CQILAB

KYUNG HEE UNIVERSITY

EDGE DIFFUSION BLOCK :

DYNAMIC MODEL WITH ENHANCED RESOLUTION IN SEMANTIC COMMUNICATION

Future Communication Project I Course

Presenter :

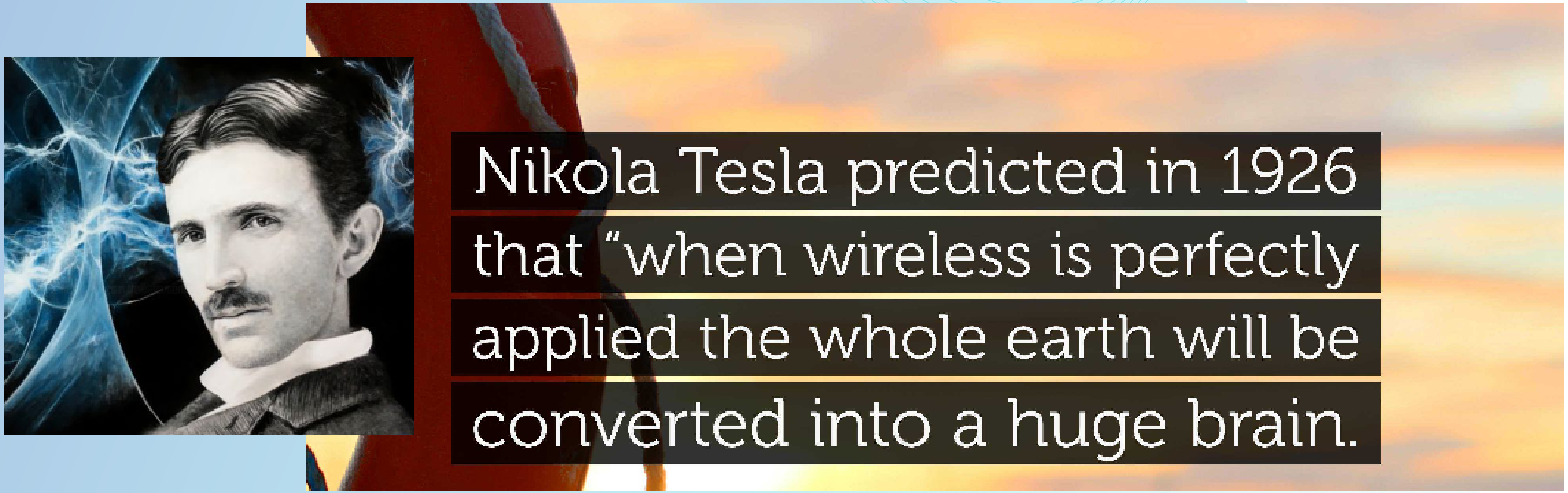
- Fadhel Hariz Dzulfikar (2022310331)
- Arfeto Brian Estadimas (2023310322)
- Roh Jae-Uk (2023315091)

Advisor :

- Professor Hyundong Shin (CQILAB)
- Yeongmin Jeong (Samsung Industry)

PRESENTATION CONTENTS

FUTURE OF CONNECTIVITY: WIRELESS AND BIG AI

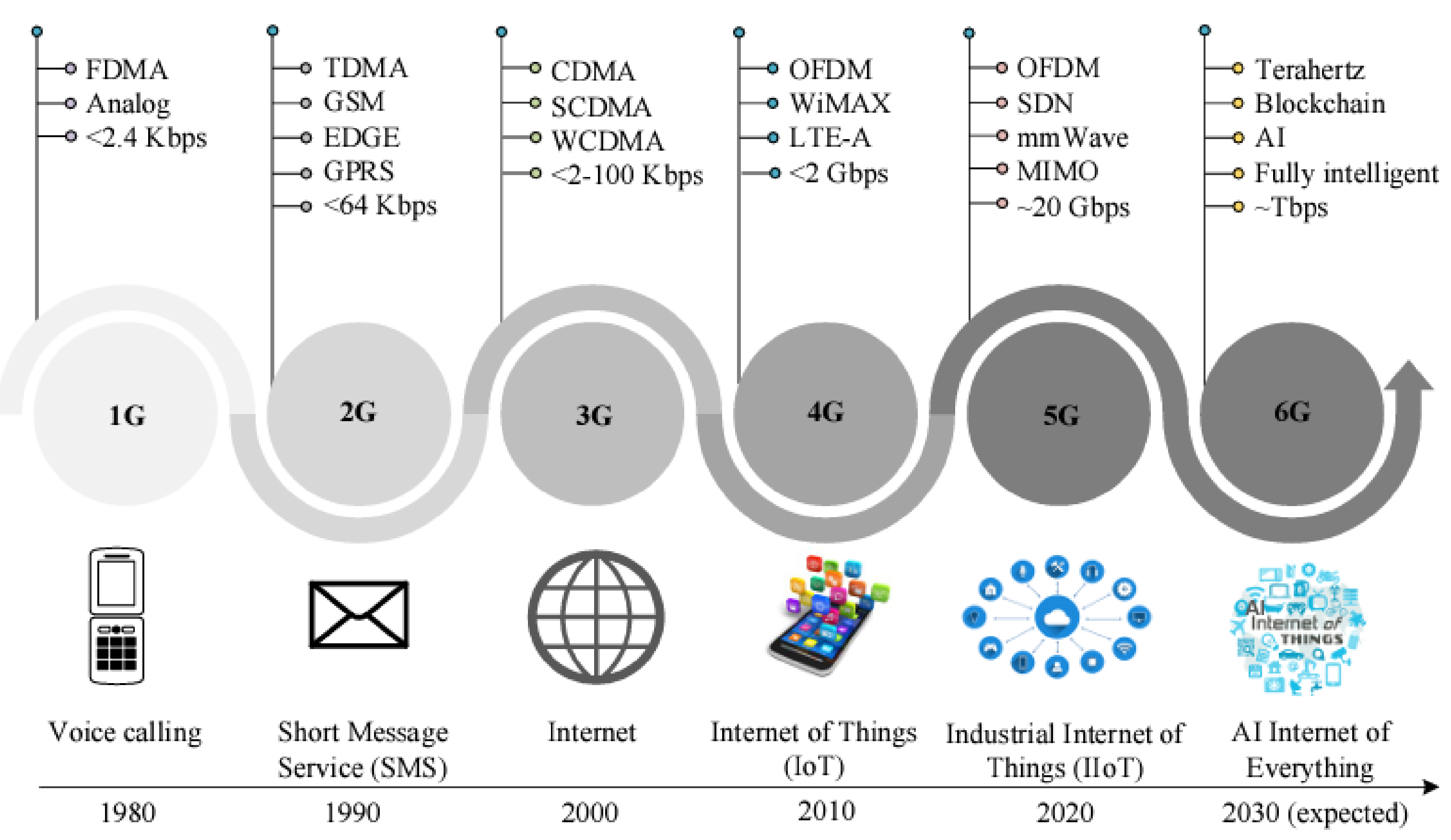


By 2030, As AI and wireless technologies continue to advance, we are witnessing the realization of Nikola Tesla's vision

FUTURE OF CONNECTIVITY: ROAD TO 6G

KPI	5G	6G
Traffic Capacity	10 Mbps/m^2	~ 1-10 Gbps/m^3
Data rate DL	20 Gbps	1 Tbps
Data rate UL	10 Gbps	1 Tbps
Uniform user experience	50 Mbps 2D everywhere	10 Gbps 3D eve
Latency (radio interface)	Up tp 1 msec	Up to 0.1 ms
Jitter	NS	1 µsec
Communication Reliability	Up to 10^{-5}	Up to 10^{-9}
Energy/bit	NS	pJ/bit
Localization precision	10 cm on 2D	1 cm on 3D

In order to meet the demands of the forthcoming 6G era, a revolutionary service paradigm is essential to achieve higher efficiency, unparalleled intelligence, and ultra-low latency. With 6G technology, we will be entering an era where we must seamlessly handle multi-modal data.

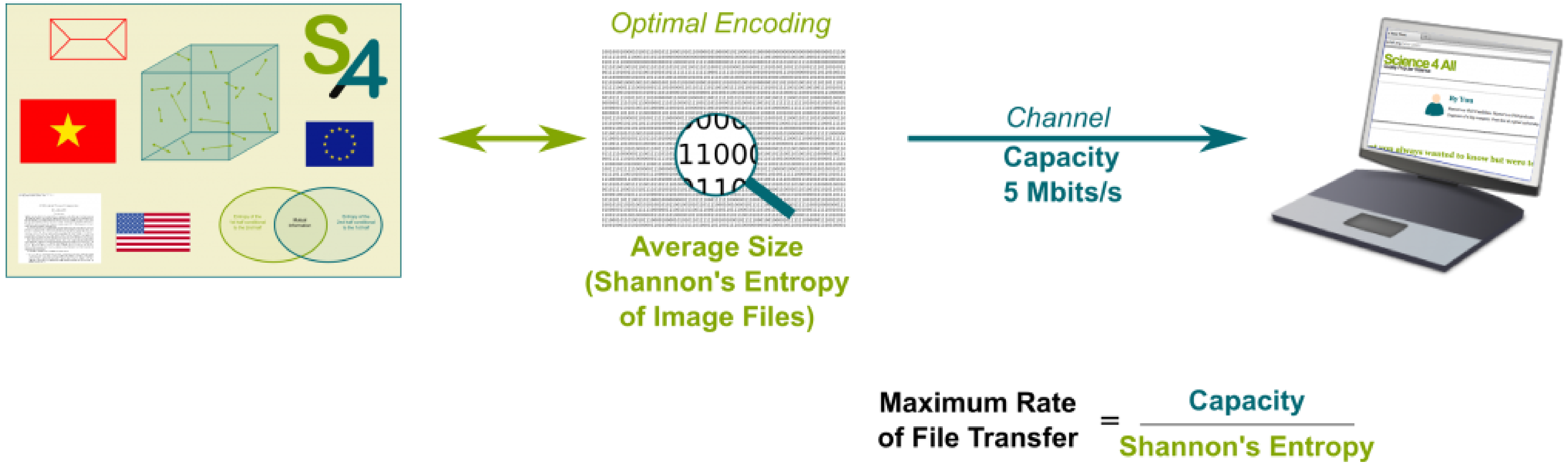


LOOKBACK INTO SHANNON..

Understanding Shannon's Limit and Information Theory

Shannon's mathematical theory of communication is the foundation of today's communications. This framework uses **bits** as the currency to exchange information. The nature of bits that hold only binary value becomes the most reliable theory because it can be translated flexibly to any physical system. Not only that, but Shannon's theory also provides comprehensive tools to meet any previous communication needs.

However, Shannon's Theory primarily focuses on the **quantity of data** transmitted, not the quality or **meaning** of the information..

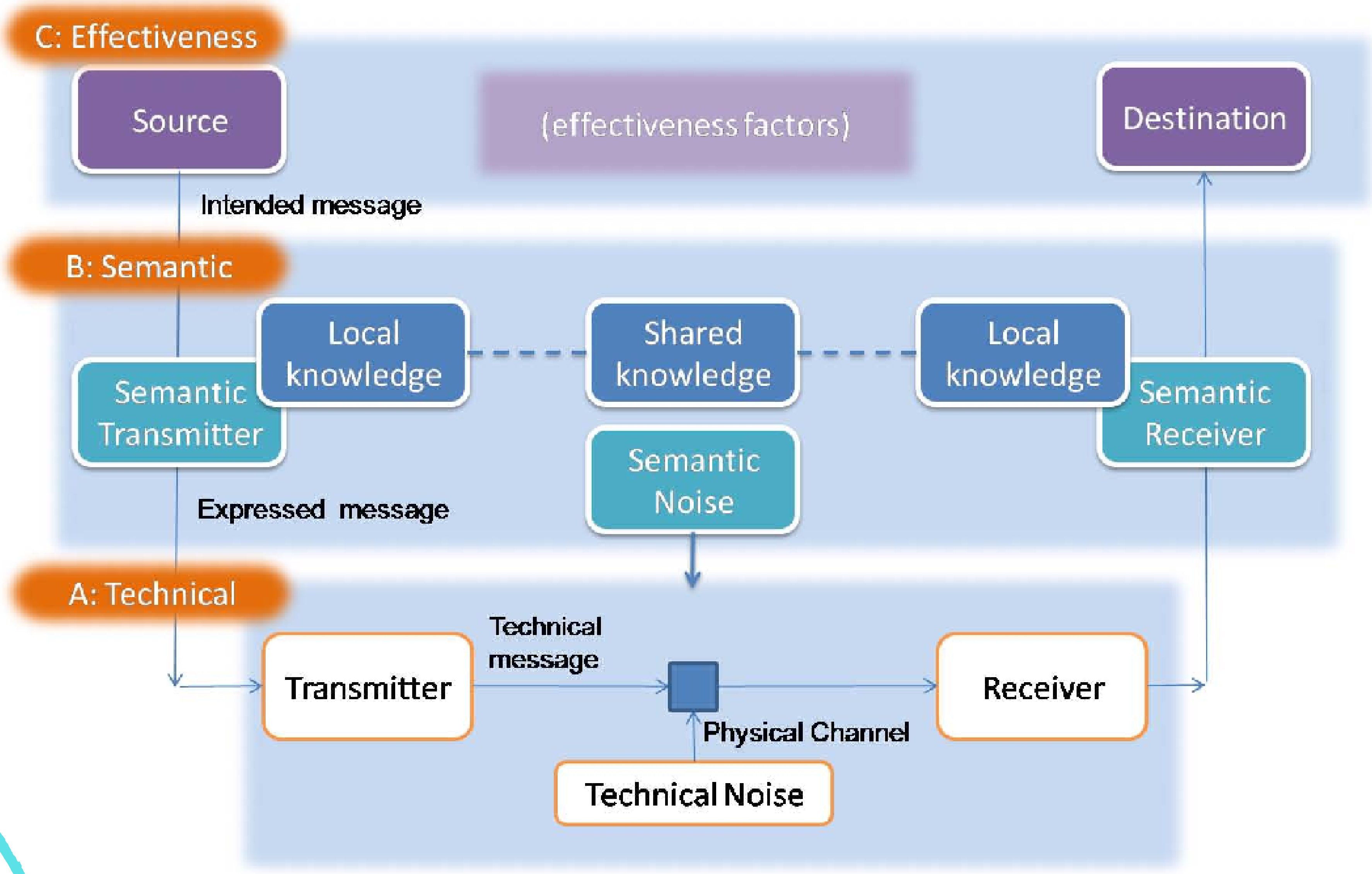


CHALLENGES: LAYERS

A. Technical Layer: How accurately can the symbols of communication be transmitted?

B. Semantic Layer: How precisely do the transmitted symbols convey the desired meaning?

C. Effectiveness Layer: How effectively does the received meaning affect conduct in the desired way?

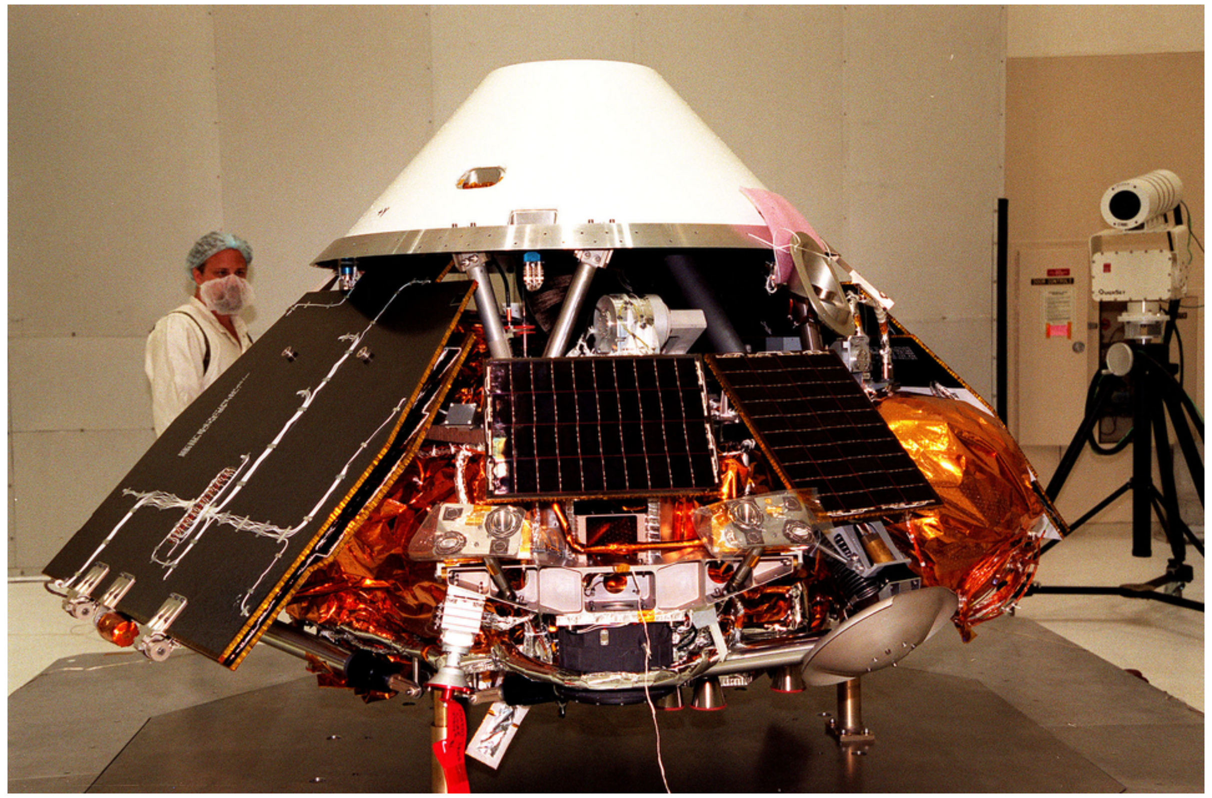


Shannon's theory established a solid and formal framework for addressing technical challenges, which has already demonstrated its proficiency in handling technical aspects.

Shannon, left deliberately aside all aspects related to semantic and effectiveness.

CHALLENGES

Why Semantic Layer of Communication is Important?
Besides Goal Oriented Communications,



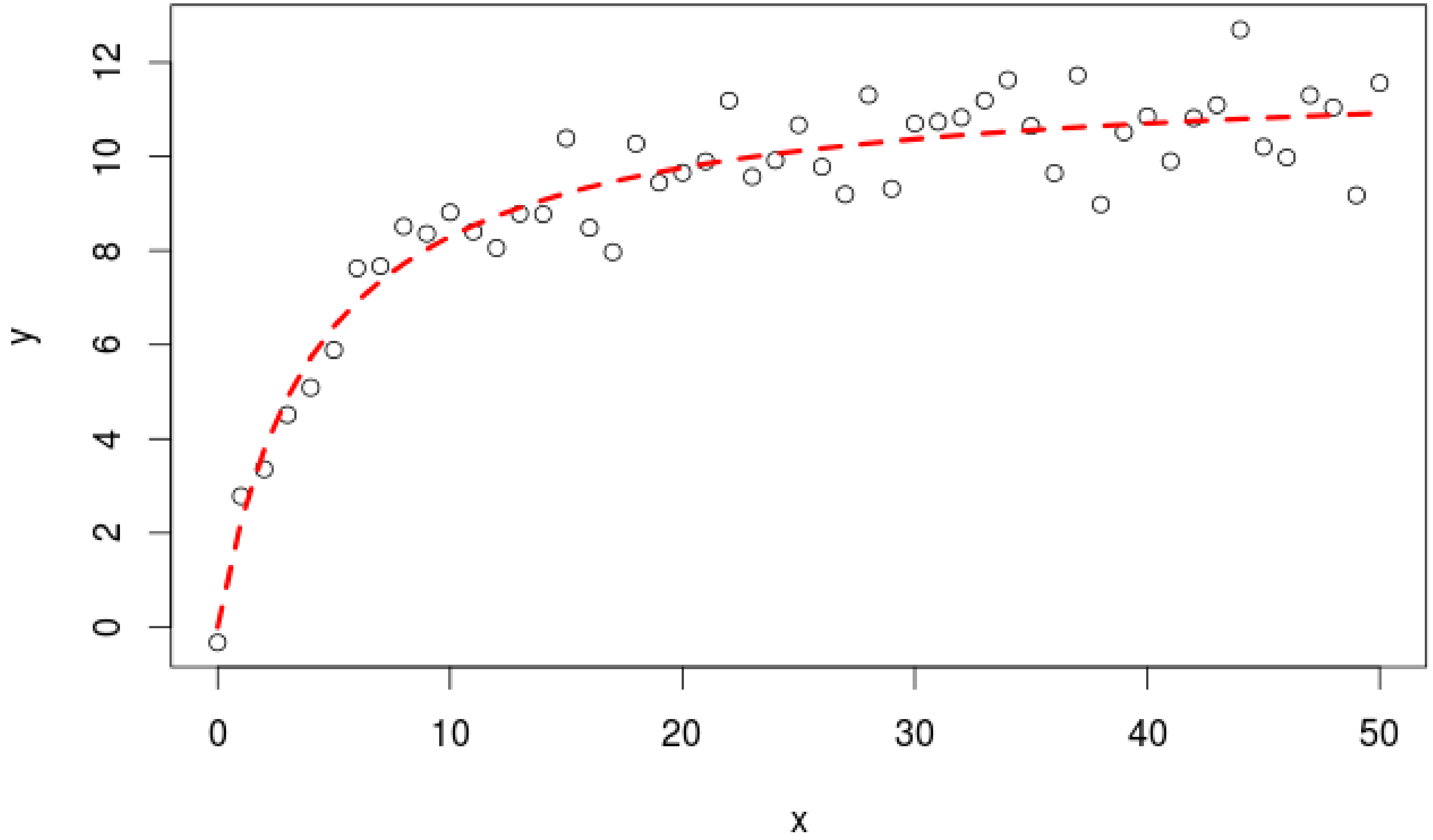
Mars Climate Orbited (1998-1999): mission failure costed 125 M$

Semantic Failure : Expressed Pounds (LbF), Interpreted Newton

Ariane 5 rocket (1996): mission failure costed $500 M

Expressed: 64-bit floating point number. Interpreted: 16-bit signed integer

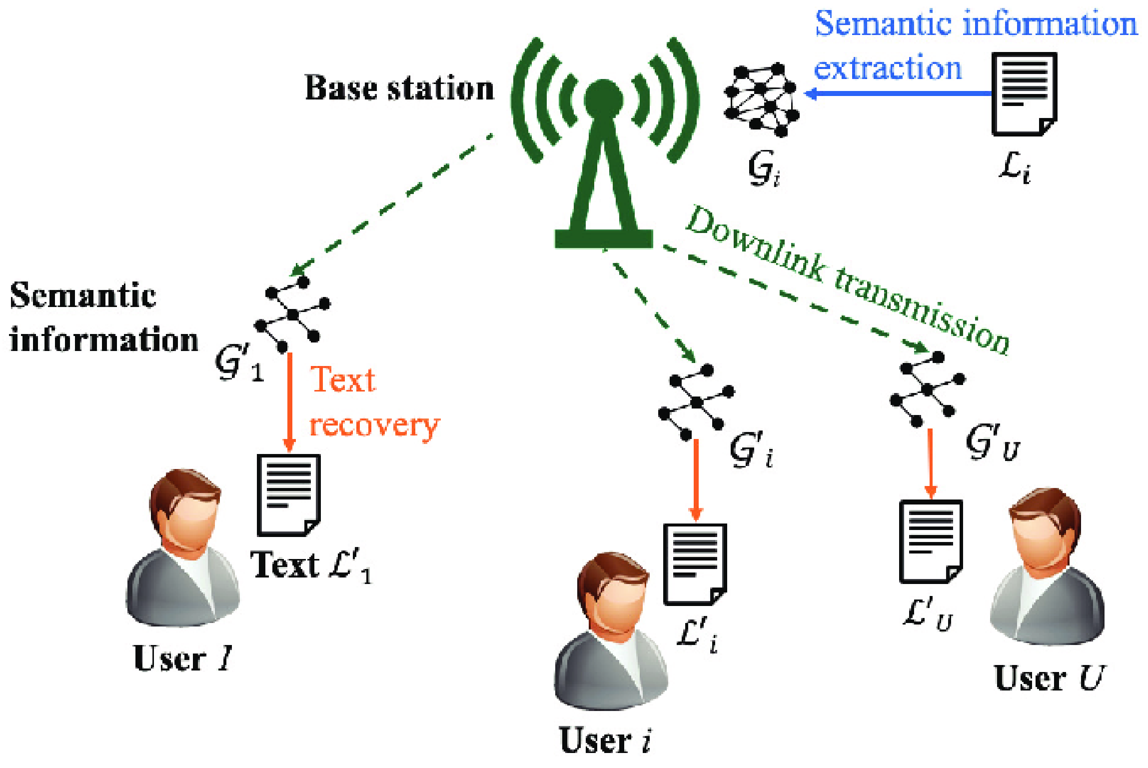
GOAL ORIENTED COMMUNICATIONS



- Goal: Learn a set of parameters θ based on observations x_i
- The observations X={x_i}-from i to N- are the outcome of a vector random variable with a probability density function p(x; θ)with vector parameter θ.
- T(X) is a function of X, such as the accuracy is the estimation of θ from T(X) is the same as that achievable using the observation X directly.

Goal Oriented &

SEMANTIC COMMUNICATION



What is the advantage?

The entropy of T(X) can be much smaller than the entropy of X

The number of bits necessary to encode T(X) may be much smaller than the number of bits to be used to encode X

The number of bits to be transmitted can be significantly decreased, with no losses in terms of inference

In general:

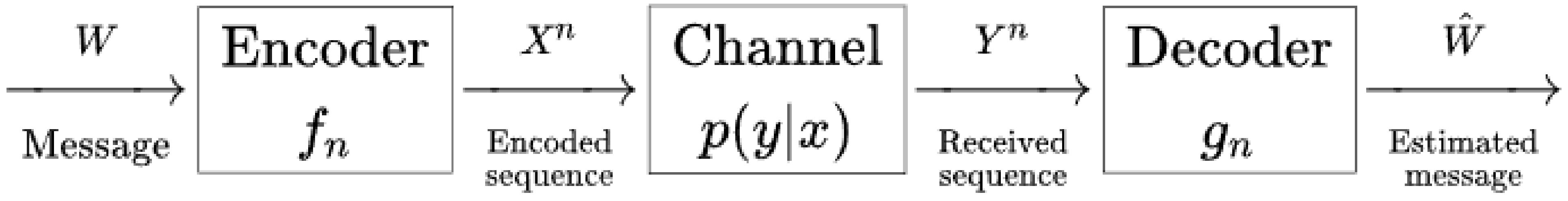
If the goal of communication is to perform some inference on the data, it is not necessary to transmit the data as they are, but it is more convenient to transmit a function of the data, which depends on the goal of the inference.

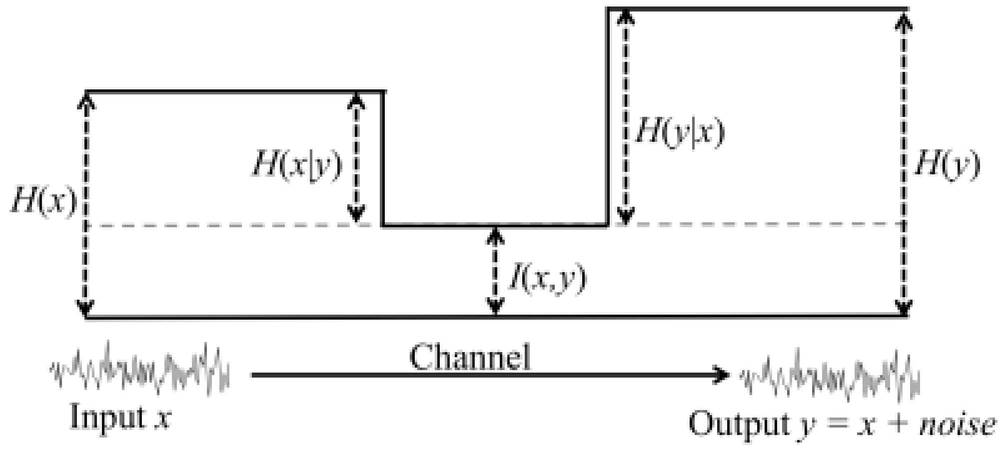
We send only what is **really relevant for the action** to be performed at the receiver side .

Data can be **significantly compacted** while leaving unaltered the performance of the inference method

Semantic Communication:

MAXIMIZING CHANNEL CAPACITY





- For such channel mutual information I(X;Y) denotes the amount of information that one random variable contain about the other random variable.

$$I(\mathbf{X};\mathbf{Y}) = h(\mathbf{X}) - h(\mathbf{Y}|\mathbf{X})$$

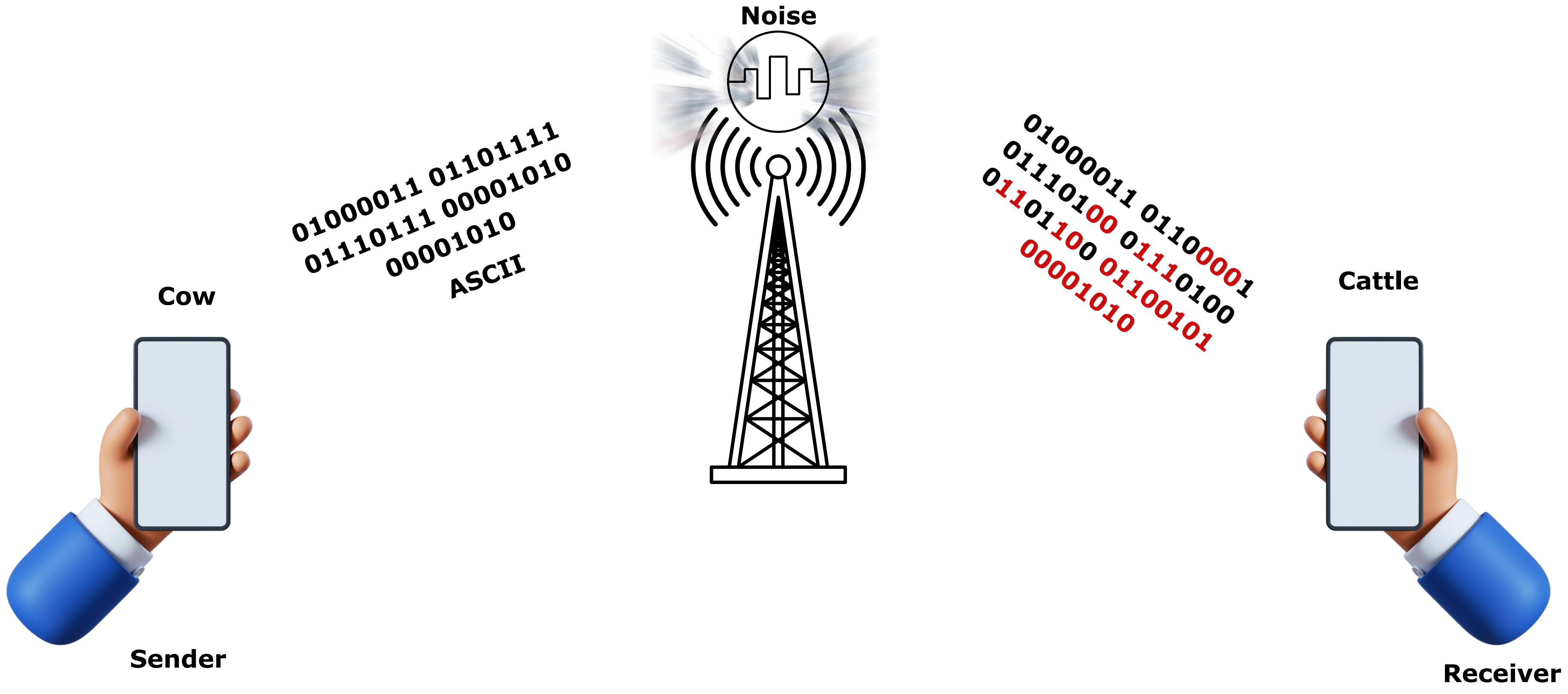
- The information capacity C is obtained by maximizing this mutual information taken over all possible input distribution P(x).

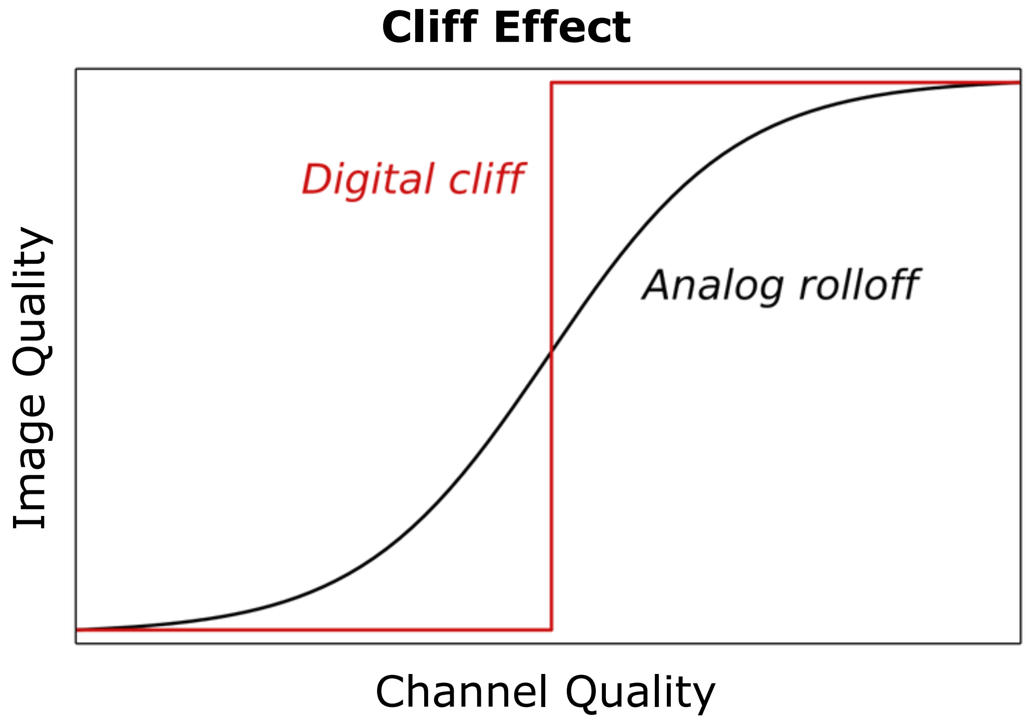
$$C = \max_{\{p(x_i)\}} I(\boldsymbol{X}, \boldsymbol{Y})$$

By incorporating semantic and goal-oriented aspects, future 6G networks can identify and transmit only the relevant information, i.e., the information that is strictly necessary to convey the meaning or accomplish the goal. This can reduce the communication overhead, latency, and energy consumption, as well as improve the quality of experience.

Goal Oriented & Semantic Communications

PRICE TO PAY





Shared KB

A shared knowledge base (KB) is a repository of information that can be accessed to interpret the information at decoder side, however it may require hardcoding.

Channel Noise

Causes quality degradation and sometimes misleading of interpretation of the message.

Cliff Effect

Phenomenon that occurs in digital communication, where the signal quality suddenly drops to zero when it falls below a certain threshold. Unlike analog communication, where the signal quality gradually degrades as the signal strength decreases

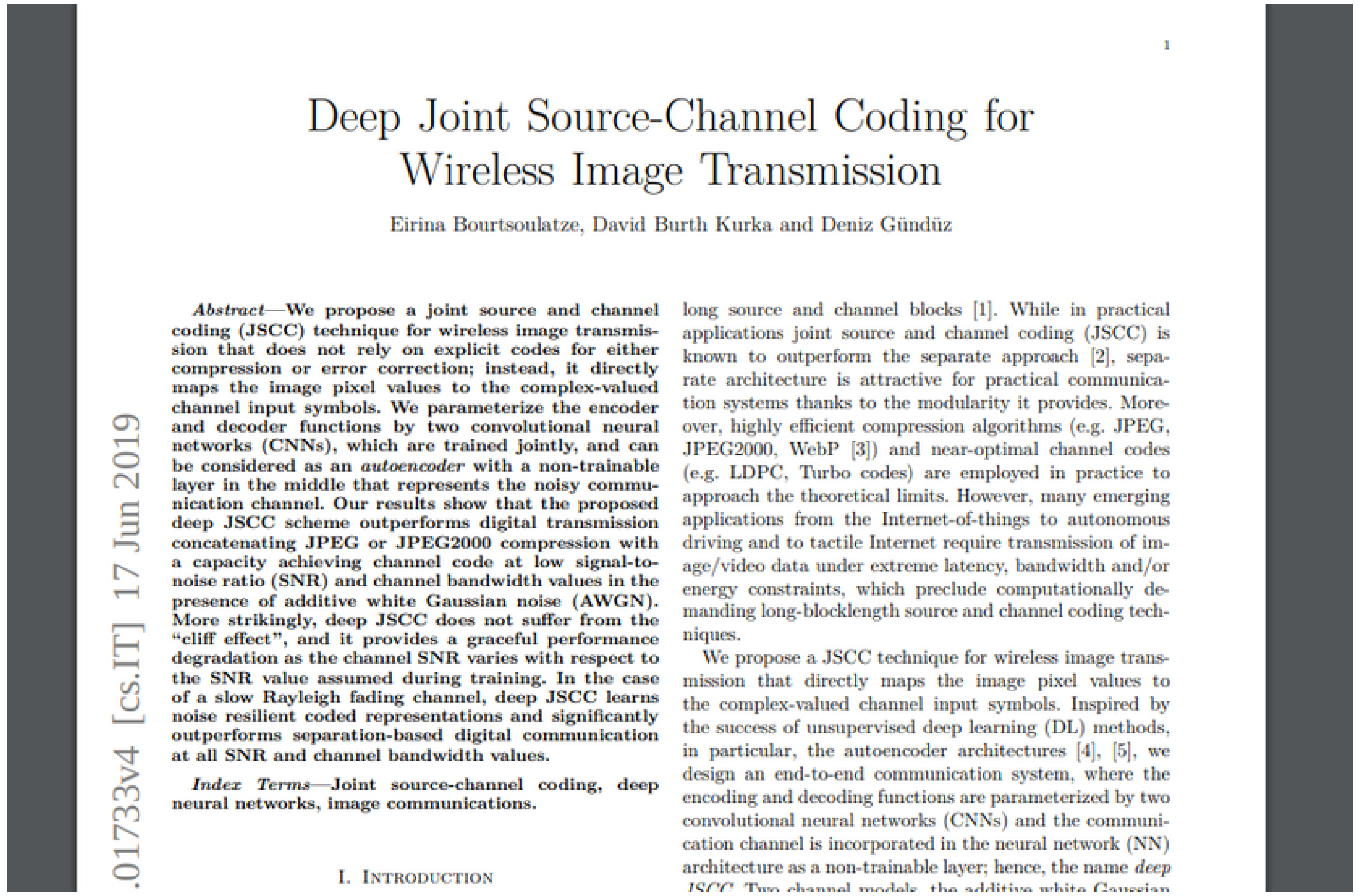
SOLUTION? AI-DRIVEN SC

Zero-shot AI-Driven SC:

JOINT SOURCE CHANNEL CODING (JSCC)

- Deep Joint Source-Channel Coding (Deep JSCC) is a technique that can solve the "cliff effect" problem and be robust to channel noise. Unlike digital communication systems that fail when the channel noise is higher than expected, Deep JSCC can still recover some useful information from the received symbols.
- Deep JSCC architecture is a wireless image transmission method that does not require explicit codes for compression or error correction. The encoder maps the input image directly to channel inputs, and the encoder and decoder functions are modeled as complementary Convolutional Neural Networks (CNNs).

DeepJSCC approach demonstrated improved metrics performance compared to conventional compression techniques such as JPEG and JPEG2000 under noisy channel conditions



1

Deep Joint Source-Channel Coding for Wireless Image Transmission

Eirina Bourtsoulatze, David Burth Kurka and Deniz Gündüz

.01733v4 [cs.IT] 17 Jun 2019

Abstract—We propose a joint source and channel coding (JSCC) technique for wireless image transmission that does not rely on explicit codes for either compression or error correction; instead, it directly maps the image pixel values to the complex-valued channel input symbols. We parameterize the encoder and decoder functions by two convolutional neural networks (CNNs), which are trained jointly, and can be considered as an *autoencoder* with a non-trainable layer in the middle that represents the noisy communication channel. Our results show that the proposed deep JSCC scheme outperforms digital transmission concatenating JPEG or JPEG2000 compression with a capacity achieving channel code at low signal-to-noise ratio (SNR) and channel bandwidth values in the presence of additive white Gaussian noise (AWGN). More strikingly, deep JSCC does not suffer from the "cliff effect", and it provides a graceful performance degradation as the channel SNR varies with respect to the SNR value assumed during training. In the case of a slow Rayleigh fading channel, deep JSCC learns noise resilient coded representations and significantly outperforms separation-based digital communication at all SNR and channel bandwidth values.

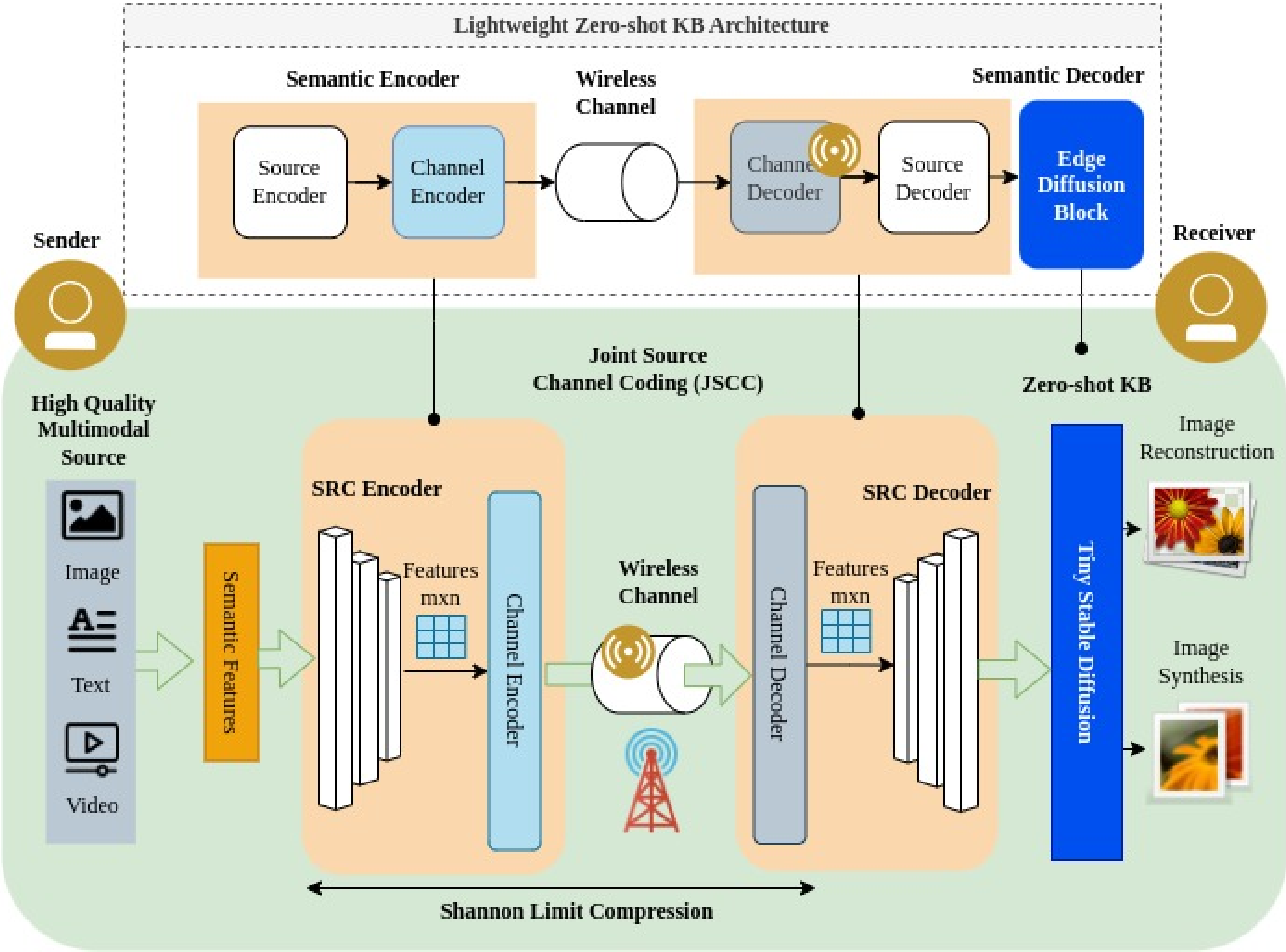
Index Terms—Joint source-channel coding, deep neural networks, image communications.

I. INTRODUCTION

long source and channel blocks [1]. While in practical applications joint source and channel coding (JSCC) is known to outperform the separate approach [2], separate architecture is attractive for practical communication systems thanks to the modularity it provides. Moreover, highly efficient compression algorithms (e.g. JPEG, JPEG2000, WebP [3]) and near-optimal channel codes (e.g. LDPC, Turbo codes) are employed in practice to approach the theoretical limits. However, many emerging applications from the Internet-of-things to autonomous driving and to tactile Internet require transmission of image/video data under extreme latency, bandwidth and/or energy constraints, which preclude computationally demanding long-blocklength source and channel coding techniques.

We propose a JSCC technique for wireless image transmission that directly maps the image pixel values to the complex-valued channel input symbols. Inspired by the success of unsupervised deep learning (DL) methods, in particular, the autoencoder architectures [4], [5], we design an end-to-end communication system, where the encoding and decoding functions are parameterized by two convolutional neural networks (CNNs) and the communication channel is incorporated in the neural network (NN) architecture as a non-trainable layer; hence, the name *deep JSCC*.

System Model:
IMAGE TRANSMISSION WITH JSCC



Problem with DeepJSCC:

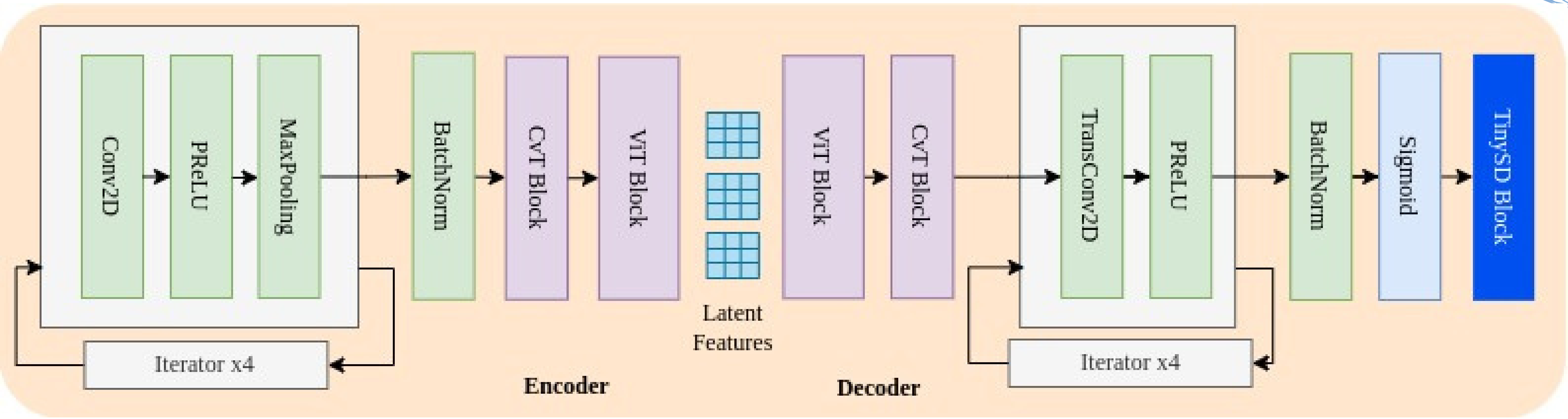
- Overfitting Data (No coherence in zero-shot)
- Low-Resolution Dataset (CIFAR-10), high vulnerability to noise

Therefore,
generative AI model (diffusion) based semantic communication system is proposed for image transmission. The system model is shown in Fig. It can be broken down into three parts.

- The transmitter extract the semantic features of their structured representations from the original image, and then extracts their compressed semantic features
- The channel is used to simulate realistic noise and heavy compression
- Generative AI at the receiver side after DNN-based semantic encoder to reconstruct the original image from the noisy, compressed features.

System Model:

VISION TRANSFORMER & CNN



- We employ a Convolutional Neural Network (CNN) and Vision Transformer (ViT)-based architecture, comprised of an encoder Fenc and a decoder Fenc, dedicated solely to the source-channel coding. The architecture of the model is illustrated in Fig 2. The semantic features, denoted as Xs, are obtained by encoding the input Xin, which could be either the input image, mask, or a batch of masks.

$$x_s = F_{enc}(x_{in}),$$

- The Vision Transformer (ViT) consists of transformer encoder layers with multi-head self-attention mechanisms, employing positional encodings and feedforward neural networks (MLP). Defined:

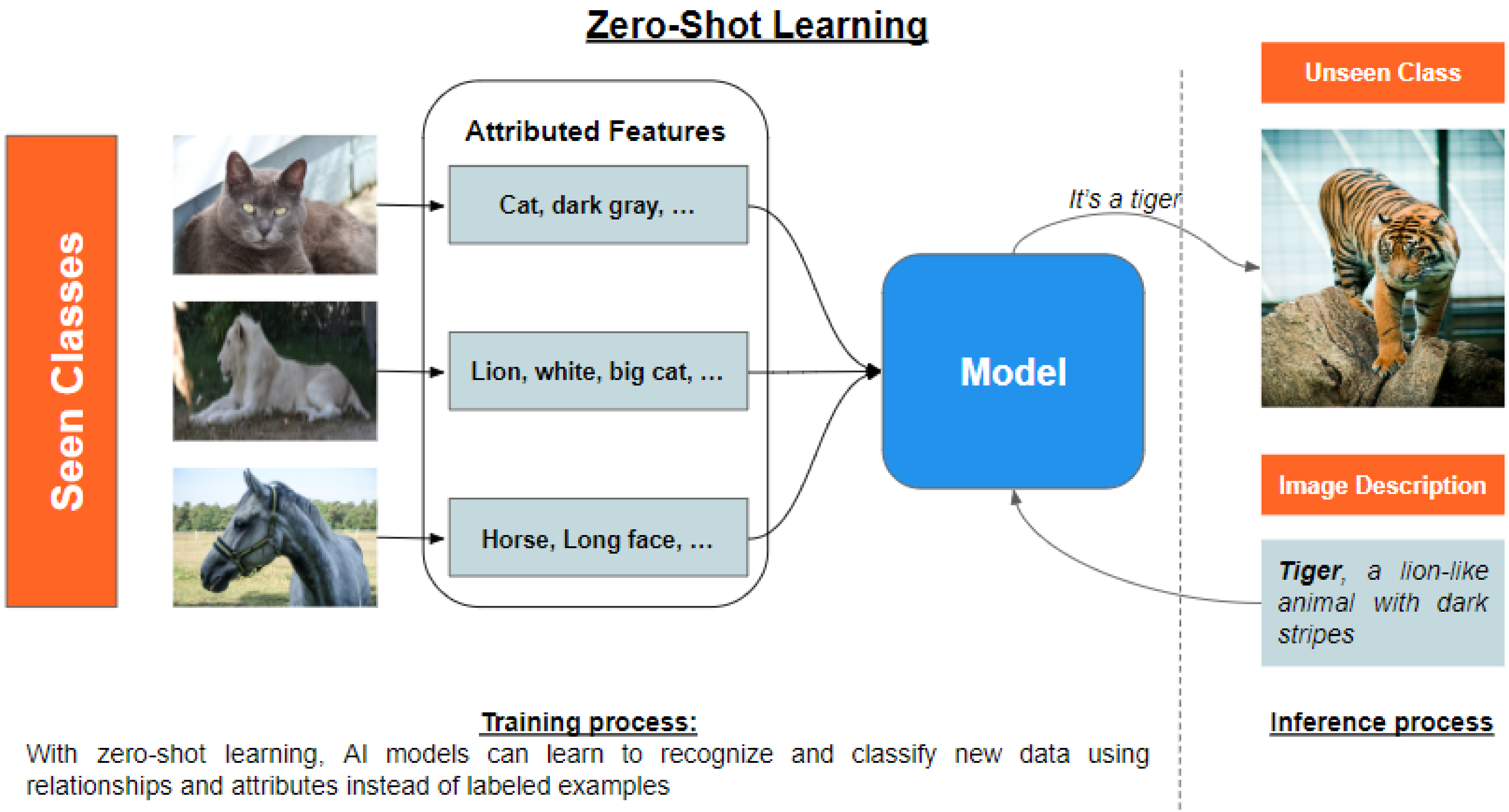
$$CR = \frac{C_o \times H_o \times W_o}{C_{in} \times H_{in} \times W_{in}},$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V$$

$$\text{MLP}(x) = \text{PReLU}(W_2\text{PReLU}(W_1 x + b_1) + b_2)$$

System Model:

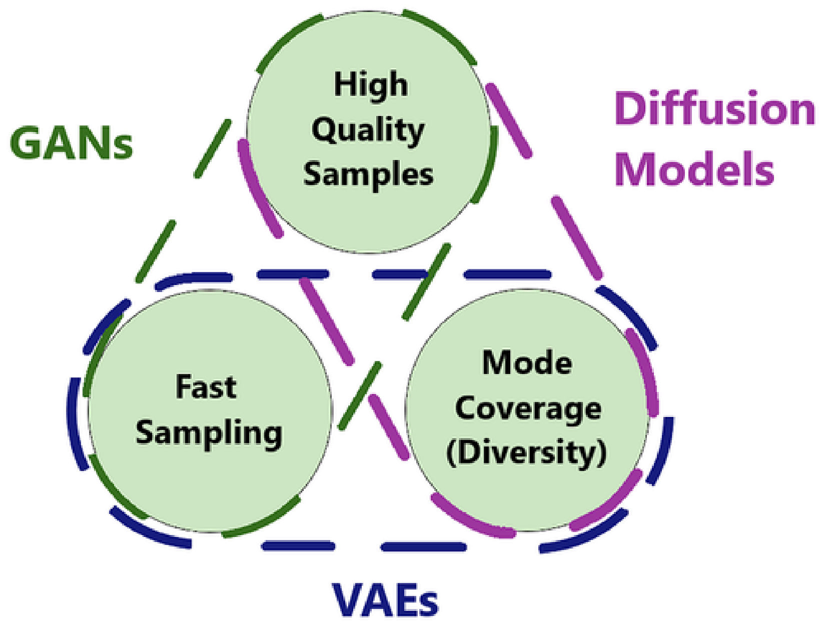
INTRODUCTION TO ZERO-SHOT

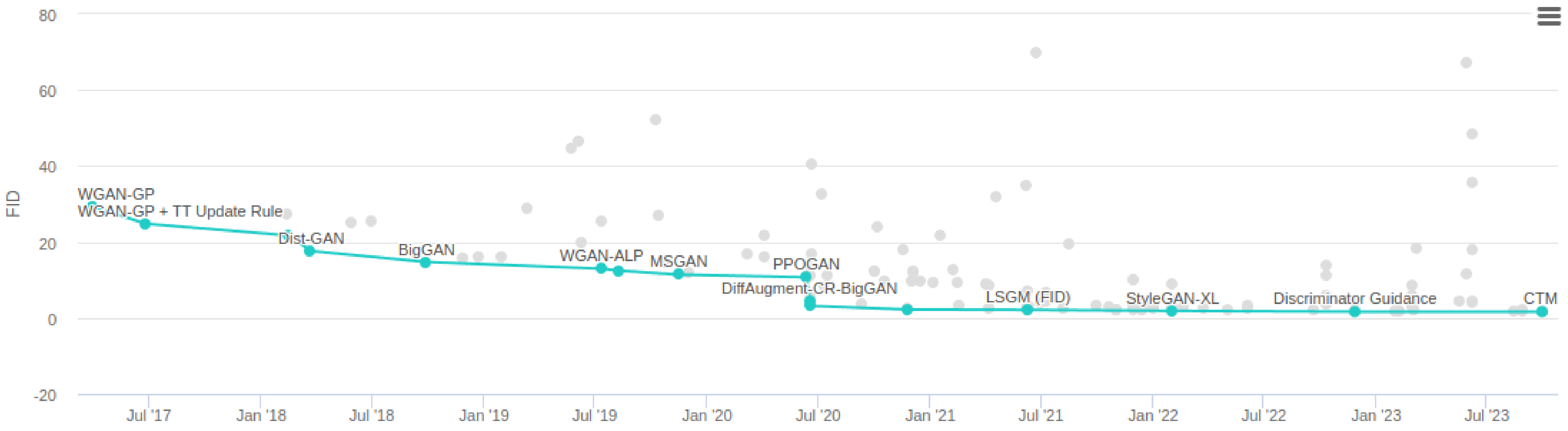


System Model:

INTRODUCTION TO DIFFUSION MODEL

- Diffusion Model is SOTA in Image Generation, beat StyleGAN and BigGAN.

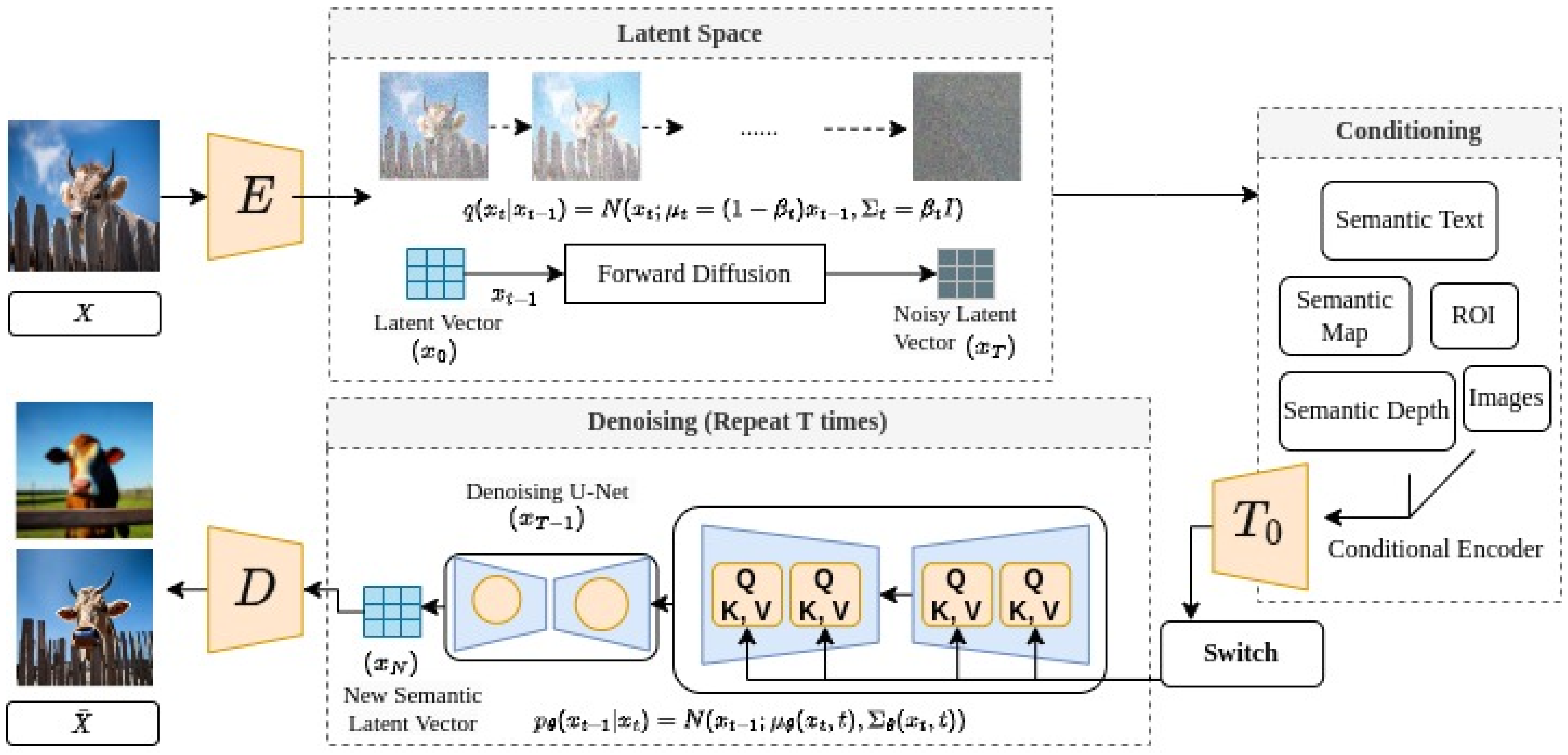




Diffusion models are generative models that work by adding noise to the data and then learning to remove it in the reverse process.

System Model:

EDGE DIFFUSION BLOCK



- To refine the process of image reconstruction block, we are employing lightweight edge diffusion model with zero-shot capabilities. Diffusion models are a class of latent variable models that learn to generate data by reversing a Markov chain that gradually adds noise to the data. The Markov chain is defined by a score function that measures the likelihood of the data given the noise level.

$$d(x_t \mid x_{t-1}) = \mathcal{N}\left(x_t; \mu_t = 1 - \beta_t x_{t-1}, \Sigma_t = \beta_t I\right)$$

With the expression for $d(x_{1:T} \mid x_0)$ is given by:

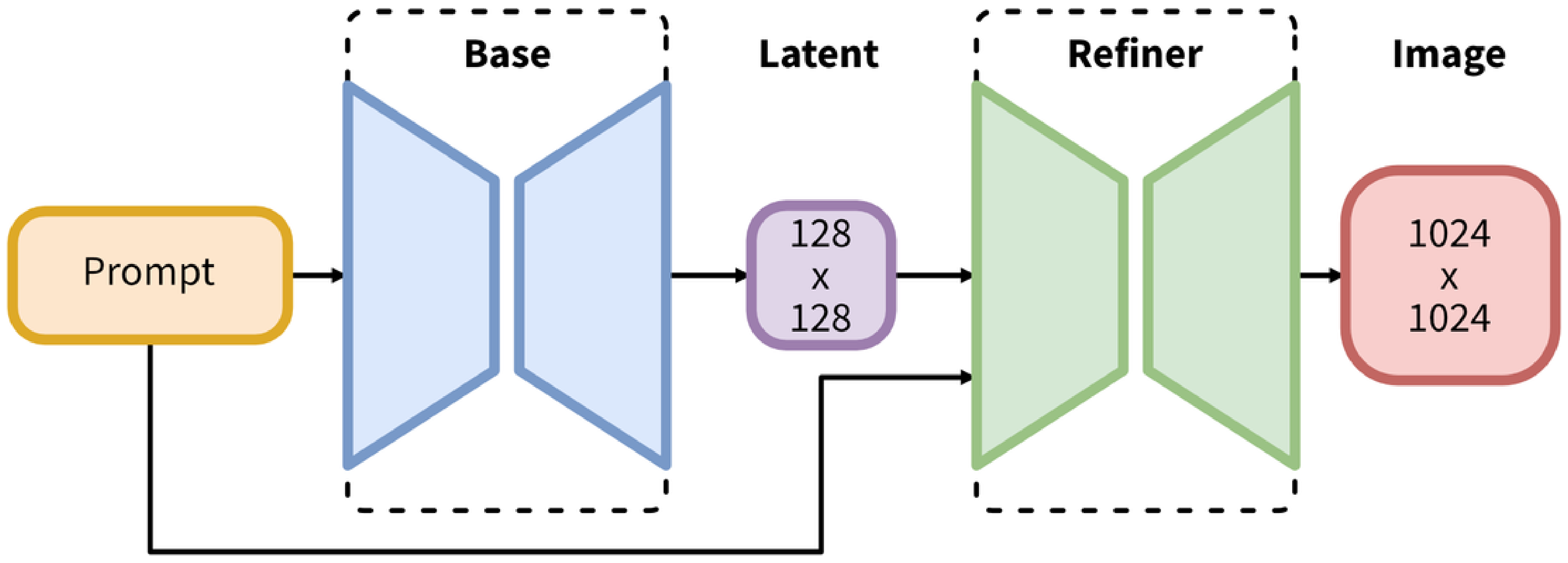
$$d(x_{1:T} \mid x_0) = \prod_{t=1}^{T} d(x_t \mid x_{t-1})$$

- The diffusion models iterate backward over this chain to learn the reverse process and thus produce high-fidelity image data from semantic text descriptions. They are trained beforehand.

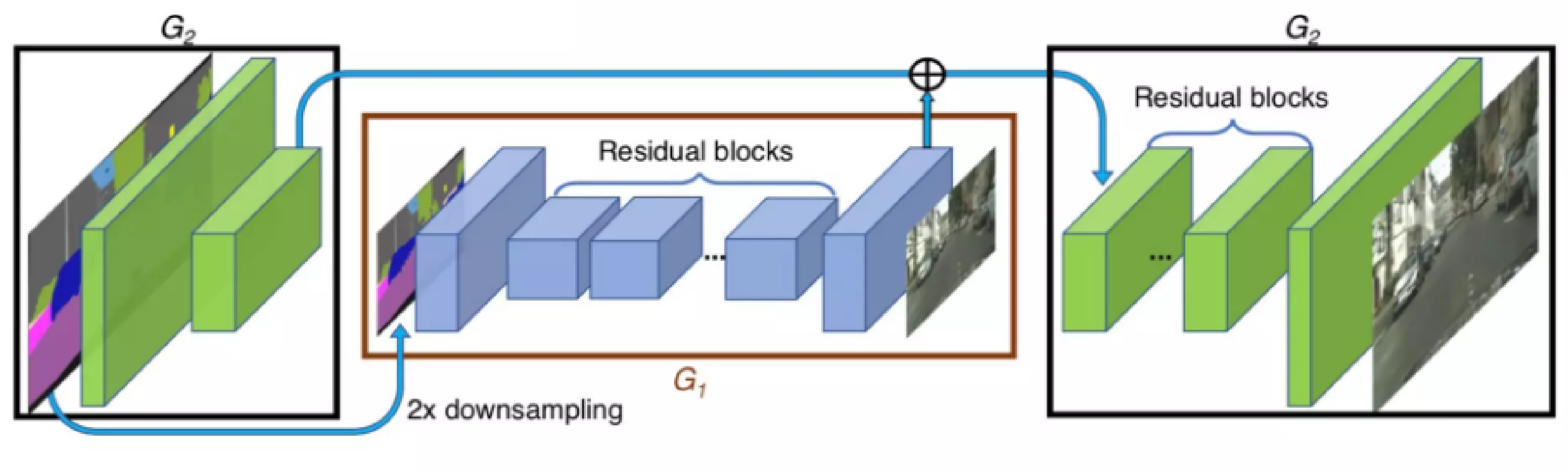
$$r_\theta(x_{t-1} \mid x_t) = \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)\right)$$

System Model:

DIFFUSION + REFINER VAE



U-Net Architecture in the previous slide represented as follows:



(Source: Jonathan Ho, "Denoising Diffusion Probabilistic Models," arXiv preprint arXiv:2006.11239 (2020), cs.LG.)

We present two models as the edge diffusion block, which is distincted by the last layers of denoising process, namely **DiffRes** and **DiffRefiner**.

- DiffRes operates without a refiner component but includes upscaling features.
- It focuses on introducing resilience during the initial stages.
- The upscale refiner is pretrained layer that focuses on bottleneck the latent representations, then provide the output of refined latent.
- This refined latent then decoded using pretrained VAE decoder to get the image back

- In the DiffRefiner, the process is similar, but with different pretrained model of refiner and VAE, which has been trained on ImageNet and Coco.

Simulations:

TRAINING, TESTING

Hyper params	Value
Epoch	1500
Dataset (Training)	Coco-Test-2017
Dataset (Testing)	Military and Civilian Vehicles
Metrics	LPIPS, SSIM
Learning Rate	0.001 (Halved/100 steps)
Loss	Cross Entropy, Focal
SNR	10, 15
Channel	AWGN (training)

We train the bottleneck of JSCC architecture, which considered as VAE with the Coco dataset with hyperparameters specified in the table.

The model is saved without any adaptation and test under different datasets to perform zero shot.

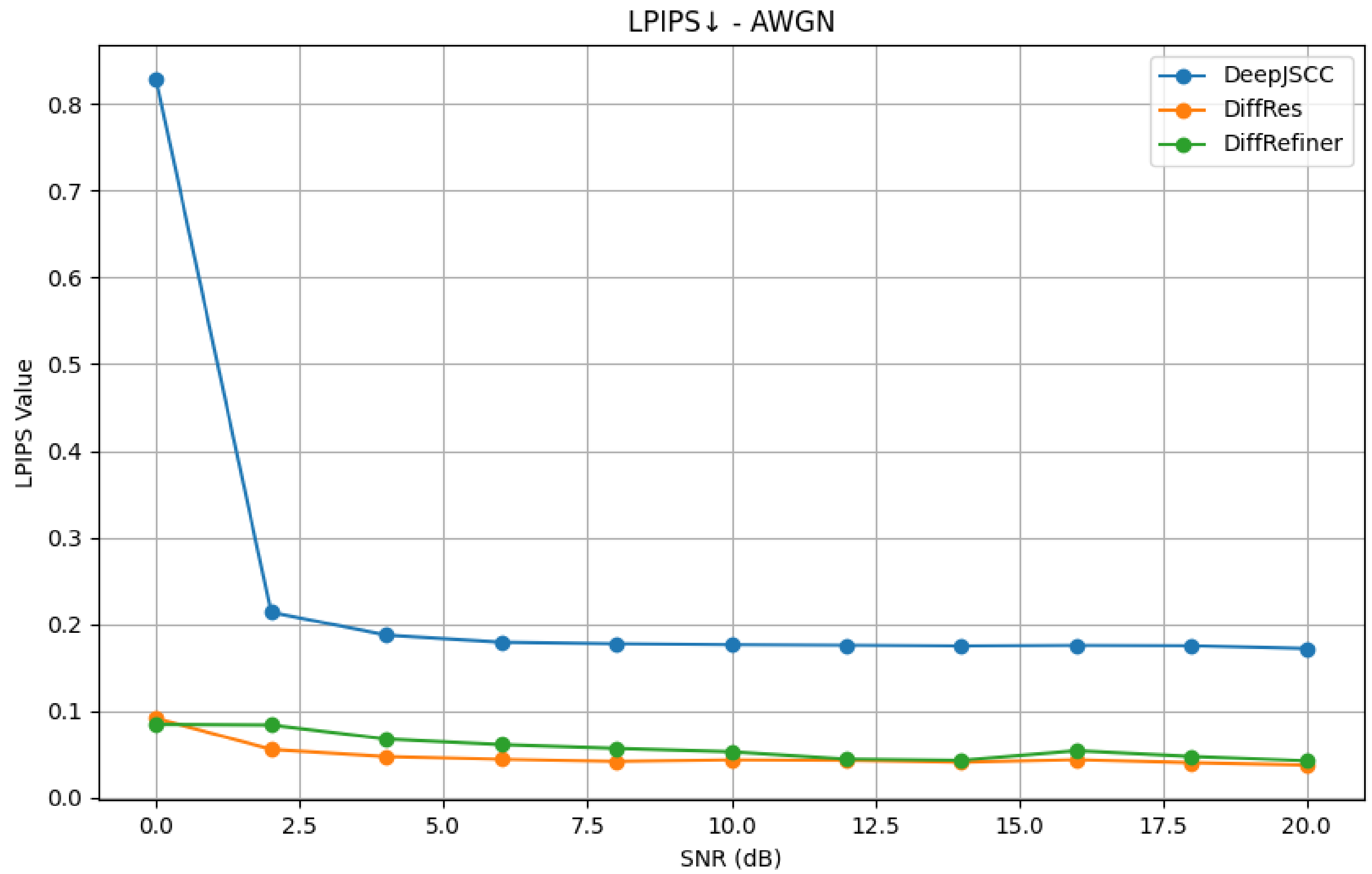
Metrics:

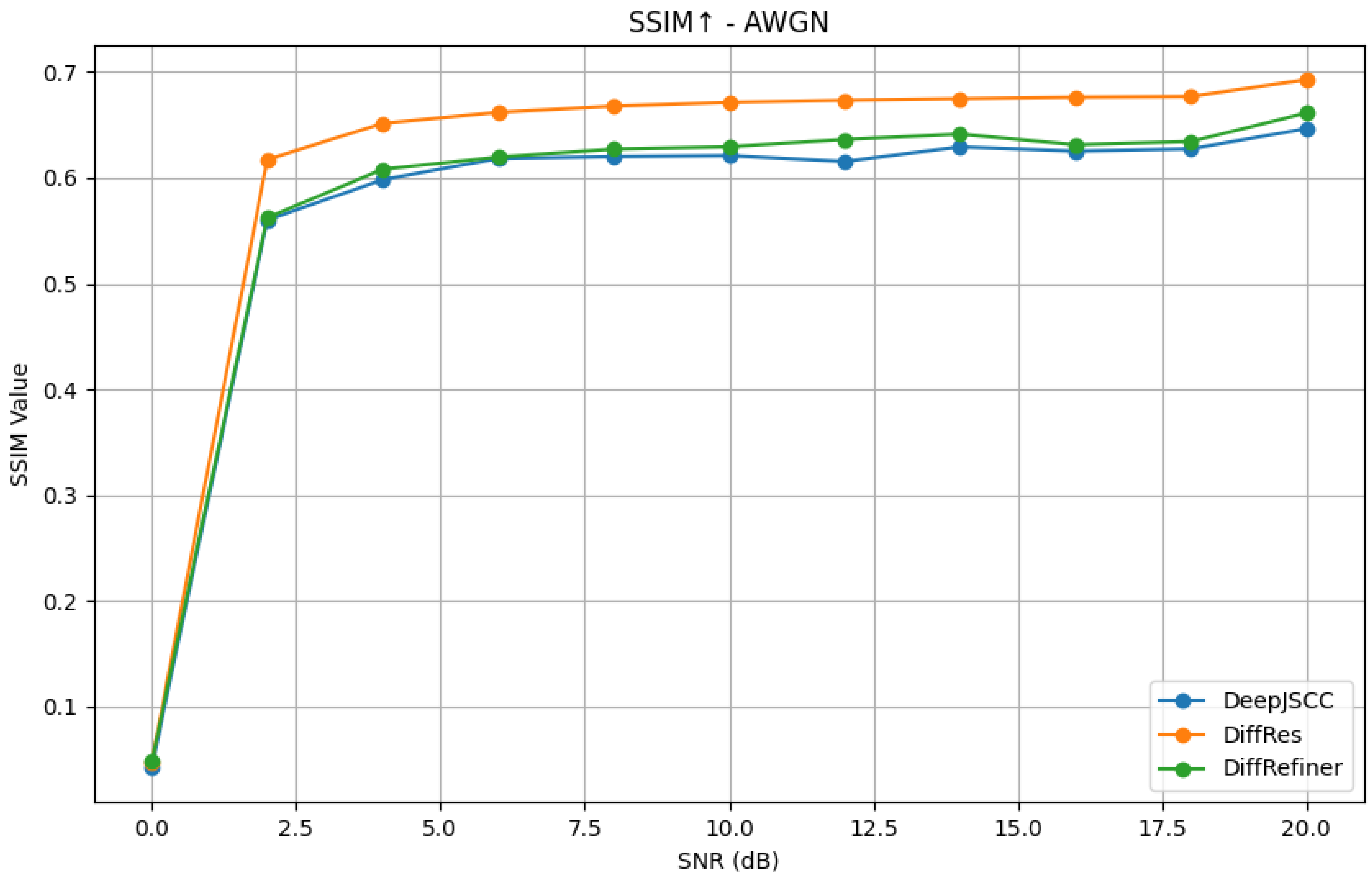
LPIPS:
Measures perceptual difference, aligning closely with human judgment. Lower LPIPS values signify a perceptually closer match to the original image.

SSIM:
Evaluates structural information between images. Higher SSIM values indicate a closer resemblance to the original image.

Case study &

SIMULATION RESULTS: AWGN



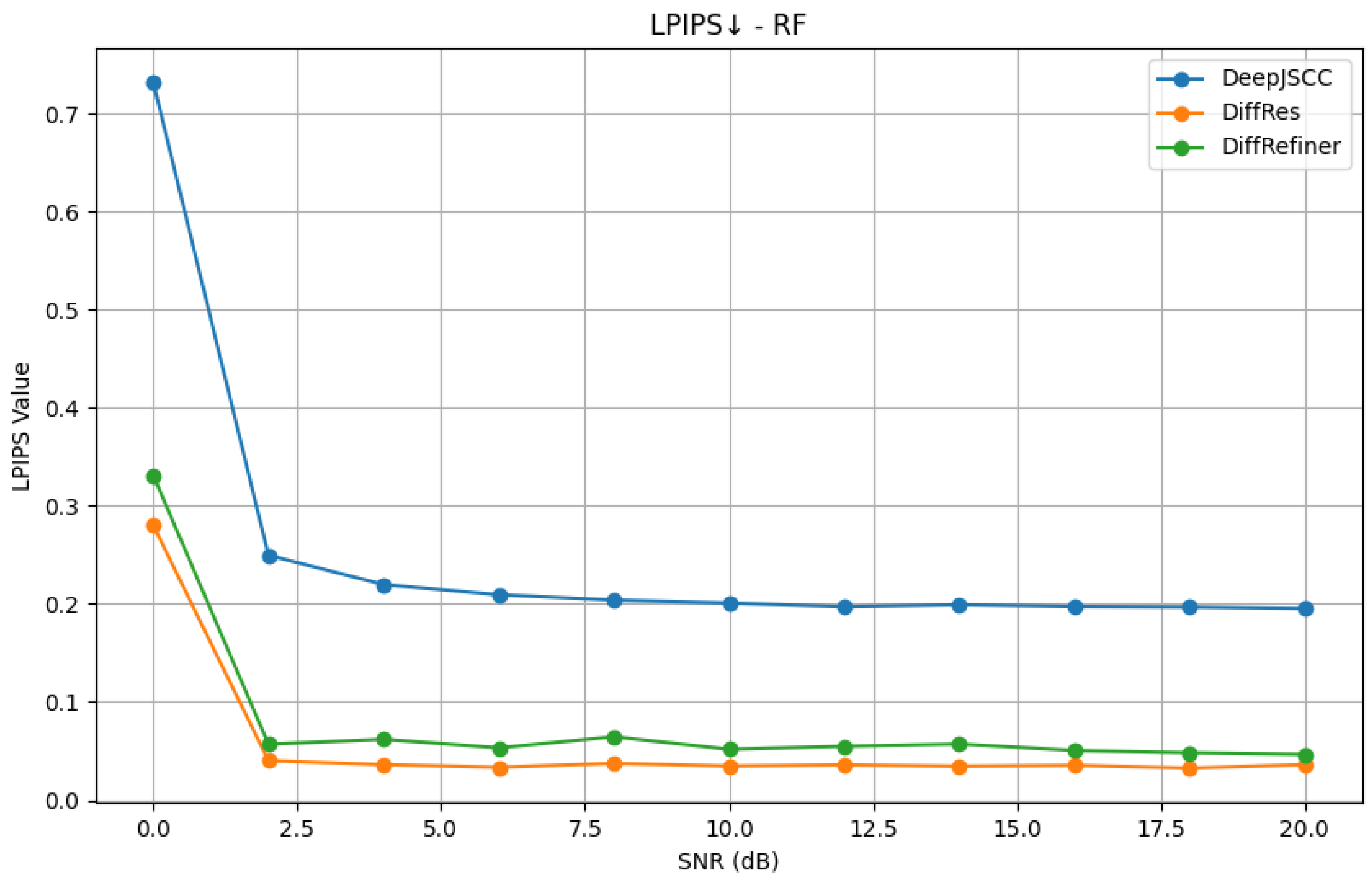


Both DiffRes and DiffRefiner achieved lower LPIPS values by 0.1 compared to the DeepJSCC.

In the case of the SSIM metric, DiffRes achieves the highest SSIM value, while DiffRefiner shows a bit improvements compared to the DeepJSCC.

Case study &

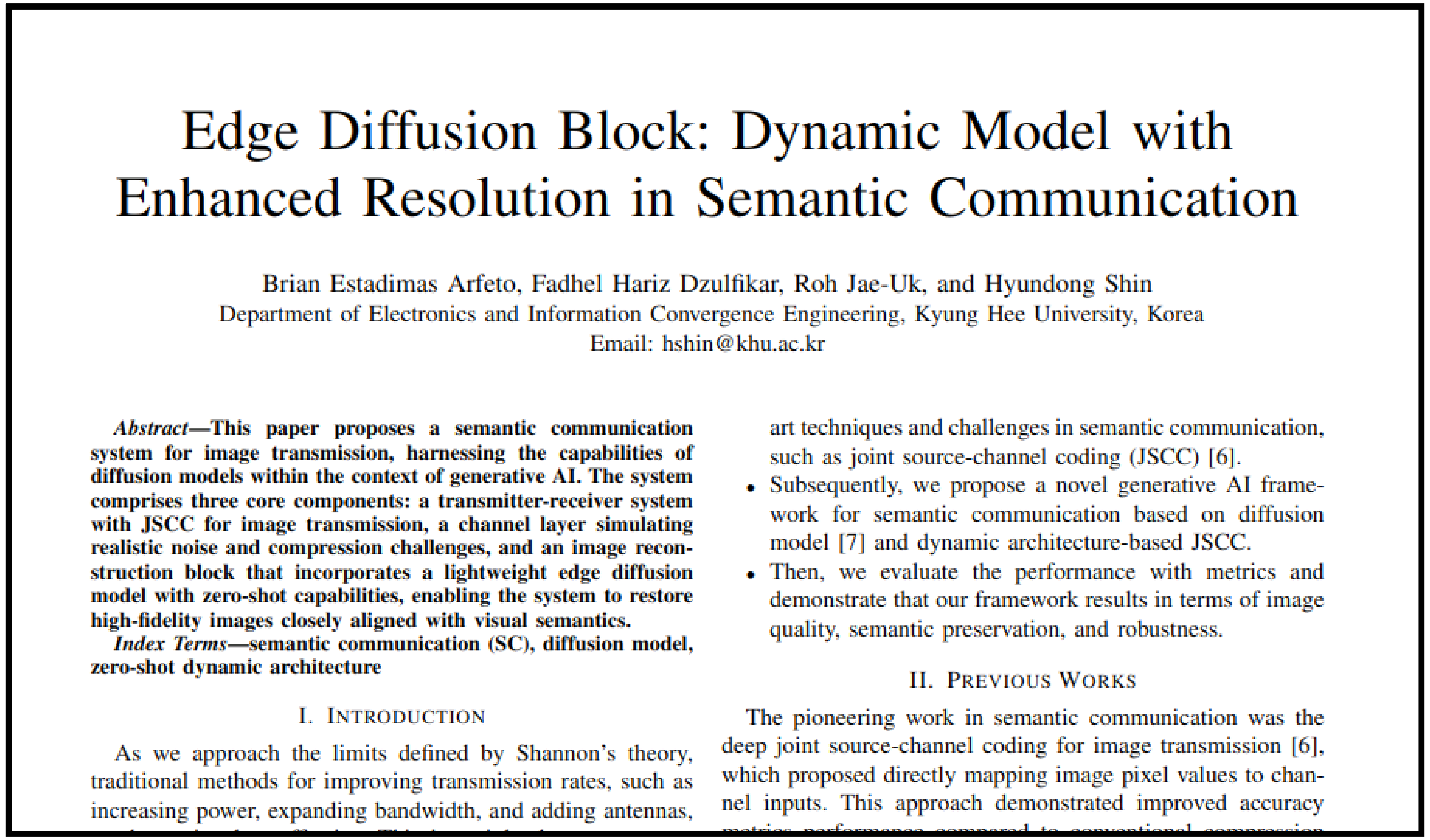
SIMULATION RESULTS: RAYLEIGH FADING



DiffRes and DiffRefiner have lower LPIPS values than the DeepJSCC model, measuring around 0.15 LPIPS.

CONCLUSIONS

Our proposed frameworks, DiffRes and DiffRefiner, excel in image transmission over noisy channels. Leveraging diffusion models and a dynamic Joint Source-Channel Coding (JSCC) architecture, our system preserves image integrity in low SNR scenarios. DiffRes optimizes initial transmission, surpassing DeepJSCC with higher SSIM values. Both frameworks demonstrate substantial reductions in LPIPS values as SNR levels increase, showcasing their effectiveness in mitigating distortions. This research contributes significantly to semantic communication, promising applications in 6G technologies. Future exploration includes the role of semantic information in broader task achievability and leveraging lightweight, sustainable architectures for further advancements.



Edge Diffusion Block: Dynamic Model with Enhanced Resolution in Semantic Communication

Brian Estadimas Arfeto, Fadhel Hariz Dzulfikar, Roh Jae-Uk, and Hyundong Shin
Department of Electronics and Information Convergence Engineering, Kyung Hee University, Korea
Email: hshin@khu.ac.kr

***Abstract*—This paper proposes a semantic communication system for image transmission, harnessing the capabilities of diffusion models within the context of generative AI. The system comprises three core components: a transmitter-receiver system with JSCC for image transmission, a channel layer simulating realistic noise and compression challenges, and an image reconstruction block that incorporates a lightweight edge diffusion model with zero-shot capabilities, enabling the system to restore high-fidelity images closely aligned with visual semantics.**

***Index Terms*—semantic communication (SC), diffusion model, zero-shot dynamic architecture**

I. INTRODUCTION

As we approach the limits defined by Shannon's theory, traditional methods for improving transmission rates, such as increasing power, expanding bandwidth, and adding antennas,

art techniques and challenges in semantic communication, such as joint source-channel coding (JSCC) [6].

- Subsequently, we propose a novel generative AI framework for semantic communication based on diffusion model [7] and dynamic architecture-based JSCC.
- Then, we evaluate the performance with metrics and demonstrate that our framework results in terms of image quality, semantic preservation, and robustness.

II. PREVIOUS WORKS

The pioneering work in semantic communication was the deep joint source-channel coding for image transmission [6], which proposed directly mapping image pixel values to channel inputs. This approach demonstrated improved accuracy

THANK YOU

감사합니다

APPENDIX

->

ROADMAP
SHARED KB SEMANTIC COMMUNICATIONS

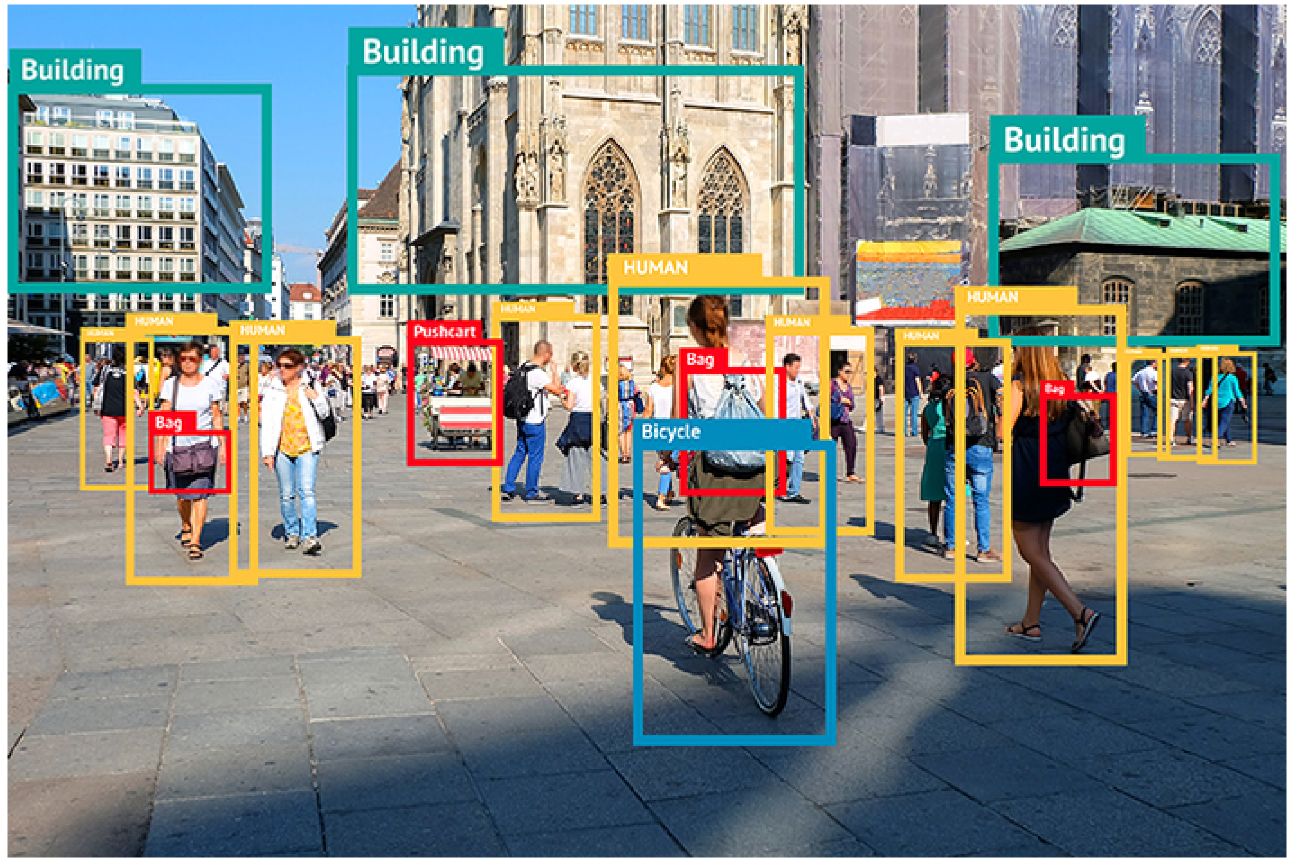
A problematic, is that DeepJSCC is trained with specific SNR rate and defined channel noise. However, in real scenario, the model should adapt in any condition.

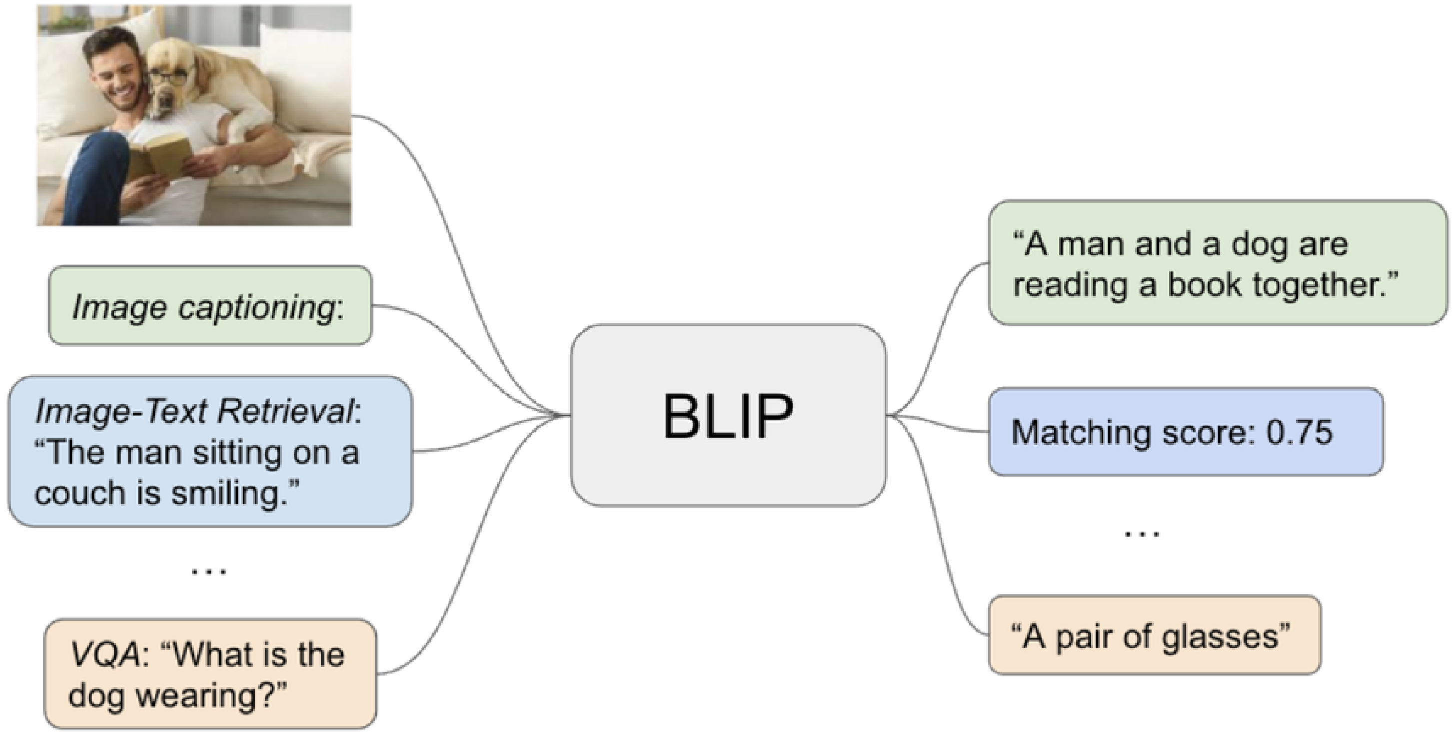
Communication KB Techniques	Knowledge Representations	Sustainability KPI	Advantages	Drawback	Evaluation Metrics
Database-driven	Relational Databases	Computational efficiency, storage efficiency, energy efficiency	Structured, easy to query	Limited to specific structures, scalability	Precision, recall, F1-score
Knowledge Graph	Graph-based	Model efficiency, energy, power, scalability	Semantic connections, scalable	Complexity in maintenance	Accuracy, Latency
Rule-based	Logic and Rules	Computational efficiency, storage efficiency	Easily extensible, transparent	Brittle, maintenance issues	Accuracy, Comprehensibility
ANN-based (end-to-end)	Neural Networks	Power, computational efficiency, model efficiency	Good for complex patterns	Requires large datasets	Accuracy, Loss, Time
Zero-shot Semantic Communication	Embeddings, Transformers	Energy efficiency, model efficiency	Fast deployment, broad applications	Limited customization	Accuracy, Transferability

Knowledge-based techniques are methods that use various forms of knowledge representations to perform tasks such as **reasoning, inference, learning**, and **decision making**. Different knowledge-based techniques have different sustainability key performance indicators (KPIs), advantages, and drawbacks.

Case Study:

MEANING EXTRACTIONS



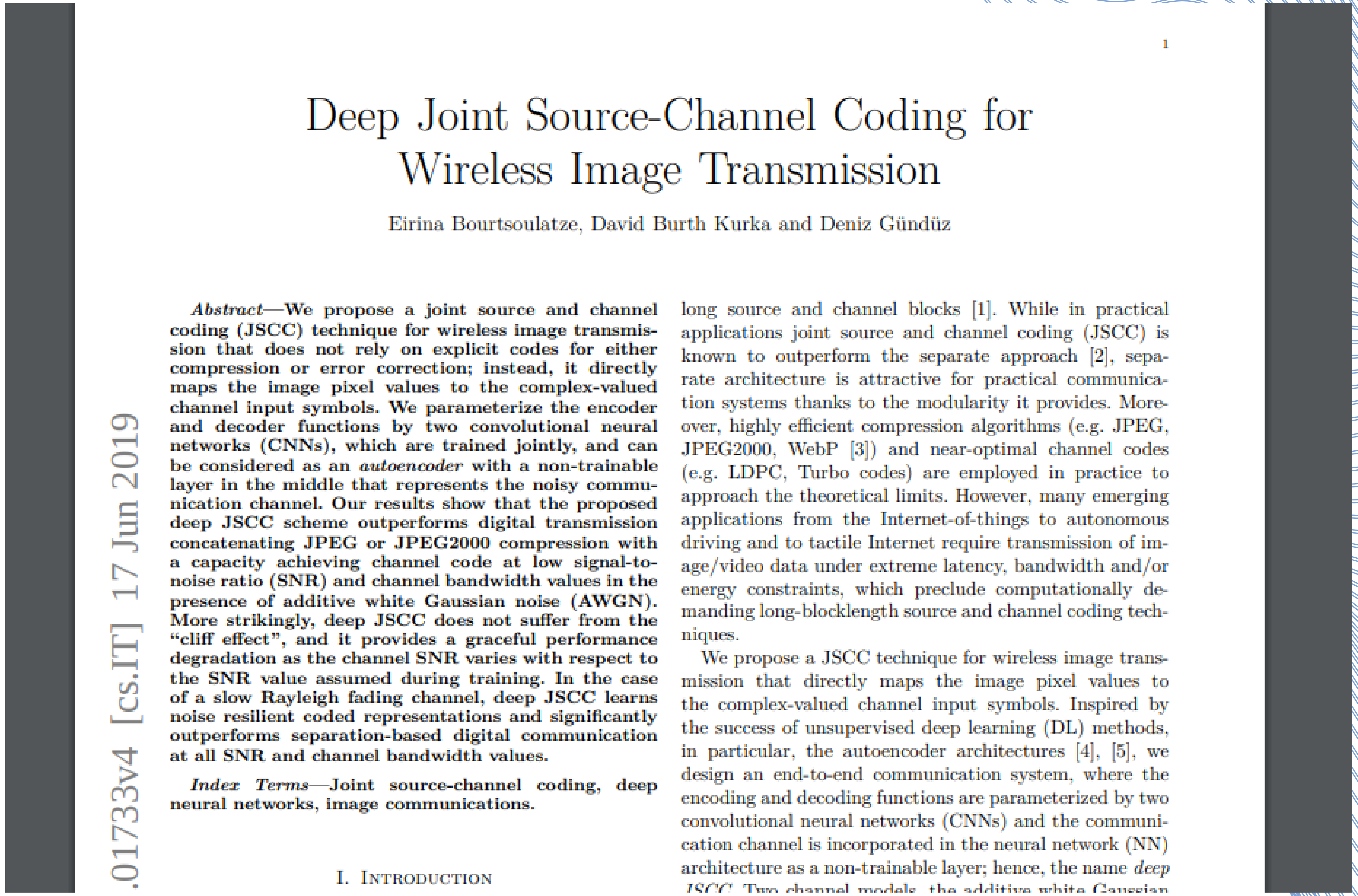


- We are employing BLIP and YOLO for image captioning and detection, for use case and for metrics.
- BLIP is a new vision-language pre-training framework that can transfer flexibly to both vision-language understanding and generation tasks. BLIP can effectively utilize noisy web data by bootstrapping the captions, where a captioner generates synthetic captions and a filter removes the noisy ones.
- YOLO is a fast and accurate object detection algorithm that can process images in real time. YOLO divides the image into a grid of cells and predicts bounding boxes and class probabilities for each cell.

DEEP-JSCC
LITERATURE REVIEW

The pioneering work in semantic communication was the deep joint source-channel coding for image transmission, which proposed directly mapping image pixel values to channel inputs.

This approach demonstrated improved PSNR performance compared to conventional compression techniques such as JPEG and JPEG2000 under noisy channel conditions



Deep Joint Source-Channel Coding for Wireless Image Transmission

Eirina Bourtsoulatze, David Burth Kurka and Deniz Gündüz

Abstract—We propose a joint source and channel coding (JSCC) technique for wireless image transmission that does not rely on explicit codes for either compression or error correction; instead, it directly maps the image pixel values to the complex-valued channel input symbols. We parameterize the encoder and decoder functions by two convolutional neural networks (CNNs), which are trained jointly, and can be considered as an *autoencoder* with a non-trainable layer in the middle that represents the noisy communication channel. Our results show that the proposed deep JSCC scheme outperforms digital transmission concatenating JPEG or JPEG2000 compression with a capacity achieving channel code at low signal-to-noise ratio (SNR) and channel bandwidth values in the presence of additive white Gaussian noise (AWGN). More strikingly, deep JSCC does not suffer from the "cliff effect", and it provides a graceful performance degradation as the channel SNR varies with respect to the SNR value assumed during training. In the case of a slow Rayleigh fading channel, deep JSCC learns noise resilient coded representations and significantly outperforms separation-based digital communication at all SNR and channel bandwidth values.

Index Terms—Joint source-channel coding, deep neural networks, image communications.

I. Introduction

long source and channel blocks [1]. While in practical applications joint source and channel coding (JSCC) is known to outperform the separate approach [2], separate architecture is attractive for practical communication systems thanks to the modularity it provides. Moreover, highly efficient compression algorithms (e.g. JPEG, JPEG2000, WebP [3]) and near-optimal channel codes (e.g. LDPC, Turbo codes) are employed in practice to approach the theoretical limits. However, many emerging applications from the Internet-of-things to autonomous driving and to tactile Internet require transmission of image/video data under extreme latency, bandwidth and/or energy constraints, which preclude computationally demanding long-blocklength source and channel coding techniques.

We propose a JSCC technique for wireless image transmission that directly maps the image pixel values to the complex-valued channel input symbols. Inspired by the success of unsupervised deep learning (DL) methods, in particular, the autoencoder architectures [4], [5], we design an end-to-end communication system, where the encoding and decoding functions are parameterized by two convolutional neural networks (CNNs) and the communication channel is incorporated in the neural network (NN) architecture as a non-trainable layer; hence, the name *deep*

PRICE TO PAY

MAXIMIZING CHANNEL CAPACITY - SHARED KB

Communication KB Techniques	Knowledge Representations	Sustainability KPI	Advantages	Drawback	Evaluation Metrics
Database-driven	Relational Databases	Computational efficiency, storage efficiency, energy efficiency	Structured, easy to query	Limited to specific structures, scalability	Precision, recall, F1-score
Knowledge Graph	Graph-based	Model efficiency, energy, power, scalability	Semantic connections, scalable	Complexity in maintenance	Accuracy, Latency
Rule-based	Logic and Rules	Computational efficiency, storage efficiency	Easily extensible, transparent	Brittle, maintenance issues	Accuracy, Comprehensibility
ANN-based (end-to-end)	Neural Networks	Power, computational efficiency, model efficiency	Good for complex patterns	Requires large datasets	Accuracy, Loss, Time
Zero-shot Semantic Communication	Embeddings, Transformers	Energy efficiency, model efficiency	Fast deployment, broad applications	Limited customization	Accuracy, Transferability

Knowledge-based techniques are methods that use various forms of knowledge representations to perform tasks such as **reasoning, inference, learning**, and **decision making**. Different knowledge-based techniques have different sustainability key performance indicators (KPIs), advantages, and drawbacks.

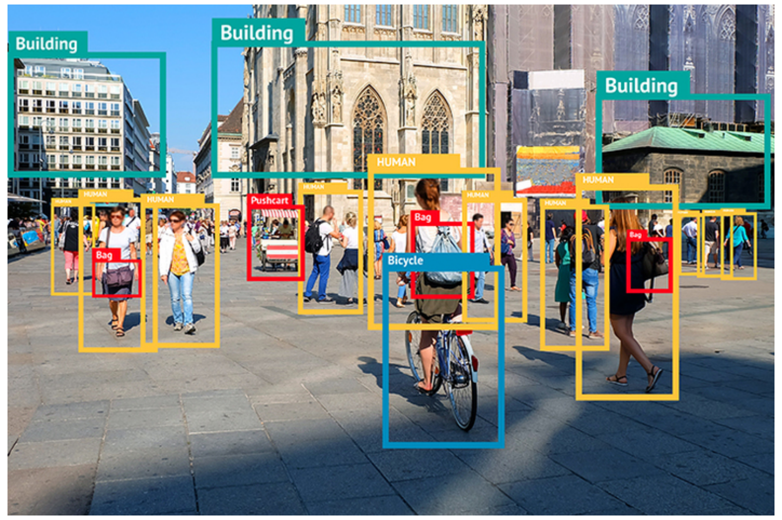
RESEARCH GOALS
GOAL + SECURED ORIENTED COMMUNICATIONS

QUANTUM

By exploiting phenomena like quantum entanglement and quantum key distribution, quantum systems can offer unparalleled security, making them particularly promising for applications where the protection of sensitive information and communication goals is paramount, such as secure government communications and financial transactions.

AI-BASED SC

Deep neural network (DNN)-based methods for semantic communication have demonstrated great potential in a range of applications. Nevertheless, a significant hurdle that requires attention is the scalability issue related to creating the knowledge base (KB). The KB is a crucial part of these systems and holds the vital information that the DNN relies on to analyze and make sense of data.



RESEARCH STRATEGY
JOINT SOURCE CHANNEL CODING

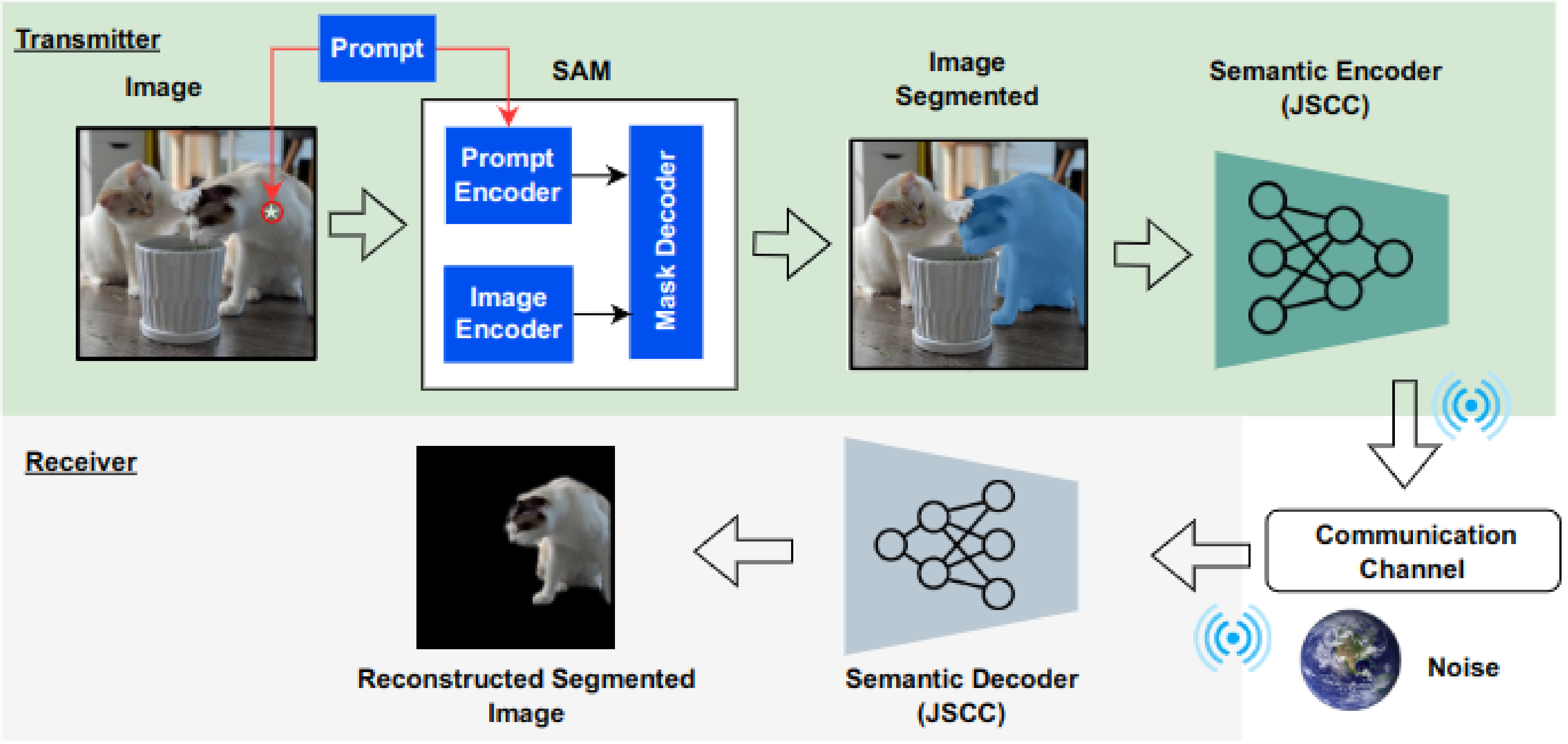
Brian Estadimas Arfeto	Fadhel Hariz Dzulfikar	Roh Jae-Uk
Implementation BigAI-JSCC Zero shot Learning	Implementation previous works (DeepJSCC and WiTT) and Quantum Channel	Performance and metrics analysis

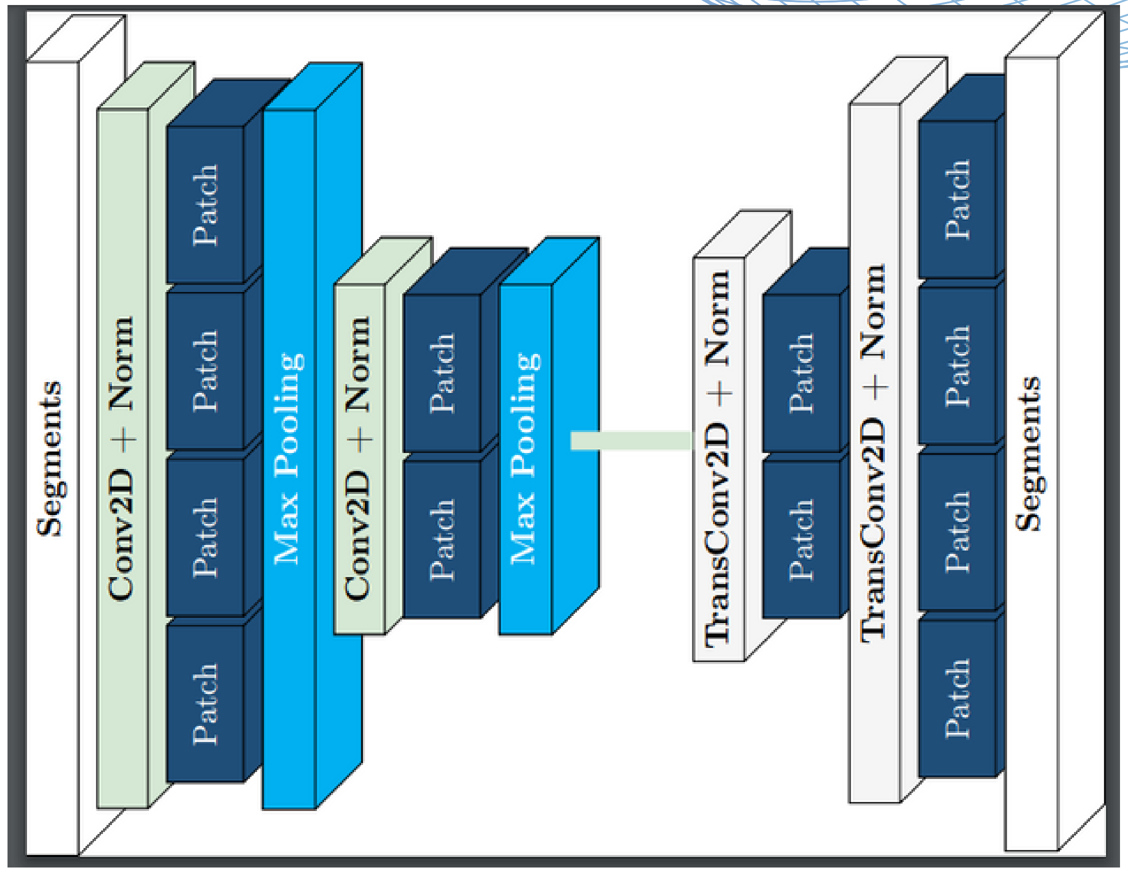
TIMELINE DIVISION

Week 1	Problem Formulation
Week 2	Set up working environment
Week 3-4	DeepJSCC simulation
Week 5-6	WiTT previous paper breakdown simulation
Week 7	JSCC encoder and decoder coding
Week 8	Midterm Presentation
Week 9	Channel simulation and estimation
Week 10-11	Foundation models data extraction and recovery breakdown
Week 12	Testing model with PSNR and LPIPS metrics
Week 13	Performance analysis metrics
Week 14	Real time testing on big COCO dataset
Week 15	Final presentation

RESEARCH METHODS

JOINT SOURCE CHANNEL CODING





Training:

To train our JSCC model, we use a convolutional neural network (CNN)-based architecture that consists of an encoder called "F_enc" and a decoder called "F_dec." This architecture is designed specifically for the source-channel coding of masks. The model's structure is depicted in Figure

The encoded features are then quantized and modulated to generate the channel symbols, denoted as Xc, which are transmitted over a noisy channel. The channel noise is modeled as additive white Gaussian noise (AWGN) with a certain signal-to-noise ratio (SNR). The received symbols, denoted as Yc, are then demodulated and decoded to reconstruct the semantic features, denoted as Xs