Lensless object classification in long wave infrared using random phase encoding

Gregory Aschenbrenner, **Kashif Usmani, Saurabh Goswami, and Bahram Javidi***
University of Connecticut, Electrical and Computer Engineering Department, Storrs, Connecticut,
United States

ABSTRACT. We introduce a lensless long wave infrared (LWIR) sensing system, utilizing double-random phase encoding. The employment of thin random phase encoding elements eliminates the need for traditional optical lenses. For object classification, convolutional neural network is used to process the speckle patterns produced by the random phase encoding, thus avoiding the reconstruction problem associated with lensless imaging. This approach is attractive for applications demanding compactness and cost-efficiency for LWIR systems. Experiments are provided to illustrate the proposed system. Our results demonstrate that this system competes well with conventional lensed LWIR imaging methods in a binary classification task under noisy conditions, where noise is not known *a priori*. To the best of our knowledge, this is the first report on such approaches in the LWIR domain.

1 Introduction

Current trends of decreasing costs of micro-bolometric sensors[1] have driven long wave infrared (LWIR) imaging to become an increasingly vital electromagnetic spectrum for industrial, commercial, and military applications.[2] Although state-of-the-art LWIR optical setups are increasingly efficient, the vast majority rely on lenses for sensing and imaging, binding them to several key limitations. Of primary consideration is the resolution of the lens, which is limited by the Raleigh criterion[3] and dependent on the size of the aperture. The issue of increasing the size of the aperture is further compounded due to the material dispersion and absorption effects in the LWIR band, leading to only a small subset of relatively expensive materials, such as chalcogenide class, germanium, or silicon, being suitable for an LWIR lens. This becomes a challenging manufacturing issue when coupled with the risk of aberration with fabricating large lenses.[4] In addition, LWIR lenses are much bulkier than their visible spectrum counterparts leading to further increased size and weight of the sensing system.

In response to these limitations, our research proposes a differing approach: the substitution of the conventional lens with a more economical and easily manufacturable thin random encoding element optimized for the LWIR band. Utilizing such an encoding element, we propose a passive thermal imaging system that depends on the temperature gradient of a scene.

In specific, we propose utilizing a lensless approach, utilizing Mie diffuser(s) to induce pseudo-random phase encoding in the LWIR band. Such lensless systems employing random phase encoding do not capture visually recognizable natural images, requiring the user to either

---

*Address all correspondence to Bahram Javidi, bahram.javidi@uconn.edu

computationally reconstruct a scene[5–10] or perform classification tasks directly on capture speckle images.[11–23] The efficacy of random phase diffusers in the classification task, particularly under coherent illumination,[11–14] is well documented in a variety of fields, such as medicine[15–20] and agriculture,[21] while also being an integral optical element in incoherent lensless reconstruction[5–10] and incoherent speckle classification.[22,23] Diffusers reduce cost, complexity, and size[9–23] compared to conventional lensed systems.

Our experimental setup utilizes two cascaded Mie diffusers, where each diffuser acts as random phase masks with feature sizes comparable to that of the wavelength. Unlike amplitude masks, phase masks do not generate the obstruction of light, thereby minimizing potential information loss.[24,25] While uniformly redundant arrays are used as an LWIR mask in an RGB-LWIR fusion system for reconstruction,[26] they are generally more complex to implement than random phase masks.

In this paper, we capture incoherent speckle images of a variety of scene configurations in the LWIR domain, classify them with a ResNet-18 convolutional neural network (CNN), and compare the performance across experiments. Specifically, we compare the binary classification of thermally self-radiating objects with both differing and similar intensity levels, imaged through both lensed and double random phase encoding (DRPE)[16] configuration(s). By capturing similar scenes in both LWIR lensed and DRPE configuration, we seek to both characterize and compare the relative performance and robustness of accordingly trained CNNs on a small dataset.

2 Methods

2.1 Systems Overview

The system we pursue is not overly complex and is based on a typical DRPE system. The system consists of a thermal scene in the object plane, two random phase diffusers, in parallel with a thermal imaging sensor, as shown in Fig. 1.

The diffuser(s) are fabricated in the lab with abrasive techniques.[24,25,27–29] The sensor is a 240×320 microbolometer at 17 μm pixel pitch. When collecting data imaged with lenses for comparison, we employ an optical lens with an effective focal length of 11 mm and an aperture ($f/\#$) of $f/1.2$. When this camera uses a lens, it is typically used for machine vision, making it an ideal candidate to use for comparisons between lensed and lensless diffuser based data. We used standard calibration across all imaging modalities to avoid changes in relative performance. A diagram of our diffuser system is shown in Fig. 1. We captured scenes at distances between 1 and 3 m. All images were acquired with the subjects positioned within a maximum angular deviation of 35 deg from the optical axis. Further, we experimentally find fixing the distance between diffusers, $z_2 = 5$ mm, yields strong scattering at the sensor. Note that, z_3 is dictated by the length of the lens tube after lens removal, that is, $z_3 = 12$ cm. An overhead view of the system configuration is shown in Fig. 2.
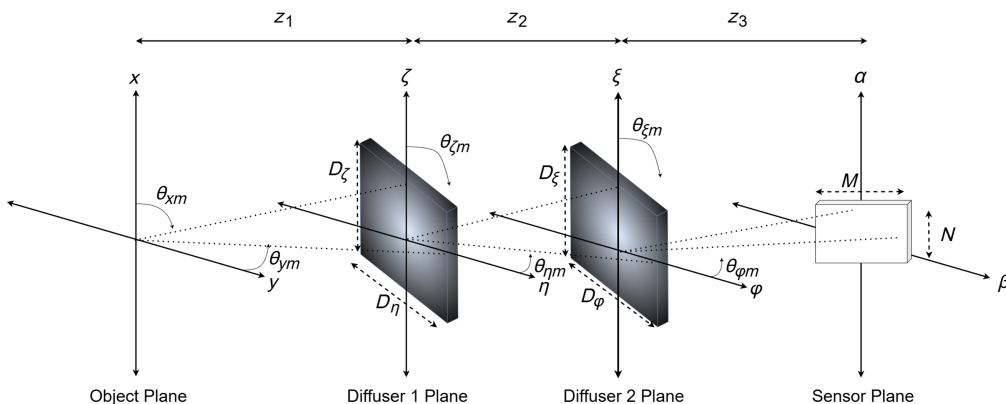

Fig. 1 Schematic diagram of optical field propagation through cascaded diffusers.
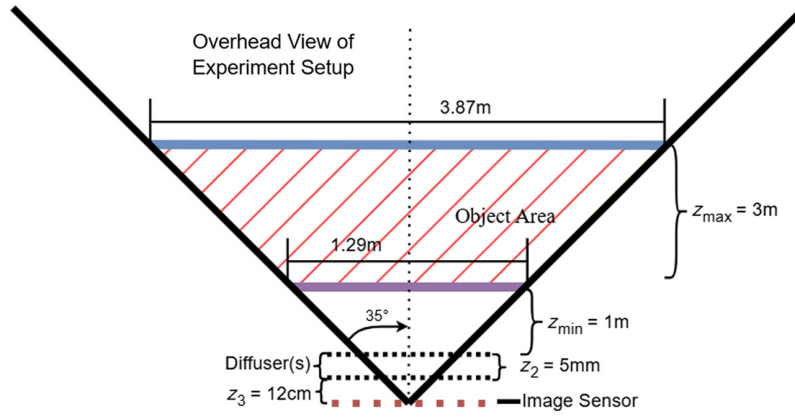
Fig. 2 Overhead view of the experimental setup.

2.2 Mathematical Description of Double Random Phase Encoding System

The optical theory for propagation follows the usual formulations for thin diffusers.[30] The following optical theory for propagation corresponds to Fig. 1. In our experiments, the object field is unknown and is denoted as $U_0(x, y)$. Due to the finite dimension $(D_\zeta, D_\eta, D_\xi, D_\varphi)$ of the rectangular diffusers and large spatial frequencies present in the Fourier transform, not all spatial frequencies can be captured.[30] The maximum spatial frequencies for the first diffuser will be denoted as (f_{xm}, f_{ym}) and the second diffuser will be denoted as $(f_{\zeta m}, f_{\eta m})$. We write our ideal low-pass filter(s), $h_{D1}(x, y)$ and $h_{D2}(x, y)$, with cut off frequencies (f_{xm}, f_{ym}) and $(f_{\zeta m}, f_{\eta m})$, respectively, where the convolution is denoted as *.

$$h_{D1}(x, y) = f_{xm} f_{ym} \operatorname{Sinc}(x f_{xm}) \operatorname{Sinc}(y f_{ym}), \tag{1}$$

$$U_0'(x, y) = U_0(x, y) * h_{D1}(x, y). \tag{2}$$

The field is then propagated at a distance z_1 to the first diffuser, where $\mathrm{F}^{-1}\{*\}$ denotes the inverse Fourier transform:

$$U_1(\zeta, \eta) = U_0'(\zeta, \eta) * \mathrm{F}^{-1}\left\{ \exp\left(\frac{j2\pi z_1}{\lambda} [1 - (\lambda f_\zeta)^2 - (\lambda f_\eta)^2]^{1/2} \right) \right\}. \tag{3}$$

The first diffusers transmittance function, $t_{D1}(\zeta, \eta)$, is given below, where $\Phi(\zeta, \eta)$ is the random phase angle uniformly distributed between $[0, 2\pi]$:

$$t_{D1}(\zeta, \eta) = \exp(j\Phi(\zeta, \eta)) \operatorname{Rect}\left(\frac{\zeta}{D_\zeta}, \frac{\eta}{D_\eta} \right). \tag{4}$$

The field through the diffuser can then be written as

$$U_1'(\zeta, \eta) = t_{D1}(\zeta, \eta) U_1(\zeta, \eta). \tag{5}$$

We repeat these steps to propagate the light to the second diffuser by taking into account the maximum spatial frequencies $(f_{\zeta m}, f_{\eta m})$, imposed by the second diffuser:

$$h_{D2}(\zeta, \eta) = f_{\zeta m} f_{\eta m} \operatorname{Sinc}(\zeta f_{\zeta m}) \operatorname{Sinc}(\eta f_{\eta m}), \tag{6}$$

$$U_1''(\zeta, \eta) = U_1'(\zeta, \eta) * h_{D2}(\zeta, \eta), \tag{7}$$

$$U_2(\xi, \varphi) = U_1''(\xi, \varphi) * \mathrm{F}^{-1}\left\{ \exp\left(\frac{j2\pi z_2}{\lambda} \right) [1 - (\lambda f_\xi)^2 - (\lambda f_\varphi)^2]^{1/2} \right\}, \tag{8}$$

$$t_{D2}(\xi, \varphi) = \exp(j\Phi(\xi, \varphi)) \operatorname{Rect}\left(\frac{\xi}{D_\xi}, \frac{\varphi}{D_\varphi} \right), \tag{9}$$

$$U_2'(\xi, \varphi) = t_{D2}(\xi, \varphi) U_2(\xi, \varphi). \tag{10}$$

Finally, we repeat these steps to propagate to the image sensor with maximum spatial frequencies $(f_{\xi m}, f_{\varphi m})$ imposed by the image sensor size:

$$h_S(\xi,\varphi) = f_{\xi m} f_{\varphi m} \operatorname{Sinc}(\xi f_{\xi m}) \operatorname{Sinc}(\varphi f_{\varphi m}), \tag{11}$$

$$U_2''(\xi,\varphi) = U_2'(\xi,\varphi) * h_S(\xi,\varphi), \tag{12}$$

$$U_S(\alpha,\beta) = U_2''(\alpha,\beta) * \mathrm{F}^{-1}\left\{ \exp\left(\frac{j2\pi z_3}{\lambda}\right)[1 - (\lambda f_\alpha)^2 - (\lambda f_\beta)^2]^{1/2} \right\}, \tag{13}$$

$$U_S'(\alpha,\beta) = U_S(\alpha,\beta)\operatorname{Rect}\left(\frac{\alpha}{M},\frac{\beta}{N}\right), \tag{14}$$

$$I(\alpha,\beta) = |U_S'(\alpha,\beta)|^2. \tag{15}$$

From Eq. (15), the continuous field is then sampled by the sensor at each discrete pixel location, where each pixel value is the average intensity over some time on that pixel. For the sake of derivations here, integration time is not accounted for. Let the pixel size be equal to (p_α, p_β) centered about discrete points $(m \times \operatorname{pix}_\alpha, n \times \operatorname{pix}_\beta)$ such that $-\frac{M}{2} \le m \le \frac{M}{2}$, $-\frac{N}{2} \le n \le \frac{N}{2}$ and $(\operatorname{pix}_\alpha, \operatorname{pix}_\beta)$ are pixel pitches along the (α,β) axis. Then, let $S(\cdot)$ be the sensor sampling function defined as

$$S(\cdot) = \operatorname{Rect}\left(\frac{\alpha - m \times \operatorname{pix}_\alpha}{p_\alpha}\right)\operatorname{Rect}\left(\frac{\beta - n \times \operatorname{pix}_\beta}{p_\beta}\right)\operatorname{Rect}\left(\frac{m}{M}\right)\operatorname{Rect}\left(\frac{n}{N}\right). \tag{16}$$

Then, the sampled intensity $I_S[m,n]$ is given below; a deviation from the monochromatic illumination is the integration over wavelength, due to an incoherent source over a wide band. Where $\mathrm{QE}(\lambda)$ is the quantum efficiency of the sensor with respect to the wavelength λ:

$$I_S[m,n] = \frac{1}{P_A}\int_{7\ \mu\mathrm{m}}^{14\ \mu\mathrm{m}}\iint I(\alpha,\beta)\mathrm{QE}(\lambda)S(\cdot)\mathrm{d}\alpha\,\mathrm{d}\beta\,\mathrm{d}\lambda, \tag{17}$$

$$P_A = \int_{7\ \mu\mathrm{m}}^{14\ \mu\mathrm{m}}\iint \mathrm{QE}(\lambda)S(\cdot)\mathrm{d}\alpha\,\mathrm{d}\beta\,\mathrm{d}\lambda. \tag{18}$$

2.3 Data Collection Procedure

The experiments pursued sought to investigate the relative performance of our lensless diffuser based system in the binary classification task under varied thermal scene conditions, when compared with a comparable lensed configuration. Two primary cases are investigated, the first case sought to gauge the performance of the lensless diffuser based system when two classes have intensities that significantly differ from one another; and the second case investigated when two classes have intensities that are similar. To achieve this, we use an iron and a beaker containing boiled water as self-radiating objects. These objects were chosen because the iron produces an intensity that is significantly higher than that of the beaker, creating a strong contrast between the classes. This, however, also means that when the two objects are put together, the resultant intensity closely resembles the intensity pattern for iron. This allows us to effectively analyze both differential and similar intensity scenarios. We conduct experiments with both imaging modalities, under both experimental setups in order to establish the performance capabilities corresponding to each of these configurations.

Further, we use self-radiating objects, such as a hot mannequin doll, a soldering iron, a kettle, and incandescent lights bulbs as "noise" objects. This is done to create variation of the background in efforts to prevent overfitting of deep learning models. The need for noisy objects arises from operating in the LWIR, where we cannot easily change the background of our scene as in the visible case. The use of noise objects is shown in Fig. 3.

Due to the lack of published research in LWIR diffuser imaging, we not only capture similar lens based imaging data to compare against our lensless diffuser setup but also image without an optical encoding element. This is to ensure that our diffuser(s) are not just acting as transmissive windows. For each experiment, we collect 150 videos for both classes, varying the background by shifting the "noise" objects and the class object. We will later add noise to the data in post-processing, to evaluate noise's effect on model performance in each imaging modality.
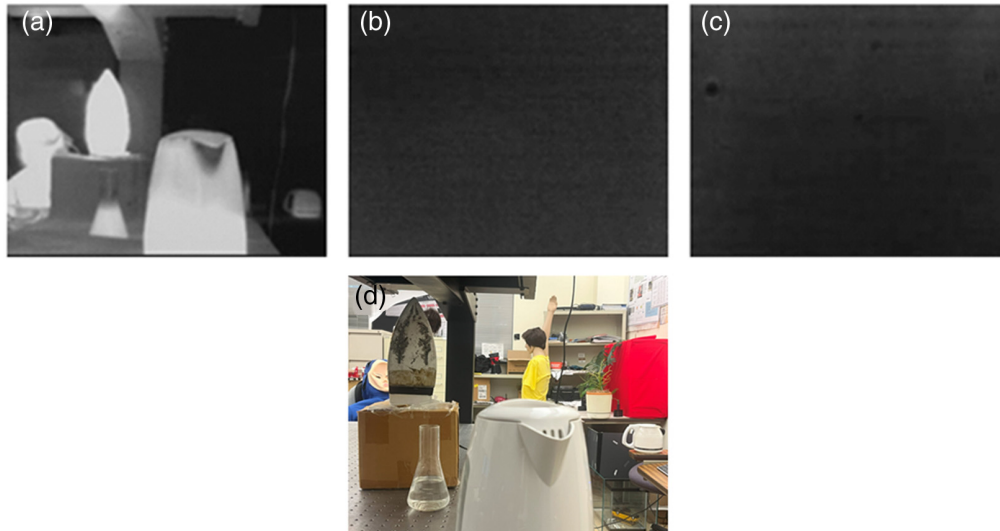
Fig. 3 Sample frames: (a) sample frame from iron class taken with lens based LWIR imaging configuration. (b) Sample frame from iron class taken with lensless diffuser(s) based configuration. (c) Sample frame from iron class taken with no diffuser and no lens configuration, and (d) sample frame taken in lens based visible imaging configuration for visualization.

2.4 Deep Learning Model

In this study, we employ a modified ResNet-18 architecture,[31] which is a base 18-layer CNN. The choice to employ a CNN over other methods, such as support vector machines (SVMs) or random forests (RFs), arises from prior research with single random phase encoding (SRPE) and DRPE systems. Under coherent conditions in SRPE systems, CNNs were shown to be more robust to noise[17] and data compression[19] when compared to initial approaches utilizing RFs or SVMs.[15,16] In addition, the use of an RF or SVM requires careful selection of appropriate features, which provides an additional refinement step in the imaging pipeline, as opposed to using CNNs, which inherently extract features from data.

Prior to training the network, images are resized to 120×160. The sole alteration(s) to the architecture involves replacing the final fully connected layer with a cascade of four fully connected layers, the first one sized 16 and incorporating a dropout of 0.3. This is followed by fully connected layers of size 8 and 4, ultimately concluding with a fully connected layer of size 2. The output of this final fully connected layer is then passed through a sigmoid activated single neuron, tailored for the binary classification task. The network is shown in Fig. 4.
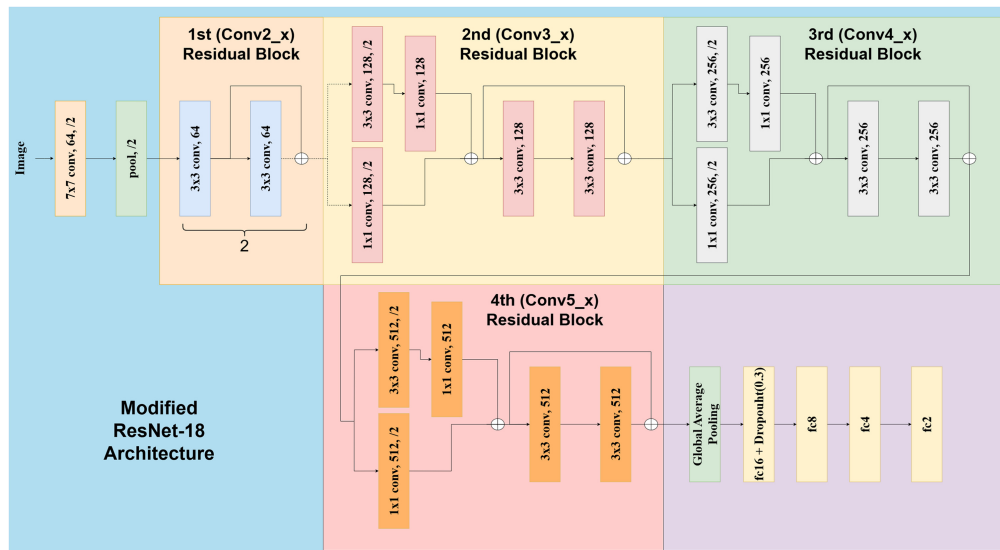

Fig. 4 Visualization of modified ResNet-18 CNN architecture utilized in the experiments.

Table 1 Optimized hyperparameters for each of the imaging modalities and noise levels.

		Hyperparameters		
Experiment	Imaging modality	Batch size	Learning rate	Optimizer
Iron versus beaker	Lens based	256	1e-3	ADAM
	Lensless diffuser	128	1e-4	ADAM
Iron versus iron + beaker	Lens based	128	1e-4	ADAM
	Lensless diffuser	256	1e-4	ADAM

To optimize the performance of the ResNet-18 based model on our LWIR dataset, we augment data and implement a hyperparameter grid search strategy. The specific data augmentation strategies employed during this study are random rotations, random reflections, random scaling, and random translations. During random rotations, each image was rotated by an angle uniformly sampled from the range 0 deg to 360 deg. To facilitate random reflection, images were flipped across horizontal and/or vertical axes. Random scaling involved either shrinking or widening of the image by a scale factor randomly selected between a range of 50% (0.5 times) and 150% (1.5 times). Further, random translation shifts the images in both horizontal and vertical directions by a random number of pixels within the range of −25 to 25 pixels.

The grid search methodology searches over optimizer type [adaptive moment estimation (ADAM) and stochastic gradient descent method (SGDM)], mini-batchsize (32, 64, 128, and 256), and learning rate (1e-3, 1e-4, 1e-5). Early stopping was set with a patience of 15 iterations and a validation frequency of 20 iterations. The model(s) trained on clean data are then evaluated for their robustness to noise in the testing phase, as detailed in Sec. 3.

We present our optimal hyperparameters for each noise level and imaging modality in Table 1.

3 Experimental Results

Training and testing of each model were performed as a five-fold cross validation. We separate data into each fold by a unique video ID, as opposed to by frame, to prevent data leakage. The motivation for the cross-validation procedure arises from collecting 1.5 s videos and then splitting them into frames, hence we cannot truly consider these data points as independent. To ensure that the model is properly learning across datapoints, we use the cross-validation procedure to ensure that different splits of the data yield similarly performing models and are not dependent on the videos being used to train.

Our study evaluates each imaging system across various performance metrics, including accuracy, precision, recall, $F1$-score, and area under the curve (AUC), across differing amounts of noise. Accuracy is simply the total number of correctly selected classes, divided by total predictions, $(TP + TN)/(P+N)$, where TP is the true positive cases, TN is the true negative cases, P is the positive cases, and N is the negative cases. Precision is a ratio of the correctly predicted positive cases to the total number of predicted positive cases and can similarly be written $TP/(TP + FP)$, where FP is the false positive cases. Recall is a proportion of correctly predicted positive cases to the total number of positive cases and can be written $TP/(TP + FN)$. $F1$-score is the harmonic mean of precision and recall. AUC provides significant information about the ROC curve in a single scalar value. AUC ranges between 0 and 1, where an AUC of 0.5 signifies random guessing, and a value of 1 indicates perfect discrimination of positive instances. In the first experiment, we use iron as our positive object, and in the second experiment, we use iron + beaker as our positive object.

3.1 Experimental Results

Table 2 introduces the average results for tracked metrics across speckle noise, mean $\mu = 0$, variance(s) $\sigma = [0, 0.01, 0.03, 0.05, 0.08]$. Both the iron versus beaker and iron versus iron + beaker experiment(s), are presented with average results across noise levels. In Figs. 4 and 5,

Table 2 Average results for each imaging modality across noise level(s) and experiments.

Experiment	Imaging modality	Average metrics				
		Accuracy (%)	F1-score	Recall	Precision	AUC
Iron versus beaker	Lens based	61.5	0.74	1	0.61	0.71
	Lensless diffuser	74.0	0.73	0.77	0.79	0.81
Iron versus iron + beaker	Lens based	64.2	0.33	0.31	0.40	0.75
	Lensless diffuser	76.7	0.78	0.81	0.76	0.83
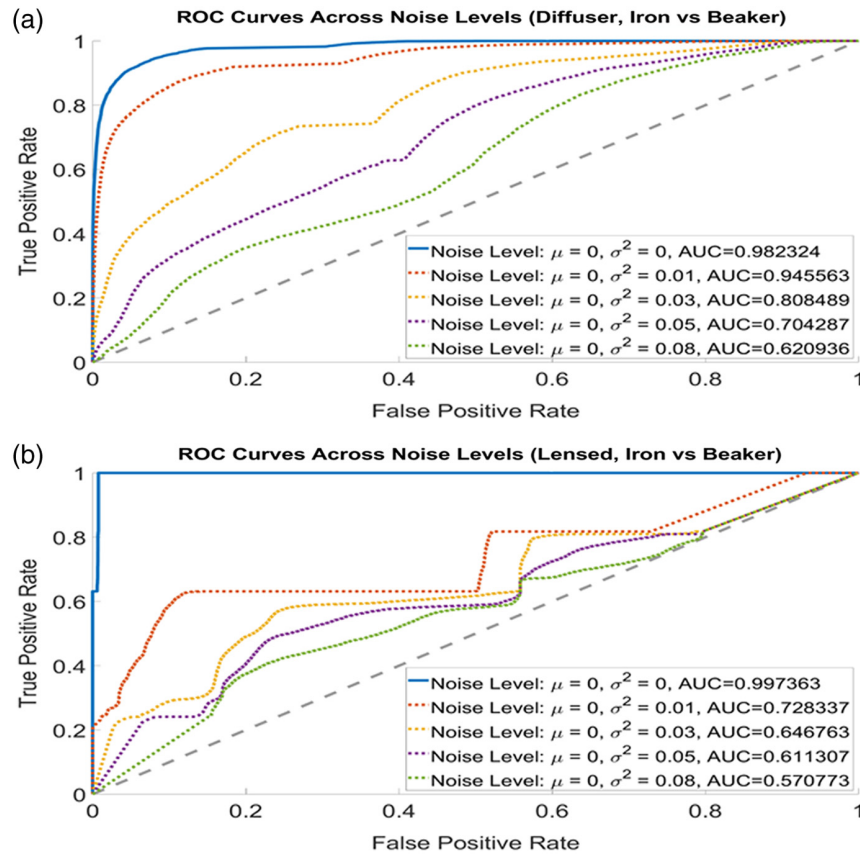
Fig. 5 ROC curves for iron versus beaker: (a) lensless diffuser based system in the presence of multiplicative speckle noise and (b) lens based system in the presence of multiplicative speckle noise, mean 0, variances [0, 0.01, 0.03, 0.05, 0.08].

we present the receiver operating characteristic (ROC) curve for the trained models for the iron versus beaker and iron versus iron + beaker experiment(s), respectively. Each ROC curve is shown for each noise level tested (Fig. 6).

3.2 Discussion

In examining the efficacy of imaging modalities under varying noise conditions in the iron versus beaker and iron versus iron + beaker experiments, we gain valuable insights into their performance. The lensless diffuser based imaging modality stands out for its robustness, maintaining high accuracy and AUC values across both experiments: with 74.0% accuracy and 0.81 AUC in the iron versus beaker experiment and 76.7% accuracy and 0.83 AUC in iron versus iron + beaker experiment. Conversely, the lens based imaging modality demonstrates vulnerability to noise. In the iron versus beaker experiment, its accuracy averages at 61.5% with a corresponding AUC
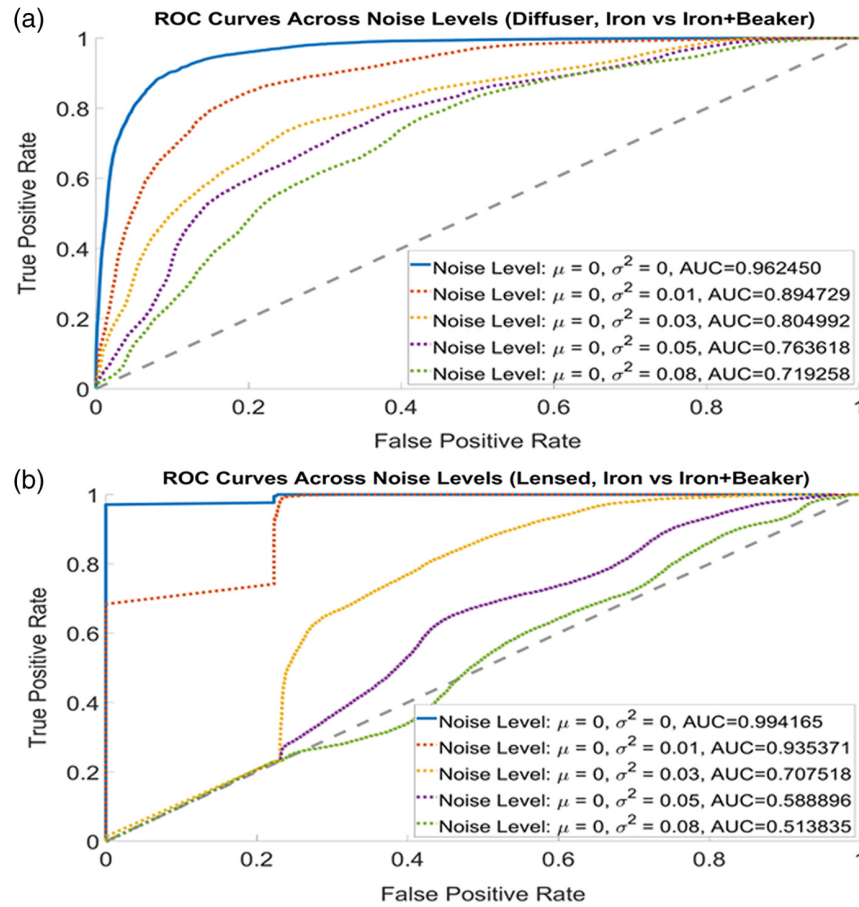
Fig. 6 ROC curves for iron versus iron + beaker. (a) Lensless diffuser based system in the presence of multiplicative speckle noise. (b) Lens based system in the presence of multiplicative speckle noise, mean 0, variances [0, 0.01, 0.03, 0.05, 0.08].

value of 0.71. Similarly, in the iron versus iron + beaker experiment, average accuracy 64.2% with a corresponding AUC value of 0.75. These results underscore the lens based imaging modality's susceptibility to noise in both experiments, exhibiting significant issues in maintaining classification accuracy and AUC.

Notably, while the lens based imaging iron versus beaker metrics are significantly higher than their lens based iron versus iron + beaker counterparts, this is due to occurrence of false positives versus false negatives in the models when noise is applied to the data. In the iron versus beaker experiment, where the beaker is the negative object, and the iron is the positive object, the addition of noise biases the detector to predict the positive object, resulting in an average recall of 1. In the iron versus iron + beaker experiment, where the iron is the negative object and iron + beaker is the positive object, the addition of noise biases the detector to predict the negative object, resulting in a comparatively low average recall of 0.31.

This does not imply either resultant model is "more" robust to noise than the other, instead simply they are biased to different sides of the decision boundary by the applied noise. Considering these concepts, it is clear that in both experiments conducted, the lensless diffuser based configuration is significantly more robust to speckle noise in *a priori* conditions, given the dataset tested.

Direct comparison to other SRPE and DRPE systems is challenging due to the unique specifics of the objects and conditions under which experiments were conducted. Each SRPE or DRPE system is often tailored to particular object or sample types, encoding schemes, and environmental conditions, making direct performance comparisons not straightforward. Comparing our DRPE results to state-of-the-art lens based systems is equally problematic. Modern computer vision architectures optimized for lens based data often utilize extensive pre-training on massive

datasets as is done with vision information transformers.[32–34] These advantages inherently skew the performance metrics, making a direct comparison unfair and not indicative of the fundamental performance differences between lens based and DRPE approaches.

For these reasons, we experimentally compare a lens based imaging system to the proposed DRPE system in similar experiments. Our results show that, given comparable dataset sizes, the DRPE system demonstrates superior robustness to speckle noise compared to traditional lens based LWIR systems. The enhanced robustness of DRPE in our specific setup underscores its potential for applications where data volume is limited, and noise resilience is critical.

4 Conclusions

This paper presents to the best of our knowledge the first lensless LWIR system for classification using a random phase encoder (diffuser). While we have used a DRPE system, a variety of diffuser configurations may be used. By implementing random phase diffusers in lieu of traditional lenses, this study showcases a method that exhibits enhanced resilience to noise while being a compact and cost-effective sensor and benefiting from ease of manufacture. This approach, as detailed in our experimental findings, underscores the inherent limitations of lens based imaging systems—particularly their vulnerability to noise when the associated CNN is not trained on noisy data. This is a particular limitation where the distribution of noise is not known *a priori*. A key insight from our analysis lies in the robust performance of the lensless diffuser-based system under varying levels of noise. The lensless diffuser LWIR system exhibited remarkable resilience, maintaining particularly high accuracy and AUC metrics in the presence of noise; a stark contrast to the lens based imaging modalities, which did not demonstrate significant robustness under similar noise conditions. This implies that the diffuser-based system's capability when coupled with CNN(s) helps mitigate the effects of noise; making it a more suitable choice for environments where noise is a significant challenge, and not known *a priori*.

In view of these experimental results, this method could be used in real-world applications where rapid deployment of compact, lightweight, low cost, and thermal imaging systems is necessary without extensive pre-training or *a priori* knowledge of noise characteristics. Such a scenario includes, but is not limited to, emergency response, security surveillance, and industrial monitoring, where quick adaptability to new environments and accurate and robust object detection are crucial. The experimental results are consistent with earlier predictions on the noise robustness of the lensless random phase encoding system.[30,35]

4.1 Future Work

Our research paves the way for future investigations into effects of different object materials on classification results, as well as enhancing the optical encoding element and investigating performance over distance. Further advancements could also be made in the machine learning pipeline by integrating more sophisticated models and preprocessing techniques or by expanding the scope to include multilabel and multiclass classification. Larger networks and datasets may be employed to increase observed metrics such as accuracy and AUC. Further improvements could involve the use of ensemble methods and the potential incorporation of explicit compressed sensing techniques.[36]

Disclosures

The authors declare no conflicts of interest.

Code and Data Availability

Data sharing is not applicable to this article, the code base nor the collected data will be publicly available due to export control restrictions. If one desires to replicate the results in our experiments, they will require an LWIR imaging sensor, not necessarily export controlled; random phase mask(s) optimized for the LWIR; follow our data collection and model fitting procedures; and properly vary scenes for uses on small datasets. Practitioners must be cautious of foregoing cross validation on small datasets collected in the manner done in this paper, as doing so will most likely cause some degree of overfitting.

References

1. T. Akin, "Low-cost LWIR-band CMOS infrared (CIR) microbolometers for high volume applications," in *IEEE 33rd Int. Conf. Micro Electro Mech. Syst. (MEMS)*, pp. 147–152 (2020).

2. L. Sagiv, S. R. Rotman, and D. G. Blumberg, "Detection and identification of effluent gases by long wave infrared (LWIR) hyperspectral images," in *IEEE 25th Convent. Electr. and Electron. Eng. in Israel*, pp. 413–417 (2008).

3. M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed., Cambridge University Press (1999).

4. J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill (1996).

5. A. Bhandari, A. Kadambi, and R. Raskar, *Computational Imaging*, The MIT Press (2022).

6. N. Antipa et al., "DiffuserCam: lensless single-exposure 3D imaging," *Optica* **5**, 1–9 (2018).

7. G. Kuo et al., "On-chip fluorescence microscopy with a random microlens diffuser," *Opt. Express* **28**, 8384–8399 (2020).

8. F. L. Liu et al., "Fourier diffuserscope: single-shot 3D Fourier light field microscopy with a diffuser," *Opt. Express* **28**, 28969–28986 (2020).

9. K. Monakhova et al., "Learned reconstructions for practical mask-based lensless imaging," *Opt. Express* **27**, 28075–28090 (2019).

10. V. Boominathan et al., "Recent advances in lensless imaging," *Optica* **9**, 1–16 (2022).

11. Y. Kashter, A. Vijayakumar, and J. Rosen, "Resolving images by blurring: superresolution method with a scattering mask between the observed objects and the hologram recorder," *Optica* **4**, 932–939 (2017).

12. K. Lee and Y. Park, "Exploiting the incoherent speckle-correlation scattering matrix for a compact reference-free holographic image sensor," *Nat. Commun.* **7**, 13359 (2016).

13. T. Ando, R. Horisaki, and J. Tanida, "Speckle-learning-based object recognition through scattering media," *Opt. Express* **23**(26), 33902–33910 (2015).

14. N. Anantrasirichai et al., "SVM-based texture classification in optical coherence tomography," in *IEEE 10th Int. Symp. Biomed. Imaging*, pp. 1332–1335 (2013).

15. B. Javidi et al., "Cell identification using single beam lensless imaging with pseudo-random phase encoding," *Opt. Lett.* **41**(15), 3663–3666 (2016).

16. B. Javidi, A. Marksman, and S. Rawat, "Automatic multicell identification using a compact lensless single and double random phase encoding system," *Appl. Opt.* **57**(7), B190–196 (2018).

17. T. O'Connor et al., "Red blood cell classification in lensless single random phase encoding using convolutional neural networks," *Opt. Express* **28**(22), 33504–33515 (2020).

18. B. Javidi, "Advances in automated disease identification with digital holography," in *Digital Hologr. and 3-D Imaging 2022*, Optica Publishing Group, p. Tu3A.1 (2022).

19. P. M. Douglass, T. O'Connor, and B. Javidi, "Automated sickle cell disease identification in human red blood cells using a lensless single random phase encoding biosensor and convolutional neural networks," *Opt. Express* **30**, 35965–35977 (2022).

20. Z. Zalevsky et al., "Simultaneous remote extraction of multiple speech sources and heart beats from secondary incoherent speckles pattern," *Opt. Express* **17**, 21566–21580 (2009).

21. A. Zdunek et al., "The bioincoherent speckle method for the investigation of agricultural crops: a review," *Opt. Lasers Eng.* **52**, 276–285 (2014).

22. X. Pan et al., "Incoherent reconstruction-free object recognition with mask-based lensless optics and the transformer," *Opt. Express* **29**, 37962–37978 (2021).

23. X. Pan et al., "Lensless inference camera: incoherent object recognition through a thin mask with LBP map generation," *Opt. Express* **29**(7), 9758–9771 (2021).

24. L. Dal Negro, "Waves in complex media," pp. 131–163, Cambridge University Press, Cambridge (2022).

25. J. Ojeda-Castañeda, "Wavefront engineering: phase conjugate masks," in *Wavefront Shaping and Pupil Engineering*, Springer Series in Optical Sciences, Vol. **235**, Springer, Berlin, Heidelberg (2021).

26. I. Reshetouski et al., "Lensless imaging with focusing sparse URA masks in long-wave infrared and its application for human detection," *Lect. Notes Comput. Sci.* **12364**, 237–253 (2020).

27. E. Hecht, *Optics*, 5th ed., Pearson Education (2017).

28. C. M. Sorensen, *Light Scattering and Absorption by Particles*, pp. 6-1–6-17, IOP Publishing (2022).

29. W. Hergert and T. Wriedt, The Mie Theory: Basics and Applications, Springer Series in Optical Sciences, Springer Berlin, Heidelberg (2012).

30. S. Goswami et al., "Assessment of lateral resolution of single random phase encoded lensless imaging systems," *Opt. Express* **31**, 11213–11226 (2023).

31. K. He et al. "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.* (2016).

32. S. Srivastava and G. Sharma, "OmniVec: learning robust representations with cross modal sharing," in *IEEE/CVF Winter Conf. Appl. of Comput. Vision (WACV)*, Waikoloa, Hawaii, pp. 1225–1237 (2024).

33. J. Yu et al., "CoCa: contrastive captioners are image-text foundation models," *Trans. Mach. Learn. Res.* 1–12 (2022).

34. M. Wortsman, "Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time," in *Proc. 39th Int. Conf. Mach. Learn.*, Vol. 162, pp. 23965–23998 (2022).

35. S. Goswami, G. Krishnan, and B. Javidi, "Robustness of lensless single random phase encoding imaging in presence of camera noise," *Opt. Express* **32**, 4916–4930 (2024).

36. A. Stern, *Optical Compressive Imaging*, CRC Press (2016).

Gregory Aschenbrenner is a PhD student at the University of Connecticut pursuing a degree in electronics, photonics, and biophotonics under Dr. Bahram Javidi. He received his BS degree in applied mathematics and BSE (honors laureate) in electrical engineering from the University of Connecticut in 2023.

Kashif Usmani is a PhD candidate at the University of Connecticut pursuing a degree in electronics, photonics, and biophotonics under Dr. Bahram Javidi. He received his BS degree in physics from Shibli National College in 2016, his MS degree in optics/optical sciences from IIT Delhi in 2018.

Saurabh Goswami is a PhD candidate at the University of Connecticut pursuing a degree in electronics, photonics, and biophotonics under Dr. Bahram Javidi. He received his BS degree in electrical engineering in 2018 and his MS degree in image processing and computer vision in 2021 from IIT Madras.

Bahram Javidi is a board of trustees distinguished professor at the University of Connecticut. He received his PhD in electrical and electronics engineering from Penn State. He has over 1100 publications, including 9 books, 58 book chapters, 540+ peer reviewed journal articles, and 520+ conference proceedings. He has been named fellow of the Institute of Electrical and Electronics Engineers (IEEE), fellow of the American Institute for Medical and Biological Engineering, fellow of the Optical Society of America, fellow of the European Optical Society, fellow of SPIE, fellow of the Institute of Physics, fellow of the Society for Imaging Science and Technology, and fellow of the Institution of Electrical Engineers. He received the fellow award by John Simon Guggenheim Foundation, the IEEE Donald G. Fink Prize Paper Award, the George Washington University's Distinguished Alumni Scholar Award, and the Humboldt and the Technology Achievement Award from SPIE.