Noise2Quality: Non-Reference, Pixel-Wise Assessment of Low Dose CT Image Quality

Ayaan Haque[a, b], Adam Wang[b], and Abdullah-Al-Zubaer Imran[b]

[a]Saratoga High School, Saratoga, CA, USA
[b]Stanford University, Stanford, CA, USA

ABSTRACT

CT image quality is reliant on radiation dose, as low dose CT (LDCT) scans contain increased noise in images. This compromises the diagnostic performance on such scans. Therefore, it is desirable to perform Image Quality Assessment (IQA) prior to diagnostic use of CT scans. Often, image quality is assessed with full-reference methods, where a LDCT is algorithmically compared against its full dose counterpart. However due to health concerns, acquiring full dose CT scans is challenging and not desirable. As an alternative, non-reference IQA (NR-IQA) can be performed. Moreover, IQA at the pixel level is important, as most IQA methods only provide a global assessment, which means localized regions of interest cannot be specifically assessed. A solution for localized-IQA is to produce visually-interpretable quality maps. Deep learning methods could be employed by leveraging computer vision techniques, such as Self-Supervised learning (SSL). In this work, we propose Noise2Quality (N2Q)—a novel self-supervised, non-reference, pixel-wise image quality assessment model to predict IQA maps from LDCTs. Self-supervised dose level prediction as an auxiliary task further improves the model performance. Our experimental evaluation both qualitatively and quantitatively demonstrates the effectiveness of the model in accurately predicting IQA maps over various baselines. *

Keywords: Computed Tomography, Self-Supervised Learning, Image Quality Assessment, Low-Dose CT

1. INTRODUCTION

Computed Tomography is one of the most fundamental medical imaging modalities, but its effectiveness is heavily reliant on the radiation dose supplied. However, research has confirmed that high radiation dose can be harmful for patients and operators.[1] Low dose CT (LDCT) is an alternative to full dose CT (FDCT), but contains higher quantities of noise that compromises image quality. Certain details required for diagnostic performance may be hidden by such noise. In order to ensure the acquired images are at expected quality, it is imperative to perform image quality assessment (IQA), such that medical practitioners can determine whether certain scans are appropriate for diagnostic analysis.

Full-reference IQA (FR-IQA) methods are generally mathematical calculations which output a metric to evaluate IQA. These conventional IQA metrics include SSIM, PSNR, MSE, FSIM, etc. However, if we want to scan patients at low dose, we will not have full dose CT (FDCT) references, meaning FR-IQA metrics cannot be applied. To address this challenge, non-reference IQA (NR-IQA) metrics have been employed, but most of these methods rely on hand-crafted feature engineering.[2] Recently, deep learning NR-IQA methods have been explored, both in the general computer vision[3,4] and medical imaging fields.

Most IQA methods perform global assessments, providing a score for the entire image. More practical for CT imaging diagnostics would be to measure the quality of a specific organ or region-of-interest (ROI) through pixel-wise local assessment. While scores can be assigned by masking off a certain region, it would be more beneficial to visualize each section of the scan in a single image. Such an assessment can be performed with an image quality assessment map (IQAM)[5] where pixel intensity denotes the quality of the pixel.

Prior deep learning work in IQA exists in both the general domain and the CT domain.[6–8] Patwari et al.[6] proposes a standard CNN which predicts conventional IQA-metrics for a given LDCT input. Li et al.[7] uses

Send correspondence to Ayaan Haque (ayaanzhaque@gmail.com)
*Our code is available at https://github.com/ayaanzhaque/Noise2Quality

a CNN to predict subjective metrics graded by experienced radiologists instead of using conventional FR-IQA metrics. Imran et al.[8] performs blind global-IQA prediction by training a model to predict various FR-IQA metrics from a given LDCT input. They also propose a noise level estimation auxiliary task. However, these methods perform global score prediction instead of localized or map assessment. For map assessment, there are few in the general domain,[9] and for CT, even fewer are available.[10] Pan et al.[9] uses a generative network to input distorted images and predict an IQA map that correlates to human vision and perception. They additionally have a second metric which regresses the generated map into a singular global score. Gao et al.[10] does not predict image quality maps, it only predicts localized scores for specific ROIs and organs. Thus, general interpretations of IQA for all regions of an image cannot be predicted and visualized at once. To our best knowledge, no methods predict CT image quality maps.

Self-Supervised Learning (SSL) is a form of unsupervised learning which produces synthetic labels from the data itself for a separate task.[11] Traditionally, self-supervised learning is performed in a two-step, pre-training process. However, it can be performed simultaneously with the downstream task. In this work, we jointly train our main task and SSL as an auxiliary task in a multitask learning framework. This sort of self-supervision can be seen in many CT denoising methods.[12,13]

In this paper, we propose Noise2Quality (N2Q), a novel NR-IQA method which leverages self-supervised learning to predict structural similarity index map (SSIM) from LDCT images. Our contributions are as follows:

- A novel deep learning approach for local IQA through image quality map prediction, as opposed to only global score prediction, by simulating multiple doses and producing ground-truth SSIM maps
- A novel self-supervised regularization dose-level estimation task to improve representation learning through multitask learning
- Experimentation proving the effectiveness of each components of our approach at accurately predicting image quality maps for localized IQA.

2. METHODS

2.1 Data

We primarily collect abdomen scans from the publicly available Mayo CT data.[14,15] The dataset includes CT scans originally acquired at routine dose level (full dose) as well as simulated quarter dose images reconstructed through inserting Poisson noise into each projection dataset. We use 20 full dose and the corresponding quarter (25%) dose CT scans. We then select the middle 50 slices from each of the scans. The dataset is split into train—10 scans (500 slices) and test—10 scans (500 slices).

2.2 Image Quality Map Generation and Dose Simulation

Our dataset originally only contains one set of LDCT and FDCT images (quarter dose). To evaluate IQA, different dose levels must be present in a single training and testing set. Thus, we synthesize arbitrary dose images by scaling the zero-mean independent noise from 25% to five additional dose levels: 5%, 10%, 50%, and 75%, described in Imran et al.[8]

Image quality assessment maps are the reference images in our training procedure. To generate image quality maps, we use SSIM map as our proxy reference. The SSIM maps are produced by calculating an SSIM value between the LDCT and FDCT for each pixel using a window of 11×11.[5] Thus, in an SSIM IQA map, each pixel intensity represents the SSIM score at that pixel. When all the scores are collected and visualized as an image, ROIs and structures can be interpreted for their image quality.

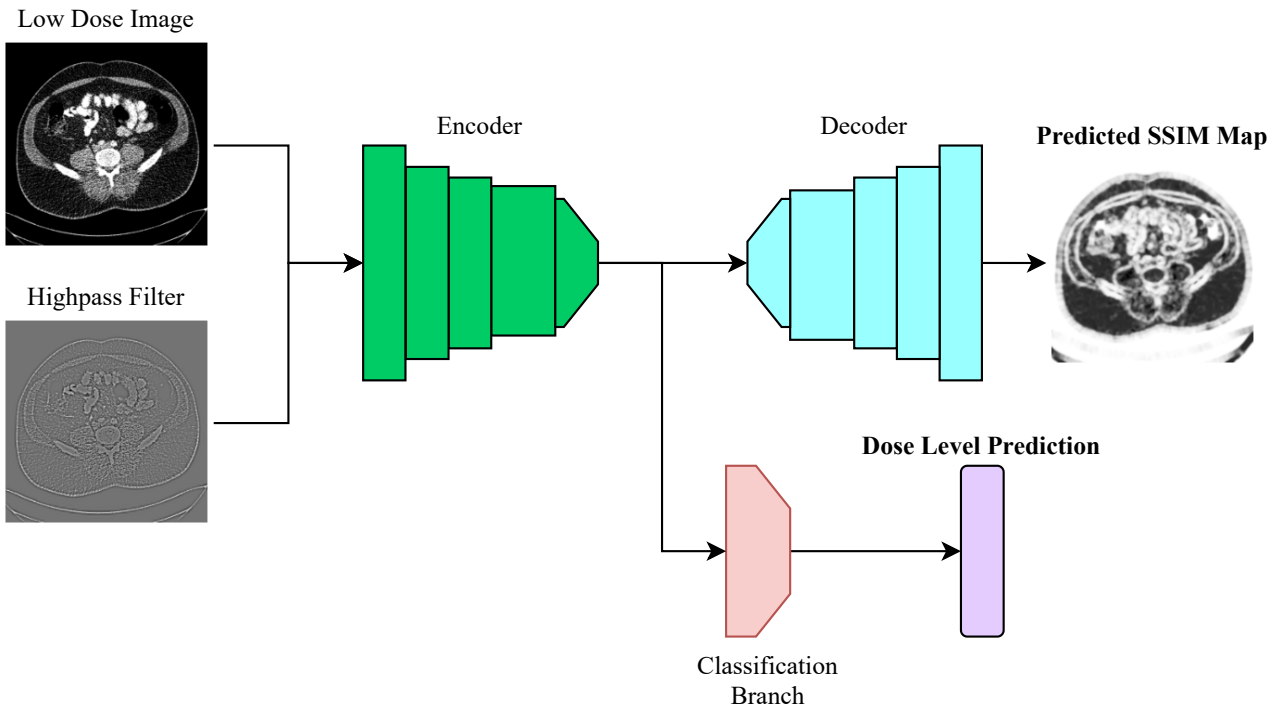

Figure 1. Schematic of the proposed N2Q model. A high pass filter of the input LDCT image is calculated and concatenated with the input image before passing to the model. The model predicts the SSIM map for local IQA as well as self-supervised dose level prediction, which is performed through the classification branch.

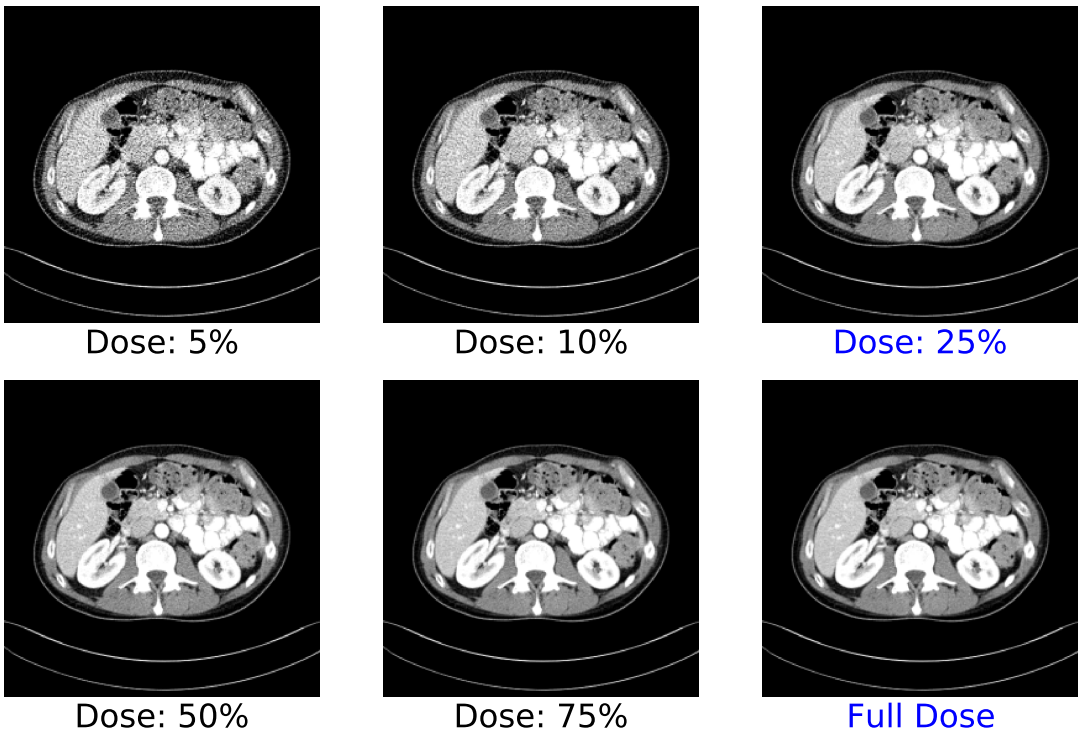

Figure 2. 6 different doses (5%, 10%, 25%, 50%, 75%, full) of a single scan are displayed. 25% and full dose is captioned in blue as they are provided by Mayo dataset.

Table 1. Results of various architectures and training procedures confirm the superiority of our proposed method (N2Q) and validates the importance of each component. The best scores are bolded.

Metrics	AE	RED-CNN	U-Net	U-HPF	U-Con	U-SSL	N2Q
SSIM	0.6761	0.7250	0.7448	0.7381	0.7484	0.7572	**0.7664**
MSE	0.1033	0.0594	0.0804	0.0979	0.0759	0.0669	**0.0686**
NRMSE	0.3226	0.2434	0.2830	0.3237	0.2794	0.2505	**0.2437**

2.3 Noise2Quality

We use a simple encoder/decoder convolutional neural network (CNN) architecture (see Fig. 1). More recent and improved architectures can be utilized, but to isolate the importance of our training procedures, we employ standard baselines. Each input LDCT image slice has a corresponding SSIM Map and our model aims to predict SSIM Maps that are as similar as possible to the ground-truth. We use entire slices instead of image patches, yielding more elegant training over other methods which perform IQA map prediction.[9] Instead of just using the LDCT image as our input, we use a high-pass filter (HPF) and LDCT concatenated input. The high-pass filter channel provides additional context for both modules of the model, improving its performance.

We simultaneously train the model to predict the dose level of the input LDCT image as a self-supervised auxiliary task. After the encoder, which is similar to a standard CNN, we add an average pooling layer and a fully connected layer to make dose level prediction as a 5 class classification task—five classes for each dose level. Each LDCT can be assigned a label based on its dose. This is a self-supervised task because the dose level information is freely available. Training in a multitask learning framework has been shown to improve performance, reduce overfitting, and improve generalization because the model must fit to two separate tasks and data representations.[16] Moreover, information learned from one task can assist in the performance of the other task. This method is does not fit the traditional self-supervised definition from general computer vision, where a network is trained on a subset of unlabeled images. The dose-level estimation task can be described as unsupervised, as all the labels for each image are synthetically generated.

Therefore, our loss has two components: IQA map estimation and dose level prediction. For quality map prediction, we compute a mean-squared error loss (MSE) between predicted quality map and the ground-truth SSIM map. For dose level prediction, we use multi-class cross-entropy (CE) loss, as we formalized our classification problem as a 5-class task. We employ a standard MSE loss instead of a hybrid loss, which have been used in recent CT imaging tasks,[17] to further isolate the importance of our primary contributions and training procedure.

Our final model objective can be represented as:

$$L(y, \hat{y}, c, \hat{c}) = L_{MSE}(\hat{y}, y) + \alpha \mathcal{L}_{CE}(\hat{c}, c), \tag{1}$$

where L_{MSE} is the standard MSE loss, L_{CE} is the cross-entropy loss, y is the ground-truth SSIM map, $\hat{y}$ is the predicted IQA map, c is the reference dose level, and $\hat{c}$ is the predicted dose level. A weight of α is assigned to the dose level prediction task.

3. RESULTS

We experimented with different architectures and training procedures. As baselines, we used a U-Net,[18] an Autoencoder (U-net without skip-connections), and a RED-CNN,[19] which is an autoencoder without downsampling, or constant convolutions of 96 filters. In our proposed approach and ablation experiments, we trained U-HPF—U-Net with only the HPF input, U-Con—U-Net trained with a concatenated HPF and LDCT input, U-SSL—U-Net trained with the dose level estimation task, and N2Q—U-Net trained on concatenated input and SSL.

Table 1 displays the performance of our various architectures and training procedures. For architectures, the U-Net has the strongest overall performance compared to the standard autoencoder and RED-CNN. When trained with only the high-pass filter as the input, the model performance is lower than with LDCT as the input. However, when the HPF is concatenated with the in the image, the performance improves over with only LDCT input. Finally, our self-supervised dose level prediction improves the performance even further, as when it is

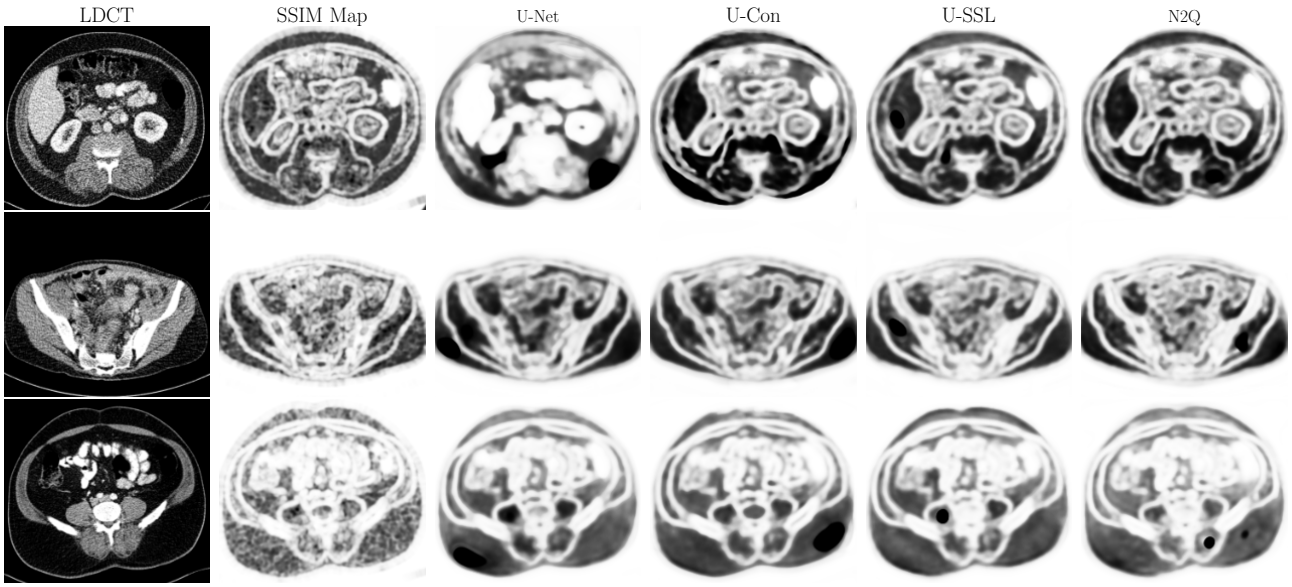

Figure 3. Visual comparison of the predicted IQA map demonstrates strong qualitative improvements from our proposed N2Q over the baseline models.

trained with either LDCT only or the concatenated input, it outperforms its counterpart without SSL. The table overall confirms the importance of each of our contributions.

Fig. 3 shows the predictions of various models on LDCT images. The brighter pixels in the IQA map denote better quality. Each component yields improved performance over its counterpart, and N2Q has the highest similarity with the reference scan. While visually the pixel-intensities and image contrast are not entirely accurate to the ground-truth, as the models tend to underestimate image quality, for the purposes of visual IQA, our model matches well to the ground-truth for indicating high-quality and low-quality areas. This also explains why the SSIM evaluation metric, which is based on luminosity and contrast, has relatively low scores across the board.

4. CONCLUSIONS

We have presented, N2Q, a self-supervised image quality assessment model which predicts SSIM image quality maps for local, pixel-wise assessment. We have introduced a novel self-supervised dose level prediction auxiliary task to improve performance through multitask learning and regularize training. Our proposed method indeed outperforms a number of baselines at producing accurate quality maps. While most works perform global-IQA assessment in medical imaging, we focus on predicting localized-IQA, which is more relevant for clinical applications. Our future research will experiment with more baseline and state-of-the-art methods, as well as using different metrics for IQA maps, such as GSSIM. For the localized and organ-specific quality estimation, organ segmentation will be leveraged. That will enable us to perform more focused and practical image quality assessment in low dose abdominal CT images.

REFERENCES

[1] Brenner, D. J. and Hall, E. J., "Computed tomography — an increasing source of radiation exposure," *New England Journal of Medicine* **357**(22), 2277–2284 (2007). PMID: 18046031.

[2] Mittal, A., Moorthy, A. K., and Bovik, A. C., "No-reference image quality assessment in the spatial domain," *IEEE Transactions on image processing* **21**(12), 4695–4708 (2012).

[3] Liu, X., van de Weijer, J., and Bagdanov, A. D., "Rankiqa: Learning from rankings for no-reference image quality assessment," in [*Proceedings of the IEEE International Conference on Computer Vision (ICCV)*], (Oct 2017).

[4] Dendi, S. V. R., Dev, C., Kothari, N., and Channappayya, S. S., "Generating image distortion maps using convolutional autoencoders with application to no reference image quality assessment," *IEEE Signal Processing Letters* **26**(1), 89–93 (2018).

[5] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing* **13**(4), 600–612 (2004).

[6] Patwari, M., Gutjahr, R., Raupach, R., and Maier, A., "Measuring ct reconstruction quality with deep convolutional neural networks," in [*International Workshop on Machine Learning for Medical Image Reconstruction*], 113–124, Springer (2019).

[7] Li, S., He, J., Wang, Y., Liao, Y., Zeng, D., Bian, Z., and Ma, J., "Blind ct image quality assessment via deep learning strategy: initial study," in [*Medical Imaging 2018: Image Perception, Observer Performance, and Technology Assessment*], **10577**, 105771A, International Society for Optics and Photonics (2018).

[8] Imran, A.-A.-Z., Pal, D., Patel, B., and Wang, A., "SSIQA: Multi-task learning for non-reference CT image quality assessment with self-supervised noise level prediction," in [*2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*], 1962–1965 (2021).

[9] Pan, D., Shi, P., Hou, M., Ying, Z., Fu, S., and Zhang, Y., "Blind predicting similar quality map for image quality assessment," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 6373–6382 (2018).

[10] Gao, Q., Li, S., Zhu, M., Li, D., Bian, Z., Lv, Q., Zeng, D., and Ma Sr, J., "Combined global and local information for blind ct image quality assessment via deep learning," in [*Medical Imaging 2020: Image Perception, Observer Performance, and Technology Assessment*], **11316**, 1131615, International Society for Optics and Photonics (2020).

[11] Jing, L. and Tian, Y., "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE transactions on pattern analysis and machine intelligence* (2020).

[12] Krull, A., Buchholz, T.-O., and Jug, F., "Noise2void-learning denoising from single noisy images," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 2129–2137 (2019).

[13] Hendriksen, A. A., Pelt, D. M., and Batenburg, K. J., "Noise2inverse: Self-supervised deep convolutional denoising for tomography," *IEEE Transactions on Computational Imaging* **6**, 1320–1335 (2020).

[14] Chen, B., Duan, X., Yu, Z., Leng, S., Yu, L., and McCollough, C., "Development and validation of an open data format for CT projection data," *Medical physics* **42**(12), 6964–6972 (2015).

[15] Yu, L., Shiung, M., Jondal, D., and McCollough, C. H., "Development and validation of a practical lower-dose-simulation tool for optimizing computed tomography scan protocols," *Journal of computer assisted tomography* **36**(4), 477–487 (2012).

[16] Haque, A., Imran, A.-A.-Z., Wang, A., and Terzopoulos, D., "Multimix: Sparingly-supervised, extreme multitask learning from medical images," in [*2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*], 693–696 (2021).

[17] Haque, A., Wang, A., and Imran, A.-A.-Z., "Window-level is a strong denoising surrogate," in [*Machine Learning in Medical Imaging*], 457–466, Springer International Publishing, Cham (2021).

[18] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*International Conference on Medical image computing and computer-assisted intervention*], 234–241, Springer (2015).

[19] Chen, H., Zhang, Y., Kalra, M. K., Lin, F., Chen, Y., Liao, P., Zhou, J., and Wang, G., "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Transactions on Medical Imaging* **36**, 2524–2535 (Dec 2017).