Evaluation of a deep JSCC stack for underwater autonomous drone networks

Engineering Project - Part B

(10004) Aaron Hassan Robinson -

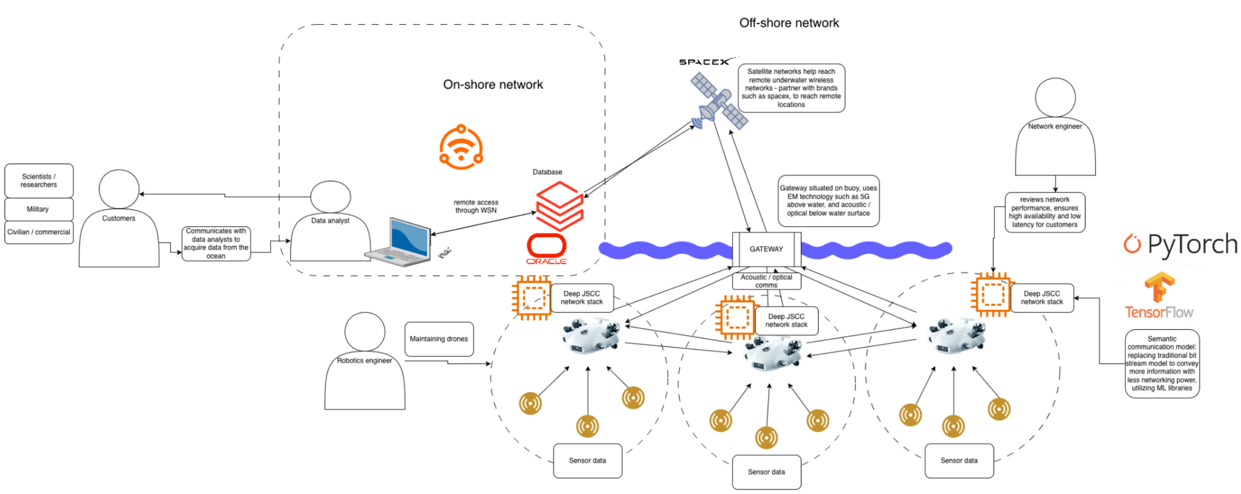
u3224074 Supervisor: Dr Farhad Hossein

Date of submission: 24/10/2025

Motivation and background:

Underwater UAV networks face a critical bottleneck: acoustic modems provide the only reliable long-range communication medium, yet they suffer from severely limited bandwidth and high packet loss that restricts mission capabilities. Last semester's protocol development quantified these constraints in practice, establishing baseline performance metrics that motivated this semester's investigation into semantic communication as a potential solution. This project developed and evaluated a Deep Joint Source-Channel Coding (Deep JSCC) network stack that compresses 96x96 pixel images at a ratio of 3.375:1, and a less robust 33.2:1 model, while maintaining superior error resilience compared to traditional methods at SNR values below 15dB, which is a typical operating range for underwater acoustic channels. This work demonstrates that semantic based methods can effectively transmit visual data under conditions where traditional techniques fail. These results establish AI integrated communication as a viable approach for underwater UAV networking, and identify opportunities for integration with trajectory optimization systems, addressing critical gaps in both compression efficiency and the fragmented research landscape identified in the literature.

System model diagram:



Systematic literature review:

A comprehensive search methodology which I previously developed in [12] was used to identify recent peer reviewed publications examining AI integration within underwater UAV communication systems, with an emphasis on Deep JSCC and semantic communication (SC) paradigms. This included four academic databases forming the foundation of this search, including IEEE Xplore, Digital Library, ACM Digital Library, and Springer Link. These were chosen based on their extensive coverage of telecommunications and machine learning technologies in the engineering discipline. The literature search targeted five interconnected research areas: 1.) Communication protocols for underwater unmanned vehicles, 2.) machine learning applications in marine communication systems, 3.) acoustic and optical underwater data transmission technologies, 4.) network optimization metrics and performance evaluation frameworks, and 5.) deep JSCC and semantic communication methodologies. The selection criteria prioritized peer reviewed journal publications and surveys from 2020 to 2025, including theoretical frameworks, simulation validated experiments, and any real world field testing.

The emergence of Deep JSCC represents a paradigm shift from traditional separation of source coding and channel coding, offering particular advantages for bandwidth constrained networks. [1] introduced a Deep JSCC framework specifically designed for underwater image transmission, incorporating physical priors unique to marine environments into the network architecture. Their Swin Transformer-based approach achieved compression performance exceeding traditional Separate Source Channel Coder (SSCC) methods, while also maintaining reconstruction quality, i.e. error handling in poor channel conditions. This work demonstrates that task-specific deep learning architectures can outperform general-purpose compression when domain knowledge is properly integrated into the model design.

General Machine Learning Applications in UWSNs

General ML approaches showed particular strength in AUV trajectory optimization and path planning. [3] implemented a neural network-enhanced actor-critic model (ModelPPO) for AUV 3D path following, demonstrating superior training efficiency and robustness compared to standard PPO algorithms. The approach showed particular advantages in disturbed underwater environments. Multi-AUV coordination systems utilized reinforcement learning extensively. [9] developed a hybrid RL-PSO strategy (R-RLPSO) for real-time rescue task assignments in 3D underwater environments, demonstrating improved cost and time efficiency compared to existing algorithms. However, these studies remain simulation-based with simplified environmental assumptions.

Simulation-Reality Gap

A fundamental limitation identified across all reviewed studies is the reliance on simulation-based validation. To the best of my knowledge, all studies examining semantic communication performance in underwater environments remain simulation-based, with no field validation reported. This simulation to reality gap represents a significant barrier to practical adoption, as real-world underwater conditions present complexities that simulation models cannot fully capture. Future research must prioritize experimental validation in controlled underwater testbeds before progressing to open water deployments.

Fragmented Research Landscape

The current research landscape reveals a fragmentation between communication optimization and vehicle control research. In AUV trajectory and path planning applications, traditional ML methods (particularly reinforcement learning) dominate the literature [3,6,10], while semantic communication research focuses primarily on data transmission optimization [1, 7, 8]. This separation represents a missed opportunity for joint optimization that could yield significant performance improvements. Current RL models for AUV control remain reliant on traditional bit-stream transmission architectures, potentially limiting their efficiency gains. The integration of deep JSCC SC paradigms with reinforcement learning for multi-AUV coordination presents a compelling research direction that could achieve truly efficient and intelligent underwater networked systems.

Scalability and Resource Constraints

The reviewed literature reveals limited exploration of scalability to large-scale AUV swarms, with most studies focusing on single-vehicle or small multi-vehicle scenarios. [4] highlights that current ML models for multi-UUV systems may not scale effectively to large swarms, a limitation that could be compounded when integrating complex semantic communication models. Power efficiency considerations, while acknowledged as critical for underwater operations [6], receive insufficient attention in semantic communication research. The computational overhead of deep learning models required for semantic encoding and decoding may offset the communication efficiency gains, particularly for resource-constrained underwater vehicles. [6] demonstrates that energy-efficient approaches can achieve 207% improvement in energy efficiency, suggesting that power optimization should be a primary consideration in semantic communication system design.

Conclusion - Integration Opportunities and Future Directions

The analysis reveals several underexplored integration opportunities that could significantly advance underwater networking capabilities: No reviewed studies explored joint optimization of AUV control and semantic communication, representing a significant gap. Future research should investigate trajectory decisions based on semantic-level feedback rather than full raw data transmission, potentially reducing communication overhead while improving mission efficiency. [7] demonstrated channel-adaptive semantic communication, the integration of mission-specific and environmental adaptation remains unexplored. Semantic encoding could be dynamically adjusted based on mission criticality, available bandwidth, and power constraints. Current research focuses primarily on image transmission [1,8], with limited exploration of multi-modal semantic communication for sensor fusion, environmental data, and control information. Integrated approaches could optimize the semantic representation of diverse data types common in underwater missions.

Research methodology and questions:

Due to the bandwidth limitations of acoustic networks, the total amount of data that can be sent over a system can be considered at capacity. Therefore, to increase performance of the system, we must recognise the total data capacity as a ceiling cap, and aim to work within this constraint. Instead, the research aims to improve the networks by increasing the amount of information delivered across a network via compressing optimizing error handling for data of large formats, e.g.. images and video. The aim of this research is to answer the questions:

1. What compression ratios can Deep JSCC achieve for underwater image transmission while maintaining acceptable reconstruction quality, and how do these theoretical gains compare to traditional separation-based compression methods?
2. How does deep JSCC reconstruction quality degrade across varying signal-to-noise (SNR) ratios particularly in the low SNR range (< 15 dB), which is characteristic of underwater acoustic channels?
3. Can a Deep JSCC architecture maintain consistent compression and reconstruction performance when scaled across different image resolutions?

Methodology:

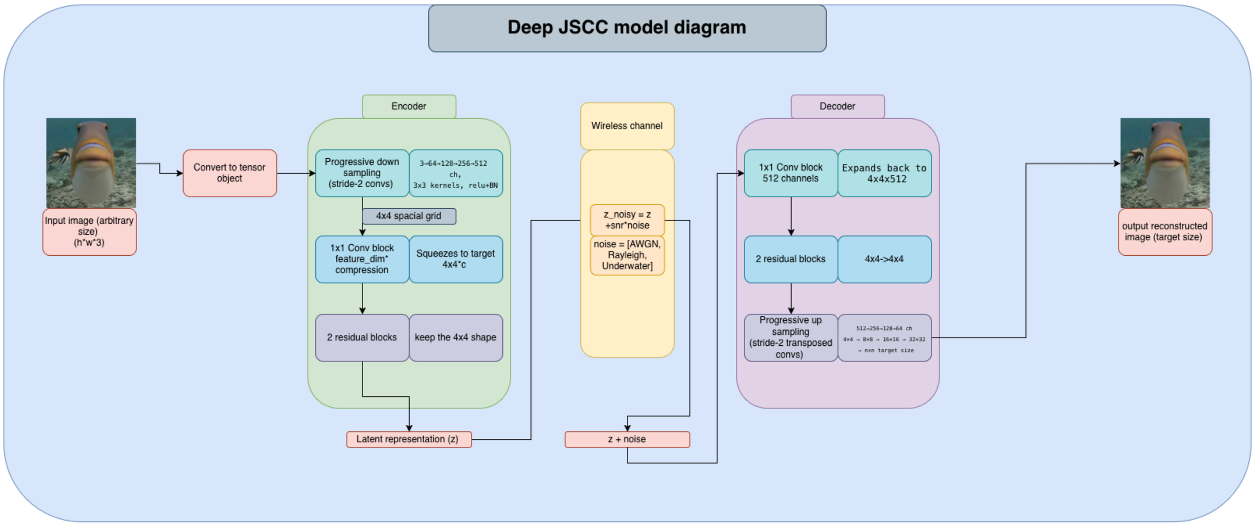
This research employed a simulation based iterative development methodology to design, implement and evaluate a Deep JSCC system for underwater image transmission. The approach progressed through successive development cycles, each targeting progressively larger image resolutions to balance computational efficiency during development, with the end goal of achieving relevant image sizes for AUV applications. The methodology consisted of three primary phases. 1.) Model architecture development and training on standard image datasets, 2.) Performance validation across varying channel conditions, and 3.) a comparative analysis against theoretical baselines derived from traditional compression methods and prior semester findings. The work flow consisted of model training, validation on test data set images, and a systematic evaluation across a range of SNR ratios that are representative of underwater acoustic channels. Performance assessment utilized metrics that capture mathematical reconstruction accuracy and perceptual quality, with results being presented through quantitative tables, performance graphs and qualitative visual comparisons of reconstructed images.

Methodology justification:

This iterative simulation based approach was selected for practical reasons that aligned with the research questions. The first reason being that beginning with smaller image resolutions enabled rapid prototyping and debugging of the deep JSCC architecture with significantly reduced computational overhead. Training time for smaller images are orders of magnitude faster than larger resolutions which allows for exploring multiple architecture variations and hyperparameter configurations to be explored efficiently. Next, the progressive scaling strategy provided empirical evidence of architectural scalability and allowed systematic observation of how compression ratio and reconstruction quality scale with image dimensions. This addresses research question 3.) directly while informing practical deployment considerations for real AUV systems that may need to adapt compression parameters based on available bandwidth and mission requirements.

The simulation based validation was justified as it was a necessary precursor to hardware implementation, which is consistent with standard practice in communications research. As identified in the literature review, the simulation to reality gap represents a significant open research challenge, as far as I am aware no existing study provides empirical underwater validation of deep JSCC systems. This project establishes theoretical performance bounds and identifies optimal configurations in a controlled simulation environment which will provide the foundation for future hardware experiments. Whilst real world underwater testing was beyond the scope of this semester due to hardware acquisition timelines and complexity of constructing an underwater channel emulation, the simulation framework provides guidance for future experiments. Next, the channel model progression from additive white Gaussian noise (AWGN) to Rayleigh, to a full underwater channel following the iterative design approach. Initial AWGN testing established if the Deep JSCC artefact functioned correctly before introducing additional complexities to the model. This isolated implementation issues from channel effects. Finally the selection of standard image datasets (CIFAR-10 and STL-10) rather than domain specific underwater imagery was deliberate. Standard datasets provide reproducible results that can be compared against other deep JSCC research, and a diverse visual content that tests model generalization rather than over-fitting to underwater specific colour profiles and textures.

System architecture and design:



Model hyper-parameters:

	Parameter	Value	Notes
0	snr_range	[1,4,7,10,13,16,19]	dB values seen during training
1	input_channels	3 (RGB)	set =1 for grayscale
2	feature_dim	256	bottleneck channels
3	compression_ratio	0.5	≡ 4×4×dim / H×W×3
4	channel_type	AWGN default	AWGN, Rayleigh, UW sim
5	learning_rate	1e-3	Adam, StepLRx0.5 30 epoch

Compression ratio:

The compression ratio was determined by the following formula:

$$R = \frac{\text{feat\_dim} \times 4 \times 4}{H \times W \times 3} \xrightarrow[96\times 96]{256\times C} 0.036\ (3.6\%)$$

Where c is the hyperparameter compression_ratio (default 0.5). Reconstruction loss was defined as:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{BHW3} \|\hat{\mathbf{x}} - \mathbf{x}\|_F^2$$

Final model implementation:

```
model = DeepJSCC_Adaptive(input_channels=3,
                    feature_dim=256,
                    compression_ratio=0.5,
                    channel_type='AWGN',
                    input_size=(96,96))      # any (H,W) if % 8
out, features, noisy_features = model(image_batch, snr_db=10)
```

Channel model:

The channel was created as a Python class that had different functions for different sound degradation environments, specifically AWGN, Rayleigh, and an underwater channel.

SNR was injected into the channel via the following formula:

$$\tilde{\mathbf{Z}} = \mathbf{Z} + \mathcal{N}(0, \sigma^2); \quad \sigma^2 = 10^{-\mathrm{SNR}_{\mathrm{dB}}/10}$$

Where z hat is the output signal. The channel could be configured for different water environments; however for the experiments in this research it was kept constant with the default parameters:

```
def underwater_channel(self, signal, snr_db, distance_m=100, medium='saltwater', temp_c=10, salinity_ppt=35, depth_m=100):
```

Iterative scaling approach:

For 96x96 images, and below, we can effectively minimise compression_ratio with little consequences on image quality. This is because of the size of the bottleneck in relation to the original image size. The bottleneck consists of a 4x4 spatial grid with channel dimension determined by the product of feature_dim and compression_ratio. With feature_dim set to 256 and compression_ratio = 0.5, the channel dimension becomes: Channels = feature_dim x compression_ratio = 256x0.5=128 channel. This means our total bottleneck size can be calculated via: Bottle neck size = 4*4*128 = 2048 float32 values. Assuming a 32-bit (4-byte) floating-point representation: Bottleneck data size = 2048x4bytes/float = 8,192kB. For 96x96 RGB images the original uncompressed size assuming 24 bit true RGB colour comes to 96x96x3x1byte/channel = 27,648 bytes (27Kb) . Therefore the compression ratio achieved is 27,648 / 8,192 = 3.375:1. This means data is compressed to roughly 29.6% of the original image size. While the 3.375:1 compression ratio is substantially lower than the theoretical maximum achievable with minimal feature dimensions as compression_ratio approaches 0, it represents a practical balance between compression efficiency and reconstruction quality for early testing. The retained 128-channel semantic representation provides sufficient capacity for the decoder to reconstruct perceptually acceptable images across a varying range of SNR conditions. Note, compression_ratio was scaled to 0.05 to demonstrate state of the art compression ratios, achieving a ratio of 33.2:1.

Data collection and experimental setup:

A range of metrics were used to assess image quality. These were conducted between the reconstructed image and the original image. These included:

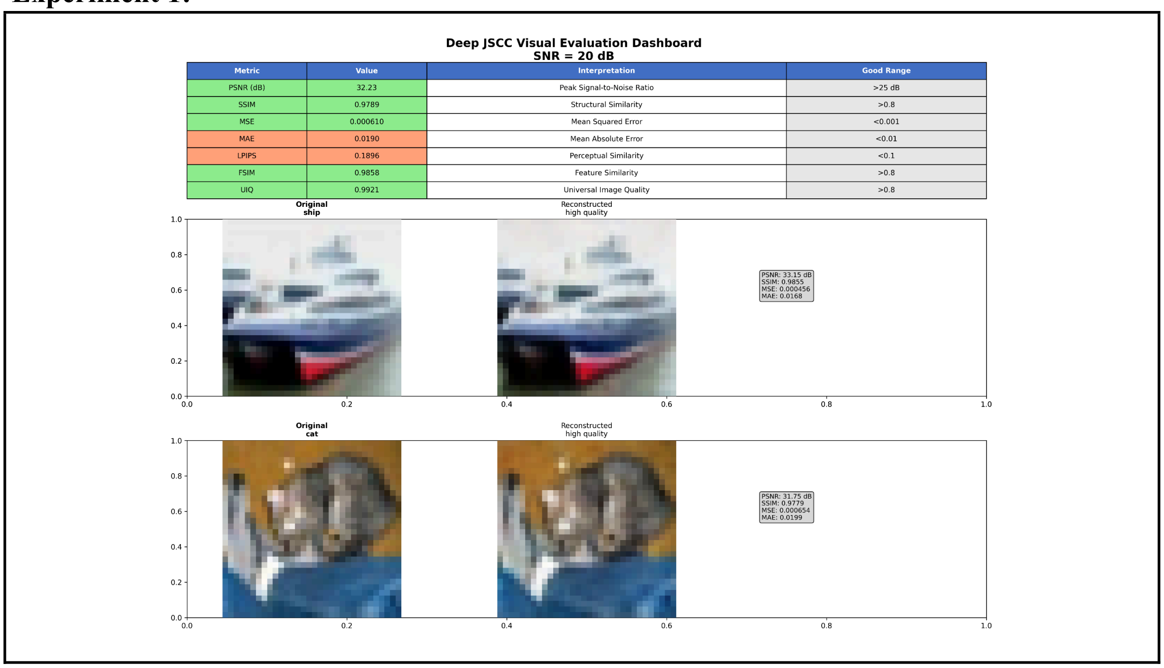
Metric	Range	Interpretation	Equation	Notes
PSNR: Peak Signal to Noise Ratio	0 - infinity	Higher score == better (dB)	$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{\mathrm{MSE}}\right)$	
SSIM: Structural Similarity Index Measure	(-1) - 1	Greater value == better (1 = identical to original image)	$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$	
MSE: Mean Square Error	0 - infinity	Smaller score == better	$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2$	
MAE: Mean Absolute Error	0 - infinity	Smaller score == better	$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\lvert x_i - y_i\rvert$	
LPIPS (approximated): Learned Perceptual Image Patch Similarity	0 - infinity	Smaller score == better (perceptual)	$\mathcal{L}_{\mathrm{lpips}} \approx \lvert\nabla I - \nabla \hat{I}\rvert_1$	Approximated with a simple gradient-domain distance
FSIM: Feature similarity Index Measure	0 - infinity	Greater score == better	$\mathrm{FSIM} = \frac{2\sum PC_1 \cdot PC_2}{\sum PC_1^2 + \sum PC_2^2}$	
UIQ: Universal Image Quality Index	(-1) - infinity	Greater UIQ == better (universal)	$\mathrm{UIQ} = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)(\bar{x}^2 + \bar{y}^2)}$	

Results and experiments

All training and evaluation was done on an Apple M3 GPU with mbs as the backend.

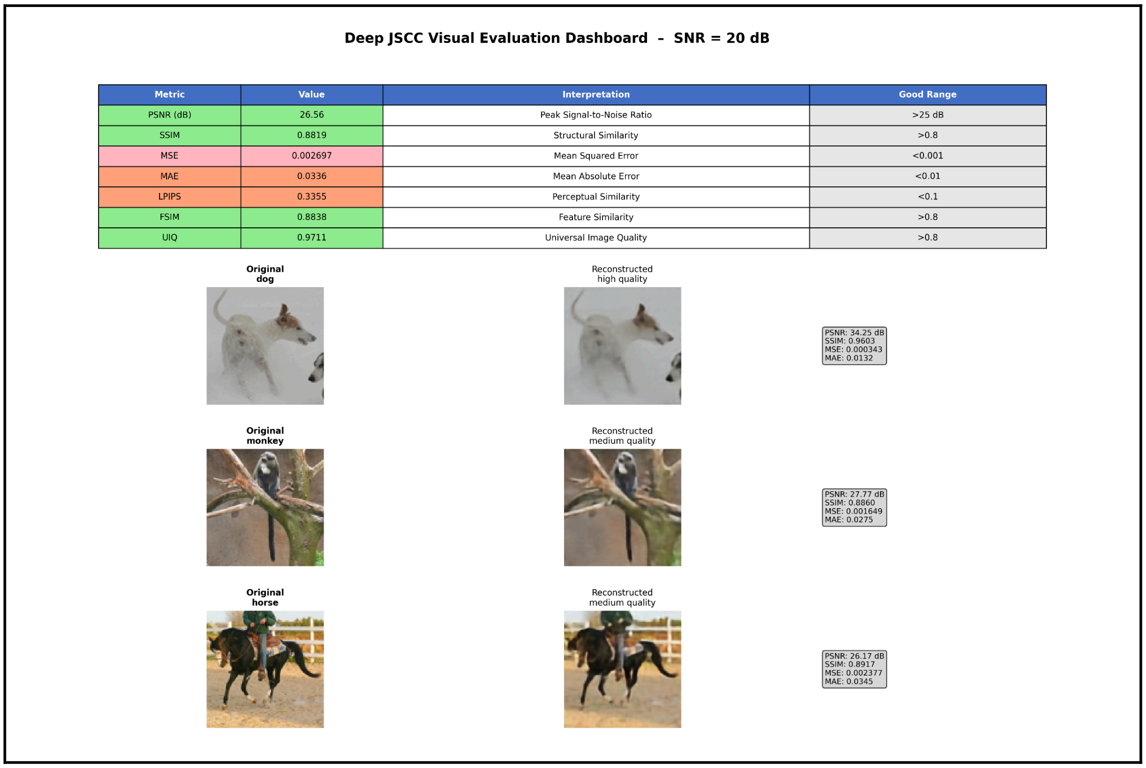
Experiment	Description
1	Evaluates performance at a constant SNR value of 20dB, at an output resolution of 32x32. The model was trained for 100 epochs with early stopping and default parameters for model and channel (AWGN) on CIFAR-10 dataset. Evaluation is done on test dataset from CIFAR-10.
2	Evaluates performance at a constant SNR value of 20dB, at an input/output resolution of 96x96. The model was trained for 100 epochs with early stopping and default parameters for model and channel (AWGN) on STL-10 dataset. Evaluation is done on a test dataset from STL-10.
3	Stress testing completely unseen handpicked images on finalised 96x96 adaptive JSCC model. A pipeline was setup to convert the image to a 96x96 image, and then convert it into the appropriate tensor object for processing by the model. Three images were hand selected: One of a fish that had clear features that could easily be extracted, a more complex coral reef picture with a different hughe than the training data, and finally a picture of University of Canberra. These were stress tested over the default SNR range.

Experiment 1:



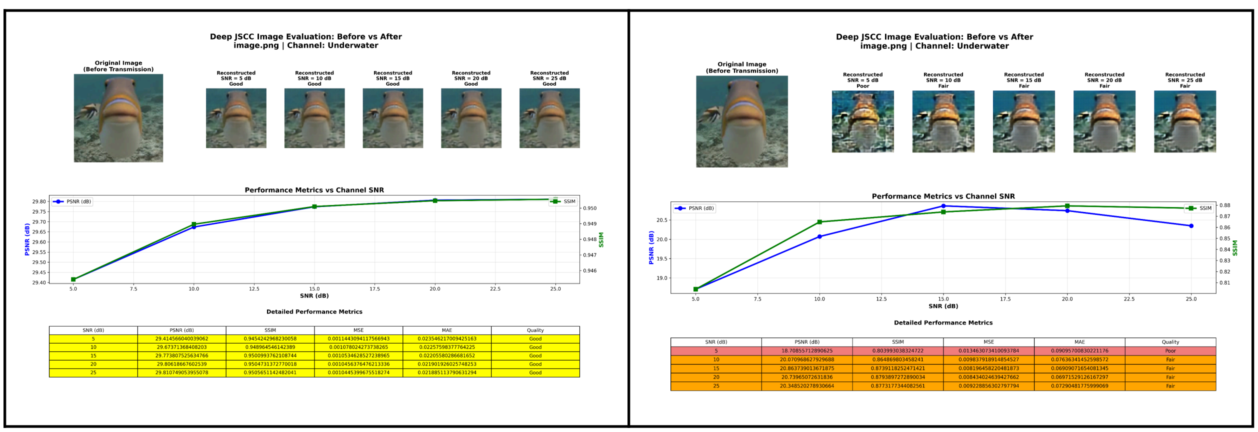
[above: figure 1, validating the adaptive JSCC model on 32x32 images from the test set of the CIFAR-10 dataset]

Experiment 2:

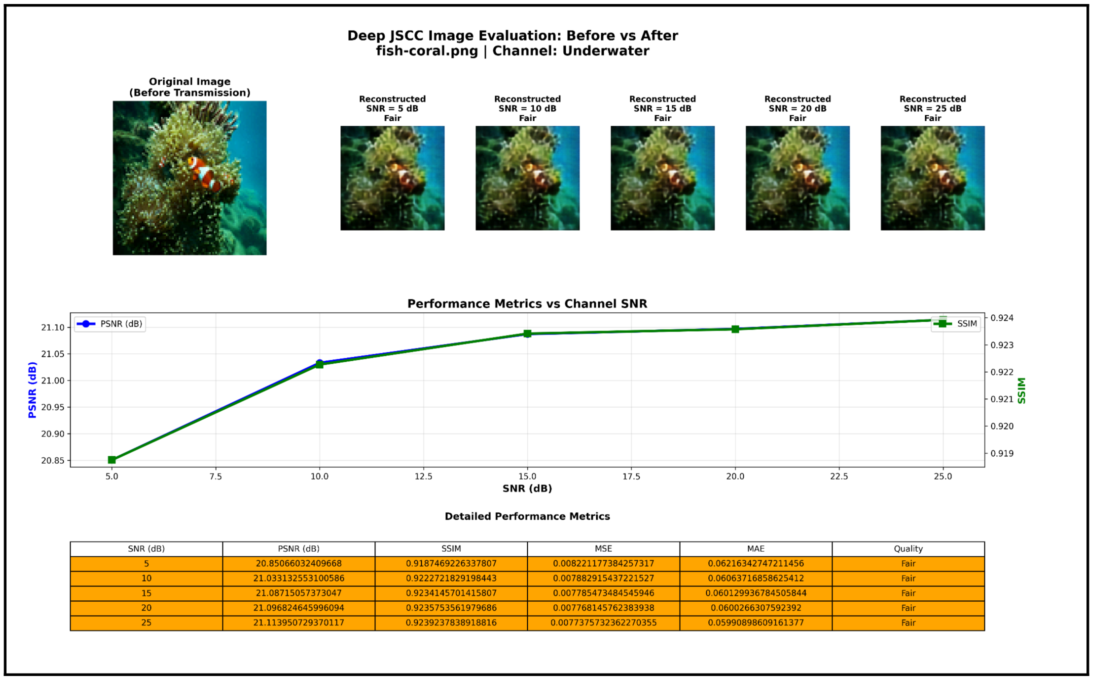


[above: figure 2, validating the adaptive JSCC model on 96x96 images from the test set of the STL-10 dataset]

Experiment 3:

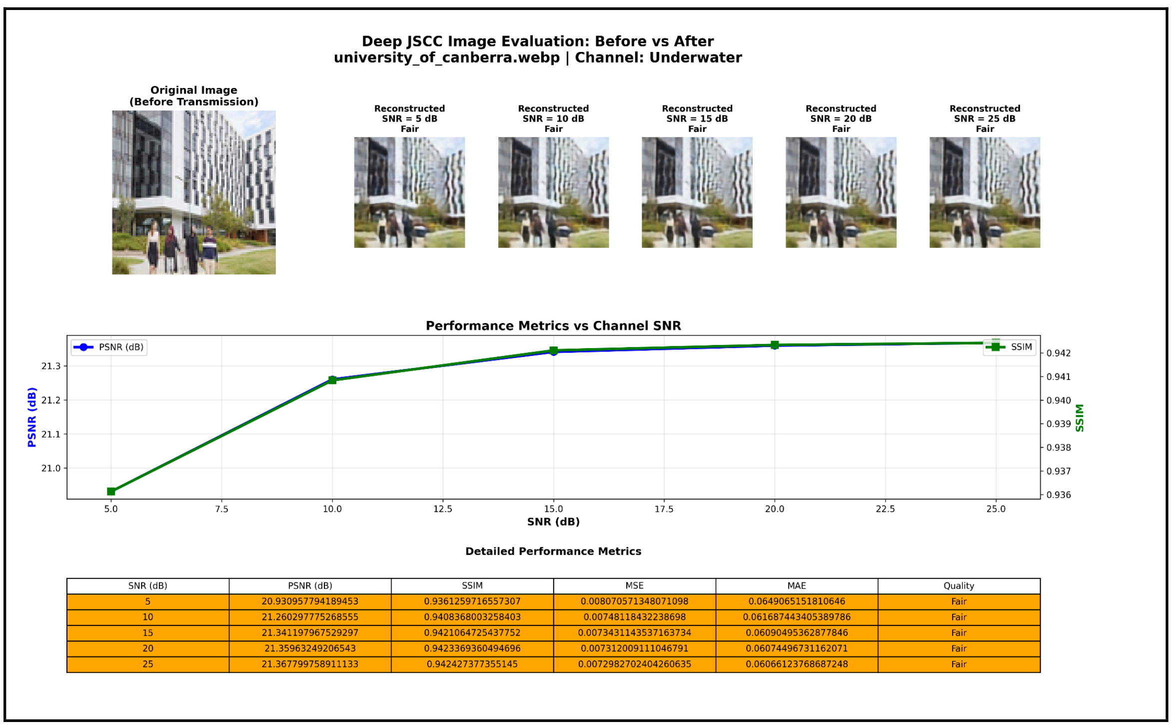


[above: figure 3, testing the network stack through underwater channel (0.5 compression ratio left, 0.05 compression ratio right]



SNR (dB)	PSNR (dB)	SSIM	MSE	MAE	Quality
5	20.85066032409668	0.9187469226337807	0.008221177384257317	0.06216342747211456	Fair
10	21.033132553100586	0.9222721829198443	0.007882915437221527	0.06063716858625412	Fair
15	21.08715057373047	0.9234145701415807	0.007785473484545946	0.060129936784505844	Fair
20	21.096824645996094	0.9235753561979686	0.007768145762383938	0.0600266307592392	Fair
25	21.113950729370117	0.9239237838918816	0.0077375732362270355	0.05990898609161377	Fair

[above: figure 4, testing the model on unseen data of the ocean through underwater channel]



SNR (dB)	PSNR (dB)	SSIM	MSE	MAE	Quality
5	20.930957794189453	0.9361259716557307	0.008070571348071098	0.0649065151810646	Fair
10	21.260297775268555	0.9408368003258403	0.00748118432238698	0.061687443405389786	Fair
15	21.341197967529297	0.9421064725437752	0.0073431143537163734	0.06090495362877846	Fair
20	21.35963249206543	0.9423369360494696	0.007312009111046791	0.06074496731162071	Fair
25	21.367799758911133	0.942427377355145	0.0072982702404260635	0.06066123768687248	Fair

[above: figure 5, testing the model on unseen data using an image of the University of Canberra]

Experiment 1 validated the models performance at 32x32 resolutions, showing that it recreates images to a high standard at a constant 20dB SNR level. Experiment 2 worked on a model trained at a 96x96 resolution and validated the model was able to reconstruct images to a high and medium standard at a constant 20dB SNR level. The results from experiment 3 were interesting. As predicted, the model performed better on the unseen data of a fish, achieving a PSNR of 29.41 and an SSIM score of 0.95 at 5 SNR. This demonstrates the extreme error handling possible via the deep JSCC model. For this another model trained to compress images via a ratio of 0.05 was also demonstrated to work however the image quality degraded, showing poor results for 5dB SNR and fair for all SNR levels above. However the trade of this is achieving a 97% reduction in data, which greatly outperforms traditional compression methods. For the image of the coral reef (figure 4), performance degraded, even at a high SNR level of 25dB the image achieved a PSNR score 21.11dB and an SSIM score of 0.91. This is suspected to be because of the green/blue hues that were unseen in the testing datasets. Visual inspection of the images shows a general capacity to recreate the image, however there is a noticeable quality loss to the human eye. It was noted that on a trial test of unseen hand-picked data of University of Canberra Campus, the model's performance severely degraded. This was hypothesized to be because of the complex patterns on the building, which the model failed to extract relevant features from. The score between the reconstructed image at SNR 25dB was 21.36dB SNR, and a maximum SSIM score of 0.94. The deep JSCC system achieved a compression ratio of 3.375:1 for 96x96 images compressing data 29.6% of original size while keeping a reasonable image quality. Whilst this raw compression ratio is less than traditional methods like JPEG which typically achieved 10-20:1, it was demonstrated that a 33.2:1 compression ratio was possible in experiment 3.

Conclusion and future works:

Further hyperparameter tuning and training is needed to improve model performance. The phenomena of abstract patterns degrading model performance observed in figure 5 alludes to the question “How can we integrate encryption into deep JSCC?” Encryption will effectively randomize pixel values entirely meaning there are no features to extract, causing the ai model to fail. It is hypothesized that for real world testing, you would develop a hybrid network stack, where a traditional header encapsulates the information the deep JSCC artifact produces, and thus, unwraps it before it is decoded by the model on the receiving end would solve this issue. It was also noted that work on drone swarms remains undocumented, and although a SSIM score of 94% still allows for a very rough copy of an image to be sent between two nodes, it is hypothesized this would severely degrade if data was relayed through a swarm of underwater drones. This research would benefit the Department of Defence, which has recently invested 1.7 billion dollars into underwater drone fleets [11].

References:

[1] Y. Peng, J. Liu, Y. Zhu, R. Zhang and F. Yuan, "Efficient Semantic Communication for Underwater Images Guided by Physical Priors," in IEEE Signal Processing Letters, vol. 32, pp. 3072-3076, 2025, doi: 10.1109/LSP.2025.3592591. [H-index=167, SJR=0.938, AVG citations/2years=4.63]

[2] F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji and V. C. M. Leung, "Enabling Massive IoT Toward 6G: A Comprehensive Survey," in IEEE Internet of Things Journal, vol. 8, no. 15, pp. 11891-11915, 1 Aug.1, 2021, doi: 10.1109/JIOT.2021.3063686. [H-index=208, SJR=2.483, AVG citations/2years=10.617]

[3] D. Ma, X. Chen, W. Ma, H. Zheng and F. Qu, "Neural Network Model-Based Reinforcement Learning Control for AUV 3-D Path Following," in IEEE Transactions on Intelligent Vehicles, vol. 9, no. 1, pp. 893-904, Jan. 2024, doi: 10.1109/TIV.2023.3282681. [H-index=74, SJR=2.821, AVG citations/2years=14.067]

[4] L. Liu, S. Zhang, L. Zhang, G. Pan and J. Yu, "Multi-UUV Maneuvering Counter-Game for Dynamic Target Scenario Based on Fractional-Order Recurrent Neural Network," in IEEE Transactions on Cybernetics, vol. 53, no. 6, pp. 4015-4028, June 2023, doi: 10.1109/TCYB.2022.3225106. [H-index=184, SJR=3.254, AVG citations/2years=10.815]

[5] M. J. Er, H. Gong, Y. Liu and T. Liu, "Intelligent Trajectory Tracking and Formation Control of Underactuated Autonomous Underwater Vehicles: A Critical Review," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 54, no. 1, pp. 543-555, Jan. 2024, doi: 10.1109/TSMC.2023.3312268. [H-index=184, SJR=3.254, AVG citations/2years=10.815]

[6] K. G. Omeke, M. Mollel, S. T. Shah, L. Zhang, Q. H. Abbasi and M. A. Imran, "Toward a Sustainable Internet of Underwater Things Based on AUVs, SWIPT, and Reinforcement Learning," in IEEE Internet of Things Journal, vol. 11, no. 5, pp. 7640-7651, 1 March1, 2024, doi: 10.1109/JIOT.2023.3319250. [H-index=208, SJR=2.483, AVG citations/2years=10.617]

[7] L. Hu, L. Yu and Z. Qin, "Deep Learning-Based Semantic Communication System for Wireless Image Transmission," in IEEE Wireless Communications Letters, vol. 14, no. 8, pp. 2391-2395, Aug. 2025, doi: 10.1109/LWC.2025.3570499. [H-index=118, SJR=2.240, AVG citations/2years=6.678]
[8] Z. Xu et al., "Integrated Object Detection and Communication for Synthetic Aperture Radar Images," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 294-307, 2025, doi: 10.1109/JSTARS.2024.3495023. [H-index=139, SJR=1.349, AVG citations/2years=6.158]
[9] J. Wu, C. Song, J. Ma, J. Wu and G. Han, "Reinforcement Learning and Particle Swarm Optimization Supporting Real-Time Rescue Assignments for Multiple Autonomous Underwater Vehicles," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 7, pp. 6807-6820, July 2022, doi: 10.1109/TITS.2021.3062500. [H-index=224, SJR=2.589, AVG citations/2years=11.531]
[10] J. Cui, G. Han, Y. Su and X. Fu, "Non-Uniform Non-Orthogonal Multicarrier Underwater Communication for Compressed Sonar Image Data Transmission," in IEEE Transactions on Vehicular Technology, vol. 70, no. 10, pp. 10133-10145, Oct. 2021, doi: 10.1109/TVT.2021.3103551. [H-index=236, SJR=2.156, AVG citations/2years=8.244]
[11] O. Caisley, “Plans for $1.7 billion underwater ‘Ghost Shark’ drone fleet unveiled,” *Abc.net.au*, Sep. 10, 2025. https://www.abc.net.au/news/2025-09-10/ghost-shark-underwater-drone-fleet-plan/105758720
[12] Aaron Hassan Robinson , “Semantic Communication versus Traditional Machine Learning for Underwater UAV Network Optimization: A Systematic Comparison of Performance and Power Efficiency” Course Rep. Engineering Project (Part B), University of Canberra, 2025.