Deep Learning-based Image Quality Assessment Metric for Quantifying Perceptual Distortions in Transmitted Images

Ijaz Ahmad
Dept. of Computer Engineering
Chosun University
Gwangju, 61452, Korea
ahmadijaz@chosun.kr

Seokjoo Shin
Dept. of Computer Engineering
Chosun University
Gwangju, 61452, Korea
sjshin@chosun.ac.kr (Corresponding author)

***Abstract*—An Image Quality Assessment (IQA) metric measures the quality degradation of an image and is used to optimize parameters of an image processing algorithm. The IQA score can be also an important indicator of the target downstream application performance. With the popularity of deep learning (DL)-based applications in resource constrained domains, most of the DL computations are outsourced to avail remotely located resources. The image data transmission for this purpose is susceptible to distortions due to the imperfect communication environment. The existing IQA metrics used to evaluate the quality of these images mainly rely on human judgment and do not account for the perceptual distortions that are responsible for the degradation in DL model performance. To address this issue, we propose a convolutional autoencoder-based IQA metric that compares images in low dimensional feature space and can be used to monitor image degradation occurred during data transmission. The simulation analysis shows that the proposed method introduces at best 0.13% error while on average 8% error compared to the application model accuracy. Importantly, the proposed IQA score coincides with the DL model performance on a downstream task and can be used to optimize the parameters of a communication system.**

Keywords—deep learning, autoencoder, image quality assessment metric

I. Introduction

An image quality assessment (IQA) metric measures the quality degradation of an image. The degradation in an image can be due to image acquisition hardware, image processing algorithm, and/or data transmission [1]. The IQA metrics that are used to quantify an image quality can be classified as algorithms that require a reference image (Full-reference methods) and algorithms that do not require a reference image (No-reference methods). The full-reference methods compare a distorted image with the pristine reference image to measure a quality score. The methods in this category are suitable to judge the quality of images that undergone certain image processing algorithms such as image compression etc. On the other hand, the no-reference methods assess the perceptual quality of an image based on expected statistical properties of an image. The traditional IQA metrics are used to optimize the parameters of a preprocessing algorithm or an image acquisition hardware to improve the image quality that is its appearance, with respect to how it is perceived by the humans. In other words, the goal of the conventional IQA metrics is to compute an objective quality score that agrees with a subjective human quality score. However, with the wide adoptability of DL in various domains, it is necessary to quantify the image quality with respect to how well a DL model performs on a target task [2], [3]. The image quality specific to a task that can provide an indication of the downstream application performance is termed as *task amenability* in [2].

DL models have achieved the state-of-the-art performance for different computer vision tasks such as classification, segmentation etc. For an efficient DL-based solution, it is important that the model training and testing are performed against the data that comes from the same target application distribution. However, when this distribution is altered by distortions such as blur and noise in the images, then the performance of the trained DL model degrades. The source of distortion might be a lossy compression algorithm or imperfect channel conditions during transmission. Lossy compression provides a solution to efficiently utilize the limited available bandwidth. The information loss in a lossy compression algorithm such as the JPEG standard, is carried out in such a way that the degradations are not human perceivable. However, these unperceivable artifacts degrade a trained DL model accuracy. For example, authors in [3]–[5] analyzed the impact of the JPEG compression algorithm on a trained DL model performance and proposed methods to circumvent this issue. An alternative solution to the lossy compression algorithm is to utilize high data rate systems. However, due to the imperfect wireless communication environment and hardware errors, the received image quality degrades significantly. The authors in [6], [7] have studied the impact of these noises on a trained DL model performance in a scenario where the images are exchanged to avail third-party owned computational resources. Their analyses have shown that a DL model accuracy suffers during inference stage, when the images were transmitted with a high data rate wireless communication system.

The aforementioned works either analyze the impact of image distortions on a trained DL model performance or proposed solutions to mitigate the performance gap. However, they do not quantify the perceptual distortions that are responsible for the degradation of a DL model performance. The related work of task amenability in [2] do not consider communication distortions. On the other hand, the existing IQA metrics are either sensitive to noise (full-reference metrics) or to changes in the distribution (no-reference metrics) [8]. Therefore, in the current study, we propose an IQA metric based on convolutional autoencoder to compare the similarity between clean and processed images in low dimensional feature space. Specifically, we have considered image distortions resulted from a communication system. For this purpose, we have implemented an Orthogonal Frequency

Division Multiplexing (OFDM)-based communication system utilizing different modulation schemes. For the downstream application task, we have considered DL-based multiclass classification problem. For the scenario, we have assumed edge-cloud collaboration to enable AI-based IoT applications. The main contributions of this work are: (1) an autoencoder-based IQA metric to quantify an image quality with respect to DL interpretibility and (2) the experiments were conducted on OFDM-based image communication system implemented with different modulation schemes and channel models.

II. System Model

We considered an edge-cloud collaboration to enable AI-based IoT applications as shown in Fig. 1. In general, there are two stages for implementing a DL-based solution. In the first stage, the model learns how to solve a problem on the available data – *training stage*. In the second stage, the trained DL model is deployed to solve real world problems related to the same domain they have previously learned – *inference stage*. The DL algorithms are compute-intensive tasks and their computation demands exceed the capability of resource-constrained edge devices. Therefore, the DL model training is carried out on the data that is available on the cloud server utilizing cloud computational resources. The trained model is then deployed on edge server, which acts as an inference engine to serve various requests. The IoT end devices, which generate the data, are resource constrained devices; therefore, the AI computations are offloaded to the edge server. The devices are connected via a wireless link implemented using an OFDM system to facilitate the high data rate demands. This data transmission is susceptible to various types of distortions such as imperfect channel conditions, hardware noises, etc. For simplicity, we assumed that there is no other source of distortions such as image quality degradation due to lossy compression, except the communication system. Our objective is to measure the perceptual distortions in wirelessly transmitted images in order to estimate the downstream AI algorithm performance.

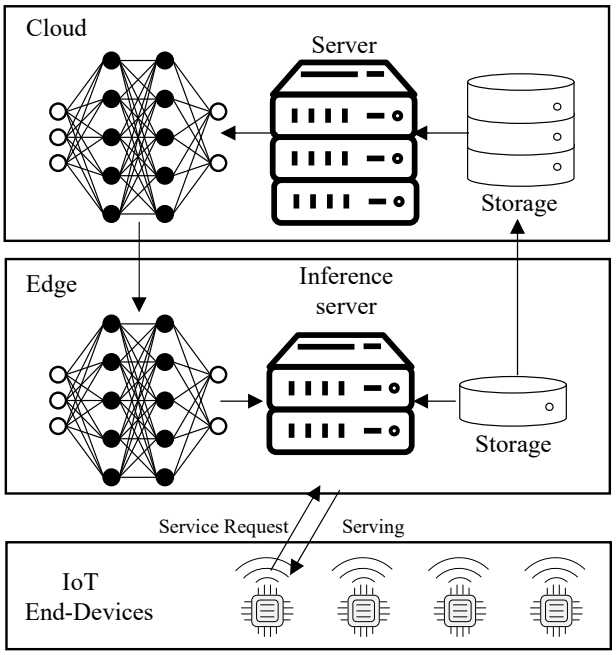

Fig. 1. Edge-cloud collaboration to enable AI-based IoT applications. The cloud server is used to train a model on its locally available data and edge is used as inference engine. The IoT end-devices send the AI computation requests and receive a result from the edge server.

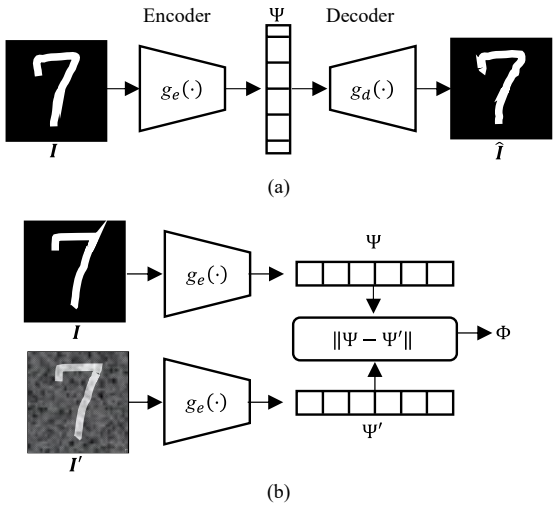

Fig. 2. The proposed convolutional autoencoder-based IQA measure. (a) is the training of the autoencoder and (b) is the feature similarity of clean and noisy images using the trained autoencoder.

III. Proposed Method

The basic principle of the proposed IQA metric is to compare the perceptual similarity of a processed image to the original image in the feature space [9]. For this purpose, a dimensionality reduction function $g(\cdot)$ such as an autoencoder [10] can be used to map an image $\boldsymbol{I} \in \boldsymbol{\mathcal{N}}^{W\times H\times C}$ in the high dimension pixel space to a feature vector in lower dimensional latent space $\Psi \in \boldsymbol{\mathcal{R}}^{w\times h}$. The proposed convolutional autoencoder as shown in Fig. 2. (a), consists of two non-linear functions: an encoder function $g_e(\boldsymbol{I})$ that maps the input image $\boldsymbol{I}$ to a feature vector Ψ, and a decoder function $g_d(\Psi)$ that reconstructs the input image from those features. The goal is to learn $g_d\left(g_e(\boldsymbol{I})\right) = \hat{\boldsymbol{I}} \approx \boldsymbol{I}$ to minimize the reconstruction loss function $\mathcal{L}(\boldsymbol{I}, \hat{\boldsymbol{I}})$, which penalizes $g_d\left(g_e(\boldsymbol{I})\right)$ for being dissimilar from $\boldsymbol{I}$. The proposed autoencoder has four convolutional layers in the encoder and the decoder, and 128 features extracted from the input image. Instead of using pooling layers to reduce the input size, the stride width is set to 2 in the last three convolutional layers. Though this increases the number of trainable parameters, but better accuracy can be achieved [11]. The trained encoder can be used to extract features from the original and processed images to compare their similarity as shown in Fig. 2. (b).

The feature vector of an image in the latent space is defined as $\Psi = [\psi_1, \psi_2, \ldots, \psi_{128}]$ and the feature vectors of N images of label l can be considered as a set of feature vectors $\bigcup_{n=1}^{N}\{\Psi_{l,n}\}$. Instead of comparing the feature vectors, a representative centroid of the feature vectors set can be defined as in [9], [12]

$$\gamma_l = \frac{1}{N}\sum_{n=1}^{N} \Psi_{l,n}\,. \tag{1}$$

The Euclidean distance $\delta(\gamma_l^o, \gamma_l^p)$ between centroids of original image features γ_l^o and processed image features γ_l^p for a label l is given by

$$\delta(\gamma_l^o, \gamma_l^p) = \sqrt{\sum_{t=1}^{128} (\gamma_{l,t}^o - \gamma_{l,t}^p)^2}. \quad (2)$$

For a specific label $\acute{l}$, the distance between $\gamma_{\acute{l}}^o$ and $\gamma_{\acute{l}}^p$ and the rest of the labels l can represent how similar the features are. The proposed IQA score (Φ) can be obtained as:

$$\Phi = 1 - f\left(\delta(\gamma_{\acute{l}}^o, \gamma_{\acute{l}}^p)\right), \quad (3)$$

where, $f(x)$ is a normalization function that maps x into the range of $[0,1]$, and Φ=1 represents the highest similarity between the two feature vectors.

IV. Results and Discussion

In this section, we implement our system model described in Section 2 and compare the transmitted image quality with respect to different quality assessment metrics. In the experiments, we have used the MNIST handwritten digit dataset [13], which consists of 60K training and 10K test images distributed between 10 classes. The images are grayscale with 28×28 dimensions. In order to avoid long runs of zeros because of the uniform black background, we have replaced the pixel values with a random value of 20. For the metrics, their score is calculated as a mean across the whole test set. In addition, our simulation parameters used for the communication system design are summarized in Table 1.

TABLE I. Summary of the simulation parameters used for the image communication system.

Parameters	Values
Modulation Techniques	BPSK, QPSK, 16-QAM, 64-QAM
Transform	FFT
Channel Model	AWGN, Rayleigh Fading
Channel Estimation	Least Square Error
IFFT/FFT Length	64
Cyclic Prefix Length	16 samples

A. Visual Quality Analysis

For visual inspection of images transmitted over different communication systems, we have considered QPSK and 64-QAM with SNR $\in \{0, 15\}$ and both channel models as examples. The standard Peppers image was transmitted over these systems and the results are shown in Fig. 3. The figure shows the transmitted and received constellation points, the received images, and histograms of the original and reconstructed images. It can be seen that the image has better appearance for QPSK than 64-QAM as the distance between the constellation points is larger and the error margin is lower. When using 64-QAM with the fading channel then the recovered image is unrecognizable. The main reason for

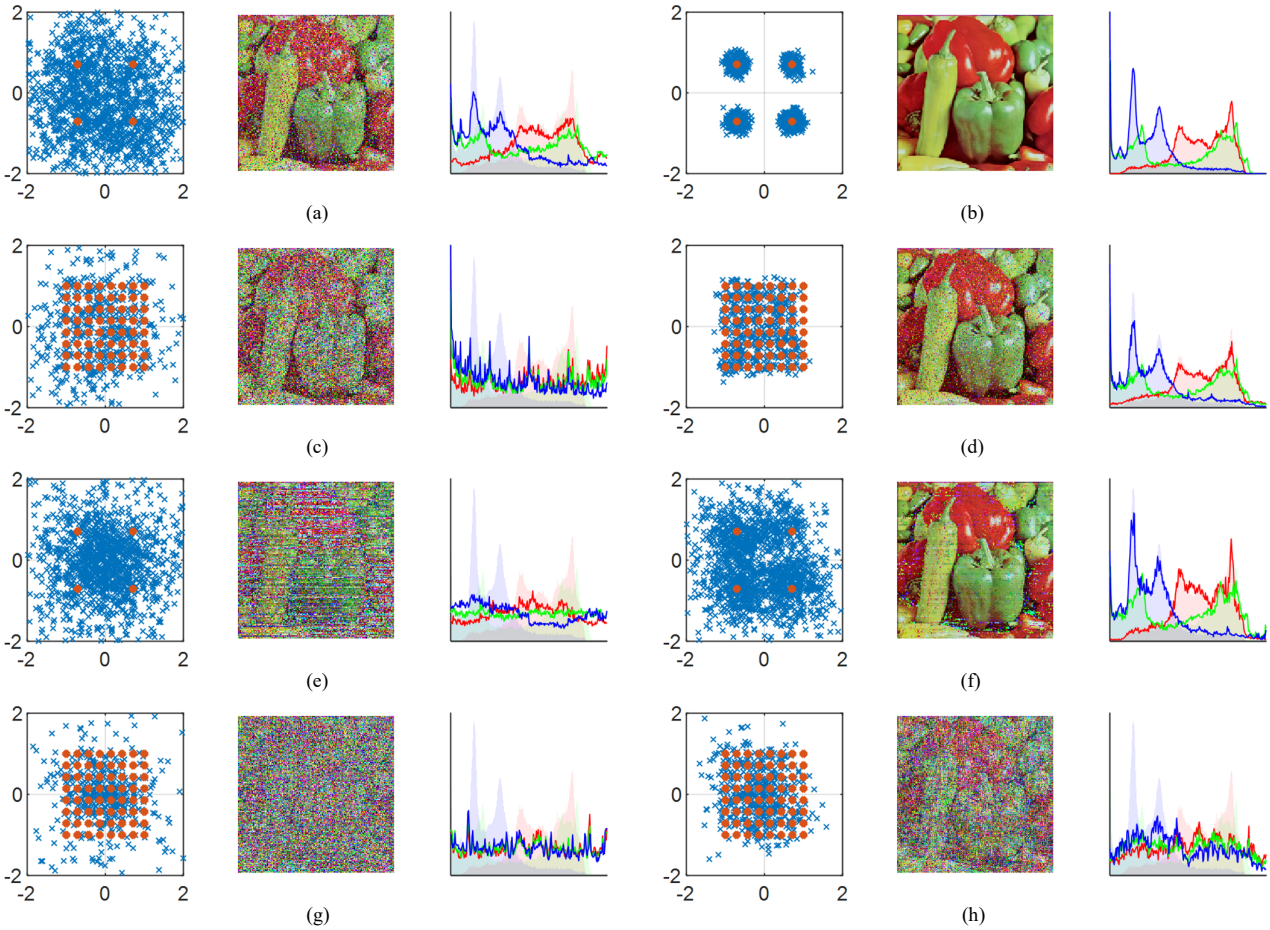

Fig. 3. Visual analysis of the images transmitted over an OFDM-based image communication system. Images were transmitted over AWGN channel in (a) – (d) and over Rayleigh channel in (e) – (f). The SNR = 0 dB in the first column and SNR = 15 dB in the second column. The modulation scheme used was QPSK in (a)(b)(e) and (f) and 64-QAM in (c)(d)(g) and (h). In the constellation diagrams, orange is the transmitted and blue is the received constellation points. The histograms for the original and received images are shown with filled area and line graph, respectively, where the color shows the histogram of its corresponding color channel in the image.

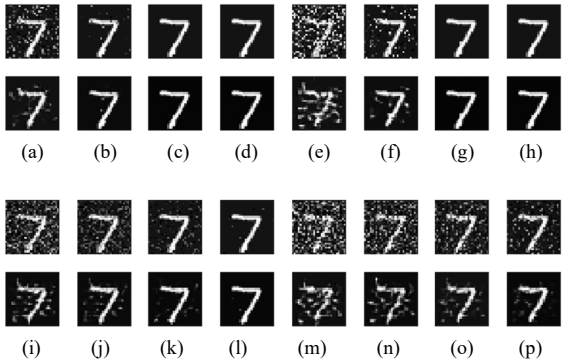

Fig. 4. Example image from the MNIST dataset transmitted over AWGN channel utilizing different modulation schemes. The proposed autoencoder reconstructed images are shown below each image. The modulation used in (a) – (d) is BPSK, (e) – (h) is QPSK, (i) – (l) is 16-QAM and (m) – (p) is 64-QAM. For each modulation SNR $\in \{0, 5, 10, 15\}$.

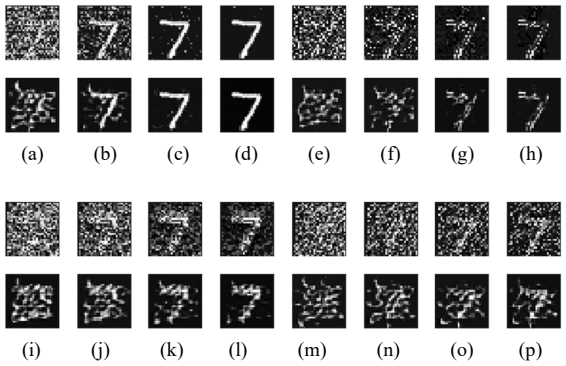

Fig. 5. Example image from the MNIST dataset transmitted over Rayleigh fading channel utilizing different modulation schemes. The proposed autoencoder reconstructed images are shown below each image. The modulation used in (a) – (d) is BPSK, (e) – (h) is QPSK, (i) – (l) is 16-QAM and (m) – (p) is 64-QAM. For each modulation SNR $\in \{0, 5, 10, 15\}$.

providing the histograms is to analyze the distorted image properties. An image histogram plots the intensity distribution as the number of pixels at each intensity level. For a natural image, the histogram is a skewed distribution concentrated at one location as shown in Fig. 3. with filled area graph for each color component. On the other hand, for noisy images the intensity distribution becomes smoother for example, as shown in Fig. 3. (c) and (e) with the line graph. In the case of Fig. 3. (a) (d) and (f), the received images have visible distortions; however, the distribution is following almost the same natural intensity distribution as that of the original image. On the other hand, for 64-QAM and Rayleigh fading channel as shown in Fig. 3. (g) and (h), the reconstructed images are noisy and have uniform distribution.

In addition, an example image from the MNSIT dataset is transmitted over all communication systems given in Table 1 and the results are shown in Fig. 4. for AWGN channel and Fig. 5. for Rayleigh fading channel. These images were used to evaluate the performance of the proposed autoencoder. In Fig. 4. and 5., each image is accompanied with the reconstructed image obtained from the proposed autoencoder. In Fig. 4. all reconstructed images are recognizable, while in Fig. 5. the learned representation from the clean images is not adequate to reconstruct images from the 16-QAM and 64-QAM transmitted images. It is worth mentioning that our objective was not to train the autoencoder for image denoising task, but to extract features from the images that are representative of the downstream task.

B. Image Quality Analysis

Bit Error Ratio (BER) is used to measure quality of the reconstructed signal on the receiver side. It is a ratio between the number of bit errors and the total number of bits. In our simulations, the BER of a system is calculated as a mean value across the whole test dataset and is given in Table 2. It can be observed that with higher SNR the BER decreases and increase in the number of bits per symbol increases the BER.

The BER measures the signal quality regardless of the signal contents. In other words, it does not measure the perceptual quality of the reconstructed images. Therefore, to get a better understanding of the image quality, we have considered Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [14] quality metrics. The PSNR measured in dB, is derived from the mean square error metric, and gives the ratio of the maximum intensity value in

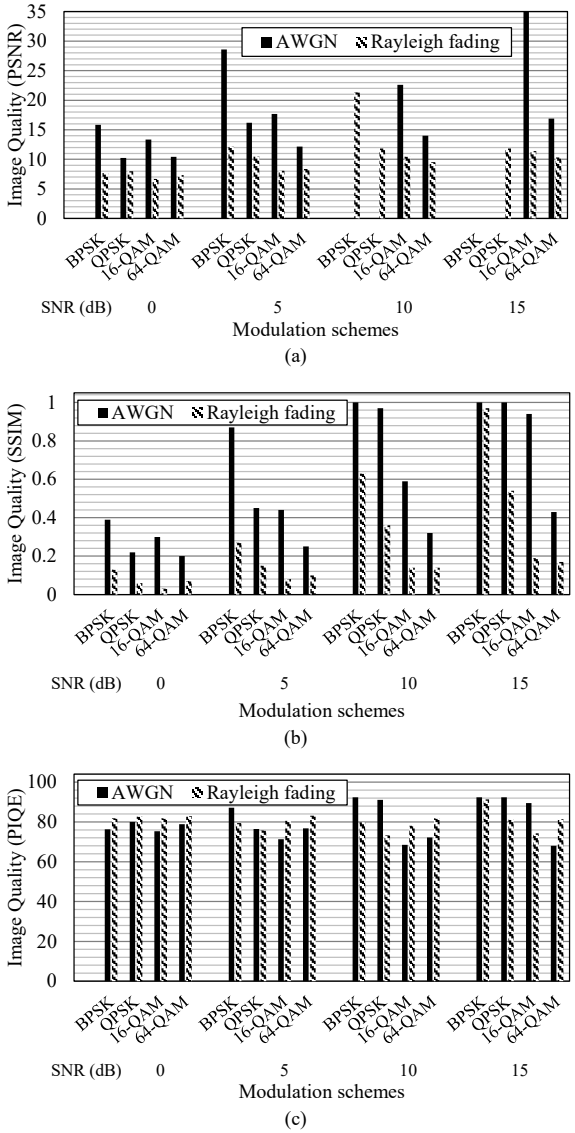

Fig. 6. Recovered image quality analysis with respect to different quality evaluation metrics. (a) is PSNR, (b) is SSIM and (c) is PIQE measure. The missing values in (a) for some modulation schemes are when PSNR=∞.

TABLE II. DL model performance correlation analysis with the image quality metrics.

Methods		SNR (dB)	BER	100 – Acc (%)	PSNR		SSIM [14]		PIQE [15]		IQA (%)
Channel Model	Modulation Scheme				*wp*	*cp*	*wp*	*cp*	*wp*	*cp*	
AWGN	BPSK	0	0.0783	1.6	15.84	15.9	2.15	2.44	76.3	73.49	9.48
		5	0.0059	0.73	28.59	28.49	8.86	8.86	87.15	87.85	1.44
		10	0	0.68	∞	∞	90	90	92.35	92.76	0
		15	0	0.68	∞	∞	90	90	92.35	92.77	0
	QPSK	0	0.2124	26.96	10.25	10.34	1.08	1.31	80.02	79.46	33.04
		5	0.055	1.93	16.2	16.29	2.6	2.84	76.45	75.66	6.3
		10	0.0012	0.68	∞	∞	15.23	15.23	91.09	91.5	0.52
		15	0	0.68	∞	∞	90	90	92.35	92.77	0
	16-QAM	0	0.3114	3.64	13.35	13.4	1.55	1.8	75.3	73.62	18.99
		5	0.2049	0.93	17.67	17.72	2.52	2.68	71.29	67.73	8.42
		10	0.0684	0.75	22.62	22.71	3.87	4.09	68.51	67.55	4.24
		15	0.0038	0.67	35.01	35.04	12.22	12.22	89.47	89.19	0.8
	64-QAM	0	0.3608	32.6	10.43	10.44	0.97	1.14	78.85	78.64	39.03
		5	0.3114	11.69	12.16	12.15	1.25	1.49	76.76	75.62	26.13
		10	0.2232	3.79	14	14.02	1.67	1.87	72.24	71.58	14.77
		15	0.0994	1.46	16.89	16.85	2.44	2.68	68	67.72	7.6
Rayleigh Fading	BPSK	0	0.2008	72.41	7.7	7.77	0.6	0.71	81.87	81.65	68.08
		5	0.0582	7.6	12.13	12.32	1.37	1.74	79.4	78.05	23.97
		10	0.0083	0.93	21.33	21.82	4.32	4.56	79.87	80.85	3.63
		15	0.0016	0.75	∞	∞	15.23	16.99	91.36	91.76	0.97
	QPSK	0	0.3539	83.35	8	8.02	0.27	0.32	82.64	82.59	89.31
		5	0.2276	56.71	10.53	10.52	0.71	0.76	75.7	75.79	68.54
		10	0.1376	27.58	11.85	12	1.94	2.01	73.23	72.66	46.96
		15	0.0966	26.6	11.85	11.98	3.37	3.37	81.19	80.48	48.25
	16-QAM	0	0.4302	88.66	6.71	6.72	0.13	0.18	81.85	82.02	91.68
		5	0.3537	80.43	8.08	8.09	0.36	0.41	80.52	80.4	83.51
		10	0.2892	54.61	10.45	10.55	0.66	0.76	78.05	77.73	71.82
		15	0.2547	35.38	11.41	11.75	0.92	1.02	74.26	74.39	57.81
	64-QAM	0	0.4428	84.87	7.33	7.34	0.32	0.32	82.99	82.95	89.03
		5	0.4138	77.27	8.37	8.4	0.46	0.51	83.15	83.03	82.83
		10	0.3988	61.29	9.52	9.54	0.66	0.76	81.87	81.57	68.95
		15	0.3979	45.07	10.31	10.35	0.81	0.97	81.37	80.84	57.12

100 – Acc: the percentage error in Accuracy, ***wp***: wrong predictions, ***cp***: correct prediction.

the image to the power of the distortion. Higher the PSNR value, similar the two images are and when two images are same then PSNR has an infinity value. Fig. 6. (a) compares the PSNR values of distorted images obtained from different communications systems. Similar to BER, the PSNR increases with the SNR values. Though the PSNR has an advantage of being simple to calculate, its score does not agree with the subjective quality score. Therefore, we considered SSIM, which takes the human perception of structure into account. It combines local image structure, luminance and contrast into a single metric score. The score ranges from 0 to 1, where larger value means high quality image. The calculated SSIM for the transmitted images is shown in Fig. 6. (b). The SSIM follows the same trend as BER and PSNR.

The metrics we have considered so far are full-reference image quality metrics, which means that they require a reference original image without any distortion to compute an objective score. However, in scenarios such as availing third-party computation services, the unavailability of reference image makes these metrics obsolete [1]. Nonetheless, there are no-reference quality metrics that evaluate the image quality by using statistical features in it. Such measure scores align well with the perceived quality of images in terms of subjective human quality score than the full-reference images. An example of such algorithms is Perception-based Image QUality Evaluator (PIQUE) proposed in [15], which is sometimes also abbreviated as (PIQE). It is a perception-based quality metric that uses arbitrary distortion in natural images for the image quality evaluation. Its value lies in the range of [0, 100] and have inverse relation with the image quality that is, high value represents lowest quality images. The PIQE values are interpreted in steps of twenties for example, 0 to 20 indicates excellent quality and 81 to 100 indicates bad quality of the image. The rest of the values can be interpreted in the same manner. In our simulations, the PIQE values obtained for measuring the transmitted image quality is shown in Fig. 6. (c). Different from other metrics, PIQE score suggests almost equally poor quality for all transmitted images. The lowest score indicating high quality image is 68. One reason for this is that PIQE is designed for natural images where there is a lot of texture in them; however, in the MNIST dataset, most part of the image is smooth with no variations in the intensity values. In addition, the PIQE score mainly relies on the statistics of the images that were used during training.

C. Deep Learning Model Performance Analysis

For the downstream application task, we have considered the handwritten digit recognition task in the MNIST dataset. For this purpose, we have implemented a convolutional neural network-based multiclass classifier. The classifier has the same architecture as the autoencoder discussed in Section 3, with an additional dense layer of 10 nodes with SoftMax activation function. The model was trained for 50 epochs with ADAM optimizer and the batch size was set to 128. During training, the validation accuracy achieved was 99.32%. Once the model was trained, its accuracy was then evaluated using test data that the model has never seen. The test accuracy achieved was similar to the validation accuracy. Next, to evaluate the model accuracy on the distorted images, the held-

out test set was transmitted over different wireless communication systems. The resultant error in accuracy is summarized in Table 2, which is obtained by (100 – Accuracy). It is given that a trained DL model performance degrades when the data distribution on which the model is tested is different from the one it was trained [5]. Therefore, it can be observed that for lower SNR and higher number of bits per symbol, the model performance degraded. Specifically, when the images were transmitted over the Rayleigh fading channel, then the model accuracy is severely reduced. The main reason for the gap in the DL model performance is that the communication noise altered the distribution of the test set.

D. IQA Correlation with DL Model Performance Analysis

To get an indication of the downstream model performance, we analyzed the correlation between image quality evaluation metrics and the DL model performance. For this purpose, we first computed the PSNR, SSIM and PIQE scores for the wrong predictions (*wp*) and correct predictions (*cp*) made by the classification model. Intuitively, an IQA should indicate lower image quality for the wrong predicted images than the correct ones. In the case of PSNR, it can be seen that the score computed for *wp* and *cp* images is same and does not indicate the DL model performance. For the SSIM metric, we first scaled its values from the range [0, 1] to $-10\log_{10}(1-\text{SSIM})$ because for similar images the SSIM values lie very closely. Compared to PSNR, there is a slight difference between the SSIM values of *wp* and *cp* images. However, the difference is not significant enough. It can be observed that both SNR and SSIM values increase across all modulation schemes when the SNR value is increased, suggesting better quality images. However, this does not coincide with the DL model performance as the values calculated for wrong and correct predictions are same. On the other hand, the PIQE values calculated for *wp* and *cp* are almost the same and as mentioned earlier, regardless of the communication system setup, PIQE computed the same score for all distorted images. For the proposed IQA metric, the score is computed on the whole test set and the percentage of wrong predictions is calculated. It can be seen that for a given modulation scheme the number of wrong predictions decreased as the SNR value is increased, which is in-lined with the task DL model performance. Overall, in predicting the downstream DL model performance, the proposed IQA introduced maximum error of 22% and minimum error of 0.13% while keeping an average error of 8% across all communication systems.

V. Conclusion

The performance of a DL model is strongly associated with the image quality. The existing IQA metrics either rely on or are optimized to coincide with human judgment. However, with the popularity of DL solutions in various domains, it is necessary to quantify perceptual loss in noisy images using machine perception. In this study, we proposed a convolutional autoencoder-based IQA metric that can be used to monitor perceptual distortions occurred in images during data transmission. Hence, the proposed IQA can be used to efficiently adjust the communication system parameters in order to enable AI-based applications in IoT.

In the future, we are interested to analyze the correlation between state-of-the-art DL model performances with the proposed IQA metric score. In the proposed IQA metric the autoencoder can be replaced with variational autoencoder and generative adversarial networks for performance improvement. In addition, we are interested to mitigate the effect of perceptual distortions resulted from a communication system on a DL model performance by proposing noise-based data augmentation method as in [3].

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government. (MSIT) (RS-2023-00278294).

References

[1] N. Ahmed and H. M. S. Asif, “Perceptual Quality Assessment of Digital Images Using Deep Features,” *cai*, vol. 39, no. 3, pp. 385–409, 2020, doi: 10.31577/cai_2020_3_385.

[2] S. U. Saeed *et al.*, “Image quality assessment for machine learning tasks using meta-reinforcement learning,” *Medical Image Analysis*, vol. 78, p. 102427, May 2022, doi: 10.1016/j.media.2022.102427.

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

[4] S. Dodge and L. Karam, “Understanding How Image Quality Affects Deep Neural Networks.” arXiv, Apr. 21, 2016. Accessed: Aug. 26, 2022. [Online]. Available: http://arxiv.org/abs/1604.04004

[5] M. Ehrlich, L. Davis, S.-N. Lim, and A. Shrivastava, “Analyzing and Mitigating JPEG Compression Defects in Deep Learning,” in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, BC, Canada: IEEE, Oct. 2021, pp. 2357–2367. doi: 10.1109/ICCVW54120.2021.00267.

[6] I. Ahmad, N. Islam, E. Kim, and S. Shin, “Performance Analysis of Cloud based Deep Learning Models in OFDM based Image Communication system,” in *Proceedings of the Korean Institute of Communications and Information Sciences (KICS) Summer Conference*, Jeju, Korea: Korean Institute of Communications and Information Sciences, Jun. 2022.

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

[8] S. Park, M. S. Ibrahim, A. Wahab, and S. Khan, “GMDM: A generalized multi-dimensional distribution overlap metric for data and model quality evaluation,” *Digital Signal Processing*, vol. 134, p. 103930, Apr. 2023, doi: 10.1016/j.dsp.2023.103930.

[9] I. Ahmad and S. Shin, “Convolutional Autoencoder for Image Quality Assessment,” presented at the 한국통신학회 학술대회논문집, 한국통신학회, 2023, pp. 1304–1305. [Online]. Available: http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE11487614

[10] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[11] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, “Striving for Simplicity: The All-Convolutional Net,” Dec. 2014.

[12] H. Salehinejad, E. Colak, T. Dowdell, J. Barfett, and S. Valaee, “Synthesizing Chest X-Ray Pathology for Training Deep Convolutional Neural Networks,” *IEEE Trans. Med. Imaging*, vol. 38, no. 5, pp. 1197–1206, May 2019, doi: 10.1109/TMI.2018.2881415.

[13] Y. LeCun, C. Cortes, and C. Burges, “MNIST handwritten digit database,” *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, vol. 2, 2010.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Trans. on Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[15] Venkatanath N, Praneeth D, Maruthi Chandrasekhar Bh, S. S. Channappayya, and S. S. Medasani, “Blind image quality evaluation using perception-based features,” in *2015 Twenty First National Conference on Communications (NCC)*, Mumbai, India: IEEE, Feb. 2015, pp. 1–6. doi: 10.1109/NCC.2015.7084843.