Analysis, Design, and Realization of Industrial IoT Networks

Magzhan Amangeldi, BSc of Industrial Automation

Submitted in fulfillment of the requirements for

the degree of Master of Science

in Electrical and Computer Engineering

NAZARBAYEV UNIVERSITY

School of Engineering and Digital Sciences

Department of Electrical and Computer Engineering

Nazarbayev University

53 Kabanbay Batyr Avenue,

Astana, Kazakhstan, 010000

Supervisor: Galymzhan Nauryzbayev

Co-supervisor: Mohammad S. Hashmi

Astana 2025

Table of Content

List of Abbreviations & Symbols

5G	Fifth Generation (Mobile Network)
AI	Artificial Intelligence
AP	Average Precision
AR	Average Recall
BERT	Bidirectional Encoder Representations from Transformers
CNN	Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
ELM	Embeddings from Language Models
FPN	Feature Pyramid Network
FRCNN	Fast Region Convolutional Neural Network
GAN	Generative Adversarial Network
GPS	Global Positioning System
IEEE	Institute of Electrical and Electronics Engineers
IMU	Inertial Measurement Unit
IIoT	Industrial Internet of Things
IoAV	Internet of Autonomous Vehicles
IoT	Internet of Things
IoU	Intersection over Union
JSCC	Joint Source-Channel Coding
MIMO	Multiple-Input Multiple-Output
ML	Machine Learning
MSE	Mean Squared Error
Mem-DeepSC	Memory-based Deep Semantic Communication
NLP	Natural Language Processing
NMT	Neural Machine Translation
QoI	Quality of Information
QoS	Quality of Service
PSNR	Peak Signal-to-Noise Ratio
ResNet	Residual Network
RoI	Region of Interest
RNN	Recurrent Neural Network

RPN	Region Proposal Network
SPAWC	Signal Processing Advances in Wireless Communications
SSIM	Structural Similarity Index Measure
UAV	Unmanned Aerial Vehicle
Ui	User Interface
Wi-Fi	Wireless Fidelity

List of Figures

Abstract

Currently, the Industrial Internet of Things (IIoT) is reshaping into highly efficient, intelligent, and context-aware systems where remote sensing and aerial monitoring are concerned. Towards realizing Unmanned Aerial Vehicle (UAV) based surveillance for disaster response and urban planning, this thesis designs and implements two semantic communication pipelines for UAV-based surveillance that address practical challenges in real-time bandwidth efficiency and geospatial reasoning. Geometric and semantics information are jointly and automatically compressed using a Joint Source Channel Coding (JSCC) autoencoder to reduce bandwidth cost while preserving information for the downstream. The second approach is detecting objects onboard the UAV, augmenting those detections with UAV telemetry data, and estimating real-world coordinates in real time.

The empirical results show that the JSCC model can compress aerial imagery with little loss of semantic features. At the same time, the onboard system has an average precision (AP) of 7.35%, with excellent localization in the region of a few meters. In particular, the second pipeline allows the UAV to send only object class labels and coordinates, which is ideal for long-distance or high-altitude operations. These contributions jointly enable designing the next generation of IIoT UAV systems that can efficiently communicate and autonomously act intelligently. Based on this, two topics are discussed in later research to explore tradeoffs, challenges, and further directions: multimodal fusion and federated semantic inference.

Chapter 1

1.1 Introduction

The development of modern industrial technologies is a key driver of advanced connectivity, real-time data analysis, and intelligent manufacturing processes in multiple sectors, such as automotive, healthcare, energy, and others. IIoT is interesting because it combines various sensors, devices, and processes in one digital environment and achieves decent resource optimization, predictive maintenance, and efficient communication over convenient industrial operations [1].

IIoT is a concept based on the intelligent digitalization of modern control systems, with a deep focus on automation, optimal communication, and human-machine interaction [2].

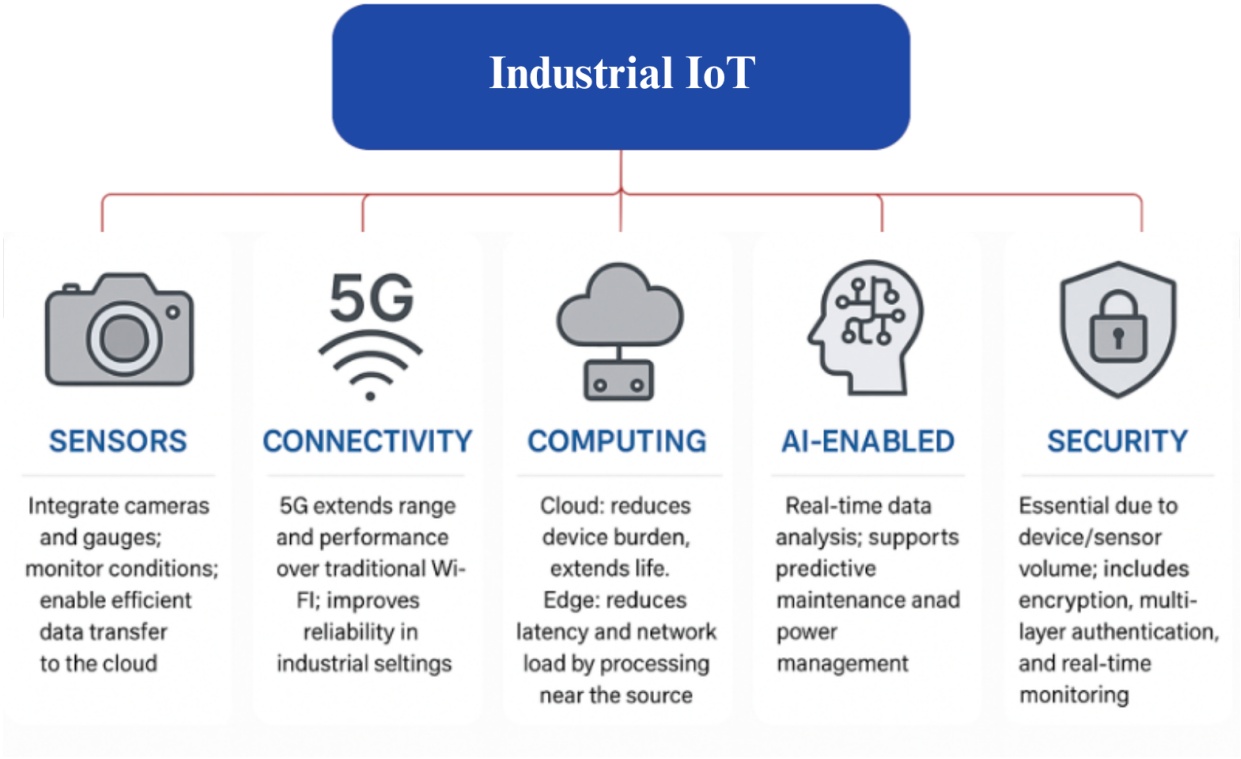

Figure 1.1 *Parts of modern IIoT networks.*

Convenient IIoT systems rely on five major parts, as shown in Figure 1.1 [3]. Classic examples of modern IIoT networks are the Internet of Autonomous Vehicles (IoAV), smart grids, smart cities, smart agriculture, etc. Nevertheless, constructing an efficient, robust and optimized network communication is a significant part of the system.

Shannon's classical communication theory became the cornerstone of modern digital networks by defining the channel capacity and optimal coding for error correction [4], [5].

Shannon introduced three fundamental problems in communication theory:

- Level A - The technical problem is centred on the message transmission problem, which is used to transmit and receive messages without errors.
- Level B - Semantic problem is associated with the issue of the accuracy of the meaning of data in the context of the transmitter and the receiver.
- Level C - The effectiveness problem focuses on the extent of communication within the IoT system. It checks whether the actions derived from the arrived information are feasible and go through the intended goal, such as optimal resource utilization or adaptation of the system's observance levels.

Traditional networks are designed to solve a technical problem without prioritizing the semantics of the transmitted messages. Nevertheless, the solution to the semantic problem has increasing relevance [6]. The core concept of semantic communication is based on using semantic features of data to transmit information. The typical case for semantic communication is sending only alarm data when needed in multi-sensor systems instead of sending raw data continuously. The alarm data could be classified as the critical or standard conditions of the device, while the data could be handled on the device and sent to the server.

However, active semantic communication can improve the efficiency of devices' transmitted data

in IIoT networks, providing more speed and accuracy for decision-making. This approach assists in reducing the bandwidth of the constrained networks and cuts down the load on the receiver side. Furthermore, Natural Language Processing (NLP) and Machine Learning (ML) allow extracting the most essential features from the raw data. It can benefit IIoT networks by prioritizing important information and the number of bits used in transmission. This thesis explores Deep Learning (DL) based optimal solutions for semantic-driven IIoT networks and modern industrial systems. The following literature review discusses the progress of semantic communication methods and comprehensive algorithms.

The evolution of traditional manufacturing and industrial systems has turned into an intelligent, interconnected ecosystem [7] thanks to the appearance of Industry 4.0. IIoT is the digital system that gives birth to this transformation, becoming the digital nervous system of smart factories and infrastructure. They are based on pervasive sensing and connectivity and intelligent data interpretation. Both industries and organizations generate vast amounts of raw data from the supply of sensors and devices that are too large to be analyzed in real-time and can yield meaningful insight [8]. The rising number of smart devices is one of the reasons that this paradigm change doesn't just have to be an efficient communicator (that provides an error-free — or perhaps near error-free communication), but it requires meaningful — where the content and the context of the transmitted information being transmitted, is as important as providing it accurately [9].

Thus, semantic communication emerges as a revolutionary gap-bridging concept in this landscape, where the idea of data transmission infuses a human-like understanding of meaning [10]. They differ from conventional models, which do not care about transmitted meaning, and semantic models ignore all information relevance, context-awareness, and inferential value. In IIoT, there is no time to waste. It is of the essence in such environments where you need to decide quickly,

one way or another: do we sound the alarms, make a call to maintain the asset, or reassign the resources. Semantic frameworks embed intelligence into the communication protocol, reducing bandwidth constraints, lowering energy consumption, and enabling distributed intelligence near the network edge.

Further, incorporating semantic communication with Artificial Intelligence (AI) and ML is a kind of leap forward in realizing adaptive and resilient IIoT systems [11]. DL systems acquire the latent features and significance hidden in complex industrial signals. JSCC, neural translation, and semantic embeddings also allow devices to interpret and infer the meaning autonomously, from which to both communicate and learn autonomously. The AI meets the semantic technologies and provides the foundation for the self-optimizing industrial environment that can anticipate faults, react to dynamic conditions, and increase overall system robustness.

Motivated by ever-emerging semantic communication trends in IIoT, this thesis develops DL-driven semantic communication frameworks for IIoT [12]. The goal is to investigate whether intelligent encoding and decoding strategies can exploit savings under resource limits to compress, prioritize, and extract high-value information. The following chapters discuss state-of-the-art semantic communication in the current literature, ML-based network optimal methods, and an implementation methodology demonstrated in multipurpose industrial default settings.

1.2 Research Questions

- How can intelligent communication methods optimize bandwidth in IIoT networks while preserving essential data for real-time decision-making?

- How can advanced computing techniques enhance the scalability, reliability, and security of IIoT networks under different industrial conditions and resource constraints?

Chapter 2 - Literature Review

2.1 Semantic Communication in IIoT

Integrating a semantic-driven approach into communication systems was suggested to improve the efficiency and importance of the data exchange process. Shannon's theory of communications examines messages as a sequence of symbols and highlights their meaning. Researchers recognized this gap in the early 1950s and presented the first theoretical framework for taking and measuring semantic information from messages [13]. The research work aimed to compute the information using bits. Over the years, efforts have been made to this foundation. In [14], the authors analyzed the relationship between information, inference, and intelligence by emphasizing how communication systems can handle the meaning of raw data. These early studies presented the relevance of the Shannon problems in networks like IIoT, where context and meaning of data play an essential role. In [15], the authors presented the Quality of Information (QoI) concepts, ensuring that the most helpful information is simultaneously delivered to the whole data. Similarly, [16] explained that using semantic relationships can preserve the precision of information while it is filtered out. This approach is mainly relevant to IIoT scenarios, where the data stream can be split into different high-level features like the indicator of machine health for further preventive maintenance and handling critical information. In [17], the authors presented the model that optimizes the transmitter and receiver algorithms to maximize successful meaning during constrained communication. This theoretical perspective highlighted the relationship between limited resources such as bandwidth and energy and the value of information.

Especially for IIoT systems, semantic-driven communication plays a crucial role because the industrial data is often context-dependent and vital. For instance, in a smart factory that contains various devices and sensors, the optimal usage of the data could emphasize the stability and reliability of the high-cost devices. Using semantic-level communication, the network can become more resilient to bandwidth constraints and enable faster decision-making on the edges. In [18], the authors presented a deep semantic communication system (DeepSC) for the specific text data, focusing on sentence meaning more than the exact words. The same concept is used in search engines such as Google, Yahoo, Yandex, etc. By leveraging semantic metrics in training like a sentence similarly or word error rate, the final model learned how to allocate limited capacity of the parts of the message. The model employs a Transformer structure using the attention mechanism to handle the meaningful features. This structure allows the ML model to understand the size of the message and its importance.

Memory-based deep semantic communication (Mem-DeepSC) achieves higher accuracy in meaning with a low word error rate by adding to the transformer memory layer [19]. This approach made the model heavier but decreased the amount of sending data between devices.

This system showed robust performance in noise channel conditions with fewer bit errors than the memoryless model. This robustness could be utilized in IIoT applications like remorse monitoring and control, where understanding the current situation is far more critical than perfectly transmitting each data byte. In addition, the devices can address data deluge issues in the networks using semantic-driven methods. In the case of multiple-user (devices) systems, this can significantly reduce redundant transmissions. This data aggregation is also highlighted in [20] for networks with many devices and a limited spectrum. These advances make semantic communication a promising paradigm for optimizing IIoT networks.

2.2 Common ML-based optimization methods

Despite using ML and DL in semantic communication, other standard methods based on the neural network show outstanding results in the optimization of IIoT networks. The reliability of heterogeneous device connectivity in dynamic conditions for IIoT networks increases when data-driven learning algorithms become implemented. Multiple studies demonstrate how ML/DL gets implemented across different points of the communication stack [2Trainingining of network-based receivers to detect signals and decode wireless transmissions demonstrates their operation as wireless systems receivers. A neural network system demonstrated its ability to detect data patterns from noisy signs, according to [23], while forming almost as well as optimal detectors yet needing no explicit channel models. In [24], the authors developed a model-driven DL-based multiple-input multiple-output (MIMO) detection, which integrates communication theory principles into Deep Neural Networks (DNNs). Developing efficient detectors that can generate better outcomes using traditional methods is based on the synchronization of a data-driven approach.

DL enables creativity in creating novel communication system designs. In [25], the authors introduced the revolutionary idea of using an autoencoder architecture to connect complete transmitter-receiver pairs, thus allowing neural networks to develop direct end-to-end message mapping schemes. The data compression and encoding process occurs at the transmitter, followed by decoding at the receiver through a joint learning mechanism that optimizes the complete communication system as an integrated whole. Later research evaluated this concept by conducting tests on wireless communication networks. A neural network-based communication link demonstrated its ability to adapt over air to channel impairments during validation tests, as reported in [26]. End-to-end learned systems could adjust their performance dynamically for IIoT networks

because they must operate in factory environments with either noisy conditions or time-varying radio interference. The ML applications extend past physical improvement work to include fundamental resource management problems crucial for IIoT implementation. A dense wireless network system serves as a real-life application of interference management. In [27], the authors trained DNNs to optimize power control management for devices to maximize network bandwidth concerning latency constraints. Due to the sizable search domains, the technique delivers better results than traditional optimization methods when managing large network devices (as found in industrial sensor networks). Likewise, DNNs could be used in scheduling, rooting, and network-slicing applications using adaptive resource allocation prediction of traffic patterns. In [28], the authors presented DL applications to physical layers and their ability to boost the reliability of communication systems through domain knowledge integration. ML has the potential within IIoT applications to improve latency and reliability by developing mechanisms to prevent fading and establishing optimal data routes between elements of a mesh network.

The utilization of ML techniques addresses unknown factors within network systems. Unpredictability inside industrial spaces includes moving machines that block communication signals. Nevertheless, fast adaptation could be achieved by combining generative models and meta-learning approaches. The authors in [29] utilized a Generative Adversarial Network (GAN) to implicitly model the wireless channel, allowing end-to-end communication systems to train without needing a formal channel model. Such modelling technique proves beneficial for networks that contend with intricate channels that are hard to model (like inside plants containing metal barriers and reflection effects). The article [30] presented an online strategy teaching communication systems to adapt to new situations using limited data quickly. A model-based system supports IIoT networks in fast adjustments following device additions or interference

pattern shifts. ML techniques integrated into system design can provide adaptive data-driven methods. Standards for implementing semantic communication exist because meaning extraction demands pattern recognition capabilities that ML excels at doing.

2.3 Semantic Encoding and Decoding

Modern AI techniques prove ideal for encoding and decoding meaning within IIoT networks that implement semantic communication. Semantic encoding converts any physical signal into a compressed form, maintaining its primary information content and decoding returns the meaningful information to the recipient. The methods used in this process borrow heavily from NLP because it seeks to detect language semantics. Neural machine translation (NMT) exemplifies how to build semantic communication by using encoder networks to produce semantic vectors from one-language inputs and decoder networks to generate equivalent meaningful outputs in different languages [31]. The authors in [32] created an attention-based NMT framework that taught sentence translation through developing semantic message representations. This principle shows direct implementation within semantic communication, where the system converts an incoming raw data stream into semantic meaning before returning it as a data stream or implementing action decisions for the receiver.

Semantic communication has also been utilized for various data types relevant to IIoT. The authors in [33] presented the first deep JSCC algorithm for text message transmission in wireless systems.

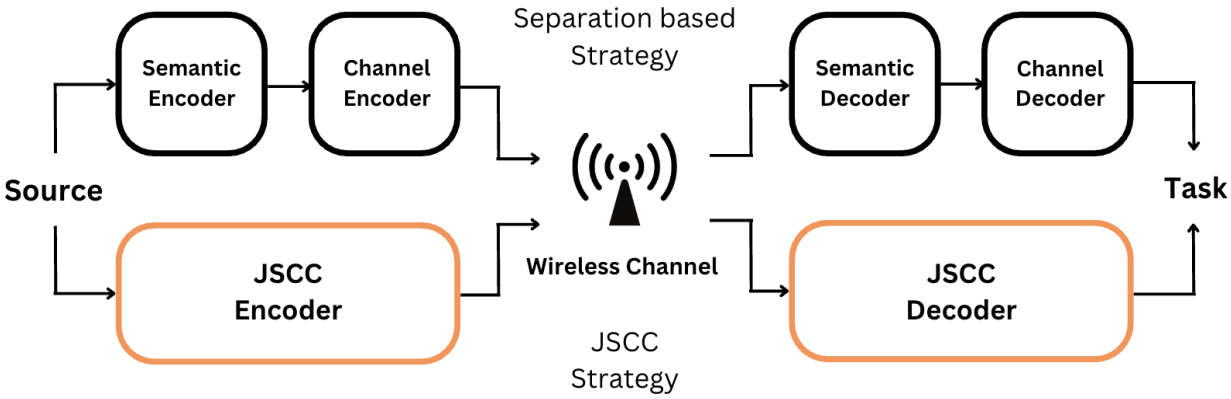

Figure 2.1 *Comparison of strategies of autoencoding.*

Figure 2.1 illustrates two alternative transmission techniques that employ either separation-based (black) or JSCC-based (orange) methods for wireless transmission of source data:

- The separation-based strategy (black) begins by subjecting the source to a semantic and channel encoder for error control operations. A semiconductor decoder regenerates the encoded bits after the channel decoder operation to produce meaningful outputs for completing various tasks.
- JSCC combines compression and error protection functions in one step through its single orange strategy that directly generates channel signals. The JSCC decoder at the receiver unifies the information recovery process to provide the result for the task.

The image reveals two distinct transmission-decoding workflows, which start with standard separated semantic and channel coding and feature a single JSCC integrated approach.

An autoencoder-like neural network system used a text-to-channel-symbol mapping technique for optimized end-to-end performance by directly learning how to protect text meaning over the transmission channel during one processing step. In [34], the authors provided wireless image

transmission using a JSCC semantic encoder and decoder. During channel degradation, the neural network provides smooth image deterioration because it selects critical semantic image features rather than precise pixel accuracy. Visual communication between a factory camera and external systems becomes enhanced through ML by enabling precise defect detection while discarding unimportant background details, reducing bandwidth usage.

An example of good semantic compression is extracting essential objects from the image, like object shapes, over the quality of the pixels. The encoding and decoding are not dependent on the type of data. The model performance only depends on the amount and quality of training data.

The key technological components for embeddings are language models with embeddings. The authors in [35] employed Word2Vec embedding to generate word vectors with adjacent placements in the space for words with contextual or semantic similarity. Semantic features for machines can be encoded using these embeddings, which enable IIoT devices to send such dense forms during data exchange. Bidirectional Encoder Representations from Transformers (BERT), along with Embeddings from Language Models (ELM), represent modern contextual embeddings that extract meaningful contextual word representations [36]. The addition of this technology lets the semantic communication system identify that the word "charge" in "battery charge low" highlights electrical energy as essential information, which differs from its importance in separate contexts for signal transmission. Based on the content, the different types of DNNs, such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), are used. Schuster and Paliwal's bidirectional RNNs enable sequence encoding through an encoder-operator that examines past and future contextual information [37] for content that requires complete sequence examination (popular in language and time-series sensor applications).

The research in [38] proved how RNNs could acquire training skills, demonstrating their ability to detect sophisticated patterns in data, while [39] used CNNs on sentence modelling through convolutional operations to detect local semantic features that may benefit command messages and IIoT event logs. ML-based semantic encoder systems unite modern technologies to convert industrial data into signals that maintain semantic value. The decoder system, which operates at an edge server or independently from it, uses a learned model to reconstruct signals while performing intelligent inference that could involve direct event detection. The communication system develops dual responsibilities by executing data interpretation and delivery processes.

2.4 Challenges in Optimizing IIoT Networks

The practical challenges in optimizing IIoT networks are:

- **Interoperability and Heterogeneity**: Industrial networks must link various devices with protocols and data formats. Integrating communication between legacy equipment and contemporary IoT devices proves difficult to implement [40]. The semantic communication infrastructure needs to achieve adaptability for translating between various data expression systems (such as alarm code from one system to status message from another). Industry-wide standardization of semantic information exchange remains an unresolved issue, which also requires the development of ML models capable of ML across different source types.
- **Data Integrity and Security**: The protection of information integrity along with trustworthiness becomes crucial when processing data through compression and inference stages [41]. IIoT decisions that involve machine shutdown depend on reliable and exclusive data. Research methods that reduce data transmission to semantics must prove the

preservation of essential information elements. The conservation of information quality through approaches based on semantic relationships improves overall quality but requires complete validation procedures. Attackers can exploit semantic communication through novel vulnerabilities by attempting to modify the intended message significance that flows between devices. Establishing safe data transmission through encryption becomes crucial for semantically encoded data because security measures need to extend to authenticate semantic information specifically.

- **Limited Bandwidth and Scalability**: Thousands of sensors and actuators within IIoT systems compete with one another to share network resources. The massive number of sensors and actuators uses too much bandwidth, although applying advanced compression does not solve the problem, especially within wireless installations [42]. The real-time application of semantic communication for numerous devices becomes computationally burdensome due to its ability to reduce unnecessary data transmission. Local servers benefit from edge computing because they manage intensive ML processing tasks but must perform complex modelling simultaneously across multiple concurrent streams efficiently. Therefore, federated learning and model compression methods are necessary to avoid congestion when deploying semantic models across the network.

Such obstacles require solutions from multiple academic fields working together. The advancements in network slicing make semantic communication possible by reserving distinct slice allocation patterns for different data flow types that range from reliable low-bandwidth semantic emergency alerts to raw data distribution when needed. Improvements in edge AI hardware technologies bring manufacturing and power facilities closer to processing complex semantic encoding/decoding algorithms within their facilities. The combination of semantic

communication with ML applications promises to fulfil industrial network requirements but requires additional solutions that address data integration, network efficiency, and compatibility problems. The research indicates that continued forward momentum has established the solid groundwork for the suggestions presented in this study.

The training process of heavyweight DNN models for specific tasks like image recognition is often complicated due to constraints in the feature extraction part [41]. The residual learning framework presented in [43] is the optimal solution for training such a DNN model. Residual Networks (ResNets) solve the degradation problem, where the training accuracy declines or saturates when network depth increases, using residual learning. The method transforms network layers from mapping to putting specific residual functions instead of creating desired underlying functions. While common networks aim to approximate underlying functions $H(x)$, ResNets focus on optimizing residual function $F(x) = H(x) - x$, and the original mapping is defined as

$$y = F(x, W_i) + x, \tag{2.1}$$

where x and y are the input and output layer vectors, and $F(x, W_i)$ is the residual function learned by nonlinear stacked layers. This formation significantly reduces the learning constraints and allows optimal training for deeper networks than traditional DNN architectures. The provided results demonstrate the impressive benefits of this learning approach. For example, in the ImageNet classification task, the proposed model executed exceptionally well results by getting a 3.57 percent error rate. Further analysis of the CIFAR-10 dataset highlighted that the ResNet can be effectively used in optimization challenges in deep architectures, such as successful training in the 100 to 1000 depth layers network architectures. Moreover, ResNet has powerful generalization capabilities in various tasks, such as object detection, image segmentation, etc., where accuracy

improvements are needed. In [44], the authors presented a Feature Pyramid Network (FPN) as an optimal solution for object detection in multiple scales. The FPN is based on the multilevel feature map using the built-in CNN pyramid architecture. Compared with traditional approaches with image pyramids, the FPN uses vertical (top-down) connections to merge high-level features with exact spatial output positioning. Figure 2.2 shows the comparison of two top-down architectures of pyramid networks:

- Upper: top-down architecture, which skips connections and predictions, is made at the finest level.
- Lower: architecture proposed FPN where predictions are independently made at all levels.

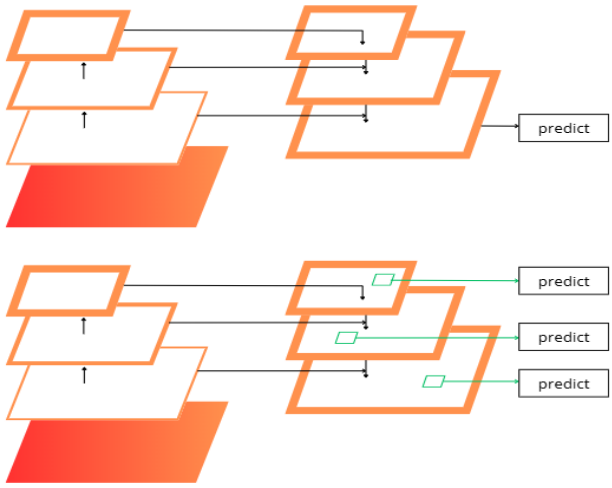

Figure 2.2 *Top-down architectures of networks.*

At first, the system uses single-pixel convolution to merge one feature map and then combines results by applying pixel-wise addition. For a given feature map C_i the process starts by reducing channel dimensions by the convolution layer and then merging with an unsampled coarser-resolution feature map P_{i+1} using element-wise addition in Eq. (2.2).

$$P_i = Conv_{3\times3}\left(Upsample\left(P_{(i+1)}\right) + Conv_{1\times1}(C_i)\right), \quad (2.2)$$

where, P_i is a feature map representation at the pyramid level i.

The two-stage object detection model Fast Region Convolutional Neural Network (FRCNN) combines proposal generation with object classification into an integrated end-to-end system, which is widely used for object detection purposes [45]. The entire input image receives feature map computations from a backbone convolutional network, which all units share. The RPN operates on feature maps by generating diverse bounding boxes at multiple scales and aspect ratios. This process is executed through its sliding mechanism. The most promising object region proposals from the first stage progress to a second stage for processing through the FRCNN detection head. The detection head operates on proposed regions and performs two simultaneous tasks: determining the object class and improving the bounding box localization accuracy.

The following equation indicates each region proposal i, FRCNN uses a multitask loss function, combining localization and classification:

$$L(p_i, u_i, t_i, v_i) = -log({p_i}^{(u_i)}) + \lambda 1[u_i \geq 1] smoothL1({t_i}^{(u_i)} - v_i), \quad (2.3)$$

where:

- ${p_i}^{(u_i)}$ is the predicted probability that i belongs to a ground-truth class u_i;
- ${t_i}^{(u_i)}$ is the predicted bounding box offset for class u_i;
- v_i is the ground-truth bounding box offsets for class u_i;
- $1[u_i \geq 1]$ is an indicator function applying one if u_i a foreground class, and zero if the background class (with label 0);
- $smoothL1$ is a robust L1 loss for bounding box regression to mitigate outliers;
- λ balances the classification and regression terms.

By sharing the same backbone CNN features between the RPN and FRCNN detection head, FRCNN achieves high accuracy while remaining computationally efficient.

The Microsoft Common Objects in COntext (MS COCO) is commonly used as the significant benchmark to evaluate the various models' object detection and segmentation performance [46]. The dataset holds over 200,000 images distributed across eighty object categories in various complex backgrounds. COCO is an industry-standard testing platform for FRCNN detection methods because its complex scenarios with small objects overlapping instances and cluttered backgrounds create evaluation challenges that accurately measure detection system functionality in real-world applications.

Sharing convolutional backbone features between stages lets FRCNN obtain more efficient training than systems that produced proposals before training. The practical implementation involves a multitask loss that maintains an equilibrium between identifying object classes, and refining box position coordinates. FRCNN demonstrates its performance using the MS COCO dataset because this diverse collection includes several small yet overlapping objects to deliver comprehensive evaluation results [47]. FRCNN serves widely as the benchmark standard in DL-based object detection algorithms because of its capacity to produce region proposals through a unified framework.

A specific level of the FPN receives Region of Interest (RoI) data based on its size during FRCNN integration. The size of RoIs determines their assignment to different feature map resolution levels, thus allowing for better scale management without input scale variations [48]. The following equation represents the feature pyramid equation with the value 224 for the canonical ImageNet pre-training size:

$$k = \left[k_0 + log_2\left(\frac{\sqrt{w \cdot h}}{224}\right)\right], \tag{2.4}$$

where:

- k is the index of the pyramid level (P_k) to which RoI is assigned;
- w is the width of RoI on the input image;
- h is the height of RoI on the input image;
- k_0 is the reference pyramid level.

Chapter 3 - Methodology

This thesis introduces a dual path method to facilitate communication-efficient image transmission and robust onboard detection and localization to provide adequate support for real-time situational awareness in UAV-based surveillance systems. Above all, this design is motivated by the need to produce a balance between geospatial accuracy, data bandwidth, and computational power onboard aerial imaging chores in practice. The proposed methodology uses communication-aware and computation-centric strategies to explore separable paradigms in semantic-driven object detection that are context-dependent to the deployment constraint.

The first pipeline considers JSCC, i.e., it compresses input imagery into a representation having relevant semantic features sufficient for robust object detection. For UAVs flying in environments where communication links are constrained, this pipeline is analogous to communication, which is constrained, and full-resolution image transfer to a ground station is infeasible. These compressed representations are then reconstituted and evaluated using high-performing detection architectures, such as FRCNN with FPN, to ensure semantic fidelity and recognition percent. Second, the second pipeline focuses on the feasibility of real-time, onboard object detection and geolocation mapping in parallel. This approach combines detection outputs with Global Positioning System (GPS), altitude, and orientation) from the UAV to produce approximate 3D georeferenced coordinates of the detected objects using Simulated real-world UAV navigation input. Thanks to this functionality, they are instrumental in search and rescue, environmental monitoring, and tactical surveillance scenarios, where GPS-tagged object identification can be acted upon immediately.

3.1 Datasets

AU-AIR Dataset

The AU-AIR [49] dataset provides a multimodal UAV dataset designed explicitly for aerial applications of scene understanding and detection tasks under unpredictable flight scenarios. The dataset contains superior video frames with bounding box annotations covering pedestrian, car, truck, and bicycle classes. AU-AIR separates itself from other datasets by uniting vision data with UAV sensor telemetry data that incorporates GPS coordinates and barometric altitude measurements with Inertial Measurement Unit (IMU) roll, pitch, and yaw readings. This dataset enables the second phase of the proposed method since it supports real-time object detection alongside geolocation functions.

Applying flight metadata to visual data frames enables researchers to generate 3D space estimates for detected objects. Geolocation features are critical for applications that operate by location since they enable understanding spatial relationships between detections and real-world objects alongside their actual detections.

NWPU VHR-10 Dataset

NWPU VHR-10 consists of Very High Resolution (VHR) optical satellite and aerial images documenting various objects from airplanes to ships and storage tanks alongside bridges and harbours [50], [51]. Research that detects objects in dense and cluttered overhead scenes uses NWPU VHR-10 because of its high spatial resolution and precise annotations. This thesis employs NWPU VHR-10 as the foundation for its first semantic compression system based on JSCC. The dataset presents challenging and realistic parameters through its large file sizes and detailed object

features, which enables adequate assessment of compressed representation quality for accurate object recognition tasks.

The research makes its information relevant for UAV transmission of crucial ground station imagery from bandwidth-restricted operations. Post-decoding detection performance assessment and reconstruction quality evaluation occur through the NWPU VHR-10 benchmark.

3.2 Proposed Methods

An investigation is carried out on two separate methodological pipelines that combine object detection functions with real-world coordinate estimates from UAV-acquired aerial imagery. The proposed methodologies focus on two separate operational scenarios inside UAV-based sensing deployments, which cover bandwidth-limited image transportation and spatial object detection at the onboard system. The object detection pipelines use DL models to process data supported by deliberate datasets made to replicate operational aerial surveillance situations. The first approach employs a JSCC autoencoder to compress high-resolution aerial images for transmission over resource-limited communication channels. The software visualizes reconstructed images to test the semantic accuracy of the JSCC data processing framework. The system suits UAV visual transfer operations towards ground stations or edge servers that function under bandwidth restrictions. The second implementation method performs geolocation estimation and onboard real-time object detection simultaneously. Flight telemetry using GPS and IMU readings allows mapping detection results from 2D space into approximate 3D world coordinates. The UAV system becomes self-operational for detecting entities by assigning their position to a universal reference frame needed for critical real-time decisions in integral environments. Through this approach, these methodologies evaluate the ability of semantic compression techniques alongside

spatial detection models in UAV systems to create scalable monitoring platforms. Figure 3.1 describes the structure of the implementation framework with two various approaches.

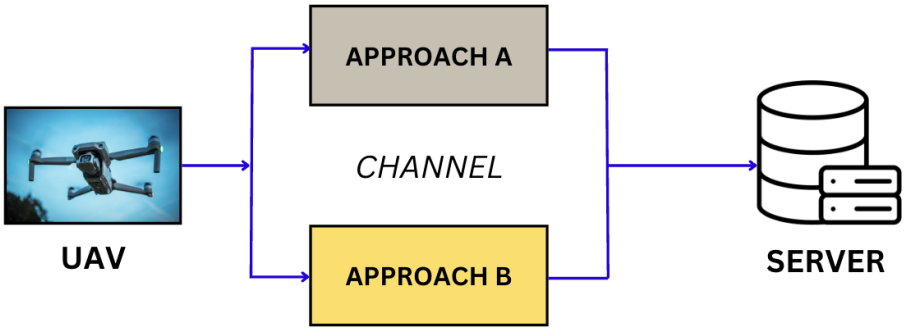

Figure 3.1 *Comparison of two approaches for UAV-to-server data transmission.*

Approach A: JSCC Autoencoder with Object Detection

A DL-based JSCC autoencoder operates within a system that combines object detection with detection accuracy analysis for compression efficiency evaluation. The system allows UAVs to transmit images or other visual data through noisy communication channels that lead to external servers for processing.

JSCC Autoencoder

A JSCC autoencoder is an end-to-end neural architecture because it connects source compression with channel robustness optimization through an integrated encoding-decoding framework. The model achieves transmission impairment resistance by integrating it into learned representations, which eliminates the requirement for traditional coding methods that combine JPEG compression and error-correcting codes.

Autoencoders implement a source-channel coding technique, comping image data through latent codes that resist channel noise degradation. The codes are forwarded across an unreliable

communication channel before being restored by the decoder. A formal description summarizes the process as follows:

$$\hat{x} = f_{dec}[(f_{enc}(x; \theta_{enc}) + n ; \theta_{dec}], \quad (3.1)$$

where:

- x is the input image (size $H \times W \times C$);
- θ_{enc} is the training parameter for the JSCC encoder;
- f_{enc} is the encoder function;
- n is the noise added by a channel;
- θ_{dec} is the training parameter for the JSCC decoder;
- f_{dec} is the decoder function;
- $\hat{x}$ is the reconstructed image.

The encoder section of training receives aerial images and generates compact latent data from their input. The model applies simulated channel noise to the representation before it resembles wireless communication transmission noise in real practice. The decoder completes rebuilding the original image with the noisy representation as input. The training purpose targets minimal reconstruction deficiencies while preserving vital semantic information for succeeding object detection

Evaluation of this component uses NWPU VHR-10 dataset features because it offers high-resolution aerial imagery with various densely arrayed target categories, including airplanes, ships and vehicles. The benchmark is an appropriate evaluation tool to analyze image reconstruction quality and feature modifications.

3.3 Object Detection

The output from the JSCC autoencoder becomes an input to process through a deep object detection framework named FRCNN with FPN enhancement. The detector was chosen because of its proven ability to detect objects at different scales, which reflects the characteristic nature of aerial views that contain targets of various sizes.

The model generates detection results through class identifications and rectangular box locations around detected objects to calculate quantitative preservation metrics of meaning after compression occurs. The detection results in inventory against different channel noise levels and compression ratios utilize mean AP (mAP) metrics for performance evaluation.

The pipeline provides a structured evaluation between semantic compression quality and object detection system performance. The system delivers essential data about UAV imagery data transmission efficiency and its ability to generate receiver-end inference outputs.

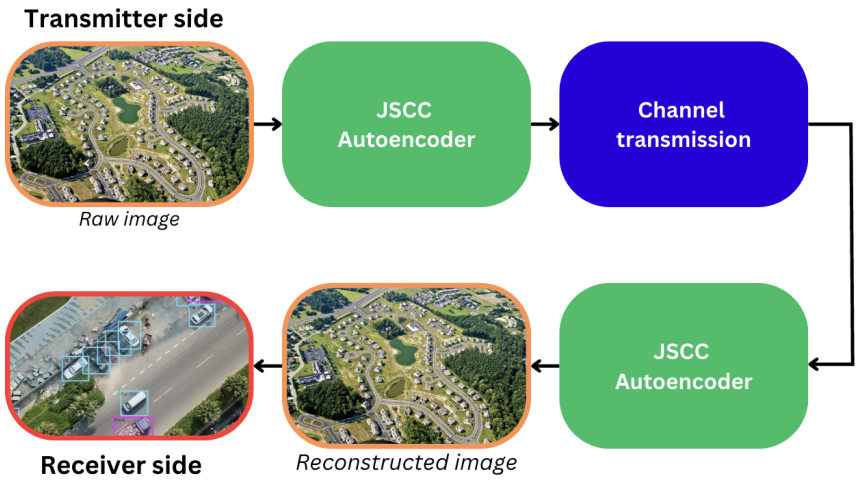

Figure 3.2 *Pipeline of the JSCC-based approach.*

Figure 3.2 presents a wireless channel transmission path for aerial images through the JSCC framework. The JSCC autoencoder receives images at the transmitter end for simultaneous compression and encoding operations. After encoding, the transmission signals proceed to the

channel. A receiver-side JSCC autoencoder operates to decode received data and generate an image reconstruction. A detection or analysis task occurs on the reconstructed image, which shows detected objects through bounding boxes according to the diagram's lower section.

Approach B: Onboard Object Detection and Real-World Coordinate Calculation

A direct object detection process occurs onboard the UAV platform before analyzing onboard telemetry data to calculate detected object geographical locations. This network operates for instantaneous detection while tracking locations when ground networks are inaccessible or crucial mission-needed geographic results are necessary for critical decisions like surveillance, disaster relief operations, and environmental studies.

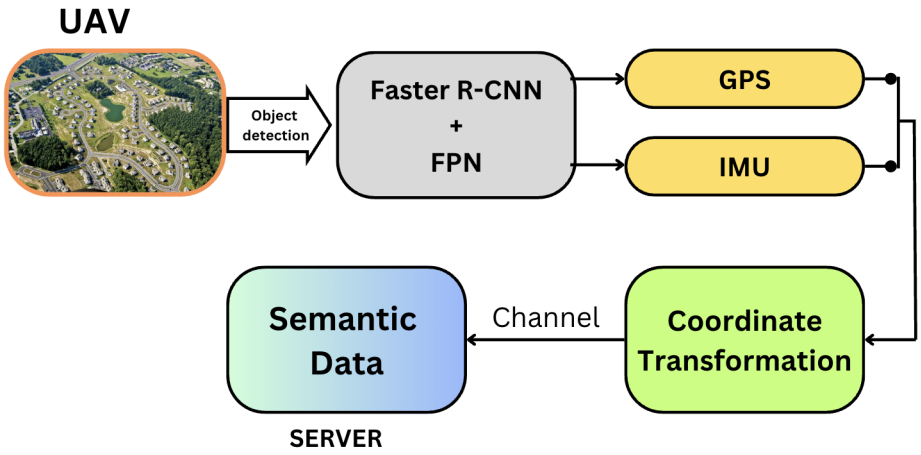

Figure 3.3 *On-UAV detection with real-world coordinate transformation.*

Figure 3.3 depicts a UAV taking an aerial image, the image being processed on board by a FRCNN with an FPN object detection model. At the same time, GPS and IMU data correlate the UAV's pose and location to objects detected. Semantic data containing the resulting detections and relevant flight parameters are transmitted to a server. The semantic detections and the UAV's flight data are combined on the server side to convert the bounding boxes to real-world coordinates or positions through a coordinate transformation module.

UAVs utilize this setup to execute real-time object detection by installing FRCNN integrated with FPN on their onboard computing unit. Real-time video stream analysis or static image processing with the front-facing camera of the UAV utilizes object detections along with image frame bounding box positions as part of the processing model.

The network removes time-consuming image transmission requirements between UAV and ground stations through local inference operations, which improves system response speed and reduces communication delays. The detection capabilities continue operating successfully through conditions that cause interruptions in network connectivity or temporary network outages. The detection architecture selection occurred because it demonstrates effective multi-scale object detection capabilities, optimal accuracy, and computer processing speed, which suits edge device deployment constraints.

The world coordinate calculations utilize 2D image-space detection projection that benefits from multiple sensor data types inside the AU-AIR dataset. The accurate UAV pose information in the dataset includes GPS coordinates, altitude values, and orientation data through the IMU, which reports roll pitch and yaw measurements.

Camera projection models and ground-plane assumptions perform the transition from 2D to 3D space by conducting a series of geometric operations. The method projects centroids from bounding boxes through reference planes starting from camera altitude and ending at camera angle to determine detected objects' physical positions. The geospatial mapping function allows the system to convert detected visual cues into precise latitude, longitude, and elevation statements for each object. A bounding box center detected in images requires projection to a global coordinate by multiplying its image coordinates with a combination of the camera intrinsic matrix estimated depth measurement and UAV roll-pitch-yaw orientation parameters.

The transformation is given by Eq. (3.2).

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = R_{\phi,\theta,\psi} \left(K^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \times d \right) + t \,, \tag{3.2}$$

where:

- (u, v) is the center of coordinates of the detected bounding box of the object in pixel space;
- K is the camera intrinsic matrix;
- d is the estimated depth or distance from the camera to the object;
- $R_{\phi,\theta,\psi}$ is the 3D rotation matrix constructed for roll ϕ, pitch θ, and yaw ψ;
- t is the translation vector that locates the camera in a world coordinate system based on GPS reference;
- $(X\ Y\ Z)^T$ are the real-world coordinates of the object.

The ability to localize in real-world settings is a critical base for UAV systems that need to execute autonomous operations requiring mission planning. The UAV obtains data about object spatial arrangements through geometry algorithms, which enables usage in operations such as area mapping and target tracking alongside coordinated flights of multiple UAVs.

3.5 Training Procedures

The section details training strategies applied to both proposed approaches by explaining preprocessing methods and architectural setups with loss functions and implementation-specific hyperparameters. The training process seeks optimal performance from each component using unique restrictions either for data transmission or real-time object detection.

Data Preprocessing

Standard image preprocessing occurs before training commences on NWPU VHR-10 and AU-AIR datasets. All images get standardized through appropriate normalization parameters for their pre-trained backbone networks, including ResNet. The resizing operations follow standards matching model input dimensions by converting images to 256×256 pixels based on the available system memory and operational difficulty.

Data augmentation techniques, including random horizontal flipping, small-angle rotation, and color transformation, improve model generalization during training. Specific augmentations add extra value to aerial datasets because objects show different orientations due to camera movement. For the JSCC autoencoder (Approach A), the training and validation subsets are created by splitting the dataset according to an 80/20 ratio. The autoencoder obtains resilient semantic feature compression through exposure to multiple object categories and scene arrangements.

Model Architectures

FRCNN is the primary object detection architecture because it provides strong accuracy and flexible features. Improved performance in aerial imagery depends on implementing the FPN framework. The FPN feature network helps the detector perform better object detection on distant small targets that frequently appear in UAV image data.

The JSCC autoencoder is a convolutional network in Approach A for encoder-decoder operation. A noisy transmission channel simulation (including the Gaussian noise layer) enables the encoder to reduce input images into latent space, and then, ultimately, the decoder brings these images back. The module exists for either standalone training or simultaneous training with an object detection model to protect vital recognition features throughout compression.

Detection training during the first phase of Approach B occurs exclusively with the AU-AIR dataset images as input. The system integrates sensor fusion during the last stage, which unites visual detections and UAV telemetry data, including GPS and IMU measurements, to produce geospatial coordinate calculations through geometric transformation operations. Detection accuracy receives improvement through the use of camera calibration parameters when these parameters become available.

Multitask Loss and Optimization

In particular, the object detection framework uses a multitask loss consisting of a classification loss (usually the cross entropy) and a regression loss (smooth L1) for bounding box refinement. This formulation allows the model to predict object categories and their spatial locations with precision at the same time.

The objective for the JSCC autoencoder is to train it to minimize the loss between the original and the reconstructed image as reconstruction loss, one of which can occur as Mean Squared Error (MSE). Some of this may be combined with a perceptual loss (for example, using a pre-trained VGG network) to encourage retention of high-level visual features necessary to perform tasks like detection later on.

In end-to-end training of the JSCC detector pipeline, a joint loss function favours reconstruction fidelity, and detection performance is used to prevent the compression from causing the loss of critical semantic information.

Hyperparameters and Implementation Details

DL with large-scale aerial image data is modelled and run with GPU-accelerated neural frameworks like PyTorch to guarantee efficient handling of such extensive data processing. We

follow standard practices on model initialization, training schedule, and regularization in terms of whether or not to take advantage of pre-trained models while also adjusting to the characteristics of pre-trained models and learning dynamics for tasks.

Also, the FRCNN detector is trained using the Adam optimizer and a learning rate of 0.00025, as we want the training to be the fastest while maintaining reasonable results. This conservative learning rate is a choice naively dictated by imposing a conservative learning rate to fine-tune the pre-trained backbone without upsetting the pre-learned representations, and it is key to the transfer learning from large datasets like ImageNet. The model is trained for 10,000 iterations before the detector is trained for the aerial imagery domain and retains its first feature extraction capabilities. The JSCC autoencoder is trained over 50 epochs using the Adam optimizer with a learning rate 0.01. The higher the learning rate, the faster autoencoder architecture can be trained from scratch; thus, in the learning paradigm of the early stage of training, its convergence is relatively faster. The main training objective reduces reconstruction error while being robust to imposed channel impairments in the form of simulated channel artifacts.

3.4 Evaluation and Metrics

Each proposed pipeline undergoes evaluation through quantitative metrics that fully match their specific methods' operational goals. Standard object detection metrics, image reconstruction quality indicators, compression efficiency measures, geospatial localization accuracy, etc., are those included in these. The evaluation shows the performance tradeoffs between accuracy levels, efficiency, and real-world operability capabilities when using UAV-based operational restrictions. The mean Average Precision (mAP) is used as the primary metric to evaluate object detection performance, as it quantifies the detector's capability of associating objects and localizing them

given the overlap between the proposed and ground truth bounding box. Using COCO-style scores, mAP is computed as AP over several Intersections over Union (IoU) thresholds ranging from 0.5 to 0.95 in steps of 0.05. Thus, mAP provides a rigorous and comprehensive measure of detection accuracy. On the other hand, VOC-style mAP is reported at a fixed IoU threshold of 0.5 and is easier but frequently used as a benchmark for object detection.

In Approach A, the analysis of the compression levels vs. detection accuracy relationship is essential. The impact of detection mAP is tracked by systematically varying the configuration of the JSCC autoencoder (determining the size of latent space or noise level in the simulated channel). Thus, it is possible to perform a quantitative analysis for a tradeoff between the benefits of bandwidth savings and the necessary preservation of semantic content for object recognition.

The quality of the reconstructed images from the JSCC autoencoder is also evaluated using standard image reconstruction criteria besides the detection metrics. MSE defines the overall pixel-wise difference between the original and reconstructed images as low-level visual fidelity. Peak Signal Noise Ratio (PSNR) is a logarithmic scale where higher values are low distortion levels in the reconstructed images. Perceived image quality is assessed by comparing structural, luminance, and contrast information and is evaluated based on the Structural Similarity Index Measure (SSIM). These metrics can verify that spatial and textural features important in detection are retained in the compressed representations. Evaluating transmission efficiency is also reported for the compression and image compression ratios.

For Approach B, real-world coordinate estimation accuracy is the focus of the evaluation. Telemetry Data such as GPS coordinates, altitude, and the UAV's orientation obtained from its IMU is used to map the bounding boxes predicted by the onboard detection model to a ground plane. Projected positions are then compared to reference ground truth data (which is available) or

evaluated as being spatially consistent by comparing the projections from consecutive frames. The error variable of interest is usually the average positional error expressed in meters, which is the difference between the estimated and actual object locations.

Additional checks are done in temporal reliability to ensure localization reliability over time, especially in different flight dynamics. In particular, this aspect is necessary when the target is in motion (e.g., for a mission scenario involving moving targets, autonomous navigation, or coordinating a team of agents).

Chapter 4 - Results

Finally, the experimental results are presented from the two approaches suggested in Chapter 3. Measurements of semantic communication and real-time geospatial inference in UAV-based object detection systems are to be made within practical bounds of limited bandwidth, computing power, and realistic variability. The results are divided into two main parts. In Part I (Section 4.1), the semantic compression pipeline of aerial images is compressed using JSCC autoencoder and then processed by an object detection model. This section investigates the tradeoff between compression level and detection quality using reconstruction metrics, MSE, PSNR, and SSIM, over different compression configurations.

The results of the onboard detection and real-world localization approach are presented in Part II (Section 4.2). Object detection is done locally on the UAV, and bounding boxes are projected onto geospatial coordinates from flight telemetry data (IMU and GPS). Detection accuracy and positional error in estimated object locations on the ground are included in the evaluation. Thus, these results show that it is possible, robust, and has concrete repercussions for integrating semantic communication and intelligent geolocation in UAV-based IIoT systems.

4.1 Part I – Semantic Compression Analysis (Approach A)

During the semantic compression pipeline training process, one of the core elements is the regularization parameter λ, which allows a tradeoff between the reconstruction of visual features and the preservation of semantic meaning. It is formulated for this purpose of optimization.

For 50 epochs, a learning rate of 0.01, the Adam optimizer was used to train the JSCC autoencoder. Training and validation loss curves in Figure 4.1 show quick convergence in the first 10 epochs

and do not overfit much. These losses imply that the model generalizes well with the dataset, and the reconstruction quality does not vary much with training.

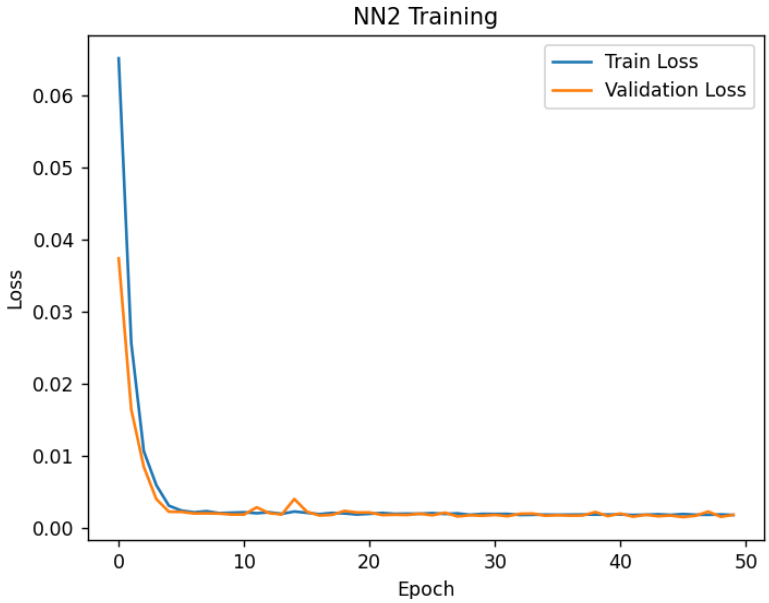

Figure 4.1 *JSCC autoencoder training and validation loss curves.*

The training formulation aims to learn the projection matrix W such that

$$\min_{W} \|X - W^T S\|_F^2 + \lambda \|S - WX\|_F^2 , \tag{4.1}$$

where:

- X is the visual feature matrix;
- W is the linear projection matrix;
- S is the semantic attribute matrix;
- λ is the regularization term.

The optimization problem is to choose the value λ which will give the best results in terms of semantics. Many compression control parameter λ values were tested to identify the best compromise between compression and image quality. The variations in MSE, PSNR, and SSIM over different λ values are shown in Figure 4.2. It is found that usually, a smaller λ leads to a better

reconstruction quality and that λ ≈ 0.05–0.1 gives the best tradeoff in terms of MSE. MSE gets less than 0.01, PSNR is above 24 dB, and the SSIM is close to 0.78, which means that the perceptual similarity to the original images is high in those settings. However, the goal is to find the optimal value of λ, and it can be seen that λ = 0.15 is the best value of λ.

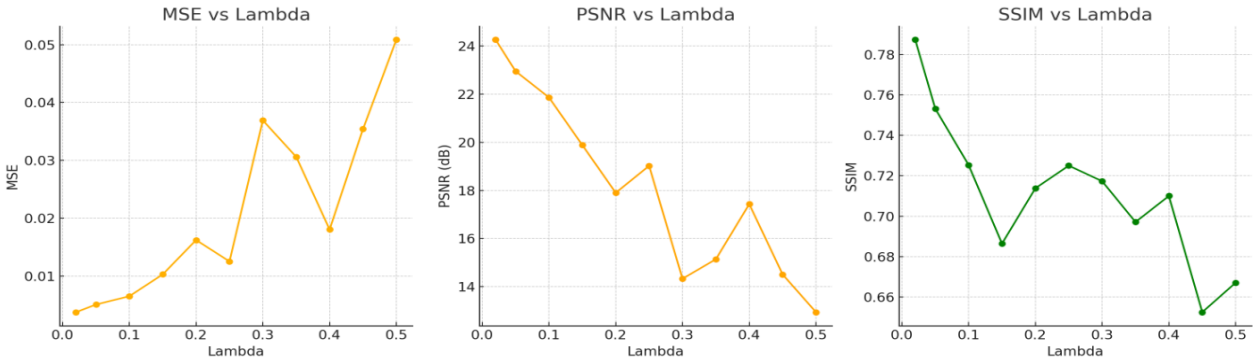

Figure 4.2 *Reconstruction metrics (MSE, PSNR, SSIM) across varying λ values.*

The reconstruction performance was finally evaluated using image-wise analysis over the test dataset, as shown in Figure 4.3. Below are summarized the metrics achieved through the tuning of λ value:

- PSNR Mean ≈ 28 dB;
- MSE Mean ≈ 0.003;
- SSIM: Mean ≈ 0.70;
- Compression Ratio ≈ 1.6.

These performance values confirm that the JSCC autoencoder can compress and reconstruct high-resolution aerial images so that the structural and semantic aspects continue to apply to proper object detection.

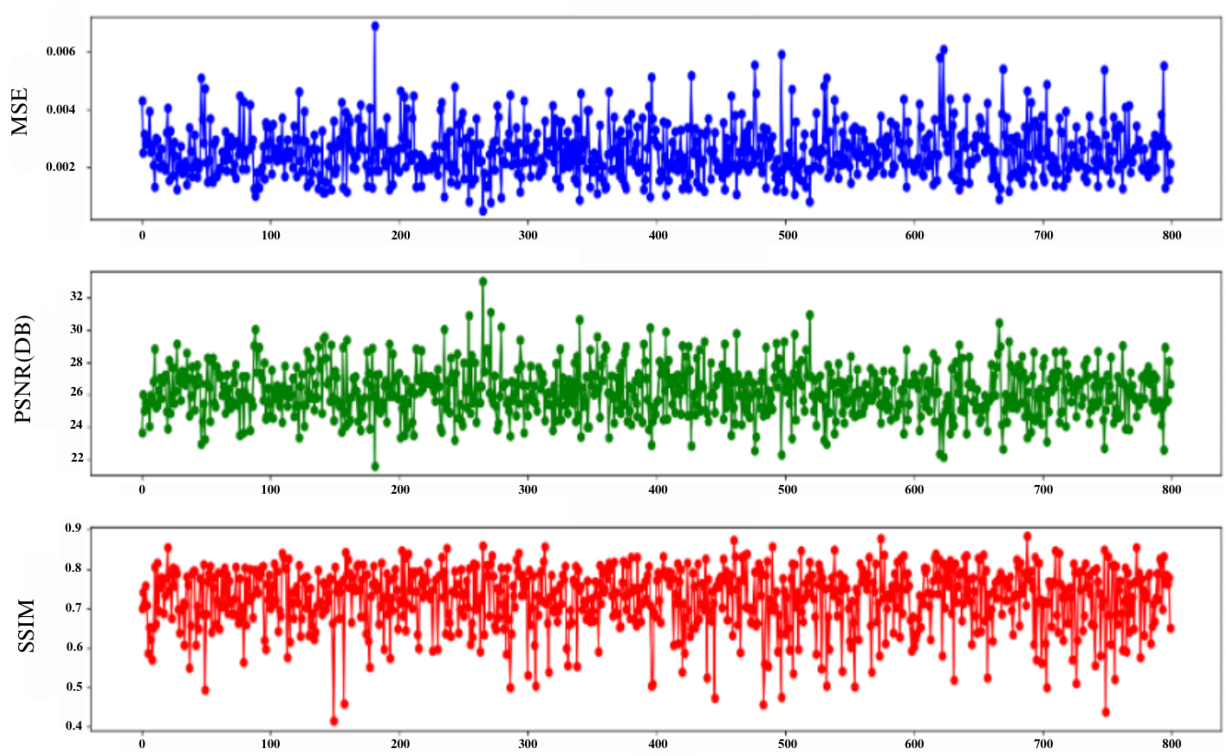

Figure 4.3 *Image-wise evaluation of MSE, PSNR, and SSIM over the test set.*

Figure 4.4 represents the visualization results of the feature extraction. The features remain the same meaning as the original image.

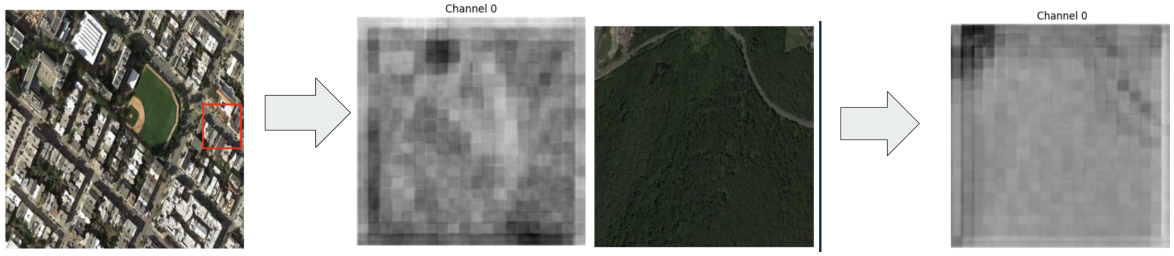

Figure 4.4 *Results of feature extraction.*

It can be observed that the features of the images contain the same pattern. The second image curve can also be found in its features. Thus, the feature visual representation of the compressed image is considered correct.

Figure 4.5 shows the results of the JSCC encoder after the compression and reconstruction. The model performs well with high-detailed images and low-detailed images.

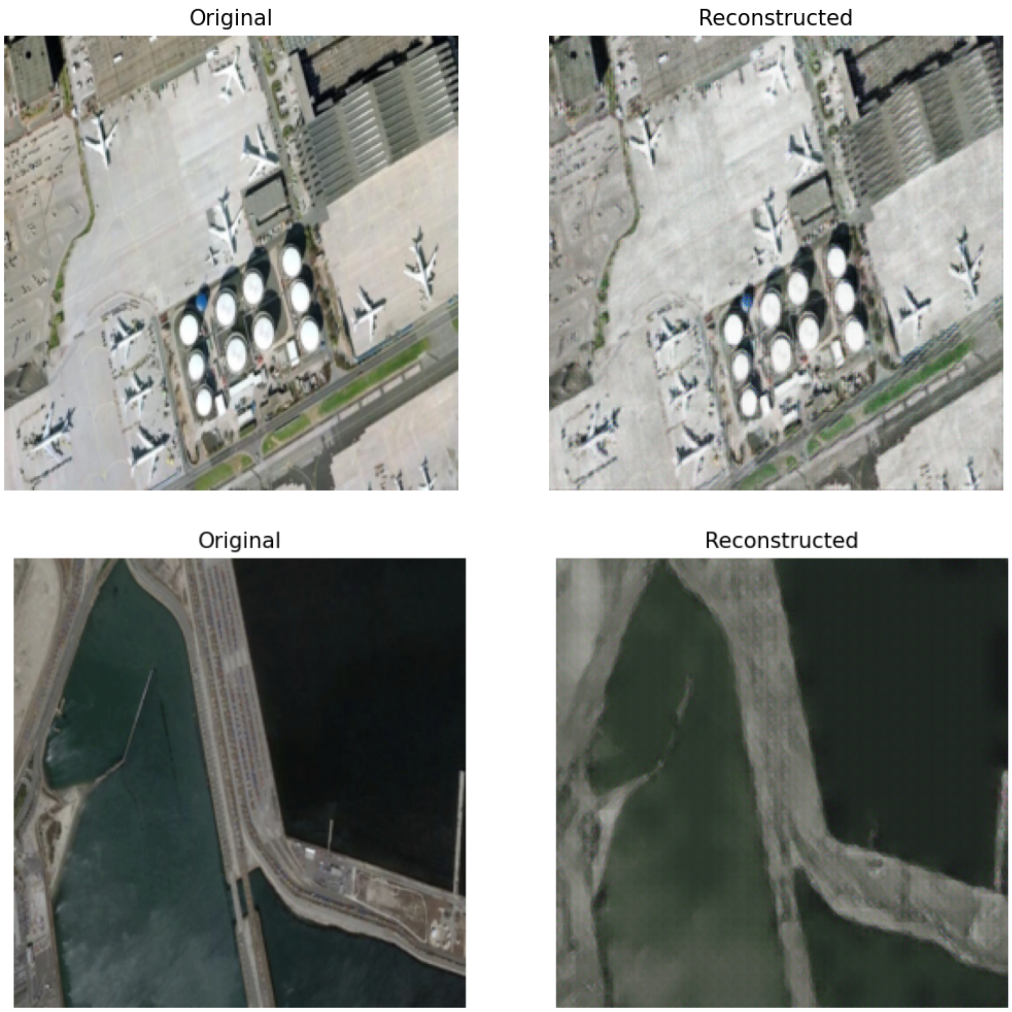

Figure 4.5 *Compression and reconstruction results.*

4.2 Part II – Onboard Object Detection and Localization (Approach B)

The results from experimentally running the second pipeline, which performs onboard object detection and approximate geolocation based on UAV-acquired imagery and flight telemetry, are presented in this section. The objective is to measure the mapping of 2D detections on real-world coordinates and the detection accuracy. The FRCNN detector algorithm was trained and fine-tuned

on the AU- AIR dataset, which was used for detection. The performance of the proposed detector is shown in Table 4.1, with various IoU thresholds in the evaluation protocols.

Table 4.1 COCO evaluation metrics for onboard object detection.

Metric	**Value**
AP@[0.50:0.95] (COCO mAP)	7.35%
AP@0.50	21.53%
AP@0.75	3.13%
AP (Small Objects)	0.44%
AP (Medium Objects)	4.72%
AP (Large Objects)	10.68%
AR@[0.50:0.95], maxDets=100	26.00%

The results suggest the detector is best at large and well-separated objects, which are more visible in UAV imagery. However, performance on small objects and dense, cluttered scenes remains a common challenge in aerial perception due to scale and resolution during performance.

Class-wise Performance Analysis

Figure 4.6 shows the per-class AP representing variance performance according to object type. However, the detector performs relatively well on "harbour", "plane", "vehicle", and "ground-track-field" but performs much less well with "container crane", "tennis court", and "basketball court". The cause of this disparity is often class imbalance, diversity of samples, and overlapping visual features.

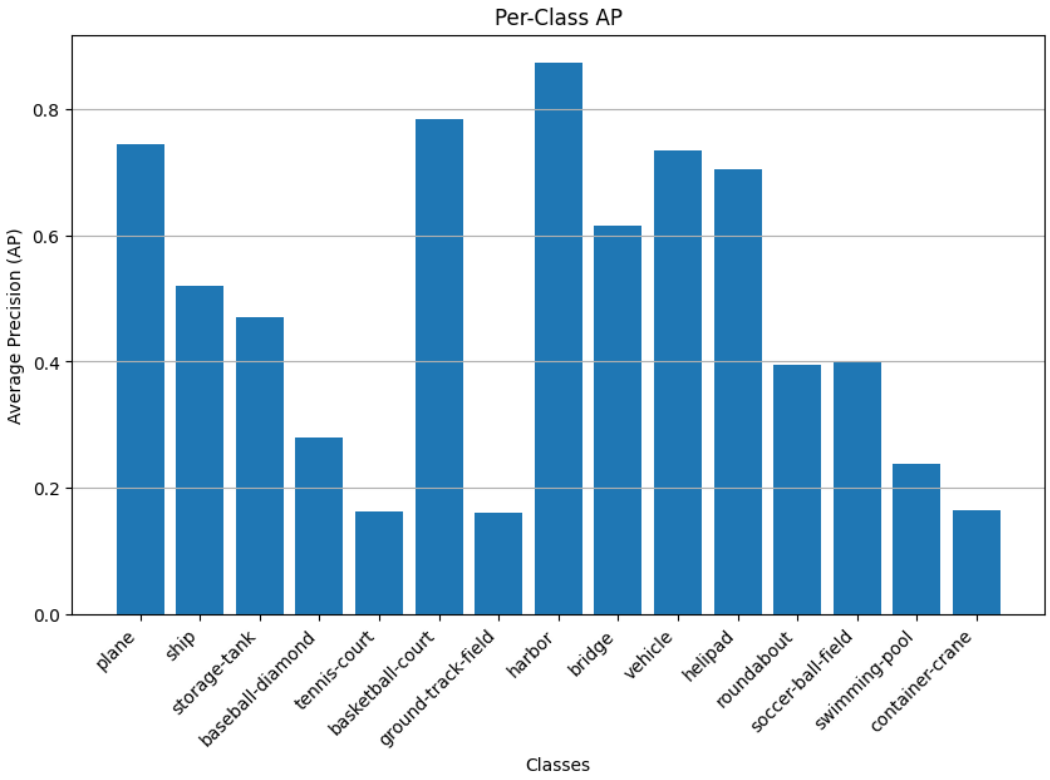

Figure 4.6 *Per-class AP for AU-AIR object categories.*

Confusion Matrix Analysis

Also, the confusion matrix in Figure 4.7 gives more insight into class-level misclassifications. Indeed, the diagonal dominance confirms accuracy predictions for many well-studied classes, and off-diagonal activity reveals confusion between visually similar categories. This supports the need for contextual reasoning and better feature extraction under real UAV conditions. Defined classes in COCO format are: plane, ship, storage tank, baseball diamond, tennis court, basketball court, ground track field, harbour, bridge, vehicle, helipad, roundabout, soccer ball field, swimming pool container crane. However, the classes of vehicles and infostructure objects are the most important. The classes for the athletic games are used to detect class mismatch errors.

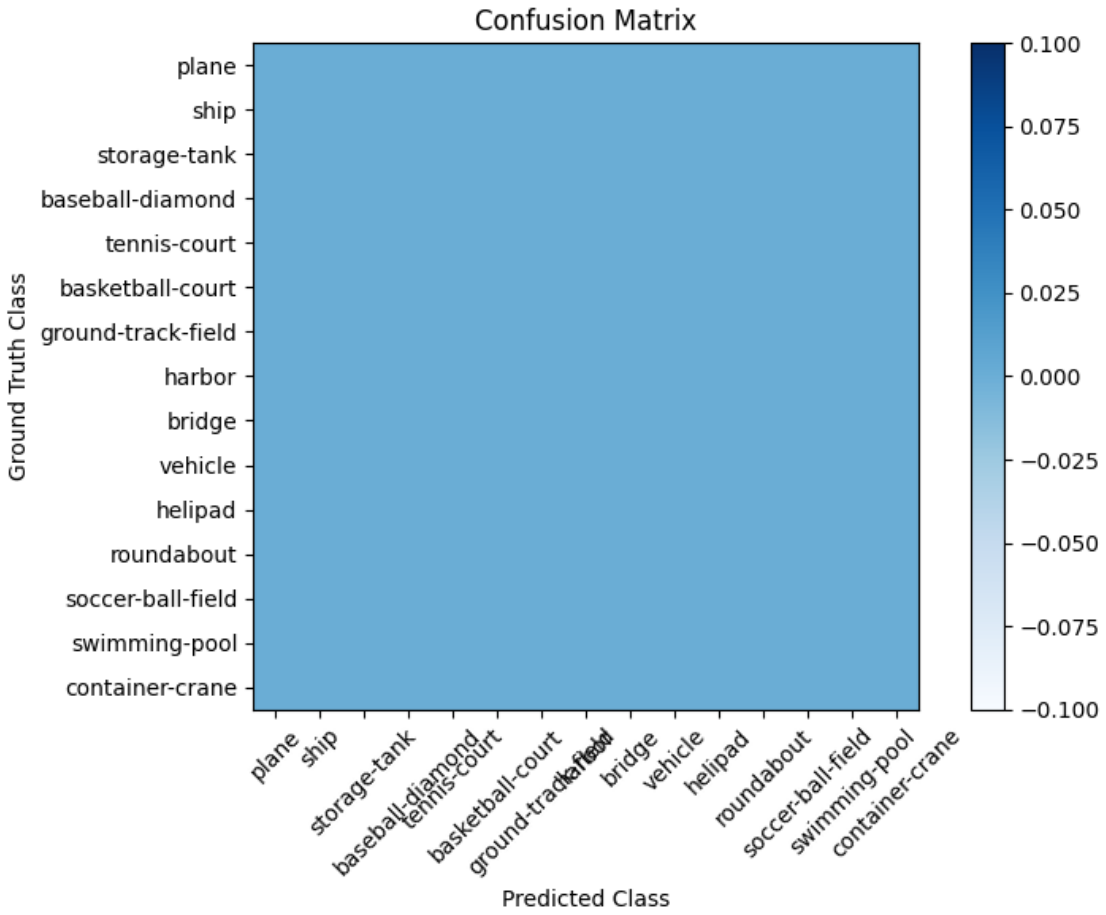

Figure 4.7 *Normalized confusion matrix for detected AU-AIR classes.*

Qualitative Detection Output

Figure 4.8 shows a sample inference result of the model, which can accurately detect multiple such as planes, ships, storage tanks, and infrastructural buildings as its bounding box confidence score. Based on the spatial alignment with scene structures, the model retains reasonable detection precision in UAV viewpoints. The visual results corroborate these quantitative results, and although they show applicability in maritime monitoring or infrastructure inspection, they also provide much-needed visual confirmation.

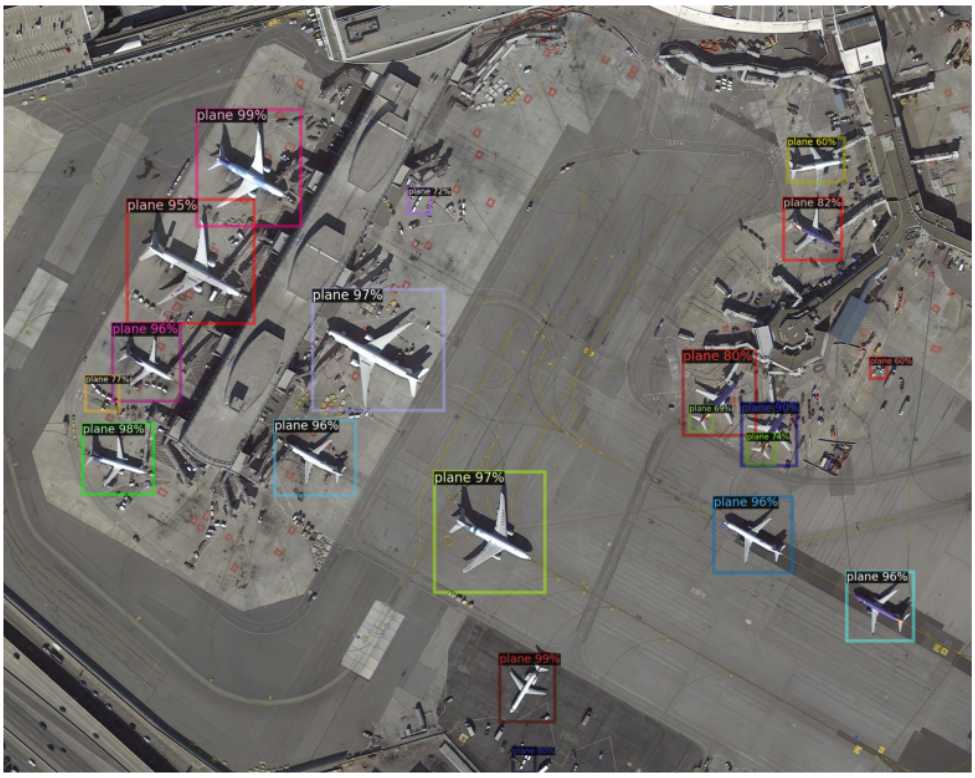

Figure 4.8 *Visual output of onboard detection with confidence scores.*

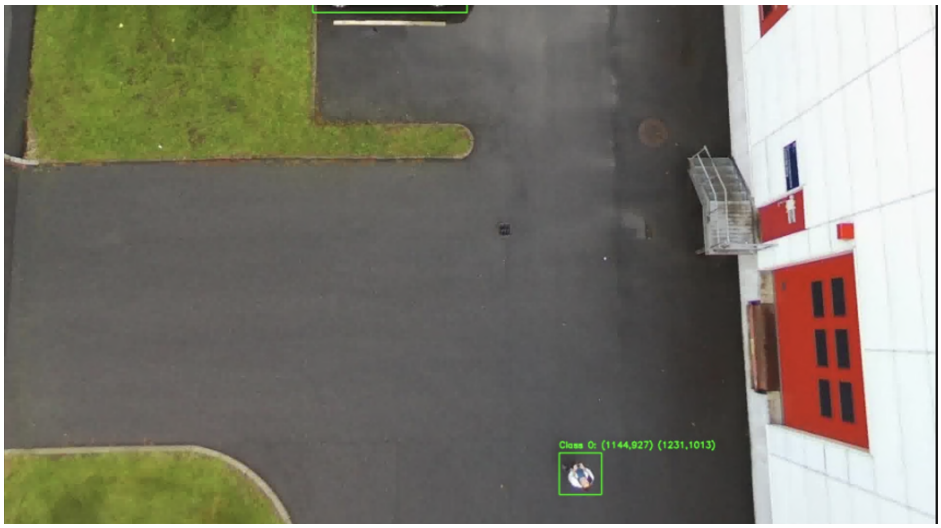

Figure 4.9 *Visual output with real-world coordinates.*

Figure 4.9 highlights the recognition with the object coordinates. In this case, it detected the object as class 0, which belongs to the people.

Chapter 5 - Conclusion

This research provided two complementary DL-based pipelines that enable efficient and intelligent aerial surveillance in IIoT applications via UAVs. Under bandwidth-constrained environments, a semantic compression strategy based on JSCC was used in the first pipeline, by which UAVs compressed compact yet semantically rich image representations to a server. The second pipeline supported real-time geolocation of onboard detected objects and integrated visual and flight telemetry data.

It is shown through extensive experimentation that the JSCC approach can have high compression ratios while keeping its visual features sufficient for reliable object detection. The selected λ-optimised model resulted in a mean PSNR of 28 dB, a mean SSIM score of 0.70, and MSE as small as 0.003, demonstrating that the compressed structures are semantically relevant.

Secondly, it was used to perform detection onboard the UAV using the AU-AIR dataset. On the other hand, the detector achieved a COCO mAP of 7.35% over the IoU thresholds and outperformed significantly on large objects (AP = 10.68%) compared to small objects. This onboard system was further expanded to perform real-world coordinate estimation, i.e. converting 2D detections into approximate 3D locations using GPS and IMU information. The ability to do this demonstrated the feasibility of autonomously and contextually monitoring UAVs in time-sensitive applications.

An essential practical takeaway from Approach B is that it is sufficient to transmit the object class labels and real-world coordinates to the server rather than full-resolution images or even original compressed features. In particular, this is useful for UAV platforms (15–30 km altitude) with expensive or severely constrained data transmission (long-range or stratospheric platforms). This paradigm drastically improves the scalability and autonomy of UAV systems operating in remote

or infrastructure-less environments by lowering the communication to a minimum of essential semantic content.

The research presents fundamental components that enable the development extensive IIoT UAV systems with intelligent capabilities. The hybrid transmission network between semantics and operational speed gives users access to enhanced performance from image delivery technologies and onboard analytic capabilities through context-based processing. The research findings verify that edge computing can run inference tasks by sending semantic coordinates and class information to preserve operational intelligence through minimal data transfer. The concepts extend UAV applications to create design patterns for multiple autonomous IIoT systems, such as robot grid systems and surveillance swarms. Only semantic-based protocols that rely on meaning instead of raw data will connect through the current approach, thus expanding future network capabilities for real-time interpretability and responsiveness. Real-world generalization remains challenging even though the research demonstrates effective results from simulated situations and controlled datasets. This thesis delivers a usable design concept with two approaches that enable the following generation of IIoT systems to achieve connectivity while acquiring semantic intelligence and contextual understanding.

Chapter 6 - Discussion and Future Work

6.1 Discussion

This thesis demonstrates from the findings that the accuracy and efficiency of semantic compression are tradeoffs in bandwidth-limited UAV systems. A scenario where image-level insights are still critical and approach A is well suited for when semantic fidelity is of the highest importance is when image-level insights are crucial and when centralized decision is required. In contrast to Approach B, Approach B focuses on edge-level processing with low latency, and therefore, inference must occur on board with minimal data offload. This approach is preferable for real-time alerting, edge autonomy, or remote environments.

However, limitations remain. The drop in detection performance as the compression becomes high or channel noise becomes severe suggests that compression-aware detectors should be more robust. Since resolution and motion artifacts from UAV flight affect the detection performance on small and medium-sized objects, the detection performance in Approach B is suboptimal. Besides, the localization error in the real world is acceptable but can be improved if the depth estimation is more accurate or if sensor fusion is used.

Semantic compression vs Exact regeneration: prioritizing meaning over fidelity in IIoT

The evaluation method for traditional compressed communication systems focuses on maintaining perfect image rebuilds with low data loss. IIoT networks and UAV-based surveillance systems experience a conceptual change in processing and transmission approaches. This fundamental question in the research asks whether exact compression methods are necessary when feature simplification leads to successful regeneration results. Regeneration methods that use semantically

meaningful features prove sufficient and superior to exact pixel matching, specifically when dealing with limited environments.

The JSCC autoencoder in approach A compressed images by maintaining important semantic content instead of precise visual elements. Evaluation metrics established the compressed images' capability to maintain sufficient structural information through SSIM scores passing 0.70 PSNR reaching 28 dB and MSE remaining at 0.003, thereby supporting downstream activities like object detection. The operational accuracy of the object detection model validated this finding since essential features needed for inference continued intact even without exact pixel reconstruction.

Implementing the semantic compression algorithm brings a substantial advantage in maintaining system operation stability during bandwidth-constrained situations, commonly occurring in UAV system field operations. The system retains priority for transmitting meaningful semantic features rather than full high-fidelity images because this approach reduces data transmission requirements — a fundamental concept for decision-supporting intelligent IIoT systems. Mission-critical situations, such as disaster relief and infrastructure inspection, gain significant advantages from swift decision-making because delays could produce profound effects. The implementation of semantic representations proves compatible with inference operations that use AI technology. FRCNN maintains its operational capability using compressed feature sets, which provide the features with sufficient information about object shape edges alongside contextual indicators. Such compression algorithms can adapt their output by considering the priority level of different information units, which reduces transmission redundancy. The research findings support sending data according to its meaningful content rather than literal data. The system requires information that allows it to function effectively rather than precisely duplicating the original scene content. These principles will guide developers toward building adaptive encoding systems edge-AI

platforms and semantic priority protocols that select between video frames, feature data, or basic labels based on priority and situation.

Adaptability of the semantic communication framework to ground-based IIoT platforms

System retraining capabilities of the semantic communication and object detection systems for ground vehicles and robotic platforms deal with fundamental design concerns associated with IIoT system generalizability. Strong evidence shows that the JSCC autoencoder and FRCNN-based detector operate as data-driven and modular components requiring minimal structural adjustment to function efficiently on alternative platforms. The JSCC autoencoder operates with image-based information passing through simulated communication exchanges. The encoder-decoder design operates without constraints from the data source under representative sample training. The retraining procedure for ground vehicle operations requires substituting aerial dataset sources (NWPU VHR-10 or AU-AIR) with terrestrial dataset collections of UAV-recorded data. The core semantic compression concept exists regardless of perspective since it maintains significant task-related information while discarding unnecessary details. The semantic autoencoder trained on street-level imagery would identify vehicle shapes, road signs, and pedestrians as essential features after exposure to this dataset domain. However, aircraft and harbour detection currently remain the priority of aerial imagery.

The object detection module (FRCNN with FPN) represents a strong architectural approach that works well for new object class and environmental training needs. This detection system performs multiple scale analysis that lets it recognize objects of varying sizes and positions that vehicle operators encounter in the field. The domain transfer process becomes highly efficient when trained detectors with small ground datasets are applied, mainly when backbone networks,

including ResNet and MobileNet, are used. This system design enables the efficient reuse of visual features while allowing detector head retraining to be the sole primary requirement.

The geospatial localization procedure from Approach B utilizes object detection data fusion with telemetry (GPS and IMU) sensors to produce 3D world coordinates and can operate effectively with ground vehicles. These algorithms produce outputs using proper extrinsic calibration methods and accurate sensor inputs. Accurate position estimation in areas without GPS access becomes possible by adding wheel odometry and LiDAR data.

The architectural aspect of this research introduces a common semantic communications framework that operates between various IIoT devices, including UAV drones, autonomous rovers and industrial robots. Smart cities warehouses and logistics systems require increasing multi-agent collaboration, which makes this approach highly important. The next version of this system should use domain adaptation methods, transfer learning methods and federated learning methods to permit agents to modify a shared model with their specific sensor inputs without breaking its semantic foundation.

The system extends beyond UAVs for its applications. This system consists of modular sections, general-purpose vision and learning models, and transferrable semantic principles, enabling seamless integration across multiple IIoT platforms following ground-based operational requirements for scalability and cross-domain knowledge acquisition.

Balancing hardware constraints and algorithmic efficiency in IIoT UAV systems

The deployment of semantic communication and onboard inference in UAV systems faces major obstacles because DL algorithms require more resources than UAV platforms usually have. The research implements the JSCC autoencoder with FRCNN, yet these methods demonstrate strong performance while requiring extensive computing power. Small to medium UAV platforms face

complex challenges while performing real-time inference because their onboard power capacity, battery capacity and memory storage remain constrained.

The design necessitates algorithmic efficiency as a solution to resolve this issue. The thesis exploited pre-trained models with FPN and multitasking capabilities to cut down redundancy in operations. Through JSCC autoencoder training, the model learned how to maintain semantic information while reconstructing data for direct end-to-end processing, minimizing the number of required network passes along with data movement requirements, which helps lower GPU energy usage specifically on NVIDIA Jetson Nano and Xavier embedded devices.

To achieve widespread deployment, more optimization measures must be implemented. The deployment efficiency of models significantly improves when practitioners implement strategies, which include MobileNet or YOLOv5n lightweight architectures, model quantization, and model pruning techniques. The method of adaptive computation allows systems to skip or simplify inference processes when performing low-priority visual frame analysis to save power during prolonged operations. The system's performance and onboard capacity tradeoffs must be balanced for optimal functionality, leading to the increased operational readiness of intelligent UAV systems within real-industrial IoT deployments.

6.2 Future Work

Future research can be conducted to improve the semantic compression pipeline and the onboard localization pipeline. One promising direction is integrating multiple modes of data sources like LiDAR, infrared, or depth sensors to make object detection robust and estimate real-time position. Temporal information across video sequences may also help improve detection consistency and reduce frame-level errors. Another improvement should be applied to detection architectures to be

onboard deployed for lightweight models (such as MobileNet or transform-based detectors) to allow real-time inference on resource-constrained UAV platforms. Moreover, adaptive transmission policies whereby the decision to send forth full detections, compressed features, or just semantic coordinates would be helpful to improve the bandwidth efficiency at different operating conditions. On a final note, federated or swarm intelligence frameworks could be deployed to collaboratively improve the model for several UAVs without data exchange across the sites, therefore enhancing system scalability and autonomy.

Bibliography

[1] Malik, P.K., Sharma, R., Singh, R., Gehlot, A., Satapathy, S.C., Alnumay, W.S., Pelusi, D., Ghosh, U. and Nayak, J., "Industrial Internet of Things and its applications in industry 4.0: State of the art," Computer Communications, vol. 166, pp. 125-139, Jan 2021.

[2] Hu, Y., Jia, Q., Yao, Y., Lee, Y., Lee, M., Wang, C., Zhou, X., Xie, R. and Yu, F.R., "Industrial Internet of Things intelligence empowering smart manufacturing: A literature review," IEEE Internet Things Journal, vol. 11, pp. 19143-19167, Jun 2024.

[3] A. D. Wyner, "Recent results in the Shannon theory," IEEE Transactions on Information Theory, vol. 20, pp. 2-10, Jan 1974.

[4] Anguita, J., Djordjevic, I., Neifeld, M., and Vasic, B., "Shannon capacities and error-correction codes for optical atmospheric turbulent channels," Journal of optical networking, vol. 4, pp. 586-601, Sep 2005.

[5] C. E. Shannon, "A mathematical theory of communication," The Bell system technical journal, vol. 5, pp. 3-55, Jan 2001.

[6] Bao, J., Basu, P., Dean, M., Partridge, C., Swami, A., Leland, W., and Hendler, J. A., "Towards a theory of semantic communication," IEEE Network Science Workshop, pp. 110-117, Jun 2011.

[7] B. Chen and J. Wan, "Emerging trends of ML-based intelligent services for Industrial Internet of Things (IIoT)," Computing, Communications and IoT Applications, pp. 135-139, Oct 2019.

[8] P. Strauß, M. Schmitz, R. Wöstmann, and J. Deuse, "Enabling of predictive maintenance in the brownfield through low-cost sensors, an IIoT-architecture and Machine Learning," IEEE International conference on big data (big data), Seattle, WA, USA, pp. 1474-1483, Dec 2018.

[9] K. C. Ravi, T. Vaishnavi, V. A. Kandaswamy, M. Dinesh, and S. Lakshmisridevi, "Revolutionizing IIot: A Smart Appliance of an Internet of Things Over Industrial Manufacturing Unit to Make Error-Free Circumstances," International Conference on Cybernation and Computation (CYBERCOM), Dehradun, India, pp. 314-319, Nov 2024.

[10] H. Xie and Z. Qin, "A lite distributed semantic communication system for Internet of Things," IEEE Journal on Selected Areas in Communications, vol. 39, pp. 142-153, Jan 2021.

[11] Khalil, R. A., Saeed, N., Masood, M., Fard, Y. M., Alouini, M. S., and Al-Naffouri, T. Y., "Deep Learning in the Industrial Internet of Things: Potentials, challenges, and emerging applications," IEEE Internet Things Journal, vol. 8, pp. 11016-11040, Jul 2021.

[12] A. Sheth, "Internet of Things to smart IoT through semantic, cognitive, and perceptual computing," IEEE Intelligent Systems, vol. 31, pp. 108-112, Apr 2016.

[13] R. Carnap and Y. Bar-Hillel, "An outline of a theory of semantic information," Research Laboratory of Electronics, MIT, Cambridge, MA, USA, Tech. Rep. 247, pp. 221-274, Oct 1952.

[14] B. H. Juang, "Quantification and transmission of information and intelligence—history and outlook," IEEE Signal Processing Magazine, vol. 28, pp. 90-101, Jul 2011.

[15] H. Baqa, N. B. Truong, N. Crespi, G. M. Lee, and F. Le Gall, "Quality of Information as an indicator of Trust in the Internet of Things," Proc. 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data (TrustCom/BigDataSE), New York, NY, USA, pp. 204-211, Aug 2018.

[16] P. Basu, J. Bao, M. Dean, and J. Hendler, "Preserving Quality of Information by using semantic relationships," Pervasive and Mobile Computing, vol. 11, pp. 188-202, Feb 2014.

[17] Wang, Y., Chen, M., Luo, T., Saad, W., Niyato, D., Poor, H. V., and Cui, S., "Performance optimization for semantic communications: An attention-based reinforcement learning approach," IEEE Journal on Selected Areas in Communications, vol. 40, pp. 2598-2613, Sep 2022.

[18] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep Learning Enabled Semantic Communication systems, "IEEE Transactions on Signal Processing, vol. 69, pp. 2663-2675, Mar 2021.

[19] Z. Qin, H. Xie, and X. Tao, "Mem-DeepSC: A Semantic Communication System with Memory," Proc. International Conference on Communications (ICC), Rome, Italy, pp. 3854-3859, May 2023.

[20] L. Deecke, L. Ruff, R. A. Vandermeulen, and H. Bilen, "Transfer-based semantic anomaly detection," International Conference on Machine Learning (ICML), pp. 2546-2558, Jul 2021.

[21] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, "Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions," IEEE Access, vol. 7, pp. 137184-137206, Dec 2019.

[22] T. Erpek, T. J. O'Shea, Y. E. Sagduyu, Y. Shi, and T. C. Clancy, "Deep learning for wireless communications," Development and Analysis of Deep Learning Architectures, S. A. R. Shah, Ed., pp. 223-266, May 2020.

[23] N. Farsad and A. Goldsmith, "Neural network detection of data sequences in communication systems," IEEE Transactions on Signal Processing, vol. 66, pp. 5663-5678, Nov 2018.

[24] H. He, C. K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," IEEE Transactions on Signal Processing, vol. 68, pp. 1702-1715, Feb 2020.

[25] M. U. Lokumarambage, V. S. S. Gowrisetty, H. Rezaei, T. Sivalingam, N. Rajatheva and A. Fernando, "Wireless end-to-end image transmission system using semantic communications," IEEE Access, vol. 11, pp. 37149-37163, Apr 2023.

[26] S. Dörner, S. Cammerer, J. Hoydis, and S. ten Brink, "Deep learning based communication over the air," IEEE Journal of Selected Topics in Signal Processing, vol. 12, no. 1, pp. 132-143, Feb 2018.

[27] Sun, H., Chen, X., Shi, Q., Hong, M., Fu, X., and Sidiropoulos, N. D., "Learning to optimize: Training Deep Neural Networks for interference management," IEEE Transactions on Signal Processing, vol. 66, no. 20, pp. 5438-5453, Oct 2018.

[28] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep learning in physical layer communications," IEEE Wireless Communications, vol. 26, pp. 93-99, Apr 2019.

[29] H. Ye, G. Y. Li, B.-H. F. Juang, and K. Sivanesan, "Channel agnostic end-to-end learning based communication systems with conditional GAN," Proc. IEEE Globecom Workshops (GC Wkshps), Abu Dhabi, UAE, pp. 1-5, Dec 2018.

[30] S. Park, O. Simeone, and J. Kang, "End-to-end fast training of communication links without a channel model via online meta-learning," Proc. IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Atlanta, GA, USA (Virtual), pp. 1-5, May 2020.

[31] L. Song, D. Gildea, Y. Zhang, Z. Wang, and J. Su, "Semantic neural machine translation using AMR," Transactions of the Association for Computational Linguistics, vol. 7, pp. 19-31, Apr 2019.

[32] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv:1409.0473, 2014. [Online]. Available: https://arxiv.org/abs/1409.0473

[33] Z. Xuan and K. R. Narayanan, "Low-delay analog joint source-channel coding with deep learning, "IEEE Transactions on Communications, vol. 71, no. 1, pp. 40-51, Jan 2023.

[34] 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, pp. 567-579, Sep 2019.

[35] U. Naseem, I. Razzak, S. K. Khan, and M. Prasad, "A comprehensive survey on word representation models: From classical to state-of-the-art word representation language models," Transactions on Asian and Low-Resource Language Information Processing, vol. 20, Art. no. 34, May 2021.

[36] P. Kamboj, S. Kumar, and V. Goyal, "Measuring and mitigating gender bias in contextualized word embeddings," IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS) pp. 1-5, Oct 2023

[37] Xu, Y., Yan, X., Wu, Y., Hu, Y., Liang, W., and Zhang, J., "Hierarchical bidirectional RNN for safety-enhanced B5G heterogeneous networks," IEEE Transactions on Network Science and Engineering, vol. 8, pp. 2946-2957, Dec 2021.

[38] A. Graves, "Generating sequences with recurrent neural networks," arXiv:1308.0850, Aug. 2013. [Online]. Available: https://arxiv.org/abs/1308.0850

[39] S. Rodzin, V. Bova, Y. Kravchenko, and L. Rodzina, "Deep learning techniques for natural language processing," in Computer Science On-line Conference, ser. Lecture Notes in Networks and Systems, vol. 423, pp. 121-130, Apr 2022.

[40] Albouq, S. S., Abi Sen, A. A., Almashf, N., Yamin, M., Alshanqiti, A., and Bahbouh, N. M., "A survey of interoperability challenges and solutions for dealing with them in IoT environment," IEEE Access, vol. 10, pp. 36416-36428, Sep 2022.

[41] Y. Yu, R. Chen, H. Li, Y. Li, and A. Tian, "Toward data security in edge intelligent IIoT," IEEE Network, vol. 33, pp. 20-26, Oct 2019.

[42] V. A. Thakor, M. A. Razzaque, and M. R. Khandaker, "Lightweight cryptography algorithms for resource-constrained IoT devices: A review, comparison and research opportunities," IEEE Access, vol. 9, pp. 28177-28193, Sep 2021.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, Lake Tahoe, NV, USA, pp. 1097-1105, Dec 2012.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 770-778, Jun 2016.

[45] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S."Feature pyramid networks for object detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 2117-2125, Jul 2017.

[46] R. Girshick, "Fast r-cnn," Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, pp. 1440-1448, Dec 2015.

[47] Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L., "Microsoft coco: Common objects in context," Computer Vision – ECCV, ser. Lecture Notes in Computer Science, Cham, Switzerland: Springer, vol. 8693, pp. 740-755, Sep 2014.

[48] Meiyin Wu and Li Chen, "Image recognition based on deep learning," 2015 Chinese Automation Congress (CAC), Wuhan, pp. 542-546, Nov 2015.

[49] I. Bozcan and E. Kayan, "AU-AIR: A Multimodal Unmanned Aerial Vehicle Dataset for Low Altitude Traffic Surveillance," IEEE International Conference on Robotics and Automation (ICRA), Paris, France (Virtual), pp. 8399-8405, Aug 2020.

[50] Su, H., Wei, S., Yan, M., Wang, C., Shi, J., and Zhang, X., "Object Detection and Instance Segmentation in Remote Sensing Imagery Based on Precise Mask R-CNN," IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Yokohama, Japan, pp. 1454-1457, Jul 2019.

[51] Su, H., Wei, S., Liu, S., Liang, J., Wang, C., Shi, J., and Zhang, X., "HQ-ISNet: High-Quality Instance Segmentation for Remote Sensing Imagery," Remote Sensing, vol. 12, Art. no. 989, Mar 2020.