Abstract

Computed Tomography (CT) is an indispensable tool for the detection and assessment of various medical conditions. This, however, comes at the cost of the health risks entailed in the usage of ionizing X-ray radiation. Using sparse-view CT aims to minimize these risks, as well as to reduce scan times, by capturing fewer X-ray projections, which correspond to fewer projection angles. However, the lack of sufficient projections may introduce significant ambiguity when solving the ill-posed inverse CT reconstruction problem, which may hinder the medical interpretation of the results. We propose a method for resolving these ambiguities, by conditioning image reconstruction on different possible semantic meanings. We demonstrate our method on the task of identifying malignant lung nodules in chest CT. To this end, we exploit a pre-trained malignancy classifier for producing an array of possible reconstructions corresponding to different malignancy levels, rather than outputting a single image corresponding to an arbitrary medical interpretation. The data-consistency of all our method reconstructions then facilitates performing a reliable and informed diagnosis (e.g. by a medical doctor).

1 Introduction

Computed tomography (CT) plays an important role in medical imaging with many applications, such as diagnosing various health conditions and devising appropriate treatment plans [30, 44]. For recording CT data, the target (e.g., a patient) is projected with X-ray radiation from various directions comprising half a circle around it, while a detector measures the attenuated radiation at the other side of the target. Measures corresponding to all projections are then organized as an array termed sinogram, from which the CT image can be reconstructed using different reconstruction methods.

However, the exposure of a patient to the ionizing X-ray radiation is known to present significant health risks such as cancer. This fuels a substantial research effort for reducing radiation exposure, for example by using sparse-view CT, where the target is radiated with
Reconstruction ambiguity. Reconstruction methods such as Filtered Back Projection (FBP) \cite{16} typically yield only one data-consistent output (a). However, there are many possible reconstructions (b) that are consistent with the data term, but differ in their pathological categorization, i.e. with respect to their malignancy. This ambiguity, which increases when using sparse-view CT, is demonstrated here for $p = 50$ projection angles.

Fewer projection angles, typically distributed uniformly around it \cite{28}. Unfortunately however, reconstructing the CT image from the recorded sparse-view data becomes an underdetermined problem, which often manifests itself as significant ambiguities in the tomographic reconstruction process.

The reconstruction of a tomographic image $x$ from a measured sinogram $f$ captured using $p$ projection angles can be formulated as a linear inverse problem of the form

$$ f = Rx + n, \quad (1) $$

where $R \in \mathbb{R}^{p \cdot d \times N}$ corresponds to the discrete Radon transform, which computes $p \cdot d$ line integrals through image $x$ (with a total number of $N$ pixels) along all projection directions\footnote{With a slight abuse of notations, we use $x$ and $f$ when referring either to the two-dimensional or to the column-stacked versions of the target image and the recorded sinogram, respectively.}. Here, $d$ is the number of pixels in the one-dimensional X-ray detector and $n \in \mathbb{R}^{p \cdot d}$ is some additive noise. As the number of projection angles $p$ decreases, the problem of recovering image $x$ in (1) becomes increasingly under-determined.

Over the years many methods attempted to tackle this challenge, typically producing a reconstruction $\hat{x}$ which strives to be as close as possible to the ground truth image $x$. However, due to the under-determined nature of the problem, there are many different valid image reconstructions $\hat{x}$ whose Radon transform $R\hat{x}$ matches the measured sinogram $f$. This is demonstrated in Fig. 1 for the case of a lung nodule captured using $p = 50$ projection angles. While all four reconstructed images on the right are consistent with the sinogram $f$ (satisfying $\frac{1}{p d} \| f - R\hat{x} \|^2 < 3 \cdot 10^{-5}$), their appearances, and more importantly their medical interpretations vary dramatically, with an increasing level of malignancy from left to right.

In this paper we point out the ambiguity that is inherent to medical data reconstruction and argue that enabling exploration of the space of consistent reconstructions, rather than producing a single arbitrary image, is essential in medical applications. We propose the first method to allow this, which enables exploring the range of possible image reconstructions $\hat{x}$ that are consistent with the measurement $f$, while potentially corresponding to different pathological findings. Our method operates by optimizing for different solutions, whose Radon transform matches with the measured sinogram while corresponding to semantically different interpretations, obtained from a pre-trained CT image classifier. In particular, we use gradient descend to minimize the data term induced by (1), as well as a term that encourages the resulting image $\hat{x}$ to be classified into different malignancy levels by a classifier.
that was trained to distinguish between malignant and benign tissues. We introduce technical novelties such as the use of an adversarially trained classifier and the sole use of energy minimization for solution exploration, which is easier and typically more stable for training than, e.g., GAN frameworks. We demonstrate our method on the case of reconstructing human lung CT images with pulmonary nodules, such that they correspond to various degrees of pathological malignancy while maintaining consistency with the measurements $f$. Nonetheless, extending our approach to other use-cases, as well as to other medical imaging modalities, would be fairly straightforward.

2 Related Work

We next provide a brief survey of both classical, as well as learning-based existing CT reconstruction techniques, followed by some background regarding adversarial attacks on neural networks, in the context of our use of a pre-trained classifier. Finally, we discuss existing methods for sampling and exploring the solution space in a few image restoration inverse problems, which are the closest to our work which aims to enable medically oriented exploration in CT reconstruction.

2.1 CT Reconstruction

Classical Techniques For an infinite number of projection angles $p$, image $x$ can be recovered from a noise-free sinogram $f$ (under some mild assumptions) using the inverse Radon transform [35]. In practical settings however, where the target is projected from a discrete set of $p$ angles and the measurements are noisy, $x$ cannot be perfectly reconstructed. One commonly used approach for approximating $x$ is the Filtered Back Projection (FBP) [16] method, which uses a discretized noise-suppressing approximation of the inverse Radon transform and yields a single arbitrary output $\hat{x}$ solving the inverse problem (1). This output typically contains artifacts, especially in sparse-view CT, where $p$ is small. To overcome this problem, many iterative approaches have been presented, that e.g. integrate sparse constrained dictionary learning [43] or use means such as nonlocal regularizers [12] or Total Variations (TV) regularization [21, 37, 38] to reduces artifacts.

In contrast to such classical techniques, we propose to use an energy minimization method that includes a trained classification network, to generate a range of data-consistent reconstructions, thus enabling access to the range of realistic solutions.

Learning-based Techniques Researchers have been applying neural networks for sparse-view CT reconstruction since the turn of the millennium [32, 40, 42]. Some methods [32, 40, 42] proposed to apply post-processing to images reconstructed using FBP, while others proposed to work in the wavelet domain [32] or use an encoder-decoder architecture [11]. Xia et al. [41] proposed a framework for tackling multiple geometries and radiation dose levels. Another line of works was designed to support the iterative reconstruction scheme [28], e.g. by using a neural network to perform projected gradient descend [13] or by taking a primal-dual approach [1]. He et al. [44] proposed to learn a prior for reconstruction using plug-and-play ADMM.

In contrast to the above works, our approach merely requires a classification (rather than a reconstruction) network, and relies on an energy minimization method to explore semantically different but similarly consistent reconstructions.
2.2 Attacks and Stability of Neural Networks

Nowadays, although neural networks are being widely used with great success, they can be vulnerable to adversarial attacks, where small changes in the input image were shown to completely alter the prediction by the neural network. Since this was first discovered by Szegedy et al. \cite{szegedy2013intriguing}, intensive research was devoted to methods that either perform adversarial attacks \cite{goodfellow2014explaining, carlini2017towards, moosavi2016deepfool} or attempt to defend against them \cite{szegedy2013intriguing, goodfellow2014explaining, moosavi2016deepfool}. For instance, Goodfellow et al. \cite{goodfellow2014explaining} showed experiments on adversarial training and proposed the Fast Gradient Sign Method (FGSM) for improving robustness against adversarial attacks. An overview of adversarial attacks and defenses can be found in \cite{szegedy2013intriguing, goodfellow2014explaining, moosavi2016deepfool}.

Note that although our method involves optimizing over the reconstructed image $\hat{x}$ to achieve the desired prediction by a pre-trained classifier (similarly to adversarial attacks), it does not entail any alteration of the recorded sinogram $f$, and instead operates on the reconstructed image $\hat{x}$ in a data consistent manner. This makes our method fundamentally different from adversarial attacks, despite the mathematical similarity manifested in (2) - note that set $P$ of possible inputs to the classifier is unbounded, which is a significant difference on the technical side as well. We emphasize that our method neither performs adversarial attacks, nor does it protects against them.

2.3 Solution Space Exploration

Recently, some of the research efforts in the image restoration community were directed toward expressing and examining the diverse space of valid solutions. Several methods proposed different approaches for sampling the solution space in image restoration problems such as denoising \cite{noh2015learning, chen2016photographic}, super-resolution \cite{kim2016accurate}, inpainting \cite{fiori2017inpainting}, deblurring \cite{kim2016accurate} and more \cite{zhu2016generative}. Other methods go beyond randomly sampling the solution space, and develop tools to enable users to explore it. These tools are typically tailored for a specific task, allowing a user to, e.g., explore the solution space in image super-resolution \cite{barnes2015deep, cao2016learning} or image decompression \cite{park2015scale}, with the latter work exploiting a pre-trained digit classifier for automatically examining the possible decodings corresponding to a compressed image of a numerical digit.

Our work aims to allow exploration (rather than just random sampling) of the space of possible medical interpretations corresponding to a recorded sparse-view CT scan. To the best of our knowledge, such active guidance techniques have not been considered in a medical context before.

3 Explorable CT Reconstruction

Exploring Data Consistent Reconstructions The goal of our work is to provide medical experts that are interpreting images for diagnostic purposes with a better understanding of the actual information the recorded data contains about the object of interest. While our idea extends to any property that can be captured by a classification (or scalar regression) network, we exemplify our method by exploring the space of possible CT reconstructions of nodules associated with different degrees of malignancy as predicted by a given classification method. An overview of our method is depicted in Fig. 2. As a malignancy classifier, we use a classification network $N_\theta : \mathbb{R}^{h \times w} \rightarrow [0, 1]$ pre-trained for classifying nodules in chest CT. The network predicts the malignancy of a nodule from the region of interest $x_C$, which
Figure 2: Overview of the proposed approach. Loss function $\mathcal{L}$ is composed of two energy terms $E_1$ and $E_2$, which take as input the Radon transform of the desired reconstruction $x$ and the output of a pre-trained classification network, respectively, as well as the desired malignancy level $m$. The input of the classification network is a crop $x_C$ from image $x$.

is manually chosen and cropped from a CT image $x$ around the nodule, as shown by the red box in Fig. 1 (a).

Because we want to predict physically plausible, i.e., data consistent, solutions only, we constrain our reconstructions to the solution space $P := \{x \mid \frac{1}{pd} \|Rx - f\|^2 \leq \delta^2\}$ of images $x$ whose sinogram $Rx$ differs from the measured data $f$ by a noise-level dependent constant $\delta$. To allow comparing reconstructions with different number of projection angles $p$, we normalize the squared $\ell^2$ norm by $1/(pd)$ (approximating the $L^2$ norm in function space more realistically). Within $P$ we explore possible solutions using a target malignancy level $m$ for our classification network $N_\theta$ by finding

$$\min_{x \in [0,1]^N} H_\varepsilon(N_\theta(x_C) - m) \text{ s.t. } x \in P,$$

where $H_\varepsilon$ is the Huber loss \[\square\] with $\varepsilon = 0.01$, which we found to work best empirically, being a trade-off between the $\ell^1$ and $\ell^2$ norms:

$$H_\varepsilon(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \leq \varepsilon, \\ \varepsilon \cdot (|a| - \frac{1}{2}\varepsilon) & \text{otherwise.} \end{cases}$$

While (2) could be optimized (at least locally) using a projected gradient descent approach, the projection is rather computationally intense such that we propose to instead consider the regularized problem

$$\min_{x \in [0,1]^N} \frac{1}{pd} \|Rx - f\|^2 + \lambda H_\varepsilon(N_\theta(x_C) - m)$$

with $\lambda$ indicating a weighting of the malignancy prediction of interest. Please refer to the Supp. material for a table summarizing all notations.
Transformations  Our goal is to produce realistically looking (rather than unnatural) CT reconstructions corresponding to malignancy levels $m$. We therefore utilize transformed versions $T_j(x_C)$, where $\{T_j\}_{j=1}^J$ is a set of natural image transformations like different rotations and scalings, which do not affect the semantic interpretation of the image. This reduces the chance of yielding an unrealistic $x_C$ that manages to “fool” the classifier $N_\theta$, as we visualize in the ablation study in the Supp. material, leading to the modified objective

$$\hat{x}(m) = \arg\min_{x \in [0,1]^N} \frac{1}{pd} \|Rx - f\|^2 + \lambda_1 \frac{1}{J} \sum_{j} H_e(N_\theta(T_j(x_C)) - m) + \lambda_2 \text{TV}(x_C),$$  \hspace{1cm} (5)$$

where we use notation $\hat{x}(m)$ to stress the dependency of the reconstructed $\hat{x}$ on our exploration parameter $m$. To further encourage smoothness, we add total variation (TV) regularization [36] to our energy function.

Soft Cropping  We empirically found that hard cropping $x$ to obtain $x_C$ often results in visible artifacts in $\hat{x}$ around the cropping boundary. To encourage a smooth transition of the crop to the remaining part of $\hat{x}$, we attenuate the gradient of $E_2$ in (5) with a Gaussian mask $G$, so that modifications to the peripheral pixels of $x_C$ are attenuated, as we visualize in an ablation study in the Supp. material. Our gradient descent update can be written as

$$x^{i+1} = x^i - \tau(\nabla E_1(x^i) + G \odot \nabla E_2(x^i)),$$  \hspace{1cm} (6)$$

where $\odot$ denotes point-wise multiplication.

Training Suitable Classification Networks  Modern image classification models can be susceptible to adversarial examples - small perturbations in the input image that cause misclassification. To further encourage the reconstruction to be meaningful and realistic, and to prevent slight imperceptible changes in $\hat{x}$ from affecting the classification by $N_\theta$, we utilize adversarial training for classifier $N_\theta$ using the Fast Gradient Sign Method [17], thus making the classifier more robust.

4 Numerical Experiments

4.1 Preparation

Data Set  For our experiments, we use the Lung Image Database Consortium Image Collection (LIDC-IDRI) [3, 13], including over 1000 cases that were annotated by four radiologists independently. There are 5 levels of grading, depending on how certain radiologists are that the nodule is malignant (or benign). To create our training annotation, we averaged the annotated levels for each CT image, and discarded the data whose classification is the closest to indeterminate (level 3), as it can be considered neither malignant nor benign.

From the remaining data, we extract those 2d slices that contain the annotated nodules and get a training dataset of 244 scans with malignant nodules and 729 scans with benign nodules. All CT images are normalized to a range of $[0,1]$. For optimizing (5), we use a validation dataset containing a total of 100 scans with 50 malignant and 50 benign cases, which was not used for training the classification network.
Classification Network To classify nodules of sparse-view CT reconstructions, we use BasicResNet, since it has shown superior results for classifying the malignancy of nodules in [2], by adapting their hyper-parameter settings (with minor changes) and training for 350 epochs using the Adam optimizer and a learning rate of 0.0005. All training input images were normalized by subtracting their mean and dividing by their standard deviation.

4.2 Solution Space Exploration

In all our experiments described below, we optimize (5) along with the Gaussian damping of (6) with a variance of 11, using gradient descent with a learning rate of 1.0 and $\lambda_1 = 1.0$ and $\lambda_2 = 0.01$ for 50000 iterations, with the stopping criterion triggered when the energy no longer decays. To start our method with an $x$ that already results in low energy, we beforehand calculated the FBP ($x_{FBP}$) of our input and minimized $E_1$ for 600 iterations using gradient descent with a learning rate of 0.0005 and a momentum of 0.9. The choice of parameters were obtained empirically. In each optimization step, we normalize the input $x_C$ of $N$ with a fixed mean and variance, taken from the FBP reconstruction.

Realistic Solution Space We explore the space of underdetermined CT reconstructions, with $p = 50$ projection angles, and consider reconstructions of different malignancies, which we control by setting the variable $m$ in (5). Fig. 3 shows an example of reconstructions of a nodule for multiple levels of malignancy, starting from a benign ($m = 0.0$) to a malignant nodule ($m = 1.0$) and several values in between. The prediction of the nodule reconstructed with FBP (a) is classified as benign with $N(x_{FBP}) = 0.2$. It can be seen that in most extreme cases and especially the cases with strong deviation from the predicted malignancy towards increasing malignancy, artifacts appear in the nodule, which here appears as black areas around the nodule. From $m = 0.5$ onward they become visually unrealistic. Numerically we found that rather small changes of $m$ differing by the prediction on the FBP reconstruction by around $\pm 0.1$ yields realistic images while still causing significant changes in the appearance of the nodule, see Fig. 3, (c)–(e). In the following, we refer to the prediction of the reconstruction by FBP as the original malignancy.

Investigation on the Residuals An important question is how much the nodules can change in their appearance and malignancy while still maintaining data consistency. Because this has to depend on the number of projections recorded in the sinogram, we consider reconstructions with 50, 100, 200, and 360 angles and optimize (5) towards malignant and benign reconstructions. Exemplary reconstructions with rather large variations of the malignancy level $m$ are illustrated in Fig. 4 for varying numbers of projection angles $p$. One can see that fewer projections tend to allow larger variations in the reconstructions, e.g. allowing

Figure 3: Nodule reconstructions using FBP (a), and various degrees of malignancy with our proposed approach in (b)–(g).
the nodule to almost disappear for $p = 50$ projections in Sample 3. For a large number of $p = 200$ projections strong deviations from the original malignancy $N_\theta(x_C)$ can lead to severe artifacts that do not correspond to a medically realistic reconstruction anymore.

The severe visual artifacts raise the question to what extent such reconstructions even remain data consistent. Therefore, we analyse the behaviour of the residual $r = |Rx - f|$, that shows the pointwise distance of the sinogram of the reconstruction $Rx$ and the measurement $f$. It is visualized in Fig. 5 for each aforementioned sample for the malignancy that is opposite to its original classification. Here, the red marking indicates the area in the residual that has an influence on the nodule in the reconstruction. We can see that a modification in the nodule (within the red marking) is easier to recognize the more projection angles were used for the reconstruction. For fewer projection angles, such as $p = 50$, it is possible to modify the nodule to another malignancy without any sign of the exploration in the residuum.

To quantify this effect we compute the mean squared error of all points insight the red track of each nodule as well as outside of it, and denote them by $e_i$ and $e_o$, respectively. The plots on the top of Fig. 5, show the difference between the interior and the exterior error $(e_i - e_o)$ as a function of the malignancy we were able to enforce on the reconstruction. As we can see, small values of $p$ allow to not only access the entire range of malignancies without compromising data fidelity, but also do not lead to any recognizable difference between the errors of rays that pass through the nodule and those that do not. As $p$ increases, the differences of errors for Sample 1 and Sample 2 increases for malignancies in opposite to their original classification. In contrast, the plot of Sample 3 shows that there also exist cases, where the difference of errors is not monotone in its classification result.

To go beyond the exemplification on three particular samples, Table 1 shows measurements on the growth of the error and the on the malignancy prediction, when optimizing towards the extrema of malignancy $N_\theta(x_C) = 1$ or $N_\theta(x_C) = 0$ for the validation dataset of 100 reconstructions. Here we differentiate between two sets of reconstructions: those whose corresponding reconstruction by FBP are classified as benign $S^B = \{x | N(x_C^{FBP}) < 0.5\}$ and

---

**Table 1:**

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 0.2$</td>
<td>$m = 0.2$</td>
<td>$m = 0.2$</td>
</tr>
<tr>
<td>$m = 0.8$</td>
<td>$m = 0.8$</td>
<td>$m = 0.8$</td>
</tr>
</tbody>
</table>

**Figure 4:** Examples of data consistent reconstructed nodules for a varying number of projection angles $p$ for $m = 0.2$ and $m = 0.8$, such that the reconstructed nodules are categorized by the classification network $N_\theta$ into different classes with respect to their malignancy.
DRÖGE, BAHAT, HEIDE, MÖLLER: EXPLORABLE CT RECONSTRUCTION

Figure 5: Residual and error measurement for three examples of data consistent reconstructed nodules for a varying number of projection angles \( \rho \). Top: difference of the interior and exterior error when optimized towards different malignancies. Bottom: Pointwise residual (in sinogram space) for different \( \rho \), when optimizing originally benign samples to become malignant and vice versa. The values of the residual are clipped at 0.1.

those whose are classified as malignant \( S^M = \{ x | N(x_{\text{FBP}}^C) \geq 0.5 \} \). Table 4 shows results for exploring reconstruction in the direction of the classification opposite to the classification obtained on the classical FBP reconstruction. We compare their mean data consistency loss, their mean prediction, and their mean distance of the interior and the exterior error \((e_i - e_o)\) for each number of projection angles \( \rho \). As we can see, the data consistency loss as well as the distance of the interior and the exterior error increase with increasing \( \rho \). This applies especially when optimizing the originally malignant classified nodules towards a benign classification. Note also, that large values of \( \rho \) make it impossible to reach extreme values of \( N_\theta(x) \) in our setting. Finally, the widely used FBP reconstructions lead to a reconstruction error that is at least four orders of magnitude higher than all explorable reconstructions. Thus, trusting a FBP reconstruction would mean that an extreme range of possible alternate solutions would have to be considered as well. For the sake of completeness, we included a more detailed table in which we optimize malignancy prediction towards a small and a large value for each set and provide standard deviations for each of the results in the Supp. material.
Table 1: Mean data consistency loss, network prediction and distance between the interior and exterior error \((e_i - e_o)\) of the residual \(r\) for originally benign nodules \(x \in S^B\) and malignant nodules \(x \in S^M\) optimized towards the most extreme opposite malignancies. For comparison the last rows show the results of the reconstructions by FBP.

### 5 Conclusion

In our paper we proposed a method for exploring the solution space of ambiguous sparse-view CT reconstruction on the classification of lung nodules in CT, by conditioning the reconstruction with a pre-trained classification network. We have shown the extent to which we can alter the perceived malignancy of lung nodules and analyzed the range of alterations for different levels of ambiguity in CT images. While current methods aim at predicting the most realistic reconstruction (typically derived from a large set of training data), we argue that an exploration towards the pathologically most and least concerning reconstruction is significantly more informative to a medical expert interpreting the images: A healthy-looking result is a stronger indication of a healthy patient when obtained by optimizing for the most pathologically concerning image, compared to when optimizing for the most realistic one. This holds true particularly in a medical context where great caution needs to be taken of any possible bias arising from the set of training images. We hope this work could set the foundations for developing and experimentally validating comprehensive frameworks that could be used in practice to assist medical patient diagnoses.

### References


[3] Samuel G Armato 3rd, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Hen-


