Supplementary Material for FreestyleRet: Retrieving Images from Style-Diversified Queries

Hao Li[1,2]*, Yanhao Jia[3,4]*, Peng Jin[1,2], Zesen Cheng[1], Kehan Li[1], Jialu Sui[5], Chang Liu[6], and Li Yuan[1,2]†

[1] School of Electronic and Computer Engineering, Peking University
[2] Peng Cheng Laboratory, Shenzhen, China
[3] College of Computing and Data Science, Nanyang Technological University
[4] Deep NeuroCognition Lab, I2R and CFAR, Agency for Science, Technology and Research, Singapore
[5] School of Science and Engineering, Chinese University of Hong Kong
[6] Department of Automation, Tsinghua University

1 Further Discussions for Related Works

1.1 Our Setting VS Domain Adaptation

The domain adaptation task and the multi-task learning (our setting) all belong to the field of Transfer Learning. Both domain adaptation and multi-task learning (our setting) aim to improve the performance from one task to another task. Thus, with all these similarities, it is vital to point out the difference between domain adaptation and multi-task learning (our setting). Here we first analyse from the data aspect:

- In the domain adaptation setting, labeled data are available only in the source domain. For the target domain, there are only a few labeled data or no labeled data.
- In the multi-task learning (our setting), labeled data are available in both the source task and the target tasks. For our style-diversified QBIR setting, we collect 10,000 images for each style modality, which is available for CLIP fine-tuning.

From the methodology aspect, prompt learning methods, including CoOP [29], CoCoOP [28], MaPLe [10], are commonly used in both domain adaptation and multi-task learning. Thus, we compare the performance of our FreestyleRet and these baselines and demonstrate the effectiveness of our proposed framework.

1.2 Our Setting VS Other Retrieval Settings

Our FreestyleRet proposes a novel retrieval setting: Image Retrieval with Style-Diversified Queries. However, during our survey of related works, we have identified several closely related retrieval tasks, including Composed Image Retrieval [22], User Generalized Image Retrieval [14], Fashion Retrieval [7], Synthesis Image Retrieval [22], and Sketch Retrieval [24]. Consequently, we summarize these tasks and highlight the differences and contributions of our novel task: Style-Diversified Image Retrieval Task in comparison to them.

Composed Image Retrieval Composed Image Retrieval (CIR) [13, 17, 22] aims to retrieve a target image based on a query composed of a reference image and a relative caption that describes the difference between the two images. Zero-shot CIR [1] is a derivative task associated with CIR, learning image-text joint features without requiring a labeled training dataset. The CIR task has been extensively studied in various Vision and Language tasks, such as visual question answering [11, 26] and visual grounding [3, 4].
Difference: The composed image retrieval focuses on retrieving natural images from composed queries (image+text) and does not consider style-diversified query inputs. However, our style-diversified retrieval setting achieves not only style-diversified query-based retrieval ability but also achieves good performance when retrieving from composed queries with various styles (sketch+text, art+text, low resolution+text).

User Generalized Image Retrieval The User Generalized Image Retrieval (UGIR) [14] is a task that retrieves natural images and text. Formally, UGIR defines data belonging to one user as a user domain, and the differences among different user domains as user domain shift. UGIR trains on a user domain and tests on various user domains to evaluate their feature generalization.
Difference: The user-generalized image retrieval task focuses on exploring the domain adaptation capability of retrieval models, where the domain refers to a natural image dataset encompassing diverse categories of objects. However, in our style-diversified retrieval setting, we adapt the domain of a wide range of image styles as queries, including natural images, sketches, artistic images, and blurry low-resolution images.

Fashion, Synthesis, and Sketch Retrieval Fashion Retrieval [5, 7, 19], Synthesis Image Retrieval [20, 22, 27], and Sketch Retrieval [12, 18] aim to retrieve from one specific class of images, including the fashion clothes, synthesis natural scenes, and sketch-based images. These tasks are applied in the search engines.
Difference: The fashion retrieval, synthesis retrieval, and sketch retrieval all focus on retrieving from single-style queries. However, our style-diversified retrieval maintains the ability to retrieve based on queries with various styles, including sketch images and synthesis art-style images.

1.3 Controllable Image Generation for Synthesis Datasets

Controllable Image Generation [2] aims to generate an image with a condition, including text, image, etc. With the development of VAE, GAN [15], and Diffusion [9, 21] models, deep-learning methods achieve outstanding quality in the visual generation domain. Due to the high image quality and the open-vocabulary ability of generative models, researchers apply generative models to synthesize images for dataset generations [16, 25]. In this paper, we apply AnimateDiff [6] for our DSR dataset generation. In order to assess the potential hallucination issues [8,23] that may arise in generative models, we introduce the hps-v2 metric to evaluate the quality of the generated images.

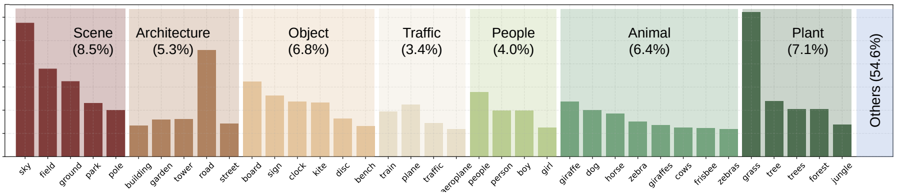

Fig. 1: Concept distribution of our DSR dataset. Our dataset exhibits a diverse distribution in different concept domains.

2 Supplements for Ablation Studies

We present the supplementary for the ablation studies to analyze the impact of the baseline setting, the prompt-token inserting strategy, and the epoch amount for our FreestyleRet framework.

2.1 Concept Distribution Analysis for our DSR Dataset

Fig. 1 shows the concept distribution of the images in our DSR dataset. Our dataset exhibits a diverse distribution in different concept domains. The top-5 concept classes in our DSR dataset are "Scene", "Architecture", "Object", "Animal", and "Plant". The sum of the proportions of the top five concept classes in the dataset accounts for 31.4%, leaving ample space for a rich diversity of other concept classes. This further substantiates the equilibrium of our dataset at the concept level.

2.2 Details for the Baseline Settings and Implementations

In the experimental part, we select the commonly-used cross-modality pre-trained models, CLIP and BLIP, as the baselines for our style-diversified retrieval task. There are three settings for the pre-trained baseline models, including zero-shot tuning, prompt tuning, and fine-tuning setting.

As shown in Table.1 in the main paper, for visual-linguistic baselines (CLIP-finetune, BLIP-finetune), we finetune all their parameters with the learning rate of 1e-6. For ImageBind and LanguageBind, we propose their zero-shot performance due to their multi-modality capability. For prompt learning baselines (VPT, CoCoOP, MaPLe), we train them on the DSR dataset with a learning rate of 1e-5 and 40 epochs.

In Tab. 1, we make extra ablations on the performance of three tuning settings. Comparing all three settings in line.1-3, the fine-tuning setting outperforms others significantly in style-diversified retrieval, including sketch, art, and low resolution. However, both prompt-tuning and fine-tuning settings have similar performance on text-image retrieval. Our FreestyleRet framework outperforms all three tuning settings in text, sketch, art, and low-resolution settings.

Table 1: Evaluation of different baseline settings. Our model surpasses zero-shot tuning, prompt tuning, and fine-tuning baselines.

#	Method	Text → Image		Sketch → Image		Art → Image		Low-Res → Image	
		R@1↑	R@5↑	R@1↑	R@5↑	R@1↑	R@5↑	R@1↑	R@5↑
1	CLIP-ZeroShot	66.1	94.7	47.5	77.3	58.5	93.7	45.0	75.7
2	CLIP-Prompt	72.2	96.4	63.6	93.6	58.2	90.4	78.8	97.1
3	CLIP-Finetune	73.8	97.4	78.4	95.6	70.2	96.6	84.0	97.1
4	BLIP-Prompt	74.3	95.3	67.1	90.9	51.1	85.3	77.2	95.8
5	BLIP-Finetune	80.1	98.7	75.3	94.8	63.7	93.4	83.7	96.3
6	**FreestyleRet(C)**	71.4	97.2	81.6	**98.0**	72.3	**98.1**	86.7	98.2
7	**FreestyleRet(B)**	**81.6**	**99.2**	**81.2**	97.1	**74.5**	97.4	**90.5**	**98.5**

Table 2: The ablation analysis for the prompt token inserting strategy. We ablate three prompt-token inserting strategies in our FreestyleRet framework, including inserting in the shallow layer, inserting in the deep layer, and inserting in both layers. Experiments show that inserting in both shallow and deep layers achieves the best.

#	**Shallow**	**Bottom**	**S→I**	**A→I**	**LR→I**
1	Random	-	78.1	70.2	85.3
2	Style Space	-	78.5	71.6	85.5
3	Gram Matrix	-	79.0	70.9	84.2
4	-	Random	78.5	70.3	83.5
5	-	Style Space	79.4	70.7	84.7
6	-	Gram Matrix	79.2	70.8	83.8
7	Style Space	Gram	**81.6**	**72.3**	**86.7**

2.3 Additional Ablation for Prompt Token Inserting Strategies

In the main paper, we conducted ablation experiments on the initialization choices and the number of prompt-tuning tokens in the prompt-tuning structure. In the supplementary material, we further performed ablation on the number of layers in the prompt tuning structure. Specifically, in Table. 2, we compared the performance of the model when only inserting prompt tokens in shallow layers, only inserting prompt tokens in deep layers, and inserting prompt tokens in both shallow and deep layers. All experiments in Table. 2 are conducted by our FreestyleRet framework on the DSR dataset. "S→I" represents sketch to image retrieval. "A→I" represents art to image retrieval. "LR→I" represents low-resolution to image retrieval.

Compare line.7 with line.1-3 and line.4-6 in Table. 2, inserting prompt tokens in both shallow and deep layers outperforms other inserting strategies. In comparison to the random initialization method (line.1&4), both style initialization (line.2&5) and gram initialization (line.3&6) result in higher accuracy. Additionally, the deep-layer prompt provides the encoder with a larger bias, contributing to a slight increase in performance compared to the shallow-layer prompt strategy.

2.4 Additional Ablation for the Model Structure

VGG Feature extractor: When extracting stylistic features from images, we employ the VGG model as our feature extractor. Given that the VGG model is structured as a stack of multiple convolutional neural networks, the selection of which layer to use for image representation is a critical issue. To address this, we conducted an ablation study (Table. 3) on the chosen number of layers and demonstrated that the third convolutional layer performs optimally. This phenomenon is also consistent with related works on image style transfer.
Clustering Iteration: In the process of style space generation, we utilize the K-Nearest Neighbors (KNN) algorithm to cluster images of various styles, thereby constructing a style space. We conducted an ablation study (Table. 3) on the clustering iterations within the KNN algorithm and verified that the optimal performance is achieved when the number of iterations is set to 300.

Table 3: The ablation analysis for the VGG feature extractor and the clustering iteration setting.

VGG Layer	**Text→I**	**Sketch→I**	KNN Iter	**Text→I**	**Sketch→I**
Conv-2	79.4	80.9	200	80.7	81.5
Conv-3	**81.6**	**82.0**	300	**81.6**	**82.0**
Conv-4	80.3	80.7	400	81.6	81.9

2.5 Epoch Analysis for the FreestyleRet

To demonstrate the fast convergence and low computational cost of our FreestyleRet framework, we conduct the epoch analysis for our FreestyleRet and visualize the performance change under different epochs training.

As shown in Fig.2, our FreestyleRet framework achieves better performance and faster convergence speed with 5-10 training epochs compared to other baselines such as prompting tuning BLIP, CLIP, and VPT models. These pre-trained baseline models need at least 50 or more training epochs to converge.

Also, we observe that text and low-resolution retrieval converge after 5 training epochs, faster than art and sketch retrieval (10 epochs). The text modal and the low-resolution style have less information gap between the natural image modality, so their performance converges faster. On the other hand, the sketch

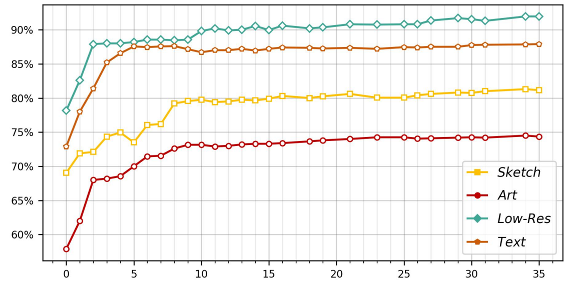

Fig. 2: The epoch analysis for our FreestyleRet framework. For the style-diversified retrieval task, our lightweight framework achieves a rather good performance under 10 epochs.

style and the art style, containing more style and textural information, require more epochs (about 10) to achieve better retrieval accuracy. Additionally, each training epoch only takes 4 minutes. The performance in the main body is an average of epoch-5, epoch-10, and epoch-20 evaluation results.

Table 4: The Text-Retrieval performance of our FreestyleRet and baselines.

#	**Method**	Image→Text		Sketch→Text		Art→Text		Low-Res→Text	
		R@1	R@5	R@1	R@5	R@1	R@5	R@1	R@5
1	CLIP*	55.2	90.8	48.4	87.9	64.5	96.8	42.6	81.8
2	BLIP*	71.4	94.9	55.5	87.0	81.0	98.8	49.2	81.8
3	VPT	52.2	91.7	45.2	87.7	52.7	94.6	44.3	84.6
4	ImageBind	73.5	96.5	56.1	88.4	82.7	99.0	42.4	73.8
5	LanguageBind	80.5	98.3	63.9	91.6	87.2	99.7	56.9	86.8
6	FreestyleRet-CLIP	71.6	98.0	66.7	**96.7**	74.4	99.1	64.1	94.8
7	FreestyleRet-BLIP	**82.8**	**99.0**	**71.0**	96.4	**86.6**	**99.7**	**69.5**	**96.9**

3 Supplements for Experimental Results

In order to comprehensively validate the superiority of our FreestyleRet model in handling the retrieval of queries with different styles, we conducted extensive experiments involving cross-modal retrieval among various style-diversified queries, including any queries to Text modality, any queries to Art modality, any queries to Sketch modality, and any queries to Low-resolution modality.

We present the performance comparison between our FreestyleRet and other baselines in Table. 4 (Any→Text), Table. 5 (Any→Art), Table. 6 (Any→Sketch), and Table. 7 (Any→Low-resolution Images). All experiments are conducted on the DSR dataset. Experimental results demonstrate that our FreestyleRet framework achieves state-of-the-art (SOTA) performance in almost all retrieval scenarios. Specifically, in complex scenarios including sketch and art style retrieval, our

Table 5: The Art-Retrieval performance of our FreestyleRet and baselines.

#	Method	Image→Art		Sketch→Art		Text→Art		Low-Res→Art	
		R@1	R@5	R@1	R@5	R@1	R@5	R@1	R@5
1	CLIP*	63.0	94.7	61.2	92.7	75.5	98.2	51.9	87.9
2	BLIP*	57.1	88.5	44.8	82.8	82.8	98.7	39.4	79.3
3	VPT	67.4	95.5	60.3	93.1	61.6	96.5	44.3	84.6
4	ImageBind	46.4	80.5	28.7	60.8	82.6	98.9	57.8	89.6
5	LanguageBind	65.8	93.2	41.1	77.7	86.7	99.2	34.8	72.0
6	FreestyleRet-CLIP	72.9	**97.8**	**66.5**	**96.2**	85.0	99.6	**62.8**	**94.1**
7	FreestyleRet-BLIP	**73.6**	97.4	63.1	94.4	**90.2**	**99.7**	60.1	92.2

Table 6: The Sketch-Retrieval performance of FreestyleRet and baselines.

#	Method	Image→Sketch		Art→Sketch		Text→Sketch		Low-Res→Sketch	
		R@1	R@5	R@1	R@5	R@1	R@5	R@1	R@5
1	CLIP*	70.5	96.1	60.5	92.9	55.0	90.8	60.4	90.9
2	BLIP*	69.8	93.5	47.6	82.8	58.6	89.8	52.3	82.8
3	VPT	71.7	96.2	62.3	92.9	49.4	88.6	63.3	91.5
4	ImageBind	54.0	81.8	38.3	71.6	56.1	88.4	26.2	52.5
5	LanguageBind	74.6	96.1	57.5	87.0	65.7	94.0	54.5	83.8
6	FreestyleRet-CLIP	77.8	**98.1**	66.5	**96.2**	72.3	97.4	68.7	**95.1**
7	FreestyleRet-BLIP	**80.5**	97.7	**66.8**	94.9	**76.6**	**97.7**	**71.1**	94.3

FreestyleRet model outperforms other baseline models by a significant margin of 6%-10% due to the integration of our style extraction module and style-based prompt tuning module.

In Table. 7, we observed that the fine-tuned BLIP model outperforms our FreestyleRet model in the retrieval of Images to low-resolution images. This is because there is a high semantic similarity between low-resolution images and natural images, and simple prompt tuning allows the baseline model to achieve good results. However, our model still surpasses the baseline in tasks involving cross-modal retrieval from other modalities to low-resolution image modalities.

4 Supplements for Case Study

As shown in Fig. 3 and Fig. 4, we add more visualization results in our supplementary material. Each sample has three images to compare the retrieval performance between our FreestyleRet and the CLIP baseline on the DSR dataset. The left images are the queries randomly selected from different styles. The middle and the right images are the retrieval results of our FreestyleRet-BLIP model and the original BLIP model, respectively.

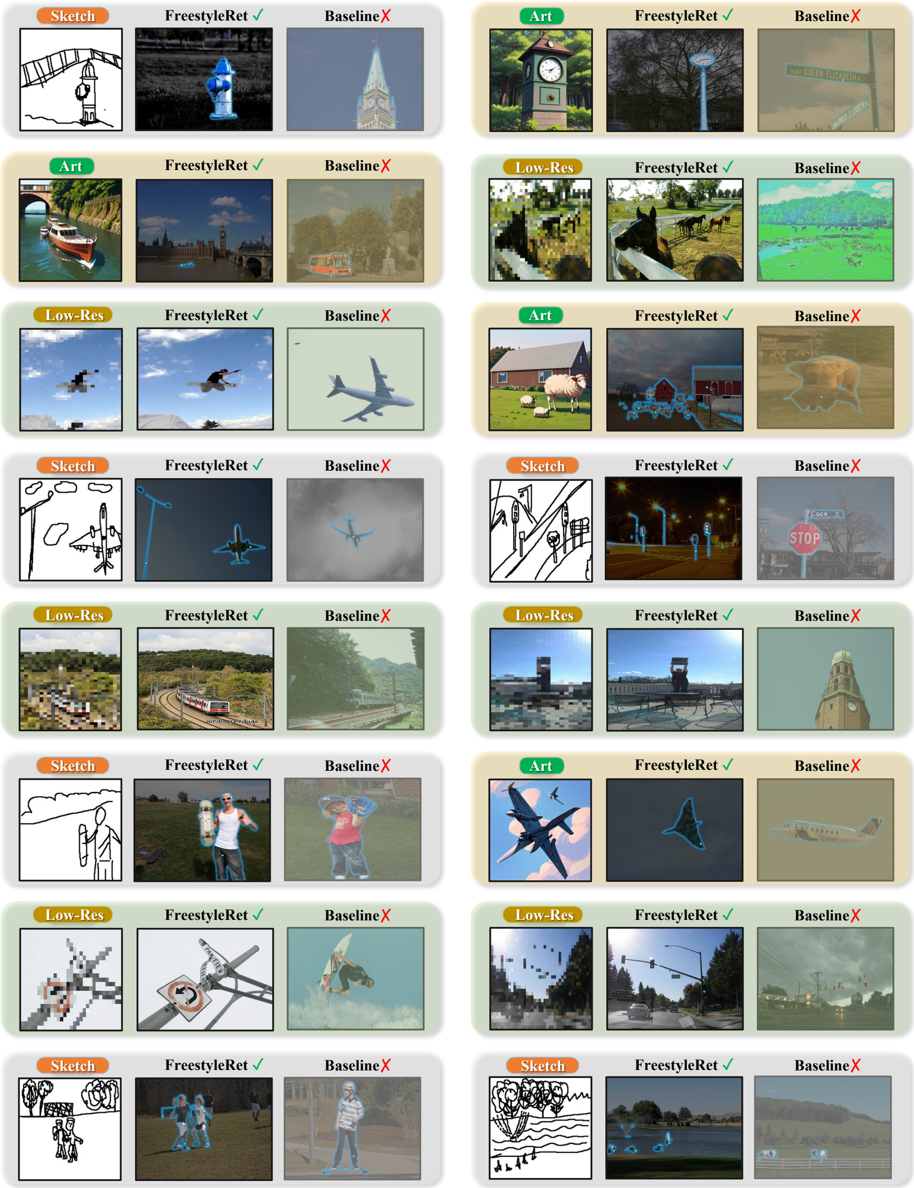

Fig. 3: The Visualization of our FreestyleRet-BLIP and the baseline BLIP model on our DSR dataset.

Table 7: The Low-Resolution Image Retrieval performance of our FreestyleRet framework and baseline models.

#	Method	Image→Low-Res		Art→Low-Res		Text→Low-Res		Sketch→Low-Res	
		R@1	R@5	R@1	R@5	R@1	R@5	R@1	R@5
1	CLIP*	79.3	97.2	53.0	89.2	46.0	82.3	59.5	92.4
2	BLIP*	**89.0**	40.8	73.9	87.0	51.5	84.4	51.4	82.3
3	VPT	75.5	95.7	56.7	90.3	45.6	85.7	61.9	91.6
4	ImageBind	59.9	83.1	25.2	49.8	42.4	73.8	30.7	56.8
5	LanguageBind	81.0	97.6	47.3	81.2	58.5	87.9	55.5	85.6
6	FreestyleRet-CLIP	80.2	97.5	62.6	**95.2**	68.7	96.6	67.4	**95.3**
7	FreestyleRet-BLIP	88.4	**98.6**	**63.9**	94.1	**76.0**	**97.5**	**71.3**	94.3

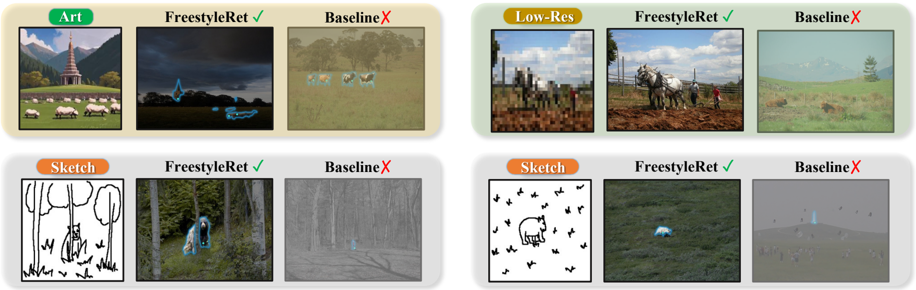

Fig. 4: The Visualization of our FreestyleRet-BLIP and the baseline BLIP model on our DSR dataset.

References

1. Baldrati, A., Agnolucci, L., Bertini, M., Del Bimbo, A.: Zero-shot composed image retrieval with textual inversion. arXiv preprint arXiv:2303.15247 (2023)
2. Cao, P., Zhou, F., Song, Q., Yang, L.: Controllable generation with text-to-image diffusion models: A survey. arXiv preprint arXiv:2403.04279 (2024)
3. Cheng, Z., Jin, P., Li, H., Li, K., Li, S., Ji, X., Liu, C., Chen, J.: Wico: Win-win cooperation of bottom-up and top-down referring image segmentation. arXiv preprint arXiv:2306.10750 (2023)
4. Cheng, Z., Li, K., Jin, P., Ji, X., Yuan, L., Liu, C., Chen, J.: Parallel vertex diffusion for unified visual grounding. arXiv preprint arXiv:2303.07216 (2023)
5. Guo, X., Wu, H., Cheng, Y., Rennie, S., Tesauro, G., Feris, R.: Dialog-based interactive image retrieval. arXiv: Computer Vision and Pattern Recognition,arXiv: Computer Vision and Pattern Recognition (May 2018)
6. Guo, Y., Yang, C., Rao, A., Wang, Y., Qiao, Y., Lin, D., Dai, B.: Animatediff: Animate your personalized text-to-image diffusion models without specific tuning. arXiv preprint arXiv:2307.04725 (2023)
7. Han, X., Wu, Z., Huang, P.X., Zhang, X., Zhu, M., Li, Y., Zhao, Y., Davis, L.S.: Automatic spatially-aware fashion concept discovery. In: Proceedings of the IEEE international conference on computer vision. pp. 1463–1471 (2017)

8. Jesson, A., Beltran-Velez, N., Chu, Q., Karlekar, S., Kossen, J., Gal, Y., Cunningham, J.P., Blei, D.: Estimating the hallucination rate of generative ai. arXiv preprint arXiv:2406.07457 (2024)
9. Jin, P., Li, H., Cheng, Z., Li, K., Ji, X., Liu, C., Yuan, L., Chen, J.: Diffusionret: Generative text-video retrieval with diffusion model. arXiv preprint arXiv:2303.09867 (2023)
10. Khattak, M.U., Rasheed, H., Maaz, M., Khan, S., Khan, F.S.: Maple: Multi-modal prompt learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19113–19122 (2023)
11. Li, H., Jin, P., Cheng, Z., Zhang, S., Chen, K., Wang, Z., Liu, C., Chen, J.: Tg-vqa: Ternary game of video question answering. arXiv preprint arXiv:2305.10049 (2023)
12. Lin, F., Li, M., Li, D., Hospedales, T., Song, Y.Z., Qi, Y.: Zero-shot everything sketch-based image retrieval, and in explainable style. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23349–23358 (2023)
13. Liu, Z., Rodriguez-Opazo, C., Teney, D., Gould, S.: Image retrieval on real-life images with pre-trained vision-and-language models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2125–2134 (2021)
14. Ma, X., Yang, X., Gao, J., Xu, C.: The model may fit you: User-generalized cross-modal retrieval. IEEE Transactions on Multimedia **24**, 2998–3012 (2021)
15. Meng, X., Huang, J., Li, Z., Wang, C., Teng, S., Grau, A.: Dedustgan: Unpaired learning for image dedusting based on retinex with gans. Expert Systems with Applications **243**, 122844 (2024)
16. Paulin, G., Ivasic-Kos, M.: Review and analysis of synthetic dataset generation methods and techniques for application in computer vision. Artificial intelligence review **56**(9), 9221–9265 (2023)
17. Saito, K., Sohn, K., Zhang, X., Li, C.L., Lee, C.Y., Saenko, K., Pfister, T.: Pic2word: Mapping pictures to words for zero-shot composed image retrieval. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19305–19314 (2023)
18. Sangkloy, P., Jitkrittum, W., Yang, D., Hays, J.: A sketch is worth a thousand words: Image retrieval with text and sketch. In: European Conference on Computer Vision. pp. 251–267. Springer (2022)
19. Sui, J., Ma, X., Zhang, X., Pun, M.O.: Dtrn: Dual transformer residual network for remote sensing super-resolution. In: IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium. pp. 6041–6044. IEEE (2023)
20. Sui, J., Ma, X., Zhang, X., Pun, M.O.: Gcrdn: Global context-driven residual dense network for remote sensing image super-resolution. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2023)
21. Tang, Z., Ren, Z., Zhao, X., Wen, B., Tremblay, J., Birchfield, S., Schwing, A.: Nerfdeformer: Nerf transformation from a single view via 3d scene flows. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10293–10303 (2024)
22. Vo, N., Jiang, L., Sun, C., Murphy, K., Li, L.J., Fei-Fei, L., Hays, J.: Composing text and image for image retrieval-an empirical odyssey. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6439–6448 (2019)
23. Wang, Y., Liu, X., Li, Y., Chen, M., Xiao, C.: Adashield: Safeguarding multimodal large language models from structure-based attack via adaptive shield prompting. arXiv preprint arXiv:2403.09513 (2024)

24. Xu, P., Huang, Y., Yuan, T., Pang, K., Song, Y.Z., Xiang, T., Hospedales, T.M., Ma, Z., Guo, J.: Sketchmate: Deep hashing for million-scale human sketch retrieval. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8090–8098 (2018)
25. Yan, Z., Yao, T., Chen, S., Zhao, Y., Fu, X., Zhu, J., Luo, D., Yuan, L., Wang, C., Ding, S., et al.: Df40: Toward next-generation deepfake detection. arXiv preprint arXiv:2406.13495 (2024)
26. Ye, Q., Cao, B., Chen, N., Xu, W., Zou, Y.: Fits: Fine-grained two-stage training for knowledge-aware question answering. arXiv preprint arXiv:2302.11799 (2023)
27. Zhao, Y., Wang, J., Qi, Q.: Mindcamera: Interactive image retrieval and synthesis. In: 2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP) (Jun 2018). `https://doi.org/10.1109/ivmspw.2018.8448722`, `http://dx.doi.org/10.1109/ivmspw.2018.8448722`
28. Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Conditional prompt learning for vision-language models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16816–16825 (2022)
29. Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Learning to prompt for vision-language models. International Journal of Computer Vision **130**(9), 2337–2348 (2022)