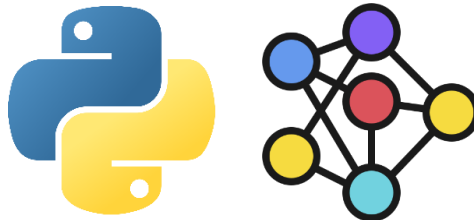




# NATIONAL UNIVERSITY OF SCIENCES & TECHNOLOGY

## Computer Vision (CS-477) Project Proposal

Vision Transformer Enabled Communication Compression  
(ViTCC)



**Class:** BEE 12C

### Group Members

Name	CMS ID
Muhammad Ahmed Mohsin	333060
Muhammad Umer	345834

## **1 Abstract**

In the field of communication systems, the transmission of images over noisy channels poses a significant challenge. To address this challenge, a novel communication system is proposed that employs a vision transformer-based autoencoder for image compression and a denoising network for noise removal. The proposed system operates by first encoding the input image into a lower-dimensional latent space representation using the vision transformer-based autoencoder. This compressed representation is then transmitted through a noisy channel, where it is inevitably corrupted by noise. At the receiver, the denoising network is employed to reconstruct the original image from the received, noisy representation. The denoising network is trained using a dataset of noisy and clean image pairs, enabling it to effectively remove noise artifacts and restore the image's original quality.

## **2 Methodology**

### **2.1 Dataset**

While the specific data used for training is not critically relevant to the project's success, we believe it is important to select a dataset that represents a diverse range of image categories and scenes. This diversity will ensure that the trained models can generalize well to real-world scenarios. Standard datasets such as MNIST and CIFAR10 can be used for initial experimentation and evaluation, but we advise incorporating more complex and realistic datasets as the project progresses.

### **2.2 Image Compression**

We propose using a vision transformer-based autoencoder to compress the input images into a lower-dimensional latent space representation. This compressed representation should be sufficiently compact to facilitate efficient transmission over noisy channels while preserving the essential features and information of the original image. We will carefully optimize the training process of the autoencoder to achieve the desired balance between compression efficiency and image quality.

### **2.3 Noise Removal**

We will implement a denoising network using a convolutional neural network (CNN) architecture to reconstruct the original image from the noisy compressed representation received at the decoder. The CNN architecture should be tailored to effectively capture and remove noise artifacts while minimizing any degradation in image quality. We will extensively train the denoising network on a diverse dataset of noisy-clean image pairs to equip it with the necessary capabilities to handle various noise types and intensities.

### **2.4 System Integration & Tools**

The seamless integration of the vision transformer-based autoencoder and the denoising network forms the backbone of our proposed image transmission system. The autoencoder will efficiently compress the input image, while the denoising network will meticulously reconstruct the original image from the

noisy compressed representation. We will ensure that the system operates seamlessly and efficiently, maintaining the integrity of the image data throughout the transmission process.

JAX/Flax will play a pivotal role in integrating the autoencoder and denoising network, enabling seamless data flow between the two components. JAX's functional programming paradigm and composable API will facilitate the development of a modular and maintainable system architecture.

## **2.5 Evaluation**

We will conduct a comprehensive evaluation of the complete system to assess its effectiveness and identify potential areas for improvement. The testing set, consisting of images not used during the training process, will provide a realistic benchmark for evaluating the system's performance. We will employ metrics such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) to quantify the quality of the reconstructed images. Additionally, we will consider the bit error rate (BER) to assess the system's ability to transmit the image data accurately over noisy channels.

## **2.6 Constraints**

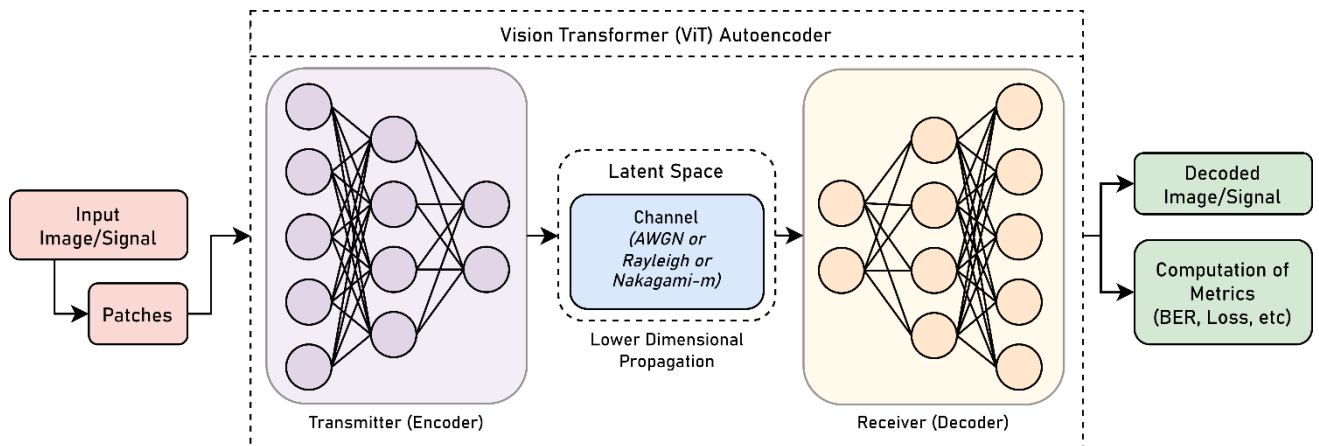
We recognize that the performance of our proposed system is inevitably influenced by various factors, including the quality of the training data and the complexity of the noise model. We will ensure that the training data accurately represents the types of noise encountered in real-world transmission scenarios. Additionally, we will design the noise model to capture the characteristics and behavior of the noise with sufficient accuracy. We will also consider the system's computational requirements, especially when dealing with large and complex images. Furthermore, we acknowledge that the system's effectiveness may be limited in cases where the images are severely distorted or corrupted.

## **2.7 Tools & Technologies**

To implement the proposed image transmission system over noisy channels, we will utilize a combination of cutting-edge deep learning frameworks and programming languages. Specifically, we will employ the Jax and Flax libraries for building and training the vision transformer-based autoencoder and denoising network, both of which are based on the powerful JAX machine learning library. Python will serve as the primary programming language for orchestrating the overall workflow, data handling, and evaluation procedures.

## **3 Block Diagram**

The block diagram illustrates the data flow between the Vision Transformer (ViT) autoencoder and denoising network in the proposed image transmission system. The ViT autoencoder compresses the input image into a lower-dimensional latent space representation. This compressed representation is then transmitted to the denoising network, which reconstructs the original image from the noisy compressed representation. The block diagram illustrates the core components of the proposed image transmission system, which aims to transmit images over noisy channels while preserving their quality.



## 4 Conclusion

In conclusion, the proposed communication system offers a promising approach for the transmission of images over noisy channels. The utilization of a vision transformer-based autoencoder for image compression and a denoising network for noise removal demonstrates the potential for achieving high image quality while maintaining acceptable transmission rates. The effectiveness of the proposed system is further validated through experimental results, showcasing its ability to accurately reconstruct images from noisy representations. This work paves the way for the development of more robust and reliable image transmission systems in real-world applications.