

Design of Novel Loss Functions for Deep Learning in X-ray CT

Obaidullah Rahman^a, Ken D. Sauer^a, Madhuri Nagare^b, Charles A. Bouman^b, Roman Melnyk^c, Jie Tang^c, and Brian Nett^c

^aDepartment of Electrical Engineering, University of Notre Dame, U.S.A.

^bSchool of Electrical and Computer Engineering, Purdue University, U.S.A.

^cGeneral Electric Healthcare, Waukesha, U.S.A.

ABSTRACT

Deep learning (DL) shows promise of advantages over conventional signal processing techniques in a variety of imaging applications. The networks' being trained from examples of data rather than explicitly designed allows them to learn signal and noise characteristics to most effectively construct a mapping from corrupted data to higher quality representations. In inverse problems, one has options of applying DL in the domain of the originally captured data, in the transformed domain of the desired final representation, or both.

X-ray computed tomography (CT), one of the most valuable tools in medical diagnostics, is already being improved by DL methods. Whether for removal of common quantum noise resulting from the Poisson-distributed photon counts, or for reduction of the ill effects of metal implants on image quality, researchers have begun employing DL widely in CT. The selection of training data is driven quite directly by the corruption on which the focus lies. However, the way in which differences between the target signal and measured data is penalized in training generally follows conventional, pointwise loss functions.

This work introduces a creative technique for favoring reconstruction characteristics that are not well described by norms such as mean-squared or mean-absolute error. Particularly in a field such as X-ray CT, where radiologists' subjective preferences in image characteristics are key to acceptance, it may be desirable to penalize differences in DL more creatively. This penalty may be applied in the data domain, here the CT sinogram, or in the reconstructed image. We design loss functions for both shaping and selectively preserving frequency content of the signal.

Keywords: Deep learning, neural network, X-ray CT, novel loss functions, spectral shaping.

1. INTRODUCTION

Artificial neural networks (ANN) have been increasingly finding success in X-ray computed tomography (CT).¹⁻¹⁰ ANN in imaging are designed by adjusting strengths of interconnections among artificial neurons with the goal of making the network's output, on the average, as close as possible to the ideal form of the image. This ideal form may be well known in training phase of the ANN, in which one may start with a perfect signal as the "target" and then corrupt it according to the character of noises and artifacts typically encountered in application. Alternatively, the target image may be imperfect, but far less afflicted with error than those encountered as measurements. In training, simple multiplicative coefficients or other representations of neural interconnections are iteratively adjusted to minimize some average measured error, or loss, between an ensemble of network-processed input data and their respective target images, as represented in Figure 1. The measured loss is backpropagated through the ANN to provide gradients to correct the connections and reduce loss, thus "learning" the inverse operator. Following training, the network may be applied to new data sets in order to reduce their content of error as described by the system's loss function. The process is, with increasing frequency, titled "deep learning" because more powerful computational resources have allowed more layers in the ANN, hence a "deeper" network.

Further author information: (Send correspondence to O.R.)

O.R.: E-mail: orahman@nd.edu

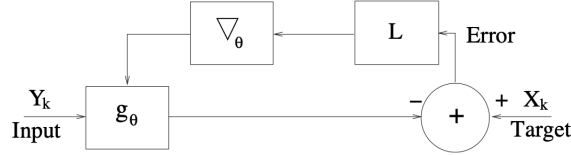


Figure 1: Training of neural network. Parameters governing system behavior are denoted by θ . The gradient of the loss function's penalization of error (L), as a function of θ , is used to improve the averaged match between target and output of network during training.

Probably the most common loss function applied has been mean-squared error. Let us define Y as the input data, which we model as a function of some ideal, target image X , or $Y = h(X)$. The task of the ANN is to extract from Y a rendering close to the unknown, ideal image. If we define $g = h^{-1}$, our training would seek to learn g to produce $X = g(Y)$. Equality is seldom achievable due to noise or other corruption, and we optimize in the sense of average, possibly weighted, error. If we use the variable k to index among training pairs, n to index entries in vectors X_k and Y_k , and θ to represent the variable parameters of the ANN, our DL-trained mapping g_θ for the mean-squared error case may be expressed in terms of

$$\theta = \underset{\theta}{\operatorname{argmin}} \sum_{k,n} w_{k,n} [X_{k,n} - (g_\theta(Y_k))_n]^2 \quad (1)$$

in which the weightings $w_{k,n}$ may be fixed in either or both variables, or may be adapted according to relative local characteristics of data. This weighted, mean-squared penalty on the standard error, $S_k \triangleq X_k - g_\theta(Y_k)$, has a number of potential advantages, including being statistically well-matched to Gaussian noise. In cases where less severe penalization of large errors is desired, squared error may be replaced by absolute error, similarly to penalty adjustment in edge-preserving regularization.

While simple norms such as expressed above provide highly useful loss metrics, it has long been recognized in the image processing community that they may be less than ideal for applications in which the final receiver for the system's output is a human observer. Various metrics for perceptual loss have been designed in hopes of optimizing the elusive human-interpreted quality of audio¹ and visual data.² For diagnostic CT imaging, in which much analysis is performed by radiologists, more subjective quality metrics are applied by the end users of the technology, and spectral content of residual noise, plateauing of image levels in low-contrast areas and other context-dependent evaluations must be addressed.

This work consists of a novel class of loss metrics which may expand the usefulness of DL in X-ray CT. We generalize the sense of optimality to

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_k L[X_k, Y_k, g_\theta(Y_k)], \quad (2)$$

where L is now a function that may capture any number of spectral and spatial characteristics in the error. In the X-ray CT arena, we may choose to improve the signal in either the sinogram domain, where measurements are made directly, or in the image domain after reconstruction by any existing algorithm. The signal and error statistics in these two differ, leading to designs tailored for each case. In the following, we describe one embodiment of the design.

2. METHOD

Conventional, point-wise mean-squared error as loss may be thought of as a flat spectral penalty. However, in cases where we wish to focus on removing artifacts with low or medium spatial frequency content, penalizing all frequencies equally may be counter-productive. Given that many well-developed, edge-preserving techniques are available for removing high-frequency noise, particularly in the image domain, low-signal correction in CT may in some cases be better served by training the network to remove errors only in lower frequencies. In this case, we propose a loss function L in Eq. (3) that may take the form

$$L[S_k] \triangleq \phi[f_1(S_k)], \quad (3)$$

where ϕ is a suitable error metric applied only within the passband of the lowpass filter f_1 . The higher frequency error becomes a “don’t care” element for the network. Alternatively, band-pass or high-pass filtering may focus loss on those portions of the error spectrum. Particularly in three-dimensional image vectors, frequencies may be treated differently along the three axes. This forms the first part of our novel loss function.

The discussion above is most commonly addressed to conventional CT imagery in two or three dimensions, in which spatial frequency has roughly equivalent meaning in all dimensions. However, the present methods are intended at least as importantly for use in the native domain of the data, the sinogram. Application of the type of loss function in Eq. (3) in the sinogram requires modeling behavior in such coordinates as row, channel and view, where the first two index in the detector panel of the CT gantry, and the last indexes the distinct rotating, two-dimensional views of patient or object. In this case, the error filtering operation will need to be spatially adapted, as statistics of both the underlying signal and the corrupting noise vary spatially in the sinogram domain.

It has been widely observed in the DL community that networks appear to have a strong tendency toward elimination of high frequencies in the output and this may occur even when the penalized loss is restricted to low frequency error as in Eq. (3). An example application is using DL for low signal correction, where some of the most problematic artifacts are of low to medium spatial frequency. Here, it may be advantageous to retain parts of the error spectrum in the output when the correction network is applied in the sinogram domain. Powerful, adaptive denoisers in the image domain can capitalize on the relatively stationary underlying image statistics to remove higher frequency noise with little damage to edge resolution. Thus, we may wish to actively discourage suppression of this part of the error signal in the first stage of processing in order to preserve both resolution and desirable texture. We propose a second part of the loss function that will penalize removal of components of the signal Y_k according to their spectral content. This component of the loss may be expressed similarly to Eq. (3), but with the argument redefined as

$$T_k \triangleq Y_k - g_\theta(Y_k) \tag{4}$$

$$L[S_k, T_k] \triangleq \phi[f_1(S_k) + \alpha f_2(T_k)], \tag{5}$$

An realization of the system is shown below in Figure 2. It includes the two loss functions discussed previously. The first loss, realized by the right branch, penalizes the error from Eq. (3) filtered by f_1 . The left branch features the error from Eq. (4), where a different portion of the spatial frequency spectrum of error within the passband of f_2 is penalized. The two types of error signals are combined before the application of the norm ϕ and the gradient for backpropagation. The weighting factor α could be any positive value, with increase resulting in more of the desired frequency components preserved in the output. The responses of filters f_1 and f_2 plus

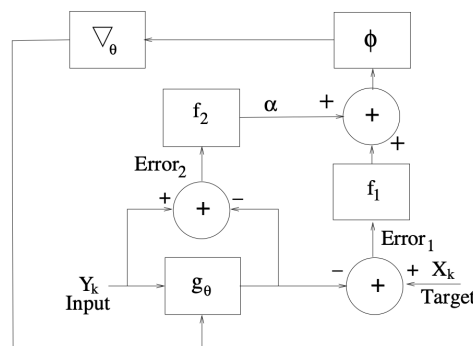


Figure 2: Training of system to encourage the output to mimic the target content as selected by filter f_1 , but refrain from removal of input signal content as selected by filter f_2

the parameter α appear to provide a great deal of control over the inference behavior of the network. In an extremely conservative case, with $f_1 = f_2 = 1.0 \forall \omega$ and $\alpha = 1$, the composite error becomes

$$X_k + Y_k - 2g_\theta(Y_k), \tag{6}$$

which will simply place the optimum output midway between the target and the input.

3. RESULTS

Parts of this method have been preliminarily tested with phantom and clinical data. Below are a few results with the latter. In this configuration, the training loss was the weighted sum of low pass (LP) filtered error between output and target, and high pass (HP) filtered error between input and output. The filters are shown in the Figure 3. The DL network was trained to operate in the original domain i.e. counts domains. Training data consisted of high-dosage Kyoto phantom scans as targets, with synthetic photon counting and electronic noise added to form input sinograms. We can see in Figure 4 the increase in fine-grain texture i.e. high frequency components in the output with increase in α .

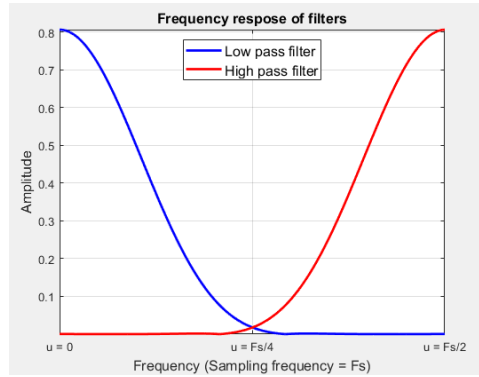
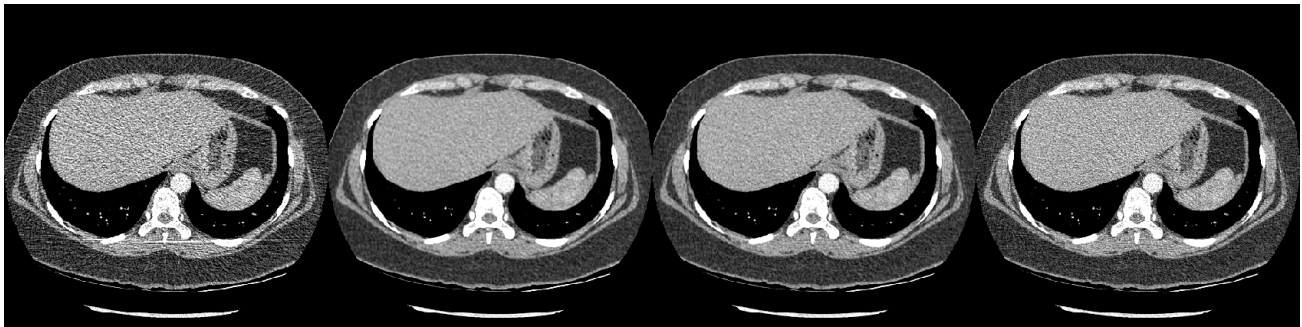
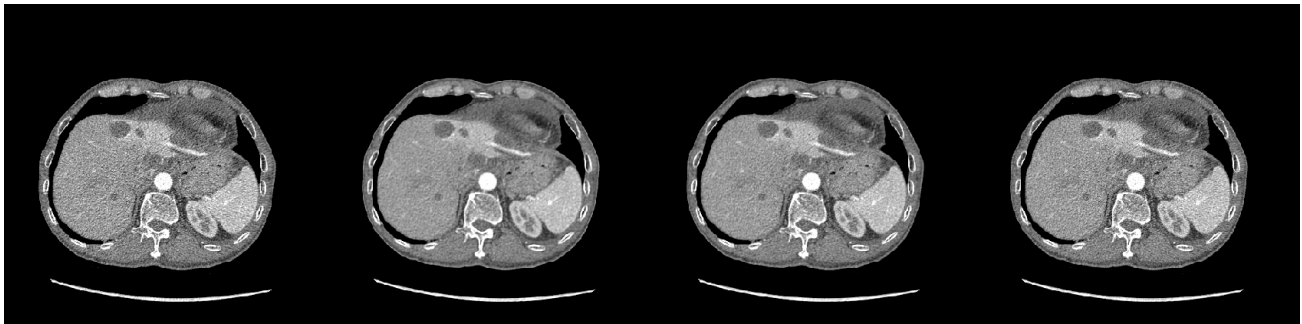


Figure 3: Filters used. The LP filter is f_1 and the HP filter is f_2



(a)



(b)

Figure 4: Reconstructed image (Left to right) Uncorrected; corrected with low pass filter loss ($\alpha = 0$); $\alpha = 0.6$; $\alpha = 0.8$.

The noise power spectra (NPS), shown in Figure 5, were measured in the liver region of reconstructed clinical images. The NPS resulting from the use of only low pass in the loss function ($\alpha = 0$) can be seen to lack much high frequency content. Use of the high pass filter on the error between the input and the output preserves some of the high frequency components, retaining resolution along with high-frequency noise. The value of α can be adjusted based on the balance between NPS qualities and noise tolerance in the image. To assess the flatness of

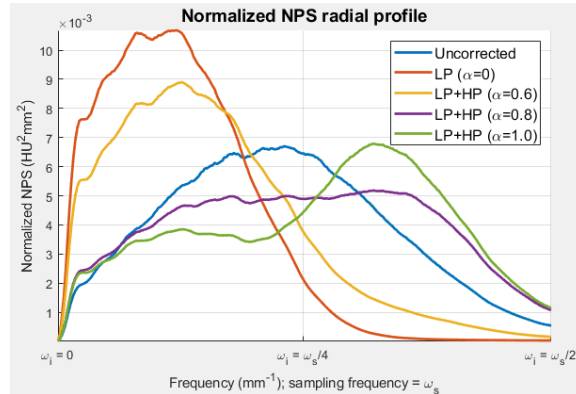


Figure 5: Normalized NPS curves

the NPS curve, entropy measurement was performed as

$$Entropy = \sum_{\omega_i=0}^{\omega_s/2} NPS(\omega_i) \log_2 \frac{1}{NPS(\omega_i)}, \tag{7}$$

where ω_i is the discrete spatial frequency and ω_s is the spatial sampling frequency. It can be seen in Table 1 that the flatness of the NPS increases with α as far as 0.8, but it suffers from excessive high frequency emphasis for α of 1.0. This case exhibits undesirable streaks in the image as well.

Flatness metric (entropy in bits of information)					
Ideal case	Uncorrected	$\alpha = 0$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1.0$
8.00	7.76	7.07	7.46	7.90	7.87

Table 1: Entropy as a measurement of flatness of NPS curves. Higher value indicates flatter, more desirable NPS

4. CONCLUSION

This paper presents a combination of two frequency-weighted loss function components for a deep network, furnishing potentially better control of the behavior of the network in removing signal corruption. The first part of the DL loss function employed here restricts training loss to lower frequency error between a target data set and the input set processed by the network. The second component of the loss ensures the preservation of select error content from the uncorrected data, with the intent of delegating any removal of that error to a later stage of processing. This results in network’s ability to retain desired traits in the data according to chosen models for training loss. In our example application, improvement in the texture of the reconstructed image was observed and confirmed with the NPS metric. Further work will test the value of this design in improving the noise/resolution trade-off in the presence of image-domain postprocessing. We have developed this novel DL loss function design for X-ray CT imaging, but it can easily find application in other areas.

REFERENCES

- [1] Ananthabhotla, I., Ewert, S., and Paradiso, J. A., “Towards a perceptual loss: Using a neural network codec approximation as a loss for generative audio models,” in *[Proceedings of the 27th acm international conference on multimedia]*, 1518–1525 (2019).

- [2] Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M. K., Zhang, Y., Sun, L., and Wang, G., “Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss,” *IEEE transactions on medical imaging* **37**(6), 1348–1357 (2018).
- [3] Suzuki, K., “Transforming projection data in tomography by means of machine learning,” (Jan. 18 2018). US Patent App. 15/646,119.
- [4] Yang, Q., Yan, P., Kalra, M. K., and Wang, G., “Ct image denoising with perceptive deep neural networks,” *arXiv preprint arXiv:1702.07019* (2017).
- [5] Ghani, M. U. and Karl, W. C., “Cnn based sinogram denoising for low-dose ct,” in [*Mathematics in Imaging*], MM2D–5, Optica Publishing Group (2018).
- [6] Lee, T.-C., Zhou, J., and Yu, Z., “Deep learning based adaptive filtering for projection data noise reduction in x-ray computed tomography,” in [*15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*], **11072**, 229–233, SPIE (2019).
- [7] Yuan, H., Jia, J., and Zhu, Z., “Sipid: A deep learning framework for sinogram interpolation and image denoising in low-dose ct reconstruction,” in [*2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*], 1521–1524, IEEE (2018).
- [8] Wang, G., Cong, W., and Qingsong, Y., “Tomographic image reconstruction via machine learning,” (Apr. 6 2021). US Patent 10,970,887.
- [9] Thibault, J.-B., Srivastava, S., Hsieh, J., Bouman Jr, C. A., Ye, D., and Sauer, K., “Image generation using machine learning,” (Sept. 21 2021). US Patent 11,126,914.
- [10] Rahman, O., Nagare, M., Sauer, K. D., Bouman, C. A., Melnyk, R., Nett, B., and Tang, J., “Mbir training for a 2.5 d dl network in x-ray ct,” in [*16th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine, Leuven, Belgium*], 19–23 (2021).