 DOI: 10.3217/978-3-85125-869-1-09

Computed Tomography Reconstruction Using Generative Energy-Based Priors

Martin Zach[1], Erich Kobler[2], and Thomas Pock[1]

***Abstract*— In the past decades, Computed Tomography (CT) has established itself as one of the most important imaging techniques in medicine. Today, the applicability of CT is only limited by the deposited radiation dose, reduction of which manifests in noisy or incomplete measurements. Thus, the need for robust reconstruction algorithms arises. In this work, we learn a parametric regularizer with a global receptive field by maximizing it's likelihood on reference CT data. Due to this unsupervised learning strategy, our trained regularizer truly represents higher-level domain statistics, which we empirically demonstrate by synthesizing CT images. Moreover, this regularizer can easily be applied to different CT reconstruction problems by embedding it in a variational framework, which increases flexibility and interpretability compared to feed-forward learning-based approaches. In addition, the accompanying probabilistic perspective enables experts to explore the full posterior distribution and may quantify uncertainty of the reconstruction approach. We apply the regularizer to limited-angle and few-view CT reconstruction problems, where it outperforms traditional reconstruction algorithms by a large margin.**

I. INTRODUCTION

Throughout the past decades, Computed Tomography (CT) has become an invaluable tool in diagnostic radiology. However, along with its ever-increasing usage have come concerns about the associated risks from ionizing radiation exposure [6]. Approaches that try to remedy this problem include hardware measures such as tube current reduction or modulation (for instance in the form of automatic exposure control [37]), adaptive section collimation [15], or angular under-sampling [11], [10]. Such measures are now standard in clinical CT systems, but require robust reconstruction algorithms.

Classical CT reconstruction algorithms include Filtered Back-Projection (FBP) [8], [18], which has been superseded by more robust iterative algebraic reconstruction techniques [36], [40] in clinical practice. In light of dose reduction, these algorithms may be equipped with prior knowledge to increase reconstruction quality of low-dose scans. Traditional, hand-crafted regularizers, such as Total Variation (TV) [35] and extensions such as Total Generalized Variation (TGV) [5], typically encode regularity assumptions of the reconstruction, such as sparsity of gradients. These hand-crafted regularizers have been used extensively and successfully in reconstruction problems [13], [26], [41], however they do not fully model the a-priori available information. To capture also higher-order image statistics, the idea of learning a regularizer from data emerged [43], [34], [23]. Although these learning-based approaches are now dominant in many fields, such models have classically focused on modeling local statistics and leave much to be desired in modeling global dependencies.

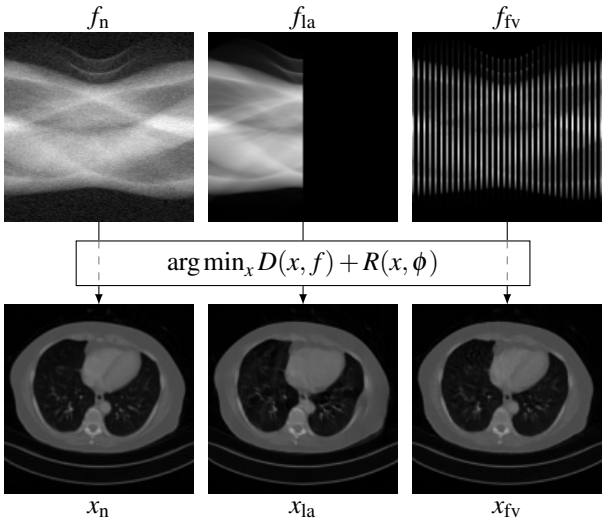

Fig. 1. Our proposed method is able to reconstruct images from noisy, limited-angle and few-view measurements (denoted by the subscripts $\mathrm{n}, \mathrm{la}, \mathrm{fv}$) satisfactorily.

From a statistical point of view, any regularizer R induces a Gibbs-Boltzmann distribution

$$p_R(x) = \frac{\exp(-R(x))}{\int_{\mathscr{X}} \exp(-R(\xi))\ \mathrm{d}\xi}, \tag{1}$$

where $\mathscr{X}$ is the space of all possible images. Ideally, samples $x \sim p_R$ should be indistinguishable from samples from the underlying reference distribution, which is hardly possible for hand-crafted regularizers.

In this work, we propose a novel generatively trained regularizer utilizing a global receptive field that yields high-quality reconstructions even in case of strong noise or heavily undersampled measurements. In Fig. 1, we show how our model is able to satisfactorily reconstruct CT images from noisy (i.e. low tube current) and incomplete (i.e. limited-angle or few-view) data without observable artifacts. In fact, using this regularizer we can synthesize naturally appearing CT images *without any* data (see Fig. 4). In contrast to feed-forward formulations [4], [12], we cast the reconstruction as a variational problem. This helps interpretability of the trained regularizer by means of analyzing its induced distribution as

[1]Institute of Computer Graphics and Vision, Graz University of Technology, 8010 Graz, Austria {martin.zach,pock}@icg.tugraz.at
[2]Institute of Computer Graphics, Johannes Kepler University Linz, 4040 Linz, Austria erich.kobler@jku.at

well as the posterior distribution of any type of reconstruction problem. We apply a trained model to limited-angle and few-view reconstruction problems, and compare our approach quantitatively and qualitatively with traditional reconstruction algorithms. In addition, we perform experiments which leverage the probabilistic nature of our approach, such as prior and posterior sampling.

To summarize, we

- define a novel network architecture capable of synthesizing natural CT images without measurement data,
- demonstrate that our regularizer outperforms classical algorithms in typical reconstruction problems, and
- show that our probabilistic approach allows to compute the pixel-wise posterior-variance, which in turn is related to uncertainty quantification.

II. Related Work

A. Learning-based CT Reconstruction

In recent years, there has been a strong shift from hand-crafted regularizers towards data-driven reconstruction schemes. The learning-based methods can be applied in the sinogram domain [4], [19], such that the final image can be reconstructed using traditional reconstruction algorithms. Alternatively, a preliminary reconstruction may be computed using the (noisy and possibly incomplete) sinogram, which can subsequently be enhanced by a trained convolutional neural network (CNN) [12]. An alternative learning-based reconstruction approach is to learn a direct mapping from the data domain to the image domain [42]. However, this requires to learn a wealth of parameters solely to compute an approximate inverse of the forward acquisition operator. Another recently popularized approach is to learn an unrolled iterative reconstruction algorithm [20], [2], [24]. Whilst the results look promising, we point out that such approaches typically assume a particular acquisition setup and, at inference time, can only be applied in settings that are very similar to the training setting.

B. Generative Models as Regularizers in Medical Imaging

Energy-based models (EBMs) have a long history in the field of image processing [25]. However, only recently some works [17], [29] have explored their generative capabilities, rivaling the performance of Generative Adversarial Networks (GANs). While GANs have been used as an implicit prior for reconstruction problems in medical imaging (e.g. [1]), to the best of our knowledge, using EBMs capable of synthesizing natural images at full-scale as regularizers in medical imaging is still largely unexplored.

III. Methodology

In this work, we represent CT images of size $n = n_w \times n_h$ pixels as vectors $x \in \mathbb{R}^n$. The subsequent analysis easily generalizes to image data in any dimensions. Acquiring n_θ projections with n_d detector elements, the post-log sinogram $f \in \mathbb{R}^m$ of size $m = n_\theta \times n_d$ is given by

$$f = Ax + \eta, \tag{2}$$

where $A\colon \mathbb{R}^n \to \mathbb{R}^m$ is the acquisition operator, and $\eta \in \mathbb{R}^m$ represents the additive measurement noise, summarizing photon statistics, thermal noise in the measurement channels, and pre-processing steps. The linear acquisition operator A is defined by the geometry of the measurement setup, and throughout this work we assume that both A and η can be characterized up to reasonable precision.

A. Bayesian Modeling

To account for measurement uncertainties and missing data in the observations f, we adopt a rigorous statistical interpretation of (2). Bayes' Theorem relates the posterior probability $p(x \mid f)$ to the data-likelihood $p(f \mid x)$ and the prior $p(x)$ by

$$p(x \mid f) \propto p(f \mid x)p(x). \tag{3}$$

Here, $p(x \mid f)$ quantifies the belief in a solution x given a datum f. In the negative log-domain, (3) is transformed to

$$E(x,f) \coloneqq D(x,f) + R(x), \tag{4}$$

where we identify the *data-fidelity* term $D\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^+$ modeling the negative log-likelihood $-\log p(f \mid x)$, and the *regularizer* $R\colon \mathbb{R}^n \to \mathbb{R}$ modeling the negative log-prior $-\log p(x)$. The *energy* $E\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ assigns a scalar $E(x,f)$ to any (x,f)-pair, and in the sense of (3) is interpreted as the negative log-posterior $-\log p(x \mid f)$.

Typically, D makes use of the forward operator A to quantify the agreement between the reconstruction of x and the measured data f. R may for instance represent the TV semi-norm [35], which is well known to favor piece-wise constant solutions. For the sake of simplicity, we assume η to be Gaussian, and consequently set $D(x,f) = \frac{1}{2\sigma^2}\|Ax - f\|^2$, where σ^2 denotes the variance of η. We discuss the choice of R in the next section.

B. Parameter Identification

Although many hand-crafted choices for R exist, such as TGV [5] or wavelet-based approaches [16], it is generally agreed upon that modeling higher order image statistics should be based on learning [43]. In contrast to the widely adopted feed-forward approaches, in this work we retain the variational structure to allow statistical interpretation. To account for the parameters, we extend (4) to

$$E(x,f,\phi) \coloneqq D(x,f) + R(x,\phi), \tag{5}$$

where $R\colon \mathbb{R}^n \times \Phi \to \mathbb{R}$ is parametrized by ϕ in the set of feasible parameters Φ. We illustrate our particular choice of R (for two-dimensional input images) in Fig. 2 and emphasize that the input image is reduced to a scalar only by means of (strided) convolutions. Here, ϕ summarizes the convolution kernels and biases, and Φ reduces to $\mathbb{R}^{n_p}$, where n_p is the total number of parameters.

The Bayesian separation of of data-likelihood and prior allows us to train our regularizer *generatively* without any measurement data as follows. We denote by p_ϕ the Gibbs-Boltzmann distribution of $R(\cdot,\phi)$ in the sense of (1), to emphasize the dependence on the parameters. Assuming

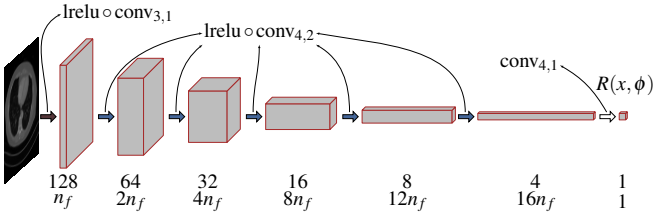

Fig. 2. Our proposed architecture follows a typical encoder structure. The subscripts specify filter size and stride and the annotations show the spatial resolution of the feature maps and the number of features.

access to a distribution p_x of reference CT images, we identify the optimal parameters ϕ^* by minimize the negative log-likelihood

$$\phi^* \in \arg\min_{\phi\in\Phi}\{\Gamma(\phi) := \mathbb{E}_{x\sim p_x}[-\log p_\phi(x)]\}. \tag{6}$$

In the seminal work of [21], it is shown that the gradient of (6) with respect to the parameters ϕ is given by

$$\nabla_1\Gamma(\phi) = \mathbb{E}_{x^+\sim p_x}[\nabla_2 R(x^+,\phi)] - \mathbb{E}_{x^-\sim p_\phi}[\nabla_2 R(x^-,\phi)], \tag{7}$$

where ∇_l denotes the gradient w.r.t. the l-th argument. We discuss the estimation of the expectations in both terms extensively in Sec. III-C.

We highlight that (6) does not require any (x, f)-pairs. That is, for training we do not require access to measurement data but only to (the usually much more ubiquitous) reference images. Moreover, a trained regularization model serves as a drop-in replacement for hand-crafted regularizers for any reconstruction problem by adapting the data-fidelity D to account for a particular forward operator A and noise statistics.

C. Model Sampling

While the first term in (7) is easily approximated given any dataset, the second term requires sampling the induced model distribution, which is known to be hard in high dimensions [7]. For any reasonably sized image $x \in \mathbb{R}^n$ computing the partition function is infeasible, hence the distribution has to be approximated using Markov Chain Monte Carlo (MCMC) techniques. In this work, we utilize the unadjusted Langevin algorithm (ULA) [32], [31], [33], which makes use of the gradient of the underlying probability density function to improve mixing times of the Markov chains. The ULA algorithm read as

$$x^k \sim \mathcal{N}(x^{k-1} + \frac{\varepsilon}{2}\nabla_1 \log p_\phi(x^{k-1}), \beta\varepsilon \mathrm{Id}_n),\ k = 1,\dots,K, \tag{8}$$

where $\mathcal{N}(\mu,\Sigma)$ denotes the normal distribution on $\mathbb{R}^n$ with mean μ and covariance Σ. $\beta, \varepsilon \in \mathbb{R}^+$ are appropriately chosen scaling parameters, and K denotes the total number of steps. To aid the convergence of the Markov chains, we further follow the idea of persistent chains [38] and use a buffer in which the states of the chains persist throughout parameter updates.

D. Experimental Setup

For all the following experiments, we set $n_f = 48$, resulting in $n_p = 12\,179\,905$ and set the ReLU leak coefficient to 0.05.

Algorithm 1: Maximum Likelihood training of an EBM. $\mathcal{U}(\mathcal{X})$ denotes the uniform distribution on $\mathcal{X}$ and each r denotes an independent sample from $\mathcal{U}([0,1])$.

Input : p_x, σ_{data}, n_{buffer}, p_{re}, K, ϕ, n_e, ε, β
Output : ϕ approximately minimizing (6)

$\mathcal{B} \leftarrow \{u_1,\dots,u_{n_{\text{buffer}}}\}$, $u_i \sim \mathcal{U}([0,1]^n)$
for $t = 1,\dots,n_e$ **do**
  $x^+ \sim (p_x * \mathcal{N}(0, \sigma^2_{\text{data}}\mathrm{Id}_n)), x^0 \sim \mathcal{B}$
  Generate x^- with (8) using x^0, ε, K, β
  if $r > p_{re}$ **then** $x_{\text{refill}} = x^-$
  else
    if $r > 0.5$ **then** $x_{\text{refill}} = x^+$
    else $x_{\text{refill}} = u \sim \mathcal{U}([0,1]^n)$
  end
  $\mathcal{B} \leftarrow \mathcal{B} \setminus \{x^0\} \cup \{x_{\text{refill}}\}$
  $\phi \leftarrow \text{Adam}(\nabla_2 R(x^+,\phi) - \nabla_2 R(x^-,\phi))$
end

We trained the regularizer on the Low Dose CT Image and Projection dataset [27], where the images were downsampled to 128×128. We optimized (7) using Adam [22] with a learning rate of 5×10^{-4} and set the first and second order momentum variables to $\beta_1 = 0.9$ and $\beta_2 = 0.999$. To stabilize training, we convolved p_x with $\mathcal{N}(0, \sigma^2_{\text{data}}\mathrm{Id}_n)$, where $\sigma_{\text{data}} = 1.5 \times 10^{-2}$. We used a batch size of 25 and a replay buffer holding 8000 images with reinitialization chance of $p_{\text{re}} = 1\,\%$. Samples in the buffer were reinitialized with an equal chance of uniform noise or samples from the data distribution. To sample p_ϕ, we ran (8) with $K = 500$, using $\varepsilon = 1$ and $\beta = 7.5 \times 10^{-3}$.[1] We summarize the training algorithm in Alg. 1.

For the reconstruction problems, we used accelerated proximal gradient descent [28], as summarized in Alg. 2 with $J = 1 \times 10^3$, $\gamma_1 = 0.5$, $\gamma_2 = 1.5^{-1}$. We solve the proximal operator $\text{prox}\colon \mathbb{R}^n \to \mathbb{R}^n$, which for $H\colon \mathbb{R}^n \to \mathbb{R}$ and $\tau \in \mathbb{R}^+$ is defined as

$$\text{prox}_{\tau H}(y) = \arg\min_x \tau H(x) + \frac{1}{2}\|x - y\|_2^2 \tag{9}$$

using 10 iterations of the conjugate gradient method. In what follows, the forward operator A assumes a parallel-beam geometry with $n_d = 362$ detectors of size 1 pixel and is discretized using the ASTRA toolbox [39]. Unless stated otherwise, η is 0.1 % Gaussian noise.

IV. Results

A. Induced Prior Distribution

For most hand-crafted regularizers, there typically exists a geometrical interpretation. For instance, it is well known that TV is related to the perimeter of the level sets of an image [9]. Hence, the influence on the reconstruction is fairly easily understood. Our regularizer can hardly be interpreted in such

[1] Similar to [30], we reparametrize the regularizer as $\frac{R}{T}$ for a-priori chosen T, such that $\varepsilon = 1$ in (8).

Algorithm 2: Accelerated proximal gradient descent with Lipschitz-backtracking.

Input : initial α, f, x^0, ϕ, J, $\gamma_1 \in (0,1)$, $\gamma_2 \in (0,1)$
Output : x^{J+1} approximately minimizing (5)

$x^1 = x^0$
for $t = 1, \dots, J$ **do**
  $\bar{x} = x^t + \frac{t}{t+3}(x^t - x^{t-1})$
  $g = \nabla_1 R(\bar{x}, \phi)$
  for *ever* **do**
    $x^{t+1} = \operatorname{prox}_{\alpha D(\cdot, f)}(\bar{x} - \alpha g)$
    $Q = R(\bar{x}, \phi) + \langle g, x^{t+1} - \bar{x} \rangle + \frac{1}{2\alpha} \left\| x^{t+1} - \bar{x} \right\|_2^2$
    if $R(x^{t+1}, \phi) \leq Q$ **then**
      $\alpha \leftarrow \alpha / \gamma_1$
      break
    else $\alpha \leftarrow \gamma_2 \alpha$
  end
end

a way, however the energy-perspective allows for a statistical analysis by means of the Gibbs-Boltzmann distribution p_ϕ.

One of the main characteristics of any distribution are the points which locally maximize the density (modes). By (1) it is easily seen that the modes of p_ϕ conincide with local minima of $R(\cdot, \phi)$. However, modes may occur as spikes in regions of generally low mass, and thus samples may represent the underlying distribution more accurately. Therefore, we inspect our regularizer by computing modes as well as samples.

We find $x \sim p_\phi$ using Langevin sampling (8) with $K = 40\,000$ steps, and find $\arg\min_x R(x, \phi)$ with Alg. 2 using $D(x, f) = 0$. In both cases, we set $x_0 \sim \mathcal{U}([0,1]^n)$. We show the trajectories of x^t during minimzation of $R(\cdot, \phi)$ and samples $x \sim p_\phi$ in Fig. 3.

The results indicate that our model is able to synthesize natural CT images without any measurement data. This is in stark contrast to other priors typically used in medical imaging (see e.g. [1, Fig. 1] for samples drawn from hand-crafted priors).

B. Limited-Angle and Few-View Reconstruction

In this section, we shift our focus towards CT reconstruction problems, where we first treat the reconstruction problem as a deterministic mapping in the maximum a-posteriori (MAP) sense. Specifically, we denote by $x^* \colon \mathbb{R}^m \to \mathbb{R}^n$ the model-optimal reconstruction identified by the mapping

$$x^*(f) \in \arg\min_x \{D(x, f) + R(x, \phi)\}. \tag{10}$$

Further, let $p_{\hat{x}}$ denote a distribution on $\mathbb{R}^n \times \mathbb{R}^m$ of (problem-dependent) (f, x)-pairs of a (noisy and incomplete) datum f and the corresponding reference image x.

To illustrate the capabilities of our trained regularizer, we first consider a limited-angle reconstruction problem. Specifically, we reconstruct an image from $n_\theta = 270$ projections uniformly spaced over the quarter-circle $\theta \in [0, \frac{\pi}{2}]$. We show qualitative results in Fig. 4 (top), where the FBP reconstruction exhibits smearing artifacts that are characteristic of limited-angle CT. Simultaneous Algebraic Reconstruction Technique (SART) [3] and additional TV regularization help remedy this problem somewhat, however the reconstruction is not satisfactory. We observe unnatural disconnected contours in the reconstruction, especially around the thorax. On the contrary, our model is capable of reconstructing a natural looking image with realistic anatomy and high level of detail. We show $\mathbb{E}_{(f,x)\sim p_{\hat{x}}}[\mathrm{PSNR}(x^*(f), x)]$ in Tab. I. The results are in accordance with the qualitative analysis, with our model improving the TV reconstruction by over 4.5 dB.

TABLE I
$\mathbb{E}_{(f,x)\sim p_{\hat{x}}}[\mathrm{PSNR}(x^*(f), x)]$ FOR LIMITED-ANGLE ($\theta \in [0, \frac{\pi}{2}]$) AND FEW-VIEW ($n_\theta \in \{100, 50, 30, 20\}$) RECONSTRUCTION.

		FBP	SART	TV	Ours
limited-angle	$\theta \in [0, \frac{\pi}{2}]$	19.05	27.72	29.67	**34.21**
few-view	$n_\theta = 100$	37.15	43.86	46.77	**49.47**
	$n_\theta = 50$	33.12	37.05	40.21	**45.06**
	$n_\theta = 30$	28.78	33.04	35.33	**41.65**
	$n_\theta = 20$	25.24	30.55	31.77	**38.48**

In contrast to limited-angle CT, in few-view CT data are acquired over the full half-circle $\theta \in [0, \pi]$. However, on this half-circle only $n_\theta \ll$ projections are sparsely acquired. In traditional reconstruction algorithms, the sparse data manifests itself as streaking artifacts around sharp contours, where subsequent projections do not properly cancel each other. Such artifacts can clearly be seen in the FBP reconstruction in Fig. 4 (bottom), where we show the results for a $n_\theta = 20$ few-view reconstruction problem. TV regularization yields a sharp and largely artifact-free image at the cost of losing almost all details. Our method can reconstruct the image satisfactorily, where artifacts are removed whilst retaining small details. Tab. I shows quantitative results, with our approach consistently beating the reference methods for all $n_\theta \in \{100, 50, 30, 20\}$ by a large margin.

C. Posterior Analysis

Instead of treating R as a point estimator in the maximum a-posteriori sense (10), the Bayesian formulation allows to explore the full posterior distribution of any given reconstruction problem. This is especially useful in the medical domain, where interpretability is of utmost importance. To this end, we perform Langevin sampling of the posterior distribution (3) with the same parameters as in training. We show some illustrative examples for limited-angle and few-view CT in Fig. 5. The figure shows samples $\xi \sim p(x \mid f, \phi) = p_\phi(x) p(f \mid x)$ from the posterior distribution associated with Eq. (5) as well as it's expectation and variance. For the limited-angle reconstruction, we observe large variance around regions of high ambiguity, where there exist no projections to define contours. Similarly, for the few-view problem, there is high variance around small structures such as the vertebrae or blood vessels in the lung. For both problems, the approximated expected value over the posterior also yields a visually

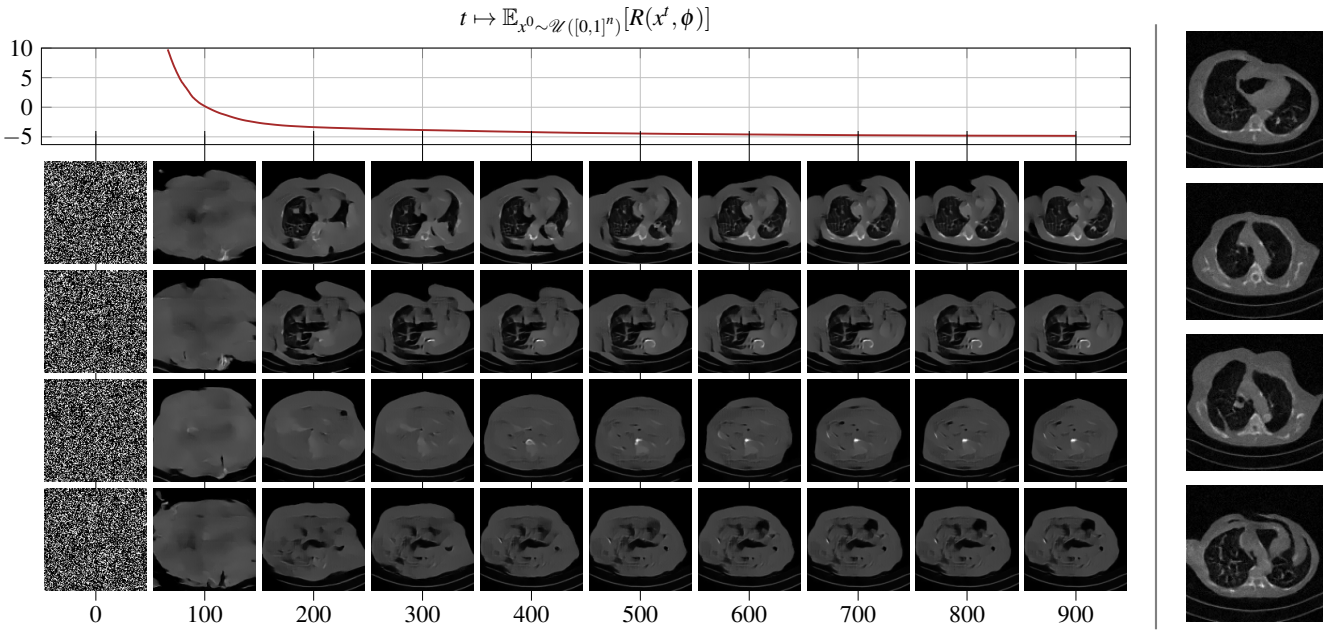

Fig. 3. Trajectories of the images from uniform noise to $\arg\min_x R(x, \phi)$ along with the corresponding $R(x^t, \phi)$ (left) and samples $x \sim p_\phi$ from the Langevin process (8) after $K = 40\,000$ steps (right).

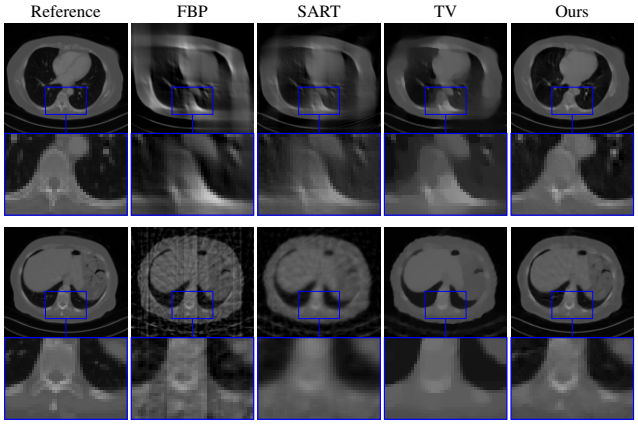

Fig. 4. Comparison between FBP, SART, TV, and our method for limited-angle ($\theta \in [0, \frac{\pi}{2}]$, top) and few-view ($n_\theta = 20$, bottom) CT reconstruction. Our model is able to faithfully reconstruct the image, whereas the other methods are not able to fully remove the smearing and streaking artifacts.

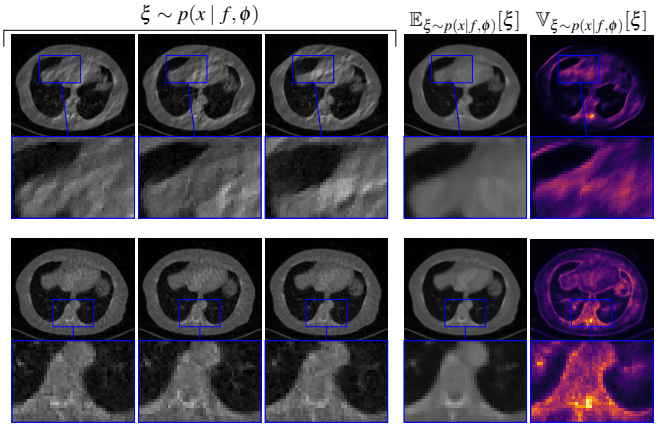

Fig. 5. Sampling the posterior of a limited-angle ($\theta \in [0, \frac{\pi}{2}]$, top) and few-view ($n_\theta = 30$, bottom) CT reconstruction problem: The three images on the left show different samples during the sampling process, the two images on the right show the expected value and variance of the posterior distribution respectively.

appealing, although somewhat over-smoothed, reconstruction.

D. Out-of-Distribution Application

1) Uncertainty Quantification Through Posterior Variance Analysis: To study how the variance relates to uncertainty, we perform the following experiment: We introduce unnatural (read: not present in the training data) structures into the image by overlaying the "cameraman" image and an example of the "grid" texture from the Describable Textures Dataset [14] on a reference scan. Subsequently, we approximate the variance of a few-view reconstruction problem using $n_\theta = 20$ views by Langevin sampling.

We show the expected value and variance over the posterior for the clean and corrupted scans in Fig. 6. Although the bulk of the cameraman shows low variance (and indeed the reconstruction looks natural in these regions), we observe high variance in unnatural regions, such as the artificially introduced corners and the tripod. Similarly, compared to the reference scan the grid overlay leads to high variance in the posterior.

In general, we believe that high posterior variance is related to model uncertainty. To be more specific, we expect high variance if the measurement data suggests structures that are not consistent with the training data. This could potentially aid in detecting pathologies in images.

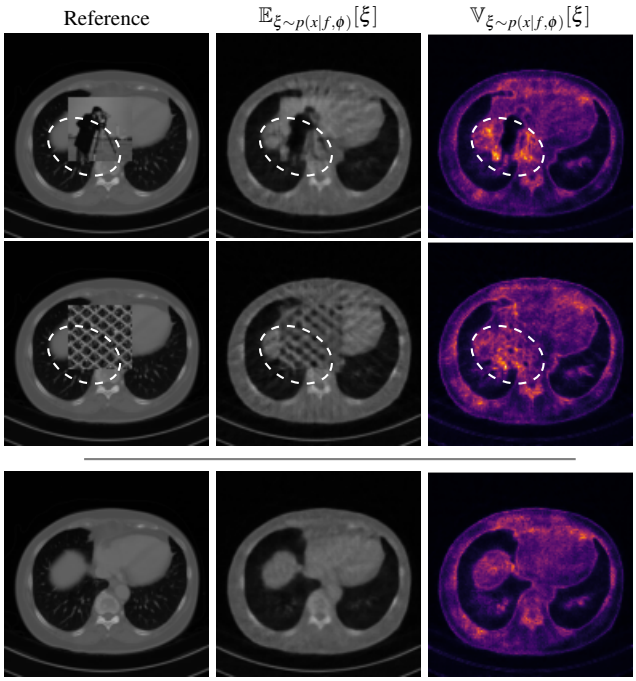

Fig. 6. Comparison of the posterior distribution of a corrupted (top two) versus clean (bottom) scan: The high variance around the corrupted regions (highlighted) relates to high model uncertainty.

2) Generalization: In Sec. IV-A we have shown how samples $x \sim p_\phi$ resemble data drawn from p_x — that is, R encodes a prior in the frequentist sense. With this, a natural question is if our proposed regularizer can be applied to reconstruction problems where the underlying distribution deviates far from p_x. To study this, we propose the following experiment: We let

$$x_\kappa = \mathrm{rot}_\kappa(x) + \eta, \tag{11}$$

where $\mathrm{rot}_\kappa \colon \mathbb{R}^n \to \mathbb{R}^n$ is the bi-linear rotation operator of angle κ and η is 10 % Gaussian noise and find

$$x^* \in \arg\min_x \frac{1}{2\sigma^2} \|x - x_\kappa\|^2 + R(x,\phi). \tag{12}$$

The results in Fig. 7 show that performance quickly deteriorates with increasing κ. This is in line with our expectations, since our regularizer models global characteristics of the reconstruction which are not rotation invariant.

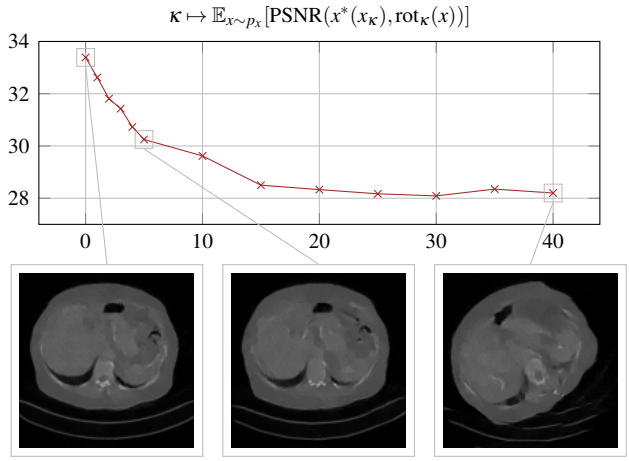

Fig. 7. Performance of the regularizer on out-of-distribution data: For denoising rotated images, the PSNR quickly decays even for small rotations.

V. CONCLUSION

In this work, we designed a parametrized regularizer utilizing a global receptive field, which we trained on full-scale CT images by maximizing their likelihood. The induced Gibbs-Boltzmann distribution of the trained regularizer strongly resembles the data distribution — that is, our model is capable of synthesizing natural CT images *without* any data. The maximum likelihood framework does not assume any particular forward acquisition operator or noise statistics, and the trained regularizer can be applied to any reconstruction problem. In limited-angle and few-view reconstruction problems, we observed significantly improved quantitative and qualitative performance when compared to classical reconstruction algorithms. Further, we were able to relate the variance in the posterior with unnatural structures in the underlying image, as is the case for certain pathologies.

In summary, we believe that learning energy-based models capable of truly capturing the underlying distribution is a very promising direction for future research. Such models yield natural reconstructions with severely undersampled and noisy data, where data consistency can be enforced with arbitrary data terms. We also want to emphasize that training requires only reconstructed images, which are typically much more ubiquitous than image-data pairs. Future work includes the extension to higher resolutions used in clinical practice today, and tackling the problem of scale- and rotation-invariance. Further, a rigorous mathematical analysis in the context of inverse problems, stability w.r.t. training and measurement data would improve the applicability in clinical practice.

REFERENCES

[1] J. Adler and O. Öktem, "Deep bayesian inversion," *arXiv preprint arXiv:1811.05910*, 2018.

[2] ——, "Learned primal-dual reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1322–1332, 2018.

[3] A. Andersen and A. Kak, "Simultaneous algebraic reconstruction technique (sart): A superior implementation of the art algorithm," *Ultrasonic Imaging*, vol. 6, no. 1, pp. 81–94, 1984.

[4] R. Anirudh, H. Kim, J. J. Thiagarajan, K. A. Mohan, K. Champley, and T. Bremer, "Lose the views: Limited angle ct reconstruction via implicit sinogram completion," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6343–6352.

[5] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, Jan. 2010.

[6] D. J. Brenner and E. J. Hall, "Computed tomography — an increasing source of radiation exposure," *New England Journal of Medicine*, vol. 357, no. 22, pp. 2277–2284, 11 2007.

[7] S. Brooks, A. Gelman, J. Galin, and X.-L. Menng, *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2011.

[8] T. M. Buzug, *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*. Springer Berlin Heidelberg, 2008.

[9] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, *An Introduction to Total Variation for Image Analysis*. De Gruyter, 2010, pp. 263–340.

[10] B. Chen, E. Kobler, M. J. Muckley, A. D. Sodickson, T. O'Donnell, T. Flohr, B. Schmidt, D. K. Sodickson, and R. Otazo, "SparseCT: System concept and design of multislit collimators," *Medical Physics*, vol. 46, no. 6, pp. 2589–2599, 5 2019.

[11] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Medical Physics*, vol. 35, no. 2, pp. 660–663, Jan. 2008.

[12] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose ct with a residual encoder-decoder convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.

[13] Z. Chen, X. Jin, L. Li, and G. Wang, "A limited-angle CT reconstruction method based on anisotropic TV minimization," *Physics in Medicine and Biology*, vol. 58, no. 7, pp. 2119–2141, Mar. 2013.

[14] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, , and A. Vedaldi, "Describing textures in the wild," in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[15] P. D. Deak, O. Langner, M. Lell, and W. A. Kalender, "Effects of adaptive section collimation on patient radiation dose in multisection spiral CT," *Radiology*, vol. 252, no. 1, pp. 140–147, July 2009.

[16] B. Dong, J. Li, and Z. Shen, "X-ray CT image reconstruction via wavelet frame based regularization and radon domain inpainting," *Journal of Scientific Computing*, vol. 54, no. 2-3, pp. 333–349, Feb. 2012.

[17] Y. Du and I. Mordatch, "Implicit generation and modeling with energy based models," in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019.

[18] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America A*, vol. 1, no. 6, pp. 612–619, 6 1984.

[19] M. U. Ghani and W. C. Karl, "Cnn based sinogram denoising for low-dose ct," in *Imaging and Applied Optics*. Optical Society of America, 2018, p. MM2D.5.

[20] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3055–3071, Nov. 2017.

[21] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 8 2002.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

[23] E. Kobler, A. Effland, K. Kunisch, and T. Pock, "Total deep variation for linear inverse problems," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[24] E. Kobler, A. Effland, T. Pock, B. Chen, and D. Sodickson, "Total deep variation for sparsect reconstruction," in *6th International Conference on Image Formation in X-Ray Computed Tomography*, 2020, 6th International Conference on Image Formation in X-Ray Computed Tomography ; Conference date: 03-08-2020 Through 07-08-2020.

[25] Y. LeCun, S. Chopra, R. Hadsell, F. J. Huang, and et al., "A tutorial on energy-based learning," in *Predicting Structured Data*. MIT Press, 2006.

[26] Y. Liu, Z. Liang, J. Ma, H. Lu, K. Wang, H. Zhang, and W. Moore, "Total variation-stokes strategy for sparse-view x-ray CT image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 33, no. 3, pp. 749–763, Mar. 2014.

[27] T. R. Moen, B. Chen, D. R. Holmes III, X. Duan, Z. Yu, L. Yu, S. Leng, J. G. Fletcher, and C. H. McCollough, "Low-dose ct image and projection dataset," *Medical Physics*, vol. 48, no. 2, pp. 902–911, 2021.

[28] Y. E. Nesterov, "A method for solving the convex programming problem with convergence rate $o(1/k^2)$," *Doklady Akademii Nauk SSSR*, vol. 269, pp. 543–547, 1983.

[29] E. Nijkamp, M. Hill, T. Han, S.-C. Zhu, and Y. N. Wu, "On the anatomy of mcmc-based maximum likelihood learning of energy-based models," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 5272–5280, Apr. 2020.

[30] E. Nijkamp, M. Hill, S.-C. Zhu, and Y. N. Wu, "Learning non-convergent non-persistent short-run mcmc toward energy-based model," in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019.

[31] G. O. Roberts and J. S. Rosenthal, "Optimal scaling of discrete approximations to langevin diffusions," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 60, no. 1, pp. 255–268, 2 1998.

[32] G. O. Roberts and R. L. Tweedie, "Exponential convergence of langevin distributions and their discrete approximations," *Bernoulli*, vol. 2, no. 4, pp. 341 – 363, 1996.

[33] P. J. Rossky, J. D. Doll, and H. L. Friedman, "Brownian dynamics as smart monte carlo simulation," *The Journal of Chemical Physics*, vol. 69, no. 10, pp. 4628–4633, 1978.

[34] S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2005, pp. 860–867 vol. 2.

[35] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[36] Y. Saad and H. A. van der Vorst, "Iterative solution of linear systems in the 20th century," *Journal of Computational and Applied Mathematics*, vol. 123, no. 1, pp. 1–33, 2000, numerical Analysis 2000. Vol. III: Linear Algebra.

[37] M. Söderberg and M. Gunnarsson, "Automatic exposure control in computed tomography – an evaluation of systems from different manufacturers," *Acta Radiologica*, vol. 51, no. 6, pp. 625–634, July 2010.

[38] T. Tieleman, "Training restricted boltzmann machines using approximations to the likelihood gradient," in *International Conference on Machine Learning*. New York, NY, USA: Association for Computing Machinery, 2008, p. 1064–1071.

[39] W. van Aarle, W. J. Palenstijn, J. D. Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The ASTRA toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, Oct. 2015.

[40] G. Wang, M. W. Vannier, and P.-C. Cheng, "Iterative x-ray cone-beam tomography for metal artifact reduction and local region reconstruction," *Microscopy and Microanalysis*, vol. 5, no. 1, p. 58–65, 1999.

[41] Y. Zhang, W.-H. Zhang, H. Chen, M.-L. Yang, T.-Y. Li, and J.-L. Zhou, "Few-view image reconstruction combining total variation and a high-order norm," *International Journal of Imaging Systems and Technology*, vol. 23, no. 3, pp. 249–255, Aug. 2013.

[42] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and R. M. S., "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, Mar. 2018.

[43] S. C. Zhu, Y. Wu, and D. Mumford, "Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling," *International Journal of Computer Vision*, vol. 27, no. 2, pp. 107–126, 1998.