文章编号：1003-0530(2024)06-1122-19

JPEG原图重构取证方法

黎溢轩[1] 李昊东[1] 曾吉申[2] 黄瑞灵[1] 黄继武*[1]
（1. 广东省智能信息处理重点实验室，广东深圳 518060；
2. 阿里巴巴集团，浙江杭州 310000）

摘 要：数字图像已被广泛用于办理各类网上业务和作为司法证据。与此同时，利用流行的图像编辑软件，普通用户就可以对图像进行语义篡改而不留下视觉痕迹。因此，对数字图像的原始性和真实性进行辨识已成为迫切的应用需求。基于元数据的图像篡改取证方法因准确性高、计算量小而得到重视。然而，原图重构技术（例如应用MagicEXIF元数据编辑器）的出现使上述方法完全失效。针对这一问题，本文提出一种JPEG图像原图重构取证方法，用于检测图像是否受到重构攻击。通过分析原图重构的过程，以及重构前后图像的像素统计特征差异，本文对深度学习隐写分析模型SRNet（Steganalysis Residual Network）进行轻量化改进：裁剪其冗余的下采样层以减少参数，引入通道注意力机制以提高对关键特征的提取能力，并采用知识蒸馏的方法进一步提升模型的准确率。进一步地，通过分析重构对不同颜色分量的影响，采用YCbCr颜色分量作为模型输入，以提高检测性能。为测试算法的性能，我们收集了由不同品牌和型号的手机拍摄的图像数据，构建了大规模重构图像数据库。实验表明，本文提出的模型在参数量显著减少的情况下性能优于流行的模型，对512×512大小的图像可取得98%以上的检测正确率，且具有良好的跨设备泛化能力。同时，通过应用迁移学习，本文方法对不同版本的重构软件也具有较好的泛化性。

关键词：图像元数据；原图重构；图像取证；重构检测；轻量化模型
中图分类号：TP391.4 **文献标识码：**A **DOI：**10.16798/j.issn.1003-0530.2024.06.012

引用格式：黎溢轩，李昊东，曾吉申，等. JPEG原图重构取证方法[J]. 信号处理，2024，40(6)：1122-1140. DOI：10.16798/j.issn.1003-0530.2024.06.012.
Reference format：LI Yixuan，LI Haodong，ZENG Jishen，et al. Forensic method for JPEG image reconstruction[J]. Journal of Signal Processing，2024，40(6)：1122-1140. DOI: 10.16798/j.issn.1003-0530.2024.06.012.

Forensic Method for JPEG Image Reconstruction

LI Yixuan[1] LI Haodong[1] ZENG Jishen[2] HUANG Ruiling[1] HUANG Jiwu*[1]
（1. Guangdong Key Laboratory of Intelligent Information Processing，Shenzhen，Guangdong 518060，China；
2. Alibaba Group，Hangzhou，Zhejiang 310000，China）

Abstract: Digital images have been widely used in various online businesses and as judicial evidence. Simultaneously，using popular image editing software，ordinary users can tamper with the image semantics without leaving visual traces. Therefore，identifying the originality and authenticity of digital images has become an urgent application requirement. Image tampering forensics based on metadata has garnered attention due to its high accuracy and minimal computational requirements. However，the emergence of original image reconstruction technologies，exemplified by tools like the MagicEXIF metadata editor，renders the aforementioned methods entirely ineffective. To solve this problem，this study

收稿日期：2023-06-15；修回日期：2023-09-17
*通信作者：黄继武 jwhuang@szu.edu.cn *Corresponding Author：HUANG Jiwu，jwhuang@szu.edu.cn
基金项目：国家自然科学基金（U19B2022）；广东省自然科学基金（2022A1515010145）
Foundation Items：The National Natural Science Foundation of China（U19B2022）；Guangdong Provincial Natural Science Foundation of China（2022A1515010145）

proposes a JPEG original image reconstruction forensics method to detect if the image was reconstructed. By analyzing the original image reconstruction process and the difference in pixel statistical characteristics of the image before and after reconstruction, this study develops a lightweight improvement on the deep learning steganography analysis model steganalysis residual network (SRNet): it cuts its redundant lower sampling layer to reduce parameters, introduces the channel attention mechanism to improve the ability to extract key features, and uses the knowledge distillation method to further improve the accuracy of the model. Furthermore, by analyzing the influence of reconstruction on different color components, the YCbCr color component is used as the model input to improve the detection performance. To test the performance of the algorithm, we collected image data captured by mobile phones of different brands and models and built a large-scale reconstruction image dataset. The experiment demonstrates that the proposed model outperforms a popular model, even with a significantly reduced number of parameters. For a 512×512 image, the proposed model achieves a detection accuracy exceeding 98% and exhibits strong cross-device generalization capability. Simultaneously, through the application of transfer learning, the proposed method also achieved good generalization for different versions of reconstructed software.

Key words: image metadata; image reconstruction; image forensics; reconstruction detection; lightweight model

1　引言

数字图像越来越多应用于网络信息服务。例如,各种证件/文档扫描件以图像形式通过网络传送或提交。与此同时,借助操作方便、功能齐全的图像编辑软件(例如Adobe Photoshop、ACDSee、美图秀秀、光影魔术手等)[1],使用者不需要专业知识就能轻松地对数字图像进行编辑,生成眼睛难以识别的伪造图像。伪造图像的滥用正在给社会和个人带来越来越大的安全隐患[2]。JPEG是当前最为流行的图像格式。针对JPEG图像的篡改取证,具有广泛的应用意义,因此也成为了图像取证领域一个重要的研究问题。

JPEG图像篡改取证是一种数字取证技术,旨在从JPEG图像中获取和分析数字证据用于检测和分析图像是否经过篡改操作,以确定JPEG图像的真实性。根据检测的依据不同,可分为基于设备物理特性的方法[3-4]、基于光照一致性的方法[5-6]、基于压缩痕迹的方法[7-10]、基于JPEG元数据(Metadata)的方法[11-12]、基于内在统计特征的方法[13-14]、基于深度学习的方法[15-17]等。

基于JPEG元数据的图像取证方法因准确性高、计算量小而得到重视,常被用于高实时性需求、大批量图像处理的实际应用场合。JPEG元数据记录了很多图像参数。这些参数因拍摄设备不同而存在差异,且它们在图像经过软件编辑后又会被相应修改。通过对比和分析图像元数据中的异常,就能识别图像真实与否[18]。目前国内外已有许多工作利用JPEG图像的元数据信息对JPEG图像的原始来源进行取证。例如,文献[12]根据JEPG图像的量化表、哈夫曼表及EXIF信息判断图像是否经过Photoshop处理;文献[19]利用JPEG元数据信息中的13个特征判断图像是否由数码设备生成;文献[20]分析了JPEG格式和元数据信息,并将缩略图和压缩质量等元数据信息应用于司法鉴定,实现JPEG的真实性取证;文献[21]根据JPEG的元数据信息包括拍摄时间、量化表、缩略图等特征,利用元数据进行JPEG图像原始性认证,检测准确,证据清楚,计算简单,确定JPEG格式图像是否经过图像处理软件处理。

但是,基于元数据的取证技术存在抗攻击能力弱的缺点。篡改后的图像元数据有可能被有经验的攻击者再次篡改,以消除编辑软件留下的痕迹,从而导致取证技术失效。以前,要利用软件篡改图像而又不在图像和元数据上留下明显的篡改痕迹,需要相当的图像处理和反取证(Anti-forensics)专业知识。近年来,出现了一款针对JPEG图像格式的专业级别元数据编辑器——MagicEXIF元数据编辑器,能对元数据进行无损修复[1]。MagicEXIF元数据编辑器提供的原图重构工具可通过用户选定的照相设备型号对图像进行模拟重新拍摄操作,从而在不修改图像内容的情况下生成无异常元数据的重构图像。该工具使用简单,不需要依赖任何JPEG相关的专业知识就能轻松完成操作。因此,篡改者可以在利用图像编辑软件对图像进行篡改后,使用MagicEXIF软件的原图重构功能伪造指定型号的相机/手机元数据信息,从而逃避现有的基于元数据的图像取证检测。因此,如何检测这种重构攻

[1] https://www.magicexif.com/help/products/editor

击，成为JPEG图像取证需要解决的一个实际问题。目前，尚未有对MagicEXIF重构进行检测的报道。尽管MagicEXIF重构检测任务与重压缩检测任务有一定的相似性，但是在应用场景和检测目的等方面存在较大的差异，并且后续实验表明，重压缩方法应用于重构任务效果并不理想。

针对上述JPEG原图重构的取证任务，本文分析了原图重构的过程以及重构前后图像特征的差异，并提出一种针对MagicEXIF的JPEG原图重构检测方法。首先，分析了重构攻击对图像造成的影响，阐明了重构检测在机理上的可行性。接着，考虑到图像隐写分析和取证分析任务的相关性，我们将基于深度学习的隐写分析残差网络SRNet[22]作为重构检测的原型网络。进一步，为了满足实际应用场景计算要求，我们对网络进行轻量化改进，应用网络剪枝加速模型推理，并引入注意力机制和应用知识蒸馏的方法进一步提升模型的准确率。在此基础上，从理论和实验研究了不同颜色空间对重构检测的有效性，选用YCbCr作为输入图像颜色空间。最后，针对MagicEXIF元数据编辑器不同版本重构功能存在差异的情形，应用迁移学习使所训练的模型适用于不同应用场景。我们建立了涵盖目前市面上8个主流手机品牌共67个手机型号的大型重构数据库，以对所提出的重构检测算法进行性能评估。实验结果表明，本文提出的模型与主流相关方法相比，在不同的测试集上均取得了最好的检测性能，且所需参数量显著少于主流的网络。跨库实验展示了所提方法对跨设备及跨重构软件版本等情况均具有良好的泛化性。

本文主要贡献如下：

1）从机理上分析了重构图像与非重构图像存在的特征差异；分析了不同彩色分量对重构检测的有效性。为JPEG原图重构的检测算法设计提供了原理支撑。

2）提出了一个基于深度学习的JPEG图像检测算法。在上述分析基础上，我们应用隐写分析网络SRNet，根据重构检测的实际需求进行了轻量化改进，并引入通道注意力机制以提高对关键特征的提取能力，随后采用知识蒸馏的方法进一步提升模型的准确率，设计了一个轻量高效的JPEG重构检测算法。本文算法是第一个公开报道的JPEG图像原图重构检测算法。

3）根据实际应用场景构建了大型重构数据库，在所构建的数据库上与主流相关方法进行对比。实验结果表明，本文算法不仅在网络所需参数量显著少于主流的网络，而且在不同的测试集上均取得了最佳的检测性能。

本文的后续内容安排如下：第2节介绍MagicEXIF元数据编辑器和原图重构工具的功能，并分析了原图重构产生的特征差异；第3节介绍所提出的重构检测算法，包括对原型网络的改进细节及颜色分量的选择；第4节展示实验结果及进行讨论分析；第5节对本文工作进行总结，并对未来的改进工作进行展望。

2 MagicEXIF原图重构及其分析

2.1 MagicEXIF元数据编辑器

JPEG元数据是指JPEG图像文件中包含的一些附加信息，如图像的分辨率、图像的创建日期、地点信息等，它们可以为JPEG图像提供背景信息，从而支持图像的处理和技术分析，部分元数据信息如表1所示。此外，JPEG元数据还可以用于认证图像的真实性，从而提高JPEG图像篡改的检测率。

表1 JEPG图像部分元数据信息标记及含义

Tab. 1 Labeling and meaning of meta data information in JEPG Images

应用标记	含义	应用标记	含义
Make	生产者	Model	型号
Orientation	方向	DateTime	日期时间
DateTime Original	创建时间	ResolutionUnit	分辨率单位
ExifIOffset	位置信息	Software	软件
YCbCr Position	色相定位	Exposure Time	曝光时间

MagicEXIF元数据编辑器是一个专业级的JPEG照片元数据查看和编辑软件。它支持读取和修改JPEG图像中的EXIF、GPS、XMP、厂商注释等元数据信息[23]，并且存有多个从真实原图样例中提取的设备元数据模板，可对图像进行重构操作。该软件既是目前最全面的EXIF查看器，也是目前唯一能无损修改元数据以实现原图重构的EXIF编辑器。

2.2 原图重构

本文所述的原图重构，是指利用MagicEXIF元数据编辑器提供的原图重构工具对图像进行处理。一般地，Photoshop等图像编辑软件在保存图像时会在元数据中自动插入软件私有信息以方便管理图像编辑历史。该类信息并不影响图像内容，但可作

为检测图像是否经过软件处理的依据。MagicEXIF元数据编辑器可以识别出这些私有信息，并通过原图重构功能将其彻底抹除。在给定重构的目标设备时，MagicEXIF的原图重构工具可根据选定设备的编码算法重新模拟一遍JPEG编码过程，在压缩时采用与所选目标设备相同的量化表，生成数据特征相同的重构图像。因此，重构图像在元数据特征和底层编码格式上均与目标设备的实拍照片没有差异。

MagicEXIF重构图像生成方式有两种：一种是篡改者对手机拍摄的原图直接进行重构操作，通过更改拍摄时间、地点及设备等元数据信息，达到伪造图像来源的目的，过程如图1所示；另一种是篡改者对手机拍摄的图像利用图像编辑软件进行修改，通过重构工具抹除软件处理的痕迹，将其伪造成未经修改的原始图像，过程如图2所示。

图1　原始图像进行重构操作

Fig. 1　Reconstructing the original image

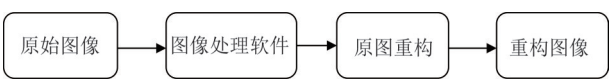


图2　原始图像经过软件编辑后进行重构操作

Fig. 2　Reconstruct the original image after software editing

因此，在对JPEG图像进行原始性辨识时，不仅需要对图像的元数据进行取证，还需要检测JPEG图像是否经过MagicEXIF元数据编辑器重构操作。

值得注意的是，MagicEXIF在v1.05版本后对重构功能进行了改进，优化了重构的算法[2]。因此，MagicEXIF v1.05版本前后的原图重构结果是存在较大差异的。本文中我们将v1.05之前的版本称为旧版本。根据我们的实验，利用新版本的原图重构工具所生成的图像数据训练得到的模型对旧版本的原图重构工具生成的图像数据检测能力很差。这也是实际应用时需要考虑的问题之一。

2.3　重构检测的机理

由于MagicEXIF的原图重构工具能恢复图像原有属性，这就导致基于元数据信息或设备指纹的原始性检测方法无法正确地识别出图像的真实来源。为分析真实应用场景中原始图像经过图像编辑软件处理后再重构过程里元数据信息和设备指纹信息的变化，我们将苹果手机拍摄的原始图像利用美图秀秀软件进行压缩模拟篡改过程，然后利用MagicEXIF元数据编辑器的重构操作指定为索尼相机拍摄的图像。如图3所示为模拟篡改过程中部分元数据信息的变化，左图代表苹果手机拍摄的图像信息，中间代表利用美图秀秀进行压缩后得到的图像信息，右图为上述压缩图像经过原图重构工具处理后的重构图像信息，可看出当原始图像经过图像编辑软件处理后，部分拍摄参数被修改且留有篡改痕迹，而经过重构操作模拟为索尼相机拍摄后，拍摄设备、拍摄时间等相关参数及软件信息均被修改替换，元数据信息中原始设备特征及图像处理软件处理后遗留的痕迹均被抹除。图4所示为模拟篡改过程中设备指纹信息的变化，由图可以看出原始图像经过图像处理软件压缩处理后，设备指纹信息发生了变化并带有图像处理软件的指纹信息，而再经过重构操作模拟压缩后，设备指纹信息再次被压缩变为索尼相机的压缩指纹，无法根据设备指纹信息正确地获取图像的真实来源，这为篡改取证带来了新的难题。

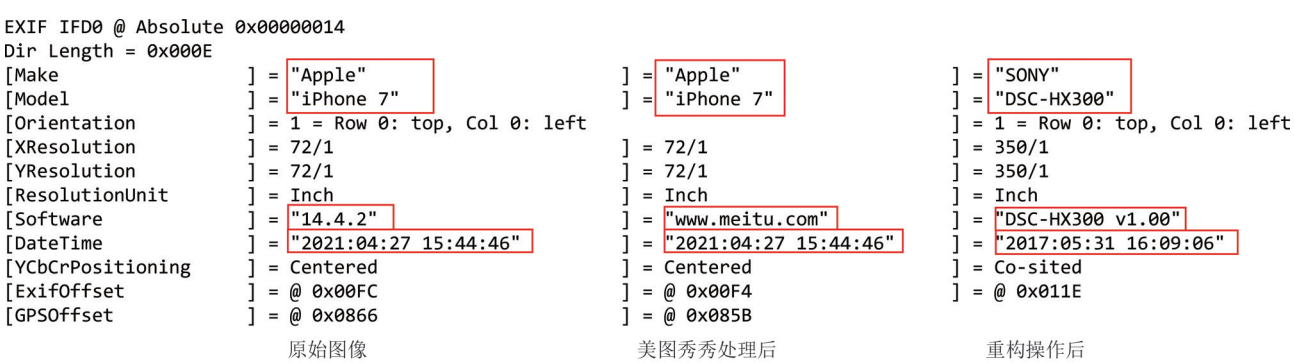


图3　图像的部分元数据信息变化

Fig. 3　Changes in partial metadata information of images

[2] https://www.magicexif.com/start/log

```
Table length = 132
----
Destination ID=0 (Luminance)
DQT, Row #0:   2   2   2   3   5   6   8  10
DQT, Row #1:   2   2   2   3   5   6   8  10
DQT, Row #2:   2   2   3   5   6   8  10  12
DQT, Row #3:   3   3   5   6   8  10  12  14
DQT, Row #4:   5   5   6   8  10  12  14  15
DQT, Row #5:   6   6   8  10  12  14  15  15
DQT, Row #6:   8   8  10  12  14  15  15  15
DQT, Row #7:  10  10  12  14  15  15  15  15
Approx quality factor = 91.94
(scaling=16.12 variance=12.56)
----
Destination ID=1 (Chrominance)
DQT, Row #0:   2   2   4   7  16  16  16  16
DQT, Row #1:   2   4   4  11  16  16  16  16
DQT, Row #2:   4   4   9  16  16  16  16  16
DQT, Row #3:   7  11  16  16  16  16  16  16
DQT, Row #4:  16  16  16  16  16  16  16  16
DQT, Row #5:  16  16  16  16  16  16  16  16
DQT, Row #6:  16  16  16  16  16  16  16  16
DQT, Row #7:  16  16  16  16  16  16  16  16
Approx quality factor = 92.03
(scaling=15.95 variance=1.27)
```

原始图像

```
Table length = 67
----
Destination ID=0 (Luminance)
DQT, Row #0:   2   1   1   2   2   4   5   6
DQT, Row #1:   1   1   1   2   3   6   6   6
DQT, Row #2:   1   1   2   2   4   6   7   6
DQT, Row #3:   1   2   2   3   5   9   8   6
DQT, Row #4:   2   2   4   6   7  11  10   8
DQT, Row #5:   2   4   6   6   8  10  11   9
DQT, Row #6:   5   6   8   9  10  12  12  10
DQT, Row #7:   7   9  10  10  11  10  10  10
Approx quality factor = 95.04
(scaling=9.93 variance=1.25)
----
Destination ID=1 (Chrominance)
DQT, Row #0:   2   2   2   5  10  10  10  10
DQT, Row #1:   2   2   3   7  10  10  10  10
DQT, Row #2:   2   3   6  10  10  10  10  10
DQT, Row #3:   5   7  10  10  10  10  10  10
DQT, Row #4:  10  10  10  10  10  10  10  10
DQT, Row #5:  10  10  10  10  10  10  10  10
DQT, Row #6:  10  10  10  10  10  10  10  10
DQT, Row #7:  10  10  10  10  10  10  10  10
Approx quality factor = 94.91
(scaling=10.18 variance=0.26)
```

美图秀秀处理后

```
Table length = 132
----
Destination ID=0 (Luminance)
DQT, Row #0:   1   1   1   1   1   1   1   1
DQT, Row #1:   1   1   1   1   1   1   1   1
DQT, Row #2:   1   1   1   1   1   1   1   1
DQT, Row #3:   1   1   1   1   1   1   1   1
DQT, Row #4:   1   1   1   1   1   1   1   1
DQT, Row #5:   1   1   1   1   1   1   1   1
DQT, Row #6:   1   1   1   1   1   1   1   1
DQT, Row #7:   1   1   1   1   1   1   1   1
Approx quality factor = 100.00
(scaling=2.99 variance=6.13)
----
Destination ID=1 (Chrominance)
DQT, Row #0:   1   1   1   1   1   1   2   2
DQT, Row #1:   1   1   1   1   1   1   2   2
DQT, Row #2:   1   1   1   1   1   1   2   2
DQT, Row #3:   1   1   1   1   1   2   2   2
DQT, Row #4:   1   1   1   1   1   2   2   2
DQT, Row #5:   1   1   1   2   2   2   2   2
DQT, Row #6:   2   2   2   2   2   2   2   2
DQT, Row #7:   2   2   2   2   2   2   2   2
Approx quality factor = 98.97
(scaling=2.06 variance=1.28)
```

重构操作后

图 4 图像的设备指纹信息变化

Fig. 4 Changes in device fingerprint information of images

JPEG 图像的元数据信息和像素特征常被作为 JPEG 图像取证的依据，因此我们尝试从元数据信息和图像像素两个方向对重构操作进行检测。

从元数据上看，原图重构工具已对图像的元数据进行重新封装，因此无法直接从元数据信息中检测出图像是否经过了原图重构操作。

从图像像素上看，重构操作并不会改变图像的语义内容，但是 JPEG 压缩是一种有损压缩，即便按照给定的压缩参数，在压缩和解压缩过程中，仍然有一定的像素信息被损失或改变，重构过程中特有的编解码过程以及使用指定设备的量化表对图像进行重新压缩，必然会使图像的像素值发生变化导致重构前后图像特征存在差异，这种差异为重构检测提供了可行性。

在图 1 所示情况中，原始设备拍摄的原始图像只经历了拍摄设备量化表的单次压缩，而重构图像经历了拍摄设备压缩和 MagicEXIF 指导设备模拟重构的两次 JPEG 压缩。在图 2 所示情况中，篡改图在拍摄设备压缩的基础上增加了图像处理软件特有量化表的二次压缩，而重构图像则是由照相设备量化表的第一次压缩、图像编辑软件的量化表二次压缩、MagicEXIF 模拟的量化表三次压缩后得到。综上可见，重构图像总比非重构图像经历多一次压缩，且最后一次压缩总是 MagicEXIF 进行的压缩。

由重构压缩引入的痕迹最终都会反映在图像像素的细微变化上。如图 5 所示，图像重构前后在视觉上没有明显变化。然而，我们在 iPhone11、Huawei P20 和 Oppo A72 三个不同品牌的手机设备拍摄的元原始图像中随机选取 100 幅图像，分别执行重构操作、图像处理软件处理、图像处理软件处理后再重构操作，得到原始图像、原始图像经过重构后的图像、原始图像经过图像处理软件处理后的图像、原始图像经过图像处理软件处理后再重构的图像四类图像各 100 幅，通过提取每幅图像的 Y 分量，计算 Y 分量中相邻像素的差值进行求和取平均，统计得到差分像素分布的概率（核密度曲线）。结果如图 6 所示，左侧代表原始图像直接进行重构操作前后的差分概率分布图，右侧代表原始图像经过图像处理软件处理后再重构操作前后的差分概率图，(a)(b)(c)分别代表原始图像来源设备品牌为苹果品牌、华为品牌和 OPPO 品牌。可以发现在不同品牌中，无论是原始图像直接进行重构，还是原始图像经过图像处理软件处理再进行重构，重构前后像素差的概率分布呈现出明显差异。这表明通过图像像素进行重构检测是可行的。

重构检测与 JPEG 图像重压缩检测有一定的相似性，二者均是通过分析图像的压缩特性进行取证分析，但仍存在较明显的区别。一方面，重压缩检测的目的在于分辨出一次压缩和二次（以及二次以上）压缩。已有的重压缩检测研究中，目标图像通常基于标准 JPEG 压缩量化表进行二次压缩[24-25]，量化器已知。重构检测是要分辨出最后一次压缩是否为重构压缩处理（本文中是 MagicEXIF），目标图像量化器不统一、编辑历史未知（包括压缩算法和压缩次数未知）、传输来源不确定，往往具有更复杂的处理/压缩历史。重构检测任务环境的问题要更

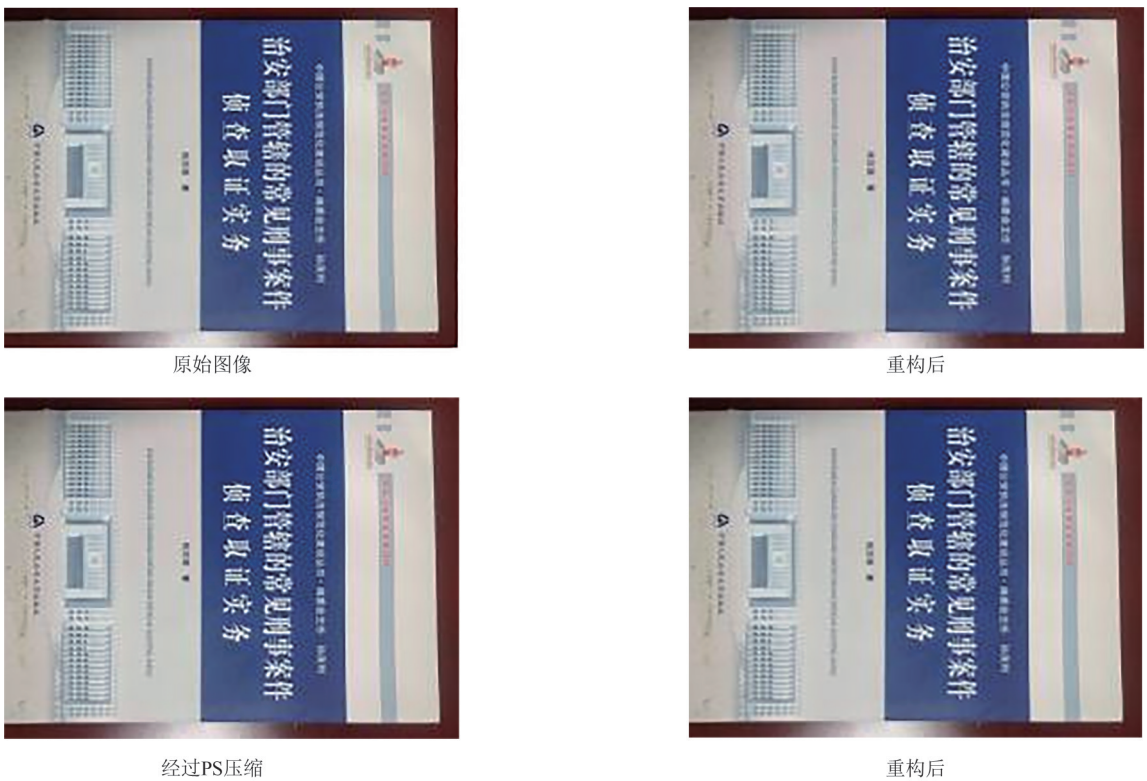


图5　重构前后图像像素变化

Fig. 5　Changes in image pixels before and after reconstruction

加开放。另一方面，已有重压缩检测主要聚焦在解决第一次压缩和第二次压缩质量因子相近，或者第一次压缩质量因子较高而第二次压缩质量因子较低的情况。重构检测任务关心的是，不论什么情况，只要最后一次压缩是重构算法引起的压缩，就需要检测出来。为了检验是否可以把重构检测作为重压缩检测的一个特例来处理，我们选取了具有代表性的重压缩检测算法[26]作为对比方法之一。实验结果表明，重压缩方法用于重构检测任务有一定效果，但性能并不理想。这也证实了本文所提出的取证方法的优越性和实用性。详细结果请参见第4.3节。

3　重构检测方法

本文提出的重构检测方法应用场景如图7所示。据上述分析，重构前后像素矩阵压缩特征存在差异，检测算法的关键即捕捉该差异特征。从这个意义上，重构检测任务与隐写分析任务相似，因此可以采用隐写分析模型进行检测。考虑在很多实际应用中，图像检测量大，图像原始性的检测需要快速完成，因此对所用模型进行轻量化，并采用注意力机制提升性能，随后采用知识蒸馏的训练方法进一步提升模型的准确率；最后，根据我们对重构前后彩色分量特征变化的分析，本文采用YCbCr作为输入颜色空间。

3.1　网络模型设计

为有效地从图像像素中提取适用于重构检测的特征，本文对基于SRNet的深度网络结构进行了改进，所设计网络模型的结构及其中的普通卷积块、残差卷积块如图8所示。改进方法由两部分组成：模型轻量化、引入通道注意力机制。

3.1.1　原型网络

考虑到重构检测与隐写分析有一定相似性，我们采用基于深度学习的隐写分析模型SRNet作为基本网络结构。该网络利用残差网络的特点模拟了传统SRM[27]的特征提取过程，在空域和JPEG域隐写分析上都取得了较好的效果。SRNet可按功能分为三个部分：

1）隐写噪声提取：该部分由网络的前7个块构成。前两层采用普通卷积块结构，利用深度网络自身的强拟合能力，增强网络的图像细节提取能力，后5层添加了ResNet[28]中提出的残差卷积结构。

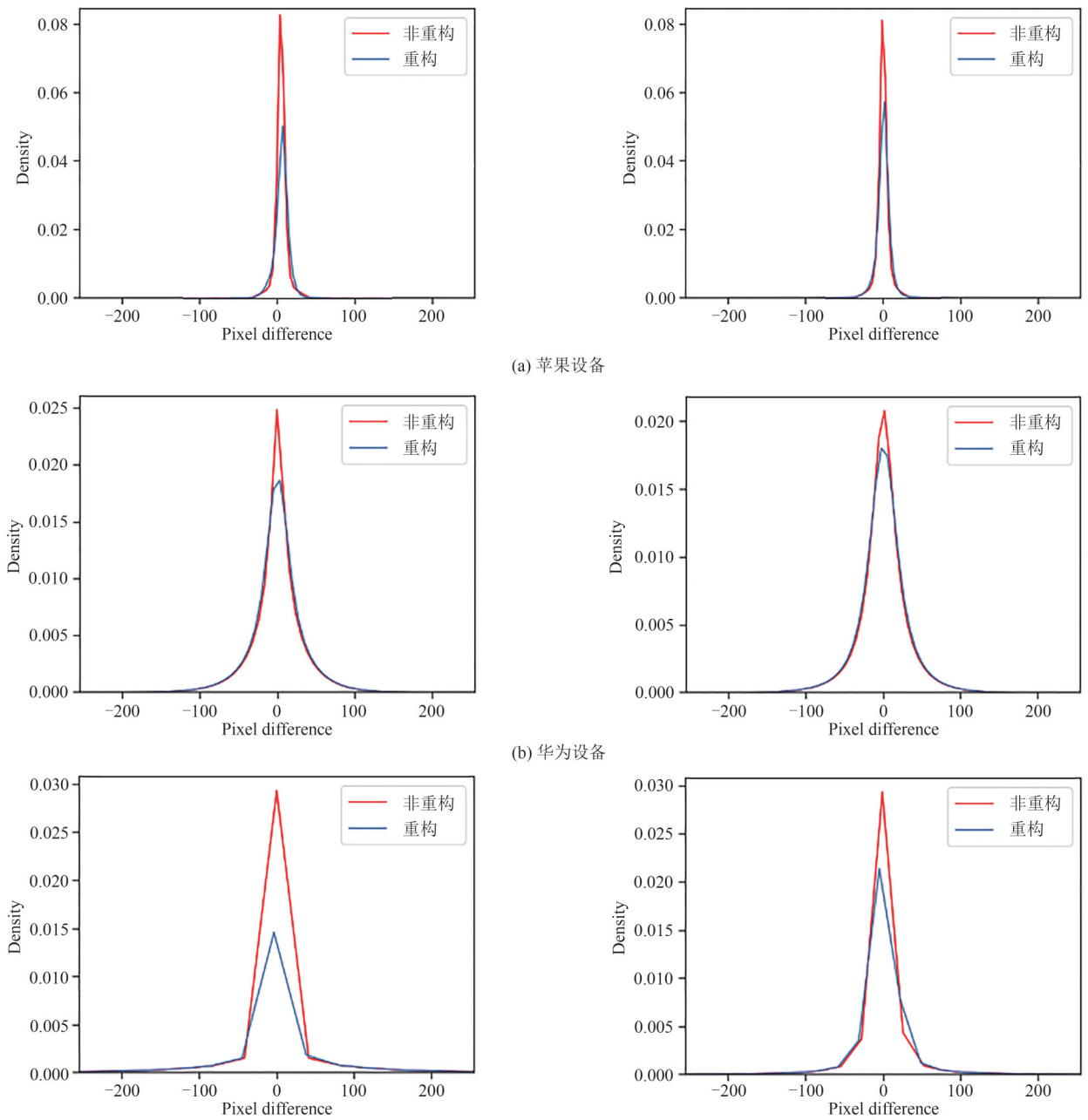


图 6 不同品牌手机拍摄的图像重构前后的差分概率图

Fig. 6 Differential probability maps of images captured by different brands of mobile phones before and after reconstruction

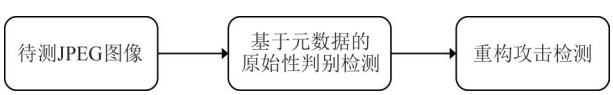


图 7 重构检测应用场景

Fig. 7 Reconstructing detection application scenarios

网络前 7 层均不采用下采样层，因为下采样池化会导致隐写噪声的衰减，不利于所需噪声特征的提取[29]。由于使用了残差结构，解决了随网络结构层数增加而带来的在反向传播中的梯度爆炸和梯度消失的问题，有利于学习到所需的“隐写噪声残差”，从而帮助网络在训练中更易得到全局最优解。

2）特征降维：该部分由网络的第 8~12 块组成。第 8~11 块采用了相同的单元块，在残差块中添加了平均池化层。第 12 个块在移除了残差结构，并添加了全局平均池化层。该部分通过池化的方法降低

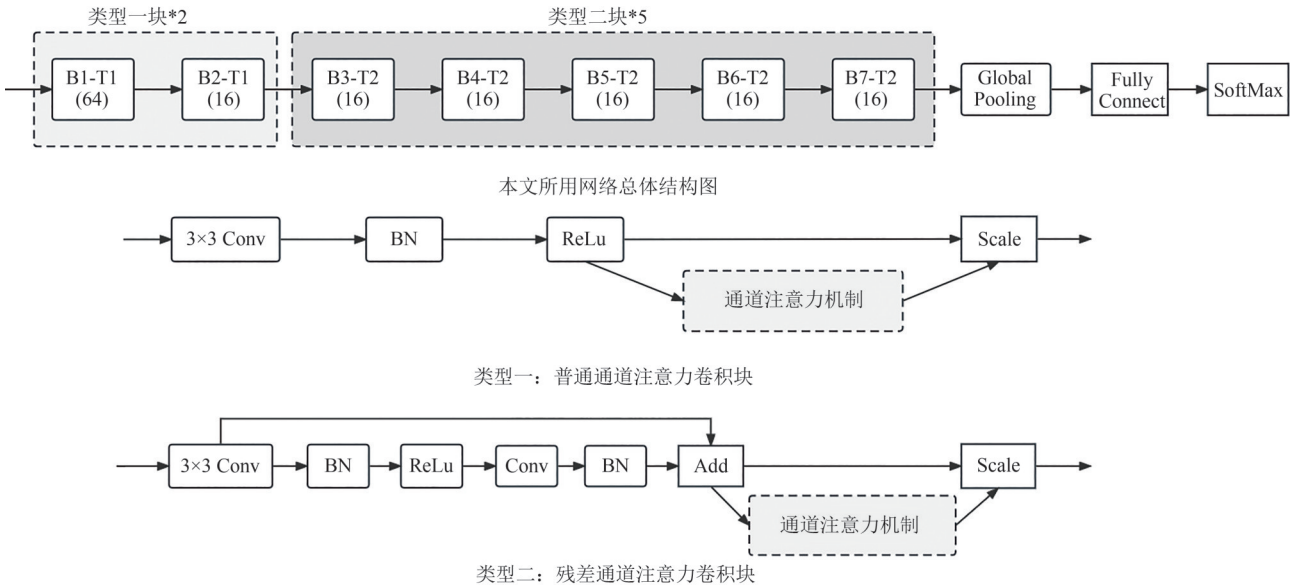


图 8　网络结构图及两种类型卷积块

Fig. 8　Network structure diagram and two types of convolutional blocks

了特征图的分辨率，同时增大了感受野。

3）线性分类器：该部分由网络的第 13 个块组成，将全局平均池化后的结果输入到全连接层（Fully Connected，FC），进而通过 Softmax 函数输出最终结果。

3.1.2　模型轻量化

SRNet在第 8~12 层采用了逐层下采样池化的方法，渐进地降低特征的分辨率。然而由于前 7 层隐写嵌入噪声提取部分为了减少信息丢失，没有采用池化层，导致下采样部分输入分辨率仍为 512×512，意味着每个下采样层都含有较大的计算量。考虑到 MagicEXIF 原图重构检测任务更关注全局空间位置信息，我们将冗余的逐层下采样层去除，保留全局平均池化层对模型分辨率进行压缩降维。这样极大地减少了模型的参数量，便于模型的落地应用。本文中将轻量化后的只保留 SRNet 前 7 层的模型称为 SRNet_L7。

3.1.3　通道注意力机制

由于直接采用全局平均池化替换第 7 层之后的卷积层，带来了残差信息的丢失问题，这会导致模型精度下降。为此，我们引入注意力机制增强模型对关键特征的提取能力，弥补残差信息丢失对模型精度影响。

近年来注意力机制在计算机视觉中起到了重要作用。注意力机制可以看作是一种网络动态选择的过程，可根据输入和注意力机制的类型自适应地对特征进行加权。常见的注意力机制有空间注意力机制（SAM）[30]、通道注意力机制（CAM）[31]、卷积块注意力机制（CBAM）[32]等。本文选用通道注意力机制，通过在通道域中生成注意力掩码，从而筛选出重要的通道并赋予更大的权重实现注意力的转移。在重构检测任务中选用通道注意力的理由如下：

1）通道注意力机制计算复杂度低，只需增加少量计算量即可获得明显的性能提升，便于模型落地部署和短时间内对大批量图像进行检测；

2）通道注意力机制通过分析不同通道分量与图像关键信息的相关性，能帮助浅层网络更好地利用全局感受野的信息，达到排除无用信息的目的，从而取得性能的提升。

3）由于重构过程是对整幅图像进行操作，重构检测任务不需要对被重构位置进行判断，因而不适合采用空间注意力机制。使用通道注意力机制能够自适应地为通道分配权重，更好地捕捉重构痕迹，从而取得更好的检测性能。

通道注意力机制结构如图 9 所示。

通道注意力机制作为一种轻量高效的提升模型性能的模块，在目前主流的 MobileNet 系列[33]、EfficientNet 系列[34]等轻量级模型中均能看到它的身影，进一步说明它对于提升模型性能的显著能力。

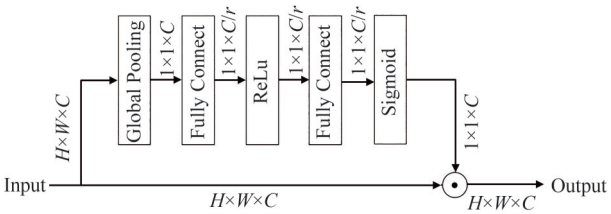


图 9 通道注意力机制

Fig. 9 Channel attention mechanism

3.1.4 知识蒸馏算法

知识蒸馏(Knowledge distillation,KD)的本质是通过神经网络迁移学习训练的方法,由于简单的网络受限于层数无法学习到深层的特征,所以先基于真实数据训练出一个大型的“教师模型”,将复杂的“教师网络”学习到的“知识”和真实的标签数据共同指导简单的“学生网络”,使所需的“学生模型”在保持低网络复杂度的同时保留接近复杂网络的性能[35]。通常来说,“教师网络”因为拥有更复杂的网络结构从而能够捕捉更深层的特征信息取得更好的分类效果,且其分类能力越好,越有利于“学生网络”能力的提升。

在一般的有监督分类任务中,常用“0”和“1”等硬标签的方式对数据进行标注,但这种表达方式包含信息量相对有限,无法表示类与类之间的关系导致模型无法取得更好的性能。而知识蒸馏将复杂度高、性能更好的“教师网络”的输出概率作为数据的软标签,引导“学生模型”进行学习。一方面,让“学生网络”学习到不同类别之间的相似性,能够在不增加复杂度的前提下提升模型的性能;另一方面,“教师网络”的辅助训练有助于“学生网络”更好地拟合。

为了能够将分类输出“软化”,该算法在 Softmax 的基础上引入了温度系数 T 来描述输出概率的软化程度,起到保留相似信息的作用,代称为 SoftmaxT。T 的取值越大,输出的类别概率越平滑,越趋近于均匀分布;而 T 的取值越小,输出越概率越接近于 onehot,导致模型训练难度增加,当 T=1 时,等同于普通 Softmax,其计算公式如下:

$$\mathrm{SoftmaxT}=\frac{\exp(z_i/T)}{\sum_j^N \exp(z_j/T)} \tag{1}$$

其中 z_i 为神经网络输出层的结果,N 为输入向量维度。

知识蒸馏算法流程图如图 10 所示。对于同一个输入图像,首先经过训练好的“教师模型”得到预测输出,然后将其使用 SoftmaxT 变换得到软化后的概率分布标签(软标签)。随后,计算“学生模型”得到的预测输出,经过 SoftmaxT 变换后与“教师模型”得到的软标签计算 KL 散度(Kullback-Leibler Divergence),用于衡量两个概率分布之间的相似性。最后将“学生模型”得到的预测输出与真实标签(硬标签)计算交叉熵函数,用于表示预测标签与真实标签之间的差值。因此,知识蒸馏算法的总损失包含两个部分,其计算表达式如下:

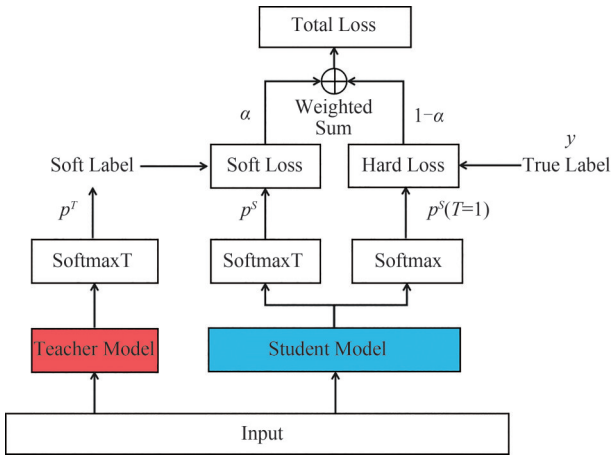


图 10 知识蒸馏网络结构图

Fig. 10 Knowledge distillation network structure

$$\mathrm{Loss}=\alpha \mathrm{L}_{\mathrm{soft}}(p^T,p^S)+(1-\alpha)\mathrm{L}_{\mathrm{hard}}(p^S,y) \tag{2}$$

$$\mathrm{L}_{\mathrm{soft}}(p^T,p^S)=-\sum p^S\log p^T \tag{3}$$

$$\mathrm{L}_{\mathrm{hard}}(p^S,y)=-\sum p^S\log y \tag{4}$$

其中,p^T 和 p^S 分别表示“教师网络”和“学生网络”的预测结果,y 为真实标签,α 为权重系数。权重系数 α 用于调整软标签和真实标签对模型总损失的影响。

在本次的研究任务中,本文选取在视觉分类任务中效果比较优越的 Vision Transformer 算法[36]。作为“教师网络”,以基于 SRNet 作轻量化后的网络作为“学生网络”,实现知识蒸馏算法的设计。

3.2 颜色空间

面向不同应用场合,数字图像可采用不同的颜色空间来表示[37]。而不同颜色空间的图像作为神经网络的输入,反映出不同的特征直接影响着网络的分类结果。因此,在上述已构建网络的基础上,需要进一步挑选合适的颜色空间从而达到预期的检测性能。对于本文的重构检测任务,我们选用 YCbCr 作为输入的颜色空间,其中 Y 分量表示亮度信息,Cb、Cr 分量分别表示 RGB 中蓝色、红色分量与亮度值的色差信息[38]。RGB 作为最常见的颜色空间,在实际应用中也最为广泛,而本文针对重构检测任务选择 YCbCr 颜色空间而不是 RGB 颜色空

间，主要有以下三个方面考虑：

1）在JPEG图像的压缩编码过程中，首要进行颜色空间的转换，由RGB颜色空间转为YCbCr颜色空间，再进行后续的采样量化编码操作[39]。在压缩过程中，由于Y分量包含更多的视觉敏感信息，通常对Cb、Cr分量进行下采样以提高压缩率，Y分量与Cb、Cr分量使用的量化表不同，这会造成压缩效应的差异。此外，如表2所示，不同拍摄设备、图像编辑软件及MagicEXIF元数据编辑器对YCbCr分量采用了不同的下采样率，因此产生不同的编码效应[40]。这些差异有助于对原始图像、利用图像编辑软件篡改后的图像、重构图像的识别。所以，采用YCbCr颜色空间能够充分利用JPEG压缩编码的特性，有利于提高重构检测性能。

表2　不同来源的图像采样率

Tab. 2　Image sampling rates from different sources

设备/软件	采样率	设备/软件	采样率
苹果	4:1:1	Sony	2:1:1
华为	4:1:1	ps	4:1:1、2:1:1
小米	4:1:1	ps_web	4:1:1、1:1:1
诺基亚	4:1:1	ACDSee	4:1:1、1:1:1
魅族	2:1:1	光影魔术手	4:1:1、2:1:1、1:1:1
Canon	2:1:1	美图秀秀	4:1:1、1:1:1
Nikon	2:1:1	MagicEXIF	4:1:1、1:1:1

2）在RGB颜色空间中，三个颜色分量的构成都与亮度信息有较大的相关性，每个分量都携带有亮度信息。在一些亮度差异较大的场景中（如高强光或黑夜等），RGB颜色空间相对来说更易受到影响[41]。而YCbCr将亮度和色度信息分离，能够减轻光照的影响，具有更好的稳定性，同时也保留了关键特征，有助于使模型拥有更好的泛化能力。

3）YCbCr颜色空间比RGB颜色空间在颜色和亮度细节上更加准确，可以提供更多的图像信息。此外，YCbCr颜色空间通过使用增益技术，能够更好地保存图像的细节。对于重构检测任务来说，当图像整体语义信息不发生变化时，如何更好地捕捉细节差异是实现检测任务的关键。

4　实验结果

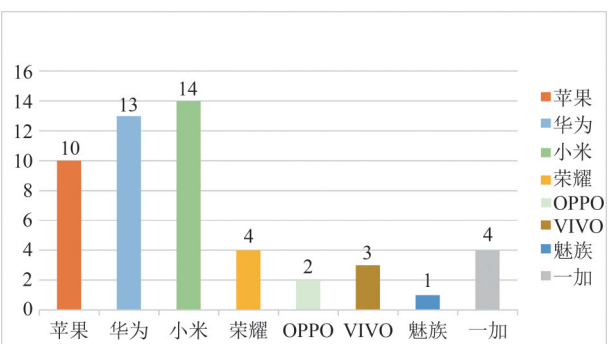
4.1　实验设置

4.1.1　数据库

由于在现实中缺乏大规模的重构图像数据，为验证所提出方法的有效性，我们需要首先构建相应的数据库。各种手机拍摄的图像，都有可能经过不同的图像处理软件处理并进行重构抹除痕迹，因而数据库要求图像内容丰富、主流手机型号齐全、涵盖主流图像处理软件。在所构建的数据库中，原始图像和经各种软件处理后的图像被认为是非重构图像，将经过MagicEXIF软件重构后的图像被认为是重构图像。

根据MagicEXIF重构检测任务，我们构建了三个数据库：设备型号最齐全的重构数据库MagicEXIF_basic、用于检测算法模型泛化能力的跨库数据库MagicEXIF_extra和利用低版本MagicEXIF原图重构工具构建的数据库MagicEXIF_old。

MagicEXIF_basic数据库。为构建全面的原始图像数据库，我们收集了市面上常见的8种手机品牌共51个手机型号，各手机型号如图11所示。实验中使用手机对纸质文件（封面、证书）及日常风景（室内、户外）等两类拍摄场景至少应用三种拍摄模式（滤镜、人像、夜景等）进行拍摄，每部手机在上述两类场景各拍摄50幅图像，得到有效的图像共5100幅。



used in the experiment

图11　实验所用的手机品牌及型号数

Fig. 11　Number of mobile phone brands and models used in the experiment

使用以下五种图像处理软件对上述拍摄的原始图像进行处理：Adobe Photoshop（“PS”）、网页版Photoshop（“PS_web”）、ACDSee、光影魔术手、美图秀秀。不同软件在保存JPEG图像时可供选择的压缩等级也各不相同，如PS有13个压缩等级，ACDSee则有100个压缩等级。为了统一选取，以PS的亮度量化因子为基准选取13种压缩等级，在其他的图像处理软件中挑选出13个最接近PS量化因子的压缩参数，不同软件的存储压缩参数如表3所示。

表 3 各软件存储压缩参数

Tab. 3 Each software stores compression parameters

软件	参数
PS	{ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9,10,11,12}
PS_web	{19,20,26,31,37,42,49,45,55,65,73,85,98}
ACDSee	{13,23,40,45,53,60,68,36,73,80,75,93,98}
光影魔术手	{46,52,62,66,70,75,81,77,84,88,92,95,99}
美图秀秀	{46,52,62,66,70,75,81,77,84,88,92,95,99}

根据参数表使用不同软件对原始图像进行压缩，不改变图像语义内容，共得到331500幅经过图像处理软件压缩的图像。

MagicEXIF元数据编辑器的原图重构工具目前支持的重构模板包括三种相机品牌（Canon、Nikon、Sony）和两种手机品牌（苹果、华为），共157个相机/手机型号。为避免数据重复冗余的同时能保证遍历到每个重构模板，将拍摄的原始图像以及经过图像处理软件压缩后的图像随机分成157组，每组图像使用对应的重构模板进行编辑，共生成336600幅重构图像。

由于MagicEXIF在v1.05版本前后原图重构的结果是存在较大差异的，本文选用v1.03和v1.10分别作为改进重构功能前后的代表版本。数据库MagicEXIF_basic采用v1.10版本生成。

MagicEXIF_extra数据库。实际应用中，待检测图像的拍摄手机型号可能不包含在训练集中。为测试该情况下算法的泛化能力，我们构建了MagicEXIF_extra数据库，该数据库包含两部分，搭建所用设备来源如表4所示：

表 4 MagicEXIF_extra图像来源

Tab. 4 MagicEXIF_ Extra image source

类型	品牌	型号
跨型号	苹果	6s, Xr
	华为	P20，Mate30
	荣耀	V10,V20
	OPPO	A72，Renoace
	VIVO	noe3, noe5
	一加	5T，7T_pro
	小米	Mi10，Note9
跨品牌	微软	Windows phone
	摩托罗拉	MotoX
	三星	Galaxy S5
	魅蓝	M1

①针对上述MagicEXIF_basic中的每个品牌，额外收集两个不在MagicEXIF_basic数据库中的型号。不参与训练，仅作为测试数据，验证模型对不同型号的跨库能力，生成跨设备重构图像共1600幅。

②收集四种不在MagicEXIF_basic数据库的品牌手机，按构建MagicEXIF_basic相同的过程，生成新的测试数据库，以验证模型对不同品牌的泛化能力，生成跨品牌重构图像共400幅。

MagicEXIF_old数据库。由于利用MagicEXIF新版本重构工具生成的图像训练的模型对低版本重构图像检测性能并不理想，我们在MagicEXIF_basic的非重构图像中随机选取两种拍摄场景各200幅，得到400幅非重构图像并使用v1.03版本进行重构，得到MagicEXIF_old数据库，由于对每幅图像进行了分块处理，实际训练中该数据库包含17200个图像块，该数据库用于迁移学习和性能检测。

4.1.2 数据预处理

由于深度网络的性能很大程度上依赖于输入数据，为了得到足够的训练数据，在输入网络之前对每幅图像进行裁剪，尽可能多地裁剪成512×512块，保证数据量充足。同时为了避免模型在训练过程中过拟合，在训练中对图像进行了数据增强。每幅512×512图像在输入网络之前，都有一定的概率会从旋转、水平翻转、镜像翻转中选取一种或多种方法进行变换。验证和测试时对图像进行TTA（Test Time Augment），对待测图像进行水平翻转，将待测图像和翻转后的图像进行预测，预测得到的两个结果取平均值作为最后的预测得分。

4.1.3 对比方法

考虑到实际的应用场景，本文选择了以下模型进行对比：主流的轻量级神经网络MobileNetV2[42]、EfficientNetB0[34]，基于机器学习的隐写分析网络SRM+SVM[43-44]，基于深度学习的隐写分析网络Zhu Net[45]，基于深度学习的重压缩检测网络Multi-domian Net[26]。

在所选对比算法中，与主流轻量级网络的对比是为了测试所提出的网络是否能兼顾速度与精度，从而满足实际应用场景的需求；隐写分析网络SRM+SVM和Zhu Net与本文提出的算法属于同领域算法，用于对比本文所用方法与同领域的算法之间的性能差异；重压缩检测算法原理与隐写分析算法接近，因此将重压缩检测性能较好的Multi-domian Net作为对比算法，一方面测试重压缩检测

算法对该任务的可行性，另一方面对比不同检测方法应用于本任务的性能差异。

4.1.4　训练设置

在输入分辨率上，由于Multi-domian Net网络设置固定输入尺寸为64×64，在训练时选取每个512×512块中间的64×64区域作为输入，其余模型以512×512尺寸图像作为输入。

所有模型基于Python3.8版本和PyTorch1.10版本搭建，使用一块NVIDIA RTX2080Ti进行训练及推理。整个训练过程可分为两个阶段：预训练和迁移训练。

预训练阶段使用MagicEXIF_basic数据库，按照8∶1∶1随机划分训练集、验证集和测试集。预训练阶段采用Kaiming初始化策略[46]对网络进行初始化，将主干网络全局平均池化后的结果输入到Softmax层来实现分类。使用二值交叉熵（Binary Cross-Entropy，BCE）作为损失函数，利用带Momentum的随机梯度下降法（Stochastic Gradient Descent，SGD）对网络进行优化，初始学习率设置为1×10^{-3}，总共设200个epoch，批大小为128，采用Cosine Annealing[47]对学习率进行动态调整，有利于避免模型陷入局部最优解。

在迁移训练阶段使用MagicEXIF_old数据库。为避免模型因迁移学习而导致原先性能大幅下降，在训练集中加入等量的MagicEXIF_basic的训练集数据。迁移训练的过程中，模型的主干网络使用预训练权重进行初始化，然后对模型进行微调。使用带Momentum的随机梯度下降法对网络进行优化，初始学习率设置为5×10^{-4}，总共设50个epoch，批大小为64，采用Cosine Annealing对学习率进行动态调整。

4.1.5　评价指标

在二分类任务中，原始标签可分为正类和负类两种情况，我们将重构样本视为正样本，非重构样本视为负样本。选用真阳率（TPR，即分类正确的正样本占正样本总数的比例）、真阴率（TNR，即分类准确的负样本占负样本总数的比例）和准确率（ACC，即分类正确的样本数占样本总数的比例）作为实验的评价指标。

4.2　消融实验

本节讨论了对网络结构的改变和颜色空间选择对重构检测任务的影响，为了分析和验证其有效性，基于MagicEXIF_basic数据库和MagicEXIF_extra数据库进行了消融实验，其中MagicEXIF_basic数据库用于训练和观察测试性能，MagicEXIF_extra数据库用于验证跨库性能。

4.2.1　网络结构的影响

为了验证剪裁SRNet第8~12层的合理性，以及引入SE通道注意力机制的必要性，与SRNet原模型进行了对比实验。我们统计了不同模型的参数量、计算浮点数（GFlops）、访存量（MB）、在MagicEXIF_basic上的测试准确率、在NVIDIA RTX 2080Ti GPU上每秒处理512×512分辨率的图像数（幅/秒）作为评价指标。表5给出了实验结果。

表5　针对网络结构的消融实验结果

Tab. 5　Results of ablation experiments targeting network structures

网络	参数量	计算量	访存量	ACC	幅/秒
SRNet	4778114	21.46	1350	0.986	63
SRNet_L7	**34498**	**9.03**	**947**	0.983	**183**
SRNet_L7_cbam	40370	9.21	968	0.988	57
SRNet_L7_se	36592	9.09	949	**0.988**	121

在表5中，SRNet表示由Jessica提出的原始网络模型，SRNet_L7表示裁剪第8~12的下采样层，SRNet_L7_cbam表示添加了CBAM注意力机制，SRNet_L7_se表示添加SE通道注意力机制，即本文所使用的模型。从结果可以看出，通过裁剪原网络第8~12层，参数量减少了99.3%，计算浮点数减少了57.9%，但模型轻量化的同时也带来了准确率的降低。因此使用轻量级的注意力机制进行弥补，对裁剪后的模型使用CBAM注意力机制和SE通道注意力机制，分类结果均提升了0.5%，比原始的SRNet提升了0.2%，也说明了空间注意力机制在该任务中作用不明显，仅用SE通道注意力机制在更轻量级的参数下达到相同的提升效果。最终的模型在NVIDIA RTX 2080Ti GPU上的测试速度较原模型提升了92%。通过网络层的删减以及注意力机制的添加，实现了更轻量更快速更准确的检测性能。

4.2.2　知识蒸馏方法的影响

为验证知识蒸馏方法对所设计的轻量级算法性能的提升效果，我们探究了不同的温度系数T和权重系数α对实验结果的性能，在MagicEXIF_basic数据库上分别训练“教师网络”、未使用和使用知识蒸馏技术的“学生网络”模型，实验结果如表6所示。

从实验结果可以看出，当温度系数T选取2，权

表 6 针对不同温度系数和权重系数的实验结果

Tab. 6 Experimental results for different temperature coefficients and weight coefficients

网络	温度系数 T	权重系数 α	ACC
Visual Transformer	-	-	99.5%
本文模型	-	-	98.8%
本文模型-KD	1	0.1	98.9%
	1	0.2	99.1%
	2	0.1	99.0%
	2	0.2	**99.2%**
	4	0.1	99.1%
	4	0.2	99.1%

重系数选取值为0.2时效果最好。基于thop库提供的函数参数量测量，“教师网络”Visual Transformer网络参数量为52M，而本文提出的轻量级网络作为“学生网络”仅0.032M，参数量仅需其千分之一，采用知识蒸馏技术优化的损失方法，“学生网络”取得了与“教师网络”相近的性能，与未经过知识蒸馏的原始网络相比取得了更好的效果，这证明了所提出的轻量级网络与知识蒸馏算法相结合的有效性。

4.2.3 不同颜色分量影响

为分析颜色分量对重构检测的影响，对RGB颜色空间及YCbCr颜色空间的各个分量进行分析。由于每个颜色分量包含的信息不同，我们考察了单一分量及混合分量对任务性能的影响，并从中选出性能最佳的组合，实验结果如表7所示。

表 7 针对不同颜色分量的消融实验结果

Tab. 7 Results of ablation experiments targeting different color components

通道	MagicEXIF_basic			MagicEXIF_extra		
	TPR	TNR	ACC	TPR	TNR	ACC
Y	0.975	0.980	0.978	0.972	0.971	0.971
Cb	0.956	0.954	0.954	0.942	0.947	0.943
Cr	0.945	0.953	0.949	0.927	0.912	0.921
CbCr	0.523	0.617	0.573	0.511	0.586	0.564
RGB	0.975	0.978	0.977	0.968	0.964	0.965
YCbCr	**0.989**	**0.986**	**0.988**	**0.985**	**0.973**	**0.979**

由表7可见，YCbCr三个分量共同作为输入时模型性能最佳。YCbCr颜色空间在MagicEXIF_basic数据库的训练准确率达到了0.988，比常规的RGB颜色空间提升了0.011，提升幅度为1.2%，在MagicEXIF_extra跨库测试集上准确率达到了0.979，比RGB颜色空间提升了0.014，提升幅度为1.5%。同时观察到单一的Y分量训练在验证集上已有较高的性能，甚至超越了RGB三分量的得分，说明Y通道信息对重构任务更加的敏感。同时跨库结果说明，YCbCr在不同场景、不同型号的拍摄设备得到的图像中拥有更高的准确率，YCbCr比RGB拥有更高的稳定性和更好的泛化能力。

4.3 重构检测性能与比较

在该部分实验中，我们将考察在常规情况下（测试设备的型号可知，包含在训练集中）模型的检测性能，并验证其对不可知设备及不同版本重构软件的泛化能力。

4.3.1 对已知设备的检测性能

在本实验中，我们将MagicEXIF_basic数据库按8:2比例随机划分为训练集和测试集，对比不同方法在MagicEXIF_basic数据库上的检测性能，测试集与训练集同源的方法视为对于可知设备的检测实验，实验结果如表8所示。可以看出，在MagicEXIF_basic的测试集中，各方法均达到了较好的效果。基于深度学习的方法比传统的机器学习方法效果更好，隐写分析网络Zhu Net和本文所用方法比主流轻量级神经网络在该任务上取得更好的效果，证明将隐写分析网络应用于该任务的有效性。本文所用方法在测试指标上与所有方法的对比中取得了最佳效果，在经过知识蒸馏训练策略后，模型的重构检测性能得到了进一步的提升。

表 8 对可知设备的实验结果

Tab. 8 Experimental results on known devices

	TPR	TNR	ACC
MobileNetV2[42]	0.973	0.971	0.972
EfficientNetB0[34]	0.981	0.989	0.984
SRM+SVM[43-44]	0.959	0.980	0.969
Zhu Net[45]	0.988	0.987	0.987
Multi-domain Net[26]	0.950	0.924	0.942
本文模型	0.989	0.987	0.988
本文模型-KD	**0.994**	**0.991**	**0.992**

4.3.2 对未知设备的泛化性能

在实际应用场景中，照相设备（特别是手机）更新换代很快，将所有手机品牌、手机型号都收集加入到训练集中是不现实的。为验证模型对于未知

设备的泛化检测能力，我们设置了MagicEXIF_extra数据库，用于验证模型分别在跨品牌、跨型号情景下的性能。实验结果如表9、表10所示。实验结果中模型均是基于MagicEXIF_basic训练集数据进行训练，原始代表在MagixEXIF_basic测试集上得分，并以此作为基准，括号数据表示跨型号、跨品牌实验与原始值的差异。

表9　跨型号实验结果

Tab. 9　Cross model experimental results

	TPR	TNR	ACC
MobileNetV2[42]	0.961 (−0.012)	0.964 (−0.007)	0.962 (−0.010)
EfficientNetB0[34]	0.975 (−0.006)	0.982 (−0.007)	0.977 (−0.007)
SRM+SVM[43-44]	0.950 (−0.009)	0.968 (−0.012)	0.959 (−0.010)
Zhu Net[45]	0.975 (−0.013)	0.971 (−0.016)	0.974 (−0.013)
Multi-domain Net[26]	0.934 (−0.016)	0.913 (−0.011)	0.929 (−0.013)
本文模型	0.973 (−0.016)	0.985 **(−0.002)**	0.979 (−0.009)
本文模型-KD	**0.989** **(−0.005)**	**0.987** (−0.004)	**0.987** **(−0.005)**

（括号内容为跨库数值与训练数值的差值）

表10　跨品牌实验结果

Tab. 10　Cross model and cross brand experimental results

	TPR	TNR	ACC
MobileNetV2[42]	0.955 (−0.018)	0.947 (−0.024)	0.951 (−0.021)
EfficientNetB0[34]	0.969 (−0.012)	0.978 (−0.011)	0.973 (−0.011)
SRM+SVM[43-44]	0.944 (−0.015)	0.953 (−0.027)	0.948 (−0.021)
Zhu Net[45]	0.970 (−0.018)	0.960 (−0.018)	0.967 (−0.020)
Multi-domain Net[26]	0.929 (−0.021)	0.909 (−0.015)	0.924 (−0.018)
本文模型	0.970 (−0.019)	0.973 (−0.014)	0.971 (−0.018)
本文模型-KD	0.985 **(−0.009)**	0.982 **(−0.009)**	0.983 **(−0.009)**

（括号内容为跨库数值与训练数值的差值）

从表9可以看出，在同品牌跨型号的测试中，EfficientNetB0表现的综合性能最好，其次是本文所提出的模型，说明本文提出的模型结构具有较好的泛化能力。当采用知识蒸馏的方法进行训练后，在TPR和TNR指标与MagicEXIF_basic测试结果相比分别降低了0.005和0.004，准确率相比只降低了0.005，说明对同品牌跨型号设备拍摄的图像保持有较高的检测准确率。

从表10的结果可以看出，跨品牌测试的结果与同品牌跨型号的测试结果呈现相似的特点，但是在跨品牌的测试中，各个指标下降得更多，说明不同品牌的设备间存在着较大的差异从而导致模型有一定的精度损失。但无论是跨型号还是跨品牌的测试中，本文模型在各个指标上均体现了较小的损失。同时，采用知识蒸馏的训练方法，能够有效提高模型的泛化能力。最后，本文所用模型在经过知识蒸馏方法训练后，在同品牌跨型号和跨品牌的测试中在准确率均保持在0.98以上，说明所用的模型和方法对不同品牌、不同型号的设备拍摄的图像仍有较强的重构检测能力。

4.3.3　对不同软件版本的泛化性能和迁移学习

根据已有实验，不同元数据编辑器版本之间存在差异，并且在查阅软件更新日志后发现v1.05版本前后的原图重构使用的内置算法是不同的。因此，以MagicEXIF_v1.03作为低版本的代表，生成了MagicEXIF_old数据库，用于跨库测试及迁移学习。为了对比不同算法的泛化能力，使用4.3.1实验所示方法在MagicEXIF_basic数据库（用版本v1.10生成）分别进行训练，并在MagicEXIF_old数据库上进行跨库测试，实验结果如表11所示。

由表11可以看出，基于MagicEXIF_basic数据库训练的各个模型，在MagicEXIF_old数据库测试

表11　MagicEXIF_old数据库测试实验结果

Tab. 11　MagicEXIF_ old dataset testing experimental results

	TPR	TNR	ACC
MobileNetV2[42]	0.021	0.974	0.498
EfficientNetB0[34]	0.043	**0.995**	0.519
SRM+SVM[43-44]	0.115	0.981	0.548
Zhu Net[45]	0.118	0.986	0.552
Multi-domain Net[26]	0.107	0.959	0.528
本文模型	0.128	0.989	0.561
本文模型-KD	**0.133**	0.992	**0.567**

的 TPR 均达到了 0.9 以上，说明对未重构图像的识别准确率仍保持有较高的水平，但是 TNR 只有 0.2 不到，对于低版本的 MagicEXIF 元数据编辑器的原图重构工具检测效果都比较差，进一步说明了实际应用场景中对模型进行迁移学习的必要性。尽管各个模型在 MagicEXIF_old 数据库上的重构准确率都很低，但所用方法在对比方法中仍取得了最优效果，证明了本文中提出方法的有效性。

为了提升模型对低版本重构工具的检测能力，我们考察了不同迁移学习策略对模型性能的影响。首先，我们使用本文提出的模型在 MagicEXIF_basic 数据库训练得到所需的预训练模型，然后对比不同的迁移学习方法的性能。方法一，仅使用 MagicEXIF_old 中 70% 的数据量对预训练模型进行迁移学习；方法二，在方法一所用数据的基础上加入等量的 MagicEXIF_basic 数据库图像共同进行迁移学习。对两种情况下训练得到的模型，将 MagicEXIF_basic 和 MagicEXIF_old 中的未训练数据作为测试集输入测试，所得结果如表 12 所示。

表 12 不同迁移学习方法实验结果

Tab. 12 Experimental results of different transfer learning methods

	MagicEXIF_basic			MagicEXIF_old		
	TPR	TNR	ACC	TPR	TNR	ACC
预训练	**0.989**	**0.987**	**0.988**	0.128	0.989	0.561
仅 old	0.723	0.974	0.859	**0.989**	**0.987**	**0.987**
混合	0.977	0.982	0.979	0.949	0.977	0.964

由表 12 中结果可以看出，仅使用 MagicEXIF_old 进行迁移学习，对于 old 数据库的检测性能有了比较明显的提升，但同时原模型的性能下降明显，在 MagicEXIF_basic 数据库中准确率相比下降了 13%。通过在 MagicEXIF_old 数据库基础上添加等量的 MagicEXIF_basic 图像数据，在尽可能保留原模型性能的前提下，对 MagicEXIF_old 重构图像的准确率从 0.128 提升到了 0.949，实现了模型对 MagicEXIF 原图重构检测的新旧版本支持，达到了更佳的综合性能，提高了模型的实用性。

据 MagicEXIF 元数据编辑器官网更新版本记载，自 MagicEXIFv1.03 起加入了原图重构工具，到目前已经更新了多个版本。本文根据网上可供下载的 MagicEXIF 元数据编辑器，分别安装了 v1.03、v1.08、v1.09 和 v1.10 四个不同版本，在 MagicEXIF_extra 的非重构数据库中随机挑选封面和风景场景各 50 幅，得到 100 幅原始图像。基于不同版本的 MagicEXIF 元数据编辑器进行重构操作，用于交叉测试不同版本之间的检测性能。上述实验已证明混合数据进行迁移学习方法的有效性，我们基于不同的模型采用表 13 中混合数据的迁移学习方法进行训练，在不同的版本中进行交叉测试，测试指标以 ACC 代表模型对不同版本的重构图像的检测准确率，实验结果如表 13 所示。

表 13 迁移学习后对不同版本的重构检测实验结果

Tab. 13 Experimental results of reconstruction detection of different versions after transfer learning

	V1.03	V1.08	V1.09	V1.10
MobileNetV2[42]	0.86	0.92	0.93	0.93
EfficientNetB0[34]	0.90	0.96	0.94	**0.98**
Zhu Net[45]	0.91	0.93	0.92	0.95
Multi-domain Net[26]	0.87	0.91	0.91	0.92
本文模型	**0.94**	**0.97**	**0.96**	0.97

从表 13 可以看出，各个模型经过混合数据的迁移学习后，对不同版本的重构图像检测准确率都有一定的提升。对于旧版本 V1.03，本文模型取得了 0.94 的最高准确率。对于 V1.08、V1.09 和 V1.10 三个新版本生成的重构图像，EfficientNet 和本文所提出的网络综合表现较好，本文模型在经过迁移学习后对于三个版本的重构图像检测准确率均达到了 0.96 以上，优于其他对比方法。

4.4 算法效率及数据规模对比实验

在该部分实验中，我们将考察本文所提出的算法与主流相关方法在检测效率方面（网络参数、CPU/GPU 上实际检测速度）进行对比，并考察不同训练数据规模对于网络性能的影响。

4.4.1 不同算法之间的效率对比

为了进一步对比所用算法与其他轻量级网络之间的差异，我们使用 thop 库提供的 profile 函数来计算统计不同网络所用的参数量、内存占用和浮点数。结果如表 14 所示。可以看出，由于 Multi-domain Net 采用 64×64 作为输入分辨率，在内存占用和浮点数参数上明显少于其他以 512×512 作为输入分辨率的网络。在相同分辨率输入的模型对比中，所用网络在参数量上占据了极大的优势，与主流的轻量级神经网络相比，参数量仅需其 1% 即可达到较好的效果，说明所用模型结构在嵌入式设备

表 14　参数量、内存占用和浮点数对比

Tab. 14　Comparison of parameter amount, memory usage, and floating point number

	参数量	内存	浮点数
MobileNetV2[42]	2.29M	389M	1.67G
EfficientNetB0[34]	16.7M	996M	8.01G
Zhu Net[45]	2.89M	419M	4.84G
Multi-domain Net[26]	3.10M	35.6M	0.36G
本文模型	0.032M	400M	5.03G

上部署具有较大的优势。在内存和浮点数上与其他网络相比差别不大。

为了更符合实际场景，我们对比了不同网络在CPU和GPU的对于不同分辨率的推理速度。对于在CPU上的测试实验，为了更好地观察实验性能，我们将所训练的模型转换为onnx格式再进行推理。对于在GPU上的测试实验，我们在单张NVIDIA RTX 2080Ti GPU进行，对比不同网络在不同分辨率下每秒处理的图像数量。由于Multi-domain Net需要固定输入分辨率，因此不作为本次实验的对比方法，实验结果如表15所示。

从表15可以看出，MobileNetV2相较于其他网络而言优势较为明显，本文所用模型在实际推理速度中与EfficientNetB0相近。实验总结，所用模型在推理速度上与EfficientNet等轻量级网络速度相近，但其所需的参数量极小，故模型所占内存空间相对而言更小，在移动端部署占据一定的优势。

表 15　在CPU、GPU测试实验结果（幅/秒）

Tab. 15　Experimental results on CPU and GPU testing（pic/s）

	CPU			GPU		
	256	512	1024	256	512	1024
MobileNetV2[42]	121	29	5	1060	464	55
EfficientNetB0[34]	88	16	3	767	214	21
Zhu Net[45]	67	10	2	678	195	30
本文模型	72	11	3	729	223	26

4.4.2　训练数据规模对性能的影响

对于轻量级模型而言，由于其网络层数低、参数小的特点，网络的特征提取能力与深层网络相比而言较弱，因而依赖于大规模的训练数据，但在实际场景中由于数据获取难度大、成本高等原因，很可能出现数据不足的情况。为了探究训练数据规模对模型性能的影响，我们在MagicEXIF_basic数据库中随机挑选25%和50%的数据作为训练集，对比不同方法在较少数据量情况下的学习情况。实验结果如表16所示。从实验结果可以看出，当训练数据量减少为25%时，各个模型的性能都有一定程度的下降，本文所用网络在与其他深度学习网络的对比中取得了最佳的性能，在ACC指标上较明显领先于Zhu Net和MobileNetV2，略高于Effcient-NetB0。当训练数据量减少为50%时，各个深度网络之间的差距减小，本文模型在ACC指标对比中仍取得了领先。结果表明本文所用模型在少量的数据训练后仍达到了较高的水平，在经过知识蒸馏方法训练后性能得到了进一步提升。

表 16　不同规模数据量进行训练实验结果

Tab. 16　Training experiment results with different scale data volumes

	25%			50%		
	TPR	TNR	ACC	TPR	TNR	ACC
MobileNet V2[42]	0.943	0.948	0.944	0.970	0.958	0.963
EfficientNet B0[34]	0.937	**0.995**	0.968	0.985	0.989	0.986
Zhu Net[45]	0.931	0.969	0.953	0.966	0.983	0.972
Multi-domain Net[26]	0.812	0.739	0.837	0.895	0.837	0.883
本文模型	0.951	0.987	0.974	0.983	**0.994**	0.987
本文模型-KD	**0.969**	0.990	**0.980**	**0.989**	0.992	**0.991**

5　结论

针对图像元数据重构对图像原始性辨识带来的安全问题，本文提出了一种JPEG原图重构取证方法。通过分析原图重构的机理及不同颜色分量对重构检测的作用，本文建立了重构检测的基本依据。基于此，通过对SRNet进行网络裁剪、引入注意力机制、应用知识蒸馏的训练方法，构建了一个有效的重构检测模型。与原型网络相比，所提网络具有更轻量、更快速、更准确的特点。大量实验结果表明，本文方法与传统轻量级深度网络、传统隐写分析方法及基于深度学习的隐写分析方法相比，能取得更高的重构检测准确率，且在跨设备、跨重构软件版本等情形下均具有良好的泛化性。

为有效应对实际应用中的大规模检测任务，还需要进一步提高本文算法的检测性能。例如，实际图像分辨率可能远高于512×512，高分辨率有助于

提升检测正确率，但也会加重计算负担。如何通过优化网络来解决这一矛盾，将是未来需要继续研究的问题。

参考文献

[1] 谢皓，张健，倪江群. 数字图像操作取证综述[J]. 信号处理，2021，37(12)：2323-2337.
XIE Hao，ZHANG Jian，NI Jiangqun. A survey of digital image operation forensics[J]. Journal of Signal Processing，2021，37(12)：2323-2337.（in Chinese）

[2] 李晓龙，俞能海，张新鹏，等. 数字媒体取证技术综述[J]. 中国图象图形学报，2021，26(6)：1216-1226.
LI Xiaolong，YU Nenghai，ZHANG Xinpeng，et al. Overview of digital media forensics technology[J]. Journal of Image and Graphics，2021，26(6)：1216-1226.（in Chinese）

[3] FERRARA P，BIANCHI T，DE ROSA A，et al. Image forgery localization via fine-grained analysis of CFA artifacts[J]. IEEE Transactions on Information Forensics and Security，2012，7(5)：1566-1577.

[4] CHEN Can，MCCLOSKEY S，YU Jingyi. Image splicing detection via camera response function analysis[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition（CVPR）. Honolulu，HI，USA. IEEE，2017：1876-1885.

[5] KE Yongzhen，QIN Fan，MIN Weidong，et al. Exposing image forgery by detecting consistency of shadow[J]. The Scientific World Journal，2014，2014：1-9.

[6] ZHANG Wei，CAO Xiaochun，ZHANG Jiawan，et al. Detecting photographic composites using shadows[C]// 2009 IEEE International Conference on Multimedia and Expo. New York，NY，USA. IEEE，2009：1042-1045.

[7] CHATTERJEE S，GARAIN U，KHAN F，et al. Compressed digital image forensics using zigzag scanned coefficients[J]. International Journal of Computer Applications，2017，173：17-27.

[8] HUANG Xiaosa，WANG Shilin，LIU Gongshen. Detecting double JPEG compression with same quantization matrix based on dense cnn feature[C]//2018 25th IEEE International Conference on Image Processing（ICIP）. Athens，Greece. IEEE，2018：3813-3817.

[9] WANG Jinwei，WANG Hao，LI Jian，et al. Detecting double JPEG compressed color images with the same quantization matrix in spherical coordinates[J]. IEEE Transactions on Circuits and Systems for Video Technology，2020，30(8)：2736-2749.

[10] BAKAS J，RAMACHANDRA S，NASKAR R. Double and triple compression-based forgery detection in JPEG images using deep convolutional neural network[J]. Journal of Electronic Imaging，2020，29(2)：023006.

[11] RHEE K H. Detection of JPEG compression forensics[J]. Journal of the Institute of Electronics and Information Engineers，2017，54(11)：107-112.

[12] KEE E，JOHNSON M K，FARID H. Digital image authentication from JPEG headers[J]. IEEE Transactions on Information Forensics and Security，2011，6(3)：1066-1075.

[13] BIANCHI T，PIVA A. Image forgery localization via block-grained analysis of JPEG artifacts[J]. IEEE Transactions on Information Forensics and Security，2012，7(3)：1003-1017.

[14] AMERINI I，BALLAN L，CALDELLI R，et al. A SIFT-based forensic method for copy-move attack detection and transformation recovery[J]. IEEE Transactions on Information Forensics and Security，2011，6(3)：1099-1110.

[15] KWON M J，YU I J，NAM S H，et al. CAT-net：Compression artifact tracing network for detection and localization of image splicing[C]//2021 IEEE Winter Conference on Applications of Computer Vision（WACV）. Waikoloa，HI，USA. IEEE，2021：375-384.

[16] BI Xiuli，WEI Yang，XIAO Bin，et al. RRU-net：The ringed residual U-net for image splicing forgery detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops（CVPRW）. Long Beach，CA，USA. IEEE，2020：30-39.

[17] ZHU Ye，CHEN Chaofan，YAN Gang，et al. AR-net：Adaptive attention and residual refinement network for copy-move forgery detection[J]. IEEE Transactions on Industrial Informatics，2020，16(10)：6714-6723.

[18] KHASHANDARAG A S，EBRAHIMIAN N. A new method for color image steganography using SPIHT and DFT，sending with JPEG format[C]//Proceedings of the 2009 International Conference on Computer Technology and Development-Volume 01. New York：ACM，2009：581-586.

[19] 陈超. JPEG图像被动取证的研究[D]. 广州：中山大学，2015.
CHEN Chao. Research on passive forensics of JPEG images[D]. Guangzhou：Sun Yat-sen University，2015.（in Chinese）

[20] 卢启萌，施少培. Exif信息在数码照片真实性鉴定中的应用[J]. 中国司法鉴定，2012(5)：86-90.
LU Qimeng，SHI Shaopei. The application of exif in digital photo authentication[J]. Chinese Journal of Forensic Sciences，2012(5)：86-90.（in Chinese）

[21] 邢文博，杜志淳. JPEG图像文件头取证[J]. 计算机时代，2019(10)：11-15，18.
XING Wenbo，DU Zhichun. JPEG image file header forensics[J]. Computer Era，2019(10)：11-15，18.（in Chinese）

[22] BOROUMAND M，CHEN Mo，FRIDRICH J. Deep residual network for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security，2019，14(5)：1181-1193.

[23] TAUBMAN D S，MARCELLIN M W. JPEG2000：Image Compression Fundamentals，Standards，and Practice[M]. Boston：Kluwer Academic Publishers，2002.

[24] POPESCU A C，FARID H. Statistical tools for digital forensics[C]//Fridrich J. International Workshop on Information Hiding. Berlin，Heidelberg：Springer，2004：128-147.

[25] HUANG Fangjun，HUANG Jiwu，SHI Yunqing. Detecting double JPEG compression with the same quantization matrix[J]. IEEE Transactions on Information Forensics and Security，2010，5(4)：848-856.

[26] ZENG Ximei，FENG Guorui，ZHANG Xinpeng. Detection of double JPEG compression using modified DenseNet model[J]. Multimedia Tools and Applications，2019，78(7)：8183-8196.

[27] FRIDRICH J，KODOVSKY J. Rich models for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security，2012，7(3)：868-882.

[28] HE Kaiming，ZHANG Xiangyu，REN Shaoqing，et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition（CVPR）. Las Vegas，NV，USA. IEEE，2016：770-778.

[29] RUAN Feng，ZHANG Xing，ZHU Dawei，et al. Deep learning for real-time image steganalysis：A survey[J]. Journal of Real-Time Image Processing，2020，17(1)：149-160.

[30] ZHU Xizhou，CHENG Dazhi，ZHANG Zheng，et al. An empirical study of spatial attention mechanisms in deep networks[C]//2019 IEEE/CVF International Conference on Computer Vision（ICCV）. Seoul，Korea（South）. IEEE，2020：6687-6696.

[31] HU Jie，SHEN Li，SUN Gang. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City，UT，USA. IEEE，2018：7132-7141.

[32] WOO S，PARK J，LEE J Y，et al. CBAM：convolutional block attention module[C]//European Conference on Computer Vision. Cham：Springer，2018：3-19.

[33] HOWARD A G，ZHU Menglong，CHEN Bo，et al. MobileNets：Efficient convolutional neural networks for mobile vision applications[EB/OL]. 2017：arXiv：1704.04861. https://arxiv.org/abs/1704.04861.

[34] TAN Mingxing，LE Q V. EfficientNet：Rethinking model scaling for convolutional neural networks[EB/OL]. 2019：arXiv：1905.11946. https://arxiv.org/abs/1905.11946.

[35] HINTON G，VINYALS O，DEAN J. Distilling the knowledge in a neural network[EB/OL]. 2015：arXiv：1503.02531. https://arxiv.org/abs/1503.02531.

[36] DOSOVITSKIY A，BEYER L，KOLESNIKOV A，et al. An image is worth 16x16 words：Transformers for image recognition at scale[EB/OL]. 2020：arXiv：2010.11929. https://arxiv.org/abs/2010.11929.

[37] 包新月，俞磊. 计算机数字图像处理常用颜色空间及转换[J]. 电子技术与软件工程，2021(7)：122-123.
BAO Xinyue，YU Lei. Common color space and conversion in computer digital image processing[J]. Electronic Technology & Software Engineering，2021(7)：122-123.（in Chinese）

[38] HEMALATHA S，DINESH ACHARYA U，RENUKA A. Comparison of secure and high capacity color image steganography techniques in RGB and YCbCr domains[J]. International Journal of Advanced Information Technology，2013，3(3)：1-9.

[39] KOJU R，JOSHI S R. Comparative analysis of color image watermarking technique in RGB，YUV，and YCbCr color channels[J]. Nepal Journal of Science and Technology，2015，15(2)：133-140.

[40] KHALILI M，ASATRYAN D. Effective digital image watermarking in YCbCr color space accompanied by presenting a novel technique using DWT[EB/OL]. 2012：arXiv：1206.4520. https://arxiv.org/abs/1206.4520.

[41] 车冬娟，周建伟，王健，等. JPEG有损压缩图像保留格式加密研究[J]. 北华航天工业学院学报，2022，32(1)：4-6.
CHE Dongjuan，ZHOU Jianwei，WANG Jian，et al. Preserving format encryption of JPEG lossy compressed image[J]. Journal of North China Institute of Aerospace Engineering，2022，32(1)：4-6.（in Chinese）

[42] SANDLER M，HOWARD A，ZHU Menglong，et al. MobileNetV2：inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City，UT，USA. IEEE，2018：4510-4520.

[43] 黄瑞灵. 对抗环境下图像原始性的辨识[D]. 深圳：深圳大学，2020.

HUANG Ruiling. Identification of image primitiveness in confrontation environment [D] Shenzhen: Shenzhen University, 2020. (in Chinese)

[44] LI Haodong, LUO Weiqi, QIU Xiaoqing, et al. Identification of various image operations using residual-based features[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(1): 31-45.

[45] ZHANG Ru, ZHU Feng, LIU Jianyi, et al. Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 1138-1150.

[46] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile. IEEE, 2016: 1026-1034.

[47] LOSHCHILOV I, HUTTER F. SGDR: Stochastic gradient descent with warm restarts[EB/OL]. 2016: arXiv: 1608.03983. https://arxiv.org/abs/1608.03983.

作者简介

黎溢轩 男，1998 年生，广东东莞人。深圳大学电子与信息工程学院硕士研究生，主要研究方向为图像取证、图像信息安全。
E-mail：2070436055@szu.edu.cn

李昊东 男，1990 年生，广东中山人。深圳大学助理教授、硕士生导师，主要研究方向为多媒体取证与安全、图像及音视频处理、模式识别、机器学习。
E-mail：lihaodong@szu.edu.cn

曾吉申 男，1993 年生，广东揭阳人。深圳大学电子与信息工程学院博士研究生，目前工作于阿里巴巴集团。主要研究方向为图像取证、数字水印、图像隐写分析、多媒体信息安全。
E-mail：jishen.zjs@alibaba-inc.com

黄瑞灵 男，1995 年生，广东梅州人。深圳大学电子与信息工程学院硕士研究生。主要研究方向为图像取证、图像信息安全。
E-mail：875054756@qq.com

黄继武 男，1962 年生，广东揭阳人。深圳大学特聘教授和博士生导师。主要研究方向为信息隐藏、图像/音频/视频信号处理、模式识别、机器学习。
E-mail：jwhuang@szu.edu.cn

（责任编辑：边熙淳）