

Deep Optimization Prior for THz Model Parameter Estimation

Tak Ming Wong*, Hartmut Bauermeister†, Matthias Kahl‡,

Peter Haring Bolívar‡, Michael Möller†, Andreas Kolb*

Center for Sensor Systems (ZESS), University of Siegen, Germany

Abstract

In this paper, we propose a deep optimization prior approach with application to the estimation of material-related model parameters from terahertz (THz) data that is acquired using a Frequency Modulated Continuous Wave (FMCW) THz scanning system. A stable estimation of the THz model parameters for low SNR and shot noise configurations is essential to achieve acquisition times required for applications in, e.g., quality control. Conceptually, our deep optimization prior approach estimates the desired THz model parameters by optimizing for the weights of a neural network. While such a technique was shown to improve the reconstruction quality for convex objectives in the seminal work of Ulyanovet al., our paper demonstrates that deep priors also allow to find better local optima in the non-convex energy landscape of the nonlinear inverse problem arising from THz imaging. We verify this claim numerically on various THz parameter estimation problems for synthetic and real data under low SNR and shot noise conditions. While the low SNR scenario not even requires regularization, the impact of shot noise is significantly reduced by total variation (TV) regularization. We compare our approach with existing optimization techniques that require sophisticated physically motivated initialization, and with a 1D single-pixel reparametrization method.

1. Introduction

Terahertz (THz) imaging is an emerging sensing technology with great potential for contact-free material analysis and non-destructive testing not only in application areas such as quality control in the semiconductor or the manufacturing industry, but also for civil security applications [5, 8, 15, 22, 26].

However, current physically-interpretable material quantities can only be derived reliably from THz data acquired by highly specialized THz spectroscopic instrumentation operating in well controlled experimental lab environments. It is, therefore, of high practical interest to enable the estimation of material-related information using 3D THz data attained with widely used THz imaging components. Frequency Modulated Continuous Wave (FMCW) THz systems [12] capture data in the frequency domain at each pixel, for which the associated physical quantities must be estimated according to the known THz data formation model. After being transferred to the spatial domain, this estimation comprises a sophisticated, non-linear optimization process. Due to the low signal strength of the widely used THz sources, it takes up to hours to acquire high Signal-to-Noise Ratio (SNR) THz image data for robust parameter estimations, and the parameter estimation for high SNR data already requires significant optimization efforts and fine tuned parameter initialization. Our approach is to improve the robustness of the parameter estimation process for lower SNR THz data.

Often image analysis and reconstruction problems, such as the THz problem stated above, are modelled such that a quantity of interest u_{xy} is extracted from measurements g_{xy} at every pixel (x, y) to match a given nonlinear data formation process A , i.e., $g_{xy} = A(u_{xy})$, yielding optimization problems of the form

$$\min_u \sum_{xy} \mathcal{L}(A(u_{xy}), g_{xy}) + R(u), \quad (1)$$

where \mathcal{L} and R are a suitable discrepancy measure (loss) and an optional regularization, respectively. As A is commonly nonlinear and the problem (1) is highly nonconvex, (1) is often solved locally with first-order descent methods.

In this paper we propose the concept of *deep optimization prior*, a novel unsupervised method to solve highly non-linear optimization problems. Deep image prior [28] is applied to problems similar to (1), but for linear inverse imaging problems yielding *convex* optimization problems

*Computer Graphics and Multimedia Systems Group

†Computer Vision Group

‡Institute for High Frequency and Quantum Electronics (HQE)

with unique global minimizers, by reparameterizing the original (image) variable u as the output of a CNN $\mathcal{N}(g; \theta)$

$$\min_{\theta} \sum_{xy} \mathcal{L}(A(\mathcal{N}(g; \theta)_{xy}), g_{xy}) + R(\mathcal{N}(g; \theta)). \quad (2)$$

Subsequently, they omit the regularizer R , stop the iteration early, and we conclude that the reconstruction obtained this way is of higher quality. Our work extends deep image prior to *non-convex* optimization problems and shows that not only the quality of the solution increases, but also the ability to find *lower energy minima*: By reparameterizing the originally *spatially uncoupled* variables u as the output of a U-net [24] acting on the data, a gradient descent algorithm is able to avoid undesirable local minima when the same algorithm on the original variables gets stuck in. Most strikingly, the quality of a classical approach (1) has a severe dependency on a good initialization with physical knowledge, while the common *random initialization* of network weights seems to be sufficient for consistently finding good local minima for (2).

Our approach comprises the following contributions:

- We propose the *deep optimization prior* concept, *i.e.*, the reparametrization of a pixel-wise non-convex THz model parameter estimation problem via a spatially coupled 3D neural network.
- We theoretically show that surjective reparametrizations can never eliminate critical points of the original cost function, but merely yield a different gradient descent path.
- We evaluate our approach and demonstrate that the deep optimization prior approach robustly reconstructs THz model parameters in low SNR and shot noise situations. It finds significantly better local minima compared to classical optimization methods and a single-pixel DNN, and it is very memory efficient.

2. THz Imaging Primer

In this section we briefly describe the THz FMCW scanning system, the acquired data, and the THz model. For further details we refer to the work of Wong *et al.* [29, 30].

Our reference FMCW THz system uses active frequency modulated THz radiation to sense reflected signals including phase shifts for around 1,000 frequency samples between 514–640 GHz. The imaging unit consists of a THz transmitter that illuminates the scene, and a receiver with optical components that capture the reflected signal. The unit is mounted on a $x - y$ translational platform which is used to scan the scene pixel-by-pixel (more details in [9]).

We denote the acquired reflected electric field amplitude and phase at lateral position $(x, y) \in \mathbb{R}^2$ with $\hat{g}((x, y), f)$. In FMCW radar signal processing, this frequency signal is

converted into time domain by a Fourier transform, yielding

$$g_t(x, y, t) = \mathcal{F}\{\hat{g}(x, y, f)\}. \quad (3)$$

Using the speed of light, $g_t(x, y, t)$ can be directly scaled to the resulting complex valued spatial 3D THz signal $g(x, y, z) \in \mathbb{C}^{n_x \times n_y \times n_z}$, where n_x, n_y, n_z is the number of vertical, horizontal and depth samples, respectively. Equivalently, we may represent g by considering the real and imaginary parts as two separate channels, resulting in a 4D real data tensor $G \in \mathbb{R}^{n_x \times n_y \times n_z \times 2}$.

After the FFT in (3), the z -direction signal envelope is an ideal sinc function as continuous spatial signal amplitude, giving rise to the physical THz model

$$A(u; z) = \hat{e} \operatorname{sinc}(\sigma(z - \mu)) \exp(-i(\omega z - \phi)) \quad (4)$$

where the THz model parameters $u = (\hat{e}, \mu, \sigma, \phi)$ relate to the electric field amplitude, the z -position of the surface, the width of the reflected pulse, and the phase of the spatial signal $g(x, y, z)$, respectively. While σ is a system parameter, the THz parameters \hat{e} and ϕ are material dependent. More details of the THz model are described in [29].

3. Prior Work

Classical Optimization Approaches Inverse problems have been studied in the microwave and THz communities for the past two decades, with widespread use of classical, mainly gradient-based optimization approaches. Early works proposed to estimate material thickness and reflective index for THz time domain spectroscopy (TDS) system using a gradient descent algorithm [10], to utilize sequential quadratic programming and the genetic algorithm for multilayer structure permittivity extraction [1], and to apply a combined genetic algorithm and gradient descent method for complex permittivity measurement for dielectric multilayer structures [23]. More recent works have proposed to use the limited-memory BFGS algorithm (LBFGS) for THz-TDS multilayer imaging [3], to use the Levenberg-Marquardt method for THz single pixel image reconstruction [4], to utilize the Gauss-Newton method for FMCW THz complex permittivity estimation of dielectric and non-magnetic materials [2]. Wong *et al.* [29] proposed to use the Trust-Region Algorithm for FMCW THz image reconstruction and deconvolution, to use a steepest gradient descent method and the Nelder-Mead algorithm for FMCW THz thickness estimation [25], and to use the MUSIC algorithm to estimate per-pixel FMCW THz model parameters [14]. Clark *et al.* [6] proposed to utilize a neural network for the initialization of Gauss-Newton optimizers, in order to solve non-linear least squares optimization problems in the application of motion stereo.

Deep Learning Based Approaches Some research has been conducted to apply deep learning methods to inverse

THz problems. *Liet al.* [16] proposed to train a convolutional neural network for THz-TDS image super-resolution. *Mao et al.* [20] proposed to train a CNN architecture for THz-TDS integrated circuit defect detection. *Wong et al.* [30] proposed a per-pixel model-based autoencoder for FMCW THz image reconstruction. *Suet et al.* [27] applied supervised deep learning by a standard U-net for THz tomographic imaging, which reconstructed 3D object from a series of corrupted THz-TDS 2D images.

1D Per-Pixel Auto-Encoder As our deep optimization prior approach extends the approach of *Wong et al.* [30], we discuss this approach in more detail. This approach directly estimates $u = (\hat{e}, \sigma, \mu, \phi)$ from the given $\mathbb{R}^{n_z \times 2}$ complex samples $g_{x,y}$ in the spatial domain utilizing a model-based autoencoder network $\mathcal{N}^{1D}(\cdot; \theta)$ and the loss function

$$\min_{\theta} \sum_{\text{training batch } b} \|A(\mathcal{N}^{1D}(G_{x,y}^b; \theta)) - G_{x,y}^b\|_2^2. \quad (5)$$

The approach uses an unsupervised training method to fit the THz model (4) using 80% of the pixels for training in (5), and tests if the resulting network can directly predict the desired parameters on the remaining 20% of the THz image pixels. The results in [30, Tab. 1, Figs. 8c) and 8d)] show that the matching quality of the model-based autoencoder is comparable to the optimization approach [29] and using the auto-encoder's output as an initialization of a classical optimizer yields superior loss values.

4. Proposed Method

The aim of THz model parameter estimation is to extract the parameters $u(x, y) = (\hat{e}(x, y), \sigma(x, y), \mu(x, y), \phi(x, y)) \in \mathbb{R}^4$ of the THz model (4) at each pixel location (x, y) such that it corresponds to the given FMCW THz measurements $G(x, y) \in \mathbb{R}^{n_z \times 2}$, *i.e.* $\min_u \mathcal{L}(A(u), G(x, y))$. Even with simple choices of the loss function \mathcal{L} such as an ℓ^2 -squared loss, the resulting fitting problem is highly nonconvex. Existing approaches [29, 30] use a simple ℓ^2 -squared loss

$$\min_u \sum_{x,y} \|A(u_{x,y}) - G_{x,y}\|_2^2, \quad (6)$$

in order to not further exacerbate the goal of a robust THz parameter estimation. Moreover, applying local first order or quasi-Newton methods to $10^5 - 10^6$ pixel optimizations (6) is already quite costly. To the best of our knowledge, no regularization approaches have been applied to this kind of THz model parameter estimation so far.

4.1. Deep Optimization Prior

Recalling the overall approach depicted in Sec. 1, the main idea is to reparameterize the unknown (image) variable $u_{x,y}$ in nonconvex optimization problems of the

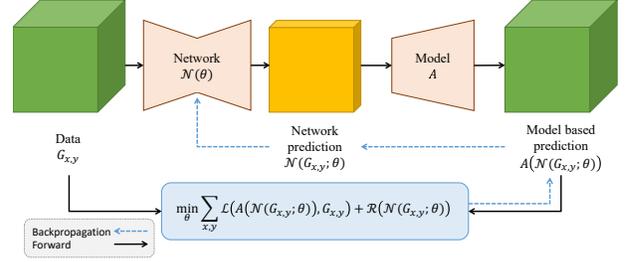


Figure 1: The proposed deep optimization prior approach: Reparametrizing $u_{x,y}$ by a network \mathcal{N} in combination with the model-based autoencoder.

form (1) by the prediction of a neural network \mathcal{N} via $u_{x,y} = \mathcal{N}(G_{x,y}; \theta)$ for network parameters, yielding a reformulated optimization problem of the form (2).

Fig. 1 illustrates the overall *deep optimization prior* network architecture comprising the combination of the reparameterization network and the model-based autoencoder. The proposed deep optimization prior approach is minimizing the loss function \mathcal{L} as an optimizer during the unsupervised *training* procedure, which is different to the unsupervised *training-then-prediction* approach proposed by [30]. The network-based reparameterization for THz model parameter estimation by combining (6) and (2) is formulated as:

$$\min_{\theta} \sum_{x,y} \|A(\mathcal{N}(G; \theta)_{x,y}) - G_{x,y}\|_2^2. \quad (7)$$

Besides the data term loss function (7), a regularization term for THz model parameter estimation can be applied. As shown in Sec. 7, the regularization improves the THz parameter estimation in the case of individual pixel failure, *i.e.* shot noise. We add a regularizing term to (7) as follows:

$$\min_{\theta} \sum_{x,y} \|A(\mathcal{N}(G; \theta)_{x,y}) - G_{x,y}\|_2^2 + \lambda \|\nabla \mathcal{N}(G; \theta)_{x,y}\|_1, \quad (8)$$

where ∇ is the gradient operator on neighborhood pixels applied to the model parameter predicted by network \mathcal{N} , weighted by the regularization coefficients λ .

4.2. 3D Model-based Autoencoder

As stated in [30], a model-based autoencoder allows unsupervised learning of measurement data by resembling an autoencoder with a learnable network based encoder and a physical model-based decoder, and therefore able to deal with measurement-specific distortions. However, during the learning phase in the per-pixel learning approach, the lateral neighborhood information is not considered. We propose

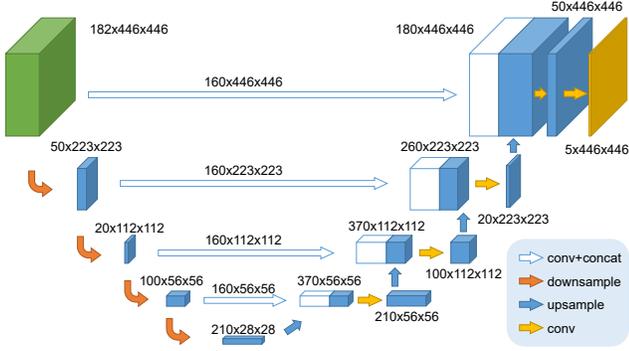


Figure 2: The proposed U-net architecture of network \mathcal{N} (example for 182 channels with 446×446 pixels) start from the data tensor $G_{x,y}$ (green box) to the desired parameter $u_{x,y} = \mathcal{N}(G_{x,y}; \theta)$ (yellow box). Blue boxes represent feature maps.

the extension to a 3D model-based autoencoder which allows unsupervised learning on the THz measurement data using the proposed deep optimization prior approach for a lateral spatial coupled optimization. Please note, that in contrast to the 1D single pixel autoencoder [30], our network-based reparameterization (Sec. 4.1) allows spatial coupling even though the THz model (4) is independent in the lateral spatial domain.

Network Architecture As the THz measurement data is a high-dimension data in the form of a 4-D tensor, the computational requirement of the network is intense. Using typical CNN architectures similar to [30] is no longer possible because of the high memory and computational requirement of the fully connected layers. Hence, we utilize a U-net-type network similar to [24]. A U-net is a kind of fully convolutional networks (FCN) [18], *i.e.* it only consists of convolutional layers, and is, therefore, computational extremely efficient, while it couples pixels in large lateral spatial regions, which is an important feature in our application.

We optimized the exact structure of the U-net architecture via an ablation study on one particular THz imaging dataset *MetalPCB*. We made two changes to the original U-net architecture [24] that improve the results for optimization purposes significantly: First, we do not double the number of channels in the encoder-part, but insert an intermediate bottleneck. Second, the number of channels that are skipping from the encoder to the decoder features at the same scale via a concatenation is increased from the commonly used 4 channels to 160 channels (see Fig. 2). For a better understanding of the effects of these changes, please consult the ablation study in the supplementary material.

In order to ensure the predicted parameters lie within the reasonable physical scale indicated by the dataset, a sigmoid transfer function is applied to the predicted parameter

$u_{x,y}$ and is linearly scaled to get $\hat{u}_{x,y} \in [u_{\min}, u_{\max}]$. Finally, the forward model $A(\hat{u}_{x,y})$ takes the projected parameter as input to generate the spatial THz model data.

In the per-pixel autoencoder [30], the network directly estimate the phase angle $\phi \in [0, 2\pi]$ linearly, raising the problem of the gradient computation across the $0 - 2\pi$ phase wrap. We solve this problem by predicting two real-valued phase components $\phi_c = \cos \phi$ and $\phi_s = \sin \phi$, and reconstructing the phase afterwards $\phi = \text{atan2}(\phi_s, \phi_c)$. Hence, the proposed U-net architecture predicts 5 parameters $\{\hat{e}, \sigma, \mu, \phi_c, \phi_s\}$ (see Fig. 2).

For a fair comparison, we adopt the per-pixel autoencoder [30] by incorporating the phase unwrapping part in evaluation in Sec. 7.1, 7.2 and 7.3.

The details of this unsupervised training procedure is given in the supplementary material.

5. Theoretical Aspects of Reparametrizations

In this section we provide a theoretical analysis of the proposed parameterization using neural networks and show that it implicitly corresponds to a variable metric optimization strategy for problem (1).

Neglecting the regularizer, problem (1) is in itself not coupled on a pixel level. For the sake of simplicity consider for now the general uncoupled problem

$$\min_{u \in \Omega} \sum_i h_i(u_i), \quad (9)$$

where $\Omega = \Omega_{i_1} \times \dots \times \Omega_{i_n}$ is the product space of the pixel-wise domains and h_i are (non-convex) cost functions at pixel i . Clearly, minimizing (9) reduces to minimizing problem h_i for each pixel i as the sum of the cost functions decouples on a pixel level. Therefore gradient descent on problem (9) corresponds to gradient descent on each of the subproblems h_i . Considering a reparametrization of the problem by a continuous function $\mathcal{N}: \Theta \rightarrow \Omega$ yields

$$\min_{\theta \in \Theta} (H \circ \mathcal{N})(\theta). \quad (10)$$

for $H(u) := \sum_i h_i(u_i)$, and thus generalizes (7). Although the problems at pixel level can share a common structure, reformulation (10) alone without knowledge of this structure is not advantageous in general due to the the preservation of local geometries, as stated in the following remark

Remark 1. Preservation of local minima. Let \hat{u} be a local minimizer of H in the range of \mathcal{N} . Then each $\hat{\theta} \in \mathcal{N}^{-1}(\hat{u})$ is also a local minimizer of $H \circ \mathcal{N}$.

Furthermore assuming differentiability of \mathcal{N} , consider a continuous interpretation of gradient descent, the *gradient flow* w.r.t H , *i.e.* a $\theta(t)$ s.t. $\theta'(t) = -\nabla(H \circ \mathcal{N})(\theta(t))$.

Then for $u(t) := \mathcal{N}(\theta(t))$ it holds

$$u'(t) = \nabla \mathcal{N}(\theta(t))^T \theta'(t) \quad (11)$$

$$= -\nabla \mathcal{N}(\theta(t))^T \nabla (H \circ \mathcal{N})(\theta(t)) \quad (12)$$

$$= -\nabla \mathcal{N}(\theta(t))^T \nabla \mathcal{N}(\theta(t)) \nabla H(u(t)). \quad (13)$$

For $\Omega \subset \mathbb{R}^d$ the matrix $M(t) := \nabla \mathcal{N}(\theta(t))^T \nabla \mathcal{N}(\theta(t)) \in \mathbb{R}^{d \times d}$ is positive semi-definite and hence $-M(t) \nabla H(u(t))$ is a descent direction. We hypothesize that for certain problem classes h_i as studied in the numerical experiments the temporally changing implicit gradient preconditioning with $M(t)$ is advantageous in terms of training dynamics. In particular, networks with a large receptive field such as a U-net typically yield dense matrices $\nabla \mathcal{N}(\theta(t))^T \nabla \mathcal{N}(\theta(t))$ and thus induce changes in predictions u_{xy} even if $\frac{\partial H}{\partial u_{xy}} = 0$.

6. Experimental Setup

Datasets We evaluate the optimization performance of classical approaches and the proposed method on synthetically simulated and measured THz datasets:

- *MetalPCB*: Measured THz image dataset from [30].
- *SynthUSAF*: Ground truth THz model parameters are synthetically generated and the raw THz data is synthesized using (4).
- *SynthObj*: Ground truth parameters are synthetically simulated from an 3D object from [11].

To simulate different noise level, we use two synthetic noise models:

- *+AWGN*: adding Additive White Gaussian Noise (AWGN) by -20 to 10 dB peak-SNR (PSNR) to the frequency domain signal, with a static background noise level and a varying signal power.
- *+ShotNoise*: On top of the AWGN, a random salt-and-pepper noise signal with 60dB higher power is added for 10% of the pixels.

All synthetic datasets are generated using the size of *MetalPCB*, i.e. $446 \times 446 \times 91$.

Choice of optimizer Commonly used optimization methods for the THz inverse problem can be categorised as (see Sec. 3): Hessian based methods (second order gradient), which include Levenberg Marquardt [21], Trust Region Algorithm [7], and LBFGS [17]; Gradient descent methods (first order gradient), which include gradient descent, and steepest gradient descent.

We optimize the deep optimization prior loss functions (7) and (8) using the AdamW optimizer as implemented in PyTorch with GPU acceleration. To ensure a fair comparison we phrase the classical optimization (1) as the minimization of a "network" that does not receive any input node, but instead only outputs the learnable parameters u to avoid any differences in implementation. As a second baseline, we additionally evaluate the LBFGS [17] optimizer

for the classical approach to exclude a systematic advantage of the specific AdamW method for optimization problems with a deeply nested structure. All formulations and optimizers are run for 1200 iterations (i.e. *full-batch epochs* in machine learning terminology). Moreover, we compare them to the per-pixel autoencoder [30]. In order to have a fair comparison, we changed the optimization algorithm of the per-pixel autoencoder from Adam in [30] to AdamW.

Initialization Descent-based nonconvex optimization methods depend on the selected parameter initialization. As *random parameter initialization* yields very bad results for classical optimization (see Sec. 7.1), we also apply the *physics based initialization* from [29] for AdamW and LBFGS. For the random network initialization, we adopted the method from [13] for the per-pixel autoencoder and for the proposed 3D autoencoder. To verify the robustness of random initialization, each setting that is related to random initialization of the model parameters or to random initialization of the network parameters is run 5 times.

Hyperparameter optimization In order to respect these physical meaning of the THz model parameters (see Sec. 2, we retain the original data scale for training and optimization. However, the large variance of numeric ranges of these parameters leads to a diverging optimal hyperparameter for network training and optimizer. We, therefore, optimize the hyperparameters via a grid search for 4 learning rates from 10^{-3} to 10^0 for all approaches individually using the *MetalPCB* dataset, and the corresponding optimal learning rate (see Tab. 1, column LR) is applied for all datasets.

For the regularization coefficients λ , we empirically maximize the coefficients but not blurring the parameter images based on visual inspection for all optimizers (LBFGS, AdamW and the proposed method) using the shot noise model. Further technical details on the choice of optimizer, initialization and hyperparameter optimization are given in the supplementary material.

7. Evaluation

In this section, we evaluate optimizers and the proposed method according to objective loss function, parameter accuracy, visual quality, and computational requirement.

7.1. Evaluation on Loss

Measurement dataset Table 1 shows the average ℓ^2 -squared loss in (7) using *MetalPCB* and *MetalPCB+AWGN* dataset by optimizers per-pixel autoencoder [30], LBFGS [17] and AdamW [19] to the proposed method.

The proposed method obtains the lowest loss for the measurement *MetalPCB* dataset. For the additional AWGN noise levels, our method is overall the best optimizer, except for a marginal gap to the per-pixel autoencoder at

Average Normalized Loss ($\times 10^{-6}$)							
Optimizer	PPAE		LBFGS		AdamW		Proposed
Initial.	Random	Physics	Random	Physics	Random	Random	
MetalPCB							
Opt. LR	0.001	0.01	0.1	0.001	0.01	0.01	
measured	3372.63	218.02	15465.89	61.32	12677.52	57.56	
MetalPCB+AWGN at PSNR Level							
-20dB	34927.94	39766.81	105352.48	36100.08	49608.79	30871.59	
-10dB	3232.18	10488.53	85814.26	7380.64	21591.09	3271.89	
0dB	408.61	1967.64	63289.90	965.00	18226.71	400.09	
10dB	112.16	240.86	27453.51	135.92	17439.26	111.22	

Table 1: Comparison of average ℓ^2 -squared loss in (7) using *MetalPCB* and *MetalPCB+AWGN* dataset by optimizers Per-pixel autoencoder, LBFGS and AdamW to the proposed 3D autoencoder. The given ℓ^2 -squared loss is normalized by the signal power. The best optimizers (lower is better) are highlighted.

Average Normalized Loss ($\times 10^{-6}$)				
Optimizer	PPAE	LBFGS	AdamW	Proposed
SynthUSAF+AWGN by PSNR level				
Optimal LR	0.001	0.01	0.01	0.01
-20dB	38843.38	38624.05	35838.50	29802.03
-10dB	19326.69	15620.10	9243.03	3058.49
0dB	8107.89	3730.37	3698.46	317.82
10dB	7750.62	1220.81	1190.69	40.82
SynthObj+AWGN at PSNR level				
Optimal LR	0.001	0.001	0.01	0.01
-20dB	30163.70	49725.60	36259.96	29729.65
-10dB	13342.05	13370.01	10202.86	3276.31
0dB	5170.08	7027.83	4032.16	387.28
10dB	2741.44	16733.44	1711.49	106.93

Table 2: Comparison of average ℓ^2 -squared loss in (7) using *SynthUSAF* and *SynthObj* datasets with AWGN.

–10dB PSNR noise level. Moreover, the minimum and maximum of multiple runs are relatively stable for the proposed method, while the per-pixel autoencoder obtains a huge variation in the measurement dataset and at –20dB noise level (see statistics in the supplementary material).

For the initialization, the classical optimizers LBFGS and AdamW obtain very high loss by using random initialization, while the proposed method achieves very good results, *i.e.*, the proposed method is robust to the initialization.

As the physics based initialization for LBFGS and AdamW is clearly a better optimizer than random based initialization, we use physics based initialization for LBFGS and AdamW from here on for the entire evaluation section.

Synthetic dataset with AWGN Table 2 shows that the proposed method achieves the lowest average loss for both synthetic datasets and at all noise levels significantly. The statistics of synthetic datasets with the AWGN noise model can be found in the supplementary material.

Datasets with shot noise model Table 3 shows the average loss with regularization in (8) using *MetalPCB*, *SynthUSAF* and *SynthObj* dataset at 0dB AWGN and 10% shot noise respectively. To make a fair comparison, all op-

Average Normalized Loss ($\times 10^{-6}$) at 0dB PSNR Level			
Optimizer	LBFGS	AdamW	Proposed
Learning Rate	1	0.1	0.01
MetalPCB+ShotNoise	2034.9	1670.4	967.3
SynthUSAF+ShotNoise	10953.1	5036.5	4831.2
SynthObj+ShotNoise	7771.8	4329.7	4271.7

Table 3: Comparison of average loss with regularization in (8) using *MetalPCB*, *SynthUSAF* and *SynthObj* datasets with shot noise model. The learning rate is selected based on optimal learning rate of *MetalPCB+ShotNoise*

timizers use the same set of regularization coefficient λ . Note, that we only include LBFGS, AdamW and the proposed method in this table as the per-pixel autoencoder cannot optimize the loss function with total variation regularization. The additional min/max values are given in the supplementary material.

As we can see, the proposed method obtains the lowest loss among all optimizers. Note that AdamW also achieves a marginally worse, second best optimizer for synthetic dataset, while the proposed method still achieves a significantly lower loss for measurement dataset *MetalPCB*.

Discussion Given the fact that all these optimizers are optimizing the same loss function (7) and (8) respectively, the result shows that the proposed method find lower energy minima by avoiding undesirable local minima stuck in the gradient descent algorithm.

7.2. Evaluation on Parameter Accuracy

In addition, we evaluate classical optimizers, the per-pixel autoencoder and the proposed method according to the parameter accuracy during optimization in (7) and (8). To estimate the model parameter accuracy, we evaluate the Root Mean Square Error (RMSE) of the estimated parameters regarding the ground truth parameters for the synthetic datasets *SynthUSAF* and *SynthObj*. Due to the variation in signal power of different noise levels, the RMSE of the model parameter \hat{e} is normalized by the signal power given by the individual dataset.

7.2.1 Synthetic Dataset with AWGN

Numerical comparison Figure 3 plots the RMSE of model parameters \hat{e} , estimated by per-pixel autoencoder, LBFGS, AdamW and the proposed method using *SynthUSAF* and *SynthObj* dataset respectively. We provide all model parameters plots in the supplementary material.

Overall, the proposed method has the most accurate (lowest RMSE) estimation for all model parameters. The per-pixel autoencoder occasionally achieves more accurate parameters at low noise level (0dB and 10dB), but the error

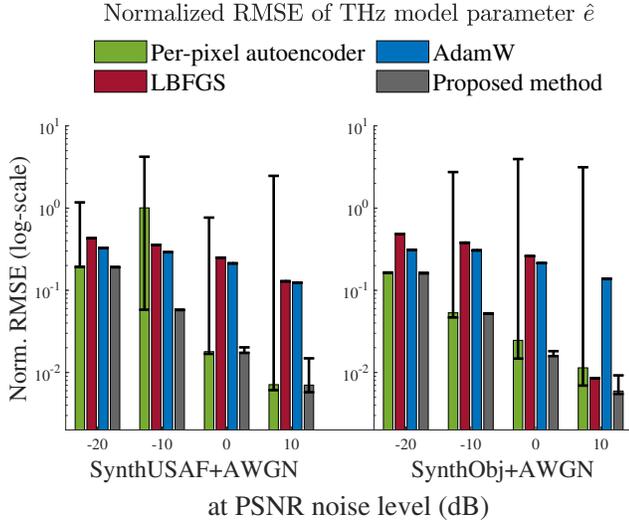


Figure 3: Comparison of RMSE of model parameter \hat{e} using dataset *SynthUSAF* and *SynthObj* at AWGN noise level from -20 to 10 dB. Bars and whiskers indicate minimum, median and maximum RMSE among 5 runs.

bars show its instability. Given all optimizers are optimizing the same loss function, the parameter accuracy gap between all 3 per-pixel optimizers and the proposed method shows that the proposed deep optimization prior approach significantly improves the accuracy of model parameter estimation by overcoming the undesired local minima.

Qualitative comparison To check the quality of the estimated model parameters, Fig. 4 depicts the corresponding model parameter images for \hat{e} and μ (the other model parameter images are given in the supplementary material). The given \hat{e} and μ parameter images are extracted from the run that obtained the median RMSE (*median quality run*) at 0 dB PSNR noise level for the *SynthUSAF* dataset, and the highest RMSE (*worst quality run*) for *SynthObj* dataset.

As we can see from the μ images for *SynthUSAF*, per-pixel autoencoder and the proposed method both estimate generally accurate parameters, while the proposed method has slightly more accurate μ values than the per-pixel autoencoder in the middle of the structure. However, when we compare the μ images of *SynthObj* obtained by the *worst run*, the proposed method performs significantly better than the per-pixel autoencoder and AdamW. This performance difference shows that the per-pixel autoencoder is sensitive to the network initialization, while the proposed method can overcome undesired local minima given the fact that both autoencoders utilize the same initialization scheme [13].

7.2.2 Synthetic Dataset with Shot Noise

Numerical comparison As we can see from Table 4, the proposed method obtains the most accurate parameters

RMSE (Median of 5 runs)			
Optimizer	LBFGS	AdamW	Proposed
SynthUSAF+ShotNoise at 0dB PSNR			
LR	1	0.1	0.01
\hat{e}	0.2125	0.1346	0.1274
μ	0.3606	0.1642	0.1393
σ	0.00877	0.00584	0.00372
ϕ	0.7067	0.4872	0.4572
SynthObj+ShotNoise at 0dB PSNR			
LR	1	0.1	0.01
\hat{e}	0.2233	0.1098	0.1079
μ	0.3083	0.2344	0.3129
σ	0.00495	0.00508	0.00435
ϕ	0.602	0.4029	0.387

Table 4: Comparison of model parameters RMSE using *SynthUSAF* and *SynthObj* datasets with shot noise model.

with a slight margin to AdamW for the *SynthUSAF* dataset, which achieves better accuracy for μ .

Qualitative comparison Fig. 4 (last two columns) shows the median run parameter images \hat{e} and μ for *SynthUSAF+ShotNoise* and *SynthObj+ShotNoise*. (Other model parameter images are shown in the supplementary material.)

By visual comparison of the *SynthUSAF* μ images, we observe that the proposed method removes shot noise significantly, while the AdamW optimizer still retains more shot noise. The visual comparison of the *SynthObj* μ images shows that the proposed method obtains less shot noise than the AdamW optimizer.

Discussion By evaluation on the parameter image quality in Sec. 7.2.1 and Sec. 7.2.2, we can see that the quality of the optimization solution is improved significantly. However, we notice that in rare cases the proposed 3D autoencoder approach generates artefacts in the corner of the single parameter image, such as in the top-left corner of the \hat{e} image for the *SynthUSAF+AWGN* dataset (Fig. 4).

7.3. Timing and Memory

Average Time in seconds				
Optimizer	PPAE	LBFGS	AdamW	Proposed
Optimizer without regularization				
MetalPCB	†5012.0+3.0	1650.3	71.9	225.2
MetalPCB+AWGN	†4339.2+3.1	1835.8	71.6	223.8
Optimizer with regularization				
MetalPCB+ShotNoise	*N/A	5503.5	223.2	367.4
SynthUSAF+ShotNoise	*N/A	5829.8	225.0	371.3
SynthObj+ShotNoise	*N/A	4853.8	225.1	371.4

† This is average training time and prediction time for PPAE

* Not available: PPAE is not available for regularization.

Table 5: Runtime comparison.

Table 5 compares the optimization time for all methods. As the per-pixel autoencoder is trained by the classical

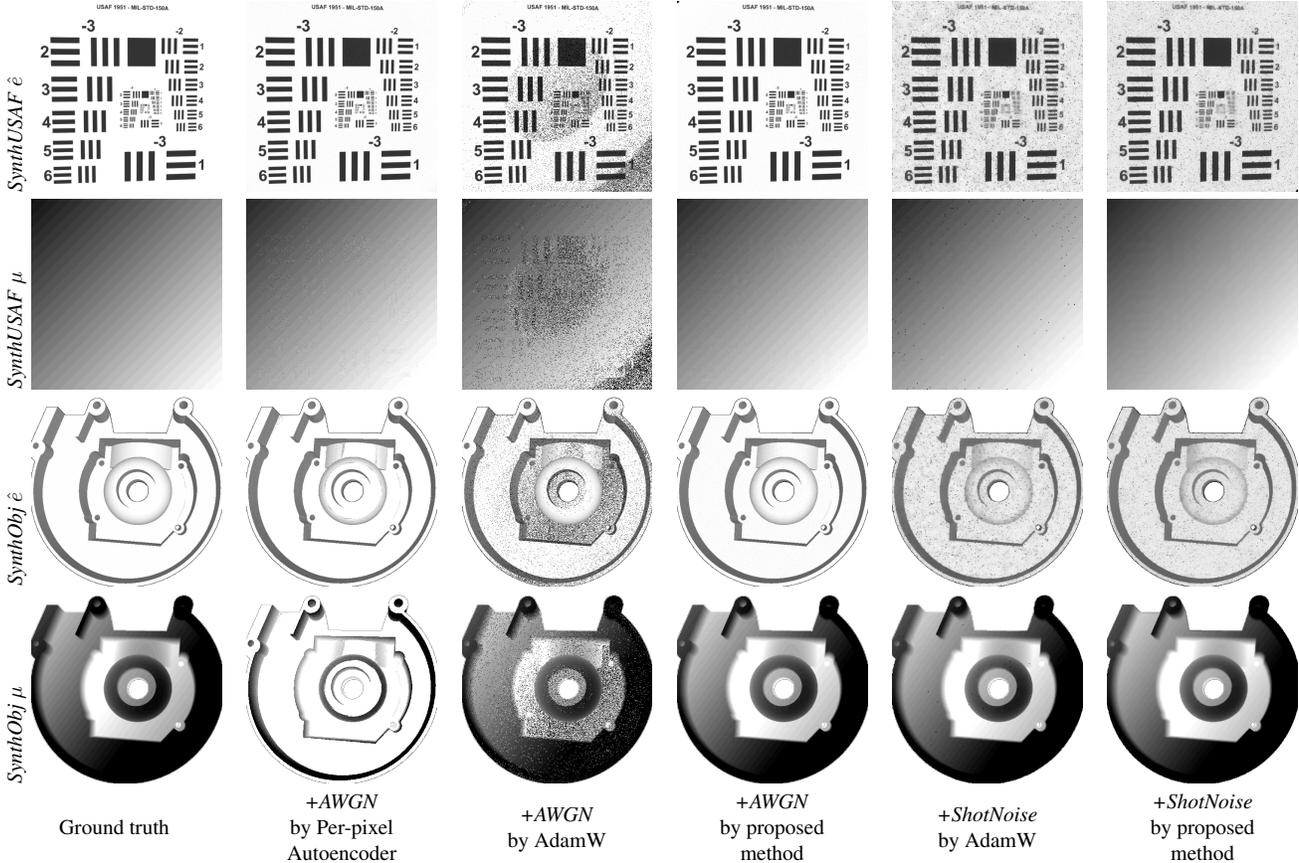


Figure 4: Comparison of model parameters \hat{e} and μ by ground truth and optimizers using *SynthUSAF* and *SynthObjd* datasets with both noise models at $0dB$ PSNR. Note that *SynthUSAF*+AWGN, *SynthUSAF*+ShotNoise and *SynthObjd*+ShotNoise parameter images are selected by the median RMSE among 5 runs respectively (*median quality run*), while *SynthObjd*+AWGN parameter images are selected by the largest RMSE (*worst quality run*).

training-then-prediction approach, the training and prediction times are stated individually. Note that datasets with the shot noise model are optimized by an additional regularization term using (8), which include local dependence.

The proposed method has a drastically improved training time compared to the per-pixel autoencoder, from 1.4 hours to 3.7 minutes, *i.e.*, a factor of 22.2.

However, we can see that AdamW is the fastest optimizer in general. It is faster than the proposed method by a factor of 3.1, when the optimization problem is pixel-wise operation *without* regularization. However, AdamW does out-perform the proposed method only by a factor of 1.6 when the optimization problem includes regularization.

To compare the memory requirement for the per-pixel autoencoder and the proposed method, we record the graphics memory requirement of the network by PyTorch Profiler. For the per-pixel autoencoder, the graphics memory requirement for a 446×446 batch size is $10.53GB$, while the proposed method only requires $2.20GB$. This shows that the proposed method is more efficient in computation and

memory requirement than the per-pixel autoencoder.

8. Conclusion

In this paper, we propose a deep optimization prior approach with application to THz model parameters estimation. Comparing to classical first order (AdamW), second order (LBFGS) optimizers and a single-pixel model-based autoencoder method, the proposed approach shows superior ability to robustly find better local optima in the highly non-convex energy landscape, and it enhances the quality of the desired model parameters. Experiments demonstrate that the proposed deep optimization prior approach robustly reconstructs THz model parameters in low SNR and shot noise situation *without* relying on sophisticated physically motivated initializations. Moreover, the proposed U-net encoder architecture results in a network that is computational and memory-wise highly efficient compared to the state-of-art 1D convolutional neural network structure. Future research will include exploiting more application to other non-convex optimization problems.

References

- [1] Michael E Baginski, Daniel L Faircloth, and Manohar D Deshpande. Comparison of two optimization techniques for the estimation of complex permittivities of multilayered structures using waveguide measurements. *IEEE transactions on microwave theory and techniques*, 53(10):3251–3259, 2005.
- [2] Jan Barowski, Marc Zimmermanns, and Ilona Rolfes. Millimeter-wave characterization of dielectric materials using calibrated fmcw transceivers. *IEEE Transactions on Microwave Theory and Techniques*, 66(8):3683–3689, 2018.
- [3] A Bose, A Kadu, H Mansour, P Wang, P Boufounos, PV Orlik, and M Soltanian. Thz multi-layer imaging via non-linear inverse scattering. In *2019 44th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, pages 1–2. IEEE, 2019.
- [4] Martin Burger, Janic Föcke, Lukas Nickel, Peter Jung, and Sven Augustin. Reconstruction methods in thz single-pixel imaging. In *Compressed Sensing and Its Applications*, pages 263–290. Springer, 2019.
- [5] Wai Lam Chan, Jason Deibel, and Daniel M Mittleman. Imaging with terahertz radiation. *Reports on progress in physics*, 70(8):1325, 2007.
- [6] Ronald Clark, Michael Bloesch, Jan Czarnowski, Stefan Leutenegger, and Andrew J Davison. Learning to solve non-linear least squares for monocular stereo. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 284–299, 2018.
- [7] Thomas F Coleman and Yuying Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on optimization*, 6(2):418–445, 1996.
- [8] Peter De Maagt, Peter Haring Bolivar, and Chris Mann. Terahertz science, engineering and systems—from space to earth applications. *Encyclopedia of RF and Microwave Engineering*, 2005.
- [9] Jinshan Ding, Matthias Kahl, Otmar Loffeld, and Peter Haring Bolívar. Thz 3-d image formation using sar techniques: simulation, processing and experimental results. *IEEE Transactions on Terahertz Science and Technology*, 3(5):606–616, 2013.
- [10] Timothy D Dorney, Richard G Baraniuk, and Daniel M Mittleman. Material parameter estimation with terahertz time-domain spectroscopy. *JOSA A*, 18(7):1562–1571, 2001.
- [11] Bertram Drost, Markus Ulrich, Paul Bergmann, Philipp Hartinger, and Carsten Steger. Introducing mvtec itodd-a dataset for 3d object recognition in industry. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 2200–2208, 2017.
- [12] Fabian Friederich, Wolff Von Spiegel, Maris Bauer, Fanzhen Meng, Mark D Thomson, Sebastian Boppel, Alvydas Lisauskas, Bernd Hils, Viktor Krozer, Andreas Keil, et al. Thz active imaging systems with real-time capabilities. *IEEE Transactions on Terahertz Science and Technology*, 1(1):183–200, 2011.
- [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.
- [14] Weidong Hu, Yade Li, Zhongde Han, Zhihao Xu, Yunzhang Zhao, and Jiaqi Ni. High-resolution range profile reconstruction method for terahertz fmcw radar. *Applied Optics*, 60(22):6400–6408, 2021.
- [15] Christian Jansen, Steffen Wietzke, Ole Peters, Maik Scheller, Nico Vieweg, Mohammed Salhi, Norman Krumbholz, Christian Jördens, Thomas Hochrein, and Martin Koch. Terahertz imaging: applications and perspectives. *Appl. Opt.*, 49(19):E48–E57, 2010.
- [16] Yade Li, Weidong Hu, Xin Zhang, Zhihao Xu, Jiaqi Ni, and Leo P Ligthart. Adaptive terahertz image super-resolution with adjustable convolutional neural network. *Optics Express*, 28(15):22200–22217, 2020.
- [17] Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.
- [18] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [19] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2018.
- [20] Qi Mao, Yunlong Zhu, Cixing Lv, Yao Lu, Xiaohui Yan, Shihan Yan, and Jingbo Liu. Convolutional neural network model based on terahertz imaging for integrated circuit defect detections. *Optics express*, 28(4):5000–5012, 2020.
- [21] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.
- [22] Daniel M Mittleman, Rune H Jacobsen, and Martin C Nuss. T-ray imaging. *IEEE Journal of selected topics in quantum electronics*, 2(3):679–692, 1996.
- [23] Maria E Requena-Pérez, Antonio Alberó-Ortiz, Juan Monzó-Cabrera, and Alejandro Díaz-Morcillo. Combined use of genetic algorithms and gradient descent optimization methods for accurate inverse permittivity measurement. *IEEE Transactions on Microwave Theory and Techniques*, 54(2):615–624, 2006.
- [24] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [25] Nina S Schreiner, Michael Bortz, Wolfgang Sauer-Greff, Ralph Urbansky, and Fabian Friederich. Optimization algorithms for accurate fmcw millimeter wave and terahertz thickness measurements. In *2020 50th European Microwave Conference (EuMC)*, pages 623–626. IEEE, 2021.
- [26] Peter H Siegel. Terahertz technology. *IEEE Transactions on microwave theory and techniques*, 50(3):910–928, 2002.
- [27] Weng-tai Su, Ta-Hsuan Chao, Shang-Hua Yang, and Chia-Wen Lin. Seeing through a black box: Toward high-quality terahertz tomographic imaging via multi-scale spatio-spectral image fusion. *arXiv preprint arXiv:2103.16932*, 2021.

- [28] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018.
- [29] Tak Ming Wong, Matthias Kahl, Peter Haring Bolívar, and Andreas Kolb. Computational image enhancement for frequency modulated continuous wave (fmcw) thz image. *Journal of Infrared, Millimeter, and Terahertz Waves*, 40(7):775–800, 2019.
- [30] Tak Ming Wong, Matthias Kahl, Peter Haring Bolívar, Andreas Kolb, and Michael Möller. Training auto-encoder-based optimizers for terahertz image reconstruction. In *German Conference on Pattern Recognition*, pages 93–106. Springer, 2019.