Proceedings of the 2nd International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/62/20240387

How pre-trained large model can help, when SAM meets image restoration

An Ping

Sun Yat-sen University (Shenzhen Campus), Shenzhen, Guangdong, China, 518107

pingan3@mail2.sysu.edu.cn

Abstract. In the process of image capture, degradation is inevitable due to noise, motion, down-sampling and so on. Therefore, image restoration is essential to improve the quality of images to enhance their visual effects and benefit downstream tasks. Adding prior knowledge can help the model better understand the image content and restoration requirements, thus improving the quality and efficiency. Semantic-level prior information can be generated by pre-trained large-scale models, such as segment anything models (SAM), and applied to a large number of downstream tasks. SAM has demonstrated powerful robustness and stability in restoration tasks, such as denoising, super-resolution, low-light enhancement, etc. Meanwhile, as an interactive component, SAM brings more control for users during the repair process. In this paper, we focus on the importance of SAM as prior information and systematically summarize a series of recent works combining SAM prior and low-level image restoration from three perspectives. In addition, we have summarized some potential problems and future directions of SAM.

Keywords: Survey, Segment Anything Model, Foundation Model, Image Restoration.

1. Introduction

In the process of image formation, transmission, reception, and processing, there are inevitably external and internal interferences, causing a certain degree of degradation to the original image. Such degradation may deteriorate image quality, making the image blurry and even masking image features, increasing the difficulty of subsequent image analysis. Therefore, restoring the original image during the pre-processing stage is an important aspect of image processing, including tasks such as image denoising [25, 26], image super-resolution [14], and image inpainting [22]. Image restoration is not only crucial in computer vision tasks, but also a crucial step in research on other image processing tasks such as medical imaging [13], remote sensing [30], and digital forensics [31]. To understand the process of image restoration, we use a degradation model to describe the mathematical relationship between the original image and the degraded image:

$$y = Ax + n. \tag{1}$$

It usually consists of two parts: a degenerate function A and an additive noise term n. The degradation function captures the effects of blurring, motion, geometric distortion, or other factors that change the content of an image. The noise term explains random fluctuations in image intensity due to sensor defects, transmission errors, or environmental conditions. Depending on the application and domain, different degradation models may be more appropriate and realistic. For example, common degradation models include atmospheric turbulence models and motion ambiguity models, both of which are models of degradation caused by general motion.

Therefore, any image restoration task can be abstracted as an optimization problem:

$$x^* = \underset{x}{\operatorname{argmin}} L(x; x_0) + \lambda\Omega(x), \tag{2}$$

where $L(x; x_0)$ is the data fidelity term, x_0 is the degraded (Low resolution, obscured, noisy) image, and $\lambda\Omega(x)$ is the regularizer.

The data term $L(x; x_0)$ can be understood as a loss function which maintains consistency between the estimated original image and the degraded image in terms of content. Different data terms have different fitting characteristics, and their selection is usually directly determined by the research task. In the latter part of the optimization problem, the regularizer $\lambda\Omega(x)$ captures the general regularity of natural images, which is used to constrain our model. From a Bayesian prior perspective, when some image information is known, we can greatly simplify the model and improve efficiency. Thus, for regularization restoration methods, the most crucial step is to find a suitable natural image prior. Due to different application fields, the image features that people pay attention to are also different, so it is necessary to construct a reasonable regularization denoising model according to the needs. In fact, over the past few decades, various denoising models based on prior knowledge have been exploited, including gradient models, sparse models and nonlocal self-similarity (NSS) models. In particular, the NSS models are popular in state-of-the-art methods [1, 13, 21, 23, 24].

In recent years, with the widespread research of deep learning in academia and industry, its application advantages in image semantic extraction, feature representation, image generation, and other aspects have become increasingly prominent, making the research of image restoration methods based on deep learning a hot topic. Deep convolutional networks are successful because of their ability to learn from large image datasets. Unlike hand-crafted priors in traditional regularization methods, learning priors is a method to train deep convolutional networks directly from data sets, which takes low-quality images as input and high-quality images as expected output, such as DnCNN [25], FFDNet [26], RDN [28], HINet [3] and NAFNet [2].

Additionally, with the popularization of large models, the accuracy and flexibility of training have been improved. In particular, more robust semantic-level priors are readily available. For instance, a powerful pretrained foundation model - segment anything model (SAM) [12] learns on 1 billion masks and 11 million images with 636 million parameters, resulting in amazing improvements in accuracy and efficiency. Moreover, SAM also has a good performance on degraded images, which makes it possible to combine SAM with image restoration.

In this paper, we review some of the existing tasks that combine SAM and image restoration. We've grouped these algorithms into three categories based on how SAM and tasks combine. Furthermore, we also analyzed and summarized the problems and future trends of SAM. We believe that these investigations can provide the latest findings and advances in image restoration.

2. Preliminaries

2.1. Prior in classical methods

Image restoration tasks are indeterminate, which means a degraded image may correspond to multiple original images. In order to reduce the solution space of the problem, we need to add constraints to better approximate the true solution. In classical image restoration tasks, degraded images are widely tackled based on prior information supplement from experience and knowledge. In the image super-resolution task, the introduction of self-similar prior information adds constraints to the reconstruction process, so as to obtain a good reconstruction effect. Assumptions about prior distributions such as gradient, sparsity, low rank, etc. are exploited, leading to the development of methods such as WNNM [7], MCWNNM [6], wavelet domain processing, and BM3D [5]. Although hand-crafted priors have high restoration quality, it usually has a significant performance degradation in the face of complex optimization problems or severe interference cases. People want to get more universal priors and capture richer statistical data from natural images, so obtaining prior models through deep learning has become a more promising research focus.

2.2. *Prior learned in deep learning methods*

Convolutional neural network has a strong expression ability, and can extract some statistical information implicitly from a large number of data, making it an effective prior in the process of restoring degraded images. For example, Zhang et al. proposed an effective denoising convolutional neural network DnCNN [25] by stacking convolutional layers, batch normalization and rectified linear unit layers, and then gradually proposed FFDNet [26] and CBDNet [8], which showed satisfactory results on both synthetic noise images destroyed by AWGN and real noise images. Other trainable deep priors, such as TNRD [4] and CSF [17], have achieved good results in terms of computational efficiency and denoising effect, but they are very sensitive to the choice of parameters and their performance deteriorates in deeper network models. In addition, some deep priors rely on specific statistics that limit their applicability, such as SinGAN [18]. However, with the advent and development of large-scale models, more general prior information can be easily obtained, thus improving the ability of image restoration.

2.3. *A more robust semantic prior*

Large-scale foundation models have become an important part of artificial intelligence. With advanced training schemes such as self-supervised training, large models can be continuously optimized and scaled from massive amounts of data and eventually achieve zero-shot generalization across a wide range of downstream tasks. The success of large-scale models in natural language processing (NLP) has inspired researchers to explore large-scale models in computer vision (CV). CLIP and ALIGN, for example, combine natural language and images through contrast learning, demonstrating excellent transferability in downstream tasks. After discovering the excellent performance of the foundation models in the NLP and CV domains, Meta recently released a powerful pre-trained foundation model called SAM. SAM utilizes a pre-training algorithm for promptable segmentation tasks that simulates a series of prompts (e.g., points, boxes, masks) for each training sample and compares the model's mask prediction with the ground truth. In addition, SAM's model design is mainly composed of image encoder, prompt encoder and lightweight mask decoder. Given the image and corresponding prompts, they are first encoded by two types of encoders and then decoded by a mask decoder. Intuitively, the pre-training task gives the model the ability to respond to any prompt within the inference time, so the downstream task can be solved by designing the appropriate prompt. More importantly, SAM demonstrates remarkable robustness without compromising image quality, including noisy images and low resolution [9]. This excellent performance allows SAM to be used as prior information for image restoration tasks.

3. How SAM can help

SAM has demonstrated to the public its strong segmentation capability and zero-shot generalization capability, which makes it versatile enough to cover a wide range of application scenarios. SAM's segmentation capabilities can provide semantic-aware prior as a basis for fine-tuning models to better serve low-level tasks. More importantly, SAM has been shown to maintain robustness on degraded images, which allows it to serve as a prior to improve restoration quality [9]. In this section, we review a series of image restoration tasks that apply SAM prior and classify them based on the functionality of SAM prior.

3.1. *SAM as features*

SAM demonstrates powerful generalization capability under zero-shot migration, opening up new possibilities in the field of intelligent image analysis. SAM can obtain all possible object masks by giving a set of prompt points covering a sufficient area. The robustness shown by SAM on degraded images can help improve the accuracy of marking targets in complex scenes, thus generating valuable semantic features. Given a degraded image $I_d \in \mathbb{R}^{H\times W\times C_{in}}$ of an input, SAM can generate a segmentation masks map $M_{sam} \in \mathbb{R}^{H\times W\times N_m}$:

$$M_{sam} = SAM(I_d), \tag{3}$$

where N_m is the number of masks. However, given these masks, specific designs are still needed to make effective use of them. In order to transform M_{sam} into a prior representation P_{sam} with more diverse and rich information, it is necessary to conduct a "feature combination" operation between the input image I_d and the semantic mapM_{sam}:

$$P_{sam} = f_r(I_d \copyright M_{sam}), \tag{4}$$

where © means to concatenate them along the channel dimension, $f_r(\cdot)$ denotes the feature combination network.

Most existing methods choose to design a lightweight plug-in to handle the correlation between prior representation and original features. The plug-in can be inserted into all levels of the network's feature layers according to task requirements. The spatial features and structure of degraded images are enhanced by the mapped semantic information, so that the image restoration network can recover more details in the reconstructed images and obtain high-quality results with good visual perception. Furthermore, the modules are trained in a fine-tuning manner, subtly addressing the balance between performance improvement and computational burden.

3.1.1. SAM prior tuning unit. The focus of image restoration tasks on large-scale foundation models is relatively limited. Xiao et al. [20] for the first time utilized the prior obtained from SAM to perform image restoration tasks. To integrate SAM priors, they propose a lightweight SAM prior tuning (SPT) unit that enhances the low-level features and spatial structure of degraded images by selectively transmitting semantic priors. By taking advantage of the rich semantic information in the mapping, the networks are able to recover more details of ground-truth images during the reconstruction process. As a plug-and-play component, SPT unit can be easily integrated into existing image restoration networks and improve performance, including CNN-based and Transform-based methods. Meanwhile, the SPT unit updates only a few trainable parameters, significantly reducing the computational burden while improving the recovery quality.

Specifically, low-quality images and semantic map M_{SAM} extracted from SAM are input to generate SAM prior representation $\mathcal{P}$. The SPT unit takes in M_{SAM}, $\mathcal{P}$, and the deep feature F_i extracted from the i-th building block as input, and then transmits the data to the feature branch and the SAM prior branch, respectively. Multiply the output features of the two branches to obtain the correlation between F_i and $\mathcal{P}$, and finally output a new feature map F_i^{spt}. To effectively combine this new feature map, it is added to the original feature map F_i using a weighting factor of α. This operation allows subsequent building blocks to take full advantage of SAM prior semantic information. The process can be demonstrated by the following equations:

$$\begin{gathered} \mathcal{P}_{i+1} = f(\mathcal{P}_i) + \mathcal{P}_i, \\ F_i^{spt} = f(F_i^{'}) * f(\mathcal{P}) + F_i^{'}, \\ F_i^n = F_i + \alpha F_i^{spt}, \end{gathered} \tag{5}$$

where $f(\cdot)$ consists of two convolution layers with ReLU activation in between, F_i' is the enhanced feature representation obtained from the input of F_i and M_{SAM}.

Through numerous experiments on image super-resolution and colour image denoising, it can be observed that the SPT unit can help to reconstruct more natural texture details and sharper edges, and the restoration results show better visual quality than the original method. For some unrealistic textures, future research could introduce prior knowledge into existing methods to improve fidelity.

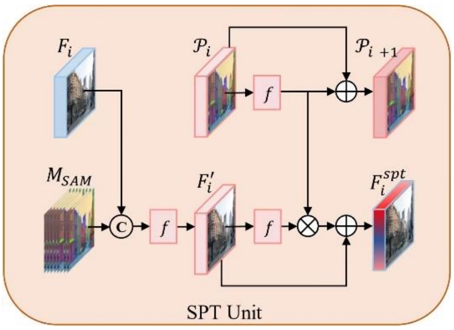

Figure 1. Illustration of the SPT unit proposed in [20].

3.1.2. SAM-guided small dehazing model. Like other image restoration tasks, dehazing tasks are also related to texture, which means different textures will degrade differently under the action of fog. The prior information provided by SAM can guide small networks to better learn the degradation process of textures and edges. Jin et al. [11] found the anti-fog capability of SAM through experiments, that is, by increasing the network parameters, the large model can potentially realize the self-adaptability to a certain degree of fog. To this end, they propose a new idea to help low-level tasks benefit from the development of large models, and successfully demonstrate the effectiveness and applicability of SAM in improving fog removal performance.

Instead of colour segmentation in other projects, they propose grayscale coding, where the output of the segmentation model is a number corresponding to the segmentation of each pixel in the image. This operation can not only clearly understand the order of segmentation, but also reduce the computational complexity of small dehazing model. To enable the model to perceive this grayscale segmentation, the input was expanded to four channels to integrate the advantages of SAM prior into the small model. For SAM-guided small dehazing model, the texture and edge details are recovered better, and the improvement was greater on a foggier dataset, such as NTIRE2020.

3.1.3. SAM-guided refinement module. Similarly, for video super-resolution (VSR) tasks, existing works ignore the valuable semantic information between enhanced frames. In order to make effective use of SAM's robust semantic prior, Lu et al. [14] proposed a simple and effective SAM-guided refinement module (SEEM), which combines SAM prior with the features of corresponding frames to generate semantic-aware features to enhance the alignment and fusion processes between frames. As a lightweight plug-in module, SEEM can be easily integrated into existing methods for versatility and extensibility. In addition, SEEM can be trained in both fully fine-tuning and efficient tuning manners, overcoming limitations of limited storage capacity.

Similar to SPT, SEEM needs to first concatenate the SAM-based prior M_{SAM} with its corresponding frame, and obtain the SAM-based representation R_{SAM} through a feature combination network consisting of two convolutional layers and one residual block. Then, the mapped feature F_m forward through the channel attention block (CAB) to adaptively readjust the scaling of channel-wise attention features F_a. Fianlly, SEEM outputs the semantic-aware feature. The process can be demonstrated by the following equations:

$$\begin{gathered} F_{cab} = F_a \odot F_m, \\ F_{seem} = F_m + F_{cab}, \end{gathered} \tag{6}$$

SEEM is used to improve the performance of EDVR based on sliding window and BasicVSR based on bidirectional recurrent, respectively. It is evident that SEEM significantly boosts the visual quality of baseline methods compared to many existing state-of-the-arts. However, SEEM occasionally produces artifacts that are non-existent in high-resolution frames, and there is a slight decrease in performance on fast-motion videos.

3.1.4. Enlighten Anything. In low-light image enhancement tasks, many unsupervised methods do not take into account the degradation of visible information in low-light scenes, which seriously affects the aggregation of complementary information. Moreover, the mainstream supervised learning methods require a large amount of pairwise training data, which may lead to overfitting. Taking advantage of SAM's high recognition accuracy in low-light environments and zero-shot learning, Zhang et al. [29] proposed Enlighten Anything (EA), which can enhance and fuse the semantic priors after SAM segmentation with low-light images to obtain fusion images with good visual perception.

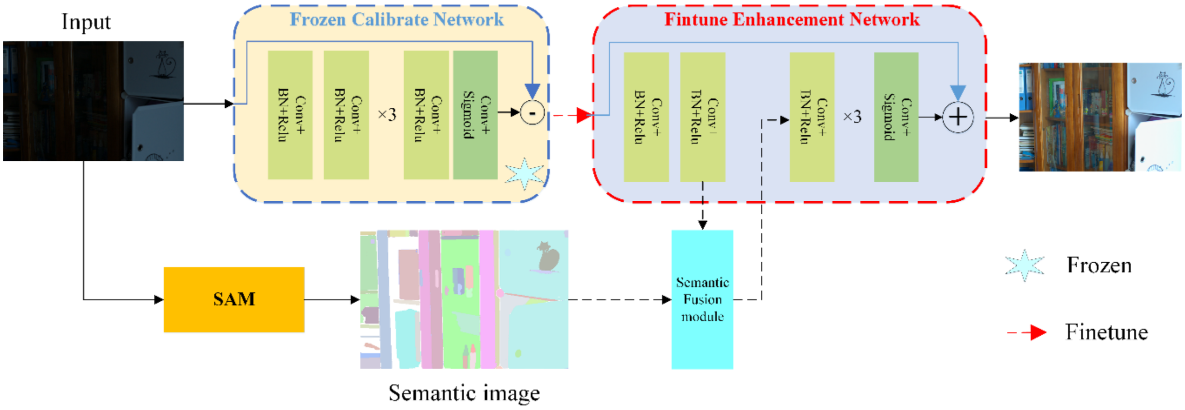

Figure 2. The structure of Enlighten Anything proposed in [29].

The structure of EA is divided into three parts, which are self-calibration module, semantic fusion module and enhancement module. To better train the framework, an unsupervised loss function is designed to constrain light estimates at each stage. Compared with the existing SOTA methods, EA's evaluation indicators are significantly improved, and the colour is more prominent in terms of visual effects. In the future, EA can increase complexity to deal with more realistic images.

3.2. SAM as an interactive component

Advances in large language models have inspired a great deal of interest in the development of fundamental models for computer vision. SAM represents a paradigm shift in image segmentation, moving away from specific types of tasks to a more flexible, generalizable model. SAM allows users to interactively guide segmentation with prompt-based input (e.g., points, boxes, masks, text). This high-quality interaction comes from a new segmentation method - a pre-trained promptable segmentation algorithm. Intuitively, the pre-training task gives the model the ability to respond to any prompt within the inference time, so the low-level tasks can be solved by designing the appropriate prompt. As a result, SAM's excellent interactivity provides users with more possibilities and versatility in many single-result restoration tasks.

3.2.1. Inpainting Anything. In order to solve the challenge of holes filling in modern image inpainting systems, Yu et al. [22], based on SAM, proposed a mask-free image inpainting called Inpaint Anything (IA). By clicking on an object in the image, SAM automatically segments and creates a mask, followed by a state-of-the-art inpainter, such as LaMa, to fill in the "holes" with contextual data. Aiming at the non-smooth boundary of the object mask, IA uses dilation operation to refine the mask. In addition, through powerful AI-Generated content (AIGC) models, IA provides a user-friendly pipeline that supports text prompts to fill in or replace any content.

Experiments show that IA can effectively inpaint images with different content, resolution and aspect ratio, which is both general and robust. In the future, IA may also support more practical functions and be used in more realistic applications.

3.2.2. Restore Anything Pipeline. Great progress has been made in image restoration methods based on deep learning. However, existing methods tend to treat the entire image as a single entity, ignoring the semantic differences in the image. In addition, most existing methods produce only a single result, which may not suit the preferences of different users. By integrating SAM into a controllable image restoration model, Jiang et al. [10] proposed Restore Anything Pipeline (RAP), a novel interactive image restoration method. By clicking on the desired object, RAP supports the use of different restoration levels to independently process objects with different texture details or different degradation types, avoiding the generation of generic results. In terms of user friendliness, RAP utilizes a flexible blind restoration framework in which the encoder extracts and predicts the degradation parameters of damaged images. Users can automatically obtain the recovery results of decoder output, and can also achieve precise control of output results by manually adjusting parameters.

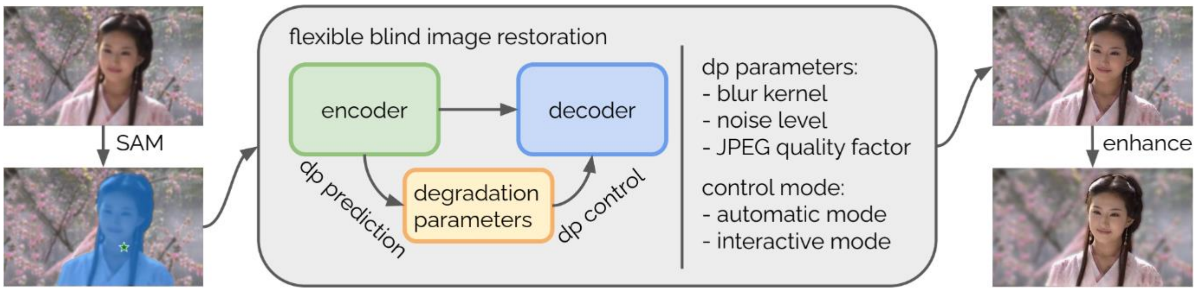

Figure 3. Restore Anything Pipeline (RAP) proposed in [10].

The versatility of RAP is demonstrated through three common restoration tasks: JPEG artifact removal, denoising, and image deblurring. Compared to the SOTA methods, RAP produces superior visual effects by restoring the foreground and background at different levels. RAP represents a promising direction in image restoration, providing users with greater control and enabling object-level image restoration.

3.3. *Medical segmentation with SAM*

Current medical image segmentation methods are often designed and trained for specific tasks, and the performance of these models can significantly degrade in new tasks. Therefore, a versatile foundation model is needed to address a wider range of applications in clinical practice. There is a lot of work that has validated the impact of SAM on medical image datasets. It was concluded that SAM performed well overall but had a higher failure rate when faced with dense small targets or bends [16]. Therefore, further work is still needed to optimize and improve the performance of SAM in medical image segmentation tasks.

3.3.1. SAM in medical images. MedSAM. By fine-tuning SAM using labelled images, Ma et al. [15] propose MedSAM, a foundation model used to segment a wide range of anatomical and pathological regions in different medical modalities. Through thorough evaluation, MedSAM has been shown to not only outperform existing state-of-the-art segmentation foundation models, but to compete with and even surpass professional models. Note that MedSAM shows great potential in specific tasks, such as tumor segmentation. By enabling accurate and efficient segmentation across a wide range of tasks, MedSAM has great potential to accelerate the personalization of treatment plans.

3.3.2. Medical SAM Adapter. Unlike the fully fine-tuning model, Wu et al. [19] propose the Medical SAM Adapter (MSA), which fine-tunes the pre-trained SAM through a technique called adaptive Parametric Efficient fine-tuning (PEFT), thus integrating medical specific domain knowledge into the segmentation model. Specifically, it works by inserting several parameter-efficient Adapter modules into the original model, and then adjusting only the Adapter parameters. This simple and efficient approach enables SAM to significantly advance in medical image segmentation and achieve state-of-the-art performance across a range of segmentation tasks, even surpassing fully fine-tuned MedSAM.

3.3.3. Input augmentation with SAM. SAMAug. Although SAM is not robust enough for medical image segmentation, Zhang et al. [27] found that SAM is helpful for medical image segmentation models as semantic prior. To this end, they propose a novel method called SAMAug, which directly uses the segmentation masks generated by SAM to enhance the original input of the model. Unlike traditional enhancement methods that work at low level, SAMAug aims to add a high-level structure to the original image to provide better semantics for subsequent medical image segmentation. In addition, compared to fine-tuned MedSAM and self-adaptive MSA, SAMAug has more efficient computation and memory overhead during training.

Specifically, firstly, the segmentation prior map $prior_{seg}$ and the boundary prior map $prior_{boundry}$ are drawn on the newly created segmentation prior map according to the masks provided by SAM, and the two are added to the second channel and the third channel of the original image x respectively (the first channel is the grayscale original image). Therefore, a new training set is obtained by input enhancement. Furthermore, in order to deal with the complex and changeable medical image data, SAMAug designed two parameter update functions and two model deployment strategies to fully utilize the semantic value of prior maps.

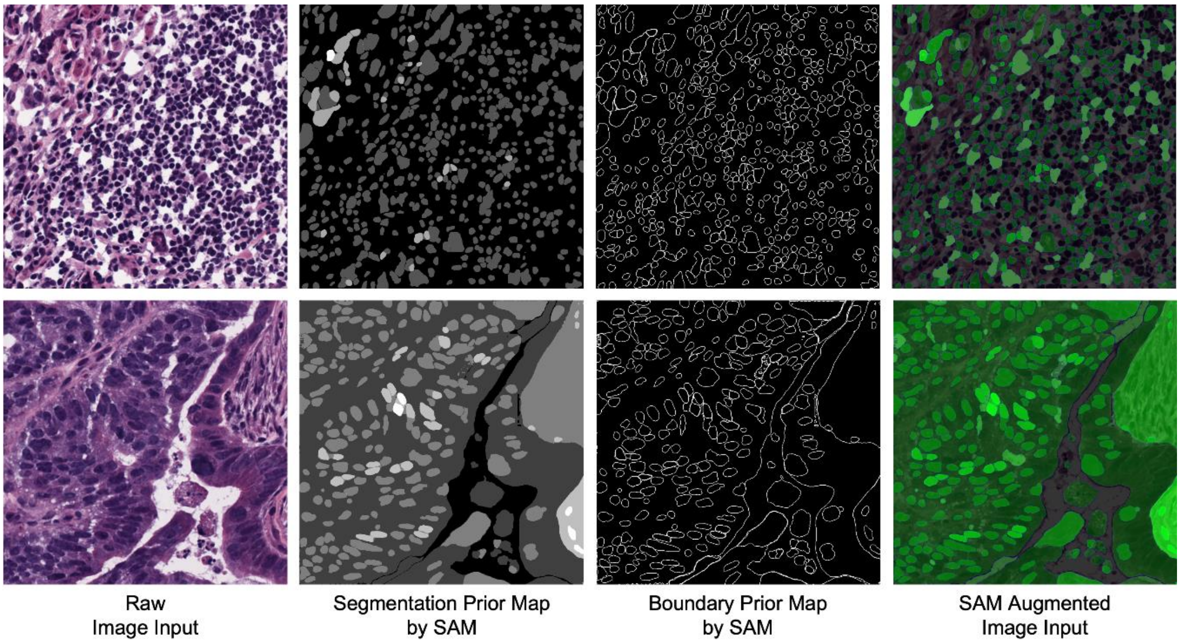

Figure 4. Visual examples of images augmented by SAMAug proposed in [27].

The experiments show that SAMAug significantly improves the results of the original medical image segmentation model. While SAM occasionally fails to give an accurate segmentation immediately, the masks and features it generates provide perceptual priors for subsequent models to generate more accurate task-specific segmentation results. In the future, SAMAug has the opportunity to make further improvements in efficiency and clinical application.

4. Conclusion and possible future directions

In this paper, we systematically review the current application of SAM in image restoration, and make an innovative summary and classification of how SAM, as a robust semantic prior and fine-tuning capability, is deployed to the training of deep networks. However, there are open questions and challenging issues awaiting further inquiry.

- The model design of SAM is relatively large, especially the high-quality model needs more computing power, which is an extra cost. If SAM can be deployed through model pruning or other miniaturization techniques in the future, SAM will become more prevalent in the deployment of downstream tasks.
- The quality of the masks generated by SAM involves multiple parameters. In pixel-level restoration tasks, the quality of the masks can seriously affect the results. Therefore, in order to generate accurate and efficient masks, the relevant parameters need to be determined by additional experiments.
- Although SAM showed robustness on degraded images, it often performed poorly on some low-quality images, especially those involving specialized data (e.g., medical lesion, industrial defect). The results may have improved with the help of the prompts, but the reliability of SAM still needs further study.

Overall, with robust and accurate segmentation capabilities, the semantic awareness provided by SAM can improve the performance of low-level models to some extent. We believe that our summary can provide some insights to guide future research to further improve the role of SAM in low-level tasks.

References

[1] Chen H, He X, Teng Q and Ren C. Single image super resolution using local smoothness and nonlocal self-similarity priors[J]. *Signal Processing: Image Communication*, 2016, 43: 68-81.

[2] Chen L, Chu X, Zhang X and Sun J. Simple baselines for image restoration[C]. 2022 *European Conf. on Computer Vision* (Springer Nature Switzerland) pp 17-33.

[3] Chen L, Lu X, Zhang J, Zhang J, Chu X and Chen C. Hinet: Half instance normalization network for image restoration[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 182-192.

[4] Chen Y and Pock T. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2016, 39(6): 1256-1272.

[5] Dabov K, Foi A, Katkovnik V and Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering[J]. *IEEE Transactions on image processing*, 2007, 16(8): 2080-2095.

[6] Gu S, Xie Q, Meng D, Zuo W, Feng X and Zhang L. Weighted nuclear norm minimization and its applications to low level vision[J]. *International journal of computer vision*, 2017, 121: 183-208.

[7] Gu S, Zhang L, Zuo W and Feng X. Weighted nuclear norm minimization with application to image denoising[C]. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014: 2862-2869.

[8] Guo S, Yan Z, Zhang K, Zuo W and Zhang L. Toward convolutional blind denoising of real photographs[C]. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019: 1712-1722.

[9] Huang Y, Cao Y, Li T, Juefei-Xu F, Lin D, Tsang I, Liu Y and Guo Q. On the robustness of segment anything[J]. *arXiv preprint arXiv*: 2305.16220, 2023.

[10] Jiang J and Holz C. Restore Anything Pipeline: Segment Anything Meets Image Restoration[J]. *arXiv preprint arXiv*: 2305.13093, 2023.

[11] Jin Z, Chen S, Chen Y, Xu Z and Feng H. Let Segment Anything Help Image Dehaze[J]. *arXiv preprint arXiv*: 2306.15870, 2023.

[12] Kirillov A, et al. Segment anything[J]. *arXiv preprint arXiv*: 2304.02643, 2023.

[13] Liu H, Guo Q, Wang G, Gupta B and Zhang C. Medical image resolution enhancement for healthcare using nonlocal self-similarity and low-rank prior[J]. *Multimedia Tools and Applications*, 2019, 78: 9033-9050.

[14] Lu Z, Xiao Z, Bai J, Xiong Z and Wang X. Can SAM Boost Video Super-Resolution?[J]. *arXiv preprint arXiv*:2305.06524, 2023.

[15] Ma J and Wang B. Segment anything in medical images[J]. *arXiv preprint arXiv*:2304.12306, 2023.

[16] Mazurowski M A, Dong H, Gu H, Yang J, Konz N and Zhang Y. Segment anything model for medical image analysis: an experimental study[J]. *Medical Image Analysis*, 2023, 89: 102918.

[17] Schmidt U and Roth S. Shrinkage fields for effective image restoration[C]. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014: 2774-2781.

[18] Shaham T R, Dekel T and Michaeli T. Singan: Learning a generative model from a single natural image[C]. *Proceedings of the IEEE/CVF international conference on computer vision*, 2019: 4570-4580.

[19] Wu J, Fu R, Fang H, Liu Y, Wang Z, Xu Y, Jin Y and Arbel T. Medical sam adapter: Adapting segment anything model for medical image segmentation[J]. *arXiv preprint arXiv*:2304.12620, 2023.

[20] Xiao Z, Bai J, Lu Z and Xiong Z. A dive into sam prior in image restoration[J]. *arXiv preprint arXiv*:2305.13620, 2023.

[21] Xu J, Zhang L, Zuo W, Zhang D and Feng X. Patch group based nonlocal self-similarity prior learning for image denoising[C]. *Proceedings of the IEEE international conference on computer vision*, 2015: 244-252.

[22] Yu T, Feng R, Feng R, Liu J, Jin X, Zeng W and Chen Z. Inpaint anything: Segment anything meets image inpainting[J]. *arXiv preprint arXiv*:2304.06790, 2023.

[23] Zeng H, Xie X, Kong W, Cui S and Ning J. Hyperspectral image denoising via combined non-local self-similarity and local low-rank regularization[J]. *IEEE Access*, 2020, 8: 50190-50208.

[24] Zha Z, Yuan X, Zhou J, Zhu C and Wen B. Image restoration via simultaneous nonlocal self-similarity priors[J]. *IEEE Transactions on Image Processing*, 2020, 29: 8561-8576.

[25] Zhang K, Zuo W, Chen Y, Meng D and Zhang L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising[J]. *IEEE transactions on image processing*, 2017, 26(7): 3142-3155.

[26] Zhang K, Zuo W and Zhang L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising[J]. *IEEE Transactions on Image Processing*, 2018, 27(9): 4608-4622.

[27] Zhang Y, Zhou T, Liang P and Chen D Z. Input augmentation with sam: Boosting medical image segmentation with segmentation foundation model[J]. *arXiv preprint arXiv*:2304.11332, 2023.

[28] Zhang Y, Tian Y, Kong Y, Zhong B and Fu Y. Residual dense network for image restoration[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2020, 43(7): 2480-2495.

[29] Zhao Q, Zhang X, Tang H, Gu C and Zhu S. Enlighten-anything: When Segment Anything Model Meets Low-light Image Enhancement[J]. *arXiv preprint arXiv*:2306.10286, 2023.

[30] Rasti B, Chang Y, Dalsasso E, Denis L and Ghamisi P. Image restoration for remote sensing: Overview and toolbox[J]. *IEEE Geoscience and Remote Sensing Magazine*, 2021, 10(2): 201-230.

[31] Muhammad N A and Fatima R. Role of image processing in digital forensics and cybercrime detection[J]. *International Journal of Computational and Innovative Sciences*, 2022, 1(1): 39-42.