

# AWED: Asymmetric Wavelet Encoder-Decoder Framework for Simultaneous Gas Distribution Mapping and Gas Source Localization

Shutong Zhong, Ming Zeng, Member, IEEE, and Liang Mao

**Abstract**—Gas distribution mapping (GDM) and gas source localization (GSL) are two crucial research areas in gas monitoring. However, due to the time-varying and non-uniform nature of gas distribution and the limitations of gas sensors, the accurate and rapid estimation of gas distribution from sparse sensor data is a challenging task. This article proposes an end-to-end model called asymmetric wavelet encoder-decoder (AWED) to address GDM and GSL from ultra-sparse sensor data. The model uses a simplified encoder and enhanced decoder, incorporating wavelet reconstruction module (WRM) to decode from both spatial and frequency domains. Additionally, a wavelet L1 loss is introduced to promote frequency domain similarity between predicted and real images. The proposed method achieves a 32x super-resolution of gas distribution maps from 7x7 sensor data to 224x224 resolution images, and achieves a gas source localization accuracy of 0.265m within a 10mx10m area. The model also exhibits fewer parameters, faster prediction speed, and better real-time performance compared to existing methods. Experiments demonstrate that the proposed method outperforms traditional interpolation and state-of-the-art deep learning-based methods in reconstructing gas distribution maps and localizing gas sources from ultra-sparse sensor data.

**Index Terms**—Gas distribution mapping, gas source localization, image super-resolution, wavelet transform

## I. INTRODUCTION

**G**AS monitoring is important for ensuring personnel safety, protecting the ecological environment, and improving production efficiency. It has a wide range of applications in various industrial fields. For example, in industries such as chemical, petroleum, and coal mining, gas leaks can lead to fires or explosions. In the power industry, the use and production of various toxic and harmful gases are involved. By monitoring these gases in real-time, we can detect anomalies promptly and take measures to prevent accidents.

Gas distribution mapping (GDM) [1],[2] and gas source localization (GSL) [2],[3],[4] are two crucial research areas in gas monitoring. By constructing gas distribution maps and predicting the location of leakage sources, we can clearly understand the distribution and concentration levels of gases in a specific area, which can guide emergency response teams to quickly determine affected areas and take effective evacuation and rescue measures. However, due to the characteristics of time-varying and non-uniform distribution of gases[5], as well as the long response and recovery times of current gas sensors,

it is challenging to quickly and accurately construct high-resolution gas distribution maps with a few sparse sampling data. Although more sensors can be used to improve the spatial resolution of gas distribution maps, this also brings high equipment costs and maintenance challenges.

Many previous works [6],[7],[8],[9] have explored the application of computational fluid dynamics (CFD) in GDM and GSL, which rely on mathematical models for numerical solutions. However, this simulation-based approach requires precise environmental geometric models. Besides, in real environments, boundary conditions such as interfaces of different materials and variable environmental conditions are difficult to accurately obtain. Additionally, the gas plume is circulated through wind which is turbulent in nature, making the highest concentration patch in the environment not the actual source location [10]. In recent years, with the development of deep learning, some researchers have begun to use neural networks to accomplish GDM and GSL tasks. For example, Winkler et al.[11] used convolutional neural network(CNN)-based encoder-decoder structures for gas distribution map super-resolution. Zhang et al.[12] used historical sensor data as multi-channel time series inputs with SE block for gas distribution map super-resolution. The SE block assigns different weights to each channel. Hu et al.[13] utilized conditional generative adversarial networks (CGAN) to estimate non-stationary turbulent flow fields in urban models with airflow velocity information. Badawi et al.[14] incorporated traditional projection-onto-convex-sets (POCS) iteration in the structure of the deep model to obtain a regularized solution that conforms to our prior knowledge of the spatio-temporal structure of the gas concentration distribution. Sun et al.[15] employed the tunable diode laser absorption spectroscopy (TDLAS) technique for accurate mapping of indoor gas distributions. Additionally, the maximum likelihood expectation maximization (MLEM) algorithm was proposed to correct concentrations in each grid, resulting in higher reconstruction accuracy. Juffry et al.[16] used a CFD model to simulate distribution images of 361 different ethanol vapor leak locations in an indoor environment. Then a multilayer perceptron (MLP) was trained to directly regress the leak source location. Bi et al.[17] combined the spatiotemporal feature extraction capabilities of CNN with the temporal processing capabilities of long short-term memory networks (LSTM) to complete the task of locating hydrogen leak sources in hydrogen refueling stations. Zhang et al.[18] used an attention mechanism to assign initial weights to the time series data. Then, an LSTM was employed

This work was supported by National Natural Science Foundation of China, Tianjin under Grant 23JCYBJC01670. (Corresponding author: Ming Zeng.)

The authors are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: zengming@tju.edu.cn).

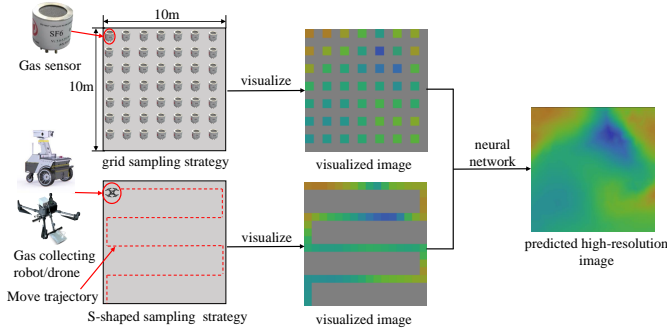


Fig. 1. Diagram of two sampling strategies.

to learn the complex correlations between the weighted data and the leak locations. Ruiz et al.[19] proposed a gas source localization method based on physics-guided neural networks (PGNN). Luo et al.[20] utilized continuous wavelet transform and modulus maxima approach to elucidate the temporal difference information embedded within gas sensor signals, and deployed gas sensors aboard a quadcopter helicopter to fulfill GSL tasks. Yan et al.[21] designed a system incorporating four metal-oxide semiconductor gas sensor modules for the detection of gas concentrations. By learning the spatiotemporal correlations within the sensor array response data, an end-to-end estimation of the odor source direction was achieved.

Despite these advances, the multiples of the above GDM super-resolution model are usually limited to 2 or 4. This implies that a large number of sampling points are needed to obtain a more accurate high-resolution image, which is difficult to achieve in practical situations. In addition, these methods have not explored the impact of different sampling strategies on the reconstruction accuracy. Besides, current work often discusses GDM or GSL tasks separately. There is no end-to-end multi-task learning model that integrates both GDM and GSL simultaneously. Furthermore, due to the strict requirements for response speed in emergency scenarios, the resource consumption and real-time performance of the model also need to be considered, which have not been adequately discussed in current research.

In real-world scenarios, there are typically two methods for collecting gas distribution data: one is through an array of sensors at fixed locations, and the other is using mobile robots or drones equipped with gas sensors to collect gas concentrations along an S-shaped trajectory. Therefore, we design two sampling strategies: grid sampling strategy and S-shaped sampling strategy, as shown in Fig.1. For the grid sampling strategy, the data from one sensor represents the gas concentration within a certain range. Since there are intervals between sensors, we add masks for these intervals (gray parts in the visualized image). Similarly, for the S-shaped sampling strategy, we add masks to the unsampled areas. Finally, a neural network is used to predict a complete high-resolution gas distribution map.

Because sensor data is very sparse (i.e.  $7 \times 7$  for grid sampling strategy), reconstructing a high-resolution image (i.e.  $224 \times 224$ ) requires a super-resolution magnification of 32 times. This is a challenging task for typical super-resolution

neural networks. Therefore, we divide the image into patches of fixed size, where each patch represents the data from one sensor, and we add masks to the unsampled patches. Consequently, we can view this problem as a combination of image super-resolution and image inpainting tasks: super-resolution within the patches and inpainting between the patches. For both of these tasks, the most commonly used network structure is the encoder-decoder architecture such as AE [22], VAE [23], U-Net [24], and Transunet [25]. The encoder-decoder architecture consists of an encoder and a decoder. The encoder compresses the input data into a low-dimensional representation and the decoder reconstructs the data from the representation.

The problem addressed in this article has certain specific characteristics. Since the input image is only sparsely sampled low-density gas data and the output is a large-scale high-resolution gas distribution map, the amount of information available for extraction in the encoder is very limited. In response to these characteristics, we propose an Asymmetric Wavelet Encoder Decoder (AWED) model. On one hand, we simplify the design of the encoder, and on the other hand, we strengthen the decoder by incorporating a wavelet reconstruction module (WRM) that decodes from both the spatial domain and the frequency domain. Additionally, we introduce a wavelet L1 loss to enhance the similarity between the predicted image and the real image in the frequency domain, thereby improving the quality of the super-resolution reconstruction.

The main contributions of this article are as follows:

- 1) A novel encoder-decoder model is proposed, which simplifies the encoder and strengthens the decoder design. The wavelet reconstruction module is introduced into the decoder to effectively reconstruct high-resolution gas distribution maps from sparse sensor data.
- 2) A wavelet L1 loss function is introduced to optimize the model's performance in the frequency domain, enhancing the similarity between the predicted and real images in the frequency domain, thereby improving the quality of the reconstructed images.
- 3) An end-to-end learning framework is constructed that can handle both gas distribution mapping and gas source localization tasks simultaneously. The framework concurrently learns and optimizes of GDM and GSL tasks within a single model, which improves the practicality and efficiency of the model.
- 4) Experimental results show that the method proposed in this article can accurately reconstruct a  $224 \times 224$  resolution gas distribution map and achieve a localization accuracy of 0.265m within a  $10\text{m} \times 10\text{m}$  space using only  $7 \times 7$  sensor data.

The structure of this article is as follows. Section II introduces some related works about image super-resolution and wavelet transform. Section III explains our Asymmetric Wavelet Encoder Decoder for simultaneous gas distribution mapping and gas source localization. Section IV is devoted to the explanation of the experimental setup and results. Section V draws a conclusion for this article.

## II. RELATED WORK

### A. Image Super-resolution

Image super-resolution aims to reconstruct a high-resolution image from a low-resolution image, which can restore the image's detail and clarity. This is particularly important for image browsing, editing, and display. Early studies involved direct statistical modeling of images [26],[27] or used spline functions as image priors, with interpolation pixels estimated by fitting nearby pixels [28],[29]. With the rapid development of deep learning technology, researchers began to apply neural networks to image super-resolution tasks. SRCNN [30] was a pioneering work that first used CNN to learn the non-linear mapping between low-resolution and high-resolution images, outperforming traditional methods. Subsequently, Kim et al.[31] deepened the SRCNN by proposing a much deeper super-resolution network called VDSR with 20 convolutional layers. RCAN [32] introduced a residual structure and built a deep model with over 400 layers. Additionally, research [33],[34] added attention mechanisms in the spatial or channel dimensions of the super-resolution model, further enhancing modeling capabilities. With the widespread application of Vision Transformer (ViT) [35] in computer vision, many researches have emerged using ViT for image super-resolution. SwinIR [36] is an image restoration model based on the Swin Transformer, which effectively captures multi-scale features in images through its hierarchical transformer structure, making it suitable for not only super-resolution but also for denoising, deblurring, and other image restoration tasks. Restormer [37] is a transformer-based model designed specifically for high-resolution image restoration, achieving high performance and low computational complexity in super-resolution tasks through efficient attention mechanisms and convolution operations. DRCT [38] proposed the dense-residual-connected transformer to eliminate spatial information loss during encoding, achieving state-of-the-art (SOTA) performance in the field of image super-resolution.

### B. Wavelet Transform

Wavelet transform possesses properties such as shift-invariance, multi-resolution analysis, edge-detection, local operations, and energy compaction, which have led to its widespread application in various image processing fields. It has shown excellent performance in image compression [39], image denoising [40], image classification [41][42], and object detection [43]. Additionally, in the field of image restoration, DeWRNet [44] utilized wavelet transform to decompose images into low-frequency and high-frequency sub-images, and employed distinct models to process them individually, which can more effectively restore and enhance details in super-resolution image. Guo et al.[45] combined wavelet transform with a fully convolutional neural network to achieve image super-resolution. Liu et al.[46] proposed Multi-level Wavelet CNN (MWCNN) for image restoration, reducing the computational resource consumption of neural networks. Yang et al.[47] used discrete wavelet transform and inverse discrete wavelet transform to replace the downsampling and upsampling operations in Unet for image dehazing, allowing the network to

focus more on edge features and color enhancement. Jeevan et al.[48] proposed a resource-efficient computer vision network architecture based on wavelet transform, which can achieve competitive performance in tasks such as image classification, image segmentation, and image super-resolution with a very small number of parameters. For the problem studied in this article, wavelet transform can provide local time-frequency information, effectively capturing the local features and texture information of the image, which is crucial for restoring and reconstructing details in image super-resolution and image inpainting tasks.

## III. METHODOLOGY

To address the GDM and GSL problems under conditions of ultra-low density sampling, this article proposes the Asymmetric Wavelet Encoder Decoder. The overall architecture of AWED, as shown in Fig.2, includes a simplified encoder, a decoder based on wavelet transform, and Distribution Mapping Head (DMH) and Source Localization Head (SLH) for simultaneous GDM and GSL tasks.

Taking the grid sampling strategy as an example, the input image  $I(I \in R^{H \times W \times 3})$  is divided into patches of size  $(H/14, W/14)$ , totaling  $14 \times 14$ , 196 patches. Among these, the pixel values of 49 patches represent the gas concentration data from 49 sensors, while the remaining 147 patches are masked, representing the unsampled areas between sensors. A binary mask image  $M(M \in R^{H \times W \times 1})$  is first extracted from the image  $I$ , then  $M$  and  $I$  are concatenated and input into AWED. It is necessary to input  $M$  into the neural network because the network needs to know which parts are missing data. In each patch, the pixel values are uniform. If traditional  $3 \times 3$  or  $5 \times 5$  convolutions are employed, the majority of the convolutional computations would be conducted on identical pixel values, which is nonsensical. Instead, AWED first employs a single downsampling step as the encoder, using a large convolutional kernel (kernel size=16, stride=16) to obtain a low-dimensional feature map  $F_l(F_l \in R^{H/16 \times W/16 \times C})(eq.(1a))$ . On one hand, large convolutional kernels reduce the number of downsampling layers, thus reducing the number of network parameters. On the other hand, large convolutional kernels can also increase the receptive field of the network. The low-dimensional feature map is then passed through the decoder to learn the completed high-dimensional features  $F_h(F_h \in R^{H/2 \times W/2 \times C/8})(eq.(1b))$ . Finally, the high-dimensional feature maps are sent to the Distribution Mapping Head and Source Localization Head to obtain the reconstructed high-resolution image  $I'(I' \in R^{H \times W \times 3})$  and the predicted leak source location  $P'(P' \in R^{1 \times 2})$  (eq.(1c) and eq.(1d)).

### A. Wavelet Transform-based Decoder

The decoder consists of 3 stacked decoder blocks, with each decoder block doubling the spatial dimensions of the input features and halving the number of channels. After passing through 3 decoder blocks, a feature map of size  $(H/2, W/2, C/8)$  is obtained. Each decoder block is composed of an upsampling module and a wavelet reconstruction module in parallel, reconstructing the gas distribution map

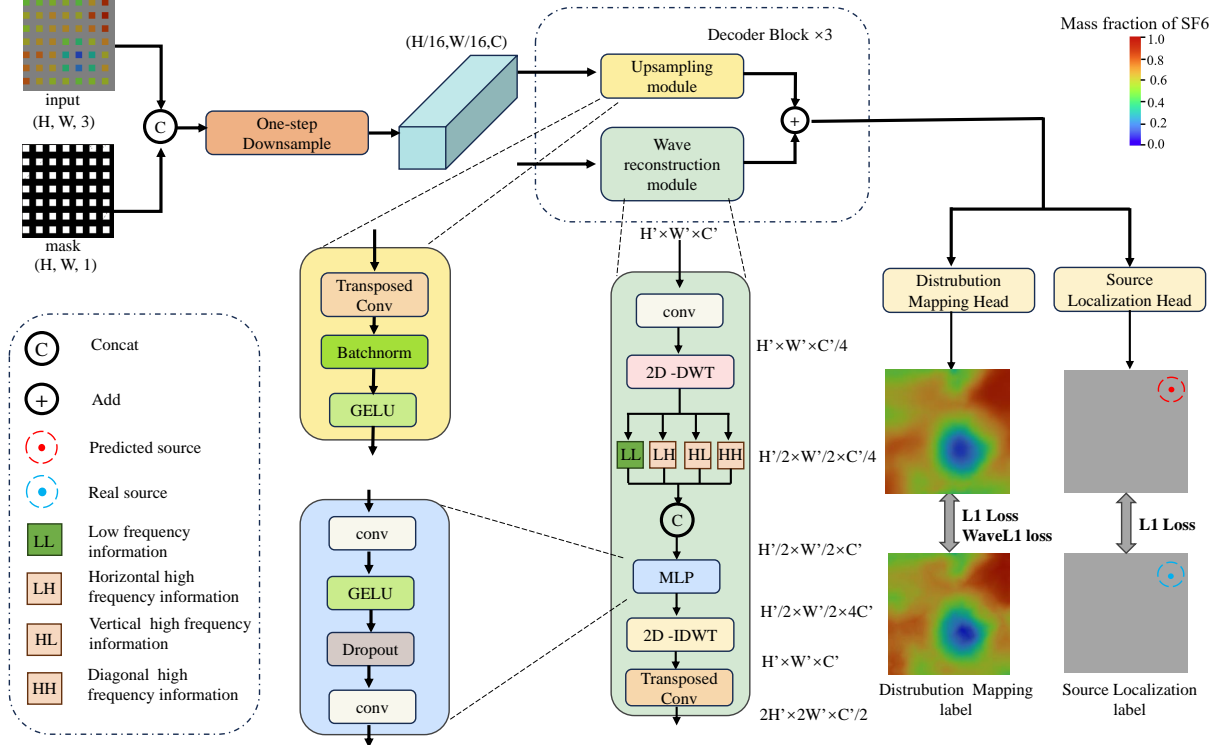


Fig. 2. Architecture of our asymmetric wavelet encoder-decoder.

$$F_l = \text{Encoder}(\text{Concat}[I, M]), I \in R^{H \times W \times 3}, M \in R^{H \times W \times 1} \quad (1a)$$

$$F_h = \text{Decoder}(F_l), F_l \in R^{H/16 \times W/16 \times C}, F_h \in R^{H/2 \times W/2 \times C/8} \quad (1b)$$

$$I' = \text{DMH}(F_h), I' \in R^{H \times W \times 3} \quad (1c)$$

$$P' = \text{SLH}(F_h), P' \in R^{1 \times 2} \quad (1d)$$

from the perspectives of both the spatial domain and the frequency domain. The upsampling module uses transposed convolutions to increase the image resolution, followed by batch normalization and the GELU activation function. The output of the upsampling module is added to the output of the wavelet reconstruction module, and the result is then fed into the next decoder block.

The core idea of the wavelet reconstruction module is to use two-dimensional discrete wavelet transform to extract the high-frequency and low-frequency parts of the image. In the wavelet domain, a convolutional neural network is used for learning missing data, followed by an inverse wavelet transform to convert the data back into an image. Finally, transposed convolution is used to upsample the image. For an input tensor  $x_{in}(x_{in} \in R^{H' \times W' \times C'})$ , the wavelet reconstruction module first uses convolution to reduce the number of channels to  $C'/4$ , resulting in a feature map  $x_1(x_1 \in R^{H'/4 \times W'/4 \times C'/4})$ (eq.(2a)). A two-dimensional discrete wavelet transform is then applied to obtain four components: LL, LH, HL, and HH, which represent the low-frequency part, the horizontal high-frequency part, the vertical high-frequency part, and the diagonal high-frequency part of the image, each with dimensions  $(H'/2, W'/2, C'/4)$ (eq.(2b)), as shown in

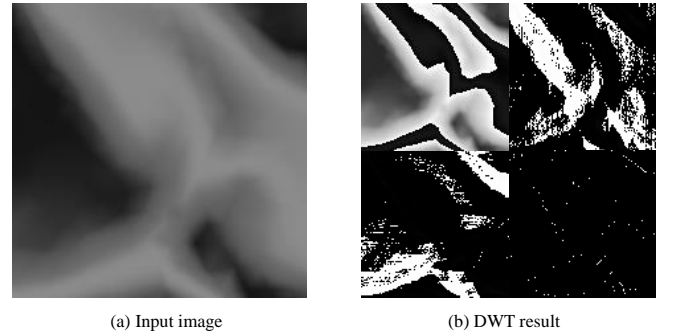


Fig. 3. Illustration of two-dimensional discrete wavelet transform.

Fig.3. Multilevel wavelet transform involves further wavelet decomposition of the low-frequency part from the first-level wavelet transform, resulting in four components of size  $(H'/4, W'/4, C'/4)$ . These four components are then concatenated along the channel dimension and processed by an MLP to obtain a feature map  $x_2(x_2 \in R^{H'/2 \times W'/2 \times 4C'})$ (eq.(2c)). Finally, an inverse wavelet transform is applied to restore the data to the size of the input image, resulting in  $x_3(x_3 \in R^{H' \times W' \times C'})$ (eq.(2d)). This is followed by a transposed con-

$$x_1 = \text{Conv}(x_{in}), x_{in} \in R^{H' \times W' \times C'}, x_1 \in R^{H' \times W' \times C' / 4} \quad (2a)$$

$$LL, LH, HL, HH = \text{DWT}(x_1), LL, LH, HL, HH \in R^{H'/2 \times W'/2 \times C' / 4} \quad (2b)$$

$$x_2 = \text{MLP}(\text{Concat}([LL, LH, HL, HH])), x_2 \in R^{H' / 2 \times W' / 2 \times 4C'} \quad (2c)$$

$$x_3 = I - \text{DWT}(x_2), x_3 \in R^{H' \times W' \times C'} \quad (2d)$$

$$x_{out} = \text{TransposedConv}(x_3), x_{out} \in R^{2H' \times 2W' \times C' / 2} \quad (2e)$$

volution to perform the upsampling operation, yielding the upsampled feature map  $x_{out}$  ( $x_{out} \in R^{2H' \times 2W' \times C' / 2}$ ) (eq.(2e)).

Traditional Fourier transforms are not effective in handling signals with abrupt changes or edges because Fourier transforms are based on global information. In contrast, wavelet transforms can provide local time-frequency information, including both low-frequency approximation parts and high-frequency detail parts. This time-frequency information is capable of effectively capturing local features and texture information of images, which is crucial for restoring and reconstructing details in super-resolution and inpainting tasks. Moreover, wavelet transforms have a natural multi-scale property, allowing for the analysis of images at different scales, which is very useful for understanding the multi-scale features of image content. Additionally, wavelet transforms can represent images as a series of wavelet coefficients, which often exhibit high sparsity in the sparse domain. Utilizing this sparsity can lead to more efficient image reconstruction, reducing computational complexity and improving reconstruction efficiency. Experiments also show that introducing a wavelet reconstruction module results in only a small increase in the number of parameters.

### B. Distrubution Mapping Head and Source Localization Head

The Distrubution Mapping Head and Source Localization Head decode the gas distribution map and the gas leak source from the feature maps respectively. After passing through 3 layers of decoder blocks, a feature map of size  $(H/2, W/2, C/8)$  is obtained. The DMH consists of a decoder block and a convolutional layer. In the DMH, the decoder block outputs a feature map of size  $(H, W, C/16)$ , which is then input into the convolutional layer to obtain a reconstructed image of size  $(H, W, 3)$ . The SLH uses an MLP to project the feature map to a 2-dimensional vector  $[x, y]$ , representing the horizontal and vertical coordinates of the leak source.

One important question is which layer of the decoder block information should be used to decode the gas leak source. A natural approach would be to place the SLH immediately after the DMH, using the reconstructed image to predict the leak source. However, our experiments have shown that this is not the optimal approach. On one hand, the accuracy of leak source prediction will be affected by reconstruction errors in the image. On the other hand, decoding from a high-resolution image quadratically increases the number of parameters in the MLP. Therefore, we choose to decode the leak source location from the feature map with lower resolution in the decoder. And we have investigated the impact of different decoder block layers on the accuracy of gas source localization and found

that decoding the leak position from the feature map of the 3rd decoder block layer achieves the best accuracy.

### C. Wavelet L1 Loss and Overall Loss

As mentioned earlier, wavelet transform can effectively capture the local features and texture information of images, which is crucial for image super-resolution and inpainting. Therefore, in addition to computing the L1 loss between the reconstructed image  $I'$  and its label image  $\hat{I}$ , we also calculate the wavelet L1 loss between them. This involves transforming  $I'$  and  $\hat{I}$  to the wavelet domain and then computing the L1 loss again. For the predicted leak source  $P'$  and the true leak source  $P$ , we calculate their L1 loss. The overall loss  $L$  is composed of the reconstruction loss  $L_{recon}$  and the localization loss  $L_{loc}$ .  $L_{recon}$  consists of the reconstruction L1 loss and the reconstruction wavelet L1 loss.

$$L = L_{recon} + L_{loc} \quad (3)$$

$$L_{recon} = \frac{1}{HW} \sum_{i=0}^H \sum_{j=0}^W |\hat{I}_{i,j} - I'_{i,j}| + \frac{4}{HW} \sum_{i=0}^{H/2} \sum_{j=0}^{W/2} |DWT(\hat{I})_{i,j} - DWT(I')_{i,j}| \quad (4)$$

$$L_{loc} = \sum_{k=0}^1 |P_k - P'_k| \quad (5)$$

In eq.(4),  $H, W$  represent the height and width of the image,  $\hat{I}_{i,j}$  and  $I'_{i,j}$  represent the pixel values of row  $i$  and column  $j$  of the ground-truth image and reconstructed image. DWT represents discrete wavelet transform. In eq.(5),  $P_k$  and  $P'_k$  represent the coordinates of ground-truth gas leak source and predicted gas leak source.

TABLE I  
SF6 SIMULATION PARAMETER SETTINGS

Parameter	Value
Density[kg/m <sup>3</sup> ]	6.52
Specific heat[J/(kg · K)]	665.18
Thermal conductivity[W/(m · K)]	0.01204
Viscosity[kg/(m · s)]	1.72e-05
Molecular weight[kg/kmol]	146.054

TABLE II  
PERFORMANCE COMPARISON OF DIFFERENT METHODS

Method	Params(M)	Grid sampling			S-shape sampling		
		L1 loss↓	SSIM↑	PSNR↑	L1 loss↓	SSIM↑	PSNR↑
linear interpolate	—	0.0368	0.9239	25.4472	0.0371	0.9041	23.9024
nearest interpolate	—	0.0416	0.8234	23.5459	0.0435	0.8234	23.4559
cubic interpolate	—	0.0357	0.9250	25.6624	0.0367	0.9147	24.8574
MAE [49]	111.00	0.0176	0.9441	30.5514	0.0370	0.9091	23.9982
TransUnet [25]	93.23	0.0159	0.9628	30.9896	0.0275	0.9383	26.0911
Segformer [50]	7.98	0.0194	0.8890	29.9072	0.0298	0.8719	26.1485
DRCT [38]	5.63	0.0182	0.9574	29.7258	—	—	—
AWED-T (ours)	<b>0.68</b>	0.0158	0.9597	31.1594	0.0289	0.9239	26.0817
AWED-B (ours)	10.30	0.0147	0.9640	31.4679	0.0253	0.9420	26.7349
AWED-L (ours)	41.15	<b>0.0143</b>	<b>0.9653</b>	<b>31.8560</b>	<b>0.0239</b>	<b>0.9461</b>	<b>27.1626</b>

TABLE III  
PERFORMANCE AND RESOURCE USAGE OF DIFFERENT METHODS

Method	Params(M)	FLOPs(G)	GPU Memory(GB)	FPS(/s)
MAE [49]	111.00	9.3	4.7	127
TransUnet [25]	93.23	24.67	13.5	92
Segformer [50]	7.98	15.75	22.0	142
DRCT [38]	5.63	2.79	2.59	121
AWED-T (ours)	<b>0.68</b>	<b>0.805</b>	<b>2.4</b>	<b>231</b>
AWED-B (ours)	10.30	12.44	7.1	184
AWED-L (ours)	41.15	49.78	12.8	101

#### IV. EXPERIMENT

##### A. Dataset Construction

The actual gas dispersion experiments are unrepeatable and uncontrollable, making it difficult to conduct a large number of repetitive experiments. Moreover, the excessive deployment of dense gas sensors is costly and inevitably affects the original gas distribution. So dense gas sampling in real-world environments is impractical. This article uses ANSYS/Fluid Flow (CFX) analysis software to simulate the diffusion and distribution of gases, with the leakage scenario set in an indoor substation. A geometric body of 10m x 10m x 5m is established, with 1.5m x 1m windows on all four sides for ventilation. Two of these windows are set as velocity inlets to simulate the air intake, while the other two were set as pressure outlets to simulate the air exhaust. For the velocity inlet, the velocity is set to 5 meters per second. For the pressure outlet, the pressure is set to one atmospheric pressure. The grid size was set to 0.2m, with a gravitational acceleration of -9.8m/s in the vertical direction. SF6 gas is chosen as the leak gas. SF6 is a commonly used insulating gas in high-voltage electrical equipment due to its excellent insulating performance and arc-extinguishing ability. However, if SF6 leaks, it can accumulate to a certain concentration in the lower space, causing staff to suffer from brain hypoxia, even rapid suffocation. Since there is no SF6 gas in the ANSYS/Fluid Flow (CFX) analysis database, the gas parameters need to be manually set, as shown in Table I. The location of the leak source and the inlet wind speed are set using the user-defined function (UDF). In UDF, we divided the 10m x 10m plane into 1m x 1m cells, resulting in a total of 100 cells. Within each cell, a point was randomly selected as the center coordinates for the leakage source.

During the computation process, we established sampling planes at the heights of 0.5 meters, and 1.5 meters. The mass fractions of SF6 at these three heights were recorded and saved. For each leakage location, we initiated sampling starting from the 30-second mark and continued to sample every 5 seconds thereafter. Each sampling session yielded distribution data for 2 different heights, resulting in a total of 12 sampling instances within 30 seconds.

In the end, 100 different leakage positions are set up, leading to the simulation of 1100 SF6 gas distribution images with labeled leak source positions. All the gas concentration data has been mapped into an RGB three-channel image as the network input because we found that three-channel RGB images can provide a richer texture and detail, allowing the network to capture more variations in details compared to single-channel image. These images are then divided into 1000 training sets and 100 test sets. To facilitate the division of patches, each image is resized to 224 x 224. Then, according to the two sampling strategies shown in Fig.1, masks are added to the images. The pixel values within each patch are set to the pixel value of the center point of the patch, representing the data from the SF6 gas sensors.

##### B. Implementation Details

The experimental environment is Ubuntu 22.04 with Python 3.10, and the GPU is GeForce RTX4060 Ti(16G)×2. The network is built with pytorch 2.1.1. The optimizer is AdamW, with a initial learning rate of 0.001. The batch size is set to 32 and the train epoch is 400. The warmup strategy (warmup iters = 40 epoch and warmup ratio=0.01) and CoinseLRScheduler learning rate scheduler are applied.

To evaluate our method, we choose L1 loss, SSIM, and PSNR as our experiment metric, which are usually used in image reconstruction and image super-resolution tasks. L1 loss calculates the absolute difference per pixel between the predicted image and the original image to measure reconstruction quality. SSIM is designed to assess the degree of distortion of an image, taking into account changes in brightness, contrast, and structure of the image. Compared with L1 loss, SSIM takes into account the visual characteristics of human eyes, so it is closer to human visual perception when assessing image quality. PSNR quantifies the difference between the



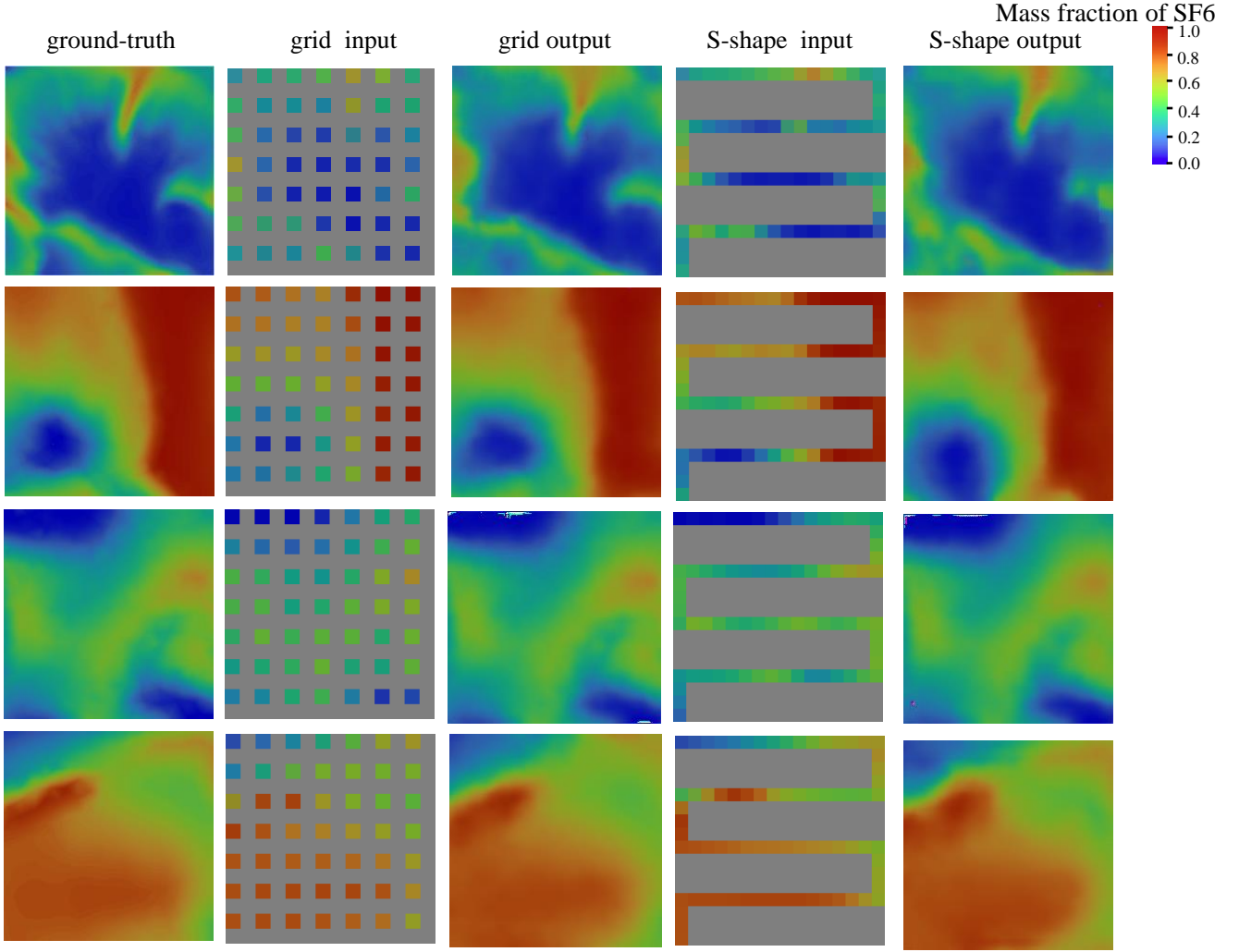


Fig. 4. Reconstruction results of AWED-B under two different sampling strategies.

reconstructed image and the original image in dB, and higher PSNR values generally correspond to better image quality.

### C. Experiment Result

1) *GDM Result*: The problem of high-resolution reconstruction of sparse gas sensor data can be regarded as an interpolation problem. Therefore, this article compares three traditional interpolation methods: linear interpolation, nearest interpolation, and cubic interpolation. In addition, four image reconstruction methods based on deep learning are also compared: MAE [49] uses the idea of mask training, adding random masks to the image and training the Transformer to predict the missing parts. TransUnet [25] and Segformer [50] are classic networks in the field of image segmentation, which are encoder-decoder structures for image-to-image tasks. DRCT [38] proposed the dense-residual-connected transformer to eliminate spatial information loss during encoding, which is currently the SOTA in the field of image super-resolution. To adapt to the task in this article, we replace the random mask strategy in MAE with the grid sampling strategy and S-shaped sampling strategy. The input of DRCT is  $7 \times 7$  sparse gas sensor

data, completing  $\times 32$  image super-resolution. For the model in this article, by changing the number of channels  $C$  in the low-dimensional feature map  $F_l$ , we construct three different-sized models: AWED-T ( $C=128$ ), AWED-B ( $C=512$ ), and AWED-L ( $C=1024$ ).

Table II presents the results of the methods mentioned above on the test set. Although traditional interpolation methods do not require training, their reconstruction errors are large. Methods based on deep learning, after being trained with a large number of samples, can reconstruct higher-quality gas distribution maps. Benefiting from the simplified encoder structure and the resource-efficient wavelet reconstruction module, AWED-T can achieve high-quality reconstruction effects with only 0.68MB of parameters. Its L1 loss is close to that of TransUnet with 93.23MB parameters and is higher than that of MAE, Segformer, and DRCT. Additionally, as the number of parameters increases, AWED-B and AWED-L achieve L1 losses of 0.0147 and 0.0143, respectively. In terms of SSIM and PSNR metrics, AWED-L also reaches the SOTA.

To comprehensively evaluate our model, we also compare several methods from the aspects of resource consumption

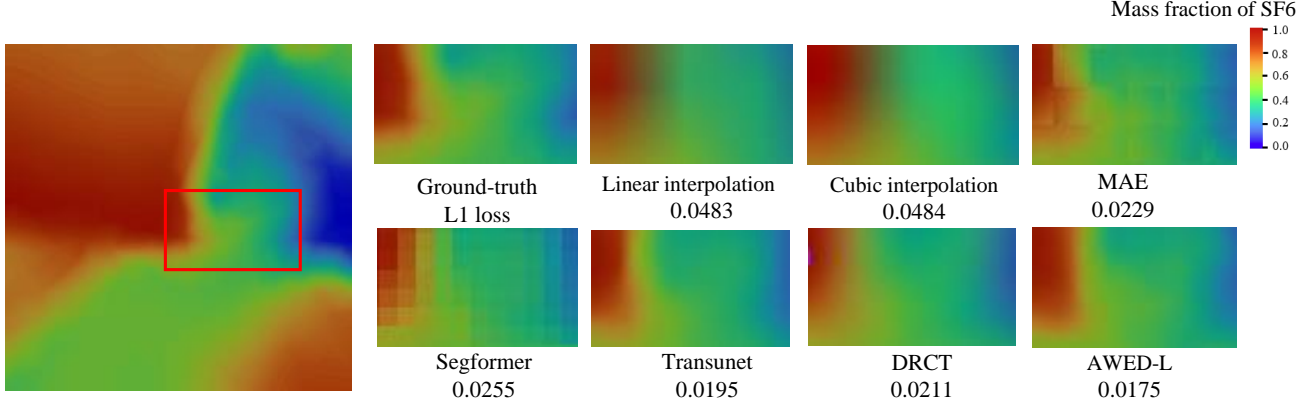


Fig. 5. Detail comparison of different methods.

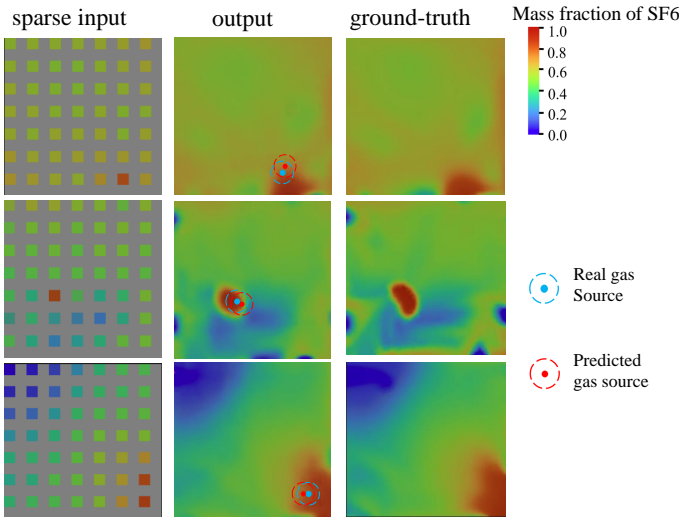


Fig. 6. Simultaneous GDM and GSL results with grid sampling strategy.

and real-time performance, as shown in Table III. FLOPs represents the number of floating-point operations, which can be used to measure the computational complexity of deep learning models. GPU Memory represents the amount of GPU memory consumed during model training. FPS represents the inference speed of the model, which measures its real-time performance. As can be seen, AWED-T has significantly lower FLOPs than other models. Training the model only requires 2.4G of GPU memory. On an RTX4060 Ti GPU, AWED-T achieves a FPS of 231, making it a resource-efficient and high-real-time model that can be easily deployed on edge computing devices. Whether for daily inspection or emergency detection, AWED-T can quickly estimate the gas distribution within the environment, aiding in rapid decision-making for emergency plans.

Fig. 4 demonstrates the reconstruction results of AWED-B in this article under two different sampling strategies. For the grid sampling, merely using the sparse sensor data from 49 points, the method can reconstruct a high-resolution gas

distribution image. For the S-shaped sampling, it is also able to accurately predict the missing data over large areas and long distances. Fig. 5 compares the details reconstructed by several methods, displaying magnified details of the reconstructed images within the red frames. It can be observed that the AWED reconstructed images exhibit superior accuracy and smoothness in the details compared to other methods.

2) *GSL Result*: As mentioned above, the geometric model simulates an indoor space of 10m x 10m, and the resolution of each gas distribution image is 224 x 224. Therefore, each pixel point can represent an area of approximately 0.045m x 0.045m. After calculating the pixel euclidean distance between the predicted leak source and the true leak source, we can convert the result to the actual distance to measure GSL precision.

In AWED, a key issue is to determine which layer of the decoder block information to feed into the SLH. Table IV explores the GSL experimental results obtained by using the output of the 3rd decoder block, the output of the 4th decoder block in the DMH, and the predicted reconstructed image as the input to the SLH. Using AWED-T, the output of decoder block 3 as the input to the SLH yields the best positioning accuracy, with a positioning error of 0.4136m. Finally, AWED-L achieves a positioning accuracy of 0.2652m.

Fig. 6 illustrates the results of simultaneously completing GDM and GSL under grid sampling. In the output image, the red dot represents the actual location of the gas leak source, while the blue dot represents the predicted location of the leak source.

TABLE IV  
COMPARISON OF LOCATION ERRORS

Model	SLH input	Location error (Euclidean distance)
AWED-T	Decoder block3	0.3348
	Decoder block4	0.4779
	Reconstructed image	0.7145
AWED-B	Decoder block3	0.2834
AWED-L	Decoder block3	<b>0.2652</b>



TABLE V  
ABLATION STUDY OF AWED-T AND AWED-B

Model	WRM	Wavelet L1	L1 loss↓	SSIM↑	PSNR↑
AWED-T	✓	✓	0.0181	0.8907	26.1118
			0.0162	0.9516	30.9440
	✓	✓	0.0175	0.9478	29.9661
			0.0158	0.9597	31.1594
AWED-B	✓	✓	0.0154	0.9594	30.9144
			0.0149	0.9618	31.3991
	✓	✓	0.0150	0.9613	31.1788
			0.0147	0.9640	31.4679

3) *Ablation Study*: To investigate the role of the wavelet reconstruction module and the wavelet L1 loss, we conduct ablation experiments using AWED-T and AWED-B. Table V demonstrates the reconstruction effects under grid sampling without incorporating WRM and Wavelet L1, without incorporating WRM, and without incorporating Wavelet L1. When the WRM and Wavelet L1 are added separately, they both contribute to reducing the L1 loss and increasing SSIM and PSNR. This suggests that both components individually enhance the reconstruction quality of the images. Furthermore, incorporating both WRM and Wavelet L1 yields the best reconstruction quality.

## V. CONCLUSION

This article proposes a method for constructing gas distribution maps and locating gas leak sources from sparse sensor data based on an asymmetric wavelet encoder-decoder model. Using a simplified encoder and an enhanced decoder, along with a wavelet reconstruction module, the method enables decoding from both the spatial domain and the frequency domain to improve reconstruction quality. A wavelet L1 loss is also proposed to calculate the L1 loss between the predicted and real images in the frequency domain, enhancing the similarity between the predicted and real images in the frequency domain. Using CFD-generated samples of SF<sub>6</sub> gas leakage distribution within a substation, an end-to-end framework is developed that simultaneously completes the tasks of constructing gas distribution maps and locating gas sources. Experiments demonstrate that with only 7x7 sensor data, the model can accurately reconstruct a 224x224 resolution gas distribution map and achieve a localization accuracy of 0.2652m within a 10m x 10m space. The model has fewer parameters, faster prediction speed, and stronger real-time performance. In the future, more diverse sampling strategies will be studied to assess their impact on model performance, optimizing sensor deployment plans, and applying the model to real gas leak detection scenarios to validate its effectiveness.

## REFERENCES

- [1] A. Gongora, J. Monroy, F. Rahbar, C. Ercolani, J. Gonzalez-Jimenez, and A. Martinoli, "Information-driven gas distribution mapping for autonomous mobile robots," *Sensors*, vol. 23, no. 12, 2023. [Online]. Available: <https://www.mdpi.com/1424-8220/23/12/5387> I
- [2] T. Kim and I. Park, "Gas distribution mapping and source localization: A mini-review," *Journal of Sensor Science and Technology*, vol. 32, no. 2, pp. 75–81, 2023. I
- [3] A. Yeon, R. Visvanathan, K. Kamarudin, S. Mamduh, A. Zakaria, and L. Kamarudin, "Gas source localization accuracy: A comparison between conventional, weighted arithmetic mean and kernel-based gas distribution mapping methods in small indoor area," in *IOP Conference Series: Materials Science and Engineering*, vol. 705, no. 1. IOP Publishing, 2019, p. 012034. I
- [4] T. Jing, Q.-H. Meng, and H. Ishida, "Recent progress and trend of robot odor source localization," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 16, no. 7, pp. 938–953, 2021. I
- [5] T. Wiedemann, A. J. Lilienthal, and D. Shutin, "Analysis of model mismatch effects for a model-based gas source localization strategy incorporating advection knowledge," *Sensors*, vol. 19, no. 3, p. 520, 2019. I
- [6] S. Mimouni, S. Fleau, and S. Vincent, "Cfd calculations of flow pattern maps and les of multiphase flows," *Nuclear Engineering and Design*, vol. 321, pp. 118–131, 2017. I
- [7] J. McNay and R. Hilditch, "Evaluation of computational fluid dynamics (cf) vs. target gas cloud for indoor gas detection design," *Journal of Loss Prevention in the Process Industries*, vol. 50, pp. 75–79, 2017. I
- [8] M. Awadalla, T.-F. Lu, Z. F. Tian, B. Dally, and Z. Liu, "3d framework combining cfd and matlab techniques for plume source localization research," *Building and Environment*, vol. 70, pp. 10–19, 2013. I
- [9] J. Waeytens, S. Durand, and S. Sadr, "Experimental validation of a cfd-based air quality sensor placement strategy to localize indoor source emissions," in *Building Simulation 2019*, 2019, p. 6p. I
- [10] Y. Fukazawa and H. Ishida, "Estimating gas-source location in outdoor environment using mobile robot equipped with gas sensors and anemometer," in *SENSORS, 2009 IEEE*. IEEE, 2009, pp. 1721–1724. I
- [11] N. P. Winkler, H. Matsukura, P. P. Neumann, E. Schaffernicht, H. Ishida, and A. J. Lilienthal, "Super-resolution for gas distribution mapping: convolutional encoder-decoder network," in *2022 IEEE International Symposium on Olfaction and Electronic Nose (ISOEN)*. IEEE, 2022, pp. 1–3. I
- [12] Y.-J. Zhang, Q.-H. Meng, T. Jing, Y.-K. Li, T.-S. Song, and Z.-Y. Deng, "Se-edsr: A deep learning method for gas distribution mapping," in *2022 41st Chinese Control Conference (CCC)*. IEEE, 2022, pp. 7076–7081. I
- [13] C. Hu, H. Kikumoto, B. Zhang, and H. Jia, "Fast estimation of airflow distribution in an urban model using generative adversarial networks with limited sensing data," *Building and Environment*, vol. 249, p. 111120, 2024. I
- [14] D. Badawi, I. Bassi, S. Ozev, and A. E. Cetin, "Deep-learning-based gas leak source localization from sparse sensor data," *IEEE Sensors Journal*, vol. 22, no. 21, pp. 20999–21008, 2022. I
- [15] Y. Sun, W. Chen, F. Li, Z. Gu, L. Feng, D. Guo, and H. Cai, "Improving gas distribution tomography: Evaluation of algorithms and fan-beam measurement geometries indoors," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2023. I
- [16] Z. H. M. Juffry, K. Kamarudin, A. H. Adom, M. F. Miskon, L. M. Kamarudin, A. Zakaria, S. Mamduh, and A. N. Abdullah, "Application of deep neural network for gas source localization in an indoor environment," *International Journal of Computers, Communications & Control*, vol. 18, no. 3, 2023. I
- [17] Y. Bi, Q. Wu, S. Wang, J. Shi, H. Cong, L. Ye, W. Gao, and M. Bi, "Hydrogen leakage location prediction at hydrogen refueling stations based on deep learning," *Energy*, vol. 284, p. 129361, 2023. I
- [18] X. Zhang, J. Shi, M. Yang, X. Huang, A. S. Usmani, G. Chen, J. Fu, J. Huang, and J. Li, "Real-time pipeline leak detection and localization using an attention-based lstm approach," *Process Safety and Environmental Protection*, vol. 174, pp. 460–472, 2023. I
- [19] V. S. P. Ruiz, P. Hinsin, T. Wiedemann, C. Christof, and D. Shutin, "Gas source localization using physics guided neural networks," *arXiv preprint arXiv:2405.04151*, 2024. I
- [20] B. Luo, Q.-H. Meng, J.-Y. Wang, and M. Zeng, "A flying odor compass to autonomously locate the gas source," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 1, pp. 137–149, 2018. I
- [21] Z. Yan, Q.-H. Meng, T. Jing, S.-W. Chen, and H.-R. Hou, "A deep learning-based indoor odor compass," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–10, 2023. I
- [22] U. Michelucci, "An introduction to autoencoders," *arXiv preprint arXiv:2201.03898*, 2022. I
- [23] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013. I
- [24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and*

- computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241. I
- [25] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, “Transunet: Transformers make strong encoders for medical image segmentation,” *arXiv preprint arXiv:2102.04306*, 2021. I, II, III, IV-C1
- [26] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Bm3d image denoising with shape-adaptive principal component analysis,” in *SPARS’09-Signal Processing with Adaptive Sparse Structured Representations*, 2009. II-A
- [27] R. Fattal, “Image upsampling via imposed edge statistics,” in *ACM SIGGRAPH 2007 papers*, 2007, pp. 95–es. II-A
- [28] R. E. Crochiere and L. R. Rabiner, “Interpolation and decimation of digital signals—a tutorial review,” *Proceedings of the IEEE*, vol. 69, no. 3, pp. 300–331, 1981. II-A
- [29] M. Unser, A. Aldroubi, and M. Eden, “Enlargement or reduction of digital images with minimum loss of information,” *IEEE Transactions on Image Processing*, vol. 4, no. 3, pp. 247–258, 1995. II-A
- [30] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*. Springer, 2014, pp. 184–199. II-A
- [31] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654. II-A
- [32] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 286–301. II-A
- [33] B. Niu, W. Wen, W. Ren, X. Zhang, L. Yang, S. Wang, K. Zhang, X. Cao, and H. Shen, “Single image super-resolution via a holistic attention network,” in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*. Springer, 2020, pp. 191–207. II-A
- [34] Y. Mei, Y. Fan, Y. Zhou, L. Huang, T. S. Huang, and H. Shi, “Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5690–5699. II-A
- [35] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020. II-A
- [36] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, “Swinir: Image restoration using swin transformer,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 1833–1844. II-A
- [37] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, “Restormer: Efficient transformer for high-resolution image restoration,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5728–5739. II-A
- [38] C.-C. Hsu, C.-M. Lee, and Y.-S. Chou, “Drct: Saving image super-resolution away from information bottleneck,” *arXiv preprint arXiv:2404.00722*, 2024. II-A, II, III, IV-C1
- [39] H. Ma, D. Liu, R. Xiong, and F. Wu, “iwave: Cnn-based wavelet-like transform for image compression,” *IEEE Transactions on Multimedia*, vol. 22, no. 7, pp. 1667–1679, 2019. II-B
- [40] S. Ruikar and D. Doye, “Image denoising using wavelet transform,” in *2010 International conference on mechanical and electrical technology*. IEEE, 2010, pp. 509–515. II-B
- [41] J. Bruna and S. Mallat, “Invariant scattering convolution networks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013. II-B
- [42] Q. Li, L. Shen, S. Guo, and Z. Lai, “Wavelet integrated cnns for noise-robust image classification,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 7245–7254. II-B
- [43] W.-Y. Hsu and W.-J. Wu, “Object detection using structure-preserving wavelet pyramid reflection removal network,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022. II-B
- [44] W.-Y. Hsu and P.-W. Jian, “Detail-enhanced wavelet residual network for single image super-resolution,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022. II-B
- [45] T. Guo, H. Seyed Mousavi, T. Huu Vu, and V. Monga, “Deep wavelet prediction for image super-resolution,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 104–113. II-B
- [46] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, “Multi-level wavelet-cnn for image restoration,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 773–782. II-B
- [47] H.-H. Yang and Y. Fu, “Wavelet u-net and the chromatic adaptation transform for single image dehazing,” in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 2736–2740. II-B
- [48] P. Jeevan, K. Viswanathan, A. Sethi *et al.*, “Wavemix: A resource-efficient neural network for image analysis,” *arXiv preprint arXiv:2205.14375*, 2022. II-B
- [49] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, “Masked autoencoders are scalable vision learners,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16 000–16 009. II, III, IV-C1
- [50] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, “Segformer: Simple and efficient design for semantic segmentation with transformers,” *Advances in neural information processing systems*, vol. 34, pp. 12 077–12 090, 2021. II, III, IV-C1