Co-salient object detection in optical remote sensing images via consensus exploration and detail perception

Yanliang Ge[1], Jiaxue Chen[1], Taichuan Liang[1], Yuxi Zhong[1], Hongbo Bi[1], and Qiao Zhang[1]

[1]Affiliation not available

September 23, 2024

Abstract

Co-salient object detection (CoSOD) in optical remote sensing images (ORSI) is an emerging extension of salient object detection (SOD), which aims to identify common salient objects from a set of related optical remote sensing images. For this mission, we carefully construct the first large-scale dataset, CoORSI. The dataset consists of 7668 elaborately selected high-quality images and target mask annotations, covering macroscopic geographic scenes such as rivers and beaches, as well as man-made targets such as airplanes and ships, in a total of 10 categories. In addition, we propose a consensus exploration and detail perception network (CEDPNet) for cosalient object detection in remote sensing images. Specifically, a collaborative object search module (COSM) is introduced to effectively integrate high-level features and obtain inter-pixel and inter-region correlation, so as to further explore and locate collaborative objects. On the basis of this module, we design a feature sensing module (FSM), which integrates difference contrast enhancement unit (DCEU) and multi-scale detail boosting unit (MBDU) to enhance the perception of salient targets. Finally, the high-level semantic information is continuously fused with the low-level detailed features to obtain the final co-salient detection maps. Extensive experimental evaluations confirm that CEDPNet has significantly superior performance compared to other competitors in Co-salient object detection in optical remote sensing images. The CoORSI dataset, model and results will be available at: https://github.com/chen000701/CEDPNet.

Co-salient object detection in optical remote sensing images via consensus exploration and detail perception

Yanliang Ge, Jiaxue Chen, Taichuan Liang, Yuxi Zhong, Hongbo Bi and Qiao Zhang

***Abstract*—Co-salient object detection (CoSOD) in optical remote sensing images (ORSI) is an emerging extension of salient object detection (SOD), which aims to identify common salient objects from a set of related optical remote sensing images. For this mission, we carefully construct the first large-scale dataset, CoORSI. The dataset consists of 7668 elaborately selected high-quality images and target mask annotations, covering macroscopic geographic scenes such as rivers and beaches, as well as man-made targets such as airplanes and ships, in a total of 10 categories. In addition, we propose a consensus exploration and detail perception network (CEDPNet) for co-salient object detection in remote sensing images. Specifically, a collaborative object search module (COSM) is introduced to effectively integrate high-level features and obtain inter-pixel and inter-region correlation, so as to further explore and locate collaborative objects. On the basis of this module, we design a feature sensing module (FSM), which integrates difference contrast enhancement unit (DCEU) and multi-scale detail boosting unit (MBDU) to enhance the perception of salient targets. Finally, the high-level semantic information is continuously fused with the low-level detailed features to obtain the final co-salient detection maps. Extensive experimental evaluations confirm that CEDPNet has significantly superior performance compared to other competitors in Co-salient object detection in optical remote sensing images. The CoORSI dataset, model and results will be available at: https://github.com/chen000701/CEDPNet.**

***Index Terms*—Co-salient object detection, optical remote sensing images, CoORSI dataset.**

I. INTRODUCTION

SALIENT object detection is a task designed to identify the most visually attractive areas in an image/video and generate pixel-level prediction maps by simulating the human visual perception system [1]–[3]. With the rapid development of remote sensing technology, the task of CoSOD in ORSI has attracted more and more attention. Unlike natural scene images taken by human photographers with hand-held cameras, optical remote sensing images are automatically collected by various remote sensors deployed on satellites or aircraft with minimal human intervention [4], [5]. Affected by various imaging conditions, remote sensing images obtained by aerial photography usually cover a large range of scenes and diverse surface types [6]. Therefore, compared with the natural scene, the remote sensing salient object detection task is often more challenging. Fig. 1 shows several typical challenging scenarios for Co-salient object detection in optical remote sensing tasks.

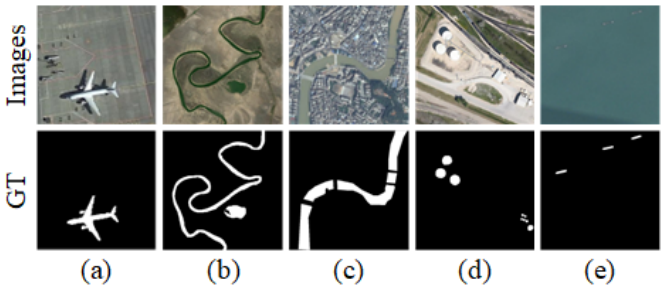

Fig. 1: Challenging optical remote sensing image scenes. (a) Shadow interference. (b) Complex topology. (c) Clutter background. (d) Multi-scale objects. (e) Tiny objects in low contrast scenes.

Recently, CoSOD has been very important in the computer vision and image processing research direction. Unlike traditional salience object detection, CoSOD utilizes the correlation between multiple images to identify common salient areas or objects [7]. Remote sensing images usually contain a lot of noise and background interference, so the information provided by a single image is often limited and incomplete. The reliability of object detection can be enhanced by using the common attributes in image groups and the complementarity between different scene images [8]. However, the task of co-salient object detection needs to avoid the interference caused by different categories of salience targets and non-salience targets of the same category, and fully explore the semantic commonality of co-salient objects, which further increases the difficulty of mining co-salience clues from optical remote sensing images [9], [10]. The Co-salient object detection in optical remote sensing task can meet the needs of multi-image analysis, improve the ability of target recognition, and can be widely used in surface monitoring [11], [12], urban planning [13] and complex scene understanding [14].

Due to the lack of specific dataset for the optical Co-salient object detection in optical remote sensing task, this paper meticulously constructs a large-scale Co-salient object detection in optical remote sensing dataset CoORSI on the basis of the existing datasets in the field of remote sensing

Manuscript received ***, ***; revised ***, ***.

Yanliang Ge, Jiaxue Chen, Taichuan Liang, Yuxi Zhong, and Hongbo Bi are with the School of Electrical Engineering and Information, Northeast Petroleum University, Daqing 163318, China (e-mail: 15804593399@139.com; chen_0024@163.com; jgsultc188@163.com; zyx197823@163.com; bhbdq@126.com).

Qiao Zhang is with the School of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China (e-mail: 15263653689@163.com).

Co-corresponding authors: Hongbo Bi and Qiao Zhang.

Fig. 2: Illustration of Co-salient object detection in optical remote sensing task. The task refers to the detection of co-occurring salient objects in multiple remote sensing images.

salient object detection, shown in Fig.2. However, there are still many problems such as background interference, diverse object scales and complex topological structure, which greatly hinder the detection performance of co-salient objects in remote sensing images. Therefore, how to explore and aggregate effective subtle clues from different images for accurate co-salient object detection is worth more efforts.

To solve above problems, this paper proposes a consensus exploration and detail perception network (CEDPNet), which is composed of two key parts: collaborative object search module (COSM) and feature sensing module (FSM). Specifically, in order to avoid the interference of diverse object scales and complex scenes, and successfully explore the clear edge and detail information, we designed a feature sensing module. FSM can obtain the representation information of salient objects from the two perspectives of difference contrast enhancement and multi-scale detail boosting, so as to accurately detect the complete morphology and structure of salient objects. In addition, in order to effectively focus on common regions or objects and avoid the impact of different categories of salient objects, we propose a collaborative object search module, which makes full use of different levels of features, deepens the understanding of image semantic information, considers the pixel correlation between image groups and enhances the attention to regional correlation. At the same time, the correlation screening mechanism is used to deepen the consistency learning within the image group, and realize the accurate localization and recognition of co-salience objects. Finally, the decoding strategy guided by multi-layer semantics is used to implement an effective aggregation of coarse-grained semantic information and fine-grained detail features, and make full use of multi-layer features fusion to explore and segment remote sensing co-salient object. The core contributions of this paper are summarized as follows:

1) **CoORSI dataset:** We propose the first large-scale and challenging dataset for Co-salient object detection in optical remote sensing, named CoORSI. The dataset is mainly composed of 10 common remote sensing scenarios, and contains 7668 remote sensing images and corresponding binary mask annotations.
2) **CEDPNet:** Based on CoORSI dataset, we propose a consensus exploration and detail perception network, which adopts the way of bidirectional perception of salient features and parallel mining of collaborative feature for refined detection and recognition of remote sensing co-salient objects.
3) **Empirical contribution:** We demonstrated the effectiveness of the proposed approach by rigorously evaluating 19 state-of-the-art models under six widely used evaluation metrics, including 12 ORSI-SOD methods and 7 CoSOD methods. The comprehensive benchmark test results show that the proposed method is significantly superior to other competitors, making it an effective solution for Co-salient object detection in optical remote sensing tasks.

II. RELATED WORK

A. Remote Sensing Datasets

A high-quality dataset can often promote the effective training of the model and improve the accuracy of the prediction. Tab. I lists four available image datasets specifically for salient object detection in optical remote sensing images.

- **ORSSD [15]:** The first pixel-level labeled dataset for salient object detection in optical remote sensing images, including 600 training images and 200 test images, laid a foundation for subsequent research.
- **EORSSD [5]:** An extended version of the original ORSSD dataset, including 1400 training images and 600 test images for a total of 2000 images and corresponding pixel-level ground truths.
- **ORSI-4199 [16]:** A comprehensive benchmark dataset of 4199 images with pixel-level annotations. While expanding the scale of the dataset, the challenging attributes of the detection objects are summarized in detail.
- **RSISOD [17]:** A massive dataset of nearly 5,000 images covering a wide range of realistic scenes, different object properties, and different types of salient objects.

TABLE I
STATISTICS OF EXISTING ORSI SOD DATASETS. WE LIST SOME DATASET INFORMATION SUCH AS YEAR, PUBLICATION (PUB.), THE NUMBER OF OBJECT TYPES (CLASS), SALIENT OBJECTS ATTRIBUTES (ATT.), DATASET NUMBER (TRAIN AND TEST).

Dataset	Year	Pub.	Class	Att.	Train	Test
ORSSD [15]	2019	TGRS	#	#	600	200
EORSSD [5]	2021	TIP	#	#	1400	600
ORSI-4199 [16]	2022	TGRS	#	9	2000	2199
RSISOD [17]	2023	TGRS	18	#	3784	1270

The above datasets are proposed for the task of salient object detection, which are usually processed on a single image, the classification of the image is usually unclear or the image contains multiple categories. Therefore they aren't suitable for co-salient object detection task. Based on the classification, statistics and screening of existing remote sensing salient object detection datasets, we construct a large-scale dataset

specifically for Co-salient object detection in optical remote sensing task, namely CoORSI. The dataset contains 7668 images covering ten categories such as aircraft, ships and rivers, which can be applied to the network training and performance evaluation of Co-salient object detection in optical remote sensing.

B. Remote Sensing Salient Object Detection

In recent years, inspired by the effective strategy of salient object detection in natural scene images (NSI), salient object detection in optical remote sensing images (RSI) has made remarkable progress. The inchoate ORSI-SOD method adopts bottom-up measure for salient object detection, which heavily relies on human prior knowledge and handmade features [18]–[21]. However it is difficult to design effective handmade features in complex scenes, and the applicable occasions are very limited.

At present, the model based on deep learning has become the main framework of ORSI-SOD, and the encoder-decoder structure based on convolutional neural network has been widely used. Li et al. [15] propose an end-to-end deep network, LV-Net, which supposes cluttered backgrounds by utilizing L-shaped and V-shaped modules to sense different scales and local details of salient objects. Luo et al. [22] adopt two interactive decoding branch architectures based on U-shaped network to further explore the correlation between semantically enhanced salient features and edge features through multi-scale attention interaction. Zheng et al. [17] introduce the boundary sensing partial decoder and the structure sensing partial decoder, and supervise the learning of the network model through the edge sensing loss and structural similarity loss. Zhao et al. [23] combine multi-scale sparse transform and pyramid-hole attention to construct an adaptive double-flow sparse encoder to enhance the global perception of local features and local details of global representation.

In addition, some work introduces additional feature maps (i.e., edge or skeleton maps) for supervision to enhance the recognition ability of co-salient object profiles and achieve more accurate object segmentation. For example, Tu et al. [16] extract multi-scale regional features of salient object through hierarchical attention modules, and embed boundary features into regional features at multiple scales to optimize both boundary and regional features. Qi et al. [24] integrate adjacent features with a nonparametric alignment strategy and use interactive lead loss to combine significance and edge detection to facilitate the detection of salient object with fuzzy edges and irregular topologies.Zhou et al. [25] propose an edge-sensing location attention unit to extract and fuse multi-layer depth features to effectively sharpen edges and locate salient objects. Gong et al. [26] use a two-level network architecture guided by edges and skeleton from coarse to fine, and enhances edge and skeleton features through spatial graph attention and spatial self-optimization. Liang et al. [27] design a multi-scale edge-embedded attention module to enhance the capture of salience objects by incorporating edge information into spatial attention maps.

The above methods provide valuable suggestions for co-salient object detection in optical remote sensing images. However, without the supervision of additional maps, most solutions are disturbed by complex background and shadow and the target to be detected is of various scales, so it is often difficult to detect objects with clear boundaries and complete structure. Therefore, we adopt the method of enhancing the differentiation between the target and the background and boosting multi-scale detail to effectively overcome the interference caused by messy background, thereby accurately identify co-salient targets.

C. Co-salient Object Detection

Early co-salient object detection base on handcrafted heuristic features and score the correlation of each pixel or region to assess the consistency between groups of images [28]–[32]. In addition, some methods use the existing salient object detection algorithm to obtain the subgraph of the initial prediction of the image, and deduce the fusion weights between the subgraphs according to some prior constraints to explore the common significance information between images [33]–[36]. For example, Zhang et al. [37] used feedback gradient information to draw more attention to discriminant co-salient features. Jin et al. [38] extracted intra-significance clues from the significance graphs of single images predicted by the off-the-shelf SOD method, and obtained inter-significance clues through correlation fusion. However, the integration of multiple clues and prior information often deviates from the expected result and is difficult to apply in real situations.

Now the end-to-end co-salient object detection method based on deep learning has gradually replaced the traditional method and continuously improved the performance of the model. Gao et al. [39] embed the collaborative attention module into the upper convolutional layer of the full convolutional network to give greater attention weight to the common salient targets. Zhang et al. [40] aggregate cross-image consistency cues by self-attention mechanism and generated dynamic kernel from consensus features to summarize fine-grained image-specific consensus object cues and coarse-grained group common knowledge. Yu et al. [41] design the democracy prototype generation module to explore the significant features of comprehensive cooperation with democracy and reduced background interference. Zhu et al. [42] introduce a circular proxy purification strategy to search for noise-free common representations by iteratively optimizing the predicted significance graphs. Zheng et al. [43] introduce memory-assisted contrast consistency learning to save and update the consistency of images from different groups in a memory queue, improving the quality and integrity of prediction maps.

Inspired by the above method, we employ a mixed method of self-attention and cross-attention to capture local features within image and effectively integrate global information, deeply exploring and mining consistent semantic information. The collaborative object search module, by leveraging correlation statistics of pixels and regions, obtains coherence clues among images, thereby maximizing the suppression of interference from unrelated targets and enabling accurate identification and positioning of co-salient objects.

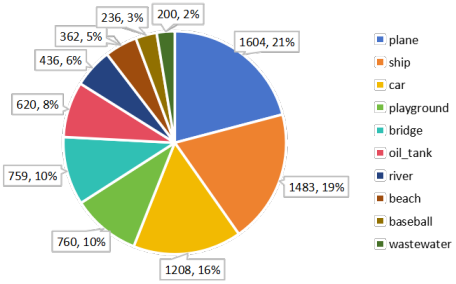

Fig. 3: Classification structure of CoORSI dataset.

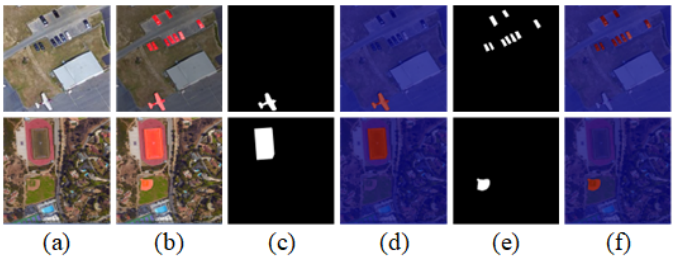

Fig. 4: Image processing. (a) Images. (b) Multiple different categories. (c) and (e) Single class objects pixel-level labeling results. (d) and (f) New annotation results visualization.

III. METHODOLOGY

In this part, the proposed dataset and network model are described in detail. In III-A, we describe the proposed dataset CoORSI. In III-B, we give the overall architecture of CEDPNet. Then, we elaborate on the core components of CEDPNet in section III-C-III-D, including feature sensing module and collaborative object search module.

A. CoORSI Dataset

Remote sensing co-salient object detection aims to detect common remote sensing objects from a set of images. The quality and diversity of datasets play a crucial role in the training and evaluation of algorithms. However, the lack of a comprehensive and high-quality dataset for co-salient object detection in optical remote sensing images limits the improvement of algorithm performance and the development of new methods. To this end, we constructed the first large-scale ORSI-CoSOD dataset by screening and processing four existing remote sensing salient object detection datasets (i.e., ORSSD [15], EORSSD [5], ORSI-4199 [16] and RSISOD [17]).

Screening criteria: The CoORSI is constructed on the premise of ensuring that the images are not damaged and the images with no salient object or unknown category are eliminated. In addition, the images in the dataset should cover different types of geographical features, environmental conditions and target objects, while ensuring that the number of various types of objects in the dataset is relatively balanced to improve the generalization ability of the model. To do this, we filtered out groups of less than 50 images and ultimately selected scenes/categories commonly found in ten remote sensing images, such as aircraft, ships, and rivers. Fig. 3 illustrates the classification structure of our proposed dataset.

Image processing: Since some existing remote sensing salient object detection datasets don't provide corresponding category labeling, we need to conduct detailed category labeling for selected images. Furthermore, we find that some images contain two or more salient objects of different categories, and the corresponding pixel-level labeling is obviously not suitable for co-salient object detection. Therefore, while checking the accuracy of image labeling, we re-calibrate the images containing multiple categories at pixel level to ensure the reliability of model training results. The related process of image processing is shown in Fig. 4. In order to ensure the accuracy and consistency of annotation, multiple team members processed 7668 images at the same time and adopted interactive inspection methods to further avoid annotation disputes.

B. Architecture Overview

The CEDPNet follows the encoder-decoder structure and its main framework is shown in Fig. 5. Given a set of related images $\mathcal{I} = \{I_n\}_{n=1}^{N}$ containing a common salient object of a certain category as input, and adopt the pre-trained EfficientNet [44] network as backbone to extract image features. Meanwhile, different strategies are used to explore the generated high-level and low-level features. In order to effectively capture and locate co-salient objects in image groups and avoid the influence of non-salient objects and backgrounds, we send the reconstructed high-level features and original features into the collaborative object search module. For low-level features, we designed a feature sensing module, which effectively alleviates the difficulty of complex background interference and small target detection through the methods of difference contrast enhancement and multi-scale detail boosting. COSM enhances feature representation of semantic consistency by exploring the similarity of pixels and regions in image groups. Finally, a simple decoder is used to aggregate the information between layers in a way of multi-layer semantic progressive guidance, which realizes the process of co-salient object from rough localization to fine detection.

C. Collaborative Object Search Module

Remote objects of the same category often have certain similarities in the appearance of objects (i.e., the size, shape, texture and scale of objects). Inspired by the literature [45], we propose a collaborative object search module to achieve accurate recognition and positioning of collaborative objects, as shown in Fig. 6. Collaborative object search module (COSM) can effectively capture the common features of collaborative objects by calculating the pixel correlation and regional correlation among the same category image groups.

COSM is mainly composed of hierarchical feature fusion block (HFFB) and collaborative feature mining block (CFMB). Hierarchical feature fusion block mainly integrates high-level features to achieve effective interaction between different levels of features, so as to enhance the ability of the model to ex-

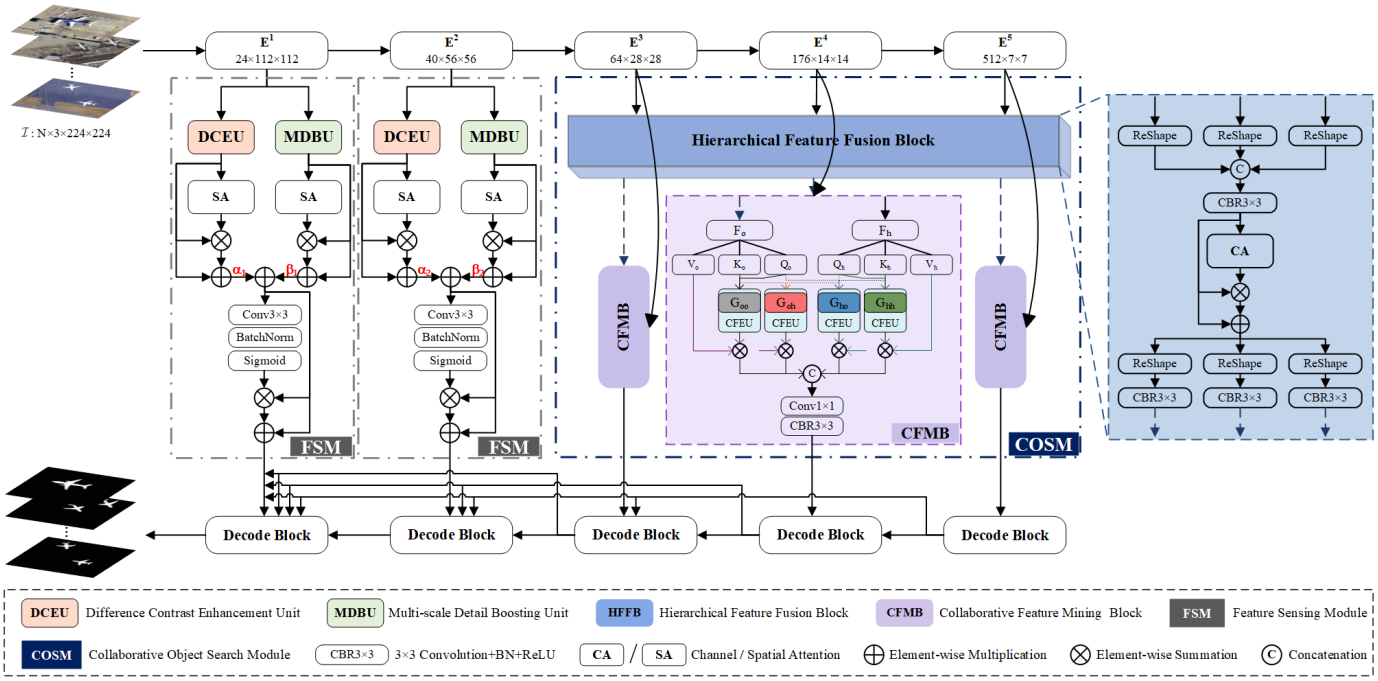

Fig. 5: The overall architecture of the CEDPNet. CEDPNet contains two main modules: feature sensing module (FSM) and collaborative object search module (COSM). COSM explores the consistency clues by counting image group pixels and regions correlation. FSM realizes the fine detection of objects by difference contrast enhancement and multi-scale detail boosting. Furthermore, fine-grained representation information is continuously aggregated under the guidance of high-level semantic information to achieve accurate location and segmentation of the remote sensing image co-salient objects.

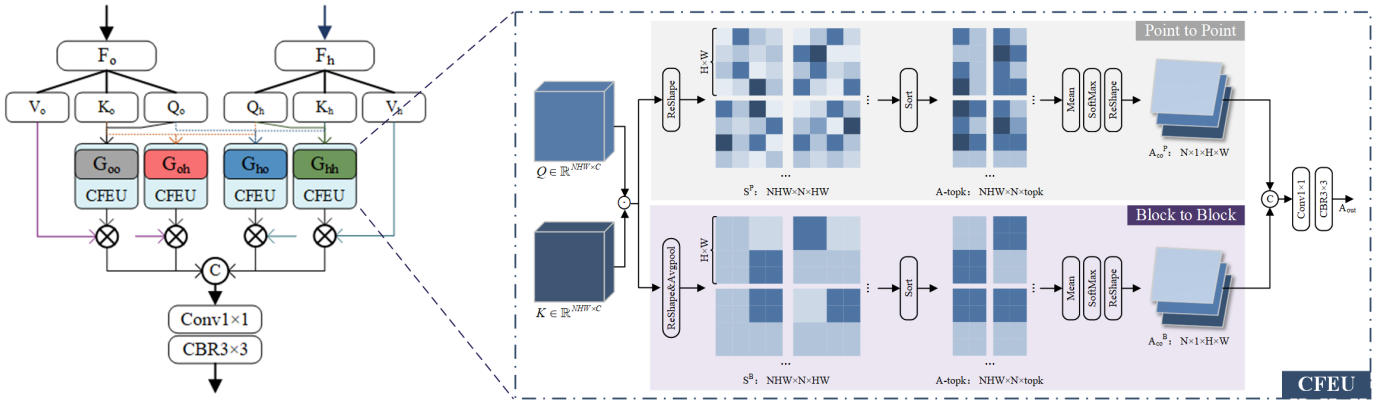

Fig. 6: The architecture of the collaborative object search module (COSM). COSM adopts hierarchical feature fusion to enhance semantic information understanding, and simulates the attention mechanism to explore the pixel-level and region-level correlations of image groups in parallel, mining the intra-group consensus to effectively locating and recognizing co-salient objects.

tract image semantic information. HFFB takes reshaping operation for high-level features $F_i \in \mathbb{R}^{N\times C_i\times H_i\times W_i}$ $(i=3,4,5)$ to make their feature size consistent. Then the channel dimension concatenation and convolution operation are used to generate feature F_c, which realizes the effective fusion of multi-layer features and improve the expression ability of features. In addition, assigning weights to each channel of the feature allows the network to focus on features that are more critical to the task. The above calculation procedure is as follows:

$$\begin{cases} F_c = CBR_{3\times 3}\left(Cat\left(\mathrm{Re}\left(F_3\right),\mathrm{Re}\left(F_4\right),\mathrm{Re}\left(F_5\right)\right)\right) \\ F_h = F_c \oplus F_c \otimes \mathcal{F}_{CA}\left(F_c\right) \end{cases} \tag{1}$$

where $\mathcal{F}_{CA}(\cdot)$ denotes the channel attention, $Re(\cdot)$ denotes the reshape operation, $Cat(\cdot)$ denotes the channel cascade operation, $\oplus$ denotes element-wise summation, $\otimes$ denotes element-wise multiplication, and $CBR(\cdot)$ denotes the combination of convolutional layer, Batch Normalization layer, and ReLU layer.

In order to better integrate features from different levels, the channel attention-weighted fusion features F_h are reshaped again and sent to CFMB together with the original side output features F_o in a specified way. The original encoder output features F_o retain the basic semantic information of co-salient object, while the features F_h processed by hierarchical fusion can effectively capture global and contextual information. For the unique and complementary feature attributes in feature maps from different sources, collaborative feature mining block simulates self-attention [46] and cross-attention [47], enhance their feature description ability, achieve the interactive fusion of different features, and further explore semantic consensus.

We take one of the encoder side output features $F_o \in \mathbb{R}^{N\times C\times H\times W}$ as an example to introduce collaborative feature mining block in detail. F_o performs linear transformation to generate features $Q_o \in \mathbb{R}^{NHW\times C}$, $K_o \in \mathbb{R}^{NHW\times C}$ and $V_o \in \mathbb{R}^{N\times C\times H\times W}$. Similarly, the feature $F_h \in \mathbb{R}^{N\times C\times H\times W}$ generated by hierarchical fusion reshaping is processed accordingly to obtain Q_h, K_h and V_h. The collaborative feature mining block simulates the self-attention mechanism, and generates the initial correlation matrix $S_{oo} \in \mathbb{R}^{NHW\times NHW}$ by matrix multiplication of K_o and Q_o. Each row of S_{oo} represents the similarity between one pixel and all pixels, and effectively captures the consensus correlation inside the feature. By simulating the cross-attention mechanism, the cross-correlation matrix $S_{oh} \in \mathbb{R}^{NHW\times NHW}$ is generated through matrix multiplication of K_o and Q_h, and represents the consensus correlation between feature graphs guided by rich semantic information. By simulating the use of diverse attention, we can enhance global context awareness, deepen the understanding of contextual information, and further improve the discriminability of consensus features. The above process can be expressed as:

$$\begin{cases} \mathrm{S}_{oo} = Q_o \odot K_o^T, S_{oh} = Q_o \odot K_h^T \\ S_{hh} = Q_h \odot K_h^T, S_{ho} = Q_h \odot K_o^T \end{cases} \tag{2}$$

Where $\odot$ denotes matrix multiplication.

In addition, the initial affinity matrix and the interactive affinity matrix are respectively fed into the collaborative feature exploration unit, which is the key to the collaborative feature mining block. The cooperative feature exploration unit explores the correlation between pixel points and region blocks in the image group respectively, and uses the pixel-level and region-level feature to coordinate location of the object to be detected.

On the one hand, CFEU uses the correlation between pixel points to model, reshaping the correlation matrix S to $S^P \in \mathbb{R}^{NHW\times N\times HW}$. Only a unique maximum similarity value is selected for each image, which lacks democratic representation and is susceptible to noise. In order to make more pixels participate in the selection stage and improve the accuracy of recognition, we take the mean of the first k with the highest similarity value as the common representation of cooperative significance, and obtain the N maximum similarity value of each pixel. Then, the average value of N maximum similarity values is taken as the co-salient probability of each pixel, and the pixel-level global affinity attention map A_{co}^P is generated by softmax and reshape operations. The calculation process of A_{co}^P is as follows:

$$A_{co}^P = \mathrm{Re}\left(\mathcal{S}\left(mean\left(topk\left(S\right)\right)\right)\right) \tag{3}$$

where $\mathcal{S}(\cdot)$ denotes the softmax operation, and $\mathrm{topk}(\cdot)$ denotes selecting the highest topk values.

On the other hand, CFEU uses the correlation between regional blocks to model, and explores the cooperative goal effectively by exploring the regional level correlation. Compared with the pixel level, regional level exploration can effectively improve the computational efficiency and ensure the structural integrity of the detected object. Specifically, the correlation matrix S is divided into several local regions, and pooling operations are taken to extract the most representative features to represent each local region. After sorting the region-level similarity, the region-level global affinity attention map A_{co}^R is obtained by the same operation as the pixel branch. The calculation process is as follows:

$$A_{co}^R = \mathrm{Re}\left(\mathcal{S}\left(mean\left(topk\left(\mathrm{avg}\left(S\right)\right)\right)\right)\right) \tag{4}$$

The global affinity concerns (A_{co}^P and A_{co}^R) generated by different methods are fused with a cascade of channels. After that, 1×1 convolution is used to achieve channel dimension reduction, and 3×3 convolution is combined to enhance the nonlinear expression ability of attention map, and then the corresponding attention map A_{out} is generated.The above calculation process is as follows:

$$A_{\text{out}} = CBR_{3\times 3}\left(Conv_{1\times 1}\left(Cat\left(A_{co}^P, A_{co}^B\right)\right)\right) \tag{5}$$

Finally, the generated attention maps were used to weight features V_o and V_h respectively to enhance the final consensus feature representation. Integrating them together, we can obtain a reliable global consistency map F_{co}. The calculation process of F_{co} is as follows:

$$\begin{cases} F_{oo} = A_{out}^{oo} \otimes V_o \\ F_{oh} = A_{out}^{oh} \otimes V_o \\ F_{hh} = A_{out}^{hh} \otimes V_h \\ F_{ho} = A_{out}^{ho} \otimes V_h \end{cases} \tag{6}$$

$$F_{co} = CBR_{3\times 3}\left(Conv_{1\times 1}\left(Cat\left(F_{oo}, F_{oh}, F_{hh}, F_{ho}\right)\right)\right) \tag{7}$$

D. Feature Sensing Module

Due to the complex background and shadow interference, it is difficult to detect the complete topological structure and smooth edge of the salient objects in remote sensing images, especially small targets. Low-level features contain rich texture and detail information, which is conducive to delineating fine-grained salient targets. To this end, we design a feature sensing module (FSM) to fully utilize the fine-grained information of remotely sensed objects. FSM consists of two branches with similar architecture based on difference contrast enhancement unit (DCEU) and multi-scale detail boosting unit (MDBU), as shown in Fig. 7 and Fig. 8. The two-branch structure is used to optimize the low-level features $f_i \in \mathbb{R}^{N\times C_i\times H_i\times W_i}$ $(i = 1, 2)$ and explicitly focus on local differences and object details.

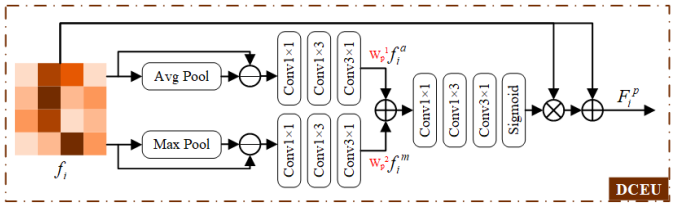

Fig. 7: The architecture of the difference contrast enhancement unit (DCEU). DCEU employs the pool-subtraction operation to enhance the differential perception.

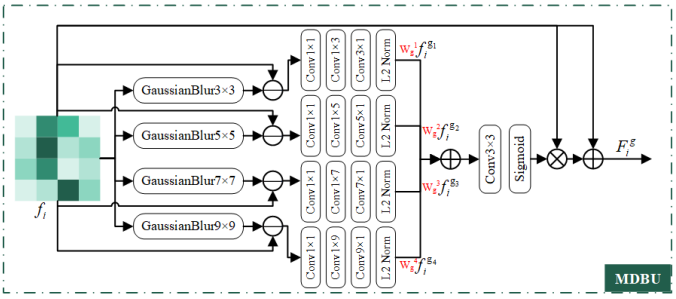

Fig. 8: The architecture of the multi-scale detail boosting unit (MDBU). MDBU implements multi-scale analysis and detail boosting by multiple Gaussian filters.

Difference Contrast Enhancement Unit: The original feature contains rich local features, such as edges and textures, while the pooling feature is smoother and may obscure some important local features. The difference value between the original feature and pooled feature can form a contrast learning mechanism to highlight local features and enhance the sensitivity of the model to edges and textures. The difference contrast enhancement unit uses the pooling and subtraction method to perceive the distinction of the low-level features output by the encoder, and then adopts the asymmetric convolution operation to identify and enhance the features in the specific direction of the image, so as to improve the accuracy of texture recognition. Mathematically, the process of calculating differential features is expressed as:

$$f_i^a = Conv_3\left(abs\left(f_i - avg\left(f_i\right)\right)\right) \tag{8}$$

$$f_i^m = Conv_3\left(abs\left(f_i - \max\left(f_i\right)\right)\right) \tag{9}$$

where $Conv\left(\cdot\right)$ denotes convolution operation, $avg\left(\cdot\right)$ denotes average pooling operation, $\max\left(\cdot\right)$ denotes max pooling operation, and $abs\left(\cdot\right)$ denotes absolute value calculation.

The contribution of differential features generated by inconsistent pooling operations to object refinement detection is not exactly the same. We balance the importance of contrasting differential features by setting learnable weight W_p^1 and W_p^2. In addition, the difference features with enhancement factors are added at the element level, and 1×1 convolution and asymmetric convolution operations are carried out to further explore the spatial clues of salient objects. After the sigmoid activation operation is performed on the output feature ${f_i}^p$, the residual multiplication operation is used to emphasize the difference-sensing feature representation and superimpose the output feature f_i on the encoder side to obtain difference-enhanced features F_i^p. The above process can be defined as:

$$\begin{cases} f_i^p = Conv\left(W_p^1 f_i^m \oplus W_p^2 f_i^a\right) \\ F_i^p = f_i \oplus f_i \otimes \sigma\left(f_i^p\right) \end{cases} \tag{10}$$

where $\sigma\left(\cdot\right)$ denotes the sigmoid activation function.

Multi-scale Detail Boosting Unit: Inspired by [48], the multi-scale detail boosting unit performs multi-scale analysis of image by applying Gaussian filtering with changed standard deviations, capturing multiple levels of features from rough to robust. Specifically, four diverse degrees of Gaussian blur are applied to the output features f_i of the encoder side to extract the corresponding detail features, and then the generated four layers of detail features are combined element by element to generate the whole detail features. However, compared with crude details, fine details may overshoot and saturate gray values while enlarging the difference of gray values near edges, resulting in loss of image details. Therefore, we assign corresponding weight factors (W_{g}^1, W_{g}^2, W_{g}^3 and W_{g}^4), to multi-scale detail features to optimize the expression of features and the enhancement effect of images. The process can be formalized as:

$$\begin{cases} f_i^{\mathrm{g}_1} = Conv_3\left(abs\left(f_i - \mathrm{G}_{3\times3}\left(f_i\right)\right)\right) \\ f_i^{\mathrm{g}_2} = Conv_5\left(abs\left(f_i - \mathrm{G}_{5\times5}\left(f_i\right)\right)\right) \\ f_i^{\mathrm{g}_3} = Conv_7\left(abs\left(f_i - \mathrm{G}_{7\times7}\left(f_i\right)\right)\right) \\ f_i^{\mathrm{g}_4} = Conv_9\left(abs\left(f_i - \mathrm{G}_{9\times9}\left(f_i\right)\right)\right) \\ f_i^{\mathrm{g}} = Conv\left(W_g^1 f_i^{g_1} \oplus W_g^2 f_i^{g_2} \oplus W_g^3 f_i^{g_3} \oplus W_g^4 f_i^{g_4}\right) \end{cases} \tag{11}$$

where $\mathrm{G}\left(\cdot\right)$ denotes Gaussian filtering operation. Similar to the difference contrast enhancement unit, after the output feature ${f_i}^{\mathrm{g}}$ performs the sigmoid activation operation, the residual multiplication operation is used to accentuate the detail feature representation and superimpose the output feature f_i on the encoder side to obtain the detail boosting feature. The above process can be defined as:

$$F_i^g = f_i \oplus f_i \otimes \sigma\left(f_i^g\right) \tag{12}$$

In order to fully integrate and extract the fine-grained features of co-salient objects, we use spatial attention [49] operation on the output features (F_i^p and F_i^g) of two branches to reduce the learning of noise and irrelevant features, so that the model can identify and understand key features more accurately. The above process can be described by the formula:

$$F_i^{p_{att}} = F_i^p \oplus F_i^p \otimes \mathcal{F}_{SA}\left(F_i^p\right) \tag{13}$$

$$F_i^{g_{att}} = F_i^g \oplus F_i^g \otimes \mathcal{F}_{SA}\left(F_i^g\right) \tag{14}$$

Where $\mathcal{F}_{SA}\left(\cdot\right)$ denotes the spatial attention.

In addition, the attention-enhancing double-branch features ($F_i^{p_{att}}$ and $F_i^{g_{att}}$) are aggregated by weight and element-by-element addition. The sigmoid operation of the double-branch aggregation features F_i is used as the weight, element-by-element multiplication and addition operations are integrated with F_i to better focus on effective detail features and obtain clear salient boundaries. The calculation process of fusion features F_i and FEM output features F_i^{out} is expressed as follows:

$$F_i = \alpha_i F_i^{p_{att}} \oplus \beta_i F_i^{g_{att}} \tag{15}$$

$$F_i^{out} = F_i \oplus F_i \otimes \sigma\left(\mathcal{F}_{BN}\left(Conv_{3\times3}\left(F_i\right)\right)\right) \tag{16}$$

Where $\mathcal{F}_{BN}(\cdot)$ denotes Batch Normalization operations.

E. Decoder and Loss Function

Decoder: In the process of decoding, we use multi-layer semantic guidance to continuously fuse high-level semantic information with low-level detailed features to obtain the final co-salient object detection prediction. In fact, our decoder consists of five similar decoding blocks, each of which integrates the convolution layer, Batch Normalization layer and ReLU layer while fusing high-level information to work cooperatively to produce high-quality segmentation results.

Loss Function: In order to accelerate the model convergence speed and improve the accuracy of detection, we use weighted binary cross entropy loss ($\mathcal{L}_{BCE}^{\omega}$) and weighted intersection loss ($\mathcal{L}_{IOU}^{\omega}$) to perform pixel-level supervision on each decoding block in the training stage. In addition, different weights are given to predictions for each level of the decoder. Thus, the total loss is defined as:

$$\mathcal{L} = \sum_{i=1}^{5} w_i \left(\mathcal{L}_{BCE}^{\omega}\left(P_i, G\right) + \mathcal{L}_{IOU}^{\omega}\left(P_i, G\right)\right) \tag{17}$$

where G is the ground truth value and P_i is the prediction of the decoder output at each level. More details of these two losses can be found in [50], [51].

IV. EXPERIMENTS AND RESULTS

In this section, we will conduct a comprehensive experimental evaluation and performance test of the proposed CEDPNet on our proposed CoORSI dataset.

A. Implementation Details

We implement our model with PyTorch toolbox [52] on NVIDIA GeForce RTX 3090 GPU. All experiments are carried out on our proposed CoORSI dataset. In addition, CEDPNet adopts EfficientNet-B5 [44] as the backbone of the model and sets the images input size to 224×224. During the training stage, the initial learning rate is 1e-4 and the batch size is adjusted to 8.

B. Evaluation Metrics

To evaluate the performance of the proposed method, six widely used metrics were employed, including mean absolute error (MAE) [53], F-measure [54], E-measure [55], S-measure [56].

MAE ($\mathcal{M}$) [53] measures the absolute difference between predicted map (S) and ground truth (G) in a pixel by pixel manner and can be formulated as:

$$\mathcal{M} = \frac{1}{H \times W} \sum_{x=1}^{H} \sum_{y=1}^{W} \left|S\left(x, y\right) - G\left(x, y\right)\right| \tag{18}$$

where W and H denote the width and height of the salient prediction map respectively.

F-measure (F_β) [54] is a common index to evaluate network performance. It is designed to evaluate the weighted harmonic mean of precision P and recall R, and is formulated as follows:

$$F_\beta = \frac{\beta^2 + 1 \times P \times R}{\beta^2 \times P + R} \tag{19}$$

where β^2 is set to 0.3. In this paper, we use max F-measure (F_β^{max}) and mean E-measure (F_β^{mean}).

E-measure [55] (E_ϕ) is a perceptual measure that simultaneously evaluates both local pixel-level matching and global image-level statistics between the predicted map and ground truth, which can be formulated as follows:

$$E_\phi = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \phi_{FM}\left(x, y\right) \tag{20}$$

where $\phi_{FM}(\cdot)$ denotes the enhanced alignment matrix. In this paper, we use max E-measure (E_ϕ^{max}) and mean E-measure (E_ϕ^{mean}).

S-measure [56] (S_α) means using area perception (S_r) and object perception (S_o) to evaluate the similarity of spatial structure, the formula is expressed as:

$$S_\alpha = \alpha S_o + \left(1 - \alpha\right) S_r \tag{21}$$

where the function of the parameter α is to regulate the balance between (S_o) and (S_r), is set to 0.5. More details please refer to the public available evaluation toolbox[1].

C. Comparison with SOTAs

In this section, we compare our proposed model in detail with other representative models. There are no published models in the field of Co-salient object detection in optical remote sensing. Therefore, we compare our proposed model with 12 recently proposed representative remote sensing salient object detection models (ACCoNet [57], CorrNet [58], ERPNet [25], SeaNet [59], GeleNet [60], SEINet [22], BSCGNet [61], MEANet [27], TSCNet [62], TLCKD-Net [63], GSANet [64], SFANet [65]) and 7 natural scene co-salient object detection models (CADC [40], DCFM [41], TCNet [66], CoRP [42], MCCL [43], GCoNet+ [67], DMT [68]). It is worth noting that the results of all these methods are generated by retraining on our CoORSI dataset according to the code published by the authors.

1) Quantitative Evaluation: The result in TABLE II shows that our proposed CEDPNet outperforms competitors on six widely used evaluation metrics on the CoORSI dataset. Compared with other models, our model not only explores collaborative objects in parallel, but also adopts the strategies of difference perception enhancement and multi-scale detail boosting to achieve fine identification of co-salient objects. Meanwhile, Fig. 9 shows the PR curve and F-measure curve of CEDPNet and other 7 CoCOD methods and 12 ORSI-SOD methods on the CoORSI dataset. It can be observed that our proposed method is superior to other methods, the PR curve of the CEDPNet is closest to the coordinate (1, 1), and the region below the F-measure curve is also the largest. Moreover, we also compare all the methods involved in terms of FLOPs and Params. It can be seen that while improving the performance of

[1]https://github.com/zzhanghub/eval-co-sod

TABLE II

THE QUANTITATIVE COMPARISON WITH 19 STATE-OF-THE-ART MODELS ON CORSI UNDER SIX WIDELY USED EVALUATION METRICS, INCLUDING MAE ($\mathcal{M}$), MAXIMUM F-MEASURE (F_β^{max}), MEAN F-MEASURE (F_β^{mean}), MAXIMUM E-MEASURE (E_ϕ^{max}), MEAN E-MEASURE (E_ϕ^{mean}), S-MEASURE (S_α). ↑ MEANS THAT THE HIGHER THE SCORE, THE BETTER. ↓ INDICATES THAT THE LOWER THE SCORE, THE BETTER. THE TOP THREE RESULTS ARE HIGHLIGHTED IN RED, GREEN, AND BLUE FOR CLEAR DISTINCTION.

Method	Publish	Backbone	FLOPs (G)	Params (M)	$\mathcal{M}\downarrow$	$F_\beta^{max}\uparrow$	$F_\beta^{mean}\uparrow$	$E_\phi^{max}\uparrow$	$E_\phi^{mean}\uparrow$	$S_\alpha\uparrow$
CADC [40]	ICCV(2021)	VGG	35.21	392.85	0.0641	0.6988	0.5937	0.8622	0.7354	0.7384
DCFM [41]	CVPR(2022)	VGG	31.62	18.67	0.0584	0.6607	0.6551	0.8317	0.8269	0.7407
CoRP [42]	TPAMI(2023)	VGG	36.57	18.27	0.0306	0.8067	0.7923	0.9169	0.9023	0.8430
MCCL [43]	AAAI(2023)	PVT	4.37	27.04	0.1132	0.4547	0.4089	0.7444	0.6570	0.6153
GCoNet+ [67]	TPAMI(2023)	VGG	36.50	18.42	0.0590	0.6031	0.5682	0.8021	0.7544	0.6934
TCNet [66]	TCSVT(2023)	VGG+ViT	40.77	69.40	0.0226	0.8434	0.8301	0.9514	0.9457	0.8731
DMT [68]	CVPR(2023)	VGG+MaskFormer	18.28	28.03	0.0410	0.7506	0.7199	0.8932	0.8671	0.7921
CorrNet [58]	TGRS(2022)	VGG	16.32	4.07	0.0757	0.5899	0.5708	0.6930	0.6716	0.6865
ACCoNet [57]	TCYB(2023)	VGG	141.26	102.55	0.0564	0.7558	0.7527	0.8602	0.8533	0.8007
ERPNet [25]	TCYB(2023)	ResNet	131.54	77.19	0.0770	0.6835	0.6572	0.8221	0.7748	0.7474
SeaNet [59]	TGRS(2023)	MobileNet	1.10	2.75	0.0810	0.6250	0.6214	0.7751	0.7687	0.7261
GeleNet [60]	TIP(2023)	PVT	4.74	25.45	0.0393	0.7575	0.7517	0.9109	0.9069	0.8247
SEINet [22]	J-STARS(2023)	EfficientNet	5.84	5.56	0.0504	0.7489	0.7461	0.8648	0.8592	0.8094
BSCGNet [61]	TGRS(2023)	VGG	66.20	26.99	0.0839	0.6050	0.5978	0.7401	0.7303	0.7205
TSCNet [62]	TCAS-II(2023)	VGG+ViT	110.58	101.20	0.0335	0.8186	0.8121	0.9246	0.9214	0.8513
MEANet [27]	ESWA(2024)	MobileNet	4.59	3.27	0.0860	0.6623	0.6590	0.7984	0.7919	0.7318
TLCKD [63]	CVIU(2024)	ResT	32.51	50.41	0.0246	0.8337	0.8104	0.9405	0.9189	0.8658
SFANet [65]	TGRS(2024)	Res2Net	5.97	25.10	0.0539	0.7517	0.7488	0.8623	0.8555	0.7958
GSANet [64]	Entropy(2024)	UniFormer	8.02	49.44	0.0206	0.8461	0.8354	0.9516	0.9462	0.8736
OUR	-	EfficientNet	15.94	30.77	0.0186	0.8545	0.8377	0.9551	0.9479	0.8819

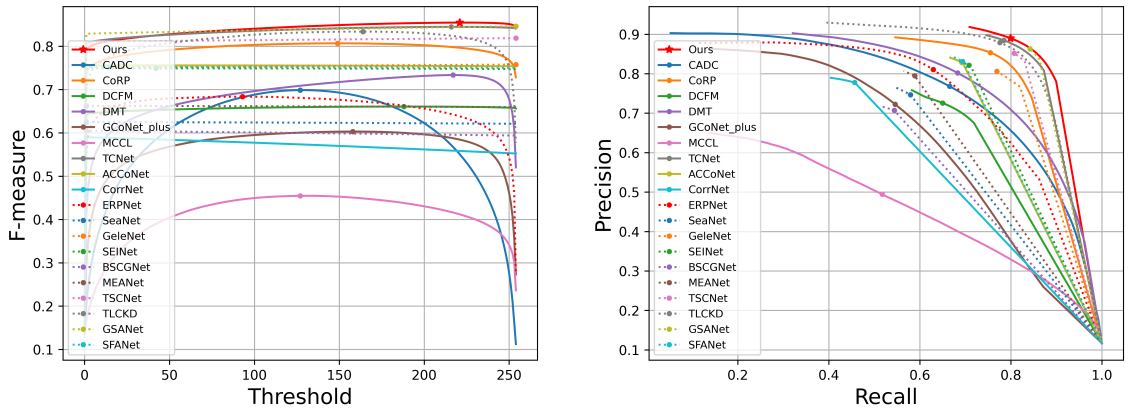

Fig. 9: Illustration of F-measure curves (left column) and PR curves (right column) on the CoORSI dataset. Please zoom in for detailed view.

the model, the parameter scale and computational complexity of CEDPNet are lower than that of the CoSOD sub-optimal model TCNet, and the overall level is moderate.

2) Qualitative Evaluation: Fig. 10 shows the visual comparison of the proposed method with other remote sensing salient object detection methods, including five challenging image groups, which intuitively embodies the excellent performance of our proposed model. It can be seen that the SOD model rarely considers category consistency, and often introduces other non-cooperative salient objects, especially in the aircraft group, where routes are mistaken for the co-salient object. However, through regional and pixel-level correlation statistics, CEDPNet avoids interference of non-cooperative salient objects and realizes accurate positioning and identification of co-salient objects. In addition, in comparison to other methods, our approach notably enhances the detection capabilities for both multi-scale and small-sized objects, ensuring that a variety of targets, such as cars and ships, are accurately identified. At the same time, our model effectively maintains the structural integrity of the object in the river image set with complex topology.

Fig. 11 shows a visual comparison of the proposed method with the co-salience object detection method, including four challenging image groups. Compared to the SOD model, CoSOD showed greater accuracy in distinguishing common salient areas. However, in challenging scenarios such as complex background and shadow interference, other co-salient object detection models may have boundary blurring and redundant noise. In contrast to others, our approach employs a dual strategy of differential perception enhancement and multi-scale detail boosting to address the interference from complex scenes and cluttered backgrounds. This enables us to accurately and comprehensively localize and segment co-salient objects within diverse complex scenarios, thereby delivering

Fig. 10: Visual comparison of the proposed model with other 12 existing state-of-the-art ORSI-SOD algorithms on proposed CoORSI datasets.

optimal visual performance.

D. Ablation Study

In this section, we perform multiple ablation experiments on the CoORSI dataset with the same parameter settings to verify the validity of the proposed model and each module. The detailed analysis is as follows.

Effectiveness of FSM. The feature sensing module aims to achieve fine-grained detection of salient objects from two perspectives of difference contrast enhancement and multi-scale detail boosting. According to TABLE III, the addition of the FSM module achieves a significant performance improvement compared to the model baseline. Specifically, the performance gains are 25.09%(0.007), 4.52%(0.036), 3.35%(0.027), 2.66%(0.025), 2.71%(0.025), 2.70%(0.023) in terms of $\mathcal{M}$, $F_\beta^{\max}$, F_β^{mean}, $E_\beta^{\max}$, E_β^{mean} and S_α, respectively.

For two sub-units in the FSM module (DCEU and MDBU), we performed ablation experiments to verify their effectiveness in the model. By removing these sub-units in turn and retraining the model, we were able to quantify their specific

Fig. 11: Visual comparison of the proposed model with other 7 existing state-of-the-art CoSOD algorithms on proposed CoORSI datasets.

TABLE III
MODULE ABLATION STUDY. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Model			CoORSI					
Baseline	FSM	COSM	$\mathcal{M}\downarrow$	$F_\beta^{max}\uparrow$	$F_\beta^{mean}\uparrow$	$E_\phi^{max}\uparrow$	$E_\phi^{mean}\uparrow$	$S_\alpha\uparrow$
✓			0.0291	0.8017	0.7901	0.9210	0.9127	0.8453
✓	✓		0.0218	0.8379	0.8166	0.9455	0.9374	0.8681
✓		✓	0.0221	0.8352	0.8110	0.9446	0.9356	0.8649
✓	✓	✓	**0.0186**	**0.8545**	**0.8377**	**0.9551**	**0.9479**	**0.8819**

TABLE IV
EXPERIMENTAL RESULTS OF FSM INTERNAL ABLATION. THE EXPERIMENT PROVES THE EFFECTIVENESS OF DCEU AND MDBU. THE BEST OF THESE RESULTS ARE SHOWN IN BOLD.

Model	CoORSI					
	$\mathcal{M}\downarrow$	$F_\beta^{max}\uparrow$	$F_\beta^{mean}\uparrow$	$E_\phi^{max}\uparrow$	$E_\phi^{mean}\uparrow$	$S_\alpha\uparrow$
w/o DCEU	0.0195	0.8545	0.8340	0.9539	0.9468	0.8799
w/o MDBU	0.0196	0.8529	0.8325	0.9534	0.9463	0.8789
OURs	**0.0186**	**0.8545**	**0.8377**	**0.9551**	**0.9479**	**0.8819**

contribution to overall performance, as shown in TABLE IV. The experimental results show that, compared with the model that removes the difference contrast enhancing unit and the multi-scale detail enhancing unit respectively, our model achieves the optimal performance in all indexes.

Effectiveness of COSM. The collaborative object search module aims to accurately locate and identify cooperative objects from two aspects of pixel and region. According to TABLE III, the addition of COSM module also achieved a marked improvement in model performance compared to the Baseline. Specifically, the performance improvements of $\mathcal{M}$, $F_\beta^{\max}$, F_β^{mean}, $E_\beta^{\max}$, E_β^{mean} and S_α are 24.05%(0.007), 4.18%(0.034), 2.65%(0.021), 2.56%(0.024), 2.51%(0.023), 2.32%(0.020), respectively.

We performed ablation experiments on two key components of COSM (hierarchical feature fusion block and collaborative feature mining block) to demonstrate their effectiveness, and the experimental results are shown in TABLE V. The experimental results show that hierarchical feature fusion block and collaborative feature mining block can promote the effective detection of co-salient objects.

TABLE V
EXPERIMENTAL RESULTS OF COSM INTERNAL ABLATION. THE EXPERIMENT PROVES THE EFFECTIVENESS OF HIERARCHICAL FEATURE FUSION BLOCK AND COLLABORATIVE FEATURE MINING BLOCK. THE BEST OF THESE RESULTS ARE SHOWN IN BOLD

Model	CoORSI					
	$\mathcal{M}\downarrow$	$F_\beta^{max}\uparrow$	$F_\beta^{mean}\uparrow$	$E_\phi^{max}\uparrow$	$E_\phi^{mean}\uparrow$	$S_\alpha\uparrow$
w/o HFFB	0.0193	0.8484	0.8295	0.9516	0.9439	0.8789
w/o CFMB	0.0210	0.8495	0.8276	0.9504	0.9432	0.8771
OURs	**0.0186**	**0.8545**	**0.8377**	**0.9551**	**0.9479**	**0.8819**

For two branches in the CFMB (pixel correlation branch and region correlation branch), we performed ablation experiments to verify their validity in the model, and the experimental results are shown in TABLE VI. The experimental results show that the consensus feature extraction between pixel correlation branches and regional correlation branches plays a positive role in co-salient object detection. In addition, in order to verify the validity of the consensus feature representation using top-k filtering at different levels, we adopted different initial parameter configurations for experimental verification. The experimental results are shown in TABLE VII. When the top-k selection of the last three layers is $(16, 4, 1)$, the model performance reaches the best. Excessive pixel selection may lead to the introduction of some unnecessary noise or redundant information, which will have a negative impact on the performance of the model.

TABLE VI
EXPERIMENTAL RESULTS OF CFMB INTERNAL ABLATION. THE EXPERIMENT PROVES THE EFFECTIVENESS OF PIXEL CORRELATION BRANCH AND REGION CORRELATION BRANCH. THE BEST OF THESE RESULTS ARE SHOWN IN BOLD

Model	CoORSI					
	$\mathcal{M}\downarrow$	$F_\beta^{max}\uparrow$	$F_\beta^{mean}\uparrow$	$E_\phi^{max}\uparrow$	$E_\phi^{mean}\uparrow$	$S_\alpha\uparrow$
w/o pixel	0.0199	0.8513	0.8308	0.9518	0.9448	0.8788
w/o region	0.0217	0.8375	0.8133	0.9455	0.9373	0.8579
OURs	**0.0186**	**0.8545**	**0.8377**	**0.9551**	**0.9479**	**0.8819**

TABLE VII
THE TOP-K INITIAL PARAMETER SETTINGS. THE BEST OF THESE RESULTS ARE SHOWN IN BOLD.

Top-K	CoORSI					
	$\mathcal{M}\downarrow$	$F_\beta^{max}\uparrow$	$F_\beta^{mean}\uparrow$	$E_\phi^{max}\uparrow$	$E_\phi^{mean}\uparrow$	$S_\alpha\uparrow$
$(1, 1, 1)$	0.0216	0.8369	0.8220	0.9452	0.9377	0.8640
$(4, 2, 1)$	0.0209	0.8350	0.8110	0.9444	0.9362	0.8646
$(16, 4, 1)$	**0.0186**	**0.8545**	**0.8377**	**0.9551**	**0.9479**	**0.8819**

Visual comparison of each proposed module. We present intuitive visualizations in Fig. 12 to confirm the effectiveness of the modules we designed. As shown in the third column, when only the baseline is used to detect cooperative objects, part of the background is mistaken for an object. Further, baselines with FSM modules enhance the awareness of salient objects, and baselines with COSM modules enhance the recognition accuracy of cooperative objects. As shown in the last column, the two modules we propose work together to more effectively achieve precise positioning and clear detection of co-salient objects from complex backgrounds.

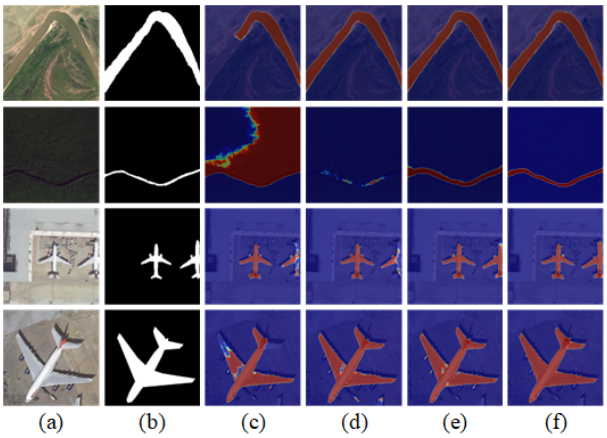

Fig. 12: Visual comparison of modules and our final model in ablation experiments. (a) Images. (b) GT. (c) Baseline. (d) Baseline+FSM. (e) Baseline+COSM. (f) Ours.

Flexibility to Various Backbones. We use different backbones instead of the original backbone for experimental verification to further prove the flexibility and adaptability of the proposed CEDPNet. As shown in TABLE VIII, the comparison results show that the model performs best when EfficientNet-B5 [44] network is used as the backbone of the model. In addition, the effectiveness of feature sensing module and collaborative object search module is also verified for different backbone networks. Our proposed CEDPNet provides an efficient solution for co-salient object detection in optical remote sensing images with its excellent performance and flexibility.

TABLE VIII
PERFORMANCE EVALUATION OF OUR CEDPNET ON DIFFERENT ENCODER BACKBONES. THE BEST OF THESE RESULTS ARE SHOWN IN BOLD.

Backbone	CoORSI					
	$\mathcal{M}\downarrow$	$F_\beta^{max}\uparrow$	$F_\beta^{mean}\uparrow$	$E_\phi^{max}\uparrow$	$E_\phi^{mean}\uparrow$	$S_\alpha\uparrow$
VGG-16 [69]	0.0211	0.8466	0.8237	0.9488	0.9409	0.8733
ResNet-50 [70]	0.0200	0.8473	0.8267	0.9503	0.9431	0.8757
Res2Net-50 [71]	0.0209	0.8486	0.8306	0.9507	0.9439	0.8786
MobileNet-V2 [72]	0.0204	0.8436	0.8237	0.9493	0.9422	0.8735
EfficientNet-B5 [44]	**0.0186**	**0.8545**	**0.8377**	**0.9551**	**0.9479**	**0.8819**

E. Failure Case Analyses

In Fig. 13, we present some representative failure cases of CEDPNet that reveal the current challenges of our approach. Case (a) and case (b) show that when multiple co-salient objects are clustered closely or narrow, there is a phenomenon

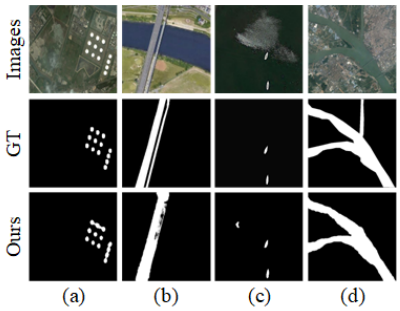

Fig. 13: Examples of failure cases.

that the segmentation regions have adhesion and fail to form a clear boundary. Case (c) shows that in the case of occlusion, the target object is confused with the background, resulting in a false identification. Case (d) shows that in the case of complex environmental interference, the contrast between the target object and the background is too low to detect, resulting in incomplete targets to be detected. In the future, we will improve the model in many aspects, including but not limited to improving the segmentation accuracy of small size or nearby objects, enhancing the processing ability of occlusion cases, and improving the detection performance of low-contrast objects in complex environments, in order to improve the overall performance and robustness of the model.

V. CONCLUSION

In this paper, we construct CoORSI, the first large-scale dataset suitable for Co-salient object detection in optical remote sensing task. At the same time, we propose a multi-layer semantic guidance network to realize the preliminary exploration of remote sensing collaborative tasks. CEDPNet enhances the fine-grained perception of salient objects by difference contrast enhancement and multi-scale detail boosting. At the same time, it adopts the correlation calculation of pixel and region level to realize the detection and identification of collaborative objects, so as to realize the accurate positioning and fine detection of remote sensing co-salient objects. In addition, quantitative and qualitative experiments demonstrate the effectiveness of the proposed CEDPNet and its superior performance over other competitors. We hope that the research presented in this work (i.e., dataset and baseline) will serve as a catalyst for further exploration in the field, bringing new insights and innovative ideas to the academic community.

ACKNOWLEDGMENTS

The work was supported by the National Natural Science Foundation of China (No. 62471124), Heilongjiang Province Natural Science Foundation (No. LH2022F005) and Young Top Talents Fund in the School of Electrical Information Engineering of Northeast Petroleum University (No. DYDQQB202204).

REFERENCES

[1] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A benchmark," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5706–5722, 2015.

[2] R. Cong, J. Lei, H. Fu, M.-M. Cheng, W. Lin, and Q. Huang, "Review of visual saliency detection with comprehensive information," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 10, pp. 2941–2959, 2019.

[3] W. Wang, Q. Lai, H. Fu, J. Shen, H. Ling, and R. Yang, "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] D. Hong, W. He, N. Yokoya, J. Yao, L. Gao, L. Zhang, J. Chanussot, and X. Zhu, "Interpretable hyperspectral artificial intelligence: When non-convex modeling meets hyperspectral remote sensing," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 2, pp. 52–87, 2021.

[5] Q. Zhang, R. Cong, C. Li, M.-M. Cheng, Y. Fang, X. Cao, Y. Zhao, and S. Kwong, "Dense attention fluid network for salient object detection in optical remote sensing images," *IEEE Transactions on Image Processing*, vol. 30, pp. 1305–1317, 2021.

[6] R. Cong, Y. Zhang, L. Fang, J. Li, Y. Zhao, and S. Kwong, "Rrnet: Relational reasoning network with parallel multiscale attention for salient object detection in optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022.

[7] D.-P. Fan, T. Li, Z. Lin, G.-P. Ji, D. Zhang, M.-M. Cheng, H. Fu, and J. Shen, "Re-thinking co-salient object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, pp. 4339–4354, 2022.

[8] D. Zhang, J. Han, C. Li, and J. Wang, "Co-saliency detection via looking deep and wide," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 2994–3002.

[9] Y. Ge, Q. Zhang, T.-Z. Xiang, C. Zhang, J. Zhang, and H. Bi, "Gsnnet: Group semantic-guided neighbor interaction network for co-salient object detection," *Computer Vision and Image Understanding*, vol. 227, p. 103611, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314222001898

[10] C. Zhang, H. Bi, T.-Z. Xiang, R. Wu, J. Tong, and X. Wang, "Collaborative camouflaged object detection: A large-scale dataset and benchmark," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2023.

[11] F. Yang, Q. Xu, and B. Li, "Ship detection from optical satellite images based on saliency segmentation and structure-lbp feature," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 602–606, 2017.

[12] P. P S, J. Soni, and B. H A, "Building extraction from remote sensing images using deep learning and transfer learning," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 3079–3082.

[13] T. Wellmann, A. Lausch, E. Andersson, S. Knapp, C. Cortinovis, J. Jache, S. Scheuer, P. Kremer, A. Mascarenhas, R. Kraemer, A. Haase, F. Schug, and D. Haase, "Remote sensing in urban planning: Contributions towards ecologically sound policies?" *Landscape and Urban Planning*, vol. 204, p. 103921, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169204620308860

[14] Y. Liu, Z. Xiong, Y. Yuan, and Q. Wang, "Transcending pixels: Boosting saliency detection via scene understanding from aerial imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.

[15] C. Li, R. Cong, J. Hou, S. Zhang, Y. Qian, and S. Kwong, "Nested network with two-stream pyramid for salient object detection in optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9156–9166, 2019.

[16] Z. Tu, C. Wang, C. Li, M. Fan, H. Zhao, and B. Luo, "Orsi salient object detection via multiscale joint region and boundary model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[17] Q. Zheng, L. Zheng, Y. Bai, H. Liu, J. Deng, and Y. Li, "Boundary-aware network with two-stage partial decoders for salient object detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023.

[18] Z. Dong, M. Wang, Y. Wang, Y. Zhu, and Z. Zhang, "Object detection in high resolution remote sensing imagery based on convolutional neural networks with suitable object scale features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 3, pp. 2104–2114, 2020.

[19] L. Zhang and K. Yang, "Region-of-interest extraction based on frequency domain analysis and salient region detection for remote sensing image," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 5, pp. 916–920, 2014.

[20] L. Zhang, K. Yang, and H. Li, "Regions of interest detection in panchromatic remote sensing images based on multiscale feature fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 12, pp. 4704–4716, 2014.

[21] L. Zhang and A. Li, "Region-of-interest extraction based on saliency analysis of co-occurrence histogram in high spatial resolution remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 5, pp. 2111–2124, 2015.

[22] H. Luo and B. Liang, "Semantic-edge interactive network for salient object detection in optical remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 6980–6994, 2023.

[23] J. Zhao, Y. Jia, L. Ma, and L. Yu, "Adaptive dual-stream sparse transformer network for salient object detection in optical remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 5173–5192, 2024.

[24] Q. Wang, Y. Liu, Z. Xiong, and Y. Yuan, "Hybrid feature aligned network for salient object detection in optical remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[25] X. Zhou, K. Shen, L. Weng, R. Cong, B. Zheng, J. Zhang, and C. Yan, "Edge-guided recurrent positioning network for salient object detection in optical remote sensing images," *IEEE Transactions on Cybernetics*, vol. 53, no. 1, pp. 539–552, 2023.

[26] A. Gong, J. Nie, C. Niu, Y. Yu, J. Li, and L. Guo, "Edge and skeleton guidance network for salient object detection in optical remote sensing images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 12, pp. 7109–7120, 2023.

[27] B. Liang and H. Luo, "Meanet: An effective and lightweight solution for salient object detection in optical remote sensing images," *Expert Systems with Applications*, vol. 238, p. 121778, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423022807

[28] Z. Tan, L. Wan, W. Feng, and C.-M. Pun, "Image co-saliency detection by propagating superpixel affinities," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 2114–2118.

[29] H. Li and K. N. Ngan, "A co-saliency model of image pairs," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3365–3375, 2011.

[30] H. Fu, X. Cao, and Z. Tu, "Cluster-based co-saliency detection," *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 3766–3778, 2013.

[31] Z. Liu, W. Zou, L. Li, L. Shen, and O. Le Meur, "Co-saliency detection based on hierarchical segmentation," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 88–92, 2014.

[32] L. Li, Z. Liu, W. Zou, X. Zhang, and O. Le Meur, "Co-saliency detection based on region-level fusion and pixel-level refinement," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*, 2014, pp. 1–6.

[33] X. Cao, Z. Tao, B. Zhang, H. Fu, and X. Li, "Saliency map fusion based on rank-one constraint," in *2013 IEEE International Conference on Multimedia and Expo (ICME)*, 2013, pp. 1–6.

[34] X. Cao, Y. Cheng, Z. Tao, and H. Fu, "Co-saliency detection via base reconstruction," *Proceedings of the 22nd ACM international conference on Multimedia*, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:15415561

[35] L. Ye, Z. Liu, J. Li, W.-L. Zhao, and L. Shen, "Co-saliency detection via co-salient object discovery and recovery," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 2073–2077, 2015.

[36] R. Huang, W. Feng, and J. Sun, "Saliency and co-saliency detection by low-rank multiscale fusion," in *2015 IEEE International Conference on Multimedia and Expo (ICME)*, 2015, pp. 1–6.

[37] Z. Zhang, W. Jin, J. Xu, and M.-M. Cheng, "Gradient-induced co-saliency detection," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*. Springer, 2020, pp. 455–472.

[38] W. Jin, J. Xu, M.-M. Cheng, Y. Zhang, and W. Guo, "Icnet: Intra-saliency correlation network for co-saliency detection," in *Neural Information Processing Systems*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:227275180

[39] G. Gao, W. Zhao, Q. Liu, and Y. Wang, "Co-saliency detection with co-attention fully convolutional network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 3, pp. 877–889, 2021.

[40] N. Zhang, J. Han, N. Liu, and L. Shao, "Summarize and search: Learning consensus-aware dynamic convolution for co-saliency detection," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4147–4156, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:238253141

[41] S. Yu, J. Xiao, B. Zhang, and E. G. Lim, "Democracy does matter: Comprehensive feature mining for co-salient object detection," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 969–978, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:247411421

[42] Z. Zhu, Z. Zhang, Z. Lin, X. Sun, and M.-M. Cheng, "Co-salient object detection with co-representation purification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 8193–8205, 2023.

[43] P. Zheng, J. Qin, S. Wang, T.-Z. Xiang, and H. Xiong, "Memory-aided contrastive consensus learning for co-salient object detection," in *AAAI Conference on Artificial Intelligence*, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:257232511

[44] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[45] Q. Fan, D.-P. Fan, H. Fu, C.-K. Tang, L. Shao, and Y.-W. Tai, "Group collaborative learning for co-salient object detection," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 12 283–12 293.

[46] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Neural Information Processing Systems*, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:13756489

[47] H. Lin, X. Cheng, X. Wu, F. Yang, D. Shen, Z. Wang, Q. Song, and W. Yuan, "Cat: Cross attention in vision transformer," *2022 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:235390617

[48] Y. Kim, Y. J. Koh, C. Lee, S. Kim, and C.-S. Kim, "Dark image enhancement based on pairwise target contrast and multi-scale detail boosting," in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 1404–1408.

[49] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[50] D.-P. Fan, G.-P. Ji, M.-M. Cheng, and L. Shao, "Concealed object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, pp. 6024–6042, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:231985788

[51] J. Wei, S. Wang, and Q. Huang, "F^3net: fusion, feedback and focus for salient object detection," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 12 321–12 328.

[52] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[53] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, and A. Borji, "Enhanced-alignment measure for binary foreground map evaluation," in *International Joint Conference on Artificial Intelligence*, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:44072899

[54] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1597–1604.

[55] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 1529–1536.

[56] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji, "Structure-measure: A new way to evaluate foreground maps," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4558–4567.

[57] G. Li, Z. Liu, D. Zeng, W. Lin, and H. Ling, "Adjacent context coordination network for salient object detection in optical remote sensing images," *IEEE Transactions on Cybernetics*, vol. 53, no. 1, pp. 526–538, 2023.

[58] G. Li, Z. Liu, Z. Bai, W. Lin, and H. Ling, "Lightweight salient object detection in optical remote sensing images via feature correlation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[59] G. Li, Z. Liu, X. Zhang, and W. Lin, "Lightweight salient object detection in optical remote-sensing images via semantic matching and

edge alignment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–11, 2023.
[60] G. Li, Z. Bai, Z. Liu, X. Zhang, and H. Ling, "Salient object detection in optical remote sensing images driven by transformer," *IEEE Transactions on Image Processing*, vol. 32, pp. 5257–5269, 2023.
[61] D. Feng, H. Chen, S. Liu, Z. Liao, X. Shen, Y. Xie, and J. Zhu, "Boundary-semantic collaborative guidance network with dual-stream feedback mechanism for salient object detection in optical remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–17, 2023.
[62] G. Li, Z. Bai, and Z. Liu, "Texture-semantic collaboration network for orsi salient object detection," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 71, pp. 2464–2468, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:265262327
[63] P. Dong, B. Wang, R. Cong, H.-H. Sun, and C. Li, "Transformer with large convolution kernel decoder network for salient object detection in optical remote sensing images," *Computer Vision and Image Understanding*, vol. 240, p. 103917, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314223002977
[64] H. Li, X. Chen, W. Yang, J. Huang, K. Sun, Y. Wang, A. Huang, and L. Mei, "Global semantic-sense aggregation network for salient object detection in remote sensing images," *Entropy*, vol. 26, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:270078371
[65] Y. Quan, H. Xu, R. Wang, Q. Guan, and J. Zheng, "Orsi salient object detection via progressive semantic flow and uncertainty-aware refinement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–13, 2024.
[66] Y. Ge, Q. Zhang, T.-Z. Xiang, C. Zhang, and H. Bi, "Tcnet: Co-salient object detection via parallel interaction of transformers and cnns," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 6, pp. 2600–2615, 2023.
[67] P. Zheng, H. Fu, D.-P. Fan, Q. Fan, J. Qin, Y.-W. Tai, C.-K. Tang, and L. Van Gool, "Gconet+: A stronger group collaborative co-salient object detector," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 9, pp. 10 929–10 946, 2023.
[68] L. Li, J. Han, N. Zhang, N. Liu, S. Khan, H. Cholakkal, R. M. Anwer, and F. S. Khan, "Discriminative co-saliency and background mining transformer for co-salient object detection," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 7247–7256.
[69] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:14124313
[70] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
[71] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 2, pp. 652–662, 2021.
[72] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.