Journal of *Imaging*

MDPI

Article

Conditional Invertible Neural Networks for Medical Imaging

Alexander Denker *, Maximilian Schmidt, Johannes Leuschner and Peter Maass

Center for Industrial Mathematics, University of Bremen, Bibliothekstr. 5, 28359 Bremen, Germany; maximilian.schmidt@uni-bremen.de (M.S.); jleuschn@uni-bremen.de (J.L.); pmaass@uni-bremen.de (P.M.)
* Correspondence: adenker@uni-bremen.de

Abstract: Over recent years, deep learning methods have become an increasingly popular choice for solving tasks from the field of inverse problems. Many of these new data-driven methods have produced impressive results, although most only give point estimates for the reconstruction. However, especially in the analysis of ill-posed inverse problems, the study of uncertainties is essential. In our work, we apply generative flow-based models based on invertible neural networks to two challenging medical imaging tasks, i.e., low-dose computed tomography and accelerated medical resonance imaging. We test different architectures of invertible neural networks and provide extensive ablation studies. In most applications, a standard Gaussian is used as the base distribution for a flow-based model. Our results show that the choice of a radial distribution can improve the quality of reconstructions.

Keywords: image reconstruction; invertible neural networks; normalizing flows

Citation: Denker, A.; Schmidt, M.; Leuschner, J.; Maass, P. Conditional Invertible Neural Networks for Medical Imaging. *J. Imaging* **2021**, *7*, 243. https://doi.org/10.3390/jimaging7110243

Academic Editors: Fabiana Zama and Elena Loli Piccolomini

Received: 31 August 2021
Accepted: 13 November 2021
Published: 17 November 2021

1. Introduction

The image reconstruction task arising in computed tomography (CT) or medical resonance imaging (MRI) can be formulated as an inverse problem. A forward operator $\mathcal{A} : X \to Y$ maps the image $x^\dagger$ to (noisy) measurements

$$y^\delta = \mathcal{A}x^\dagger + \epsilon, \tag{1}$$

where $\epsilon \in Y$ describes the noise. Research in inverse problems has mainly focused on developing algorithms for obtaining stable reconstructions of the true image $x^\dagger$ in the presence of noise. In recent years, data-driven methods have been increasingly used in research and applications to solve inverse problems [1]. The choice of methods ranges from post-processing approaches [2], unrolling iterative schemes as neural network layers [3,4], and learned regularization terms [5] to complete learning of an inversion model from data [6]. However, many data-driven methods only give a point estimate of the solution as output. However, especially for ill-posed inverse problems, an estimation of the uncertainties is essential. In order to incorporate uncertainties arising in the inversion process, the reconstruction process can be interpreted in a statistical way as a quest for information [7,8]. Instead of approximating a single point estimate, we are interested in the entire conditional distribution $p(x|y^\delta)$ of the image given the noisy measurement data. Traditionally, methods such as Markov chain Monte Carlo [9] or approximate Bayesian computation [10] have been used to estimate the unknown conditional distribution. However, these methods are often computationally expensive and unfeasible for large-scale imaging problems. A new approach is the application of deep generative models for this task. In general, the goal of a deep generative model is to learn a surrogate model for the unknown distribution based on samples. Well-known approaches from the field of generative networks are variational autoencoders (VAEs) [11,12] and generative adversarial networks (GANs) [13]. Recently, flow-based generative models [14] were introduced, which use an invertible transformation to learn a continuous probability density. One of the

advantages is that flow-based models allow exact likelihood computation, thus allowing for maximum likelihood training.

1.1. Related Work

A variety of neural network methods have been proposed to analyze inverse problems [1]. We are especially interested in methods that can estimate the uncertainties arising in the inversion process. Several approaches have been developed in the past, e.g., Bayesian neural networks can be combined with deep learning models [15], or conditional GANs can be used to learn the unknown posterior density implicitly [16]. Recently, flow-based models have been used to learn a surrogate model for the unknown posterior. These flow-based models are often implemented using invertible neural networks. They have been used to predict oxygen saturation in tumors [17], image colorization [18], day-to-night translation [19], reconstruction of the grazing incidence in X-ray fluorescence [20], or the identification of the permeability field of an oil reservoir [21]. There is also the first application for computed tomography [22,23]. Recent work also studied the application of stochastic normalizing flows to inverse problems [24]. Our work builds on the concept of conditional invertible neural networks (cINNs) as introduced in [18], but our focus lies on medical image reconstruction.

1.2. Contributions

Prior work on cINNs for inverse problems dealt mainly with image-to-image problems [18,19] or lower-dimensional applications [17]. These cINNs are implemented using two components: an invertible neural network used for the normalizing flows and a conditioning network used to extract features from the conditional input. This conditioning network does not have to be invertible and is often implemented as a convolutional neural network (CNN). In our work, we expand these concepts to inverse problems in medical imaging, where the topology of the measurement space and the image space differ significantly. In CT reconstruction, the measurements are line integrals over the image domain. In MR imaging, the measurements can be interpreted in the frequency domain. This creates an additional challenge, as CNNs are built to take advantage of local relationships and often fail when there are global relationships in the measurements. We address this problem by integrating a traditional reconstruction operator into the conditioning network of the cINN. For the problem of CT reconstruction, we use the filtered back-projection (FBP) operator, and for MRI, we use the zero-filled inverse Fourier transform. Further, we experiment with two different invertible neural network architectures found in literature: the multi-scale architecture popularized in the *Real NVP* framework [25] and an invertible UNet, as proposed by Etmann et al. [26]. Additionally, we propose the use of a different base distribution, a radial Gaussian distribution, instead of the widely used standard normal distribution.

2. Materials and Methods

In this section, we introduce normalizing flows and discuss how flow-based models can be implemented. We describe building blocks for invertible neural networks and how they can be used for conditional normalizing flows. In the last part of this section, we explain the different architectures used for the experiments.

2.1. Deep Generative Models

The aim of generative modeling is to build a model using a dataset that represents the underlying distribution of the data. There are two distinct goals in generative modeling. The first is to approximate the probability density function of given samples (i.e., density estimation). The second goal is to generate new data samples distributed according to the distribution (i.e., sampling). The term deep generative modeling is used when the underlying model is implemented using neural networks. In recent years, a wide variety of powerful methods have been proposed. These can be broadly grouped into

latent-variable models, autoregressive models [27,28], and normalizing flows (NFs) [29]. The latent-variable models include implicit models, such as generative adversarial networks (GANs) [13] and variational autoencoders (VAEs) [11,12]. These latent-variable models work by specifying a lower-dimensional latent space and learning a conditional distribution to sample from the image space. GANs are trained using a critic or discriminator network in an adversarial scheme. It was recently shown that GANs have the ability to produce realistic-looking images [30]. However, it is not possible to compute the likelihood with a GAN. VAEs induce a noisy observation model and utilize a lower bound to the exact likelihood function for training. So, it is only possible to evaluate an approximation to the exact likelihood. Additionally, the noisy observation model often leads to blurry-looking images. For autoregressive models (ARMs), the joint distribution is factorized into a product of conditional distributions using the product rule. Using this factorization, neural networks are used to model the dependencies. In this way, the likelihood of an ARM can be computed exactly, but sampling from such a model can be slow. Recently, score-based generative models were proposed [31], which are trained to approximate the gradient of the density and rely on Langevin dynamics for sampling. Models based on the concept of NFs have the advantage of allowing exact likelihood calculation, thus offering the possibility to use a maximum likelihood training and a fast sampling procedure. In distinction to VAEs, they are invertible by design and have no reconstruction loss. Recently, stochastic NFs [32] were introduced, which interweave the deterministic invertible transformations of an NF with stochastic sampling, promising more expressive transformations. For more information, we refer to the recent review article by Ruthotto and Haber [33].

2.2. Application of Generative Models to Inverse Problems

Inverse problems can be studied from a statistical point of view [8]. In this interpretation, we are interested in the conditional distribution $p(x|y^\delta)$ of the unknown image x given the measurement data y^δ, the so-called posterior. Using Bayes' theorem, this posterior can be decomposed into a prior $p(x)$ and the likelihood $p(y^\delta|x)$:

$$p(x|y^\delta) \propto p(y^\delta|x)p(x) \tag{2}$$

For a given noise model, the likelihood $p(y^\delta|x)$ can be evaluated using the forward model $\mathcal{A} : X \to Y$ [34]. The prior $p(x)$ encodes information about the image. Deep generative models are usually incorporated in two ways: learning a model for the prior $p(x)$ [35] or learning a model for the full posterior distribution $p(x|y^\delta)$ [19,22]. To explore the posterior distribution, other point estimates can be used. Commonly, the maximum a posteriori (MAP) estimate

$$\begin{aligned} \hat{x} &= \arg\max_{x \in X} p(x|y^\delta) \\ &= \arg\max_{x \in X} \log(p(y^\delta|x)) + \log(p(x)) \end{aligned} \tag{3}$$

or the pointwise conditional mean $\mathbb{E}[x|y^\delta]$ is used as a reconstruction, and the pointwise conditional variance $\mathrm{Var}[x|y^\delta]$ is used to assess the uncertainty. As computing the conditional mean and the conditional variance would require solving a high-dimensional integral, we use an approximation to estimate both moments as

$$\widehat{\mathbb{E}[x|y^\delta]} = \frac{1}{N}\sum_{i=1}^{n} x_i \quad \text{and} \quad \widehat{\mathrm{Var}[x|y^\delta]} = \frac{1}{n}\sum_{i=1}^{N}(x_i - \widehat{\mathbb{E}[x|y^\delta]})^2, \tag{4}$$

with N i.i.d. samples $\{x_i\}$ drawn from the trained model. In our experiments, we focus on directly learning a model for the full posterior $p(x|y^\delta)$.

2.3. Normalizing Flows

The concept of NFs is based on the work of Tabak and Turner [14]. Flow-based models are constructed using two components: a base distribution and an invertible transformation. Let $\mathbf{z}$ be a random variable with a known probability density function $p_{\mathbf{z}}$. This distribution is called the base distribution and should be simple to evaluate and sample from. The second component is a transformation $T_\theta : X = \mathbb{R}^n \to Y = \mathbb{R}^n$, which is parametrized by θ. This transformation has to be invertible, and both T_θ and T_θ^{-1} have to be differentiable. This particular class of functions is called a diffeomorphism. The base distribution $p_{\mathbf{z}}$ induces a distribution via the invertible transformation T_θ on the image space $\mathbf{x} = T_\theta(\mathbf{z})$. Using the change-of-variable theorem, it is possible to evaluate the likelihood of this induced distribution:

$$p_\theta(x) = p_{\mathbf{z}}(T_\theta^{-1}(x))|\det J_{T_\theta^{-1}}(x)|. \tag{5}$$

Here, $J_{T_\theta^{-1}}(x)$ denotes the Jacobian of T_θ^{-1}. In some cases, it may be advantageous to express (5) using the Jacobian of T_θ:

$$p_\theta(x) = p_{\mathbf{z}}(T_\theta^{-1}(x))|\det J_{T_\theta}(T_\theta^{-1}(x))|^{-1}. \tag{6}$$

This exact formulation of the probability density offers the possibility to fit the parameters θ of the NF using maximum likelihood estimation [36]. Assume that we have a dataset of i.i.d. samples $\{x^{(i)}\}_{i=1}^N$ from an unknown target distribution; then, this objective is used for training the NF:

$$\begin{aligned} \max \mathcal{L}(\theta) &= \sum_{i=1}^{N} \log(p_\theta(x^{(i)})) \\ &= \sum_{i=1}^{N} \Big(\log p(T_\theta^{-1}(x^{(i)})) + \log|\det J_{T_\theta}(T_\theta^{-1}(x^{(i)}))| \Big). \end{aligned} \tag{7}$$

This maximum likelihood objective is equivalent to minimizing the Kullback–Leibler divergence between the unknown target distribution and the induced distribution of the flow-based model [29].

The key challenge is to build an expressive invertible transformation T_θ. For this purpose, two essential properties of diffeomorphisms can be exploited. Diffeomorphisms are composable, i.e., if T_1 and T_2 are invertible and differentiable, then, the same holds for $T_2 \circ T_1$. Further, it is possible to decompose the computation of the inverse and the Jacobian determinant:

$$(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-2} \text{ and } \det J_{T_2 \circ T_1}(z) = \det J_{T_2}(T_1(z)) \cdot \det J_{T_1}(z) \tag{8}$$

This allows us to build a complex transformation as a concatenation of simple transformations. We start by defining a base distribution for $\mathbf{z_0}$. Using the concatenated $T_\theta = T_K \circ \cdots \circ T_1$, we can compute the probability density of $\mathbf{x} = \mathbf{z_K} = T_\theta(\mathbf{z_0})$ via

$$p_\theta(z_K) = p_{\mathbf{z_0}}(T_\theta^{-1}(z_K)) \prod_{k=1}^{K} |\det J_{T_k}(T_k^{-1}(z_k))|^{-1} \tag{9}$$

with $z_{k-1} = T_k^{-1}(z_k)$. This composition of transformations leads to the name normalizing flow [29]. The transformations T_i are a critical part of this formulation. We need transformations that:

- are easily invertible,
- offer an efficient calculation of the logarithm of the Jacobian determinant,

and are still expressive enough to approximate complex distributions. Several different models offer invertibility and tractable determinants, e.g., planar flows [37], residual

flows [38,39], or Sylvester flows [40]. We focus on a class of models that are based on so-called coupling layers [36,41]. Besides the invertibility of the transformations, the stability of the inverse pass must also be taken into account. Behrmann et al. [42] showed that typical normalizing flow building blocks can become highly unstable and, therefore, numerically non-invertible.

2.4. Invertible Neural Networks

Invertible neural networks consist of layers that guarantee an invertible relationship between their input and output. Therefore, they are ideally suited to be used as normalizing flow. There is also the advantage that the intermediate activations do not have to be stored during backpropagation in training. Compared to regular neural networks, the memory consumption decreases considerably, so more extensive networks or batch sizes can be realized. For both CT [26,43] and MRI [44], there are already invertible architectures that actively use this property.

The main building blocks of invertible neural networks used in this work are the so-called coupling layers [36,41]. Coupling layers are invertible by design and have block triangular Jacobians, which allow for an efficient calculation of the logarithm determinant. The main idea of a coupling layer is that the input is split into two parts, where one part is transformed, whereas the other is left unchanged. It is crucial to implement some mixing or permutation between coupling layers for all dimensions to influence one another. In imaging applications, invertible spatial downsampling operations are also integrated into the network [17,25,26,45].

2.4.1. Coupling Layers

Let $x \in \mathbb{R}^n$ and I_1, I_2 disjoint partitions of $\{1, \dots, n\}$ with $|I_1| = d$ and $|I_2| = n - d$. Then, a coupling layer is defined via

$$\begin{aligned} y_{I_1} &= x_{I_1} \\ y_{I_2} &= G(x_{I_2}, M(x_{I_1})), \end{aligned} \tag{10}$$

where $G : \mathbb{R}^{n-d} \times \mathbb{R}^{n-d} \to \mathbb{R}^{n-d}$ is called the coupling law, which has to be invertible with respect to the first argument. The function $M : \mathbb{R}^d \to \mathbb{R}^{n-d}$ is the coupling function, which does not need to be invertible and can be implemented as an arbitrary neural network. Two main types of coupling functions have been studied in the literature: additive coupling functions and affine coupling functions. Additive coupling, as used in [36], follows this design:

$$\begin{aligned} y_{I_1} &= x_{I_1} \\ y_{I_2} &= x_{I_2} + M(x_{I_1}) \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} x_{I_1} &= y_{I_1} \\ x_{I_2} &= y_{I_2} - M(y_{I_1}). \end{aligned} \tag{11}$$

A more flexible type of coupling is affine coupling [25]. Affine coupling layers introduce an additional scaling function to the translation of the additive coupling layer. In this way, a scale $s(x)$ and a translation $t(x)$ are learned, i.e., $M(x) = [s(x), t(x)]$:

$$\begin{aligned} y_{I_1} &= x_{I_1} \\ y_{I_2} &= x_{I_2} \odot \exp(s(x_{I_1})) + t(x_{I_1}) \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} x_{I_1} &= y_{I_1} \\ x_{I_2} &= \exp(-s(y_{I_1})) \odot (y_{I_2} - t(y_{I_1})) \end{aligned} \tag{12}$$

Instead of choosing $\exp(\cdot)$, sometimes other functions that are non-zero everywhere are used. Because one part of the input is unchanged during the forward pass of a coupling layer, we get a lower block triangular structure for the Jacobian matrix:

$$\frac{\partial y}{\partial x} = \begin{pmatrix} I_m & 0 \\ \frac{\partial y_{I_2}}{\partial x_{I_1}} & \frac{\partial y_{I_2}}{\partial x_{I_2}} \end{pmatrix}. \tag{13}$$

This allows us to compute the determinant as $\det\left(\frac{\partial y}{\partial x}\right) = \det\left(\frac{\partial y_{I_2}}{\partial x_{I_2}}\right)$, which drastically reduces the computational complexity. For additive coupling layers, this further reduces to the identity matrix, i.e., they have a unit determinant. Affine coupling layers have a diagonal structure in the block:

$$\det\left(\frac{\partial y_{I_2}}{\partial x_{I_2}}\right) = \exp\left(\sum_{i \in I_1} s(x_1)_i\right). \tag{14}$$

However, as $s(x_1)$ is already evaluated in the forward pass, computing the determinant does not involve significant computational effort. The special structure of the Jacobian highlights the fact that some parts of the input are not processed and have no influence on each other. It is essential to include some permutation or mixing of dimensions in order to build an expressive sequence of coupling layers.

2.4.2. Channel Mixing and Downsampling

For each coupling layer, the input is split into two parts, and only one-half is processed. For image data, this splitting is usually done in the channel dimension. Let $u \in \mathbb{R}^{c \times h \times w}$ be an image with c channels. We choose c_1, c_2 such that $c_1 + c_2 = c$. The image is then split into two parts, $u_{I_1} \in \mathbb{R}^{c_1 \times h \times w}$ and $u_{I_2} \in \mathbb{R}^{c_2 \times h \times w}$. In earlier works, the permutation after each coupling layer was implemented as a fixed random channel shuffling [36]. In the *Glow* architecture, an improvement was seen when using fixed 1×1 convolutions instead of simple permutations [45]. These fixed convolutions can be seen as a generalization of random shuffling. Another central part of invertible neural networks in imaging applications is invertible downsampling operations, i.e., reduction of the spatial dimensions of image data. The standard downsampling operations in CNNs, such as pooling layers or strided convolutions, are inherently non-invertible, as they reduce the dimensionality of the image. Invertible downsampling operations reduce the spatial dimension while simultaneously increasing the number of channels, thus keeping the overall dimensionality the same. Let $u \in \mathbb{R}^{c \times h \times w}$ be an image with c channels, where both the height h and the width w are even. An invertible downsampling operation halves both spatial dimensions and quadruples the number of channels, i.e., $\tilde{u} \in \mathbb{R}^{4c \times h/2 \times w/2}$. There are three main types of invertible downsampling operations used in the literature. The first is checkerboard downsampling, which is a simple rearrangement of the image pixels [46]. A more advanced type of downsampling is haar downsampling, introduced in [17], which uses the 2D haar transform to decompose the image into average channels and vertical, diagonal, and horizontal components. These two downsampling operations are illustrated in Figure 1. Recently Etmann et al. introduced a learnable invertible downsampling operation [26].

Figure 1. Input image (**a**) and output of checkerboard downsampling (**b**) and haar downsampling (**c**). Inspired by [26].

2.5. Base Distribution

In most applications, a standard n-dimensional Gaussian $\mathbf{z} \sim \mathcal{N}(0, I)$ is chosen as the base distribution, which leads to the following log-likelihood:

$$\log(p_{\mathbf{z}}(z)) = -\frac{1}{2}\|z\|_2^2 - \frac{n}{2}\log(2\pi). \tag{15}$$

The second term is constant with respect to z and can be dropped during training. It has been observed that the likelihood of flow-based models sometimes exhibits artifacts, i.e., out-of-distribution data are often assigned a higher likelihood than training data [47]. In [48], the authors suggested that this behavior is due to the difference between the *high-likelihood set* and the *typical set* in high-dimensional Gaussian distributions. For a standard Gaussian, the region of the highest density is at its mean, but the typical set is at a distance of $\sqrt{d}$ away from the mean. In [49], the authors addressed this problem for Bayesian neural networks and chose a *radial* Gaussian distribution where the typical set and high-density region coincided. This radial Gaussian was formulated in hyperspherical coordinates, where the radius is distributed according to a half-normal distribution, i.e., $r = |\hat{r}|$ with $\hat{r} \sim \mathcal{N}(0,1)$, and all angular coordinates follow a uniform distribution over the hypersphere. We use this radial distribution as a base distribution for training flow-based models. This radial distribution leads to the following log-likelihood:

$$\ln p_{\mathbf{z}}(z) = \ln\left(\frac{\sqrt{2}}{\sqrt{\pi}S_n}\right) - (n-1)\ln(\|z\|_2) - \frac{\|z\|_2^2}{2}, \tag{16}$$

where S_n is the surface of the n-dimensional unit sphere. The derivation can be found in Appendix A.1. Sampling is nearly as efficient as for the standard Gaussian distribution. First, a point $x \sim \mathcal{N}(0, I_n)$ is sampled and normalized. This point is then scaled using a radius $r = |\hat{r}|$ with $\hat{r} \sim \mathcal{N}(0,1)$.

Other base distributions have also been considered in the literature. Hagemann and Neumayer used a Gaussian mixture model as a base distribution, which led to higher-quality samples, especially in multi-modal applications [50].

2.6. Conditional Normalizing Flow

Let $\mathbf{x}$ and $\mathbf{y}$ be two random variables over two spaces, X and Y. For our applications, we always use $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$. The goal of conditional density estimation is to approximate the conditional probability distribution $p(x|y)$ given an i.i.d. data set $\{(x^{(i)}, y^{(i)})\}_{i=1}^N$ of input–output pairs sampled from the joint distribution $p(x,y)$. We use a conditional normalizing flow (CNF) to build a probabilistic model $p_\theta(x|y)$ to approximate the unknown conditional distribution $p(x|y)$ [17,51]. A CNF consists of a transformation $T_\theta : Z \times Y \to X$ that has to be invertible with respect to the first argument, and both $T_\theta(\cdot; y)$ and $T_\theta^{-1}(\cdot; y)$ have to be differentiable for every $y \in Y$. By choosing a base distribution $p_{\mathbf{z}}$, the CNF model induces a probability distribution, and the density can be evaluated via the change-of-variable method:

$$p_\theta(x|y) = p_{\mathbf{z}}(T_\theta^{-1}(x;y))\left|\det\left(\frac{\partial T_\theta^{-1}(x;y)}{\partial x}\right)\right|. \tag{17}$$

We use $J_{T_\theta^{-1}}(x;y) = \frac{\partial T_\theta^{-1}(x;y)}{\partial x}$ as a shorthand notation for the Jacobian matrix. Fitting the parameters θ of the CNF can be done using a maximum likelihood loss:

$$\begin{aligned}\max_\theta \mathcal{L}(\theta) &= \sum_{i=1}^N \log(p_\theta(x^{(i)}|y^{(i)})) \\ &= \sum_{i=1}^N \left(\log p(T_\theta^{-1}(x^{(i)};y^{(i)})) + \log\left(\left|\det J_{T_\theta^{-1}}(x^{(i)};y^{(i)})\right|\right)\right).\end{aligned} \tag{18}$$

We use the same trick as for the NF and implement the CNF as a concatenation of simple invertible building blocks.

Conditional Coupling Layers

Conditional coupling layers are the primary way of constructing expressive CNF models. They can be seen as an extension of the original coupling layers and were introduced in [18] for modeling conditional image densities. For a conditional coupling layer, we extend the coupling function M to take the measurements y^δ as an additional input. Let $x \in \mathbb{R}^n$ be the input, $y^\delta \in R^m$ the measurements, and I_1, I_2 disjoint partitions of $\{1, \dots, n\}$ with $|I_1| = d$ and $|I_2| = n - d$. Then, a conditional coupling layer is defined by

$$\begin{aligned} y_{I_1} &= x_{I_1} \\ y_{I_2} &= G(x_{I_2}, M(x_{I_1}, y^\delta)) \end{aligned} \tag{19}$$

where $G : \mathbb{R}^{n-d} \times \mathbb{R}^{n-d} \to \mathbb{R}^{n-d}$ is called the coupling law, which has to invertible with respect to the first argument. Function $M : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^{n-d}$ is the coupling function. Conditional coupling layers offer the same advantages as regular coupling layers, i.e., a block triangular Jacobian and analytical invertibility. In our experiments, we mainly use conditional affine coupling layers, i.e., replacing $s(x_{I_1})$ and $t(x_{I_1})$ with $s(x_{I_1}, y^\delta)$ and $t(x_{I_1}, y^\delta)$. For any fixed conditional input y^δ, the conditional normalizing flow is invertible.

Another way of introducing the conditional input y^δ into the model is to use a conditional base distribution [51]. In this approach, the base distribution can be modeled as a normal distribution where the mean and variance are functions of y^δ, i.e., $p(\mathbf{z}|y^\delta) = \mathcal{N}(\mathbf{z}; \mu(y^\delta), \sigma^2(y^\delta))$. Both the mean and variance function can be parametrized as a neural network and trained in parallel to the flow-based model.

2.7. Conditioning Network

Instead of directly using the measurements y^δ as an additional input to the conditional coupling layer, a conditioning network H is used, which transforms the y^δ to $h = H(y^\delta)$ [18,51]. The motivation behind this is that the conditioning network can learn to extract essential features. This decouples the feature extraction and the density modeling. It is possible to either use a fixed, pre-trained network H or to train the conditioning network parallel to the CNF. This conditioning network is often implemented as a big CNN. As convolutional networks are built to exploit equivariance in natural images, they are not ideally suited for CT or MRI measurement data. Instead, we implemented this conditioning network as a model-based inversion layer $\mathcal{A}^\dagger$, which maps from the measurement space to the image space, concatenated with a post-processing CNN to extract features from this initial reconstruction.

Depending on the structure of the conditioning network, an additional loss term for this network can be used during training. One option is to compare the output of H to the ground-truth data and thereby train a second reconstruction path within the whole cINN. The goal is to get a single high-quality reconstruction from the conditioning network and cover the uncertainties, e.g., from ambiguous solutions, in the sampled reconstruction from the CNF. During inference, the output from the conditioning network and the CNF can be combined to create the final reconstruction.

2.8. Multi-Scale Architecture

Unlike other latent variable models, such as GANs or VAEs, flow-based models work with a full-dimensional base distribution. This is necessary to ensure bijectivity. However, it is also expensive in terms of both memory cost and computational complexity to propagate the full-dimensional image through the network. A typical architecture for flow-based models is the *multi-scale architecture* [25]. This architecture combines coupling blocks, downsampling, and splitting operations. A part of the intermediate representation is split off and directly forwarded to the output for each scale. This combination of splitting and feed-forwarding creates a hierarchy of features and reduces the computational effort. We

visualize this architecture in Figure 2. In our experiments, we always use downsampling of factor 2 after each scale. A multi-scale architecture with L scales can be described by:

$$\begin{aligned} x^0 &= x \\ (z^{i+1}, x^{i+1}) &= f^{i+1}(x^i, H^i(y^\delta)) \\ z^L &= f^L(x^{L-1}, H^{L-1}(y^\delta)) \\ z &= (z^1, \dots, z^L). \end{aligned}$$

Each f^i consists of a coupling $\rightarrow$ downsampling $\rightarrow$ coupling $\rightarrow$ splitting operation.

The *multi-scale architecture* follows the NICE and Real-NVP framework [25,36] and is related to the i-RevNet architecture [46]. However, in i-RevNet, the authors refrained from splitting the dimensions in their bijective architecture.

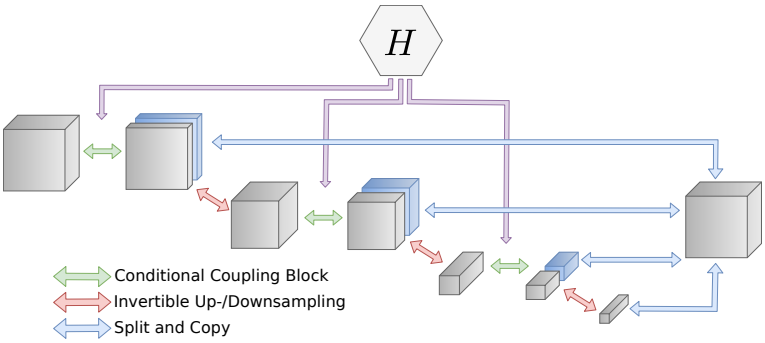


Figure 2. Multi-scale architecture with conditioning network H. The conditioning network processes the conditioning input y^δ and outputs this to the respective conditional coupling layer.

2.9. Invertible UNet

With the iUNet, we follow the work of Etmann et al. [26]. The idea is to adapt the concept of the UNet architecture [52] and replace all common layers with their invertible counterparts. In addition, we introduce a conditioning network H, which also has a UNet structure. In this case, the layers do not have to be invertible. Network H uses the same spatial down- and upsampling scales as the iUNet. At each scale, the current activation $H^i_{u;d}$ is used as conditioning for the respective block $f^{i+1}_{d;u}$ in the iUNet. Note that the direction of the UNet is inverse to the iUNet, since it starts from measurement $y^\delta \in Y$ and maps to X. A representation of the whole network is shown in Figure 3. For an architecture with L scales, we get:

$$\begin{aligned} x^0_d &= x \\ (c^{i+1}, x^{i+1}_d) &= f^{i+1}_d(x^i_d, H^i_u(y^\delta)), && i = 0, \dots, L-2 \\ x^L_d &= f^L_d(x^{L-1}_d, H^{L-1}_u(y^\delta)) \\ x^L_u &= x^L_d \\ x^{i-1}_u &= f^i_u((x^i_u, c^i), H^i_d(y^\delta)), && i = L, \dots, 1 \\ z &= x^0_u \end{aligned}$$

where indices d, u denote the down- and upsampling paths, respectively. Block f^i_d consists of coupling $\rightarrow$ downsampling $\rightarrow$ split, f^L_d is just coupling, and f^i_u is upsampling $\rightarrow$ concat $\rightarrow$ coupling. Compared with the multi-scale architecture, the iUNet concatenates the splits step-by-step in the upsampling path and not all together in the last layer.

The conditioning UNet H creates outputs in the image domain X. Therefore, we can introduce an additional conditioning loss, as proposed in Section 2.7. Specifically, we use

$$\min_{\theta} -\log(p_\theta(x|y^\delta)) + \alpha \,\mathrm{MSE}(H(y^\delta), x), \tag{20}$$

where $\alpha \geq 0$ is a weighting factor. Note that one can also use a pre-trained UNet with fixed parameters as conditioning and benefit from the advantages of the CNF in comparison to a simple post-processing approach.

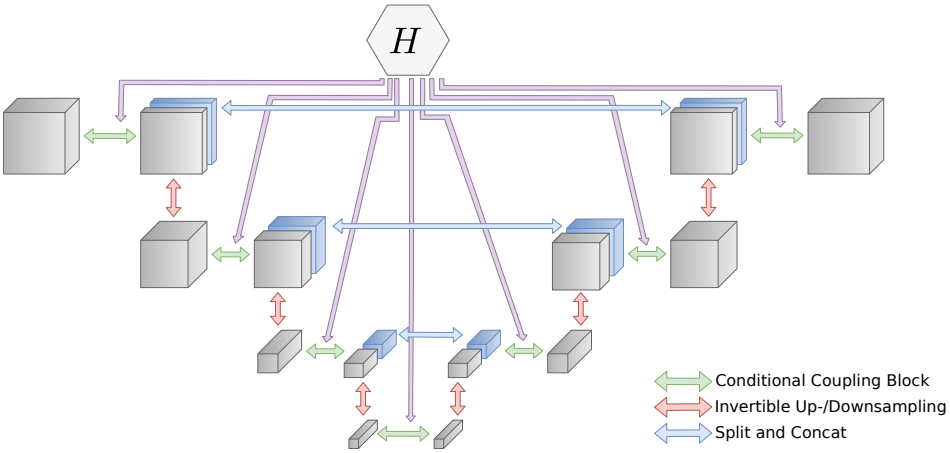


Figure 3. End-to-end invertible UNet with conditioning network H. The conditioning network processes the conditioning input y^δ and outputs this to the respective conditional coupling layer.

3. Experimental Setup

In this section, we present three different applications used to evaluate different architectures for conditional flow-based models. In the first example, we study compressed sensing with Gaussian measurements on the popular MNIST dataset [53]. The other two applications cover essential aspects of medical imaging: accelerated magnetic resonance imaging and low-dose computed tomography. In these two medical imaging scenarios, different sources introduce uncertainty into the reconstruction process. We have an undersampling case in accelerated MRI, i.e., we have fewer measurements than necessary according to the Nyquist–Shannon sampling theorem. So, a strong prior is needed for a good reconstruction. The challenge in low-dose CT is that the lower radiation dose leads to a worse signal-to-noise ratio. Although we are in an oversampling case, the reconstruction is complicated by a more significant amount of noise.

Our source code is publicly available at https://github.com/jleuschn/cinn_for_imaging (last accessed: 16 November 2021).

3.1. Compressed Sensing

As an initial example, we study a similar setup to that in [54]. The goal is the recovery of an image from Gaussian measurements. We evaluate our models on the popular MNIST [53] dataset, which consists of 28×28-size images of handwritten digits. MNIST contains 60,000 training images and 10,000 test images. We split the 60,000 training images into 50,000 for training the CNF model and 10,000 for validation. The forward operator is a matrix $\mathcal{A} \in \mathbb{R}^{m \times n}$. It has independent Gaussian entries with zero mean and variance $1/m$, i.e., $\mathcal{A}_{i,j} \sim \mathcal{N}(0, 1/m)$. We use $m = 196, n = 784$, i.e., 4 times downsampling. We added 10% relative noise to the simulated measurements. In this experiment, we want to study the influence of the inversion layer in the conditioning network H. We use the generalized inverse $\mathcal{A}^\dagger = \mathcal{A}^+$ and a TV-regularized solution $\mathcal{A}^\dagger = (\mathcal{A}^T\mathcal{A} + \lambda\nabla^T\nabla)\mathcal{A}^T$ with a regularization parameter $\lambda = 0.02$. We further use the same neural network architecture for both the conditional invertible network and the conditioning network for both choices of $\mathcal{A}^\dagger$. The cINN was implemented as a multi-scale architecture with

two learnable downsampling operations. The exact implementation can be found in Appendix A.2.

3.2. Computed Tomography

When describing the propagation of radiation through biological tissue, two processes have to be considered: absorption and scattering. For high-energy X-ray beams, the scattering effect is usually neglected. The forward problem in parallel-beam computed tomography can then be described by the 2D Radon transform [55]:

$$Ax(s,\varphi) = \int_{\mathbb{R}} x\left(s\begin{bmatrix}\cos(\varphi)\\ -\sin(\varphi)\end{bmatrix} + t\begin{bmatrix}-\sin(\varphi)\\ \cos(\varphi)\end{bmatrix}\right) \mathrm{d}t, \tag{21}$$

where x is the spatial varying mass absorption coefficient, which depends on tissue type and density. The Radon transform corresponds to the log-ratio between the source intensity and the measured intensity.

For continuous, noise-free measurements, the filtered back-projection (FBP) in combination with the Ram-Lak filter gives the exact inversion formula [56]. In general, recovering the image is a mildly ill-posed problem in the sense of Nashed [57,58]. This means that slight deviations in the measurement, e.g., noise, can lead to significant changes in the reconstruction. The influence of the noise can be reduced by choosing an adequate filter for the FBP. Another challenge arises from the discretization of real measurements, which can lead to artifacts in the FBP reconstruction. Over the years, a number of different reconstruction methods, such as algebraic reconstruction techniques [59] (ART) and total variation (TV) regularization [60], were introduced to compensate for the drawbacks of the FBP. Recently, deep learning approaches extended the choice of methods to push the boundaries on image quality for low-dose, sparse-angle, and limited-angle measurements [2,3,23,61].

In our experiments, we use the LoDoPaB-CT dataset [62] to replicate the challenges that arise from low-dose CT measurements. The dataset contains over 40,000 normal-dose, medical CT images from the human thorax from around 800 patients. Poisson noise is used to simulate the corresponding low-dose measurements. See Figure 4 for an example of a simulated low-dose measurement, an FBP reconstruction, and the ground-truth image. LoDoPaB-CT has a dedicated test set that we use for the evaluation and comparison of our models. In addition, there is a special challenge set with undisclosed ground-truth data. We evaluate the best model from our experiments on this set to allow for a comparison with other reconstruction approaches. The challenge results can be found on the online leaderboard (https://lodopab.grand-challenge.org/evaluation/challenge/leaderboard/, last accessed: 16 November 2021).

Figure 4. Reconstruction and measurements for the low-dose LoDoPaB-CT data.

3.3. Magnetic Resonance Imaging

We will now briefly introduce MRI and the considered simple model, following the description in [63], to which we refer the reader for more details, including limitations of the model.

In MRI, one measures the radio-frequency (RF) responses of nuclei (e.g., protons) to RF pulses while applying different external magnetic fields in order to obtain a density image. A strong static magnetic field is applied, which causes the resonance frequency of the nuclei to be within the RF range. Pulses at this frequency are emitted using an RF transmitting coil, triggering RF response signals detected by an RF receiving coil. For spatial encoding, configurable magnetic gradient fields $G = (G_x, G_y, G_z)$ are applied that change the applied magnetic field and thereby the resonance frequency depending on the location. During a scan, different gradient fields G are selected for each repetition of a pulse sequence.

A simple model for the measured receive coil signal in each repetition is given by

$$y(t) = \int x(r) \exp(-2\pi i k(t) \cdot r) \, \mathrm{d}r, \qquad k(t) = \gamma \int_0^t G(\tau) \, \mathrm{d}\tau,$$

where x is the spatial signal density (i.e., the image) and k specifies a position in the so-called k-space, which coincides with the Fourier space. The choice of G determines the trajectory of k for this repetition. By collecting samples from multiple repetitions, one can obtain a complete Cartesian sampling of the k-space that satisfies the Nyquist–Shannon sampling theorem. This enables (approximate) reconstruction via the inverse fast Fourier transform (IFFT).

A major limiting factor is the time-consuming measurement process, which directly depends on the number of repetitions required to obtain a full sampling of the k-space. While using fewer repetitions accelerates the process, it leads to an underdetermined reconstruction problem and can introduce artifacts due to the missing frequencies. In order to reconstruct from undersampled measurement data, prior information needs to be incorporated. Additionally, measurements are noisy in practice, further increasing reconstruction ambiguity, since all solutions matching the measured data within the noise level would be plausible. This strengthens the requirement of prior information.

In our experiments, we used the emulated single-coil measurements from the NYU fastMRI database [64,65]. The fully sampled measurements were retrospectively subsampled to simulate accelerated MRI data. See Figure 5 for an example of a subsampled measurement, a zero-filled IFFT reconstruction, and the ground truth obtained from the full measurement. We used an acceleration factor of 4, i.e., only 25% of frequencies were kept. Undersampling was performed by selecting 8% of the lowest frequencies and randomly adding higher frequencies until the acceleration factor was reached. The public dataset consists of a training part and a validation part. In total, the training dataset includes 973 volumes (34,742 slices) and the validation dataset includes 199 volumes (7135 slices). Additionally, there is a private test set that consists of 108 volumes (3903 slices). For this private test set, only the undersampled measurements are available, and the models can only be evaluated on the official fastMRI website (https://fastmri.org/, last accessed: 16 November 2021). Our best model can be found on the public leaderboard for "Single-Coil Knee", allowing for comparison with other approaches (our submission is named "cINN v2"). The fastMRI dataset includes scans from two different pulse sequences: coronal proton-density weighting with (PDFS) and without (PD) fat suppression. We trained our models on the full dataset, but used the distinction into PD and PDFS for evaluation on the validation set.

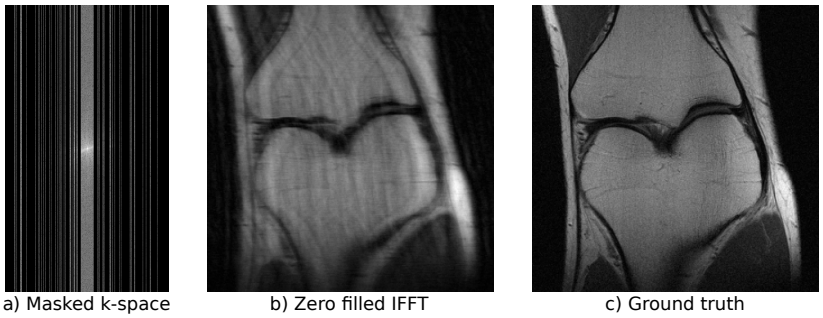


Figure 5. Measurements and reconstruction for the single-coil fastMRI data.

4. Results

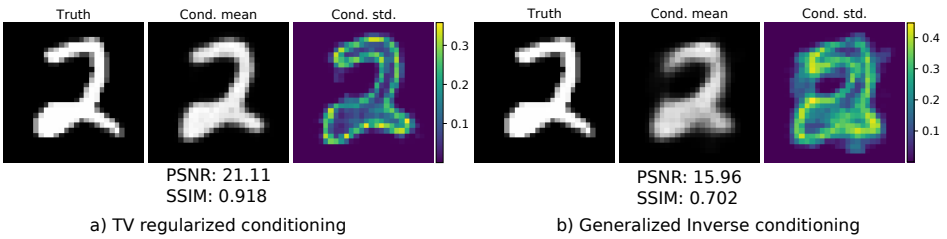
In this section, we present the results of the three different experimental setups. The focus here is on LoDoPaB-CT and fastMRI. For these use cases, we compare different architectures and ablations during training. To assess the performance, we evaluate the peak-signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) [66] on the datasets. The PSNR is strongly related to the mean squared error and expresses the ratio of the maximum possible value to the reconstruction error. In general, a higher PSNR corresponds to a better reconstruction. The SSIM compares the overall image structure, including luminance and contrast, of the reconstruction and the ground-truth image. A detailed definition of the evaluation metrics can be found in Appendix A.3.

4.1. Compressed Sensing

Both models were trained using the Adam optimizer [67] until convergence with a fixed learning rate of 1×10^{-4}. The final model was chosen as the best model regarding the negative log-likelihood on the validation set. The conditional mean was used as reconstruction, and we evaluated both the PSNR and SSIM for the entire test set. The results can be seen in Table 1. The TV-regularized solution as the conditioning input leads to a drastic improvement both in terms of PSNR and SSIM. A visual comparison of one reconstruction is given in Figure 6. One can see that the reconstruction using the TV-regularized solution fits way better to the original ground-truth image. In addition, the conditioned standard deviation is more centered towards the edges of the number. The reconstruction using the generalized inverse as a conditioning input is much smoother and more blurry. The conditional standard deviation is not so focused on specific features on the image. Lastly, we illustrated samples from both models in Figure 7. The samples drawn from the model using the TV-regularized conditioning input look much more realistic.

Table 1. Mean and standard deviation of the PSNR and SSIM for compressed sensing on the MNIST test dataset. The conditioned mean was computed with 100 samples.

Compressed Sensing on MNIST				
	$\mathcal{A}^\dagger = \mathcal{A}^+$		$\mathcal{A}^\dagger = (\mathcal{A}^T\mathcal{A} + \lambda\nabla^T\nabla)\mathcal{A}^T$	
	PSNR	SSIM	PSNR	SSIM
Multi-scale cINN	17.32 ± 2.05	0.752 ± 0.084	19.89 ± 2.54	0.868 ± 0.063

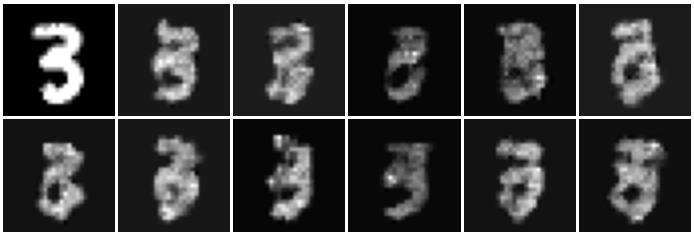


PSNR: 21.11
SSIM: 0.918
a) TV regularized conditioning

PSNR: 15.96
SSIM: 0.702
b) Generalized Inverse conditioning

Figure 6. Conditioned mean and standard deviation for the different inversion layers.

a) Samples from Posterior with TV reconstruction conditioning



b) Samples from Posterior with generalized Inverse conditioning

Figure 7. Samples from the posterior learned by the cINN. The ground-truth sample is shown in the upper-left corner. In (**a**), we used the conditioning based on the TV-regularized reconstruction, and in (**b**), the conditioning was chosen as the generalized inverse. It can be seen that individual samples from the generalized inverse conditioning do not look realistic.

4.2. Computed Tomography

First, we investigate different conditioning networks for the multi-scale architecture. Based on these results, we compare the multi-scale network to the iUNet. The experiments also include variations in the target distribution and the loss function. The results regarding the different conditioning networks can be found in Table 2. The overall results on the LoDoPaB-CT test set are shown in Table 3.

For all comparisons between the multi-scale architecture and iUNet, a unified setting was used. Both networks had a similar size (2.9 Mio. for the iUNet and 3.2 Mio. for the multi-scale architecture). We used five scales for all networks. The inversion model inside the conditioning was the filtered back-projection (FBP). For the iUNet additive coupling layers and for the multi-scale architecture, affine coupling layers were used. Channel permutation after each coupling layer was implemented using fixed 1×1 convolutions [45]. Gradient descent steps with the Adam optimizer [67], an initial learning rate of 1×10^{-4}, and a reduction factor of 0.8 on plateaus were performed during training. The best parameter configuration for each setting was chosen based on the lowest negative log-likelihood on the validation set.

4.2.1. Architecture of Conditioning Network

We tested three different architectures for the conditioning network in the multi-scale cINN model. The first architecture (average pooling) consisted of one initial learned convolutional layer to blow up the number of channels, followed by average pooling operations to reduce the spatial dimensions to the correct size. In the next architecture (CNN), the one initial convolutional layer was replaced by a fully convolutional neural network. The last architecture (ResNet) used residual connections and replaced all average pooling operations with strided convolutional layers. All models were trained using the same initialization with the Adam optimizer [67]. We evaluated all three choices on the LoDoPaB test set, and the results can be seen in Table 2. In our experiments, increasing the complexity of the conditioning network also increased the reconstruction quality in terms of SSIM and PSNR. We suspect that this increase in quality is related to the fact that a more extensive conditioning network can extract a larger amount and more essential features from the conditioning input.

Table 2. Influence of the type of conditioning network for the multi-scale cINN. The mean and standard deviation of the PSNR and SSIM were evaluated on the full LoDoPaB test set using 1000 samples for the cond. mean.

LoDoPaB-CT			
Model	**Cond. Network**	**PSNR**	**SSIM**
Multi-scale	Average Pooling	33.15 ± 3.64	0.806 ± 0.156
	CNN	34.64 ± 4.18	0.826 ± 0.160
	ResNet	35.07 ± 4.34	0.831 ± 0.160

Based on these results, we chose the ResNet conditioning for the following experiments. Note that we reduced the number of parameters of the multi-scale cINN in the other experiments to be comparable with the iUNet model and shorten the time for training. Overall, this has only a minor effect on the reconstruction quality.

4.2.2. Base Distribution

It has been proven that under reasonable conditions for the true density, any base distribution can be used for normalizing flows [29]. However, the question arises of whether some distributions are more suitable than others. We study two different choices for the base distribution: a standard Gaussian distribution used in most flow-based models and a radial Gaussian, as discussed in Section 2.5. As we are interested in the conditional mean in most applications, sample efficiency is vital for the practical implementation and evaluation of a flow-based model.

Table 3 shows mixed results for the different base distributions. While the iUNet benefits from the choice of the radial Gaussian distribution, the performance is worse for the multi-scale model. Nevertheless, the difference in PSNR and SSIM is only minor in this test. However, we could observe a difference in the quality and deviation during the sampling process for a single reconstruction. Networks that were trained with the radial distribution could produce high-quality reconstructions from a single sample. On the other hand, the standard deviation between each sampled reconstruction is significantly smaller than for the models with normal distribution. This can also be seen in the standard deviation plots in Figure 8. Overall, models trained with the radial distribution can use fewer samples for the conditional mean to achieve good reconstructions. In Table 3, we used 100 samples to compute the conditioned mean. Results for 1000 samples can be found in Table A1. Overall, the additional effort of sampling 10 times more data does not justify the small gain in image quality in this experiment.

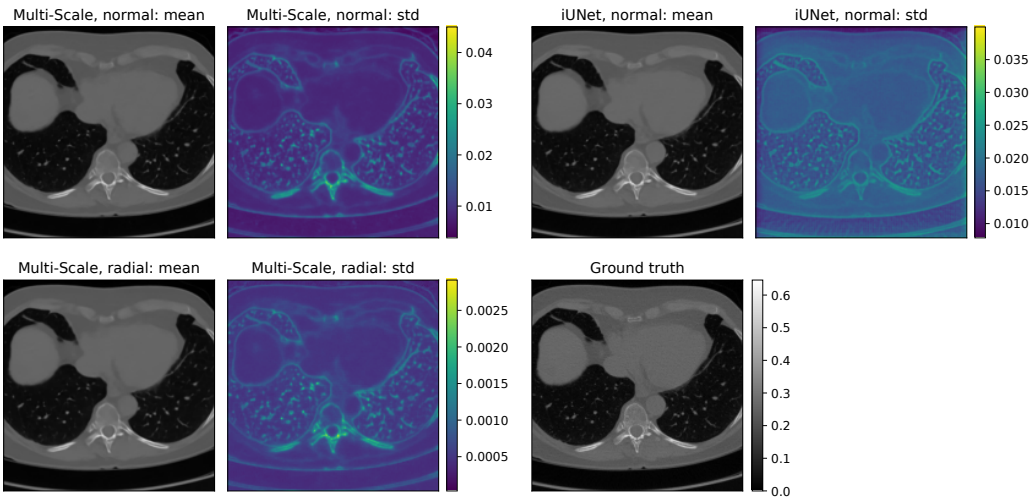


Figure 8. Cond. mean and point-wise standard deviation for the iUNet and the multi-scale architecture on the LoDoPaB-CT data.

Table 3. Mean and standard deviation of the PSNR and SSIM for the LoDoPaB-CT test set. Conditioned mean computed with 100 samples. Unless stated otherwise, training noise was applied and no cond. loss was used.

LoDoPaB-CT				
Model	**Base Distribution**	**Train Noise**	**PSNR**	**SSIM**
Multi-scale	Normal	Yes	34.94 ± 4.24	0.829 ± 0.157
		No	34.92 ± 4.26	0.829 ± 0.158
	Radial	Yes	34.89 ± 4.29	0.823 ± 0.161
		No	34.65 ± 4.25	0.829 ± 0.161
iUNet	Normal	Yes	34.65 ± 4.11	0.805 ± 0.151
		No	34.48 ± 3.96	0.824 ± 0.153
	Radial	Yes	34.58 ± 4.40	0.830 ± 0.158
		No	34.57 ± 4.40	0.830 ± 0.158
Cond. Loss				
iUNet	Normal	Yes	34.88 ± 4.17	0.809 ± 0.148
		No	34.65 ± 4.11	0.805 ± 0.151
	Radial	Yes	34.99 ± 4.39	0.825 ± 0.157
		No	34.58 ± 4.40	0.830 ± 0.158

4.2.3. Training with Additional Noise

In most image datasets, pixel values can only take a specific, discrete range of values. Training a continuous flow-based model on discrete data can lead to artifacts, i.e., the model allocates arbitrary high-likelihood values to the discrete values [68]. In order to circumvent this problem, it is common to add a small amount of noise to the data to get a continuous distribution. This process is called dequantization and, in recent reviews, was done on all image datasets [69]. We found that this problem was not as severe for the medical imaging datasets studied in this paper; e.g., the LoDoPaB-CT dataset already used a dequantization of the discrete HU values. There is, however, a different problem with medical imaging datasets used for image reconstruction. Since there are no real ground-truth images available, high-quality reconstructions are used for training. For LoDoPaB-CT, reconstruction from normal-dose CT measurements and for fastMRI reconstruction from fully sampled MRI measurements are used instead [62,65]. These reconstructions are not free of noise, so we use an additional dequantization step and add random Gaussian noise in the order of the background noise to the training images. As an ablation, we add random Gaussian noise with zero-mean and a variance of 0.005 to the ground-truth images during training. We have chosen these values to correspond to the empirical background noise in the ground-truth images.

In Table 3, results for the multi-scale network and the iUNet with and without additional training noise are shown. For both architectures, the additional noise results in the same or a slightly improved PSNR. Concerning SSIM, the models achieve the same or a marginally lower score with the additional training noise. Overall, due to the high number of images in the dataset (lower overfitting risk) and the existing dequantization, there is no clear benefit from the additional noise in this case.

4.2.4. Training with Conditional Loss

As described in Section 2.9, the final output of the conditional network for the iUNet is in the image domain X. As an ablation, we added a supervised mean squared error loss to the negative log-likelihood term (see Equation (20)) during the training using a weighting factor $\alpha = 1.0$. This additional loss could guide the conditional network to learn more relevant features.

The results for the iUNet are given in the lower part of Table 3. The network benefits from the additional loss on the output of the conditioning network. However, like for all regularization terms, putting too much weight on the conditioning loss interferes with the primary objective of the cINN model. The performance deteriorates in this case. The loss also has a direct impact on the intermediate representations of the conditioning UNet. They shift from feature selection to the reproduction of complete reconstructions. An example is shown in Figure A1 in Appendix A.4.

4.2.5. Sample Refinement

Using cINN, we are able to sample realistic-looking CT reconstruction. However, we have no guarantees that the sample explains the data y, i.e., $AT_\theta(y, z) \approx y$. In order to fulfill this data consistency constraint, we use an additional refinement based on a variational Tikohonov formulation:

$$\hat{x} \in \arg\min_x \|Ax - y\|_2^2 - \lambda \log p_\theta(x|y). \tag{22}$$

We solve for $\hat{x}$ using an iterative scheme and use as initialization our sample $T_\theta(y, z)$ from the cINN. In our experiments, only using the maximum posterior solution as a reconstruction often results in artifacts in the reconstructed image. Therefore, we transitioned to the penalized version in Equation (22). An important topic is the choice of the parameter λ. In Table 4, the results for both the iUNet and the multi-scale architecture are given. Increasing the weighting factor λ from 0 to 1.0 leads to an improvement in terms of PSNR and SSIM for both the multi-scale architecture and the iUNet. However, further increasing the factor λ leads again to a deterioration in most cases.

In total, the reconstruction quality with the sample refinement is worse than for the conditional mean approach. Therefore, we stick to the conditional mean reconstruction technique for the following experiments on the fastMRI dataset.

Table 4. Mean and standard deviation for sample refinement on LoDoPaB for the first 100 samples of the test set. Minimized Equation (22) for 100 iterations with a learning rate 1×10^{-4}. The initial value was one sample from our model $x_0 = T_\theta^{-1}(z, y^\delta)$.

		LoDoPaB-CT	
Model	λ	**PSNR**	**SSIM**
Multi-scale	0	32.02 ± 3.18	0.742 ± 0.135
	0.01	32.10 ± 3.21	0.749 ± 0.137
	0.1	32.56 ± 3.40	0.766 ± 0.142
	1.0	33.03 ± 3.58	0.783 ± 0.148
	10.0	32.97 ± 3.56	0.784 ± 0.149
iUNet	0	32.16 ± 3.12	0.731 ± 0.126
	0.01	32.31 ± 3.19	0.737 ± 0.128
	0.1	32.83 ± 3.41	0.759 ± 0.135
	1.0	32.98 ± 3.45	0.765 ± 0.136
	10.0	32.88 ± 3.40	0.756 ± 0.133

4.3. Magnetic Resonance Imaging

The results for the two architectures, multi-scale and iUNet, for different configurations are presented in Table 5. Example reconstructions and point-wise standard deviations between samples for the best models are shown in Figure 9. For all configurations, the models were trained using the Adam optimizer [67], and the initial learning rate of 1×10^{-4} was reduced by a factor of 0.8 on plateaus. The final model was chosen as the best model regarding the negative log-likelihood on the validation set. As the ground-truth images for the fastMRI test set are not publicly available, we report the PSNR and SSIM on the

validation data in Table 5. Further, following the evaluation in [65], we present the results subdivided into PD and PDFS.

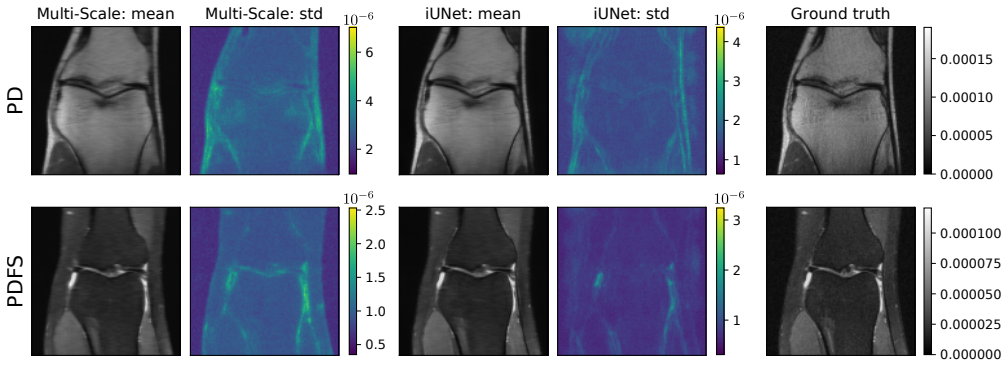


Figure 9. Cond. mean and point-wise standard deviation for the best-performing multi-scale architecture and iUNet on the fastMRI data. Both networks use the radial base distribution and no additional training noise, and the iUNet is trained with conditional loss.

Both networks were implemented such that the number of parameters was comparable (2.5 Mio. for the iUNet and 2.6 Mio. for the multi-scale network). We used five scales for all networks. For the iUNet additive coupling layers and for the multi-scale architecture, affine coupling layers were used. Channel permutation after each coupling layer was implemented using fixed 1×1 convolutions [45]. The conditioning network for the iUNet was based on a UNet architecture. For the multi-scale network, we used an architecture based on a ResNet. Both used the zero-filled IFFT as model-based inversion layer.

Table 5. Mean and standard deviation for the fastMRI dataset. Conditioned mean computed with 100 samples. Unless otherwise specified, no additional training noise and no cond. loss were used.

fastMRI						
Model	**Base Distribution**	**Train Noise**	**PSNR**		**SSIM**	
			PD	**PDFS**	**PD**	**PDFS**
Multi-scale	Normal	Yes	29.15 ± 6.25	23.18 ± 8.20	0.777 ± 0.086	0.536 ± 0.105
		No	28.54 ± 6.52	20.92 ± 9.87	0.776 ± 0.086	0.536 ± 0.105
	Radial	Yes	31.84 ± 3.56	25.76 ± 5.92	0.760 ± 0.092	0.515 ± 0.107
		No	32.07 ± 2.34	26.54 ± 2.73	0.764 ± 0.090	0.522 ± 0.103
iUNet	Normal	No	27.85 ± 1.38	25.76 ± 2.10	0.622 ± 0.052	0.474 ± 0.096
	Radial	No	31.89 ± 2.43	25.94 ± 2.86	0.732 ± 0.107	0.432 ± 0.126
Cond. Loss						
iUNet	Normal	Yes	27.91 ± 1.35	25.83 ± 2.12	0.628 ± 0.054	0.474 ± 0.096
		No	27.85 ± 1.38	25.76 ± 2.10	0.622 ± 0.052	0.474 ± 0.096
	Radial	Yes	31.62 ± 2.26	26.04 ± 2.81	0.730 ± 0.096	0.469 ± 0.110
		No	31.89 ± 2.43	25.94 ± 2.86	0.732 ± 0.107	0.432 ± 0.126

4.3.1. Base Distribution

As with the LoDoPaB dataset, we investigate the influence of the target distribution. The results in Table 5 show that switching from the standard Gaussian distribution to the radial Gaussian leads to an improvement in terms of PSNR for nearly all configurations on the fastMRI dataset. This is in contrast to the observations on the LoDoPaB dataset, where the differences are only minor. However, note that the PSNR and SSIM values for fastMRI are calculated based on the maximum value range of a whole scan, compared to a slice-based choice on LoDoPaB (cf. Appendix A.3). Therefore, the values between the two

experiments cannot be directly compared. Nevertheless, the radial Gaussian appears to be a good choice on fastMRI.

We also undertook a small study to assess the influence of the number of samples (100 vs. 1000) on the reconstruction quality. The results matched with the extensive comparison on the LoDoPaB dataset: The additional sampling contributes little to the image quality while using substantially more computational resources. On fastMRI, we could also observe higher PSNR and SSIM values for single-sample reconstruction from models with radial Gaussian base distribution.

The performance of models trained with the normal Gaussian base distribution highly depends on a sufficient number of samples for the reconstruction. On the other hand, an increase in the number of samples usually leads to an equivalent increase in the computing time.

4.3.2. Training with Additional Noise

We follow the same noising strategy as for the LoDoPaB-CT data and add random Gaussian noise with zero mean and a variance of 0.005 to the ground-truth images during training. For the multi-scale architecture, we observe an improvement for the standard Gaussian and a decline for the radial Gaussian base distributions. We noticed instabilities during the training for the iUNet. Therefore, only values without additional noise are given in Table 5.

4.3.3. Training with Conditional Loss

The results for the inclusion of the conditional loss term are given in the lower part of Table 5. On fastMRI, introducing this additional term to the loss function only gives a slight improvement in terms of PSNR and SSIM. In fact, we also observe a minor deterioration for the iUNet trained using the radial Gaussian base distribution on the PD case for fastMRI.

5. Discussion

In this work, we studied various configurations of conditioned flow-based models on different datasets. The focus of our research was to determine best practices for the use of cINN models for reconstruction tasks in CT and MRI. The two networks used, multi-scale and iUNet, showed comparable performance in many cases. The results demonstrate that a crucial part of the cINN models is the design of the conditioning network. A precise model-based inversion layer and a subsequent extensive neural network can provide diverse features for the CNF. In particular, the model-based layer forms an interesting basis for combining mathematical modeling and data-driven learning. This can go much further than the FBP and Fourier models used here.

The choice of the base distribution also has a significant impact on the model's performance. The radial Gaussian proved to be a valuable alternative to the normal Gaussian distribution, primarily in reducing the reconstruction time by needing fewer samples for the conditioned mean and avoiding common problems with high-dimensional distributions. For the noising during training and the additional conditioning loss, on the other hand, there is no clear recommendation. The additional noise might help on small datasets, where it acts as a data augmentation step. The conditioning loss requires extra tuning of the weighting factor. More promising, therefore, might be the use of a pre-trained reconstruction network whose parameters are frozen for use in cINN.

The experiments also indicated that the training of cINN models does not always run without problems. Although invertible neural networks are analytically invertible, it is possible to encounter instabilities in some situations, and the networks may become numerically non-invertible. Furthermore, in this work, we used the conditional mean as a reconstruction method for most of the experiments. However, other choices are possible. In the following, we will address these topics in more detail.

5.1. Stability

Recently, it was noted that due to stability issues, an extensive invertible neural network could become numerically non-invertible in test time due to rounding errors [42]. We observed this problem when evaluating iUNets with affine coupling layers. In Figure 10, we show the loss during training and an example reconstruction after training. It can be observed that even when the training looks stable, one can get severe artifacts on unknown test images. We did not observe this problem for the multi-scale architecture. Affine coupling layers can have arbitrary large singular values in the inverse Jacobian matrix, which leads to an unstable inverse pass. This effect is known as *exploding inverse* [42]. For increasing stability in the iUNets, we suggest using additive coupling blocks in this architecture.

In addition, the inclusion of additional training noise led to severe instability in our experiments with the iUNet on the fastMRI data. We did not obtain any meaningful reconstructions for this case. In contrast, these issues occurred with neither the multi-scale architecture on fastMRI nor the iUNet on LoDoPaB-CT.

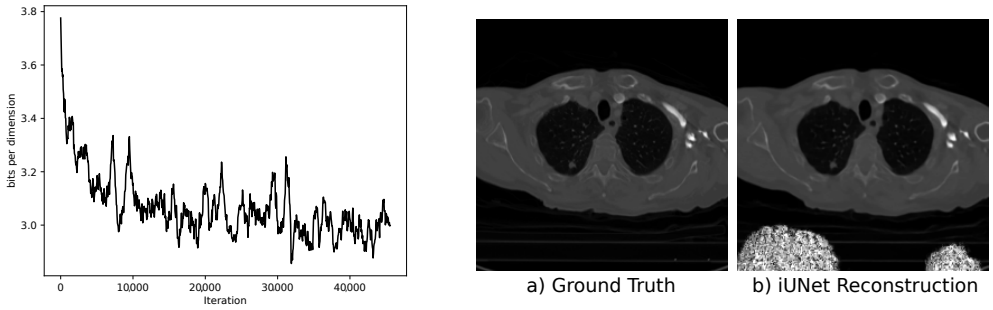


Figure 10. (**Left**) Moving average of loss during training. (**Right**) Ground-truth image from the LoDoPaB test dataset and the corresponding iUNet reconstruction. The pixels in white visualize exploding values in the reconstruction.

5.2. Reconstruction Method

A trained cINN offers us the possibility to explore the full posterior. However, for evaluating the reconstruction quality of our models, we use the conditioned mean as a point estimate. This was also done in prior work for computed tomography reconstruction [22,23], but it would be interesting to explore different choices of estimates. In Section 4.2.5, we evaluated a penalized version of the maximum posterior estimate. For the LoDoPaB-CT dataset, this results in a lower PSNR and SSIM compared to using the conditioned mean. However, one could combine the idea of the conditioned mean and the sample refinement to combine samples that have a low, regularized data discrepancy (cf. Equation (22)).

6. Conclusions

This work explored different architectures and best practices for applying conditional flow-based methods to medical image reconstruction problems in CT and MRI. Our experiments included two popular, invertible network designs. The iUNet [26] architecture is inspired by the UNet [52], which is used extensively in imaging applications. The multi-scale architecture is used in all major normalizing flow frameworks, such as Glow [45] or NICE [36]. The invertible architectures were combined with a conditioning network, which extracts various features from the measurements for the reconstruction process. This cINN framework combines the advantages of memory-efficient invertible networks and normalizing flows for uncertainty estimation with a versatile reconstruction model. Additionally, it provides a direct way to combine model-based and data-driven approaches in a single model.

The use of cINN models for medical image reconstruction is in its beginning stages, and many possible improvements should be explored. We investigated the radial Gaussian distribution as an alternative to the normal Gaussian base distribution. Our experiments

show that it can be beneficial in many cases. A promising next direction is the development of novel invertible network architectures from existing approaches. For applications in medical image reconstruction, state-of-the-art deep learning methods are based on unrolled iterative methods [4]. In [23], an extensive evaluation of the LoDoPaB-CT dataset was performed, and the best-scoring deep learning method was an unrolled learned primal–dual algorithm [3]. These unrolled iterative methods can be made invertible [43,44], but are currently only used for memory-efficient backpropagation. In further work, we want to evaluate whether invertible iterative architectures can be integrated into flow-based models.

Author Contributions: Conceptualization, A.D., M.S., and J.L.; Software, A.D., M.S., and J.L.; supervision, P.M.; project administration, P.M.; writing—original draft preparation, A.D., M.S., and J.L.; writing—review and editing, A.D., M.S., and J.L.; visualization, A.D.; funding acquisition, P.M. All authors reviewed, finalized, and approved the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: J.L., M.S., A.D., and P.M. were funded by the German Research Foundation (DFG; GRK 2224/1). J.L. and M.S. additionally acknowledge support from the DELETO project funded by the Federal Ministry of Education and Research (BMBF, project number 05M20LBB). A.D. further acknowledges support from the Klaus Tschira Stiftung via the project MALDISTAR (project number 00.010.2019).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Radial Density

High-dimensional normal distributions do not behave intuitively, as known from the low-dimensional settings. Sampling from a normal distribution gives mostly samples in the *typical set*. For an n-dimensional normal distribution $\mathcal{N}(\mu, \sigma^2)$, this typical set has a distance of $\sigma\sqrt{n}$ from the expected value μ. In [49], the authors proposed an n-dimensional radial Gaussian density in hyperspherical coordinates where:

- The radius r is distributed according to a half-normal distribution,
- All angular components $\varphi_1, \ldots, \varphi_{n-2} \in [0, \pi], \varphi_{n-1} \in [0, 2\pi]$ are uniformly distributed, yielding equal probability density at every point on the surface of the n-dimensional sphere.

Our derivation of the likelihood closely follows that of [49]. We assume that all dimensions are independently distributed. For the radius r, we get the density:

$$p(r; \theta) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad \text{for } r \geq 0. \tag{A1}$$

Let v be a point on the unit sphere. We want every point on the unit sphere to be equally likely, i.e.,

$$p(v) = \frac{1}{S_n} \quad \text{with } S_n = 2\frac{\pi^{n/2}}{\Gamma(n/2)}, \tag{A2}$$

where S_n is the surface of the n-dimensional unit sphere. We can get the density for the radial components $p(\phi_1, \ldots, \phi_{n-1})$ by solving

$$p(v)\,\mathrm{d}A = \frac{1}{S_n}\,\mathrm{d}A = p(\phi_1, \ldots, \phi_{n-1})\,\mathrm{d}\phi_1 \ldots \mathrm{d}\phi_{n-1}. \tag{A3}$$

Here, $\mathrm{d}A$ is the surface element:

$$\mathrm{d}A = \mathrm{d}\phi_{n-1} \prod_{i=1}^{n-2} \sin(\phi_i)^{n-1-i} \, \mathrm{d}\phi_1 \ldots \mathrm{d}\phi_{n-2}. \tag{A4}$$

Solving (A3) leads us to the density:

$$p(\phi_1, \phi_2, \ldots, \phi_{n-1}) = \frac{1}{S_n} \prod_{i=1}^{n-2} \sin(\phi_i)^{n-1-i}. \tag{A5}$$

Setting $\sigma = 1$ for the radial components gives us the full density in hyperspherical coordinates:

$$p_\epsilon(\epsilon = (r, \phi_1, \phi_2, \ldots, \phi_{n-1})) = \frac{\sqrt{2}}{\sqrt{\pi}} \exp(-r^2/2) \frac{1}{S_n} \prod_{i=1}^{n-2} \sin(\phi_i)^{n-1-i}. \tag{A6}$$

For our experiments, we are always working in Cartesian coordinates, so one has to do a final transformation $x = f(\epsilon)$ and use the change-of-variables theorem. The Jacobian of the n-dimensional spherical coordinate transformation is known:

$$\left| \det \left(\frac{\partial f(\epsilon)}{\partial \epsilon} \right) \right| = r^{n-1} \prod_{i=1}^{n-2} \sin(\phi_i)^{n-1-i}. \tag{A7}$$

Finally, we get

$$\begin{aligned} p_x(x) &= \frac{\sqrt{2}}{\sqrt{\pi} S_n} \exp(-r^2/2) \prod_{i=1}^{n-2} \sin(\phi_i)^{n-1-i} \left(r^{n-1} \prod_{i=1}^{n-2} \sin(\phi_i)^{n-1-i} \right)^{-1} \\ &= \frac{\sqrt{2}}{\sqrt{\pi} S_n r^{n-1}} \exp(-r^2/2) \\ &= \frac{\sqrt{2}}{\sqrt{\pi} S_n \|x\|^{n-1}} \exp(-\|x\|^2/2) \end{aligned} \tag{A8}$$

as our radial Gaussian density.

Appendix A.2. Architecture for the Compressed Sensing Example

The multi-scale architecture used for the compressed sensing experiments used in Section 3.1 consists of two learnable downsampling operations, each followed by a conditional coupling block. After a flatten layer, a last dense conditional coupling block is used.

cINN	Output size
Learnable downsampling	$4 \times 14 \times 14$
Level 1 conditional section	$4 \times 14 \times 14$
Learnable downsampling	$16 \times 7 \times 7$
Level 2 conditional section	$16 \times 7 \times 7$
Flatten	784
Split: 656 to output	128
Level 3 dense-conditional section	128

In the conditional coupling section, we use an affine coupling layer and implement the scale s and translation t using a small convolutional neural network with either 1×1

convolution or 3×3 convolutions. After each affine coupling layer, we use a fixed 1×1 convolution to permute the dimensions. For the dense coupling section, we use a simple random permutation of the dimensions and affine coupling layers with dense subnetworks s and t.

Conditional section	
Affine coupling (CNN-subnet with 1×1 kernel)	$8\times$
1×1 convolution	
Affine coupling (CNN-subnet with 3×3 kernel)	
1×1 convolution	

Dense conditional section	
Random permutation	$3\times$
Affine coupling (Dense-subnetwork)	

Appendix A.3. Evaluation Metrics

Appendix A.3.1. Peak-Signal-to-Noise Ratio

The peak-signal-to-noise ratio (PSNR) is measured by a log-scaled version of the mean squared error (MSE) between the reconstruction $\hat{x}$ and the ground-truth image $x^\dagger$. The PSNR expresses the ratio between the maximum possible image intensity and the distorting noise

$$\mathrm{PSNR}\left(\hat{x}, x^\dagger\right) := 10 \log_{10}\left(\frac{L^2}{\mathrm{MSE}(\hat{x}, x^\dagger)}\right), \quad \mathrm{MSE}\left(\hat{x}, x^\dagger\right) := \frac{1}{n}\sum_{i=1}^{n}\left|\hat{x}_i - x_i^\dagger\right|^2$$

In general, higher PSNR values are an indication of a better reconstruction. The maximum image value L can be chosen in different ways. For the MNIST and the LoDoPab-CT dataset, we compute the value per slice as $L = \max(x^\dagger) - \min(x^\dagger)$. For evaluation on the fastMRI dataset, we choose L as the maximum value per patient, i.e., per 3D volume.

Appendix A.3.2. Structural Similarity

The structural similarity (SSIM) [66] compares the overall image structure of the ground truth and reconstruction. It is based on assumptions about human visual perception. Results lie in the range $[0, 1]$, with higher values being better. The SSIM is computed through a sliding window at M locations

$$\mathrm{SSIM}\left(\hat{x}, x^\dagger\right) := \frac{1}{M}\sum_{j=1}^{M}\frac{\left(2\hat{\mu}_j\mu_j + C_1\right)\left(2\Sigma_j + C_2\right)}{\left(\hat{\mu}_j^2 + \mu_j^2 + C_1\right)\left(\hat{\sigma}_j^2 + \sigma_j^2 + C_2\right)}.$$

Here, $\hat{\mu}_j$ and μ_j are the average pixel intensities, $\hat{\sigma}_j$ and σ_j are the variances, and Σ_j is the covariance of $\hat{x}$ and $x^\dagger$ at the j-th local window. Constants $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$ stabilize the division. Just as with the PSNR metric, the maximum image value L can be chosen in different ways. We use the same choices as specified in the previous section.

Appendix A.4. Additional Figures

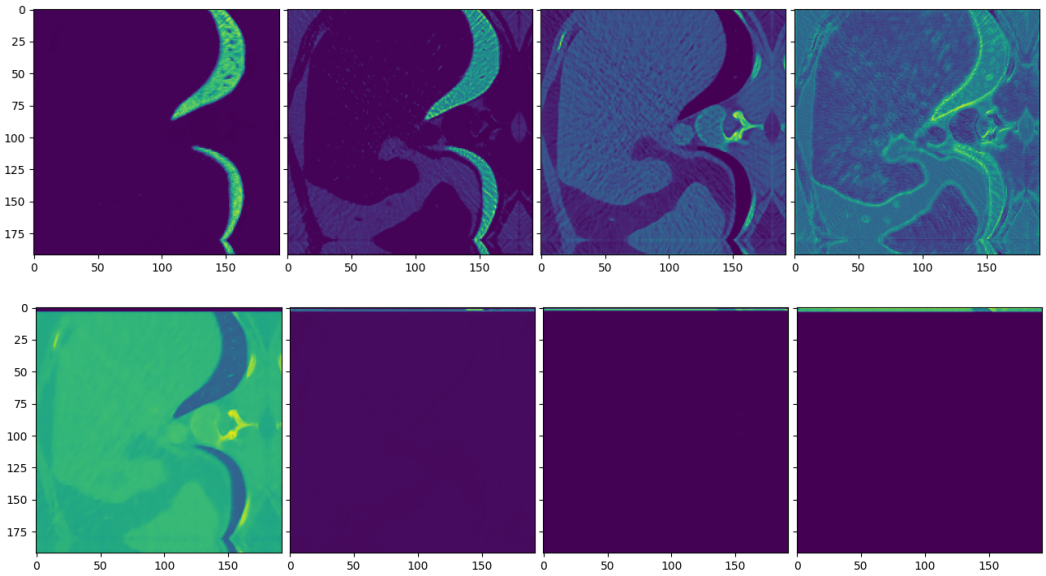


Figure A1. Intermediate activations from a single layer in a conditioning UNet model. (**Top**) cINN Model trained with just the log-likelihood. (**Bottom**) cINN model trained with an additional loss for the conditioning UNet. One can observe that the conditioning network focuses on different parts of the image if no special loss is used. Otherwise, it produces activations that are close to the final reconstruction. In addition, there are many empty activations.

Appendix A.5. Additional Results

Table A1. Mean and standard deviation of the PSNR and SSIM for the LoDoPaB-CT test set. Conditioned mean computed with 1000 samples. Unless stated otherwise, training noise was applied and no cond. loss was used.

LoDoPaB-CT				
Model	**Base Distribution**	**Train Noise**	**PSNR**	**SSIM**
Multi-scale	Normal	Yes	34.99 ± 4.26	0.830 ± 0.158
		No	34.97 ± 4.28	0.830 ± 0.157
	Radial	Yes	34.89 ± 4.29	0.823 ± 0.161
		No	34.65 ± 4.25	0.829 ± 0.161
iUNet	Normal	Yes	34.69 ± 4.13	0.806 ± 0.151
		No	34.98 ± 4.19	0.823 ± 0.148
	Radial	Yes	34.75 ± 4.23	0.819 ± 0.153
		No	34.57 ± 4.40	0.830 ± 0.158
Cond. Loss				
iUNet	Normal	Yes	34.92 ± 4.19	0.810 ± 0.148
		No	34.69 ± 4.13	0.806 ± 0.151
	Radial	Yes	34.99 ± 4.39	0.825 ± 0.159
		No	34.75 ± 4.23	0.819 ± 0.153

References

1. Arridge, S.; Maass, P.; Öktem, O.; Schönlieb, C.B. Solving inverse problems using data-driven models. *Acta Numer.* **2019**, *28*, 1–174. [CrossRef]
2. Jin, K.H.; McCann, M.T.; Froustey, E.; Unser, M. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process.* **2017**, *26*, 4509–4522. [CrossRef] [PubMed]
3. Adler, J.; Öktem, O. Learned Primal-Dual Reconstruction. *IEEE Trans. Med. Imaging* **2018**, *37*, 1322–1332. [CrossRef]

4. Adler, J.; Öktem, O. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Probl.* **2017**, *33*, 124007. [CrossRef]
5. Lunz, S.; Schönlieb, C.; Öktem, O. Adversarial Regularizers in Inverse Problems. In Proceedings of the Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems (NeurIPS 2018), Montreal, QC, Canada, 3–8 December 2018.
6. Zhu, B.; Liu, J.Z.; Cauley, S.F.; Rosen, B.R.; Rosen, M.S. Image reconstruction by domain-transform manifold learning. *Nature* **2018**, *555*, 487–492. [CrossRef] [PubMed]
7. Tarantola, A.; Valette, B. Inverse problems = quest for information. *J. Geophys.* **1982**, *50*, 159–170.
8. Kaipio, J.; Somersalo, E. *Statistical and Computational Inverse Problems*; Springer: New York, NY, USA, 2005; Volume 160. [CrossRef]
9. Martin, J.; Wilcox, L.C.; Burstedde, C.; Ghattas, O. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM J. Sci. Comput.* **2012**, *34*, A1460–A1487. [CrossRef]
10. Sunnåker, M.; Busetto, A.G.; Numminen, E.; Corander, J.; Foll, M.; Dessimoz, C. Approximate bayesian computation. *PLoS Comput. Biol.* **2013**, *9*, e1002803. [CrossRef]
11. Rezende, D.J.; Mohamed, S.; Wierstra, D. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. In Proceedings of the 31th International Conference on Machine Learning (ICML 2014), Beijing, China, 21–26 June 2014; Xing, E.P., Jebara, T., Eds.; PMLR: Cambridge, MA, USA, 2014; Volume 32, pp. 1278–1286.
12. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. In Proceedings of the 2nd International Conference on Learning Representations (ICLR 2014), Banff, AB, Canada, 14–16 April 2014.
13. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, Montreal, QC, Canada, 8–13 December 2014.
14. Tabak, E.G.; Turner, C.V. A family of nonparametric density estimation algorithms. *Commun. Pure Appl. Math.* **2013**, *66*, 145–164. [CrossRef]
15. Barbano, R.; Zhang, C.; Arridge, S.; Jin, B. Quantifying model uncertainty in inverse problems via bayesian deep gradient descent. In Proceedings of the 2020 IEEE 25th International Conference on Pattern Recognition (ICPR), Virtual Event, 10–15 January 2021; pp. 1392–1399. [CrossRef]
16. Adler, J.; Öktem, O. Deep Bayesian Inversion. *arXiv* **2018**, arXiv:1811.05910.
17. Ardizzone, L.; Kruse, J.; Rother, C.; Köthe, U. Analyzing Inverse Problems with Invertible Neural Networks. In Proceedings of the 7th International Conference on Learning Representations (ICLR 2019), New Orleans, LA, USA, 6–9 May 2019.
18. Ardizzone, L.; Lüth, C.; Kruse, J.; Rother, C.; Köthe, U. Guided Image Generation with Conditional Invertible Neural Networks. *arXiv* **2019**, arXiv:1907.02392.
19. Ardizzone, L.; Kruse, J.; Lüth, C.; Bracher, N.; Rother, C.; Köthe, U. Conditional Invertible Neural Networks for Diverse Image-to-Image Translation. In Proceedings of the Pattern Recognition (DAGM GCPR 2020), Tübingen, Germany, 28 September–1 October 2020; Akata, Z., Geiger, A., Sattler, T., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 373–387. [CrossRef]
20. Andrle, A.; Farchmin, N.; Hagemann, P.; Heidenreich, S.; Soltwisch, V.; Steidl, G. Invertible Neural Networks Versus MCMC for Posterior Reconstruction in Grazing Incidence X-Ray Fluorescence. In Proceedings of the Scale Space and Variational Methods in Computer Vision—8th International Conference (SSVM 2021), Virtual Event, 16–20 May 2021; Elmoataz, A., Fadili, J., Quéau, Y., Rabin, J., Simon, L., Eds.; Volume 12679 Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2021; pp. 528–539. [CrossRef]
21. Anantha Padmanabha, G.; Zabaras, N. Solving inverse problems using conditional invertible neural networks. *J. Comput. Phys.* **2021**, *433*, 110194. [CrossRef]
22. Denker, A.; Schmidt, M.; Leuschner, J.; Maass, P.; Behrmann, J. Conditional Normalizing Flows for Low-Dose Computed Tomography Image Reconstruction. In Proceedings of the ICML Workshop on Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models, Vienna, Austria, 18 July 2020.
23. Leuschner, J.; Schmidt, M.; Ganguly, P.S.; Andriiashen, V.; Coban, S.B.; Denker, A.; Bauer, D.; Hadjifaradji, A.; Batenburg, K.J.; Maass, P.; et al. Quantitative Comparison of Deep Learning-Based Image Reconstruction Methods for Low-Dose and Sparse-Angle CT Applications. *J. Imaging* **2021**, *7*, 44. [CrossRef] [PubMed]
24. Hagemann, P.; Hertrich, J.; Steidl, G. Stochastic Normalizing Flows for Inverse Problems: A Markov Chains Viewpoint. *arXiv* **2021**, arXiv:2109.11375.
25. Dinh, L.; Sohl-Dickstein, J.; Bengio, S. Density estimation using Real NVP. In Proceedings of the 5th International Conference on Learning Representations (ICLR 2017), Toulon, France, 24–26 April 2017.
26. Etmann, C.; Ke, R.; Schönlieb, C.B. iUNets: Learnable invertible up-and downsampling for large-scale inverse problems. In Proceedings of the 2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP), Espoo, Finland, 21–24 September 2020; pp. 1–6. [CrossRef]
27. Van den Oord, A.; Kalchbrenner, N.; Kavukcuoglu, K. Pixel Recurrent Neural Networks. In Proceedings of the 33nd International Conference on Machine Learning (ICML 2016), New York City, NY, USA, 19–24 June 2016; Balcan, M., Weinberger, K.Q., Eds.; PMLR: Cambridge, MA, USA, 2016; Volume 48, pp. 1747–1756.

28. Van den Oord, A.; Kalchbrenner, N.; Espeholt, L.; Kavukcuoglu, K.; Vinyals, O.; Graves, A. Conditional Image Generation with PixelCNN Decoders. In Proceedings of the Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, Barcelona, Spain, 5–10 December 2016.
29. Papamakarios, G.; Nalisnick, E.T.; Rezende, D.J.; Mohamed, S.; Lakshminarayanan, B. Normalizing Flows for Probabilistic Modeling and Inference. *arXiv* **2019**, arXiv:1912.02762.
30. Brock, A.; Donahue, J.; Simonyan, K. Large Scale GAN Training for High Fidelity Natural Image Synthesis. In Proceedings of the 7th International Conference on Learning Representations (ICLR 2019), New Orleans, LA, USA, 6–9 May 2019.
31. Song, Y.; Sohl-Dickstein, J.; Kingma, D.P.; Kumar, A.; Ermon, S.; Poole, B. Score-Based Generative Modeling through Stochastic Differential Equations. In Proceedings of the 9th International Conference on Learning Representations (ICLR 2021), Virtual Event, 3–7 May 2021.
32. Wu, H.; Köhler, J.; Noé, F. Stochastic Normalizing Flows. In Proceedings of the Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020 (NeurIPS 2020), Virtual, 6–12 December 2020.
33. Ruthotto, L.; Haber, E. An introduction to deep generative modeling. *GAMM-Mitteilungen* **2021**, *44*, e202100008. [CrossRef]
34. Dashti, M.; Stuart, A.M. The Bayesian Approach to Inverse Problems. In *Handbook of Uncertainty Quantification*; Springer International Publishing: Cham, Switzerland, 2017; pp. 311–428. [CrossRef]
35. Asim, M.; Daniels, M.; Leong, O.; Ahmed, A.; Hand, P. Invertible generative models for inverse problems: Mitigating representation error and dataset bias. In Proceedings of the 37th International Conference on Machine Learning (ICML 2020), Virtual Event, 13–18 July 2020; Daumé, H., III, Singh, A., Eds.; PMLR: Cambridge, MA, USA, 2020; Volume 119, pp. 399–409.
36. Dinh, L.; Krueger, D.; Bengio, Y. NICE: Non-linear Independent Components Estimation. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, 7–9 May 2015.
37. Rezende, D.J.; Mohamed, S. Variational Inference with Normalizing Flows. In Proceedings of the 32nd International Conference on Machine Learning (ICML 2015), Lille, France, 6–11 July 2015; Bach, F.R., Blei, D.M., Eds.; PMLR: Cambridge, MA, USA, 2015; Volume 37, pp. 1530–1538.
38. Behrmann, J.; Grathwohl, W.; Chen, R.T.Q.; Duvenaud, D.; Jacobsen, J. Invertible Residual Networks. In Proceedings of the 36th International Conference on Machine Learning (ICML 2019), Long Beach, CA, USA, 9–15 June 2019; Chaudhuri, K., Salakhutdinov, R., Eds.; PMLR: Cambridge, MA, USA, 2019; Volume 97, pp. 573–582.
39. Chen, T.Q.; Behrmann, J.; Duvenaud, D.; Jacobsen, J. Residual Flows for Invertible Generative Modeling. In Proceedings of the Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019.
40. Van den Berg, R.; Hasenclever, L.; Tomczak, J.M.; Welling, M. Sylvester Normalizing Flows for Variational Inference. In Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence (UAI 2018), Monterey, CA, USA, 6–10 August 2018; Globerson, A., Silva, R., Eds.; AUAI Press: Arlington, VA, USA, 2018; pp. 393–402.
41. Gomez, A.N.; Ren, M.; Urtasun, R.; Grosse, R.B. The Reversible Residual Network: Backpropagation Without Storing Activations. In Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017.
42. Behrmann, J.; Vicol, P.; Wang, K.; Grosse, R.B.; Jacobsen, J. Understanding and Mitigating Exploding Inverses in Invertible Neural Networks. In Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS 2021), Virtual Event, 13–15 April 2021; Banerjee, A., Fukumizu, K., Eds.; PMLR: Cambridge, MA, USA, 2021; Volume 130, pp. 1792–1800.
43. Rudzusika, J.; Bajic, B.; Öktem, O.; Schönlieb, C.B.; Etmann, C. Invertible Learned Primal-Dual. In Proceedings of the NeurIPS 2021 Workshop on Deep Learning and Inverse Problems, Online, 13 December 2021.
44. Putzky, P.; Welling, M. Invert to Learn to Invert. In Proceedings of the Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019.
45. Kingma, D.P.; Dhariwal, P. Glow: Generative Flow with Invertible 1x1 Convolutions. In Proceedings of the Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems (NeurIPS 2018), Montreal, QC, Canada, 3–8 December 2018.
46. Jacobsen, J.; Smeulders, A.W.M.; Oyallon, E. i-RevNet: Deep Invertible Networks. In Proceedings of the 6th International Conference on Learning Representations (ICLR 2018), Vancouver, BC, Canada, 30 April–3 May 2018.
47. Nalisnick, E.; Matsukawa, A.; Teh, Y.W.; Gorur, D.; Lakshminarayanan, B. Do Deep Generative Models Know What They Don't Know? In Proceedings of the 7th International Conference on Learning Representations (ICLR 2019), New Orleans, LA, USA, 6–9 May 2019.
48. Nalisnick, E.T.; Matsukawa, A.; Teh, Y.W.; Lakshminarayanan, B. Detecting Out-of-Distribution Inputs to Deep Generative Models Using a Test for Typicality. *arXiv* **2019**, arXiv:1906.02994.
49. Farquhar, S.; Osborne, M.A.; Gal, Y. Radial Bayesian neural networks: Beyond discrete support in large-scale Bayesian deep learning. In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS 2020), Virtual Event, 26–28 August 2020; PMLR: Cambridge, MA, USA, 2020; Volume 108, pp. 1352–1362.
50. Hagemann, P.; Neumayer, S. Stabilizing Invertible Neural Networks Using Mixture Models. *arXiv* **2020**, arXiv:2009.02994.
51. Winkler, C.; Worrall, D.E.; Hoogeboom, E.; Welling, M. Learning Likelihoods with Conditional Normalizing Flows. *arXiv* **2019**, arXiv:1912.00042.

52. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), Munich, Germany, 5–9 October 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
53. LeCun, Y.; Cortes, C.; Burges, C.C.J. The MNIST Handwritten Digit Database. Available online: http://yann.lecun.com/exdb/mnist/ (accessed on 30 April 2020).
54. Genzel, M.; Macdonald, J.; März, M. Solving Inverse Problems with Deep Neural Networks—Robustness Included? *arXiv* **2020**, arXiv:2011.04268.
55. Radon, J. On the determination of functions from their integral values along certain manifolds. *IEEE Trans. Med. Imaging* **1986**, *5*, 170–176. [CrossRef] [PubMed]
56. Buzug, T.M. *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*; Springer: Berlin/Heidelberg, Germany, 2008. [CrossRef]
57. Nashed, M.Z. A new approach to classification and regularization of ill-posed operator equations. In *Inverse and Ill-Posed Problems*; Engl, H.W., Groetsch, C.W., Eds.; Academic Press: Cambridge, MA, USA, 1987; pp. 53–75. [CrossRef]
58. Natterer, F. *The Mathematics of Computerized Tomography*; SIAM: Philadelphia, PA, USA, 2001. [CrossRef]
59. Gordon, R.; Bender, R.; Herman, G.T. Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography. *J. Theor. Biol.* **1970**, *29*, 471–481. [CrossRef]
60. Sidky, E.Y.; Pan, X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Phys. Med. Biol.* **2008**, *53*, 4777. [CrossRef] [PubMed]
61. Bubba, T.A.; Galinier, M.; Lassas, M.; Prato, M.; Ratti, L.; Siltanen, S. Deep Neural Networks for Inverse Problems with Pseudodifferential Operators: An Application to Limited-Angle Tomography. *SIAM J. Imaging Sci.* **2021**, *14*, 470–505. [CrossRef]
62. Leuschner, J.; Schmidt, M.; Baguer, D.O.; Maass, P. LoDoPaB-CT, a benchmark dataset for low-dose computed tomography reconstruction. *Sci. Data* **2021**, *8*, 109. [CrossRef]
63. Doneva, M. Mathematical models for magnetic resonance imaging reconstruction: An overview of the approaches, problems, and future research areas. *IEEE Signal Process. Mag.* **2020**, *37*, 24–32. [CrossRef]
64. Knoll, F.; Zbontar, J.; Sriram, A.; Muckley, M.J.; Bruno, M.; Defazio, A.; Parente, M.; Geras, K.J.; Katsnelson, J.; Chandarana, H.; et al. fastMRI: A Publicly Available Raw k-Space and DICOM Dataset of Knee Images for Accelerated MR Image Reconstruction Using Machine Learning. *Radiol. Artif. Intell.* **2020**, *2*, e190007. [CrossRef]
65. Zbontar, J.; Knoll, F.; Sriram, A.; Murrell, T.; Huang, Z.; Muckley, M.J.; Defazio, A.; Stern, R.; Johnson, P.; Bruno, M.; et al. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. *arXiv* **2019**, arXiv:1811.08839.
66. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
67. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, 7–9 May 2015.
68. Uria, B.; Murray, I.; Larochelle, H. RNADE: The real-valued neural autoregressive density-estimator. In Proceedings of the Advances in Neural Information Processing Systems 26: Annual Conference on Neural Information Processing Systems 2013, Lake Tahoe, NV, USA, 5–8 December 2013.
69. Kobyzev, I.; Prince, S.; Brubaker, M. Normalizing flows: An introduction and review of current methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3964–3979. [CrossRef]