Expert Systems With Applications 229 (2023) 120315

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

MDM-CPS: A few-shot sample approach for source camera identification

Bo Wang [a], Jiayao Hou [a], Fei Wei [b,*], Fei Yu [a], Weiming Zheng [a]

[a] School of Information and Communication Engineering, Dalian University of Technology, Dalian, Liaoning, 116024, PR China
[b] School of Computing, National University of Singapore, 117417, Singapore

ARTICLE INFO

Keywords:
Source camera identification
Few-shot sample databases
Multi-distance measures
Coordinate pseudo-label selection

ABSTRACT

The purpose of source camera identification (SCI) is to identify the source device of target images, so as to ensure the source reliability of digital images. However, most state-of-the-art results require sufficient training samples which are hard to obtain in practice. In this work, we propose an approach based on multi-distance measures and coordinate pseudo-label selection (MDM-CPS) approach to solve the problem of few-shot sample databases. Based on semi-supervised learning, this approach iteratively expands and updates the labeled database. Our approach drastically reduces the interference of noisy pseudo-labels in training and ensures highly-confident prediction of the pseudo-label samples. Through comprehensive experiments, our approach has achieved the best performance in few-shot sample scenarios of the common benchmark databases (i.e., Dresden database and VISION database) in the field of source camera identification.

1. Introduction

Image is one of the most important information mediums. However, an enormous number of photo editing applications make it easy to tamper with images. In order to ensure the authenticity and reliability of image information, digital image forensics technology has played an important role. Considering the difficulty in obtaining digital watermark information in judicial forensics scenarios (Dadkhah et al., 2017; Lv, Xia, Zhao, Qiao, & Zhu, 2021), passive forensics technology is usually more appropriate. Source camera identification (SCI) has a significant position in the field of digital image passive forensics. The practical significance of this work is that in the actual judicial evidence-gathering cases, the forensics personnel can further confirm the suspect's crime tools by identifying the generating equipment of suspicious images, and provide important evidence for the follow-up investigation. In addition, there are also many cases in the market where images taken by SLRs cameras are used to fake propaganda by impersonating images taken by mobile phones. Therefore, the development of camera source identification technology is of great significance for judicial evidence collection and maintaining social stability. Some recent research shows that with the rapid development of technology, the research on source camera identification technology has very important practical significance (Akbari et al., 2022; Bernacki, 2020). Due to the difference in device hardware (e.g. sensor, processor) (Gupta & Tiwari, 2018; Li, Li, & Guan, 2018; Li, Lin, Kotegar, et al., 2021) and software (e.g. the build-in image generation algorithms in the image creating process) (Roy, Chakraborty, Sameer, & Naskar, 2017; Wang, Zhong, & Li, 2019; Xu, Wang, Zhou, Xi, & Wang, 2016), traces may be left in the images.

SCI problem has been widely studied in the past. However, the existing approaches can only obtain good model performance when the number of training set samples is sufficient. The model performance is poor when the training data sets are insufficient (a.k.a., few-shot sample databases). It is worth noting that in actual judicial forensics scenarios, obtaining sufficient labeled samples is usually expensive and time-consuming and sometimes impossible. Furthermore, in some short-term judicial forensics cases, it is not realistic to artificially produce a large number of data sets with labeled samples. This problem is closely related to the few-shot problem in the context of machine learning. Typical solutions include data expansion (Hu, Yang, Liu, Liu, & Wang, 2021; Wu et al., 2021), data enhancement (Osahor & Nasrabadi, 2022; Vu, Luong, Le, Simon, & Iyyer, 2021; Zhou, Zheng, Tang, Jian, & Yang, 2022), semi-supervised learning (Ding, Wang, Caverlee, & Liu, 2022; Huang, Geng, Jiang, Deng, & Xu, 2021; Zhang et al., 2022), etc.

In this paper, we propose the MDM-CPS approach to solve the problem of the labeled few-shot sample data set in source camera identification. The contributions of our research are summarized as follows:

- We use the multiple distance measures to expand the few-shot sample database, so as to solve the problem that the deep learning

* Corresponding author.
E-mail addresses: bowang@dlut.edu.cn (B. Wang), 32009157@mail.dlut.edu.cn (J. Hou), feiwei@nus.edu.sg (F. Wei), yf1999@mail.dlut.edu.cn (F. Yu), zwm32109144@mail.dlut.edu.cn (W. Zheng).

https://doi.org/10.1016/j.eswa.2023.120315
Received 3 February 2023; Received in revised form 11 April 2023; Accepted 28 April 2023
Available online 13 May 2023

approach based on coordinate pseudo-label selection does not perform well in few-shot sample sets.
- We use the coordinate pseudo-label selection approach to iteratively update the subset of positive and negative pseudo labels which are less noisy, where the coordinate attention blocks has the spatial long-range feature information interactions, thus effectively solving the problem of insufficient information in few-shot sample sets.
- We conducted comprehensive and reliable experiments on two important threshold parameters to ensure the best model performance. The experimental results also show that the proposed MDM-CPS approach is superior to other existing approaches.

2. Related works

Source Camera Identification. Generally speaking, the existing source camera identification approaches are mostly realized from the two following aspects: comparing the differences in the inherent characteristics of the camera hardware, and comparing the differences in the processing algorithms of the camera software.

In terms of camera hardware characteristics, Lukas, Fridrich, and Goljan (2006) first used sensor pattern noise (SPN) for source camera identification, which was caused by the inhomogeneity of the image sensor and the defects of the sensor manufacturing process. Zhao, Zheng, Qiao, and Xu (2019) designed a PRNU feature classifier based on a weight function to achieve feature dimensionality reduction. Bruni, Tartaglione, and Vitulano (2021) used the coherence of PRNU weighted estimations to measure the similarity of different regions of images, and demonstrated that this method is also robust to images from social networks. Hsiao, Takenouchi, Kikuchi, Sakiyama, and Miura (2021) proposed a secure pairing framework based on PRNU, extending the traditional device classification model to model, device, and other levels. Besides, Convolutional neural networks (CNNs) can also be used to infer camera fingerprint information (Freire-Obregón, Narducci, Barra, & Castrillon-Santana, 2019). Bennabhaktula, Alegre, Karastoyanova, and Azzopardi (2022) used the hierarchical classification (brand, model and equipment) ConvNets model based on homogeneous patches, and proved that the performance of this framework is better than single classifier.

Besides, depending on the post-processing algorithms of different camera models, the image source can be identified based on the camera's built-in image processing algorithms. Bayram, Sencar, Memon, and Avcibas (2005) proposed to use the traces left by the color filter array (CFA) interpolation algorithm of cameras for model classification. In addition, in the past 20 years, research on the use of CFA features in the field of image forensics has never stopped. Many studies (Akiyama, Tanaka, & Okutomi, 2015; Ferrara, Bianchi, De Rosa, & Piva, 2012; Huang & Suzuki, 2022; Suzuki & Kyochi, 2020; Swaminathan, Wu, & Liu, 2007) have shown that CFA features can be used as a basis for interpolation algorithms to distinguish different camera models, and a large number of experimental results have proven the effectiveness of CFA features. Deng, Gijsenij, and Zhang (2011) proposed an automatic white balance (AWB) approximation algorithm for the first time. Xu et al. (2016) combined local binary pattern (LBP) and local phase quantization (LPQ) for multiple feature classification. Villalba, Orozco, López, and Castro (2016) proposed a technology based on sensor noise and video-key-frame-based wavelet transform extraction. Rahim and Foozy (2020) proposed texture features based on gray level co-occurrence matrix (GLCM) and gray level run-length matrix (GLRLM). López, Orozco, and Villalba (2021) proposed block-based enhanced sensor fingerprint and studied the problem of source camera identification in an open scenario.

Few-shot Scenario. However, in the few-shot scenario, the performance of the approaches mentioned above reduces significantly. In order to solve the problem, Tan et al. (2015) used the set projection (EP) approach based on semi-supervised learning to construct multiple prototype sets to realize the expansion of few-shot sample sets. Liang, Zhang, and Zhang (2021) used the attention multi-source fusion few-shot learning (AMF-FSL) approach. Sameer and Naskar (2020) used the deep siamese network to maximize the training space of the model by inputting pairs of samples, so as to achieve data enhancement. Wu et al. (2021) expand few-shot sample sets by generating virtual samples based on mega-Trent-diffusion (MTD). Wang, Hou, Ma, Wang, and Wei (2022) expanded the labeled few-shot sample data set based on multiple distances, and then calibrated the pseudo labels with SVM self-correction mechanism to improve model performance. In summary, the key to solving the few-shot scenario is to expand the amount of information of few-shot sample data sets.

Unlike the existing few-shot sample scenario approaches, we creatively combine the interpretability of traditional features (i.e. CFA features, and provide visualization for their interpretability) with the optimized learning ability of deep learning networks. Furthermore, the interpretable properties of traditional features can filter pseudo-label samples more scientifically and effectively, while deep learning methods can obtain better feature learning capabilities. Finally, our MDM-CPS approach was proposed.

3. The MDM-CPS approach

In order to make full use of more detailed information on the few-shot samples, this chapter proposes an approach of few-shot sample source camera identification based on multiple distance measures and coordinate pseudo-label selection, named the MDM-CPS approach. The complete MDM-CPS approach framework diagram is shown in Fig. 1. This approach is based on semi-supervised learning, initially expanding the labeled database through multiple distance measures, and then iteratively updating the labeled database after finishing coordinate pseudo-label selection. Our goal is to expand the data by adding reliable pseudo-label samples to the few-shot sample database, so as to obtain higher source camera identification accuracy.

3.1. Phase I: The multiple distance measures module

This module in this paper is based on multiple distance measures in paper (Wang et al., 2022) and has been improved in our approach: By setting a new adaptive threshold n for the number of pseudo labels, the number of pseudo-label samples is reduced while obtaining a higher pseudo-label accuracy, making the deep learning model information based on pre-expanded data sets training more accurate.

Fig. 2 shows the frame diagram of feature extraction and multiple distance measures used to expand few-shot sample databases. After extracting the effective features of the sample, we use multiple distance measures to expand the labeled few-shot sample set. The specific approach is shown in Algorithm 1. For each dimension of the extracted sample features, we calculate multiple distance measures (Manhattan distance, Euclidean distance and Chebyshev distance) between each labeled sample and all unlabeled sample, and sort these distance measures in ascend order to obtain m unlabeled samples closest to each labeled sample under different measure rules. After that, in order to ensure the reliability of these samples, we separately count the times of each unlabeled sample for each class, and then select the first n unlabeled samples and assign them labels of the same class, thus completing the expansion of labeled few-shot sample set. These selected high reliability samples are pseudo-label samples.

Among them, we use CFA image statistical features as effective features. More details of the CFA features will be described in Section 4: Algorithm related details.

As mentioned before, the model of the deep learning algorithm cannot be fully trained due to insufficient data in few-shot scenarios, which impacts the accuracy of source identification. To solve this problem, we expanded the few-shot sample database using multiple distance measures to provide sufficient training samples for the subsequent

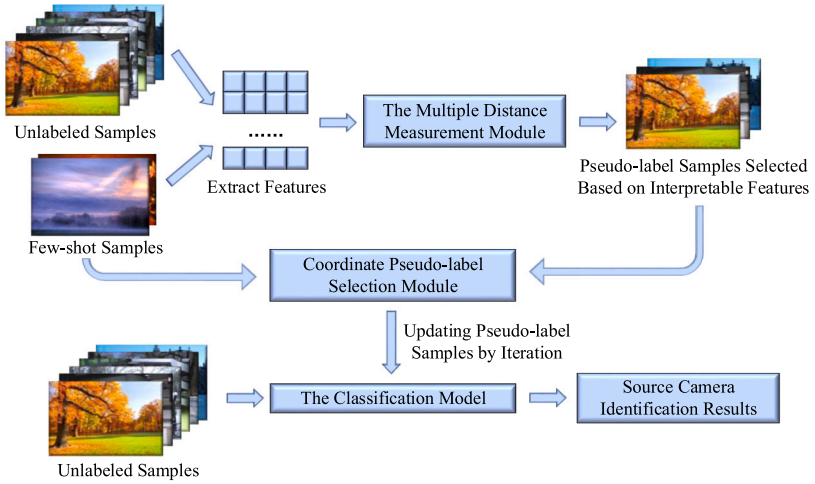


Fig. 1. The framework of MDM-CPS approach.

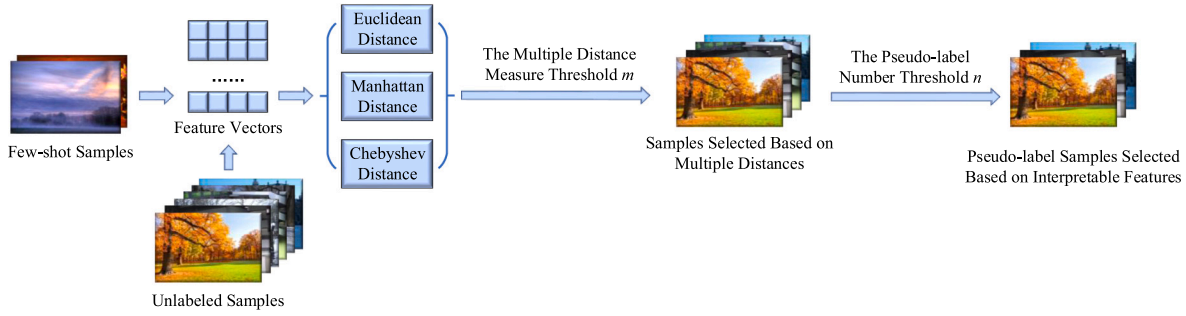


Fig. 2. The framework for feature extraction and multiple distance measures.

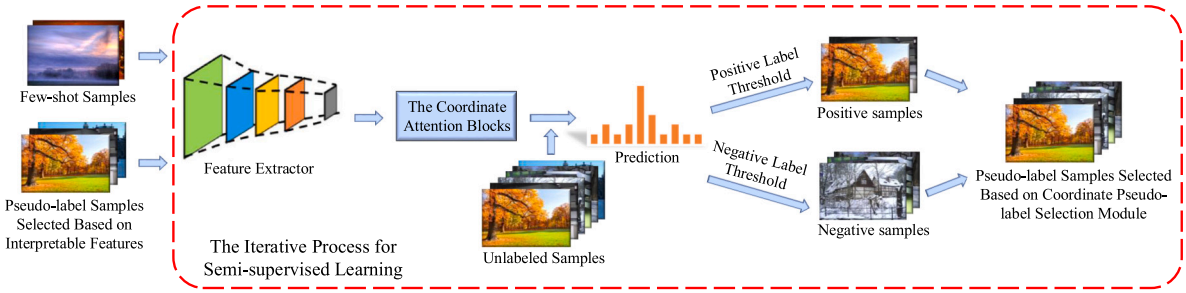


Fig. 3. The framework for coordinate pseudo-label selection module.

coordinate pseudo-label selection. However, note that as the number of labeled samples is limited, expanding over-numbered pseudo-label samples may cause issues such as incorrectly pseudo-label samples and impact the model performance. Accordingly, we limit the number of pseudo-label samples for each class in a controlled manner to ensure the accuracy of selecting pseudo-label samples.

3.2. Phase II: The coordinate pseudo-label selection module

After expanding at least n samples per camera category, we get pseudo-label samples selected based on interpretable features and multiple distance measures and include them in the existing few-shot sample database to build a new training sample database. Next, we use the coordinate pseudo-label selection module for semi-supervised learning. The framework diagram for the coordinate pseudo-label selection module is shown in Fig. 3. The coordinate attention module in Hou, Zhou, and Feng (2021) has the spatial long-range feature information interactions, and more details will be given in Section 4: Algorithm-related details. During iterations, we update the less-noisy pseudo-label sets (both positive and negative) by selecting the pseudo-label samples with higher confidence probability.

For multiple classification tasks, the traditional learning approach is to use the correct labels of samples. However, in the semi-supervised learning field, since the pseudo labels are sometimes inaccurate, the positive learning approach may provide some wrong information. As the training proceeds, the model will gradually fit noise labels, thereby reducing the performance. Fortunately, the negative learning approach can make full use of the information of noise labels (Kim, Yim, Yun, & Kim, 2019). Eqs. (1) and (2) show the cross-entropy loss function of positive learning and negative learning approaches respectively:

$$\mathcal{L}(f, y) = -\sum_{k=1}^{c} \boldsymbol{y}_k \log \boldsymbol{p}_k, \tag{1}$$

Algorithm 1 Pseudo-label samples selection algorithm based on multiple distance measures

Symbols:
$\mathcal{S}_u$: Set of unlabeled samples
$\mathcal{L}$: Set of few-shot labels
$\mathcal{S}_p$: Set of selected effective pseudo-label samples
$\mathcal{S}_l$: Set of samples with label $l \in \mathcal{L}$
$\mathcal{M}$: Collection of multiple distance measures
m: The multiple distance measure threshold for each labeled sample
n: The pseudo-label number threshold for each class
1. Extract feature vectors of all samples, including all labeled and unlabeled samples;
2. Calculate multiple distance measure vectors:
for $i \in \mathcal{M}$ and $s_l \in \{S_l : l \in \mathcal{L}\}$ **do**
Form multiple distance measure vectors $D_{s_l,i} = \{d^{(i)}_{s_l,s_u} : s_u \in S_u\}$
Sort $D_{s_l,i}$ in ascend order and take the first m entries for each sample to form a new vector $D'_{s_l,i}$
end for
3. Selecting effective pseudo-label samples:
for $l \in \mathcal{L}$ **do**
for $s_u \in S_u$ **do**
Count for s_u in $\{D'_{s_l,i} : s_l \in S_l, i \in \mathcal{M}\}$ and form count vector $C_{s_u,i}$
end for
Sort $C_{s_u,i}$ in ascend order and form the sample set S_p corresponding to the first n entries for each class
for $s_p \in S_p$ **do**
Label s_p with pseudo-label l (Note: these pseudo-label samples will be expanded to the corresponding few-shot sample sets)
end for
end for

$$\mathcal{L}(f, \bar{y}) = -\sum_{k=1}^{c} \overline{\boldsymbol{y}}_k \log\left(1 - \boldsymbol{p}_k\right), \tag{2}$$

where $\boldsymbol{y}_k$ is the real label, and $\boldsymbol{p}_k$ is the probability prediction distribution of the model output after the softmax activation function. $\overline{\boldsymbol{y}}_k$ is the complementary label in negative learning approach. For multiple classification tasks, the complementary label provides the information that the training samples do not belong to some classes. One sample can have multiple complementary labels. The closer the model predicts to the complementary label, the greater the loss of the negative learning approach, so that the model can make use of the supplementary label information.

For each sample $u \in \mathcal{U}$, the corresponding pseudo-label vector $\boldsymbol{g}^{(i)}$ can be determined according to the confidence thresholds for positive and negative labels (Rizve, Duarte, Rawat, & Shah, 2021):

$$g_k^{(i)} = \mathbb{1}\left[p_k^{(i)} \geq \tau_p\right] + \mathbb{1}\left[p_k^{(i)} \leq \tau_n\right], \tag{3}$$

where τ_p and τ_n are the confidence thresholds for positive and negative labels of each camera class respectively ($\tau_p, \tau_n \in (0, 1)$). If the probability score is sufficiently high, this sample is likely to belong to the current class, then the positive label is selected; On the contrary, if the probability score is sufficiently low, this sample is likely not to belong to the current class, then the negative label is selected. $\mathbb{1}$ means that this value is 1 when the probability prediction score is not lower than the confidence threshold, otherwise, it becomes 0.

Based on the high confidence pseudo-label vector $g_c^{(i)}$, we filter out the noise labels with insufficient confidence, so as to effectively reduce the noise interference during training. For multi-label classification, the cross-entropy loss based on the confidence threshold for positive and negative labels is modified to Eq. (4):

$$\mathcal{L}\left(\tilde{\boldsymbol{y}}_{\boldsymbol{k}}^{(i)}, \hat{\boldsymbol{y}}_{\boldsymbol{k}}^{(i)}, \boldsymbol{g}_{\boldsymbol{k}}^{(i)}\right) = -\frac{1}{s^{(i)}} \sum_{k=1}^{K} g_k^{(i)} \left[\tilde{y}_k^{(i)} \log\left(\hat{y}_k^{(i)}\right) + \left(1 - \tilde{y}_k^{(i)}\right) \log\left(1 - \hat{y}_k^{(i)}\right)\right] \tag{4}$$

where $s^{(i)}$ is the number of pseudo labels selected by sample i, $\hat{y}_k^{(i)}$ is the probability prediction output of the model, and $\tilde{y}_k^{(i)}$ is the pseudo labels assigned to unlabeled samples. By using a subset of pseudo-labels with low noise for iterative training, the accuracy of the pseudo-label samples is further improved, so the overall performance of the model is improved.

4. Algorithm related details

4.1. CFA features

For cost reasons, each pixel point data collected by CMOS/CCD image sensor in the camera provides only one color data (Red, Green, or Blue, i.e. RGB channel) for each pixel point, while the CFA interpolation algorithm can effectively restore the color information of the three channels. The CFA interpolation algorithms used by different camera manufacturers are generally different (Swaminathan et al., 2007). For each pixel and all pixels in its neighborhood, the interpolation model is calculated based on Eq. (5).

$$\mathbf{G} = a_1^g g_1 + \cdots + a^g_{(2k+1)^2-1} g_{(2k+1)^2-1} + a_1^r r_1 + \cdots + a^r_{(2k+1)^2} r_{(2k+1)^2} + a_1^b b_1 + \cdots + a^b_{(2k+1)^2} b_{(2k+1)^2} \tag{5}$$

Respectively, a_i^r, a_i^g and a_i^b are the CFA interpolation coefficient weights of the red, green and blue channels in the color images. r_α^k, g_α^k and b_α^k are the α interpolation coefficient of the neighborhood of the k pixel in these channels. According to the color distribution of Bayer CFA mode, the characteristic matrix of CFA interpolation coefficients can be obtained by solving the mean and variance of the G interpolation coefficients of R and B sampling points, as well as the R and B interpolation coefficients of two neighboring G sampling points.

It is worth noting that some studies have shown that it is effective to distinguish different camera models based on CFA interpolation features (Akiyama et al., 2015; Bayram et al., 2005; Ferrara et al., 2012; Huang & Suzuki, 2022; Suzuki & Kyochi, 2020; Swaminathan et al., 2007). This is because CFA features are extracted based on CFA interpolation algorithms for different camera models. In the spatial domain, the sum of the distances between the elements of multiple vectors can represent differences between feature vectors, and the larger the difference in distances, the greater the difference in CFA interpolation approaches. In addition, in the experimental section, we also provided a visual version of CFA features to further prove the rationality of distinguishing camera models based on CFA features.

4.2. Multiple distance measures

We use Manhattan distance (i.e., one norm), Euclidean distance (i.e., two norm), and Chebyshev distance (i.e., infinite norm) to calculate the spatial similarity between the CFA features of samples. The assignment of pseudo labels is realized by finding points with similar features in the spatial domain. The three distance measures are calculated in Eqs. (6), (7) and (8):

$$d_1 = |x_1 - x_2| + |y_1 - y_2| \tag{6}$$

$$d_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{7}$$

$$d_\infty = \max(|x_2 - x_1|, |y_2 - y_1|) \tag{8}$$

4.3. Semi-supervised learning

In many cases, it is too expensive or impossible to collect many labeled samples, however, a large number of unlabeled samples are available. Semi-supervised learning (SSL) is an effective approach at this time. In the semi-supervised learning approach, we can train the classifiers on labeled few-shot databases and predict the unlabeled samples. These unlabeled data predictions can be used as "pseudo labels" in the iteration of the classifiers, and a more stable and reliable network model can be obtained finally. Compared with the approach of only using labeled few-shot samples, semi supervised learning can use unlabeled samples to generate better decision boundaries, and therefore the training model has better performance.

4.4. Coordinate attention blocks

In the task of source camera identification, the networks need to focus on fingerprint information inside the training images. In paper (Hou et al., 2021), coordinate attention blocks are global-level attention mechanism, however, this module is not used for tasks in the field of camera source identification. We use the global feature information interaction capabilities of coordinate attention blocks to enable deep learning models to learn more comprehensive information about input images, thereby further improving the performance of our approach.

In order to encourage the interaction of attention blocks with long-range feature information in the spatial domain, the coordinate attention blocks generate a pair of direction-aware feature maps by aggregating the features of the input image along two spatial directions. After that, the direction-aware feature maps are weighted with the shared 1×1 convolution kernel and then weighted with different 1×1 convolution kernels of the two spatial directions. By multiplying the attention weights of the two spatial directions, the attention weights of each feature point can be obtained. In the process of coordinate attention blocks work, each feature point obtains the global attention weight and the attention weight along the two spatial directions, so as to realize the attention information interactions in the global and local spatial directions.

5. Experimental results and analysis

5.1. Experimental image databases and settings

In order to fully evaluate the performance of the MDM-CPS approach proposed in this paper, we have selected the most widely used public databases in source camera identification: Dresden database (Gloe & Böhme, 2010) and VISION (Shullani, Fontani, Iuliani, Al Shaya, & Piva, 2017) database.

In this experiment, we select 14 different brands of cameras (in Dresden database) and 11 different brands of cameras (in VISION database), as shown in Tables 1 and 2. Considering the influence of the number of labeled few-shot samples on the accuracy of source camera identification, we select the number of labeled samples ranging from 5, 10, 15, 20 and 25 in each class in the comparison experiment. Besides, the test database consists of 130–438 unlabeled samples in each class, with a total of 2791 (in the Dresden database) and 2163 (in the VISION database) image samples. Based on the multi-distance measures, we finally select n pseudo-label samples for each class from these unlabeled samples to expand the few-shot sample databases and then use the cooperative pseudo-label selection module to train a model to get the final source camera identification accuracy results.

Table 1
Database in experiments (Dresden).

Camera model	Abbr.	Number of samples
Agfa_DC-504	A1	167
Canon_PowerShotA640	C1	188
Casio_EX-Z150	C2	181
FujiFilm_FinePixJ50	F1	209
Kodak_M1063	K1	463
Nikon_CoolPixS710	N1	186
Olympus_mju_1050SW	O1	202
Panasonic_DMC-FZ50	P1	262
Pentax_OptioA40	P2	169
Praktica_DCZ5.9	P3	209
Ricoh_GX100	R1	192
Rollei_RCP-7325XS	R2	198
Samsung_L74wide	S1	231
Sony_DSC-H50	S2	284

Table 2
Database in experiments (VISION).

Camera model	Abbr.	Number of samples
Apple_iPad2	A1	171
Asus_Zenfone2Laser	A2	209
Huawei_Ascend	H1	155
Lenovo_P70A	L1	216
LG_D290	L2	227
Microsoft_Lumia640LTE	M1	187
OnePlus_A3000	O1	287
Samsung_GalaxyS3	S1	207
Sony_XperiaZ1Compact	S2	215
Wiko_Ridge4G	W1	253
Xiaomi_RedmiNote3	X1	311

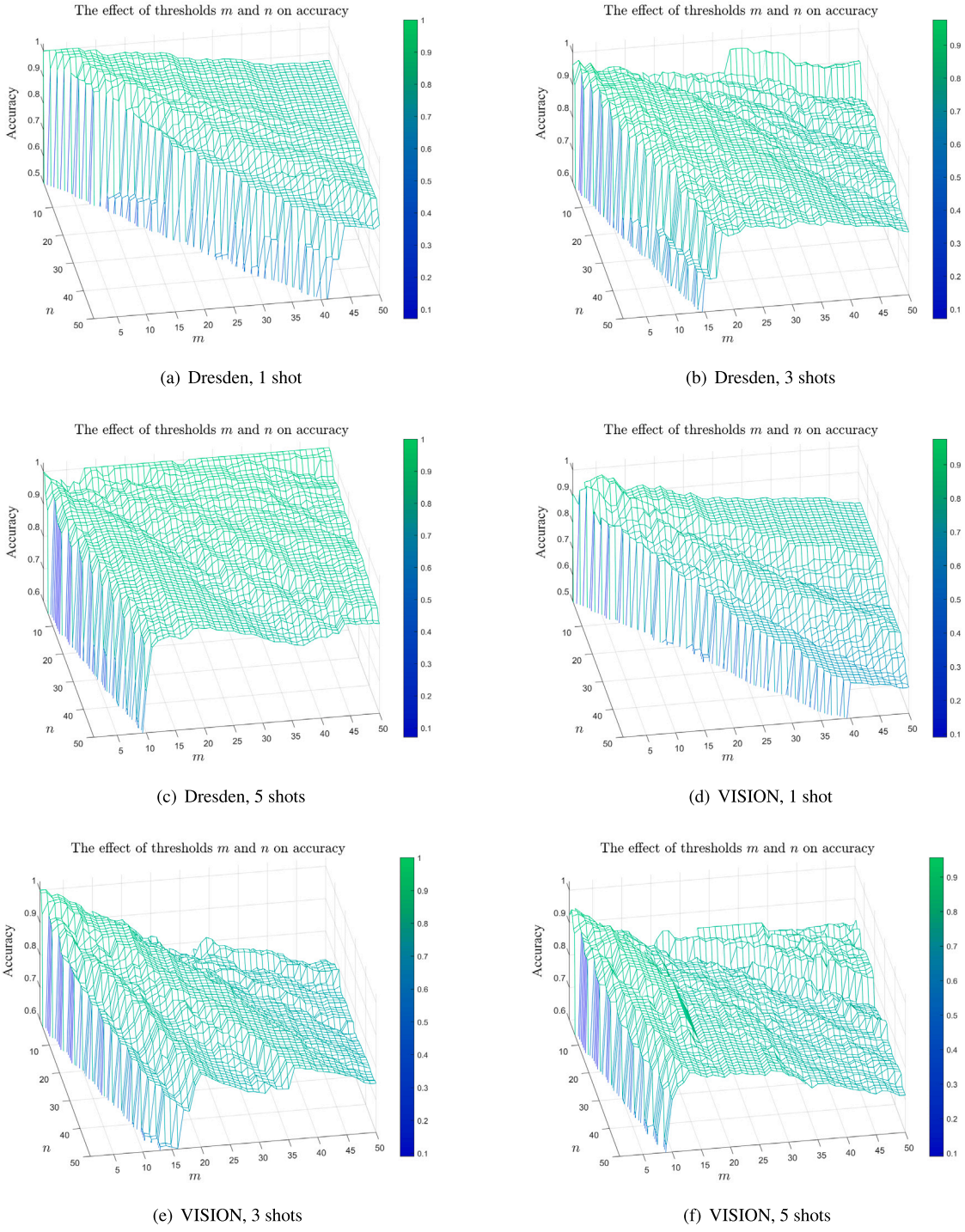
5.2. Algorithm performance evaluation and analysis

In the part of the multi-distance measure, in order to ensure the accuracy of the pseudo labels of the expansion of the few-shot sample set, we need to match the optimal parameter values of the thresholds m and n. When the number of few-shot samples in each class is 1, 3 and 5, and the value of m and n range from 1 to 50, we conducted sufficient comparative experiments. The results are shown in Fig. 4. When $m = n$, the accuracy rate of pseudo labels can reach the maximum point approximately, and when the value of n is smaller, it means that the number of pseudo-label samples finally selected is smaller and the accuracy rate is higher. Considering the balance between the number of expanded pseudo-label samples and the accuracy rate of the pseudo labels, we finally choose $m = n = 10$.

In the comparative experiment, we mark the approach using only the pseudo-label selection module as PS, and the approach using the pseudo-label selection module with coordinate attention as CPS. In addition, in order to demonstrate the effectiveness of the multi-distance measures, we mark the approach using the multi-distance measures and the pseudo-label selection module as MDM-PS and finally mark our proposed approach with all modules as MDM-CPS. The results of comparative experiments on multiple databases are shown in Figs. 5 and 6.

From the results of multiple experimental databases, each approach we proposed has a positive improvement in performance, which shows similar results on multiple databases. In addition, the smaller the number of few-shot samples, the more obvious the performance improvement of few-shot sample database expansion based on multi-distance measures, which shows that data expansion is an effective solution for the deep learning algorithm in the case of few-shot samples. The experimental results show that our proposed approach is widely applicable to solve the problem of few-shot sample source camera identification.

At the same time, in order to verify the stability of the model in an extremely few-shot sample environment, we carried out a series of



(a) Dresden, 1 shot

(b) Dresden, 3 shots

(c) Dresden, 5 shots

(d) VISION, 1 shot

(e) VISION, 3 shots

(f) VISION, 5 shots

Fig. 4. Source camera identification accuracy versus the multiple distance measure threshold m and the pseudo-label number threshold n on multiple databases.

stability experiments with the proposed MDM-CPS approach. When the number of labeled samples in each class is quite small (i.e., only one sample in each camera class), the quality of the selected samples will have a great impact on the final classification accuracy. Therefore, we conducted 20 groups of random repeated experiments when there was only one sample in each class and averaged the experimental results, as shown in Figs. 7 and 8. The experimental results show that our approach has reliable source camera identification performance in the case of extremely few-shot samples.

In addition, in order to evaluate the performance of our proposed MDM-CPS approach in the field of source camera identification, we have made fair comparison with other existing approaches under the same database and experimental settings (Sameer & Naskar, 2020; Tan et al., 2015; Wang et al., 2022; Wu et al., 2021). The experimental results in Table 3 show the superiority of our approach in the field of source camera identification.

In addition, in order to test the performance of our MDM-CPS approach in more complex situations, we re-selected all camera classes including different camera brands and models for each database. We selected 27 classes (in the Dresden database) and 35 classes (in the VISION database) to simulate the more complex situations in the actual judicial forensics scene. The experimental results are shown in Fig. 9,

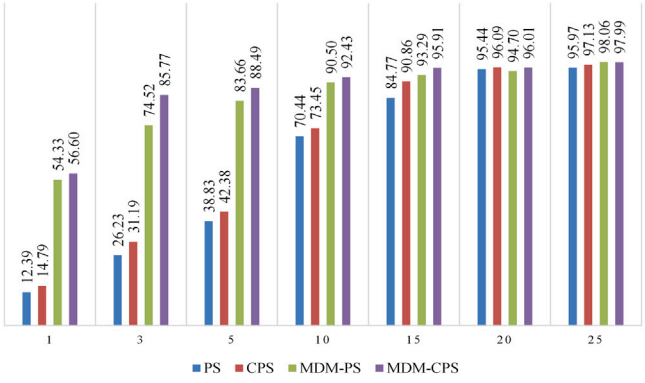


Fig. 5. Source camera identification accuracy for different approaches (%) (Dresden Database).

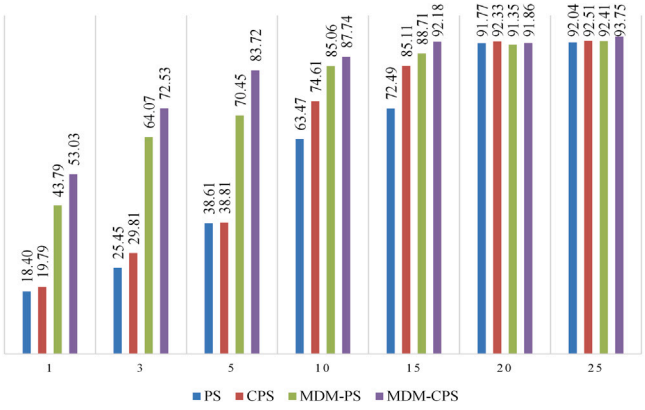


Fig. 6. Source camera identification accuracy for different approaches (%) (VISION Database).

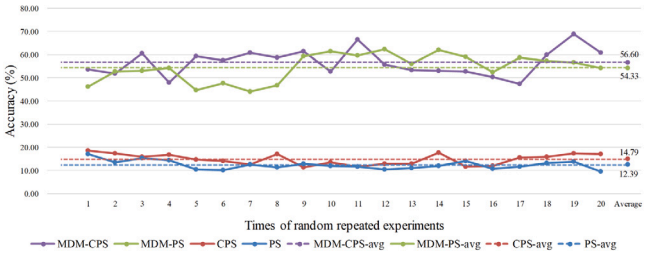


Fig. 7. Stability analysis with 1 sample per class in Dresden database (%).

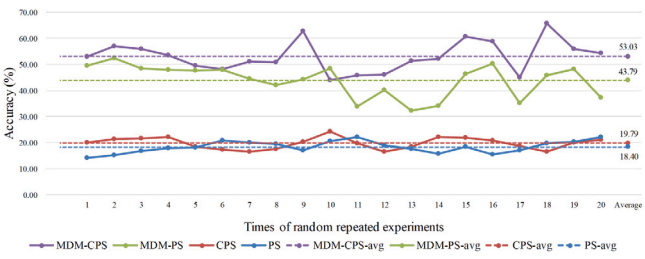


Fig. 8. Stability analysis with 1 sample per class in VISION database (%).

which fully proves the powerful performance of our model in dealing with complex multiple classification problems.

We compared traditional machine-learning-based Multi-DS approach, our proposed MDM-CPS approach, and the CPS approach of directly using semi-supervised deep learning. We accurately recorded

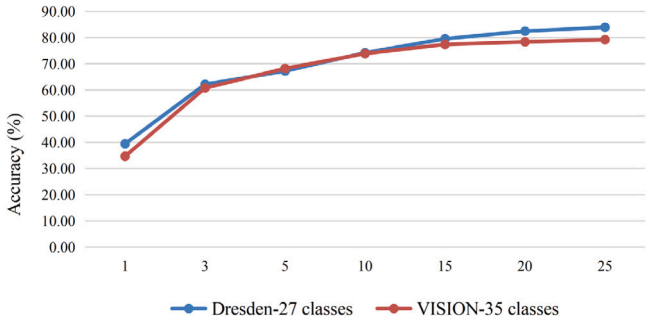


Fig. 9. Source camera identification accuracy for all camera models in multiple databases (%).

Table 3
Source camera identification accuracy compared with the existing approaches (%).

Approach	Dresden	VISION
Tan et al. (2015)	73.84	79.94
Wu et al. (2021)	75.16	80.49
Sameer and Naskar (2020)	85.30	75.20
Wang et al. (2022)	86.08	85.56
MDM-CPS	**92.43**	**87.74**

the running time of each group of experiments. The experimental results are shown in Table 4. From the experimental results, traditional machine learning based methods are faster in training classification models, while our approach is based on deep learning, which needs to iteratively update the pseudo-label sample data set with a large number of high-quality samples and requires many rounds of iterative training to obtain the best results. Therefore, deep learning models have better feature learning capabilities and better camera model classification performance. In addition, the training time of the Multi-DS-based machine learning method with respect to the number of samples is linearly increasing, with a time complexity of O(n). The CPS approach has a time complexity of O(n) when training based on few-shot sample sets, and the speed of growth of the MDM-CPS approach also has the time complexity of O(n) when training based on pre-expanded few-shot sample sets.

In Section 4.1, we discuss the effectiveness of CFA features. Here, we further demonstrate that CFA features can be used for source camera identification tasks through the experiment visualization results of CFA features. We considered two situations and visualized the CFA features separately: (1) CFA visualization results of photos taken by different camera models in the same scene; (2) CFA visualization results of photos taken by the same camera model in different scenes. The visualization experiment results are shown in Figs. 10 and 11.

From the experimental results, the distribution of CFA features in the spatial domain is most affected by the camera model, whereas variations in the actual captured scenes hardly affect the distribution of CFA features, This also explains why the source camera identification results from images with similar CFA features are consistent.

6. Conclusion

The performance of existing source camera identification approaches will deteriorate in few-shot sample source forensics scene. In this work, we propose the MDM-CPS approach to solve this problem. This approach expands the few-shot sample labeled database through the multi-distance measures and then uses the coordinate pseudo-label selection module to achieve the pseudo-label iterative learning of high confidence prediction for the expanded database, so as to reduce the interference of noise pseudo labels. The experimental results show that our proposed approach can effectively improve the accuracy of source camera identification when solving the few-shot sample problem, and

Table 4
Comparison of time complexity of different approaches (minutes).

Method and database	The number of labeled training samples per class														
	5			10			15			20			25		
	Phase I	Phase II	Total	Phase I	Phase II	Total	Phase I	Phase II	Total	Phase I	Phase II	Total	Phase I	Phase II	Total
Multi-DS, Dresden	\	\	1.30	\	\	2.58	\	\	3.90	\	\	5.14	\	\	6.45
Multi-DS, VISION	\	\	1.07	\	\	2.16	\	\	3.22	\	\	4.31	\	\	5.42
CPS, Dresden	6.50	122.01	128.61	13.01	122.95	135.96	19.53	124.03	143.56	26.05	124.35	150.40	32.56	125.03	157.59
CPS, VISION	5.10	94.03	99.13	10.20	94.08	104.28	15.30	94.15	109.45	20.41	94.22	114.63	25.51	94.33	119.84
MDM-CPS, Dresden	20.03	122.03	142.06	26.70	123.44	150.14	33.46	124.87	158.33	40.20	125.55	165.75	46.83	128.01	174.84
MDM-CPS, VISION	15.67	94.03	109.70	20.83	94.63	115.46	26.05	95.28	121.33	31.26	96.01	127.27	36.41	96.48	132.89

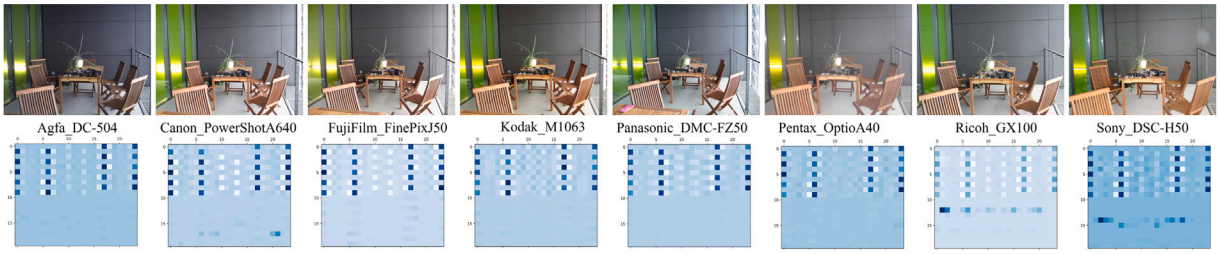


Fig. 10. CFA visualization results of photos taken by different camera models in the same scene.

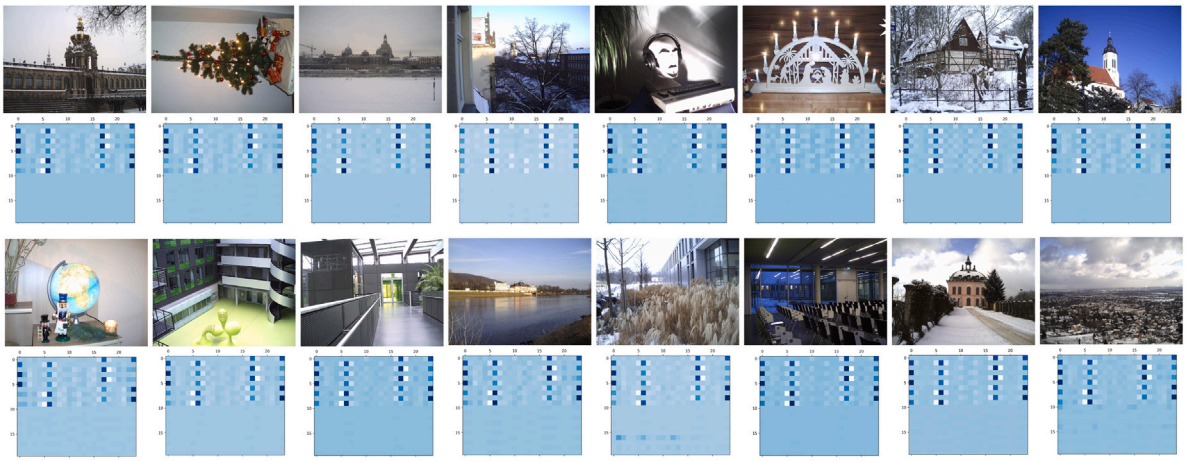


Fig. 11. CFA visualization results of photos taken by the same camera model (Agfa_DC-504) in different scenes.

our approach is superior to other existing approaches. Besides, in the case of extremely few-shot samples, our model can also ensure reliable performance, which provides a practical solution for the actual judicial forensics problems.

CRediT authorship contribution statement

Bo Wang: Conceptualization, Formal analysis, Investigation, Writing – review & editing, Supervision. **Jiayao Hou:** Conceptualization, Methodology, Validation, Investigation, Writing – original draft, Visualization. **Fei Wei:** Validation, Formal analysis, Data curation. **Fei Yu:** Investigation, Data curation. **Weiming Zheng:** Investigation, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. U1936117, No. 62106037, No. 62076052), the Science and Technology Innovation Foundation of Dalian (No. 2021JJ12GX018), the Application Fundamental Research Project of Liaoning Province (2022JH2/101300262), the Open Project Program of the National Laboratory of Pattern Recognition (NLPR), China (No. 202100032), the Fundamental Research Funds for the Central Universities, China (DUT21GF303, DUT20TD110, DUT20RC(3)088), and the Major Program of the National Social Science Foundation of China (No. 19ZDA127).

All authors have read and agreed to the manuscript.

References

Akbari, Y., Al-maadeed, S., Elharrouss, O., Khelifi, F., Lawgaly, A., & Bouridane, A. (2022). Digital forensic analysis for source video identification: A survey. *Forensic Science International: Digital Investigation*, *41*, Article 301390.

Akiyama, H., Tanaka, M., & Okutomi, M. (2015). Pseudo four-channel image denoising for noisy CFA raw data. In *2015 IEEE international conference on image processing* (pp. 4778–4782). IEEE.

Bayram, S., Sencar, H., Memon, N., & Avcibas, I. (2005). Source camera identification based on CFA interpolation. In *IEEE international conference on image processing 2005, Vol. 3* (pp. III–69). IEEE.

Bennabhaktula, G. S., Alegre, E., Karastoyanova, D., & Azzopardi, G. (2022). Camera model identification based on forensic traces extracted from homogeneous patches. *Expert Systems with Applications*, *206*, Article 117769.

Bernacki, J. (2020). A survey on digital camera identification methods. *Forensic Science International: Digital Investigation*, *34*, Article 300983.

Bruni, V., Tartaglione, M., & Vitulano, D. (2021). Coherence of PRNU weighted estimations for improved source camera identification. *Multimedia Tools and Applications*, 1–24.

Dadkhah, S., Köppen, M., Jalab, H. A., Sadeghi, S., Manaf, A. A., & Uliyan, D. M. (2017). Electromagnetismlike mechanism descriptor with Fourier transform for a passive copy-move forgery detection in digital image forensics. In *ICPRAM* (pp. 612–619).

Deng, Z., Gijsenij, A., & Zhang, J. (2011). Source camera identification using auto-white balance approximation. In *2011 international conference on computer vision* (pp. 57–64). IEEE.

Ding, K., Wang, J., Caverlee, J., & Liu, H. (2022). *Meta propagation networks for graph few-shot semi-supervised learning*. AAAI.

Ferrara, P., Bianchi, T., De Rosa, A., & Piva, A. (2012). Image forgery localization via fine-grained analysis of CFA artifacts. *IEEE Transactions on Information Forensics and Security*, *7*(5), 1566–1577.

Freire-Obregón, D., Narducci, F., Barra, S., & Castrillon-Santana, M. (2019). Deep learning for source camera identification on mobile devices. *Pattern Recognition Letters*, *126*, 86–91.

Gloe, T., & Böhme, R. (2010). The'dresden image database'for benchmarking digital image forensics. In *Proceedings of the 2010 ACM symposium on applied computing* (pp. 1584–1590).

Gupta, B., & Tiwari, M. (2018). Improving source camera identification performance using DCT based image frequency components dependent sensor pattern noise extraction method. *Digital Investigation*, *24*, 121–127.

Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 13713–13722).

Hsiao, A., Takenouchi, T., Kikuchi, H., Sakiyama, K., & Miura, N. (2021). More accurate and robust PRNU-based source camera identification with 3-step 3-class approach. In *International workshop on digital watermarking* (pp. 87–101). Springer.

Hu, X., Yang, Z., Liu, G., Liu, Q., & Wang, H. (2021). Virtual label expansion-highlighted key features for few-shot learning. In *2021 international joint conference on neural networks* (pp. 1–10). IEEE.

Huang, K., Geng, J., Jiang, W., Deng, X., & Xu, Z. (2021). Pseudo-loss confidence metric for semi-supervised few-shot learning. In *2021 IEEE/CVF international conference on computer vision* (pp. 8651–8660). IEEE Computer Society.

Huang, L., & Suzuki, T. (2022). Weighted wavelet-based spectral-spatial transforms for CFA-sampled raw camera image compression considering image features. In *ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing* (pp. 1850–1854). IEEE.

Kim, Y., Yim, J., Yun, J., & Kim, J. (2019). Nlnl: Negative learning for noisy labels. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 101–110).

Li, R., Li, C.-T., & Guan, Y. (2018). Inference of a compact representation of sensor fingerprint for source camera identification. *Pattern Recognition*, *74*, 556–567.

Li, C.-T., Lin, X., Kotegar, K. A., et al. (2021). Beyond PRNU: Learning robust device-specific fingerprint for source camera identification. arXiv preprint arXiv:2111.02144.

Liang, X., Zhang, Y., & Zhang, J. (2021). Attention multisource fusion-based deep few-shot learning for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 8773–8788.

López, R. R., Orozco, A. L. S., & Villalba, L. J. G. (2021). Compression effects and scene details on the source camera identification of digital videos. *Expert Systems with Applications*, *170*, Article 114515.

Lukas, J., Fridrich, J., & Goljan, M. (2006). Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, *1*(2), 205–214.

Lv, X., Xia, Y., Zhao, J., Qiao, P., & Zhu, B. (2021). Research on key technologies of digital multimedia passive forensics. In *2021 7th international conference on systems and informatics* (pp. 1–5). IEEE.

Osahor, U., & Nasrabadi, N. M. (2022). Ortho-shot: Low displacement rank regularization with data augmentation for few-shot learning. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 2200–2209).

Rahim, N., & Foozy, C. F. M. (2020). Source camera identification for online social network images using texture feature. In *International conference on soft computing and data mining* (pp. 283–296). Springer.

Rizve, M. N., Duarte, K., Rawat, Y. S., & Shah, M. (2021). In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. arXiv preprint arXiv:2101.06329.

Roy, A., Chakraborty, R. S., Sameer, V. U., & Naskar, R. (2017). Camera source identification using discrete cosine transform residue features and ensemble classifier. In *CVPR workshops* (pp. 1848–1854).

Sameer, V. U., & Naskar, R. (2020). Deep siamese network for limited labels classification in source camera identification. *Multimedia Tools and Applications*, *79*(37), 28079–28104.

Shullani, D., Fontani, M., Iuliani, M., Al Shaya, O., & Piva, A. (2017). VISION: A video and image dataset for source identification. *EURASIP Journal on Information Security*, *2017*(1), 1–16.

Suzuki, T., & Kyochi, S. (2020). Variable macropixel spectral-spatial transforms with intra-and inter-color decorrelations for arbitrary RGB CFA-sampled raw images. *IEEE Signal Processing Letters*, *27*, 466–470.

Swaminathan, A., Wu, M., & Liu, K. R. (2007). Nonintrusive component forensics of visual sensors using output images. *IEEE Transactions on Information Forensics and Security*, *2*(1), 91–106.

Tan, Y., Wang, B., Li, M., Guo, Y., Kong, X., & Shi, Y. (2015). Camera source identification with limited labeled training set. In *International workshop on digital watermarking* (pp. 18–27). Springer.

Villalba, L. J. G., Orozco, A. L. S., López, R. R., & Castro, J. H. (2016). Identification of smartphone brand and model via forensic video analysis. *Expert Systems with Applications*, *55*, 59–69.

Vu, T., Luong, M.-T., Le, Q., Simon, G., & Iyyer, M. (2021). STraTA: Self-training with task augmentation for better few-shot learning. In *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 5715–5731).

Wang, B., Hou, J., Ma, Y., Wang, F., & Wei, F. (2022). Multi-DS strategy for source camera identification in few-shot sample data sets. *Security and Communication Networks*, *2022*.

Wang, B., Zhong, K., & Li, M. (2019). Ensemble classifier based source camera identification using fusion features. *Multimedia Tools and Applications*, *78*(7), 8397–8422.

Wu, S., Wang, B., Zhao, J., Zhao, M., Zhong, K., & Guo, Y. (2021). Virtual sample generation and ensemble learning based image source identification with small training samples. *International Journal of Digital Crime and Forensics (IJDCF)*, *13*(3), 34–46.

Xu, B., Wang, X., Zhou, X., Xi, J., & Wang, S. (2016). Source camera identification from image texture features. *Neurocomputing*, *207*, 131–140.

Zhang, B., Ye, H., Yu, G., Wang, B., Wu, Y., Fan, J., et al. (2022). Sample-centric feature generation for semi-supervised few-shot learning. *IEEE Transactions on Image Processing*, *31*, 2309–2320.

Zhao, Y., Zheng, N., Qiao, T., & Xu, M. (2019). Source camera identification via low dimensional PRNU features. *Multimedia Tools and Applications*, *78*(7), 8247–8269.

Zhou, J., Zheng, Y., Tang, J., Jian, L., & Yang, Z. (2022). FlipDA: Effective and robust data augmentation for few-shot learning. In *Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: long papers)* (pp. 8646–8665).