基于两通道深度卷积神经网络的图像隐藏方法

段新涛*① 王文鑫① 李 磊① 邵志强① 王鲜芳② 秦 川③
①(河南师范大学计算机与信息工程学院 新乡 453007)
②(河南工学院计算机科学与技术学院 新乡 453003)
③(上海理工大学光电与计算机工程学院 上海 200093)

摘 要：现有的基于深度卷积神经网络(DCNN)实现的图像信息隐藏方法存在图像视觉质量差和隐藏容量低的问题。针对此类问题，该文提出一种基于两通道深度卷积神经网络的图像隐藏方法。首先，与以往的隐藏框架不同，该文提出的隐藏方法中包含1个隐藏网络和2个结构相同的提取网络，实现了在1幅载体图像上同时对2幅全尺寸秘密图像进行有效的隐藏和提取；其次，为了提高图像的视觉质量，在隐藏网络和提取网络中加入了改进的金字塔池化模块和预处理模块。在多个数据集上的测试结果表明，所提方法较现有的图像信息隐藏方法在视觉质量上有显著提升，载体图像PSNR和SSIM分别提高了3.75 dB和3.61%，实现的相对容量为2，同时具有良好的泛化能力。

关键词：图像信息隐藏；深度卷积神经网络；金字塔池化；预处理
中图分类号：TN911.73; TP309.2 **文献标识码**：A **文章编号**：1009-5896(2022)05-1782-10
DOI: 10.11999/JEIT210280

Image Hiding Method Based on Two-Channel Deep Convolutional Neural Network

DUAN Xintao① WANG Wenxin① LI Lei① SHAO Zhiqiang①
WANG Xianfang② QIN Chuan③
①(*College of Computer and Information Engineering*, *Henan Normal University*, *Xinxiang* 453007, *China*)
②(*College of Computer Science and Technology*, *Henan Institute of Technology*, *Xinxiang* 453003, *China*)
③(*School of Optical-Electrical and Computer Engineering*, *University of Shanghai for Science and Technology*, *Shanghai* 200093, *China*)

Abstract: The existing image information hiding methods based on Deep Convolutional Neural Networks (DCNN) have the problems of poor image visual quality and low hiding capacity. Addressing such issues, an image hiding method based on a two-channel deep convolutional neural network is proposed. First, different from the previous hiding framework, the hiding method proposed in this paper includes one hiding network and two revealing networks with the same structure, and two full-size secret images can be effectively hidden and revealed at the same time is realized. Then, to improve the visual quality of the image, an improved pyramid pooling module and a preprocessing module are added to the hiding and revealing network. The test results on multiple data sets show that the proposed method has a significant improvement in visual quality compared with existing image information hiding methods. The Peak Signal-to-Noise Ratio (PSNR) and Structural SIMilarity (SSIM) values are increased by 3.75 dB and 3.61 % respectively, a relative capacity of 2 and good generalization ability are achieved.

Key words: Image information hiding; Deep Convolutional Neural Networks (DCNN); Pyramid pooling; Preprocessing

收稿日期：2021-04-06；改回日期：2021-09-13；网络出版：2021-12-22
*通信作者： 段新涛 duanxintao@htu.edu.cn
基金项目：国家自然科学基金(U1904123, 61672354, 62072157), 教育人工智能与个性化学习河南省重点实验室基金
Foundation Items: The National Natural Science Foundation of China (U1904123, 61672354, 62072157), The Key Laboratory Foundation of Artificial Intelligence and Personalized Learning in Education of Henan Province

1　引言

信息隐藏主要是指通过将秘密信息嵌入载体的信息冗余中，利用人类感官上的不敏感性实现秘密通信。通常信息隐藏方法有3点特征，即有效嵌入容量、抗检测的安全性和抗失真的鲁棒性[1]。传统的信息隐藏方法主要面向安全性和鲁棒性，例如最低有效位匹配(Least Significant Bit Matching, LSBM)[2]，Filler等人[3]提出的接近仿真最优编码的网格码(Syndrome Trellis Codes, STC)，并基于此构建的一系列最小化失真框架下的称作安全隐写术的信息隐藏方法[4,5]。但此类传统方法一般是利用专业人士设计的代价函数，将秘密消息嵌入载体图像的高频或纹理区域，因此容易受到嵌入容量的制约。

随着深度卷积神经网络(Deep Convolutional Neural Networks, DCNN)在各个领域的兴起，近些年来催生出许多基于DCNN实现的图像隐写术[6–8]。这其中包含一些面向容量的图像隐写术[9–12]，此类方法并不需要完美编码，但要求提取图像和原始图像尽可能相似，因此将此类隐写术称作“数据隐藏”更为恰当[1]。例如，Baluja[9]提出了一种较为直观的编码器-解码器图像数据隐藏框架，包含预处理网络、隐藏网络和提取网络，尺寸为200×200的彩色RGB秘密图像在预处理网络中经卷积层处理后通道数先升高至50再降至30，输出含有高频纹理信息和边缘信息的7通道特征张量，随后和尺寸相同的彩色RGB载体图像级联后一同输入隐藏网络，得到含有秘密图像的隐写图像，再利用提取网络从隐写图像中提取出秘密图像。随后，Chen等人[10]提出了一种称为两通道深度隐藏网络(Two-channel Deep Hiding Network, TDHN)的模型，实现了将一幅彩色图像隐藏至另一幅相同尺寸大小的彩色图像中，达到了1的相对容量，为了提高隐写图像的视觉质量，Chen等人[10]还设计了4种模块用于该隐藏模型，其中Maxmod模块用于提取载体图像的高频纹理特征，Avgmod模块用于保持秘密图像的完整性，Inception Resnet模块用生成隐写图像，Inception模块[11]用于从隐写图像中提取出秘密图像。Yu[12]提出了一种基于CycleGAN[13]结构设计的隐藏模型(Attention Based Data Hiding, ABDH)，利用类激活映射(Class Activation Mapping, CAM)技术[14]，通过预训练的注意力模型提取出载体图像的注意力掩模(Attention Mask, AM)，将载体图像、秘密图像和注意力掩模一同输入目标图像生成模型后得到隐写图像，隐写图像输入秘密图像生成模型实现秘密图像的提取。在训练中，载体图像判别模型-目标图像生成模型与秘密图像判别模型-秘密图像生成模型分别构成对抗关系。此外，由于CycleGAN本身是基于源域到目标域之间的风格转换，而在该图像数据隐藏方法中则需避免从原始载体图像中提取出秘密图像的风险，因此Yu[12]在损失函数中加入不一致性损失，在训练中迫使秘密图像生成模型从载体图像中提取出的图像与原始秘密图像保持不一致。

上述基于DCNN实现的图像数据隐藏方法已在安全方面取得较好的性能，但在视觉效果和嵌入容量方面仍存在不足。针对此类不足，基于深度学习的理论基础，本文提出通过引入改进的金字塔池化模块和预处理模块来改善隐藏网络的隐藏与提取性能，并实现了对两幅秘密图像的隐藏和提取。此外，本文还提出了一种新的评价指标，以更加客观地衡量图像隐藏方法中载体图像的修改率和秘密图像的提取率。本文的主要贡献在于：

(1) 提出了一种新的基于DCNN实现的图像数据隐藏模型，通过应用改进的金字塔池化模块和预处理模块有效地提高了隐写图像的视觉效果和抗检测性，并实现了2的相对容量；

(2) 针对现有图像数据隐藏方法中图像修改率和提取率衡量指标存在的不足，提出了一种新的评价指标；

(3) 除训练所用的ImageNet数据集外，本文的隐藏模型还在多个不同数据集上表现出了良好的泛化能力，可以实现对遥感图像等特殊用途的图像进行有效的隐写和提取。

本文的其余部分将对本文的工作做进一步的介绍：第2节将详细介绍本文提出的图像数据隐藏方法，第3节是实验结果和分析，第4节将对本文工作的全部内容进行总结。

2　图像数据隐藏方法

这部分将对整体框架、改进的金字塔池化模块和预处理模块、隐藏网络和提取网络作详细的介绍。

2.1　整体框架

本文的图像数据隐藏框架如图1所示，在隐藏网络中，发送方将1幅载体图像和2幅秘密图像同时输入2个预处理模块后得到2个张量，再经过下采样-金字塔池化-上采样3个阶段处理，最终得到含有秘密图像的隐写图像，接收方接收到隐写图像后利用2个提取网络分别恢复出2幅不同的提取图像，2个提取网络结构相同，均由下采样-金字塔池化-上采样构成。整个隐藏和提取过程表示为

$$\begin{aligned} C' &= H(C, S_1, S_2) \\ &= H_{\mathrm{Up}}(H_{\mathrm{Pyramid}}(H_{\mathrm{Down}}(\mathrm{Prep_C}(C), \mathrm{Prep_S}(S_1, S_2)))) \end{aligned} \tag{1}$$

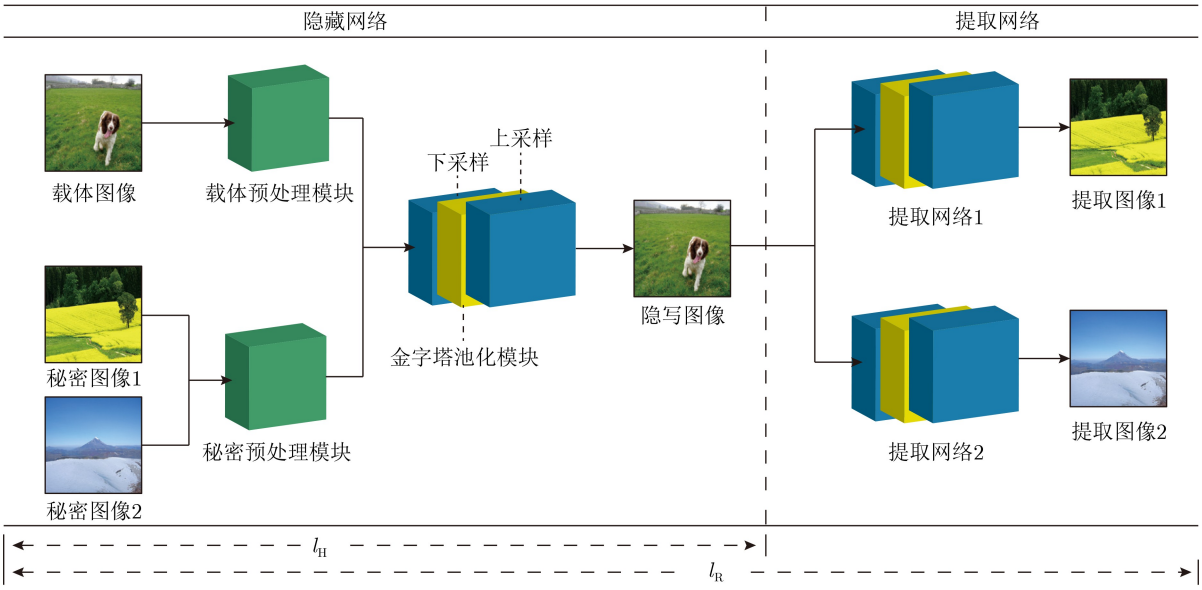


图 1 隐藏框架

$$S'_1 = R_1(C') = R_{1,\mathrm{Up}}(R_{1,\mathrm{Pyramid}}(R_{1,\mathrm{Down}}(C'))) \tag{2}$$

$$S'_2 = R_2(C') = R_{2,\mathrm{Up}}(R_{2,\mathrm{Pyramid}}(R_{2,\mathrm{Down}}(C'))) \tag{3}$$

其中，C, S_1和S_2分别表示载体图像、秘密图像1和秘密图像2，C'表示隐写图像，S'_1和S'_2表示提取图像1和提取图像2。$\mathrm{Prep_C}(*)$和$\mathrm{Prep_S}(*)$分别表示载体预处理模块和秘密预处理模块，$H(*)$, $R_1(*)$和$R_2(*)$分别表示隐藏网络和2个结构相同的提取网络。式中带有Down, Pyramid和Up下标的函数分别表示对应网络的下采样、金字塔池化和上采样3个阶段。

训练中的总损失函数包含隐藏损失l_{H}和提取损失l_{R}，如式(4)—式(6)。l_{H}用于计算C和C'之间的误差，l_{R}用于计算S和S'之间的误差。如图1下方所示，l_{H}仅对隐藏网络的参数更新产生约束，但l_{R}对整个隐藏和提取网络的参数更新均产生约束。

$$l_{\mathrm{H}} = \frac{1}{N}\sum_{i=1}^{N}\|H(C,S_1,S_2)-C\|^2 \tag{4}$$

$$l_{\mathrm{R}} = \frac{1}{N}\sum_{i=1}^{N}\|R_1(C')-S_1\|^2 + \frac{1}{N}\sum_{i=1}^{N}\|R_2(C')-S_2\|^2 \tag{5}$$

$$l_{\mathrm{Sum}} = l_{\mathrm{H}} + \varphi\cdot l_{\mathrm{R}} \tag{6}$$

其中，l_{Sum}是总损失函数，ϕ是超参数，N是每批次输入模型中(C,S_1,S_2)组合的个数。式(5)中l_{R}包含的2项分别是2个提取网络的损失函数。本文使用均方差MSE作为损失函数，见式(7)。

$$\mathrm{MSE} = \frac{1}{W\times H}\sum_{i=1}^{W}\sum_{j=1}^{H}(X_{i,j}-Y_{i,j})^2 \tag{7}$$

其中，X表示原始图像(载体图像或秘密图像)，Y表示经隐藏模型处理得到的修改图像(隐写图像或提取图像)。

2.2 模块

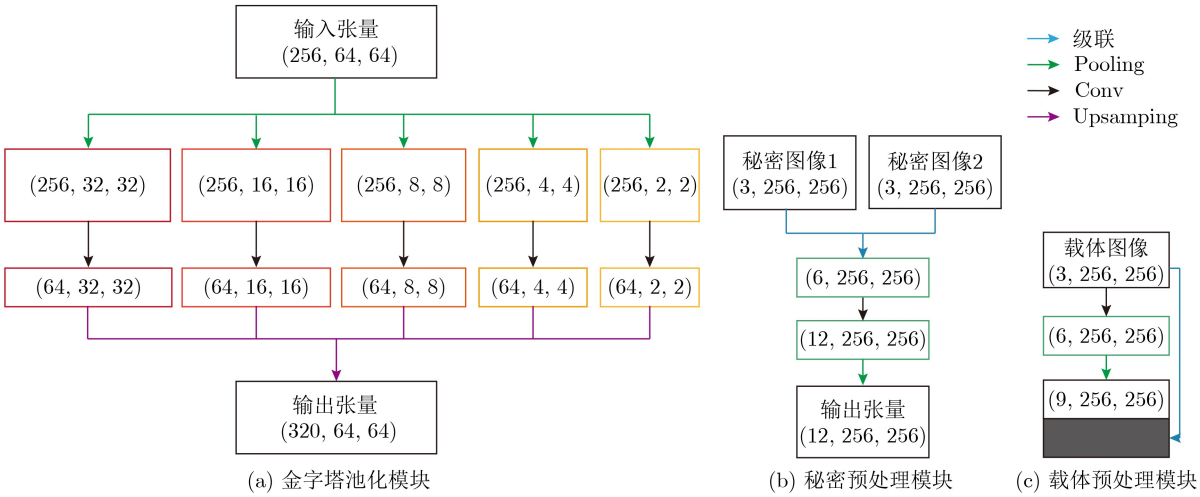
Zhou等人[15]指出DCNN的感受野远比理论域要小，尤其是在DCNN的较深层处，此类情况通常会导致很多网络无法充分融合先前的重要全局特征，进而影响网络的性能。为解决这一问题，Zhao等人[16]提出了一种金字塔池化模块，该模块通过融合不同比例下的特征，将多种规模的特征信息结合起来，从而减少不同子区域之间上下文信息的丢失。基于此，本文尝试在隐藏模型中加入金字塔池化模块以提高隐写模型生成图像的质量，但在实验中发现原始模块中的等级和大小在隐藏模型中并不能满足隐写任务的需求，因此本文尝试将原始的金字塔池化模块的等级从4种增加至5种，并将每个等级的大小从$1\times1\sim6\times6$修改为$2\times2\sim32\times32$，如图2(a)。下文图2、图3的$(D,\ W,\ H)$分别表示图像的通道数、宽度和高度。

同时，本文在隐藏网络中引入了预处理模块，与文献[10]中的两通道预处理模块不同，本文仅用1次池化和2次卷积提取特征，尽可能地减少整个网络的参数量和计算量，如图2(b)和图2(c)所示。其中，2幅秘密图像级联后输入秘密预处理模块，通过卷积层提高通道数后，再经全局平均池化层得到含有全局特征的输出张量，类似地，载体图像输入载体预处理模块后也是通过卷积层提高通道数，再

经全局最大池化层提取相应的高频特征[10]。由于隐藏网络的最终目标是生成与载体图像尽可能相似的隐写图像，因此本文在载体预处理模块中增加了一个跳跃连接，将载体图像与载体预处理模块的输出张量进行级联，以减少在后续特征提取中载体图像有效特征的丢失。

2.3 隐藏网络和提取网络

除金字塔池化模块和预处理模块外，隐藏网络的整体结构类似U-Net[17]中的编码器-解码器结构，如图3上。另外本文也使用了跳跃连接用于将前期卷积层中的一些特征张量映射到后期的卷积层中，这有助于降低重要特征的丢失。相较于隐藏网络，提取网络并不包含预处理模块，但保留了下采样-金字塔池化-上采样阶段，如图3下。尽管$R_1(*)$和$R_2(*)$的结构相同，但由于在训练中两者的训练目标(提取图像)并不是同一幅秘密图像(S_1和S_2)，因此最终得到的参数并不相同。此外，如图2、图3所示，各个模块、隐藏网络以及提取网络中的单一层



(a) 金字塔池化模块　(b) 秘密预处理模块　(c) 载体预处理模块

图 2 改进模块

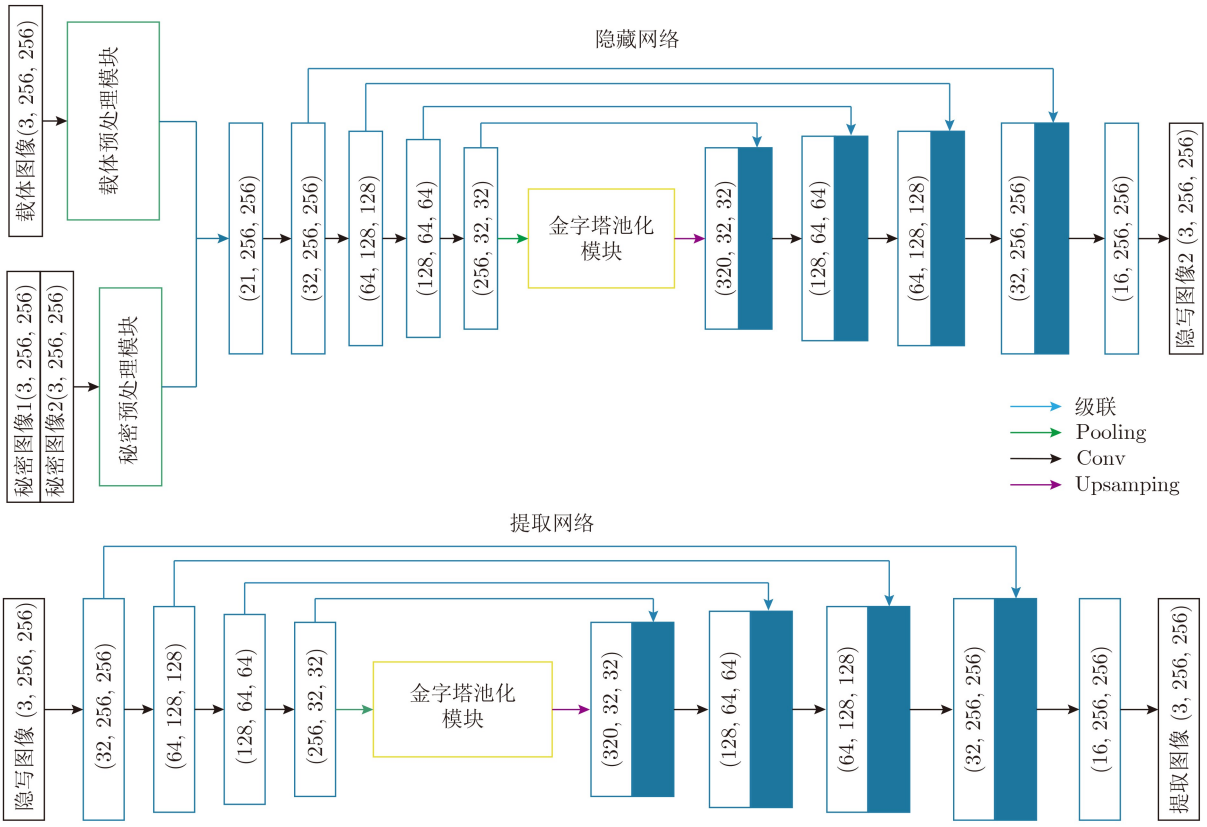


图 3 隐藏网络和提取网络

中均包含卷积、BatchNorm和ReLU激活函数，其中涉及对特征图的尺寸大小进行减半或翻倍操作的层卷积核大小为4×4，其余层均为3×3。

3 实验和分析

为保证隐藏模型的训练效果，本文从ImageNet数据集中随机挑选了100000幅图像作为训练用的数据集，其中98000幅图像用作训练集，验证集和测试集各1000幅。另外，训练、验证及测试中所产生的载体图像和秘密图像的组合均是随机生成。训练的总epochs设为200，初始学习率设为1×10^{-3}，ϕ设置为0.5，优化器使用Adam。实验所用的程序是Python 3.6，实现隐藏模型所用机器学习库是pytorch。实验所用GPU设备的配置为GeForce RTX 2070 8 GB。

3.1 图像隐藏效果及分析

实验中，本文方法和文献[9]中实现相同嵌入容量的方法进行了比较，如图4。文献[9]在图像色彩还原方面效果较差，尤其是红框区域，详见图5。而本文方法无论是隐写图像还是提取图像的视觉效果均明显优于文献[9]，尤其是当载体残差亮度提高至5倍时文献[9]的方法就显示出较为明显的像素修改痕迹。

除主观视觉效果外，本文还使用峰值信噪比PSNR[18]和结构相似性SSIM[19]对生成图像的质量进行客观评价。PSNR属于一种评价图像的客观标准，值越大越好，一般值在30以上即认为修改后的图像与原图像在视觉上相差不大。SSIM是从亮度、对比度以及结构3个方面来对两幅图像之间的相似性进行客观评价的，取值范围介于0到1之间，越高越好。具体计算方法见式(8)、式(9)。

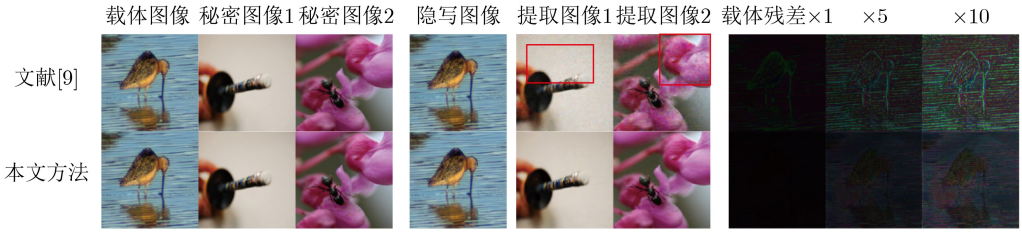
$$\mathrm{PSNR}=10\cdot\lg\left(\frac{\mathrm{MAX}_{\mathrm{I}}^{2}}{\mathrm{MSE}}\right) \tag{8}$$

其中，$\mathrm{MAX}_{\mathrm{I}}^{2}$是图像可能的最大像素值，$\mathrm{MAX}_{\mathrm{I}}=2^{n}-1$，如果图像中的像素由8位二进制表示，则$n$=8。

$$\mathrm{SSIM}\left(x,y\right)=\frac{\left(2\mu_x\mu_y+c_1\right)\left(2\sigma_{xy}+c_2\right)}{\left(\mu_x^2+\mu_y^2+c_1\right)\left(\sigma_x^2+\sigma_y^2+c_2\right)} \tag{9}$$

其中，μ代表对应下标的均值，σ^2代表对应下标的方差，c_1和c_2是两个常数。

针对图4中的2组图像本文测试了其PSNR和SSIM，如表1所示，使用本文方法所生成图像的指标均高于文献[9]，尤其是载体图像的PSNR和



(a) 示例1

(b) 示例2

图 4 主观效果对比

图 5 放大效果对比

SSIM的最高值分别达到了38.17 dB和99.28%，秘密图像对应的指标最高值分别达到了40.13 dB和98.11%。此外，本文方法各项指标的平均值较文献[9]分别提高了3.75 dB/3.61%, 4.57 dB/5.86%和4.56 dB/6.19%。

3.2 消融实验

为检查改进的金字塔池化模块和预处理模块是否对本文图像隐藏模型性能的提高具有帮助，本文利用消融实验在多个方面对模块的性能进行检测，并以此进一步了解本文提出方法的可行性。在此实验中分别去掉了金字塔池化模块和预处理模块，在保证其他指标不变更的情况下，分别对模型进行重新训练，随后在客观指标和抗检测性等方面进行比较。

针对最基本的客观指标PSNR和SSIM，本文使用测试集的1000幅图像分别对对照组模型进行了测试，结果见表2。表2中*Prep表示去掉预处理模块，*Pyramid表示去掉金字塔池化模块。通过对测试集的1000幅图像进行测试，很明显在客观指标上对照组都未能达到原始图像隐藏模型的水平。其中，对照组*Pyramid在大多数指标上都低于原始模型和对照组*Prep，其原因可能在于本文提出的隐藏网络和提取网络均使用了金字塔池化模块，这也正说明金字塔池化模块对本文隐藏模型性能提高的重要性。

此外，针对消融实验中不同对照组的安全性本文也作了分析。StegExpose[20]是一种针对LSB隐写术检测的隐写分析工具，实验中使用StegExpose的分析结果如图6，纵轴正阳率表示分析正确，横轴负阳率表示分析错误，黑色虚线表示隐写分析结果等同于随机猜测，图中每一组ROC (Receiver Operating Characteristic) 曲线的下方面积(Area Under Curve, AUC)见表3第2列，数值越小越好。此外，还使用SRNet[21]对本文模型进行了隐写分析，SRNet是一种基于DCNN实现的隐写分析网络，可以对目标图像的空域和JPEG域进行有效的盲隐写分析检测，分析结果见表3第3列。从表3的数据来看，本文方法在基于DCNN的隐写分析下

表 1　与文献[9]比较

图像	方法	载体图像-隐写图像		秘密图像1-提取图像1		秘密图像2-提取图像2	
		PSNR(dB)	SSIM(%)	PSNR(dB)	SSIM(%)	PSNR(dB)	SSIM(%)
图4(a)	文献[9]	31.95	94.44	29.68	83.90	26.93	78.43
	本文方法	**34.48**	**99.28**	**40.13**	**97.91**	**32.60**	**98.22**
图4(b)	文献[9]	31.08	95.19	28.72	93.01	33.22	88.67
	本文方法	**38.17**	**98.47**	**37.16**	**98.11**	**34.35**	**97.28**
平均值	文献[9]	32.32	94.81	30.40	90.70	30.55	90.29
	本文方法	**36.07**	**98.42**	**34.97**	**96.56**	**35.11**	**96.48**

表 2　消融实验的PSNR和SSIM比较

图像	载体图像-隐写图像		秘密图像1-提取图像1		秘密图像2-提取图像2	
	PSNR(dB)	SSIM(%)	PSNR(dB)	SSIM(%)	PSNR(dB)	SSIM(%)
ImageNet	**36.07**	**98.42**	**34.97**	**96.56**	**35.11**	**96.48**
*Prep	35.66	96.89	34.23	95.12	34.44	95.17
*Pyramid	34.86	95.65	33.85	94.93	34.10	96.27

表 3　隐写分析结果

隐藏模型	StegExpose AUC	SRNet 隐写分析准确率
本文方法	**0.5533**	**0.6844**
文献[9]	–	0.6975
*Prep	0.6223	0.7195
*Pyramid	0.5698	0.6994

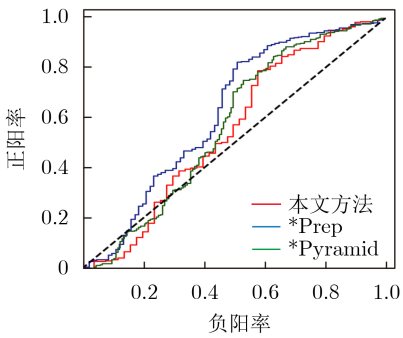


图 6 StegExpose隐写分析结果

的表现略有不足，这是由于此类方法的隐藏容量大，隐写图像难以保持不被察觉[7,9]。但在消融实验中，本文模型的SRNet隐写分析准确率比对照组都要低，这说明本文提出的改进模块对于提高整个模型的安全性具有一定的积极作用，同时该准确率还低于文献[9]，因此本文方法比文献[9]具有相对更高的安全性。

3.3 嵌入容量分析

由于本文提出的隐藏模型将秘密图像的数量提高至2幅，因此在嵌入容量方面要高于现有的基于DCNN实现的图像数据隐藏方法。通常图像数据隐藏的嵌入容量使用相对容量进行衡量，详见式(10)。通过和几种现有的方法进行比较，本文提出的隐藏模型的嵌入容量远大于其他方法，详见表4。

$$\text{相对容量}=\frac{\text{绝对容量}}{\text{隐写图像大小}} \tag{10}$$

其中，绝对容量是指每幅载体图像中嵌入秘密消息的大小，即秘密图像的大小。

3.4 修改率和提取率

为了对隐藏方法中的图像像素值修改幅度进行量化，文献[10]提出了一种针对图像修改率和提取率的计算方法，如式(11)、式(12)所示。其计算原理在于，衡量单个像素上像素值修改幅度的平均值。但当某点像素值由0修改为1时，该式中微小常数ε的值将对计算结果产生较大影响，且该点的像素值并不能代表该点像素值可能达到的最大修改幅度。

$$\text{修改率}=\frac{1}{W\times H}\sum_{j=1}^{H}\sum_{i=1}^{W}\frac{|C'(i,j)-C(i,j)|}{C(i,j)+\varepsilon} \tag{11}$$

$$\text{提取率}=1-\frac{1}{W\times H}\sum_{j=1}^{H}\sum_{i=1}^{W}\frac{|S'(i,j)-S(i,j)|}{S(i,j)+\varepsilon} \tag{12}$$

其中，W表示图像的宽度，H表示图像的高度，C'表示隐写图像，C表示载体图像，S'表示提取图像，S表示秘密图像。

基于上述原因本文提出了一种新的计算方式，如式(13)—式(15)。此处需要强调的是，尽管修改率和提取率是基于像素值修改幅度计算而来的，但其仅能表达图像在经过隐藏模型处理后平均每个像素上被修改的程度，并不能直接代表视觉效果的好坏。如表5所示，在载体图像修改率和秘密图像提取率等指标上本文方法基本保持了与文献[9]所提方法的同等水平，部分指标还略有优势。结合表1中的数据，说明与文献[9]相比本文方法在像素修改幅度保持同等水平的基础上实现了更优的隐藏效果。

$$\text{修改率}=\frac{1}{W\times H\times D}\sum_{k=1}^{D}\sum_{j=1}^{H}\sum_{i=1}^{W}\frac{|C'(i,j,k)-C(i,j,k)|}{\delta} \tag{13}$$

$$\text{提取率}=1-\frac{1}{W\times H\times D}\sum_{k=1}^{D}\sum_{j=1}^{H}\sum_{i=1}^{W}\frac{|S'(i,j,k)-S(i,j,k)|}{\delta} \tag{14}$$

$$\delta=\max\left(\text{img}(i,j,k),(255-\text{img}(i,j,k))\right) \tag{15}$$

其中，D表示图像的通道数，δ表示某点像素值可能达到的最大修改幅度。

表 4 嵌入容量比较

方法	绝对容量(Byte)	隐写图像大小(Byte)	相对容量
文献[21]	18.3～135.4	64×64	1.49×10^{-3}～1.10×10^{-2}
文献[22]	1535～4300	1024×1024	1.46×10^{-3}～4.10×10^{-3}
文献[23]	26214～104857	512×512	1×10^{-1}～4×10^{-1}
文献[10]	3×224×224	3×224×224	1
文献[12]	3×256×256～3×512×512	3×512×512	2.5×10^{-1}～1
本文方法	6×256×256	3×256×256	2

表 5 修改率和提取率比较(%)

图像	方法	载体图像修改率	秘密图像1提取率	秘密图像2提取率
图4(a)	文献[9]	**1.35**	98.65	97.44
	本文方法	1.68	**99.30**	**98.18**
图4(b)	文献[9]	1.20	98.47	**98.86**
	本文方法	**1.09**	**98.97**	98.47
ImageNet平均值	文献[9]	1.80	98.02	97.62
	本文方法	**1.61**	**98.99**	**98.92**

3.5 泛化能力分析

泛化能力，通常是指机器学习算法对具有同一规律的训练集以外的数据，在经过训练后的网络模型中也能得到合理的输出。除训练用的ImageNet数据集外，本文还选取了CeleA, COCO, VOC2012, AID和UCMerced LandUse共5个数据集，并分别随机抽取1000幅图像对隐藏模型进行测试。其中，CeleA为人脸数据集，COCO和VOC2012属于自然图像数据集，AID和UCMerced LandUse为遥感图像数据集，考虑到在应用场景中遥感图像一般不会作为载体图像参与信息隐藏行为，因此在对AID和UCMerced LandUse两组数据集的测试中本文使用了自然图像作为载体。测试结果如图7所示，在5组测试集的测试结果中图像的还原度均表现良好，残差图中也无明显的秘密图像轮廓。表6展示了在5组测试集中的客观指标平均值，尤其是CeleA中的部

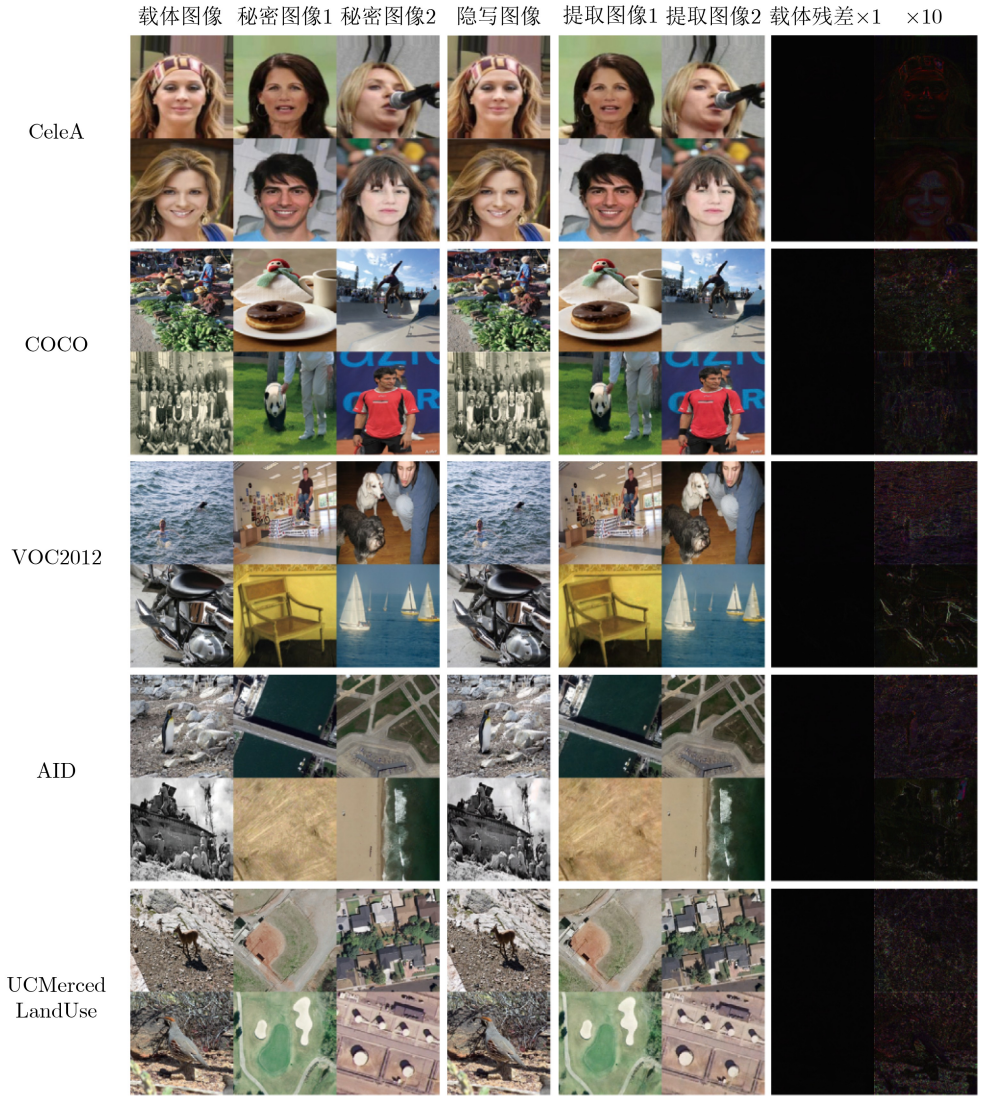


图 7 不同数据集的测试结果

表 6　5组数据集的测试结果

数据集	载体图像-隐写图像			秘密图像1-提取图像1			秘密图像2-提取图像2		
	PSNR(dB)	SSIM(%)	修改率(%)	PSNR(dB)	SSIM(%)	提取率(%)	PSNR(dB)	SSIM(%)	提取率(%)
CeleA	35.81	96.40	1.83	36.18	96.90	98.22	36.24	97.50	98.24
COCO	34.03	96.41	2.22	33.29	93.79	97.62	33.82	94.98	97.76
VOC2012	34.06	96.48	2.23	33.45	93.54	97.66	34.10	94.80	97.80
AID	34.83	97.29	2.07	32.35	93.28	96.75	34.12	95.22	97.65
UCMerced Land Use	34.53	96.88	2.16	31.25	92.52	96.44	32.12	94.76	96.85

分指标超越了隐藏模型在ImageNet测试集中的表现。此外，表6显示载体图像的修改率介于1.83%到2.23%之间，秘密图像的提取率介于96.44%到98.24%之间，5组数据集中产生图像的平均修改率和提取率均保持在了较好水平，因此本文模型具有良好的泛化能力。

4 结束语

本文提出了一种新的图像隐藏和提取方法，通过对改进后的金字塔池化模块和预处理模块的应用，实现了在1幅载体图像上对2幅全尺寸秘密图像的隐藏和提取，并验证了隐藏模型在多个方面的良好性能。具体地，本文的模型有以下特点：(1)具有良好的视觉效果，通过本文提出的隐藏模型获得的隐写图像和两幅提取图像，峰值信噪比PSNR/结构相似性SSIM的最高值可以分别达到38.17 dB/99.28%和40.13 dB/98.11%，测试集中的平均值分别达到了36.07 dB/98.42%, 34.97 dB/96.56%和35.11 dB/96.48%；(2)具有高达2的相对容量；(3)通过对CeleA, COCO等多种不同类型数据集的测试，结果显示具有良好泛化能力，可以胜任不同场景的应用，例如传递遥感图像的军事用途等。

参 考 文 献

[1] ZHANG Chaoning, LIN Chenguo, BENZ P, *et al.* A brief survey on deep learning based data hiding, steganography and watermarking[EB/OL]. https://arxiv.org/abs/2103.01607, 2021.

[2] KER A D. Improved detection of LSB steganography in grayscale images[C]. The 6th International Workshop on Information Hiding (IH), Toronto, Canada, 2004: 97–115. doi: 10.1007/978-3-540-30114-1_8.

[3] FILLER T, JUDAS J, and FRIDRICH J. Minimizing additive distortion in steganography using Syndrome-Trellis Codes[J]. *IEEE Transactions on Information Forensics and Security*, 2011, 6(3): 920–935. doi: 10.1109/TIFS.2011.2134094.

[4] FRIDRICH J and FILLER T. Practical methods for minimizing embedding impact in steganography[C]. SPIE 6505, Security, Steganography, and Watermarking of Multimedia Contents IX, San Jose, USA, 2007: 13–27. doi: 10.1117/12.697471.

[5] FRIDRICH J, GOLJAN M, LISONEK P, *et al.* Writing on wet paper[C]. SPIE 5681, Security, Steganography, and Watermarking of Multimedia Contents VII, San Jose, USA, 2005: 328–340. doi: 10.1117/12.583160.

[6] TANG Weixuan, LI Bin, BARNI M, *et al.* An automatic cost learning framework for image steganography using deep reinforcement learning[J]. *IEEE Transactions on Information Forensics and Security*, 2021, 16: 952–967. doi: 10.1109/TIFS.2020.3025438.

[7] ZHANG Chaoning, BENZ P, KARJAUV A, *et al.* UDH: Universal Deep Hiding for steganography, watermarking, and light field messaging[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 10223–10234.

[8] LUO Xiyang, ZHAN Ruohan, CHANG Huiwen, *et al.* Distortion agnostic deep watermarking[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 2020: 13545–13554. doi: 10.1109/CVPR42600.2020.01356.

[9] BALUJA S. Hiding images within images[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(7): 1685–1697. doi: 10.1109/TPAMI.2019.2901877.

[10] CHEN Feng, XING Qinghua, and LIU Fuxian. Technology of hiding and protecting the secret image based on two-channel deep hiding network[J]. *IEEE Access*, 2020, 8: 21966–21979. doi: 10.1109/ACCESS.2020.2969524.

[11] SZEGEDY C, VANHOUCKE V, IOFFE S, *et al.* Rethinking the inception architecture for computer vision[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016: 2818–2826. doi: 10.1109/CVPR.2016.308.

[12] YU Chong. Attention based data hiding with generative adversarial networks[C]. The 34th AAAI Conference on Artificial Intelligence (AAAI), New York, USA, 2020: 1120–1128. doi: 10.1609/aaai.v34i01.5463.

[13] ZHU Junyan, PARK T, ISOLA P, *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017: 2242–2251. doi: 10.1109/ICCV.2017.244.

[14] ZHOU Bolei, KHOSLA A, LAPEDRIZA À, *et al.* Learning deep features for discriminative localization[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016: 2921–2929. doi: 10.1109/CVPR.2016.319.

[15] ZHOU Bolei, KHOSLA A, LAPEDRIZA À, *et al.* Object detectors emerge in deep scene CNNs[EB/OL]. http://arxiv.org/abs/1412.6856, 2015.

[16] ZHAO Hengshuang, SHI Jianping, QI Xiaojuan, *et al.* Pyramid scene parsing network[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017: 6230–6239. doi: 10.1109/CVPR.2017.660.

[17] RONNEBERGER O, FISCHER P, and BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]. The 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 2015: 234–241. doi: 10.

1007/978-3-319-24574-4_28.

[18] HORÉ A and ZIOU D. Image quality metrics: PSNR vs. SSIM[C]. The 20th International Conference on Pattern Recognition, Istanbul, Turkey, 2010: 2366–2369. doi: 10.1109/ICPR.2010.579.

[19] YE Yuanxin, SHAN Jie, BRUZZONE L, *et al*. Robust registration of multimodal remote sensing images based on structural similarity[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(5): 2941–2958. doi: 10.1109/TGRS.2017.2656380.

[20] BOEHM B. StegExpose - A tool for detecting LSB steganography[EB/OL]. http://arxiv.org/abs/1410.6656, 2014.

[21] BOROUMAND M, CHEN Mo, and FRIDRICH J. Deep residual network for steganalysis of digital images[J]. *IEEE Transactions on Information Forensics and Security*, 2019, 14(5): 1181–1193. doi: 10.1109/TIFS.2018.2871749.

[22] WU Kuochen and WANG C. Steganography using reversible texture synthesis[J]. *IEEE Transactions on Image Processing*, 2015, 24(1): 130–139. doi: 10.1109/TIP.2014.2371246.

[23] YANG Jianhua, LIU Kai, KANG Xiangui, *et al*. Spatial image steganography based on generative adversarial network[OL]. http://arxiv.org/abs/1804.07939, 2018.

段新涛：男，1972年生，副教授，研究方向为图像信息隐藏、图像盲取证、深度学习、盲源分离等.

王文鑫：男，1994年生，硕士生，研究方向为图像信息隐藏、深度学习.

李　磊：男，1995年生，硕士生，研究方向为图像信息隐藏、深度学习.

邵志强：男，1996年生，硕士生，研究方向为图像信息隐藏、深度学习.

王鲜芳：女，1969年生，教授，研究方向为人工智能与模式识别、数据挖掘、机器学习及应用.

秦　川：男，1980年生，教授，研究方向为多媒体信息安全、数字图像处理、信息隐藏、AI安全、深度学习、密文域信号处理、数字取证.

责任编辑：马秀强