

基于 UNetFormer 的隐藏图像特殊噪声去除方法

文 来¹ 阴明旭¹ 黎润霖¹ 刘 瑞¹ 文 强¹ 马 龙¹
WEN Lai YIN Mingxu LI Runlin LIU Rui WEN Qiang MA Long

摘 要

针对被提取出的含特殊噪声隐藏图像, 文章设计了一种基于 UNet 的自注意力机制隐藏图像去噪网络。该网络利用 UNet 结构对图像进行尺度变换, 采用卷积操作实现全局去噪。网络主干嵌入了的滑窗自注意力机制优化版, 利用其滑动窗口感知能力对图像进行局部去噪。通过添加分割和双采样操作, 使其能够获取高分辨率的图像, 并且有效解决图像棋盘效应。基于隐藏图像噪声数据集的去噪实验结果显示, 该网络能够有效去除隐藏图像噪声, 有效恢复图像的纹理细节, 验证了文章方法的优异去噪效果。

关键词

隐藏图像; 图像去噪; UNet; 自注意力机制

doi: 10.3969/j.issn.1672-9528.2025.06.029

0 引言

随着数字技术和互联网的飞速发展, 图像、音频、视频等数字媒体已成为信息传播的主要载体。然而, 信息在传输和存储过程中面临着安全威胁, 如未经授权的访问、窃取、篡改和非法复制等。为保护信息的机密性、完整性和可用性, 信息加密隐藏技术^[1]应运而生。将信息隐藏到图像中, 即利用图像作为载体, 将秘密信息嵌入其中, 使其对人类视觉系统不可察觉。这种技术利用了数字图像的冗余性和人类视觉系统对细微变化的不敏感性, 实现了在不明显改变图像质量的情况下传递和存储秘密信息。

然而当载体图像被噪声破坏时, 被提取出的隐藏图像也会相应地携带噪声, 并且破坏程度更大, 因此需要特定的方法进行噪声恢复, 使得隐藏图像重新可读。在神经网络尚未广泛用于图像处理之前, 研究人员提出了一种位修复方法。该方法检测被椒盐噪声破坏的载体图像像素, 对隐藏图像中从这些受损点提取的位标记为不可信, 并利用邻近像素的可信位, 针对不同情况采用不同的恢复策略。通过这种方法, 像素中失真的高位得到准确恢复, 有效提高了图像的恢复质量^[2]。

文献 [3] 提出了一种新的迭代滤波器, 用于检测和去除椒盐噪声。该滤波器根据图像的噪声密度进行决策, 既能滤除噪声, 又能保留图像的细节。每一步都使用固定大小的窗口, 确保过滤过程的相关性。基于多过程的最小-最大平均池化的滤波器, 也能够用于去除传输过程中出现的椒盐噪声^[4]。

使用上述方法恢复后, 图像还会存在比较明显的模糊和扭曲, 使得隐藏图像恢复不完整, 因此本文提出了一种基于 UNet 的自注意力机制隐藏图像噪声去除网络, 针对载密图像在被不同强度噪声破坏下的隐藏图像进行恢复, 并且采用滑窗自注意力机制优化主干网络, 使得图像细节恢复更完整。

1 UNet 自注意力机制隐藏图像噪声去除网络

UNet 网络具有精确的定位能力, 通过对称的 U 形结构和跳跃连接, 将编码器中的特征直接传递给解码器, 融合了低层次的细节信息和高层次的语义信息, 实现了像素级别的精确分割。对小样本数据也有良好的适应性, 最初为医学图像分割设计^[5], 能够在训练数据有限的情况下取得优异的性能。UNet 是端到端的全卷积神经网络, 直接从原始输入图像输出分割结果, 简化了训练和推理过程。通过多层次的特征提取和融合, UNet 同时捕获了全局上下文和局部细节信息, 增强了对目标形状和边缘的识别能力。并且其结构简单易于实现和调整, 适用于处理高分辨率图像, 计算效率高, 被广泛应用于医学图像、遥感图像和自然图像的分割等多个领域。

1.1 Swin Transformer 自注意力机制

Swin Transformer 引入了一种用于计算机视觉任务的层次化 Transformer 架构, 其自注意力机制基于移动窗口来计算。具体而言, 图像首先被划分为不重叠的局部窗口, 在每个窗口内计算自注意力, 以捕获局部的特征信息。随后, 在后续层中, 窗口的划分方式会发生位移, 使得窗口之间有所重叠, 这种移动窗口的方法使得相邻窗口之间能够进行信息交互。通过这种设计, Swin Transformer 在保留局部细节的同时, 有效捕获了全局上下文信息, 且计算复杂度随着图像大小线性增长, 因而在处理高分辨率图像时具有更高的效率和可扩展

1. 吉利学院智能科技学院 四川成都 641423

[基金项目] 四川省大学生创新训练计划项目 (S202412802263)

性。其网络核心结构如图 1 所示。

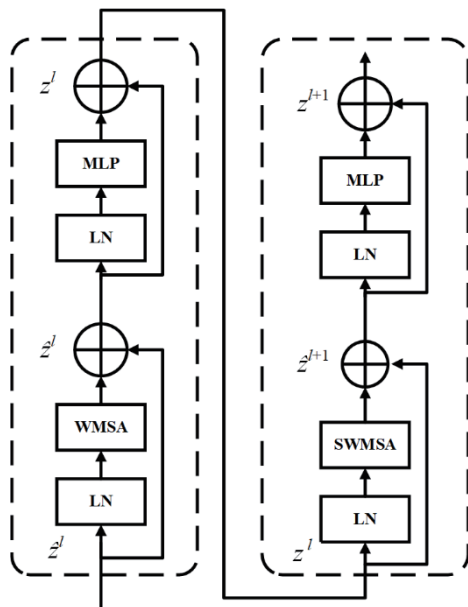


图 1 Swin Transformer 核心结构

本文采用的是 Swin Transformer v2^[6] 是上述核心结构的优化版本，其网络结构对比图如图 2 所示。其引入了对数间隔的连续位置偏置，使模型能够更好地适应不同分辨率的输入，提升了对高分辨率图像的泛化能力。采用了后归一化替代了前归一化，提高了深层网络的训练稳定性。引入余弦相似度来计算注意力权重，取代了传统的点积注意力机制。这种方法对于模型输入的振幅不敏感，有助于减少极端注意力值的发生。

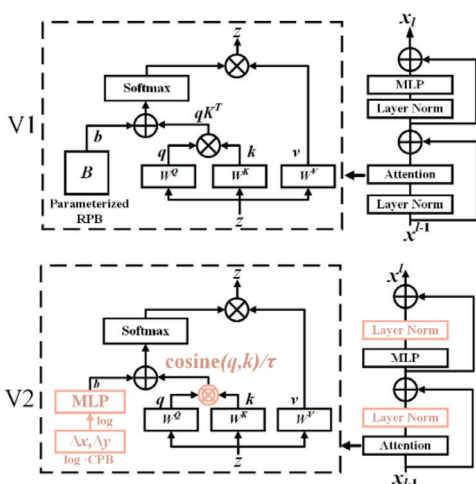


图 2 Swin Transformer v2 结构对比

1.2 模型结构

本文提出了一种融合 Swin Transformer v2 与 UNet 的图像去噪网络。整体框架继承了 UNet 的上下采样结构，用于特征提取与分辨率恢复，同时引入 Swin Transformer v2 和上下采样模块，以提高去噪效果。

其具体流程为：输入噪声图像后，首先经过 3×3 卷积层，再通过 Swin Transformer 和上下采样模块，接着通过一次下采样卷积，特征图尺寸缩小至原来的一半。该操作重复两次，使特征图的尺寸最终缩减为输入图的 $1/4$ 。

然后，缩减后的特征图继续通过 Swin Transformer v2 与上下采样模块处理，并与之前对应的下采样特征图做残差操作。经过上采样卷积逐步恢复特征图的分辨率，重复两次后，最终恢复至原始输入的尺寸。最后通过一层 3×3 卷积得到输出图像，即去噪后的预测图像。网络模型如图 3 所示。

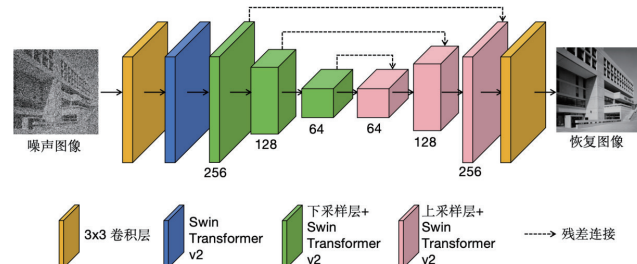


图 3 本文提出的网络结构

与此同时，在原有 UNet 网络使用跳跃连接的基础上，本文改进使用残差连接，有效缓解深层网络中的梯度消失问题，确保梯度能够稳定地传播到前面的层，更有效地整合来自不同层次的信息，从而提高模型对特征的表达能力，尤其在去噪等任务中，能够保持原有特征并在此基础上增强去噪效果。直接学习复杂映射可能会导致特征退化问题。残差连接通过学习残差而不是直接学习映射，使得训练更加容易，并减少深度模型中的特征退化。并且，在使用残差连接后，网络参数明显减少，不会有过多冗余的特征图在网络中训练，大幅减少了训练时间和资源成本。

1.3 评价指标

PSNR（峰值信噪比）和 SSIM（结构相似性指数）是用于评估图像质量的两个常用指标，特别在图像去噪和图像恢复等任务中。PSNR 用于衡量原始图像与恢复图像之间的差异。其值越高，表示图像质量越好，与原始图像的差异越小。其公式表示为：

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right) \quad (1)$$

式中：MAX 为图像像素的最大值；MSE 为均方误差，是原始图像与重建图像之间的平均像素差平方和。

SSIM 用于衡量原始图像和恢复图像在结构上的相似性，考虑亮度、对比度以及结构等多个方面。与 PSNR 不同，SSIM 更接近于人类视觉系统对图像质量的判断。SSIM 的值范围在 0 到 1 之间，1 表示两个图像完全相同。计算时，SSIM 通过亮度、对比度和结构分量的组合来对图像进行评估，从而更好地反映图像的视觉质量。其计算公式为：

$$\text{SSIM}(x, y) = \frac{(2u_x u_y + c_1)(2\sigma_{xy} + c_2)}{(u_x^2 + u_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2)$$

式中： u_x 和 u_y 表示图像 x 和 y 的均值； σ_x 和 σ_y 表示图像 x 和 y 的标准差； σ_{xy} 表示 x 和 y 之间的协方差； c_1 和 c_2 表示稳定常数。

1.4 损失函数

本文选用自定义损失函数进行图像去噪。整体自定义损失函数的公式为：

$$L_{\text{total}} = l_{\text{MSE}} + 0.05 l_{\text{MS_SSIM}} \quad (3)$$

式中： L_{total} 表示整体自定义损失函数项； l_{MSE} 表示 MSE 均方误差损失； $l_{\text{MS_SSIM}}$ 表示多尺度相似度误差损失。

MSE 计算公式为：

$$l_{\text{MSE}} = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n (f(x)_j^i - y_j^i)^2 \quad (4)$$

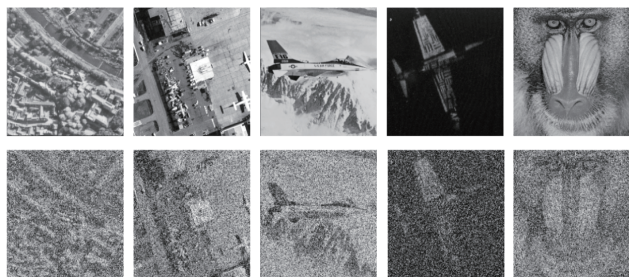
MS_SSIM 误差损失计算如公式为：

$$l_{\text{MS_SSIM}} = 1 - \prod_{m=1}^M \left(\frac{2\mu_x \mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \right)^{\beta_m} \left(\frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \right)^{\gamma_m} \quad (5)$$

式中： M 表示不同的尺度； μ_x 和 μ_y 分别表示预测图像与原始图像的均值； σ_x 和 σ_y 表示两者的标准差； σ_{xy} 表示两者的协方差；常数 c_1 和 c_2 用于确保计算稳定，防止分母过小导致除零； β_m 和 γ_m 表示相对重要性。选择 MS_SSIM 作为损失函数而非 SSIM，是因为 SSIM 只对整幅图像进行评估，可能忽略局部区域的差异。而 MS_SSIM 通过多尺度分析，能更好地捕捉图像中不同区域的失真程度，使得去噪性能更加稳定和出色。

2 数据集

训练集部分图片如图 4 所示。



注：上行为原图；下行为提取出的含特殊噪声图

图 4 部分训练集

训练集中包含在 10%、20%、30%、40% 和 50% 脉冲噪声干扰下，提取每种强度各 5 000 张具有不同纹理细节的隐藏图像，和一般噪声不同，提取出的图像噪声密度更大。图像大小为 256 px×256 px。验证集由另外 100 张不属于训练

集的灰度图像组成，图像大小同样为 256 px×256 px。所有图像的像素值均从 [0, 255] 归一化到 [0, 1]，然后用于训练和验证。

3 实验及结果

3.1 实验环境及参数配置

网络模型构建在 Tensorflow-gpu 2.6.0 平台上，实验环境为 Windows 系统（64 位）Python 3.9，利用 NVIDIA GeForce 4090（24 GB 显存）显卡进行训练和测试。实验中，为了使网络更接近最优解，优化器采用 SGD，并设置迭代次数为 100。初始学习率设定为 0.001，每 10 轮迭代后逐步递减，最终降至 0.000 1。其余参数保持默认，批次大小设为 16，图片初始输入大小为 256 px×256 px，损失函数为自定义。

3.2 实验结果及分析

本文通过大量实验客观验证了所设计网络在图像去噪方面的有效性。针对测试集，所有测试图片均未包含在训练集和验证集中。实验选用了“Lena”图、“Boat”图和“Baboon”图作为隐藏图像，“House”图作为载体图像，采用 LSB 信息隐藏策略将隐藏图像嵌入载体图像生成载密图像。在生成的载密图像上分别添加 10%、30% 和 50% 强度的脉冲噪声以进行干扰测试。最后，实验分别使用 3×3 和 5×5 的中值滤波、各参考文献中的方法以及本文设计的去噪方法对图像进行恢复处理，恢复效果如图 5 所示。结果表明，本文设计的去噪网络在图像表现出色。



(a) 直接提取



(b) 3×3 中值滤波



(c) 5×5 中值滤波



(d) 文献 [3]

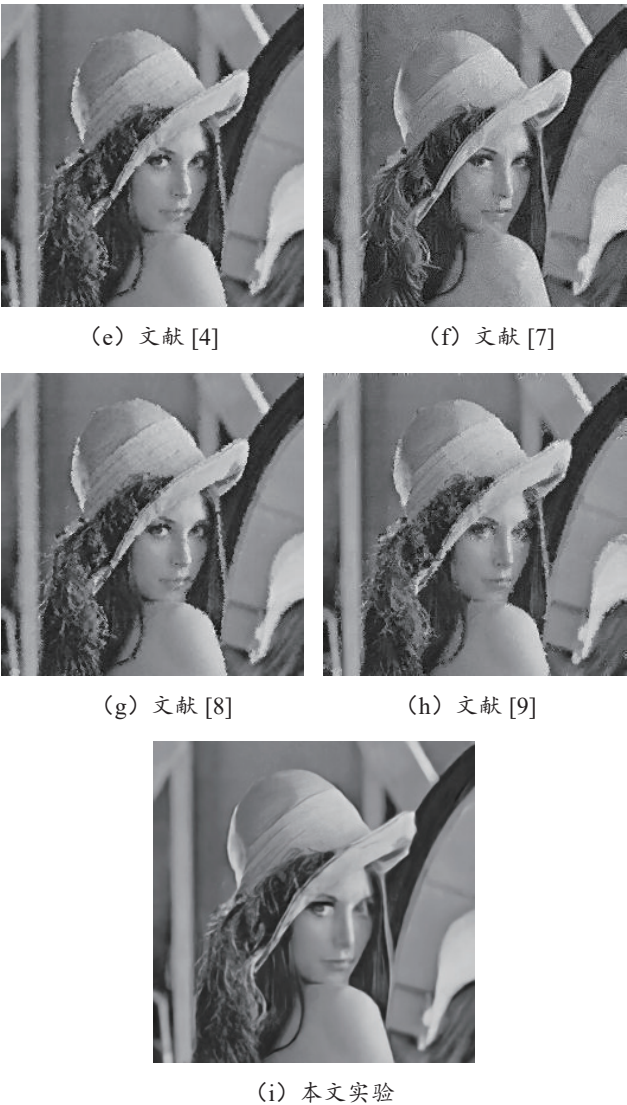


图 5 载密图像在强度为 50% 椒盐噪声干扰下的“Lena”效果图

从图 5 可以看出，本文所采用的方法和设计的网络在去噪效果上表现优异，无论从主观视觉效果还是具体的 PSNR 数值上，都明显优于其他实验方法。从对比实验图中可以看到，大多数对比方法的结果图像无法完全去除噪声或存在明显模糊，而本文方法和网络的恢复图像在细节上较为完整，且 PSNR 数值较高，展现出其去噪能力的优势。

本文选取了 100 张纹理细节不同、受噪声干扰的隐藏图像，并使用表 1 中列出的不同方法，对噪声强度从 0.1 至 0.5 的隐藏图像进行去噪恢复，随后计算其 PSNR 均值。表中加粗的数据代表各方法中的最高值。从表中对比数据可以看出，该方法具有最佳实验效果。由此可见，本文设计的去噪网络在各项指标上均优于其他对比方法，展现出最优的修复效果。

表 1 隐藏图像修复后的 PSNR 均值

载密图像添加噪声强度	10%	20%	30%	40%	50%
直接提取	17.861 3	14.937 1	13.349 5	12.200 1	11.357 1
3×3 中值滤波	26.884 4	24.179 5	22.301 9	20.902 6	19.417 9
5×5 中值滤波	25.067 2	22.406 3	20.534 4	18.896 5	17.936 7
文献 [3]	24.220 3	21.575 6	19.410 3	17.220 9	15.599 4
文献 [4]	29.110 4	25.319 6	22.804 7	21.511 4	20.078 5
文献 [7]	32.494 3	28.530 3	26.063 3	24.556 1	23.416 7
文献 [8]	30.940 7	27.116 3	24.048 9	22.921 2	21.316 3
文献 [9]	29.384 2	26.474 3	23.103 4	21.674 6	20.231 5
文献 [10]	32.189 4	29.350 1	27.497 0	26.016 8	24.573 4
文献 [11]	36.103 4	32.683 7	30.544 3	28.830 4	27.122 9
本文方法	36.508 3	33.822 1	30.884 1	29.401 4	28.661 1

4 结语

本文针对含噪声的隐藏图像，设计了一种基于 UNet 的自注意力机制的去噪网络。该网络利用 UNet 结构实现多尺度变换，并通过卷积操作进行全局去噪。主干网络嵌入了滑动自注意力机制，凭借滑动窗口的感知能力进行局部去噪。此外，网络中加入了分割与双采样操作，以便获取高分辨率图像，有效解决了图像中的棋盘效应问题。基于含噪声隐藏图像数据集的实验结果表明，该网络能够显著去除噪声，恢复图像纹理细节，验证了本文方法的优越去噪性能。然而，使用神经网络对图像进行修复工作，往往会呈现出过拟合（过平滑）现象，某些重要的细节特征可能会被抹平，因此如何设计一个能恢复出更具细节的噪声图像网络是今后研究关注的重点。

参考文献：

[1] 张笑. 图像加密技术综述[J]. 网络安全技术与应用, 2021(7): 35-36.

[2] WANG W Q, YE J Y, WANG T Q, et al. Reversible data hiding scheme based on significant-bit-difference expansion[J]. IET image process, 2017, 11(11): 1002-1014.

[3] SHARMA N, SOHI P J S, GARG B, et al. A novel multilayer decision based iterative filter for removal of salt and pepper noise[J]. Multimedia tools and applications, 2021,80: 26531-26545.

[4] SATTI P, SHARMA N, GARG B. Min-Max average pooling based filter for impulse noise removal[J]. IEEE signal processing letters, 2020,27:1475-1479.

- [5] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Berlin: Springer, 2015: 234-241.
- [6] LIU Z, HU H, LIN Y T, et al. Swin Transformer v2: scaling up capacity and resolution[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 12009-12019.
- [7] ZHANG X Q, LIANG T, TANG Z J, et al. Restoration of embedded image from corrupted stego image [J]. Signal processing, 2012, 92(7):1691-1698.
- [8] VEERAKUMAR T, SUBUDHI B N, ESAKKIRAJAN S. Empirical mode decomposition and adaptive bilateral filter approach for impulse noise removal[J]. Expert systems with applications, 2019, 121: 18-27.
- [9] NADEEM M, HUSSAIN A, MUNIR A, et al. Removal of random valued impulse noise from grayscale images using

quadrant based spatially adaptive fuzzy filter[J]. Signal processing, 2020, 169:107403.

- [10] DONG Y, ZHANG X Q, YU C Q, et al. Secret image restoration with convex hull and elite opposition-based learning strategy[J]. IEEE signal processing letters, 2023, 30: 195-199.
- [11] ZHANG X Q, LI X L, TANG Z J, et al. Noise removal in embedded image with bit approximation[J]. IEEE transactions on knowledge and data engineering, 2020, 34(3): 1359-1369.

【作者简介】

文来（2002—），男，四川资阳人，本科在读，研究方向：机器学习、计算机科学与技术。

阴明旭（1998—），男，四川广安人，硕士研究生，助教，研究方向：机器学习、数字图像处理、智能驾驶感知。

（收稿日期：2024-12-31 修回日期：2025-05-09）

（上接第 115 页）

4 结语

本文围绕复杂软件分支漏洞的智能预测问题，提出了一种基于改进支持向量机的预测方法。该方法不仅提高了预测的准确率，还降低了计算复杂度，为软件安全漏洞的实时预测提供了有力支持。展望未来，加强跨领域合作，推动软件安全漏洞预测技术在实际软件系统中的应用落地，将是提升软件安全水平的关键所在。

参考文献：

- [1] 刘亚鹏. 基于深度学习的开源软件安全漏洞预测方法研究[J]. 计算机应用文摘, 2024, 40(11):110-112.
- [2] 张珣, 张楚童, 艾孜孜·吐尔逊, 等. 一种局部时空图卷积方法及其在网络漏洞预测的应用[J]. 科技导报, 2023, 41(13):67-75.
- [3] 鲁法明, 唐梦凡, 包云霞, 等. 目标导向的多线程程序UAF漏洞预测方法[J]. 软件学报, 2023, 34(7):3043-3063.
- [4] 郑炜, 刘程远, 吴潇雪, 等. 基于知识图谱的跨项目安全缺陷报告预测方法[J]. 软件学报, 2024, 35(3):1257-1279.
- [5] 胡雨涛, 王溯远, 吴月明, 等. 基于图神经网络的切片级漏洞检测及解释方法[J]. 软件学报, 2023, 34(6):2543-2561.
- [6] 李彤彤, 王诗蕊, 张耀方, 等. 面向工控系统漏洞的多维属

性评估[J]. 计算机工程与科学, 2023, 45(2):261-268.

- [7] 罗宏芳, 王春枝. 基于贝叶斯攻击图的光网络安全态势预测研究[J]. 激光杂志, 2023, 44(8):134-138.
- [8] 丁朝晖, 张伟, 杨国玉, 等. 工业控制系统网络攻击预测技术研究[J]. 电子技术应用, 2023, 49(1):86-90.
- [9] 李秋月, 韩道军, 张磊, 等. 基于分层注意力网络和积分梯度的细粒度漏洞检测方法[J]. 计算机科学, 2024, 51(12): 326-333.
- [10] 张俊, 李山山, 李磊, 等. 基于残差门控图卷积网络的源代码漏洞检测[J]. 计算机工程与应用, 2023, 59(22):293-299.

【作者简介】

唐玲（1984—），女，湖南长沙人，硕士，讲师，研究方向：软件测试、模糊测试。

马凌（1981—），女，吉林乾安人，硕士，教授，研究方向：计算机应用技术。

喻幸（1981—），男，湖北汉川人，硕士，高级工程师，研究方向：信息技术应用。

（收稿日期：2025-02-21 修回日期：2025-06-11）