Autoencoder Based Image Quality Metric for Modelling Semantic Noise in Semantic Communications

P. Samarathunga, T. Fernando, V. Gowrisetty, T. Atulugama and A. Fernando

Semantic communication has attracted significant attention as a key technology for emerging 6G communications. Though it has lots of potentials specially for high volume media communications, still there is no proper quality metric for modelling the semantic noise in semantic communications. This paper proposes an autoencoder based image quality metric to quantify the semantic noise. An autoencoder is initially trained with the reference image to generate the encoder decoder model and calculate its latent vector space. Once it is trained, a semantically generated/received image is inserted to the same autoencoder to create the corresponding latent vector space. Finally, both vector spaces are used to define the Euclidean space between two spaces to calculate the Mean Square Error between two vector spaces, which is used to measure the effectiveness of the semantically generated image. Results indicate that the proposed model has a correlation coefficient of 88% with the subjective quality assessment. Furthermore, the proposed model is tested as a metric to evaluate the image quality in conventional image coding. Results indicate that the proposed model can also be used to replace conventional image quality metrics such as PSNR,SSIM,MSSIM,UQI, VIFP, and SSC whereas these conventional metrics completely failed in semantic noise modelling.

Introduction: Semantic communication has led to renewed interest in solving "the semantic problem" in communication systems rather than further pursuing the limits of solutions to "the technical problem", giving rise to the concept of semantic communications, which has now become an active field of research [1],[2]. The main idea behind semantic communications is that, enabled by shared prior knowledge, a machine can identify the meaning of a message based on a semantic representation of the original message. An allegory for this concept in human terms would be one human being able to reconstruct a vivid picture in their mind of an event that they cannot see but can hear about from a radio broadcast or can read about from a printed book. Though, current semantic research is mainly limited to text and speech transmissions, there are some preliminary image transmissions on semantic communications through error-prone channels [3] is presented in the recent past. However, there is no any objective quality metric available for modelling semantic noise in the semantic channel for image transmission applications. This paper proposes an autoencoder based objective semantic quality evaluation model for quantifying the semantic noise in a semantic image transmission system.

Related Work: Due to the advancement of Machine Learning (ML) and the exponential growth of media applications, it is expected that semantic communication will become the centrepiece of designing end-to-end media communication systems, mainly for Machine-to-Machine communications and 6G [4],[5],[6]. Semantic communication considers integrating the meaning of the data into various tasks related to processing and transmitting data, which represents a major change from the traditional Shannon paradigm [2]. Semantic communication is mainly supported by ML and Artificial Intelligence, more specifically deep learning techniques, which allow machines to comprehend information and extract the semantic, or meaning of the information, mimicking the functionality of the human brain. While some initial semantic communication research on text, audio and image transmission has been reported, there is no any model available for quantifying the semantic noise which is the main criteria for determining the success of the semantic communication system.
There are a few existing semantic quality metrics available for text and speech transmission including semantic obviousness, semantic similarity measurement based on knowledge mining, and self-supervised contrastive projection learning [7],[8],[9]. Semantic communication system evaluation uses a semantic similarity measure [10] that combines semantic accuracy and completeness of recovered text. Recently, a perceptual impact of semantic content on image quality is founded on the concept of semantic obviousness [7]. This method extracts two types of features: one for capturing local image characteristics and another for measuring semantic obviousness. Self-supervised contrastive projection learning is a key concept proposed by researchers to evaluate the semantic similarity in single-particle diffraction images [8]. Dimensionality reduction is one such strategy, which results in embeddings with semantic meaning that is consistent with physical intuition. Additionally, researchers have extended the knowledge in Artificial Neural Networks (ANN) to assessing semantic similarity. This research introduces a feature-based approach that leverages artificial ANNs to simulate the human similarity ranking process [9]. However, none of these methods can be used for semantically generated images, since the concepts of semantics used in the papers and semantic communication have a significant gap. In response, this paper proposes first such model which can quantify the level of semantic noise in a semantically generated image, which is a crucial factor in evaluating the effectiveness of image based semantic communication system.

Proposed Framework: Figure 1 illustrates the proposed framework for estimating the semantic noise of the semantic communication system. As shown in figure 1, the proposed autoencoder (presents in Figure 2) is trained with the original or reference (undistorted) image and its latent vector (V_o) is generated. Once it's trained, any semantically generated image or quantised image is considered as the input to the same autoencoder, and the new latent vector (V_1) is derived. Finally, the Euclidean space between the two vectors (V_o and V_1) is considered to generate the Mean Square Error between the vector space as presented in Equation (1),

$$AEQM = \left(\frac{\sum_{i=1}^{N}(V_{io}-V_{i1})^2}{N}\right) \tag{1}$$

where V_{io}, V_{i1} *and* N are latent vector of the original image, latent vector of the distorted image and the size of the latent vector space respectively.

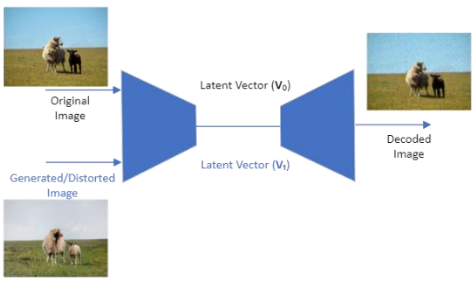

Fig. 1 Autoencoder-based semantic noise model framework.

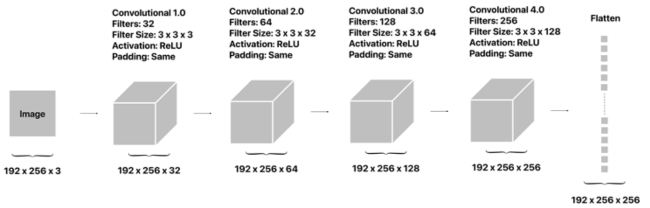

Fig.1*a* Encoder

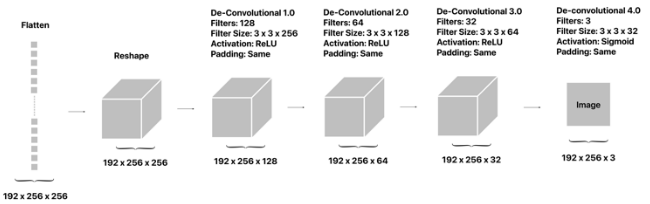

Fig.1*b* Decoder

Fig. 2 Proposed autoencoder architecture used in the proposed quality model.

Figure 2 presents the autoencoder implemented in the proposed quality metric introduced in Figure 1. The input layer has the form of (192 x 256 x 3) pixels and convolutional and de-convolutional layers are used to define the architecture of the autoencoder. Four convolutional layers with a rectified linear unit (ReLU) activation function are formed as encoder layers. The decoder layers include four de-convolutional layers with a ReLU activation function, followed by a final convolutional layer with a sigmoid activation function. The autoencoder model is then defined as a sequential model by combining the encoder and decoder layers. Finally, the autoencoder model is compiled with the Adam optimizer with a learning rate of 0.001 and a binary cross-entropy loss function before training the model. In order to optimize the performance and the complexity, the above hyperparameters are selected based on a series of experiments. Since the proposed codec uses four convolutional layers, it has the capability of capturing the image features accurately, which is considered as the foundation of the proposed quality metric. Since semantic communication uses the semantic meaning rather than the pixel level distortions in an image, the proposed model has the capability of evaluating the semantic noise accurately.

Though the above model assumes an image size of 192x256, it should be noted that this can be extended to any image size, since it's independent of the spatial resolution of the image.

Results: The proposed Auto Encoder Quality Model (AEQM) is tested with 11 different image categories (Spatial Index ranges from low to high) to find out how it is performed against existing most popular image quality metrics (Peak Signal-to-Noise Ratio (PSNR), Universal Quality Image Index (UQI), Visual Information Fidelity (VIFP), Structural Similarity Index (SSIM), Spatial Correlation Coefficient (SCC), Multi-scale Structural Similarity Index (MSSIM)). Table 1 illustrates the performance comparisons between the AEQM and the above metrics for 11 different image groups with different quantization artefacts generated from a JPEG codec (Level of quantization of 5% -100% are considered during this experiment). Table 1 also presents the corresponding subjective quality assessments (DSQA-Double Stimulus Quality Assessment) with 50 subjects. Results clearly show that AEQM is highly correlated with the subjective scores, like the standard image quality metrics considered (range of all metrics are provided in Table 2).

Table 1: Performance of AEQM in modelling quantization noise.

Quality Metric	Quantization Level					
	Q5	**Q10**	**Q25**	**Q50**	**Q75**	**Q100**
PSNR	23.074	25.345	28.058	30.020	32.21	39.49
UQI	0.9750	0.9850	0.991	0.994	0.996	0.998
VIFP	0.2025	0.2858	0.397	0.475	0.553	0.824
SSIM	0.6498	0.7436	0.842	0.892	0.927	0.985
SCC	0.1426	0.2361	0.379	0.488	0.583	0.887
MSSIM	0.870	0.9303	0.969	0.983	0.990	0.998
AEQM	0.0001	0.000043	0.000016	0.000007	0.000003	0.0000002
Subject. Score	2.51	2.98	3.46	4.12	4.45	4.95

Table 2: Performance of AEQM in modelling semantic noise.

Quality Metric	Quality Score	Lowest Score	Highest Score	Correlation Coefficient
PSNR	12.799	0 dB	∞ dB	30%
UQI	0.779	0	1	31%
VIFP	0.059	0	1	21%
SSIM	0.345	0	1	22%
SCC	0.029	0	1	6%
MSSIM	0.385	0	1	25%
AEQM	0.001	1	0	88%
Subjective Score	4.891	0	5	N/A

Finally, the performance of AEQM is investigated for semantically generated images in modelling the semantic noise/distortions. The models proposed in [3], [11] are considered in generating semantically communicated images at the receiver. Generative Adversarial Network (GAN) generated images and reference images are used in the model proposed as shown in the Figure 1 and Figure 2 in calculating the AEQM. For the comparison purpose, same images are considered in conventional quality metrics calculations and Table 2 illustrates the performance comparisons. As before, subjective experiments (DSQA) with 50 subjects are conducted in verifying the proposed objective quality metric. Results indicate that AEQM has a very high correlation coefficient of 88% against the subjective scores, while all other conventional metrics performed extremely poor. Conventional image quality metrics are designed for measuring the quantization artifacts of the image rather than the semantics of it, while proposed metric considers both quantization artifacts and semantics of the images. The proposed encoder has the capability of extracting the semantics of the image rather than only the statistics of the image and compare against the original image, which leads to its superior performance. Though AEQM is computationally expensive compared to other metrics considered, it can be considered as an objective quality metric in modelling the semantic noise in semantic communications due to its outstanding performance.

Conclusions: In this paper, an autoencoder based objective quality metric is proposed for modelling semantic noise in semantic communications. The autoencoder is trained using an undistorted image, and its latent vector is compared against the latent vector of the distorted or generated image in the semantic communication system. Vector spaces are used in calculating the Mean Square Error between the two vector spaces and generate a model for quantifying the semantic noise. Results indicate that the proposed AEQM model exhibits a very high correlation (88%) against the subjective quality assessment in quantifying the semantic noise and outperforms traditional image quality metrics by a significant margin. In the future, the proposed model will be further developed in modelling the semantic noise in semantic video communications.

P. Samarathunga, T. Fernando, V. Gowrisetty, T. Atulugama and A. Fernando (Department of Computer and Information Sciences, University of Strathclyde, Glasgow G1 1XQ, UK)
E-mail: Anil.Fernando@strath.ac.uk

References

[1] C. E. Shannon, "A mathematical theory of communication," The Bell System Technical Journal, 1948, vol. 27, no. 3, pp. 379-423.
[2] C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication," University of Illinois Press, 1949, 3rd ed.
[3] M. U. Lokumarambage, V. S. Gowrisetty, H. Rezaei, T. Sivalingam, N. Rajatheva, and A. Fernando, "Wireless End-to-End Image Transmission System Using Semantic Communications," IEEE Access, 2023, vol. 37149-37163. DOI: 10.1109/ACCESS.2023.3266656.
[4] E. C. Strinati and S. Barbarossa, "6G Networks: Beyond Shannon Towards Semantic and Goal-Oriented Communications," arXiv, 2021, Feb. 17. Available: http://arxiv.org/abs/2011.14844. Accessed: Apr. 19, 2023. DOI: https://doi.org/10.48550/arXiv.2011.14844.
[5] P. Zhang, W. Xu, H. Gao, K. Niu, X. Xu, X. Qin, C. Yuan, Z. Qin, H. Zhao, J. Wei, and F. Zhang, "Toward Wisdom-Evolutionary and Primitive-Concise 6G: A New Paradigm of Semantic Communication Networks," Engineering, 2022, vol. 8, pp. 60-73. DOI: https://doi.org/10.1016/j.eng.2021.11.003.
[6] P. Dong, Q. Wu, X. Zhang, and G. Ding, "Edge semantic cognitive intelligence for 6G networks: Novel theoretical models, enabling framework, and typical applications," China Communications, 2022, vol. 19, no. 8, pp. 1-14. DOI: https://doi.org/10.23919/jcc.2022.08.001.
[7] P. Zhang, W. Zhou, L. Wu, and H. Li, "SOM: Semantic obviousness metric for image quality assessment," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, June 07-12, 2015, pp. 2394-2402. DOI: 10.1109/CVPR.2015.7298853.
[8] J. Zimmermann, F. Beguet, D. Guthruf, B. Langbehn, and D. Rupp, "Finding the semantic similarity in single-particle diffraction images using self-supervised contrastive projection learning," npj Computational Materials, 2023,vol. 9, no. 1,pp. 1-9.DOI: 10.1038/s41524-023-00966-0.
[9] W. Li, R. Raskin, and M. F. Goodchild, "Semantic similarity measurement based on knowledge mining: an artificial neural net approach," International Journal of Geographical Information Science, 2012,vol.26,no.8, pp. 1415-1435. DOI: 10.1080/13658816.2011.635595.
[10] Y. Wang, M. Chen, W. Saad, T. Luo, S. Cui, and H. V. Poor, "Performance Optimization for Semantic Communications: An Attention-based Learning Approach," 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 2021, pp. 1-6. DOI: 10.1109/GLOBECOM46510.2021.9685056.
[11] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic Image Synthesis With Spatially-Adaptive Normalization," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach,CA,USA,2019, pp. 2332-2341. DOI: 10.1109/CVPR.2019.00244.