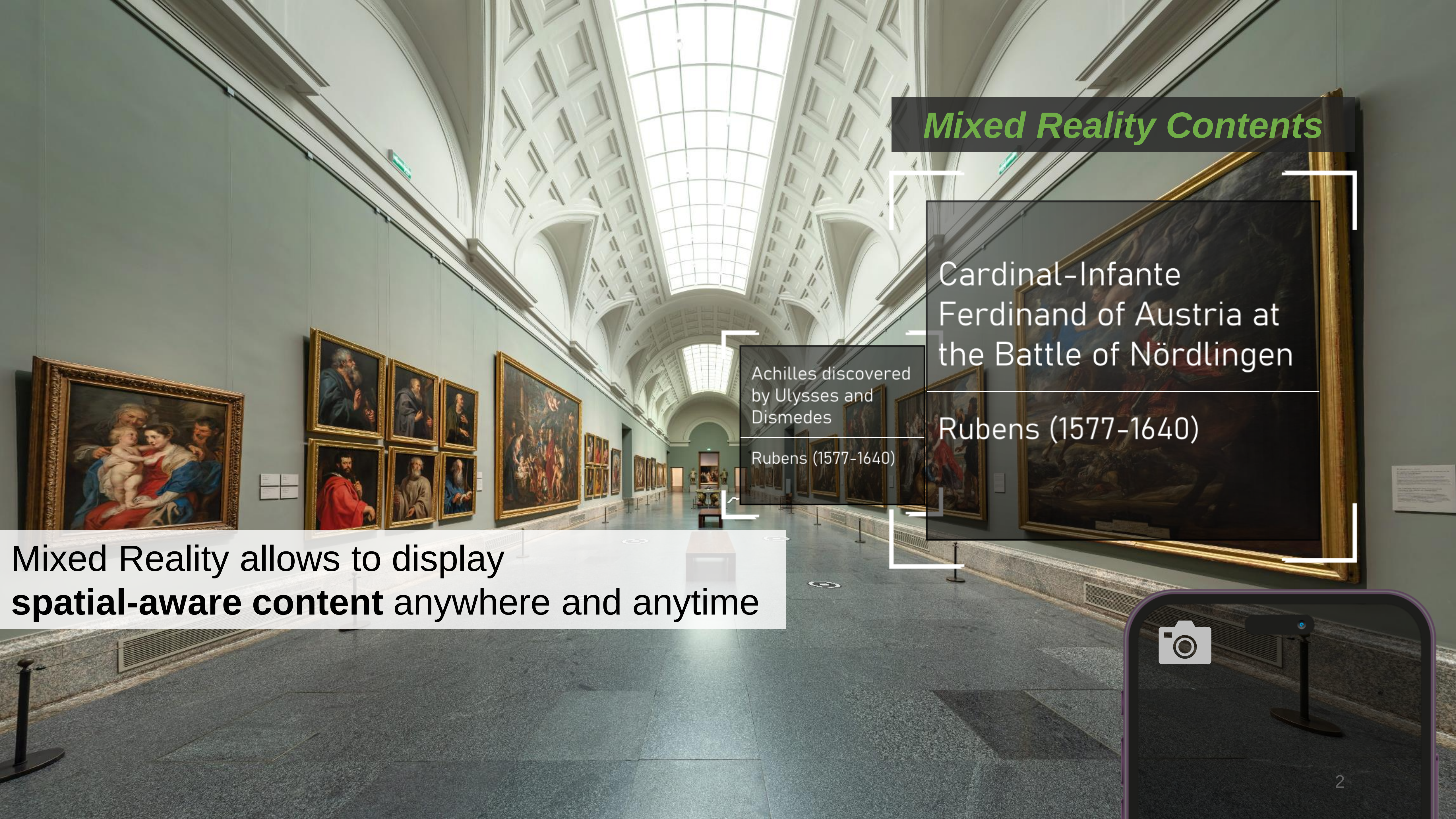
Practical Optical Camera Communication Behind Unseen and Complex Backgrounds

Rui Xiao[1], Leqi Zhao[1], Feng Qian[2], Lei Yang[2], and Jinsong Han[1]

[1] Zhejiang University

[2]Ant Group

Mixed Reality Contents



Mixed Reality allows to display **spatial-aware content** anywhere and anytime

Problem: How to **align** these **virtual contents** on **precise** objects or locations?

Low Overhead Solution?


Book Store


Shopping Center


Museum
One thing in common – *Lights!*


Grocery

Optical Camera Communication (OCC)

Feature 1 - ***Location-awareness***: inherently ***links data*** to light’s ***location***

Feature 2 - ***Pervasiveness***: LED lights and cameras are pervasive

0101010101
01010110…

What if we reuses ***Lights*** as ***transmitters***, and ***cameras*** as ***receivers***…

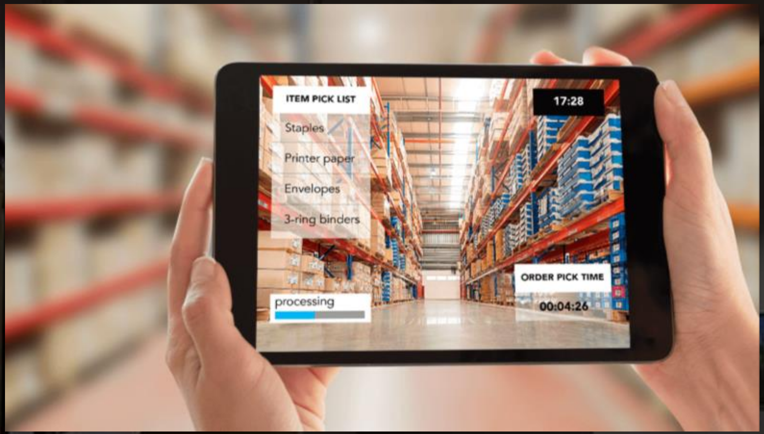
One thing in common – ***Lights!***

Modulated Blinking

Optical Camera Communication (OCC)

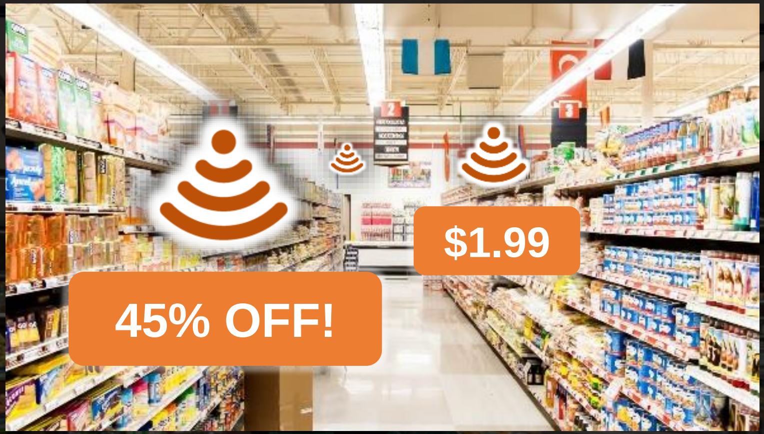
Feature 1 - ***Location-awareness***: inherently ***links data*** to light’s ***location***

Feature 2 - ***Pervasiveness***: LED lights and cameras are pervasive



MR Content Delivery

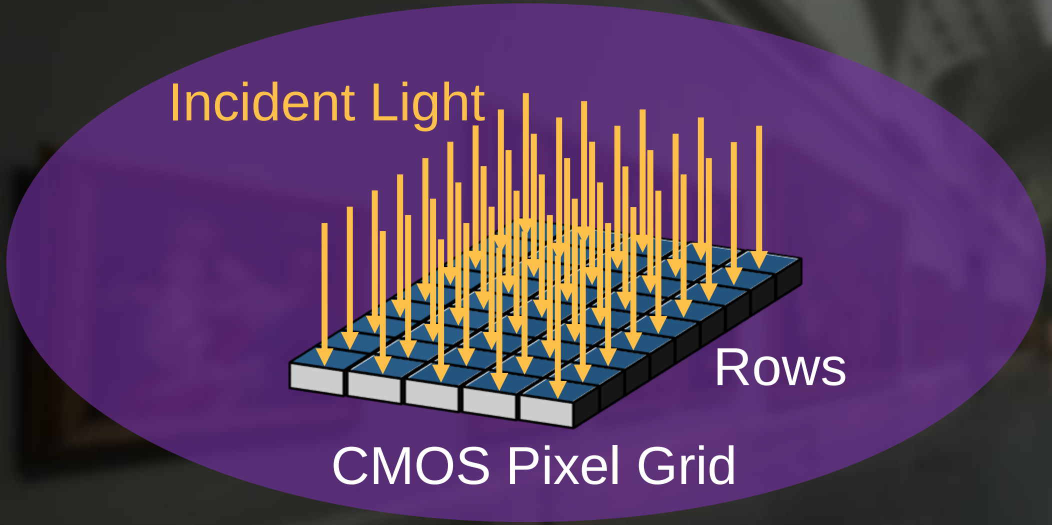
Indoor Localization

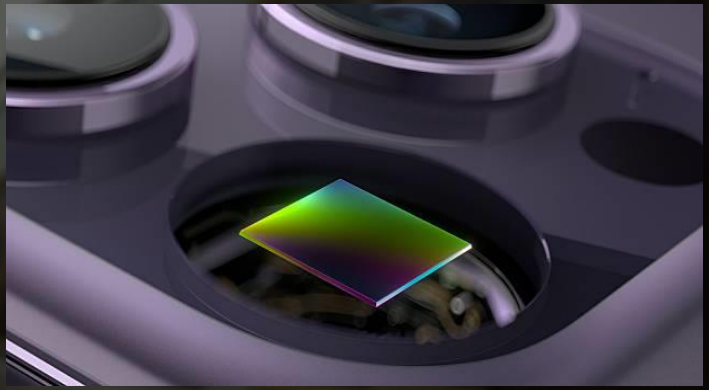


Pervasive Connectivity

OCC with Rolling Shutter CMOS

- Rolling shutter exposes a frame ***column by column***

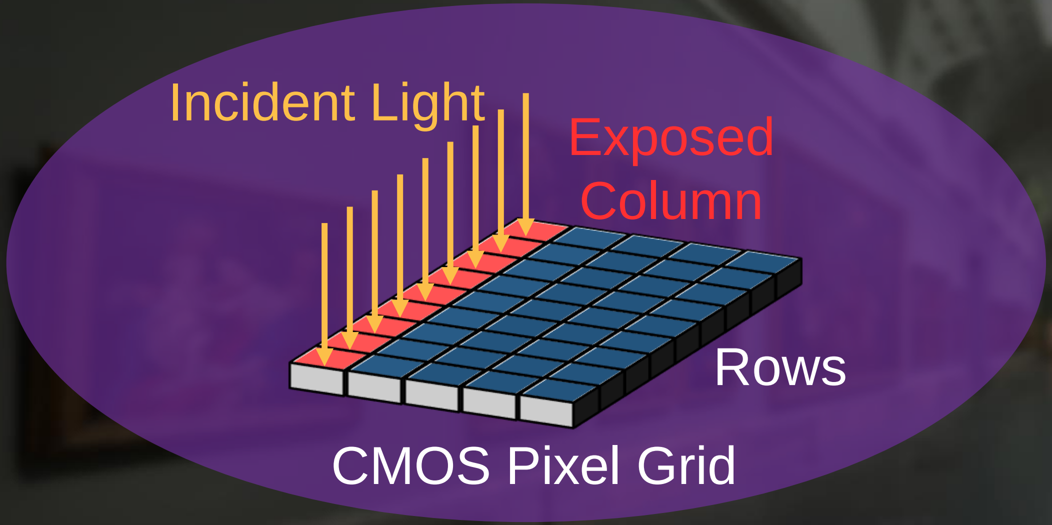




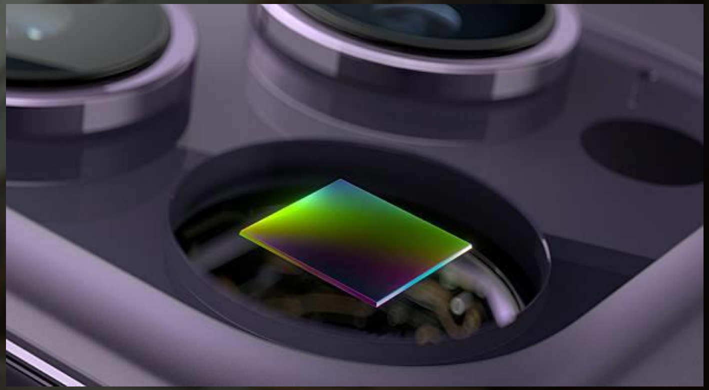
Rolling Shutter CMOS

OCC with Rolling Shutter CMOS

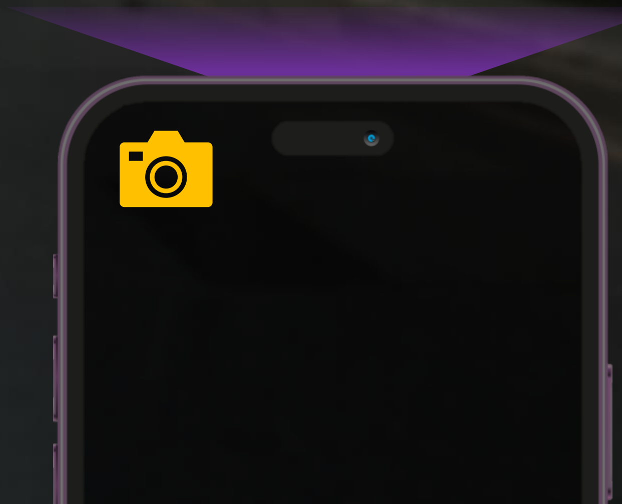
- Rolling shutter exposes a frame ***column by column***



CMOS Pixel Grid

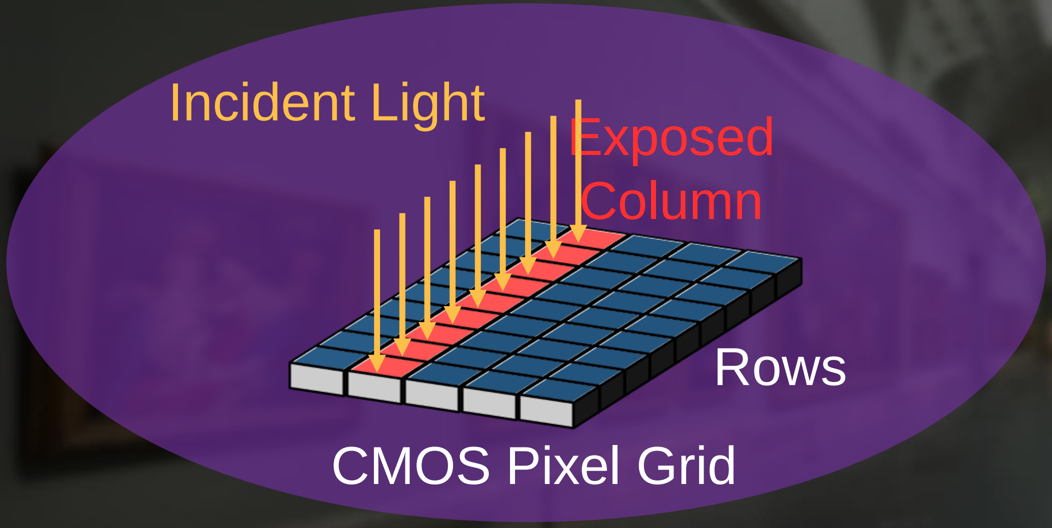


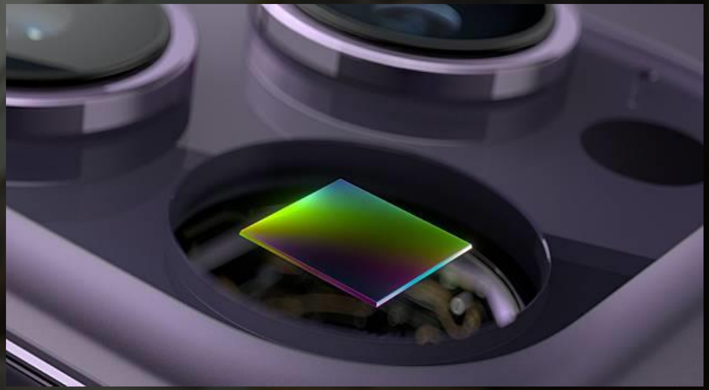
Rolling Shutter CMOS



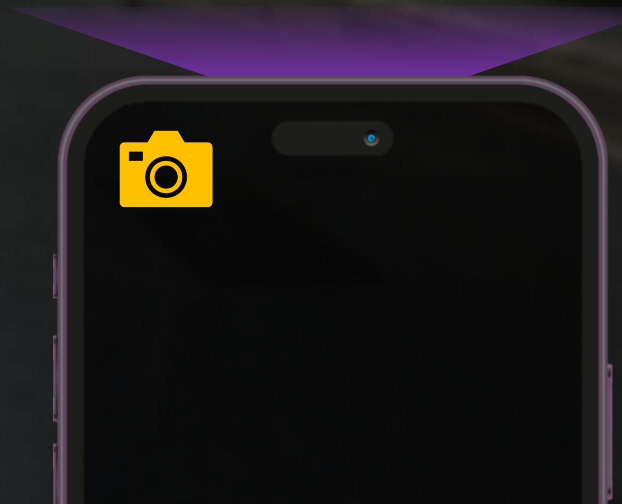
OCC with Rolling Shutter CMOS

- Rolling shutter exposes a frame ***column by column***



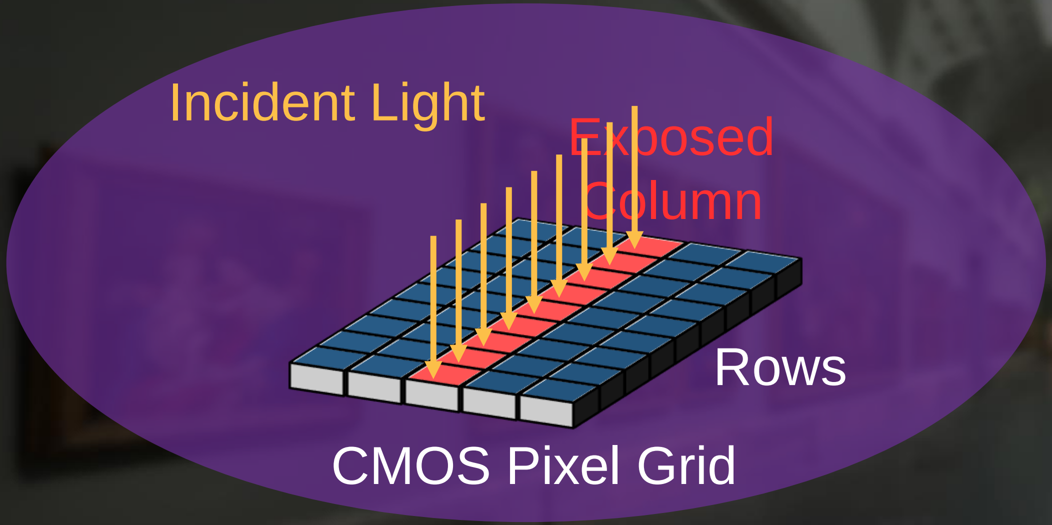


Rolling Shutter CMOS

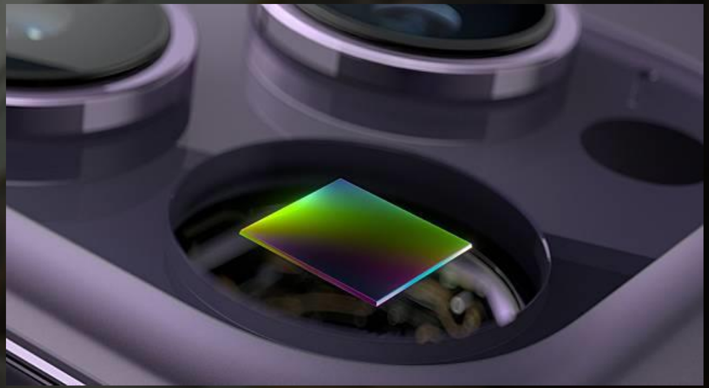


OCC with Rolling Shutter CMOS

- Rolling shutter exposes a frame ***column by column***

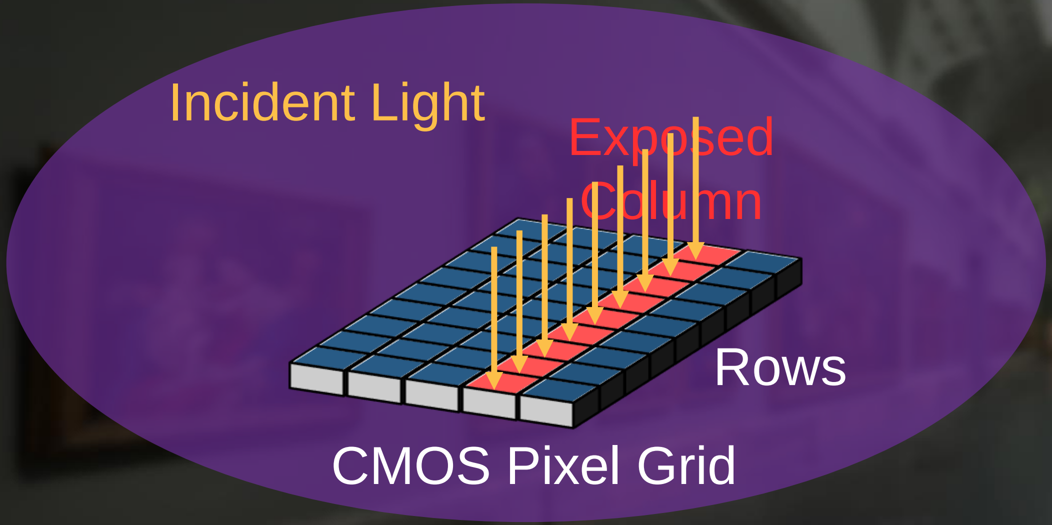


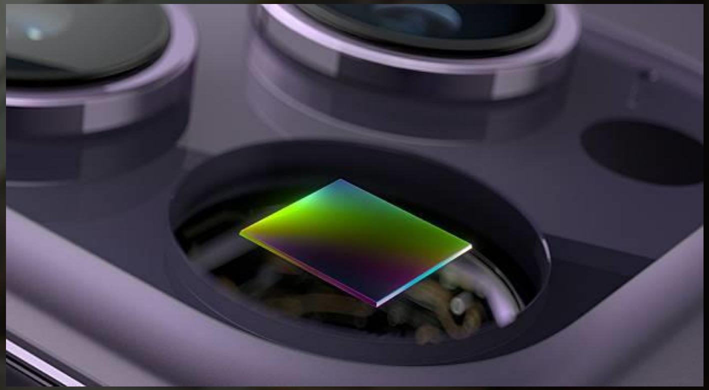
CMOS Pixel Grid


Rolling Shutter CMOS

OCC with Rolling Shutter CMOS

- Rolling shutter exposes a frame ***column by column***

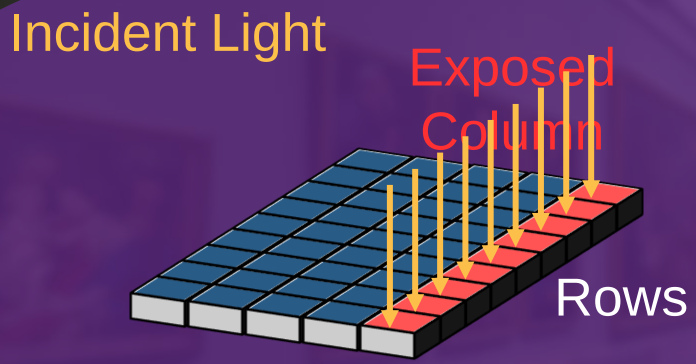




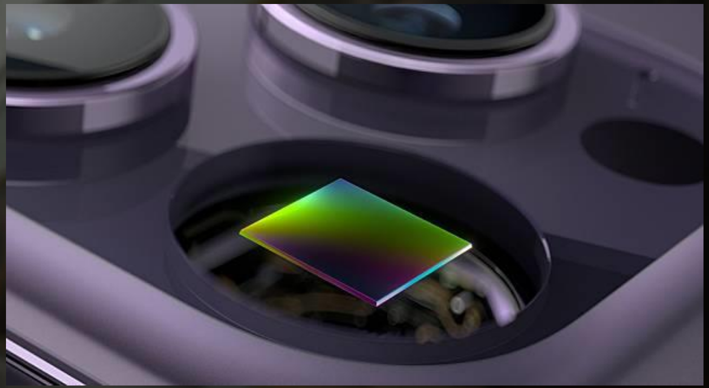
Rolling Shutter CMOS

OCC with Rolling Shutter CMOS

- Rolling shutter exposes a frame ***column by column***



CMOS Pixel Grid

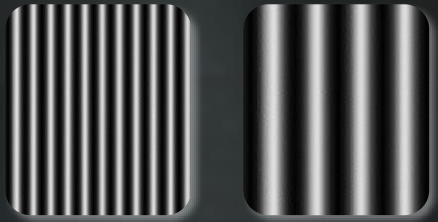


Rolling Shutter CMOS

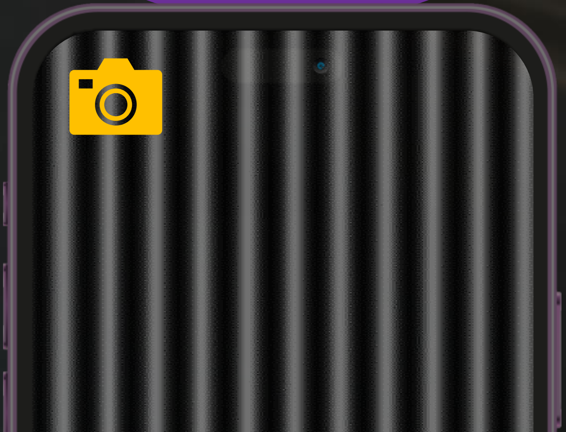
Modulated Blinking (~KHz)

e.g. FSK

Modulated Stripes



Bit 0 **Bit 1**



Issues with Existing OCC Designs

Signal-to-Noise Ratio (SNR) → Symbol Error Rate (SER)

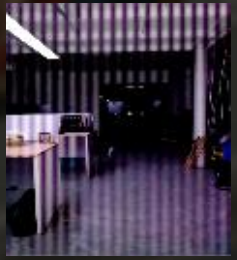
Issues with Existing OCC Designs (Cont.)

① 

-12.8dB

low SNR but ***clean background***

- Limited distance (~0.4m)
- Often difficult to find such a clean reflector

② 

1.1dB

complex background but ***high SNR***

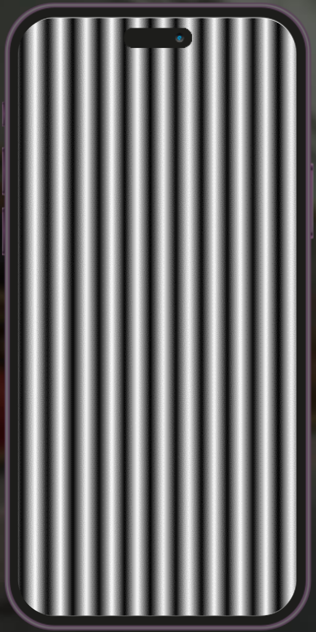
- Still limited distance (~1.4m)
- Pronounced video degradation

WinkLink!

Can we build a novel OCC system that works under *low SNR* with any *unseen complex backgrounds* ?

Issues with Existing OCC Designs (Cont.)

SNR: 20 dB



Traditional OCC

vs.

SNR: -20 dB

WinkLink

Under Low SNR ✔

Complex Background ✔

WinkLink!

Can we build a novel OCC system that works under *low SNR* with any *unseen complex backgrounds* ?

Challenge 1: Dynamic Background Entanglement

- The entanglement can be modeled as:

=

+

Video Frame $\boldsymbol{I}$ | Background $\boldsymbol{B}$ | OCC Signals $\boldsymbol{O}$

(what we capture) | ***(what we want)***

- Disentangling Signal O is an ***ill-posed*** problem.
- The background B is ***dynamic***, varying across different frame I.

Solution: DNN-based Signal Extraction

- The entanglement can be modeled as:

Video Frame I

=

Background B

+

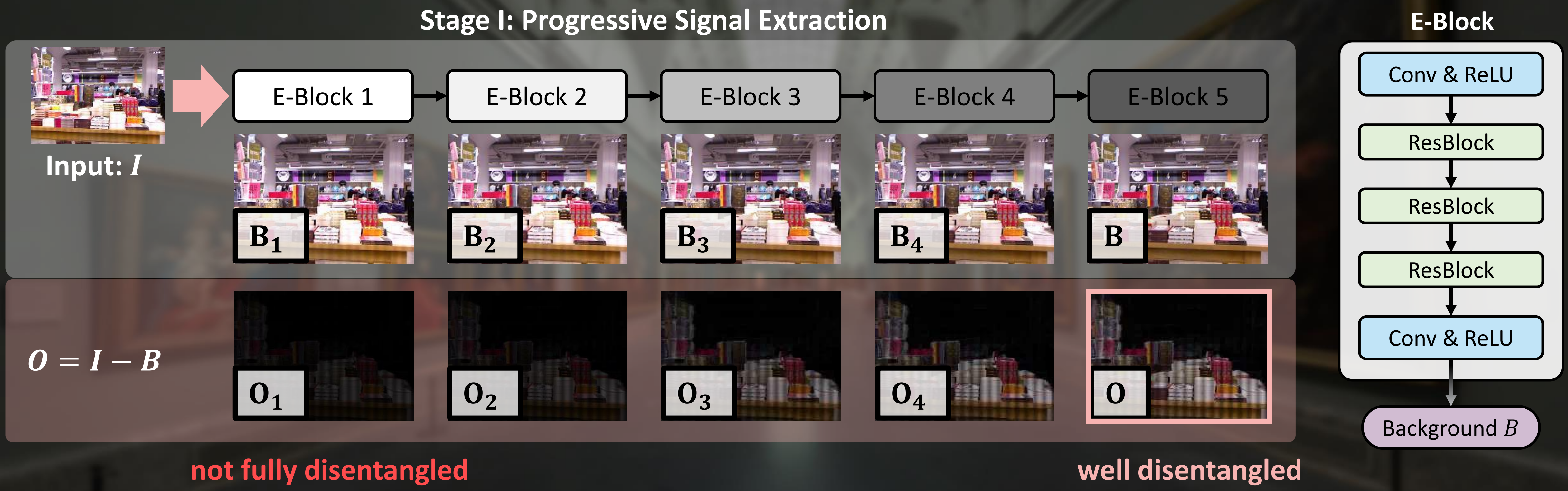
OCC Signals O

- Key insight:

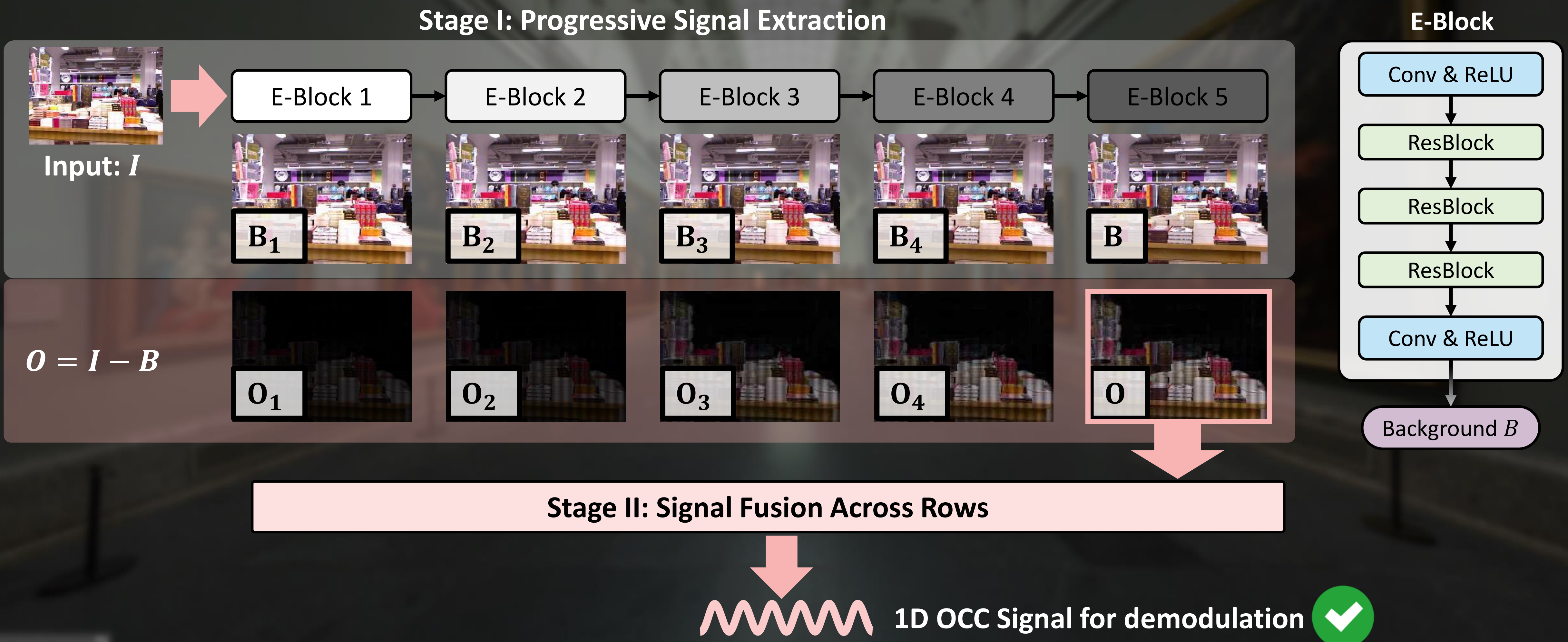
1. ***DNN can handle ill-posed problem*** by implicitly enforcing constraints
2. Replication of signals across rows → Spatial correlation

   (DNNs excel at capturing spatial correlation)

Solution: DNN-based Signal Extraction



Solution: DNN-based Signal Extraction



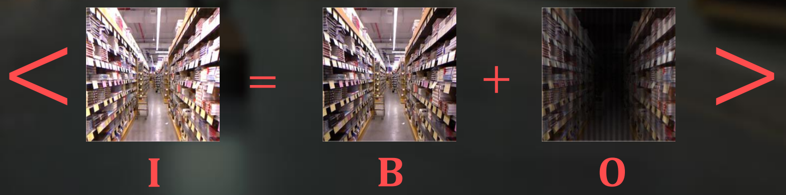
Challenge 2: Laborious Training Data Preparation

Generalizability of WinkLink

1. Unseen backgrounds
2. Diverse stripe patterns
3. Varying SNRs

↑

Diversity of Training Data



- We require a **large dataset** of **paired** $< \boldsymbol{I}, \boldsymbol{B}, \boldsymbol{O} >$ with above diversities.
- Manual assembly of such a diverse dataset is ***time-intensive*** and ***impractical***.

Solution: Scalable Training Data Synthesis

$<$  $=$  $+$  $>$

I **B** **O**

- Target: synthesize **paired** $< \boldsymbol{I}, \boldsymbol{B}, \boldsymbol{O} >$ -- *easy to scale*

  while ***minimizing the gap*** between synthetic and real data
- ***Key components*** for precise synthesis:
  - the *light reflection* model under the Lambertian assumption
  - the *light attenuation* on varying distances

Solution: Scalable Training Data Synthesis

- Input:

Depth map
(only for training)

Background B

- Output:

Frame I

Signal O

Solution: Scalable Training Data Synthesis

- Input:

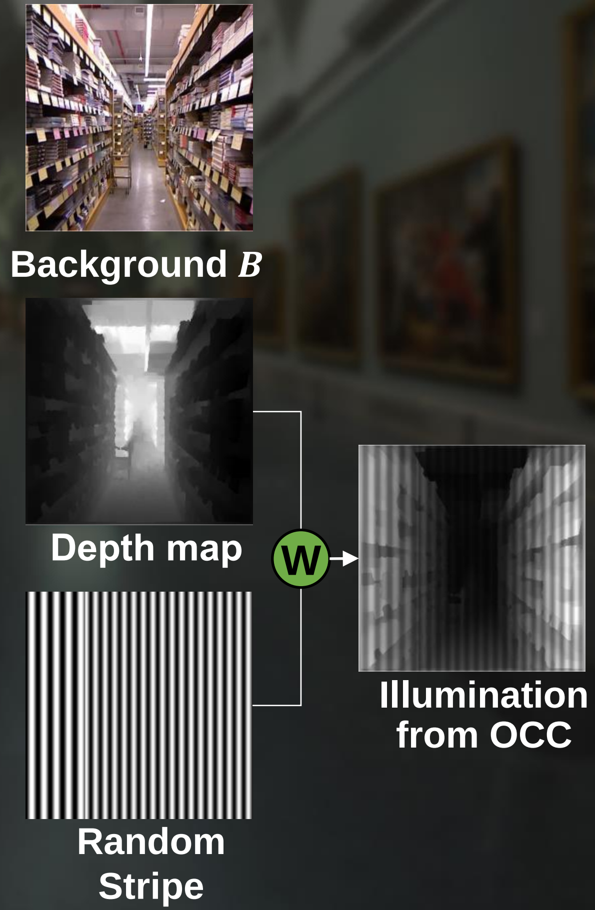
Depth map
(only for training)

Background B

- Output:

Frame I

Signal O



-- modeling light attenuation

W : Weight by Depth

Solution: Scalable Training Data Synthesis

- Input:

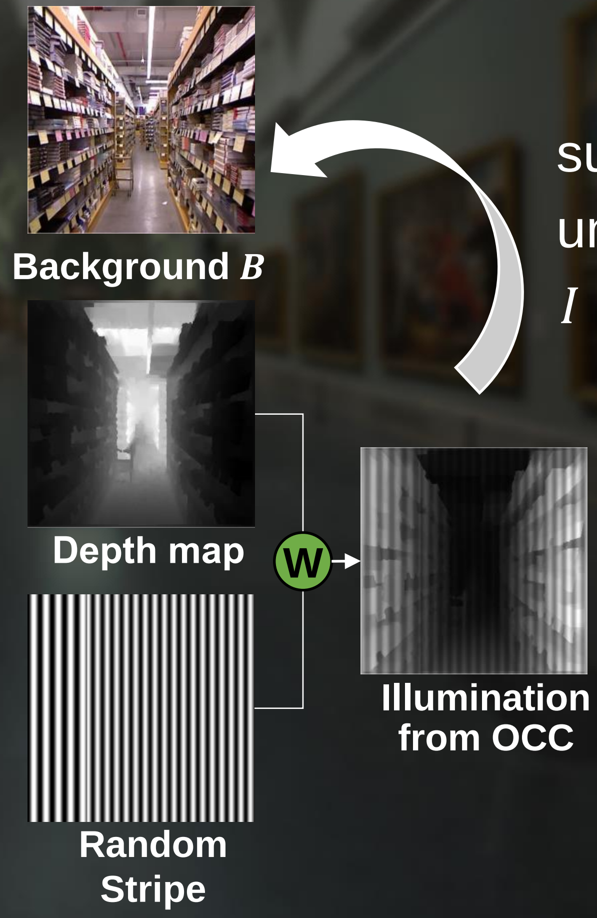

Depth map
(only for training)


Background B

- Output:


Frame I


Signal O



superimpose
under **Lambertian assumption**

$$I = Reflectance \times Illumination$$

W : Weight by Depth

Solution: Scalable Training Data Synthesis

- Input:

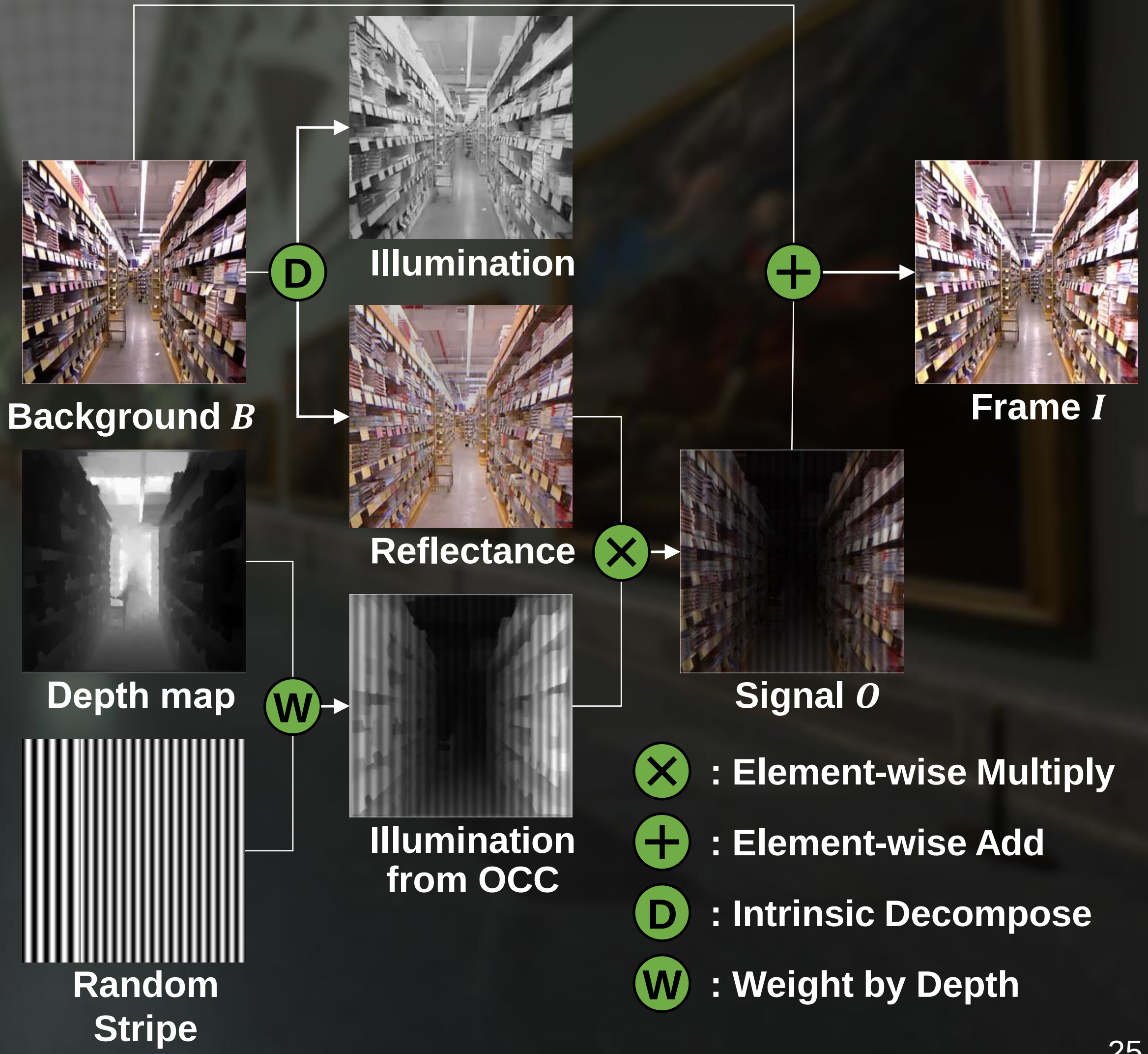
Depth map
(only for training)

Background B

- Output:

Frame I

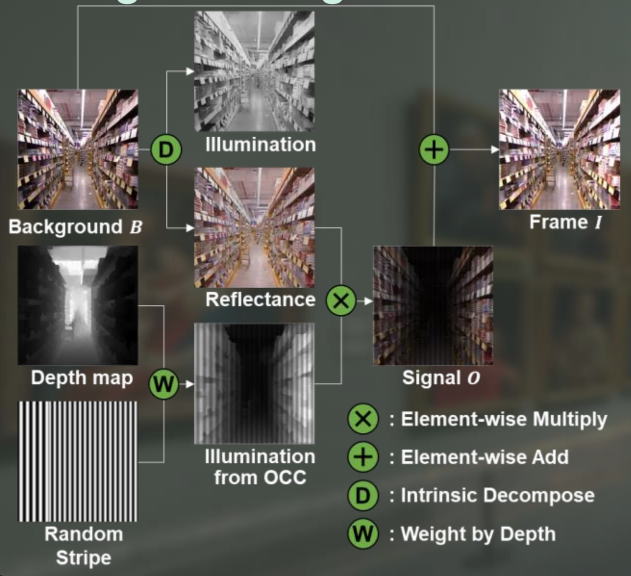
Signal O



- Source dataset: NYU Depth V2 dataset
- We synthesize **7245** frames.

Design Overview of *WinkLink*

Solving Challenge 2



Transmitter

Bootstrapping

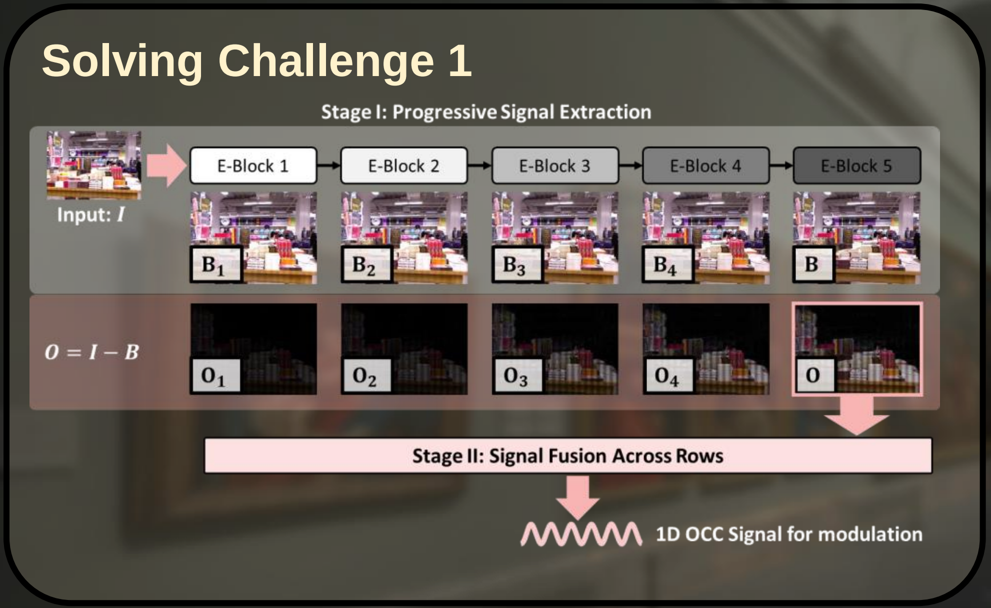
Deployment

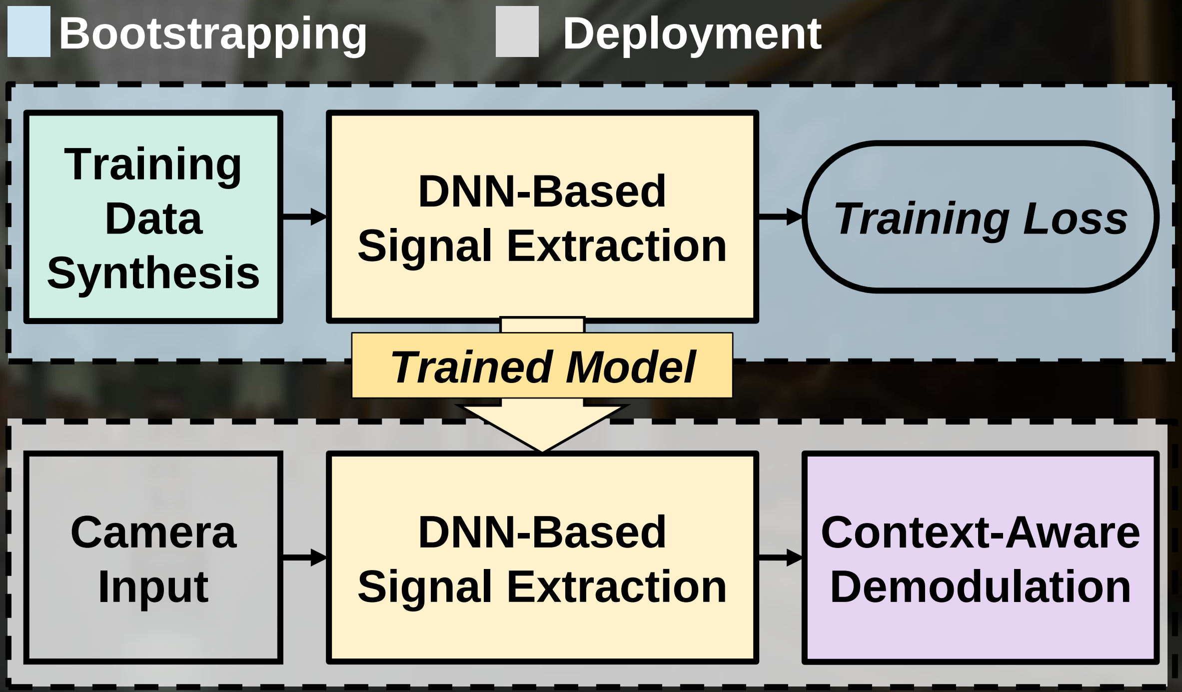
Training Data Synthesis

Camera Input

Receiver

Design Overview of *WinkLink*





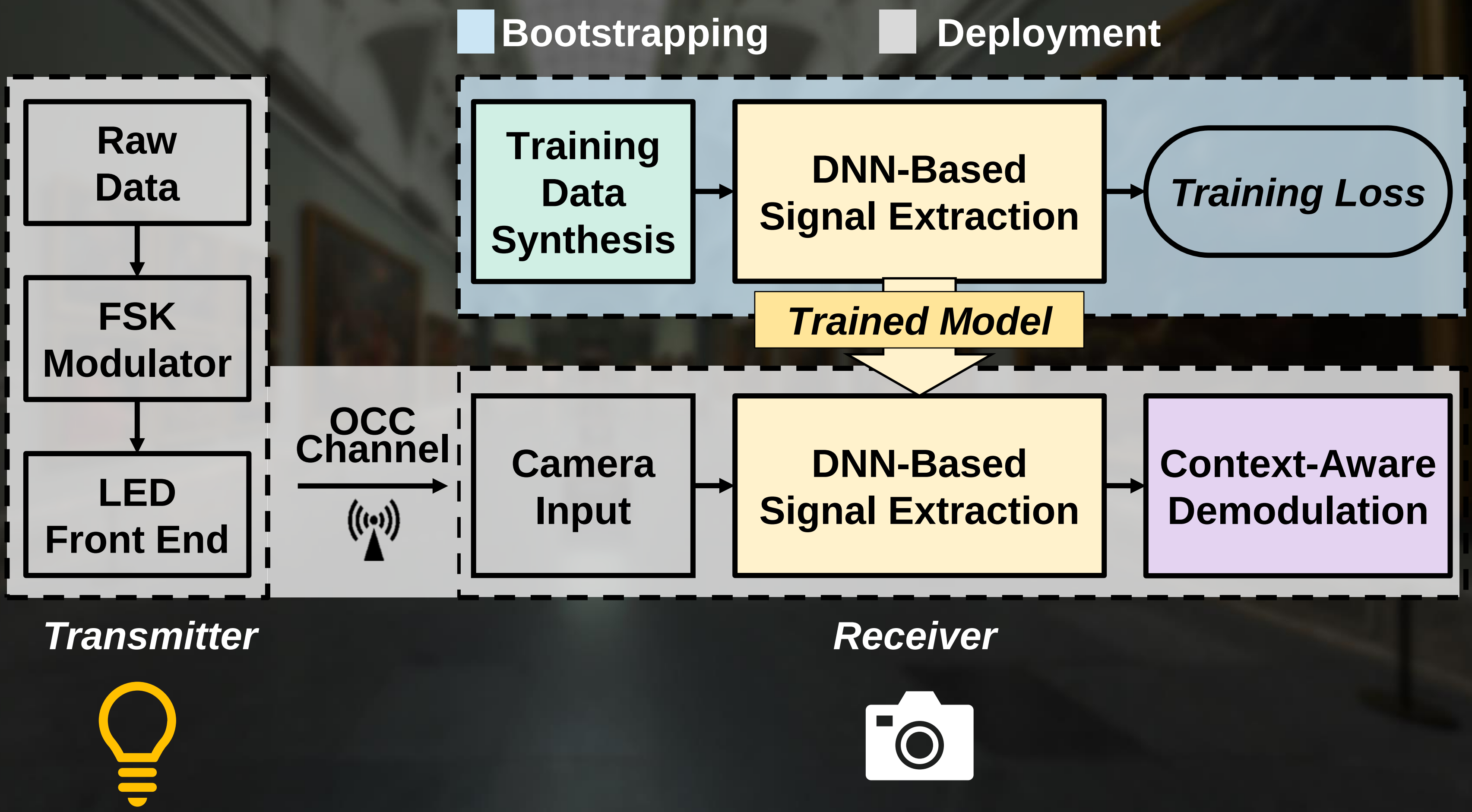
Transmitter

Receiver

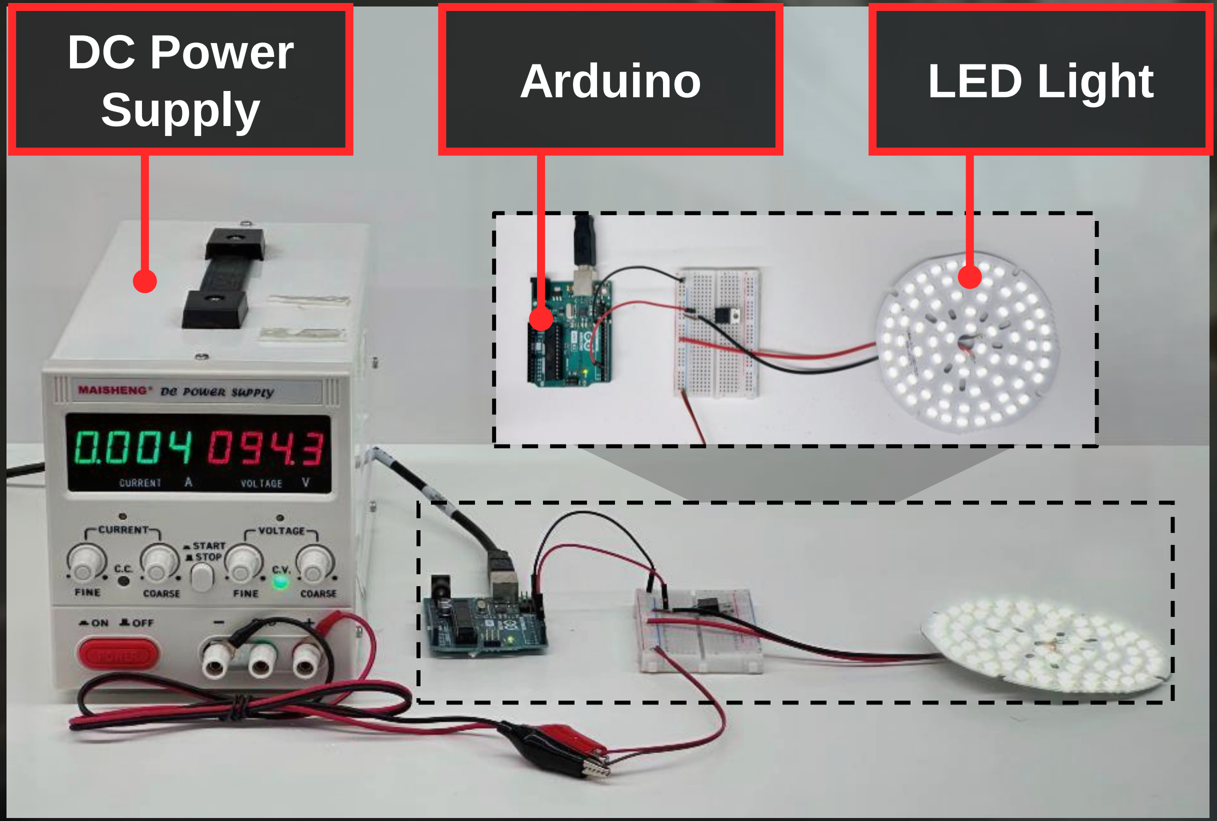
Design Overview of *WinkLink*

Under Low SNR ✔

Complex Background ✔

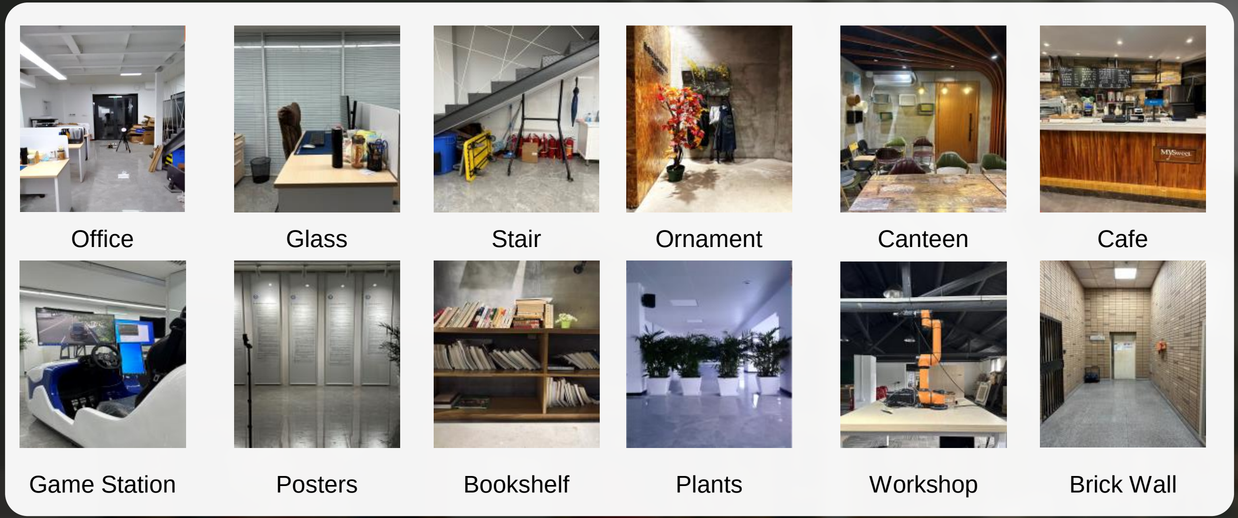


Prototype and Implementation



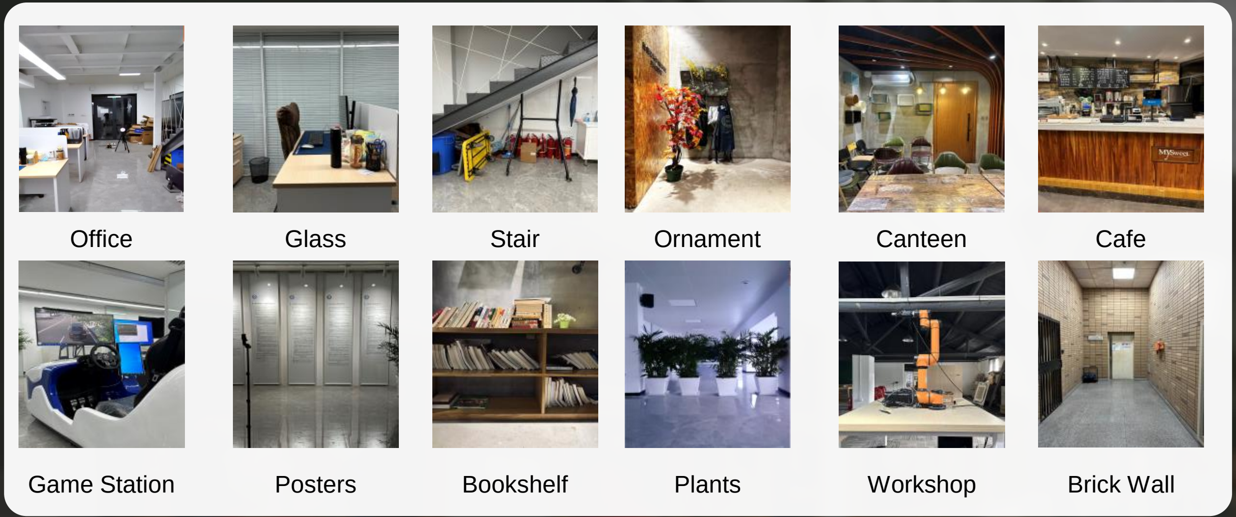
- Transmitter
  - 5 watt LED
  - Modulation: 4-FSK
  - Each symbol: 1/60 seconds (120 bps)
- Receiver
  - Phone (iPhone/Samsung/Huawei)
  - Frame Rate: 60 FPS
  - Resolution: 512x512

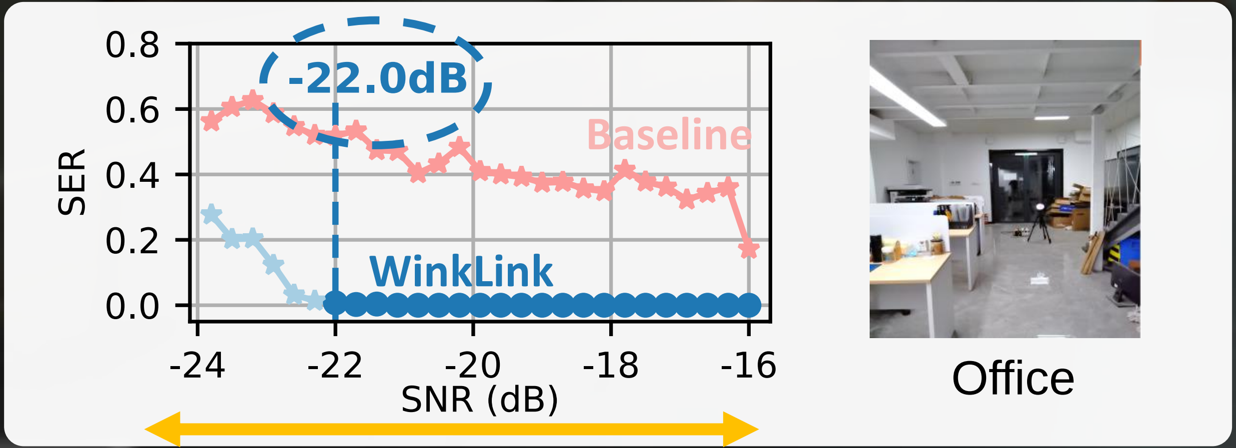
Overall Performance



- Test Dataset
  - 12 unseen environments
  - 30K frames per environment
    - ***genuinely captured*** and ***not synthetic***

Overall Performance

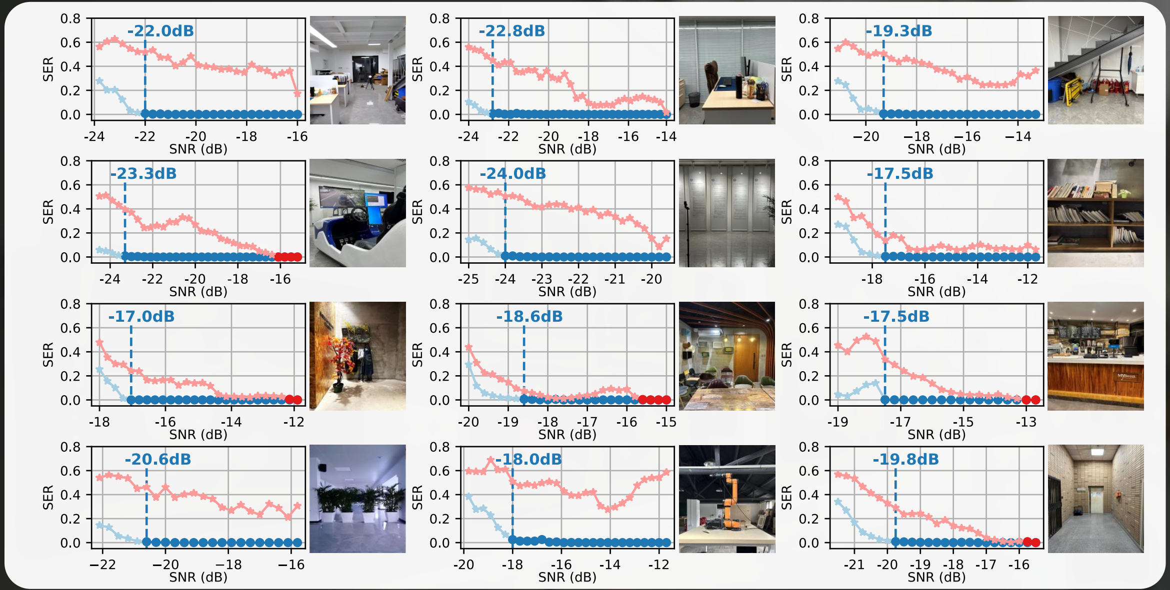




Environment 1/12

- Test Dataset
  - 12 unseen environments
  - 30K frames per environment
    - ***genuinely captured*** and ***not synthetic***
    - **SNR variation** by adjusting **LED power** from 0 to 5 watts
- WinkLink vs. Baseline
- Metric: **mSNR** – minimum SNR at which SER drops below 0.01

Overall Performance



- Test Dataset
  - 12 unseen environments
  - 30K frames per environment
    - ***genuinely captured*** and ***not synthetic***
    - **SNR variation** by adjusting **LED power** from 0 to 5 watts
- WinkLink vs. Baseline
- Metric: **mSNR** – minimum SNR at which SER drops below 0.01

➢ Results

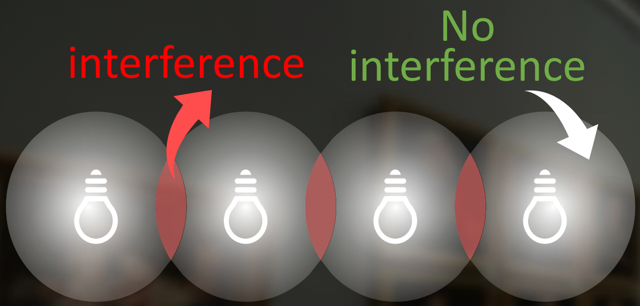
- **WinkLink** shows an average mSNR of **-20 dB**
- Consistently outperforms baseline with a **5.8 dB** SNR gain

Summary of Other Evaluation Results

- Works with **three distinct phone models**: iPhone 14 Pro, Huawei P40 Pro, and Samsung Galaxy S21
- Performs well when the user is **moving** at speed of 2m/s (under **dynamic backgrounds**)
- Works at a distance up to **11 meters** with a **10 watt LED**
- Has **minimized interference** on concurrent vision applications (e.g., object detection)

Discussion

- Interference between Multiple OCC Links



Space Division

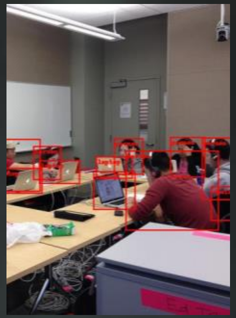
Light steering

Frequency Division

- Integrated Sensing and Communication on Vision

Sensing with Vision

Segmentation



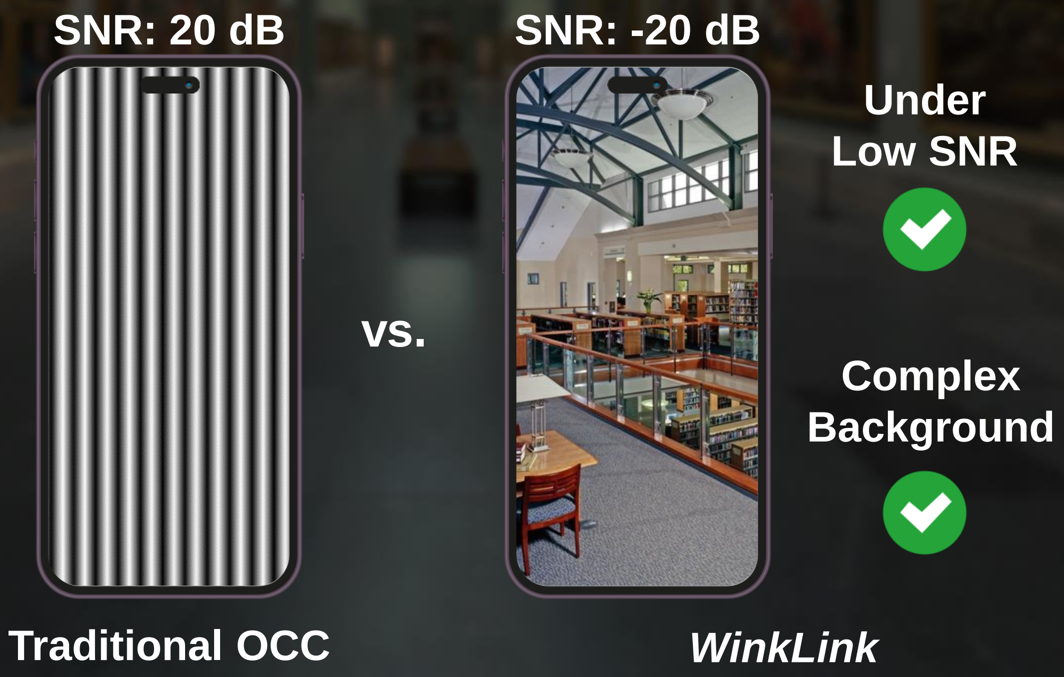
Detection

Communication with WinkLink

1001010111010100 10 ⋯

Conclusion

- We propose WinkLink, a novel OCC system that operates under **unseen and complex backgrounds**, while maintaining **low-SNR requirements**.
- We hope to explore the **integrated sensing and communication** on vision domain.



Thank you!

Homepage

Mail: ruixiao24@zju.edu.cn

Rui Xiao

PhD candidate at Zhejiang University

Mobile, Wireless, Sensing, Security

I'm seeking a ***post-doctoral position*** starting in ***Fall 2025***.
Please feel free to contact me!