Complex & Intelligent Systems (2022) 8:2371–2381
https://doi.org/10.1007/s40747-021-00633-1

ORIGINAL ARTICLE

Hybrid attention network with appraiser-guided loss for counterfeit luxury handbag detection

Jianbiao Peng[1,2] · Beiji Zou[1,2] · Xiaoyu He[3] · Chengzhang Zhu[1,2,4,5]

Received: 26 August 2021 / Accepted: 17 December 2021 / Published online: 2 February 2022

Abstract

Recently, convolutional neural networks have shown good performance in many counterfeit detection tasks. However, accurate counterfeit detection is still challenging due to the following three issues: (1) fine-grained classification, (2) class imbalance, and (3) high imitation samples. To address these issues, we propose a hybrid attention network (HANet) for counterfeit luxury handbag detection. In HANet, a hybrid attention module is first designed. Compared with existing methods that directly use classic CNNs for counterfeit detection, the HA module jointly uses a channel attention unit and a spatial attention unit to learn important information on both the channel and spatial dimensions. The HA modules can be easily integrated into the ResNet architecture to enhance the discriminative representation ability of CNNs, so as to help the network find subtle differences between the real and counterfeit products. In addition, an appraiser-guided loss is proposed to train HANet. Considering the factor of class imbalance and high imitation samples, the proposed loss gives the counterfeit class a higher weighting, and meanwhile gives the high imitation samples a much higher weighting. The proposed loss introduces the knowledge of appraisers, which allows HANet to not only treat real and counterfeit samples relatively fairly, but also pay more attention to the learning of difficult samples. To evaluate the performance of our method, we have constructed a well-benchmarked luxury handbag dataset. On this dataset, the performance of HANet, ResNet50, and the state-of-the-art attention methods is compared. The results demonstrate that HANet achieve superior performance against all its competitors.

Keywords Counterfeit detection · Convolutional neural networks · Attention mechanism · Luxury handbag · High imitation samples

Introduction

Counterfeit goods are a massive worldwide problem, which directly affects almost all high-value products [1,2]. According to a report by the Organization for Economic Cooperation and Development (OECD) [3], illegal trafficking of counterfeit goods accounted for 3.3% of the world trade in 2016, which was approximately $509 billion, up from 2.5% in 2013. Moreover, these figures are only based on 2016 customs seizure data, and do not include counterfeit products produced and consumed in various countries, or pirated products distributed through the Internet. At present, counterfeit goods have appeared in many industries, ranging from luxury handbags, perfume, and machine components to chemical products [4–6]. More seriously, some counterfeit goods even threaten the safety of human life, such as counterfeit auto parts, medical equipment with incorrect parameters, counterfeit tablets, counterfeit baby milk powder, etc. [7–9]. Moreover, counterfeit goods also seriously affect social security. As we all know, the profits made by counterfeiters in various markets have become one of the important sources of funds for illegal and potentially harmful activities around the world [10].

✉ Chengzhang Zhu
chzhzhu@csu.edu.cn

1 School of Computer Science and Engineering, Central South University, Changsha 410083, China

2 Hunan Engineering Research Center of Machine Vision and Intelligent Medicine, Central South University, Changsha 410083, China

3 School of Automation, Central South University, Changsha 410083, China

4 The College of Literature and Journalism, Central South University, Changsha 410083, China

5 Center for Intelligent Media and Communication Research, The College of Literature and Journalism, Central South University, Changsha 410083, China

Fighting against counterfeit goods is a protracted and never-ending struggle. So far, many methods based on overt and covert technologies have been proposed for counterfeit detection [11–16]. The commonly used overt anti-counterfeiting technologies include holograms [17], bar codes, watermarks, color-changing inks, and sequential product numbers. This type of method relies heavily on the verification details that exist on the object surface, and these details can easily be reverse-engineered or removed by counterfeiters. Covert anti-counterfeiting technologies include security inks, digital watermarks, biological, chemical or microscopic taggants [18], QR or RFID tags, etc. The covert methods are usually more accurate than the explicit methods and can provide a stronger guarantee of authenticity. However, these solutions are often costly and even difficult to be adopted by commodity manufacturers. In many markets for high-value goods, manufacturers may oppose adding implicit solutions to goods, especially luxury goods, fashion, or art.

Recently, with the rapid improvement of the computing ability of GPU devices, convolutional neural networks (CNNs), as a typical algorithm in the field of deep learning, have shown extremely powerful performance in various image analysis tasks [19–21]. For example, in the famous ImageNet image classification competition, the recognition ability of CNNs has surpassed that of humans [19]. In the field of medical image analysis, the diagnosis performance of deep models trained by CNNs is superior to that of professional physicians [22–24]. Inspired by these successful practices, some studies have begun to apply CNNs for the detection of counterfeit goods. Among these methods, their main research objects are counterfeit currencies [25,26], counterfeit medicines [27,28], and counterfeit luxury handbags [29,30]. By collecting a large number of specific product images and annotating their labels (i.e., real or counterfeit), existing CNN-based methods tend to directly use CNNs to learn the difference between the real and counterfeit classes and thereby realizing the detection of counterfeit products.

Although these CNN-based methods have achieved good performance in some specific counterfeit detection tasks, their performance is still limited due to the following three issues: (1) fine-grained classification: real and counterfeit products are sometimes very similar in image appearance, and CNNs used for natural scene image classification usually lack the ability to capture these extremely subtle differences; (2) class imbalance: the number of real product images is far more than counterfeit ones; (3) high imitation products: they are difficult samples in the CNN training process, which may mislead the feature learning process and deteriorate the final counterfeit detection performance.

To solve the above problems, we propose a hybrid attention network (named HANet) with appraiser-guided loss for luxury handbag detection. Compared with existing methods that directly use classic CNNs for counterfeit detection, we propose a novel hybrid attention module, called HA module. The HA module jointly uses a channel attention unit and a spatial attention unit to learn important information on the channel and spatial dimensions, which can enable CNNs to automatically locate the discriminative regions of real and counterfeit products. In addition, an appraiser-guided loss is proposed to train HANet. Considering the factor of class imbalance, the proposed loss gives the counterfeit class a higher weighting based on the ratio of class distribution. Then, for counterfeit samples belonging to high imitations, the proposed loss further increases the weight of them. The reason why the proposed loss is called the appraiser-guided loss is that the specific determination of whether a certain counterfeit sample is a high imitation product is made by the discussion of multiple experts. With the proposed appraiser-guided loss, HANet could treat real and counterfeit samples relatively fairly during the training process, and meanwhile pay more attention to the learning of difficult samples (i.e., high imitation products).

The main contributions of this paper are summarized as follows:

- A novel HA module is proposed to learn important information on both channel and spatial dimensions, which can make CNNs focus more on the discriminative regions of real and counterfeit products. To the best of our knowledge, this is the first attempt to apply the attention mechanism in CNNs for counterfeit detection.
- A new loss is proposed to train a counterfeit detection model under the condition of class imbalance and high imitation products. The proposed loss incorporates the expert knowledge of appraisers and gives a higher weighting to high imitations samples, thus promoting the model's learning of difficult samples during the training process.
- A large luxury handbag dataset has been collected and well annotated, including four brands (i.e., Chanel, Gucci, Louis Vuitton, and Prada) and 74,916 images. On the constructed dataset, the effectiveness of HANet is demonstrated by comparing with ResNet and state-of-the-art attention methods.

Related work

Counterfeit detection based on CNNs

Due to their powerful end-to-end feature learning capabilities, CNNs have been widely used for counterfeit detection. In these studies, most of them study counterfeit currencies [25,26,31–33] or medicines [27,28,34–36], and a few investigate counterfeit luxury handbags [29,30]. For example, Desai et al. [31] proposed a method combining CNN and

Generative Adversarial Network (GAN) to detect counterfeit India currency. Kamble et al. [32] trained a CNN model to identify counterfeit currency on handheld devices such as smartphones and tablets. Rahmad et al. [33] used K-nearest neighbor (KNN) and CNN for counterfeit currency detection, and demonstrated that the detection performance of CNN is higher than that of KNN. Zheng et al. [34] proposed a general method to detect counterfeit drugs based on a siamese network structure. Mishra et al. [35] used a variety of methods including support vector machine (SVM), logistic regression, linear regression, and CNN for high-accuracy counterfeit drug detection. Ferdosi et al. [36] used a VGG-16 model with transfer learning technique to build a non-invasive identification system for drug brand classification and counterfeit detection.

Regarding the detection of counterfeit luxury bags, there are only two related published studies. The first study was proposed by Sharma et al. [29]. They used two classifiers (SVM and an 8-layer CNN) to detect the authenticity of a variety of physical objects, including 20 types of leather, 120 types of fabrics, 10 types of paper, 10 types of plastic surfaces, 2 authentic NFL jerseys, and 2 types of Viagra pills. The experimental results show that the recognition performance of CNN is higher than that of SVM. In addition, they stated that their method has been deployed to verify the authenticity of luxury handbags. An obvious limitation of this method is the requirement of a special imaging equipment to display the details of products, which is often unfriendly to the end users. Target at this problem, Serban et al. [30] proposed a more friendly counterfeit detection system for luxury handbags. This system developed multiple CNN models (i.e., VGG-16 models) to detect the authenticity of different parts of the handbag. Regarding the counterfeit detection of Louis Vuitton (LV) handbags, the counterfeit detection parts include buckles, etiquettes, and textures. In its actual use stage, this system requires users to upload the specified part image according to the prompt, and then applies the corresponding CNN model to test its authenticity score.

In summary, existing CNN-based methods for counterfeit detection mainly focus on the studies of counterfeit currencies and medicines. In the few published papers that support counterfeit luxury handbag detection, they only verified the performance of their methods on a limited dataset, such as some physical objects and LV handbags. As a result, the performance of their methods on multiple brands of luxury handbags is not sure. In addition, their methods are simple CNN models, such as 8-layer CNN or VGG-16. Due to the limited feature representation capabilities, their performance on some complex counterfeit detection tasks may be challenged.

Image classification based on attention mechanism

The attention mechanism in deep learning is similar to the attention mechanism of human vision, i.e., selecting key points from a large amount of information while ignoring other irrelevant information. It has brought many breakthroughs in the fields of natural language processing [37,38] and computer vision [39,40]. Recently, the attention mechanism has also been used to improve the feature representation ability of CNNs in large-scale image classification tasks [41,42]. For example, Wang et al. [41] proposed a residual attention network called RAN, which is stacked by multiple attention modules in the residual network. Each attention module consists of a trunk branch and a mask branch, where the trunk branch is used for feature processing, and the mask branch employs an encoding–decoding structure to learn the corresponding attention feature map. Hu et al. [42] proposed a new attention-based network called SENet, which consists of a series of squeeze and excitation (SE) blocks. The SE block explicitly uses two fully connected (FC) layers to learn the importance of each channel, where the first FC layer compresses the original feature map in the channel dimension and the second FC layer restores the number of channels to the original size.

Because the attention mechanism has the characteristics of helping the network locate the discriminative and meaningful regions on the image, it has also been applied to fine-grained image classification tasks [43–46]. For example, Zheng et al. [43] proposed a part learning method based on multi-attention CNNs for fine-grained classification, in which part generation and feature learning can reinforce each other. Peng et al. [44] proposed an object-partial attention model for weakly supervised fine-grained image classification, in which the object-level attention locates the object of the image, and the partial-level attention selects the discriminative part of the object. Zhang et al. [45] proposed a residual attention network for skin lesion classification, which contains an attention residual module that combines residual learning and spatial attention mechanisms. Huynh et al. [46] proposed an attention mechanism based on dense attributes for fine-grained classification of small samples. This mechanism can focus on the most relevant image regions for each attribute and obtain attribute-based features accordingly.

Although the attention mechanism has many successful applications in the field of fine-grained image classification, its performance has not yet been verified in counterfeit detection tasks. As described in “Introduction”, counterfeit detection tasks belong to the fine-grained classification. Therefore, to further boost the performance of the counterfeit detection, it is necessary to incorporate the attention mechanism into the counterfeit detection method.

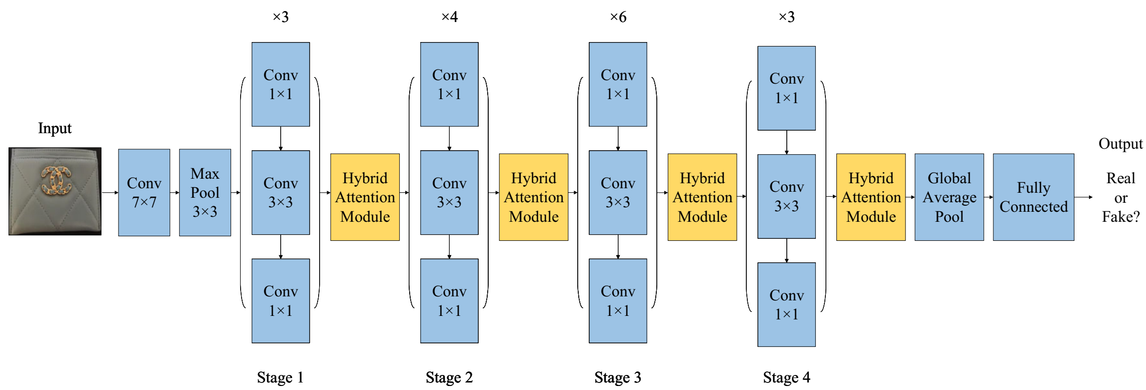


Fig. 1 Overview of the proposed HANet

Proposed approach

Overview

Figure 1 shows the overall framework of HANet. The input of HANet is a part image of a luxury handbag, and the output is the counterfeit detection prediction of this image. Specifically, HANet includes a convolutional layer, a max pooling layer, a series of the combination of residual blocks and the hybrid attention (HA) modules, a global average pooling layer, and a fully connected layer. Overall, HANet adopts the macro-structure of ResNet, i.e., adopting the skip connections and residual learning mechanism. Therefore, similar to ResNet, HANet could also be stacked deeply while avoiding the vanishing gradient problem. Different from ResNet, HANet uses the attention mechanism to learn important information on both channel and spatial dimensions to optimize the features at different stages, with the aim of improving the discriminative representation ability of the network. Specifically, the HA module is inserted after the residual blocks of each stage of the network. It optimizes the output features of the current stage's residual block and sends the optimized features to the next stage to continue feature extraction. In addition, considering the issues of class imbalance and high imitation samples, a new loss that incorporates the knowledge of appraisers is proposed to train HANet. The details of the HA module and the loss are introduced in the following.

HA module

As discussed in "Introduction", since counterfeit detection is a fine-grained classification task, existing CNN-based methods may be difficult to capture the differences between real and counterfeit classes, thereby leading to limited recognition performance. To this end, we designed an HA module with a hybrid attention mechanism to explore important information on channel and spatial dimensions, aiming at enhancing the discriminative representation capabilities of the network and achieving more accurate counterfeit detection.

The specific structure of the HA module is shown in Fig. 2. Its input is the feature map output by the residual blocks, denoted as $F \in \mathbb{R}^{C\times H\times W}$ (C, H, and W represent the channel number, height, and width of the feature, respectively). First, a 1×1 regular convolution is performed on F to generate $F_c \in \mathbb{R}^{C\times H\times W}$ and $F_s \in \mathbb{R}^{C\times H\times W}$, respectively. Subsequently, F_c and F_s are processed by a channel attention unit and a spatial attention unit, respectively. These two attention units are introduced in the following.

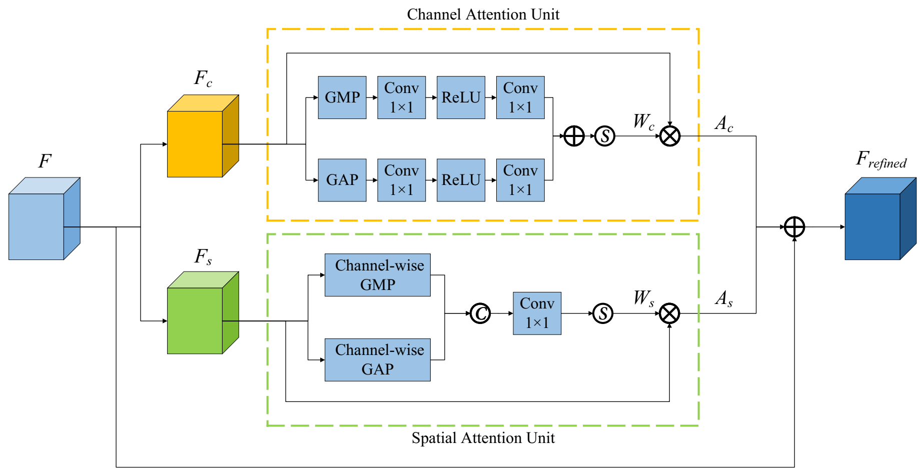
For the channel attention unit, the input feature map F_c is processed by a global max pooling (GMP) layer and a global average pooling (GAP) layer on the spatial dimension to obtain two $C \times 1 \times 1$ feature maps. Then, they are, respectively, sent to a multi-layer perceptron (MLP) composed of two 1×1 convolutional layers. In the MLP, the number of output channels of the first convolutional layer is C/r (r is the reduction rate which is set to 16 in this paper), and the number of output channels of the second convolutional layer is C. Then, the two feature maps output by the MLP are fused by an element-wise addition. Then, a sigmoid activation operation is performed to generate the channel-wise attention weight, denoted as $W_c \in \mathbb{R}^{C\times 1\times 1}$. Finally, W_c and F_c are element-wise multiplied to generate the channel-wise attention feature map A_c. The operations of the entire channel attention unit can be summarized as the following two formulas:

$$W_c = \sigma(w_{m1}(w_{m0}(\mathrm{GMP}^s(F_c))) + w_{a1}(w_{a0}(\mathrm{GAP}^s(F_c)))) \quad (1)$$

$$A_c = W_c \odot F_c, \quad (2)$$

where $\mathrm{GMP}^s(\cdot)$ and $\mathrm{GAP}^s(\cdot)$ represent global average and global max pooling operations on the spatial dimension, respectively, w_{m0} and w_{m1} represent two 1×1 convolutions following the GMP layer, w_{a0} and w_{a1} represent two 1×1 convolutions following the GAP layer, $\sigma(\cdot)$ represents the

Fig. 2 Structure of the HA module



sigmoid function, and $\odot$ represents the element-wise multiplication operation.

Regarding the spatial attention unit, the input feature map F_s is processed by a GMP layer and a GAP layer on the channel dimension to obtain two $1 \times H \times W$ feature maps. Subsequently, the two feature maps are concatenated on the channel dimension, and the number of channels is reduced to 1 via a 1×1 convolution. Then, a sigmoid function is performed on this single-channel feature map to generate the spatial-wise attention weight, denoted as $W_s \in \mathbb{R}^{1\times H\times W}$. Finally, the spatial-wise attention feature map A_s is generated through the element-wise multiplication of W_s and F_s. The operation of the entire channel attention unit can be summarized as the following two formulas:

$$W_s = \sigma(w_2(\mathrm{Cat}(\mathrm{GMP}^c(F_s), \mathrm{GAP}^c(F_s)))) \tag{3}$$

$$A_s = W_s \odot F_s, \tag{4}$$

where $\mathrm{GMP}^c(\cdot)$ and $\mathrm{GAP}^c(\cdot)$ represent global average and global max pooling operations on the channel dimension, respectively, Cat represents the feature concatenation operation on the channel dimension, and w_2 represents the 1×1 convolution.

Finally, the output of the HA module is the weighted summation of A_c, A_s, and F

$$F_{\mathrm{refined}} = F + \alpha \cdot A_c + \beta \cdot A_s, \tag{5}$$

where $F_{refined}$ represents the final refined feature map, and α and β are two learnable weighting factors to adjust the contributions of A_c and A_s in F_{refined}, respectively. α and β are initially set to 0, and their values can be adaptively learned during the model training process.

From the above descriptions, it can be observed that the proposed HA module jointly uses the channel and spatial attention mechanisms to learn and capture important information of the input feature map. In the literature, there are also some existing hybrid attention methods that integrate both channel and spatial attention mechanisms, such as CBAM [47] and DANet [48]. For the input feature map, CBAM first executes a channel attention module, and then applies a spatial attention module to the output feature map of the channel attention module to obtain the final refined feature map. Compared with CBAM that uses a serial channel-spatial attention structure, the HA module uses two parallel attention units to implement the channel and spatial attention mechanisms, so that the learning process of the two attention mechanisms can be decoupled. DANet has parallel position and channel attention modules, which is similar to our HA module. However, the self-attention operations in these two modules are computationally expensive, resulting in slow model training and inference speed. Compared with DANet, the proposed HA module uses GMP and GAP operations, which can greatly reduce model parameters and lower the computational complexity. Moreover, the output of the HA module not only contains the channel and spatial attention feature maps, but also includes the original input feature map. In this way, the convergence of model training can be guaranteed, and the representation ability of the network can also be enhanced.

Loss function

In the existing methods for counterfeit detection, most of them use the original cross-entropy loss for model training. However, an obvious fact is that there exists heavy class imbalance in counterfeit detection tasks, i.e., there are far more real products than counterfeit products. In addition, there also exists a certain number of high imitations that are easily confused with the real ones, which are called hard negative samples. It is well known that the problem of class imbalance will cause the model to be biased toward classes with more samples, and the difficult samples will mislead the model training process. To solve these two problems, we propose a appraiser-guided loss. The proposed loss first gives a higher weighting to the non-representative class (i.e., the counterfeit class) to alleviate the impact of class imbalance on training. Then, a much higher weighting is given to the samples annotated as high imitations to make the model focus more on these hard negative examples. Given that the number of real and counterfeit samples in the training data are M and N, respectively, the proposed appraiser-guided loss is defined as follows:

$$L_{\mathrm{AG}} = -\frac{1}{M+N}\sum_{i=1}^{M+N}\omega(x_i)(y_i \cdot log(\phi(x_i;\theta)) + (1-y_i)\cdot \log(1-\phi(x_i;\theta))), \tag{6}$$

where x_i and y_i represent an input image and its label, $\phi(x_i;\theta)$ represents the predicted result with the input of x_i (θ represents the parameters of the network and $\phi(\cdot)$ represents the input to output mapping of the network), and $\omega(x_i)$ is a weighting factor function according to x_i. Specifically, $\omega(x_i)$ is defined as follows:

$$\omega(x_i) = \begin{cases} \frac{M+N}{M}, & x_i \in \mathbb{P}, \\ \frac{M+N}{N}, & x_i \in \mathbb{N}_{\text{normal}}, \\ \left(\frac{M+N}{N}\right)^k, & x_i \in \mathbb{N}_{\text{hard}}, \end{cases} \tag{7}$$

where $\mathbb{P}$, $\mathbb{N}_{\text{normal}}$, and $\mathbb{N}_{\text{hard}}$ represent the sets of real samples, counterfeit samples that do not belong to high imitations (i.e., general counterfeit samples), and counterfeit samples that belong to high imitations (i.e., high imitation samples). To give high imitations a greater loss weighting, $k > 1$. Here, we set k to 1.5 in all experiments of this paper.

From Eqs. 6 and 7, we can draw the following two conclusions: (1) in the proposed loss, since $M > N$, the general counterfeit samples receive a higher loss weighting than the real samples; (2) since $k > 1$, the high imitation samples obtain a higher loss weighting than the general counterfeit samples. From the above two conclusions, the proposed appraiser-guided loss has the potential to address the issues of class imbalance and high imitations in counterfeit detection tasks.

Analysis of principle

As described before, HANet is a kind of attention-based CNN method designed for counterfeit detection. In the following, we compare HANet with existing CNN-based counterfeit detection methods and other attention methods:

- For existing CNN-based counterfeit detection methods, they usually directly employ classic CNNs for classification. However, the real and fake products are sometimes very similar in image appearance, and the CNNs designed for natural scene image classification tasks may lack the ability to capture these subtle differences. To address this issue, an HA module is designed to learn important information on both channel and spatial dimensions, which can be easily integrated into the ResNet architectures and help the network locate the discriminative regions of real and counterfeit products.
- Existing attention methods can be simply divided into channel, spatial, and hybrid attention methods. The first two kinds of methods learn importance on channel and spatial dimensions, respectively, while the hybrid ones aim at learning important information on both two dimensions. CBAM [47] and DANet [48] are two representatives of hybrid attention methods. As discussed in “HA Module”, compared with CBAM, HANet could obtain better channel and spatial attention maps using two parallel attention units to decouple the two attention learning mechanisms. Compared with DANet that also uses two parallel attention units to implement the channel and spatial attention mechanisms, HANet has advantages in model size and computational complexity.
- In addition, HANet differs from existing CNN-based counterfeit detection methods in the loss function. For most CNN-based counterfeit detection methods, they usually use the original cross-entropy loss for model training. However, their training performance may be deteriorated due to the challenges of heavy class imbalance and high imitation samples. To deal with these two challenges, an appraiser-guided loss is proposed to train HANet, which first gives the counterfeit class a higher weighting based on the ratio of class distribution and then further increases the weight of counterfeit samples belonging to high imitations. By doing so, HANet could treat real and counterfeit samples relatively fairly during the training process and pay more attention to the learning of difficult samples.

Experimental setup

Dataset

To verify the performance of the proposed algorithm, we have collected and well-labeled a luxury handbag dataset. To collect this dataset, we have designed an online program for the verification of luxury bags. In this program, the users can upload the pictures of luxury bags that need to be authenticated according to the requirements of our program, and then, the authenticity of the uploaded pictures is verified by the appraisers from a third-party professional appraisal agency. Through this online verification program, we have

Fig. 3 Some sample images of the dataset. These images all belong to Chanel. The images on the left and right sides belong to real and counterfeit classes, respectively, and each row of the images represents an identification part

Table 1 Details of the Luxury Handbag Dataset

	Bag	Button	Coding	Embossing	Front	Hasp	Label	Lock	Logo	Sign	Tag	Zipper	Total
Chanel													
All	/	/	1689	/	5260	994	/	/	/	3817	3698	/	15,458
Real	/	/	1268	/	3547	661	/	/	/	2589	2640	/	10,705
Counterfeit	/	/	421	/	1713	333	/	/	/	1228	1058	/	4753
Gucci													
All	1722	727	1873	/	4500	/	/	/	/	3292	2583	/	14,697
Real	1479	640	1628	/	3784	/	/	/	/	2821	2162	/	12,514
Counterfeit	243	87	245	/	716	/	/	/	/	471	421	/	2183
Louis Vuitton													
All	/	6262	4444	/	6588	/	/	5103	/	5253	/	6481	34,131
Real	/	4664	3254	/	4610	/	/	3646	/	3743	/	4565	24,482
Counterfeit	/	1598	1190	/	1978	/	/	1457	/	1510	/	1916	9649
Prada													
All	/	247	/	317	2601	/	1171	/	979	2911	1247	1157	10,630
Real	/	202	/	254	1880	/	854	/	791	1997	875	823	7676
Counterfeit	/	45	/	63	721	/	317	/	188	914	372	334	2954

collected a large number of pictures of luxury bags, and each picture has the ground-truth label of real or counterfeit. Overall, the dataset has 74,916 high-quality luxury handbag images, covering four luxury brands. The four luxury brands are Chanel, Gucci, Louis Vuitton, and Prada, which contain 15,458, 14,697, 34,131, and 10,630 images, respectively. Some sample images of the dataset are shown in Fig. 3. In this dataset, the images are all part images of luxury handbags labeled by appraisals. The real and counterfeit labels are repeatedly confirmed by multiple appraisers. According to the general rules of luxury bag identification, there are 12 general identification parts: bag, button, coding, embossing, front, hasp, label, lock, logo, sign, tag, and zipper. Different brands of luxury handbags may have different identification parts. For example, for CHANEL, its identification parts include coding, front, hasp, sign, and tag. For GUCCI, its identification parts include bag, button, coding, front, sign, and tag. For each brand, the number of images of different parts and the number of real and counterfeit images are given in Table 1. It can be seen from the table that the number of real samples is obviously more than that of the counterfeit ones. The class imbalance makes the counterfeit detection task more difficult.

Algorithms for comparison

As mentioned earlier in 'Proposed approach', the proposed HANet is built by integrating the channel and spatial attention mechanisms into ResNet. Therefore, the algorithms used for comparison include the baseline (ResNet50 [19]) and attention methods, including SENet50 [42], RAN50 [41], ARL-CNN50 [45], and CBAM50 [47]. Among these attention methods, SENet50 belongs to the channel attention methods, RAN50 and ARL-CNN50 belong to spatial attention methods, and CBAM50 belongs to the hybrid channel and spatial attention methods. It is worth mentioning that these attention methods and the proposed HANet all use the macro-architecture of ResNet50, so the depth of the network is basically the same, which can ensure the fairness of the comparisons in our experiments. The number of parameters and computational complexity of these algorithms are shown in Table 2. We use the indicator of floating point operations

Table 2 Comparisons of different attention methods in terms of the number of parameters (Params) and floating point operations (FLOPs)

	Params (M)	FLOPs (G)
ResNet50 [19]	23.512	5.377
SENet50 [42]	26.027	5.379
RAN50 [41]	31.900	6.300
ARL-CNN50 [45]	23.512	5.377
CBAM50 [47]	26.029	5.381
HANet	25.027	5.379

Table 3 Classification accuracy of HANet, the baseline (ResNet50), and the attention methods (SENet50, RAN50, ARL-CNN50, and CBAM50) on the Chanel brand of the luxury handbag dataset

	Coding	Front	Hasp	Sign	Tag
ResNet50 [19]	0.923	0.868	0.880	0.873	0.947
SENet50 [42]	0.930	0.875	0.886	0.877	0.953
RAN50 [41]	0.934	0.880	0.890	0.879	0.955
ARL-CNN50 [45]	0.934	0.876	0.887	**0.884**	0.958
CBAM50 [47]	0.940	0.882	0.895	0.883	0.962
HANet	**0.945**	**0.887**	**0.896**	0.883	**0.968**

For each identification part, the highest performance is highlighted in boldface

(FLOPs) to represent the computational complexity of a deep learning model. Note that these two indicators are calculated on the input image with the resolution of 256 × 256 × 3. It can be seen from Table 2 that compared with the original ResNet50, the increase in the number of parameters and computational complexity of ARL-CNN50 is negligible, while that of RAN50 is the largest among these attention methods. Compared with CBAM50, which is also a hybrid attention method, HANet has relatively fewer parameters and lower computational complexity.

Implementation details

In the training phase, for each brand of the luxury handbag, we need to train an independent counterfeit detection model for each identification part. According to the conventions of machine learning, we divided the original dataset with the ratio of 3:1:1, i.e., 3/5, 1/5, and 1/5 data were used for model training, validation, and test, respectively. Next, we adjusted the image size to 256 × 256. To avoid the over-fitting problem, online data augmentations were implemented, including random rotation ([40; +40]), zoom (90%–110% of width and height), and horizontal and vertical flips. The SGD algorithm was adopted as the optimizer, with the momentum of 0.9, weight decay of 0.0005, and batch size of 16. The learning rate was initially set to 0.001, and halved every 50 epochs. To speed up the training process, we loaded the pre-trained parameters on ImageNet of the original ResNet. The maximum number of training epochs was set to 300. We retained the model with the best performance on the validation set, and used the model to test on the corresponding test set to obtain the final performance.

Results and analysis

Compared with the baseline and attention methods

First, we compared the performance of HANet with that of the baseline (ResNet50) and attention methods on the luxury handbag dataset. The experimental results of Chanel, Gucci, Louis Vuitton, and Prada are shown in Tables 3, 4, 5, and 6, respectively. From Table 3, compared with all the competitors, HANet achieves the highest accuracy on four identification parts of Chanel (coding, front, hasp, and tag), and the second highest accuracy on Sign. Compared with the baseline, HANet improves the accuracy by 2.2%, 1.9%, 1.6%, 1.0%, and 2.1% on Chanel's coding, front, hasp, sign, and tag, respectively. From Table 4, it indicates that HANet achieves the highest accuracy on all the six identification parts of Gucci. Compared with the baseline, HANet improves the accuracy by 2.3%, 1.9%, 1.9%, 2.2%, 2.1%, and 2.1% on Gucci's bag, button, coding, front, sign, and tag, respectively. From Table 5, among all the methods, HANet obtains the highest accuracy on the button, coding, lock, sign, and zipper of Louis Vuitton, and ranks the second place on front. Specifically, HANet improves the accuracy by 1.7%, 2.1%, 1.4%, 1.7%, 1.9%, and 1.8% against the baseline on Louis Vuitton's button, coding, front, lock, sign, and zipper, respectively. From Table 6, it shows that HANet achieves the highest accuracy on six identification parts of Prada (button, embossing, label, sign, tag, and zipper), and the second highest accuracy on other two identification parts (front and logo). Compared with the baseline, HANet improves the accuracy by 1.9%, 2.1%, 1.5%, 1.0%, 1.1%, 1.9%, 1.8%, and 2.2% on the Prada's button, embossing, front, label, logo, sign, tag, and zipper, respectively. To summarize, the above results suggest that HANet has better performance than the baseline and attention methods on our collected luxury handbag dataset. In particular, compared with CBAM50 that also belongs to a hybrid attention method, HANet not only achieves superior performance, but also has fewer parameters and lower computational complexity (Table 2).

Ablation study

HANet includes two main components: the HA module and the appraiser-guided loss. To investigate their effectiveness, we conducted the ablation experiments on each of them.

Table 4 Classification accuracy of HANet, the baseline (ResNet50), and the attention methods (SENet50, RAN50, ARL-CNN50, and CBAM50) on the Gucci brand of the luxury handbag dataset

	Bag	Button	Coding	Front	Sign	Tag
ResNet50 [19]	0.935	0.929	0.903	0.897	0.878	0.900
SENet50 [42]	0.943	0.934	0.911	0.906	0.889	0.908
RAN50 [41]	0.945	0.940	0.913	0.910	0.889	0.913
ARL-CNN50 [45]	0.944	0.939	0.915	0.910	0.891	0.912
CBAM50 [47]	0.949	0.944	0.915	0.914	0.893	0.915
HANet	**0.958**	**0.948**	**0.922**	**0.919**	**0.899**	**0.921**

For each identification part, the highest performance is highlighted in boldface

Table 5 Classification accuracy of HANet, the baseline (ResNet50), and the attention methods (SENet50, RAN50, ARL-CNN50, and CBAM50) on the Louis Vuitton brand of the luxury handbag dataset

	Button	Coding	Front	Lock	Sign	Zipper
ResNet50 [19]	0.884	0.920	0.894	0.856	0.938	0.876
SENet50 [42]	0.893	0.927	0.900	0.864	0.943	0.883
RAN50 [41]	0.896	0.931	0.904	0.868	0.945	0.885
ARL-CNN50 [45]	0.895	0.930	0.896	0.864	0.946	0.886
CBAM50 [47]	0.899	0.933	**0.910**	0.870	0.949	0.889
HANet	**0.901**	**0.941**	0.908	**0.873**	**0.957**	**0.894**

For each identification part, the highest performance is highlighted in boldface

Table 6 Classification accuracy of HANet, the baseline (ResNet50), and the attention methods (SENet50, RAN50, ARL-CNN50, and CBAM50) on the Prada brand of the luxury handbag dataset

	Button	Embossing	Front	Label	Logo	Sign	Tag	Zipper
ResNet50 [19]	0.949	0.899	0.856	0.809	0.924	0.891	0.849	0.890
SENet50 [42]	0.958	0.906	0.864	0.814	0.929	0.897	0.853	0.894
RAN50[41]	0.963	0.911	0.867	0.816	0.933	0.903	0.853	0.897
ARL-CNN50 [45]	0.957	0.910	0.868	0.814	0.934	0.900	0.857	0.894
CBAM50 [47]	0.967	0.913	**0.872**	0.818	**0.940**	0.901	0.861	0.901
HANet	**0.968**	**0.920**	0.871	**0.819**	0.935	**0.910**	**0.867**	**0.912**

For each identification part, the highest performance is highlighted in boldface

Tables 7 and 8 give the ablation results on Gucci and Louis Vuitton of the luxury handbag dataset, respectively. In these experiments, HANet without the HA module represents the variant obtained by deleting the HA modules in HANet (the appraiser-guided loss is retained), and HANet without the appraiser-guided loss represents the variant obtained by replacing the appraiser-guided loss with regular binary cross-entropy loss.

(1) *Effectiveness of the HA module*: To valid the effectiveness of the HA module, we compared the results of HANet and HANet without the HA module. From Table 7, it suggests that the application of the HA module improves the accuracy by 1.6%, 1.4%, 1.3%, 1.6%, 1.4%, and 1.7% on Gucci's bag, button, coding, front, sign, and tag, respectively. From Table 8, it indicates that the usage of the HA module improves the accuracy by 1.4%, 1.3%, 1.0%, 1.0%, 1.6%, and 1.3% on Louis Vuitton's button, coding, front, lock, sign, and zipper, respectively. In summary, the above results demonstrate the effectiveness of the HA module in HANet.

(2) *Effectiveness of the appraiser-guided loss*: We also compared the results of HANet and HANet without the appraiser-guided loss. From Table 7, the employment of the appraiser-guided loss improves the accuracy by 0.5%, 0.3%, 0.4%, 0.4%, 0.3%, and 0.4% on Gucci's bag, button, coding, front, sign, and tag, respectively. From Table 8, it shows that the usage of the appraiser-guided loss improves the accuracy by 0.2%, 0.3%, 0.1%, 0.1%, 0.4%, and 0.3% on Louis Vuitton's button, coding, front, lock, sign, and zipper, respectively. In summary, the above results demonstrate the effectiveness of the appraiser-guided loss in HANet.

Conclusion

In this paper, we proposed HANet for counterfeit detection in luxury handbag images. In HANet, an HA module with a hybrid attention mechanism was first designed. Compared with existing methods that directly use classic CNNs for counterfeit detection, the HA module jointly uses a channel attention unit and a spatial attention unit to learn important information on both the channel and spatial dimensions. The proposed HA module can be easily integrated into the

Table 7 Ablation study for the HA module and the appraiser-guided loss on the Gucci brand of the luxury handbag dataset

	Bag	Button	Coding	Front	Sign	Tag
HANet without the HA module	0.942	0.934	0.909	0.903	0.885	0.904
HANet without the appraiser-guided loss	0.953	0.945	0.918	0.915	0.896	0.917
HANet	**0.958**	**0.948**	**0.922**	**0.919**	**0.899**	**0.921**

For each identification part, the highest performance is highlighted in boldface

Table 8 Ablation study for the HA module and the appraiser-guided loss on the Louis Vuitton brand of the luxury handbag dataset

	Button	Coding	Front	Lock	Sign	Zipper
HANet without the HA module	0.887	0.928	0.898	0.863	0.941	0.881
HANet without the appraiser-guided loss	0.899	0.938	0.907	0.873	0.953	0.891
HANet	**0.901**	**0.941**	**0.908**	**0.874**	**0.957**	**0.894**

For each identification part, the highest performance is highlighted in boldface

ResNet architecture to help the network find subtle differences between the real and counterfeit products. In addition, an appraiser-guided loss was proposed to train HANet. Considering the factors of class imbalance and high imitation samples, the proposed loss gives the counterfeit class a higher weighting and gives the high imitation samples a much higher weighting. The proposed loss introduces the knowledge of appraisers, which allows HANet to not only treat real and counterfeit samples relatively fairly, but also pay more attention to the learning of difficult samples. We evaluated the performance of HANet on our self-constructed dataset, a large and well-benchmarked luxury handbag dataset. The results showed that HANet could achieve superior performance against the state-of-the-art methods.

In the future, we plan to design a more effective attention module to further improve the performance of counterfeit detection. Moreover, we also intend to collect as more data as we can to verify the generalization performance of our model.

Acknowledgements This work was supported by the Scientific and Technological Innovation Leading Plan of High-Tech Industry of Hunan Province (Grant Number: 2020GK2021), the National Natural Science Foundation of China (Grant Number: 61902434), and the Natural Science Foundation of Hunan Province, China (Grant Number: 2019JJ50826).

References

1. Maaz M, Ali T (2020) How counterfeits goods are destroying brand reputation. J Eng Econ Dev 6(2):38–49
2. Antonopoulos GA, Hall A, Large J, Shen A (2020) Counterfeit goods fraud: an account of its financial management. Eur J Crim Policy Res 26(3):357–378
3. EUIPO OECD (2019) Trends in trade in counterfeit and pirated goods: illicit trade
4. Staake T, Thiesse F, Fleisch E (2012) Business strategies in the counterfeit market. J Bus Res 65(5):658–665
5. Chaudhry P, Zimmerman A (2013) The global growth of counterfeit trade. In: Protecting your intellectual property rights. Springer, pp 7–31
6. Kalyoncuoglu S, Sahin B et al (2017) Moderating role of materialism in the effect of perceived value on purchase intention of counterfeits of luxury brands. Int J Mark Stud 9(4):76–88
7. ElAmrawy F, ElAgouri G, Elnoweam O, Aboelazayem S, Farouk EM, Nounou MI (2016) Adulterated and counterfeit male enhancement nutraceuticals and dietary supplements pose a real threat to the management of erectile dysfunction: A global perspective. J Diet Suppl 13(6):660–693
8. Chow D (2010) Counterfeiting as an externality imposed by multinational companies on developing countries. Va J Int Law 51:785
9. Kangaspunta K, Musumeci M (2013) Trafficking in counterfeit goods. In: Handbook of transnational crime and justice, p 101
10. Cho S-H, Fang X, Tayur S (2015) Combating strategic counterfeiters in licit and illicit supply chains. Manuf Serv Oper Manag 17(3):273–289
11. Ngo YH, Li D, Simon GP, Garnier G (2011) Paper surfaces functionalized by nanoparticles. Adv Colloid Interface Sci 163(1):23–38
12. Sharma A, Subramanian L, Brewer EA (2011) Paperspeckle: microscopic fingerprinting of paper. In: Proceedings of the 18th ACM conference on Computer and communications security, pp 99–110
13. Chafee Z (1921) The reacquisition of a negotiable instrument by a prior party. Columbia Law Rev 21(6):538–553
14. Soares M, Fracalossi DM, de Freitas LEL, Rodrigues MS, Redig JC, Mouriño JLP, Seiffert WQ, do Nascimento Vieira F (2015) Woven label; weave label; woven & printed labels; waterproof fabric mark. Revista Brasileira de Zootecnia 44(10):343–349
15. van Renesse RL (2000) Synergistic combination of document security techniques. In: Optical security and counterfeit deterrence techniques III, vol 3973. International Society for Optics and Photonics, pp 126–138

16. Gradinarova G, Janyan A (2011) Motor simulation and verbal association in idiom-idiom verification: effects of imageability. In: Proceedings of the annual meeting of the cognitive science society, vol 33
17. Miao J, Ding X, Zhou S, Gui C (2019) Fabrication of dynamic holograms on polymer surface by direct laser writing for high-security anti-counterfeit applications. IEEE Access 7:142926–142933
18. Gooch J, Daniel B, Abbate V, Frascione N (2016) Taggant materials in forensic science: a review. TrAC Trends Anal Chem 83:49–54
19. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
20. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
21. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
22. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118
23. Zhang Y, Li Z, Jiao P, Zhu S (2021) Two-stage stochastic programming approach for limited medical reserves allocation under uncertainties. Complex Intell Syst 1–11
24. Li S, Liu B, Li S, Zhu X, Yan Y, Zhang D (2021) A deep learning-based computer-aided diagnosis method of x-ray images for bone age assessment. Complex Intell Syst 1–11
25. Naresh KS, Singal G, Sirikonda S, Nethravathi R (2020) A novel approach for detection of counterfeit Indian currency notes using deep convolutional neural network. In: IOP conference series: materials science and engineering, vol 981. IOP Publishing, p 022018
26. Hoang V-D, Vo H-Th (2018) Hybrid discriminative models for banknote recognition and anti-counterfeit. In: 2018 5th NAFOSTED conference on information and computer science (NICS). IEEE, pp 394–399
27. Ishiyama R, Takahashi T, Makino K, Kudo Y, Kooper M, Abbink D (2019) Medicine tablet authentication using fingerprints of ink-jet printed characters. In: 2019 IEEE international conference on industrial technology (ICIT). IEEE, pp 871–876
28. Alsallal M, Sharif MS, Al-Ghzawi B, al Mutoki SMM (2018) A machine learning technique to detect counterfeit medicine based on x-ray fluorescence analyser. In: 2018 international conference on computing, electronics & communications engineering (iCCECE). IEEE, pp 118–122
29. Sharma A, Srinivasan V, Kanchan V, Subramanian L (2017) The fake vs real goods problem: microscopy and machine learning to the rescue. In: Proceedings of the 23rd ACM sigkdd international conference on knowledge discovery and data mining, pp 2011–2019
30. Şerban A, Ilaş G, Poruşniuc G-C (2020) Spotthefake: an initial report on a new cnn-enhanced platform for counterfeit goods detection. arXiv:2002.06735
31. Desai S, Rajadhyaksha A, Shetty A, Gharat S (2021) Cnn based counterfeit Indian currency recognition using generative adversarial network. In: 2021 International conference on artificial intelligence and smart systems (ICAIS). IEEE, pp 626–631
32. Kamble K, Bhansali A, Satalgaonkar P, Alagundgi S (2019) Counterfeit currency detection using deep convolutional neural network. In: 2019 IEEE Pune section international conference (PuneCon). IEEE, pp 1–4
33. Rahmad C, Rohadi E, Lusiana RA (2021) Authenticity of money using the method knn (k-nearest neighbor) and cnn (convolutional neural network). In: IOP conference series: materials science and engineering, vol 1073. IOP Publishing, p 012029
34. Zheng A-B, Yang H-H, Pan X-P, Yin L-H, Feng Y-C (2020) On-site identification of counterfeit drugs based on near-infrared spectroscopy Siamese-network modeling. IEEE Access 9:3195–3206
35. Mishra AK, Essop MH (2019) Low-cost spectrogram based counterfeit medicine detection. arXiv:1904.07152
36. Ferdosi BJ, Sakib MA, Islam MS, Dhar J (2021) Identifying counterfeit medicine in bangladesh using deep learning. In: International conference on human-centered intelligent systems. Springer, pp 46–55
37. Wang Y, Huang M, Zhu X, Zhao L (2016) Attention-based lstm for aspect-level sentiment classification. In: Proceedings of the 2016 conference on empirical methods in natural language processing, pp 606–615
38. Zhou X, Wan X, Xiao J (2016) Attention-based lstm network for cross-lingual sentiment classification. In: Proceedings of the 2016 conference on empirical methods in natural language processing, pp 247–256
39. Xiao T, Xu Y, Yang K, Zhang J, Peng Y, Zhang Z (2015) The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 842–850
40. Haut JM, Paoletti ME, Plaza J, Plaza A, Li J (2019) Visual attention-driven hyperspectral image classification. IEEE Trans Geosci Remote Sens 57(10):8065–8080
41. Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, Wang X, Tang X (2017) Residual attention network for image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3156–3164
42. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132–7141
43. Zheng H, Fu J, Mei T, Luo J (2017) Learning multi-attention convolutional neural network for fine-grained image recognition. In: Proceedings of the IEEE international conference on computer vision, pp 5209–5217
44. Peng Y, He X, Zhao J (2017) Object-part attention model for fine-grained image classification. IEEE Trans Image Process 27(3):1487–1500
45. Zhang J, Xie Y, Xia Y, Shen C (2019) Attention residual learning for skin lesion classification. IEEE Trans Med Imaging 38(9):2092–2103
46. Huynh D, Elhamifar E (2020) Fine-grained generalized zero-shot learning via dense attribute-based attention. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 4483–4493
47. Woo S, Park J, Lee J-Y, Kweon S (2018) Cbam: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV), pp 3–19
48. Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, Lu H (2019) Dual attention network for scene segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3146–3154