{n} \sum_{i=1}^n (\mu^{*,i} - \frac{1}{n} \sum_{j=1}^n \mu^{*,j})^2}, \quad (4)$$

Evolution Ensemble: Instead of training multiple networks or parameters to make the ensemble, MB-Diff allows running the inference multiple times and fusing the obtained evolutions. However, simply averaging the predicted identities from multiple evolutions is not effective, as the used MSE loss without activation constrains the predicted identities to be around 0 or 1, unlike the Sigmoid function which would limit the identities to a range between 0 and 1. Therefore,

Table 1. Comparison of skin lesion segmentation with different approaches on the ISIC-2016 and PH² datasets. The averaged scores of both sets are presented respectively.

Method	ISIC-2016 [8]				PH ² [13]			
	IoU↑	Dice↑	ASSD↓	HD95↓	IoU↑	Dice↑	ASSD↓	HD95↓
U-Net++ [24]	81.84	88.93	15.01	44.83	81.26	88.99	15.97	46.66
CA-Net [7]	80.73	88.10	15.67	44.98	75.18	84.66	21.06	64.53
TransFuse [23]	86.19	92.03	10.04	30.33	82.32	89.75	15.00	39.98
TransUNet [3]	84.89	91.26	10.63	28.51	83.99	90.96	12.65	33.30
XBound-Former [18]	87.69	93.08	8.21	21.83	85.38	91.80	10.72	26.00
MedSegDiff [22]	83.39	89.85	12.38	31.23	82.21	89.73	13.53	36.59
MB-Diff (Ours)	88.87	93.78	7.19	18.90	87.12	92.85	9.16	22.95

we employ the max vote algorithm to obtain the final segmentation map. In this algorithm, each pixel is classified as a lesion only if its identity sum across all n evolutions is greater than a threshold value τ . Finally, the segmentation map is generated as $y^* = (\sum_{i=1}^n y^{*,i}) \geq \tau$.

3 Experiment

3.1 Datasets and Evaluation Metrics

Datasets: We use two publicly available skin lesion segmentation datasets from different institutions in our experiments: the ISIC-2016 dataset and the PH² dataset. The ISIC-2016 dataset [8] is provided by the International Skin Imaging Collaboration (ISIC) archive and consists of 900 samples in the public training set and 379 samples in the public validation set. As the annotation for its public test set is not currently available, we additionally collect the PH² dataset [13], which contains 200 labeled samples and is used to evaluate the generalization performance of our methods.

Evaluation Metrics: To comprehensively compare the segmentation results, particularly the boundary delineations, we employ four commonly used metrics to quantitatively evaluate the performance of our segmentation methods. These metrics include the *Dice* score, the *IoU* score, Average Symmetric Surface Distance (*ASSD*), and Hausdorff Distance of boundaries (95–th percentile; *HD95*). To ensure fair comparison, all labels and predictions are resized to (512×512) before computing these scores, following the approach of a previous study [18].

3.2 Implementation Details

For the diffusion model hyper-parameters, we use the default settings of the plain diffusion model, which can be found in the supplementary materials. Regarding

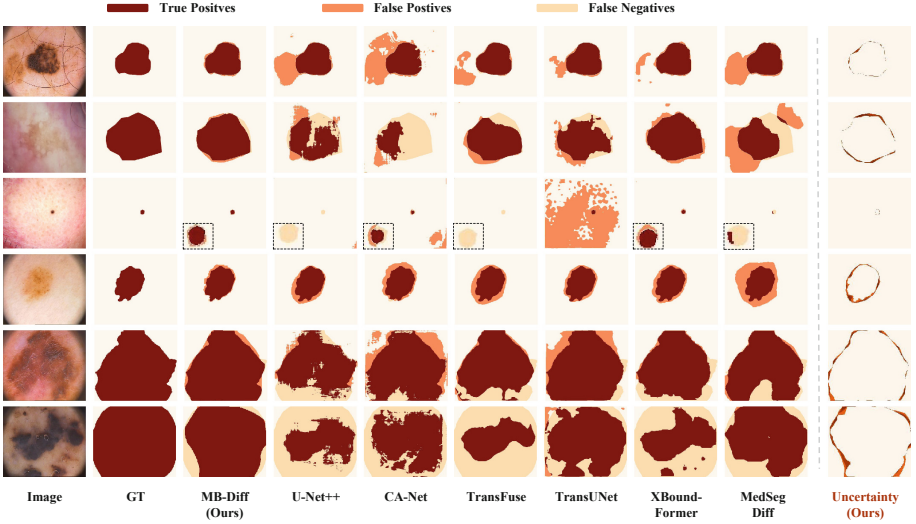


Fig. 2. Visual comparison of our method and the SOTAs. The first three rows are samples from the ISIC-2016 validation set and the last three rows are from the PH² dataset. We highlight the small lesions using dotted boxes in the third row.

the training parameters, we resize all images to (256×256) for efficient memory utilization and computation. We use a set of random augmentations, including vertical flipping, horizontal flipping, and random scale change (limited to $0.9 \sim 1.1$), to augment the training data. We set the batch size to 4 and train our model for a total of 200,000 iterations. During training, we use the AdamW optimizer with an initial learning rate of $1e-4$. For the inference, we set $n = 4$ and $\tau = 2$ considering the speeds.

3.3 Comparison with State-of-the-Arts

We majorly compare our method to the latest skin lesion segmentation models, including the CNN-based and transformer-based models, i.e., U-Net++ [24], CA-Net [7], TransFuse [23], TransUNet [3], and especially the boundary-enhanced method, X-BoundFormer [18]. Additionally, we evaluate our method against MedSegDiff [22], a recently released diffusion-based model, which we re-trained for 200,000 steps to ensure a fair comparison.

The quantitative results are shown in Table 1, which reports four evaluation scores for two datasets. Though the parameters of CNNs and Transformers are selected with the best performance on ISIC-2016 validation set and the parameters of our method are selected by completing the 200,000 iterations, MB-Diff still achieves the 1.18% IoU improvement and 0.7% Dice improvement. Additionally, our predicted boundaries are closer to the annotations, as evidenced by the ASSD and HD95 metrics, which reduce by 1.02 and 1.93 pixels, respectively. When compared to MedSegDiff, MB-Diff significantly outperforms it in

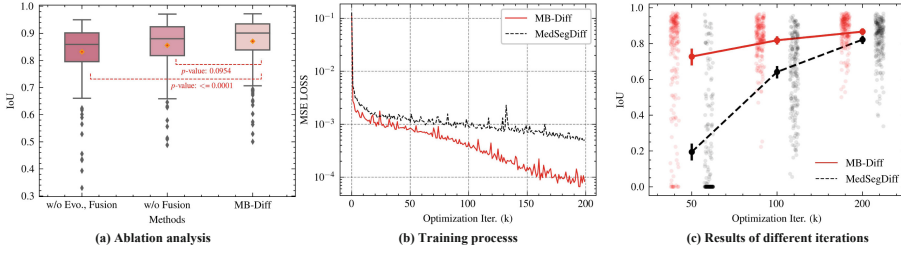


Fig. 3. Detailed analysis of our method, including the ablation analysis (a) and the comparison to the other diffusion-based method (b, c).

all metrics. Moreover, our method shows a larger improvement in generalization performance on the PH² dataset, indicating its better ability to handle new data.

We present a visual comparison of challenging samples in Fig. 2, including three samples from the ISIC-2016 validation set and three from the PH² dataset. These samples represent edge cases that are currently being studied in the community, including size variation, boundary ambiguity, and neighbor confusion. Our visual comparison reveals several key findings: (1) MB-Diff consistently achieves better segmentation performance on small and large lesions due to its thorough learning of boundary evolution, as seen in rows 3, 5, and 6. (2) MB-Diff is able to produce correct boundaries even in cases where they are nearly indistinguishable in human perception, eliminating the need for further manual adjustments and demonstrating significant practical value. (3) MB-Diff generates fewer false positive segmentation, resulting in cleaner predictions that enhance the user experience.

Furthermore, we provide a visualization of evolution uncertainties in Fig. 2, where deeper oranges indicate larger uncertainties. It is evident that most regions with high uncertainties correspond to false predictions. This information can be used to guide human refinement of the segmentation in practical applications, ultimately increasing the AI’s trustworthiness.

3.4 Detailed Analysis of the Evolution

In this subsection, we make a comprehensive analysis to investigate the performance of each component in our method and compare it to the diffusion-based model, MedSegDiff. The results of our ablation study are presented in Fig. 3(a), where “w/o Evo” refers to using image features to directly train a segmentation model with FPN [11] architecture and “w/o Fusion” means no evolution fusion is used. To ensure a fair comparison, we average the scores of multiple evolutions to represent the performance of “w/o Fusion”. The results demonstrate that our evolutionary approach can significantly improve performance, and the evolution uncertainty-based fusion strategy further enhances performance. Comparing our method to MedSegDiff, the training loss curve in Fig. 3(b) shows that our method converges faster and achieves smaller losses, indicating that our multi-scale image guidance is more effective than that of MedSegDiff. Furthermore, we evaluate our

method's performance using parameters saved at different iterations, as shown in Fig. 3(c). Our results demonstrate that our method has competitive performance at 50k iterations versus MedSegDiff at 200k iterations and our method at 100k iterations has already outperformed well-trained MedSegDiff.

4 Conclusion

In this paper, we introduced the medical boundary diffusion (MB-Diff) model, which is a novel approach to segment skin lesions. Our proposed method formulates lesion segmentation as a boundary evolution process with finite timesteps, which allows for efficient and accurate segmentation of skin lesions. To guide the boundary evolution towards the lesions, we introduce an efficient multi-scale image guidance module. Additionally, we propose an evolution uncertainty-based fusion strategy to yield more accurate segmentation. Our method is evaluated on two well-known skin lesion segmentation datasets, and the results demonstrate superior performance and generalization ability in unseen domains. Through a detailed analysis of our training program, we find that our model has faster convergence and better performance compared to other diffusion-based models. Overall, our proposed MB-Diff model offers a promising solution to accurately segment skin lesions, and has the potential to be applied in a clinical setting.

Acknowledgement. This work is supported by the National Key Research and Development Program of China (2019YFE0113900).

References

1. Azad, R., Asadi-Aghbolaghi, M., Fathy, M., Escalera, S.: Attention Deeplabv3+: multi-level context attention mechanism for skin lesion segmentation. In: Bartoli, A., Fusiello, A. (eds.) ECCV 2020. LNCS, vol. 12535, pp. 251–266. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-66415-2_16
2. Cao, W., et al.: ICL-Net: global and local inter-pixel correlations learning network for skin lesion segmentation. *IEEE J. Biomed. Health Inf.* **27**(1), 145–156 (2022)
3. Chen, J., et al.: TransUNet: Transformers make strong encoders for medical image segmentation. arXiv preprint [arXiv:2102.04306](https://arxiv.org/abs/2102.04306) (2021)
4. Czolbe, S., Arnavaz, K., Krause, O., Feragen, A.: Is segmentation uncertainty useful? In: Feragen, A., Sommer, S., Schnabel, J., Nielsen, M. (eds.) IPMI 2021. LNCS, vol. 12729, pp. 715–726. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78191-0_55
5. DeVries, T., Taylor, G.W.: Leveraging uncertainty estimates for predicting segmentation quality. arXiv preprint [arXiv:1807.00502](https://arxiv.org/abs/1807.00502) (2018)
6. Dhariwal, P., Nichol, A.: Diffusion models beat GANs on image synthesis. *Adv. Neural Inf. Process. Syst.* **34**, 8780–8794 (2021)
7. Gu, R., et al.: Ca-net: Comprehensive attention convolutional neural networks for explainable medical image segmentation. *IEEE Trans. Med. Imaging* **40**(2), 699–711 (2020)

8. Gutman, D., et al.: Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). arXiv preprint [arXiv:1605.01397](https://arxiv.org/abs/1605.01397) (2016)
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **33**, 6840–6851 (2020)
10. Li, H., et al.: Dense deconvolutional network for skin lesion segmentation. *IEEE J. Biomed. Health Inf.* **23**(2), 527–537 (2018)
11. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125 (2017)
12. Mehrtash, A., Wells, W.M., Tempany, C.M., Abolmaesumi, P., Kapur, T.: Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. *IEEE Trans. Med. Imaging* **39**(12), 3868–3878 (2020)
13. Mendonça, T., Ferreira, P.M., Marques, J.S., Marcal, A.R., Rozeira, J.: PH 2-A dermoscopic image database for research and benchmarking. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5437–5440. IEEE (2013)
14. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: *International Conference on Machine Learning*, pp. 8162–8171. PMLR (2021)
15. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695 (2022)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
17. Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A.: *Cancer statistics, 2022*. CA: Cancer J. Clin. **72**(1), 7–33 (2022)
18. Wang, J., et al.: XBound-former: toward cross-scale boundary modeling in transformers. *IEEE Trans. Med. Imaging* **42**(6), 1735–1745 (2023)
19. Wang, J., Wei, L., Wang, L., Zhou, Q., Zhu, L., Qin, J.: Boundary-aware transformers for skin lesion segmentation. In: de Bruijne, M., et al. (eds.) *MICCAI 2021*. LNCS, vol. 12901, pp. 206–216. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_20
20. Wang, W., et al.: Pyramid vision transformer: a versatile backbone for dense prediction without convolutions. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 568–578 (2021)
21. Wu, H., Pan, J., Li, Z., Wen, Z., Qin, J.: Automated skin lesion segmentation via an adaptive dual attention module. *IEEE Trans. Med. Imaging* **40**(1), 357–370 (2020)
22. Wu, J., Fang, H., Zhang, Y., Yang, Y., Xu, Y.: MedSegDiff: Medical image segmentation with diffusion probabilistic model. arXiv preprint [arXiv:2211.00611](https://arxiv.org/abs/2211.00611) (2022)
23. Zhang, Y., Liu, H., Hu, Q.: TransFuse: fusing transformers and CNNs for medical image segmentation. In: de Bruijne, M., et al. (eds.) *MICCAI 2021*. LNCS, vol. 12901, pp. 14–24. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_2
24. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: a nested u-net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) *DLMIA/ML-CDS -2018*. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1