

Toward Medical Deepfake Detection: A Comprehensive Dataset and Novel Method

Shuaibo Li¹, Zhaohu Xing¹, Hongqiu Wang¹, Pengfei Hao¹, Xingyu Li¹,
Zekai Liu¹, and Lei Zhu^{1,2} 

¹ The Hong Kong University of Science and Technology (Guangzhou)

² The Hong Kong University of Science and Technology
leizhu@ust.hk

Abstract. The rapid advancement of generative AI in medical imaging has introduced both significant opportunities and serious challenges, especially the risk that fake medical images could undermine healthcare systems. These synthetic images pose serious risks, such as diagnostic deception, financial fraud, and misinformation. However, research on medical forensics to counter these threats remains limited, and there is a critical lack of comprehensive datasets specifically tailored for this field. Additionally, existing media forensic methods, which are primarily designed for natural or facial images, are inadequate for capturing the distinct characteristics and subtle artifacts of AI-generated medical images. To tackle these challenges, we introduce **MedForensics**, a large-scale medical forensics dataset encompassing six medical modalities and twelve state-of-the-art medical generative models. We also propose **DSKI**, a novel **D**ual-**S**tage **K**nowledge **I**nfusing detector that constructs a vision-language feature space tailored for the detection of AI-generated medical images. DSKI comprises two core components: 1) a cross-domain fine-trace adapter (CDFA) for extracting subtle forgery clues from both spatial and noise domains during training, and 2) a medical forensic retrieval module (MFRM) that boosts detection accuracy through few-shot retrieval during testing. Experimental results demonstrate that DSKI significantly outperforms both existing methods and human experts, achieving superior accuracy across multiple medical modalities.

Keywords: Medical Forensics · Medical Image Analysis · Trustworthy AI in Medical Imaging · Healthcare Security · Dataset

1 Introduction

The rapid advancement of deep learning accelerated the integration of artificial intelligence (AI) in medicine. However, challenges such as the high cost of medical data collection, privacy regulations, and limited availability of annotated medical datasets have hindered progress. Recently, generative AI techniques, especially denoising diffusion models, have been utilized to generate realistic, diverse, and true-to-distribution medical imaging data, to augment healthcare model training [15,13,27]. These AI-generated datasets have been beneficial in

improving tasks like classification, segmentation, and cross-modal translation. Despite these advancements, the increasing quality and volume of synthetic medical images pose significant risks to healthcare systems. Prior studies [20,12] have demonstrated that even medical experts can be misled by the high visual realism of AI-generated images. Thus, developing reliable and effective methods to detect these AI-generated medical images is crucial for mitigating risks and ensuring patient safety.

While AI-generated image detection has been widely studied in natural and facial images [19,23,25,21,17,16], few studies have focused on medical deepfakes. Applying methods intended for natural or facial images to the medical domain is suboptimal due to the unique nature of medical images. AI-generated medical images aim to replicate physiological phenomena and anatomical structures [15,13,27,28,26]. Compared to natural or facial images, these forged medical images often contain more subtle and localized clues (e.g., irregular low-level textures, unrealistic anatomical and pixel statistics), which are less semantically meaningful. Additionally, the variety of modalities and structures in medical images makes detection even more difficult. Current medical forensic methods [1,2] are still in the early stage, primarily addressing images manipulated by traditional tools or GANs, but remain ineffective against the hyper-realistic images produced by advanced AI models like Diffusion Models. In this paper, we focus on detecting AI-generated medical images created by various state-of-the-art medical generative models.

A major challenge in medical forensics is the lack of large-scale datasets of AI-generated medical images, due to the diversity of medical modalities and the complexity of generative models. To address this, we introduce MedForensics, a large-scale dataset of high-quality medical images generated by twelve leading models across six modalities, providing a key benchmark for medical forensics. Additionally, we propose the Dual-Stage Knowledge Infusing Detector (DSKI), designed to distinguish AI-generated medical images. In the training stage, a cross-domain fine-tune adapter (CDFA) captures forensic clues in the spatial and noise domains, using an inception module to extract multi-scale artifacts and a constrained CNN to model low-level pixel statistics. In the testing stage, a Medical Forensic Retrieval Module (MFRM) enhances detection performance and scalability. Experimental results show that DSKI outperforms state-of-the-art methods and human experts in detecting medical deepfakes, offering a crucial solution to mitigate the risks posed by AI-generated images and safeguard healthcare systems.

2 The Proposed MedForensics Dataset

2.1 Dataset Details

To advance the development of medical forensic detectors and assess their ability to distinguish AI-generated from real medical images, we introduce MedForensics, a large-scale dataset comprising 116,000 medical images, with an equal

Table 1. Details of our proposed MedForensics dataset.

Model	Modality	Classification	Pairs Number	Real Image Source
U-KAN-U	Ultrasound	Breast	5000	BUSI, UNS
U-KAN-E	Endoscope	Colorectum	5000	Kvasir, CVC, CV3D, ETIS
U-KAN-H	Histopathology	Histopathology	5000	GlaS, LC
SegGuidedDiff-M	MRI	Breast	5000	DBCM
SegGuidedDiff-C	CT	Neck-to-pelvis	5000	CT-ORG
Cheff	X-ray	Chest	5000	MaChEX
EFN	Ultrasound	Heart	4000	CAMUS
MONAI-MGD	MRI	Brain	5000	BraTS
DiffEcho	Ultrasound	Heart	4000	CAMUS
ArSDM	Endoscope	Colorectum	5000	Kvasir, CVC, CV3D, ETIS
Hyper-GAN	MRI	Brain	5000	BraTS
ICVAE	MRI	Brain	5000	BraTS

number of real and AI-generated fake images. MedForensics spans six medical modalities: Ultrasound, Endoscopy, Histopathology, MRI, CT, and X-ray, covering a diverse range of real-world forensic scenarios. We employ 12 state-of-the-art (SOTA) models from nine medical image generation studies, including U-KAN [15], SegGuidedDiff [13], Cheff [27], EFN [24], MONAI-MGD [18], DiffEcho [4], ArSDM [9], Hyper-GAN [29], and ICVAE [8]. Each generator produces 5,000 synthetic images, except for DiffEcho and EFN, which produce 4,000 images. Real images are primarily sourced from the corresponding generative model’s training set to maintain a balanced distribution of real and fake images. In cases where training sets were small (e.g., U-KAN-H’s GlaS dataset with only 165 images), we supplement with images from larger datasets of the same modality [5,6,3,11,27,14]. The dataset is standardized at the 256×256 resolution, and the images are split into training and testing sets with an 80/20 split. With its large volume, diverse modalities, and inclusion of advanced SOTA models, MedForensics provides a comprehensive resource for developing and evaluating medical forensic detection methods. Table 1 details the dataset composition, while Figure 1 illustrates examples of the generated images.

2.2 Fake Medical Image Generators

Diffusion-based Model: Diffusion models have recently emerged as the most advanced architecture in medical image generation. Notably, U-KAN [15] integrates Kolmogorov-Arnold Networks (KANs) into the noise predictor of the diffusion model, with three variants that generate highly realistic ultrasound, endoscopy, and histopathology images, namely U-KAN-U, U-KAN-E, and U-KAN-H. SegGuidedDiff [13] enables anatomically controllable image generation by adhering to a multi-class anatomical segmentation mask during each sampling step. SegGuidedDiff-M and SegGuidedDiff-C generate high-quality breast MRI and abdominal/neck-to-pelvis CT images. Cheff [27] utilizes a cascaded latent diffusion model to generate state-of-the-art chest radiographs, while EFN [24] employs Denoising Diffusion Probabilistic Models (DDPM) guided by cardiac semantic labels to generate high-quality ultrasound images. MONAI-MGD, part

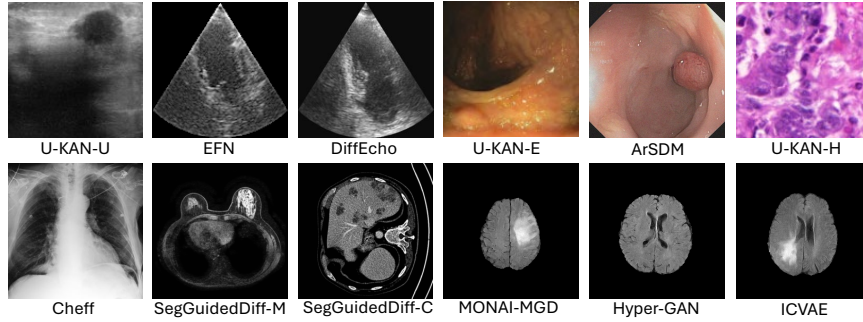


Fig. 1. Visualization of AI-generated medical images from the proposed MedForensics dataset, covering six imaging modalities and twelve generative models.

of the widely used MONAI [18] library, generates brain MRI images. DiffEcho [4] produces realistic echocardiography images, and ArSDM [9] applies an adaptive refinement semantic diffusion model to generate colonoscopy images.

GAN-based Model: Over the past decade, Generative Adversarial Networks (GANs) have led to substantial advancements in the quality of image generation. Hyper-GAN [29] constructs a multi-contrast MRI image translation model that adapts to various MR contrast types, enabling high-fidelity brain MRI generation.

VAE-based Model: Variational Autoencoders (VAEs) learn a latent space to encode and reconstruct data. ICVAE [8] separates the input data’s embedding space from conditioning variables, ensuring that generated brain MRI features are independent of conditioning factors, thus increasing output diversity.

3 Method

3.1 Overview of DSKI

As shown in Figure 2, DSKI is built upon the pre-trained CLIP [22], which has been trained on an extensive dataset of 400 million image-text pairs, providing a robust vision-language feature space suitable for a variety of downstream tasks. CLIP has also demonstrated sufficient effectiveness in media forensics tasks [19], which makes it an ideal foundation for our approach. While the original CLIP space performs well in distinguishing the authenticity of natural images, its high-level visual semantics are insufficient for medical image forensics due to the lack of fine-grained details and medical-specific information. To address this, DSKI performs a two-stage medical knowledge infusing to adapt the original CLIP feature space into an adequate, generalized medical forgery discrimination space. The training stage is centered around the cross-domain fine-trace adapter (Section 3.2), while the testing phase utilizes the forensic retrieval module (Section 3.3) to enhance detection accuracy further.

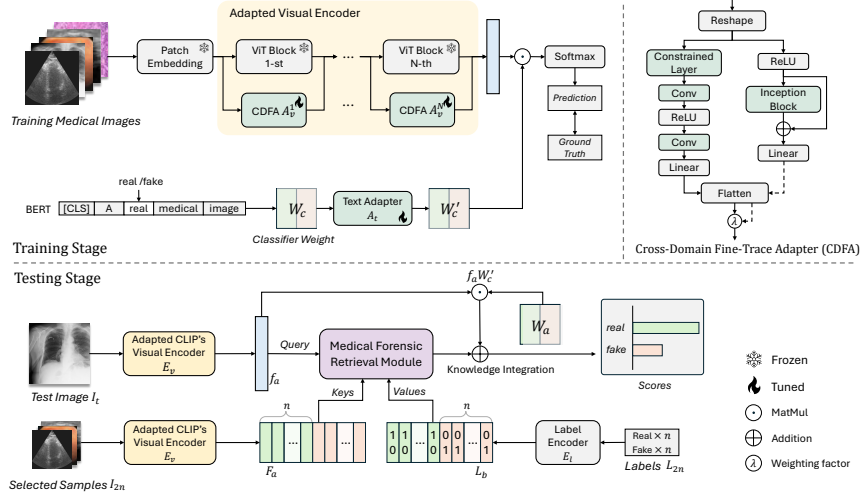


Fig. 2. Overview of the DSKI framework. The training stage employs a Cross-Domain Fine-Trace Adapter (CDFA) to inject medical forensic knowledge into the CLIP backbone via spatial and noise feature streams. During testing, a Medical Forensic Retrieval Module (MFRM) retrieves few-shot knowledge from a feature bank to enhance detection performance.

3.2 Training Stage: Cross-domain Fine-trace Adapter (CDFA)

Our training set includes real and fake medical images, each paired with natural language prompts (e.g., "A real/fake medical image"). The vanilla CLIP employs the text encoder to convert the prompts P_c into classifier weights W_c , and the visual encoder extracts image features f . We introduce visual adapters $A_v^{(i)}(\cdot)$ and a text adapter $A_t(\cdot)$ to fine-tune the pre-trained CLIP. Each visual adapter (e.g., CDFA) is placed in parallel with the MLP in the i -th transformer block. The CDFA has two streams that capture fine-grained forgery traces from both the spatial and noise domains. In the noise domain, constrained convolution layers and ReLU layer help learn abnormal pixel relationships while suppressing semantic content. In the spatial domain, an inception module with parallel convolutions (1×1 , 3×3 , 5×5) captures multi-scale forensic artifacts, which are crucial for interpreting medical images [16]. The two streams are defined as follows:

$$\hat{f}_s^{(i)} = \text{Conv} \left(\text{ReLU} \left(\text{ConstrainedConv} \left(f^{(i)} \right) \right) \right), \quad (1)$$

$$\hat{f}_n^{(i)} = \text{InceptionConv} \left(\text{ReLU} \left(f^{(i)} \right) \right), \quad (2)$$

where $f^{(i)}$ is the input to the visual adapter in the i -th transformer block, and $\hat{f}_s^{(i)}$, $\hat{f}_n^{(i)}$ are the outputs from the spatial and noise streams. The outputs of both streams are fused using a learnable scale factor λ , forming the final medical forensic feature:

$$\hat{f}^{(i)} = \hat{f}_s^{(i)} + \lambda \hat{f}_n^{(i)}. \quad (3)$$

The adapter output $\hat{f}^{(i)}$ is added to the corresponding MLP. The text adapter $A_t(\cdot)$ consists of two sequential linear layers applied after the text encoder. At this point, we obtain the new image feature f' and classifier weights W' . During training, the weights of $A_v(\cdot)$ and $A_t(\cdot)$ are optimized using the binary cross-entropy loss:

$$\mathcal{L}(\theta) = -\frac{1}{N} \sum_{n=1}^N [y_n \log(p_n) + (1 - y_n) \log(1 - p_n)], \quad (4)$$

where N is the total number of training samples, y_n is the ground truth label for the n -th sample (0 for real, 1 for fake), and p_n is the predicted probability of the sample being classified as the positive class.

3.3 Testing Stage: Medical Forensic Retrieval Module (MFRM)

After the training stage, DSKI collects cross-domain, multi-scale artifacts to form a comprehensive fine-grained view for medical image forensics. To enhance detection robustness and accuracy, we design MFRM, inspired by [30]. The MFRM is built upon a cache feature bank. Using the fully adapted CLIP model from the training stage, we randomly select a small sample set I_{2n} from the completed dataset, containing n real and n synthetic medical images with corresponding labels L_{2n} . The adapted visual encoder E_v extracts the L2-normalized features $F_a \in \mathbb{R}^{2n \times C}$. The ground-truth labels L_{2n} are converted into a binary vector $L_b \in \mathbb{R}^{2n \times 2}$ with a one-hot encoder E_l . F_a and L_b serve as keys and values, forming the feature bank, which stores the new medical forensic knowledge extracted from the few-shot samples. Given a test image I_t , its L2-normalized feature $f_a \in \mathbb{R}^{1 \times C}$ is extracted by the adapted CLIP visual encoder and is used as a query in the MFRM to retrieve relevant information from the feature bank. The affinity scores $A_s \in \mathbb{R}^{1 \times 2n}$ between the query and the keys are computed as:

$$A_s = \exp(-\alpha(1 - f_a F_a^T)), \quad (5)$$

where α is a modulation hyperparameter. The prediction of the MFRM can be obtained with $A_s L_b$. Then, the knowledge retrieved from the MFRM is combined with prior knowledge ($f_a W_a^T \in \mathbb{R}^{1 \times 2}$, where W_a is the adapted classifier weights) from the adapted CLIP model to yield the final logits:

$$p_f = \beta A_s L_b + f_a W_a^T, \quad (6)$$

where β is the residual ratio. This retrieval-based MFRM not only enhances synthetic medical image detection by integrating new knowledge without re-training, but also provides scalability for our model. As new medical image synthesis frameworks emerge, DSKI can seamlessly integrate few-shot, newly labeled samples into the feature bank during testing, boosting its ability to detect fake images from new frameworks.

Table 2. Cross-modal detection Acc/AP of compared methods on the MedForensics dataset. The bold value indicates the best performance.

Methods	Ultrasound	Endoscope	Histopathology	MRI	CT	X-ray	Mean
UniFD [19]	85.7/78.8	80.8/86.7	84.2/78.2	61.6/54.5	54.6/66.1	63.6/67.4	71.8/72.0
AEROBLAD [23]	73.1/65.8	68.7/73.9	55.6/51.4	64.8/81.6	56.3/56.8	57.9/56.6	62.7/64.4
NPR [25]	71.8/76.5	56.2/58.4	75.4/74.7	51.5/56.1	74.7/73.3	52.2/63.7	63.6/67.1
F3-Net [21]	72.3/77.2	70.8/71.5	59.8/64.3	55.9/57.8	65.6/72.3	53.5/58.0	63.0/66.9
DFH [2]	72.8/73.5	62.4/68.3	75.2/72.4	59.3/58.3	67.2/70.3	64.8/64.7	67.0/67.9
RECCE [7]	59.8/62.3	54.7/61.0	64.6/69.7	69.0/69.5	51.4/62.8	63.3/85.1	60.5/68.4
MedNet [1]	63.2/60.7	58.3/54.5	64.3/62.3	70.9/68.7	69.7/69.4	64.4/66.7	65.1/63.7
Ours	98.9/93.4	91.9/91.4	97.8/93.8	84.9/89.1	92.7/93.6	83.4/91.6	91.6/92.2

4 Experiments

4.1 Experiments Setup

We evaluated the performance of our proposed method on the MedForensics (see Section 2). The compared methods fall into three categories: state-of-the-art natural image forensics (UniFD [19], AEROBLADE [23], NPR [25]), facial image forensics (F3-Net [21], RECCE [7]), and medical image forensics (MedNet [1], DFH [2]). All compared methods were trained on the MedForensics training set for fair comparison. We evaluated all models on the MedForensics dataset using Accuracy (Acc) and Average Precision (AP), following standard media forensics protocols.

Implementation Details. We initialize the backbone (ViT-L/14) using weights provided by [22], and the remaining parts are randomly initialized. We train the model for 50 epochs with a batch size of 32 on 4 NVIDIA GTX3090 GPUs. We adopt SGD as the optimizer, with a momentum of 0.9, a weight decay of 0.005, and an initial learning rate of 1e-4. Meanwhile, the visual adapters are placed in the ViT blocks at $i = 7, 15, 23$. α and β are set to 0.1 and 10 for blending adjustment. In testing, the number of selected samples n is empirically set to 16.

4.2 Comparison Results

Quantitative Results Comparison. Table 2 summarizes the detection performance of different methods across various modalities. Thanks to the dual-stage aggregation of medical forensic knowledge, DSKI significantly outperforms other methods in all modalities and metrics. Specifically, DSKI shows an improvement of over 20% in Acc and AP compared to UniFD [19] using the original CLIP feature space. Unlike methods [1,2], which suffer degraded performance in detecting fake images (with relatively poor quality) from outdated GANs, DSKI handles a broader range of architectures and outperforms these methods by a significant margin.

Turing Test (Blinded Expert Evaluation). To evaluate the real-world applicability of our method, we conducted a Turing test with three medical specialists—a radiologist, a pathologist, and a gastroenterologist. We randomly selected

Table 3. Blinded expert evaluation: human vs. DSKI performance.

Detectors	Ultrasound, MR, CT, X-ray	Histopathology	Endoscope
Radiologist	72.3/78.7	-	-
Pathologist	-	74.1/78.9	-
Gastroenterologist	-	-	68.5/72.0
DSKI	89.8/90.1	95.8/93.7	92.0/90.7

50 synthetic and 50 real images from each modality in the MedForensics dataset and asked the experts to assess their authenticity. Table 3 shows that the experts had difficulty distinguishing real from fake images due to the similarity in pathological features. In contrast, our DSKI effectively identified the fake images by capturing low-level medical forgery artifacts.

4.3 Ablation Study

We conducted ablation experiments on the core components (Table 4). Integrating both modules significantly enhances performance. Each component plays a distinct role: CDFA captures fine-grained artifacts during training, while MFRM utilizes a retrieval mechanism to inject medical forensic knowledge from few-shot samples during testing. Table 5 summarizes the evaluation of different CDFA feature streams, with the full configuration achieving optimal results. We also tested the DSKI’s scalability on 1,000 fake UWF fundus images generated by an unseen framework [10] and 1000 real images. DSKI demonstrated strong performance (Acc = 86.6%, AP = 89.4%) even without adding new samples. Adding a few extra samples to the feature bank improved Acc by 2.1% and AP by 2.7%, showcasing its scalability to emerging threats.

Table 4. Ablation study for core components (CDFA and MFRM) in DSKI. **Table 5.** Ablation study for feature streams (spatial and noise) in CDFA.

CDFA	MFRM	Acc	AP	Spatial	Noise	Acc	AP
×	×	70.9	72.6	×	×	75.3	78.9
✓	×	86.0	89.8	✓	×	80.3	82.4
×	✓	75.3	78.9	×	✓	82.6	83.3
✓	✓	91.6	92.2	✓	✓	91.6	92.2

5 Conclusion

This paper tackles the growing threat of AI-generated medical deepfakes with two key contributions. First, we introduce MedForensics, a high-quality, large-scale dataset for medical forensics, covering six modalities and generated by

twelve state-of-the-art models. Second, we propose the Dual-Stage Knowledge Infusing (DSKI) detector, a two-stage method that enhances synthetic medical image detection. Experimental results show that DSKI outperforms both state-of-the-art methods and human experts across multiple modalities, offering a more robust solution for detecting medical deepfakes. This work facilitates the development of trustworthy healthcare systems by providing a comprehensive dataset and an effective detection framework.

Acknowledgments. This work is supported by the Guangdong Science and Technology Department (No. 2024ZDZX2004).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Albahli, S., Nawaz, M.: Mednet: Medical deepfakes detection using an improved deep learning approach. *Multimedia Tools and Applications* **83**(16), 48357–48375 (2024)
2. Alsaheel, A., Alhassoun, R., Alrashed, R., Almatrafi, N., Almallouhi, N., Albahli, S.: Deep fakes in healthcare: How deep learning can help to detect forgeries. *Computers, Materials & Continua* **76**(2) (2023)
3. Anna Montoya, Hasnin, k.s.W.C.y.: Ultrasound nerve segmentation (2016), <https://kaggle.com/competitions/ultrasound-nerve-segmentation>
4. Ashrafian, P., Yazdani, M., Heidari, M., Shahriari, D., Hacıhaliloglu, I.: Vision-language synthetic data enhances echocardiography downstream tasks (2024)
5. Bobrow, T.L., Golhar, M., Vijayan, R., Akshintala, V.S., Garcia, J.R., Durr, N.J.: Colonoscopy 3d video dataset with paired depth from 2d-3d registration. *Medical Image Analysis* p. 102956 (2023)
6. Borkowski, A.A., Bui, M.M., Thomas, L.B., Wilson, C.P., DeLand, L.A., Mastorides, S.M.: Lung and colon cancer histopathological image dataset (lc25000). *arXiv preprint arXiv:1912.12142* (2019)
7. Cao, J., Ma, C., Yao, T., Chen, S., Ding, S., Yang, X.: End-to-end reconstruction-classification learning for face forgery detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 4113–4122 (2022)
8. Delgado, J.M.D., Oyedele, L.: Deep learning with small datasets: using autoencoders to address limited datasets in construction management. *Applied Soft Computing* **112**, 107836 (2021)
9. Du, Y., Jiang, Y., Tan, S., Wu, X., Dou, Q., Li, Z., Li, G., Wan, X.: Arsdm: colonoscopy images synthesis with adaptive refinement semantic diffusion models. In: *International conference on medical image computing and computer-assisted intervention*. pp. 339–349. Springer (2023)
10. Fang, Z., Yu, X., Zhou, G., Zhuang, K., Chen, Y., Ge, R., Wang, C., Jia, G., Wu, Q., Ye, J., et al.: Lpuwf-ldm: Enhanced latent diffusion model for precise late-phase uwf-fa generation on limited dataset. *Expert Systems with Applications* **270**, 126471 (2025)
11. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., de Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: *MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II* 26. pp. 451–462. Springer (2020)

12. Ju, Z., Zhou, W.: Vm-ddpm: Vision mamba diffusion for medical image synthesis. arXiv preprint arXiv:2405.05667 (2024)
13. Konz, N., Chen, Y., Dong, H., Mazurowski, M.A.: Anatomically-controllable medical image generation with segmentation-guided diffusion models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 88–98. Springer (2024)
14. Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.M., Grenier, T., et al.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. *IEEE transactions on medical imaging* **38**(9), 2198–2210 (2019)
15. Li, C., Liu, X., Li, W., Wang, C., Liu, H., Liu, Y., Chen, Z., Yuan, Y.: U-kan makes strong backbone for medical image segmentation and generation. arXiv preprint arXiv:2406.02918 (2024)
16. Li, S., Ma, W., Guo, J., Xu, S., Li, B., Zhang, X.: Unionformer: Unified-learning transformer with multi-view representation for image manipulation detection and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12523–12533 (2024)
17. Li, S., Xu, S., Ma, W., Zong, Q.: Image manipulation localization using attentional cross-domain cnn features. *IEEE Transactions on Neural Networks and Learning Systems* **34**(9), 5614–5628 (2021)
18. MONAI: Monai model zoo, <https://monai.io/model-zoo.html>
19. Ojha, U., Li, Y., Lee, Y.J.: Towards universal fake image detectors that generalize across generative models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 24480–24489 (2023)
20. Prezja, F., Paloneva, J., Pölönen, I., Niinimäki, E., Äyrämö, S.: Deepfake knee osteoarthritis x-rays from generative adversarial neural networks deceive medical experts and offer augmentation potential to automatic classification. *Scientific Reports* **12**(1), 18573 (2022)
21. Qian, Y., Yin, G., Sheng, L., Chen, Z., Shao, J.: Thinking in frequency: Face forgery detection by mining frequency-aware clues. In: European conference on computer vision. pp. 86–103. Springer (2020)
22. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PmLR (2021)
23. Ricker, J., Lukovnikov, D., Fischer, A.: Aeroblade: Training-free detection of latent diffusion images using autoencoder reconstruction error. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9130–9140 (2024)
24. Stojanovski, D., Hermida, U., Lamata, P., Beqiri, A., Gomez, A.: Echo from noise: synthetic ultrasound image generation using diffusion models for real image segmentation. In: International Workshop on Advances in Simplifying Medical Ultrasound. pp. 34–43. Springer (2023)
25. Tan, C., Zhao, Y., Wei, S., Gu, G., Liu, P., Wei, Y.: Rethinking the up-sampling operations in cnn-based generative network for generalizable deepfake detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 28130–28139 (2024)
26. Wang, H., Yang, G., Zhang, S., Qin, J., Guo, Y., Xu, B., Jin, Y., Zhu, L.: Video-instrument synergistic network for referring video instrument segmentation in robotic surgery. *IEEE Transactions on Medical Imaging* (2024)

27. Weber, T., Ingrisch, M., Bischl, B., Rügamer, D.: Cascaded latent diffusion models for high-resolution chest x-ray synthesis. In: Pacific-Asia conference on knowledge discovery and data mining. pp. 180–191. Springer (2023)
28. Xing, Z., Ye, T., Yang, Y., Liu, G., Zhu, L.: Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 578–588. Springer (2024)
29. Yang, H., Sun, J., Yang, L., Xu, Z.: A unified hyper-gan model for unpaired multi-contrast mr image translation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24. pp. 127–137. Springer (2021)
30. Zhang, R., Zhang, W., Fang, R., Gao, P., Li, K., Dai, J., Qiao, Y., Li, H.: Tip-adapter: Training-free adaption of clip for few-shot classification. In: European conference on computer vision. pp. 493–510. Springer (2022)