arXiv:2302.02276v1 [cs.MM] 5 Feb 2023

基于隐写特征增强与图注意力学习的 JPEG 隐写分析

Qiyun Liu、Zhiguang Yang 和 Hanzhou Wu*

2023年2月7日

摘要

图像隐写分析旨在判断载体图像是否包含隐藏信息。由于 JPEG 是社交网络中最常用的图像格式，因此，对 JPEG 图像进行隐写分析也成为最迫切需要研究的课题。然而，为了检测 JPEG 图像中是否隐藏了秘密信息，现有的大多数算法都结合了流行的计算机视觉相关网络进行设计，却没有考虑图像隐写分析中出现的关键特性。至关重要的是，在表示学习过程中，作为极弱信号的隐写信号能够得到增强。受此启发，本文提出了一种用于 JPEG 隐写分析的新型表示学习算法，主要由图注意力学习模块和特征增强模块组成。图注意力学习模块旨在避免因卷积神经网络的局部特征学习而造成的全局特征损失，以及依赖堆叠深层网络来扩展感知域的问题。特征增强模块用于防止卷积层堆叠削弱隐写信息。此外，本文采用基于大规模数据集进行预训练的方式初始化网络权重，从而增强网络提取判别性特征的能力。我们建议使用 ALASKA2 对在 BOSSBase+BOWS2 上训练的模型进行预训练。实验结果表明，所提出的算法在检测准确率方面优于以往方法，验证了所提方法的优越性和适用性。

1 引言

作为信息隐藏的重要分支， [1], 隐写术被称为一种艺术，即在不显著扭曲载体信号的情况下，将秘密信息嵌入载体信号中，例如数字图像和视频。隐写术具有隐蔽性，因此可用于有效解决数字通信中的安全问题。一般来说，任何类型的数字信号都可以用作隐写术的载体。然而，从应用角度来看，数字图像是最常用的隐写载体，因为它易于编辑，且具有大量可用于容纳秘密信息的冗余空间。因此，现有的大多数隐写方法都是针对数字图像设计的 [2]。数字图像有多种格式，其中 JPEG 是社交网络上最流行的图像格式，因为它在保持较小存储空间的同时提供了较高的视觉质量。将 JPEG图像用于隐写术，可以借助大量正常的社交活动轻松掩盖隐写通信的存在，从而使 JPEG隐写成为热门研究领域。

作为隐写术的对抗技术，隐写分析也得到了广泛研究，旨在揭示数字图像中是否存在隐藏信息。现有的大多数隐写分析方法都将数字图像中隐藏信息的检测建模为二分类问题，即判断给定图像是否包含秘密信息。JPEG隐写的快速发展，使JPEG隐写分析成为亟待探索的研究方向。早期的JPEG隐写分析方法 [3, 4, 5] 遵循传统的机器学习框架，需要人工设计统计特征，并可能采用集成策略。近年来，随着深度卷积神经网络（CNNs）的普及，越来越多的研究 [6, 7, 8, 9] 致力于将CNNs在计算机视觉领域取得的成功应用于图像隐写分析。一般而言，基于CNN的图像隐写分析包含三个

*通讯作者，电子邮件：h.wu.phd@ieee.org

阶段 [10]。首先，从给定图像中提取一定数量的噪声残差。然后，根据这些噪声残差确定一组高维特征向量。最后，利用二分类器将图像分类为载体图像或含密（即包含秘密信息的图像）。得益于CNNs强大的学习能力，上述三个阶段可以通过端到端的方式实现。在图像噪声残差的提取方面，许多基于CNN的图像隐写分析器会利用先验信息引导模型收敛 [9, 11, 12, 13]。例如，Huang等人 [13] 最近提出了一种隐写分析网络，将选择通道应用于JPEG图像的空间域。Boroumand等人 [14] 提出了SRNet，以降低网络对先验信息的依赖。该架构将大幅扩展的前端与多个残差连接相结合，无需传统的人工设计即可提取噪声残差，从而取得了出色的检测性能。

在探索深度学习用于隐写分析的适用性时，有必要考虑隐写原理。空间域隐写会直接改变原始空间域中的图像像素。在数据嵌入过程中，不同图像中待嵌入位置的像素各不相同。擅长挖掘局部模式的神经网络可能会记住特定的嵌入模式，最终损害训练模型的泛化能力。相比之下，JPEG隐写通常通过修改量化 DCT 系数，将秘密信息嵌入 DCT域。变换回空间域后，DCT 系数的修改会扩展到相应 8×8 块中的所有空间像素。通过学习空间域中不同块之间的统计不一致性，JPEG隐写可能比空间域隐写更容易被检测到。鉴于图神经网络（GNNs） [15] 具有对对象间统计依赖关系进行建模的强大能力，我们希望进一步探索图表示学习在JPEG隐写分析中的适用性。图神经网络近年来蓬勃发展，因为它们在处理非欧几里得数据方面具有优势，越来越多的研究人员也开始将其应用于视觉任务 [16]。由于图表示学习能够将图像建模为完全图以进行全局分析，因此可以更好地把握整体统计特征及其关联。我们的先前工作已证明，图表示学习可以与空间图像隐写分析相结合 [10]，因此亟需进一步研究其在JPEG隐写分析中的适用性。

以往基于卷积神经网络的研究针对适用于 JPEG隐写分析 的网络结构进行了一系列探索。尽管这些研究融入了现有的深度学习设计理念，却忽略了图像隐写分析所涉及的关键因素。网络加深过程中出现的特征消失问题一直未受到重视。在图像隐写分析中，待识别的隐写特征极其微弱，而卷积和池化也会抑制微弱信号。网络中连续堆叠的卷积与池化层会逐层削弱微弱的隐写特征，其中一些甚至可能消失。另一方面，卷积神经网络通过增加深度，扩展卷积核局部感知的感受野。在这一过程中，不可避免地会丢失一些全局信息，这也可以理解为整幅图像中所有信息之间的相互关系。我们认为，以上两个因素是现有隐写分析模型性能受限的主要原因。

基于上述分析，我们认为，设计能够更好地利用隐写特征的网络结构时，需要考虑两个关键问题。第一，如何防止微弱的隐写信号被网络本身抑制；第二，如何利用隐写信号的全局特征。为此，本文开发了一种结合卷积神经网络与图神经网络优势的全新 JPEG隐写分析 深度学习框架。针对第一个问题，我们采用特征增强模块作为基本单元，移除池化层，并将卷积的步幅降至 1，以更好地增强微弱的隐写信号。为了解决如何利用全局特征的问题，我们将图像输入图注意力网络，以学习像素级注意力系数。这种设计能够从全局视角进一步有效提取残差特征向量，并有效学习隐写特征。此外，通过使用大规模数据集对网络权重进行初始化的预训练，可以加快模型收敛，并增强网络提取判别性特征的能力。我们建议使用 ALASKA2 对采用 BOSSBase+BOWS2 训练的模型进行预训练。实验表明，在不同载荷下，本文提出的方法在检测准确率方面优于以往方法。

本文其余结构安排如下：第2节详细介绍提出的隐写分析方法，随后在第3节给出广泛的实验结果与分析。最后，我们在第4节总结本文并进行讨论。

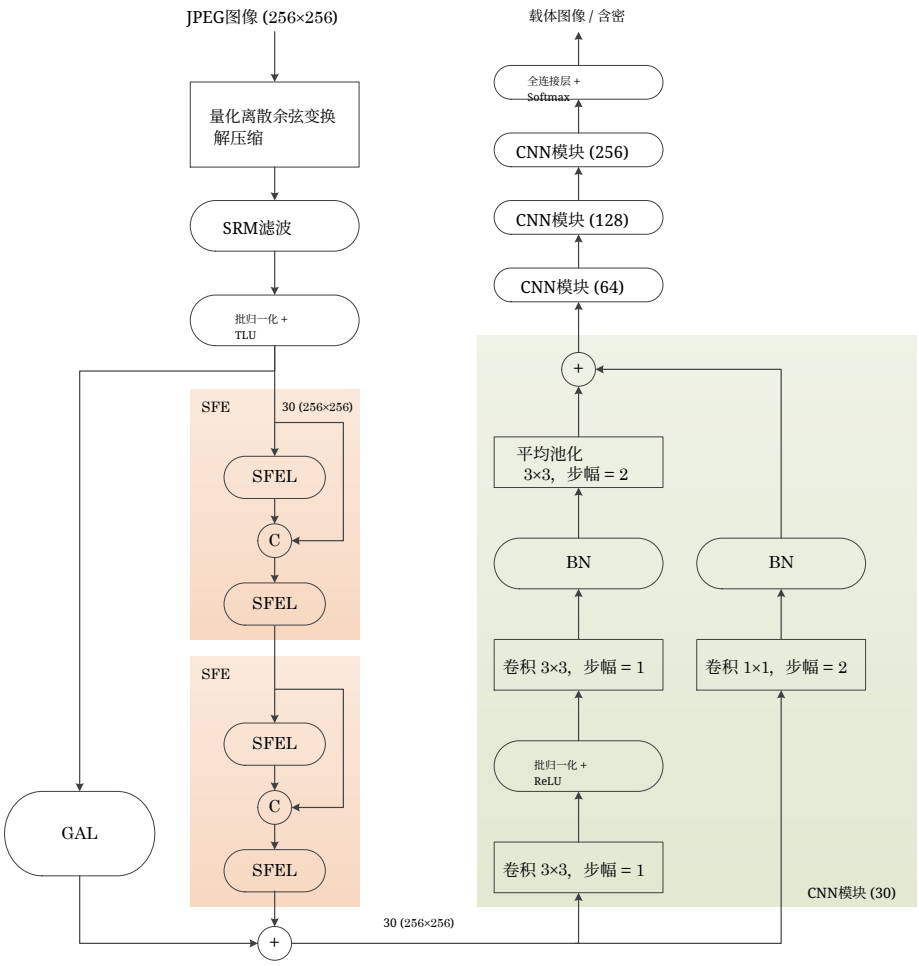

图1：提出的网络架构 JPEG-GraphNet 示意图。

2 提出的方法

2.1 概述

所提出的网络架构被命名为 JPEG-GraphNet，因为该架构利用图表示学习进行 JPEG图像隐写分析。如图1 所示，JPEG-GraphNet 的整体流程包括三个阶段，即特征提取（左侧）、特征学习（右下方）和特征分类（右上方）。特征提取的目的是提取原始隐写特征，并进一步增强这些特征，以便在特征学习阶段进行深度表示学习。特征学习的目的是为特征分类阶段提取判别性特征，其目标是判断输入是否为含密图像。

具体而言，特征提取由三个模块组成，即预处理、隐写特征增强（SFE）和图注意力学习（GAL）。预处理步骤从输入 JPEG图像的解压缩版本中提取隐写残差。随后，将类噪声残差输入 SFE 模块进行局部特征增强，并输入 GAL模块进行全局特征增强。两种增强后的特征将合并，并输入后续的特征学习模块；该模块由四个结构相似的 CNN模块组成。特征学习模块用于将接收到的隐写特征表示为高维特征。最后一步是特征分类：高维特征经过一个配备 softmax函数的全连接层，最终输出用于表示图像分类结果的独热向量。下文将详细介绍 JPEG-GraphNet。

2.2 特征提取

2.2.1 预处理

通过预处理将领域知识引入 JPEG-GraphNet，该预处理包含量化 DCT 解压缩、SRM 高通滤波和专用激活函数。从数学上讲，给定一幅尺寸为 $h \times w$ 的 JPEG图像，其中 h 和 w 均为 8 的倍数，在我们的实验中二者均为 256。设 C $=\{c_{i,j}^{n,m}\}$ 为量化 DCT 系数矩阵，其中 c^n、$_i{}^m$、$_j$ 是第 $(n,\ m)$ 个 8×8 块中的第 $(i,\ j)$ 个元素，$0 \le i$、$j < 8$、$0 \le n < h/8$，且 $0 \le m < w/8$。设 Q $=\{q_{i,j}^{n,m}\}$ 为相应的量化矩阵。量化前 C 的 DCT 系数矩阵可近似为 C 与 Q 的逐元素乘积，即 D $=\{d_{i,j}^{n,m}\}$，其中 d^n、$_i{}^m$、$_j = c^n$、$_i{}^m$、$_j$ q^n、$_i{}^m$、$_j$。值得注意的是，对于任意 $n \ne n'$ 或 $m \ne m'$，通常有 q^n、$_i{}^m$、$_j = q^{n'}$、$_i{}^{m'}$、$_j$。为确定空间像素（即解压缩后的亮度通道），需执行逆 DCT 变换，即，

$$f_{x,y}^{n,m} = \frac{1}{4}\sum_{u=0}^{7}\sum_{v=0}^{7} g_u g_v d_{u,v}^{n,m} \cos\frac{(2x+1)u\pi}{16}\cos\frac{(2y+1)v\pi}{16}, \tag{1}$$

其中，$g_z = 1/\sqrt{2}$ 当 $z = 0$ 时，$g_z = 1$ 否则。换言之，$\mathbf{C}$ 的空间亮度矩阵可表示为 $\mathbf{F} = \{f_{i,j}^{n,m}\}$。量化 DCT 解压缩的目标是确定 $\mathbf{F}$。为了根据 $\mathbf{F}$ 确定 $\mathbf{D}$，可以应用 DCT 变换，即，

$$d_{u,v}^{n,m} = \frac{1}{4} g_u g_v \sum_{x=0}^{7}\sum_{y=0}^{7} f_{x,y}^{n,m} \cos\frac{(2x+1)u\pi}{16}\cos\frac{(2y+1)v\pi}{16}. \tag{2}$$

因此，$\mathbf{C}$ 可以通过应用 c^n、$_i{}^m$、$_j = d^n$、$_i{}^m$、$_j$ $/q^n$、$_i{}^m$、$_j$ 来确定。已有大量研究表明，高通滤波对于图像隐写分析至关重要，有助于捕获隐写伪影。作为残差提取器，高通滤波能够抑制图像内容并提高隐写分析的噪声信号比。高通滤波可以通过一组可训练卷积核实现。虽然我们始终可以自行初始化可训练卷积核，但众所周知，如果采用随机初始化，低复杂度的卷积神经网络通常无法在相对较小的数据集上学习到良好的残差提取器。因此，我们建议以手工设计的方式初始化高通滤波卷积核。具体而言，本文采用 SRM（空间富模型的简称） [17]滤波策略，因其性能优越。具体来说，使用文献 5×5 中所述的总计 30 个尺寸为 [17, 14]的基本线性滤波器（即“spam”滤波器及其对称版本）来初始化卷积核。它们以$\mathbf{F}$ 为输入，并输出总计 30 个尺寸为 $h \times w$ $(\times 1)$ 的残差特征图。

神经网络中节点的激活函数根据输入定义该节点的输出。激活函数可以为神经网络引入非线性，从而大幅增强特征表示能力。激活函数有多种选择，例如双曲正切、Sigmoid函数和修正线性单元（ReLU） [18]。在传统的图像隐写分析中，人工设计的特征提取器通常具有某种对称性，例如，直方图分箱所在的区间关于零点对称。尽管这些特征会进一步融合，以实现高性能的隐写分析，但这一点启发我们在初始特征提取中引入特征对称性，以获得更好的特征学习效果。受此启发，我们希望在早期阶段使用的激活函数能够将输入映射到一个对称区间。此外，根据经验，将激活函数设计为硬曲线而非软曲线，更有利于特征学习。基于这一思路，我们建议在 JPEG-GraphNet 中使用一种称为截断线性单元（TLU）的激活函数，即在 SRM滤波之后应用 TLU。TLU 可表示为 [12]:

$$\mathrm{TLU}(x) = \begin{cases} -T, & x \in (-\infty, -T) \\ x, & x \in [-T, T] \\ T, & x \in (T, +\infty) \end{cases} \tag{3}$$

其中，T 是用于界定输出范围的预设阈值。此外，为加速模型收敛，默认采用批归一化（BN） [19] 。

2.2.2 隐写特征增强

为进一步增强预处理步骤生成的残差图中的含密信息，我们提出了一个隐写特征增强（SFE）模块，如图所示，该模块使用了两次

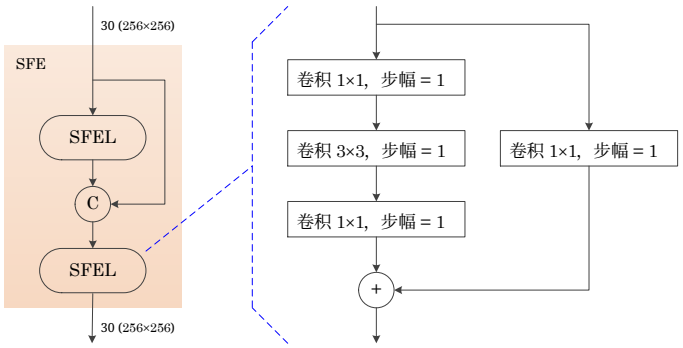

图2：SFE模块的结构信息。

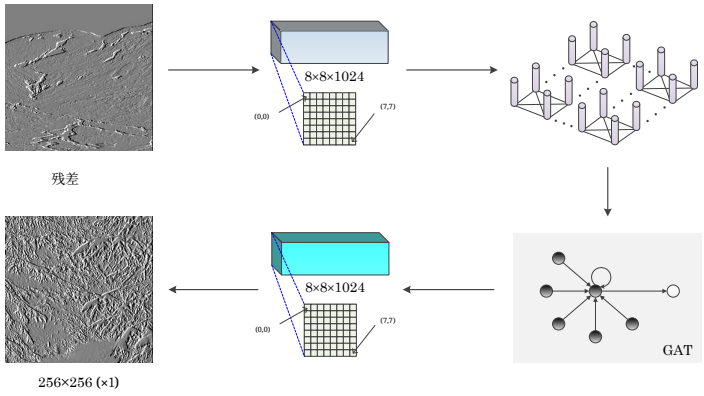

图3：GAL的详细信息，其中分辨率为 256×256，仅作示例。

1。SFE模块的详细结构如图2所示，从中可以看出引入了两个SFE层（SFEL）。SFEL的设计原理受到改进用于JPEG隐写分析的EfficientNet这一思路的启发 [20]：在早期CNN层中通过步幅或池化降低特征图分辨率，会抑制类似噪声的隐写信号，同时增强图像内容，从而对隐写分析检测准确率产生负面影响。因此，SFEL的设计遵循两项关键原则：一方面，移除池化以减少特征损失；另一方面，所有卷积层的步幅均设为1，并使用配备 1×1 卷积的捷径连接处理特征，从而尽可能保留浅层特征信息。

由于网络在浅层学习到的重要特征很容易在更深层中消失，为了防止浅层特征消失，一种简单的方法是随着网络逐渐加深而设置更多捷径连接。如果浅层特征在传递过程中消失，它们仍可通过捷径连接直接传递到更深层。我们在SFE中使用“concat”将浅层特征转化为深层特征，从而保留更多隐写信息。

2.2.3 图注意力学习

为了挖掘全局隐写信息以实现高效的隐写分析，我们引入了图注意力学习（GAL）模块，其细节如图3所示。具体而言，给定一幅 JPEG图像，首先对其进行预处理以生成一组残差。随后，将每个残差划分为互不重叠的8×8块，并叠加成一个维度为 $8 \times 8 \times (hw/64$的特征图。例如，如果图像的宽度和高度均为 256，则特征图的尺寸为 $8 \times 8 \times 1024$，如图 3所示。我们构建一个由 64 个节点组成的完全图，每个节点对应一个空间点。完全图中的每个节点还关联一个长度为 $hw/64$的特征向量。换言之，上述特征图将被拆分为总计$hw/64$ 个特征向量，并将每个特征向量分配给对应于同一空间点的节点

作为特征向量。通过这种方式，我们可以构建一个带有节点特征的完全图，其中节点数为 64，边数为 2016，且每个特征向量的维度为 $hw/64$。随后将该图输入图注意力网络进行进一步处理。尽管图注意力网络值得进一步研究，但为简化起见，本文采用 [21] 中介绍的图注意力网络，因为其性能优越。网络结构的详情请参见 [21] 。简而言之，该网络以关联一组节点特征的图作为输入，并输出一组新的节点特征作为结果。图注意力学习的层数设为 2。每个新特征向量的维度与输入特征向量的维度相同。最后，对所有这些特征向量进行处理，构成一个尺寸为 256×256的新特征图。通过堆叠新的特征图，输出将扩展至 30 个通道，并与 SFE 生成的特征图相加。

2.3 特征学习与分类

经过上述隐写特征处理后，输出进入特征学习模块。为降低特征图的维度，隐写特征学习层采用池化和步长为 2 的残差捷径。如图1右侧的 CNN 模块所示，采用 3×3，并在两层卷积层后进行步幅为 2 的平均池化；同时，残差捷径也采用核大小为 1、步幅为 2 的卷积层，将输出特征图的宽度和高度同时缩减至输入的一半。关于本文所用 CNN 模块的设计，读者可参阅 SRNet 的特征学习部分 [14], , 其中设计了两种用于降维的模块，即“type3”和“type4”。本文按照 type3 设计 CNN 模块。

我们使用四个 CNN模块逐步降低维度。如图1所示，括号中的数字表示该模块所有卷积层中的卷积核数量。值得强调的是，前三个 CNN模块的结构完全相同，而最后一个 CNN模块移除了残差捷径，并将末尾的平均池化替换为全局池化，从而将特征降至一维。最后一部分由一个标准的 256 维全连接层和一个 Softmax线性分类器组成，用于输出最终概率。

3 实验结果与分析

3.1 实验设置

实验所用的基础训练数据集来自两个图像源，即 BOSSbase v1.01[22] 和 BOWS2 [23], , 每个图像源均包含 10,000 张分辨率为 512×512的灰度图像。由于计算资源有限，所有涉及的图像均经过中心裁剪，分辨率调整为 256×256。训练集包含 14,000 对载体图像-隐写图像对（从 BOSSbase 中随机选取 4,000 对，从 BOWS2 中选取 10,000 对）。包含 1,000 对 BOSSbase 载体图像-隐写图像对的验证集用于选择超参数；最终测试结果则通过对测试集中的 5,000 对 BOSSbase 载体图像-隐写图像进行分类得出。我们在 J-UNIWARD [24] 和 UED-JC [25] 中评估所提出的隐写分析模型的性能，载荷范围为每个非零 AC DCT系数 0.1 至 0.5 比特（bpnzac），质量因子（QF）为 75 和 95。

我们使用 PyTorch 进行模拟，并通过单个 TITAN RTX 24 GB GPU 加速。采用随机梯度下降优化器 Adamax [26] ，每批次使用 16 对载体图像-含密图像。批归一化参数通过指数移动平均进行学习，衰减率为 0.9。在预训练中，滤波器权重使用 He 初始化器进行初始化，并采用 2×10^{-4} L_2 正则化。滤波器偏置设为 0.2，且不进行正则化。训练初始化使用预训练网络权重。对于全连接分类器层，权重初始化为均值为零、标准差为 0.01 的高斯分布，且不设置偏置。

3.2 预训练策略

预训练可视为一种初始化网络权重的方式。正确初始化参数对于训练至关重要，尤其是在小载荷隐写分析等困难任务中。在自然图像分类问题中，对模型进行预训练可以使其

提取构成自然图像基本元素的滤波器（例如边缘、纹理、周期性模式等）。因此，我们认为，在更大的隐写分析数据集上进行预训练可以提升模型提取隐写信息的能力。

预训练数据集为 ALASKA2 数据集 [27] ，该数据集用于近期的 Kaggle 竞赛[1]。该数据集包含采用三个质量因子（75、90 和 95）压缩的彩色图像。每种质量因子下，均有 25,000 张分辨率为 512×512 的图像。我们使用 J-UNIWARD 和 UED-JC 生成了隐写图像，载荷为 0.5 至 0.1 bpnzac，质量因子为 75 和 95。全部 25,000 对图像均用于模型预训练。

我们首先在预训练数据集上针对不同的隐写算法训练模型 100 个 epoch。随后，在基础训练数据集上进行 350,000 次迭代（400 个 epoch）训练，初始学习率为 $r_1 = 0.001$；之后将学习率降低至 $r_2 = 0.0001$，再额外训练 87,500 次迭代（100 个 epoch）。最后，选取最近 100,000 次迭代中验证准确率最高时对应的模型参数作为训练结果。

为了验证哪种预训练方式最能有效提高隐写分析检测器的检测准确率，我们设置了 ALASKA2 预训练数据集，采用三种不同的嵌入率：0.2、0.4 和 0.5 bpnzac，以及两种图像质量：质量因子75 和 QF95。我们使用上述三种设置生成基于 J-UNIWARD 的数据集，并将这三个数据集预训练得到的模型参数用于后续训练；后续训练采用相同设置，但将数据集替换为 BOSSBase+BOWS2。模型的检测性能通过测试集上的总分类错误概率来衡量 $P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}})$，其中 P_{FA} 和P_{MD} 分别表示误报概率和漏检概率。

实验结果如表 1所示。可以看出，在训练集嵌入率相同的情况下，预训练集的嵌入率越高，最终模型的效果越好。我们推测，较高嵌入率的预训练集可能会向模型“传授”更多先验知识，从而帮助模型后续识别全新的图像。因此，在所有后续实验中，我们均使用嵌入率为 0.5 bpnzac 的 ALASKA2 作为预训练集。

表 1：在 ALASKA2 上预训练、在 BOSSbase 系列上使用 J-UNIWARD 算法训练的提出的隐写分析模型的检测错误 P_E。

设置	ALASKA2 0.5 bpnzac/质量因子75	ALASKA2 0.4 bpnzac/质量因子75	ALASKA2 0.2 bpnzac/质量因子75
BOSSBase+BOWS2 0.5 bpnzac/质量因子75	.0442	-	-
BOSSbase+BOWS2 0.4 bpnzac/质量因子75	.0901	.1230	-
BOSSbase+BOWS2 0.2 bpnzac/质量因子75	.1492	.1679	.1823

3.3 性能评估

为了与当前在 BOSSBase 和 BOWS2 联合数据集上最先进的方法进行比较，我们将提出的方法与 SRNet 进行了比较 [14] 以及 Xu 提出的网络 [8], ，本文将其称为 J-XuNet。此处比较的所有卷积神经网络检测器均按照相应论文中的描述进行训练。所有检测器均使用与本文完全相同的数据集进行训练，采用 PyTorch 实现，并在单个图形处理器上运行。如表2所示，在多种嵌入率和多种嵌入算法下，提出的方法取得了最佳检测准确率。

表2：提出的方法与以往方法在 J-UNIWARD 和 UED-JC 下、质量因子75（左）和95（右）时，对 bpnzac 中五种载荷的检测错误 P_E。

嵌入方法	模型	QF75					QF95				
		0.5	0.4	0.3	0.2	0.1	0.5	0.4	0.3	0.2	0.1
J-UNIWARD	J-XuNet	.0776	.1207	.1895	.2892	.4310	.2543	.3162	.4246	.4512	.4812
	SRNet	.0575	.1070	.1235	.1929	.3621	.1268	.1845	.2831	.3774	.4367
	提出的	.0442	.0901	.1049	.1492	.3014	.1083	.1595	.2387	.3050	.4021
UED-JC	J-XuNet	.0473	.0887	.1108	.1372	.2944	.0983	.1592	.2091	.2983	.3958
	SRNet	.0113	.0789	.0985	.1188	.1923	.0700	.0978	.1362	.2054	.3369
	提出的	.0105	.0521	.0728	.1003	.1739	.0623	.0804	.1122	.1815	.2950

[1] https://www.kaggle.com/competitions/alaska2-image-steganalysis/overview

图 4 中模型训练期间的准确率曲线表明，在相同的训练时长内，提出的模型具有更好的准确率表现和更高的收敛准确率。从准确率曲线可以看出，提出的模型在训练初期就能达到高于其他模型的准确率，这表明预训练确实为模型提供了先验知识，并帮助模型更快收敛。在后续训练过程中，其测试准确率始终高于其他模型，这也验证了我们提出的 SFE 和 GAL 模块能够有效增强隐写信息的提取能力。SFE 高效的特征增强模式与 GAL 的图注意力学习相结合，使网络能够更好地传递关键的弱隐藏信号，从而使提出的模型始终保持较高的准确率。

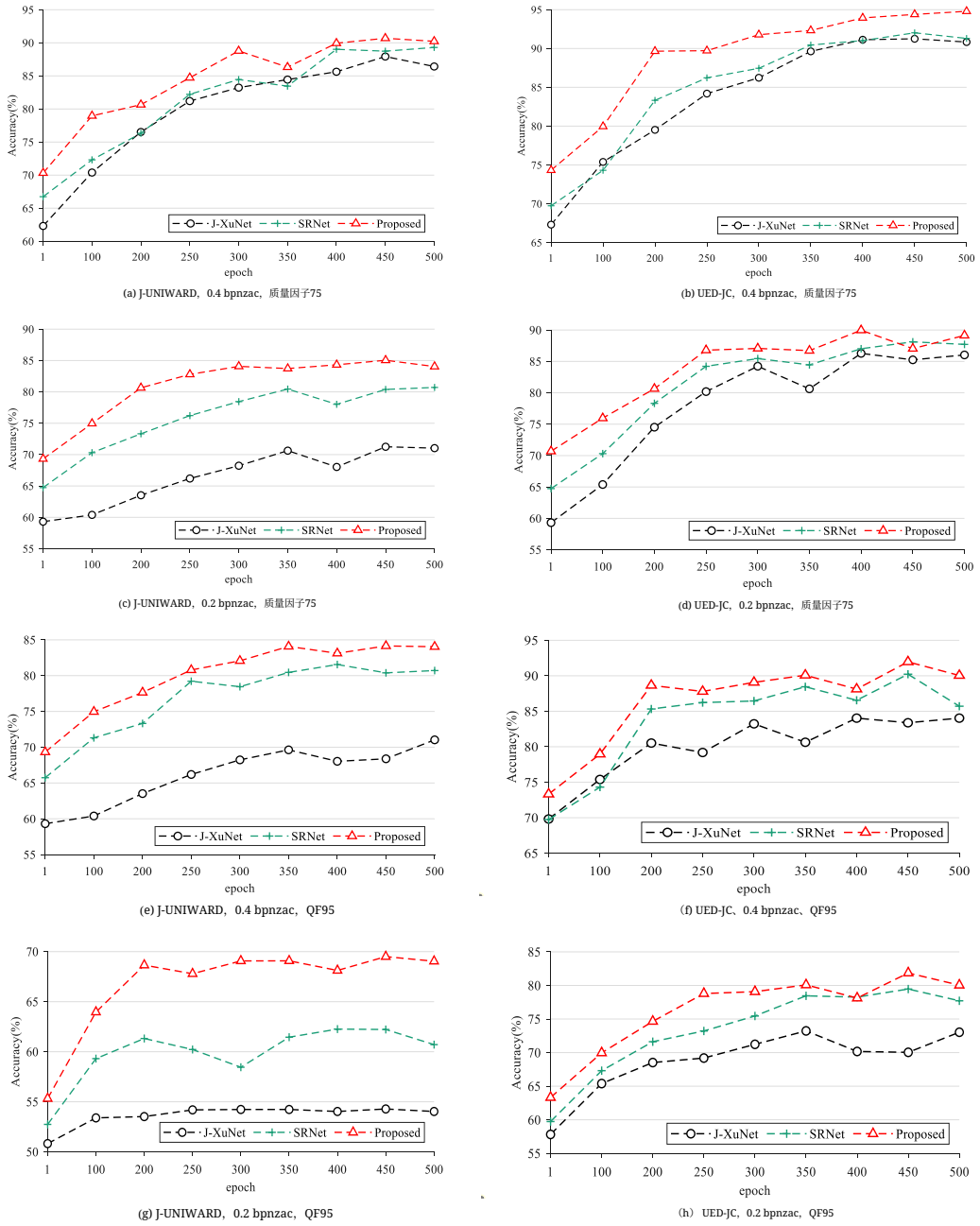

图4：验证集上的检测准确率曲线。

3.4 消融研究

为验证提出的SFE模块和GAL模块的有效性，我们进一步进行了消融实验。我们在更具代表性的0.4 bpnzac嵌入率下进行了消融实验，表3 展示了与标准模型相比，移除SFE模块、移除GAL模块以及同时移除两者时的检测错误率。从表中数据可以看出，移除SFE模块或GAL模块都会导致模型的检测错误率上升；但移除SFE后，模型的准确率比移除GAL后骤降约10%，因此SFE在隐写信息增强中更为重要，而GAL可能凭借提取全局信息信号的能力，辅助SFE进一步提升性能。同时移除SFE和GAL后，模型的检测准确率大幅下降，这表明二者在隐写信息提取方面都非常重要，也验证了我们提出的这两个模块的有效性。

表3：与标准模型相比，不使用SFE模块、不使用GAL模块以及两者都不使用时的检测错误 P_E 。

	J-UNIWARD 质量因子75	J-UNIWARD QF95	UED-JC 质量因子75	UED-JC QF95
提出的	.0901	.1595	.0521	.0804
不使用 SFE	.0232	.3076	.0203	.2897
不使用 GAL	.0133	.1830	.0109	.1433
不使用 SFE 和 GAL	.3987	.4483	.3024	.4075

4 结论

本文结合卷积神经网络中隐写信号传递较弱的特点与图注意力神经网络的全局化优势，设计了一种更适合图像隐写信号特征传递与增强的方法。总体而言，与其他方法相比，本文提出的模块能够更好地解决网络层堆叠导致隐写特征减弱的问题。此外，该方法还能够更充分地利用隐写信号的全局特征。经过在更大数据集上的预训练后，所得模型与现有优秀方法相比，也能够取得更高的检测准确率。随着图神经网络的普及，许多无监督研究不断涌现。未来，我们将探索把无监督学习应用于图像隐写分析的方法。

致谢

本研究　获得 CCF-腾讯犀牛鸟青年教师开放科研基金的资助　　.

参考文献

[1] F. Petitcolas、R. Anderson 和 M. Kuhn。信息隐藏——综述。*IEEE*会议论文集，第87卷，第7期，第1062-1078页，1999年。[2] H. Wu、H. Wang、H. Zhao 和 X. Yu。基于图论方法的多层分配隐写。《多媒体工具与应用》，第74卷，第18期，第8171-8196页，2015年。[3] V. Holub、J. Fridrich。利用非抽取离散余弦变换进行JPEG隐写分析的低复杂度特征。*IEEE*信息取证与安全汇刊，第10卷，第2期，第219-228页，2015年。[4] X. Song、F. Liu、C. Yang、X. Luo 和 Y. Zhang。利用二维Gabor滤波器分析自适应JPEG隐写。*ACM*信息隐藏与多媒体安全研讨会，第15-23页，2015年。[5] V. Holub 和 J. Fridrich。用于JPEG图像隐写分析的相位感知投影模型。*IS&T*电子成像，媒体水印、安全与取证，第9409卷，第94090T页，2015年。

[6] G. Xu、H. Wu 和 Y. Shi。用于隐写分析的卷积神经网络结构设计。IEEE信号处理快报，第23卷，第5期，第708-712页，2016年。[7] G. Xu、H. Wu 和 Y. Shi。用于隐写分析的CNN集成：一项实证研究。ACM信息隐藏与多媒体安全研讨会论文集，第103-107页，2016年。[8] G. Xu。用于检测J-UNIWARD的深度卷积神经网络。ACM信息隐藏与多媒体安全研讨会论文集，第67-73页，2017年。[9] A. Su、X. He 和 X. Zhao。基于采用高斯偏导数滤波器的ResNeXt的JPEG隐写分析。《多媒体工具与应用》，第80卷，第3期，第3349-3366页，2021年。[10] Q. Liu、L. Zhou 和 H. Wu。用于空间图像隐写分析的图表示学习。IEEE多媒体信号处理国际研讨会论文集，第1-5页，2022年。[11] F. Liu、X. Zhou、X. Yan、Y. Lu 和 S. Wang。通过多样化滤波器和挤压激励卷积神经网络进行图像隐写分析。《数学》，第9卷，第2期，第189页，2021年。[12] J. Ye、J. Ni 和 Y. Yi。用于图像隐写分析的深度学习层次表示。IEEE信息取证与安全汇刊，第12卷，第11期，第2545-2557页，2017年。[13] J. Huang、J. Ni、L. Wan 和 J. Yan。用于JPEG隐写分析的低模型复杂度定制卷积神经网络。ACM信息隐藏与多媒体安全研讨会论文集，第198-203页，2019年。[14] M. Boroumand、M. Chen 和 J. Fridrich。用于数字图像隐写分析的深度残差网络。IEEE信息取证与安全汇刊，第14卷，第5期，第1181-1193页，2019年。[15] J. Zhou、G. Cui、S. Hu、Z. Zhang、C. Yang、Z. Liu、L. Wang、C. Li 和 M. Sun。图神经网络：方法与应用综述。AI Open，第1卷，第57-81页，2020年。[16] K. Han、Y. Wang、J. Guo、Y. Tang 和 E. Wu。Vision GNN：一幅图像胜过一个节点图。神经信息处理系统会议论文集，15页，2022年。[17] J. Fridrich 和 J. Kodovsky。用于数字图像隐写分析的丰富模型。IEEE信息取证与安全汇刊，第7卷，第3期，第868-882页，2012年。[18] V. Nair 和 G. E. Hinton。修正线性单元提升受限玻尔兹曼机性能。国际机器学习会议论文集，第807-814页，2010年。[19] S. Ioffe 和 C. Szegedy。批归一化：通过减少内部协变量偏移加速深度网络训练。国际机器学习会议论文集，第448-456页，2015年。[20] Y. Yousfi、J. Butora、J. Fridrich 和 C. F. Tsang。改进用于JPEG隐写分析的EfficientNet。ACM信息隐藏与多媒体安全研讨会论文集，第149-157页，2021年。[21] P. Veli˘ckovi´c、G. Cucurull、A. Casanova、A. Romero、P. Li`o 和 Y. Bengio。图注意力网络。国际学习表征会议论文集，2018年。[22] P. Bas、T. Filler 和 T. Pevn´y。“攻破我们的隐写系统”：组织BOSS的方方面面。信息隐藏国际研讨会论文集，第59-70页，2011年。[23] P. Bas 和 T. Furon。BOWS-2。http://bows2.ec-lille.fr，2007年7月。[24] V. Holub、J. Fridrich 和 T. Denemark。适用于任意域隐写术的通用失真函数。《EURASIP信息安全期刊》，2014年。[25] L. Guo、J. Ni 和 Y. Q. Shi。用于高效JPEG隐写的均匀嵌入。IEEE信息取证与安全汇刊，第9卷，第5期，第814-825页，2014年。[26]D. P. Kingma 和 J. Ba。Adam：一种随机优化方法。arXiv预印本 arXiv:1412.6980，2014年。[27] R. Cogranne、Q. Giboulot 和 P. Bas。ALASKA#2：以真实图像挑战隐写分析领域的学术研究。IEEE信息取证与安全国际研讨会论文集，第1-5页，2020年。