Medical Image Synthesis via Fine-Grained Image-Text Alignment and Anatomy-Pathology Prompting

Wenting Chen[1], Pengyu Wang[2], Hui Ren[3], Lichao Sun[4], Quanzheng Li[3], Yixuan Yuan[2*], and Xiang Li[3⋆]

[1]City University of Hong Kong [2]The Chinese University of Hong Kong
[3]Massachusetts General Hospital and Harvard Medical School
[4]Lehigh University

Abstract. Data scarcity and privacy concerns limit the availability of high-quality medical images for public use, which can be mitigated through medical image synthesis. However, current medical image synthesis methods often struggle to accurately capture the complexity of detailed anatomical structures and pathological conditions. To address these challenges, we propose a novel medical image synthesis model that leverages fine-grained image-text alignment and anatomy-pathology prompts to generate highly detailed and accurate synthetic medical images. Our method integrates advanced natural language processing techniques with image generative modeling, enabling precise alignment between descriptive text prompts and the synthesized images' anatomical and pathological details. The proposed approach consists of two key components: an anatomy-pathology prompting module and a fine-grained alignment-based synthesis module. The anatomy-pathology prompting module automatically generates descriptive prompts for high-quality medical images. To further synthesize high-quality medical images from the generated prompts, the fine-grained alignment-based synthesis module pre-defines a visual codebook for the radiology dataset and performs fine-grained alignment between the codebook and generated prompts to obtain key patches as visual clues, facilitating accurate image synthesis. We validate the superiority of our method through experiments on public chest X-ray datasets and demonstrate that our synthetic images preserve accurate semantic information, making them valuable for various medical applications.

1 Introduction

In the medical field, high-quality medical images are scarce and difficult to access due to data privacy concerns and the labor-intensive process of collecting such data [10]. This scarcity of medical images can hinder the development and training of artificial intelligence (AI) models for various medical applications, such

[⋆] Corresponding authors: Yixuan Yuan (yxyuan@ee.cuhk.edu.hk), Xiang Li (xli60@mgh.harvard.edu) as diagnosis [26], segmentation [19,21,25,7,6,14,20], report generation [3], image synthesis [5,8,4], detection [27], and abnormality classification. One solution to overcome this challenge is to use medical image synthesis techniques to generate synthetic data that can replace or supplement real medical images.

Several chest X-ray generation methods have been investigated to mitigate these issues, which can be categorized into three main groups: generative adversarial networks (GAN) based [22,28,16], diffusion based [2,1], and transformer based [17,18] methods. Madani *et al.* [22] and Zhang *et al.* [28] utilize unconditional GANs to synthesize medical images as a form of data augmentation to improve segmentation and abnormality classification performance. To leverage medical reports, some diffusion-based methods [2,1] take the impression section of medical reports and random Gaussian noise as input for chest X-ray generation, ignoring the finding section that includes more detailed descriptions. To consider more details in medical reports, several transformer-based methods [17,18] take both finding and impression sections of medical reports as input to synthesize chest X-rays. However, current methods generate medical images based on the given ground-truth report from the dataset, which may not fully describe all the details of the medical image. In fact, medical images contain different anatomical structures (lobe, heart, and mediastinal) and pathological conditions (opacity, effusion, and consolidation), which are important for clinical diagnosis. As a result, the generated medical images often lack this detailed information. Thus, there is a need for a medical image synthesis method that can generate high-quality medical images with detailed anatomical and pathological descriptions.

Another significant challenge for current medical image synthesis methods is the substantial inter-modal gap between medical images and reports. Medical images, comprising thousands of pixels, visualize rich textures and colors, while medical reports consist of only a few sentences to summarize the findings and impressions of the medical images. This disparity leads to a great imbalance in the amount of information contained in each modality, resulting in a large inter-modal gap between medical reports and images [12]. As a result, the generated medical images may not accurately reflect the content of the corresponding medical reports, as the synthesis models struggle to bridge this information gap. Furthermore, the limited information provided in the medical reports may not be sufficient to guide the synthesis of highly detailed and accurate medical images, which are crucial for clinical diagnosis and decision-making. Thus, it is necessary to develop techniques that can effectively mitigate the information imbalance and minimize the inter-modal gap between medical reports and images. By doing so, the synthesized medical images can better capture the detailed anatomical structures and pathological conditions described in the medical reports, leading to more reliable and informative synthetic data for various medical applications.

To address these issues, we propose a novel medical image synthesis model that leverages the capabilities of fine-grained image-text alignment and anatomy-pathology prompts to generate highly detailed and accurate synthetic medical

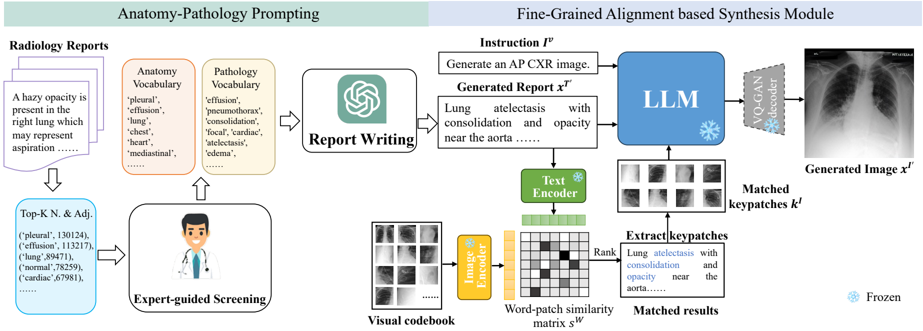

Anatomy-Pathology Prompting
Fine-Grained Alignment based Synthesis Module
Radiology Reports
A hazy opacity is present in the right lung which may represent aspiration ……
Top-K N. & Adj.
('pleural', 130124), ('effusion', 113217), ('lung',89471), ('normal',78259), ('cardiac',67981), ……
Expert-guided Screening
Anatomy Vocabulary
'pleural', 'effusion', 'lung', 'chest', 'heart', 'mediastinal', ……
Pathology Vocabulary
'effusion', 'pneumothorax', 'consolidation', 'focal', 'cardiac', 'atelectasis', 'edema', ……
Report Writing
Instruction I^v
Generate an AP CXR image.
Generated Report $x^{T'}$
Lung atelectasis with consolidation and opacity near the aorta ……
LLM
VQ-GAN decoder
Generated Image $x^{I'}$
Text Encoder
Matched keypatches k^I
Extract keypatches
Visual codebook
Image Encoder
Word-patch similarity matrix s^W
Rank
Lung atelectasis with consolidation and opacity near the aorta……
Matched results
Frozen

Fig. 1. The overview of the proposed method. It consists of an anatomy-pathology prompting module to generate descriptive reports with given anatomy and pathology words, and a fine-grained alignment based synthesis module using fine-grained image-text alignment to facilitate image generation.

images. Our approach consists of two key components: an **anatomy-pathology prompting** and a **fine-grained alignment based synthesis module**. The **anatomy-pathology prompting** aims to automatically generate descriptive reports for high-quality medical images. It first constructs the anatomy and pathology vocabularies from radiology reports under the guidance of radiologists, and then employs GPT-4 to write reports based on the given vocabularies. This ensures that the generated reports contain comprehensive and accurate descriptions of the anatomical structures and pathological conditions present in the medical images. To further synthesize high-quality medical images from the generated reports, we introduce a **fine-grained alignment based synthesis module**. This module pre-defines a visual codebook containing multiple patches commonly observed in the radiology dataset and performs fine-grained alignment between the generated reports and the visual codebook. Through this alignment, the module extracts the most matched keypatches that provide visual clues for the large language model (LLM) during the synthesis process. The LLM takes the generated reports, keypatches, and instructions as input and outputs visual tokens, which are then decoded by a VQ-GAN decoder to produce the final synthetic medical images. We conduct extensive experiments on publicly available chest X-ray (CXR) datasets to validate the superiority of our method compared to existing approaches. Furthermore, we perform semantic analysis on both real and synthetic images to demonstrate that our synthetic images preserve accurate semantic information, including anatomical structures and pathological conditions, making them valuable for various medical applications.

2 Method

2.1 Anatomy-Pathology Prompting

Since current methods struggle to synthesize medical images with complex anatomical structures (lobe, heart, and mediastinal) and pathological conditions (opacity, effusion, and consolidation), we introduce an anatomy-pathology prompting to automatically generate descriptive reports for high-quality medical image generation. This prompting module contains two main steps, including the design of anatomy and pathology vocabularies and prompts generation.

Designing Anatomy and Pathology Vocabularies. As illustrated in Fig. 1, we have developed anatomy and pathology vocabularies to extract instance-level anatomical and pathological terms from radiological reports and images. Recognizing that anatomical and pathological terms are typically nouns and adjectives, we employ a word filter to extract all nouns and adjectives from the impression and findings sections of reports in the MIMIC-CXR dataset [15]. We then select the top-K nouns and adjectives based on their occurrence frequencies. Finally, under expert guidance, we manually remove any remaining non-medical nouns and adjectives that GPT-4 is unable to filter out, and categorize the screened words into anatomy and pathology vocabularies according to their medical attributes. The number of words in anatomy and pathology vocabularies is 75 and 44, respectively. We demonstrate the word frequency of the anatomy and pathology vocabularies, as shown in Fig. 2.

Prompts Generation. With the anatomy and pathology vocabularies, we employ GPT4 to automatically generate the medical reports. Specifically, we first provide the vocabularies to GPT4 and require it to randomly select N and M words from anatomy and pathology vocabularies, respectively, which can be combined as the findings. Then, these words are passed to GPT4 to write a report with reasonable findings for a chest X-ray image. To let GPT4 write reports as our requirement, we use the following instructions.

```
anatomy_list = [‘pleural’, ‘lung’, ......,‘neck’, ‘junction’]
pathology_list = [‘effusion’, ‘pneumothorax’, ......, ‘diffuse’, ‘streaky’]
Here are two lists of anatomy and pathology for chest X-rays. Please write some findings
that only include 2 words from the anatomy list and 2 from the pathology list, and
do not write any negative sentences in the findings. These four words can be randomly
selected from the two lists, respectively. Please ensure the findings are reasonable for
a chest x-ray in real medical scenarios. The output should be in 50 words. Here is an
example:
anatomy_list = [‘heart’, ‘diaphragm’]
pathology_list = [‘effusion’, ‘opacity’]
Findings: Presence of opacity observed near the heart and diaphragm regions suggestive
of effusion.
Please generate the output in the following format:
anatomy_list = [‘word1’, ‘word2’]
pathology_list = [‘word3’, ‘word4’]
Findings:
```

This instruction example requires GPT4 to use two words from anatomy and pathology vocabularies, respectively. Actually, we can use more than two words and set N and M for the number of words we used in anatomy and pathology vocabularies, respectively. Then, we collect the anatomy-pathology

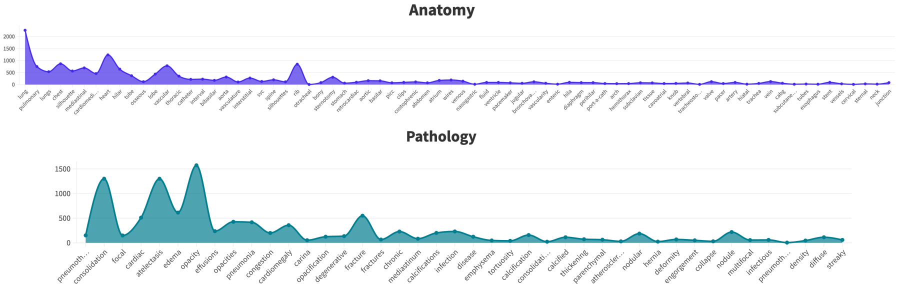


Fig. 2. The word frequency of the anatomy and pathology vocabularies.

prompts generated by GPT4, where each prompt contains an anatomy word list (e.g. `[‘heart’, ‘diaphragm’]`), a pathology word list (e.g. `[‘effusion’, ‘opacity’]`), and a generated report (e.g. `Presence of opacity observed near the heart and diaphragm regions suggestive of effusion.`). With these generated anatomy-pathology prompts, we can provide the synthesis model descriptive reports with detailed anatomical structures and pathological conditions.

2.2 Fine-Grained Alignment based Synthesis Module

Since there is an information imbalance and the inter-modal gap between medical reports and images, we devise a fine-grained alignment based synthesis module to leverage the fine-grained image-text alignment to facilitate image generation. The fine-grained alignment between medical reports and visual codebook to obtain matched keypatches as a clue for image synthesis. This module includes three steps for medical image synthesis, i.e. visual codebook construction, keypatches extraction, and image synthesis.

Visual Codebook Construction. To construct a visual codebook, we first identify the most common patches in the training set images and designate them as keypatches. This process involves matching patches from CXR images with textual tokens from their corresponding medical reports. We select the top κ_1 CXR-report pairs that exhibit the highest report-to-CXR similarities, denoted as s^T. For each selected CXR-report pair, we calculate the maximum similarity between each textual token and the image patches, resulting in word-patch maximum similarity scores. The embeddings of textual tokens and image patches are extracted by the pre-trained text and encoders [3], respectively. These scores are then ranked, and the patches corresponding to the top κ_2 similarities are extracted and included in the visual codebook as keypatches. Each keypatch in the codebook consists of the patch itself and its associated features.

Keypatches Extraction. With the visual codebook, we establish a correspondence between the features of keypatches and the textual tokens of the generated report. This is achieved by matching the features of each keypatch in the visual codebook with the textual tokens, resulting in the creation of a word-patch similarity matrix, denoted as $s^W \in \mathbb{R}^{(\kappa_1 \times \kappa_2) \times K}$, where K represents the total number of textual tokens in the report. To identify the keypatches that are most relevant to the generated report, we perform a ranking operation on the word-patch similarity matrix along the dimension of keypatches. For each textual token, we select the top κ_3 keypatches with the highest word-patch similarity scores. Finally, we extract the features of these selected keypatches, denoted as k^I, which serve as a compact representation of the visual information most closely associated with the textual content of the generated report.

Image Synthesis. After acquiring the keypatches, we employ a frozen VQ-GAN encoder [11] E to transform the matched keypatches k^I into image tokens $E(k^I)$. These image tokens are then fed into a pre-trained large language model (LLM)[3] along with the instruction and the generated report. The input to the LLM follows an instruction-following format. By providing the LLM with the instruction, generated report, and image tokens of the keypatches, we enable the model to predict image tokens that correspond to the desired CXR image. Finally, the predicted image tokens are decoded using the VQ-GAN decoder, resulting in the generation of the CXR image $x^{I'}$. This process leverages the power of the pre-trained LLM to interpret the textual instruction and report, while utilizing the visual information encoded in the keypatches to guide the generation of a realistic and coherent CXR image.

By adopting the fine-grained alignment based synthesis module, we can generate high-quality medical images with the detailed anatomical structures and pathological conditions described in the medical reports.

3 Experiments and Results

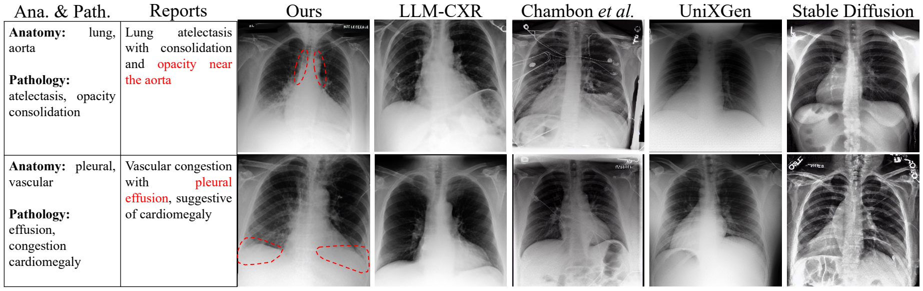
3.1 Experiment Setting

Datasets. In our experiments, we utilize two widely used publicly available chest X-ray datasets: MIMIC-CXR [15] and OpenI [9]. The MIMIC-CXR dataset is a large-scale dataset consisting of 473,057 images and 206,563 corresponding medical reports from 63,478 patients. We adhere to the official dataset splits, which allocate 368,960 samples for training, 2,991 for validation, and 5,159 for testing. On the other hand, the OpenI dataset is smaller in size, containing 3,684 report-image pairs. The dataset is divided into 2,912 samples for training and 772 for testing.

Implementation and Metrics. We use the pre-trained image encoder, text encoder and LLM [3] in the fine-grained alignment synthesis module. The pre-trained VQ-GAN model [11] is adopted to encode image patches to image tokens, and decode the image tokens to images. All the models are frozen in the framework. To assess the image quality, we use the Fréchet inception distance (FID) [13] and Natural Image Quality Evaluator (NIQE) [23]. The lower values indicate the better performance.

Table 1. Comparison of report-to-CXR generation performance on the MIMIC-CXR and the OpenI datasets.

	MIMIC-CXR		OpenI	
Methods	FID ↓	NIQE ↓	FID ↓	NIQE ↓
Stable diffusion [24]	14.5194	5.7455	11.3305	5.7455
Chambon *et al.* [2]	12.7408	4.4534	8.2887	4.4534
RoentGen [1]	13.1979	5.1286	6.5666	5.1286
UniXGen [17]	14.0569	6.2759	7.5210	6.2759
LLM-CXR [18]	11.9873	4.5876	5.9869	4.5876
Ours	8.8213	4.1138	5.7455	4.1138



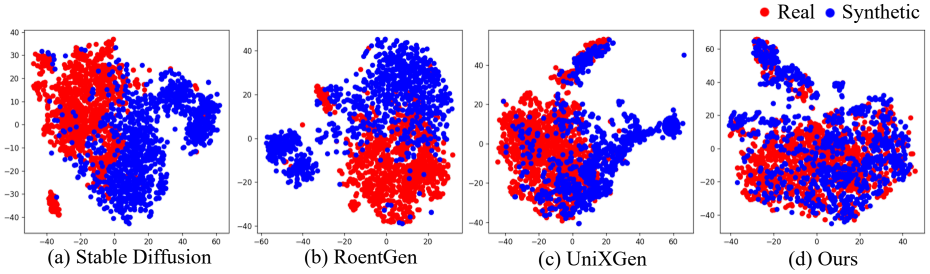
Ana. & Path.	Reports
Anatomy: lung, aorta **Pathology:** atelectasis, opacity consolidation	Lung atelectasis with consolidation and opacity near the aorta
Anatomy: pleural, vascular **Pathology:** effusion, congestion cardiomegaly	Vascular congestion with pleural effusion, suggestive of cardiomegaly

Fig. 3. The generated chest X-ray images of the MIMIC-CXR dataset with highlighted regions.

3.2 Comparison with State-of-the-Arts

We conducted a quantitative comparison of our method with state-of-the-art text-to-image generation methods, such as Stable Diffusion [24], and report-to-CXR generation approaches, including Chambon *et al.* [2], RoentGen [1], UniXGen [17], and LLM-CXR [18]. As shown in Table 1, our method achieves the highest FID scores on both datasets, demonstrating its superior performance in generating CXR images with descriptive reports. To further investigate the high-level feature distribution of the generated CXR images, we randomly selected 1,000 cases from the test set and performed t-SNE visualization on both real and synthetic CXR images from the MIMIC-CXR dataset. Fig. 4 illustrates that while the synthetic CXR images generated by current methods exhibit notable differences from the real ones, our method produces images that nearly overlap with the real images in the t-SNE visualization, highlighting its exceptional ability to generate highly realistic CXR images.

Fig. 3 presents a comparison of CXR images generated by our method and existing approaches on both the MIMIC-CXR and OpenI datasets. In the first example, our proposed method successfully synthesizes the 'opacity near the aorta' described in the input report, while other methods struggle to generate this specific feature. This observation highlights the superior capability of our



(a) Stable Diffusion (b) RoentGen (c) UniXGen (d) Ours

Fig. 4. The t-SNE visualization of the real and synthetic CXR images on the MIMIC-CXR dataset.

Table 2. Anatomy and pathology classification performance (%) comparison of MIMIC-CXR dataset and CXR images generated by our method.

	Anatomy		Pathology		Overall	
Data source	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
MIMIC-CXR	91.21	78.17	**92.19**	74.42	91.59	76.74
Ours	**94.74**	**83.88**	92.11	**77.02**	**93.74**	**81.27**

method in producing highly realistic and accurate CXR images that faithfully reflect the content of the corresponding reports.

3.3 Semantic Analysis

To further analyze the semantic information of the synthetic images, we pre-train a classifier on the MIMIC-CXR dataset for the multi-label anatomy and pathology classification. Then, we test the classification performance of the real and synthetic images. In Table 2, we show the classification performance for the test set of the MIMIC-CXR dataset and CXR images generated by our method. Our method significantly outperforms the real data by a large margin with an accuracy of 2.15%, implying our synthetic data with accurate semantic information about anatomical structures and pathological conditions. Moreover, we also show the performance of each category for anatomy and pathology classification. As visualized in Fig. 5, our method achieves higher precision than the real data in most categories. These indicate the medical images generated by our method preserve more semantic information in terms of anatomy and pathology.

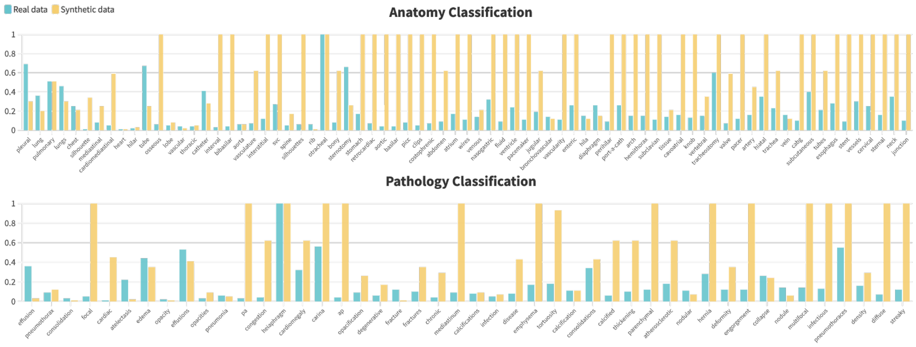


Fig. 5. Anatomy and pathology classification performance of each category. Each column shows the precision score.

4 Conclusion

To synthesize high-quality medical images with detailed anatomical and pathology information, we introduce a medical image synthesis model to generate anatomy-pathology prompts and highly detailed medical images. In order to provide the descriptive reports with anatomy and pathology information, we design an anatomy-pathology prompting to establish anatomy and pathology vocabularies and employ GPT4 to automatically generate reports. With the descriptive reports, we devise a fine-grained alignment based synthesis module to perform alignment between the reports and pre-defined visual codebook to obtain matched keypatches. Moreover, this module utilizes the LLM and VQ-GAN to convert reports, instructions, and matched keypatches to synthetic images.

Acknowledgement. This work was supported by the Hong Kong Research Grants Council (RGC) General Research Fund 14204321.

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

References

1. Chambon, P., Bluethgen, C., Delbrouck, J.B., Van der Sluijs, R., Połacin, M., Chaves, J.M.Z., Abraham, T.M., Purohit, S., Langlotz, C.P., Chaudhari, A.: Roentgen: vision-language foundation model for chest x-ray generation. arXiv preprint arXiv:2211.12737 (2022)
2. Chambon, P., Bluethgen, C., Langlotz, C.P., Chaudhari, A.: Adapting pretrained vision-language foundational models to medical imaging domains. arXiv preprint arXiv:2210.04133 (2022)

3. Chen, W., Li, X., Shen, L., Yuan, Y.: Fine-grained image-text alignment in medical imaging enables cyclic image-report generation. arXiv preprint arXiv:2312.08078 (2023)
4. Chen, W., Liu, J., Chow, T.W., Yuan, Y.: Star-rl: Spatial-temporal hierarchical reinforcement learning for interpretable pathology image super-resolution. IEEE Trans. Med. Imag. (2024)
5. Chen, W., Liu, Y., Hu, J., Yuan, Y.: Dynamic depth-aware network for endoscopy super-resolution. IEEE J. Biomed. Health Inform. **26**(10), 5189–5200 (2022)
6. Chen, W., Yu, S., Ma, K., Ji, W., Bian, C., Chu, C., Shen, L., Zheng, Y.: Tw-gan: Topology and width aware gan for retinal artery/vein classification. Med. Image Anal. **77**, 102340 (2022)
7. Chen, W., Yu, S., Wu, J., Ma, K., Bian, C., Chu, C., Shen, L., Zheng, Y.: Tr-gan: Topology ranking gan with triplet loss for retinal artery/vein classification. In: MICCAI. pp. 616–625. Springer (2020)
8. Chen, W., Zhao, W., Chen, Z., Liu, T., Liu, L., Liu, J., Yuan, Y.: Mask-aware transformer with structure invariant loss for ct translation. Med. Image Anal. **96**, 103205 (2024)
9. Demner-Fushman, D., Kohli, M.D., Rosenman, M.B., Shooshan, S.E., Rodriguez, L., Antani, S., Thoma, G.R., McDonald, C.J.: Preparing a collection of radiology examinations for distribution and retrieval. JAMIA **23**(2), 304–310 (2016)
10. El Jiani, L., El Filali, S., et al.: Overcome medical image data scarcity by data augmentation techniques: A review. In: ICM. pp. 21–24. IEEE (2022)
11. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis. In: CVPR. pp. 12873–12883 (2021)
12. Henning, C.A., Ewerth, R.: Estimating the information gap between textual and visual representations. In: ICMR. pp. 14–22 (2017)
13. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. NeurIPS **30**, 6629–6640 (2017)
14. Ji, W., Chen, W., Yu, S., Ma, K., Cheng, L., Shen, L., Zheng, Y.: Uncertainty quantification for medical image segmentation using dynamic label factor allocation among multiple raters. In: MICCAI on QUBIQ workshop. vol. 2 (2020)
15. Johnson, A.E., Pollard, T.J., Berkowitz, S.J., Greenbaum, N.R., Lungren, M.P., Deng, C.y., Mark, R.G., Horng, S.: Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports. Scientific data **6**(1), 317 (2019)
16. Karbhari, Y., Basu, A., Geem, Z.W., Han, G.T., Sarkar, R.: Generation of synthetic chest x-ray images and detection of covid-19: A deep learning based approach. Diagnostics **11**(5), 895 (2021)
17. Lee, H., Kim, W., Kim, J.H., Kim, T., Kim, J., Sunwoo, L., Choi, E.: Unified chest x-ray and radiology report generation model with multi-view chest x-rays. arXiv preprint arXiv:2302.12172 (2023)
18. Lee, S., Kim, W.J., Ye, J.C.: Llm itself can read and generate cxr images. arXiv preprint arXiv:2305.11490 (2023)
19. Liu, J., Guo, X., Yuan, Y.: Graph-based surgical instrument adaptive segmentation via domain-common knowledge. IEEE Trans. Med. Imag. **41**(3), 715–726 (2021)
20. Liu, J., Guo, X., Yuan, Y.: Prototypical interaction graph for unsupervised domain adaptation in surgical instrument segmentation. In: MICCAI. pp. 272–281. Springer (2021)

21. Liu, J., Zhang, Y., Chen, J.N., Xiao, J., Lu, Y., A Landman, B., Yuan, Y., Yuille, A., Tang, Y., Zhou, Z.: Clip-driven universal model for organ segmentation and tumor detection. In: ICCV. pp. 21152–21164 (2023)
22. Madani, A., Moradi, M., Karargyris, A., Syeda-Mahmood, T.: Chest x-ray generation and data augmentation for cardiovascular abnormality classification. In: Medical imaging 2018: Image processing. vol. 10574, pp. 415–420. SPIE (2018)
23. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal Process. Lett. **20**(3), 209–212 (2012)
24. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR. pp. 10684–10695 (2022)
25. Wenting, C., Jie, L., Yixuan, Y.: Bi-vlgm: Bi-level class-severity-aware vision-language graph matching for text guided medical image segmentation. arXiv preprint arXiv:2305.12231 (2023)
26. Wu, J., Yu, S., Chen, W., Ma, K., Fu, R., Liu, H., Di, X., Zheng, Y.: Leveraging undiagnosed data for glaucoma classification with teacher-student learning. In: MICCAI. pp. 731–740. Springer (2020)
27. Yang, X., Li, X., Li, X., Chen, W., Shen, L., Li, X., Deng, Y.: Two-stream regression network for dental implant position prediction. Expert Syst. with Appl. **235**, 121135 (2024)
28. Zhang, T., Fu, H., Zhao, Y., Cheng, J., Guo, M., Gu, Z., Yang, B., Xiao, Y., Gao, S., Liu, J.: Skrgan: Sketching-rendering unconditional generative adversarial networks for medical image synthesis. In: MICCAI. pp. 777–785. Springer (2019)