



Article

AAU-Net: Attention-Based Asymmetric U-Net for Subject-Sensitive Hashing of Remote Sensing Images

Kaimeng Ding ^{1,2} , Shiping Chen ³ , Yu Wang ⁴, Yueming Liu ², Yue Zeng ^{1,*} and Jin Tian ¹

¹ Jinling Institute of Technology, Nanjing 211169, China; dkm@jit.edu.cn (K.D.); jim.tian@jit.edu.cn (J.T.)

² State Key Laboratory of Resource and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; liuym@lreis.ac.cn

³ CSIRO Data61, Sydney, NSW 1710, Australia; Shiping.Chen@data61.csiro.au

⁴ Changjiang Nanjing Waterway Bureau, Nanjing 210011, China; chc@cjinhdj.com

* Correspondence: zengy@jit.edu.cn

Abstract: The prerequisite for the use of remote sensing images is that their security must be guaranteed. As a special subset of perceptual hashing, subject-sensitive hashing overcomes the shortcomings of the existing perceptual hashing that cannot distinguish between “subject-related tampering” and “subject-unrelated tampering” of remote sensing images. However, the existing subject-sensitive hashing still has a large deficiency in robustness. In this paper, we propose a novel attention-based asymmetric U-Net (AAU-Net) for the subject-sensitive hashing of remote sensing (RS) images. Our AAU-Net demonstrates obvious asymmetric structure characteristics, which is important to improve the robustness of features by combining the attention mechanism and the characteristics of subject-sensitive hashing. On the basis of AAU-Net, a subject-sensitive hashing algorithm is developed to integrate the features of various bands of RS images. Our experimental results show that our AAU-Net-based subject-sensitive hashing algorithm is more robust than the existing deep learning models such as Attention U-Net and MUM-Net, and its tampering sensitivity remains at the same level as that of Attention U-Net and MUM-Net.

Keywords: security of remote sensing images; deep learning; subject-sensitive hashing; integrity authentication; perceptual hash; U-Net



Citation: Ding, K.; Chen, S.; Wang, Y.; Liu, Y.; Zeng, Y.; Tian, J. AAU-Net: Attention-Based Asymmetric U-Net for Subject-Sensitive Hashing of Remote Sensing Images. *Remote Sens.* **2021**, *13*, 5109. <https://doi.org/10.3390/rs13245109>

Academic Editors: Amir Hussain, Ahmed Al-Dubai, William (Bill) J Buchanan, Jonathan Wu, Kaizhu Huang, Bin Luo, Jin Tang, Wadli Boulila and Adel M. Alimi

Received: 13 October 2021

Accepted: 13 December 2021

Published: 16 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the development of Earth observation (EO) technology, remote sensing (RS) images have been widely used in many fields, such as environmental science [1,2], urban planning [3,4], disaster monitoring [5,6], agriculture [7,8], and surveying and mapping [9,10]. However, the use of RS images has an implicit premise, i.e., the security of the RS image must be guaranteed. If an RS image is tampered with during transmission and storage, the content of the RS image will change or may even be distorted. If a user uses a tampered RS image, his/her analysis results obtained from the tampered RS image would either be not accurate enough, or incorrect, both of which may lead to a wrong and dangerous decision in some applications; if the user is not sure whether the RS image has been tampered with or not, the value of that image can be greatly reduced, or it even becomes useless.

A set of examples of tampered RS images is shown in Figure 1. For tampered RS images, as shown in Figure 1b, it is difficult to determine whether they have been tampered with, even if they are compared with the original images, as shown in Figure 1a. Furthermore, it is also difficult to compare the “suspicious” RS images with the original RS images. Therefore, the data integrity of RS images must be guaranteed so that the RS images can be trusted and used with peace of mind. To address the above problem, some authentication technologies have been developed to ensure the integrity of RS images, including cryptography, digital watermarking, block chains, and perceptual hashing.

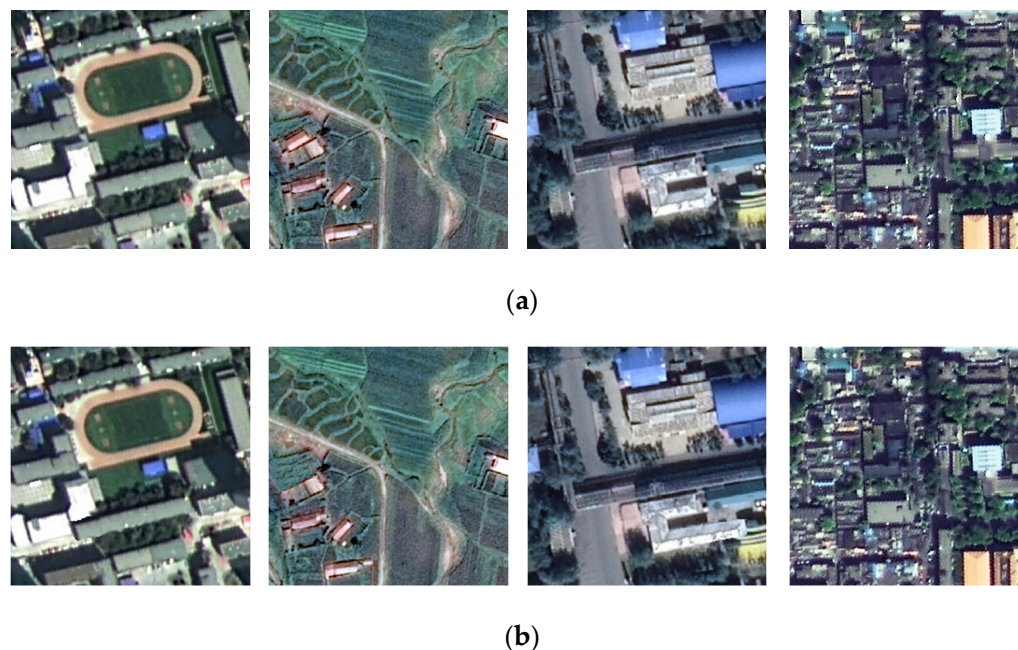


Figure 1. Examples of content tampering of RS images: (a) original RS image; (b) tampered RS images.

Among the existing authentication technologies, cryptography technology authenticates data at the binary level and is too sensitive to changes in data. Therefore, cryptography technology has many shortcomings in the authentication of RS images. For example, after the format conversion of remote sensing images, the binary changes greatly, but its content does not change. Digital watermarking technology requires certain modifications to the RS image, which is not allowed in many cases. Moreover, digital watermarking, similarly to cryptography, cannot identify whether the content of the RS image has changed. Block chain technology has strong application prospects in RS image authentication, but the block chain relies on cryptography technology in the realization process, which inherits the deficiencies of cryptography technology in RS image authentication. Perceptual hash [11,12] overcomes the shortcomings of cryptography technologies and digital watermarking technology and has strong application prospects. Some scholars have carried out related research on perceptual hash authentication algorithms of RS images, leading to some encouraging results [13–15].

As the spatial resolution of RS images becomes higher and higher, the performance of the existing perceptual hash algorithms gradually encounters a bottleneck. For example, the robustness of existing perceptual hashing algorithms to high-resolution RS images needs to be further strengthened. At the same time, as RS images are used in more fields, new security issues have arisen. For example, users pay different levels of attention to different contents of RS images in different applications: hydrologists tend to pay more attention to the information of rivers and lakes in the images; survey workers pay more attention to the roads, bridges, and buildings; and users who study the extraction of data related to moving objects are more interested in the features of moving objects, such as airplanes, ships, and vehicles. However, the current mainstream perceptual hash technology cannot achieve “subject-sensitive” integrity authentication. As a branch of perceptual hash, subject-sensitive hashing [16] can enable “subject-sensitive” integrity authentication. However, the existing subject-sensitive hashing algorithms have the following shortcomings:

1. The differences in the various bands of RS images are not taken into account. Each band of the RS image reflects the spectral information within the band. As a result, there are certain differences among different bands. However, the current subject-sensitive hashing algorithms do not make the best use of these differences [16].
2. The tampering sensitivity of the existing subject-sensitive hash still needs further improvement. Compared with perceptual hash, subject-sensitive hashing further

distinguishes robustness into “subject-related tampering” and “subject-unrelated tampering”. However, the existing subject-sensitive hash cannot sufficiently distinguish the above two types of tampering sensitivities.

3. While perceptual hashing methods (including subject-sensitive hashing) have better robustness than cryptographic methods, the robustness of existing subject-sensitive hashing algorithms still has room for improvement. For example, the robustness of the existing methods for the lossy data compression of RS images is still not ideal.

To address the above problems, taking advantage of the characteristics of the attention mechanism, in that it can focus attention on the important features of the image [17,18], we constructed an attention-based asymmetric U-Net (AAU-Net) for the extraction of subject-sensitive features of RS images. The attention mechanism was first successfully applied in the field of natural language processing [19] and later successfully applied to image processing and other fields [20,21]. According to the characteristics of subject-sensitive hashing that distinguishes “subject-related features” and “subject-unrelated features”, where the extracted features are as robust as possible and the feature image does not need to be the same size as the original image, the AAU-Net adopts an asymmetric network structure, which adds the attention mechanism to the decoder part and multi-scale input to the encoder part.

In this paper, we propose a novel attention-based asymmetric U-Net (AAU-Net) for the subject-sensitive hashing of RS images. Our AAU-Net improves the robustness of features and demonstrates obvious asymmetric structure characteristics by combining the attention mechanism and the characteristics of subject-sensitive hashing. On the basis of AAU-Net, we propose a subject-sensitive hash algorithm based on band feature fusion, taking into account the differences between the characteristics of each band of RS images. In particular, we make the following key contributions:

1. In view of the shortcomings of existing subject-sensitive hashing algorithms, such as the inability to distinguish between subject-related tampering and subject-unrelated tampering, we introduce the attention mechanism into the research of subject-sensitive hashing to better realize the integrity authentication of RS images.
2. Combining the characteristics of subject-sensitive hashing, a network named attention-based asymmetric U-Net (AAU-Net) is proposed to extract the subject-sensitive features of the bands of RS images, which has good tampering sensitivity and robustness.
3. There are certain differences in the content of different bands of RS images, while the existing subject-sensitive hash algorithm does not take the differences into account. To overcome this problem, a subject-sensitive hashing algorithm that integrates the features of each band of RS images is proposed based on our AAU-Net.

The composition of this paper is as follows: Section 2 reviews the related work. Section 3 presents the proposed AAU-Net and subject-sensitive hashing algorithm in detail. Section 4 presents the experiment in detail and provides a discussion. Finally, we conclude this paper in Section 5.

2. Related Work

In a broad sense, “hash” means “one-way mapping”. Perceptual hashing, also known as perceptual hash, is a subset of the generalized “hash”, and it is a one-way mapping from multimedia datasets to digests. Perceptual hash can uniquely map multimedia data with the same perceptual content to a digital digest and meet the requirements of perceptual robustness and safety. According to different multimedia objects, perceptual hash can be subdivided into image perceptual hash, video perceptual hash, and audio perceptual hash. In this paper, perceptual hash generally refers to image perceptual hashing unless otherwise specified.

Suppose the original RS image is denoted as I , the RS image with the same effective content as I but not necessarily the same in binary representation is represented as I' , the

image perceptual hash function is denoted as $PH()$, and the output result of $PH()$ is the perceptual hash sequence denoted as ph ; then,

$$PH(I) = ph_I \quad (1)$$

$$PH(I') = ph_{I'} \quad (2)$$

$$ph_I = ph_{I'} \quad (3)$$

Suppose the function $D()$ is used to measure the difference between two hash sequences, I'' represents an image with a certain different content from the content of I , and T is the threshold; then,

$$D(PH(I), PH(I'')) > T \quad (4)$$

Compared to cryptographic hash, perceptual hash is more suitable for the integrity authentication of RS images: in the process of the distribution, transmission, and use of RS images, the effective content information carried by the RS images is not changed after format conversion, data compression, etc.; only the information carrier is changed. For the integrity authentication of RS images, more attention should be paid to the integrity of the content carried, rather than whether the carrier itself has changed.

Subject-sensitive hashing [16] is a special case of perceptual hashing, which can realize the “subject-sensitive” authentication of RS images; that is, it divides the content tampering of RS images into two types: “subject-related tampering” and “subject-unrelated tampering”.

The relationship between hash, cryptographic hash, and perceptual hash is shown in Figure 2. In the category of generalized hash, although perceptual hash has many similarities with cryptographic hash, the difference is very obvious: cryptographic hash generates a hash sequence based on the binary representation of data, while perceptual hash generates a hash sequence based on the perceptual content of multimedia data.

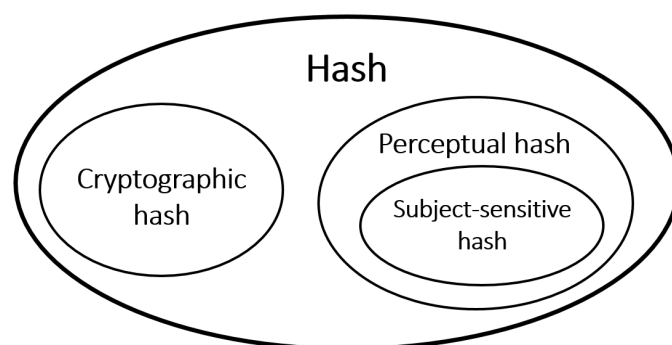


Figure 2. The relationship between hash, cryptographic hash, and perceptual hash.

Here, we take “buildings” as the subject to compare subject-related tampering with subject-unrelated tampering—that is, tampering related to a building is “subject-related tampering”, and tampering that is not related to a “building” is “subject-unrelated tampering”, as shown in Figure 3. The comparison shows that although the size of tampering areas in Figure 3b,c is roughly the same, it is clear that the subject-related tampering shown in Figure 3b is more destructive to RS images, as the tampering in Figure 3b directly affects the user’s analysis results of RS images. However, conventional perceptual hash algorithms cannot achieve this “subject-sensitive” integrity authentication and cannot distinguish between these two types of tampering. Subject-sensitive hashing overcomes this shortcoming of perceptual hashing.



Figure 3. Comparison of subject-related tampering and subject-unrelated tampering: (a) original RS image; (b) subject-related tampered RS image; (c) subject-unrelated tampered RS image.

Compared with traditional perceptual hash, the difficulty of subject-sensitive hashing is to distinguish between “subject-related tampering” and “subject-unrelated tampering”—that is, to achieve “subject-sensitive” integrity authentication of the RS image. However, it is difficult to implement subject-sensitive hashing based on traditional image processing technology: traditional image processing technology struggles to achieve “subject-sensitive” feature extraction, and it also finds it difficult to define “subject-sensitive” features. Fortunately, the rise of deep learning provides a feasible way to achieve subject-sensitive hashing: through the learning of specific samples, deep neural networks have the ability to extract “subject-sensitive” features. Compared with traditional methods, deep learning has excellent feature expression capabilities [22–24] and can extract the hidden high-level features of RS images [25]. Methods based on deep learning can automatically learn more essential features from training samples, reduce the complexity of artificially designed features, and provide a feasible way to achieve “subject-sensitive” authentication.

The framework of the deep learning-based subject-sensitive hashing algorithm is shown in Figure 4: first, on the basis of analyzing the application of RS images, a training sample set is constructed that meets the requirements of integrity authentication; then, a suitable deep neural network is built and trained to extract the subject-sensitive features of RS images; in the process of generating a subject-sensitive hash sequence, the subject-sensitive features of the preprocessed RS image are extracted through the trained deep neural network model, and the extracted features are dimensionally reduced, coded, and enhanced to generate the final subject-sensitive hash sequence.

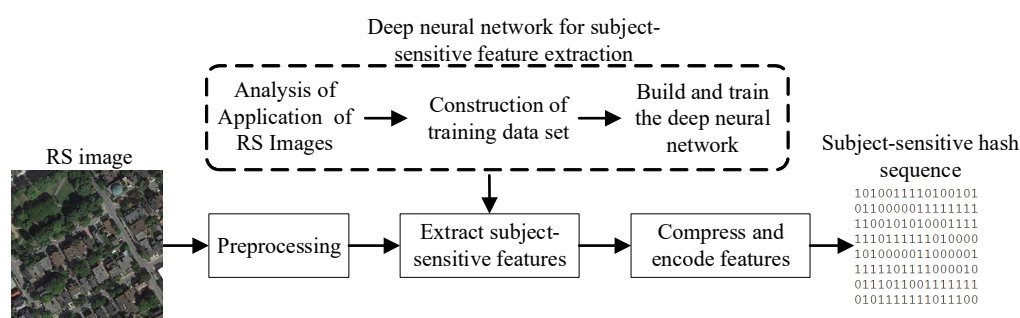


Figure 4. Framework of subject-sensitive hashing algorithm based on deep learning.

The process of constructing the training sample set is generally divided into two steps: the first step is to modify the existing training dataset to generate samples “directly related to the subject”: for example, generating training samples based on the existing building-oriented training dataset; the second step is to manually draw some robust training samples: for example, manually drawing training samples with “false edges” removed. The preprocessing enables the RS image to meet the input requirements of the deep neural network; the subject-sensitive hash sequence can be expressed in a binary or hexadecimal format. Among the above steps, the most critical step is to design a

suitably deep neural network model, which will directly affect the performance of the subject-sensitive hash algorithm.

The information regarding various bands of RS images often has certain differences, but the existing perceptual hash algorithms (including subject-sensitive hashing) for RS images often do not take this difference into account. The existing perceptual hash algorithms for RS images [13–16] generally perform a grayscale operation on the original image in the preprocessing stage and then extract the characteristics of the obtained grayscale image. This obviously does not take into account the band differences of RS images.

On the other hand, affected by the correlation of bands, there is a lot of information redundancy between the various bands of RS images [26–28]. The subject-sensitive hash algorithm used for RS image authentication should have excellent abstraction and computational efficiency, which requires the subject-sensitive hash algorithm to reduce the redundancy between the bands as much as possible while taking into account the differences between each band. To solve the above problems, drawing lessons from the idea of band fusion, we separately extract the features of each band of remote sensing images and then fuse the feature images. Remote sensing image fusion technology can coordinate two or more different wavebands to generate new images with more information [29,30].

Feature extraction based on deep learning can effectively overcome the limitations of traditional RS image processing methods [31] and provide a feasible way to carry out subject-sensitive hashing. In the related methods of deep learning, the attention mechanism enables the network to dynamically select a subset of input attributes in a given input–output pair setting to improve the accuracy of decision making [32].

The attention mechanism has been widely used in tasks such as natural language description [33], machine translation [34], image feature extraction [35,36], and image classification [37,38]. In essence, the attention mechanism is a weight probability distribution mechanism that assigns larger weights to important content and smaller weights to other content. This enables the attention mechanism to focus on finding useful information that is significantly related to the current output in the input data, in order to highlight the features related to the prediction, thereby improving the accuracy of the model's prediction. As the attention mechanism focuses on the features that are of interest to the user, and the features that the user is interested in are the key to the extraction of “subject-sensitive” features and the realization of subject-sensitive hashing, it is feasible to apply the attention mechanism to the task of subject-sensitive hashing.

In this research, we refer to the attention gate (AG) proposed in [39] and combine the characteristics of subject-sensitive hashing to propose a deep neural network named attention-based asymmetric U-Net for subject-sensitive hashing of RS images.

3. Proposed Method

In this section, the proposed AAU-Net is introduced in detail, and then the subject-sensitive hashing algorithm based on AAU-Net is discussed.

3.1. AAU-Net: Attention-Based Asymmetric U-Net

3.1.1. Asymmetric Network Architecture of AAU-Net

Although we are not the first to combine the attention mechanism with U-Net, our AAU-Net is more suitable for the subject-sensitive hashing of RS images. To achieve subject-sensitive authentication, it is not the case that the more information extracted by AAU-Net, the better, but rather the more information related to the subject extracted, the better. The attention gate pays more attention to useful salient features and can learn to suppress irrelevant areas [17,21,39], which is very effective for subject-sensitive feature extraction.

The architecture of the proposed AAU-Net is shown in Figure 5, which is somewhat similar to that of the original U-Net [40], but it is also obviously different. The asymmetry structure of AAU-Net is mainly reflected in the following aspects:

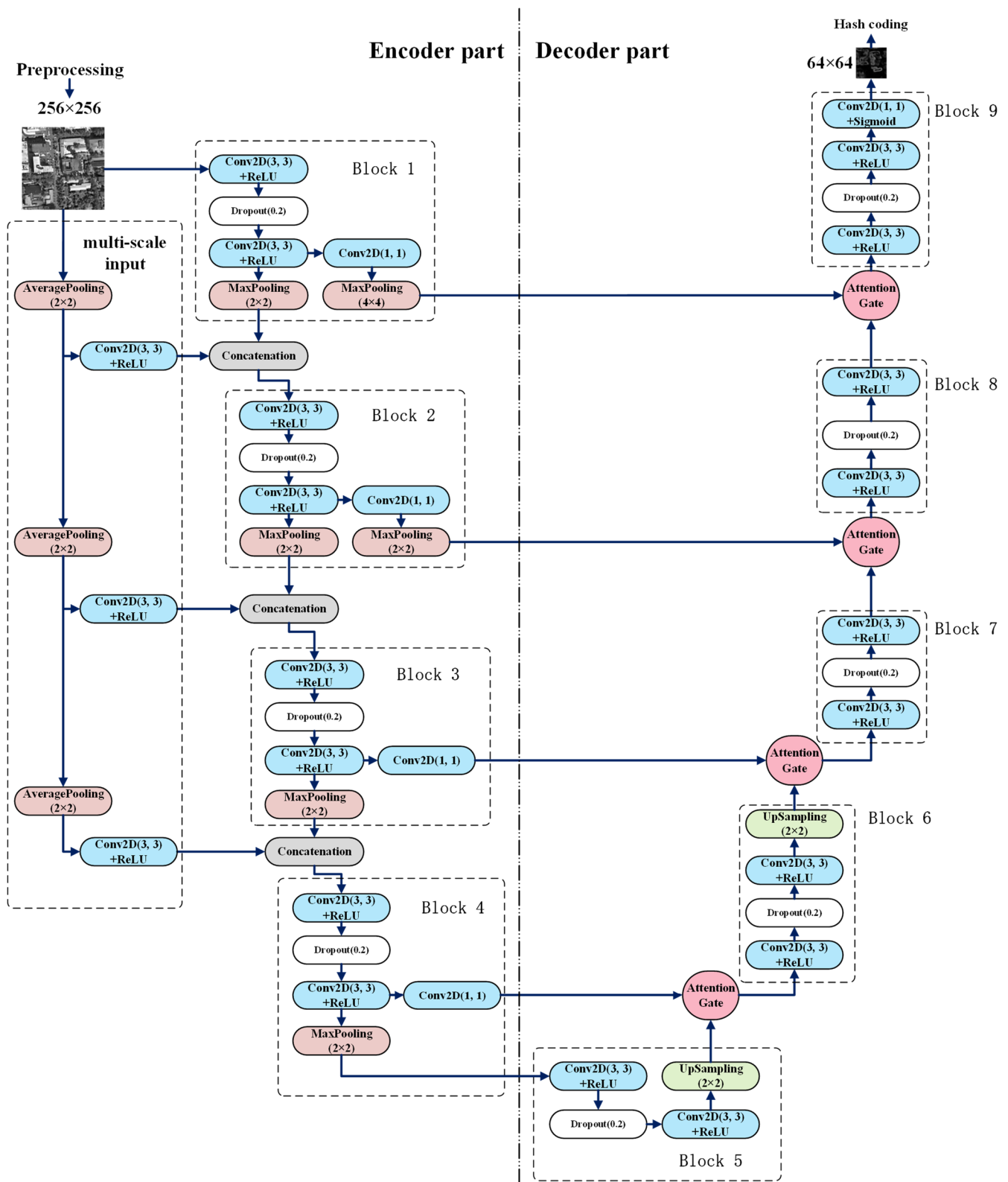


Figure 5. Detailed structure of the proposed network.

1. The encoder and decoder parts of the network structure are asymmetrical. In the original U-Net and variant networks based on U-Net (such as TernaUSNet [41], Res-UNet [42], MultiResUNet [43], and Attention U-Net [39]), the level of pooling in the encoder stage and the level of upsampling in the decoder stage are generally the same,

while in AAU-Net, the level of pooling is greater than the level of upsampling: after upsampling in block 6, there is no upsampling in block 7, block 8, and block 9. If there are more upsampling operations, the pixel-level noise (false features composed of several pixels) will also increase.

2. The asymmetry of input and output. The size of the input image of AAU-Net is 256×256 , but the output is 64×64 . This not only helps to reduce information redundancy, simplifying the downsampling operation of the algorithm in the feature processing stage, but also improves the robustness of the subject-sensitive hashing algorithm. In fact, even the traditional perceptual hash algorithm uses the method of reducing the image resolution in the image preprocessing stage to increase the robustness of the algorithm. After all, the lower the image resolution, the less useless information it contains.
3. The continuous pooling operation and stride convolution in the original U-Net cause the loss of some spatial information [44], while the multi-scale input can greatly reduce the loss of spatial information [45], thereby improving the accuracy of the model. In AAU-Net, we perform downsampling processing on the original input image and build a multi-scale input (image pyramid) to reduce the loss of spatial information due to convolution and pooling operations, so that each layer of the encoding stage can learn more rich features, and then improve the model's ability to extract subject-sensitive features.
4. In the decoder stage of AAU-Net, an attention gate is added to suppress irrelevant areas, while the encoder does not introduce the attention mechanism, which is similar to Attention U-Net [39]. Although the multi-scale input of the encoder helps to extract rich information, it also increases the possibility of extracting useless features such as noise. The attention mechanism causes AAU-Net to suppress the useless features as much as possible.

3.1.2. Attention Gate

In AAU-Net, attention gates are mainly used in each module of the decoding stage to suppress irrelevant areas, which is similar to Attention U-Net [39]. The structure diagram of the attention gate [39,46,47] is shown in Figure 6.

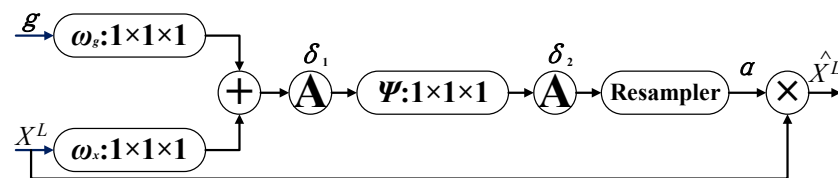


Figure 6. Diagram of the attention gate.

In Figure 6, g represents the gating vector acting on the input pixel, which contains context information and is used to reduce the response of lower-level features; X^L represents the initial feature map of the input; δ_1 and δ_2 represent the activation functions; α is the attention coefficient related to a specific task, which is used to suppress the expression of features not related to the specific task; and \hat{X}^L represents the output feature map, which is obtained by multiplying the attention coefficient α and X^L . The calculation formula of the attention coefficient α is as follows:

$$\alpha = \delta_2(\Psi^T(\delta_1(\omega_x^T X^L + \omega_g^T g + b_1)) + b_2) \quad (5)$$

where b_1 and b_2 represent bias, and Ψ and ω represent $1 \times 1 \times 1$ convolution operations.

It should be noted that the activation functions δ_1 and δ_2 are not limited to specified activation functions. Δ_1 and δ_2 can be ReLU or Elu, LeakyReLU, PReLU, or other activation functions. In our experiment, after repeated experiments, combined with the conclusions in [39,46,47], δ_1 was determined as ReLU, and δ_2 was determined as the sigmoid function.

3.1.3. Loss Function

Subject-sensitive features account for a small proportion of RS images; that is, this is unbalanced sample learning. Therefore, the loss function of AAU-Net, which is known as the cost function, should be adapted to this feature learning with “a small proportion of effective features”. Focal loss (FL) [48] is very suitable for the learning of subject-sensitive features. FL is a modified version of cross-entropy (CE) loss [49] and seeks to balance between easy and hard samples. Our AAU-Net uses the α -balanced variant of focal loss [50] as the loss function, as follows:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (6)$$

In the formula, p_t represents the estimated probability of the model being a positive class at this pixel, while α_t can balance the ratio of positive and negative samples, and the selection of α_t needs to consider the value of γ . The optimal values of the two influence each other, so the optimization of α_t and γ needs to be combined and adjusted. In the experiment, combined with the conclusions of [16,49,50], we determined that $\alpha_t = 0.25$ and $\gamma = 2$ are the optimal settings, which makes the subject-sensitive hash algorithm relatively optimal in terms of robustness and tampering sensitivity.

3.2. Subject-Sensitive Hashing Based on AAU-Net

3.2.1. The Process of Generating Subject-Sensitive Hash Sequences

The existing sensing hashing algorithms for RS images more or less refer to the image sensing hashing algorithms. However, compared with ordinary images, RS images often contain multiple independent bands, and different bands have clear physical meanings and carry different spectral information; moreover, ordinary images are mainly grayscale images (single band) or color images (three bands), while the number of bands of RS images is not unique and may be more than three bands. The existing perceptual hash algorithm for RS images mainly does not take into account the band characteristics of RS images.

To overcome the above problems, our AAU-Net-based subject-sensitive hashing algorithm has some differences from the existing perceptual hash algorithm (including the subject-sensitive hashing algorithm) [11–16]: our algorithm separately extracts the subject-sensitive features of each band of RS images and then performs feature fusion, instead of graying the image and then extracting features. Taking the RS image with 3 bands as an example, our algorithm flow is shown in Figure 7.

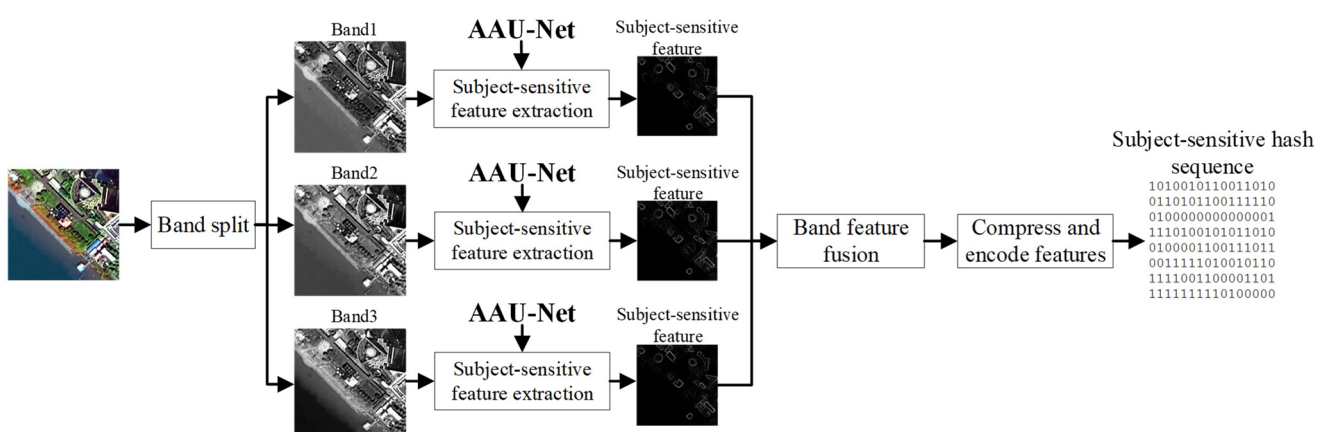


Figure 7. Flow chart of subject-sensitive hashing algorithm based on AAU-Net (taking a 3-band image as an example).

Step 1. Image preprocessing based on band split.

Assuming that the RS image (denoted as I) has N bands, each band of the original RS image is represented by a single-band image independently and is adjusted to the input

size of AAU-Net—that is, 256×256 pixels. The band image after preprocessing is denoted as I_k , where k is a positive integer less than N .

If the RS image is too large, the image can be divided into grids first, and the divided grid unit can be used for the next step, similarly to the method of [13,14,16].

Step 2. Subject-sensitive feature extraction of images based on AAU-Net. This step uses the trained AAU-Net to extract the subject-sensitive features of the preprocessed band image I_k , and the output result is a 64×64 grayscale image which is denoted as F_k .

Step 3. The fusion of the features of each band. The existing subject-sensitive hash algorithm generally does not have this step, and this is the biggest difference between this algorithm and the existing algorithm. Pixel-level image fusion can better retain the original information, which is conducive to subsequent image analysis and feature extraction [51,52]. We construct an adaptive weighted feature fusion method which determines the weighting coefficient according to information entropy:

First, calculate each weighting coefficient α_k according to the information entropy of each band, as shown in (7), where E represents the information entropy of the band image, k represents the band number, and N represents the number of bands:

$$\alpha_k = \frac{E_k}{\sum_{i=1}^N E_k} \quad (7)$$

Next, let the fusion result be denoted as F ; then, each pixel $F(i, j)$ of F is as shown in (8):

$$F(i, j) = \sum_{k=1}^N \alpha_k F_k(i, j) \quad (8)$$

Among them, $F_k(i, j)$ represents the value of the corresponding pixel in F_k .

Step 4. Compress and encode the fused feature image F to generate the subject-sensitive hash sequence of the original RS image. Here, we use the PCA-based feature dimensionality reduction method to extract the high-level values of the principal components as the subject-sensitive feature. After the subject-sensitive features are normalized and quantified, the binary sequence obtained is the subject-sensitive hash sequence.

The above process is described in pseudocode in Algorithm 1.

Algorithm 1: Subject-sensitive hash algorithm of an RS image.

Input: Storage path of the RS image

Output: Subject-sensitive hash sequence

```

1   $I = \text{read}(\text{Storage path of the RS image})$ 
2   $\text{Band}_1, \text{Band}_2, \dots, \text{Band}_N = \text{split}(I)$ 
3   $I_1, I_2, I_3 \dots, I_N = \text{pre-process}(\text{Band}_1, \text{Band}_2, \dots, \text{Band}_N)$ 
4  for  $I_k$  in  $(I_1, I_2, I_3 \dots, I_N)$ 
5       $F_k = \text{AAU-Net}(I_k)$ 
6  end for
7   $F = \sum \alpha_k F_k$ 
8  Subject-sensitive feature = PCA( $F$ )
9  Subject-sensitive hash sequence = Normalization and Quantification (Subject-sensitive feature)
```

3.2.2. Integrity Authentication Process

The authentication process of the subject-sensitive hashing algorithm based on AAU-Net is realized by comparing the normalized hamming distance [53] of the hash sequence. The detailed process is as follows:

Step 1. Generate the subject-sensitive hash sequence of the original RS image I , denoted as hash_1 . hash_1 is stored and transmitted as an attachment to the RS image I , in case it is necessary to authenticate the integrity of the RS image.

Step 2. For the image to be authenticated (denoted as I'), use the same process in Section 3.2.1 to generate its subject-sensitive hash sequence, denoted as $hash_2$.

Step 3. Calculate the normalized hamming distance [53] between $hash_1$ and $hash_2$, denoted as Dis :

$$Dis = \left(\sum_{i=1}^{stringlength} |hash1(i) - hash2(i)| \right) / stringlength \quad (9)$$

where $stringlength$ is the length of the hash sequence in binary format.

Step 4. Compare Dis with the set threshold T . If Dis is greater than or equal to T , the content of the RS image I' to be authenticated has been tampered with; if Dis is less than T , the content of the RS image I' has not been tampered with—that is, the RS image I' has passed the authentication.

The above process is described in pseudocode in Algorithm 2.

Algorithm 2: Integrity authentication process.

Input: T , $Hash_1$ and I'

Output: Authentication result

```

1   $Hash_2$  = Subject-sensitive_hash ( $I'$ )
2   $Dis$  = Normalized_hamming_distance ( $Hash_1$ ,  $Hash_2$ )
3  If ( $Dis > T$ )
4      Integrity authentication failed
5  else
6      Passed integrity authentication
7  end if
```

4. Experiment and Discussion

4.1. Performance Evaluation Metrics

For perceptual hashing, evaluation metrics include robustness, tampering sensitivity, high efficiency (computing efficiency), and security. Among them, robustness and tampering sensitivity are the two most important evaluation metrics. As a special case of perceptual hashing, the evaluation metrics of subject-sensitive hashing can refer to the evaluation metrics of perceptual hash, and the evaluation component of subject-sensitive authentication should be added.

In this paper, we mainly used the following metrics to evaluate the performance of subject-sensitive hashing algorithms based on the evaluation metrics of perceptual hash:

(1) Performance of robustness.

Robustness refers to the proportion of data whose hash sequence changes below the threshold after the image in the test dataset undergoes the operation of “not changing the image content”. We further divided the robustness into two categories: subject-related robustness and subject-unrelated robustness. The metrics for evaluating robustness are defined as

$$RO(T) = \frac{Num_{RO}}{Num_{Total_RO}} \quad (10)$$

where T represents a preset threshold, Num_{Total_RO} represents the number of images for robustness testing, and Num_{RO} represents the number of images whose hash sequence change is higher than the threshold T .

(2) Performance of sensitivity to tampering.

Sensitivity to tampering, also known as tampering sensitivity, was used to test the ability of the subject-sensitive hash algorithm to detect malicious tampering. In this paper, the proportion of tampering detected successfully under different thresholds was used

to describe the tampering sensitivity of the algorithm, which is the same as perceptual hashing. It is defined as follows:

$$SE(T) = \frac{Num_{SE}}{Num_{Total_SE}} \quad (11)$$

where Num_{SE} represents the number of images whose hash sequence change is lower than the preset threshold T . As with robustness testing, tampering sensitivity testing also needs to be performed under multiple different thresholds.

(3) Other evaluation metrics.

Other evaluation metrics include computing efficiency, algorithm security, and abstraction. Computing performance means that the algorithm calculates the hash sequence of the image as quickly as possible while ensuring robustness and tampering sensitivity; algorithm security requires that the effective information of the original image cannot be obtained from the hash sequence; abstraction means that the length of the hash sequence is as short as possible, because the shorter the hash sequence, the more conducive it is to storage and use.

4.2. Datasets and Experimental Settings

4.2.1. Datasets

In this paper, three types of datasets were used to train AAU-Net and evaluate the performance of the subject-sensitive hashing algorithm based on AAU-Net: the datasets for model training, the datasets for robustness testing, and the datasets for tampering sensitivity testing. For the dataset used for model training, we used the same training dataset as MUM-Net [16], which combines the existing WHU building dataset [54] and the method of manually drawing robust edge images. The datasets used for robustness testing and tampering sensitivity testing are discussed in Sections 4.4.1 and 4.4.2, respectively.

4.2.2. Implementation Details

We implemented our AAU-Net network using Keras (Tensorflow as backend) on a GPU workstation equipped with an Intel I7-9700K CPU @3.60 GHz, 32 GB DDR4 RAM, and NVIDIA RTX 2080Ti GPU (11 G memory). In the training process of AAU-Net, the batch size was set to 4, and the number of epochs was set to 100; ReLU and sigmoid were selected as the activation functions of the attention gate of AAU-Net, since we found through experimental comparison that if δ_1 is ReLU and δ_2 is sigmoid, the tampering sensitivity and robustness of the algorithm are better than other activation functions (such as Elu, LeakyReLU, and PReLU).

During the experiment, we selected the following models for comparison with AAU-Net: original U-Net [40], M-Net [55], MUM-Net [16], MultiResU-Net [43], Attention U-Net [39], Attention ResU-Net [56], and Attention R2U-Net (integration of recurrent residual U-Net (R2UNet) [57] and Attention U-Net [39]).

To prove the effectiveness of AAU-Net's asymmetric input and output, we employed Multi-scale Attention U-Net (MA-U-Net) [58,59] as a comparison model. The network structure of MA-U-Net is similar to that of Attention U-Net, with multi-scale input. The biggest difference between MA-U-Net and our AAU-Net is that each block of the network decoding part of MA-U-Net contains an upsampling layer—that is, the input and output of MA-U-Net are symmetrical.

For each comparison model, the same process as in Section 3.2.1 was used to construct the subject-sensitive hashing algorithm, which is compared with the AAU-Net-based algorithm in the following sections. Although the lightweight model semi-U-Net proposed in [60] can also be used to implement subject-sensitive hashing, its main purpose is to reduce the weight of the model and improve the computational performance. The robustness and tampering sensitivity of subject-sensitive hashing algorithms based on semi-U-Net are not outstanding, so we do not discuss semi-U-Net as a comparison model in this paper.

Moreover, as the algorithm flow in this paper is quite different from the algorithm in [60], the algorithm in this paper is not comparable with the subject-sensitive hashing algorithm in [60].

4.3. Examples of Integrity Authentication

To more intuitively illustrate the integrity authentication of the subject-sensitive hashing algorithm based on AAU-Net, we chose the RS image shown in Figure 8a as an example to compare the performance of our algorithm and each comparison algorithm.



Figure 8. Examples of authentication for RS image: (a) the original RS image; (b,c) operation of content retention (TIFF format to BMP format, lossy data compression); (d–f) subject-unrelated image content changes; (g–j) subject-related image content changes.

In Figure 8, Figure 8a is the original RS image (TIF format), and the others can be divided into 3 groups: Figure 8b,c are images after format conversion (TIFF format to BMP format) and lossy compression (90% JPEG compression), respectively; Figure 8d–f are examples of subject-unrelated tampering—that is, the content change of the image has nothing to do with the building; Figure 8g–j are subject-related content tampering: Figure 8g has a building added, Figure 8h has been maliciously smeared, in Figure 8i, a building has been razed, and a building in Figure 8j is replaced.

We used the subject-sensitive hashing algorithm based on AAU-Net and the subject-sensitive hashing algorithm based on the contrast models in Section 4.2.1 to generate the hash sequence of each image in Figure 8. Each subject-sensitive hashing algorithm differs only in the model it uses, and the other processes remain the same. The normalized distance between the subject-sensitive hash sequences of each image in Figure 8b–j and the original RS image (i.e., Figure 8a) is shown in Table 1.

It can be seen from Table 1 that all of the subject-sensitive hashing algorithms based on each deep learning model can realize “subject-sensitive” integrity authentication. That is, a certain tolerance is maintained for subject-unrelated image changes, but the subject-sensitive hash sequence changes more drastically for subject-related image changes. The results of the integrity authentication based on Table 1, when the threshold T was set to 0.03, are shown in Table 2.

It can be seen from Table 2 that our AAU-Net-based subject-sensitive hashing algorithm is the best-performing algorithm compared to algorithms based on other deep learning models. The AAU-Net-based algorithm not only shows a certain degree of robustness to subject-unrelated image content changes but also detects all malicious tampering.

The U-Net-based algorithm, the M-Net-based algorithm, the MultiResU-Net-based algorithm, and the Attention R2U-Net-based algorithm can also detect malicious tampering in Figure 8, but they are not robust enough to subject-unrelated content changes; the Attention ResU-Net-based algorithm, the MA-U-Net-based algorithm, and the Attention U-Net-based algorithm are more robust to subject-unrelated content changes but fail to detect all malicious tampering in Figure 8.

Table 1. Normalized hamming distance of the algorithms based on different models.

Tampering Test	Figure 8b	Figure 8c	Figure 8d	Figure 8e	Figure 8f	Figure 8g	Figure 8h	Figure 8h	Figure 8h
U-Net-based algorithm	0.0	0.046875	0.05859375	0.02734375	0.05859375	0.20703125	0.12109375	0.078125	0.03515625
M-Net-based algorithm	0.0	0.15234375	0.01171875	0.0234375	0.01953125	0.24609375	0.24609375	0.171875	0.1328125
MultiResU-Net-based algorithm	0.0	0.09375	0.015625	0.0390625	0.0625	0.2578125	0.23046875	0.06640625	0.13671875
MUM-Net-based algorithm	0.0	0.20703125	0.0234375	0.03515625	0.0625	0.11328125	0.20703125	0.2578125	0.01171875
Attention ResU-Net-based algorithm	0.0	0.02734375	0.0	0.015625	0.0	0.07421875	0.234375	0.015625	0.0
Attention R2U-Net-based algorithm	0.0	0.10546875	0.10546875	0.0234375	0.125	0.1796875	0.26953125	0.1640625	0.0703125
Attention U-Net-based algorithm	0.0	0.0625	0.0078125	0.02734375	0.01171875	0.125	0.23046875	0.1015625	0.01171875
MA-U-Net-based algorithm	0.0	0.078125	0.03125	0.09375	0.02734375	0.140625	0.1796875	0.21875	0.08203125
AAU-Net-based algorithm	0.0	0.0234375	0.0078125	0.00390625	0.0078125	0.11328125	0.2890625	0.06640625	0.0390625

Table 2. Result of the integrity authentication based on Table 1 (threshold T set to 0.03).

Tampering Test	Figure 8b	Figure 8c	Figure 8d	Figure 8e	Figure 8f	Figure 8g	Figure 8h	Figure 8h	Figure 8h
U-Net-based algorithm	Success	Fail	Fail	Success	Fail	Success	Success	Success	Success
M-Net-based algorithm	Success	Fail	Success	Success	Success	Success	Success	Success	Success
MultiResU-Net-based algorithm	Success	Fail	Success	Fail	Fail	Success	Success	Success	Success
MUM-Net-based algorithm	Success	Fail	Success	Fail	Fail	Success	Success	Success	Fail
Attention ResU-Net-based algorithm	Success	Success	Success	Success	Success	Success	Success	Fail	Fail
Attention R2U-Net-based algorithm	Success	Fail	Fail	Success	Fail	Success	Success	Success	Success
Attention U-Net-based algorithm	Success	Fail	Success	Success	Success	Success	Success	Success	Fail
MA-U-Net-based algorithm	Success	Fail	Fail	Fail	Success	Success	Success	Success	Success
AAU-Net-based algorithm	Success	Success	Success	Success	Success	Success	Success	Success	Success

In the actual integrity authentication process, the threshold T is often set according to the requirements of the RS image application environment, training sample set, authentication intensity, and other requirements, and different thresholds T often result in different authentication results. For example, if the authentication intensity is required to be high, the threshold T should be set relatively low; on the contrary, if robustness is required, the threshold T should be set relatively high. On the basis of Tables 1 and 2, the value of the threshold T was set to 0.02 and 0.05, respectively, and the corresponding integrity authentication results are shown in Tables 3 and 4, respectively.

It can be seen from Table 3 that a smaller threshold T reduces the robustness of the algorithms. At the same time, it can be seen from Table 4 that a larger threshold T greatly improves the robustness of the algorithms, but the sensitivity to tampering of the algorithms also decreases.

Of course, testing a set of examples can only provide an intuitive explanation and does not guarantee the performance of the algorithm. Next, we test and compare AAU-Net-based subject-sensitive hashing algorithms in more depth.

Table 3. Result of the integrity authentication based on Table 1 (threshold T set to 0.02).

Tampering Test	Figure 8b	Figure 8c	Figure 8d	Figure 8e	Figure 8f	Figure 8g	Figure 8h	Figure 8h	Figure 8h
U-Net-based algorithm	Success	Fail	Fail	Fail	Fail	Success	Success	Success	Success
M-Net-based algorithm	Success	Fail	Success	Fail	Success	Success	Success	Success	Success
MultiResU-Net-based algorithm	Success	Fail	Success	Fail	Fail	Success	Success	Success	Success
MUM-Net-based algorithm	Success	Fail	Fail	Fail	Fail	Success	Success	Success	Fail
Attention ResU-Net-based algorithm	Success	Fail	Success	Fail	Success	Success	Success	Fail	Fail
Attention R2U-Net-based algorithm	Success	Fail	Fail	Fail	Fail	Success	Success	Success	Success
Attention U-Net-based algorithm	Success	Fail	Success	Fail	Success	Success	Success	Success	Fail
MA-U-Net-based algorithm	Success	Fail	Fail	Fail	Fail	Success	Success	Success	Success
AAU-Net-based algorithm	Success	Fail	Success	Success	Success	Success	Success	Success	Success

Table 4. Result of the integrity authentication based on Table 2 (threshold T set to 0.05).

Tampering Test	Figure 8b	Figure 8c	Figure 8d	Figure 8e	Figure 8f	Figure 8g	Figure 8h	Figure 8h	Figure 8h
U-Net-based algorithm	Success	Success	Fail	Success	Fail	Success	Success	Success	Fail
M-Net-based algorithm	Success	Fail	Success	Success	Success	Success	Success	Success	Success
MultiResU-Net-based algorithm	Success	Fail	Success	Success	Fail	Success	Success	Success	Success
MUM-Net-based algorithm	Success	Fail	Success	Success	Fail	Success	Success	Success	Fail
Attention ResU-Net-based algorithm	Success	Success	Success	Success	Success	Success	Success	Fail	Fail
Attention R2U-Net-based algorithm	Success	Fail	Fail	Success	Fail	Success	Success	Success	Success
Attention U-Net-based algorithm	Success	Fail	Success	Success	Success	Success	Success	Success	Fail
MA-U-Net-based algorithm	Success	Fail	Success	Fail	Success	Success	Success	Success	Success
AAU-Net-based algorithm	Success	Success	Success	Success	Success	Success	Success	Success	Fail

4.4. Performance Analysis

4.4.1. Performance of Robustness

Differing from deep learning applications such as image segmentation and image classification, the testing of subject-sensitive hashing algorithms based on deep learning requires a large amount of data for testing to make the results more reliable.

In this paper, we constructed an RS image dataset named $Datasets_{10000}$ to test the robustness of the algorithm. The number of RS images contained in $Datasets_{10000}$ is 10,000, which means

$$\text{card}(Datasets_{10000}) = 10,000 \quad (12)$$

The images of $Datasets_{10000}$ are stored in TIF format and generated from the GaoFen-2 (GF-2) satellite, DOTA [61], and some test data in [16,60]. In addition, the size of each image in $Datasets_{10000}$ is adjusted to 256×256 pixels.

Unlike ordinary images which focus on visual effects, RS images have higher requirements regarding accuracy. There are not as many content retention operations for RS images as for ordinary images. Data format conversion, digital watermark embedding, and data compression are common retention operations for RS images. Among them, digital watermarking technology is often used to protect the copyright of RS images [62]. Data compression can reduce the amount of data to reduce storage space and improve their transmission and storage efficiency [63]. Therefore, we tested and compared the robustness of the algorithm from the above aspects.

First, we tested the robustness of the algorithm to image compression. Here, we used portable network graphics (PNG) compression technology to perform lossy compression on the images in $Datasets_{10000}$ and used the percentage of images whose hash sequence changes are lower than the threshold to describe the robustness of the algorithm. In the actual application of integrity authentication, due to factors such as different training datasets and different authentication strengths, the threshold setting is often not unique. Therefore, we counted the results under different thresholds, as shown in Table 5.

Table 5. Robustness test comparison of PNG lossy compression.

	$T = 0.02$	$T = 0.03$	$T = 0.05$	$T = 0.10$	$T = 0.20$	$T = 0.30$
U-Net-based algorithm	64.2%	75.0%	85.2%	96.6%	99.3%	99.9%
M-Net-based algorithm	71.6%	81.0%	91.9%	98.0%	99.6%	100%
MultiResU-Net-based algorithm	87.8%	93.0%	97.5%	99.2%	99.9%	100%
MUM-Net-based algorithm	63.0%	74.8%	88.5%	97.2%	100%	100%
Attention ResU-Net-based algorithm	94.1%	96.4%	98.3%	99.9%	100%	100%
Attention R2U-Net-based algorithm	53.0%	64.8%	78.9%	92.0%	97.8%	100%
Attention U-Net-based algorithm	91.6%	94.5%	97.8%	99.4%	99.9%	100%
MA-U-Net-based algorithm	90.7%	92.2%	98.5%	99.8%	100%	100%
AAU-Net-based algorithm	96.1%	97.8%	98.8%	99.8%	100%	100%

It can be seen from the results in Table 5 that our AAU-Net-based algorithm has the best robustness under different thresholds. Especially when the threshold T is set relatively low, such as $T = 0.02$, the performance of the algorithm based on AAU-Net is still relatively good, which is better than the algorithm based on other deep learning models.

It should be noted that the process of each comparison algorithm is the same, except that the deep neural network models used to extract subject-sensitive features are different. In other words, the difference in algorithm performance is caused by different models. However, the process of each algorithm in [16,60] is quite different from the process of the algorithm in this paper: the algorithm in this paper extracts the features of each band of the RS image separately, while the algorithms in [16,60] perform feature extraction on the gray image after band fusion; moreover, the algorithm in this paper performs a more compact compression coding on the subject-sensitive features.

Next, we tested and compared the robustness of each algorithm to joint photographic experts group (JPEG) compression. Here, we performed 95% JPEG compression on the images in *Datasets*₁₀₀₀₀ using opencv 2.4.13, with C++ as the programming language. The robustness test results of each algorithm are shown in Table 6.

Table 6. Robustness test comparison of JPEG compression (95% JPEG).

	$T = 0.02$	$T = 0.03$	$T = 0.05$	$T = 0.10$	$T = 0.20$	$T = 0.30$
U-Net-based algorithm	63.3%	75.1%	88.9%	96.7%	99.8%	100%
M-Net-based algorithm	70.2%	80.1%	92.2%	97.6%	99.7%	100%
MultiResU-Net-based algorithm	76.3%	86.0%	94.7%	98.1%	99.6%	100%
MUM-Net-based algorithm	79.6%	90.9%	97.6%	99.9%	100%	100%
Attention ResU-Net-based algorithm	86.0%	91.1%	96.1%	99.3%	99.9%	100%
Attention R2U-Net-based algorithm	58.5%	69.2%	83.0%	95.5%	99.3%	100%
Attention U-Net-based algorithm	85.9%	92.6%	97.1%	98.9%	100%	100%
MA-U-Net-based algorithm	87.6%	93.0%	97.4%	99.6%	100%	100%
AAU-Net-based algorithm	73.6%	81.0%	93.3%	98.8%	99.8%	100%

It can be seen from Table 6 that the AAU-Net-based algorithm is less robust to JPEG compression than the Attention U-Net-based [56] algorithm, the MA-U-Net-based [58,59] algorithm, and the MUM-Net-based [16] algorithm, but it is stronger than the U-Net-based [40] algorithm and the Attention R2U-Net-based [57] algorithm. It can be seen from Table 6 that MUM-Net is more robust than U-Net and M-Net, which is consistent with the conclusion of [16].

The application of digital watermarking technology in RS images is used for the copyright identification and data tracking of RS images. The RS images before and after the watermark is embedded are generally considered to be consistent in content. For RS images, the embedding position of the digital watermark can be a single band or multiple bands, while our subject-sensitive hashing algorithm extracts the features of each band of the RS image. Therefore, we separately tested the robustness of our algorithm to

digital watermarks in the case of single-band watermarking and multi-band watermarking, instead of testing the robustness of the watermark type indiscriminately.

For a single-band watermark, we took a spatial watermark as an example to embed 32-bit information into the RS images in *Datasets*₁₀₀₀₀ and calculated the normalized hamming distance between the hash sequences of the watermark-embedded RS image and the original RS image. The robustness test results are shown in Table 7.

Table 7. Robustness test comparison of watermark embedding (32 bits embedded in single band).

	$T = 0.02$	$T = 0.03$	$T = 0.05$	$T = 0.10$	$T = 0.20$	$T = 0.30$
U-Net-based algorithm	99.6%	99.8%	99.8%	100%	100%	100%
M-Net-based algorithm	100%	100%	100%	100%	100%	100%
MultiResU-Net-based algorithm	100%	100%	100%	100%	100%	100%
MUM-Net-based algorithm	100%	100%	100%	100%	100%	100%
Attention ResU-Net-based algorithm	100%	100%	100%	100%	100%	100%
Attention R2U-Net-based algorithm	99.7%	99.9%	100%	100%	100%	100%
Attention U-Net-based algorithm	99.8%	99.8%	99.8%	99.9%	100%	100%
MA-U-Net-based algorithm	99.8%	99.9%	99.9%	100%	100%	100%
AAU-Net-based algorithm	99.8%	100%	100%	100%	100%	100%

It can be seen from Table 7 that the robustness of each algorithm to single-band watermarking is relatively ideal. This is mainly because the single-band watermarking algorithm only modifies one band of the RS image, and the other bands are not changed, while each subject-sensitive hashing algorithm extracts all the bands of the RS image to generate the hash sequence, which is affected by the single-band watermark to a lesser extent.

For the robustness test of a multi-band watermark, we embedded 24-bit watermarks on the three bands of the RS image in *Datasets*₁₀₀₀₀. The test results are shown in Table 8.

Table 8. Robustness test comparison of watermark embedding (24 bits embedded in 3 bands).

	$T = 0.02$	$T = 0.03$	$T = 0.05$	$T = 0.10$	$T = 0.20$	$T = 0.30$
U-Net-based algorithm	61.7%	81.8%	85.6%	95.8%	99.3%	100%
M-Net-based algorithm	68.7%	88.5%	90.03%	97.5%	99.5%	100%
MultiResU-Net-based algorithm	87.5%	92.1%	96.8%	99.1%	99.9%	100%
MUM-Net-based algorithm	62.1%	74.3%	88.2%	97.3%	99.7%	100%
Attention ResU-Net-based algorithm	93.6%	96.1%	98%	99.8%	100%	100%
Attention R2U-Net-based algorithm	48.7%	60.2%	77.5%	92.0%	98.5	100%
Attention U-Net-based algorithm	90.1%	94.2%	97.3%	99.0%	99.9%	100%
MA-U-Net-based algorithm	90.6%	96.0%	98.2%	98.9%	99.9%	100%
AAU-Net-based algorithm	95.1%	97.1%	98.3%	99.6%	99.9%	100%

It can be seen from Table 8 that our subject-sensitive hashing algorithm based on AAU-Net performs best, and it is not only better than traditional deep neural networks such as U-Net and M-Net but also better than other attention mechanism-based deep neural networks, such as Attention U-Net, Attention ResUNet, Attention R2U-Net, and MA-U-Net.

Subject-sensitive hashing should be robust to the subtle operations that do not change the subject-related content of the RS image. Here, we simulated the above-mentioned subtle subject-unrelated tampering by randomly setting a few pixels of the RS image to 0. In detail, four pixels from each RS image in *Datasets*₁₀₀₀₀ were randomly selected and set to 0 (the same position of the three bands was modified). Table 9 shows the robustness test results of subtle subject-unrelated tampering.

It can be seen from Table 9 that our AAU-Net-based subject-sensitive hashing algorithm has the best robustness for subtle subject-unrelated tampering among all of the compared algorithms. Even when the threshold T is small, it can still maintain good robustness. Other models such as U-Net, M-Net, and MUM-Net can only guarantee a

certain robustness when the threshold T is large, but if the threshold T is set too large, the algorithm's tampering sensitivity will be weakened.

Table 9. Robustness test comparison of subtle subject-unrelated tampering.

	$T = 0.02$	$T = 0.03$	$T = 0.05$	$T = 0.10$	$T = 0.20$	$T = 0.30$
U-Net-based algorithm	52.0%	64.2%	90.6%	93.4%	98.6%	100%
M-Net-based algorithm	53.5%	63.4%	78.5%	92.8%	98.6%	100%
MultiResU-Net-based algorithm	79.7%	88.5%	96.0%	98.7%	99.8%	99.9%
MUM-Net-based algorithm	42.1%	52.3%	71.1%	90.4%	98.3%	99.9%
Attention ResU-Net-based algorithm	81.5%	86.6%	92.1%	96.4%	98.9%	100%
Attention R2U-Net-based algorithm	19.5%	27.0%	41.3%	64.2%	86.1%	99.9%
Attention U-Net-based algorithm	52.0%	64.2%	80.6%	93.4%	98.6%	100%
MA-U-Net-based algorithm	87.4%	93.2%	97.1%	99.3%	99.7%	100%
AAU-Net-based algorithm	90.9%	93.7%	97.4%	99.4%	100%	100%

Based on the robustness test results in Tables 5–9, we can draw the following conclusions: the robustness of our AAU-Net-based algorithm is the best among various comparison algorithms; the most widely used models, such as U-Net, MA-U-Net, and Attention U-Net, can also be used to implement subject-sensitive hashing, but the robustness is not as good as that of AAU-Net; Attention R2U-Net has the worst robustness and is not recommended for subject-sensitive hashing.

In this study, we do not consider operations such as image scaling and image rotation as content retention operations, since they obviously change the content of the RS image.

4.4.2. Performance of Sensitivity to Tampering

While having good algorithm robustness, the subject-sensitive hashing algorithm should also have good tampering sensitivity (also known as “sensitivity to tampering”)—that is, it can detect the content tampering of RS images. We tested the tampering sensitivity of each comparison algorithm from the following aspects:

First of all, we performed random content tampering on the RS image of *Datasets*₁₀₀₀₀, and performed tampering sensitivity testing on the tampered RS image. The size of the tampering area is 16×16 pixels, and the position is random, which can simulate the subject-related tampering of the RS image in reality as effectively as possible. Figure 9 shows a set of examples of the above RS image tampering, and Table 10 shows the tampering detection results corresponding to Figure 9, in which the normalized hamming distance between the hash sequences of the RS images before and after tampering is presented. The overall experimental statistical results of each algorithm are shown in Table 11.

It can be seen from Table 10 that each comparison algorithm can detect the tampering example shown in Figure 9 if the threshold T is set to 0.02. If the threshold T is set to 0.05, the Attention ResU-Net-based algorithm and the MA-U-Net-based algorithm fail to detect the tampering shown in Figure 9g,h, while the Attention U-Net-based algorithm fails to detect the tampering shown in Figure 9h.

Although the experimental data in Table 11 show that the tampering sensitivity based on the Attention R2U-Net-based algorithm is the best, the experiment in Section 4.4.1 shows that the robustness of Attention R2U-Net is too poor to be suitable for subject-sensitive hashing. The tampering sensitivity of our AAU-Net-based algorithm is only slightly weaker than that of the Attention R2U-Net-based algorithm, while it has been proved in Section 4.4.1 that the robustness of our AAU-Net algorithm is the best among all comparison algorithms.

In order to further analyze the tampering sensitivity of each algorithm to subtle subject-related tampering, we reduced the size of the tampering area on the basis of the above tampering sensitivity test—that is, we used a smaller 8×8 pixel size tamper instead of a 16×16 pixel size tamper to test the tampering sensitivity of each comparison algorithm. The experimental results are shown in Table 12.

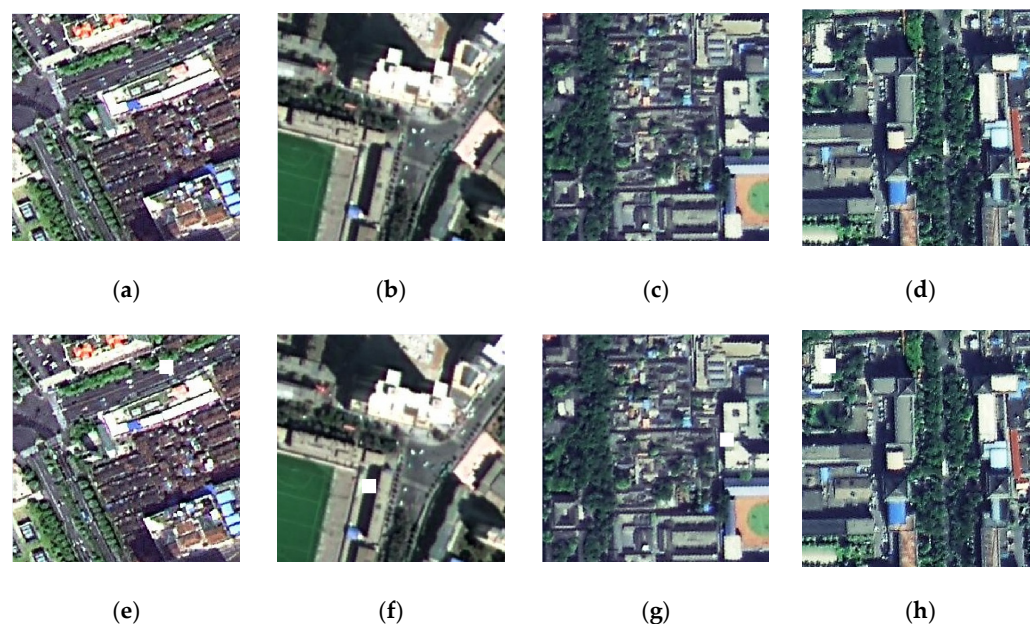


Figure 9. Examples of comparisons before and after subject-related tampering: (a) original image 1; (b) original image 2; (c) original image 3; (d) original image 4; (e) tampered image 1; (f) tampered image 2; (g) tampered image 3; (h) tampered image 4.

Table 10. Normalized hamming distance of each algorithm corresponding to Figure 9.

	Image 1	Image 2	Image 3	Image 4
U-Net-based algorithm	0.2500	0.2539	0.1210	0.1093
M-Net-based algorithm	0.1484	0.2148	0.2578	0.2500
MultiResU-Net-based algorithm	0.2421	0.2421	0.1054	0.2734
MUM-Net-based algorithm	0.1914	0.2773	0.0546	0.1367
Attention ResU-Net-based algorithm	0.1250	0.2421	0.0312	0.0234
Attention R2U-Net-based algorithm	0.1992	0.2695	0.0664	0.2187
Attention U-Net-based algorithm	0.1875	0.2656	0.2421	0.0312
MA-U-Net-based algorithm	0.1562	0.2382	0.0468	0.1484
AAU-Net-based algorithm	0.1523	0.1796	0.2109	0.0703

Table 11. Tampering sensitivity comparison of subject-related tampering (16×16 pixel tampering area).

	$T = 0.02$	$T = 0.03$	$T = 0.05$	$T = 0.10$	$T = 0.20$	$T = 0.30$
U-Net-based algorithm	98.9%	98.0%	96.2%	86.4%	59.2%	0.9%
M-Net-based algorithm	99.2%	99.0%	96.2%	87.0%	60.7%	1.3%
MultiResU-Net-based algorithm	99.1%	97.7%	92.4%	79.3%	54.3%	1.5%
MUM-Net-based algorithm	98.8%	98.5%	95.3%	78.7%	42.8%	1.6%
Attention ResU-Net-based algorithm	83.9%	76.7%	57.5%	34.7%	12.1%	0.7%
Attention R2U-Net-based algorithm	99.9%	99.9%	99.9%	98.0%	81.5%	1.8%
Attention U-Net-based algorithm	98.9%	98.0%	96.2%	86.4%	59.2%	0.9%
MA-U-Net-based algorithm	99.2%	98.1%	93.9	76.2%	38.9%	1.1%
AAU-Net-based algorithm	99.5%	99.2%	96.4%	87.0%	64.2%	1.6%

Table 12. Tampering sensitivity comparison of subject-related tampering (8×8 pixel tampering area).

	$T = 0.02$	$T = 0.03$	$T = 0.05$	$T = 0.10$	$T = 0.20$	$T = 0.30$
U-Net-based algorithm	97.0%	94.9%	85.6%	61.4%	31.1%	0.8%
M-Net-based algorithm	96.3%	94.3%	89.2%	66.8%	33.2%	0.3%
MultiResU-Net-based algorithm	84.0%	78.3%	62.5%	35.0%	11.8%	0.3%
MUM-Net-based algorithm	95.0%	91.0%	82.2%	56.2%	14.9%	0.4%
Attention ResU-Net-based algorithm	67.9%	58.1%	40.7%	23.1%	7.8%	0.3%
Attention R2U-Net-based algorithm	98.8%	98.1%	95.4%	80.9%	52.8%	1.2%
Attention U-Net-based algorithm	93.7%	91.3%	84.4%	68.6%	36.2%	0.6%
MA-U-Net-based algorithm	88.8%	85.5%	74.2%	48.4%	16.8%	9%
AAU-Net-based algorithm	92.8%	90.9%	83.0%	65.6%	29.5%	1.1%

It can be seen from Table 12 that the tampering sensitivity of the Attention R2U-Net-based algorithm is still the best among all algorithms. The tampering sensitivity of our AAU-Net-based algorithm is close to that of the Attention U-Net-based algorithm, the MA-U-Net-based Algorithm, and the MUM-Net-based algorithm, but weaker than that of the Attention R2U-Net-based algorithm. Therefore, further study of the model's detection of subtle tampering is one of our next key tasks.

Subject-sensitive hashing should also be able to detect subject-unrelated tampering, but the detection strength will be lower than that of subject-related tampering. To test the sensitivity of each comparison algorithm to subject-unrelated tampering, we selected some of the RS images in *Datasets*₁₀₀₀₀ for manual tampering and combined them with the image dataset used to test subject-unrelated tampering in [16] to construct a new RS image dataset, which is named *Datasets*₄₀₀, as it has 400 sets of RS images. The tampering sensitivity test results of each comparison algorithm on *Datasets*₄₀₀ are shown in Table 13.

Table 13. Tampering sensitivity comparison of subject-unrelated tampering with 400 RS images.

	$T = 0.02$	$T = 0.03$	$T = 0.05$	$T = 0.10$	$T = 0.20$	$T = 0.30$
U-Net-based algorithm	88.75%	84.5%	74.5%	43.25%	16.0%	0.25%
M-Net-based algorithm	69.5%	60.0%	36.5%	14.75%	4%	0%
MultiResU-Net-based algorithm	73.5%	65.25%	45.75%	15.5%	4.0%	0%
MUM-Net-based algorithm	87.25%	82.25%	68.5%	38.25%	7.5%	0%
Attention ResU-Net-based algorithm	30.75%	25.5%	14.5%	6.5%	1.5%	0%
Attention R2U-Net-based algorithm	89.5%	84.5%	73%	47.25%	21.25%	0.25%
Attention U-Net-based algorithm	82.0%	76.5%	60.25%	28.75%	12.25%	0%
MA-U-Net-based algorithm	87.8%	80.9%	65.1%	36.9%	12.5%	0.1%
AAU-Net-based algorithm	82.75%	80.25%	65.5%	29.5%	10.5%	0%

Comparing Tables 11–13, it can be seen that the sensitivity of each comparison algorithm to the subject-unrelated tampering decreased, which shows that each algorithm maintains a certain degree of robustness to the subject-unrelated tampering. Of course, if necessary, this tampering can still be detected as long as the threshold is lowered.

In order to show the subject-unrelated tampering more clearly, we selected four sets of subject-unrelated tampering examples in *Datasets*₄₀₀, as shown in Figure 10, and the corresponding experimental results are shown in Table 14.

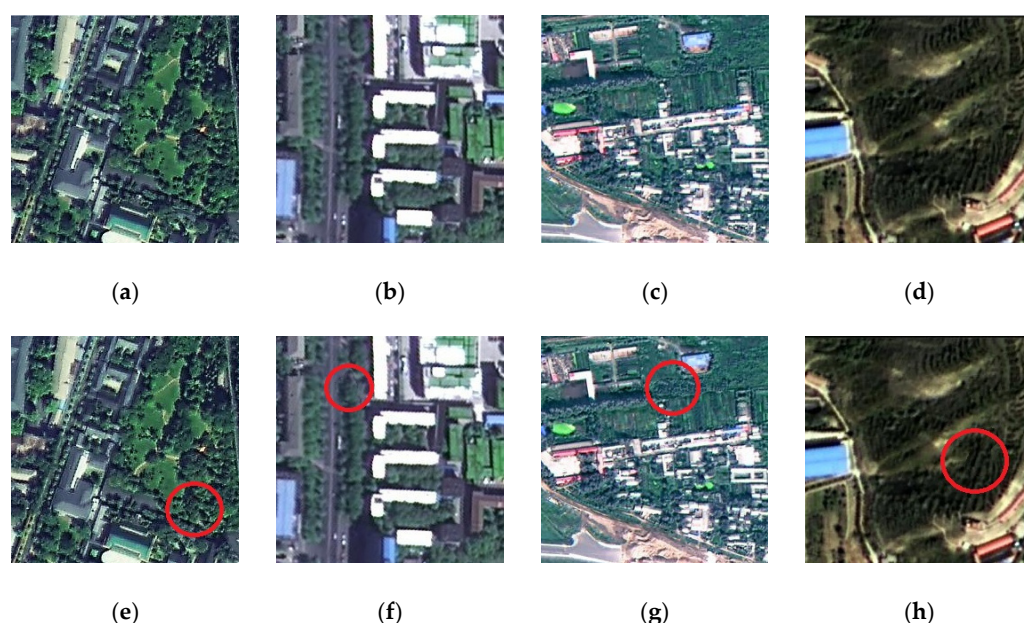


Figure 10. Examples of RS images in *Datasets*₄₀₀: (a–d) the original RS images 1–4; (e–h) the tampered RS images 1–4 (subject-unrelated tampering).

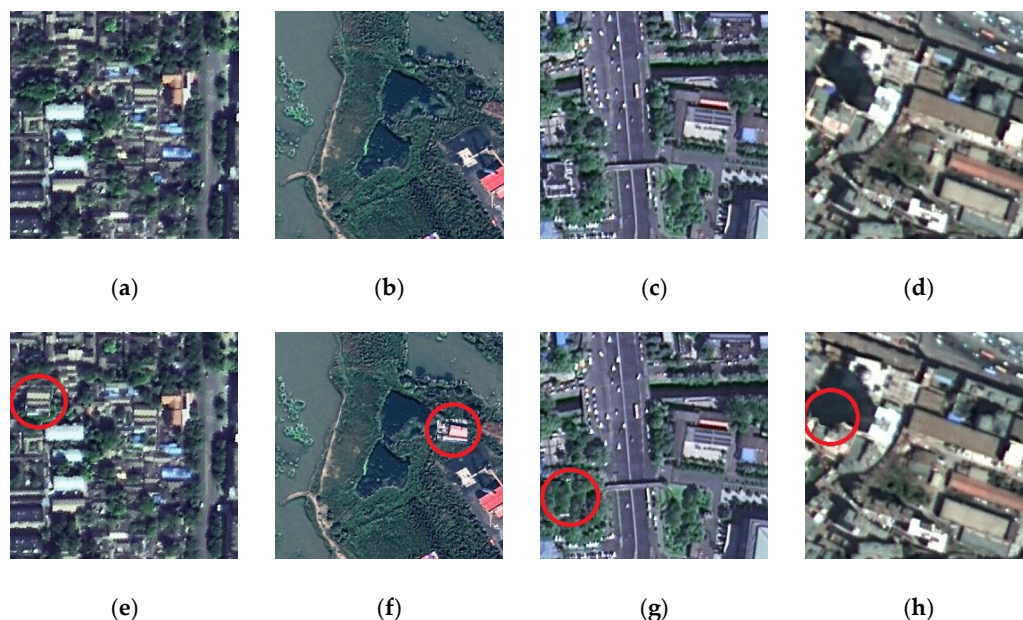
Table 14. Normalized hamming distance of each algorithm corresponding to Figure 10.

	Image 1	Image 2	Image 3	Image 4
U-Net-based algorithm	0.0273	0.0234	0.0742	0.1054
M-Net-based algorithm	0.0117	0.0	0.0078	0.0
MultiResU-Net-based algorithm	0.0429	0.0	0.0	0.0156
MUM-Net-based algorithm	0.0468	0.0	0.0	0.0195
Attention ResU-Net-based algorithm	0.0117	0.0	0.0	0.0
Attention R2U-Net-based algorithm	0.0429	0.0429	0.0156	0.0390
Attention U-Net-based algorithm	0.0156	0.0	0.0390	0.0429
MA-U-Net-based algorithm	0.0234	0.0078	0.0469	0.0664
AAU-Net-based algorithm	0.0078	0.0039	0.0195	0.0234

In the examples shown in Figure 10, the size of the tampered area is larger than the area of the tampering example in Figure 9, which is 16×16 pixels, but the tampered area in Figure 10 is subject unrelated, such as woods and shrubs, and the change in the corresponding subject-sensitive hash sequence should not be as drastic as that seen in the example in Figure 9, as shown in Table 14.

It can be seen from Table 14 that our AAU-Net-based algorithm and other comparison algorithms are significantly less sensitive to subject-unrelated tampering, but, if necessary, this tampering can still be detected by lowering the threshold.

Corresponding to subject-unrelated tampering, we specially constructed an RS image dataset with building contents that had been tampered with to further test the sensitivity of each comparison algorithm to subject-related tampering. This RS image dataset selects 200 original RS images from *Datasets*₁₀₀₀₀ and is named *Datasets*₂₀₀. Figure 11 shows four sets of RS images in *Datasets*₂₀₀, and the corresponding experimental results are shown in Table 15.

**Figure 11.** Examples of RS images in *Datasets*₂₀₀: (a–d) the original RS images 1–4; (e–h) the tampered RS images 1–4.

It can be seen from Table 15 that even if the threshold T is set to a small value (such as 0.02), each algorithm can still detect subject-related tampering.

A more comprehensive tampering sensitivity test on *Datasets*₂₀₀ is shown in Table 16. Comparing Tables 14 and 16, it can be seen that each algorithm is highly sensitive to building-related tampering. Moreover, our AAU-Net-based algorithm is slightly more sensitive to

tampering than the U-Net-based algorithm and the Attention U-Net-based algorithm, but inferior to the MUM-Net-based algorithm and the Attention R2U-Net-based algorithm.

Table 15. Normalized hamming distance of each algorithm corresponding to Figure 11.

	Image 1	Image 2	Image 3	Image 4
U-Net-based algorithm	0.1171	0.2265	0.0820	0.1484
M-Net-based algorithm	0.2421	0.1796	0.2539	0.1484
MultiResU-Net-based algorithm	0.1367	0.1757	0.1718	0.1015
MUM-Net-based algorithm	0.2226	0.1914	0.1132	0.2265
Attention ResU-Net-based algorithm	0.0273	0.0625	0.0546	0.0351
Attention R2U-Net-based algorithm	0.1054	0.1875	0.0898	0.2070
Attention U-Net-based algorithm	0.1835	0.2656	0.2656	0.1445
MA-U-Net-based algorithm	0.2148	0.2265	0.2304	0.2109
AAU-Net-based algorithm	0.1875	0.2929	0.1601	0.1796

Table 16. Tampering sensitivity comparison of subject-related tampering with 200 RS images.

	$T = 0.02$	$T = 0.03$	$T = 0.05$	$T = 0.10$	$T = 0.20$	$T = 0.30$
U-Net-based algorithm	96.0%	94.0%	85.5%	63.5%	30.0%	1.0%
M-Net-based algorithm	98.0%	97.0%	93.5%	76.5%	37.5%	0.5%
MultiResU-Net-based algorithm	90.5%	86.5%	72.0%	39.5%	12.5%	0.0%
MUM-Net-based algorithm	99.0%	98.5%	94.0%	70.0%	31.0%	0.0%
Attention ResU-Net-based algorithm	57.0%	47.5%	32.5%	16.5%	5.0%	0.5%
Attention R2U-Net-based algorithm	99.5%	98.5%	95.0%	82.5%	52.5%	1.0%
Attention U-Net-based algorithm	96.0%	94.0%	85.0%	63.5%	30.0%	0.0%
MA-U-Net-based algorithm	96.9%	94.8%	91.1%	68.2%	28.2%	0.5%
AAU-Net-based algorithm	96.5%	94.0%	87.0%	67.0%	30.0%	1.0%

Comparing Tables 11 and 16, we can see that each algorithm (not just our AAU-Net-based algorithm) is more sensitive to the random tampering of 16×16 pixels than building-related tampering. This is mainly because the random tampering of 16×16 pixels and the building-related tampering are both subject-related tampering, and the random tampering of 16×16 pixels will damage the information regarding the building in the RS image more strongly.

4.4.3. Analysis of Security

Subject-sensitive hashing requires the same security requirements as perceptual hashing, and both are required to meet the “one-way” security requirements; that is, the effective information of the original data cannot be obtained from the hash sequence.

Since the interpretability of deep neural networks is difficult, which means it is difficult for the current deep neural network to interpret the decision of the model from a human perspective [64,65], it is difficult to inversely obtain the effective information of the original RS image from the subject-sensitive hash sequence, which satisfies the security requirement of being “one-way”. Moreover, our algorithm fuses the features of each band before the feature compression and encoding, which further enhances the “one-way” nature of subject-sensitive hashing.

4.5. Discussion

Subject-sensitive hashing is a theory developed on the basis of perceptual hashing, which can be regarded as a special case of perceptual hashing. However, due to this concept only being proposed relatively recently, there are still many imperfections in the existing subject-sensitive hashing algorithms. In view of the existing main subject-sensitive hashing problems, such as an “inability to perceive content changes in different bands” and “insufficient algorithm robustness”, this study proposes a deep neural network named AAU-Net based on the attention mechanism for subject-sensitive hashing of RS images.

Due to the combination of characteristics of subject-sensitive hashing that requires the subject-sensitive features to be as robust as possible, the structure of AAU-Net presents

asymmetric characteristics: a multi-scale input is added in the encoder stage of AAU-Net to extract richer information from RS images, and attention gates are added to each module in the decoder stage to enhance the model's ability to extract subject-related features, which ultimately significantly improves the performance of the subject-sensitive hashing algorithm; furthermore, in the encoding stage of AAU-Net, there are four pooling layers, while there are only two upsampling layers in the decoding stage.

In the theoretical analysis and experiments of this paper, we used building information as an example to construct a training dataset and a dataset used to test the performance of subject-sensitive hashing algorithms. From the experimental results in Sections 4.3 and 4.4, we can draw the following conclusions:

(1) Robustness.

The asymmetric structure of AAU-Net makes it more robust than the most widely used models, such as U-Net, MA-U-Net, and Attention U-Net. Although the robustness of JPEG compression is slightly weaker than that of Attention U-Net, AAU-Net is better than the seven compared models in other robustness tests. The robustness of Attention ResUNet is second only to AAU-Net. Moreover, Attention R2U-Net is not recommended for subject-sensitive hashing due to its poor robustness.

(2) Tampering sensitivity.

Although the robustness of Attention R2U-Net is the worst among all the models, its tampering sensitivity is the best, according to the experimental results. This cannot change the fact that Attention R2U-Net is not suitable for subject-sensitive hashing, but the tampering sensitivity of Attention R2U-Net can be used as a benchmark for other models. The tampering sensitivity of AAU-Net is basically the same as that of Attention U-Net, the MA-U-Net-based algorithm, and MUM-Net, and second only to Attention R2U-Net. Additionally, Attention ResUNet's tampering sensitivity is relatively poor.

(3) Security.

The security of each comparison algorithm depends on the one-way nature caused by the inexplicability of the deep neural network, meaning there is no obvious difference in the security of each model.

Differing from the existing perceptual hash algorithms for RS images which perform the grayscale operation in the preprocessing stage, our algorithm extracts the features of each band and then fuses the feature images. Although this increases the computational cost of the algorithm, it makes the algorithm of this paper better than the existing subject-sensitive hashing algorithms (such as the algorithm in [16]). The most prominent performance is that for operations that only change one band, for which the robustness of our algorithm is greatly improved. For example, comparing Tables 7 and 8, it can be seen that our algorithm is far more robust to single-band watermarking algorithms than multi-band watermarking algorithms.

5. Conclusions and Future Work

In this research, we introduced the attention mechanism into subject-sensitive hashing and proposed an attention-based asymmetric U-Net (AAU-Net). Due to the combination of the characteristics of subject-sensitive hashing, such as the extracted features being as robust as possible, AAU-Net presents obvious asymmetric structure characteristics compared to Attention U-Net. The results show that the asymmetric structure of AAU-Net makes it more robust than existing deep learning models such as Attention U-Net and MUM-Net, and the tampering sensitivity is basically the same as that of Attention U-Net and MUM-Net.

However, the subject-sensitive hashing algorithm based on AAU-Net in this paper still has certain limitations, mainly including the weak robustness to JPEG compression, and the insufficient ability to detect subtle subject-related tampering. Therefore, the future focus of this research includes two aspects: improving the sensitivity of the algorithm

to subtle subject-related tampering, and improving the robustness of the algorithm to JPEG compression.

Author Contributions: K.D. conceived the idea and designed the scheme under the guidance of S.C.; Y.L. and Y.Z. assisted with the study design and the experiments; J.T. and Y.W. participated in the collection and collation of experimental data. All the authors reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by grants from (a) the National Natural Science Foundation of China (Grant Nos. 41801303, 42101428); (b) the Scientific Research Fund of Jinling Institute of Technology (Grant Nos. jit-fhxm-201604, jit-b-201520); and (c) the Qing Lan Project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Our training data set is constructed on the basis of WHU Building Dataset, which is available from this link: <http://gpcv.whu.edu.cn/data> (accessed on 12 December 2021). The codes used for this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Qu, Y.H.; Zhu, Y.Q.; Han, W.C.; Wang, J.D.; Ma, M.G. Crop Leaf Area Index Observations with a Wireless Sensor Network and Its Potential for Validating Remote Sensing Products. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 431–444. [\[CrossRef\]](#)
2. Mukesh, S.B.; Komal, C.H.; Rustam, P.; Alexander, K. Eco-environmental quality assessment based on pressure-state-response framework by remote sensing and GIS. *Remote Sens. Appl. Soc. Environ.* **2021**, *23*, 100530.
3. Tu, W.; Zhang, Y.T.; Li, Q.Q.; Mai, K.; Cao, J.Z. Scale Effect on Fusing Remote Sensing and Human Sensing to Portray Urban Functions. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 38–42. [\[CrossRef\]](#)
4. Huang, X.; Liu, H.; Zhang, L.P. Spatiotemporal Detection and Analysis of Urban Villages in Mega City Regions of China Using High-Resolution Remotely Sensed Imagery. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3639–3657. [\[CrossRef\]](#)
5. Natalia, R.; Conrado, M.; Rudorff, M.K.; Gustavo, O. Remote sensing monitoring of the impact of a major mining wastewater disaster on the turbidity of the Doce River plume off the eastern Brazilian coast. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 349–361.
6. Santos, L.B.L.; Carvalho, T.; Anderson, L.O.; Rudorff, C.M.; Marchezini, V.; Londe, L.R.; Saito, S.M. An RS-GIS-Based Comprehensive Impact Assessment of Floods—A Case Study in Madeira River, Western Brazilian Amazon. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1614–1617. [\[CrossRef\]](#)
7. Del’Arco, S.I.; Feitosa, R.Q.; Achanccaray, P.M.; Dias, M.B.L.; Alfredok, J.; Schultz, B.; Pinheiro, L.E. Campo Verde Database: Seeking to Improve Agricultural Remote Sensing of Tropical Areas. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 369–373. [\[CrossRef\]](#)
8. Steele-Dunne, S.C.; McNairn, H.; Monsivais-Huetero, A.; Judge, J.; Liu, P.W.; Papathanassiou, K. Radar Remote Sensing of Agricultural Canopies: A Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2249–2273. [\[CrossRef\]](#)
9. Liu, Q.; Zhai, G.J.; Lu, X.S. Integrated land-sea surveying and mapping of intertidal zone based on high-definition remote sensing images and GIS technology. *Microprocess. Microsyst.* **2021**, *82*, 103937. [\[CrossRef\]](#)
10. Eylül, M.; Marius, R.; Christian, G.; Lars, T.W. Countrywide mapping of trees outside forests based on remote sensing data in Switzerland. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *100*, 102336.
11. Niu, X.M.; Jiao, Y.H. An Overview of Perceptual Hashing. *Acta Electron. Sin.* **2008**, *36*, 1405–1411.
12. Du, L.; Ho, A.T.S.; Cong, R. Perceptual hashing for image authentication: A survey. *Sig. Process. Image Comm.* **2020**, *81*, 115713. [\[CrossRef\]](#)
13. Ding, K.M.; Zhu, Y.T.; Zhu, C.Q.; Su, S.B. A perceptual Hash Algorithm Based on Gabor Filter Bank and DWT for Remote Sensing Image Authentication. *J. China Railw. Soc.* **2016**, *38*, 70–76.
14. Zhang, X.G.; Yan, H.W.; Zhang, L.M.; Wang, H. High-Resolution Remote Sensing Image Integrity Authentication Method Considering Both Global and Local Features. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 254. [\[CrossRef\]](#)
15. Ding, K.; Chen, S.; Meng, F. A Novel Perceptual Hash Algorithm for Multispectral Image Authentication. *Algorithms* **2018**, *11*, 6. [\[CrossRef\]](#)
16. Ding, K.; Liu, Y.; Xu, Q.; Lu, F. A Subject-Sensitive Perceptual Hash Based on MUM-Net for the Integrity Authentication of High Resolution Remote Sensing Images. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 485. [\[CrossRef\]](#)
17. Zhang, X.; Nie, G.Z.; Huang, W.X.; Liu, W.X.; Ma, B.; Lin, C.W. Attention-guided image captioning with adaptive global and local feature fusion. *J. Vis. Commun. Image Represent.* **2021**, *78*, 103138. [\[CrossRef\]](#)
18. Chen, Z.; Li, D.; Fan, W.; Guan, H.; Wang, C.; Li, J. Self-Attention in Reconstruction Bias U-Net for Semantic Segmentation of Building Rooftops in Optical Remote Sensing Images. *Remote Sens.* **2021**, *13*, 2524. [\[CrossRef\]](#)

19. Kim, J.; Chi, M. SAFFNet: Self-Attention-Based Feature Fusion Network for Remote Sensing Few-Shot Scene Classification. *Remote Sens.* **2021**, *13*, 2532. [\[CrossRef\]](#)
20. Yu, J.K.; Yang, D.D.; Zhao, H.S. FFANet: Feature fusion attention network to medical image segmentation. *Biomed. Signal Process. Control* **2021**, *69*, 102912. [\[CrossRef\]](#)
21. Zhu, Y.S.; Zhao, C.Y.; Guo, H.Y.; Wang, J.Q.; Zhao, X.; Lu, H.Q. Attention CoupleNet: Fully Convolutional Attention Coupling Network for Object Detection. *IEEE Trans. Image Process.* **2019**, *28*, 113–126. [\[CrossRef\]](#)
22. Chen, X.; Weng, J.; Lu, W.; Xu, J.; Weng, J. Deep Manifold Learning Combined with Convolutional Neural Networks for Action Recognition. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 3938–3952. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Li, S.; Song, W.; Fang, L.; Chen, Y.; Ghamisi, P.; Benediktsson, J.A. Deep Learning for Hyperspectral Image Classification: An Overview. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6690–6709. [\[CrossRef\]](#)
24. Yuan, X.; Gu, Y.; Wang, Y.; Yang, C.; Gui, W. A Deep Supervised Learning Framework for Data-Driven Soft Sensor Modeling of Industrial Processes. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 4737–4746. [\[CrossRef\]](#)
25. Du, P.J.; Bai, X.Y.; Tan, K.; Xue, Z.H.; Samat, A.; Xia, J.S.; Li, E.Z.; Su, H.J.; Liu, W. Advances of Four Machine Learning Methods for Spatial Data Handling: A Review. *J. Geovis. Spat. Anal.* **2020**, *4*, 13. [\[CrossRef\]](#)
26. Wei, J.B.; Huang, Y.K.; Lu, K.; Wang, L.Z. Nonlocal Low-Rank-Based Compressed Sensing for Remote Sensing Image Reconstruction. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1557–1561. [\[CrossRef\]](#)
27. Cheng, Q.; Liu, H.Q.; Shen, H.F.; Wu, P.H.; Zhang, L.P. A Spatial and Temporal Nonlocal Filter-Based Data Fusion Method. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4476–4488. [\[CrossRef\]](#)
28. Liu, F.Y.; Chen, Z.Z. An Adaptive Spectral Decorrelation Method for Lossless MODIS Image Compression. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 803–814. [\[CrossRef\]](#)
29. Kulkarni, S.C.; Rege, P.P. Pixel level fusion techniques for SAR and optical images: A review. *Inf. Fusion* **2020**, *59*, 13–29. [\[CrossRef\]](#)
30. Peng, Y.D.; Li, W.S.; Luo, X.B.; Du, J.; Gan, Y.; Gao, X.B. Integrated fusion framework based on semicoupled sparse tensor factorization for spatio-temporal-spectral fusion of remote sensing images. *Inf. Fusion* **2021**, *65*, 21–36. [\[CrossRef\]](#)
31. Yang, X.; Li, S.S.; Chen, Z.C.; Chanussot, J.; Jia, X.P.; Zhang, B.; Li, B.P.; Chen, P. An attention-fused network for semantic segmentation of very-high-resolution remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *177*, 238–262. [\[CrossRef\]](#)
32. Zhao, Z.P.; Bao, Z.T.; Zhang, Z.X.; Deng, J.; Cummins, N.; Wang, H.; Tao, J.H.; Schuller, B. Automatic Assessment of Depression from Speech via a Hierarchical Attention Transfer Network and Attention Autoencoders. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 423–434. [\[CrossRef\]](#)
33. Ji, Z.; Li, S.J. Multimodal Alignment and Attention-Based Person Search via Natural Language Description. *IEEE Internet Things J.* **2020**, *7*, 11147–11156. [\[CrossRef\]](#)
34. Zhang, B.; Xiong, D.Y.; Xie, J.; Su, J.S. Neural Machine Translation With GRU-Gated Attention Model. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 4688–4698. [\[CrossRef\]](#)
35. Lu, X.Q.; Wang, B.Q.; Zheng, X.T. Sound Active Attention Framework for Remote Sensing Image Captioning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 1985–2000. [\[CrossRef\]](#)
36. Wang, B.; Wang, C.G.; Zhang, Q.; Su, Y.; Wang, Y.; Xu, Y.Y. Cross-Lingual Image Caption Generation Based on Visual Attention Model. *IEEE Access* **2020**, *8*, 104543–104554. [\[CrossRef\]](#)
37. Zhu, M.H.; Jiao, L.C.; Liu, F.; Yang, S.Y.; Wang, J.N. Residual Spectral-Spatial Attention Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 449–462. [\[CrossRef\]](#)
38. Xing, X.H.; Yuan, Y.X.; Meng, M.Q.H. Zoom in Lesions for Better Diagnosis: Attention Guided Deformation Network for WCE Image Classification. *IEEE Trans. Med. Imaging* **2020**, *39*, 4047–4059. [\[CrossRef\]](#)
39. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999. Available online: <https://arxiv.org/abs/1804.03999> (accessed on 20 May 2018).
40. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
41. Iglovikov, V.; Shvets, A. TernaUSNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation. *arXiv* **2018**, arXiv:1801.05746. Available online: <https://arxiv.org/abs/1801.05746> (accessed on 17 January 2018).
42. Xiao, X.; Lian, S.; Luo, Z.; Li, S. Weighted Res-UNet for High-Quality Retina Vessel Segmentation. In Proceedings of the 9th International Conference on Information Technology in Medicine and Education (ITME), Hangzhou, China, 19–21 October 2018; pp. 327–331.
43. Ibtehaz, N.; Rahman, M.S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Net.* **2020**, *121*, 74–87. [\[CrossRef\]](#)
44. Gu, Z.W.; Cheng, J.; Fu, H.Z.; Zhou, K.; Hao, H.Y.; Zhao, Y.T.; Zhang, T.Y.; Gao, S.H.; Liu, J. CE-Net: Context Encoder Network for 2D Medical Image Segmentation. *IEEE Trans. Med. Imaging* **2019**, *38*, 2281–2292. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Fu, H.; Cheng, J.; Xu, Y.; Wang, D.W.K.; Liu, J.; Cao, X. Joint Optic Disc and Cup Segmentation Based on Multi-Label Deep Network and Polar Transformation. *IEEE Trans. Med. Imaging* **2018**, *37*, 1597–1605. [\[CrossRef\]](#)
46. Guo, M.; Liu, H.; Xu, Y.; Huang, Y. Building Extraction Based on U-Net with an Attention Block and Multiple Losses. *Remote Sens.* **2020**, *12*, 1400. [\[CrossRef\]](#)

47. Tong, X.; Wei, J.; Sun, B.; Su, S.; Zuo, Z.; Wu, P. ASCU-Net: Attention Gate, Spatial and Channel Attention U-Net for Skin Lesion Segmentation. *Diagnostics* **2021**, *11*, 501. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2999–3007.
49. Le, N.; Bui, T.; Vo-Ho, V.-K.; Yamazaki, K.; Luu, K. Narrow Band Active Contour Attention Model for Medical Segmentation. *Diagnostics* **2021**, *11*, 1393. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [\[CrossRef\]](#)
51. Li, S.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. *Inf. Fusion* **2017**, *33*, 100–112. [\[CrossRef\]](#)
52. Lin, S.Z.; Han, Z.; Li, D.W.; Zeng, J.C.; Yang, X.L.; Liu, X.W.; Liu, F. Integrating model- and data-driven methods for synchronous adaptive multi-band image fusion. *Inf. Fusion* **2020**, *54*, 145–160. [\[CrossRef\]](#)
53. Ding, K.M.; Zhu, C.Q.; Lu, Q. An adaptive grid partition based perceptual hash algorithm for remote sensing image authentication. *Wuhan Daxue Xuebao* **2015**, *40*, 716–720.
54. Ji, S.P.; Wei, S.Y. Building extraction via convolutional neural networks from an open remote sensing building dataset. *Acta Geod. Cartogr. Sin.* **2019**, *48*, 448–459.
55. Adiga, V.; Sivaswamy, J. FPD-M-net: Fingerprint Image Denoising and Inpainting Using M-Net Based Convolutional Neural Networks. In *Inpainting and Denoising Challenges*; Springer: Cham, Switzerland, 2019; pp. 51–61.
56. Zhao, S.; Liu, T.; Liu, B.W.; Ruan, K. Attention residual convolution neural network based on U-net (AttentionResU-Net) for retina vessel segmentation. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *440*, 032138. [\[CrossRef\]](#)
57. Alom, M.Z.; Yakopcic, C.; Hasan, M.; Taha, T.M.; Asari, V.K. Recurrent residual U-Net for medical image segmentation. *J. Med. Imaging* **2019**, *6*, 014006. [\[CrossRef\]](#)
58. Hu, J.J.; Song, Y.; Zhang, L.; Bai, S.; Yi, Z. Multi-scale attention U-net for segmenting clinical target volume in graves' ophthalmopathy. *Neurocomputing* **2021**, *427*, 74–83. [\[CrossRef\]](#)
59. Chattopadhyay, S.; Basak, H. Multi-scale Attention U-Net (MsAUNet): A Modified U-Net Architecture for Scene Segmentation. *arXiv* **2020**, arXiv:2009.06911. Available online: <https://arxiv.org/abs/2009.06911> (accessed on 15 September 2020).
60. Ding, K.M.; Su, S.B.; Xu, N.; Jiang, T.T. Semi-U-Net: A Lightweight Deep Neural Network for Subject-Sensitive Hashing of HRRS Images. *IEEE Access* **2021**, *9*, 60280–60295. [\[CrossRef\]](#)
61. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 3974–3983.
62. Yuan, G.H.; Hao, Q. Digital watermarking secure scheme for remote sensing image protection. *China Commun.* **2020**, *17*, 88–98. [\[CrossRef\]](#)
63. Zhang, J.; Sang, L.; Li, X.P.; Wang, H.; Li, Y.S. Design and Implementation of Raw Data Compression System for Subsurface Detection SAR Based on FPGA. *J. Geovis. Spat. Anal.* **2020**, *4*, 2. [\[CrossRef\]](#)
64. Gao, X.; Mu, T.; Goulermas, J.Y.; Thiyaalingam, J.; Wang, M. An Interpretable Deep Architecture for Similarity Learning Built Upon Hierarchical Concepts. *IEEE Trans. Image Process.* **2020**, *29*, 3911–3926. [\[CrossRef\]](#)
65. Wu, C.; Gales, M.J.F.; Ragni, A.; Karanasou, P.; Sim, K.C. Improving Interpretability and Regularization in Deep Learning. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 256–265. [\[CrossRef\]](#)