Applied Intelligence (2026) 56:331
https://doi.org/10.1007/s10489-026-07351-6

A few-shot sample method for source camera identification based on residual information distillation

Huimin Liu[1] · **Jiayao Hou**[1] · **Jiaqi Chi**[1] · **Bo Wang**[1]

Received: 12 January 2026 / Accepted: 23 June 2026

Abstract

The objective of source camera identification (SCI) is to ascertain the originating device of target images, ensuring the reliability of digital image sources. However, current state-of-the-art techniques often rely on an abundant supply of training samples, which proves challenging to acquire in practical settings. To solve it, we propose a novel approach called residual information distillation (RID) to address the issue of few-shot sample databases. By leveraging the concept of knowledge transfer across datasets, our method optimizes model performance in scenarios with a limited number of training samples. The proposed approach involves training a teacher model using publicly available datasets that offer a wealth of samples. Subsequently, the knowledge encapsulated in the teacher model's weights is distilled into a student model, which is trained on few-shot sample datasets, to facilitate more effective training. Extensive experiments on Dresden, VISION, and SOCRatES datasets show that RID outperforms state-of-the-art methods. In the 10-shot setting, our method achieves 89.07%, 85.64%, and 79.08% accuracy, with notable improvements over existing approaches. It also maintains strong performance in 5-shot, cross-network, and cross-task distillation scenarios. The proposed RID effectively alleviates overfitting and improves generalization, making it practical for real-world few-shot forensic applications.

Keywords Source camera identification · Few-shot sample · Residual information distillation

1 Introduction

With the rapid development of digital imaging technology and multimedia communication, images have become one of the most important information carriers in daily life, judicial investigation, and network security. However, the widespread availability of easy-to-use image editing software has greatly reduced the difficulty of image tampering, forgery, and malicious dissemination. Such behaviors not only damage information credibility but also breed various illegal activities and security incidents. Under this background, digital image forensics technology has gradually become a research focus, providing critical technical support for copyright protection, evidence authentication, and cybercrime investigation.

In practical judicial forensics and real-world scenarios, most image data do not carry pre-embedded digital watermarks or authentication information [1, 2]. As a result, active forensics methods relying on prior information are difficult to apply, making passive digital image forensics the mainstream solution in actual detection. As a core branch of passive forensics, SCI aims to determine the specific imaging device that generates a given image by mining unique and stable device fingerprints. It plays an irreplaceable role in source tracing, evidence confirmation, and case investigation, and has important theoretical value and practical significance.

Over recent years, extensive studies have been devoted to source camera identification. According to the technical paradigm, existing methods can be divided into two categories: traditional handcrafted feature-based methods and

✉ Bo Wang
bowang@dlut.edu.cn

Huimin Liu
1292202003@qq.com

Jiayao Hou
769724342@qq.com

Jiaqi Chi
0703chi@mail.dlut.edu.cn

[1] School of Information and Communication Engineering, Dalian University of Technology, No.2 Linggong Road, Ganjingzi District, Dalian 116024, Liaoning, China

Published online: 06 July 2026

deep learning-based methods. Traditional methods mainly rely on camera inherent traces, such as sensor pattern noise (SPN), color filter array (CFA) interpolation artifacts, and statistical features, combined with classifiers (e.g., SVM) to complete identification [3–5]. Deep learning-based methods automatically learn discriminative device fingerprints through convolutional neural networks (CNNs), achieving superior performance under large-scale annotated datasets [6, 7]. Despite considerable progress, nearly all state-of-the-art methods heavily depend on sufficient training samples to obtain generalized feature representations.

Unfortunately, in real-world forensic applications, collecting large-scale annotated camera images is often restricted by limited manpower, material resources, and scene constraints. This leads to a critical few-shot learning challenge in source camera identification: only a very small number of labeled samples (e.g., 5 to 25 samples per class) are available for model training. As verified in Table 1, when training samples are extremely scarce, both traditional machine learning and deep learning models suffer from severe performance degradation, overfitting, and poor generalization ability. This bottleneck greatly limits the practical deployment of existing SCI methods in real forensics scenarios.

To alleviate the few-shot problem in SCI, researchers have proposed several strategies, including virtual sample generation, data augmentation, semi-supervised learning, and feature expansion [8]. Although these methods improve the utilization of limited samples to a certain extent, they still fail to introduce external generalized prior knowledge, resulting in limited performance improvement. More importantly, these methods are inherently unsuitable for the SCI task due to the unique characteristics of camera fingerprints. Common data augmentation operations such as rotation, flipping, scaling and color jittering alter the subtle noise patterns that constitute camera fingerprints, destroying the device-specific traces that SCI relies on [8]. Virtual sample generation methods based on generative models tend to learn dominant image content rather than weak noise-level features, making it difficult to produce authentic camera fingerprint information [9]. Semi-supervised learning methods require large amounts of unlabeled data from the same distribution, which is often unavailable in forensic scenarios due to privacy constraints and evidence confidentiality. Knowledge distillation provides a feasible solution to transfer rich knowledge from a pre-trained teacher model to a lightweight student model. However, conventional knowledge distillation is restricted to same-layer supervision and cannot achieve efficient cross-level knowledge interaction, making it difficult to fully mine and transfer discriminative camera fingerprint information.

To address the above challenges and limitations, this paper proposes a Residual Information Distillation framework for few-shot source camera identification. In this framework, the core concept of residual information refers to the difference between the feature representations of corresponding layers in the teacher and student models, which captures the discriminative camera fingerprint knowledge that the student model has not yet learned. The core idea is to transfer generalized device knowledge learned from large-scale datasets to the few-shot student model via cross-level residual distillation, adaptive feature fusion, and hierarchical context loss supervision. The main contributions of this work are summarized as follows:

- To tackle the information insufficiency problem in few-shot SCI, we propose a novel RID framework. It adopts a cross-dataset knowledge transfer paradigm to migrate generalized camera fingerprint knowledge from large-scale public datasets to few-shot student models, effectively alleviating overfitting and poor generalization.
- We design three synergistic components to enhance distillation efficiency: a cross-level residual distillation mechanism that breaks traditional same-layer supervision limitation and enables deep knowledge interaction; an attention-based fusion module to adaptively aggregate cross-level features with dimension mismatch; and a hierarchical context loss module that decomposes multi-level information via spatial pyramid pooling, combined with a weighted loss function for stable optimization.
- Extensive experiments on Dresden, VISION and SOCRatES datasets show that RID achieves state-of-the-art performance. It obtains 89.07%, 85.64% and 79.08% accuracy in the 10-shot setting, with up to 3.14% improvement over existing methods, and maintains strong robustness in 5-shot, cross-network and cross-task scenarios.

The structure of this paper is as follows: In Section 2, we delve into related work. Section 3 provides an overview of the methodology employed in this study. The analysis and

Table 1 Accuracy of source identification in the case of few-shot sample datasets (%)

Methods\Sample number per class	5	10	15	20	25	>100
LBP+SVM [8]	45.64	71.24	78.97	83.29	85.45	85.45
CFA+SVM [8]	49.75	69.99	78.67	82.68	84.89	92.53
CNNs [6]	29.83	50.95	63.85	75.68	85.64	99.82

experimental results are detailed in Section 4. Lastly, Section 5 encapsulates the discussion and conclusion.

2 Related work

To clarify why the proposed Residual Information Distillation framework is a natural and necessary next step, this section reviews three interconnected areas: (i) source camera identification methods that define the target task and its characteristic camera-fingerprint cues; (ii) few-shot learning strategies that have been explored to alleviate the data-scarcity challenge in SCI; and (iii) knowledge distillation and transfer learning approaches that provide a principled way to inject external generalized knowledge into low-data models. We then summarize the residual gaps that motivate RID.

2.1 Source camera identification

In general, the current source camera identification methods primarily focus on two aspects: analyzing the disparities in the inherent characteristics of camera hardware and evaluating the variances in camera software processing algorithms.

In terms of camera hardware characteristics, Lukas et al. [10] pioneered the use of SPN for SCI. They initially employed wavelet filters to denoise images and extract content information, then obtained residual noise images reflecting the hardware characteristics of camera sensors by subtracting content information from original images, achieving image source attribution based on these residuals. Zhao et al. [11] designed a weight-function-based PRNU feature classifier for dimensionality reduction. Bruni et al. [12] found that PRNU estimates extracted from natural images exhibit stronger robustness than those from flat-field regions, which is particularly relevant for images shared on social media. Hsiao et al. [13] proposed a PRNU-based secure pairing framework, extending traditional device classification to brand, model, and individual device levels.

Beyond traditional handcrafted features, CNNs have emerged as powerful tools for inferring camera fingerprint information [14]. Bennabhaktula et al. [15, 16] developed hierarchical classification ConvNets based on homogeneous patches, demonstrating that this framework outperforms single classifiers. More recently, Zheng et al. [17] proposed ADF-Net, which incorporates a lightweight SE-BRB module for efficient feature extraction, achieving a favorable balance between model accuracy and computational efficiency. Elharrouss et al. [7] further introduced PDC-ViT, which combines pixel difference convolution with vision transformers to capture subtle forensic traces beyond the receptive field of standard CNNs.

For camera software processing characteristics, source identification relies on traces left by built-in image processing pipelines, which vary across different camera models. Bayram et al. [18] first proposed utilizing traces from camera CFA interpolation algorithms for model classification. Over the past two decades, extensive research [19–23] has validated that CFA features can effectively distinguish between camera models with different interpolation algorithms. Villalba et al. [24] proposed a method combining sensor noise and wavelet transform extraction from video key frames. Sychandran et al. [25] designed a novel architecture integrating convolutional layers and residual blocks for camera model and sensor grade identification. Lopez et al. [26] developed a block-based enhanced sensor fingerprint method and evaluated its performance in open-set SCI scenarios. Furthermore, Han et al. [27] introduced a dual-branch contrastive learning framework that leverages spatial-frequency consistency constraints to strengthen camera fingerprint representations, significantly improving feature robustness against common image distortions.

Two observations emerge from the above review and connect this subsection to the next one. First, regardless of whether handcrafted or learned features are used, modern SCI methods all rely on extracting weak, device-specific traces (sensor noise, ISP artefacts, CFA interpolation residues), which are easily overwhelmed by dominant scene content. Second, both classifier-based and CNN-based pipelines achieve their reported accuracy only when sufficiently many labeled images per device are available, often on the order of hundreds. As a result, when only a few samples per device can be collected—which is the typical situation in real forensic casework—the very feature representations that make SCI work are the first to deteriorate, motivating the dedicated study of few-shot strategies discussed next.

2.2 Few-shot learning for source camera identification

While the aforementioned source camera identification methods have achieved remarkable performance on large-scale annotated datasets, their generalization ability and identification accuracy degrade significantly in few-shot scenarios, where only 5–25 labeled samples per camera model are typically available in real forensic applications.

Classic few-shot learning methods can be generally categorized into three paradigms: metric-based methods [28, 29], optimization-based methods [30], and data augmentation-based methods [31]. These paradigms have been widely applied to the SCI task, and researchers have proposed various domain-specific solutions.

To address this issue, Tan et al. [32] used the set projection (EP) method based on semi-supervised learning to

construct multiple prototype sets, enabling the expansion of few-shot sample sets. Liang et al. [33] used the attention multi-source fusion few-shot learning (AMF-FSL) method. Sameer et al. [34] used a deep siamese network to maximize the training space of the model by inputting pairs of samples, thereby achieving data augmentation. Wu et al. [9] expanded few-shot sample sets by generating virtual samples based on mega-Trent-diffusion (MTD). Wang et al. [8, 35] expanded the labeled few-shot sample dataset based on multiple distances and then calibrated the pseudo-labels using an SVM self-correction mechanism to improve the model's performance. Wang et al. [36] proposed the MDM-CPS approach based on multi-distance measures and semi-supervised learning, which iteratively expands and updates the labeled few-shot sample database.Overall, existing few-shot SCI methods mainly focus on expanding the information content of limited sample datasets, but they rarely introduce external generalized prior knowledge from large-scale datasets, which limits their performance improvement in extreme few-shot scenarios.

Beyond the few-shot learning methods specifically designed for SCI, recent advances in optimization-assisted machine learning and hybrid intelligent systems have also influenced the development of robust few-shot frameworks. Metaheuristic optimization algorithms such as Mountain Gazelle Optimizer [37] and Galactic Swarm Optimization [38] have been successfully applied to optimize neural network training processes, improving both accuracy and efficiency. Hybrid learning approaches that combine different feature extraction modules have shown enhanced generalization ability in data-scarce scenarios [39]. Furthermore, advanced data processing techniques such as the HENOS oversampling strategy [40] have addressed the challenge of imbalanced data distribution, which is also a common issue in forensic applications. These methodological advances have inspired the design of our RID framework, particularly in terms of multi-module synergy, weighted loss function optimization, and adaptive feature aggregation.

Despite their differences, the few-shot strategies surveyed above share a common premise: they attempt to make better use of the few labeled samples that are available, either through sample expansion, pseudo-labeling, metric learning, or optimization-aware training. None of them, however, addresses the more fundamental issue that a few-shot SCI dataset simply does not contain enough information to characterize a generalized camera fingerprint. Moreover, the data augmentation, virtual generation, and semi-supervised pipelines that work well for ordinary image classification can actively harm SCI, because the rotations, flips, colour jittering, and generative reconstructions they rely on alter or smooth out the very noise-level traces that SCI depends on. This naturally raises the question of whether external generalized prior knowledge can be transferred into the few-shot SCI model from outside the dataset itself, which is exactly what knowledge distillation and transfer learning aim to provide.

2.3 Knowledge distillation and transfer learning for low-data scenarios

Knowledge distillation, originally proposed by Hinton et al. [41], transfers "dark knowledge" from a high-capacity teacher model to a lightweight student model by encouraging the student to match the soft probability distribution of the teacher. Subsequent work generalized this idea from logit-level matching to feature-level matching: FitNets [42] aligned intermediate hidden representations, attention transfer [43] matched spatial attention maps, and relational distillation [44] preserved sample-pair relationships. More recently, Chen et al. [45] proposed knowledge review distillation, which allows deeper teacher layers to supervise shallower student layers and demonstrates that cross-level supervision is more informative than strict same-layer alignment. These advances have shown that distillation can effectively compress, regularize, and inject prior knowledge into student models without requiring additional labeled data.

Transfer learning is a complementary direction that aims to reuse knowledge learned from a source domain on a related target domain [46, 47]. In low-data scenarios, transfer learning has been used to pre-train backbones on large auxiliary datasets and then fine-tune them on small target datasets. The intuition is that low-level features (edges, textures, noise patterns) learned from rich data generalize across tasks, while only the task-specific head needs to be re-learned with limited samples. This paradigm has been particularly successful in medical image analysis, remote sensing, and other domains where annotation is expensive.

However, when these general-purpose distillation strategies are directly imported into SCI, two specific limitations emerge. First, conventional logit-level and same-layer feature distillation force the student to mimic the entire feature space of the teacher, which is wasteful: most of the teacher's representation is dominated by scene content, while only a small residual component carries the discriminative camera fingerprint. Second, camera fingerprints are intrinsically multi-scale—low-level sensor noise lives in shallow layers while ISP-induced artefacts live in deeper layers—so strict same-layer supervision cannot transmit the full hierarchy of forensic cues. To the best of our knowledge, no existing distillation framework explicitly targets the residual, cross-level, fingerprint-oriented transfer that few-shot SCI fundamentally needs. The proposed RID framework is therefore positioned at the intersection of these three areas:

it formulates the transferred quantity as residual information to remain faithful to the forensic essence of SCI, performs cross-level distillation to respect the hierarchical nature of camera fingerprints, and operates within a few-shot teacher–student protocol that injects generalized prior knowledge without altering the original noise-level feature distribution. The remainder of the paper develops this framework in detail.

3 Proposed method

The proposed RID framework presents three core innovations tailored for few-shot SCI. Unlike conventional knowledge distillation that transfers full features via strict same-layer supervision, RID adopts a residual-based paradigm, transferring only the discriminative camera fingerprint knowledge the student lacks to avoid redundant learning. It further breaks the layer-wise alignment limitation with cross-level residual transfer, enhanced by ABF and HCL modules for comprehensive multi-scale knowledge transfer. Optimized for few-shot SCI, RID leverages external generalized knowledge to alleviate overfitting without destroying subtle device traces, unlike data augmentation-based approaches. This method assists the few-shot sample model in better training and effectively alleviates the issues of overfitting and poor generalization performance often encountered by deep learning models in few-shot sample scenarios. The overall framework of the RID method is illustrated in Fig. 1.

3.1 Cross-level residual distillation module

Cross-level residual distillation (CRD) is the core of our RID framework. Notably, the concept of "residual information" in our method has an inherent connection with the core principle of traditional SCI. In classical forensic approaches, camera fingerprints (e.g., Photo Response Non-Uniformity, PRNU) are extracted as the noise residual of an image: the dominant image content is subtracted to retain only the subtle, device-specific sensor traces. Analogously, the residual information in RID is defined as the difference between teacher and student feature representations, which subtracts the general visual knowledge already learned by the student from the teacher's comprehensive knowledge, leaving precisely the discriminative camera fingerprint knowledge that the student lacks. This design aligns perfectly with the forensic nature of SCI, as it explicitly targets the extraction and transfer of the weak device-specific traces critical for accurate identification.

Traditional knowledge distillation adopts a strict layer-by-layer supervision paradigm, which fails to transfer cross-level complementary knowledge. This is particularly problematic for SCI, as camera fingerprints are multi-scale

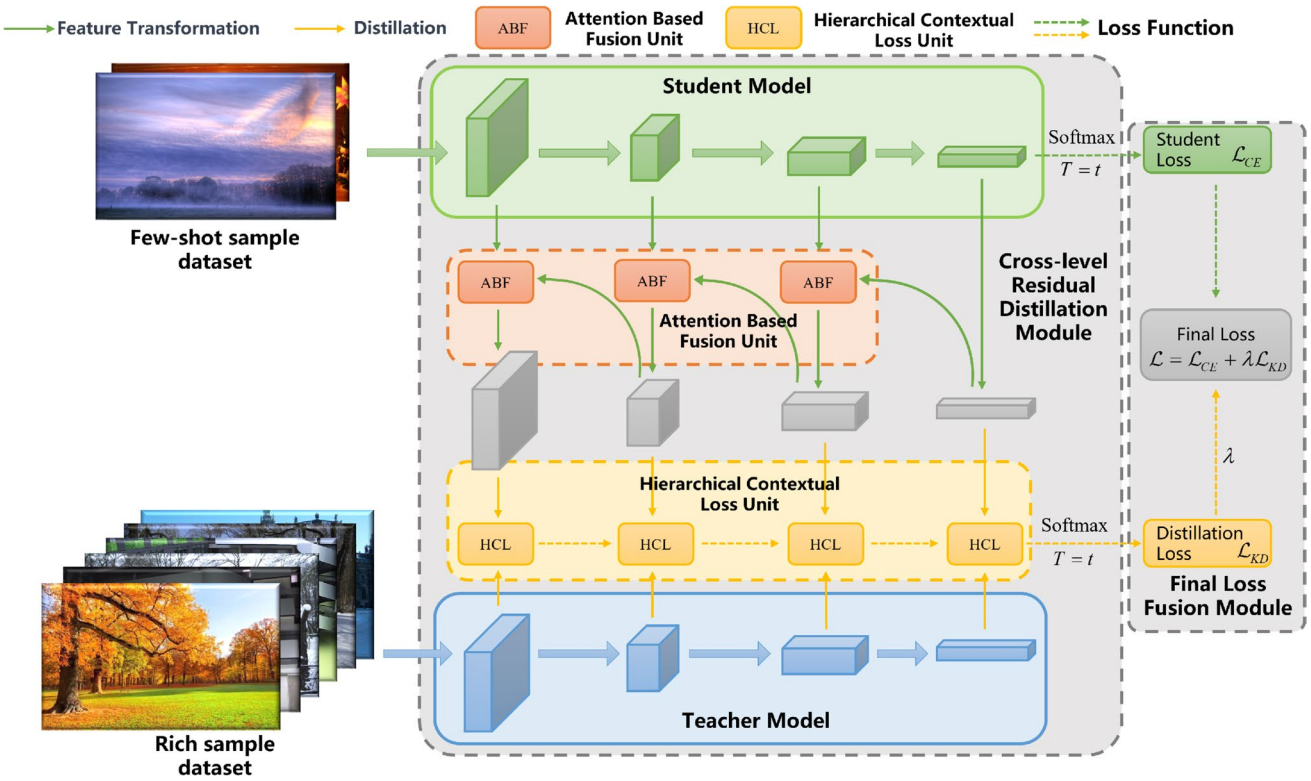

Fig. 1 The framework of RID method

weak features: shallow CNN layers encode low-level sensor noise patterns, while deep layers capture high-level ISP processing artifacts. Standard same-layer distillation leads to incomplete learning of hierarchical fingerprints, and this limitation is further amplified in few-shot scenarios where the student model cannot independently learn comprehensive representations.

To address it, our CRD module [45] introduces cross-level residual information interaction, allowing each student layer to learn from the corresponding and all shallower teacher layers. This enables comprehensive transfer of hierarchical camera fingerprint knowledge. The structure of the CRD module is illustrated in Fig. 2.

In the CRD module, the teacher model is first trained until convergence, and the parameter information of each hidden layer is saved in the model. This information is used to provide data information based on a rich sample dataset for the subsequent distillation process.

After obtaining the teacher model, this study employs cross-level residual information to distill the student model. Specifically, the concept of residual information originates from the composition of the distillation loss function $\mathcal{L}_{KD}$ in each layer of the HCL module. Unlike traditional methods that only compute distillation loss between the same hidden layers, $\mathcal{L}_{KD}$ in the CRD module focuses not only on the differences between the same layers but also on the student's shallow-level information. The cross-level information of the student is fused through the ABF module to achieve adaptive aggregation. In other words, each layer of the teacher model simultaneously supervises the corresponding and shallower layers of the student model.

In the CRD module, there are three notable issues to consider: the "deep-to-shallow" feature problem in CNN networks, the transmission of residual information in the "teacher-student" model, and the task discrepancy between the "teacher-student" models [41]. We explicitly link each challenge to the corresponding design solution in the RID framework as follows:

The deep-to-shallow feature problem in CNN networks: CNN networks often capture global features in deeper layers due to a larger receptive field, while local features are captured in shallower layers [48]. In the CRD module, it is important to address how to effectively transfer information from deep layers to shallow layers to ensure comprehensive knowledge transfer. Therefore, we design the cross-level residual distillation mechanism that allows each teacher layer to supervise not only the corresponding student layer but also all shallower student layers, enabling

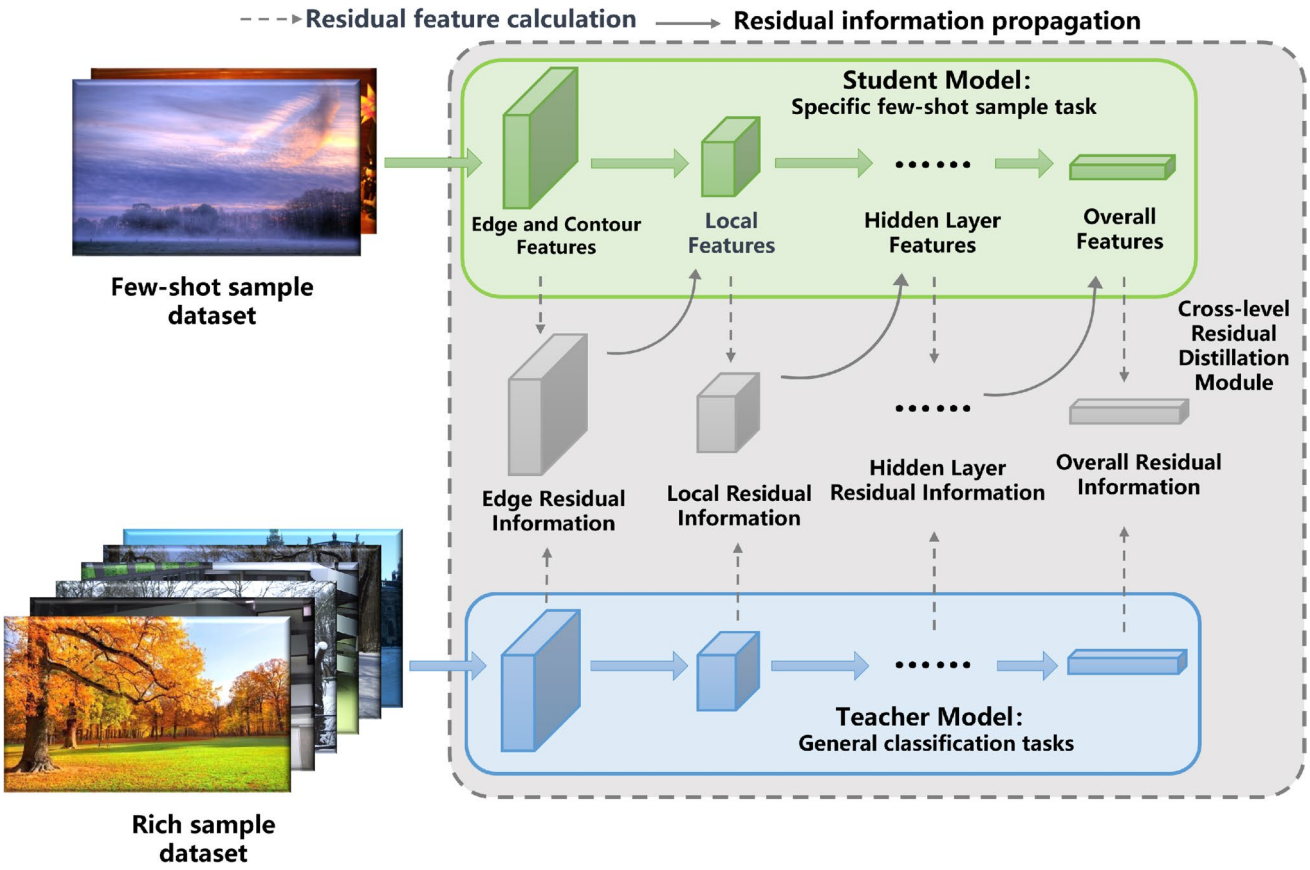

Fig. 2 The cross-level residual distillation module

the transfer of both deep semantic features and shallow edge/texture features.

The transmission of residual information in the "teacher-student" model: the CRD module focuses on distilling residual information from the teacher model to the student model. It is crucial to design an effective mechanism for transmitting this residual information, allowing the student model to benefit from the knowledge of the teacher model. Therefore, we propose the Attention Based Fusion (ABF) module to adaptively aggregate cross-level residual information, and the Hierarchical Context Loss (HCL) module to compute distillation loss at different abstraction levels, ensuring efficient and accurate residual information transmission.

The task discrepancy between the "teacher-student" models: the teacher and student models may have differences in their learning tasks or objectives. It is important to consider how to handle these task discrepancies and align the learning objectives of both models to facilitate effective knowledge distillation. Therefore, we adopt a weighted final loss function $\mathcal{L} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{KD}$ that combines the student's own prediction loss ($\mathcal{L}_{CE}$) and the distillation loss ($\mathcal{L}_{KD}$), allowing the model to balance learning the target SCI task with absorbing teacher knowledge.

In the end, the CRD module achieves the transfer of cross-level knowledge by continuously delivering the residual information between the corresponding network layers of the teacher and student models. This enables the student model to learn from the deeper layers of the network and facilitates knowledge transfer across layers.

3.2 Attention based fusion module

The structure of the ABF module is illustrated in Fig. 3. The main function of this module is to adaptively aggregate cross-level features with mismatched resolutions and channel dimensions, generating discriminative fused features for subsequent residual distillation.

Given a shallow feature map $F_s \in \mathbb{R}^{C_s \times H_s \times W_s}$ from the j-th layer of the student model and a deep feature map $F_d \in \mathbb{R}^{C_d \times H_d \times W_d}$ from the $(j+1)$-th layer, where $H_s = 2H_d$, $W_s = 2W_d$, and $C_s \neq C_d$, the ABF module first performs feature dimension alignment to resolve the mismatch in spatial resolution and channel number. Specifically, we upsample the deep feature to the spatial size of the shallow feature using bilinear interpolation:

$$F_d^{\text{up}} = \text{Upsample}(F_d, \text{size} = (H_s, W_s)) \tag{1}$$

and then adjust its channel dimension to match F_s via a 1×1 convolution layer followed by batch normalization and ReLU activation:

$$F_d^{\text{align}} = \text{Conv}_{1\times1}(F_d^{\text{up}}) \in \mathbb{R}^{C_s \times H_s \times W_s} \tag{2}$$

After dimension alignment, we concatenate the shallow feature F_s and the aligned deep feature F_d^{align} along the channel dimension to form a combined feature tensor:

$$F_{\text{cat}} = \text{Concat}(F_s, F_d^{\text{align}}) \in \mathbb{R}^{2C_s \times H_s \times W_s} \tag{3}$$

To adaptively weight the contributions of different channels, we apply a channel-wise attention sub-module to F_{cat}. First, global average pooling is used to aggregate spatial information into a channel-wise statistic vector:

$$z_c = \frac{1}{H_s W_s} \sum_{i=1}^{H_s} \sum_{j=1}^{W_s} F_{\text{cat}}(c, i, j), \quad c = 1, 2, \ldots, 2C_s \tag{4}$$

where $z = [z_1, z_2, \ldots, z_{2C_s}] \in \mathbb{R}^{2C_s}$ captures the global response of each channel. Next, two fully connected (FC) layers with a bottleneck structure are employed to model non-linear channel-wise dependencies:

$$s = \sigma\left(\text{FC}_2\left(\text{ReLU}\left(\text{FC}_1(z)\right)\right)\right) \in \mathbb{R}^{2C_s} \tag{5}$$

where σ denotes the sigmoid activation function, $\text{FC}_1 \in \mathbb{R}^{C_s \times 2C_s}$ reduces the channel dimension by half to introduce non-linearity and reduce computational cost, and $\text{FC}_2 \in \mathbb{R}^{2C_s \times C_s}$ restores the original channel dimension. Finally, the learned attention weights s are broadcasted to

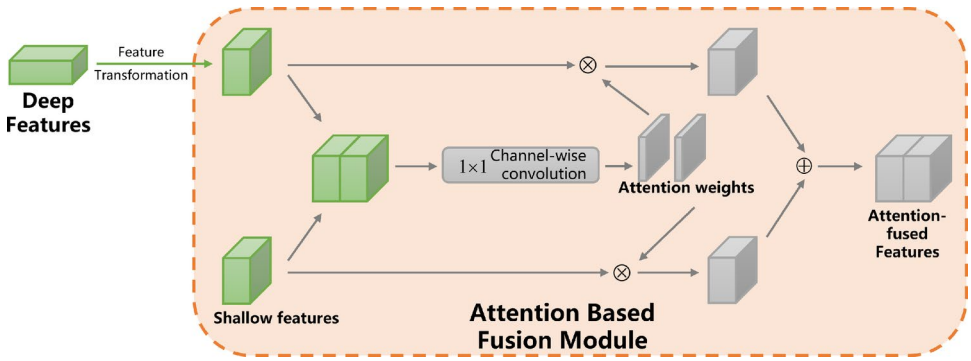

Fig. 3 The attention based fusion module

the spatial dimensions and multiplied element-wise with the concatenated feature map F_{cat}. The weighted feature map is then split into two parts corresponding to the original shallow and deep features:

$$F_{\text{weighted}} = F_{\text{cat}} \otimes s \tag{6}$$

$$F^{s}_{\text{out}} = F_{\text{weighted}}[:, 1 : C_s, :, :], \quad F^{d}_{\text{out}} = F_{\text{weighted}}[:, C_s + 1 : 2C_s, :, :] \tag{7}$$

where $\otimes$ denotes element-wise multiplication. The final output of the ABF module is obtained by summing these two weighted feature maps, which integrates both shallow edge/texture information and deep semantic information:

$$F_{\text{out}} = F^{s}_{\text{out}} + F^{d}_{\text{out}} \in \mathbb{R}^{C_s \times H_s \times W_s} \tag{8}$$

It is worth noting that the ABF module can generate different attention maps based on the input features, allowing for dynamic aggregation of the two feature maps. The advantage of this aggregation method lies in the adaptive and superior nature of attention-based fusion compared to direct concatenation and summation. This is because the two feature maps come from different stages of the network, and each contains diverse content information. As mentioned earlier, lower-level and higher-level features emphasize different types of information, and attention maps can effectively aggregate them in a more reasonable manner.

Furthermore, the ABF module employs channel-wise attention weighting [49] and does not introduce complex pixel-level attention. This is because, for features at different levels, their pixel-level significance and types of content emphasis differ. Applying the same pixel-level attention mechanism to both would blur the cross-level differences. From another perspective, the core of the RID method lies in the introduction of information from the teacher model, rather than the connection among student models. Hence, the ABF module does not require complex pixel-level attention mechanisms.

3.3 Hierarchical context loss module

The schematic diagram of the HCL module is illustrated in Fig. 4. The main function of this module is to observe the differences between the student attention-fusion features based on the ABF module and the corresponding teacher network layer features. It then calculates the distillation loss $\mathcal{L}_{KD}$.

In the HCL module, our work utilizes spatial pyramid pooling to perform pooling operations at different levels, effectively separating knowledge information into different levels of contextual information [50]. This method allows for better extraction of more information at different abstraction levels. Specifically, this paper first uses spatial pyramid pooling to extract knowledge from the student attention-fusion features based on the ABF module and the corresponding teacher network layer features at different levels. Then, it calculates the Euclidean distance $\mathcal{L}_2$ to measure the dissimilarity between the two types of features. Finally, it sums up all the distance losses to obtain the final distillation loss $\mathcal{L}_{KD}$.

It is worth noting that it is reasonable to use the Euclidean distance $\mathcal{L}_2$ as the loss function between pooled feature maps. This is because using $\mathcal{L}_2$ distance effectively transfers information between features at the same level.

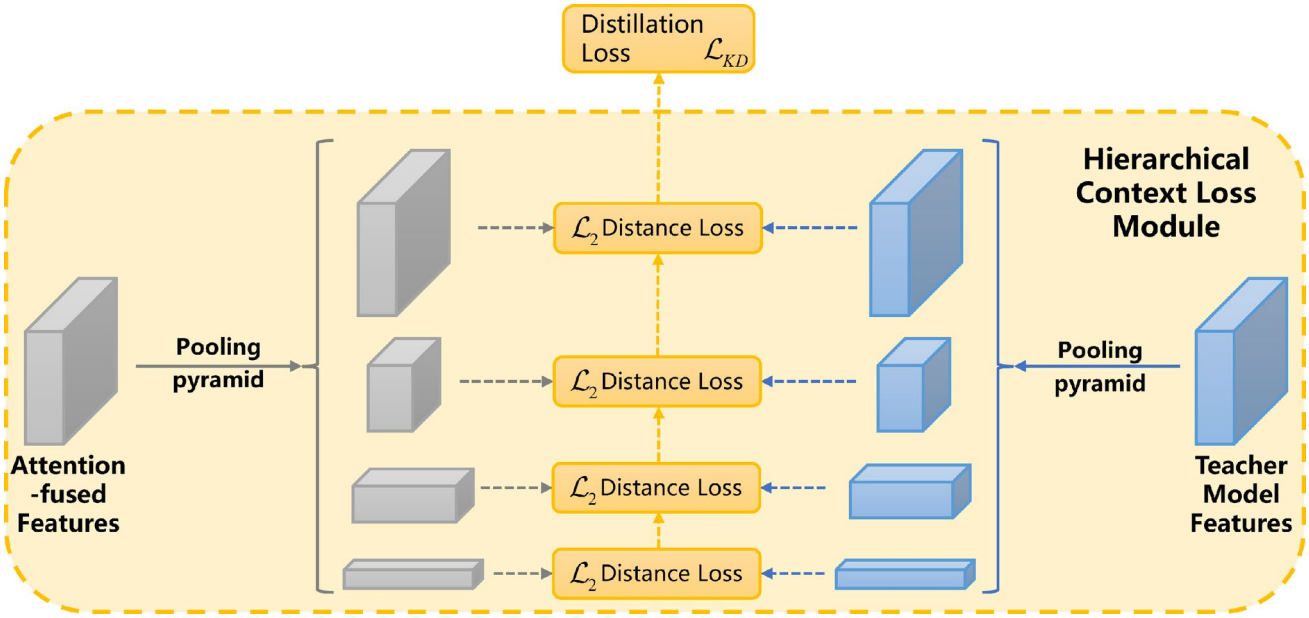

Fig. 4 The hierarchical context loss module

Additionally, the use of a pyramid structure aggregates information from different levels, allowing the student model to learn information across multiple dimensions that the teacher model has learned. Another reason for using this structure is that calculating the global $\mathcal{L}_2$ distance on individually pooled features is insufficient to transfer complex-level information.

3.4 Final loss fusion module

In the overall framework of the method, this paper introduces the form of the loss function, where the final loss function $\mathcal{L}$ is a weighted combination of the cross-entropy loss function $\mathcal{L}_{CE}$ generated by the student model's output and the distillation loss function $\mathcal{L}_{KD}$ obtained through the HCL module. The design of the distillation loss function $\mathcal{L}_{KD}$ is the core part of the final loss.

For the classical task of cross-level residual distillation, the loss function for the single-layer distillation of the student network can be expressed as follows:

$$\mathcal{L}_{SKD} = \sum_{j=1}^{i} \mathcal{L}_2(\mathrm{F}_s^i, \mathrm{F}_t^j) \tag{9}$$

where F_s^i and F_t^j represent the features extracted by the student and teacher models, respectively, at the *i*-th and *j*-th layers of the network. $\mathcal{L}_2$ denotes the function used to calculate the Euclidean distance. In other words, this paper utilizes the features of the teacher model at the previous *i*-th level to guide the student model at the *j*-th level, thereby achieving the transfer of residual information. Considering all student network layers, the loss function can be modified as:

$$\mathcal{L}_{MKD} = \sum_{i=1}^{n} \sum_{j=1}^{i} \mathcal{L}_2(\mathrm{F}_s^i, \mathrm{F}_t^j) \tag{10}$$

However, this formula introduces a large number of computation steps, resulting in a time complexity of $\mathcal{O}(n^2)$. Firstly, for a student network with *n* stages, the cumulative number of computations required is $k = n(n-1)/2$. Such a learning process can be simplified. Secondly, due to significant differences in information among different stage features, we employ the idea of feature fusion to better capture multiple levels of features.

When incorporating the feature fusion idea, the calculation method for the multi-step $\mathcal{L}_2$ distance loss will be modified accordingly. The ABF module starts the calculation from the deepest level of the student model and performs feature fusion with the previous layer. It then calculates the distillation loss $\mathcal{L}_{KD}$ with the teacher model. In the next calculation, these two layers of features will undergo the same feature fusion operation and distillation loss $\mathcal{L}_{KD}$ calculation with the adjacent shallower layer. In other words, the entire process is performed recursively.

Therefore, in this paper, the formula needs to be reversed and the summation needs to be approximated by introducing the ABF module to transform the sum of $\mathcal{L}_2$ distances of multi-level features into the sum of $\mathcal{L}_2$ distances of fused features. This can be expressed by:

$$\mathcal{L}'_{MKD} = \sum_{j=1}^{n} \sum_{i=j}^{n} \mathcal{L}_2(\mathrm{F}_s^i, \mathrm{F}_t^j) \tag{11}$$

$$\sum_{i=j}^{n} \mathcal{L}_2\left(\mathrm{F}_s^i, \mathrm{F}_t^j\right) \approx \mathcal{L}_2\left(\mathcal{U}\left(\mathrm{F}_s^j, \mathrm{F}_s^{j+1}, \cdots, \mathrm{F}_s^n\right), \mathrm{F}_t^j\right) \tag{12}$$

Here, the fusion operator $\mathcal{U}$ is explicitly defined as a hierarchical fusion process implemented by the ABF module: starting from the top-level feature F_s^n, we iteratively fuse it with the lower-level feature F_s^{n-1} using the ABF module (with F_s^{n-1} as the lower-level feature and F_s^n as the higher-level feature, undergoing resolution upsampling and channel alignment via a 1×1 convolution), then fuse the resulting output with F_s^{n-2}, and so on until F_s^j. The final output of $\mathcal{U}$ is a fused feature with the same spatial size and channel number as F_s^j, which integrates information from all student features from level *j* to *n*.

After replacing the initial accumulated features with the fused features for $\mathcal{L}_2$ distance calculation, this paper obtains the final residual distillation loss outputted by the HCL module. This can be expressed by:

$$\mathcal{L}_{KD} = \mathcal{L}_2\left(\mathrm{F}_s^n, \mathrm{F}_t^n\right) + \sum_{j=n-1}^{1} \mathcal{L}_2\left(\mathcal{U}\left(\mathrm{F}_s^j, \mathrm{F}_s^{j+1}, \cdots, \mathrm{F}_s^n\right), \mathrm{F}_t^j\right) \tag{13}$$

The separate calculation of the *n*-th layer feature in the equation above is necessary because this layer serves as the starting round for the recursive calculation and does not involve cross-level feature fusion. By designing the equation above, the time complexity for the cumulative $\mathcal{L}_2$ distance calculation is reduced from $\mathcal{O}(n^2)$ with double-layer accumulation to $\mathcal{O}(n)$ for a single layer. For a concrete example, consider a student network with $n = 56$ layers (e.g., ResNet-56). The original multi-layer distillation loss would require $n(n-1)/2 = 56 \times 55/2 = 1540$ pairwise feature comparisons per sample, resulting in prohibitive computational cost. In contrast, our revised formula with the ABF module only requires $n = 56$ fusion and comparison operations per sample, reducing the computation by more than 27 times. This achieves a more efficient and reliable loss calculation, which aligns with the attention-fusion idea of the ABF module.

Finally, this paper combines the cross-entropy loss function $\mathcal{L}_{CE}$ from the student model's output with the distillation loss function $\mathcal{L}_{KD}$ obtained through the HCL module by taking a weighted sum to obtain the designed final loss function. The weight coefficient is denoted as λ, and the final loss function is as shown below:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{KD} \tag{14}$$

With this, the paper has completed the design of the complete loss function for the RID method, and Fig. 5 provides a detailed description of the calculation method for the loss function at each distillation stage. It is worth noting that another hyperparameter in knowledge distillation is the temperature coefficient T, which represents the degree of softening of the model's predicted output labels. Similar to the design of the distillation process, a higher temperature indicates a purer and more concentrated extraction. In knowledge distillation, a higher temperature coefficient corresponds to "softer" predicted output labels generated by the teacher model, which have higher information entropy and extract more general knowledge.

The relationship between the probability predicted output q_i and the temperature coefficient T is given by:

$$q_i = \frac{\exp\left(z_i/T\right)}{\sum_j \exp\left(z_j/T\right)} \tag{15}$$

In the above discussion, the paper provides a detailed description of two crucial hyperparameters, λ and the temperature coefficient T, and also conducts comprehensive comparative experiments in the subsequent experimental section to demonstrate the impact of these two hyperparameters on the accuracy of few-shot sample source camera identification tasks. Our work optimizes model parameters within a reasonable range to achieve the best performance.

4 Experiments

To demonstrate the superiority of the proposed method, extensive experiments were conducted on three publicly available datasets using the RID method introduced in this paper. A comparison with existing methods was performed to validate the performance improvement of the proposed method. Thorough comparative experiments were conducted on important parameters, and detailed experiments were conducted in various complex scenarios.

4.1 Introduction to the datasets and experimental settings

The experiments in this study were conducted on three publicly available datasets for digital image source identification, including the classic camera dataset Dresden[51], as well as the mobile phone datasets VISION[52] and SOCRatES[53]. These datasets contain images captured from multiple cameras or mobile phone brands, models, and individuals, and the images cover various scenes to simulate diverse shooting conditions in real-life scenarios. As a result, these datasets have gained wide recognition among forensic investigators in the field of source camera identification. Next, this paper will provide detailed introductions and descriptions of these three datasets (Figs. 6 , 7 and 8).

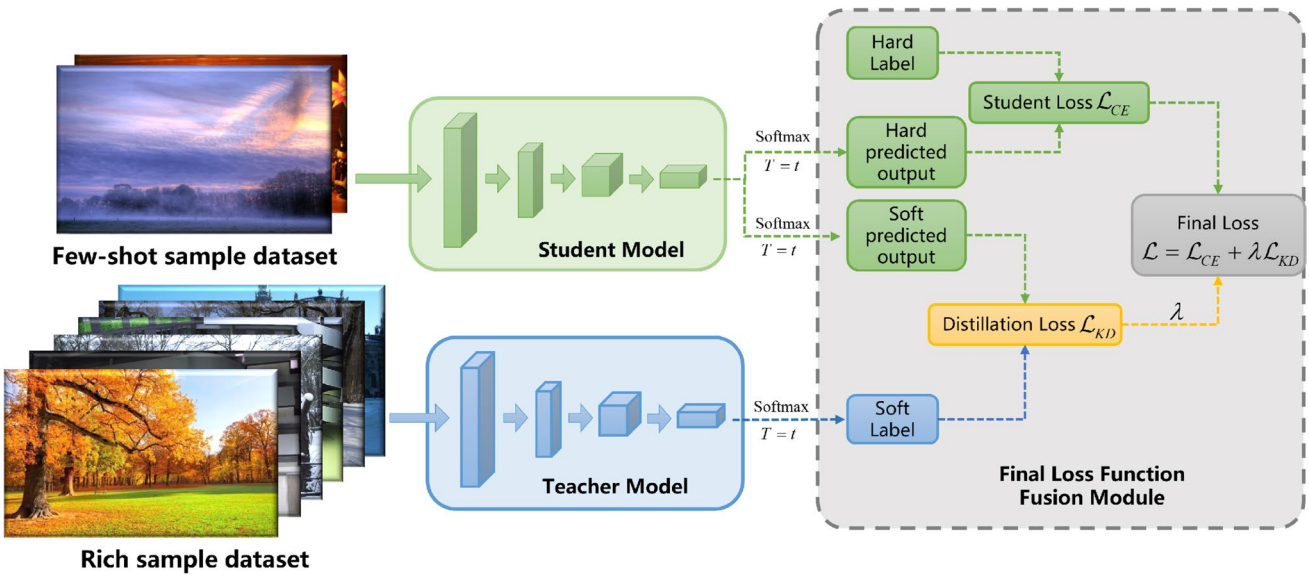

Fig. 5 The final loss function fusion module

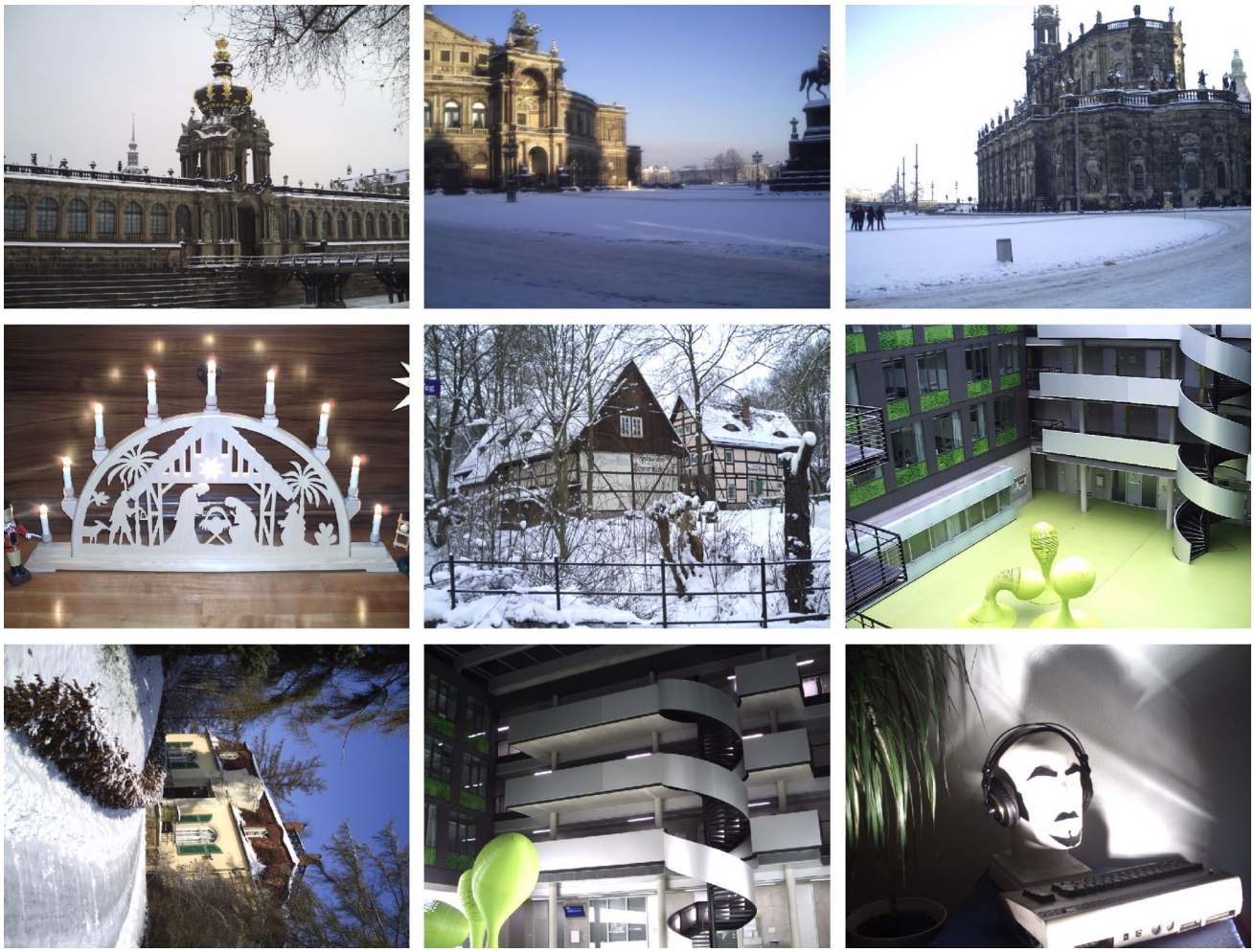

Fig. 6 Partial sample images of Dresden image database

As shown in Fig. 8, the SOCRatES dataset is also designed for mobile source forensics. It was publicly released in 2019 and consists of 103 mobile devices, including 15 different smartphone brands and 60 different models. The dataset contains over 9,700 images and 1,000 videos. Similar to the VISION dataset, SOCRatES provides a wide range of images and videos captured by various smartphones, offering valuable resources for research and development in the field of source identification and forensic analysis.

Among the three datasets, regarding the selection of the number of data categories, the paper selected 14 different camera models from various brands in the Dresden dataset, 11 different smartphone models from various brands in the VISION dataset, and 10 different smartphone models from 5 different brands in the SOCRatES dataset. It is worth noting that the selection of these image categories is consistent with other existing works [8, 9, 32, 34, 54] to ensure the fairness of cross-method comparisons.

Furthermore, in the context of a few-shot sample environment, we have also chosen to remain consistent with existing work by selecting 5-25 samples for each category of images, ensuring the rigor of the few-shot sample environment setup. The information, including the camera brand, model, quantity, resolution, and other details, is presented in Tables 2, 3, and 4. In these tables, "S/N" is an abbreviation for serial number, and ABBR stands for abbreviation. Considering the impact of the number of labeled samples on the accuracy of source identification, the paper compared the performance with varying numbers of labeled training samples in each category, including 5, 10, 15, 20, and 25 samples. In addition, the test datasets consisted of 2,791 images (in the Dresden dataset), 2,163 images (in the VISION dataset), and 1,927 images (in the SOCRatES dataset), with 130-438 unlabeled samples in each category.

In addition, to reduce the impact of uncertainties such as distortion on the experimental results while ensuring consistent resolution, this paper performed a random crop of a region within the center area 512×512 with a size of 224×224 in each round of the experiments. Furthermore, this paper used classification accuracy and time complexity

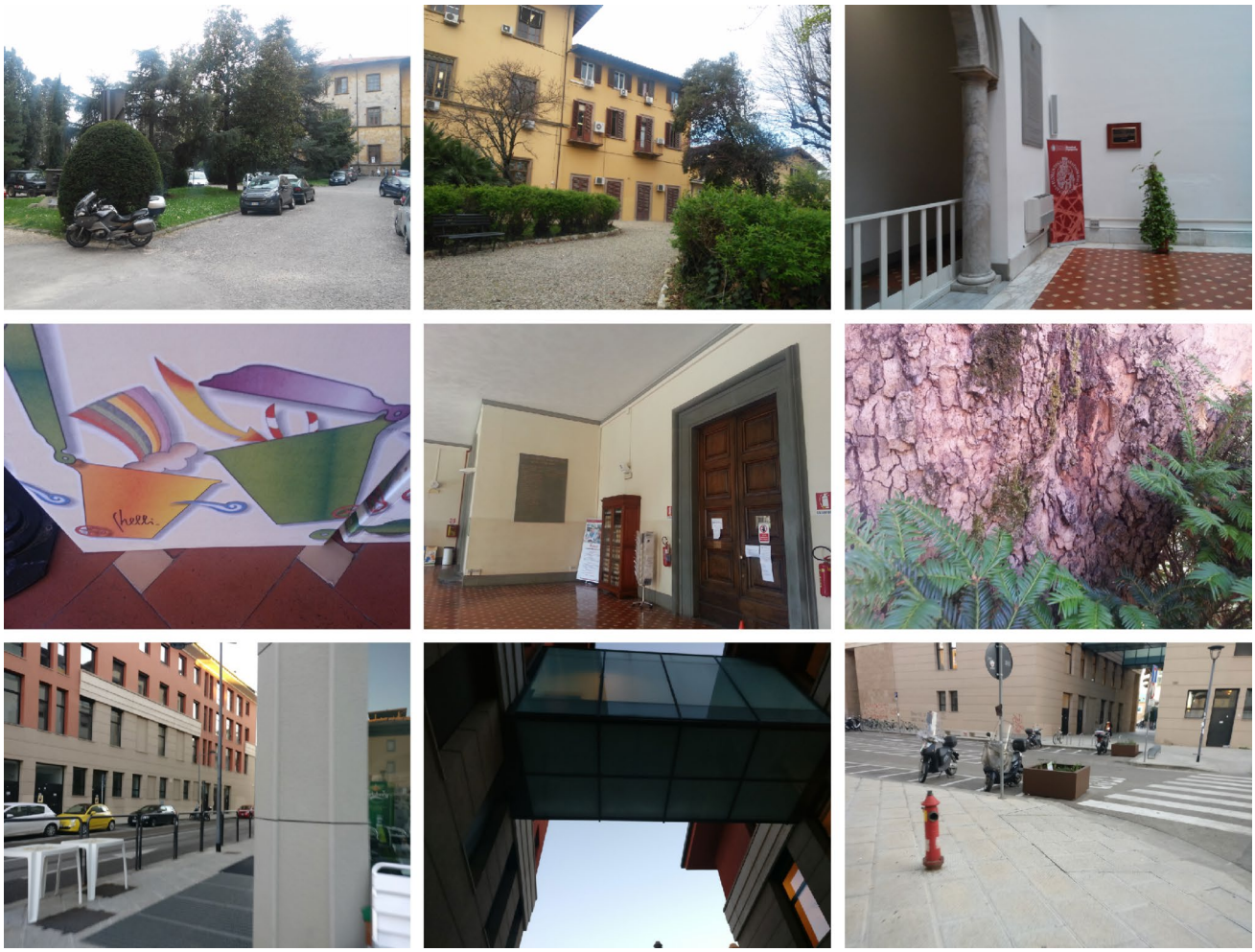

Fig. 7 Partial sample images of VISION database

as evaluation metrics to assess the performance of each method. These metrics were employed to measure the effectiveness and efficiency of the methods.

In order to test the performance of the proposed RID method, this study conducted fair comparisons with other existing methods under the same dataset and experimental settings. To simulate real-world application scenarios and enhance the robustness of the proposed method, we use different datasets in the training process of teacher model and student model. To be more specific, we use Dresden, VISION and Dresden datasets for teacher model training while VISION, Dresden and SOCRatES datasets are used in student model training in one-to-one correspondence.

4.2 Comparative experiment of classification accuracy

As is shown in Table 5, through the comparative experiments with other existing methods, the results demonstrate that the proposed RID method outperforms the existing methods under the same experimental configurations (i.e., 14 categories for Dresden dataset, 11 categories for VISION dataset, and 10 categories for SOCRatES dataset, each with 10 few-shot samples per category). Therefore, the method based on the supplementation of model knowledge information proves to be one of the most effective approaches in addressing the issue of few-shot sample problems, and it shows superior performance compared to other existing methods across the three datasets.

However, the proposed RID methods based on deep learning and knowledge distillation in this paper provide superior source identification performance at the expense of longer model training time. Therefore, deep learning methods achieve better performance improvements at the cost of increased training time, which is a noticeable difference compared to traditional machine learning methods. The comparison of model training time for different methods on the Dresden dataset is shown in Table 6.

The aforementioned experimental results once again demonstrate that the RID method, despite consuming more

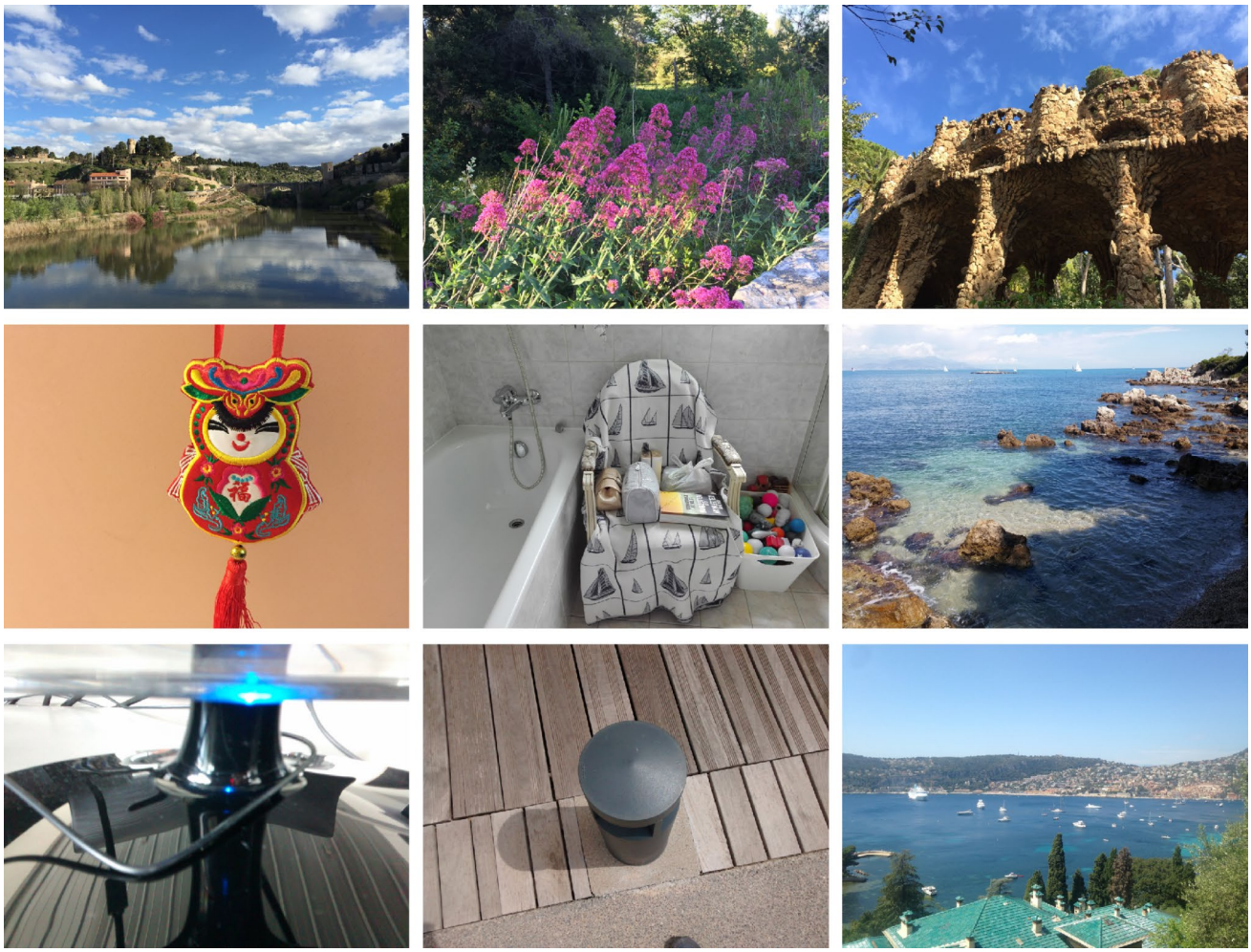

Fig. 8 Partial sample images of SOCRatES database

Table 2 Sample information for Dresden dataset in the experiment

S/N	Camera model	Quantity	Image size
1	Agfa_DC-504	167	4032 × 3024
2	Canon_PowerShotA640	188	3648 × 2736
3	Casio_EX-Z150	181	3264 × 2448
4	FujiFilm_FinePixJ50	209	3264 × 2448
5	Kodak_M1063	463	3664 × 2748
6	Nikon_CoolPixS710	186	4352 × 3264
7	Olympus_mju_1050SW	202	3648 × 2736
8	Panasonic_DMC-FZ50	262	3648 × 2736
9	Pentax_OptioA40	169	4000 × 3000
10	Praktica_DCZ5.9	209	2560 × 1920
11	Ricoh_GX100	192	3648 × 2736
12	Rollei_RCP-7325XS	198	3072 × 2304
13	Samsung_L74wide	231	3072 × 2304
14	Sony_DSC-H50	284	3456 × 2592

Table 3 Sample information for VISION dataset in the experiment

S/N	Camera model	ABBR	Quantity	Image size
1	Apple_iPad2	A1	171	960 × 720
2	Asus_Zenfone2Laser	A2	209	3264 × 1836
3	Huawei_Ascend	H1	155	3264 × 2448
4	Lenovo_P70A	L1	216	4780 × 2704
5	LG_D290	L2	227	3264 × 2448
6	Microsoft_Lumia640LTE	M1	187	3264 × 1840
7	OnePlus_A3000	O1	287	4640 × 3480
8	Samsung_GalaxyS3	S1	207	3264 × 2448
9	Sony_XperiaZ1Compact	S2	215	5248 × 3936
10	Wiko_Ridge4G	W1	253	3264 × 2448
11	Xiaomi_RedmiNote3	X1	311	4608 × 2592

model training time, outperforms current state-of-the-art methods on the three datasets and possesses distinct advantages in various scenarios. We believe there are several reasons why the RID method requires more model training time. Firstly, deep learning frameworks inherently involve continuous convolution operations on samples to extract features and require gradient updates, which inherently consume time. Additionally, during the transfer and distillation of knowledge across datasets, further comparison

Table 4 Sample information for SOCRatES dataset in the experiment

S/N	Camera model	Quantity	Image size
1	Apple iPhone 5s	180	3264×2448
2	Apple iPhone 6	200	3264×2448
3	Apple iPhone 6s	180	4032×3024
4	Apple iPhone 7	200	4032×3024
5	Asus Zenfone 2	193	4096×3072
6	LG G3	190	4160×3120
7	Motorola Moto G	180	2592×1944
8	Samsung Galaxy A3	200	4128×2322
9	Samsung Galaxy S5	180	5312×2988
10	Samsung Galaxy S7 Edge	224	4032×3024

Table 5 Comparison experimental results with other existing approaches under the same conditions, 10-shot samples per class (%). Bold values indicate the best performance in each column.

Methods \ Datasets	Dresden	VISION	SOCRatES
EP [55]	73.84	79.94	63.91
MTD-EM [9]	75.16	80.49	64.84
Deep Siamese Network [34]	85.30	75.20	\
Multi-DS [8]	86.08	85.56	67.00
PCEP [54]	87.06	84.84	75.94
RID(Ours)	**89.07**	**85.64**	**79.08**
ΔAcc	**2.01**	**0.08**	**3.14**

Table 6 Comparison of time complexity of different approaches in Dresden database (minutes)

Methods \ Datasets	5	10	15	20	25
EP [55]	1.34	2.39	3.46	4.68	5.87
MTD-EM [9]	0.43	1.31	2.63	4.43	6.63
Multi-DS [8]	1.07	2.16	3.22	4.31	5.42
Multi-PCEP [54]	2.26	3.65	5.96	8.72	11.26
RID(Ours)	136.15	143.56	150.96	158.37	165.78

of residual information between the teacher and student models is needed to update the student model, which also requires a certain amount of time. In our experiments, the method using knowledge distillation after training showed a significant performance improvement in few-shot sample situations compared to direct training and outperformed existing methods. Therefore, our approach of using residual information for knowledge distillation is justified.

It is important to note that the inference time of RID is comparable to that of existing deep learning-based methods (less than 10 ms per 224×224 image on an RTX 4090 GPU), which does not affect its practical deployment in real-world forensic applications. The longer training time is a one-time cost that can be offset by several optimization strategies. First, lightweight teacher models can be adopted, replacing the current ResNet-56 teacher model with lightweight architectures such as MobileNetV2 or ShuffleNetV2, which can reduce training time by 40%–60% while maintaining most of the performance. Second, distillation pruning can be applied, where structured pruning is performed on the teacher model before distillation to remove redundant parameters and layers, reducing the computational cost of feature extraction and residual comparison. Third, quantization acceleration can be utilized, employing 16-bit floating-point (FP16) or 8-bit integer (INT8) quantization during both training and inference to significantly speed up matrix operations without noticeable accuracy loss. Finally, feature pre-caching can be implemented, where the features of the teacher model on the training dataset are pre-extracted and cached to avoid repeated feature extraction during each training epoch, which can reduce training time by approximately 30%.

4.3 Hyperparameters analysis experiments for RID

To investigate the impact of important hyperparameters in the RID method, we conduct comprehensive sensitivity analysis experiments on the weight coefficient λ (balancing prediction loss and distillation loss) and the temperature coefficient T. The experimental results are shown in Fig. 9. The weight coefficient λ controls the relative importance of the distillation loss $\mathcal{L}_{KD}$ compared to the student's own cross-entropy loss $\mathcal{L}_{CE}$. A smaller λ weakens the supervisory effect of the teacher model, while a larger λ may overwhelm the student's task-specific learning.

We evaluate the performance of RID with λ ranging from 1 to 512 across all three datasets and all few-shot settings (5–25 samples per class). The optimal λ value is consistently 64 across all three datasets, achieving the highest accuracy of 89.07%, 85.64%, and 79.08% on Dresden, VISION, and SOCRatES, respectively. Furthermore, RID exhibits strong stability across a wide range of λ values: when λ varies from 32 to 128, the accuracy degradation is less than 1.0% on all datasets, demonstrating the robustness of our framework to hyperparameter changes. Notably, the sensitivity of λ increases as the number of training samples decreases. In the 5-shot setting, the optimal λ shifts slightly to 128 as stronger teacher supervision is required, while in the 25-shot setting, $\lambda = 32$ achieves comparable performance to $\lambda = 64$. This observation is consistent with the intuition that fewer training samples require more external knowledge from the teacher model.

The temperature coefficient T controls the softening degree of the teacher's predicted output labels. A higher T produces softer labels with higher information entropy, which helps extract more general knowledge.

The experimental results show that T has a relatively small impact on classification accuracy compared to λ. The optimal value is consistently $T = 2$ across all three datasets and all few-shot settings. When T varies from 1 to 8, the maximum accuracy degradation is only 0.87%, further

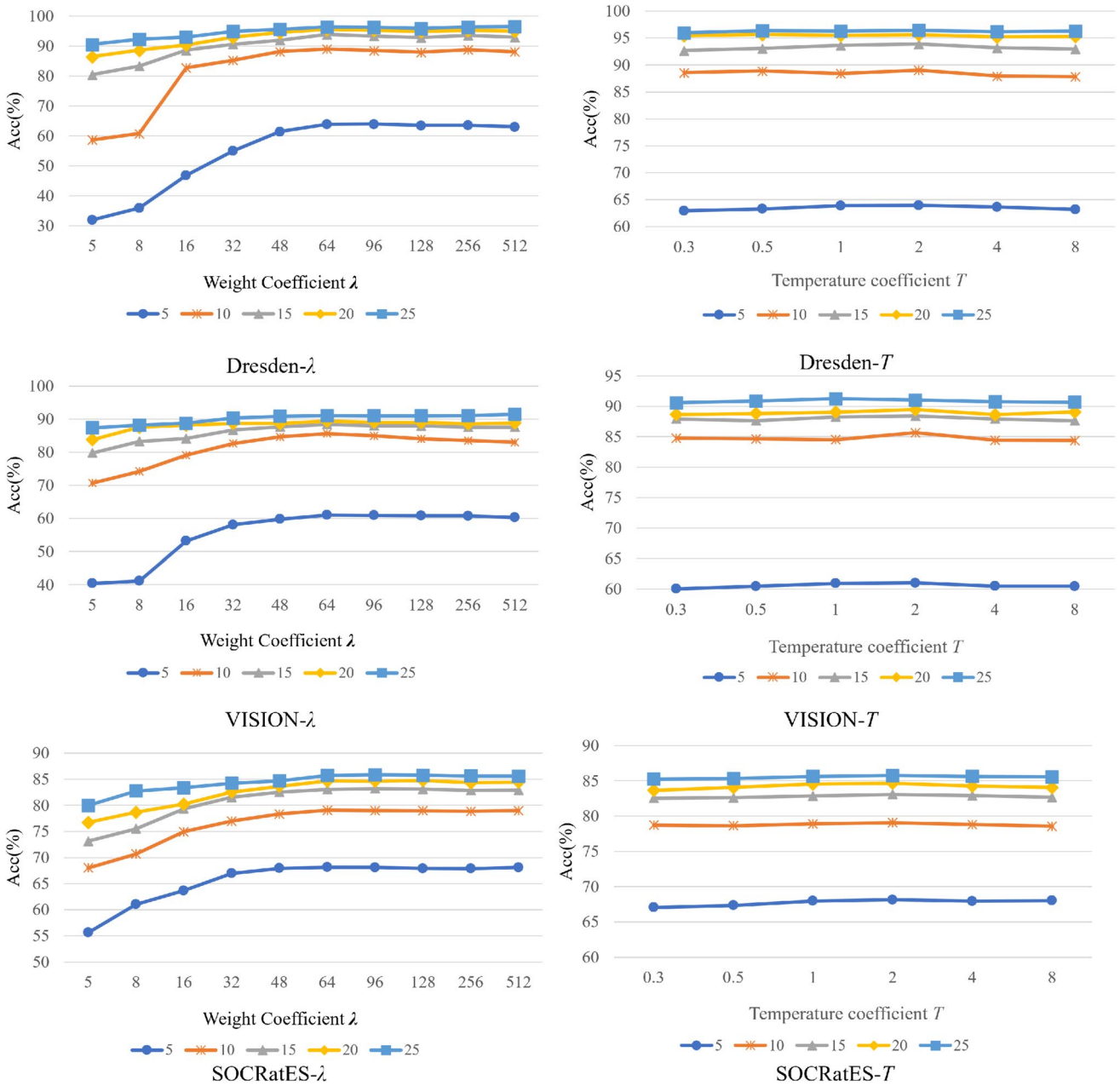

Fig. 9 Comparison experiment of distillation loss weighting coefficient and temperature coefficient in multiple databases

confirming the stability of the RID framework. This minor impact can be attributed to the cross-dataset training paradigm, where the class indices of the teacher and student datasets do not directly correspond, reducing the effectiveness of soft label distillation.

Overall, the parameter analysis experiments demonstrate that the RID framework is robust to hyperparameter variations, with stable performance across a wide range of λ and T values. The optimal hyperparameters ($\lambda = 64, T = 2$) are determined through comprehensive cross-dataset and cross-shot evaluations.

4.4 Cross-network distillation experiments for RID

In this experiment, the teacher model is set as ShuffleNetV2. In addition to being the same as the teacher model, the student model includes ResNet networks with different numbers of layers, such as ResNet20, ResNet32, ResNet56, and ResNet110. The results are shown in Fig. 10.

According to the results of the cross-network distillation experiment, the use of knowledge transfer strategies significantly improves the accuracy in the case of few-shot sample sizes. In the scenario where each class has only 5 few-shot

samples, this method achieves performance improvements of up to 34.28%, 30.20%, and 40.43% on the three datasets, respectively. Furthermore, it is observed that the performance of cross-network distillation is slightly lower than that of within-network distillation. This can be attributed to the fact that the teacher and student models have different datasets, and the additional replacement operations introduce larger knowledge disparities between them, leading to a decrease in accuracy. However, this cross-network distillation method still outperforms the strategy without distillation.

Additionally, as the depth of the student network increases, the learning capacity of the student model gradually improves. However, there is a trade-off between underfitting and overfitting. The highest performance is achieved at ResNet56, indicating that an appropriate depth of the student network is also an important factor affecting accuracy.

In addition, this study also generated a confusion matrix on the VISION dataset when each class contained 25 samples, as shown in Fig. 11. The final accuracy of this confusion matrix is 91.03%, corresponding to the results of the RID method on the VISION dataset with 25 samples as shown in Fig. 10. Specifically, the overall identification accuracy reaches 91.03% under the 25-shot setting, demonstrating the strong generalization ability of the RID method. Most smartphone models achieve an identification accuracy above 85%. The vast majority of misclassifications occur between models of the same brand. For example, Sony Xperia Z1 Compact (S2) has a 10.00% misclassification rate to Samsung Galaxy S3 (S1), and Huawei Ascend (H1) has a 9.90% misclassification rate to Xiaomi Redmi Note 3 (X1). This is because smartphones from the same brand often use similar image signal processor (ISP) algorithms and sensor configurations, resulting in highly similar camera fingerprints. Models with unique hardware configurations achieve nearly 100% identification accuracy, such as Asus Zenfone 2 Laser (A2) (99.30%) and Microsoft Lumia 640 LTE (M1) (99.45%). This further confirms that the RID method can effectively capture device-specific hardware traces. No cross-brand misclassifications are observed in the confusion matrix, indicating that the proposed method can reliably distinguish between different smartphone brands even in few-shot scenarios.

4.5 Cross-task distillation experiments for RID

The purpose of this experiment is to examine the impact of knowledge transfer in both the same-task and cross-task scenarios. The CIFAR10 dataset is a classic image classification dataset with 60,000 images divided into 10 sample classes. The training and testing sets are split in a 5:1 ratio, and all images have the same resolution. Therefore, to perform cross-domain knowledge transfer, the teacher model is trained on the CIFAR10 dataset and then applied to the source camera identification task to obtain the comparison results. To match the feature dimension and task objective between the CIFAR10 general classification task and the SCI forensic task, we perform the following adaptation on the CIFAR10-trained teacher model:

We remove the final 10-class classification layer of the original CIFAR10 teacher model, which is designed for general object classification and not applicable to camera fingerprint extraction. We add a new linear projection layer with input dimension equal to the output dimension of the teacher's last convolutional layer and output dimension equal to the feature dimension of the student model (consistent with the same-task distillation setting). During the entire cross-task distillation process, all parameters of the CIFAR10 teacher model are frozen, and only the newly added linear projection layer is fine-tuned for 10 epochs on the SCI training set to align the feature distribution. This prevents catastrophic forgetting of the general visual knowledge learned by the teacher model on CIFAR10. The experimental results are shown in Fig. 12, where "V" represents the VISION dataset (the tested smartphone model dataset), "C" represents the CIFAR10 dataset (the image content classification dataset), "D" represents the Dresden dataset (the camera model dataset), "S" represents the SOCRatES dataset (the trained smartphone model dataset), and the arrow "→" denotes the distillation process.

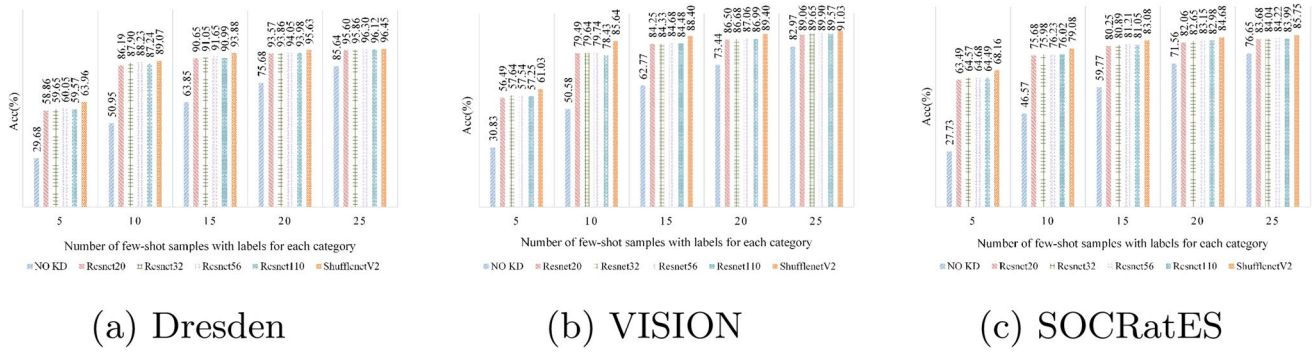

Fig. 10 Cross-Network Distillation Experiment in multiple databases

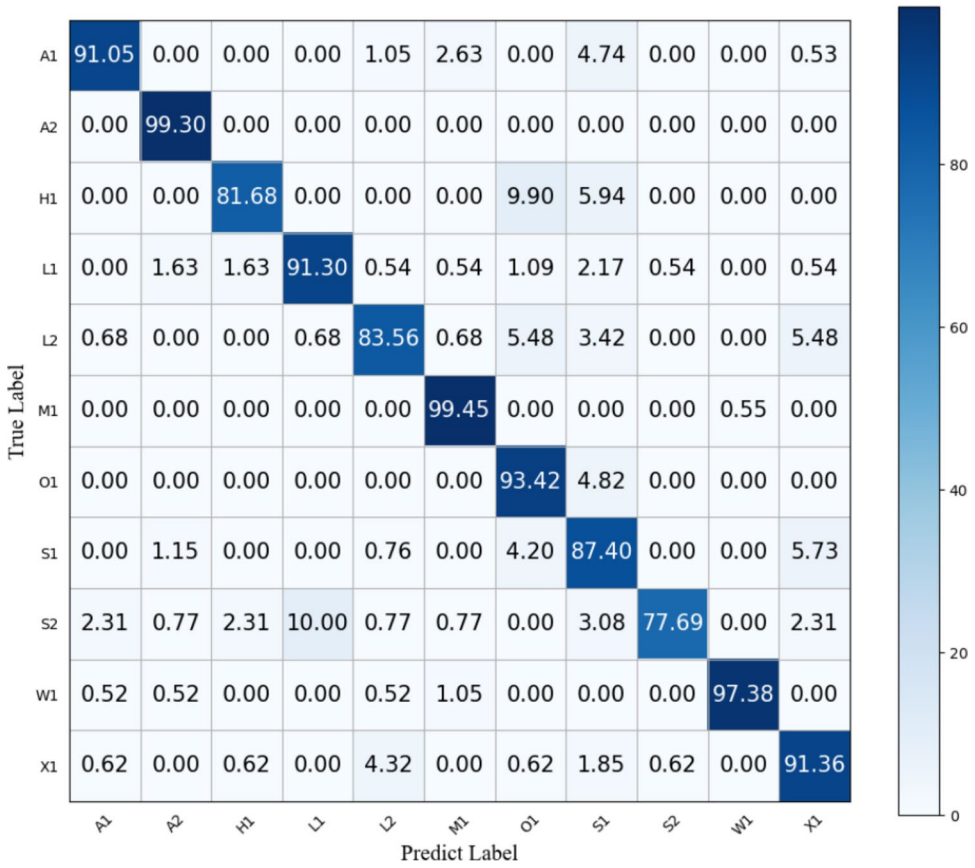

Fig. 11 Confusion matrix (VISION dataset) with 25 samples for each class

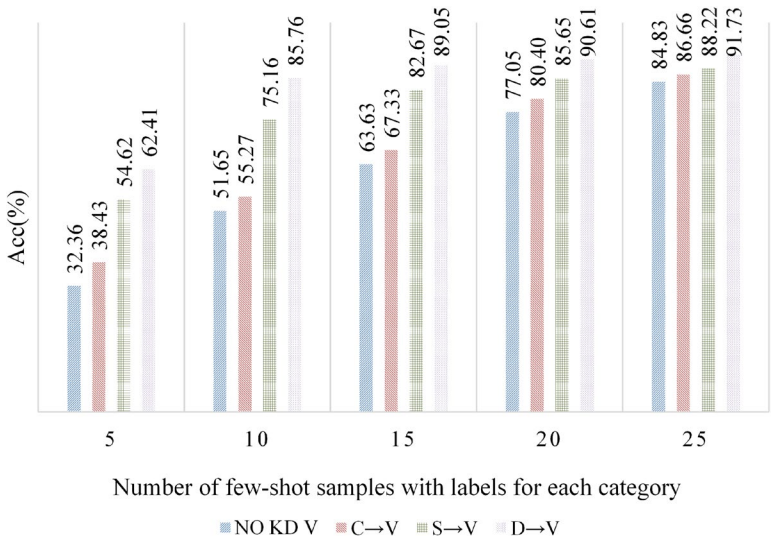

Fig. 12 Cross-task distillation experiment

The experiment consists of four scenarios for comparative analysis, including: No distillation used($NO\ KD\ V$), Cross-Domain Knowledge Transfer($C \rightarrow V$), Same-Domain Knowledge Transfer($S \rightarrow V, D \rightarrow V$). By comparing the results across these four scenarios, the paper aims to analyze the effectiveness of different knowledge transfer approaches and understand their impact on the performance in complex settings.

According to the results of the cross-task distillation experiment, although the performance improvement using different types of task datasets for knowledge transfer is not significant, it is still effective compared to the no distillation learning method. In the scenario where each category has only 5 few-shot samples, the RID method can still achieve a 6.07% increase in accuracy. This indicates that the proposed teacher knowledge transfer method has been successfully

applied to the student model. In the case of same-domain tasks, using the SOCRatES and Dresden datasets as teacher model datasets for distilling the student model based on the VISION dataset results in accuracy improvements of 22.26% and 30.05% respectively. Therefore, distillation among same-domain tasks shows more significant performance improvements.

4.6 Ablation study on cross-level residual distillation

To comprehensively evaluate the contribution of each component in the proposed RID framework, we conduct systematic ablation experiments on the VISION dataset. The RID framework consists of three core modules: CRD, ABF, and HCL. We progressively evaluate the performance by enabling different combinations of these modules, starting from the baseline model (no distillation) to the full RID framework.

Effect of individual modules We first evaluate the contribution of each module independently. As shown in Table 7, the baseline model without any distillation achieves an average accuracy of 77.71%. When only the ABF module is enabled, the performance improves to 80.20%, demonstrating that attention-based feature fusion can provide modest gains even without cross-level residual information. Similarly, enabling only the HCL module yields 79.35% average accuracy, indicating that hierarchical context loss contributes to better knowledge transfer. However, both ABF and HCL show limited improvements when used independently, as they lack the cross-level residual features that are essential for effective knowledge distillation in few-shot scenarios.

Effect of CRD as the core module In contrast, enabling only the CRD module results in a substantial performance boost to 85.19% average accuracy, representing a 7.48% improvement over the baseline. This significant gain confirms that cross-level residual distillation is the core component of the RID framework, as it enables the student model to effectively capture hierarchical camera fingerprint knowledge from the teacher model. The CRD module alone outperforms both ABF-only and HCL-only configurations by large margins (4.99% and 5.84% respectively), demonstrating its critical role in addressing the few-shot SCI challenge.

Effect of module combinations When combining CRD with either ABF or HCL, we observe further performance improvements. The CRD+ABF combination achieves 87.22% average accuracy, while CRD+HCL reaches 86.92%. Both combinations show consistent gains over CRD-only across all few-shot settings, with improvements of 2.03% and 1.73% respectively. This indicates that ABF and HCL effectively complement the CRD module by adaptively aggregating cross-level features and refining the distillation process through hierarchical context modeling.

Effect of the full RID framework Finally, when all three modules are enabled together, the full RID framework achieves the best performance of 88.62% average accuracy. The synergistic effect of combining all three modules results in a 3.43% improvement over CRD-only and a 10.91% improvement over the baseline, confirming that each component contributes meaningfully to the overall performance.

Overall, the ablation study clearly demonstrates that CRD is the core module providing the majority of performance gains; ABF and HCL are effective auxiliary modules that enhance the distillation process when combined with CRD; and the full RID framework achieves the best performance through the synergistic integration of all three components.

4.7 Complex class experiment for RID

This work investigates the impact of the number of categories on the performance of the RID method by evaluating its performance. To avoid the issue of mismatched categories, the study selected 27 categories in the Dresden and VISION datasets, and 20 categories in the SOCRatES dataset. Knowledge distillation was performed using both the camera model dataset and the smartphone model dataset. The experimental results are shown in Fig. 13.

The analysis of the image source identification accuracy in complex categories reveals that the RID method performs

Table 7 Modular ablation study results on VISION dataset (%). Bold values indicate the best performance under each shot setting and the highest average accuracy across all ablation configurations

CRD	ABF	HCL	10-shot	15-shot	20-shot	25-shot	Avg. Acc.
✗	✗	✗	70.45	76.82	80.39	83.17	77.71
✗	✓	✗	73.26	79.45	82.71	85.38	80.20
✗	✗	✓	72.18	78.53	81.92	84.76	79.35
✓	✗	✗	81.37	84.62	86.71	88.05	85.19
✓	✓	✗	83.92	86.75	88.53	89.68	87.22
✓	✗	✓	83.58	86.41	88.26	89.42	86.92
✓	✓	✓	**85.64**	**88.40**	**89.40**	**91.03**	**88.62**

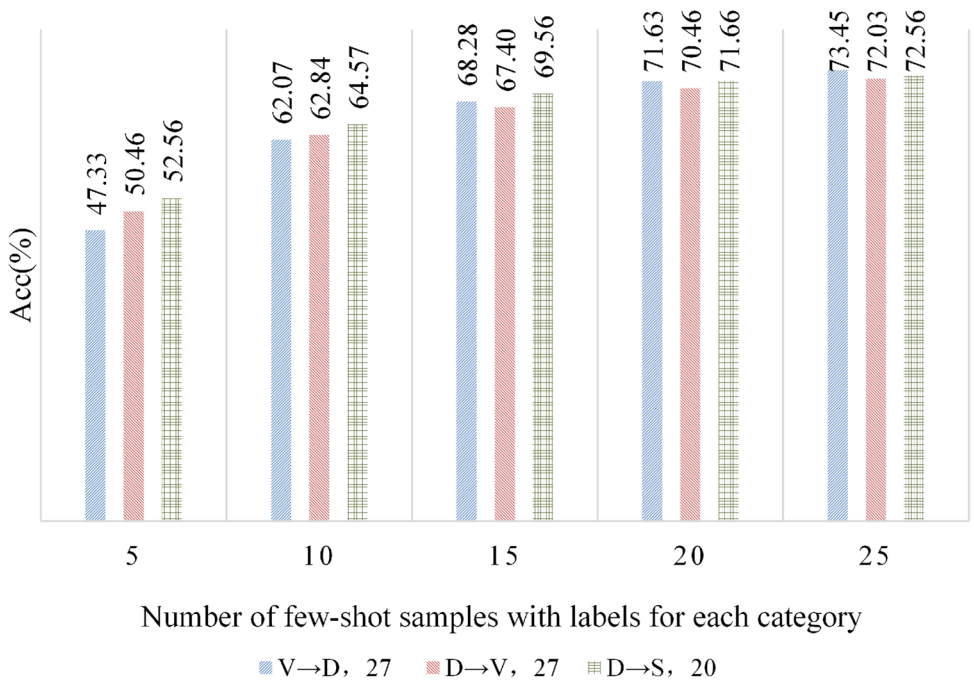

Fig. 13 Accuracy of source camera identification in complex class situations

well even in more complex category scenarios. For instance, when the number of categories increases to at least 20, with only 5 few-shot samples per category, the image source identification accuracy of the RID method on the three datasets can still be maintained at around 50%.

5 Conclusion

This paper proposes a few-shot sample source camera identification method called Residual Information Distillation (RID) to address the practical challenge of source camera identification in few-shot sample scenarios. The RID method employs a knowledge transfer method within the context of cross-dataset knowledge transfer. It first trains a teacher model on other publicly available large-scale datasets and then utilizes a cross-layer residual distillation structure to dynamically supervise and distill the information of each layer and deeper levels of the student model, enabling knowledge transfer and distillation. Extensive experimental results demonstrate the superior performance of the knowledge transfer approach in few-shot sample source camera identification. By leveraging the knowledge from the teacher model, the RID method outperforms state-of-the-art methods in terms of performance. Despite its effectiveness, RID still has limitations: it relies on large-sample datasets for pre-training the teacher model and has longer training time than traditional methods, and has not been verified on forged or compressed images. In future work, we will optimize efficiency by lightweight teacher models and distillation pruning, and extend RID to forged images, compressed images, and open-set identification to enhance its practicality in real forensic scenarios.

Acknowledgements This work is supported by the Application Fundamental Research Project of Liaoning Province (No. 2025JH2/101330123) and the National Natural Science Foundation of China (No. 62372452).

Author Contributions Huimin Liu: Conceptualization, Methodology, Software, Formal analysis, Validation, Visualization, Writing - original draft, Writing - review & editing. Jiayao Hou: Conceptualization, Methodology, Formal analysis, Writing - review & editing. Jiaqi Chi: Conceptualization, Methodology, Writing - review & editing. Bo Wang: Conceptualization, Methodology, Writing - review & editing, Supervision, Resources.

Data Availability The datasets generated and/or analysed during the current study are available in the GitHub repository, https://github.com/DLUTAIS/RID.git.

Code Availability The source code of the proposed method is available at https://github.com/DLUTAIS/RID.git.

Declarations

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.

References

1. Dadkhah S, Köppen M, Jalab HA, Sadeghi S, Manaf AA, Uliyan DM (2017) Electromagnetismlike mechanism descriptor with fourier transform for a passive copy-move forgery detection in digital image forensics. ICPRAM, pp 612–619

2. 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 (ICSAI). IEEE, pp 1–5
3. Xu B, Wang X, Zhou X, Xi J, Wang S (2016) Source camera identification from image texture features. Neurocomputing 207:131–140
4. Roy A, Chakraborty RS, Sameer VU, Naskar R (2017) Camera source identification using discrete cosine transform residue features and ensemble classifier. CVPR Workshops, pp 1848–1854
5. Wang B, Zhong K, Li M (2019) Ensemble classifier based source camera identification using fusion features. Multimed Tools Appl 78(7):8397–8422
6. Ma N, Zhang X, Zheng H-T, Sun J (2018) Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 116–131
7. Elharrouss O, Akbari Y, Almadeed N, Al-Maadeed S, Khelifi F, Bouridane A (2025) Pdc-vit: Source camera identification using pixel difference convolution and vision transformer. Neural Comput Appl 37(9):6933–6949
8. Wang B, Hou J, Ma Y, Wang F, Wei F (2022) Multi-ds strategy for source camera identification in few-shot sample data sets. Secur Commun Netw 2022
9. 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. Intern J Digit Crime Forens (IJDCF) 13(3):34–46
10. Lukas J, Fridrich J, Goljan M (2006) Digital camera identification from sensor pattern noise. IEEE Trans Inf Forensics Secur 1(2):205–214
11. Zhao Y, Zheng N, Qiao T, Xu M (2019) Source camera identification via low dimensional prnu features. Multimed Tools Appl 78(7):8247–8269
12. Bruni V, Tartaglione M, Vitulano D (2021) Coherence of prnu weighted estimations for improved source camera identification. Multimed Tools Appl 1–24
13. 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. Springer, pp 87–101
14. Freire-Obregón D, Narducci F, Barra S, Castrillon-Santana M (2019) Deep learning for source camera identification on mobile devices. Pattern Recogn Lett 126:86–91
15. Bennabhaktula GS, Alegre E, Karastoyanova D, Azzopardi G (2022) Camera model identification based on forensic traces extracted from homogeneous patches. Expert Syst Appl 206:117769
16. Huan S, Liu Y, Yang Y, Law N-FB (2024) Camera model identification based on dual-path enhanced convnext network and patches selected by uniform local binary pattern. Expert Syst Appl 241:122501
17. Zheng H, You C, Wang T, Ju J, Li X (2024) Source camera identification based on an adaptive dual-branch fusion residual network. Multimed Tools Appl 83(6):18479–18495
18. 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. IEEE, pp 69
19. Swaminathan A, Wu M, Liu KR (2007) Nonintrusive component forensics of visual sensors using output images. IEEE Trans Inf Forensics Secur 2(1):91–106
20. Ferrara P, Bianchi T, De Rosa A, Piva A (2012) Image forgery localization via fine-grained analysis of cfa artifacts. IEEE Trans Inf Forensics Secur 7(5):1566–1577
21. 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 (ICIP). IEEE, pp 4778–4782
22. 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 Process Lett 27:466–470
23. 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 (ICASSP). IEEE, pp 1850–1854
24. Villalba LJG, Orozco ALS, López RR, Castro JH (2016) Identification of smartphone brand and model via forensic video analysis. Expert Syst Appl 55:59–69
25. Sychandran C, Shreelekshmi R (2024) Sccrnet: a framework for source camera identification on digital images. Neural Comput Appl 36(3):1167–1179
26. López RR, Orozco ALS, Villalba LJG (2021) Compression effects and scene details on the source camera identification of digital videos. Expert Syst Appl 170:114515
27. Han Z, Yang Y, Zhang J, Li Y, Liu Y, Law N-FB (2025) A contrastive learning-based heterogeneous dual-branch network for source camera identification. Neurocomputing 130406
28. Vinyals O, Blundell C, Lillicrap T, Kavukcuoglu K, Wierstra D (2016) Matching networks for one shot learning
29. Snell J, Swersky K, Zemel RS (2017) Prototypical networks for few-shot learning
30. Finn C, Abbeel P, Levine S (2017) Model-agnostic meta-learning for fast adaptation of deep networks
31. Zhang H, Cisse M, Dauphin YN, Lopez-Paz D (2017) mixup: Beyond empirical risk minimization
32. 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. Springer, pp 18–27
33. Liang X, Zhang Y, Zhang J (2021) Attention multisource fusion-based deep few-shot learning for hyperspectral image classification. IEEE J Selec Topics Appl Earth Observ Remote Sens 14:8773–8788
34. Sameer VU, Naskar R (2020) Deep siamese network for limited labels classification in source camera identification. Multimed Tools Appl 79(37–38):28079–28104
35. Wang B, Wu S, Wei F, Wang Y, Hou J, Sui X (2022) Virtual sample generation for few-shot source camera identification. J Inf Secur Appl 66:103153
36. Wang B, Hou J, Wei F, Yu F, Zheng W (2023) Mdm-cps: A few-shot sample approach for source camera identification. Expert Syst Appl 229:120315
37. Karabiyik MA, Turkoglu B, Asuroglu T (2025) Optimizing artificial neural networks using mountain gazelle optimizer. Access IEEE 13(000):50464–50479
38. Uymaz O, Turkoglu B, Kaya E, Asuroglu T (2024) A novel diversity guided galactic swarm optimization with feedback mechanism. IEEE Access 12:108154–108175
39. Karaaslan M, Turkoglu B, Kaya E, Asuroglu T (2024) Voice analysis in dogs with deep learning: Development of a fully automatic voice analysis system for bioacoustics studies. Sensors 24(24):7978
40. Turkoglu B, Karabiyik MA, Asuroglu T Henos: A henon map-based chaotic oversampling strategy for imbalanced data classification. IEEE Access 13
41. Hinton G, Vinyals O, Dean J (2015) Distilling the knowledge in a neural network. Stat 1050:9
42. Romero A, Ballas N, Kahou SE, Chassang A, Gatta C, Bengio Y (2015) Fitnets: Hints for thin deep nets. Intern Conf Learn Represent
43. Zagoruyko S, Komodakis N (2017) Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. Intern Conf Learn Represent

44. Park W, Kim D, Lu Y, Cho M (2019) Relational knowledge distillation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3967–3976
45. Chen P, Liu S, Zhao H, Jia J (2021) Distilling knowledge via knowledge review. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 5008–5017
46. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. IEEE
47. Shi X, Lv F, Seng D, Zhang J, Xing B (2020) Visualizing and understanding graph convolutional network. Multimed Tools Appl 3:1–21
48. Cozzolino D, Verdoliva L (2019) Noiseprint: A cnn-based camera model fingerprint. IEEE Trans Inf Forens Secur PP(99):1–1
49. 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
50. Zhao H, Shi J, Qi X, Wang X, Jia J (2017) Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2881–2890
51. 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
52. Shullani D, Fontani M, Iuliani M, Shaya OA, Piva A (2017) Vision: a video and image dataset for source identification. EURASIP J Inf Secur 2017(1):1–16
53. Galdi C, Hartung F, Dugelay J-L (2019) Socrates: A database of realistic data for source camera recognition on smartphones. In: ICPRAM, pp 648–655
54. Wang B, Yu F, Ma Y, Zhao H, Hou J, Zheng W (2023) Pcep: Few-shot model-based source camera identification. Mathematics 11(4):803
55. Tan Y, Wang B, Li M, Guo Y, Kong X, Shi Y (2016) Camera source identification with limited labeled training set. In: Digital-forensics and watermarking: 14th international workshop, IWDW 2015, Tokyo, Japan, October 7-10, 2015, Revised Selected Papers 14. Springer, pp 18–27