On Learning the Invisible in Photoacoustic Tomography with Flat Directionally Sensitive Detector[*]

Bolin Pan[†] and Marta M. Betcke[‡]

Abstract. In photoacoustic tomography (PAT) with flat sensor, we routinely encounter two types of limited data. The first is due to using a finite sensor and is especially perceptible if the region of interest is large relative to the sensor or located farther away from the sensor. In this paper, we focus on the second type caused by a varying sensitivity of the sensor to the incoming wavefront direction which can be modelled as binary i.e. by a cone of sensitivity. Such visibility conditions result, in the Fourier domain, in a restriction of both the image and the data to a bow-tie, akin to the one corresponding to the range of the forward operator. The visible wavefrontsets in image and data domains, are related by the wavefront direction mapping. We adapt the wedge restricted Curvelet decomposition, we previously proposed for the representation of the full PAT data, to separate the visible and invisible wavefronts in the image. We optimally combine fast approximate operators with tailored deep neural network architectures into efficient learned reconstruction methods which perform reconstruction of the visible coefficients and the invisible coefficients are learned from a training set of similar data.

Key words. learned image reconstruction, compressed sensing, Curvelet transform, photoacoustic tomography, fast Fourier methods, limited-view

AMS subject classifications. 94A08, 97R40, 94A12, 92C55, 65T50, 68U10

1. Introduction. Photoacoustic Tomography (PAT) is a hybrid imaging modality that can deliver high resolution *in-vivo* images of optical energy absorbed upon an illumination with a laser-generated near infra-red light pulse [59, 11, 45, 57, 65, 60]. PAT harnesses the photoacoustic effect to encode optical contrast onto broadband ultrasonic waves thereby circumventing the depth and spatial resolution limitations of purely optical imaging techniques.

If the data is measured over a surface surrounding the region of interest for long enough time and using omnidirectional sensors, the reconstruction can be obtained via *time reversal* which amounts to a single solve of a wave equation; see Section 2 for details. In practice, the limited-view problems are introduced due to spatially and/or directionally restricted sampling of the ultrasound (US) waves. An example of the former is a finite size linear/planar sensor placed on one side of the domain; the latter most US sensors' incapability of recording wavefronts impinging on the detector at angles beyond $\pm\theta_{\max}$, $\theta_{\max} < \pi/2$. Another type of limited-view problems result from restricting the length of the recorded time series, a necessity arising in a case of trapping sound speeds. As in this work we focus on constant speed of sound, we do not consider this last type of limited-view problems.

For limited-data problems due to sparse sampling rather than limited-view, compressed sensing (CS) approaches that iteratively minimise a penalty function combining an explicit model of US propagation with constraints on the image as regularisation [39, 10, 8, 34, 40, 15,

52, 46, 29, 31, 27, 28, 30, 47] have proven to provide significantly better reconstructions than *time reversal* and more general Neumann series based iterative methods which lack inherent regularisation. A shared drawback of all these methods is high computational complexity of iterating with the forward and adjoint/inverse operators. This shortcoming has been addressed by two step methods which reconstruct complete data first and then obtain the image through single time reversal [34, 13, 46], which however usually sacrifice some reconstruction quality for efficiency.

For the considered limited-view problems the situation is direr. Time reversal and Neumann series methods as expected only reconstruct the visible singularities producing images with characteristic limited-view artefacts and lack noise suppression properties of CS methods [63]. Performance of compressed sensing depends on the choice of the prior, e.g. while total variation (TV) yields overall better reconstructions than Curvelet sparsity (reasons for which are discussed in Subsection 6.3), even for TV the directions of singularities missing from the data result in inherent blur in the reconstructed images.

A decade after the breakthrough in Deep Neural Networks, Deep Learning (DL) techniques are ubiquitous in tomographic imaging [42, 41, 35, 1, 66, 2, 9]. In comparison to CS, the learned approaches i) need fewer applications of the forward/adjoint operator which leads to more efficient reconstruction algorithms, ii) have the ability to extract prior information about the image from a training set of similar images which makes them particularly suitable for e.g. medical image reconstruction where patient studies fit such scenario. DL approaches to PAT image reconstruction mainly fall in one of the two categories: i) learned post-processing of a simple reconstruction such as e.g. pseudo-inverse, ii) learned iterative reconstruction (also referred to as model based learning).

The learned post-processing usually employs a network with many layers and learnable parameters capable of encoding complicated image priors. The learned post-processing of the universal back-projection [61] was studied in [6, 7, 5, 51, 50, 44], of the first iteration of an averaged time reversal in [53] and of time reversal in [33]. All but the last work use standard U-Net architecture, while the last uses a dense U-Net (a U-Net with dense blocks and skip connections). The main advantage of learned post-processing is that an initial simple reconstruction can be decoupled from the iterative training process. This obviously limits the impact of the physical model on the reconstruction procedure and hence makes the process less robust and more dependent on the training set used.

Learned iterative reconstruction approaches are based on unrolling of iterative solvers and parametrising the proximal map with neural networks, usually much smaller than those used in learned post-processing, e.g. [54, 17, 16, 14, 64]. Due to repeated application of the forward and adjoint/approximate inverse operators on the training set, learned iterative reconstruction methods for 3D PAT suffer from unreasonable training times. To tackle this, a greedy training for learned gradient schemes for 3D PAT was suggested in [38]. The same authors later considered use of efficient Fourier domain approximate forward model to accelerate the training process [36], which is also the direction we take in this work.

Other proposed approaches include learned pre-processing of the data prior to reconstruction, fully learned methods completely bypassing the physics of the model, and methods that aim to learn some representation of the regulariser for use in variational framework, we refer to [37] for a review of such approaches specific to PAT.

1.1. Related Work. In a series of papers Frikel and Quinto [47, 27, 30, 28, 48] characterised the visibility of singularities in X-ray tomography and in [63, 31] in PAT with finite omnidirectional flat sensor ($\theta_{max} = \pi/2$). The limitations of CS approaches with Curvelet sparsity regularisation were discussed in [27, 28] in the context of X-ray tomography.

Motivated by these works, a number of authors considered implications of the visible and invisible singularities for learning methods in limited-angle tomography in X-ray and in general [19, 18, 3]. The approach proposed in [3] capitalizes on the microlocal properties of the Radon transform to inpaint the singularities extracted from the data, using the network proposed in [4], into the reconstructed image. In [18] the authors reinterpret the unrolled ISTA iteration as CNN layers which coefficients and structure are inferred from the convolutional properties of Fourier integral operators and pseudo-differential operators and apply it to limited-angle X-ray tomography. The present work is closest in spirit to the approach proposed in [19] for the limited-angle X-ray tomography which promotes the idea to separate the visible part of the image, which can be stably reconstructed in Shearlets frame using CS approach, from the invisible part of the image which needs to be learned. The main benefits of such approach are i) a clear separation between what is stably reconstructed from the measured data and the prior that is learned from the training set and ii) decoupling of the reconstruction of the visible which requires iterating with the forward/adjoint operator from learning the invisible which does not, as the invisible is contained in the null space of the discrete forward operator. A similar idea was proposed from an algebraic perspective as null space learning in [49]. Directional frames like Curvelets or Shearlets have the significant advantage of providing a microlocal representation of the null space and its complement.

1.2. Motivation. In this work we focus on the limited-view problem due to varying sensitivity of the sensor to the direction of the wavefront impinging on the detector. The directional sensitivity can be modelled as binary with a maximal impingement angle $\pm\theta_{max}$, $\theta_{max} < \pi/2$, for the wavefronts to be registered, resulting in a cone of sensitivity around the direction normal to the sensor. We observe this problem shares many similarities with the limited-angle parallel X-ray, in particular the decomposition into visible/invisible singularities in the Fourier domain corresponds to the partition of the angular range of the singularities regardless of their location, exactly as for the parallel X-ray transform. We note that this is not the case for PAT with a flat sensor of finite size where the angular range of visible singularities varies with the position of the singularity.

The decomposition into visible and invisible singularities for limited-angle parallel X-ray geometry underpins the result in [28] on visible singularities only recovery via compressed sensing in Curvelets or Shearlets frames. Based on this result, the authors of [19] proposed to reconstruct the visible and learn the invisible in Shearlet frame for the limited-angle parallel X-ray tomography. These results along with the above observation of similarities between the limited-angle parallel X-ray CT and AT with flat sensor with limited sensitivity angle, *limited-angle PAT* for short, motivate our work.

1.3. Contribution. Using the wavefront mapping (**WfM**) [46], we obtain the correspondence between the visible and invisible wavefront directions in PAT image and PAT data. This allows us the observation that the limited sensitivity angle restricts the data range in a way that enables the use of efficient Fourier domain forward/adjoint and pseudoinverse operators

derived in Subsections 3.1 and 3.3 and Appendix A.

Due to the visible/invisible parts inducing an angular range partition in the limited-angle PAT, we can adapt the wedge restricted Curvelet transform proposed in [46] for a sparse PAT data representation, to a sparse representation of the visible/invisible components of the PAT image (see Subsection 4.2) and recover the sparse wedge restricted Curvelet coefficients of the visible image component via CS with the data fidelity term using the aforementioned limited-angle PAT Fourier operators. The recovered visible component of the PAT image is then represented by its Coronae coefficients, visible Coronae coefficients for short (see Subsections 4.4 and 4.5), resulting in a minimal representation which perfectly matches the split into the visible/invisible and inherits the multiscale structure of the Curvelet frame. These visible Coronae coefficients are fed into a Coronae-Net, a tailored U-Net inspired architecture, designed and trained to fill in the invisible component of the PAT image given the visible component.

1.4. Outline. The reminder of the paper is organized as follows. Section 2 briefly introduces the time domain formulation of the forward, adjoint and inverse problems in PAT. In Section 3 we summarise the existing results on Fourier domain PAT and stipulate the limited-angle Fourier PAT operators. The new, to best of our knowledge, derivation of the adjoint in Fourier domain is presented in Appendix A. In Section 4 we introduce the various multiscale representations used throughout the paper, discuss their roles and mutual relations and justify their suitability for representation of the initial pressure in limited-angle PAT. In Section 5 we outline our framework for reconstructing the visible via the solution of a CS problem using the Fourier domain limited-angle PAT operators and a sparse representation in a fully wedge restricted Curvelet frame, and learning the invisible via the proposed Coronae-Net working on Coronae coefficients of the visible and invisible parts of the initial pressure image. We also extend the framework to real world scenario with imperfect visible coefficients via residual learning, ResCoronae-Net. We illustrate the performance of our framework in both perfect and imperfect learning scenarios and its generalisation potential on a fully synthetic ellipse data set in Section 6 and on a more realistic DRIVE retina vessel data set in Section 7. Section 8 provides conclusions and outlook.

2. Photoacoustic Tomography. With several assumptions on the tissue properties [58], the photoacoustic forward problem can be modelled as an initial value problem for the free space wave equation

(A)
$$\underbrace{\left(\frac{1}{c^2(\mathbf{x})}\frac{\partial^2}{\partial t^2} - \nabla^2\right)}_{:=\square^2}p(t,\mathbf{x}) = 0, \quad (t,\mathbf{x}) \in (0,T) \times \mathbb{R}^d,$$
$$p(0,\mathbf{x}) = p_0(\mathbf{x}),$$
$$p_t(0,\mathbf{x}) = 0,$$

where $p(t,\mathbf{x}) \in \mathcal{C}^\infty((0,T) \times \mathbb{R}^d)$ is a time dependent acoustic pressure recorded for a finite time T, $p_0(\mathbf{x}) \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ is its initial value, and $c(\mathbf{x}) \in \mathcal{C}^\infty(\mathbb{R}^d)$ is the speed of sound in the tissue.

The photoacoustic inverse problem recovers the initial pressure $p_0(\mathbf{x})$ in the region of interest Ω on which $p_0(\mathbf{x})$ is compactly supported, from time dependent measurements

$$g(t, \mathbf{x}_\mathcal{S}) = \omega(t)p(t, \mathbf{x}_\mathcal{S}), \quad (t, \mathbf{x}_\mathcal{S}) \in (0, T) \times \mathcal{S}$$

at a set of locations $\mathbf{x}_\mathcal{S} \in \mathcal{S} \subset \mathbb{R}^d$, e.g. the boundary of Ω, where $\omega \in \mathcal{C}_0^\infty(0, T)$ is a temporal smooth cut-off function. It amounts to a solution of the following initial value problem for the wave equation with constraints on the surface [26]

$$\begin{aligned}
&\square^2 q(t, \mathbf{x}) = 0, && (t, \mathbf{x}) \in (0, T) \times \mathbb{R}^d, \\
\textbf{(TR)} \quad &q(0, \mathbf{x}) = 0, \; q_t(0, \mathbf{x}) = 0, \\
&q(t, \mathbf{x}) = g(T - t, \mathbf{x}_\mathcal{S}), && (t, \mathbf{x}_\mathcal{S}) \in (0, T) \times \mathcal{S},
\end{aligned}$$

evaluated at T, $q(T, \mathbf{x})$, also referred to as a *time reversal*. For non-trapping smooth $c(\mathbf{x})$, in 3D, the solution of (**TR**) is the solution of the PAT inverse problems if T is chosen large enough so that $g(t, \mathbf{x}) = 0$ for all $\mathbf{x} \in \Omega$, $t \geq T$ and the wave has left the domain Ω. Assuming that the measurement surface $\mathcal{S}$ surrounds the region of interest Ω containing the support of initial pressure p_0, the wave equation (**TR**) has a unique solution.

When no complete data is available (meaning that some singularities have not been observed at all as opposed to the case where partial energy has been recorded, see e.g. [56]), p_0 is usually recovered in a variational framework, including some additional penalty functional corresponding to prior knowledge about the solution. Iterative methods for solution of such problems require application of the adjoint operator which amounts to the solution of the following initial value problem with a time varying mass source [10]

$$\begin{aligned}
\textbf{(A}^\star\textbf{)} \quad &\square^2 q(t, \mathbf{x}) = \begin{cases} \frac{\partial}{\partial t} g(T - t, \mathbf{x}_\mathcal{S}), & (t, \mathbf{x}_\mathcal{S}) \in (0, T) \times \mathcal{S} \\ 0 & \text{everywhere else} \end{cases} \\
&q(0, \mathbf{x}) = 0, \\
&q_t(0, \mathbf{x}) = 0,
\end{aligned}$$

evaluated at T, $q(T, \mathbf{x})$.

3. Fourier Domain Formulation of PAT. The operators defined in this section map between Fourier transforms of the initial pressure p_0 and the data g. When using Fourier domain PAT operators in conjunction with Curvelet frame (which is also defined and computed via the Fourier transform), all the computations can be carried out directly in the Fourier domain without need to compute intermediate Fourier transforms.

3.1. Forward, Adjoint and Inverse PAT Operators in the Fourier Domain. We consider a PAT geometry with a line sensor in 2D (a plane sensor in 3D). Let $\mathbf{x} = (\mathbf{x}_\perp, \mathbf{x}_\mathcal{S}) \in \mathbb{R} \times \mathbb{R}^{d-1}$ and the sensor hyperplane be $\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{x}_\perp = 0\}$. Then $\mathbf{x}_\mathcal{S}$ describes the component parallel to the sensor and $\mathbf{x}_\perp$ the component perpendicular to the sensor. Assuming half space symmetry w.r.t. the sensor plane with a constant speed of sound $c(\mathbf{x}) = c$, the solution of the wave equation (**A**) in the Fourier domain can be explicitly written as [24]

$$(3.1) \qquad \hat{p}(ct, \mathbf{k}) = \cos(c|\mathbf{k}|t)\hat{p}_0(\mathbf{k}),$$

where $\hat{u}$ denotes the Fourier transform of u in some or all variables and $\mathbf{k} \in \mathbb{R}^d$ is the Fourier domain wave vector. Using (3.1), we can obtain the following *forward* mapping between the initial pressure $p_0(\mathbf{x}_\perp, \mathbf{x}_\mathcal{S})$, $(\mathbf{x}_\perp, \mathbf{x}_\mathcal{S}) \in \mathbb{R} \times \mathbb{R}^{d-1}$ and the PAT data $g(ct, \mathbf{x}_S)$, $(ct, \mathbf{x}_S) \in \mathbb{R} \times \mathbb{R}^{d-1}$, with $g(ct, \mathbf{x}_S) = p(ct, \mathbf{x}_S)^1$, in the Fourier domain [24, 43, 62, 13]

$$(\hat{\mathbf{A}}) \qquad \hat{g}(\omega/c, \mathbf{k}_\mathcal{S}) = \frac{\omega/c}{\sqrt{(\omega/c)^2 - |\mathbf{k}_\mathcal{S}|^2}} \hat{p}_0 \left(\sqrt{(\omega/c)^2 - |\mathbf{k}_\mathcal{S}|^2}, \mathbf{k}_\mathcal{S} \right),$$

where $\mathbf{k} = (\mathbf{k}_\perp, \mathbf{k}_\mathcal{S})$ are the Fourier domain counterparts to the ambient space variables $\mathbf{x} = (\mathbf{x}_\perp, \mathbf{x}_\mathcal{S})$. In the above formulae the time t, its frequency ω and $\mathbf{k}_\perp$ are understood as non-negative quantities, with symmetries defining the functions on their negatives. In particular, formula (3.1) has an implicit assumption of even symmetry of $p_0(\mathbf{x}_\perp, \mathbf{x}_\mathcal{S})$, $g(ct, \mathbf{x}_S)$ in $\mathbb{R}^d$ w.r.t. $\mathbf{x}_\perp = 0$, $ct = 0$, respectively,

$$(3.2) \qquad \hat{p}_0(-\mathbf{k}_\perp, \mathbf{k}_S) = \hat{p}_0(\mathbf{k}_\perp, \mathbf{k}_S),$$
$$\hat{g}(-\omega/c, \mathbf{k}_S) = \hat{g}(\omega/c, \mathbf{k}_S).$$

For the right hand side of equation $(\hat{\mathbf{A}})$ to be well-defined and bounded, the expression under the square root should be bounded away from zero for all non-zero wave vectors i.e. $\mathcal{R}_{\mathbf{A}} = \left\{ (\omega/c, \mathbf{k}_\mathcal{S}) \in \mathbb{R}^d : |\mathbf{k}_\mathcal{S}| < |\omega/c| \right\} \cup \mathbf{0}$, resulting in a bow-tie shaped range of photoacoustic operator in the sound-speed normalised Fourier space, see Figure 2 (b). We note that $(\hat{\mathbf{A}})$ turned around underlies the inverse transform (which is stable) [43], but as forward transform the factor blows up and regularisation is required [36].

As already alluded, the inverse PAT Fourier operator follows immediately from $(\hat{\mathbf{A}})$,

$$(\hat{\mathbf{A}}^{-1}) \qquad \hat{p}_0(\mathbf{k}_\perp, \mathbf{k}_\mathcal{S}) = \frac{|\mathbf{k}_\perp|}{|\mathbf{k}|} \hat{g}(|\mathbf{k}|, \mathbf{k}_\mathbf{S}),$$

with the change of variables induced by the dispersion relation in the wave equation $(\omega/c)^2 = |\mathbf{k}|^2 = |\mathbf{k}_\mathcal{S}|^2 + |\mathbf{k}_\perp|^2$,

$$(3.3) \qquad \mathbf{k} = (\mathbf{k}_\perp, \mathbf{k}_S) \leftarrow \left(\mathrm{sign}\,(\omega) \sqrt{(\omega/c)^2 - |\mathbf{k}_S|^2}, \mathbf{k}_S \right),$$
$$(\omega/c, \mathbf{k}_S) \leftarrow \left(\mathrm{sign}\,(\mathbf{k}_\perp) \left(|\mathbf{k}| = \sqrt{|\mathbf{k}_\perp|^2 + |\mathbf{k}_S|^2} \right), \mathbf{k}_S \right).$$

The even symmetry (3.2) is inherited by the map (3.3).

Using the definition of the adjoint, the adjoint PAT Fourier operator was derived in the Appendix A

$$(\hat{\mathbf{A}}^*) \qquad \hat{p}_0(\mathbf{k}_\perp, \mathbf{k}_\mathcal{S}) = \hat{g}(|\mathbf{k}|, \mathbf{k}_\mathbf{S}).$$

We note that both inverse $(\hat{\mathbf{A}}^{-1})$ and adjoint $(\hat{\mathbf{A}}^*)$ operators only differ by the scaling factor present in the inverse and absent from the adjoint, while the underlying change of variables is the same.

---

[1]Note, that the constant speed of sound c is used to normalise the time/frequency in our Fourier formulation of the operators, the non-normalised (physical) formulation of the forward operator would have an extra factor $1/c$ on the right hand side of $\hat{\mathbf{A}}$ (s.a. [46]) which in turn would affect the scaling of the other operators. It has the consequence that g, as defined here, needs to be scaled by $1/c$ to correspond to physical PAT data which is what we do in simulations.

3.2. Wavefront Direction Mapping. The wavefront mapping (**WfM**) was derived in [46] observing that the change of coordinates (3.3) defines a one-to-one map between the frequency vectors $\mathbf{k} = (\mathbf{k}_\perp, \mathbf{k}_S) \in \mathbb{R}^d$ and $(\omega/c, \mathbf{k}_S) \in \mathcal{R}_\mathbf{A}$. This mapping is illustrated in Figure 1. Figure 1 (a) shows the wavefronts (in blue) and their corresponding wavefront vectors in the ambient space $\bar{\theta} = (-\cos\theta, \sin\theta)$ (on the left) and in the data space $\bar{\beta} = (\cos\beta, -\sin\beta)$ (on the right). Note, that these wavefront vectors have the same direction (and hence angles[2]: $\bar{\theta}$, $\bar{\beta}$) as their frequency domain counterparts $\mathbf{k}_{\bar{\theta}}$, $\mathbf{k}_{\bar{\beta}}$ which are depicted in Figure 1 (b). With

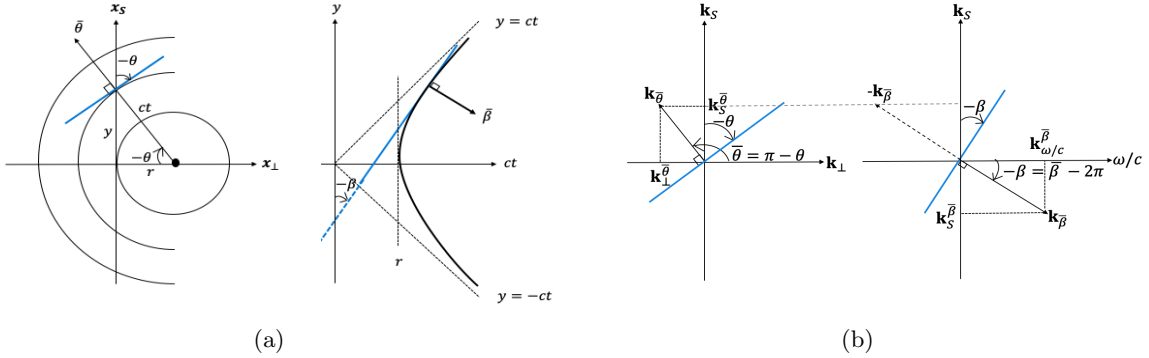(a)   (b)

Figure 1: (a) Wavefront direction mapping between the ambient space $\bar{\theta}$ (left), and the data space, $\bar{\beta}$ (right) on an example of a 2D spherical wave. $-\theta$ and $-\beta$ are the angles that the wavefronts (in blue) make with the sensor (the negative signs are the consequence of the right handed coordinate system). (b) The Fourier domain counterpart, the wavefront vector mapping between $\mathbf{k}_{\bar{\theta}}$ (left), and $\mathbf{k}_{\bar{\beta}}$ (right).

the notation in Figure 1, we deduce the map between the wavefront vector angles $\bar{\theta}$ and $\bar{\beta}$ considering the corresponding frequency domain vectors

$$(3.4) \qquad \tan\bar{\beta} = \frac{\mathbf{k}_\mathcal{S}^{\bar{\beta}}}{\mathbf{k}_{\omega/c}^{\bar{\beta}}} = \frac{-\mathbf{k}_\mathcal{S}^{\bar{\theta}}}{|\mathbf{k}_{\bar{\theta}}| = \sqrt{(|\mathbf{k}_\perp^{\bar{\theta}}|^2 + |\mathbf{k}_\mathcal{S}^{\bar{\theta}}|^2}} = \sin(-\bar{\theta}).$$

We can express the wavefront mapping (3.4) in terms of the angles the wavefronts make with the detector $-\theta$, $-\beta$, using their relation with wavefront vector angles $\bar{\theta} = \pi - \theta$, $\bar{\beta} = 2\pi - \beta$ and basic trigonometric identities

$$(\mathbf{WfM}) \qquad\qquad \beta = \arctan(\sin\theta),$$

where $\theta \in (-\pi/2, \pi/2)$ and $\beta \in (-\pi/4, \pi/4)$ for the un-mirrored initial pressure. The derivation was presented in $\mathbb{R}^d$, $d = 2$ for simplicity, $d = 3$ follows analogously by treating both detector coordinates the same, see also [46].

---

[2]For simplicity of notation, we do not distinguish between the ambient/data space vectors $\bar{\theta}, \bar{\beta}$ and their angles in the right handed coordinate system. If the vector or its angle is meant depends on the context.

3.3. Limited-angle PAT Fourier Operators.
According to the wavefront mapping (**WfM**) we have

$$\tan \beta_{\max} = \sin \theta_{\max},$$

which yields the correspondence between the visible ambient space wavefronts $[-\theta_{\max}, \theta_{\max}]$, $0 \leq \theta_{\max} < \pi/2$ and the visible data space wavefronts $\beta \in [-\beta_{\max}, \beta_{\max}]$, $0 \leq \beta_{\max} < \pi/4$. Thus we only need to restrict the range of θ in equation ($\hat{\mathbf{A}}$) to formulate the limited-angle PAT Fourier forward operator

$$(\hat{\mathbf{A}}_\angle) \quad \hat{g}(\omega/c, \mathbf{k}_S) = \frac{\omega/c}{\sqrt{(\omega/c)^2 - |\mathbf{k}_S|^2}} \hat{p}_0 \left(\sqrt{(\omega/c)^2 - |\mathbf{k}_S|^2}, \mathbf{k}_S \right), \quad |\mathbf{k}_S| \leq \sin(\theta_{\max})|\mathbf{k}|.$$

($\hat{\mathbf{A}}_\angle$) maps from the domain $\mathcal{D}_{\hat{\mathbf{A}}}^\angle = \left\{ (\mathbf{k}_\perp, \mathbf{k}_S) \in \mathbb{R}^d : |\mathbf{k}_S| \leq \sin(\theta_{\max})|\mathbf{k}| \right\}$, $\sin \theta_{\max} < 1$ onto the narrowed bow-tie shaped range $\mathcal{R}_{\hat{\mathbf{A}}}^\angle = \left\{ (\omega/c, \mathbf{k}_S) \in \mathbb{R}^d : |\mathbf{k}_S| \leq |\omega/c| \tan \beta_{\max} \right\}$, $\tan \beta_{\max} < 1$, where again we extend the map by even symmetry for $\omega/c < 0$.

A Fourier domain illustration of the visibility in ambient space for $\theta_{\max} = \pi/3$ is shown in Figure 2 (a), and the corresponding narrowing down of the bow-tie shaped range in the Fourier transformed data space is highlighted with the red cut-off line in Figure 2 (b). With
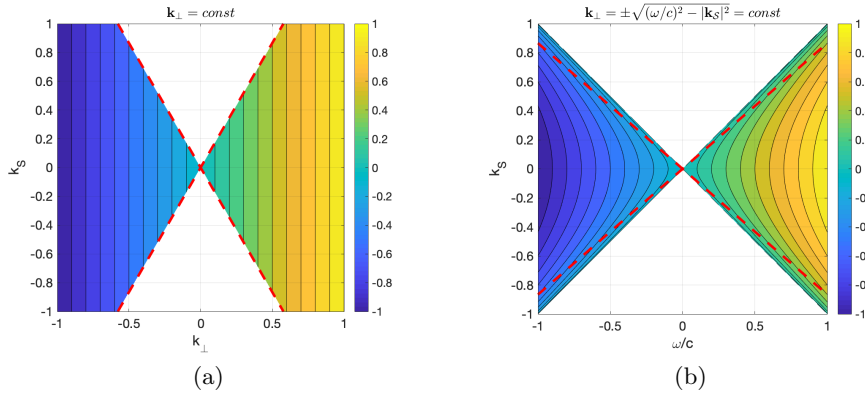

(a)

(b)

Figure 2: Contour plot of $\mathbf{k}_\perp = const$ over the bow-tie shaped (a) domain of the limited-angle PAT Fourier forward operator, $\theta_{\max} = \pi/3$, (b) range of the PAT Fourier forward operator with the further narrowed range of the limited-angle PAT Fourier forward operator highlighted with the red dashed line.

this restriction $|\theta| \leq \theta_{\max} < \pi/2$, the denominator of ($\hat{\mathbf{A}}_\angle$) is bounded away from 0,

$$(3.5) \qquad (\omega/c)^2 - |\mathbf{k_S}|^2 \geq (1 - \tan^2 \beta_{\max})(\omega/c)^2 = (\omega/c)^2 \cos^2 \theta_{\max} > 0,$$

and the limited-angle PAT Fourier forward operator can be stably evaluated. We note, that our restriction is equivalent to an upper bound on the factor (suggested as regularisation in [36])

$$(3.6) \qquad\qquad \frac{\omega/c}{\sqrt{(\omega/c)^2 - |\mathbf{k}_S|^2}} = \frac{|\mathbf{k}|}{|\mathbf{k}_\perp|} = \frac{1}{\cos \theta} \leq D,$$

with the choice of $D = (\cos\theta_{\max})^{-1}$.

We can now analogously define the limited-angle PAT Fourier inverse operator, $\mathcal{R}_{\mathbf{A}}^{\angle} \to \mathcal{D}_{\mathbf{A}}^{\angle}$

$$(\hat{\mathbf{A}}_{\angle}^{-1}) \qquad\qquad \hat{p}_0(\mathbf{k}_\perp, \mathbf{k}_S) = \frac{|\mathbf{k}_\perp|}{|\mathbf{k}|}\hat{g}\left(|\mathbf{k}|, \mathbf{k_S}\right), \quad |\mathbf{k}_S| \le \sin(\theta_{\max})|\mathbf{k}|,$$

and the limited-angle PAT Fourier adjoint operator, $\mathcal{R}_{\mathbf{A}}^{\angle} \to \mathcal{D}_{\mathbf{A}}^{\angle}$

$$(\hat{\mathbf{A}}_{\angle}^{*}) \qquad\qquad \hat{p}_0(\mathbf{k}_\perp, \mathbf{k}_S) = \hat{g}\left(|\mathbf{k}|, \mathbf{k_S}\right), \quad |\mathbf{k}_S| \le \sin(\theta_{\max})|\mathbf{k}|.$$

We note, that by Schwartz's Paley-Wiener theorem a Fourier transform of a compactly supported function is an entire function, thus in principle the function can be recovered from the knowledge of its Fourier transform on any open set, a.k.a. the continuous limited-angle operator is injective on compactly supported functions. However, a numerical realisation of such procedure would be highly unstable. In practice the functions p_0, g are discretized on a grid in $\mathbb{R}^d$ and a discrete fast Fourier transforms along with interpolation are used for evaluation of the operators. Then the discrete limited-angle forward operator has a non-trivial null space which corresponds to the invisible part of the image. In what follows, we use the same notation for the continuous and discrete operators, as their meaning is clear from the context.

The effect of the limited-angle PAT Fourier forward model in data domain is demonstrated in Figure 3 on the 4-disk phantom with line sensor on top. The disks are placed inside the north sector of the square domain so to eliminate any effect of the finite sensor. As the full data Fourier forward operator ($\hat{\mathbf{A}}$) is not well defined due to the blow up of the factor for the frequencies approaching the boundary of the open range (large bow-tie in Figure 2 (b)), we have to limit the factor which corresponds to limiting the sensitivity angle away from $\pi/2$ (albeit only by $\epsilon > 0$, very small). Thus the full angle is chosen as $\pi/2 - \epsilon$. The nearly full-angle data ($\theta_{\max} = \pi/2 - \epsilon$, $\epsilon > 0$, small) shown in (b) is dominated by slanted lines. These slanted lines align with the wavefront mapping of $\pi/2 - \epsilon$ and make an angle $\arctan(\sin(\pi/2-\epsilon)) \approx \pi/4$ with the sensor. We suppose that the limited, but still large factor, in the nearly full-data forward operator has the effect of adding these slanted singularities, akin to the singularities added by a sharp data cut-off in limited-view scenarios. Furthermore, the slanted lines wrap around along the horizontal dimension due to rectangular aspect ratio of the domain (the represented vertical frequencies are higher than the represented horizontal frequencies). For the limited-angle data ($\theta_{\max} = \pi/4$)[3] in (c) these artefacts are significantly reduced as the narrower bow-tie i) effects a stronger limit on the factor, ii) is effectively a high frequency cut-off resulting in a smoothing of the time-space domain data. Smoothing is a common strategy to reduce artefacts due to singularities added by the sharp cut-off of time-space domain data (e.g. due to the finite size sensor). This is consistent with microlocal analysis and was also proposed by Frikel and Quinto in [31] for limited sensor PAT. While there is an apparent similarity between (d) the limited-angle inversion $p_0^{\text{Linear}} = \hat{\mathbf{A}}_{\angle}^{-1}g_\angle$ and (e) the limited-angle adjoint $p_0^{\text{Adj}} = \hat{\mathbf{A}}_{\angle}^{*}g_\angle$, we point out the noticeable difference in contrast (s.a. (f) for the visualisation of their difference) due to the frequency dependent factor $|\mathbf{k}|/|\mathbf{k}_\perp|$ present in the inverse but not in the adjoint.

---

[3] In real data acquisition scenarios with Fabry Perròt sensor, $\theta_{\max}$ is empirically found to be $\approx \pi/4$.

A peculiarity of the half space geometry with planar/line sensor, is that each wavefront is measured exactly once i.e. we measure the wavefront travelling towards the detector and the wavefront in the opposite direction is accounted for by assuming the mirror symmetry of the data w.r.t. the detector hyperplane. Thus neither the limited sensitivity angle nor the limited sensor result in any partially measured singularities (all or nothing) and hence no improvement on linear reconstruction can be expected by a Neumann series type approaches.
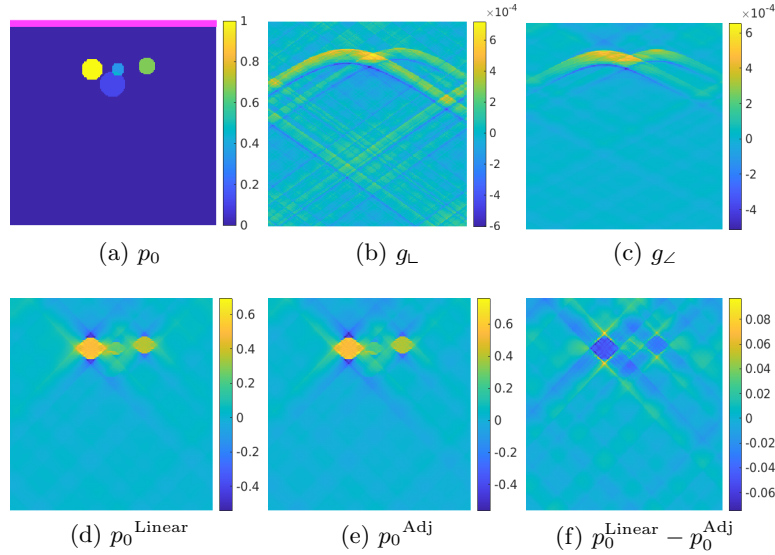

Figure 3: 4-disk Phantom - Limited-angle PAT Fourier operators: (a) p_0: 4 disks in the north sector of a square domain with a line detector on top (in pink), (b) full-angle data $g_\llcorner$ obtained via PAT Fourier forward operator $g_\llcorner = \hat{\mathbf{A}} p_0$ (a.k.a. $\theta_{\max} \approx \pi/2$) and (c) limited-angle data $g_\angle = \hat{\mathbf{A}}_\angle p_0$ with $\theta_{\max} = \pi/4$, (d) direct linear inversion of limited-angle data, $g_\angle$, $p_0^{\text{Linear}} = \hat{\mathbf{A}}_\angle^{-1} g_\angle$, (e) backprojection of $g_\angle$, $p_0^{\text{Adj}} = \hat{\mathbf{A}}_\angle^* g_\angle$ and (f) difference between the direct linear inversion and the backprojection of $g_\angle$, $p_0^{\text{Linear}} - p_0^{\text{Adj}}$.

4. Multiscale Representation of Photoacoustic Initial Pressure. In this section we discuss the multiscale representations used throughout the paper. We start with the standard Curvelet transform and introduce its restriction to the wedge of directions, fully wedge restricted Curvelet transform. The fully wedge restricted Curvelet transform can provide microlocal properties preserving representation of either the null space of the discrete limited-angle operator (invisible directions) or its complement (visible directions), depending on the choice of the wedge. Finally, we introduce a Coronae decomposition which is effectively the scale (frequency band) decomposition underpinning the Curvelet transform used in Section 5.2.2 to construct a U-Net with a multiscale structure matching this of the Curvelet frame.

4.1. Curvelets. The Curvelet transform [20] is a multiscale pyramid with many directions and positions at each scale. Figure 4 (a) shows the Curvelet induced tiling of the Fourier domain in 2D, with Curvelet window functions supported near a trapezoidal wedge with the

orientation θ_l, $\theta_l \in [-\pi, \pi)$ and the scale j. The corresponding Curvelet envelope function in spatial domain is aligned along a ridge of length $2^{-j/2}$ and width 2^{-j}. The wedges/envelopes become finer with increasing scales which lends the Curvelet frame the ability to resolve singularities along curves [23], [22], [20].

We introduce the Curvelet transform following the continuous presentation in [20]. For each scale j, orientation $\theta_l = 2\pi/L \cdot 2^{-\lfloor j/2 \rfloor} \cdot l, l \in \mathbb{Z}^{d-1}$ such that $\theta_l \in [-\pi, \pi)^{d-1}$, where $L \cdot 2^{\lfloor j_0/2 \rfloor}$ is the number of angles at the second coarsest scale $j = j_0 + 1$, j_0 even, the Curvelet coefficients of $u : \mathbb{R}^d \to \mathbb{R}$, are computed as

$$(\mathbf{C}) \qquad C_{j,l}(\mathbf{a}) = \int_{\mathbb{R}^d} \hat{u}(\mathbf{k}) \tilde{U}_{j,\theta_l}(\mathbf{k}) e^{i\mathbf{x}_\mathbf{a}^{(j,l)} \cdot \mathbf{k}} d\mathbf{k},$$

which is a projection of the Fourier transform $\hat{u}(\mathbf{k})$ on the corresponding frame element function $\tilde{U}_{j,\theta_l}(\mathbf{k}) \exp(-i\mathbf{k} \cdot \mathbf{x}_\mathbf{a}^{(j,l)})$ with $\tilde{U}_{j,\theta_l}(\mathbf{k})$ a trapezoidal frequency window with scale j and orientation θ_l, and a spatial domain centre at $\mathbf{x}_\mathbf{a}^{(j,l)} = (a_1 \cdot 2^{-j}, a_2 \cdot 2^{-j/2}, \ldots, a_d \cdot 2^{-j/2}) \in \mathbb{R}^d$ corresponding to the sequence of translation parameters $\mathbf{a} = (a_1, \ldots, a_d) \in \mathbb{Z}^d$. Note that the coarse scale Curvelets are isotropic Wavelets corresponding to translations of an isotropic low pass window $\tilde{U}_0(\mathbf{k})$. The window functions at all scales and orientations form a smooth partition of unity on the relevant part of $\mathbb{R}^d$. For real function u, we use the symmetric version of the Curvelet transform corresponding to the Hermitian symmetry of the wedges θ_l and $\theta_l + \pi$ in the frequency domain. There are two different ways to evaluate the integral $(\mathbf{C})$ efficiently. In this paper, we use the implementation via wrapping i.e. the digital coronisation with shears introduced in [20]. This fast digital Curvelet transform is a numerical isometry. We refer to [20] for further details and to Curvelab package[4] for the 2D and 3D implementations.

4.2. Fully Wedge Restricted Curvelets. The wedge restriction of the Curvelet transform was introduced in [46] and we refer to that paper and the accompanying codes[5] for the details of the general construction in $\mathbb{R}^d$. Here we restrict the presentation to the symmetric version in 2D as it is sufficient to explain the idea of full wedge restriction and its generalisation to $d > 2$ follows the same principles as the generalisation of the wedge restriction to higher dimensions.

The 2D *symmetric wedge restricted Curvelet transform* restricts the Curvelet orientations to the symmetric double wedge (bow-tie) $\theta_l \in W := [-\theta^w, \theta^w] \cup [-\theta^w + \pi, \theta^w + \pi]$ with $\theta^w \in (0, \pi/2)$ as illustrated in Figure 4 (b), where the gray region corresponds to Curvelets with orientations inside the bow-tie range.

The drawback of the original formulation of the wedge restricted Curvelet transform are the isotropic Wavelet window functions at the coarse scale inherited from the original Curvelet transform which result in non-directional coarse scale coefficients. To overcome this shortcoming, we enforce the split into in and out of wedge frequencies also at the coarse scale j_0 applying a binary bow-tie filter to the coarse scale isotropic Curvelet coefficients[6], and henceforth we

---

[4]http://www.curvelet.org/software.html

[5]https://github.com/BolinPan/Wedge_Restricted_Curvelet

[6]The isotropic Curvelet coefficients are projections on the translations of the smooth isotropic low pass window $\tilde{U}_0(\mathbf{k})$. Thus our construction results in a smoothed bow-tie restriction, however, all the frequency windows and higher scales are also smooth, thus smooth restriction is applied consistently to all scales.

refer to this construction as a *fully wedge restricted Curvelet transform*[7].

Formally we can write the fully wedge restricted Curvelet transform $\breve{\Psi} : \mathbb{R}^n \to \mathbb{R}^{\breve{N}}$, $\breve{N} < N$ as a composition of the standard Curvelet transform $\Psi : \mathbb{R}^n \to \mathbb{R}^N$ and a projection operator $P_W : \mathbb{R}^N \to \mathbb{R}^{\breve{N}}$, $\breve{\Psi} = P_W \Psi$

$$(4.1) \qquad P_W : \mathbb{R}^N \to \mathbb{R}^{\breve{N}}$$

$$P_W(C_{j,l}(\mathbf{a})) = \begin{cases} C_{j,l}(\mathbf{a}), & \theta_{j,l} \in W, \, j \geq j_0 \\ \mathbf{0}, & \text{otherwise}, \end{cases}$$

where the angles $\theta_{j,l}$ are the discrete wedge orientations (essentially corresponding to the wedge centres) and we extend this notation to the coarse scale j_0 as $\theta_{j_0,l} = \text{atan2}(\mathbf{x_a}^{(j_0,l)})$. We note the fully wedge restricted Curvelet transform is an isometry when defined on the restriction to the range of $\breve{\Psi}^\dagger$ i.e. $\breve{\Psi} : \mathbb{R}^n|_{\text{range}(\breve{\Psi}^\dagger)} \to \mathbb{R}^{\breve{N}}$, where range$(\breve{\Psi}^\dagger)$ corresponds to the set W in the Fourier domain. Figure 4 (b,c) illustrate in Fourier domain the difference between the ranges of the wedge restricted and fully wedge restricted 2D Curvelet transforms.
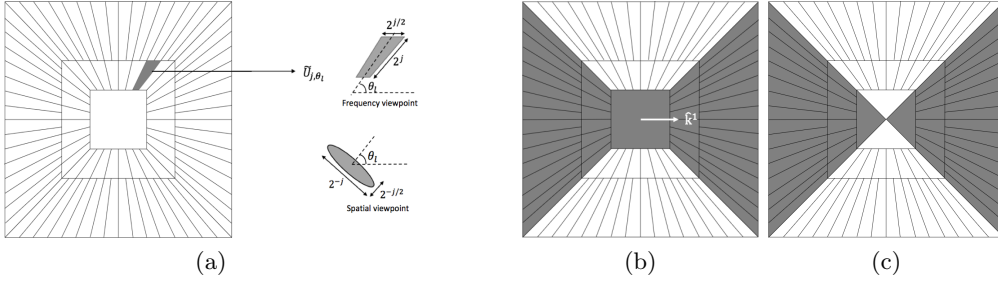

Figure 4: (a) An example of a 2D frequency domain Curvelet tiling using 3 scales. The highlighted wedge, magnified view in top right, corresponds to the frequency window $\tilde{U}_{j,\theta_l}$ at scale j with orientation θ_l near which a Curvelet is supported. The orientation of the envelope of the corresponding Curvelet function in the spatial domain is shown in bottom right corner. Induced (b) wedge restricted (c) fully wedge restricted Curvelet transforms for $\theta_{\max} = \pi/4$. The respective bow-tie shaped ranges are coloured gray.

4.3. Fully Wedge Restricted Curvelet Representation of Initial Pressure. Curvelets provide an almost optimally sparse representation if the function $u : \mathbb{R}^d \to \mathbb{R}$ is smooth away from (piecewise) $\mathcal{C}^2$ singularities [22, 55, 25]. More precisely, the error of the best s-term approximation, u_s, (corresponding to taking the s largest in magnitude coefficients) in Curvelet frame decays as [22]

$$(4.2) \qquad ||u - u_s||_2^2 \leq \mathcal{O}((\log s)^3 \cdot s^{-2}).$$

---

[7]The code is available from https://github.com/BolinPan/CoronaeNet

Furthermore, Curvelets have been shown to sparsely propagate through the wave equation [21] essentially being translated along the trajectories of the Hamiltonian flow with initial conditions corresponding to the spatial domain centre and the direction of the Curvelet. For constant speed of sound, this implies that the angle θ at which the Curvelet impinges on the detector is the same as the Curvelet orientation in the decomposition of the initial pressure p_0. A direct consequence of the restriction of the angle $|\theta| \leq \theta_{\max}$ is the restriction to the *visible* initial pressure p_0, which corresponds to symmetric fully wedge restricted Curvelet transform with directions $\theta_l \in W^\theta_{\max}$, $W^\theta_{\max} := [-\theta_{\max}, \theta_{\max}] \cup [-\theta_{\max} + \pi, \theta_{\max} + \pi]$ with $\theta_{\max} \in (0, \pi/2)$. On the other hand, the restriction of the impingement angle, $|\theta| \leq \theta_{\max}$, translates via (**WfM**) into the restriction of the angle β, $|\beta| \leq \arctan(\sin\theta_{\max})$, implying narrowing of the bow-tie shaped range of the PAT forward operator to $W^\beta_{\max} := [-\beta_{\max}, \beta_{\max}] \cup [-\beta_{\max} + \pi, \beta_{\max} + \pi]$ with $\beta_{\max} = \arctan(\sin\theta_{\max}) \in (0, \pi/4)$.

We now reinterpret (**C**) to obtain the symmetric fully wedge restricted Curvelet transform for representation of the *visible* initial pressure

$$(\mathbf{FWRC}) \quad \check{C}_{j,l}(\mathbf{a}_\perp, \mathbf{a}_\mathcal{S}) = \iint_{\mathbb{R}^d} \hat{u}(\mathbf{k}_\perp, \mathbf{k}_\mathcal{S}) \tilde{U}_{j,\theta_{j,l}}(\mathbf{k}_\perp, \mathbf{k}_\mathcal{S}) e^{i(\mathbf{x}_\mathbf{a} \cdot (\mathbf{k}_\perp, \mathbf{k}_\mathcal{S}))} d\mathbf{k}_\perp d\mathbf{k}_\mathcal{S}, \quad \theta_{j,l} \in W_{\max},$$

with the following notation:

$(\mathbf{k}_\perp, \mathbf{k}_\mathcal{S})$ ambient/image domain frequency $\mathbf{k} = (\mathbf{k}_\perp, \mathbf{k}_\mathcal{S})$;

$\theta_{j,l}$ a discrete direction of a trapezoidal frequency window $\tilde{U}_{j,\theta_{j,l}}(\mathbf{k}_\perp, \mathbf{k}_\mathcal{S})$. The tiling is computed using standard Curvelet transform on a cuboid domain, followed by the projection P_W on the bow-tie shaped set $W_{\max}$; We remark, that efficient implementation would bypass the computation of the wedges outside $W_{\max}$, but it would require modifications to the Curvelet Toolbox functions.

$W_{\max}$ the restriction of $\theta_{j,l}$ to those in the bow-tie $W_{\max} := [-\theta_{\max}, \theta_{\max}] \cup [-\theta_{\max} + \pi, \theta_{\max} + \pi]$ with $\theta_{\max} \in (0, \pi/2)$ effects the projection on the *visible* range of initial pressure in the Fourier domain;

$(\mathbf{a}_\perp, \mathbf{a}_\mathcal{S})$ the grid of spatial translations: $\mathbf{a}_\mathcal{S}$ parallel to the detector $\mathcal{S}$, $\mathbf{a}_\perp$ perpendicular to $\mathcal{S}$ (note that Curvelet transform via wrapping uses one grid per each quadrant at each scale, a.k.a. $\mathbf{x}_\mathbf{a}^{j,\mathbb{Q}} = (2^{-j}\mathbf{a}_\perp^\mathbb{Q}, 2^{-j/2}\mathbf{a}_\mathcal{S}^\mathbb{Q})$, $\mathbb{Q} = \{\mathbb{N}, \mathbb{W}, \mathbb{S}, \mathbb{E}\}$).

We note that in practice, the function $p_0 : \mathbb{R}^d \to \mathbb{R}$ is discretised on an n point grid in $\mathbb{R}^d$, yielding a vector $p_0 \in \mathbb{R}^n$, and we apply the symmetric discrete Curvelet transform $\Psi : \mathbb{R}^n \to \mathbb{R}^N$. The discrete counterparts of the wedge restricted $\tilde{\Psi}$ and fully wedge restricted $\check{\Psi}$ transforms follow analogously.

We compare the standard Curvelet transform Ψ, wedge restricted Curvelet transform $\tilde{\Psi}$ with fully wedge restricted Curvelet transform $\check{\Psi}$ for $\theta_{\max} = \pi/4$ on the 4-disk phantom (shown in Figure 3 (a)). We use 3 scales and 32 angles (at the 2nd coarsest level) for the standard Curvelet transform, yielding wedge restricted transforms with 16 angles (at the 2nd coarsest level). Both the wedge restricted Curvelet transform and the fully wedge restricted Curvelet transforms exclude the wedges out of the bow-tie range $W_{\max}$ at the higher $j > j_0$ scales; see Figure 5 (b,c). The standard Curvelet transform and the wedge restricted Curvelet transform have the same coarse scale coefficients while the fully wedge restricted Curvelet transform enforces the split of the coarse scale into visible $\theta_{j,l} \in W_{\max}$ and invisible $\theta_{j,l} \notin W_{\max}$ to obtain the visible initial pressure only; see Figure 5 (d-f).

The image domain effect of projection induced by all three Curvelet transforms are shown in Figure 6. Figure 6 (c,f) illustrates that the projection corresponding to the fully wedge restricted Curvelet transform corresponds to the visible initial pressure.

(a) $|\Psi p_0|$ (b) $|\tilde{\Psi} p_0|$ (c) $|\breve{\Psi} p_0|$

(d) $|\Psi p_0|_{\mathrm{C}}$ (e) $|\tilde{\Psi} p_0|_{\mathrm{C}}$ (f) $|\breve{\Psi} p_0|_{\mathrm{C}}$
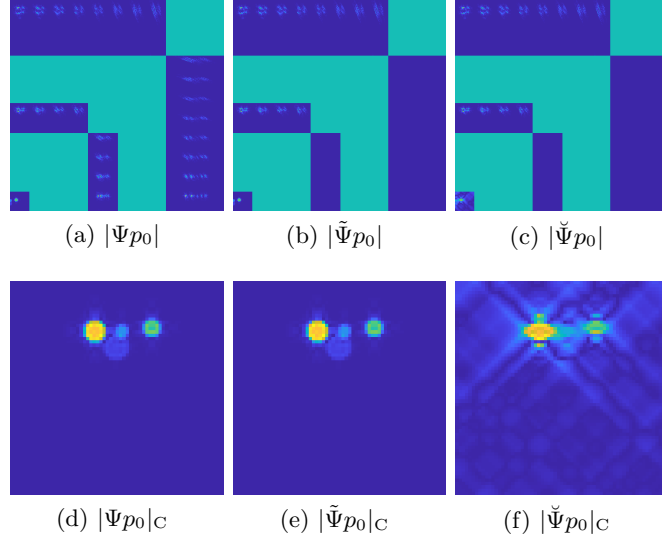
Figure 5: 4-disk Phantom - Curvelet coefficients of 4-disk phantom ($\theta_{\max} = \pi/4$ for wedge restriction): (a) standard, (b) wedge restricted, (c) fully wedge restricted Curvelet transform coefficient magnitudes (showing top-right quarter of scale-normalised coefficient magnitudes); (d-f) magnitudes of the corresponding coarse scale coefficients.

4.4. Coronae Decomposition and Reconstruction. The Coronae decomposition is effectively the scale (frequency band) decomposition underpinning the Curvelet transform. The Coronae decomposition is computed in the Fourier domain but the Coronae coefficients at each scale are images obtained with inverse scale restricted Fourier transform. We make use of the same smooth partition of unity filters as used in Curvelet Toolbox.

We henceforth denote the filters at the scale j with $\mathbb{L}_j : \mathbb{C}^{n_j \times m_j} \to \mathbb{C}^{n_{j-1} \times m_{j-1}}$ the low-pass filter followed by the restriction to the range $\mathbb{C}^{n_{j-1} \times m_{j-1}}$, with $\mathbb{L}_j^{\wedge} : \mathbb{C}^{n_j \times m_j} \to \mathbb{C}^{n_j \times m_j}$ the same filter but upsampled to $\mathbb{C}^{n_j \times m_j}$ via zero padding (equivalent to $\mathbb{L}_j$ without the range restriction) and with $\mathbb{H}_j : \mathbb{C}^{n_j \times m_j} \to \mathbb{C}^{n_j \times m_j}$, the high-pass filter which completes the partition of unity $\mathbb{H}_j = \sqrt{1 - (\mathbb{L}_j^{\wedge})^2}$. Here $n_j \times m_j$ is the size of the image $p_j \in \mathbb{R}^{n_j \times m_j}$ at the scale j

$$(4.3) \qquad n_j \times m_j = \left(\frac{n_J}{3 \cdot 2^{J-j-1}} + 2 \cdot \lfloor \frac{n_J}{3 \cdot 2^{J-j}} \rfloor + 1 \right) \times \left(\frac{m_J}{3 \cdot 2^{J-j-1}} + 2 \cdot \lfloor \frac{m_J}{3 \cdot 2^{J-j}} \rfloor + 1 \right),$$

for $j_0 \leq j < j_0 + J$ and $n_J \times m_J$ is the size of the original image, p_{j_0+J}.

The Coronae decomposition is obtained by the application of the filter pair $(\mathbb{L}_j, \mathbb{H}_j)$: $\mathbb{H}_j(p_j)$ returns the high-pass component with the same size as p_j and $\mathbb{L}_j(p_j)$ the low-pass

(a) $\Psi^\dagger \Psi p_0$ (b) $\tilde{\Psi}^\dagger \tilde{\Psi} p_0$ (c) $\breve{\Psi}^\dagger \breve{\Psi} p_0$

(d) $\Psi^\dagger \Psi p_0 - p_0$ (e) $\tilde{\Psi}^\dagger \tilde{\Psi} p_0 - p_0$ (f) $\breve{\Psi}^\dagger \breve{\Psi} p_0 - p_0$
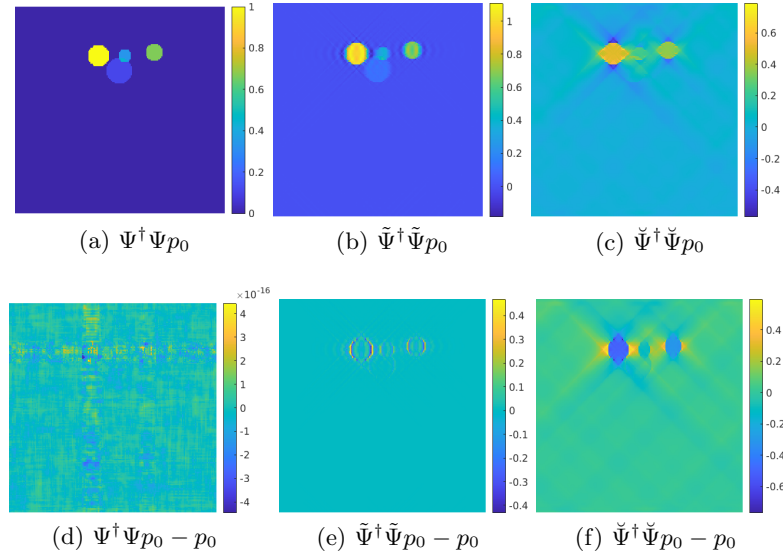
Figure 6: 4-disk Phantom - Visualisation of the image domain projections effected by the various Curvelet transforms ($\theta_{\max} = \pi/4$ for wedge restriction): (a) standard Ψ, (b) wedge restricted $\tilde{\Psi}$, (c) fully wedge restricted $\breve{\Psi}$; (d-f) projections on the corresponding orthogonal complements in $\mathbb{R}^d$.

component with the size of p_{j-1}; see Figure 7 for visualisation of the effect of 1-level of Coronae decomposition of the 4-disk phantom.

(a) p_0 (b) $\mathbb{H}_{\mathrm{F}}(p_0)$ (c) $\mathbb{L}_{\mathrm{F}}(p_0)$
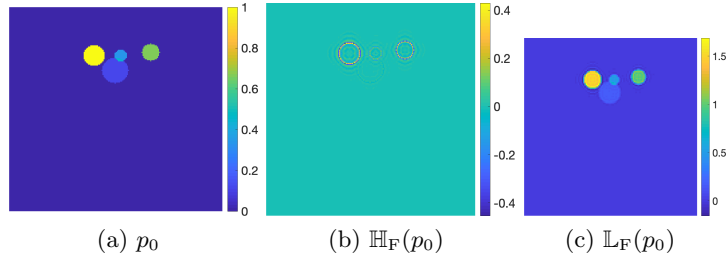
Figure 7: 1-level Coronae decomposition of 4-disk phantom: (a) p_0, (b) high-pass component $\mathbb{H}_{\mathrm{F}}(p_0)$, and (c) low-pass component $\mathbb{L}_{\mathrm{F}}(p_0)$.

The signal can be recovered by the Coronae reconstruction. The Coronae reconstruction upsamples the low-pass component via zero-padding and sums up the padded low-pass component with the corresponding high-pass component in the Fourier domain to recover the original signal (denoted with $\oplus$).

Figure 8 (left) shows the schematic of the Fourier domain computations involved in the Coronae decomposition (top row) and reconstruction (bottom row). For visualisation pur-

poses the overlapping smooth partition of unity filters[8] were replaced with discontinuous non-overlapping partition of unity by indicator functions of the low- and high-band.

Based on the 1-level Coronae decomposition obtained with the pair $(\mathbb{L}_j, \mathbb{H}_j)$ we can form a filter bank recursion for a multi-level Coronae decomposition

$$(4.4) \qquad ((\mathbb{L}_{j-1}(\mathbb{L}_j), \mathbb{H}_{j-1}(\mathbb{L}_j)), \mathbb{H}_j), \quad j = j_0 + J, ..., j_0 + 1,$$

and reconstruction

$$(4.5) \qquad (\mathbb{L}_j \leftarrow \mathbb{L}_{j-1} \oplus \mathbb{H}_{j-1}) \oplus \mathbb{H}_j, \quad j = j_0 + 1, ..., j_0 + J,$$

Figure 8 (right) displays a 3 scale filter bank where the red and green arrows indicate the Coronae decompositions and reconstructions, respectively.
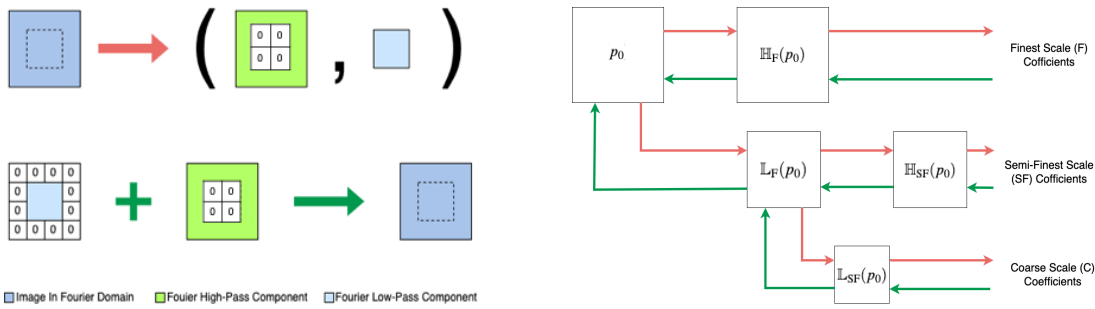

Figure 8: Left: 1-level Coronae decomposition in 2D Fourier domain. Coronae decomposition (top), and reconstruction using zero padding (bottom). Note, that the interior of the green square in Fourier domain is populated with zeros. Right: a filter bank with 3 scales effecting 2-level Coronae decomposition ($\rightarrow$) and reconstruction ($\leftarrow$).

4.5. Coronae Decomposition vs Curvelet Decomposition. The relation of the Coronae and Curvelet decompositions can be formally stated in the Fourier domain as

$$(\mathbf{QC}) \qquad \mathbb{L}_j \hat{Q}_j(\mathbf{k}_j) = \sum_l \sum_{\mathbf{a}_{j,l}} \tilde{U}_{j,\theta_l}(\mathbf{k}_j) C_{j,l}(\mathbf{a}_{j,l}) e^{-i\mathbf{x}_{\mathbf{a}}^{(j,l)} \cdot \mathbf{k}_j},$$

where $\mathbf{k}_j$ is the frequency domain vector restricted to the scale j, $j \geq j_0$[9] and $\mathbf{x}_{\mathbf{a}}^{(j,l)}$ are the Curvelet centers at scale j and angle $\theta_{j,l}$. In particular, when using implementation via wrapping we have one grid per quadrant $\mathbf{x}_{\mathbf{a}}^{(j,l)} = \mathbf{x}_{\mathbf{a}}^{(j,\mathbb{Q}(l))}$.

Applying the Coronae transform to the visible / invisible parts of the image (with the maximal sensitivity angle $\theta_{\max}$) corresponds to restricting the sum in (**QC**) to the visible

---

[8]Such filters require appropriate wrapping of the low pass component $\mathbb{L}_j(p_j)$ before Fourier inversion at scale $j - 1$.

[9]Note that the overlapping smooth partition of unity filters yield larger $\mathbf{k}_j$ than the 0-1 non-overlapping filters.

wedge $l : \theta_{j,l} \in W_{\max}$ or to its complement $l : \theta_{j,l} \notin W_{\max}$ for invisible coefficients. This is equivalent to the fully wedge restricted Curvelet transform with the same wedge $W_{\max}$ for visible coefficients and again its complement for invisible coefficients.

The computation of Coronae coefficients of the visible/invisible bypasses the computation of the Curvelet transform. After Fourier domain application of the bow-tie shaped filter to extract the visible $\mathrm{atan2}(\mathbf{k}) \in W_{\max}$ and its complement the invisible $\mathrm{atan2}(\mathbf{k}) \notin W_{\max}$ parts of the image, we use the filters bank defined in Subsection 4.4 to obtain the Coronae decomposition Q of each visible/invisible component directly. Figure 9 (g-i) shows 2-level Coronae decomposition coefficients of the visible part of the 4-disk phantom.
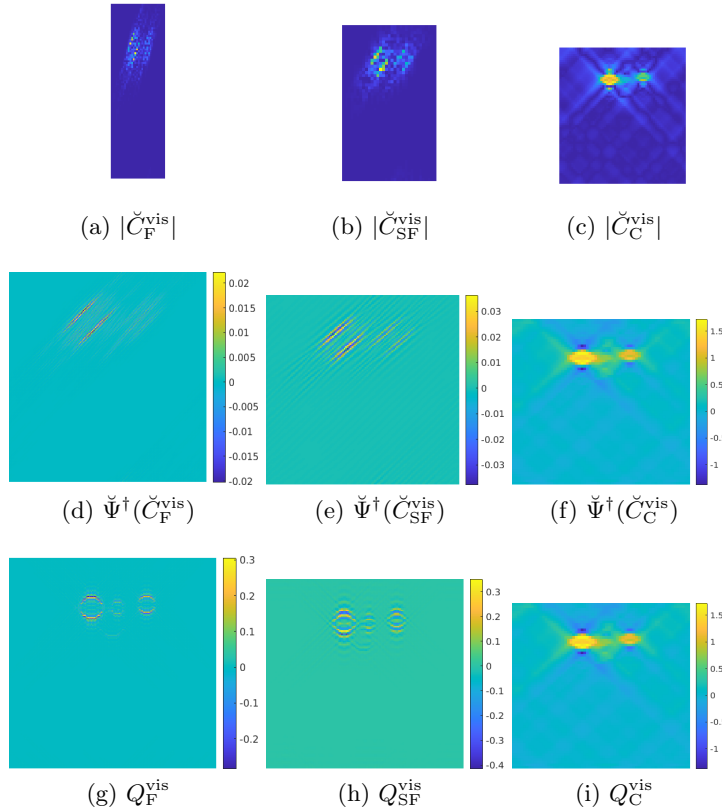

(a) $|\check{C}_{\mathrm{F}}^{\mathrm{vis}}|$ (b) $|\check{C}_{\mathrm{SF}}^{\mathrm{vis}}|$ (c) $|\check{C}_{\mathrm{C}}^{\mathrm{vis}}|$

(d) $\check{\Psi}^{\dagger}(\check{C}_{\mathrm{F}}^{\mathrm{vis}})$ (e) $\check{\Psi}^{\dagger}(\check{C}_{\mathrm{SF}}^{\mathrm{vis}})$ (f) $\check{\Psi}^{\dagger}(\check{C}_{\mathrm{C}}^{\mathrm{vis}})$

(g) $Q_{\mathrm{F}}^{\mathrm{vis}}$ (h) $Q_{\mathrm{SF}}^{\mathrm{vis}}$ (i) $Q_{\mathrm{C}}^{\mathrm{vis}}$

Figure 9: 4-disk Phantom - Representations of the visible ($\theta_{\max} = \pi/4$) with 3 scales {C, SF, F}: (a-b) amplitudes of fully wedge restricted Curvelet coefficients in one visible trapezoidal window at scales F, SF (corresponding blue tiles in the Curvelet coefficient display) and (c) amplitudes of all visible coarse scale C coefficients; (d-f) image domain representation of the coefficients in (a-c); (g-i) visible Coronae coefficients at all scales.

5. Reconstructing the Visible and Learning the Invisible. The fully wedge restricted Curvelet transform introduced in Subsections 4.2 and 4.3, allows us to decompose the initial pressure $p_0 \in \mathbb{R}^n$ into Curvelet coefficients that belong to either the visible $f^{\mathrm{vis}} \in \mathbb{R}^{\check{N}}$ or the

invisible $f^{\text{inv}} \in \mathbb{R}^{\check{N}}$ initial pressure (in the sense that the other coefficients are set to 0)

$$\textbf{(VID)} \qquad \begin{aligned} f^{\text{vis}} &= \check{\Psi} p_0, \\ f^{\text{inv}} &= \check{\Psi}^{\perp} p_0. \end{aligned}$$

Given the limited-angle data $g_\angle \in \mathbb{R}^m$ contaminated with additive white noise, the *visible* initial pressure p_0 can be reconstructed by reconstructing the *visible* coefficients of p_0 in the fully wedge restricted Curvelet frame. The latter is a standard compressed sensing $\ell_2 - \ell_1$ minimisation problem with the ℓ_2 data fidelity term involving the aforementioned limited-angle PAT Fourier forward operator. As the invisible coefficients are in the null space of the discrete forward operator i.e. $f^{\text{inv}} \in \textbf{ker}(\hat{\textbf{A}}_\angle)$ [28], f^{inv} can not be reconstructed from $g_\angle$. Instead, given a representative training set, we can fill in the invisible coefficients using a trained CNN. We refer to such two step method as *reconstructing the visible and learning the invisible.*

5.1. Reconstructing the Visible. In this step, we recover the visible part of initial pressure $p_0 \in \mathbb{R}^n$ from noisy measurement $g_\angle \in \mathbb{R}^m$ in a fully wedge restricted Curvelet frame. To this end we solve

$$\textbf{(VR)} \qquad \tilde{f}^{\text{vis}} \in \arg\min_f \frac{1}{2}||\hat{\textbf{A}}_\angle \check{\Psi}^\dagger f - \hat{g}_\angle||_2^2 + \tau ||\Lambda f||_1,$$

where $\hat{\textbf{A}}_\angle : \mathcal{D}_{\textbf{A}}^\angle \to \mathcal{R}_{\textbf{A}}^\angle$ is the discrete limited-angle PAT Fourier forward operator, $\check{\Psi}^\dagger : \mathbb{R}^{\check{N}} \to \mathbb{R}^n$ is the left inverse (and an adjoint) of the fully wedge restricted Curvelet transform $\check{\Psi}$[10], f are the coefficients of p_0 in $\check{\Psi}$, $\tilde{f}^{\text{vis}}$ are the reconstructed visible Curvelet coefficients, and the regularisation parameter τ is made scale dependent by multiplication with $\Lambda = \text{diag}(2^{\textbf{j}-2})$, where $\textbf{j} \geq j_0 \textbf{1}$ is a vector containing the scales of the frame elements $\check{\Psi}$. We propose to solve (VR) using **FISTA** [12] with a pre-computed Lipschitz constant $L = ||\check{\Psi}\hat{\textbf{A}}_\angle^\dagger \hat{\textbf{A}}_\angle \check{\Psi}^\dagger||_2$.

5.2. Learning the Invisible. Armed with the visible reconstruction f^{vis}, we now learn the invisible coefficients by a tailored U-Net. As the invisible coefficients are in the null space of the discrete forward operator, there is no need/benefit to include the forward and adjoint operators into the network. Our network is based on a U-Net architecture but it works on *Coronae coefficients* and uses the matching *Coronae Decomposition* in lieu of downsampling and the *Coronae Reconstruction* in lieu of upsampling. Hence, we term it *Coronae-Net* (CorNet). The details of CorNet architecture are depicted in Figure 10. The inputs of CorNet are visible Coronae coefficients which are generated from the reconstructed visible fully wedge restricted Curvelet coefficients. The visible Coronae coefficients are mapped to a latent representation via a series of convolutions, non-linearities and high/low-pass filters on the down branch of the CorNet. The correlations between the visible and invisible parts of the image are encoded in this latent representation during training and allow prediction of the invisible part of the image. The decoder branch of the CorNet, decodes and upsamples both visible and invisible Coronae coefficients to the highest scale output.

---

[10]As already mentioned in Section 3 all the computations can be executed directly in the Fourier domain. For simplicity we use the same notation for transforms (Ψ, $\tilde{\Psi}$, $\check{\Psi}$), which act on the Fourier transform $\hat{p}_0$ as for those that act on p_0 itself.

5.2.1. Rationale for Using Coronae Coefficients as Network Input. There are two disadvantages to usage of Curvelet coefficients as a network input. Firstly, to preserve the locality of the CNN convolutions (with kernels with localized support, frequently 2 or 3 pixels in each dimension), they need to be applied in a domain where local convolutions are meaningful. A vanilla realisation of a Curvelet based input could entail computing an image domain representation of each individual Curvelet $\breve{\Psi}^\dagger(C_{j,l}(\mathbf{a}_{j,l}))$ which, even when using scale appropriate image sizes, would result in a prohibitive number of image inputs. A better alternative is to use all coefficients corresponding to one trapezoidal wedge window (the blue tiles of the Curvelet coefficient plot in Figure 5 (a-c), visualised for one window per scale in Figure 9 (a-c)), or their image domain representations in Figure 9 (d-f). However, these are still numerous due to, parabolic scaling induced, angles doubling every second scale. Furthermore, the size of the trapezoidal window and hence the corresponding coefficients varies even within the scale which would require additional processing e.g. zero-padding when working with the coefficients (though not needed when working with their image domain representations Figure 9 (d-f)). Finally, it is not clear that training CNNs with small-support kernels would yield good results when working with rather non-smooth and not highly localised (at least not respective to their size) coefficients per window inputs as those in Figure 9 (a-c).

We are however, interested in the split into visible and invisible parts rather than in individual (or per window) Curvelet coefficients. This is most economically represented by the Coronae decomposition applied to the visible/invisible parts of the image (see Figure 9 (g-i)) which results in a two channel scale-wise input (visible and invisible channel per scale). Moreover, it is apparent from Figure 9 that the Coronae decomposition is also parsimonious at a higher level feature representation, as it is not reliant on a superposition for final representation unlike Curvelet decomposition. Thus we can reasonably expect it will be easier to learn.

5.2.2. Coronae-Net. We propose the following CorNet which we construct with a depth matching the number of scales of the Curvelet decomposition. The input at each scale j consists of 2 channels: the first channel is the Coronae coefficient $Q_j(p^{\mathrm{vis}})$ of the visible and the second channel will be used to predict the unknown Coronae coefficient $Q_j(p^{\mathrm{inv}})$ of the invisible[11]. At the input stage the latter could be initialised with $\mathbf{0}$ but in practice adding random white noise in training is preferable. In the encoder branch, scale by scale the convolutions and nonlinearities extract and correlate the features of the visible and invisible Coronae coefficients to effect the prediction of the invisible Coronae coefficients. The Coronae decomposition is used to pass the low-pass components downward and high-pass components directly to the decoder via the skip connection. On the up-branch the visible and predicted invisible features are decoded via a symmetric cascade of convolutions and nonlinearities. The Coronae reconstruction is used to assemble the high-pass components passed through the skip connection with the low-pass components obtained from the deeper layer. We highlight the asymmetry in the CorNet input and output, while we input the Coronae coefficients scale-wise, the output Coronae coefficients are upsampled (via 0-padding) and returned at the finest

---

[11]The invisible input is not strictly necessary, however it allows to explicitly inform the network about the structure of the invisible Coronae coefficients, which we deem beneficial due to minimal representation of the visible/invisible and perfect matching of the network scales by the Coronae decomposition.

scale. Furthermore, the output is multi-headed, as popular in multi-task learning framework, which allows us to assign different weights to coefficients at different scales (and possibly even to visible and invisible channels).

Although the output contains both visible and invisible, the loss function is only applied to the invisible Coronae coefficients and can be written as

$$\textbf{(PL)} \qquad \mathcal{L}(\mathcal{NN}_{\mathrm{Cor}}(Q^{\mathrm{P\,vis}}), \tilde{Q}^{\mathrm{P\,inv}}) = \sum_{j=j_0}^{j_0+J} \left(\alpha_j \cdot \mathrm{MSE}\left(\mathcal{NN}_{\mathrm{Cor}}(Q_j^{\mathrm{P\,vis}}), \tilde{Q}_j^{\mathrm{P\,inv}} \right) \right),$$

where the perfect visible Coronae coefficients (original size) $Q_j^{\mathrm{P\,vis}}$ are the input to the Cor-Net, $\mathcal{NN}_{\mathrm{Cor}}$, $\tilde{Q}_j^{\mathrm{P\,inv}}$ are the upsampled to the finest-scale perfect/reference invisible Coronae coefficients, and α_j are some fixed scale dependent weights. The total loss (**PL**) is summed over the training set of in/output pairs of $Q^{\mathrm{P\,vis}}$ and $\tilde{Q}^{\mathrm{P\,inv}}$.
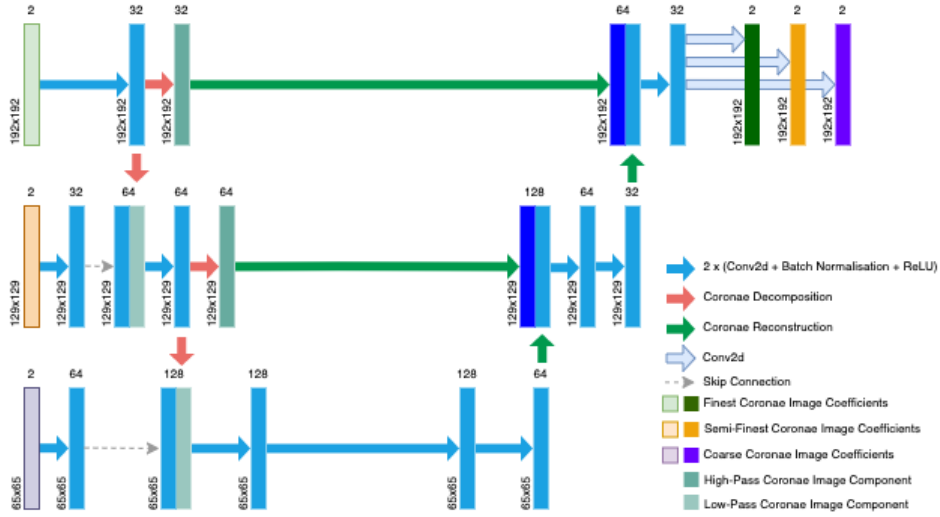

Figure 10: Coronae-Net - Architecture here shown with 3 scales for learning the invisible Coronae coefficients in 2D. The input is scale-wise and has 2 channels per scale: the first channel is the visible Coronae coefficient at the corresponding scale and the second channel will be used to predict the unknown invisible Coronae coefficient at the corresponding scale. The Coronae decomposition and reconstruction are used as downsampling and upsampling in the network, respectively.

Recall the invisible Curvelet coefficients are in the null space of the discrete limited-angle PAT Fourier forward operator. The same holds for invisible Coronae coefficients, c.f. Subsection 4.5, thus learning the invisible with CorNet from perfect visible coefficients also follows the null space learning paradigm as the scheme presented in [49], but our chosen representation is essentially optimal for both the representation of the visible/invisible parts of the image and preservation of the microlocal properties of the PAT forward operator.

5.2.3. ResCoronae-Net. The main limitation of the perfect/null space learning is the assumption of availability of perfect visible coefficients. In realistic scenarios the reconstructed visible Coronae coefficients Q^{vis} will suffer from artefacts introduced by the reconstruction procedure. In our case, the Fourier operators contain an error due to interpolation between image and data domain grids. Therefore, we propose a residual version of the Coronae-Net, *ResCoronae-Net* (ResCorNet), to learn an update to both visible and invisible Coronae coefficients. We use the same **0** or a random noise initialisation for the invisible coefficients as for CorNet, thus the residual connection has no effect for invisible coefficients. We note that we need to upsample our inputs to the finest scale to match the output size before the residual connection (the diagram of ResCoronae-Net can be found in supplementary material Figure SM1). The loss function used for the residual learning penalises errors in both visible and invisible Coronae coefficients

$$(\textbf{IPL}) \qquad \mathcal{L}(\mathcal{NN}_{\mathrm{ResCor}}(Q^{\mathrm{I_{vis}}}), \tilde{Q}^{\mathrm{P_{all}}}) = \sum_{j=j_0}^{j_0+J} \left(\alpha_j \cdot \mathrm{MSE}\left(\mathcal{NN}_{\mathrm{ResCor}}(Q_j^{\mathrm{I_{vis}}}), \tilde{Q}_j^{\mathrm{P_{all}}} \right) \right),$$

where the reconstructed imperfect visible Coronae coefficients (original size) $Q_j^{\mathrm{I_{vis}}}$ are the input to the ResCorNet, $\mathcal{NN}_{\mathrm{ResCor}}$, $\tilde{Q}_j^{\mathrm{P_{all}}}$ are the upsampled to the finest-scale perfect/reference visible and invisible Coronae coefficients, and α_j some fixed scale dependent weights. Again, the total loss (**IPL**) is summed over the training set of in/output pairs of $Q^{\mathrm{I_{vis}}}$ and $\tilde{Q}^{\mathrm{P_{all}}}$.

6. Synthetic Data. In this section, we evaluate the performance of our proposed reconstruction scheme and compare it with various other reconstruction methods. We focus on two learning scenarios: 1) *Perfect Learning*, learns the invisible Coronae coefficients from the given perfect visible Coronae coefficients via CorNet with loss function (**PL**); 2) *Imperfect Learning*, learns the update of the visible and the invisible Coronae coefficients, from the imperfect visible Coronae coefficients, constructed from the solution of (**VR**), via ResCorNet with loss function (**IPL**).

6.1. Experimental Setup. We evaluate our proposed method on an *ellipse data set*. The data consists of 3000 synthetic images of superposition of 15 to 20 ellipses with randomly drawn location, orientation and contrast. The images are of resolution $n_{\mathbf{x}_\perp} \times n_S = 192 \times 192$ assuming a uniform voxel size $h_{\mathbf{x}} = 10\mu\mathrm{m}$. The locations of the ellipses are randomly chosen within the upper half of the image to reduce (while not eliminate completely) the limited sensor effect. The individual images are normalised between 0 and 1 (see Figure 11 (a) for a representative image sample from the ellipse data set).

We assume homogeneous speed of sound $c = 1500\mathrm{m/s}$. The pressure time series is recorded every $h_t = 66.67\mathrm{ns}$ by a line sensor with sensitivity angle $\theta_{\max} = \pi/4$ on top of the domain with image matching resolution i.e. $h_{\mathbf{x}} = 10\mu\mathrm{m}$, which we can interpret as number of time-steps × number of sensors shaped PAT data volume, 272×192.

We use the fully wedge restricted Curvelet transform with 3 scales and 16 angles in the wedge $W_{\max}$ (at tthe che 2nd coarse scale) for the sparse representation of initial pressure p_0. The parameters of the transform determine the structure of the networks such as depth (equal to the number of scales, here 3), and the input/output sizes.

We trained all the networks presented in this section in the same manner to make a fair comparison. 2400 images are used for training, 300 images for validation and 300 for testing. We chose Adam as the optimizer with Xavier initialisation [32] and batch size of 2. We trained all networks for 200 epochs with 20 epochs early stopping to avoid over-fitting. The initial learning rate is set to $1 \cdot 10^{-3}$ with a cosine decay. Each network is trained in PyTorch on a single Tesla K40c GPU.

6.2. Perfect Learning. We start with an idealised experiment where we learn the invisible Coronae coefficients from the perfect visible Coronae coefficients. The perfect visible/invisible image components and their Coronae coefficients are computed directly via projection e.g. $p_0^{\mathrm{P_{vis}}} = \breve{\Psi}^\dagger \breve{\Psi} p_0$, $p_0^{\mathrm{P_{inv}}} = p_0 - \breve{\Psi}^\dagger \breve{\Psi} p_0$ followed by the Coronae decomposition of $p_0^{\mathrm{P_{vis}}}, p_0^{\mathrm{P_{inv}}}$ for each image p_0 in the training set.

The variation in magnitudes and sizes of Coronae coefficients across scales, suggests use of larger weights for higher scales. The choice of $\alpha_{\mathrm{F}} = 2$, $\alpha_{\mathrm{SF}} = 2$ and $\alpha_{\mathrm{C}} = 1$ in (**PL**) led to the best network performance. We quantitatively compare the results of CorNet with standard variational approaches and U-Net based learned post-processing on the ellipse test data, in terms of MSE, PSNR and SSIM averaged over the test set.

We compare our results to two reference learned post-processing methods, based on U-Net architecture, trained to remove limited-angle artefacts. $\mathcal{NN}_{\mathrm{U\text{-}Net}}(p_0^{\mathrm{P_{vis}}})$ takes as an input the perfect visible component $p_0^{\mathrm{P_{vis}}}$ and is trained on $(p_0^{\mathrm{P_{vis}}}, p_0)$ pairs with MSE loss. $\mathcal{NN}_{\mathrm{U\text{-}Net}}(\tilde{Q}_j^{\mathrm{P_{vis}}})$ takes as an input the upsampled to the finest scale (via zero-padding in the Fourier domain) perfect visible Coronae coefficients $\tilde{Q}_j^{\mathrm{P_{vis}}}$ (a.k.a. multichannel input at highest scale only, one channel per scale) and is trained on $(\tilde{Q}_j^{\mathrm{P_{vis}}}, \tilde{Q}_j^{\mathrm{P_{all}}})$ pairs with MSE loss. The diagrams of the reference U-Nets can be found in supplementary material.

Table 1: Ellipses - Map from perfect visible to invisible. Imaging metrics averaged over the test set.

	MSE	PSNR	SSIM
$p_0^{\mathrm{P_{vis}}}$	$5.8311 \cdot 10^{-3} \pm 2.0110 \cdot 10^{-3}$	22.6029 ± 1.5361	0.3612 ± 0.0547
$p_0^{\mathrm{P_{inv}}}$	$1.5827 \cdot 10^{-2} \pm 5.0783 \cdot 10^{-3}$	18.2206 ± 1.3726	0.2729 ± 0.0574
$\mathcal{NN}_{\mathrm{U\text{-}Net}}(p_0^{\mathrm{P_{vis}}})$	$1.2207 \cdot 10^{-4} \pm 5.4004 \cdot 10^{-5}$	39.5088 ± 1.7885	0.9889 ± 0.0044
$\mathcal{NN}_{\mathrm{U\text{-}Net}}(\tilde{Q}_j^{\mathrm{P_{vis}}})$	$5.1026 \cdot 10^{-5} \pm 2.4257 \cdot 10^{-5}$	43.3695 ± 1.9730	0.9907 ± 0.0038
$\mathcal{NN}_{\mathrm{Cor}}(Q_j^{\mathrm{P_{vis}}})$	$\mathbf{3.4599 \cdot 10^{-5} \pm 1.4536 \cdot 10^{-5}}$	$\mathbf{44.9678 \pm 1.7715}$	$\mathbf{0.9919 \pm 0.0030}$

Figure 11 visualises the reconstruction quality for one of the test images. The first row of Figure 11 displays the initial pressure p_0 along with its decomposition into perfect visible $p_0^{\mathrm{P_{vis}}}$ and perfect invisible $p_0^{\mathrm{P_{inv}}}$ parts. The second row of Figure 11 illustrates the learned reconstructions (for plotting only, we restrict the pixel values to [0,1], with values out of this range mapped to the nearest interval end). Although all three methods perform very well, the zoomed-in insets in Figure 11 reveal our proposed scheme yields superior reconstruction in terms of invisible boundary (green inset), ellipse shapes and faint contrasts (red inset). On average, $\mathcal{NN}_{\mathrm{Cor}}(Q_j^{\mathrm{P_{vis}}})$ achieves lowest MSE, highest PSNR and highest SSIM in Table 1
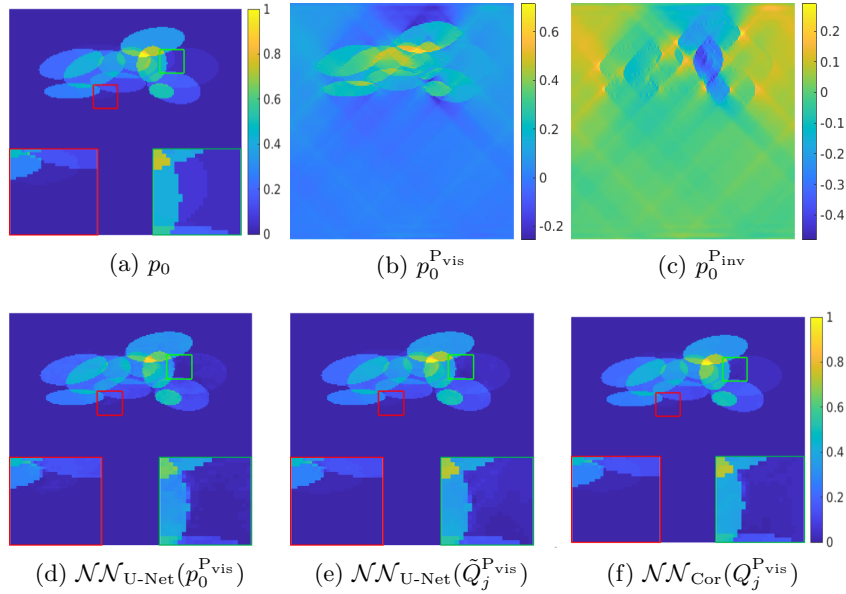
Figure 11: Ellipses - Visualisation of performance of perfect learning on one test image ($\theta_{\max} = \pi/4$): (a) initial pressure p_0, (b) perfect visible component of p_0, (c) perfect invisible component of p_0, (d-f) different learned reconstructions.

which is consistent with better, in eye ball metric, reconstructions produced by the CorNet[12]. However, we observe all methods in (d-f) fail to recover the faint ellipse on top of even fainter large ellipse and overlapping with ellipses with higher contrasts (highlighted in green inset). The low contrast boundary of the faint ellipse is difficult to pick up for the networks. We encountered a similar scenario in Figure 7, where the boundary of the faint blue disc is almost invisible in the high-pass component $\mathbb{H}(p_0)$ while the boundaries of the ellipses with higher contrast are still perceptible.

6.3. Imperfect Learning. Next, we consider the more practical scenario of correcting the visible and learning the invisible Coronae coefficients on the ellipse data set from limited-angle PAT data with additive white noise with standard deviation $\sigma = 2.5 \cdot 10^{-4}$. We reconstruct the visible Curvelet coefficients $\tilde{f}^{\mathrm{vis}}$ solving (**VR**) for regularisation parameter $\tau = 2.5 \cdot 10^{-4}$ performing 50 FISTA iterations. From the point of view of the noise level this corresponds to an early stopping, however the convergence of FISTA has already slowed down from the theoretical rate $\mathcal{O}(1/k^2)$ due to interpolation errors in the forward operator, thus 50 FISTA iterations constitute a reasonable computing time / accuracy trade-off and provide a good starting point for learning the correction. This scenario employs ResCorNet with the scale dependent loss weights $\alpha_{\mathrm{F}} = 2$, $\alpha_{\mathrm{SF}} = 2$ and $\alpha_{\mathrm{C}} = 1$ in (**IPL**). Again, we compare the results obtained with ResCorNet against and residual variants of U-Net learned post-processing methods on the test subset of the ellipse data.

---

[12]The invisible Coronae coefficients $\tilde{Q}_j^{\mathrm{L_{inv}}}$ learned using CorNet can be found in supplementary material.

We compare our results to five reference reconstruction methods applied to noisy limited-angle data $g_\angle$. $p_0^{\text{Linear}} = \hat{\mathbf{A}}_\angle^{-1} g_\angle$: linear inversion. $p_{0_{\ell_1}}^{\text{I}_{\text{vis}}}$: recovery of the sparse representation of the visible part of p_0 in the fully wedge restricted Curvelet frame via **(VR)**. $p_{0_{\text{TV}+}}^{\text{I}_{\text{vis}}}$: total variation regularised solution with non-negativity constraint [8]. And two learned post-processing methods based on ResU-Net architecture trained to remove artefacts due to limited sensitivity angle and errors in the visible part of the image/Coronae coefficients. $\mathcal{NN}_{\text{ResU-Net}}(p_0^{\text{I}_{\text{vis}}})$: takes as an input the imperfect visible coefficients $p_0^{\text{I}_{\text{vis}}}$ and is trained on $(p_0^{\text{I}_{\text{vis}}}, p_0)$ pairs with MSE loss. $\mathcal{NN}_{\text{ResU-Net}}(\tilde{Q}_j^{\text{I}_{\text{vis}}})$: takes as an input the (upsampled to the finest scale via zero-padding in the Fourier domain) imperfect visible Coronae coefficients $\tilde{Q}_j^{\text{I}_{\text{vis}}}$ and is trained on $(\tilde{Q}_j^{\text{I}_{\text{vis}}}, \tilde{Q}_j^{\text{P}_{\text{all}}})$ pairs with MSE loss. The diagrams of the reference ResU-Nets can be found in supplementary material.

Table 2: Ellipses - Map from imperfect visible to corrected visible and learned invisible. Imaging metrics averaged over test set. Note, that in the top two rows, perfect visible component $p_0^{\text{P}_{\text{vis}}}$ is used as a ground truth while, in the remaining rows, we use the original image p_0.

	MSE	PSNR	SSIM
p_0^{Linear}	$1.5031 \cdot 10^{-4} \pm 6.0914 \cdot 10^{-5}$	38.5706 ± 1.7347	0.8296 ± 0.0460
$p_{0_{\ell_1}}^{\text{I}_{\text{vis}}}$	$9.5904 \cdot 10^{-5} \pm 3.6300 \cdot 10^{-5}$	40.4581 ± 1.5361	0.9577 ± 0.0090
p_0^{Linear}	$6.0860 \cdot 10^{-3} \pm 2.2048 \cdot 10^{-3}$	22.4308 ± 1.5593	0.3284 ± 0.0436
$p_{0_{\ell_1}}^{\text{I}_{\text{vis}}}$	$5.9660 \cdot 10^{-3} \pm 2.1752 \cdot 10^{-3}$	22.5209 ± 1.5700	0.3463 ± 0.0580
$p_{0_{\text{TV}+}}^{\text{I}_{\text{vis}}}$	$8.6550 \cdot 10^{-5} \pm 4.4774 \cdot 10^{-5}$	41.0863 ± 1.9422	$\mathbf{0.9861 \pm 0.0059}$
$\mathcal{NN}_{\text{ResU-Net}}(p_0^{\text{I}_{\text{vis}}})$	$1.4197 \cdot 10^{-4} \pm 6.3716 \cdot 10^{-5}$	38.8494 ± 1.7685	0.9586 ± 0.0133
$\mathcal{NN}_{\text{ResU-Net}}(\tilde{Q}_j^{\text{I}_{\text{vis}}})$	$9.4125 \cdot 10^{-5} \pm 3.7918 \cdot 10^{-5}$	40.5950 ± 1.7027	0.9705 ± 0.0074
$\mathcal{NN}_{\text{ResCor}}(Q_j^{\text{I}_{\text{vis}}})$	$\mathbf{8.1092 \cdot 10^{-5} \pm 3.0781 \cdot 10^{-5}}$	$\mathbf{41.2081 \pm 1.6141}$	0.9755 ± 0.0069

Table 3: Ellipses - Timing and resources for imperfect learning experiment.

	Train	Execution	Parameter
p_0^{Linear}	-	1m52s	0
$p_{0_{\ell_1}}^{\text{I}_{\text{vis}}}$	-	16m23s	1
$p_{0_{\text{TV}+}}^{\text{I}_{\text{vis}}}$	-	14m12s	1
$\mathcal{NN}_{\text{ResU-Net}}(p_0^{\text{I}_{\text{vis}}})$	6h05m	~ 16s	467554
$\mathcal{NN}_{\text{ResU-Net}}(\tilde{Q}_j^{\text{I}_{\text{vis}}})$	8h12m	~ 24s	468838
$\mathcal{NN}_{\text{ResCor}}(Q_j^{\text{I}_{\text{vis}}})$	18h48m	~ 24s	1001670

The image quality measures averaged over the 300 test images are listed in Table 2, specifically MSE, PSNR and SSIM w.r.t. the provided ground-truth images. The reconstruction quality is illustrated on an example from the test set in Figure 12. Despite $p_{0_{\text{TV}+}}^{\text{I}_{\text{vis}}}$ achieving highest SSIM, the reconstructed boundaries of the ellipses highlighted in the white inset in Figure 12 (c) are incomplete (total variation apparently fails to completely recover the invisible boundaries) and considerably worse than for the learning based reconstructions (d-f). We note that the linear reconstruction p_0^{Linear} in Figure 12 (a) and $p_{0_{\ell_1}}^{\text{I}_{\text{vis}}}$ in Figure 12 (b) look similar with a small difference in contrast. The linear inversion results in slightly less clear ellipse boundaries and noisier image than the non-linear ℓ_1 reconstruction; see blue inset
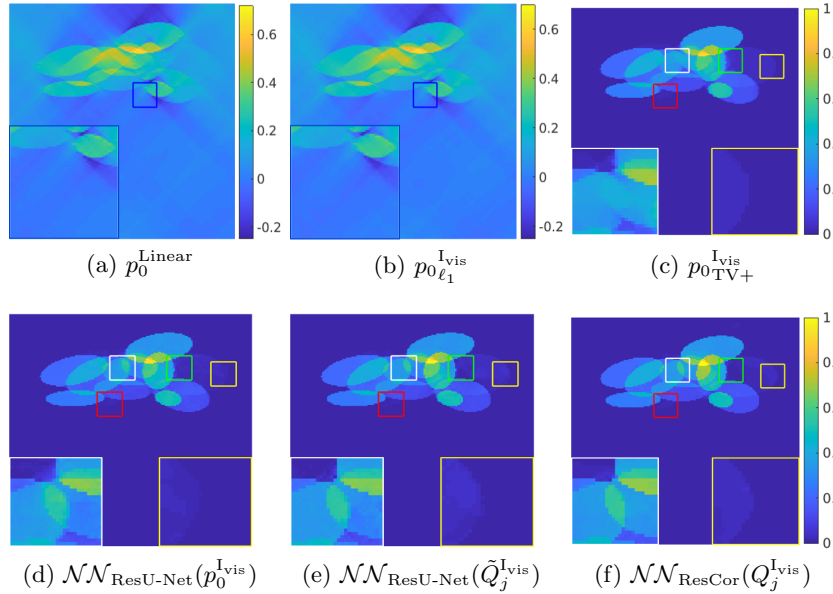
Figure 12: Ellipses - Visualisation of performance of the imperfect learning on noisy limited-angle data $g_{\angle}$, $\theta_{\max} = \pi/4$, for one test image: (a) linear inversion, (b) reconstructed imperfect visible component via (**VR**), (c) total variation regularised solution with non-negativity constraint, (d-f) different learned reconstructions.

in Figure 12 (a,b). We recall, that the linear inversion here is able to recover the visible component due to limited sensitivity angle (and even limited sensor) completely, thus any difference should be due to higher robustness of the non-linear reconstruction to noise. This is consistent with the latter's lower MSE, higher PSNR and higher SSIM compared to the linear inversion. The second row in Figure 12 (restricted to range [0,1]) shows the results of the post-processing with ResU-Nets and ResCorNet, respectively. Although the ResU-Net based methods in Figure 12 (d,e) estimate the invisible boundaries reasonably well they exhibit some high frequency noise. Upon closer inspection (white and yellow insets in Figure 12) and quantitative analysis in Table 2, our proposed ResCorNet seems to strike the best balance between the recovery of invisible boundaries while preserving the piece-wise constant interior of the ellipses. Similarly, as in the perfect learning scenario, all three methods in (d-f) fail to recover the faint green framed ellipse.

The timing and resources are listed in Table 3 (Note, that $p_{0\ell_1}^{\mathrm{I_{vis}}}$ and $p_{0\mathrm{TV}+}^{\mathrm{I_{vis}}}$ are computed using Matlab Parallel Computing Toolbox[13]). Clearly $\mathcal{NN}_{\mathrm{ResU\text{-}Net}}(p_0^{\mathrm{I_{vis}}})$ is fastest to train and execute, while $\mathcal{NN}_{\mathrm{ResCor}}(Q_j^{\mathrm{I_{vis}}})$ requires longest training time. The main reason is due to ResCorNet's larger number of trainable parameters: 1) the preceding convolutions at each scale are paramount to extract the features from the scale-wise Coronae coefficients input; 2) the Coronae coefficients sizes are determined by the partition of unity filters underpinning

---
[13]https://www.mathworks.com/products/parallel-computing.html

Curvelet decomposition, which result in larger sizes of layer inputs at each but the finest scale than those obtained from max-pooling. It is noteworthy, that at prediction stage, the approaches learning the Coronae coefficients $\mathcal{NN}_{\text{ResU-Net}}(\tilde{Q}_j^{\text{I}_{\text{vis}}})$ and $\mathcal{NN}_{\text{ResCor}}(Q_j^{\text{I}_{\text{vis}}})$ only have a slight overhead compared to $\mathcal{NN}_{\text{ResU-Net}}(p_0^{\text{I}_{\text{vis}}})$, and are much faster than the variational reconstruction approaches.

6.4. Generalisation. We now consider the generalisation properties of (Res)CorNet. To this end we chose a 4-disk phantom depicted in Figure 13. The 4-disk phantom is not in the ellipse data set (in the sense of how the set is generated) while, a disk being a special case of an ellipse, has similar features such as convexity and boundaries of C^2 type. In this experiment, the visible Curvelet coefficients are obtained solving (**VR**) for $\tau = 2.5 \cdot 10^{-4}$ with 50 FISTA iterations. Again the white noise with standard deviation $\sigma = 2.5 \cdot 10^{-4}$ is added to the corresponding limited-angle data with $\theta_{\max} = \pi/4$. We apply the ResCorNet trained on the ellipse data set to the 4-disk phantom. This scenario corresponds to training on a complicated data set and generalising our model to a less complex image with some shared and some unseen characteristics. We note that the training set contains many ellipses at least partially outside the north sector of the domain (inside which the limited sensor has no bearing on the reconstruction) while our 4-disk phantom lies strictly inside. The main limitation of the ResU-Net based reconstructions in Figure 13 (a,b) are streak like artefacts and failure to recover the invisible boundary of the large blue disk highlighted in the magenta inset. Our proposed reconstruction method using ResCorNet in Figure 13 (c) provides a notable improvement in the definition of the invisible boundary as well as the reduction in streak like background artefacts, which is consistent with its lowest MSE, highest PSRN and SSIM in Table 4. However, even in (c) parts of the invisible boundary of the faint large blue disk are missing (both left and right sides). We believe these artefacts could be due to the training set containing images of many overlapping ellipses, which means that examples of boundaries against the background are rare, only present for exterior ellipses. Furthermore, we note that the recovered shape has some resemblance with an ellipse. Both observations testify to limits of generalisation. The remaining disks are almost perfectly reconstructed by all methods. Overall, our proposed method while not without limitations is clearly superior in this generalisation scenario both in eye ball metric and the quantitative analysis in Table 4.

Table 4: 4-disk - Generalisation - Ellipse learned map from imperfect visible to corrected visible and learned invisible.

	MSE	PSNR	SSIM
$\mathcal{NN}_{\text{ResU-Net}}(p_0^{\text{I}_{\text{vis}}})$	$2.0000 \cdot 10^{-4}$	36.9776	0.9398
$\mathcal{NN}_{\text{ResU-Net}}(\tilde{Q}_j^{\text{I}_{\text{vis}}})$	$1.0000 \cdot 10^{-4}$	39.8984	0.9446
$\mathcal{NN}_{\text{ResCor}}(Q_j^{\text{I}_{\text{vis}}})$	$\mathbf{2.9599 \cdot 10^{-5}}$	$\mathbf{45.2885}$	$\mathbf{0.96866}$

7. Realistic Data. In this section we evaluate performance of our method on a realistic vessel data set. Our data set of vessel images is generated by randomly cropping images

(a) $\mathcal{NN}_{\text{ResU-Net}}(p_0^{\text{I}_{\text{vis}}})$ (b) $\mathcal{NN}_{\text{ResU-Net}}(\tilde{Q}_j^{\text{I}_{\text{vis}}})$ (c) $\mathcal{NN}_{\text{ResCor}}(Q_j^{\text{I}_{\text{vis}}})$
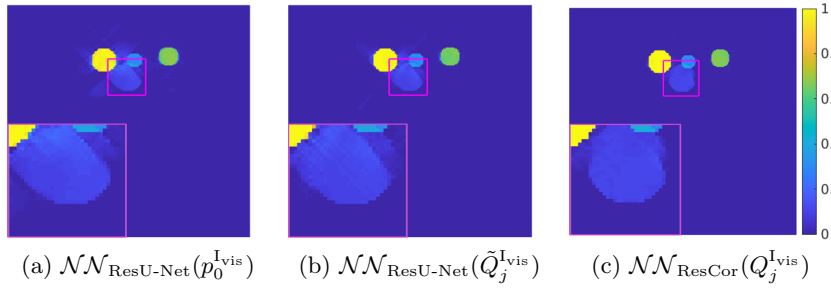
Figure 13: 4-disk Generalisation - Visualisation of generalisation performance of correcting the visible and learning the invisible from ellipse data to 4-disk phantom ($\theta_{\max} = \pi/4$): (a-c) different learned reconstructions.

from the DRIVE data set for retinal vessel segmentation[14]. The re-sampled to 192×96 and normalised to $[0,1]$ vessel images constitute the top half of the test image, with bottom half all 0, resulting in test images of size 192×192, see Figure 14. We assume homogeneous speed of sound $c = 1500$m/s. The pressure time series is recorded every $h_t = 66.67$ns by a line sensor, placed on top of the domain, with image matching resolution i.e. $h_{\mathbf{x}} = 10\mu$m resulting in a 272×192 PAT data volume. The sensitivity angle $\theta_{\max} = 2\pi/9$ is slightly smaller (and so is the visibility cone) than in the previous examples. 2000 images are used for training, 200 for validation and 200 for testing.

We perform the visible reconstruction (**VR**) in the fully wedge restricted Curvelet frame with 3 scales and 12 angles in the visible wedge (at the 2nd coarsest scale) for the regularisation parameter $\tau = 1 \cdot 10^{-4}$ with 50 FISTA iterations. As we are in the imperfect learning scenario, we train the matching ResCorNet to minimise the loss (**IPL**) with weights $\alpha_{\text{F}} = 2$, $\alpha_{\text{SF}} = 2$ and $\alpha_{\text{C}} = 1$ in the same regime as outlined in Section 6.

The Coronae coefficient based learning $\mathcal{NN}_{\text{ResU-Net}}(\tilde{Q}_j^{\text{I}_{\text{vis}}})$ and $\mathcal{NN}_{\text{ResCor}}(Q_j^{\text{I}_{\text{vis}}})$ outperforms image based learning $\mathcal{NN}_{\text{ResU-Net}}(p_0^{\text{I}_{\text{vis}}})$ as clearly visible in reconstructions in Figure 14. $\mathcal{NN}_{\text{ResU-Net}}(p_0^{\text{I}_{\text{vis}}})$ suffers from strong limited-angle artefacts while $p_{0\,\text{TV}+}^{\text{I}_{\text{vis}}}$ fails to recover the invisible vessel edges completely. Zooming in Figure 14 (red and green insets) corroborates that ResCorNet produces smooth background, sharp and detailed vessel edges and well matched contrasts. Some of these are also present in the other reconstructions but ResCorNet is unique in achieving all these goals simultaneously. This is reminiscent of ResCorNet performance for the ellipse data set (see Subsection 6.3).

The quantitative analysis is summarised in Table 5. Most notably, we observe a large performance leap across all the measures between the methods learning Coronae coefficients and those learning images which reaffirms the benefits of Coronae representation in this context.

8. Conclusion. The paper rigorously builds a framework where images are nearly optimally sparsely represented in a fully wedge restricted Curvelet frame allowing splitting into invisible/visible components, which are optimally matched with the null space/its comple-

---

[14]https://drive.grand-challenge.org/

Table 5: Vessels - Map from imperfect visible to corrected visible and learned invisible. Imaging metrics averaged over test set. Note, that in the top two rows, perfect visible component $p_0^{\mathrm{P_{vis}}}$ is used as a ground truth while, in the remaining rows, we use the original image p_0.

	MSE	PSNR	SSIM
p_0^{Linear}	$1.3135 \cdot 10^{-3} \pm 5.0308 \cdot 10^{-4}$	29.1241 ± 1.6788	0.8836 ± 0.0529
$p_{0\,\ell_1}^{\mathrm{I_{vis}}}$	$2.5338 \cdot 10^{-4} \pm 1.4832 \cdot 10^{-4}$	36.5491 ± 2.2176	0.9537 ± 0.0330
$p_{0\,\ell_1}^{\mathrm{I_{vis}}}$	$1.0201 \cdot 10^{-2} \pm 6.6000 \cdot 10^{-3}$	20.7758 ± 2.8440	0.7947 ± 0.1086
$p_{0\,\mathrm{TV}+}^{\mathrm{I_{vis}}}$	$4.3110 \cdot 10^{-3} \pm 3.0828 \cdot 10^{-3}$	24.7764 ± 3.3682	0.9398 ± 0.0292
$\mathcal{NN}_{\mathrm{ResU\text{-}Net}}(p_0^{\mathrm{I_{vis}}})$	$4.5594 \cdot 10^{-2} \pm 1.7357 \cdot 10^{-2}$	13.7066 ± 1.6040	0.6562 ± 0.1104
$\mathcal{NN}_{\mathrm{ResU\text{-}Net}}(\tilde{Q}_j^{\mathrm{I_{vis}}})$	$\mathbf{7.4186 \cdot 10^{-4} \pm 6.6170 \cdot 10^{-4}}$	32.3976 ± 2.9230	0.9540 ± 0.0717
$\mathcal{NN}_{\mathrm{ResCor}}(Q_j^{\mathrm{I_{vis}}})$	$7.5803 \cdot 10^{-4} \pm 7.0400 \cdot 10^{-4}$	$\mathbf{32.4036 \pm 3.0706}$	$\mathbf{0.9611 \pm 0.0714}$

(a) p_0 (b) $p_{0\,\ell_1}^{\mathrm{I_{vis}}}$ (c) $p_{0\,\mathrm{TV}+}^{\mathrm{I_{vis}}}$

(d) $\mathcal{NN}_{\mathrm{ResU\text{-}Net}}(p_0^{\mathrm{I_{vis}}})$ (e) $\mathcal{NN}_{\mathrm{ResU\text{-}Net}}(\tilde{Q}_j^{\mathrm{I_{vis}}})$ (f) $\mathcal{NN}_{\mathrm{ResCor}}(Q_j^{\mathrm{I_{vis}}})$
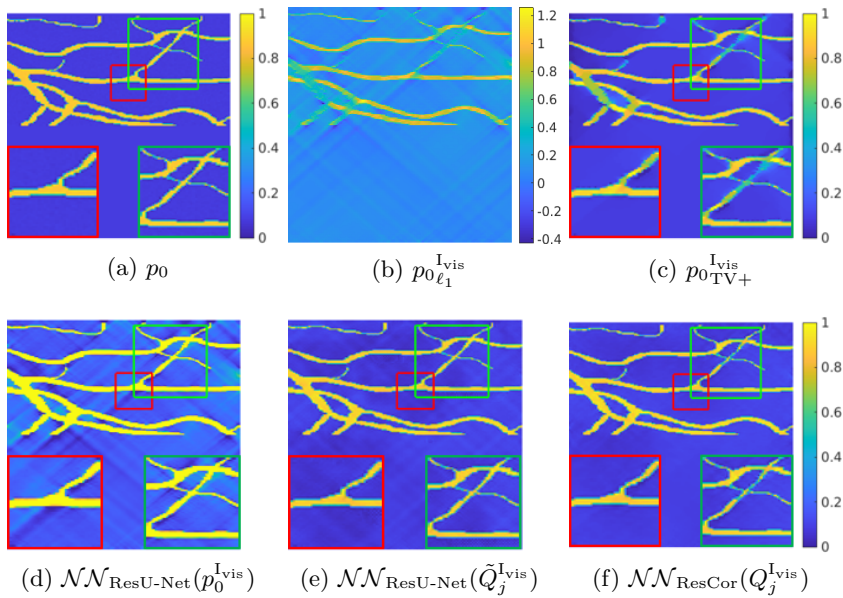
Figure 14: Vessels - Visualisation of performance of correcting the visible and learning the invisible on noisy limited-angle data $g_\angle$, $\theta_{\mathrm{max}} = 2\pi/9$, for one test image: (a) initial pressure p_0 (ground truth), (b) imperfect visible component reconstruction via **(VR)**, (c) total variation regularised solution with non-negativity constraint, (d-f) different learned reconstructions.

ment of the discrete forward operator. All computations can be performed efficiently in the Fourier domain. The network architecture working with Coronae coefficients of the visible and invisible image components is carefully designed to match both the visible/invisible split and the multiscale decomposition induced by the frame. Even, the sole learning of Coronae coefficients, as evidenced by our reference U-Net working on Coronae coefficients, already results in a substantial improvement on an image based learned post-processing using a standard U-Net. The resulting scheme implements the *reconstructing the visible and learning the invisible* strategy which decouples the forward model from the training process at no detri-

ment to reconstruction accuracy. This results in a stark performance benefit and a better interpretability compared to the model based learning which is considered the gold standard (at least in terms of image quality) among learned reconstructions.

We believe that our framework for this particular class of limited-view problems in photoacoustic tomography will further interest in applying similar ideas to other inverse problems, in particular application to the limited-angle parallel X-ray CT is immediate. As, with all learned methods, performance of our methods strongly depends on appropriateness and quality of the training data. Finally, we note that extension of our method to 3D is conceptually straightforward and, as no expensive forward/adjoint operators are evaluated in training, it is computationally feasible.

Appendix A. Adjoint PAT Fourier Operator. By the definition, the adjoint PAT Fourier operator $(\hat{\mathbf{A}}^*)$ satisfies the equality

$$(A.1) \qquad \langle \hat{g}, \hat{\mathbf{A}} \hat{p}_0 \rangle = \langle \hat{\mathbf{A}}^* \hat{g}, \hat{p}_0 \rangle$$

for any real and even w.r.t. first variable functions $\hat{g}(\omega/c, \mathbf{k}_S)$, $\hat{p}(\omega/c, \mathbf{k}_S)$, $(\omega/c, \mathbf{k}_S) \in \mathcal{R}_{\mathbf{A}} \subset \mathbb{R}^d$ and $\hat{p}_0(\mathbf{k}_\perp, \mathbf{k}_S)$, $(\mathbf{k}_\perp, \mathbf{k}_S) \in \mathbb{R}^d$ with sufficient regularity and decay (essentially in Schwarz space over $\mathbb{R}^d$) so that all the operations are well defined and $\langle \cdot, \cdot \rangle$ is the standard L_2 scalar product which we restrict to $\mathcal{R}_{\mathbf{A}}$ for functions over $(\omega/c, \mathbf{k}_S) \in \mathcal{R}_{\mathbf{A}}$.

Substituting the action of the forward operator $\hat{\mathbf{A}}$ on $\hat{p}_0$, $\hat{p}(\omega/c, \mathbf{k}_S) = \hat{\mathbf{A}} \hat{p}_0(\mathbf{k}_\perp, \mathbf{k}_S)$, on the left hand side of (A.1), we obtain

$$(A.2) \quad \langle \hat{g}, \hat{\mathbf{A}} \hat{p}_0 \rangle = \langle \hat{g}(\omega/c, \mathbf{k}_S), \hat{p}(\omega/c, \mathbf{k}_S) \rangle$$
$$= 2 \int_{\mathbb{R}_{\geq 0}} \int_{\mathbb{R}^{d-1}} \hat{g}(\omega/c, \mathbf{k}_S) \frac{\omega/c}{\sqrt{(\omega/c)^2 - |\mathbf{k}_S|^2}} \hat{p}_0 \left(\sqrt{(\omega/c)^2 - |\mathbf{k}_S|^2}, \mathbf{k}_S \right) d(\omega/c) d\mathbf{k}_S$$

where we used the even symmetry in the first variable of $\hat{g}(\omega/c, \mathbf{k}_S)$, $\hat{p}(\omega/c, \mathbf{k}_S)$, which in turn implies even symmetry of $\hat{p}_0(\mathbf{k}_\perp, \mathbf{k}_S)$ (used later), to restrict the the first integral to non-negative real axis $\mathbb{R}_{\geq 0}$.

In the next step we make use of change of variables induced by the dispersion relation in the wave equation (3.3), $(\mathbf{k}_\perp, \mathbf{k}_S) \leftarrow \left(\sqrt{(\omega/c)^2 - |\mathbf{k}_S|^2}, \mathbf{k}_S \right)$, $\mathbf{k}_\perp \geq 0$, assumed continued with even symmetry for $\mathbf{k}_\perp < 0$. The Jacobian of the change of coordinates yields

$$d\mathbf{k}_\perp = \frac{\omega/c}{\sqrt{(\omega/c)^2 - |\mathbf{k}_S|^2}} d(\omega/c).$$

Executing the change of coordinates in the last line of (A.2), the factor gets absorbed into the volume element and we are left with

$$\langle \hat{g}, \hat{\mathbf{A}} p_0 \rangle = \iint_{\mathbb{R}^d} \hat{g}(|\mathbf{k}|, \mathbf{k}_S) \hat{p}_0(\mathbf{k}_\perp, \mathbf{k}_S) d\mathbf{k}_\perp d\mathbf{k}_S$$
$$= \langle \hat{g}(|\mathbf{k}|, \mathbf{k}_S), \hat{p}_0(\mathbf{k}_\perp, \mathbf{k}_S) \rangle,$$

where the factor 2 is accounted for by the extension of the integration domain of the product of even functions from $\mathbb{R}_{\geq 0}$ back to $\mathbb{R}$. Comparing the above equation to the right hand side of the defining equality (A.1), we immediately read out the action of the adjoint operator on $\hat{g}(\omega/c, \mathbf{k}_S)$ as

$$(\text{A.3}) \qquad \hat{\mathbf{A}}^* \hat{g}(\omega/c, \mathbf{k}_S) = \hat{g}(|\mathbf{k}|, \mathbf{k}_S).$$

REFERENCES

[1] J. ADLER AND O. ÖKTEM, *Solving ill-posed inverse problems using iterative deep neural networks*, Inverse Problems, 33 (2017), p. 124007.

[2] J. ADLER AND O. ÖKTEM, *Learned primal-dual reconstruction*, IEEE transactions on medical imaging, 37 (2018), pp. 1322–1332.

[3] H. ANDRADE-LOARCA, G. KUTYNIOK, O. ÖKTEM, AND P. PETERSEN, *Deep microlocal reconstruction for limited-angle tomography*, arXiv preprint arXiv:2108.05732, (2021).

[4] H. ANDRADE-LOARCA, G. KUTYNIOK, O. OKTEM, AND P. C. PETERSEN, *Extraction of digital wavefront sets using applied harmonic analysis and deep neural networks*, SIAM Journal on Imaging Sciences, 12 (2019), pp. 1936–1966.

[5] S. ANTHOLZER, M. HALTMEIER, R. NUSTER, AND J. SCHWAB, *Photoacoustic image reconstruction via deep learning*, in Photons Plus Ultrasound: Imaging and Sensing 2018, vol. 10494, International Society for Optics and Photonics, 2018, p. 104944U.

[6] S. ANTHOLZER, M. HALTMEIER, AND J. SCHWAB, *Deep learning for photoacoustic tomography from sparse data*, Inverse problems in science and engineering, 27 (2019), pp. 987–1005.

[7] S. ANTHOLZER, J. SCHWAB, AND M. HALTMEIER, *Deep learning versus ℓ_1-minimization for compressed sensing photoacoustic tomography*, in 2018 IEEE International Ultrasonics Symposium (IUS), IEEE, 2018, pp. 206–212.

[8] S. ARRIDGE, P. BEARD, M. BETCKE, B. COX, N. HUYNH, F. LUCKA, O. OGUNLADE, AND E. ZHANG, *Accelerated high-resolution photoacoustic tomography via compressed sensing*, Physics in Medicine & Biology, 61 (2016), p. 8908.

[9] S. ARRIDGE, P. MAASS, O. ÖKTEM, AND C.-B. SCHÖNLIEB, *Solving inverse problems using data-driven models*, Acta Numerica, 28 (2019), pp. 1–174.

[10] S. R. ARRIDGE, M. M. BETCKE, B. T. COX, F. LUCKA, AND B. E. TREEBY, *On the adjoint operator in photoacoustic tomography*, Inverse Problems, 32 (2016), p. 115012.

[11] P. BEARD, *Biomedical photoacoustic imaging*, Interface focus, 1 (2011), pp. 602–631.

[12] A. BECK AND M. TEBOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM journal on imaging sciences, 2 (2009), pp. 183–202.

[13] M. M. BETCKE, B. T. COX, N. HUYNH, E. Z. ZHANG, P. C. BEARD, AND S. R. ARRIDGE, *Acoustic wave field reconstruction from compressed measurements with application in photoacoustic tomography*, IEEE Transactions on Computational Imaging, 3 (2017), pp. 710–721.

[14] Y. E. BOINK, C. BRUNE, AND S. MANOHAR, *Robustness of a partially learned photoacoustic reconstruction algorithm*, in Photons Plus Ultrasound: Imaging and Sensing 2019, vol. 10878, International Society for Optics and Photonics, 2019, p. 108781D.

[15] Y. E. BOINK, M. J. LAGERWERF, W. STEENBERGEN, S. A. VAN GILS, S. MANOHAR, AND C. BRUNE, *A reconstruction framework for total generalised variation in photoacoustic tomography*, ArXiv e-prints, (2017).

[16] Y. E. BOINK, S. MANOHAR, AND C. BRUNE, *A partially-learned algorithm for joint photo-acoustic reconstruction and segmentation*, IEEE transactions on medical imaging, 39 (2019), pp. 129–139.

[17] Y. E. BOINK, S. A. VAN GILS, S. MANOHAR, AND C. BRUNE, *Sensitivity of a partially learned model-based reconstruction algorithm*, PAMM, 18 (2018), p. e201800222.

[18] T. A. BUBBA, M. GALINIER, M. LASSAS, M. PRATO, L. RATTI, AND S. SILTANEN, *Deep neural networks for inverse problems with pseudodifferential operators: an application to limited-angle tomography*, SIAM Journal on Imaging Sciences, (2021).

[19] T. A. BUBBA, G. KUTYNIOK, M. LASSAS, M. MÄRZ, W. SAMEK, S. SILTANEN, AND V. SRINIVASAN, *Learning the invisible: A hybrid deep learning-shearlet framework for limited angle computed tomography*, Inverse Problems, 35 (2019), p. 064002.

[20] E. CANDES, L. DEMANET, D. DONOHO, AND L. YING, *Fast discrete curvelet transforms*, multiscale modeling & simulation, 5 (2006), pp. 861–899.

[21] E. J. CANDES AND L. DEMANET, *The curvelet representation of wave propagators is optimally sparse*, Communications on Pure and Applied Mathematics, 58 (2005), pp. 1472–1528.

[22] E. J. CANDÈS AND D. L. DONOHO, *New tight frames of curvelets and optimal representations of objects with piecewise c2 singularities*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 57 (2004), pp. 219–266.

[23] E. J. CANDES AND D. L. DONOHO, *Continuous curvelet transform: I. resolution of the wavefront set*, Applied and Computational Harmonic Analysis, 19 (2005), pp. 162–197.

[24] B. T. COX AND P. C. BEARD, *Fast calculation of pulsed photoacoustic fields in fluids using k-space methods*, The Journal of the Acoustical Society of America, 117 (2005), pp. 3616–3627.

[25] D. L. DONOHO, *Sparse components of images and optimal atomic decompositions*, Constructive Approximation, 17 (2001), pp. 353–382.

[26] D. FINCH AND S. K. PATCH, *Determining a function from its mean values over a family of spheres*, SIAM journal on mathematical analysis, 35 (2004), pp. 1213–1240.

[27] J. FRIKEL, *A new framework for sparse regularization in limited angle x-ray tomography*, in 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2010, pp. 824–827, https://doi.org/10.1109/ISBI.2010.5490113.

[28] J. FRIKEL, *Sparse regularization in limited angle tomography*, Applied and Computational Harmonic Analysis, 34 (2013), pp. 117–141.

[29] J. FRIKEL AND M. HALTMEIER, *Efficient regularization with wavelet sparsity constraints in photoacoustic tomography*, Inverse Problems, 34 (2018), p. 024006, https://doi.org/10.1088/1361-6420/aaa0ac, https://doi.org/10.1088/1361-6420/aaa0ac.

[30] J. FRIKEL AND E. T. QUINTO, *Characterization and reduction of artifacts in limited angle tomography*, Inverse Problems, 29 (2013), p. 125007, https://doi.org/10.1088/0266-5611/29/12/125007, https://doi.org/10.1088/0266-5611/29/12/125007.

[31] J. FRIKEL AND E. T. QUINTO, *Artifacts in incomplete data tomography with applications to photoacoustic tomography and sonar*, SIAM Journal on Applied Mathematics, 75 (2015), pp. 703–725, http://www.jstor.org/stable/24511467.

[32] X. GLOROT AND Y. BENGIO, *Understanding the difficulty of training deep feedforward neural networks*, in Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.

[33] S. GUAN, A. A. KHAN, S. SIKDAR, AND P. V. CHITNIS, *Fully dense unet for 2-d sparse photoacoustic tomography artifact removal*, IEEE journal of biomedical and health informatics, 24 (2019), pp. 568–576.

[34] M. HALTMEIER, T. BERER, S. MOON, AND P. BURGHOLZER, *Compressed sensing and sparsity in photoacoustic tomography*, Journal of Optics, 18 (2016), p. 114004.

[35] K. HAMMERNIK, T. KLATZER, E. KOBLER, M. P. RECHT, D. K. SODICKSON, T. POCK, AND F. KNOLL, *Learning a variational network for reconstruction of accelerated mri data*, Magnetic resonance in medicine, 79 (2018), pp. 3055–3071.

[36] A. HAUPTMANN, B. COX, F. LUCKA, N. HUYNH, M. BETCKE, P. BEARD, AND S. ARRIDGE, *Approximate k-space models and deep learning for fast photoacoustic reconstruction*, in International Workshop on Machine Learning for Medical Image Reconstruction, Springer, 2018, pp. 103–111.

[37] A. HAUPTMANN AND B. T. COX, *Deep learning in photoacoustic tomography: Current approaches and future directions*, Journal of Biomedical Optics, 25 (2020), p. 112903.

[38] A. HAUPTMANN, F. LUCKA, M. BETCKE, N. HUYNH, J. ADLER, B. COX, P. BEARD, S. OURSELIN, AND S. ARRIDGE, *Model-based learning for accelerated, limited-view 3-d photoacoustic tomography*, IEEE transactions on medical imaging, 37 (2018), pp. 1382–1393.

[39] C. HUANG, K. WANG, L. NIE, L. V. WANG, AND M. A. ANASTASIO, *Full-wave iterative image reconstruction in photoacoustic tomography with acoustically inhomogeneous media*, IEEE transactions on medical imaging, 32 (2013), pp. 1097–1110.

[40] N. Huynh, F. Lucka, E. Zhang, M. Betcke, S. Arridge, P. Beard, and B. Cox, *Sub-sampled fabry-perot photoacoustic scanner for fast 3d imaging*, in Photons Plus Ultrasound: Imaging and Sensing 2017, vol. 10064, International Society for Optics and Photonics, 2017, p. 100641Y.

[41] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, *Deep convolutional neural network for inverse problems in imaging*, IEEE Transactions on Image Processing, 26 (2017), pp. 4509–4522.

[42] E. Kang, J. Min, and J. C. Ye, *A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction*, Medical physics, 44 (2017), pp. e360–e375.

[43] K. P. Köstli, M. Frenz, H. Bebie, and H. P. Weber, *Temporal backward projection of optoacoustic pressure transients using fourier transform methods*, Physics in Medicine & Biology, 46 (2001), p. 1863.

[44] H. Lan, D. Jiang, C. Yang, and F. Gao, *Y-net: a hybrid deep learning reconstruction framework for photoacoustic imaging in vivo*, arXiv preprint arXiv:1908.00975, (2019).

[45] L. Nie and X. Chen, *Structural and functional photoacoustic molecular tomography aided by emerging contrast agents*, Chemical Society Reviews, 43 (2014), pp. 7132–7170.

[46] B. Pan, M. Betcke, S. Arridge, F. Lucka, B. Cox, N. Huynh, P. Beard, and E. Zhang, *Photoacoustic reconstruction using sparsity in curvelet frame: Image versus data domain*, IEEE Transactions on Computational Imaging, (2021).

[47] E. T. Quinto, *Singularities of the x-ray transform and limited data tomography in $\mathbb{R}^2$ and $\mathbb{R}^3$*, SIAM Journal on Mathematical Analysis, 24 (1993), pp. 1215–1225.

[48] E. T. Quinto, *Artifacts and visible singularities in limited data x-ray tomography*, Sensing and Imaging, 18 (2017), p. 9.

[49] J. Schwab, S. Antholzer, and M. Haltmeier, *Deep null space learning for inverse problems: convergence analysis and rates*, Inverse Problems, 35 (2019), p. 025008.

[50] J. Schwab, S. Antholzer, and M. Haltmeier, *Learned backprojection for sparse and limited view photoacoustic tomography*, in Photons Plus Ultrasound: Imaging and Sensing 2019, vol. 10878, International Society for Optics and Photonics, 2019, p. 1087837.

[51] J. Schwab, S. Antholzer, R. Nuster, and M. Haltmeier, *Real-time photoacoustic projection imaging using deep learning*, arXiv preprint arXiv:1801.06693, (2018).

[52] J. Schwab, S. Pereverzyev Jr, and M. Haltmeier, *A galerkin least squares approach for photoacoustic tomography*, SIAM Journal on Numerical Analysis, 56 (2018), pp. 160–184.

[53] H. Shan, G. Wang, and Y. Yang, *Accelerated correction of reflection artifacts by deep neural networks in photo-acoustic tomography*, Applied Sciences, 9 (2019), p. 2615.

[54] H. Shan, C. Wiedeman, G. Wang, and Y. Yang, *Simultaneous reconstruction of the initial pressure and sound speed in photoacoustic tomography using a deep-learning approach*, in Novel Optical Systems, Methods, and Applications XXII, vol. 11105, International Society for Optics and Photonics, 2019, p. 1110504.

[55] J.-L. Starck, E. J. Candès, and D. L. Donoho, *The curvelet transform for image denoising*, IEEE Transactions on Image Processing, 11 (2002), pp. 670–684.

[56] P. Stefanov and G. Uhlmann, *Thermoacoustic tomography with variable sound speed*, Inverse Problems, 25 (2009), p. 075011.

[57] K. S. Valluru, K. E. Wilson, and J. K. Willmann, *Photoacoustic imaging in oncology: translational preclinical and early clinical experience*, Radiology, 280 (2016), pp. 332–349.

[58] K. Wang and M. A. Anastasio, *Photoacoustic and thermoacoustic tomography: image formation principles*, in Handbook of Mathematical Methods in Imaging, Springer, 2011, pp. 781–815.

[59] L. V. Wang, *Multiscale photoacoustic microscopy and computed tomography*, Nature photonics, 3 (2009), pp. 503–509.

[60] J. Xia and L. V. Wang, *Small-animal whole-body photoacoustic tomography: a review*, IEEE Transactions on Biomedical Engineering, 61 (2013), pp. 1380–1389.

[61] M. Xu and L. V. Wang, *Universal back-projection algorithm for photoacoustic computed tomography*, Physical Review E, 71 (2005), p. 016706.

[62] Y. Xu, D. Feng, and L. V. Wang, *Exact frequency-domain reconstruction for thermoacoustic tomography. i. planar geometry*, IEEE transactions on medical imaging, 21 (2002), pp. 823–828.

[63] Y. Xu, L. V. Wang, G. Ambartsoumian, and P. Kuchment, *Reconstructions in limited-view thermoacoustic tomography*, Medical physics, 31 (2004), pp. 724–733.

[64] C. Yang, H. Lan, and F. Gao, *Accelerated photoacoustic tomography reconstruction via recurrent in-*

ference machines, in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 6371–6374.

[65] Y. Zhou, J. Yao, and L. V. Wang, *Tutorial on photoacoustic tomography*, Journal of biomedical optics, 21 (2016), p. 061007.

[66] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, *Image reconstruction by domain-transform manifold learning*, Nature, 555 (2018), pp. 487–492.