Highlights

Let Segment Anything Help Image Dehaze

Zheyan Jin,Shiqi Chen,Yueting Chen,Zhihai Xu,Huajun Feng

- We discovered and demonstrated the emergence of anti-fog capabilities in large-scale image segmentation models, which were not innately present in the dataset or training process but are achieved through large-scale datasets and large-scale models.

- Through grayscale coding and channel expansion, we propose a new framework for transferring the advantages of the large model to the low-level visual dehaze task engaging with small-scale data and small models, which accelerate the adaptation of specific dehaze results.

- We carry out a comprehensive experiment for evaluating the proposed method and comparing the impact of different model sizes on the final dehaze results under different fog scenarios.

Let Segment Anything Help Image Dehaze⋆,⋆⋆

Zheyan Jin[a,*,1], Shiqi Chen[a], Yueting Chen[a,2] (Co-ordinator), Zhihai Xu[a] and Huajun Feng[a]

[a]*Zhejiang University, West Lake District 38 Zhejiang University Road, HangZhou, 315000, China*

ARTICLE INFO

ABSTRACT

The large language model and high-level vision model have achieved impressive performance improvements with large datasets and model sizes. However, low-level computer vision tasks, such as image dehaze and blur removal, still rely on a small number of datasets and small-sized models, which generally leads to overfitting and local optima. Therefore, we propose a framework to integrate large-model prior into low-level computer vision tasks. Just as with the task of image segmentation, the degradation of haze is also texture-related. So we propose to detect gray-scale coding, network channel expansion, and pre-dehaze structures to integrate large-model prior knowledge into any low-level dehazing network. We demonstrate the effectiveness and applicability of large models in guiding low-level visual tasks through different datasets and algorithms comparison experiments. Finally, we demonstrate the effect of grayscale coding, network channel expansion, and recurrent network structures through ablation experiments. Under the conditions where additional data and training resources are not required, we successfully prove that the integration of large-model prior knowledge will improve the dehaze performance and save training time for low-level visual tasks.

1. Introduction

Image dehazing is one of the important computer vision tasks, which removes the haze interference by using image restoration algorithms, allowing for better subsequent calculations. As the research on dehazing algorithms continues to advance, such as the appearance of complex scenes with thick haze, non-uniform haze, and complex lighting conditions, which small-sized models are difficult to handle well. The mainstream progress in dehazing algorithms is mainly based on transformer models with enhanced parameter counts Song, He, Qian and Du (2022) and image-domain adaptation methods Yi, Ma, Zhang, Liu and Wu (2022) respectively. These algorithms often require a large number of datasets. However, dehazing models often lack reliable real-world data sets, and it is also difficult to separately train different large models for different dehazing tasks.

With the continuous development of large-scale models, the emergence of many abilities beyond data itself in large language and large segmentation models. These abilities have been achieved through the joint improvement of data quantity and network scale. However, image dehazing can not achieve significant improvements through large-scale

and high-quality model training. It is difficult for the low-level vision dehazing model to benefit from the large model with large data size and continuous rapid development. Therefore, we hope that image dehazing networks based on small models and small data can be improved through the large image segmentation models. This allows large models to enhance the ability of image dehazing.

We have discovered an emergent self-adaptive ability of large-scale image segmentation models in the image domain. Even if the training dataset does not contain images specifically for fog, by increasing the network parameter, the large model can quickly compensate for the performance impact of fog on segmentation. Therefore, for various fog scenarios, we use different parameter-count large-scale segmentation models to guide the encoding image dehazing model. And through the method of grayscale coding and channel expansion, the small dehazing network can learn the dehazing ability of the large segmentation model. The above content enables the application of powerful generalization anti-haze capabilities of large models in small dehaze networks.

Our main innovations and contributions are as follows:

- We discovered and demonstrated the emergence of anti-fog capabilities in large-scale image segmentation models, which were not innately present in the dataset or training process but are achieved through large-scale datasets and large-scale models.

- Through grayscale coding and channel expansion, we propose a new framework for transferring the advantages of the large model to the low-level visual dehaze task engaging with small-scale data and small models, which accelerate the adaptation of specific dehaze results.

- We carry out a comprehensive experiment for evaluating the proposed method and comparing the impact of different model sizes on the final dehaze results under different fog scenarios.

⋆⋆The second title footnote which is a longer text matter to fill through the whole text width and overflow into another line in the footnotes area of the first page.

This note has no numbers. In this work we demonstrate a_b the formation Y_1 of a new type of polariton on the interface between a cuprous oxide slab and a polystyrene micro-sphere placed on the slab.

*Corresponding author
**Principal corresponding author
✉ 11930051@zju.edu.cn (Z. Jin); chenyt@zju.edu.cn (Y. Chen); xuzh@zju.edu.cn (Z. Xu); fenghj@zju.edu.cn (H. Feng)
ORCID(s): 0000-0001-8466-7520 (Z. Jin)

[1]This is the first author footnote. but is common to third author as well.
[2]Another author footnote, this is a very long footnote and it should be a really long footnote. But this footnote is not yet sufficiently long enough to make two lines of footnote text.
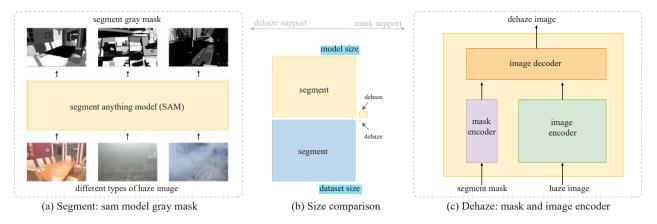

(a) Segment: sam model gray mask   (b) Size comparison   (c) Dehaze: mask and image encoder

Figure 1: Main pipeline. (a) Segment model: Put haze image input into a large-scale segmentation model, and output the segmentation result in grayscale encoding. By utilizing the emergence capability of large models, it can also handle haze images that have not been trained before. **(b) Size comparison:** Both in terms of network size and dataset size, the largest existing dehazing model and segmentation large model are several orders of magnitude apart. Therefore, the emergence transparency ability of large models and the image domain translation ability of the dehazing model can help each other. **(c) Dehaze:** grayscale encoded segmentation mask is added to the encoder part of the image dehazing network through the encoder-decoder structure.

2. Related Works

2.1. Image dehazing

Image dehazing is a low-level computer vision task. It aims to eliminate the negative effects of haze scattering on images. Tang, Yang and Wang (2014) used a random forest regression model to estimate the amount of haze. He (2009) was the first to use a dark channel prior to removing haze. Deep learning-based dehazing methods can be divided into two main categories: calculation of intermediate parameters and direct end-to-end training. The former estimates intermediate parameters and then inputs them into an atmospheric degradation model to calculate the final clean image. Later models tend to learn the mapping from hazy to clean images directly. Cai, Xu, Jia, Qing and Tao (2016) presented a full convolution neural network (CNN) called DehazeNet for image dehazing. The model accepts hazy images as input and produces a transmission map as its output. Ren, Pan, Zhang, Cao and Yang (2020) proposed a multi-scale deep neural network to estimate the transmission value. Chen, He, Fan, Liao and Hua (2019) proposed a threshold fusion network that utilizes a generative adversarial network (GAN) for image dehazing, solving the common unreal unreality issue. Song et al. (2022) applied a larger parameter-count transformer structure to the dehazing field and achieved better results. Zheng, Zhan, He, Dong and Du (2023) used the domain changes between different hazy images from the same dataset for image dehazing.

Compared to normal image dehazing, the thick haze, non-uniform haze, and night scene haze are more complex, and the research started later. Ancuti, Ancuti, Sbert and Timofte (2019) released a real-world dataset of thick haze outdoors. Ancuti, Ancuti and Timofte (2020) released a real-world dataset of non-uniform haze outdoors. Jing, Yang and Wang (2014) proposed the NDIM algorithm, which performs colour correction after estimating the colour

features of the incident light. Yu, Tan and Brown (2015) distinguished between atmospheric light, haze light, glow light, and different types of sources of light, and proposed an algorithm based on glow special processing and night-time different source recognition. Ancuti, Ancuti, Vleeschouwer and Bovik (2016) proposed a multi-scale artificial light patch pyramid network to adapt to night-time haze environments. Jing, Yang, Shuai, Yu and Chang (2017) believed that the local maximum intensity of each colour channel in night-time images was mainly contributed by environmental lighting, and proposed the MRP algorithm with maximum reflection first-order prior. Zhang, Cao, Zha and Tao (2020) based on scene geometry, then simulated the light and object reflection rates in two dimensions. A new method and benchmark testing method for haze rendering images were proposed.

2.2. Segmentation and large language model

Unlike image dehazing, which is a low-level task in computer vision, semantic segmentation is a high-level task in computer vision to classify each pixel in an image into a particular class or object. The goal is to generate a dense pixel segmentation map, where each pixel is assigned to a specific class or object. Some examples of benchmark datasets for this task include Cityscapes Cordts, Omran, Ramos, Rehfeld, Enzweiler, Benenson, Franke, Roth and Schiele (2016), PASCAL VOC Everingham, Van Gool, Williams, Winn and Zisserman (2010), and ADE20K Zhou, Zhao, Puig, Fidler, Barriuso and Torralba (2017). Models are typically evaluated using metrics such as average intersection over union (average IoU) and pixel accuracy measures.

Recently, some studies have explored the relationship between image coding models and large language models. Pretrained large language models on massive datasets available online have revolutionized the generalization of natural language processing (NLP) Brown, Mann, Ryder, Subbiah, Kaplan, Dhariwal, Neelakantan, Shyam, Sastry, Askell et al.

(2020a). These models can generalize to properties beyond the distribution of the data. These models can even defeat end-to-end trained fine-tuning models with a certain probability when there is a lack of dataset. Empirical trends indicate that this behavior improves as the size of the model increases, and the size of the dataset and the scale of the training model also significantly affect the performance of the model. Aligning pairs of text and images online is an excellent work, for example, CLIP Radford, Kim, Hallacy, Ramesh, Goh, Agarwal, Sastry, Askell, Mishkin, Clark et al. (2021) and ALIGN Jia, Yang, Xia, Chen, Parekh, Pham, Le, Sung, Li and Duerig (2021) use contrastive learning to train text and image encoders that align both pairs. After improving, such encoding methods can generalize to zero-sample scenarios for new image contexts and data. Such coding methods also effectively collaborate with other modules that enable fundamental image tasks, such as image generation by DALL·E Ramesh, Pavlov, Goh, Gray, Voss, Radford, Chen and Sutskever (2021).

2.3. The haze effects on semantic segmentation

There have been previous studies that focus on the relationship between image dehazing and image segmentation. Haze can cause degradation in the segmentation capabilities of existing models.

Some research focuses on the segmentation performance of the network can be improved after image dehazing. After passing the dehazing pipeline, the distribution of the image is closer to the distribution of the normal segmentation training dataset, thereby improving segmentation and detection capabilities. Li, Peng, Wang, Xu and Feng (2017) found that dehaze images provided better detection and segmentation results.

Other researchers hope to adapt the data environment of fog conditions by directly improving detection and segmentation networks. Lee, Son and Kwak (2022) proposed a region mapping filter to improve the segmentation capabilities of images in haze-contaminated scenes. Sakaridis, Christos, Dai, Dengxin, Gool and Luc (2018) and Hahner, Dai, Sakaridis, Zaech and Van Gool (2019) respectively proposed a method of training segmentation networks in haze-contaminated scenes by synthesizing haze-contaminated pipelines.

3. Methodology

3.1. Large model can overcome haze degradation

We believe that large-scale models can emerge with unexpected capabilities. For example, Brown, Mann, Ryder, Subbiah and Amodei (2020b) generative pre trained transformer (GPT) trained extensively can perform text interaction such as dialogues. Due to their sufficient parameters and wide distribution of image color and contrast participation, large-scale segmentation networks can potentially achieve some degree of domain adaptability to haze. As the previous paragraph mentioned, other researchers have used methods such as image dehaze or domain adaptation to achieve the

same results. However, the anti-haze capabilities of large models are automatically obtained, rather than manually optimized. The existing segmentation large models and dehaze models exhibit several orders of magnitude differences in their model parameters and training datasets. In terms of image numbers, the dataset size of segmentation large models is four orders of magnitude larger than that of dehaze models. Meanwhile, the parameter size of segmentation large models is two orders of magnitude larger than that of dehaze networks. The detailed comparison is shown in Fig.1. We use the segment anything model (SAM) Kirillov, Mintun, Ravi, Mao, Rolland, Gustafson, Xiao, Whitehead, Berg, Lo et al. (2023) as our large segmentation model. Its training data volume exceeds 1 billion masks, and network parameters are close to 3GB, and the occupied video memory is close to 50GB when tested on the dehaze dataset. If large models can perform precise segmentation under haze conditions, this can assist small dehaze networks in various types of haze.

3.2. Haze, image texture and segmentation

The traditional dehaze formula is as follows,

$$I(x) = R(x)t(x) + L(x)(1 - t(x)), \quad (1)$$

where x is the position of the pixel, $I(x)$ is the signal received by the camera pixel, $R(x)$ is the signal emitted by the object itself, $L(x)$ is the atmospheric global illumination, and $t(x)$ is a transmission rate. The transmission rate formula is as follows,

$$t(x) = e^{-\beta \cdot d(x)}, \quad (2)$$

where $d(x)$ is the distance from the object to the camera, β is the attenuation coefficient, and e shows that the attenuation is in exponential.

However, many haze images are not homogeneous and the transmittance is not linearly related to depth, the $t(x)$ is more complex. Additionally, $L(x)$ is influenced by the properties of the fog and the light. The variable environment of fog makes Eq.2 often fail, and it is difficult to directly calculate $R(x)$ from the input of the sensor $I(x)$ in a linear manner.

If a large model can output precise segmentation results, Eq.(1) can be transformed into Eq.(3).

$$I(x) = R_{mask}(x)R_{mask2texture}(x)t(x) + L(x)(1 - t(x)), \quad (3)$$

where the $R_{mask}(x)$ represents the segmentation mask output by a large-scale segmentation model under ideal conditions, while $R_{mask2texture}(x)$ represents the transition matrix from the segmentation mask to the no haze image. Since we ultimately need $R(x)$, once we have the guidance of $R_{mask}(x)$, it will be easier to fit and calculate $R_{mask2texture}(x)$ using Eq.(3). Then calculate $R(x)$ using Eq.(4),

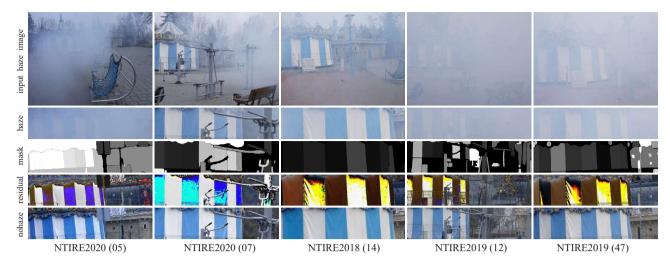$$R(X) = R_{mask}(x)R_{mask2texture}(x). \quad (4)$$

Figure 2: The relationship between segmentation and texture in haze images. The figure displays different types of haze scenarios, input haze images, segmentation masks, residuals, and no haze images. The segmentation result can separate similar haze degradation areas under different haze conditions. The numbers in parentheses in the figure indicate the image position of the dataset.

In addition to relying on the dehaze formula, we can also subjectively understand the relationship between image segmentation and image texture degradation under fog conditions. As shown in Fig.2, different objects often have different textures, and different textures will suffer from different degeneration under the effect of fog. The segmentation mask provides texture edge information and segments similar degeneration parts together, thus better guiding small networks to learn similar degeneration processes.

3.3. Grayscale coding of segmentation results

The output of the segmentation model is a digital number corresponding to the segmentation of each pixel in the image. In other segmentation projects, the segmentation result is often visualized by converting the image to color, with different colors used to represent different types of segments in an image. Converting segments in an image to color requires three channels, which increases the computational complexity of subsequent small models for dehazing. Therefore, we propose to combine segmentation on a grayscale channel as shown in Fig.1(a). Since large segmentation models can segment all objects in the image, there often exist many objects in an image. As a result, an image is often divided into many parts in segmentation result. According to our experiments on a dehazing dataset, an image can be segmented into more than 130 parts at most, while the majority of images have a segmentation number between 30 and 127 (half of 255). We propose a grayscale coding method, shown in Algorithm 1, which converts the segmentation result into a grayscale image. There are several advantages to this method. First, try to fill the 1-255 grayscale space (0 is the result of no segmentation). From dark to bright, it can be clearly understood the output order of each segmentation result. There are brighter areas in the segmentation results of the same image, indicating more segmentation, and the network performance is better. Using grayscale coding helps

---

Algorithm 1: Grayscale coding of segmentation

Input: The output results of the segmentation network *masks*, including segmentation types *id* and segmentation pixel distribution data *area*.

Output: Grayscale coding segment mask *segmask*

for *id, area* in **enumerate(masks)** **do**
  area becomes a matrix with a segmentation result of 1;
  if *id < 127* **then**
    $segmask = segmask + area \times 2 \times (id + 1)$;
  end
  else
    $segmask = segmask + area \times (2 \times (255 - id) - 1)$;
  end
end
segmask from matrix to single channel image;
return *segmask*;

---

to feel the segmentation effect more subjectively, as shown in Fig.4.

3.4. Image dehaze model with segmask

Common dehazing models typically have an input and output of three channels. The input fog image has RGB three channels, and the output is the dehazed image has RGB three channels. To enable the model to perceive grayscale encoded segmentation masks, we need to expand the input channels to four channels by adding a grayscale channel and putting the segmentation masks in the new grayscale channel. The output channels remain at three channels. This causes the encoder part of the network to have an additional channel and expand in size, while the decoder part remains unchanged,
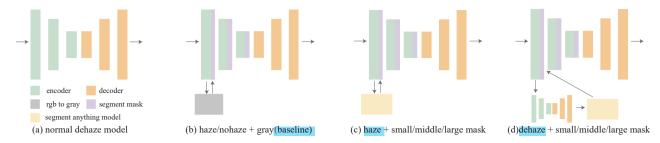
Figure 3: Network structure of image dehaze model with mask. The figure shows the network structure of different methods compared in our experiments. Each color block of the network structure has a one-to-one correspondence with Fig.1.

as shown in Fig.1(c). It is important to note that the two encoders are not separate, but rather have different channels. This expansion channel method can be used by most dehazing networks, but after increasing the number of channels, the parameters increase, and the original network structure and the improved structure have different parameters, making direct comparison unfair. To remove the impact of parameter changes, we generated grayscale images from the RGB channels into the new grayscale channel as the baseline. Since the concentration and distribution of haze are not uniform, the difficulty of different dehazing datasets is different, and this baseline also serves as a baseline for evaluating the difficulty of the dataset itself. We placed the fourth channel of the network's encoder into the following types, all corresponding network structures are shown in Fig.3.

We will carry out the following types of dehazing network structures with different masks, where gray refers to the grayscale image from the RGB image. Among them, small/middle/large refers to different model sizes in the SAM model. Large refers to vit_h, which is the largest model of the SAM model. Middle refers to vit_l, medium scale. Small refers to vit_b, which is the smallest model in SAM.

- **haze gray**: Baseline. The grayscale of the haze image is directly sent to the network as a segmask.

- **nohaze gray**: Directly send the grayscale image of the fog-free image as segmask to the network. Is the absolute ideal ceiling for experiments. Used to evaluate data and network capabilities.

- **haze + small/middle/large**: After the fog image is segmented by different size models(small/middle/large), it is sent to the network as a segmask.

- **nohaze + small/middle/large**: After the no-haze image is segmented by different size models, it is sent to the network as a segmask. Used in evaluating segmentation network performance and dehazing training.

- **dehaze + small/middle/large**: Dehazing first, and the result of dehazing image segmentation by different size models is sent to the network as a segmask. Used in the actual dehazing test.

4. Data Preparation

There are many existing dehaze datasets available, and we need to choose data that covers various scenarios such as indoor and outdoor, thick fog, thin fog, and non-uniform fog. This will allow us to test the effectiveness of our method under different fog conditions. At the same time, we should strive to use real-shot datasets as much as possible to achieve a closer match to real-world scenarios. We have chosen RESIDE to represent simulated thin fog outdoors, and RE-VIDE Zhang, Dong, Pan, Zhu, Tai, Wang, Li, Huang and Wang (2021) to represent thick fog indoors. NTIRE2018I Ancuti, Ancuti, Timofte and Vleeschouwer (2018a) and NTIRE2018O Ancuti, Ancuti, Timofte and Vleeschouwer (2018b) represent real-world fog captured indoors and outdoors, while NTIRE2019 Ancuti et al. (2019) represents outdoor thick fog and NTIRE2020 Ancuti et al. (2020) represents outdoor non-uniform haze.

Due to the existence of different types of haze, there is a relationship between the deterioration of different haze types. We analyze the guidance ability of large segment models for small dehaze models based on different types of dehaze data. The dehaze result in different fog scenarios can show the relationship between the degradation of different fog types and the segmentation mask.

To compare the impact of non-uniform haze and different haze concentrations on the segmentation network, we used Unreal Engine 5 EPIC (2022b) to render and generate different concentrations of fog to evaluate the segmentation ability of the network under haze conditions. The simulation images were rendered based on the volume fog EPIC (2022a), and the strength of the fog concentration varied with the intensity of the volume fog. As the fog concentration increased, the image became increasingly polluted.

5. Experimental Assessment

5.1. The haze effects on segmentation

The general detection and segmentation datasets do not use datasets with haze for training. The dataset used for the SAM large model training does not have any data related to haze. Therefore, the degradation caused by haze will have a negative impact on the performance of the segmentation network.
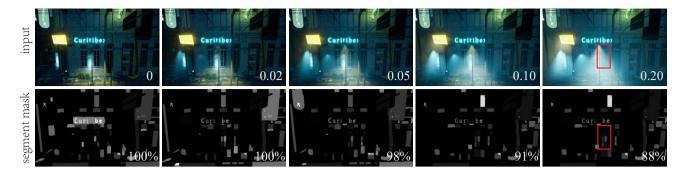
Figure 4: Different concentrations of haze and the performance of segmentation networks. The images in the top row are input images with haze, and the numbers in the lower right corner are the fog concentration. The images in the bottom row are gray scale masks obtained after passing the segmentation network, and the numbers in the lower right corner are the percentage of segmentations compared to the no-haze case. As the fog concentration increases, the order of segmentation changes, and the mask that is segmented last is brighter. It is worth noting that in the red box, even if the fog is not evenly distributed, the model can still correctly segment each window position. Large-model segmentation networks can provide texture guidance for image dehazing.

Table 1
Quantitatively compare the segmentation performance of models of different sizes under different datasets

Method	nohaze small	hurtle	haze middle	nohaze large	haze large	
RESIDE	80 (61%)	77 (58%)	131 (99%)	127 (96%)	132 (100%)	130 (98%)
REVIDE	55 (65%)	50 (59%)	82 (96%)	74 (87%)	85 (100%)	75 (88%)
NTIRE2018O	53 (76%)	51 (73%)	69 (98%)	45 (64%)	70 (100%)	47 (67%)
NTIRE2018I	52 (54%)	48 (50%)	88 (92%)	87 (91%)	96 (100%)	90 (94%)
NTIRE2019	32 (49%)	8 (12%)	61 (94%)	15 (23%)	65 (100%)	16 (25%)
NTIRE2020	47 (52%)	33 (36%)	94 (103%)	59 (65%)	91 (100%)	57 (63%)

We quantitatively and qualitatively measured the negative impact of haze concentration on the performance of the segmentation network. Since the existing dehaze datasets do not contain different concentrations and non-uniform haze images, we used Unreal Engine 5 EPIC (2022b) to render different concentrations of smog data. We passed these different concentrations of data through the segmentation large model separately. The baseline was an image without fog. As the fog concentration increased from 0 to 0.2, which corresponds to going from no haze to very thick haze, the detection rate gradually decreased from 100% to 88%. As shown in the red box area in Fig.4, due to the ability of large models, in severe conditions of uneven thick fog, a large-scale segmentation model can still correctly segment each glass of the window.

5.2. Model size on haze image segmentation

We compared the segmentation ability of different size models for different haze datasets in Fig.5. The left side represents different types of haze, including indoor thick fog REVIDE, outdoor non-uniform fog NTIRE2020, outdoor thick fog NTIRE2019, and thin fog RESIDE. As the size of the segmentation model increases, the number of segmentation categories and accuracy will also increase. A mask that appears more white in the image represents more segments that can be extracted. The right side represents the

segmentation ability of different size segmentation without haze. We found that for RESIDE, a relatively simple haze data set, the segmentation results with fog are often better than those without fog. In more difficult haze scenarios, the larger the model size, the better the segmentation accuracy. It is worth noting that when the parameter number of the segmentation model increases by one level, it often directly offsets the effects of haze. For example, the effect of haze + middle is often better than nohaze + small. This subjectively demonstrates the role of the emerging anti-haze ability of the segmentation network in various complex haze scenarios. For the large model performance on RESIDE is poor, mainly because the small model can already perform effectively in less severe haze scenarios, which is also repeated in the comparison of dehaze results later.

We also conducted quantitative performance statistics on all dehaze datasets under different size segmentation models, as shown in Table 1. All data is the average number of segments per image after the segmentation network segments the entire dehaze dataset. In general, small segmentation models are sufficient for simple fog scenes. For moderate-difficulty scenes, increasing the model scale can overcome the side effects of haze. For extremely difficult fog scenes, increasing the model scale can alleviate the side effects of haze. The fog changes on the RESIDE dataset are simple. Simple haze cannot cause the degradation of large-model
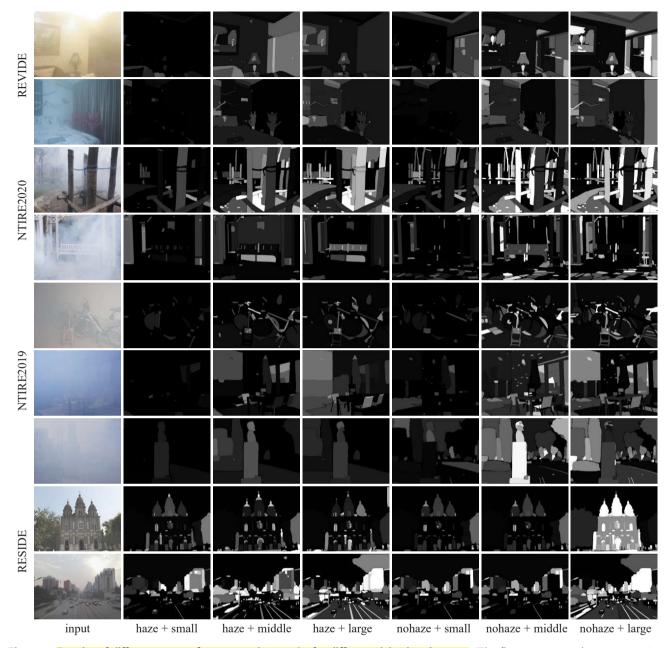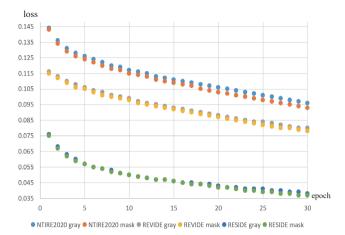
Figure 5: Results of different types of segmentation masks for different dehazing datasets. The figure presents the segmentation results of hazy images in different dehazing datasets from top to bottom. The selected images correspond to the dehazing results in Fig.7. The results from left to right are the hazy input images, different size models segmentation results with haze and dehazing.

segmentation results. The large-model segmentation results of haze images and clear images are not different. At the same time, neither size of the segmentation large model shows a significant improvement in subsequent image dehazing. REVIDE, NTIRE2018O, and NTIRE2018I maintain similar segmentation results, indicating that fog can cause some impact on segmentation, but increasing the scale of the segmentation network can overcome the negative effects of fog. NTIRE2019 and NTIRE2020 show the uniqueness of these two datasets. NTIRE2019 is a dataset of dense fog, and the difference between the presence and absence

of fog is very dramatic. No matter how the scale increases, the segmentation model cannot use extremely weak or even nonexistent signals. However, as the size of the network increases, some weaker but still detectable signals are still correctly segmented by the large model. Segmentation results can improve by 100% or more. NTIRE2020 is an unevenly distributed dehaze dataset. Uneven fog is a challenge for small dehaze models, but for large segmentation models, uneven fog interference has almost no effect, as shown in Fig.5. The gradient haze border does not affect the segmentation effect. At the same time, the performance of
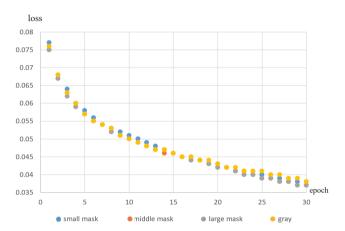
Figure 6: The use of large models can accelerate the dehaze model training. The above figure shows the convergence results of different data sets when trained with and without large segment model masks. The below figure shows the convergence results of the same data set when trained with different sizes of large segment model.

large models directly segmenting uneven fog is also superior to the performance of small models after dehaze processing.

5.3. Segmentation mask accelerates training

Due to the ability of large model segmentation networks to emerge with anti-fog performance, they can enhance the effectiveness of dehaze. We also found the segmentation mask can improve the convergence speed of dehaze network training, as shown in Fig.6. We believe that the segmentation results produced by large models can guide small dehaze networks. Large models segment the same texture variations into different regions, helping dehaze networks accelerate training. All of our experimental structures demonstrate this property. As shown in the blue and red points in the upper figure of Fig.6, compared to the other curves, the loss of the blue and red points converges slowly, representing more complex and difficult to process dehaze datasets under fog conditions. However, red dots fall faster than blue dots, the segmentation mask can significantly accelerate training.

Moreover, this acceleration effect is also present even on simpler dehaze datasets. Enlargement of the blue and green lines in the upper figure of Fig.6 gives the result in the lower part. Different-scale segmentation models can generate slight perturbations, but in general, the segmentation mask of large models accelerates faster than small models and the baseline.

5.4. Comparison of qualitative dehazing results

Fig.7 shows an example of the results of dehazing by different masks. The images in the Fig.7 are the local enlargements of the corresponding regions in Fig.5. The segmentation masks in Fig.5 can also be compared with the corresponding regions in Fig.7. For simple datasets such as RESIDE, existing dehaze networks can easily perform well. However, when faced with thick fog and non-uniform haze scenes, our method shows significant subjective improvements. As the quality of the segmentation mask improves, the quality of the dehaze processing also rapidly increases. Both the color and the details of the edge are restored better. In REVIDE, the green vase in the red box and the air conditioning socket are better dehaze due to the segment mask guidance, resulting in better texture details and colors. In NTIRE2020, non-uniform haze is prone to causing errors in dehaze restoration, such as the trunk in Fig.7 third row being unable to maintain the same color by haze+gray baseline. However, with the guidance of the segmentation model, the trunk can be correctly dehaze. After the pre-dehaze step and segmentation, the texture detail defog network can also perform some restoration of the texture in the third row of NTIRE2020 red box. In the NTIRE2019 three rows of displayed images, due to the guidance of the segmentation model, the bicycle frame, the yellow chair, and the difficult to process branch nodes are also better resolved. It is worth noting that once the pre-de-fog step is incorrect, it can also have side effects. For example, in the third row of NTIRE2019 dehaze+small, if the pre-dehaze step is incorrect, like haze+gray, the segmentation network will also be incorrect and ultimately lead to incorrect dehaze results. However, from a subjective visual perspective, the dehaze results benefit from the anti-fog segmentation ability of large models.

5.5. Comparison of quantitative dehazing results

We also quantitatively compared the improvement effects of our method under different levels of fog. Overall, the results were consistent with subjective perceptions, as shown in Table2. The data was sorted from high to low based on the haze gray mask baseline's PSNR, representing the difficulty of different datasets. The olive color represents the best result, while the teal color represents the second-best result. It can be clearly seen that the color results are diagonally distributed, indicating that our method performs better and improves more for more difficult dehaze datasets and relatively little for easier dehaze datasets. For the simple RESIDE, there was no improvement because the dehazing small network itself can perform well. The NTIRE2018I and NTIRE2018O were relatively moderate, and our method
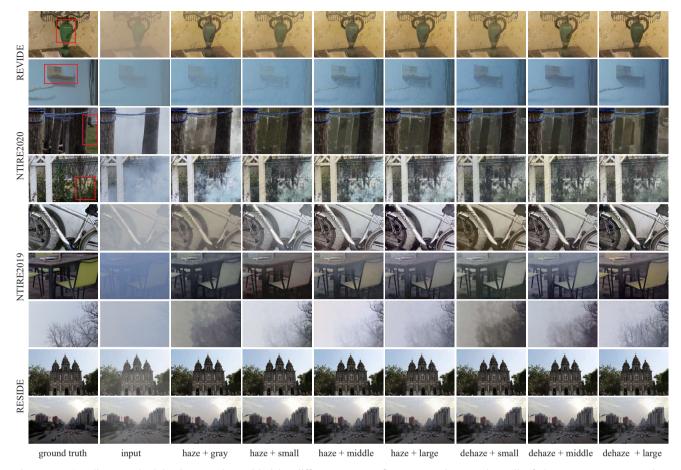
Figure 7: Small network dehazing results guided by different types of segmentation masks. All of the images correspond to Fig.5 in the previous paragraph, which is a magnified version of the previous image. From top to bottom, they represent different datasets of different types of fog. From left to right, they show different dehaze results. The red boxes highlight relevant areas that are worth attention. In scenes that are difficult for existing dehaze networks to handle, it can be seen that as the scale of the segmentation model increases, the quality of the dehaze results improves significantly.

Table 2
Quantitative results comparing the universality of different datasets (best and second best)

Dataset	gray	small	middle	large
RESIDE (haze)	33.81(0%) / 0.988	33.72(1%) / 0.988	32.46(17%) / 0.988	32.65(14%) / 0.982
RESIDE (nohaze/dehaze)	**57.12 / 0.999**	33.67(16%) / 0.988	33.70(13%) / 0.988	32.84(12%) / 0.988
NTIRE2018I (haze)	28.42(8%) / 0.938	29.05(0%) / 0.946	27.97(13%) / 0.961	27.73(16%) / 0.956
NTIRE2018I (nohaze/dehaze)	**53.38 / 0.998**	28.47(7%) / 0.914	27.35(22%) / 0.951	27.31(22%) / 0.929
NTIRE2018O (haze)	26.03(1%) / 0.795	26.06(1%) / 0.796	26.08(1%) / 0.794	26.01(2%) / 0.794
NTIRE2018O (nohaze/dehaze)	**49.90 / 0.997**	26.09(1%) / 0.793	26.14(0%) / 0.793	26.15(0%) / 0.796
REVIDE (haze)	22.03(17%) / 0.878	22.42(12%) / 0.880	22.66(9%) / 0.880	22.61(10%) / 0.883
REVIDE (nohaze/dehaze)	**42.66 / 0.998**	22.58(10%) / 0.873	22.86(6%) / 0.882	23.41(0%) / 0.888
NTIRE2020 (haze)	19.90(13%) / 0.658	20.52(5%) / 0.652	20.54(5%) / 0.652	20.69(3%) / 0.659
NTIRE2020 (nohaze/dehaze)	**50.11 / 0.999**	20.33(7%) / 0.655	20.68(3%) / 0.655	20.94(0%) / 0.658
NTIRE2019 (haze)	16.46(17%) / 0.415	17.15(8%) / 0.578	17.32(6%) / 0.575	17.62(2%) / 0.575
NTIRE2019 (nohaze/dehaze)	**49.51 / 0.999**	17.26(7%) / 0.545	17.38(5%) / 0.587	17.84(0%) / 0.587

could improve by 1% to 8%. For medium-level difficulty dehaze datasets, larger-sized segmentation networks quickly produce saturated effects and cannot continue to provide effective texture segmentation. For NTIRE2020, NTIRE2019, and REVIDE, which were more difficult scenarios, our

method could improve by 13% to 17%. The increase in segmentation model scale can improve dehazing performance.

Additionally, we also studied the optimal effect of using large models to guide dehazing, as shown in the bold part of Table 2. We applied a grayscale image without haze as a

Table 3
Quantitative results comparing the universality of different networks (best and second best)

RESIDE	nohaze	haze gray	haze small	de. small	haze mid.	de. midd.	haze large	de. large
Unet10M	**41.79**	28.05(2%)	27.78(5%)	27.65(7%)	27.91(4%)	28.22(0%)	27.43(10%)	27.32(11%)
Uformr60M	**45.64**	30.24(4%)	29.55(13%)	30.59(0%)	29.55(13%)	29.73(10%)	29.88(8%)	30.11(6%)
Restor.100M	**57.12**	33.81(0%)	33.72(1%)	33.67(2%)	32.46(17%)	33.70(1%)	32.65(14%)	32.84(12%)

NTIRE2019	nohaze	haze gray	haze small	de. small	haze mid.	de. midd.	haze large	de. large
Unet10M	**25.49**	14.15(1%)	14.11(1%)	14.11(1%)	14.10(1%)	14.13(1%)	14.21(0%)	14.17(0%)
Uformr60M	**26.28**	15.04(5%)	15.19(3%)	15.27(2%)	15.32(2%)	15.45(0%)	15.42(0%)	15.45(0%)
Restor.100M	**49.51**	16.46(17%)	17.15(8%)	17.26(7%)	17.32(6%)	17.38(5%)	17.62(3%)	17.84(0%)

mask to the network for dehazing, and the network could quickly learn and accept it. It appears that there is much potential to explore with existing large models. This method is also used in the following section to evaluate the learning ability of different small dehaze networks.

5.6. Mask awareness of dehazing network

To fully demonstrate the network's learning ability and independence from prior knowledge of dehazing, we did not choose various custom-designed models specifically for dehazing, such as Dehazeformer Song et al. (2022) and C2PNet Zheng et al. (2023). We believe these networks may be more suitable for certain types of hazy scenes. Instead, we selected general models for image restoration algorithms with different sizes, represented by Unet Ronneberger, P.Fischer and Brox (2015), Uformer Wang, Cun, Bao, Zhou, Liu and Li (2022), and Restormer Zamir, Arora, Khan, Hayat, Khan and Yang (2022).

The size of the model is crucial for the ability of a large-scale segmentation model, and it is also important for a small-scale dehaze model to accept the ability of a large-scale segmentation model. Although large-scale segmentation models can transmit anti-fog ability to defog networks, different scales of defog networks exhibit different learning abilities under the same training resources. As shown in Table 3, 10M, 60M, and 100M represent the scale of different models. For simple dehaze datasets, the improvement caused by the segmentation mask is not apparent regardless of the scale of the dehaze network. For more complex scenarios of haze, bigger size and more complex small-scale dehazing models that have better fitting abilities can often learn the anti-haze performance of large-scale segmentation networks. Furthermore, larger-scale dehaze models can better perceive the improvement of large-scale segmentation models. To fully reflect the learning relationship between large and small models, the previous demonstrations have all been tested using the largest-scale dehaze model.

6. Conclusion

In this article, we discover and prove the emergence of anti-fog ability in large-dataset, large-parameter segmentation model. We apply this ability to small dehaze models. First, we gray-scale encode the segmentation results of the large-scale segmentation model. Then, we input these segmentation masks into a new channel of the small dehaze model to achieve better dehaze results. At the same time, we also discover the acceleration effect of large model masks on the training of the dehaze network. We analyze the impact of different fog scenarios, dehaze datasets, large-capacity segmentation model sizes, and small dehaze model sizes on dehaze results. For complex fog scenes, large models can help small dehaze models better handle them. A wide range of experimental results show that our method performs well.

Overall, we enable the large semantic segmentation models to help small dehaze models that cannot be trained with large parameters and large models. This allows small models to enjoy the various development advantages of large models, while also without requiring large models to be specifically optimized for small-model tasks. We propose a new method for large models to help low-level computer vision, allowing low-level visual tasks to benefit from the development of large models.

CRediT authorship contribution statement

Yueting Chen: Data curation, Writing - Original draft preparation.

References

Ancuti, C., Ancuti, C.O., Vleeschouwer, C.D., Bovik, A., 2016. Nighttime dehazing by fusion, in: IEEE International Conference on Image Processing.

Ancuti, C.O., Ancuti, C., Sbert, M., Timofte, R., 2019. Dense haze: A benchmark for image dehazing with dense-haze and haze-free images. arXiv .

Ancuti, C.O., Ancuti, C., Timofte, R., 2020. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

Ancuti, C.O., Ancuti, C., Timofte, R., Vleeschouwer, C.D., 2018a. I-haze: a dehazing benchmark with real hazy and haze-free indoor images .

Ancuti, C.O., Ancuti, C., Timofte, R., Vleeschouwer, C.D., 2018b. O-haze: A dehazing benchmark with real hazy and haze-free outdoor images, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al., 2020a. Language models are few-shot learners. Advances in neural information processing systems 33, 1877–1901.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Amodei, D., 2020b. Language models are few-shot learners .

Cai, B., Xu, X., Jia, K., Qing, C., Tao, D., 2016. Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing 25, 5187–5198.

Chen, D., He, M., Fan, Q., Liao, J., Hua, G., 2019. Gated context aggregation network for image dehazing and deraining, in: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV).

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3213–3223.

EPIC, 2022a. https://docs.unrealengine.com/4.27/en-us/building-worlds/fogeffects/.

EPIC, 2022b. https://docs.unrealengine.com/5.0/en-us.

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The pascal visual object classes (voc) challenge. International Journal of Computer Vision 88, 303–338.

Hahner, M., Dai, D., Sakaridis, C., Zaech, J.N., Van Gool, L., 2019. Semantic understanding of foggy scenes with purely synthetic data. IEEE .

He, K., 2009. Single image haze removal using dark channel prior .

Jia, C., Yang, Y., Xia, Y., Chen, Y.T., Parekh, Z., Pham, H., Le, Q., Sung, Y.H., Li, Z., Duerig, T., 2021. Scaling up visual and vision-language representation learning with noisy text supervision, in: International Conference on Machine Learning, PMLR. pp. 4904–4916.

Jing, Z., Yang, C., Shuai, F., Yu, K., Chang, W.C., 2017. Fast haze removal for nighttime image using maximum reflectance prior, in: IEEE Conference on Computer Vision Pattern Recognition.

Jing, Z., Yang, C., Wang, Z., 2014. Nighttime haze removal based on a new imaging model, in: 2014 IEEE International Conference on Image Processing (ICIP).

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al., 2023. Segment anything. arXiv preprint arXiv:2304.02643 .

Lee, S., Son, T., Kwak, S., 2022. Fifo: Learning fog-invariant features for foggy scene segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 18911–18921.

Li, B., Peng, X., Wang, Z., Xu, J., Feng, D., 2017. Aod-net: All-in-one dehazing network, in: Proceedings of the IEEE international conference on computer vision, pp. 4770–4778.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al., 2021. Learning transferable visual models from natural language supervision, in: International conference on machine learning, PMLR. pp. 8748–8763.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., Sutskever, I., 2021. Zero-shot text-to-image generation, in: International Conference on Machine Learning, PMLR. pp. 8821–8831.

Ren, W., Pan, J., Zhang, H., Cao, X., Yang, M.H., 2020. Single image dehazing via multi-scale convolutional neural networks with holistic edges. International Journal of Computer Vision 128, 240–259.

Ronneberger, O., P.Fischer, Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer. pp. 234–241. URL: http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a. (available on arXiv:1505.04597 [cs.CV]).

Sakaridis, Christos, Dai, Dengxin, Gool, V., Luc, 2018. Semantic foggy scene understanding with synthetic data. INTERNATIONAL JOURNAL OF COMPUTER VISION .

Song, Y., He, Z., Qian, H., Du, X., 2022. Vision transformers for single image dehazing. arXiv e-prints .

Tang, K., Yang, J., Wang, J., 2014. Investigating haze-relevant features in a learning framework for image dehazing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H., 2022. Uformer: A general u-shaped transformer for image restoration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 17683–17693.

Yi, X., Ma, B., Zhang, Y., Liu, L., Wu, J., 2022. Two-step image dehazing with intra-domain and inter-domain adaptation. Neurocomputing 485, 1–11.

Yu, L., Tan, R.T., Brown, M.S., 2015. Nighttime haze removal with glow and multiple light colors, in: IEEE International Conference on Computer Vision.

Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., 2022. Restormer: Efficient transformer for high-resolution image restoration, in: CVPR.

Zhang, J., Cao, Y., Zha, Z.J., Tao, D., 2020. Nighttime dehazing with a synthetic benchmark, in: Proceedings of the 28th ACM International Conference on Multimedia, Association for Computing Machinery, New York, NY, USA. p. 2355–2363. URL: https://doi.org/10.1145/3394171.3413763, doi:10.1145/3394171.3413763.

Zhang, X., Dong, H., Pan, J., Zhu, C., Tai, Y., Wang, C., Li, J., Huang, F., Wang, F., 2021. Learning to restore hazy video: A new real-world dataset and a new method, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9239–9248.

Zheng, Y., Zhan, J., He, S., Dong, J., Du, Y., 2023. Curricular contrastive regularization for physics-aware single image dehazing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5785–5794.

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A., 2017. Scene parsing through ade20k dataset, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 633–641.

Author biography with author photo. Author biography. Author biography. Author biography. Author biography.