

# 面向语义通信的非线性变换编码

张平<sup>1</sup>, 戴金晟<sup>2</sup>, 张育铭<sup>2</sup>, 王思贤<sup>2</sup>, 秦晓琦<sup>1</sup>, 牛凯<sup>1</sup>

(1. 北京邮电大学网络与交换技术国家重点实验室, 北京 100876; 2. 北京邮电大学泛网无线通信教育部重点实验室, 北京 100876)

**摘要:** 针对经典通信系统模块设计分离、处理范式受限等因素限制了端到端信息传输性能持续提升的问题, 提出了面向语义通信的非线性变换编码传输新框架。首先, 基于变分理论推导了语义通信端到端率失真优化准则。据此, 设计了非线性变换来提取信源数据在语义隐空间的紧致表征, 并通过语义变分熵建模引导实现了变速率非线性联合信源信道编码。实验表明, 语义非线性变换编码能显著提升端到端数据传输性能及鲁棒性, 是实现语义通信的关键技术之一。

**关键词:** 语义通信; 非线性变换; 非线性编码; 变分熵建模; 率失真优化

**中图分类号:** TN92

**文献标志码:** A

**DOI:** 10.11959/j.issn.1000-436x.2023087

## Nonlinear transform coding for semantic communications

ZHANG Ping<sup>1</sup>, DAI Jincheng<sup>2</sup>, ZHANG Yuming<sup>2</sup>, WANG Sixian<sup>2</sup>, QIN Xiaoqi<sup>1</sup>, NIU Kai<sup>1</sup>

1. The State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China  
2. The Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China

**Abstract:** The modular design and limited processing mechanism of traditional communication systems limit the continuous improvement of end-to-end data transmission capability. For this reason, a new nonlinear transform coding framework for semantic communications was proposed. First, an end-to-end rate distortion optimization criterion for semantic communication was derived based on variational theory. Based on this, a nonlinear transform was designed to extract the compact representation of source data in the semantic latent space, and variable-rate nonlinear joint source-channel coding was implemented through the guidance of variational entropy model. Experiments show that semantic nonlinear transform coding can significantly improve the end-to-end data transmission performance and robustness, and is one of the key technologies to catalyze future semantic communications.

**Keywords:** semantic communications, nonlinear transform, nonlinear coding, variational entropy modeling, rate-distortion optimization

## 0 引言

20 世纪 50 年代, 信息论创始人香农和信息论著名学者 Weaver 针对通信的本原进行了深层次思考<sup>[1]</sup>, 将通信问题分成技术(语法)层、语义层、效用(语用)层 3 个层次。其中, 基于统计概率的经典通信属于最基本的语法层。从通信技术演进历

程可以看出, 香农在 1948 年创立信息论<sup>[2]</sup>以后, 经典通信在信息压缩、信息传输、信息处理等方面快速发展; 但随着经典通信的快速进步, 其发展也遇到多重瓶颈。在信息压缩层面, 目前语音、图像、视频压缩编码技术已经逼近信息熵或率失真函数理论极限, 继续压缩复杂度开销巨大且收效甚微; 在信息传输层面, 当前信道传输能力已经逼近香农

收稿日期: 2022-12-19; 修回日期: 2023-04-11

通信作者: 戴金晟, daijincheng@bupt.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62293481, No.92067202, No.62001049, No.62071058)

**Foundation Item:** The National Natural Science Foundation of China (No.62293481, No.92067202, No.62001049, No.62071058)

信道容量极限，开拓频谱、提高功率、增加天线抬升容量限的代价巨大，难以可持续发展；在信息处理层面，随着后摩尔时代的到来，芯片算力提升空间殆尽。综上，经典通信发展遭遇信息压缩、传输、处理三重瓶颈，亟待另辟蹊径，进行理论与技术创新。

回顾通信问题的 3 个层次，由于历史局限，香农在 1948 年仅考虑消息符号的传输，忽略了语义信息。而信息论著名学者 Brillouin 在 1962 年即指出：香农信息论忽略信息的意义是因为当年通信工程的特殊需要，但并不意味人们要永远忽略。香农在 1948 年提出的经典信息论仅考虑消息符号的可靠传输，不考虑在更高的语义层对通信系统进行设计。得益于人工智能在算法、算力、数据方面的巨大进步，近年来逐渐兴起的语义信息传输正在让通信技术回归本原，它们试图将本地智能作为系统设计的组成部分，从而实现从技术层到语义层的跨层升级<sup>[3-4]</sup>。基于语义信息表征和压缩编码问题的探索<sup>[5-6]</sup>，通过挖掘信息本身的新维度，利用信源信道的联合设计，赋能端到端通信系统的整体性能跃升<sup>[7-10]</sup>。

从技术层具体来看，香农信息论指导下的经典通信系统设计在研究范畴、研究层次与研究维度方面存在局限，一定程度上限制了信息通信系统性能的持续提升。如图 1 所示，主要体现为以下 4 个方面。

1) 模块分离设计。依据香农信源信道分离定理，在编码码长、时延、复杂度均不受限的条件下，信源压缩与信道传输模块可以进行分离设计，两者独立的优化等价于端到端系统全局优化。过去几十年，该原理被应用于通信系统的各个模块（如信道编码、信号调制等），简化了系统设计。但实际系统的编码码长、时延、复杂度均受到限制，尤其是对于实时性要求较高的新业务通信系统，由于编码

码长和编译码复杂度等限制更严格，模块分离设计相比于系统联合优化设计有明显性能损失。

2) 处理范式受限。现有通信系统的编译码、调制解调等模块大多采用线性处理，虽然该范式简化了系统设计，且在一定条件下可以推导得到模块最优处理的解析表达式，但线性处理范式限制了模块处理能力的提升，无法进一步满足系统性能提升需求。以深度学习为代表的人工智能技术正是利用了神经网络的非线性处理机制，显著提升了信息处理能力，为通信系统各模块处理机制的创新提供了思路。

3) 优化准则单一。经典通信系统以香农信息论为理论基础，以准确传输数据或精确传送信号波形为目标，而其中承载的内容信息是什么以及信息被如何使用并未受到特别关注。在此背景下，通信系统大多数模块的数学推导都基于高斯噪声假设，因此模块优化设计的优化准则大多为最小均方误差（MSE），如信道估计、信号检测、信道编译码等均是如此。然而，当考虑范畴更广的端到端通信系统时，简单的 MSE 准则无法准确匹配信源侧人类主观感知体验或智能机器任务表现，导致全系统优化效率降低。

4) 先验知识不足。经典信源压缩与信道传输模块均基于统计概率特征进行构建，没有考虑具体通信场景中的先验知识来辅助提升端到端传输性能，尤其没有涉及更高层的语义先验信息，也没有考虑信息传输过程推进中先验知识的更新。以上先验知识形态、使用、更新几方面的不足在一定程度上制约了系统性能的进一步提升。

鉴于此，本文在通信系统引入非线性联合信源信道处理和多元目标引导的端到端优化 2 个新机制来提升端到端传输性能，克服传统通信系统的技术瓶颈。通过对非线性变换编码传输系统的深入研究，具体回答语义特征如何提取、如何传输、如何

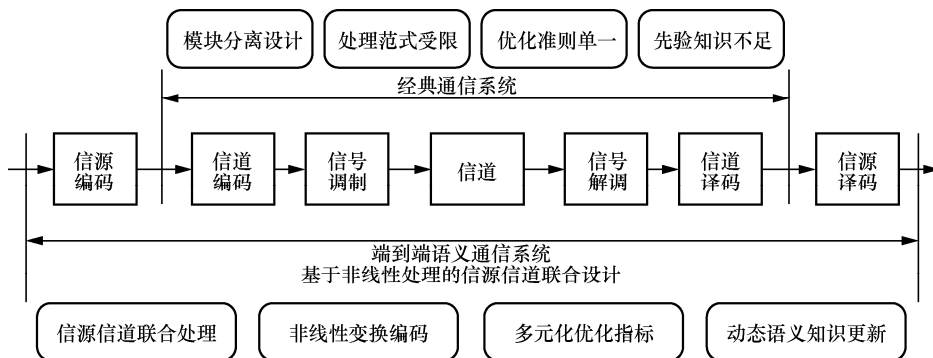


图 1 经典通信系统与端到端语义通信系统对比

使用三大核心问题。具体来看, 本文的主要研究工作如下。

1) 提出了非线性变换编码传输系统框架, 基于变分理论推导了语义通信的端到端传输速率-失真 (RD, rate-distortion) 二元约束优化准则。据此, 建立了语义通信系统各模块的基本数学模型, 用于指导模块间的匹配设计, 实现端到端联合优化。进一步设计了面向人类感知体验的端到端传输速率-失真-感知 (RDP, rate-distortion-perceptual) 三元约束优化准则, 通过引入生成模型, 有效提升端到端数据传输的人类感知体验。

2) 通过非线性变换, 得到原始数据在语义隐空间的紧致表征。通过在语义隐空间引入变分熵建模, 估计数据的语义特征熵分布, 实现对数据不同区域复杂度的准确评估。据此, 设计了变速率的联合信源信道编码 (JSCC, joint source-channel coding) 方法, 实现了通信传输资源和语义表征信息量之间的准确匹配, 获得了系统层的编码增益。

3) 开展了广泛的实验验证, 结果表明, 所提语义编码传输方法在端到端传输性能与鲁棒性方面明显优于现有工程应用的编码传输方法, 验证了语义通信系统的有效性。

## 1 经典信源信道编码技术

在过去几十年中, 大量学术界和工业界研究致力于信源压缩与信道传输技术, 这是构成端到端通信系统最主要的 2 个部分, 也是香农信息论的 2 个主要分支。本节首先回顾了信源压缩的经典信源编码技术, 总结了几条重要准则; 然后回顾了用于可靠传输的信道编码技术, 并基于此引出了联合信源信道编码技术。

### 1.1 信源编码技术

信源压缩的核心思路是利用数据分布的不等概率和上下文相关性, 对出现概率更高的数据符号分配更短的代表序列, 对出现概率更低的数据符号分配更长的代表序列, 可以实现有效数据压缩。香农的无失真信源编码定理和限失真信源编码定理在事实上给出了信源压缩算法设计的几条重要准则。

1) 尽可能准确估计信源分布。相比于真实分布, 更准确的估计分布将会使它们之间的相对熵尽可能小, 从而降低信源编码的码率, 提高压缩效率。

2) 变长编码。香农最优码长分配准则表明任意

信源样本对应的码长应正比于估计分布得到的信息量 (因此, 对熵的估计也被称为熵模型), 本质上则需要对样本进行变长编码, 从而有效提高信源编码效率。

3) 率失真优化。广泛使用的限失真信源编码设计本质上是编码码率与失真之间的折中。

依据上述信息论分析, 信源编码技术主要使用矢量量化 (VQ, vector quantization)、熵约束矢量量化 (ECVQ, entropy-constrained vector quantization) 和变换编码 (TC, transform coding) 3 种范式。理论上, VQ、ECVQ 是最优压缩方法, 但两者的设计复杂度随信源维度呈指数级上升, 只能处理低维简单信源, 难以实用化。为解决上述两类矢量量化方法的问题, 学术界提出变换编码的范式, 成为当前信源编码的主流方法。TC 的核心思路是将原始信源变换到新的变换域进行表征, 由于变换域中表征的各维度被较好地解除了相关性, 可采用简单的标量量化对各维度分别进行处理, 并根据所估计的熵模型进行熵编码, 从而大幅降低复杂度。

大多数围绕 TC 的理论均建立在信源服从联合高斯分布的基本假设上, 此时可以推导得到变换函数的最优解。经典信源编码技术基本采用线性变换编码 (LTC, linear transform coding) 的架构, 其中的解析变换及合成变换函数是显式手工设计的。例如, 现代视频编码标准中广泛采用的离散余弦变换 (DCT, discrete cosine transform) 结合熵编码的架构是典型的 LTC 范式。然而, LTC 只能提取到信源数据的浅层表征, 难以对复杂信源高效解相关, 导致熵模型无法逼近信源的真实分布。鉴于此, 学术界提出了非线性变换编码 (NTC, nonlinear transform coding) 技术<sup>[11]</sup>, 其利用深度神经网络来构建解析变换和合成变换函数, 熵模型也利用深度神经网络来构建。通过端到端学习, NTC 中的熵模型可以有效逼近真实信源分布, 获得更好的 RD 性能。从人工智能表征学习角度看, 基于深度神经网络的非线性变换操作提取了数据更紧致的语义表征, 利用语义隐空间的熵建模对信源语义特征做了很好的估计, 从而得到精准的变长熵编码, 实现对信源的高效压缩。

### 1.2 信道编码技术

香农关于信道传输的有噪信道编码定理 (香农第二定理) 是经典通信技术的理论基础。

有噪信道编码定理指出, 如果信息传输速率  $R$

小于信道容量  $C$ ，则总可找到一种编码方式，使当编码序列足够长时传输差错任意小；反之，则不存在使差错任意小的编码。

信道编码定理在事实上也给出了信息传输设计的几条重要准则。

1) 动态码率控制。信道编码码率  $R_c$  与信号调制阶数  $m$  需要随信道动态变化进行调整，以保证  $R < C$ 。

2) 多轮增量传输。若信道译码出错，可以尝试重复增量传输部分编码比特，补偿性能损失。本质上是在调整全局编码码率，以保证其小于即时信道容量。

根据上述定理，香农首先给出了分离式的信源信道编码范式，即先将信源数据压缩为最有效表征，再设计合适的信道编码来传输该表征。香农证明，当信源和信道均达到渐进均分特性（AEP, asymptotic equipartition property）时，分离设计不会影响系统整体性能的最优性，这是著名的“信源信道分离定理”。该定理为后续通信系统的工程研究提供了重要指导，并在理论上证明信源压缩技术和信道传输技术可以分离优化，极大降低了通信系统工程设计的复杂度。

信道编码的设计主要目标是在有限码长下提高纠错能力，逼近香农信道容量极限。目前，5G 所采用的信道编码方式主要有两类。

1) 低密度奇偶校验（LDPC, low-density parity-check）码<sup>[12]</sup>：主要用于数据信道的数据传输，码长一般较长。

2) 极化码<sup>[13]</sup>：主要用于控制信道的信令传输，码长一般较短。

此外，根据香农第二定理，信道编码码率  $R_c$  与信号调制阶数  $m$  的组合应该随着实时信道状态自适应调整，即无线通信中的链路自适应技术。其主要目的是根据链路反馈的信道状态信息（CSI），选择合适的  $m$  与  $R$  的组合，使其传输平均误块率能稳定在一定数值范围内。此外，信道编码还会与重传技术结合，形成混合自动重传请求（HARQ, hybrid automatic repeat request）技术，这是链路传输可靠性的重要保障。

### 1.3 联合信源信道编码技术

尽管信源信道分离设计长期以来就是通信系统设计的基本范式，本文也观察到一些似乎违反“分离定理”的生活常识。一个简单的例子是通过随机删除信道传输文本信息，可以寻找最高效的信

源编码先将原始文本变为表征（如比特序列），并将其通过信道编码后经信道传输。但是，当信道条件恶劣导致信道译码发生差错时，信源译码将会出现严重错误，导致原始文本语义内容无法被理解。反之，如果直接向信道发送原始文本，可能会失去部分字符，然而依然可以大致理解文本的含义。类似地，人脑可以从叠加了白噪声的语音信号中分辨出说话人所表达的内容。这些例子的本质原因是信源的冗余天然适配了信道条件。这一观察启发了联合信源信道编码技术的研究，这对实际端到端传输系统的设计具有重要意义。

依据上述理论分析，香农提出了联合信源信道编码的概念，可以设计一个编码器，直接将信源数据映射为信道传输的符号。联合信源信道编码理论上能够在有效提取信源表征的同时，更合理地分配信道资源，使系统达到端到端性能最优。但香农并未给出联合信源信道编码设计的统一理论指导，依靠专家经验和手工设计难以得到同时满足高效率信源压缩和高性能差错控制的编码器，之后研究者对 JSCC 方案的探索大多数也仅适用于具有明确分布的简单信源，没有形成能够工程化应用的成熟编码传输方案。

## 2 语义非线性编码传输

语义编码传输将表征学习、信源编码、信道编码三者匹配融合设计，实现端到端数据高效率、高保真传输。表征学习提取信源的语义特征，联合信源信道编码实现对语义表征的高效、鲁棒传输。整个链路的各模块均采用非线性处理范式实现，提升语义表征与传输能力。

### 2.1 直接编码传输

第一类语义非线性编码传输系统是端到端一体设计。如图 2 所示，发送端将语义特征提取、信源信道编码封装为一个编码器模块，接收端将信源信道译码、语义特征融合封装为一个译码器模块。编译码器均采用深度神经网络（DNN, deep neural network）实现，获得非线性处理增益。该方法称为深度联合信源信道编码（Deep JSCC, deep joint source-channel coding）<sup>[14-21]</sup>。本质上，该结构源于深度自编码器（AE, auto-encoder）结构，但端到端训练编译码器的过程考虑了中间瓶颈层数据经过信道传输引入的失真，从而增加了编译码器对抗信道噪声、衰落等不利因素的能力。

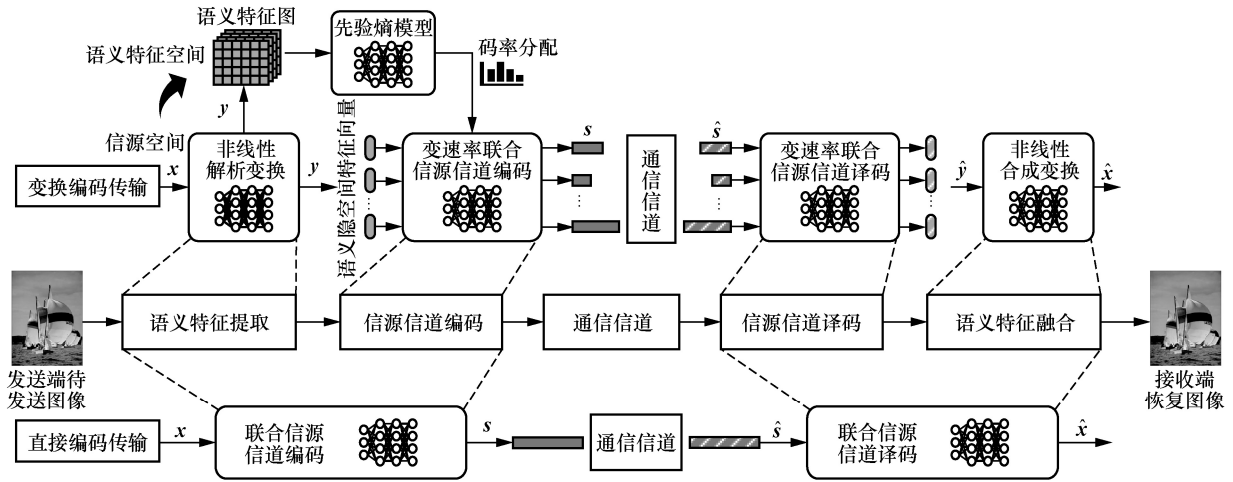


图 2 两类语义非线性编码传输系统结构对比

从香农信息论角度来看，Deep JSCC 可以视为信源高维矢量到信道传输低维矢量的映射。以图像信源为例，Deep JSCC 的操作逻辑为将一张图像表示为像素点组成的一个  $m$  维矢量  $\mathbf{x} \in \mathbb{R}^m$ ，通过一个基于 DNN 的编码函数映射  $\mathbf{s} = f_e(\mathbf{x}; \phi_f)$  到一个  $k$  维的信道输入矢量  $\mathbf{s} \in \mathbb{R}^k$ ，其中  $\phi_f$  表示构成编码器的 DNN 参数集合。通常有  $k < m$ ， $\rho = \frac{k}{m}$  为信道带宽比 (CBR, channel bandwidth ratio) [16]，表示信道输入维度与信源维度的比值。序列  $\mathbf{s}$  直接经由信道发送，信道向传输的符号序列  $\mathbf{s}$  引入随机误差，记为  $W(\cdot; \mathbf{v})$ ，信道参数封装在  $\mathbf{v}$  中，接收符号序列为  $\hat{\mathbf{s}} = W(\mathbf{s}; \mathbf{v})$ ，转移概率为  $p_{\hat{s}_i | s_i}(\hat{s}_i | s_i)$ 。以典型的加性白高斯噪声 (AWGN) 信道为例，信道传输方程为  $\hat{\mathbf{s}} = W(\mathbf{s}; \sigma^2) = \mathbf{s} + \mathbf{n}$ ，其中，噪声分量  $\mathbf{n}$  服从高斯分布  $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_k)$ ， $\sigma^2$  是噪声功率。接收端采用解码函数  $\hat{\mathbf{x}} = f_d(\hat{\mathbf{s}}; \theta_f)$  从接收符号序列  $\hat{\mathbf{s}}$  中恢复出信源序列  $\hat{\mathbf{x}}$ 。Deep JSCC 的编码和解码函数  $f_e$  和  $f_d$  可使用卷积神经网络 (CNN, convolutional neural network) 来构建。

Deep JSCC 端到端通信系统操作流程如图 3 所示，其优化目标为最小化信源与接收端重建信源的误差，其端到端优化损失函数为

$$L(\phi_f, \theta_f) = \mathbb{E}_{x \sim p_x} d(\mathbf{x}, \hat{\mathbf{x}}) \quad (1)$$

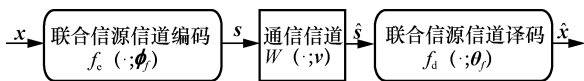


图 3 Deep JSCC 端到端语义通信系统操作流程

以 Deep JSCC 为代表的直接编码传输方法本质上采用了“定长编码”思路。当编译码器端到端训练好之后，对任意样本  $\mathbf{x}$ ，其信道传输符号数量  $k$  都是定值，且没有考虑信源数据内部在语义内容复杂性上的差异，导致系统编码效率低下。从已有实验结果来看，以 Deep JSCC 为代表的直接编码传输方式仅对极低维度的图像（如 CIFAR10 数据集中 32 像素  $\times$  32 像素的图像）端到端传输问题可以获得比经典信源信道方法更好的传输性能，核心原因是目前工程应用的主流图像、视频信源编码方法（如 BPG、VTM 等）均是针对高分辨率图像、视频设计的，并不针对极低维度信源做优化，导致 Deep JSCC 等直接编码传输方法相对而言获得了一定性能优势，但这种优势不具有普适性。

## 2.2 变换编码传输

不同于上述直接编码传输思路，本节从变换编码的角度出发，构建更高效的端到端语义编码传输系统框架 [22]，核心思路是根据信源内容分布实现与之匹配的变速率非线性联合信源信道编码。

### 2.2.1 系统架构

如图 2 所示，该系统架构继承了 NTC 与 JSCC 的双重优势，被称为非线性变换信源信道编码 (NTSCC, nonlinear transform source-channel coding)。在发送端，非线性解析变换  $g_a(\cdot; \phi_g)$  用于提取信源样本  $\mathbf{x}$  的深层语义特征，构成隐空间语义特征图  $\mathbf{y}$ 。 $\mathbf{y}$  将会继续被送入变速率联合信源信道编码  $f_e(\cdot; \phi_f)$  得到信道输入矢量  $\mathbf{s}$ 。给定信道传输函数  $W(\cdot; \mathbf{v})$ ，则接收符号序列为  $\hat{\mathbf{s}} = W(\mathbf{s}; \mathbf{v})$ 。接收端先将  $\hat{\mathbf{s}}$  送入变速率联合信源信道译码  $f_d(\cdot; \theta_f)$  恢复得到关于隐

空间语义特征图的估计  $\hat{\mathbf{y}}$ ，再送入非线性合成变换  $g_s(\cdot; \theta_g)$  进行语义特征融合重构信源数据  $\hat{\mathbf{x}}$ 。整个 NTSCC 系统的流程为

$$\mathbf{x} \xrightarrow{g_a(\cdot)} \mathbf{y} \xrightarrow{f_c(\cdot)} \mathbf{s} \xrightarrow{W(\cdot)} \hat{\mathbf{s}} \xrightarrow{f_d(\cdot)} \hat{\mathbf{y}} \xrightarrow{g_s(\cdot)} \hat{\mathbf{x}} \quad (2)$$

NTSCC 系统的关键是：语义隐空间表征矢量  $\mathbf{y}$  会被送入先验熵模型  $p_y$ ，计算得到语义表征  $\mathbf{y}$  每个维度上取值对应的概率，从而获得原始数据  $\mathbf{x}$  在语义隐空间  $\mathbf{y}$  上的信息量分布。 $p_y$  实际是由 DNN 所定义的一个参数化分布，参数集合记为  $\boldsymbol{\varphi}$ 。先验熵模型  $p_y(\mathbf{y}; \boldsymbol{\varphi})$  的计算式为

$$p_y(\mathbf{y}; \boldsymbol{\varphi}) = \prod_i \underbrace{\left( p(y_i; \boldsymbol{\varphi}^{(i)}) * \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right) \right)}_{p_{y_i}(\cdot; \boldsymbol{\varphi}^{(i)})} (y_i) = \prod_i p_{y_i}(y_i; \boldsymbol{\varphi}^{(i)}) \quad (3)$$

其中， $p(y_i; \boldsymbol{\varphi}^{(i)})$  表示由参数  $\boldsymbol{\varphi}^{(i)}$  确定的关于  $y_i$  的熵模型， $*$  表示卷积操作， $\mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$  表示  $\left(-\frac{1}{2}, \frac{1}{2}\right)$  范围内的均匀分布，卷积形成的概率分布  $p_{y_i}(\cdot; \boldsymbol{\varphi}^{(i)})$  表示关于  $y_i$  的代理熵模型。卷积均匀分布的目的是使  $p_{y_i}(\cdot; \boldsymbol{\varphi}^{(i)})$  在任何  $y_i$  取值点上的概率都落在 (0,1) 范围内，保证后续信息量计算的数值稳定。

在先验熵模型的引导下，每个信源样本  $\mathbf{x}$  所对应的信道输入矢量  $\mathbf{s}$  的总维度是动态变化的，并且可以实现每个语义隐空间的表征矢量  $\mathbf{y}_i$  对应不同的编码码率，即  $\mathbf{y}_i$  编码得到的传输矢量  $\mathbf{s}_i$  的维度不一样，可以依据  $\mathbf{y}_i$  所对应的熵大小来确定。语义隐空间的先验熵模型结合联合信源信道编码的码率分配策略，使 NTSCC 实现变速率编码传输，这是其相对于 Deep JSCC 直接编码传输方案获得大幅性能提升的本质原因。

### 2.2.2 变分建模

从端到端系统角度来看，面向数据传输任务的语义通信目标是使接收端恢复的数据分布与发送端的真实数据分布尽可能一致，从而不仅实现数据元素级别的恢复，而且追求全局视野的感知体验优化，这符合深度学习“生成模型”的思想。因此，可以从变分建模的角度，推导出语义通信编码传输的率失真优化准则。

从变分自编码器 (VAE, variational auto-encoder) 角度来看，图 4 所示的 NTSCC 系统对以下几个概率分布进行了建模。

1)  $q_{\hat{\mathbf{s}}|\mathbf{x}}(\hat{\mathbf{s}}|\mathbf{x})$  表示给定信源样本  $\mathbf{x}$  时，关于接收符号序列  $\hat{\mathbf{s}}$  的条件概率分布。 $q_{\hat{\mathbf{s}}|\mathbf{x}}(\hat{\mathbf{s}}|\mathbf{x})$  由非线性解析变换  $g_a(\cdot; \boldsymbol{\varphi}_g)$ 、非线性联合信源信道编码  $f_c(\cdot; \boldsymbol{\varphi}_f)$ 、通信信道  $W(\cdot; \boldsymbol{\nu})$  共同决定，因此  $q_{\hat{\mathbf{s}}|\mathbf{x}}$  是关于参数  $\{\boldsymbol{\varphi}_g, \boldsymbol{\varphi}_f, \boldsymbol{\nu}\}$  的概率模型。

2)  $p_{\hat{\mathbf{s}}}(\hat{\mathbf{s}})$  表示关于接收符号序列  $\hat{\mathbf{s}}$  的可学习的先验概率分布。在给定非线性解析变换  $g_a(\cdot; \boldsymbol{\varphi}_g)$ 、非线性联合信源信道编码  $f_c(\cdot; \boldsymbol{\varphi}_f)$ 、通信信道  $W(\cdot; \boldsymbol{\nu})$  后， $p_{\hat{\mathbf{s}}}(\hat{\mathbf{s}})$  是由语义隐空间熵模型  $p_y(\mathbf{y}; \boldsymbol{\varphi})$  及信道转移概率  $p_{\hat{\mathbf{s}}|\mathbf{s}}(\hat{\mathbf{s}}|\mathbf{s})$  共同确定的，其计算式为

$$p_{\hat{\mathbf{s}}}(\hat{\mathbf{s}}) = \sum_{\mathbf{y}} p_{\hat{\mathbf{s}}|\mathbf{s}}(\hat{\mathbf{s}}|f_c(\mathbf{y}; \boldsymbol{\varphi}_f)) p_y(\mathbf{y}; \boldsymbol{\varphi}) \quad (4)$$

3)  $p_{\mathbf{x}|\hat{\mathbf{s}}}(\mathbf{x}|\hat{\mathbf{s}})$  表示接收端得到接收符号序列  $\hat{\mathbf{s}}$  时，关于恢复数据  $\mathbf{x}$  的条件概率分布。该概率分布由非线性联合信源信道译码  $f_d(\cdot; \theta_f)$ 、非线性合成变换  $g_s(\cdot; \theta_g)$  及失真  $d$  共同确定。例如，当使用均方误差 (MSE) 作为失真  $d$  时， $p_{\mathbf{x}|\hat{\mathbf{s}}}(\mathbf{x}|\hat{\mathbf{s}})$  为

$$p_{\mathbf{x}|\hat{\mathbf{s}}}(\mathbf{x}|\hat{\mathbf{s}}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}(\hat{\mathbf{s}}), \tau^2 \mathbf{I}) \quad (5)$$

其中， $\boldsymbol{\mu}(\hat{\mathbf{s}}) = g_s(f_d(\hat{\mathbf{s}}; \theta_f); \theta_g)$  是高斯分布的均值，即模型推理过程中的信源的估计值  $\hat{\mathbf{x}} = \boldsymbol{\mu}(\hat{\mathbf{s}})$ ， $\tau^2$  为

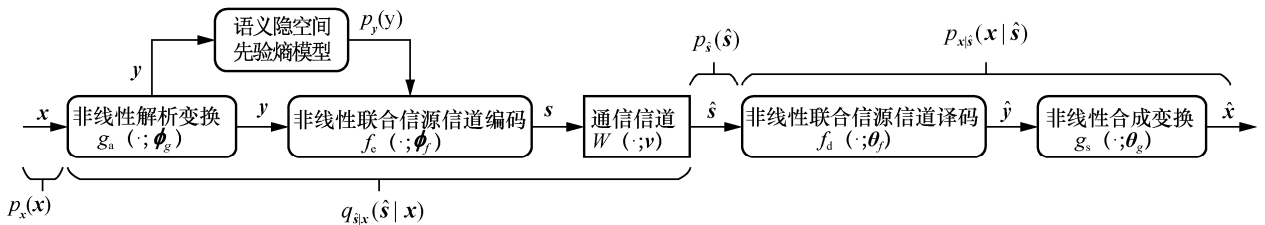


图 4 NTSCC 端到端语义通信流程及其对应的概率建模

常数，是端到端传输的能量约束。

根据上述几个参数化建模的概率分布，结合变分建模理论，NTSCC 系统端到端优化目标为最小化两组联合概率分布之间的 KL (Kullback-Leibler) 散度  $\text{KL}[(p_x q_{\hat{s}|x}) \parallel (p_{\hat{s}} p_{x|\hat{s}})]$ ，即

$$\begin{aligned} & \min_{\phi_g, \phi_f, \theta_g, \theta_f, \varphi} \text{KL}[(p_x q_{\hat{s}|x}) \parallel (p_{\hat{s}} p_{x|\hat{s}})] = \\ & \min_{\phi_g, \phi_f, \theta_g, \theta_f, \varphi} \mathbb{E}_{x \sim p_x} \mathbb{E}_{\hat{s} \sim q_{\hat{s}|x}} \log \frac{p_x(\mathbf{x}) q_{\hat{s}|x}(\hat{\mathbf{s}}|\mathbf{x})}{p_{\hat{s}}(\hat{\mathbf{s}}) p_{x|\hat{s}}(\mathbf{x}|\hat{\mathbf{s}})} = \\ & \min_{\phi_g, \phi_f, \theta_g, \theta_f, \varphi} \mathbb{E}_{x \sim p_x} \mathbb{E}_{\hat{s} \sim q_{\hat{s}|x}} [\log q_{\hat{s}|x}(\hat{\mathbf{s}}|\mathbf{x}) - \log p_{\hat{s}}(\hat{\mathbf{s}}) - \\ & \log p_{x|\hat{s}}(\mathbf{x}|\hat{\mathbf{s}})] + \underbrace{\mathbb{E}_{x \sim p_x} \log p_x(\mathbf{x})}_{\text{常数}_1} = \\ & \min_{\phi_g, \phi_f, \theta_g, \theta_f, \varphi} \underbrace{\mathbb{E}_{x \sim p_x} \mathbb{E}_{\hat{s} \sim q_{\hat{s}|x}} \log q_{\hat{s}|x}(\hat{\mathbf{s}}|\mathbf{x})}_{\text{常数}_2} + \\ & \mathbb{E}_{x \sim p_x} \mathbb{E}_{\hat{s} \sim q_{\hat{s}|x}} \underbrace{[-\log p_{\hat{s}}(\hat{\mathbf{s}})]}_{\text{传输速率}} + \underbrace{[-\log p_{x|\hat{s}}(\mathbf{x}|\hat{\mathbf{s}})]}_{\text{端到端传输失真}} + \text{常数}_1 \quad (6) \end{aligned}$$

经过推导发现，NTSCC 系统优化目标本质上是寻求平均传输速率  $\mathbb{E}_{x \sim p_x} \mathbb{E}_{\hat{s} \sim q_{\hat{s}|x}} [-\log p_{\hat{s}}(\hat{\mathbf{s}})]$  与平均端到端失真  $\mathbb{E}_{x \sim p_x} \mathbb{E}_{\hat{s} \sim q_{\hat{s}|x}} [-\log p_{x|\hat{s}}(\mathbf{x}|\hat{\mathbf{s}})]$  之间的折中。注意到，速率项中的  $p_{\hat{s}}(\hat{\mathbf{s}})$  是由语义空间  $\mathbf{y}$  的熵模型  $p_y(\mathbf{y}; \varphi)$  确定的，据此可以推导得到语义通信编码传输系统的一般化设计准则，即最小化信道传输速率-失真 (RD) 损失函数

$$L_{\text{RD}} = \mathbb{E}_{x \sim p_x} \left[ \sum_i \underbrace{-\eta_i \log p_{y_i}(y_i; \varphi^{(i)})}_{k_i} + \lambda d(\mathbf{x}, \hat{\mathbf{x}}) \right] \quad (7)$$

其中， $k_i$  表示语义特征矢量  $\mathbf{y}_i$  经过信源信道编码后对应的符号矢量  $\mathbf{s}_i$  的维度，可视为  $\mathbf{s}_i$  通过信道传输所占用的带宽； $\eta_i$  表示从  $\mathbf{y}_i$  的熵到信道传输符号数量  $k_i$  的比例放缩系数；当无特殊兴趣区域时，全局范围  $\eta_i$  设置为相同值  $\eta_i = \eta, \forall i$ 。超参数  $\lambda$  控制总传输带宽开销  $k = \sum_i k_i$  与失真  $d$  之间的折中， $\lambda$  越大，

优化得到的 NTSCC 模型信道带宽开销  $k$  越大，对

应的端到端传输失真  $d$  越小，反之亦然； $d(\mathbf{x}, \hat{\mathbf{x}})$  表示信源样本  $\mathbf{x}$  与重构样本  $\hat{\mathbf{x}}$  之间的误差度量，一般表示客观误差度量，例如，在图像传输问题中，取 MSE 表示发送端信源图像  $\mathbf{x}$  与接收端重构图像  $\hat{\mathbf{x}}$  对比的峰值信噪比 (PSNR, peak signal to noise ratio)； $d$  也可以度量信源图像  $\mathbf{x}$  与重构图像  $\hat{\mathbf{x}}$  之间的多尺度结构相似性 (MS-SSIM, multi-scale structural similarity) [23]。图 5 给出了不同  $\lambda$  取值下传输图像带宽分配结果，从图 5 可以看出，带宽分配与图像内容复杂度有显著关系。

### 2.2.3 语义感知增强

对于语义通信传输任务来说，有时更期待提高接收端重构图像  $\hat{\mathbf{x}}$  的全局主观感知质量，此时，可以在式(8)的 RD 损失函数中加入感知失真评价项。以图像传输为例，除了文献[24]中给出的感知质量优化方法，也需要在系统架构中引入条件生成对抗网络 (CGAN, conditional generative adversarial network) [25] 模块，即对重构图像  $\hat{\mathbf{x}}$  引入鉴别器  $D$ ，进一步增强图像主观感知质量。此时，端到端优化的损失函数为

$$L_{\text{RD}} = \mathbb{E}_{x \sim p_x} \left[ \underbrace{\sum_i -\eta_i \log p_{y_i}(y_i; \varphi^{(i)})}_{\text{信道传输速率}} + \underbrace{\beta_M \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}_{\text{客观 MSE 度量}} + \underbrace{\beta_L d_{\text{LPIPS}}(\mathbf{x}, \hat{\mathbf{x}})}_{\text{主观 LPIPS 度量}} + \underbrace{\beta_D (-\log D(\hat{\mathbf{x}}|\hat{\mathbf{y}}))}_{\text{CGAN 鉴别器打分度量}} \right] \quad (8)$$

其中，构成失真评价项的成分除了客观 MSE 度量外，还有基于 DNN 建模的学习感知图像块相似度 (LPIPS) [24] 失真度量，以及 CGAN 鉴别器  $D$  给出的打分度量， $(\beta_M, \beta_L, \beta_D)$  控制 3 种失真度量之间的比重权衡。

### 2.2.4 模块实现

针对上述 NTSCC 系统中几个重要的变换及编

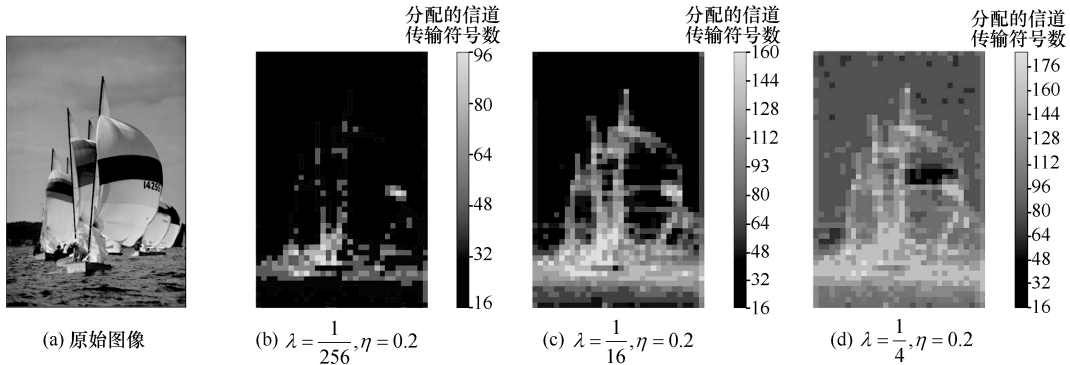


图 5 传输图像带宽分配结果

码模块, 本节给出几类典型的实现方式。以下设计的 DNN 实现结构均针对图像信源, 其他类型信源需针对性地再设计。

1) 变换模块  $g_a$  与  $g_s$ 。如图 6 所示, 2 个变换模块均采用 ViT (Vision Transformer) 结构<sup>[26]</sup>来实现, 对于一张 RGB 图像  $x \in \mathbb{R}^{h \times w \times 3}$ , 其首先会被分为  $l_1 = \frac{h}{2} \times \frac{w}{2}$  个互不重叠的像素块, 每个像素块的维度是  $2 \times 2 \times 3$ , 可以用一个块嵌入矢量表征。从 ViT 结构的输入来看, 嵌入矢量被视为 token, 将这些 token 对应的像素块位置按照从左到右下的顺序排序, 得到 token 序列  $(x_1, \dots, x_{l_1})$ 。经过一个全连接 (FC) 层线性变换后, 每个像素块 token 变为一个  $c$  通道的嵌入矢量。

接着,  $N_1$  层 Transformer 块将作用于这  $l_1$  个嵌入矢量的 token<sup>[26]</sup>。如图 6 所示, 每个 Transformer 块是一个序列输入序列输出过程, 它由一个多头自注意力 (MHSA, multi-head self-attention) 层、一个具有跳跃连接和层归一化 (LayerNorm, layer normalization) 的前馈层组成<sup>[27-28]</sup>。Transformer 块各层的操作为

$$O_1 = X + \text{MHSA}(X) \quad (9)$$

2) 联合信源信道编码模块  $f_e$  和  $f_d$ 。非线性编码  $f_e$  与  $f_d$  的 DNN 架构如图 7 所示,  $f_e$  根据熵模型  $p_{y_i}(\cdot; \varphi^{(i)})$  以可变速率传输语义隐表征矢量  $y_i$ 。特别地,  $f_e$  首先将  $y$  划分为块嵌入矢量序列  $(y_1, y_2, \dots, y_{l_1})$ , 每个  $y_i$  都是一个  $c$  维矢量。学习到

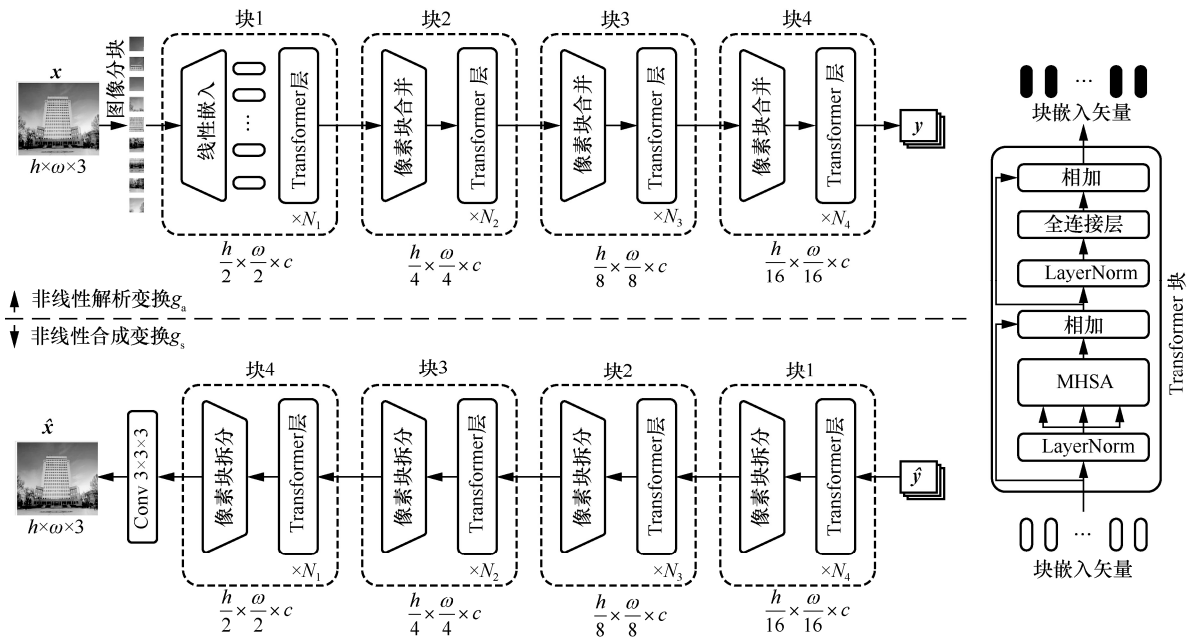


图 6 非线性变换  $g_a$  与  $g_s$  的 DNN 架构

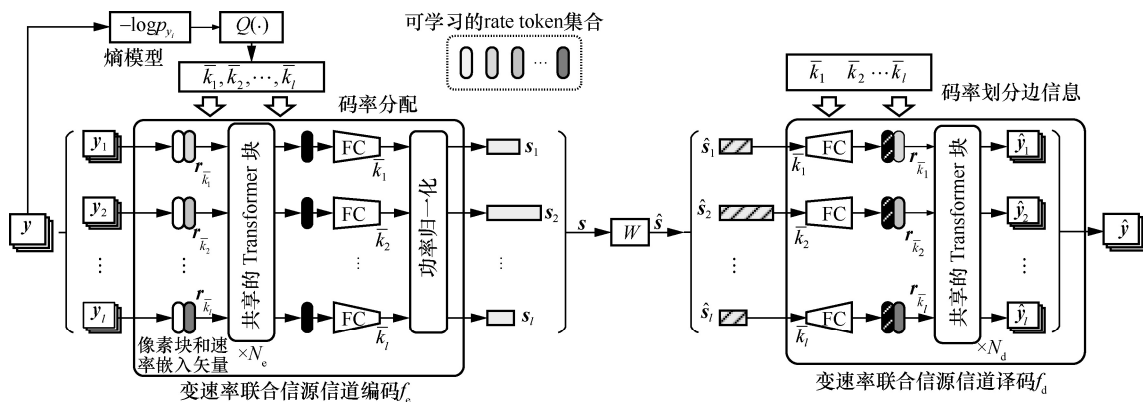


图 7 非线性编码  $f_e$  与  $f_d$  的 DNN 架构

的熵模型  $-\log p_{y_i}(\mathbf{y}_i)$  是  $\mathbf{y}_i$  的  $c$  维上的熵之和, 这样便得到了语义空间  $\mathbf{y}$  的信息密度分布。因此, 用于发送  $\mathbf{y}_i$  的带宽成本, 如 OFDM 子载波数量  $\bar{k}_i$  可以确定为

$$\bar{k}_i = Q(k_i) = Q(-\eta_i \log p_{y_i}(\mathbf{y}_i; \boldsymbol{\phi}^{(i)})) \quad (10)$$

其中, 先验熵模型  $p_{y_i}(\mathbf{y}_i; \boldsymbol{\phi}^{(i)})$  通过式(3)可得,  $Q$  表示标量量化, 其包括  $2^q$  ( $q=1, 2, \dots$ ) 个整数, 并且量化值集合为  $\mathcal{V} = \{v_1, v_2, \dots, v_{2^q}\}$ 。因此, 预定的  $q$  bit 应当作为额外的边信息被发送, 以通知接收端每个语义隐表征矢量  $\mathbf{y}_i$  分配多少带宽。

为了灵活地将  $\mathbf{y}_i$  编码到一个  $\bar{k}_i$  维的信道输入矢量  $\mathbf{s}_i$ , 一种直观的方案是将解码器  $f_c$  划分为  $l$  个部分, 每个部分负责编码一个嵌入矢量  $\mathbf{y}_i$ 。为了实现这一点, 需要训练  $2^q$  对子编码器和子译码器。这种方式将导致巨大的训练复杂性和存储成本。并且, 它忽略了  $\mathbf{y}_i$  之间的上下文相关性, 导致系统性能折损。因此, 本文将动态神经网络结构<sup>[29]</sup>引入 Transformer 块中, 以实现 JSCC 编码器  $f_e$ 。与固定计算图和参数的静态网络模型相比, 动态神经网络可以使网络结构或参数适应不同的输入, 从而在性能、计算效率等方面带来显著优势。这里, 本文使用  $N_c$  组共享 Transformer 块进行特征提取, 并使用 FC 层将  $\mathbf{y}_i$  编码到  $\bar{k}_i$  维矢量。该过程包含以下 2 个步骤。

① 通过自注意力机制将每个  $\mathbf{y}_i$  的信道带宽成本  $\bar{k}_i$  指示给共享 Transformer 块。

② 使用 FC 层将对应于  $\mathbf{y}_i$  的 Transformer 块的输出转化为  $\bar{k}_i$  维矢量。

受 Transformer 位置编码思路的启发, 本文设计了可学习的 rate token 集合  $\mathcal{R} = \{\mathbf{r}_{v_1}, \mathbf{r}_{v_2}, \dots, \mathbf{r}_{v_{2^q}}\}$  表示速率信息。如图 7 所示, 每一个  $\mathbf{y}_i$  在送入 Transformer 块之前将与其相应的 rate token 相加。共享 Transformer 块的输出可以学习以适应  $\mathbf{y}_i$  的熵, 并且使随后的 FC 层可以高效地调整输出维度。在接收端, 不同长度的信道接收矢量  $\hat{\mathbf{s}}_i$  先经由 FC 层重塑为同一维度, 然后分别添加 rate token, 再发送到  $N_c$  组共享 Transformer 块中, 得到重构的语义隐空间表征矢量  $\hat{\mathbf{y}}_i$ 。

### 2.2.5 语义非线性变换编码实现过程总结

前文以图像信源为例介绍了 NTSCC 的模块细节, NTSCC 框架也可以灵活应用在语音<sup>[30]</sup>、视频<sup>[31]</sup>

等不同信源传输任务中, 具有语义编码传输的普适性。本质上, 该框架包含 2 个模块: 语义特征提取(变换)模块和语义特征编码(联合信源信道编码)模块。变换模块通过解除信源各维度之间的相关性, 有效提取语义表征。针对不同种信源和不同传输任务, 变换模块可以使用不同的网络模型部署。提取到信源的语义特征之后, 联合信源信息编码模块将赋予其信道语义信息, 并将两者编码为适合信道传输的符号。本文通过变换模块中得到的语义熵模型, 以端到端优化的方式得到变速率联合信源信道编码器, 实现语义特征的编码传输。

## 3 实验结果与分析

### 3.1 实验配置

为了验证所提语义非线性变换编码传输系统的端到端传输性能, 本节将比较 NTSCC 系统与前文提到的直接编码传输 Deep JSCC 系统和经典信源信道编码传输系统的性能。为了量化所提系统对高维复杂信源的编码传输能力, 实验使用 RGB 图像数据进行无线传输。实验的具体配置如下。

1) 数据集。NTSCC 的训练集是从 Open Images 数据集<sup>[32]</sup>中随机采样的 500 000 张图像, 训练过程中, 这些图像被随机裁剪为 256 像素  $\times$  256 像素。为了量化端到端图像传输能力, 实验使用分辨率不同、内容各异的图像数据集进行测试, 图像尺寸从小到大的测试集依次为: CIFAR10 (50 000 张训练图像和 10 000 张测试图像, 32 像素  $\times$  32 像素)、Kodak (24 张图像, 768 像素  $\times$  512 像素)、CLIC2021 (60 张图像, 2 048 像素  $\times$  1 890 像素)。

2) 对比方案。端到端的图像传输包括信源信道编码和无线信道传输 2 个主要模块。因此, 对比方案包括所提的 Deep JSCC 方案和经典分离式信源信道编码方案。分离式信源信道编码方案采用了压缩能力依次增强的图像信源编码方案 JPEG<sup>[33]</sup>、JPEG2000<sup>[34]</sup>、BPG (H.265 视频编码标准的帧内图像编码方案), 并结合实际应用的 5G 标准 LDPC 信道编码(分别记为 JPEG + LDPC、JPEG2000 + LDPC 和 BPG + LDPC)。本文考虑 AWGN 信道和衰落信道下测试性能, 在实际部署中, 为了与先前工作一致, 将信道输入序列中的 2 个连续实符号转换为一个复信道输入符号, 并添加复高斯噪声。

3) 度量指标和损失函数。实验使用广泛应用的

像素级度量指标（如 PSNR 和 MS-SSIM<sup>[22]</sup>）和最近兴起的基于深度学习的感知度量指标（如 LPIPS<sup>[23]</sup>）对所提出的 NTSCC 模型和其他端到端传输模型进行性能评估。PSNR 对应于像素级的 L2 欧氏距离，因此当评估模型在 PSNR 上的表现时，将失真  $d$  设置为信源图像  $x$  和重构图像  $\hat{x}$  间的均方误差 MSE。当评估 MS-SSIM 指标时，失真  $d$  被设置为  $1 - MS-SSIM$  以最小化。更高的 PSNR/MS-SSIM 指数意味着更好的传输表现。然而，即便 PSNR 和 MS-SSIM 被广泛用作经典图像质量评价指标，它们仍旧是简单且固定的函数，难以反映人类感知的诸多细微差别。为了更贴近语义通信目标，本文进一步采用了基于深度学习的 LPIPS 指标作为量化图像传输效果的语义感知损失。LPIPS 取值范围为 0~1，LPIPS 值越小表示损失越少。

4) 模型训练细节。在训练阶段，使用学习率为  $10^{-4}$  的 Adam 优化器，mini-batch 的大小设置为 10。所有模型训练与部署均以 Pytorch 为架构实现，模型使用单块 RTX 3090 GPU 训练。

### 3.2 性能分析

#### 3.2.1 PSNR 性能

图 8 展示了 SNR=10 dB 的 AWGN 信道上 PSNR 与 CBR 的关系。对于经典分离式方案，JPEG + LDPC、JPEG2000 + LDPC、BPG + LDPC 均采用 2/3 速率(4 069, 6 144)的 LDPC 码和 16 进制正交幅度调制 (QAM)。结果表明，NTSCC 在各个 CBR 上大多优于 Deep JSCC，并且两者的性能差距会随着图像分辨率和 CBR 的增大而逐渐扩大。与分离式方案相比，NTSCC 在 3 个数据集上均超过了实际中的 JPEG+LDPC、JPEG2000 + LDPC 和 BPG + LDPC 方案。

图 9(a)和图 9(b)展示了 PSNR 与 AWGN 信道不同 SNR 的关系，CBR 设置为  $\rho = \frac{k}{m} = \frac{1}{16}$ 。对于分离式方案，通过评估 LDPC 码率和调制的不同组合的性能，得到了在每个 SNR 下最佳性能配置的包络。由于 NTSCC 方法学习到了一种速率匹配机制，因而可以通过微调模型训练超参数，确保其最大信道带宽比低于  $\frac{1}{16}$ ，以达到公平的比较。结果表明，NTSCC 系统表现出显著性能提升，相较于 Deep JSCC、JPEG+LDPC 和 JPEG2000+LDPC 至少提升 1 dB。此外，如图 9(c)所示，当 SNR 逐渐降低时，NTSCC 表现出和 Deep JSCC 同样平滑的性能下降，然而基于

分离式的 BPG+LDPC 传输方案的性能出现陡降（被称为“悬崖效应”），这是因为信道译码出现差错导致信源译码出现明显差错传播效应。

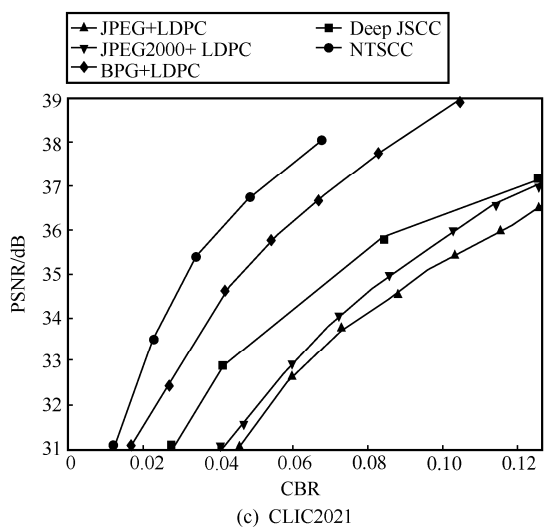
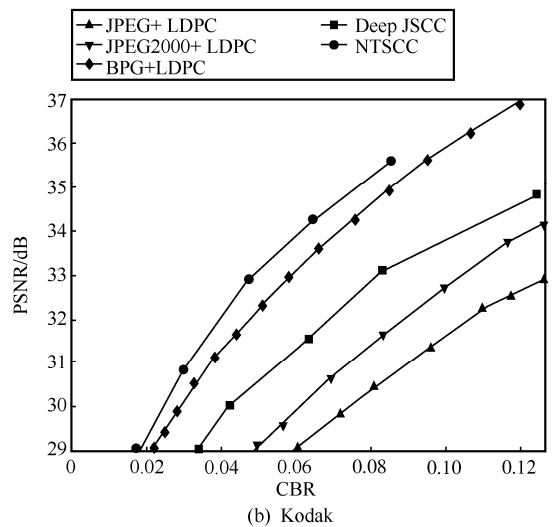
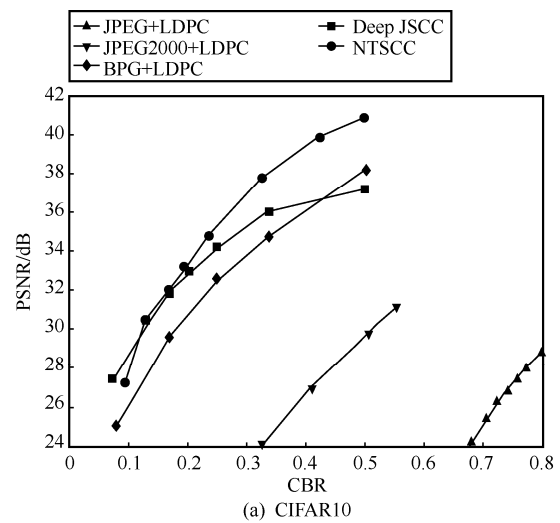
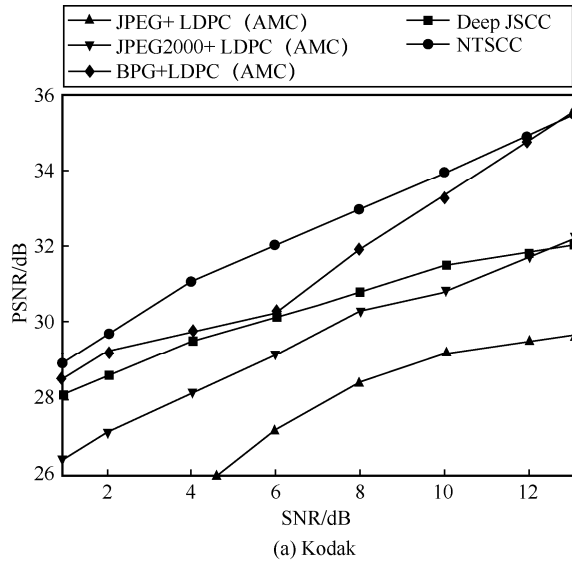
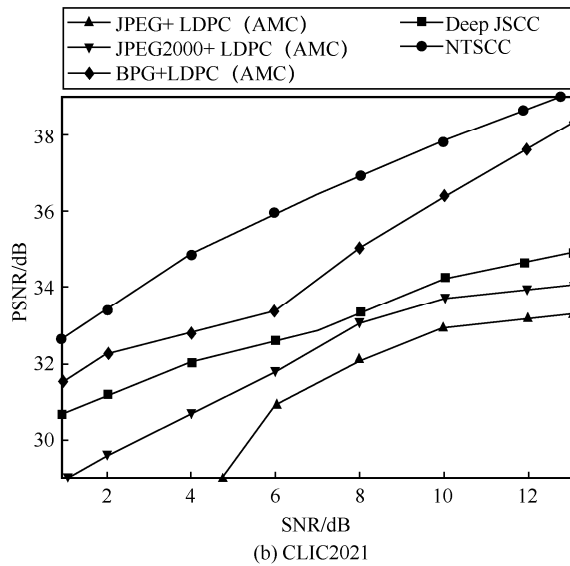


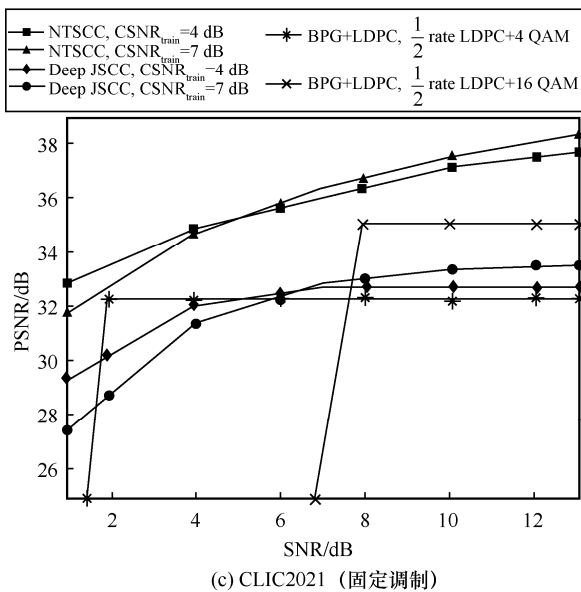
图 8 SNR=10 dB 的 AWGN 信道上 PSNR 与 CBR 的关系



(a) Kodak



(b) CLIC2021



(c) CLIC2021 (固定调制)

图 9 PSNR 与 AWGN 信道不同 SNR 的关系

### 3.2.2 MS-SSIM 性能

图 10(a)和图 10(b)展示了 SNR=10 dB 的 AWGN 信道上 MS-SSIM 与 CBR 的关系;图 10(c)和图 10(d)展示了 MS-SSIM 与 AWGN 信道不同 SNR 的关系, CBR 设置为  $\rho = \frac{k}{m} = \frac{1}{16}$ 。由于 MS-SSIM 的值介于 0 (最差) 和 1 (最好) 之间, 并且绝大多数值高于 0.9, 本文将 MS-SSIM 值的单位转换为 dB 以提高易读性。结果表明, NTSCC 很大程度上优于其他方案, 并且在高 CBR 区域具有更高的性能增益。与 PSNR 指标下的结果相比, 易发现 BPG + LDPC 的结果普遍差于深度学习驱动的语义通信方案。

### 3.2.3 LPIPS 性能

除了上述 PSNR 和 MS-SSIM 失真度量外, 面向语义通信目标, 本文进一步使用以人类视觉感知为导向的 LPIPS 损失函数训练 NTSCC 模型, LPIPS 度量是反映人类感知体验的主流指标。此外, 如前文所述, 当 NTSCC 模型以人类感知体验度量为目标时, 模型训练的损失函数参考式 (8)。图 11 展示了 SNR=10 dB 的 AWGN 信道上 LPIPS 与 CBR 的关系, LPIPS 值越小表示失真越小。对于端到端语义通信传输方案 (Deep JSCC 和 NTSCC), 图 11 在括号中标记了 PSNR 和 MS-SSIM 以指示模型的训练目标。NTSCC (Perceptual) 曲线表示如式(8)所示的 RDP 感知损失函数来优化模型。显然, 感知优化的 NTSCC 在性能上远优于其他方案。

### 3.2.4 衰落信道性能

图 12 展示了实际衰落信道条件下, BPG+LDPC 方案与 NTSCC 方案传输性能的对比。其中, 衰落信道采样自 COST2100 信道模型<sup>[35]</sup>, 衰落系数会随着帧数而变化。图 12(a)展示了衰落信道 SNR 变化曲线。图 12(b)展示了 BPG+LDPC 编码调制方案以及对应的 QP 值。图 12(c)展示了衰落信道传输表现, 表示在 CBR = 0.04 的约束条件下, 2 个方案每帧的 PSNR 值。实验结果表明, 即使配置了自适应调制编码 (AMC) 机制, 由于悬崖效应, 具有有限数量量化参数的信源信道分离式调制方案也难以灵敏应对 SNR 变化的衰落信道。相比之下, NTSCC 框架具有更好的灵活性和更高的编码增益。

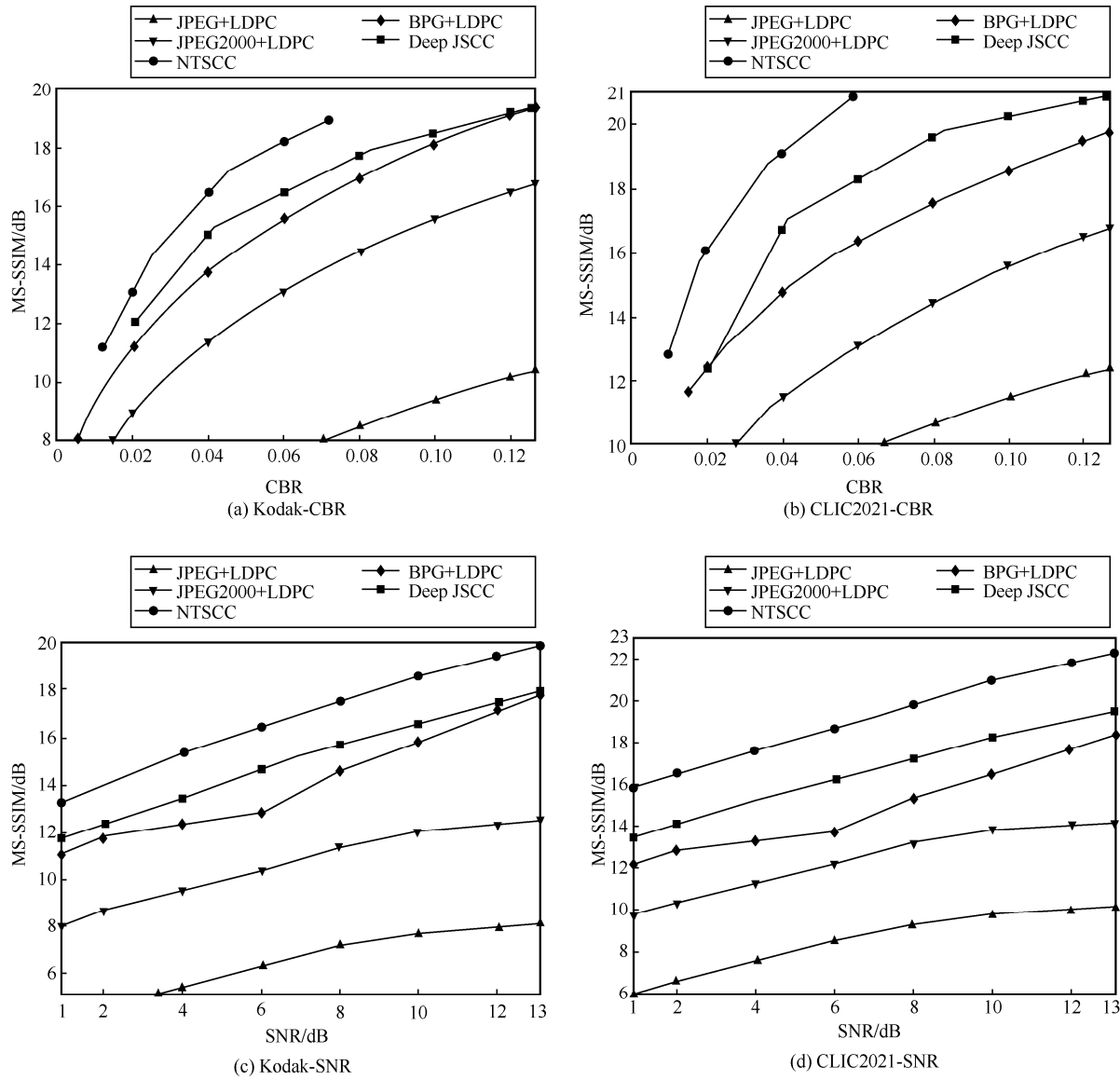


图 10 MS-SSIM 与 CBR 和 SNR 的关系

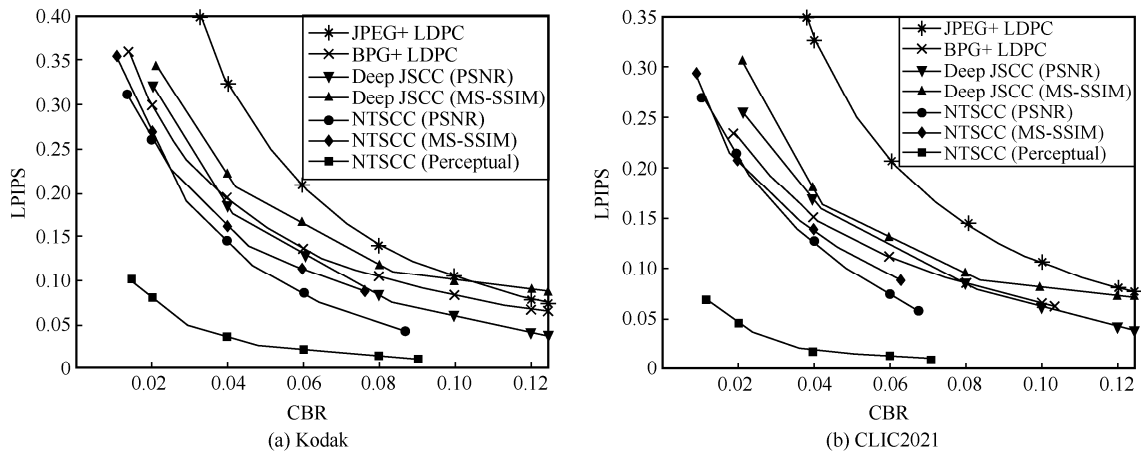


图 11 SNR = 10 dB 的 AWGN 信道上 LPIPS 与 CBR 的关系

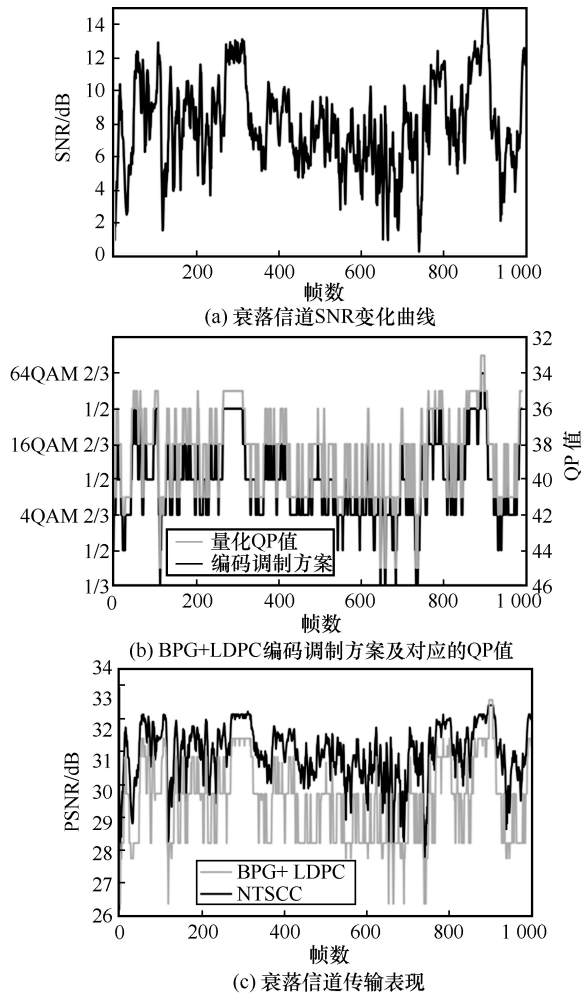


图12 实际衰落信道条件下, BPG+LDPC方案与NTSCC方案传输性能的对比如

## 4 结束语

本文提出了面向语义通信的端到端非线性变换编码传输新框架。首先, 基于变分理论推导出了语义通信端到端率失真优化准则。据此, 设计了非线性变换来提取信源数据在语义隐空间的紧致表征, 并通过语义变分熵建模引导实现了变速率联合信源信道编码。实验表明, 语义非线性变换编码能显著提升端到端数据传输性能及鲁棒性, 是实现语义通信的关键技术之一。

本文提出的语义编码传输框架目前主要面向高保真/优人类感知体验的端到端数据传输设计, 未来可进一步扩展到机器类智能任务主导的端到端语义通信场景, 具有广阔的研究前景。

## 参考文献:

[1] WARREN W. Recent contributions to the mathematical theory of communication[J]. ETC: A Review of General Semantics, 2017,

74(1/2): 136-157.

[2] SHANNON C E. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948, 27(3): 379-423.

[3] ZHANG P, XU W J, GAO H, et al. Toward wisdom-evolutionary and primitive-concise 6G: a new paradigm of semantic communication networks[J]. Engineering, 2022, 8: 60-73.

[4] QIN Z J, TAO X M, LU J H, et al. Semantic communications: principles and challenges[J]. arXiv Preprint, arXiv: 2201.01389, 2022.

[5] 石光明, 李莹玉, 谢雪梅. 语义通讯: 智能时代的产物[J]. 模式识别与人工智能, 2018, 31(1): 91-99.

SHI G M, LI Y Y, XIE X M. Semantic communications: outcome of the intelligence era[J]. Pattern Recognition and Artificial Intelligence, 2018, 31(1): 91-99.

[6] 石光明, 肖泳, 李莹玉, 等. 面向万物智联的语义通信网络[J]. 物联网学报, 2021, 5(2): 26-36.

SHI G M, XIAO Y, LI Y Y, et al. Semantic communication networking for the intelligence of everything[J]. Chinese Journal on Internet of Things, 2021, 5(2): 26-36.

[7] NIU K, DAI J C, YAO S S, et al. A paradigm shift toward semantic communications[J]. IEEE Communications Magazine, 2022, 60(11): 113-119.

[8] 牛凯, 戴金晟, 张平, 等. 面向6G的语义通信[J]. 移动通信, 2021, 45(4): 85-90.

NIU K, DAI J C, ZHANG P, et al. 6G-oriented semantic communications[J]. Mobile Communications, 2021, 45(4): 85-90.

[9] DAI J C, ZHANG P, NIU K, et al. Communication beyond transmitting bits: semantics-guided source and channel coding[J]. IEEE Wireless Communications, 2022: doi.org/10.1109/MWC.017.2100705.

[10] 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.

[11] BALLÉ J, CHOU P A, MINNEN D, et al. Nonlinear transform coding[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 15(2): 339-353.

[12] RICHARDSON T, KUDEKAR S. Design of low-density parity check codes for 5G new radio[J]. IEEE Communications Magazine, 2018, 56(3): 28-34.

[13] ARIKAN E. Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels[J]. IEEE Transactions on Information Theory, 2009, 55(7): 3051-3073.

[14] FARSAID 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). Piscataway: IEEE Press, 2018: 2326-2330.

[15] CHOI K, TATWAWADI K, GROVER A, et al. Neural joint source-channel coding[C]//International Conference on Machine Learning. New York: PMLR, 2019: 1182-1192.

[16] BOURTSOULATZE E, KURKA D B, GÜNDÜZ D. Deep joint source-channel coding for wireless image transmission[C]//Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2019: 4774-4778.

[17] GÜNDÜZ D, KURKA D B, JANKOWSKI M, et al. Communicate to learn at the edge[J]. IEEE Communications Magazine, 2020, 58(12): 14-19.

- [18] KURKA D B, GÜNDÜZ D. Bandwidth-agile image transmission with deep joint source-channel coding[J]. IEEE Transactions on Wireless Communications, 2021, 20(12): 8081-8095.
- [19] XU J L, 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.
- [20] WANG S X, YANG K, DAI J C, et al. Distributed image transmission using deep joint source-channel coding[C]//Proceedings of 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2022: 5208-5212.
- [21] WANG J, WANG S X, DAI J C, et al. Perceptual learned source-channel coding for high-fidelity image semantic transmission[J]. arXiv Preprint, arXiv: 2205.13120, 2022.
- [22] DAI J C, WANG S X, TAN K L, et al. Nonlinear transform source-channel coding for semantic communications[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(8): 2300-2316.
- [23] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment[C]//Proceedings of the Thirty Seventh Asilomar Conference on Signals, Systems & Computers. Piscataway: IEEE Press, 2004: 1398-1402.
- [24] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 586-595.
- [25] MIRZA M, OSINDERO S. Conditional generative adversarial nets[J]. arXiv Preprint, arXiv: 1411.1784. 2014.
- [26] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[J]. arXiv Preprint, arXiv: 2010.11929. 2020.
- [27] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2022: 9992-10002.
- [28] BA J L, KIROS J R, HINTON G E. Layer normalization[J]. arXiv Preprint, arXiv: 1607.06450. 2016.
- [29] HAN Y Z, HUANG G, SONG S J, et al. Dynamic neural networks: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(11): 7436-7456.
- [30] XIAO Z X, YAO S S, DAI J C, et al. Wireless deep speech semantic transmission[J]. arXiv Preprint, arXiv: 2211.02283. 2022.
- [31] WANG S X, DAI J C, LIANG Z J, et al. Wireless deep video semantic transmission[J]. IEEE Journal on Selected Areas in Communications, 2023, 41(1): 214-229.
- [32] KUZNETSOVA A, ROM H, ALLDRIN N, et al. The open images dataset v4[J]. International Journal of Computer Vision, 2020, 128(7): 1956-1981.
- [33] WALLACE G K. The JPEG still picture compression standard[J]. IEEE Transactions on Consumer Electronics, 1992, 38(1): 30-44.
- [34] MARCELLIN M W, GORMISH M J, BILGIN A, et al. An overview of JPEG2000[C]//Proceedings of Data Compression Conference. Piscataway: IEEE Press, 2002: 523-541.
- [35] LIU L F, OESTGES C, POUTANEN J, et al. The COST 2100 MIMO channel model[J]. IEEE Wireless Communications, 2012, 19(6): 92-99.

## [作者简介]



张平(1959-),男,陕西汉中,人,博士,中国工程院院士,北京邮电大学教授、博士生导师,主要研究方向为先进移动通信系统。



戴金晟(1992-),男,湖南常德人,博士,北京邮电大学副教授、博士生导师,主要研究方向为语义通信、信源信道编码、智能信号处理等。



张育铭(1999-),男,河南南阳人,北京邮电大学硕士生,主要研究方向为语义通信等。



王思贤(1997-),男,安徽宿州人,北京邮电大学博士生,主要研究方向为语义通信、信源信道编码、计算机视觉等。



秦晓琦(1988-),女,北京人,博士,北京邮电大学副教授,主要研究方向为下一代无线网络、智能高效机器型通信等。



牛凯(1976-),男,山西文水人,博士,北京邮电大学教授,主要研究方向为信息论与极化码、智能信号处理等。