



# 基于深度学习的图像语义通信系统

## Deep Learning-Based Image Semantic Communication System

张振国/ZHANG Zhenguo, 杨倩倩/YANG Qianqian,  
贺诗波/HE Shibo

(浙江大学, 中国 杭州 310058)  
(Zhejiang University, Hangzhou 310058, China)

DOI: 10.12142/ZTETJ.202302011

网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20230411.1400.010.html>

网络出版日期: 2023-04-11

收稿日期: 2023-02-06

**摘要:** 语义通信是一种新颖的通信方式, 可通过传输数据的语义信息提高带宽效率。提出一种用于无线图像传输的系统。该系统基于深度学习技术开发并以端到端 (E2E) 的方式进行训练。利用深度学习实现语义特征的提取和重建, 在发送端提取信源信息不同类型和不同形式的语义特征, 在接收端融合各类型语义特征进行目标语义恢复。仿真结果表明, 与基准模型相比, 所提模型在信道环境恶劣的情况下, 具有更好的重建精度。

**关键词:** 端到端通信; 多级语义通信; 图像压缩; 图像传输

**Abstract:** Semantic communications is a novel form of communication that improves bandwidth efficiency by transmitting semantic information about data. A system for wireless image transmission is introduced, which is developed based on deep learning techniques and trained by an end-to-end (E2E) approach. Deep learning is used to extract and reconstruct semantic features, extract different types and forms of semantic features of source information at the sending end, and fuse various types of semantic features at the receiving end for target semantic recovery. The simulation results show that compared with the benchmark model, the proposed model has better reconstruction accuracy under the bad channel environment.

**Keywords:** end-to-end communication; multi-level semantic communication; image transmission

当前, 中国正大力发展信息产业。无线通信技术的快速发展为智慧城市、高清视频、自动驾驶、远程医疗等带来产业变革。基于物联网的智能感知网络迅速发展, 在提供便利的同时也带来了庞大的无线通信数据<sup>[1]</sup>。数据通信方式不再受限于人-人通信, 而是转变为以目标为导向的通信方式<sup>[2]</sup>。随着人工智能的快速发展, 基于深度学习的通信为克服传统通信困难提供新思路。基于深度学习的架构在通信系统中取得了显著成果, 达到甚至超过传统方案的性能。基于香农定理的1G到5G技术已经无法满足当前的通信需求, 基于深度学习的语义通信为下一代无线通信技术带来更多可能<sup>[3-5]</sup>。深度学习已被用于优化基于分离模块设计的传统通信, 如信源编解码器、信道编解码器和调制解调模块<sup>[6-8]</sup>。文献[9]建立端到端 (E2E) 的信源信道联合编码 (JSCC) 通信系统, 该系统能够有效应对传统通信系统中的瓶颈问题。

E2E 通信系统的发射器和接收器由深度神经网络 (DNN) 组成。编码器学习输入数据的特征向量, 并通过无线信道将复数符号发送到解码器进行目标重建。

与传统通信相比, 语义通信侧重于传递信息的含义, 而不是符号的精确传输。区别于传统通信系统利用误码率 (BER) 或符号错误率 (SER) 评估通信结果, 语义通信系统通过最小化输入和重建信息之间的语义损失, 来恢复接收器处的信源信息<sup>[9]</sup>。现有的语义通信系统联合设计发射器和接收器, 以实现更好的传输效率和鲁棒性<sup>[10]</sup>。深度学习在语义通信模型中得到广泛的应用, 具有良好的特征提取和学习能力, 能够对信源包含的语义信息进行提取和传输<sup>[11]</sup>。相较于数字通信通信模型, 基于深度学习的通信模型不会出现“悬崖效应”。当前, 语义通信主要基于深度学习进行开发和探索, 并且已经取得一定的效果<sup>[12]</sup>。基于深度学习的语义通信系统显示出巨大的潜力, 能够有效传输不同类型的信息。随着物联网设备的大量部署, 以目标为导向的通信方式<sup>[2]</sup> (如人-机、机-机) 不断涌现, 这给边缘设备带来巨大的通

**基金项目:** 国家自然科学基金 (62201505); 浙江大学-新加坡科技设计大学创新、设计与创业联盟项目 (202102); 中央大学基础研究基金 (2021FZZX001-20)

信压力。语义通信只传输目标需要的信息，大大减少数据通信量，提高通信效率。语义通信将成为下一代物联网无线通信技术的重要组成部分。

信道噪声干扰是影响无线通信系统性能的主要因素之一，因此提高通信系统应对噪声环境的鲁棒性是传统通信和语义通信的共同目标。数字通信方案通过增加信道编码量来提高系统的抗噪能力，使通信量急剧增加。当前，基于深度学习的通信系统通过DNN缓解噪声对系统的干扰，同时平衡系统的通信量。文献[13]提出了一种用于通用无线信道的基于深度学习的端到端通信系统，其中信源编码、信道编码、调制解调等传统模块被DNN所取代。该端到端系统以数据驱动的方式成功地利用各种相关性，获得较好的结果。文献[14]提出了一种基于强化学习的方法，在不知道信道传递函数或信道状态信息（CSI）的情况下优化发送端DNN。在实际系统中，由于存在信令和检测方案的损伤、硬件缺陷、变化的信道条件等情况，经过训练的网络所得到的信道与用于训练网络的信道显著不同。文献[15]使用随机扰动方法设计了一个无信道模型的端到端通信框架。该模型在真实信道中训练基于深度学习的通信系统，不需要对信道模型进行任何假设。文献[16]开发了一种基于条件对抗生成网络（StyleGAN）的方法，用于构建端到端通信系统，其中StyleGAN用于构建信道效应模型。该模型将与导频数据相对应的接收信号作为调节信息的一部分，在接收机处获得信道状态信息，并将其用于信号检测。

近年来，研究人员提出了多种基于深度学习的语义通信系统，旨在提高在噪声信道上自然语言传输的性能。文献[17]提出了一种用于自然语言的JSCC系统，通过最小化E2E失真，实现比传统方法更好的通信性能。文献[18]提出了两种方案来处理有限的数据速率问题：自适应传输方案和广义数据表示方案。其中，自适应传输方案可以在不同信道条件下以均方误差（MSE）约束最大化数据速率；广义数据表示方案取代了one-hot表示，以获得更高的数据速率。文献[19]则提出了一种新型语义通信系统，通过信道估计辅助训练实现文本重建。与文本语义通信系统相比，语音信号难以提取和表示基本的语义特征。研究人员使用多种方法来开发用于语音信号传输的语义通信系统，文献[20]提出了一种基于深度学习的语音通信系统。该系统通过联合训练语义和信道编解码器，学习和提取语音特征，同时利用注意力机制减轻实际通信场景中的信道失真和衰减，使系统获得更好的通信性能。文献[21]采用波向量（wav2vec）来提取音频语义特征，并通过强化学习（RL）提高特征提取的准确性。

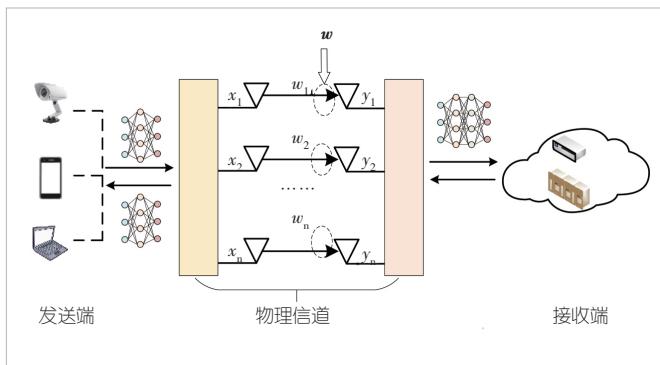
针对图像信息的压缩和传输问题，基于深度学习的图像

压缩技术已经成为研究热点。一系列的图像无线信号通信系统相继产生。文献[22]提出了一种基于深度学习的无线图像传输系统，实现基于E2E的JSCC，使用峰值信噪比（PSNR）和结构相似性指数（SSIM）测量重建图像的质量。在JSCC的基础上，另一种具有信道反馈的图像重建方案DeepJSCC-f，通过接收来自接收器的信道反馈，进一步提高图像重建精度<sup>[23]</sup>。文献[24]提出了基于正交频分复用（OFDM）数据路径的JSCC方案，用于多路径衰落信道的无线图像传输。该通信方案通过整合专家知识实现了更好的性能。文献[25]开展了物联网（IoT）设备用于图像传输的研究，通过将两个DNN的联合传输识别方案部署在设备端，在识别精度上比传统方案更好。文献[26]提出了联合特征压缩和传输系统，以处理边缘服务器上有限的计算资源。该方案不仅提高了E2E的可靠性，而且降低了计算复杂度。此外，M. JANKOWSKI等提出了基于自编码器严格约束的设备边缘通信系统，在有限的计算能力下实现了更好的分类精度<sup>[27]</sup>。文献[28]提出了一种新的航空图像传输范式，在无人机端部署一个轻量级模型，用于感知图像和信道条件的语义传输模块，在接收端通过计算能力更强的基站对收到的信息进一步处理，提高分类精度。文献[29]提出了基于深度学习的用于传输单模态和多模态数据的多用户语义通信系统，通过在编码器和解码器层之间添加连接来融合多模态数据，实现图像检索和视觉问答等功能。然而，当前仍缺乏对于图像语义的评估准则，导致语义通信系统的图像恢复仍需要使用传统的图像重建准则来衡量系统性能。

语义通信作为下一代通信技术的重要组成部分，为新一代的通信变革提供新方法。目前，语义通信发展仍处于探索阶段。作为下一代通信技术的重要组成部分，语义需要理论和技术的双重发展。目前，语义通信在数学理论证明方面仍未得到普遍认可，在技术方面深度学习的机理和可解释性仍然有待研究。本文是在技术层面对图像语义通信的一次探索，通过简单的通信环境假设，验证通信系统的可行性；通过基于深度学习的图像信息挖掘，探索信息内容的语义性对信息传递的影响。实验结果表明，本文所提出的语义通信系统具有较好的鲁棒性，特别是在有限带宽条件下更具优势。这表明高级语义信息在图像传输中具有优势。

## 1 语义系统模型

图像语义通信模型的一般框架如图1所示。与传统通信系统相同，语义通信系统包含发送端、无线信道和接收端3个部分。两者的主要区别在于语义通信系统所采用的编解码方式。语义通信系统将信源信道的编解码作为一个整体。信



▲图1 图像语义通信系统

道编解码在应对噪声的同时对信源编码信息的重要程度进行衡量，为其中重要的信息分配更多的符号位以进行编码。针对信源中的语义内容，语义编码对有含义的信息进行编码表达，其过程实际就是对信源中语义概念的高度抽象与压缩。为提高图像无线通信的高效性和准确性，本文提出一种面向语义的图像通信系统，该系统通过联合信源语义编解码和信道编解码的方式，提高系统的通信效率和鲁棒性。

### 1.1 编码器

编码器由两部分构成：语义特征提取器和联合语义信道编码器。具体过程为：首先，编码器的输入图像 $S$ 由归一化层预处理，使得每个元素都在 $[0,1]$ 范围内；其次，通过多个基于神经网络的多级语义特征提取器提取输入图像的不同语义特征；最后，利用联合语义信道编码器将语义特征编码为符号，通过物理通道传输到接收器，传输向量 $x$ ：

$$x = T_\beta(T_\alpha(S)), \quad (1)$$

其中， $T_\alpha(\cdot)$ 为多级语义特征提取网络，网络参数为 $\alpha$ ； $T_\beta(\cdot)$ 为联合语义信道编码器，网络参数为 $\beta$ 。

本文所提方案在一个广泛使用的物理信道——加性白高

斯噪声（AWGN）信道上进行模型测试。则解码器上接收到的信号 $y$ 为：

$$y = x + w, \quad (2)$$

其中， $w$ 为圆对称高斯分布的独立同分布向量， $w \sim \mathcal{CN}(0, \sigma^2 I)$ ， $\sigma^2$ 为信道的平均噪声功率， $I$ 为单位矩阵。

### 1.2 解码器

解码器也由两部分组成：联合语义信道解码器和图像重建模块。联合语义信道解码器减轻信号在AWGN信道的噪声干扰，并恢复多级语义特征。图像重建模块融合不同层次的语义信息并重建目标图像。反归一化层将每个元素重新缩放为图像像素值（0~255）。联合语义信道解码器和图像重建模块的参数为 $\xi$ 和 $\eta$ ，接收器重建图像为：

$$\hat{S} = R_\eta(R_\xi(y)), \quad (3)$$

其中， $R_\xi(\cdot)$ 和 $R_\eta(\cdot)$ 分别为联合语义信道编码器和图像重建模块， $y$ 为从信道接收到的信号。

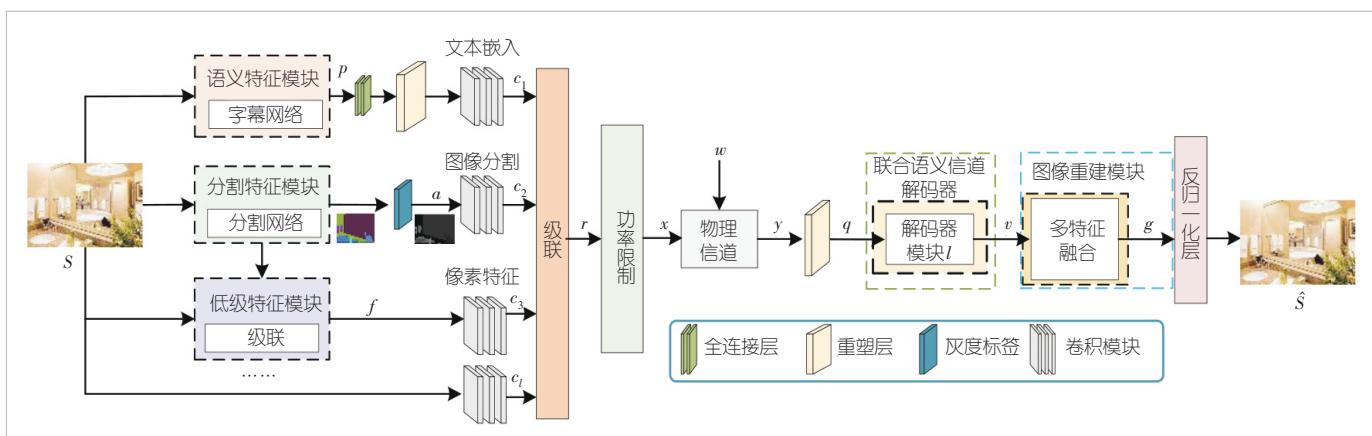
图像语义通信系统的目标是最小化输入图像 $S$ 和重建图像 $\hat{S}$ 之间的平均误差，均方差（MSE）用来评估 $S$ 和 $\hat{S}$ 之间的失真：

$$\mathcal{L} = \frac{1}{N} \sum_{k=1}^N d(S_k, \hat{S}_k), \quad (4)$$

其中， $d(S, \hat{S}) = \frac{1}{n} \|S - \hat{S}\|^2$ 为均方误差分布， $N$ 为样本数。

### 2 多级图像语义通信模型

本文提出一种基于深度学习的无线图像传输语义通信系统，如图2所示。其中，多级语义特征提取器用于提取不同级别的语义特征。高级语义信息包含图像的抽象性和通用性



▲图2 图像语义通信系统的整体架构

指标，低级语义信息包含图像的局部细节语义信息<sup>[30]</sup>。该系统通过基于深度学习的特征提取器提取信源特征，并通过与语义信道的联合训练给不同的信息赋予不同的权重。语义信道编码器和解码器联合在接收器处成功恢复这些语义特征，并通过图像重建模块对多级语义信息进行融合并重构目标图像。

## 2.1 多级语义特征提取器

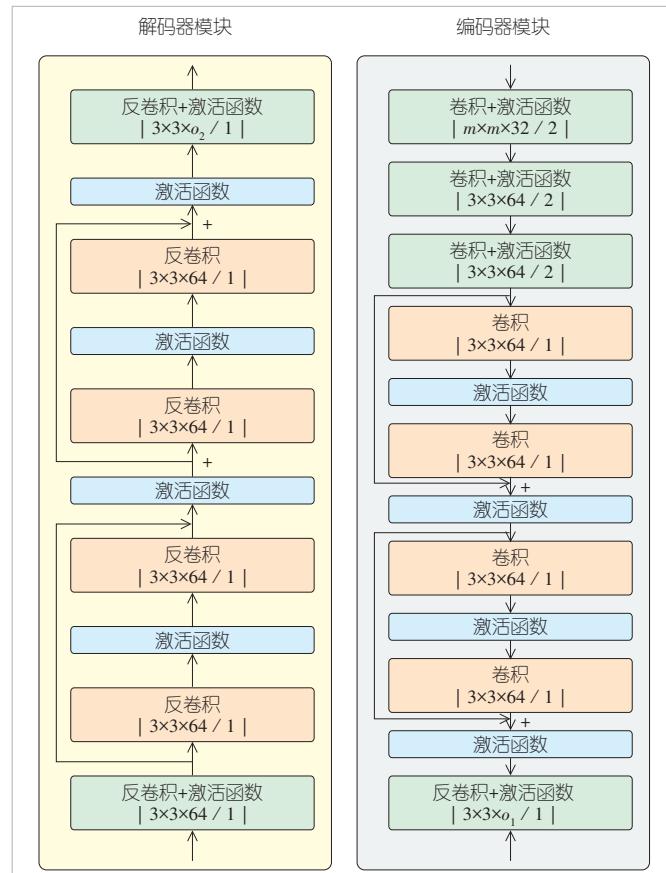
对于输入图像  $S \in \mathbb{R}^{b \times h \times w \times 3}$ ，首先通过归一化层将像素值映射到[0,1]范围，其中  $b$ 、 $h$  和  $w$  分别为图像的批数量、高和宽，3 表示图像对应的图像通道数。一个批次的图像数据会被送入多级语义特征提取器。该提取器包含3个模块：语义特征模块、分割特征模块和低级特征模块。不同模块分别提取不同层级和不同形式的图像语义信息。语义特征模块包含一个预训练的图像字幕模型，该模型由 ResNet-152 模型<sup>[31]</sup>和长短期记忆网络（LSTM）<sup>[32]</sup>层组成，用于提取图像文本形式的高级语义信息  $p \in \mathbb{R}^{\frac{b \times 2 \cdot h \cdot w}{t^2} \times 1}$ ，其中  $t$  为下采样因子。文本形式语义特征是带有图像纹理信息的文本嵌入信息。为了方便理解，本文称这些特征为“文本形式”。分割特征模块通过预训练的图像分割<sup>[33]</sup>模型获取图像分割形式的高级语义信息  $a \in \mathbb{R}^{b \times h \times w \times 1}$ 。低级特征模块由两部分组成：联合部分和直取部分。其中，联合部分通过 concatenate 按通道维度将归一化图像和分割特征相结合，并输出  $f \in \mathbb{R}^{b \times h \times w \times 1}$ ；直取部分为直接输入归一化的图像特征。低级特征模块主要用于图像细节特征的提取，为高级语义信息做细节补充，实现更丰富的信息重构。超参数  $t$  控制低级特征传输信息量，该参数决定通信量和目标信息的重建质量。

## 2.2 联合语义信道编解码

通过多级语义特征提取器可获得4种类型的目标语义信息特征：文本形式特征  $p$ 、图像分割特征  $a$ 、联合特征  $f$  和图像归一化低级特征。语义特征  $p$  首先通过全连接层进行维度压缩，进而通过整形层和编码模块输出  $c_1 \in \mathbb{R}^{b \times \frac{h}{t} \times \frac{w}{t} \times 1}$ ，其中  $\times 1$  表示维度信息。语义特征  $a$  和  $f$  分别输入两个独立的编码器模块，并输出不同的特征维度信息  $c_2 \in \mathbb{R}^{b \times \frac{h}{t} \times \frac{w}{t} \times 1}$  和  $c_3 \in \mathbb{R}^{b \times \frac{h}{t} \times \frac{w}{t} \times 3}$ 。 $c_3$  和  $c_l$  除了最后一层的输出通道数不同外，其他结构均相同。编码器模型  $c_l$  获取图像细节信息特征  $c_l \in \mathbb{R}^{b \times \frac{h}{t} \times \frac{w}{t} \times l}$ 。联合层将获取的不同内容的特征  $c_1, \dots, c_l$  进

行联合和压缩，并输出  $r \in \mathbb{R}^{b \times \frac{h}{t} \times \frac{w}{t} \times e}, e = 1, 2, \dots, 3 \times l - 4$ ；进而通过功率归一化层生成  $k$  个数传输符号  $x$ ，同时这些符号的发射功率低于给定值，其中  $x \in \mathbb{E}^{b \times k}$ ， $k = \frac{h}{t} \times \frac{w}{t} \times e$ 。信道带宽压缩比为  $k/n = \frac{3 \times t^2}{g}$ ， $g = 1, \dots, 3 \times l - 4$ ，该值可根据信道环境自适应调整。高级语义信息是目标重建的增强信息。当  $e = 1$  时，联合语义信道编码输出为  $c_3 \in \mathbb{R}^{b \times \frac{h}{t} \times \frac{w}{t} \times 1}$ 。

接收端的重塑层将接收到的信号重组为  $q_i, i = 1, \dots, l$  个语义特征，其尺寸为  $b \times \frac{h}{t} \times \frac{w}{t} \times (3 \times l - 4)$ 。当连接接收到的符号的实部和虚部时，最后一个维度加倍，如  $q_1 \in \mathbb{R}^{b \times \frac{h}{t} \times \frac{w}{t} \times 2}$ 。同样， $q_i$  使用第  $3i - 1$  和  $3i + 1$  元素的最后一个维度， $q_i \in \mathbb{R}^{b \times \frac{h}{t} \times \frac{w}{t} \times 6}$ 。每个  $q_i, i = 1, \dots, l$  都输入到解码器模块，其具有相同的卷积核尺寸 ( $m=3$ )，如图3所示。每个解码器的输出语义特征为  $d_i \in \mathbb{R}^{b \times \frac{2h}{t} \times \frac{2w}{t} \times o}, i = 1, \dots, l$ ，其中  $o$  为输出通道数。卷积神经网络的架构由  $m \times m \times o/st$  参数构



▲图3 编码器模块、解码器模块结构示意图

成，其中  $m$ 、 $o$  和  $st$  分别是卷积核尺寸、通道输出数量和步长。每个卷积层之后为广义的归一化变换层（包括广义除数归一化（GDN）和逆 GDN（IGDN））。

### 2.3 图像重建模块

图像重建模块需要融合不同形式和内容的语义特征，完成不同语义内容之间的相互补充，通过注意力机制深度挖掘融合信息，进而将融合特征重建为目标图像。首先借助双特征融合模块对两种形式的高级语义特征进行融合，即双特征融合模块通过交叉结构和通道注意力机制（CA）学习输入特征；然后通过像素上采样模块对特征信息升维，其中像素上采样模块由卷积层和像素上采样层构成；最后将相同维度的高级语义信息和低级细节补充信息进行级联操作，通过残差网络对融合后的信息进行提取并重建目标图像。残差网络由反卷积层和 PReLU 激活函数构成（最后一层为 sigmoid 激活函数），其网络结构如图 4 所示。在图像重建模块中，不同形式特征生成的粗糙图像含有不同的成分，例如比较平滑的低频信息和充满边缘、纹理的高频信息。同时，卷积层的每个过滤器都包含一个局部感受野，其输出无法利用局部信息之外的上下文信息。因此，通过通道注意力机制改变特征权重，能够提高重要信息的权重占比。通道注意力机制的网络结构如图 5 所示。

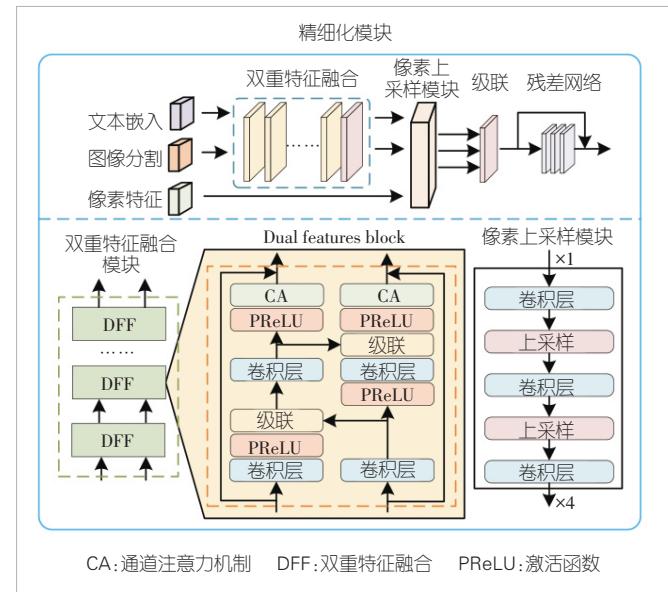
## 3 仿真结果与分析

### 3.1 仿真参数

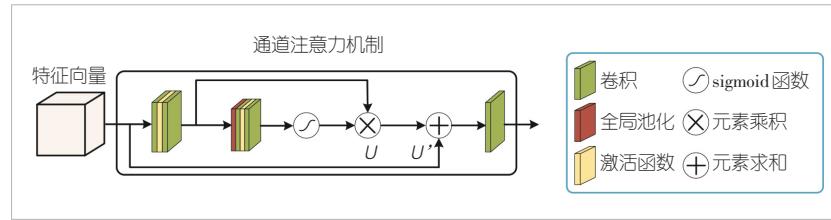
根据设计模型特点，需要使用具有多类型标签的数据集对模型进行训练。本文使用 MSCOCO<sup>[34]</sup> 和 ADE20K<sup>[35]</sup> 数据集训练和评估所提出的系统。MSCOCO 数据集包含 123 287 张图像（82 783 张用于训练，40 504 张用于测试），每张图像都包含 5 个不同的字幕标题。ADE20K 数据集包含 27 574 张图像、150 个语义标签，图像高度和宽度至少为 512 像素。训练模型时，每个图像都被裁剪为固定大小： $h = 128$ ， $w = 128$ 。对于文本形式特征，本文使用带有文本标签的 MSCOCO 数据集来训练语义特征模块 CaptionNet<sup>[24-25]</sup>，并使用带有分割标签的 ADE20K 数据集来训练系统的其余部分（冻结语义特征模块）。最后，本文在 Kodak<sup>[36]</sup> 图像数据集上测试所提出的图像无线通信系统。Kodak 数据集总共包含 24 张固定尺寸  $768 \times 512$  的图像。

所提模型在 Pytorch<sup>[37]</sup> 中实现。并使用 Adam 算法<sup>[38]</sup> 进行优化。学习率设置为 0.000 1，

批次大小设置为 32，下采样因子  $t = 8$ 。本文采用现有的基于深度学习的方法（DeepJSCC 和 Aided Deep-JSCC<sup>[39]</sup>），以及传统的基于分离的数字传输方案（JPEG）作为基准进行比较，并使用 PSNR 和 SSIM 指标来评估性能。PSNR 表示信号最大可能功率和影响其表示精度的破坏性噪声功率的比值， $PSNR = 10 \log_{10} \frac{MAX^2}{MSE}$  dB，其中  $d(x,y)$  为均值误差，MAX 表示图像点颜色的最大数值。SSIM 使用 3 个标准度量图像：亮度、对比度和结构。 $SSIM = \frac{(2\mu_s\mu_{\hat{s}} + C_1)(2\sigma_{s\hat{s}} + C_2)}{(\mu_s^2 + \mu_{\hat{s}}^2 + C_1)(\sigma_s^2 + \sigma_{\hat{s}}^2 + C_2)}$ ，其中  $\mu$  为亮度， $\sigma$  为对比度， $C_1$  和  $C_2$  为常数。Aided Deep-JSCC 是一种用于无线传感器网络的分布式语义通信方案。其中，每个传感器观察并编码一个公共图像，并通过无线信道独自将其发送给接收器。接收器融合来自不同传感器的信息并重建原始图像。该方案类似于本文所提方案，即从原始图像中提取多个版本的语义信息，并将其发送给接收者以用于信源信息的重建。根据香农分离定理，传输速率必须降低信道容量，通过物理信道传输的信源信息需要满足  $R \leq \frac{k}{n} \log_2(1 + SNR)$ ，其中  $R$  为信源信



▲图 4 图像重建模块结构示意图



▲图 5 通道注意力机制网络结构示意图

息的传输速率,  $\log_2(1 + SNR)$  为信道容量,  $k$  为信道维度,  $n$  为图像维度。信源信息在物理信道上可靠传输的最大速率  $R_{\max}$ 。传统的图像压缩方案具有最小压缩码率  $R_{\min}$ , 这是接收端重建目标图像的极限。如果  $R_{\min} > R_{\max}$ , 则接收器无法重建输入图像。 $SNR = 10 \lg \frac{P_s}{P_n}$  为信噪比, 其中  $P_s$  为信号功率,  $P_n$  为噪声功率。

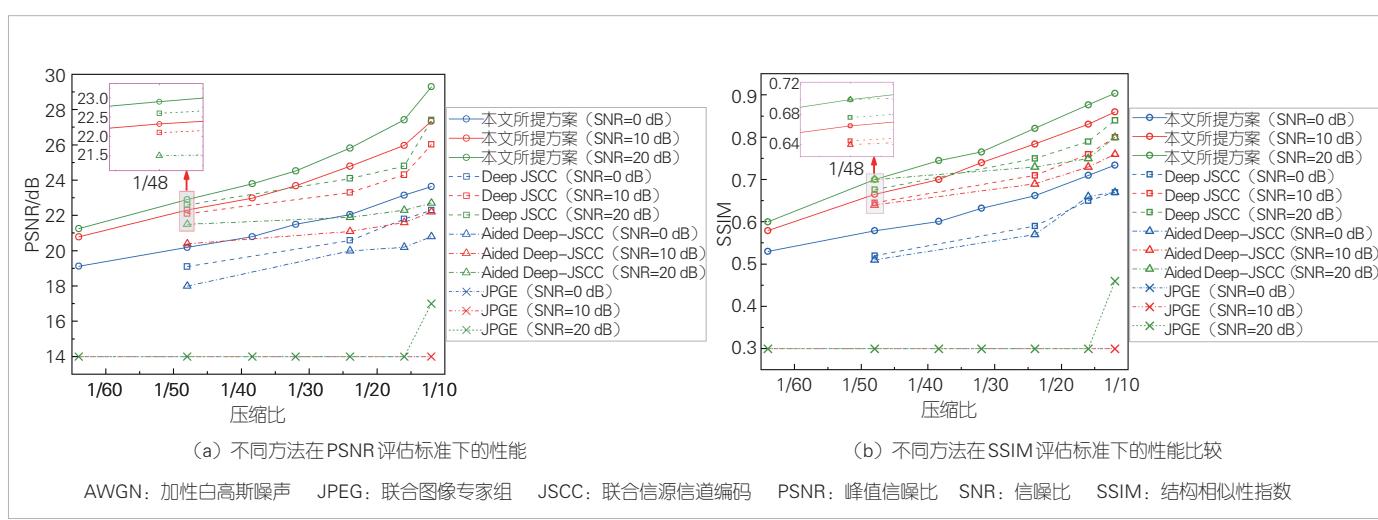
### 3.2 结果分析

图 6 对比了不同压缩比下图像重建 PSNR 的性能表现。可以看出, 本文所提算法优于其他深度学习方案和传统方案。同时, 基于深度学习的图像通信系统不会因“悬崖效应”(信道条件低于某个阈值, 接收器无法恢复传输的图像)而遭受大幅的性能下降。数字传输方案在较差的信道环境中( $SNR < 10$  dB)和低压缩比( $k/n < 1/10$ )条件下引发系统崩溃, 而基于深度学习的系统仍然可以完成目标信息的重建。本文所提方案中的联合语义信道编解码方案能够均衡分配信源信息编码和抗噪编码字符数, 使得在信号压缩和重建过程中能够更好地表达语义信息。在低压缩比信道情况较好的环境中, 本文所提方案性能与 DeepJSCC 模型相近。这是由于: 在低压缩比下, 发射端没有足够的带宽来传达更详细的图像细节信息, 导致系统重建目标在细节信息的刻画上不足。随着压缩比的增大, 本文所提算法性能与其他方案之间的差距逐渐拉大。这进一步体现出所提模型在抗噪和信息恢复上的优势。

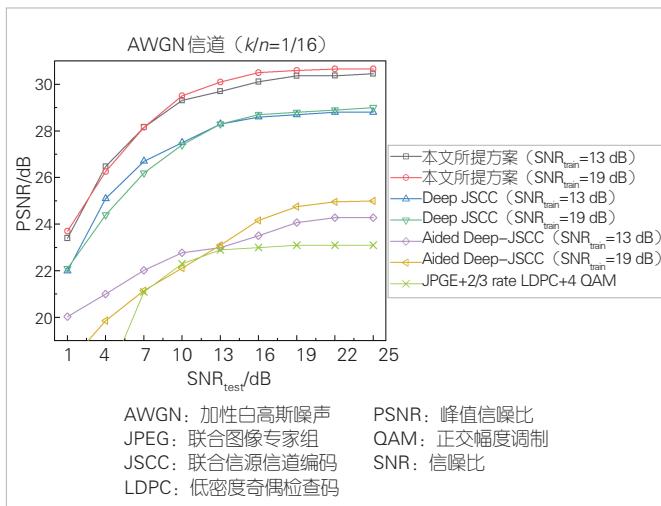
图 6 同时展示了不同方法在 SSIM 评估标准下的性能比较。SSIM 从图像结构方面反映了原始图像和重建图像之间的相似性。本文所提方案在低压缩比及较差的信道条件下

(例如,  $SNR=0$ ) 的性能显著优于其他方案。Aided Deep-JSCC 系统将 SSIM 和 MSE 作为损失函数, 在 SSIM 指标上具有更好的性能。本文所提方法的图像重建模块对高级语义信息进行深度挖掘, 并利用残差网络对融合后的多级信息进行学习, 进而提高图像重建质量; 在低压缩比信道环境较好的情况下, 依然能够达到与 Aided Deep-JSCC 方案相同的性能(本文所提方案中并未使用 SSIM 作为损失函数)。在高压缩比情况下, Aided Deep-JSCC 的重建质量最差。这是由于: Aided Deep-JSCC 方案中的每个传感器都独立于图像进行编码, 在不同编码器之间引入了语义信息冗余, 造成不必要的带宽浪费。而本文所提出的方案以联合的方式提取多级语义信息, 强制不同编码器获得的语义信息不同, 相对提高了不同压缩比在较差物理信道情况下( $SNR < 10$  dB) SSIM 的评分, 进一步证明了所提方案在带宽有限、信道条件恶劣情况下的优势。

图 7 比较了本文所提方案和基准方法在 Kodak 数据集上的 PSNR 性能表现。在 AWGN 信道环境下压缩比  $k/n$  设置为  $1/16$ , 数字通信方案信源编解码采用 JPEG, 信道编解码为低密度奇偶检查码(LDPC), 使用 4 符号正交幅度调制(QAM)数字调制方案。如图 7 所示, 本文所提方案优于其他基于深度学习的方法, 表现出多级语义信息在无线图像传输上的优势。同时, 采用数字通信方案的图像传输质量远低于基于深度学习的方法。造成该现象的主要原因是传统方案对信道变化引起的信道错误率比较敏感。由于可用于压缩的位数是固定的, 当信道环境达到一定阈值后, 数字通信方案图像重建质量不会再提高。在  $SNR_{test} < SNR_{train}$  的情况下, 本文所提系统不会受到数字传输方案中的“悬崖效应”的影响。相反, 本文所提方案随着信噪比值的降低, 性能平滑地



▲图 6 AWGN 信道上具有不同压缩比的 PSNR 和 SSIM 性能比较



▲图7 不同图像无线传输方案在Kodak数据集上PSNR的性能比较

下降，其能够平衡压缩比和鲁棒性之间的重要层面。如果模型以高SNR值训练，则系统的PSNR性能主要由带宽压缩比决定，反之亦然。本文所提方法的PSNR性能，比Deep JSCC高 $1.37\sim1.96$  dB，比Aided Deep-JSCC高 $3.37\sim7.39$  dB，比传统方法高 $6.8\sim9.06$  dB。相较于Aided Deep-JSCC的指标波动性，本文所提方法在面对不同信道环境时具有更平滑的性能表现。如图8所示，相比于静态信道环境，本文所提方案的图像重建质量仍具有较好表现，这说明网络具有估计信道状态的能力。

#### 4 结束语

本文提出了一种基于深度学习的无线图像传输语义通信系统，与其他基于深度学习和基于分离的数字传输方案相比，性能表现优异。仿真结果表明，所提出的语义通信系统的有效性和鲁棒性均优于其他方案。

随着无线通信技术的快速发展，未来6G高效、智能的

通信方式，将给工业生产和大众生活带来颠覆性的变革。语义通信作为新一代通信方式，将助力6G发展。然而，有关语义通信的关键技术和理论基础仍需要进一步发展和完善。本文仅作为语义通信的一次尝试和探索。语义通信技术的发展还需要更多的专家学者共同推进。

#### 参考文献

- [1] 徐英姿, 刘原, 时梦然, 等. 语义在通信中的应用综述 [J]. 电信科学, 2022, 38 (S1): 43–59. DOI: 10.11959/j.issn.1000-0801.2022107
- [2] 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. DOI: 10.11959/j.issn.1000-436x.2022117
- [3] CALVANESE STRINATI E, BARBAROSSA S. 6G networks: beyond Shannon towards semantic and goal-oriented communications [J]. Computer networks, 2021, 190: 107930. DOI: 10.1016/j.comnet.2021.107930
- [4] STRINATI E C, BELOT D, FALEMPIN A, et al. Toward 6G: from new hardware design to wireless semantic and goal-oriented communication paradigms [C]//Proceedings of ESSCIRC 2021 – IEEE 47th European Solid State Circuits Conference (ESSCIRC). IEEE, 2021: 275–282. DOI: 10.1109/ESSCIRC53450.2021.9567793
- [5] WANG C X, HUANG J, WANG H M, et al. 6G oriented wireless communication channel characteristics analysis and modeling [EB/OL]. [2023-01-10]. <https://arxiv.org/abs/2007.13958>
- [6] UMAIR M M, MILAN S, GIUSEPPE V, et al. Improved deep distributed light field coding [J]. IEEE open journal of circuits and systems, 2021, 2: 325–337
- [7] YE H, LI G Y, JUANG B H. Power of deep learning for channel estimation and signal detection in OFDM systems [J]. IEEE wireless communications letters, 2017, 7(1): 114–117. DOI: 10.1109/LWC.2017.2757490
- [8] WANG Y, LIU M, YANG J, et al. Data-driven deep learning for automatic modulation recognition in cognitive radios [J]. IEEE transactions on vehicular technology, 2019, 68(4): 4074–4077. DOI: 10.1109/TVT.2019.2900460
- [9] TSE D, VISWANATH P. Fundamentals of wireless communication [M]. Cambridge: Cambridge University Press, 2005
- [10] BASU P, BAO J, DEAN M, et al. Preserving quality of information by using semantic relationships [J]. Pervasive and mobile computing, 2014, 11: 188–202. DOI: 10.1016/j.pmcj.2013.07.013
- [11] ZHOU Q Y, LI R P, ZHAO Z F, et al. Semantic communication with adaptive universal transformer [J]. IEEE wireless communications letters, 2022, 11(3): 453–457. DOI: 10.1109/LWC.2021.3132067
- [12] YE H, LI G Y, JUANG B H. Deep learning based end-to-end wireless communication systems without pilots [J]. IEEE transactions on cognitive communications and networking, 2021, 7(3): 702–714. DOI: 10.1109/TCNN.2021.3061464
- [13] YE H, LIANG L, LI G Y, et al. Deep learning-based end-to-end wireless communication systems with conditional GANs as unknown channels [J]. IEEE transactions on wireless communications, 2020, 19(5): 3133–3143. DOI: 10.1109/TWC.2020.2970707
- [14] AOUDIA F A, HOYDIS J. Model-free training of end-to-end communication systems [J]. IEEE journal on selected areas in communications, 2019, 37(11):



▲图8 在SNR=13 dB和压缩比为1/16的情况下不同方法的可视化比较

- 2503–2516. DOI: 10.1109/JSAC.2019.2933891
- [15] RAJ V, KALYANI S. Backpropagating through the air: deep learning at physical layer without channel models [J]. IEEE communications letters, 2018, 22(11): 2278–2281. DOI: 10.1109/LCOMM.2018.2868103
- [16] YE H, LI G Y, JUANG B H F, et al. Channel agnostic end-to-end learning based communication systems with conditional GAN [C]//Proceedings of 2018 IEEE Globecom Workshops (GC Wkshps). IEEE, 2018: 1–5. DOI: 10.1109/GLOCOMW.2018.8644250
- [17] FARSAZ N, RAO M, GOLDSMITH A. Deep learning for joint source-channel coding of text [C]//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018: 2326–2330. DOI: 10.1109/ICASSP.2018.8461983
- [18] CHEN X, CHENG J L, ZHANG Z C, et al. Data-rate driven transmission strategies for deep learning-based communication systems [J]. IEEE transactions on communications, 2020, 68(4): 2129–2142. DOI: 10.1109/TCOMM.2020.2968314
- [19] XIE H Q, QIN Z J, LI G Y, et al. Deep learning enabled semantic communication systems [J]. IEEE transactions on signal processing, 2021, 69: 2663–2675. DOI: 10.1109/TSP.2021.3071210
- [20] WENG Z Z, QIN Z J. Semantic communication systems for speech transmission [J]. IEEE journal on selected areas in communications, 2021, 39(8): 2434–2444. DOI: 10.1109/JSAC.2021.3087240
- [21] TONG H N, YANG Z H, WANG S H, et al. Federated learning for audio semantic communication [J]. Frontiers in communications and networks, 2021, 2: 43. DOI: 10.3389/frcmn.2021.734402
- [22] BOURTSOULATZE E, BURTH KURKA D, GÜNDÜZ D. Deep joint source-channel coding for wireless image transmission [J]. IEEE transactions on cognitive communications and networking, 2019, 5(3): 567–579. DOI: 10.1109/TCCN.2019.2919300
- [23] KURKA D B, GÜNDÜZ D. DeepJSCC-f: deep joint source–channel coding of images with feedback [J]. IEEE journal on selected areas in information theory, 2020, 1(1): 178–193. DOI: 10.1109/JSAIT.2020.2987203
- [24] YANG M Y, BIAN C H, KIM H S. Deep joint source channel coding for wireless image transmission with OFDM [C]//Proceedings of ICC 2021 – IEEE International Conference on Communications. IEEE, 2021: 1–6. DOI: 10.1109/ICC42927.2021.9500996
- [25] 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. DOI: 10.1109/ACCESS.2019.2920929
- [26] JANKOWSKI M, GÜNDÜZ D, MIKOŁAJCYK K. Wireless image retrieval at the edge [J]. IEEE journal on selected areas in communications, 2021, 39(1): 89–100. DOI: 10.1109/JSAC.2020.3036955
- [27] JANKOWSKI M, GÜNDÜZ D, MIKOŁAJCYK K. Joint device–edge inference over wireless links with pruning [C]//Proceedings of 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2020: 1–5. DOI: 10.1109/SPAWC48557.2020.9154306
- [28] KANG X, SONG B, GUO J, et al. Task-oriented image transmission for scene classification in unmanned aerial systems [J]. IEEE transactions on communications, 2022, 70(8): 5181–5192. DOI: 10.1109/TCOMM.2022.3182325
- [29] XIE H Q, QIN Z J, TAO X M, et al. Task-oriented multi-user semantic communications [J]. IEEE journal on selected areas in communications, 2022, 40(9): 2584–2597. DOI: 10.1109/JSAC.2022.3191326
- [30] LIU X X, XU Q Y. Adaptive attention-based high-level semantic introduction for image caption [J]. ACM transactions on multimedia computing, communications, and applications, 2020, 16(4): 1 – 22. DOI: 10.1145/3409388
- [31] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016: 770–778. DOI: 10.1109/CVPR.2016.90
- [32] YU Y, SI X S, HU C H, et al. A review of recurrent neural networks: LSTM cells and network architectures [J]. Neural computation, 2019, 31(7): 1235–1270. DOI: 10.1162/neco\_a\_01199[PubMed]
- [33] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network [C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 6230–6239. DOI: 10.1109/CVPR.2017.660
- [34] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: common objects in context [EB/OL]. [2023-01-10]. <https://arxiv.org/abs/1405.0312>
- [35] ZHOU B L, ZHAO H, PUIG X, et al. Scene parsing through ADE20K dataset [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 5122–5130. DOI: 10.1109/CVPR.2017.544
- [36] FRANZEN R. Kodak lossless true color image suite [EB/OL]. [2023-01-10]. <https://r0k.us/graphics/kodak/>
- [37] PASZKE A, GROSS S, MASSA F, et al. Pytorch: an imperative style, high-performance deep learning library [EB/OL]. [2023-01-10]. <https://arxiv.org/abs/1912.01703>
- [38] KINGMA D P, BA J. Adam: a method for stochastic optimization [EB/OL]. [2023-01-10]. <https://arxiv.org/abs/1412.6980>
- [39] YAN J T, HUANG J H, HUANG C. Deep learning aided joint source–channel coding for wireless networks [C]//Proceedings of 2021 IEEE/CIC International Conference on Communications in China (ICCC). IEEE, 2021: 805–810. DOI: 10.1109/ICCC52777.2021.9580301

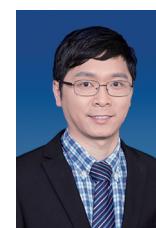
## 作者简介



张振国，浙江大学在读博士研究生；主要研究方向为基于深度学习的语义通信技术。



杨倩倩，浙江大学百人计划研究员、IMT-2030无线AI推进组成员，IEEE VTC 2022、IEEE ICC 2023等旗舰会议的技术委员会成员/分会主席，《IEEE Transactions on Wireless Communications》《IEEE Journal on Selected Areas in Communications》等刊物的审稿人；主要研究领域为无线通信、信息理论、语义通信；获中国科技产业化促进会科技创新奖一等奖等奖项；发表论文40余篇。



贺诗波，浙江大学控制科学与工程学院院长聘教授，工业信息物理融合系统省部共建协同创新中心主任、工程师学院双副院长、工业控制研究所副所长，2014年入选国家青年人才计划、浙江大学百人计划，IEEE高级会员、IET Fellow；获得教育部自然科学奖一等奖、IEEE通信学会亚太区杰出青年研究学者奖、中国自动化学会首届优秀博士学位论文奖，并获IEEE iThings、IEEE RFID-TA、IEEE WCNC、IEEE PIMRC最佳论文奖等奖项。