OpEnCam: Lensless Optical Encryption Camera

Salman S. Khan, Xiang Yu, Kaushik Mitra, Manmohan Chandraker, Francesco Pittaluga

***Abstract*—Lensless cameras multiplex the incoming light before it is recorded by the sensor. This ability to multiplex the incoming light has led to the development of ultra-thin, high-speed, and single-shot 3D imagers. Recently, there have been various attempts at demonstrating another useful aspect of lensless cameras - their ability to preserve the privacy of a scene by capturing encrypted measurements. However, existing lensless camera designs suffer numerous inherent privacy vulnerabilities. To demonstrate this, we develop the first comprehensive attack model for encryption cameras, and propose OpEnCam– a novel lensless optical encryption camera design that overcomes these vulnerabilities. OpEnCam encrypts the incoming light before capturing it using the modulating ability of optical masks. Recovery of the original scene from an OpEnCam measurement is possible only if one has access to the camera's encryption key, defined by the unique optical elements of each camera. Our OpEnCam design introduces two major improvements over existing lensless camera designs - (a) the use of two co-axially located optical masks, one stuck to the sensor and the other a few millimeters above the sensor and (b) the design of mask patterns, which are derived heuristically from signal processing ideas. We show, through experiments, that OpEnCam is robust against a range of attack types while still maintaining the imaging capabilities of existing lensless cameras. We validate the efficacy of OpEnCam using simulated and real data. Finally, we built and tested a prototype in the lab for proof-of-concept.**

***Index Terms*—Lensless imaging, visual privacy, inverse problems.**

I. Introduction

Lensless imaging is a computational imaging modality that replaces the lens of a conventional camera with a thin optical mask and computation. The addition of the mask, which no longer has the focusing ability of the lens, leads to multiplexing of the incoming light prior to capture. This multiplexing of light has been exploited to develop ultra-thin cameras for 2D [1], [2] and 3D imaging [2], [3] high-speed imaging [4] and hyper-spectral imaging [5]. Another interesting aspect of lensless cameras that has been relatively less explored is their ability to preserve privacy in the optical domain itself before capture. This is made possible by the ability of lensless cameras to perform computation in the optical space, with operations defined by their unique mask patterns and arrangements.

With camera-based technologies becoming increasingly integrated into every aspect of our lives, the potential for leakage of sensitive visual information is ever-increasing. This is leading to significant privacy and security concerns from consumers, citizens, and governments. The standard approach to address these concerns is to encrypt sensitive image data at the sensor level, after image capture, via software or specialized hardware. However, most cameras lack such specialized hardware. Further, both software and hardware-based solutions may still be susceptible to data sniffing attacks that gain access to sensitive data before encryption. Thus, protecting the privacy of a scene through optical processing before capture is currently the need of the hour.

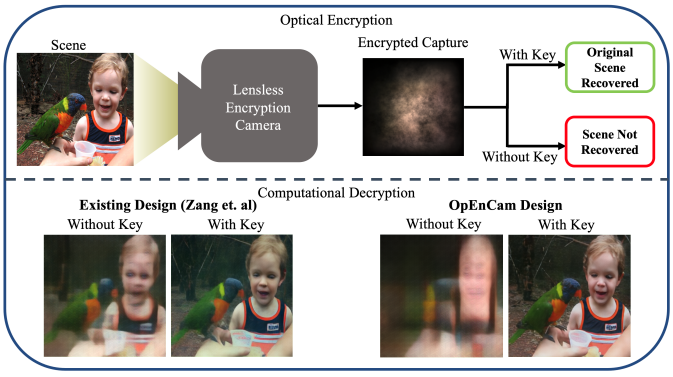

Fig. 1. **Optical Encryption.** Optical encryption camera encodes the scene optically prior to capture. Decryption is possible computationally using a key that is unique to the hardware of each encryption camera. Existing optical encryption designs are susceptible to powerful learning-based attacks even without the key. Proposed OpEnCam design prohibits such learning-based attacks effectively.

While optical image encryption has been around for many decades [6], most existing solutions require coherent illumination for encoding and decoding, making them unsuitable for deployment in unconstrained real-world environments. Further, incoherent illumination solutions that have been proposed are susceptible to ciphertext-only attacks [7] or have yet to be realized with real hardware due to the experimental nature of the optics [8].

A few works like [9]–[11] have used existing lensless cameras [1] to show optical encryption-based privacy-preserving applications. However, these works are fundamentally limited by the inherent design flaws of the existing lensless cameras, which were designed for imaging applications with no privacy considerations. In this work, we look at the design of existing lensless cameras and make two significant modifications to it that allows us to improve their privacy-preserving ability prior to capture. Our novel lensless camera design is called OpEnCam – Optical Encryption Camera. In OpEnCam, we first introduce a novel double mask design using two co-axially placed masks – one at the top of the bare sensor placed flush against it and the other a few millimeters above it. Such a double mask design makes lensless cameras significantly

S. S. Khan is with Rice University, Houston, TX, USA - 77005
X. Yu is with Amazon, USA
K. Mitra is with the Indian Institute of Technology Madras, Chennai, India - 600036
M. Chandraker and F. Pittaluga are with the NEC Labs America, San Jose, CA, USA - 95110
This paper was produced by the IEEE Publication Technology Group. They are in Piscataway, NJ.

less susceptible to various attacks without compromising their imaging ability. Secondly, we introduce a novel design strategy based on signal processing heuristics for the two masks used in our system that further enhances the privacy of our system. Finally, we show through experiments that our proposed OPENCAM design still has the imaging ability of existing lensless cameras while preserving privacy.

To summarize, we make the following contributions:

1) A novel framework for generating lensless computational cameras that encrypt image data before image capture, using two co-axially placed masks - a scaling mask placed flush against the bare sensor, and a multiplexing mask placed a few millimeters from the sensor.
2) Novel heuristically designed random key-generators for the two masks, providing a distinct mask pattern for each camera. We show that these key-generators are robust against various forms of ciphertext-only attacks and known plaintext attacks.
3) A novel transformer-based keyed decryption approach.
4) We validate our results both in simulation and real data collected using an inexpensive proof-of-concept prototype built in our lab.

II. BACKGROUND

A. Optical Cryptography

A cryptographic system is judged to be reliably secure only if it can successfully survive a rigorous evaluation via cryptanalysis. The goal of cryptanalysis is to identify potential weaknesses in encryption mechanisms that would allow someone to decode the encrypted data, even without possessing the confidential key. Technological progress in the realm of optical cryptography has spurred the introduction of various specialized cryptanalysis approaches, such as chosen plaintext attacks (CPA) [12], [13], known plaintext attacks (KPA) [14], [15], and ciphertext-only attacks (COA) [16]–[18].

In a CPA scenario, the attacker can choose arbitrary plaintexts to be encrypted and then examine the resulting ciphertexts. The primary objective is to use this information to discover a weakness in the encryption algorithm or even to deduce the secret key used for encryption. In contrast, in a KPA scenario, the attacker does not have the luxury of choosing the plaintexts that are encrypted. Instead, the attacker must work with pre-existing pairs of plaintexts and ciphertexts. Despite this limitation, a successful KPA could potentially reveal enough about the encryption scheme or key to decrypt other ciphertexts encrypted with the same key or to weaken the overall security of the cryptosystem. Finally, in a COA, the attacker has access solely to the ciphertext—that is, the encrypted data—without any accompanying plaintext or other additional information. The goal of the attacker is to deduce either the plaintext, the encryption key, or details about the encryption algorithm, based solely on the available ciphertexts. Ciphertext-only attacks are considered one of the most challenging forms of cryptanalysis because the attacker has minimal information to work with. Unlike chosen CPA or KPA, where the attacker has more control or information, COA provides the least amount of leverage for the attacker. If a cryptographic algorithm can resist a ciphertext-only attack, it's generally considered to be quite secure.

One of the earliest works on optical encryption is [6]. The authors place independent white uniformly distributed phase masks in the input and Fourier planes of a 4f correlator to encode images. However, this approach only works for coherent illumination, and is also susceptible to autocorrelation-based COAs and impulse-based CPAs [19]. [19] improves on [6] by adding a third independent white uniformly distributed phase mask in the image plane of the 4f correlator to inhibit impulse-based CPAs [19]. In [20], the authors place a random white uniformly distributed phase mask in the aperture plane of a camera to produce an incoherent imaging system for optical image encryption and demonstrate that such a system is susceptible to autocorrelation-based COAs. In [8], the authors simulate an experimental 3D printed optical fiber bundle [21] at the imaging plane of a camera to produce pixel shuffling. However, such optical fiber bundles have only been shown for parallel fibers, which are not suitable for pixel shuffling [21]. Our framework employs standard optical masks that are cost-effective and easily mass-produced, and we construct a prototype encryption camera. Works like [7], [9]–[11] use single optical masks for encryption. Although these methods are practical, they are not robust to various decryption attacks. In comparison, our solution is not only practical but also robust to these attacks. It should also be noted that none of the above existing works provide a complete analysis of the different forms of attacks that need to be considered for an optical encryption system, especially the powerful learning-based attacks that exploit hidden priors in data.

B. Lensless Imaging

In a conventional single-mask lensless camera like [1], [2], the lens of the camera is replaced by a thin optical mask placed at some distance from the sensor that modulates the incoming light. For a scene X, the measurement recorded by a sufficiently large sensor Y is given by:

$$Y = P * X + N, \tag{1}$$

where $*$ is the full-size convolutional operator (no cropping due to finite sensor size), P is the point spread function (PSF), and N is additive noise. The PSF is the response of the camera to a point source, and it depends on the mask pattern. The optical mask can be implemented using an amplitude mask that attenuates the incoming light [1] or a phase mask that modulates based on diffraction [2]. Due to the large PSF size and limited angular response of the pixels [2], [3], the convolution follows a zero-padded boundary condition. Existing works like [1]–[3] use a similar model for thin 2D and 3D lensless imaging. These lensless imagers were developed for imaging applications, and their ability to perform optical encryption has not been fully explored.

FlatCam [1] is a lensless camera that places a separable coded amplitude mask above a bare sensor array to enable a thin and flat form-factor imaging device, which can simulate a conventional camera by reconstructing conventional images from coded measurements. DiffuserCam [3] and PhlatCam

[2] replace the coded amplitude mask from FlatCam [1] with a coded phase mask for improved light efficiency and reconstruction quality. Spectral DiffuserCam [5] exploits the multiplexing ability of lensless imagers to do hyper-spectral imaging. OPENCAM is similar to these methods in that a coded optical mask is used to capture coded measurements. However, OPENCAM design goes a step further - our aim is not only to enable high-quality reconstruction of conventional images but also to prevent decryption attacks. To this end, we employ a second optical scaling mask positioned flush against the bare sensor array and propose novel mask designs for the multiplexing and scaling masks.

C. Image Reconstruction

Image reconstruction is a core problem in computational imaging, and plays a key role in lensless imaging. The problem of scene reconstruction involves the computational recovery of the latent scene from a lensless capture. Conventional approaches for lensless image reconstruction are mostly based on regularized least squares [1], [3], [4], [22]. More recently, deep-learning-based reconstruction methods have proposed [23]–[26]. We employ both regularized least squares and deep-learning-based reconstruction methods.

III. ATTACK MODEL

Since optical elements behave as linear operators [27], optical encryption cameras will always be susceptible to chosen-plaintext attacks in which an attacker uses a large set of ciphertext-plaintext pairs to estimate the encryption function. Naturally, this limits the practical use of optical encryption cameras to settings where attackers lack physical access to the camera location. Accordingly, in this paper, we assume that attackers do not have physical access to the camera locations, but can access a camera's stream. In other words, attackers have access to ciphertexts (measurements) from a camera, but not their corresponding plaintext (raw images), except for three special cases which we detail below. This is a reasonable assumption for many settings, such as security cameras in homes and offices.

In this paper, we consider four types of attacks: ciphertext-only attacks (COA) and three special cases of known plaintext attacks, namely, impulse known plaintext attack (I-KPA), uniform known plaintext attack (U-KPA), and uniform-impulse known plaintext attack (UI-KPA). The reason we consider the known impulse/uniform plaintext settings is that bright impulse-like illumination sources and uniform backgrounds often appear naturally in scenes and may be recognizable in the corresponding encrypted sensor measurement. Hence, attackers may gain access to an approximate impulse or uniform response of the sensor, without having physical access to the sensor.

Let $C : \mathbb{R}^{W_1 \times H_1 \times 3} \times \mathbb{R}^{W_2 \times H_2 \times M} \rightarrow \mathbb{R}^{W_3 \times H_3 \times 3}$ denote an optical encryption camera that maps a plaintext (scene) $X \in \mathbb{R}^{W_1 \times H_1 \times 3}$ and key $K \in \mathbb{R}^{W_2 \times H_2 \times M}$ to the ciphertext (measurement) $Y = C(K, X)$. The goal of an attacker is to recover the plaintext X from ciphertext Y with knowledge of function C, but partial or no knowledge of key K.

For each of the four attack types, we train a Dense Prediction Transformer (DPT) [28] D to learn the inverse mapping from $B = \{Y, \hat{X}, \hat{K}\}$ for arbitrary key K to the corresponding plaintext X, where Y denotes a single cipherext, $\hat{X}$ an estimate of the plaintext, and $\hat{K}$ an estimate of the key. For the COA, no estimates $\hat{X}$ and $\hat{K}$ are available, so B reduces to Y. For the KPAs, the methods for estimating $\hat{X}$ and $\hat{K}$ depend on the design of the targeted optical encryption camera (See section V for additional details.)

To train D, we generate a random key for each ciphertext sample in each batch and use a loss consisting of a combination of an L1 pixel loss and an L2 perceptual loss (as in [29], [30]) over the outputs of layers *relu1_1*, *relu2_2*, and *relu3_3* of VGG16 [31] pre-trained for image classification on the ImageNet [32] dataset. Concretely, the loss is given by

$$\mathcal{L}_D = w_1 ||D(B) - X||_1 + w_2 \sum_{i=1}^{3} ||\phi_i(D(B)) - \phi_i(X)||_2^2, \quad (2)$$

where $w_1 = 0.5$, $w_2 = 1.2$, and $\phi_1 : \mathbb{R}^{H \times W \times 3} \rightarrow \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times 64}$, $\phi_2 : \mathbb{R}^{H \times W \times 3} \rightarrow \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times 128}$, and $\phi_3 : \mathbb{R}^{H \times W \times 3} \rightarrow \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times 256}$ denote the layers *relu1_1*, *relu2_2*, and *relu3_3*, respectively, of the pre-trained VGG16 network.

IV. OPENCAM

A. Imaging Model (Encryption Process)

We propose a novel double-mask sensor design with two co-axially placed masks - an optical scaling mask $S \in \mathbb{R}^{W_3 \times H_3 \times 3}$ positioned flush against the sensor array and an optical multiplexing mask positioned a few millimeters above the scaling mask. The scaling mask is implemented via an amplitude mask. For the multiplexing mask, an amplitude or phase mask can be used, as described in the previous section. Let $P \in \mathbb{R}^{W_2 \times H_2 \times 3}$ denote the point-spread-function (PSF) produced by the multiplexing mask. Then, P and S constitute the "encryption key" of the system, and the encrypted measurements $Y \in \mathbb{R}^{W_3 \times H_3 \times 3}$ captured by the sensor are given by

$$Y = S \cdot (P * X) + N, \quad (3)$$

where $X \in \mathbb{R}^{W_1 \times H_1 \times 3}$ denotes the scene and $N \in \mathbb{R}^{W_3 \times H_3 \times 3}$ the sensor noise. It should be noted again that $*$ represents a full-sized convolution with zero-padded boundary conditions similar to the model used in existing thin lensless cameras.

Compared to single-mask lensless designs [1], [2], [10], [11], [20], our novel double-mask design has a larger key space, which makes decryption attacks more challenging, and also inhibits impulse and uniform known plain text attacks. For the single mask design, the encryption key (i.e., PSF P) is equal to the impulse response of the camera, so the IKPA completely compromises the system. This is not the case for our double-mask design, as the scaling mask makes the system shift-variant, so the impulse response only reveals the "effective" encryption key for a single scene pixel.

Imaging a uniform scene. We refer to the response of the camera to a uniform scene as Uniform Scene Response (USR).

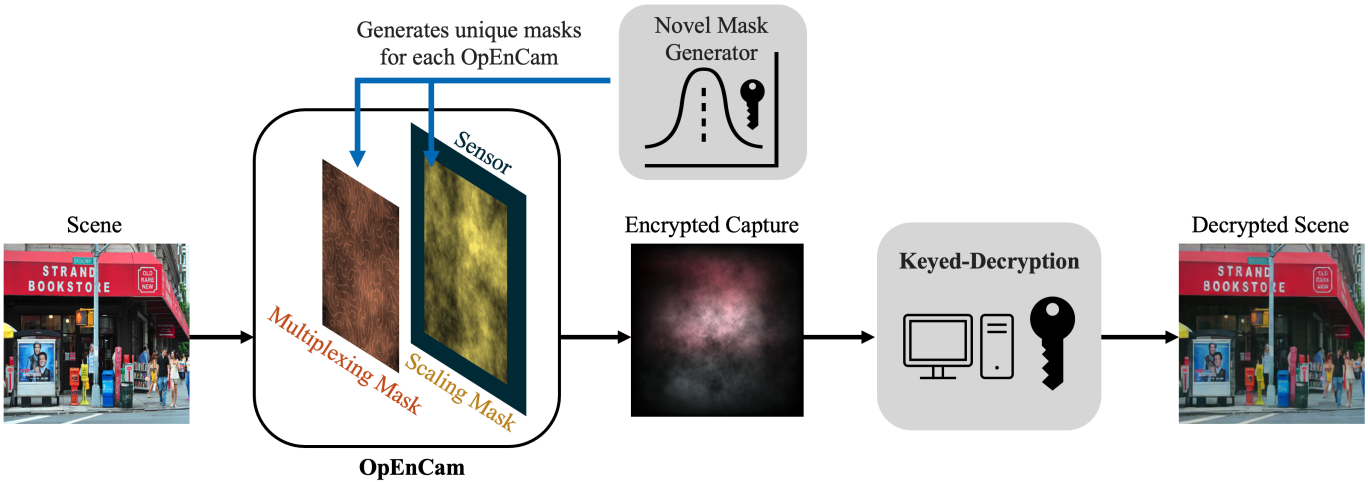

Fig. 2. **OpEnCam Framework.** Our **Op**tical **En**cryption **Cam**eras encrypt image data before image capture using optical masks. Recovery of decrypted image data is possible only if you have access to the camera's encryption key, which is defined by the unique optical elements of each camera. The optical mask patterns are generated by a novel mask generator that is robust against various ciphertext-only attacks.

Because of the structure of OpEnCam forward imaging model, it is natural to wonder if a USR, due to scenes like a plain wall, can reveal the scaling mask S up to a scale. For a uniform scene to reveal the scaling mask S, $P*X$ must also be uniform. However, due to full-sized, linear convolution with zero-padded boundary conditions, $P*X$ is never uniform, even for a plain wall scene. Therefore, it is not possible to reveal S from a USR.

B. Mask Design (Key Generation)

For both single and double-mask optical encryption cameras, the design of the masks is critical for good performance, i.e., for enabling high-quality keyed decryption while also thwarting decryption attacks. Consider the design with a single multiplexing mask. The mask should have the following desirable properties:

1) PSF should contain directional filters for all angles to enable high-quality lensless reconstruction [2].
2) The autocorrelation of the PSF should not be an impulse. The PSFs for which the autocorrelation is an impulse, are susceptible to autocorrelation-based COAs [20]. These attacks exploit the fact that taking the autocorrelation of encoded measurements eliminates the mask component if the autocorrelation of the PSF is an impulse. This reduces the COA problem to recovering an image from its autocorrelation.
3) The PSF produced by the multiplexing mask should not be binary. Binary PSFs may be revealed to an attacker when a point source appears in the scene. Although our double-mask design inherently provides protection against such attacks, it is still not desirable for the PSF to be compromised, as it reduces the key size significantly. Consider a 1D case of binary PSF p, and positive scaling mask s. The measurement $y(n)$ recorded at the sensor due to a single point source is given by,

$$y(n) = \begin{cases} s(n), & \text{if } n \in \mathcal{J} \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

where $\mathcal{J}$ is the set of all pixel locations for which $p(n)$ is 1. Simple thresholding of y i.e. $y > 0$, reveals p and reduces the size of the key.

Keeping the above criteria in mind, we propose a novel design for the multiplexing mask. The PSF due to OpEnCam multiplexing mask is given by,

$$P = \alpha P_{colr} + (1-\alpha)P_{cont}, \tag{5}$$

where P_{cont} is a binary contour PSF obtained from Perlin noise [33] proposed in [2], P_{colr} is a colored noise with certain roll-off. Perlin contours are known for high-quality lensless imaging as shown in [2], [23] but have impulse-like autocorrelations. On the hand, P_{colr} is smoother and has non-impulse autocorrelation with varying roll-off. We fix the Perlin feature size, randomize the permutation vector for Perlin noise, and correspondingly the contour PSF P_{cont}. The length of the permutation vector is the same as the height or width of the PSF. To generate P_{colr}, we generate colored noise with Power Spectral Density (PSD) given by,

$$H_\beta(u,v) = \frac{1}{(u^2+v^2)^{(\frac{\beta}{2})}}. \tag{6}$$

The corresponding colored noise $P_{colr}(\beta)$ is given by,

$$P_{colr}(\beta) = \mathcal{F}^{-1}(\sqrt{H_\beta}e^{j\theta}). \tag{7}$$

Here $\mathcal{F}^{-1}$ is the inverse FFT, θ is random white noise sampled from $\mathcal{U}(0,1)$ and β is a scalar hyper-parameter that controls the roll-off of the colored noise. We randomly sample β from $\mathcal{U}(1,10)$. Finally, the linear combination co-efficient α is chosen uniformly at random. Apart from the above criteria, it should also be noted that PSF must be non-negative as we can not subtract light.

For the scaling mask S, we use colored noise again without the Perlin contours. For a given pair of (S,P), the color of the noise (β) is the same for both S and P_{colr} to avoid attacks due to filtering of the either component. It should be noted that S has to be positive as we do not want to throw information

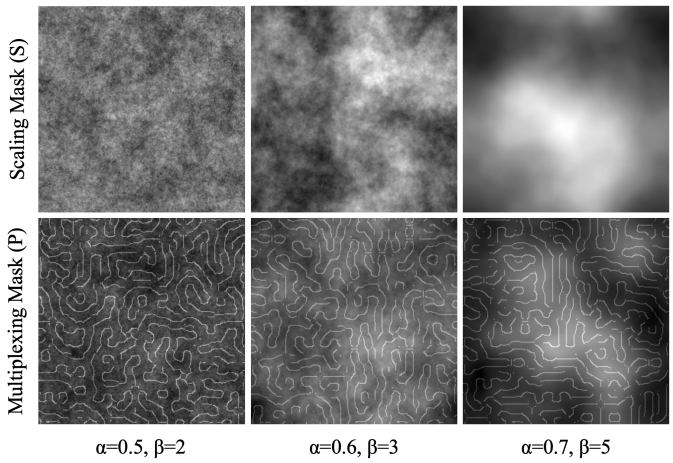

Fig. 3. **Generated Mask Samples.** Each column shows a scaling mask and multiplexing mask pair generated from our key generator. Each of these is randomly generated. Moreover, the underlying Perlin contours in the multiplexing mask PSF are also randomized.

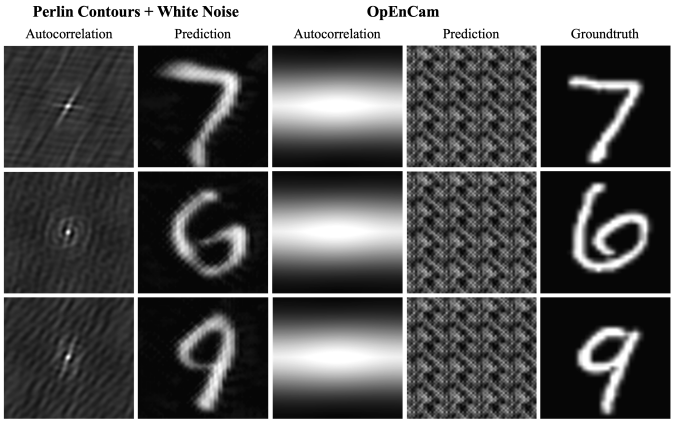

Fig. 4. **Autocorrelation Analysis.** It is not possible to reconstruct the underlying plaintext/scene from the measurement autocorrelation for the OpEnCam design P_O. While the reconstruction is possible for P_W (Perlin Contours + White noise).

from any of the sensor pixels. Some samples of the generated mask patterns are shown in fig. 3.

C. Effect of Colored Noise PSF - Autocorrelation Analysis

In the above subsection, we claimed that autocorrelation of the multiplexing mask should not be an impulse and, as a result, settled on a combination of Perlin contour and colored noise for the multiplexing PSF of OpEnCam. In this experiment, we validate this claim by comparing the performance of two different multiplexing mask designs under autocorrelation-based ciphertext-only attack. The mask designs being compared are (a) $P_O = \alpha P_{colr} + (1-\alpha)P_{cont}$, (b)$P_W = \alpha P_{white} + (1-\alpha)P_{cont}$, where P_{white} is white noise PSF with impulse autocorrelation. P_O is the OpEnCam design while P_W is a corresponding version with white noise used instead of colored noise for PSF. To perform this experiment, we first simulate measurements using each of the mask generators and perform autocorrelation of the simulated measurements. We then feed this autocorrelation to a U-Net, which then learns to estimate the underlying scene or plaintext from the autocorrelation. Given, that the autocorrelation of P_W is close to an impulse (because autocorrelations of P_{white} and P_{cont} are impulse), U-Net finds it easy to learn the mapping from the measurement autocorrelation to the underlying scene for P_W, while it fails to do so for P_O. We perform this experiment on MNIST dataset. Fig. 4 shows the visual results for the experiment. As can be seen, it is possible to reconstruct the underlying scene from the autocorrelation of the measurement for P_W, while this is not the case for P_O i.e. our OpEnCam design.

Encryption Design	PSNR ↑	SSIM ↑	LPIPS ↓
PhlatCam [2]	**23.10**	**0.81**	**0.30**
Multi-pinhole [10]	22.70	0.79	**0.30**
FlatCam [9]	20.68	0.74	0.40
Random Binary [11]	20.22	0.77	0.35
Random Speckle [7]	20.43	0.71	0.44
OpEnCam	<u>22.90</u>	<u>0.80</u>	<u>0.32</u>

TABLE I
Image Recovery via Keyed Decryption. The reconstruction quality of **OpEnCam** is at par with state of the art lensless image reconstruction quality. Best performance is made bold while the second best performance is underlined.

D. Image Reconstruction (Keyed Decryption)

The decryption module accepts three inputs, an encrypted image Y, a scaling mask S, and a PSF P due to the multiplexing mask M; and returns a decrypted image $\hat{X}$. For an ideal system, decryption can be achieved by simply applying the inverse of the two optical computations. However, in practice, to account for noise, cropping, and imperfect calibration, we instead employ a double-step estimation process to recover the scene X. In the first step, we employ Tikhonov regularized least squares to recover an initial estimate $\hat{X_T}$ of image X:

$$\hat{X}_T = \arg\min_X ||Y_N - P * X||_F^2 + \gamma ||X||_F^2 \tag{8}$$

Here Y_N is the measurement after scaling normalization i.e. $Y_N = \frac{Y}{S+\epsilon}$. $\hat{X}_T$ has a closed-form solution and can be implemented in the Fourier domain as Wiener filtering.

In the second step, we feed our initial estimate $\hat{X}_T$ plus measurement Y, and multiplexing mask PSF P to a Dense Prediction Transformer (DPT) [28] $D : \mathbb{R}^{W_1 \times H_1 \times 7} \rightarrow \mathbb{R}^{W_1 \times H_1 \times 3}$, which we train via stochastic gradient descent to produce a refined version $\hat{X}$, given input $B = \{Y, \hat{X}_T, P\}$. As in eq. (2), we use a combination of L1 loss and VGG feature loss to learn the mapping with $w_1 = 0.5$ and $w_2 = 1.2$. We found that this transformer-based refinement performed slightly better than the U-Net-based refinement typically used in lensless imaging works like [23], [34], [35].

V. Experimental Results

A. Preliminaries

1) Baselines: We compare the performance of OpEnCam design with other mask-based passive optical encryption designs as well as existing lensless camera designs. Our baselines

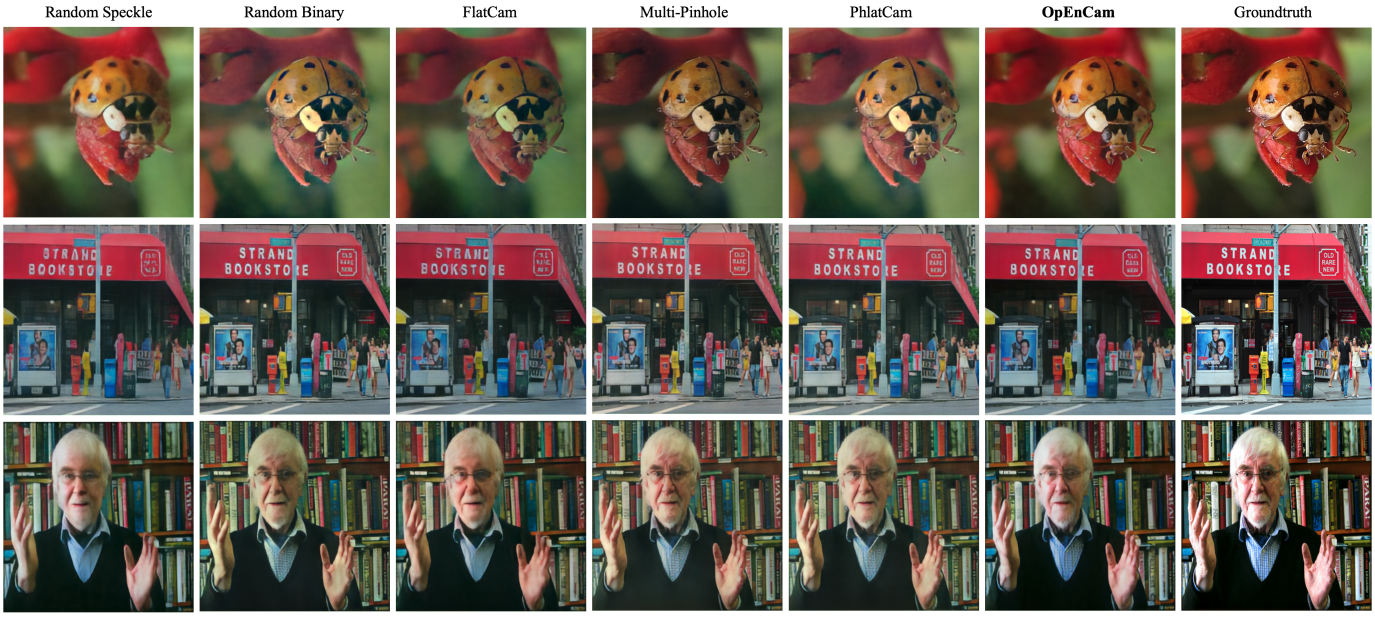

Fig. 5. **Keyed Decryption.** The first five columns show the keyed-decryption performance for existing single mask systems. Keyed decryption using PhlatCam (Boominathan *et al.* [2]) is currently state of the art for lensless imaging. OpEnCam keyed decryption performance matches that of existing mask designs for lensless imaging and optical encryption despite the addition of a scaling mask to enhance privacy.

for optical encryption include multi-pinhole mask generator (Ishii *et al.* [10]), random binary mask generator (Wang *et al.* [11]) and random speckle mask generator (Zang *et al.* [7]) We also compare against PhlatCam [2] lensless camera mask generator, which uses a Perlin contour PSF multiplexing mask, and FlatCam [1], [9], which uses separable Maximum Length Sequences (MLS) for mask patterns.

2) Dataset and Implementation: We validate the efficacy of our proposed framework via simulation and real-world experiments. For training keyed-decryption and attack models described in this section, we use the Places365 dataset [36]. We use a subset of ImageNet [37] for quantitative evaluation of the keyed-decryption and attack models. Given that each encryption approach (including OpEnCam) has its own mask generator that randomly generates the corresponding mask patterns, we simulate the captured measurement/ciphertext using these randomly generated mask patterns using the above datasets. Once simulated, we feed these measurements (along with the mask patterns for keyed-decryption) to the corresponding neural networks for final predictions. During testing, we generate predictions for a large number of mask patterns generated from each mask generator and report the average performance.

B. Keyed Decryption

Given that high-quality scene recovery using the key or mask patterns defines the usefulness of our system, we first experimentally validate the performance of keyed decryption transformer. We compare the keyed-decryption performance of OpEnCam against PhlatCam [2], Multi-pinhole [10], FlatCam [9], Random Binary [11] and Random Speckle [7]. We use the same keyed-decryption strategy as described in Section IV-D for a fair comparison. The keyed-decryption network uses Tikhonov reconstructions described in Equation 8 along with the measurement and the multiplexing mask as input and produces the final reconstructions. We use DPT [28] as the transformer. We compute the average Peak Signal to Noise ratio (PSNR), Structural Similarity Index Measure (SSIM) and Learned Perceptual Image Patch Similarity (LPIPS) for a subset of ImageNet dataset as described in [23]. Table I shows the average performance. Higher PSNR, SSIM and lower LPIPS indicate better reconstruction quality. The performance of OpEnCam is at par with state of the art lensless image reconstruction performance [2], [34] while marginally outperforming existing encryption methods like [7], [9]–[11]. A similar trend is observed in visual results shown in fig. 5.

Encryption Design	PSNR ↓	SSIM ↓	LPIPS ↑
PhlatCam [2]	<u>18.10</u>	<u>0.63</u>	<u>0.57</u>
Multi-pinhole [10]	19.29	0.65	0.48
FlatCam [9]	19.00	0.65	0.54
Random Binary [11]	18.73	0.64	0.56
Random Speckle [7]	19.42	0.65	0.54
OpEnCam- Single Mask	18.22	0.64	<u>0.57</u>
OpEnCam- Double Mask	**16.28**	**0.59**	**0.60**

TABLE II
Transformer-based Blind COA. Double mask **OpEnCam** provides the best security as seen from the reconstruction quality. Large gap between single and double mask OpEnCam highlights the improvement offered by the scaling mask in terms of privacy. Best performance is made bold while the second best performance is underlined.

C. COA: Ciphertext Only Attack

We train a Dense Prediction Transformer(DPT) [28] to learn a mapping from a single ciphertext/measurement to the

Fig. 6. **Blind COA.**We show some visual reconstruction results for two scenes. Double-mask OPENCAM reconstructions are barely recognizable, while existing approaches reveal a significant amount of texture.

corresponding plaintext/scene for each of the mask generators using the loss function described in 2. A secure optical encryption system should be robust against these powerful data-driven transformer-based attacks. We compare the scene reconstruction performance for different optical encryption strategies. Among the existing works, we compare against PhlatCam [2], random speckle [7], multi-pinhole [10], random binary [11] and FlatCam [9]. We also compare against single-mask OPENCAM i.e. the OPENCAM design without the scaling mask to highlight the importance of the double-mask system. Lower PSNR, SSIM and higher LPIPS indicate lower reconstruction quality and better security. In fig. 8, we show the privacy-utility tradeoff plot. From the plot, it can be observed that OPENCAM clearly outperforms existing optical encryption and lensless imaging designs. The reconstructions from OPENCAM, shown in fig. 6, are barely recognizable, indicating that it is extremely challenging to recover the original scene from a single encoded measurement captured using OPENCAM when the key or mask patterns are not available. Table II shows the quantitative comparison of the different encryption approaches. Double Mask OPENCAM clearly outperforms all other existing optical encryption and lensless imaging designs. The large gap between OPENCAM single and double mask models suggests that the scaling mask not only makes the system shift-variant but also adds another layer of privacy to existing mask-based lensless camera designs.

D. I-KPA: Impulse - Known Plaintext Attack

An optical encryption camera with a single multiplexing mask is susceptible to an Impulse based Known Plaintext Attack (I-KPA). More specifically, if a bright source appears in a scene, an approximate point spread function (PSF) due to the multiplexing mask is revealed. Since for single mask-based encryption, the PSF is the key, the attacker can use a bright-source measurement to decrypt ciphertexts. Moreover, as pointed out in Section IV-B, binary masks are more susceptible to these bright source attacks even with a scaling mask. In this experiment, we show that our double-mask encryption of OPENCAM is robust against these bright source attacks.

Encryption Design	PSNR ↓	SSIM ↓	LPIPS ↑
PhlatCam [2]	19.04	0.76	0.38
Multi-pinhole [10]	22.61	0.85	0.22
FlatCam [9]	18.79	0.66	0.52
Random Binary [11]	19.28	0.72	0.43
Random Speckle [7]	19.04	0.65	0.53
PhlatCam - Double Mask	18.28	0.67	0.50
OPENCAM- Single Mask	<u>18.02</u>	<u>0.66</u>	<u>0.57</u>
OPENCAM- Double Mask	**16.28**	**0.59**	**0.60**

TABLE III
IMPULSE - KPA. ATTACKER HAS ACCESS TO AN APPROXIMATE PSF P CORRESPONDING TO A SINGLE BRIGHT SOURCE. DOUBLE-MASK OPENCAM OUTPERFORMS ALL EXISTING DESIGNS. PHLATCAM EVEN WITH A SCALING MASK (PHLATCAM-DOUBLE) IS SUSCEPTIBLE TO THIS ATTACK DUE TO ITS SPARSE BINARY PSF. BEST PERFORMANCE IS MADE BOLD WHILE THE SECOND BEST PERFORMANCE IS UNDERLINED.

To perform this experiment, we first simulate a measurement of a scene with a bright source. The relative intensity of the bright source with respect to the rest of the scene is 10^3. Assuming this measurement as the approximate PSF, we apply our keyed decryption pipeline of Section IV-D. We compare against the baselines described above: PhlatCam [2], multi-pinhole [10], random binary [11], FlatCam MLS [9]. To highlight the importance of our double-mask system we also compare it against single-mask version of OPENCAM without the scaling mask. Furthermore, to highlight the disadvantage of binary masks like that of PhlatCam under point source attacks, we compare against a modified version of PhlatCam where we add an additional scaling mask. We show the privacy-utility tradeoff plot for the I-KPA in fig. 8. Visual reconstruction results are shown in fig. 7. It can be seen that OPENCAM with scaling mask significantly outperforms all existing mask approaches. Moreover, despite the use of scaling mask, PhlatCam-Double Mask is unable to provide any privacy due to the binary nature of its multiplexing mask. Corresponding visual results for a few important approaches are also compared in the same figure. We show the quantitative results for the I-KPA in table III.

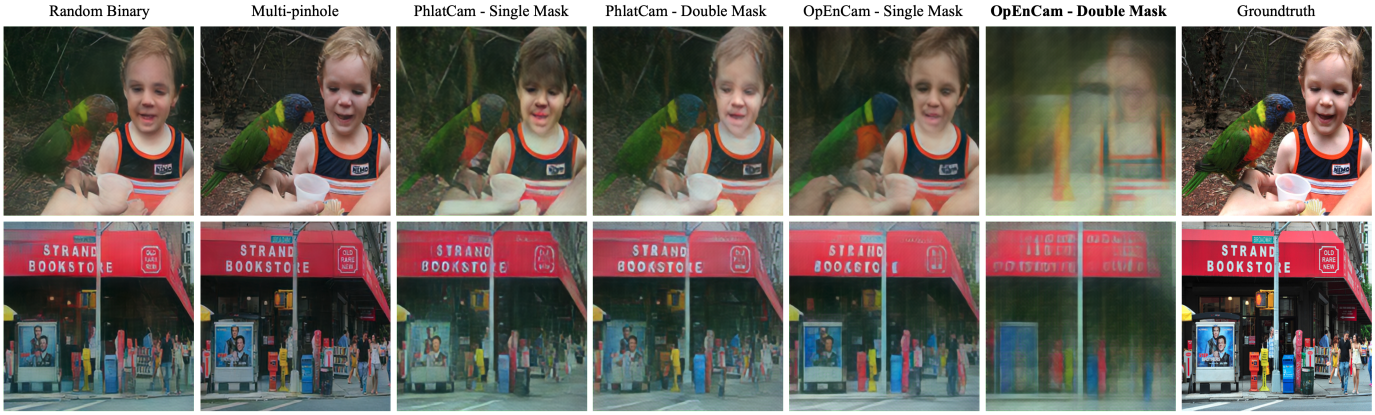

Fig. 7. **Impulse - KPA.** We show some visual reconstruction results for two scenes. Double-mask OPENCAM reconstructions are barely recognizable, while other designs are completely susceptible to this attack, including Double-masked PhlatCam.

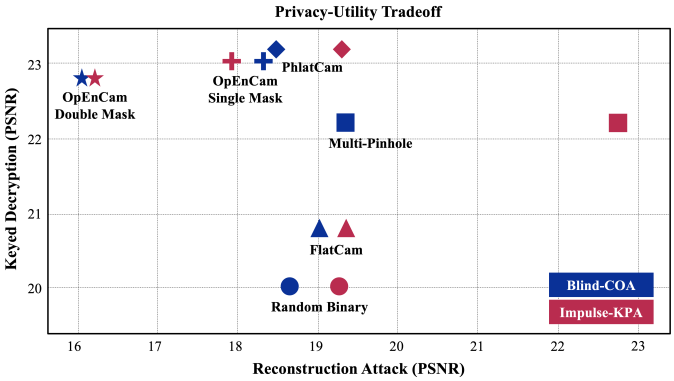

Fig. 8. **Privacy-Utility Plots.** We show the privacy-utility plot for blind COA and Impulse-KPA. For both attacks, double-masked OPENCAM, outperforms other designs and occupies the desired top-left corner.

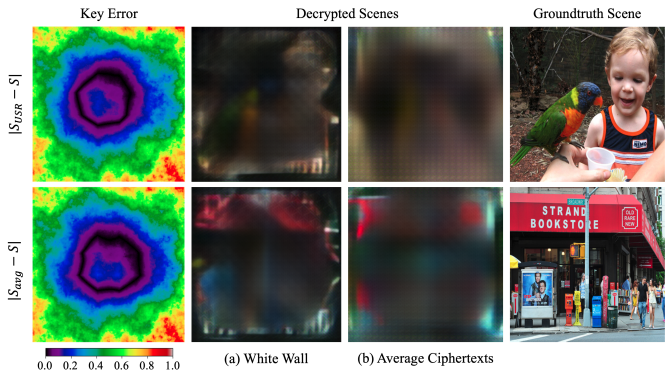

Fig. 9. **Uniform - KPA.** (a) From a USR, (b) from ciphertexts. It can be seen that both the USR and the average of measurements fail to reveal the underlying hidden scaling mask leading to poor plain text estimations.

E. U-KPA: Uniform - Known Plaintext Attack

1) Attack due to uniform scene: To verify that a uniform scene response (USR) doesn't reveal the scaling mask, we approximated the scaling mask S as $S_{USR} = Y_{USR}$, where Y_{USR} is a measurement for an all-ones scene. We used S_{USR} to remove the scaling mask component from the test measurement and decrypted it using the blind COA transformer trained for single-mask OPENCAM. We show the decrypted result and the error $|S_{USR} - S|$ in fig. 9(a). It can be seen that S_{USR} is not an accurate approximation of S.

2) Attack due to multiple measurements: Another question that may arise is whether an average of diverse measurements reveal the scaling mask. To verify this, we assume access to $N = 12000$ diverse OPENCAM measurements of scenes from the Places365 dataset and estimate the scaling mask S as $S_{avg} = (1/N)\sum_{j=1}^{N} Y^{(j)}$. We then used S_{avg} to remove the scaling mask component from the measurement and decrypted it using the transformer trained for single-mask OPENCAM. We show the results in fig. 9(b). It can be seen that S_{avg} is not an accurate approximation of S. This attack assumes that under the condition $N \to \infty$, S_{avg} will be approximately a USR.

F. UI-KPA: Uniform-Impulse Known Plaintext Attack

In this attack, the attacker tries to jointly estimate the scaling mask S and PSF P and then uses them to decrypt the ciphertext. We assume the attacker has access to the approximate impulse response for a scene with a dominating bright source, and additional information in the form of a white-wall-measurement. We then solve the optimization,

$$\{\hat{S}, \hat{P}\} = \arg\min_{S,P} \sum_{i=1}^{2} ||Y_i - S \odot (P * X_i)||_F^2, \qquad (9)$$

to estimate the keys. Y_1 is a USR, and Y_2 is the captured impulse response, while X_1 and X_2 are all-ones images and impulse images respectively. Y_1 can also be an average measurement corresponding to a large diverse set of natural scenes. Once S and P are estimated, the attacker uses the keyed-decryption framework of Section IV-D to decrypt the ciphertexts/measurements. We show the results for different relative intensities of the point source with respect to the rest of the scene in fig. 10. For cases where the relative intensity of the bright source is less than 10^4, it is not possible to accurately estimate the scaling mask and the PSF, and as a result, the quality of the decrypted scene is poor. For relative intensities beyond 10^4, it is possible to reconstruct the scene

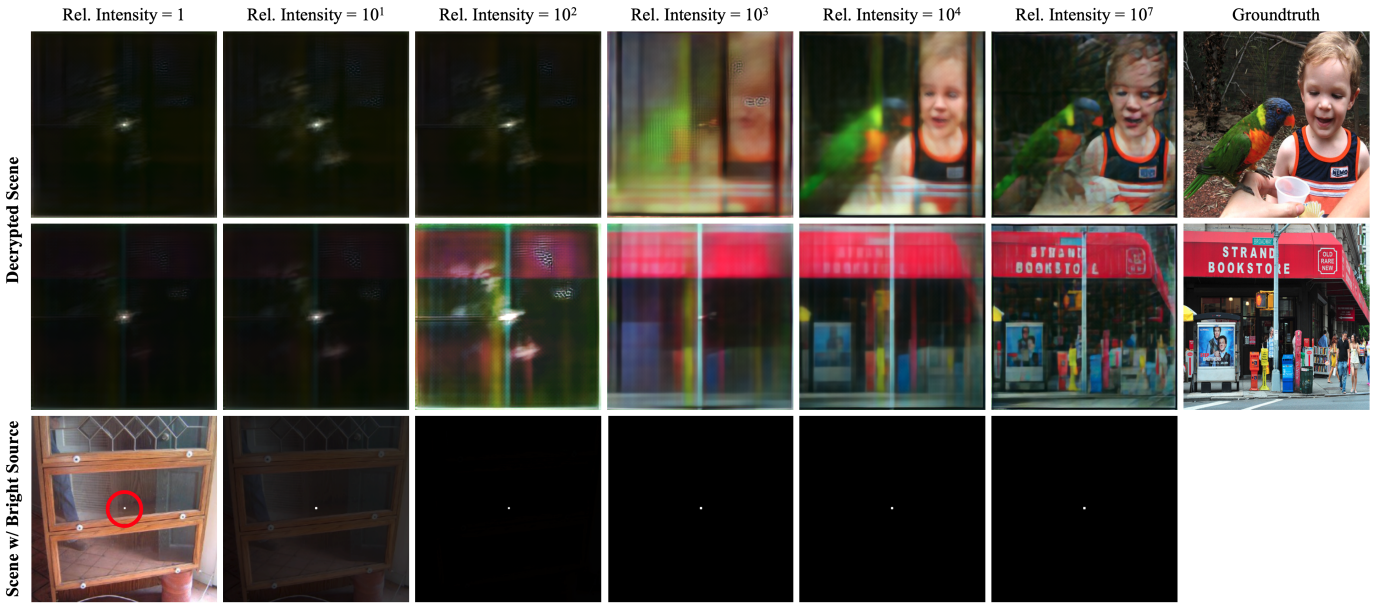

Fig. 10. **Uniform-Impulse - KPA.** (Top Rows) We show the decrypted scenes for varying relative intensities of the bright source. (Bottom Row) We show the scene with a point source (encircled) of varying relative intensity. OPENCAM offers privacy for a wide range of practical relative point source intensities.

to a reasonable extent. However, such large relative intensities of point sources imply dark-room-like environments which are less likely to occur in the attack model we have assumed i.e. the attacker doesn't have access to the physical location of the camera. Nevertheless, we acknowledge that this could be a limitation of OPENCAM and leave improving security for such scenarios as future work.

G. Real Experimental Results

For the real-world experiments, we construct a working prototype OpEnCam using Basler Ace4024-29uc machine vision camera. To implement the multiplexing mask, we use an amplitude mask printed using a conventional office printer and place it over a Perlin contour phase mask. The scaling mask S essentially implements a spatially varying exposure pattern for each sensor pixel in OPENCAM. Given that placing the scaling mask flush against the bare sensor requires breaking the sensor's protective covering, we implement its effect through exposure bracketing. That is, we quantize the scaling mask into 16 levels and then capture the measurement due to the multiplexing mask at 16 different relative exposure values given by the 16 levels of quantization. Finally, using the quantized scaling mask, we blend the measurement. We capture the experimental scaling mask (S_{exp}) by displaying a white image on a monitor and blending the measurements captured at 16 different exposure values without the multiplexing mask. We then capture the experimental PSF (P_{exp}) in the same way, but with the multiplexing mask, i.e., we capture the measurements for a point source at 16 different exposure values and blend them. The true PSF is then the normalized version of P_{exp} i.e., $P_{true} = P_{exp}/(S_{exp}+\epsilon)$. Finally, we display images of natural scenes on the monitor, capture the corresponding exposure stack, and blend it to obtain the experimental measurements. For keyed decryption on these experimental data, we use the Tikhonov regularized reconstruction (Eq. (8)) to obtain the intermediate scene estimates. Since these estimates are generally noisy, we then use a denoising U-Net to refine them. For blind transformer-based attack, we use the DPT trained in Section V-C on the experimental measurements. We show the visual results in fig. 11. These results demonstrate the efficacy of our approach – recovery of the original scene from an OpEnCam measurement is possible only if one has access to the camera's encryption key.

VI. CONCLUSION AND DISCUSSION

We propose a novel design for lensless cameras called OPENCAM to perform optical encryption. At the core of our method is the double-mask imaging model which implements shift-variant forward encryption process. Moreover, using ideas derived from signal processing, we design the mask patterns of our system to not only allow high-quality lensless imaging but also be robust against various forms of attacks. Through extensive experiments, we showed that our proposed OPENCAM design is robust against powerful learning-based ciphertext-only attacks and three special cases of known plaintext attacks. Finally, we build an OPENCAM prototype that allows us to validate our claims through promising preliminary results. Using a photolithography-based phase mask printing process [2], our prototype's performance can significantly improve, and we plan to implement this in the future.

Although preliminary results from OPENCAM show promising results, it should be noted that, like existing optical encryption systems, OPENCAM also has its limitations. Accessing the hardware or allowing extensive control of the scene has the potential to reveal the key. However, as pointed out through our experiments, the OPENCAM design is still an improvement over the existing optical encryption system and is a step towards more secure imaging.

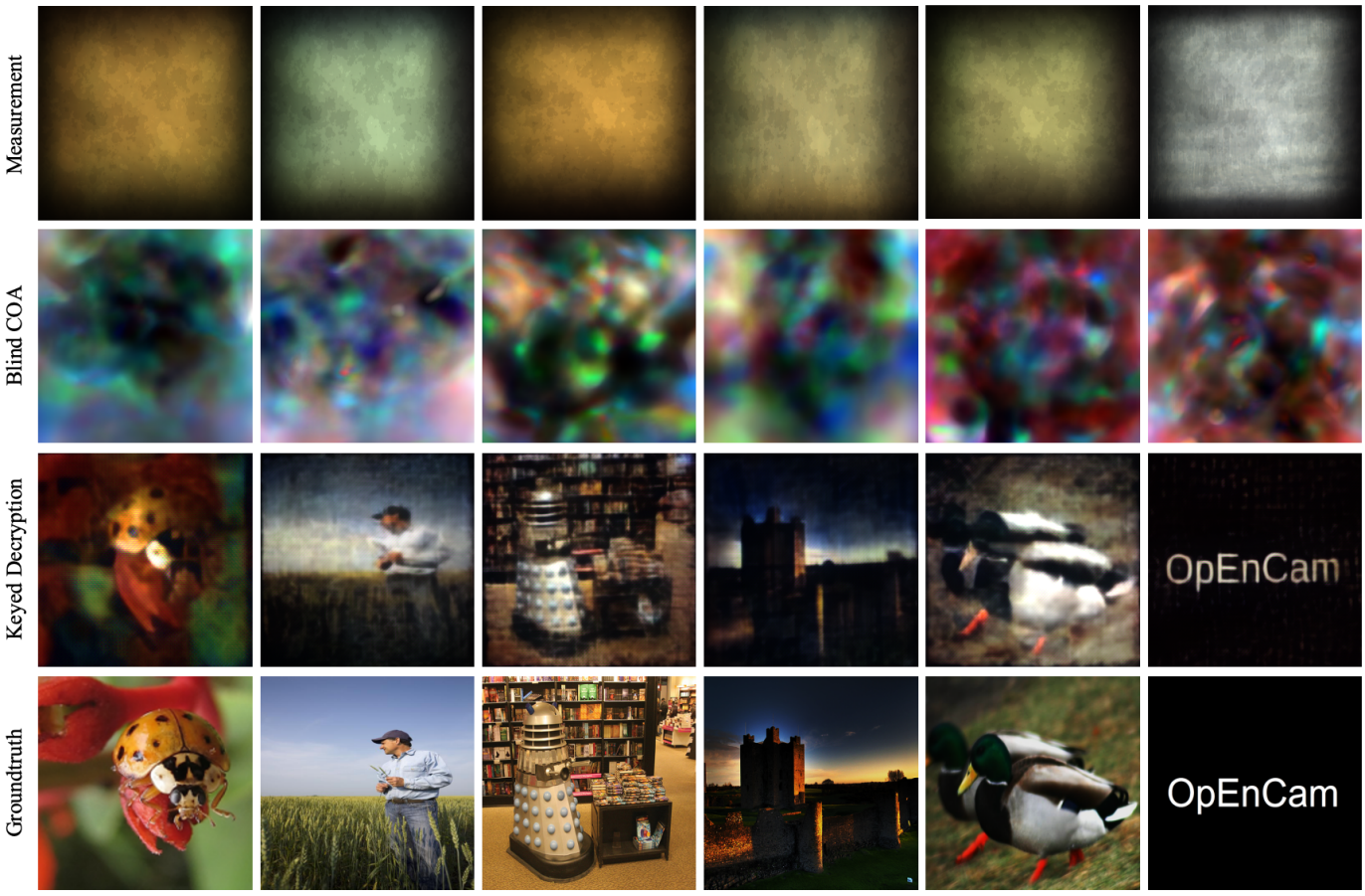

Fig. 11. **Real Results from OpEnCam Prototype.** The top row shows OpEnCam measurements. The second row shows the reconstructions from a DPT-based blind COA. The third row shows keyed reconstructions. The bottom row shows the ground-truth scene.

In future, it would be interesting to look into the data-driven design of OpEnCam systems and using multiple-masked lensless cameras for other applications.

References

[1] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk, "Flatcam: Thin, lensless cameras using coded aperture and computation," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 384–397, 2016. 1, 2, 3, 6

[2] V. Boominathan, J. K. Adams, J. T. Robinson, and A. Veeraraghavan, "Phlatcam: Designed phase-mask based thin lensless camera," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 7, pp. 1618–1629, 2020. 1, 2, 3, 4, 5, 6, 7, 9

[3] N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "Diffusercam: lensless single-exposure 3d imaging," *Optica*, vol. 5, no. 1, pp. 1–9, 2018. 1, 2, 3

[4] N. Antipa, P. Oare, E. Bostan, R. Ng, and L. Waller, "Video from stills: Lensless imaging with rolling shutter," in *2019 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2019, pp. 1–8. 1, 3

[5] K. Monakhova, K. Yanny, N. Aggarwal, and L. Waller, "Spectral diffusercam: lensless snapshot hyperspectral imaging with a spectral filter array," *Optica*, vol. 7, no. 10, pp. 1298–1307, 2020. 1, 3

[6] P. Refregier and B. Javidi, "Optical image encryption based on input plane and fourier plane random encoding," *Optics letters*, vol. 20, no. 7, pp. 767–769, 1995. 1, 2

[7] J. Zang, Z. Xie, and Y. Zhang, "Optical image encryption with spatially incoherent illumination," *Opt. Lett.*, vol. 38, no. 8, pp. 1289–1291, Apr 2013. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-38-8-1289 1, 2, 5, 6, 7

[8] J. Byrne, B. Decann, and S. Bloom, "Key-nets: Optical transformation convolutional networks for privacy preserving vision sensors," in *British Machine Vision Conference (BMVC)*, 2020. 1, 2

[9] T. Nguyen Canh and H. Nagahara, "Deep compressive sensing for visual privacy protection in flatcam imaging," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0. 1, 2, 5, 6, 7

[10] Y. Ishii, S. Sato, and T. Yamashita, "Privacy-aware face recognition with lensless multi-pinhole camera," in *European Conference on Computer Vision*. Springer, 2020, pp. 476–493. 1, 2, 3, 5, 6, 7

[11] Z. W. Wang, V. Vineet, F. Pittaluga, S. N. Sinha, O. Cossairt, and S. Bing Kang, "Privacy-preserving action recognition using coded aperture videos," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019. 1, 2, 3, 5, 6, 7

[12] X. Peng, H. Wei, and P. Zhang, "Chosen-plaintext attack on lensless double-random phase encoding in the fresnel domain," *Optics letters*, vol. 31, no. 22, pp. 3261–3263, 2006. 2

[13] M. Liao, D. Lu, W. He, and X. Peng, "Optical cryptanalysis method using wavefront shaping," *IEEE Photonics Journal*, vol. 9, no. 1, pp. 1–13, 2017. 2

[14] X. Peng, P. Zhang, H. Wei, and B. Yu, "Known-plaintext attack on optical encryption based on double random phase keys," *optics letters*, vol. 31, no. 8, pp. 1044–1046, 2006. 2

[15] U. Gopinathan, D. S. Monaghan, T. J. Naughton, and J. T. Sheridan, "A known-plaintext heuristic attack on the fourier plane encryption algorithm," *Optics Express*, vol. 14, no. 8, pp. 3181–3186, 2006. 2

[16] X. Peng, H.-Q. Tang, and J.-D. Tian, "Ciphertext-only attack on double random phase encoding optical encryption system," 2007. 2

[17] C. Zhang, M. Liao, W. He, and X. Peng, "Ciphertext-only attack on a joint transform correlator encryption system," *Optics express*, vol. 21, no. 23, pp. 28 523–28 530, 2013. 2

[18] X. Liu, J. Wu, W. He, M. Liao, C. Zhang, and X. Peng, "Vulnerability to ciphertext-only attack of optical encryption scheme based on double

random phase encoding," *Optics express*, vol. 23, no. 15, pp. 18 955–18 968, 2015. 2

[19] M. Liao, S. Zheng, S. Pan, D. Lu, W. He, G. Situ, X. Peng *et al.*, "Deep-learning-based ciphertext-only attack on optical double random phase encryption," *Opto-Electronic Advances*, vol. 4, no. 5, p. 05200016, 2021. 2

[20] M. Liao, W. He, D. Lu, and X. Peng, "Ciphertext-only attack on optical cryptosystem with spatially incoherent illumination: from the view of imaging through scattering medium," *Scientific Reports*, vol. 7, no. 1, pp. 1–9, 2017. 2, 3, 4

[21] Y. Wang, J. Gawedzinski, M. E. Pawlowski, and T. S. Tkaczyk, "3d printed fiber optic faceplates by custom controlled fused deposition modeling," *Optics Express*, vol. 26, no. 12, pp. 15 362–15 376, 2018. 2

[22] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 83–91, 2008. 3

[23] S. S. Khan, V. Sundar, V. Boominathan, A. Veeraraghavan, and K. Mitra, "Flatnet: Towards photorealistic scene reconstruction from lensless measurements," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 3, 4, 5, 6

[24] D. Bagadthey, S. Prabhu, S. S. Khan, D. T. Fredrick, V. Boominathan, A. Veeraraghavan, and K. Mitra, "Flatnet3d: intensity and absolute depth from single-shot lensless capture," *JOSA A*, vol. 39, no. 10, pp. 1903–1912, 2022. 3

[25] X. Pan, X. Chen, S. Takeyama, and M. Yamaguchi, "Image reconstruction with transformer for mask-based lensless imaging," *Optics Letters*, vol. 47, no. 7, pp. 1843–1846, 2022. 3

[26] J. D. Rego, K. Kulkarni, and S. Jayasuriya, "Robust lensless image reconstruction via psf estimation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 403–412. 3

[27] G. Wetzstein, A. Ozcan, S. Gigan, S. Fan, D. Englund, M. Soljačić, C. Denz, D. A. Miller, and D. Psaltis, "Inference in artificial intelligence with deep optics and photonics," *Nature*, vol. 588, no. 7836, pp. 39–47, 2020. 3

[28] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 179–12 188. 3, 5, 6

[29] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *CVPR*, 2017, pp. 4681–4690. 3

[30] A. Dosovitskiy and T. Brox, "Generating images with perceptual similarity metrics based on deep networks," in *Advances in Neural Information Processing Systems*, 2016, pp. 658–666. 3

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015. 3

[32] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *CVPR*, 2009, pp. 248–255. 3

[33] K. Perlin, "Improving noise," in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, 2002, pp. 681–682. 4

[34] S. S. Khan, V. Adarsh, V. Boominathan, J. Tan, A. Veeraraghavan, and K. Mitra, "Towards photorealistic reconstruction of highly multiplexed lensless images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7860–7869. 5, 6

[35] K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," *Optics express*, vol. 27, no. 20, pp. 28 075–28 090, 2019. 5

[36] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 6

[37] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015. 6