IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 May 2024, accepted 5 June 2024, date of publication 11 June 2024, date of current version 19 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3412796

RESEARCH ARTICLE

A Semantic Segmentation and Distortion Correction of Fragment Perforation Small Target

JUNCHAI GAO, HAORUI HAN, AND HANSHAN LI
School of Electronic and Information Engineering, Xi'an Technological University, Xi'an 710021, China

Corresponding author: Junchai Gao (junchaigao1212@yeah.net)

This work was supported in part by the Key Science and Technology Program of Shaanxi Province under Grant 2023-YBGY-342, and in part by the National Natural Science Foundation of China under Grant 62073256.

ABSTRACT In order to solve the superimposed image distortion caused by explosion shock wave and shooting tilt in warhead explosion test, and the problem of small target of the fragment perforation missed detection, a method combining image correction and small target semantic segmentation is proposed. Based on mesh planarization and homomorphic inverse projection transformation, the distortion of fragment perforation image is corrected. For the reliability and accuracy of small fragment perforation segmentation, based on a encoder-decoder structure, a DSA-EF(Double Self-Attention and Embedded Fusion) semantic segmentation net is proposed. In the encoder part, swin transformer block is as the backbone network to extract the multi-scale features, especially the small fragment perforation features, and the receptive field expansion block is used to fuse the multi-scale features in order to enhance features representation. In the decoder part, the channel attention mechanism of ECA block is introduced to consider the correlation between different channels of multi-scale target feature, and remove interference from redundant information, while retain shallow small fragment perforations. Through simulation experiments, this method can effectively correct the superimposed image distortion of the target plate. The ablation experiments verify the specific impact of each component on model performance, and proved the effectiveness of each modules for the entire method. Compared with five semantic segmentation methods of FCN and Seg Former et al., it shows that this method has better segmentation performance, effectively solves the problem of small fragment perforation target recognition, and improves the segmentation reliability and accuracy of small target.

INDEX TERMS Image correction, small target detection, attention mechanism, image segmentation, area calculation.

I. INTRODUCTION

In the study of the damage probability of warhead explosion [1], [2], the most commonly used method is static explosion experiment, which reflects the performance of warhead by collecting various parameters of fragments generated by explosion, such as flying speed, angle, position and so on. At present, the most commonly used methods for parameter test of fragments are divided into two categories, one is the contact measurement method, and the other is the noncontact test [3].

For the contact test method is widely used because of its simple and intuitive reaction to the results of the experiment. The main method used in the shooting range is the target plate test method [4]. This method sets the target plate around the explosion range of the warhead. After detonating the warhead, the fragment hits the target plate, and the damage performance of the warhead is reflected by counting the parameter information such as the location, distribution density and perforation size of the fragment. The target plate test methods commonly used in domestic shooting range include comb target, spherical target, ring target and so on. The comb target is used to test the velocity of fragments [5], [6], and the spherical target and ring target are used to count the

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

distribution density of fragments with manual measurement, which have problems such as low statistical efficiency and huge workload. In particular, manual measurement cannot accurately measure irregular perforation area. Subsequently, the development of computer vision [7] has provided a new way method for fragments perforation area measurement, and the research on target detection of fragment target images has gradually increased. Gao and Zhang [8] used photogrammetric technique to measure of the warhead detonation center, and the 3D point cloud of the detonation field shape is reconstructed. Finally, the fragment dispersion parameters are calculated.

For the noncontact test method, such as camera intersection measurement method [9] forms an intersection detection plane by arranging two CCD cameras in front of the target plate to detect the passing target. However, due to the high requirements for the site and environment, the complexity of the camera layout, and the inability to test the fragment group, it is not suitable for static explosion experiments. Watson et al. [10] uses two synchronized high-speed cameras with split images, to analyze the image sequences by identifying fragments and their trajectories and finally measure fragments' 3D velocities in the hypervelocity process. Sequeira et al. [11] through the icosahedral imaging system to determine mass and volume features of fragments with shape-from-silhouette 3D reconstruction. Hu et al. [12] utilized high-speed stereo photography with background difference algorithm to extract the center and area of each fragment in the image sequence.

Both in the target plate contact test method, and in the camera noncontact test method, the camera imaging technology is used. However, it is often difficult to effectively measure the damage of the target plate due to the influence of measurement conditions, environment and irregularity of the perforation area. If there is an error in the calculation of the perforation area, it will have a serious impact on the calculation of the damage probability. Therefore, it is of great significance to accurately calculate the perforation area of the target plate for studying the damage probability.

But most of the above image processing methods are based on conventional image processing algorithms, which mainly rely on professional knowledge and experience to design, and they are interpretable with better performance in some specific scenes, but the generalizability is weak. Because the fragment target acquisition is in the outdoor natural scene, the scene changes with time and space, the generalization performance of the processing algorithm is weak, which seriously affects the normal running of the target detection system. By learning the inherent laws and representation levels of the sample data, deep learning can make the machine have the ability of analysis and learning like a human, and can realize target detection. For the general depth learning target detection model, the basic backbone neural network (VGG series and Resnet series) has several times of down-sampling processing. if the classification and regression operations are performed in the feature layer after several layers of sampling processing, the receptive field of the small fragments feature mapping back to the original image may be larger than the size of the small fragment in the original image, resulting in poor detection effect. At the same time, the size of small fragments in the feature map is only single-digit pixels, and the number of extracted features is few. Therefore, this paper studies the small target segmentation algorithm. At the same time, because the target image deforms with the deformation of target plate and the geometric deformation of projective imaging, which seriously affects the accuracy of target location area calculation, an image distortion correction algorithm is designed.

II. RELATED WORK

Image distortion correction is the premise of accurate measurement of target, so domestic and foreign scholars have done a lot of research on this aspect. For image distortion correction, most of them is for lens distortion correction [13], [14]. Considering the geometric distortion when shooting the target image, He et al [15] proposed a target plate image preprocessing based on inverse perspective transformation, the distortion of target plate image in the open country is overcome. For the problem of shooting angle distortion, which affects the measurement, identification and classification of PV modules, Wang et al. [16] established a distortion correction model, based on matching a method of image distortion correction for photovoltaics module without calibration is studied, which improves the image visual correction effect and data accuracy. For the target plate deform to curved surface, the correction of the curved surface belongs to the surface parameterization method, which is based on the 3D point cloud to parameterized expansion. Based on the graph theory and the intelligent optimization method, Zhang [17] unfolds the curved surface of the ship's outer plate to get the plate which is easy to process and manufacture. Based on the quasi conformal mapping, an effective method of planar region parameterization is proposed to keep the mapped area and angle unchanged as far as possible [18].

Target segmentation is the key to the calculation of target parameters and target tracking, so a lot of research has been done in this area. There are two main methods for object segmentation: traditional image processing and artificial intelligence deep learning. In the early days, Li [19] proposed to use image processing methods such as edge detection to realize the identification and analysis of fragment perforation, but only rough statistics were made on the position of the target hole, and no more parameter information was obtained. He et al. [15] applied 2D-OTSU robust threshold segmentation based on mean-median combined filtering for the fragment damage area, carry out target damage area recognition. Song [20] proposed a fragment target image recognition based on improved Otsu threshold algorithm. The optimal threshold of the algorithm is solved by processing the gradient image of the fragment perforation image, but there are still problems that cannot be universally applied and poor stability.

With the popularity of artificial intelligence algorithms [21], [22], [23], because of its outstanding performance in image classification, target detection and semantic segmentation, it has overcome the problems of poor generalization ability and low efficiency of traditional image segmentation algorithms. Therefore, the research on target detection of fragment target images based on deep learning is gradually increasing. He et al. [24] proposed SPPNet, which adds a spatial pyramid pool (SPP) layer between convolution layer and full connection layer, it is used to solve the problem of feature extraction in CNN (Convolutional Neural Network, CNN), which requires the input image size to be fixed, and the detection speed is 38 times that of R-CNN. Girshick et al. proposed Fast-RCNN and Faster-RCNN, which improved the detection accuracy by 11.5% and 18.8%, respectively [25], [26]. Liu et al. [27] proposed Swin Transformer method, which used self-attention, combined hierarchical and shifted windows to reduce complex of target detection.

For small target segmentation, because it contains too little information, so it contains too few discriminant features, and the number of small target sample is too few in the data set. which makes the trained network model has poor detection performance for the small target detection. In recent years, various methods have been proposed to solve the problem of small target detection, it can be divided into five categories: multi-scale feature learning, generative adversarial learning, context learning, data set enhancement and weight redistribution methods. Multi-scale feature learning is mainly represented by the FPN (Feature Pyramid Networks, FPN) detection algorithm proposed by Tsungyi Lin et al, which has a good detection effect for small targets [28], the main idea is to combine low-level spatial information and high-level semantic information to enhance small target features. Reference [29], [30], and [31] are the representative methods of generative adversarial learning. The main idea is to obtain high-resolution images or features of small targets based on sub-map or geo-resolution features of targets through GAN adversarial network. The main methods of context learning are reference [32], [33], and [34]. The main idea of context learning is to use the environmental information of small targets or the relationship between small objects and other easily detected objects to assist the detection of small targets. The main representative methods of data set enhancement is reference [35], the main idea is to solve the problem of small target image and small target instance imbalance in data set, and increase the number of small targets. The representative methods of weight redistribution is reference [36] and [37], which makes the network pay more attention to the training of targets by increasing the weight in the calculation of the loss function, this allows the network to treat targets of different scales more equally.

Especially for small fragment perforation target segmentation, Lei et al. [38] realizes the recognition of fragment targets through the joint training of region generation network (RPN) and Fast R-CNN, but the detection speed is slow, and the utilization rate of shallow features is not high; Xu [39] proposed a method based on improved Faster-RCNN fragment parameter measurement. By improving the feature extraction structure and increasing the residual module, the utilization rate of shallow features is improved, but there is a problem that small targets cannot be accurately detected. Wei [40] proposed to modify the SSD backbone network and introduce attention mechanism to improve the detection rate of small fragments.

The methods of image distortion correction and targets segmentation are actively explored in the world, but the methods of image distortion correction do not take into account the image distortion caused by shooting angle and targets distortion together, moreover the small target segmentation method still has the problem of missing detection to the small target. Therefore how to use a single two-dimensional image to correct the image distortion caused by shooting angle and target deformation together quickly and effectively, and how to reliably detect small targets is an urgent problem to be solved. Based on the above discussion, the image of the target region is meshed and planed according to the small deformation degree of the target and the approximate plane in the small region, a simple inverse perspective mapping model is established to match the vertices of the planar region images with the vertices of the real region, which is solved by SVD decomposition. For the segmentation of small target with fragment damage region, based on the encoding and decoding network the attention mechanism is adopted, Swin Transformer is introduced into the coding part to extract multi-scale features, and based on RFE receptive field expansion module the multi-scale features are fused, then, a multi-scale feature extraction and fusion Encoder were designed to extract multi-scale targets features rapidly. At the same time, the attention mechanism of ECA-Net was introduced into the decoder, the attention-enhanced decoder is designed to enhance the ability of distinguishing small target from background and segment small target image.

The following outlines the main contributions of our research to highlight the importance of our work:

- An image correction method combining mesh planarization and homomorphic inverse perspective transform model is proposed to realize image distortion caused by shooting angle and target physical deformation, and to optimize the input image quality.
- In order to solve the problem that small fragment perforation are easily missed in the image segmentation, a double self-attention mechanism embedded fusion (DSA-EF)semantic segmentation network based encode decode network is designed.
- To extract multi-scale features and improve the features extraction efficiency, Swin Transformer is introduced into the encoding part of the DSA-EF semantic segmentation network as the backbone network, and the RFE receptive field expansion module is combined to fuse multi-scale features, a multi-scale feature extraction and fusion Encoder is designed to extract the features of multi-scale targets quickly and improve the accuracy and stability of small target detection.

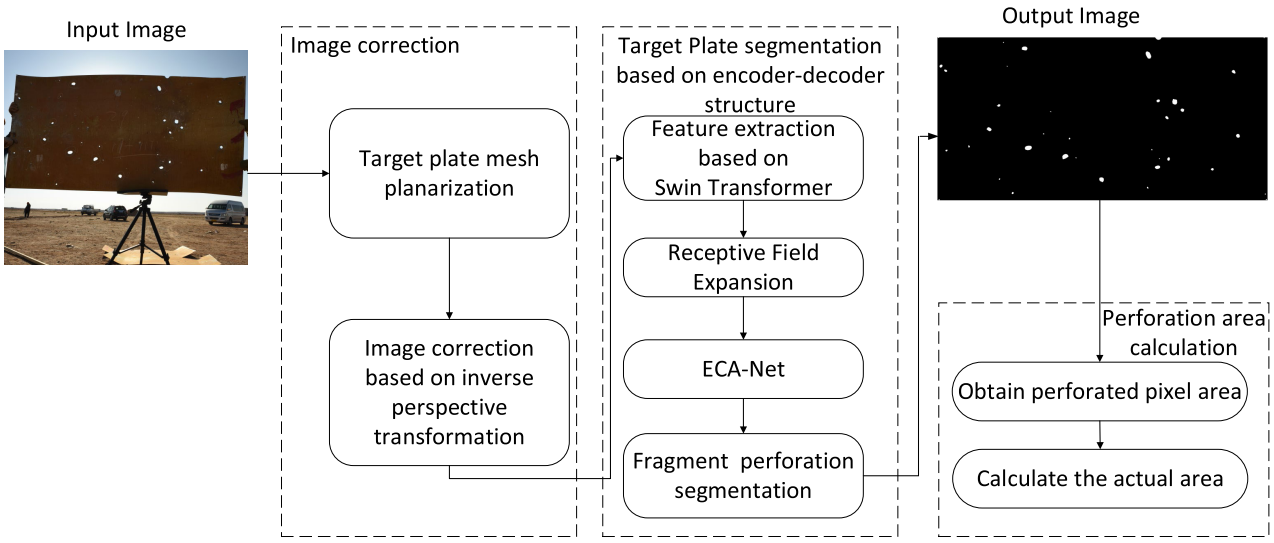


FIGURE 1. Fragment perforation of target plate area calculation algorithm.

• In order to improve the attention of small targets, the efficient channel attention mechanism is added to enhance the feature information of small fragments in the feature decoding stage, enhance the attention to the small target features and retain shallow features.

The paper includes the following outline. Section II introduces the related work in this field at home and abroad, Section III introduces the method. Section IV introduces the implementation details of image correction. Section V introduces the implementation details of the segmentation algorithm. Section VI introduces the results. The conclusions of this study are shown in Section VII.

III. TARGET PLATE DAMAGE IMAGE ANALYSIS AND FRAGMENT PERFORATION SEGMENTATION SCHEME DESIGN

The damage effect of warhead explosion is generally measured by the damage of equivalent target plate. Shock wave and many high-speed fragments will be formed after warhead explosion, and the target plate will be affected by these two types of damage elements. This paper mainly studies the damage effect of fragments attacking equivalent target plate after warhead explosion. The damage of the equivalent target plate by the fragments generated after the warhead explosion is mainly determined by the fragment breakdown velocity, the specific kinetic energy of the fragment and the effective fragment density when the fragment reaches the equivalent target plate. The input image shown in Fig. 1 is the equivalent target damage image taken at the scene after the test.

It can be seen from Fig. 1 that because the image of the target plate is affected by the shooting angle, and there is a physical deformation caused by the explosion impact on the target surface, that is, the target magnification at different positions is different, these factors seriously affect the accuracy of the area measurement of the target plate damage area. Therefore, it is necessary to process the target plate image for enhancement, correct the target plate image, eliminate the background interference in the image, and restore the true size of the fragment perforation. Because the fragment perforation in the image belongs to the small size target, and the fragment perforation size is different, this puts forward the requirement of the segmentation algorithm, which requires both the recognition of the perforation region and the accurate segmentation the edge of the fragment perforation position. Therefore, the deep learning algorithm is used to extract the characteristics of the fragment perforation at different scales, and the fragment perforation region is identified and segmented. Then the relationship between the number of imaging pixels of the target plate and the physical parameters of the actual target plate is calculated, and the area of the damage area is calculated.

IV. IMAGE CORRECTION OF TARGET PLATE UNDER COMPLEX CONDITIONS

Due to the impact of the warhead explosion in the actual test, the metal target plate will produce physical deformation [41]. So there is an impact on the fragment perforation measurement due to the deformation of the target plate in the captured target plate image, and because of the inevitable existence of artificial tilt angle in the shooting process, there is a certain geometry deformation in the captured target plate image. Therefore, it is necessary to correct the target plate image to eliminate the influence of target plate deformation, as shown in Fig.2. Based on the idea of mesh planarization and the method of homomorphic inverse perspective transformation between the planes, this paper proposes a correction method. By dividing the target image into finite parts for processing, each sub-region becomes a simple part. Because the deformation of the target plate is mainly reflected in the lateral deformation, this paper divides the target plate image horizontally and equally into M of mesh regions. When M is large enough, the whole target plate image region can be

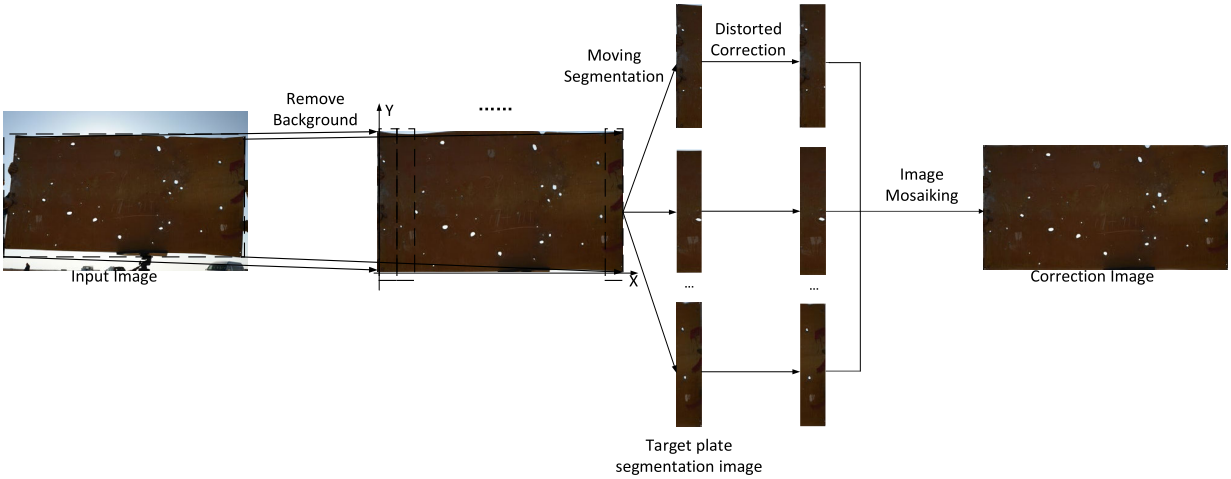


FIGURE 2. Image correction process.

approximately divided into multiple approximate mesh plane regions. For the mesh plane of target plate image decomposition, combined with the corresponding decomposition of physical target plane, homomorphic perspective projection transformation method is used to correct the target plate image by mesh region. The essence of inverse perspective transformation is to project the image to a new visual plane, that is, to use the three-dimensional space coordinate transformation to map the original photographic image into a standard image to realize the correction of the image.

The coordinate system is established with the lower left of the physical target plate and the target plate image as the origin, then the physical target plate homogeneous coordinate is set as $A = (X, Y, 1)$, and the pixel homogeneous coordinates of the target plate image is set as $A' = (X', Y', 1)$.The perspective transformation between the physical target plate and the target plate image with homomorphic matrix $a_{3\times3}$ is as following [42]:

$$(X, Y, 1) \cong (X', Y', 1) \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{1}$$

Matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ defines rotation transformation and scaling transformation. $\begin{pmatrix} a_{13} & a_{23} \end{pmatrix}^T$ defines the translation vector, $\begin{pmatrix} a_{31} & a_{32} \end{pmatrix}$ defines the projection vector.

So, there is:

$$\begin{cases} X = \dfrac{a_{11}X' + a_{21}Y' + a_{31}}{a_{13}X' + a_{23}Y' + a_{33}} \\ Y = \dfrac{a_{12}X' + a_{22}Y' + a_{32}}{a_{13}X' + a_{23}Y' + a_{33}} \end{cases} \tag{2}$$

It can be seen from (2) that there are 9 unknowns in the homomorphic matrix $a_{3\times3}$, but since it's in a homogeneous coordinate system, the homomorphic matrix $a_{3\times3}$ has only 8 degrees of freedom, so let $a_{33} = 1$, then eight equations are needed to solve the parameters in the perspective projection transformation matrix $a_{3\times3}$ simultaneously. One pair of matching points determines two equations, so four pairs of matching points are required.

Since the physical target plate is a regular rectangle of known size, the size of the mesh plane of each physical target plate can be calculated when the number of regions in the mesh planarization is determined. So take four vertices of each physical target grid plane as feature matching points. The pixel size of target plate image can also be measured by image segmentation. The size of each grid-planarization regions can also be calculated when the number of it is determined. Suppose that for the matched target plate image mesh plane and the physical target mesh plane, L, W are the length and width of the mesh plane in the target plate image, and L_{real}, W_{real} are the length and width of the mesh plane in the real target plate. Then the four corner vertices of the image target plate mesh plane are $D'_0(0, 0, 1), D'_1(L, 0, 1), D'_2(0, W, 1), D'_3(L, W, 1)$, and the four corner vertices of the actual target plate mesh plane are $D_0(0, 0, 1), D_1(L_{real}, 0, 1), D_2(0, W_{real}, 1), D_3(L_{real}, W_{real}, 1)$.

The four corresponding coordinate point pairs before and after the homomorphic inverse perspective transformation can be substituted into the equation (2), A set of 8 equations is set up, which can be solved by SVD decomposition to get transform matrix $a_{3\times3}$ of planar homomorphic inverse projection of each target plate image grid. The corrected target plate image can be obtained by transforming the grid plane of each target plate image according to the calculated homomorphic inverse projection transformation matrix $a_{3\times3}$ according to equation (1).

V. FRAGMENT PERFORATION OF TARGET PLATE REGION SEGMENTATION ALGORITHM DESIGN

It is an important part of measurement to identify and segment the fragment perforation region from the corrected target image by using the algorithm. However, different from the general target segmentation requirements, the image of the

target plate fragment perforation region mainly has the following problems: First, the region of fragment damage is a small object category compared to the entire target plate region; the size of the fragment perforation region is different, which requires the segmentation algorithm to have better segmentation performance, so as to effectively retain the edge features of the fragment perforation, find and segment the fragment perforations of different sizes in the whole image; there are many false objects in the image, such as the traces on the target plate itself, the marks, etc., which will interfere with the recognition process of the network.

For the above problems, in this paper, a semantic segmentation method DSA-EF Net for target plate image damage fragment perforation based on multi-scale feature fusion is proposed. The algorithm is based on the network structure of encoder-decoder [21], [22], [23]. In order to ensure the accuracy of small fragment perforation feature extraction, Swin Transformer (SA) is added to the encoder part to extract image features. It can not only effectively retain the shallow features of the target plate image fragment perforation region, but also extract the deep features well, so that the encoder network can extract the multiscale fragment perforation region features in the target plate image more reliably and quickly. Moreover, the Receptive Field Expansion (RFE) module is added, by introducing dilated convolution to realize the fusion of multi-scale fragment perforation features, so that the network can obtain more semantic information to improve the generalization ability and optimize the performance of feature extraction network. In addition, in the decoder part, the Efficient Channel Attention (ECA) mechanism is adopted to retain the shallow features of the image to the greatest extent and ensure the efficiency of small target recognition. The DSA-EF net is shown in Fig. 3:

A. MULTI-LEVEL FEATURE EXTRACTION ENCODER

In order to obtain more shallow features from the target plate image, the Swin Transformer model [43] is used to extraction a hierarchical design. First, the target plate image to be processed is input into the marker segmentation layer, and it is divided into non-overlapping parts of equal size. The feature extraction process of the Swin Transformer neural network is divided into two stages. In the first stage, the linear embedding layer is used to project these segmented original features into the specified dimension to form a subgraph, and then put it into the Swin Transformer module to extract features, which can obtain the shallow features of the target plate image. In the second stage, the feature image is fused by down-sampling to reduce the image resolution and reduce the number of channels. Then the Swin Transformer module is input again to extract features, and the second stage operation is repeated continuously. In this way, the target plate image is continuously reduced in dimension and deeper features are obtained. The Swin Transformer module structure is shown in Fig. 4.

The W-MSA (Windows Multi-head Self Attention) module introduces the window mechanism, which reduces the amount of calculation; the SW-MSA (Shifted Window-Multi-head Self Attention) module introduces cross-window links, which solves the problem that feature interaction cannot be carried out across windows. The calculation process of Swin Transformer block is shown in (3).

$$\begin{cases} \widehat{D}_l = \text{W-MSA}[\text{LN}(D_{l-1})] + D_{l-1} \\ D_l = \text{MLP}[\text{LN}(\widehat{D}_l)] + \widehat{D}_l \\ \widehat{D}_{l+1} = \text{SW-MSA}[\text{LN}(D_l)] + D_l \\ D_{l+1} = \text{MLP}[\text{LN}(\widehat{D}_{l+1})] + \widehat{D}_{l+1} \end{cases} \tag{3}$$

D_l and D_{l+1} are different levels of output; MLP is the multi-layer perceptron; LN is the normalization operation; W-MSA is the window attention mechanism; W-MSA is the moving window attention mechanism.

B. RECEPTIVE FIELD EXPANSION MODULE

In order to enhance the ability of the network to extract local correlation and structural information in the target image, this paper proposes to add a receptive field expansion module generation that combines dilated convolution [44] and residual connection in the feature extraction process. The structure of receptive field expansion (RFE) module is shown in Fig. 5.

Due to the particularity of Swin Transformer feature data, the data format needs to be processed first, and then the data is sent to parallel dilated convolution model. The parallel dilated convolution designed in this paper consists of three 3*3 dilated convolution layers with different dilated rates. This module allows the features extracted by each dilated convolution with different sampling rates to be processed in a separate branch, and finally fuses the features on different branches to generate results. The design of different expansion rate modules can not only construct convolution kernels with different receptive fields to obtain multi-scale object information, but also avoid gridding effect. Finally, the activated data is used as output. In this way, the module can obtain image features of different scales and fuse them together to help the model better understand the target.

After the initial 3*3 convolution kernel is expanded, the receptive field is $\left[\,3*3\ 5*5\ 7*7\,\right]$. Using convolution kernels with multiple expansion coefficients on the same feature for feature extraction can obtain different scale information components, which is conducive to enhancing feature richness and accuracy of multi-scale feature extraction of images. At the same time, each branch shares weights and decrease the number of network parameters. Finally, the three dilated convolution branches are merged to perform channel splicing and dimensionality reduction on the extracted features to prevent information redundancy.

C. ATTENTION FUSION DECODER

The attention mechanism plays a great role in extracting data features, which makes the model more sensitive to target

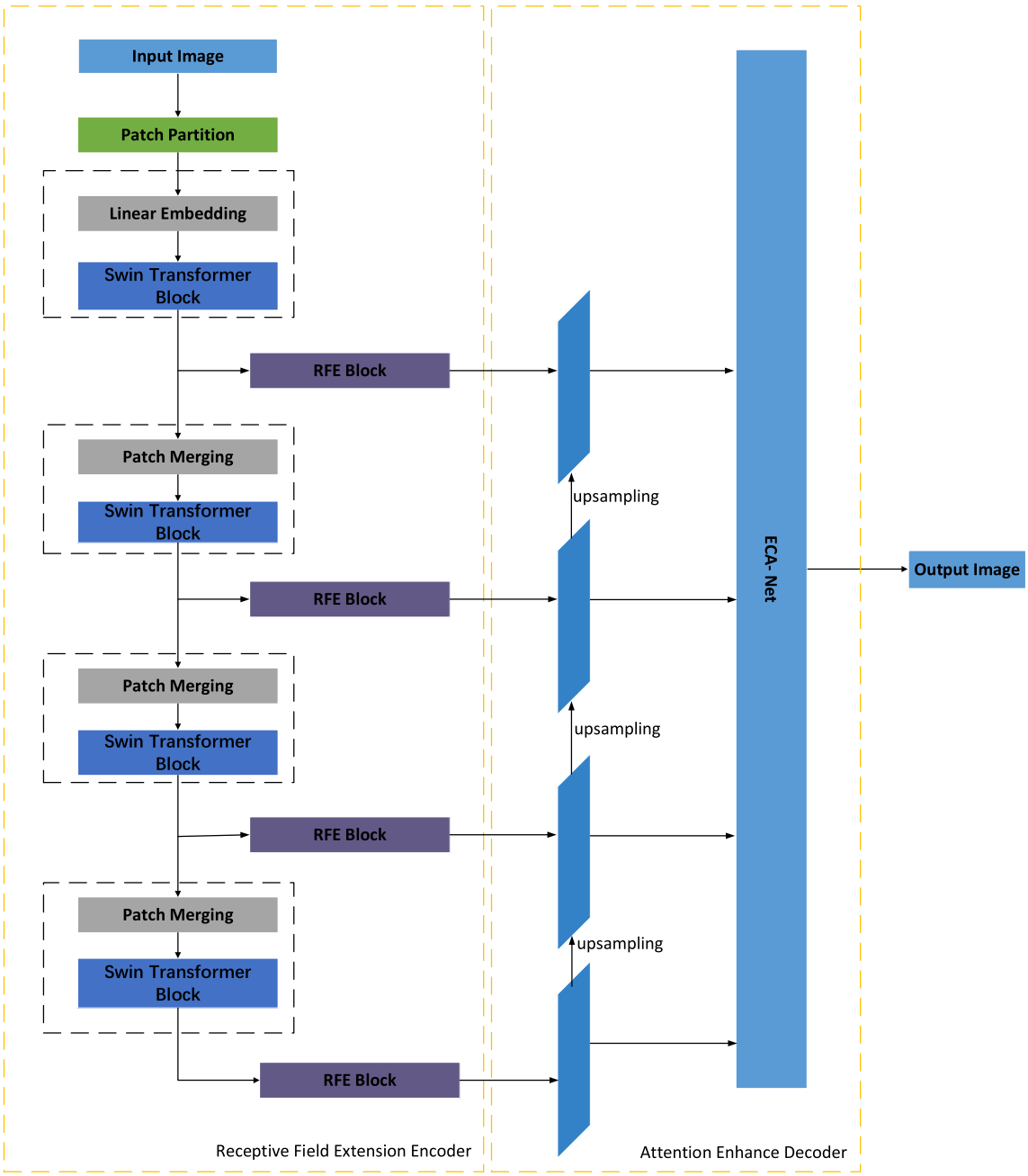


FIGURE 3. DSA-EF Net model.

information when processing target images, so it is easier to retain shallow features and improve model performance. In this paper, the ECA mechanism is introduced at the decoder, and the global adaptive attention weight calculation is introduced into the channel dimension of the four different levels of target feature images output by the network. By increasing the weight of the shallow feature map, the network 's ability to distinguish small targets is increased.

The most used attention mechanism is the squeeze and incentive network (SE-Net), and the calculation of SE-Net is more complicated and requires more computing resources. Because SE-Net uses dimensionality reduction operations, it affects the learning effect. Therefore, this paper uses the efficient channel attention network (ECA-Net) [45], [46] as an attention module, which can significantly reduce the complexity of the model and avoid the impact on the learning effect.

ECA-Net is an improvement of SE-Net. It proposes a non-dimensional local cross-channel interaction strategy, and its cross-channel dimension interaction is realized by one-dimensional convolution. This can reduce the computational complexity while ensuring that the performance does not decrease. It uses an average pooling layer instead of the full connection layer in the original SE module to realize the

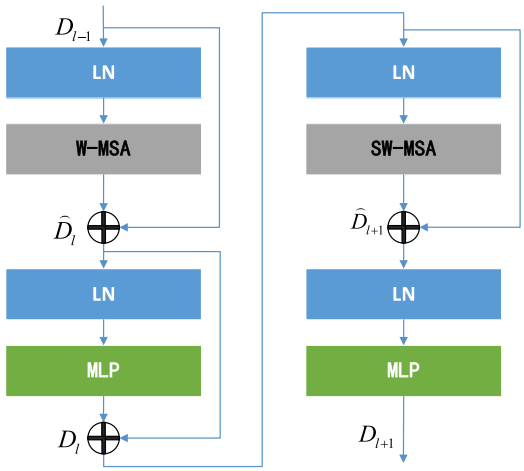


FIGURE 4. Swim Transformer module structure.

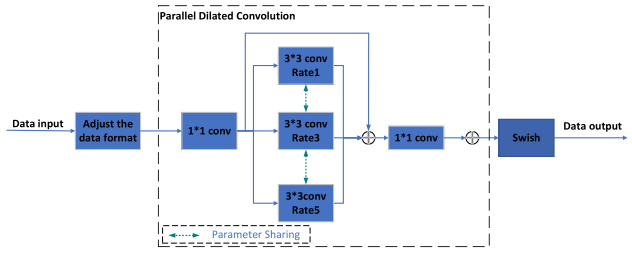


FIGURE 5. Receptive field expansion module structure.

extraction and mapping of input features. The process of ECA-Net is shown in Fig. 6:

In this paper, the target plate feature image with a length of L and a width of W is transmitted through C channels of the encoder network to compress it in the spatial dimension to obtain T, as shown in (4):

$$T_{avg} = \mathrm{GAP}(T) = \frac{1}{H \times W} \sum_{i=1}^{W} \sum_{j=1}^{H} T_{ij} \tag{4}$$

Among them: $T_{ij} \in R^{(C)}$ is the full channel feature of T at (i, j), and then after the global average pooling of the acquired features, all the features of the target image of size 1*1 are obtained.

Then, the local interaction between K adjacent channels is realized by using one-dimensional convolution of size K. The size of parameter K can be calculated according to the number of input channels, as shown in (5):

$$K = \frac{\log_2 C + 1}{2} \tag{5}$$

After calculation, this paper uses one-dimensional convolution with a size of 2 to achieve local interaction between adjacent channels. After convolution, the feature map is obtained. Secondly, the Sigmoid function is used to activate it to obtain the weight coefficient.

$$\rho = \sigma(f^K(T^c_{avg})) \tag{6}$$

Here, σ is the Sigmoid activation function, and f^K is a one-dimensional convolution operation of size K. By multiplying the weight coefficient with the input feature element T one by one, a new output feature F' is obtained.

VI. EXPERIMENT AND RESULT ANALYSIS

A. DATASET CONSTRUCTION

In this paper, 100 target plate images of different sizes taken from different directions and different environments are selected as the data set, and 15 images of data set are randomly selected as the verification set. The original image resolution is too large to be directly input into the model. Therefore, the data set is cut into 512 × 512pixel size images. Finally, a total of 1731 training sets, 515 validation sets, and 500 test sets are generated.

The following principles are followed when constructing the dataset:

- The data set created should reflect the state of the target plate after the experiment in the actual environment;
- The data set should be able to reflect the real environment at the time of shooting;
- The images in the data set should be suitable for labeling to improve the accuracy of the trained model;
- The constructed data set should be applicable to the current computing equipment and meet the requirements of semantic segmentation methods for data quantity and quality.

Firstly, this paper uses the open-source data annotation tool Labelme to label the target plate images, which mainly includes two categories. The target surface is classified into complex background except perforation. Then, the corresponding single-channel grayscale label is generated according to the file; in order to reduce the burden of hardware calculation, this paper cuts the image samples and grayscale la-bels into pixel-sized image blocks. The data and labels after cutting are enhanced to meet the data requirements of network training and improve the quality of training. Finally, the proportion is divided into 4528 training sets, 565 validation sets, and 550 test sets. In addition, the data set is mainly used to train the semantic segmentation method DSA-EF, so that the trained model can accurately extract and segment the perforation from it in practical application, and improve the accuracy and efficiency.

B. EXPERIMENTAL ENVIRONMENT AND PARAMETER CONFIGURATION

Windows11 operating system, NVIDIA GeForce RTX 4060 Laptop GPU, i5-11400H CPU, i5-13500HX, 16GB memory; the software used is Matlab R2022a, Python3.9, and the deep learning framework used is Tensorflow2.6.

The algorithm in this paper uses the stochastic gradient descent method with momentum, the momentum is set to 0.9, and the weight attenuation rate is 0.0005. Learning rate is set to 0.01. Decay rate is set to 0.96. Global step is set to 20000. Decay rate is set to 100.

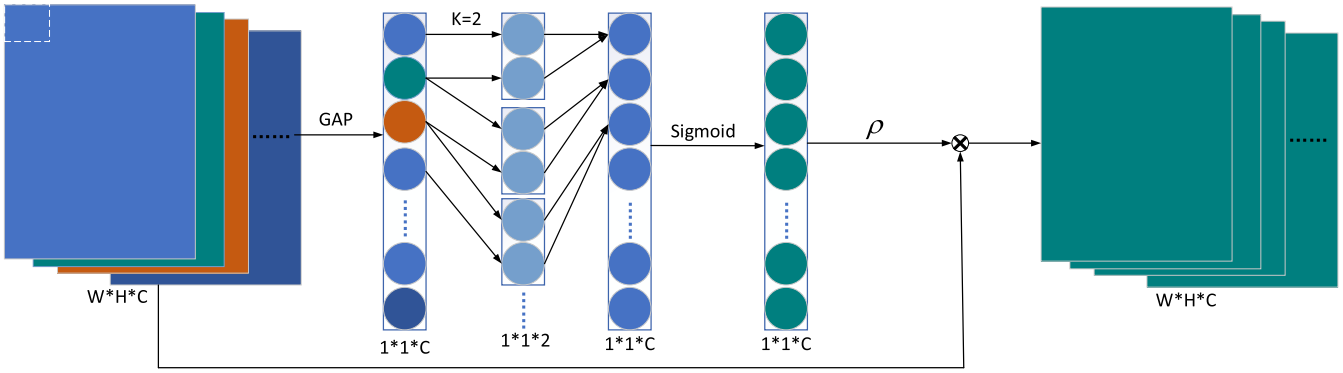
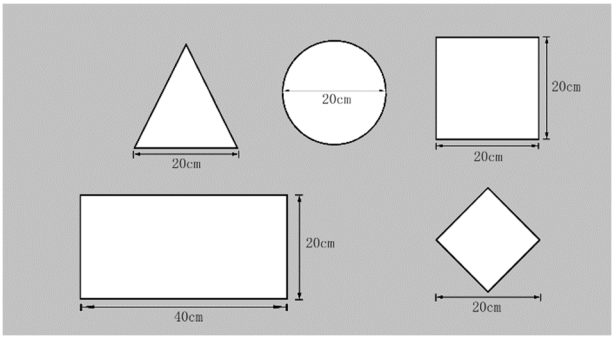
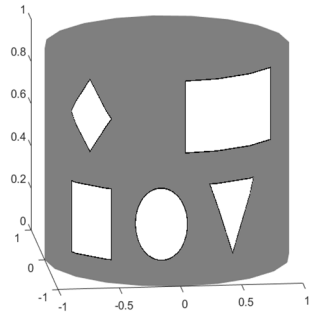


FIGURE 6. ECA-Net.



(a) Standard target plate simulation



(b) Bending target plate simulation

FIGURE 7. Object of target plate measurement simulation.

TABLE 1. Accuracy analysis of simulation measurement results of standard target plate.

No.	Geometry	Actual area (cm^2)	DSA-EF calculation (cm^2)	Error rate (%)
1	triangle	173.21	167.61	3.23
2	roundness	314.16	300.62	4.31
3	square	400	397.79	0.56
4	rectangle	800	787.55	1.55
5	rhombus	200	195.5	2.23

TABLE 2. Results of ablation experiments.

Encoder Network	Decoder Network	ISPRS Vaihingen				Real Dataset
		IoU%			MIoU/%	MIoU/%
		Impervious Surface	Building	Low Vegetation		
Swin Transformer	Basic Fusion Network	82.97	79.98	83.40	83.95	84.01
Swin Transformer + RFE	Basic Fusion Network	85.57	82.13	86.42	86.11	86.32
Swin Transformer	ECA-NET	83.29	82.40	86.87	85.94	85.76
Swin Transformer + RFE	ECA-NET	86.98	84.58	88.65	88.29	87.95

C. VALIDATION EXPERIMENT

Because the flight angle and velocity of fragments generated by warhead explosion are random, the fragment perforation area on large metal target plate is often irregularly distributed, and the perforation is mostly irregular in shape, and the perforation edge is mostly burred and not smooth. Obviously, the accuracy verification of the measurement results in this paper cannot be measured manually. Therefore, this paper proposes to design a standard target plate as shown in Fig. 7 (a), draw a standard-sized hole on it, and project it to a cylindrical surface as shown in Fig. 7 (b), so as to realize the simulation of the tilt and deformation of the target plate, and then calculate the original image area. The projected image area is used as the theoretical value, and then compared with the measurement results in this paper to verify the measurement accuracy of this method.

TABLE 3. The comparison results of different methods for target plate data set.

Method category	Algorithm	mPA/%	Dice/%	Accuracy/%	MIoU/%	JAC/%
CNN-base	FCN	82.97	79.98	92.40	83.95	84.01
	DeepLabV3+	85.57	82.13	95.42	86.11	86.32
Transformer-base	SETR	83.29	82.40	95.87	85.94	85.76
	Segmenter	86.98	84.58	97.65	88.29	87.95
	Seg-Former	86.91	83.33	97.24	87.96	87.60

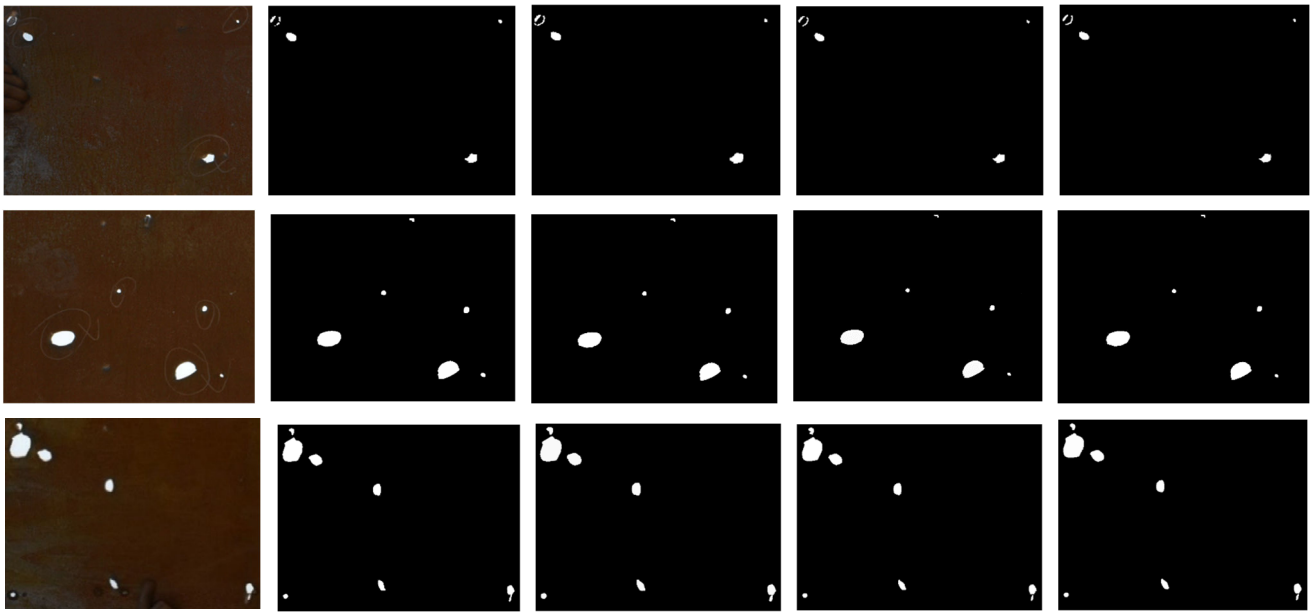


FIGURE 8. Partial visualization results of different methods for real datasets.

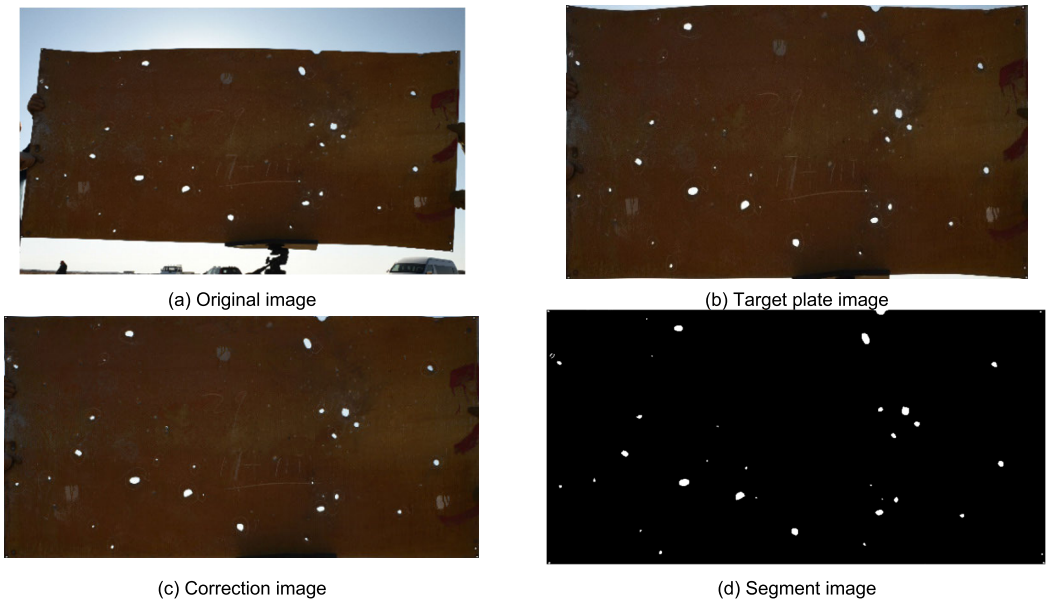


FIGURE 9. Sequence image of actual target plate experimental process.

D. ABLATION EXPERIMENT

In order to verify the effectiveness of this method, this paper has carried out ablation experiments. Therefore, this paper has conducted experimental verification on high-resolution remote sensing image dataset ISPRS Vaihingen and a measured data set. Compare the influence of RFE module and

ECA-Net module on the experimental results, and analyze the performance of each part. In the experiment, choose Swin Transformer as the backbone network. In order to verify the performance of the proposed network, RFE module and ECA-NET are inserted into the typical target detection network framework for comparative experiments.

From the table, it can be seen that after adding the RFE module to the basic Swin Transformer model, the MIoU on the two data sets is increased by 1.86 % and 2.13 % compared with that before. After adding ECA-Net in the decoder part, the MIoU on the two data sets is increased by 2.28 % and 2.19 % compared with that before. The MIoU of the proposed algorithm DSA-EF Net on ISPRS Vaihingen and private data sets is 88.06 % and 85.03 % respectively. Compared with the original Swin Transformer network MIoU, it increased by 4.44 % and 3.71 % respectively, and was superior to other models.

E. CONTRAST EXPERIMENT

This paper compares the algorithm with the current implemented and representative mainstream segmentation methods, including classical segmentation methods based on DCNN such as FCN, DeepLabV3+, and segmentation methods based on Transformer structure such as SETR, Segmenter, Seg-Former. The parameters and environment configured during the experiment are consistent. The results are shown in Table 3.

It can be seen from the table that the mPA, Dice, accuracy, MIoU, and Jaccard rates corresponding to the method in this paper are 3.15 %, 3.46 %, 3.36 %, 2.75 %, and 2.84 % higher than DeepLabV3 with the best performance in the DCNN-based method, and 1.81 %, 2.26 %, 1.58 %, 1.51 %, and 1.56 % higher than the SegFormer method with the best performance in the Transformer-based method. Because the convolution operation has translation and other degeneration, CNN is better at capturing large and regular target information, so the effect in this data set is not ideal. The method in this paper is better in the recognition of target categories with smaller scales and irregular shapes. The analysis is as follows: the pooling operation of DCNN will lead to the gradual disappearance of the target features in the deepening of the network, while Swim Transformer uses automatic learning to perform window feature mapping, which will not cause the loss of information; in this paper, the RFE module is introduced into the encoder to assist the network in feature extraction and fusion; Finally, for the design of the decoder, this method designs a multi-scale feature fusion decoder based on ECA-Net, which integrates the features of the target in all output layers of the network and can effectively supplement the detailed information of the target. Therefore, the proposed method can better extract and utilize the features of the target, and the model segmentation have better performance.

This paper will conduct visual analysis on the mainstream segmentation methods based on the Transformer structure and our method on some test sets, as shown in Figure 8.

TABLE 4. Measurement results of fragment perforation area and perimeter of target plate.

No.	Area (mm^2)	Perimeter (mm)
1	247.62	56.10
2	188.44	49.34
3	98.79	35.98
4	92.42	35.84
5	470.91	79.10
6	212.84	57.42
7	26.54	17.51
8	21.58	15.30
9	54.44	26.05
10	116.86	37.49
11	640.08	91.66
12	899.99	109.41
13	38.64	21.09
14	70.39	28.48
15	801.87	105.48
16	604.37	88.53
17	1008.55	118.78
18	267.61	57.83
19	785.96	103.02
20	329.10	64.30
21	262.30	58.27
22	40.12	21.22
23	23.58	15.89
24	506.03	81.76
25	247.94	55.66
26	72.94	30.87
27	342.83	67.51
28	401.07	70.21
29	184.96	48.12

Figure 9 is the image of the metal target plate taken after the actual experiment and the effect group diagram after the preprocessing stage. Fig. 9 (a) is the actual shooting picture. It can be seen that the target plate in the picture is not horizontal, and there are many interference backgrounds. After the spatial inverse perspective transformation, the geometric deformation caused by the angle during shooting is corrected. As shown in Fig. 9 (b), it not only removes the redundant background, but also minimizes the impact of geometric deformation. Through the finite element moving transformation method, the deformation correction of the target plate is carried out, as shown in Fig. 9 (c), which effectively suppresses the influence of the target plate deformation on the measurement of the perforation area of the target plate image, and effectively deals with the background that cannot be completely eliminated during the tilt correction. The filtering

not only maintains the sharpness of the perforation edge, but also minimizes the influence of the metal texture and stain noise of the target plate. Fig. 9 (d) is the image segmented by the algorithm. It can be seen that the recognition effect is good, the perforation position can be effectively identified, and the segmentation edge is clear.

After the target fragment perforation area is segmented, it is necessary to accurately measure and calculate the number of pixels in each fragment perforation area and the physical unit area of a single pixel. The breakdown area of the metal target plate damaged by fragments is mostly random in shape and irregular holes on the edges. Therefore, this paper uses pixel traversal to measure the area of the irregular area of the target plate damage image. The traversal method is to traverse and extract the pixels of all damaged areas, and the total number of pixels P_{sum} can be obtained by summing the number of pixels identified in each damaged area of the target image. Then, the average physical area of a single pixel is calculated from the physical area of the target plate and the image resolution, so as to calculate the physical actual area S of each damaged area, as shown in (7), where L and W represent the length and width of the corrected target plate in the image, respectively, and L_{real} and W_{real} represent the length and width of the actual target plate.

$$S = P_{sum} \times \frac{L \times W}{L_{real} \times W_{real}} \tag{7}$$

In the actual experiment, the size of the target plate is. The final measurement results are shown in Table 3:

VII. CONCLUSION

In order to solve the problem that the metal target image has superimposed distortion and the perforation of the fragment target image is difficult to measure accurately, this paper designs a set of metal target damage image recognition process combining image processing and deep learning. Aiming at the problem of target plate image tilt, the tilt correction of target image based on inverse perspective transformation is proposed. Aiming at the problem of physical deformation of the target plate caused by explosion shock, proposed a correction method based on mesh planarization and homomorphic inverse projection transformation, and realized the correction of the target plate deformation. In order to better identify small targets, proposed an algorithm of DSA-EF semantic segmentation net for target fragment perforation region segmentation and recognition based on deep learning. In the algorithm, to improve the feature extraction network, the receptive field expansion module and the efficient attention mechanism are added, and the shallow features of the image are fully utilized to achieve the purpose of fragment small target recognition. Finally, the paper designs validation experiments to demonstrate the effectiveness of our method. Meanwhile, through designs the ablation experiment to verify the effectiveness of this method, and designed the comparative experiment to comparing with existing methods. The analysis proves the feasibility and accuracy of the method proposed in this paper in the identification and measurement of perforation in the damage area of metal fragment target. The actual test results show that the measurement accuracy of this method meets the actual needs.

Since the receptive field expansion module proposed in this paper contains a certain amount of convolution operations, the calculation amount and parameter amount of the model are slightly larger. How to effectively reduce the complexity and parameter amount of the model under the premise of ensuring the segmentation performance is the next research direction.

REFERENCES

[1] B. S. Elveli, O. Vestrum, K. O. Hauge, T. Berstad, T. Børvik, and V. Aune, "Thin steel plates exposed to combined ballistic impact and partially confined airblast loading," *Eng. Failure Anal.*, vol. 144, Feb. 2023, Art. no. 106943.
[2] P. N. Verma and K. D. Dhote, "Characterising primary fragment in debris cloud formed by hypervelocity impact of spherical stainless steel projectile on thin steel plate," *Int. J. Impact Eng.*, vol. 120, pp. 118–125, Oct. 2018.
[3] F. Xu, "Research on key technologies of damage power field testing system software," M.S. thesis, Dept. Autom., North Univ. China, Shanxi, China, 2013.
[4] J. H. Wu and J. Liu, "Development status of fragment parameter test technology in warhead static explosion field," *J. Ordnance Equip. Eng.*, vol. 40, no. 10, pp. 104–110, 2019.
[5] J. J. Zhou, D. R. Kong, and C. D. Xu, "Integrated fragment velocity and dispersion characteristics measurement system," *Electron. Meas. Technol.*, vol. 45, no. 12, pp. 133–140, 2022.
[6] J. P. Wu, M. J. Qiao, and Z. G. Yan, "An overview of the development of warhead fragment field parameter test technology," *J. Ordnance Equip. Eng.*, vol. 40, no. 5, pp. 105–109, 2019.
[7] M. Nixon and A. Aguado, *Feature Extraction and Image Processing for Computer Vision*. New York, NY, USA: Academic, 2019.
[8] J. Gao and X. Zhang, "Fragment perforation spatial localization measurement method and calculation analysis by using photogrammetry technology," *IEEE Access*, vol. 11, pp. 102092–102102, 2023.
[9] Y. Ren, "Multi-target impact coordinate test based on dual linear array image analysis," M.S. thesis, Dept. Weapon Ind. Mil. Technol., Xi 'an University Technol., Xi'an, China, 2019.
[10] E. Watson, N. Kunert, R. Putzar, H.-G. Maas, and S. Hiermaier, "Four-view split-image fragment tracking in hypervelocity impact experiments," *Int. J. Impact Eng.*, vol. 135, Jan. 2020, Art. no. 103405.
[11] J. Sequeira, "3D reconstruction of naturally fragmenting warhead fragments," Ph.D. dissertation, Fac. Sci., Stellenbosch Univ., South Africa, 2023.
[12] P. Y. Hu, "Warhead fragments motion trajectories tracking and spatio-temporal distribution reconstruction method based on high-speed stereo photography," in *Defence Technology*. Amsterdam, The Netherlands: Elsevier, 2024.
[13] Y. J. Zhang, "Camera calibration," in *3-D Computer Vision: Principles, Algorithms and Applications*. Singapore: Springer Nature, 2023, pp. 37–65.
[14] Y. Pi, M. Wang, B. Yang, and Z. Gao, "Robust camera distortion calibration via unified RPC model for optical remote sensing satellites," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5627815.
[15] M. He, H. Wang, and Y. Huang, "Research on field metal fragment target measurement based on image processing," *J. Ballistics*, vol. 33, no. 4, pp. 83–90, 2021.
[16] J. W. Wang, Y. C. Zhu, S. Wang, and K. Li, "Research on distortion correction method of aerial photovoltaic module images," *Hunan Electric Power*, vol. 43, no. 4, pp. 74–79, 2023.
[17] W. Y. Zhang, J. Q. Zhou, F. S. Sun, and X. Yu, "Research on geometric expansion of shell plate of surface ship based on graph theory and intelligent optimization method," *Shipbuilding China*, vol. 63, no. 6, pp. 255–263, 2022.
[18] Z. Lyu, G. P. T. Choi, and L. M. Lui, "Bijective density-equalizing quasi-conformal map for multiply connected open surfaces," *SIAM J. Imag. Sci.*, vol. 17, no. 1, pp. 706–755, Mar. 2024.

[19] M. F. Li, C. He, and Z. Y. Ling, "The application of image recognition technology in the analysis of warhead static explosion test," *Sichuan Ordnance J.*, vol. 35, no. 6, pp. 101–104, 2014.
[20] Y. Song, "Research on target detection method of fragment impact image," M.S. thesis, Dept. Weapon Ind. Mil. Technol., Suzhou Univ. Sci. Technol., Jiangsu, China,2022.
[21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[22] Y. Ji, H. Zhang, Z. Zhang, and M. Liu, "CNN-based encoder–decoder networks for salient object detection: A comprehensive review and recent advances," *Inf. Sci.*, vol. 546, pp. 835–857, Feb. 2021.
[23] I. Häggström, C. R. Schmidtlein, G. Campanella, and T. J. Fuchs, "DeepPET: A deep encoder–decoder network for directly solving the PET image reconstruction inverse problem," *Med. Image Anal.*, vol. 54, pp. 253–262, May 2019.
[24] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Jul. 2015.
[25] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[27] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.
[28] T. Y. Lin, P. Dollár, R. Girshick, and K. He, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2117–2125.
[29] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "SOD-MTGAN: Small object detection via multi-task generative adversarial network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 206–221.
[30] J. A. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, "Perceptual generative adversarial networks for small object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1222–1230.
[31] J. Noh, W. Bae, W. Lee, J. Seo, and G. Kim, "Better to follow, follow to be better: Towards precise supervision of feature super-resolution for small object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9724–9733.
[32] S. Bell, C. L. Zitnick, K. Bala, and R. Girshick, "Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2874–2883.
[33] X. Tang, D. K. Du, Z. He, and J. Liu, "PyramidBox: A context-assisted single shot face detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 797–813.
[34] H. Han, J. Y. Gu, Z. Zhang, J. F. Dai, and Y. C. Wei, "Relation networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3588–3597.
[35] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, and K. Cho, "Augmentation for small object detection," 2019, *arXiv:1902.07296*.
[36] G. Liu, J. Han, and W. Rong, "Feedback-driven loss function for small object detection," *Image Vis. Comput.*, vol. 111, Jul. 2021, Art. no. 104197.
[37] A. Fateh, R. T. Birgani, M. Fateh, and V. Abolghasemi, "Advancing multilingual handwritten numeral recognition with attention-driven transfer learning," *IEEE Access*, vol. 12, pp. 41381–41395, 2024.
[38] J. B. Lei, Z. M. Wang, and J. Li, "Research on target detection of fragment image based on Faster R-CNN," *Foreign Electron. Meas. Technol.*, vol. 40, no. 1, pp. 70–74, 2021.
[39] S. Xu, "Measurement of warhead fragment parameters based on vision," M.S. thesis, School Electron. Inf. Eng., Xi 'an Univ. Technol., Xi'an, China, 2022.
[40] Q. Wei, Z. Q. He, and Y. L. Wang, "Research on fragment recognition method of static explosion field based on DenseNet and attention mechanism," *Ordnance Equip. Eng.*, vol. 44, no. 2, pp. 259–265, 2023.
[41] S. E. Rigby, A. Tyas, R. J. Curry, and G. S. Langdon, "Experimental measurement of specific impulse distribution and transient deformation of plates subjected to near-field explosive blasts," *Experim. Mech.*, vol. 59, no. 2, pp. 163–178, Feb. 2019.
[42] K. Kornél, "2D and 3D perspective transformations," *Comput. Graph.*, vol. 14, no. 1, pp. 117–124, Jan. 1990.
[43] A. M. Rekavandi, S. Rashidi, F. Boussaid, S. Hoefs, E. Akbas, and M. bennamoun, "Transformers in small object detection: A benchmark and survey of state-of-the-art," 2023, *arXiv:2309.04902*.
[44] I. Khalfaoui-Hassani, T. Pellegrini, and T. Masquelier, "Dilated convolution with learnable spacings," 2021, *arXiv:2112.03740*.
[45] E. M. Bakr, A. El-Sallab, and M. Rashwan, "EMCA: Efficient multiscale channel attention module," *IEEE Access*, vol. 10, pp. 103447–103461, 2022.
[46] J. Wang, J. Yu, and Z. He, "DECA: A novel multi-scale efficient channel attention module for object detection in real-life fire images," *Appl. Intell.*, vol. 52, no. 2, pp. 1–14, 2022.

JUNCHAI GAO received the B.S. degree in educational technology from Shaanxi Normal University, Xi'an, China, in 1995, the M.S. degree in testing and measurement techniques and instruments from Xi'an Technological University, Xi'an, in 2006, and the Ph.D. degree in weapons science and technology from Northwestern Polytechnical University, Xi'an, in 2018.

She is currently an Associate Professor with Xi'an Technological University, where she is engaged in research on image processing, computer vision, and visual navigation technology. She has published more than ten articles in academic journals indexed in well-reputed databases, such as the Engineering Index and Science Citation Index. She also holds six patents and won more than ten science and technology progress awards.

HAORUI HAN was born in Xi'an, Shaanxi, China, in 2000. He received the B.S. degree from the North University of China, Taiyuan, China, in 2022. He is currently pursuing the master's degree with Xi'an Technological University, Xi'an.

His research interests include control theory and control engineering, such as image recognition in target plates and research on sky screen targets.

HANSHAN LI received the B.S. degree in electronic measurement and instrumentation from Xi'an Shiyou University, Xi'an, China, in 2001, the M.S. degree in testing and measurement techniques and instruments from Xi'an Technological University, Xi'an, in 2004, and the Ph.D. degree in testing and measurement techniques and instruments from Northwestern Polytechnical University, Xi'an, in 2010.

He is currently a Professor and a Ph.D. Supervisor with Xi'an Technological University, where he is engaged in research and development of photoelectricity detection, measurement and control technology, dynamic object test technology, image processing technology, and target damage assessment. He has published more than 90 articles in academic journals indexed in well-reputed databases, such as Science Citation Index. He holds 35 patents and won more than 19 science and technology progress awards.