Signal, Image and Video Processing
https://doi.org/10.1007/s11760-024-03105-4

ORIGINAL PAPER

A lightweight multi-task learning network based on key area guidance for counterfeit detection

Yaotian Yang[1] · Yu Yang[1] · Linna Zhou[1] · Jixin Zou[2]

Received: 2 January 2024 / Revised: 17 February 2024 / Accepted: 21 February 2024
© The Author(s) 2024

Abstract
Counterfeit detection traditionally relies on manual efforts, but manual detection efficiency is notably low. The accuracy of deep learning methods is challenging because of the insufficient samples, so it is crucial to allow the model to learn effective representation at a lower training cost. Given the above problems, we proposed a lightweight multi-task learning method that employs an uncomplicated auxiliary task to enhance the main task's attention and reduce the training sample requirements. A key area guidance algorithm is designed to construct the auxiliary task, disturbing key image areas to generate new samples and training the auxiliary task to recognize the disturbance. This guides the main task in discerning authenticity from these key areas. Additionally, a tailored data preprocessing strategy was designed to improve the method's performance further. Achieving an impressive 98.8% accuracy in identifying various counterfeiting points, our method outperforms existing advanced methods. Importantly, the method significantly reduces training costs. Even with an 80% reduction in the sample size, the method maintains a 92.1% accuracy, demonstrating minimal performance degradation compared to alternative methods.

Keywords Counterfeit detection · Multi-task learning · Key area guidance · Small sample size

1 Introduction

With the rapid development of online shopping, the problem of counterfeit products has become increasingly severe and a global problem. For example, the Chinese police cracked a case of counterfeit products in 2021, more than 340,000 items, such as counterfeit clothing and luggage, evaluating more than 300 million RMB [1]. Obviously, counterfeit goods harm the legitimate interests of consumers, and should be able to detect in a simple way. However, the detection task, especially related to expensive goods, can be only conducted by experts even by far. It means poor efficiency when encountering a tremendous number of goods.

There have been a few studies of automatic counterfeit detection. Traditional machine learning methods manually design features based on domain knowledge [2], which requires more time and workforce. The deep learning method must require a large-scale dataset, and a small sample size cannot guarantee its performance. However, counterfeit detection differs from defect detection, fake face detection, and other tasks. The difficulty of sample collection leads to the inability to structure a large-scale dataset, and the variety of sample series leads to the weakness of multi-series detection performance. Given the above problems, we propose a counterfeit detection network based on key area guidance and multi-task learning, and conduct experiments on counterfeit luxury goods as an example. The authenticity of samples is judged by the shape differences between the genuine and counterfeit samples. The main contributions of this paper are summarized as follows:

Firstly, the multi-task learning mechanism is introduced in counterfeit detection for the first time. The single-task method requires large amounts of data to find tiny differences between genuine and counterfeit samples. However, a well-designed multi-tasking method can facilitate the learn-

✉ Yu Yang
yangyu@bupt.edu.cn

✉ Jixin Zou
zoujixin163@163.com

Yaotian Yang
yangyaotian@bupt.edu.cn

Linna Zhou
zhoulinna@bupt.edu.cn

[1] School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China

[2] Institute of Forensic Science of China, Beijing 100038, China

Published online: 20 March 2024

ing process. We design a simple auxiliary task that is easy to learn from key areas, and helps the attention of the main task quickly focus on key locations, and thus reduce the requirement for data volume.

Secondly, a lightweight multi-task architecture is designed for counterfeit detection. For the sake of a compact architecture, the supporting relationship between tasks is leveraged and so the feature extraction network is shared. The parameters are optimized to simultaneously minimizing a disturbance detection loss and authenticity identification loss. In addition, a sample generation algorithm, called KAG, is designed by a way of disturbing key areas. As a result, a super-dataset is constructed for training both main and auxiliary tasks, free of collecting additional samples.

Finally, an image preprocessing strategy named FWD is proposed to avoid deformation interference. Images are usually normalized into same size and heavy distortion will be introduced if original ones with diverse aspect ratio. This will significantly affect performance as fine-grained differences in key objects are also confused. With the FWD strategy, the sample image is first filled into a shape of square, which is proved to enhance the learning.

The rest of this paper is organized as follows: The second section discusses the related work that can be used for counterfeit detection. The proposed method and performance analysis are detailed separately in the third and the fourth section. In the fifth section, summation and future study are presented.

2 Related work

Some related studies on counterfeit detection have been published in recent years. The entropy team was the first to apply deep learning to counterfeit detection [3]. The microscopic features of genuine products have unique attributes that can be used for identification. The application was constrained due to the dependence of specialist equipment. Tang's team used object detection and text recognition to identify samples [4] and developed the "Bao Xiaojian" counterfeit detection system. Wang et al. designed a lightweight CNN authentication model to identify texture material and font print of Gucci's black labels [5]. Arguing that both global and local information should be exploited for better performance, a two-stage method [6] extract features both from a whole word and its separate characters. All these works keep on improving the ability to identify non-significant differences. Actually, according to our research, a counterfeit detection application should rather be recognized as a fine-grained classification task due to two facts. Firstly, the discrimination between genuine and counterfeit gets weaker as upgrade of counterfeit craftsmanship. Secondly, the style of genuine goods varies among different series.

With aware of the similarity to the fine-grained classification task, techniques including attention mechanism, fine-grained classification and large visual model, are analyzed too. Wang et al. proposed ECA Net [7], which enhances the recognition ability of the model by introducing an attention module. Fine-grained classification networks [8–13] designed various mechanisms to push models to focus on critical visual areas thus addressing the issue of large intra-class differences and tiny inter-class differences. Since the Vision Transformer [14], many ViTs are put forward and achieved impressive improvements. Especially, the SwinTransformer [15], which introduces a shift window mechanism to improve performance while saving computational costs, is widely chosen as a backbone network to promote visual presentation. Although these methods are useful in learning the tiny differences between genuine and counterfeit goods, none of them can tolerate a small dataset.

Seeking a promotion over fine-grained discrimination, multi-task methods [16–28] are also taken account in. A multi-task learning focuses its design on parameter-sharing strategies for different tasks and for improving task correlation. Demonstrated by studies [29–31], appropriate auxiliary tasks can effectively promote the learning process of the other tasks. Although the multi-task learning mechanism has the potential to solve the problem of counterfeit detection with insufficient samples, the current research is limited in application of recommendation systems, and the research on counterfeit detection is still blank.

The proposed method addresses the aforementioned issues. It is composed of a targeted counterfeit detection network based on task characteristics and is able to learn fine-grained divergences on the challenging dataset with multiple categories and small samples in each class.

3 The proposed approach

Multi-series goods with similar but various visual characteristics are usually encountered in counterfeit detection task. What contradicts requirement of multi-classification is the poor number of samples. It is almost impossible to train a separate model for each series. Therefore, we decide to build a multi-classification model and leverage an auxiliary task assisting in promote fine-grained discrimination ability. The auxiliary task should not only be easy to learn, but also should be beneficial to attract the attention onto the critical visual areas and should not increase samples requirement.

The overall architecture, shown in Fig. 1, composes five stages: key area guidance, feature extraction, main task branch, auxiliary task branch, and loss aggregation. The label of the main task and the auxiliary task is denoted as $\{(y_i^1, y_i^2)|y_i^k \in \{0, 1\}, k \in \{1, 2\}\}$

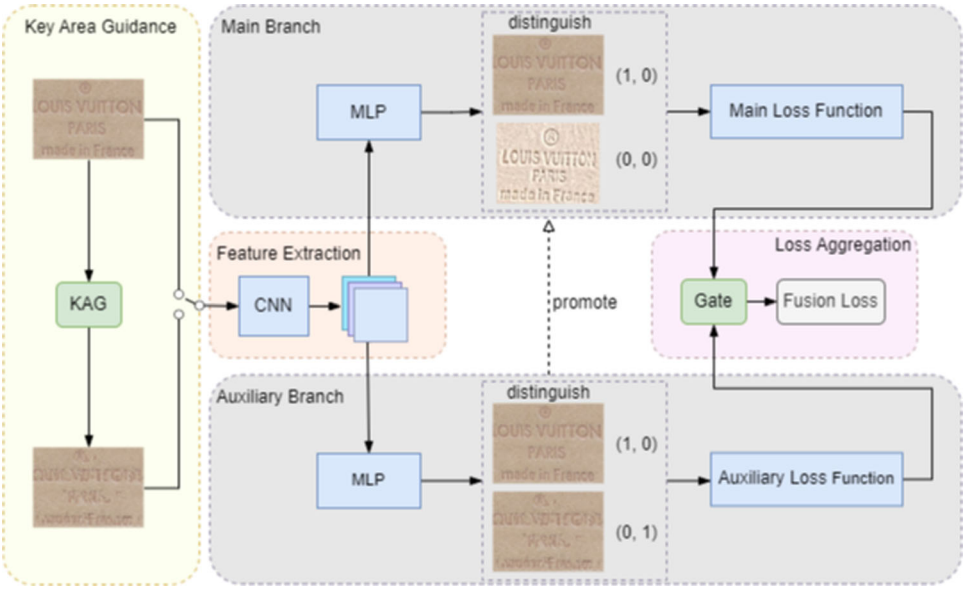


Fig. 1 The overall network architecture of the proposed method. The original images and the disturbed version generated by the KAG algorithm are sampled randomly, each has two attributes and then the corresponding feature is obtained by the feature extraction network. the main branch and the auxiliary branch make judge in turn on whether it is counterfeit and whether it is disturbed. In particular, no auxiliary branch is needed during the test process, and the model is lighter

Algorithm: KAG Algorithm

Input: Original image m, segment degree N;
Output: Scrambled image a;

1: Divide m into $N \times N$ patches dynamically;
2: for r in N:
3: for c in $N-1$:
5: $patch_{(r,c)} \in \{m_{(n,n)} \mid 0 < n < N\}$;
6: $patch_{(r,c+1)} \in \{m_{(n,n)} \mid 0 < n < N\}$;
7: $p =$ Bernoulli(0.5);
8: if $p = 1$:
9: $temp = patch_{(r,c)}$;
10: $patch_{(r,c)} = patch_{(r,c+1)}$;
11: $patch_{(r,c+1)} = temp$;
12: Output the scrambled image as a.

Fig. 2 The schematic diagram of KAG algorithm

3.1 Key area guidance

As aforementioned analysis, the auxiliary task need guide the attention of the main task. Inspired by some image augmentation techniques [32–34], we propose a simple but effective algorithm named KAG to enhance the focus of the main task by disturbing image patches. We also observed that DCL-Net [35] used a similar technique to drive the model to focus on detail differences, which confirms the validity of our method.

The KAG algorithm takes unprocessed raw images as input, the input image is first divided into several patches of the same size by segment degree, the segment degree N is used to control the granularity of image segmentation, where the input image is adaptively segmented into $N \times N$ patches. Then adjacent patches are randomly replaced, and the replacement condition follows the Bernoulli distribution with a probability of 0.5.

The effects of the KAG algorithm are mainly reflected in two aspects: sample quantity and sample attributes. In terms of quantity, the algorithm alters the character morphology of the generated new samples, increasing the available samples and reducing the model's reliance on sample quantity. Regarding attributes, regardless of whether the original sample is genuine or counterfeit, it will become counterfeit after being perturbed by the algorithm. Importantly, the non-core authentication attributes such as sample color, material texture, and brightness remain largely unchanged before and after KAG processing, while the character morphology changes, prompting the network to focus on the core areas.

The principle of the algorithm is shown in Fig. 2. In the KAG algorithm, adjacent patches of original samples are shuffled to simulate various counterfeit samples. This allows the network to recognize multiple patterns of counterfeit samples instead of being limited to a single pattern, so the problem of poor model performance caused by large intra-class differences can be relieved. Simultaneously, it indirectly increases the sample size.

The LV leather tag is taken as an example in Fig. 3, in which (a) is the original image, and (b) is the scrambled image generated by the KAG algorithm. It magnified the dif-

(a) original image (b) scrambled image

Fig. 3 The sample produced by KAG

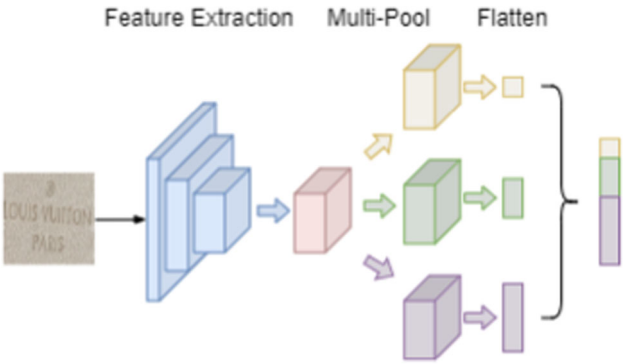


Fig. 4 The feature extractor

ferences of characters and simulated a rough counterfeiting process, while ensuring that the character space position is not excessively disturbed.

3.2 Feature extraction stage

In the application of counterfeit detection, fine-grained visual divergences need to be learned, which means deep network is required. Similar with many novel designs, ResNet50 [36] is chosen for feature extraction as shown in Fig. 4. The residual structure of ResNet50 transmits the shallow features to the deep layer by skip connection and combines the shallow texture and deep semantic of the input image, which is pretty beneficial. However, it is not necessarily the case that more complex networks yield better performance. Subsequent experiments have shown that ResNet50 outperforms other backbone networks.

Due to the high similarity between counterfeit products and genuine ones, relying on residual networks is inadequate. Therefore, we have incorporated multi-scale features by leveraging spatial pyramid pooling [37] to enhance the model's capability in capturing diverse scales of information. Different pooling windows are designed to capture different details of characters. In the specific implementation, three scales of adaptive pooling are applied for each feature map. The formula is as follows:

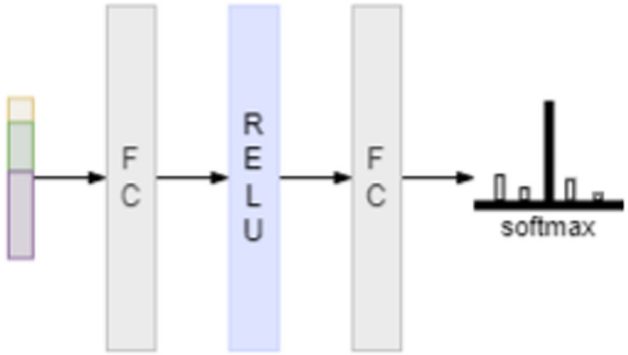


Fig. 5 The classifier

$$f_{i,j}^{t,l+1} = \text{average}\Big(f_{i,j}^{t,l}[\text{i} \times \frac{w^l}{w^{l+1}}; (\text{i}+1) \times \frac{w^l}{w^{l+1}}, \text{j} \times \frac{w^{l+1}}{h^{l+1}}; (\text{j}+1) \times \frac{w^{l+1}}{h^{l+1}}]\Big) \tag{1}$$

$$F = \text{concat}\big(f_{i,j}^{1,l+1}, f_{i,j}^{2,l+1}, f_{i,j}^{3,l+1}\big) \tag{2}$$

where $f \in \mathbb{R}^{w \times h}$ represents the input feature map, $t \in \{1, 2, 4\}$ represents three pooling levels, and i and j represent the element coordinates of the output feature map. w^l and h^l represent the width and height of the current input respectively. Finally, the three output feature maps are flattened into 1D vectors and concatenated to obtain the output feature vector.

3.3 Task branch

In the task branch, feature vectors are sent to the main and auxiliary branch, respectively. The auxiliary task recognizes the character disturbance generated by the KAG algorithm, and the main task discriminates whether samples are genuine or counterfeit ones. As the disturbance caused by the KAG algorithm are much more significant than the divergences between the genuine and counterfeit samples, so the auxiliary task can be easily optimized, thus guiding the attention of the main task.

It is found that in the proposed architecture, a simple classifier is enough to work well. The two task branches are implemented similarly by the classifier shown in Fig. 5, mainly composed of two fully connected layers. ReLU is applied in the middle, and the classification result is finally mapped to a probability distribution with the sum of 1 by softmax. The calculation method of softmax is as follows:

$$y_i = \frac{e^{p_i}}{\sum_{i=1}^{n} e^{p_i}} \in (0, 1) \tag{3}$$

where n denotes the total number of categories, $p_i \in \mathbb{R}^n$ is the prediction probability of the ith category of the model output, and y_i represents the probability of the ith category calculated by softmax, $\sum_{i=1}^{n} y_i = 1$.

3.4 Loss aggregation

The auxiliary task seeks optimum parameter to minimize its predication error of perturbation. According to that, the cross-entropy defined as follows is chosen as the loss function to evaluate a disturbance detection loss. For the main task, a same type of loss function is selected to reflect authenticity identification loss.

$$L(s, y) = -\sum_{i=1}^{c} s_i \log y_i \tag{4}$$

where c denotes the total number of categories. In the main task, it equals the number of categories for all genuine and counterfeit samples and is two in the auxiliary task. s_i is the label after one-hot coding, y_i represents the prediction normalized by softmax. The closer between prediction and ground truth, the smaller the loss is.

As the two tasks hold different importance, the aggregation needs to be carefully designed. Aiming to make the model thoroughly learn from the main and auxiliary task, the two losses are integrated by a weight coefficient α as follows:

$$L_{\text{fusion}} = \alpha \cdot L_{\text{main}} + (1 - \alpha) \cdot L_{\text{aux}} \tag{5}$$

where L_{main} and L_{aux} are separately denoted as the authenticity identification and perturbation detection loss. The optimization of the weight coefficient is discussed in the fourth part.

3.5 Preprocessing strategy

Images are always resized into shapes of uniform square before being fed to a DL network and therefore the caused deformation of objects may eliminate the none-significant visual differences between genuine and counterfeit samples. To address the issue, a preprocessing strategy named FWD is proposed for counterfeit detection. The FWD fills up an image before zooming an image and thus maintain its aspect ratio.

Figure 8 shows the strategy with different colors. Illustrated sequentially from left to right, are the original images, the directly resized ones, and those filled with black, gray, white, random colors, and the adaptive average color adjacent to the edge, respectively. Obviously, directly resizing the sample lengthened the letters and reduced the letter spacing. Such deformations are usually enough to challenge a state-of-the-art model of classification task. While with the FWD strategy, the impact of deformation can be mitigated.

4 Result and analysis

To evaluate the proposed method, the comparison among attention networks, fine-grained classification algorithm, ViT models and our model is presented. Besides, impact and selection of important designs are discussed too. In seek for an insight interpretation over the presented mechanism, t-SNE and GradCAM are applied.

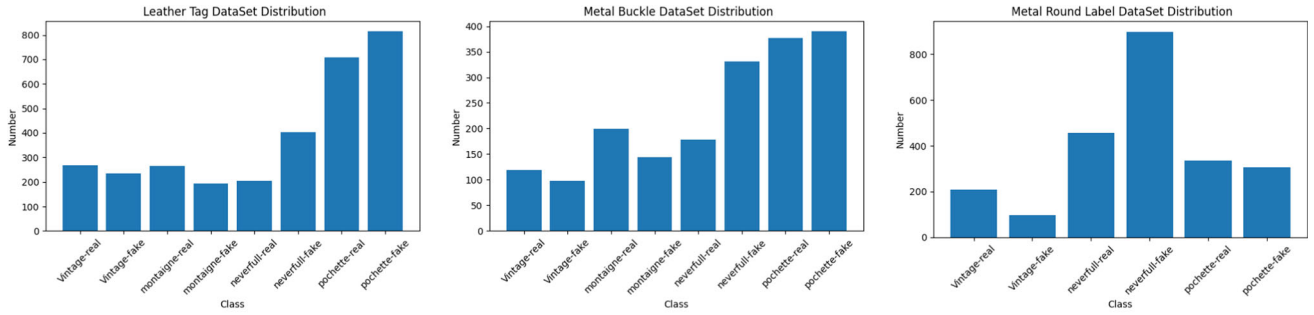
4.1 Experimental setup

In our method, the network has been trained for 200 epochs, and the batch size is set to 24. The SGD optimizer is used in the training process, momentum is set to 0.9 and weight decay is set to 0.001. The initial learning rate is 0.001, the cosine annealing strategy is employed, the attenuation cycle is ten epochs, and the minimum learning rate is 0.0005. α is set to 0.5. In the KAG algorithm, the segment degree is 20 by default. The images are automatically resized to 448×448 before being fed into the model. In the FWD strategy, the gray color is applied by default.

4.2 Dataset

Data need preprocessing before being further analyzed. Firstly, techniques like Gamma correction and auto contrast enhancement are applied over data captured by various devices to gain better visual quality. Secondly, an object location network, such as YOLOv7 [38] in our case, is leveraged to segment objects including leather tags, metal buckles, and metal round labels. The objection location network was pretrained with handcraft cut samples. Through the above steps, we extract sample images from the entire bag image for subsequent experiments.

We conducted experiments on the leather tag, metal buckle, and metal round label datasets, respectively. Each dataset contains luxury images from different series of LV brands in the real world. Each series includes two categories: genuine and counterfeit. The number of each category is diverse as a result of the tough collecting task. Consequently, it challenges the abilities of the models under unbalanced categories. The samples of these three datasets total in 3095, 1836, and 2300, respectively. The distribution over series in each dataset is shown in Fig. 6, and some samples are illustrated in Fig. 7. Each dataset is disjointly segmented into three subsets with a ratio of 8:1:1, namely the training set, the valid set, and the test set.



(a) leather tag dataset (b) metal buckle dataset (c) metal round label dataset

Fig. 6 The distribution of the three datasets

Fig. 7 Some samples in the three datasets

Fig. 8 The FWD strategy with various colors

4.3 Comparison and analysis

We carried out experiments on three datasets and evaluated the performance from three metrics: precision, recall, and f1-score. The results shown in Table 1 indicate the method works well on almost all conditions. The f1-score shows it recognizes the counterfeit metal buckles better than genuine ones. Although the model performs variously in different series but the metal round labels are still be well recognized.

The confusion matrixes on the test sets are illustrated in Fig. 9. The results indicate the good performance of our model in multi-series detection. It can accurately recognize the authenticity of multi-series samples, only a few samples are incorrectly identified as other categories.

Research on automated detection of luxury goods is scarce, our study primarily revolves around the key technologies involved in luxury detection, selecting state-of-the-art algorithms for comparison. Table 2 shows the average accuracy of each method on the three datasets. KAG-MTLN outperforms other methods, with a highest accuracy of 98.8%. Furthermore, the FWD strategy was discussed on various models. The strategy works well on the leather tag dataset, but not on others where square-shaped samples are predominant. Our experiments validate that the FWD strategy is more suited for samples with a larger aspect ratio.

Different training costs were compared in these methods. We extracted 100% (3095), 80% (2476), 60% (1857), 40% (1238), and 20% (619) of the samples from the leather tag dataset for experiments. In Fig. 10, the horizontal and vertical coordinates represent the sample size and test accuracy, respectively. At an 80% reduced sample size, all networks, except KAG-MTLN, saw notable accuracy drops due to limited data. KAG-MTLN outperformed MMAL Net by 4.8%, ECA Net by 46.1%, CBAM by 41.7%, API-Net by 9.6%, MHEM by 14.3%, and SwinTransformer by 31.8%. Our findings highlight that our method can significantly reduce the training cost with maintained performance.

4.4 Ablation experiment

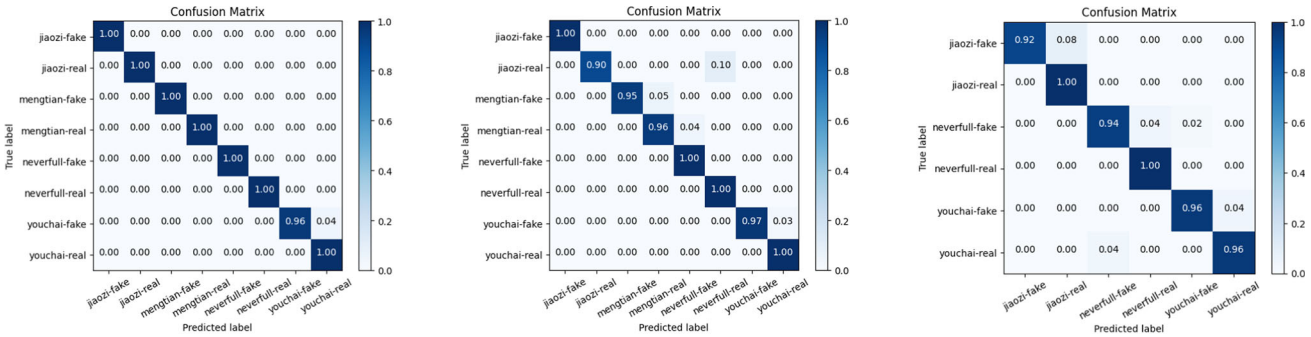
We further discussed the effectiveness of our method by adequate ablation experiments on the leather tag dataset.

4.4.1 Analysis over segment degree

To assess the effect of segment degree in the KAG algorithm, we experimented with different values. Table 3 shows that the optimum value is 20. In Fig. 11, we display samples with segment degrees of 20 and 50. A higher segment degree yields smaller patch divisions, and too large values will excessively destroy letter detail, leading to reduced performance.

Table 1 Performance metrics on three datasets

Category	Leather tag			Metal buckle			Metal round label		
	Precision (%)	Recall (%)	F1 (%)	Precision (%)	Recall (%)	F1 (%)	Precision (%)	Recall (%)	F1 (%)
Vintage-fake	100.0	100.0	100.0	100.0	100.0	100.0	100.0	91.7	95.7
Vintage-real	100.0	100.0	100.0	100.0	90.0	94.7	95.8	100.0	97.9
Montaigne-fake	100.0	100.0	100.0	100.0	94.7	97.3	98.9	93.9	96.3
Montaigne-real	100.0	100.0	100.0	95.7	95.7	95.7	90.0	100.0	94.7
Neverfull-fake	100.0	100.0	100.0	96.8	100.0	98.4	92.0	95.8	93.9
Neverfull-real	100.0	100.0	100.0	96.0	100.0	98.0	96.2	96.2	96.2
Pochette-fake	100.0	95.7	97.8	100.0	97.0	98.5	–	–	–
Pochette-real	94.8	100.0	97.3	97.3	100.0	98.6	–	–	–



(a) leather tag (b) metal buckle (c) metal round label

Fig. 9 The confusion matrix of three datasets

Table 2 The comparison of the average accuracy between KAGMTLN and other advanced methods on three datasets where "+FWD" means to use the FWD strategy

Method	Leather tag (%)	Metal buckle (%)	Metal round label (%)
ECA Net [7]	89.6	85.9	88.7
ECA Net (+FWD)	89.6	85.3	89.1
CBAM [39]	89.9	85.3	88.3
CBAM (+FWD)	90.5	86.4	90.0
SwinTransformer [15]	78.9	88.0	89.1
SwinTransformer (+FWD)	82.3	81.5	89.1
MMAL-Net [13]	97.5	97.3	97.4
MMAL-Net (+FWD)	97.8	92.9	96.1
API-Net [40]	98.4	96.2	**98.3**
API-Net (+FWD)	98.7	95.7	97.8
MHEM [41]	96.8	95.1	96.5
MHEM (+FWD)	98.4	96.2	96.5
KAG-MTLN	97.2	**98.6**	98.0
KAG-MTLN (+FWD)	**98.8**	97.8	95.9

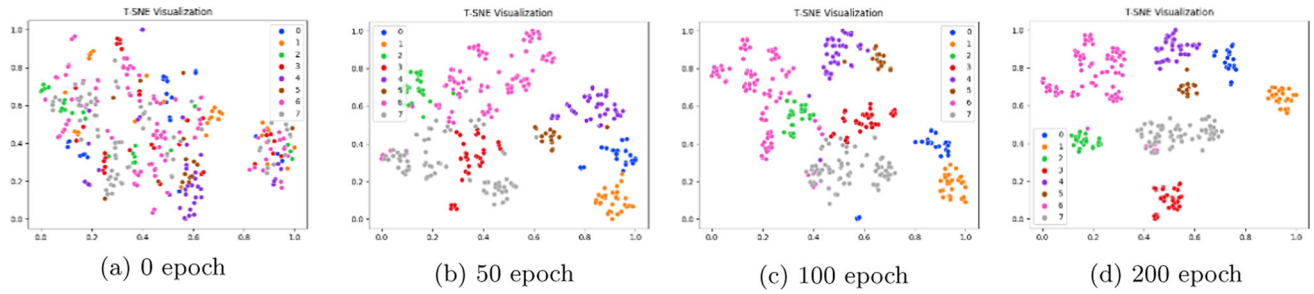
The best result is highlighted in bold

4.4.2 Different colors in FWD

To examine the impact of different image padding colors, we conducted experiments using black, white, gray, random, and uniform colors. The results, summarized in Table 4, indicate that gray and uniform colors yield the best performance.

4.4.3 Weight coefficient of loss function

To analyze the effect of weight coefficients in the loss function, we explored different values, as shown in Table 5. Optimal performance is achieved with a main coefficient of 0.5, striking a balance between the main and auxiliary tasks.



(a) 0 epoch (b) 50 epoch (c) 100 epoch (d) 200 epoch

Fig. 12 The visualization of t-SNE under different training epochs

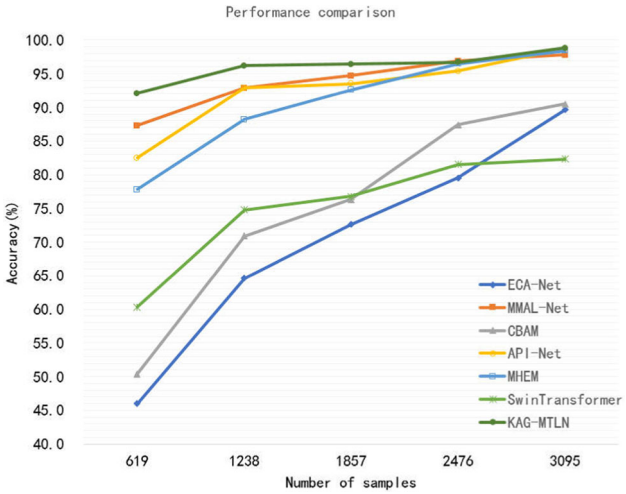


Fig. 10 The comparison of various models under different sample sizes. Each sample set contains multiple series

Table 3 performance of the KAG algorithm with different segment degrees

Segment degree	10	20	30	40	50
Acc	97.9%	98.8%	98.2%	97.2%	96.0%

Fig. 11 The impact of different segment degrees, the three images are the original image, the generated image with segment degrees of 20 and 50, respectively

Table 4 The performance comparison of different padding colors

Color	None	Black	White	Gray	Random	Uniform
Acc	97.2%	97.2%	98.5%	98.8%	98.2%	98.8%

Table 5 The performance under different main weight coefficients

α	0.9	0.8	0.7	0.6	0.5
Acc	97.6%	97.9%	97.9%	97.9%	98.8%

Table 6 Accuracy of different backbone models

Backbone	Acc (%)
VGG19	95.7
ResNet50	98.8
ResNet101	97.8
EfficientNet-B2	97.5

When the main coefficient is set to 1, the model becomes a single-task architecture.

4.4.4 Selection of backbone network

To investigate the impact of different backbone networks, we utilized a range of networks for feature extraction, as detailed in Table 6. These results confirm that our method's performance remains robust across diverse backbones, underscoring its reliability.

4.4.5 Effectiveness analysis of feature

To validate feature distinguishability, we employed t-SNE [42] to visualize the output of the final layer in the feature extraction stage at different training epochs. In Fig. 12, each colors signify a distinct category. Due to diverse processes by different counterfeiters, goods within the same category exhibit varied variations. Consider the purple dots, while distinct from other colors, they form multiple clusters. Overall, KAG-MTLN's extracted features are highly differentiated, effectively mitigating interference from intra-class differences.

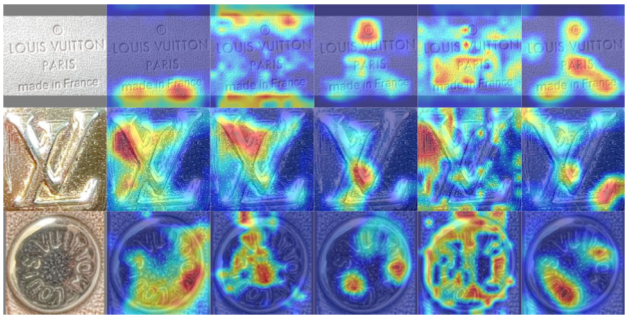


Fig. 13 Partial heatmap of the model on the test set, with each column sequentially from the original image, ECA Net, CBAM, MMAL Net, SwinTransformer, KAG-MTLN

Table 7 The effectiveness of auxiliary branch. "-auxiliary" represents the removal of the auxiliary branch

Method	Acc (%)
KAG-MTLN	98.8
-auxiliary	96.7

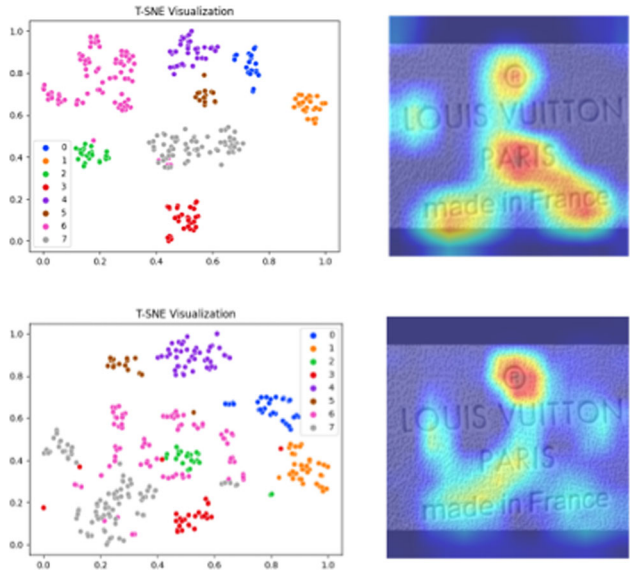


Fig. 14 The two lines are the t-SNE graph and heatmap before and after removing the auxiliary branch, respectively

4.4.6 Effectiveness analysis of key area

To assess the model's attention on key areas, GradCAM [43] was utilized to highlight focus regions. Figure 13 compares heatmaps from different models on the same sample. Authenticity indicators, such as the circular R mark for leather tags, end serif and inflection point for metal buckles, and circular font for metal round labels, are discerned. The heatmaps demonstrate our method's ability to discriminate based on specific key areas.

4.4.7 Effectiveness analysis of multi-task

To validate the effect of the auxiliary task, we conducted experiments by excluding the auxiliary branch. The results in Table 7 demonstrate that the auxiliary branch significantly improves the learning performance of the main branch, resulting in a 2.1% higher accuracy in our multi-task architecture compared to the single-task counterpart.

Simultaneously, the visualization results are contrasted before and after removing the auxiliary branch. In Fig. 14, the features extracted by the multi-task method are more clustered and differentiated and can focus on more details.

5 Conclusion

In conclusion, this paper proposes a counterfeit detection network based on key area guidance and multi-task learning. The experiments indicate that our method achieves superior performance with reduced training cost. Through visual analysis, the key areas of the sample can be highlighted by the method effectively. Furthermore, our method is not confined to the presented architecture, it can readily extend existing single-task methods to multi-task methods. Future work will focus on refining fusion strategies between diverse tasks and enhancing the key area guidance algorithm. Our method provides an efficient deep learning solution for intelligent counterfeit detection, contributing to fight against counterfeit products and safeguarding the legitimate rights and interests of consumers.

Author's contribution YY(Yaotian Yang) and YY(Yu Yang) conducted preliminary research and designed the method. LZ provided technical guidance. YY(Yaotian Yang) and YY(Yu Yang) carried out experiments. JZ provided and processed the data. YY(Yaotian Yang) wrote the manuscript. YY(Yu Yang) and LZ revised the manuscript. All authors reviewed the final version of the manuscript

Funding This work was supported by the National Key R&D Program of China (2022YFC3300803), the Natural Science Foundation of China (Grant No.62172053), the 111 Project (Grant No. B21049), and Open Foundation of Guizhou Provincial Key Laboratory of Public Big Data (2018BDKFJJ019).

Data availability The dataset and codes utilized in this paper are currently not publicly accessible; however, interested parties may acquire them by contacting the authors through reasonable requests.

Declarations

Conflict of interest The authors declare that they have no conflict of interest

Ethical approval Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Daping, L.: Shandong police releases ten major cases of cracking down on intellectual property infringement crimes. Prod. Reliab. Rep, 10–11 (2022)
2. Sharma, A., Srinivasan, V., Kanchan, V., Subramanian L.: The fake versus 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 (2017)
3. Sharma, A., Subramanian, L., Brewer, E.A.: Paperspeckle: microscopic fingerprinting of paper. In: Proceedings of the 18th ACM Conference on Computer and Communications Security, pp. 99–110 (2011)
4. Tang, Z., Wu, C., Lu, Y.: Training methods, systems, and equipment for item identification models (2019)
5. Wang, B.: Research adn application of real or fake label appraisal based on deep learning. Master's thesis, Xi'an University of Science and Technology (2020)
6. Peng, J., Zou, B., Zhu, C.: A two-stage deep learning framework for counterfeit luxury handbag detection in logo images. Sign. Image Video Process. **17**(4), 1439–1448 (2023)
7. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11534–11542 (2020)
8. Zheng, H., Fu, J., Zha, Z., Luo, J.:Looking for the devil in the details: learning trilinear attention sampling network for fine-grained image recognition. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5007–5016 (2019)
9. Bera, A., Wharton, Z., Liu, Y., Bessis, N., Behera, A.: Sr-gnn: spatial relation-aware graph neural network for fine-grained image categorization. IEEE Trans. Image Process **31**, 6017–6031 (2022)
10. Bera, A., Wharton, Z., Liu, Y., Bessis, N., Behera, A.: Sr-gnn: spatial relation-aware graph neural network for fine-grained image categorization. IEEE Trans. Image Process. **31**, 6017–6031 (2022)
11. Sun, H., He, X., Peng, Y.: Sim-trans: structure information modeling transformer for fine-grained visual categorization. In: Proceedings of the 30th ACM International Conference on Multimedia, pp. 5853–5861 (2022)
12. Ardhendu, B., Zachary, W., Hewage, P.R.P.G., Bera A.: Context-aware attentional pooling (cap) for fine-grained visual classification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 929–937 (2021)
13. Zhang, F., Li, M., Zhai, G., Liu, Y. Multi-branch and multi-scale attention learning for fine-grained visual categorization. In: MultiMedia Modeling: 27th International Conference, MMM 2021, Prague, Czech Republic, June 22–24, 2021, Proceedings, Part I 27, pp. 136–147. Springer (2021)
14. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. et al: An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
15. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)
16. Long, M., Cao, Z., Wang, J., Yu, P.S.: Learning multiple tasks with multilinear relationship networks. Adv. Neural Inf. Process. Syst. **30** (2017)
17. Misra, I., Shrivastava, A., Gupta, A., Hebert martial: cross-stitch networks for multi-task learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3994–4003 (2016)
18. Lu, Y., Kumar, A., Zhai, S., Cheng, Y., Javidi, T., Feris,R.: Fully-adaptive feature sharing in multi-task networks with applications in person attribute classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5334–5343 (2017)
19. Ma, J., Zhao, Z., Yi, X., Chen, J., Hong, L., Chi, E.H.: Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1930–1939 (2018)
20. Tang, H., Liu, J., Zhao, M., Gong, X.: Progressive layered extraction (ple): a novel multi-task learning (mtl) model for personalized recommendations. In: Proceedings of the 14th ACM Conference on Recommender Systems, pp. 269–278 (2020)
21. Liebel, L.: Marco: auxiliary tasks in multi-task learning. arXiv preprint arXiv:1805.06334 (2018)
22. Park, S., Lee, J., Kim, E.: Resource-efficient multi-task deep learning using a multi-path network. IEEE Access **10**, 32889–32899 (2022)
23. Ruiz, C., Alaíz, C.M., Dorronsoro, J.R.: Convex multi-task learning with neural networks. In International Conference on Hybrid Artificial Intelligence Systems, pp. 223–235. Springer (2022)
24. Cheng, G., Dong, L., Cai, W., Sun, C.: Multi-task reinforcement learning with attention-based mixture of experts. IEEE Robot. Autom. Lett. (2023)
25. Gondere, M.S., Schmidt-Thieme, L., Sharma, D.P., Scholz, R.: Multi-script handwritten digit recognition using multi-task learning. J. Intell. Fuzzy Syst. **43**(1), 355–364 (2022)
26. Rotman, G., Reichart, R.: Multi-task active learning for pre-trained transformer-based models. Trans. Assoc. Comput. Linguist. **10**, 1209–1228 (2022)
27. Yifan, X., Cui, Y., Jiang, X., Yin, Y., Ding, J., Li, L., Dongrui, W.: Inconsistency-based multi-task cooperative learning for emotion recognition. IEEE Trans. Affect. Comput. **13**(4), 2017–2027 (2022)
28. Gibson, J., Atkins, D.C., Creed, T.A., Imel, Z., Georgiou, P., Narayanan, S.: Multi-label multi-task deep learning for behavioral coding. IEEE Trans. Affect. Comput. **13**(1), 508–518 (2019)
29. Kung, P.-N., Yin, S.-S., Chen, Y.-C., Yang, T.-H. , Chen, Y.-N.: Efficient multi-task auxiliary learning: selecting auxiliary data by feature similarity. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 416–428 (2021)
30. Qiao, K., Liang, J., Liu, Z., Kunjie, Yu., Yue, C., Boyang, Q.: Evolutionary multitasking with global and local auxiliary tasks for constrained multi-objective optimization. IEEE/CAA J. Autom. Sin. **10**(10), 1951–1964 (2023)
31. Feng, Q., Chen, S.: Learning multi-tasks with inconsistent labels by using auxiliary big task. Fronti. Comput. Sci. **17**(5), 175342 (2023)
32. Chen, P., Liu, S., Zhao, H., Jia, J.: Gridmask data augmentation. *ArXiv*, abs/2001.04086, (2020)

33. Devries, T., Taylor, G.W: Improved regularization of convolutional neural networks with cutout. arxiv:1708.04552 (2017)
34. Kumar Singh, K., Jae Lee, Y.: Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3524–3533 (2017)
35. Chen, Y., Bai, Y., Zhang, W., Mei,T.: Destruction and construction learning for fine-grained image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5157–5166 (2019)
36. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
37. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Patt. Anal. Mach. Intell. **37**(9), 1904–1916 (2015)
38. Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y.M.: Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7464–7475 (2022)
39. Woo, S., Park, J., Lee, J.-Y., Kweon I.S.: Cbam: convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19 (2018)
40. Zhuang, P., Wang, Y., Qiao, Yu.: Learning attentive pairwise interaction for fine-grained classification. Proc. AAAI conf. Artif. Intell. **34**, 13130–13137 (2020)
41. Liang, Y., Zhu, L., Wang, X., Yang, Y.: Penalizing the hard example but not too much: a strong baseline for fine-grained visual classification. IEEE Trans. Neural Net. Learn. Syst. (2022). https://doi.org/10.1109/TNNLS.2022.3213563
42. van der Maaten, L., Hinton, G.E.: Visualizing data using t-sne. J. Mach. Learn. Res. **9**, 2579–2605 (2008)
43. Selvaraju, R.R, Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.