LEARNING THE INVISIBLE: A HYBRID DEEP LEARNING-SHEARLET FRAMEWORK FOR LIMITED ANGLE COMPUTED TOMOGRAPHY

TATIANA A. BUBBA, GITTA KUTYNIOK, MATTI LASSAS, MAXIMILIAN MÄRZ, WOJCIECH SAMEK, SAMULI SILTANEN, AND VIGNESH SRINIVASAN

ABSTRACT. The high complexity of various inverse problems poses a significant challenge to model-based reconstruction schemes, which in such situations often reach their limits. At the same time, we witness an exceptional success of data-based methodologies such as deep learning. However, in the context of inverse problems, deep neural networks mostly act as black box routines, used for instance for a somewhat unspecified removal of artifacts in classical image reconstructions. In this paper, we will focus on the severely ill-posed inverse problem of limited angle computed tomography, in which entire boundary sections are not captured in the measurements. We will develop a hybrid reconstruction framework that fuses model-based sparse regularization with data-driven deep learning. Our method is *reliable* in the sense that we only learn the part that can provably not be handled by model-based methods, while applying the theoretically controllable sparse regularization technique to the remaining parts. Such a decomposition into *visible* and *invisible* segments is achieved by means of the shearlet transform that allows to resolve wavefront sets in the phase space. Furthermore, this split enables us to assign the clear task of inferring unknown shearlet coefficients to the neural network and thereby offering an *interpretation* of its performance in the context of limited angle computed tomography. Our numerical experiments show that our algorithm significantly surpasses both pure model- and more data-based reconstruction methods.

1. INTRODUCTION

Due to increased computational power and advanced mathematical understanding, there is a growing interest in solving severely ill-posed inverse problems. The goal is to recover an unknown quantity from indirect measurements, where typically only few of them are acquired and the reconstruction process is highly sensitive to modelling errors and noise. Traditional inversion methods are based on complementing the insufficient and corrupted measurement data by mathematical models, which impose *a priori* information on the solutions. Such methods include Tikhonov regularization, Bayesian inversion, and inversion algorithms based on partial differential equations or applied harmonic analysis.

However, sometimes the ill-posedness renders it very difficult to robustly recover specific parts of the target. A prominent example is the inverse problem of limited angle computed tomography (limited angle CT), where the missing part of the wavefront set of the target can be read off the measurement geometry [72, 26, 67]. In some medical applications, it is enough to consider slices of the reconstruction where the stable part of the wavefront set reliably provides the clinically important boundaries of tissues. For example, in [73] the spatial position of microcalcifications in the breast can be recovered, and the slice considered in [51] provides a low-dose X-ray examination for dental implant planning. However, any new method that is able to recover the missing part of the wavefront set more reliably would improve the quality of those reconstructions and lead to unprecedented applications of limited angle tomography.

(T. A. Bubba, M. Lassas, S. Siltanen) DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF HELSINKI, 00014 HELSINKI, FINNLAND

(G. Kutyniok) DEPARTMENT OF MATHEMATICS AND DEPARTMENT OF ELECTRICAL ENGINEERING & COMPUTER SCIENCE, TECHNISCHE UNIVERSITÄT BERLIN, 10623 BERLIN, GERMANY

(M. März) DEPARTMENT OF MATHEMATICS, TECHNISCHE UNIVERSITÄT BERLIN, 10623 BERLIN, GERMANY

(W. Samek, V. Srinivasan) DEPARTMENT OF VIDEO CODING AND ANALYTICS, FRAUNHOFER HEINRICH HERTZ INSTITUTE, 10587 BERLIN, GERMANY

E-mail addresses: `tatiana.bubba@helsinki.fi, kutyniok@math.tu-berlin.de, matti.lassas@helsinki.fi, maerz@math.tu-berlin.de, wojciech.samek@hhi.fraunhofer.de, samuli.siltanen@helsinki.fi, vignesh.srinivasan@hhi.fraunhofer.de`.

Currently, we witness a tremendous success of data-based methodologies such as deep neural networks for a wide range of problems, for example, speech recognition [39], the game of Go [80] or image classification [52] and many more. The underlying philosophy is agnostic in the sense that no explicit data model is specified, but vast amounts of training data are used to infer an implicit proxy. During the last years, also the area of inverse problems is increasingly impacted by machine learning approaches, in particular, by deep learning (see, e.g., [86, 10, 47, 44, 3, 36]). However, at this time, neural networks are mostly used as black boxes that are for instance trained for an unspecific image enhancement of direct inversions, or for a replacement of iteration steps in optimization algorithms.

In this paper, we develop a framework for solving the inverse problem of limited angle CT by combining model-based sparse regularization using shearlets with a data-driven deep neural network approach. The key idea of our hybrid method goes back to Quinto's fundamental visibility analysis of limited angle CT based on *microlocal analysis* [72]. We utilize sparse regularization with shearlets for splitting the (wavefront set of the) data into a *visible part,* recoverable by classical model-based methods, and an *invisible part*, that is provably not contained in the measured data. Precisely this part is sought to be recovered by an inference in the shearlet domain by means of a trained neural network. Such an estimation of unknown shearlet coefficients is highly dependent on a faithful model-based reconstruction of its visible counterpart. Therefore, the focus of our work is on a moderate missing wedge, i.e., where at most as much information needs to be inferred as it is available to classical methods on the visible part.

1.1. **Shearlets and Sparse Regularization.** Given an ill-posed inverse problem $y = \mathcal{R} f + \eta$, where $\mathcal{R} : X \to Y$ with suitable spaces X and Y, and η models measurement noise, Tikhonov regularization provides an approximate solution $f^\lambda \in X$, $\lambda > 0$, by minimizing the functional

$$J_\lambda(f) := \| \mathcal{R} f - y \|^2 + \lambda \cdot \mathcal{P}(f), \quad f \in X,$$

with $\mathcal{P}(f)$ being a penalty term, promoting desired properties in the solution f^λ. Sparse regularization is then based on the common paradigm that for each class of data, there exists a sparsifying representation system [16, 13, 19, 22]. In the considered situation, one would assume that there exists a system $(\psi_\mu)_\mu \subseteq X$ such that the sparsity promoting ℓ^1-norm of the coefficient vector $(\langle f, \psi_\mu \rangle)_\mu$ is small, therefore allowing to choose the regularization term as

$$\mathcal{P}(f) = \|(\langle f, \psi_\mu \rangle)_\mu\|_1.$$

Let us now focus on inverse problems in imaging. In this situation, wavelet systems [61] are suboptimal, since it is known that – due to their isotropic nature – they are not capable of providing optimally sparse approximations of images under the well-accepted assumption that images are governed by edges, hence anisotropic features. *Shearlets* [56, 54] are representation systems specifically designed for multivariate data that are optimally adapted to such anisotropic structures. As such they can be seen as a further development of *curvelets* [12], which were the first system that allowed to provide optimally sparse approximations of *cartoon-like images* - a mathematical abstraction of real-world images. Shearlets build upon the same ideas, however, they additionally offer the benefit of an unified treatment of the continuous and discrete situation allowing for faithful implementations [54]. Shearlets have already been very successfully applied to various inverse problems, such as denoising [21], CT [15], phase retrieval [59] or inverse scattering [55].

To be a bit more precise, the elements of a *shearlet system* $\{\psi_{j,k,\boldsymbol{m}}\}_{(j,k,\boldsymbol{m}) \in \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}^2}$ are parametrized by a scale parameter j, a directional parameter k, and a parameter for the position $\boldsymbol{m}$. In the sequel, we will denote the associated transform by $\mathrm{SH}_\psi(f)$, i.e.,

$$\mathrm{SH}_\psi(f) = (\langle f, \psi_{j,k,\boldsymbol{m}} \rangle)_{(j,k,\boldsymbol{m}) \in \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}^2}$$

(for more details see Section 2.2.2). Similarly, a *continuous shearlet system* $\{\psi_{a,s,\boldsymbol{t}}\}_{(a,s,\boldsymbol{t}) \in \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}^2}$ can be defined by considering continuous indexing parameters. One striking property of the continuous version of shearlets – which will be crucial for our approach – is their ability to resolve the wavefront set of generalized functions [30, 53]. Roughly speaking, a wavefront set consists of the positions of the singular support of a generalized function f together with their directions, and is a subset of the so-called *phase space* $\mathbb{R}^2 \times \mathbb{P}_1$. Considering the decay properties of the shearlet coefficients $(\langle f, \psi_{a,s,\boldsymbol{t}} \rangle)_{(a,s,\boldsymbol{t}) \in \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}^2}$ as $a \to 0$ yields precisely those position-direction pairs $(\boldsymbol{t}, s)$ which constitute the wavefront set of f.

1.2. **Neural Networks and Inverse Problems.** Artificial neural networks were originally introduced in 1943 by McCulloch and Pitts as an approach to develop learning algorithms by mimicking the human brain [64]. Their main goal at that time was the development of a theoretical approach to artificial intelligence. However, the limited amount of data and the lack of high performance computers prevented the training of networks with many layers.

By now these two obstacles are overcome and we have massive amounts of training data as well as a tremendously increased computing power available, thereby allowing the training of deep neural networks. This is one of the reasons why neural networks have recently seen such a spectacular comeback with impressive performance results in applications such as game playing (AlphaGo), image classification, speech recognition, to name a few [80, 39, 52]. From a mathematical perspective, a deep neural network in an idealized form is a high-dimensional function $\mathcal{NN} : \mathbb{R}^n \to \mathbb{R}^d$ of the form

$$\mathcal{NN}(\boldsymbol{x}) = W_L(\sigma(W_{L-1}(\sigma(\ldots(\sigma(W_1(\boldsymbol{x})))\ldots)))), \tag{1.1}$$

with the W_j being affine-linear functions and $\sigma : \mathbb{R} \to \mathbb{R}$ being the (non-linear) activation function applied componentwise. In a nutshell, the goal of *deep learning* is to approximate an (unknown) structural relation between the input space $\mathbb{R}^n$ and the output space $\mathbb{R}^d$ with $\mathcal{NN}$. This task is achieved by determining the affine-linear functions W_j from the knowledge of training examples $(\boldsymbol{x}_i, \boldsymbol{y}_i)_{i=1}^N \subseteq \mathbb{R}^n \times \mathbb{R}^d$ following the underlying relation.

One should stress that various special cases exist, with convolutional neural networks (CNNs) [27, 57] being the most prominent architectures in the context of imaging. However, most of the related research is performance driven, while developing a mathematical foundation mostly plays a secondary role. Despite the lack of a complete theoretical understanding, deep learning is currently penetrating various areas of applied mathematics. This is in particular true for the area of inverse problems in imaging sciences where sophisticated model-based approaches, which used to be the previous state of the art, are now outperformed.

Some approaches train a deep neural network directly for the inversion from noisy measurements $\boldsymbol{y} = \mathcal{R}\boldsymbol{f} + \boldsymbol{\eta}$, based on a collection of training samples $(\boldsymbol{y}_i, \boldsymbol{f}_i)_{i=1}^N$ following the considered forward model, e.g., [69] for CT and [86] for denoising and inpainting[1]. Other typical works aim at explicitly incorporating knowledge about the forward model $\mathcal{R}$ into the reconstruction process. This is for instance achieved by preprocessing the measurements $\boldsymbol{y}_i$ with a model-based inversion, e.g., [44, 47, 70] for CT, or [84] in the case of magnetic resonance imaging. Finally, recent approaches insert deep networks into iterative reconstruction schemes, for instance by unrolling the steps and casting them as a network [29], or by replacing some of the proximal operators by a CNN [4, 87, 65]. Still, all these methodologies have in common that the entire reconstructed image has undergone transformations by one or more neural networks, during which control over the applied modifications might have been lost. We refer the interested reader to the reviews [3, 62] for a more detailed discussion on the use of deep neural networks in the context of inverse problems.

1.3. **A Bit of History: Limited Angle Computed Tomography.** Limited angle tomography appears frequently in practical applications, such as dental tomography [50], damage detection in concrete structures [38], breast tomosynthesis [89], or electron tomography [5]. Given the original image $f \in L^1(\mathbb{R}^2)$, a simplified, mathematical model of the general tomographic data acquisition process is given by the *Radon transform*

$$\mathcal{R}f(\theta, s) = \int_{L(\theta,s)} f(\boldsymbol{x}) dS(\boldsymbol{x}), \tag{1.2}$$

where $\theta \in [-\pi/2, \pi/2)$, $s \in \mathbb{R}$ and

$$L(\theta, s) := \left\{ \boldsymbol{x} \in \mathbb{R}^2 : x_1 \cos(\theta) + x_2 \sin(\theta) = s \right\}$$

denotes the line with normal direction θ and distance s to the origin, with dS being the 1-dimensional Lebesgue measure along it; e.g., [23, 67]. The underlying geometric setup is sketched in Figure 1.

Due to physical constraints on the measurement device, the Radon transform $\mathcal{R}f$ is often not known on the entire angular range $\theta \in [-\pi/2, \pi/2)$, but only on a subinterval $[-\phi, \phi]$ with $\phi < \pi/2$. We will

---

[1]In order to highlight the difference between continuous objects and their discretizations, we will use boldface letters to denote matrices and finite-dimensional vectors in $\mathbb{R}^n$ ($n \geq 2$).
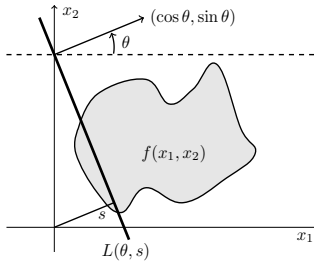
Figure 1: Geometric setup of the Radon transform defined in Equation (1.2).

indicate such a *missing wedge* by the notation

$$\mathcal{R}_\phi f := \mathcal{R} f\big|_{[-\phi,\phi]\times\mathbb{R}}.$$

The task of *limited angle CT* is to recover an approximation of f from its noisy measurements

$$y = \mathcal{R}_\phi f + \eta, \tag{1.3}$$

where η models deterministic and/or random measurement errors. There exists an abundance of inversion strategies for (limited angle) CT and we will now briefly review some of the most relevant ones for our work.

Due to the limited angular range, not all features of the measured object f are captured under $\mathcal{R}_\phi$ [72] and the resulting inverse problem is severely ill-posed [17]. Therefore, classical methods, in particular the filtered backprojection (FBP), are known to yield suboptimal performance, although still being very popular in applications, mostly due to computational performance.

In the case of low-dose CT, it has been demonstrated that sparse regularization methods allow for accurate reconstructions from fewer tomographic measurements than usually required by standard methods such as the FBP [79, 45, 75, 78, 46]. Often total variation (TV), which enforces gradient sparsity, is used as a simple but very effective prior, but also wavelets [60, 73, 49], curvelets [11, 25] and shearlets [15, 9, 33] have been successfully applied. However, to the best of our knowledge, shearlets have not yet been considered for ℓ^1-regularization in limited angle CT. Although advanced sparsity-based variational schemes define effective regularization methods, the amount of missing data in limited angle CT is typically so severe that certain features remain impossible to reconstruct and streaking artifacts appear [25].

There already exist a few approaches to exploit deep learning for solving the limited angle CT problem. All of the three following methods have in common that they are essentially based on a direct inversion, followed by a "denoising" procedure - as such, one of the most straightforward ways to tackle an inverse problem. In [88], a shallow convolutional network is trained to remove artifacts in FBP reconstructions. Additionally to the postprocessing with a variational network, [34] makes use of a second neural network for correcting inhomogeneities in the projection domain. Intriguing results are achieved in the work of Gu and Ye [31]: similar as in [44], a so-called U-Net CNN [76] is trained to improve the FBP reconstruction. However, based on the insight that the artifacts in limited angle tomography posses a directional nature, the CNN processes the directional wavelet coefficients of the FBP image. While all of these three methods yield impressive results given the substantial amount of missing projections, potential drawbacks can be summarized as follows:

- The remarkable post-processing capabilities of neural networks come with a flavor of *alchemy*: a somewhat unspecified removal of artifacts in the FBP reconstructions can be observed. However, it remains unclear to what extent the resulting image has been modified by a CNN and therefore how reliable the reconstructions are. We regard this as particularly critical for medical applications.
- While post-processing FBP data is computationally attractive, it might not be an optimal choice in terms of reconstruction quality, since the FBP solution is heavily contaminated with artifacts and potentially blurry - a flaw that might be amplified by post-processing with a U-Net architecture [35].
- The advantage of regularizing with an anisotropic system, which allows for an extraction of visibility information, is not fully exploited.

1.4. **Our Contribution.** The main objective of our approach is to design a reconstruction framework for limited angle CT, where deep learning is solely applied to those parts of the inverse problem that are provably not contained in the measured data. This will ensure a maximal amount of reliability and interpretability of our results. Interestingly, such an hybrid approach clearly outperforms previous data-based methods.

Let us now foremost focus on edge information, which in the distributional situation refers to the wavefront set of an image f. The fundamental visibility analysis for limited angle CT by Quinto [72] allows to distinguish which singularities can be accurately reconstructed and which are not contained in the measurements. Put simply, the dividing criteria is whether an edge is tangent to an acquired line $L(\theta, s)$ or not (see also Theorem 2.2 and Visibility Principle 2.3). Singularities belonging to the first type are referred to as *visible* while the others are *invisible*. We can conclude that some parts of the wavefront set can be robustly recovered – the *visible part* – and some parts not – the *invisible part*. Thus, a complete recovery of all singularities of f could be regarded as an inpainting problem on its wavefront set. Our original intention for this work was to *handcraft* a variational prior that promotes the completion of the gaps during the reconstruction. Although such rules are quite intuitive, their mathematical formalization turned out to be surprisingly difficult. Consequently, the goal of our work is to estimate the invisible part by applying a deep neural network that is specifically trained for this task. Such an inference of invisible information from the knowledge of its visible counterpart is feasible, since the wavefront sets of typical images follows similar structural patterns in the phase space.

As mentioned before, we intend to apply advanced model-based methods for a recovery of the reliable boundary information. Thus, the first step of our algorithm consists in solving a sparse regularization problem, which is conceptually of the following form (see Section 3.1 and Algorithm 4.1 for more details):
Step 1 - Recover the Visible:
$$\boldsymbol{f}^* := \operatorname{argmin}_{\boldsymbol{f}} \frac{1}{2} \|\boldsymbol{\mathcal{R}}_\phi \boldsymbol{f} - \boldsymbol{y}\|_2^2 + \lambda \cdot \|\mathbf{SH}_\psi(\boldsymbol{f})\|_1$$
Thereby, $\boldsymbol{\mathcal{R}}_\phi$ and $\mathbf{SH}_\psi$ denote finite dimensional approximations of their continuous version $\mathcal{R}$ and SH_ψ. Promoting sparsity with the ℓ^1-norm in the shearlet domain allows to characterize the visible parts of the boundaries. Indeed, similar as in [25], we observe that there exists a partition of the shearlet parameter set $\Lambda = \mathcal{I}_{\mathtt{vis}} \cup \mathcal{I}_{\mathtt{inv}}$ approximately satisfying the following:

- for $(j, k, \boldsymbol{m}) \in \mathcal{I}_{\mathtt{inv}}$: $\mathbf{SH}_\psi(\boldsymbol{f}^*)_{(j,k,\boldsymbol{m})} \approx 0$,
- for $(j, k, \boldsymbol{m}) \in \mathcal{I}_{\mathtt{vis}}$: $\mathbf{SH}_\psi(\boldsymbol{f}^*)_{(j,k,\boldsymbol{m})} \approx \mathbf{SH}_\psi(\boldsymbol{f})_{(j,k,\boldsymbol{m})}$.

The shearlet coefficient tensor $\mathbf{SH}_\psi(\boldsymbol{f}^*)$ resembles the previously discussed phase space, in which the missing wedge of limited angle CT causes gaps in the wavefront set. Inpainting them can be rephrased as an estimation of the shearlet coefficients associated with $\mathcal{I}_{\mathtt{inv}}$ - a task that shall be accomplished by an artificial deep neural network.

The network is trained to generate an estimation of the invisible shearlet coefficients $\mathbf{SH}_\psi(\boldsymbol{f})_{\mathcal{I}_{\mathtt{inv}}}$, when the coefficient tensor $\mathbf{SH}_\psi(\boldsymbol{f}^*)$ is given as input.
Step 2 - Learn the Invisible (LtI):

$$\mathcal{N}\mathcal{N}: \ \mathbf{SH}_\psi(\boldsymbol{f}^*) \ \longrightarrow \ \text{[neural network]} \ \longrightarrow \ \boldsymbol{F}\left(\overset{!}{\approx} \mathbf{SH}_\psi(\boldsymbol{f})_{\mathcal{I}_{\mathtt{inv}}}\right)$$

The architecture, referred to as *PhantomNet* – a network that learns phantom-like[2] coefficients; see Section 4.3.3 and Figure 7 for details – is a modified U-Net CNN [76], which is a popular choice in the field of inverse problems, e.g. [47, 44].

Having an estimation of the invisible part of the wavefront set at hand, the final step consists in fusing both parts and mapping the output back to the image domain via the inverse shearlet transform:
Step 3 - Combine both Parts:
$$\boldsymbol{f}_{\mathtt{LtI}} = \mathbf{SH}_\psi^{-1}\left(\mathbf{SH}_\psi(\boldsymbol{f}^*)_{\mathcal{I}_{\mathtt{vis}}} + \boldsymbol{F}\right)$$

---

[2]*"Phantom:* Something apparently seen, heard, or sensed, but having no physical reality"; definition from [1].

Concluding, the deep neural network is only used to infer the invisible shearlet coefficients, hence to estimate only the truly invisible boundary information. The visible part is entirely treated by the well understood method of sparse regularization with shearlets, which increases the overall reliability of our reconstructions. By assigning a clear task to the neural network, namely estimating invisible edge information, we gain a deeper understanding of our hybrid reconstruction framework. Furthermore, our network takes a rather accurate reconstruction of the visible coefficients as input, making the estimation of the invisible information easier. Additionally, the central question of how well our results generalize to unrelated testing data is only relevant on the invisible part. These advantages however come with a grain of salt due to the computational complexity of ℓ_1-minimization, which is dominating the running time of our approach.

1.5. **Expected Impact.** We anticipate our results to have the following impacts:

- *Limited Angle CT.* We propose a reconstruction framework that allows to complete the gaps in the wavefront set caused by the missing wedge of limited angle CT. We demonstrate that deep neural networks are capable of inferring the invisible parts in the shearlet domain.
- *Hybrid Methods.* Our numerical experiments in Section 5 show that our hybrid method outperforms both, traditional model-based reconstruction schemes and more data-oriented methods. Thus, it supports the often advocated strategy to "take the best out of both worlds", and gives evidence to the potential of such combined approaches.
- *Interpretable Deep Learning.* Our results reveal a possibility of utilizing deep learning in a more controlled manner by applying it precisely to the part – here coined the "invisible" part – which defies any model-based approach. In this sense, the reconstruction method allows for a comprehensible interpretation, where machine learning is only used for inferring lost information. If a theory for inpainting with deep learning became available, our framework might allow for a transfer of these results to limited angle CT.

One should also stress that this concept might also be applicable to other inverse problems, predominantly those with a substantial amount of missing or distorted data.

1.6. **Outline.** The paper is organized as follows. Section 2 is devoted to reviewing the theoretical background of limited angle CT and the shearlet transform. In Section 3, we detail a key idea of our approach, namely the decomposition of the data (in the phase space) into a visible and an invisible part by means of ℓ^1-regularization with shearlets. Our algorithmic approach that infers the invisible wavefront information by a deep neural network is introduced in Section 4, where we also give some background information on deep learning and discuss our network architecture PhantomNet. Finally, we demonstrate the performance of our methodology by a series of numerical experiments (see Section 5). Concluding remarks and future perspectives are briefly summarized in Section 6.

2. Theoretical Background

In this section, we summarize the theoretical concepts that are essential for our proposed recovery framework. We first discuss results from microlocal analysis that explain which edge information is available in the acquired data and then introduce the reader to shearlets, which will be the key ingredient to access the visible information.

2.1. **Visibility of Singularities in Limited Angle CT.** There is a body of work based on microlocal analysis that gives a precise description which singularities are visible in the limited data and which singularities cannot be determined [72, 26, 68]. Since these insights will be central for our proposed reconstruction architecture we will briefly summarize some of those in the following. First, the notion of wavefront sets is required, which allows to simultaneously describe the location and direction of a singularity of a function f. It is based on a localized correspondence of smoothness and rapid decay in the Fourier domain.
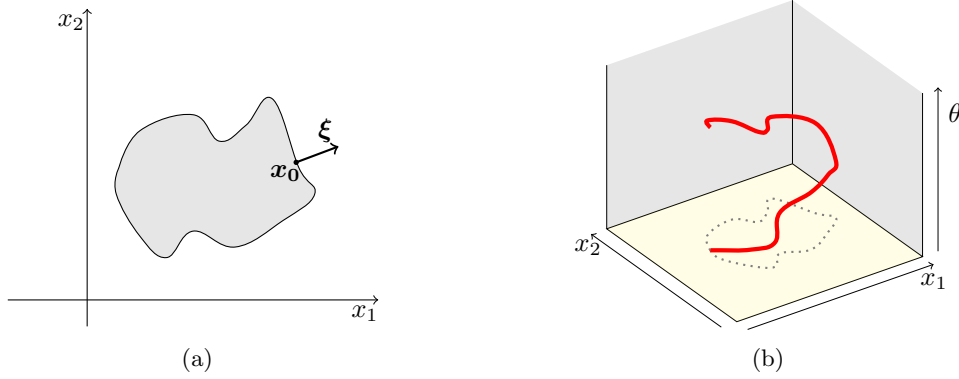
Figure 2: (a) Visualization of a point $(\boldsymbol{x}_0, \boldsymbol{\xi}_0) \in \mathrm{WF}(f)$ where $f = \mathbb{I}_D$ for a set $D \subseteq \mathbb{R}^2$ with smooth boundary. Such indicator functions $\mathbb{I}_D$ are examples of conormal distributions [41] of which the wavefront set $\mathrm{WF}(f)$ is contained in the conormal bundle $N^*(S)$ of a smooth surface S (or a curve in the 2-dimensional case). The bundle $N^*(S)$ consists of the points $(\boldsymbol{x}, \boldsymbol{\xi})$ where $\boldsymbol{x} \in S$ and $\boldsymbol{\xi}$ is normal to S. For a curve $S \subset \mathbb{R}^2$ the conormal bundle $N^*(S)$ is a 2-dimensional submanifold of the 4-dimensional space $T^*(\mathbb{R}^2)$, that is, the wavefront set of f is contained in a smooth, low dimensional subset of the phase space.

Definition 2.1 ([40]). Let $f \in L^2_{\mathrm{loc}}(\mathbb{R}^2)$, i.e., f is square integrable on every compact subset of $\mathbb{R}^2$. Let $\boldsymbol{x}_0 \in \mathbb{R}^2$ and $\boldsymbol{\xi} \in \mathbb{R}^2 \setminus \{\boldsymbol{0}\}$. Then f is said to be *smooth at $\boldsymbol{x}_0$ in the direction $\boldsymbol{\xi}$*, if there exists a smooth cut-off function $\phi \in C_c^\infty(\mathbb{R}^2)$ such that $\phi(\boldsymbol{x}_0) \neq 0$ and an open cone $V_{\boldsymbol{\xi}} \subseteq \mathbb{R}^2$ containing $\boldsymbol{\xi}$ such that given $N \in \mathbb{N}$ there exists a C_N with

$$\left| \widehat{\phi \cdot f}(\boldsymbol{\xi}) \right| \leq C_N (1 + \|\boldsymbol{\xi}\|)^{-N},$$

for all $\boldsymbol{\xi} \in V_{\boldsymbol{\xi}}$, where $\widehat{\phi \cdot f}$ denotes the Fourier transform of the product $\phi \cdot f$. Furthermore, the *wavefront set of f* is defined by

$$\mathrm{WF}(f) := \left\{ (\boldsymbol{x}_0, \boldsymbol{\xi}) \in \mathbb{R}^2 \times \mathbb{R}^2 : f \text{ is not smooth at } \boldsymbol{x}_0 \text{ in the direction } \boldsymbol{\xi} \right\}.$$

The wavefront set is a subset of the cotagent space $T^*(\mathbb{R}^2)$.

The wavefront set is often visualized in the *phase space*, i.e., in the set of position-orientation pairs $(\boldsymbol{x}_0, \theta)$, where $\boldsymbol{x}_0 \in \mathbb{R}^2$ and θ is in the real projective space $\mathbb{P}_1$ in $\mathbb{R}^2$ (freely identified with $[0, \pi)$). $\mathrm{WF}(f)$ can be seen as a subset of the phase space, encoding the positions and directions in which f is non-smooth, see Figure 2 for a visualization of a simple example.

The previous description of positions and directions of singularities, in the sense of Definition 2.1, allows for a characterization of their visibility in computed tomography.

Theorem 2.2 ([72, 71]). *Let $f \in L^2_{\mathrm{loc}}(\mathbb{R}^2)$ and $L_0 = L(\theta_0, s_0)$ be a line in the plane. Let $(\boldsymbol{x}_0, \boldsymbol{\xi}) \in \mathrm{WF}(f)$ such that $\boldsymbol{x}_0 \in L_0$ and $\boldsymbol{\xi} \in \mathbb{R}^2$ is a normal vector to L_0. Then the following holds.*

  (i) The singularity of f at $(\boldsymbol{x}_0, \boldsymbol{\xi})$ causes a unique singularity in $\mathrm{WF}(\mathcal{R} f)$ at (θ_0, s_0).
  (ii) Singularities of f not tangent to $L(\theta_0, s_0)$ do not cause singularities in $\mathcal{R} f$ at (θ_0, s_0).

The implications of the previous theorem for limited angle CT are colloquially summarized in the following general principle [71]:

Visibility Principle 2.3

  i) A singularity of f that is tangent to a line contained in the limited angle data set should be "easy" to reconstruct. We will refer to such boundaries as *visible*.
  ii) Singularities of f that are not tangent to a line in the limited angle data set should be impossible to reconstruct. They are referred to as *invisible*.

Let us point out that the information which boundaries are (in-)visible is completely determined by the measurement setup, i.e., by the sampled angular range $[-\phi, \phi]$. Therefore it is known *a priori* and can be used for reconstruction purposes.

2.2. **Shearlets.** In the following, we will give a brief introduction to shearlet frames and discuss their properties when used as an ℓ_1-regularizer for limited angle CT. For the sake of readability we aim at conveying the general ideas and refer the interested reader to the cited literature for more details and precise formulations of the described results.

2.2.1. *The Continuous Shearlet Transform.* The basic construction of shearlets is based on applying three different operations to a well chosen *generator function* $\psi \in L^2(\mathbb{R}^2)$, obtaining elements of the form

$$\psi_{a,s,\boldsymbol{t}} = |\det M_{as}|^{1/2} \cdot \psi\left(M_{as}(\cdot - \boldsymbol{t})\right),$$

where $\boldsymbol{t} \in \mathbb{R}^2$ encodes *translations* and $(a, s) \in \mathbb{R}_+ \times \mathbb{R}$ controls the *parabolic scaling matrix* A_a and the *shearing matrix* S_s via the composite matrix

$$M_{as} := A_a^{-1} S_s^{-1} = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}^{-1} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}^{-1}.$$

The *continuous shearlet transform* is then defined as the mapping

$$L^2(\mathbb{R}^2) \ni f \mapsto \mathcal{SH}_\psi f(a, s, \boldsymbol{t}) = \langle f, \psi_{a,s,\boldsymbol{t}} \rangle, \quad (a, s, \boldsymbol{t}) \in \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}^2.$$

Thus, $\mathcal{SH}_\psi$ analyzes the function f around the location $\boldsymbol{t}$ at different resolutions and orientations encoded by the scale and shearing parameters a and s, respectively. The (asymptotic) orientation of such a shearlet $\psi_{a,s,\boldsymbol{t}}$ is visualized in Figure 3(a). Throughout this work, it is assumed that $\hat{\psi}$ has compact support such as in the case of a classical shearlet in [53].

The continuous shearlet transform has become a well studied research object in the last decade: it turns out that it exhibits an unwanted directional bias, which is circumvented by considering the so-called *cone-adapted* shearlet transform (see also next section). Furthermore, it can be shown that, under mild conditions, the shearlet system forms a continuous frame, implying for instance that a reconstruction of f from its shearlet transform is possible [30].

Of particular importance for our work are the results of [30, 53], in which it is shown that the continuous shearlet transform allows resolving the wavefront set of distributions by analyzing the decay properties of the continuous shearlet transform. Due to the cone-adaption, the precise statements are of rather technical nature and we will only give the general principle here: assume that $\boldsymbol{\xi} \in \mathbb{R}^2 \setminus \{\boldsymbol{0}\}$ with $\xi_2/\xi_1 \in [-1, 1]$. It can be shown that f is smooth at $\boldsymbol{x_0}$ in the direction $\boldsymbol{\xi}$, if and only if there is an open neighbourhood U of $(\xi_2/\xi_1, \boldsymbol{x_0})$ such that

$$|\mathcal{SH}_\psi f(a, s, \boldsymbol{t})| \in \mathcal{O}(a^k), \quad \text{as } a \to 0,$$

for all $k \in \mathbb{N}$, with the $\mathcal{O}(\cdot)$-term uniform over $(s, \boldsymbol{t})$ in U. Similar results hold true for other directions $\boldsymbol{\xi} \in \mathbb{R}^2 \setminus \{\boldsymbol{0}\}$. Overall we can summarize that:

> *The wavefront set of f is resolved by distinguishing different decay rates of its continuous shearlet transform.*

2.2.2. *The Discrete Shearlet System.* Starting from the continuous transform, the goal is now to derive a discrete system of functions that allows to encode anisotropic features in the digital realm. Such a *discrete shearlet system* can be formally obtained by sampling the parameter space $\mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}^2$ on a discrete subset. The so-called *regular discrete shearlet system* associated with $\psi \in L^2(\mathbb{R}^2)$ is defined by

$$\text{SH}(\psi) := \left\{ \psi_{j,k,\boldsymbol{m}} = 2^{(3j)/4} \psi(S_k A_{2^j} \cdot - \boldsymbol{m}), \text{ for } j, k \in \mathbb{Z}, \boldsymbol{m} \in \mathbb{Z}^2 \right\}. \tag{2.1}$$

Furthermore, the *discrete shearlet transform* is defined as the mapping

$$L^2(\mathbb{R}^2) \ni f \mapsto \text{SH}_\psi f(j, k, \boldsymbol{m}) = \langle f, \psi_{j,k,\boldsymbol{m}} \rangle, \quad (j, k, \boldsymbol{m}) \in \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}^2.$$

We point out that the essence of shearlets lies in the fact that the shearing matrix S_k preserves the integer lattice for $k \in \mathbb{Z}$, which is desired for a numerical implementation. Under mild assumptions on
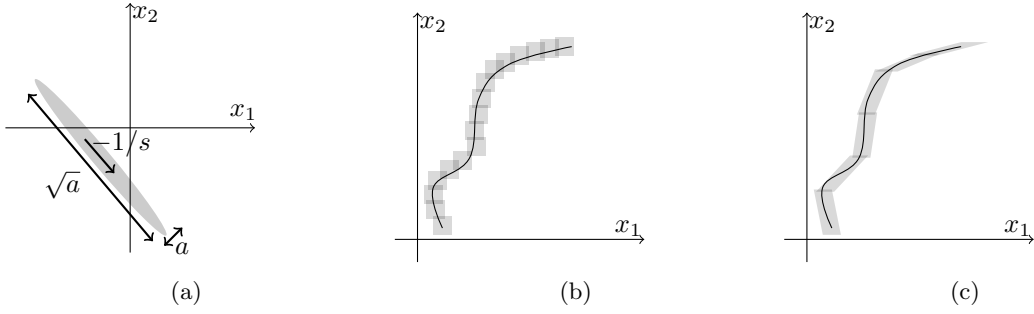
Figure 3: (a) Visualization of the (asymptotic) orientation of a shearlet $\psi_{a,s,t}$ for $a \to 0$. (b) shows a covering of a boundary section with isotropic wavelet elements [61] and (c) with anisotropic shearlet elements.

the generator ψ, it can be shown that the system $\mathrm{SH}(\psi)$ forms a *Parseval frame* [14] of $L^2(\mathbb{R}^2)$, giving rise to the shearlet representation

$$f = \sum_{(j,k,\boldsymbol{m}) \in \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}^2} \langle f, \psi_{j,k,\boldsymbol{m}} \rangle \, \psi_{j,k,\boldsymbol{m}} = \mathrm{SH}_\psi^T(\mathrm{SH}_\psi(f)). \tag{2.2}$$

Similar as in the case of the continuous shearlet transform, the discrete system also exhibits an unwanted *directional bias*. This side effect is visualized in Figure 4(a), where the Fourier domain support of various elements in $\mathrm{SH}(\psi)$ corresponding to different values of j and k is shown. By partitioning the Fourier space into vertical and horizontal conic regions, denoted by $\mathcal{C}_v$ and $\mathcal{C}_h$ respectively, together with a separate low frequency part $\mathcal{L}$, a more uniform tiling is achieved, see Figure 4(b). This is reflected in the following basic definition of *cone-adapted discrete shearlet systems*.

Definition 2.4. Let $\phi, \psi \in L^2(\mathbb{R}^2)$. Then the *cone-adapted discrete shearlet system* is defined by

$$\mathrm{SH}(\phi, \psi) = \Phi(\phi) \cup \Psi(\psi) \cup \tilde{\Psi}(\tilde{\psi}),$$

where

$$\Phi(\phi) := \left\{ \psi_{0,0,\boldsymbol{m},0} := \phi(\cdot - \boldsymbol{m}) : \boldsymbol{m} \in \mathbb{Z}^2 \right\},$$

$$\Psi(\psi) := \left\{ \psi_{j,k,\boldsymbol{m},1} := 2^{(3j)/4} \psi(S_k A_j \cdot - \boldsymbol{m}) : j \in \mathbb{N}_0, k \in \mathbb{Z}, |k| \le 2^{\lceil j/2 \rceil}, \boldsymbol{m} \in \mathbb{Z}^2 \right\},$$

$$\tilde{\Psi}(\tilde{\psi}) := \left\{ \psi_{j,k,\boldsymbol{m},-1} := 2^{(3j)/4} \tilde{\psi}(S_k^T \tilde{A}_j \cdot - \boldsymbol{m}) : j \in \mathbb{N}_0, k \in \mathbb{Z}, |k| \le 2^{\lceil j/2 \rceil}, \boldsymbol{m} \in \mathbb{Z}^2 \right\},$$

with $\tilde{\psi}(x_1, x_2) := \psi(x_2, x_1)$ and $\tilde{A}_j = \mathrm{diag}(2^{j/2}, 2^j) \in \mathbb{R}^{2 \times 2}$. For ease of notation we introduce the index set

$$\Lambda := \{(j, k, \boldsymbol{m}, \iota) : j \in \mathbb{N}_0, k \in \mathbb{Z}, |\iota|j \ge j \ge 0, |k| \le |\iota|2^{\lceil j/2 \rceil}, \boldsymbol{m} \in \mathbb{Z}^2, \iota \in \{1, 0, -1\}\}.$$

The *cone-adapted discrete shearlet transform* is then defined as the mapping

$$L^2(\mathbb{R}^2) \ni f \mapsto \mathrm{SH}_{\psi,\phi} \, f(j, k, \boldsymbol{m}, \iota) = (\langle f, \psi_{j,k,\boldsymbol{m},\iota} \rangle)_{(j,k,\boldsymbol{m},\iota) \in \Lambda}.$$

In the previous definition, the function ψ is referred to as *shearlet generator*. Its corresponding systems $\Psi(\psi)$ and $\tilde{\Psi}(\tilde{\psi})$ essentially differ in the reversed roles of the input variables and therefore correspond to the horizontal and vertical conic region, respectively. Note that by restricting the range for the shearing variable k on each cone, the orientations of the resulting functions are distributed more equally. This can be seen in the Fourier tiling of the cone-adapted system, which is depicted in Figure 4(c). Finally, ϕ is referred to as the *shearlet scaling function* and it is associated to the low frequency part $\mathcal{L}$, since it is chosen to have compact frequency support near the origin.

There are various extensions and refinements of this basic definition available in the literature, notably for instance the construction of compactly supported shearlets. We refrain from giving further details and refer the interested reader to [54] and references therein. Instead, we conclude this excursion by stating a stylized approximation theorem for discrete shearlet frames, which shows that shearlets are an optimal sparsifying transform for a particular class of natural images.
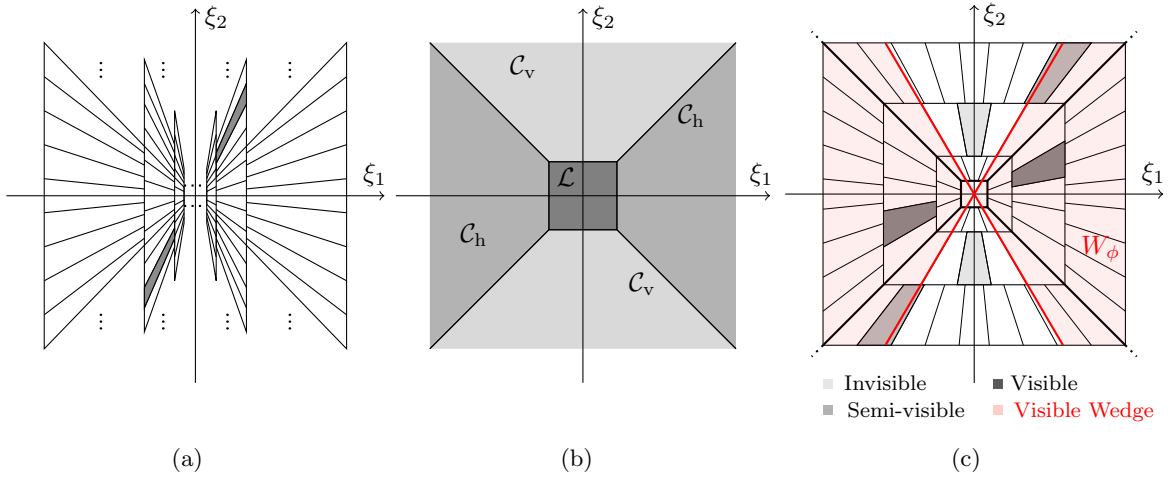
Figure 4: Illustration of tilings in the Fourier domain. (a) shows the frequency support of elements in the regular shearlet system for different values of j and k and thereby reveals a directional bias. (b) indicates how the Fourier domain is separated in two conic regions and a low frequency part. In (c), the tiling of the cone-adapted discrete shearlet system is visualized. Additionally, the visible wedge W_ϕ of Definition 3.1 is shown in red, together with examples for (in-)visible shearlets (see Section 3.2).

Theorem 2.5 ([32]). *Let* $\mathrm{SH}(\phi, \psi)$ *be a variant of the cone-adapted shearlet system (see* [32] *for precise definition). Let* $f \in L^2(\mathbb{R}^2)$ *be a* cartoon-like *function, i.e.,* $f = f_0 + f_1 \cdot \mathbb{I}_B$, *where* $B \subseteq [0,1]^2$ *is a set with* ∂B *being a closed* C^2-*curve with bounded curvature and* $f_i \in C^2(\mathbb{R}^2)$ *with* $\mathrm{supp}\, f_i \subseteq [0,1]^2$ *and* $\|f_i\|_{C^2} \leq 1$.

Let f_N *be a nonlinear* N-*term approximation obtained by summing over the* N *largest shearlet coefficients* $\langle f, \psi_{j,k,\boldsymbol{m},\iota} \rangle$ *in* (2.2). *Then there exists a constant* $C > 0$, *independent of* f *and* N, *such that*

$$\|f - f_N\|_2^2 \leq C N^{-2} \log^3 N, \quad as\ N \longrightarrow \infty.$$

Ignoring the additional log-factors, the previous theorem reveals that shearlets allow for a $O(N^{-2})$ approximation rate of cartoon-like functions, which is proven to be the optimal rate [18]. It is also achieved by other anisotropic systems such as curvelets [12], but out of reach for isotropic wavelet systems [61]. An intuitive explanation of this fact can be found in Figures 3(b) and 3(c): the anisotropic scaling and the shearing allow to capture geometric features more efficiently than isotropic wavelet systems.

3. The Concept of Visible and Invisible Coefficients

While the concepts of the previous section are infinite-dimensional and of rather abstract nature, a central question is how shearlets can help to access the visible part of the wavefront set in a practical manner. We will argue in this section that, at least heuristically, ℓ^1-minimization allows us to distinguish between *visible* and *invisible shearlet coefficients*.

3.1. ℓ^1-**Analysis Minimization.**

Motivated by the observation that shearlets define an efficient sparsifying transform for images with anisotropic features, they are becoming an increasingly popular choice for sparsity based regularization of inverse problems [54, 55, 59, 21]. In particular the sparsifying effect of ℓ^1-regularization has been an active field of research over the last two decades, often leading to state of the art results for the inversion from limited data. Under the label *compressed sensing* many powerful recovery guarantees for subsampled random measurements have been derived [13, 19, 24].

In this work, we propose to use the shearlet system in an *analysis based variational prior* for reconstructing reliable visible shearlet coefficients. This means that for obtaining an approximation f^* from the measurements y in (1.3), we solve the convex optimization problem

$$f^* \in \mathrm{argmin}_{f \geq 0}\, \|\mathrm{SH}_{\psi,\phi}(f)\|_{1,w} + \frac{1}{2}\|\mathcal{R}_\phi f - y\|_2^2, \tag{3.1}$$

where $\|x\|_{1,w} = \sum_j w_j |x_j|$ denotes the weighted ℓ^1-norm of $x \in \ell^1(\Lambda)$ with weights $w \in \ell^2(\Lambda, \mathbb{R}_+)$. Such a weight vector balances the influence of the shearlet regularizer and the ℓ^2-data fidelity term, subsuming the usual regularization parameter. In most tomographic problems, it is known a priorily that the desired image f is non-negative and including this constraint into (3.1) leads to superior reconstruction results.

3.2. **Visibility of Shearlets.** We now discuss the implications of the visibility principle of Section 2.1 on the obtained shearlet coefficients $\mathrm{SH}_{\psi,\phi}(f^*)$. From to the Visibility Principle 2.3 we infer that the variational approach of (3.1) should only be able to reconstruct boundaries which are visible in the limited angle data set. In terms of shearlet coefficients this means that coefficients corresponding to shearlets aligned with invisible boundaries of a solution f^* are negligible. Note that the visible boundaries are completely determined by the measured angular range $[-\phi, \phi]$, which can be conveniently expressed in the frequency domain via the Fourier slice theorem [67]. This motivates the following definition, which certainly only makes sense for bandlimited shearlet constructions, i.e., when $\hat{\psi}$ has compact support.

Definition 3.1. Let $\mathrm{SH}(\phi, \psi)$ be a bandlimited, cone-adapted discrete shearlet system and $\phi \in (0, \pi/2)$. Then, the *visible wedge* is defined by

$$W_\phi := \left\{ \xi \in \mathbb{R}^2 : \xi = r \cdot (\cos\omega, \sin\omega)^T, r \in \mathbb{R}, |\omega| \le \phi \right\}.$$

Furthermore, we define the *invisible shearlet indices* by

$$\mathcal{I}_{\mathtt{inv}} := \left\{ (j, k, \boldsymbol{m}, \iota) \in \Lambda : \mathrm{supp}\, \hat{\psi}_{j,k,\boldsymbol{m},\iota} \cap W_\phi = \emptyset \right\}, \tag{3.2}$$

and the *visible indices* by $\mathcal{I}_{\mathtt{vis}} := \Lambda \backslash \mathcal{I}_{\mathtt{inv}}$.

A visualization of the geometry described in Definition 3.1 can be found in Figure 4(c). The notion of *invisible coefficients* was originally coined by Frikel in [25] for the case of curvelet frames, an anisotropic function system similar to shearlets, but based on rotation instead of shearing. It was proven that for invisible curvelet elements $\psi_j \in L^2(\mathbb{R}^2)$ it holds true that $\psi_j \in \ker \mathcal{R}_\phi$. This property was then used for dimension reduction of the *synthesis-based* ℓ^1-regularization

$$z^* \in \mathrm{argmin}_z \|z\|_{1,w} + \frac{1}{2} \left\| \mathcal{R}_\phi \left(\sum_{j \in \Lambda} \psi_j z_j \right) - y \right\|_2^2,$$

where Λ denotes the index set of the considered curvelet frame $(\psi_j)_{j \in \Lambda}$. It was shown that the coefficients associated to invisible curvelets satisfy $z_j^* = 0$, which can be immediately used to obtain an equivalent, smaller dimensional problem. Due to the similarities of shearlets and curvelets, these arguments directly translate to shearlet frames, as already pointed out in [25].

We expect that a similar statement holds true for the analysis-based minimization of (3.1), i.e., for an invisible shearlet index $(j, k, \boldsymbol{m}, \iota) \in \mathcal{I}_{\mathtt{inv}}$ of a solution f^* of (3.1) it holds $\langle f^*, \psi_{j,k,\boldsymbol{m},\iota} \rangle \approx 0$. Although a theoretical analysis of this conjecture appears to be more complicated and therefore beyond the scope of this work, we have chosen the analysis formulation for our work for several reasons: first, it is known that analysis based methods usually yield better reconstruction quality in imaging applications. Furthermore, since the optimization is over the image domain, the resulting problem is of smaller dimension and therefore more efficient solvers exist. Also, the analysis formulation allows to naturally include the non-negativity constraints, which is fundamental in tomography applications. Finally, the analysis point of view is closer related to the characterization of wavefront sets with the continuous shearlet transform, which forms the theoretical foundation of our approach.

Before we present a numerical example that justifies the terminology of (in-)visibile shearlet elements, we first discuss a handy relaxation of Definition 3.1 in the following remark.

Remark 3.2. For shearlets with $\mathrm{supp}\, \hat{\psi}_{j,k,\boldsymbol{m},\iota} \not\subseteq W_\phi$ and $\mathrm{supp}\, \hat{\psi}_{j,k,\boldsymbol{m},\iota} \cap W_\phi \ne \emptyset$, the (in-)visibility attribution can be less clear. While $\mathcal{R}_\phi\, \psi_{j,k,\boldsymbol{m},\iota} \ne 0$ in such a case, the contribution can still be negligible if most of the support lies outside the visible wedge W_ϕ. In our numerical experiments, we therefore relax the condition of (3.2) by classifying a shearlet as visible if its orientation, determined by the shearing and anisotropic scaling, corresponds to a visible direction of $\mathcal{R}_\phi$. This principle is visualized in Figure 4(c), where the *semi-visible* shearlet would be classified as visible since most of its support lies inside the visible wedge.

In particular for a *fanbeam* scanning geometry, the definition via the visible wedge breaks down. In this case, we propose a generalization of (3.2) by setting

$$\mathcal{I}_{\mathtt{vis}} = \left\{ (j, k, \boldsymbol{m}, \iota) : \| \mathcal{R}_\phi \, \psi_{j,k,\boldsymbol{m},\iota} \|_2 > Q_j(\phi/\pi) \right\},$$

where $Q_j(p)$ denotes the p-quantile of all norms of the projected shearlets at scale j.

In the following, we will justify the terminology of *(in-)visible* shearlet indices by a simple numerical simulation. We are considering noisy Radon measurements $y = \mathcal{R}_{50^\circ} f + \eta$ of a circle f, which is displayed in Figure 5(a)[3]. Figure 5(b) shows a standard FBP reconstruction, which suffers from strong streaking artifacts and blurry edges. Such degradations are mostly avoided in the ℓ^1-regularized solution of (3.1), which is plotted in Figure 5(c). Note, that the visible edges are recovered almost perfectly, however, the horizontal, invisible boundary sections cannot be retrieved.

In order to validate the concept of (in-)visible coefficients, we form the image

$$\mathrm{SH}_{\psi,\phi}^T(\mathrm{SH}_{\psi,\phi}(f^*)_{\mathcal{I}_{\mathtt{vis}}} + \mathrm{SH}_{\psi,\phi}(f)_{\mathcal{I}_{\mathtt{inv}}}), \tag{3.3}$$

i.e., the visible coefficients of the ℓ^1-solution f^* are combined with the (in practice certainly unknown) invisible coefficients of the ground truth signal f[4]. From the strong resemblance of Figure 5(d) with the original image, we conclude that the visible coefficients of (3.1) resolve the visible boundary information almost perfectly, whereas the invisible coefficients do not convey any relevant information.

The replacement strategy of Equation (3.3) can be interpreted as having access to an *oracle* that produces invisible coefficients $\mathrm{SH}_{\psi,\phi}(f)_{\mathcal{I}_{\mathtt{inv}}}$, given the visible ones $\mathrm{SH}_{\psi,\phi}(f^*)_{\mathcal{I}_{\mathtt{vis}}}$ as an input. While such a perfect prediction is certainly to much to hope for, we will show in this work, that deep neural networks can be used to obtain an accurate estimation of the invisible coefficients. Indeed, Figure 5(f) shows the result obtained by combining the visible ℓ^1-coefficients $\mathrm{SH}_{\psi,\phi}(f^*)_{\mathcal{I}_{\mathtt{vis}}}$ with invisible coefficients that have been inferred by a CNN trained on a collection of ellipses; see Section 5 for further details.

We would like to point out that the previous visibility interpretation of the coefficients is not valid for the classical filtered backprojection solution f_{FBP}. The result of the analogous oracle replacement

$$\mathrm{SH}_{\psi,\phi}^T(\mathrm{SH}_{\psi,\phi}(f_{\mathrm{FBP}})_{\mathcal{I}_{\mathtt{vis}}} + \mathrm{SH}_{\psi,\phi}(f)_{\mathcal{I}_{\mathtt{inv}}}) \tag{3.4}$$

is shown in Figure 5(e). It reveals that the visible coefficients of f_{FBP} are tainted with artifacts and blurry edges, making such a separation into visible and invisible coefficients impossible.

Concluding this discussion, we state the following visibility principle:

Visibility Principle 3.3

We can split the shearlet coefficients $\mathrm{SH}_{\psi,\phi}(f^*)$ of (3.1) into a set of *visible* and *invisible* coefficients:
  (1) The visible coefficients $\mathrm{SH}_{\psi,\phi}(f^*)_{\mathcal{I}_{\mathtt{vis}}}$ carry reliable information about edges that are possible to reconstruct according to the visibility principle of Section 2.1.
  (2) The invisible coefficients $\mathrm{SH}_{\psi,\phi}(f^*)_{\mathcal{I}_{\mathtt{inv}}}$ are penalized by the ℓ^1-norm and do not contain relevant information.

Finally, we emphasize the close relationship between the shearlet coefficients $\mathrm{SH}_\psi(f^*)$ obtained by solving (3.1) and the phase space representation of microlocal analysis: recall that the wavefront set information can be extracted by analyzing the decay properties of the continuous shearlet transform. In terms of the discrete shearlet system a rapid decay of the continuous transform manifests as *sparsity* of the associated shearlet coefficients. It is therefore quite natural to access this information via the sparsity promoting effect of ℓ^1-minimization. As desired, the coefficients of shearlets which are not aligned with smooth directions of f are *"pushed to 0"* by the ℓ^1-norm. When sorted properly, the coefficients belonging to one particular scale are reminiscent of a discretized version of the wavefront set of f. In particular, they obey similar structural properties as wavefront sets in the phase space. A visualization of this observation can be found in Figure 6: a stylized plot of the finest scale coefficients of Figure 5's circle is shown in

---

[3]All objects of this computational example are certainly finite-dimensional, e.g., $f \in \mathbb{R}^{512 \times 512}$. For the sake of clarity, we chose to stick to the continuous notation until introducing our digitalized reconstruction framework in the next section.

[4]For $x \in \ell^2(\Lambda)$ and a set $\mathcal{I} \subseteq \Lambda$, $x_{\mathcal{I}} \in \ell^2(\Lambda)$ shall denote the vector with $x_{\mathcal{I}}(i) = x(i)$ for $i \in \mathcal{I}$ and $x_{\mathcal{I}}(i) = 0$ otherwise.
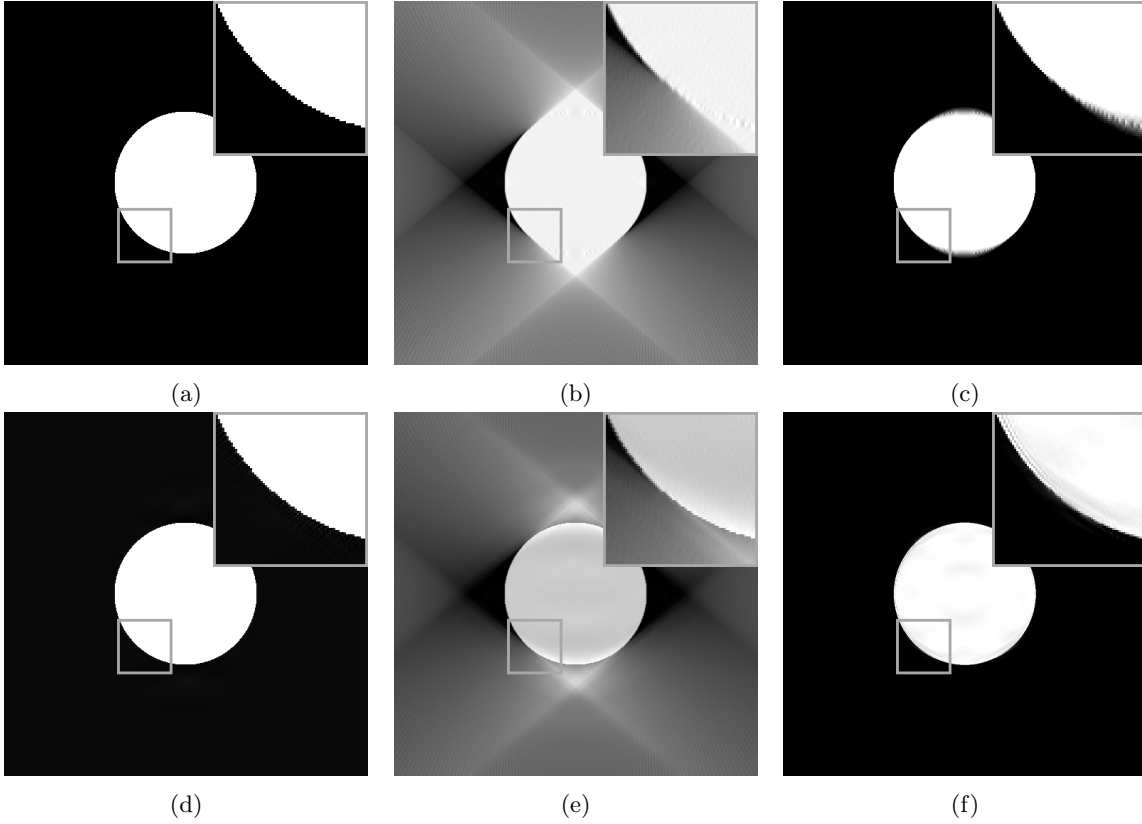
Figure 5: A simulation visualizing the Visibility Principle 3.3 using noisy $\mathcal{R}_{50°}$ measurements, i.e., a missing wedge of $80°$. (a) shows a simple choice for f and (b) its standard FBP reconstruction. (c) displays the ℓ^1-regularized shearlet solution of (3.1). In (d) and (e) we plot the results of the oracle replacement with perfect invisible coefficients as described in (3.3) and (3.4), respectively. Subplot (f) shows a reconstruction, where the invisible shearlet coefficients are inferred by a neural network. Note that the dynamic range of the plots is modified for better contrast.

Figure 6(a); cf. the phase space visualization of Figure 2(b). As to be expected, the coefficients $\mathrm{SH}_\psi(f^*)$ of Figure 6(b) follow the same pattern on the visible part, however, there are holes corresponding to the invisible coefficients. Finally, Figure 6(c) shows the coefficients of the deep learning based solution of Figure 5(f). Indeed, the neural network seems to pick up the phase space structure and accurately estimates the invisible information.

4. Proposed Reconstruction Method: Learning the Invisible (LtI)

In this section, we define our hybrid recovery framework that makes use of an artificial neural network to learn the invisible information that cannot be retrieved from the measured data. We first introduce and discuss the general reconstruction workflow and then give more details on supervised learning of shearlet coefficients and on our particular CNN architecture.

4.1. **Algorithm.** After suitable discretization, we are given the finite-dimensional measurement vector

$$\boldsymbol{y} = \boldsymbol{\mathcal{R}}_\phi \boldsymbol{f} + \boldsymbol{\eta} \in \mathbb{R}^m,$$

where $\boldsymbol{f} \in \mathbb{R}^{n^2}$ denotes the (unknown) discrete and vectorized image, $\boldsymbol{\mathcal{R}}_\phi \in \mathbb{R}^{m \times n^2}$ describes a discretized version of $\mathcal{R}_\phi$ and $\boldsymbol{\eta} \in \mathbb{R}^m$ models the measurement noise. We propose the following recovery scheme for finding a reconstruction $\boldsymbol{f}_{\mathtt{LtI}} \in \mathbb{R}^{n^2}$ of $\boldsymbol{f}$:

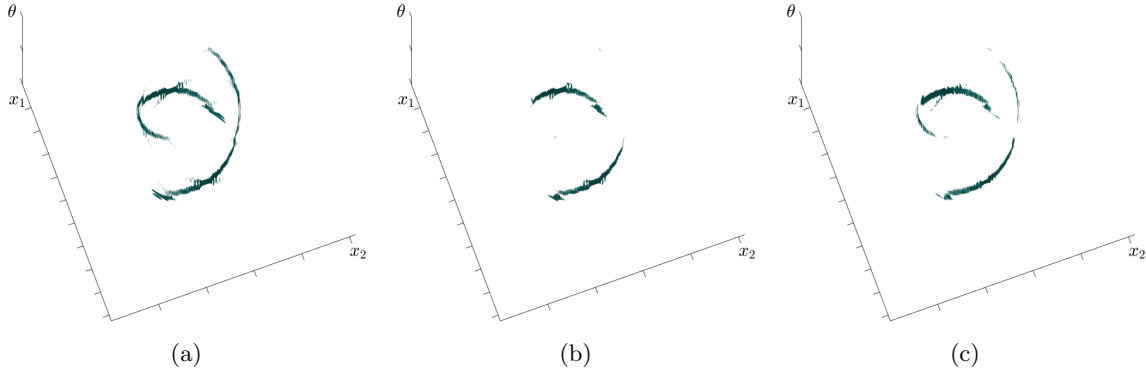(a)   (b)   (c)

Figure 6: (a) shows a stylized visualization of the shearlet coefficients of the circle in Figure 5(a). (b) displays the coefficients of the ℓ^1-analysis solution f^*, revealing holes on the invisible part. (c) shows the coefficients of the reconstruction in Figure 5(f), i.e., after the invisible coefficients have been inferred by a neural network.

Algorithm 4.1: Learning the Invisible (LtI)

Step 1: Obtain the visible coefficients $\mathbf{SH}(f^*)_{\mathcal{I}_{\mathrm{vis}}}$ via a nonlinear reconstruction

$$f^* \in \operatorname{argmin}_{f \geq 0} \|\mathbf{SH}(f)\|_{1,\boldsymbol{w}} + \frac{1}{2}\|\mathcal{R}_\phi\, f - y\|_2^2, \tag{4.1}$$

where $\mathbf{SH} \in \mathbb{R}^{J \cdot n^2 \times n^2}$ denotes a digitalized version of $\mathrm{SH}_{\psi,\phi}$ with J decomposition subbands.

Step 2: Apply a CNN, denoted by $\mathcal{NN}_{\boldsymbol{\theta}}$, that is trained to estimate the invisible coefficients from the visible ones, i.e., determine the coefficients

$$\boldsymbol{F} = \mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(f^*)) \quad (\approx \mathbf{SH}(f)_{\mathcal{I}_{\mathrm{inv}}}).$$

Step 3: Combine the visible and the learned invisible coefficients in a reliable manner by setting

$$\boldsymbol{f}_{\mathtt{LtI}} = \mathbf{SH}^T(\mathbf{SH}(f^*)_{\mathcal{I}_{\mathrm{vis}}} + \boldsymbol{F}).$$

A schematic workflow of the proposed method can be found in Figure 7. For solving the optimization problem (4.1) there is an abundance of possibilities. In this work, we are using the alternating direction method of multipliers (ADMM), as detailed in Appendix A. After a discussion of our approach in the next section, we will describe the particular architecture of $\mathcal{NN}_{\boldsymbol{\theta}}$ and the procedure of learning the parameter vector $\boldsymbol{\theta}$ in the remainder of this chapter.

4.2. Motivation and Discussion. In the following, we motivate the proposed hybrid reconstruction scheme by relating it to Visibility Principle 3.3, and discuss some of its properties.

4.2.1. *Motivation.* The missing wedge of $\mathcal{R}_\phi$ results in a lack of directional information in the measured data, which is eventually responsible for artifacts and missing image features in model-based reconstruction methods. By solving the ℓ^1-analysis minimization (4.1), we gain access to shearlet coefficients that correspond to reliable image features. The remaining invisible coefficients cannot be retrieved from the measured data. However, the obtained coefficient tensor $\mathbf{SH}(f^*)$ somewhat resembles a discretized version of the phase space that is interspersed with holes on the invisible parts. For natural images, the visible coefficients are highly structured allowing for an inference of the invisible sections (cf. discussion of Section 3.2 and Figure 2). Thus, it appears to be a natural choice to apply machine learning techniques for the estimation in *Step 2* of our proposed method.

Recently, CNNs have shown to be very effective for computer vision tasks such as image classification [52] or segmentation [76, 43], but also in the context of inverse problems, e.g., [10, 86, 44, 3, 4, 36]. Based on the intuition that low-dose CT artifacts possess a directional nature, [47, 31] proposed a post-processing of the FBP's directional wavelet coefficients. The subbands of the wavelet decomposition are

thereby treated as different channels of the FBP image and a CNN is trained to remove their artifacts. Most, if not all, of such post-processing methods are based on so-called U-Net architectures [76]. For the estimation of the invisible coefficients in *Step 2*, we use a modification of a similar architecture that we call *PhantomNet*, see Section 4.3.3.

Our hybrid approach that combines model-based visible coefficients and learned invisible coefficients as detailed in *Step 3* is accompanied by particular features that we will briefly discuss the next three paragraphs.

4.2.2. *Interpretability.* By incorporating the neural network into a model-based approach, our proposed scheme offers a clear interpretation of its post-processing abilities in the context of limited angle CT. In *Step 2*, the network $\mathcal{N}\mathcal{N}_{\boldsymbol{\theta}}$ estimates the invisible coefficients from the knowledge of the previously reconstructed visible coefficients. The underlying principle of such an estimation is that the shearlet coefficients of natural images obey specific structural rules, similar to a wavefront set in the phase space. During the training over a particular class of images (cf. Section 4.3.1), the parameter vector $\boldsymbol{\theta}$ captures these general structural properties in the shearlet domain (cf. Section 4.3.1 and Figure 6). When applying to fresh testing data the neural network estimates the invisible coefficients according to these rules. We wish to emphasize that *Step 2* can also be interpreted as a *3D-inpainting problem:* the invisible parts of the reshaped shearlet coefficient tensor $\mathbf{SH}(\boldsymbol{f}^*) \in \mathbb{R}^{n \times n \times J}$ are sought to be inpainted by the neural network $\mathcal{N}\mathcal{N}_{\boldsymbol{\theta}}$.

Overall, the split into visible and invisible shearlet coefficients and the dedication of the CNN to solely infer the invisible ones clarifies the post-processing capabilities of neural networks. In contrast, in [47, 31, 44], deep learning is merely used as a black box tool for a somewhat unspecified removal of artifacts in the FBP or its coefficients.

4.2.3. *Reliability.* While our hybrid approach entrust the CNN with the transparent task of estimating invisible coefficients, it remains unclear to what extent this is actually possible. As we have argued previously, despite the success of deep learning, there is up to date no profound understanding under which assumptions on the training data, the neural network architecture, and other design choices, an accurate inference is feasible.

Our proposed hybrid reconstruction method alleviates these issues from another perspective: by keeping the visible coefficients of the nonlinear reconstruction for the final image formation in *Step 3*, we limit the influence of the neural network on the final reconstruction to a minimum. The information that we can reconstruct via the well-understood and model-based ℓ^1-minimization of (4.1) directly contributes to the formation of $\boldsymbol{f}_{\texttt{LtI}}$. Only the part that is provably not contained in the measured data is estimated by a CNN.

In particular for medical applications, it might be unsatisfactory to process an image by a CNN in order to remove its artifacts, while having no control over the applied modifications. In our approach, the impact of the *"black-box CNN"* on the final reconstruction is constrained to the smallest possible extent. We believe that such an entanglement of model and data-based methods compromises between performance and reliability: the reconstructions greatly benefit from the abilities of neural networks for the estimation of the invisible part, whereas the visible boundaries are kept as reliable as possible.

4.2.4. *Performance.* While the two previous characteristics are mostly of conceptual nature, our numerical experiments reveal superior reconstruction quality when compared to other methods. In particular, we observe a remarkable generalization when our method is applied to different testing data. This is largely due to the fact that the visible coefficients are reconstructed by an advanced model-based method and therefore make for a better initialization of the input for the neural network. Furthermore, generalization is only of relevance on the invisible part of the wavefront set, since the final image formation of *Step 3* is based on the model-based reconstruction of the visible part.

When a CNN is used for post-processing an FBP reconstruction [44, 47, 31], a lot of its *expressiveness* is needed for removing the streaking artifacts. This is particularly important for limited angle CT, where depending on the size of the missing wedge, the streaking artifacts are severe. In contrast, an initialization with the nonlinear reconstruction of (4.1) is contaminated far less with unwanted artifacts and the network is allowed to focus on learning the invisible edge information. Additionally, it is well known that ℓ^1-minimization leads to sharper boundaries when compared to FBP images. This effect
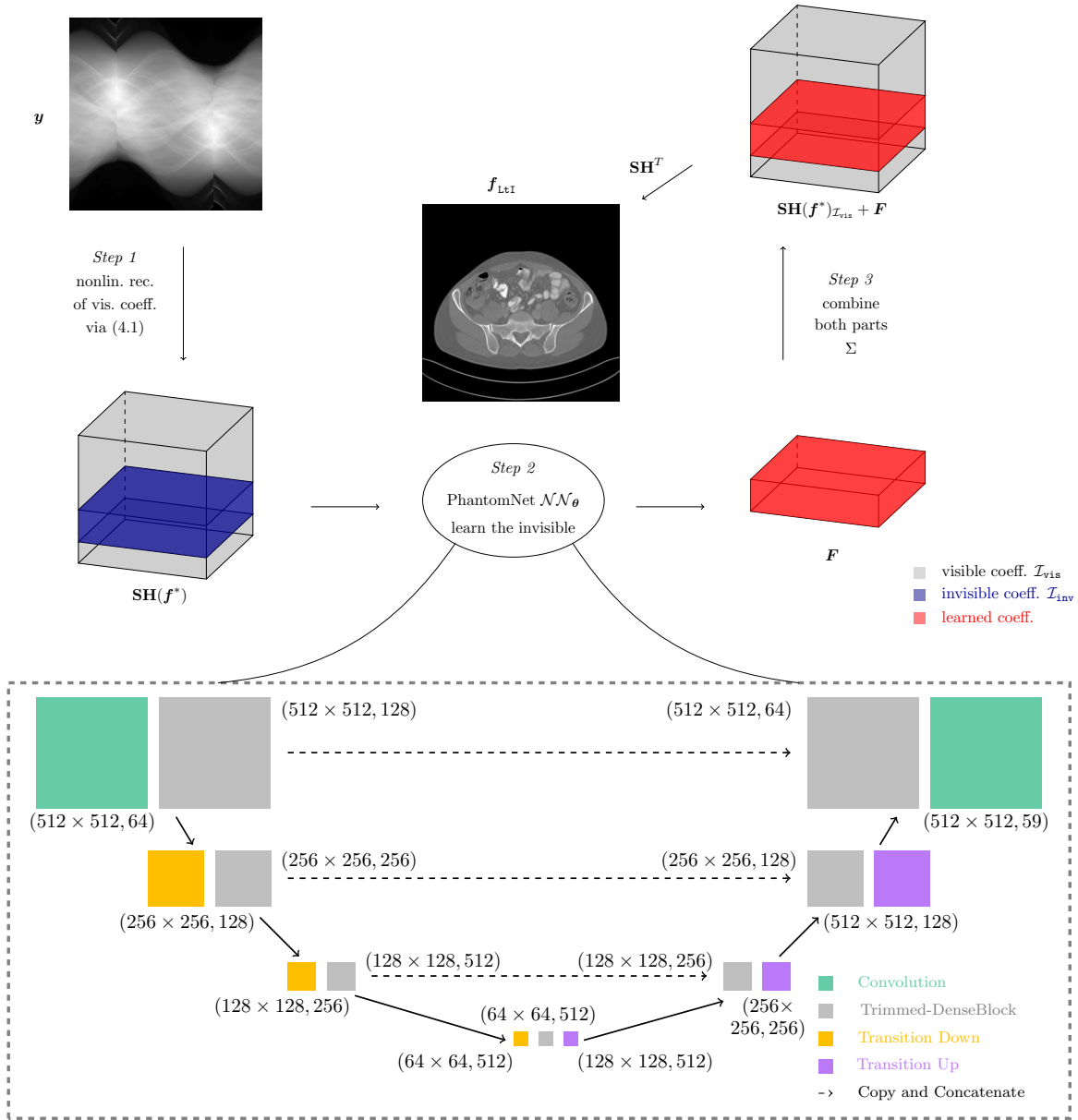
Figure 7: A schematic workflow of the proposed reconstruction framework *LtI* (see Algorithm 4.1), which learns the invisible shearlet coefficients for limited angle tomography. The lower part depicts the architecture PhantomNet. The output shape of each layer is denoted by (size $\times$ size, channels). The input to PhantomNet is of shape $(512 \times 512, 59)$. We choose $n = 4$ layers in each TDB, except for the center TDB where $n = 8$. Thus, the overall number of layers is $6 \times 4 + 8 + 8 = 40$.

might be amplified by processing with a U-Net architecture [35]. However, it is completely avoided on the visible part by the combination proposed in *Step 3*.

4.3. **Learning the Invisible with PhantomNet.** Given the visible shearlet coefficients of the non-linear reconstruction, we are now briefly describing a machine learning framework for the estimation of the invisible coefficients by means of a CNN. The overall goal is to find a (non-linear) mapping $\mathcal{NN}_{\boldsymbol{\theta}} : \mathbb{R}^{n \times n \times J} \to \mathbb{R}^{n \times n \times J}$, parametrized by a (high-dimensional) vector $\boldsymbol{\theta}$, that ideally satisfies the relation

$$\mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}^*)) \approx \mathbf{SH}(\boldsymbol{f})_{\mathcal{I}_{\texttt{inv}}}.$$

We first give a short introduction to the general statistical learning framework and then describe the particular CNN architecture PhantomNet that we use for our experiments in Section 5.

4.3.1. *(Supervised) Learning of Invisible Coefficients.* For a mathematical formalization of learning invisible coefficients, we regard the tuple $(\boldsymbol{f}, \boldsymbol{f}^*) \in \mathbb{R}^{n^2} \times \mathbb{R}^{n^2}$ as a random variable with a joint probability distribution p. Ideally, we would like to find a parameter vector $\boldsymbol{\theta}$ that allows for an estimation of the invisible coefficients with respect to p. This could for instance be achieved by minimizing the expected risk

$$\min_{\boldsymbol{\theta}} \left(\mathbb{E}_{(\boldsymbol{f}, \boldsymbol{f}^*) \sim p} \left\| \mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}^*)) - \mathbf{SH}(\boldsymbol{f})_{\mathcal{I}_{\text{inv}}} \right\|_{\boldsymbol{w}, 2}^2 \right), \tag{4.2}$$

where $\boldsymbol{w} \in \mathbb{R}^{\#\mathcal{I}_{\text{inv}}}$ is a vector of weights, accounting for instance for the fact that shearlet coefficients come in different orders of magnitude depending on their scale. Note that the ℓ^2-loss is only computed on the invisible coefficients that are sought to be learned. In principle, any other loss function or an additional regularization term, such as the sparsity promoting ℓ^1-norm of $\mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}^*))$, could be beneficial. For the sake of brevity, we will stick to the basic form of (4.2).

In practice, computing the expectation with respect to p is not possible. Instead, we are typically given a finite set of independent drawings $(\boldsymbol{f}_1, \boldsymbol{f}_1^*), \ldots, (\boldsymbol{f}_N, \boldsymbol{f}_N^*)$ and consider the minimization of the empirical risk

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{j=1}^{N} \left\| \mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}_j^*)) - \mathbf{SH}(\boldsymbol{f}_j)_{\mathcal{I}_{\text{inv}}} \right\|_{\boldsymbol{w}, 2}^2. \tag{4.3}$$

Depending on the properties of $\mathcal{NN}_{\boldsymbol{\theta}}$, the optimization problem is in general non-convex. In the case of neural networks, typically some form of gradient descent is used, where the gradients are calculated via *backpropagation* [77]. Computing the gradient for the sum over the entire training set in (4.3) is often not feasible for large-scale problems due to memory limitations. To circumvent this problem, *stochastic* or *minibatch gradient descent* is used, in which the gradient is approximated over smaller, randomly selected batches of training examples [28, Chpt. 8].

The final performance (i.e., the *generalization*) of the trained neural network $\mathcal{NN}_{\boldsymbol{\theta}}$ is evaluated on a separate set of independent drawings, the so-called *test set*, that were not previously used for the optimization of $\boldsymbol{\theta}$ in (4.3).

4.3.2. *Convolutional Neural Networks.* The main building blocks of CNNs are convolutional layers of the following form: let $\boldsymbol{I} = [\boldsymbol{I}_1, \ldots, \boldsymbol{I}_{c_1}] \in \mathbb{R}^{k \times k \times c_1}$ be an input array, where $k \in \mathbb{N}$ is referred to as the *spatial dimension* and $c_1 \in \mathbb{N}$ as the number of input *channels*. For a desired number of output channels $c_2 \in \mathbb{N}$ and $i \in \{1, \ldots, c_2\}$ let $\boldsymbol{w}^i = [\boldsymbol{w}_1^i, \ldots, \boldsymbol{w}_{c_1}^i] \in \mathbb{R}^{s \times s \times c_1}$ denote a convolutional filter with kernel size $s \in \mathbb{N}$ and $b_i \in \mathbb{R}$ a bias. Then, the i-th output channel of the convolutional layer is given by

$$\boldsymbol{o}_i = \sigma \left(\sum_{j=1}^{c_1} \boldsymbol{I}_j * \boldsymbol{w}_j^i + b_i \right), \tag{4.4}$$

where $*$ denotes a 2D-convolution and $\sigma : \mathbb{R} \to \mathbb{R}$ is a non-linear (*activation*) function. With an abuse of notation, the application of σ and the addition is thereby applied elementwise.

State of the art CNNs typically consist of dozens or hundreds of concatenated convolutional layers which are regularly alternated with *pooling* layers [28, Chpt. 9.3]. Their main features are a relatively low number of parameters (when compared with fully connected NNs) and the translation invariance due to the convolutional structure. The set of all free parameters, i.e., the convolutional weights and biases of all layers, are collected in the parameter vector $\boldsymbol{\theta}$. There is a plethora of variations and extensions of this basic definition and we refer the interested reader to [28] for more details on this subject. A precise description of the *PhantomNet*-architecture used for our experiments will be given in the following section.

4.3.3. *PhantomNet.* Our architecture, which we refer to as *PhantomNet*, is largely based on U-Net - a CNN that was introduced in [76] for biomedical image segmentation. In general, U-Nets consist of an encoder and a decoder similar to autoencoders [28, Chpt. 14]. The encoder takes the input and maps it to a latent compressed representation while the decoder up-samples towards the output. The U-Net architecture enables passing the high resolution feature maps directly from the encoder to the decoder. For the concatenation of the feature maps in the decoder, the architecture is kept symmetric with respect to the encoder, such that the overall network architecture resembles a 'U'; cf. Figure 7.

The encoder and the decoder of PhantomNet consist of the following **building blocks:**

- Trimmed-DenseBlocks (TDB): the defining feature of PhantomNet are its modificated *Dense-Blocks*. DenseBlocks were introduced in [42] as densely connected groups of layers and can be seen as an extension of the popular *residual networks (ResNets)* [37]. A residual block bypasses the non-linear transformation by an identity function, i.e.,

$$\boldsymbol{x}_{n+1} = H_{n+1}(\boldsymbol{x}_n) + \boldsymbol{x}_n, \tag{4.5}$$

where H_{n+1} represents the non-linear transformation (4.4) at the $n+1$-th layer and $\boldsymbol{x}_n$ denotes the output of the n-th layer. This shortcut connection enforces the layer to learn something new compared to the input. Additionally, the residual connections help for a faster convergence during the optimization. [82] shows that the ResNets address the problem of vanishing gradients in very deep networks by using short paths.

  DenseBlocks (DB) generalize this idea in the sense that their layers have connections from all previous layers - i.e., the input to the current layer consists of the feature maps of all the layers in the DenseBlock before it:

$$\boldsymbol{x}_{n+1} = H_{n+1}([\boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_n]), \tag{4.6}$$

where $[\cdot]$ denotes the concatenation of the arrays $\boldsymbol{x}_i$. Each layer in the block consists of a 3×3 convolution with a stride of 1 followed by a *hyperbolic tangent (tanh)* non-linear activation φ. The outputs of the layer are c_2 feature maps, which is defined as the *growth rate*. The output of the DB is a concatenation of the output of all the layers in it (of size $n \times c_2$), concatenated with the input $\boldsymbol{x}_0$. TDBs are DBs, in which the input feature maps to a block, i.e. $\boldsymbol{x}_0$, are not passed to the output of the block. In each TDB, we choose $n \in \{4, 8\}$ layers and a growth rate $c_2 \in \{16, 32, 64, 128\}$, such that there are fewer feature maps in the initial blocks and more of them towards the latent representation; cf. Figure 7.
- TransitionDown (TD): the encoder of the PhantomNet compresses the input to a latent representation. To enable the contracting path, the TransitionDown consists of a 3×3 convolution with a stride of 1 followed by *max pooling* operation with a stride of 2. This brings down the resolution of the feature maps by a factor of 2.
- TransitionUp (TU): the decoder upscales the latent representation using the transpose of a 3×3 convolution with a stride of 2.

PhantomNet consists of 4 TDBs on the encoder and 3 TDBs on the decoder, as shown in Figure 7. Feature maps of the first 3 TDBs on the encoder are passed through the skip connections and concatenated with the respective feature maps before given as input to the respective TDB at the decoder. Following the first 3 TDBs are TDs which bring down the resolution of the feature maps by a factor of 2. On the decoder, there are TU blocks after the TDB blocks, which make use of the convolution transpose operation to upscale the resolution of the feature maps by a factor of 2.

The PhantomNet architecture is inspired from [43], albeit with differences. While [43] uses DenseBlocks with batch normalization and rectified linear units as non linear activations, PhantomNet makes use of Trimmed-DenseBlocks with just *tanh* activation. The main difference between TDBs and DBs is that in TDBs the input feature maps to a block are not passed to the output of the block. Since the output feature maps from a TDB are passed through a skip connection to the decoder, there is a considerable reduction in the size of the network which is beneficial for computational expense. The residual connections in the block already force each layer to learn useful representations. The trimming of the input feature maps to the output does no harm and is found to work better for problems where sparse information has to be learned.

PhantomNet takes the entire stack of shearlet coefficients $\mathbf{SH}(\boldsymbol{f}^*)$ as an input tensor. Similar to [47, 31], the subbands of $\mathbf{SH}(\boldsymbol{f}^*)$ (corresponding to directional features at different scales) are thereby treated as different input channels. Note that PhantomNet is a *fully convolutional neural network* [58]. This means that, in contrast to classification neural networks, the last layer is not a fully connected layer but also convolutional. Therefore, PhantomNet is able to process inputs of arbitrary spatial dimension, which will be used during the training stage.

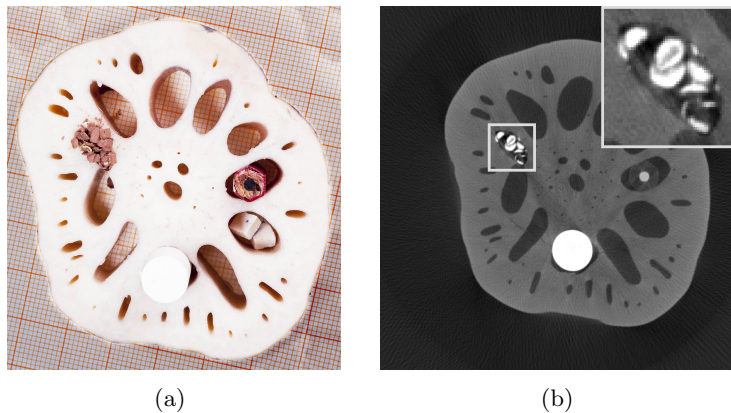(a)   (b)

Figure 8: The image (a) shows a photography of the signal used in the experiments Lotus-60° and Lotus-75°. In (b), a reference scan with full angular measurements is displayed, in which the plotting window is slightly adapted for better contrast.

5. Experiments and Results

In this section, we evaluate the performance of the proposed reconstruction scheme by comparing with classical and learning based reconstruction methods. For thorough testing we consider a combination of different measurement setups and different types of simulated and measured data.

5.1. Preliminaries.

Let us begin by describing the considered experimental scenarios and giving details on the implementation of the used operators, the training procedure and the methods that we compare our results with.

5.1.1. *Experimental Scenarios.* We consider three different types of data for evaluating our proposed method:

(1) The *ellipsoid dataset* consists of 2000 synthetic images of ellipses, where the number, locations, sizes and the intensity gradients of the ellipses are chosen at random. Using the Matlab function `radon`, we simulate noisy measurements for a missing wedge of 80°. To avoid an *inverse crime* [66] the measurements are simulated at a higher resolution and then downsampled for an image resolution of 512×512. 1600 images are used for training, 200 images for validation and 200 for testing. This experimental setup is referred to as Experiment Ellipses-50°.

(2) In a more realistic setup, we are working on human abdomen scans provided by the *Mayo Clinic* for the AAPM Low-Dose CT Grand Challenge [63]. The data consists of 10 patients resulting in 2378 images of size 512×512 with a slice thickness of 3mm. We use 9 patients for training (2134 slices) and 1 patient for testing (244 slices). We remark that this is not completely consistent with the setup used in [31], where 1 training patient was used for validation (330 slices). Noisy measurements are simulated with Astra [81] using a *fanbeam* geometry corresponding to missing wedges of 60° and 30°. We will refer to these scenarios as Mayo-60° and Mayo-75°, respectively.

(3) For testing the generalization properties of our method, we furthermore make use of real data from a scan of a *lotus root* [8]. This setup is referred to as Lotus-60° and Lotus-75°, respectively. A reference image without missing wedge can be found in Figure 8. Note, that the fanbeam geometry of the Mayo data is chosen such that it matches the specifications of the lotus scan.

5.1.2. *Operators.* For an implementation of the discrete limited angle operator $\mathcal{R}_\phi$ we use the standard `radon` routine of Matlab or a fanbeam geometry of the Astra toolbox [81], which fits to the geometry described in [8]. For all our experiments we are using a discrete, bandlimited shearlet system generated with the toolbox [83]. The resulting system has $J = 5$ scales, resulting in a transformation $\mathbf{SH} \in \mathbb{R}^{59 \cdot 512^2 \times 512^2}$, i.e., the shearlet coefficient cube $\mathbf{SH}(\boldsymbol{f})$ has 59 subbands of size 512×512.

5.1.3. *Network Training.* Training PhantomNet is performed using Tensorflow [2] with an Adam optimizer [48] and a learning rate (step size) of 10^{-4}. In order to converge to a *good* local minimizer of (4.3), i.e., one with a small generalization error, neural networks typically require a large number of training samples. Since the ellipsoid and Mayo data sets only consist out of ~ 2000 images, we make use of additional data augmentation techniques to help the network's convergence: for each tensor of size $512 \times 512 \times 59$, a random $320 \times 320 \times 59$-patch is sampled on-the-fly and given as input to the fully convolutional PhantomNet. We find that such an on-the-fly sampling defines an effective regularization method. For the evaluation on the test set, the full array is given as input to the network.

Since the shearlet coefficients naturally come in different orders of magnitude on each scale, we observe that weighting higher scales with larger weights in (4.3) (corresponding to small images features) significantly improves the neural network's performance, cf. the weighting strategy for the ℓ^1-minimization in Appendix A.

5.1.4. *Compared Methods.* We compare our reconstruction results with a variety of classical and learning based methods, which we describe briefly in the following list:

$\boldsymbol{f}_{\text{FBP}}$: Standard filtered backprojection with a 'shepp-logan' or 'ram-lak' filter, as provided by Matlab and Astra, respectively.

$\boldsymbol{f}^*$: The ℓ^1-regularized shearlet solution of (4.1).

$\boldsymbol{f}_{\text{TV}}$: Total variation regularized solution with non-negativity constraint, i.e., a solution of (4.1) where $\mathbf{SH}$ is replaced by a discrete gradient operator.

$\mathcal{NN}_{\boldsymbol{\theta}}(\boldsymbol{f}_{\text{FBP}})$: Post-processing of $\boldsymbol{f}_{\text{FBP}}$ with PhantomNet, i.e., $\mathcal{NN}_{\boldsymbol{\theta}}$ is trained to remove artifacts in $\boldsymbol{f}_{\text{FBP}}$. Besides the different CNN architecture, this method resembles the one proposed in [44].

$\mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}_{\text{FBP}}))$: Post-processing of FBP's shearlet coefficients with PhantomNet, i.e., $\mathcal{NN}_{\boldsymbol{\theta}}$ is trained to remove artifacts in the coefficient domain. This method is closely related to the one proposed in [31].

$\boldsymbol{f}_{[31]}$: The actual method of [31].

We wish to emphasize that all deep learning based comparison methods do not distinguish between visible and invisible information. Furthermore, residual learning is deployed, meaning that the networks learn the difference between input and output.

5.1.5. *Similarity Measures.* For an assessment of image quality, we are using several quantitative measures, such as the relative error (RE) given by

$$\|\boldsymbol{f}_{\text{LtI}} - \boldsymbol{f}\|_2 \, / \, \|\boldsymbol{f}\|_2 \, ,$$

where $\boldsymbol{f}$ denotes the reference image and $\boldsymbol{f}_{\text{LtI}}$ its reconstruction. Furthermore, we consider the peak signal-to-noise ratio (PSNR) and the structured similarity index (SSIM) [85] provided by Matlab. Finally, we are reporting the Haar wavelet-based perceptual similarity index (HaarPSI) that was recently proposed in [74].

5.2. **Results.** In the following, we will report and discuss the results of our numerical experiments.

5.2.1. *Ellipses-*$50°$. The average image quality measures of the 200 test images are reported in Table 1. Note that for the Ellipses-$50°$ setup we did not compare with [31], since their networks have only been trained for a fanbeam geometry and for smaller missing wedges. A visualization of the reconstruction quality for one of the test images is given in Figure 9. Due to the large missing wedge of $80°$, the FBP image in Figure 9(b) is heavily contaminated with streaking artifacts and contrast changes. Using ℓ^1-minimization in Figure 9(c), it is possible to reduce such artifacts significantly, however, the invisible boundaries are certainly not recoverable. The second row of Figure 9 reveals the effect of post-processing with CNNs: in all three methods, the invisible boundaries are well estimated by PhantomNet. Although the methods in (d) and (e) do a remarkable job, the zoomed parts of Fig 9(f) and the values in Table 1 reveal an advantage of our proposed framework. For better distinction of the learning based methods, the plots (g)-(i) show the absolute value of the difference images with respect to the ground truth. Subplot (g) and (h) reveal that $\mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}_{\text{FBP}}))$ and in particular $\mathcal{NN}_{\boldsymbol{\theta}}(\boldsymbol{f}_{\text{FBP}})$ mainly struggle with removing background fluctuations of $\boldsymbol{f}_{\text{FBP}}$. The error plot in Fig 9(i) shows that due to the ℓ^1-minimization this issue is less prominent for our proposed solution, where mostly high frequency information of the invisible edges is missing. For the sake of brevity we omitted plotting $\boldsymbol{f}^*$, which is very similar to $\boldsymbol{f}_{\text{TV}}$. However,

Method	RE	PSNR	SSIM	HaarPSI
$\boldsymbol{f}_{\text{FBP}}$	0.84	17.16	0.12	0.18
$\boldsymbol{f}^*$	0.22	28.76	0.94	0.47
$\boldsymbol{f}_{\text{TV}}$	0.21	29.54	0.95	0.54
$\mathcal{NN}_{\boldsymbol{\theta}}(\boldsymbol{f}_{\text{FBP}})$	0.19	30.20	0.54	0.75
$\mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}_{\text{FBP}}))$	0.18	30.52	0.78	0.72
$\boldsymbol{f}_{\text{LtI}}$	**0.09**	**36.96**	**0.96**	**0.86**

Table 1: Comparison of reconstruction methods for Ellipses-50°. The similarity values are averaged over the images in the test set. An example is displayed in Figure 9.

Method	Mayo-60°				Mayo-75°			
	RE	PSNR	SSIM	HaarPSI	RE	PSNR	SSIM	HaarPSI
$\boldsymbol{f}_{\text{FBP}}$	0.47	17.16	0.40	0.32	0.31	21.23	0.48	0.46
$\boldsymbol{f}_{\text{TV}}$	0.18	25.88	0.85	0.37	0.10	30.99	0.85	0.64
$\boldsymbol{f}^*$	0.17	26.34	0.85	0.40	0.09	31.58	0.92	0.65
$\boldsymbol{f}_{[31]}$	0.25	23.06	0.61	0.34	0.22	23.92	0.69	0.44
$\mathcal{NN}_{\boldsymbol{\theta}}(\boldsymbol{f}_{\text{FBP}})$	0.15	27.40	0.78	0.52	0.10	31.22	0.81	0.81
$\mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}_{\text{FBP}}))$	0.16	26.80	0.74	0.52	0.06	35.190	0.90	0.82
$\boldsymbol{f}_{\text{LtI}}$	**0.08**	**32.77**	**0.93**	**0.73**	**0.04**	**39.77**	**0.96**	**0.90**

Table 2: Comparison of reconstruction methods for Mayo-60°and Mayo-75°. The values are averaged over all slices of the test patient. Examples are shown in Figure 10 and Figure 11, respectively.

since the class Ellipses-50° consists out of simple, almost piecewise constant images, $\boldsymbol{f}_{\text{TV}}$ is a very effective reconstruction method, even when compared CNN-based methods. A similar observation was made in [44] in the context of low-dose CT.

5.2.2. *Mayo-60° and Mayo-75°.* Table 2 shows the average quality measures on the test patient of Mayo-60° and Mayo-75°, respectively. A visualization of two different slices of the test patient can be found in Figure 10 and Figure 11.

In the case of a missing wedge of 60°, the overall body shape shows clear differences for all considered methods. When compared to $\boldsymbol{f}_{\text{FBP}}$, the regularization methods in Figure 10(c) and 10(d) succeed in removing artifacts, however, the invisible boundaries are certainly not reconstructable. The learning-based methods displayed in the images (e)-(g) of Figure 10 estimate the invisible boundaries quite well, yet, there are unwanted fluctuations visible. Our proposed scheme reconstructs the invisible boundary almost perfectly. In particular, it is the only method that finds the correct round shape in the upper zoomed section.

For a smaller missing wedge of 30°, Figure 11 and Table 2 (right-hand side) show that the differences between the compared methods are becoming less prominent. The learning-based methods of the images (f)-(h) of Figure 11 are visually almost indistinguishable on mid- and large-scale features. However, for small details the zoomed part reveals a clear advantage of our proposed method. While surprisingly $\mathcal{NN}_{\boldsymbol{\theta}}(\boldsymbol{f}_{\text{FBP}})$ has a small edge over $\mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}_{\text{FBP}}))$ on Mayo-60°, here, it is indeed the other way round as reported in [31]. Finally, we remark that in this example, the staircasing effect of TV-regularization is visible on the zoomed part of Figure 11(c) and avoided by relaying on shearlets as in Figure 11(d).

5.2.3. *Lotus-60° and Lotus-75°.* In the experiments of this section, we evaluate the generalization performance by applying the neural networks trained on the Mayo data to the scan of the lotus root. Let us explicitly point out that the neural networks have only been trained on the Mayo data and have *not* been retrained on any other data set. Although there is a vague similarity between human abdomen scans and a sliced lotus root, this constitutes a difficult task for a neural network since it cannot rely on the strong correlations between scans of human patients.

(a) $\boldsymbol{f}$

(b) $\boldsymbol{f}_{\text{FBP}}$
RE: 0.85, HaarPSI: 0.18

(c) $\boldsymbol{f}_{\text{TV}}$
RE: 0.29, HaarPSI: 0.42

(d) $\mathcal{NN}_{\boldsymbol{\theta}}(\boldsymbol{f}_{\text{FBP}})$
RE: 0.19, HaarPSI: 0.74

(e) $\mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}_{\text{FBP}}))$
RE: 0.22, HaarPSI: 0.67

(f) $\boldsymbol{f}_{\text{LtI}}$
RE: 0.11, HaarPSI: 0.81

(g) $|\boldsymbol{f} - \mathcal{NN}_{\boldsymbol{\theta}}(\boldsymbol{f}_{\text{FBP}})|$

(h) $|\boldsymbol{f} - \mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}_{\text{FBP}}))|$
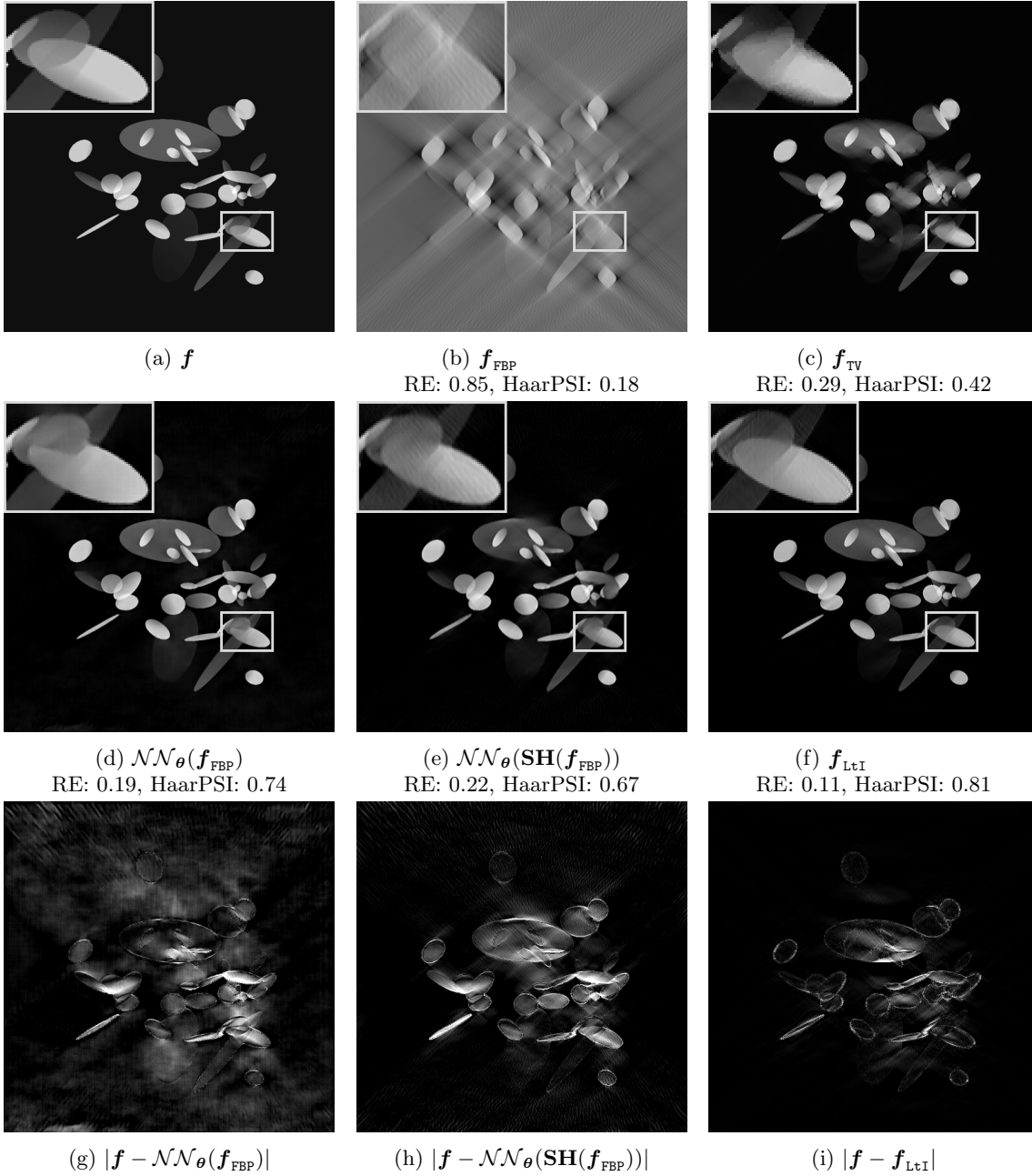
(i) $|\boldsymbol{f} - \boldsymbol{f}_{\text{LtI}}|$

Figure 9: Visualization of the results for one test image in Ellipses-50°. The last row shows the absolute value of the difference plots with respect to the ground truth image in the same plotting window. See Table 1 for averaged similarity measures over the test set.

In Figure 12, we have visualized the reconstruction results for Lotus-60°. Note that a reference scan can be found in Figure 8 and that for the sake of brevity we have omitted the result of $\mathcal{NN}_{\boldsymbol{\theta}}(\boldsymbol{f}_{\text{FBP}})$, which was very similar to $\mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}_{\text{FBP}}))$. The model-based reconstruction methods of the first row are no surprise: non-linear ℓ^1-regularization yields considerably better reconstructions than $\boldsymbol{f}_{\text{FBP}}$, with $\boldsymbol{f}_{\text{TV}}$ and $\boldsymbol{f}^*$ being of similar quality. The learning based methods of the images (d) and (e) reveal that both CNNs do not generalize well across different data sets. The network of $\boldsymbol{f}_{[31]}$ suffers from contrast changes, does not remove more streaking artifacts than ℓ^1-regularization and shows fluctuations at the invisible boundaries. The PhantomNet architecture of Figure 12(e) seems to cope even worse when working on $\boldsymbol{f}_{\text{FBP}}$. In contrast, our proposed scheme reconstructs the invisible outer shape of the lotus
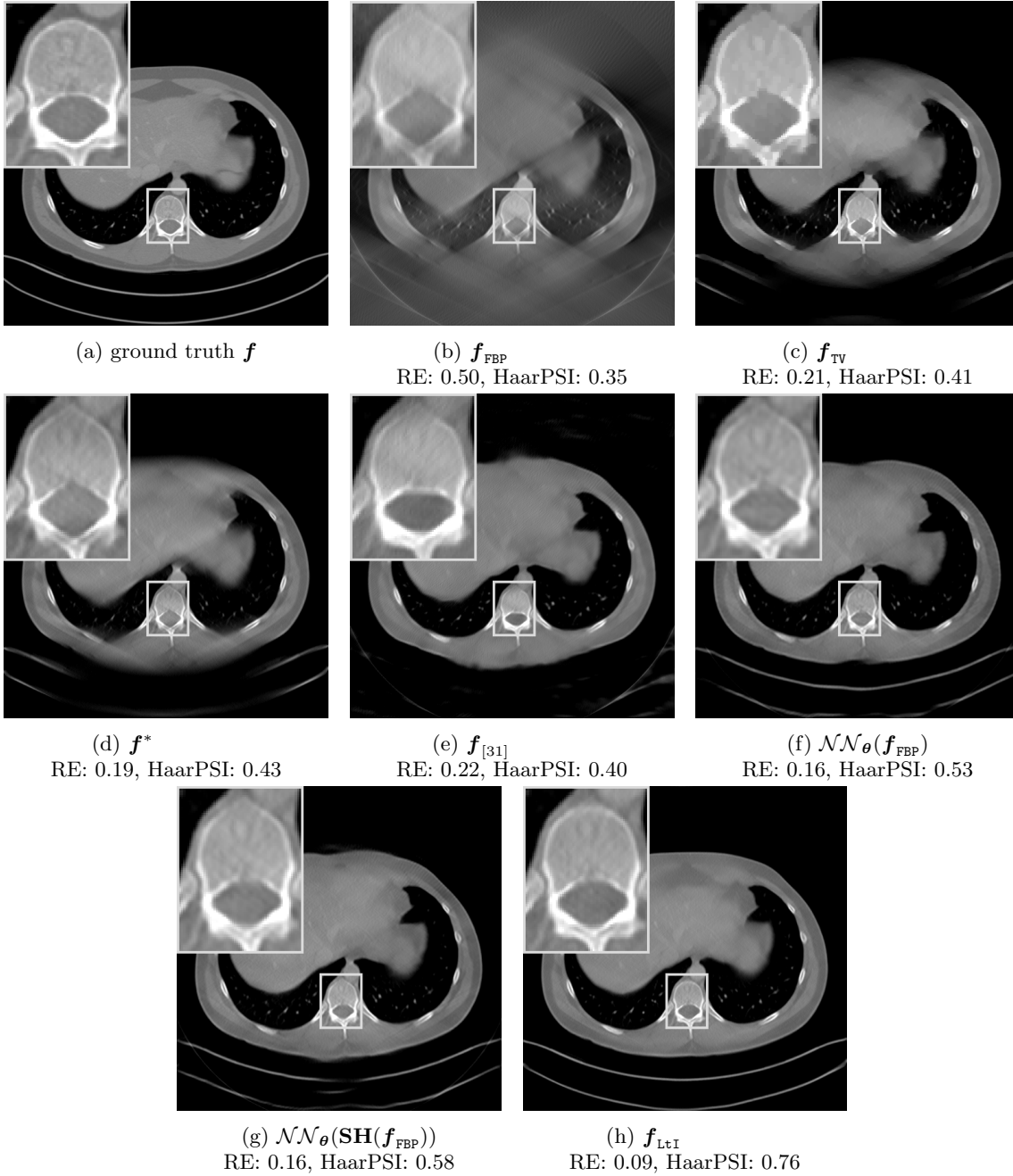
(a) ground truth $\boldsymbol{f}$

(b) $\boldsymbol{f}_{\text{FBP}}$
RE: 0.50, HaarPSI: 0.35

(c) $\boldsymbol{f}_{\text{TV}}$
RE: 0.21, HaarPSI: 0.41

(d) $\boldsymbol{f}^*$
RE: 0.19, HaarPSI: 0.43

(e) $\boldsymbol{f}_{[31]}$
RE: 0.22, HaarPSI: 0.40

(f) $\mathcal{NN}_{\boldsymbol{\theta}}(\boldsymbol{f}_{\text{FBP}})$
RE: 0.16, HaarPSI: 0.53

(g) $\mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}_{\text{FBP}}))$
RE: 0.16, HaarPSI: 0.58

(h) $\boldsymbol{f}_{\text{LtI}}$
RE: 0.09, HaarPSI: 0.76

Figure 10: Comparison for one slice of the test patient of Mayo-60°, i.e., coresponding to a missing wedge of 60°. The plotting window is slightly adapted for better contrast. See Table 2 for averaged similarity measures across all slices of the test patient.

almost perfectly and improves on the streaking artifacts of $\boldsymbol{f}^*$. Such a remarkable generalization reveals the power of our method: relying on the well reconstructed visible coefficients of $\mathbf{SH}(\boldsymbol{f}^*)$ allows for a more precise estimation of the invisible ones. The reliable combination of the visible and inferred invisible coefficients ensures that the CNN $\mathcal{NN}_{\boldsymbol{\theta}}$ does not "waste" its expressiveness on denoising the visible coefficients.

Similar, but less drastic observations hold true for the experiment Lotus-75°, displayed in Figure 13. The methods in (d) and (e) mostly succeed in estimating the invisible coefficients, however, there are still severe fluctuations visible. When comparing the result of Figure 13(f) with the quality achieved
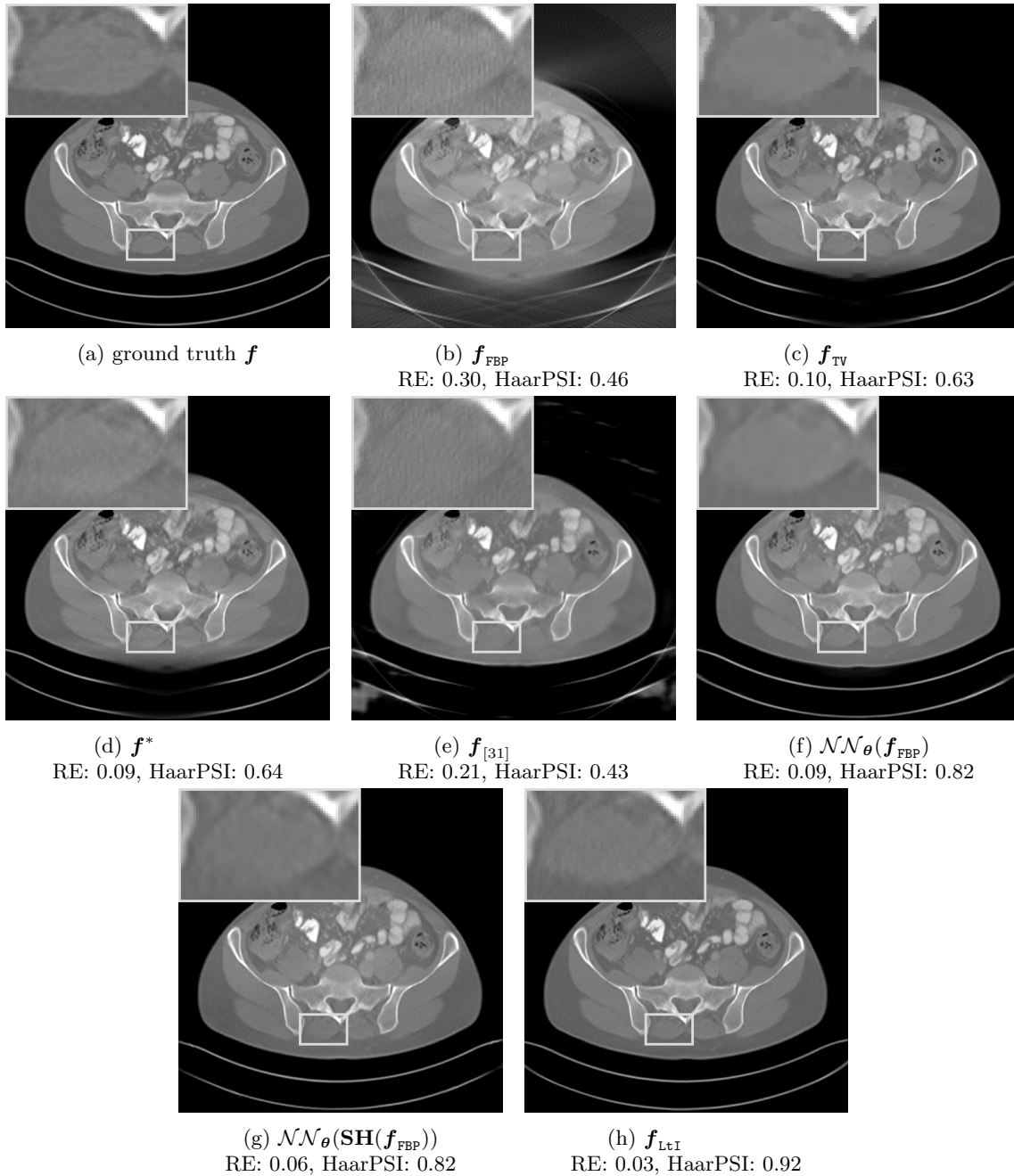
(a) ground truth $\boldsymbol{f}$

(b) $\boldsymbol{f}_{\text{FBP}}$
RE: 0.30, HaarPSI: 0.46

(c) $\boldsymbol{f}_{\text{TV}}$
RE: 0.10, HaarPSI: 0.63

(d) $\boldsymbol{f}^*$
RE: 0.09, HaarPSI: 0.64

(e) $\boldsymbol{f}_{[31]}$
RE: 0.21, HaarPSI: 0.43

(f) $\mathcal{NN}_{\boldsymbol{\theta}}(\boldsymbol{f}_{\text{FBP}})$
RE: 0.09, HaarPSI: 0.82

(g) $\mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}_{\text{FBP}}))$
RE: 0.06, HaarPSI: 0.82

(h) $\boldsymbol{f}_{\text{LtI}}$
RE: 0.03, HaarPSI: 0.92

Figure 11: Comparison for one slice of the test patient of Mayo-75°, i.e., coresponding to a missing wedge of 30°. The plotting window is slightly adapted for better contrast. See Table 2 for averaged similarity measures across all slices of the test patient.

on the test patient in Figure 11(f), a decrease in performance is visible. Nonetheless, keeping in mind that $\sim 16\%$ of the measurements are missing and that $\mathcal{NN}_{\boldsymbol{\theta}}$ has never "seen" a lotus root before, the reconstruction $\boldsymbol{f}_{\text{LtI}}$ remains noteworthy.

6. Conclusion

In the present paper, we have introduced a reconstruction method for limited angle CT where the missing gaps in the wavefront set are closed by means of a deep neural network. We have shown that the visible boundary parts can be accessed by ℓ^1-regularization with shearlets, a directional sensitive
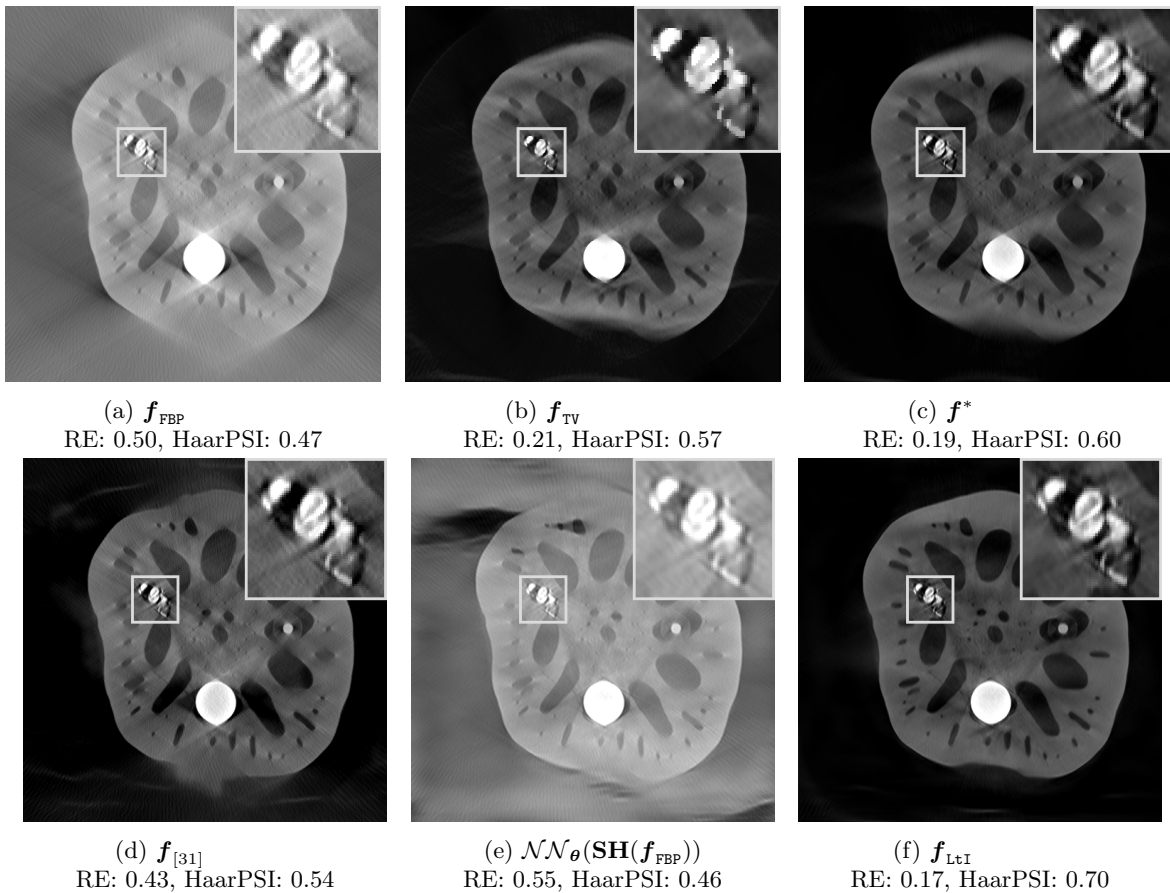
(a) $\boldsymbol{f}_{\text{FBP}}$
RE: 0.50, HaarPSI: 0.47

(b) $\boldsymbol{f}_{\text{TV}}$
RE: 0.21, HaarPSI: 0.57

(c) $\boldsymbol{f}^*$
RE: 0.19, HaarPSI: 0.60

(d) $\boldsymbol{f}_{[31]}$
RE: 0.43, HaarPSI: 0.54

(e) $\mathcal{NN}_{\boldsymbol{\theta}}(\mathbf{SH}(\boldsymbol{f}_{\text{FBP}}))$
RE: 0.55, HaarPSI: 0.46

(f) $\boldsymbol{f}_{\text{LtI}}$
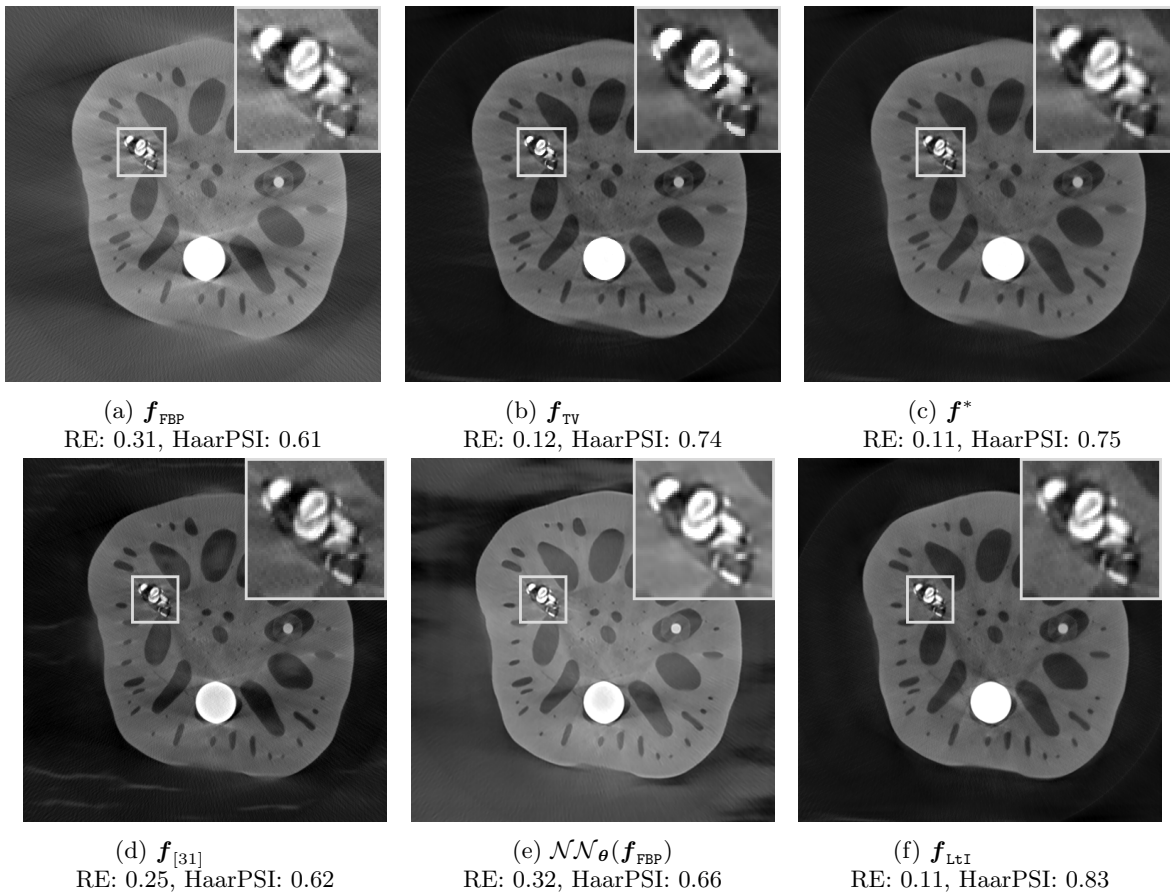RE: 0.17, HaarPSI: 0.70

Figure 12: Evaluation of the generalization properties on Lotus-60°. The neural networks have been trained on Mayo-60° and are then tested on the Lotus-60° measurements. A reference image can be found in Figure 8. The plotting window is slightly adapted for better contrast.

function system that is proven to resolve wavefront sets. Based on such an accurate reconstruction of the visible shearlet coefficients, we have trained a U-Net like architecture PhantomNet for an estimation of the unknown, invisible coefficients.

The close coupling of an advanced model-based method and a custom-tailored learning task implies an increased reliability of the final results and a better understanding of the post-processing capabilities of deep neural networks in the context of limited angle CT (cf. the discussion in Section 4.2). We have furthermore shown superior reconstruction quality when compared with classical methods and less model-oriented deep learning algorithms.

We hope that our framework for limited angle CT will spark further interest in applying similar ideas to other inverse problems. We anticipate that this might be particularly fruitful for problems, where larger amounts of data are not acquired in the measurements, as it is for instance the case in region of interest and exterior tomography [6]. Furthermore, we have observed that we could improve the learning phase by optimizing over a more sophisticated loss function defined over the image domain or by adding additional regularizers. It might be worthwhile to further pursue this line of thought in the future.

(a) $\boldsymbol{f}_{\mathrm{FBP}}$
RE: 0.31, HaarPSI: 0.61

(b) $\boldsymbol{f}_{\mathrm{TV}}$
RE: 0.12, HaarPSI: 0.74

(c) $\boldsymbol{f}^{*}$
RE: 0.11, HaarPSI: 0.75

(d) $\boldsymbol{f}_{[31]}$
RE: 0.25, HaarPSI: 0.62

(e) $\mathcal{NN}_{\boldsymbol{\theta}}(\boldsymbol{f}_{\mathrm{FBP}})$
RE: 0.32, HaarPSI: 0.66

(f) $\boldsymbol{f}_{\mathrm{LtI}}$
RE: 0.11, HaarPSI: 0.83

Figure 13: Evaluation of the generalization properties on Lotus-75°. The neural networks have been trained on Mayo-60° and are then tested on the Lotus-75° data. A reference image can be found in Figure 8. The plotting window is slightly adapted for better contrast.

References

1. "phantom", *Collins dictionary,* `https://www.collinsdictionary.com/`, Retrieved: 2018-10-26.

2. M. Abadi, P. Barham, J. Chen, et al., *TensorFlow: A System for Large-Scale Machine Learning*, Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI), vol. 16, 2016, pp. 265–283.

3. J. Adler and O. Öktem, *Solving ill-posed inverse problems using iterative deep neural networks*, Inverse Probl. **33** (2017), no. 12, 124007.

4. _____ , *Learned Primal-dual Reconstruction*, IEEE T. Med. Imaging **37** (2018), no. 6, 1322–1332.
5. W. Baumeister, R. Grimm, and J. Walz, *Electron tomography of molecules and cells*, Trends Cell Biol. **9** (1999), no. 2, 81–85.
6. L. Borg, J. Frikel, J. S. Jorgensen, and E. T. Quinto, *Analyzing Reconstruction Artifacts from Arbitrary Incomplete X-ray CT Data*, arXiv preprint arXiv:1707.03055 (2018).
7. S. Boyd, N. Parikh, E. Chu, B. Peleato, et al., *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Found. Trends Mach. Learn. **3** (2011), no. 1, 1–122.
8. T. A. Bubba, A. Hauptmann, S. Huotari, J. Rimpeläinen, et al., *Tomographic X-ray data of a lotus root filled with attenuating objects*, arXiv preprint arXiv:1609.07299 (2016).
9. T. A. Bubba, F. Porta, G. Zanghirati, and S. Bonettini, *A nonsmooth regularization approach based on shearlets for Poisson noise removal in ROI tomography*, Appl. Math. Comput. **318** (2018), 131 – 152.
10. H. C. Burger, C. J. Schuler, and S. Harmeling, *Image denoising: Can plain neural networks compete with BM3D?*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 2392–2399.
11. E. J. Candès and D. L. Donoho, *Curvelets and reconstruction of images from noisy radon data*, Proc. SPIE 4119, Wavelet applications in signal and image processing VIII, vol. 4119, 2000, pp. 108–118.
12. _____ , *New Tight Frames of Curvelets and Optimal Representations of Objects with Piecewise C^2 Singularities*, Commun. Pur. Appl. Math. (2002), 219–266.
13. E. J. Candes, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inf. Theor. **52** (2006), no. 2, 489–509.
14. O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhäuser, Boston, 2003.
15. F. Colonna, G. Easley, K. Guo, and D. Labate, *Radon transform inversion using the shearlet representation*, Appl. Comput. Harmon. A. **29** (2010), no. 2, 232 – 250.
16. I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Commun. Pur. Appl. Math. **57** (2004), no. 11, 1413–1457.
17. M. E. Davison, *The Ill-Conditioned Nature of the Limited Angle Tomography Problem*, SIAM J. Appl. Math. **43** (1983), no. 2, 428–448.
18. D. L. Donoho, *Sparse Components of Images and Optimal Atomic Decompositions*, Constr. Approx. **17** (2001), no. 3, 353–382.
19. _____ , *Compressed sensing*, IEEE Trans. Inf. Theory **52** (2006), no. 4, 1289–1306.
20. J. Douglas and H. H. Rachford, *On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables*, T. Am. Math. Soc. **82** (1956), no. 2, 421–439.
21. G. R. Easley, D. Labate, and F. Colonna, *Shearlet-Based Total Variation Diffusion for Denoising*, IEEE T. Image Process. **18** (2009), no. 2, 260–268.
22. M. Elad, *Sparse and redundant representations: From theory to applications in signal and image processing*, Springer, New York, 2010.
23. C. L. Epstein, *Introduction to the Mathematics of Medical Imaging*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2008.
24. S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, Applied and Numerical Harmonic Analysis, Birkhäuser, 2013.
25. J. Frikel, *Sparse regularization in limited angle tomography*, Appl. Comput. Harmon. A. **34** (2013), no. 1, 117–141.
26. J. Frikel and E. T. Quinto, *Characterization and reduction of artifacts in limited angle tomography*, Inverse Probl. **29** (2013), no. 12, 125007.
27. K. Fukushima, *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*, Biol. Cybern. **36** (1980), no. 4, 193–202.
28. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016, http://www.deeplearningbook.org.
29. K. Gregor and Y. LeCun, *Learning fast approximations of sparse coding*, International Conference on Machine Learning (ICML), 2010, pp. 399–406.
30. P. Grohs, *Continuous shearlet frames and resolution of the wavefront set*, Monats. Math. **164** (2011), no. 4, 393–426.
31. J. Gu and J. C. Ye, *Multi-scale wavelet domain residual learning for limited-angle CT reconstruction*, Procs Fully3D (2017), 443–447.
32. K. Guo and D. Labate, *Optimally Sparse Multidimensional Representation Using Shearlets*, SIAM J. Math. Anal. **39** (2007), no. 1, 298–318.
33. _____ , *Optimal recovery of 3D X-ray tomographic data via shearlet decomposition*, Adv. Comput. Math. **39** (2013), no. 2, 227–255.
34. K. Hammernik, T. Würfl, T. Pock, and A. Maier, *A Deep Learning Architecture for Limited-Angle Computed Tomography Reconstruction*, Bildverarbeitung für die Medizin (BVM), Springer, 2017, pp. 92–97.
35. Y. Han and J. C. Ye, *Framing U-Net via Deep Convolutional Framelets: Application to Sparse-view CT*, IEEE T. Med. Imaging **37** (2018), no. 6, 1418–1429.
36. A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, et al., *Model based learning for accelerated, limited-view 3D photoacoustic tomography*, IEEE T. Med. Imaging **37** (2018), no. 6, 1382–1393.
37. K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
38. K. A. Heiskanen, H. C. Rhim, and P. J. M. Monteiro, *Computer simulations of limited angle tomography of reinforced concrete*, Cement Concrete Res. **21** (1991), no. 4, 625–634.

39. G. Hinton, L. Deng, D. Yu, G. E. Dahl, et al., *Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups*, IEEE Signal Proc. Mag. **29** (2012), no. 6, 82–97.

40. L. Hörmander, *The analysis of linear partial differential operators. I. Distribution theory and Fourier analysis. Reprint of the 2nd edition 1990*, Springer, Berlin, 2003.

41. ———, *The analysis of linear partial differential operators. III. Pseudo-differential operators. Reprint of the 1994 edition.*, Springer, Berlin, 2007.

42. G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, *Densely Connected Convolutional Networks*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269.

43. S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, *The One Hundred Layers Tiramisu: Fully Convolutional Densenets for Semantic Segmentation*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1175–1183.

44. K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, *Deep Convolutional Neural Network for Inverse Problems in Imaging*, IEEE T. Image Process. **26** (2017), no. 9, 4509–4522.

45. J. S. Jørgensen, S. B. Coban, W. R. B. Lionheart, S. A. McDonald, and P. J Withers, *SparseBeads data: benchmarking sparsity-regularized computed tomography*, Meas. Sci. Technol. **28** (2017), no. 12, 124005.

46. J. S. Jørgensen and E. Y. Sidky, *How little data is enough? Phase-diagram analysis of sparsity-regularized X-ray computed tomography*, Phil. Trans. R. Soc. A **373** (2015), no. 2043, 20140387.

47. E. Kang, J. Min, and J. C. Ye, *A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction*, Med. Phys. **44** (2017), no. 10, 360–375.

48. D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv preprint arXiv:1412.6980 (2014).

49. E. Klann, E. T. Quinto, and R. Ramlau, *Wavelet methods for a weighted sparsity penalty for region of interest tomography*, Inverse Probl. **31** (2015), no. 2, 025001.

50. V. Kolehmainen, S. Siltanen, S. Järvenpää, J. P. Kaipio, et al., *Statistical inversion for medical x-ray tomography with few radiographs: II. Application to dental radiology*, Phys. Med. Biol. **48** (2003), no. 10, 1465.

51. V. Kolehmainen, A. Vanne, S. Siltanen, S. Järvenpää, et al., *Bayesian inversion method for 3D dental X-ray imaging*, Elektrotech. Inftech. **124** (2007), no. 7-8, 248–253.

52. A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Imagenet Classification with Deep Convolutional Neural Networks*, Advances in Neural Information Processing Systems (NIPS), 2012, pp. 1097–1105.

53. G. Kutyniok and D. Labate, *Resolution of the Wavefront Set Using Continuous Shearlets*, T. Am. Math. Soc. **361** (2009), no. 5, 2719–2754.

54. ———, *Shearlets: Multiscale Analysis for Multivariate Data*, Birkhäuser Basel, 2012.

55. G. Kutyniok, V. Mehrmann, and P. C. Petersen, *Regularization and numerical solution of the inverse scattering problem using shearlet frames*, J. Inverse Ill.-Pose. P. **25** (2017), no. 3, 287–309.

56. D. Labate, W.-Q. Lim, G. Kutyniok, and G. Weiss, *Sparse Multidimensional Representation using Shearlets*, Proc. SPIE 5914, Wavelets XI, vol. 5914, 2005, pp. 5914 – 5923.

57. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, et al., *Backpropagation Applied to Handwritten Zip Code Recognition*, Neural Comput. **1** (1989), no. 4, 541–551.

58. J. Long, E. Shelhamer, and T. Darrell, *Fully Convolutional Networks for Semantic Segmentation*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431–3440.

59. S. Loock and G. Plonka, *Phase retrieval for Fresnel measurements using a shearlet sparsity constraint*, Inverse Probl. **30** (2014), no. 5, 055005.

60. I. Loris, G. Nolet, I. Daubechies, and T. Dahlen, *Tomographic inversion using ℓ^1-norm regularization of wavelet coefficients*, Geophys. J. Int. **170** (2007), no. 1, 359–370.

61. S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd edition ed., Elsevier, Amsterdam, 2009.

62. M. T. McCann, K. H. Jin, and M. Unser, *Convolutional Neural Networks for Inverse Problems in Imaging: A Review*, IEEE Signal Proc. Mag. **34** (2017), no. 6, 85–95.

63. C. McCollough, *Tu-fg-207a-04: Overview of the low dose ct grand challenge*, Med. Physs **43** (2016), no. 6 Part 35, 3759–3760.

64. W. S. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, Bull. Math. Biophys. **5** (1943), no. 4, 115–133.

65. T. Meinhardt, M. Moeller, C. Hazirbas, and D. Cremers, *Learning Proximal Operators: Using Denoising Networks for Regularizing Inverse Imaging Problems*, International Conference on Computer Vision (ICCV), 2017.

66. J. L. Mueller and S. Siltanen, *Linear and Nonlinear Inverse Problems with Practical Applications*, vol. 10, Society for Industrial and Applied Mathematics (SIAM), 2012.

67. F. Natterer, *The Mathematics of Computerized Tomography*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2001.

68. L. V. Nguyen, *How strong are streak artifacts in limited angle computed tomography?*, Inverse Probl. **31** (2015), no. 5, 055003.

69. P. Paschalis, N. D. Giokaris, A. Karabarbounis, G. K. Loudos, et al., *Tomographic image reconstruction using artificial neural networks*, Nucl. Instrum. Methods **527** (2004), no. 1, 211 – 215.

70. D. M. Pelt and K. J. Batenburg, *Fast Tomographic Reconstruction From Limited Data Using Artificial Neural Networks*, IEEE T. Image Process. **22** (2013), no. 12, 5238–5251.

71. E. D. Quinto, *Artifacts and Visible Singularities in Limited Data X-Ray Tomography*, Sens. Imaging **18** (2017), no. 1, 1–14.

72. E. T. Quinto, *Singularities of the X-Ray Transform and Limited Data Tomography in $\mathbb{R}^2$ and $\mathbb{R}^3$*, SIAM J. Math. Anal. **24** (1993), no. 5, 1215–1225.

73. M. Rantala, S. Vanska, S. Jarvenpaa, M. Kalke, et al., *Wavelet-based reconstruction for limited-angle X-ray tomography*, IEEE T. Med. Imaging **25** (2006), no. 2, 210–217.

74. R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, *A Haar Wavelet-Based Perceptual Similarity Index for Image Quality Assessment*, Signal Process. Image **61** (2018), 33–43.

75. L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelrieß, *Improved total variation-based CT image reconstruction applied to clinical data*, Phys. Med. Biol. **56** (2011), no. 6, 1545.

76. O. Ronneberger, P. Fischer, and T. Brox, *U-net: Convolutional Networks for Biomedical Image Segmentation*, International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2015, pp. 234–241.

77. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning representations by back-propagating errors*, Nature **323** (1986), no. 6088, 533.

78. E. Y. Sidky, C.-M. Kao, and X. Pan, *Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT*, J. X-ray Sci. Technol. **14** (2006), no. 2, 119–139.

79. E. Y. Sidky and X. Pan, *Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization*, Phys. Med. Biol. **53** (2008), no. 17, 4777–807.

80. D. Silver, A. Huang, C. J. Maddison, A. Guez, et al., *Mastering the game of Go with deep neural networks and tree search*, Nature **529** (2016), no. 7587, 484.

81. W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, et al., *Fast and flexible X-ray tomography using the ASTRA toolbox*, Opt. Express **24** (2016), no. 22, 25129–25147.

82. A. Veit, M. J. Wilber, and S. Belongie, *Residual Networks Behave Like Ensembles of Relatively Shallow Networks*, Advances in Neural Information Processing Systems (NIPS), 2016, pp. 550–558.

83. F. Voigtlaender, *Shearlet implementation*, https://github.com/dedale-fet/alpha-transform, Accessed: 2018-07-24.

84. S. Wang, Z. Su, L. Ying, X. Peng, et al., *Accelerating magnetic resonance imaging via deep learning*, IEEE International Symposium on Biomedical Imaging (ISBI), 2016, pp. 514–517.

85. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image quality assessment: from error visibility to structural similarity*, IEEE T. Image Process. **13** (2004), no. 4, 600–612.

86. J. Xie, L. Xu, and E. Chen, *Image Denoising and Inpainting with Deep Neural Networks*, Advances in Neural Information Processing Systems (NIPS), 2012, pp. 341–349.

87. Y. Yang, J. Sun, H. Li, and Z. Xu, *Deep ADMM-Net for Compressive Sensing MRI*, Advances in Neural Information Processing Systems (NIPS), 2016, pp. 10–18.

88. H. Zhang, L. Li, K. Qiao, L. Wang, et al., *Image Prediction for Limited-angle Tomography via Deep Learning with Convolutional Neural Network*, arXiv preprint arXiv:1607.08707 (2016).

89. Y. Zhang, H. P. Chan, B. Sahiner, J. Wei, et al., *A comparative study of limited–angle cone-beam reconstruction methods for breast tomosynthesis*, Med. Phys. **33** (2006), no. 10, 3781–3795.

Appendix A. Solving the Analysis Formulation

We solve the minimization problem (4.1) by applying the alternating direction method of multipliers (ADMM) [20, 7]. First, we rewrite (4.1) in the equivalent form

$$\min_{\boldsymbol{f},\boldsymbol{z}} \rho_0 \cdot \left(\frac{1}{2} \left\| \boldsymbol{\mathcal{R}}_\phi \boldsymbol{f} - \boldsymbol{y} \right\|_2^2 + \left\| \Pi_1 \boldsymbol{z} \right\|_{1,\boldsymbol{w}} + \iota_{\geq 0}(\Pi_2 \boldsymbol{z}) \right) \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{f} + \boldsymbol{B}\boldsymbol{z} = \boldsymbol{0},$$

where $\rho_0 > 0$, $\boldsymbol{A}^T = (\rho_1 \, \mathbf{SH}^T, \rho_2 \boldsymbol{I}_{n^2}) \in \mathbb{R}^{n^2 \times (J+1)n^2}$, $\boldsymbol{B} = \operatorname{diag}(-\rho_1 \mathbb{1}_{Jn^2}, -\rho_2 \mathbb{1}_{n^2}) \in \mathbb{R}^{(J+1)n^2 \times (J+1)n^2}$ for $\rho_1, \rho_2 > 0$, $\mathbb{1}_k \in \mathbb{R}^k$ is the vector with all components being 1, and Π_1, Π_2 denote the projections onto the first Jn^2 and the last n^2 entries, respectively. Introducing ρ_0 might seem superfluous, but it serves as a conditioning parameter later on. Then, the scaled form of [7] for $F(\boldsymbol{f}) = \rho_0/2 \left\| \boldsymbol{\mathcal{R}}_\phi \boldsymbol{f} - \boldsymbol{y} \right\|_2^2$ and $G(\boldsymbol{z}) = \rho_0(\left\| \Pi_1 \boldsymbol{z} \right\|_{1,\boldsymbol{w}} + \iota_{\geq 0}(\Pi_2 \boldsymbol{z}))$ results in the following iterates

$$\boldsymbol{f}^{k+1} := \operatorname{argmin}_{\boldsymbol{f}} \left(F(\boldsymbol{f}) + \rho/2 \left\| \boldsymbol{A}\boldsymbol{f} + \boldsymbol{B}\boldsymbol{z}^k + \boldsymbol{u}^k \right\|_2^2 \right) \tag{A.1}$$

$$\boldsymbol{z}^{k+1} := \operatorname{argmin}_{\boldsymbol{z}} \left(G(\boldsymbol{z}) + \rho/2 \left\| \boldsymbol{A}\boldsymbol{f}^{k+1} + \boldsymbol{B}\boldsymbol{z} + \boldsymbol{u}^k \right\|_2^2 \right) \tag{A.2}$$

$$\boldsymbol{u}^{k+1} := \boldsymbol{u}^k + \boldsymbol{A}\boldsymbol{f}^{k+1} + \boldsymbol{B}\boldsymbol{z}^{k+1},$$

where $\rho > 0$. Step (A.1) results in solving the linear system

$$\left(\rho_0 \boldsymbol{\mathcal{R}}_\phi^T \boldsymbol{\mathcal{R}}_\phi + \rho \rho_1^2 \mathbf{SH}^T \mathbf{SH} + \rho \rho_2^2 \boldsymbol{I}_{n^2} \right) \boldsymbol{f}$$
$$= \rho_0 \boldsymbol{\mathcal{R}}_\phi^T \boldsymbol{y} + \rho \rho_1^2 \mathbf{SH}^T (\Pi_1 \boldsymbol{z}^k - \Pi_1 \boldsymbol{u}^k/\rho_1) + \rho \rho_2^2 (\Pi_2 \boldsymbol{z}^k - \Pi_2 \boldsymbol{u}^k/\rho_2).$$

The proximal step of (A.2) decouples into the following two seperate operations

$$\Pi_1 \boldsymbol{z}^{k+1} = \text{shrink}\left(\mathbf{SH}(\boldsymbol{f}^{k+1}) + \Pi_1\boldsymbol{u}^k/\rho_1, \frac{\rho_0\boldsymbol{w}}{\rho\rho_1^2}\right)$$

and

$$\Pi_2\boldsymbol{z}^{k+1} = \max(\boldsymbol{f}^{k+1} + \Pi_2\boldsymbol{u}^k/\rho_2, \boldsymbol{0}),$$

where *shrink* denotes to element-wise *soft-thresholding*, i.e., for $\boldsymbol{a} \in \mathbb{R}^n$ and $\boldsymbol{b} \in \mathbb{R}^n_+$ we set

$$(\text{shrink}(\boldsymbol{a}, \boldsymbol{b}))_i = \begin{cases} \max(|a_i| - b_i, 0)\frac{a_i}{|a_i|}, & \text{if } a_i \neq 0 \\ 0, & \text{else.} \end{cases}$$

After substituting $\rho\rho_i^2$ by ρ_i, and $\Pi_i\boldsymbol{u}^k/\rho_i$ by $\Pi_i\boldsymbol{u}^k$ and using that $\mathbf{SH}^T\mathbf{SH} = \boldsymbol{I}_{n^2}$, we obtain the overall algorithm

$$\boldsymbol{f}^{k+1} := \left(\rho_0\,\boldsymbol{\mathcal{R}}_\phi{}^T\boldsymbol{\mathcal{R}}_\phi + (\rho_1 + \rho_2)\boldsymbol{I}_{n^2}\right)^{-1} \tag{A.3}$$

$$\left(\rho_0\,\boldsymbol{\mathcal{R}}_\phi{}^T\boldsymbol{y} + \rho_1\,\mathbf{SH}^T(\Pi_1\boldsymbol{z}^k - \Pi_1\boldsymbol{u}^k) + \rho_2(\Pi_2\boldsymbol{z}^k - \Pi_2\boldsymbol{u}^k)\right)$$

$$\Pi_1\boldsymbol{z}^{k+1} := \text{shrink}\left(\mathbf{SH}(\boldsymbol{f}^{k+1}) + \Pi_1\boldsymbol{u}^k, \frac{\rho_0\boldsymbol{w}}{\rho_1}\right)$$

$$\Pi_2\boldsymbol{z}^{k+1} := \max(\boldsymbol{f}^{k+1} + \Pi_2\boldsymbol{u}^k, \boldsymbol{0})$$

$$\Pi_1\boldsymbol{u}^{k+1} := \Pi_1\boldsymbol{u}^k + \mathbf{SH}(\boldsymbol{f}^{k+1}) - \Pi_1\boldsymbol{z}^{k+1}$$

$$\Pi_2\boldsymbol{u}^{k+1} := \Pi_2\boldsymbol{u}^k + \boldsymbol{f}^{k+1} - \Pi_2\boldsymbol{z}^{k+1}.$$

The overparametrization by ρ_i is used to balance out data fidelity and regularization for solving the linear system (A.3). Although the algorithm converges independently of the choice of these parameters, setting them properly may lead to significant speed-up of convergence. Note, that it is not necessary to solve (A.3) up to full precision [7]. In all our experiments, we are using the conjugate gradient method with the previous iterate $\boldsymbol{f}^k$ as a warm start for finding an approximate solution. The ADMM is known to converge to modest precision within a few tens of iterations, which is usually enough for imaging purposes [7]. The speed of the overall algorithm is mostly dominated by cost of applying $\boldsymbol{\mathcal{R}}_\phi{}^T\boldsymbol{\mathcal{R}}_\phi$ for solving (A.3).

In all our experiments, we fix $\rho_2 = 1$, such that only ρ_0, ρ_1 and the weight $\boldsymbol{w}$ are left as hyperparameters. We initialize the algorithm with $\boldsymbol{f}^0 := \boldsymbol{\mathcal{R}}_\phi{}^T\boldsymbol{y}$, $\boldsymbol{z}^0 := \boldsymbol{0}$ and $\boldsymbol{u}^0 := \boldsymbol{0}$ and stop after 50 iterations. For the different experiments we are choosing the following parameter setup, which were found by manual tuning:

- $\rho_0 = 0.02, \rho_1 = 0.1$ and $w_j = 3^j/400$, (Ellipses-50°)
- $\rho_0 = 0.50, \rho_1 = 0.1$ and $w_j = 2^j/400$, (Mayo-60°)
- $\rho_0 = 0.08, \rho_1 = 0.5$ and $w_j = 2^j/72$, (Mayo-75°+ Lotus-75°)
- $\rho_0 = 0.01, \rho_1 = 0.1$ and $w_j = 2^j/40$, (Lotus-60°)

where j denotes the j-th shearlet scale. A reconstruction of a 512×512 image with a missing wedge of 60° takes about 2.5 minutes using an Intel i7 processor and 16GB RAM.