Abstract: In today’s modern communication society, the authenticity of digital media has never been of such importance as it is now. In this aspect, the reliability of digital images is of paramount importance because images can be easily manipulated by means of sophisticated software, such as Photoshop. Splicing tampering is a commonly used photographic manipulation for modifying images. Detecting splicing tampering remains a challenging task in the area of image forensics. A new multitask model based on attention mechanism, densely connected network, Atrous Spatial Pyramid Pooling (ASPP) and U-Net for locating splicing tampering in an image, AttDAU-Net, was proposed. The proposed AttDAU-Net is basically a U-Net that incorporates the spatial rich model filtering, an attention mechanism, an ASPP module and a multitask learning framework, in order to capture more multi-scale information while enlarging the receptive field and improving the detection precision of image splicing tampering. The experimental results on the datasets of CASIA1 and CASIA2 showed promising performance metrics for the proposed model ($F_1$-scores of 0.7736 and 0.6937, respectively), which were better than other state-of-the-art methods for comparison, demonstrating the feasibility and effectiveness of the proposed AttDAU-Net in locating image splicing tampering.

Keywords: image authenticity; image forgery; detection of splicing tampering; attention mechanism; multitask learning

MSC: 68T07; 98A08

1. Introduction

With the popularity of photographic equipment such as digital cameras and smartphones, people are sharing a lot of images on the Internet. However, with the increasing functionality of image editing software such as Photoshop, it is easy for illegal users to tamper with images and deceive the public [1]. The number of image forgeries is so high that people often doubt the authenticity of commercial propaganda, photo contests, court forensics [2], etc. Digital image forensics, including the active forensics and passive forensics, aim to detect such image forgeries. In the technology of active forensics, images are embedded with authentication information such as digital watermarking and digital signature. When an image is tampered with to some extent, the embedded authentication information is also tampered with and can be detected. From the perspective of resistance to attacks, digital watermarking can be further divided into robust watermarking and fragile watermarking. Robust watermarking is capable of resisting attacks to some extent and is often used for copyright protection [3], while fragile watermarking is very sensitive such that even slight changes in the image may damage the watermark [4]. The technology of digital signature is a mathematical scheme used to show the authenticity of a digital image or document [5]. The recipient of the image can be confident that the message was
created by a known sender and has not been altered during the process of transmission. The passive forensics for digital images identify the authenticity and integrity of an image by analyzing the inherent attributes of the image. Since the tampering process will inevitably leave some traces of attribute destruction, these traces can be used to determine the authenticity of the image and locate the tampered areas. Passive forensics does not rely on any pre-embedded information for tampering detection, which is a more realistic forensic method in the face of large amounts of image tampering in various image application fields. Therefore, this manuscript focuses on passive forensics of digital images.

Currently, digital image forensics can be divided into two categories: (1) classical feature extraction-based method and (2) deep learning-based method. The first category includes a method based on noise patterns, a method based on JPEG domain features and an interpolation model based on color filter arrays. Usually, images from different origins should have different noise patterns produced by image sensors or post-processing tools. Based on this, in 2020, Liu et al. [6] proposed a novel splicing tampering detection method by analyzing the noise discrepancy to locate splicing tempering. Specifically, Liu proposed an adaptive singular value decomposition to estimate local noise and vicinity noise descriptors to locate splicing tampering. Experimental results showed that Liu’s method was able to locate multiple objects from different origins. Principal component analysis (PCA) was also used to estimate the noise level. Zeng et al. [7] first conducted a block-by-block noise level estimation using a PCA-based algorithm, and then segmented the tampered area by k-means clustering. Experimental results showed the superiority of Zeng’s method in practical splicing, with small noise discrepancies between the original and tampered areas. The method in JPEG domain assumes that an image obtained from JPEG compression usually has a block artifact. If two regions in an image come from different JPEG compression sources, the partitioned sub-blocks may not be correctly aligned [8]. Zhang et al. [9] utilized higher-order co-occurrence statistics to model the underlying dependencies of the JPEG-quantized DCT coefficients and proposed an efficient algorithm to reveal the traces of non-aligned double JPEG compressed images. Another feature-based method is that based on color filter array (CFA) interpolation. Different cameras are likely to use different CFA interpolation modes, which can be used to detect image splicing. Wang et al. [10] proposed a progressive image splicing tampering detection method that can detect the position and shape of the spliced region. Wang first used a covariance matrix to recover the image R, G and B channels and utilized the inconsistencies of the CFA interpolation patterns to extract forensic features. These forensic features were then used to perform a coarse-grained detection, with the textures being used to perform a fine-grained detection. Finally, an edge smoothing was applied to implement the precise image splicing positioning.

However, the above-mentioned classical methods tend to target a specific tampering mode and have certain limitations. For example, the method based on noise patterns cannot detect image tampering when multiple lossy compressions are severally applied to the tampered image; the detection method in JPEG domain is not applicable to uncompressed images or images compressed by other compression methods. The method based on CFA interpolation detection assumes that the tampered region and the background come from two different cameras. In real life, it is often difficult to determine the tampering mode for a tampered image, so the classical detection methods are no longer applicable.

Deep learning [11] has made significant progress in image forgery detection by enabling machines to imitate the human brain, such as hearing and thinking, and solving many complex pattern recognition problems [12–16]. With the development of convolutional neural networks (CNNs) [17] and fully convolutional networks (FCNs) [18], semantic image segmentation has found a wide range of applications in image forgery detection. In 2016, Bayar et al. [19] proposed a deep learning approach for general image forgery detection using CNNs. Specifically, Bayer developed a new form of convolutional layer capable of automatically learning manipulation detection features directly from the training data, which could detect several different manipulations with high accuracy. Rao
et al. [20] proposed a new CNN-based image forgery detection algorithm for learning hierarchical representations from input images. The weights in the first layer of the CNN were initialized with a high-pass filter used in the spatial rich model (SRM) [21], which served as a regularizer to suppress the effects of image content and capture subtle tampering artifacts. Experimental results on several public datasets have demonstrated the superiority of the proposed CNN-based model over other state-of-the-art methods. In 2019, Wu [22] proposed a novel and unified end-to-end fully deep convolutional neural architecture, ManTra-Net, for performing both detection and localization for different types and combined image manipulations. It first extracts manipulation traces and then identifies abnormal areas by means of the differences between a local feature and its reference feature. Extensive experimental results have demonstrated the generalization ability, robustness and superiority of ManTra-Net, not only in single types of manipulations/forgeries, but also in their combinations, and even in unknown types. In 2020, Bi et al. [23] proposed a splicing forgery detection method in two steps (a coarse-to-refined CNN and a diluted adaptive clustering) to extract the differences in image properties between un-tampered and tampered regions from image patches. After locating the suspicious forgery regions in the first step, the final forgery regions were detected in the second step. Experimental results showed Bi’s two-step model achieved promising results compared to state-of-the-art splicing forgery detection methods, even under various attacks.

Despite the great progress in deep learning-based image manipulation detection, there are currently many challenges and issues to be addressed. The application of deep learning in image tampering detection is still a relatively new research area; for example, the performance of detection has still space to improve, and a very deep neural network has the risk of overfitting. In this manuscript, we aim to overcome these shortcomings in two aspects: (1) a proposal of a multitask learning network; and (2) a U-Net-like architecture that combines an attention mechanism (AM), a densely connected neural network (DenseNet), and an Atrous Spatial Pyramid Pooling (ASPP), namely AttDAU-Net. The main contributions of this manuscript are as follows. (1) The proposed multitask learning network enables a simultaneous two-task learning of both tampered area detection and tampered boundary detection, (2) the incorporation of the SRM filters realizes an efficient residual noise extraction and hence facilitates the tampered area detection, (3) the ASSP introduced adapts the model to tampered regions of various sizes and shapes and (4) the channel and spatial attention mechanism utilized makes the proposed model focus on an important subset of the feature map and capture informative features. Experimental results on popular open datasets CASIA1.0 and CASIA2.0 showed promising results of performance metrics such as detection precision, recall and $F_1$-score. Specifically, the precision and recall of the proposed multitask learning model, AttDAU-Net, were better than most other methods for comparison, while the $F_1$-score was better than all other methods for comparison. In addition, the proposed AttDAU-net exhibits some robustness to image compression and blurring attacks.

The rest of this manuscript is organized as follows. Section 2 reviews some fundamentals and related work. Section 3 describes the proposed method. Simulation results and evaluation are presented in Section 4. Finally, conclusions are drawn in Section 5.

2. Fundamentals and Related Work

Convolutional neural network, attention mechanism, residual noise extraction and multitask learning are the important steps of the proposed model in this paper. A brief review of these frameworks is presented as follows.

2.1. Convolutional Neural Network

The convolutional neural network (CNN) [11,17] is a special type of neural network that is usually computationally efficient and applicable to image-related tasks, such as image classification, target detection, object segmentation and medical scenarios. From AlexNet [17] to ConvNeXt [24], CNN has experienced more than a decade of development
history. A CNN is generally composed of stacked convolutional layers with learnable non-linear activation layers, pooling layers and a fully connected layer. The convolutional layer applies a number of convolutional filters to the image. The filter slides over the image and performs a weighted sum to produce a single value in the output feature map. The pooling layer downsamples the convolution results extracted by the convolution layer to reduce the dimensionality of the feature map. The commonly used pooling methods are max pooling, average pooling and stochastic pooling. The filter is also named as a kernel with learnable coefficients and biases. Multiple filters result in multiple channels of the output feature map. The activation function provides the nonlinear transformation capability required by the network. In recent years, important CNN achievements include U-Net [25], deep residual network (ResNet) [26], DenseNet [27], ASPP [28] and AM [29]. An atrous convolution generates features with large receptive fields without damaging the spatial resolution, while ASPP concatenates several atrous-convolved features with different dilation rates to produce multi-scale features.

2.2. Attention Mechanism

Attention plays an important role in human perception. Humans always selectively focus on the salient parts when observing the whole scene. Attention mechanisms give CNNs the ability to focus on a subset of a feature map. As a result, they allow CNNs to approximate more complicated functions. In principle, there are three kinds of attention mechanisms, namely spatial attention, channel attention and spatial-channel attention mechanisms. Woo et al. [29] proposed a known convolutional block attention module (CBAM) that derives attention maps along both channel and spatial dimensions, with the attention maps being used to refine the features in the input feature map. Experimental results on ImageNet-1K, MS COCO and VOC 2007 datasets showed consistent improved classification performance. Following the same idea of fusing channel and spatial attentions, in 2021, Gan et al. [30] proposed a global attention mechanism (GAM), which combines channel and spatial attention modules and integrates different convolutions in the GAM. Gan proposed a new global attention network, GAU-Net, by combining GAM modules with U-Net. Experiments on the brain tumor segmentation dataset BraTS2018 showed that GAU-Net increased the mean intersection over union (mIoU) from 0.65 to 0.75, with number of network parameters accounting for only 5.4% of that of U-Net. Another kind of spatial attention mechanism is the feature pyramid attention (FPA) [31] network proposed by Li et al. in 2018 to exploit the impact of global contextual information on semantic segmentation. An FPA module was introduced on each decoder layer to provide global context as guidance for low-level features to localize detailed category information. A new mIoU record of 84.0% on the PASCAL VOC2012 dataset was achieved by the FPA-based model.

In recent years, attention mechanisms have made important breakthroughs in areas such as image classification, target detection and natural language processing, and have proven to be beneficial in improving model performance in many application scenarios.

2.3. Residual Noise Extraction

Unlike common semantic segmentation models that focus on semantic image content, an image tampering detection model typically learns the difference between tampered and untampered regions. This is somewhat similar to image steganalysis, which concentrates on hiding information rather than the image content itself. In early 2012, for an image $x(i, j), i = 0, 1, 2, \ldots, M - 1, j = 0, 1, 2, \ldots, N - 1$, Fridrich et al. [21] proposed an SRM for computing the residual image noise component:

$$r(i, j) = \hat{r}(i, j) - cx(i, j)$$  \hspace{1cm} (1)

where $\hat{r}(i, j)$ is an estimation of $cx(i, j)$ defined over the neighborhood of $x(i, j)$ and $c$ is the residual order. The advantage of using residual values instead of pixel values is that there is a large suppression of the image content. To improve the sensitivity of the residuals at
spatial discontinuities such as edges and textures, the dynamic range of the residuals is narrowed by a quantization, round-off and truncation operation:

\[ r(i, j) \leftarrow \text{trunc}_T \left( \text{round} \left( \frac{r(i, j)}{q} \right) \right) \]  

(2)

where \( q > 0 \) is the quantization step and \( T \) is the truncation threshold. The best performance is achieved when \( q \in [c, 2c] \). Based on the results of (2), SRM extracts the nearby co-occurrence information as the final features.

In 2018, Zhou et al. [32] found that sufficiently good performance was only achieved by using these SRM filters, as follows:

\[
K_1 = \frac{1}{4} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 2 & -4 & 2 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad
K_2 = \frac{1}{12} \begin{bmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & -1 \end{bmatrix}, \quad
K_3 = \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.
\]  

(3)

2.4. Multitask Learning

Multitask learning (MTL) refers to multiple related learning tasks by exploiting useful information among them [33]. In 1997, Caruana [34] defined multitask learning as an approach to inductive transfer that improves generalization by leveraging domain information contained in the training datasets of related tasks. All tasks help each other in learning a shared representation in parallel. In this pioneering work, Caruana demonstrated multitask learning in three domains, namely an ALVINN-like [35] road-following domain, a real data domain collected using a robot-mounted camera, and a medical decision-making domain. Since multitask learning can be used in many different kinds of domains with different learning algorithms, Caruana predicted that there would be many applications of MTL to real-world problems.

In the context of deep learning, MTL learns shared representations from multitask datasets. In 2020, Vandenhende et al. [36] classified deep MTL architectures into hard and soft parameter sharing (PS) architectures, as shown in Figure 1. In the hard PS (Figure 1a), the parameter set is divided into shared and task-specific parameters, and MTL models using the hard PS generally consist of a shared encoder that branches out into task-specific heads. PS exists in the lower layers. After the PS layers, different tasks correspond to different branches, and these tasks are trained in parallel with each other. Therefore, the model is not restricted to learning a single task, but multiple application scenarios, thus enhancing the generalization capability of the model to a large extent. In the soft PS (Figure 1b), each task is assigned with its own model and parameters, and the feature sharing mechanism processes cross-talk. The soft sharing mechanism is flexible and does not require task-dependent assumptions. However, since each task has its own model, more parameters are required, and these parameters are set empirically, which has limitations in practical use. In contrast, hard sharing is simpler to implement, as it is still the most popular MTL architecture.
In recent years, deep MTL models have developed rapidly in computer vision and natural language processing. MTL leverages task-specific information contained in related task branches to improve the performance of each task. For the detection of splicing tampered images, the traces of tampering in an image are generally manifested as unnatural transitions in the tampered edges and inconsistencies in the residual noise from the tampered areas. These are two different but related tasks. Therefore, in this manuscript, we choose the commonly used hard PS as the MTL architecture to simultaneously learn two tasks, namely the detection of tampered boundary and the detection of tampered area, in order to obtain the optimal model performance.

3. AttDAU-Net—Proposed Multitask Splicing Tampering Detection Model

3.1. The Model

We propose a multitask splicing tampering detection model based on AM, DenseNet, ASPP and U-Net, named AttDAU-Net, whose structure is shown in Figure 2. In general, AttDAU-Net is a U-Net-like architecture consisting of three parts, namely a PS encoder (PS network) and two task-specific decoders (the tampered region detection network and the tampered region’s boundary detection network). An SRM and a normal CNN are placed parallel to each other in the front of the PS network. The SRM filter extracts the residual noise in the tampered images while suppressing the interference of semantic information in the original image. The normal CNN completes an initial feature extraction. The outputs of the SRM and CNN are concatenated and then fetched into a stack of three densely connected blocks (DenseBlocks) consisting of 6, 12 and 16 layers, respectively. The introduction of DenseBlocks alleviates the vanishing-gradient problem, strengthens feature propagation, encourages feature reuse and greatly reduces the number of network parameters. All these characteristics help to capture rich tampered semantic features. A transition module follows each DenseBlock for adapting the feature map size to the next DenseBlock. In the PS network, the features are extracted and shared by the two subsequent tasks. Through the PS, the informative data of both tasks are embedded in the same semantic space, which helps to reduce the risk of network overfitting.

![Figure 1. Two deep MTL architectures: (a) hard parameter sharing, and (b) soft parameter sharing.](image-url)
The tampered region detection network consists of a 12-layer pre-DenseBlock, an ASPP module and a three-step expansive path (lower horizontal path in Figure 2). The pre-DenseBlock continues the down-sampling following the previous three DenseBlocks in the encoder. The ASPP is introduced for adapting the model to tampered regions of different sizes and shapes because the atrous convolution in the ASPP has various dilation rates and facilitates the acquisition of multiscale receptive fields and multiscale tampering-related information. Each step in the expansive path consists of a feature map upsampling module (implemented by transposed convolution), a concatenation with the corresponding global attention map from the GAU module [30], and a stack of two convolutional layers. The last block is a $1 \times 1$ convolutional layer that restores the single-channel feature map output.

The boundary detection network of the tampered region consists of an FPA module and an upsampling block. Through the FPA module, a better feature representation of tampered images was learned from the output feature map of the PS network, and a final binary image presenting the boundary of the tampered region was obtained after a 4-times upsampling.

3.2. Loss Functions

For the MTL hard parameter sharing mechanism, the two tasks have their own branches and their own loss functions that are back-propagated to the PS layers during the process of network training. The total loss is the weighted sum of the two loss functions:

$$L_{loss} = L_{region} + \alpha L_{edge}$$

where $L_{region}$ is the loss function for the tampered region detection-specific task, $L_{edge}$ is the loss function for the tampered region’s boundary detection-specific task, and $\alpha$ is the
balance factor, which is empirically set to 0.25. The binary cross-entropy function was chosen for both $L_{\text{region}}$ and $L_{\text{edge}}$.

4. Results and Discussion
4.1. Development Environment, Experimental Settings and Datasets

The proposed AttDAU-Net was implemented in a workstation under Windows 10 with an i9-10920X CPU, 64 GB of RAM, two GeForce RTX 3090Ti GPUs and 24 GB of video memory. The open PyTorch was chosen as the deep learning library. The optimizer used the stochastic gradient descent method, with the initial learning rate set to 0.05, the moment set to 0.9, and the weight decay set to 0.0005. The learning rate decay strategy adopted the fixed decay strategy, being half of the previous stage after every 10 training epochs. The total number of training epochs was 100.

The datasets used were open datasets commonly employed in the area of image tampering detection, namely CASIA1.0 and CASIA2.0 [37]. The label images for the boundary detection of tampered regions were obtained by performing the mathematical morphological operations of dilation and contraction on the label images used for tampered region detection. Of the 1721 images in CASIA1.0, 921 were tampered images with corresponding labels and 800 were real images. The tampered regions have various shapes, such as circles, triangles and rectangles. Some examples of images in CASIA1.0 are shown in Figure 3.

![Figure 3. Examples of images in CASIA1.0: (a,b) real images, (c) tampered images and (d) label images.](image)

CASIA2.0 is larger than CASIA1.0 and includes 5123 tampered images with corresponding labels and 7491 real images. Among the tampered images, there are 1760 splicing tampered images. Columbia is a splicing-only tampered image dataset, which includes 183 real images and 180 tampered images. CASIA1.0 was used for training, and CASIA2.0 was used for testing.

4.2. Performance Evaluation Metrics

The performance evaluation metrics for tampered region detection in this paper are Precision ($P$), Recall ($R$, Sensitivity) and $F_1$-score. They are defined as:

$$P = \frac{TP}{TP + FP}$$  \hspace{1cm} (5)

$$R = \frac{TP}{TP + FN}$$  \hspace{1cm} (6)

$$F_1 = \frac{2PR}{P + R}$$  \hspace{1cm} (7)
where \( TP \) (True positive) is the number of successfully detected tampered pixels, \( FP \) (False positive) is the number of real pixels that are mis-detected as tampered pixels, and \( FN \) (False negative) is the number of tampered pixels that are mis-detected as real pixels.

Precision is the ratio of true positives over all predictive positives, while Recall is the ratio of true positives over all positives. The \( F_1 \)-score takes both Precision and Recall into account. The \( F_1 \)-score is often a better measure to use when one of Precision and Recall is high and the other is low, as it balances these two.

### 4.3. Comparative Results

To verify the performance of the proposed splicing tampering detection model AttDAU-Net, extensive tests were conducted using different tampering detection methods, and the comparative results are presented in Table 1, Figures 4h and 5h. In Table 1, ELA [38] is an error level analysis method that aims to detect differences between tampered and real images by detecting regions with different compression ratios in a JPEG image. Ye’s method [39] is a simple passive approach for detecting the inconsistencies of blocking artifacts caused by JPEG compression. Ye proposed an effective blocking artifact measure to reveal forgeries of digital images. FCNS [18], DeepLabV3 [28], PSPNet [40] and U-Net [25] are currently popular semantic segmentation models. DAU-Net is the same as the proposed AttDAU-Net, but without an attention mechanism. As shown in Table 1, for the CASIA1.0 and CASIA2.0 datasets, the proposed image tampering detection model, AttDAU-Net, had the best \( F_1 \)-score performance among all the methods used for comparison. ELA had the highest Recall at the cost of low Precision. DAU-Net had the best Precision at the cost of low Recall. Compared to DAU-Net, the proposed model increased the Recall on CASIA1.0 and CASIA2.0 by 8.36% and 2.79%, respectively. This indicates that the attention mechanism introduced in the proposed model plays an important role in improving the performance of the model. Figures 4a and 5a show the four tampered regions in the four images, respectively. As can be seen in Figures 4h and 5h, all four tampered regions were successfully detected.

---

![Figure 4](https://example.com/figure.png)

**Figure 4.** Ablation results of two images from CASIA1.0: (a) real images, (b) tampered images, (c) label image, (d) basic model, (e) basic model + SRM, (f) basic model + GAU, (g) basic model + SRM + GAU and (h) AttDAU-Net.
Figure 5. Ablation results of two images from CASIA2.0: (a) real images, (b) tampered images, (c) label image, (d) basic model, (e) basic model + SRM, (f) basic model + GAU, (g) basic model + SRM + GAU and (h) AttDAU-Net.

Table 1. Comparative results of different tampering detection methods on CASIA1.0 and CASIA2.0.

| Methods     | CASIA1.0       | CASIA2.0       |
|-------------|----------------|----------------|
|             | Precision | Recall | F1-Score | Precision | Recall | F1-Score |
| ELA         | 0.1242    | 0.9147  | 0.2188   | 0.0971    | 0.8864  | 0.1751   |
| Ye’s method | 0.2305    | 0.8272  | 0.3605   | 0.1784    | 0.7798  | 0.2904   |
| FCNS        | 0.7763    | 0.4610  | 0.5785   | 0.6482    | 0.4251  | 0.5135   |
| PSPNet      | 0.7119    | 0.4953  | 0.5842   | 0.4372    | 0.3775  | 0.4052   |
| DeepLabV3   | 0.7487    | 0.3738  | 0.4987   | 0.5525    | 0.3760  | 0.4475   |
| U-Net       | 0.7607    | 0.5337  | 0.6273   | 0.7404    | 0.4654  | 0.5716   |
| DAU-Net     | 0.8365    | 0.6765  | 0.7481   | 0.7707    | 0.6114  | 0.6819   |
| AttDAU-Net  | 0.7876    | 0.7601  | 0.7736   | 0.7582    | 0.6393  | 0.6937   |

4.4. Ablation Study

An ablation study is often used to examine the importance of each component of a deep learning model to the whole. From the ablation study and the individual components’ performance test, one can observe and analyze the influence of each separate module on the whole model, and meantime, identify the most important enhancement components or some modules that have little impact on the performance of the model, in order to simplify the model and improve efficiency. Five groups of tests were conducted to test the proposed model. The results of the ablation experiments are shown in Table 2, Figures 4 and 5. In Table 2, the basic model refers to the simplest model by removing the SRM, GAU and boundary detection sub-net (BD-Net) from the AttDAU-Net. As can be seen in Figures 4 and 5, the proposed AttDAU-Net not only detected all tampered regions of all images, but also obtained the best visual detection results among all the methods used for comparison.

As can be seen from Table 2, the addition of the SRM filter greatly improved the detection precision of CASIA1.0 and CASIA2.0 by 6.3% and 7.1%, respectively, compared to the basic model, since the SRM filter on the front side could amplify the residual noise and textures in the tampered regions. The addition of the GAU module improved the Recall and F1-score to some extent, as the low-level features became more sensitive after being highlighted in the attention module, facilitating the screening of important tampering-related information. The boundary detection task led to an increase in F1-score by about 2.7% and 3.4% on CASIA1.0 and CASIA2.0, respectively, because the two tasks shared parameters in the PS network and were trained simultaneously and complemented each other, that is, they helped to effectively alleviate overfitting and ultimately improve model performance.
Table 2. Comparative results by removing different components from the basic model.

| Method                  | CASIA1.0 |               | CASIA2.0 |               |
|-------------------------|----------|---------------|----------|---------------|
|                         | Precision| Recall        | F1-Score | Precision     | Recall        | F1-Score     |
| Basic model             | 0.7159   | 0.6539        | 0.6834   | 0.7673        | 0.5393        | 0.6334       |
| Basic model + SRM       | 0.7868   | 0.6472        | 0.7102   | 0.8311        | 0.5501        | 0.6620       |
| Basic model + GAU       | 0.7316   | 0.6672        | 0.6979   | 0.7680        | 0.5675        | 0.6527       |
| Basic model + SRM + GAU | 0.7954   | 0.7081        | 0.7492   | 0.7381        | 0.6141        | 0.6704       |
| AttDAU-Net              | 0.7876   | 0.7601        | 0.7736   | 0.7582        | 0.6393        | 0.6937       |

4.5. Robustness to Compression and Blurring Attacks

To verify the robustness of the proposed model against compression and blurring, we conducted JPEG image compression and Gaussian blurring on the tampered images. Table 3 presents the tampered region detection results under different compression quality factors and different standard deviations of Gaussian kernels.

Table 3. Tampering detection results of CASIA1.0 under compression and blurring attacks.

| Attack        | Metrics | Compression (Quality Factor) | Gaussian Blurring (Standard Deviation) |
|---------------|---------|------------------------------|----------------------------------------|
|               |         | 95%                          | 90%                                   | 80%          | 70%          | 0.5          | 1.0          | 1.5          | 2.0          |
| Precision     |         | 0.7795                       | 0.7699                                 | 0.6143       | 0.5667       | 0.7742       | 0.7548       | 0.7362       | 0.7265       |
| Recall        |         | 0.6827                       | 0.5765                                 | 0.4830       | 0.4962       | 0.7554       | 0.7618       | 0.7571       | 0.7563       |
| F1-score      |         | 0.7279                       | 0.6593                                 | 0.5408       | 0.5291       | 0.7647       | 0.7583       | 0.7465       | 0.7411       |

As can be seen from Table 3, compared with the results in Table 1, the tampering detection results of the proposed model on CASIA1.0 retained relatively high Precision, Recall and F1-score values under different levels of image compression and blurring. This robustness indicates that through parameter sharing, the MTL network can reduce the risk of overfitting and hence improve the robustness against various attacks.

5. Conclusions

Splicing tempering is one of the commonly encountered image manipulations. Detection of image splicing tampering has never been of such importance as it is now. This manuscript proposes a new MTL model, AttDAU-Net, based on FPA and GAM. AttDAU-Net integrates U-Net, SRM filter and ASSP with channel and spatial attention mechanisms in the MTL model, so as to capture more important information and improve image slicing tampering detection performance. The MTL aims to integrate the two tasks of tampered region detection and boundary detection by embedding the data information of both tasks into a single semantic space, thus reducing the risk of network overfitting. Experimental results on popular open datasets in the area of tampering detection demonstrate that the proposed AttDAU-Net outperforms several other common tampering detection methods. The ablation study shows the effectiveness of the components introduced in the basic model, such as the SRM filter and the GAU module. Experimental results on popular open datasets such as CASIA1.0 and CASIA2.0 showed promising results of performance metrics such as detection precision, recall and F1-score. The precision (0.7876 and 7582) and recall (0.7671 and 0.6393) of the proposed multitask learning model, AttDAU-Net, were better than most other methods for comparison, respectively for CASIA1.0 and CASIA2.0, while the F1-scores (0.7736 and 0.6937) were better than all other methods for comparison, respectively for CASIA1.0 and CASIA2.0. In addition, the proposed AttDAU-Net exhibits some robustness to image compression and blurring attacks. For further research in the future, the data imbalance could be taken into account and this work would be extended to simultaneous detection of multiple tampering types.
Author Contributions: Conceptualization, P.Z. and N.Z.; methodology, P.Z. and J.W.; software, L.T. and J.W.; validation, L.T., N.Z. and J.W.; formal analysis, P.Z.; investigation, Y.L. and J.W.; resources, Y.L. and J.W.; data curation, L.T.; writing—original draft preparation, P.Z.; writing—review and editing, J.W. and N.Z.; visualization, L.T.; supervision, N.Z.; project administration, J.W.; funding acquisition, J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, Grant Nos. 61662047 and 62041106.

Data Availability Statement: Implementation code for Python is available upon request.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

References
1. Barad, Z.J.; Goswami, M.M. Image forgery detection using deep learning: A survey. In Proceedings of the 6th International Conference on Advanced Computing and Communication Systems, Tamil Nadu, India, 6–7 March 2020.
2. Thakur, R.; Rohilla, R. Recent advances in digital image manipulation detection techniques: A brief review. Forensic Sci. Int. 2020, 312, 110331. [CrossRef][PubMed]
3. Xiang, S.J.; Ruan, G.Q.; Li, H.; He, J. Robust watermarking of databases in order-preserving encrypted domain. Front. Comput. Sci. 2021, 16, 162804. [CrossRef]
4. Shehab, A.; Elhoseny, M.; Muhammad, K.; Sangaiah, A.K.; Yang, P.; Huang, H.; Hou, G. Secure and robust fragile watermarking scheme for medical images. IEEE Access 2018, 6, 10269–10278. [CrossRef]
5. Fang, W.; Chen, W.; Zhang, W.; Pei, J.; Gao, W.; Wang, G. Digital signature scheme for information non-repudiation in blockchain: A state-of-the-art review. EURASIP J. Wirel. Commun. Netw. 2020, 1, 1–15. [CrossRef]
6. Liu, B.; Pun, C.M. Locating splicing forgery by adaptive-SVD noise estimation and vicinity noise descriptor. Neurocomputing 2020, 387, 172–187. [CrossRef]
7. Zeng, H.; Zhan, Y.; Kang, X.; Lin, X. Image splicing localization using PCA-based noise level estimation. Multimed. Tools Appl. 2017, 76, 4783–4799. [CrossRef]
8. Bianchi, T.; Piva, A. Image forgery localization via block-grained analysis of JPEG artifacts. IEEE Trans. Inf. Forensics Secur. 2012, 7, 1003–1017. [CrossRef]
9. Zhang, Y.; Song, W.; Wu, F.; Han, H.; Zhang, L. Revealing the traces of nonaligned double JPEG compression in digital Images. Optik 2020, 204, 164196. [CrossRef]
10. Wang, X.; Wang, Y.; Lei, J.; Li, B.; Wang, Q.; Xue, J. Coarse-to-fine-grained method for image splicing region detection. Pattern Recognit. 2022, 122, 108347. [CrossRef]
11. LeCun, Y.; Bengio, Y.; Hinton, G.E. Deep learning. Nature 2015, 521, 436–444. [CrossRef]
12. Hu, A.; Razmjooy, N. Brain tumor diagnosis based on metaheuristics and deep learning. Int. J. Imaging Syst. Technol. 2021, 31, 657–669. [CrossRef]
13. Sharma, P.; Diwakar, M.; Choudhary, S. Application of edge detection for brain tumor detection. Int. J. Comput. Appl. 2012, 58, 359–364. [CrossRef]
14. Arabahmadi, M.; Farahbaksh, R.; Rezaazadeh, J. 12-Deep learning for smart healthcare—A survey on brain tumor detection from medical imaging. Sensors 2022, 22, 1960. [CrossRef]
15. Tian, Q.; Wu, Y.; Ren, X.; Razmjooy, N. A new optimized sequential method for lung tumor diagnosis based on deep learning and converged search and rescue algorithm. Biomed. Signal Process. Control 2021, 68, 102761. [CrossRef]
16. Soomro, T.A.; Zheng, L.; Afifi, A.J.; Ali, A.; Soomro, S.; Yin, M.; Gao, J. Image segmentation for MR brain tumor detection using machine learning: A Review. IEEE Rev. Biomed. Eng. 2022, early access. [CrossRef]
17. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. Commun. ACM 2017, 60, 84–90. [CrossRef]
18. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2015, 39, 640–651.
19. Bayar, B.; Stamm, M.C. A deep learning approach to universal image manipulation detection using a new convolutional layer. In Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, New York, NY, USA, 20–22 June 2016.
20. Rao, Y.; Ni, J. A deep learning approach to detection of splicing and copy-move forgeries in images. In Proceedings of the 2016 IEEE International Workshop on Information Forensics and Security, Abu Dhabi, United Arab Emirates, 4–7 December 2016.
21. Fridrich, J.; Kodovsky, J. Rich models for steganalysis of digital images. IEEE Trans. Inf. Forensics Secur. 2012, 7, 868–882. [CrossRef]
22. Wu, Y.; Abdalmageed, W.; Natarajan, P. ManTra-Net: Manipulation tracing network for detection and localization of image forgeries with anomalous features. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019.
23. Xiao, B.; Wei, Y.; Bi, X.; Li, W.; Ma, J. Image splicing forgery detection combining coarse to refined convolutional neural network and adaptive clustering. Inf. Sci. 2020, 511, 172–191. [CrossRef]

24. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11976–11986. [CrossRef]

25. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015; Lecture Notes in Computer Science; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [CrossRef]

26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 20–24 June 2016.

27. Huang, G.; Liu, Z.; Maaten, L.V.D. Densely connected convolutional networks. arXiv 2018, arXiv:1608.06993v5.

28. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the 2018 European Conference on Computer Vision, Munich, Germany, 8–14 September 2018.

29. Woo, S.H.; Park, J.C.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the 2018 European Conference on Computer Vision, Munich, Germany, 8–14 September 2018.

30. Gan, X.; Wang, L.; Chen, Q.; Ge, Y.; Duan, S. GAU-Net, U-Net based on global attention mechanism for brain tumor segmentation. J. Phys. Conf. Ser. 2021, 1861, 012041. [CrossRef]

31. Li, H.; Xiong, P.; An, J.; Wang, L. Pyramid attention network for semantic segmentation. arXiv 2018, arXiv:1805.10180.

32. Zhou, P.; Han, X.; Morariu, V.I.; Davis, L.S. Learning rich features for image manipulation detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.

33. Zhang, Y.; Yang, Q. An overview of multitask learning. Natl. Sci. Rev. 2018, 5, 30–43. [CrossRef]

34. Caruana, R. Multitask learning. Mach. Learn. 1997, 28, 41–75. [CrossRef]

35. Pomerleau, D.A. ALVINN, an Autonomous Land Vehicle in a Neural Network; School of Computer Science at Research Showcase @ CMU, Carnegie Mellon University: Pittsburgh, PA, USA, 1989.

36. Vandenhende, S.; Georgoulis, S.; Van Gansbeke, W.; Proesmans, M.; Dai, D.; Van Gool, L. Multitask learning for dense prediction tasks: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 2022, 44, 3614–3633. [CrossRef] [PubMed]

37. Dong, J.; Wang, W.; Tan, T. Casia image tampering detection evaluation database. In Proceedings of the 2013 IEEE China Summit and International Conference on Signal and Information Processing, Beijing, China, 1–5 July 2013.

38. Krawetz, N. A Picture’s Worth: Digital Image Analysis and Forensics. Available online: https://www.hackerfactor.com/papers/bh-usa-07-krawetz-wp.pdf (accessed on 20 March 2016).

39. Ye, S.; Sun, Q.; Chang, E. Detecting digital image forgeries by measuring inconsistencies of blocking artifacts. In Proceedings of the 2007 IEEE Conference on Multimedia and Expo, Beijing, China, 2–5 July 2007. [CrossRef]

40. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 16–20 July 2017.