February 2024

Master's Degree Thesis

Study on Robustness of Deep Learning Against Physical Layer Impairments in Wireless Communication Systems

Graduate School of Chosun University

Department of Computer Engineering

NAZMUL ISLAM

Study on Robustness of Deep Learning Against Physical Layer Impairments in Wireless Communication Systems

무선통신시스템에서 물리계층 장애에 대한 딥러닝 견고성에 관한 연구

February 23, 2024

Graduate School of Chosun University

Department of Computer Engineering

NAZMUL ISLAM

Study on Robustness of Deep Learning Against Physical Layer Impairments in Wireless Communication Systems

Advisor: Prof. Seokjoo Shin, Ph.D.

A thesis submitted in partial fulfillment of the requirements for a Master of Computer Engineering degree.

October 2023

Graduate School of Chosun University

Department of Computer Engineering

NAZMUL ISLAM

NAZMUL ISLAM의 석사학위논문을 인준함

위원장 모상만 (인)

위 원 강문수 (인)

위 원 신석주 (인)

2023 년 12 월

조선대학교 대학원

TABLE OF CONTENTS

ABBREVIATIONS

5G	Fifth Generation
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASK	Amplitude Shift Keying
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
CIR	Channel Impulse Response
CNN	Convolutional Neural Network
CP	Cyclic Prefix
CV	Computer Vision
dB	Decibel
DFT/IDFT	Discrete Fourier Transform/Inverse Discrete Fourier Transform
DL	Deep Learning
DCT/IDCT	Discrete Cosine Transform
DNN	Deep Neural Network
DWT	Discrete Wavelet Transform
E_b/N_0	Energy per bit to noise power spectral density ratio
ELU	Exponential Linear Unit
FDM	Frequency Division Multiplexing
FFT/IFFT	Fast Fourier Transform/Inverse Fast Fourier Transform
FLC	Fixed length Coding
FLOPs	Floating-point Operations per Seconds
FN	False Negative
FP	False Positive

FSK	Frequency Shift Keying
Gbps	Gigabyte per second
GP	Gray Code penalty
Grad-CAM	Gradient-weighted Class Activation Mapping
GTSRB	German Traffic Sign Recognition Benchmark
ICI	Inter Carrier Inference
IEEE	Institute of Electrical and Electronics Engineers
IID	Independent and Identically Distributed
IoT	Internet of Things
IQ	In-phase and Quadrature
ISI	Inter Symbol Interference
ITS	Intelligent Transport System
JPEG	Joint Photographic Experts Group
JSCC	Joint Source-Channel Coding
LTE	Long-Term Evolution
LS	Least Square
M-QAM	M-ary Quadrature Amplitude Modulation
Mbps	Megabyte per second
MC-CDMA	Multi-carrier code-division multiple access
MHz	Megahertz
MIMO	Multiple Input Multiple Output
ML	Machine Learning
MMSE	Minimum Mean Square Error
MS–SSIM	Multiscale Structural Similarity Index Measure
OFDM	Orthogonal Frequency Division Multiplexing
P/S	Parallel to Serial
PHY	Physical Layer

PIQUE	Perception based Image Quality Evaluator
PSK	Phase Shift Keying
PSNR	Peak-Signal-to-Noise Ratio
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
ReLU	Rectified linear unit
RF	Radio Frequency
RGB	Red Green Blue
RQAM	Rectangular constellation M-QAM
ROI	Region-of-Interest
S/P	Serial to Parallel
SGD	Stochastic Gradient Descent
SQAM	Square constellation M-QAM
SSIM	Structural Similarity Index Measure
SVM	Support Vector Machine
t-SNE	t-distributed stochastic neighbor embedding
TN	True Negative
TP	True Positive
VLC	Variable-length Coding
WAV	Waveform Audio File Format
Wi-Fi	Wireless Fidelity
WIMAX	Worldwide Interoperability for Microwave Access
XQAM	Cross-constellation M-QAM
ZIP	Zone Improvement Plan code

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Study on Robustness of Deep Learning Against Physical Layer Impairments in Wireless Communication Systems

Nazmul Islam
Advisor: Prof. Seokjoo Shin, Ph.D.
Department of Computer Engineering
Graduate School of Chosun University

The Internet of Things (IoT) ecosystem demands fast, reliable, and efficient image data transmission to enable real-time Computer Vision (CV) applications. To fulfill these demands, the Orthogonal Frequency Division Multiplexing (OFDM)-based communication system has been widely utilized due to its parallel transmission scheme, higher spectral efficiency and data rate. However, in OFDM-based image communication systems, various impairments occur due to different physical layer (PHY) blocks such as source coding schemes, complex channel models, data transformation methods, modulation techniques, and the number of antennas, as well as interferences such as Inter-symbol Interferences (ISI) and Inter-carrier Interferences (ICI). These impairments can heavily corrupt the recovered images, making them unsuitable for downstream applications. Although OFDM communication systems have been thoroughly studied for several decades, and each PHY block is approaching maximum statistical and mathematical optimization, impairments still persist in the current digital communication systems. To address these challenges, researchers are exploring the use of artificial intelligence (AI) and Deep Learning (DL) in communication systems and downstream applications to improve overall system performance. DL is robust against

certain image perturbations and can execute CV tasks even on heavily distorted images.

The impairments from the PHY corrupt pixel values independently, but certain intrinsic properties of the image, such as spatial information, may remain intact. In DL, these properties can be extracted as multidimensional features in the convolution layers and interpreted in the top layers, enabling efficient DL-based CV applications like image classification. Therefore, in this study, we first evaluate various OFDM-based image communication systems and the effects of various PHY impairments on image quality. Subsequently, we analyze the robustness of DL models against these impairments for downstream CV applications. For DL-based downstream applications, we consider traffic sign recognition in Intelligent Transport System (ITS) environments in smart cities. Our analysis shows that the EfficientNetV2-based model achieves an accuracy range of 70%–90% across OFDM-based image communication systems with high impairments. Additionally, leveraging different data augmentation techniques improves accuracy by up to 18%.

요약

무선통신시스템에서 물리계층 장애에 대한 딥러닝 견고성에 관한 연구

이슬람 나즈물
지도교수: 신석주
컴퓨터공학과
조선대학교 대학원

IoT(사물인터넷) 생태계는 실시간 컴퓨터 비전(CV) 응용 프로그램을 가능하게 하기 위한 빠르고 신뢰적이며 효율적인 이미지 데이터 전송을 필요로 한다. 이러한 요구를 충족하기위해 직교 주파수 분할 다중화(OFDM) 기반 통신 시스템은 병렬 전송 기법을 활용하고 높은 주파수 효율성과 데이터 전송률 덕분에 널리 사용되고 있다. 그러나 OFDM 기반 이미지 통신 시스템에서는 소스 코딩 방식, 복잡한 채널 모델, 데이터 변환 방법, 변조 기술 및 송수신 안테나 수와 같은 물리 계층 (PHY)에서 야기되는 다양한 손상이 발생하며 교차 심볼 간 간섭(ISI), 교차 캐리어 간 간섭(ICI)과 같은 간섭도 존재한다. 이러한 손상은 복구된 이미지 품질을 심하게 손상시켜 하향 스트리밍 응용 서비스에 적합하지 않을 수 있다. OFDM 기반 통신시스템은 수십 년 동안 심도있게 연구되었기때문에 각각의 PHY 블록은 통계적 혹은 수학적으로 최적화에 근사화되어가고 있지만 디지털 통신시스템에서의 손상(impairment)은 여전히 통신의 장애로 여겨진다. 이러한 손상을 해소하기위해 연구자들은 통신시스템 및 하향 스트리밍 서비스에서 인공 지능(AI)과 딥 러닝(DL)의 활용을 탐구하여 전반적인 시스템 성능을 향상시키고 있다. DL은 특

정 이미지 왜곡에 강하며 심하게 왜곡된 이미지에서도 CV 작업을 실행할 수 있다.

PHY에서 비롯된 손상은 픽셀 값에 독립적으로 손상을 초래하지만 이미지의 공간 정보와 같은 특정 내재적 속성이 유지될 수 있다. DL에서는 이러한 속성들이 합성곱 계층에서 다차원 특징으로 추출되어 최상위 계층에서 해석될 수 있으며, 이를 통해 효율적인 DL 기반 CV 응용 프로그램인 이미지 분류와 같은 작업이 가능하다. 따라서 본 연구에서는 먼저 다양한 OFDM 기반 이미지 통신시스템과 다양한 PHY 손상이 이미지 품질에 어떠한 영향을 미치는 지에 대해서 평가한다. 그 후에 DL 모델이 이러한 손상에 대한 하향 CV 응용 서비스에서의 강건성을 분석한다. DL 기반 하향 응용 서비스에서는 스마트 시티의 지능형 교통 시스템(ITS) 환경에서의 교통 표지판 인식을 고려하였다. 분석 결과, EfficientNetV2 기반 모델은 높은 손상이 있는 OFDM 기반 이미지 통신시스템에서 70%에서 90%의 정확도 범위를 달성함을 보여주었다. 또한 다양한 데이터 증강 기술을 활용하면 정확도가 최대 18% 향상됨을 입증하였다.

.

I. INTRODUCTION

The rapid advancement of technology in recent years has led to the development of a vast network of interconnected devices and systems known as the Internet of Things (IoT) ecosystem, which has transformed the way we interact with the physical world. From smart homes to connected vehicles, IoT technology has enabled smart devices to communicate with each other, share information, and make decisions based on the shared data. These smart devices, sensors, and actuators have enabled us to collect vast amounts of data from our surroundings to develop intelligent systems. Image communication system is at the center of IoT ecosystem to enable Computer Vision (CV) applications such as traffic sign recognition, event monitoring, medical image analysis, and surveillance systems. Most of these systems heavily rely on the efficient and reliable transmission of visual data between devices, which requires high data rate and link capacity to handle the large amounts of image data generated by the IoT devices in real-time.

The emergence of Fifth Generation (5G) communication systems aims to cater to these requirements, with potential advancements in network capacity, spectral efficiency and infrastructure evolution compared to previous generations [1]. One such innovation is the Orthogonal Frequency Division Multiplexing (OFDM) system, which is recognized to mitigate multipath fading, and provide high data rate with high order modulation and parallel transmission schemes [2], [3]. The OFDM communication system uses both frequency and time domain multiplexing to transmit data over multiple sub-carriers. This division of the data stream into several subcarriers allows for parallel transmission, leading to more efficient use of the available bandwidth, and enabling high-speed data transmission with minimal interference and noise. Since its inception, it has become popular and has been adapted to many practical wireless communication systems and standards such as WiMAX, IEEE 802.11, Wi-Fi, and LTE.

Downstream IoT applications demand robust communication systems that offer high data rates and bandwidth efficiency while ensuring image quality retention. While high data rate OFDM-based image communication systems can facilitate the fast transmission of images, there are

significant challenges in processing the vast amount of image data with added noise introduced by the utilization of different OFDM systems and the environment itself. Various factors such as complex channel models, data transformation methods, modulation techniques, the number of antennas, as well as interferences such as Inter-symbol Interferences (ISI) and Inter-carrier Interferences (ICI) can heavily corrupt the recovered images. To address these challenges, researchers are exploring the use of artificial intelligence (AI) and deep learning (DL) to improve the performance of CV applications in IoT ecosystem. AI and DL have emerged as powerful tools for processing and analyzing large amounts of data in real-time for various CV applications, which can extract meaningful information from image data to achieve state of the art performance [4].

1.1. Thesis Statement

Although different physical layer (PHY) blocks in high data-rate OFDM-based image communication systems may introduce various impairments to the received image data, DL are robust against certain impairments and can significantly enhance the reliability of downstream applications that depend on such communication systems for image transmission.

1.2. Objectives and Contributions

OFDM-based wireless communication systems and DL-based CV applications (such as image classification tasks) have been extensively studied in their respective areas; however, their combination has not been thoroughly explored as per literature review, as they are two different areas in the science field. Therefore, this research aims to bridge the gap between the two and conduct a multi-tier analysis of OFDM-based image communication systems for DL applications while availing third-party computational resources. The objectives and contributions of the thesis can be further categorized as:

Research Objective 1: Implement OFDM-based image communication system incorporating various techniques within the PHY blocks to improve overall communication system. Subsequently, evaluate the quality of the output signal and received images, under the influence of impairments

from the PHY block.

Research Question 1: To what extent do various techniques applied in PHY blocks impact the data rate, recovered image quality, and system complexity?

Research Contribution 1: Implemented OFDM-based image communication systems utilizing various source coding/decoding, modulation/demodulation, transformation, channel modelling, and channel estimation techniques in the PHY blocks. Evaluated the performance of the systems in terms of data rate, efficiency, transmission environment and system complexity. Additionally, analyzed the quality of the received images transmitted through these OFDM communication systems.

Research Objective 2: Images received from the various communication systems mentioned in objective 1 can be heavily distorted due to the PHY impairments during transmission. DL are robust against certain distortions and can be utilized for downstream CV applications. Therefore, our second objective is to analyze the robustness of DL on the received images from the OFDM communication systems mentioned in objective 1 for the DL-based downstream task.

Research Question 2: Can heavily distorted image data, corrupted by PHY impairments during transmission, be suitable for DL-based downstream applications?

Research Contribution 2: Conducted robustness analysis of DL-based downstream CV applications in smart cities, specifically, traffic sign recognition in Intelligent Transport System (ITS) environments. The performances of two distinct DL models are analyzed on recovered images from various OFDM-based image communication systems implemented in contribution 1. Additionally, we have utilized different augmentation techniques to improve the performance of the DL models for the downstream tasks.

1.3. Publications and Thesis Outline

The main chapters of this thesis are derived from the research paper published in journals and conferences during my Master's program. The thesis chapters along with the respective publications

are given below:

Chapter 1: This chapter includes the introduction, thesis statement, objectives, contributions and thesis outline along with the publications.

Chapter 2: This chapter provides the preliminaries for the foundation of the research. This includes PHY blocks in OFDM communication system and DL for CV applications.

Chapter 3: This chapter provides the existing related studies of OFDM-based communication system, and robustness of DL on image perturbation.

Chapter 4: This chapter provides the results and discussion of the contributions. It presents the performance analysis of various OFDM-based communication systems and the robustness of DL against impairments from these systems. The analysis and results in this section have been published in below papers (latest first):

P1. **N. Islam** and S. Shin, "Impact of Source Coding on Downstream AI Applications", IEEE International Conf. on Artificial Intelligence in Information and Communication (ICAIIC), Osaka, Japan, Feb. 2024 *(In progress)*

P2. **N. Islam** and S. Shin, "Robust Deep Learning Models for OFDM-Based Image Communication Systems in Intelligent Transportation Systems (ITS) for Smart Cities", Electronics, vol. 12, no. 11, p. 2425, May 2023.

P3. **N. Islam**, I. Ahmad and S. Shin, "Robustness of Deep Learning Enabled IoT Applications Utilizing Higher Order QAM in OFDM Image Communication System", IEEE International Conf. on Artificial Intelligence in Information and Communication (ICAIIC), Bali, Indonesia, Feb. 2023, pp. 630–635.

P4. I. Ahmad, **N. Islam** and S. Shin," Performance Analysis of Cloud-based Deep Learning Models on Images Recovered without Channel Correction in OFDM System", IEEE Asia Pacific Conference on Communication (APCC), Jeju, Republic of Korea, Oct. 2022, pp.

255-259.

P5. I. Ahmad, **N. Islam**, E. Kim and S. Shin, “Performance Analysis of Cloud based Deep Learning Models in OFDM based Image Communication system”, KICS Summer Conference (하계종합학술발표회), Jeju, Republic of Korea, Jun. 2022, pp. 0500-0501.

Chapter 5: This chapter concludes the thesis with key findings and provides possible future directions.

1.4. Other Publications

Besides the mentioned papers, I have contributed to the following publications as first author:

P6. **N. Islam** and S. Shin, “Deep Learning in Physical Layer: A Review on Data Driven End-to-End Communication Systems and their Enabling Semantic Applications” *(In progress)*

P7. **N. Islam** and S. Shin, “Evaluation of Neural Demappers for Trainable Constellation in an End-to-End Communication System", IEEE International Conference on Ubiquitous and Future Networks (ICUFN), Paris, France, Jul. 2023, pp. 39-42.

P8. **N. Islam** and S. Shin, “Review of Deep Learning-based Malware Detection for Android and Windows System”, The 33rd Joint Conference on Communications and Information (JCCI 2023), Yeosu, Republic of Korea, Apr. 2023.

II. BACKGROUND

2.1. OFDM Communication System

2.1.1. Fundamentals of OFDM System

Orthogonal Frequency-Division Multiplexing (OFDM) is an advanced modulation technique that operates on multicarrier principles. This allows a single, high-rate data stream to be conveyed across multiple lower rate subcarriers. This is done by subdividing a high-rate bit-stream into numerous parallel bitstreams, each at a lower rate, and modulating each using orthogonal sub-carriers. Frequency-Division Multiplexing (FDM) employs a systematic approach in the utilization of the frequency band. It segments the band into multiple non-overlapping frequency sub-channels, ensuring that there is no inter-channel interference due to spectral overlap. This approach, although efficient in minimizing interference, doesn't optimize the spectrum's usage. A novel idea emerged in the mid-1960s, proposing the use of parallel data transmission combined with FDM using overlapping sub-channels, known as OFDM [5].

FDM is a fundamental technique used in telecommunications, where multiple signals are transmitted simultaneously over a single communication channel or line. In FDM, each signal is assigned a unique frequency range or sub-channel, ensuring that each signal can be transmitted without interfering with the others. This allows for the simultaneous transmission of various data streams over a common pathway. The mathematical representation of an FDM signal is as follows:

$$S(t) = \sum_{n=1}^{N} s_n(t)\cos\left(2\pi f_n t + \varphi_n\right), \quad (1)$$

where $S(t)$ is the composite signal, $s_n(t)$ represents each individual signal, f_n is the carrier frequency of each signal, φ_n is the phase offset of each signal.

OFDM evolved from the basic concept of FDM. OFDM is a special case of FDM, where the sub-carriers are carefully chosen to be orthogonal to each other. The term 'orthogonal' refers to a precise mathematical relation between the carrier frequencies in the system. In OFDM, the sub-

carriers are sine waves, which are mathematically orthogonal. This orthogonality is key in preventing interference between the sub-carriers, allowing them to be closely spaced, and hence, making the transmission more spectrum efficient. Orthogonality is maintained in OFDM by ensuring that the cross-correlation between any two sub-carriers is zero, as represented by the equation:

$$\int_0^T \cos(2\pi f_1 t)\cos(2\pi f_2 t) = 0;\ for\ f_1 \neq f_2 \tag{2}$$

This orthogonal nature of the sub-carriers in OFDM allows for the precise demodulation of signals at the receiver end, as long as synchronization is meticulously maintained. The OFDM signal is represented as:

$$S(t) = \sum_{n=1}^{N-1} s_n \cdot e^{j(2\pi f_n t + \varphi_n)}, \tag{3}$$

where $S(t)$ is the composite signal, $s_n(t)$ represents each individual signal, f_n is the carrier frequency of each signal, φ_n is the phase offset of each signal.

2.1.2. Advantages of OFDM communication System

OFDM is an advanced modulation scheme used primarily in high-speed digital communication systems, enhancing the overall performance and efficiency of wireless networks. It has various advantages that make it popular for implementation in practical digital communication systems and standards. One of the main advantages of OFDM is its ability to handle high-speed data streams efficiently. It transforms the data from a serial stream to parallel streams, increasing the duration of data symbols on each sub-carrier. This approach mitigates Inter-Symbol Interference (ISI) caused by the dispersive nature of radio channels. Furthermore, to combat ISI, OFDM integrates guard interval, which reduces ISI and negates the necessity for a complex equalizer at the receiver. OFDM is also resistant against multi-path delay spread, a significant cause of ISI. It minimizes delay spread by the conversion of high data-rate symbols into low data-rate, increasing the symbol duration.

OFDM ensures more effective utilization of spectral resources compared to traditional FDM.

Unlike FDM, where the frequency band is divided into multiple individual sub-carriers for transmitting parallel data streams, OFDM allows these sub-carriers to overlap spectrally while maintaining their orthogonality, ensuring optimized spectrum usage without causing interference.

The robustness of OFDM against frequency selective fading or narrowband interference is a significant attribute of OFDM. Unlike single-carrier systems where a single fade could compromise the entire link, OFDM's multicarrier nature ensures that only a handful of subcarriers are affected, allowing for the utilization of error correction coding to amend these minor interferences.

Another advantage of OFDM is its adaptability to asymmetric data services, commonly seen in wireless networks. By altering the number of sub-channels, OFDM can support different transmission rates for uplink and downlink, aligning with the dynamic needs and considerations of mobile communication systems. Such flexibility fosters its compatibility with various access methods, allowing for the development of multi-carrier platforms such as Multi-Carrier Code Division Multiple Access (MC-CDMA) and frequency-hopping OFDM. This adaptability enables simultaneous data transmission from multiple users, enhancing the system's resistance to narrow-band interference by restricting its impact to limited sub-carriers. The interoperability and seamless integrability of OFDM systems with various modulation techniques makes it the ideal choice for hybrid systems as well, such as massive Multiple Input Multiple Output (MIMO) system [6].

2.2. PHY blocks in OFDM Communication System

OFDM is a parallel transmission system that allows spectra of subchannel to overlap with subcarrier tones and coherent detection maximizing the available bandwidth. The PHY of the OFDM communication system is shown in Figure 1.

In the OFDM system, serial high data-rate sub-streams are divided into low data-rate sub-streams and encoded onto orthogonal subcarriers. The spectral shape of the subcarriers is such that the discrete subchannels at the orthogonal subcarrier frequencies have a spectrum of 0 and there are no Inter Carrier Interference (ICI) among the subchannels. As seen in Figure 1, the encoded bit stream

is mapped onto a complex-valued in-phase and quadrature (IQ) constellation plane and converted to IQ data according to the chosen modulation techniques in the modulation block. Deployed modulation techniques will be further discussed in Section 2.2.2. The serial data stream is then converted into parallel data stream and a frequency domain OFDM symbol, $X(k)$, is generated by inserting training data and guard bands into the IQ data. The training data are the pilot symbols (known data-sequence) that are inserted in the data stream to carry out channel estimation [7]. For pilot insertion, there are multiple techniques based on the location and order of pilots in the transmission. We have used comb-type pilot insertion with linear interpolation as it provides improved results [7]. In comb-type pilot insertion, pilots are inserted into specific OFDM subcarriers that are continuously transmitted throughout the communication.

The frequency domain OFDM symbol, $X(k)$, is then transformed into the time-domain OFDM symbol, $x_t(n)$, using inverse transformation followed by parallel-to-serial (P/S) transmission.

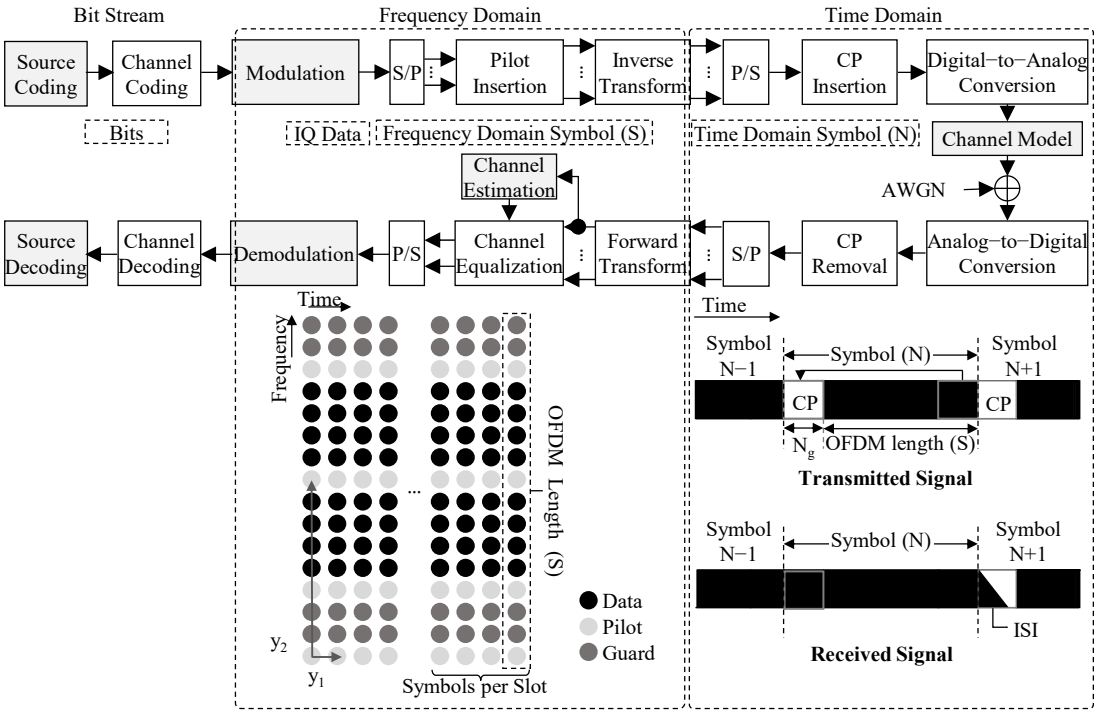

Figure 1. Basic block diagram of OFDM PHY. The frequency–domain and time domain symbol representation is also illustrated.

The transformation techniques will be further discussed in Section 2.2.3. In the next step, a guard interval known as a cyclic prefix (CP), is sliced from the end of $x_t(n)$ symbol and prepended to $x_t(n)$ to form the full time-domain OFDM symbol, $x_{cp}(n)$, which extends the portion of the OFDM symbol in a cyclic manner. The CP mitigates the Inter-Symbol Interference (ISI) and ICI from multipath radio channels and neighboring OFDM symbols, and it is selected to be larger than the anticipated delay spread. OFDM symbol with CP length N_g is given as:

$$x_{cp}(n) = \begin{cases} x_t(N+n), & -N_g \leq n \leq \cdots, N-1 \\ x_t(n), & otherwise \end{cases}. \quad (4)$$

The signal, $x_{cp}(n)$, is then elevated to the radio frequency (RF) by the RF front end and transmitted through the air via wireless channel. During the transmission over the air, the signal experiences the effect of multiple paths and fading effect that creates duplicate signals with different time lags, phases, and amplitudes at the receiver. Applied wireless channel models will be further discussed in Section 2.2.4. The radio signal is then captured and re-converted to IQ samples by the receiver front end. The carrier synchronizer recovers the time-domain OFDM symbols at the receiver. The CP is then removed from $y_{cp}(n)$ to yield the remaining IQ samples, $y(n)$, which is given by:

$$y(n) = \begin{cases} y_{cp}(n), & -N_g \leq n \leq \cdots, N-1 \\ y_{cp}(n+N_g), & otherwise \end{cases}. \quad (5)$$

The symbols are then converted to parallel stream and is transformed into the frequency–domain OFDM symbol, $Y(k)$, using transformation. The symbol $Y(k)$ undergoes channel estimation and channel equalization to produce an estimate of the transmit frequency–domain IQ data, $\hat{X}(k)$. The channel estimation compares the received data sequence and pilot symbols with the previously known ones to obtain an estimate of the Channel Impulse Response (CIR) and perform channel correction. Channel estimation techniques used in the experiment will be further discussed in Section 2.2.5. This estimate is then demodulated into soft bits (log-likelihood), $\tilde{b}$, which are further converted into binary output bits, $\hat{b}$, by the channel decoder before passing it to the next communication layer.

2.2.1. Source Coding

Source coding is the process of assigning binary sequences to elements of a specific alphabet. In this context, the term "alphabet" refers to a collection of symbols, known as ‘letters’. The set of these binary sequences is termed a ‘code’, while the individual sequences within the set are known as ‘codewords’ [9]. Fixed-length coding (FLC) is a type of source coding scheme where every symbol from the source alphabet is represented by an equal number of bits. For instance, an image with color channel Red, Green, and Blue (RGB) will have pixel value ranging from 0 to 255. When using fixed-length coding where each codeword has n bits, the total codewords or reconstruction levels M equals 2^n, where n is the length of bits, therefore each pixel value (as $2^8 = 256$ possible values) is assigned a fixed length of 8 bits. While fixed-length coding is simple and straightforward, it is not efficient. Lower pixel values are assigned with fixed 8-bit even when they have shorter codeword compared to higher pixel values [1], [9]. Additionally, redundant bits are added to the codewords with smaller code lengths to make them fixed length. Errors in these redundant bits, especially in the most significant ones, from the communication channel can drastically alter pixel values and heavily distort the recovered images.

In practical scenarios, image formats usually incorporate advanced techniques, such as compression and variable-length coding (VLC), to minimize the data size, especially for intricate or large images. Standards like JPEG combine transformations, quantization, and variable-length coding techniques like Huffman coding to attain notable compression ratios. For many applications, variable-length coding methods like Huffman coding are favored over fixed-length coding, especially for data compression. The advantage is that variable-length coding can utilize the varying probabilities of symbols in data, enabling more compressed data representations. Huffman coding is a popular method for lossless data compression. This method involves the use of variable-length prefix coding to represent data in such a way that frequently occurring symbols are assigned shorter codes, while less common symbols are assigned longer codes. This algorithm is efficient in reducing data redundancy, thereby optimizing the data for storage or transmission.

To implement Huffman coding, first the frequency or probability of each symbol in the message or dataset is calculated. Symbols are then sorted based on their calculated frequencies, forming the basis for the construction of the Huffman tree, a binary tree where each leaf node represents a symbol. The tree is constructed by repetitively selecting two symbols (or trees) with the lowest frequencies and combining them to form a new binary tree, the frequency of which is the sum of the two selected symbols (or trees). This procedure continues iteratively until a single tree containing all symbols is formed. Next, the binary code for each symbol is generated. where moving left assigns a '0' and moving right assigns a '1' to a symbol's code. Consequently, this leads to the creation of a unique binary code for each symbol, ensuring that no code acts as a prefix to another, ensuring that the encoded data can be uniquely decoded. For instance, a data source that produces symbols A, B, C, and D with probabilities 0.4, 0.3, 0.2, and 0.1, respectively. After performing Huffman coding, symbol A is represented by the code '0', B by '10', C by '110', and D by '111'.

The average length L of the Huffman code can be expressed as:

$$L = \sum_{i=1}^{M} p(x_i) l(x_i), \tag{6}$$

where, $p(x_i)$ is the probability of source symbol x_i, $l(x_i)$ is the length of the Huffman codeword for symbol x_i, M is the total number of unique source symbols. In contrast to fixed-length coding, Huffman coding minimizes the average codeword length, hence reducing the redundancy in the encoded data, which results in data compression. The efficiency of Huffman coding is often measured in terms of entropy $H(X)$, which represents the lower bound on the average length of encoding symbols from a source X. The entropy of a source is given by:

$$H(X) = -\sum_{i=1}^{M} p(x_i) log_2 p(x_{i)}. \tag{7}$$

In an ideal scenario where the probability distribution of the source symbols is known and matches the assumptions made during the Huffman coding process, the efficiency of the code can be

close to the entropy of the source. However, the performance of Huffman coding decreases when there is a mismatch between the actual and assumed probability distributions [1], [9]. Moreover, adaptive Huffman coding can be utilized when the source statistics are unknown or non-stationary. In adaptive Huffman coding, the Huffman tree is updated dynamically as symbols are being encoded or decoded, allowing the code to adapt to the varying symbol probabilities. This makes it more robust in practical scenarios where the statistics of the source output are not known in advance, but it adds complexity to the encoding and decoding process.

Error resilience is another consideration in source coding. In Huffman coding, an error in a single bit can lead to the misinterpretation of multiple symbols, making the error propagate through the decoded output. Error-correcting codes such as Hamming codes can be incorporated into the Huffman coding scheme to improve error resilience. A Hamming code can correct single-bit errors by adding redundant parity bits to the data. For a k-bit data word, the number of parity bits required, r, is determined by the equation:

$$2^r \geq k + r + 1 \tag{8}$$

Though, the trade-offs between compression efficiency, complexity, and error resilience must be considered when utilizing VLC, in real-world applications, Huffman coding plays a significant role in file compression techniques such as ZIP files and multimedia compression formats like JPEG and MP3 [9]. It is effective in reducing data size without loss of information, which makes it crucial for data compression.

2.2.2. Digital Modulation

Digital modulation is a technique used in communication systems to transmit information, encoded as digital data, using an analog signal. Different modulation schemes, such as amplitude, frequency, and phase modulation, are applied to alter the carrier wave, propagating the information within it for transmission. These modulated signals can traverse through the communication channel, encountering various impairments such as noise, attenuation, and interference. Despite these

challenges, the receiver's demodulation process aims to accurately recover the original information from the modulated signals.

Most practical OFDM communication systems exhibit passband frequency responses, and to achieve high throughput with bandwidth and power efficiency, digital passband modulation is frequently employed in these systems. Digital passband modulation schemes can be categorized according to the variation in amplitude, phase, or frequency of the transmitted signal with respect to the message signal, such as ASK, PSK, FSK, and QAM. Binary digital modulation is one of the simplest forms of digital modulation, where the modulating data signal consists of two distinct levels, representing binary '0' or '1'. A widely recognized binary modulation technique is Binary Phase Shift Keying (BPSK), which conveys data by altering the phase of the carrier wave. On the other hand, M-ary digital modulation is a more sophisticated technique that allows the transmission of more than one bit of information in each symbol. In M-ary modulation, "M" symbolizes the number of unique states or levels the modulated signal can have. For example, Quadrature Amplitude Modulation (QAM) and Phase Shift Keying (PSK) are common M-ary modulation techniques [1]. With a higher number of states, M-ary modulation can transmit more bits per symbol, enabling the transfer of data at a much higher rate compared to binary modulation, making it suitable for high-speed data transmission applications.

Various M-ary modulation techniques are commonly used in digital communications. QPSK utilizes a four-point constellation diagram, uniformly distributed around a circular path. Due to its four phases, QPSK has the ability to transmit two bits per symbol, using gray coding to optimize the bit error rate (BER). The symbols of the constellation design using sine and cosine is essential for the transmission. The representation of QPSK signal for the nth phase is given by:

$$s(t) = \sqrt{\frac{2E}{T}} \cos\left(2\pi f_c t + (2n-1)\frac{\pi}{4}\right), \qquad n = 1,2,3,4, \tag{9}$$

where E is the energy per symbol, T is the period of the signal, f_c is the carrier frequency, and n is the index variable that takes values 1, 2, 3, or 4, representing each of the four possible phase shifts

corresponding to phase shifts of $\pi/4$, $3\pi/4$, $5\pi/4$ and $7\pi/4$, respectively in QPSK. This can be represented in IQ plane. The in-phase (I) function of the signal is:

$$\varphi_I = \sqrt{\frac{2}{T}}\cos(2\pi f_c t), \tag{10}$$

and the quadrature-phase (Q) function is:

$$\varphi_Q = \sqrt{\frac{2}{T}}\sin(2\pi f_c t). \tag{11}$$

The signal constellation design of QPSK consist of 4 points in the IQ with coordinates $\pm\sqrt{E/2}$, $\pm\sqrt{E/2}$, and the total power is split between the two carriers equally.

M-QAM is an advanced modulation technique that varies both amplitude and phase of the transmitted signal simultaneously, resulting in greater bandwidth and power efficiency which can be extended to advanced systems, such as MIMO [2], [3], [11]. Additionally, high data rate can be achieved when utilizing higher order M-QAM; however, using higher order M-QAM for higher data rates comes with a cost of higher BER, cell-to-cell interference, smaller coverage area, and hardware complexity [2], [3], [11]. Despite these challenges, M-QAM is extensively used in various IEEE standards and wireless communication systems, such as 5G digital video broadcast communications, satellite communications, Wi-Fi, WiMAX, VDSL, and more. A few applications with different order of M-QAM for modern communication standards are given in Table 1 [12].

Table 1. Applications utilizing M-QAM Modulation for modern communication standards.

Communication Systems	M-QAM Utilized
IEEE 802.11n/g/ad/ay	16, 64
IEEE 802.16m (WiMAX 2)	16, 64
IEEE 802.11ac/af/ah	16, 64,256
DVB-T2 (Digital Video Broadcasting-Terrestrial)	16, 64,256
IEEE 802.22b	16, 64,256
TS 36.331 (Release 14-LTE Advanced Pro)	16, 64,256
TS 36.331 (Release 15-5G support)	16, 64,256, 1024
IEEE 802.11ax (Wi-fi 6)	16, 64,256, 1024

The M-QAM signal waveforms are composed of two independently amplitude-modulated carriers that are in quadrature to each other as shown below [13]:

$$s(t) = I(t)\cos(2\pi f_c t) - Q(t)\sin(2\pi f_c t), 0 \le t \le T, \quad (12)$$

where T time duration of symbol, f_c is the carrier frequency, and $I(t)$ and $Q(t)$ are the respective IQ components of the signal.

Wireless communication systems aim to achieve high data-rate transmission with efficient utilization of limited bandwidth for the best quality of service (QoS), and high spectral and power efficiency. This can be achieved by reducing the average transmit power of the constellation design in the modulation schemes [2], [3], [11]. Most of the M-QAM constellation designs proposed decades ago (from early 1960s) are still being used in commercial communication systems. M-QAM has the unique capability to encode information in both amplitude and phase of the transmitted signal, allowing higher spectral efficiency. This enables more bits to be encoded per symbol for a given average energy in the constellation design [2], [3]. Rectangular M-QAM (RQAM) constellation design is widely used due to its improved error performance and use of a simple maximum likelihood detection method, which reduces system complexity [12], [14]. Square M-QAM (SQAM) is a type of RQAM constellation that is optimized for even-length symbols (16-QAM, 64-QAM, 256-QAM, and 1024-QAM) [15]. The distribution of constellation points (symbols) in SQAM forms a perfect square lattice with equal rows and columns and maximizes the minimum Euclidean distance between points for a given average symbol power. In SQAM, each constellation points shares boundaries with a maximum of four neighboring points and differs by only one bit, therefore it has perfect gray coding, with a Gray Code penalty (GP) value 1 [14]. The GP was introduced in [14] as the average difference in bits between adjacent symbols in a constellation:

$$G_P = \frac{1}{M}\sum_{i=1}^{M} G_P^{S_i}, \quad (13)$$

where $M = 2^k$, k is the length of symbol and $G_P^{S_i}$is the GP for the i^{th} data symbol, S_i. In SQAM,

the k bits of the serial data stream are represented on a two-dimensional constellation design using gray coding. The $I(t)$ and $Q(t)$ distributed over the set $\{\pm d, \pm 3d, \dots, \pm(M-1)d\}$ independently, where $2d$ is the adjacent Euclidean distance between the two constellation points given by [13]:

$$d = \sqrt{\frac{3log_2 M \cdot E_b}{(M^2-1)}}, \tag{14}$$

where E_b is the bit energy and M is the order of QAM. Bit Error Rate (BER) is a vital factor in evaluating the effectiveness of constellation designs in modulation schemes. Later sections will cover the analysis of the experimental results, but the theoretical bit error calculation of SQAM is given by [13]:

$$B_e = \frac{1}{log_2\sqrt{M}} \sum_{i=1}^{log_2\sqrt{M}} B_e(i), \tag{15}$$

where $B_e(i)$ is the bit error probability of the i^{th} bit.

When designing constellation points with odd length of symbols (32-QAM, 128-QAM, 512-QAM) in RQAM, the constellation points are usually distributed to form a perfect rectangle, either with a horizontal shape parallel to the in-phase axis or with a vertical shape parallel to the imaginary axis, both having the same average energy; however, this is not desirable as it increases the peak and average powers. To overcome this, [14] proposes SQAM with the outer corner constellation points relocated for odd length of symbols forming a cross shape constellation design referred to as cross-constellation M-QAM (XQAM). XQAM reduces the peak and average energy and provides at least 1 dB gain over prefect rectangular shape constellation design [14], [16]. The generalized GP for higher order XQAM constellation design can be given as $\left(1+\frac{1}{\sqrt{2M}}+\frac{1}{3M}\right)$ [16]. For the BER calculations, authors in [14] proposed an approximate expression which was reproduced by authors in [16] for XQAM as:

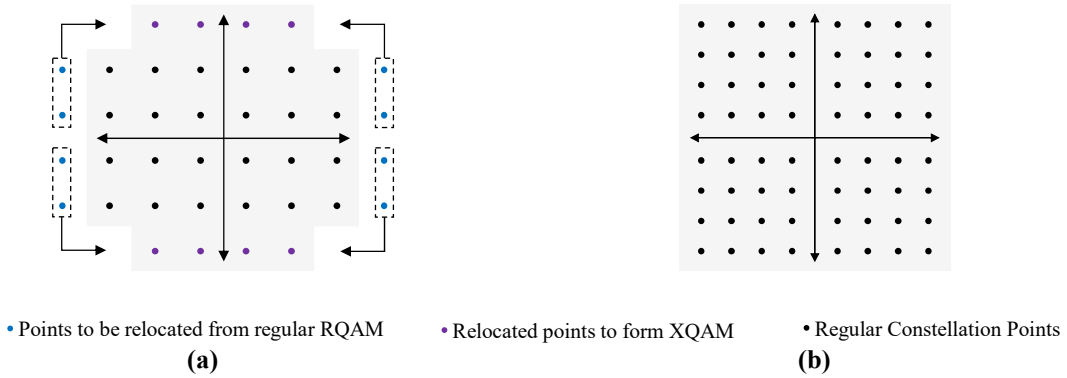

Figure 2. Constellation design for M-QAM modulation. (a) XQAM for odd length of symbol (32-QAM). (b) SQAM for even length of symbol (64-QAM).

$$B_e(M) \cong G_p \frac{A_n}{log_2 M} \frac{1}{2} erfc\left(\frac{d}{\sqrt{N_0}}\right), \tag{16}$$

where G_p is the GP, A_n is the average number of nearest neighbors of a constellation point [given by$\left(4 - \left(\frac{6}{\sqrt{2M}}\right)\right)$ for XQAM of M≥ 32], N_0 is the noise density, $2d$ is the adjacent Euclidean distance between the two constellation points and $erfc(\cdot)$ is the error function defined as:

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt. \tag{17}$$

The decision region for SQAM is simple because the IQ component has distinct vertical and horizontal decision regions. However, for XQAM constellations, only horizontal and vertical decision regions are not sufficient as their corner constellation points are not available. In XQAM, the end columns symbols are relocated to new cross-type positions in the constellation, forming three types of symbols, edge symbols, corner symbols, and interior symbols [14], [16], as shown in Figure 2. The inner symbols form a closed square decision region, while the edge symbols form semi-infinite rectangles. The decision region for the corner symbols forms a 45-degree angle along the horizontal and vertical axes. Due to this irregular pattern, some of the noisy symbols are not transformed into the mapped decision region, resulting in higher bits in error for XQAM constellation design [2], [12].

Higher order M-QAM utilizes a greater number of bits per symbol and can significantly

Table 2. Parameters and bandwidth efficiency of XQAM and SQAM constellation designs.

M-QAM	Bits per Symbol (k)	PAPR	GP	Bandwidth Efficiency ↑
16 +	4	1.80	1	640 Mbps
32 *	5	1.70	1.166	800 Mbps
64 +	6	2.333	1	960 Mbps
128 *	7	2.073	1.065	1.12 Gbps
256 +	8	2.647	1	1.28 Gbps
512 *	9	2.28	1.039	1.44 Gbps
1024 +	10	2.81	1	1.6 Gbps

* XQAM constellation; + SQAM constellation; ↑ Per 160 MHz Channel.

increase the data rate. The potential capacity of a system utilizing higher order M-QAM can be calculated by using Shannon's channel capacity formula [17] as:

$$C = Blog_2\left(1 + \frac{S}{N}\right) \text{ bits/s}, \tag{18}$$

where C is the channel capacity in bits per second, B is the bandwidth of the channel in Hertz, S is the signal power, and N is the noise power. Using Equation (18) to express Shannon bound as bandwidth efficiency, $\eta = \frac{C}{B}$, it can be expressed as:

$$\eta = log_2\left(1 + \frac{S}{N}\right) \text{bits/s/Hz}. \tag{19}$$

In IEEE 802.11ax (Wi-Fi 6) standard, 160 MHz is the maximum allocated channel width [18]. Considering that, we can calculate the bandwidth efficiency using Equation (19) for higher order M-QAM, as given in Table 2. It shows that with higher order M-QAM, the data rate increases significantly. Additionally, the table also shows PAPR and GP values for higher order M-QAM calculated in [12]. From the table, we can observe that XQAM exhibits higher GP compared to SQAM due to its irregular pattern; however, SQAM has higher PAPR compared to XQAM, due to its full lattice structure [12]. For the experiment we have considered higher order M-QAM (16-QAM to 1024-QAM) modulation with both SQAM and XQAM constellation design for the OFDM-based image communication system.

2.2.3. Transformation

Transformation techniques are used in OFDM to modulate and demodulate the data on the subcarriers. In the transmitter, after the data symbols are mapped onto the subcarriers in the frequency domain, an inverse transformation is applied that converts the frequency-domain symbols into a time-domain signal which can be transmitted over the communication channel. At the receiver, the received time-domain signal is converted back to the frequency domain to demodulate the data symbols from the subcarriers using forward transformation which facilitates the extraction of the data symbols from the received signal. There are various Transformation techniques such as, Discrete Fourier Transform (DFT) [19], Fast Fourier Transform (FFT) [20] and Discrete Cosine Transform (DCT) [21].

In DFT-OFDM systems, the frequency domain OFDM symbol, $X(k)$, is first transformed into the time-domain OFDM symbol, $x_t(n)$, using an N-point inverse discrete Fourier transform (IDFT). IDFT transformation reduces the system complexity of parallel symbols and removes any pulse shift that occurred in the modulation process [22]. The IDFT equation is given below:

$$x_t(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k) e^{\frac{j2\pi kn}{N}}, \tag{20}$$

where $j = \sqrt{-1}, N$ is the DFT length, $X(k)$ is the frequency–domain OFDM symbol for $k = 0, 1, \cdots, N-1$, and $x_t(n)$ is the resultant time-domain OFDM symbol. At the receiver, discrete Fourier transform (DFT) is used to convert the time-domain signal back to the frequency domain signal, $Y(k)$. The DFT transformation is given by:

$$Y(k) = \frac{1}{N}\sum_{n=0}^{N-1} y(n) e^{\frac{-j2\pi kn}{N}}, \tag{21}$$

where $y(n)$ is the received signal after removing the CP. The FFT is an algorithm to compute the DFT efficiently, and it does not have a distinct equation of its own. The FFT algorithm computes the same result as the DFT but in a more efficient manner. The equation remains the same, but it reduces

the number of computations required from $O(N^2)\ to\ O(NlogN)$, where N is the number of data points.

DCT-OFDM is similar to DFT, and the operation takes place in a similar manner with inverse operation at the transmitter end and forward operation at the receiver end. One fundamental difference lies in their basis functions. DFT employs a mix of sine and cosine functions as its basis, allowing it to handle both amplitude and phase information, resulting in a complex output. On the other hand, DCT exclusively uses cosine functions, resulting in a purely real output, thus lacking phase information, which inherently offers symmetry in its output. This could reduce complexity in signal processing and imbalance in the in-phase/quadrature imbalance. Additionally, DCT concentrates the signal energy into lower frequency components, which can be advantageous in certain scenarios, such as image and audio transmissions over OFDM systems. At the transmitter, the IDCT is given by:

$$x_t(n) = \frac{1}{N}\sum_{k=0}^{N-1} \alpha(k)X(k)\cos\left(\frac{\pi k(2n+1)}{2N}\right), \tag{22}$$

where:

$$\alpha(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 0 \\ \sqrt{\frac{2}{N}}, & k = 1, 2, \dots, N-1 \end{cases}. \tag{23}$$

the forward DCT at the receiver is given by:

$$Y(k) = \alpha(k)\sum_{n=0}^{N-1} y(n)\cos\left(\frac{\pi k(2n+1)}{2N}\right), \tag{24}$$

DFT is more commonly used in OFDM compared to DCT due to its capacity to handle both the magnitude and phase information of signals.

2.2.4. Wireless Channel

A simplified wireless channel is well specified in [7], [23]. The time-domain received signal over a wireless channel can be defined as:

$$y_{cp}(n) = x_{cp}(n) \otimes h(n) + w(n), \tag{25}$$

where $y_{cp}(n), x_{cp}(n), h(n), w(n)$, and $\otimes$ are the received signal, transmitted signal, channel impulse response, AWGN, and convolution operator, respectively, in the time domain. The frequency–domain received signal is obtained by DFT transformation of (5) as:

$$Y(k) = X(k) \cdot H(k) + W(k), \tag{26}$$

where $Y(k), X(k), H(k), W(k)$, and $\cdot$ are the received signal, transmitted single, channel impulse response, AWGN, and element-wise product, respectively, in the frequency domain. AWGN with zero mean Gaussian distribution and uniform spectral density is often used to model the additive noise (for example, electric noise, thermal noise, interference) in the OFDM system. Therefore, the AWGN, $w(n)$, from Gaussian distribution with zero mean and standard deviation, is given by:

$$w(n) = \begin{cases} \sigma \cdot \mathcal{N}_N(0, \sigma), For\ real \\ \sigma \cdot [\mathcal{N}_N(0,1) + j\mathcal{N}_N(0, \sigma)], For\ complex\ ' \end{cases}, \tag{27}$$

where $\mathcal{N}_N(0, \sigma)$ is the Gaussian noise vector with 0 mean and standard deviation $\sigma = \frac{N_0}{2}$, and N_0 is the power spectral density.

Signal transmitted over wireless channel experience reflection, diffraction, and scattering, leading to multiple versions of the same signal arriving at the receiver with different amplitude, phase, and delay. This phenomenon is known as the multipath fading effect and can be modelled as linear finite impulse-response (FIR) filter give as [24]:

$$y_t = \sum_{l=0}^{L-1} h_l \cdot x_{(t-l)}, \text{where}, h_l = \sum_{K=1}^{K} \sqrt{\Omega_k} z_k\, sinc\left(\frac{\tau_k}{T_s} - l\right), \tag{28}$$

vector Ω and τ are the power-delay profile (PDP) of fading process, and z_k and T_s are the complex-valued variable and sampling period of the discrete signal, respectively. L are the filter length chosen,

such that $|h_l|$ is small for $l \geq L$ or $l < 0$. The effect of different channels on frequency–domain symbols of the proposed 16-QAM OFDM system is shown in Figure 3.

For the multipath Rayleigh Fading channel, the real and imaginary part of z_k are independent and identically distributed (IID) Gaussian random variables, therefore $|z_k|^2$ follows the Rayleigh Fading distribution. K is the number of paths in multipath fading channel and if $K = 1$ it is said to be flat fading with only on path, whereas, if $K > 1$ it is frequency selective fading with multipath interference. For the experiment, we have considered simple AWGN channel and Rayleigh Fading channel for the OFDM-based image communication system.

2.2.5. Channel Estimation

Channel estimation is crucial to obtain information about the Channel Impulse Response (CIR). Pilot-aided channel estimation is often used at the receiver by inserting known data sequences (pilots) in the form of block, comb, or scatter patterns to sample the channel distortions [25], [26]. These pilots are recognizable at the receiver and are either constant or low auto-correlation sequences, such as Zadoff–Chu sequences. For channel correction, the received bit sequence is compared with the known ones at the receiver to estimate the CIR, which is then equalized to mitigate the channel distortions [25], [27].

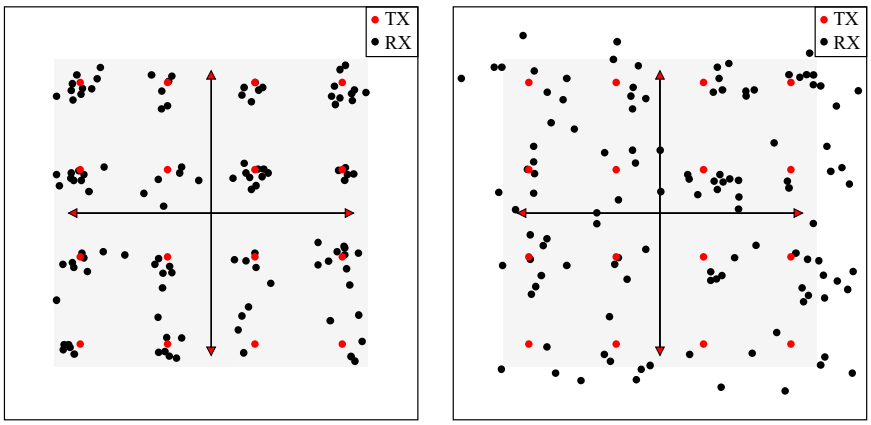

Figure 3. Effect of various channel models. System with 16-QAM Modulation using LS Channel Estimation at E_b/N_0 15 dB over (a) AWGN and (b) Rayleigh Fading Channel

The most common pilot-aided channel estimation method is the least square (LS) technique. The LS method increases the sampling rate of the channel frequency responses at the pilot subcarriers, and then uses interpolation, such as linear, spline, or cubic, to estimate the Channel Impulse Response (CIR) at the data subcarriers [7], [23], [26]. The LS estimator of the channel is given by:

$$\hat{X}(k) = \frac{Y(k)}{\hat{H}_p^{LS}}, \tag{29}$$

$$\text{where, } \hat{H}_p^{LS} = \mathcal{F}\left(\frac{Y_p}{X_p}\right). \tag{30}$$

$\hat{X}(k)$is the estimated channel, $\hat{H}_p^{LS}$ contains the LS estimate, $Y_p \in \mathbb{C}^{N_p \times 1}$ is the received pilots, $X_p \in \mathbb{C}^{N_p \times N_p}$ is a matrix with known transmitted pilot symbols on its diagonals, and $\mathcal{F}(\cdot)$ is the interpolation operation, such as linear, cubic, spline, etc. [25]. The LS algorithm is simple because it disregards noise and ICI, therefore, it is calculated with minimal complexity without any channel statistics knowledge; however, it results in high means square error (MSE).

Minimum Mean Square Error (MMSE) channel estimation technique utilizes the second-order statistics of the channel conditions (prior channel knowledge) and performs channel estimation [28], [29]. It uses channel auto convenience to reduce the MSE, as shown below:

$$\hat{H}_p^{MMSE} = R_H[R_H + (X_p X_p^H)^{-1}]^{-1}\hat{H}_p^{LS}, \tag{31}$$

where R_H is the frequency–domain correlation matrix at the pilot-symbols with $R_H = E[H_p {H_p}^{\dagger}]$ and the channel coefficient matrix can be obtained via interpolation [28]. In our experiment we have considered both LS and MMSE channel estimation with linear interpolation. Linear interpolation utilizes the frequency response of two neighboring pilots to estimate the frequency response of the data subcarriers in between them, therefore, it has simple computational complexity [28]. Overall, MMSE channel estimation outperforms LS estimator, especially in lower E_b/N_0 regions. However, as the number of subcarriers in an OFDM system increases, the computational complexity of the MMSE estimator also increases due to the computations required for matrix inversions [30].

2.3. DL for Image Classification Task

2.3.1. DL Basics

The Artificial Neural Network (ANN) can learn from examples and use that knowledge to classify new, unseen data. This is referred to as a *classification* task and ANNs are particularly well-suited to this task [31], [32]. The capability to classify new unseen data (known as *generalization*) make ANNs highly effective in solving classification problems where the desired outcome cannot be pre-determined [32].

ANNs consist of input and output layers, with one or more hidden layers made up of interconnected nodes called neurons. The most common type of layer is the fully connected layer, where every neuron is connected to every other neuron in the previous and next layer. The *activation function* determines the transmission of information between these neurons in the layers. To effectively train an ANN, a loss is calculated between the predicted outputs and the actual outputs. The training algorithms tracks and updates the weights of the neurons to minimize the loss over time, using *optimization* techniques. Moreover, *backpropagation* algorithm is used to propagate error information from the last layer to the first layer for weight modification and train the model in iteration or epoch. During training, the model assesses its performance on validation data (which is distinct from the training data) to fine-tune the hyperparameters and improve the model architecture. If the model is not trained enough, it is referred to as *underfitting*, meaning that it did not learn the training set well enough. On the other hand, *overfitting* occurs when the network has learned the training set too well, causing it to struggle with new data [33]. Different regularization techniques are adapted to mitigate the underfitting and overfitting problem. The final evaluation of the model's generalization is performed on unseen data, known as testing or inferencing.

A Deep Neural Network (DNN) is a type of ANN with multiple hidden layers. The term "deep" refers to the number of sequential layers within the network, indicating the number of times the input data passes through the transfer functions of sequential layers. This results in a more complex model

capable of capturing and representing high-level abstractions and patterns in the data. DNNs are designed to learn and make predictions based on large amounts of complex data, by processing and transforming the data iteratively through multiple hidden layers [4].

2.3.2. Convolutional Neural Network

Convolutional neural network (CNN) is a type of DNN that was introduced in [34] to provide an efficient learning method for images. The CNN architecture mostly consists of a convolutional layer, pooling layer, activation function, fully connected layer, and output layer. The convolutional layer is used for extracting features from input data. It utilizes a filter (represented by a small weight matrix) that slides over the input data, performing element-wise multiplication and addition to produce a set of *feature maps*. These filters have adjustable weights, learned through backpropagation, and are used to recognize local patterns in the input data. The convolutional layer is effective in learning features that are invariant to shifts in the input, reducing model complexity and improving model generalization. The size, stride, and padding of the filters can be adjusted to control the spatial dimensions of the output feature maps. By using multiple convolutional layers, the model can learn hierarchical representations of the input data, making it effective for tasks such as image classification and object detection.

A convolution layer consists of a set of F filters known as *depth*. The weights of the filters are $W^f \in \mathbb{R}^{a\times b}, f = 1, \dots, F$. These weights generate feature map, $Y^f \in \mathbb{R}^{n'\times m'}$ from an input matrix $X \in \mathbb{R}^{n\times m}$ as per the given convolution:

$$Y_{i,j}^f = \sum_{k=0}^{a-1}\sum_{l=0}^{b-1} W_{a-k,b-l}^f \cdot X_{1+s(i-1)-k,1+s(j-1)-l}, \tag{32}$$

where s is called *stride* with value ≥ 1, $n' = 1 + \lfloor (n + a - 2)/s \rfloor$ and $m' = 1 + \lfloor (m + b - 2)/s \rfloor$. It is assumed X is zero-padded which denotes, $X_{i,j} = 0$ for all $i \notin [1, n]$ and $j \notin [1, m]$. Furthermore, in image processing, X is usually a three-dimensional tensor; therefore, the weights are three-dimensional as well and applies on all input channels simultaneously. After

obtaining the feature maps from a convolution layer, a *pooling layer* performs down-sampling on the feature maps obtained from the previous convolution layer. The purpose of pooling is to reduce the spatial dimensions of the feature maps, making the network more computationally efficient, and invariant to translations in the input image. Therefore, the pooling layers partitions Y into $p \times p$ regions with values representing the most important information from the set and computes a single output value using pooling techniques such as max-pooling and average-pooling, which, respectively, take the maximum or average value from the features in the pooling window as the output before passing it to the next layer.

The *activation function* (non-linearity layer) in CNN introduces non-linearity into the network, which benefits the stacking of multiple layers in a network. The activation function is typically applied to each element of the input vector and is used to determine the output of the network, often in a binary format, such as 0 or 1. The resulting non-linearity enables deep neural networks to model complex, non-linear systems and provides a powerful tool for pattern recognition tasks. Previously, *tanh* and *sigmoid* activation functions were popular; however, with the understanding that most data are centered around zero, newer techniques, such as rectified linear units (ReLU) and exponential linear units (ELU), have become prevalent because they offer non-linear behavior near zero [35], [36].

A *fully connected layer* (also known as a dense layer) in CNN is connected to all neurons in the previous layer and next layer. It takes the output from the convolutional layers and applies a matrix multiplication operation, followed by a bias offset, to produce a final prediction for the image classification. The final layer in CNN is usually a fully connected layer, which produces a vector of values, each representing a probability for a given class (category that data sample belongs to). More details on CNN can be referred to in [4].

Augmentation Techniques: Improving the generalization of the model and mitigating the overfitting problem is a major challenge in the deep learning field, including in CNN models. Data augmentation is a method that addresses this challenge by improving the sufficiency and diversity of

the training data. It is used to artificially expand the size of the training dataset by generating modified versions of images in the dataset, creating a more diverse and comprehensive representation of the real-world data. The goal is to make the model more robust to variations in the data and reduce overfitting by increasing the diversity of the training set. Common techniques include random rotation, scaling, flipping, cropping, and adding noise to the images. Additionally, data augmentation helps to reduce bias in the model by creating new data that helps the model better capture the nuances and variations in the real-world data. LeNet-5 [37] was one of the first CNN applications that utilized image data augmentations for handwritten digit classification. The authors in [38] presents various existing methods and promising developments of data augmentation in DL.

III. RELATED WORK

3.1. OFDM-Based Image Communication System

Orthogonal Frequency Division Multiplexing (OFDM) is a standard communication system that has been adapted across various technologies and standards. In this section, we will discuss a few of the related works that utilize OFDM communication system for image transmission. The authors in [39] investigated the transmission of gray scale images using an OFDM system with various even-ordered M-QAM modulations (from 4-QAM to 256-QAM) and channels with different fading and shadowing parameters. The results demonstrate that lower order M-QAM modulation provides a higher quality of recovered image due to a lower Bit Error Rate (BER). Authors in [40] evaluated the performance of Quadrature Phase Shift Keying (QPSK), 4-QAM, 16-PSK, 16-QAM modulation for an OFDM-based image transmission system over an Additive White Gaussian Noise (AWGN) channel. Their results showed that 16-QAM provided the best performance in terms of image quality and BER. Authors in [41] investigates the integration of Discrete Wavelet Transform (DWT) and Fast Fourier Transform (FFT)-based OFDM with QPSK modulation for the transmission of grayscale images in AWGN and Rayleigh channels. They propose the integration of two adaptive filtering techniques, least mean squares (LMS) and recursive least squares (RLS), to reduce BER and concludes LMS outperforms RLS in terms of noise pattern detection and efficient recovery of the modulating signal. Authors in [42] proposed a new technique for transmitting images over underwater time-dispersive fading channels, called Progressive Zero-Padding (PZP)-OFDM using QPSK modulation, which improves the BER in underwater environments. Authors in [43] analyzed the BER performance of image transmission in non-OFDM and OFDM communication systems under various channel conditions (AWGN, Rician and Rayleigh Fading channels) and modulation techniques (Binary-PSK, QPSK, 16-QAM). Their results show that OFDM has better performance compared to non-OFDM, and overall BPSK in AWGN had the lowest BER. Authors in [44] have implemented a real-time practical OFDM system for image transmission using Raspberry Pi (RPi) and Pluto software-defined radios (PlutoSDR). The performance of the system was evaluated by

comparing the BER for different modulation schemes, including BPSK, QPSK, 16-PSK, and 256-PSK, where BPSK had the best performance. Authors in [45] investigated the impact of channel estimation for OFDM based image transmission through AWGN channel with various modulation schemes (BPSK, QPSK, 8-PSK, and 16-QAM), and concluded that Least Square (LS) method significantly improved the quality of the restored images. Authors in [46] employed two types of OFDM systems, FFT-OFDM and DWT-OFDM, to transmit compressed images using Discrete Cosine Transform (DCT), Wavelet (WAV) transform, and compressive sensing methods using different modulation techniques (BPSK, QPSK, 16-QAM, and 64-QAM) over AWGN channel. Their results showed that 16-QAM DWT-OFDM outperformed FFT-OFDM with lower Mean Square Error (MSE) while avoiding the use of a cyclic prefix. Authors in [47] compared different modulation techniques (QPSK, 16-QAM, and 64-QAM) for transmitting grayscale and Red Green Blue (RGB) images over OFDM image communication system under varying channel conditions (AWGN and Rayleigh Fading), and filters (Wiener, Median, and No Filter). Their results show QPSK in AWGN channel had better recovered images using a median filterer for a lower Signal-to-Noise Ratio (SNR) value. Authors in [48] investigated the effect of improved tone reservation (ITR) on the peak-to-average power ratio (PAPR) of a 16-QAM-OFDM system for transmitting images over AWGN channel. Their results show that ITR can effectively reduce PAPR, leading to improved image quality and fewer transmission errors. The authors in [49] and [50] performed a comprehensive analysis of image communication under various conditions in Single User Multiple Input Multiple Output (SU-MIMO) and Massive MIMO OFDM systems, respectively. They evaluated the effect of various channel conditions (AWGN, Rayleigh), modulation techniques (BPSK, QPSK, 8-PSK, 16-PSK, 32-PSK, 64-PSK), antenna configurations, number of users, and transformation techniques (FFT, Fractional Fourier Transform, DWT, DCT) on image communication in 5G networks. The results from the former study indicated that using the DCT transformation technique improved the BER compared to FFT and increasing the number of received antennas led to higher quality recovered images. The later study indicated that DWT outperformed FRFT and FFT in terms of BER; however, as the number of users continued to increase in the Massive MIMO system, the quality of

the recovered image decreased irrespective of the transformation technique used. Additionally, both studies showed that lower-order PSK modulation produced better results overall.

There have been several studies of image transmission across OFDM communication system with DL based channel coding and decoding techniques for CV applications. For instance, [51] first proposed deep joint source-channel coding (JSCC) for wireless image transmission by combining the source and channel coding into a single auto-encoder structure. It optimized source compression and error correction coding through back-propagation, which outperformed conventional schemes (separate source and channel coding) under AWGN and Rayleigh flat fading channels. This scheme was further extended to larger neural networks [52], feedback network [53], progressive transmission [54], and additional attention modules [25,26]. Motivated by model-based deep learning, DL-based JSCC were also extended to OFDM image communication system with fading channels in [27,28]. The former study fed the neural network-encoded image as frequency domain OFDM baseband symbols and used adversarial measures to further improve the quality of reconstructed images. The latter utilized double attention mechanism to better map image features to subchannels and ensure important features were transmitted over the high-quality subchannels. Both [27,28] outperformed conventional separate coding schemes.

The aforementioned research works focused on OFDM-based image communication systems; however, they have not considered, multi-tier analysis of the blocks in the OFDM PHY that encompasses higher-order modulation schemes, various channel models and channel estimation. Additionally, it is not yet practical to adapt PHY as a DL architecture. Furthermore, DL application of the received images for a specific task has not been considered, only system performance and image quality has been studied. The authors in [59]–[61] attempted to mitigate this gap and considered DL-based applications, and analyzed the performance of DL models on received images from various OFDM communication systems. Authors in [59] determined that images from FFT-based OFDM were better for image transmission for cloud-based DL applications, as opposed to DCT-based systems. Authors [60] investigated the impact of channel correction on DL models in the

OFDM-based image communication system and showed that channel correction had a major impact on the quality of recovered images, although DL models still produced acceptable results without it (with a lesser degree of improvement). Authors in [61] evaluated the impact of higher order M-QAM on the performance of the DL model. Their results showed that the DL model accuracy was lower on images from higher order M-QAM compared to lower order M-QAM, but the overall accuracy was still relatively high and suitable for applications that are unaffected by small fluctuations in DL accuracy. The studies have investigated the DL-based applications for image communication systems; however, they have not considered multi-tier analysis of the OFDM communication system, including fading channels, modulation techniques, and channel estimation methods. Additionally, they have not evaluated different DL models, nor have they exploited any model training techniques, such as data augmentation, to improve the DL model accuracy. Therefore, for this study we consider multi-tier analysis of various OFDM-based image communication systems and robustness analysis of multiple DL models on images recovered from these systems to enable CV applications in ITS. Therefore, the focus of this thesis is on the PHY of OFDM communication systems, and the MIMO systems are left for future work.

3.2. Robustness of DL Models on Image Perturbation

Robustness analysis of DL models on noisy images from different communication systems for CV applications are also considered in this study. In this section, we will discuss the related work based on robustness of DL models on images with different types of noise perturbation. Despite the success of Deep Neural Network (DNN) with complex and high-dimensional image data, they are still not robust against image perturbations. Authors in [62] were the first to discover the vulnerability of DNNs to certain input perturbations, which result in significant discontinuities in their outputs. These perturbations were named *adversarial examples* and were found to cause a wide range of DL models to misclassify, regardless of the different model architecture or training data. The authors also examined several interesting properties of neural networks, including their capacity for learning intricate functions, the existence of adversarial examples that can confuse the network, and their

ability to generalize and extrapolate. Authors in [63] carry out further studies on adversarial models and present defense systems, such as adversarial training, input transformations, and robust optimization to minimize the effect of adversarial examples.

Among various image perturbation, visible noise on digital images is a type of perturbation resulting from perceptible alterations made to the image pixels, and it can have an impact on DL-based image classification tasks. The introduction of visible noise in digital images during the processes of its acquisition, encoding, transmission, and processing is an inherent consequence of the utilization of electronic components, particularly sensors and actuators [64], [65]. A comprehensive overview of digital image noise models can be found in [65], which highlights various types of noise that can impact digital images, such as additive Gaussian noise, quantization noise, color quantization with dither, salt–pepper noise, Rayleigh noise, gamma noise, uniform noise (white noise), and Poisson noise. The paper also discusses methods for modeling and analyzing these noises using probability density functions and statistical models, and common techniques for removing them, such as spatial filtering, frequency filtering, and wavelet denoising.

DL has made remarkable advancements in various domains, particularly in image classification. However, the accuracy of these models degrades when images are subjected to such distortions and noise [66]. Authors in [66] evaluated various factors that contribute to image quality, including resolution, noise, and blur, and evaluated their individual impacts on DNNs. The outcomes indicated that noise and blur have a stronger impact on DNNs than other factors. The authors also presented strategies, such as pre-processing and data augmentation, for improving DNN performance on low-quality images. Authors in [67] evaluated the robustness of Convolutional Neural Networks (CNN) to various types of image degradations, such as Gaussian noise, salt and pepper noise, Joint Photographic Experts Group (JPEG) compression, and motion-blur. The results showed that CNNs are generally more robust to image degradation compared to traditional methods, but they still have limitations. To address these limitations, the authors proposed a CNN architecture called a capsule network, which was shown to be more robust to degradation. Authors in [68] investigate the effect

of color information on the robustness of CNNs for image classification tasks. They have compared the performance of CNNs trained on color images and grayscale images when subjected to degradation, such as Gaussian noise and JPEG compression. The results show that CNNs trained on color images are more robust to degradation compared to those trained on grayscale images, due to the additional cues provided by the color information. In CV tasks, CNNs have shown outstanding performance when applied to identically and independently distributed data. However, they are vulnerable to changes in data distribution [62] and color corruption [69] as they only use local features [70], [71]. Authors in [72] proposed Vision Transformers (ViT) to mitigate this challenge by considering global image context and reduced bias towards local textures, resulting in robustness towards occlusions. Although the aforementioned studies have shown promising results in improving the accuracy and robustness of CNN models against various noise perturbation on the images, they have not considered various communications systems and the performance of DL models on images with heavy noise perturbations introduced by these communication systems. Therefore, in current study we focus on the performance of CNN models against noise from various OFDM-based image communication systems for CV applications in ITS.

IV. RESULTS AND DISCUSSION

4.1. Evaluation Matrices

4.1.1. Image Communication System Analysis Matrices

BER: The Bit Error Rate (BER) is a standard performance metric in digital wireless communication systems to measure the quality of the reconstructed signal at the receiver. It calculates the ratio of incorrect bits to the total number of bits transmitted, typically expressed as a ratio. The BER is a crucial measure in determining the maximum achievable data rate for a given system design, considering the presence of noise and other sources of interference in the channel.

PSNR: Peak Signal-to-Noise Ratio (PSNR) is an important metric to evaluate the quality of recovered images in image communication systems. Unlike BER, which indicates the performance of the communication system in bit level, PSNR provides a measure of the perceptual quality of the images by comparing the received image with the original transmitted image. PSNR is a full-reference image quality measure that uses an objective approach based on explicit numerical criteria, such as comparisons with ground truth or prior knowledge expressed in terms of statistical parameters and tests. The PSNR value is measured in Decibels (dB) and approaches infinity as the mean square error (MSE) approaches zero, implying that a higher PSNR value indicates a higher-quality image. A low PSNR value, on the other hand, signifies significant numerical differences between the original and recovered images [73].

MS-SSIM: The MS-SSIM (multi-scale structural similarity index) is another full-reference image metric developed to compare the quality of an input image to that of a reference image with no distortion. Developed by Wang et al. [73], the MS-SSIM is correlated with the quality perception of the human visual system (HVS). It works by aggregating the inner similarity indexes obtained from multiple spatial scales (resolutions) to estimate the overall similarity between the input and reference images. MS-SSIM is an improved metric of single-scale SSIM (SS-SSIM), resolving the limitations of being only suitable for limited visual contexts and unable to account for the vast visual

diversity [74]. MS-SSIM is also measured in dB and the values are usually represented in log scale to observe high quality results, $-10log_{10}(1-M)$, where M is the MS-SSIM score.

The full-reference image quality measures such as PSNR, MS-SSIM, require a reference image with no distortion to calculate an objective quality score. However, in cases where images are generated, or no reference image is available, non-reference image quality can be used. Perception-based Image QUality Evaluator (PIQUE) is a non-reference perception-based evaluation matric that uses arbitrary distortion for image quality measurement in natural images [75]. PIQUE value is in the range of [0 to 100], and the score is interpreted in steps of twenties, for example, 0 to 20 means excellent and 81 to 100 means bad quality of the image.

4.1.2. DL Performance Analysis Matrices

In DL, performance analysis is critical to understand the performance of the DL model, in tasks such as classification or regression. Various metrics help in evaluating the performance of these models.

Accuracy is a fundamental DL performance metric. It quantifies the overall correctness of the model by calculating the proportion of true predictions among the total number of instances.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} . \tag{33}$$

A true positive (TP) is an outcome where the model correctly predicts the positive class, while a true negative (TN) is an outcome where the model correctly predicts the negative class. False positives (FP) are instances where the model incorrectly predicts the positive class, and false negatives (FN) are instances where the model incorrectly predicts the negative class.

Precision and recall (sensitivity) are particularly crucial in imbalanced datasets. Precision refers to the proportion of true positive predictions among all instances that the model predicted as positive. It represents the reliability of the model in predicting positive instances. Recall, or sensitivity, indicates the ability of the model to correctly identify positive instances from actual

positives, helping understand the model's coverage of the positive class in the dataset.

$$Precision = \frac{TP}{TP + FP}, \tag{34}$$

$$Recall = \frac{TP}{TP + FN}. \tag{35}$$

F1-score uses precision and recall evaluating the performance of the model. It is particularly useful when the class distribution is imbalanced, ensuring that both false positives and false negatives are considered in the evaluation. The F1 score ranges from 0 to 1, where a score of 1 indicates a perfect model and a score of 0 indicates a poor model. It gives more weight to lower numbers; thus, a model can only achieve a high F1 score if both precision and recall are high.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \tag{36}$$

There are various visual representations to evaluate the Models' performance. A confusion matrix is a comprehensive table used for performance evaluation. It comprises of all four elements: TP, TN, FP, and FN. This matrix provides a more detailed view of the model's performance, helping identify the types and sources of errors made.

The generalizability of the DL model can also be observed using the t-distributed stochastic neighbor embedding (t-SNE) [76] visualization algorithm. It is a probabilistic dimensionality reduction algorithm which maps high-dimensional feature points into low-dimensional feature space (typically 2D or 3D) while preserving the similarities. Features close to each other in the high-dimensional space are mapped closer to each other in the lower-dimensional space with high probability. Points that are close together in the t-SNE plot are likely to have similar features or attributes (in terms of color, texture, or shape), whereas points that are far apart are likely to be dissimilar.

Gradient-weighted Class Activation Mapping (Grad-CAM) heat map is another visualization algorithm that is used to visualize the attention region for a CNN based DL models while making

predictions [77], [78]. It uses the gradient of the final convolutional layer with respect to the predicted class to weigh the layer activations, and then averages these weighted values over the spatial dimensions to obtain a class activation map. The map shows the crucial region of the image that is used to make the prediction by the CNN model.

In the following section, we present the performance analysis of the study. First, we discuss the performance of various OFDM image communication systems and the effect of PHY impairments on image quality. Then, we discuss the environment considered for the downstream DL application, and finally the robustness of DL against these impairments for downstream CV applications.

4.2. Performance Analysis of OFDM Image Communication System

In this section, we discuss the performance of various OFDM image communication systems and the effect of PHY impairments on image quality.

4.2.1. Transformation and QPSK Modulation

For one of the simulation analyses, we have considered FFT and DCT based OFDM implementation for their low computational complexity. Figure 4. illustrates quality of the recovered images in terms of PSNR (dB). The image communication system based on DCT-OFDM has

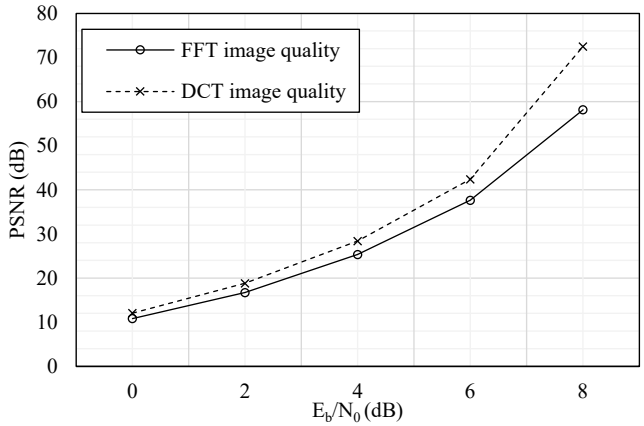

Figure 4. PSNR of recovered images using different transformation techniques. Communication system is OFDM using QPSK modulation over AWGN channel.

preserved higher quality of the recovered images than FFT-OFDM. At E_b/N_0 8dB, around 24% improvement in image quality was observed by using DCT Transformation.

Though DCT transformation retains better image quality as per analysis, it is not a standard technique used in OFDM. DFT/IDFT or FFT/ IFFT are standard transformation techniques for frequency-time domain conversion in OFDM communication system. These methods are well-established and often used in practice due to their effectiveness in managing multi-carrier modulation. They aid in creating orthogonal subcarriers in the OFDM for efficient use of the available spectrum, which is essential for reducing Inter-Carrier Interference (ICI). They also facilitate the system's robustness against multi-path propagation effects and aid in coping with frequency-selective fading. For the rest of the simulation IDFT/DFT will be utilized as transformation techniques due to the aforementioned benefits.

For modulation techniques, QPSK has been initially utilized in the simulation. In QPSK, the phase of the carrier signal is varied to transmit data, with each symbol carrying 2 bits of information. This is a low data-rate transmission and may not be suitable for downstream IoT applications that demand high data-rate for real time communication. Therefore, for the rest of the experiment we will be utilized higher order QAM for high data-rate and advantages mentioned in Section 2.2.2.

4.2.2. Source Coding and Channel Model

Images can be compressed using either lossless or lossy methods. While lossless compression aims to preserve image quality, lossy compression ensures higher compression savings [10]. However, increased compression savings have adverse effects on DL performance [8]. Therefore, instead of compression savings, this study mainly focuses on evaluating the image quality recovered from OFDM systems using FLC and VLC. In the experiment, for FLC, the corresponding decimal values of each pixel are coded to a fixed 8-bit length codeword. On the other hand, for VLC, the decimal values are coded as variable length codes according to the length of their respective

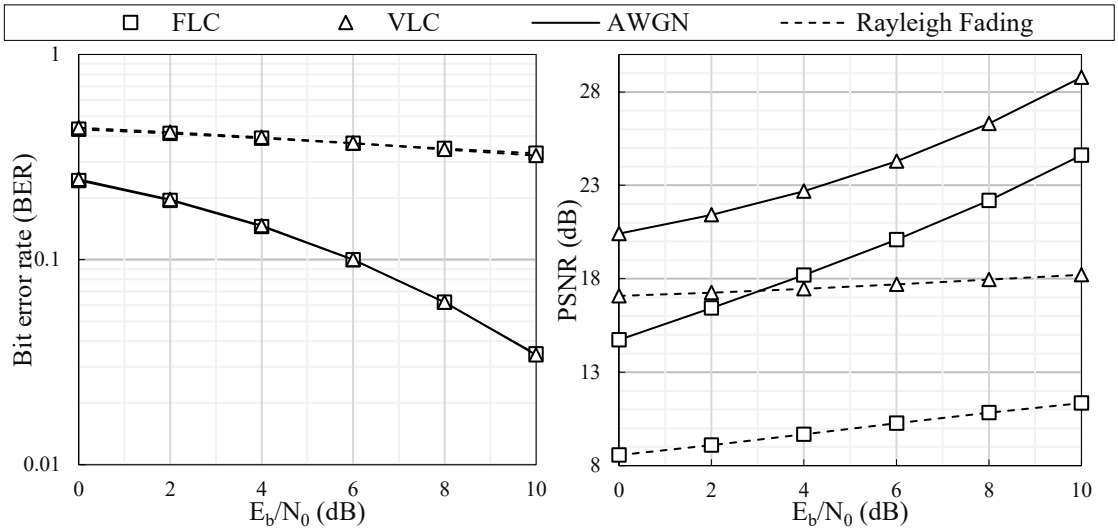

Figure 5. Performance analysis using different source coding schemes. (a) is BER, and (b) is PSNR of 16-QAM OFDM.

codewords. The information about the length of each codeword for every pixel is transmitted as side information to the receiver for successful decoding. Although transmitting side information may increase the overall system overhead, it avoids the synchronization problems commonly experienced in standard VLC coding, such as Huffman coding.

The BER analysis (on a logarithmic scale) for the proposed OFDM-based image communication system over both AWGN and Rayleigh fading channels is illustrated in Figure 5(a). Over the AWGN channel, there is a significant drop in BER as E_b/N_0 increases. At an E_b/N_0 of 10dB, the system achieved the lowest BER of 0.03. Throughout the experiment, FCL and VLC achieved identical BER. In contrast, for the system over the Rayleigh fading channel, the BER does not decrease significantly at higher E_b/N_0 values compared to the AWGN channel. The lowest BER recorded was 0.32 at an E_b/N_0 of 10dB. Similar to AWGN channel, both FLC and VLC yield identical BER values throughout the experiment. The reason for that could be multifold. Firstly, BER is the rate of number of bits in error, which remain consistent regardless of the number of bits transmitted using the different source coding scheme. Secondly, although FLC and VLC inherently differ in terms of average code length, their efficiencies are comparable over the given channel conditions,

which did not particularly favor one coding scheme over the other. Lastly, the inherent characteristics of the AWGN and Rayleigh fading channels did not differentiate between the two different coding techniques, resulting in identical values.

The quality of the recovered images in terms of PSNR from the proposed systems over AWGN and Rayleigh fading channels is presented in Figure 5(b). Images recovered from the AWGN channel exhibit high quality, which further improves with an increase in E_b/N_0. Using VLC, as the E_b/N_0 increased, the PSNR improved by 41%, reaching a peak of 28.78dB at an E_b/N_0 of 10dB. A similar trend is observed with system using FLC, where the PSNR improved by 67%, attaining a maximum of 24.62dB at the same E_b/N_0. However, there is a significant difference between the image qualities from systems using FLC and VLC. Though the BER remain consistent between the two source coding schemes as observed previously, the error in the bits can drastically change a pixel value depending on its codeword position as discussed in Section 2.2.1., resulting in difference in image quality. Throughout the experiment, VLC consistently achieved better image quality compared to FLC, with improvement of highest 39% and lowest 17% observed at 0 dB and 10dB, respectively. Images from the communications system over a Rayleigh fading channel were heavily distorted with noise, which was observed in the PSNR analysis as well. At E_b/N_0 10dB, the highest PSNR of 18dB and 11dB was observed by the system using VLC and FLC, respectively. There was no significant improvement in image quality even with the increase in E_b/N_0. Throughout the experiment, the PSNR value improved only by 1dB and 2dB for VLC and FLC, respectively. Similar to the AWGN channel, there was a significant difference in image quality between FLC and VLC schemes. VLC achieved the highest improvement of 98% at E_b/N_0 0dB and lowest 64% at E_b/N_0 10dB, compared to FLC.

Figure 6. presents sample images recovered from various systems for visual analysis. Based on visual inspection, VLC retains better image quality compared to FLC, especially over the Rayleigh fading channel. Using FLC over Rayleigh fading channel, the information in the Region of Interest (ROI) is invisible at lower E_b/N_0 regions.

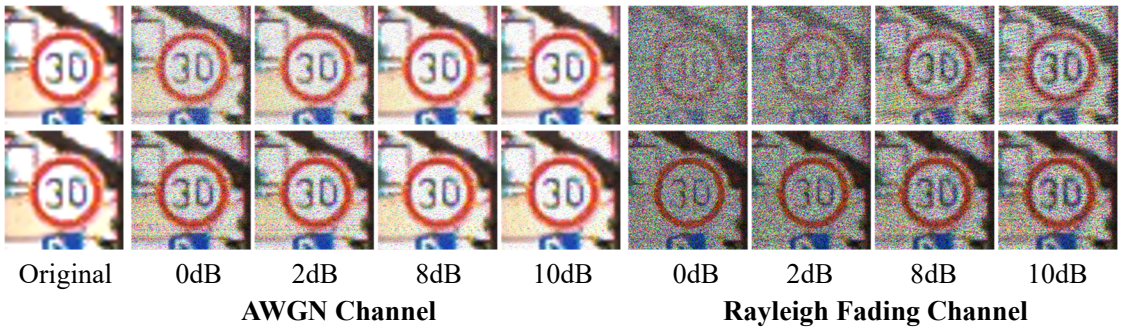

Figure 6. Sample images from communication systems. Images at various E_b/N_0 using FLC (top) and VLC (bottom) over the two channels is shown.

To summarize the parameters in the experiment, FLC and VLC are utilized for source coding. The modulation technique used is 16-QAM (rectangular constellation) with a symbol length of 4. The data stream for the system comprises 64 subcarriers, 16 CP, and 4 pilots (using comb-type pilot insertion). IDFT and DFT are used for the transformation techniques. LS with linear interpolation is employed for channel estimation, and zero-forcing is applied for channel equalization. For the channel model, AWGN and a Rayleigh fading channel are considered for wireless transmission.

4.2.3. Channel Correction

Another experiment was carried out to analyze the effect of channel correction on the OFDM communication system and recovered image quality. For this purpose, two OFDM systems were implemented for image communication. The first OFDM system was implemented with channel correction (withCC) using Least Square channel estimation technique. and second without channel correction (withoutCC). For the pilot insertion, comb-type pilot based estimation was used. As for the rest simulation parameters, 16-QAM for modulation, DFT/IDFT for transformation and simple AWGN channel for the channel model were utilized. The system's efficiency is being considered in terms of bit error rate (BER) and image quality metrics, such as PSNR and PIQUE.

For proposed OFDM based image communication systems, namely withCC and withoutCC, BER analysis is shown in Figure 7. The BER value for an image is obtained by taking the average of

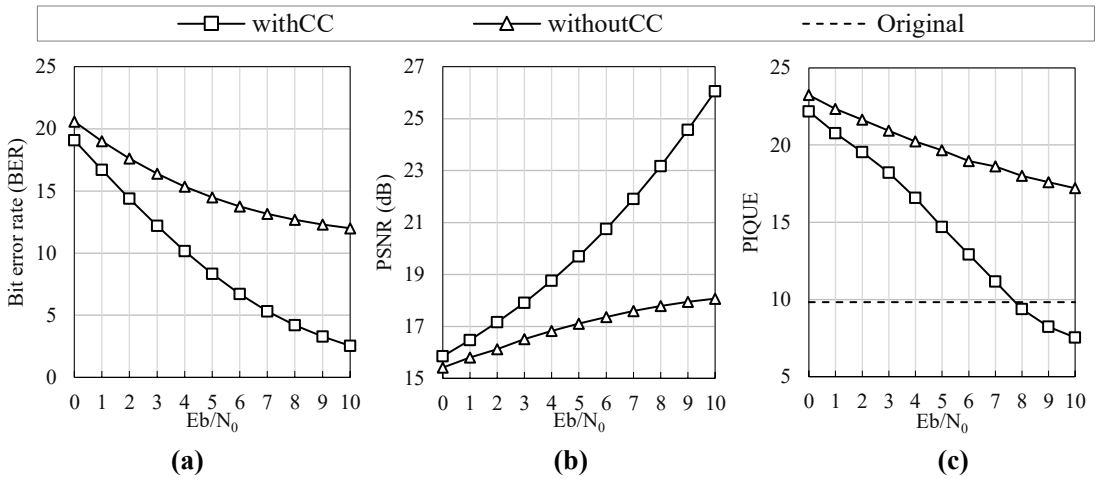

Figure 7. Performance analysis of system with and without channel correction. Analysis in terms of (a) BER, (b) PSNR, and (c) PIQUE in 16-QAM OFDM over AWGN channel.

OFDM data sub-carriers. The BER presented in Figure 7 (a). is the mean across all the images in test dataset. The BER of both systems improves with higher E_b/N_0. For $E_b/N_0=10$, there is up to 17% and 9% BER reduction compared to $E_b/N_0=0$ for withCC and withoutCC systems, respectively. However, such higher E_b/N_0 is not feasible in practice. When comparing the BER of the two systems, withCC has better signal quality than withoutCC system, due to the channel correction.

The BER measure does not provide any information about perceptual quality of recovered images. Therefore, in analysis we have considered PSNR and PIQUE to quantify quality of the images. Figure 7(b). shows the PSNR value, taken as the mean across all the images in test dataset. Similar to BER, the PSNR value also improves with higher E_b/N_0. In withCC system, there is a significant difference between the PSNR values for each E_b/N_0 as compared to withoutCC. The PSNR value increases up to 11 dB and 3 dB for withCC and withoutCC systems, respectively. Overall, withCC system is able to retain higher quality images than its counterpart. The measure of statistical features in terms of PIQUE of the proposed systems is shown in Fig.5 (b). The values are the mean across all the images in the test dataset. In general, when natural images datasets are constructed, conditions under which the images are acquired may differ, which sometimes result in image distortions. Therefore, Figure 7(c). also shows quality of the original images, which is used as

a baseline to compare other methods with it. Following the same trend as PSNR, PIQUE score improves with higher E_b/N_0 for both systems. For the given range of E_b/N_0, PIQUE is improved up to 15 and 6 for withCC and withoutCC systems, respectively. The withCC system achieves PIQUE same as original image after 7 dB whereas, withoutCC system does not reach closer to the original images throughout the experiments. In addition, for excellent quality of images, $E_b/N_0 \geq 3$ should be used when OFDM system has channel correction and $E_b/N_0 \geq 5$ should be used when there is no channel correction. For withCC systems' $E_b/N_0 \geq 8$ the recovered images have surpassed quality of the original pristine images. The reason is that PIQUE measure is designed for simple distortions resulting from image compression and additive noises, and not for the case when images have complex distortions occurred during communication.

Channel correction has a major impact on the quality of signal and recovered images even on simple AWGN channel. Furthermore, in OFDM communication system, channel correction is essential for mitigating the effects of multipath and frequency-selective fading, ensuring that the transmitted signals are accurately received by correcting distortions and interferences. It improves the system's adaptability, synchronization, and overall signal-to-noise ratio, resulting in more robust and reliable wireless communications. For the rest of the experiment, we utilized channel correction block, considering LS and MMSE channel estimation techniques.

4.2.4. Channel Estimation, M-QAM and Channel Models

For the next experiments we have considered various OFDM communication systems for image transmission. Within the system, we have considered higher order M-QAM modulation to attain a high data rate, at the cost of added noise in the images. Additionally, we have considered various channel models and channel estimation techniques to evaluate the noise and image quality under different system conditions. Multipath AWGN channel and Rayleigh Fading channel with AWGN were used as channel models to simulate real-world environmental conditions. Simple LS and more complex MMSE channel estimation were used to evaluate the trade-off between image quality and system complexity. The simulation parameters for the OFDM-based image

Table 3. Simulation parameters of the OFDM-based image communication system.

Parameters	Values
Source Coding	Variable Length Coding
Modulation Technique	M-QAM
Modulation Order	$M \in \{16, 32, 64, 128, 256, 512, 1024\}$
Transformation	IDFT/DFT
Length of Symbol (k)	$log_2(M)$
Subcarriers (S)	$n \times 2^n$
Cyclic Prefix (CP)	$S/4$
Pilots (P)	$CP/4$
Pilot Insertion	Comb-type
Channel Model	AWGN, Rayleigh Fading
Channel Estimation	LS, MMSE
Channel Equalization	Zero-Forcing

communication system considering, higher order M-QAM, different channel estimation and different channel model is given in Table 3.

To evaluate the performance of the OFDM-based image communication system, and output image quality, we will categorize the next discussion based on different channel models. This approach aims for clarity in presenting the performance outcomes in various environments. Specifically, the evaluation will be divided into two main sections: Evaluation over AWGN Channel and Evaluation over Rayleigh Fading Channel.

4.2.4.1. Over AWGN Channel

Figure 8 (a). illustrates the BER in logarithmic scale versus E_b/N_0 for the proposed OFDM-based image communication system over AWGN channel, using higher order M-QAM and different channel estimation techniques. The graph indicates that lower order M-QAM experiences a sharp drop in BER with higher E_b/N_0, characterized by the waterfall trend, whereas the drop in the higher order M-QAM is insignificant across all E_b/N_0 values due to the increased number of error bits. At E_b/N_0 20 dB, 16-QAM and 1024-QAM have reached their lowest mean BER of 2.25×10^{-5} and 0.05 using MMSE channel estimation, respectively. For channel estimation analysis, MMSE outperforms LS channel estimation across all M-QAM, and for 16-QAM at E_b/N_0 20 dB, the use of MMSE resulted in the highest improvement (approximately 40 times lower BER) compared to LS channel

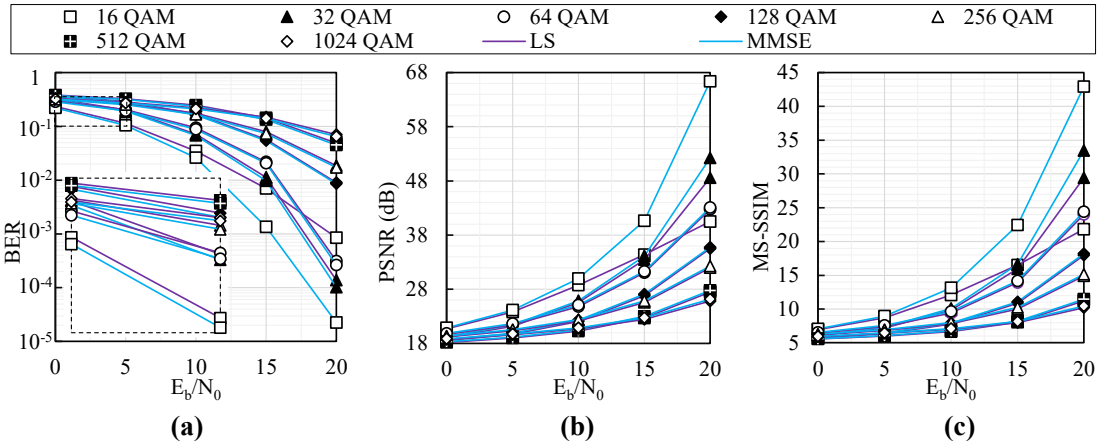

Figure 8. Performance analysis of various OFDM systems over AWGN channel. Analysis in terms of (a) BER, (b) PSNR, and (c) MS-SSIM with respect to various E_b/N_0.

estimation. Overall, the 16-QAM system using MMSE channel estimation has the lowest average BER and 1024-QAM system using LS estimation has the highest average BER.

The image− quality measure in terms of PSNR of the recovered images for the proposed systems is shown in Figure 8 (b). The PSNR values follow a similar pattern to the BER and improve with higher E_b/N_0; however, when using higher order M-QAM, the PSNR values decrease indicating the decline in image quality across all E_b/N_0. Additionally in high E_b/N_0 regions, the quality of the image improves significantly for lower order M-QAM with a steep upward incline compared to higher order M-QAM, which has a more flattened incline. At E_b/N_0 20 dB, 16-QAM and 1024-QAM have obtained their highest quality images with mean PSNR value of 66 dB and 26 dB using MMSE channel estimation, respectively. With respect to channel estimation, MMSE outperformed LS across all M-QAM, and 16-QAM at E_b/N_0 20 dB achieved highest improvement of 26 dB when utilizing MMSE channel estimation; however, with the higher order M-QAM, the improvement due to different channel estimation is less significant. MS-SSIM evaluation metrics (in logarithmic scale) have almost identical trends with PSNR, as illustrated in Figure 8(c). For higher E_b/N_0 the MS-SSIM improves significantly for lower order M-QAM, however, the improvement is slow for the higher order M-QAM. Lower order M-QAM has a higher MS-SSIM score and demonstrates significant

improvement with increased E_b/N_0, while higher order M-QAM does not exhibit a significant improvement in the image quality across E_b/N_0. At E_b/N_0 20 dB, 16-QAM and 1024-QAM have obtained their highest quality images with mean MS-SSIM values 43 dB and 10 dB using MMSE channel estimation, respectively. In terms of channel estimation, MMSE performs better than LS for all M-QAM (16-QAM at E_b/N_0 20 dB achieved an improvement of 20 dB using MMSE); however, for higher order M-QAM, the difference in performance between the two techniques is not substantial. Overall, the 16-QAM system with MMSE channel estimation had the best image quality retention and 1024-QAM with LS had the worst image quality retention throughout the experiments in terms BER, PSNR, and MS-SSIM evaluation matric.

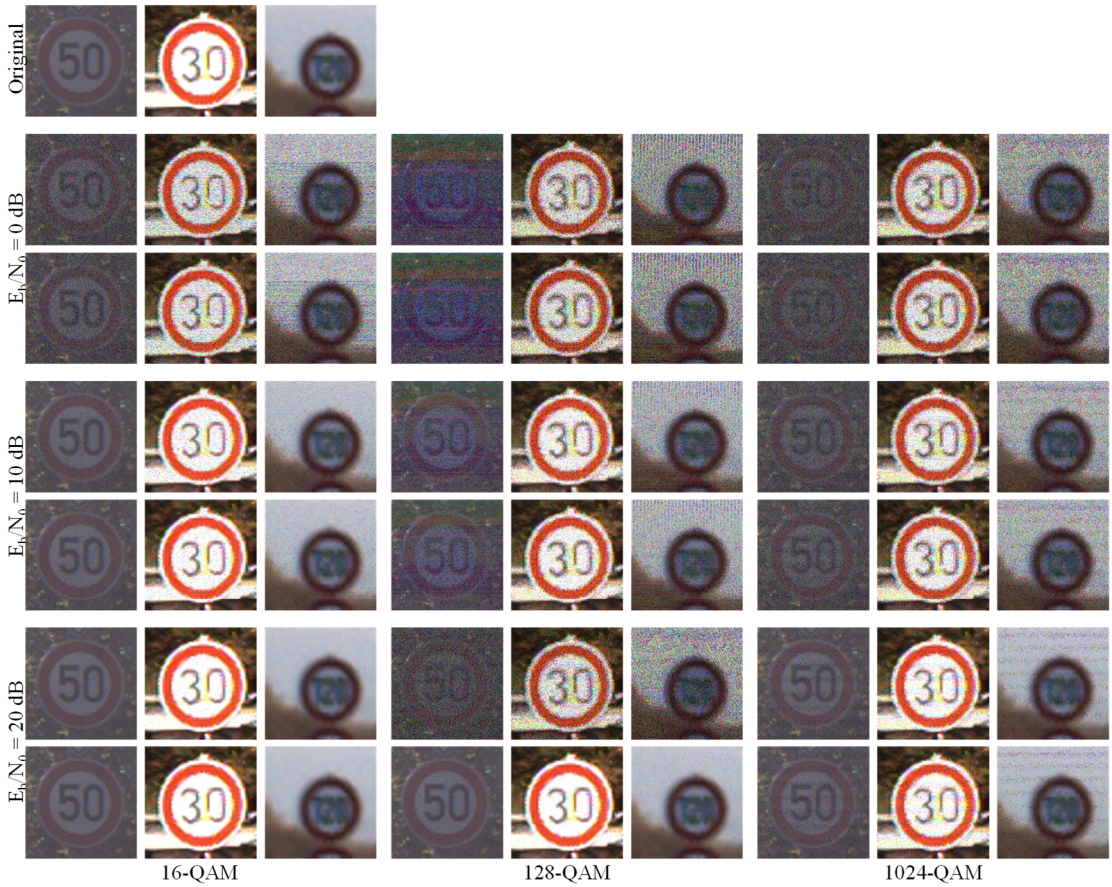

Figure 9. Sample recovered images from OFDM system over AWGN channel. The top row of each E_b/N_0 represents received images from system using LS channel estimation, whereas the bottom row images are from systems using MMSE channel estimation.

For visual analysis, Figure 9, shows three example images from the dataset and the distortions from different communication systems when transmitted over AWGN channel. From the visual inspection, the noise increases across higher order M-QAM; however, with increase in E_b/N_0, the image quality still improves significantly for lower order M-QAM (16-QAM) compared to higher order M-QAM (128-QAM and 1024-QAM). Although the images may be corrupted at lower E_b/N_0, the information in the region of interest (ROI) remains visible even with additional noise in the surrounding regions due to the way the traffic sign are designed with different color contrast [79], [80]. In terms of different channel estimation techniques, there is no visible distinction between the LS and MMSE as the images are almost visually identical in comparison to the two channel estimation techniques.

4.2.4.2. Over Rayleigh Fading Channel

The BER performance of the OFDM-based image communication system with higher order M-QAM modulation and various channel estimation techniques over the Rayleigh Fading channel is presented in Figure 10(a). The BER is presented on a logarithmic scale and is plotted against the E_b/N_0 for the proposed systems. From the graph we can see that there is no significant improvement in the BER across E_b/N_0 compared to the systems over simple AWGN channel as discussed in Section 4.2.4.1, this is due to heavy noise and distortions introduce by the fading effect of the Rayleigh Fading channel. The lowest BER was recorded with 16-QAM system utilizing MMSE channel estimation, achieving values of 0.41 to 0.18 across E_b/N_0 0 dB to 20 dB, respectively. The highest BER was obtained by 512-QAM with only a small drop from 0.47 to 0.40 across the E_b/N_0. In terms of channel estimation, MMSE performed better than LS for all M-QAM throughout the experiment and the difference in performance was significant at the lower E_b/N_0 region due to its prior channel knowledge. Additionally, 16-QAM leveraged the greatest improvement in BER (difference of 0.1 at E_b/N_0 20 dB) when using MMSE channel estimation, however, the difference in the performance was insignificant for rest of the M-QAM with respect to different channel estimation techniques. Overall, the BER increases with a higher order of M-QAM and MMSE channel estimation performed

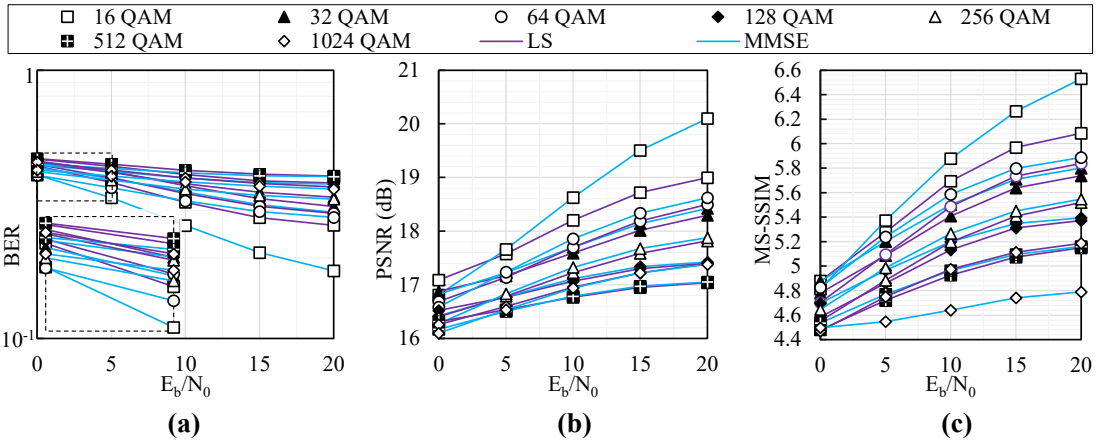

Figure 10. Performance analysis of various OFDM systems over Rayleigh Fading channel. Analysis in terms of (a) BER, (b) PSNR, and (c) MS-SSIM with respect to various E_b/N_0.

better than LS channel estimation throughout the experiment. 16-QAM has the lowest average BER using MMSE channel estimation and 512-QAM has the highest average BER using LS channel estimation across E_b/N_0.

The PSNR evaluation on the received images is shown in Figure 10 (b). and it shows a similar pattern to BER. The PSNR value improves with higher E_b/N_0, however it decreases with higher order M-QAM across all E_b/N_0 due to increased error bits that subsequently resulted in lower image quality. The 16-QAM and 1024-QAM systems demonstrated their best image quality using MMSE channel estimation at E_b/N_0 20 dB, with average PSNR values of 20 dB and 17 dB, respectively. In terms of channel estimation, in contrast to the BER analysis, the PSNR value of images from systems using LS was better than MMSE from E_b/N_0 0 dB to 5 dB; however, after E_b/N_0 5 dB, all the M-QAM systems performed better with MMSE channel estimation, and the highest quality improvement of the image due to different channel estimation was observed by the 16-QAM system (improvement of 1 dB). Throughout the experiments, the 16-QAM system with MMSE channel estimation maintained the highest level of image quality retention (except at E_b/N_0 0 dB), while the 512-QAM system with LS estimation exhibited the lowest image quality. Like PSNR, the MS-SSIM values exhibit some improving trends with higher E_b/N_0 but show a decline in image quality for higher order

M-QAM as shown in Figure 10 (c). The 16-QAM and 1024-QAM systems had the highest MS-SSIM score at E_b/N_0 20 dB, with scores of 6.5 and 5.18 while using MMSE and LS channel estimation, respectively. In terms of channel estimation, all the M-QAM systems performed better with MMSE channel estimation across E_b/N_0 5 dB to 20 dB, except for 1024-QAM, where MMSE channel estimation showed significant drop compared to LS. Overall, the 16-QAM system with MMSE channel estimation maintained the highest MS-SSIM score, while the 1024-QAM system with MMSE estimation had the lowest throughout the experiment.

Figure 11. shows recovered images from communication system over Rayleigh Fading channel. Compared to the visual inspection of images from AWGN channel discussed in Section 4.2.4.1, the images from the Rayleigh Fading channel are heavily distorted and most of the ROI information at the lower E_b/N_0 are not clearly visible. The amount of noise on the images decreases slightly with higher E_b/N_0, and that is sufficient to make the information on the ROI visible due to the robust design of the traffic signs with different color contrast [79], [80]. Therefore, for higher E_b/N_0, the images may be heavily distorted but the information in the ROI (speed limit) is still slightly visible, and the DL models can exploit that and extract the features from the ROI to predict the signs correctly, which will be further discussed in Section 4.4.3.3. Another trend that is observed is that the images from the M-QAM systems with odd length of symbol (128-QAM) have reduced the brightness compared to the even length of the symbol, which retained brightness the same as the original image. In terms of channel estimation, the images at E_b/N_0 0 dB from system utilizing MMSE channel estimation (bottom row) is clearer compared to LS channel estimation (top row) across the M-QAM, where the information in the ROI is not visible at all for the LS channel estimation. Therefore, it indicates that MMSE retains better quality images for lower E_b/N_0 compared to LS in terms of human visual perspective.

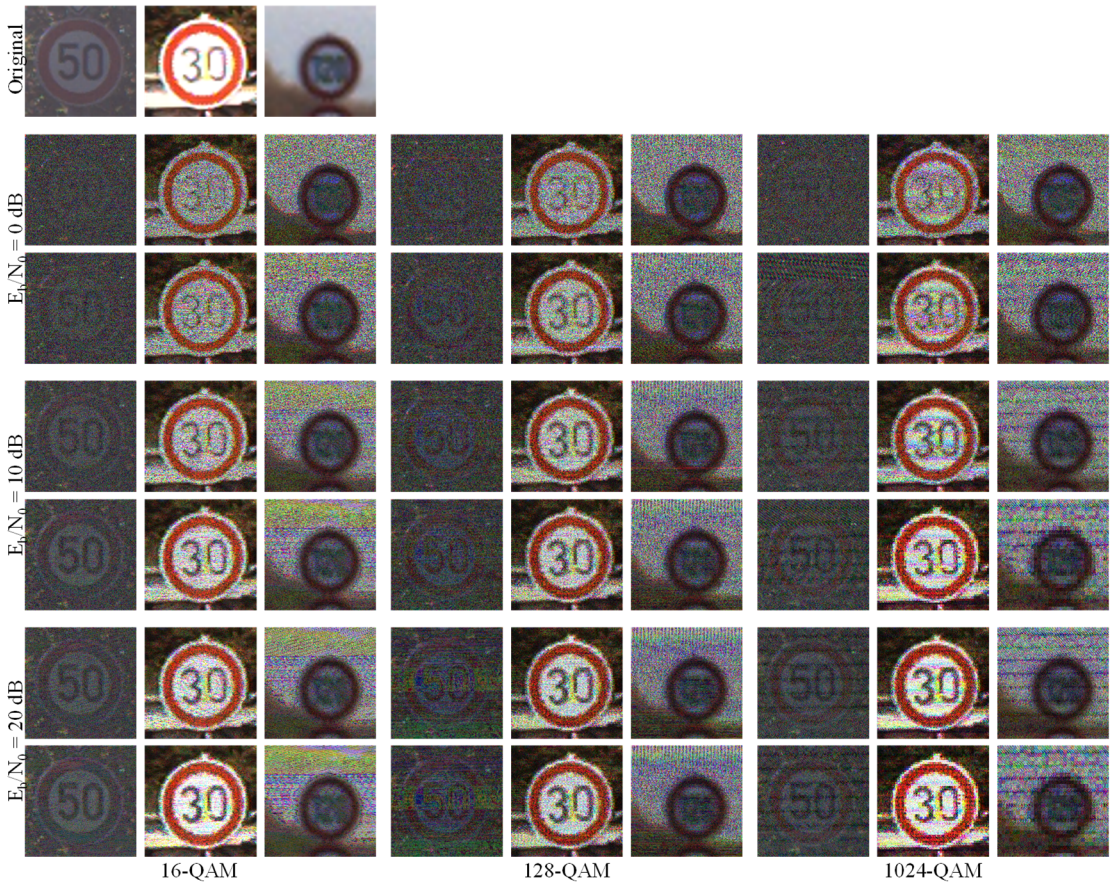

Figure 11. Sample recovered images from OFDM system over Rayleigh Fading channel. The top row of each E_b/N_0 represents received images from system using LS channel estimation, whereas the bottom row images are from systems using MMSE channel estimation.

4.3. Downstream DL Application: Traffic Sign Recognitions

The integration of advanced information and communication technologies along with Internet of Things (IoT) devices has brought a significant transformation to the Intelligent Transportation System (ITS) in smart cities. The incorporation of these technologies has resulted in synchronized transport networks, improved driving experience, optimized traffic management, and facilitated intelligent vehicular applications [81], [82]. The image communication system lies at the core of the ITS, enabling applications that directly connect vehicles with traffic infrastructures and management systems [83]. One such application of traffic sign recognition in ITS was considered in this

experiment. Traffic sign recognition is an essential CV application of an ITS environment and the collection and recognition of these signs are critical to ensure smooth traffic flow and prevention of accidents [83]. Effective traffic sign recognition in ITS requires both the collection of real-time images using high data rate communication systems, and the accurate recognition of these images using AI-enabled systems, respectively. In this experiment, the performance analysis of both these aspects was considered.

Furthermore, to effectively deploy AI-enabled system in an ITS environment for CV applications, a robust computing infrastructure is required for handling large amounts of image data, processing it in real-time, and performing complex computations [84]–[86]. Edge servers and cloud servers are two potential solutions that can meet these requirements [86]. Edge servers are located close to the data source, and can process data locally, leading to faster processing, reduced latency, and improved reliability. Cloud servers, on the other hand, provide vast computing power and storage

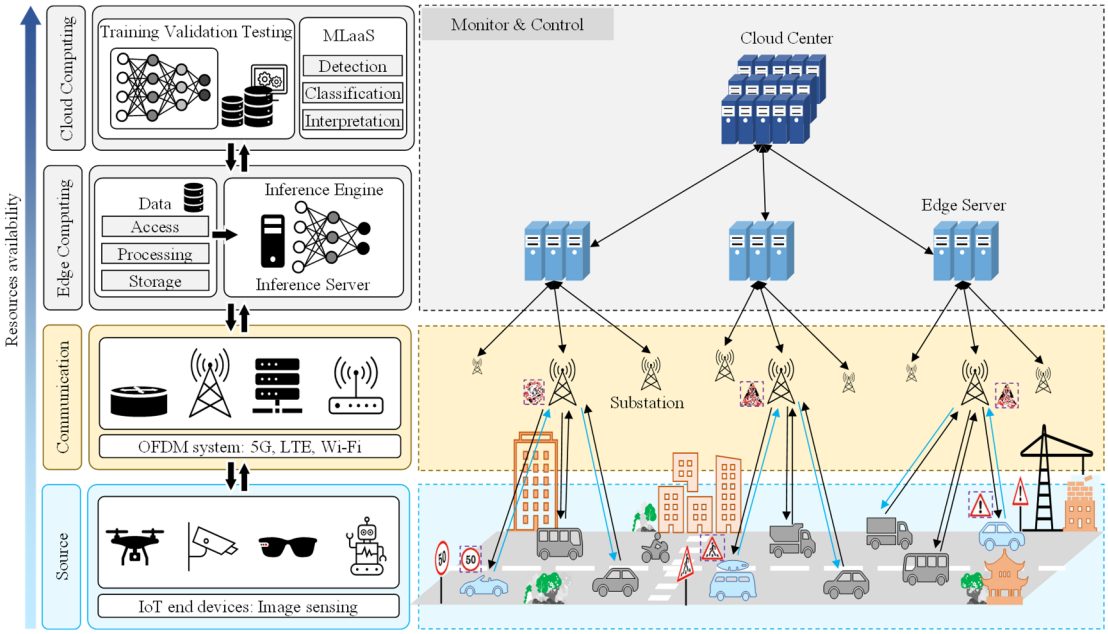

Figure 12. Schematic of downstream DL application. It illustrates an overview of used case scenario utilizing Edge–Cloud collaboration for deep learning-based CV applications in (a) generic IoT ecosystem and (b) ITS environment within a Smart City.

capacity, making them ideal for data-intensive tasks such as DL model training. By combining the computing power of edge and cloud servers, it is possible to process and analyze data in real-time while minimizing latency and reducing communication costs forming and edge–cloud collaboration [86]. Figure 12. shows such an infrastructure for the traffic sign recognition application in ITS, as considered in this experiment. Images are first transmitted to the edge server through various OFDM-based image communication systems, where they are introduced to noise and distortions from the system. The edge server then carries out inferencing on the received distorted images using a trained model, which were trained in the cloud server using a public dataset. The cloud center has several models that are task specific to enable Machine Learning as a Service (MLaaS) [86], [87]. In this scenario, the edge server requested a model for classification task to carry out traffic sign recognition in the ITS environment.

4.4. DL Models

In this section we discuss the robustness of DL models on images recovered from different communication systems presented in Section 4.2. We briefly describe the CNN models utilized in the experiment and the motivation of the choice, followed by the robustness analysis of the individual models with and without augmentation techniques under the influence of the different communication systems.

4.4.1. ResNet152V2

ResNet152V2 is a CNN model belonging to the ResNet family, commonly used for image classification tasks in computer vision applications [88], [89]. It is a more advanced and intricate version of the ResNet model with a total of 152 layers and utilizes skip connections (referred to as residual connections) which make it easier for the model to learn from lower layers and improve overall performance. ResNet152V2 employs a bottleneck design, which reduces the number of parameters required in each layer while maintaining the same level of representation power, making it a more efficient model compared to the earlier models in the series. Additionally, ResNet152V2

Table 4. Parameters of the DL models used in the experiments.

	ResNet152V2	EfficientNetV2-B0
Model Input Size:	224 × 224	
Top Layers:	Max Pool, 2048 Flatten, 2048 Fully Connected, 4096 * Fully Connected, 4096 * Dropout, 4096 Fully Connected, 7 +	Average Pool, 1280 Flatten, 1280 Fully Connected, 256 * Fully Connected, 7 +
Total parameter:	83,534,343	6,249,047
Trainable parameter:	83,390,599	6,188,439
Non-trainable parameter:	143,744	60,608

* ReLU Activation + Softmax Activation.

uses better weight initialization to prevent the vanishing and exploding gradient problems, and normalization approaches, by combining batch normalization and weight normalization, to improve the stability and speed of training [88], [89]. ResNet152V2 also utilizes different training techniques, such as augmentation and stochastic depth to improve the model's generalizability, making it more potent and reliable for image classification tasks compared to other models in the family [88], [89]. Overall, ResNet152V2 performs better than its predecessor and other CNN models in terms of accuracy while requiring fewer parameters. However, it is a very deep and complex model with many layers which demands significant computational resources. These demands can be fulfilled by edge-cloud collaboration. Therefore, for our experiment we have considered ResNet152V2 architecture for feature extraction with a proposed classifier for the classification task, as shown in Figure 13, and parameters of the model are mentioned in Table 4.

4.4.2. EfficientNetV2.B0

EfficientNet [90] is a family of CNN models that are designed to achieve state-of-the-art accuracy with highly efficient use of computational resources. Since its inception, it has become popular for computer vision applications due to its ability to achieve high accuracy on image classification tasks while using fewer parameters and floating-point operations (FLOPs) than other popular CNN models. EfficientNet utilizes neural architecture search (NAS) to design the baseline

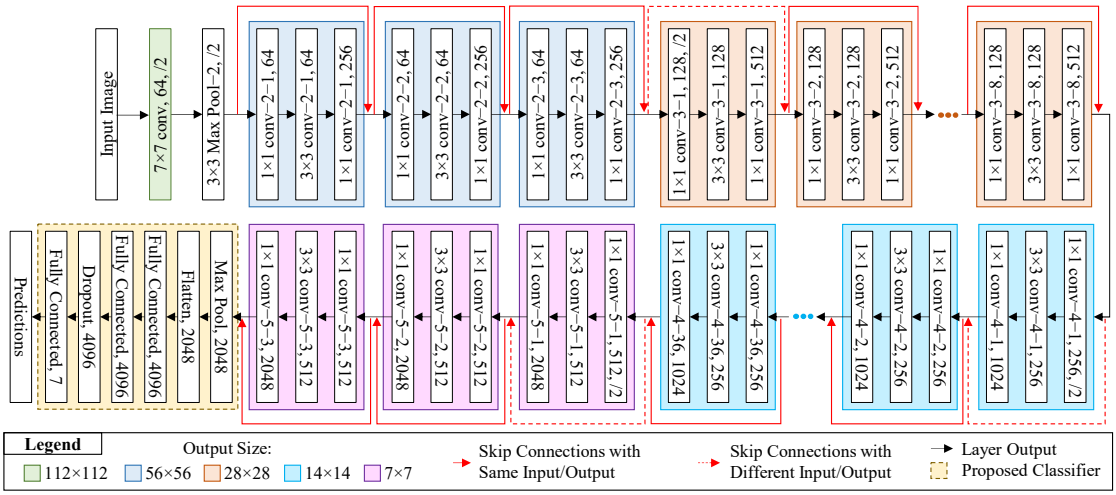

Figure 13. ResNet152-based proposed DL model for traffic sign recognition in ITS. For a layer '1×1 conv−2−1, 64', 1×1 is the filter size, 2−1 is module number, and 64 is the number of filters. The ellipses show that the blocks are repeated.

model, EfficientNet-B0, which has a better trade-off between parameters and accuracy. The model is then uniformly scaled up in terms of depth, width, and resolution to obtain a family of models, ranging from EfficientNet-B0 (the smallest model) to EfficientNet-B7 (the largest model). Despite using depth-wise convolution to achieve superiority in terms of the number of parameters and FLOPS, EfficientNetV1 had limitations in fully utilizing accelerators, which limited its training and inference speed. EfficientNetV2 [91] mitigates these limitations while ensuring parameter efficiency. It proposes three solutions, first, to adjust the size and regularization progressively during training; second, a non-uniform scaling strategy to add more layers in later stages; and third, the proposed Fused-MBConv in the early stage to improve training speed (introducing a small overhead on parameters and FLOPs). Therefore, in our experiment we have used EfficientNetV2-B0 to achieve state-of-the-art accuracy while ensuring parameter efficiency using the smallest model (B0) in the family and leveraging faster training and inference time (V2). The EfficientNetV2-B0 architecture (adapted from [91], [92]) with our proposed classifier is shown in Figure 14. and the parameters are mentioned in Table 4.

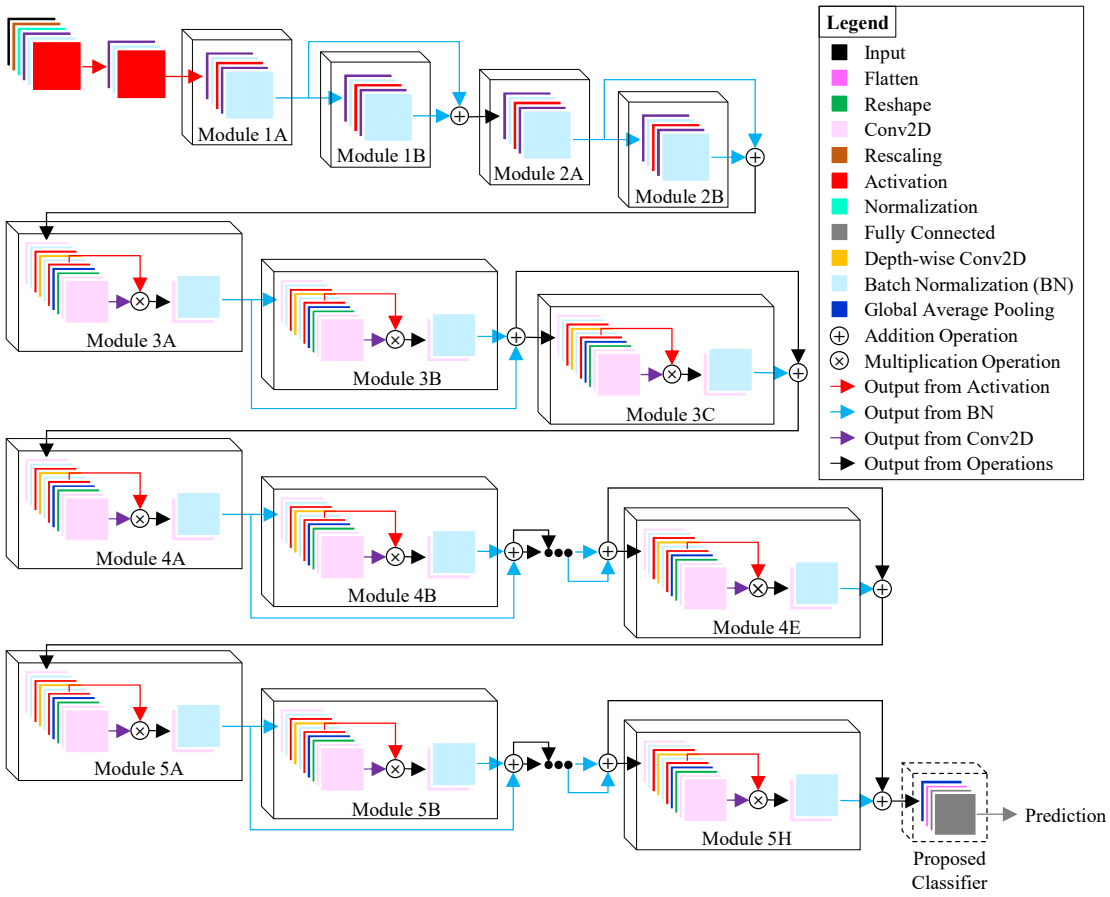

Figure 14. EfficientNetV2-B0-based proposed DL model for traffic sign recognition in ITS.

For this study, we have considered the German Traffic Sign Recognition Benchmark (GTSRB) dataset with speed limit signs of seven different classes, 30, 50, 60, 70, 80, 100, and 120. The dataset was divided into 80% for training, 10% for validation, and 10% for testing. For training, Stochastic Gradient Descent (SGD) optimizer and cross entropy-loss was used for both models, with learning rates ranging from 0.01 to 0.000001 and 0.1 to 0.0001 for ResNet152V2 and EfficientNetV2-B0, respectively, using the reduce learning rate (ReduceLR) technique. Furthermore, various augmentation techniques, such as 20-degree rotation, horizontal and vertical shift with factor 0.1, nearest fill mode, and 0.25 to 1.25 factor zoom and brightness, were applied to achieve high validation accuracy during training (as discussed in Section 2.3.2).

4.4.3. Performance Analysis of DL Models

In the experiment, ResNet152V2 achieved 98% training accuracy, and EfficientNetV2-B0 achieved 100% training accuracy. The training accuracy is comparable to related classification study on GTSRB dataset using ResNet152 [93] and EfficientNetV2 [94] family which achieved 96% and 98%, respectively. The validation accuracy of the models showed significant difference with and without applying augmentation techniques in the experiment. ResNet152V2 achieved 84% validation accuracy without augmentation, indicating overfitting and poor generalization; however, when the augmentation techniques were applied, the validation accuracy improved to 96%. Similarly, EfficientNetV2-B0 achieved 95% validation accuracy without augmentation, which improved to 99% with augmentation. Both the training and validation of the two models were completed on pristine (original) images. The inference (test) accuracy of the models on noisy images from different communication systems is further discussed.

4.4.3.1. Over AWGN Channel

The DL model performance in terms of classification accuracy on the reconstructed images from various communication systems over AWGN channel is discussed in this section. The dotted lines illustrate the model accuracy achieved without applying any augmentation techniques during training, whereas the solid lines represent the accuracy achieved with augmentation techniques (WA) applied during training. The accuracy of the DL model corresponds to the mean value across all inferencing images for the specific M-QAM and E_b/N_0.

Performance analysis of ResNet152V2 over AWGN channel: When augmentation techniques are not applied during the training of ResNet152V2 model, the accuracy achieved on the clean (original) images is 84%, and the overall accuracy of the model is low, as shown in Figure 15. This is due to the model not being able to learn and generalize well without augmentation techniques. From the graph, we can observe that the accuracy of the model decreases significantly with higher order M-QAM in the lower E_b/N_0 regions; however, for the higher E_b/N_0 regions the model is able to generalize well and achieves almost same accuracy as it performed on the original images across

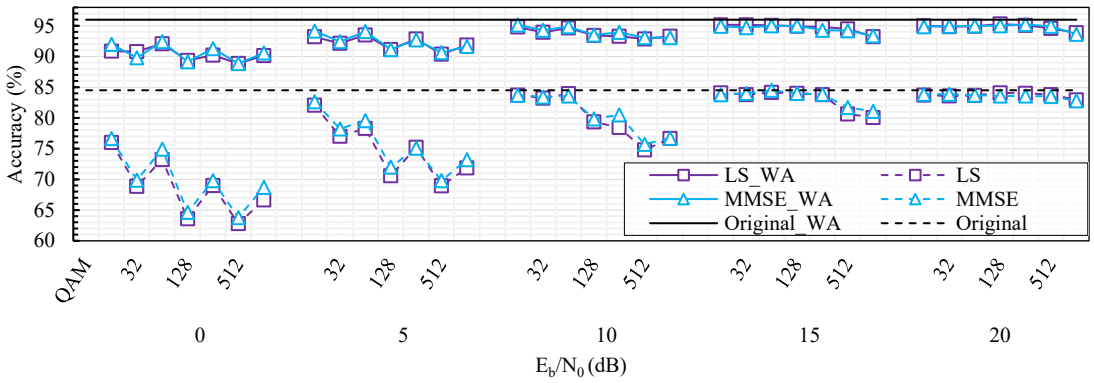

Figure 15. Performance of ResNet125V2 on images transmitted over AWGN channel. For order of M-QAM, M∈{16,32,⋯,1024} across each E_b/N_0.

all M-QAM (except 512-QAM and 1024-QAM). We can see a clear distinction between the accuracy of the model on images from M-QAM systems with odd length of symbol and even length of symbol with a zig-zag pattern in the lower E_b/N_0 regions. There is a sharp drop in the accuracy of the model on images from M-QAM systems with odd length of symbol (32-QAM, 128-QAM, and 512-QAM) which is due to the increased error bits, as discussed in Section 2.2.2. The highest difference in accuracy on images across different M-QAM was observed at E_b/N_0 0 dB, with 14% difference between highest accuracy (16-QAM) and lowest accuracy (512-QAM), this difference is less than 2% for E_b/N_0 20 dB. For channel estimation, accuracy was better on systems using MMSE channel estimation; however, the improvement is insignificant. Overall, the lowest accuracy observed without augmentation is 63% on images from 512-QAM system using LS channel estimation at E_b/N_0 0 dB. On the other hand, images from 16-QAM to 256-QAM achieved a highest accuracy of 84% (same as on the original images) for E_b/N_0 15 dB and 20 dB.

When augmentation techniques were applied, the ResNet152V2 was able to learn better and the accuracy of the model on clean (original) images improved to 96%. The impact of M-QAM with odd and even length of symbol is still observed, however the overall accuracy of the model improved significantly across all M-QAM. There is no drastic drop in accuracy across higher order M-QAM and the difference in the accuracy is less than 4% for all E_b/N_0. This is due to the model being able to generalize well across all M-QAM as it was trained with different augmentation techniques.

Accuracy with respect to channel estimation techniques has overlapping patterns, indicating that there is no significant impact when using more complex channel estimation (such as MMSE) on the accuracy of the DL model. With augmentation, the lowest accuracy obtained was 89% for 512-QAM system at E_b/N_0 0 dB; however, beyond E_b/N_0 10 dB the accuracy of the model was greater than 93% across all M-QAM and different channel estimation techniques.

Performance analysis of EfficientNetV2-B0 over AWGN channel: In contrast to ResNet152V2, EfficientNetV2-B0 achieved 95% accuracy on the clean (original) images even without applying any augmentation techniques during training as shown in Figure 16., this is due to its better learnability, as discussed in Section 4.4.2. However, there is a drop in the accuracy of models on images across higher order M-QAM. At E_b/N_0 0 dB, the highest drop in accuracy of 6% was observed on images from 16-QAM (highest accuracy) to 512-QAM (lowest accuracy); however, the difference in accuracy across M-QAM systems becomes smaller with increase in E_b/N_0. For E_b/N_0 20 dB the accuracy curve almost flattens (indicating no difference in accuracy across higher order M-QAM), and the accuracy was almost same as it was on the original images. This indicates that the model can generalize the noise on the images across all systems and provide better accuracy even for higher order M-QAM systems.

Furthermore, applying augmentation techniques during training drastically improved the EfficientNetV2-B0 model and it was able to predict almost all the clean images correctly, achieving

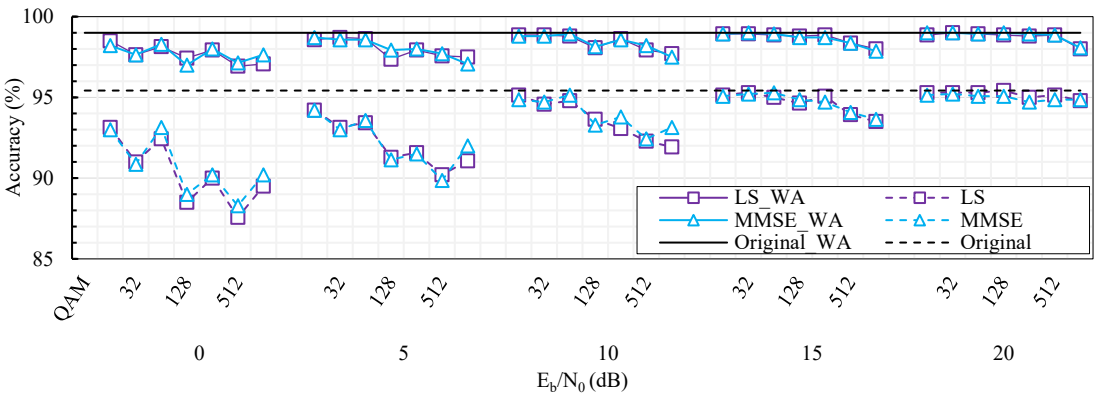

Figure 16. Performance of EfficientNetV2-B0 on images transmitted over AWGN channel.

99% accuracy. The difference in accuracy on images from higher order M-QAM is less than 2% with only little distinctions between M-QAMs with odd and even length of symbol. When applying augmentation, the lowest accuracy obtained was 97% on images from 512-QAM at 0 dB. However, beyond E_b/N_0 10 dB, the model accuracy on images from most of the systems was same as it performed on the original images (except 128-QAM at 10 dB, 512-QAM and 1024-QAM at 15 dB, and 1024-QAM at 20 dB). Throughout the experiment, the accuracy from system with different channel estimation had overlapping patterns and there were no clear indications that complex channel estimation (such as MMSE) improves the model accuracy. In comparison to the two DL models, EfficientNetV2-B0 outperformed ResNet152V2 across all image communication systems, regardless of whether augmentation was applied or not, this is due it its robustness as discussed in Section 4.4.2.

4.4.3.2. Over Rayleigh Fading Channel

In the evaluation of communication systems and image quality discussed in Section 4.2, we have observed severe degradation in performance when the Rayleigh Fading channel is utilized, as compared to employing only a simple AWGN channel model. We can see a similar trend in the accuracy of the DL models; however, when the DL models utilized augmentation techniques during training, they have achieved favorable results even on the Rayleigh Fading channel, as shown in figures in this section.

Performance analysis of ResNet152V2 over Rayleigh Fading Channel: When using ResNet152V2 without applying augmentation techniques during training, the accuracy on the images from the Rayleigh Fading channel achieved very low accuracy, as shown in Figure 17. The accuracy of the DL model dropped on images from higher order M-QAM systems and there was no significant improvement with higher E_b/N_0, in fact the difference in accuracy on the images from lower order M-QAM and higher order M-QAM was more significant in the higher E_b/N_0 region. The difference in accuracy on images from 16-QAM (highest accuracy) to 1024-QAM (lowest accuracy) was 8% at E_b/N_0 0 dB; however, the difference was 21% at E_b/N_0 20 dB. This is because the image quality in lower order M-QAM improved more significantly compared to those in higher order M-QAM with

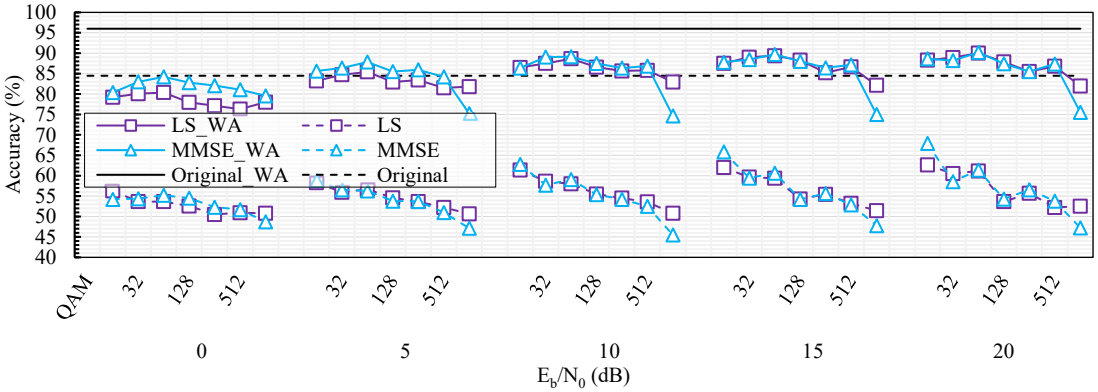

Figure 17. Performance of ResNet125V2 on images transmitted over Rayleigh Fading.

the increase in E_b/N_0. Additionally, it also indicates that the model was not able to generalize well on the images from different M-QAM systems. The highest accuracy obtained without augmentation was 68% at E_b/N_0 20 dB for 16-QAM system with MMSE channel estimation and the lowest accuracy of 46% was obtained at E_b/N_0 10 dB on images from the 1024-QAM system with MMSE channel estimation. For analysis on different channel estimations, there was no significant difference in the improvement in accuracy (except for 16-QAM and 1024-QAM) due to different channel estimation, as it had overlapping trends throughout the experiment. Images from 16-QAM system using MMSE channel estimation achieved almost 5% better accuracy compared to LS at E_b/N_0 15 dB and 20 dB; however, images from 1024-QAM achieved almost same improvement in accuracy using LS channel estimation compared to MMSE for E_b/N_0 higher than 5dB.

The ResNet152V2 model trained using augmentation techniques performed much better even on the multipath Rayleigh Fading channel. Overall, the model accuracy improved, and after E_b/N_0 10 dB the model accuracy was greater than 86% across all systems (except 1024-QAM). The highest accuracy of 90% was obtained by 64-QAM at E_b/N_0 15 dB and 20 dB and the lowest accuracy of 75% was obtained at E_b/N_0 10 dB on images from 1024-QAM system, both using MMSE channel estimation. Overall, the difference in accuracy on images across higher order M-QAM was less than 4% (except 1024-QAM) which indicates the model was able to generalize well on images even from higher order M-QAM system. In terms of the effect of channel estimation techniques, there were

distinct differences in the accuracy of the model on images from systems utilizing different channel estimation techniques in the lower E_b/N_0 regions. From E_b/N_0 0–10 dB the accuracy on images from systems using MMSE channel estimation was greater compared to LS channel estimation, and at E_b/N_0 0 dB this difference was the highest with almost 5% improvement in accuracy for 64-QAM, 128-QAM, 256-QAM and 512-QAM. The gap gradually decreased for E_b/N_0 higher than 10 dB, and an overlapping pattern (except 1024-QAM) of accuracy on images from systems utilizing both channel estimation techniques was observed.

Performance analysis of EfficientNetV2-B0 over Rayleigh Fading Channel: Like ResNet152V2, the accuracy of EfficientNetV2-B0 also decreases on images from systems using the Rayleigh Fading channel compared to simple AWGN channel, as shown in Figure 18. Without applying augmentation techniques during training, the model was not able to generalize well, and the accuracy of the models decreases significantly with higher order M-QAM. There is a slight improvement in accuracy of the model on images from lower order M-QAM with the increase in E_b/N_0; however, accuracy on images from higher order M-QAM shows very little improvement. Without augmentation, the highest accuracy of 90% was obtained on images from the 16-QAM system at E_b/N_0 20 dB and the lowest accuracy of 70% was obtained on images from 1024-QAM at E_b/N_0 5 dB, using MMSE channel estimation. There was no significant difference in improvement by using different channel estimation techniques, except on images from the 1024-QAM system, where the model on images from system utilizing LS obtained approximately 8% greater accuracy compared to MMSE at E_b/N_0 greater than 5 dB. The model performed slightly better on systems utilizing MMSE for 32-QAM, 64-QAM, 128-QAM, and 512-QAM at E_b/N_0 0 dB, however this improvement is insignificant considering the overall performance throughout the experiment.

The EfficientNetV2-B0 model trained with augmentation techniques had drastic improvement across all systems. The model could generalize well since the difference in accuracy across M-QAM

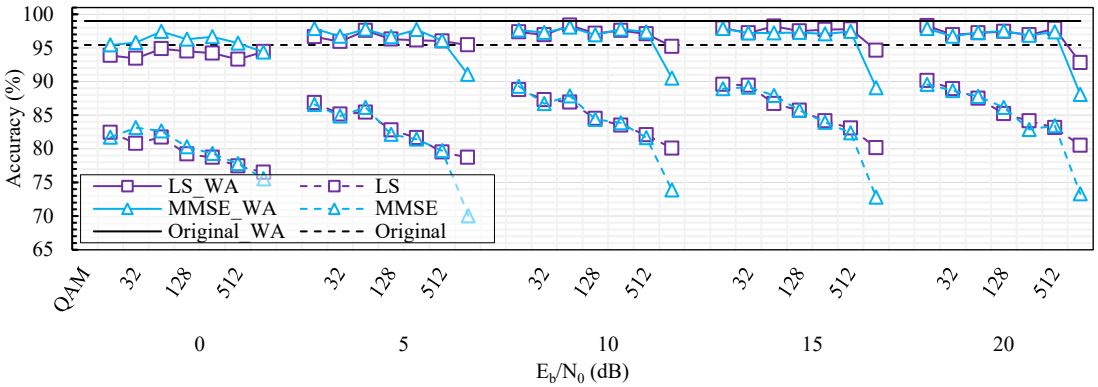

Figure 18. Performance of EfficientNetV2-B0 on images transmitted over Rayleigh Fading.

systems is less than 2% (except 1024-QAM) at all E_b/N_0. Throughout the experiment, the accuracy of the model was greater than 95% across all systems (except 1024-QAM). The highest accuracy of 99% was obtained on images from 16-QAM system at E_b/N_0 20 dB, whereas the lowest accuracy of 88% was obtained on images from 1024-QAM system at E_b/N_0 20 dB, both using MMSE channel estimation techniques. The improvement in accuracy on images from systems using different channel estimation was observed only at E_b/N_0 0 dB, where accuracy was approximately 4% better on images from systems using MMSE compared to LS across all systems (except 16-QAM and 1024-QAM). However, at higher E_b/N_0 regions, different channel estimations did not have a significant impact on the accuracy, except for 1024-QAM, where accuracy on images from a system using LS channel estimation had approximately 5% greater accuracy compared to MMSE. Which indicates that the effect of using complex channel estimation (such as MMSE) improves the model accuracy at lower E_b/N_0 regions; however, it can have contrasting effect at higher E_b/N_0 regions and higher order M-QAM.

Comparing the performance of the two models, EfficientNetV2-B0 had greater accuracy on images from systems over the Rayleigh Fading channel, compared to ResNet152V2. The accuracy of the EfficientNetV2-B0 model trained using augmentation was from 88 to 99%, whereas for ResNet152V2 it was between 75% and 90% across all systems.

4.4.3.3. Visual Analysis of DL Generalizability

The generalizability of the DL model across the M-QAM system using different channel estimation can be further observed using the t-distributed stochastic neighbor embedding (t-SNE) [76] visualization algorithm, as shown in Figure 19. It is a probabilistic dimensionality reduction algorithm which maps high-dimensional feature points into low-dimensional feature space (typically 2D or 3D) while preserving the similarities. Features close to each other in the high-dimensional space are mapped closer to each other in the lower-dimensional space with high probability. Points that are close together in the t-SNE plot are likely to have similar features or attributes (in terms of color, texture, or shape), whereas points that are far apart are likely to be dissimilar. From Figure 19., it can be observed that the same class points from the DL model trained with augmentation are close together and do not have many overlapping points of different class (color) across E_b/N_0 0 dB, 10 dB, and 20 dB for the M-QAM systems. Whereas points from the DL model trained without

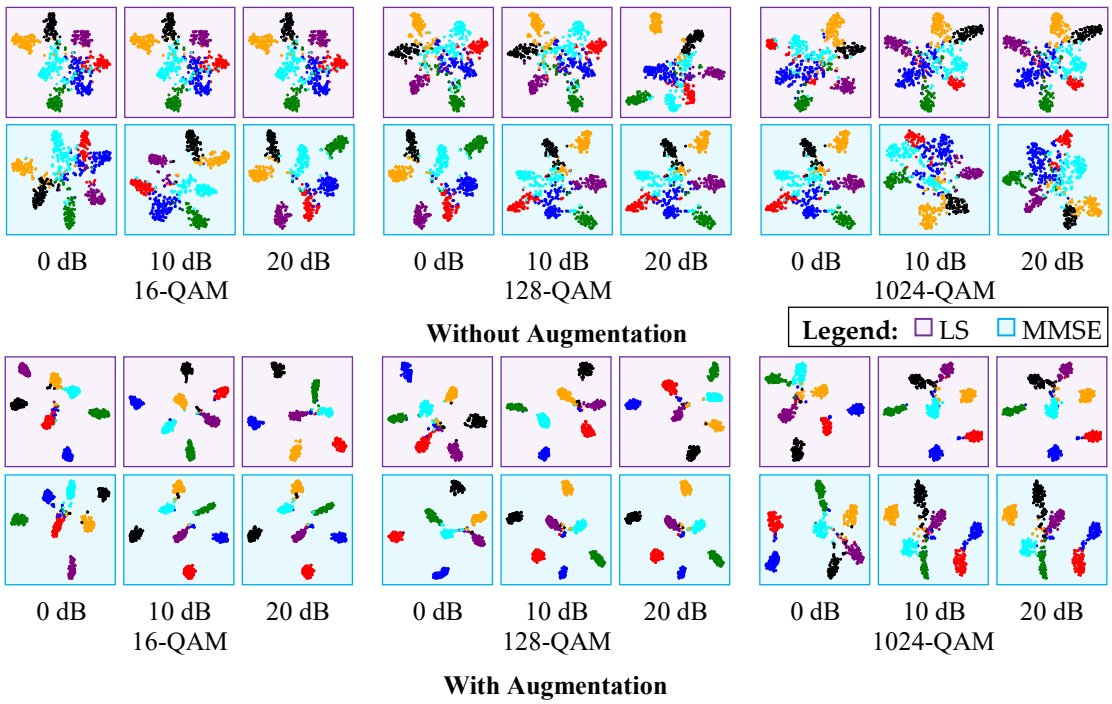

Figure 19. Two-dimensional t-SNE plot. It shows feature space analysis of EfficientNetV2-B0 model on recovered images over Rayleigh Fading channel. Each color of the dots represents one of the seven classes chosen from the GTSRB dataset.

augmentation are more scattered and overlap with other classes. There are no significant differences in class separability for different channel estimation used across the M-QAM systems. Overall, DL model trained with augmentation provides better separability between the seven classes and their features, resulting in better classification performance across all communication systems compared to DL model trained without augmentation.

Gradient-weighted Class Activation Mapping (Grad-CAM) heat map is another visualization algorithm that is used to visualize the attention region for a CNN while making predictions [77], [78]. It uses the gradient of the final convolutional layer with respect to the predicted class to weigh the layer activations, and then averages these weighted values over the spatial dimensions to obtain a class activation map. The map shows the crucial region of the image that is used to make the prediction by the CNN model. Figure 20. shows the Grad-CAM visualization of EfficientNetV2-B0 model prediction on images recovered from different communication systems. The images that were correctly predicted has Grad-CAM heat map concentrated in the ROI region of the traffic sign,

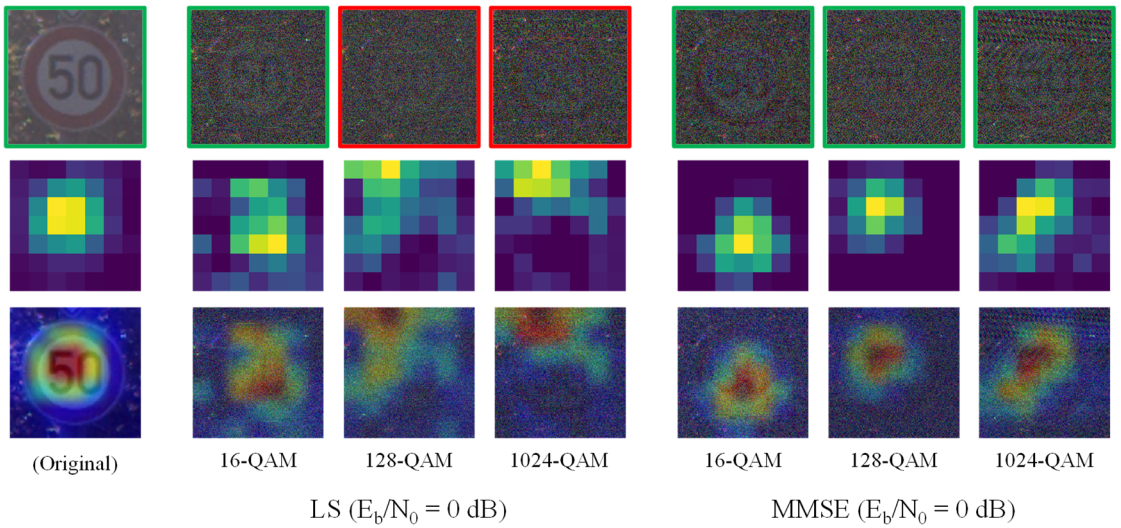

Figure 20. Grad-CAM visualizations. Analysis of EfficientNetV2-B0 model (with Augmentation) on sample image re-covered from different OFDM-based image communication systems under the Rayleigh Fading channel. The recovered image from different communication system (top), Grad-CAM heat map of the DL model's attention region (middle) and superimposed heat map over the image (bottom) is shown. Images with green border are correctly predicted and images with red border are incorrectly predicted.

whereas the two images that were wrongly predicted has the Grad-CAM heat map outside the ROI, therefore it could not extract the important features from the ROI to make the accurate prediction for the traffic sign.

For the t-SNE and Grad-CAM analysis, we visualize the performance of EfficientnetV2-B0 on noisy images recovered from OFDM communication system using the Rayleigh Fading channel only. The reason for this is that we have observed from Section 4.4.3 that EfficientnetV2-B0 is superior and more robust compared to ResNetV152, and the Rayleigh Fading channel replicates the real-world environment inducing more noise on the images.

For the experiment, the simulation of the various image communication systems and the BER analysis was conducted using Python 3.9.7, and the image quality analysis (PSNR and MS-SSIM) was performed using MATLAB R2021b. The performance analysis of the DL models was carried out using Python 3.9.7 and TensorFlow 2.2.0. The experiments were carried out jointly on Intel Core i5-4590 CPU and NVIDIA Tesla V100 (32 GB Memory) GPU.

V. CONCLUSION AND FUTURE WORK

5.1. Conclusion

In this study, we have analyzed the performance of DL models on images transmitted over various OFDM wireless communication systems for CV applications. Specifically, we have considered physical impairment from various techniques implemented in source coding, modulation, channel estimation, transformation, and channel model block in OFDM communication system. The main objective was to achieve a higher data rate to enable real time downstream CV applications while maintaining the overall communication system complexity. In general, the utilization of a higher order M-QAM in the fading channel environment leads to heavily corrupted image data in the output, regardless of the channel estimation techniques. However, our results have shown that the feature extractor of a DL model can be robust against these distortions with suitable data augmentation techniques, thereby improving the model generalizability across the higher order M-QAM. In other words, this trained feature extractor can extract meaningful features even from very noisy images, which can be utilized for downstream tasks, such as traffic sign recognition in ITS.

5.2. Future Work

Integration of advanced communication systems like MIMO, NOMA, and other hybrid systems significantly improve communication efficiency, and reliability that is crucial in IoT ecosystem. Additionally, transmissions in an IoT environment consist of a diverse array of data types such as image, video, audio, and text. Therefore, multi-modal downstream DL applications, integrated with advanced communication systems, can be an extension of the current study.

The present study implemented legacy OFDM PHY for communication system. An interesting future direction is implementing PHY as a DL architecture for end-to-end learning of the communication system. This is a paradigm shift from traditional model-driven (block-based) communication system to data-driven communication system, which would enable intelligent applications such as semantic communication system to enable beyond 5G networks.

REFERENCES

[1] J. G. Proakis, *Digital communications*. McGraw-Hill, Higher Education, 2008.

[2] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital communication*. Springer Science & Business Media, 2012.

[3] M. K. Simon and M.-S. Alouini, *Digital communication over fading channels: a unified approach to performance analysis*. in Wiley series in telecommunications and signal processing. New York: John Wiley & Sons, 2000.

[4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. in Adaptive computation and machine learning. Cambridge, Massachusetts: The MIT Press, 2016.

[5] R. Prasad, *OFDM for wireless communications systems*. Artech House, 2004. Accessed: Oct. 13, 2023. [Online].

[6] P. Banelli, S. Buzzi, G. Colavolpe, A. Modenini, F. Rusek, and A. Ugolini, "Modulation formats and waveforms for 5G networks: Who will be the heir of OFDM?: An overview of alternative modulation schemes for improved spectral efficiency," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 80–93, 2014.

[7] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. on Broadcast.*, vol. 48, no. 3, pp. 223–229, Sep. 2002, doi: 10.1109/TBC.2002.804034.

[8] I. Ahmad and S. Shin, "A novel hybrid image encryption–compression scheme by combining chaos theory and number theory," *Signal Processing: Image Communication*, vol. 98, p. 116418, Oct. 2021, doi: 10.1016/j.image.2021.116418.

[9] K. Sayood, *Introduction to data compression*. Morgan Kaufmann, 2017.

[10] I. Ahmad and S. Shin, "Noise-cuts-Noise Approach for Mitigating the JPEG Distortions in Deep Learning," in *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Bali, Indonesia: IEEE, Feb. 2023, pp. 221–226. doi: 10.1109/ICAIIC57133.2023.10067012.

[11] M. Ergen, *Mobile broadband: including WiMAX and LTE*. New York, NY : Springer, 2009.

[12] P. K. Singya, P. Shaik, N. Kumar, V. Bhatia, and M.-S. Alouini, "A Survey on Higher-Order QAM Constellations: Technical Challenges, Recent Advances, and Future Trends," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 617–655, 2021, doi: 10.1109/OJCOMS.2021.3067384.

[13] Kyongkuk Cho and Dongweon Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, Jul. 2002, doi: 10.1109/TCOMM.2002.800818.

[14] J. Smith, "Odd-Bit Quadrature Amplitude-Shift Keying," *IEEE Trans. Commun.*, vol. 23, no. 3, pp. 385–389, Mar. 1975, doi: 10.1109/TCOM.1975.1092806.

[15] C. Campopiano and B. Glazer, "A Coherent Digital Amplitude and Phase Modulation Scheme," *IEEE Trans. Commun.*, vol. 10, no. 1, pp. 90–95, Mar. 1962, doi: 10.1109/TCOM.1962.1088634.

[16] P. K. Vitthaladevuni, M.-S. Alouini, and J. C. Kieffer, "Exact BER computation for cross QAM constellations," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 3039–3050, Nov. 2005, doi: 10.1109/TWC.2005.857997.

[17] C. E. Shannon, "A mathematical theory of communication," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, Jan. 2001, doi: 10.1145/584091.584093.

[18] B. Bellalta, "IEEE 802.11 ax: High-efficiency WLANs," *IEEE Wireless Communications*, vol. 23, no. 1, pp. 38–46, 2016.

[19] S. Weinstein and P. Ebert, "Data transmission by frequency-division multiplexing using the

discrete Fourier transform," *IEEE transactions on Communication Technology*, vol. 19, no. 5, pp. 628–634, 1971.
[20] E. O. Brigham and R. E. Morrow, "The fast Fourier transform," *IEEE spectrum*, vol. 4, no. 12, pp. 63–70, 1967.
[21] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE transactions on Computers*, vol. 100, no. 1, pp. 90–93, 1974.
[22] T.-D. Chiueh and P.-Y. Tsai, *OFDM Baseband Receiver Design for Wireless Communications*. New York, NY: John Wiley & Sons, 2008.
[23] Y. Shen and E. Martinez, "Channel Estimation in OFDM Systems." Freescale semiconductor application note, 2006.
[24] M. C. Jeruchim, P. Balaban, K. S. Shanmugan, and M. C. Jeruchim, Eds., *Simulation of communication systems: modeling, methodology, and techniques*, 2nd ed. in Information technology--transmission, processing, and storage. New York: Kluwer Academic/Plenum Publishers, 2000.
[25] J.-J. van de Beek, "Synchronization and channel estimation in OFDM systems," Doctoral Thesis, Luleå University of Technology, Luleå, Sweden, 1998.
[26] Ji-Woong Choi and Yong-Hwan Lee, "Optimum pilot pattern for channel estimation in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2083–2088, Sep. 2005, doi: 10.1109/TWC.2005.853891.
[27] W. Kim, Y. Ahn, J. Kim, and B. Shim, "Towards deep learning-aided wireless channel estimation and channel state information feedback for 6G," *J. Commun. Netw.*, vol. 25, no. 1, pp. 61–75, Feb. 2023, doi: 10.23919/JCN.2022.000037.
[28] X. Yi and C. Zhong, "Deep Learning for Joint Channel Estimation and Signal Detection in OFDM Systems," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2780–2784, Dec. 2020, doi: 10.1109/LCOMM.2020.3014382.
[29] J. A. Cortes, F. J. Canete, and L. Diez, "Channel estimation for OFDM-based indoor broadband power line communication systems," *J. Commun. Netw.*, pp. 1–16, 2023, doi: 10.23919/JCN.2022.000056.
[30] F. Wang, "Pilot-based channel estimation in OFDM system," M.S. thesis, University of Toledo, OH, USA, 2011.
[31] B. Yegnanarayana, *Artificial neural networks*, 11. print. New Delhi: Prentice Hall of India, 2005.
[32] S. Lawrence, C. Giles, and A. Tsoi, "What Size Neural Network Gives Optimal Generalization? Convergence Properties of Backpropagation," University of Maryland, College Park, MD, USA, Tech. Rep. CS-TR-3617, 1996.
[33] Y. Chauvin, "Generalization performance of overtrained back-propagation networks," in *Neural Networks*, vol. 412, L. B. Almeida and C. J. Wellekens, Eds., in Lecture Notes in Computer Science, vol. 412. , Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 45–55. doi: 10.1007/3-540-52255-7_26.
[34] Y. LeCun, "Generalization and network design strategies," in *Connectionism in perspective*, Zurich, Switzerland: Elsevier, 1989.
[35] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)." arXiv, Feb. 22, 2016. Accessed: Feb. 19, 2023. [Online]. Available: http://arxiv.org/abs/1511.07289
[36] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 807–814.

[37] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[38] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *J Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.

[39] A. Tuli, N. Kumar, K. Kanchan Sharma, and S. T. Sharma, “Image Transmission using M-QAM OFDM System over Composite Fading Channel,” *IOSR Journal of Electronics and Communication Engineering IOSR-JECE*, vol. 9, no. 5, pp. 69–77, 2014.

[40] D. Krishna and M. S. Anuradha, “Image Transmission through OFDM System under the Influence of AWGN Channel,” *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 225, p. 012217, Aug. 2017, doi: 10.1088/1757-899X/225/1/012217.

[41] A. Mannan and A. Habib, “Adaptive processing of image using DWT and FFT OFDM in AWGN and Rayleigh channel,” in *2017 International Conference on Communication, Computing and Digital Systems (C-CODE)*, Islamabad, Pakistan: IEEE, Mar. 2017, pp. 346–350. doi: 10.1109/C-CODE.2017.7918955.

[42] H. Esmaiel and D. Jiang, “Progressive ZP-OFDM for Image Transmission Over Underwater Time-Dispersive Fading Channels,” in *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, Southend, United Kingdom: IEEE, Aug. 2018, pp. 226–229. doi: 10.1109/iCCECOME.2018.8658958.

[43] A. Agarwal, B. S. Kumar, and K. Agarwal, “BER Performance Analysis of Image Transmission Using OFDM Technique in Different Channel Conditions Using Various Modulation Techniques,” in *Computational Intelligence in Data Mining*, vol. 711, H. S. Behera, J. Nayak, B. Naik, and A. Abraham, Eds., in Advances in Intelligent Systems and Computing, vol. 711. , Singapore: Springer Singapore, 2019, pp. 1–8. doi: 10.1007/978-981-10-8055-5_1.

[44] J. Patel and M. Seto, “Live RF Image Transmission using OFDM with RPi and PlutoSDR,” in *2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, London, ON, Canada: IEEE, Aug. 2020, pp. 1–5. doi: 10.1109/CCECE47787.2020.9255670.

[45] V. Rajesh and A. R. Abdul Rajak, “Channel estimation for image restoration using OFDM with various digital modulation schemes,” *J. Phys.: Conf. Ser.*, vol. 1706, no. 1, p. 012076, Dec. 2020, doi: 10.1088/1742-6596/1706/1/012076.

[46] Z. H. Al-Shably and Z. M. Hussain, “Performance of FFT-OFDM versus DWT-OFDM under Compressive Sensing,” *J. Phys.: Conf. Ser.*, vol. 1804, no. 1, p. 012087, Feb. 2021, doi: 10.1088/1742-6596/1804/1/012087.

[47] M. J. Mohsin, W. K. Saad, B. J. Hamza, and W. A. Jabbar, “Performance analysis of image transmission with various channel conditions/modulation techniques,” *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 3, pp. 1158–1168, 2020.

[48] Z. N. Ghanim and B. M. Omran, “OFDM PAPR reduction for image transmission using improved tone reservation,” *Int. J. Electr. Comput. Eng.(2088-8708)*, vol. 11, pp. 416–423, 2021.

[49] L. Kansal, G. S. Gaba, N. Chilamkurti, and B.-G. Kim, “Efficient and Robust Image Communication Techniques for 5G Applications in Smart Cities,” *Energies*, vol. 14, no. 13, p. 3986, Jul. 2021, doi: 10.3390/en14133986.

[50] L. Kansal, S. Berra, M. Mounir, R. Miglani, R. Dinis, and K. Rabie, “Performance Analysis of Massive MIMO-OFDM System Incorporated with Various Transforms for Image Communication in 5G Systems,” *Electronics*, vol. 11, no. 4, p. 621, Feb. 2022, doi: 10.3390/electronics11040621.

[51] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, “Deep joint source-channel coding for wireless image transmission,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, 2019.

[52] D. Burth Kurka and D. Gündüz, “Joint Source-Channel Coding of Images with (not very) Deep Learning,” in *International Zurich Seminar on Information and Communication (IZS 2020)*, Zurich, Switzerland: ETH Zurich, Feb. 2020, pp. 90–94. doi: 10.3929/ETHZ-B-000402967.

[53] D. B. Kurka and D. Gunduz, “DeepJSCC-f: Deep Joint Source-Channel Coding of Images With Feedback,” *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 178–193, May 2020, doi: 10.1109/JSAIT.2020.2987203.

[54] D. B. Kurka and D. Gunduz, “Successive Refinement of Images with Deep Joint Source-Channel Coding,” in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Cannes, France: IEEE, Jul. 2019, pp. 1–5. doi: 10.1109/SPAWC.2019.8815416.

[55] J. Xu, B. Ai, W. Chen, A. Yang, P. Sun, and M. Rodrigues, “Wireless Image Transmission Using Deep Source Channel Coding With Attention Modules,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 2315–2328, Apr. 2022, doi: 10.1109/TCSVT.2021.3082521.

[56] M. Ding, J. Li, M. Ma, and X. Fan, “SNR-Adaptive Deep Joint Source-Channel Coding for Wireless Image Transmission,” in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, ON, Canada: IEEE, Jun. 2021, pp. 1555–1559. doi: 10.1109/ICASSP39728.2021.9414037.

[57] M. Yang, C. Bian, and H.-S. Kim, “OFDM-guided deep joint source channel coding for wireless multipath fading channels,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, pp. 584–599, 2022.

[58] H. Wu, Y. Shao, K. Mikolajczyk, and D. Gunduz, “Channel-Adaptive Wireless Image Transmission With OFDM,” *IEEE Wireless Commun. Lett.*, vol. 11, no. 11, pp. 2400–2404, Nov. 2022, doi: 10.1109/LWC.2022.3204837.

[59] I. Ahmad, N. Islam, E. Kim, and S. Shin, “Performance Analysis of Cloud based Deep Learning Models in OFDM based Image Communication system,” in *Proceedings of the Korean Institute of Communication Sciences Conference*, Jeju Island, Korea, Republic of: KICS, 2022, pp. 500–501.

[60] I. Ahmad, N. Islam, and S. Shin, “Performance Analysis of Cloud-based Deep Learning Models on Images Recovered without Channel Correction in OFDM System,” in *2022 27th Asia Pacific Conference on Communications (APCC)*, Jeju Island, Korea, Republic of: IEEE, 2022, pp. 225–259.

[61] N. Islam, I. Ahmad, and S. Shin, “Robustness of Deep Learning enabled IoT Applications Utilizing Higher Order QAM in OFDM Image Communication System,” in *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Bali, Indonesia: IEEE, Feb. 2023, pp. 630–635. doi: 10.1109/ICAIIC57133.2023.10067100.

[62] C. Szegedy *et al.*, “Intriguing properties of neural networks.” arXiv, Feb. 19, 2014. Accessed: Jul. 27, 2022. [Online]. Available: http://arxiv.org/abs/1312.6199

[63] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and Harnessing Adversarial Examples.” arXiv, Mar. 20, 2015. Accessed: Feb. 19, 2023. [Online]. Available: http://arxiv.org/abs/1412.6572

[64] K.-S. Kim, H.-Y. Lee, and H.-K. Lee, “Spatial error concealment technique for losslessly compressed images using data hiding in error-prone channels,” *J. Commun. Netw.*, vol. 12,

no. 2, pp. 168–173, Apr. 2010, doi: 10.1109/JCN.2010.6391373.
[65] A. K. Boyat and B. K. Joshi, “A Review Paper: Noise Models in Digital Image Processing,” 2015, doi: 10.48550/ARXIV.1505.03489.
[66] S. Dodge and L. Karam, “Understanding how image quality affects deep neural networks,” in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, Lisbon, Portugal: IEEE, Jun. 2016, pp. 1–6. doi: 10.1109/QoMEX.2016.7498955.
[67] P. Roy, S. Ghosh, S. Bhattacharya, and U. Pal, “Effects of Degradations on Deep Neural Network Architectures.” arXiv, Jun. 26, 2019. Accessed: Feb. 19, 2023. [Online]. Available: http://arxiv.org/abs/1807.10108
[68] K. De and M. Pedersen, “Impact of Colour on Robustness of Deep Neural Networks,” in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, BC, Canada: IEEE, Oct. 2021, pp. 21–30. doi: 10.1109/ICCVW54120.2021.00009.
[69] D. Hendrycks and T. Dietterich, “Benchmarking Neural Network Robustness to Common Corruptions and Perturbations.” arXiv, Mar. 28, 2019. Accessed: Feb. 19, 2023. [Online]. Available: http://arxiv.org/abs/1903.12261
[70] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, “ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness.” arXiv, Nov. 09, 2022. Accessed: Feb. 19, 2023. [Online]. Available: http://arxiv.org/abs/1811.12231
[71] W. Brendel and M. Bethge, “Approximating CNNs with Bag-of-local-Features models works surprisingly well on ImageNet.” arXiv, Mar. 20, 2019. Accessed: Feb. 19, 2023. [Online]. Available: http://arxiv.org/abs/1904.00760
[72] M. Naseer, K. Ranasinghe, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, “Intriguing Properties of Vision Transformers.” arXiv, Nov. 25, 2021. Accessed: Feb. 19, 2023. [Online]. Available: http://arxiv.org/abs/2105.10497
[73] A. Hore and D. Ziou, “Image Quality Metrics: PSNR vs. SSIM,” in *2010 20th International Conference on Pattern Recognition*, Istanbul, Turkey: IEEE, Aug. 2010, pp. 2366–2369. doi: 10.1109/ICPR.2010.579.
[74] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multiscale structural similarity for image quality assessment,” in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, Pacific Grove, CA, USA: IEEE, 2003, pp. 1398–1402. doi: 10.1109/ACSSC.2003.1292216.
[75] N. Venkatanath, D. Praneeth, M. C. Bh, S. S. Channappayya, and S. S. Medasani, “Blind image quality evaluation using perception based features,” in *2015 twenty first national conference on communications (NCC)*, IEEE, 2015, pp. 1–6. Accessed: Oct. 14, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7084843/
[76] L. Van der Maaten and G. Hinton, “Visualizing data using t-SNE.,” *Journal of machine learning research*, vol. 9, no. 11, 2008.
[77] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning Deep Features for Discriminative Localization,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 2921–2929. doi: 10.1109/CVPR.2016.319.
[78] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” 2016, doi: 10.48550/ARXIV.1610.02391.
[79] A. De La Escalera, L. E. Moreno, M. A. Salichs, and J. M. Armingol, “Road traffic sign detection and classification,” *IEEE transactions on industrial electronics*, vol. 44, no. 6, pp.

848–859, 1997.
[80] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, “A system for traffic sign detection, tracking, and recognition using color, shape, and motion information,” in *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.*, IEEE, 2005, pp. 255–260.
[81] L. Cai, J. Pan, L. Zhao, and X. Shen, “Networked Electric Vehicles for Green Intelligent Transportation,” *IEEE Comm. Stand. Mag.*, vol. 1, no. 2, pp. 77–83, 2017, doi: 10.1109/MCOMSTD.2017.1700022.
[82] S. A. A. Shah, E. Ahmed, M. Imran, and S. Zeadally, “5G for Vehicular Communications,” *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 111–117, Jan. 2018, doi: 10.1109/MCOM.2018.1700467.
[83] K. Zhang, S. Leng, X. Peng, L. Pan, S. Maharjan, and Y. Zhang, “Artificial Intelligence Inspired Transmission Scheduling in Cognitive Vehicular Communications and Networks,” *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1987–1997, Apr. 2019, doi: 10.1109/JIOT.2018.2872013.
[84] D. Kim, S. Park, J. Kim, J. Y. Bang, and S. Jung, “Stabilized adaptive sampling control for reliable real-time learning-based surveillance systems,” *J. Commun. Netw.*, vol. 23, no. 2, pp. 129–137, Apr. 2021, doi: 10.23919/JCN.2021.000009.
[85] N. Kiran, C. Pan, S. Wang, and C. Yin, “Joint resource allocation and computation offloading in mobile edge computing for SDN based wireless networks,” *J. Commun. Netw.*, vol. 22, no. 1, pp. 1–11, Feb. 2020, doi: 10.1109/JCN.2019.000046.
[86] I. Ahmad and S. Shin, “Perceptual Encryption-based Privacy-Preserving Deep Learning in Internet of Things Applications,” in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju Island, Korea, Republic of: IEEE, Oct. 2022, pp. 1817–1822. doi: 10.1109/ICTC55196.2022.9952589.
[87] M. Ribeiro, K. Grolinger, and M. A. M. Capretz, “MLaaS: Machine Learning as a Service,” in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, Miami, FL, USA: IEEE, Dec. 2015, pp. 896–902. doi: 10.1109/ICMLA.2015.152.
[88] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition.” arXiv, Dec. 10, 2015. Accessed: Feb. 21, 2023. [Online]. Available: http://arxiv.org/abs/1512.03385
[89] K. He, X. Zhang, S. Ren, and J. Sun, “Identity Mappings in Deep Residual Networks.” arXiv, Jul. 25, 2016. Accessed: Feb. 21, 2023. [Online]. Available: http://arxiv.org/abs/1603.05027
[90] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” presented at the International conference on machine learning, PMLR, 2019, pp. 6105–6114.
[91] M. Tan and Q. Le, “Efficientnetv2: Smaller models and faster training,” in *International Conference on Machine Learning*, PMLR, 2021, pp. 10096–10106.
[92] I. Ahmad and S. Shin, “A Perceptual Encryption-Based Image Communication System for Deep Learning-Based Tuberculosis Diagnosis Using Healthcare Cloud Services,” *Electronics*, vol. 11, no. 16, p. 2514, Aug. 2022, doi: 10.3390/electronics11162514.
[93] Y. Li, M. Li, B. Luo, Y. Tian, and Q. Xu, “DeepDyve: Dynamic Verification for Deep Neural Networks,” in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, Oct. 2020, pp. 101–112. doi: 10.1145/3372297.3423338.
[94] Z. Zhang, M. Ye, Y. Xie, and Y. Liu, “PG-Prnet: A Lightweight Parallel Gated Feature Extractor Based on An Adaptive Progressive Regularization Algorithm,” 2022.

ACKNOWLEDGMENTS

Baba, whom I lost not long ago, continues to profoundly influence my life. I attribute all my achievements and positive traits to him. It was only after his passing that I realized he had meticulously shaped me into a Man, preparing me for the world. His teachings and ideals still guide me in his absence, and he is so deeply rooted in my thoughts and emotions that words cannot express. Ma, on the other hand, is the epitome of strength and resilience. During COVID-19, she faced immense tragedies, losing both her parents and her husband within the shortage of two years. Continuing this series of unfortunate events, just a month after I arrived in South Korea, she was diagnosed with cancer, which she bravely fought for over a year. Her resilience in tackling these challenges, never breaking down and always instilling hope in me to this day, remains an unwavering source of inspiration. This acknowledgment of Baba and Ma is a true testament to remind me in years to come of my roots, and the legacy I carry in my veins.

I would like to express my heartfelt gratitude to Prof. Shin for granting me the opportunity to pursue my master's degree under his supervision. This incredible opportunity has changed the course of my life. In the face of numerous challenges during admission, such as COVID-19, my personal issues, and travel restrictions, Prof. Shin deferred my admission for three semesters, and was extremely supportive. During the program, Prof. Shin's approach encouraged self-reliance and critical thinking, inspiring me to find solutions to technical and research problems independently. His insightful and constructive feedback has been instrumental, and greatly enhanced the quality of my work. Prof. Shin, not just as an academic supervisor, but as a human being, has left an indelible mark on my personal and professional journey. I am grateful to have crossed paths with him, and I extend my sincerest appreciation for his knowledge, guidance, and support.

I would like to thank all my friends and colleagues at Chosun University. I am grateful to Sifat Rezwan, who first informed me about the MS position and guided me to come to South Korea. Though he left South Korea by the time I arrived, he truly laid out the path for me. I would like to thank my dear friend, colleague, and mentor, Dr. Ijaz Ahmad for the endless hours of teaching,

polishing, guidance, conversation, laughing, advice, ranting, eating, walking, photography, and so on. My accomplishments in the last two years have been the $\frac{\pi^2}{6}$ of it all. Yet, this is just the beginning of the equation, which he would often quote as *"we are not at that point of our relationship yet"*. He might not be convinced by the *consistency, clarity,* and *accuracy* of this analogy, but I am just laying it out on the table for diverse scientific interpretation – that's just how we are. I am very thankful to Iftekharul Islam Shovon for helping me settle in South Korea, and always being there when I needed him. I would like to thank Pulok Tarafder for his positive influence, the hangouts, and most importantly, for partnering me as a back-pair in Jokgu — a game I have newly learned and come to love. I would like to thank Manal and Nabi for really creating a part of home here after they have arrived. It has taken us back to the adda and life from college and undergrad, with Dr. Ayesha being the joyful, helpful, and lively soul at the heart of it all. It has only been the last few months, yet the shared moments and experiences resonate deeply, outweighing the brevity of the time spent together. I would also like to thank Manal, Dr. Ayesha, Nabi, Ameera, Cham, Abel, Minwuye, Gabriel, and Villiana, for all the hours of basketball, football, taekwondo, and great times spent together.

My friends who are not around me but have been vital throughout this journey are Yeasin for accompanying me virtually through the tough nights, Abir for constantly checking on me, Udoy for making me feel like I always have someone and all the memes, Emad for taking me back to happy places with countless hours of laughs, Maha for always listening to me, and *EEE boiz* for making me feel home is where the heart is.

Last but not least, I am forever grateful to my dear wife Manal Mosharaf. She gives me purpose and meaning in everything I do. She has been pivotal in making me more active socially, mentally, and physically. Our time in South Korea has been filled with *Ghibli moments* and *snowfall feelings*, which I will cherish and hold dearly for the rest of time. She is the reason I strive to better myself every day, in every way. With her I conclude, like I would do in life till death.

-Nazmul Islam