

GasUpper: Gas-Aware Upsampling for Enhanced Gas Segmentation

Yuting Lu^{1,2}, Xiaoyu Wang^{1,2}, Jingyi Cui^{1,2}, Le Yang³, Shunzhou Wang⁴, Yongqiang Zhao², and Binglu Wang^{1,2} ✉

ABSTRACT

Segmenting greenhouse gases from hyperspectral images can provide detailed information regarding their spatial distribution, which is significant for the monitoring of greenhouse gases. However, accurate segmentation of greenhouse gases is a challenging task due to two main reasons: (1) Diversity: greenhouse gases vary in concentration, size, and texture; (2) Camouflage: the boundaries between greenhouse gases and the surrounding background are blurred. Existing methods primarily focus on designing new modules to address the above challenges, often neglecting the design of the upsampling method within the model, which is crucial for achieving accurate segmentation. In this work, we propose Gas-Aware Upsampling (GasUpper), a novel and efficient upsampling method tailored for greenhouse gas segmentation. Specifically, we first generate a coarse segmentation mask during the upsampling process. Based on the roughly segmented gas and background, we then extract the global features of the gas and combine them with the original features to obtain de-camouflaged feature map that include both the global characteristics of the gas and the local details of the image. This de-camouflaged feature map serves as the foundation for subsequent point sampling. Finally, we utilize the de-camouflaged feature map to generate upsampling coordinate offsets, enabling the model to adaptively adjust the sampling regions based on the content during the sampling process. We conduct comprehensive evaluations by replacing the upsampling method in various segmentation approaches with GasUpper on two hyperspectral datasets. The results indicate that GasUpper consistently and significantly enhances the performance across all segmentation models (0.08%–9.44% Intersection over Union (IoU), 0.47%–6.26% Accuracy), outperforming other upsampling methods.

KEYWORDS

computer vision; greenhouse gas monitoring; hyperspectral image segmentation; upsampling method

Currently, the global environmental situation is becoming increasingly severe, and extreme weather events caused by climate change are increasing, including heat waves, floods, droughts, and wildfires^[1]. These phenomena pose significant threats to human lives and livelihoods, resulting in substantial human, economic, and environmental losses. Researches^[2,3] indicate that the release of greenhouse gases like carbon dioxide, methane, and nitrous oxide plays a key role in driving these environmental and climatic disruptions. Therefore, effectively monitoring and measuring these greenhouse gas emissions has become a crucial topic in environmental science and technology research^[4].

Hyperspectral image, with its excellent spectral resolution and rich band information, can capture the unique spectral characteristics of greenhouse gases. Consequently, leveraging the abundant spectral information within hyperspectral images to segment greenhouse gases provides a novel pathway for monitoring their emissions^[5-7]. It is a challenging task, however, due to two major reasons. Firstly, influenced by atmospheric conditions and environmental factors, as well as the varying chemical properties and sources of greenhouse gases, these gases often typically exhibit different distributions in hyperspectral images, such as concentrations, size, and texture, as shown in

Fig. 1a. The second is the camouflage effect of greenhouse gases. Under conditions of low gas concentrations or specific lighting and meteorological influences, greenhouse gases often exhibit spectral features, textures, and morphologies that seamlessly blend into the environmental backdrop in images, as shown in Fig. 1b. Especially in urban environments, shadows and reflections from buildings, as well as complex terrains, produce features similar to those of greenhouse gases, causing the boundaries between greenhouse gases and the background to become blurred and indistinct, making them difficult to identify. These issues result in the inaccurate segmentation of greenhouse gases.

While numerous methodologies have focused on refining modules to address the above challenges, they have inadvertently neglected the significance of upsampling operations^[8-10]. Feature upsampling holds a pivotal role in modern hyperspectral image segmentation models. Firstly, upsampling is crucial for resolution and spatial information recovery. The upsampling operation can gradually restore the low-resolution feature maps output by the encoder module to the original image resolution, enabling the decoder to produce segmentation results with clear contours and accurate positions^[11-14]. Secondly, in the decoder module, upsampling operations typically upsample low-resolution feature maps to high resolution to fuse them with feature maps from the

1 School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China

2 School of Automation, Northwestern Polytechnical University, Xi'an 710072, China

3 School of Electronics and Control Engineering, Chang'an University, Xi'an 710064, China

4 Electronic and Computer Engineering, Peking University Shenzhen Graduate School, Shenzhen 518055, China

Address correspondence to [Binglu Wang, wb1921129@gmail.com](mailto:Binglu.Wang_wb1921129@gmail.com)

© The author(s) 2025. The articles published in this open access journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

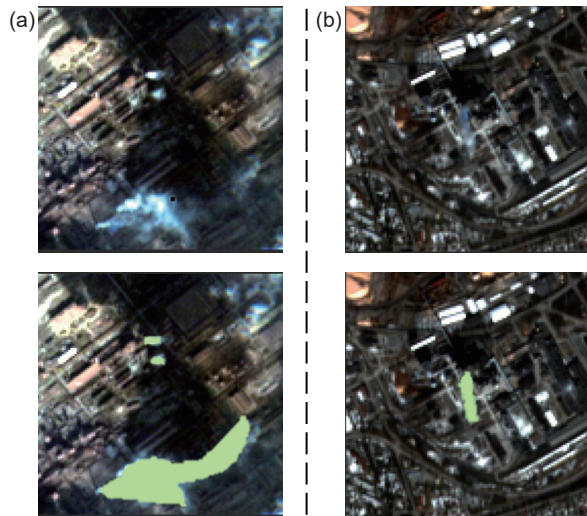


Fig. 1 Examples of greenhouse gases in hyperspectral images. We select three spectral bands from the hyperspectral images for visual display. (a) Greenhouse gases with different spatial distributions. (b) Greenhouse gases with blurred boundaries against the surrounding environment.

encoder module, thereby fully leveraging feature information at different levels^[15-18]. The most widely used feature upsampling operators are the nearest neighbor and bilinear interpolations, which determine new features according to the distance of spatial distance. However, nearest neighbor and bilinear interpolations only consider sub-pixel neighborhoods, failing to capture the rich semantic information of the gas which is usually regionally distributed on images. Some other learnable upsamplers introduce learnable parameters in upsampling. For example, the deconvolution^[19] layer works as an inverse operator of a convolution layer, which learns a set of instance-agnostic upsampling kernels. However, the deconvolution operator applies the same kernel over the entire image regardless of content,

limiting its ability to distinguish greenhouse gases from the surrounding environment. Pixel shuffle^[20] utilizes convolution to increase the number of preceding channels and then reshapes the feature maps to enhance resolution, but it also struggles with a fixed upsampling pattern. Overall, the aforementioned upsampling methods employ fixed patterns to process all regions uniformly, lacking the capability to perform adaptive processing based on distinctions between objects and background, as shown in Fig. 2. This limitation becomes particularly evident when dealing with hyperspectral image segmentation, where fine-grained discrimination between greenhouse gases and their surroundings is crucial. Camouflaged greenhouse gases often blend seamlessly into their backgrounds, necessitating an upsampling approach that can dynamically adapt to the subtle differences in textures, colors, and shapes that differentiate gases from their environment.

To overcome the aforementioned limitations, this paper proposes a novel feature upsampling method called Gas-Aware Upsampling (GasUpper), which enhances the distinguishability of gas features and background features and dynamically adjusts the upsampling areas based on image content to improve segmentation accuracy. The working mechanism of GasUpper is illustrated in Fig. 2. Specifically, GasUpper first distinguishes between the target (gas-containing regions) and background regions during the upsampling process and enhances the features of both to generate the de-camouflaged feature map. The enhancement of gas region features helps the model better recognize the targets by amplifying their prominent features. Meanwhile, the enhancement of background features aids in reducing mis-segmentation and mitigating the interference from background noise. This approach significantly enhances the distinguishability between the greenhouse gas regions and the background. Then, a dynamic adaptive sampling method is introduced to process the de-camouflaged feature map features, which generates adaptive upsampling coordinate offset values

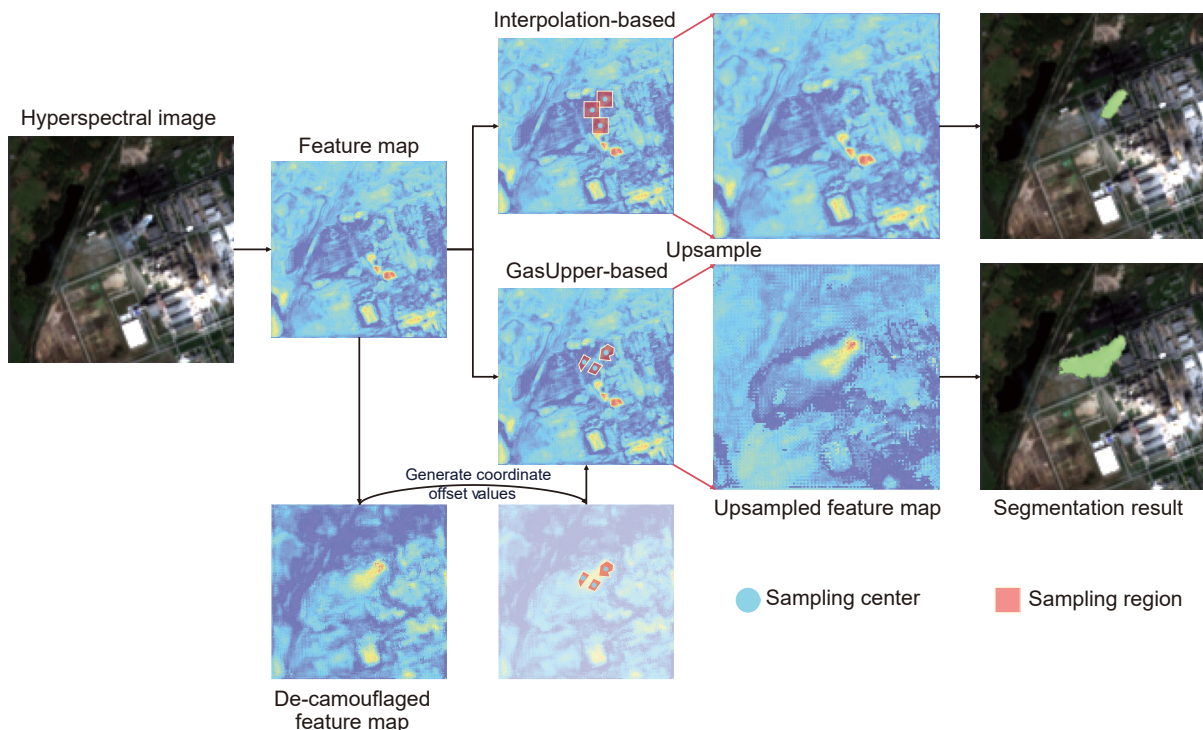


Fig. 2 Illustration of GasUpper working mechanism. Compared to interpolation-based upsampling methods with fixed sampling rules, GasUpper can dynamically adjust the sampling regions based on image content.

based on varying input features. Unlike traditional methods that rely on fixed sampling patterns, GasUpper can represent the distribution of gas, and magnify and capture the differences between gas regions and background regions, enabling the model to better predict segmentation results.

To validate the effectiveness of GasUpper, we conducted extensive experiments on two benchmark datasets. The results demonstrate that GasUpper outperforms traditional upsampling methods, achieving remarkable improvements in segmentation accuracy. Specifically, GasUpper can improve the performance of segmentation models by up to 9.44% Intersection over Union (IoU) on the industrial plume feature dataset^[21], and by up to 6.26% Accuracy on the power generation dataset^[22]. The significant improvements achieved demonstrate that GasUpper is an efficient feature upsampling method with the potential to serve as a foundation for future research.

In summary, this paper makes the following contributions:

- We propose a novel feature upsampling method, GasUpper, specifically designed for greenhouse gas segmentation. At each position, GasUpper can leverage the enhanced content information to predict the upsampling coordinate offset values, achieving more accurate segmentation masks.
- We introduce two key modules for GasUpper: Camouflage Removal Module (CRM) and Gas-Aware Adaptive Point Sampling Module (GAPSM). Specifically, CRM captures clues of greenhouse gases camouflaged in the environment by combining global gas features with local detailed information, thereby achieving camouflage removal for greenhouse gases. GAPSM utilizes the de-camouflaged feature map to generate adaptive sampling coordinate offset values, which guide the upsampling process.
- Through extensive experiments, we demonstrate that GasUpper substantially improves segmentation accuracy compared with other upsampling methods, e.g., 9.44% IoU improvement on the industrial plume feature dataset, making it a promising upsampling method.

1 Related Work

1.1 Hyperspectral image segmentation

The field of hyperspectral image segmentation has seen significant advancements in recent years. Early efforts, such as the edge-embedded marker-based watershed algorithm proposed by Li et al.^[23], utilized edge information to enhance segmentation accuracy and boundary precision. This approach was further refined by the same group with an improved watershed algorithm that incorporated pre- and post-processing techniques to mitigate over-segmentation issues^[24]. As deep learning progressed, Kampffmeyer et al.^[25] designed a deep convolutional neural network tailored for semantic segmentation in hyperspectral images, focusing on accurate mapping in urban environments while addressing class imbalance challenges. Building on this, Su and Zhang^[26] introduced methods to evaluate segmentation performance from both local and global perspectives, providing quantifiable insights into error distribution. FusionNet^[27] further advanced this domain by combining sea-land segmentation with ship detection through deep convolutional networks, integrating edge-aware regularization to refine segmentation outcomes. Innovations continued with Kemker et al.^[28], who developed a novel framework for hyperspectral image segmentation, employing synthetic data for network initialization to counter the challenges of label scarcity. Rusyn et al.^[29] made significant contributions by

addressing the specific challenge of cloud forecasting through cloud image segmentation techniques. Motiyani et al.^[30] proposed a new clustering method that utilizes k-means to sequentially perform feature reduction, segmentation, and clustering on hyperspectral images, aiming to improve segmentation accuracy. Medellín et al.^[31] introduced a new method that combines the Arbitrary Segment Modeling (M-SAM) with the Spectral Angle Mapping (H-SAM) algorithm for hyperspectral image semantic segmentation. Wang and Zhang^[32] proposed a multi-scale differential feature optimization network that leverages spatial information relationships and high-low semantic features between bi-temporal images to improve recognition accuracy.

To further enhance segmentation accuracy and efficiency in high-resolution imagery, various innovative strategies have been proposed. Convolutional networks have been effectively leveraged to extract subtle features and define precise boundaries, as demonstrated in dynamic multicontext segmentation^[33] and boundary loss optimization^[34]. Additionally, novel deep learning architectures such as ResUNet-a^[35] and HQ-ISNet^[36] have been specifically designed to address the complexities of instance segmentation in hyperspectral images. The incorporation of attention mechanisms, as exemplified by studies like MANet^[37] and UNetFormer^[38], has significantly improved the capture of both local and global contextual information, leading to enhanced segmentation outcomes. Pan et al.^[39] has designed a set of Deep Dual-Resolution Networks (DDRNs) for real-time and accurate semantic segmentation, which consist of a deep dual-resolution backbone and an enhanced low-resolution context information extractor.

These works have also tackled critical issues such as sparse annotations^[40], domain shifts^[41], and the need for real-time processing^[38], thus advancing the capabilities of semantic segmentation techniques in hyperspectral imaging. Despite these advancements, many existing hyperspectral image segmentation methods still rely on conventional upsampling techniques with fixed sampling rules. Unlike other segmented targets, greenhouse gases in hyperspectral images exhibit diverse and camouflaged spatial properties, requiring a more carefully designed approach to upsample to achieve precise and reliable results.

1.2 Upsampling methods

Upsampling is a critical technique in signal and image processing, which is mainly used to upscale image data from lower to higher resolutions.

In deep learning and computer vision, it plays a pivotal role in pixel-level tasks such as image segmentation, aiming to reconstruct high-resolution images or feature maps from their lower-resolution versions. Traditional upsampling methods, such as interpolation, rely on fixed algorithms that perform uniform interpolation between pixel points to increase resolution. These approaches often disregard the semantic content of the images or feature maps, focusing solely on mathematical manipulation of pixel values^[42]. While methods like deconvolution^[43] can introduce undesirable checkerboard artifacts^[43], and unpooling techniques^[44] can result in the loss of essential positional information, thereby hindering accurate reconstruction of original data.

In recent years, dynamic upsampling approaches, including CARAFE^[45], FADE^[46], SAPA^[47], and DySample^[48] have gained traction for their ability to incorporate learning-based adjustments during upsampling. These advanced methods preserve finer details while reducing noise and inconsistencies.

Despite their advantages, most existing upsampling methods treat objects and backgrounds uniformly, which will lead to

blurred edges in the resulting images. Our objective is to develop an upsampling method that differentiates between gases and backgrounds, providing sharper and more distinct boundaries in areas with rich edge information.

2 Methodology

2.1 Overview

Greenhouse gases in atmospheric environments often exhibit diverse and camouflaged properties, varying shapes, and indistinct boundaries, making precise segmentation extremely challenging. While existing methods have made significant strides in addressing these challenges, they often rely on fixed upsampling patterns, which can struggle to adapt to the subtle differences between gas regions and their surroundings. To overcome these limitations, we introduce GasUpper, a novel upsampling method that dynamically adjusts the sampling process based on the content of the input feature map, enabling more accurate and refined gas segmentation. The key innovation of GasUpper lies in its two-stage adaptive mechanism: the CRM distinguishes between gas-containing regions and the background by combining global gas features with local image details, while the GAPSM dynamically adjusts the sampling regions based on the content of the input feature map. This adaptive approach enables GasUpper to capture subtle differences in texture, color, and shape, which are often overlooked by conventional methods, making it particularly effective in addressing the challenges of greenhouse gas segmentation, such as the camouflage effect and diverse spatial distributions of gases in hyperspectral images. By combining global context with local detail and introducing a dynamic adaptive sampling mechanism, GasUpper represents a significant advancement over existing upsampling methods, offering a more accurate solution for hyperspectral image segmentation.

The overall framework of GasUpper is shown in Fig. 3. We first feed the original image into the encoder to generate an initial feature map. Next, we process the feature map through the CRM to produce a global feature map. Then we combine the global feature map with the original image feature through a residual connection to result in a de-camouflaged feature map that integrates both the global context of greenhouse gas and the background feature, as well as the local details of the original image. CRM allows the model to dynamically distinguish whether the sampling area corresponds to the background or the gas of interest. Finally, we upsample the original image feature to the target size by the GAPSM.

2.2 Camouflage removal module

Upsampling is a process of gradually restoring spatial information

by enhancing the resolution of an image to recover more details. However, due to the often blurred boundaries between greenhouse gases and their surrounding environment, confusion can easily arise during the upsampling process, leading to a loss of detail in the enlarged image of greenhouse gases. Therefore, to better restore the details in the image after upsampling, it is necessary to accurately distinguish between greenhouse gases and the background during this process.

The CRM is designed to distinguish between camouflaged greenhouse gas and background regions, generating a de-camouflaged feature map X_D that integrates both gas and background features. Specifically, given a feature map X of size $C \times H \times W$, we first use a 1×1 convolution layer to predict a segmentation mask, where the mask has two channels representing gas and background, respectively. To determine the probability of each pixel belonging to gas or background, we apply a softmax normalization to the mask to generate a probability map. This process can be expressed as

$$M = \text{Softmax}(f(X)) \quad (1)$$

where $M \in \mathbb{R}^{2 \times H \times W}$ denotes the computed segmentation mask and $f(\cdot)$ represents the convolution function.

Once the segmentation mask M is generated, it is decomposed into a gas mask M_G and a background mask M_B in the channel dimension. However, at this stage, both the gas and background masks contain continuous values. To further clarify whether the region is gas or background, we apply a threshold τ to binarize these probabilities, identifying confident gas and background localizations for enhancement while leaving uncertain pixels near boundaries for subsequent processing. The choice of the threshold τ is crucial for effective upsampling. If τ is set excessively high, gas regions with strong camouflage characteristics may be overlooked. Conversely, a low τ could introduce noise by misclassifying the background as gas. Through experiments, we determined that the optimal threshold for effectively balancing these trade-offs is 0.75.

Due to the blurred boundary between greenhouse gases and the background in hyperspectral images, global features are needed to provide overall semantic information, such as spectral response, intensity distribution, and spatial proportions. These features help the model distinguish between gas and background regions. To achieve this, we introduce global average pooling to generate global feature vectors. We first apply a weighting operation to the feature map, reducing noise and enhancing the model's attention on distinguishing between gas and background. And then, we average the weighted feature map across spatial dimensions. The approach extracts comprehensive gas and background features without being influenced by spatial positioning. The resulting global feature vectors are denoted as V_G and V_B , which represent

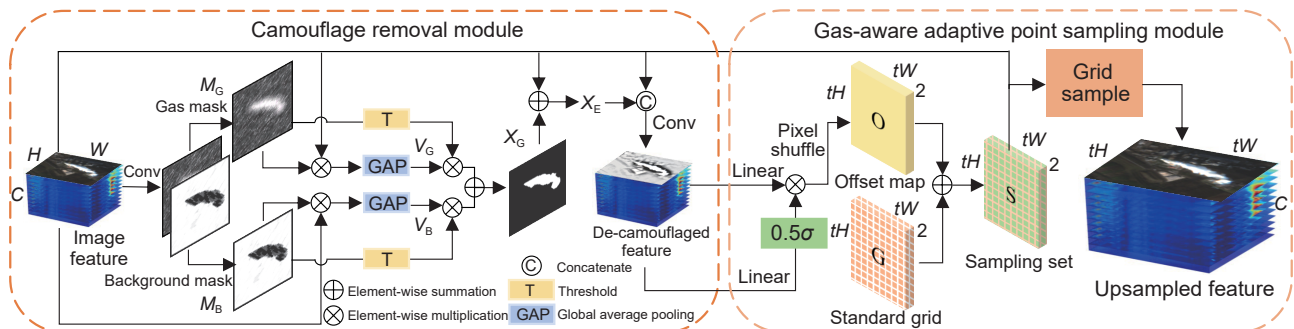


Fig. 3 Overall framework of the GasUpper. GasUpper comprises two main components: the CRM and the GAPSM. The CRM enhances feature maps by distinguishing gas from background, while the GAPSM adjusts the upsampling grid based on pixel characteristics.

the global features of the entire gas and background regions, respectively:

$$V_G = \text{GAP}(X \odot M_G) \quad (2)$$

$$V_B = \text{GAP}(X \odot M_B) \quad (3)$$

where $\text{GAP}(\cdot)$ represents the global average pooling operation.

Next, the binarized gas mask and background mask are combined with the respective gas and background feature vectors to generate the global feature map X_G . This operation enhances feature information by emphasizing the most relevant regions:

$$X_G = (V_G \cdot (M_G > \tau) + V_B \cdot (M_B > \tau)) \quad (4)$$

where τ is the threshold.

Subsequently, to enhance gradient flow and inject global context information without losing the local details of the original image, we apply a residual connection between the global feature map X_G and the original feature map X to produce an enhanced feature map X_E :

$$X_E = X + X_G \quad (5)$$

Given the camouflage characteristics of greenhouse gases in natural atmospheric environments, we further enhance the feature representation to generate a de-camouflaged feature map X_D by concatenating the feature map X_E with the original feature map X . This operation enables the model to adaptively select and combine features across multiple levels, preserving local detail information. However, this approach increases the channel dimension, so we apply a fusion convolution layer to reduce the number of channels and flexibly adjust the weights of different feature channels, achieving a more adaptive feature fusion. The entire process can be expressed by

$$X_D = F(\text{concat}(X, X_E)) \quad (6)$$

where $\text{concat}(\cdot)$ represents concatenation along the channel dimension and F denotes the convolutional layer used to fuse the

concatenated features.

2.3 Gas-aware adaptive point sampling module

After obtaining the de-camouflaged feature map X_D from the CRM, we need to generate the sampling point set for the target image based on the feature map as shown in Fig. 4. Given an upsampling factor t , we first use a linear layer with input and output channel sizes C and $2s^2$ on the feature map X_D to compute offset values for each pixel. To constrain the magnitude of the offset, we apply a sigmoid function to compress the output values to the range $[0, 1]$. These values are then scaled by 0.5 to further limit the offset range to $[0, 0.5]$ to ensure that sampling points remain within a reasonable vicinity. This operation produces an offset map O with dimensions $2t^2 \times H \times W$, and then a pixel shuffle operation is applied to reshape O into dimensions $2 \times tH \times tW$. Next, the offset map O is combined with the initial sampling positions grid G of the feature map X to generate the final set of sampling points S . The entire process can be expressed as

$$O = \text{Sigmoid}(\text{Linear}(X_D)) \times 0.5 \quad (7)$$

$$S = O + G \quad (8)$$

The sampling set $S \in \mathbb{R}^{2 \times tH \times tW}$ represents the position of each pixel in the target feature relative to the input feature. However, to generate the target size image feature $X' \in \mathbb{R}^{C \times tH \times tW}$, we utilize a grid sampling function that maps the coordinates in S to the corresponding positions in the input feature $X \in \mathbb{R}^{C \times H \times W}$. For each sampling position, the value is computed using bilinear interpolation, which estimates the value by considering the four nearest integer pixel values in the input feature.

$$X' = \text{GridSample}(X, S) \quad (9)$$

As shown in Fig. 5, the target values obtained through the original bilinear upsampling method differ significantly from those generated by the GAPSM's offset-based bilinear interpolation. In origin bilinear upsampling, the position of each

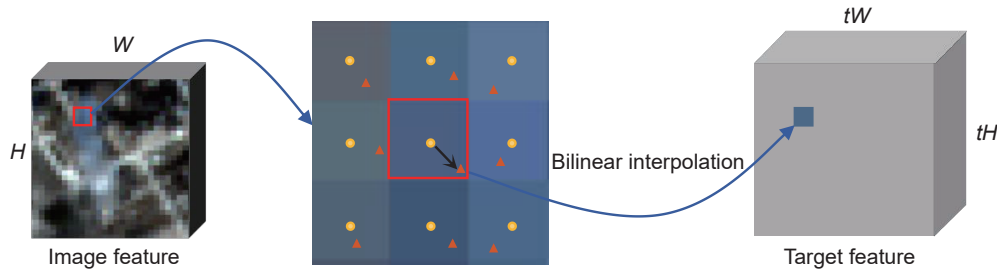


Fig. 4 Illustration of interpolation based on sampling set generated by gas-aware adaptive point sampling. The yellow circles represent the original sampling points, while the red triangles indicate the new sampling points.

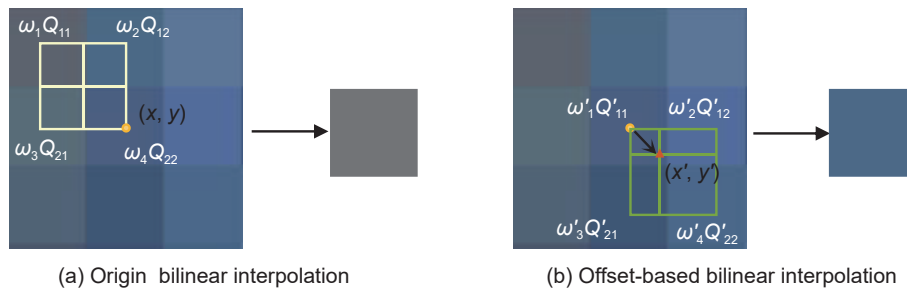


Fig. 5 Comparison with bilinear interpolation upsampling method. Q represents the values at the four neighboring grid points, w denotes the interpolation weights corresponding to these grid points, and (x, y) indicates the sampling point.

target pixel is fixed and mapped to the original features based on a regular scaling ratio, ignoring the relative positional relationships of neighboring points Q_{ij} . The sampling points are solely influenced by the fixed scaling factor. In contrast, the new sampling point (x', y') generated by our method dynamically adjust their weights w'_i based on predicted offsets. This allows the sampling process to adapt to the content of the input feature.

2.4 Training and inference

Training process. During training, since decoders typically have multi-layer structure, our method predicts a segmentation mask at each layer. And for each layer, we calculate cross-entropy loss, which improves the quality of the masks and guides the model to progressively and precisely distinguish between gas and background. These individual losses are then aggregated to derive the final training loss, utilizing the supervisory signals from all decoder layers. The loss calculation can be expressed as

$$L_{\text{total}} = \sum_{i=1}^N L_{\text{CE}}(P_i, G_i) \quad (10)$$

where N represents the number of decoder layers, $L_{\text{CE}}(\cdot, \cdot)$ denotes the cross-entropy loss function, P_i represents the prediction result of the i -th layer, and G_i represents the ground truth segmentation mask.

Inference process. During inference, we utilize the parameters learned during training to dynamically generate pixel-wise sampling coordinate offsets from the input low-resolution features. And then adopt the adaptive point upsampling to produce the high-resolution output.

3 Experiment

3.1 Datasets

Among the available public datasets, we choose two high-quality gas segmentation datasets: the Characterization of Industrial Smoke Plumes from Remote Sensing Data (CISPRSD)^[21] and the Power Generation Data Set (PGDS)^[22].

CISPRSD. This dataset was collected from ESA's Sentinel-2 earth-observing satellite constellation^[21]. Each raster image contains all 13 spectral-band channels from the calibrated Level-2A reflectances and is cropped to a dimension of 1200 meters on each side, equivalent to 120×120 pixels. Low-resolution channels are resampled to the highest available resolution (10 m/pixel). The images are categorized into gas types for segmentation. Based on the methodology of Mommert et al.^[21], we select 996 images for training and 214 images for testing, with both sets padded to a size of 512×512 pixels.

PGDS. This dataset is also acquired from ESA's Sentinel-2 earth-observing satellite constellation^[22]. Each image contains 13 spectral bands and has dimensions of either 120×120 pixels or 300×300 pixels. The images are labeled into gas categories for segmentation. Following the approach of Hanna et al.^[22], we select 1498 images for training and 73 images for testing, with all images padded to 512×512 pixels.

3.2 Instantiation of GasUpper

To evaluate the effectiveness of our proposed method, we integrate the GasUpper into various segmentation models and conducted extensive experiments on CISPRSD and PGDS. We select a diverse set of recent and well-performing segmentation models for comparison. Each model is individually trained and finely tuned

to establish robust baselines, ensuring a thorough evaluation. Specifically, we select three representative models from CNN-based methods and two representative models from ViT-based methods, based on their performance metrics and architectural diversity. The selection provides a comprehensive assessment of the GasUpper method across different model types. We replace the original upsampling methods with GasUpper to validate its effectiveness.

U-Net + GasUpper. In the U-Net model^[49], the encoder employs a layer-by-layer downsampling strategy to capture image features at different scales. Specifically, in its four levels, features are first extracted using 3×3 convolutions, followed by 2×2 max pooling operations to reduce the spatial dimensions of the feature maps. During the subsequent upsampling part, the feature map dimensions are gradually restored to their original sizes.

U-Net employs a 2×2 transposed convolution kernel for this upsampling strategy, which effectively doubles the height and width of the feature maps, restoring them to their original dimensions. Finally, the upsampled feature maps are concatenated with the corresponding downsampled feature maps to recover spatial information. Since the scale of each upsampling operation in U-Net is consistent, we can directly replace the four upsampling operations in the U-Net model with GasUpper, setting the upsampling scale to 2. This ensures that the feature maps in each decoder stage are uniformly magnified by a factor of two, achieving the same resolution as the corresponding encoder stage feature maps. The process is illustrated in Fig. 6a.

DeepLabv3 + GasUpper. In DeepLabv3+^[50], the upsampling component is enhanced by the Atrous Spatial Pyramid Pooling (ASPP) technology. The ASPP module processes high-level features, which are then reduced in dimension using a 1×1 convolution. These features are upsampled by a factor of four using bilinear interpolation and fused with corresponding low-level features from the encoder. Following this, an additional convolution layer refines the features, and the refined feature map is upsampled again to the original input size using bilinear interpolation, producing a detailed segmentation map.

To improve this process, we replace the two bilinear interpolation upsampling operations with GasUpper, as shown in Fig. 6b. This replacement maintains consistent parameters but leverages GasUpper's dynamic and precise upsampling capabilities. The first replacement occurs after the 1×1 convolution, dynamically adjusting the upsampling to ensure spatial integrity and enhance feature map quality. The second replacement is at the final upsampling stage.

DDRNet + GasUpper. The DDRNet^[39] employs a dual-resolution network, where the high-resolution branch maintains a $1/8$ resolution to preserve fine details, and the low-resolution branch gradually reduces to $1/16$, $1/32$, and $1/64$ to capture global context. Specifically, the low-resolution features are progressively upsampled to the high-resolution size using $2 \times$ bilinear interpolation. In both the high-resolution branch and the fused feature map, $8 \times$ upsampling generates intermediate segmentation predictions for training supervision. Finally, the fused $1/8$ resolution features are upsampled $8 \times$ to the original input resolution for the final segmentation output. As shown in Fig. 6c, we replace DDRNet's bilinear upsampling with GasUpper, while maintaining consistent upsampling factors to ensure comparability.

SegFormer + GasUpper. In contrast to U-Net, the SegFormer model^[51] upsampling procedure does not use a fixed scaling factor. Instead, it incorporates multi-scale feature fusion. The process accepts feature maps from four different scales, originating from

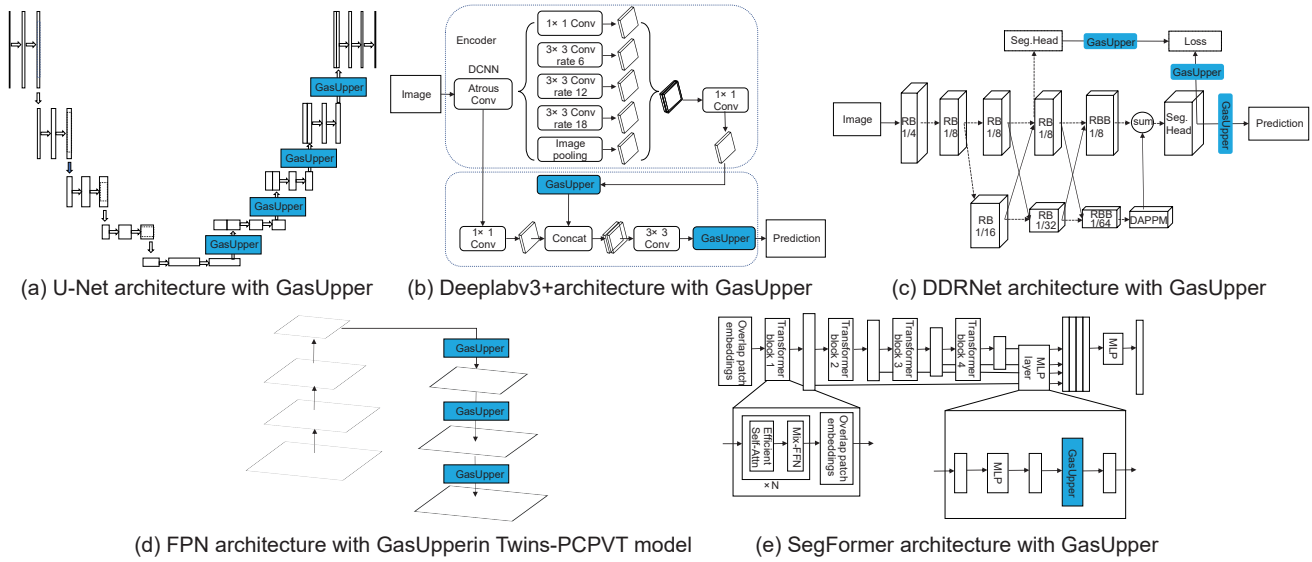


Fig. 6 Illustration of GasUpper integrated into the segmentation models.

the backbone (1/4, 1/8, 1/16, 1/32). Subsequently, it uses bilinear interpolation to upsample these feature maps to the size of the initial stage (1/4 scale). In our approach, we replace the original bilinear interpolation upsampling with GasUpper. As illustrated in Fig. 6d, we dynamically set the target scales for upsampling to (1, 2, 4, 8), resulting in feature maps that are upsampled to a resolution of 1/4 of the original image size.

Twins-PCPVT + GasUpper. The Twins-PCPVT model^[52] utilizes a Feature Pyramid Network (FPN)^[53], which consists of two main components: a down-up path and a top-down path. The down-up path extracts hierarchical features through a series of standard convolutional operations, while the top-down path facilitates feature fusion via bilinear interpolation upsampling and skip connections. Specifically, bilinear interpolation is employed for upsampling to align the dimensions of the low-level feature maps. Following this, a 1×1 convolution adjusts the channel numbers to ensure consistency between the upscaled and low-level feature maps. Subsequently, element-wise addition is used for feature fusion. As illustrated in Fig. 6e, the traditional bilinear interpolation upsampling can be seamlessly replaced with the GasUpper method without additional operations. This replacement maintains the upsampling scale at 2, ensuring consistency with previous scales.

3.3 Implementation details

Experimental setup. We follow the training schedule outlined in MMSegmentation^[54]. For a fair comparison of method performance, all experiments are implemented with PyTorch and run on RTX 3090 Ti. The training involves a total of 80 000 iterations. We utilize the SGD optimizer, configured with a momentum of 0.9 and a weight decay of 0.0005. The batch size is set to 4, and the initial learning rate is 0.001. Data augmentation is applied by randomly scaling images between 0.5 and 2.0, followed by random cropping and flipping. Additionally, image edges are padded to ensure square input images.

Evaluation metrics. To clearly demonstrate the segmentation performance for gas regions, we evaluate the model using the IoU, Accuracy (Acc), F1-Score, and Precision metrics only for the gas regions, as recommended by Lu et al.^[47] IoU measures the overlap between the predicted and ground truth values, while Acc reflects the proportion of correctly classified pixels out of the total. In addition, the F1-Score, which balances precision and recall, is used

to evaluate the model's ability to correctly identify gas regions while minimizing false positives and false negatives. Precision, on the other hand, quantifies the proportion of correctly predicted gas pixels relative to all pixels predicted as gas, providing insight into the model's reliability in avoiding false detections. These metrics collectively offer a comprehensive evaluation of the model's segmentation performance. Furthermore, to assess computational efficiency and model complexity, we measure Floating Point Operations (FLOPs) and Parameters (Params), which provide insights into the computational load and memory usage during runtime.

3.4 Experimental results

Results of the instantiation models. The performance results of the GasUpper-based instance models evaluated on the CISPRSD and PGDS datasets are comprehensively presented in Table 1. Following the replacement of upsampling with GasUpper, the U-Net model exhibits a substantial improvement across both datasets. On the CISPRSD dataset, U-Net with GasUpper achieves a 5.09% increase in IoU and a 5.26% increase in Accuracy. Similarly, on the PGDS dataset, U-Net with GasUpper shows a 7.68% improvement in IoU and an 8.16% increase in Accuracy. These significant improvements underscore the efficacy of GasUpper in enhancing the segmentation capabilities of the U-Net model, particularly in handling the complex and camouflaged nature of greenhouse gases in hyperspectral images. For the DeepLabv3+ model, the integration of GasUpper also yields notable improvements. These results highlight the adaptability of GasUpper to different model architectures.

The SegFormer model, which is based on a transformer architecture, also benefits significantly from the integration of GasUpper. On the CISPRSD dataset, SegFormer with GasUpper achieves a 5.57% increase in IoU and a 6.66% increase in Accuracy. On the PGDS dataset, the improvements are more modest but still significant, with a 0.73% increase in IoU and a 1.62% increase in Accuracy. These results demonstrate that GasUpper is not only effective in CNN-based models but also in transformer-based architectures, which are increasingly popular in modern segmentation tasks. The Twins-PCPVT model, which utilizes a Feature Pyramid Network (FPN) for multi-scale feature fusion, also shows consistent improvements with GasUpper. Although the improvements in IoU were modest, the consistent

Table 1 Gas segmentation results on the CISPRSD and PGDS.

Backbone	Dataset	Method	IoU (%)	Acc (%)	F1-score (%)	Precision (%)	FLOPs	Params
U-Net-S	CISPRSD	U-Net ^[49]	67.94	76.17	80.91	86.28	0.204×10 ¹²	28.99×10 ⁶
		U-Net+GasUpper	73.03 (+5.09)	81.43 (+5.26)	84.41 (+3.50)	87.63 (+1.35)	0.214×10 ¹²	32.60×10 ⁶
	PGDS	U-Net ^[49]	61.24	66.87	75.96	87.92	0.204×10 ¹²	28.99×10 ⁶
		U-Net+GasUpper	68.92 (+7.68)	75.03 (+8.16)	81.60 (+5.64)	89.42 (+1.50)	0.214×10 ¹²	32.60×10 ⁶
ResNet-50	CISPRSD	DeepLabv3plus ^[50]	73.02	79.64	84.41	89.78	0.177×10 ¹²	41.22×10 ⁶
		DeepLabv3plus+GasUpper	74.29 (+1.27)	86.72 (+6.63)	85.25 (+0.84)	83.83 (−5.95)	0.179×10 ¹²	76.19×10 ⁶
	PGDS	DeepLabv3plus ^[50]	74.15	80.89	85.16	89.89	0.177×10 ¹²	41.22×10 ⁶
		DeepLabv3plus+GasUpper	76.27 (+2.12)	83.36 (+2.47)	85.93 (+0.77)	91.91 (+2.02)	0.179×10 ¹²	76.19×10 ⁶
DDRNet	CISPRSD	DDRNet ^[39]	75.07	81.62	85.76	90.34	5.735×10 ⁶	4.729×10 ⁹
		DDRNet+GasUpper	76.30 (+1.23)	82.34 (+0.72)	86.56 (+0.80)	91.23 (+0.89)	5.889×10 ⁶	5.132×10 ⁹
	PGDS	DDRNet ^[39]	74.25	79.65	85.22	91.64	5.735×10 ⁶	4.729×10 ⁹
		DDRNet+GasUpper	74.71 (+0.46)	79.69 (+0.04)	85.52 (+0.30)	92.27 (+0.63)	5.889×10 ⁶	5.132×10 ⁹
MiT-B	CISPRSD	SegFormer ^[51]	62.86	73.11	77.19	81.75	8.116×10 ⁹	3.73×10 ⁶
		SegFormer+GasUpper	68.43 (+5.57)	79.77 (+6.66)	81.26 (+4.07)	82.80 (+1.05)	11.248×10 ⁹	4.74×10 ⁶
	PGDS	SegFormer ^[51]	75.96	82.55	86.41	87.56	8.116×10 ⁹	3.73×10 ⁶
		SegFormer+GasUpper	76.69 (+0.73)	84.17 (+1.62)	86.79 (+0.38)	89.79 (+2.23)	11.248×10 ⁹	4.74×10 ⁶
PCPVT-S	CISPRSD	Twins ^[52]	76.18	85.81	86.48	87.16	44.491×10 ⁹	27.85×10 ⁶
		Twins+GasUpper	76.31 (+0.13)	86.90 (+1.09)	86.56 (+0.08)	86.22 (−0.94)	45.093×10 ⁹	28.15×10 ⁶
	PGDS	Twins ^[52]	77.30	84.58	87.20	89.99	44.491×10 ⁹	27.85×10 ⁶
		Twins+GasUpper	77.76 (+0.46)	86.05 (+1.47)	87.49 (+0.29)	88.97 (−1.02)	45.093×10 ⁹	28.15×10 ⁶

gains in Accuracy and F1-score across both datasets suggest that GasUpper enhances the model's ability to correctly classify gas regions, even in challenging scenarios. Finally, the DDRNet, which is designed for real-time and accurate semantic segmentation, also benefits from the integration of GasUpper. These improvements not only validate the broad applicability of GasUpper across different model architectures but also highlight its potential for improving the performance of state-of-the-art segmentation models.

Furthermore, as shown in Table 1, GasUpper introduces additional FLOPs and parameters compared to the baseline models. For instance, in the U-Net architecture, GasUpper increases the FLOPs from 0.204×10¹² to 0.214×10¹² and the number of parameters from 28.99×10⁶ to 32.60×10⁶. Despite these increases in computational complexity, the performance improvements justify the additional cost, particularly in applications where accurate segmentation of greenhouse gases is critical. For example, U-Net with GasUpper achieves a 5.09% increase in IoU and a 5.26% increase in Accuracy on the CISPRSD dataset. The increased computational cost of GasUpper can be attributed to its dynamic adaptive sampling mechanism and the Camouflage Removal Module, which enhance the model's ability to distinguish between gas regions and complex backgrounds. While these modules introduce additional FLOPs and parameters, they are crucial for addressing the unique challenges of greenhouse gas segmentation, such as the camouflage effect and blurred boundaries in hyperspectral images.

Comparison with other upsampling methods. In order to thoroughly evaluate the validity of our proposed GasUpper in comparison to several widely used upsampling techniques in hyperspectral image segmentation, including nearest neighbor, bilinear interpolation, transposition convolution, and pixel shuffle, we conduct a series of comprehensive experiments using U-Net

and Twins as baseline models. Each of these upsampling operators is systematically integrated into the U-Net and Twins to replace its standard upsampling method, and the modified U-Net and Twins models were trained and tested on the CISPRSD.

The results of these comparative experiments are detailed in Table 2. Our proposed GasUpper method outperforms other upsampling operations on most metrics. Specifically, in the context of U-Net, GasUpper achieved an IoU of 73.03%, which represents a significant improvement over the other methods. Similarly, when applied to the Twins model, GasUpper delivered an IoU of 76.31%, outperforming the next best method by a substantial margin.

In addition to the quantitative improvements, the qualitative results presented in Fig. 7 further validate the advantages of GasUpper. The visual comparisons, conducted using U-Net as the baseline model, demonstrate that GasUpper consistently generates clearer and more accurate segmentation maps compared to other upsampling methods, such as nearest neighbor, bilinear interpolation, deconvolution, and pixel shuffle. As shown in Fig. 7, GasUpper effectively reduces mis-segmentation errors, particularly in regions where gas plumes overlap with complex backgrounds, such as urban structures or natural terrain. For instance, in cases where traditional upsampling methods fail to distinguish between gas regions and background noise, GasUpper successfully identifies and isolates the gas plumes with higher precision. This is achieved through its dynamic adaptive sampling mechanism, which adjusts the sampling regions based on the content of the input feature map, thereby enhancing the model's ability to capture subtle differences in texture, color, and shape between gas regions and their surroundings. Furthermore, the visual results highlight GasUpper's ability to preserve fine-grained details in the segmentation maps, which is critical for applications requiring high spatial accuracy, such as environmental monitoring

Table 2 Gas segmentation results of different upsampling methods on the CISPRSD using U-Net and Twins model, with the best performance highlighted in bold.

Model	Method	IoU (%)	Acc (%)	F1-score (%)	Precision (%)	FLOPs	Params
U-Net	Nearest	68.00	74.77	80.95	88.25	0.204 $\times 10^{12}$	28.992 $\times 10^6$
	Bilinear	67.94	76.17	80.91	86.28	0.204 $\times 10^{12}$	28.992 $\times 10^6$
	Deconv	71.07	79.50	83.09	87.02	0.234 $\times 10^{12}$	39.788 $\times 10^6$
	Pixel Shuffle	71.09	79.00	83.10	87.65	0.359 $\times 10^{12}$	79.135 $\times 10^6$
	GasUpper	73.03	81.43	84.41	87.63	0.214 $\times 10^{12}$	32.603 $\times 10^6$
Twins	Nearest	73.32	83.97	84.60	85.25	44.469 $\times 10^9$	27.849 $\times 10^6$
	Bilinear	76.18	85.81	86.48	87.16	44.491 $\times 10^9$	27.849 $\times 10^6$
	Deconv	76.25	86.95	86.53	86.11	45.418 $\times 10^9$	28.243 $\times 10^6$
	GasUpper	76.31	86.90	86.56	86.22	45.093 $\times 10^9$	28.147 $\times 10^6$

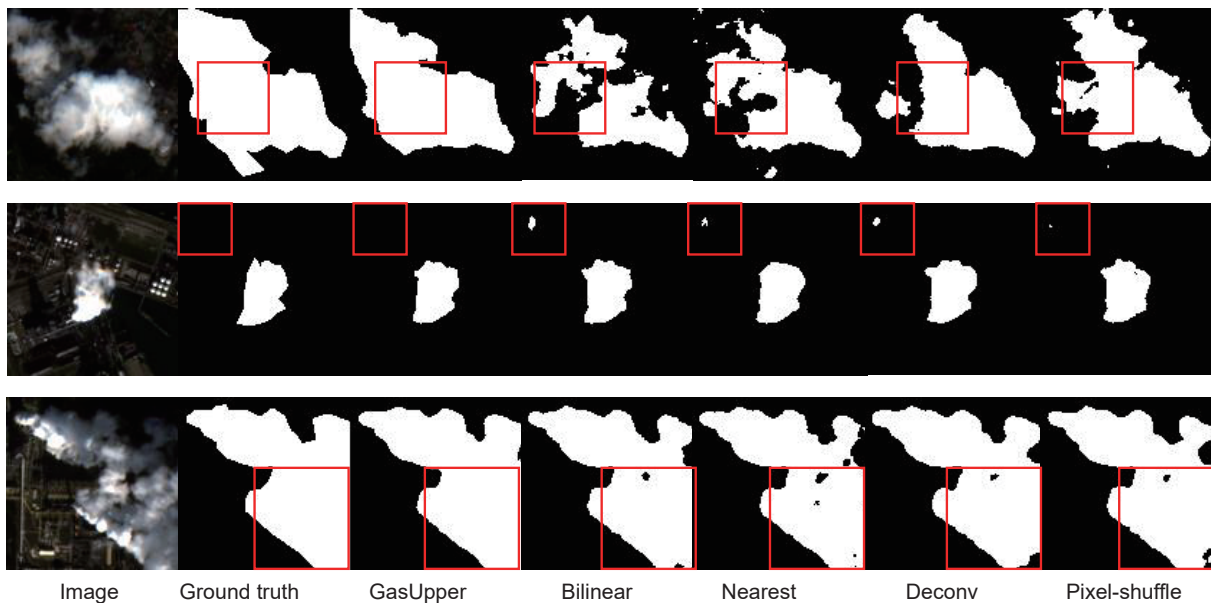


Fig. 7 Qualitative results of different upsampling operators on hyperspectral image segmentation tasks. Note that only the RGB channels of the hyperspectral image are visualized for reference purposes.

and greenhouse gas emission tracking. In contrast, other upsampling methods often produce segmentation maps with blurred or fragmented gas regions, leading to inaccuracies in the final output. The qualitative analysis thus underscores GasUpper's effectiveness in handling the unique challenges of greenhouse gas segmentation, making it a promising approach for hyperspectral image analysis tasks. These visual comparisons, combined with the quantitative metrics, provide strong evidence of GasUpper's effectiveness and its potential to advance the state-of-the-art in gas segmentation.

3.5 Ablation studies

In this section, we conduct ablation experiments using U-Net as the baseline network on the CISPRSD. The primary objective is to assess the impact of crucial modules and identify the optimal settings for several key parameters.

Effect of camouflage removal module. To assess the effectiveness of the CRM, we conduct a detailed comparative analysis under controlled training conditions. By maintaining identical training schemes across all experiments, we ensure that any observed differences in performance could be directly attributed to the incorporation of CRM. The results presented in Table 1, clearly demonstrate that the CRM significantly enhances

the model performance in the Gas category. Specifically, the IoU and Acc for the Gas category show notable improvements of 8.09% and 3.1%. These results strongly validate the role of CRM in enhancing the segmentation capabilities of our method.

Choice of parameters. We begin by investigating the impact of different scaling factors on the performance of our proposed method. Specifically, we evaluate the performance using scaling factors of 0.25, 0.5, and 0.75. The results, as depicted in Table 3, indicate that a scaling factor of 0.5 yields the best results, achieving an IoU of 73.03%. Additionally, we explore the influence of various threshold settings on the performance of method. We test thresholds set at 0.25, 0.5, and 0.75. As shown in Table 3, the threshold of 0.75 provides the optimal performance.

4 Limitation and Future Work

While our proposed GasUpper method significantly improves the performance of greenhouse gas segmentation in hyperspectral images, there are still several limitations that need to be addressed. Firstly, in the process of dealing with greenhouse gas camouflage effects, GasUpper has not been integrated into specialized Camouflage Object Detection (COD)^[55-57] models for testing. Due to the limited space of this article, readers are referred to Fan et al.⁷

Table 3 Ablation studies: Key module and parameters analysis.

Ablation		IoU (%)	Acc (%)
CRM	×	64.94	78.32
	√	73.03	81.42
Scaling factor	0.25	73.03	81.73
	0.50	73.03	81.42
	0.75	68.33	73.17
Threshold	0.25	72.72	80.98
	0.50	72.24	79.45
	0.75	73.03	81.42

s researches^[38,39] for more detailed progress on COD. Secondly, the computational efficiency of GasUpper needs to be further optimized. The integration of global features and the adaptive sampling process may introduce additional computational overhead.

In future research, we intend to integrate GasUpper with state-of-the-art camouflage object detection models. This integration will involve thorough testing and validation to ensure that GasUpper enhances the detection capabilities of these models, particularly in complex and challenging scenarios where gases are difficult to discern. Additionally, to improve the computational efficiency of GasUpper, we will focus on optimizing the integration of global features and the adaptive sampling process. This will involve exploring more efficient algorithms for feature extraction and reducing the complexity of the adaptive sampling mechanism.

5 Conclusion

We propose the Gas-Aware Upsampling method that enhances feature representation during the upsampling process by integrating gas global and image local information, and adaptively adjusts the sampling regions based on the image content, thereby achieving more accurate greenhouse gas segmentation. GasUpper demonstrates consistent performance improvements across multiple benchmark datasets, particularly with significant enhancements in IoU (0.08%–9.44%) and Acc (0.47%–6.26%). Looking ahead, GasUpper's applicability could extend beyond environmental monitoring to fields such as medical imaging and multispectral satellite image analysis.

Article History

Received: 28 November 2024; Revised: 23 January 2025; Accepted: 5 February 2025

References

- [1] P. Mangal, A. Rajesh, and R. Misra, Big data in climate change research: Opportunities and challenges, in *Proc. Int. Conf. Intelligent Engineering and Management (ICIEM)*, London, UK, 2020, pp. 321–326.
- [2] S. Kirschke, P. Bousquet, P. Ciais, M. Saunois, J. G. Canadell, E. J. Dlugokencky, P. Bergamaschi, D. Bergmann, D. R. Blake, L. Bruhwiler, et al., Three decades of global methane sources and sinks, *Nat. Geosci.*, vol. 6, no. 10, pp. 813–823, 2013.
- [3] R. K. Pachauri and L. A. Meyer, *Climate Change 2014: Synthesis Report*. Geneva, Switzerland: IPCC, 2014.
- [4] E. D. Sherwin, J. S. Rutherford, Z. Zhang, Y. Chen, E. B. Wetherley,

- P. V. Yakovlev, E. S. F. Berman, B. B. Jones, D. H. Cusworth, A. K. Thorpe, et al., US oil and gas system emissions from nearly one million aerial site measurements, *Nature*, vol. 627, no. 8003, pp. 328–334, 2024.
- [5] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, SpectralFormer: Rethinking hyperspectral image classification with transformers, *IEEE Trans. Geosci. Remote Sens.*, vol. 60, p. 5518615, 2021.
- [6] H. Zhang, H. Chen, G. Yang, and L. Zhang, LR-Net: Low-rank spatial-spectral network for hyperspectral image denoising, *IEEE Trans. Image Process.*, vol. 30, pp. 8743–8758, 2021.
- [7] J. Zhang, Z. Cai, F. Chen, and D. Zeng, Hyperspectral image denoising via adversarial learning, *Remote Sens.*, vol. 14, no. 8, p. 1790, 2022.
- [8] C. Chen, Y. Wang, N. Zhang, Y. Zhang, and Z. Zhao, A review of hyperspectral image super-resolution based on deep learning, *Remote. Sens.*, vol. 15, no. 11, p. 2853, 2023.
- [9] M. Zhang, C. Zhang, Q. Zhang, J. Guo, X. Gao, and J. Zhang, ESSAformer: Efficient transformer for hyperspectral image super-resolution, in *Proc. IEEE/CVF Int. Conf. Computer Vision*, Paris, France, 2023, pp. 23016–23027.
- [10] M. Rostami and A. A. Beheshti Shirazi, Hyperspectral image super-resolution via learning an undercomplete dictionary and intra-algorithmic postprocessing, *IEEE Trans. Geosci. Remote Sens.*, vol. 61, p. 5512115, 2023.
- [11] D. P. Fan, T. Zhou, G. P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, Inf-Net: Automatic COVID-19 lung infection segmentation from CT images, *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2626–2637, 2020.
- [12] D. P. Fan, G. P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, PraNet: parallel reverse attention network for polyp segmentation, in *Proc. 23rd Int. Conf. Medical Image Computing and Computer Assisted Intervention*, virtual, 2020, pp. 263–273.
- [13] D. P. Fan, G. P. Ji, G. Sun, M. M. Cheng, J. Shen, and L. Shao, Camouflaged object detection, in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2020, pp. 2777–2787.
- [14] H. Mei, G. P. Ji, Z. Wei, X. Yang, X. Wei, and D. P. Fan, Camouflaged object segmentation with distraction mining, in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 8772–8781.
- [15] G. P. Ji, Y. C. Chou, D. P. Fan, G. Chen, H. Fu, D. Jha, and L. Shao, Progressively normalized self-attention network for video polyp segmentation, in *Proc. 24th Int. Conf. Medical Image Computing and Computer Assisted Intervention*, Strasbourg, France, 2021, pp. 142–152.
- [16] Y. Lv, J. Zhang, Y. Dai, A. Li, B. Liu, N. Barnes, and D. P. Fan, Simultaneously localize, segment and rank the camouflaged objects, in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 11591–11601.
- [17] J. Pei, T. Cheng, D. P. Fan, H. Tang, C. Chen, and L. Van Gool, OSFormer: one-stage camouflaged instance segmentation with transformers, in *Proc. 17th European Conf. Computer Vision (ECCV)*, Tel Aviv, Israel, 2022, pp. 19–37.
- [18] Y. Lv, J. Zhang, Y. Dai, A. Li, N. Barnes, and D. P. Fan, Toward deeper understanding of camouflaged object detection, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 7, pp. 3462–3476, 2023.
- [19] H. Noh, S. Hong, and B. Han, Learning deconvolution network for semantic segmentation, in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, Santiago, Chile, 2015, pp. 1520–1528.
- [20] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D.

- Rueckert, and Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 1874–1883.
- [21] M. Mommert, M. Sigel, M. Neuhausler, L. Scheibenreif, and D. Borth, Characterization of industrial smoke plumes from remote sensing data, arXiv preprint arXiv: 2011.11344, 2020.
- [22] J. Hanna, M. Mommert, L. M. Scheibenreif, and D. Borth, Multitask learning for estimating power plant greenhouse gas emissions from satellite imagery, in *Proc. 35th Conf. Neural Information Processing Systems (NeurIPS)*, Montreal, Canada, 2021, pp. 739–758.
- [23] D. Li, G. Zhang, Z. Wu, and L. Yi, An edge embedded marker-based watershed algorithm for high spatial resolution remote sensing image segmentation, *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2781–2787, 2010.
- [24] B. Li, M. Pan, and Z. Wu, An improved segmentation of high spatial resolution remote sensing image using Marker-based Watershed Algorithm, in *Proc. 20th Int. Conf. Geoinformatics*, Hong Kong, China, 2012, pp. 1–5.
- [25] M. Kampffmeyer, A. B. Salberg, and R. Jenssen, Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, Las Vegas, NV, USA, 2016, pp. 680–688.
- [26] T. Su and S. Zhang, Local and global evaluation for remote sensing image segmentation, *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 256–276, 2017.
- [27] D. Cheng, G. Meng, S. Xiang, and C. Pan, FusionNet: edge aware deep convolutional networks for semantic segmentation of remote sensing harbor images, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 12, pp. 5769–5783, 2017.
- [28] R. Kemker, C. Salvaggio, and C. Kanan, Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning, *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 60–77, 2018.
- [29] B. Rusyn, R. Kosarevych, O. Lutsyk, and V. Korniy, Segmentation of atmospheric clouds images obtained by remote sensing, in *Proc. 14th Int. Conf. Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, Lviv-Slavske, Ukraine, 2018, pp. 213–216.
- [30] H. Motiyani, P. K. Mali, and A. Mehta, Hyperspectral image segmentation, feature reduction and clustering using k-means, in *Proc. Int. Conf. Computing, Communication, and Intelligent Systems (ICCCIS)*, Greater Noida, India, 2022, pp. 389–393.
- [31] A. Medellin, D. Grabowsky, D. Mikulski, and R. Langari, Sam-Sam - A novel approach to hyperspectral image semantic segmentation, in *Proc. 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Athens, Greece, 2023, pp. 1–5.
- [32] J. Wang and L. Zhang, The multiscale differential feature optimization networks for remote sensing image change detection, *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 17, pp. 16847–16859, 2024.
- [33] K. Nogueira, M. Dalla Mura, J. Chanussot, W. R. Schwartz, and J. A. dos Santos, Dynamic multicontext segmentation of remote sensing images based on convolutional networks, *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7503–7520, 2019.
- [34] A. Bokhovkin and E. Burnaev, Boundary loss for remote sensing imagery semantic segmentation, in *Proc. 16th Int. Symp. Neural Networks (ISNN 2019)*, Moscow, Russia, 2019, pp. 388–401.
- [35] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data, *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, 2020.
- [36] H. Su, S. Wei, S. Liu, J. Liang, C. Wang, J. Shi, and X. Zhang, HQ-ISNet: High-quality instance segmentation for remote sensing imagery, *Remote Sens.*, vol. 12, no. 6, p. 989, 2020.
- [37] R. Li, S. Zheng, C. Zhang, C. Duan, J. Su, L. Wang, and P. M. Atkinson, Multiattention network for semantic segmentation of fine-resolution remote sensing images, *IEEE Trans. Geosci. Remote Sens.*, vol. 60, p. 5607713, 2021.
- [38] L. Wang, R. Li, C. Zhang, S. Fang, C. Duan, X. Meng, and P. M. Atkinson, UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery, *ISPRS J. Photogramm. Remote Sens.*, vol. 190, pp. 196–214, 2022.
- [39] H. Pan, Y. Hong, W. Sun, and Y. Jia, Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes, *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 3, pp. 3448–3460, 2023.
- [40] Y. Hua, D. Marcos, L. Mou, X. X. Zhu, and D. Tuia, Semantic segmentation of remote sensing images with sparse annotations, *IEEE Geosci. Remote. Sens. Lett.*, vol. 19, p. 8006305, 2021.
- [41] H. Wang, C. Tao, J. Qi, R. Xiao, and H. Li, Avoiding negative transfer for semantic segmentation of remote sensing images, *IEEE Trans. Geosci. Remote Sens.*, vol. 60, p. 4413215, 2022.
- [42] P. Thevenaz, T. Blu, and M. Unser, Interpolation revisited [medical images application], *IEEE Trans. Med. Imaging*, vol. 19, no. 7, pp. 739–758, 2000.
- [43] A. Odena, V. Dumoulin, and C. Olah, Deconvolution and checkerboard artifacts, *Distill*, vol. 1, no. 10, p. e3, 2016.
- [44] M. D. Zeiler and R. Fergus, Visualizing and understanding convolutional neural networks, in *Proc. 13th European Conf. Computer Vision (ECCV)*, Zurich, Switzerland, 2019, pp. 818–833.
- [45] J. Wang, K. Chen, R. Xu, Z. Liu, C. C. Loy, and D. Lin, CARAFE: Content-aware ReAssembly of FEatures, in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, Seoul, Republic of Korea, 2019, pp. 3007–3016.
- [46] H. Lu, W. Liu, H. Fu, and Z. Cao, FADE: Fusing the assets of decoder and encoder for task-agnostic upsampling, in *Proc. 17th European Conf. Computer Vision (ECCV)*, Tel Aviv, Israel, 2022, pp. 231–247.
- [47] Hao Lu, Wenze Liu, Zixuan Ye, Hongtao Fu, Yuliang Liu, and Zhiguo Cao, Sapa: Similarity-aware point affiliation for feature upsampling, in *Proc. 36th Conf. Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, USA, 2022, pp. 20889–20901.
- [48] W. Liu, H. Lu, H. Fu, and Z. Cao, Learning to upsample by learning to sample, in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, Paris, France, 2023, pp. 6027–6037.
- [49] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in *Proc. 18th Int. Conf. Medical Image Computing and Computer Assisted Intervention*, Munich, Germany, 2015, pp. 234–241.
- [50] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in *Proc. 15th European Conf. Computer Vision (ECCV)*, Munich, Germany, 2018 pp. 833–851.
- [51] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, Segformer: Simple and efficient design for semantic segmentation with transformers, in *Proc. 35th Conf. Neural Information Processing Systems (NeurIPS)*, virtual, 2021, pp. 12077–12090.

- [52] X. Chu, Z. Tian, Y. Wang, B. Zhang, H. Ren, X. Wei, H. Xia, and C. Shen, Twins: Revisiting the design of spatial attention in vision transformers. in *Proc. 35th Conf. Neural Information Processing Systems (NeurIPS)*, Virtual, 2021, pp. 9355–9366.
- [53] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, Feature pyramid networks for object detection, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 936–944.
- [54] Contributors. MMSegmentation, MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark, 2020.
- [55] A. Khan, M. Khan, W. Gueaieb, A. El Saddik, G. De Masi, and F. Karray, CamoFocus: Enhancing camouflage object detection with split-feature focal modulation and context refinement, in *Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 2024, pp. 1423–1432.
- [56] G. P. Ji, D. P. Fan, Y. C. Chou, D. Dai, A. Liniger, and L. Van Gool, Deep gradient learning for efficient camouflaged object detection, *Mach. Intell. Res.*, vol. 20, no. 1, pp. 92–108, 2023.
- [57] J. Zhao, X. Li, F. Yang, Q. Zhai, A. Luo, Z. Jiao, and H. Cheng, FocusDiffuser: perceiving local disparities for camouflaged object detection, in *Proc. 18th European Conf. Computer Vision (ECCV)*, Milan, Italy, 2024, pp. 181–198.
- [58] D. P. Fan, G. P. Ji, M. M. Cheng, and L. Shao, Concealed object detection, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6024–6042, 2022.
- [59] D. P. Fan, G. P. Ji, P. Xu, M. M. Cheng, C. Sakaridis, and L. Van Gool, Advances in deep concealed scene understanding, *Vis. Intell.*, vol. 1, no. 1, p. 16, 2023.