

WIRELESS SEMANTIC COMMUNICATIONS

CONCEPTS, PRINCIPLES, AND CHALLENGES

EDITED BY

YAO SUN | LAN ZHANG

DUSIT NIYATO | MUHAMMAD ALI IMRAN



WILEY

Wireless Semantic Communications

Wireless Semantic Communications

Concepts, Principles, and Challenges

Edited by

Yao Sun

University of Glasgow
Glasgow, UK

Lan Zhang

Michigan Technological University
Houghton, USA

Dusit Niyato

Nanyang Technological University
Singapore

Muhammad Ali Imran

University of Glasgow
Glasgow, UK

WILEY

This edition first published 2025
© 2025 John Wiley & Sons Ltd.

All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

The right of Yao Sun, Lan Zhang, Dusit Niyato and Muhammad Ali Imran to be identified as the authors of the editorial material in this work has been asserted in accordance with law.

Registered Offices

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty

While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data applied for:

Hardback: 9781394223305

Cover Design: Wiley

Cover Image: © Yuichiro Chino/Getty Images

Set in 9.5/12.5pt STIXTwoText by Straive, Chennai, India

Contents

List of Contributions *xiii*

Preface *xvii*

1	Intelligent Transceiver Design for Semantic Communication	1
	<i>Yiwen Wang, Yijie Mao, and Zhaohui Yang</i>	
1.1	Knowledge Base	1
1.1.1	Basic Structures of the Knowledge Base	1
1.1.2	KB-Assisted Single-Modal SC	1
1.1.3	KB-Assisted Multimodal SC	2
1.2	Source and Channel Coding	4
1.2.1	Preliminaries of Source and Channel Encoders/Decoders	5
1.2.2	Different Transceiver Design in SC	5
1.3	Multiuser SC	7
1.3.1	Interference Management Approaches in SC	7
1.3.2	Different MA-Assisted SC	7
1.3.2.1	OFDMA-Assisted SC	8
1.3.2.2	SDMA-Assisted SC	8
1.3.2.3	NOMA-Assisted SC	8
1.3.2.4	RSMA-Assisted SC	10
1.3.2.5	MDMA-Assisted SC	11
1.3.2.6	DeepMA-Assisted SC	12
1.4	Transceiver Design for Single-Modal and Multimodal Data	13
1.4.1	Text	13
1.4.2	Image	14
1.4.3	Speech	14
1.4.4	Video	14
1.4.5	Multimodal	15
1.5	Challenges and Future Directions	16
1.5.1	Internet of Things System	16
1.5.2	Unmanned Aerial Vehicles System	16
1.5.3	Holographic Telepresence	17
1.5.4	Intelligently Transport System	17
1.5.5	Smart Cities	17
	References	18

2	Joint Cell Association and Spectrum Allocation in Semantic Communication Networks	23
	<i>Le Xia, Yao Sun, and Muhammad Ali Imran</i>	
2.1	Introduction	23
2.2	Semantic Communication Model	26
2.2.1	Background Knowledge Matching Model	26
2.2.2	Semantic Channel Model in the PKM-Based SC-Net	28
2.2.3	Semantic Channel Model in the IKM-Based SC-Net	30
2.2.4	Basic Network Topology of SC-Net	31
2.3	Optimal CA and SA Solution in the PKM-Based SC-Net	32
2.3.1	Problem Formulation	32
2.3.2	Optimal Solution for UA	33
2.3.3	Optimization Solution for SA	34
2.4	Optimal CA and SA Solution in the IKM-Based SC-Net	35
2.4.1	Problem Formulation	35
2.4.2	Problem Transformation with Semantic Confidence Level	36
2.4.3	Solution Finalization for CA and SA	37
2.5	Numerical Results and Discussions	38
2.5.1	Performance Evaluations in the PKM-Based SC-Net	39
2.5.2	Performance Testing in the IKM-Based SC-Net	42
2.6	Conclusions	44
	References	45
3	An End-to-End Semantic Communication Framework for Image Transmission	47
	<i>Lei Feng, Yu Zhou, and Wenjing Li</i>	
3.1	Introduction	47
3.1.1	Related Works	48
3.1.2	Contribution	49
3.2	The End-to-End Image Semantic Communication Framework Driven by Knowledge Graph	50
3.2.1	Overview	50
3.2.2	Semantic Encoding	50
3.2.2.1	Entity Detection	52
3.2.2.2	Relationship Detection	53
3.2.3	Semantic Transmission	54
3.2.4	Semantic Decoding	57
3.2.4.1	Entity Mapping	57
3.2.4.2	Relationship Acquisition	57
3.3	Semantic Similarity Measurement	59
3.3.1	Graph-to-Graph Semantic Similarity (GGSS)	59
3.3.1.1	Wasserstein Distance	59
3.3.1.2	Gromov–Wasserstein Distance	60
3.3.2	Image-to-Image Semantic Similarity (IISS)	61
3.4	Simulation	62

3.5	Conclusion	63
	References	64
4	Robust Semantic Communications and Privacy Protection	67
	<i>Xuefei Zhang</i>	
4.1	Motivation and Introduction	67
4.2	Robust Semantic Communication	68
4.2.1	A Robust Semantic Communication Model for Image Transmission	69
4.2.1.1	Generation of Semantic Noise	69
4.2.1.2	Robust Semantic Communication Model	70
4.2.1.3	Performance Analysis	71
4.2.2	A Robust Semantic Communication Model for Speech Transmission	74
4.3	Knowledge Discrepancy-Oriented Privacy Protection for Semantic Communication	75
4.3.1	Knowledge Discrepancy and Privacy Issue	75
4.3.2	Privacy Protection Model Against Knowledge Discrepancy	76
4.3.2.1	Overview	76
4.3.2.2	Knowledge Inference Module	77
4.3.2.3	Path Cutting-Off Module	80
4.3.3	Performance Analysis	81
4.4	Conclusion	84
	References	85
5	Interplay of Semantic Communication and Knowledge Learning	87
	<i>Fei Ni, Bingyan Wang, Rongpeng Li, Zhifeng Zhao, and Honggang Zhang</i>	
5.1	Introduction	87
5.2	Basic Concepts and Related Works	88
5.2.1	Introduction to the KG	88
5.2.2	Knowledge Representation and Reasoning in SemCom	90
5.3	A KG-enhanced SemCom System	91
5.3.1	System Model	91
5.3.2	The Design of the KG-Enhanced SemCom Receiver	93
5.3.3	The Training Methodology	94
5.3.4	Simulation Results	95
5.3.4.1	Dataset and Parameter Settings	95
5.3.4.2	Numerical Results	96
5.4	A KG Evolving-based SemCom System	99
5.4.1	System Overview	99
5.4.2	The Design of the KG Evolving-Based SemCom Receiver	99
5.4.3	The Training Methodology	100
5.4.4	Simulation Results	101
5.4.4.1	Dataset and Parameter Settings	101
5.4.4.2	Numerical Results	103
5.5	LLM-assisted Data Augmentation for the KG Evolving-Based SemCom System	104
5.5.1	Description of the LLM-Assisted Data Augmentation Solution	104

5.5.2	Simulation Results	104
5.5.2.1	Dataset	104
5.5.2.2	Numerical Results	105
5.6	Conclusion	105
	References	106
6	VISTA: A Semantic Communication Approach for Video Transmission	109
	<i>Chengsi Liang, Xiangyi Deng, Yao Sun, Runze Cheng, Le Xia, Dusit Niyato, and Muhammad Ali Imran</i>	
6.1	Introduction	109
6.2	Video Transmission Framework in VISTA	110
6.3	SLG-Based Transceiver Design in VISTA	111
6.3.1	Semantic Segmentation Module	113
6.3.1.1	Object Detection	113
6.3.1.2	Trajectory Prediction	113
6.3.1.3	SLG Construction	114
6.3.1.4	Frame Sampling	114
6.3.2	JSCC Module	114
6.3.3	Frame Interpolation Module	115
6.4	Simulation Results and Discussions	116
6.4.1	Simulation Setting	116
6.4.2	VISTA Framework Performance Evaluation	116
6.5	Conclusions	120
	References	120
7	Content-Aware Robust Semantic Transmission of Images over Wireless Channels with GANs	123
	<i>Xuyang Chen, Daquan Feng, Qi He, Yao Sun, and Xiang-Gen Xia</i>	
7.1	Introduction	123
7.2	System Model	124
7.2.1	Semantic Encoder	125
7.2.2	Quantizer	125
7.2.3	Wireless Channel	126
7.2.4	Semantic Decoder	126
7.3	System Architecture	127
7.4	Experimental Results	127
7.4.1	Datasets and Settings	127
7.4.2	Performance Metrics	128
7.4.3	Results Analysis	128
7.5	Conclusion	130
	References	130
8	Semantic Communication in the Metaverse	133
	<i>Yijing Lin, Zhipeng Gao, Hongyang Du, Jiacheng Wang, and Jiakang Zheng</i>	
8.1	Introduction	133
8.2	Related Work	134

8.2.1	Semantic Communication in Metaverse	135
8.2.2	Authenticity of Semantic Information in Metaverse	135
8.2.3	Efficiency of SemCom in Metaverse	136
8.3	Unified Framework for SemCom in the Metaverse	137
8.3.1	Framework Overview	138
8.3.1.1	Semantic Module	138
8.3.1.2	AIGC Module	139
8.3.1.3	Rendering Module	139
8.3.2	SemCom: Bridging Meaning in the Metaverse	139
8.3.2.1	Integration of SemCom and AIGC	139
8.3.2.2	SemCom and Metaverse	141
8.3.2.3	AIGC and Metaverse	141
8.3.2.4	SemCom, AIGC, and Metaverse	141
8.3.3	Research Gaps	141
8.3.3.1	Dishonest Semantic Transformation	141
8.3.3.2	Inefficient Resource Utilization	142
8.4	Zero-Knowledge Proof-Based Semantic Verification	142
8.4.1	Training-Based Semantic Attack Mechanism	142
8.4.1.1	Extraction	143
8.4.1.2	Loss	143
8.4.1.3	Optimization Strategy	144
8.4.2	ZKP-Based Semantic Verification Mechanism	144
8.4.2.1	Semantic Extraction	145
8.4.2.2	Transformation for Security	145
8.4.2.3	ZKP Extraction for Verifiable Transformation	145
8.4.2.4	Generation of Zero-Knowledge Proof	146
8.4.2.5	Zero-Knowledge Proof Verification	146
8.4.3	Security Analysis	146
8.4.3.1	Completeness	146
8.4.3.2	Soundness	147
8.4.3.3	Zero Knowledge	147
8.5	Diffusion Model-Based Resource Allocation	147
8.5.1	Joint Optimization Problem Formulation	147
8.5.1.1	Semantic Extraction	149
8.5.1.2	AIGC Inference	149
8.5.1.3	Graphics Rendering	149
8.5.1.4	Formulation	150
8.5.2	Strategic Joint Resource Allocation	150
8.5.2.1	Diffusion Model	150
8.5.2.2	Resource Allocation	151
8.6	Simulation Results	152
8.6.1	Authenticity	152
8.6.2	Efficiency	154
8.7	Future Directions	155
8.7.1	Incentives for Semantic Information	156
8.7.2	Privacy Preserving for SemCom in the Metaverse	156

8.7.3	Multitask Semantic Extraction Across Virtual Worlds	156
8.7.4	Lightweight Semantic Communication Before Transmission	157
8.7.5	Challenges in Practical Deployment	157
8.8	Conclusion	157
	References	158
9	Large Language Model-Assisted Semantic Communication Systems	163
	<i>Shuaishuai Guo, Yanhu Wang, Biqian Feng, and Chenyuan Feng</i>	
9.1	Introduction	163
9.2	SSSC Using Pretrained LLMs	165
9.2.1	System Model	165
9.2.2	Problem Formalization	166
9.2.3	Problem Transformation	167
9.2.4	Projected Gradient Descent Optimization	168
9.2.5	Convergence and Computational Complexity Analysis	169
9.2.6	Experimental Results	170
9.3	SIAC Using Pretrained LLMs	171
9.3.1	System Model	171
9.3.2	Semantic Importance Quantification Using Generative LLMs like ChatGPT	173
9.3.3	Semantic Importance Quantification Using Discriminative LLMs like BERT	174
9.3.4	Overhead Analysis	174
9.3.4.1	Communication Content Is Generated by Generative LLMs like ChatGPT	174
9.3.4.2	Communication Content Has Already Been Pre-generated by Other Sources	174
9.3.4.3	Communication Content Is Generated by Other Sources in Real-Time	175
9.3.5	Semantic Importance-aware Power Allocation	175
9.3.6	Experimental Results	176
9.4	Future Direction of Using LLMs: Semantic Correction	178
9.4.1	ChatGPT-powered Semantic Correction	179
9.4.2	BERT-Powered Semantic Correction	179
9.5	Conclusion	180
	Acronyms	180
	References	181
10	RIS-Enhanced Semantic Communication	183
	<i>Bohao Wang, Ruopeng Xu, Zhaohui Yang, and Chongwen Huang</i>	
10.1	RIS-Empowered Communications	183
10.1.1	What Is RIS?	183
10.1.2	RIS-Assisted Communication System	183
10.2	Beamforming Design for RISs Enhanced Semantic Communications	184
10.2.1	System Model	184
10.2.2	Problem Formulation	186
10.2.3	Algorithm Design	187
10.2.4	Simulation Results	188
10.3	Privacy Protection in RIS-Assisted Semantic Communication System	189
10.4	AI for RIS-Assisted Semantic Communications	191

10.4.1	RIS-Based On-the-Air DDNN Semantic Communications	191
10.4.2	Encoding and Decoding of RIS-aided Semantic Communications	192
10.5	Conclusion	195
	Acronyms	195
	References	196
	Index	199

List of Contributions

Xuyang Chen

College of Electronics and Information
Engineering
Shenzhen University
Shenzhen
China

Runze Cheng

James Watt School of Engineering
University of Glasgow
Glasgow
Lanarkshire
UK

Xiangyi Deng

Glasgow College
University of Electronic Science and
Technology of China
Chengdu
Sichuan
China

Hongyang Du

School of Computer Science and Engineering
Nanyang Technological University
Singapore

Daquan Feng

College of Electronics and Information
Engineering
Shenzhen University
Shenzhen
China

Lei Feng

The State Key Laboratory of Networking and
Switching Technology
Beijing University of Posts and
Telecommunications
Beijing
China

Biqian Feng

Department of Electronic Engineering
Shanghai Jiao Tong University
China

Chenyuan Feng

Eurecom
France

Zhipeng Gao

State Key Laboratory of Networking and
Switching Technology
Beijing University of Posts and
Telecommunications
Beijing
China

Shuaishuai Guo

School of Control Science and Engineering
Shandong University
Jinan
China

and

Shandong Key Laboratory of Wireless
Communication Technologies
Shandong University
Jinan
China

Qi He

National Key Laboratory of Science and
Technology on Communications
University of Electronic Science and
Technology of China
Chengdu
China

Chongwen Huang

College of Information Science and Electronic
Engineering
Zhejiang University
Hangzhou
Xihu Region
China

Muhammad Ali Imran

James Watt School of Engineering
University of Glasgow
Glasgow
Lanarkshire
UK

Rongpeng Li

College of Information Science and Electronic
Engineering
Zhejiang University
Hangzhou
China

Wenjing Li

The State Key Laboratory of Networking and
Switching Technology
Beijing University of Posts and
Telecommunications
Beijing
China

Chengsi Liang

James Watt School of Engineering
University of Glasgow
Glasgow
Lanarkshire
UK

Yijing Lin

State Key Laboratory of Networking and
Switching Technology
Beijing University of Posts and
Telecommunications
Beijing
China

Yijie Mao

School of Information Science and Technology
ShanghaiTech University
Shanghai
China

Fei Ni

College of Information Science and Electronic
Engineering
Zhejiang University
Hangzhou
China

Dusit Niyato

School of Computer Science and Engineering
Nanyang Technological University
Singapore

Yao Sun

James Watt School of Engineering
University of Glasgow
Glasgow
Lanarkshire
UK

Bingyan Wang

College of Information Science and Electronic Engineering
Zhejiang University
Hangzhou
China

Bohao Wang

College of Information Science and Electronic Engineering
Zhejiang University
Hangzhou
Xihu Region
China

Jiacheng Wang

School of Computer Science and Engineering
Nanyang Technological University
Singapore

Yanhu Wang

School of Control Science and Engineering
Shandong University
Jinan
China

and

Shandong Key Laboratory of Wireless Communication Technologies
Shandong University
Jinan
China

Yiwen Wang

School of Information Science and Technology
ShanghaiTech University
Shanghai
China

Le Xia

James Watt School of Engineering
University of Glasgow
Glasgow
Lanarkshire
UK

Xiang-Gen Xia

Department of Electrical and Computer Engineering
University of Delaware
Newark, DE
USA

Ruopeng Xu

College of Information Science and Electronic Engineering
Zhejiang University
Hangzhou
Xihu Region
China

Zhaohui Yang

College of Information Science & Electronic Engineering
Zhejiang University
Hangzhou
Xihu Region
China

Honggang Zhang

College of Information Science and Electronic Engineering
Zhejiang University
Hangzhou
China

and

Zhejiang Lab
Hangzhou
China

Xuefei Zhang

School of Information and Communication Engineering
Beijing University of Posts and Telecommunications
Beijing
China

Zhifeng Zhao

College of Information Science and Electronic
Engineering
Zhejiang University
Hangzhou
China

and

Zhejiang Lab
Hangzhou
China

Jiakang Zheng

School of Electronic and Information
Engineering
Beijing Jiaotong University
Beijing
China

Yu Zhou

The State Key Laboratory of Networking and
Switching Technology
Beijing University of Posts and
Telecommunications
Beijing
China

Preface

Tremendous traffic demands have increased in current wireless networks to accommodate for the upcoming pervasive network intelligence with a variety of advanced smart applications. In response to the ever-increasing data rates along with stringent requirements for low latency and high reliability, it is foreseeable that available communication resources like spectrum or energy will gradually become scarce in the upcoming years. Combined with the almost insurmountable Shannon limit, these destined bottlenecks are, therefore, motivating us to hunt for bold changes in the new design of future networks, i.e., making a paradigm shift from bit-based traditional communication to context-based semantic communication.

The concept of semantic communication was first introduced by Weaver in his landmark paper, which explicitly categorizes communication problems into three levels, including the technical problem at the bit level, the semantic problem at the semantic level, and the effectiveness problem at the information exchange level. Nowadays, the technical problem has been thoroughly investigated in the light of classical Shannon information theory, while the evolution toward semantic communication is just beginning to take shape, with the core focus on meaning delivery rather than traditional bit transmission.

Concretely, semantic communication first refines semantic features and filters out irrelevant content by encoding the semantic information (i.e., semantic encoding) at the source, which can greatly reduce the number of required bits while preserving the original meaning. Then, the powerful semantic decoders are deployed at the destination to accurately recover the source meaning from received bits (i.e., semantic decoding), even if there are intolerable bit errors at the syntactic level. Most importantly, through further leveraging matched background knowledge with respect to the observable messages between source and destination, users can acquire efficient exchanges for the desired information with ultralow semantic ambiguity by transmitting fewer bits.

While semantic communication offers these attractive and valuable benefits, it also faces many challenges. For example, when considering the computing limitation of terminal devices, personalized background knowledge, as well as unstable wireless channel conditions, how to design semantic encoder and decoder should be a challenging issue. Moreover, in the networking layer, it is nontrivial to seek the optimal wireless resource management strategy to optimize its overall network performance in a semantics-aware manner. Therefore, before fully enjoying the superiorities of semantic communications, this book would like to explore the following fundamental issues: (i) How many benefits can be achieved by using semantic communication? (ii) How much cost (mainly consumed resources) is incurred to guarantee the required performance, such as

semantic ambiguity? and (iii) How can we maximize the benefits of semantic communications applied to different wireless networks with constraints of cost?

This book will explore recent advances in the theory and practice of semantic communication. In detail, the book covers the following aspects:

- (1) Principles and fundamentals of semantic communication.
- (2) Transceiver design of semantic communications.
- (3) Resource management in semantic communication networks.
- (4) Semantic communication applications to vertical industries and some typical communication scenarios.

Chapter 1 delves into the transceiver design for semantic communications. Specifically, we first summarize established designs for key components in semantic communications, with a key focus on the knowledge base and semantic encoders/decoders crucial for single-user semantic communications. Our discussion then extends to multiuser SC, specifically emphasizing the synergy between various multiple-access schemes and semantic communications, including orthogonal multiple access (OMA), space-division multiple access (SDMA), non-orthogonal multiple access (NOMA), rate-splitting multiple access (RSMA), and model-division multiple access (MDMA). In the end, we explore various applications of semantic communications and analyze the potential alterations in transceiver design required for these applications.

Chapter 2 studies semantic communications from a networking perspective, particularly focusing on the upper layer. Our primary objective is to investigate optimal wireless resource management strategies within the semantic communications-enabled network (SC-Net) to enhance overall network performance in a semantics-aware manner. This entails addressing the unique challenge of ensuring background knowledge alignment between multiple mobile users (MUs) and multitier base stations (BSs). Efficient resource management remains paramount within the SC-Net, offering numerous benefits such as guaranteeing high-quality SemCom services and enhancing spectrum utilization. By devising effective resource allocation strategies, we aim to optimize network performance and facilitate seamless communication within the SC-Net ecosystem.

Chapter 3 innovatively proposes the transformation theory of the semantic domain–spatial domain, projecting the knowledge graph onto a three-dimensional tensor in the spatial domain. By mapping the entity’s semantic ambiguity with the intensity of discrete points, the knowledge graph is reconstructed from background knowledge libraries and three-dimensional tensors at the receiving end. Additionally, this chapter proposes a graph-to-graph semantic similarity (GGSS) metric based on graph optimal transport theory to evaluate the similarity of semantic information before and after transmission, as well as a semantic-level image-to-image semantic similarity (IISS) metric that aligns with human perception. Finally, we demonstrate the effectiveness and rationality of the framework through simulations.

Chapter 4 first introduces the problem that neural networks in semantic communication are very vulnerable to adversarial attacks, then proposes robust semantic communication systems for image and speech transmission. Meanwhile, this chapter discusses the privacy issue caused by the difference of knowledge bases between transmitter and receiver in semantic communication and proposes a knowledge discrepancy-oriented privacy protection (KDPP) method for semantic communication to reduce the risk of privacy leakage while retaining high data utility.

Chapter 5 investigates the means of knowledge learning in semantic communication with a particular focus on the utilization of Knowledge Graphs (KGs). Specifically, we first review

existing efforts that combine semantic communication with knowledge learning. Subsequently, we introduce a KG-enhanced semantic communication system, wherein the receiver is carefully calibrated to leverage knowledge from its static knowledge base for ameliorating the decoding performance. Contingent upon this framework, we further explore potential approaches that can empower the system to operate in evolving a knowledge base more effectively. Furthermore, we investigate the possibility of integration with large language models (LLMs) for data augmentation, offering additional perspective into the potential implementation means of semantic communication. Extensive numerical results demonstrate that the proposed framework yields superior performance on top of the KG-enhanced decoding and manifests its versatility under different scenarios.

Chapter 6 presents a novel framework called VISTA (VIdeo transmission over Semantic communication Approach) for video transmission by exploiting semantic communications. VISTA comprises three key modules: the semantic segmentation module and the frame interpolation module, responsible for semantic encoding and decoding, respectively, and the joint source-channel coding (JSCC) module, designed for SNR-adaptive wireless transmission.

Chapter 7 proposes a content-aware robust semantic communication framework for image transmission based on generative adversarial networks (GANs). Specifically, the accurate semantics of the image are extracted by the semantic encoder and divided into two parts for different downstream tasks: regions of interest (ROI) and regions of non-interest (RONI). By reducing the quantization accuracy of RONI, the amount of transmitted data volume is reduced significantly. During the transmission process of semantics, a signal-to-noise ratio (SNR) is randomly initialized, enabling the model to learn the average noise distribution. The experimental results demonstrate that by reducing the quantization level of RONI, transmitted data volume is reduced up to 60.53% compared to using globally consistent quantization while maintaining comparable performance to existing methods in downstream semantic segmentation tasks. Moreover, our model exhibits increased robustness with variable SNRs.

Chapter 8 first proposes an integrated framework for bridging meanings of semantic information in the Metaverse to achieve efficient interaction between physical and virtual worlds. This chapter then presents a Zero Knowledge Proof-based verification mechanism to secure the authenticity of the extracted information. This chapter also introduces a diffusion model-based resource allocation mechanism to maximize the utility of resources. Simulation results are presented to validate the authenticity and efficiency of the proposed mechanisms. Additionally, this chapter discusses future directions to further advance SemCom in the metaverse.

Chapter 9 discusses the method of leveraging large language model (LLM) to assist semantic communication systems. Specifically, this chapter first discusses leveraging LLM to define the semantic loss of communications, based on which a signal-shaping method is proposed to minimize the semantic loss for semantic communications with a few message candidates. Then, this proposed a more generalized method to quantify the semantic importance of a word/frame using LLM and investigate semantic importance-aware communications (SIAC) to reliably convey the semantics with limited communication and network resources. Finally, this chapter points out the future direction of using LLM for semantic correction. Experiments are conducted to verify the effectiveness of leveraging LLM to assist semantic communications.

Chapter 10 explores cutting-edge advancements in RIS for semantic communication. It delves into three pivotal areas: optimizing beamforming in RIS-aided systems for enhanced communication in complex digital environments like the Metaverse; employing physical layer strategies

for robust privacy protection in semantic communication systems; and leveraging deep learning for advanced interpreting and prioritizing data and encoding and decoding semantic transmission in wireless communications. These topics collectively highlight the potential of RIS in transforming communication paradigms, emphasizing efficiency, security, and intelligent data processing in semantic communication.

Yao Sun
Lan Zhang
Dusit Niyato
Muhammad Ali Imran
April 2024