



OPEN

Conditional generative learning for medical image imputation

Ragheb Raad¹, Deep Ray², Bino Varghese³, Darryl Hwang³, Inderbir Gill⁴,
Vinay Duddalwar³ & Assad A. Oberai¹✉

Image imputation refers to the task of generating a type of medical image given images of another type. This task becomes challenging when the difference between the available images, and the image to be imputed is large. In this manuscript, one such application is considered. It is derived from the dynamic contrast enhanced computed tomography (CECT) imaging of the kidneys: given an incomplete sequence of three CECT images, we are required to impute the missing image. This task is posed as one of probabilistic inference and a generative algorithm to generate samples of the imputed image, conditioned on the available images, is developed, trained, and tested. The output of this algorithm is the “best guess” of the imputed image, and a pixel-wise image of variance in the imputation. It is demonstrated that this best guess is more accurate than those generated by other, deterministic deep-learning based algorithms, including ones which utilize additional information and more complex loss terms. It is also shown that the pixel-wise variance image, which quantifies the confidence in the reconstruction, can be used to determine whether the result of the imputation meets a specified accuracy threshold and is therefore appropriate for a downstream task.

Image imputation or image synthesis refers to the task of generating or synthesizing missing images using available data, which is often in the form of other types of images. In medical imaging, these techniques address tasks like generating images of one type (FLAIR MR image, for example) given images of a different type (T2 MR image), generating missing slices in a three-dimensional stack of slices of an organ, and generating an image at a specific time-point in a temporal sequence of images obtained from contrast-enhanced imaging modalities. Generally speaking, if the available images are “close” to the missing image, the image synthesis/imputation task is easier. For example, if the temporal sequence includes ≈ 50 images, wherein the sequential changes are small, then the missing image is easily approximated by a weighted sum of its neighbors in the sequence. On the other hand, if the available information is sparse, the image imputation task is challenging. For example, when the entire temporal sequence contains only a few images (say 4) and the difference between each image is significant, relying on neighboring images alone to infer a missing image is not an option. In this case, the image imputation algorithm must learn the complex dependencies between images in the sequence from an independent set of training data, and then apply this knowledge to the case of interest. These types of image imputation problems are the focus of this manuscript.

In particular, we consider the dynamic contrast enhanced computed tomography (CECT) imaging of the kidneys. In this modality, an intravenous contrast agent is injected into the subject and CT images are acquired at different time-points leading to pre-contrast, corticomedullary (30–40 s after injection), nephrographic (100 s after injection), and excretory (5–10 minutes after injection) phase images¹. A complete sequence consists of all images at all four time-points. In some cases, one or more these images may be missing and may need to be imputed. For instance, subject motion during the exam could blur some images rendering them of little clinical value, or, under a clinical protocol a subject with a renal tumor might undergo CECT imaging where the pre-contrast image is not recorded. However, each image in the CECT sequence is important and has clinical value. For example, the pre-contrast, corticomedullary and nephrographic images are all required to evaluate intensity enhancement and washout within the tumor, kidney and other organs, which is in turn used in evaluation and diagnosis². Also, in the excretory phase, the renal pelvis is clearly visualized and its location relative to the tumor can be determined. This information is useful to a surgeon performing nephrectomy to remove the tumor^{3,4}.

Deep learning has found significant applications in tackling image synthesis/imputation problems of the type described above. Specifically, algorithms based on a U-net architecture that map images of one type to another

¹Aerospace and Mechanical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089, USA. ²Department of Mathematics, University of Maryland, College Park, MD 20742, USA. ³Radiology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA. ⁴Urology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA. ✉email: aoberai@usc.edu

via convolutional layers have been remarkably successful. Among these are a class of algorithms that typically employ adversarial learning and are closely related to generative adversarial networks (GANs)⁵. This includes the PIX2PIX algorithm which performs image transformation using pair-wise image data and adversarial loss augmented with a reconstruction term⁶. In CycleGAN⁷ these ideas are extended by adding a cycle consistency loss and removing the stringent requirement of working with pairwise images. Algorithms like the StarGAN⁸ and RadialGAN⁹ extend these ideas to image transformation across multiple domains with a single generator network that uses a specific code which carries information about each domain. CollaGAN¹⁰ provides a similar mapping across multiple domains by relying on multiple consistency losses. Recently, algorithms based on the transformer architecture have also been successfully applied to problems in medical imaging^{11,12}.

There has been a growing realization that algorithms for medical imaging should also provide some estimates of the uncertainty in their outputs. Stated simply, this means that given an image of one type they should account for the fact that there may be an ensemble of images that are consistent with the given input image. Thus they should either generate this ensemble or quantify the heterogeneity within it. This has been accomplished for images translation of one type to another¹³, de-noising of MR images¹⁴, and enhancement of the brain MR images¹⁵. Along these lines, in this manuscript, for the case of imputing CECT images, we recognize that there may be whole collection of likely complete sequences that are consistent with a given incomplete sequence. The algorithm presented in this manuscript generates this ensemble and produces a “mean” image that is shown to be more accurate than other comparable methods. It also produces a standard deviation image that quantifies the uncertainty in the prediction thus providing a measure of confidence to the user.

In the probabilistic framework described in this work, we treat both the incomplete and complete image sequences as random vectors. Then, using data which consists of complete sequences and their incomplete counterparts (generated by decimating an image at random), we train a conditional generative adversarial (cGAN) network^{16–18}. This network takes samples from the joint distribution of two random vectors and learns to efficiently sample from the conditional distribution. That is, given an instance of one of these vectors, it generates samples of the other vector conditioned on that instance (see Fig. 1). In our case, the “instance” is the incomplete CECT image sequence and the samples generated are the complete CECT sequences that are consistent with this incomplete sequence. From these samples we extract the desired imputed image, and compute the pixel-wise mean and standard deviation. The mean imputed image provides our best guess for the missing image, and the standard deviation image quantifies the uncertainty in our prediction. Through rigorous testing we show that the mean image is generally more accurate than images produced by methods that do not account for the probabilistic nature of the problem. We also demonstrate that the standard deviation image is correlated with the error in the imputation and can be used to quantify the confidence in the imputed image.

Deep learning-based image imputation techniques have recently been used for imputing and synthesizing CT images. This includes generating CT images for data augmentation to eventually improve the performance of a CT-image based classifier^{19–21}. It also includes algorithms for generating CECT images at a single time point for the lungs²², and the kidneys²³, where the latter study uses the concept of neural transfer for improved performance. Recently, several GAN-based approaches were implemented and tested for imputing renal CECT images at different time-points²⁴. The approaches tested (in the present work) include several standard algorithms and two novel methods, ReMIC and DiagnosisGAN, which were shown to be the most accurate. In ReMIC²⁵,

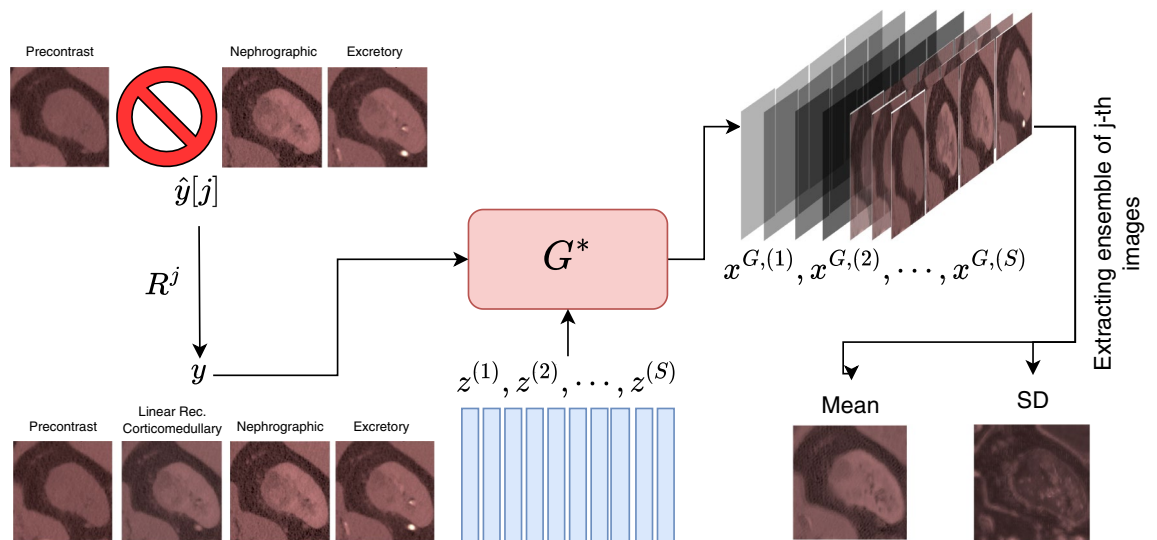


Figure 1. A schematic diagram of the imputation algorithm. In this illustration, we have assumed that the corticomedullary image is missing from the sequence ($j = 2$). The first step involves constructing a linear regression-based guess of the missing image through the operator R^j . This approximate sequence, and a sequence of random vectors $z^{(i)}$, are used as input to the fully-trained generator, G^* . The generator produces an ensemble of likely complete sequences wherein each member is denoted by $x^{G,(i)}$. These are used to calculate the pixel-wise mean and standard deviation (SD) images. The best guess to the imputed image is extracted from the former and the latter is used to determine the confidence in the imputation.

a representational disentanglement scheme for multi-domain image completion was used to improve the performance of the algorithm, whereas in DiagnosisGAN²⁴ in addition to the CECT images themselves, other sources of information, like segmentation mask for the tumor, and the knowledge cancer sub-type, were used to improve the performance of the method. In the “Results” section of this manuscript we compare our algorithm with these methods and conclude that our method is more accurate, and at the same time provides estimates of confidence in the imputation task.

We remark that in an earlier work²⁶ we presented a probabilistic method for imputing CECT images where we utilized GANs to learn the prior distribution^{27,28} of a complete sequence of CECT images. This prior was combined with a likelihood term driven by a measured incomplete sequence to set up a Bayesian estimate for the probability distribution of the complete sequence. This inference problem was then solved by advanced Markov-Chain Monte Carlo (MCMC) methods. In the present approach, in contrast to this, we utilize a conditional GAN to directly learn and sample from the conditional distribution. This leads to an algorithm that is much more efficient. As a result, it can be applied to reconstructing slices that include the tumor, the kidney, and the surrounding tissue, whereas the earlier work was limited to segmented images of only the tumor. Thus the work presented in the study differs in methodology (cGAN to directly learn the conditional distribution in contrast to GAN to learn the prior) and is more numerically efficient and widely applicable.

The remainder of this manuscript is organized as follows. In the following section, we present the results obtained by applying our algorithm to incomplete sequence of CECT images. Thereafter, we discuss these results and their medical relevance. Finally, in the “Methods” Section we describe our algorithm in detail.

Results

Patient data

The study population consists of patients who had renal masses diagnosed on abdominal CECT scans and underwent resection at USC between May 2007 and September 2018. The pathology of the masses was confirmed after resection, and the patients were identified through a query of a surgical database. Patients without evaluable preoperative imaging or missing any of the four time-points of the CECT study were excluded. The final cohort included 370 patients, and three-dimensional regions of interest of the renal masses were manually segmented by two senior radiologists using Synapse 3D software²⁹. The original images were 512×512 pixels with a pixel size of 0.9765 mm in each direction. These images were cropped to a size of 128×128 by selecting a square centered on the tumor centroid. From the total data, 296 subjects were used for training (around 80%), 37 subjects were used for validation (around 10%), and 37 subjects were used for testing the algorithm (around 10%). The training data was augmented by rotating each image by ± 10 , ± 20 and ± 30 degrees and by shifting it in the horizontal and vertical directions. This yielded at most eleven images for each original image. Data from Hounsfield units (within the range $(-3024, 3071)$) was normalized to the range $(-1, 1)$ by first clipping values below -150 and above 1000 and then linearly transforming the Hounsfield scale to the normalized range.

Image imputation results

Figures 2, 3 and 4 display the results of the image imputation algorithm for six subjects selected from among the 37 test subjects. These subjects were selected to highlight the diversity of the type of renal CECT images that the algorithm can be applied to. Results for all 37 subjects are available in³⁰. Each figure represents results from two subjects, and for each subject, the first row displays true images, the second row contains the mean imputed images generated by the cGAN, and the third row contains the images of the pixel-wise standard deviation generated by the cGAN. The columns represent the four time-points of the CECT exam. The mean imputed image may be interpreted as the “best guess” generated by the cGAN, while the standard deviation image represents the spatial distribution in the uncertainty in this imputation.

To generate the imputed images, the algorithm assumes that the true image for a specific time-point is missing and needs to be predicted by making use of the images at the three other time-points. For example, for a given subject, images in the first column in rows 2 and 3 are generated by assuming that the true image for the pre-contrast time-point is missing and needs to be imputed by using the true images at the other three time-points. Similarly, images in the second, third, and fourth columns are obtained by imputing missing images for the corticomedullary, nephrographic, and excretory time-points, respectively, while making use of the true images from complementary set of time-points. Therefore, each figure demonstrates the capacity of the cGAN algorithm to impute images for the four distinct CECT time-points.

In Table 1, we perform a quantitative comparison of the performance on the cGAN with an improved variant of the popular PIX2PIX⁶ algorithm (described in the “Methods” Section) which may be thought of as a deterministic version of the cGAN approach. We implement this method, training and testing it using the same image data used for the cGAN. From this Table it is clear that the cGAN method outperforms the PIX2PIX algorithm for all time-points. It is also interesting to note that it is most accurate in imputing pre-contrast images and least accurate in imputing corticomedullary images.

In the same table, we compare the performance on the cGAN and PIX2PIX algorithms implemented in our study with the performance of the two top-performing algorithms reported in a recent CECT renal image imputation study²⁴. These algorithms are

1. DiagnosisGAN²⁴, which is a generative adversarial algorithm like the cGAN. However unlike the cGAN approach, it does not provide any information regarding the confidence in the imputation and further it requires segmented tumor images to improve its performance.
2. ReMIC²⁵, which relies on a representational disentanglement scheme for multi-domain image completion. When compared with the cGAN approach, the ReMIC algorithm is more complex and relies on different

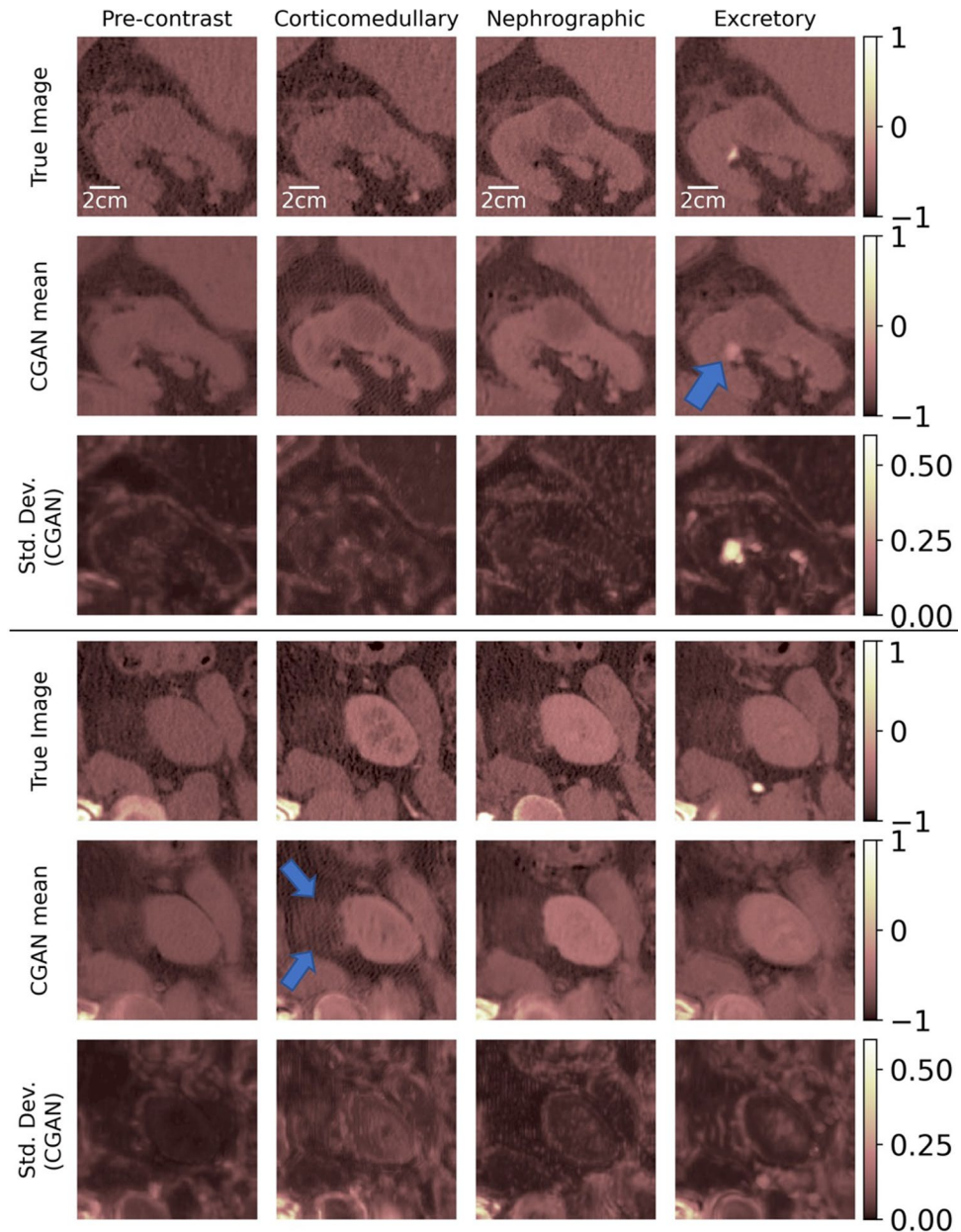


Figure 2. True and imputed images for Subjects 1 and 2.

types of losses which include image and latent domain consistency losses, adversarial losses, and reconstruction losses. In contrast to the cGAN approach, the ReMIC approach is also a deterministic approach and does not yield any information regarding confidence in the imputed images.

The interested reader is referred to the “Methods” section and the original references for further details on these methods. From Table 1, which reports the values of the normalized mean squared error (NMSE), the structural similarity index measure (SSIM) and the peak signal-to-noise ratio (PSNR) averaged over all time-points and all subjects, we conclude that for both these metrics the cGAN approach is the most accurate.

Standard deviation and uncertainty

While the cGAN-based method provides accurate imputation results, its distinguishing feature is the ability to draw an ensemble from the posterior distribution rather than just a single most likely sample. This ensemble can be utilized to compute statistics that offer insight into the confidence in the imputed image. We demonstrate this by computing the estimated pixel-wise standard deviation in the imputed images in Figures 2, 3 and 4 (last row). These images offer a spatial depiction of the level of uncertainty present in the imputed results. The higher the value of standard deviation in zone of pixels, the larger the ensemble variation and uncertainty in that zone.

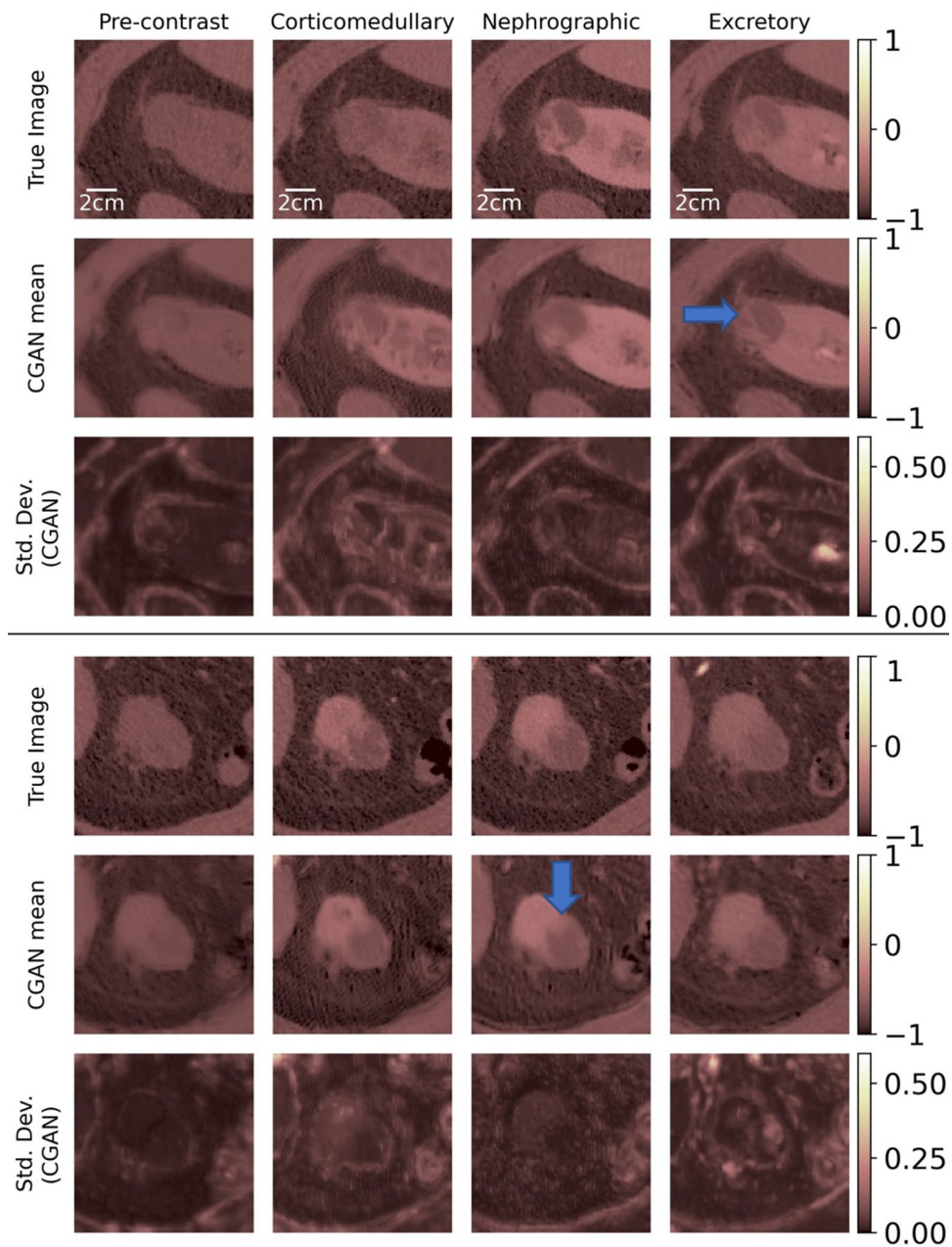


Figure 3. True and imputed images for Subjects 3 and 4.

We investigate the relation between the standard deviation computed by our algorithm and the true error of the image imputed by the cGAN. If a positive correlation is identified, then the standard deviation may be utilized as an indicator of the error in the imputed image, thus serving as a powerful tool for the end-user to eliminate imputed images that are likely to be inaccurate. In particular, we determine whether the total value of standard deviation (summed over all pixels) can be used to classify a given imputed image as being acceptable. We set a threshold of NMSE = 0.1 as a criterion and bin each imputed image into “acceptable” and “not acceptable” classes. Thereafter, we quantify the performance of the total standard deviation as a surrogate for performing this classification. The results are summarized in the receiver operating characteristic (ROC) curve shown in Fig. 5. For this curve, the corresponding the area under the curve (AUC) value is 0.8825. Based on this we conclude then that the total standard deviation is predictive of whether a given imputation is sufficiently accurate.

Discussion

In images for Subject 1 (Fig. 2) we observe a well marginated tumor which is predominantly endophytic (growing within the kidney rather than protruding out) and has density similar to that of soft tissue. This tumor was a surgically proven renal cell carcinoma. The generated images demonstrate the lesion and the margins at each time-point. The subtle nodular peripheral enhancement noted in the true corticomedullary and nephrographic

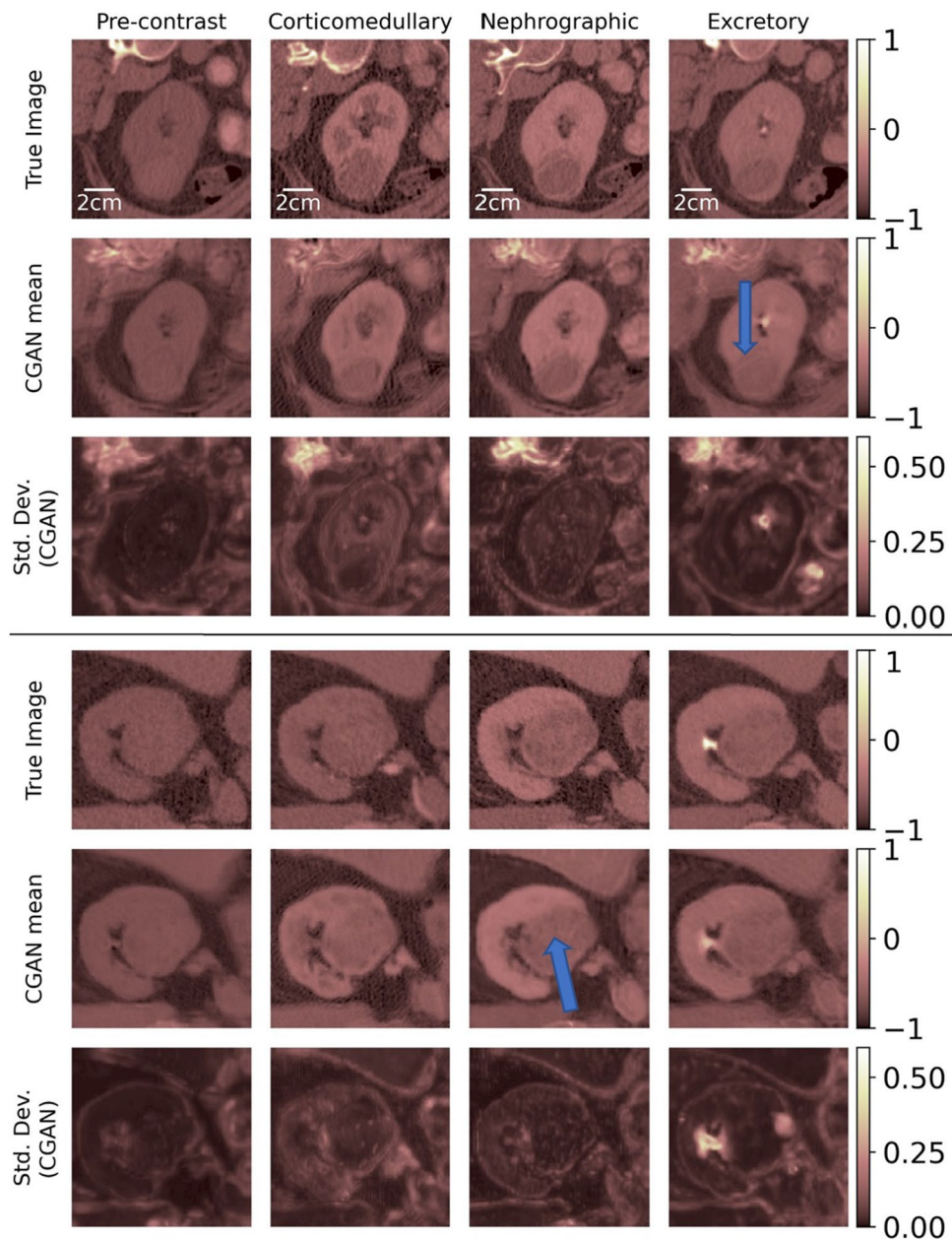


Figure 4. True and imputed images for Subjects 5 and 6.

phases is also clearly seen in the generated images. In addition, in the imputed excretory phase image the location of the adjacent calyx (arrow) where urine collects from that portion of the kidney is anatomically consistent with the true image. We note that the depiction of the tumor, its margins and its relationship with the calyx and other intra-renal structures are important for surgical planning. All these features are well reproduced in the imputed images.

In images for Subject 2 (Fig. 2), we observe a patient with an exophytic (protruding out of the kidney) tumor which was an angiomyolipoma. The imputed images reproduce the density of the tumor and its relationship with the kidney very well. The depiction of the density of the tumor, viz. a density that is predominantly the same as the density of fat tissue, enables the accurate diagnosis of this specific tumor via imaging.

The tumor in Subject 3 is a complex cystic (with roughly the density of a fluid) mass with multiple irregular nodular enhancing septations within it (Fig. 3). The complexity, nodularity of the septations are important diagnostic features used by radiologists to characterize cystic tumors using the Bosniak scoring system^{31–34}. This was a Bosniak 4 tumor indicating (and proven) malignant cystic renal carcinoma. We note that these important features are reproduced in the imputed images.

Subject 4 is another example of a complex cystic tumor where the nodular peripheral enhancement from the anterior margin of the tumor leads to the radiologic characterization of this tumor as cystic renal carcinoma.

	Pre-contrast	Corticomedullary	Nephrographic	Excretory	Combined
SSIM					
cGAN	0.7262	0.6232	0.6353	0.6607	0.6614 (0.0845)
PIX2PIX	0.7022	0.6057	0.5992	0.6111	0.6295 (0.0927)
DiagnosisGAN ²⁴	0.6105	0.6549	0.6408	0.6210	0.6243 (0.1045)
ReMIC ²⁴	0.6260	0.6453	0.6252	0.6368	0.6333 (0.1012)
PSNR					
cGAN	28.43	26.66	26.95	25.16	26.80 (2.91)
PIX2PIX	27.25	25.84	26.02	23.89	25.75 (3.05)
DiagnosisGAN ²⁴	–	–	–	–	20.07 (2.06)
ReMIC ²⁴	–	–	–	–	20.33 (2.02)
NMSE					
cGAN	0.1043	0.1302	0.1297	0.1587	0.1307
PIX2PIX	0.1184	0.1472	0.1457	0.1832	0.1486

Table 1. Performance of the cGAN and PIX2PIX algorithms on test data (values averaged over 37 subjects). Also included is the performance of DiagnosisGAN and ReMIC²⁴, as reported by the authors. The metrics considered include Structural Similarity Index Measure (SSIM), Peak Signal-to-Noise Ratio (PSNR) and Normalized Mean Square Error (NMSE). Note that the cGAN and PIX2PIX algorithms were trained on different data set than the DiagnosisGAN and ReMIC algorithms. Further, for these two algorithms the authors did not provide PSNR values for each phase and NMSE values for any phase. The best algorithm for each metric for each image type is shown in bold font.

We note that this enhancement is reproduced faithfully in the corticomedullary and nephrographic imputed images (Fig. 3). Subject 5 displays a complex cystic tumor where the thickened margin is well seen in the imputed images (Fig. 4).

Subject 6 displays a tumor that is a predominantly hypo-dense (low density compared to the adjacent normal renal tissue) with multiple internal components (arrow) which are well seen in the imputed images (Fig. 4). This is a specific type of renal carcinoma, viz. papillary renal cell carcinoma where the density of the tumor is key to the diagnosis and is faithfully reproduced in the imputed images.

One of the distinguishing features of the method developed in this study is its ability to produce a distribution of imputed images conditioned on the images available at other time-points. This allows us to compute an image of the pixel-wise standard deviation for each imputed image. At the local level, the standard deviation image (shown in third row for each subject) highlights the regions where the uncertainty in the imputation is high. We note that the regions of high uncertainty tend to be at the interface between the kidney and abdominal cavity, and between other organs and the abdominal cavity.

The total intensity of the standard deviation image also contains useful information. In particular, as shown in Fig. 5 we note that this value is good indicator of whether the NMSE of an imputed image is small or large. We note that it can be used to classify (AUC = 0.88825) whether the NMSE for an imputed image will be above or below a threshold value of 0.1. This is particularly useful for downstream clinical tasks where imputed images with large standard deviation, and hence low-confidence, can be disregarded by the clinician. In addition, the standard deviation images themselves may be used to communicate to the clinician the regions where the uncertainty in the imputation is the largest. If these regions correspond to regions that are important in reaching a clinical decision (the tumor and its vicinity, for example), then a clinician may ascribe a lower confidence to their final decision.

Finally, we note that even though the main feature of our approach is to quantify the confidence in the imputation results, it produces images with less error than other state-of-the-art methods. In Table 1, its performance is compared with an enhanced PIX2PIX approach that uses the same training and test data. For each time-point and in all three metrics we observe that the cGAN based approach incurs smaller errors. In the same table, it is also compared with other leading methods, albeit on different training and test data. Here too, it quite clearly outperforms the other methods, even though these methods utilize additional information (like tumor segmented images) and contain losses that are more complex. We believe that primary reason for the better performance of our approach is its stochastic nature. By virtue of this, the mean imputed image averages over any predictions that are outliers and does not let them strongly influence the final estimate.

Methods

Data for this study was extracted from an Institutional Review Board (IRB) approved Kidney Mass Data and Specimen Collection project. Informed consent for the repository was obtained by the USC IRB consistent with §45 CFR 46.116(f). The study was conducted in accordance with USC policies, IRB policies, and federal regulations. Subject privacy and confidentiality were protected according to applicable HIPAA, and USC IRB policies and procedures.

We begin by introducing some mathematical notation. Define the product space $\mathbb{R}_4^N := \mathbb{R}^{N \times N \times N \times N}$, where $N = 128 \times 128$ is the size/resolution of each image. Let $x = [x_1, x_2, x_3, x_4] \in \Omega_X \subset \mathbb{R}_4^N$ be the sequence of CECT images for a patient at the four time-points. Let $\hat{y}[j] \in \Omega_Y \subset \mathbb{R}_3^N$ be the patient's CECT image sequence with the

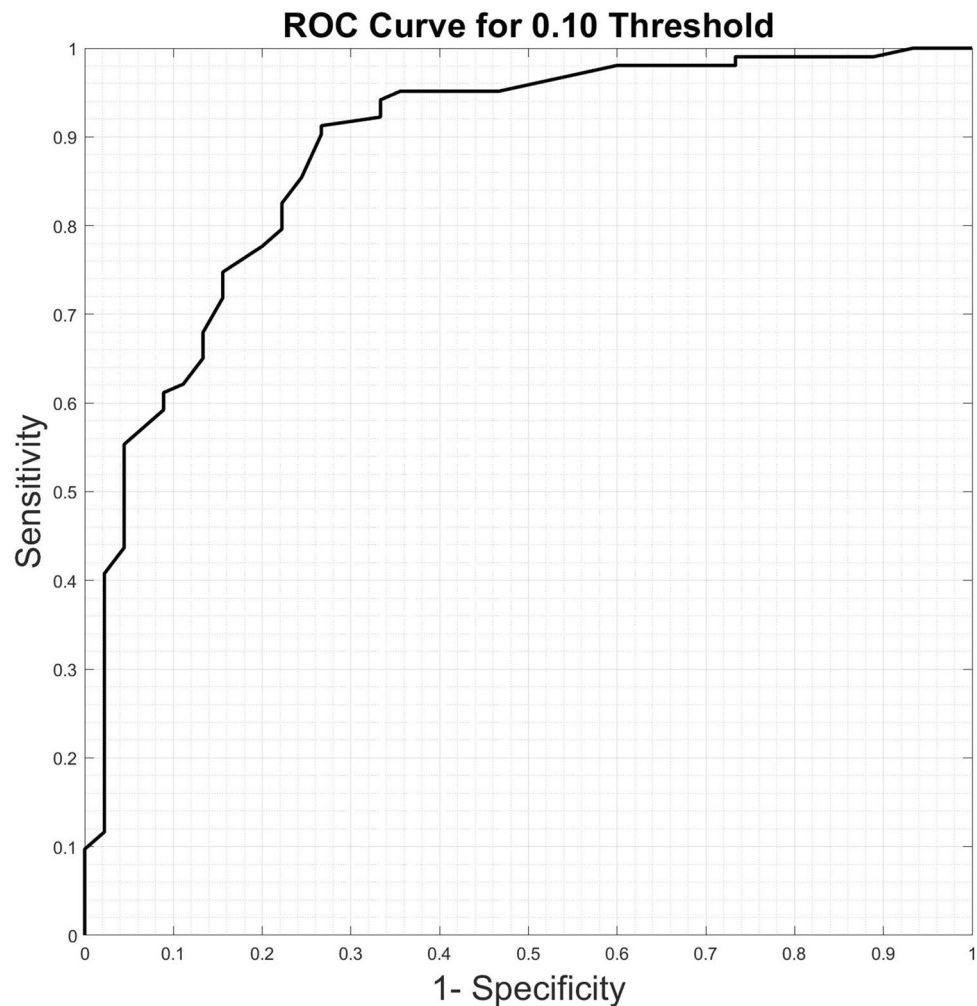


Figure 5. ROC curve for classifying a given imputed image as acceptable.

j -th image in the sequence missing. The missing image is replaced by a simple linear reconstruction using the map $R^j : \Omega_Y \subset \mathbb{R}_4^N \rightarrow \Omega_Y \subset \mathbb{R}_4^N$, with a mapping existing for each $1 \leq j \leq 4$ (see Supplementary Note 1). Note that the linear reconstruction of the missing image only serves as an initial guess, and is typically unable to represent several desired features and intensity variations of the true image. We use the notation $y = [y_1, y_2, y_3, y_4] \in \Omega_Y \subset \mathbb{R}_4^N$ to denote the final measurement where an incomplete sequence has been filled in with the linear approximator. We are interested in finding an appropriate x given the measurement y .

To accommodate for the fact the reconstructed x may not be unique, we consider a Bayesian formulation where sequences x and y are modelled using random variable X and Y , respectively. We are thus interested in finding the conditional probability distribution $P_{X|Y}$ given an incomplete measurement $Y = y$, and generating samples from this distribution. Furthermore, we want to learn this distribution by only working with a finite set of paired samples $\{x^{(i)}, y^{(i)}\}$ drawn from the joint probability distribution P_{XY} . This is achieved using a cGAN which comprises two networks, namely the generator G and the critic D . The generator is given by the mapping $G : \Omega_Z \times \Omega_Y \rightarrow \Omega_X$ with $x^G = G(z, y)$, where z is a realization of the N_Z -dimensional latent random variable Z defined on Ω_Z . The latent vector is typically chosen to follow a simple distribution P_Z , such as a multivariate Gaussian, which is easy to sample from. The role of the G is to generate samples (given $Y = y$) from the learned distribution $P_{X|Y}^G$ which are similar to samples from the true conditional. The critic is given by the mapping $D : \Omega_X \times \Omega_Y \rightarrow \mathbb{R}$, with its role being to distinguish between true joint samples $(x, y) \sim P_{XY}$ and fake joint samples (x^G, y) where x^G is a fake sample generated by G .

We consider a particular cGAN variant, known as the Wasserstein cGAN¹⁷, which makes use of the following loss function

$$\mathcal{L}(D, G) := \mathbb{E}_{\substack{(x, y) \sim P_{XY} \\ z \sim P_Z}} [D(x, y) - D(G(z, y), y)]. \quad (1)$$

The two networks are trained simultaneously by solving the following minmax problem

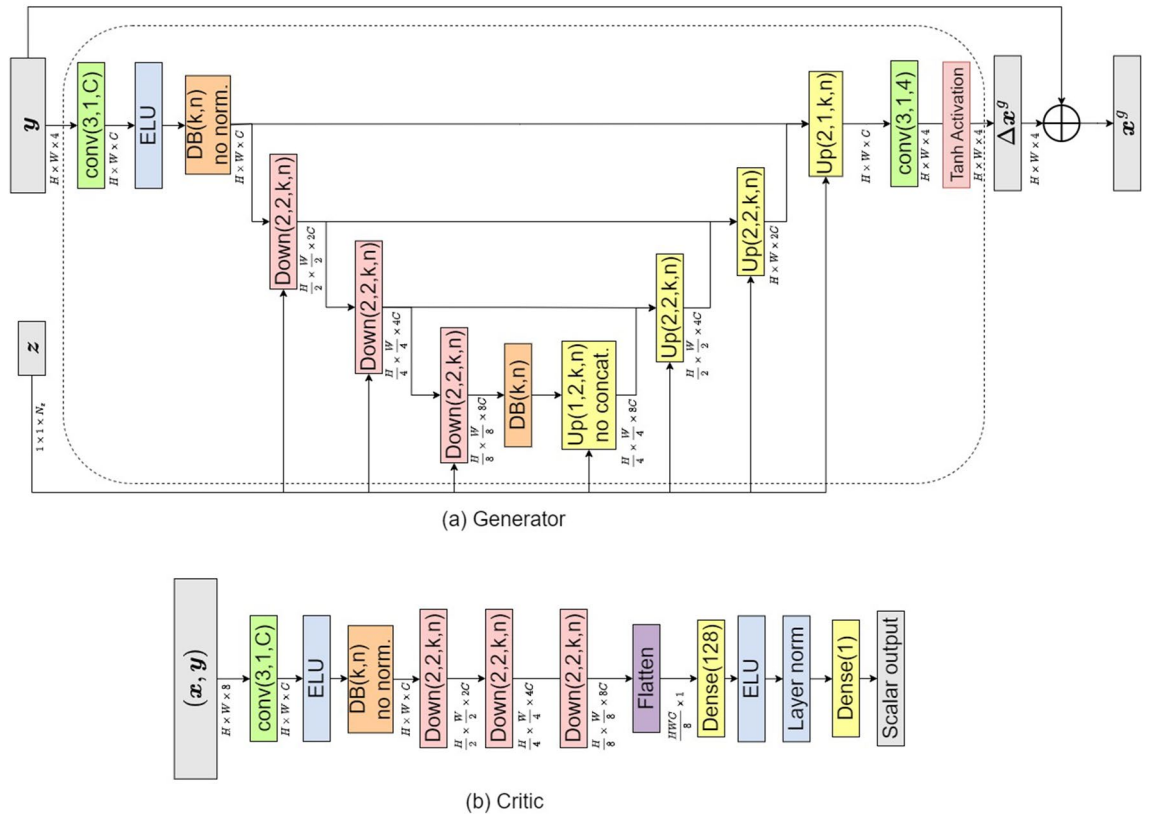


Figure 6. Architecture of generator and critic used in the conditional GAN.

$$D^*(G) = \arg \max_D \mathcal{L}(D, G), \quad G^* = \arg \min_G \mathcal{L}(D^*(G), G). \tag{2}$$

Under the assumption that the critic is 1-Lipschitz and that the maximization problem is solved perfectly, it can be shown that finding the optimal generator is equivalent to minimizing the mean (with respect to the marginal distribution P_Y) Wasserstein-1 distance between $P_{X|Y}$ and $P_{X|Y}^G$ ¹⁷

$$G^* = \arg \min_G \mathbb{E}_{y \in P_Y} [W_1(P_{X|Y}(\cdot|y), P_{X|Y}^G(\cdot|y))]$$

where W_1 is Wasserstein-1 metric³⁵. The Lipschitz constraint on the critic can be weakly imposed using a gradient penalty term while training D ^{17,18} (also see Supplementary Note 3).

In practice, we do not know the exact form of P_{XY} , however, we assume access to samples $\{x^{(i)}\}_{i=1}^K$ of complete image sequences drawn from the marginal P_X of X . For each of these samples, we drop the j -th image in the sequence and use the linear reconstructor R^j to construct the corresponding y . Note that four such y 's can be constructed for each x . This leads to the dataset $\mathcal{S} = \{(x^{(i)}, y^{(i)})\}_{i=1}^M$ with $M = 4K$ samples, where each pair can be seen as a sample from the joint distribution P_{XY} . Using these samples, the expectations in (1) are approximated by empirical averages.

Once the cGAN is trained, given a new incomplete measurement $\tilde{y}[j]$, we can use the trained G^* to generate an ensemble of probable x 's and evaluate their pixel-wise statistics

$$\mathbb{E}_{x \sim P_{X|Y}^G} [f(x)] \approx \frac{1}{S} \sum_{i=1}^S f(G^*(z^{(i)}, R^j(\tilde{y}[j]))) , \quad z^{(i)} \sim P_Z \tag{3}$$

for any continuous, bounded function f on Ω_X . By setting $f(x) = x$ in (3) we can evaluate the mean prediction denoted by \bar{x} , which serves as our best guess for the complete sequence. Choosing $f(x) = (x - \bar{x})^2$ in (3), we can evaluate the pixel-wise variance of the learned posterior distribution. This variance (or rather the standard deviation) can be used to quantify the uncertainty in the reconstructed sequence. The schematic of the reconstruction algorithm is shown in Fig. 1. Note that if the j -th image is missing, we are typically only interested in statistics of the j -th images of the generated ensemble.

cGAN architecture

The architecture of the generator and critic is based on those considered in¹⁸. The generator G has a U-Net architecture, as shown in Figure 6a, taking as input the measured sequence y (after linear reconstruction of the

missing phase) and the latent variable z . The latent information is injected at every scale of the contracting and expanding branches of the U-Net using conditional instance normalization (CIN)³⁶, which has two advantages: i) the latent dimension N_Z can be chosen independently of N_X or N_Y , thus allowing for significant dimension reduction, and ii) stochasticity is introduced at all scales of the U-Net, which overcomes the issue of mode collapse¹⁸.

It is typical to use residual blocks to introduce non-linearity in the U-Net, which is what was also done in¹⁸. However, in the present work, we make use of dense blocks since they lead to superior performance compared to residual blocks while reducing the number of trainable parameters³⁷. In Fig. 6, the dense blocks are denoted by $DB(k, n)$, where n corresponds to the number of sub-blocks in the dense block, while k denotes the number of output features in all but the last sub-block. $Down(p, q, k, n)$ denotes a down-sampling block, which coarsens the input spatial resolution by a factor of p , while increasing the number of channels by a factor of q . The parameters k, n correspond to the dense block used in the down-sampling block. Similarly, $Up(p, q, k, n)$ denotes the up-sampling block, which refines the spatial resolution by a factor of p and decreases the number of channels by a factor of q . Further details about the constitutive blocks of the U-Net are given in the Supplemental Note 2.

Another difference between the present architecture as compared to the U-Net in¹⁸ is the use of an outer skip connection, which adds the output of the U-Net to the measurement sequence y . Thus, the U-Net's output can be interpreted as a pixel-wise perturbation Δx^G of all the images in the sequence, which is added to the input measurement y to predict a complete sequence x^G . Such skip connections are routinely used in U-Nets trained for image de-noising applications^{38,39}.

As shown in Fig. 6b, the architecture of the Critic D comprises dense block-based down-sampling, followed by fully connected network that gives a scalar output. Since the latent variable is not used to evaluate the critic, CIN is replaced by layer normalization⁴⁰. The original Wasserstein cGAN¹⁷ used a more complicated and specialized critic to overcome the mode collapse. However, as discussed and demonstrated in¹⁸, the injection of sufficient stochasticity in the generator via CIN allows the use of a simpler critic architecture.

PIX2PIX: a deterministic algorithm

To demonstrate the benefits of a stochastic imputation model, we compare the results of the proposed cGAN with a PIX2PIX GAN⁶. This is considered the standard model to use in image-to-image translation. The PIX2PIX model in our paper is modified to resemble our cGAN model for a fair comparison. The generator of the PIX2PIX approach is required to not only fool the critic but to also make a prediction that is (point-wise) close to the ground truth. This motivated the augmentation of l_1 distance term in the generator loss⁶, which is the approach we follow as well. The generator for the PIX2PIX model is given by the map $G : \Omega_Y \rightarrow \Omega_X$ and does not use a latent variable. Thus, a single x^G is generated for a given y , unlike our cGAN model. The loss function for PIX2PIX is given by

$$\mathcal{L}(D, G) = \mathbb{E}_{(x, y) \sim P_{XY}} [D(x, y)] - D(G(y), y). \quad (4)$$

To train PIX2PIX, the following minmax problem is solved

$$D^*(G) = \arg \max_D \mathcal{L}(D, G), \quad G^* = \arg \min_G (\mathcal{L}(D^*(G), G) + \lambda_1 \mathbb{E}_{(x, y) \sim P_{XY}} [\|G(y) - x\|_1]) \quad (5)$$

The architectures of the networks will be similar those used in the cGAN approach with the exception that batch normalization⁴¹ is used instead of CIN in the generator.

Other generative models

We briefly describe two existing deep generative models that have been developed for medical image imputation tasks. The first is DiagnosisGAN²⁴, which is a specialized GAN model that can simultaneously generate missing image in CT sequences and classify the cancer subtype. In this model, the generator takes an incomplete multi-phase CT sequence as input and generates a synthesized volume for the missing phase. The training objective function is composed of several loss terms, including the adversarial loss, a reconstruction loss, a lesion segmentation loss, and cancer subtype classification loss.

Another approach is known as the Representational disentanglement scheme for Multi-domain Image Completion (ReMIC²⁵), which is a multi-domain completion and segmentation framework. The ReMIC model consists of a content encoder, which is shared across all domains (FLAIR or MRI for example). There are also domain-specific style encoders, and generators. In addition to an adversarial loss, the framework includes an image consistency loss for visible domains, latent consistency loss, and reconstruction loss for the generated images. Additionally, the framework employs a representational learning approach, where a segmentation generator follows the content code for a unified image generation and segmentation.

We remark that compared to these two models, the proposed cGAN has a much simpler architecture and objective function. Further, the cGAN generator is capable of generating an ensemble of possible missing images for an incomplete CT sequence instead of a single reconstructed image. The ensemble statistics enable us to quantify the uncertainty in the reconstruction.

Once the cGAN is trained, generating samples for a given sequence of incomplete images is relatively inexpensive. In particular, for the problem considered in this manuscript, the generation of 800 images took 100 seconds of wall clock time on two p100 GPUs. Correspondingly, the deterministic approach (PIX2PIX) takes only a small fraction of this time since it generates only a single image. We note that even though the stochastic approach generates many more images and therefore takes more time to produce the results, the overall computational

burden in both approaches (stochastic and deterministic) was dominated by the time spent for training the network, which took around 90 hours in both cases.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 28 April 2023; Accepted: 21 December 2023

Published online: 02 January 2024

References

- Lee, W.-K. *et al.* Imaging of renal cell carcinoma. *Société Internationale d'Urologie J.* **3**, 407–423 (2022).
- Campbell, S. C. *et al.* Renal mass and localized renal cancer: Evaluation, management, and follow-up: Aua guideline: Part i. *J. Urol.* **206**, 199–208 (2021).
- Tsili, A. C. *et al.* The role of imaging in the management of renal masses. *Eur. J. Radiol.* **141**, 109777 (2021).
- Abou Elkassem, A. M. *et al.* Role of imaging in renal cell carcinoma: A multidisciplinary perspective. *Radiographics* **41**, 1387–1407 (2021).
- Goodfellow, I. *et al.* Generative adversarial networks. *Commun. ACM* **63**, 139–144 (2020).
- Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* <https://doi.org/10.1109/cvpr.2017.632> (2017).
- Zhu, J.-Y., Park, T., Isola, P. & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. in *Proceedings of the IEEE international conference on computer vision*, 2223–2232 (2017).
- Choi, Y. *et al.* Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8789–8797 (2018).
- Yoon, J., Jordon, J. & Schaar, M. Radialgan: Leveraging multiple datasets to improve target-specific predictive models using generative adversarial networks. in *International Conference on Machine Learning*, 5699–5707 (PMLR, 2018).
- Lee, D., Kim, J., Moon, W.-J. & Ye, J. C. Collagan: Collaborative gan for missing image data imputation. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2487–2496 (2019).
- Dalmaz, O., Yurt, M. & Çukur, T. Resvit: Residual vision transformers for multimodal medical image synthesis. *IEEE Trans. Med. Imaging* **41**, 2598–2614 (2022).
- Liu, J. *et al.* One model to synthesize them all: Multi-contrast multi-scale transformer for missing data imputation. *IEEE Trans. Med. Imaging* (2023).
- Upadhyay, U., Chen, Y. & Akata, Z. Robustness via uncertainty-aware cycle consistency. *Adv. Neural. Inf. Process. Syst.* **34**, 28261–28273 (2021).
- Upadhyay, U., Chen, Y., Hepp, T., Gatidis, S. & Akata, Z. Uncertainty-guided progressive gans for medical image translation. in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III* **24**, 614–624 (Springer, 2021).
- Tanno, R. *et al.* Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion MRI. *Neuroimage* **225**, 117366 (2021).
- Mirza, M. & Osindero, S. Conditional generative adversarial nets. [arXiv:abs/1411.1784](https://arxiv.org/abs/1411.1784) (2014).
- Adler, J. & Öktem, O. Deep bayesian inversion. <https://doi.org/10.48550/ARXIV.1811.05910> (2018).
- Ray, D., Ramaswamy, H., Patel, D. V. & Oberai, A. A. The efficacy and generalizability of conditional GANs for posterior inference in physics-based inverse problems. *Numer. Algebra Control Optim.* <https://doi.org/10.3934/naco.2022038> (2022).
- Frid-Adar, M. *et al.* Gan-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **321**, 321–331 (2018).
- Lee, H. *et al.* Classification of focal liver lesions in CT images using convolutional neural networks with lesion information augmented patches and synthetic data augmentation. *Med. Phys.* **48**, 5029–5046 (2021).
- Toda, R. *et al.* Synthetic CT image generation of shape-controlled lung cancer using semi-conditional InfoGAN and its applicability for type classification. *Int. J. Comput. Assist. Radiol. Surg.* **16**, 241–251 (2021).
- Liu, J. *et al.* Dyefreenet: Deep virtual contrast CT synthesis. in *Simulation and Synthesis in Medical Imaging: 5th Intl. Workshop, SASHIMI 2020, Lima, Peru, Oct. 4, 2020, Proceedings* **5**, 80–89 (Springer, 2020).
- Seo, M. *et al.* Neural contrast enhancement of CT image. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 3973–3982 (2021).
- Uhm, K.-H., Jung, S.-W., Choi, M. H., Hong, S.-H. & Ko, S.-J. A unified multi-phase CT synthesis and classification framework for kidney cancer diagnosis with incomplete data. *IEEE J. Biomed. Health Inform.* **26**, 6093–6104. <https://doi.org/10.1109/jbhi.2022.3219123> (2022).
- Shen, L. *et al.* Multi-domain image completion for random missing input data. *IEEE Trans. Med. Imaging* **40**, 1113–1122. <https://doi.org/10.1109/tmi.2020.3046444> (2021).
- Raad, R. *et al.* Probabilistic medical image imputation via deep adversarial learning. *Eng. Comput.* **38**, 3975–3986 (2022).
- Patel, D. V. & Oberai, A. A. Gan-based priors for quantifying uncertainty in supervised learning. *SIAM/ASA J. Uncertain. Quantif.* **9**, 1314–1343 (2021).
- Patel, D. & Oberai, A. A. Bayesian inference with generative adversarial network priors. Preprint at *arXiv preprint arXiv:1907.09987* (2019).
- Synapse 3d, <https://healthcaresolutions-us.fujifilm.com/enterprise-imaging/synapse-3d>.
- Conditional GAN medical imputation. <https://github.com/RaghebRaad400/Conditional-GAN-Medical-Imputation/>.
- Israel, G. M. & Bosniak, M. A. How i do it: Evaluating renal masses. *Radiology* **236**, 441–450. <https://doi.org/10.1148/radiol.2362040218> (2005).
- Israel, G. M. & Bosniak, M. A. An update of the Bosniak renal cyst classification system. *Urology* **66**, 484–488. <https://doi.org/10.1016/j.urology.2005.04.003> (2005).
- Silverman, S. G. *et al.* Bosniak classification of cystic renal masses, version 2019: An update proposal and needs assessment. *Radiology* **292**, 475–488. <https://doi.org/10.1148/radiol.2019182646> (2019).
- Warren, K. S. & McFarlane, J. The Bosniak classification of renal cystic masses. *BJU Int.* **95**, 939–942. <https://doi.org/10.1111/j.1464-410x.2005.05442.x> (2005).
- Villani, C. *Optimal Transport: Old and New. Grundlehren der mathematischen Wissenschaften* (Springer, 2008).
- Dumoulin, V., Shlens, J. & Kudlur, M. A learned representation for artistic style. Preprint at [arXiv:1610.07629](https://arxiv.org/abs/1610.07629) (2017).
- Huang, G., Liu, Z., van der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).

38. Guan, S., Khan, A. A., Sikdar, S. & Chitnis, P. V. Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal. *IEEE J. Biomed. Health Inform.* **24**, 568–576. <https://doi.org/10.1109/JBHI.2019.2912935> (2020).
39. Gurrola-Ramos, J., Dalmau, O. & Alarcón, T. E. A residual dense u-net neural network for image denoising. *IEEE Access* **9**, 31742–31754. <https://doi.org/10.1109/ACCESS.2021.3061062> (2021).
40. Ba, J. L., Kiros, J. R. & Hinton, G. E. Layer normalization. Preprint [arXiv:1607.06450](https://arxiv.org/abs/1607.06450) (2016).
41. Ioffe, S. & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML15*, 448–456 (JMLR.org, 2015).

Acknowledgements

The support from ARO grant W911NF2010050 and the Ming-Hsieh Institute at USC is acknowledged.

Author contributions

R.R. implemented trained and tested the cGAN algorithms. D.R. developed and implemented the cGAN algorithms. B.V. was responsible for creating the patient data pipeline. D.H. transformed and manipulated subject image data for the purpose of the study. I.G. provided expert urological guidance while formulating and writing the study. V.D. provided expert radiological guidance in formulating and writing the study. A.A.O. oversaw the development of the algorithms, the conduct of the study, and played a major role in writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-50566-7>.

Correspondence and requests for materials should be addressed to A.A.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024