Practical Optical Camera Communication Behind Unseen and Complex Backgrounds

Rui Xiao[1], Leqi Zhao[1], Feng Qian[2], Lei Yang[2], Jinsong Han[1]*
[1]Zhejiang University, Zhejiang, China
[2]Ant Group, China
ruixiao24@zju.edu.cn, zhaoleqi@zju.edu.cn, youzhi.qf@antgroup.com,
yl149505@antgroup.com, hanjinsong@zju.edu.cn

ABSTRACT

Optical camera communication (OCC) holds potential for location-aware data transfer, facilitating applications such as localization and overlaying digital content for mixed reality experiences. However, existing OCC designs commonly require a clean background for reliable demodulation, rendering its use disruptive and impractical. To this end, we propose *WinkLink*, a novel OCC system capable of robust transmission behind complex backgrounds, even under low signal-to-noise ratio (SNR) conditions. We address the key challenge of extracting subtle signals in the lossy OCC channel by designing a two-stage deep neural network and a context-aware demodulation protocol. The proposed system is trained solely on a synthesized dataset yet generalizes effectively to unseen real-world backgrounds. Through experiments in 12 diverse environments, we demonstrate that *WinkLink* successfully transmits OCC signals under a low SNR of -20 dB, achieving a substantial 5.8 dB SNR gain. This low SNR translates to an extended distance to 5.5× of baseline (11m with a 10W LED transmitter) and negligible interference on concurrent vision applications. Finally, *WinkLink* proves its efficacy even when the device is moving, i.e., dynamic backgrounds, making it ready for deployment on mobile devices.

CCS CONCEPTS

• **Networks** → **Wireless access networks**; • **Computing methodologies** → **Computer vision**; • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; **Mixed / augmented reality**.

KEYWORDS

Optical Camera Communication; Low Signal-to-Noise Ratio; Mobile Interaction

ACM Reference Format:
Rui Xiao[1], Leqi Zhao[1], Feng Qian[2], Lei Yang[2], Jinsong Han[1]*. 2024. Practical Optical Camera Communication Behind Unseen and Complex Backgrounds. In *The 22nd Annual International Conference on Mobile Systems, Applications and Services (MOBISYS '24), June 3–7, 2024, Minato-ku, Tokyo, Japan.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3643832.3661866

*Jinsong Han is the corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
MOBISYS '24, June 3–7, 2024, Minato-ku, Tokyo, Japan
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0581-6/24/06
https://doi.org/10.1145/3643832.3661866

1 INTRODUCTION

The concept of converging the physical and digital worlds has captured substantial attention from both industry and academia. A trending example is the rise of mixed reality (MR) apps, which offer users a truly immersive interaction experience [1, 16, 67]. While integrating both worlds, a key problem to be addressed is to create a robust mechanism for *registering virtual content* on specific objects or at precise locations, without necessitating extensive training or intricate setup [46, 51, 64].

To create such a *cyber-physical hyperlink*, optical camera communication (OCC) emerges as a compelling technology [19, 34, 87]. OCC transforms everyday LED lights into transmitters, leveraging mobile device cameras as receivers. Distinguished from high-speed RF links, OCC inherently links received data to the transmitter's identity or location. By repurposing light, it eliminates the need for an additional wireless front end, effectively rendering any object near light as a portal to the digital realm. As a result, OCC extends the boundaries of connectivity for common camera-equipped mobile devices, such as phones, tablets, and MR headsets.

However, despite its intriguing potential, the practical deployment of OCC faces challenges. The operation principle of OCC is to encode data streams through modulating light blinking, which is captured by cameras as stripe patterns for demodulation. These stripes overlay the original photo, or *background*, where the background acts as "interference" for OCC [37, 81, 84]. Especially in complex backgrounds, the stripe pattern becomes significantly distorted. Consequently, current OCC implementations rely on a common assumption: *a clean background for demodulation*, so as to remove this interference. However, this assumption imposes a significant constraint on deployment. Specifically, capturing a clean background implies that the camera must be very close to a clean reflector, typically within 40 cm [81]. That is, to receive OCC data, the users have to go to a *pre-defined, small* region, requiring additional guidance and considerable user effort. It also disrupts the user experience, particularly when integrated with MR apps, conflicting with the goal of creating an immersive experience.

Although previous approaches, e.g., CORE-Lens [50], attempt to enable OCC in the presence of complex backgrounds, their practical implementation is constrained by the high signal-to-noise ratio (SNR) requirement, which demands OCC signals to be significantly stronger than ambient light. Therefore, the communication distance is still confined, i.e., around 1.4 meters. Meanwhile, the resulting strong stripe pattern is obtrusive, significantly degrading video

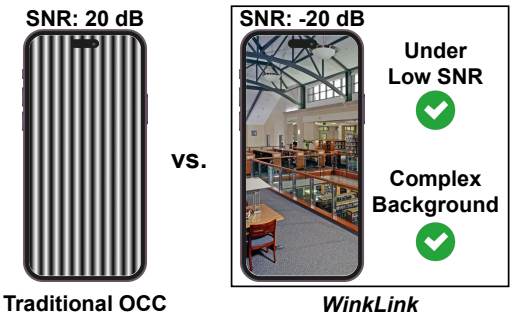


Figure 1: Figure depicts a comparison between *WinkLink* and traditional OCC. *WinkLink* features transmission behind unseen and complex backgrounds under low SNR conditions.

quality and adversely affecting concurrent vision applications. Finally, the effectiveness of CORE-Lens-like approaches is reliant on training the system with specific backgrounds, rendering them *unsuitable for unseen backgrounds* and diminishing practicality.

To unlock a seamless experience, we propose *WinkLink*, an OCC system that supports robust transmission behind *unseen complex* backgrounds even under *low-SNR* conditions. Figure 1 provides a comparative illustration of the usage scenarios between traditional OCC and *WinkLink*. *WinkLink* liberates users from the constraints of clean or trained backgrounds, facilitating easy deployment. Its low-SNR characteristic minimizes interference with concurrent vision applications. With a modest 10-watt light, *WinkLink* exhibits an impressive transmission range exceeding 11 meters. Additionally, our design accommodates *dynamic* backgrounds, enhancing device mobility by eliminating the requirement for a consistent background for decoding.

The development of *WinkLink* faces two primary challenges. The first challenge is to extract subtle signals from unseen, dynamic, and complex backgrounds, a task that can be formulated as an ill-posed problem [5]. What further complicates this challenge is the uneven entanglement of the signals with the background. Specifically, objects in the backgrounds carry different amounts of signals due to their diverse reflectance and distances from the light source [29]. The second challenge is the delayed response of the signal, leading to the lossy OCC channel. This lossiness is primarily caused by video compression algorithms, such as H.264 [28], which introduce temporal dependencies between frames, resulting in a delay in signal representation. This delay becomes particularly pronounced under low-SNR conditions. Therefore, determining the energy boundaries for accurate demodulation becomes challenging.

To address the first challenge, we leverage a fundamental property of OCC-encoded images: the replication of signals within each row of each color channel, leading to channel-wise and spatial-wise correlations. Exploiting these correlations, we constrain the ill-posed problem by formulating a loss in deep neural network (DNN), which is also known for its capability to capture intricate global correlations in images [90]. To ensure the scalability of *WinkLink*, especially in unseen backgrounds, we design a data synthesis scheme based on an analysis of the light propagation model, encompassing critical factors like object reflectance and distance. By

Figure 2: Figure depicts the example OCC applications.

leveraging online images to synthesize train dataset, *WinkLink* generalizes well to real-world OCC frames under unseen backgrounds without requiring the manual collection of even a single image. Finally, we overcome the challenge of delayed signal response with a context-aware demodulation protocol. While traditional methods sorely depend on the signal intensity of the current frame for demodulation, we additionally consider the intensity disparity between consecutive frames. This inclusion of the first-order derivative is effective in alleviating the adverse impacts of video compression and bolstering transmission accuracy.

We implement *WinkLink*[1] and evaluate its performance by conducting real-world experiments across 12 diverse environments. Our evaluation, based on a substantial dataset of over 520,000 OCC-encoded video frames captured by phones, showcases *WinkLink*'s superior performance compared to previous approaches. Specifically, our findings indicate that: *WinkLink* 1) achieves symbol error rates below 0.01 for various unseen backgrounds with an average SNR of -20 dB, showing an SNR gain of over 5.8 dB, 2) exhibits robustness in dynamic motion scenarios, 3) achieves a 5.5× improvement in communication distance, and 4) induces minimal interference with concurrent vision applications. Through this work, we hope to demonstrate the feasibility of simplifying communication deployment through machine learning assistance, thereby suggesting a new avenue for providing system and networking support for spatial and pervasive computing on mobile devices.

2 PROBLEM FORMULATION

Before detailing *WinkLink*'s design, we briefly introduce OCC. Its core concept, SNR, is then defined and visualized. We then highlight the issues with existing designs through preliminary experiments and present our design goals.

2.1 OCC Preliminary

Optical camera communication (OCC) is a subset of visible light communication [55, 89], aligning with the IEEE 802.15.7 standard [10]. OCC stands out by utilizing commonly available commercial cameras found in phones, tablets, and other mobile devices as receivers. As shown in Figure 2, OCC has been employed in various applications, including indoor localization [60], MR content delivery [32, 64], and providing pervasive connectivity for low-end IoT devices [25].

Rolling Shutter Mechanism. Most existing OCC systems rely on the *rolling shutter mechanism* to receive the modulated LED blinking. It is a characteristic feature of CMOS image sensors in commercial cameras [47]. Instead of exposing the entire frame simultaneously,

[1]The source code and the synthetic dataset are released at https://github.com/ruixiao24/winklink-mobisys2024.

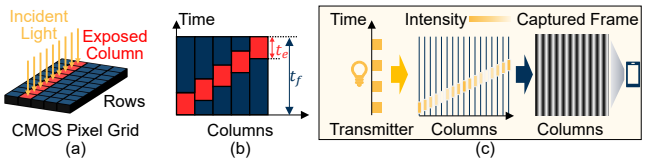


Figure 3: (a) and (b) illustrate that CMOS exposes the frame column by column, which acts like a sliding-window sampler. (c) illustrates that under RS-FSK modulation, the transmitter emits light periodically, which results in spatial periodicity in the captured frame.

rolling shutter sensors expose one column at a time, sequentially, to reduce caching-related overhead, as illustrated in Figure 3 (a) and (b). The intensity of each column can be understood as an integration or a moving-average filter across the exposure duration t_e. Assuming the sampling begins at time t, the intensity of column $A(t)$ can be expressed as:

$$A(t) = \int_{-\infty}^{\infty} I_L(\tau) g_r(\tau - t) d\tau. \quad (1)$$

Here, $I_L(\tau)$ represents the light intensity at time τ, and $g_r(\tau - t)$ is a gate (rectangular) function that equals 1 during the interval $(t, t + t_e)$ and 0 otherwise. This sequential recording mechanism empowers cameras to capture rapidly blinking light from the transmitter, imprinting it onto the captured image as stripes.

RS-FSK Modulation. A prevalent modulation technique in OCC is rolling-shutter frequency-shift keying (RS-FSK) modulation, wherein distinct symbols are conveyed by different frequencies of LED blinking (e.g., 1 kHz for symbol 0 and 1.5 kHz for symbol 1) [37]. As shown in Figure 3 (c), the temporal periodicity of the OCC transmitter's blinking translates into spatial periodicity within the captured image. Let $\mathcal{F}$ denote the set of frequencies used for modulation. Consequently, each symbol can encode $\lfloor \log_2 |\mathcal{F}| \rfloor$ data bits. In this paper, *WinkLink* employs a 4-FSK modulation scheme, where each symbol corresponds to 2 data bits.

2.2 SNR in OCC

We introduce the concept of the signal-to-noise ratio (SNR), which represents the relative strength of the received OCC signal and the accompanying background. When quantified in decibels (dB), the SNR is defined as:

$$SNR_{dB} = 10 \log_{10} \left(P_{signal} / P_{noise} \right), \quad (2)$$

where P_{signal} and P_{noise} denote the power of OCC signal and the background, respectively. In the context of OCC, the signal power is reflected in the value of individual pixels within images. Therefore, P_{signal} and P_{noise} are computed by summing all pixels corresponding to a symbol in images.

Why SNR? SNR plays a central role in OCC, as ***SNR fundamentally governs decoding complexity***. While various factors contribute to decoding complexity, SNR effectively ***encapsulates these variables***. Transmission power, distance, ambient light, camera settings, and other factors, indirectly impact decoding complexity by initially shaping SNR, solidifying SNR's paramount importance. Essentially, achieving transmission under lower SNR implies the capability to transmit with reduced power, over longer distances, in stronger ambient light, and with more flexible camera settings.

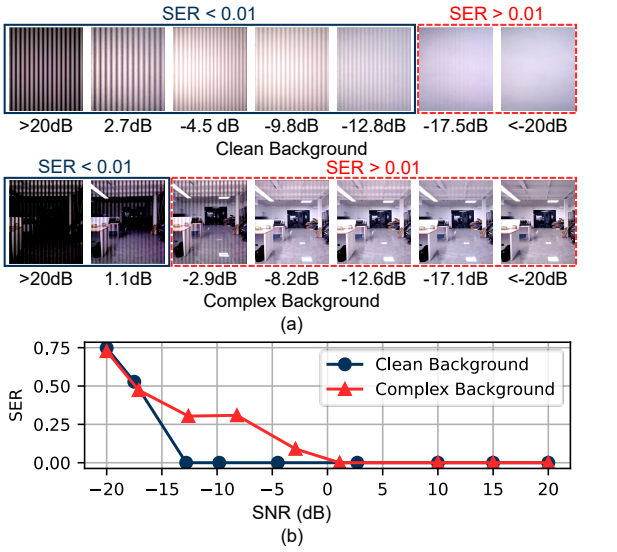


Figure 4: Figure depicts (a) received frames at different SNRs under clean and complex backgrounds, and (b) their respective SERs.

Visualizing Different SNRs. To visually illustrate the SNR in OCC, we display images across varying SNR levels in Figure 4(a). The noise power P_{noise} is computed when the OCC light is turned off. When the OCC light is activated, we calculate $P_{noise} + P_{signal}$. Subtracting these values yields P_{signal} and P_{noise} separately for SNR computation. This subtraction is applied *solely for SNR computation* and *not for decoding*. As SNR decreases, the stripe pattern gradually weakens. The SNR across these images is manipulated by adjusting the strength of ambient light and the OCC light source. It is important to note that for a given SNR, there are multiple parameter combinations that can result in the same SNR. However, this visualization offers a basic understanding of how video frames appear under different SNR conditions.

2.3 Issues with Existing Designs

To illustrate this, we conduct preliminary transmission experiments and present results in Figure 4(b). When demodulating on clean backgrounds, the signals are evenly distributed across the images. By summing up all the rows to aggregate the signals, the symbol error rate (SER) is lower than 0.01 for signals with SNR higher than -12.8 dB. However, under complex backgrounds, the current summing-up design fails to provide reliable results. The SER exceeds 0.01 when SNR drops below 1.1 dB. Therefore, existing designs require either a clean background or a high SNR for robust transmission, leading to various deployment limitations.

The requirement for clean backgrounds necessitates placing the camera in proximity to a clean reflector, with the distance limited to within 40 cm [81]. Note that digital zoom proves ineffective in addressing this concern, as it introduces signal loss during image

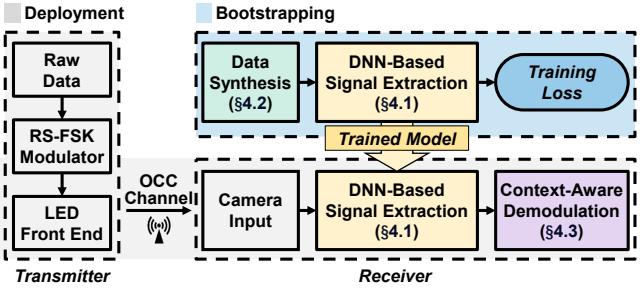


Figure 5: Figure depicts *WinkLink*'s design overview. During the *Bootstrapping Phase*, *WinkLink* trains the signal extraction model with synthesized data. In *Deployment Phase*, users utilize *WinkLink* on their mobile devices to demodulate real-world OCC frames.

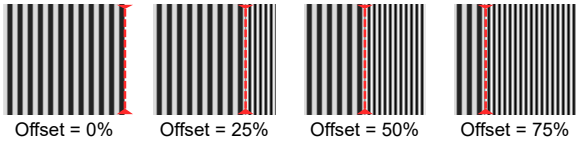


Figure 6: Figure depicts the mixed symbol effect with different amounts of offset.

cropping [37]. Moreover, this setup is often impractical, such as attempting to find such a reflector in a fully stocked grocery store.

The requirement for high SNR again constrains distance, observed at 1.4 meters in CORE-Lens [50], as light intensity attenuates rapidly with distance, following the inverse square law (intensity $\propto 1/\text{distance}^2$) [29]. Besides, the strong stripes lead to pronounced video degradation caused by prominent stripe patterns. This degradation significantly impairs core vision applications such as video recording and vision-based sensing, applicable to both line-of-sight[2] and non-line-of-sight OCC designs. Consequently, it renders vision application and OCC transmission mutually exclusive.

2.4 Design Goal of *WinkLink*

Given the above limitations, our goal is to design an OCC system capable of deployment in complex backgrounds under low-SNR conditions. The design strives for scalability on unseen backgrounds. Meanwhile, mobility and portability are essential objectives, requiring *WinkLink* to operate well in mobile scenarios, such as in handheld or moving cameras. This is also more challenging due to dynamic backgrounds, making techniques like background subtraction difficult to apply. The ultimate aim is to facilitate the practical and reliable deployment of OCC in real-world scenarios.

3 OVERVIEW AND CHALLENGES

We first present the design overview of *WinkLink* (§3.1) and present its main technical challenges (§3.2).

[2]In line-of-sight OCC designs, the background typically appears to be pure dark in order to correctly expose the stripes from lights, rendering it incompatible with concurrent vision application [19, 37, 87].

3.1 *WinkLink* Overview

We now provide a brief overview of *WinkLink*. The block diagram of *WinkLink* is shown in Figure 5, composed of an LED transmitter and a camera as the receiver. On the *Transmitter* side, the input data bit stream is modulated using the RS-FSK due to its robustness under low-SNR conditions.

The design of *Receiver* can be divided into two phases - *Bootstrapping* and *Deployment Phases*. *Bootstrapping Phase* occurs offline, entailing the generation of a substantial synthetic training dataset encompassing diverse scenarios (§4.2). This dataset serves as the basis for training a deep signal extraction model capable of extracting subtle OCC signals from encoded images (§4.1). In the *Deployment Phase*, *WinkLink* operates online, leveraging the trained model to demodulate the reflected OCC signals without requiring direct line-of-sight between the transmitter and receiver [50, 81], even in low-SNR conditions and against unseen backgrounds. After capturing the images from the camera, the encoded signals are extracted from the video frames using the trained model. The signals are then demodulated via the Context-Aware Demodulation module (§4.3).

3.2 Design Challenges

Designing *WinkLink* involves three main challenges.

3.2.1 Challenge I: Dynamic Signal Extraction. Low SNR results in a subtle stripe pattern in images, posing a highly challenging extraction task. This problem is *mathematically ambiguous* or *ill-posed* [5]. It can be formulated as to decompose the observed image $\mathbf{I}$ into the linear combination of the signal-free background $\mathbf{B}$ and the signal layer $\mathbf{O}$:

$$\mathbf{I}(\mathbf{x}) = \mathbf{B}(\mathbf{x}) + \mathbf{O}(\mathbf{x}), \tag{3}$$

where $\mathbf{x} = (x, y)$ is a 2D vector representing the coordinates (x, y) of a pixel's position in the image. Moreover, the background $\mathbf{B}$ is dynamic, varying across different image $\mathbf{I}$. To address this challenge, we employ a DNN-Based Signal Extraction module for extracting the signals (§4.1).

3.2.2 Challenge II: Scalable Training Data Preparation. High-quality training data is essential for the scalability of systems utilizing machine learning, including *WinkLink*. *WinkLink*'s scalability can be decomposed into four dimensions: ① unseen backgrounds, ② diverse stripe frequencies, ③ varying SNRs, and ④ different degrees of mixed-symbol effects[3] as illustrated in Figure 6. To achieve this scalability, it becomes imperative to build a dataset that incorporates these four types of diversities. However, it is laborious to obtain the ground-truth signal layer $\mathbf{O}$, especially under low-SNR conditions. The manual assembly of such a diverse dataset is time-intensive and impractical to conduct. *WinkLink* tackles this challenge with a Training Data Synthesis module (§4.2).

3.2.3 Challenge III: Delayed Signal Response. The reaction of signals exhibits a temporal lag. To illustrate this phenomenon, we measure the intensity transition between continuously transmitting symbol 1 and continuously transmitting symbol 0 in Figure 7 and 8. Ideally, the intensity associated with symbol 1 should instantly

[3]Mix-symbol effect refers to the scenario where a single image frame captures a combination of two or more consecutive transmitted symbols, typically arising from the asynchronization, or time offset, between the LED light and the camera receiver.

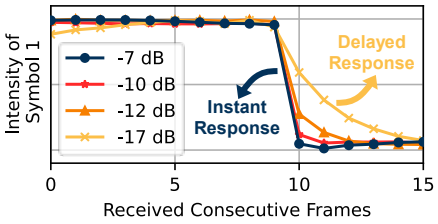


Figure 7: Figure depicts the delay of falling edges under different SNRs.

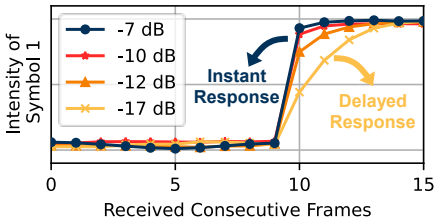


Figure 8: Figure depicts the delay of rising edges under different SNRs.

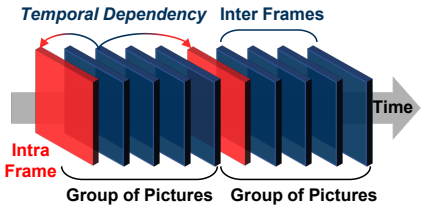


Figure 9: Figure depicts the temporal dependency between frames caused by video compression.

increase/decrease upon its appearance/disappearance. However, the observed images reveal a delayed response spanning up to five frames, particularly pronounced in low-SNR conditions, indicating an inverse relationship between response speed and OCC signal SNR. This delay issue is attributed to the compression process in video encoding. Specifically, video frames are categorized into intra frames and inter frames (Figure 9). Inter frames are reliant on the data within the intra frames, introducing *temporal dependency* and thus causing a lag in the signal response. Therefore, accurately determining intensity boundaries for demodulation poses a significant challenge. Meanwhile, recording uncompressed video is impractical due to latency, storage costs, and the need for *WinkLink* to operate concurrently with vision applications. We address this challenge by adopting a Context-Aware Demodulation module (§4.3).

4 DETAILED DESIGN

In this section, we present the design details of *WinkLink*'s modules shown in Figure 5.

4.1 DNN-based Signal Extraction

This module takes as input the encoded image $\mathbf{I}(\mathbf{x})$ taken by cameras and outputs one-dimensional OCC signals denoted as $\mathbf{y} \in \mathbf{R}^{1\times W}$, where W is the width of images. To address this ill-posed problem (formulated in §3.2.1), we implicitly enforce constraints using DNNs through the formulation of their associated loss functions [14]. Indeed, DNNs are particularly effective in our context due to a critical feature of OCC-encoded images: the replication of signals within each row of each color channel, resulting in *channel-wise and spatial-wise correlations*. DNNs excel at capturing such intricate global correlations in images. To enhance the model's scalability to unseen backgrounds, we adopt a two-stage network architecture, which first extracts the signal layer $\mathbf{O}$ from image $\mathbf{I}$, and then fuses $\mathbf{O}$ into the one-dimensional signals $\mathbf{y}$, as shown in Figure 10.

Stage I: Progressive Signal Extraction. The goal of the first stage is to extract the signal layer $\mathbf{O}(\mathbf{x})$ from the input image $\mathbf{I}(\mathbf{x})$. To achieve this, we introduce a stripe extraction model denoted as f. This model takes $\mathbf{I}$ as input and aims to output the background $\mathbf{B}$. Consequently, the signal layer $\mathbf{O}$ can be obtained using $\mathbf{O} = \mathbf{I} - \mathbf{B}$, which serves as the input of Stage II.

The primary design problem here is to devise an effective model structure to efficiently extract subtle stripes. To address this, we adopt the concept of **progressive signal extraction**. Specifically, we utilize a series of extraction blocks, termed E-Blocks, to gradually extract stripes from the input image $\mathbf{I}$. Each E-Block follows an identical structure, comprising two convolutional layers (one at the beginning and one at the end) and three stacked residual blocks (ResBlocks) [21]. Each ResBlock consists of two 3x3 convolutional layers with 32 filters, followed by ReLU activation. The ResBlock uses a skip connection, adding the original input to the second convolutional layer's output. To enhance parameter efficiency and reduce storage overhead, we introduce **cross-layer parameter sharing** [35]. Specifically, all E-Blocks share the same set of weights, making one E-Block work iteratively. This approach significantly reduces the number of parameters in the overall model, resulting in a compact model size of approximately 2.5 megabytes.

When formulating the optimization objective to minimize the dissimilarity between the output $f(\mathbf{I})$ and the background component $\mathbf{B}$, we consider **both spatial and frequency domain** distances. This consideration arises from recognizing the **column-wise periodicity** characteristics of OCC signals. The devised loss function, $\mathcal{L}_1$, is expressed as follows:

$$\mathcal{L}_1 = ||f(\mathbf{I}) - \mathbf{B}||_2 + ||F(f(\mathbf{I})) - F(\mathbf{B})||_2, \tag{4}$$

where $||\cdot||_2$ is the Euclidean distance and F represents the 1D discrete Fourier transform (DFT) conducted along the horizontal direction to capture column-wise periodicity [31]. The output of the last E-Block yields the signal layer $\mathbf{O}$, serving as Stage II's input.

Stage II: Signal Fusion. After obtaining the signal layer $\mathbf{O}$ from Stage I, fusing signals across rows to generate one-dimensional OCC signals $\mathbf{y}$ remains challenging due to the *unevenness of information*. Specifically, different pixels in $\mathbf{O}$ convey varying amounts of signals due to differences in object reflectance, distance from the transmitter, etc [4], making simple row summation unsuitable. To address this, we introduce a second DNN model g dedicated to producing clean 1D OCC signals $\mathbf{y}$. Its optimization objective is to minimize the difference between the output and the ground truth signal $\mathbf{y}_{gt}$, which can be expressed as:

$$\mathcal{L}_2 = ||g(\mathbf{O}) - \mathbf{y}_{gt}||_2. \tag{5}$$

In the second model g, the input $\mathbf{O}$ goes through six consecutive ResBlocks, followed by a convolutional layer with Tanh activation [57]. After each block, 1D max pooling is applied along the vertical dimension to distill and ultimately compress $\mathbf{O}$ into the output signal $\mathbf{y}$.

The two stages are trained **end-to-end** [39], optimizing jointly through the sum of loss functions $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$. We use the Adam optimizer with a learning rate of 1e-04 for training.

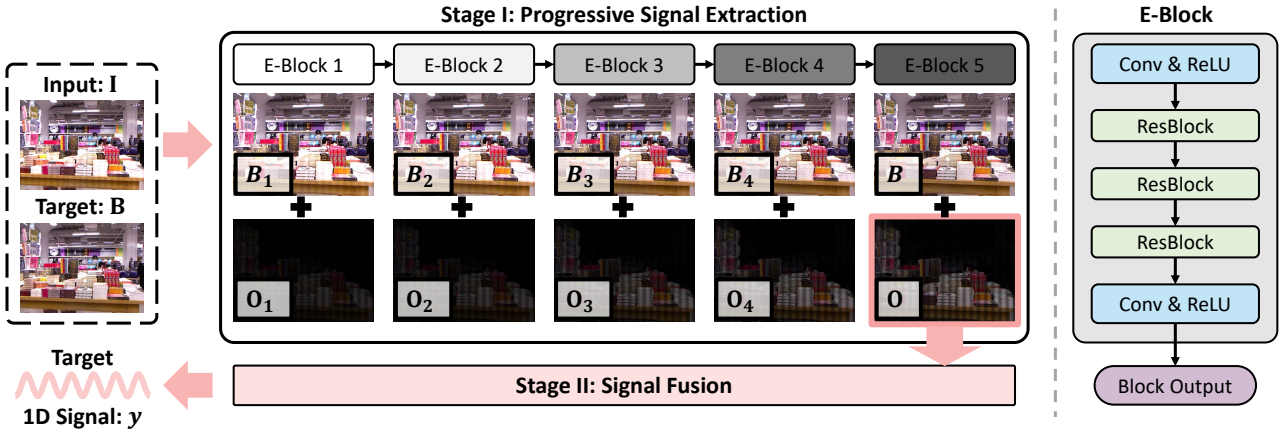


Figure 10: Figure depicts *WinkLink*'s two-stage signal extraction model (§4.1). *WinkLink* first extracts the signal layer O from the encoded image I in Stage I, and then fuses O into the one-dimensional signals y in Stage II.

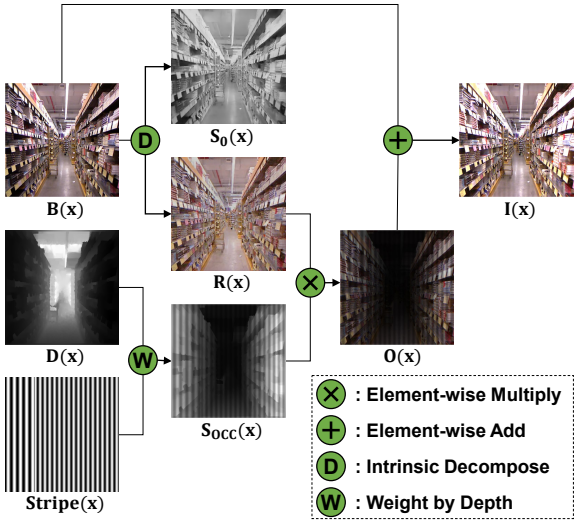


Figure 11: Figure depicts *WinkLink*'s Training Data Synthesis module (§4.2). This module synthesizes OCC-encoded image $\mathbf{I}(\mathbf{x})$ with the known background image $\mathbf{B}(\mathbf{x})$ and signal $\mathbf{y}_{gt}$ serving as ground truth for training.

4.2 Scalable Training Data Synthesis

This module generates a synthetic dataset to train the DNN model. To ensure its effectiveness on real-world test data, we minimize the gap between synthetic and real data by leveraging precise light propagation modeling, which accounts for (1) the *light reflection* model under the Lambertian assumption [4] and (2) the *light attenuation* on varying distances. To make *WinkLink* generalize well in real world, this process encompasses the four vital types of diversities discussed in §3.2.2. This module is only present in *Bootstrapping Phase* and is not required to be deployed on user devices. Notably, we completely eliminate the necessity for manual image collection for training, as *the entire dataset is synthesized.*

The workflow of the module is depicted in Figure 11. Its inputs comprise individual images, serving as background components **B**, along with their corresponding depth maps **D**, sourced from an online dataset [66]. The module's outputs consist of synthesized encoded images **I** with ground truth stripe $\mathbf{y}_{gt}$ and background **B**.

Stripe Pattern Generation. The first step is to generate the 1D signal $\mathbf{y}_{gt}$ with given frequencies, following Equation 1. This process accounts for two diversities: OCC frequency diversity and mixed-symbol diversity. The former involves selecting frequencies from a uniform distribution between 1 KHz and 3 KHz. To introduce mixed-symbol diversity, a border point is randomly chosen among columns. Then, distinct frequencies are randomly assigned to the left and right sides of this border point. The resulting $\mathbf{y}_{gt}$ is then expanded vertically to create the 2D stripe pattern $\mathbf{Stripe}(\mathbf{x})$, corresponding to the received frame under clean backgrounds.

Illumination Map Computation. Our second step is to employ the stripe pattern $\mathbf{Stripe}(\mathbf{x})$ and the depth map $\mathbf{D}(\mathbf{x})$ as inputs to compute the illumination map $\mathbf{S}_{\mathrm{OCC}}(\mathbf{x})$. This map characterizes the received light intensity from the OCC transmitter at each pixel. Leveraging the depth map $\mathbf{D}(\mathbf{x})$, which provides depth information for each pixel, along with a randomly chosen transmitter point r and the principles of the inverse square law [29], we compute $\mathbf{S}_{\mathrm{OCC}}(\mathbf{x})$ as:

$$\mathbf{S}_{\mathrm{OCC}}(\mathbf{x}) = \mathbf{Stripe}(\mathbf{x}) / |\mathbf{D}(\mathbf{x}) - r|^2 . \tag{6}$$

Note that the receiver device is *not required* to possess depth-sensing capabilities during deployment, as the depth map is solely utilized for training data synthesis and is not a requisite when deploying *WinkLink* on mobile devices.

Coupling Illumination with Input Image. Finally, we generate the encoded image $\mathbf{I}(\mathbf{x})$ by superimposing $\mathbf{S}_{\mathrm{OCC}}(\mathbf{x})$ onto the input image $\mathbf{B}(\mathbf{x})$. This process takes into account the fact that when light interacts with a surface, energy loss occurs, leading to a transformation of the original wavelength directed toward the observer. To simulate this process, we utilize the Lambertian assumption [4]. Under this assumption, an image captured by a camera can be modeled as the element-wise product of a reflectance component **R** and

illumination component S, i.e.:

$$\mathbf{I}(\mathbf{x}) = \mathbf{R}(\mathbf{x})\mathbf{S}(\mathbf{x}), \tag{7}$$

The illumination component S can be further decomposed into $\mathbf{S}(\mathbf{x}) = \mathbf{S_0}(\mathbf{x}) + k * \mathbf{S_{OCC}}(\mathbf{x})$, where $\mathbf{S_0}(\mathbf{x})$ and $\mathbf{S_{OCC}}(\mathbf{x})$ correspond to illumination arising from ambient light and the OCC transmitter, respectively. The parameter k is to accommodate various SNR situations, yielding:

$$\mathbf{I}(\mathbf{x}) = \mathbf{R}(\mathbf{x})(\mathbf{S_0}(\mathbf{x}) + k * \mathbf{S_{OCC}}(\mathbf{x})). \tag{8}$$

Considering that $\mathbf{B}(\mathbf{x}) = \mathbf{R}(\mathbf{x})\mathbf{S_0}(\mathbf{x})$, we have:

$$\mathbf{I}(\mathbf{x}) = \mathbf{B}(\mathbf{x}) + k * \mathbf{R}(\mathbf{x})\mathbf{S_{OCC}}(\mathbf{x}). \tag{9}$$

We obtain $\mathbf{R}(\mathbf{x})$ from $\mathbf{B}(\mathbf{x})$ via intrinsic decomposition [14], thus obtaining the output $\mathbf{I}(\mathbf{x})$. Finally, we manipulate the value of k to achieve SNR variation across the range from -14 dB to -25 dB, thereby attaining diversity in SNR levels.

4.3 Context-Aware Demodulation

This module takes as input the extracted 1D signals $\mathbf{y}$ from the DNN model to output a demodulated bit stream, utilizing a context-aware demodulation approach. The key insight is that, despite the variability of absolute OCC intensity due to delayed signal response (§3.2.3), the increase/decrease in intensity can assist as an additional indicator for demodulation. By combining the absolute intensity with its first-order derivative between consecutive frames, we significantly enhance the transmission robustness under the lossy channel (as shown in §5.3.2). This process unfolds as follows.

For each symbol window denoted as K, a 1D fast Fourier transform (FFT) is applied to compute the intensities P_i^K of frequency components for each symbol, where $i = 1, \ldots, 4$, representing distinct frequency components. Subsequently, z-normalization is conducted on the intensities of each symbol across all received symbols [61]. The normalized symbol intensity, denoted as $\bar{P}_i^K$, is defined as $\bar{P}_i^K = (P_i^K - \mu_i)/\sigma_i$, wherein $\mu_i = \sum_k P_i^K/N_s$ represents the mean of intensities across all symbol windows, N_s is the total number of symbols, and $\sigma_i = \sqrt{\sum_k (P_i^K - \mu_i)^2/(N_s - 1)}$ represents the standard deviation across all symbol windows.

We then determine the symbol based on the intensity change. Specifically, when demodulating window K, we calculate the first-order derivative $\Delta\bar{P}_i^K = \bar{P}_i^K - \bar{P}_i^{K-1}$. If $\Delta\bar{P}_i^K$ exceeds a predefined threshold denoted as $thresh_i$, that is, $\Delta\bar{P}_i^K > thresh_i$, we recognize symbol i as the decoded symbol. In practice, the threshold is set as $thresh_i = 0.5 * \sigma_{\Delta\bar{P}_i}$, where $\sigma_{\Delta\bar{P}_i}$ represents the standard deviation of $\Delta\bar{P}_i^K$. The value 0.5 *remains a fixed parameter* and does not change. If multiple frequency components satisfy $\Delta\bar{P}_i^K > thresh_i$, we demodulate the symbol as the frequency component with the highest value, i.e., $i = \arg\max_i \Delta\bar{P}_i^K$. Finally, in cases where no frequency component has a first-order derivative exceeding the threshold, we resort to demodulation by comparing the absolute value of the normalized symbol intensity and select $i = \arg\max_i \bar{P}_i^K$. By doing so, *WinkLink* achieves robust demodulation under the lossy OCC channel.

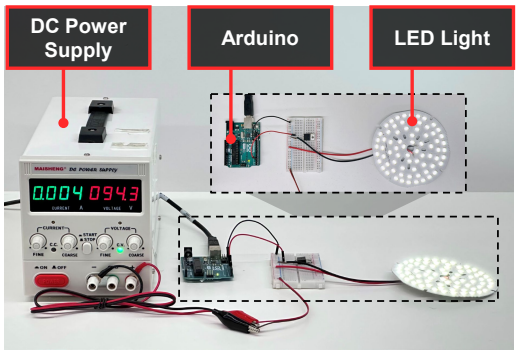


Figure 12: Figure depicts the prototype of *WinkLink* transmitter. The LED light is modulated using an Arduino Uno and powered by a DC power supply.

5 EVALUATION

We present the evaluation of *WinkLink* through comprehensive real-world experiments, demonstrating its effectiveness.

5.1 Experimental Setup

Prototype Implementation. As shown in Figure 12, we implement the prototype with a low-cost Arduino Uno board [3]. It is connected to a MOSFET component. The Uno's microcontroller has hardware support for pulse width modulation (PWM) and thus can be used to generate square waves with specific frequencies. We implement 4-FSK modulation by adjusting the timer register to vary the OCC signals from 1020 Hz to 1980 Hz, maintaining a 50% duty cycle. The LED light is powered by a DC power supply.

Train Dataset Synthesis. *WinkLink*'s signal extraction network is trained on a synthetic dataset derived from the NYU Depth V2 dataset [66], containing 1,449 dense RGB-depth image pairs from diverse indoor scenes. We leverage all 1,449 images for varied backgrounds and synthesize each into 5 training samples, introducing variations in stripe frequency, SNR, and symbol mixtures. Stripe frequencies are randomly set between 1 KHz and 3 KHz, with SNRs from -14 dB to -25 dB, resulting in a dataset of 7,245 images. Training is conducted using PyTorch on a server with dual NVIDIA RTX 3090 GPUs over 200 epochs.

Test Dataset Collection. We collect test dataset with a total of 520,000 distinct frames under 12 diverse environments. During collection, each symbol persists for 1/60 seconds, enabling a transmission rate of 120 bits per second (bps) with 4-FSK modulation. Our default receiver is a Huawei P40 Pro phone, utilizing the mcpro24fps app for video recording [27, 53]. We also evaluate *WinkLink* across varying phone models, transmission rates, distances, and motion speeds (§5.4). For video recording, we set a frame rate of 60 frames per second (FPS) and an exposure time of 1/1200 seconds, using auto ISO[4] for well-lit images, at a resolution of 1920×1080 pixels. Frames are resized to 512×512 to fit *WinkLink*'s signal extraction model.

[4]Based on the principles of the exposure triangle [58], we adjust ISO to balance the light for fixed exposure settings. For example, a scene well-exposed at 1/120 second and ISO 100 can maintain its exposure quality at 1/1200 second with an ISO of 1000, compensating for the reduced light due to the shorter exposure time. Devices like the Huawei P40 Pro, with a maximum ISO of 6400 (expandable in low light), demonstrate the ability to adeptly adjust ISO for various lighting conditions.

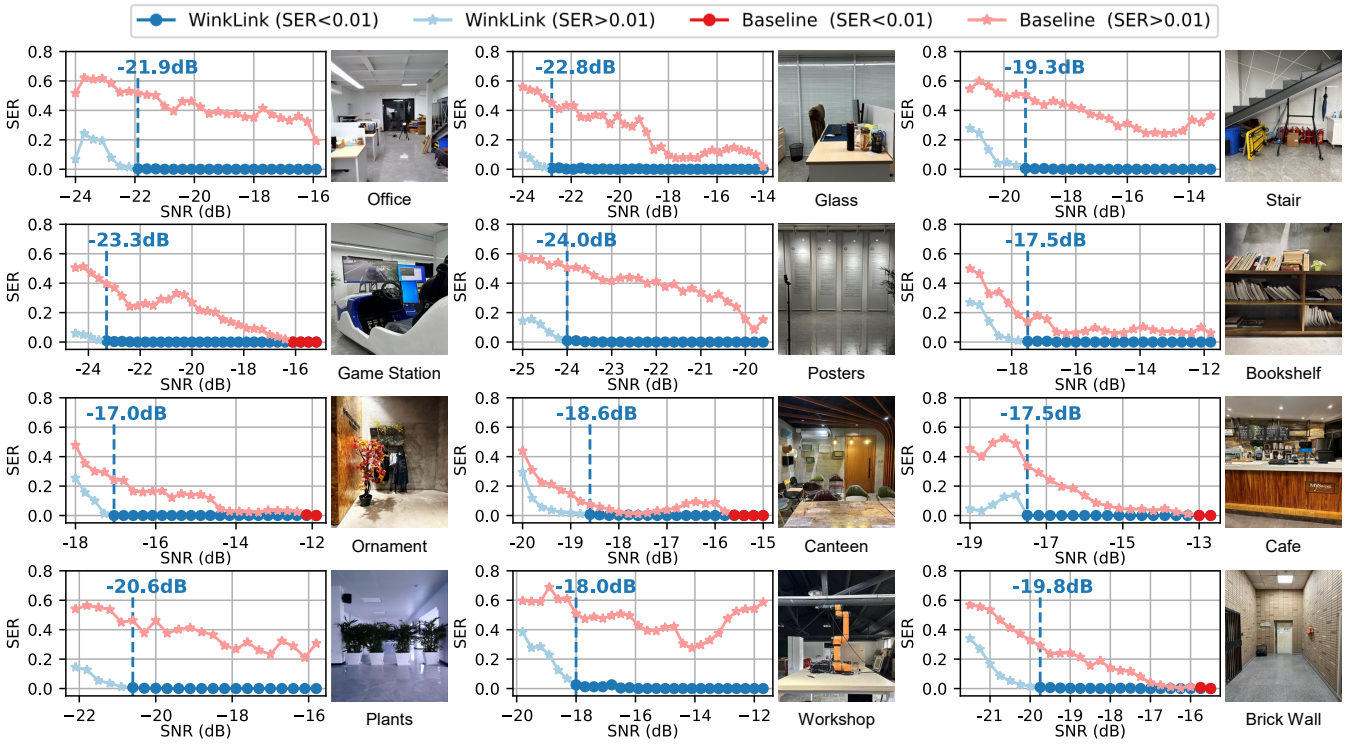


Figure 13: Figure depicts *WinkLink*'s overall performance in 12 diverse environments with mSNRs annotated.

Note that *WinkLink* is trained upon an online dataset, meaning that all the backgrounds in the tests are unseen. We conduct this study upon the approval of our institution's Institutional Review Board.

Performance Metrics. We define two metrics for evaluation: the Symbol Error Rate (SER) and the minimum Signal-to-Noise Ratio (mSNR). **SER** is computed using the formula $SER = N_{error}/N_{symbol}$, where N_{error} represents the number of symbol errors and N_{symbol} represents the total number of transmitted symbols. **mSNR** denotes the minimum SNR value at which the SER drops below 0.01, which is a valuable indicator of an acceptable SNR threshold.

5.2 Overall Performance

5.2.1 Data Preparation. To ensure diversity and minimize bias, our data collection encompasses a varied dataset. We assess *WinkLink*'s performance through experiments in 12 distinct environments, as shown in Figure 13, featuring a range of textures, materials, and transmitter & receiver positions (distances from 2.9m to 5.9m). Ambient light is kept constant within each environment but varies across the 12 environments, ranging from 40 to 800 lux. We collect 30,000 frames per environment under different SNRs, totaling 360,000 unique frames for analysis. SNR variations are achieved by adjusting LED's power from 0 to 5 watts. It is important to note that all backgrounds are ***unseen*** for *WinkLink*, and all the test frames are ***genuinely captured*** by phone cameras and ***not synthetic***.

5.2.2 Overall Results. We demonstrate *WinkLink*'s overall communication performance under unseen and complex backgrounds under low-SNR conditions. We set the *baseline* to be previous FSK systems, which involves summing up the rows to aggregate signal for demodulation [37, 60]. It is chosen due to its superior performance in low-SNR conditions compared to other modulation designs. Figure 13 depicts *WinkLink*'s performance in each individual environment, showing SER curves with respect to the SNR. We also denote the mSNRs of *WinkLink* on the plots.

Our experiments reveal that *WinkLink* achieves an average mSNR of -20.0 dB. While mSNR varies across different environments due to the impact of background characteristics, such as texture and light distribution, *WinkLink consistently outperforms the baseline* in all scenarios and across all SNRs. The baseline only achieves SERs lower than 0.01 in five environments, indicating poor performance in the other seven environments where mSNR is not applicable. In where mSNR is applicable, *WinkLink* offers a noteworthy SNR gain higher than 5.8 dB, highlighting its capability of robust transmission behind arbitrary backgrounds. The lowered SNR requirement significantly mitigates the deployment constraints encountered by prior OCC systems and substantially enhances the feasibility of practical OOC deployment.

5.3 Performance of System Modules

We now evaluate the individual modules of *WinkLink*.

5.3.1 DNN-based Signal Extraction Module. We evaluate this module (§4.1) in the following three aspects.

Sub Network Design Effectiveness. In evaluating the DNN model's design, we conducted an ablation study comparing two baseline scenarios: *WinkLink* without the iterative E-Block design (D1), where

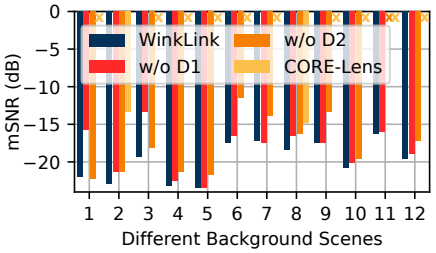


Figure 14: Figure depicts mSNRs of different signal extraction models.

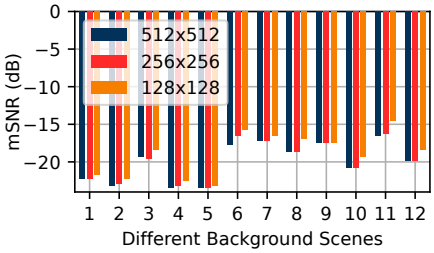


Figure 15: Figure depicts mSNRs under different input sizes.

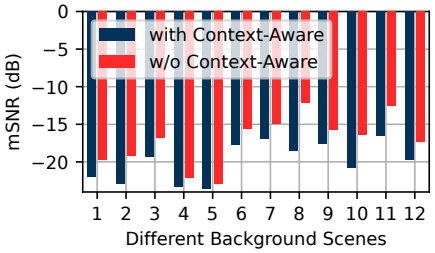


Figure 16: Figure depicts the effect of context-aware demodulation.

only one E-Block pass is performed, and *WinkLink* without its two-stage framework (D2), employing only Stage II. For D2, Stage II processes encoded images as inputs to directly produce 1D signals, optimizing solely for L2 loss. The mSNRs are depicted in Figure 14. Conditions where the SER consistently exceeds 0.01 (i.e., mSNR not applicable) are marked with '✗' on the plot. The results demonstrate the effectiveness of the two design components, with an average mSNR enhancement of 1.5 dB and 2.5 dB, respectively.

Comparison with CORE-Lens. We also compare our overall network design with the VAE-GAN network used in CORE-Lens [50], which we retrained with our synthetic dataset due to the unavailability of its original data and weights. In Figure 14, VAE-GAN shows ineffectiveness in all but one scenario, indicating its *inadequacy in low-SNR conditions*. This performance disparity can be attributed to the differing core design logic: CORE-Lens uses a *generative model*, while *WinkLink* relies on *image translation*, which is more effective in low-SNR conditions as it allows the model to learn and adapt to diverse backgrounds, converging effectively without the need for a prohibitively large dataset.

Impact of Input Size. By default, we resize images to 512×512 for network input. To investigate the impact of input size on *WinkLink*'s performance, we examine three sizes: 512×512, 256×256, and 128×128. The corresponding mSNR values are shown in Figure 15. Smaller sizes result in average mSNR increases of merely 0.13 dB and 1.08 dB, respectively. This highlights *WinkLink*'s compatibility with small input sizes, allowing for reduced processing time without significant performance loss (detailed in §5.5).

5.3.2 Context-Aware Demodulation Module. Recall that the main goal of this module (§4.3) is to address the delayed signal response by using the first-order derivative of symbol intensity for demodulation. We now compare *WinkLink*'s performance with and without the module. Without the module, we resort to using only the intensity for demodulation. In Figure 16, the inclusion of this module consistently yields better results compared to its absence, with an average mSNR improvement of 2.84 dB, highlighting the importance of *WinkLink*'s Context-Aware Demodulation module.

5.4 Differing Experimental Conditions

We evaluate *WinkLink*'s performance across various experimental conditions in the first environment, with SNR variations achieved by altering LED power between 0 and 5 watts, consistent with §5.2.

5.4.1 Varying Camera Devices. We evaluate *WinkLink*'s efficacy across three distinct phone models: iPhone 14 Pro (iOS 16), Huawei P40 Pro (Harmony OS), and Samsung Galaxy S21 (Android 12), with respective readout durations of 11.7 μs, 4.96 μs, and 11.3 μs [2, 27, 63]. Video recordings are conducted through the Protake app for iPhone, and the mcpro24fps app for Huawei and Samsung devices [53, 68]. As illustrated in Figure 17, the mSNRs are -21.9 dB, -19.4 dB, and -19.8 dB, for the three phones, respectively. This consistent performance is attributed to *WinkLink*'s independence from device-specific knowledge.

5.4.2 Varying Motion Speed. *WinkLink*'s functionality extends to scenarios involving cameras in motion with dynamic backgrounds. We conduct experiments under varying camera velocities: 0.5 m/s, 1 m/s, and 2 m/s—representative of ordinary walking speeds. The SERs are measured across six SNRs levels from -15.5 dB to -22.6 dB, presented in Figure 18. Notably, the SERs are consistently below 0.001 until the SNR decreases to -22.4 dB, which is below the mSNR of stationary camera scenario. *WinkLink*'s resilience on receiver mobility also indicates its capability to manage scenarios with moving users or objects, due to its ability to handle background variations. Receiver mobility, which alters every pixel of the background, presents a higher challenge than scenarios with moving objects that only impact a portion of the scene. The core of *WinkLink* is to employ a single-frame strategy, extracting signals based on individual frames rather than using multi-frame techniques that depend on a constant background [19, 37]. Consequently, this approach ensures *WinkLink*'s consistent performance amidst the dynamic variations common in real-world environments.

5.4.3 Varying Transmission Rates. We assess *WinkLink*'s performance across different transmission rates by altering symbol durations while maintaining a consistent video frame rate of 60 FPS. With symbol lengths of 1/30s, 1/60s, and 1/120s, we achieve transmission rates of 60 bps, 120 bps, and 240 bps, respectively. The results in Figure 19 show mSNRs of -22.7 dB, -22.6 dB, and -17.2 dB. Transmitting at 240 bps requires a higher SNR due to symbol loss caused by frame gaps, a recognized issue with phone cameras, particularly pronounced when symbol duration is shorter than frame duration [23]. However, -17.2 dB in 240 bps is already significantly lower than the mSNR of baseline in 120 bps, showing *WinkLink*'s effective demodulation under higher transmission rates.

5.4.4 Varying Communication Range. To highlight the benefits of *WinkLink*'s reduced SNR requirements, we evaluate its performance over different distances using LEDs of 5 watts and 10 watts. The ambient light intensity is held constant at 450 lux, reflecting typical

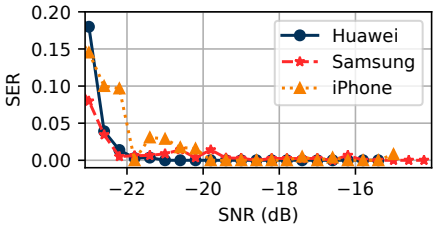


Figure 17: Figure depicts *WinkLink*'s performance using different phones as the receiver.

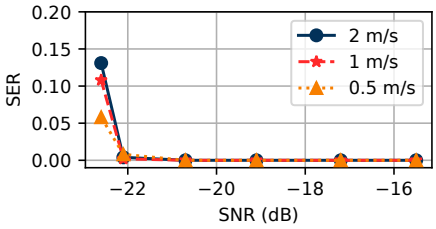


Figure 18: Figure depicts *WinkLink*'s performance under different movement speeds.

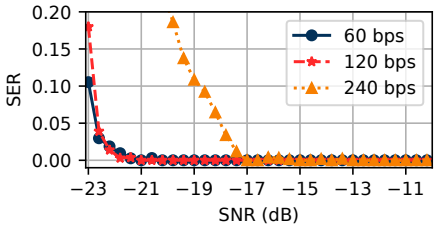
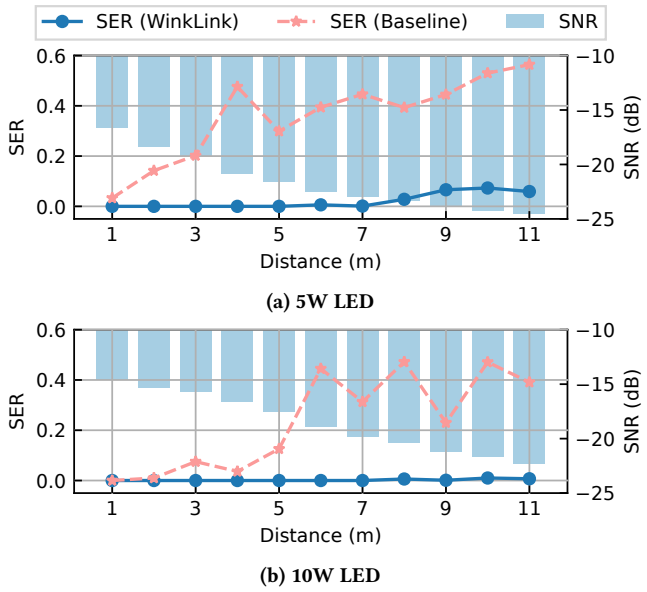


Figure 19: Figure depicts *WinkLink*'s performance under varying transmission rates.



(a) 5W LED

(b) 10W LED

Figure 20: Figure depicts *WinkLink*'s performance at different distances when the LED is (a) 5W and (b) 10W.

indoor conditions. We compare *WinkLink* with the same baseline as §5.2, presenting SERs and SNRs across distances in Figure 20. With the 5-watt LED, *WinkLink* consistently attains SERs below 0.01 up to a distance of 7 meters, with an SNR of -22.2 dB. For the 10-watt LED, SERs stay below 0.007 up to 11 meters – the maximum distance achievable within our laboratory setting. These results highlight *WinkLink*'s capability to maintain reliable communication over extended ranges, fitting well within the scope of moderate room sizes. In larger environments such as museums, where separate lighting is common for exhibits, *WinkLink* can leverage these individual light sources for effective communication across the space. Additionally, employing larger power LEDs or multiple LED lights can extend the transmission range even further.

5.5 Storage Overhead and Running Time

We evaluate the storage overhead and running time of *WinkLink*'s demodulation process, considering varying input sizes and three distinct device types: an RTX 3090 GPU, an Intel Xeon 4210R CPU, and a Huawei P40 Pro phone. The results are presented in Table 1.

Table 1: Storage overhead and running time.

Input Size	Time			Model Size
	GPU	CPU	Phone	
512 × 512	1.36ms	45.9ms	1491ms	2.68MB
256 × 256	0.36ms	11.7ms	348ms	2.49MB
128 × 128	**0.09ms**	**3.38ms**	**77ms**	**2.45MB**

Table 2: Experiment setups for the four vision tasks.

	DataSet	# Images	Model
Image Classification	ImageNet [15]	50,000	ResNet [21]
Object Detection	CoCo [48]	5,000	Faster R-CNN [62]
Instance Segmentation	CoCo [48]	5,000	Mask R-CNN [20]
Semantic Segmentation	CoCo [48]	5,000	FCN [65]

Upon GPU execution, the processing time amounts to a mere 1.36 milliseconds. Remarkably, upon phone execution, employing an input size of 128 × 128 reduces the processing time to 77 milliseconds, achieving low latency on mobile edge. Referring to observation in §5.3.1, where adopting 128 × 128 incurs a small mSNR increase of 1.08 dB – an acceptable trade-off, we recommend its implementation for on-device inference scenarios. Techniques like model compression can further expedite on-device inference [24, 33, 49] (see §6 for further discussion). Conversely, when potent edge servers are available, larger input sizes can be utilized for enhanced demodulation robustness [91]. Meanwhile, the model's compact size of 2.5 MB ensures a small storage overhead. These attributes position *WinkLink* as a practical and deployable solution.

5.6 Interference on Vision Systems

With *WinkLink*'s low-SNR characteristics, we envision its operation with concurrent vision applications. We evaluate OCC's interference with vision systems by examining four representative vision tasks: image classification, object detection, instance segmentation, and semantic segmentation. For each task, we utilize accuracy, box mean average precision (MAP), mask MAP, and mean intersection-over-union as evaluation metrics, and employ ResNet, Faster R-CNN, Mask R-CNN, and FCN as the respective vision

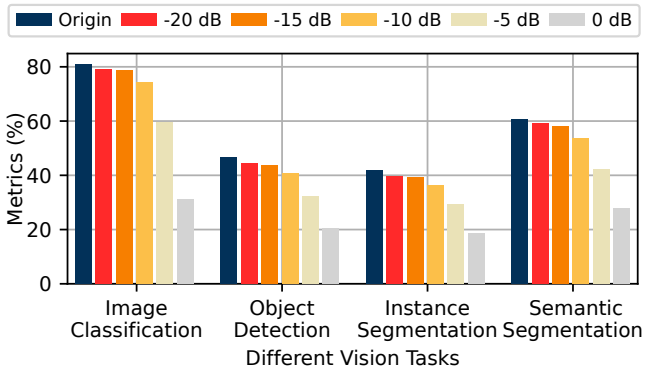


Figure 21: Figure depicts the inference of OCC on four representative vision tasks at different SNR levels.

models [20, 21, 62, 65]. Dataset details for each task are outlined in Table 2. We synthesize OCC signals on images, spanning SNRs from -20 dB to 0 dB with increments of 5 dB. The results, illustrated in Figure 21, show that as SNRs increase to -10 dB, -5 dB, and 0 dB, the average relative degradation over four tasks is 11.2%, 29.2%, and 56.8%, respectively. However, at -20 dB, i.e., *WinkLink*'s operational SNR, the average relative degradation is only 3.4%, showing that *WinkLink* has minimal interference with vision tasks and enables effective concurrent operation.

6 DISCUSSION

We now present important discussion points of *WinkLink*.

Domain Gap Between Synthetic and Real Data. While we have included most of critical factors of light propagation, we acknowledge potential *domain gap* between synthetic and real data, particularly concerning real-world noise and distortions. Indeed, rendering a scene under specific light condition remains a challenging problem in the field of computer graphics [79]. While domain gap does affect model robustness, it is manageable because *WinkLink* works well in 12 diverse unseen environments, demonstrating its ability to learn transferable knowledge from synthetic data to handle the majority of real-world cases. In rare corner cases, practical solutions such as fine-tuning the model based on real captured instances, are readily implementable. The current model, trained solely over synthetic data, provides a solid base for further refinement.

Data Rate and Collaborative Connectivity. The upper bound of *WinkLink*'s transmission capacity is constrained by its ***low-SNR*** characteristics, as stipulated by Shannon's capacity theorem [11]. Currently, *WinkLink* enables location-aware data transfer at 240 bps, reusing existing light infrastructure to support spatial localization and pervasive connectivity [34, 37, 64]. It can facilitate applications such as redeeming discount vouchers in grocery stores or monitoring power consumption in smart homes, with optional visualization through MR. *WinkLink* can also facilitate seamless connection establishment through the transmission of short-lived tokens. These moderate-size payloads can be directly delivered through *WinkLink*. However, *WinkLink* is not a one-size-fits-all solution for connectivity. For applications requiring larger data transfer, such as delivering menus in restaurants or interactive dinosaur models in museums, we envision *WinkLink* transmitting IDs for cloud service lookup, followed by fetching larger payloads over faster connections like cellular networks. Their operation in distinct spectrum enables easy coexistence [62], allowing each technology to complement and leverage its unique strengths.

Computational Cost. While high-resolution video recording and on-device DNN inference are resource-intensive, *WinkLink* effectively reduces these demands by supporting low-resolution inputs like 128x128, lowering computational load and saving battery life. As demonstrated in §5.5, the inference latency is 0.09 ms on cloud and 77 ms on-device, with potential for optimization. On-device DNN inference optimization is an active research field. At the model level, key strategies include model compression [6, 18, 22] and neural architecture search [69, 75]. At the system level, leveraging heterogeneous processors, cache optimization, and adaptive offloading [30, 36, 52] holds great promise. These ongoing efforts, coupled with evolving mobile processor capabilities, especially in GPUs and NPUs, will significantly enhance inference speeds.

Interference between Multiple OCC Links. The extended communication range of *WinkLink* may lead to interference between adjacent OCC links. However, the *inherent space division property* of OCC serves as a natural countermeasure [19, 34, 84]. Physical proximity to a specific light can enhance its signal strength, effectively prioritizing its content and thereby minimizing interference. Techniques like beam steering, e.g., using devices such as light shades to focus the LED beam, further refine signal targeting and minimize overlap. Adjusting light intensity and employing frequency division, which allocates unique frequency bands to different signals, are additional effective strategies to mitigate interference and ensure coexistence among neighboring transmitters.

Integrated Sensing and Communication on Vision. *WinkLink* utilizes the vision modality for communication while minimizing impacts on vision applications (§5.6). Therefore, *WinkLink* sheds new light on the integration of sensing and communication (ISAC) in the vision region. Under *WinkLink*, OCC becomes a valuable additional feature of the camera, enhancing its value without imposing limitations on its original purpose. Future work will aim to expand these integrated capabilities. A potential issue, particularly in MR settings, is the visibility of stripe patterns, which could impact user experience. To solve this, note that *WinkLink*'s DNN model not only extracts OCC signals but also clears stripe patterns and provides a clean background. This dual-functionality suggests a promising avenue for development, potentially enhancing video quality by removing visual disturbances while extracting OCC signals.

7 RELATED WORK

We now present related work with *WinkLink*.

Optical Camera Communication. Researchers have focused on enhancing the throughput of OCC through modulation optimization [9, 25, 80]. Recent advancements have extended OCC to under-screen cameras [82, 83]. The LED-camera channel has been explored in diverse applications including indoor localization [34, 84, 92], and hand pose reconstruction [87]. The closest to our work is CORE-Lens [50], which enables object recognition under OCC transmission. While CORE-Lens operates at a limited distance of 1.4 meters and with a restricted set of six trained backgrounds (e.g., a pink

wallpaper), *WinkLink* excels in low-SNR transmission under arbitrary unseen complex backgrounds, extending the operational range while ensuring seamless compatibility with four standard vision applications without modifications.

Communication and Sensing with Light. Beyond OCC, light has been utilized as a communication medium in various contexts, such as LED-photodiode channel [55, 89], screen-camera channel [26, 40, 56, 59, 71, 86], backscatter communication [70, 73, 76, 77], and air-water communication [7, 8]. Researchers have also leveraged the radio-frequency side channel to enhance or intercept light communication [12, 13, 54]. Besides, light has been employed as a sensing medium, enabling applications like indoor position and orientation estimation [17, 38, 74, 85], inertial tracking [88], glucose and blood pressure monitoring [42, 72, 78], gaze tracking [44, 45], and human sensing [41, 43].

8 CONCLUSION

We propose *WinkLink*, a novel OCC system that operates under unseen and complex backgrounds, all while maintaining low-SNR requirements. Through our implementation and extensive real-world evaluation across 12 diverse environments comprising 520,000 frames, we demonstrate the effectiveness of *WinkLink* in adapting to dynamic surroundings, covering longer distances, and minimizing interference with vision applications, establishing *WinkLink* as a practical and reliable OCC system for real-world scenarios.

ACKNOWLEDGMENTS

We sincerely thank our anonymous reviewers and our shepherd, Professor Xia Zhou, for their valuable feedback. This paper is supported by the National Natural Science Foundation of China under grant U21A20462 and 62372400, "Pioneer" and "Leading Goose" R&D Program of Zhejiang under grant No. 2023C01033.

REFERENCES

[1] Apple. 2023. Apple Vision Pro. https://www.apple.com/apple-vision-pro/.
[2] Apple. 2023. iPhone 14 Pro. https://www.apple.com/iphone-14-pro/.
[3] Arduino. 2023. Arduino Uno Rev3. https://store.arduino.cc/products/arduino-uno-rev3.
[4] Jonathan T. Barron and Jitendra Malik. 2015. Shape, Illumination, and Reflectance from Shading. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 8 (2015), 1670–1687.
[5] Sai Bi, Xiaoguang Han, and Yizhou Yu. 2015. An L_1 image transform for edge-preserving smoothing and scene-level intrinsic decomposition. *ACM Trans. Graph.* 34, 4 (2015), 78:1–78:12.
[6] Han Cai, Chuang Gan, Tianzhe Wang, Zhekai Zhang, and Song Han. 2020. Once-for-All: Train One Network and Specialize it for Efficient Deployment. In *8th International Conference on Learning Representations, ICLR'20.*
[7] Charles J. Carver, Qijia Shao, Samuel Lensgraf, Amy Sniffen, Maxine Perroni-Scharf, Hunter Gallant, Alberto Quattrini Li, and Xia Zhou. 2022. Sunflower: locating underwater robots from the air. In *The 20th Annual International Conference on Mobile Systems, Applications and Services, MobiSys'22.* ACM, 14–27.
[8] Charles J. Carver, Tian Zhao, Hongyong Zhang, Kofi M. Odame, Alberto Quattrini Li, and Xia Zhou. 2020. AmphiLight: Direct Air-Water Communication with Laser Light. In *17th USENIX Symposium on Networked Systems Design and Implementation, NSDI'20.* USENIX Association, 373–388.
[9] Chun-Ling Chan, Hsin-Mu Tsai, and Kate Ching-Ju Lin. 2017. POLI: Long-Range Visible Light Communications Using Polarized Light Intensity Modulation. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys'17.* ACM, 109–120.
[10] C/LAN/MAN LAN/MAN Standards Committee. 2018. IEEE Standard for Local and metropolitan area networks–Part 15.7: Short-Range Optical Wireless Communications. https://standards.ieee.org/ieee/802.15.7/6820/.
[11] Thomas M. Cover and Joy A. Thomas. 2006. *Elements of Information Theory.* Wiley-Interscience, USA.
[12] Minhao Cui, Yuda Feng, Qing Wang, and Jie Xiong. 2020. Sniffing visible light communication through walls. In *The 26th Annual International Conference on Mobile Computing and Networking, MobiCom'20.* ACM, 30:1–30:14.
[13] Minhao Cui, Qing Wang, and Jie Xiong. 2021. RadioInLight: doubling the data rate of VLC systems. In *The 27th Annual International Conference on Mobile Computing and Networking, MobiCom'21.* ACM, 615–627.
[14] Partha Das, Sezer Karaoglu, and Theo Gevers. 2022. PIE-Net: Photometric Invariant Edge Guided Network for Intrinsic Image Decomposition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR'22.* IEEE, 19758–19767.
[15] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'09.* IEEE Computer Society, 248–255.
[16] Danilo Gasques, Janet G. Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Yifei Zhang, Wanze Xie, Xinming Zhang, Konrad Davis, Michael Yip, and Nadir Weibel. 2021. ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telementoring. In *CHI Conference on Human Factors in Computing Systems, CHI'21.* ACM, 662:1–662:14.
[17] Muhammad Kumail Haider, Yasaman Ghasempour, and Edward W. Knightly. 2018. Search Light: Tracking Device Mobility using Indoor Luminaries to Adapt 60 GHz Beams. In *Proceedings of the 19th ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc'18.* ACM, 181–190.
[18] Song Han, Huizi Mao, and William J. Dally. 2016. Deep Compression: Compressing Deep Neural Network with Pruning, Trained Quantization and Huffman Coding. In *4th International Conference on Learning Representations, ICLR'16.*
[19] Jie Hao, Yanbing Yang, and Jun Luo. 2016. CeilingCast: Energy efficient and location-bound broadcast through LED-camera communication. In *35th Annual IEEE International Conference on Computer Communications, INFOCOM'16.* IEEE, 1–9.
[20] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. 2017. Mask R-CNN. In *IEEE International Conference on Computer Vision, ICCV'17.* IEEE Computer Society, 2980–2988.
[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR'16.* IEEE Computer Society, 770–778.
[22] Yihui He, Ji Lin, Zhijian Liu, Hanrui Wang, Li-Jia Li, and Song Han. 2018. AMC: AutoML for Model Compression and Acceleration on Mobile Devices. In *15th European Conference on Computer Vision, ECCV'18,* Vol. 11211. Springer, 815–832.
[23] Yuki Hokazono, Aoi Koizuka, Guibing Zhu, Makoto Suzuki, Yoshiaki Narusue, and Hiroyuki Morikawa. 2021. IoTorch: Reliable LED-to-Camera Communication Against Inter-Frame Gaps and Frame Drops. *IEEE Trans. Mob. Comput.* 20, 2 (2021), 550–564.
[24] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *CoRR* abs/1704.04861 (2017).
[25] Pengfei Hu, Parth H. Pathak, Xiaotao Feng, Hao Fu, and Prasant Mohapatra. 2015. ColorBars: increasing data rate of LED-to-camera communication using color shift keying. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies, CoNEXT'15.* ACM, 12:1–12:13.
[26] Wenjun Hu, Hao Gu, and Qifan Pu. 2013. LightSync: unsynchronized visual communication over screen-camera links. In *The 19th Annual International Conference on Mobile Computing and Networking, MobiCom'13.* ACM, 15–26.
[27] Huawei. 2023. HUAWEI P40 Pro. https://consumer.huawei.com/en/phones/p40-pro/.
[28] ITU. 2021. H.264: Advanced video coding for generic audiovisual services. https://www.itu.int/rec/T-REC-H.264.
[29] John David Jackson. 1975. *Classical electrodynamics; 2nd ed.* Wiley, New York, NY.
[30] Fucheng Jia, Deyu Zhang, Ting Cao, Shiqi Jiang, Yunxin Liu, Ju Ren, and Yaoxue Zhang. 2022. CoDL: efficient CPU-GPU co-execution for deep learning inference on mobile devices. In *The 20th Annual International Conference on Mobile Systems, Applications and Services, MobiSys'22.* ACM, 209–221.
[31] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. 2021. Focal Frequency Loss for Image Reconstruction and Synthesis. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV'21.* IEEE, 13899–13909.
[32] Mahmudur Khan and Jacob Chakareski. 2019. Visible Light Communication for Next Generation Untethered Virtual Reality Systems. In *17th IEEE International Conference on Communications Workshops, ICC Workshops 2019.* IEEE, 1–6.
[33] Yong-Deok Kim, Eunhyeok Park, Sungjoo Yoo, Taelim Choi, Lu Yang, and Dongjun Shin. 2016. Compression of Deep Convolutional Neural Networks for Fast and Low Power Mobile Applications. In *4th International Conference on Learning Representations, ICLR'16.*
[34] Ye-Sheng Kuo, Pat Pannuto, Ko-Jen Hsiao, and Prabal Dutta. 2014. Luxapose: indoor positioning with mobile phones and visible light. In *The 20th Annual International Conference on Mobile Computing and Networking, MobiCom'14.* ACM, 447–458.

[35] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2020. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. In *8th International Conference on Learning Representations, ICLR'20*.

[36] Stefanos Laskaridis, Stylianos I. Venieris, Mário Almeida, Ilias Leontiadis, and Nicholas D. Lane. 2020. SPINN: synergistic progressive inference of neural networks over device and cloud. In *The 26th Annual International Conference on Mobile Computing and Networking, MobiCom'20*. ACM, 37:1–37:15.

[37] Hui-Yu Lee, Hao-Min Lin, Yu-Lin Wei, Hsin-I Wu, Hsin-Mu Tsai, and Kate Ching-Ju Lin. 2015. RollingLight: Enabling Line-of-Sight Light-to-Camera Communications. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys'15*. ACM, 167–180.

[38] Liqun Li, Pan Hu, Chunyi Peng, Guobin Shen, and Feng Zhao. 2014. Epsilon: A Visible Light Based Positioning System. In *Proceedings of the 11th USENIX Symposium on Networked Systems Design and Implementation, NSDI'14*. USENIX Association, 331–343.

[39] Rongjie Li, Songyang Zhang, and Xuming He. 2022. SGTR: End-to-end Scene Graph Generation with Transformer. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR'22*. IEEE, 19464–19474.

[40] Tianxing Li, Chuankai An, Xinran Xiao, Andrew T. Campbell, and Xia Zhou. 2015. Real-Time Screen-Camera Communication Behind Any Scene. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys'15*. ACM, 197–211.

[41] Tianxing Li, Chuankai An, Tian Zhao, Andrew T. Campbell, and Xia Zhou. 2015. Human Sensing Using Visible Light Communication. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, MobiCom'15*. ACM, 331–344.

[42] Tianxing Li, Derek Bai, Temiloluwa Prioleau, Nam Bui, Tam Vu, and Xia Zhou. 2020. Noninvasive glucose monitoring using polarized light. In *The 18th ACM Conference on Embedded Networked Sensor Systems, SenSys'20*. ACM, 544–557.

[43] Tianxing Li, Qiang Liu, and Xia Zhou. 2016. Practical Human Sensing in the Light. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys'16*. ACM, 71–84.

[44] Tianxing Li, Qiang Liu, and Xia Zhou. 2017. Ultra-Low Power Gaze Tracking for Virtual Reality. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems, SenSys'17*. ACM, 25:1–25:14.

[45] Tianxing Li and Xia Zhou. 2018. Battery-Free Eye Tracker on Glasses. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking, MobiCom'18*. ACM, 67–82.

[46] Wanwan Li, Changyang Li, MinYoung Kim, Haikun Huang, and Lap-Fai Yu. 2023. Location-Aware Adaptation of Augmented Reality Narratives. In *Proceedings of the 2023 Conference on Human Factors in Computing Systems, CHI'23*. ACM, 33:1–33:15.

[47] Chia-Kai Liang, Li-Wen Chang, and Homer H. Chen. 2008. Analysis and Compensation of Rolling Shutter Effect. *IEEE Trans. Image Process.* 17, 8 (2008), 1323–1330.

[48] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *European Conference on Computer Vision, ECCV'14*, Vol. 8693. Springer, 740–755.

[49] Baoyuan Liu, Min Wang, Hassan Foroosh, Marshall F. Tappen, and Marianna Pensky. 2015. Sparse Convolutional Neural Networks. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR'15*. IEEE Computer Society.

[50] Ziwei Liu, Tianyue Zheng, Chao Hu, Yanbing Yang, Yimao Sun, Yi Zhang, Zhe Chen, Liangyin Chen, and Jun Luo. 2022. CORE-lens: simultaneous communication and object recognition with disentangled-GAN cameras. In *The 28th Annual International Conference on Mobile Computing and Networking, MobiCom'22*. ACM, 172–185.

[51] Mayra Donaji Barrera Machuca, Álvaro Cassinelli, and Christian Sandor. 2020. Context-Based 3D Grids for Augmented Reality User Interfaces. In *The 33rd Annual ACM Symposium on User Interface Software and Technology, UIST'20*. ACM, 73–76.

[52] Akhil Mathur, Nicholas D. Lane, Sourav Bhattacharya, Aidan Boran, Claudio Forlivesi, and Fahim Kawsar. 2017. DeepEye: Resource Efficient Local Execution of Multiple Deep Vision Models using Wearable Commodity Hardware. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys'17*. ACM, 68–81.

[53] mcpro24fps. 2023. Professional manual video camera app. https://www.mcpro24fps.com/.

[54] Muhammad Sarmad Mir, Minhao Cui, Borja Genovés Guzmán, Qing Wang, Jie Xiong, and Domenico Giustiniano. 2023. LeakageScatter: Backscattering LiFi-leaked RF Signals. In *Proceedings of the 24th ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc'23*. ACM.

[55] Muhammad Sarmad Mir, Borja Genovés Guzmán, Ambuj Varshney, and Domenico Giustiniano. 2021. PassiveLiFi: rethinking LiFi for low-power and long range RF backscatter. In *The 27th Annual International Conference on Mobile Computing and Networking, MobiCom'21*. ACM, 697–709.

[56] Viet Nguyen, Yaqin Tang, Ashwin Ashok, Marco Gruteser, Kristin J. Dana, Wenjun Hu, Eric Wengrowski, and Narayan B. Mandayam. 2016. High-rate flicker-free screen-camera communication with spatially adaptive embedding. In *35th Annual IEEE International Conference on Computer Communications, INFOCOM'16*. IEEE, 1–9.

[57] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. 2018. Activation Functions: Comparison of trends in Practice and Research for Deep Learning. *CoRR* abs/1811.03378 (2018).

[58] PolarPro. 2023. The Three Elements of the Exposure Triangle. https://www.polarpro.com/blogs/polarpro/the-three-elements-of-the-exposure-triangle.

[59] Kun Qian, Yumeng Lu, Zheng Yang, Kai Zhang, Kehong Huang, Xinjun Cai, Chenshu Wu, and Yunhao Liu. 2021. AIRCODE: Hidden Screen-Camera Communication on an Invisible and Inaudible Dual Channel. In *18th USENIX Symposium on Networked Systems Design and Implementation, NSDI'21*. USENIX Association, 457–470.

[60] Niranjini Rajagopal, Patrick Lazik, and Anthony Rowe. 2014. Visual light landmarks for mobile devices. In *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks, IPSN'14*. IEEE/ACM, 249–260.

[61] Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. 2012. Searching and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD'12*. 262–270.

[62] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Annual Conference on Neural Information Processing Systems, NIPS'15*.

[63] SAMSUNG. 2023. Galaxy S21. https://www.samsung.com/us/smartphones/galaxy-s21-5g/buy/.

[64] Rahul Anand Sharma, Adwait Dongare, John Miller, Nicholas Wilkerson, Daniel Cohen, Vyas Sekar, Prabal Dutta, and Anthony Rowe. 2020. All that GLITTERs: Low-Power Spoof-Resilient Optical Markers for Augmented Reality. In *19th ACM/IEEE International Conference on Information Processing in Sensor Networks, IPSN'20*. IEEE, 289–300.

[65] Evan Shelhamer, Jonathan Long, and Trevor Darrell. 2017. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 4 (2017), 640–651.

[66] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. 2012. Indoor Segmentation and Support Inference from RGBD Images. In *12th European Conference on Computer Vision, ECCV'12*, Vol. 7576. Springer, 746–760.

[67] Maximilian Speicher, Brian D. Hall, and Michael Nebeling. 2019. What is Mixed Reality?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI'19*. ACM, 537.

[68] Apple Apps Store. 2023. Protake = Mobile Cinema Camera. https://apps.apple.com/us/app/protake-mobile-cinema-camera/id1498431506.

[69] Mingxing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan, Mark Sandler, Andrew Howard, and Quoc V. Le. 2019. MnasNet: Platform-Aware Neural Architecture Search for Mobile. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR'19*. IEEE, 2820–2828.

[70] Ambuj Varshney, Andreas Soleiman, Luca Mottola, and Thiemo Voigt. 2017. Battery-free Visible Light Sensing. In *Proceedings of the 4th ACM Workshop on Visible Light Communication Systems, VLCS@MobiCom'17*. ACM, 3–8.

[71] Anran Wang, Zhuoran Li, Chunyi Peng, Guobin Shen, Gan Fang, and Bing Zeng. 2015. InFrame++: Achieve Simultaneous Screen-Human Viewing and Hidden Screen-Camera Communication. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys'15*. ACM, 181–195.

[72] Edward Jay Wang, Junyi Zhu, Mohit Jain, TienJui Lee, Elliot Saba, Lama Nachman, and Shwetak N. Patel. 2018. Seismo: Blood Pressure Monitoring using Built-in Smartphone Accelerometer and Camera. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI'18*. ACM, 425.

[73] Purui Wang, Lilei Feng, Guojun Chen, Chenren Xu, Yue Wu, Kenuo Xu, Guobin Shen, Kuntai Du, Gang Huang, and Xuanzhe Liu. 2020. Renovating road signs for infrastructure-to-vehicle networking: a visible light backscatter communication and networking approach. In *The 26th Annual International Conference on Mobile Computing and Networking, MobiCom'20*. ACM, 6:1–6:13.

[74] Yu-Lin Wei, Chang-Jung Huang, Hsin-Mu Tsai, and Kate Ching-Ju Lin. 2017. CELLI: Indoor Positioning Using Polarized Sweeping Light Beams. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys'17*. ACM, 136–147.

[75] Bichen Wu, Xiaoliang Dai, Peizhao Zhang, Yanghan Wang, Fei Sun, Yiming Wu, Yuandong Tian, Peter Vajda, Yangqing Jia, and Kurt Keutzer. 2019. FBNet: Hardware-Aware Efficient ConvNet Design via Differentiable Neural Architecture Search. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR'19*. IEEE, 10734–10742.

[76] Yue Wu, Purui Wang, Kenuo Xu, Lilei Feng, and Chenren Xu. 2020. Turboboosting Visible Light Backscatter Communication. In *Proceedings of the 2020 Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication, SIGCOMM'20*. ACM, 186–197.

[77] Kenuo Xu, Chen Gong, Bo Liang, Yue Wu, Boya Di, Lingyang Song, and Chenren Xu. 2022. Low-Latency Visible Light Backscatter Networking with RetroMU-MIMO. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems, SenSys'22*. ACM, 448–461.

[78] Yinan Xuan, Colin Barry, Jessica De Souza, Jessica H Wen, Nick Antipa, Alison A Moore, and Edward J Wang. 2023. Ultra-low-cost mechanical smartphone attachment for no-calibration blood pressure measurement. *Scientific Reports* 13, 1 (2023), 8105.

[79] Bangbang Yang, Yinda Zhang, Yijin Li, Zhaopeng Cui, Sean Fanello, Hujun Bao, and Guofeng Zhang. 2022. Neural rendering in a room: amodal 3D understanding and free-viewpoint rendering for the closed scene composed of pre-captured objects. *ACM Trans. Graph.* 41, 4 (2022), 101:1–101:10.

[80] Yanbing Yang, Jie Hao, and Jun Luo. 2017. CeilingTalk: Lightweight Indoor Broadcast Through LED-Camera Communication. *IEEE Trans. Mob. Comput.* 16, 12 (2017), 3308–3319.

[81] Yanbing Yang, Jiangtian Nie, and Jun Luo. 2017. ReflexCode: Coding with Superposed Reflection Light for LED-Camera Communication. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking, MobiCom'17*. ACM, 193–205.

[82] Hanting Ye and Qing Wang. 2021. SpiderWeb: Enabling Through-Screen Visible Light Communication. In *The 19th ACM Conference on Embedded Networked Sensor Systems, SenSys'21*. ACM, 316–328.

[83] Hanting Ye, Jie Xiong, and Qing Wang. 2023. When VLC Meets Under-Screen Camera. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services, MobiSys'23*. ACM, 343–355.

[84] Chi Zhang and Xinyu Zhang. 2016. LiTell: robust indoor localization using unmodified light fixtures. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, MobiCom'16*. ACM, 230–242.

[85] Chi Zhang and Xinyu Zhang. 2017. Pulsar: Towards Ubiquitous Visible Light Localization. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking, MobiCom'17*. ACM, 208–221.

[86] Kai Zhang, Chenshu Wu, Chaofan Yang, Yi Zhao, Kehong Huang, Chunyi Peng, Yunhao Liu, and Zheng Yang. 2018. ChromaCode: A Fully Imperceptible Screen-Camera Communication System. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking, MobiCom'18*. ACM, 575–590.

[87] Xiao Zhang, Griffin Klevering, Juexing Wang, Li Xiao, and Tianxing Li. 2023. RoFin: 3D Hand Pose Reconstructing via 2D Rolling Fingertips. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services, MobiSys'23*. ACM, 330–342.

[88] Tian Zhao, Yu-Lin Wei, Wei-Nin Chang, Xi Xiong, Changxi Zheng, Hsin-Mu Tsai, Kate Ching-Ju Lin, and Xia Zhou. 2018. Augmenting Indoor Inertial Tracking with Polarized Light. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys'18*. ACM, 362–375.

[89] Tian Zhao, Kevin Wright, and Xia Zhou. 2016. The darkLight rises: visible light communication in the dark. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, MobiCom'16*. ACM, 2–15.

[90] Xun Zhou, A. Kai Qin, Maoguo Gong, and Kay Chen Tan. 2021. A Survey on Evolutionary Construction of Deep Neural Networks. *IEEE Trans. Evol. Comput.* 25, 5 (2021), 894–912.

[91] Zhi Zhou, Xu Chen, En Li, Liekang Zeng, Ke Luo, and Junshan Zhang. 2019. Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing. *Proc. IEEE* 107, 8 (2019), 1738–1762.

[92] Shilin Zhu and Xinyu Zhang. 2017. Enabling High-Precision Visible Light Localization in Today's Buildings. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys'17*. ACM, 96–108.