UDC-VIT: A Real-World Video Dataset for Under-Display Cameras


Kyusu Ahn[3*]


JiSoo Kim[1]


Sangik Lee[4†]


HyunGyu Lee[2]


Byeonghyun Ko[2]


Chanwoo Park[2]


Jaejin Lee[1,2]

[1] Dept. of Data Science, Graduate School of Data Science, Seoul National University

[2] Dept. of Computer Science and Engineering, Seoul National University

[3] CAE Team, Research Center, Samsung Display Co., Ltd.

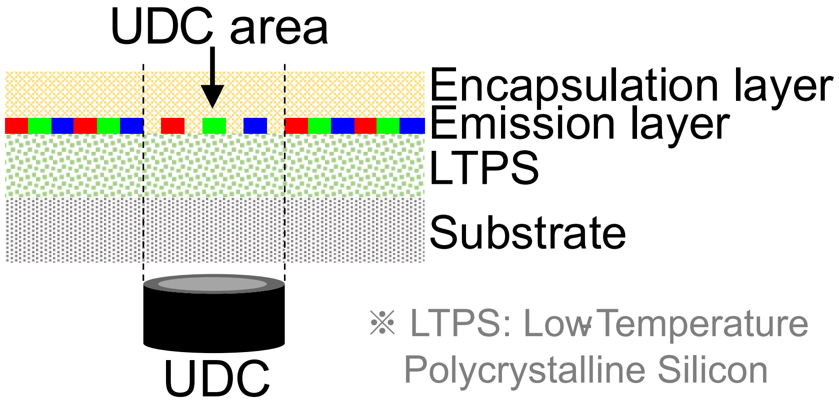
[4] Mobile Display Electronics Development Team, Samsung Display Co., Ltd.

* This work was done when he was a Ph.D. student at Seoul National University.

† This work was done when he was an M.S. student at Seoul National University.

Under-Display Camera (UDC)

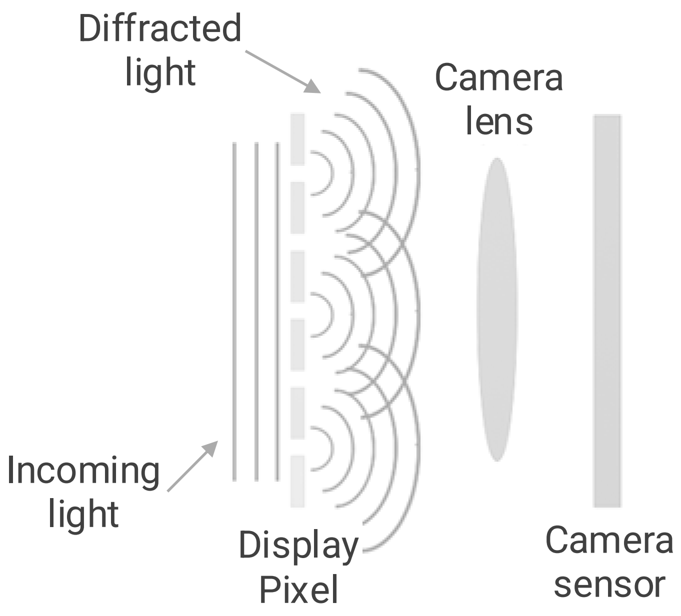
- An imaging system where the camera is positioned beneath the display
- Use the UDC area as a display space and take pictures when the camera operates



※ LTPS: Low Temperature Polycrystalline Silicon

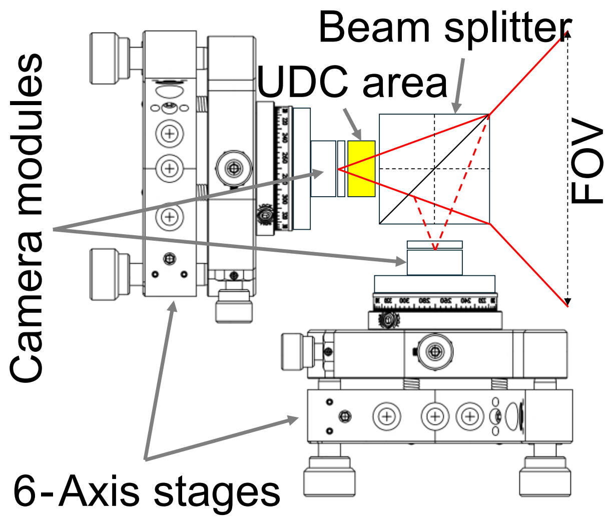
Motivation

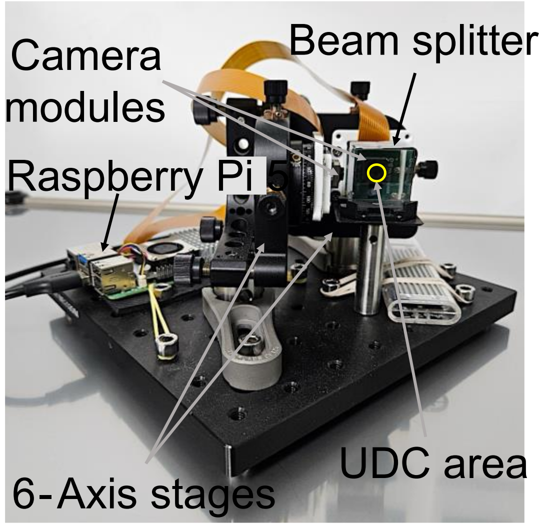
- **UDC degradation**
  - The pixels diffract the light traveling through the camera lens
  - Complex degradations can occur in a single image or video frame



Video Capturing System

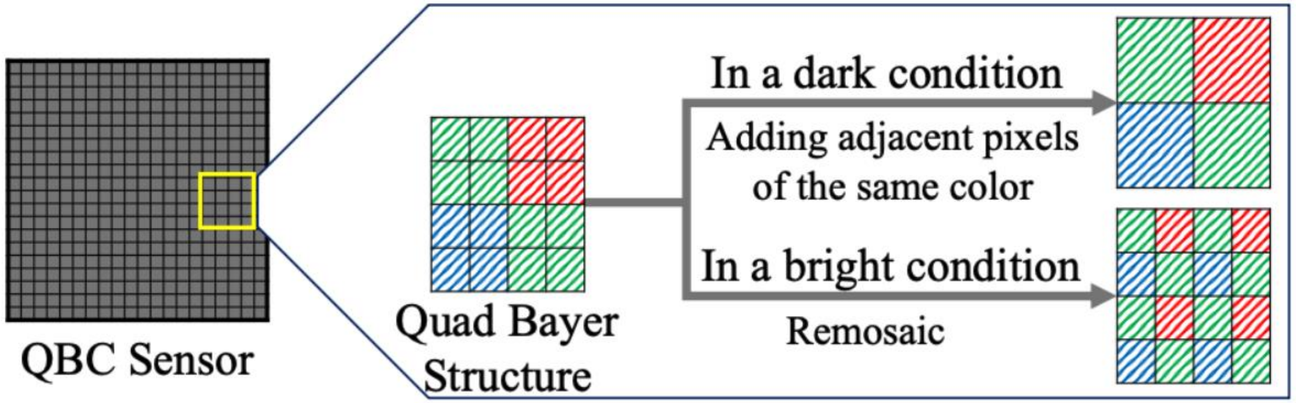
- **Constructing a real-world UDC video dataset**
  - Finding a matching pair of UDC-distorted and ground-truth videos with high alignment accuracy
  - Synchronize the time for all frames when capturing videos





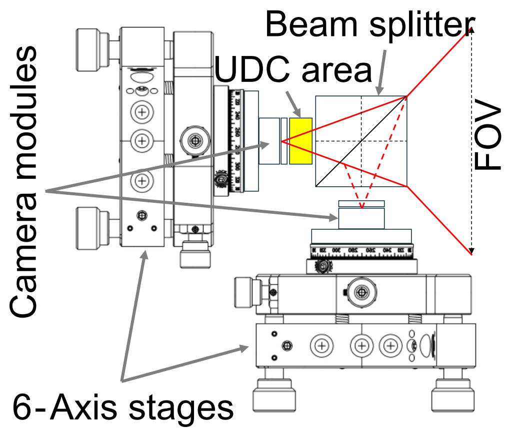
Camera Module

- **Quad Bayer Coding (QBC)**
  - In low-light conditions, four **adjacent pixels are grouped** to reduce noise
  - In bright conditions, the sensor **reverts the pixels** to the Bayer structure



Beam Splitter & Optical Mount

- **Beam splitter**
  - Capture the same scene
  - Align the two cameras to the beam splitter's split fields of view (FOV)



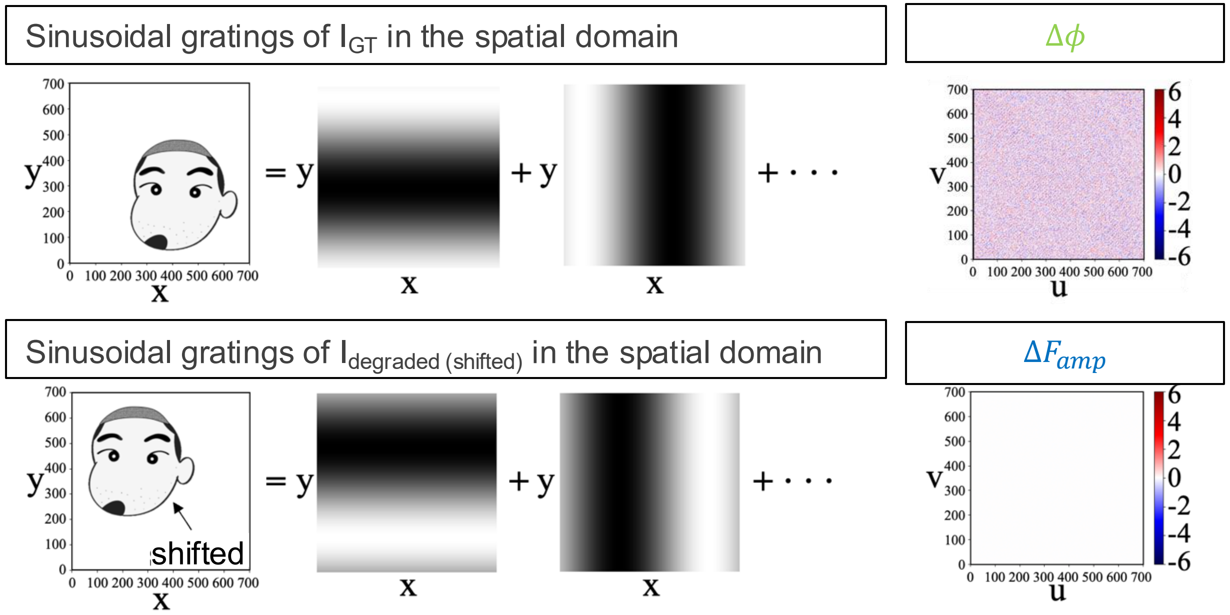
- **Optical Mount**
  - Each camera module is mounted on a K6XS mount
  - *Shift*, *rotate*, and *tilt* across the six axes to align their FOV

Controller

- **Dual-camera with Raspberry Pi 5**
  - Synchronize the two cameras by using MPI barriers
  - Frame difference < 0.5 fps (8 msec)
  - Excluding fast-motion scenes (e.g., cars)

Obtaining Aligned Video Pairs

- **Alignment using DFT**
  - Measure similarity in both spatial and frequency domains
  - Metrics: ΔMSE, amplitude difference (ΔF_{amp}), phase difference ($\Delta\phi$)
  - Phase consistency ($\Delta\phi$) is key for alignment



Alignment Accuracy

- **Alignment accuracy**
  - Pseudo-real attains a PCK value of 58.75% at α = 0.01
  - **UDC-VIT** achieves consistently high PCK values (92.12–99.69%)

Dataset	Alignment Required	Alignment Method	PCK (Ratio of correctly aligned keypoints to the total number)			
			$\alpha = 0.003$	$\alpha = 0.01$	$\alpha = 0.03$	$\alpha = 0.10$
Pseudo-real	✓	AlignFormer	**N/A**	**58.75**	**95.08**	**99.93**
UDC-SIT	✓	DFT	**93.67**	**97.26**	**98.56**	**99.35**
VidUDC33K			99.65	99.82	99.84	99.90
UDC-VIT	✓		**85.10**	**98.65**	**99.22**	**99.64**
UDC-VIT	✓	DFT	**92.12**	**98.95**	**99.32**	**99.69**

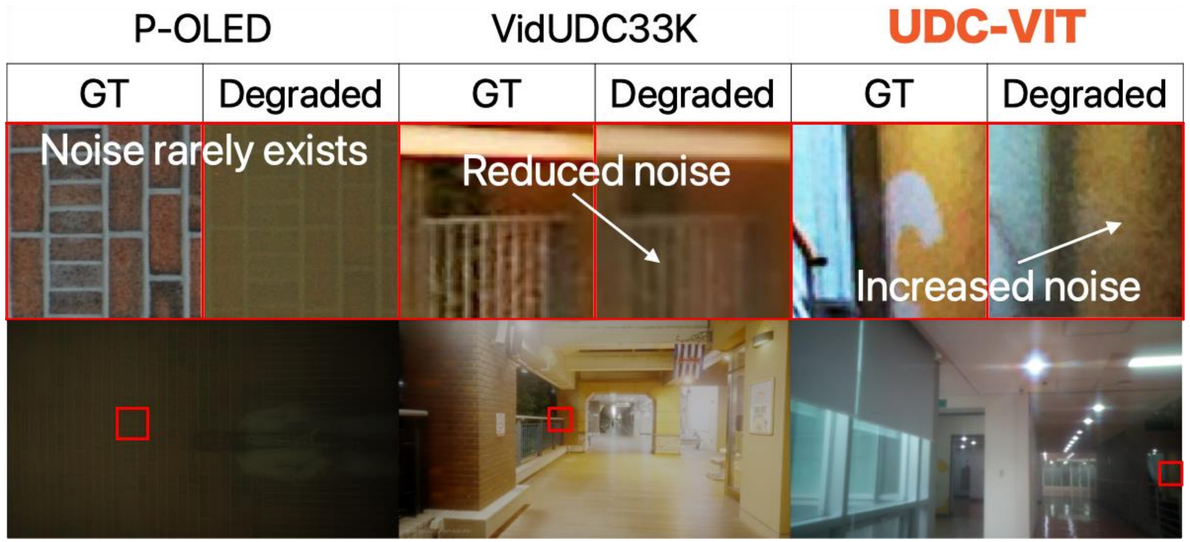
Dataset Comparison

- The first real-world UDC video dataset
- Contain flare and face recognition scenarios

Dataset	Type	Scene	Dataset size	Flare presence	Variant flares	Face recognition	Publicly available	Publication
PexelsUDC-T/P	**Video**	Synthetic	160×100 (16,000)					arXiv '23
VidUDC33K	**Video**	Synthetic	677×50 (33,850)	✓			✓	AAAI '24
UDC-VIT	**Video**	**Real**	647×180 (116,460)	✓	✓	✓	✓	ICCV '25

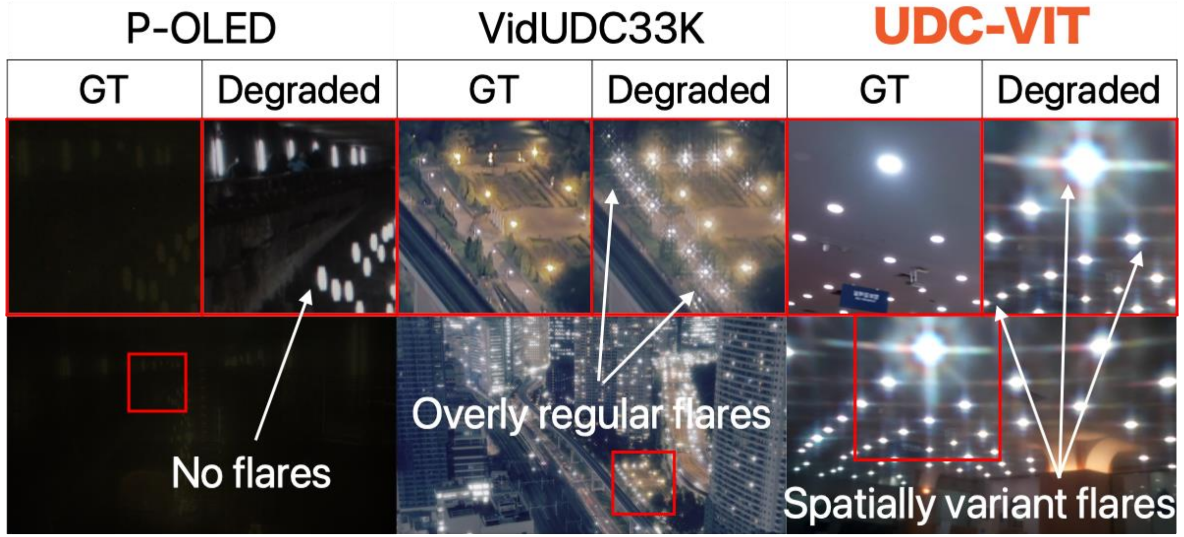
Dataset Comparison (Cont'd)

- **Noise and transmittance decrease**
  - Display panel reduces transmittance → amplifies noise under low light
  - UDC-VIT captures real transmittance loss & digital noise
  - Sensor type (e.g., QBC) also affects the noise pattern



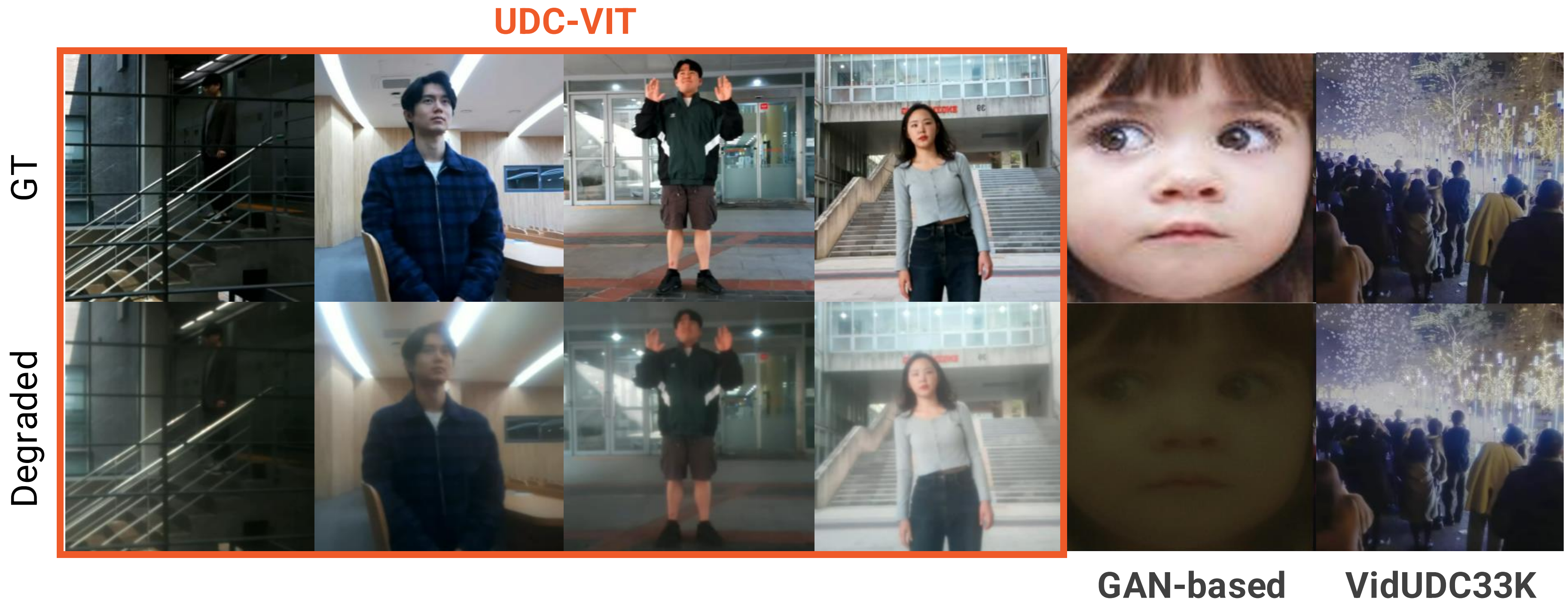
Dataset Comparison (Cont'd)

- **Variant Flare**
  - *Spatially* variant flares
  - *Temporally* variant flares
  - *Light source* variant flares



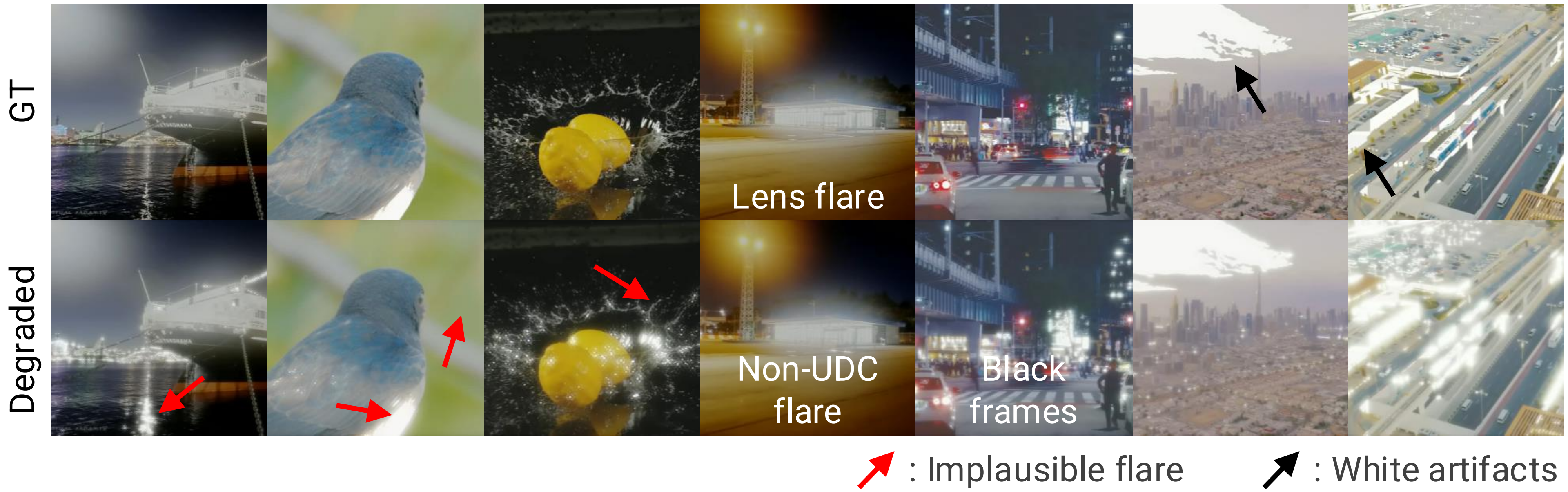
Dataset Comparison (Cont'd)

- **Face recognition**
  - Previous datasets only include limited human representations
  - UDC-VIT includes 64.6% human videos with diverse motions and angles



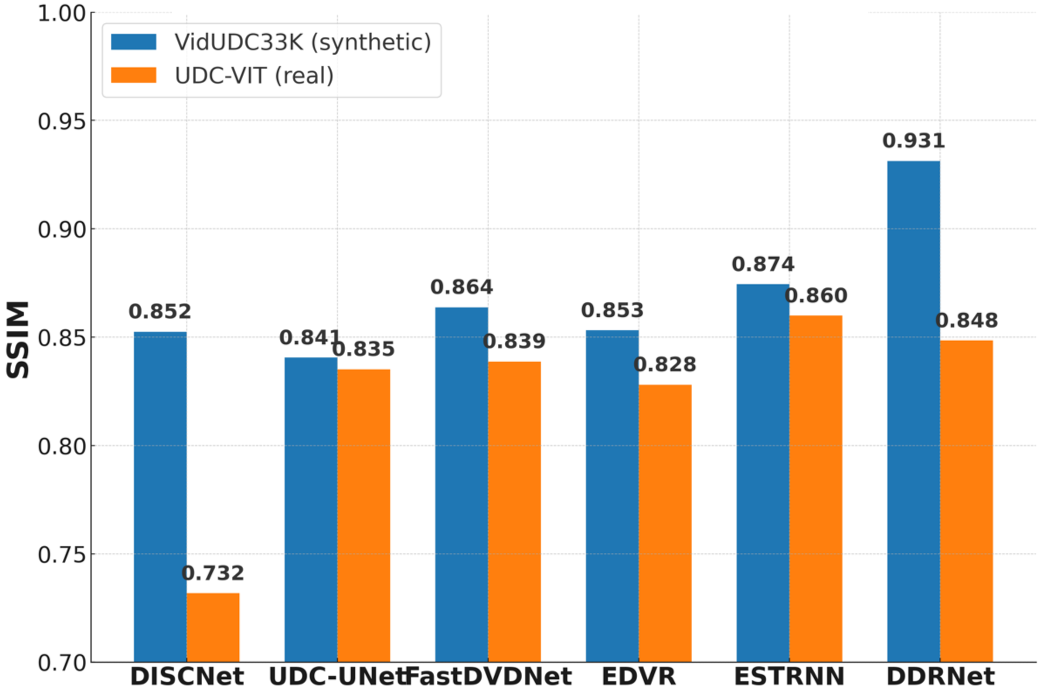
Dataset Comparison (Cont'd)

- **Strange scenes in existing datasets (e.g., VidUDC33K)**
  - Synthetic datasets often contain unrealistic artifacts
  - White/black artifacts due to invalid transformations



Effects on Restoration Models

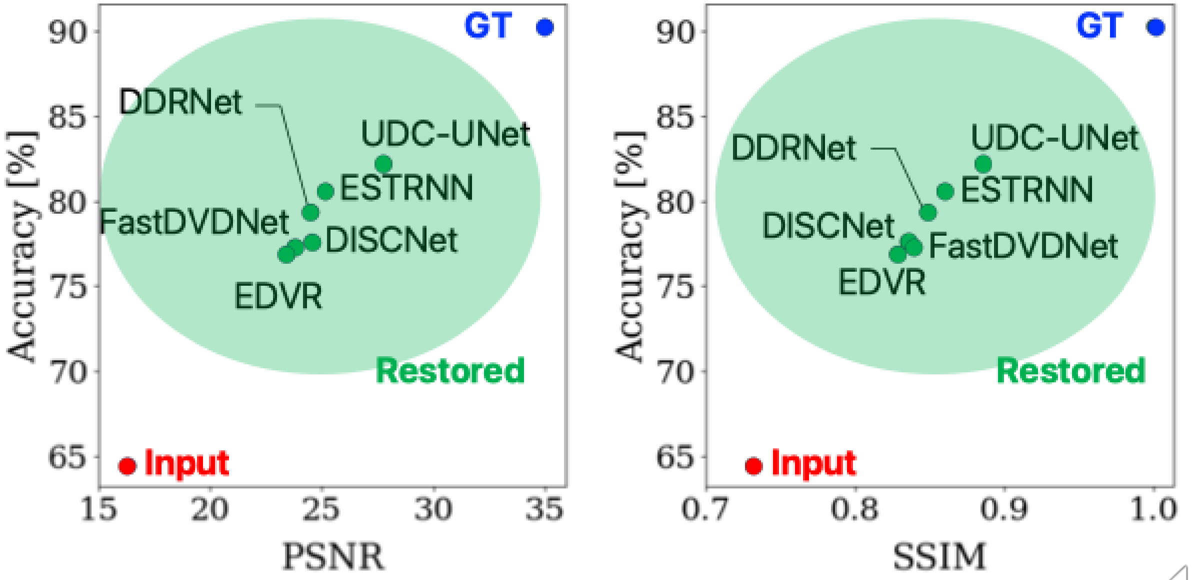
- **Key Findings:**
  - Rankings differ between synthetic (VidUDC33K) and real (UDC-VIT)
  - Residual CNNs (e.g., UDC-Unet and ESTRNN) show better frame consistency
  - DDRNet strong on synthetic, but less dominant on real data



Face Recognition (FR)

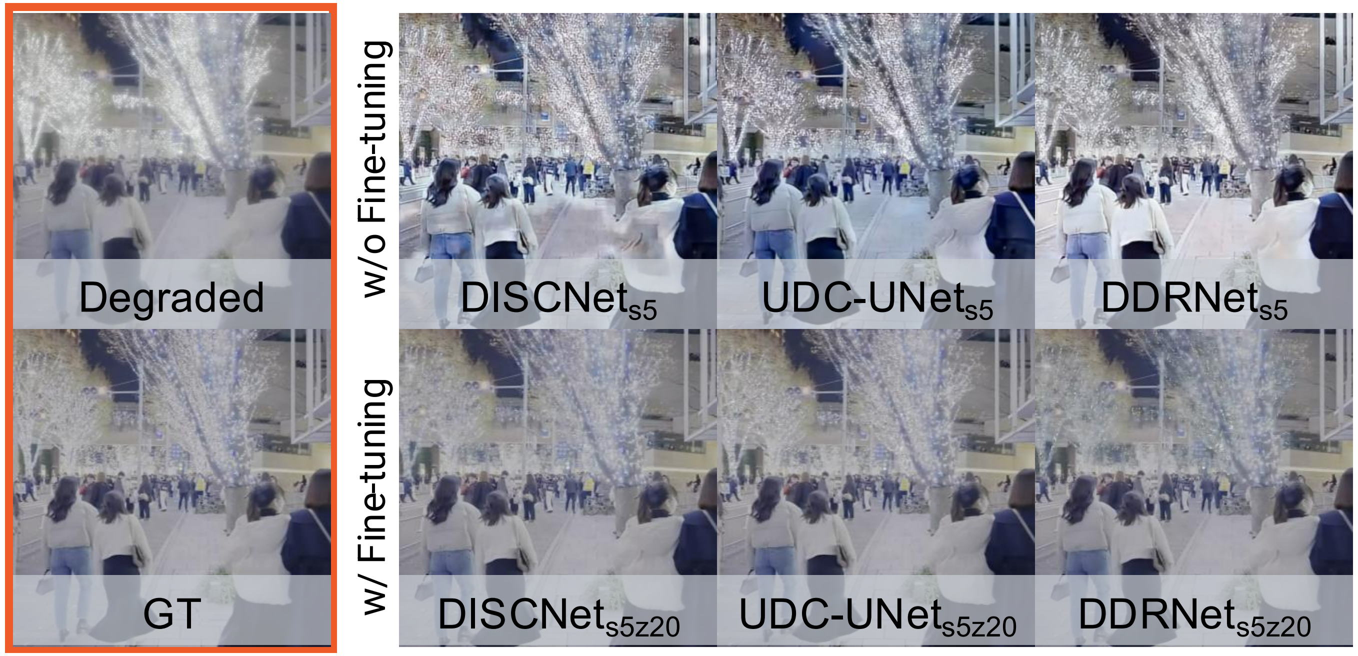
- **Setup**
  - 7 FR models (VGG-Face, Facenet, OpenFace, DeepFace, DeepID, Dlib, and ArcFace)
  - 600 balanced pairs (49.2% same / 50.8% different people)

- **Results**
  - Input (red): lowest accuracy
  - Restored (green): improved & clustered
  - GT (blue): highest accuracy



Fine-tuning

- **Fine-tuning on VidUDC33K**
  - UDC-VIT pre-training → better fine-tuning on VidUDC33K
  - Fine-tuning adapts to VidUDC33K-specific synthetic flare & degradations



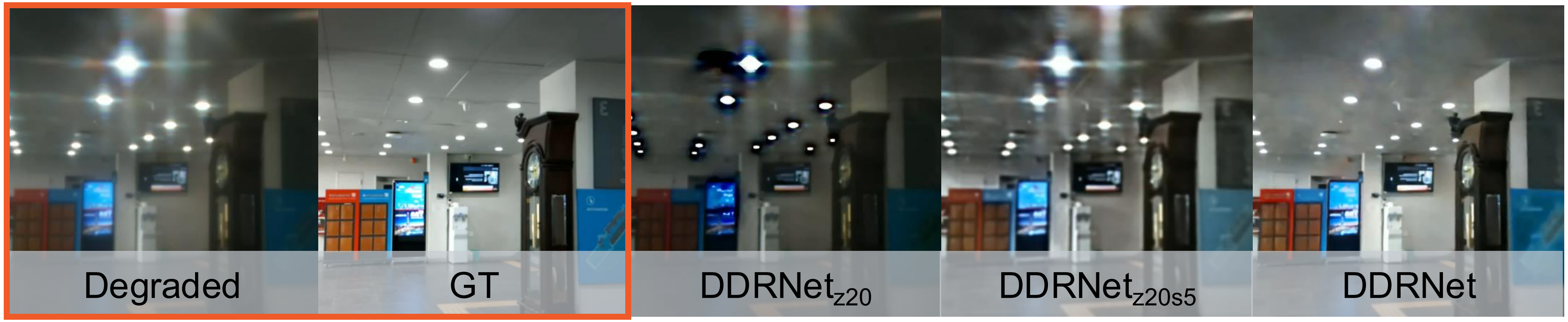
Pre-trained on UDC-VIT

Fine-tuned on VidUDC33K

Fine-tuning (Cont'd)

- **Fine-tuning on UDC-VIT**
  - Synthetic VidUDC33K pretraining → poor generalization to actual degradation in UDC-VIT
  - Pre-training on real UDC-VIT → crucial for complex degradations

UDC-VIT



	Degraded	GT	DDRNet_{z20}	DDRNet_{z20s5}	DDRNet
Pretrained	-	-	VidUDC33K	VidUDC33K	-
Fine-tuned	-	-	-	UDC-VIT	-
Trained	-	-	-	-	UDC-VIT

UDC-VIT: A Real-World Video Dataset for Under-Display Cameras

Paper **GitHub** **Project site**

For more details, scan the QR code above