Mathematical Problems of Computer Science 62, 93–111, 2024.

doi:10.51408/1963-0124

UDC 004.932.721

A Multispectral Decomposition and Frequency-Based Framework for Salient Object Detection in Remote Sensing Images

Hayk. A. Gasparyan

Yerevan State University, Yerevan, Armenia
e-mail: hayk.gasparyan@ysu.am

Abstract

Salient object detection (SOD) aims to identify the most visually prominent objects in images, crucial for tasks like image segmentation, visual tracking, autonomous navigation, and photo cropping. While SOD has been extensively studied in natural scene RGB images, detecting salient objects in remote sensing images remains underexplored due to varying spatial resolutions and complex scenes.

This paper presents a novel framework for SOD called Multispectral Decomposition Network (**MSD-Net**) in remote sensing 3-band RGB images, combining Multispectral Decomposition and Frequency-based Saliency detection. The framework includes three key steps: (i) Multispectral Decomposition: Decomposing a 3-band RGB image into 32 multispectral bands to enhance feature capture across spectral domains; (ii) Synthetic RGB Reconstruction: Using a new entropy-based measure to select the most informative bands in salient regions by analyzing frequency domain and constructing synthetic RGB image; and (iii) Saliency Fusion and Object Detection: training a segmentation network on the fusion of synthetic RGB image and input image for improved accuracy. Comprehensive evaluations of public datasets demonstrate that the proposed method performs better than state-of-the-art (SOTA) models and offers a robust solution for detecting salient objects in complex remote sensing images by integrating multispectral and frequency-based techniques.

Keywords: Saliency map, Object detection, Multispectral decomposition, Band selection, Remote sensing, Entropy.

Article info: Received 10 October 2024; sent for review 19 October 2024; accepted 26 November 2024.

Acknowledgments: This work was partly supported by the ADVANCE Research Grant provided by the Foundation for Armenian Science and Technology, which was funded by Sarkis and Nune Sepetjians. I also thank Professor S. Agaian for his invaluable guidance and support throughout this project.

Automatic monitoring systems utilizing remote sensing technologies, such as satellite and UAV imagery, have encountered significant challenges in recent years. Remote sensing image processing [1] has numerous applications, including environmental monitoring, surveillance, military operations, autonomous navigation, and visual tracking. Image content analysis is critical across all these applications, encompassing object detection, localization, segmentation, and classification. The complexities and challenges in these tasks stem from varying environmental conditions, inconsistent image quality, and the diverse range of objects or regions requiring analysis. To address these challenges, recent approaches have leveraged properties of the human visual system. Humans possess an innate ability to automatically identify regions of interest within complex scenes through the visual attention mechanism. Inspired by this capability, salient object detection (SOD) enables computers to simulate this behavior, allowing them to detect the most prominent and important objects or regions in a scene automatically. SOD's adaptability and efficiency make it valuable in various applications, including foreground annotation, image enhancement, segmentation, image quality assessment, and video summarization.

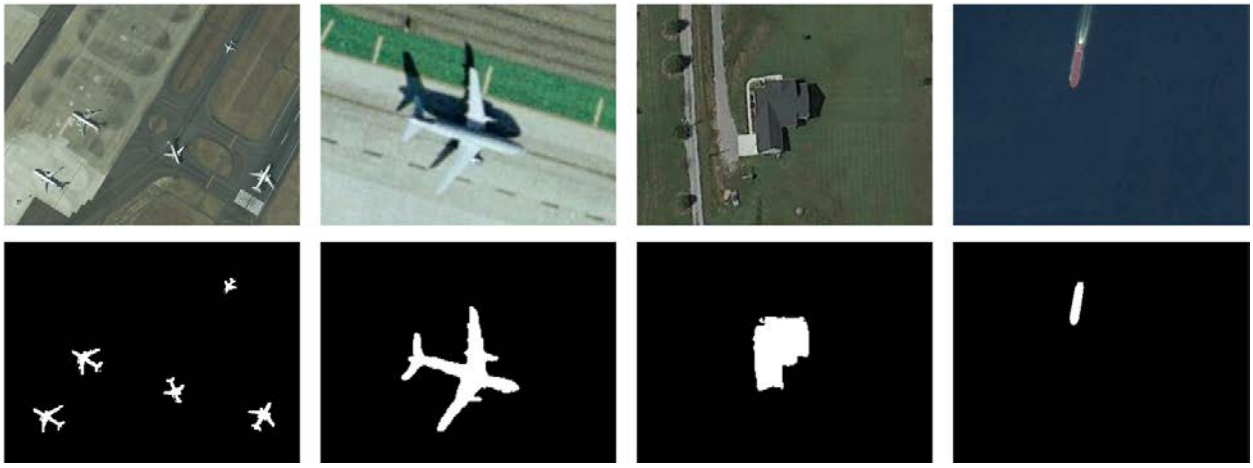

Fig. 1. Examples of SOD. Input and expected outputs are in the first and second rows, respectively.

While recent advances in signal processing have somehow solved these issues by allowing systems to detect predefined classes of objects with high accuracy, a more intricate problem arises in saliency object detection (SOD). Unlike standard detection tasks, where the system is searching for known objects, SOD involves identification of unknown objects or regions of interest [2]. Some example images of salient objects and their corresponding masks are illustrated in Figure 1. In [3], the authors have conducted an excellent review of the challenges in SOD and existing solutions, as well as their pros and cons. While SOD has been extensively studied in natural scene images, its application in remote sensing images brings more challenges, such as varying spatial resolutions, highly heterogeneous and complex scenes, and the presence of background clutter and noise in an image.

Early studies in salient object detection (SOD) showed promising results by using basic, handcrafted features like image contrast and background information to identify important regions in images. These early methods provided a foundation for the development of SOD techniques, and a detailed review of these non-deep learning approaches can be found in [4]. An interesting method proposed by [5] involved extracting spectral residuals from the frequency domain by analyzing the log spectrum of the image. This process helped to create a saliency map in the spatial domain. However, these methods had some limitations: (i) they struggled with complex textures

and fine details, which led to blurred results, and (ii) it was sensitive to noise and image artifacts, which affected their accuracy. Another recent algorithm, based on image contrast, was proposed by [6]. This method used global contrast to detect saliency by separating large objects from their surroundings. It assigned similar saliency values to similar regions, which allowed the entire object to be highlighted evenly. The saliency of each region was mainly based on its contrast with nearby areas, while distant contrasts had less influence. Although this method had difficulties with low image contrast, it was hard to distinguish objects from their backgrounds. These early methods were limited, especially in handling complex textures and low-contrast images. Further advancements were needed to improve their effectiveness and reliability.

With the success of deep learning technologies in computer vision, an increasing number of deep learning-based SOD methods [7] have emerged. Early deep SOD models generally utilized multi-layer perceptron (MLP) classifiers to predict saliency scores based on deep features extracted from individual pixels. These models significantly outperformed traditional, non-deep learning SOD methods. However, the MLP-based models were limited in their ability to capture spatial information effectively, as they lacked the structure to account for spatial dependencies across the image.

Table 1. Existing methods and limitations.

	SRS	**GCR**	**Conv-NN**	**ViT**	**MSD-Net**
Simple and efficient	+	+	-	-	+
Can handle complex structures	-	+	+	+	+
Robust to noise and contrast variations	+	-	+	+	+
High accuracy	-	-	+	+	+
Does not require large training data	+	+	-	-	+
Low computational complexity	+	+	+	-	+

Inspired by the success of fully convolutional networks (FCNs) [8] in semantic segmentation, more recent deep SOD methods have shifted toward using FCN-based architectures. These approaches incorporate advanced backbones like VGGNet [9], ResNet [10], and MobileNet [11], allowing end-to-end spatial saliency representation learning. By leveraging the strengths of these convolutional networks, modern SOD models can efficiently predict saliency maps while maintaining spatial coherence, significantly improving both accuracy and computational efficiency compared to earlier methods. Visual transformer-based architectures, such as ViT [13], have recently demonstrated significant potential in segmentation tasks. Several methods have leveraged these architectures to propose transformer-based saliency detection approaches [14]. Furthermore, novel advancements in convolutional networks have emerged, achieving state-of-the-art (SOTA) performance in salient object detection. For instance, GSANet [15] introduced the Semantic Detail Embedding Module (SDEM), which explores the relationships between multi-level features. It adaptively combines shallow texture details with deeper semantic information to efficiently aggregate information entropy in salient regions. Despite these advancements, these architectures have limitations, such as the quadratic computational complexity of visual transformers and the dependency on large-scale pixel-wise human annotations, making them less practical in specific scenarios. To develop effective SOD methods, it is crucial to address the complexities of feature extraction, minimize irrelevant data, and enhance precision in challenging environments such as low-light conditions, complex backgrounds, or high-noise environments. Table 1 summarizes the limitations of existing approaches.

This paper aims to overcome the primary limitations of current SOD methods by improving feature extraction and precision to provide a robust solution for saliency detection across diverse and complex environments. The key contributions of our work include:

1. Novel **Entropy-Based Band Selection Measure** to quantify the information within each spectral band regarding salient objects. It guides the reconstruction of a synthetic RGB image, enhancing the visibility and clarity of salient objects compared to the original image.
2. **Novel Framework for** salient object detection in remote sensing applications, leveraging multispectral decomposition and spectral frequency analysis. Specifically:
    a. We employ a multispectral decomposition technique to distribute the image's information across various spectral bands, effectively filtering out irrelevant details such as noise, background clutter, or non-salient objects and retaining only pertinent information for the segmentation process.
    b. We integrate the selected spectral bands with a segmentation network fused with the original image, improving the network's capacity to identify and delineate salient objects accurately.
3. **The presented Method** has been rigorously evaluated against several SOTA approaches using performance metrics and benchmark datasets. Furthermore, we tested its performance on additional image sets, demonstrating the framework's generalization capability across different domains and environmental conditions.

This comprehensive evaluation provides (i) strong evidence of the developed saliency detection method's effectiveness and robustness and (ii) demonstrates improved precision and adaptability in various contexts. This adaptability ensures that MSD-Net can be effectively applied in diverse and complex environments.

2. Framework for Image Enhancement and Segmentation

The proposed framework begins with a histogram equalization-based enhancement applied to the input image to improve its contrast, followed by gamma correction to adjust the brightness by raising the pixel values to the power of gamma. This pre-processing step enhances visibility and prepares the image for further analysis.

The core algorithm is divided into two main branches. The first branch generates a guidance saliency map, which provides high-level information about potential object locations and shapes while guiding further processing. The second branch decomposes the image into multiple spectral bands, distributing information across different bands to facilitate the selection of relevant data while minimizing noise or irrelevant details that could hinder the detection task. The next phase involves band selection, where the framework measures the similarity between the guidance saliency map and each spectral band. The top three bands from the "R," "G," and "B" spectrums are selected based on their relevance and combined to create a synthetic RGB image. This synthesized image is then summed with the original input image and passed to the segmentation network, which produces the final output mask, indicating the segmented regions of interest.

Fig. 2 illustrates the overall workflow of the framework. The following chapters explain each component's role in image enhancement and segmentation.

2.1 Preprocessing Step: Enhancement of an Image

The goal of image enhancement techniques is to improve the characteristics and quality of an image so that the resulting image looks better than the original when evaluated against specific criteria. Image enhancement is crucial in various image processing applications, including digital photography, medical image analysis, computer vision, remote sensing, object recognition, optical character recognition, fingerprint recognition, industrial automation, face recognition, and scientific visualization. It serves as a vital preprocessing step for numerous image-processing applications and vision systems [16]. Several image enhancement algorithms have been developed recently [16-24], which can be categorized into two main classes: spatial-domain processing and transform-domain processing. **Spatial-domain methods** operate directly on pixel values. Representative methods in this category include gray-level histogram techniques, histogram equalization, adaptive histogram equalization like Contrast Limited Adaptive Histogram Equalization (CLAHE), adaptive gamma correction, human visual system-based methods, unsharp masking, ratio image methods, fuzzy entropy approaches, empirical mode decomposition-based methods, partitioned iterated function systems, linear filters, among others (see details in [25]).

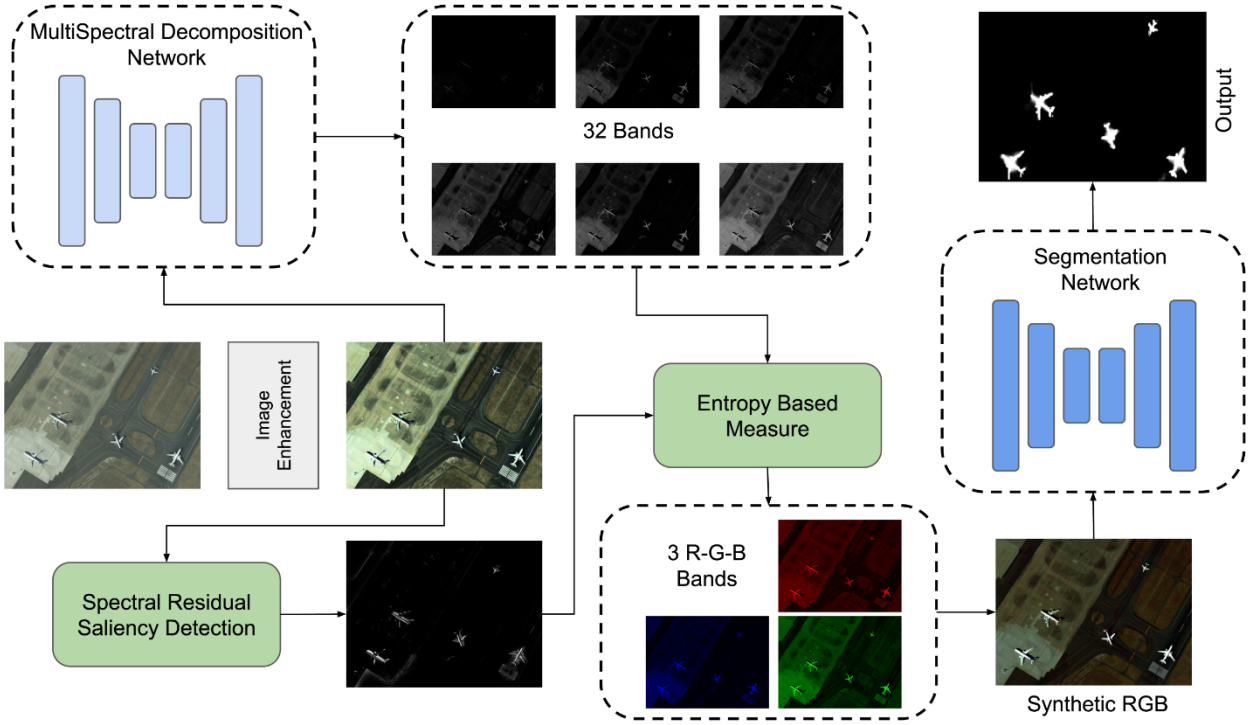

Fig. 2. Overall architecture and workflow of MSD-Net.

This article uses a combined **CLAHE and gamma correction method** as a preprocessing step.

$$I_{enh} = \left(CLAHE(I)\right)^{\gamma} \; where \; \gamma = 1.5.$$

where I is the input image, and I_{enh} is the enhanced output. This approach leverages the strengths of both techniques to enhance image quality effectively:

1. **Contrast Limited Adaptive Histogram Equalization (CLAHE) [25]:** CLAHE improves local contrast by applying histogram equalization to small regions (tiles) of the image rather than the entire image. This method limits contrast amplification to prevent noise enhancement, making fine details more visible without over-saturating the image.
2. **Gamma Correction [26]**: Gamma correction adjusts the brightness of an image by applying a non-linear transformation to the pixel intensity values. It corrects the non-linear way humans perceive light and color, ensuring that the image has appropriate luminance levels—neither too dark nor too bright.

Combined benefit: By integrating CLAHE and gamma correction, we aim to enhance both the local contrast and overall brightness of the image:

- **Step 1**: Apply CLAHE to the input image to enhance local contrast. This step emphasizes edge details and textures, making subtle features more discernible.
- **Step 2**: Perform gamma correction on the CLAHE-processed image. Adjust the gamma value to fine-tune the image brightness according to the application's specific requirements.

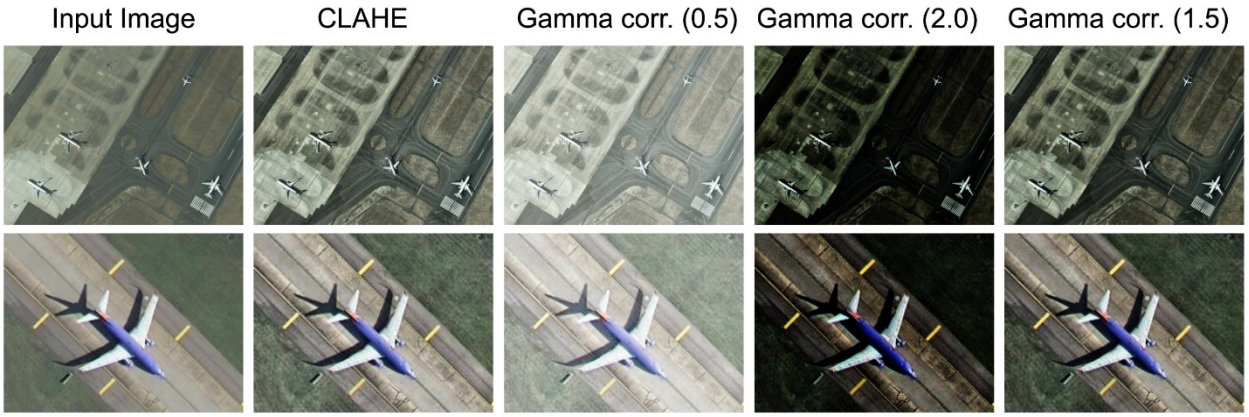

Fig. 3. Input image; CLAHE enhanced; gamma corrected with different gamma parameters.

This combined method enhances fine details while maintaining proper brightness and contrast levels, making an image more suitable for further processing or analysis information during decomposition and more accurate detection of proposal regions by the spectral residual algorithm. Figure 3 demonstrates the effect of CLAHE and gamma correction with different gamma parameters. Experiments showed the best gamma to be selected 1.5, as it does not over-enhance or under-enhance the image. In Fig. 4, spectral residual saliency algorithm is used before (b) and after enhancement (c). We can observe the difference of saliency masks compared to ground truth masks. This difference shows the effect of the enhancement preprocessing part.

2.2 Spectral Residual Saliency Object Detection

The algorithm [5] is designed to detect salient regions in an image by analyzing its spectral properties. The key intuition behind this approach is that salient regions are distinguished from the surrounding background in terms of their spectral characteristics. By working in the spectral domain (using the Fourier Transform), the algorithm can efficiently highlight these regions by identifying and manipulating the spectral residual, which captures the unique, non-redundant

information in the image. The algorithm begins with median blurring with kernel size 5 to reduce noise while preserving edges. Next is the Fourier Transform step: the image $I(x, y)$ is transformed into the frequency domain:

$$F(u, v) = F\big(I(x, y)\big)$$

yielding complex values with amplitude and phase information. A logarithm transformation is applied to the magnitude spectrum, followed by smoothing kernel convolution.

$$A(u, v) = |F(u, v)|$$

$$L(u, v) = log\ \big(A(u, v)\big)$$

$$S(u, v) = h * L(u, v)$$

Fig. 4. Spectral residual saliency (SRS) detection: (a) input image, (b) output of SRS w/o enhancement, (c) output of SRS after enhancement, (d) ground-truth mask.

The spectral residual is computed by subtracting the smoothed spectrum from the original and is exponentiated and combined with the original phase to reconstruct the frequency domain:

$$R(u,v) = L(u,v) - S(u,v),$$

$$M(u,v) = e^{R(u,v)}, \quad F'(u,v) = M(u,v) \cdot e^{j\theta(u,v)},$$

$$Output(u,v) = g(x) \cdot F^{-1}\big(F'(u,v)\big).$$

An inverse Fourier transform followed by a gaussian filter $g(x)$ with $(\sigma = 8)$ generates the **Saliency Map**. There is a final optional step, which subtracts saliency map from original image to get an anomaly map, but we do not use that step in our article. Fig. 4 illustrates some examples of spectral residual algorithms.

2.3 Multispectral Decomposition

Multispectral and hyperspectral images play a crucial role in understanding the physical attributes of objects in an image. While RGB images are limited to three channels (red, green, and blue),

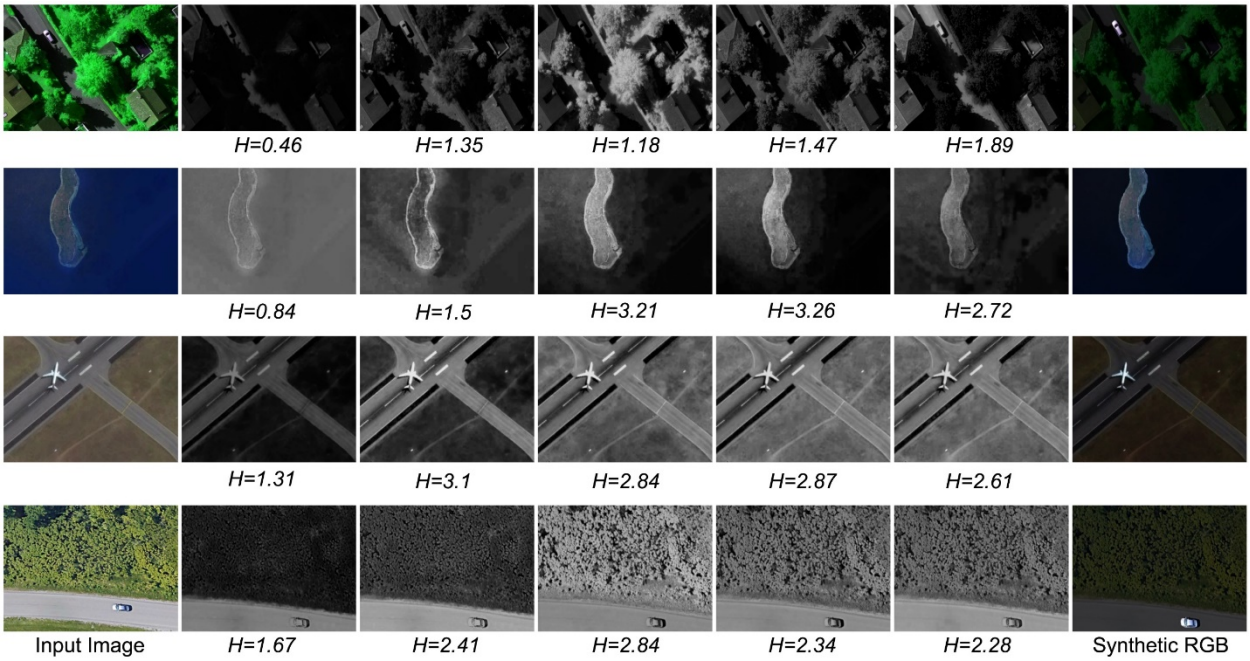

Fig. 5. From left to right: input image, multispectral decomposition bands (indices 1, 7, 15, 22, 30) with entropy measures below, Synthetic reconstructed RGB

multispectral images capture data across a broader range of wavelengths, typically tens to hundreds of spectral bands. This expanded range allows for a more detailed analysis of material properties, surface textures, and object distinctions that might not be visible in standard RGB images.

As obtaining multispectral images can be costly, a significant amount of research aimed at developing methods to predict multispectral information from standard RGB images. Therefore, considerable interest has been in constructing datasets that facilitate RGB-to-multispectral conversions (predictions) through deep learning and other techniques [27]. One such method is the Multi-stage Spectral-wise Transformer (MST++) [28], known for its high accuracy and low computational complexity. The MST++ architecture is based on a convolutional autoencoder

(CAE) design consisting of multiple stages of spectral-wise transformers. Each stage includes an encoder, a bottleneck, and a decoder, with spectral-wise attention blocks (SABs) and deconvolution layers. Skip connections between the encoder and decoder are employed to preserve important spatial information throughout the transformation process. The output of this module consists of 32 spectral bands, with the first 11 bands capturing information from the red channel, the next 11 bands for the green channel, and the remaining 10 bands for the blue channel.

MST++ is employed as the decomposition network in this work, which converts an RGB image into 32 spectral bands, providing a richer, more informative spectral representation of the scene. This arrangement allows for an efficient distribution of information across the channels, ensuring that the decomposition captures subtle variations and essential features in each color channel. Fig. 5 shows an example of anRGB image and some corresponding bands after conversion to a multispectral image. It is easy to see that certain bands contain more information about salient objects than others. The goal is to identify and select the most informative bands to be used as supplementary input for the segmentation module [29]. It can reduce the noise and other information that can bring false positives during the segmentation.

2.4 Novel Entropy-Based Band Selection Measure

Definition 1: To efficiently identify the most informative spectral bands, we compute the entropy-based band selection measure (H) for each band k using the following formula:

$$H^{(k)} = \sum_{i=0}^{N} \sum_{j=0}^{M} w_{ij} H_{ij}^{k},$$

where the entropy H_{ij}^{k} of each block is calculated.

$$H_{ij}^{k} = 20\left(-p_{ij}^{k}\, log\, \left(p_{ij}^{k}\right)\right),$$

- $p_{ij}^{k} \approx 0.5$: indicates a balance between AC and DC components, meaning the block has both structure (variation) and intensity, which suggests high information content.
- $p_{ij}^{k} \approx 0$: indicates that the block is homogeneous with slight variation (dominated by DC), meaning low information content.
- $p_{ij}^{k} \approx 1$: indicates that the block is dominated by high-frequency noise or excessive variation without meaningful structure (dominated by AC), also leading to low information content.

H_{ij}^{k} is the entropy calculated for the ij-th block of the k-th band, and w_{ij} is the average value of the corresponding ij block in the guidance map. H_{ij}^{k} is calculated with the following steps:

1. For each block B_{ij} of the image, a Fourier Transform is performed to obtain the DC and AC components.

$$F_{ij} = FFT\left(B_{ij}\right), \;\; F_{ij}^{shift} = FFTShift(F_{ij})$$

$$DC_{ij} = |F_{ij}^{shift}(0,0)|^{\beta}, \quad AC_{ij} = \sum_{x=1,y=1}^{k} |F_{ij}^{shift}(x,y)|^{\alpha}$$

α and β coefficients are selected experimentally at 0.6 and 2, respectively.

2. A probability value p_{ij}^{k} is computed from the ratio of the AC and DC components from k-th band.

$$p_{ij}^{k} = \frac{AC_{ij}^{k}}{AC_{ij}^{k} + DC_{ij}^{k}}.$$

After calculating $H^{(k)}$ for each k-th band, the top 3 bands are selected with the highest scores from each range (1-10, 11-21, 22-32). The selected bands construct a new synthetic RGB image that captures the salient object information more effectively than the original image. Some example **bands** and their corresponding **entropy scores** are illustrated in Figure 5. The final column shows the RGB image reconstructed from the selected bands. This synthetic image and the original image are merged by taking their average, as some features can be lost in synthetic RGB, which can be crucial for segmentation.

2.5 Segmentation Network Module

For the segmentation module, the merged image is passed through DeepLabV3 [30] network with a ResNet50 backbone. DeepLabV3 is a well-known standard in image segmentation tasks. It is part of a family of segmentation architectures that employ atrous convolution and multi-scale context aggregation to capture fine details in images. These architectures are widely used due to their efficiency and accuracy in pixel-level predictions. In this work, the network was chosen primarily to validate the concept of the proposed framework rather than to focus on optimizing segmentation performance, as it provides a robust and reliable baseline for evaluating the effectiveness of the approach.

2.6 Loss Functions

For training the network, we utilize two loss functions: Binary Cross Entropy (BCE) and Mean Squared Logarithmic Error (MSLE).

$$L_{BCE} = \frac{1}{N}\sum_{i=0}^{N} y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i),$$

$$L_{MSLE} = \frac{1}{N}\sum_{i=0}^{N} (\log(1 + p_i) - \log(1 + y_i))^2 .$$

BCE is commonly used for binary classification tasks, and in our case, it helps classify all pixels as either salient or non-salient regions. MSLE, similar to Mean Squared Error (MSE), introduces a logarithmic transformation to reduce the impact of large outliers, effectively treating them on the same scale as smaller values. This property makes MSLE particularly useful for creating a balanced model that is robust to noise and outliers.

2. Experimental Results

3.1 Dataset

To train and evaluate the proposed framework, we selected the most suitable benchmark dataset for optical remote sensing SOD. The first publicly available SOD dataset was ORSSD, introduced by [31]. It includes 600 training and 200 testing images, each with pixel-wise annotations for salient regions. Despite its importance to the SOD field, this dataset had limitations, particularly due to the small amount of data. To address this issue, [15] introduced an extended version called the EORSSD dataset. This dataset adds 1,200 optical remote sensing images collected from Google Earth to the existing ORSSD dataset, encompassing more complex scenes, objects, and regions. Pixel-wise saliency maps were generated using Photoshop tools, resulting in overall 2,000 images with ground-truth annotations (1,400 for training and 600 for testing). The EORSSD dataset presents several challenges: (i) multiple objects can appear in one single image, (ii) object sizes in optical remote sensing imagery (RSI) vary significantly due to the diverse satellite and airborne imaging platforms, making small object detection particularly difficult, and (iii) the dataset

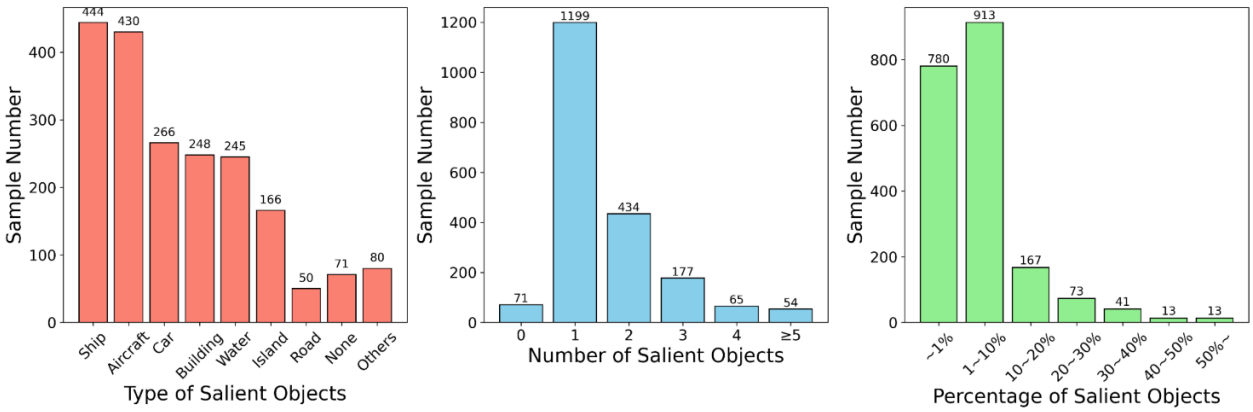

Fig. 6. Statistics about datasets object types, sizes, and counts.

includes a variety of objects such as buildings, streets, ships, aircraft, cars, water bodies, islands, and roads. In summary, the EORSSD dataset is diverse and challenging, making it a valuable resource for training and evaluating SOD models in complex remote sensing scenarios. Some statistics about the dataset are presented in Figure 6.

We evaluate MSD-Net on other remote sensing scenarios as well. To show the generalizability of the proposed method, we also evaluate it on images from the NWPU-RESISC45 [32] dataset, which was initially intended for remote sensing image scene classification. Besides that, we show the performance on **solar panels** images taken from the PV01 dataset [33], **without** providing any training example to the network.

3.2 Evaluation Metrics

To quantitatively evaluate the proposed method, we calculate **four** metrics, with different settings: adaptive, mean, and max *S*-measure (S_α) [34], mean absolute error (MAE), adaptive, mean and max *E*-measure (E_ξ) [35] and adaptive, mean and max *F*-measure (F_β) [36]. The *S*-measure

calculates object similarity considering the structural similarity between the predicted and ground truth masks.

$$MAE = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}|Sal(i,j) - GT(i,j)|, \quad S_\alpha = \alpha \times S_O + (1-\alpha) \times S_r$$

$$E_\xi = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\xi_s(i,j), \quad F_\beta = \frac{(1+\beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall}$$

α = 0.5, S_O and S_r are object and region similarities, respectively. *MAE* calculates the mean absolute distance of predicted and actual saliencies. The *E*-measure is an improved metric designed to calculate the degree of correspondence between global averages and individual local pixels. ξ_s is the enhanced alignment matrix, capturing pixel-level matching and image-level statistics. Finally, *F*-measure calculates the weighted harmonic mean of *Precision* and *Recall*. β is the weight coefficient and is set at 0.3 in our experiments. For each measure, we have three settings: adp (adaptive), mean, and max.

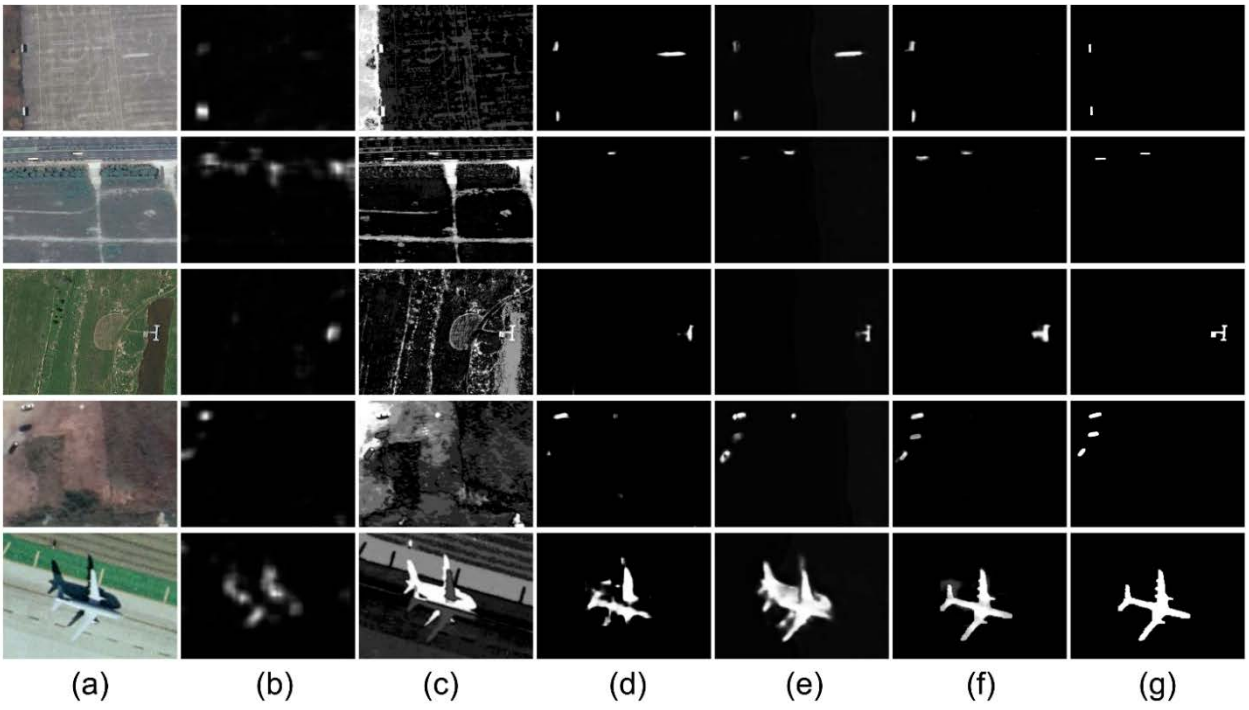

Fig. 7. Comparison of MSD-Net with others. (a) input image, (b) SRS, (c) GCR, (d) DeepLabV3, (e) GCANet, (f) MSD-Net, (g) ground truth.

Table 2. Quantitative comparison of proposed method against others on eorssd dataset.

	S-Measure ↑	MAE ↓	adpEM ↑	meanEM ↑	maxEM ↑	adpFM ↑	meanFM ↑	maxFM ↑
SRS	0.485	0.178	0.647	0.524	0.612,	0.323	0.192,	0.253
GCR	0.568	0.158	0.484	0.577	0.670	0.204	0.330	0.403
DeepLabV3	0.826	0.018	0.826	0.874	0.902	0.602	0.682	0.711
GSANet	0.801	0.025	0.834	0.856	0.871	0.616	0.67	0.689
MSD-Net	**0.841**	**0.017**	**0.854**	**0.881**	**0.912**	**0.637**	**0.703**	**0.731**

3.3. Experiments Setup

For a fair comparison, we train all the deep learning-based methods in our dataset split and evaluate with the same code and pipeline. For all training, the Adam optimizer was used with a learning rate 10^{-4}. Two loss functions have equal coefficients during the training. Random horizontal and vertical flips, rotations, and shifts were used for data augmentation. Batch size and epochs were selected 16 and 200, accordingly. The patch size for the Fourier Transform was set to K = 16, and for the entropy measure, K = 10. These parameter values were chosen based on extensive computer simulations and experimental results.

3.4 Comparison with Other Methods

For the comparison with other methods, we choose 2 non-deep learning-based algorithms, including spectral residual SOD (SRS) [5] and Global Contrast-based SOD (GCR) [6]. While they have successfully found salient objects in some simple cases, they fail if some challenges are present in images, such as complex background scenes or low contrast. To this end, we also compare 2 deep learning based SOTA models: one for general semantic segmentation task (DeepLabV3) [30], and another trained exactly for SOD task (GCANet) [15]. As mentioned, for fair comparison, we use the code they published and train ourselves on our data and our experiment settings. Despite the promising results and improvements compared with non-deep learning methods, they still have some limitations. Visual comparison of the proposed framework with other methods is presented in Fig. 7. On the contrary, MSD-Net has successfully detected the salient objects and has better boundaries, compared to those having non clear object boundaries, false positive detected pixels, as well as missing some parts of objects. While SRS (Fig. 7-b) detected the approximate location of salient objects, it smoothed them and lost a lot of details. On the contrary, GCR (Fig. 7-c) has not lost any details and processed textures well, but it has a lot of false positive cases. Deep learning-based methods have shown better performance. [15] and [30] have false positive cases on the first and fourth images and missed one object in the second image and part of the object in the third image. The fifth image is smooth, but details are lost in both cases. On the other hand, MSD-Net successfully managed to detect better masks of salient objects. Besides qualitative comparison, we also evaluate our method quantitatively using the metrics defined above. Table 2 shows that MSD-Net shows better performance compared with others on all metrics.

We demonstrate the generalizability of the MSD-Net by running it on other images taken from the dataset introduced in [32]. Although this dataset does not provide ground truth masks, as it is not designed for saliency object detection (SOD), we observe visually good masks on various image types. While quantitative evaluation is not possible in this case, we can conduct a qualitative assessment (see Fig. 8-a). We also evaluate the performance of our method on **out-of-distribution**

solar panel images (Fig. 8-b), demonstrating strong generalization capability. In future work, we aim to further improve the accuracy and efficiency of panel detection.

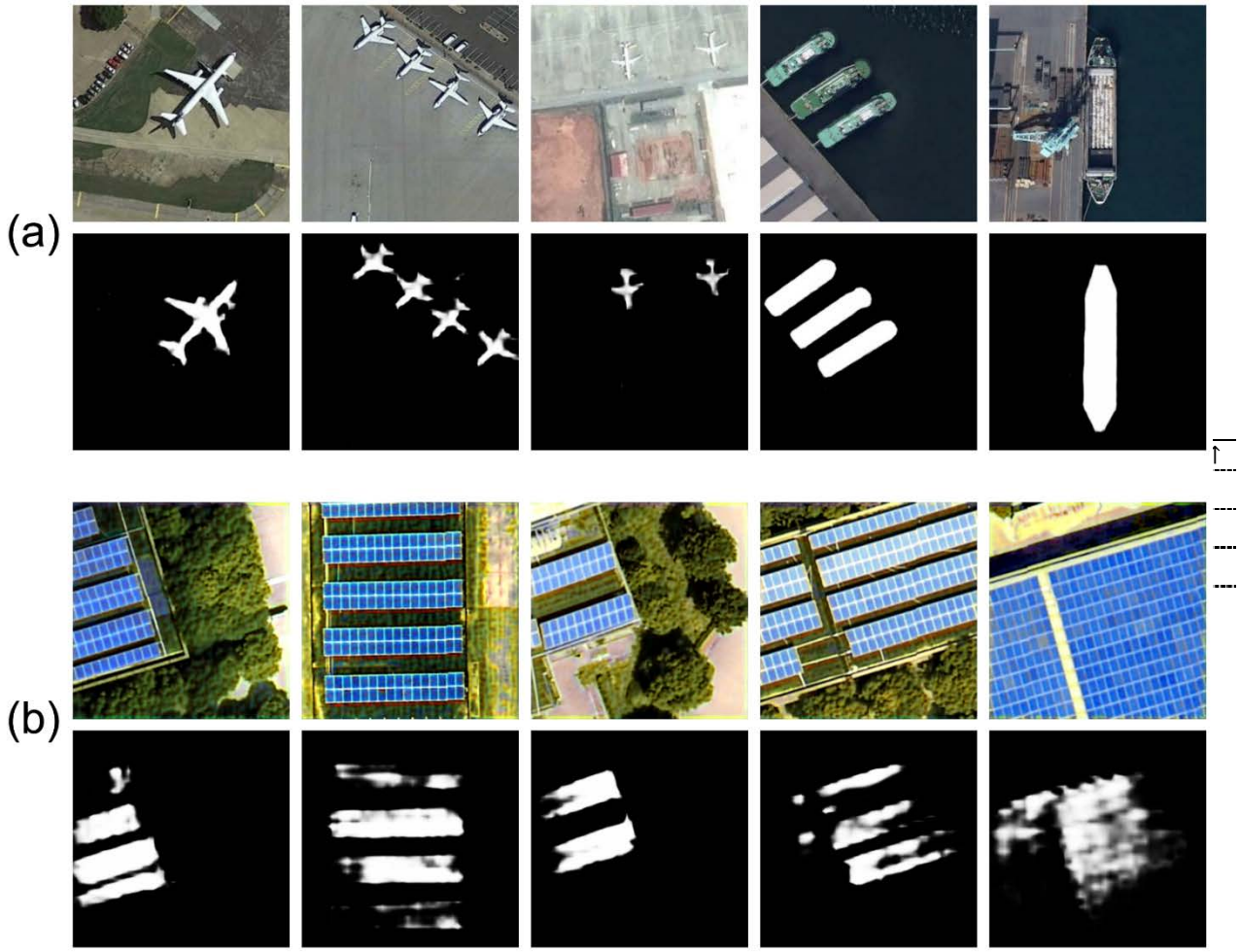

Fig. 8. Predictions of the MSD-Net on samples of **(a)** NWPU-RESISC45 dataset, **(b)** PV01 dataset.

3.5 Ablation Study

To investigate the effectiveness of each component, we first train the segmentation network without applying any pre-processing or synthetic RGB reconstruction steps, establishing a baseline. We then incrementally add components to the pipeline. Second, we integrate a single branch using the spectral residual saliency map, which is generated and fused with the input image to guide the segmentation network in more easily identifying salient objects. This addition improves the metrics a little. Finally, we incorporate the decomposition module, which results in

Table 3. Ablation study analysis

	S-Measure ↑	**MAE ↓**	**adpEM ↑**	**meanEM ↑**	**maxEM ↑**	**adpFM ↑**	**meanFM ↑**	**maxFM ↑**
Segm. only	0.826	0.018	0.826	0.874	0.902	0.602	0.678	0.711
Segm. + guide	0.832	0.018	0.841	0.881	0.912	0.622	0.682	0.720
MSD-Net	**0.841**	**0.017**	**0.854**	**0.886**	**0.915**	**0.637**	**0.703**	**0.731**

the highest performance scores when using the full pipeline. The metric values for each scenario are presented in Table 3, demonstrating the contribution and effectiveness of each block and branch in MSD-Net.

4. Conclusion

In conclusion, this paper presents MSD-Net, a novel framework for salient object detection (SOD) in remote sensing RGB images. MSD-Net enhances feature representation and improves detection accuracy in complex remote sensing scenarios using multispectral decomposition and frequency-based saliency detection techniques. Additionally, we introduce an entropy-based similarity measure for effective band selection and synthetic RGB reconstruction. Experimental results on the EORSSD dataset demonstrate that MSD-Net significantly outperforms state-of-the-art methods on public datasets. Furthermore, we evaluate the framework on various datasets and conduct an ablation study to analyze the contribution of each component.

References

[1] L. Zhang and Li. Zhang, "Artificial intelligence for remote sensing data analysis: A review of challenges and opportunities", *IEEE Geoscience and Remote Sensing Magazine,* vol. 10, no. 2, pp. 270-294, 2002.

[2] X. Wang et al., "Salient object detection: a mini review", *Frontiers in Signal Processing*, vol. 4, 1356793, 2024.

[3] W. Wang et al., "Salient object detection in the deep learning era: An in-depth survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 3239-3259, 2021.

[4] A. Borji et al., "Salient object detection: A benchmark", *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5706-5722, 2015.

[5] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach", *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, DOI: 10.1109/CVPR.2007.383267

[6] Cheng Ming-Ming, et al., "Global contrast based salient region detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no.3, pp. 569-582, 2014.

[7] Li. Guanbin and Yu. Yizhou, "Visual saliency based on multiscale deep features", *Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition*. 2015.

[8] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, DOI: 10.1109/CVPR.2015.7298965

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556 (2014).

[10] K. He et al., "Deep residual learning for image recognition", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, DOI: 10.1109/CVPR.2016.90

[11] A. Howard, "Mobilenets: Efficient convolutional neural networks for mobile vision applications", arXiv preprint arXiv:1704.04861, 2017.

[12] G. Fang et al., "Video saliency detection using object proposals", IEEE Transactions on

Cybernetics, vol. 48, no.11, pp. 3159-3170, 2017.
[13] A. Dosovitskiy, “An image is worth 16x16 words: Transformers for image recognition at scale”, arXiv preprint arXiv:2010.11929, 2020.
[14] N. Liu et al., “Visual saliency transformer”, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.
[15] Q. Zhang et al., “Dense attention fluid network for salient object detection in optical remote sensing images”, *IEEE Transactions on Image Processing*, vol. 30, pp. 1305-1317, 2020.
[16] S. C. Nercessian, K. A. Panetta and S. S. Agaian, “Non-Linear Direct Multi-Scale Image Enhancement Based on the Luminance and Contrast Masking Characteristics of the Human Visual System”, IEEE Transactions on Image Processing, vol. 22, no. 9, pp. 3549-3561, 2013.
[17] J. Xia, K. Panetta and S. Agaian, “Wavelet transform coefficient histogrambased image enhancement algorithms”, Proc. SPIE 7708, 770812, 2010.
[18] A. Grigoryan, J.Jenkinson and S. Agaian, “Quaternion Fourier transform based alpha-rooting method for color image measurement and enhancement”, *Signal Processing*, vol. 109, pp. 269-289, 2015.
[19] E. Wharton, K. Panetta and S. Agaian, “Human visual system based multihistogram equalization for non-uniform illumination and shadow correction”, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1, pp. I-729–I-732, 2007.
[20] S. Nercessian, S. Agaian and K. Panetta, “An image similarity measure using enhanced human visual system characteristics”, *Proceedings of the SPIE, Mobile Multimedia/Image Processing, Security, and Applications*, vol. 8063, p. 806310, 2011.
[21] S. Agaian, “Visual morphology”, *Proc. IS&T/SPIE's Symposium on Electronic Imaging Science & Technology*, vol. 3304, pp. 153–163, 1999.
[22] R. Kogan, S. Agaian and K. Panetta, “Visualization using rational morphology and zonal magnitude reduction”, *IX Proceedings of IS&T/SPIE's Symposium on Electronic Imaging Science & Technology*, San Jose, CA, vol. 3304, pp. 153–163, 1998.
[23] H. D. Cheng, Y.-H. Chen and Y. Sun, “A novel fuzzy entropy approach to image enhancement and thresholding”, *Signal Process*, vol. 75, pp. 277–301, 1999.
[24] S. Agaian, B. Silver and K. Panetta, “Transform coefficient histogrambased image enhancement algorithms using contrast entropy”, IEEE Trans. Image Process., vol. 16, pp. 741–758, 2007.
[25] A. M. Reza, “Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancemen”, *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, vol. 38, pp. 35-44, 2004.
[26] T. Trongtirakul, S. Agaian and S. Wu, “Adaptive Single Low-Light Image Enhancement by Fractional Stretching in Logarithmic Domain”, *IEEE Access*, vol. 11, pp. 143936-143947, 2023.
[27] B. Arad et al., “Ntire 2022 spectral recovery challenge and data set”, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.
[28] Y. Cai et al., “Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction”, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.
[29] H. Gasparyan, T. Davtyan and S. Agaian, “A novel framework for solar panel segmentation from remote sensing images: Utilizing Chebyshev transformer and hyperspectral decomposition”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, 2024, DOI: 10.1109/TGRS.2024.3386402

[30] Li.-Ch. Chen, “Rethinking atrous convolution for semantic image segmentation”, arXiv preprint arXiv:1706.05587, 2017.
[31] Ch. Li et al., “Nested network with two-stream pyramid for salient object detection in optical remote sensing images”, I*EEE Transactions on Geoscience and Remote* Sensing, vol. 57, no. 11, pp. 9156-9166, 2019.
[32] G. Cheng, J. Han and Xi. Lu, “Remote sensing image scene classification: Benchmark and state of the art”, *Proceedings of the IEEE 105.10*, pp. 1865-1883, 2017.
[33] H. Jiang et al., “Multi-resolution dataset for photovoltaic panel segmentation from satellite and aerial imagery”, *Earth System Science Data 13.11*, pp. 5389-5401, 2021.
[34] D.-P. Fan et al., “Structure-measure: A new way to evaluate foreground maps”, *Proceedings of the IEEE International Conference on Computer Vision*, 2017. DOI: 10.1109/ICCV.2017.487
[35] D.-P. Fan et al., “Enhanced-alignment measure for binary foreground map evaluation”, arXiv preprint arXiv:1805.10421, 2018.
[36] R. Achanta et al., “Frequency-tuned salient region detection”, IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009.
[37] B. Silver, S. Agaian and K. Panetta, “Contrast entropy based image enhancement and logarithmic transform coefficient histogram shifting”, *Proceedings of IEEE ICASSP*, pp. 633–636, 2005.

Բազմասպեկտրալ տրոհում և հաճախականության վրա հիմնված շրջանակ հեռահաղորդակցման Պատկերներում ակնհայտ օբյեկտների հայտնաբերման համար

Հայկ Ա. Գասպարյան

Երևանի պետական համալսարան, Երևան, Հայաստան
e-mail: hayk.gasparyan@ysu.am

Ամփոփում

Ակնհայտ օբյեկտների հայտնաբերումը (SOD) նպատակ ունի լուսանկարներում հայտնաբերել ամենաակնառու օբյեկտները, ինչը կարևոր է այնպիսի խնդիրների համար, ինչպիսիք են՝ պատկերների սեգմենտացիան, տեսողական հետևումը, ինքնավար նավիգացիան և լուսանկարների կրճատումը: Թեև SOD-ը լայնորեն ուսումնասիրվել է բնական տեսարանների RGB պատկերներում, հեռահաղորդակցման պատկերներում ակնառու օբյեկտների հայտնաբերումը մնում է չհետազոտված՝ փոփոխական տարածական չափերի և բարդ տեսարանների պատճառով:

Այս հոդվածը ներկայացնում է SOD-ի նոր շրջանակ, որը կոչվում է Multispectral Decomposition Network (MSD-Net)՝ հեռահաղորդակցման 3-շերտ RGB պատկերներում, որը համատեղում է բազմասպեկտրային

տրոհումը և հաճախականության վրա առաջնայնության հայտնաբերումը: Շրջանակը ներառում է երեք հիմնական քայլեր. (i) Բազմասպեկտրալ տրոհում. 3-շերտավոր RGB պատկերի տրոհում 32 բազմասպեկտրային գոտիների՝ սպեկտրային տիրույթներում հատկանիշների գրավումը ուժեղացնելու համար; (ii) Սինթետիկ RGB-ի վերակառուցում. էնտրոպիայի վրա հիմնված նոր չափման կիրառում՝ նշանավոր շրջաններում առավել տեղեկատվական գոտիներ ընտրելու համար՝ վերլուծելով հաճախականության տիրույթը և կառուցելով սինթետիկ RGB պատկեր; և (iii) Saliency Fusion and Object Detection. սեգմենտավորման ցանցի ուսուցում վերակառուցված պատկերի և մուտքային պատկերի միաձուլման վրա՝ բարելավված ճշգրտության համար: Հանրային տվյալների հավաքածուների համապարփակ գնահատումը ցույց է տալիս, որ առաջարկվող մեթոդն ավելի լավ է գործում, քան ժամանակակից (SOTA) մոդելները և առաջարկում է կայուն լուծում բարդ հեռահաղորդակցման պատկերներում ակնհայտ օբյեկտները հայտնաբերելու համար՝ ինտեգրելով բազմասպեկտրային և հաճախականության վրա հիմնված տեխնիկաներ:

Բանալի բառեր՝ ակնհայտության քարտեզ, օբյեկտների հայտնաբերում, բազմասպեկտրային տրոհում, գոտու ընտրություն, հեռահաղորդակցում, էնտրոպիա

Мультиспектральное разложение и частотная основа для выделения заметных объектов на изображениях дистанционного зондирования

Айк А. Гаспарян

Ереванский государственный университет, Ереван, Армения
e-mail: hayk.gasparyan@ysu.am

Аннотация

Обнаружение заметных объектов (SOD) направлено на идентификацию наиболее визуально выделяющихся объектов на изображениях, что важно для задач таких, как сегментация изображений, визуальное отслеживание, автономная навигация и кадрирование фотографий. Хотя SOD активно изучалась в изображениях естественных сцен в RGB, обнаружение заметных объектов на изображениях дистанционного зондирования остается малоизученным из-за изменчивости пространственных разрешений и сложности сцен.

В данной работе представлен новый фреймворк для SOD, называемый Сетью Мультиспектрального Разложения (MSD-Net) в 3-полосных RGB

изображениях дистанционного зондирования, объединяющий мультиспектральное разложение и обнаружение заметности на основе частот. Фреймворк включает три ключевых шага: (i) Мультиспектральное разложение: разложение 3-полосного RGB изображения на 32 мультиспектральные полосы для улучшения захвата признаков через спектральные домены; (ii) Синтетическая RGB реконструкция: использование новой меры на основе энтропии для выбора наиболее информативных полос в заметных регионах путем анализа частотного домена и построения синтетического RGB изображения; и (iii) Слияние заметности и обнаружение объектов: обучение сегментационной сети на слиянии выбранных полос и входного изображения для повышения точности. Обширные оценки на публичных наборах данных показывают, что предложенный метод превосходит существующие модели и предлагает надежное решение для обнаружения заметных объектов на сложных изображениях дистанционного зондирования, интегрируя мультиспектральные и частотно-ориентированные техники.

Ключевые слова: карта очевидности; обнаружение объектов; мультиспектральное разложение; выбор полосы; дистанционное зондирование; энтропия