

会议名称	IMT-2030(6G)语义通信任务组第三次会议	会议地点	北京
提交单位	北京邮电大学	会议时间	2024 年 3 月 30 日

面向图像信源传输的非线性变换信源信道编码 (NTSCC) 提案

作者: 梁子鉴, 王昌硕, 戴金晟, 王思贤, 谭凯林, 司中威, 秦晓琦, 牛凯, 张平

1. 背景

近年来, 现代无线通信技术的发展正不可避免的陷入瓶颈。一方面, 以 LDPC 码、极化码为代表的信道编码技术已经可以逼近香农信道容量极限, 与此同时, 以 BPG、H.265/HEVC 为代表的限失真信源编码技术也逐渐逼近压缩极限, 只能依靠大量复杂度堆叠以及更多的能源消耗来获得进一步的性能提升, 导致难以可持续发展[1]。另一方面, 随着智能化应用场景和各种下游任务的涌现, 以传输符号准确度为目标的经典通信系统无法高效适配这些新型场景, 亟待新型通信技术的应用[2]。这些问题都迫使人们考虑更高层次的解决方法, 即从语义通信的角度设计编码传输方法。

目前, 语义通信技术的考虑大多从联合信源信道编码的角度出发, 通过联合考虑信源的语义特征以及信道的统计特性, 对编码传输方法进行优化设计, 实现端到端的性能提升。其中, 最为典型的方法是深度联合信源信道编码 (Deep Joint Source-Channel Coding, DJSCC) [3]方案。它基于深度神经网络, 利用直接编码的方法, 实现从信源符号到信道符号的直接映射。这种方法被广泛应用于图像[4]、视频[5]、文本[6]、语音[7]等多种模态信源的语义编码传输方法以及传输场景当中, 且均具有优异的性能表现。尽管这种方法相对于传统通信方法获得了有效的性能提升, 并且表现出了强大的鲁棒性, 但这种方法在较高的信源维度、较高的信道带宽比 (Channel Bandwidth Ratio, CBR) 以及较高的信噪比的条件下将无法获得性能优势。这是由于这种方法不能有效的根据信源的不同实现变长速率的编码与传输, 导致在传输过程中出现带宽的不必要浪费。

本提案提出了一种基于非线性变换编码的语义编码方案, 即非线性变换信源信道编码 (Nonlinear Transform Source-Channel Coding, NTSCC) [8]。其首先通过非线性解析变换, 将信源符号映射到语义特征向量空间, 利用熵模型实现针对于语义特征的速率估计, 并依照估计的速率信息, 利用联合信源信道编码实现可变速率传输。通过面向率失真形式的损失函数进行优化训练, 相比于 DJSCC 方案, 本提案所提语义通信方法可以大幅度提高语义传输的效率与可靠性, 并且获得相比于传统信源信道分离编码方案, 在不同信噪比、不同 CBR 的条件下均可获得显著的性能增益。

会议名称	IMT-2030(6G)语义通信任务组第三次会议	会议地点	北京
提交单位	北京邮电大学	会议时间	2024 年 3 月 30 日

2. 基本框架

2.1 NTC 的基本框架

本提案提出的 NTSCC 的基本框架建立在 Ballé 在 2018 年提出的 NTC 框架的基础上[9]。

NTC 的基本框架如图 1 所示：

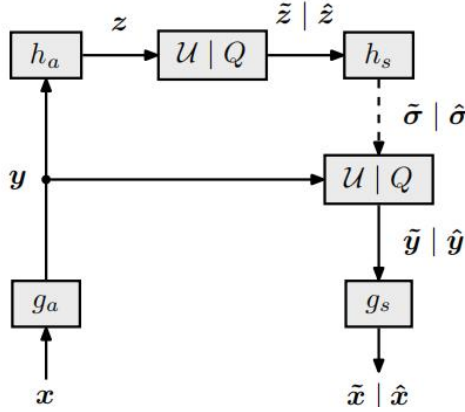


图 1 NTC 的基本框架

NTC 中采用了间接编码的方法，即引入了解析变换，将信源数据映射到隐空间，反映信源数据的语义信息。由于变换后的隐变量 y 并没有完全解除相关性，因此，引入了另一个解析变换模块以进一步解除相关性，得到隐变量 z ，也被称作超先验。利用超先验作为条件，可以为隐空间中特征向量的提供速率估计，以计算编码速率。其损失函数为：

$$L = \mathbb{E}_{x \sim p_x(x)} \left[\lambda \left(-\log p_{\tilde{y}|\tilde{z}}(\tilde{y}|\tilde{z}) - \log p_{\tilde{z}}(\tilde{z}) \right) + d(\mathbf{x}, \hat{\mathbf{x}}_{\text{NTC}}) \right] \quad (1)$$

其中 λ 为速率项与失真项之间的平衡因子。利用损失函数，可以令 NTC 模型面向率失真优化进行优化。实验表明，NTC 的性能具有与理论上最优的矢量量化相近的性能，且处理复杂度大幅度降低。本提案以 NTC 为基本框架，结合联合信源信道编码的思想，将其扩展到无线通信中，形成了 NTSCC 框架。

2.2 NTSCC 的基本框架

本提案提出的 NTSCC 框架采用了非线性变换编码的方法，即通过非线性解析变换将信源数据映射到隐空间，反映信源数据的语义信息，然后再进行信源信道联合编码，将语义信息传输到无线信道当中。

图 2 为 NTSCC 的基本结构示意图

会议名称	IMT-2030(6G)语义通信任务组第三次会议	会议地点	北京
提交单位	北京邮电大学	会议时间	2024 年 3 月 30 日

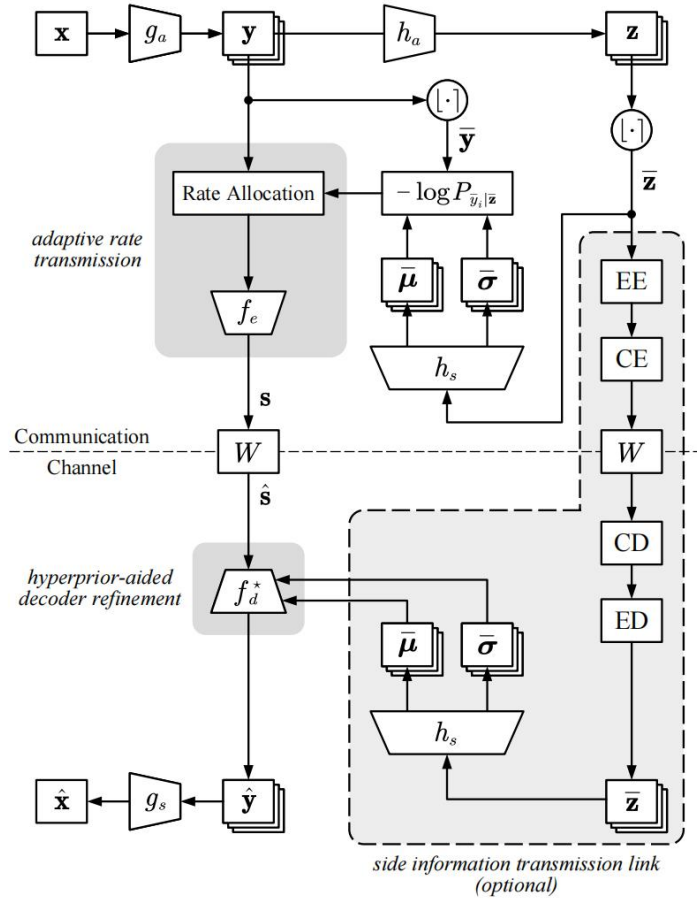


图 2 NTSCC 的基本结构示意图

总体来看，图 2 的上半部分为 NTSCC 的发射机，中间为信道，下半部分是 NTSCC 的接收机。信源为图像信源，传输之前进行分块操作，再送入发射机中。在发射机中，一方面，将 NTC 中的解析变换作为 DJSCC 编码前的一种“预编码”，提取源语义特征作为潜在表示。DJSCC 随后在这个潜在的空间上进行操作；另一方面，引入超先验解析变换，对潜在的语义变量进行进一步解相关，作为边信息，并通过超先验综合变换估计每个图像块的信息量，用于指导联合信源信道编码的输出维度，起到传输速率自适应的效果。接收机可以在边信息的指导下进行信源信道联合译码，再通过综合变换重构信源图像块，也可以无需边信息，直接进行联合信源信道编码之后再重构信源。

接下来对图 2 表示的 NTSCC 框架每个模块的作用进行详细介绍：

- 解析变换模块 g_a ：解析变换 g_a 的作用是利用非线性变换，将信源 \mathbf{x} 映射到语义特征空间中，并通过相关性的去除，解析获得隐式特征 \mathbf{y} ；
- 综合变换模块 g_s ：综合变换 g_s 的作用是利用对隐变量的估计 $\hat{\mathbf{y}}$ ，对信源数据进行重

会议名称	IMT-2030(6G)语义通信任务组第三次会议	会议地点	北京
提交单位	北京邮电大学	会议时间	2024 年 3 月 30 日

建 (综合), 输出重建后的信源 $\hat{\mathbf{x}}$;

- 超先验解析变换模块 h_a : 超先验解析变换 h_a 用于对隐变量 \mathbf{y} 进一步地进行解相关, 输出隐变量 \mathbf{z} , 其中 \mathbf{z} 也被称作超先验;
- 超先验综合变换模块 h_s : 超先验综合变换 h_s 以具有随机性的隐变量 $\tilde{\mathbf{z}}$ (或量化后的隐变量) 为条件, 为隐式特征 \mathbf{y} 输出相应的均值序列和方差序列, 以建立混合高斯分布, 并在 $y_i - \frac{1}{2}$ 到 $y_i + \frac{1}{2}$ 的范围内计算概率值。
- 熵模型: 经过超先验解析变换模块 h_a 与超先验综合变换模块 h_s 的处理后, 熵模型利用 $-\log P(\mathbf{y}_i | \tilde{\mathbf{z}})$ 计算每个图像块的信息量, 作为联合信源信道编码的指导, 可以对其传输速率进行控制;
- 可变速率 DJSCC 模块: DJSCC 编码器 f_e 以隐变量 \mathbf{y} 为输入, 熵模型计算出的 $-\log P(\mathbf{y}_i | \tilde{\mathbf{z}})$ 为指导, 为每个图像块进行不同输出维度的降维变换, 输出变长的序列 \mathbf{s} ; DJSCC 解码器 f_d 以变长传输序列 $\hat{\mathbf{s}}$, 为每个图像块进行重建, 并拼接组合成隐变量的估计值 $\hat{\mathbf{y}}$;
- 信道编码模块 CE 和熵编码模块 EE (可选): 用于对边信息 \mathbf{z} 进行信源和信道编码;
- 信道译码模块 CD 和熵译码模块 ED (可选): 用于对边信息 \mathbf{z} 进行信源和信道译码。

需要说明的是, 经过量化后的超先验 $\tilde{\mathbf{z}}$, 既可以传输到接收端, 作为编解码器共享的边信息, 辅助进行信源的重建, 也可以不将其传输到接收端, 仅在得知每个图像块的信道符号长度的条件下, 恢复隐式特征和信源信息。若传输 $\tilde{\mathbf{z}}$ 时, 接收端可以联合地利用接收的符号 $\hat{\mathbf{s}}$, 与 $\tilde{\mathbf{z}}$ 经过超先验综合变换 h_s 后得到的均值 $\bar{\boldsymbol{\mu}}$ 和方差 $\bar{\boldsymbol{\sigma}}$ 作为联合信源信道解码器 f_d 的输入, 以获得更好的重建性能。在这种情况下, 为了确保边信息的可靠传输, 信道编码 (CE 和 CD) 应采用先进的接近信道容量的编码, 如低密度校验码 (LDPC) 码[9]或极化码[10]等。相应地, 不传输 $\tilde{\mathbf{z}}$ 可以一定程度的节省信道带宽成本, 但解码质量会有所下降。

2.3 NTSCC 的模型结构设计

接下来对本提案中 NTSCC 框架里的各个模块的实现方案进行介绍。

NTSCC 的解析变换与综合变换的网络结构原则上可以选用任何一种与信源数据类型相符的网络结构, 本提案针对的信源数据类型是图像, 因此选用了图像数据处理能力较强

会议名称	IMT-2030(6G)语义通信任务组第三次会议	会议地点	北京
提交单位	北京邮电大学	会议时间	2024 年 3 月 30 日

的 Swin-Transformer 结构[11]。其输入为一系列图形块对应的嵌入序列，即先进行图像分块，把图像分成一系列的 2×2 的小块，再对每个图像块的像素进行高维度嵌入变换，每个图像块对应输出一个对应的矢量序列。通过多层 Swin-Transformer 结构，对输入图像进行解相关变换。如图 3 所示。

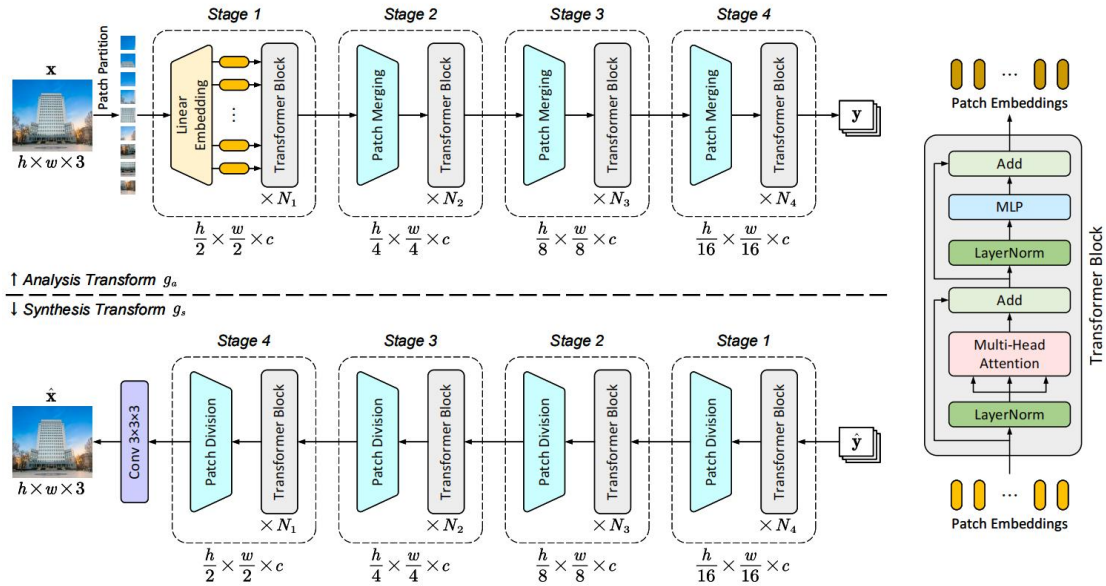


图 3 解析变换与综合变换的网络结构设计

熵模型中的超先验解析变换和综合变换模块采用卷积神经网络结构进行实现，如图 4 所示，是由一系列的卷积层和 ReLU 激活函数组成的。

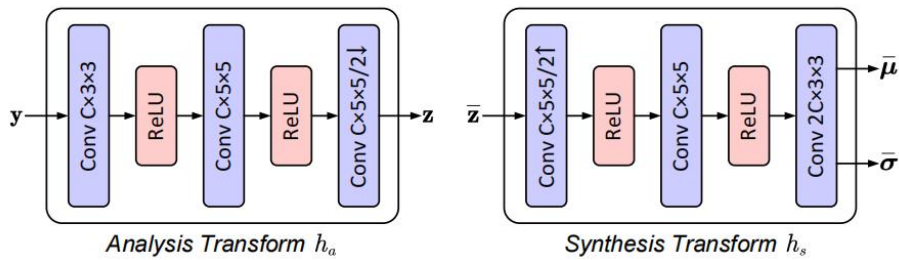


图 4 超先验解析变换与超先验综合变换的网络结构设计

联合信源信道编译码模块的模型结构如图 5 所示。为了实现对隐变量的每一个块进行不同维度的降维变换，在编码器内部设计了一系列相同维度输入（维度等于隐变量的通道数）、不同维度输出的全连接层组，用于进行降维，并在解码器中设计了结构与之对称的全连接层组。在实际使用时，依据熵模型导出的每个块的信息量以及当前的信道条件，获

会议名称	IMT-2030(6G)语义通信任务组第三次会议	会议地点	北京
提交单位	北京邮电大学	会议时间	2024 年 3 月 30 日

得在编码器端的每个块所选择的全连接层的输出维度，并利用对应的全连接层进行降维处理，并将获得的变速率序列送入信道中。在接收端获得序列后，就可以利用每个块相应的维度进行数据重建。此外，为了实现模型对不同速率输出的适应能力，在进行变速率降维与变速率重建的前后，分别采用一个速率感知的 Transformer 结构进行自适应优化。经过变速率 JSCC 结构的变换，就可以将隐变量 \mathbf{y} 在熵模型的指导下进行变速率传输，实现更加高效的语义信息传输。

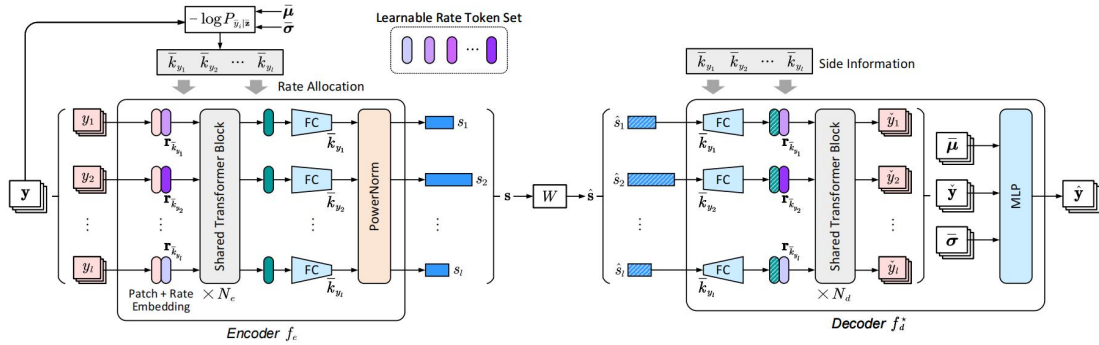


图 5 DJSCC 网络结构设计

3 理论分析

NTSCC 的优化问题也可以表述为如图 6 所示的变分模型。

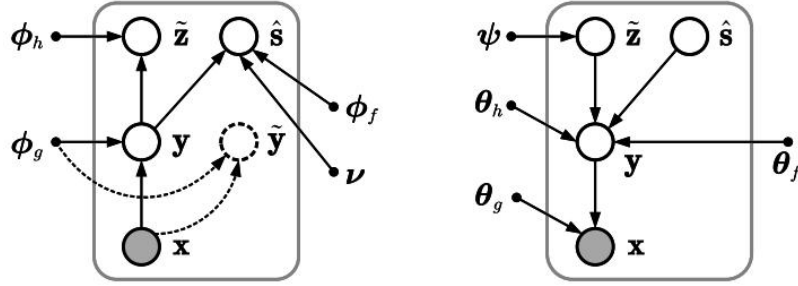


图 6 NTSCC 的变分模型

概率生成模型代表 JSCC 解码器和综合变换，近似推理模型对应于解析变换和 JSCC 编码器。推理模型的目标是构造一个参数化的概率密度 $q_{\hat{s}, \tilde{z}|\mathbf{x}}$ ，来近似真实的后验概率分布 $p_{\hat{s}, \tilde{z}|\mathbf{x}}$ 。为了使两者尽可能接近，需要最小化两个概率分布之间的 KL 散度：

$$\min_{\phi_g, \phi_h, \phi_f, \theta_g, \theta_h, \theta_f} \mathbb{E}_{\mathbf{x} \sim p_{\mathbf{x}}} D_{\text{KL}} [q_{\hat{s}, \tilde{z}|\mathbf{x}} \| p_{\hat{s}, \tilde{z}|\mathbf{x}}] \quad (2)$$

经过化简，最小化 KL 散度等价于最小化公式

会议名称	IMT-2030(6G)语义通信任务组第三次会议	会议地点	北京
提交单位	北京邮电大学	会议时间	2024 年 3 月 30 日

$$\min_{\phi_g, \phi_h, \phi_f, \theta_g, \theta_h, \theta_f} \mathbb{E}_{\mathbf{x} \sim p_{\mathbf{x}}} \mathbb{E}_{\hat{\mathbf{s}}, \tilde{\mathbf{z}} \sim q_{\hat{\mathbf{s}}, \tilde{\mathbf{z}}|\mathbf{x}}} \left[\underbrace{\log q_{\hat{\mathbf{s}}, \tilde{\mathbf{z}}|\mathbf{x}}(\hat{\mathbf{s}}, \tilde{\mathbf{z}}|\mathbf{x})}_{\text{边信息的速率}} - \underbrace{\log p_{\tilde{\mathbf{z}}}(\tilde{\mathbf{z}})}_{\text{传输速率}} - \underbrace{\log q_{\hat{\mathbf{s}}|\tilde{\mathbf{z}}}(\hat{\mathbf{s}}|\tilde{\mathbf{z}})}_{\text{传输速率}} - \underbrace{\mathbb{E}_{\mathbf{y} \sim p_{\mathbf{y}|\hat{\mathbf{s}}, \tilde{\mathbf{z}}}} [\log p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y})]}_{\text{加权失真}} \right] + \text{const} \quad (3)$$

第一项中

$$q_{\hat{\mathbf{s}}, \tilde{\mathbf{z}}|\mathbf{x}}(\hat{\mathbf{s}}, \tilde{\mathbf{z}}|\mathbf{x}) = \prod_i \mathcal{N}(\hat{s}_i | s_i, \sigma_n^2) \cdot \prod_j \mathcal{U}\left(\tilde{z}_j | z_j - \frac{1}{2}, z_j + \frac{1}{2}\right) \quad (4)$$

其中, $\mathbf{y} = g_a(\mathbf{x}; \phi_g)$, $\mathbf{s} = f_e(\mathbf{y}; \phi_f)$, $\mathbf{z} = h_a(\mathbf{y}; \phi_h)$ 。由于均匀分布的宽度是恒定的, 高斯分布也只与信道有关, 因此第一项不会对训练产生影响, 从而可以在损失函数中省去。

第二项表示边信息 $\tilde{\mathbf{z}}$ 的传输速率, 第三项表示表示语义特征在边信息给定条件下的速率。两项之和表示一共需要传输的信息的总速率。

第四项表示恢复 $\hat{\mathbf{x}}$ 的对数似然, 即综合变换 g_s 的输出, 它由 DJSCC 解码器 f_d 的输出加权后构成。在高斯分布的假设下, 该项相当于均方误差。

因此, NTSCC 的优化可以看作是一个传输率失真的优化问题, 损失函数可以写作

$$L = \mathbb{E}_{\mathbf{x} \sim p_{\mathbf{x}}(\mathbf{x})} \left[\lambda \left(\underbrace{-\eta \log P_{\tilde{\mathbf{y}}|\tilde{\mathbf{z}}}(\tilde{\mathbf{y}}|\tilde{\mathbf{z}})}_{k_y} - \underbrace{\frac{\log P_{\tilde{\mathbf{z}}}(\tilde{\mathbf{z}})}{C_z}}_{k_z} \right) + d(\mathbf{x}, \hat{\mathbf{x}}_{\text{NTSCC}}) + d(\mathbf{x}, \hat{\mathbf{x}}_{\text{NTC}}) \right] \quad (5)$$

其中, $\mathbf{y} = g_a(\mathbf{x}; \phi_g)$, $\mathbf{z} = h_a(\mathbf{y}; \phi_h)$, $\hat{\mathbf{s}} = W(f_e(\mathbf{y}; \phi_f); \mathbf{v})$, $(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}) = h_s(\tilde{\mathbf{z}}; \theta_h)$, $\hat{\mathbf{y}} = f_d^{\hat{\mathbf{a}}}(\hat{\mathbf{s}}, \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}; \theta_{f^{\hat{\mathbf{a}}}})$, $\hat{\mathbf{x}} = g_s(\hat{\mathbf{y}}; \theta_g)$, 超参数 η 为一个比例系数, 控制估计速率的权重, C_z 表示传输量化后超先验 $\tilde{\mathbf{z}}$ 的信道的信道容量。总的传输信道带宽项上的拉格朗日乘数 λ 决定了传输速率和端到端失真之间的权衡。基于之前的分析, 条件熵模型 $P_{\tilde{\mathbf{y}}|\tilde{\mathbf{z}}}(\tilde{\mathbf{y}}|\tilde{\mathbf{z}})$ 和超先验的熵模型 $p_{\tilde{\mathbf{z}}}(\tilde{\mathbf{z}})$ 可以做因式分解。为了使用梯度下降方法来优化 NTSCC 模型, 类似于 NTC 中的做法, 将量化变量 \mathbf{y} 和 \mathbf{z} 的过程放松为向其中的每一项 y_i 和 z_i 添加均匀噪声的过程:

$$\tilde{y}_i = y_i + o_i, \tilde{z}_i = z_i + o_j, o_i \sim \mathcal{U}\left[-\frac{1}{2}, \frac{1}{2}\right] \quad (6)$$

其中, o 为服从均匀分布的随机变量。

会议名称	IMT-2030(6G)语义通信任务组第三次会议	会议地点	北京
提交单位	北京邮电大学	会议时间	2024 年 3 月 30 日

综上所述, KL 散度的最小化相当于联合优化非线性变换模型和 DJSCC 模型, 以实现优化 NTSCC 的端到端传输 R-D 性能。

4 实验结果

4.1 实验准备

- 数据集: 本提案在 RGB 小型图像数据集 CIFAR10[12]、中型数据集 Kodak[13]和大型数据集 CLIC2021[14]上的进行了无线图像传输的试验。这些数据集被广泛用于衡量图像相关算法的性能。用于训练 NTSCC 模型的数据集由从 Open Images[15]数据集中采样的 50 万张图像组成。在模型训练过程中, 用于训练的图像被随机裁剪成 256×256 个图像块。
- 对比方案: 本提案的对比方案包括 DJSCC 方案和经典的信源信道分离编码方案。其中信源信道分离编码方案采用 BPG 和 LDPC 码的方法来实现, 并映射为复信号在高斯信道中进行传输。
- 评估指标: 主要选用三种评估指标来度量传输性能的差异, 分别是峰值信噪比 (peak signal-to-noise ratio, PSNR)、多尺度结构相似度指数 (multi-scale structural similarity index, MS-SSIM) [16]和基于学习的感知图像块相似性 (learned perceptual image patch similarity, LPIPS) [17]。其中, PSNR 和 MS-SSIM 是目前最广泛使用的指标, 但它们并不能很好的解释人类感知的许多细微差别。为了更接近于语义通信的目的, 本提案进一步采用新兴的基于深度学习的 LPIPS 度量作为感知损失来量化图像传输性能, 它可以模仿人类的感知评估过程来给出 LPIPS 损失评分。

4.2 结果分析

- PSNR: 图 7 显示了在信道信噪比为 10 dB 的 AWGN 信道上各种传输方法的 R-D 结果, 其中失真度量选用 PSNR 测量。为了实现不同的 R-D 权衡, 采用了 $\lambda \in \{1024, 256, 64, 16, 4\}$ 训练的 NTSCC 模型, 并设置 $\eta = 0.2$ 。对于信源信道分离编码方案, 采用了一个 $2/3$ 的速率 (4096,6144) 的 LDPC 码与 16QAM 调制作为“BPG+LDPC”方案的配置。表 1 提供了 BD-CBR 和 BD-PSNR 的结果, 其中相对度量“BD-X”已被广泛用于评估不同的图像/视频压缩系统[18]的性能。其中, BD-CBR 指的是在相同的 PSNR 下, 相比基线算法“BPG+LDPC”, 平均节省的 CBR 的百分比; BD-PSNR 表示

会议名称	IMT-2030(6G)语义通信任务组第三次会议	会议地点	北京
提交单位	北京邮电大学	会议时间	2024 年 3 月 30 日

在相同的 CBR 下的平均 PSNR 增益。结果显示，本提案所提出的 NTSCC 方法基本上对所有信道带宽比都优于 DJSCC，且其性能差距随着图像数据集的分辨率和信道带宽比的增加而增大。与信源信道分离编码方法相比，NTSCC 与“BPG+NTSCC”方案相比具有明显优势，并且在三种数据集上都超越了实际的“BPG+LDPC”方案。

表 1 BD-CBR 和 BD-PSNR 性能比较

方法	CIFAR10 数据集		Kodak 数据集		CLIC2021 数据集	
	BD-CBR	BD-PSNR(dB)	BD-CBR	BD-PSNR(dB)	BD-CBR	BD-PSNR(dB)
NTSCC	-28.91%	2.64	-17.96%	0.81	-28.09%	1.31
BPG+Capacity	-21.41%	1.77	-22.91%	0.97	-22.91%	1.09
BPG+LDPC	0.00%	0	0.00%	0	0.00%	0
DJSCC	-22.71%	1.71	45.48%	-1.55	54.31%	-1.76

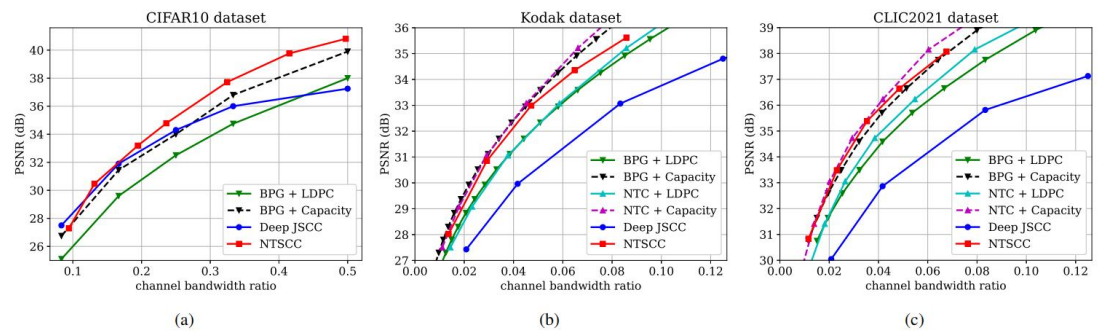


图 7 PSNR 随平均 CBR 的变化曲线图 (SNR=10 dB)

- MS-SSIM: 图 8(a)和图 8(b)显示了在信噪比为 10 dB 的 AWGN 信道上的 MS-SSIM 的 RD 性能。图 8(c)和图 8(d) 显示了 MS-SSIM 性能与信噪比的变化，其中平均 CBR 设置为 1/16。表 2 提供了 BD-CBR 和 BD-MS-SSIM 的结果.由于 MS-SSIM 的取值范围为 0（最差）到 1（最好），且大部分数值在 0.9 以上，因此这里采用 dB 为单位。为了实现不同的 R-D 权衡，采用了 $\lambda \in \{1, 1/4, 1/16, 1/32, 1/128\}$ 训练的 NTSCC 模型，并设置 $\eta = 0.4$ 。 λ 值越低导致带宽比 R 越大。结果显示，NTSCC 方案的性能优于传统的分离信源信道编码方案和 DJSCC 方案。并在高分辨率图像和高 CBR 区域上取得了较大的改进。

表 2 BD-CBR 和 BD-MS-SSIM 性能比较

方法	Kodak 数据集	CLIC2021 数据集
----	-----------	--------------

会议名称	IMT-2030(6G)语义通信任务组第三次会议	会议地点	北京
提交单位	北京邮电大学	会议时间	2024 年 3 月 30 日

	BD-CBR	BD-MS-SSIM(dB)	BD-CBR	BD-MS-SSIM(dB)
NTSCC	-43.01%	2.29	-64.27%	4.18
BPG+Capacity	-20.99%	1.30	-22.92%	1.09
BPG+LDPC	0.00%	0	0.00%	0
DJSCC	-21.47%	1.01	-31.77%	1.77

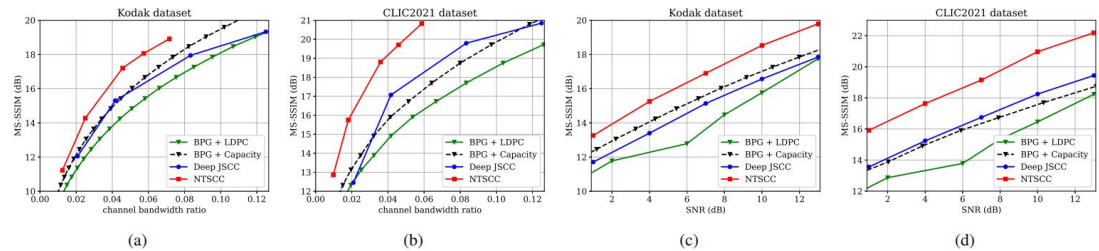


图 8 MS-SSIM 随平均 CBR 与信噪比的变化曲线图

- LPIPS: 用 LPIPS 指标作为损失函数中的失真项, 对 NTSCC 进行训练后, 在 Kodak 和 CLIC2021 两种数据集下进行测试, 测试结果如图 9 所示。可以观察到, NTSCC 可以以更低的信道带宽成本实现更高的视觉质量。由于 LPIPS 指标更加贴近人眼的主观感受, 因此所提出的 NTSCC 方法可以更好地支持未来的语义通信。

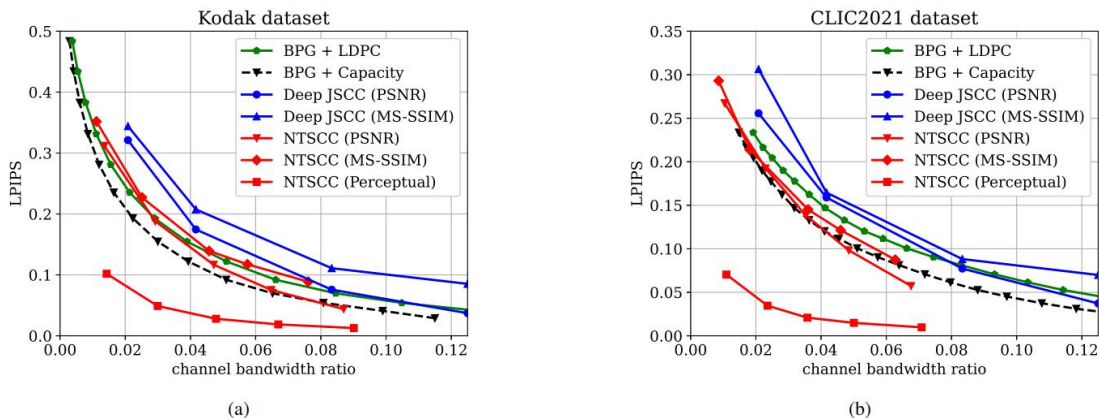


图 9 LPIPS 感知性能随平均 CBR 的变化曲线图 (SNR=10dB)

4.3 重建效果示例

为了直观地展示感知优化的效果, 图 10 和图 11 展示了测试集上图像重建结果。其中, 第一列显示了原始图像。第二列显示了在原始图像中截取的一个图像块。第三列到第六列显示了在信噪比为 10 dB 的 AWGN 信道上使用不同传输方案的重建图像, 括号中的指标

文稿编号: IMT-2030-Semantic_2024xxx

会议名称	IMT-2030(6G)语义通信任务组第三次会议	会议地点	北京
提交单位	北京邮电大学	会议时间	2024 年 3 月 30 日

表示模型训练目标失真项。红色数字和蓝色数字表示与“BPG+LDPC”方案相比，带宽成本增加和节省的百分比。







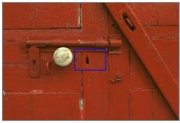














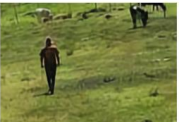


	Original	BPG + LDPC	Deep JSCC (PSNR)	NTSCC (PSNR)	NTSCC (Perceptual)
 512 × 768	 <i>R</i> / PSNR (dB)	 0.043 (0%) / 34.22	 0.042 (-2%) / 31.10	 0.038 (-12%) / 34.22	 0.039 (-9%) / 33.62
 512 × 768	 <i>R</i> / PSNR (dB)	 0.023 (0%) / 31.97	 0.042 (+83%) / 31.71	 0.022 (-4%) / 32.27	 0.023 (0%) / 31.46
 512 × 768	 <i>R</i> / PSNR (dB)	 0.029 (0%) / 23.12	 0.042 (+45%) / 23.67	 0.026 (-10%) / 23.44	 0.027 (-7%) / 22.99
 1512 × 2016	 <i>R</i> / PSNR (dB)	 0.020 (0%) / 23.71	 0.021 (+5%) / 25.09	 0.020 (0%) / 26.67	 0.018 (-10%) / 25.50

图 10 不同方案下的图片重建效果对比 (度量指标为 PSNR)







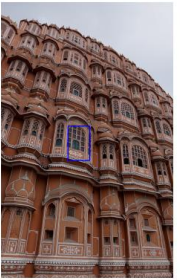





	Original	BPG + LDPC	Deep JSCC (MS-SSIM)	NTSCC (MS-SSIM)	NTSCC (Perceptual)
 1499 × 1002	 <i>R</i> / MS-SSIM (dB)	 0.022 (0%) / 12.84	 0.021 (-5%) / 11.67	 0.019 (-14%) / 14.78	 0.022 (0%) / 15.45
 2048 × 1365	 <i>R</i> / MS-SSIM (dB)	 0.042 (0%) / 14.93	 0.042 (0%) / 15.85	 0.025 (-40%) / 14.89	 0.033 (-21%) / 16.68

图 11 不同方案下的图片重建效果对比 (度量指标为 MS-SSIM)

会议名称	IMT-2030(6G)语义通信任务组第三次会议	会议地点	北京
提交单位	北京邮电大学	会议时间	2024 年 3 月 30 日

从上述示例中，可以观察到在感知损失下训练的 NTSCC 模型可以以更低的信道带宽成本实现更高的视觉质量。特别是，它有效地避免了伪影，即原本被拍摄物体并不存在而在图像上却出现各种形态的影像。同时，NTSCC 具有对更多细节的高保真重建，而传统的“BPG+LDPC”方案则达不到上述效果。

5 总结

本提案提出了一种基于非线性变换编码的语义编码方案 NTSCC。该方案通过非线性解析变换，将信源符号映射到语义特征向量空间，利用熵模型实现针对于语义特征的速率估计，并依照估计的速率信息，利用联合信源信道编码实现可变速率传输。NTSCC 面向率失真形式的损失函数进行优化训练，且失真度量更贴近人类的感知。因此，本提案所提语义通信方法可以显著提高语义传输的效率与可靠性，既能够保持较高的主观感知质量，又显著节省了信号传输带宽，体现出双重性能优势。相比于传统信源信道分离编码与标准的 DJSCC 方案，NTSCC 在不同信噪比、不同信道带宽比的条件下均可获得显著的性能增益。NTSCC 框架在未来的语义通信系统中有广阔的应用前景。

参考文献

- [1] 张平, 牛凯, 田辉, 等. 6G 移动通信技术展望[J]. Journal on Communications, 2019.
- [2] 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.
- [3] Bourtsoulatz E, Kurka D B, Gündüz D. Deep joint source-channel coding for wireless image transmission[J]. IEEE Transactions on Cognitive Communications and Networking, 2019, 5(3): 567-579.
- [4] Wang J, Wang S, Dai J, et al. Perceptual learned source-channel coding for high-fidelity image semantic transmission[C]//GLOBECOM 2022-2022 IEEE Global Communications Conference. IEEE, 2022: 3959-3964.
- [5] Wang S, Dai J, Liang Z, et al. Wireless deep video semantic transmission[J]. IEEE Journal on Selected Areas in Communications, 2022, 41(1): 214-229.
- [6] Xie H, Qin Z, Li G Y, et al. Deep learning enabled semantic communication systems[J]. IEEE Transactions on Signal Processing, 2021, 69: 2663-2675.
- [7] Xiao Z, Yao S, Dai J, et al. Wireless deep speech semantic transmission[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.
- [8] Dai J, Wang S, Tan K, et al. Nonlinear transform source-channel coding for semantic

文稿编号: IMT-2030-Semantic_2024xxx

会议名称	IMT-2030(6G)语义通信任务组第三次会议	会议地点	北京
提交单位	北京邮电大学	会议时间	2024 年 3 月 30 日

- communications[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(8): 2300-2316.
- [9] Ballé J, Minnen D, Singh S, et al. Variational image compression with a scale hyperprior[C]//International Conference on Learning Representations. 2018. Richardson T, Kudekar S. Design of low-density parity check codes for 5G new radio[J]. IEEE Communications Magazine, 2018, 56(3): 28-34.
- [10] Arıkan 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.
- [11] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
- [12] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images[J]. 2009.
- [13] Kodak PhotoCD dataset, URL: <http://r0k.us/graphics/kodak/>, 1993.
- [14] CLIC 2021: Challenge on learned image compression, URL: <http://compression.cc>, 2021.
- [15] Kuznetsova A, Rom H, Alldrin N, et al. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale[J]. International journal of computer vision, 2020, 128(7): 1956-1981.
- [16] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment[C]//The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. Ieee, 2003, 2: 1398-1402.
- [17] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 586-595.
- [18] Bjontegaard G. Calculation of average PSNR differences between RD-curves[J]. ITU SG16 Doc. VCEG-M33, 2001.