Uncertainty-aware GAN with Adaptive Loss for Robust MRI Image Enhancement

Uddeshya Upadhyay
IIT-Bombay
uddeshya@cse.iitb.ac.in

Viswanath P. Sudarshan
TCS Research
viswanath.pamulakantysudarshan@tcs.com

Suyash P. Awate
IIT-Bombay
suyash@cse.iitb.ac.in

Abstract

Image-to-image translation is an ill-posed problem as unique one-to-one mapping may not exist between the source and target images. Learning-based methods proposed in this context often evaluate the performance on test data that is similar to the training data, which may be impractical. This demands robust methods that can quantify uncertainty in the prediction for making informed decisions, especially for critical areas such as medical imaging. Recent works that employ conditional generative adversarial networks (GANs) have shown improved performance in learning photo-realistic image-to-image mappings between the source and the target images. However, these methods do not focus on (i) robustness of the models to out-of-distribution (OOD)-noisy data and (ii) uncertainty quantification. This paper proposes a GAN-based framework that (i) models an adaptive loss function for robustness to OOD-noisy data that automatically tunes the spatially varying norm for penalizing the residuals and (ii) estimates the per-voxel uncertainty in the predictions. We demonstrate our method on two key applications in medical imaging: (i) undersampled magnetic resonance imaging (MRI) reconstruction (ii) MRI modality propagation. Our experiments with two different real-world datasets show that the proposed method (i) is robust to OOD-noisy test data and provides improved accuracy and (ii) quantifies voxel-level uncertainty in the predictions.

1. Introduction and Related Work

The image-to-image translation methods that generate images of the target view, given images of the source view, are often *ill-posed* problems as unique one-to-one mapping may not exist between the source and target views. Learning-based methods proposed in this context [18, 21, 12] evaluate the performance on test data that is similar to training data. Such models suffer from severe degradation in the performance when presented with out-of-distribution (OOD)-noisy data [11, 19, 24]. In critical applications, like in medical imaging, in addition to robustness to OOD-noisy data, it is important to quantify the uncertainty in the predictions made by the model to aid clinical decisions [26, 34]. Particularly, problems such as enhancing the quality of a given medical image and synthesizing medical images of target modality given a source modality benefit from the uncertainty estimation by allowing risk assessment in the predicted images [37, 26, 34]. In this work, we use two critical tasks in medical image analysis that can broadly be posed as image-to-image translation problems to demonstrate the efficacy of the proposed methods: (i) undersampled magnetic resonance imaging (MRI) reconstruction and (ii) MRI modality propagation, i.e., synthesizing T2 weighted (T2w) MRI from T1 weighted (T1w) MRI.

MRI k-space (a 2D complex-valued space based on 2D Fourier transform of the slices) data acquisition is a time-consuming process as the speed is dependent on hardware and physiological constraints [25, 32, 44]. Typically, faster acquisitions are realized by undersampling the k-space. However, it leads to blurring or aliasing effects in the MRI image. We formulate the reconstruction of high-quality MRI scans using undersampled k-space data as a translation from low-quality MRI (lqMR) to high-quality MRI (hqMR) scans (lqMR $\rightarrow$ hqMR). Similarly, modality propagation is of interest because routine clinical protocols acquire images with multiple contrasts which are critical for better diagnostics. Acquiring multiple contrasts increases scanning time. T1w MRI and T2w MRI are among the most commonly acquired contrasts and synthesis of T2w from T1w is posed as a translation problem (T1w $\rightarrow$ T2w).

Recently, conditional generative adversarial networks (GANs) have shown substantially improved performance for the above-mentioned tasks in comparison to other learning-based methods [33, 28, 6, 35, 40, 39, 38]. Such methods condition the generator (that generates the target view) with a source view and the task of the discriminator is to discriminate between the generated target samples from the ground-truth target samples through adversarial loss term. Additionally, some methods also impose a

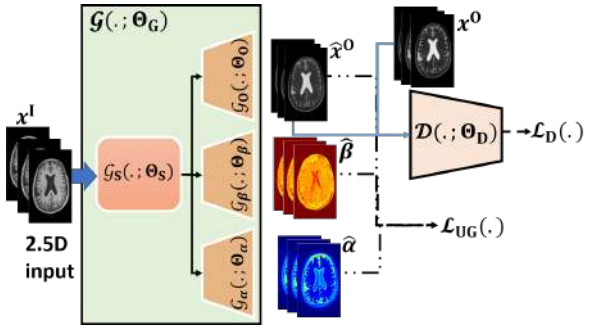

Figure 1. **Model schema.** $\hat{x}^{\mathrm{O}}$, $\hat{\beta}$, and $\hat{\alpha}$ denote the estimated output image, shape, and the scale parameter of the GGD prior respectively. The parameters governing output $\hat{x}^{\mathrm{O}}$ are a combination of the shared weights (across all three branches) in $\mathcal{G}_{\mathrm{S}}$ and $\hat{x}^{\mathrm{O}}$-specific weights in $\mathcal{G}_{\mathrm{O}}$. Similarly, $\hat{\alpha}$ and $\hat{\beta}$ depend on parameters in $\mathcal{G}_\alpha$ and $\mathcal{G}_\beta$, respectively, in addition to shared weights in $\mathcal{G}_{\mathrm{O}}$ (refer Equation 11 – 13).

cycle-consistency penalty to constraint the ill-posed translation problem [45, 21, 12]. Along with improved accuracy, quantifying uncertainty can potentially aid in clinical decision making, especially in the presence of OOD-noisy data, while improving the accuracy of subsequent tasks such as image segmentation, classification, etc [37]. In the general context of medical imaging, OOD-noisy perturbations could be systemic (scanner-related) and/or physiological. Hence, beyond reliable image synthesis during inference, a thorough analysis of the robustness of the network along with quantification of uncertainty in the predicted images is crucial for clinical translation of synthesis frameworks [42, 27, 29]. Designing learning strategies that are robust to large errors (i.e., outliers whose residuals follow heavy-tailed distributions) is a well-studied problem in statistics, optimization, and parameter estimation [2, 10, 13, 5].

Quasi norm-based loss function, ℓ_q, use a generalized notion of the norm, where q (often called the shape parameter) can be tuned empirically for robust learning [36, 1]. Moreover, such empirical methods often require the q to be fixed spatially. This work proposes a novel loss function that automatically learns the shape parameter q that can vary spatially for the image dataset along with the scale parameter that allows us to estimate per-voxel uncertainty.

2. Methods

We describe our proposed GAN framework and discuss specific details pertaining to the two applications: (i) MRI reconstruction using undersampled k-space data posed as a quality enhancement (QE) task and (ii) Modiality Propagation (MP). Our proposed solution for these two applications can easily be extended to other image-to-image translation problems both with MRI as well as other modalities.

Let the set of input images be represented by X^{I} and the corresponding set of output images be represented by X^{O}. Let x^{I} and x^{O} represent elements from the set of input and output images (i.e., $x^{\mathrm{I}} \in X^{\mathrm{I}}$ and $x^{\mathrm{O}} \in X^{\mathrm{O}}$), respectively. We learn the mapping from input to output using T pairs of co-registered training data: $\mathcal{X} := \{(x_t^{\mathrm{I}}, x_t^{\mathrm{O}})\}_{t=1}^{T}$. The input and output images for QE and MP are described below.

Undersampled MRI reconstruction (QE). Here, a mapping is learned from a low-quality MRI image (lqMR) (obtained via zero-filled inverse Fourier transform (IFT) of the undersampled k-space data) to the corresponding high-quality MRI image (hqMR) (obtained from fully-sampled k-space). Therefore, for QE, $(X^{\mathrm{I}}, X^{\mathrm{O}})$ becomes $(X^{\mathrm{LQ}}, X^{\mathrm{HQ}})$. Let operator $\mathcal{F}$ represent the 2D Fourier transform and matrix H represent the sampling mask to perform undersampling in k-space. Then, the undersampled MRI is related to fully-sampled MRI via $x^{\mathrm{LQ}} = \mathcal{F}^{-1}(\mathcal{F}(x^{\mathrm{HQ}}) \odot H)$. With a fixed sampling mask, accounting for noisy k-space acquisitions, corrupted by complex Gaussian noise (say η_{K}), that represent OOD-noisy data [41, 46, 20], the OOD-noisy input sample (say $x^{\mathrm{LQ}}_{\mathrm{OOD}}$) can be modeled by,

$$x^{\mathrm{LQ}}_{\mathrm{OOD}} = \mathcal{F}^{-1}((\mathcal{F}(x^{\mathrm{HQ}}) + \eta_{\mathrm{K}}) \odot H) \tag{1}$$

Modality propagation (MP). Here, we learn a mapping from T1w images to corresponding T2w images. In this context, $(X^{\mathrm{I}}, X^{\mathrm{O}})$ can be represented by the set of T1w MR and T2w MR images, i.e., $(X^{\mathrm{T1}}, X^{\mathrm{T2}})$. Unlike QE, the OOD-noisy input sample (say $x^{\mathrm{T1}}_{\mathrm{OOD}}$) during inference is obtained by adding Gaussian noise in the image-space (say η_{I}) to the T1w MRI input image (say x^{T1}). The addition of Gaussian noise in image-space is consistent with [9], i.e.,

$$x^{\mathrm{T1}}_{\mathrm{OOD}} = x^{\mathrm{T1}} + \eta_{\mathrm{I}} \tag{2}$$

2.1. Model

This work, inspired by conditional GANs, proposes a novel GAN-based framework which (i) employs learnable quasi-norm loss, and (ii) estimates uncertainty maps, as described below. Let $\mathcal{G}(\cdot;\theta^{\mathrm{G}})$ and $\mathcal{D}(\cdot;\theta^{\mathrm{D}})$ represent our generator and discriminator respectively. The input to the generator is $x^{\mathrm{I}} \in X^{\mathrm{I}}$, the predicted image is $\hat{x}^{\mathrm{O}} = \mathcal{G}(x^{\mathrm{I}};\theta^{\mathrm{G}})_O = \mathcal{G}_{\mathrm{O}}(\mathcal{G}_{\mathrm{S}}(x^{\mathrm{I}};\theta^{\mathrm{S}});\theta^{\mathrm{O}})$ (i.e., output of O-branch of $\mathcal{G}$ as shown in Figure 1), and the ground-truth is $x^{\mathrm{O}} \in X^{\mathrm{O}}$. Each image consists of K pixels. We denote j^{th} pixel in i^{th} image, say z, as z_{ij}.

Prior works typically formulate the above task as a regression and the distribution of residuals between the predictions and the ground-truths is assumed to be isotropic standard Gaussian distribution (zero mean and fixed standard deviation). However, some of the limitations of this approach include: (i) not being able to model the outliers/corruptions in the dataset, as residuals due to outliers tend to follow heavy-tailed distributions. (ii) the assumption of fixed standard deviation does not account for heteroscedastic uncertainty in their modeling. Recent works

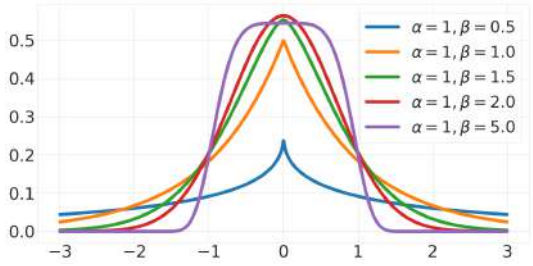

Figure 2. **Generalized Gaussian Distribution**. PDF for GGD with differnt shape parameters (β). Lower β corresponds to heavy-tailed distribution.

model the heteroscedastic uncertainty by assuming that the variance is data-dependent [15]. We improve upon the assumptions made by prior works by (i) modeling the residuals to follow a *generalized Gaussian distribution* (GGD) with zero mean, which is $\frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})}e^{-\left(\frac{|\epsilon-0|}{\alpha}\right)^{\beta}}$, and (ii) allowing the shape parameter (β) and the scale parameter (α) to vary spatially. This allows us to learn appropriate quasi-norms at all spatial locations, as well as estimating uncertainty at every output pixel. Therefore, for the i^{th} image, the residual at pixel location j, ϵ_{ij} (between the predicted value $\hat{x}^{\mathrm{O}}_{ij}$ and ground-truth value x^{O}_{ij}), follows GGD, i.e.,

$$\hat{x}^{\mathrm{O}}_{ij} := x^{\mathrm{O}}_{ij} + \epsilon_{ij} \tag{3}$$

$$\epsilon_{ij} \sim \frac{\beta_{ij}}{2\alpha_{ij}\Gamma(\frac{1}{\beta_{ij}})}e^{-\left(\frac{|\epsilon_{ij}-0|}{\alpha_{ij}}\right)^{\beta_{ij}}} \tag{4}$$

$$\hat{x}^{\mathrm{O}}_{ij} \sim \frac{\beta_{ij}}{2\alpha_{ij}\Gamma(\frac{1}{\beta_{ij}})}e^{-\left(\frac{|(\mathcal{G}_{\mathrm{O}}(\mathcal{G}_{\mathrm{S}}(x^{\mathrm{I}};\theta^{\mathrm{S}});\theta^{\mathrm{O}}))_{ij}-x^{\mathrm{O}}_{ij}|}{\alpha_{ij}}\right)^{\beta_{ij}}} \tag{5}$$

GGD is capable of modelling heavy-tailed distributions including the Gaussian and Laplace PDFs as shown in Figure 2. Here $\alpha_{ij} > 0$ is the scale parameter, $\beta_{ij} > 0$ denotes the shape parameter, and $\Gamma(\cdot)$ is the standard gamma function. In our formulation all the ϵ_{ij}'s are independent but not necessarily identically distributed as α_{ij} and β_{ij} may vary spatially. Hence, the likelihood is,

$$P(\mathcal{X}|\Theta) := \prod_{i=1,j=1}^{i=T,j=K} \frac{\beta_{ij}}{2\alpha_{ij}\Gamma(\frac{1}{\beta_{ij}})}e^{\left(\frac{-|(\mathcal{G}(x^{\mathrm{I}};\theta^{\mathrm{G}})_{O})_{ij}-x^{\mathrm{O}}_{ij}|}{\alpha_{ij}}\right)^{\beta_{ij}}} \tag{6}$$

Therefore, log-likelihood is,

$$\log P(\mathcal{X}|\Theta) = \sum_{i=1,j=1}^{i=T,j=K} -\left(\frac{|\hat{x}^{\mathrm{O}}_{ij} - x^{\mathrm{O}}_{ij}|}{\alpha_{ij}}\right)^{\beta_{ij}} + \log\frac{\beta_{ij}}{2\alpha_{ij}} - \log\Gamma(\frac{1}{\beta_{ij}}) \tag{7}$$

where Θ represents the collection of network parameters. In this way, to improve the robustness of the network, we

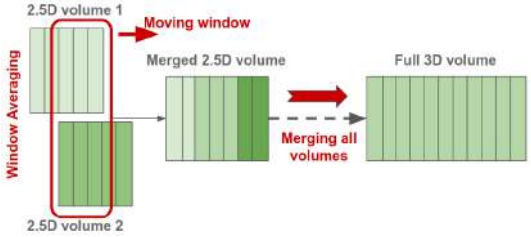

Figure 3. **Merging scheme.** Moving window average to merge overlapping 2.5D volumes to produce a smooth 3D volume.

predict (i) $\hat{x}^{\mathrm{O}}_{ij}$ (the predicted value at every pixel location, also the mean of GGD), (ii) $\hat{\alpha}_{ij}$ (an estimate for the true α_{ij} at every pixel location), and (iii) $\hat{\beta}_{ij}$ (an estimate for the true β_{ij} at every pixel location). Hence, the proposed robust quasi-norm based loss is

$$\mathcal{L}_{\mathrm{U}}(\{\hat{x}^{\mathrm{O}}_{ij}\},\{\hat{\alpha}_{ij}\},\{\hat{\beta}_{ij}\},\{x^{\mathrm{O}}_{ij}\}) := -\log P(\mathcal{X}|\Theta) \tag{8}$$

$\forall 1 \leq i \leq T$ and $1 \leq j \leq K$. Where $\{\hat{x}^{\mathrm{O}}_{ij}\}$ indicates the set of all pixel values over all the images in the dataset, similarly for others. In addition to the above uncertainty-aware fidelity loss term, the adversarial term depending upon $\mathcal{D}(\cdot;\theta^{\mathrm{D}})$ is defined in terms of binary cross-entropy between the true and predicted probability vectors for each of the generated image (say $\hat{y}_i = \mathcal{D}(\hat{x}^{\mathrm{O}}_i;\theta^{\mathrm{D}})$) and ground-truth image (say $y_i = \mathcal{D}(x^{\mathrm{O}}_i;\theta^{\mathrm{D}})$), respectively. The binary cross entropy loss is given by, $\mathcal{L}_{\mathrm{CE}}(\hat{y}_i, y_i) = -\sum_c [y_{ic}\log(\hat{y}_{ic}) + (1-y_{ic})\log(1-\hat{y}_{ic})]$, where y_{ic} represents the c^{th} element in the output probability vector obtained from $\mathcal{D}$ with input x^{O}_i. Hence, $\mathcal{G}(\cdot;\theta^{\mathrm{G}})$ minimizes the loss $\mathcal{L}_{\mathrm{UG}}(\{\hat{x}^{\mathrm{O}}_{ij}\},\{\hat{\alpha}_{ij}\},\{\hat{\beta}_{ij}\},\{x^{\mathrm{O}}_{ij}\})$ given by,

$$\mathcal{L}_{\mathrm{U}}(\{\hat{x}^{\mathrm{O}}_{ij}\},\{\hat{\alpha}_{ij}\},\{\hat{\beta}_{ij}\},\{x^{\mathrm{O}}_{ij}\}) + \lambda\sum_i \mathcal{L}_{\mathrm{CE}}(\mathcal{D}(\hat{x}^{\mathrm{O}}_i;\theta^{\mathrm{D}}),1) \tag{9}$$

On the other hand, $\mathcal{D}(\cdot;\theta^{\mathrm{D}})$ minimizes $\mathcal{L}_D(\{\hat{x}^{\mathrm{O}}_i\},\{x^{\mathrm{O}}_i\})$ given by,

$$\sum_i \mathcal{L}_{\mathrm{CE}}(\mathcal{D}(x^{\mathrm{O}}_i;\theta^{\mathrm{D}}),1) + \mathcal{L}_{\mathrm{CE}}(\mathcal{D}(\hat{x}^{\mathrm{O}}_i;\theta^{\mathrm{D}}),0) \tag{10}$$

We use the strategy elucidated in [28, 8] for training.

The architecture of the proposed generator is inspired by U-Net [31]. We modify the U-Net such that the last convolutional layer is split into three, i.e., three independent convolutional layers attached to the penultimate block in U-Net (as shown in Figure 1). This model predicts $\hat{x}^{O}$, $\hat{\alpha}$, and $\hat{\beta}$. The discriminator network ($\mathcal{D}(\cdot;\theta^{\mathrm{D}})$ in Figure 1) is a typical CNN architecture as described in [6]. Although image synthesis benefits from 3D processing, training CNNs with 3D images poses substantial challenges such as increased computational demand, cost, and training time. Several works use patch-based approaches to circumvent this issue.

However, the final reconstruction from patches might result in artifacts [33]. In this work, we design our model to accept 2.5D input as described in [4] and produce a 2.5D output, this allows the model to exploit 3D like neighborhOOD-noisy information while learning, resulting in better quality outputs. The final 3D output is obtained by joining (via averaging) overlapping 2.5D volumes as shown in Figure 3.

2.2. Training and Testing Scheme

All the networks were trained using Adam optimizer [16] by sampling mini-batches of size 16. The initial learning rate was set to $2e^{-4}$ and cosine annealing was used to decay the learning rate with epochs. The λ for MP and SR (Equation 9) was set to $7e^{-4}$ and $1e^{-3}$ respectively. For numerical stability, the proposed network produces $\frac{1}{\hat{\alpha}}$ instead of $\hat{\alpha}$. The positivity constraint on the output is enforced by applying the ReLU activation function at the end of the three output layers in the network (Figure 1). In addition to the network-generated outputs ($\hat{x}^{\mathrm{O}}$, $\hat{\alpha}$, $\hat{\beta}$), we also compute aleatoric and epistemic uncertainty maps denoted by $\hat{\sigma}_{\text{aleatoric}}$ and $\hat{\sigma}_{\text{epistemic}}$, respectively. While $\hat{\sigma}_{\text{epistemic}}$ captures the uncertainty in the model parameters, $\hat{\sigma}_{\text{aleatoric}}$ captures the data-dependent uncertainty, we combine both to produce $\hat{\sigma}$. To capture the epistemic uncertainty, multiple forward passes (R forward passes) are done for every single input image at inference with dropouts activated. In this way, the final output of the proposed framework, for a given input x^{I} along with the uncertainty maps are given by,

$$\hat{x}^{\mathrm{O}}_r = \mathcal{G}_{\mathrm{O}}(\mathcal{G}_{\mathrm{S}}(x^{\mathrm{I}};\theta^{\mathrm{S}}_r);\theta^{\mathrm{O}}_r) \tag{11}$$

$$\hat{\alpha}_r, \hat{\beta}_r = \mathcal{G}_{\alpha}(\mathcal{G}_{\mathrm{S}}(x^{\mathrm{I}};\theta^{\mathrm{S}}_r);\theta^{\alpha}_r), \mathcal{G}_{\beta}(\mathcal{G}_{\mathrm{S}}(x^{\mathrm{I}};\theta^{\mathrm{S}}_r);\theta^{\beta}_r) \tag{12}$$

$$\hat{x}^{\mathrm{O}}, \hat{\alpha}, \hat{\beta} = \frac{1}{R}\sum_{r=1}^{r=R}\hat{x}^{\mathrm{O}}_r, \frac{1}{R}\sum_{r=1}^{r=R}\hat{\alpha}_r, \frac{1}{R}\sum_{r=1}^{r=R}\hat{\beta}_r \tag{13}$$

$$\hat{\sigma}^2_{\text{aleatoric}}, \hat{\sigma}^2_{\text{epistemic}} = \frac{\hat{\alpha}^2\Gamma(3/\hat{\beta})}{\Gamma(1/\hat{\beta})}, \frac{\sum_{r=1}^{r=R}(\hat{x}^{\mathrm{O}}_r - \hat{x}^{\mathrm{O}})^2}{R} \tag{14}$$

$$\hat{\sigma}^2 = \hat{\sigma}^2_{\text{aleatoric}} + \hat{\sigma}^2_{\text{epistemic}} \tag{15}$$

where r denotes the index for r^{th} forward pass through the network and $\{\theta^{\mathrm{S}}_r, \theta^{\mathrm{O}}_r, \theta^{\alpha}_r, \theta^{\beta}_r\}$ denotes the parameters that are instantiated due to dropouts on the r^{th} forward pass, as described in [7]. In our experiments R was set to 50.

3. Datasets and Experiments

For the two applications QE and MP, we describe the *in vivo* training data and generation of OOD-noisy test data.

3.1. Datasets

We use the following two datasets: (i) proprietary dataset consisting of registered T1w MRI scans for QE, and (ii) publicly available IXI dataset (https://brain-development.org/ixi-dataset/) for MP. The dataset in (i) consists of T1w MRI scans with 1 mm ^{3} isotropic resolution from 100 subjects. We consider these images as the HQ-MR images for the QE experiment. The methodology to obtain LQ-MR images is shown in Figure 4. For LQ data acquisition, we sample only 30% of the k-space using the mask in Figure 4, resulting in an acceleration of $> 3\times$. Subsequently, the LQ images are obtained as the inverse Fourier transform (IFT) of the undersampled k-space. We use training, validation, and test split of 70, 10, and 20 non-overlapping subjects respectively. For each subject, we use 50 mid-brain axial slices. The dataset in (ii) consists of multi-contrast MR images from several healthy subjects, collected across three different scanners. For MP, we use T1w and T2w contrast images which were co-registered (intra-subject) using ANTS [17]. We employ a training, validation, and test split of 250, 50, and 100 non-overlapping subjects respectively and we use 70 axial slices from the mid-brain region per subject.

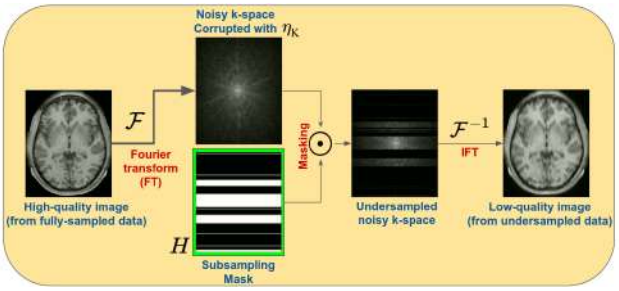

Figure 4. **Retrospective data generation for QE.**. MRI undersampling is achieved by acquiring limited frequencies in k-space.

3.2. Experiments

	QE			MP		
Noise levels	$\hat{\alpha}$ Mean (std)	$\hat{\beta}$ Mean (std)	PSNR Mean (std)	$\hat{\alpha}$ Mean (std)	$\hat{\beta}$ Mean (std)	PSNR Mean (std)
NL0	0.007 (0.003)	0.61 (0.005)	31.22 (1.564)	0.012 (0.001)	1.12 (0.002)	26.18 (1.334)
NL1	0.006 (0.003)	0.60 (0.005)	31.00 (1.597)	0.011 (0.002)	1.05 (0.005)	25.75 (1.767)
NL2	0.006 (0.003)	0.59 (0.006)	29.73 (1.487)	0.011 (0.002)	0.98 (0.003)	25.27 (1.137)
NL3	0.006 (0.003)	0.57 (0.009)	28.02 (1.424)	0.011 (0.003)	0.91 (0.003)	24.35 (1.624)

Table 1. **Trends across noise levels.** overall variations in scale parameter ($\hat{\alpha}$), beta parameter ($\hat{\beta}$), and the output PSNR values across all the noise levels for QE and MP.

OOD-noisy data in MRI. MRI data acquisition depends on several factors that may affect the quality of the MRI scans leading to a wide distribution of datasets. In practice, the dataset used for learning algorithms will not cater to all the possible variations in the distributions; hence, it is imperative to design learning algorithms that are robust to a wide-range of OOD-noisy data. In particular, the signal-to-noise ratio (SNR) of the MRI k-space data (and consequently the image) depends on various factors such as field strength, pulse sequence, tissue characteristics, number of receiver coils and their sensitivities, scan/physiological parameters, etc. [41, 46, 30, 22]. Hence, for both QE and MP,

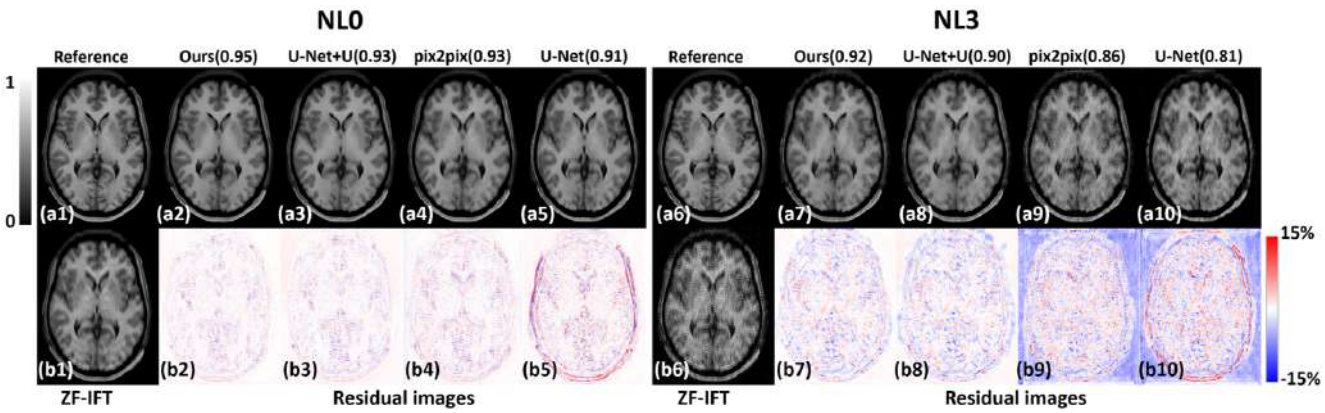

Figure 5. **Results for QE.** Predicted HQ images with input LQ images at two different noise-levels (NLs), **(b1):** NL0 (high SNR, same NL as training-set), and **(b5):** NL3 (low SNR, highest simulated NL). **(a2)–(a5) and (a7)–(a10):** Predicted images for all the methods at NL0 and NL3, respectively. **(b2)–(b5) and (b7)–(b10):** Corresponding residual images. SSIM values are indicated within paranthesis.

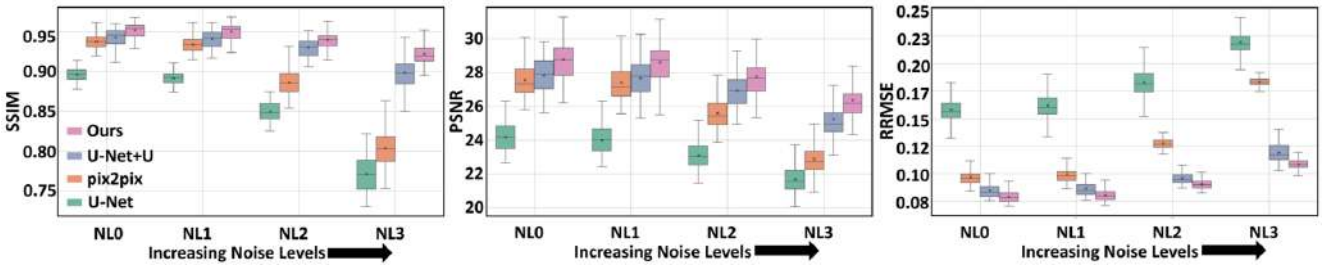

Figure 6. **Quantitative assessment for QE.** SSIM, PSNR, and RRMSE values for all the methods at multiple noise-levels (NL0 to NL3). At each NL, 50 mid-brain slices from each subject (20 test-subjects) were evaluated (i.e., 1000 slices).

we generate OOD-noisy MRI data that captures variations in noise levels as described below. We consider the noise-level (NL) of the training set to be NL0 for both QE and MP. We model *three* increasing noise-levels, NL1–NL3, of OOD-noisy degradations for the test data. Degradations are in (i) k-space for QE and (ii) image-space for MP.

QE. Having trained the network using images at NL0, at inference, we evaluate the robustness of all the methods at *three* increasing levels of noise in the k-space. In practice, this is obtained by increasing the magnitude of complex Gaussian noise η_K in the k-space and as described in Equation 1. In our experiments, sampling mask H is fixed (i.e., fixing the k-space trajectory [23, 20], Figure 4), acquiring only 30% of the k-space. The PSNR values of the LQ images at NL0–NL3, averaged across the test set, were around $\{21, 18, 16, 14\}$ dB with respect to HQ images.

MP. Here we employ two kinds of OOD-noisy test data: (i) evaluation at *three* increasing levels of noise in the image-space and (ii) inclusion of *unseen* synthetic lesions. The image-space degradations are obtained by increasing the magnitude of Gaussian noise η_I Equation 2. For the lesions related experiment, we add lesions in both the input (T1w) and the reference images (T2w) based on the segmented lesion masks available from the BRATS 2020 dataset (https://www.med.upenn.edu/cbica/brats2020/data.html).

4. Results and Discussion

In this work, for a fair evaluation, we modify the baselines (originally in 2D) to use a 2.5D-style training strategy. We use structural similarity index (SSIM) [43], peak signal-to-noise ratio (PSNR) and relative root mean squared error (RRMSE) for quantitative comparison. RRMSE between two images a and b is defined as $\text{RRMSE}(a, b) = \|a - b\|_F / \|a\|_F$, where $\|\cdot\|_F$ indicates Frobenius norm.

Quantitative and Qualitative Evaluations for QE. We use the following state-of-the art methods for comparison: (i) **U-Net** as used in [44], (ii) **pix2pix** as used in [44, 14], and (iii) **U-Net with uncertainty (U-Net+U)** as used in [44]. Finally, our proposed method, **URGAN (ours)**. Our uncertainty model is based on *generalized Gaussian distributions* allowing our loss function to be tuned automatically, unlike that of (U-Net+U) as described in [44]. All the baseline models are trained with the loss function proposed in their respective formulations. Figure 5 shows the predicted images for QE at NL0 and NL3, for a representative test slice. At NL0, the predicted images from all the methods (Figure 5 (a2) – (a5)) have low residual error and are comparable among each other. However, at NL3, pix2pix and U-Net generate images with artifacts (Figure 5 (b4) - (b5)). Whereas, the predicted images from methods that model uncertainty (Ours and U-Net+U) show a superior recovery of structure and contrast in addition to

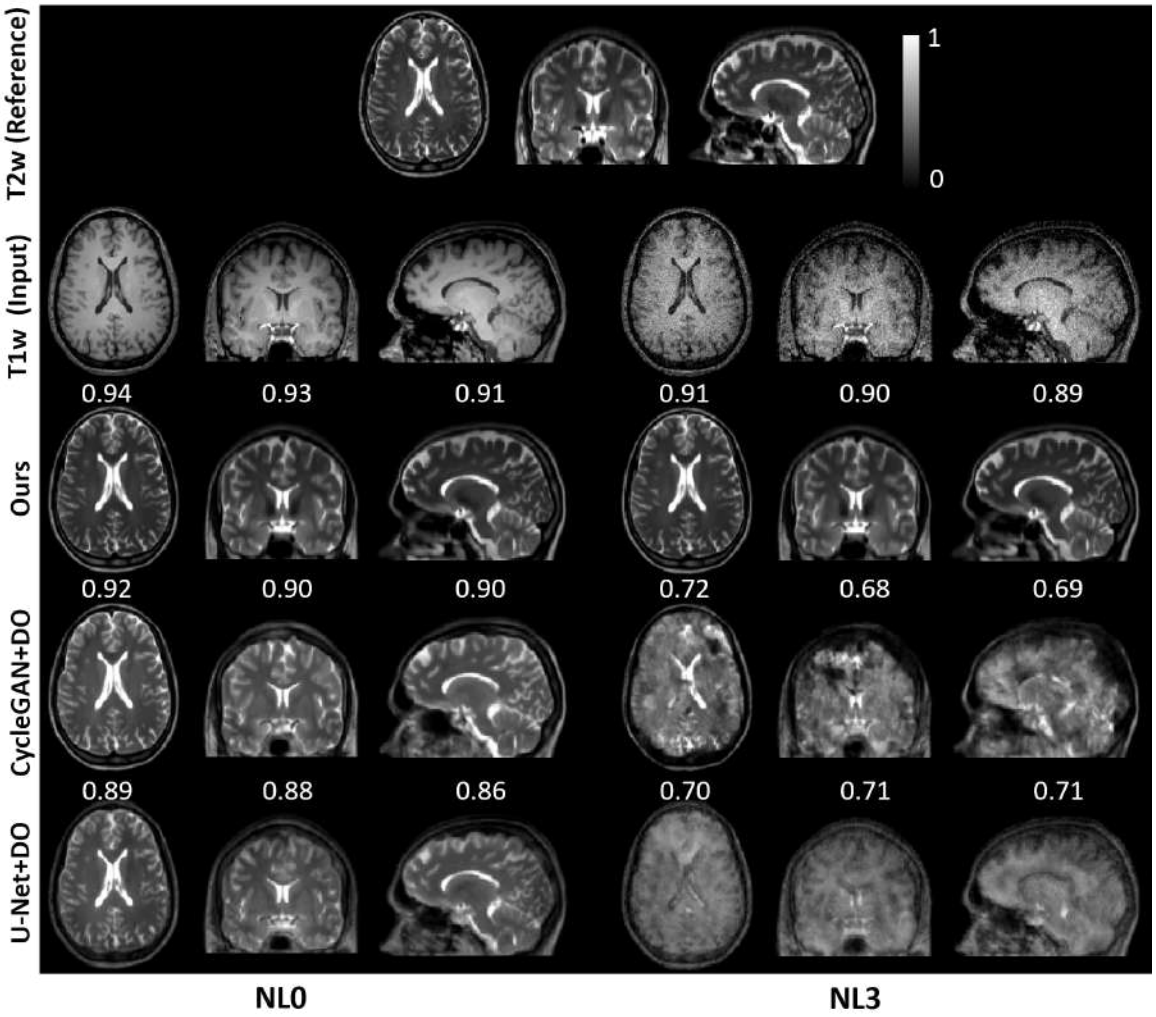

Figure 7. **Qualitative assessment for MP using orthogonal views.** Results on two different input noise-levels (NLs): NL0 (High SNR, same NL as training-set), NL3 (Low SNR, highest simulated NL). SSIM values for each slice embedded.

the removal of artifacts (Figure 5 (a2-a3 and a7-a8)). The quantitative evaluation across all the OOD-noisy test data (Figure 6) shows that the performance of all the baseline methods suffers substantially at higher degradation levels, emphasizing the improved robustness due to the proposed GGD-based uncertainty model. Table 1 shows that with increasing NLs, the mean of $\hat{\beta}_{ij}$s decreases, explained in 4.1.

Qualitative and Qualitative Evaluations for MP. We use the following state-of-the-art methods for comparison: (i) **U-Net** as used in [3, 4] that is adapted to work with unimodal input (T1w) and predicts the unimodal output (T2w). (ii) **CycleGAN** from [6]. Here we used a U-Net for the generator (as it led to better performance). (iii) We improve the robustness of U-Net (from [3, 4]) to OOD-noisy noisy input data by employing dropouts (DO) during both training and inference, called **U-Net+DO**. (iv) Similarly, we employ dropouts for CycleGAN (from [6]) giving us **CycleGAN+DO**. We perform multiple forward passes and take the mean to get the final outputs of models with dropouts activated at inference. Both U-Net and U-Net+DO are trained using the pixel-wise ℓ_1 loss; CycleGAN and CycleGAN+DO use an additional adversarial loss with cycle consistency as in [6].

Figure 7 shows the MP orthogonal views through the 3D output obtained from our method as well as the best performing baselines. U-Net+DO and CycleGAN+DO show improved performance in comparison to the non-DO counterparts as evident from Figure 9. Similar to QE, at NL0, outputs of all methods are comparable. However, at higher NLs, the proposed method outperforms all the baselines substantially, across all metrics (Figure 7 and 9). Figure 8 shows the predicted images for a representative test axial input slice along with the residuals for the proposed model, U-Net+DO, and CycleGAN+DO at two different NLs (NL0 and NL3). At NL0, the predicted images from all the methods (Figure 8 (a2) – (a4) appear closer in structure and con-

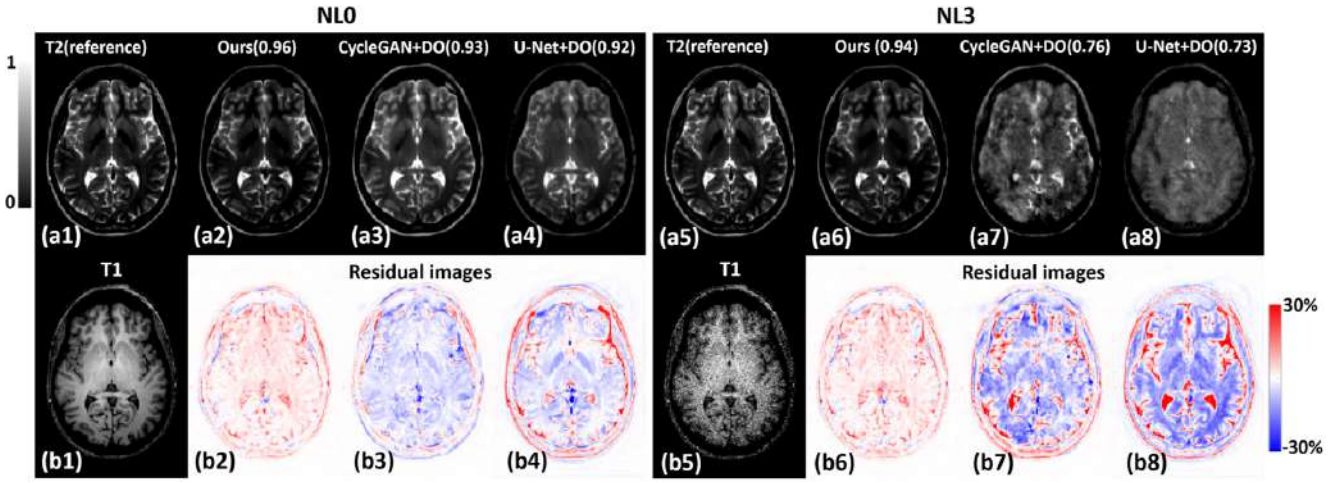

Figure 8. **Results for MP (axial view).** Predicted T2w images with input T1w images at two different noise-levels (NLs), **(b1):** NL0 (high SNR, same NL as training-set), and **(b5):** NL3 (low SNR, highest simulated NL). **(a2)–(a4) and (a6)–(a8):** Predicted (T2w) for all the methods at NL0 and NL3, respectively. **(b2)–(b4) and (b6)–(b8):** Corresponding residuals. SSIM values are indicated within paranthesis.

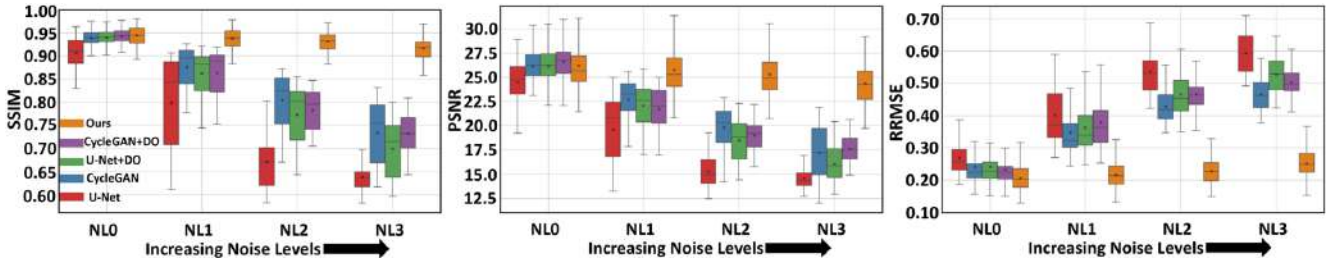

Figure 9. **Quantitative assessment for MP.** SSIM, PSNR, and RRMSE values for all the methods at multiple noise-levels (NL0 to NL3). At each NL, 70 mid-brain slices from each subject (100 test-subjects) were evaluated (i.e., 7000 slices).

trast to the ground truth (Figure 8 (a1)). Our method shows the least residual (Figure 8 (b2)) compared to that of CycleGAN+DO (b3) and U-Net+DO (b4). At NL3, our method (Figure 8 (a6)) outperforms both CycleGAN+DO (Figure 8 (a7)) and U-Net+DO (Figure 8 (a8)) in terms of visual quality as well as SSIM values. All baselines show poor synthesis of structure, contrast, and limited removal of noise. Our method removes significant amount of noise, recover structure, contrast and texture, and is closer in appearance to the reference image (Figure 8 (a5)).

4.1. Uncertainty quantification

Along with improved accuracy of the predicted images, we demonstrate the efficacy of estimating uncertainty maps, ($\hat{\sigma}$, $\hat{\sigma}_{\text{aleatoric}}$, $\hat{\sigma}_{\text{epistemic}}$). Uncertainty maps are derived from the outputs of the network: $\hat{x}^{\text{O}}$, $\hat{\alpha}$, and $\hat{\beta}$ as described in Equation 14-15. Scatter plots in Figure 11 show that for both QE and MP, the residuals between the predictions and the ground-truths correlate positively with the uncertainty obtained from our framework, i.e., high overall residuals correspond to high overall uncertainty for the image. Therefore, higher overall uncertainty may be used as a proxy to infer about increased imperfections in reconstruction/synthesis. Interestingly, Figure 11 also show that high residuals correlate negatively with predicted shape-parameter ($\hat{\beta}$) that is consistent with the fact that lower shape-parameter values correspond to heavy-tailed distributions for residuals, i.e., outliers.

We also employ a testing scenario for MP, where the incoming test subject has an unseen pathology (a lesion). In this case too, we use the network which was trained on healthy individuals only. Figures 10 (a1) and (a2) show the input (T1w) and the reference (T2w) image with simulated lesions. The sub-figures (i) and (ii) show results on two representative slices. The predicted image ($\hat{x}^{\text{T2}}$, Figure 10 (a3)), the scale parameter of the GGD ($\hat{\alpha}$, Figure 10 (b1)), and the corresponding shape parameter ($\hat{\beta}$, Figure 10 (b2)), are the outputs of the network. The derived uncertainty map is shown in Figure 10 (b3), obtained using Equations 13 – 15. For the background region, both the scale and shape maps show values that are less spatially varying. While the scale map ($\hat{\alpha}$) captures edges in the image at a coarser level, the shape map ($\hat{\beta}$) captures subtle features in the image. In all the slices intensity values in the $\hat{\beta}$ image, within the lesion region are significantly lower than other regions, indicating outliers. We observe that the absolute residual image has peak values in and around the lesion. This is expected as the training data is devoid of such pathological cases.

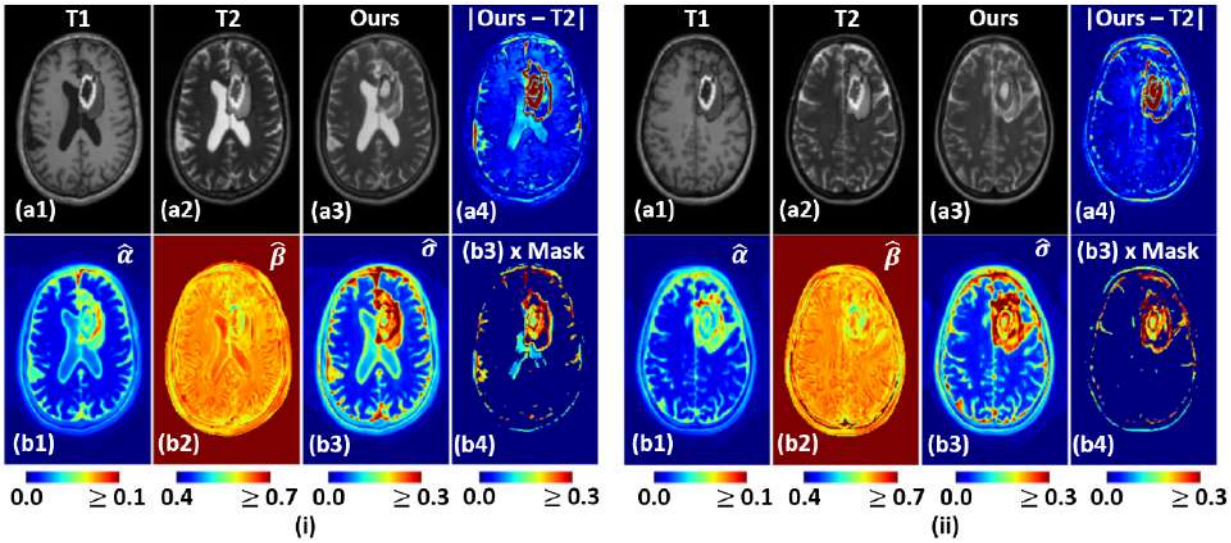

Figure 10. **Uncertainty quantification on OOD-noisy data (synthetic lesion added to MP test data). Subfigures (i) – (ii)** show results from two representative slices. **(a1)-(a2):** Input T1w (x^{T1}) and the corresponding reference T2w (x^{T2}). **(a3)-(a4):** Predicted image ($\hat{x}^{\text{T2}}$) and absolute error map. **(b1)-(b2):** The learned scaling ($\hat{\alpha}$) and shape ($\hat{\beta}$) parameter maps. **(b3):** Uncertainty map ($\hat{\sigma}$). **(b4):** Masked aleatoric uncertainty map (refer Section 4.1).

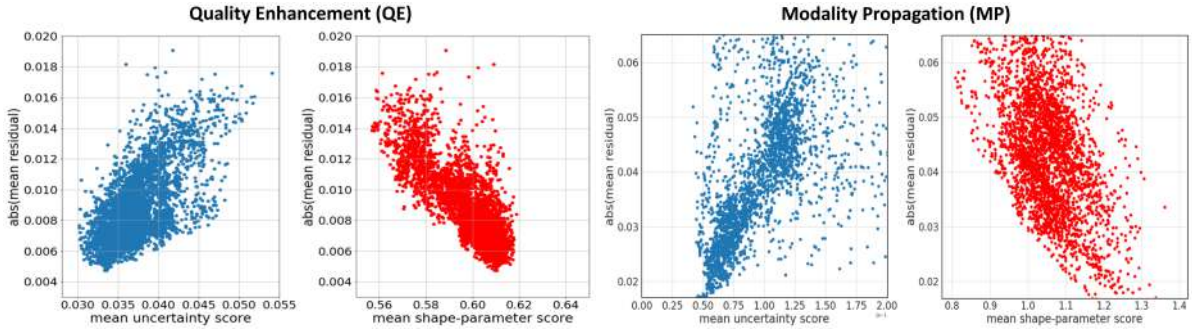

Figure 11. **Correlation between residual, uncertainty, and shape-parameter. Data from all NLs.** Each point indicates an image, uncertainty score and shape-parameter score is calculated for an image by taking mean of $\hat{\sigma}$ and $\hat{\beta}$ over all the pixels.

The regions with "high" intensity values within the uncertainty map ($\hat{\sigma}$, Figure 10 (b3)) show high residual values as well (Figure 10 (a4)). In addition to showing peak values of the uncertainty maps in regions with high residual error, the uncertainty map is able to capture minor variations in the predicted images, especially at the skull boundary and gyri. To make the uncertainty map clinically relevant, we identify "high" residual values as the ones that are above a pre-defined threshold, say τ. The value of τ is fixed as follows: we choose τ to be slightly above the standard deviation of the highest noise-level at which the network provided reasonably accurate reconstructed images (NL3). In this paper, we choose $\tau = 0.17$, which is slightly above the estimated standard deviation of the noise at NL2. Figure 10 (b4) shows the product of the uncertainty map ($\hat{\sigma}$) and the binary mask obtained by thresholding the absolute error map at $\tau = 0.17$. Clearly, the resultant uncertainty map (Figure 10 (b4)) reflects pixels with the highest predicted uncertainty. Hence, uncertainty map can be a proxy for residual map (*that is not available at inference*).

Conclusion. In this work, we proposed a GAN-based framework with adaptive quasi-norm loss functions for improved robustness to unseen perturbations on test data. We demonstrated the efficacy of our network on two key applications arising in the field of medical imaging, namely undersampled MRI reconstruction and modality propagation. Enhanced output produced using this method after post-processing can serve in clinical decision making based. We compared our network with state-of-the-art networks for both applications and found that the performance of our network is substantially better (both qualitatively and quantitatively) at high noise levels, and comparable at noise levels similar to the training set. While our experiments with perturbations related to scanners (noise level study) showed that our framework can generate predicted images with low residual, the experiments related to physiological perturbations (lesion study) showed that our uncertainty maps can be proxy for the residual maps.

References

[1] JT Barron. A general and adaptive robust loss function. In *IEEE Conf Comp Vis Patt Recog. (CVPR)*, pages 4331–4339, 2019. 2

[2] MJ Black and A Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *Int J Comp Vis*, 19(1):57–91, 1996. 2

[3] A Chartsias, T Joyce, M Giuffrida, and S Tsaftaris. Multimodal MR synthesis via modality-invariant latent representation. *IEEE Trans Med Imag.*, 37(3):803–814, 2017. 6

[4] KT Chen, E Gong, M Carvalho, J Xu, A Boumis, M Khalighi, K Poston, S Sha, M Greicius, E Mormino, et al. Ultra-low-dose 18F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. *Radiol.*, 290:649, 2019. 4, 6

[5] D Cox and DV Hinkley. *Theoretical statistics*. CRC Press, 1979. 2

[6] S Dar, M Yurt, L Karacan, A Erdem, E Erdem, and T Çukur. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Trans Med Imag.*, 38(10):2375–2388, 2019. 1, 3, 6

[7] Y Gal and Z Ghahramani. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *Int Conf Mach Lear. (ICML)*, pages 1050–1059, 2016. 4

[8] I Goodfellow. NIPS 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*, 2016. 3

[9] H Gudbjartsson and S Patz. The Rician distribution of noisy MRI data. *Magn Res Med.*, 34(6):910–914, 1995. 2

[10] T Hastie, R Tibshirani, and M Wainwright. *Statistical learning with sparsity: the lasso and generalizations*. CRC press, 2015. 2

[11] YC Hsu, Y Shen, H Jin, and Z Kira. Generalized odin: Detecting out-of-distribution image without learning from out-of-distribution data. In *IEEE Conf Comp Vis Patt Recog. (CVPR)*, June 2020. 1

[12] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *European Conf Comp Vis.(ECCV)*, pages 172–189, 2018. 1, 2

[13] PJ Huber et al. The 1972 Wald lecture robust statistics: A review. *The Annals of Mathematical Statistics*, 43(4):1041–1067, 1972. 2

[14] P Isola, JY Zhu, T Zhou, and AA Efros. Image-to-image translation with conditional adversarial networks. In *IEEE Conf Comp Vis Patt Recog. (CVPR)*, pages 1125–1134, 2017. 5

[15] A Kendall and Y Gal. What uncertainties do we need in Bayesian deep learning for computer vision? In *Adv Neural Info Proc Syst.*, pages 5574–5584, 2017. 3

[16] D Kingma and J Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4

[17] A Klein, J Andersson, B Ardekani, J Ashburner, B Avants, MC Chiang, GE Christensen, D Collins, J Gee, P Hellier, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage*, 46(3):786–802, 2009. 4

[18] C Ledig, L Theis, F Huszár, J Caballero, A Cunningham, A Acosta, A Aitken, A Tejani, J Totz, Z Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE Conf Comp Vis Patt Recog. (CVPR)*, pages 4681–4690, 2017. 1

[19] K Lee, K Lee, H Lee, and J Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In *Adv Neural Info Proc Sys.*, pages 7167–7177, 2018. 1

[20] J Liu and D Saloner. Accelerated MRI with CIRcular cartesian UnderSampling (CIRCUS): a variable density Cartesian sampling strategy for compressed sensing and parallel imaging. *Quant Imag Med Surgery.*, 4(1):57, 2014. 2, 5

[21] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In *Adv Neural Info Proc Sys.*, pages 700–708, 2017. 1, 2

[22] M Lustig, D Donoho, and JM Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Res Med.*, 58(6):1182–1195, 2007. 4

[23] R Mezrich. A perspective on k-space. *Radiol.*, 195(2):297–315, 1995. 5

[24] SM Moosavi-Dezfooli, A Fawzi, O Fawzi, and P Frossard. Universal adversarial perturbations. In *IEEE Conf Comp Vis Patt Recog. (CVPR)*, pages 1765–1773, 2017. 1

[25] David Moratal, A Vallés-Luch, Luis Martí-Bonmatí, and Marijn E Brummer. k-space tutorial: an MRI educational tool for a better understanding of k-space. *Biomed Imag and Interven J.*, 4(1), 2008. 1

[26] T Nair, D Precup, D Arnold, and T Arbel. Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. *Med Imag Anal.*, 59:101557, 2020. 1

[27] T Nair, D Precup, D Arnold, and T Arbel. Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. *Med Imag Anal.*, 59:101557, 2020. 2

[28] D Nie, R Trullo, J Lian, L Wang, C Petitjean, S Ruan, Q Wang, and D Shen. Medical image synthesis with deep convolutional adversarial networks. *IEEE Trans Biomed Engg.*, 65(12):2720–2730, 2018. 1, 3

[29] D Ravì, A Szczotka, S Pereira, and T Vercauteren. Adversarial training with cycle consistency for unsupervised super-resolution in endomicroscopy. *Med Imag Anal.*, 53:123–131, 2019. 2

[30] PM Robson, AK Grant, AJ Madhuranthakam, R Lattanzi, DK Sodickson, and CA McKenzie. Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions. *Magn Res Med.*, 60(4):895–907, 2008. 4

[31] O Ronneberger, P Fischer, and T Brox. U-net: Convolutional networks for biomedical image segmentation. In *Int Conf Med Imag Comput Comput-Assist Intervention. (MICCAI)*, pages 234–241. Springer, 2015. 3

[32] J Schlemper, J Caballero, J Hajnal, A Price, and D Rueckert. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE Trans Med Imag.*, 37(2):491–503, 2017. 1

[33] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Ann Rev Biomed Engg.*, 19:221–248, 2017. 1, 4

[34] J Sjölund, A Eklund, E Özarslan, M Herberthson, M Bånkestad, and H Knutsson. Bayesian uncertainty quantification in linear models for diffusion MRI. *NeuroImage*, 175:272–285, 2018. 1

[35] Viswanath P. Sudarshan, Uddeshya Upadhyay, Gary F. Egan, Zhaolin Chen, and Suyash P. Awate. Towards lower-dose pet using physics-based uncertainty-aware multimodal learning with robustness to out-of-distribution data. *Medical Image Analysis*, 73, 2021. 1

[36] D Sun, S Roth, and M Black. Secrets of optical flow estimation and their principles. In *IEEE Conf Comp Vis Patt Recog.*, pages 2432–2439. IEEE, 2010. 2

[37] R Tanno, D Worrall, A Ghosh, E Kaden, S Sotiropoulos, A Criminisi, and D Alexander. Bayesian image quality transfer with CNNs: exploring uncertainty in dMRI super-resolution. In *Int Conf Med Image Comput Computer-Assist Intervention. (MICCAI)*, pages 611–619. Springer, 2017. 1, 2

[38] Uddeshya Upadhyay and Suyash P. Awate. A mixed-supervision multilevel gan framework for image quality enhancement. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, 2019. 1

[39] Uddeshya Upadhyay and Suyash P. Awate. Robust super-resolution gan, with manifold-based and perception loss. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019. 1

[40] Uddeshya Upadhyay, Viswanath P. Sudarshan, and Suyash P. Awate. Quest for medisyn: Quasi-norm based uncertainty estimation for medical image synthesis. In *ICML 2020 Workshop on Uncertainty & Robustness in Deep Learning*, 2020. 1

[41] Patrick Virtue and Michael Lustig. The empirical effect of Gaussian noise in undersampled MRI reconstruction. *Tomography*, 3(4):211, 2017. 2, 4

[42] G Wang, W Li, M Aertsen, J Deprest, S Ourselin, and T Vercauteren. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomput.*, 338:34–45, 2019. 2

[43] Z Wang, A Bovik, H Sheikh, and E Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Imag Proc.*, 13(4):600–12, 2004. 5

[44] Z Zhang, A Romero, MJ Muckley, P Vincent, L Yang, and M Drozdzal. Reducing uncertainty in undersampled MRI reconstruction with active acquisition. In *IEEE Conf Comp Vis Patt Recog. (CVPR)*, pages 2049–2058, 2019. 1, 5

[45] JY Zhu, T Park, P Isola, and AA Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE Int Conf Comp Vis. (ICCV)*, Oct 2017. 2

[46] Y Zhu, W Shen, F Cheng, C Jin, and G Cao. Removal of high density Gaussian noise in compressed sensing MRI reconstruction through modified total variation image denoising method. *Heliyon*, 6(3):e03680, 2020. 2, 4