DOI:10.19651/j.cnki.emt.2416948

基于 Swin Transformer 的图像语义通信系统*

孙洋舟[1,2] 严天峰[1,2,3] 孙文灏[1,2] 汤春阳[3] 王映植[1,2]
(1.兰州交通大学电子与信息工程学院 兰州 730070;2.甘肃省无线电监测及定位行业技术中心 兰州 730070;
3.丝路梵天(甘肃)通信技术有限公司 兰州 730030)

摘 要:语义通信是一种旨在传递语义信息的通信方式,其通过可以有效的减少冗余和传输数据量等特点。目前语义通信的研究仅处于起步阶段,更多的理论研究有助于推动语义通信系统的真正实施。实现语义通信的核心技术端到端信源信道联合编码在过去几年中取得了长足的进步,语义图像也得到了发展。为解决计算效率过低、语义特征提取不足等问题,本文设计了一款新的神经网络 JSCC。具体而言,受 Swin Transformer 在视觉任务中的优异表现的启发,首次将 Swin-Transformer 模块与残差网络相结合,设计出基于 Swin Transformer 的图像语义通信系统。为了解决传统的 CNN 对图像特征提取效率欠佳等问题,引入注意力残差网络模块初步提取图像语义特征,然后通过 Swin Transformer 进一步对图像语义特征进行提取。通过实验的结果的验证,与已有方案相比,本文所提出的方案在 PSNR 取得了高于 2 dB 的性能提升,在 MS-SSIM 性能上取得了 5%以上的性能提升。

关键词:语义通信;Swin Transformer;信源信道联合编码
中图分类号:TN914 **文献标识码**:A **国家标准学科分类代码**:510.5025

Image semantic communication system based on swin transformer

Sun Yangzhou[1,2] Yan Tianfeng[1,2,3] Sun Wenhao[1,2] Tang Chunyang[3] Wang Yingzhi[1,2]
(1. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;
2. Gansu Province Radio Monitoring and Positioning Industry Technology Center, Lanzhou 730070, China;
3. Silk Road Brahma Communication Technology, Lanzhou 730030, China)

Abstract: Semantic communication is a type of communication designed to convey semantic information, which is characterized by the fact that it can effectively reduce redundancy and the amount of transmitted data. Currently the research on semantic communication is only in its infancy, and more theoretical research can help to promote the real implementation of semantic communication systems. The core technology for realizing semantic communication, end-to-end joint source channel coding, has made great progress in the past few years, and semantic images have also been developed. In order to solve the problems of computational inefficiency and insufficient semantic feature extraction, a new neural network JSCC is designed in this paper. Specifically, inspired by the excellent performance of Swin Transformer in visual tasks, the Swin-Transformer module is combined with residual networks for the first time, and a Swin Transformer-based image semantic communication system. In order to solve the problems such as the poor efficiency of traditional CNN for image feature extraction, the attention residual network module is introduced to extract the image semantic features initially, and then the image semantic features are further extracted by Swin Transformer. Through the verification of the experimental results, compared with the existing schemes, the proposed scheme in this paper achieves higher than 2 dB performance improvement in PSNR and more than 5% performance improvement in MS-SSIM performance

Keywords: semantic communication;Swin Transformer;source channel joint coding

0 引 言

随着信息时代的到来,信息逐渐成为人们生活的一部分。然而,传统的语法信息压缩和传输正接近香农信息论的极限,目前的容量和其他性能改进方法不能支持未来中通信的持续发展。早在 1948 年,香农在他的论文中提到了

收稿日期:2024-09-20
* 基金项目:甘肃省科技重大专项项目(22ZD6GA041)、甘肃省拔尖人才项目(6660030102)资助

信息传播的 3 个层次，即技术问题、语义问题和效应问题。但是，根据当时的技术水平，无法很好地解决语义和效果的问题。随着人工智能的兴起，自然语言处理和计算机视觉技术不断改进，人们可以从信息中提取语义，这为下一代通信系统提供了可能[1]。

在 20 世纪 40 年代，Weaver 就已经开始研究语义，开创了语义通信的研究先河，并进一步刻画了交际的句法、语义和语用特征。Carnap 在 1953 年提出了一个具有命题逻辑的语义信息理论，他们还使用了语义信息的概率度量。Barwise 和 Perry 将语义信息理论扩展到情景逻辑，Floridi 解决了矛盾无法正确测量的问题。D′Alfonso 采用了真实相似性的概念来量化语义信息，以支持更广泛的用例。但是由于条件不足，语义通信一直处于理论阶段。

目前，人工智能也得到了快速的发展，使得语义知识的提取成为可能[2-3]，并以此发展出语义通信技术来探索通信中的语义问题[4]。随着信源信道联合编码（joint source-channel coding，JSCC）在通信系统上的应用与优化。对于图像传输任务，目前使用卷积神经网络（convolutional neural network，CNN）骨干的神经网络 JSCC 及其变体可以产生超越基于分离的经典方法的端到端图像传输性能。语义通信也展现出来了更好的前景[5]。在基于深度学习的端到端通信和自然语言处理技术的推动下，语义编码（解码）和信道编码（解码）可以通过深度神经网络来实现[6]。

Bourtsoulatze 等[7]将深度学习网络引入到图像压缩和通信系统中，提出了一种信源信道联合编码模型深度信源信道联合编码（deep joint source-channel coding，Deep-JSCC）。Lee 等[8]考虑了一个简单的图像传输场景，一个物联网设备发送图像到服务器完成识别任务，使用深度神经网络作为信道编码器和解码器来实现 JSCC。Xu 等[9]在有 SNR 反馈时考虑点对点的图像传输系统，将广泛应用于计算机视觉的注意力机制融入特征提取。随着 Transformer 在视觉任务中的发展[10-11]，考虑到图像数据有更多的空间冗余，Hu 等为图像分类任务提出了资源节约型特征提取模型，在编码过程中，使用带有视觉 transformer（vision transformer，ViT）结构的掩码自编码器[12]（masked auto-encoder，MAE），并采用一个对称编码解码器结构，MAE 可以从部分观测中重构一个图像。受 ViT 在处理图像干扰方面的鲁棒性启发[13]，Yoo 等[14]提出了一种基于 ViT 的语义通信模型 SemViT，并对结果进行了深入分析，以了解图像语义通信系统的工作原理以及 ViT 的优势所在。后续对 ViT 网络进行改良，在图像处理任务中有很大的提升[15-16]。基于 ViT 架构，Liu 等[17]提出的 Swin Transformer 在视觉任务中更加出色，与 ViT 相比，在减少计算量的同时，拥有更好的适应性和建模能力。更多的理论研究有助于推动语义通信系统的真正实施，因此本文基于 Swin Transformer 提出了新的语义通信系统，探讨 Swin Transformer 在图像的语义通信中的性能。本文的主要工作如下：

1）在经典语义通信系统的基础上，本文基于 Swin Transformer 提出了一种更为高效准确的语义通信系统，在减少计算量的同时，有效的提升了语义特征提取的效率。

2）简单的 CNN 网络对语义特征提取有限，因此本文设计了一种效率更高的注意力机制[18]与残差模块[19]相结合的架构。通过引入通道注意力机制和多尺度特征融合[20-21]，进一步优化网络结构，设计出一款注意力残差网络模块，从而提高语义特征提取的效率。

3）在引入峰值信噪比和多尺度结构相似度分析指标的同时，在 AWGN、Rayleigh、Rician 信道下，基于 Swin Transformer 的语义通信系统的性能指标有较大提升。

1　基础理论

1.1　窗口多头注意力机制

Swin Transformer 网络架构在 ViT 网络中引入了窗口多头注意力机制（windows multi-head self-attention，W-MSA）和移动窗口多头注意力机制（shifted windows multi-head self-attention，SW-MSA），通过对特征图划分为不同的独立区域，在小区域内进行提取特征，在减少计算量的同时，提升语义特征提取的效率。

如图 1 所示，与 MSA 区别在于 W-MSA 将特征图按照 M×M 划分为大小相同的独立子区域，然后在每个子区域内做多头注意力机制（multi-head self-attention，MSA）。

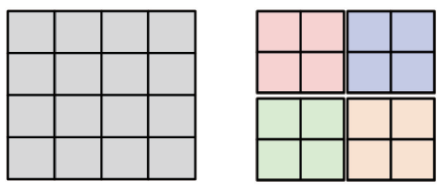

图 1　W-MSA 特征图

Fig. 1　W-MSA characterization diagram

如图 1 所示，与 MSA 区别在于 W-MSA 将特征图按照 M×M 划分为大小相同的独立子区域，然后在每个子区域内做 MSA。MSA 表达式为：

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}+B\right)\boldsymbol{V} \tag{1}$$

$$MSA(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = Concat(head_1,\cdots,head_h)W^{\mathrm{O}} \tag{2}$$

$$head_i = Attention(\boldsymbol{Q}\boldsymbol{W}_i^{\boldsymbol{Q}},\boldsymbol{K}\boldsymbol{W}_i^{\boldsymbol{K}},\boldsymbol{V}\boldsymbol{W}_i^{\boldsymbol{V}}) \tag{3}$$

其中，$\mathbf{W}_i^{Q} \in \boldsymbol{R}^{d_{bmodel}\times d_k}$，$\mathbf{W}_i^{K} \in \boldsymbol{R}^{d_{bmodel}\times d_k}$，$\mathbf{W}_i^{V} \in \boldsymbol{R}^{d_{bmodel}\times d_k}$，每个 headi 的 Qi、Ki、Vi 可以通过 $\mathbf{W}_i^{Q}$、$\mathbf{W}_i^{K}$、$\mathbf{W}_i^{V}$ 的映射得到，$\mathbf{W}_i^{O} \in \boldsymbol{R}^{hd_v\times d_{bmodel}}$，通过 WO 将拼接提取到的特征进行融合。

W-MSA 虽提高了子区域中的特征提取的效率，但同时也隔绝了各个子区域之间的信息传递，而 SW-MSA 可以让相邻子区域内的信息进行相互的传递。如图 2 所示，

SW-MSA 先进行窗口偏移,然后进行区域划分获得新的 M×M 子区域。W-MSA 与 SW-MSA 的组合使用,可以减少计算量和提升特征提取效率的同时,杜绝了各个子区域间信息的不传递。

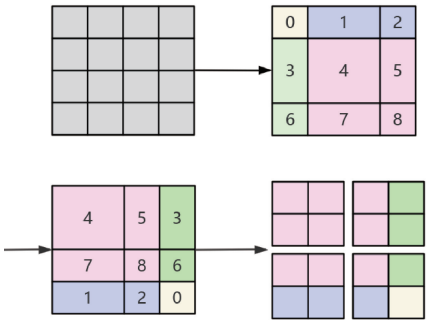

图 2 SW-MSA 特征图

Fig. 2 SW-MSA characterization map

1.2 Swin Transformer Block

如图 3 所示,本文采用 Swin Transformer 模块作为语义特征提取的核心模块,其包括 Swin Transformer Block 和 Patch Embbeding 两部分构成。首先,Swin Transformer 通过自注意力机制能够有效的提取图像语义特征,可增强模型的特征提取能力。其次,Transformer 具有全局的感受野,Swin Transformer 属于 Transformer 在图像处理中的变体,在图像处理任务中有优异的表现。Swin Transformer Block 通过将 Transformer 模块中的 MSA 替换为基于移位窗口的注意力机制,连续的自注意层之间的窗口分区的移位提供了它们之间的连接,显著增强了建模能力。最后,Swin Transformer 具有较高的灵活性和扩展性,模块化的设计允许它以与其他神经网络层或模块集成。

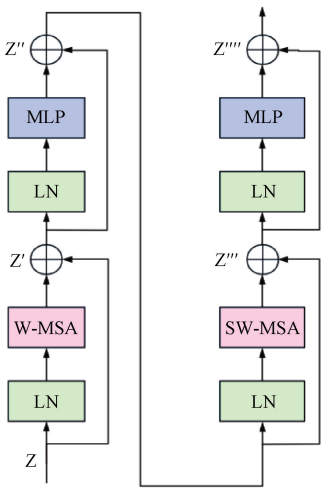

图 3 Swin-Transformer Block 结构图

Fig. 3 Swin-Transformer Block structure diagram

Swin-Transformer Block 原理表达式如下:

$$\boldsymbol{Z}' = \mathrm{W-MSA}(\mathrm{LN}(\boldsymbol{Z})) + \boldsymbol{Z} \tag{4}$$

$$\boldsymbol{Z}'' = \mathrm{MLP}(\mathrm{LN}(\boldsymbol{Z}')) + \boldsymbol{Z}' \tag{5}$$

$$\boldsymbol{Z}''' = \mathrm{SW-MSA}(\mathrm{LN}(\boldsymbol{Z}'')) + \boldsymbol{Z}'' \tag{6}$$

$$\boldsymbol{Z}'''' = \mathrm{MLP}(\mathrm{LN}(\boldsymbol{Z}''')) + \boldsymbol{Z}''' \tag{7}$$

其中,$\boldsymbol{Z}$ 是输入 Swin-Transformer Block 模块的特征矩阵,$\boldsymbol{Z}''''$是 Swin-Transformer Block 模块输出的特征矩阵,W-MSA 与 SW-MSA 为基于窗口的注意力机制模块,LN 为归一化层,MLP 为多层感知机。

1.3 注意力残差模块

ResNet 是由 He 等提出的经典卷积神经网络,如图 4 所示,为残差网络的残差连接方式,其由卷积层、批量归一化层(batch normalization,BN)、ReLU 激活函数构成。残差结构与 BN 层的结合成功解决了解决梯度消失或者梯度爆炸问题,其网络的特征提取和精确度也有明显的提升。

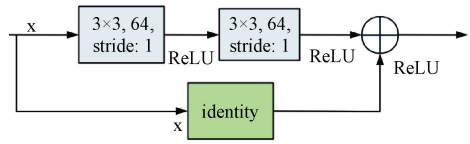

图 4 残差模块结构图

Fig. 4 The residual module structure diagram

其原理表达式为:

$$y = F(\boldsymbol{x}, \{W_i\}) + W_s\boldsymbol{x} \tag{8}$$

其中,$F(x, \{W_i\}+x)$为残差网络待学习的映射,x 为输入残差网络的矩阵,W_s 将输入矩阵与残差网结构输出矩阵进行维度匹配,通常使用 1×1 卷积核进行维度匹配。

如图 5 所示,本文通过在两个残差模块之间引入通道注意力机制,改进残差网络,使得模型更好地关注局部语义特征和细节语义特征。

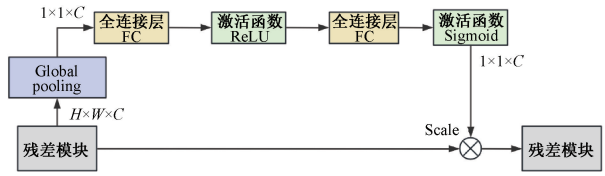

图 5 注意力残差模块结构图

Fig. 5 Attention residual module structure diagram

由残差模块初步提取语义特征的时候,容易忽略局部特征而丢失重要的语义特征,而注意力机制可以有效的改良这一缺陷[22]。通道注意力原理如式(9)~(11)所示。

$$\boldsymbol{z} = F_{global\ pooling}(\boldsymbol{x}) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{x}(i,j) \tag{9}$$

$$\boldsymbol{s} = \sigma(W_2\delta(W_1\boldsymbol{x})) \tag{10}$$

$$\boldsymbol{x}' = F_{scale}(\boldsymbol{s}, \boldsymbol{x}) \tag{11}$$

其中,$F_{\mathrm{global\ pooling}}$ 为全局池化操作,W_1 和 W_2 为全连接层的映射,δ 为 ReLU 激活函数,σ 为 sigmoid 激活函数,F_{scale} 将获得的通道权重矩阵 $\boldsymbol{s}$ 与输入的特征矩阵 $\boldsymbol{x}$ 进行矩阵相乘。通过通道注意力机制,学习各通道的权重,尽可能的保留图像中语义特征。

1.4 信道仿真

本文通过对 Rayleigh 信道、Rician 信道、AWGN 信道

进行仿真来模拟通信过程中信道产生的噪声。

Rayleigh 信道的数学模型如下：

$$y(t)=h(t)\cdot x(t) \tag{12}$$

$$h(t)=A_h e^{j\varphi_h(t)} \tag{13}$$

其中，$h(t)$是 Rayleigh 信道的复数形式，其幅度 A_h 和相位 $\varphi_h(t)$都是随机的且相互独立的，$x(t)$为传入信道的信号。

AWGN 信道的数学表达式可以简化为：

$$y(t)=x(t)+n(t) \tag{14}$$

$n(t)$是加性高斯噪声，其在每个时间点 t 上都是独立且服从高斯分布的。高斯噪声 $n(t)$具有均值为零，方差为一定值，且噪声功率谱密度是常数，表示为 N_0。

Rician 信道的数学模型为：

$$h_{\text{Richan}}=\sqrt{\frac{K}{K+1}}+\sqrt{\frac{K}{K+1}}h_{Rayleigh} \tag{15}$$

其中，K 是反映信号衰落程度的参数。

2　基于 Swin Transformer 的图像语义通信系统

如图 6 所示，为本文所提出的基于 Swin Transformer 的图像语义通信系统（image semantic communication system based on swin transformer，SemSWT）。

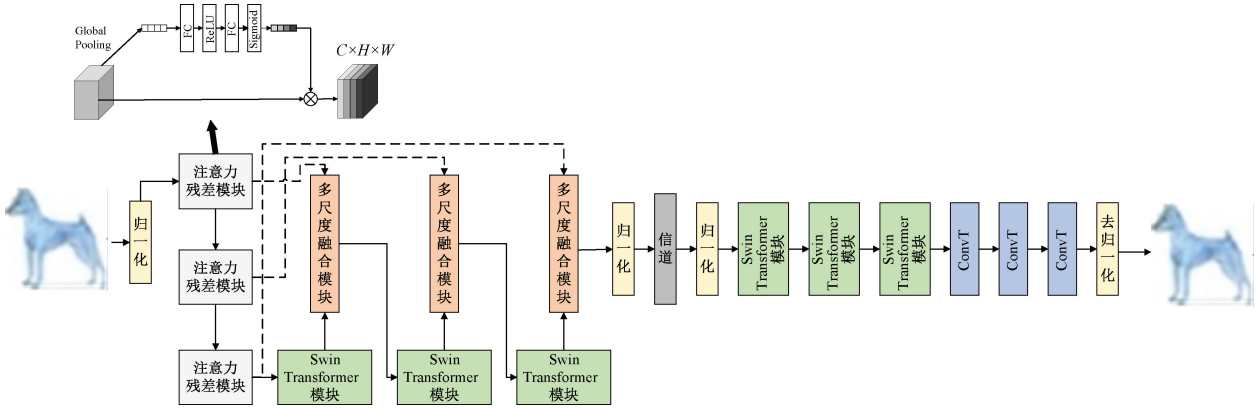

图 6　SemSWT 框架图

Fig. 6　SemSWT framework diagram

其遵循语义通信系统的典型自动编码器设计。它包含联合信源信道编码器、信道、信源信道联合译码器 3 个部分组成。通过仿真 3 种信道来验证本文模型的性能，编码器、译码器如下文所示。

2.1　信源信道联合编码器

为了提高模型的语义特征提取效率，本文设计的编码器由 1.3 节中所设计的 ResNet 网络和 Swin Transformer 网络相结合。

如图 7 所示，为本文 SemSWT 系统的编码器结构图，编码器由归一化层、3 组注意力残差模块与 3 组 Swin Transformer 模块构成，并且加入多尺度特征融合。首先，将数据通过归一化层进行数据归一化消除数据的差异，提高模型的训练精度。然后将图片数据利用注意力残差模块进行特征提取，为了适应图片，由 3×3 的卷积核、ReLU 激活函数构成的残差模块进行语义特征提取，然后与传入残差模块的矩阵通过 1×1 的卷积核进行维度匹配后的矩阵进行融合获得语义特征矩阵，然后将提取到的语义特征通过通道注意力机制分配给各通道特征权重，学习语义特征矩阵中重要的语义信息。经过 3 组注意力残差模块，初步提取语义特征。然后将语义特征矩阵送入到由 W-MSA 和 SW-MSA 为核心的 Swin Transformer 模块中提取语义特征，将注意力残差模块初步提取的语义特征通过窗口划分为多个窗口进行 MSA 进行语义特征提取。将第一个残

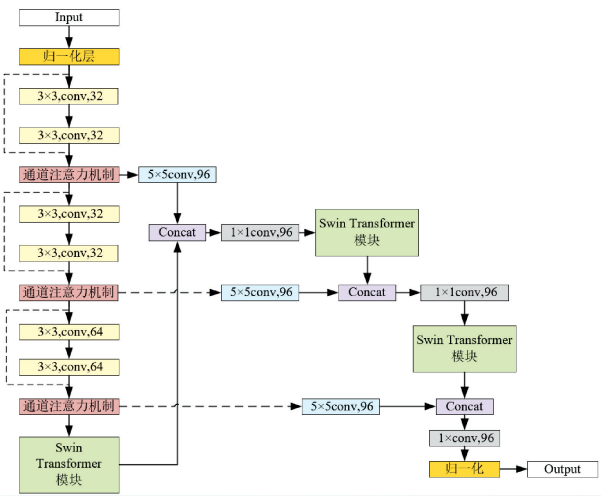

图 7　编码器结构图

Fig. 7　Encoder structure diagram

差模块的输出通过 5×5 的卷积和，然后与第 1 个 Swin Transformer 模块的输出进行拼接，由 1×1 的卷积核进行语义特征融合，送入第 2 个 Swin Transformer 模块。依次类推，获得多层融合语义特征矩阵。最后通过归一化层将语义特征矩阵归一化，最终获得由编码器映射的语义特征矩阵。连续的 W-MSA 和 SW-MSA，在增加语义特征的提取效率的同时，提升了运算效率，充分的语义特征提取使

得建模能力获得了极大的提升,最后将获得的语义特征矩阵送入到仿真信道中。

2.2 信源信道联合译码器

基于 Swin Transformer 出色的建模能力,本文设计的解码器由 Swin Transformer 模块和上采样反卷积(transposed convolution,ConvT)模块相结合,对通过仿真信道后的加噪语义特征矩阵进行解码。首先,将加噪语义特征矩阵通过 3 组 Swin Transformer 模块进行解码,然后通过上采样 ConvT 模块和 PReLU 激活函数获得解码后的图片 $X' \in R^{H \times W \times 3}$。

如图 8 所示,为本文 SemSWT 系统的解码器的结构图,解码器由归一化层、3 组 Swin Transformer 模块、ConvT 模块、PReLU 激活函数与去归一化层构成。将通过仿真信道的语义特征矩阵输入到解码器中,首先通过归一化层将数据归一化,提高模型的收敛速度。然后通过 3 层 Swin Transformer 模块进行解码,通过 3 层反卷积层 ConvT 和激活函数 PReLU 激活函数,最后通过去归一化层,使得解码后的数据恢复到原始的数据范围[0, 255],获得解码后的语义图像。

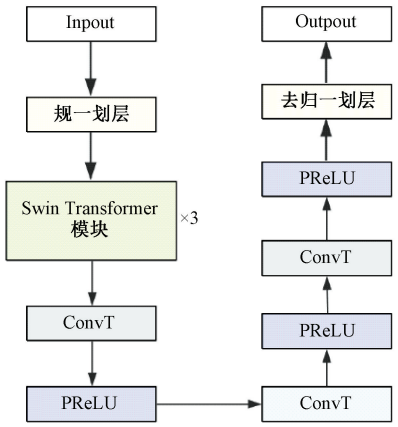

图 8 解码器结构图

Fig. 8 Decoder structure diagram

2.3 评价指标

本文采取峰值信噪比(peak signal-to-noise ratio,PSNR)和多尺度结构相似度(multi-scale structural similarity,MS-SSIM)作为本文的评价指标来验证所提出方案的性能。

PSNR 衡量原始图像与解码后的图像两者的相似程度的客观指标。其计算公式如下:

$$\mathrm{PSNR} = 10\lg\left(\frac{MAX^2}{MSE}\right) \tag{16}$$

其中,MAX 为图像像素的最大值,8 位采样点的图像最大值为 225。MSE 为原始图像与解码后的图像之间的方差。

MS-SSIM 通过多尺度测量 SSIM 来获得总体评价,计算公式如下:

$$\mathrm{MS\text{-}SSIM} = [l(x,y)]^{\alpha_M} \cdot \prod_{j=1}^{M} [c(x,y)]_j^{\beta} \cdot [s(x,y)]^{\gamma_j} \tag{17}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{18}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{19}$$

$$l(x,y) = \frac{2\mu_x + \mu y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{20}$$

其中,M 为多尺度的总数,x 表示所传输的图片,y 表示解码重构后的图片,$\sigma_x\sigma_y$ 分别表示传输图片和解码器重构后图片的方差,σ_{xy} 表示两者的协方差,μ 表示均值,C_1、C_2、C_3 为防止分母为 0 的约束常数。

3 实验与结果分析

3.1 实验环境与数据集

实验环境:本文实验在 Windows11 64 位操作系统、Intel i7-13650HX 处理器、NVIDIA GeForce RTX4060 Laptop GPU 独立显卡的计算机上进行。编程框架基于 PyTorch 框架,在 Python3.10 实现。

数据集:本文使用由 University of Toronto Computer Science 发布的 CIFAR10 数据集上进行,其包含 60 000 张 32×32 的彩色图像,包含飞机、轮船、汽车等十个类别,每个类别由 5 000 张训练图像和 1 000 张测试图像构成,为了防止模型过拟合,提高样本的多样性,通过旋转图片、增加图片的对比度等手段,对已有图像进行数据增广,扩充为 100 000 张彩色图像。

训练方法:本文通过端到端的方法对模型进行训练,在训练过程中,批大小设置为 128,使用学习率为 1×10^{-4} 的 Adam 优化器进行训练,对模型训练 600 个 Epochs。在训练过程中,在 AWGN、Rayleigh、Rician 信道训练模型,同时信道模型的信噪比均设置为 25 dB。

3.2 结果分析

为了验证本文提出的 SemSWT 重构的语义图像的效果,本文将 SemSWT 与 DeepJSCC、SemViT 两种方法进行对比,得到重构的语义图像,如图 9 所示,展示了在信道信噪比为 15 dB 时,SemSWT 与已有方案对传输图片重构的语义通信,可以看出,SemSWT 能获得更好的语义特征,更好的还原图像的语义信息,SemSWT 重构的语义图像能够很好的还原出语义图像,可以更好的传输图像的语义信息。而 DeepJSCC 与 SemViT 在细节方面还原的效果欠缺,在重构语义图像时还会产生色彩不同的斑点,影响图像的重构质量。

本文与 DeepJSCC、SemViT 两种方法进行对比。本文将信噪比(signal-to-noise ratio,SNR)设置为 25 dB 训练的模型分别在 Rayleigh 信道、Rician 信道、AWGN 下进行测试,从 SNR 为 25 dB 开始,通过将 SNR 降低 5 dB,直至 SNR 为 0 dB,并且在极低信噪比 SNR 为 −5 dB 和 −3 dB

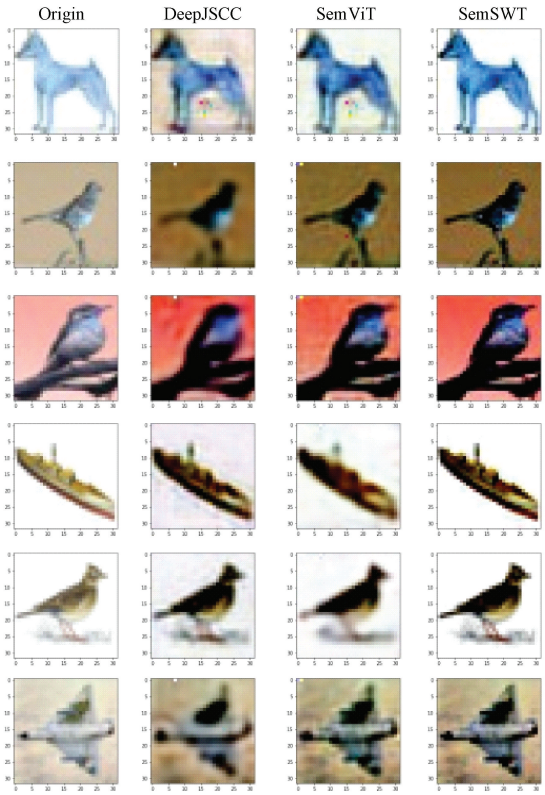

图 9　SNR=10 dB 不同方案重构的语义图像

Fig. 9　Semantic images reconstructed with different schemes for SNR=10 dB

也进行了测试。获得模型的 PSNR 性能和 MS-SSIM，其结果如下文所示。

为了验证本文提出的 SemSWT 重构的语义图像的效果，如表 1 所示，展示了在 SNR 为 25 dB 时客观指标 PSNR 与 MS-SSIM 结果，在 AWGN、Rayleigh、Rician 信道 PSNR 指标分别提升了 2.48、2.91、2.89 dB，MS-SSIM 指标分别提升了 7.44%、7.58%、7.49%。可以看出，SemSWT 能更好的获得语义特征，更好的还原图像的语义信息，有更好的应用价值。

表 1　不同模型的实验结果

Table 1　Experimental results of different models

模型	信道	PSNR/dB	MS-SSIM/%
DeepJSCC	AWGN	28.13	77.53
	Rayleigh	26.19	76.59
	Rician	26.28	76.41
SemViT	AWGN	29.34	82.47
	Rayleigh	28.33	81.95
	Rician	28.68	82.11
SemSWT	AWGN	**31.82**	**89.91**
	Rayleigh	**31.24**	**89.53**
	Rician	**31.57**	**89.60**

为了验证本文提出的 SemSWT 可以对图像更好的还原语义，本文将 SemSWT 与 DeepJSCC、SemViT 两种方法进行对比，得到语义图像重构的 PSNR 指标。在 AWGN、Rayleigh、Rician 信道的实验结果分别如图 10(a)、(b)、(c)所示，本文提出的 SemSWT 的 PSNR 指标都优于 DeepJSCC、SemViT 两种方法，且在信噪比小于 15 dB 时，

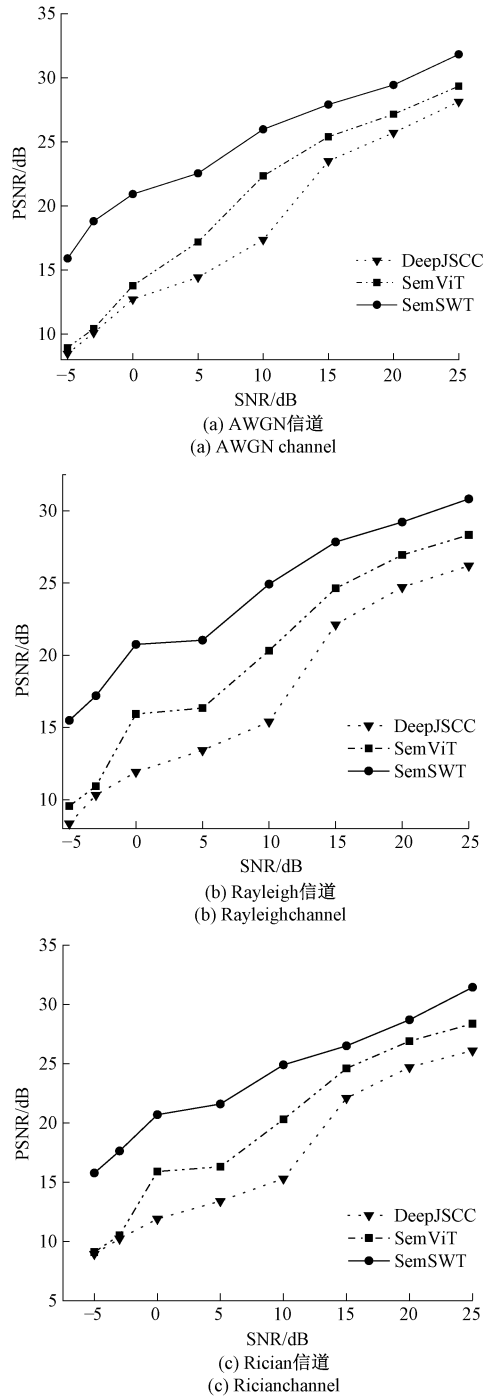

图 10　不同信道下 PSNR 性能与 SNR 关系图

Fig. 10　Plot of PSNR performance versus SNR for different channels

本文方法表现出更优的效果,SemSWT 在信噪比在−5 dB 时,PSNR 指标也有 15 dB 的性能,其值高于 DeepJSCC、SemViT 方法 5 dB。反映出 SemSWT 网络能够应用于图像语义通信,有一定的使用价值。

单一的 PSNR 指标并不能很好的说明重构后的语义图像的效果,为了进一步验证本文提出的 SemSWT 在语义图像上的性能,如图 11(a)、(b)、(c)所示,展示了在 AWGN、Rayleigh、Rician 信道下 MS-SSIM 与 SNR 的关系。本文提出的 SemSWT 的 MS-SSIM 指标都优于 DeepJSCC、SemViT 两种方法。本文提出的 SemSWT 随着信噪比的降低,性能会达到临界值,不会出现性能上的断崖式降低,并且在−5 dB 时也有 52.05%的相似度。反映出 SemSWT 网络应用于图像语义通信时有更强的鲁棒性,可以在信噪比变化较大时,也有较好的性能。

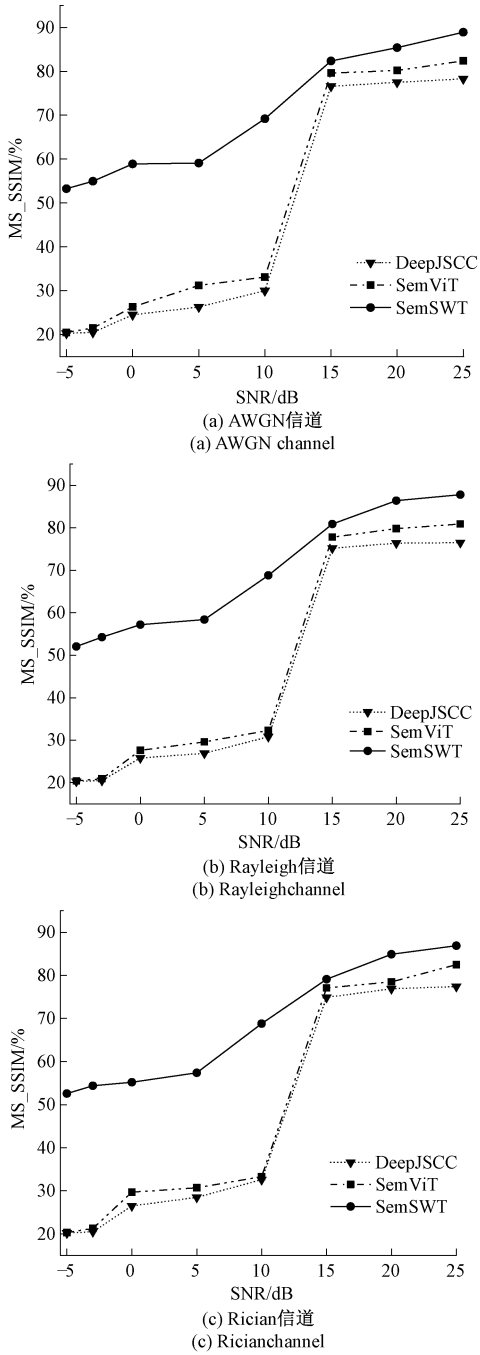

图 11 不同信道下 MS-SSIM 性能与 SNR 关系图

Fig. 11 Plot of MS-SSIM performance versus SNR for different channels

3.3 消融实验

为了进一步验证本文提出模型的可行性与有效性,本文在 SNR=25 dB 的 AWGN 信道下分别对残差连接与通道注意力以及多尺度特征融合进行消融实验,实验结果如表 2 所示,在单独加入残差连接时,PSNR 提升了 0.8 dB,MS-SSIM 提升了 1.2%,在单独使用通道注意力机制时,PSNR 提升了 0.6 dB,MS-SSIM 提升了 1.1%,当同时使用残差连接与通道注意力时,PSNR 提升了 1.5 dB,MS-SSIM 提升了 2.6%,在此基础上添加多尺度融合,PSNR 提升了 0.31 dB,MS-SSIM 提升了 0.68%。由此可知,本文所使用的残差注意力模块对于模型的性能具有较好的提升效果。

表 2 消融实验结果

Table 2 Results of ablation experiments

方法			PSNR/dB	MS-SSIM/%
通道注意力	残差模块	多尺度融合		
—	—	—	29.93	86.32
√	—	—	30.51	87.45
—	√	—	30.72	87.54
√	√	—	31.51	88.93
√	√	√	31.82	89.71

4 结 论

基于 Swin Transformer 的原理,本文提出一种基于 Swin Transformer 的图像语义通信系统 SemSWT。系统包括编码器、解码器以及模拟信道。在发射端,将基于注意力机制的残差模块与 Swin Transformer 模块相结合,提高了语义通信的效率的。通过对 Rayleigh 信道、Rician 信道、AWGN 信道仿真进行加噪来模拟真实信道环境。在接收端,通过 Swin Transformer 模块和上采样进行解码,然后通过反卷积还原图片。实验结果表面,在 AWGN、Rayleigh 以及 Rician 信道中,SemSWT 在 PSNR 取得了至少 2 dB 的性能提升,特别是在 0 dB 处,也保持 PSNR 在 20 dB。在 MS-SSIM 性能上,SemSWT 取得了 55%的性能,在 SNR 的降低到−5 dB 时也有 52.05%的性能。在未来的工作中,会进一步的优化模型,以期在便携式设备中部署。

参考文献

[1] 刘传宏,郭彩丽,杨洋,等. 面向智能任务的语义通信:

理论、技术和挑战[J]. 通信学报，2022，43(6)：41-57.
LIU CH H，GUO C L，YANG Y，et al. Intelligent task-oriented semantic communications：Theory，technology and challenges [J]. Journal on Communications，2022，43(6)：41-57.
[2] 李天放，孙一宸，于明鑫，等. 结合语义分割与跨模态差分特征补偿的红外与可见光图像融合方法[J]. 电子测量与仪器学报，2024，38(7)：34-45.
LI T F，SUN Y CH，YU M X，et al. Infrared and visible image fusion method integrating semantic segmentation and cross-modality differential feature compensation[J]. Journal of Electronic Measurement and Instrumentation，2024，38(7)：34-45.
[3] 李云飞，张巧芬，王桂棠，等. 基于改进 PointNet++的室内点云语义分割模型[J]. 国外电子测量技术，2023，42(12)：63-69.
ZHAO Y F，ZHANG Q F，WANG G T，et al. Improved PointNet++ model for in door point cloud semantic segmentation [J]. Foreign Electronic Measurement Technology，2023，42(12)：63-69.
[4] 秦志金，赵茨茨，李凡，等. 多模态语义通信研究综述[J]. 通信学报，2023，44(5)：28-41.
QIN ZH J，ZHAO T T，LI F，et al. Survey of research on multimodal semantic communication[J]. Journal on Communications，2023，44(5)：28-41.
[5] 张平，牛凯，姚圣时，等. 面向未来的语义通信：基本原理与实现方法[J]. 通信学报，2023，44(5)：1-14.
ZHANG P，NIU K，YAO SH SH，et al. Semantic communications for future：Basic principle and implementation methodology [J]. Journal on Communications，2023，44(5)：1-14.
[6] SEBASTIAN D，SEBASTIAN S，JOKAB H，et al. Deep learning based communication over the air[J]. IEEE Journal of Selected Topics in Signal Processing，2017，12(1)：132-143.
[7] BOURTSOULATZE E，KURKA D B，GUNDUZ D. Deep joint source-channel coding for wireless image transmission[J]. IEEE Transactions on Cognitive Communications and Networking，2019，5(3)：567-579.
[8] LEE C H，LIN J W，CHEN P H，et al. Deep learning constructed joint transmission recognition for internet of things [J]. IEEE Access，2019，7：76547-76561.
[9] XU J，AI B，CHEN W，et al. Wireless image transmission using deep source channel coding with attention modules[J]. IEEE Transactions on Circuits and Systems for Video Technology，2022，32(4)：2315-2328.
[10] 刘龙，方檡炫，张梦璇，等. 基于 Transformer 特征关联融合小目标检测算法研究[J]. 信号处理，2024，40(8)：1-26.
LIU L，FANG J X，ZHANG M X，et al. Research on feature association and fusion small target detection algorithm based on transformer[J]. Journal of Signal Processing，2024，40(8)：1-26.
[11] 刘华咏，黄聪，金汉均. 注意力增强的视觉 Transformer 图像检索算法[J]. 电子测量技术，2023，46(23)：50-55.
LIU H Y，HUANG C，JIN H J. Image retrieval method with attention-enhanced visual Transformer[J]. Electronic Measurement Technology，2023，46(23)：50-55.
[12] HE K M，CHEN X L，XIE S N，et al. Masked autoenco-ders are scalable vision learners [C]. Conference on Computer Vision and Pattern Recognition，Piscataway：IEEE Press，2022：15979-15988.
[13] DOSOVITSKIY A，BEYERR L，KOLESNIKOV A，et al. An image is worth 16×16 words：transformers for image recognition at scale [J]. International Conference on Learning Representations. ArXiv preprint arXiv：2010.1192，2020.
[14] YOO H J，DAI L L，KIM S K，et al. On the role of ViT and CNN in semantic communications：Analysis and prototype validation[J]. IEEE Access，2023，11：71528-71541.
[15] 龙伟军，郭宇轩，徐艺卓，等. 基于二分匹配 Transformer 的 SAR 图像检测[J]. 信号处理，2024，40(9)：1648-1658.
LONG W J，GUO Y X，XU Y ZH，et al. SAR image detection based on the bipartite matching transformer[J]. Journal of Signal Processing，2024，40(9)：1648-1658.
[16] 刘铁，段勇. 融合 CNN 和 Transformer 的机器人室内场景识别[J]. 电子测量与仪器学报，2023，37(5)：223-229.
LIU T，DUAN Y. Robot indoor scene recognition based on fusion of CNN and Transformer[J]. Journal of Electronic Measurement and Instrumentation，2023，37(5)：223-229.
[17] LIU Z，LIN Y，CAO Y，et al. Swin transformer：Hierarchical vision transformer using shifted windows[C]. 2021 IEEE/CVF International Conference on Computer Vision，2021：9992-10002.
[18] HU J，SHEN L，SUN G，et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2017，42(8)：2011-2023.
[19] HE K M，ZHANG X，REN S，et al. Deep residual learning for image recognition [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition，2016：770-778.
[20] 魏秀业，程海吉，贺妍，等. 基于特征融合与 ResNet 的行星齿轮箱故障诊断[J]. 电子测量与仪器学报，2022，36(5)：213-222.
WEI X Y，CHENG H J，HE Y，et al. Fault diagnosis of planetary gearboxes based on feature fusion and ResNet[J]. Journal of Electronic Measurement and Instrumentation，2022，36(5)：213-222.
[21] 彭宝玉，曹立佳. 多尺度融合注意力机制改进 U-Net 实现肺部感染区域分割[J]. 国外电子测量技术，2023，42(10)：177-183.
PENG B Y，CAO L J. Multi-scale fusion attention mechanism improves U-Net to achieve lung infection area segmentation [J]. Foreign Electronic Measurement Technology，2023，42(10)：177-183.
[22] 袁嘉辉，刘蕊，梁虹，等. 基于 SE-ResNet34 的红火蚁巢穴判别模型[J]. 电子测量技术，2023，46(23)：97-104.
YUAN J H，LIU R，LIANG H，et al. Red fire ant nest classification model based on SE-ResNet34[J]. Electronic Measurement Technology，2023，46(23)：97-104.

作者简介

孙洋舟，硕士研究生，主要研究方向为语义通信等。
E-mail：2480213924@qq.com

严天峰（通信作者），教授，主要研究方向为无线电通信与监测。
E-mail：yantianfeng@163.com