

東京大学大学院新領域創成科学研究科

複雑理工学専攻

令和4年度

修士論文

**Toward Faster and Accurate Post-disaster Damage  
Assessment: Development of End-to-End Building Damage  
Detection Framework with Super-resolution Architecture**

2022年9月2日提出

指導教員 吉川 一朗 教授

富 宣超

**Toward Faster and Accurate Post-disaster Damage  
Assessment: Development of End-to-End Building Damage  
Detection Framework with Super-resolution Architecture**

**Xuanchao Fu**

47-206131

Complexity Science and Engineering

Completed: September 2022

Keywords: *Remote Sensing, Deep Learning, Super-resolution, Semantic segmentation, Building damage, Natural disasters*

**Abstract**

Building damage detection (BDD) using satellite images has been frequently adopted as an essential reference for post-disaster rescue, whereas its timeliness is significantly impacted by the long revisit time of high-resolution remote sensing satellites. Therefore, a reliable super-resolution method which is optimized for accurate and detailed BDD is important for advancing the BDD analysis even when we can use only low-resolution images captured by nanosatellites after a disaster. Based on Super-Resolution Generative Adversarial Network (SRGAN) and U-Net convolutional network, an efficient and novel BDD framework is proposed in this research for obtaining upsampled BDD results from low-resolution post-disaster images. We trained the framework using two disasters from the xBD dataset and tested three different structures. The results show that our training structure based on an end-to-end framework successfully generated super-resolution BDD maps from low-resolution images, which performed significantly better than those from a two-stage training structure.

# Contents

<i>Acknowledgement .....</i>	<i>1</i>
<b>1. Introduction .....</b>	<b>2</b>
<b>1.1 Background .....</b>	<b>2</b>
1.1.1 Building damage detection (BDD) with satellite images .....	2
1.1.2 CNN model-based BDD framework.....	3
1.1.3 Applying nanosatellite images to BDD.....	4
1.1.4 CNN-based Super-resolution model .....	6
<b>1.2 Purpose and Overview .....</b>	<b>8</b>
<b>2. Dataset and data processing.....</b>	<b>10</b>
<b>2.1 xBD dataset .....</b>	<b>10</b>
<b>2.2 Data pre-processing .....</b>	<b>12</b>
2.2.1 Image cropping with undersampling.....	12
2.2.2 EfficientNetV2-based cloud coverage image screening.....	12
2.2.3 LR image generation based on Gaussian blur and Bicubic .....	14
2.2.4 Dataset segmentation and data augmentation .....	16
<b>2.3 Dataset structure.....</b>	<b>16</b>
<b>3. Framework structure and module definition.....</b>	<b>18</b>
<b>3.1 Overall structure of BDD framework.....</b>	<b>18</b>
<b>3.2 Structure of Seg module.....</b>	<b>19</b>
3.2.1 Structure of BDD-U-Net .....	19

3.2.2 Loss function of BDD-U-Net .....	20
<b>3.3 Structure of SR module.....</b>	<b>22</b>
3.3.1 Two different types of SR modules .....	22
3.3.2 Structure and loss function of SRGAN .....	22
3.3.3 Structure and loss function of BDD-SRGAN.....	25
<b>3.4 Structure of proposed three BDD frameworks .....</b>	<b>27</b>
3.4.1 Structure of SR-B framework.....	27
3.4.2 Structure of BSR-B framework .....	29
3.4.3 Structure of EEBSR-B framework.....	30
<b>3.5 Structure of the control group BDD framework.....</b>	<b>31</b>
<b>3.6 Training and evaluation.....</b>	<b>33</b>
3.6.1 Training conditions and equipment.....	33
3.6.2 Overall F1-score .....	34
3.6.3 PSNR and SSIM.....	35
<b>4. Results and Discussions .....</b>	<b>37</b>
<b>4.1 Results on tsunami-dataset .....</b>	<b>37</b>
4.1.1 Overall F1-score results .....	37
4.1.2 PSNR/SSIM results.....	38
4.1.3 Evaluation of the output of Seg modules.....	39
4.1.4 Evaluation of the output of SR modules.....	42
<b>4.2 A discussion on the advantage of the frameworks .....</b>	<b>44</b>
4.2.1 Spatial attention mechanism and SAM module.....	44
4.2.2 Analysis of attentional maps.....	46

<b>4.3 A discussion on the generalizability of the EEBSR-B framework.....</b>	<b>48</b>
4.3.1 Wildfire and hurricane datasets .....	48
4.3.2 Results on wildfire and hurricane datasets .....	48
<b>4.4 A discussion on the expandability of the EEBSR-B framework .....</b>	<b>50</b>
4.4.1 ESRGAN-based EEBSR-B framework.....	50
4.4.2 Results of EEBSR-B based on SRGAN or ESRGAN.....	51
<b>5. Conclusions.....</b>	<b>53</b>
<b>Reference .....</b>	<b>55</b>

# Acknowledgement

First and foremost, I am most grateful to my supervisor, Prof. Ichiro Yoshikawa, whose useful suggestions, incisive comments, and constructive criticism have contributed greatly to the completion of this thesis. Without his illuminating instruction and patience, this thesis could not have reached its present form.

I would like to express my most profound appreciation to Dr. Toru Kouyama (National Institute of Advanced Industrial Science and Technology) and Lecturer Dr. Naoto Yokoya for their valuable inputs and suggestions throughout this research.

I am also grateful to thank Lecturer Dr. Kazuo Yoshioka and Prof. Takeshi Imamura, for their encouragement and insightful comments.

A special thanks also to all members of the Geoinformation Service Research Team of the National Institute of Advanced Industrial Science and Technology for their helpful comments and suggestions. Another special thanks to the team that operates and maintains the AI Bridging Cloud Infrastructure (ABCI) system for providing stable and high-quality computing resources for the experimental sessions in this thesis.

I thank my fellow labmates in Yoshikawa: Mr. Hang Yang, Mr. Fumiharu Suzuki, Mr. Huiyang Lin, Mr. Yi Yang and Ms. Jie Yang, for the stimulating discussions and for all the knowledge we have had in the last two years.

# **1. Introduction**

## **1.1 Background**

### **1.1.1 Building damage detection (BDD) with satellite images**

With the impact of climate change and human activities, the intensity and frequency of natural disasters have been increasing in recent years, with upwards of 300 serious natural disasters occurring each year in the last decade. The impacts caused by these disasters bring serious challenges to the development of the economy and the safety of human life and property. The damage level of buildings in various natural disasters typically has a strong spatial correlation with human casualties and economic losses, and thus there is an urgent need for high-accurate building damage detection (BDD) approach with low response time to support post-disaster rescue and response.

In recent years, remote sensing based BDD approaches have been extensively researched and applied with increasing imaging performance, i.e., higher resolution less than 1 m, which can resolve even individual houses. In many types of remote sensing data, the pre/post-disaster dual-phase optical remote sensing images are currently the most widely used in BDD tasks which can both realistically reflect the ground surface details with high-resolution (HR) and extract more information of building damage from the changes of pre/post-disaster images.

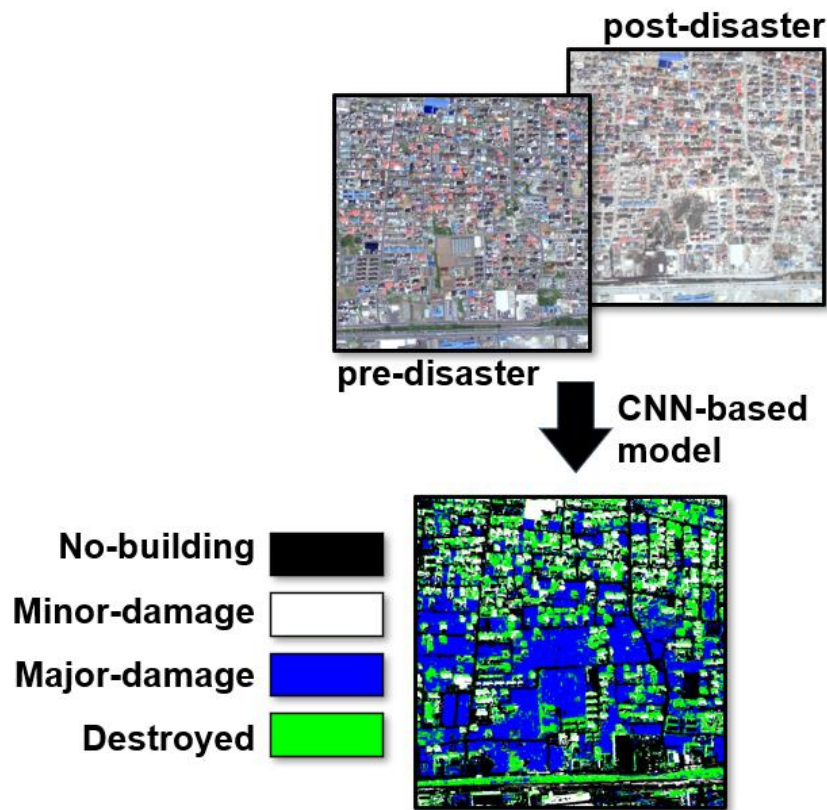
As another aspect of BDD, in recent years, increasingly nano and small satellites have been launched (more than 1,000 satellites per year), and the constellation of satellites

has been realized. In such a situation, continuous monitoring of hazard damage can be done, while the faster decision for planning the next observation based on earlier satellite observation is important. The human decision may be a bottleneck for such an operation, and thus it is important to establish an automated process in a ground station, such as the Kashiwa-Campus site, which can provide “human-like” results and replace human planning. From this viewpoint, it is natural to consider utilizing artificial intelligence technology.

### **1.1.2 CNN model-based BDD framework**

In recent years, convolutional neural network (CNN)-based models have been widely used in BDD tasks for remote sensing images, as shown in Figure 1.1, the current conventional CNN-based BDD framework generally requires pre- and post-disaster HR optical remote sensing images as input and outputs high accuracy semantic segmentation results with building damage information.

And the U-Net Convolutional Network has even performed well in BDD practice after various severe disasters in recent years and has become a standard structure of BDD frameworks. For instance, Ermlick et al. proposed a multimodal U-Net model with better performance in the BDD task and showed that this structure deserves further exploration. Other studies using different variants of U-Net have also yielded positive results.



**Figure 1.1 Conventional CNN model-based BDD framework**

### **1.1.3 Applying nanosatellite images to BDD**

Despite the improvement in accuracy of the CNN-based BDD framework for pre/post-disaster optical images, the low frequency of obtaining Post-disaster HR (more than 2 days for <1 m resolution, Table 1.1) images is a serious challenge after disasters.

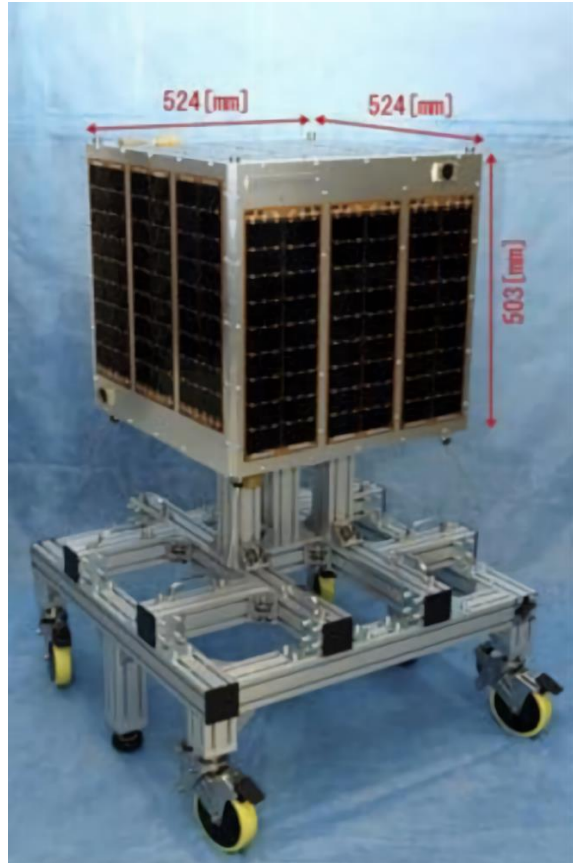
Current studies related to BDD and building detection mainly use HR optical remote sensing images with ground sample distance (GSD) less than 1m. On the other hand, images with GSD of more than 1m are rarely used since it is difficult to distinguish the contours of buildings, and images with GSD of 3-5m are even difficult to distinguish the presence of buildings, while the images can be obtained every day.

Satellite Name	Revisit Time	Maximum GSD
WorldView-2	> 2 days	0.46m
WorldView-3	> 2 days	0.31m
WorldView-4	> 2 days	0.31m
GeoEye-1	> 2 days	0.41m
IKONOS	> 3 days	0.82m
Pleiades-1A/1B	> 1 days	0.7m

**Table 1.1** The information of typical optical remote sensing satellites

As in Table 1.1, HR satellites typically require a revisit time of 2-3 days (or more), which means that it takes at least 48-72 hours to get a high accuracy building damage map of the disaster area with the current BDD framework. However, 24 hours after a severe disaster is usually regarded as the golden window to minimize fatalities by rescue. Therefore, the timeliness of CNN-based BDD framework based on HR pre/post-disaster images is severely inadequate in practical applications.

In this current situation, it becomes more feasible to acquire LR images within 24 hours or even less after a disaster based on a cluster of nanosatellites, and to obtain BDD results using pre-disaster HR images and post-disaster LR images. For example, the Hodoyoshi series (as shown in Figure 1.2) of nanosatellites developed under the leadership of the University of Tokyo is capable of taking optical remote sensing images up to GSD = 6.7m.



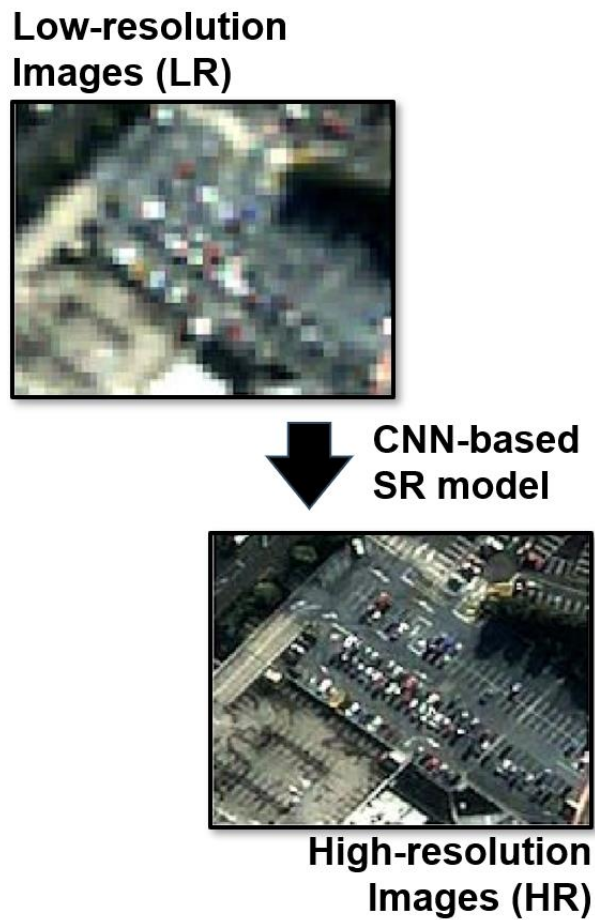
**Figure 1.2 Appearance and dimensions of the nanosatellite Hodoyoshi-1 (Nano-Satellite Center, 2012)**

The lower unit price and launch costs make it possible to deploy a large number of similar nanosatellites, and if hundreds of similar nanosatellites are deployed in orbit, there is a greater possibility of reducing the time to acquire LR images to less than 24 hours, even taking into account the effects of weather. However, although nanosatellites can quickly acquire post-disaster LR images, it is still a challenge to obtain highly accurate BDD results with pre-disaster HR images and post-disaster LR images that differ greatly in resolution.

#### **1.1.4 CNN-based Super-resolution model**

For the above challenge, the super-resolution techniques, which have been dramatically

improved in recent years, have the potential to resolve the problem of the low resolution of images acquired by nanosatellites. As shown in Figure 1.3, trained CNN-based super-resolution models can generate HR images based on the input LR images with more natural and high confidence relative to traditional methods.



**Figure 1.3 Conventional CNN model-based BDD framework (Müller et al., 2020)**

CNN-based super-resolution models have also been widely used in recent years for super-resolution in remote sensing images, for example, Courtrai et al. (2020) have shown that super-resolution generative adversarial network (SRGAN) can improve the accuracy of aircraft detection from remote sensing images. While Zhang et al. (2021)

proposed a two-stage framework performs high accuracy building detection from LR remote sensing images. These studies show that SR methods such as SRGAN can enhance potential of LR post-disaster images as an imitation of HR post-disaster images.

## 1.2 Purpose and Overview

In this research, we assume a scenario where HR pre-disaster images are available but only LR images captured by nanosatellites are available as post disaster images at a time point within 24 hours after a disaster due to the reasons such as the long revisit time of HR satellites and weather. Our target is to design a BDD framework effective in this scenario.

Under this situation, we can use only LR images as post-disaster images. We propose a novel End-to-End BDD framework and two non-End-to-End frameworks for comparison based on SRGAN and pre/post-disaster double-modal U-Net (BDD-U-Net). In the non-End-to-End frameworks, the SR part and BDD part will be trained separately (i.e., two-step training framework), while in an end-to-end training framework SR part and BDD part will be trained alternately. We compare the Overall F1-score results of the three frameworks on the dataset of the Indonesian tsunami and evaluate the images output from the SR modules of the three frameworks with PSNR/SSIM/visual effects. After identifying the best performing framework among the three frameworks based on the experimental results, we tried to analyze why the framework is advantageous through a series of small-scale experiments and demonstrated that the framework performs well in other types of disasters and has a certain degree of expandability.

The major contributions in this research are as follows:

(1) Proposed an Overall End-to-End BDD framework structure with input of pre-disaster HR images and post-disaster LR images.

(2) For the above framework, designed a SRGAN-based super-resolution model and a pre/post-disaster double-modal semantic segmentation model.

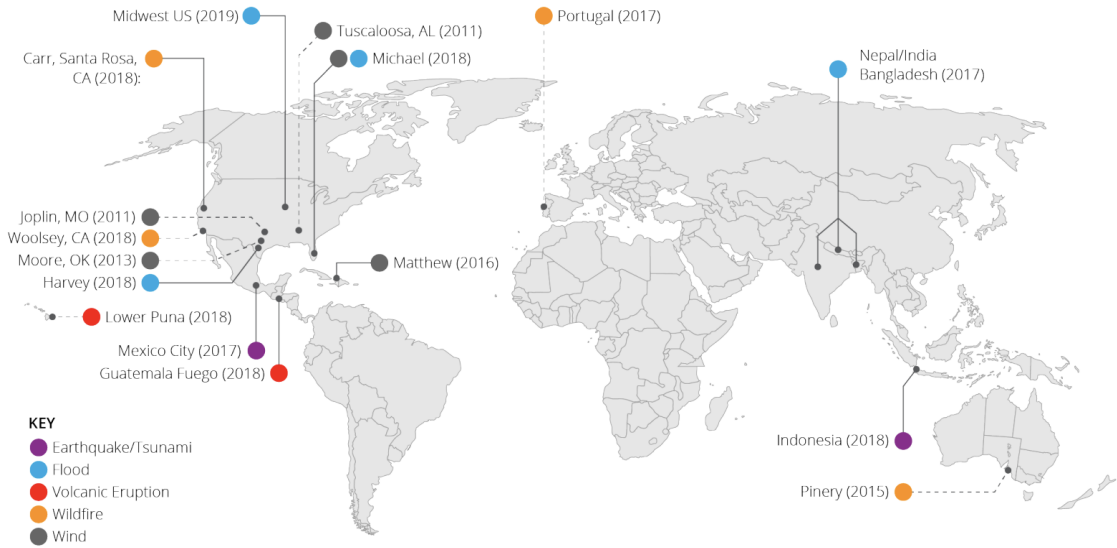
(3) Compared three different BDD frameworks with different structures based on the above overall structure and modules, including two frameworks for two-step training and one framework for end-to-end training.

(4) Analyzed the reason for the advantage of the best performing BDD framework through small-scale experiments and verified its generalizability and expandability.

## 2. Dataset and data processing

### 2.1 xBD dataset

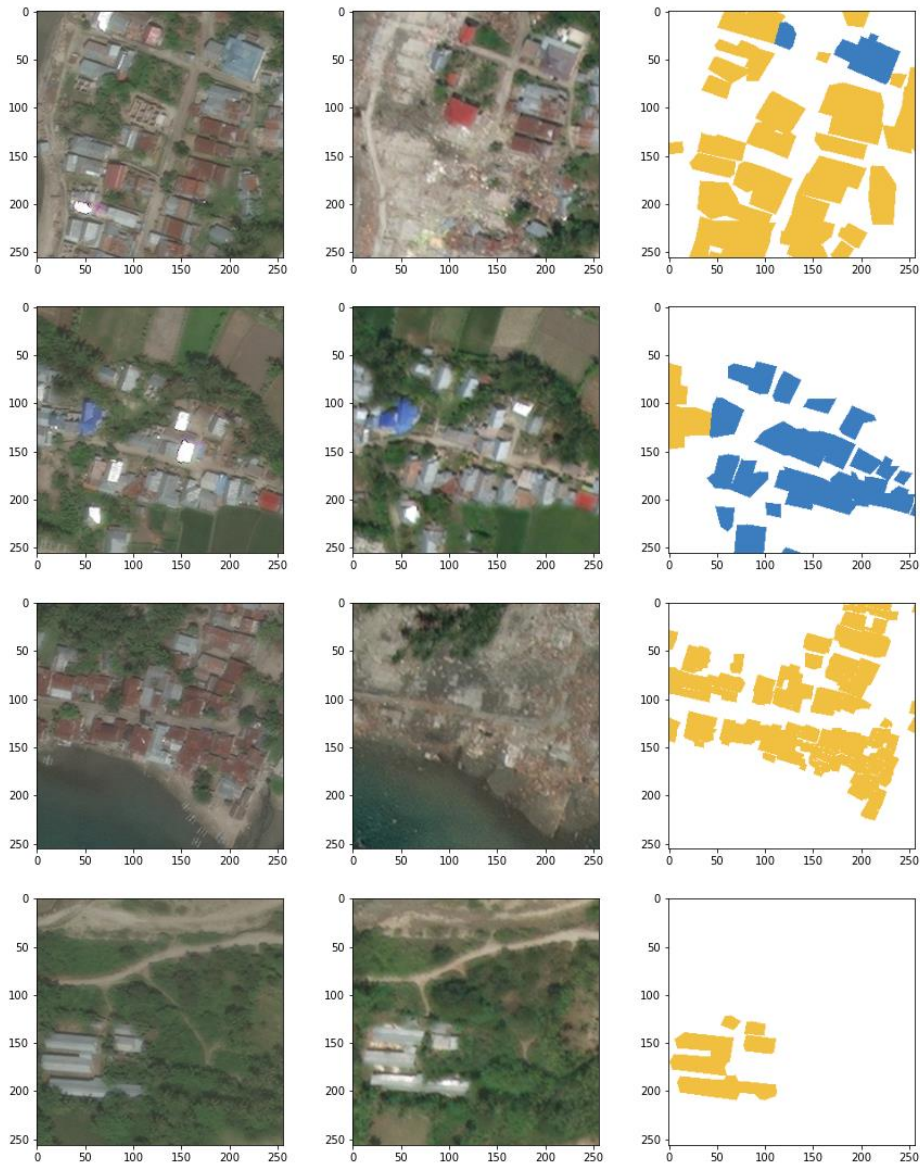
xBD dataset is a widely known natural disaster building damage dataset, which includes labels of pre-disaster building location and labels of post-disaster damage level for 850,736 buildings from six different disasters (as shown in Figure 2.1). xBD dataset provides paired pre/post-disaster HR (GSD = 0.5m) three-band RGB optical images, it is well suited to simulate the “without post-disaster HR images” scenario, which we assumed.



**Figure 2.1 Disaster types and disasters represented in xBD around the world (Gupta et al., 2019)**

For better quantitative evaluation of the performance of the BDD frameworks for tsunami disaster cases, we selected images of two disasters from the dataset, the 2018 Sunda Strait tsunami and 2018 Sulawesi earthquake and tsunami, and reclassified the building damages levels from the original four types (no damage, minor damage, major

damage, destroyed) to two types (no damage, damaged or destroyed), with the reason that the semantic segmentation task of six types is inherently too difficult to effectively represent the performance differences between the BDD frameworks proposed in this research (For example, severely damaged and slightly damaged buildings may look almost identical to the image, but the label was created with reference to other data).



**Figure 2.2 Examples of building labels in the reclassified xBD dataset, from left to right: pre-disaster, post-disaster, label (blue - undamaged, yellow - damaged or destroyed)**

## **2.2 Data pre-processing**

### **2.2.1 Image cropping with undersampling**

As the first step of data preprocessing, we cropped the original 1024x1024 size images in the xBD dataset to 256x256 size images in order to facilitate the input of the CNN-based model for training (For example, inputting 16 random images of size 256 will result in a more stable convergence of the CNN due to the richer types of features it contains, compared to inputting one image of size 1024). The specific cropping method is to first crop each 1024x1024 size image into 16 256x256 size images, then determine whether there are buildings in each 256x256 size image based on its label and save the images with buildings in them. The reason is that the xBD dataset is only guaranteed to have buildings in each 1024x1024 size image, and a direct crop would produce a large number of 256x256 size images without buildings, which would have a negative impact on the training of the model.

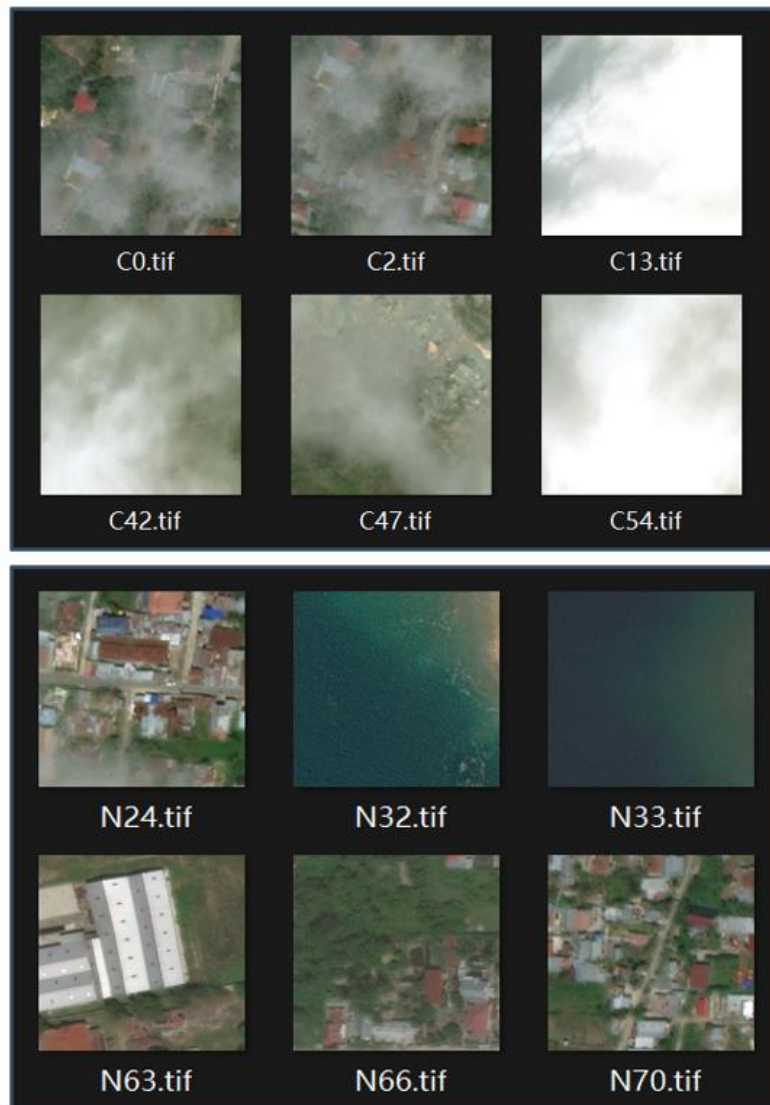
After the above processing, we obtained a dataset of 256x256 size while ensuring that there are buildings in each image (pre-disaster-HR, post-disaster-HR, label).

### **2.2.2 EfficientNetV2-based cloud coverage image screening**

The above image of the 256x256 size dataset still has a problem that will seriously affect the training of deep learning models. That is, the xBD dataset ensures that there are no images with high cloud coverage in the 1024x1024 size image, but many images with high cloud coverage in the 256x256 image after cropping, which will have a

significant impact on the training of super-resolution and semantic segmentation models.

To solve the problem, we manually selected a small number of samples with high cloud coverage from the dataset. We made a small image classification dataset with the same number of images with low cloud coverage (The criteria for judging high cloud coverage/low cloud coverage are shown in Figure 2.3).



**Figure 2.3 Examples of high cloud coverage images / low cloud coverage images  
(upper image: high cloud coverage, lower image: low cloud coverage)**

We selected EfficientNetV2, the image classification model currently considered to have the best performance and trained and tested it with the cloud-coverage image classification dataset mentioned above, finally obtaining more than 95% accuracy of high cloud-coverage/low cloud-coverage image classification.

Based on the trained EfficientNetV2 model, we filtered out high cloud coverage images for the whole dataset, which specifically classifies all pre-disaster and post-disaster images separately as to whether the cloud coverage is too high. If one of the pre-disaster and post-disaster has too much cloud coverage, that group of data (pre-disaster, post-disaster, label) was removed from the dataset.

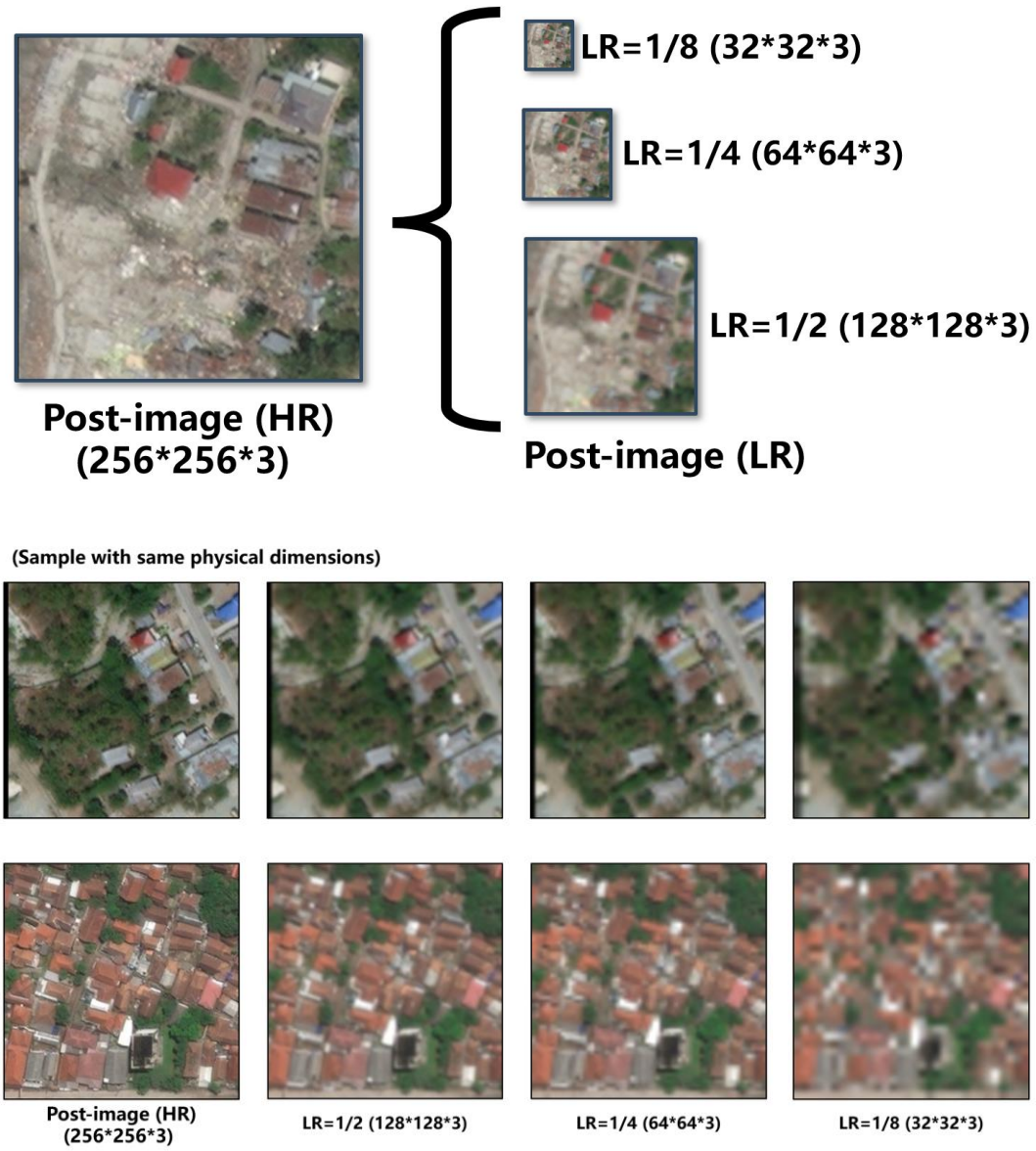
After the above processing, we obtained 256 by 256 size dataset with less high cloud coverage images (pre-disaster-HR, post-disaster-HR, label).

### **2.2.3 LR image generation based on Gaussian blur and Bicubic**

Although the above dataset can already be used as training/validation/testing data for the semantic segmentation model, because the super-resolution model is also available in the BDD framework proposed in this research, a low-resolution version needs to be produced for each post-disaster image in the dataset.

In this research, we chose the method of Gaussian blurring followed by Bicubic downsampling to produce a low-resolution version of each post-disaster image. The images obtained by this method proved closer to the authentic low-resolution images. We used the above method to produce low-resolution versions of each post-disaster

image at 1/2, 1/4, and 1/8 times the resolution, with the dimensions and appearance shown in Figure 2.4.



**Figure 2.4 Size and appearance of low-resolution versions of post-disaster images**

After the above processing, we obtained a 256x256 size dataset with a low-resolution version of the image with less high cloud coverage images (pre-disaster-HR, post-disaster-HR, post-disaster-LR, label).

#### **2.2.4 Dataset segmentation and data augmentation**

To facilitate the training/validation/testing of the super-resolution and semantic segmentation models, we created a training set, validation set, and test set with the data number ratio of 6/2/2. After segmenting the dataset, the last issue that negatively affects the training is that the three types of labels, no-building/undamaged-building/damaged or destroyed-building, account for 75.70%/18.65%/5.63%, respectively, in the whole dataset, and such an unbalanced ratio may have an impact on the performance of the semantic segmentation model.

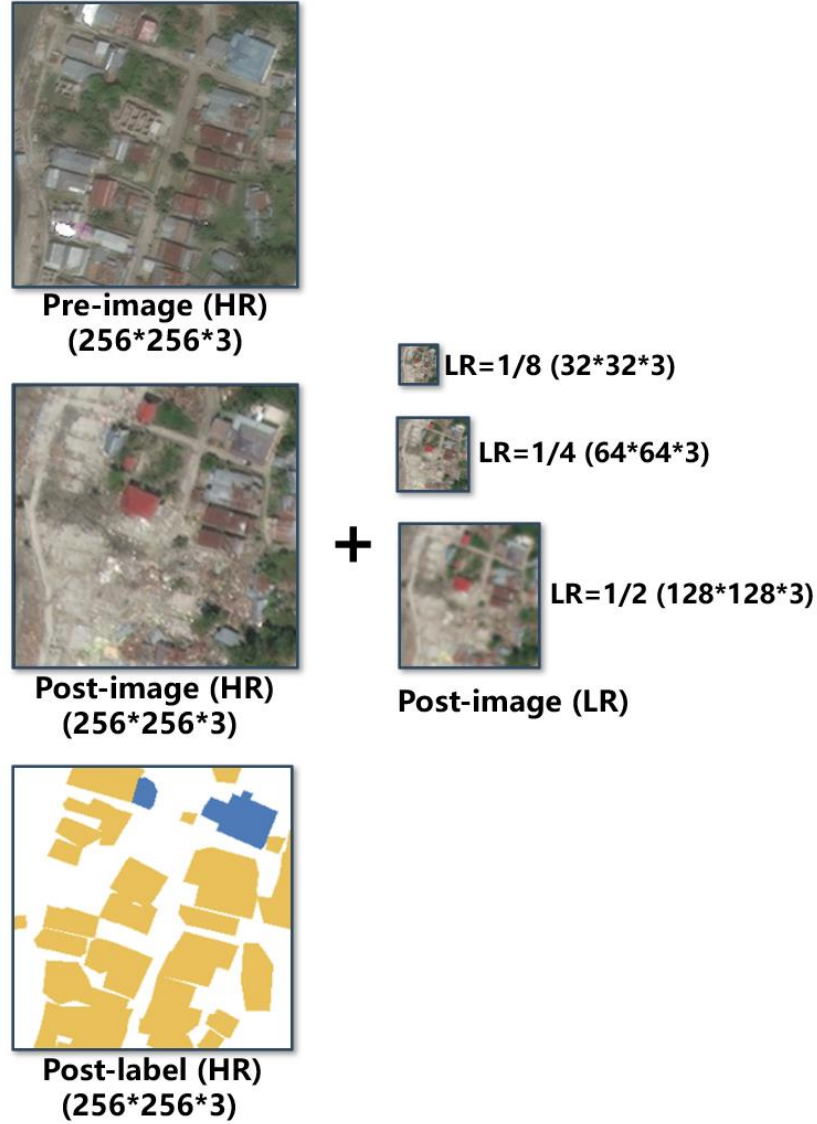
To address this issue, we chose two methods to reduce the impact caused by sample imbalance. The first method is to perform undersampling data augmentation on the training set by rotating and flipping the groups in the training set that contain damaged or destroyed buildings (90/180/270-degree rotation and X-axis/Y-axis flip), thus increasing the proportion of damaged building samples in the training set. The second method will be introduced when the loss function of the semantic segmentation model is introduced.

### **2.3 Dataset structure**

After the above four steps, the structure of the dataset is shown in Figure 2.5. Each group of data in the dataset includes the pre-disaster HR image, the post-disaster HR image, three low-resolution versions of the post-disaster HR image (1/2, 1/4, and 1/8), and the semantic segmentation labels.

The pre/post-disaster HR images and semantic segmentation labels are used as

training/validation/testing data for the semantic segmentation model, while the post-disaster HR images and the three low-resolution versions of the post-disaster HR images are used as training/validation/testing data for the super-resolution model.



**Figure 2.5 The group of data in the processed dataset consists of pre-disaster HR, post-disaster HR, post-disaster LR (1/2, 1/4 and 1/8), and label**

The dataset that completes the above preprocessing is used in our main experiments, which we will refer to as the tsunami-dataset.

### 3. Framework structure and module definition

#### 3.1 Overall structure of BDD framework

This section introduces the overall structure of our BDD framework. As illustrated in Figure 3.1, our proposed framework consists of the SR and Seg modules. The SR module at the first part is used to transform the LR post-disaster images into SR post-disaster images of the same resolution as the HR images, while the Seg module at the second part is used to evaluate the damage of buildings from the paired HR pre-disaster images and SR post-disaster images.

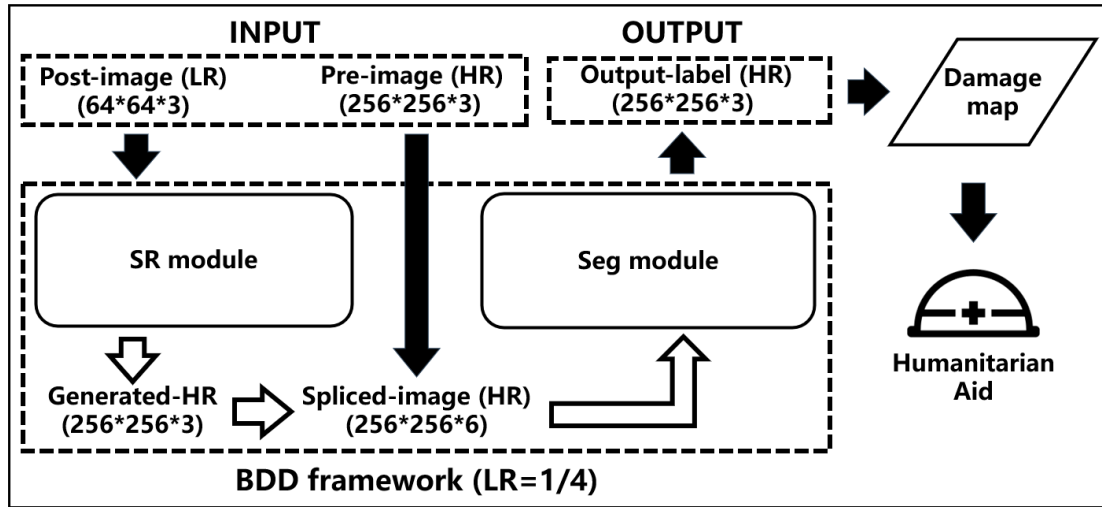


Figure 3.1 Overall structure of the BDD framework (with LR=1/4)

Based on this overall structure, we tested with three BDD frameworks with different module structures and training processes in this paper to find an appropriate structure for our purpose. The specific structures will be described in detail in the subsequent sections.

## 3.2 Structure of Seg module

### 3.2.1 Structure of BDD-U-Net

As illustrated in Figure 3.2, we construct a double-modal model based on the U-Net structure, we call it as “BDD-U-Net” as the Seg module in the framework. This structure can concatenate features of both pre-disaster and post-disaster images extracted from each convolutional layer.

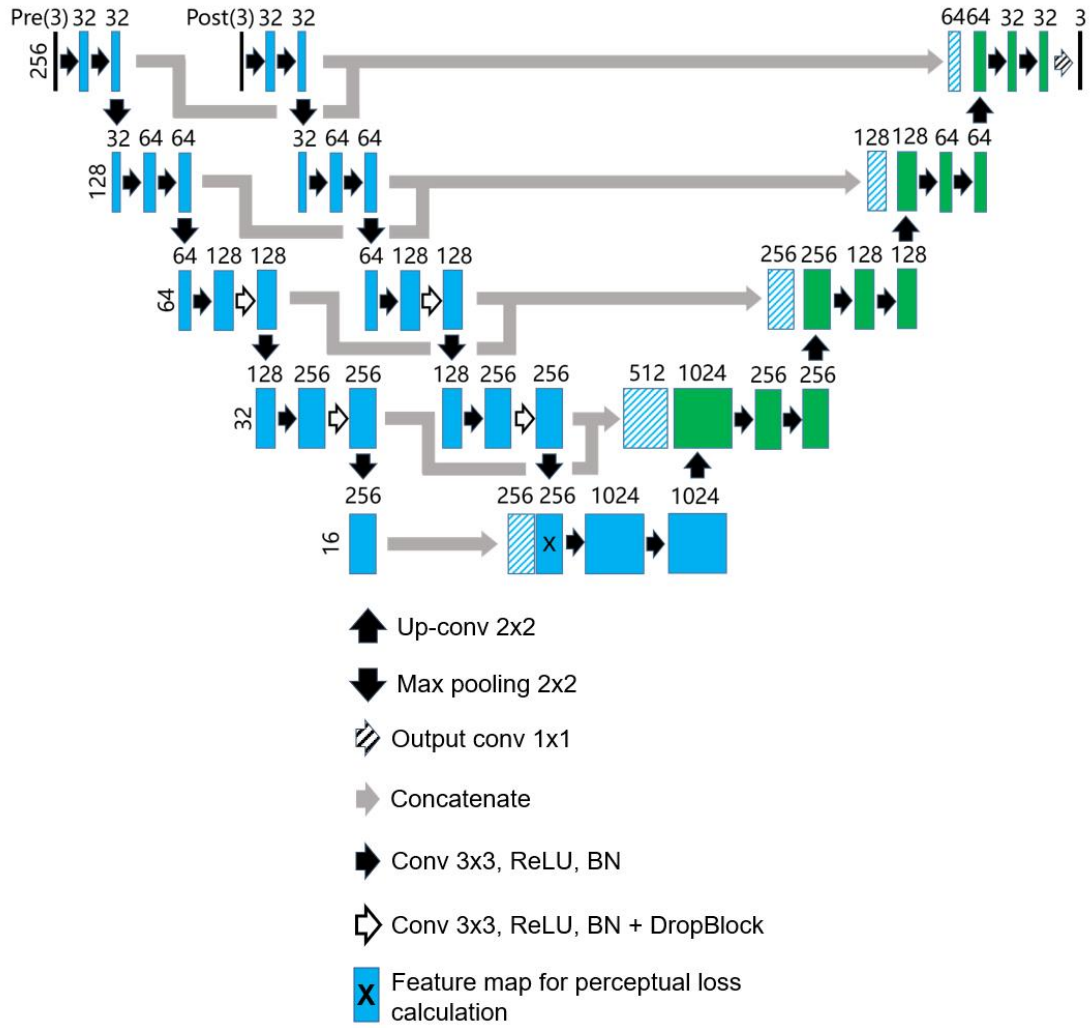


Figure 3.2 Structure of the BDD-U-net

It has been confirmed that the structure is well able to perform the building detection and damage assessment tasks simultaneously. Our BDD-U-Net also uses the DropBlock layer, which motivates the network to learn more robust and efficient features by randomly dropping local feature maps. It is worth noting that in the structure of BDD-U-Net, there is a feature map labeled with the “X” symbol, which has an essential role in the training of the SR module, as will be explained in the introduction of the SR module.

We use the same structure of the Seg module in all three frameworks, but the training processes will be somewhat different due to the difference in the framework structure.

### **3.2.2 Loss function of BDD-U-Net**

Section 2.2.4 introduced that the tsunami-dataset has a sample imbalance problem. In addition to the imbalance in the number of samples, the damaged and destroyed buildings are inherently more difficult to distinguish for the semantic segmentation model. To solve these problems, we use Focal loss to train the Seg module.

Focal loss is a loss function to deal with sample imbalance, which focuses on the point of adding weights to the losses corresponding to the samples according to the ease of sample discrimination, i.e., adding a smaller weight to the easily distinguishable samples and a larger weight to the difficult samples. Focal loss is a modification of the cross-entropy loss function, and the standard cross-entropy loss function in the two-category classification case is as in Formula 3.1.

$$L = -y \log y' - (1 - y) \log (1 - y') = \begin{cases} -\log y', & y = 1 \\ -\log (1 - y'), & y = 0 \end{cases} \quad (3.1)$$

The  $y'$  in Formula 3.1 is the output of the model after the activation function, so its value is between 0 and 1. It can be seen that in the case of two-category classification, the ordinary cross-entropy loss function has a minor loss for positive samples with higher output probability. For negative samples, the smaller the output probability is, the smaller the loss is. In this case, the loss function is slow in the iterative process of a large number of simple samples and may not be optimized to the optimum.

$$L_{fl} = \begin{cases} -\alpha(1 - y')^\gamma \log y', & y = 1 \\ -(1 - \alpha)y'^\gamma \log (1 - y'), & y = 0 \end{cases} \quad (3.2)$$

For the above problem, as shown in Formula 3.2, the Focal loss first adds the gamma factor to the original one, where  $\gamma > 0$  can reduce the loss of easily classified samples and make the model focus more on the difficult samples. In addition, Focal loss also adds a balancing factor-alpha to balance the unevenness of the sample itself by reducing the weight of the more easily determined negative sample in the loss calculation.

When training the Seg module, the parameters of Focal loss were set to  $\gamma = 2$  and  $\alpha = 0.25$  (no-building and undamaged-building are considered as negative sample), which are the best parameters recommended in the Focal loss paper.

### 3.3 Structure of SR module

#### 3.3.1 Two different types of SR modules

The SRGAN models used as SR modules in this paper are divided into two types: SRGAN and BDD-SRGAN. Both models have the same structure of generator and discriminator as the original paper of SRGAN, with the difference that the loss function of BDD-SRGAN is modified from the original one.

#### 3.3.2 Structure and loss function of SRGAN

The first type of SR module is the original SRGAN model. There are three parts in the original SRGAN model, the generator, the discriminator, and the pre-trained VGG19 image classification model used to calculate the VGG loss (also called perceptual loss). The generator and discriminator use the same structure as in the SRGAN paper, whose internal structure is shown in Figure 3.3.

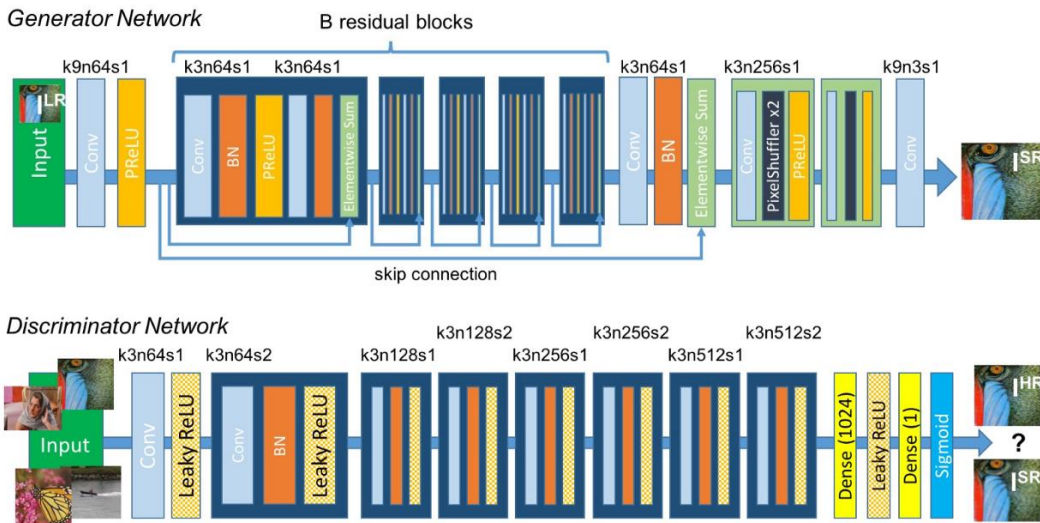
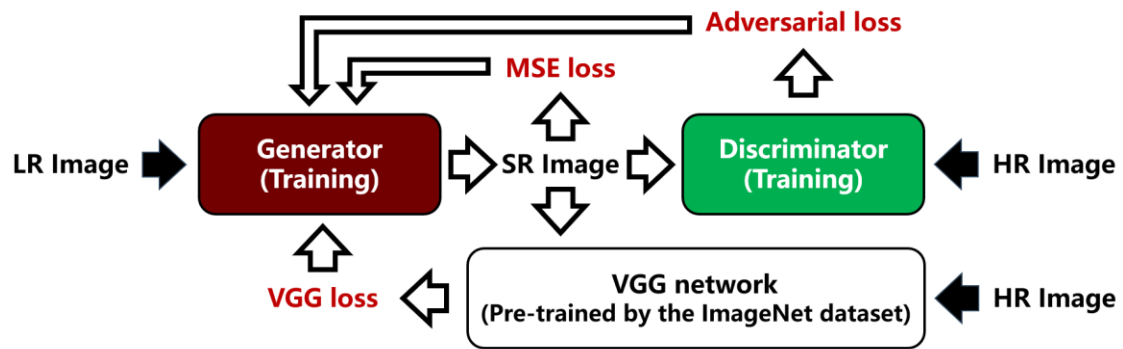


Figure 3.3 The generator and discriminator of SRGAN (Ledig et al., 2017)

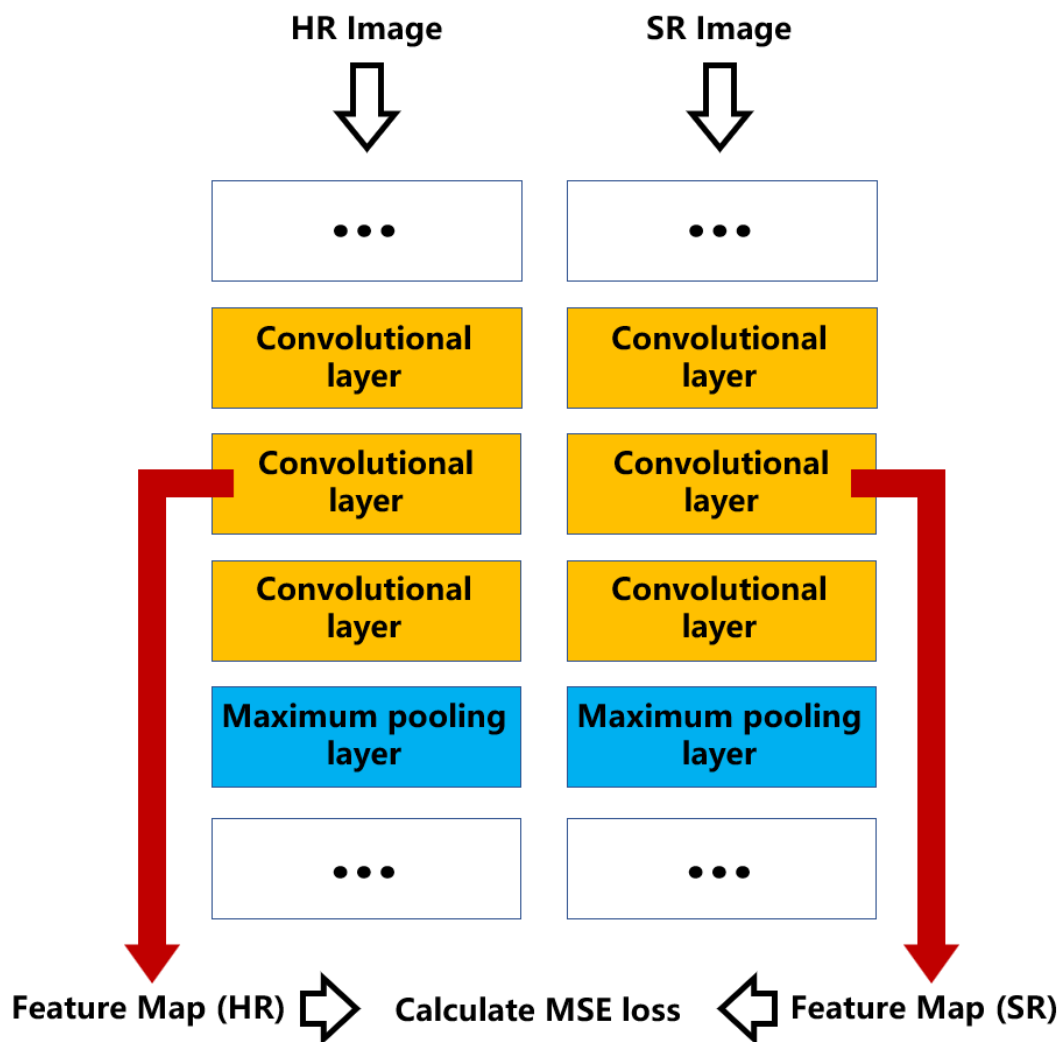
During the SRGAN's training, the LR image is fed into the generator, the SR image (the "fake" HR image generated by the generator) is obtained, and then the SR image is used to calculate three kinds of losses for training the generator. The process of calculating the loss of the generator in training for SRGAN is shown in Figure 3.4.



**Figure 3.4 Loss calculation for generator (SRGAN)**

The first one is MSE loss, which is obtained by calculating Mean Squared Error directly with the HR image, and the MSE loss is used to control that the generated SR image does not differ too much from the HR image at the pixel level.

The second is VGG loss, which is calculated by inputting both HR and SR images to the VGG19 model pre-trained on the ImageNet dataset and then calculating the Mean Squared Error of the feature maps at the same location of both models, as shown in Figure 3.5. The SRGAN in this experiment uses the feature map obtained from the 16th convolutional layer of VGG19 to calculate VGG loss. VGG loss is used to control the generator to generate SR images that are as visually realistic as possible and produce richer details.



**Figure 3.5 The calculation of VGG loss in the original SRGAN**

The third one is the discriminator-based Adversarial loss, which is calculated by inputting SR images into the discriminator, and the discriminator classifies them as HR or SR images. The accuracy of the classification is taken as the Adversarial loss. The Adversarial loss makes the generator generate more realistic SR images to better "trick" the discriminator.

$$L_{total} = MSE\ loss + 0.006 * VGG\ loss + 0.001 * adversarial\ loss \quad (3.3)$$

Based on the above three kinds of losses, the loss calculation formula for the generator used to train the generator is shown in Formula 3.3, where the coefficients are all taken to be the same values as in the original SRGAN paper. Since only the generator is used as the SR module in SRGAN after training, the loss used to train the discriminator is not introduced in detail and is calculated in the same way as the original SRGAN.

### 3.3.3 Structure and loss function of BDD-SRGAN

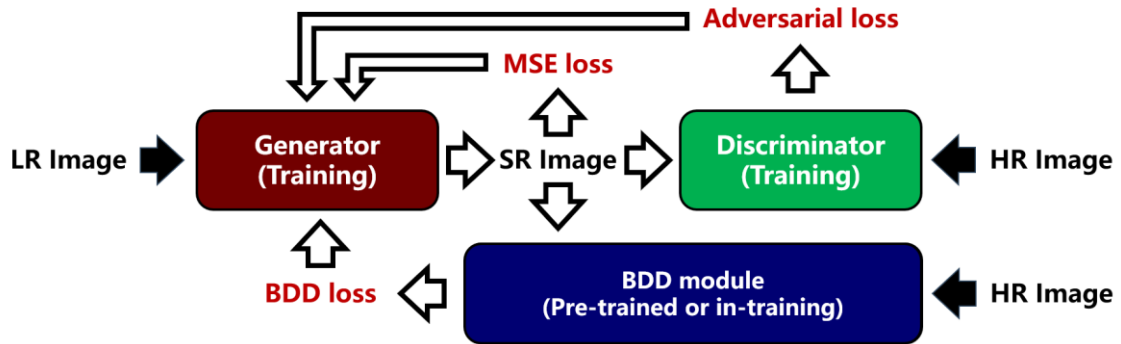
Although SRGAN is an excellent super-resolution model, there are two significant shortcomings of using SRGAN as an SR module in the BDD framework proposed in this research.

The first shortcoming is that the original SRGAN uses a VGG model pre-trained on the ImageNet dataset, which contains many different types of images but almost no remote sensing images. This leads to the fact that when inputting post-disaster remote sensing images of SR and HR into the pre-trained VGG19, the convolutional layers in the model may have difficulty extracting the practical features from them, thus reducing the effectiveness of VGG loss.

The second shortcoming is that using the pre-trained VGG model to calculate the VGG loss of SRGAN makes the SR module and Seg module of the BDD framework isolated, without any information transfer between the two modules, and also without a way to end-to-end train the whole framework.

To solve the above two shortcomings, we redesigned the structure and loss function of SRGAN. We named the improved model BDD-SRGAN, whose process of calculating the loss of the generator in training is shown in Figure 3.6.

The original SRGAN uses one of the feature maps of the VGG19 network pre-trained on ImageNet for calculating the perceptual loss, while in BDD-SRGAN, the above VGG19 network is replaced with the last feature map in the encoder of the post-disaster image in the pre-trained or in-training BDD-U-Net (represented in Figure 3.2 as the 16x16x256 size feature map derived from the post image, with the “X” symbol). For easy distinction, we refer to the VGG loss calculated using the above method as the BDD loss.



**Figure 3.6 Loss calculation for generator (BDD-SRGAN)**

Compared with the original SRGAN, we believe that BDD-SRGAN, which uses BDD loss to replace VGG loss, is more suitable as the SR module in the BDD framework. On the one hand, because BDD-U-Net can extract the valuable features of the post-disaster

HR images better than VGG19, the BDD loss obtained by inputting both SR and HR post-disaster images into BDD-U-Net can obtain more practical information than VGG loss. On the other hand, since BDD loss can be calculated by in-training BDD-U-Net, the SR module and Seg module of the BDD framework using BDD-SRGAN as the SR module can be trained simultaneously, which provides the possibility of end-to-end training of the BDD framework.

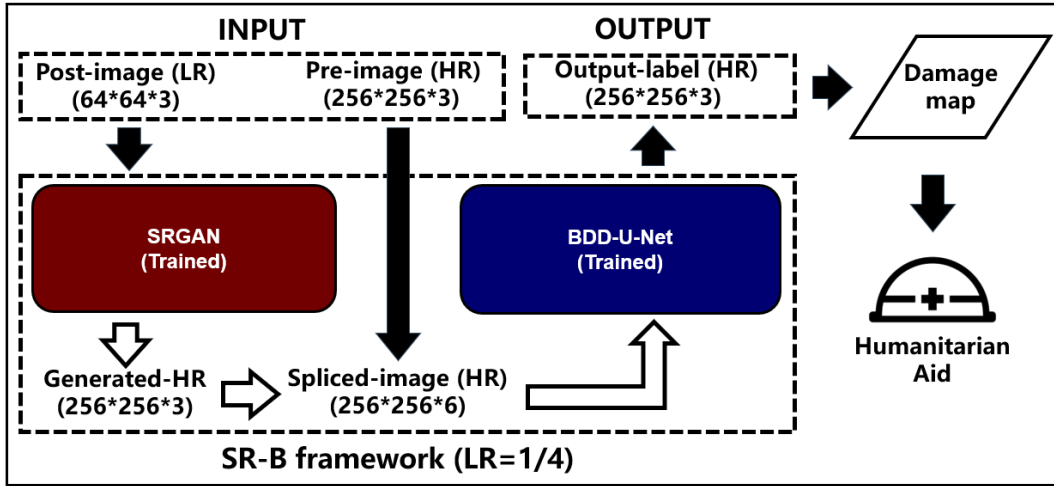
$$L_{total} = MSE\ loss + 0.006 * BDD\ loss + 0.001 * adversarial\ loss \quad (3.4)$$

With replacing the VGG loss with BDD loss, the loss calculation formula for the generator used to train BDD-SRGAN is shown in Formula 3.4, where the coefficients are still taken as the values in the original SRGAN paper.

### 3.4 Structure of proposed three BDD frameworks

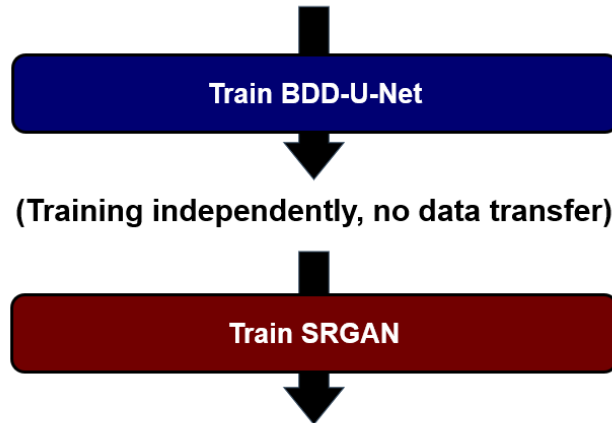
#### 3.4.1 Structure of SR-B framework

In Section 3.4, three BDD frameworks with different structures, modules, and training processes are introduced, whose structures are based on the overall structure of the BDD framework introduced above, while their modules are based on one Seg module (BDD-U-Net) and two SR modules (SRGAN, BDD-SRGAN) introduced above.



**Figure 3.7 Structure of the SR-B framework**

The first is the SRGAN/BDD-U-Net (SR-B) framework, where the SR module uses the original SRGAN, and the Seg module uses BDD-U-Net. The structure is shown in Figure as shown in Figure 3.7.



**Figure 3.8 Training process of SR-B framework**

The training process of the SR-B framework is shown in Figure 3.8. When training the SR-B framework, BDD-U-Net as the Seg module and SRGAN as the SR module are

trained completely independently with no data transfer between them during the training process.

### 3.4.2 Structure of BSR-B framework

The second is the BDD-SRGAN/BDD-U-Net (BSR-B) framework, in which the SR module uses BDD-SRGAN, and the Seg module uses BDD-U-Net. The structure is shown in Figure 3.9.

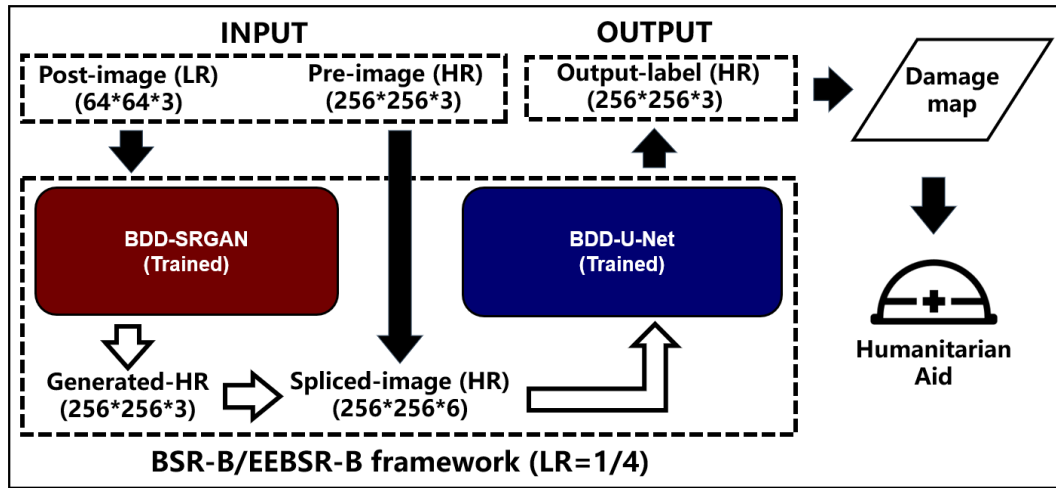
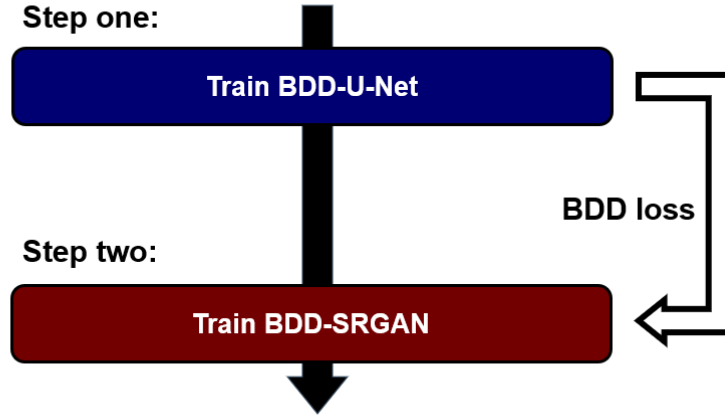


Figure 3.9 Structure of the BSR-B and EEBSR-B framework

The training process of the BSR-B framework is shown in Figure 3.10, where the BDD-U-Net is trained first and then use the feature map of the pre-trained BDD-U-Net to calculate the BDD loss in BDD-SRGAN (i.e., BDD-U-Net is fixed while BDD-SRGAN is trained), which can be used to supervise the SR module to generate SR post-disaster images that better meet the requirements of the Seg module.



**Figure 3.10 Training process of BSR-B framework**

### **3.4.3 Structure of EEBSR-B framework**

The third is the BDD-SRGAN/BDD-U-Net framework with end-to-end training (EEBSR-B), in which the SR module uses BDD-SRGAN and the Seg module uses BDD-U-Net. The structure is shown in Figure 3.9.

The training process of the EEBSR-B framework is shown in Figure 3.11. In contrast to the above two frameworks, the training process of EEBSR-B is characterized by first training the Seg module and then the SR module when training each batch, and the Seg module is trained using the post-disaster HR images output from the SR module instead of the real post-disaster HR images. At the same time, BDD-SRGAN uses the feature map of the BDD-U-Net during training to calculate the BDD loss.

We try to use this end-to-end training process to enable better information transfer between the Seg module and the SR module while tending the SR module to generate SR images that are more favorable to the current Seg module from the first batch and

continuing this state after that. This training process is inspired by the alternating training of generator and discriminator when training a GAN.

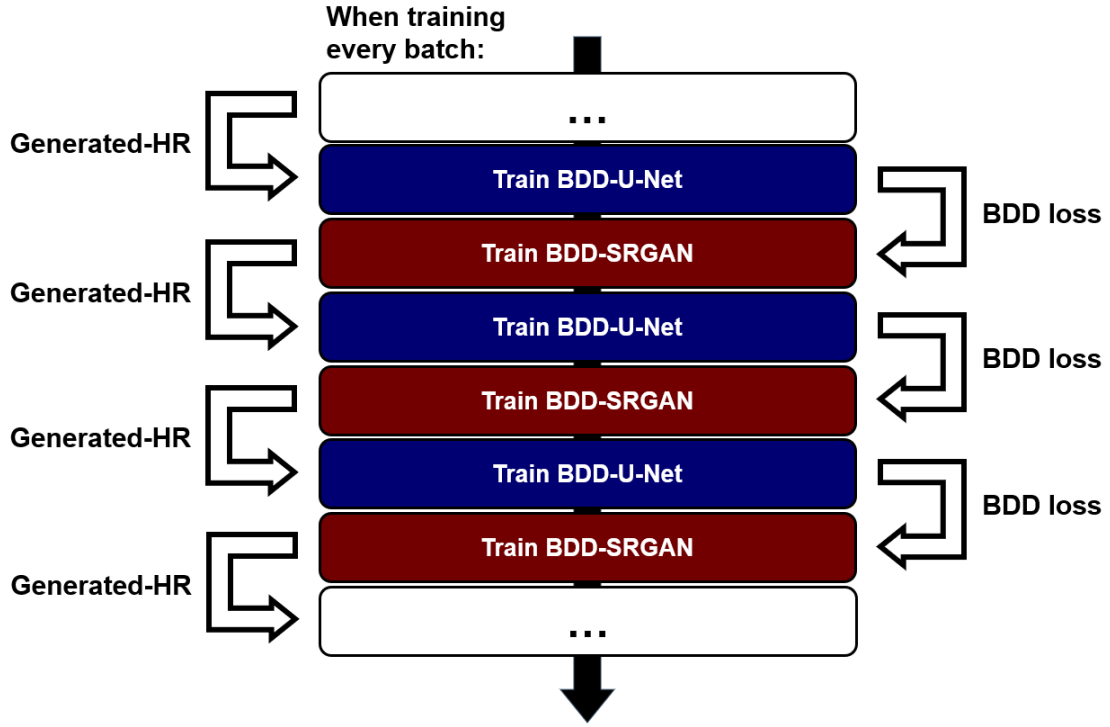


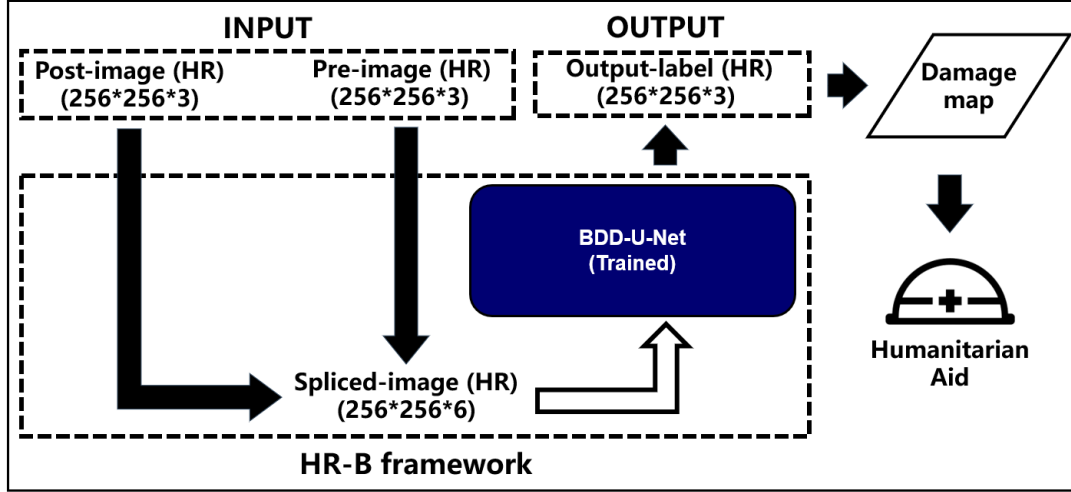
Figure 3.11 Training process of EEBSR-B framework

### 3.5 Structure of the control group BDD framework

In addition to the three frameworks proposed in this research introduced in Section 3.4, we also set up two control group frameworks based on the same overall BDD framework structure in order to better evaluate our proposed frameworks.

The structure of the first control group framework is shown in Figure 3.12. To confirm how much performance difference the three frameworks based on pre-disaster HR and post-disaster LR images proposed in this study have with the BDD framework using pre-disaster HR and post-disaster HR images in the ideal state, the framework removes

the SR module and directly inputs the pre-disaster HR and post-disaster HR images into the Seg module. We refer to this framework as HR-B, while the training process is such that only the Seg module is trained separately.



**Figure 3.12 Structure of the HR-B framework**

The structure of the second control group framework is shown in Figure 3.13. To compare the performance difference between the CNN-based super-resolution model and the traditional upsampling method in the BDD framework, which replaces the SR module with the Bicubic module and upsamples the post-disaster LR images directly. We refer to this framework as BIC-B. Because the Bicubic upsampling module does not need to be trained, the training process of this framework is such that only the Seg module is trained separately.

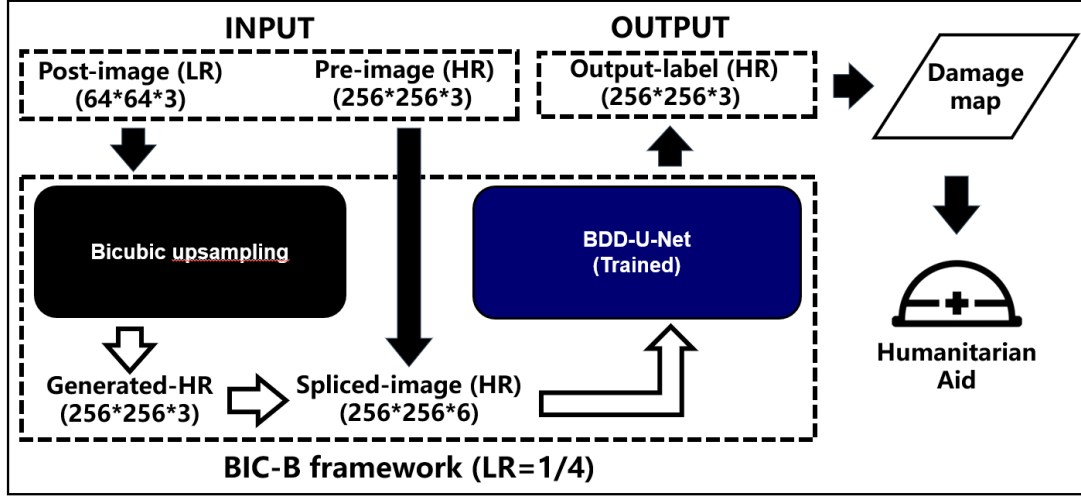


Figure 3.13 Structure of the BIC-B framework

## 3.6 Training and evaluation

### 3.6.1 Training conditions and equipment

To ensure the validity of the controlled experiments, the modules of all five frameworks above were trained using the same training conditions and equipment. We chose Adam with a learning rate of 0.0003 as the optimizer and set the batch size to 16. All three frameworks were trained with 128 epochs in NVIDIA A100-SXM4 and tested with the same equipment.

In the training process using the tsunami-dataset, we followed the dataset segmentation results introduced in Section 2.2.4 to use 60% of the data as the training set, while 20% of the data were used as the validation and test sets, respectively. In training every framework, during the training process of SR module independently, we saved the model with the lowest  $L_{total}$  on the validation set as the training result. For training the

Seg module independently, we saved the model with the highest Pixel Accuracy in the validation set as the training result. In the End-to-end training process, we saved the model of SR and Seg module with the highest Pixel Accuracy of Seg module on the validation set as the best End-to-End training result. The calculation of  $L_{total}$  for the SR module is shown in Section 3.3, and the calculation of Pixel Accuracy is shown in Formula 3.5.

$$Pixel\ Accuracy = \frac{total\ pixel\ classified\ correctly}{total\ number\ of\ pixel} \quad (3.5)$$

### 3.6.2 Overall F1-score

We chose the Overall F1-score as the main evaluation criterion for the performance of the Seg module and the BDD framework. For each sample category, the F1-score is calculated as shown in in Figure 3.14 and Formulas 3.6, 3.7, and 3.8, while the Overall F1-score is the average of the F1-scores of all categories.

		Predicted Class	
		Positive	Negative
Actual Class	Positive	tp	fn
	Negative	fp	tn

Figure 3.14 Definition of tp, fp, fn, tn

$$Precision = \frac{tp}{tp+fp} \quad (3.6)$$

$$Recall = \frac{tp}{tp+fn} \quad (3.7)$$

$$F1-score = \frac{2 * Recall * Precision}{Recall + Precision} \quad (3.8)$$

For example, in the category of undamaged buildings, Precision indicates the probability that the pixels of undamaged buildings among all pixels are correctly classified. In contrast, Recall indicates the proportion of pixels of truly undamaged buildings to all pixels judged as undamaged buildings. The F1-score is the harmonic average of the above two, which is considered to be a high reference evaluation criterion for semantic segmentation.

### 3.6.3 PSNR and SSIM

We chose PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity) as objective evaluation criteria for the output images of the SR module since these two metrics are commonly used to evaluate the performance of super-resolution models.

PSNR can be simply calculated by the mean squared difference MSE, which aims to measure the similarity of two images at the pixel level. Although the evaluation results usually appear inconsistent with the primary human perception due to the variability of human visual characteristics, it is still a reference evaluation criterion. SSIM, on the other hand, is an index that measures the degree of structural similarity between two images by comparing three indicators: brightness, contrast, and structure. Compared with PSNR, SSIM is relatively more consistent with human visual characteristics in

evaluating image quality.

However, it is worth noting that in recent studies related to super-resolution models, when evaluating the visual effect of super-resolution results, the values of PSNR and SSIM are still frequently found to be poorly correlated with the visual effect scores obtained from large-scale questionnaires, so this research will also evaluate the output of SR module by visual effect in addition to PSNR and SSIM.

## 4. Results and Discussions

### 4.1 Results on tsunami-dataset

#### 4.1.1 Overall F1-score results

In this section, we compare the training results of our proposed three frameworks and two control group frameworks at LR sizes of 1/2, 1/4 and 1/8. Their Overall F1-score results are shown in Table 4.1.

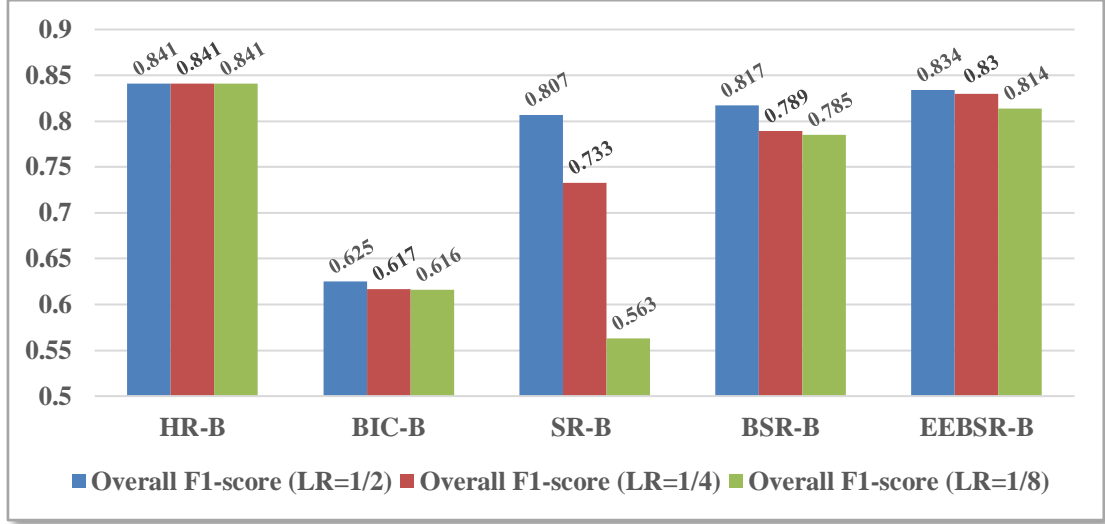
Framework	Overall F1-score (LR=1/2, 1/4, 1/8)
HR-B	0.841
BIC-B	0.625 / 0.617 / 0.616
SR-B	0.807 / 0.733 / 0.563
BSR-B	0.817 / 0.789 / 0.785
<b>EEBSR-B</b>	<b>0.834 / 0.830 / 0.814</b>

**Table 4.1 Overall F1-score obtained by different frameworks on the test set of the tsunami-dataset**

The results show that the BSR-B framework has a higher Overall F1-score than the SR-B framework, which demonstrates that the pre-trained Seg module can supervise the SR module to generate SR images that are more in line with its requirement.

At the same time, EEBSR-B outperforms all other frameworks except HR-B in terms of Overall F1-score results, as can be seen from Table 4.2, where EEBSR-B's Overall F1-

score is substantially ahead of BIC-B for  $LR = 1/2$ ,  $1/4$  and  $1/8$  times resolution, and reaches a level close to that of HR-B.



**Table 4.2 Overall F1-score obtained by different frameworks on the test set of the tsunami-dataset (Histogram)**

#### 4.1.2 PSNR/SSIM results

In this section, we compare the PSNR/SSIM metrics of the SR module output results for our three proposed frameworks and the BIC-B framework at LR sizes of  $1/2$ ,  $1/4$ , and  $1/8$ , as shown in Table 4.3.

Although EEBSR-B outperforms other frameworks other than HR-B in terms of Overall F1-score, the PSNR and SSIM of the images generated by its SR module of EEBSR-B do not always outperform the two-stage training frameworks of SR-B or BSR-B. The reason can explain this is that End-to-End training enables the SR module of EEBSR-B to be optimized to obtain more accurate BDD results (Overall F1-score) rather than

obtaining SR images that are closer to HR images in terms of PSNR and SSIM benchmarks.

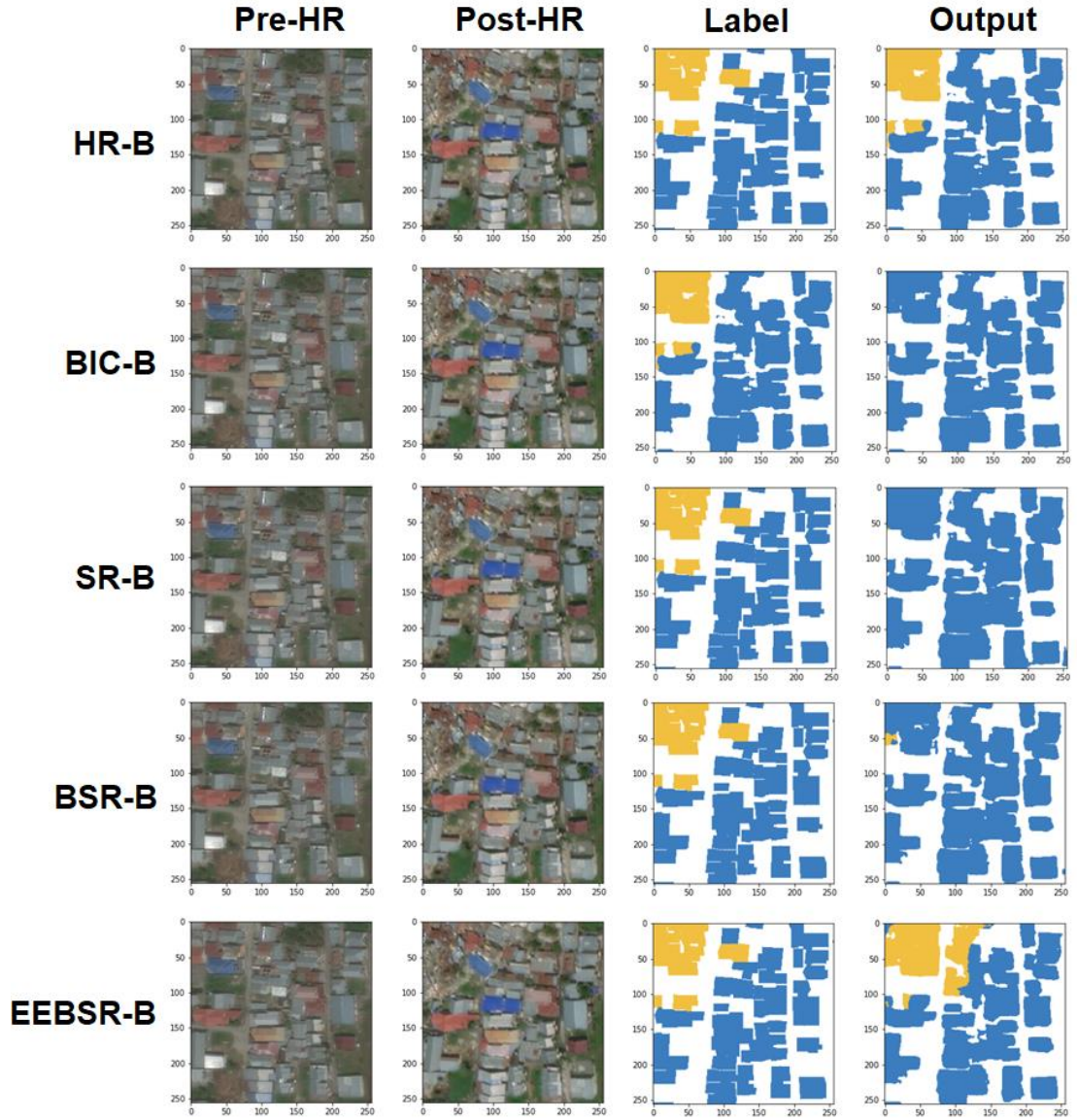
Framework	PSNR (LR=1/2, 1/4, 1/8)	SSIM (LR=1/2, 1/4, 1/8)
HR-B	$\infty$	1
BIC-B	32.63/30.99/27.17	0.868/0.818/0.705
SR-B	<b>36.37</b> /29.33/25.04	0.938/ <b>0.796</b> /0.567
BSR-B	31.83/ <b>29.81</b> /26.65	0.918/0.795/ <b>0.679</b>
<b>EEBSR-B</b>	36.22/28.13/ <b>26.67</b>	<b>0.939</b> /0.700/0.672

**Table 4.3 PSNR and SSIM obtained by different frameworks on the test set of the tsunami-dataset**

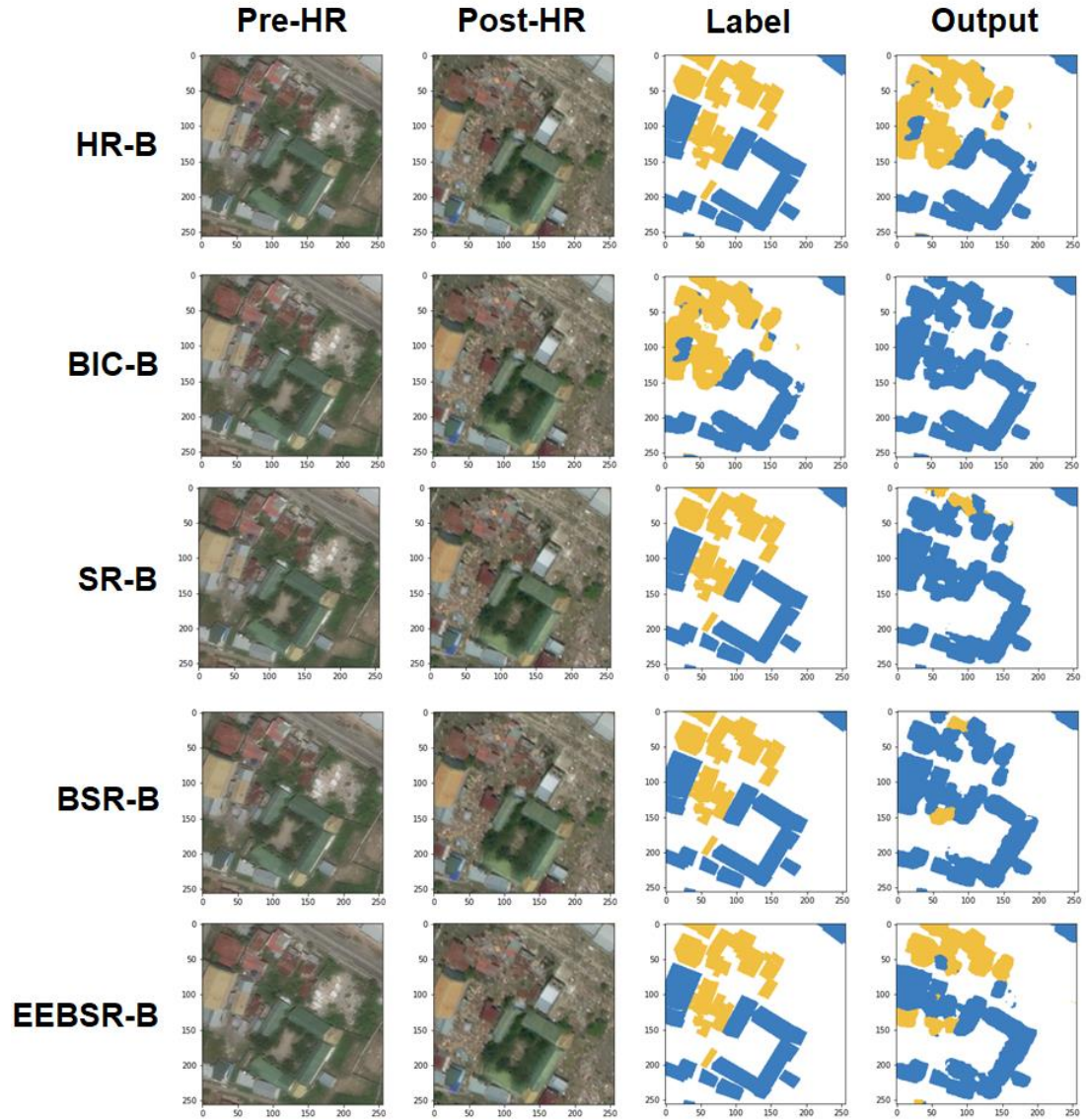
### 4.1.3 Evaluation of the output of Seg modules

In this section, we perform a subjective evaluation of representative images in the results of the Seg module output for our proposed three frameworks and two control group frameworks at LR size of 1/4, as shown in Figures 4.1 and 4.2.

By observing the output results of the Seg module, we found that the BIC-B framework is very difficult to determine the damage to the building, although it can accurately locate the building location, while the SR-B and BSR -B frameworks, although more accurate than BIC-B in determining the damage of buildings, still determine many damaged or destroyed buildings as undamaged buildings.

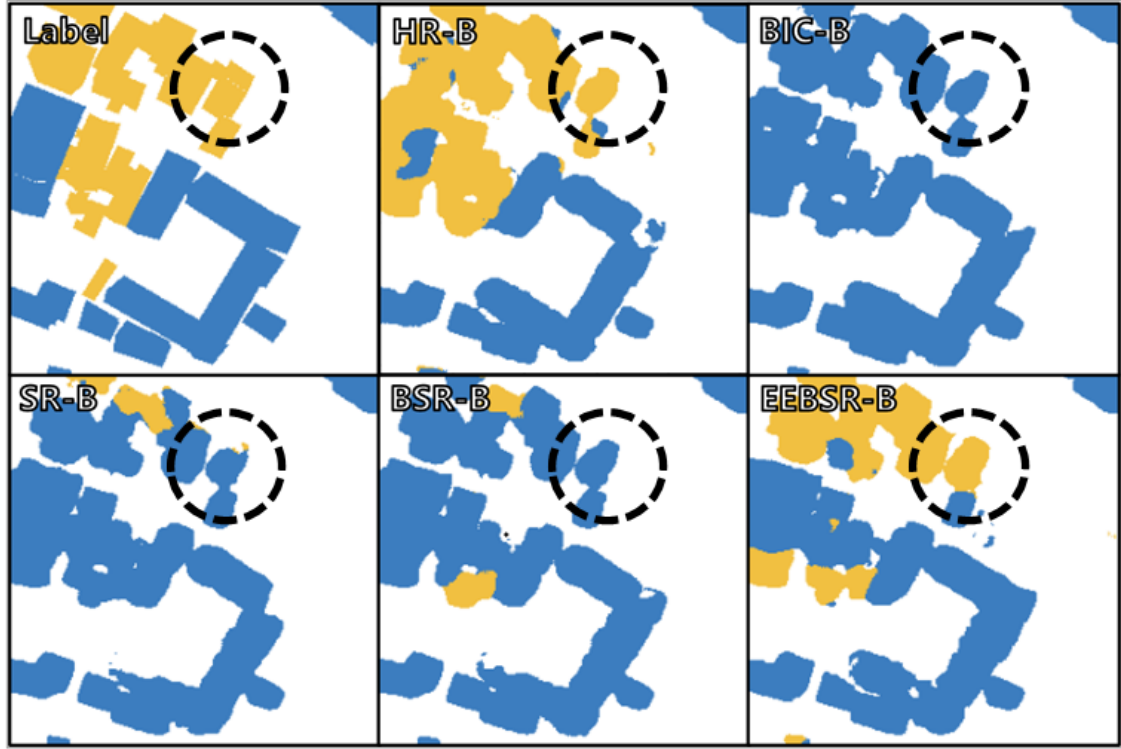


**Figure 4.1 Results of Seg module output for all frameworks (Example 1, LR =1/4)**



**Figure 4.2 Results of Seg module output for all frameworks (Example 2, LR =1/4)**

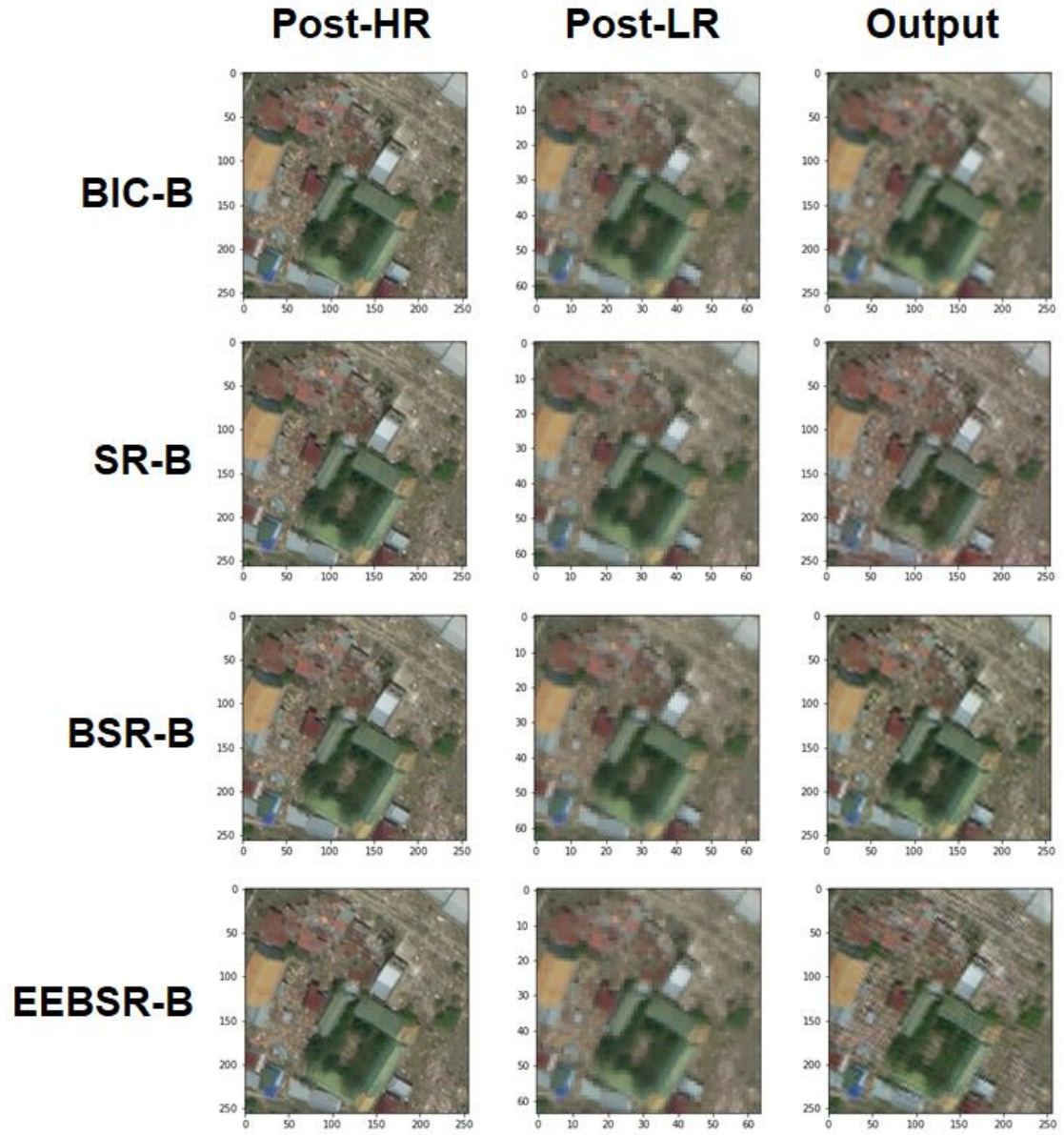
In contrast, the EEBSR-B framework achieves almost the same performance as HR-B in determining the damaged buildings, as shown in Figure 4.3, where only the EEBSR-B and HR-B frameworks can make correct determinations for the damaged buildings in the circles. The reasons why the EEBSR-B framework achieves such a significant advantage will be discussed in the subsequent section.



**Figure 4.3** Semantic segmentation results of damaged buildings in circles for Seg modules of five different frameworks (images from Example 2, LR = 1/4)

#### 4.1.4 Evaluation of the output of SR modules

In this section, we perform a subjective evaluation of representative images from the output of the SR module for our proposed three frameworks and BIC-B framework at LR size of 1/4, as shown in Figure 4.4.



**Figure 4.4** The upsampling result in the BIC-B framework and the output of the SR module of the SR-B, BSR-B and EEBSR-B frameworks (LR = 1/4)

By observing the output results of the SR modules of different frameworks, we found that the SR module of BSR-B framework seems to generate images that are visually closest to the HR images, while the images generated by the SR module of SR-B framework do not seem to be as rich in detail as those of BSR-B.

The images generated by the SR module of the EEBSR-B framework have some unnatural textures and are visually less similar to the HR images, but when the images are compared after zooming in (as in Figure 4.5), we find that the images generated by EEBSR-B seem to better reproduce the building boundaries, which may be one of the reasons for its advantage in the Overall F1-score.



**Figure 4.5** The upsampling result in the BIC-B framework and the output of the SR module of the SR-B, BSR-B and EEBSR-B frameworks (zoomed in, LR = 1/4)

## **4.2 A discussion on the advantage of the frameworks**

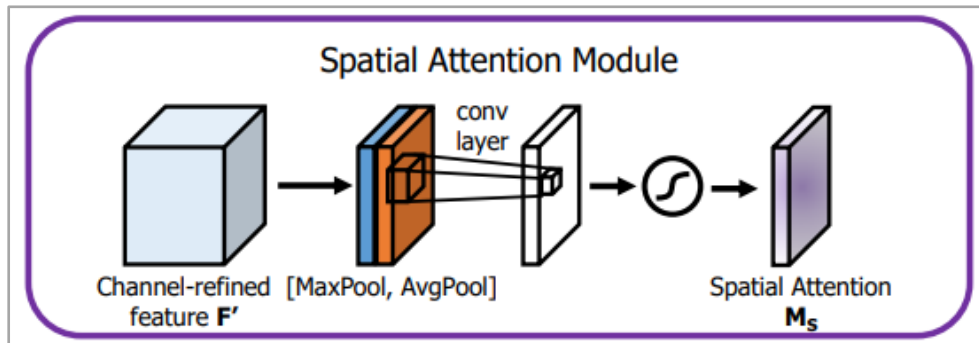
### **4.2.1 Spatial attention mechanism and SAM module**

In Section 4.1.3, we found that BSR-B and EEBSR-B can better determine the damage level of buildings relative to other frameworks, and especially the EEBSR-B framework

has a significant performance advantage.

To explain the phenomenon more objectively, in this section, we try to introduce the spatial attention mechanism into the SR module of the frameworks and observe the focus areas of the SR module of SR-B, BSR-B, and EEBSR-B frameworks by outputting attention maps.

The spatial attention mechanism is an attention mechanism that has been widely used in computer vision in recent years. The core thought is that not all areas in an image are equally important in contributing to the task, and only a part of the task-relevant areas is necessary to be cared for. It is worth noting that the spatial attention mechanism does not force the model to focus on areas not originally of interest but rather induces the model to pay more attention to areas already of interest.

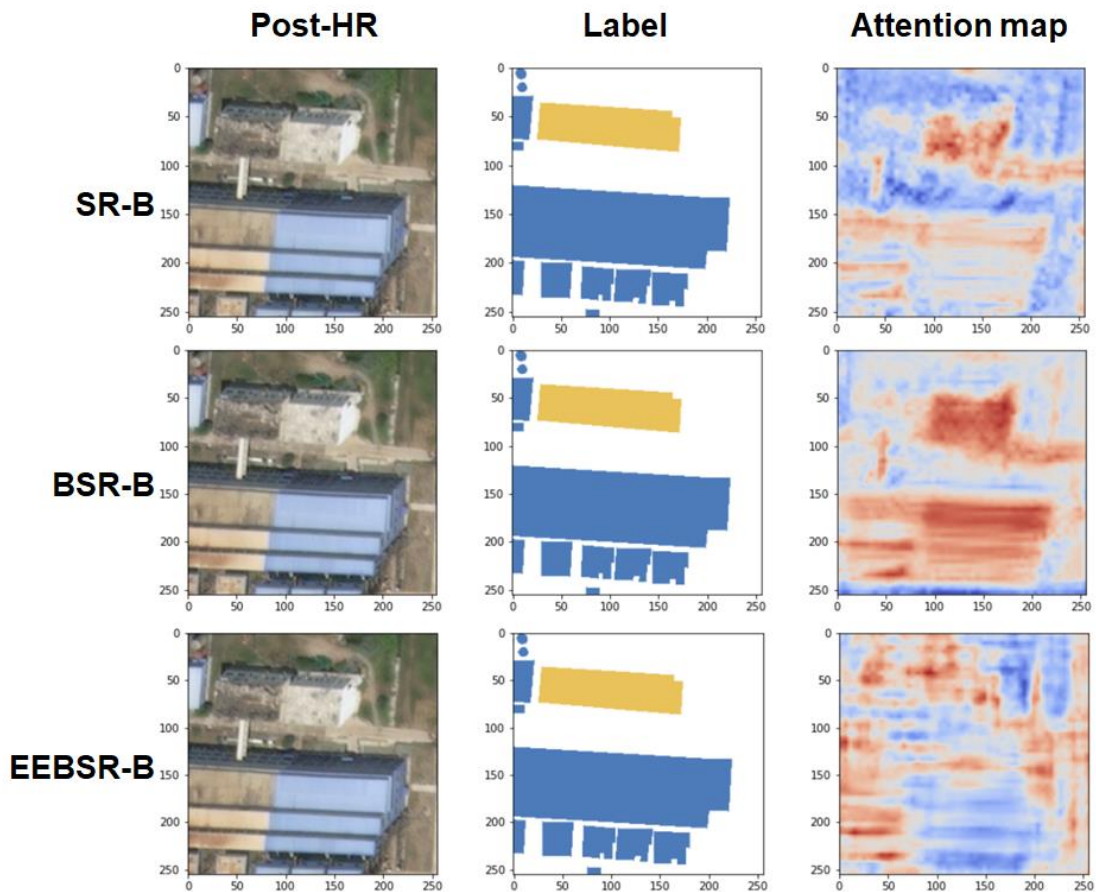


**Figure 4.6 Structure of SAM in CBAM (Woo et al., 2018)**

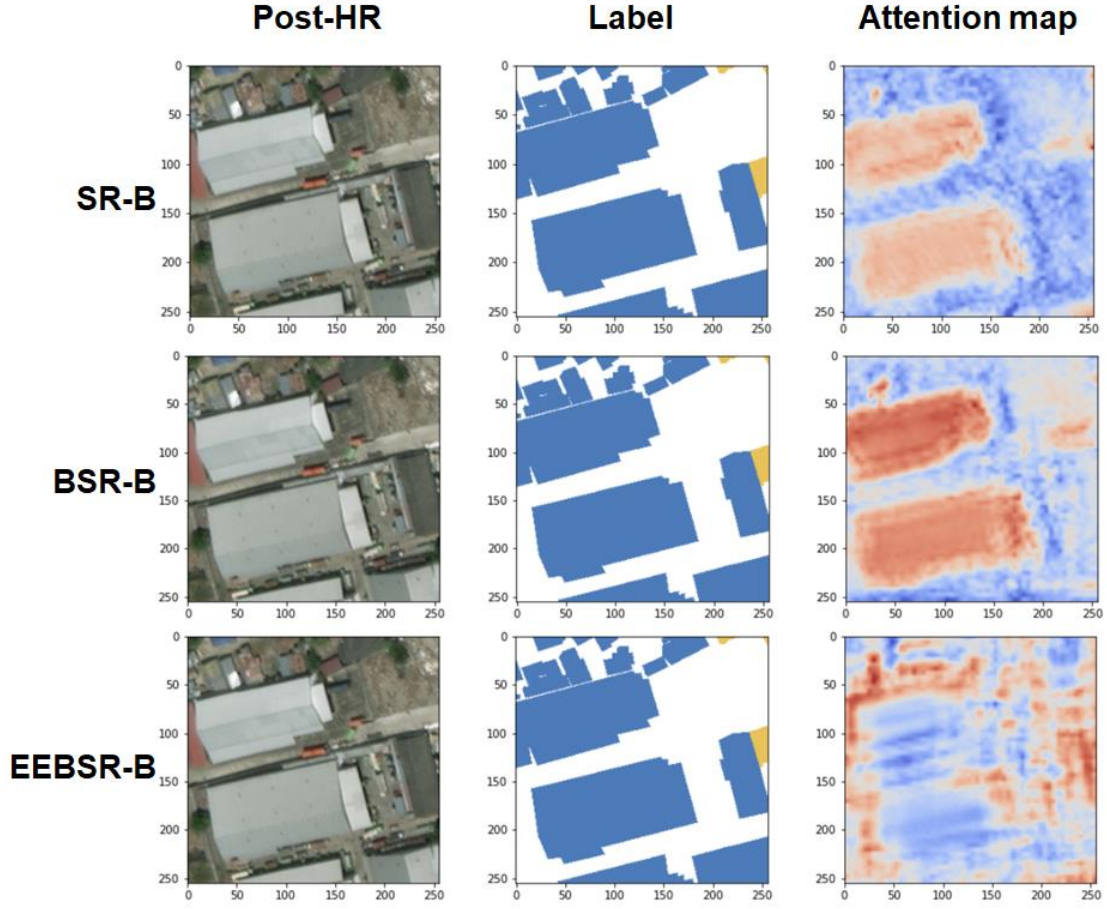
In this research, we chose SAM (Spatial Attention Module) from CBAM (Convolutional Block Attention Module), which was proposed in 2018, as the spatial attention module (the structure is shown in Figure 4.6) and inserted it into SRGAN and BDD-SRGAN in all five ResBlocks, which makes the output of attention maps possible.

### 4.2.2 Analysis of attentional maps

We trained the SR-B, BSR-B, and EEBSR-B frameworks with the SAM module inserted on the tsunami-dataset in the same configuration as in Section 3.6.1 and calculated and visualized the average of the attention maps from all five SAMs in SRGAN or BDD-SRGAN during testing. Examples of the more representative results of the attention maps are shown in Figures 4.7 and 4.8.



**Figure 4.7 Results of the attention maps output for the SR module of the SR-B, BSR-B and EEBSR-B frameworks (Example 1, LR = 1/4)**



**Figure 4.8 Results of the attention maps output for the SR module of the SR-B, BSR-B and EEBSR-B frameworks (Example 2, LR = 1/4)**

By observing the attention maps output by the SR modules of the three frameworks, we found that the SR module of SR-B pays a certain degree of attention to the building area. In contrast, the SR module of BSR-B pays more attention to the building area compared to SR-B. However, the case of EEBSR-B is more different from SR-B and BSR-B, with its SR module paying more attention to the boundary of the building than to the building itself, which is especially significant in the case of larger buildings in the image.

Based on the above experimental results, we believe that the advantage of the EEBSR-B framework is that its SR module focuses more on rebuilding the boundary of the

building rather than the building itself, which improves the performance of the BDD framework.

## **4.3 A discussion on the generalizability of the EEBSR-B framework**

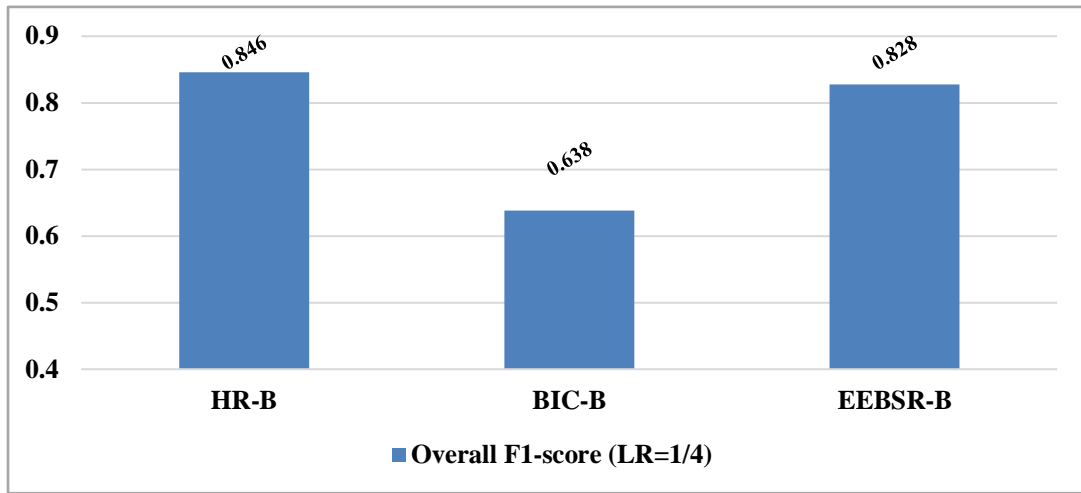
### **4.3.1 Wildfire and hurricane datasets**

Although the EEBSR-B framework achieves a significant advantage on the tsunami-dataset, whether the framework can achieve the same advantage on other types of natural disaster datasets is an important part of validating its application potential.

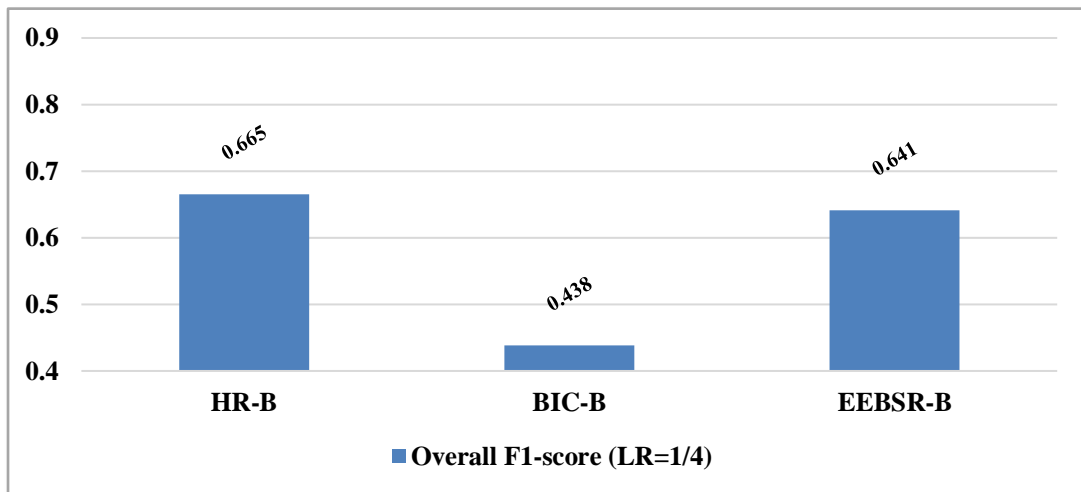
To validate the generalizability of the EEBSR-B framework, we selected two other types of natural disasters from the xBD: the 2018 California wildfires and the 2016 Category 5 Hurricane Matthew and created datasets with the same data processing methods as in Sections 2.21 - 2.24, which are called the woolsey-fire-dataset and the hurricane-matthew-dataset.

### **4.3.2 Results on wildfire and hurricane datasets**

We trained the HR-B, BIC-B and EEBSR-B frameworks with LR of 1/4 on the two datasets described above with the same training configuration as in Section 3.6.1 to verify whether EEBSR-B can keep its advantage over BIC-B. The results are shown in Tables 4.4 and 4.5.



**Table 4.4 Overall F1-score obtained by HR-B, BIC-B and EEBSR-B frameworks on the test set of woolsey-fire-dataset (LR = 1/4)**



**Table 4.5 Overall F1-score obtained by HR-B, BIC-B and EEBSR-B frameworks on the test set of hurricane-matthew-dataset (LR = 1/4)**

The complexity of woolsey-fire-dataset is similar to tsunami-dataset, and also has similar sample proportions. The above experimental results show that the EEBSR-B framework still obtains an Overall F1-score close to HR-B on woolsey-fire-dataset, significantly outperforming the BIC-B framework.

The hurricane-matthew-dataset, on the other hand, is significantly more complex than the tsunami-dataset due to the inclusion of images from multiple Caribbean countries. In this case, the above experimental results show that the EEBSR-B framework still obtains a substantial lead over the BIC-B framework in the Overall F1-score on the hurricane-matthew-dataset.

Based on the above experimental results, we believe that the EEBSR-B framework has a certain degree of generalizability that enables it to significantly outperform the BIC-B framework in different kinds of disaster scenarios, demonstrating its application potential.

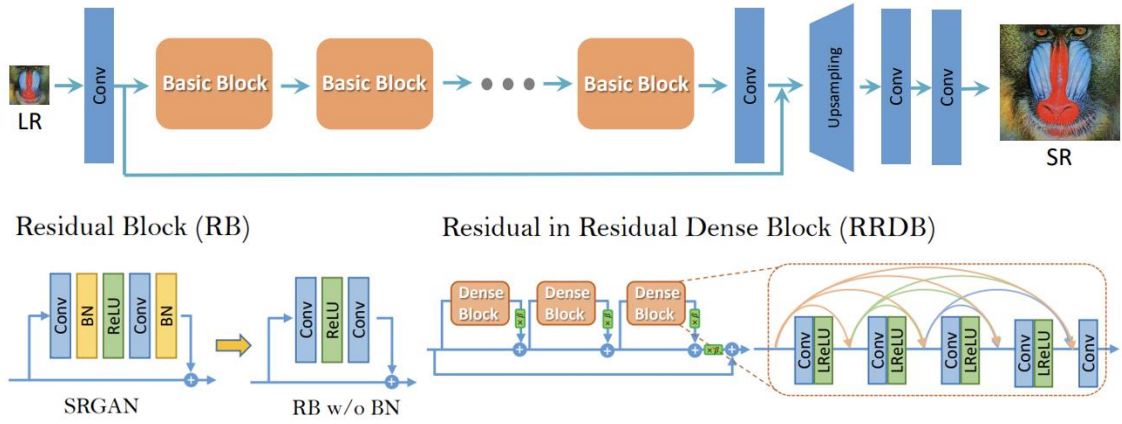
## **4.4 A discussion on the expandability of the EEBSR-B framework**

### **4.4.1 ESRGAN-based EEBSR-B framework**

After verifying the generalizability of the EEBSR-B framework, its extensibility is also a critical issue. In this research, we would like to believe that the advantages of the EEBSR-B framework are derived from the improved logic of BDD loss, and the training logic of the End-to-End training, rather than the specific advantages obtained under the current structure of this particular framework of EEBSR-B.

Therefore, to verify the extensibility of the EEBSR-B framework, in this section we choose ESRGAN, a modified model of SRGAN that was proposed in 2018, instead of SRGAN as the basic structure of the SR module of the EEBSR-B framework and compare it with the EEBSR-B framework that uses SRGAN as the basic structure of the

SR module. The generator and discriminator in the SR module with ESRGAN as the basic structure use the same structure as in the ESRGAN paper, while its loss function also uses the calculation method and coefficients in the ESRGAN paper, but the VGG loss is also replaced with BDD loss, and the location of the feature map used to calculate the BDD loss is the same as that used in BDD-SRGAN.

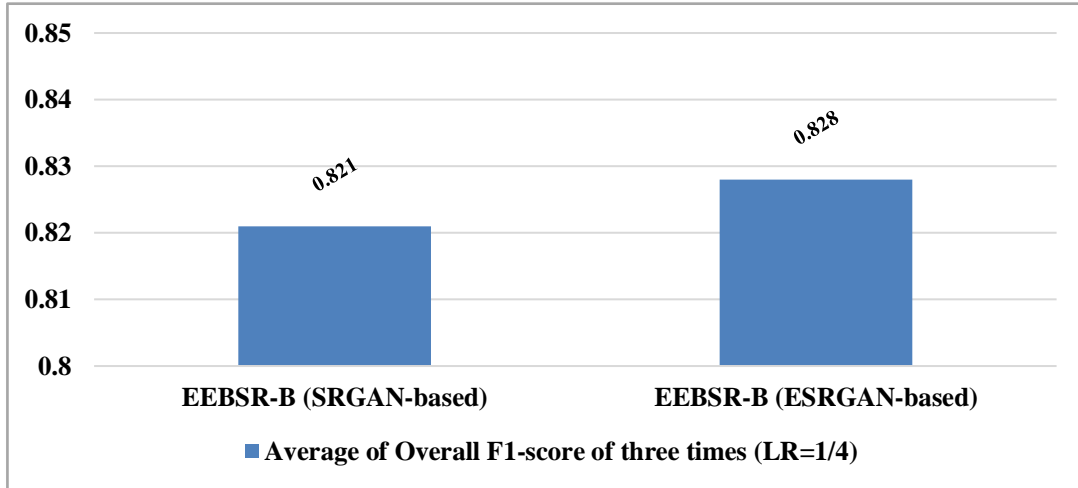


**Figure 4.9 Structure of ESRGAN's generator (Wang et al., 2018)**

#### 4.4.2 Results of EEBSR-B based on SRGAN or ESRGAN

Since the ESRGAN-based EEBSR-B framework requires a large amount of computational resources and longer time, in order to more objectively verify the expandability of EEBSR-B, we changed the number of epochs from 128 to 64 based on the training configuration in Section 3.6.1, and trained the SRGAN- and ESRGAN-based EEBSR-B frameworks three times with LR of 1/4 on tsunami-dataset with the above training configuration. The average of three trainings of the Overall F1- score results are shown in Table 4.6.

The above results show that the ESRGAN-based EEBSR-B framework slightly outperforms the SRGAN-based EEBSR-B framework. Although the difference between these two is relatively small (less than 0.01), this result indicates the advantage of the ESRGAN-based EEBSR-B framework in the case of three replicated experiments. However, since the difference is small, we can say that SRGAN-based EEBSR-B framework is sufficiently accurate without huge computational cost which is required for ESRGAN-based EEBSR-B. Depending on computational resource, we may select the more optimum one.



**Table 4.6 Overall F1-score obtained by the EEBSR-B framework based on SRGAN and ESRGAN on the test set of tsunami-dataset (LR = 1/4)**

Based on the above experimental results, we have reasons to believe that the advantages of the EEBSR-B framework come from the right direction of our design of the loss function and training process, so we believe that the framework has high scalability, which means that the performance of the framework can be improved by replacing the SR module and Seg module with better performance.

## 5. Conclusions

In this research, we have attempt to develop a high-performance, time-efficient building damage detection framework that uses post-disaster low-resolution images taken by nanosatellites and pre-disaster high-resolution images as inputs to address the issue of the lack of timeliness of traditional building damage detection frameworks caused by the untimely acquisition of high-resolution remote sensing images within 24 hours after a large-scale natural disaster due to satellite revisit times and weather conditions.

To construct the above framework, we first designed an overall structure of a building damage detection framework with pre-disaster high-resolution images and post-disaster low-resolution images as inputs and designed an SRGAN-based super-resolution model and a pre-disaster/post-disaster double-modal semantic segmentation model based on U-Net for the above global structure. For the above overall structure, we designed an SRGAN-based super-resolution model as the super-resolution module in the framework and a U-Net-based pre/post-disaster double-modal semantic segmentation model as the semantic segmentation module in the framework.

Based on the above overall structure and modules, we compared three BDD frameworks with different structures and training processes, including two two-step training frameworks and one End-to-End training framework, and finally verified that the EEBSR-B framework has significant performance advantages over other frameworks. With low-resolution images at 1/2, 1/4, and 1/8 times the resolution, the performance of the EEBSR-B framework is all close to that of the conventional BDD framework using high-resolution pre/post-disaster images.

We also analyze the reason for the advantage of the EEBSR-B framework, as well as the generalizability and expandability of the framework, through a series of small-scale experiments. Based on the experimental results, we believe that the super-resolution module in EEBSR-B focuses more on the generation of building boundaries, which improves the overall performance of the framework, while the framework has a certain degree of generalizability and expandability.

This research confirms that obtaining high-accuracy building damage detection results using only post-disaster low-resolution images taken by nanosatellites and pre-disaster high-resolution images is practically feasible and potentially useful for future post-disaster response efforts. We have reasons to believe that with the rapid development of deep learning super-resolution techniques and nanosatellite technology, designing nanosatellite data-based BDD frameworks with better performance in terms of accuracy and timeliness is of significant importance and is an important topic for future works.

## Reference

- [1] Tan, Ling, et al. "Can we detect trends in natural disaster management with artificial intelligence? A review of modeling practices," *Natural Hazards* 107.3, 2389-2417, 2021.
  
- [2] Bai, Yanbing, et al. "A framework of rapid regional tsunami damage recognition from post-event TerraSAR-X imagery using deep neural networks." *IEEE Geoscience and Remote Sensing Letters* 15.1 (2017): 43-47.
  
- [3] Nex, Francesco, et al. "Structural building damage detection with deep learning: Assessment of a state-of-the-art CNN in operational conditions." *Remote sensing* 11.23 (2019): 2765.
  
- [4] Hao, Hanxiang, et al. "An attention-based system for damage assessment using satellite imagery." 2021 *IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021.
  
- [5] Ermlick, William, et al. "Natural Disaster Building Damage Assessment Using a Two-Encoder U-Net." *International Symposium on Visual Computing*. Springer, Cham, 2020.
  
- [6] Wu, Chuyi, et al. "Building damage detection using U-Net with attention mechanism from pre-and post-disaster remote sensing datasets." *Remote Sensing* 13.5 (2021): 905.
  
- [7] Bai, Yanbing, et al. "Pyramid pooling module-based semi-siamese network: A

benchmark model for assessing building damage from xBD satellite imagery datasets." *Remote Sensing* 12.24 (2020): 4055.

[8] Ming, D. P., Qun Wang, and J. Y. Yang. "Spatial scale of remote sensing image and selection of optimal spatial resolution," *Journal of remote sensing* 12.4, 529-537, 2008.

[9] Totani, Tsuyoshi, et al. "Thermal design procedure for micro-and nanosatellites pointing to earth." *Journal of Thermophysics and Heat Transfer* 28.3 (2014): 524-533.

[10] Totani, Tsuyoshi, et al. "Verification of Rapid Thermal Design Approach Using Design and Flight Data of Hodoyoshi-1 Microsatellite." *46th International Conference on Environmental Systems*, 2016.

[11] Courtrai, Luc, Minh-Tan Pham, and Sébastien Lefèvre. "Small object detection in remote sensing images based on super-resolution with auxiliary generative adversarial networks," *Remote Sensing* 12.19, 3152, 2020.

[12] Zhang, Lixian, et al. "Making low-resolution satellite images reborn: a deep learning approach for super-resolution building extraction," *Remote Sensing* 13.15, 2872, 2021.

[13] Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4681-4690, 2017.

- [14] Gupta, Ritwik, et al. "xbd: A dataset for assessing building damage from satellite imagery," arXiv preprint arXiv:1911.09296, 2019.
- [15] García, Salvador, and Francisco Herrera. "Evolutionary undersampling for classification with imbalanced datasets: Proposals and taxonomy." *Evolutionary computation* 17.3 (2009): 275-306.
- [16] Tan, Mingxing, and Quoc Le. "Efficientnetv2: Smaller models and faster training." *International Conference on Machine Learning*. PMLR, 2021.
- [17] Tao, Xin, et al. "Detail-revealing deep video super-resolution." *Proceedings of the IEEE International Conference on Computer Vision*. 2017.
- [18] Li, Peijun, Haiqing Xu, and Jiancong Guo. "Urban building damage detection from very high resolution imagery using OCSVM and spatial features." *International Journal of Remote Sensing* 31.13 (2010): 3393-3409.
- [19] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.
- [20] Santurkar, Shibani, et al. "How does batch normalization help optimization?." *Advances in neural information processing systems* 31 (2018).
- [21] Dong, Weisheng, et al. "Nonlocally centralized sparse representation for image

restoration,” IEEE transactions on Image Processing 22.4, 1620-1630, 2012.

[22] Ghiasi, Golnaz, Tsung-Yi Lin, and Quoc V. Le. “Dropblock: A regularization method for convolutional networks,” arXiv preprint arXiv:1810.12890, 2018.

[23] Müller, Markus U., et al. "Super-resolution of multispectral satellite images using convolutional neural networks." arXiv preprint arXiv:2002.00580 (2020).

[23] Ahmed, Tashin, and Noor Hossain Nuri Sabab. "Classification and understanding of cloud structures via satellite images with EfficientUNet." SN Computer Science 3.1 (2022): 1-11.

[25] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017.

[26] Kim, Byeongjoon, et al. "A performance comparison of convolutional neural network-based image denoising methods: The effect of loss functions on low-dose CT images." Medical physics 46.9 (2019): 3906-3923.

[27] Ondruska, Peter, et al. "End-to-end tracking and semantic segmentation using recurrent neural networks." arXiv preprint arXiv:1604.05091 (2016).

[28] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems 27 (2014).

- [29] Huynh-Thu, Quan, and Mohammed Ghanbari. "Scope of validity of PSNR in image/video quality assessment." *Electronics letters* 44.13 (2008): 800-801.
- [30] Hore, Alain, and Djemel Ziou. "Image quality metrics: PSNR vs. SSIM." 2010 20th international conference on pattern recognition. IEEE, 2010.
- [31] Yacouby, Reda, and Dustin Axman. "Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models." *Proceedings of the first workshop on evaluation and comparison of NLP systems*. 2020.
- [32] Chu, Qi, et al. "Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism." *Proceedings of the IEEE international conference on computer vision*. 2017.
- [33] Woo, Sanghyun, et al. "Cbam: Convolutional block attention module." *Proceedings of the European conference on computer vision (ECCV)*. 2018.
- [34] Chen, Hao, and Zhenwei Shi. "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection." *Remote Sensing* 12.10 (2020): 1662.
- [35] Spialek, Matthew L., J. Brian Houston, and Kyle C. Worley. "Disaster communication, posttraumatic stress, and posttraumatic growth following Hurricane Matthew." *Journal of health communication* 24.1 (2019): 65-74.

[36] Wang, Xintao, et al. "Esrgan: Enhanced super-resolution generative adversarial networks." Proceedings of the European conference on computer vision (ECCV) workshops. 2018.