

Real-time Terahertz Compressive Optical-Digital Neural Network Imaging

**SHAO-HSUAN WU,¹ SEYED MOSTAFA LATIFI,^{1,2} CHIA-WEN LIN,^{3,4,5}
AND SHANG-HUA YANG^{1,4,5,*}**

¹*Institute of Electronics Engineering, National Tsing Hua University, Hsinchu, 30013, Taiwan*

²*School Engineering, University of Liverpool, Liverpool, L69 7ZL, UK*

³*The Institute of Communications Engineering, NTHU, Hsinchu, Taiwan*

⁴*Department of Electrical Engineering, National Tsing Hua University, Hsinchu, 30013, Taiwan*

⁵*Terahertz Optics and Photonics Center, National Tsing Hua University, Hsinchu, 30013, Taiwan*

*shanghua@ee.nthu.edu.tw

Abstract: Terahertz (THz) waves offer unique advantages, including the ability to penetrate non-metallic and non-polarized materials while maintaining bio-safety. However, current THz imaging systems face significant challenges due to hardware constraints, leading to information loss during the conversion of THz waves into digital data. To address these challenges, we present a hybrid optical-digital neural network for THz computational imaging. This approach employs a physical encoder, optical neural network (ONN), to transform and reduce the dimensionality of physical signals, effectively compressing them to accommodate the limitations of THz sensor arrays. Once the compressed signals are captured and converted into digital data by the THz sensor array, a jointly trained digital neural network (DNN) reconstructs the signals into desired or original form. In our experiments, we optimized the parameters of both the ONN and the DNN. The proposed THz computational imaging system demonstrated enhanced imaging quality, an imaging area four times larger than the physical sensor array, and diffraction-free imaging. Furthermore, we achieved real-time imaging at a rate of 2 frames per second.

1. Introduction

Over the past decades, terahertz (THz) technology has rapidly advanced due to its unique properties, such as non-ionizing, non-destructive [1–3], offering a chemically informative spectrum [4], and its ability to penetrate non-metallic and non-polar materials [5]. These characteristics make THz technology particularly well-suited for non-invasive imaging applications, facilitating its adoption in various fields, including biomedical imaging [6–8], quantum sensing and imaging [9], chemical identification [10, 11], and circuit inspection [12]. Consequently, THz imaging has become increasingly prominent in sectors like bio-informatics, industrial inspection, and security screening. With the growing demand for THz imaging, the development of real-time, high-throughput imaging systems is crucial. Real-time THz imaging not only enhances the practicality of current applications but also enables the capture and analysis of dynamic, time-sensitive data, unlocking new possibilities across diverse fields.

The initial approach to THz imaging used a 2-axis mechanical stage for raster scanning, allowing signal collection from the region of interest [13]. Although this method delivers high-quality THz images with wavelength-scale spatial resolution [14], the data acquisition process is time-consuming, as the duration increases linearly with image size. For instance, scanning a centimeter-scale object can take minutes to hours, with higher spatial resolutions extending the time required even further. To achieve practical imaging speeds in real-world applications, focal-plane array (FPA) sensors have been proposed, integrating multiple THz sensors into 1D or 2D array formats. Common types of FPA sensors for THz imaging include time-resolved sensor arrays, field-effect transistor (FET) arrays, and microbolometer arrays. Time-resolved sensor arrays typically consist of THz photoconductive antennas (PCAs) or electro-optic (EO) sensors paired with a pulsed THz emitter and a femtosecond laser in a

47 pump-probe measurement setup, enabling real-time imaging [15, 16]. However, those approaches
48 requires a high-power, bulky femtosecond laser and mechanical delay stages, which increase
49 the system's size, cost, and complexity, limiting its practical applicability. THz FET array
50 sensors couple the incident THz waves with plasma waves between source and drain, producing
51 high-speed currents proportional to THz signal strength [17–19]. These sensors feature a
52 compact design and operate at room temperature. However, scaling THz FETs to large arrays
53 continues to present significant fabrication challenges, and the limited space-bandwidth product
54 remains a longstanding issue [20]. The microbolometer FPA, offering a higher space-bandwidth
55 product, consists of a bolometric sensor array that absorbs incoming THz energy and converts it
56 into the current signals [21]. The configuration of microbolometer FPA can be scalable, with
57 commercial arrays achieving 388×288 pixels [22], enabling laser-free THz imaging at tens of
58 frames per second. However, microbolometer FPA faces a significant challenge: they are very
59 sensitive to diffraction-induced artifacts. These artifacts inherently arise from wave propagation
60 and can lead to substantial information loss. Additionally, the degree of diffraction-induced
61 artifacts is significantly influenced by the relationship between the wavelength and the size of
62 the geometrical features of the tested object, particularly $\frac{\text{aperture size}}{\lambda} \approx 1$, leading to more
63 pronounced diffraction patterns in these conditions. Consequently, real-time THz imaging
64 systems based on FPAs face considerable challenges in reconstructing interior information of
65 objects at the millimeter scale.

66 To balance imaging speed, image quality, and information reconstruction, several computational
67 THz imaging approaches have been proposed [23, 24]. THz compressed sensing (CS) has emerged
68 as a powerful computational imaging technique that overcomes the Nyquist sampling limit [25]
69 and significantly reduces data acquisition time compared to raster scanning methods. To date, a
70 6-fps THz CS video with a single-pixel detector has been demonstrated [26]. However, high-speed
71 THz CS using spatial light modulators faces a trade-off between speed, resolution, and field of
72 view, as increasing the imaging area reduces resolution due to fixed pixel counts. Recently, Lin
73 et al. proposed a physical computational imaging approach utilizing an Optical Neural Network
74 (ONN) constructed from diffractive optical elements to process information while the optical
75 signal carrier propagates [27]. The ONN can implement high-dimensional correlated functions
76 between input and output optical wavefronts and perform certain degrees of computation, thereby
77 alleviating significant computational burden on the computer side [28]. However, ONN-based
78 THz imaging systems heavily depend on the sensing capabilities of THz sensors [27, 29–31],
79 which limits their performance to the specifications of the sensors and the format of the captured
80 data.

81 Here, We propose a compressive THz computational imaging framework that combines a
82 hybrid optical-digital neural network, leveraging the strengths of ONNs for light-speed processing,
83 THz FPAs for high-speed data acquisition, and DNNs for signal decompression and information
84 retrieval. In this system, the ONN transforms and encodes multidimensional signals into lower
85 dimensions, reducing hardware complexity and the volume of data collected. The DNN then
86 decompresses these low-dimensional signals into higher-dimensional data, extracting object
87 information. This hybrid imaging system offers several benefits: (1) Video-rate imaging:
88 Utilizing light-speed processing and 10s-fps data capturing capabilities. (2) Reduced hardware
89 constraints: Minimizing dependence on hardware specifications through signal transformation. (3)
90 Diffraction-free imaging: Using DNNs to reconstruct signals into high-dimensional information.
91 By transforming large data volumes into compact tokens, the system ensures efficient data
92 transfer, acquisition, and processing, allowing low-pixel-number THz sensor arrays to capture
93 multi-scale information from THz signals. This methodology can adapt existing computational
94 imaging modalities and 3D-printed THz device designs, making it suitable for various real-world
95 applications in industrial inspection, security screening, material development, bioinformatics,
96 and space exploration.

97 **2. Principle**

98 In this section, we introduce the working principle of physical-digital signal dimensional
99 transformation. The ONN serves as the encoder, transforming multi-scale information into a lower-
100 dimensional space in the desired data format. The DNN functions as the decoder, decompressing
101 intensity-based THz signals into a higher dimension and reconstructing the object information.
102 Next, we will next illustrate the methodology for designing and identifying the encoder and
103 decoder functions. This process involves balancing multiple objectives, including dimensionality
104 reduction, the preservation of meaningful latent representations, accurate reconstruction, and
105 robust generalization. Additionally, ensuring the encoder and decoder can work seamlessly in
106 different environments, physical and digital. Following this, we will demonstrate the training
107 steps for our proposed hybrid neural network THz computational imaging system.

108 *2.1. Optical-digital data compression and decompression*

109 In our proposed imaging framework, the core concept is signal dimension transformation. Here,
110 "signal dimension" refers to the types of signals carried by the propagated electromagnetic wave
111 (e.g., amplitude, phase, polarization), rather than the numerical structure of the data. In Fig. 1(a),
112 a conventional imaging system is depicted, where the dimensionality of the displayed image is
113 determined by the detection capabilities of the sensor. For example, if bolometric sensors are
114 used, only the amplitude dimension can be recorded, while the phase, polarization, and frequency
115 dimensions may be lost. This loss leads to a degradation in the quality and accuracy of the
116 reconstructed information. In Fig. 1(b), a computational imaging system is illustrated. Through
117 multiple sensor modules or multifunctional sensors, this system can capture a broader range of
118 signal dimensions. For instance, when using THz PCA as the sensor, it record amplitude, phase,
119 and polarization dimensions across time, space, and frequency domains. The object information
120 is then extracted using physics-guided computational methodologies. However, capturing more
121 data dimensions results in a rapid increase in data size and complexity, significantly complicating
122 the inverse problem of extracting object information. As a result, improvements in image
123 quality are often achieved at the cost of heavy computational load and extended processing times.
124 In Fig. 1(c), we represent the architecture of our proposed imaging system. A physical encoder is
125 designed to preprocess the wave characteristics—phase, amplitude, and polarization—across
126 time, space, and frequency domains, enabling the construction of a transformation function,

$$f_e : x \in \mathbb{R}^N \rightarrow \mathbb{R}^K.$$

127 This transformation compresses the high-dimensional signal (N) into a lower-dimensional
128 representation (K), aligning it with the conversion limitations of the sensor while retaining
129 essential information. The information obtained by the sensor converts signals into digital forms,
130 and this captured data is referred to as compressed imaging. Mathematically, this is represented
131 in a latent space (z). A decoder is then used to decompress captured digital data (z) to reconstruct
132 data that represent and display the original high-dimensional (N) information, which can be
133 expressed as:

$$f_d : z \in \mathbb{R}^K \rightarrow \mathbb{R}^N.$$

134 Through this physical-digital signal transformation, our framework overcomes the limitations
135 of conventional sensors, minimizing information loss during the conversion into digital data
136 and enabling higher-dimensional data representation. In the following subsections, we will
137 provide further details on determining suitable encoder and decoder functions, along with their
138 mathematical models, within the context of the proposed imaging framework.

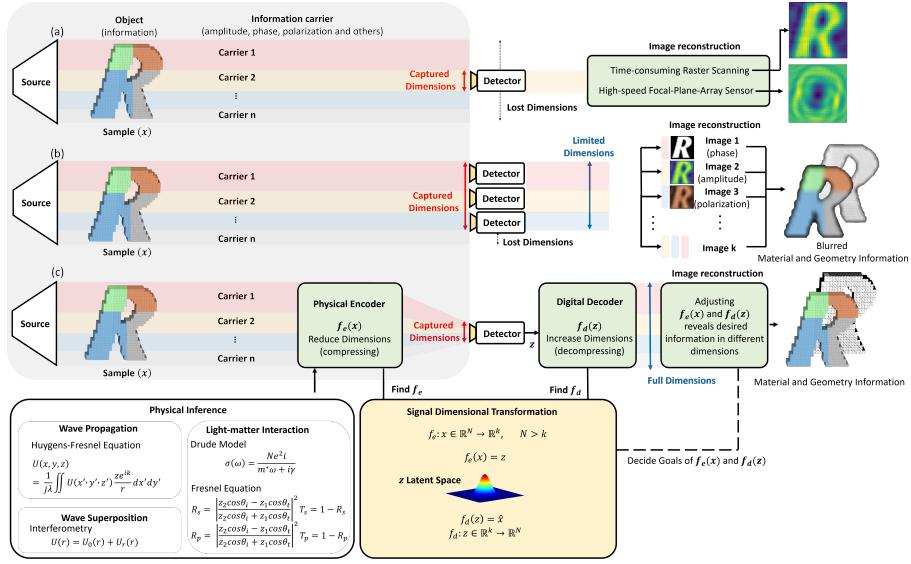


Fig. 1. (a) A conventional imaging system that utilizes a sensor module to capture specific information from the optical carrier wave. (b) A computational imaging system that employs multiple sensor modules or multifunctional sensors to capture a broader range of signal dimensions. Object information is then extracted using physics-guided computational methodologies. (c) A hybrid optical-digital imaging system incorporating a physical encoder and a digital decoder. The physical encoder compresses the complete object information from full dimensions to a lower dimension, while the digital decoder decompresses this information back to full dimensions and reconstructs the object information.

139 2.2. Hybrid optical-digital neural network: Design

140 2.2.1. Optical neural network

141 An ONN is used to manipulate the optical beam characteristics spatially, temporally, or spectrally
 142 to perform physical computations. As an optical processor, ONN is designed to control
 143 the amplitude, phase, or polarization of the optical beam through wave propagation, wave
 144 superposition, and light-matter interactions at the speed of light. This primarily aims to achieve
 145 the encoded wavefront profiles along the propagation path.

146 To calculate the interference and propagation of light, the Angular Spectrum Method (ASM)
 147 is employed. ASM models the optical field as a superposition of infinitely many point sources,
 148 each radiating outward according to the principles of optical field propagation. This approach
 149 allows for the calculation of the optical field distribution after propagation, expressed as:

$$E(x, y, z) = \iint_{-\infty}^{\infty} \hat{E}(k_x, k_y; 0) e^{i[k_x x + k_y y \pm k_z z]} dk_x dk_y \quad (1)$$

150 where z is chosen as the propagation axis, and $E(x, y, z)$ denotes the optical field at a specific
 151 point in space, represented by the coordinates (x, y, z) . The term $\hat{E}(k_x, k_y; 0)$ corresponds to
 152 the two-dimensional Fourier transform of the complex optical field $E(x, y, z)$ at the initial plane
 153 $z = 0$, and can be expressed as:

$$\begin{aligned} \hat{E}(k_x, k_y; z) &= \mathcal{F}\{E(x, y, z)\} \\ &= \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} E(x, y, z) e^{i[k_x x + k_y y \pm k_z z]} dk_x dk_y \end{aligned} \quad (2)$$

154 This Fourier transform describes the optical field in the spatial frequency domain, where k_x and
 155 k_y represent the spatial frequencies along the x - and y - axes, respectively, and k_z captures the
 156 propagation of wave along the z -axis. The inverse Fourier transform, used to reconstruct the
 157 optical field from its angular spectrum, is expressed as:

$$E(k_x, k_y; z) = \iint_{-\infty}^{\infty} \hat{E}(k_x, k_y, k_z) e^{i[k_x x + k_y y]} dk_x dk_y \quad (3)$$

158 Thus, the angular spectrum \hat{E} after propagating a distance z can be related to its initial form at
 159 $z = 0$ by:

$$\hat{E}(k_x, k_y; z) = \hat{E}(k_x, k_y; 0) e^{\pm k_z z} \quad (4)$$

160 Here, k_z is the longitudinal wavenumber, and the exponential term $e^{\pm i k_z z}$ accounts for the phase
 161 evolution of the wave during propagation along the z -axis. The relationship between k_x , k_y ,
 162 and k_z is governed by the Helmholtz equation, $(\nabla^2 + k^2)E(x, y, z) = 0$, where k is the total
 163 wavenumber defined as $k = (\omega/c)n$, with ω being the angular frequency, c the speed of light,
 164 and n the refractive index of the medium. The longitudinal wavenumber k_z can be expressed as:

$$k_z \equiv \sqrt{k^2 - k_x^2 - k_y^2} \quad \text{with } \text{Im}\{k_z\} \geq 0 \quad (5)$$

165 If k_z becomes purely imaginary, the wave is evanescent, decaying exponentially along the z -axis
 166 while propagating in the transverse directions.

167 Utilizing Eq. (1) to Eq. (5), the optical field after passing through an ONN can be calculated.
 168 The ONN utilized in our work comprises several diffractive layers, each layer characterized by a
 169 unique transmittance function $P_L(x, y)$, which governs the modulation of the optical field at the
 170 L -th layer. This transmittance function can be expressed as:

$$P_L(x, y) = \exp [ik(n(\lambda) + i\kappa(\lambda) - n_r)T_L(x, y)], \quad (6)$$

171 where $n(\lambda)$ and $\kappa(\lambda)$ are the real and imaginary parts of the refractive index of the ONN medium,
 172 n_r is the refractive index of the surrounding medium, and $T_L(x, y)$ is the thickness profile of
 173 the layer. These thicknesses are optimized using deep learning to achieve the desired optical
 174 functionality. The total optical field after propagating through m diffractive layers is given by:

$$U_m(x, y) = \mathcal{F}^{-1} \left\{ \mathcal{F}\{U_0(x, y)\} \prod_{L=1}^m [P_L(x, y) e^{ik_z z_{gap}}] \right\}. \quad (7)$$

175 Here, $U_0(x, y)$ is the initial optical field, $P_L(x, y)$ represents the transmittance of the L -th layer,
 176 and z_{gap} is the distance between layers. The field at the final plane, $U_f(x, y)$, is detected and
 177 represents the output of the ONN. For short, ONNs utilize carefully designed diffractive layers
 178 and the principles of light propagation to perform complex optical computations, all at the speed
 179 of light.

180 2.2.2. Finding Co-design encoder and encoder

181 To determine the optimal encoding ($f_e(\cdot)$) and decoding ($f_d(\cdot)$) functions for our proposed
 182 computational imaging system, we employed the training methodology of variational autoencoders
 183 (VAEs). VAEs have emerged as one of the most prominent approaches in unsupervised
 184 learning for modeling complex, high-dimensional data distributions. As part of the family of
 185 probabilistic graphical models, VAEs employ variational Bayesian methods to learn latent variable
 186 representations of data distributions [32]. VAEs are highly suited for real-time THz imaging due
 187 to their ability to efficiently process multi-dimensional data. VAEs compress complex spatial,
 188 spectral, and temporal information into a compact latent space, enabling high-quality image

189 reconstruction from sparse measurements, reducing the need for high-dimensions sensors. Their
 190 probabilistic framework enhances robustness against noise and uncertainty, ensuring accurate
 191 object analysis while simplifying hardware requirements by compensating for missing data
 192 computationally. Compared to other models, explicit probabilistic of VAEs modeling makes them
 193 more reliable and interpretable, addressing the key challenges of THz imaging with efficiency
 194 and precision.

195 The fundamental goal of a VAE is to maximize the marginal likelihood of the observed data x ,
 196 denoted as $p_\theta(x)$. This marginal likelihood is expressed as:

$$p_\theta(x) = \int p_\theta(x|z)p(z) dz \quad (8)$$

197 In this equation, z represents the latent variable generated by the encoder, and $p(z)$ denotes the
 198 prior distribution over the latent variable. The term $p_\theta(x|z)$ represents the conditional distribution
 199 of the observed data x given the latent variable z , parameterized by θ , which corresponds to the
 200 parameters of decoder. Significantly, the dimensionality of z is designed to be smaller than that
 201 of x , enabling a compact representation that aids in efficient processing and storage. Since the
 202 marginal likelihood in Eq. (8) is intractable to compute directly due to the integral over the latent
 203 variable z , VAEs utilize an approximate posterior distribution, denoted $q_\phi(z|x)$, to approximate
 204 the true posterior $p_\theta(z|x)$. This variational approximation enables efficient training of VAEs
 205 using stochastic optimization techniques.

206 Training a VAE involves minimizing a total loss function that combines two terms: the
 207 reconstruction error and the Kullback-Leibler (KL) divergence between the approximate posterior
 208 $q_\phi(z|x)$ and the prior $p(z)$. The total loss function is defined as:

$$L(\theta, \phi; x) = -\mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x|z)] + \text{KL}(q_\phi(z|x) \| p(z)) \quad (9)$$

209 In this equation, the first term, $-\mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x|z)]$, represents the reconstruction error, which
 210 measures how well the decoder can reconstruct the input data x from the latent variable z .
 211 The second term, $\text{KL}(q_\phi(z|x) \| p(z))$, enforces regularization by ensuring that the approximate
 212 posterior remains close to the prior distribution, preventing overfitting and ensuring that the
 213 latent space adheres to the structure assumed by the prior. During each back propagation step,
 214 the parameters θ (decoder parameters) and ϕ (encoder parameters) are updated to minimize this
 215 total loss, and this optimization can be expressed as:

$$\begin{aligned} \theta^*, \phi^* &= \arg \min_{\theta, \phi} L(\theta, \phi; x), \\ \theta &\leftarrow \theta - \eta \frac{\partial L(\theta, \phi; x)}{\partial \theta}, \quad \phi \leftarrow \phi - \eta \frac{\partial L(\theta, \phi; x)}{\partial \phi} \end{aligned} \quad (10)$$

216 where θ^* and ϕ^* are the optimal weights of the encoder and decoder, achieving the lowest loss. The
 217 partial derivatives of θ^* and ϕ^* are mutually dependent, and thus the model trains both parameters
 218 jointly, highlighting the joint training process of the VAE. Through this optimization process, the
 219 model learns a latent space representation of the data that balances accurate reconstruction with
 220 a smooth, regularized latent space.

221 2.3. Hybrid optical-digital neural network: Implementation

222 In Section 2.2.2, introduce a mathematical training-process methodology to simultaneously train
 223 the optical and DNN as the encoder and decoder, respectively, ensuring they can compress
 224 and decompress data while preserving the same distribution of the original information. To
 225 facilitate this training, an optical-digital simulation environment was developed. This environment
 226 incorporates the theoretical framework described by ASM for modeling beam propagation and
 227 ONN operations, which account for interactions among diffractive layers, phase modulation, and

absorption effects. The simulation environment was implemented using Python (v3.8.18), with a spatial resolution defined by pixel dimensions equivalent to half the operational wavelength. This ensures subwavelength accuracy, enabling precise modeling of the intricate THz beam interactions and facilitating reliable training of the network under realistic conditions.

An overview of the hybrid optical-digital neural network within this simulation framework is provided in Fig. 2. The input to the system is randomly selected from either the training or testing dataset and consists of grayscale images of size 32×32 pixels. To mitigate edge effects and improve the stability of the propagation model, the input data is zero-padded with 48 pixels along all sides, resulting in a total input size of 128×128 pixels. ONN comprises several diffractive layers, each characterized by trainable parameters represented as 128×128 -pixel weight matrices. These layers operate at a resolution of half the wavelength per pixel to ensure fine-grained control of wavefront modulation. The primary role of the ONN is to encode and compress the input signal into the central portion of the resultant beam. This compression aligns the output signal with the effective detection area of the real-world THz FPA sensors. The THz sensor, designed to function within a physical sensor area of 12×10 mm, corresponds to an effective array size of approximately 16×13 pixels. By leveraging the spatial compression achieved by the diffractive layers, the system efficiently captures essential signal information while minimizing redundancy. This approach reduces the computational overhead of the hybrid architecture and optimizes the trade-off between data processing demands and system performance. This simulation-based training framework serves as a robust platform for advancing the capabilities of hybrid optical-digital neural networks.

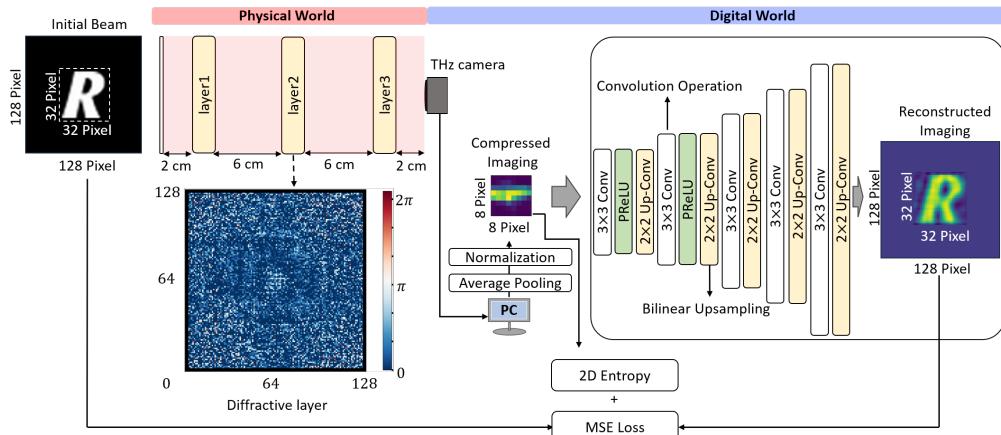


Fig. 2. The architecture of a hybrid optical-digital neural network imaging system is illustrated. In the physical domain, the angular spectrum method is employed to model beam propagation along the transmittance axis. The setup parameters include a distance of 2 cm (13.334 wavelengths) between the positions of the tested objects and the first diffractive layer, a spacing of 6 cm (40 wavelengths) between each diffractive layer, and a 2 cm gap between the third diffractive layer and the THz sensor array. The THz sensor array captures the compressed optical signal and converts it to digital data for processing in the digital domain. A digital neural network is then utilized to decode the captured data and reconstruct the object information. The optimization of the optical and digital neural network weights is achieved through the 2D entropy of the compressed imaging and the mean squared error (MSE) loss between the input (initial beam) and output (reconstructed image), which helps calculate the error value.

As the THz sensor array converts THz waves into electric signals, these signals can be digitally processed. Once converted, average pooling is applied to downsample the image to an 8×8 pixel

resolution. Following downsampling, the image is normalized to a range of 0 to 1. We refer to this downsampled, normalized image as the compressed image. The compressed image is then input into a DNN to decode and reconstruct the original image. This DNN comprises convolutional layers, bilinear upsampling layers and activation functions to achieve nonlinear transformations. In the loss function, mean square error (MSE) is used to measure the discrepancy between the central 32×32 pixels of the original image and those of the reconstructed image. Additionally, to avoid extreme sparsity that would impose excessively high alignment accuracy requirements, 2D entropy is employed to concentrate the signal within the compressed image. The total loss function is therefore defined as the sum of MSE and 2D entropy. After calculating the error value, the Adam optimizer, with a learning rate of 0.0001, is used to compute the gradients of the weights and update them accordingly. The model is trained over 4000 epochs with a batch size of 8 and a weight decay of 0.1.

3. Experiment and Results

This study involves two critical stages, validation in a virtual environment and implementation in the real world. In the first stage, the training process is conducted in a simulated, idealized environment where the ONN and DNN are jointly trained. This joint training process focuses on achieving an effective latent representation and accurate reconstruction under controlled conditions. Several significant parameters within the ONN and DNN need to be optimized during this stage. The second stage transitions from the simulated environment to real-world implementation. Unlike the idealized simulation, the real-world experiment introduces additional challenges, such as the non-ideal characteristics of the ONN devices, the non-ideal performance of the THz sensor array, and the imperfect alignment of THz waves in practical scenarios. These factors can significantly impact the performance of system, and in some cases, even produce detrimental effects. To address these challenges, this study systematically investigated optimal system configurations, implemented strategies to mitigate the uncertainties of the THz sensor array, and successfully demonstrated the feasibility and effectiveness of the proposed hybrid optical-digital neural network computational imaging system under practical conditions.

3.1. Results of simulated hybrid optical-digital neural network

In Section 2.3, we introduced the framework of the proposed hybrid optical-digital neural network imaging system, and in Section 2.2, we detailed the optimization method for the model. For the training process, a dataset comprising 6,656 uppercase alphabet letter images (A-Z), with equal samples per category, was utilized, obtained from Kaggle [33]. Each image was preprocessed and standardized to 32×32 pixels. To evaluate the generalization ability of model, a separate dataset of lowercase letters (26 categories, 8 samples per category) was used, providing insights into the performance of model on previously unseen data.

While Section 2.3 outlined the some parameters of model, such as the learning rate and optimization configurations, these parameters are typically effective across most scenarios and do not significantly impact training performance in a specialized context, which means that we can use the optimal value from other work of field. However, other parameters play a critical role in influencing the training performance and overall efficiency of the model. To achieve optimal performance, we systematically trained the model under different configurations, aiming to identify the best combination of parameters. The key factors explored in these experiments included: (1) The number of diffractive layers: varying the number of layers to determine their impact on encoding and compression efficiency in the hybrid system. (2) The number of convolutional layers in DNN: assessing how the depth of the DNN affects its capacity to generalize and complement the optical components. (3) Channel number of convolutional layers in DNN: exploring the impact of channel size on feature extraction and representation capacity. (4) Activation functions in the DNN: examining the effect of incorporating activation functions

versus using linear transformations only. By experimenting with these parameters, we aimed to identify the optimal configuration that maximized model performance while maintaining computational efficiency. The results from these simulations provide critical insights into the trade-offs and synergies between different architectural components in the hybrid neural network.

In Fig. 3, the results of these experiments are summarized. Fig. 3(a) illustrates the test performance with varying numbers of diffractive layers in the ONN, ranging from 1 to 5 layers. Beyond 5 layers, increased absorption degrades the system signal-to-noise ratio (SNR), leading to a decrease in the quality of compressed images captured by the THz sensor array. Therefore, we do not extend beyond this point. The results indicate that the performance across 3-5 configurations is relatively close, with the configurations using 4 and 5 diffractive layers achieving slightly better outcomes. However, the improvement is not substantial. Additionally, considering the potential challenges posed by increased absorption and the alignment complexity in real-world experiments with additional diffractive layers, we opted for a configuration with 3 diffractive layers to balance performance and practical implementation feasibility. Next, we evaluated the impact of varying the number of convolutional layers in the DNN. As shown in Fig. 3(b), increasing the number of convolutional layers generally improves performance, as deeper models can approximate more complex functions and relationships. While configurations with 15 convolutional layers show a slight improvement compared to those with 10 layers, it shows a trend toward overfitting after around 2000 training epochs. Additionally, deeper networks increase computational time, with each additional layer adding around 0.03 seconds per operation. Since the system is intended for real-time imaging, a balance between performance and efficiency is crucial. Based on these observations, we selected a configuration with 3 diffractive layers and 10 convolutional layers. Furthermore, we explored the effect of varying the number of channels in each convolutional layer. The results, shown in Fig. 3(c), demonstrate that increasing the number of channels improves performance. This improvement is attributed to the augmented capacity of the model to capture and represent a broader and more diverse range of features from the input data. However, using 64 channels exhibits a tendency toward overfitting and does not yield a significant performance improvement compared to 32 channels. Therefore, we chose to use 32 channels for the convolutional layers. To evaluate the performance of nonlinear and linear models, Fig. 3(d) compares the results of a nonlinear model (with an activation function, AC) and a linear model (without AC). The nonlinear model demonstrates significantly superior performance compared to the linear model.

Therefore, The final configuration of the hybrid neural network includes 3 diffractive layers in the ONN, 10 convolutional layers in the DNN, 32 channels per convolutional layer, and activation functions in the DNN. This setup achieves a balance between robust performance and computational efficiency, making it suitable for real-time imaging applications. Fig. 3(e) showcases the reconstructed test images generated by the optimized hybrid optical-digital neural network with the aforementioned parameter setting. The figure includes 26 test images of lowercase alphabet letters alongside their reconstructed counterparts. The model achieved an average MSE of 21.5 and an average peak signal-to-noise ratio (PSNR) of 15.4, demonstrating its effectiveness in generalizing to previously unseen data and validating its suitability for practical applications.

3.2. Experimental results of hybrid optical-digital neural network

In the hybrid optical-digital neural network THz imaging system, the configuration, as illustrated in Fig. 4(a), integrates components from right to left in the schematic. The THz continuous-wave source comprises a microwave signal generator (model SMB100A, Rohde & Schwarz), coupled with a frequency multiplier (model WR5.1-VNAX, Virginia Diodes) and equipped with a conical horn antenna (model WR-5.1, Virginia Diodes) to deliver a highly directional THz beam with a 10 dBm emitting power. This arrangement produces coherent radiation at 0.2 THz by setting the

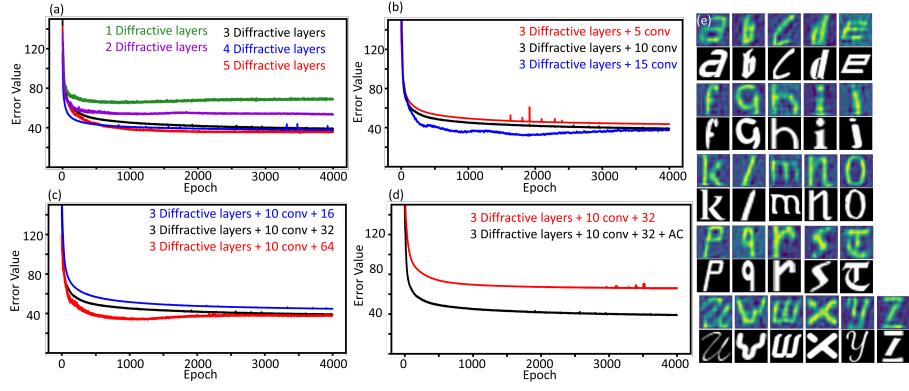


Fig. 3. (a) Testing error values across varying numbers of diffractive layers (1, 2, 3, 4, and 5) in ONN. (b) Comparison of error values for different numbers of convolutional layers (5, 10, and 15) in the model with 3 diffractive layers. (c) Effect of increasing the number of channels per convolutional layer (16, 32, and 64) in the model with 3 diffractive layers and 10 convolutional layers. (d) Impact of including an activation function (AC) versus no activation function in the model with 3 diffractive layers, 10 convolutional layers, and 32 channels per layer. (e) Simulation results of the model with 3 diffractive layers in the ONN, 10 convolutional layers, and 32 channels per layer, including an activation function.

348 microwave signal generator to 16.7 GHz and subsequently multiplying the frequency twelvefold
 349 via the frequency multiplier.

350 Our ONN is designed to operate with far-field THz waves. It is positioned at a distance of
 351 15 cm (equivalent to 100 wavelengths, 100×1.5 mm) away from the THz source, ensuring the
 352 far-field condition is satisfied for accurate beam propagation analysis using the ASM. The weights
 353 embedded within the three diffractive layers of the ONN are obtained through the training process,
 354 as detailed in Section 2.3. These layers were fabricated using a masked stereolithography (MSLA)
 355 3D printer (Phrozen, Mighty 4K) with a 20/52- μm axial/lateral resolution, enabling the precise
 356 implementation of the trained weights across the diffractive layers. The 3D printing material
 357 used is a low-absorption-coefficient, high-refractive-index composite resin in the THz frequency
 358 range based on methacrylate oligomers with TiO_2 nanoparticles [34]. Additionally, we have
 359 incorporated an offset thickness into all diffractive layers to enhance mechanical stability without
 360 interfering with the phase-based network operations. To validate the material behavior of the
 361 3D-printed THz composite material, we employed a THz time-domain spectroscopy (THz-TDS)
 362 system to measure their absorption and dielectric characteristics (as shown in Fig. 4(b)). It
 363 illustrates the measured mean refractive index of 1.70 and a mean absorption loss of 0.25 cm^{-1}
 364 for the three diffractive layers. The simulated and fabricated diffractive layers are compared
 365 in Fig. 4(c), where randomly selected areas were tested for printing accuracy, showing an average
 366 discrepancy of 2.3 μm . Furthermore, to evaluate the efficacy of ONN to manipulate the THz
 367 beam characteristics, three sample patterns (y, R, g) were then fabricated and tested in the
 368 experimental setup. Their compressed imaging was calculated in a simulated environment, and
 369 the fabricated ONN was used to perform compressed imaging in a real-world environment. The
 370 result, presented in Fig. 4(d), demonstrates that the experimental compressed images closely
 371 match the simulated compressed images, achieving an average PSNR of 22.63. These validate that
 372 our fabrication process and simulations are consistent with real-world performance, demonstrating
 373 the performance and reliability of the designed ONN module.

374 After the THz beam passing through three diffractive layers, the amplitude and phase

375 information is compressed into intensity form and captured by a THz bolometric sensor array
376 with a resolution of 384×288 pixels and a total area of $1.344 \times 1.008 \text{ cm}^2$. Despite being the
377 most mature and commercially available option, the THz bolometric sensor array has time-vary
378 thermal drift issues [35–37]. To mitigate this, we applied digital processing methods, including
379 average pooling and downsize the THz bolometric captured images to 8×8 pixel, for effectively
380 reducing the impact of thermal drift. That means the ONN compresses the dimensionality of
381 the image from 1024 pixels/dimensions to 64 pixels/dimensions. Subsequently, the THz sensor
382 array converts these signals into digital intensity form. Normalization is then applied to ensure
383 consistency and compatibility with subsequent processing steps. The compressed intensity data
384 is then fed into a DNN, whose framework follows the structure outlined in Section 3.1. Serving
385 as the decoder, the DNN is jointly trained with the ONN to ensure seamless integration. The
386 jointly trained DNN interprets the encoded signals from the ONN, decoding them to perform
387 high-dimensional transformations. The DNN processes and decodes the compressed imaging
388 data to produce an output at 128×128 pixels. In Fig. 5, we present the experimental results of the
389 hybrid optical-digital neural network THz imaging system. To validate the system, we used three
390 distinct letters—“y,” “R,” and “g.” The reconstructed images, shown in Fig. 5 (a), demonstrate
391 high fidelity with no apparent diffraction artifacts, clearly displaying distinguishable few-cycle-
392 wavelength letters. The average PSNR and MSE of the experimentally reconstructed images
393 are 10.5 and 65.5, respectively. Additionally, the system successfully covers a $2.4 \times 2.4 \text{ cm}^2$
394 field-of-view, which is four times larger than the coverage area of the sensor array, without any
395 lensing system in use.

396 Furthermore, by leveraging the video-rate capability of the THz bolometric sensor array, our
397 imaging system achieves real-time imaging at a frame rate of 2 frames per second. Notably, the
398 demonstrated frame rate is just constrained by our current hardware setup. Utilizing advanced
399 graphics processing unit (GPU) modules can further enhance the frame rate, enabling much
400 faster processing speed and real-time imaging capabilities. For demonstrating real-time imaging
401 capabilities of the proposed system in real-world scenario, a 5-mm-thick screwdriver was used as
402 a sample and moved at various angles, as shown in Fig. 5 (b) and Supplementary Video 1, to
403 test the imaging system. In Supplementary Video 2, the results display a distinct line that shifts
404 in response to the movement of the sample. Additionally, we performed the same experiment
405 without the ONN to verify that the observed moving line is indeed a response specific to the
406 working principle of the hybrid optical-digital neural network imaging system. In Supplementary
407 Video 3, results without the ONN show an inability to decode any structure or distinguishable
408 object, confirming the effectiveness of the hybrid configuration.

409 4. Conclusion

410 This work introduces a compressive optical-digital neural network framework that transforms
411 signal dimensions, enabling high-speed data acquisition and information reconstruction for
412 real-time THz imaging. Within this framework, the ONN serves as the physical encoder, designed
413 to compress full-dimensional THz signals into a lower-dimensional format as the optical signal
414 carriers propagate through the ONN. Specifically, we transformed the signal from amplitude-phase
415 dimension to intensity dimension. In the space domain, we further reduced the pixel count
416 from 1024 to 64. These allow the use of a high-frame-rate, power-detection THz camera to
417 rapidly record the compressed, low-dimensional signals in digital format. The DNN acts as the
418 digital decoder, decompressing the signal back to its original dimensionality to extract object
419 information. Noteworthily, the key idea of this computational framework is that the ONN not
420 just handles the majority of the computational processing at light speed but encodes full object
421 information into a latent space that is easy to record. This significantly enhances both data
422 acquisition and image reconstruction speeds, satisfying the essential requirements for real-world
423 THz imaging tasks.

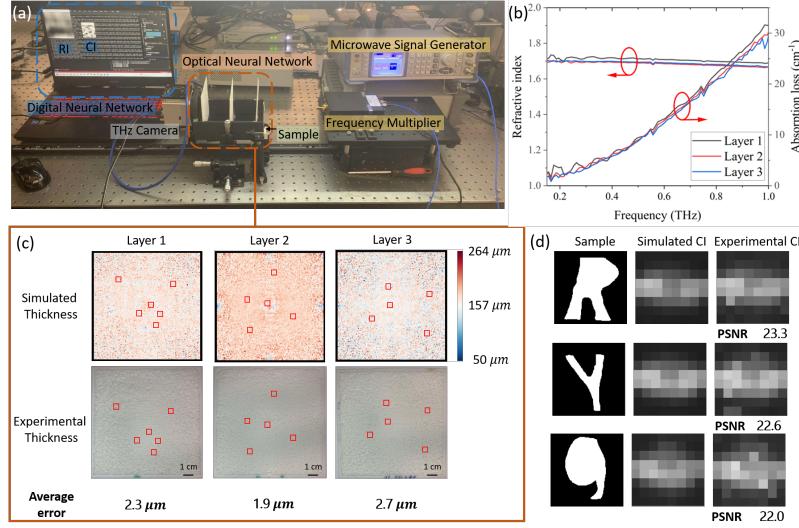


Fig. 4. (a) Experimental setup of the hybrid optical-digital neural network THz imaging system. The THz source consists of a microwave signal generator ($f = 16.7$ GHz) connected to a 12-fold frequency multiplier and a THz horn antenna. The emitted THz wave passes through the sample before reaching ONN, which is composed of three 3D-printed diffractive layers. A THz bolometric sensor array records the compressed THz intensity signal and converts it into a digital format. This digital signal is subsequently processed by a DNN, highlighted in the blue box on the laptop screen, which displays the compressed image (CI) and reconstructed image (RI). (b) The measured absorption loss and refractive index of the three 3D-printed diffractive layers. (c) Comparison of the simulated and the fabricated diffractive layers. (d) Comparison of the ground truth images, simulated compressed images, and experimental compressed images of the “y” “R” and “g” samples.

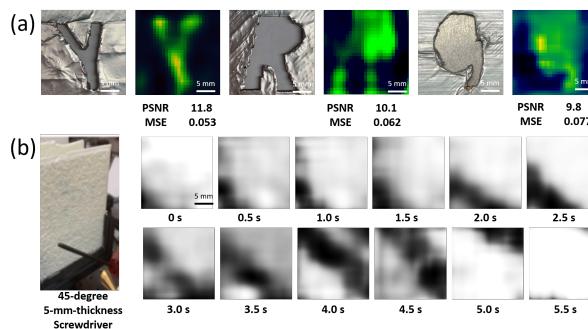


Fig. 5. (a) Reconstructed object information obtained from a hybrid optical-digital neural network THz imaging system. The left column showcases optical images of the sample objects, while the right column displays the corresponding reconstructed THz images. (b) Real-time reconstructed THz images of a moving screwdriver, captured with a time resolution of 0.5 seconds.

424 To effectively train both the physical encoders and digital decoders, we developed a digital
425 environment that simulates the physical behavior of THz beam propagation and light-matter
426 interactions between the THz beam and ONN diffractive layers. We employed the training
427 methodology of variational autoencoders to optimize encoder and decoder functionalities. (Add:
428 summarize experimental results and discussion of fig 3). To transition the ONN from digital to
429 physical environments, we developed a 3D-printable resin composite material with outstanding
430 dielectric properties, including a refractive index of 1.70 and an absorption coefficient of 0.25
431 cm^{-1} at 0.2 THz. Using a high-precision MSLA 3D printer, we implemented the physical
432 encoder with deep sub-wavelength scale lateral/axial printing resolution, ensuring almost identical
433 geometry and functionality between the 3D-printed ONN and its digital design. This compressive
434 optical-digital neural network framework has successfully demonstrated its capability to transform
435 signal dimensions and reconstruct high-quality THz images. (Add: summarize experimental
436 results and discussion of fig 5). As expected, the data processing speed of the ONN, coupled
437 with the rapid data recording capabilities of the THz camera and the efficient data reconstruction
438 speed of the compact DNN model, enabled us to achieve real-time imaging at 2 frames per
439 second on a GPU-free laptop.

440 It is worth noting that our proposed imaging system is not confined to the THz range but
441 can operate across other regions of the electromagnetic spectrum. For instance, it can be
442 applied to gigahertz-to-visible waves for high-data imaging communication and compressed
443 nonlinear infrared signals for video-rate imaging, as these domains share same optical propagation
444 principles. Moreover, although we demonstrated the co-design of ONN and DNN, the framework
445 is flexible enough to incorporate alternative model architectures, such as residual networks [38]
446 or transformer-based models [39]. Although this study primarily focuses on enhancing imaging
447 quality, the hybrid optical-digital compression method also facilitates the transformation of
448 physical signals into lower-dimensional representations. This transformation enables the jointly
449 trained DNN to effectively extract valuable information, such as material properties, phase details,
450 and polarization characteristics.

451 **Funding.** National Science and Technology Council, Taiwan (NSTC 112-2221-E-007-089-MY3).

452 **Supplementary Material.**

453 **Supplementary Video**

456 **Acknowledgments.** The authors thank Mr Pierre Talbot (INO) and their NTHU colleagues, including
457 Professors Yi-Chun Liu and Kai-Ming Feng , along with fellows Ms Yun-Ting Tseng, Mr Tsung-Han Wu,
458 Mr Yi-Chun Hung, Mr Pouya Torkaman, Mr Han-Yu Liang, Mr Chun-Yu Kao, and Mr Chia-Ming Mai for
459 their valuable support and insightful suggestions.

460 **Disclosures.** The authors declare no conflicts of interest.

461 **Data Availability.** Data underlying the results presented in this paper are not publicly available at this time
462 but may be obtained from the authors upon reasonable request.

463 **References**

- 464 1. G. C. Walker, E. Berry, N. N. Zinov'ev, A. J. Fitzgerald, R. E. Miles, J. M. Chamberlain, and M. A. Smith, “Terahertz
465 imaging and international safety guidelines,” Med. Imaging 2002: Phys. Med. Imaging **4682**, 683–690 (2002).
- 466 2. E. Berry, “Risk perception and safety issues,” J. Biol. Phys. **29**, 263–267 (2003).
- 467 3. R. H. Clothier and N. Bourne, “Effects of THz exposure on human primary keratinocyte differentiation and viability,”
468 J. Biol. Phys. **29**, 179–185 (2003).
- 469 4. P. U. Jepsen, D. G. Cooke, and M. Koch, “Terahertz spectroscopy and imaging—modern techniques and applications,”
470 Laser & Photonics Rev. **5**, 124–166 (2011).
- 471 5. M. Naftaly, A. Foulds, R. Miles, and A. Davies, “Terahertz transmission spectroscopy of nonpolar materials and
472 relationship with composition and properties,” Int. J. Infrared Millim. Waves **26**, 55–64 (2005).
- 473 6. M. Wan, J. J. Healy, and J. T. Sheridan, “Terahertz phase imaging and biomedical applications,” Opt. & Laser
474 Technol. **122**, 105859 (2020).

- 475 7. X. Yang, X. Zhao, K. Yang, Y. Liu, Y. Liu, W. Fu, and Y. Luo, "Biomedical applications of terahertz spectroscopy
476 and imaging," *Trends biotechnology* **34**, 810–824 (2016).
- 477 8. Y. Zhang, C. Wang, B. Huai, S. Wang, Y. Zhang, D. Wang, L. Rong, and Y. Zheng, "Continuous-wave thz imaging
478 for biomedical samples," *Appl. Sci.* **11**, 71 (2020).
- 479 9. M. Kutas, B. Haase, P. Bickert, F. Rixinger, D. Molter, and G. von Freymann, "Terahertz quantum sensing," *Sci.
480 Adv.* **6**, eaaz8065 (2020).
- 481 10. K. Kawase, Y. Ogawa, Y. Watanabe, and H. Inoue, "Non-destructive terahertz imaging of illicit drugs using spectral
482 fingerprints," *Opt. express* **11**, 2549–2554 (2003).
- 483 11. J. F. Federici, B. Schulkin, F. Huang, D. Gary, R. Barat, F. Oliveira, and D. Zimdars, "Thz imaging and sensing for
484 security applications—explosives, weapons and drugs," *Semicond. science technology* **20**, S266 (2005).
- 485 12. S.-H. Park, J.-W. Jang, and H.-S. Kim, "Non-destructive evaluation of the hidden voids in integrated circuit packages
486 using terahertz time-domain spectroscopy," *J. Micromechanics Microengineering* **25**, 095007 (2015).
- 487 13. B. B. Hu and M. C. Nuss, "Imaging with terahertz waves," *Opt. letters* **20**, 1716–1718 (1995).
- 488 14. M. Koch, S. Hunsche, P. Schumacher, M. Nuss, J. Feldmann, and J. Fromm, "Thz-imaging: a new method for density
489 mapping of wood," *Wood Sci. Technol.* **32**, 421–427 (1998).
- 490 15. X. Li, D. Mengu, N. T. Yardimci, D. Turan, A. Charkhesht, A. Ozcan, and M. Jarrahi, "Plasmonic photoconductive
491 terahertz focal-plane array with pixel super-resolution," *Nat. Photonics* **18**, 139–148 (2024).
- 492 16. Q. Wu, T. Hewitt, and X.-C. Zhang, "Two-dimensional electro-optic imaging of thz beams," *Appl. Phys. Lett.* **69**,
493 1026–1028 (1996).
- 494 17. A. Lisauskas, M. Bauer, S. Boppel, M. Mundt, B. Khamaisi, E. Socher, R. Venckevičius, L. Minkevičius, I. Kašalynas,
495 D. Seliuta *et al.*, "Exploration of terahertz imaging with silicon mosfets," *J. Infrared, Millimeter, Terahertz Waves* **35**,
496 63–80 (2014).
- 497 18. A. Lisauskas, D. Glaab, H. Roskos, E. Oejefors, and U. Pfeiffer, "Terahertz imaging with si mosfet focal-plane arrays,"
498 in *Terahertz Technology and Applications II*, vol. 7215 (SPIE, 2009), pp. 119–129.
- 499 19. S. Nadar, H. Videlier, D. Coquillat, F. Teppe, M. Sakowicz, N. Dyakonova, W. Knap, D. Seliuta, I. Kašalynas, and
500 G. Valušis, "Room temperature imaging at 1.63 and 2.54 thz with field effect transistor detectors," *J. Appl. Phys.* **108**
501 (2010).
- 502 20. X. Li, J. Li, Y. Li, A. Ozcan, and M. Jarrahi, "High-throughput terahertz imaging: progress and challenges," *Light.
503 Sci. & Appl.* **12**, 233 (2023).
- 504 21. B. S. Karasik, A. V. Sergeev, and D. E. Prober, "Nanobolometers for thz photon detection," *IEEE Trans. on Terahertz
505 Sci. Technol.* **1**, 97–111 (2011).
- 506 22. C. Chevalier, F. Duchesne, L. Gagnon, B. Tremblay, M. Terroux, F. Généreux, J.-E. Paultre, F. Provençal,
507 Y. Desroches, L. Marchese *et al.*, "Introducing a 384x288 pixel terahertz camera core," in *Terahertz, RF, Millimeter,
508 and Submillimeter-Wave Technology and Applications VI*, vol. 8624 (SPIE, 2013), pp. 80–87.
- 509 23. W.-T. Su, Y.-C. Hung, P.-J. Yu, C.-W. Lin, and S.-H. Yang, "Physics-guided terahertz computational imaging: A
510 tutorial on state-of-the-art techniques," *IEEE Signal Process. Mag.* **40**, 32–45 (2023).
- 511 24. Y.-C. Hung, W.-T. Su, T.-H. Chao, C.-W. Lin, and S.-H. Yang, "Terahertz deep learning fusion computed tomography,"
512 *Opt. Express* **32**, 17763–17774 (2024).
- 513 25. W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, "A single-pixel terahertz
514 imaging system based on compressed sensing," *Appl. Phys. Lett.* **93** (2008).
- 515 26. R. I. Stantchev, X. Yu, T. Blu, and E. Pickwell-MacPherson, "Real-time terahertz imaging with a single-pixel detector,"
516 *Nat. communications* **11**, 2535 (2020).
- 517 27. X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using
518 diffractive deep neural networks," *Science* **361**, 1004–1008 (2018).
- 519 28. T. Fu, J. Zhang, R. Sun, Y. Huang, W. Xu, S. Yang, Z. Zhu, and H. Chen, "Optical neural networks: progress and
520 challenges," *Light. Sci. & Appl.* **13**, 263 (2024).
- 521 29. Ç. İşıl, T. Gan, F. O. Ardic, K. Mentesoglu, J. Digani, H. Karaca, H. Chen, J. Li, D. Mengu, M. Jarrahi *et al.*,
522 "All-optical image denoising using a diffractive visual processor," *Light. Sci. & Appl.* **13**, 43 (2024).
- 523 30. B. Bai, X. Yang, T. Gan, J. Li, D. Mengu, M. Jarrahi, and A. Ozcan, "Pyramid diffractive optical networks for
524 unidirectional image magnification and demagnification," *Light. Sci. & Appl.* **13**, 178 (2024).
- 525 31. J. Li, Y. Li, T. Gan, C.-Y. Shen, M. Jarrahi, and A. Ozcan, "All-optical complex field imaging using diffractive
526 processors," *Light. Sci. & Appl.* **13**, 120 (2024).
- 527 32. L. Pinheiro Cinelli, M. Araújo Marins, E. A. Barros da Silva, and S. Lima Netto, "Variational autoencoder," in
528 *Variational Methods for Machine Learning with Applications to Deep Networks*, (Springer, 2021), pp. 111–149.
- 529 33. Kaggle, "Kaggle datasets," (2023). Accessed: 2023-11-03.
- 530 34. S. M. Latifi, P.-J. Yu, and S.-H. Yang, "High refractive index material for 3d printing thz passive devices," in *2024
531 49th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, (IEEE, 2024), pp. 1–2.
- 532 35. G. Dai, M. Li, X. He, L. Du, B. Shao, and W. Su, "Thermal drift analysis using a multiphysics model of bulk silicon
533 mems capacitive accelerometer," *Sensors Actuators A: Phys.* **172**, 369–378 (2011).
- 534 36. H. Li, Y. Zhai, Z. Tao, Y. Gui, and X. Tan, "Thermal drift investigation of an soi-based mems capacitive sensor with
535 an asymmetric structure," *Sensors* **19**, 3522 (2019).
- 536 37. M. Santo Zarnik, D. Rocak, and S. Macek, "Residual stresses in a pressure-sensor package induced by adhesive
537 material during curing: a case study," *Sensors Actuators A: Phys.* **116**, 442–449 (2004).

- 538 38. M. Shafiq and Z. Gu, "Deep residual learning for image recognition: A survey," *Appl. Sci.* **12**, 8972 (2022).
- 539 39. S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM*
540 computing surveys (CSUR) **54**, 1–41 (2022).