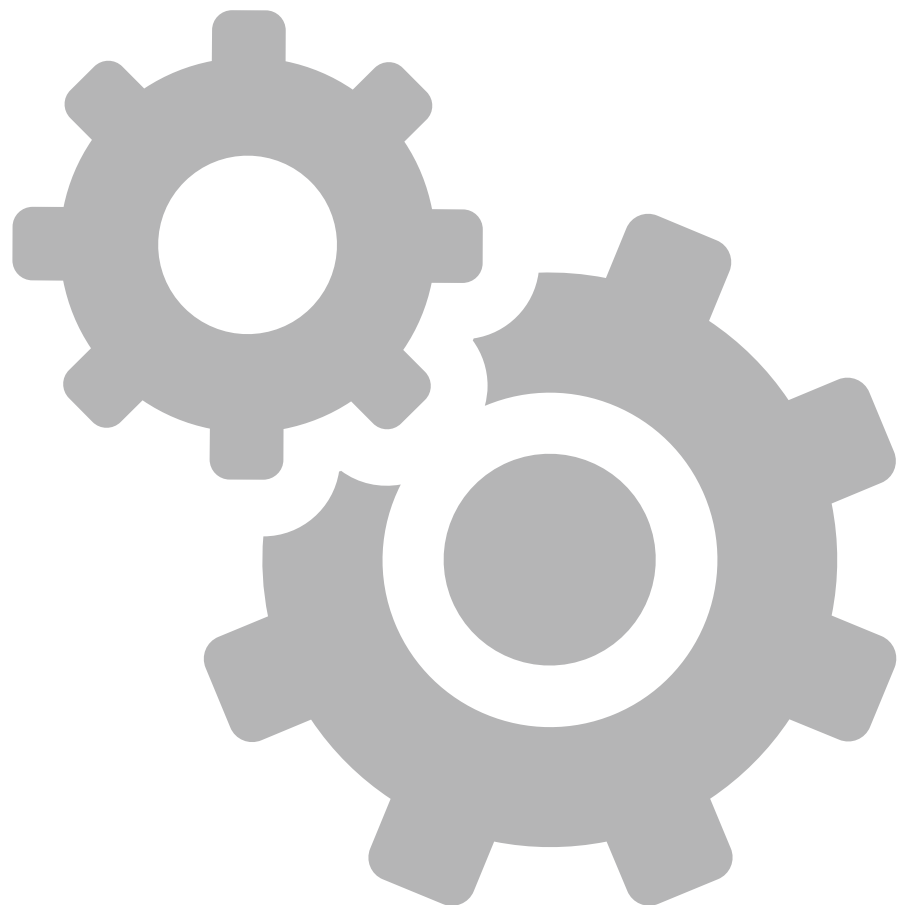


AI辅助的标准兼容型语义图像传输框架

吴泳澎、许铭楷、韩雪

上海交通大学

2025年8月



1

研究背景

2

研究挑战

3

研究内容

4

工作总结



上海交通大學

SHANGHAI JIAO TONG UNIVERSITY

PART ONE

研究背景

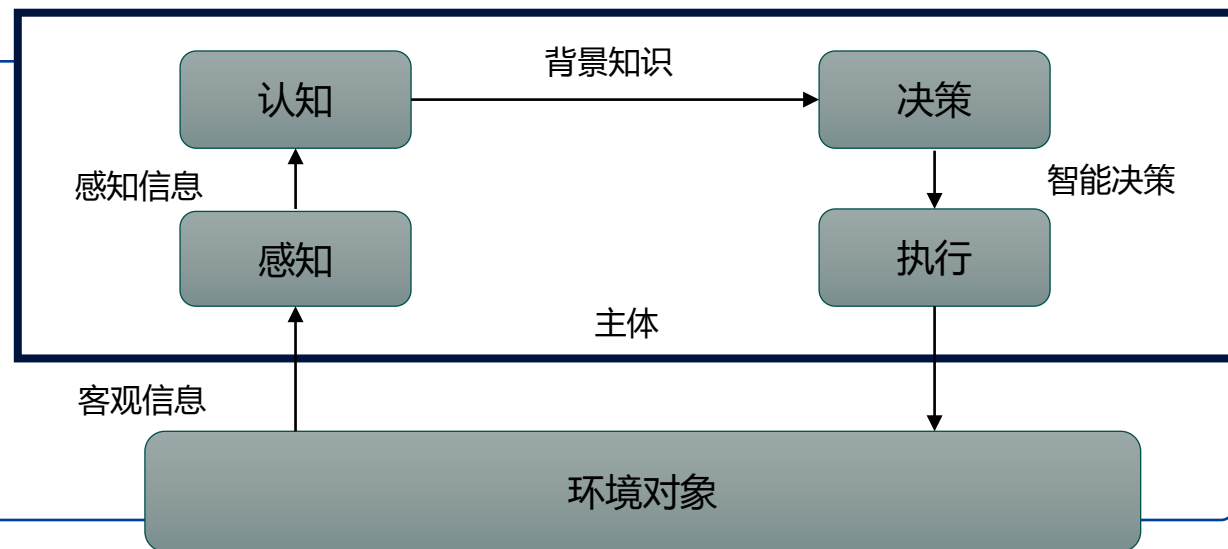
语义信息与语义通信概述



语义信息的定义：

是**感知信息** (Perceived information) 的一部分

- 基于使用者的认知与目标的抽象概念
- 是串联起语法与语用信息的桥梁

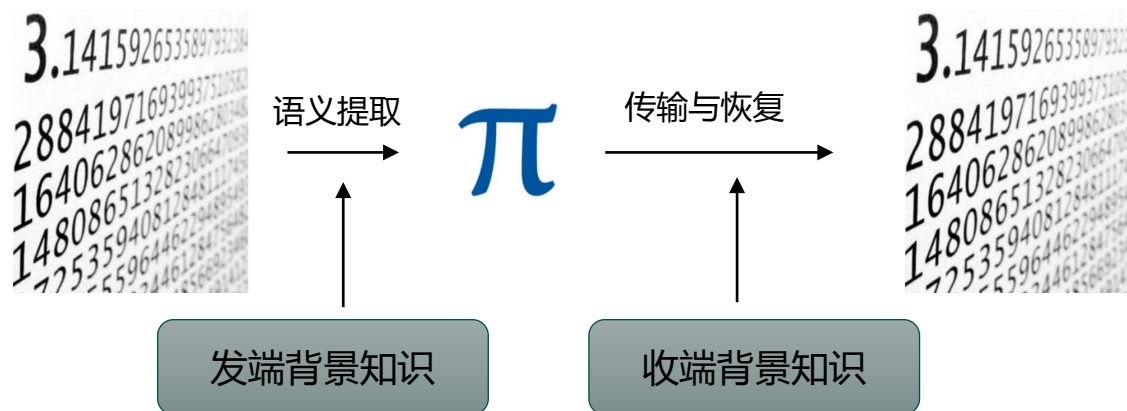


图：信息生态系统模型 [c.f. Zhong 2017]

语义通信的定义：

是传递语义信息的通信范式，6G内生智能的必要组件

- 高效的提取/传输语义信息，提升数据应用的可靠性和效率
- 是场景/任务导向的通信范式
- “智慧演化和原生简约”的“智简”网络

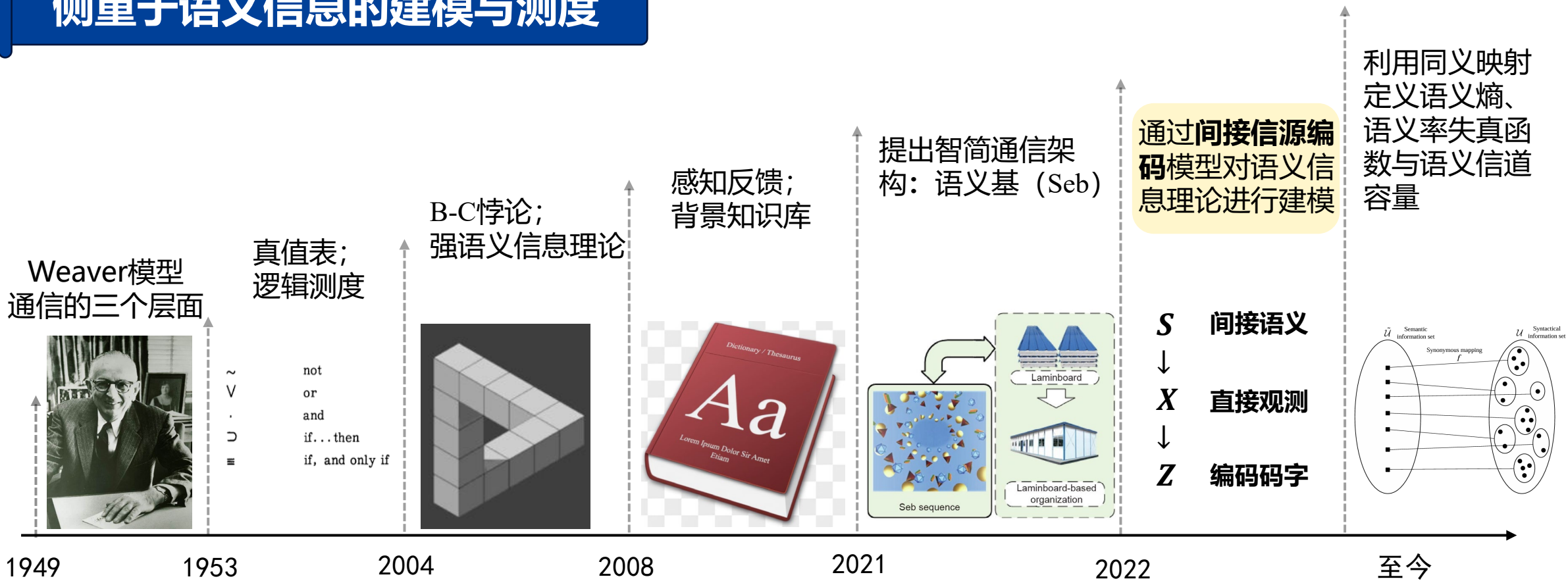


图：语义通信范式示意

语义信息理论的研究进展

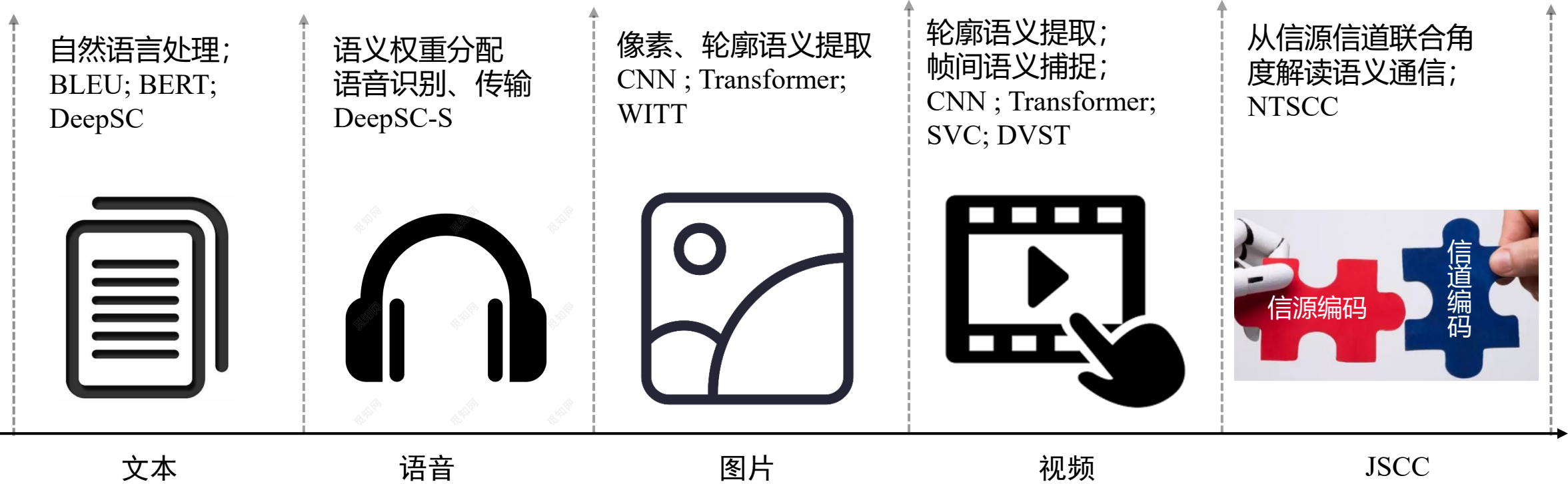


侧重于语义信息的建模与测度





侧重不同模态的语义提取与传输





上海交通大學

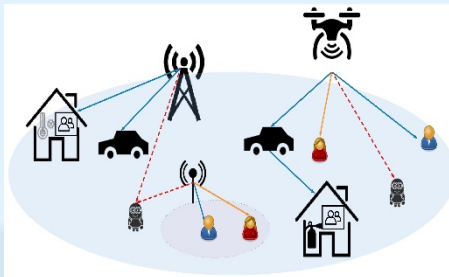
SHANGHAI JIAO TONG UNIVERSITY

PART TWO

研究挑战

无线视觉语义传输面临挑战

人机互联规模
日益扩大



需兼顾人眼视觉与机器视觉

通信场景目标
复杂多样



需兼顾重建保真与任务准确

现有协议硬件
缺乏兼容



难以充分发挥
现有通信架构

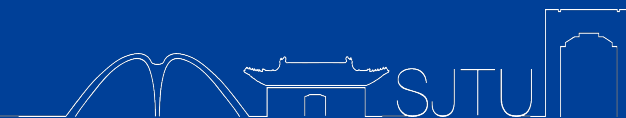
标准兼容语义
通信技术研究



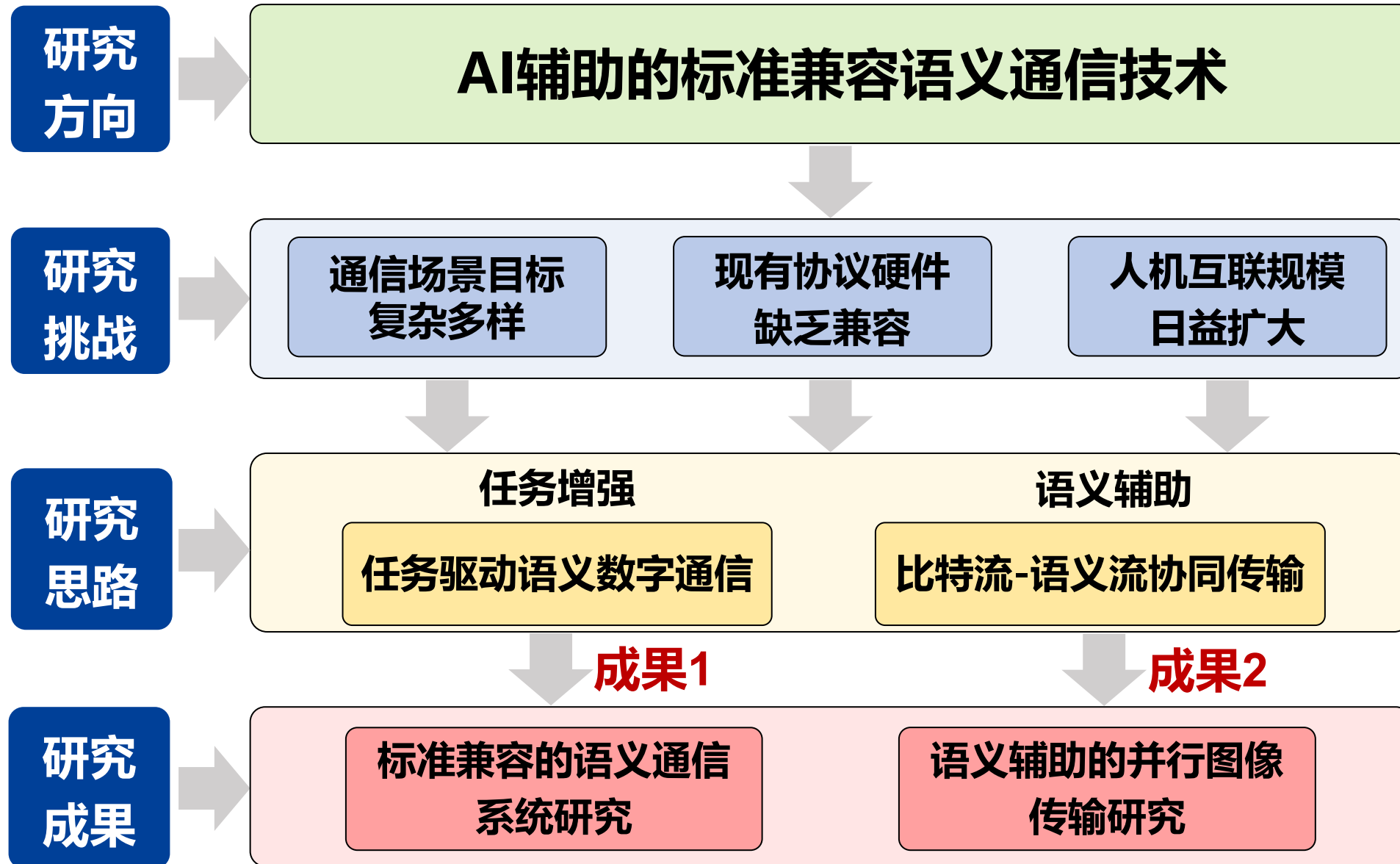
上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

PART THREE

研究内容



研究思路



研究内容 I: 标准兼容的语义通信系统研究

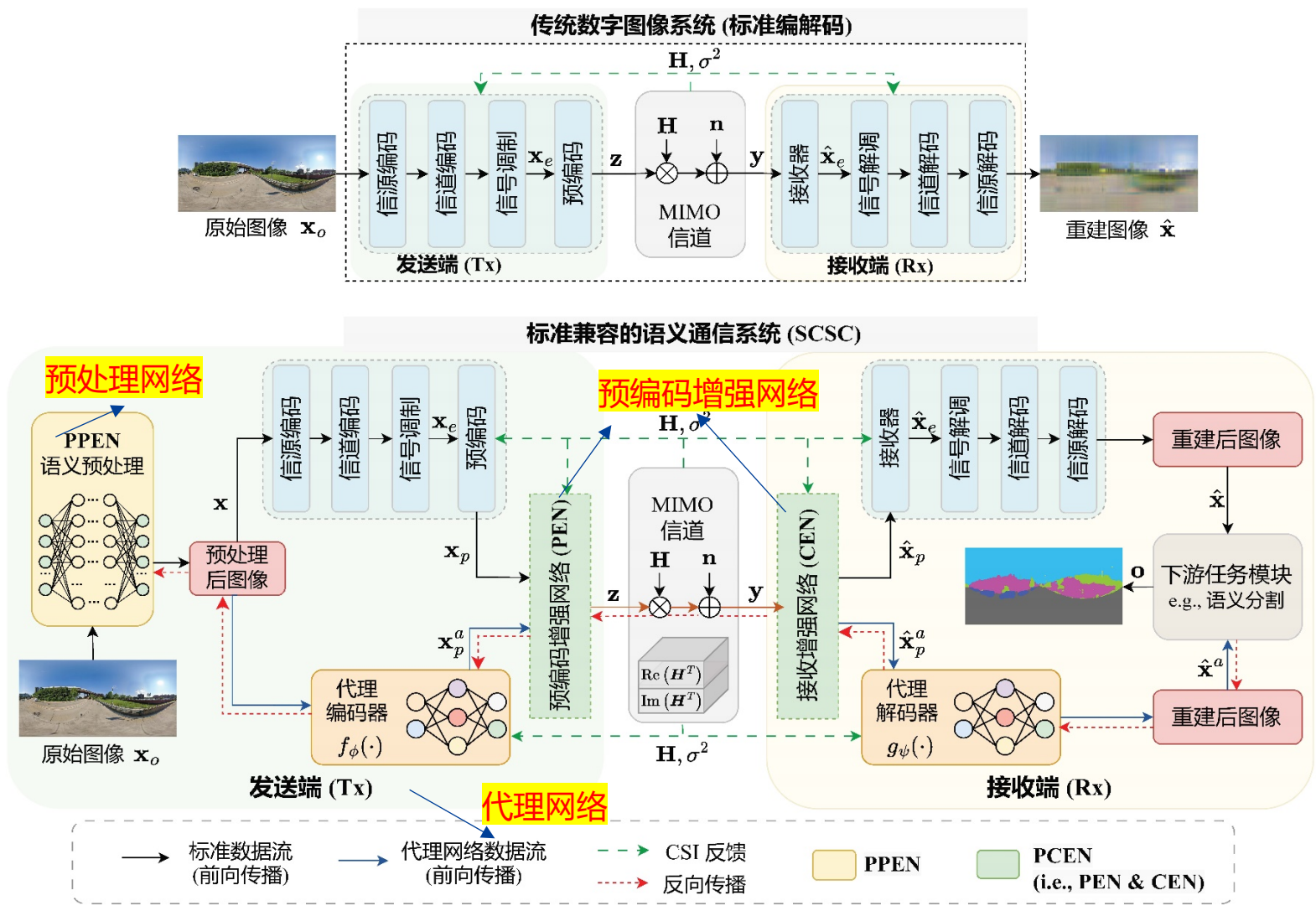


图: 传统标准的数字通信系统 & 标准兼容的语义通信系统

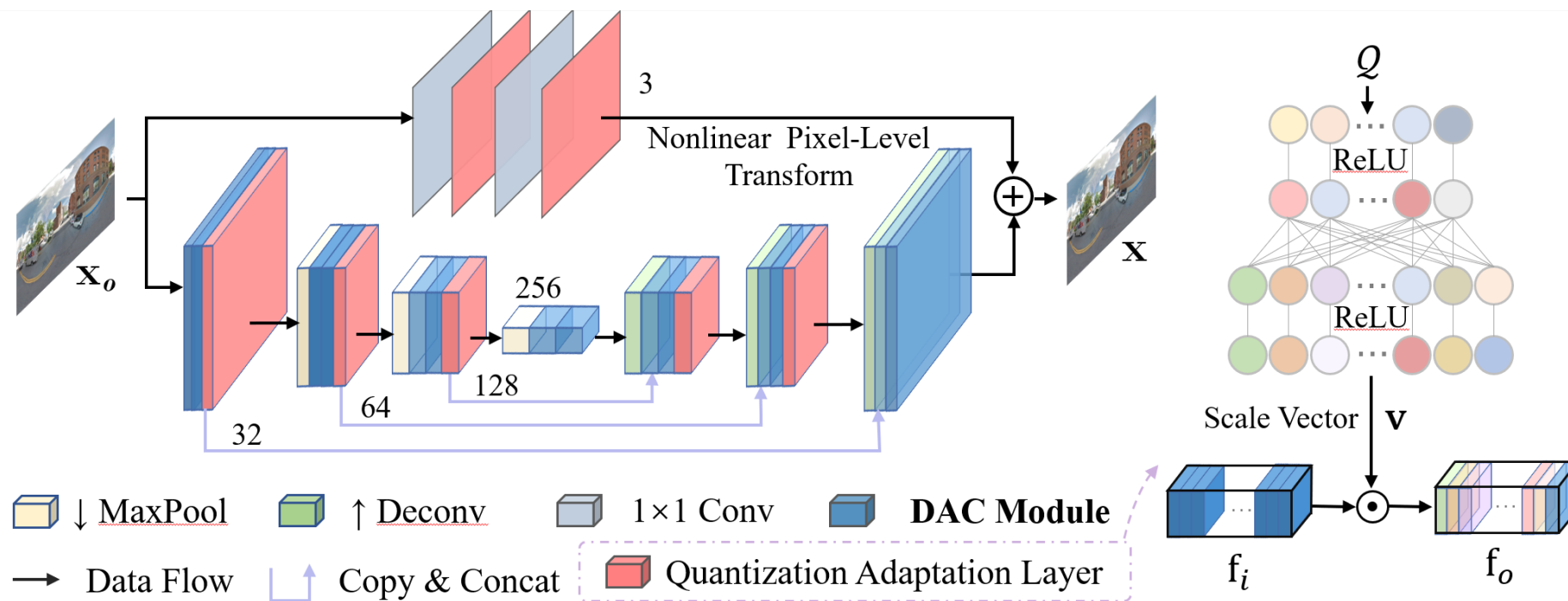
核心难题:

- ❑ 现有语义通信框架依赖于复杂的端到端网络来获得编码增益
- ❑ 信道建模较为理想, 对MIMO信道及**有限字符**信号研究较少
- ❑ 缺少语义通信实际应用的标准化协议和相应的硬件设计

研究内容:

- ❑ 基于数字图像通信系统, 设计标准兼容的语义通信系统
- (预处理, 预编码增强 & 代理网络)**

研究内容 I: 语义预处理网络



图：预处理网络结构图

整体网络结构

- X_o : 信源图像
- 1st 分支: 使用1×1个卷积层和量化自适应层来实现非线性像素级变换
- 2st 分支: 用于准确的语义特征提取

量化自适应层

- 标准源压缩具有不同的压缩比, 即不同量化参数, 在预处理模块中引入量化自适应层, 在实际应用中针对特定的压缩比生成滤波后的图像X

研究内容 I: 失真补偿网络



失真补偿网络:

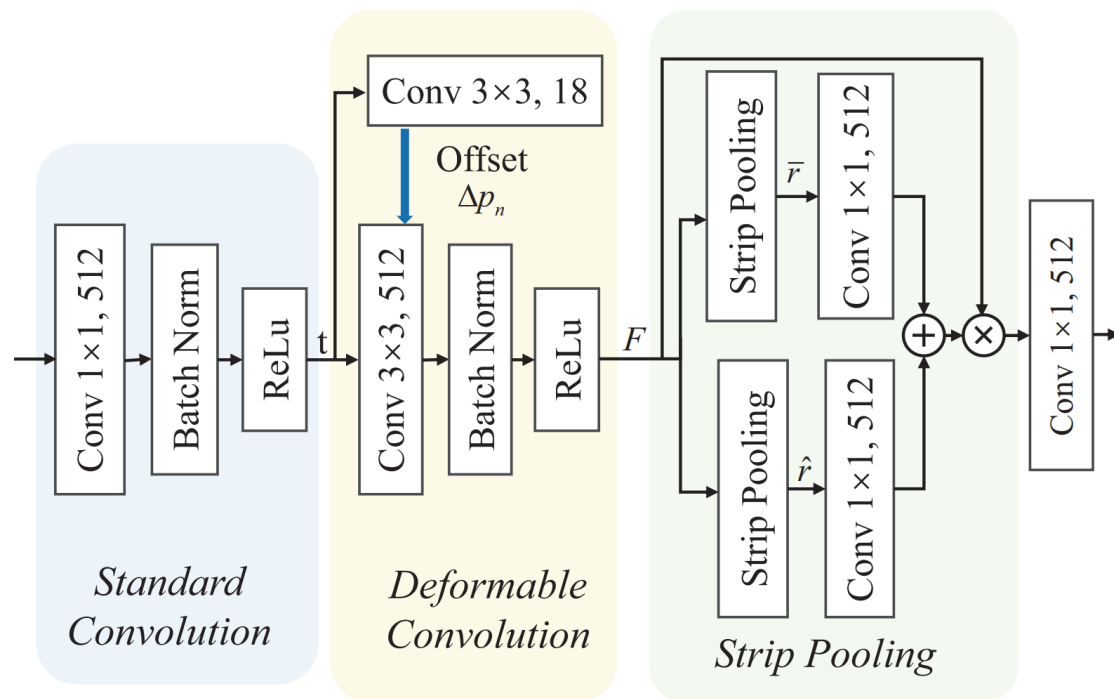


图: 失真补偿网络结构图

可变形卷积:

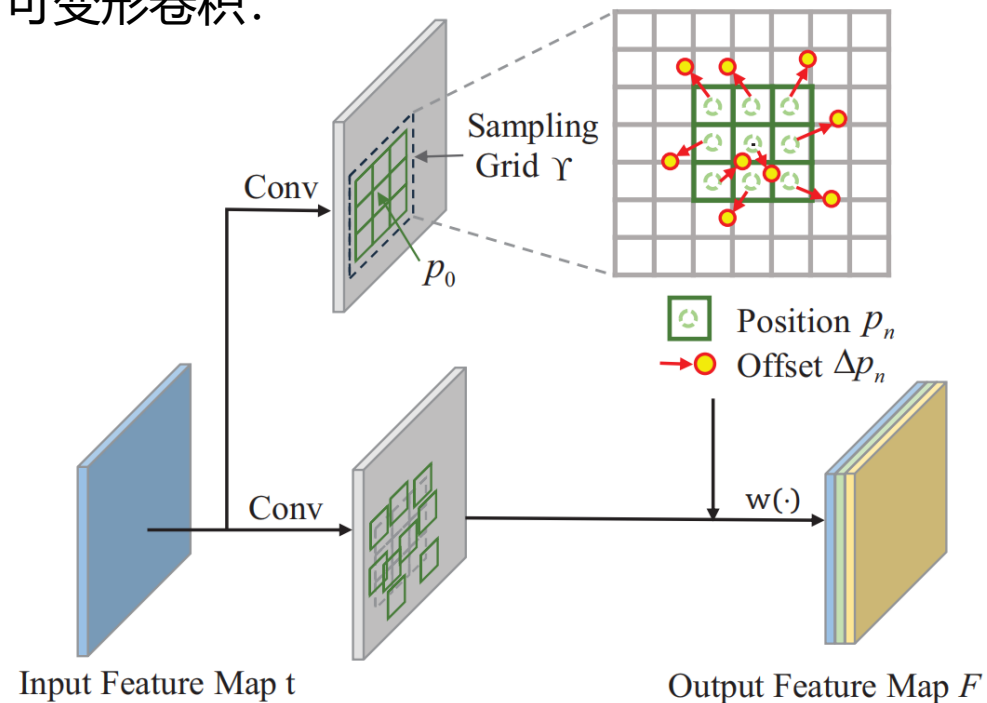
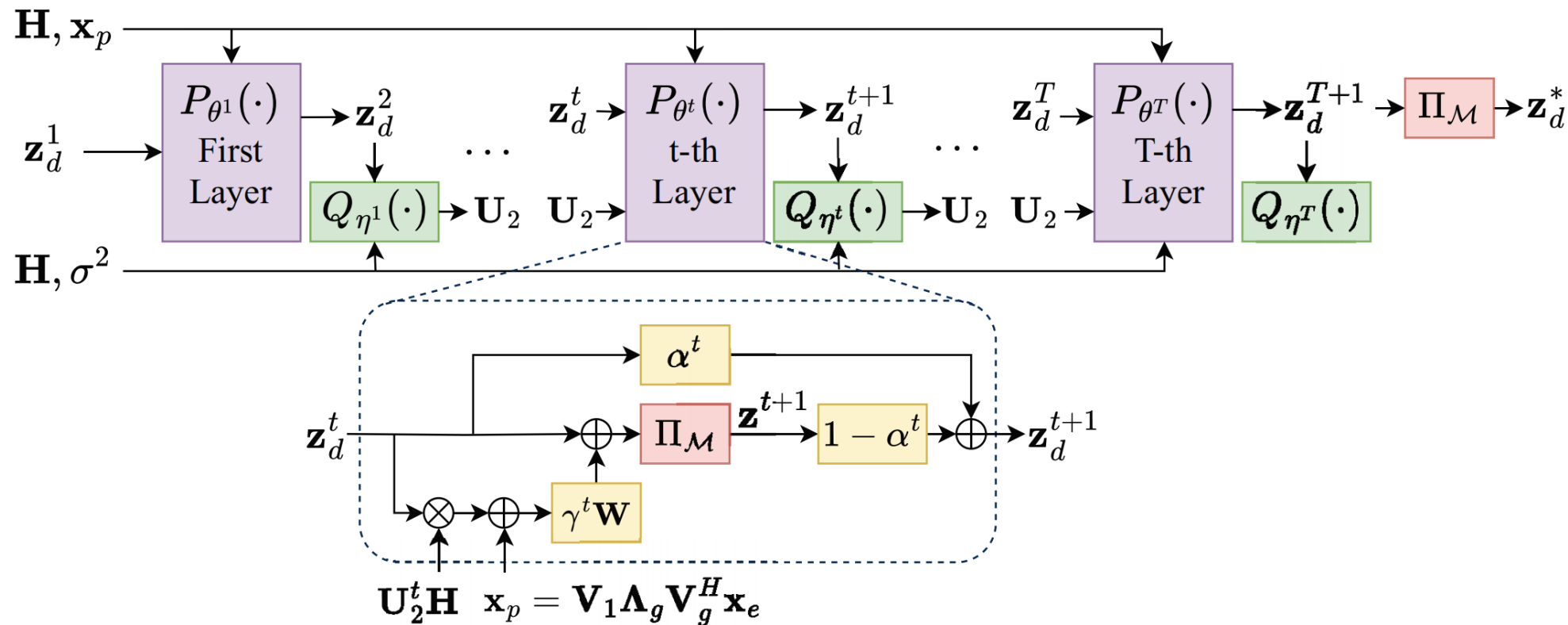


图: 可变形卷积网络示意图

可变形卷积

- 能更好地适应不规则的目标形状, 捕捉局部特征, 提高模型对物体形变的鲁棒性。

研究内容 I: 预编码增强网络



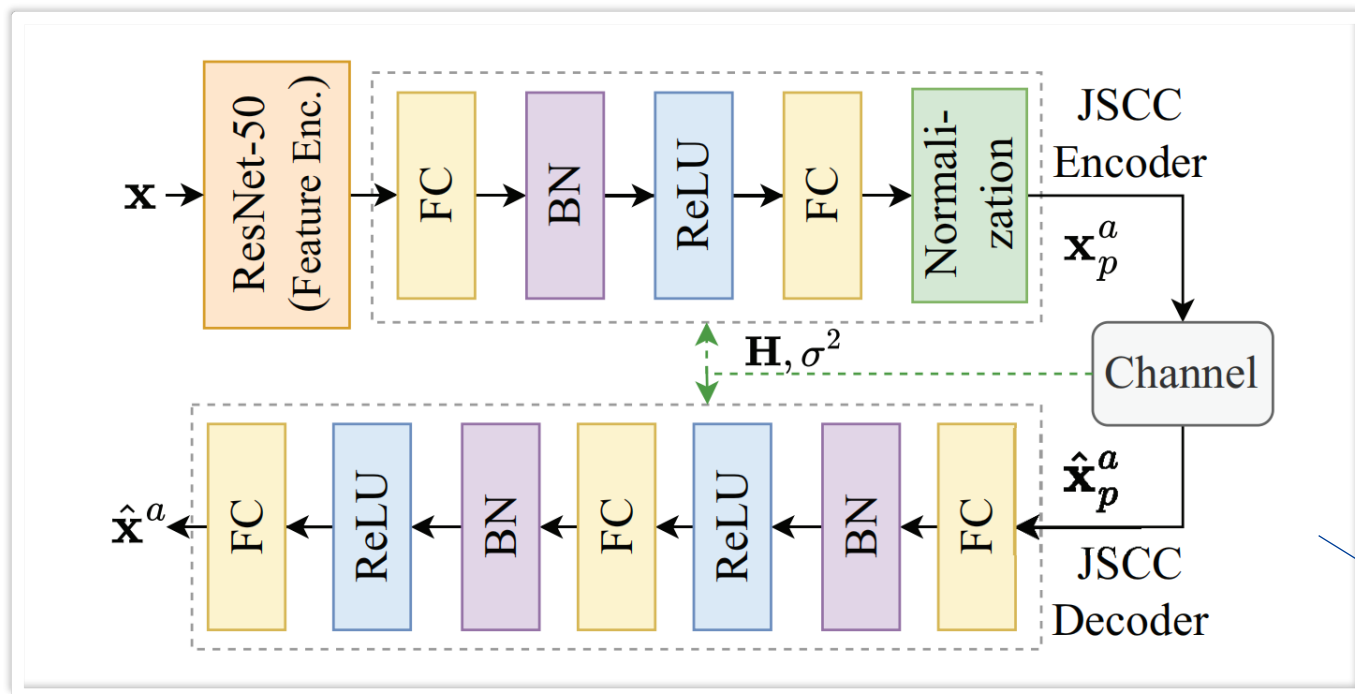
图：预编码增强网络结构图

- 将一组可训练的参数展开到网络结构中，降低模型的复杂度，减少参数数量
- 使用 θ 和 η 的交替优化方法来实现优化；

损失函数：

$$\begin{aligned}
 L_{pcen} &\triangleq \mathbb{E}_{x_p, n} \{ ||x_p - \hat{x}_p||^2 \} \\
 &= \mathbb{E}_{x_p, n} \{ ||x_p - Q_{\eta}(HP_{\theta}(x_p) + n)||^2 \}
 \end{aligned}$$

研究内容 I: 代理网络



图：代理网络结构图

模块思路：

□ 提出代理网络为了解决标准信源编解码器、信道编解码器等一些列不可微问题，实现模型端到端的训练。

采用了经典的Deep JSCC网络[DJSCC]

网络设计

为了保证前向传播中代理网络能够很好地近似数字编解码器，两者的重构质量应尽量相似，所以使用最小化两路数据流的重建图像的MSE作为损失函数。使用代理网络的梯度进行反向传播。

研究内容 I: MIMO衰落信道下相关数值结果

参数设定

- 数据集: Cityscapes & CVRG-Pano
- 下游任务模块: *ERF* – PSPNet [Yang]
- 评估矩阵: PSNR/MS-SSIM/mIoU
- 损失函数:

$$L \triangleq \lambda_1 L_{pcen} + \lambda_2 L_{seg} + \lambda_3 L_{pre}$$

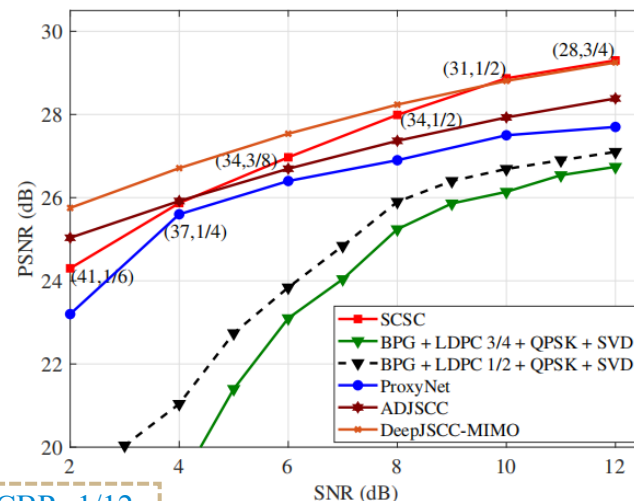
- 预编码增强网络损失函数:

$$L_{pcen} \triangleq \mathbb{E}_{x_p, n} \{ ||x_p - \hat{x}_p||^2 \}$$

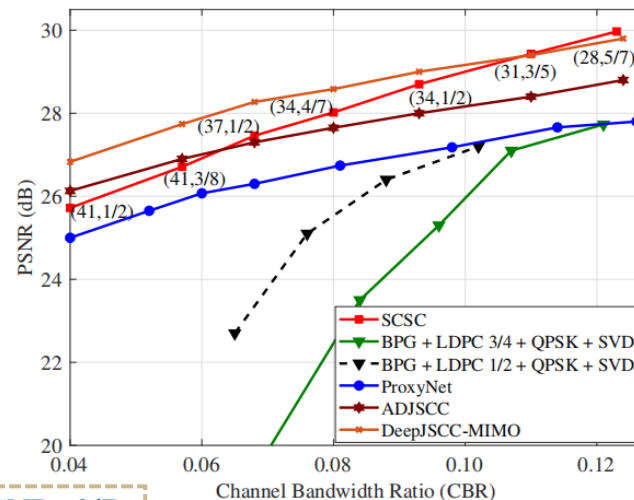
- 语义分割损失函数:

$$\mathcal{L}_{seg} = 1 - \frac{1}{B} \sum_{b=1}^B \frac{o_g^b \cap o_d^b}{o_g^b \cup o_d^b}$$

- 预处理网络损失函数 L_{pre} : 原图和预处理之后图像间的MSE损失

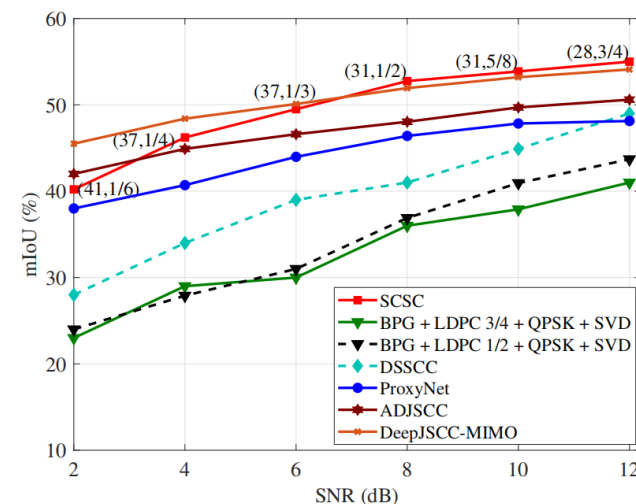


(a) PSNR versus SNR

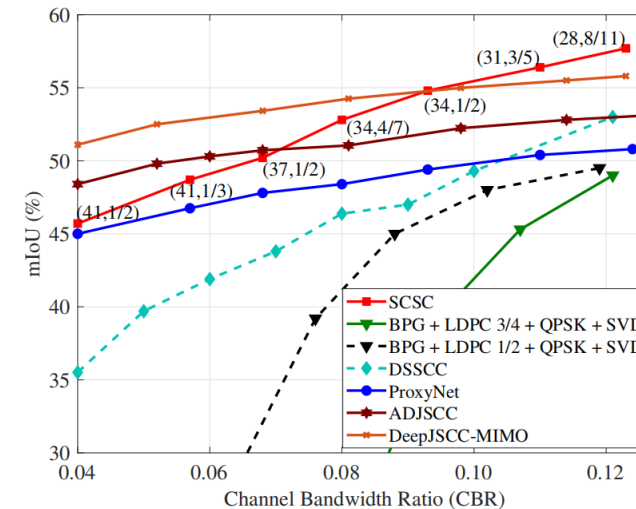


(b) PSNR versus CBR

图: PSNR结果对比



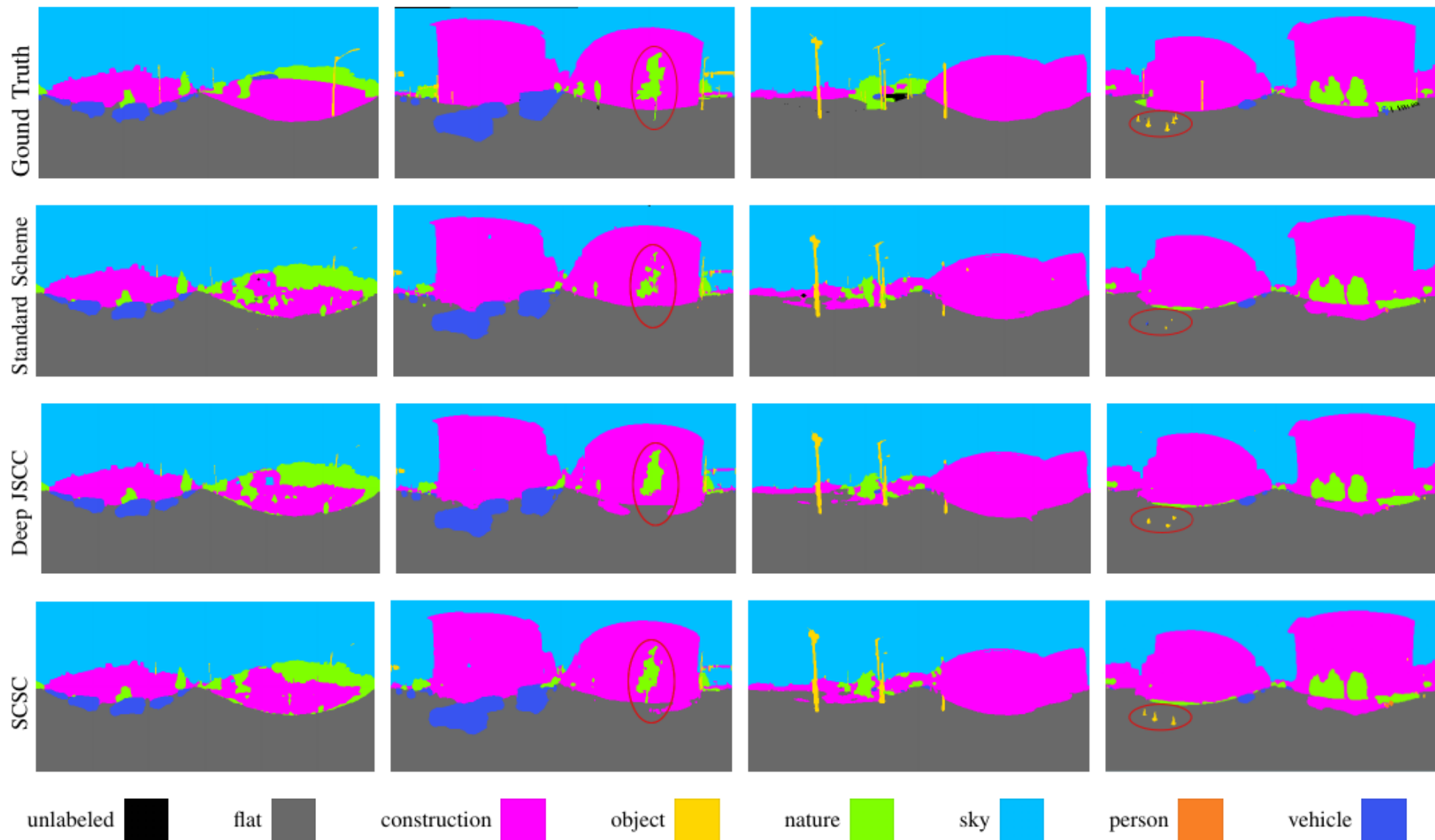
(a) mIoU versus SNR



(b) mIoU versus CBR

图: 语义分割mIoU结果对比

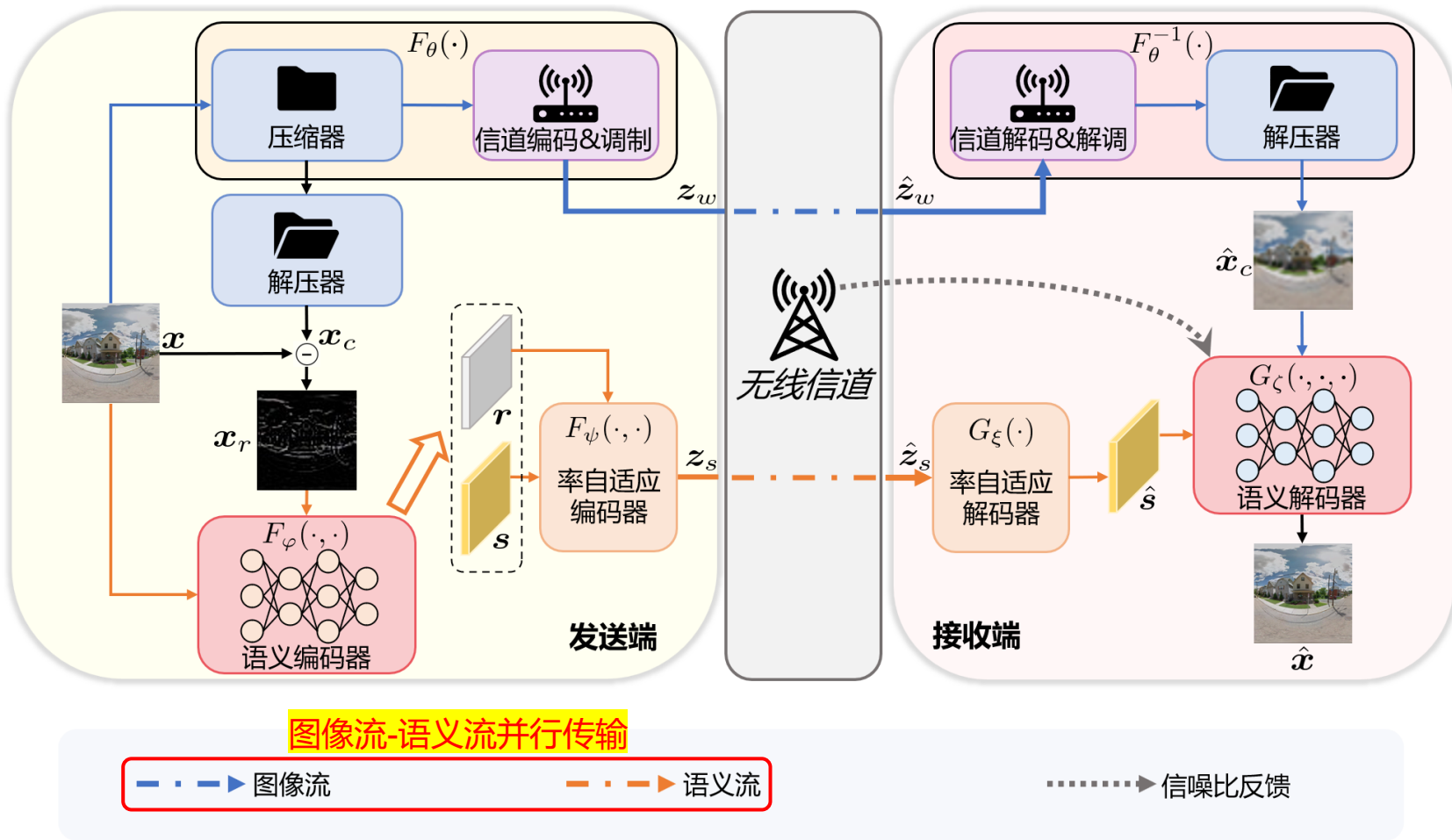
研究内容 I: MIMO衰落信道下可视化结果



结果分析:

语义分割实验中, 所提方案的结果准确度较高; 相较传统数字通信方案, 所提框架可以达到29%信道带宽 (CBR) 开销的减少。

研究内容 II：语义辅助的并行图像传输研究



图：ParaSC系统模型

核心难题

传统数字通信架构：

- ❑ 存在“悬崖效应”，难以克服信道环境动态变化引起的性能损失
- ❑ 香农信息论无法精确描述实际有限码长场景

基于AI的端到端语义通信架构：

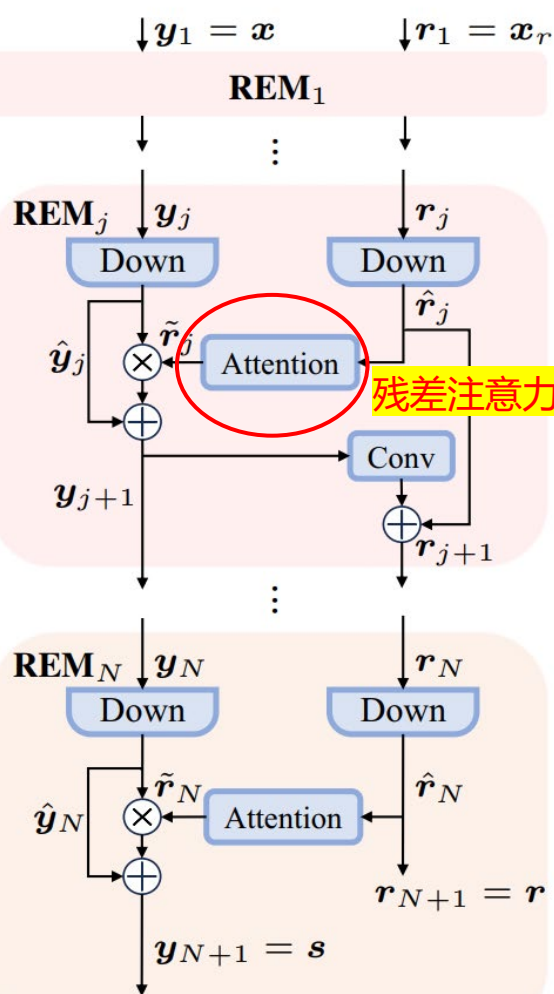
- ❑ 多采用固定码长设计，缺乏灵活的码率调控能力
- ❑ 缺少语义通信实际应用的标准化协议和相应的硬件设计

解决方案

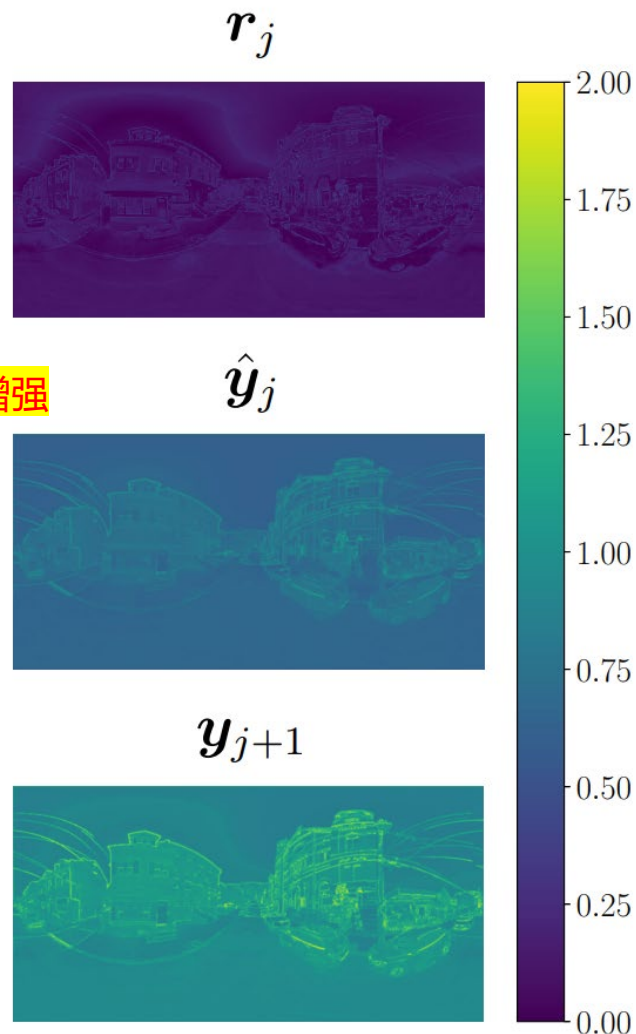
- ❑ 将**传统标准**和**AI语义**结合，提出双流并行图像传输框架**ParaSC**

❑ **核心组件：**残差语义编码、语义速率自适应、融合语义解码

研究内容 II: 残差语义增强编码



图：语义编码器结构



图：残差注意力增强可视化

整体结构:

- 基于CNN构建的轻量语义提取网络
- 网络输入：源图像 x 和残差 x_r
- 残差表示传统信源编解码器 (JPEG) 压缩图像的信息损失
- 采用两路结构，分别提取源图像语义特征和残差特征
- 两路间通过**残差注意力增强模块**交互

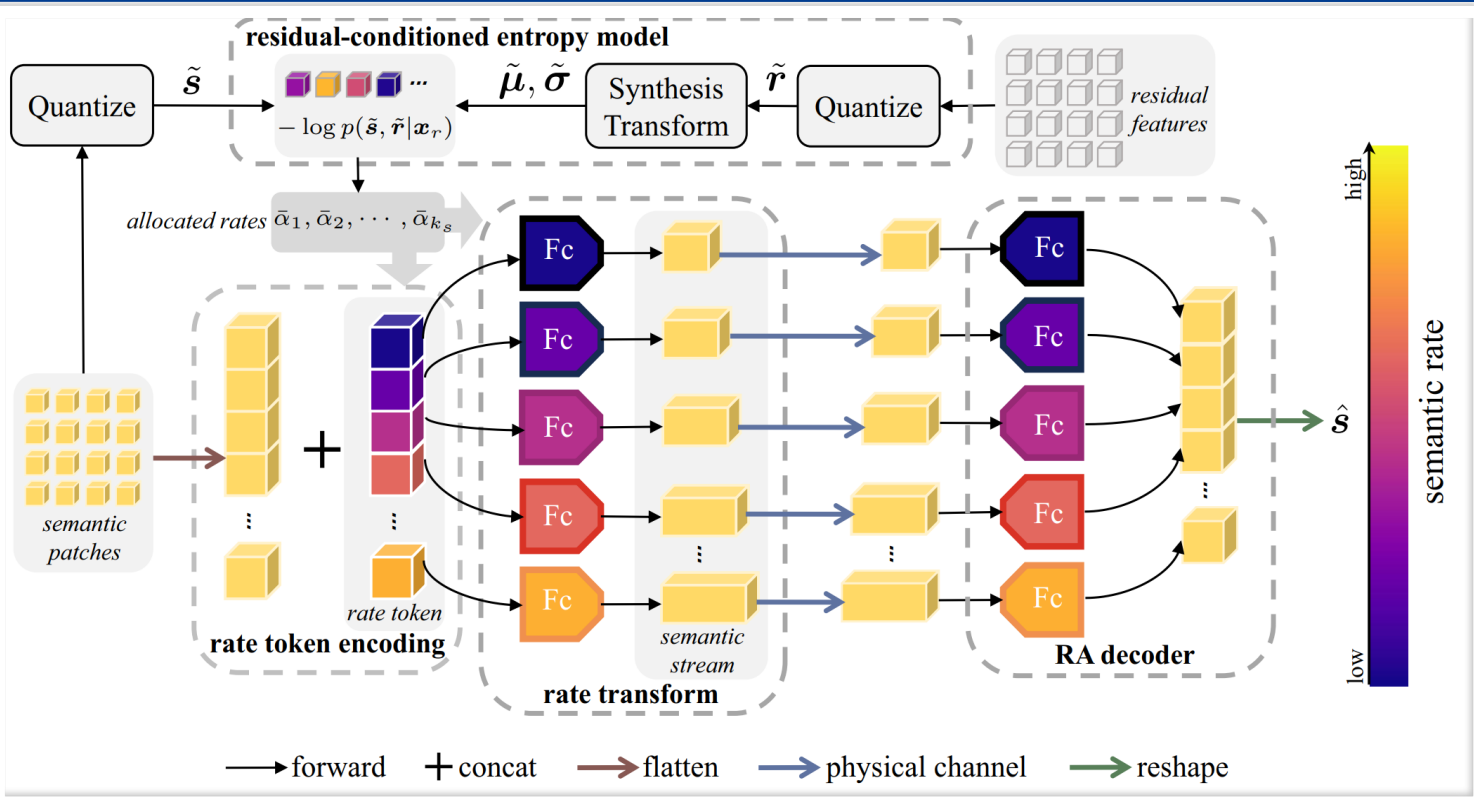
残差注意力增强:

- 单层卷积+Sigmoid激活，生成注意力分布图
- 将传统图像编码信息损失融入语义提取，获得增强语义表征

可视化:

- 高频分量激活，低频区域近似不变

研究内容 II: 残差条件的可变速率语义传输



图：语义速率自适应编解码模块结构

率-失真形式损失函数:

$$\mathcal{L}(\varphi, \psi, \xi, \zeta | \theta) = \mathcal{L}_{RD} + \lambda_2 \mathcal{L}_{perc},$$

图像感知

$$\mathcal{L}_{RD}(\varphi, \psi, \xi, \zeta | \theta) =$$

$$\mathbb{E}_{p(x)} \mathbb{E}_{q(\hat{z}_s, \tilde{r} | x, x_r)} [d(x, \hat{x}) - \lambda_1 \log p(\tilde{s}, \tilde{r} | x_r)].$$

图像失真

条件语义速率

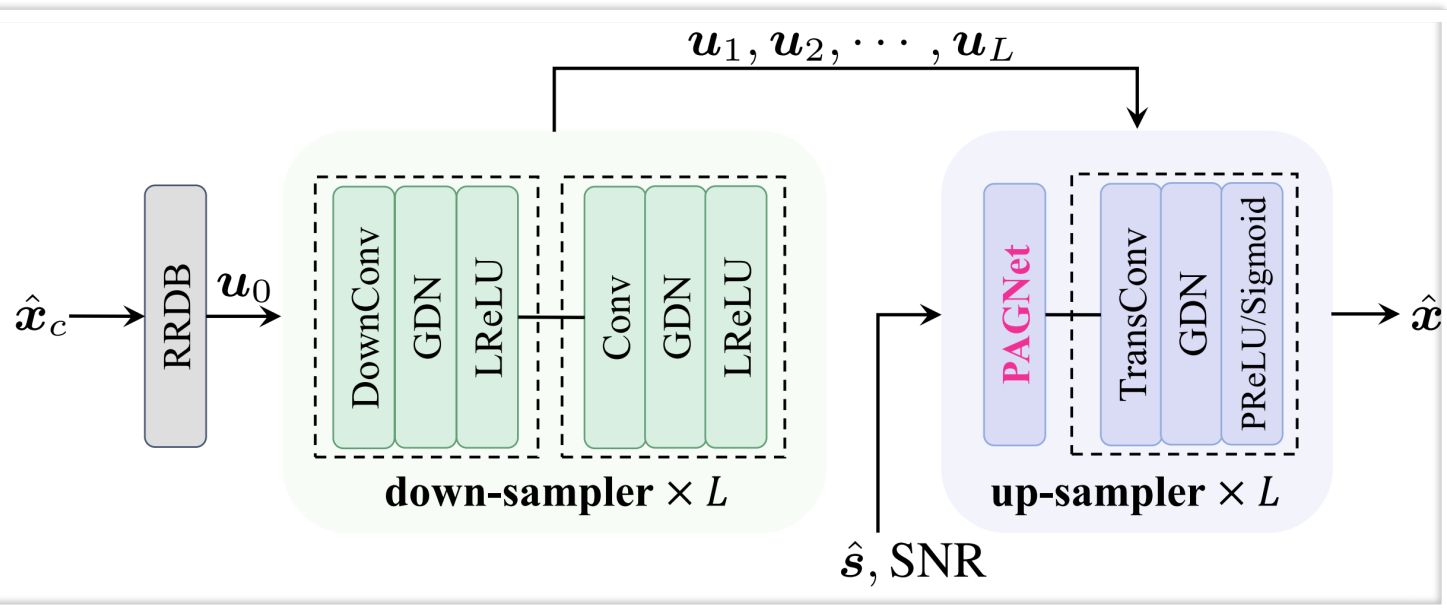
动机：利用残差进一步去除语义特征冗余，减少语义传输开销

- ❑ 图像流包含了源图像的大部分信息，语义流需要与之互补
- ❑ JPEG编码质量越高，需要额外传输的语义信息越少
- ❑ 用残差来刻画图像流的互补信息

方法：以残差为条件建立语义熵模型，实现变速率语义传输

- ❑ 条件熵模型基于高斯分布，参数由残差特征决定
- ❑ 通过条件变分自编码器 (CVAE) 导出优化目标
- ❑ 将估计熵作为最小语义传输长度，通过全连接层进行变换

研究内容 II：图像-语义融合解码



图：图像-语义融合解码模块结构

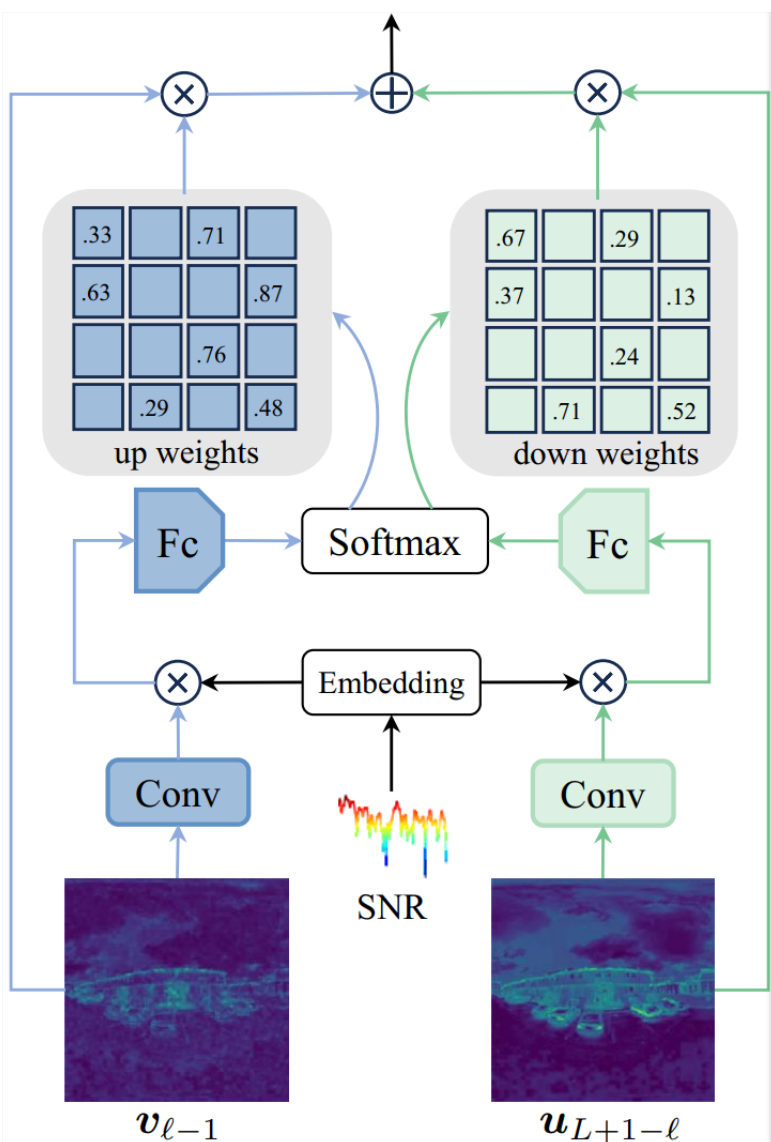
在接收端融合双流信息，以重建图像

- $\hat{x}_c$ ：传统分支重建图像，来自**图像流**
- $\hat{s}$ ：接收语义特征，来自**语义流**
- SNR：信道信噪比

基于U-Net结构，对 $\hat{x}_c$ 先下采样后上采样，在上采样过程中分层融入语义信息

- RRDB(Residual-in-Residual Dense Block)对图像流进行预处理
- 应用“卷积-归一化-激活”子结构，高效提取图像深层特征
- 提出**并行聚合网络(PAGNet)**应用于上采样阶段，根据SNR自适应融合双流信息，增强本框架对信道噪声的鲁棒性

研究内容 II: 并行聚合网络



图：并行聚合网络结构

动机：抑制“悬崖效应”对图像重建的负面影响

- 当信道容量显著高于传统分支码率时，传统解码图像质量高，双流可以同等参与重建
- 当信道容量接近或低于传统分支码率时，传统解码图像质量显著下降，应让语义流主导重建

方法：像素级别动态加权聚合双流

- 通过CNN识别图像流和语义流的噪声模式
- 将SNR嵌入，通过Softmax函数生成聚合权重
- 自适应保留传统解码图像中的清晰区域，同时抑制噪声区域

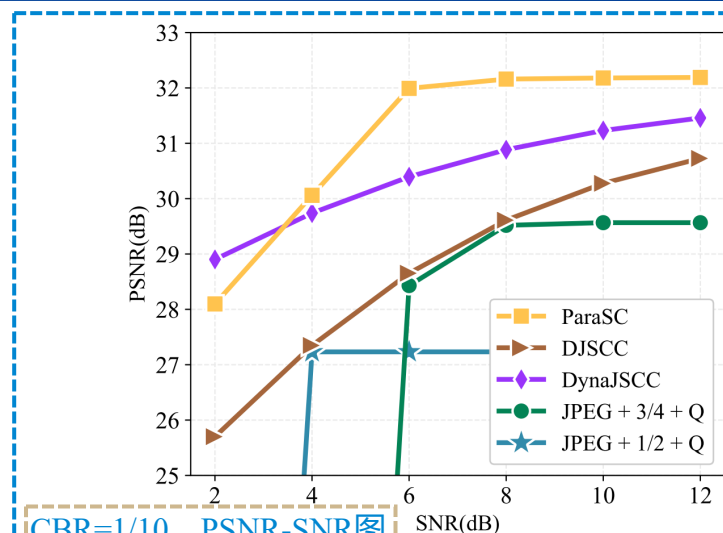
研究内容 II: 无线信道数值仿真结果

参数设定

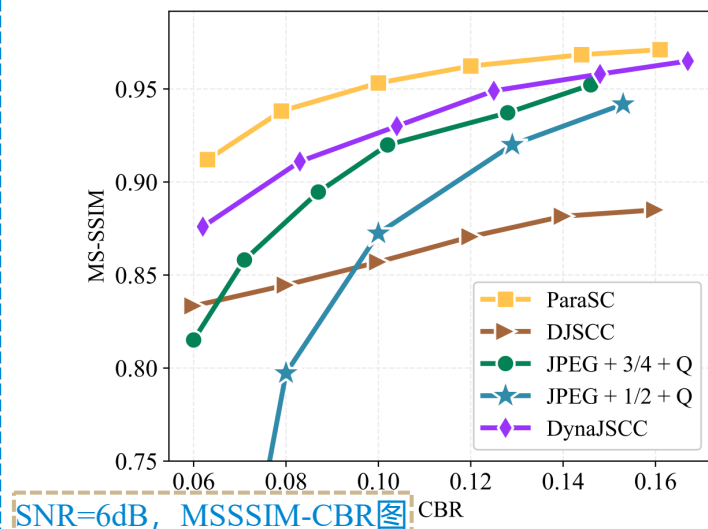
- 数据集: Kodak24 & CVRG-Pano
- 评价指标: PSNR/MS-SSIM/LPIPS
- 基线:
 - 传统分离式信源信道编码方法
 - 基于AI的端到端语义通信方法

结果分析:

- 与传统图像通信方法相比, 所提出的方法成功克服了“悬崖效应”, 在动态信道环境下性能稳健
- 与AI驱动的方法相比, 所提出的方法在多数信道、带宽情况下性能更为优越

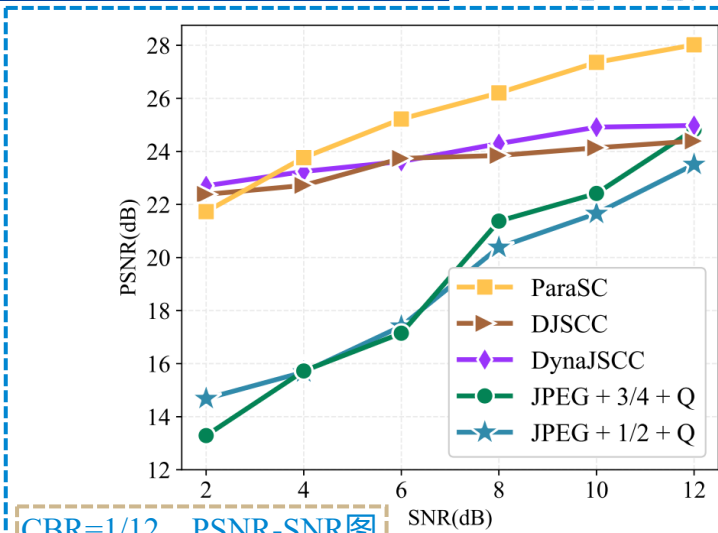


CBR=1/10, PSNR-SNR图

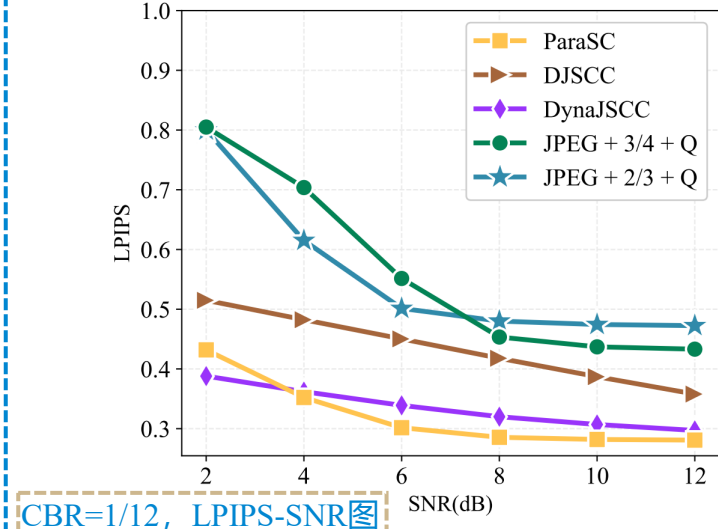


SNR=6dB, MSSSIM-CBR图

图: AWGN信道结果对比



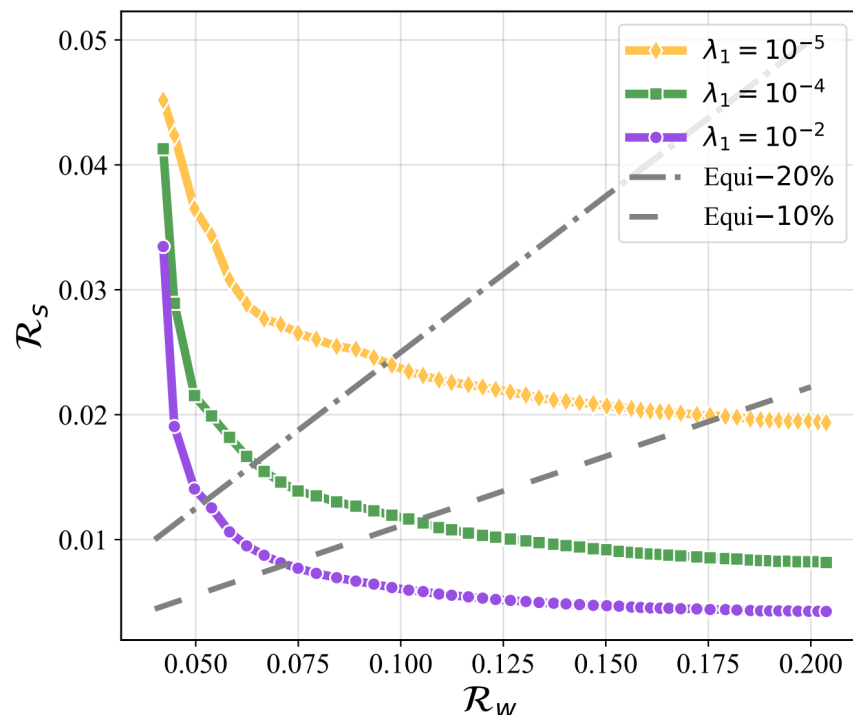
CBR=1/12, PSNR-SNR图



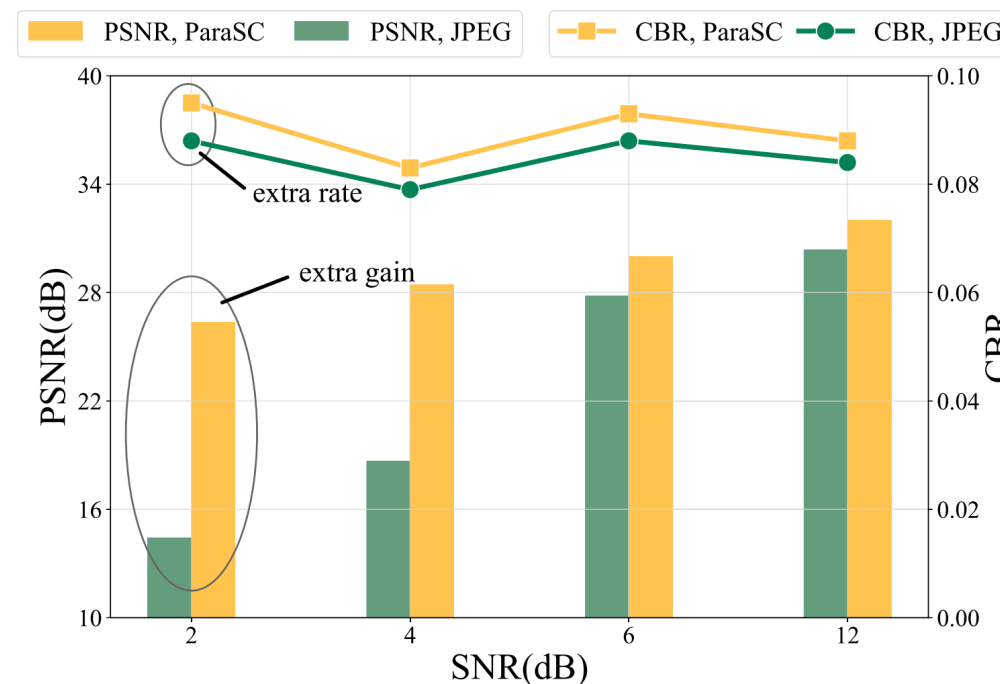
CBR=1/12, LPIPS-SNR图

图: Rayleigh衰落信道结果对比

研究内容 II：语义流消融实验



图：语义流-图像流权衡关系



图：语义分支消融

结果分析：

- **轻量语义开销：**对多数情况下的图像流带宽，语义带宽占比将低于总带宽的**10%**
- **性能增益：**低SNR下，以**8%**的额外语义带宽换取**82%**的PSNR增益



上海交通大学

SHANGHAI JIAO TONG UNIVERSITY

PART FOUR

工作总结



1、提出模块化语义通信框架（SCSC）通过可解释的预处理增强网络（PPEN）与预编码增强网络（PCEN）实现语义信息的高效传输。并且研究了代理网络突破非可微模块的端到端训练瓶颈，验证了系统在动态信道环境下的强泛化能力。

2、并行流语义通信框架（ParaSC）围绕“传统主导+语义辅助”设计，基于传统SSCC编码条件生成可变速率语义流，辅助传统架构比特流同时传输，并在接收机采用并行聚合网络（PAGNet）自适应融合双流信息，使系统在维持语义轻量化的同时对信道波动呈现良好鲁棒性。

3、具体相关成果如下：

[1] X. Han, Y. Wu, Z. Gao, B. Feng, Y. Shi, D. Gündüz, and W. Zhang, “SCSC: A novel standards-compatible semantic communication framework for image transmission,” Early Access, IEEE Trans. Commun., Jan. 2025.

[2] M. Xu, Y. Wu, Y. Shi, X.-G. Xia, M. Debbah, W. Zhang, P. Zhang, “Semantic-aided Parallel Image Transmission Compatible with Practical System”, Early Access, IEEE Trans. Wireless Commun., May. 2025.



- [1] P. Zhang, W. Xu, H. Gao, K. Niu, X. Xu, X. Qin, C. Yuan, Z. Qin, H. Zhao, J. Wei, and F. Zhang, “Toward wisdom-evolutionary and primitive-concise 6G: A new paradigm of semantic communication networks,” *Engineering*, vol. 8, pp. 60–73, Jan. 2022.
- [2] T. -Y. Tung and D. Gündüz, “Deep joint source-channel and encryption coding: Secure semantic communications,” in *IEEE Int. Conf. Commun. (ICC)*, Rome, Italy, May 2023, pp. 5620–5625.
- [3] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, “Semantic communications: Principles and challenges,” *arXiv:2201.01389*, 2021.
- [4] M. Jankowski, D. Gündüz and K. Mikolajczyk, “Wireless image retrieval at the edge,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 89–100, Jan. 2021.
- [5] S. Orhan and Y. Bastanlar, “Semantic segmentation of outdoor panoramic images,” *Signal Image Video Process.*, vol. 16, pp. 643–650, Aug. 2021.
- [6] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. Med. Image Comput.-Assisted Intervention (MICCAI)*, Munich, Germany, Oct. 2015, pp. 234–241.
- [7] R. Franzen. Kodak24 dataset. (1993). [Online]. Available: [http: //r0k.us/graphics/kodak/](http://r0k.us/graphics/kodak/)



谢谢！

