结构-语义-残差协同的图像压缩框架：面向隐私保护与高保真重建

方芯[1]，王伟[2]，贺丽君[1]，李凡[1]
（1. 西安交通大学电子与信息学部，710049，西安；2. 中国移动通信集团陕西有限公司，710077，西安）

摘要：为在极低码率条件下同时实现图像隐私保护与高保真重建，提出了一种结构-语义-残差协同的图像压缩框架。该方法从图像结构、语义与残差3个层面进行表征提取：利用边缘图作为结构信息进行主干压缩；提取去互信息处理的图像文本描述作为语义先验进行指导；将包含敏感区域信息的残差信号通过可信通道传输。在解码端，一方面结合边缘图与语义信息生成去隐私化图像，另一方面利用边缘图与残差信息恢复高保真原始图像，从而实现隐私安全与视觉保真的协同优化。结果表明：在CLIC数据集上的对比实验中，所提方法在实现高保真图像重建的同时，可节省13%～32%的码率，并能够分路生成去隐私化图像；消融实验验证了各个分支结构及约束机制的有效性，表明该框架能够在隐私保护与图像重建质量之间实现合理权衡。

关键词：图像压缩；极低码率；隐私保护；高保真
中图分类号：TP391 **文献标志码**：A
DOI：10.7652/xjtuxb202607017 **文章编号**：0253-987X(2026)07-0185-11

A Structure-Semantics-Residual Collaborative Image Compression Framework: Towards Privacy Protection and High-Fidelity Reconstruction

FANG Xin[1], WANG Wei[2], HE Lijun[1], LI Fan[1]
(1. Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;
2. China Mobile Communications Group Shaanxi Co., Ltd., Xi'an 710077, China)

Abstract: To achieve both image privacy protection and high-fidelity reconstruction at ultra-low bitrates, a structure-semantics-residual collaborative image compression framework is proposed. In this framework, representations are extracted from three distinct levels (i.e., image structure, semantics and residual): structural information is captured via edge maps for backbone compression; semantic priors are derived from image captions with mutual information removed; and residual signals containing sensitive region information are transmitted via a trusted channel. At the decoder, de-identified images are generated by integrating edge maps with semantic information, while original high-fidelity images are restored using edge maps and residual signals, thereby achieving synergistic optimization of privacy protection and visual fidelity. Results demonstrate that through the comparative experiments on the CLIC (Challenge on Learned Image Compression) dataset, the proposed method reduces bitrate by 13%—32% while maintaining high-quality reconstruction. Additionally, branch-wise generation of de-identified images is supported. Furthermore, the effectiveness of the individual branches and constraint mechanisms is verified through ablation experiments, indicating that a reasonable trade-off between privacy

收稿日期：2025-07-27。 作者简介：方芯(1998—)，女，博士生；李凡(通信作者)，男，教授，博士生导师。 基金项目：国家自然科学基金资助项目(62471376)。
网络出版时间：2025-12-03 网络出版地址：https://link.cnki.net/urlid/61.1069.t.20251203.1426.008

protection and image reconstruction quality is realized by the proposed framework.

Keywords: image compression; ultra-low bitrate; privacy protection; high-fidelity

图像数据的高效存储与传输需要图像压缩技术的有力支持。传统图像压缩[1-3]主要依赖信号变换与手工设计的编码框架，重视码率与重建图像像素失真之间的均衡。在此基础上，研究者们又进行了针对性的改进。例如，文献[4]提出了一种结合码率预测和失真优化的高精度自适应码率控制的图像压缩算法，实现了带宽受限压缩系统的定码率压缩。然而，该方法通常依赖固定的变换结构和启发式设计，难以适应复杂图像内容的非线性结构，限制了其在极低码率下的性能。

基于深度神经网络的图像压缩通过神经网络自动学习最优的特征表示和编码策略。文献[5]提出了基于变分自编码器(VAE)的压缩框架，利用可学习的分析与合成变换网络构建图像表征。文献[6]进一步引入上下文注意机制与残差块，提升了模型对图像细节与结构的保持能力。部分方法借助生成对抗网络(GAN)重建高质量图像[7-8]，弥补低比特率下的视觉损失。也有方法利用大型多模态模型(LMM)，通过模态转换和高质量图片实现极低码率下的图像语义压缩[9-11]。然而，这些方法较少关注图像内容所承载的语义层信息及其潜在的隐私风险。图像数据通常包括个体身份特征[12-13]、环境细节等敏感信息。有研究者尝试在压缩过程中分离隐私和非隐私信息，通过转换、形变或替换等手段归障敏感细节[14-19]。例如，在特征空间添加差分隐私噪声实现身份匿名[20]。

由于多数隐私保护方法与图像压缩割裂，在实际应用中难以协同优化，或严重破坏图像的结构完整性与视觉质量，或难以实现对隐私与非隐私区域的有效分离，于是研究者开始将隐私保护融合到压缩过程中[21-24]。生成式模型的发展为隐私图像压缩提供了新的研究思路。边缘图是一种直观且有效的隐私保护表征形式，其仅保留物体轮廓而不包含敏感细节，因此难以支持对原始图像的精确重建。受文本辅助边缘生成图像的启发[9]，借此二者代替图像是本文实现隐私保护压缩的重要思路。

针对上述问题，为了实现高效压缩与隐私保护的统一，本文进行了集成创新与设计优化：从结构、语义、残差 3 个层面对图像进行解耦表达。以边缘图作为压缩友好的结构表示；以去互信息的文本描述辅助语义生成，引导去隐私图像的高层语义重建；以包含敏感细节的残差信息在可信通道中传输，辅助高保真还原。在解码端，采用分路协同模型分别完成去隐私图像的合成与原图的重建。从而在压缩率、视觉质量与隐私安全之间取得良好平衡。

1 结构-语义-残差协同压缩框架

1.1 算法总述

如图 1 所示，本文提出的结构-语义-残差协同压缩框架将图像内容解耦为边缘、文本和残差，分别进行压缩与组合建模，并在解码端分路重建，同时支持隐私保护和高保真重建。具体地，在编码端，原始图像首先经边缘检测网络(HED)[25]提取边缘图像。同时利用对比语言-图像预训练模型(CLIP)[26]模型从原图中提取去互信息的文本描述。在解码过程中，边缘图与文本描述通过扩散模型生成去隐私图像。边缘图像与重建残差通过生成对抗网络生成高保真图像。

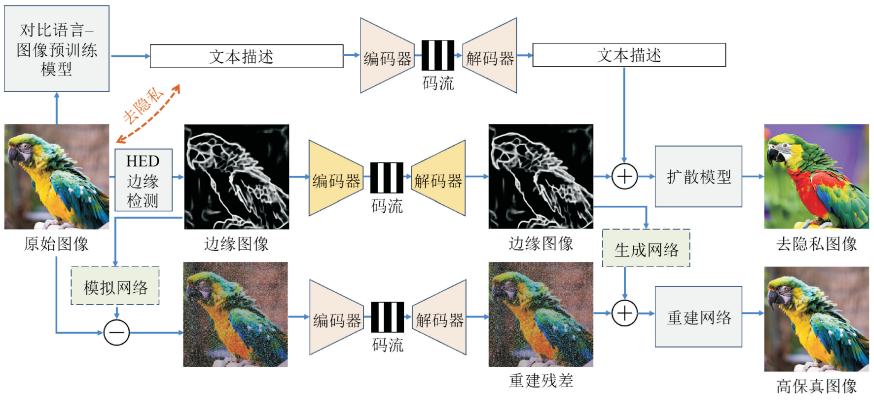


图 1 结构-语义-残差协同的图像压缩框架总览

Fig. 1 Overview of structure-semantic-residual collaborative image compression framework

1.2　基于提示反演的文本去隐私建模

为实现图像隐私内容的有效保护与可控重建，本文引入一种基于提示反演(PI)的方法对文本去互信息进行建模。该机制旨在生成文描述的同时抑制其与图像边缘的重叠性，从而有效实现图像文本描述与边缘结构的解耦。

给定一幅原始图像 $\boldsymbol{x}$，使用预训练的 CLIP 模型提取其图像嵌入 $\boldsymbol{v}=\mathrm{CLIP}_i(\boldsymbol{x})\in\mathbf{R}^d$，其中 $\mathrm{CLIP}_i(\cdot)$ 表示 CLIP 的图像编码器，$\mathbf{R}^d$ 表示 d 维实数向量空间。对于文本词元(token)序列 $\boldsymbol{t}$，其对应的文本嵌入为 $\boldsymbol{u}=\mathrm{CLIP}_t(\boldsymbol{t})\in\mathbf{R}^d$，其中 $\mathrm{CLIP}_t(\cdot)$ 表示 CLIP 的文本编码器。为在嵌入空间中接近 $\boldsymbol{v}$，则图像嵌入与文本嵌的对齐损失 L_{a} 为

$$L_{\mathrm{a}}=1-\cos(\boldsymbol{u},\boldsymbol{v})=1-\frac{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}}{\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|}\tag{1}$$

式中：$\cos(\boldsymbol{u},\boldsymbol{v})$ 表示余弦相似度计算；$\|\cdot\|$ 表示向量的范数。

该优化过程通过梯度下降搜索最优的 token 组合，使生成的文本描述与图像语义保持对齐。为保证文本的可解码性，在每次迭代中将连续嵌入向量投影到最近的合法 token 嵌入集合，从而恢复离散的 token 序列，并最终转换为自然语言文本进行无损编码。

为实现文本与结构的解耦，研究希望生成的文本表示 $\boldsymbol{u}$ 与边缘图中的内容尽可能无关。理论上希望最小化文本提示 $\boldsymbol{t}$ 与边缘图 $\boldsymbol{x}_{\mathrm{e}}$ 之间的互信息 $I(\boldsymbol{t},\boldsymbol{x}_{\mathrm{e}})$，以避免文本描述泄漏边缘结构中的隐私细节。由于 $I(\boldsymbol{t},\boldsymbol{x}_{\mathrm{e}})$ 本身难以直接计算，故使用一种基于嵌入空间余弦相似度进行近似估计。设边缘图为 $\boldsymbol{x}_{\mathrm{e}}=\mathrm{HED}(\boldsymbol{x})$，其中 $\mathrm{HED}(\cdot)$ 代表整体嵌套边缘检测网络。对应的 CLIP 嵌入为 $\boldsymbol{v}_{\mathrm{e}}=\mathrm{CLIP}_i(\boldsymbol{x}_{\mathrm{e}})$，研究希望所生成的文本嵌入 $\boldsymbol{u}$ 与结构嵌入 $\boldsymbol{v}_{\mathrm{e}}$ 的相似性尽可能低，从而实现信息分离。在解耦文本与边缘的同时，去除冗余信息和互信息。

在高维嵌入空间下，直接估计互信息的难度较大。已有许多工作表明，利用对比损失或相似度度量在一定条件下可以提供互信息的可计算上界或下界[27-28]。其中，文献[27]系统分析了多种变分下界，证明了信息噪声对比估计等基于相似度的对比损失本质上是互信息的可计算下界，给出了如下形式

$$I(\boldsymbol{X};\boldsymbol{Y})\geqslant E\left[\frac{1}{K}\sum_{i=1}^{K}\lg\left(\frac{\mathrm{e}^{f(x_i,y_i)}}{\sum_{j=1}^{K}\mathrm{e}^{f(x_i,y_j)}/K}\right)\right]\tag{2}$$

式中：$I(\boldsymbol{X};\boldsymbol{Y})$ 表示随机变量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的互信息；$E(\cdot)$ 为期望运算；$f(\cdot)$ 为打分函数；K 为锚样本 x_i 配置的 1 个正样本 y_i 与 $K-1$ 个负样本 y_j(含正样本在内)的集合大小。令 $s(x,y)=f(x,y)-\lg K$，则有

$$\frac{\mathrm{e}^{f(x_i,y_i)}}{1/K\sum_{j=1}^{K}\mathrm{e}^{f(x_i,y_j)}}=\frac{\mathrm{e}^{s(x_i,y_i)}}{\sum_{j}\mathrm{e}^{s(x_i,y_j)}}\tag{3}$$

将式(3)代回式(2)并把正负样本分开记号，便可得到与信息噪声对比估计形式等价的写法

$$I(\boldsymbol{X};\boldsymbol{Y})\geqslant E\left[\lg\frac{\mathrm{e}^{s(x,y)}}{\mathrm{e}^{s(x,y)}+\sum_{j}\mathrm{e}^{s(x,y_j)}}\right]\tag{4}$$

式(2)与式(4)仅差一个常数平移($-\lg K$)的记号变化，本质上是同一个下界。当 s 取为点积或温度缩放的余弦相似度时，即可得到常用的相似度损失与互信息下界的对应。

定义打分函数为经温度参数缩放的余弦相似度

$$s(\boldsymbol{t},\boldsymbol{x}_{\mathrm{e}})=\frac{1}{\tau}\cos(\boldsymbol{u},\boldsymbol{v}_{\mathrm{e}})\tag{5}$$

式中：τ 为缩放相似度控制相似度分布的平滑程度。由式(4)可知，降低正样本对 $(\boldsymbol{t},\boldsymbol{x}_{\mathrm{e}})$ 的相似度会使右端下界项减小，从而与降低互信息的优化目标保持一致。基于此，本文采用如下可计算的代理目标

$$\tilde{I}(\boldsymbol{t},\boldsymbol{x}_{\mathrm{e}})\propto\cos(\boldsymbol{u},\boldsymbol{v}_{\mathrm{e}})=\frac{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}_{\mathrm{e}}}{\|\boldsymbol{u}\|\,\|\boldsymbol{v}_{\mathrm{e}}\|}\tag{6}$$

式中：$\tilde{I}(\boldsymbol{t},\boldsymbol{x}_{\mathrm{e}})$ 为代理互信息，则训练目标为最小化式(6)以实现文本-边缘的去互信息建模。

此外，在高斯嵌入的简化假设下，互信息与相关系数(单位化后即余弦相似度)满足单调关系

$$I(\boldsymbol{z}_1,\boldsymbol{z}_2)=-\frac{1}{2}\lg(1-\rho^2);\ \rho=\cos(\boldsymbol{z}_1,\boldsymbol{z}_2)\tag{7}$$

式中：$\boldsymbol{z}_1$ 和 $\boldsymbol{z}_2$ 代表两个随机变量；ρ 表示相关系数。最小化 $\cos(\boldsymbol{u},\boldsymbol{v}_{\mathrm{e}})$ 亦会单调降低 $I(\boldsymbol{t},\boldsymbol{x}_{\mathrm{e}})$。综合式(2)、(4)的变分下界和式(7)的单调性分析，最小化嵌入向量之间的余弦相似度可近似达到降低互信息的效果。可见，前述目标可通过最小化文本与边缘的余弦相似度实现。

为了生成高质量的去隐私图像，如前文所述，设计了语义对齐损失函数 L_{a} 控制文本描述与原始图像语义相关性，但这步操作同样有泄漏隐私的风险。为了防止在语义对齐的过程中编码过多与隐私相关的信息，引入了一个隐私抑制项，用一种信息抽象控制的方式来抑制文本描述细节性语义。设置一组语义抽象标签 $\bar{\boldsymbol{a}}$，其中包含如“car”，“dog”等高度抽象的文本概念。生成的文本嵌入应与这些语义抽象标签 $\bar{\boldsymbol{a}}$ 相似，同时远离细节语义嵌入。通过这种约束，

鼓励提示反演将细节描述转化为抽象表达，从而促进生成文本描述的去隐私化。

将上述项引入目标函数，并最终构造提示反演的整体优化目标函数 L_{P}

$$L_{\mathrm{P}}=\lambda(1-\cos(\boldsymbol{u},\boldsymbol{v}))+\beta\cos(\boldsymbol{u},\boldsymbol{v}_{\mathrm{e}})+(1-\cos(\boldsymbol{u},\bar{\boldsymbol{a}})) \tag{8}$$

式中：λ 和 β 为权重参数；第一项为语义对齐损失，旨在保证生成图像的语义一致性并提升图像质量；第二项为结构去冗余损失，用于去除边缘结构与文本描述之间的互信息冗余；第三项为语义抽象化去隐私损失，通过将细节化描述转化为抽象表达，隐式去除其中难以直接定义的隐私语义。

综上，本文在提示反演过程中引入了语义对齐项、结构去冗余项和语义抽象化去隐私项，在保持文本表达能力的同时显著抑制其对图像细节的编码倾向，从而在语义表达与隐私保护之间实现良好平衡。

1.3 去隐私图像生成

在获得了经过提示反演优化的文本描述以及边缘图像后，对前者施加 Lempel-Ziv 无损压缩，对后者进行非线性变换编码(NTC)[29]。在解码端，采用基于稳定扩散模型的条件控制网络作为生成去隐私图像的扩散模型。条件控制网络是在稳定扩散模型框架基础上增加空间调控分支的一种条件扩散网络，其核心思想是在文本提示条件之外，引入由外部图像(解码后的边缘图 $\hat{\boldsymbol{x}}_{\mathrm{e}}$) 提取的结构信息，作为附加条件控制生成过程，从而在语义受控的同时严格保持空间布局和轮廓一致性。

如图 2 所示，在网络结构上，条件控制网络由两部分组成：一部分是稳定扩散模型主干网络(含文本编码器、U 形卷积神经网络(U-Net)以及去噪预测头)，负责根据文本提示建模扩散过程；另一部分是与之并行的条件卷积网络，用于对输入的边缘图 $\hat{\boldsymbol{x}}_{\mathrm{e}}$ 进行多层卷积特征提取，输出与主干网络同尺度的条件特征。这些条件特征在主干 U-Net 的每一层通过残差或适配模块注入，使得扩散模型在每一步迭代时不仅接收文本条件，还接收空间结构条件，进而控制输出图像的布局。

这种双条件注入的扩散生成过程可形式化为

$$\begin{cases}\boldsymbol{h}^{(l)}=\mathrm{UNet}^{(l)}(\boldsymbol{h}^{(l-1)},\boldsymbol{\gamma}^{(l)})\\ \boldsymbol{\gamma}^{(l)}=\mathrm{CondNet}(\hat{\boldsymbol{x}}_{\mathrm{e}})\end{cases} \tag{9}$$

式中：$\boldsymbol{h}^{(l)}$ 表示扩散模型在第 l 层的中间特征；$\boldsymbol{\gamma}^{(l)}$ 为由条件网络提取并注入的边缘图特征；$\mathrm{UNet}^{(l)}(\cdot)$ 代表第 l 层的 U-Net 网络；$\mathrm{CondNet}(\cdot)$ 代表条件卷积网络。

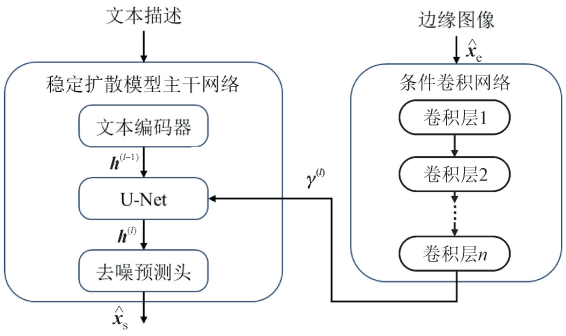


图 2 控制网络网络结构示意图

Fig. 2 Schematic diagram of the control net architecture

通过在扩散网络各层持续注入条件特征，模型得以在采样过程中被“锁定”在与输入边缘图一致的结构空间内。与传统文本-图像生成方法相比，条件控制网络的优势在于：在保持内容可控的同时，大幅增强了图像生成过程中的结构一致性，从而保证去隐私图像在空间层面保持与原图一致的人物姿态、场景轮廓等信息。在推理阶段，将前述提示反演生成的文本描述 $\boldsymbol{t}$ 与压缩解码后的边缘图 $\hat{\boldsymbol{x}}_{\mathrm{e}}$ 一同输入至扩散模型中得到去隐私图像

$$\hat{\boldsymbol{x}}_{\mathrm{s}}=\mathrm{Con}(\hat{\boldsymbol{x}}_{\mathrm{e}},\boldsymbol{t}) \tag{10}$$

式中：$\mathrm{Con}(\cdot)$ 代表条件控制网络。需要注意的是，在训练阶段，虽然使用了原图作为参考进行判别器训练，但在推理阶段，重建完全基于边缘图和去隐私文本生成，原图不再参与，从而实现真正的隐私隔离式语义图像合成。

1.4 高保真图像生成

对于无需隐私保护的解码端用户，本文设计了另一条原图重建支路，用于实现原始图像 $\boldsymbol{x}$ 的高保真还原。该重建过程以边缘图像和残差信息为输入，旨在最大限度地恢复图像的细粒度视觉信息。

在编码端，为了提取可恢复的残差信息，首先模拟基于边缘图的图像生成过程。具体地，将边缘图输入模拟原图重建的生成网络中，生成模拟重建图像 $\tilde{\boldsymbol{x}}_{\mathrm{s}}$，然后通过将原始图像与模拟重建图像作差得到残差信息 $\boldsymbol{x}_{\mathrm{r}}$，即 $\boldsymbol{x}_{\mathrm{r}}=\boldsymbol{x}-\tilde{\boldsymbol{x}}_{\mathrm{s}}$。模拟重建图像 $\tilde{\boldsymbol{x}}_{\mathrm{s}}$ 的生成过程为

$$\tilde{\boldsymbol{x}}_{\mathrm{s}}=G_{\mathrm{e}}(\boldsymbol{x}_{\mathrm{e}}) \tag{11}$$

式中：G_{e} 代表编码端的模拟生成网络，其目的在于直接模拟解码端的初步重建过程，减少重建回传占用的网络带宽以及缩短传输时延。生成模拟重建图像 $\tilde{\boldsymbol{x}}_{\mathrm{s}}$ 的损失函数 L_{s} 如下

$$L_{\mathrm{s}}=\lambda_1\|\boldsymbol{x}-\tilde{\boldsymbol{x}}_{\mathrm{s}}\|_2^2+\lambda_2\mathrm{grad}(\boldsymbol{x},\tilde{\boldsymbol{x}}_{\mathrm{s}}) \tag{12}$$

式中：第一项代表像素级别的二范式损失；第二项代图像梯度损失，确保生成图像的边缘一致性，grad(•) 为梯度损失函数；λ_1 和 λ_2 为各项权重，由经验得出。

编码端的模拟网络与解码端的生成网络具有相同的网络结构及网络参数，均由基于 U-Net[30] 的生成器和全卷积判别器组成，其中生成器的结构如图 3 所示。由 8 组跳跃连接的下采样和上采样层组成，跳跃连接可以使边缘图中的结构轮廓在浅层被提取后直接传递至图像生成阶段，防止边缘模糊、错位或消失。

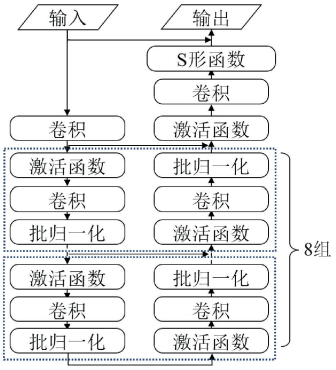


图 3　生成网络结构

Fig. 3　Architecture of generative network

残差信息 $\boldsymbol{x}_{\mathrm{r}}$ 由端到端的压缩网络[5] 进行压缩，具体的网络结构如图 4 所示。该网络由分析变换、熵编码模块和合成变换组成，中间利用熵编码模块进行压缩。编码损失函数 L_{r} 如下

$$L_{\mathrm{r}}=\lambda_3\left\|\boldsymbol{x}_{\mathrm{r}}-\hat{\boldsymbol{x}}_{\mathrm{r}}\right\|_2^2+\sum_i-\mathrm{lb}\ p_{\bar{y}_i}(\tilde{y}_i\mid\sigma_i)\tag{13}$$

式中：$\hat{\boldsymbol{x}}_{\mathrm{r}}$ 为 $\boldsymbol{x}_{\mathrm{r}}$ 的估计值；第一项代表重建的二范式损失；第二项代表码率估计，通过控制 λ_3 的变化可以实现不同失真配比下的压缩效果；$\tilde{y}_i$ 代表量化后的潜在变量；σ_i 为标准差；$p_{\bar{y}_i}$ 表示用单位均匀分布卷积的高斯分布建模量化误差。

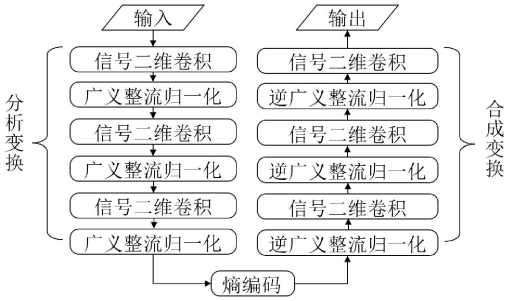


图 4　残差信息编解码网络结构

Fig. 4　Architecture of residual information codec network

在得到模拟重建图像 $\tilde{\boldsymbol{x}}_{\mathrm{s}}$ 与重建残差 $\hat{\boldsymbol{x}}_{\mathrm{r}}$ 后，将其按通道拼接并送入重建网络以生成高保真重建图像 $\hat{\boldsymbol{x}}$，重建的损失函数 L_{c} 如下

$$L_{\mathrm{c}}=\lambda_4\left\|\boldsymbol{x}-\hat{\boldsymbol{x}}\right\|_2^2+\lambda_5\sum_i\left\|\mu_1\Phi_i(\boldsymbol{x})-\mu_2\Phi_i(\hat{\boldsymbol{x}})\right\|_2^2+\lambda_6\mathrm{grad}(\boldsymbol{x},\hat{\boldsymbol{x}})\tag{14}$$

式中：第一项代表重建的二范式损失；第二项代表感知损失，$\Phi_i(\cdot)$ 是由视觉几何组网络(VGG19)[31] 提取的第 i 层特征；第三项代表图像梯度损失；λ_4、λ_5、λ_6、μ_1、μ_2 为各项损失的权重，由经验得出。

2　实验结果与分析

2.1　数据集

为验证所提出的结构-语义-残差协同压缩框架的有效性，在 CLIC2020 图像压缩基准数据集上开展实验评估。该数据集由高分辨率自然图像构成，涵盖丰富的纹理、复杂的结构细节以及多样化的拍摄场景，广泛用于学习型图像压缩方法的性能测试与比较。本文在专业图像子集上进行训练，在移动图像子集上进行测试，所有图像均采用中心裁剪并缩放至固定分辨率 512×512。

2.2　实验设置

对于编码端边缘图像的获取，采用稳定扩散模型提供的整体嵌套边缘检测器进行提取，用基于非线性变换编码的压缩方法；文本描述通过对 CLIP 模型进行微调获得，并采用 Lempel-Ziv 无损压缩算法进行编码；残差信息由编码端模拟解码结构图像后与原图之间的差值计算得到，并使用文献[5]中提出的端到端网络进行压缩。在解码端，扩散模型采用条件控制网络架构；模拟生成网络与重建网络均采用基于 U-Net 的生成器，结合基于图像块的生成对抗网络判别器的条件判别器构建生成对抗框架。语义抽象标签 $\bar{\boldsymbol{a}}$ 集合为{“person”，“car”，“room”，“animal”，“lake”，“tree”，“mountain”}，用于覆盖常见的语义类别。本文算法的去隐私图像压缩模块在配备 24 GB 显存的 GeForce RTX 3090 GPU 上进行微调训练；高保真重建模块在配备 11 GB 显存的 GeForce RTX2080Ti GPU 上进行训练。去隐私压缩部分的批大小设置为 1，$\lambda=5.0$，$\beta=1.0$；高保真压缩部分的批大小设置为 4，优化器为自适应矩估计(Adam)优化算法，初始学习率为 0.002，生成器的丢弃率设置为 0.5，感知损失选取 VGG19 网络

的 conv2_1 和 conv3_1 层的特征，$\mu_1=1$，$\mu_2=0.5$。$\lambda_1=\lambda_2=1\,000$，$\lambda_4=5\,000$，$\lambda_5=0.001$，$\lambda_6=100$。$\lambda_3$ 用于调整压缩比，取值在 0.000 1～0.000 001。

2.3 对比算法

为验证所提方法的有效性，本文选取多种具有代表性的图像压缩算法作为对比基线，涵盖传统标准和深度学习方法。传统压缩标准包括联合图像专家组标准(JPEG)[1]、JPEG2000 图像压缩标准[2]、更优便携图形格式(BPG)[32]以及多功能视频编码标准(VVC)[33]。基于深度学习的压缩方法包括全局-局部潜变量混合模型(GLLMM)[34]和基于生成式模型的隐私保护图像压缩框架(PICS)[9]。

2.4 隐私保护压缩性能

为验证本文提出的隐私保护图像压缩方法的有效性，采用自然图像质量评价指标(N_{IQE})、码率(B_{pp})、峰值信噪比(P_{SNR})以及基于 CLIP 的语义相似度(S_{CLIP})作为评价指标。其中，N_{IQE}衡量生成图像的自然感，P_{SNR}代表重建图像与原始图像在像素级的保真度，S_{CLIP}用于评估图像与其语义描述在特征空间的相似度，体现抽象语义信息的保持程度。核心思想是：在控制 N_{IQE}相近的条件下，比较不同方法在 B_{pp}、P_{SNR}和 S_{CLIP}方面的表现，以此验证本文方法在保证视觉质量和高压缩率，同时能有效去除细节隐私，并保持较高的任务支持性，同时降低隐私泄漏风险。实验的主要目标是实现高压缩率、良好的视觉感知质量和细节隐私的有效抑制，并保留对下游任务有用的高层语义信息。

为了能够直观、统一反映不同方法在隐私保护与语义保留等多维度上的综合表现，并避免仅凭单一指标无法全面反映实际隐私保护效果，参照文献[10]设计了一个综合评价指标，将 N_{IQE}、B_{pp}、P_{SNR}和 S_{CLIP}融合为统一评分。由于 3 个指标的量纲和取值范围不同，首先对其进行归一化处理，其中 N_{IQE}、B_{pp}、P_{SNR}越低表示效果越好，因此归一化后取反以保持一致性。S_{CLIP}越高越优，归一化后保持原方向。设 N'、B'、P'、C' 分别表示 N_{IQE}、B_{pp}、P_{SNR}、S_{CLIP}归一化后的得分，则综合指标的平均得分为

$$A=nN'+bB'+pP'+cC' \tag{15}$$

式中：n、b、p、c 分别为 N_{IQE}、B_{pp}、P_{SNR}、S_{CLIP} 归一化得分的权重参数，用于平衡视觉感知、语义保持、隐私保护和压缩效率的影响。本文在未引入额外偏好时采用均衡策略，即 $n=b=p=c=0.25$，以综合反映 4 个维度的性能表现。该指标越接近 1，表示整体性能越优。

表 1 为对比算法的隐私保护压缩性能。可以看出，本文算法综合得分最高，为 0.75 分，显著优于传统压缩标准(JPEG[1]、JPEG2000[2]、BPG[22]、VVC[33])和基于深度学习的方法(GLLMM[34]、PICS[9])。这一结果充分证明了本文方法在兼顾视觉质量、压缩效率、语义保持和隐私保护的多目标优化中具有明显优势。具体而言，本文算法在压缩率方面表现突出，B_{pp}最低仅为 0.029 bit/像素，显示出极高的码率压缩能力。在视觉质量方面，N_{IQE}为 11.718，优于大多数对比方法，保证了生成图像的自然视觉效果。针对隐私保护，本文方法的 P_{SNR}次低(10.03 dB)，反映了对图像细节的有效模糊与去除，从而达到保护敏感信息的目的。虽然 S_{CLIP}略低于其他方法，但依然保持了较高的语义一致性，保证了图像的抽象语义信息得以保留，证明去隐私化后的图像还能支持后续的视觉分析任务。

表 1 隐私保护压缩性能

Table 1 Privacy-preserving compression performance

算法	N_{IQE}	B_{pp}/(bit·像素$^{-1}$)	P_{SNR}/dB	S_{CLIP}	平均得分
原始图像	12.500	12.657 0	∞	1.000	
JPEG[1]	13.432	1.320 1	34.445	0.976	0.41
JPEG2000[2]	12.306	0.996 5	47.887	0.998	0.51
BPG[32]	14.558	0.1706	28.066	0.891	0.48
VVC[33]	13.461	0.089 8	27.431	0.880	0.58
GLLMM[34]	13.794	0.633 3	36.433	0.985	0.51
PICS[9]	12.514	0.031 6	9.785	0.800	0.70
本文算法	11.718	0.029 9	10.031	0.784	0.75

图 5 为不同方法在相似主观视觉质量 N_{IQE} 的重建结果。为便于观察，对整体图进行局部标记并放大。从整体感知效果来看，各方法在视觉自然性方面差异不大，均能够生成外观自然、结构完整的图像。这说明在相同感知质量约束下，各方法的表现主要体现在隐私敏感细节的处理方式以及语义信息保留程度。具体来看，对比方法在保持视觉自然性的同时，细节信息基本完整，存在敏感特征泄漏的风险；压缩率的提升在一定程度上弱化了部分纹理细节，但仍保留了大量可识别的区域。相比之下，本文提出的方法在边缘结构与场景布局上与原始图像保持了较高的一致性，同时显著弱化了高频纹理及相关细节。尽管细节被移除，但整体语义信息仍得以保持，从图像中仍可清晰辨别场景类别和主要对象，保证了视觉分析任务的可用性。

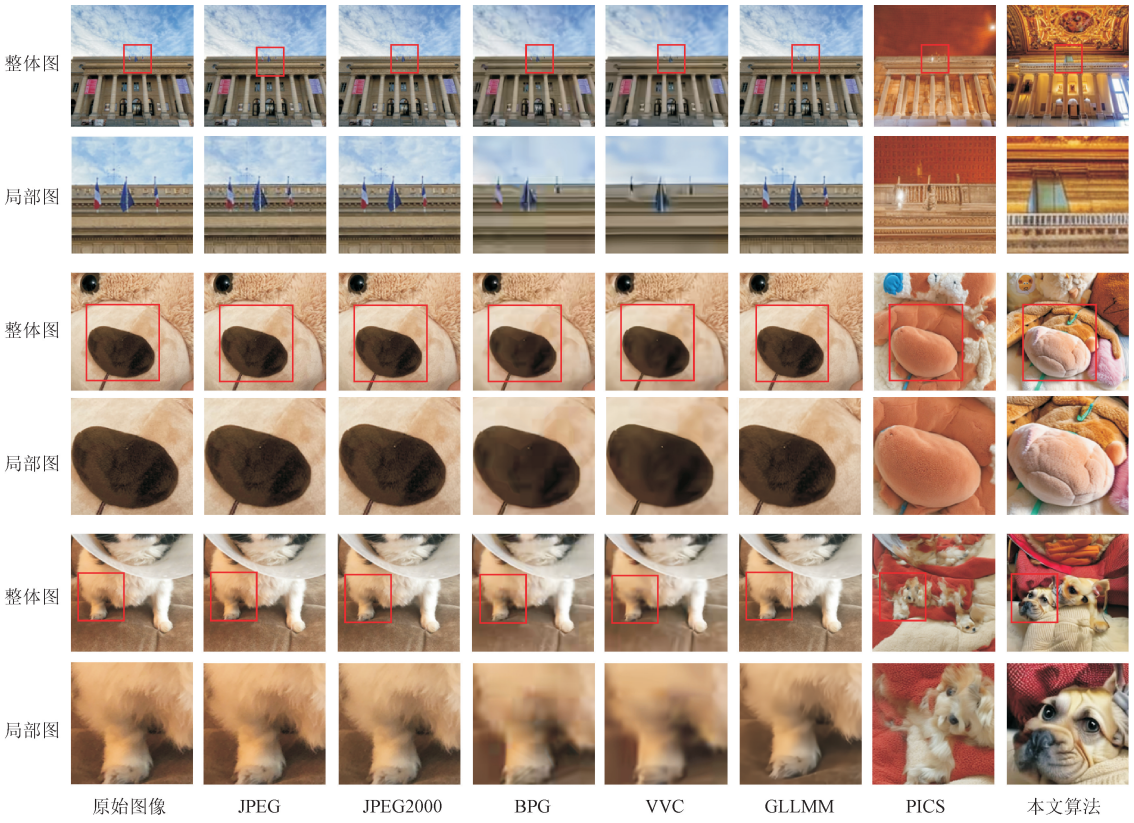


图 5　不同方法在相似自然图像质量评价指标条件下的去隐私化重建结果对比（$N_{IQE}\approx12$）

Fig. 5　Comparison of privacy-preserving reconstruction results under similar natural image quality evaluation ($N_{IQE}\approx12$)

2.5　高保真压缩性能

本节评估了基于“边缘图＋残差信息”的高保真重建性能。实验中，在相近的码率约束下，将本文方法与多种主流压缩算法进行比较，采用 P_{SNR}、结构相似性指数（S_{SIM}）和基于学习的感知图像相似度（S_{LPIP}）对重建质量进行量化，同时提供可视化结果以进行主观分析。

表 2 为不同压缩方案在多项指标下的高保真重建性能对比。可以看出，传统压缩标准在低码率下整体重建质量有限。例如，JPEG 算法在码率为 0.096 bit/像素时，仅获得 19.39 dB 的 P_{SNR} 和 0.296 的 S_{LPIP}，表现出明显的感知失真。相比之下，JPEG2000 和 BPG 算法在相似码率范围内显著提升了 P_{SNR}（分别达到 25.04、25.19 dB），但其感知质量指标 S_{LPIP} 仍维持在 0.15 以上，说明细节纹理保真度不足。VVC 算法在 S_{LPIP} 上略优于 BPG 算法，但 P_{SNR} 仅为 23.45 dB，显示出在感知与信号质量之间的权衡。基于深度学习的 GLLMM 算法[34]在码率为 0.100 6 下取得 28.39 dB 的 P_{SNR} 和 0.103 5 的 S_{LPIP}，相较传统方法在感知质量上有明显优势。同样是基于生成式模型的 PICS 算法，在码率为 0.031 6 下 P_{SNR} 仅 9.785 dB，且 S_{SIM} 也仅有 0.185 3。在相似甚至更低码率条件下，本文方法表现出压倒性优势。当仅传输边缘图与少量残差信息码率的码率为 0.068 0 时，本文方法即可实现 33.22 dB 的 P_{SNR}、0.971 的 S_{SIM} 和仅 0.007 1 的 S_{LPIP}，感知失真几乎可以忽略。

可以看出，在进行高保真重建并生成去隐私化图像的同时，本文方法可节省 13%～32%的码率。值得注意的是，与低码率区间相比，较高码率(0.235)下部分客观指标表现略有变化。这主要源于模型的损失设计与解码策略在不同码率区间的侧重点差异。具体而言，本模型的感知与语义损失项以及扩散解码策略主要针对低码率场景进行优化。在码率为 0.235 bit/像素时，该配置更强调感知质量与语义一致性，因此在 P_{SNR}、S_{SIM} 等失真指标上会出现轻微波动。

表 2 高保真重建性能

Table 2 High-fidelity reconstruction performance

算法	边缘码率/(bit·像素$^{-1}$)	残差码率/(bit·像素$^{-1}$)	重建码率/(bit·像素$^{-1}$)	P_{SNR}/dB	S_{SIM}	S_{LPIP}
JPEG[1]			0.096	19.388	0.550	0.296
JPEG2000[2]			0.079	25.038	0.672	0.177
BPG[32]			0.067	25.192	0.704	0.153
VVC[33]			0.064	23.453	0.676	0.102
GLLMM[34]			0.100	28.390	0.801	0.103
PICS[9]			0.031	9.785	0.185	0.525
本文算法(高残差码率)	0.025	0.209	0.235	31.853	0.962	0.016
本文算法(低残差码率)	0.025	0.042	0.068	33.221	0.971	0.006

图 6 为不同方法在相似码率范围内的重建结果可视化对比。整体来看，传统方法(如 JPEG 和 JPEG2000)在低码率下容易产生明显的平滑伪影和细节缺失，纹理区域表现尤为不足。BPG 和 VVC 在结构保持方面略有改善，但在复杂纹理和边缘区域仍存在一定模糊感。基于深度学习的 GLLMM[34]在感知质量上优于传统方法，能够恢复大量细节。相比之下，本文提出的基于“边缘图＋残差”重建策略在结构一致性和细节还原上表现更为出色，即使在极低码率下也能保持较高的感知保真度，边缘轮廓清晰、纹理自然，与定量指标的对比结果相一致。

图 6 不同方法在相近码率范围内的高保真重建结果对比($B_{\mathrm{pp}}\approx0.1$)

Fig. 6 Comparison of high-fidelity reconstruction of different methods under similar bitrates($B_{\mathrm{pp}}\approx0.1$)

2.6　消融实验

为了进一步验证本文方法在去隐私效果上的有效性，设计了消融实验，考察在文本生成过程中是否施加去隐私约束对重建结果的影响。分别在保留去隐私约束和去除该约束的条件下生成对应的图像，并通过定量指标及平均得分进行对比，结果见表 3。

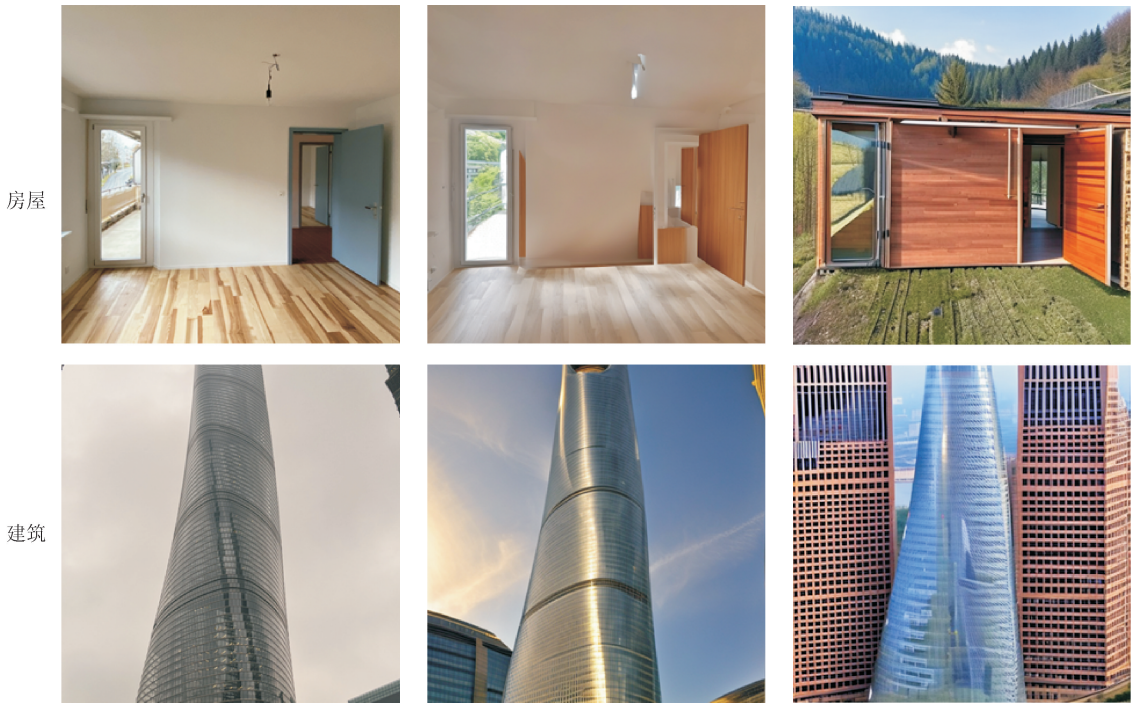
表 3　去隐私消融实验结果对比

Table 3　Comparison of privacy removal ablation experiments

方案	N_{IQE}	B_{pp}/(bit·像素$^{-1}$)	P_{SNR}/dB	S_{CLIP}	平均得分
原始图像	12.500	12.657	∞	1.000	
未去隐私	12.175	0.028	10.683	0.797	0.38
去隐私	11.718	0.029	10.031	0.784	0.50

表 3 表明，在生成的图像无参考视觉质量方面，是否增加去隐私约束的影响并不大，而增加去隐私约束的生成效果更好，可能是因为施加了一组抽象语义标签，在边缘生成图像的过程中文本提示的指向性更强，生成的内容不会被无关的文本内容干扰，因此自然图像质量评价指标更低。由 P_{SNR} 和 S_{CLIP} 可以看出，未施加去隐私约束时，生成图像中会出现与原始图像高一致性的细节（P_{SNR} 相比去隐私的高 0.652 dB），可能导致敏感信息泄漏；而在施加去隐私约束后，生成图像在保持了一定程度的语义一致性（S_{CLIP} 仅下降 0.013）的同时，有效去除了与隐私相关的细节特征，从而降低了隐私泄漏风险。这一结果验证了本文提出的隐私约束机制在保护敏感信息方面的关键作用。

除了定量分析去隐私约束的效果外，本文算法还给出了在“文本＋边缘”的去隐私分支上消融实验的可视化结果。如图 7 所示，当未引入去隐私约束时，生成图像虽然在整体结构上与原图保持一致，但仍保留了部分敏感细节，例如图中的房间布局风格、建筑外表等，存在隐私泄漏的风险。而在加入去隐私约束后，模型在生成过程中更倾向于弱化隐私相关的细节，仅保留必要的语义信息与场景结构，并且加入了一些混淆内容，使得输出图像在视觉感知上不易恢复身份特征，同时仍保持语义一致性，满足下游分析任务（如分类、检测）的需求。这一结果进一步说明，去隐私约束不仅提升了隐私保护能力，还实现了在隐私与可用性之间的有效平衡。



(a)原始图像　　(b)无去隐私约束　　(c)有去隐私约束

图 7　去隐私消融实验结果

Fig. 7　Ablation study results on privacy removal

3 结 论

本文针对视觉通信中隐私泄漏风险与高效传输需求，提出了一种基于结构-语义-残差协同的图像压缩方法。所提框架通过“边缘、文本、残差信息”三分支协同传输，分别实现隐私保护与高保真重建两类场景的需求。

(1)在隐私保护分支中，边缘结构结合去互信息文本生成的图像，显著削弱了细节隐私并保留语义一致性，有效支撑下游分析任务。

(2)在高保真重建分支中，通过引入残差信息并采用基于 GAN 的生成模型，本文方法在低码率下实现了接近原始图像的重建质量，兼顾压缩效率与视觉保真度。

(3)实验验证了本文所提出的方法能在节省 13%～32%码率进行高保真重建的同时生成去隐私化图像，并通过消融实验进一步证明去隐私约束机制的有效性。

参考文献：

[1] Wallace G K. The JPEG still picture compression standard [J]. Communications of the ACM, 1991, 34(4): 30-44.

[2] Christopoulos C, Skodras A, Ebrahimi T. The JPEG2000 still image coding system: an overview [J]. IEEE Transactions on Consumer Electronics, 2000, 46(4): 1103-1127.

[3] Sullivan G J, Ohm J R, Han W J, et al. Overview of the high efficiency video coding (HEVC) standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(12): 1649-1668.

[4] 宋蓓蓓，何帆，马穗娜，等. 一种高精度的自适应码率控制图像压缩算法 [J]. 西安交通大学学报，2022，56(2)：198-206.
Song Beibei, He Fan, MA Suina, et al. An image compression algorithm with high-precision adaptive rate control [J]. Journal of Xi'an Jiaotong University, 2022, 56(2): 198-206.

[5] Ballé J, Laparra V, Simoncelli E P. End-to-end optimized image compression [EB/OL]. (2017-03-03) [2025-07-27]. https://arxiv.org/abs/1611.01704.

[6] Cheng Zhengxue, Sun Heming, Takeuchi M, et al. Learned image compression with discretized Gaussian mixture likelihoods and attention modules [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 7936-7945.

[7] Iwai S, Miyazaki T, Sugaya Y, et al. Fidelity-controllable extreme image compression with generative adversarial networks [C]//2020 25th International Conference on Pattern Recognition (ICPR). Piscataway, NJ, USA: IEEE, 2021: 8235-8242.

[8] Agustsson E, Tschannen M, Mentzer F, et al. Generative adversarial networks for extreme learned image compression [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019:221-231.

[9] Lei E, Uslu Y B, Hassani H, et al. Text+ sketch: image compression at ultra low rates [EB/OL]. (2023-06-01) [2025-07-27]. https://arxiv.org/abs/2306.00040.

[10] Li C, Lu G, Feng D, et al. MISC: ultra-low bitrate image semantic compression driven by large multimodal model [J]. IEEE Transactions on Image Processing, 2025, 34(1): 335-349.

[11] 张云飞，贺丽君，王子溪，等. 面向机器视觉任务的多尺度图像特征压缩算法 [J]. 西安交通大学学报，2023，57(12)：1-10.
Zhang Yunfei, He Lijun, Wang Zixi, et al. Image multi-scale feature compression algorithm for machine vision tasks [J]. Journal of Xi'an Jiaotong University, 2023, 57(12): 1-10.

[12] 马彩霞，贾春福，蔡智鹏，等. 面向对抗样本干扰的人脸识别隐私保护方案 [J]. 西安电子科技大学学报，2025，52(2)：190-200.
Ma Caixia, Jia Chunfu, Cai Zhipeng, et al. Privacy-preserving face recognition against adversarial sample perturbations [J]. Journal of Xidian University, 2025, 52(2): 190-200.

[13] 周浩，戴华，杨庚，等. 基于生物特征识别的隐私保护可验证联邦学习 [J]. 计算机学报，2025，48(8)：1848-1869.
Zhou Hao, Dai Hua, Yang Geng, et al. Privacy-preserving verifiable federated learning based on biometric recognition [J]. Chinese Journal of Computers, 2025, 48(8): 1848-1869.

[14] Rakhmawati L. Image privacy protection techniques: a survey [C]//TENCON 2018: 2018 IEEE Region 10 Conference. Piscataway, NJ, USA: IEEE, 2018: 0076-0080.

[15] Yu Jun, Zhang Baopeng, Kuang Zhengzhong, et al. iPrivacy: image privacy protection by identifying sensitive objects via deep multi-task learning [J]. IEEE

Transactions on Information Forensics and Security, 2017, 12(5): 1005-1016.

[16] 闫光辉，刘婷，张学军，等. 抵御背景知识推理攻击的服务相似性位置 k 匿名隐私保护方法 [J]. 西安交通大学学报，2020，54(1)：8-18.
Yan Guanghui, Liu Ting, Zhang Xuejun, et al. Service similarity location k anonymity privacy protection scheme against background knowledge inference attacks [J]. Journal of Xi'an Jiaotong University, 2020, 54(1): 8-18.

[17] Liu Chi, Zhu Tianqing, Zhang Jun, et al. Privacy intelligence: a survey on image privacy in online social networks [J]. ACM Computing Surveys, 2023, 55(8): 161.

[18] 李璇，李德华，杨智，等. 图像中人脸隐私度的定量评估研究 [J]. 计算机与数字工程，2019，47(10)：2550-2555.
Li Xuan, Li Dehua, Yang Zhi, et al. Quantifying privacy levels of faces in images [J]. Computer & Digital Engineering, 2019, 47(10):2550-2555.

[19] Shan S, Wenger E, Zhang Jiayun, et al. Fawkes: protecting privacy against unauthorized deep learning models [C]//Proceedings of the 29th USENIX Conference on Security Symposium. USA: USENIX Association, 2020: 1589-1604.

[20] 周治平，钱新宇. 一种面向深度神经网络的差分隐私保护算法 [J]. 电子与信息学报，2022，44(5)：1773-1781.
Zhou Zhiping, Qian Xinyu. Differential privacy algorithm under deep neural networks [J]. Journal of Electronics & Information Technology, 2022, 44(5): 1773-1781.

[21] Zhang Yaofang, Fang Yuchun, Cao Yiting, et al. RB-GAN: realistic-generation and balanced-utility GAN for face de-identification [J]. Image and Vision Computing, 2024, 141: 104868.

[22] Wu Yifan, Yang Fan, Ling Haibin. Privacy-protective-GAN for face de-identification [EB/OL]. (2018-06-23) [2025-07-27]. https://arxiv.org/abs/1806.08906.

[23] Ruchaud N, Dugelay J L. JPEG-based scalable privacy protection and image data utility preservation [J]. IET Signal Processing, 2018, 12(7): 881-887.

[24] Zhang Bo, Xiao Di, Li Xiuling, et al. Privacy-preserving image compressed sensing by embedding a controllable noise-injected transformation for IoT devices [J]. Signal Processing, 2023, 210: 109055.

[25] Xie Saining, Tu Zhuowen. Holistically-nested edge detection [C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2015: 1395-1403.

[26] Bianchi F, Attanasio G, PISONI R, et al. Contrastive language-image pre-training for the Italian language [EB/OL]. (2021-08-19)[2025-07-27]. https://arxiv.org/abs/2108.08688.

[27] Poole B, Ozair S, van den Oord A, et al. On variational bounds of mutual information [C]//Proceedings of the 36th International Conference on Machine Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2019: 5171-5180.

[28] Belghazi M I, Baratin A, Rajeswar S, et al. MINE: mutual information neural estimation [EB/OL]. (2021-08-14)[2025-07-27]. https://arxiv.org/abs/1801.04062.

[29] Ballé J, Chou P A, Minnen D, et al. Nonlinear transform coding [J]. IEEE Journal of Selected Topics in Signal Processing, 2021, 15(2): 339-353.

[30] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C]// Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham, Germany: Springer International Publishing, 2015:234-241.

[31] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10)[2025-07-27]. https://arxiv.org/abs/1409.1556.

[32] Bellard F. BPG image format [EB/OL]. [2025-07-26]. https://bellard.org/bpg/.

[33] ISO/IEC 23090-3: 2024 Information technology-coded representation of immersive media: part 3 versatile video coding [S].

[34] Fu Haisheng, Liang Feng, Lin Jianping, et al. Learned image compression with Gaussian-Laplacian-logistic mixture model and concatenated residual modules [J]. IEEE Transactions on Image Processing, 2023, 32: 2063-2076.

（编辑　亢列梅）