Machine Learning Approaches with Real-world Priors for Imaging Factor Manipulation

Zhixiang Wang
Advisor: Shin'ichi Satoh
The University of Tokyo

UTokyo

Response to Pre-defense Comments

- The use of “shortcut learning” is confused, consider other title/story
  - We have changed the title and story of our method into “machine learning approaches with real-world priors”


Sato-sensei | Zheng-sensei

- What is the connection between imaging factor manipulation and image manipulation
  - We have added a discussion

Aizawa-sensei

- Figures are a little bit abstract
  - We have modified
- There is no video comparisons for first work (chapter 2)
  - We have added

Matsui-sensei

Everyone Takes Images/Videos

Keep Moment

Create Video

Video Conference

But, Taking Good Ones is Not Easy

Distortions Exist in Imaging Process

Low
Quality

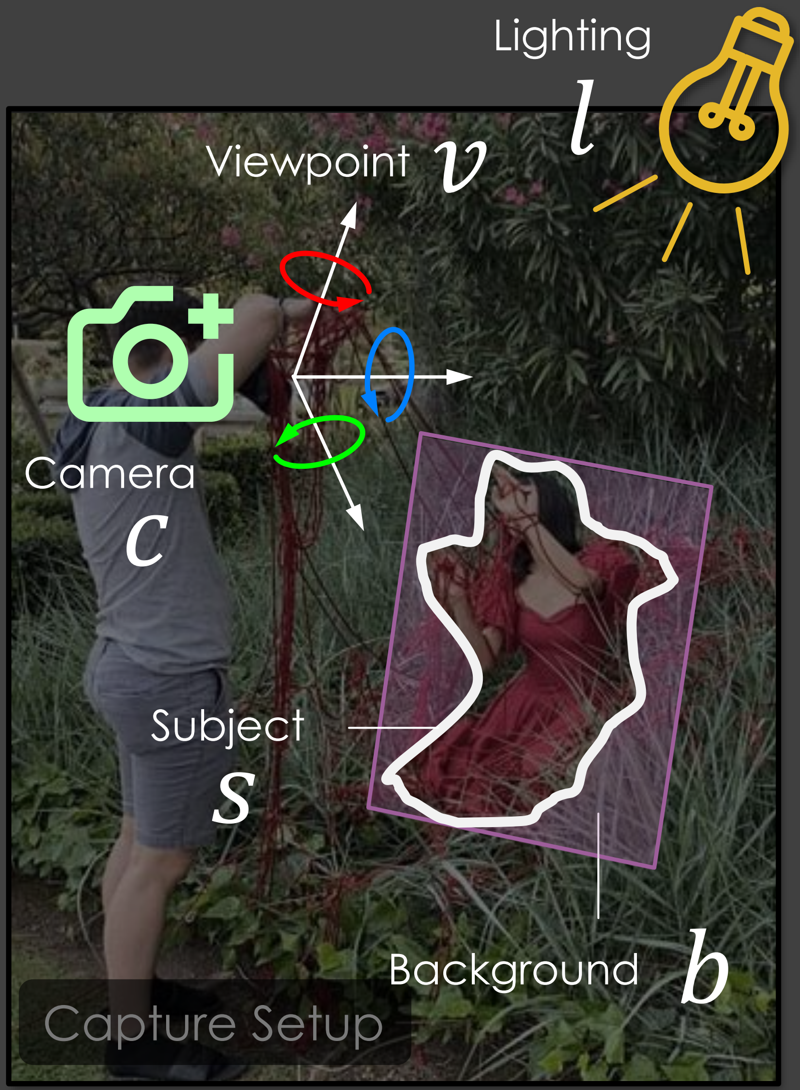
Bad
Lighting

Bad
Viewpoint

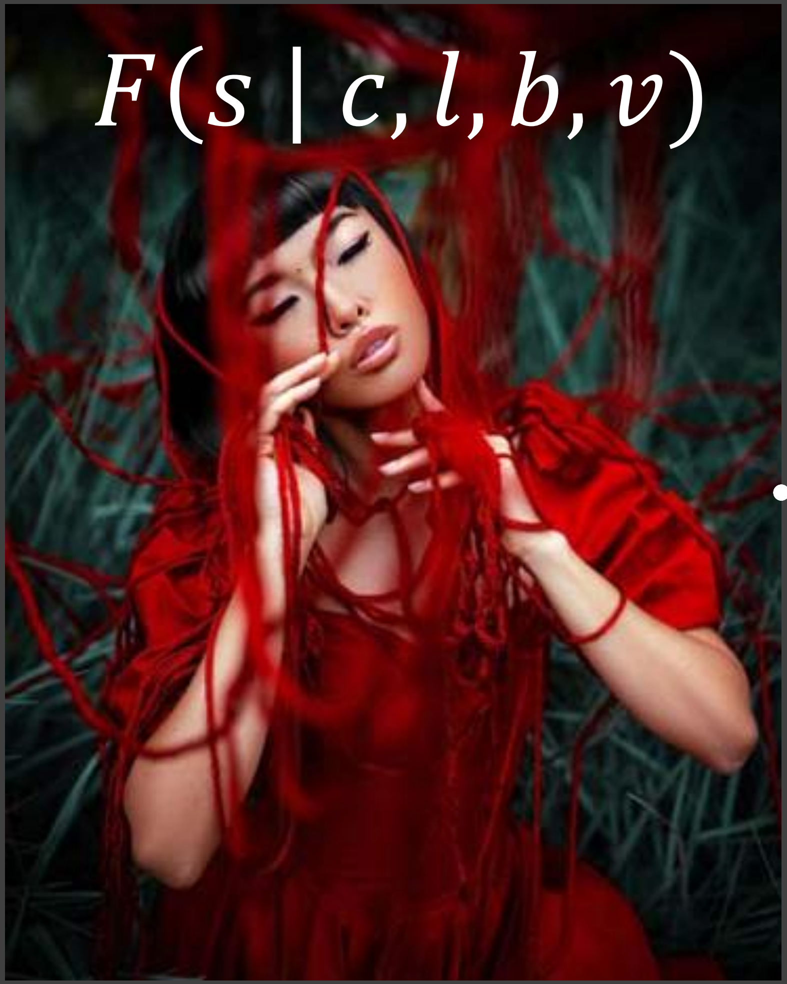
Bad
Composition

Imaging Process Has Various Factors

=

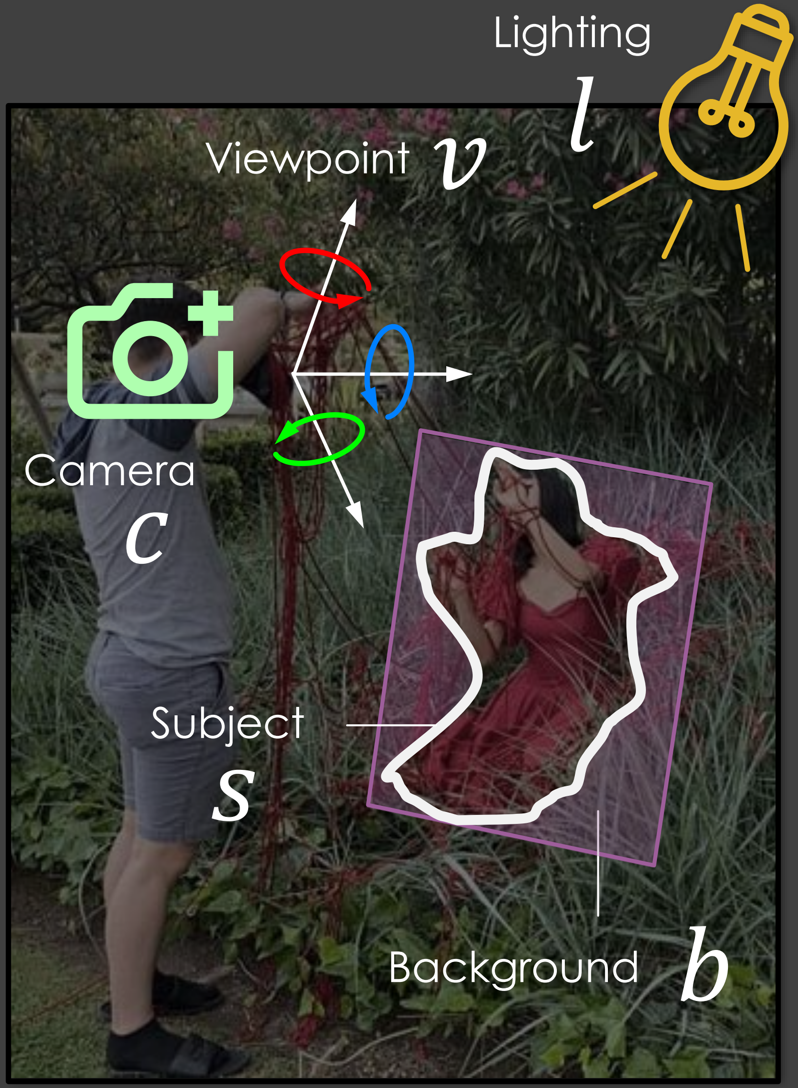


Imaging Factors Affect Images

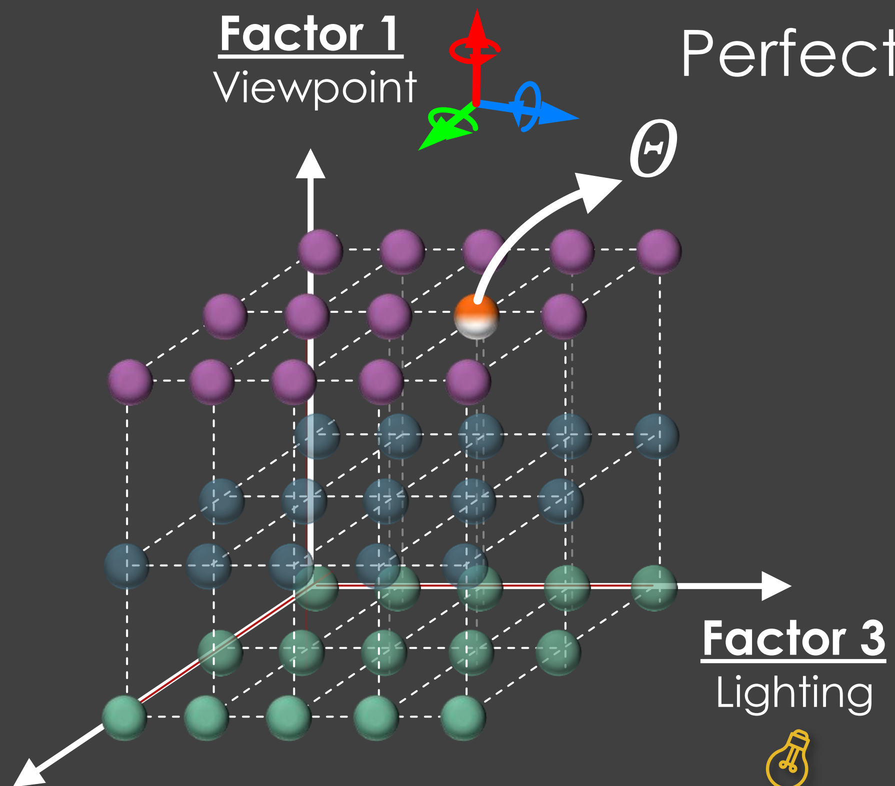


Scene Factor
- Subject: girl

Imaging Factors
- Background
- Camera
  - Sensor
  - Lens
  - Shutter
  - Aperture
- Viewpoint
- Lighting



Perfectly Setting Factors Is Not Trivial



Factor 2
Camera

Perfect Setting

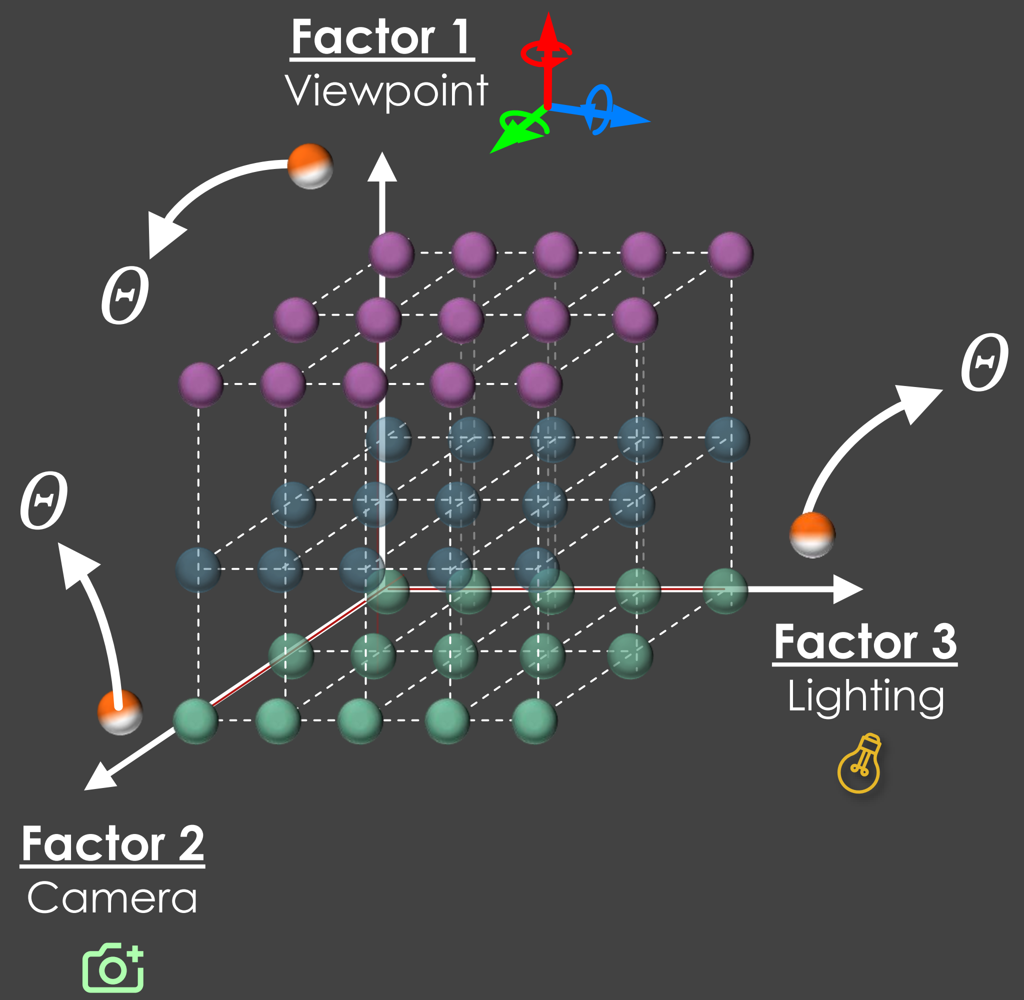
Experts

$$F(s \mid \underbrace{c, l, b, v}_{\text{Samplings } \Theta})$$

- Various factors need to control
  - Expertise and Multiple Attempts
- Some settings are hard to reach
  - Inflexible and Expensive Hardware

Hardware

Less-than-ideal Imaging Factors



$$F(s \mid \underbrace{c, l, b, v}_{\text{Samplings } \Theta})$$

→ **Distorted** and **Unsatisfactory** Captures

Device Lighting Viewpoint Background

How to Alleviate These Distortions?

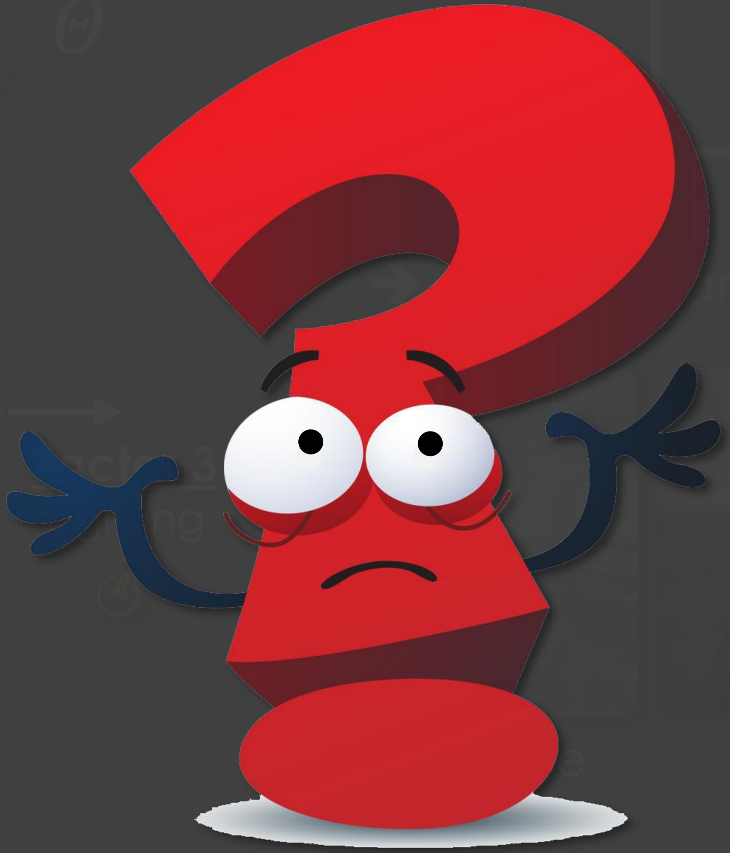
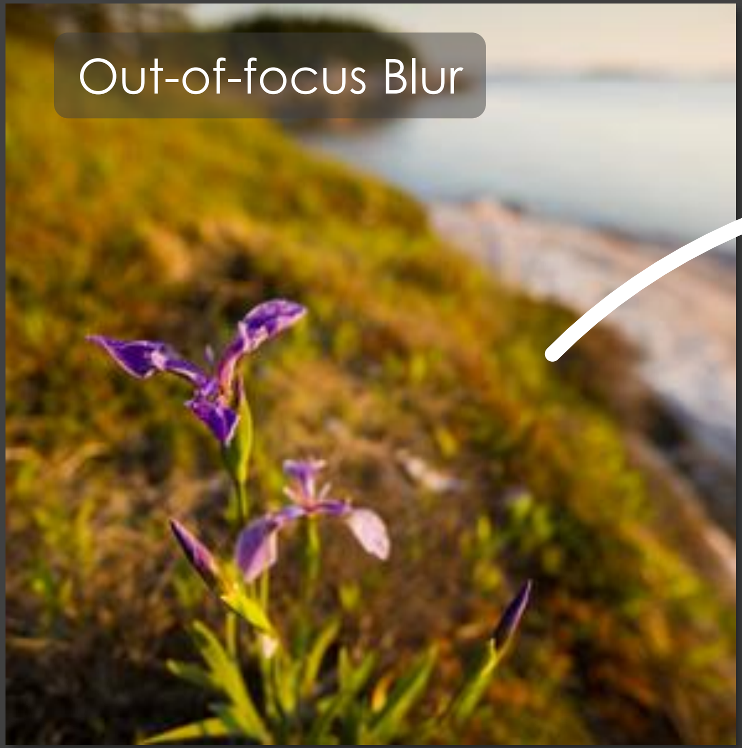
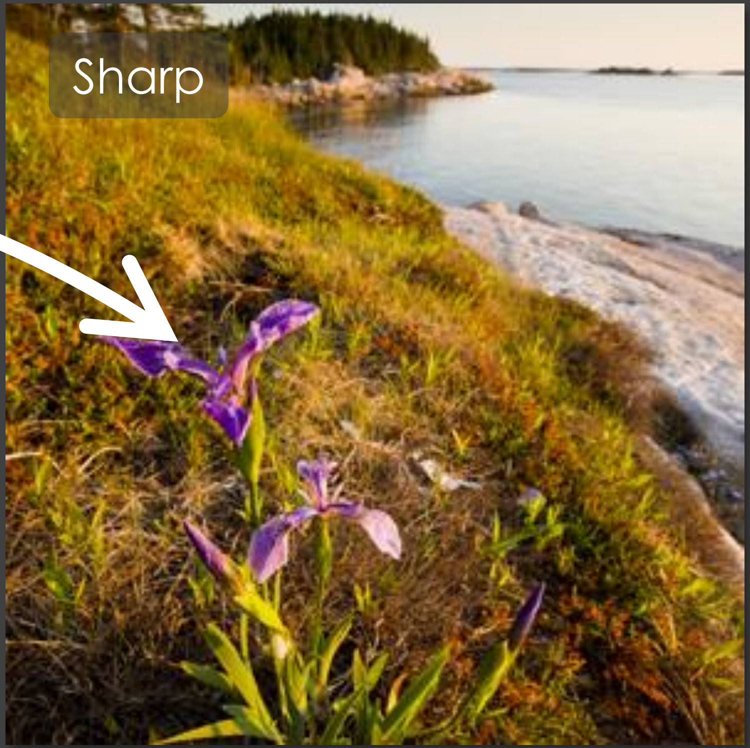


Image Manipulation



"Bad/Undesired" Image or Video

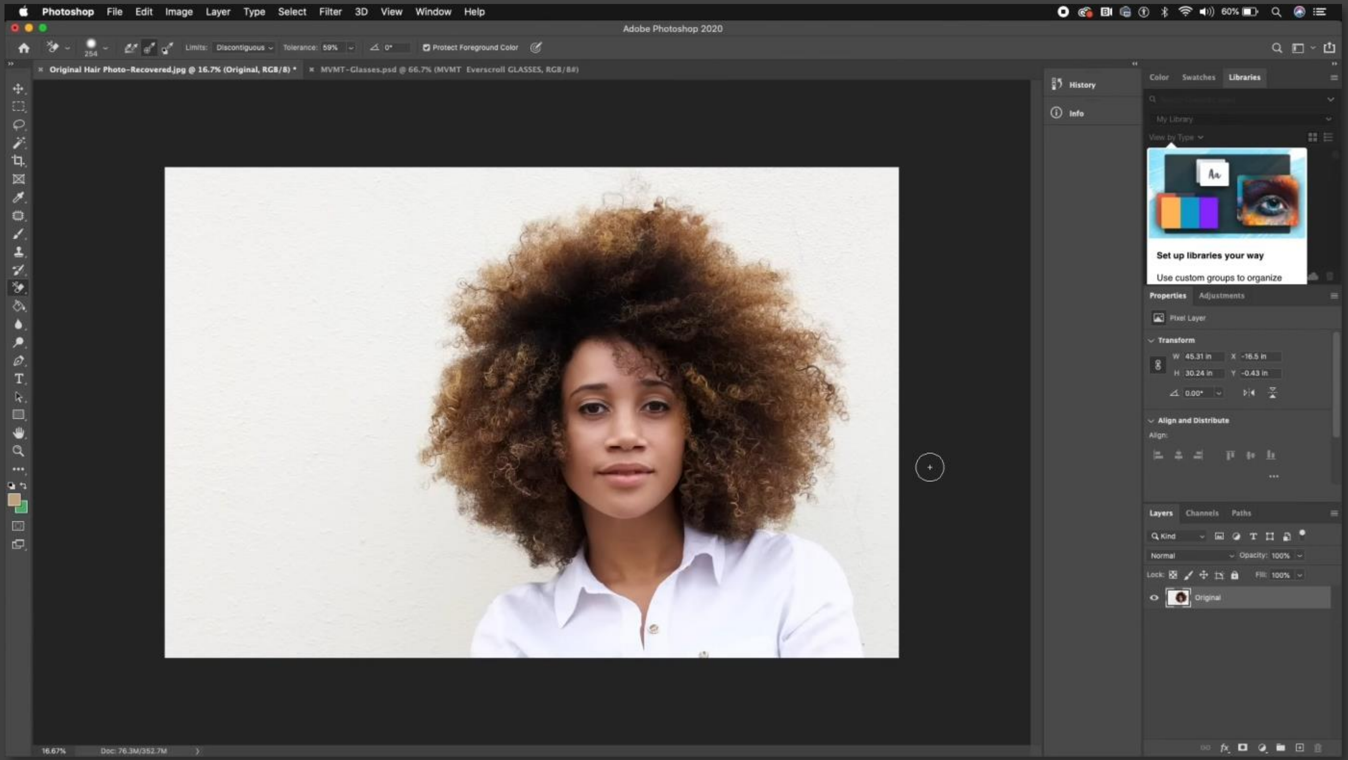
Post-Processing



"Good/Desired" Image or Video

Interactive Image Manipulation

Photoshop

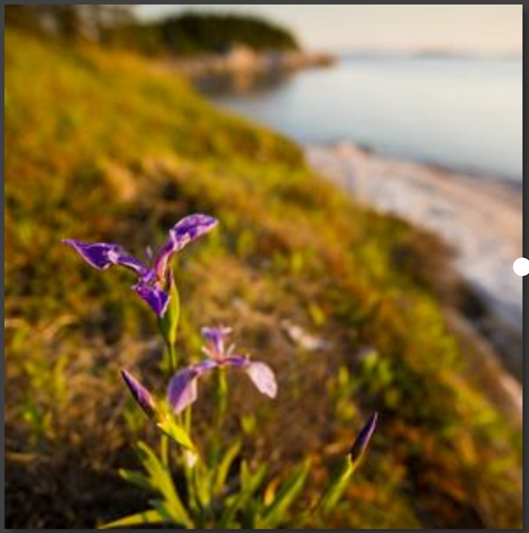


Interactive tool

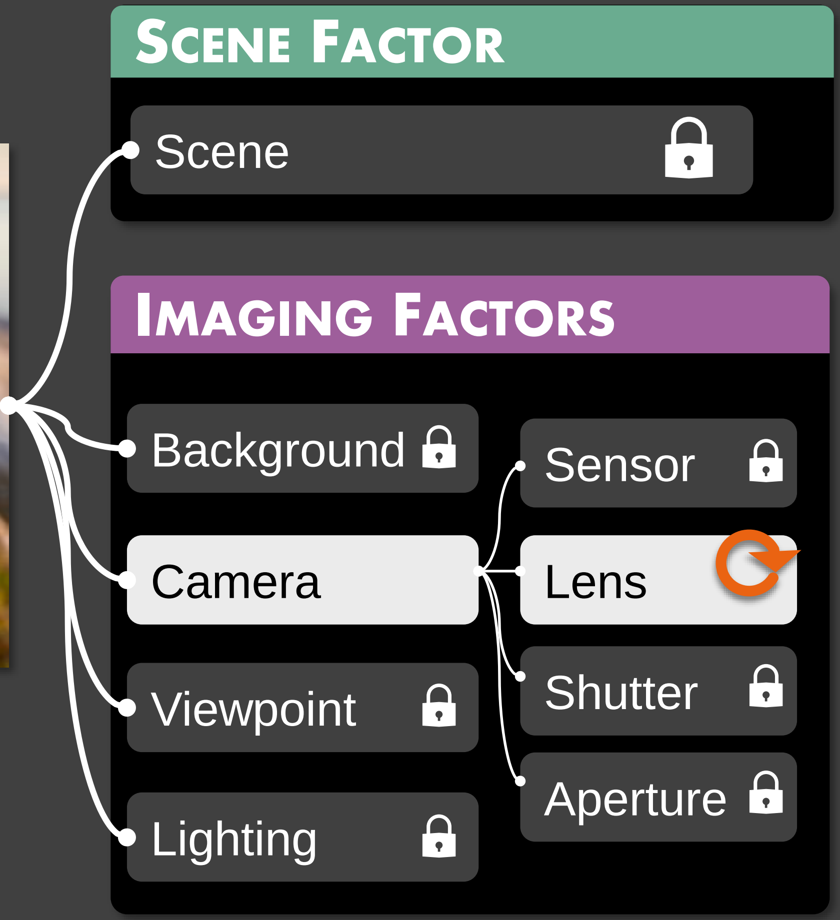
- ❑ Need efforts to master
- ❑ Not automatic
- ❑ Require sufficient observation
- ❑ Not physically plausible

Imaging Factor Manipulation

Automatically Manipulate Less-than-ideal Factors



"Bad/Undesired"



AUTOMATICALLY

"Good/Desired"

Manipulation

Imaging Factor Manipulation

Automatically Manipulate Less-than-ideal Factors

VS

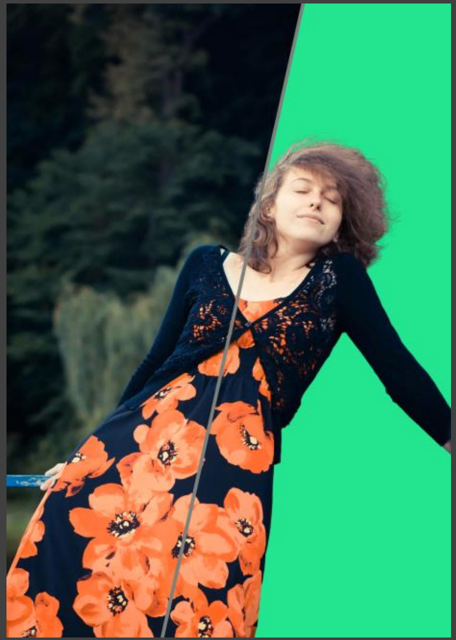
Interactive tool

- ~~Need efforts to master~~
- ~~Not automatic~~
- ~~Require sufficient observation~~
- ~~Not physically plausible~~

Our techniques

- No effort to master
- Automatic
- No sufficient observation, OK
- Physically plausible

Three Manipulations in Our Thesis



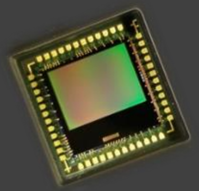
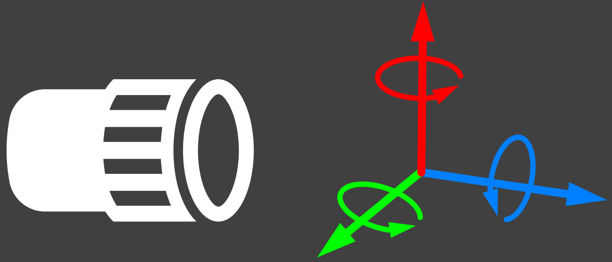


Image Sensor

Chapter 2



Viewpoint and Lens

Chapter 3

Background

Chapter 4

Relation to Automatic Image Manipulation

Chen et al, CVPR'2018

Brooks et al, CVPR'2023

Image Manipulation

Image Enhancement

Transformation

Filtering

Image Editing

Imaging Factor Manipulation

Image Restoration

Image Morphing

Saharia et al, TPAMI'22

Similarity

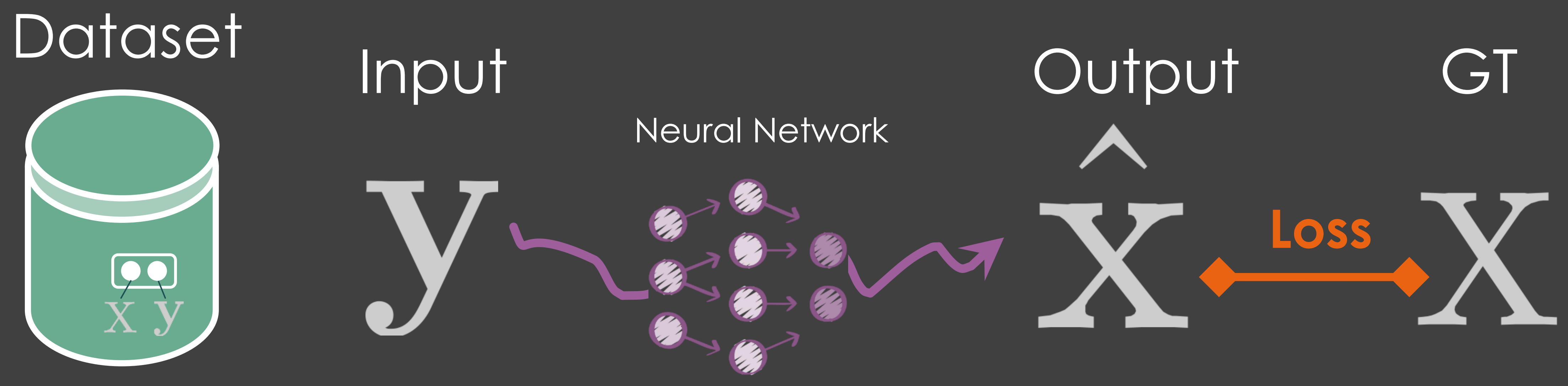
➔ Has overlap with image editing and restoration

Difference

➔ Recover the scene from imperfect observations

➔ Edit imaging factors rather than others (e.g., scene/content)

The Most Popular Approach



End-to-end **Fully Supervised** Learning

Wang et al, Deep Learning for Image Super-Resolution: A Survey, TPAMI'19
Li et al, Low-Light Image and Video Enhancement Using Deep Learning: A Survey, TPAMI'22
Zhang et al, Deep Image Deblurring: A Survey, IJCV'22
Wang et al, Deep Learning for HDR Imaging: State-of-the-Art and Future Trends, TPAMI'22

⚠ Problems

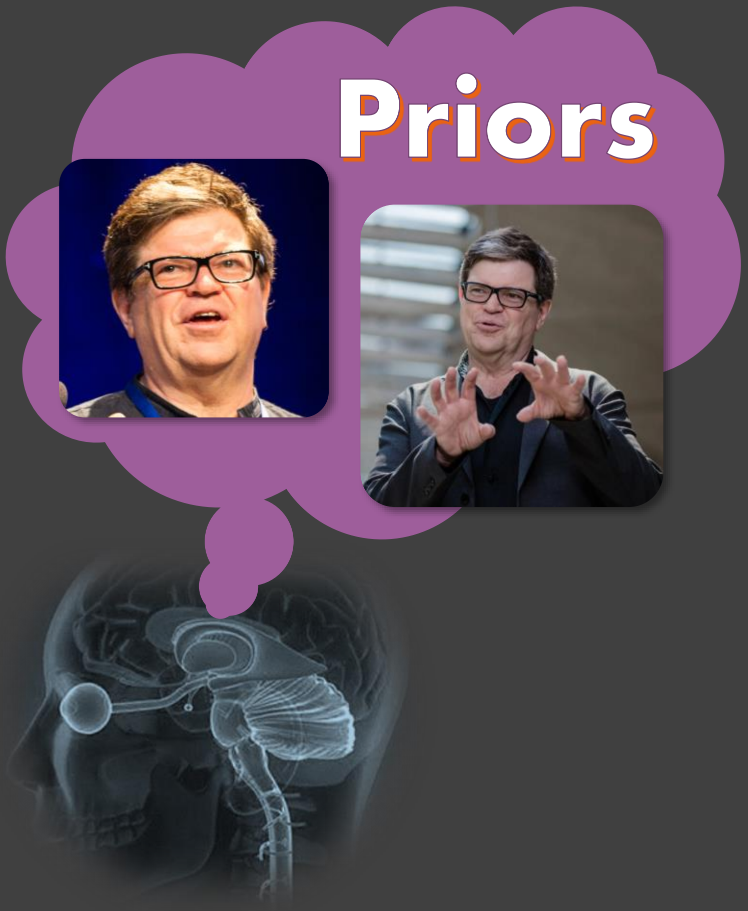
- **Learning structure:** Ineffective for seriously ill-posed problems
- **Required dataset:** Tedious data with perfect label

Inspiration from Human Perception

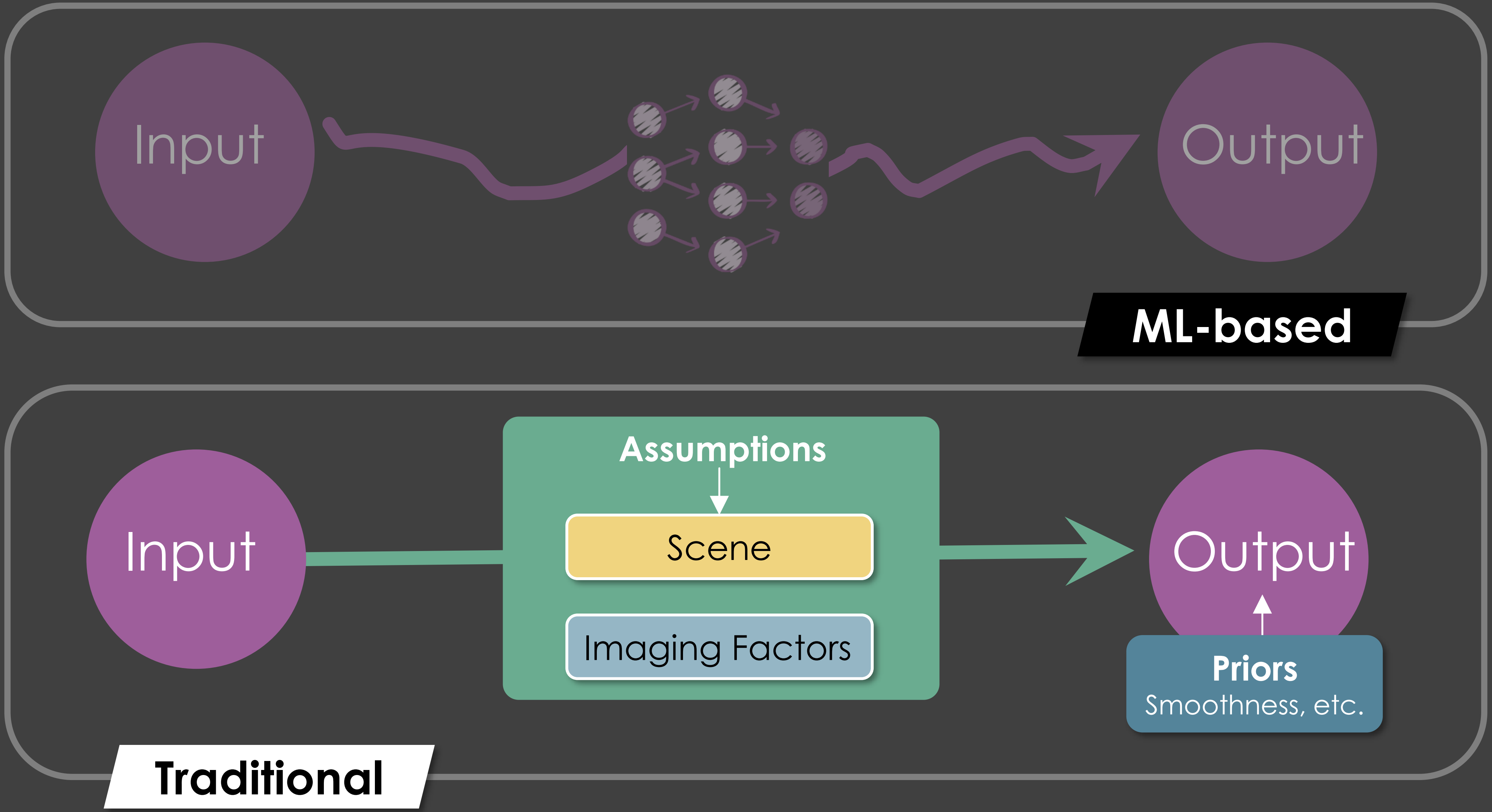
Can you **see** this image clearly?

Inspiration from Human Perception

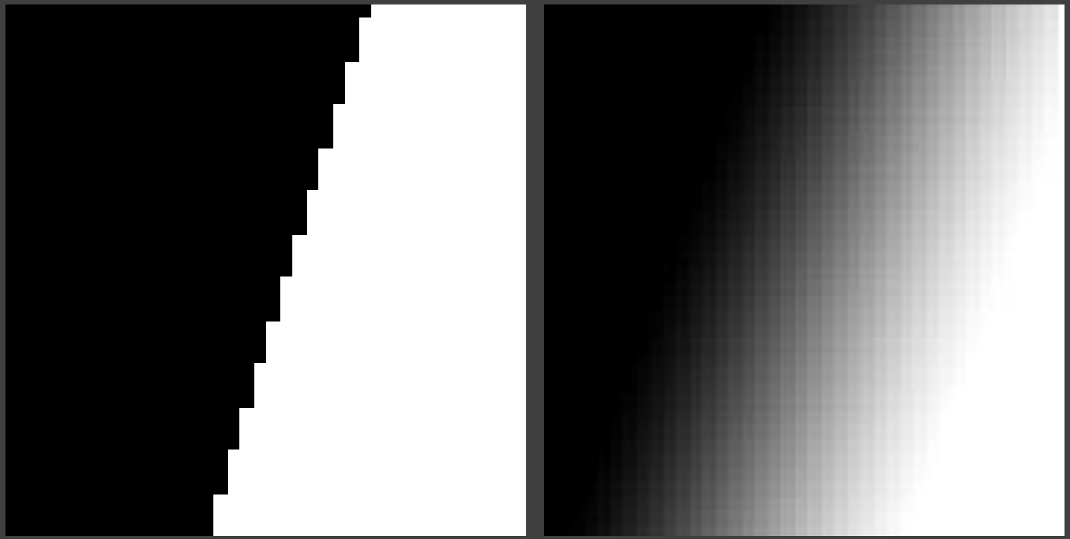
But if we know it's a photo of **Yann LeCun**



Traditional Approaches



Handcrafted Image Priors

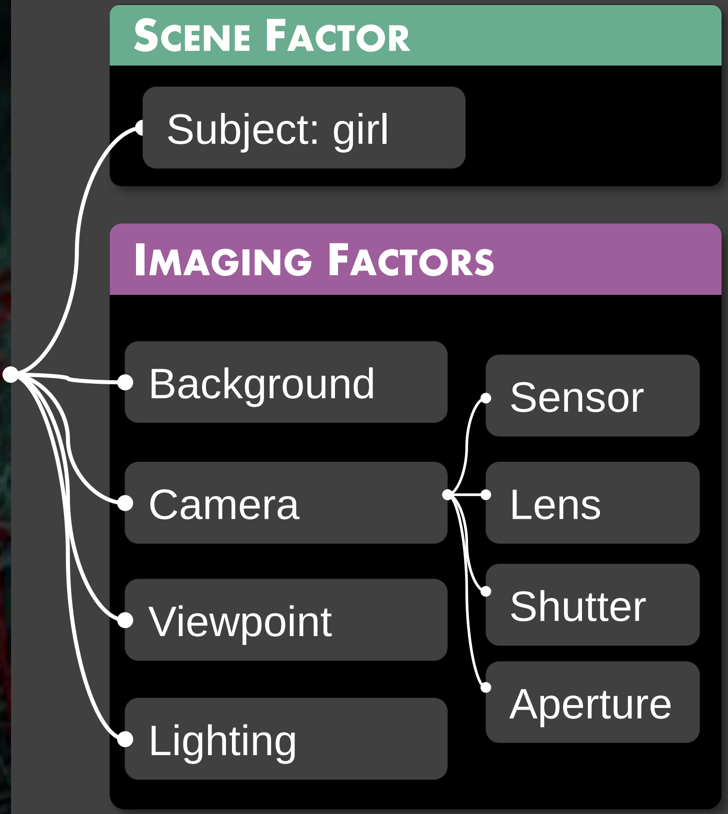


Smoothness Prior

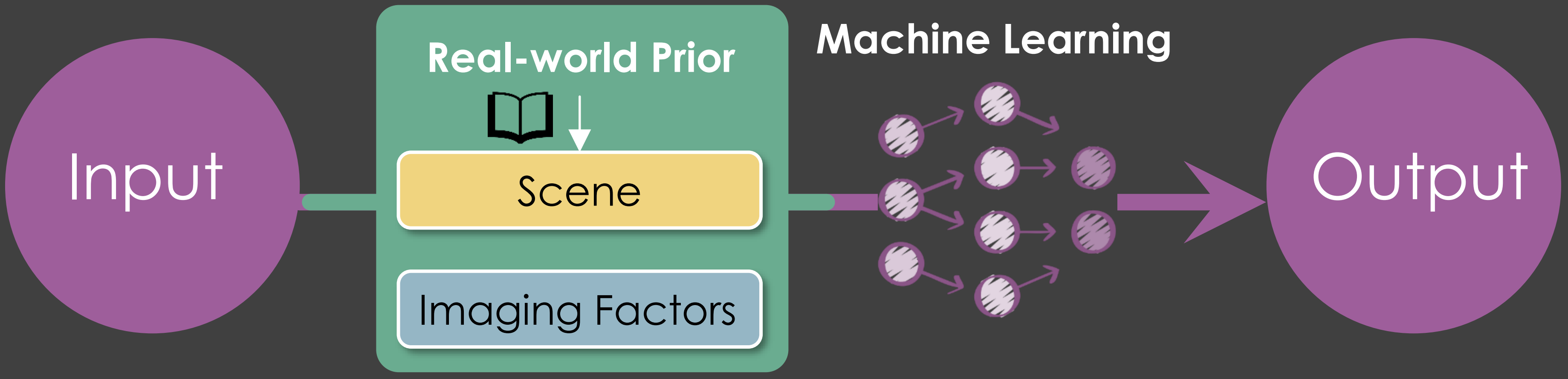
Dark Channel Prior
He, CVPR’ 09

Images Are More Than Just Pixels

Images Are More Than Just Pixels

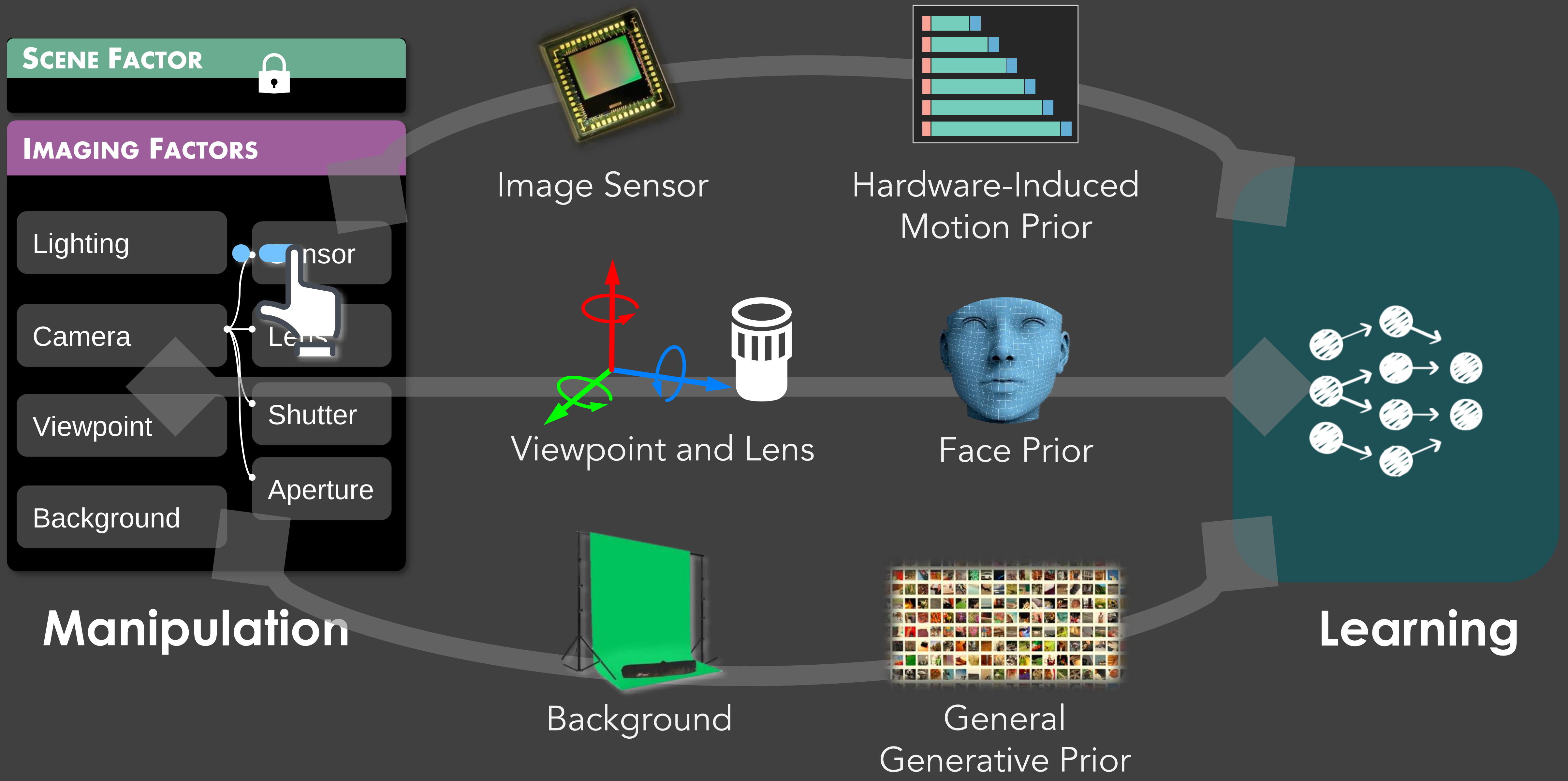


Machine Learning Approaches with Real-world Prior



Real-world Priors vs Image Prior
- Priors of scenes
- Priors that are not handcrafted

Benefits
- Change the learning structures
- Solve severe ill-posed problems
- Reduce the requirement to data


Scene Factor
Imaging Factors
Lighting
Camera
Viewpoint
Background
Shutter
Aperture
Image Sensor
Hardware-Induced Motion Prior
Viewpoint and Lens
Face Prior
Background
General Generative Prior
Manipulation
Learning

Social Impact 1 – Machine Vision with Lower Costs, Enhanced Perception

Autonomous Driving

Robot

Social Impact 2 – Bringing Cinematic Filming Capabilities to Everyone's Phone

Image credit: Thai Nguyen @Unsplash and DALLE2

Social Impact 3 – Empowering Memories

"My Father passed away yesterday, please blur/remove the background"

"My eyes are always distorted and warped in my phone selfies …"

"Fix perspective distortion please"

"I love this photo of my girls...hate the background. Will tip $20 to the best one."

PsR r/PhotoshopRequest

Social Impact 4 – Psychological

"People tend to view the inevitably warped stance of self-taken (i.e., hand-held) self-portraits as a new universal standard in appearance"

— Ward et al, 2018

Distorted

Undistorted

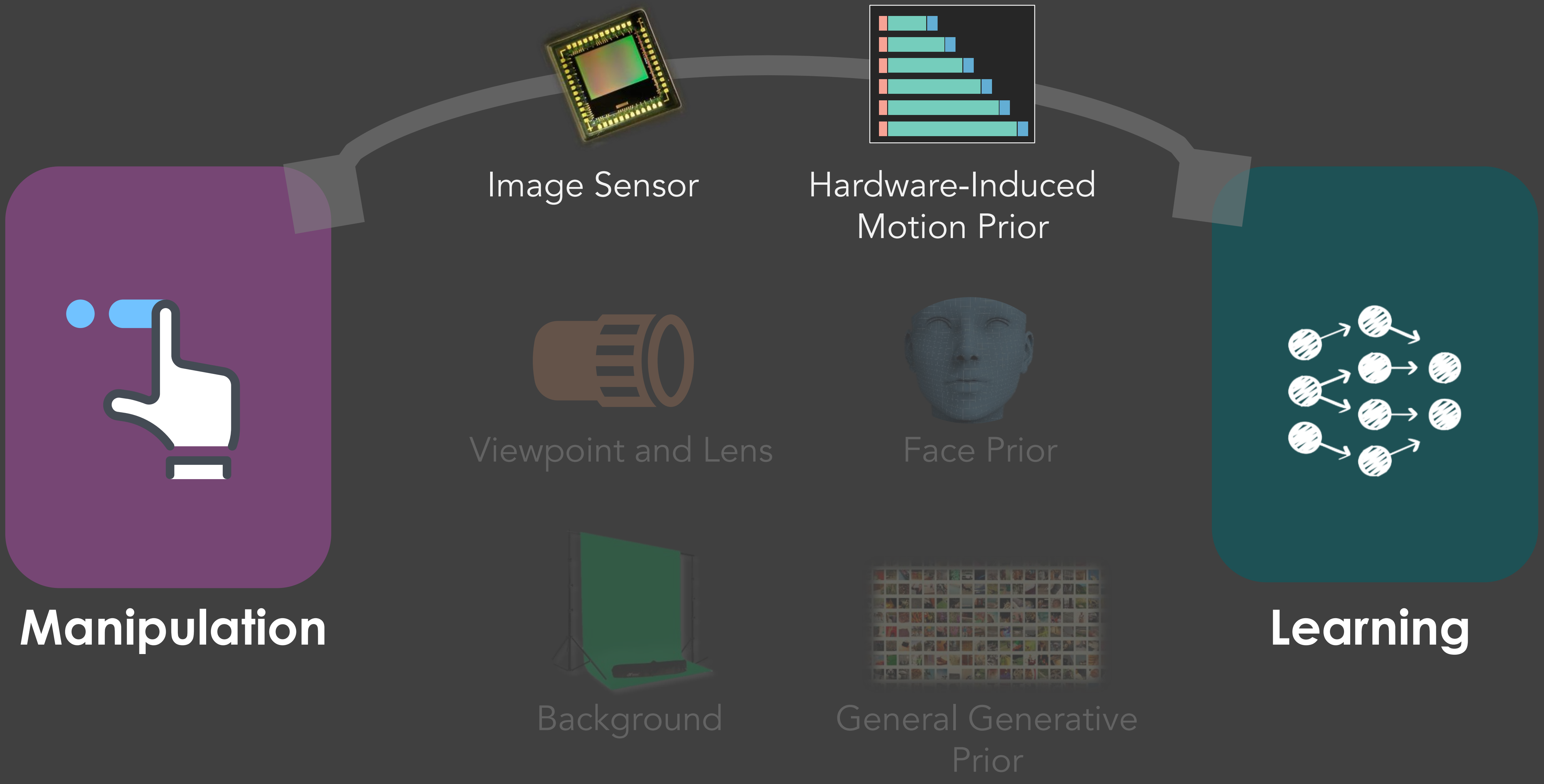
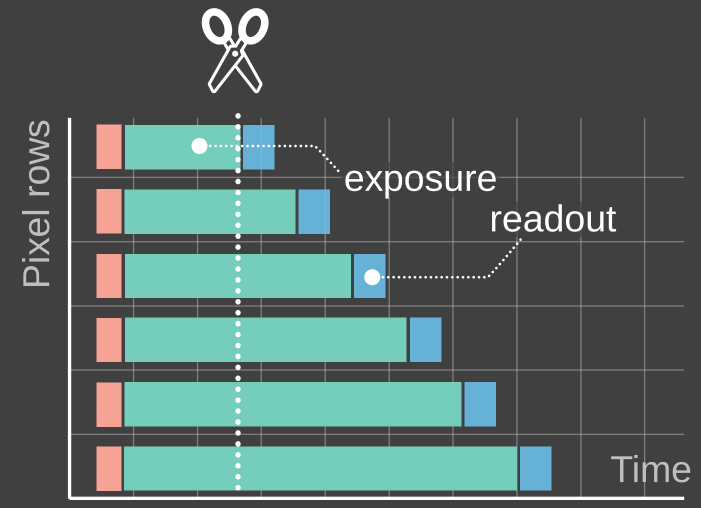


Image Sensor
Hardware-Induced Motion Prior
Viewpoint and Lens
Face Prior
Background
General Generative Prior
Manipulation
Learning

Image Sensor



Neural Global Shutter

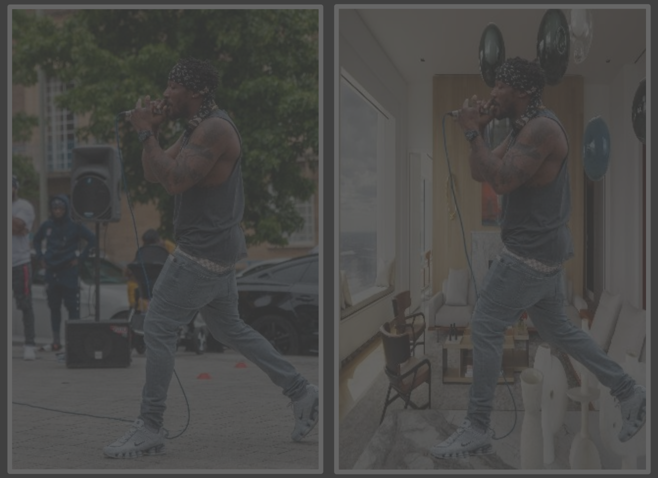
Wang et al, CVPR 2022

Viewpoint + Lens

Portrait Distortion Correction

Wang et al, IJCV 2024

Background



Matting by Generation

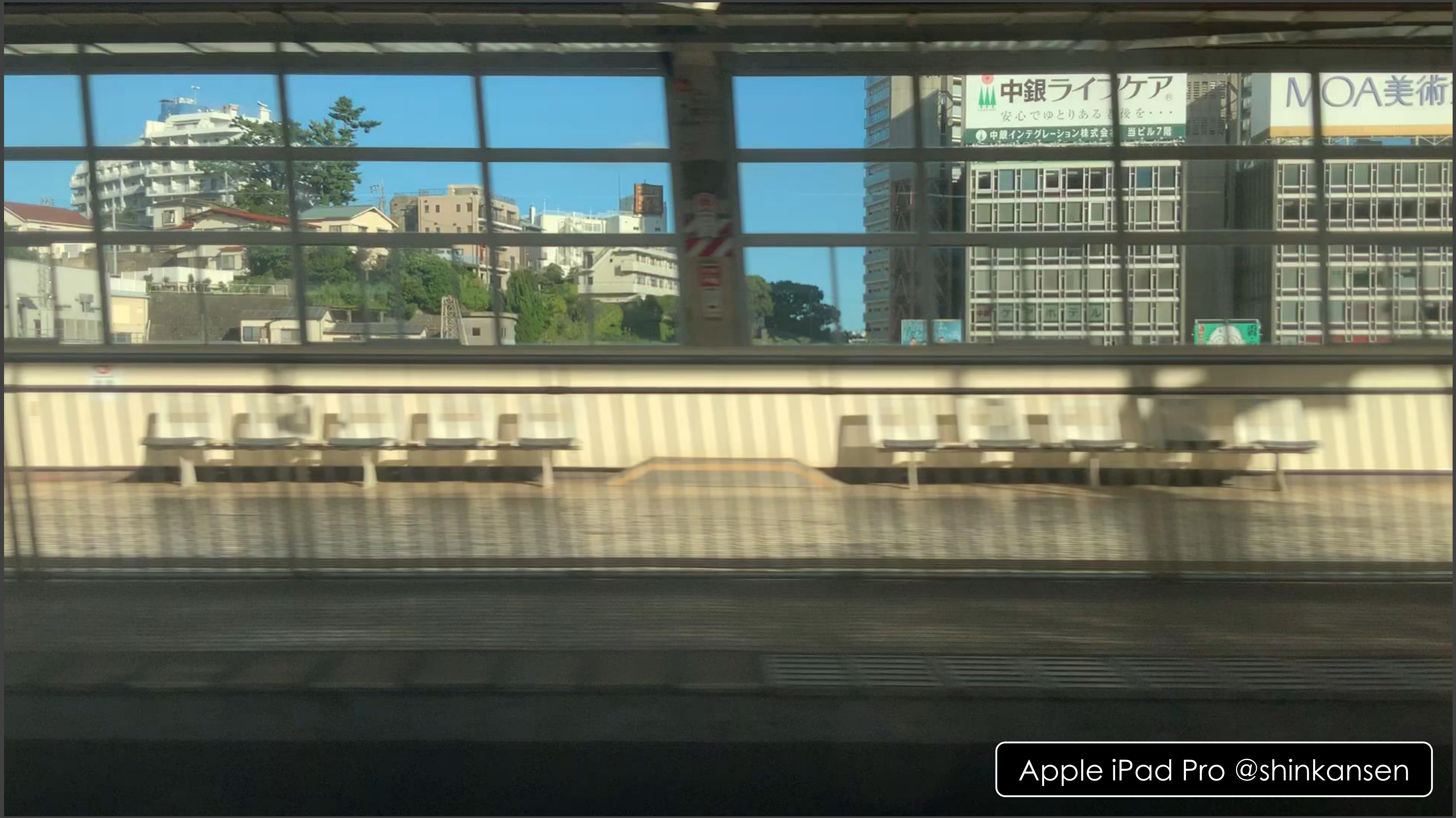
Wang et al, SIGGRAPH 2024

New Learning Structures

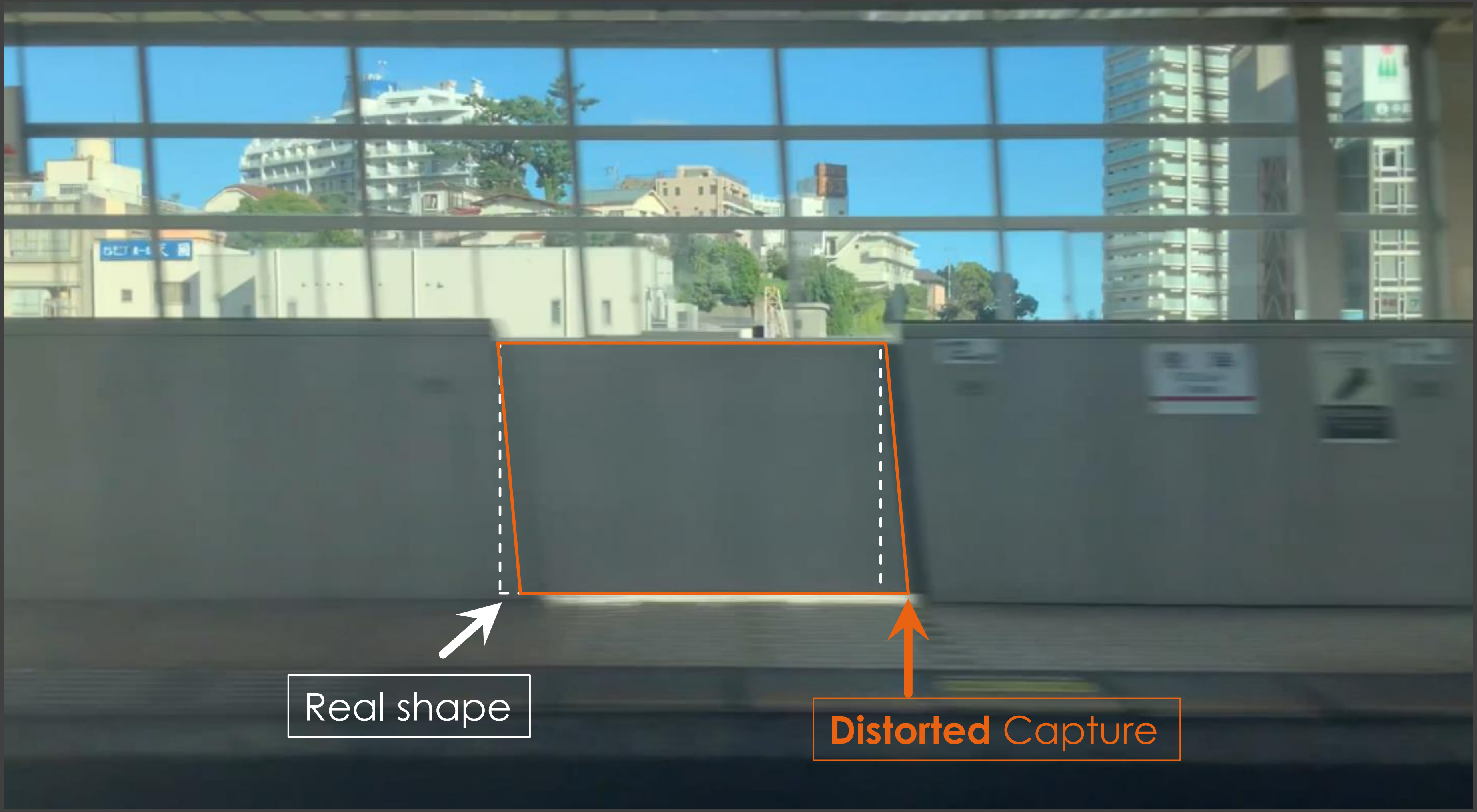
Warping → Deblurring

…	
CVPR'20	DSUR
CVPR'21	JCD
…	

Ours


Apple iPad Pro @shinkansen

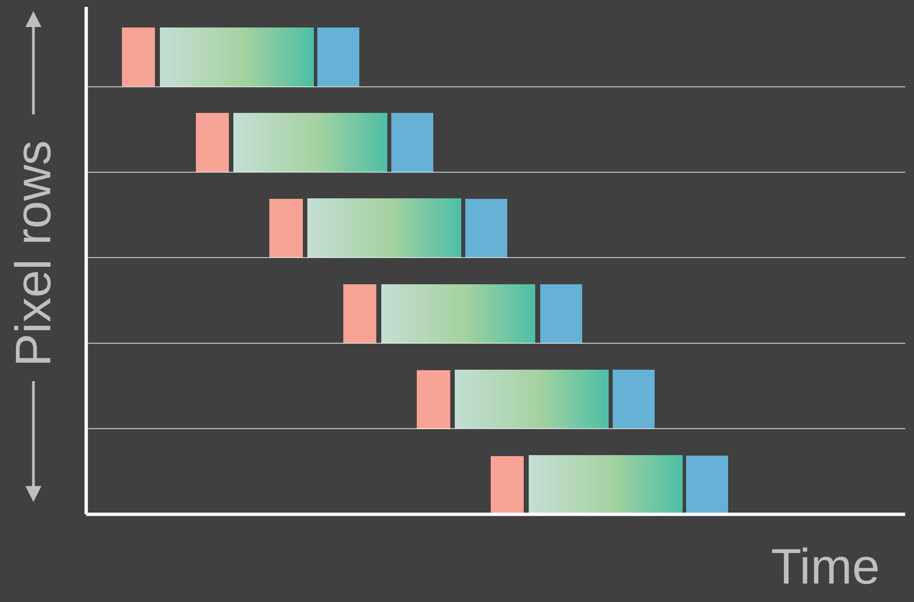
Strong Geometric Distortion

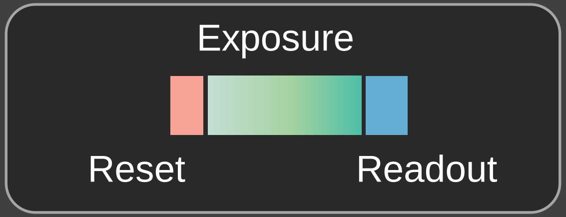


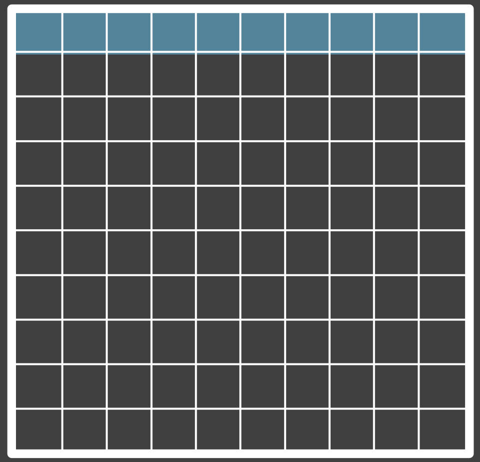
Rolling Shutter Image Sensor

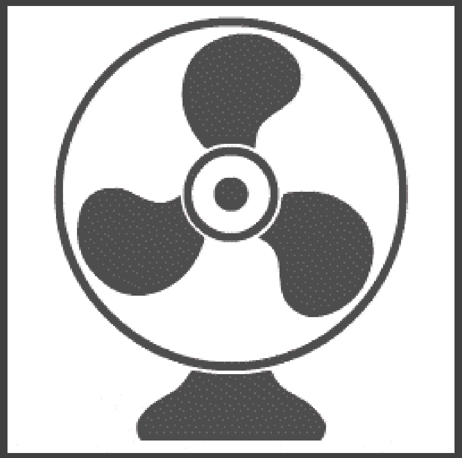
Row-by-row Exposure

- Pros: Cheap, Fast, and Widely Used
- Cons: Distortion under Fast Movement









Annoying Rolling Shutter Distortion

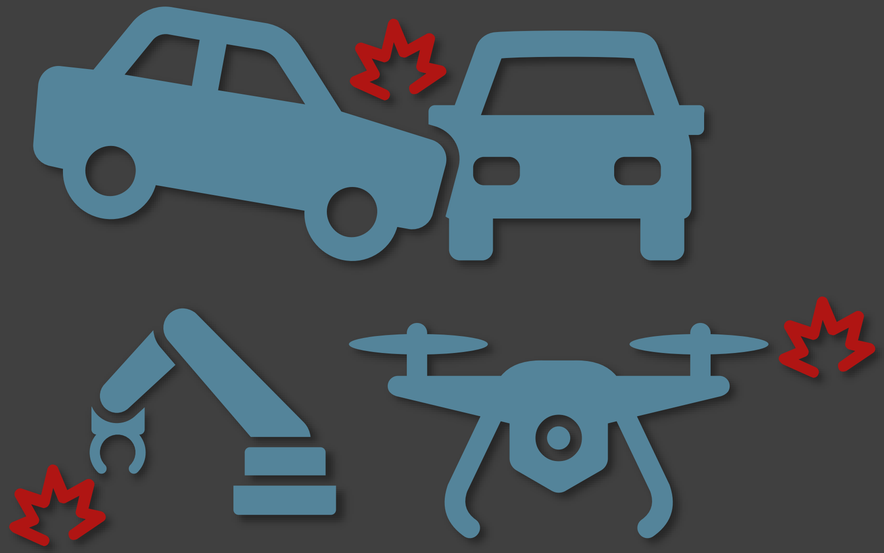
Degrade Image Fidelity

Risk Real-world Applications

Real

Distorted Capture

Autonomous Driving

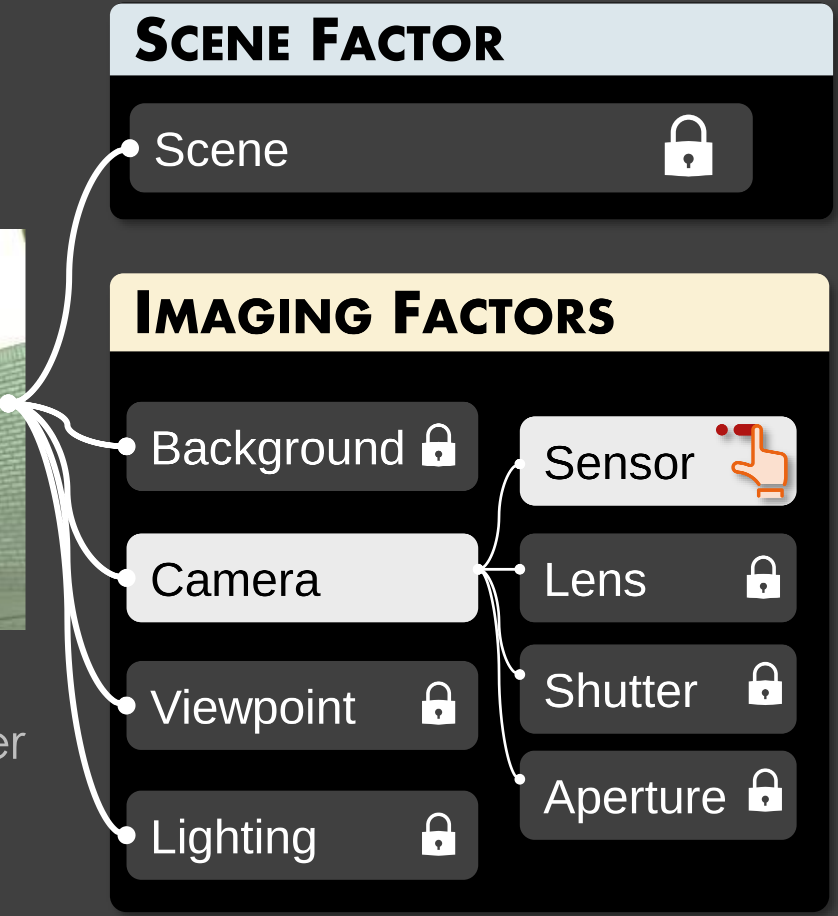


Robot

UAV

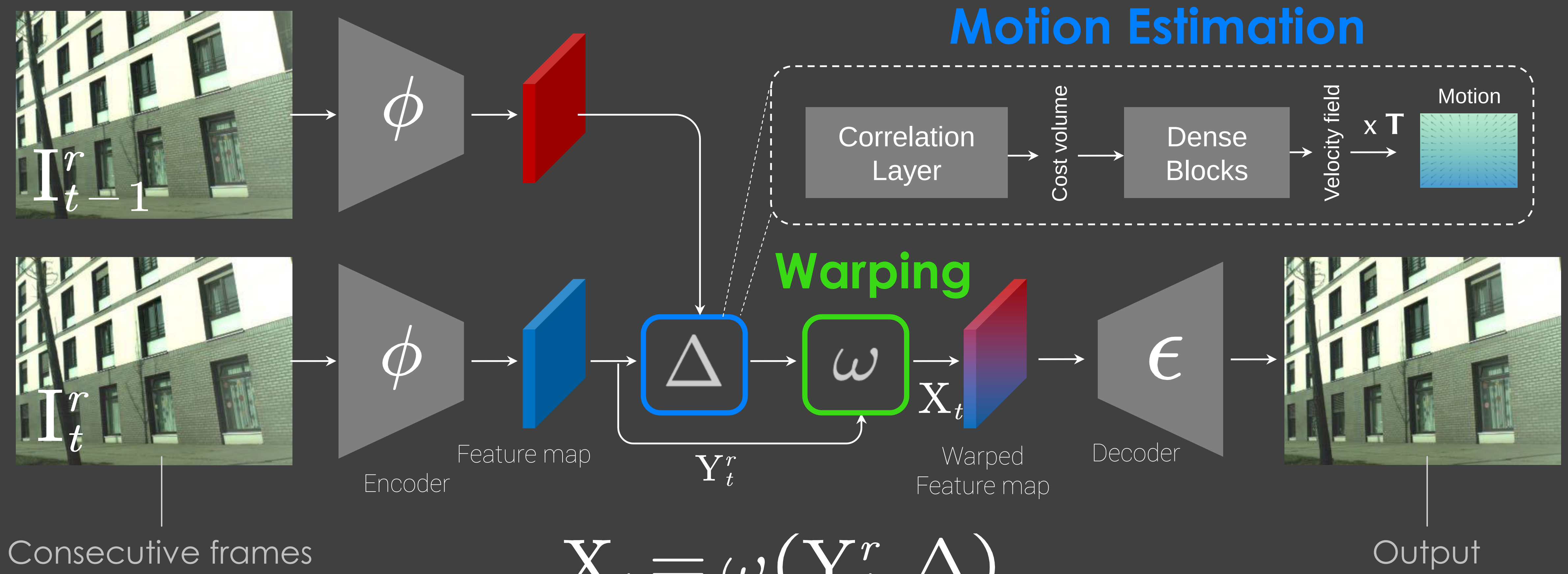
Manipulate Camera Sensor

Dynamic Scene
Captured by Rolling Shutter



Dynamic Scene
Captured by Global Shutter

Existing Methods – Warping-based



$$\mathrm{X}_t = \omega(\mathrm{Y}_t^r, \Delta)$$

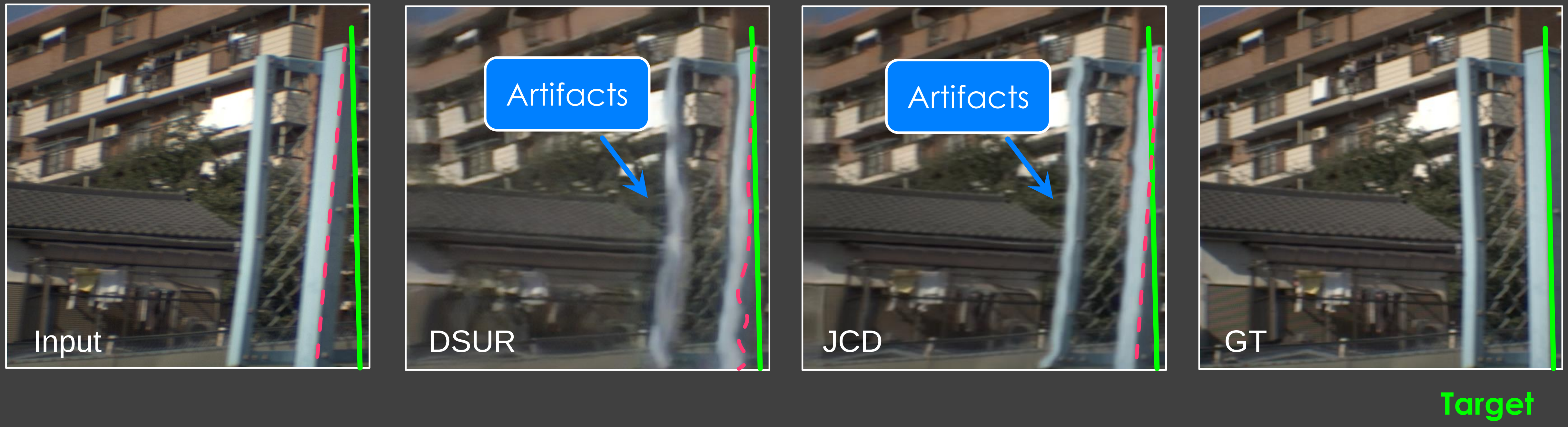
output feature

input feature

Motion field

Liu et al, DSUR, CVPR’20
Zhong et al, JCD, CVPR’21

Limitations of Existing Methods



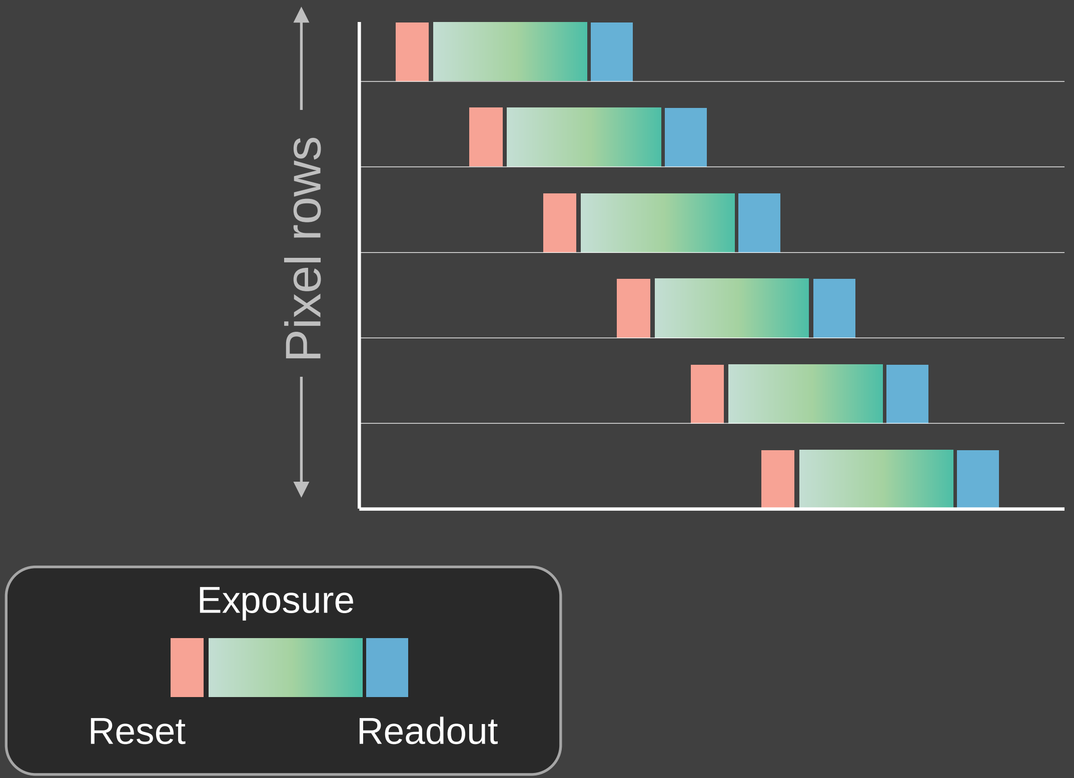
Fail to **large** and **complex** motion

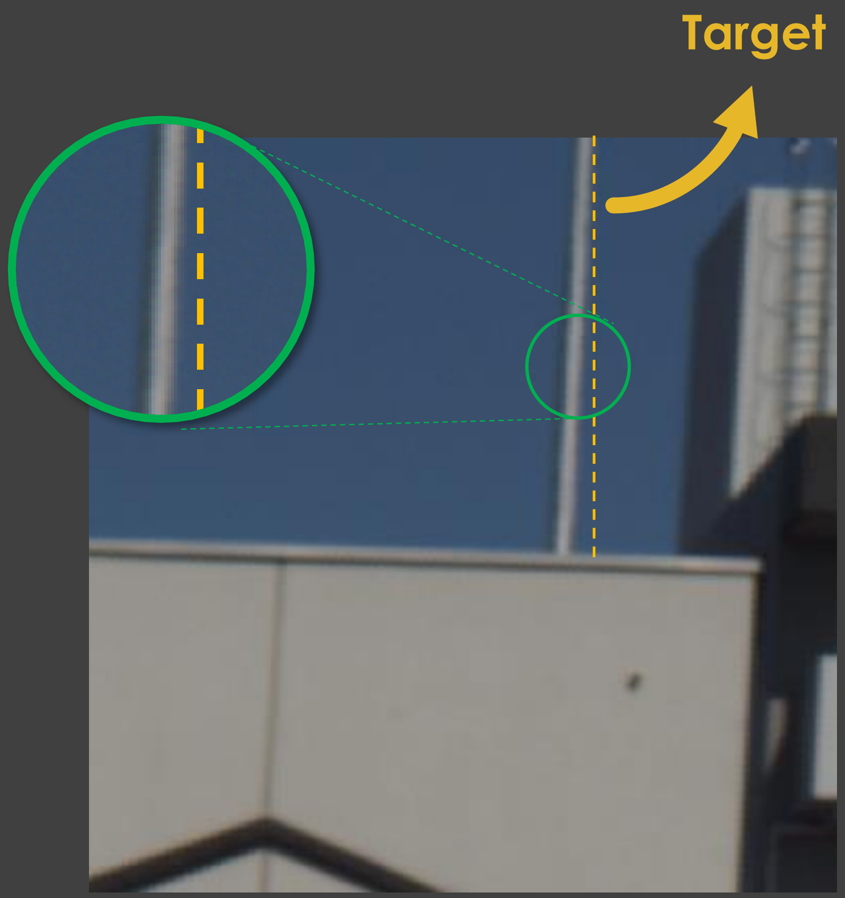
Bring **artifacts**

Liu et al, DSUR, CVPR’20
Zhong et al, JCD, CVPR’21

No Prior

rolling shutter



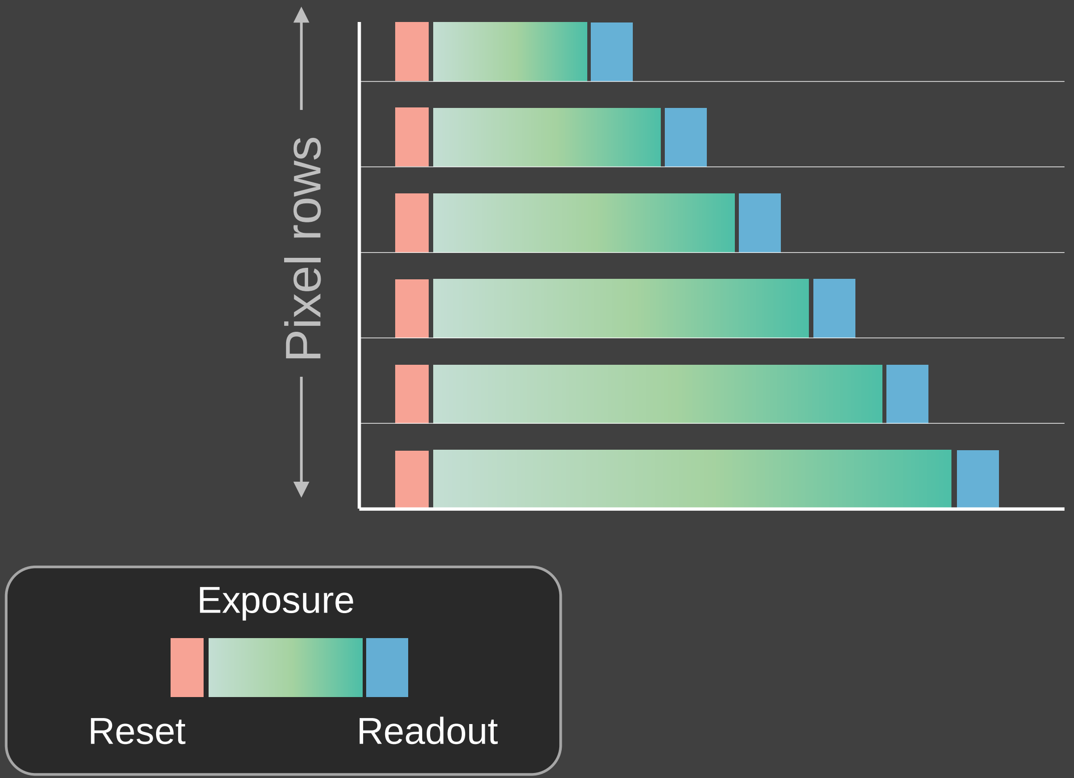


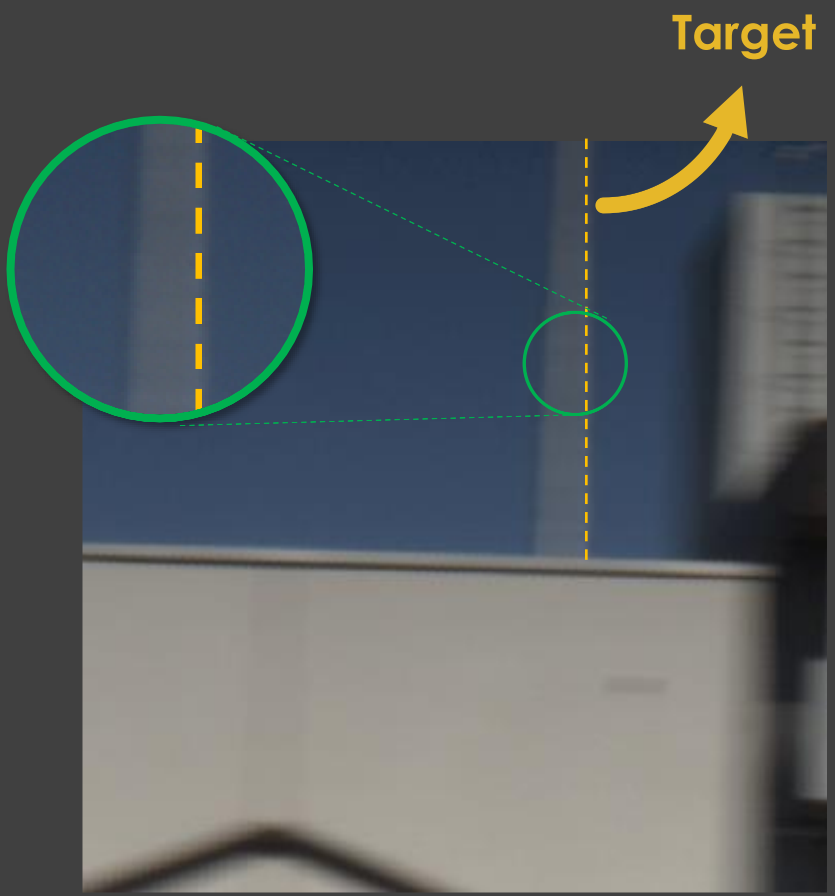
Displacement

Motion Prior Induced by Hardware

No **hardware** change

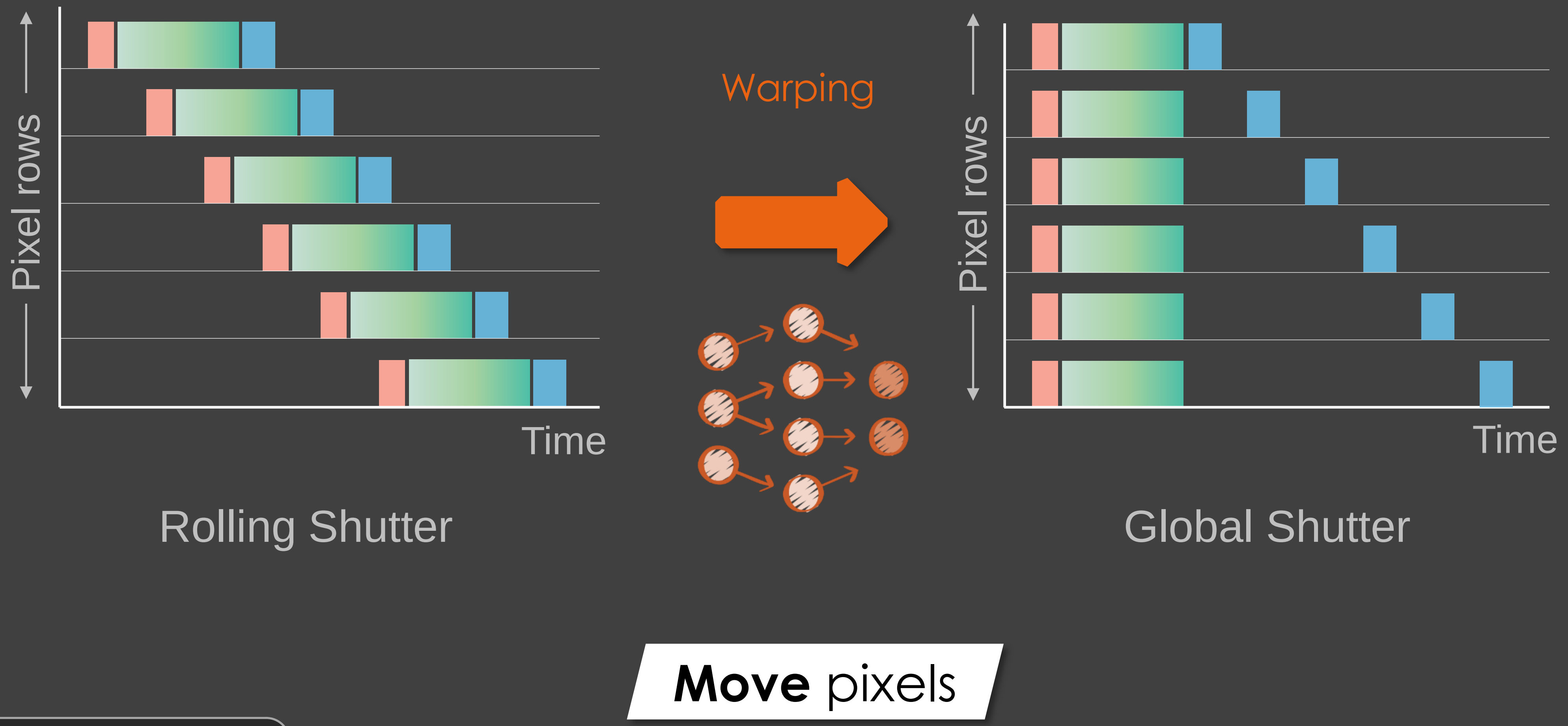
rolling shutter with global reset

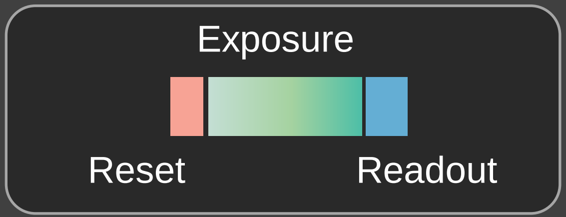




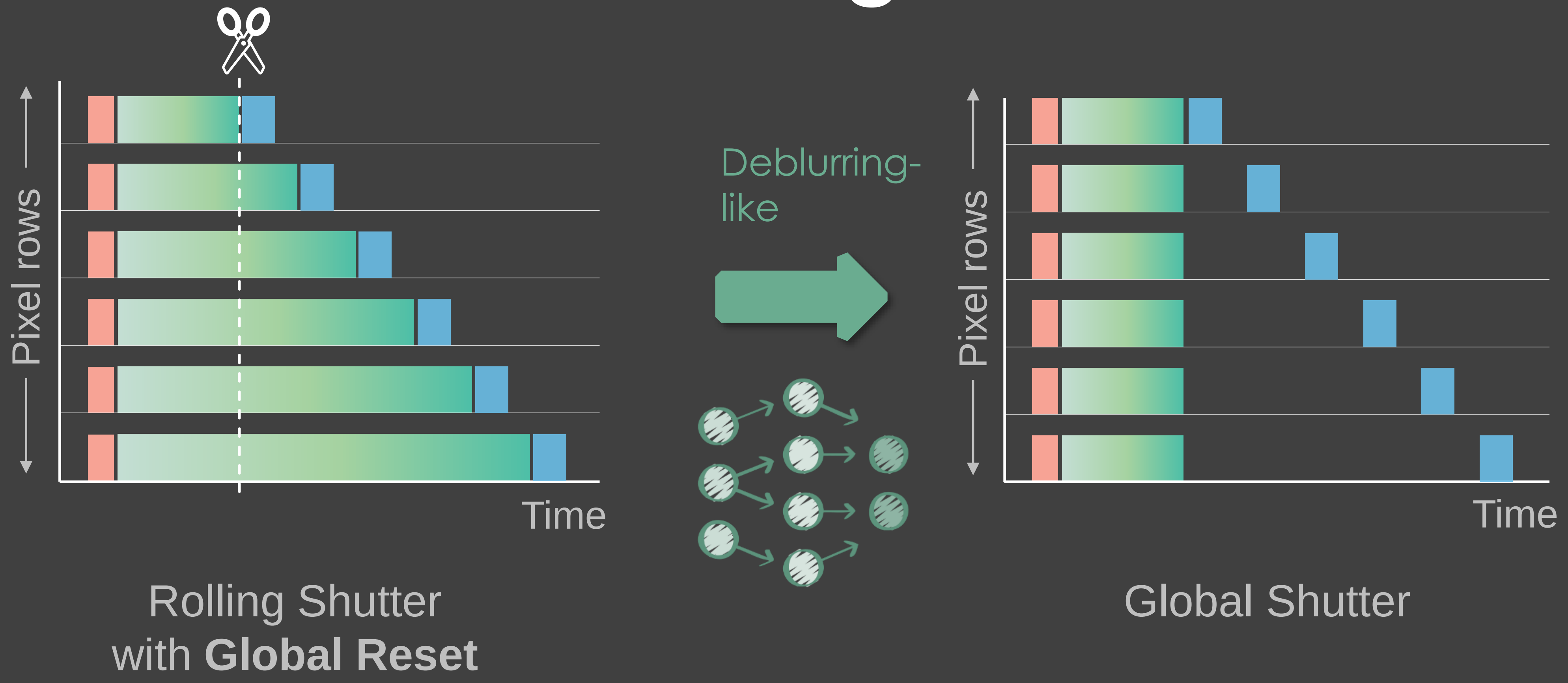
Row-wise blur

Traditional Learning Structure

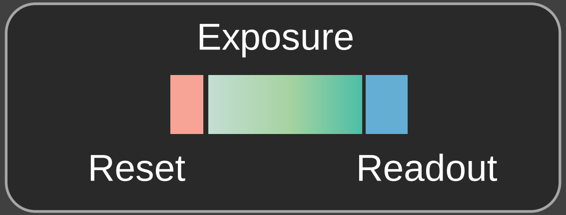




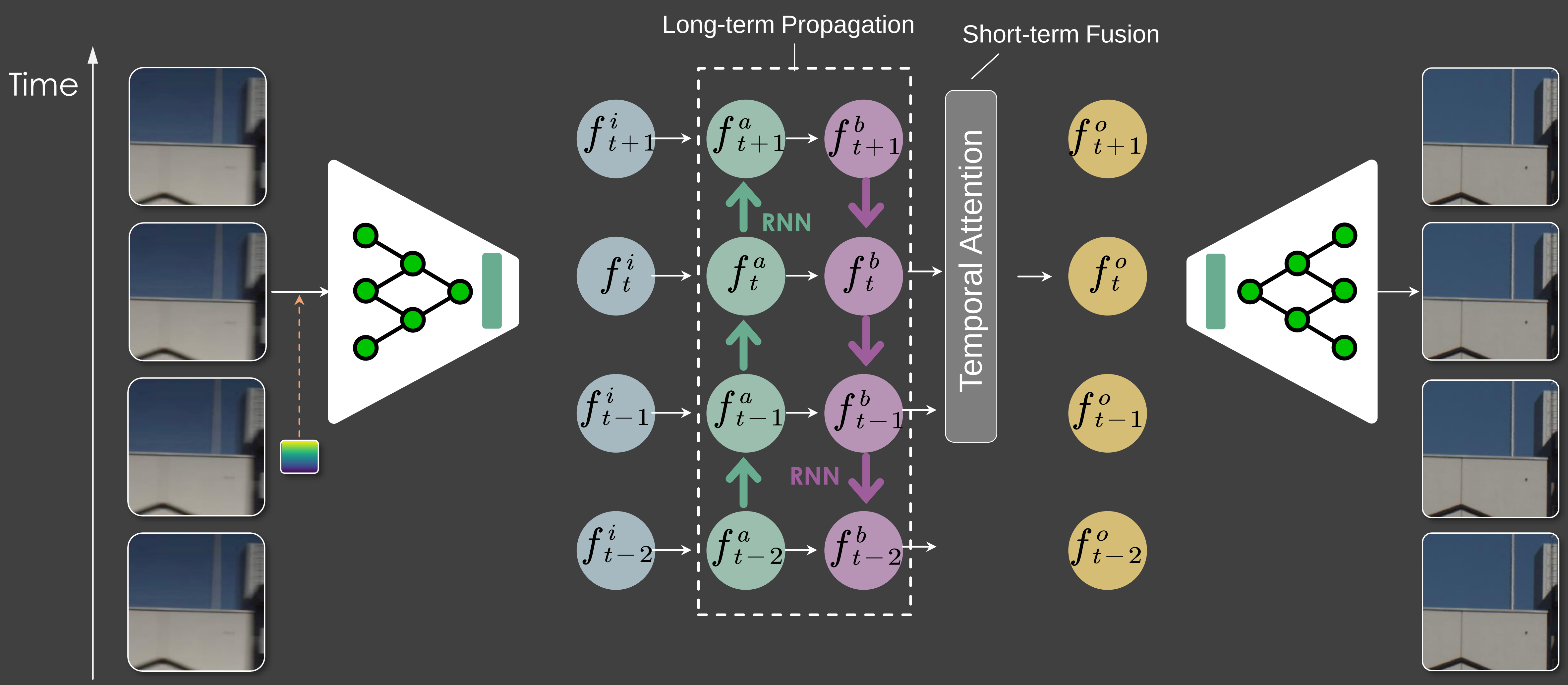
New Learning Structure



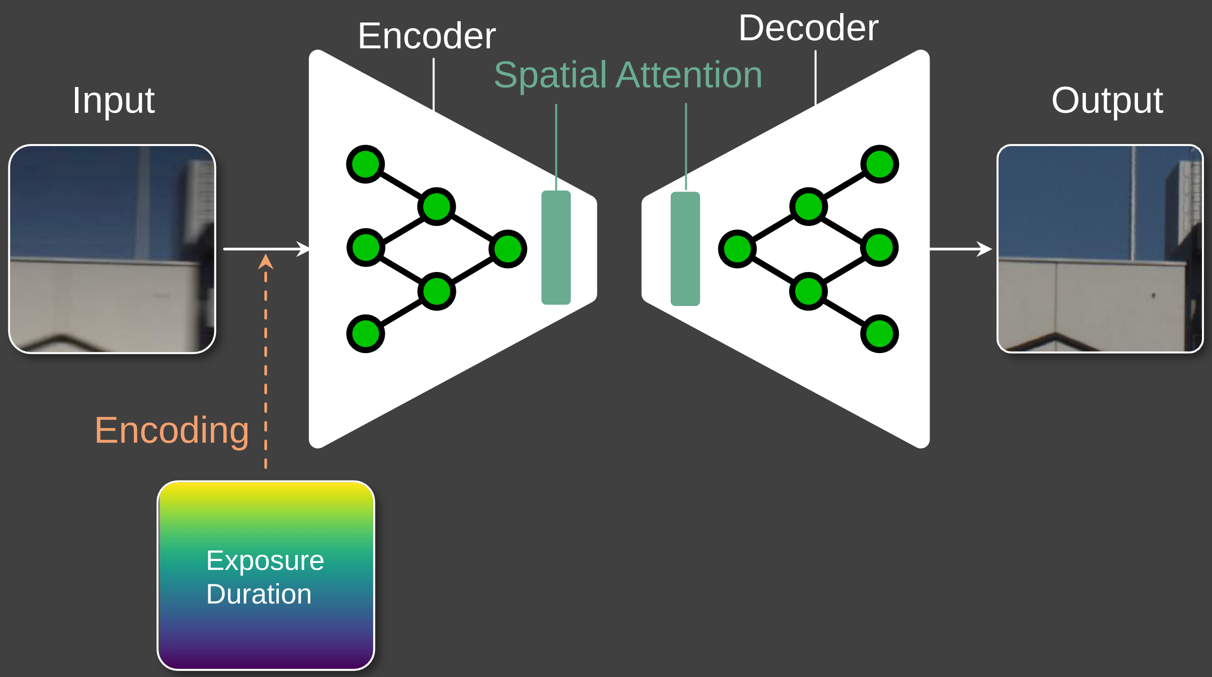
Remove over-exposure



Learning Algorithm

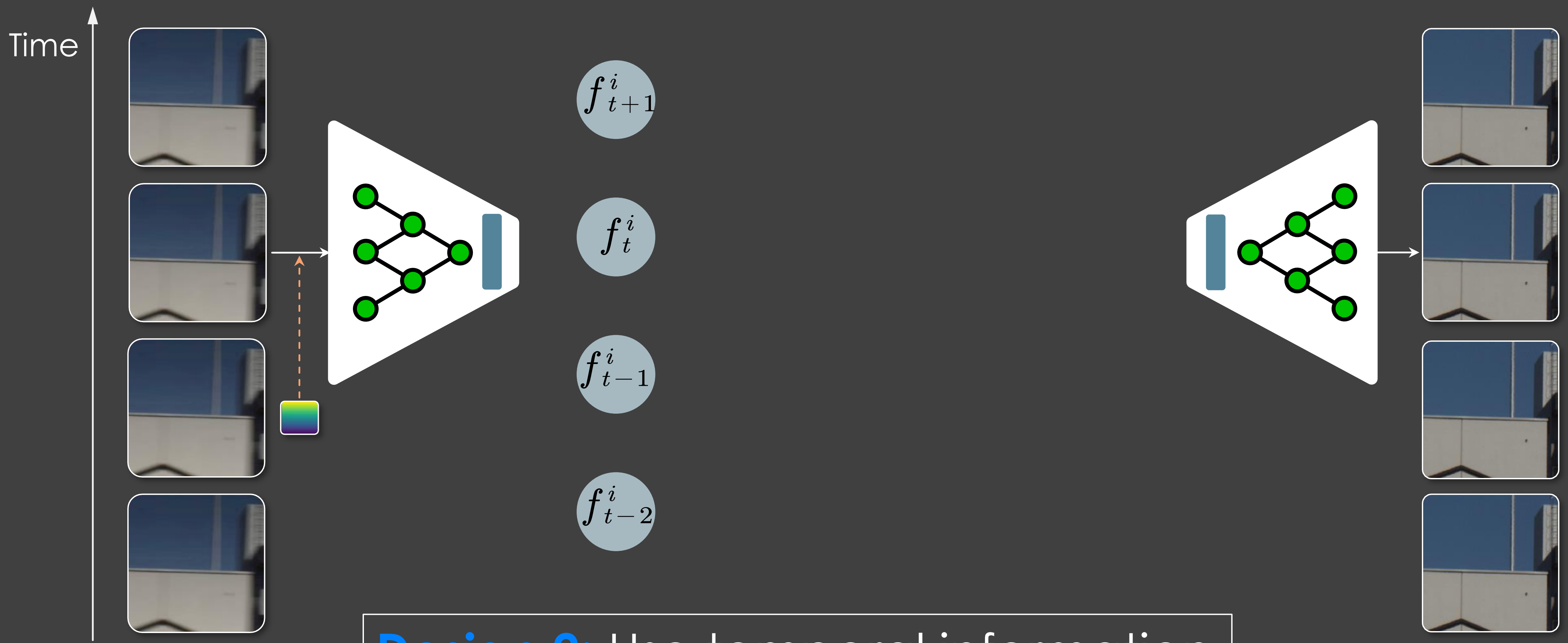


Learning Algorithm



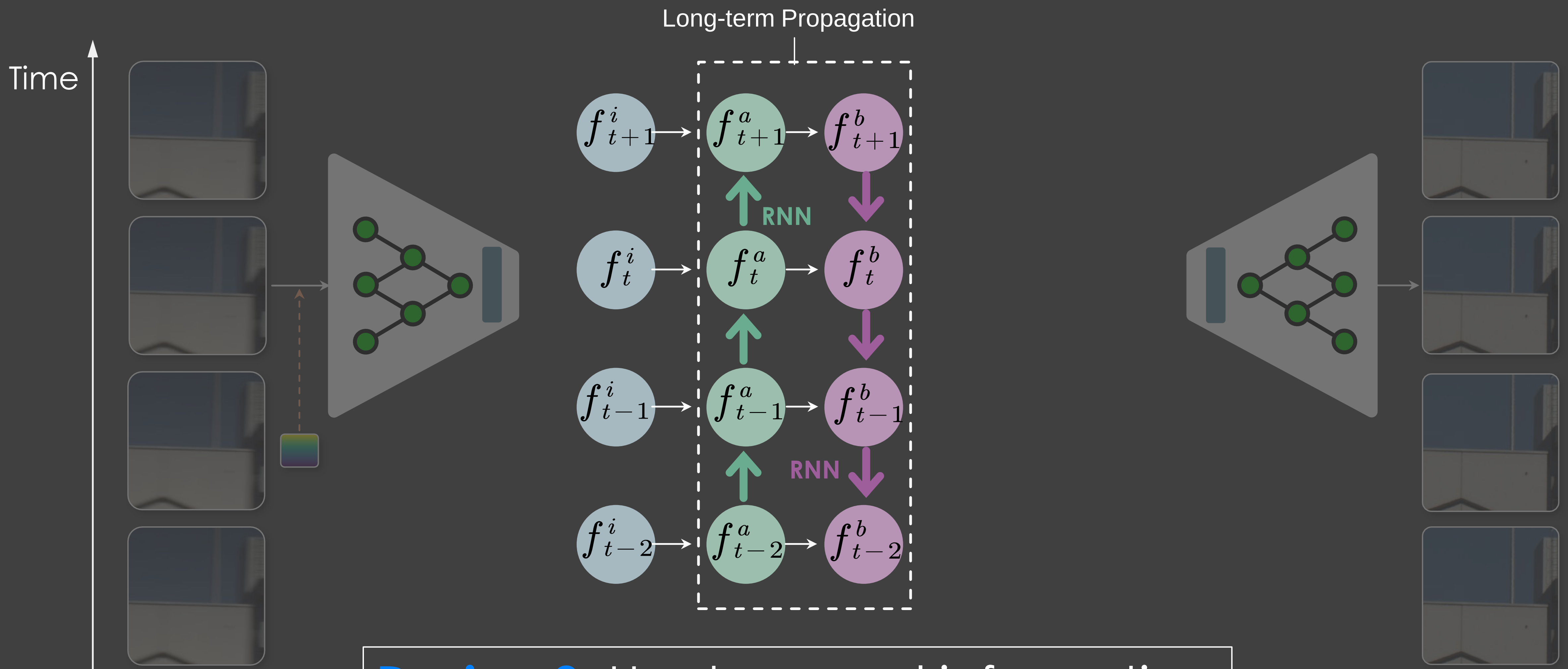
Design 1: Attend to row-wise degradations

Learning Algorithm



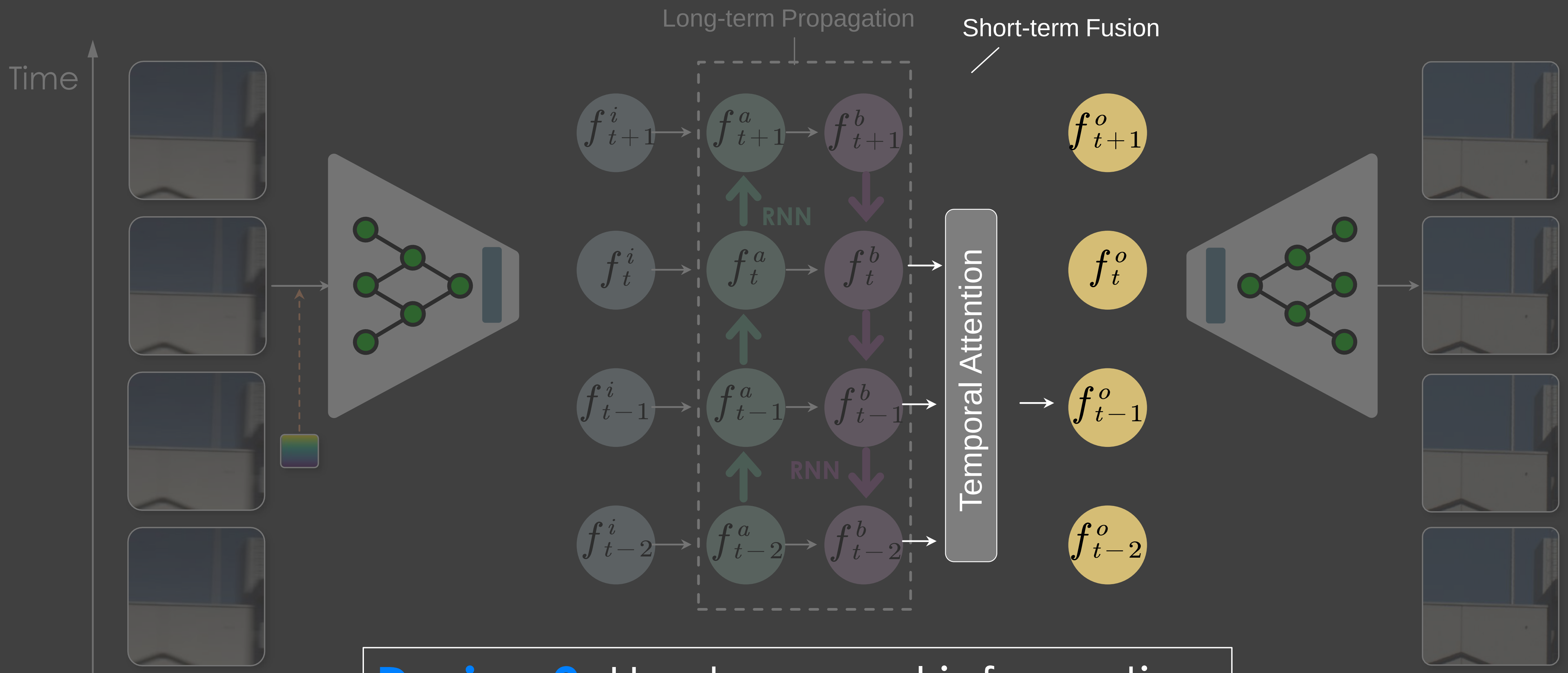
Design 2: Use temporal information

Learning Algorithm



Design 2: Use temporal information

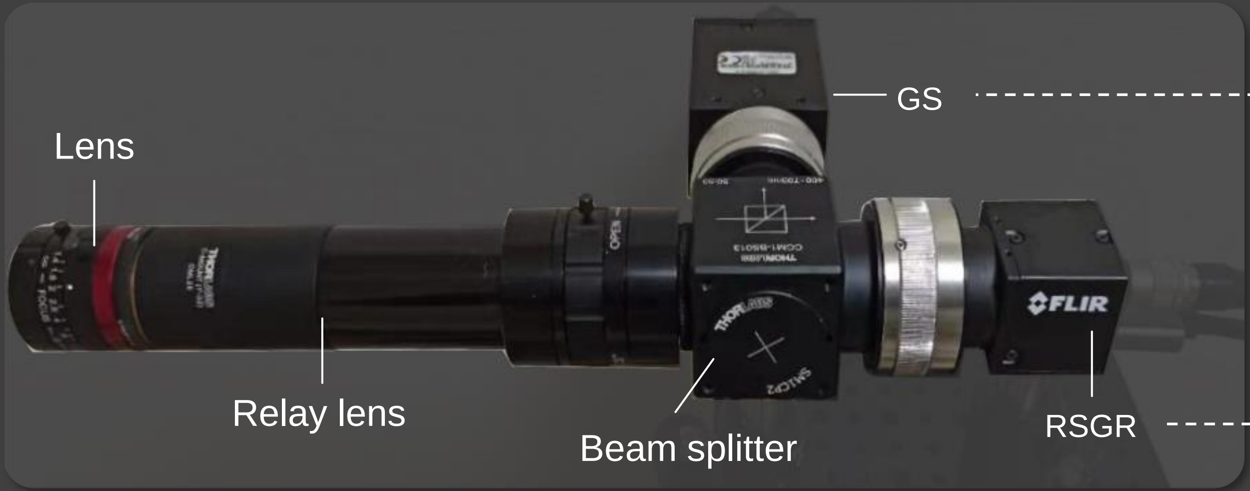
Learning Algorithm

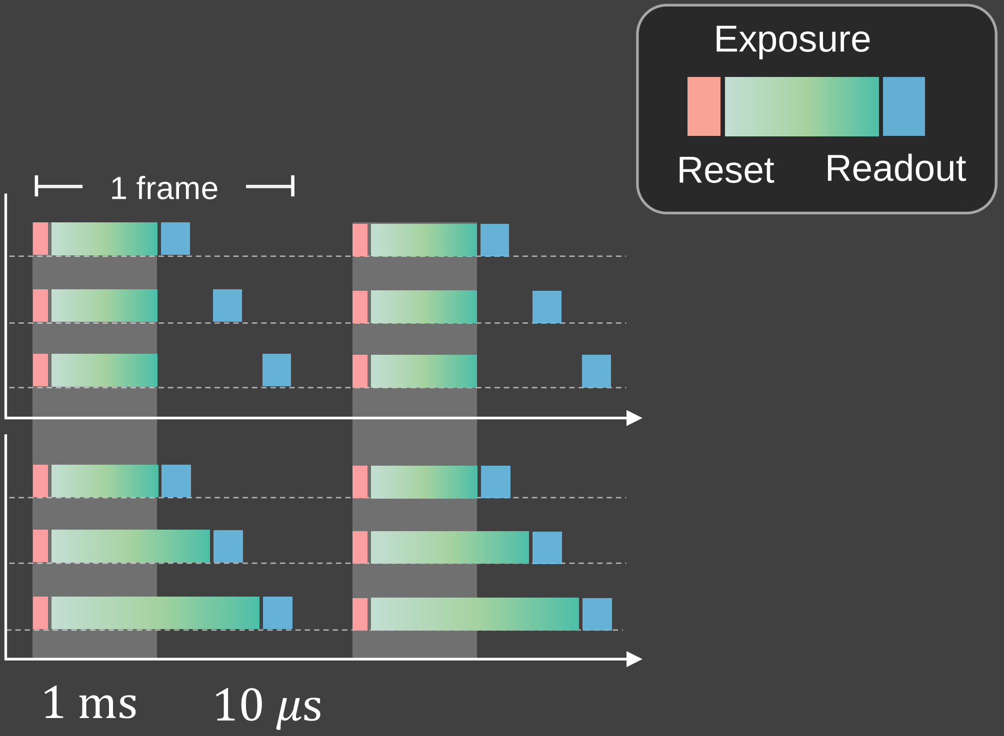


Design 2: Use temporal information

Real Camera System

Synchronized Cameras



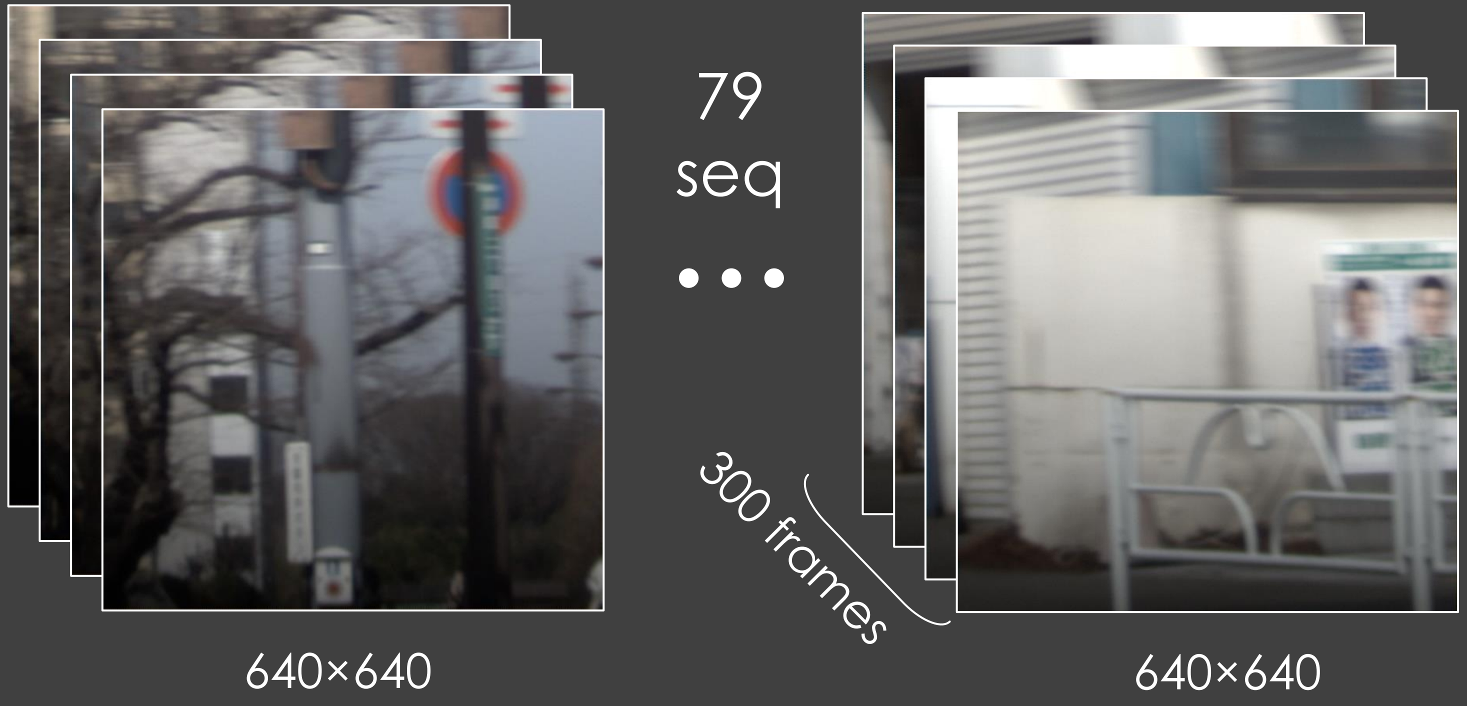


RSGR

GS

Dataset Detail

- 79 Video Sequences
- 300 frames per seq
- 640×640 resolution
- Ground truth per frame
- Outdoor, street



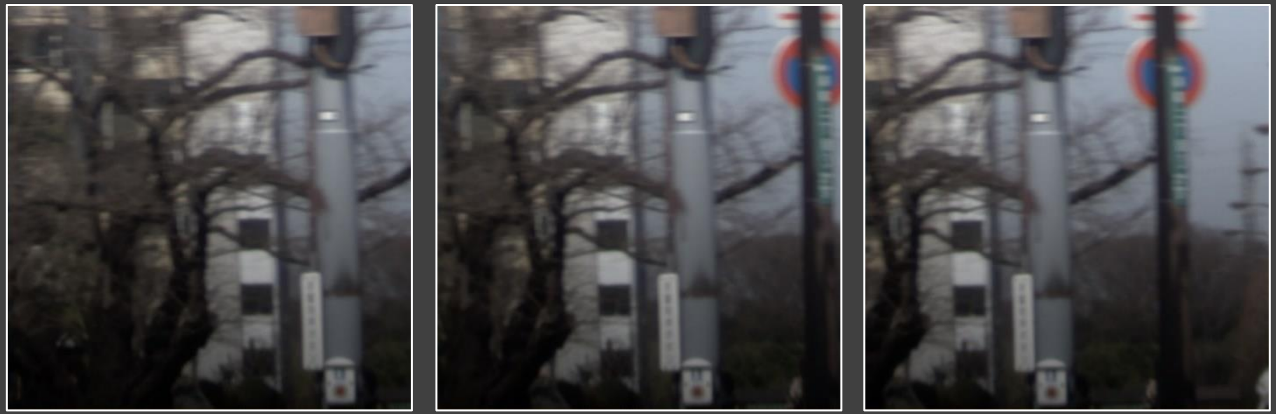
27 seq for training

52 seq for evaluation

Dataset Detail

- 79 Video Sequences
- 300 frames per seq
- 640×640 resolution
- Ground truth per frame
- Outdoor, street
  - Camera motion
  - Scene motion
  - Mixture

Camera motion



Mixed motion

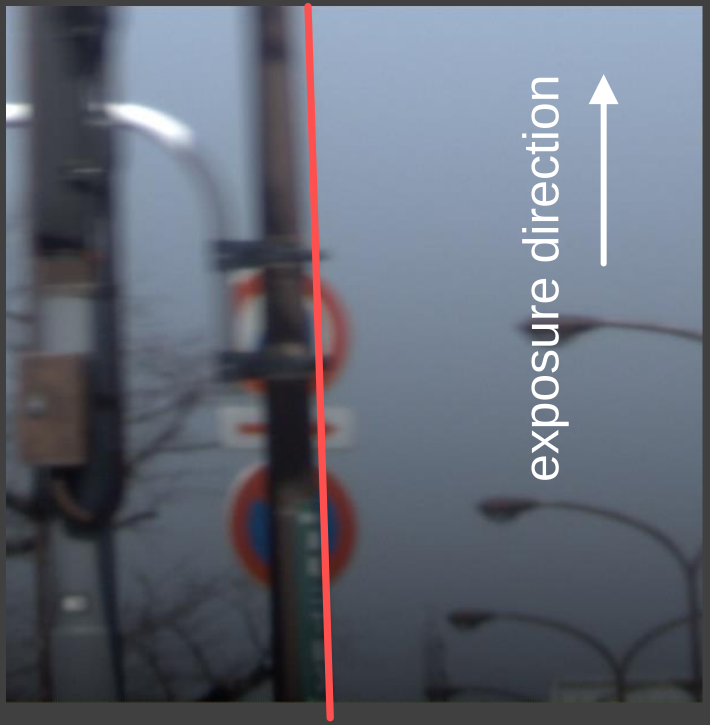
Results – Video

Input

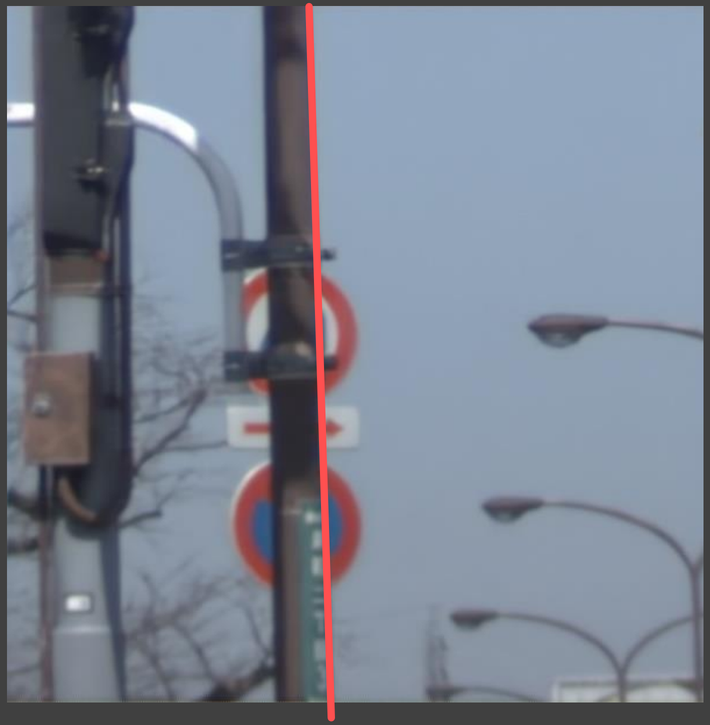
Output

Reference

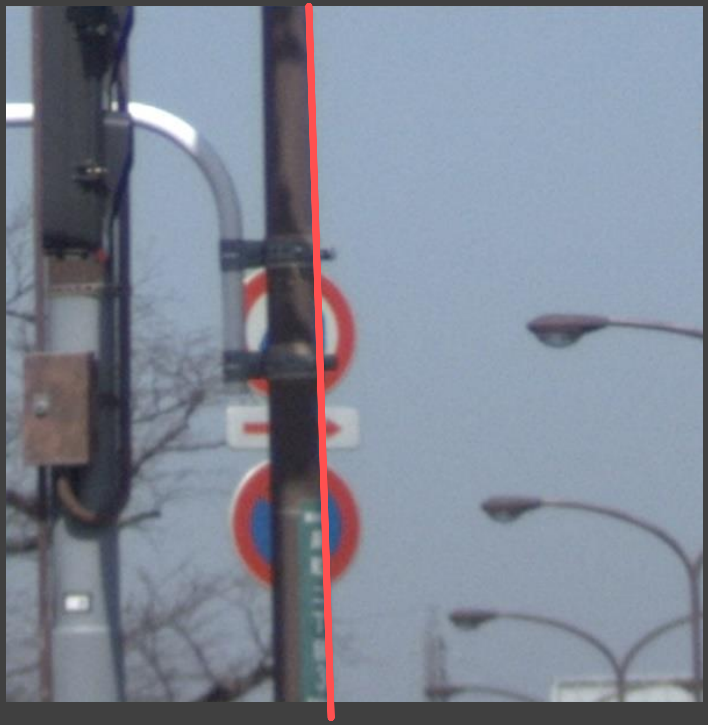
Results – Selected Frame



Input

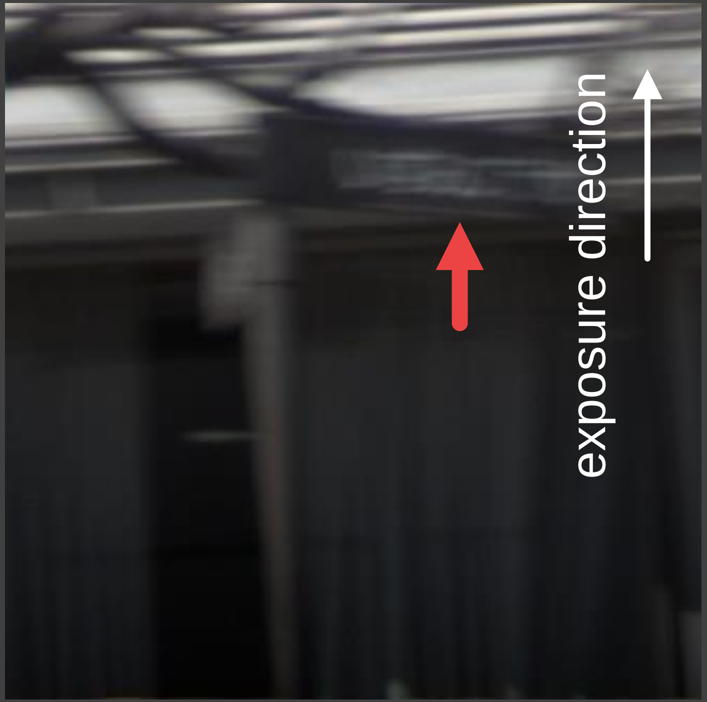


Output



Reference

Results – Selected Frame



Input

Output

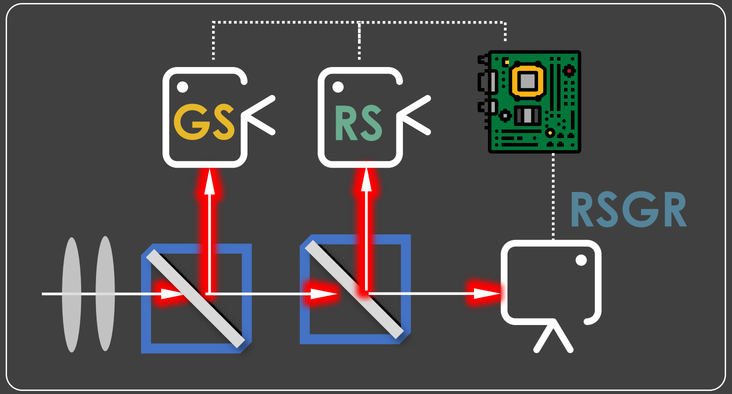
Reference

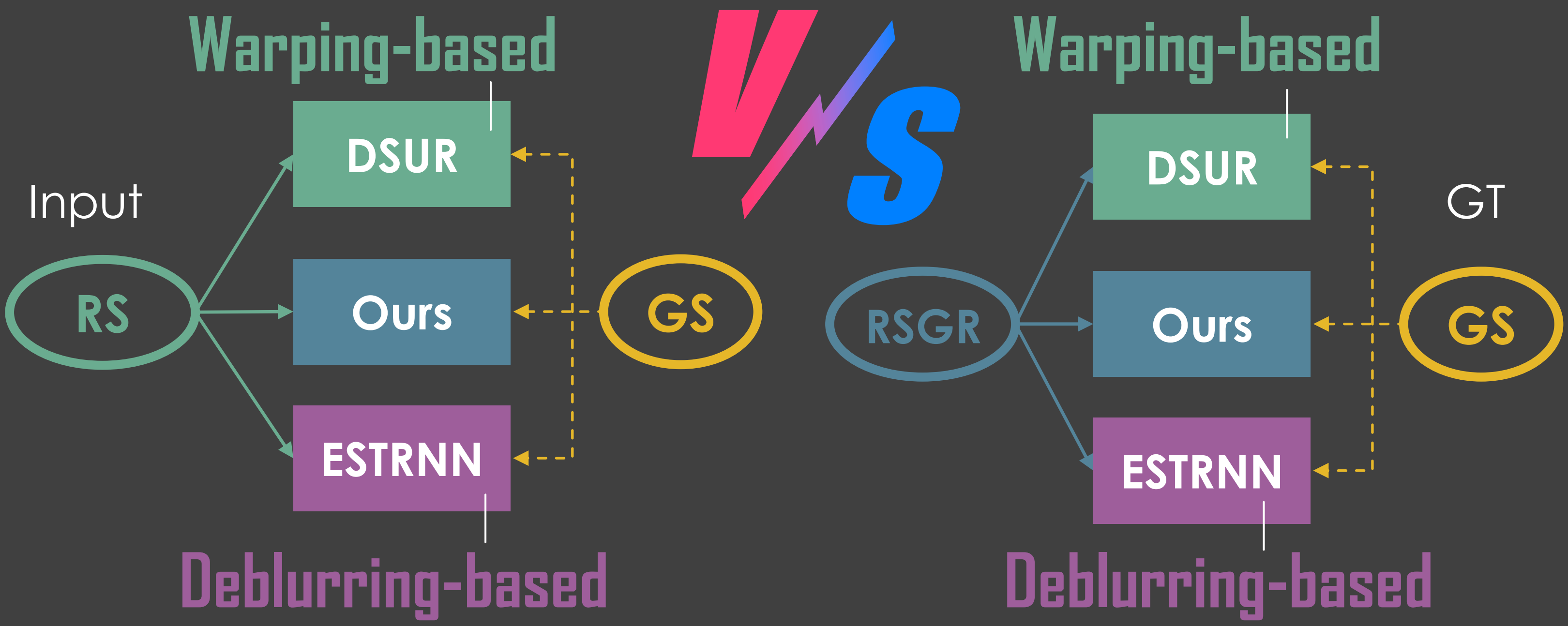
Results – Complex Degradations

Degradations Caused by Camera + Scene Motion

Input Output Reference

Evaluation – Setup





85 sequences for train

14 sequences for validation

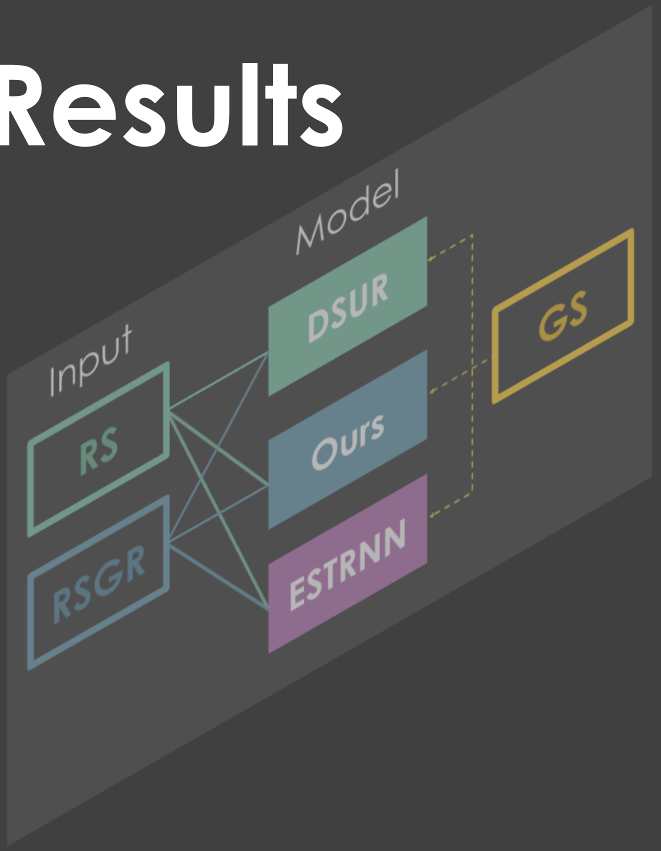
49 sequences for test

Metrics **PSNR** and **SSIM**

Liu et al, DSUR, CVPR'20 Zhong et al, ESTRNN, ECCV'20

Evaluation – Quantitative Results

Method	Type	PSNR↑ RS	PSNR↑ RSGR	SSIM↑ RS	SSIM↑ RSGR
Input	None	16.1612	17.3206	0.5356	0.6696
DSUR	RS correction	20.0274	22.5732	0.6883	0.7873
ESTRNN	Deblurring	19.3529	22.2271	0.6986	0.7974
Ours	RSGR correction	22.9542	27.8586	0.7870	0.8822



RSGR+[any model] **outperforms** RS+[any model]

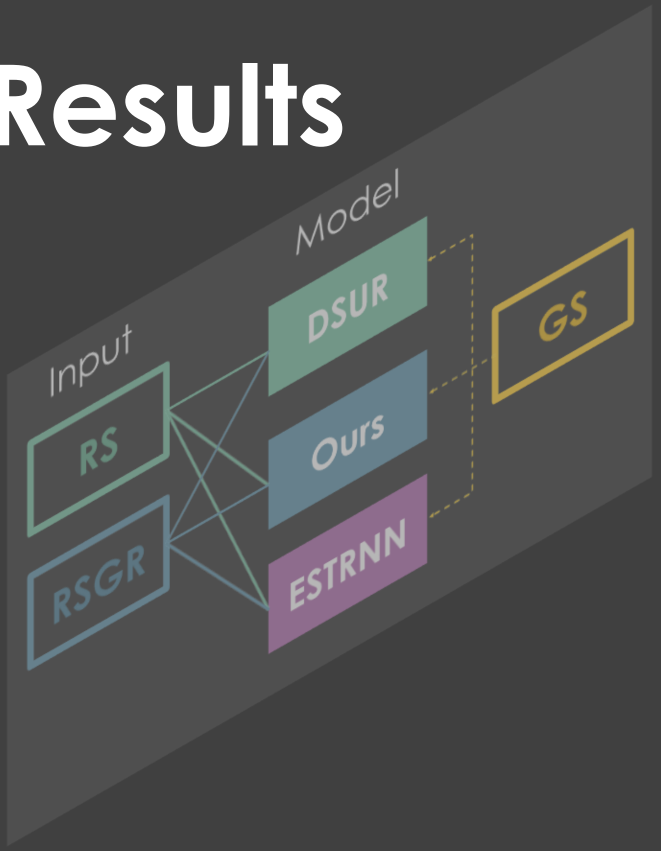
RSGR is easier to correct than RS

Liu et al, DSUR, CVPR'20
Zhong et al, ESTRNN, ECCV'20

Evaluation – Quantitative Results

Method	Type	PSNR↑ RS	PSNR↑ RSGR	SSIM↑ RS	SSIM↑ RSGR
Input	None	16.1612	17.3206	0.5356	0.6696
DSUR	RS correction	20.0274	22.5732	0.6883	0.7873
ESTRNN	Deblurring	19.3529	22.2271	0.6986	0.7974
Ours	RSGR correction	22.9542	27.8586	0.7870	0.8822

+7.8dB **+0.2**



RSGR+Ours achieves the **best** score

Liu et al, DSUR, CVPR'20
Zhong et al, ESTRNN, ECCV'20

Evaluation – Qualitative Results

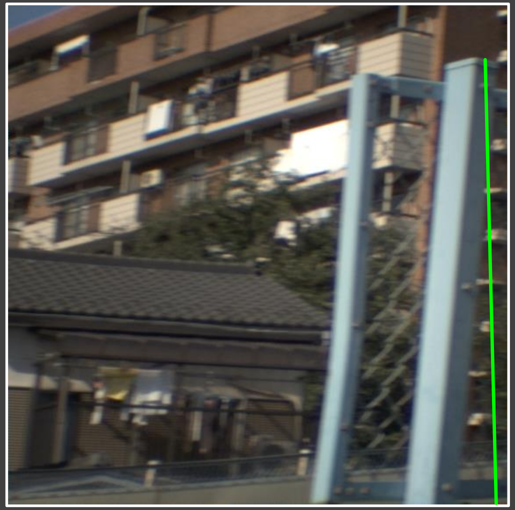
RS+DSUR

RSGR+Ours

Ground Truth

Liu et al, DSUR, CVPR'20

Evaluation – Qualitative Results

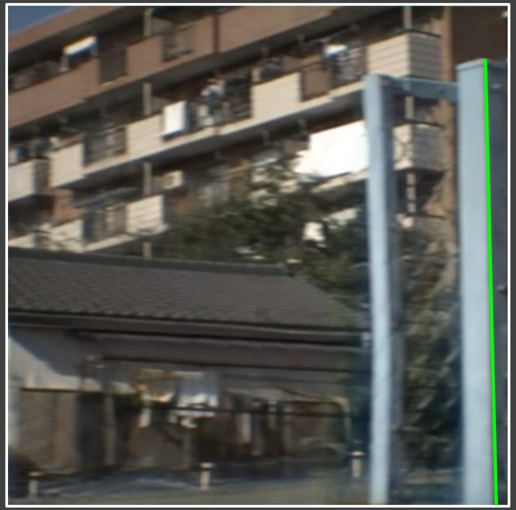


Input RS

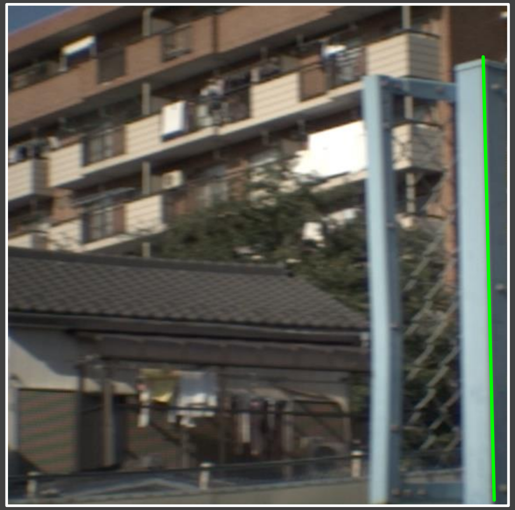
RS+DSUR

RSGR+Ours is **closer** to target than RS+DSUR

Input RSGR



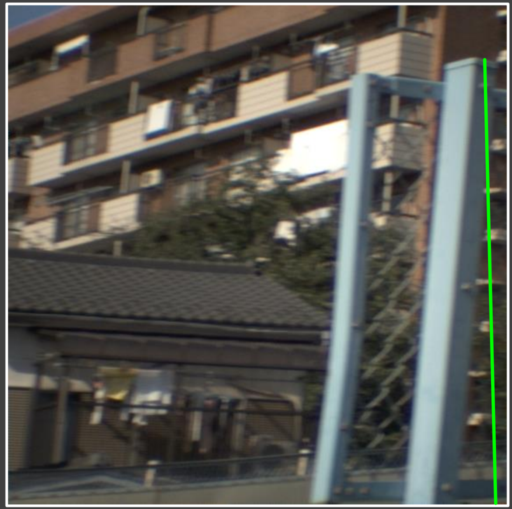
RSGR+Ours



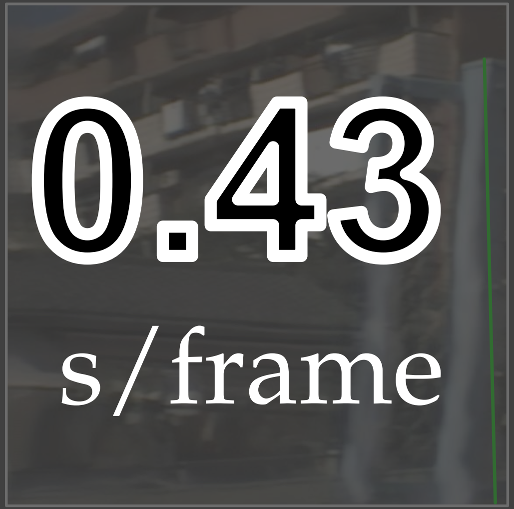
Target

Liu et al, DSUR, CVPR'20

Evaluation – Results

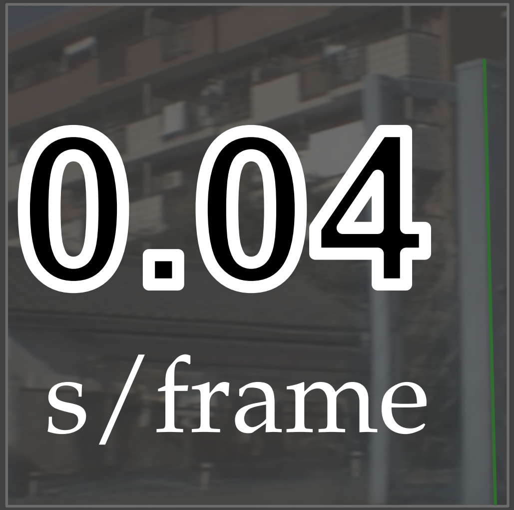


Input RS



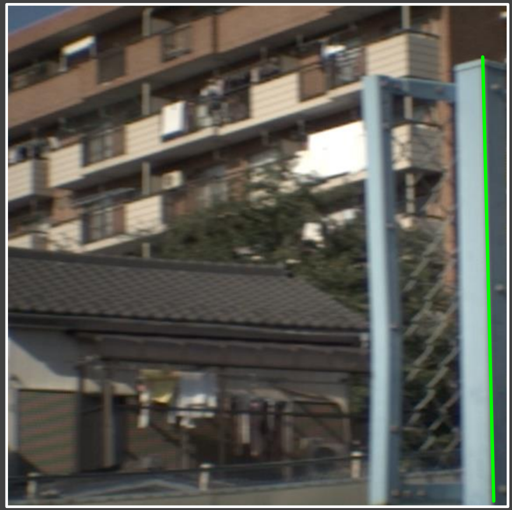
RS+DSUR

Input RSGR



RSGR+Ours

RSGR+Ours is **10x faster** than RS+DSUR



Target

Liu et al, DSUR, CVPR’20

Summary – Neural Global Shutter

- We are the first to use the RSGR hardware feature in academics. This feature induces motion priors
- With the RSGR hardware feature, we convert traditional RS correction problem into a deblurring-like one
- We develop an effective algorithm
- We build a system to capture data for supervised training

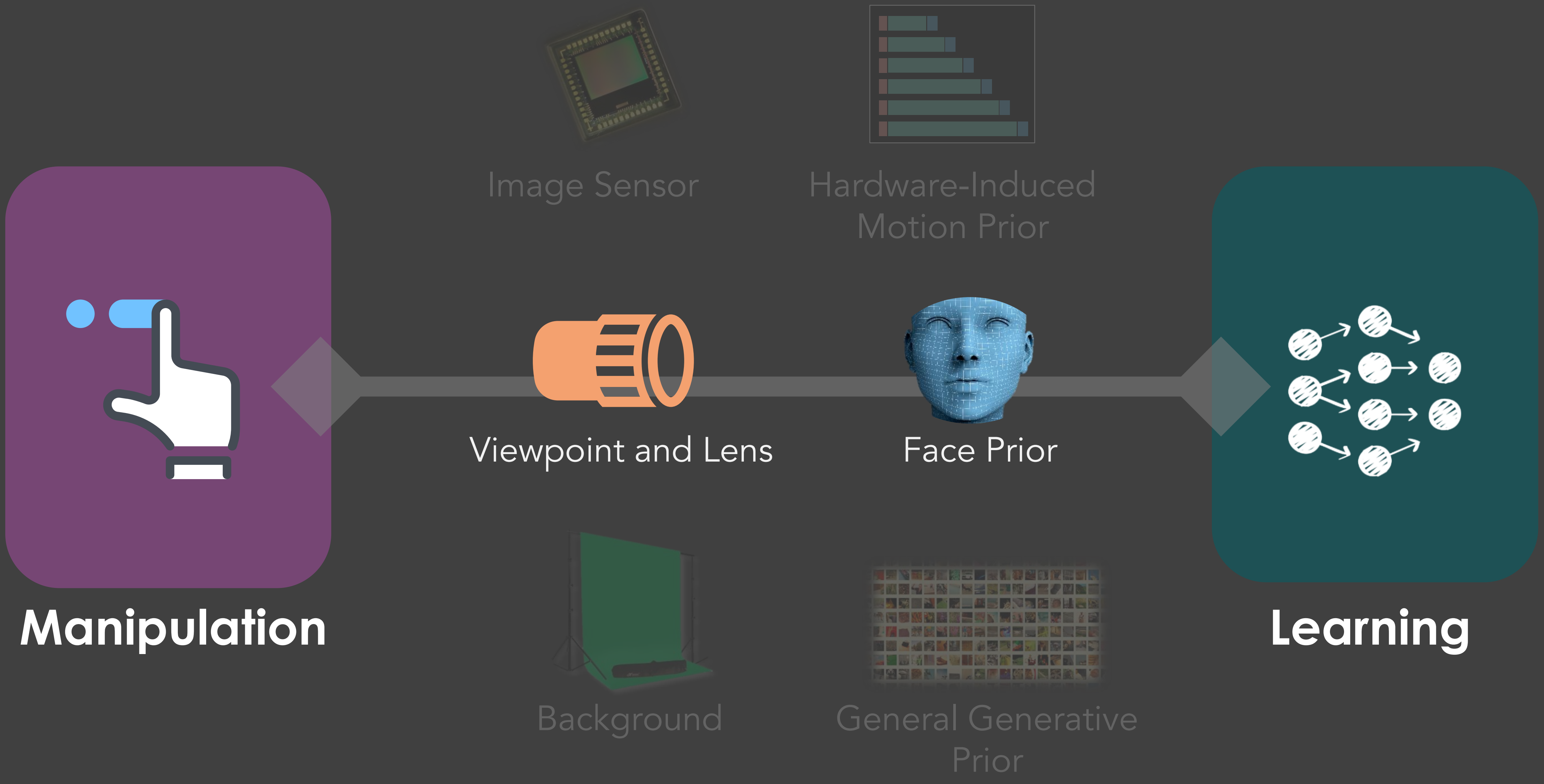
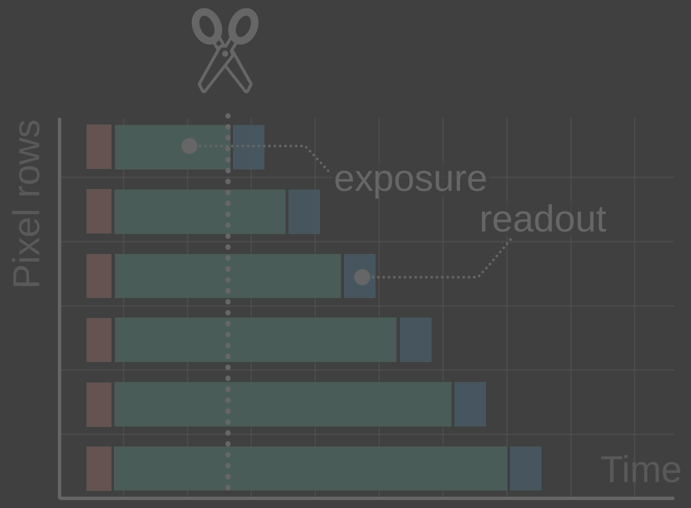


Image Sensor
Hardware-Induced Motion Prior
Manipulation
Viewpoint and Lens
Face Prior
Learning
Background
General Generative Prior

Image Sensor



Neural Global Shutter

Wang et al, CVPR 2022

Viewpoint + Lens

Portrait Distortion Correction

Wang et al, IJCV 2024

Background

Matting by Generation

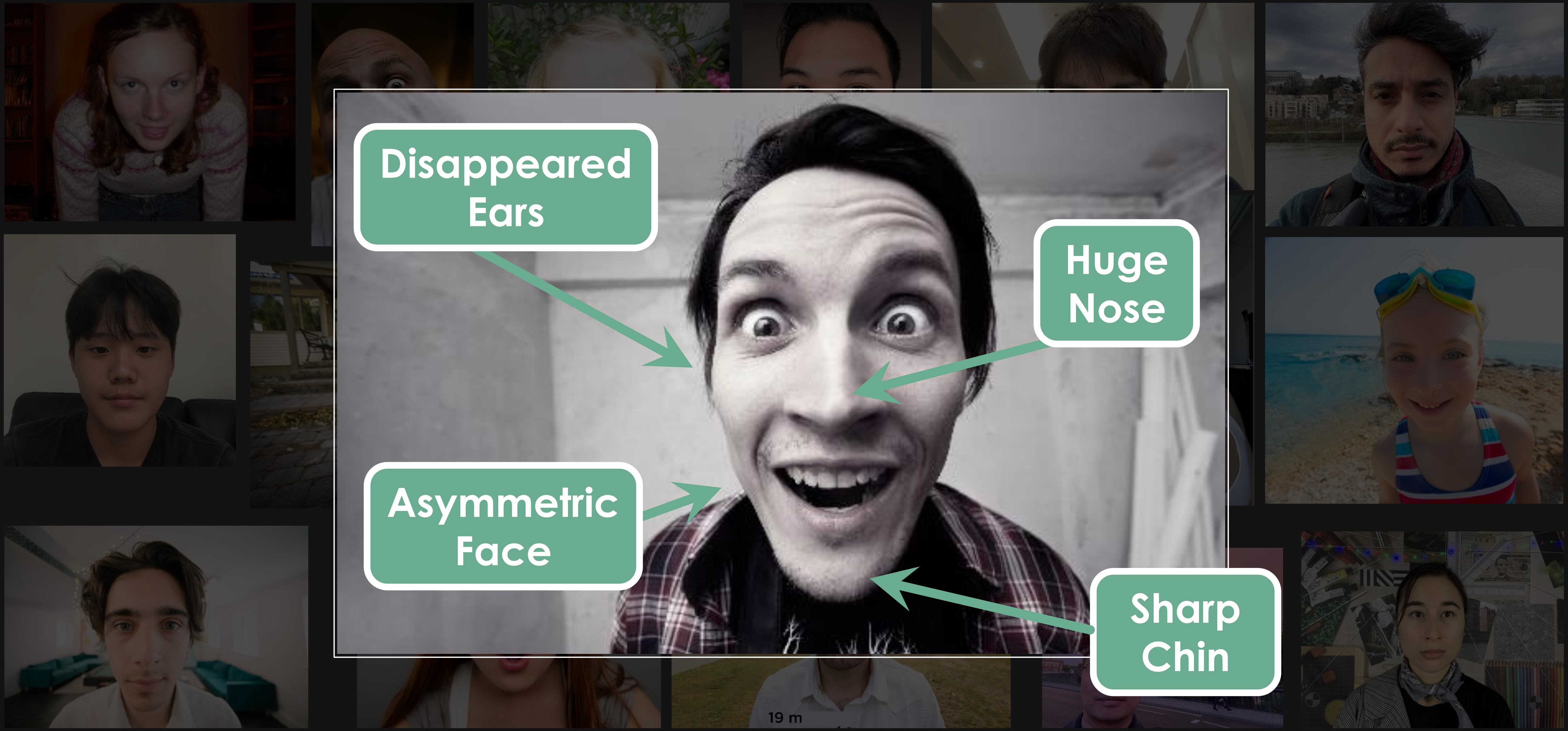
Wang et al, SIGGRAPH 2024

New Learning Structures

Warping → Rendering

SIGGRAPH'16	Fried et al.
ICCV'19	Zhao et al.

Ours


Disappeared Ears
Huge Nose
Asymmetric Face
Sharp Chin
19 m

Short Camera-to-Subject Distance

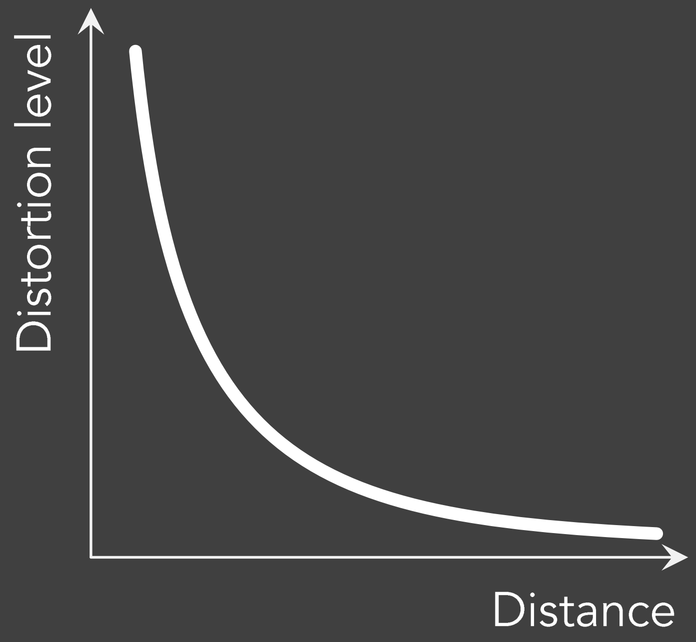
Perspective Projection

Depth Variation

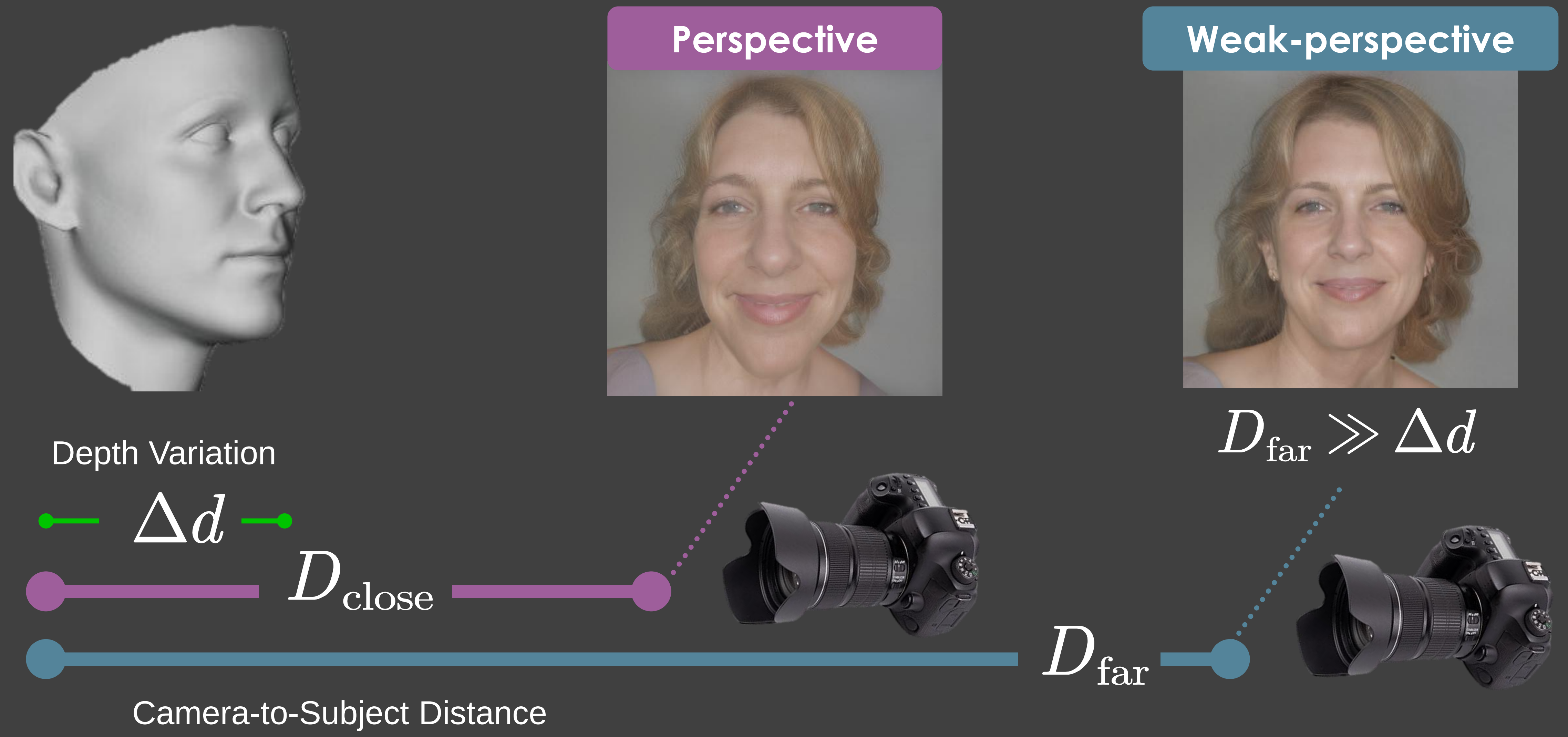
Δd

D_{close}

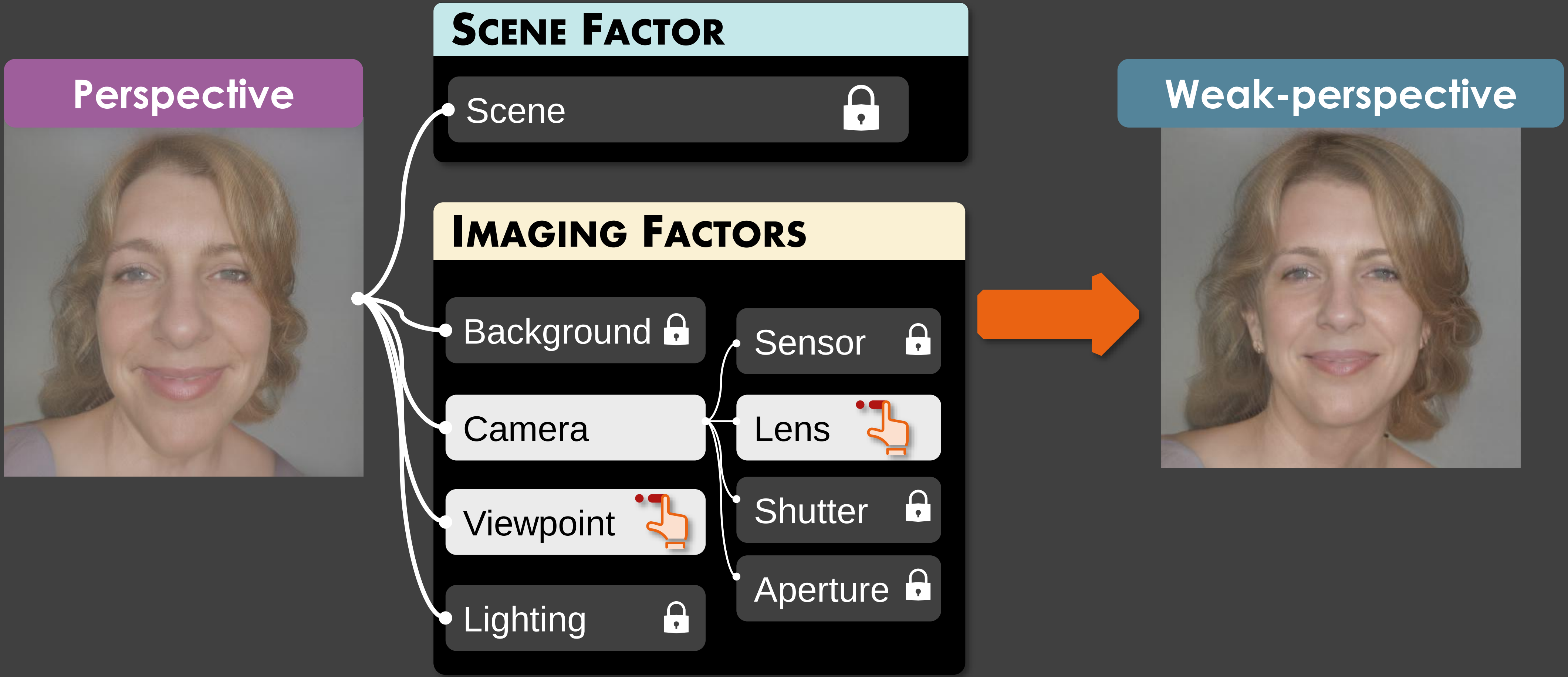
Camera-to-Subject Distance



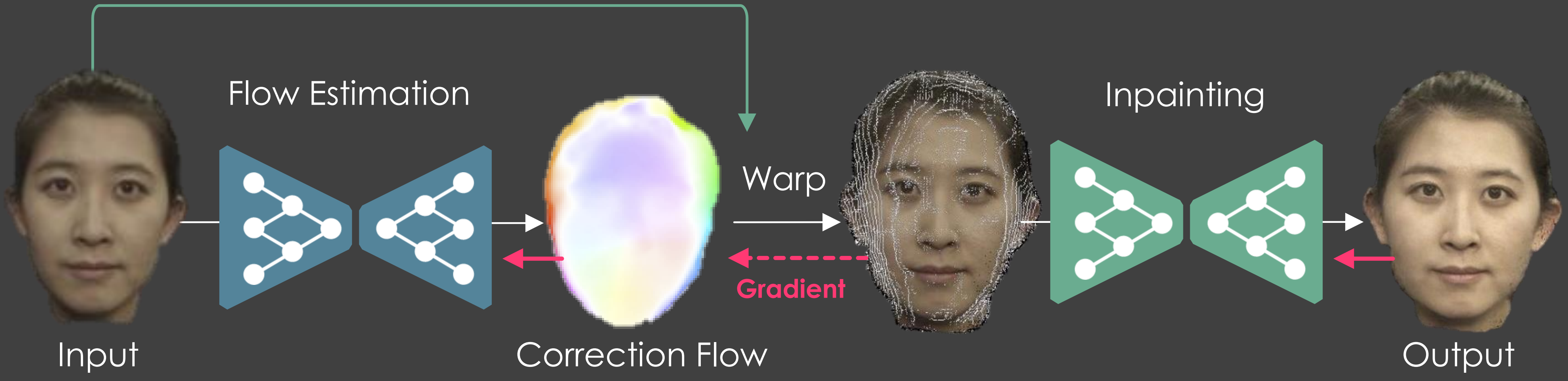
Weak-perspective Projection



Manipulate Viewpoint and Lens



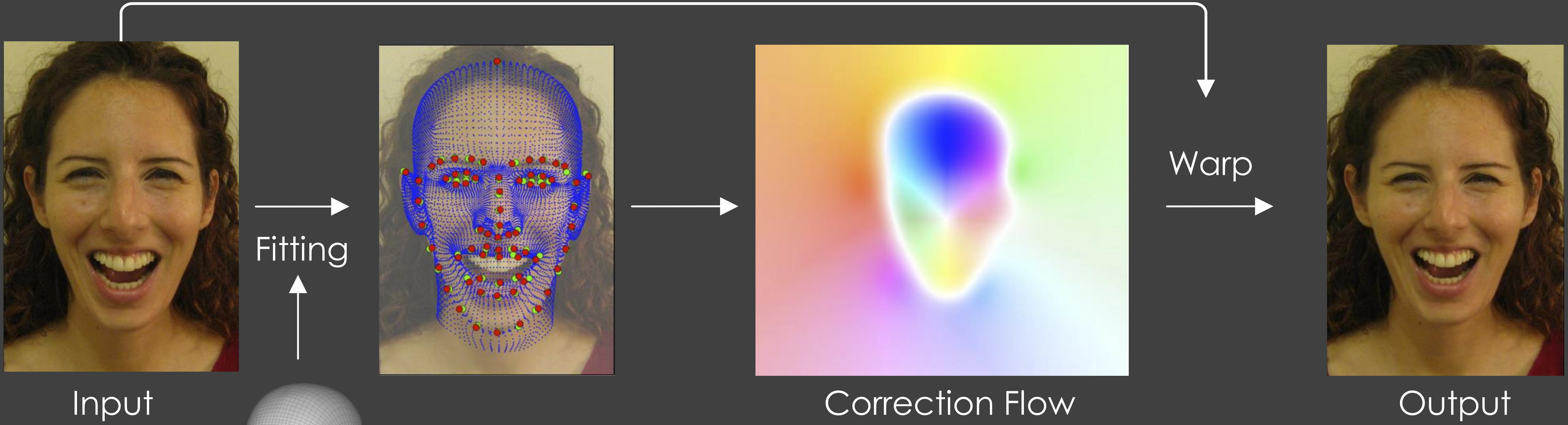
Existing Methods – Warping-based

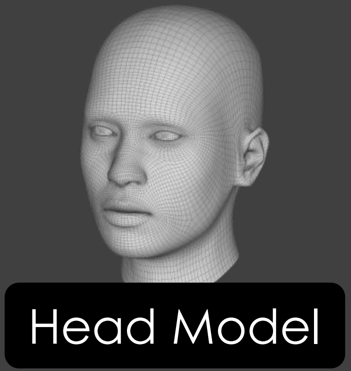


Learning-based flow estimation

Zhao et al, ICCV'19

Existing Methods – Warping-based

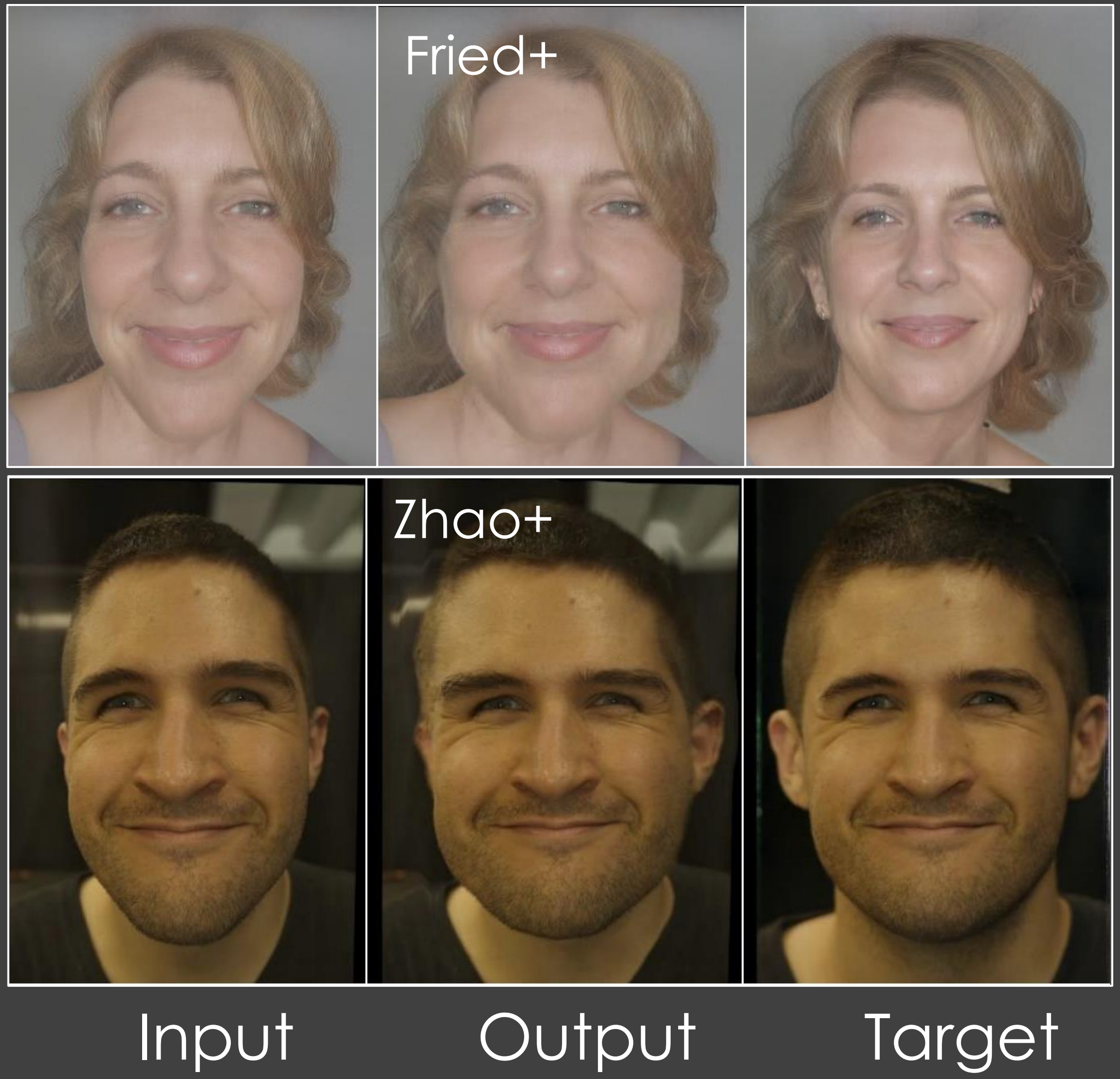




Reconstruction-based flow estimation

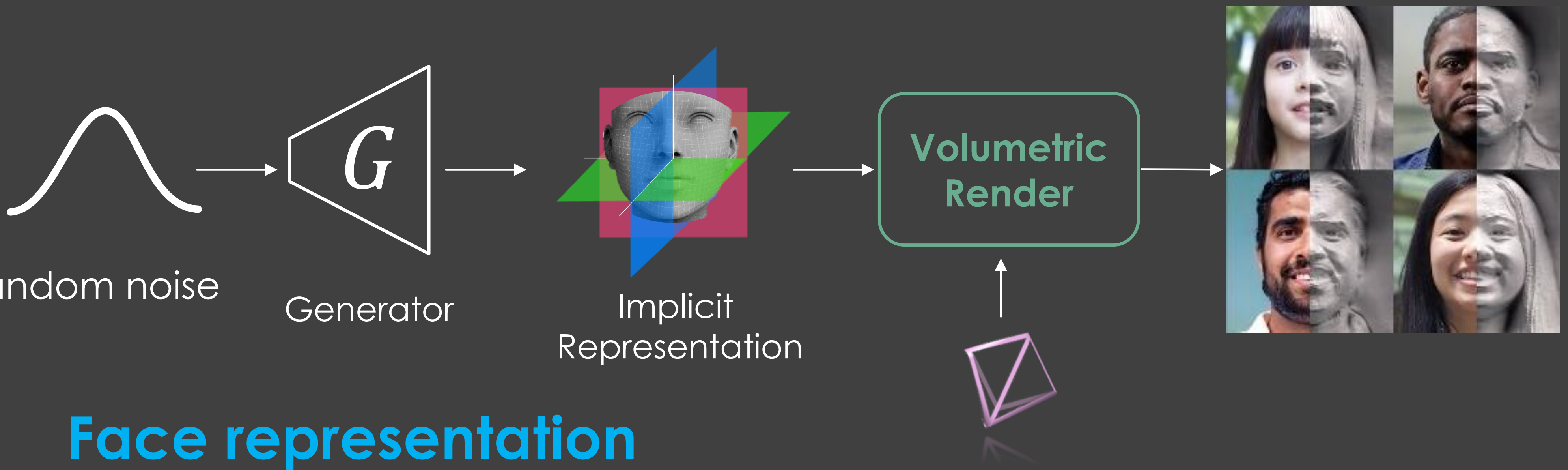
Fried et al, SIGGRAPH'16

Limitations of Existing Methods



- **Flow warping only repeats existing pixels**
  - CANNOT reveal occluded regions
    - Invisible ear, cheek, neck …
  - CANNOT deal with serious distortion
    - When camera-to-face distance is 20–40cm
  - Not 3D-aware
    - Face shape is flawed
- **Learning-based method (Zhao+) is worse**
  - Require a lot of training data
  - Hard to generalize
  - CANNOT continuously change

Real-world Prior Induced by 3D GANs



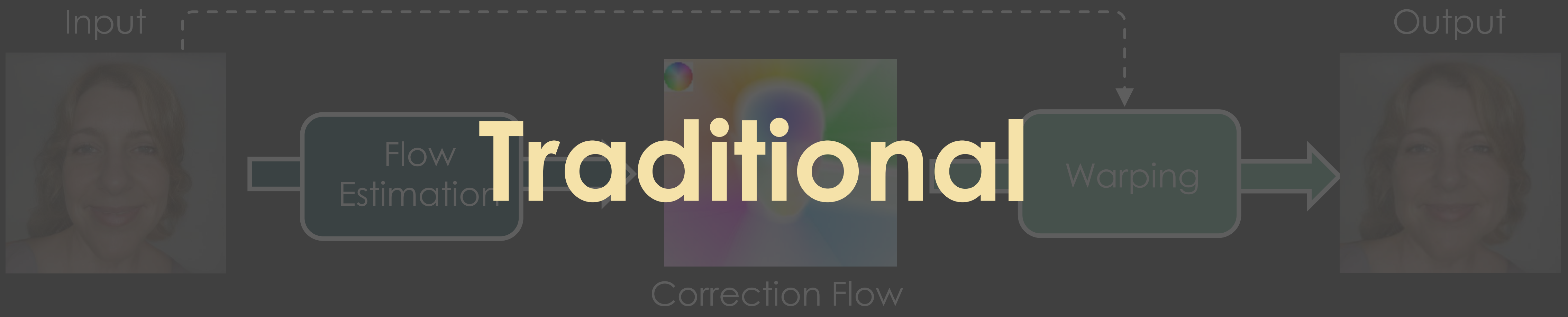
Face representation

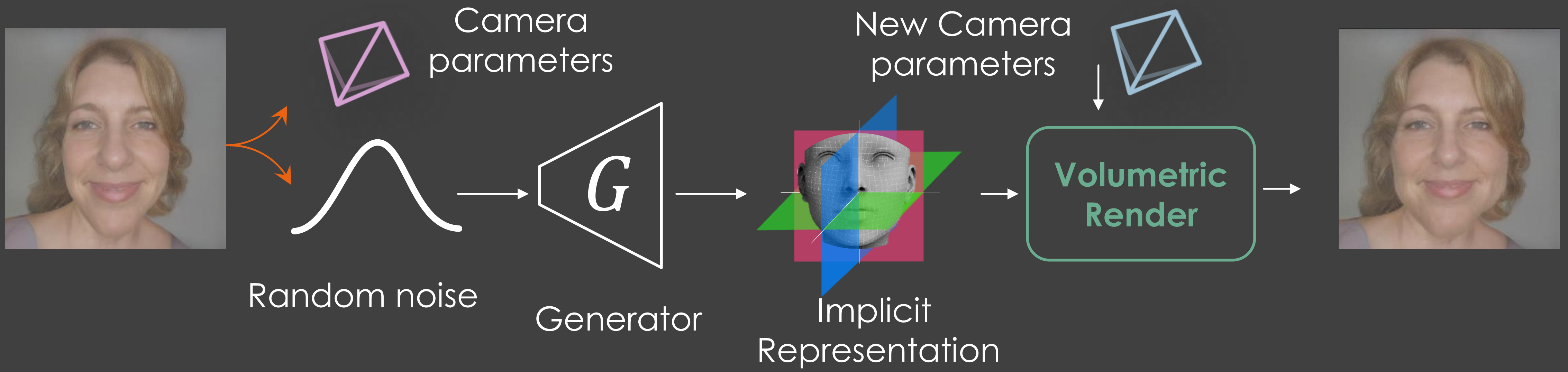
Camera parameter

Intrinsic and Extrinsic

Chan et al, EG3D, CVPR’22

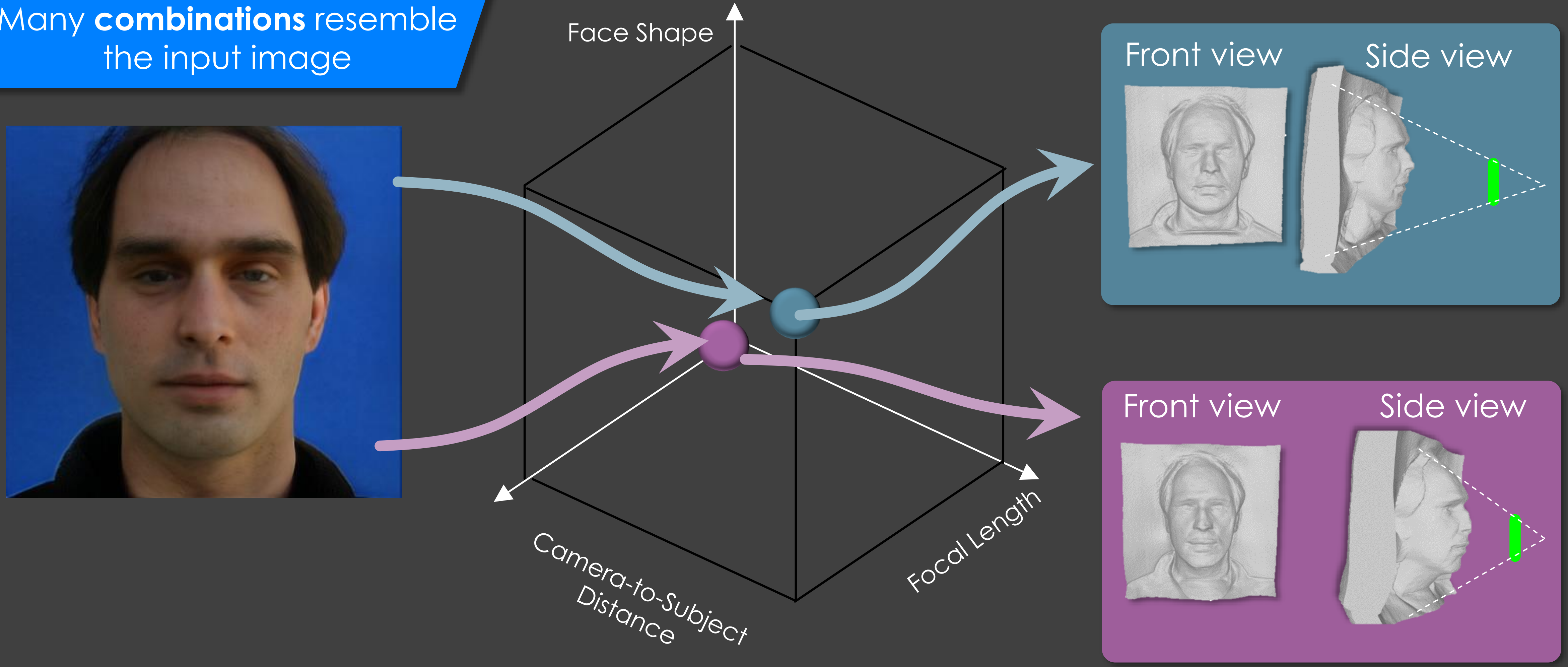
New Learning Structure





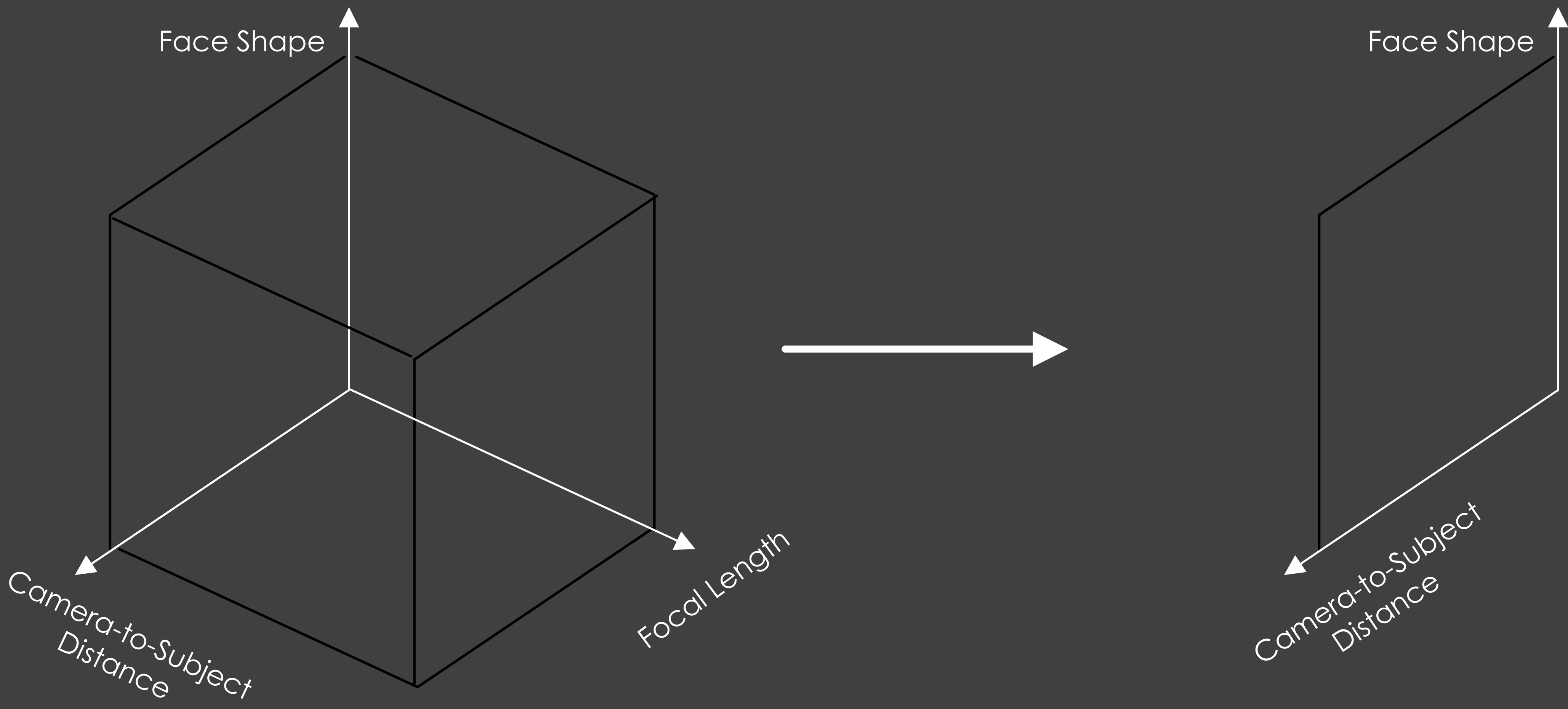
Challenge I: Ambiguity

Many **combinations** resemble the input image



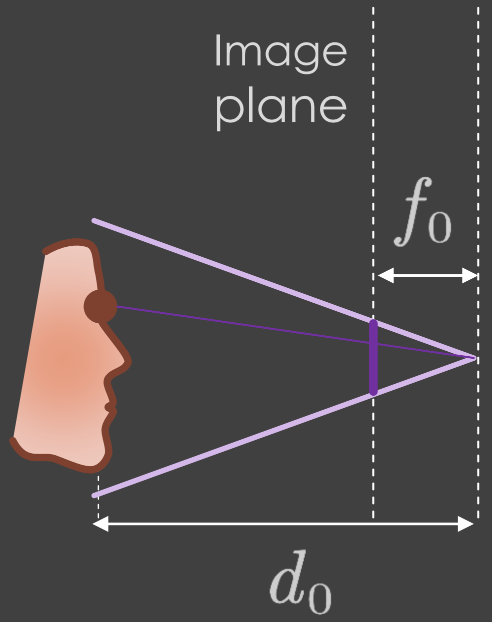
Perspective-aware 3D GAN Inversion

Focal Length Re-parameterization



Perspective-aware 3D GAN Inversion

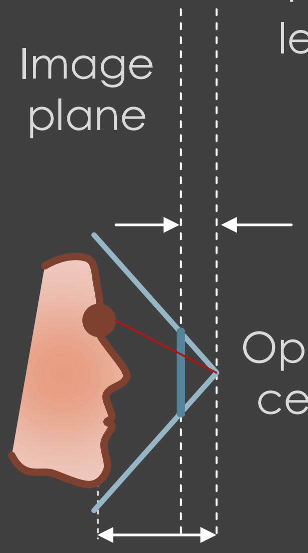
Focal Length Re-parameterization



Camera-to-Subject Distance

Translation change

$$\Delta \mathbf{t} = \begin{pmatrix} \Delta t_x \\ \Delta t_y \\ \Delta t_z \end{pmatrix}$$



$$d = d_0 + \Delta t_z$$

Camera-to-Subject Distance

Compensate approximation error

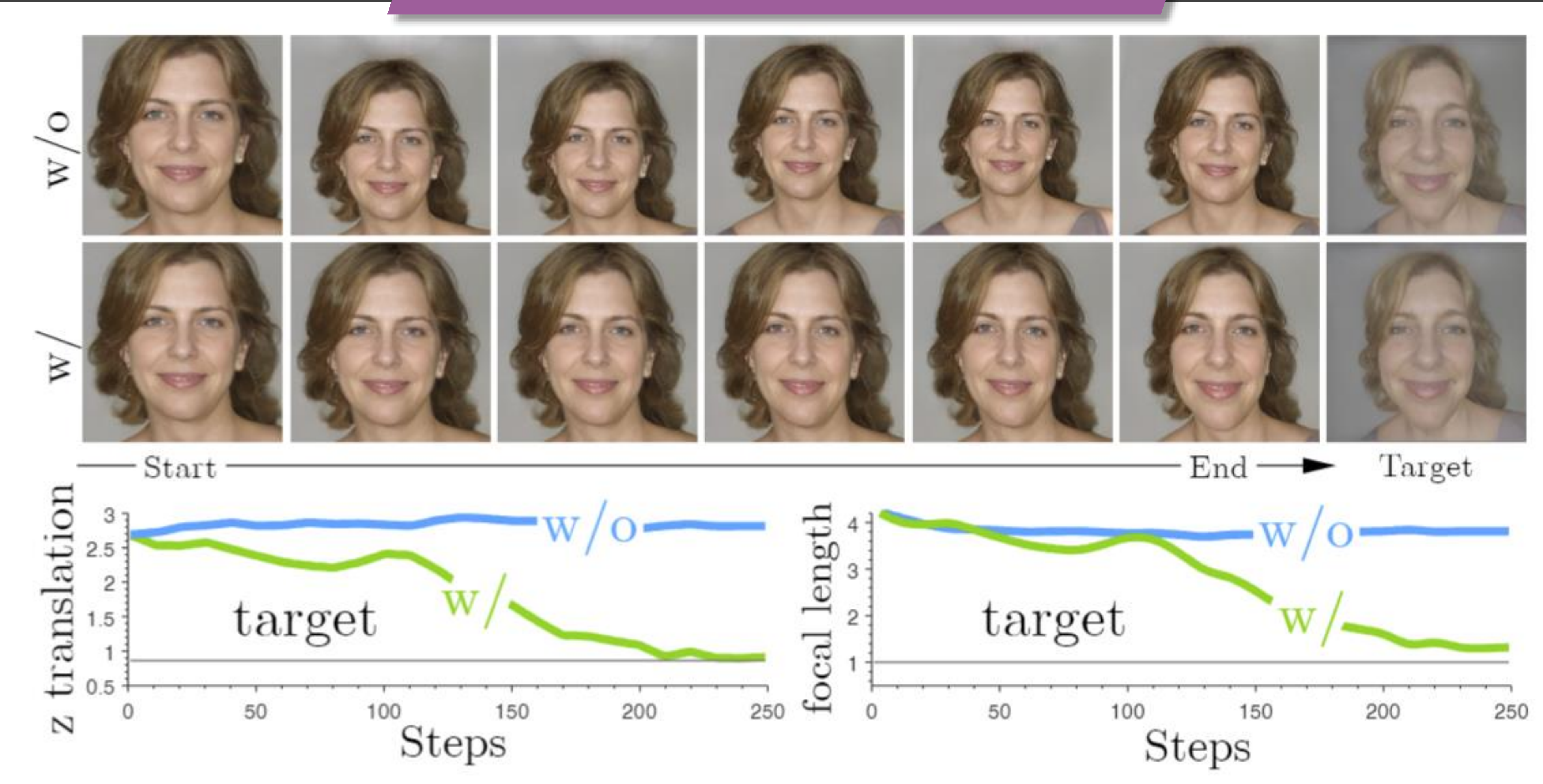
Learnable para.

$$f = \alpha \boxed{\frac{d}{d_0} f_0}$$

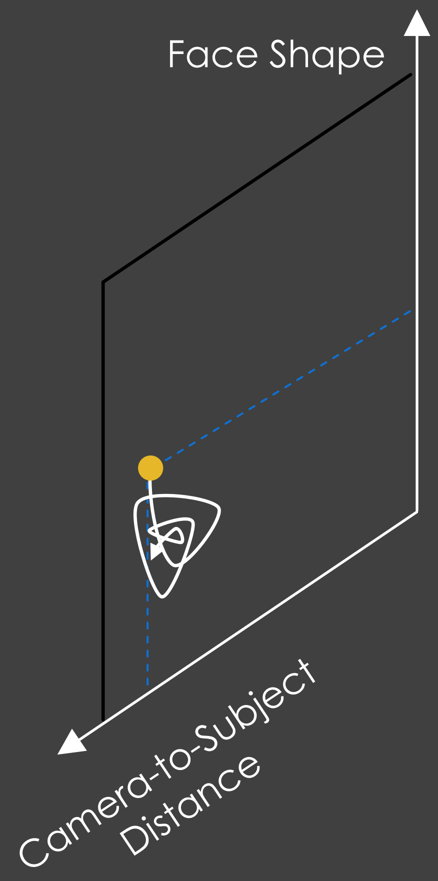
Approximation

Perspective-aware 3D GAN Inversion

Focal Length Re-parameterization



Challenge 2: Different Convergences



Face is **easier** to fall into **sub-optimum** when camera is incorrect

Face parameter $\in \mathrm{R}^{512\times 14}$

Camera parameter $\in \mathrm{R}^{K}, K \ll 512$

input

optimize face with incorrect cam

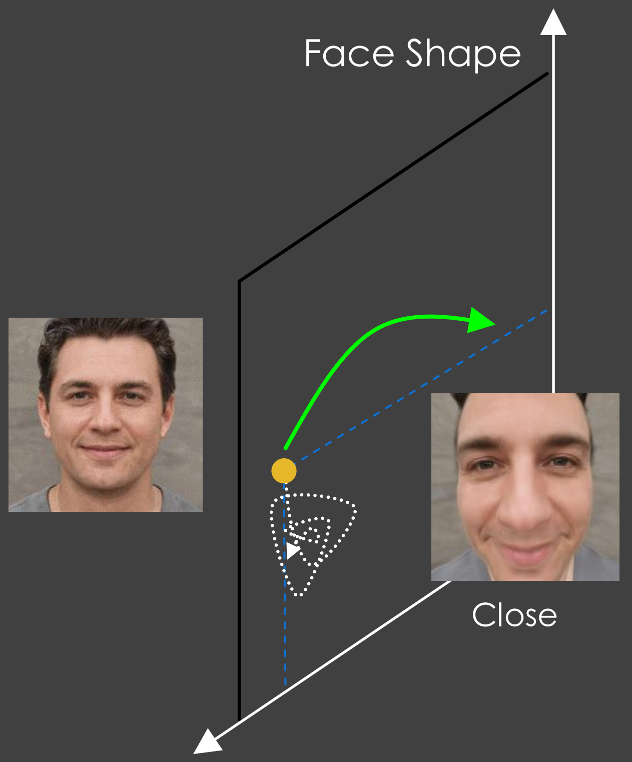
jointly optimize face and cam

reference

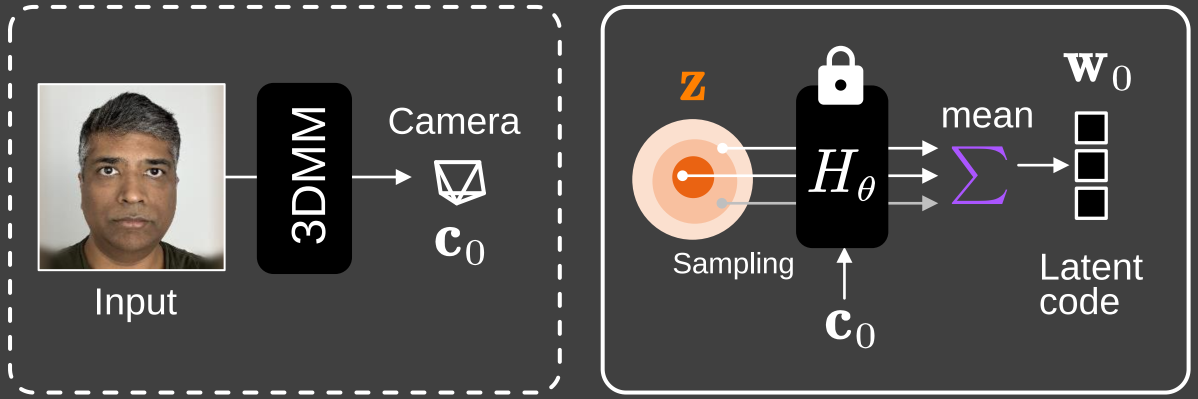
After reprojection

Perspective-aware 3D GAN Inversion

Better initialization

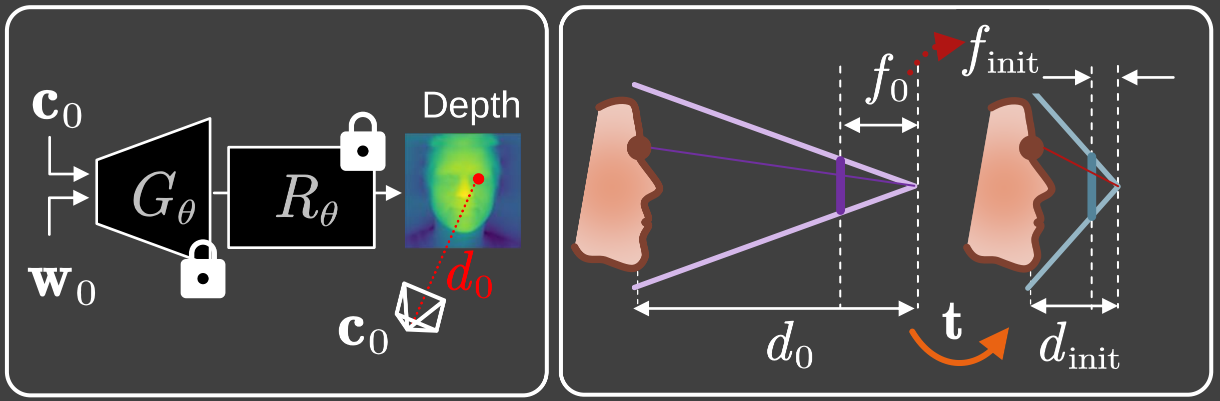


Original initialization



Get init. cam $\mathbf{c}_0$

Get init. face para. $\mathbf{w}_0$

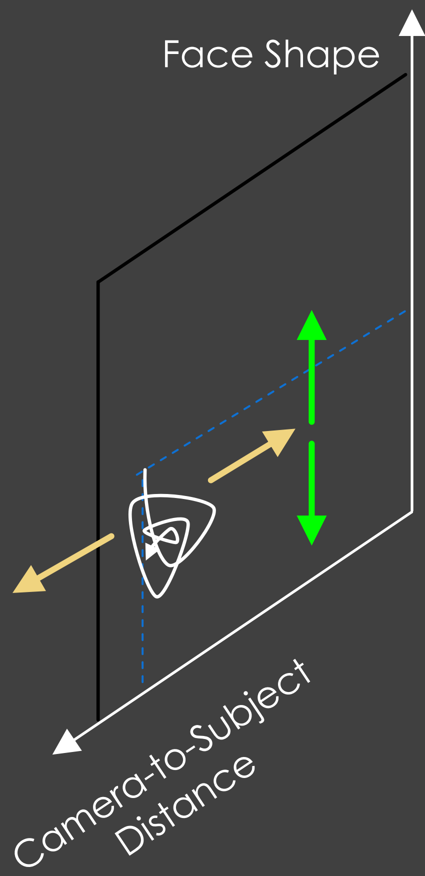


Get eyes' distance d_0

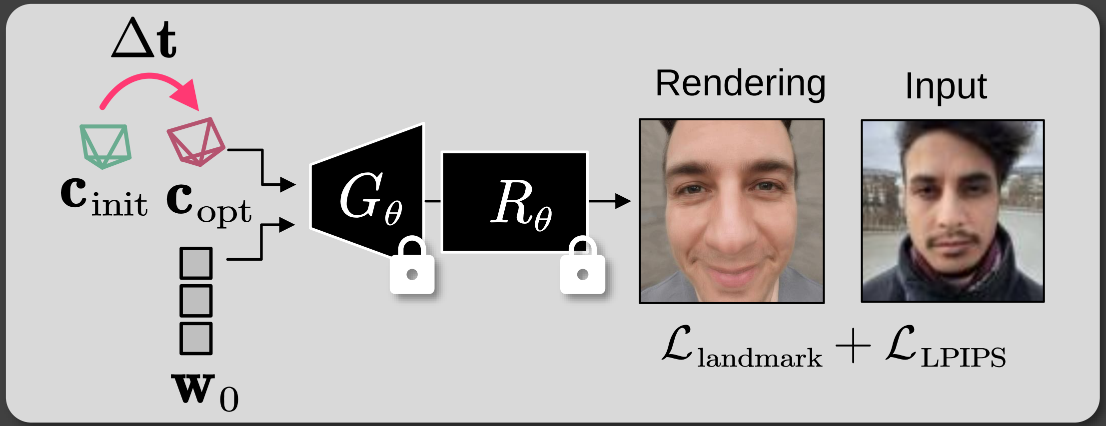
Modify $\mathbf{c}_0 \rightarrow \mathbf{c}_{\text{init}}$

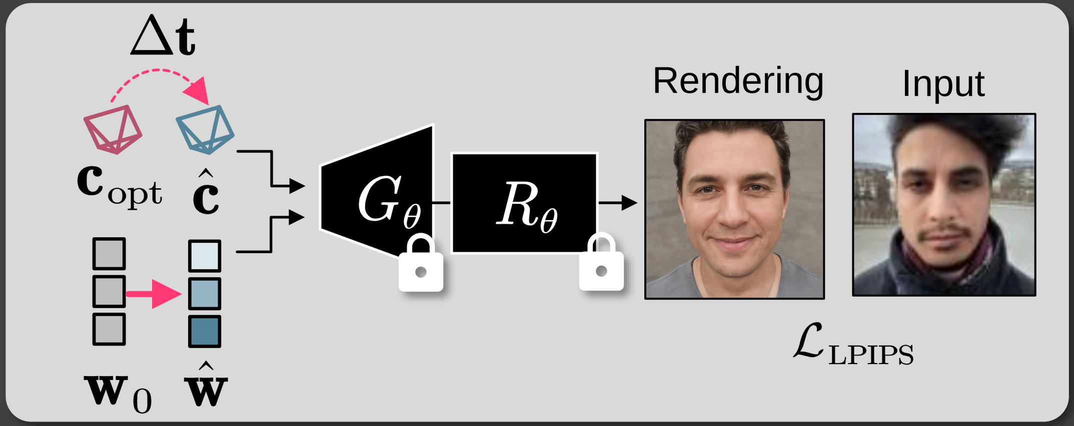
Perspective-aware 3D GAN Inversion

Optimization Scheduling



Optimize camera with mean face

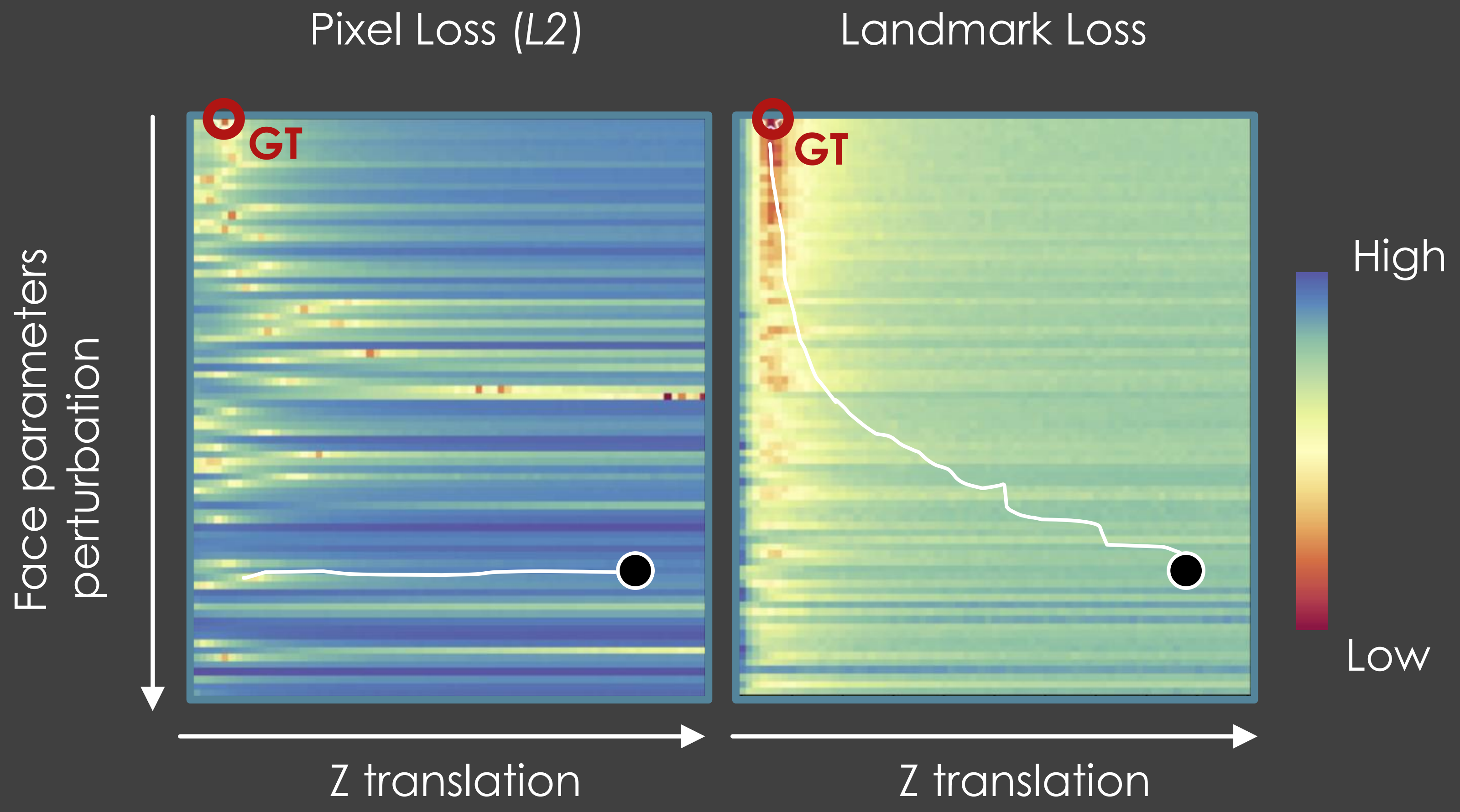




Optimize face

Challenge 3: Ambiguity from Loss

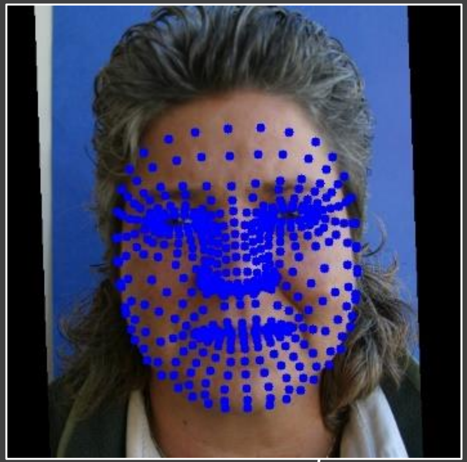
Pixel loss is **less effective** for perspective changing



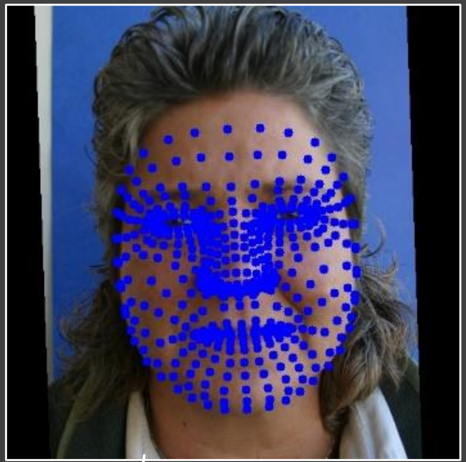
Perspective-aware 3D GAN Inversion

Uncertainty-based Geometric Loss

prediction



input



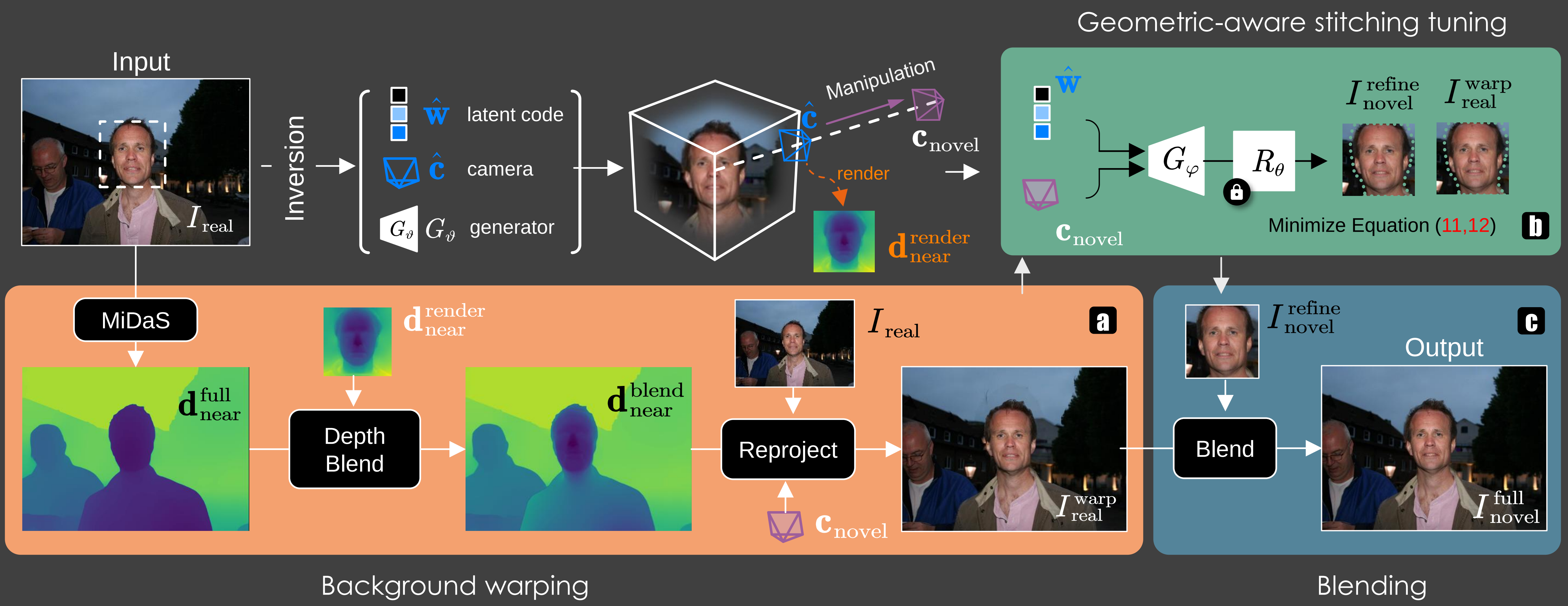
m_i
Landmark

m_i'
Landmark

Learnable Parameter

$$\sum_{i=1}^{\|\mathcal{M}\|}\left(\underbrace{\log\left(\sigma_i^2\right)}_{\textbf{Uncertainty}\text{ term}}+\frac{\|m_i-m_i'\|_2^2}{2\sigma_i^2}\right)$$

Full-frame Processing System



Results – Mesh

Distorted Input

HFGI3D

TriplaneNet

Ours

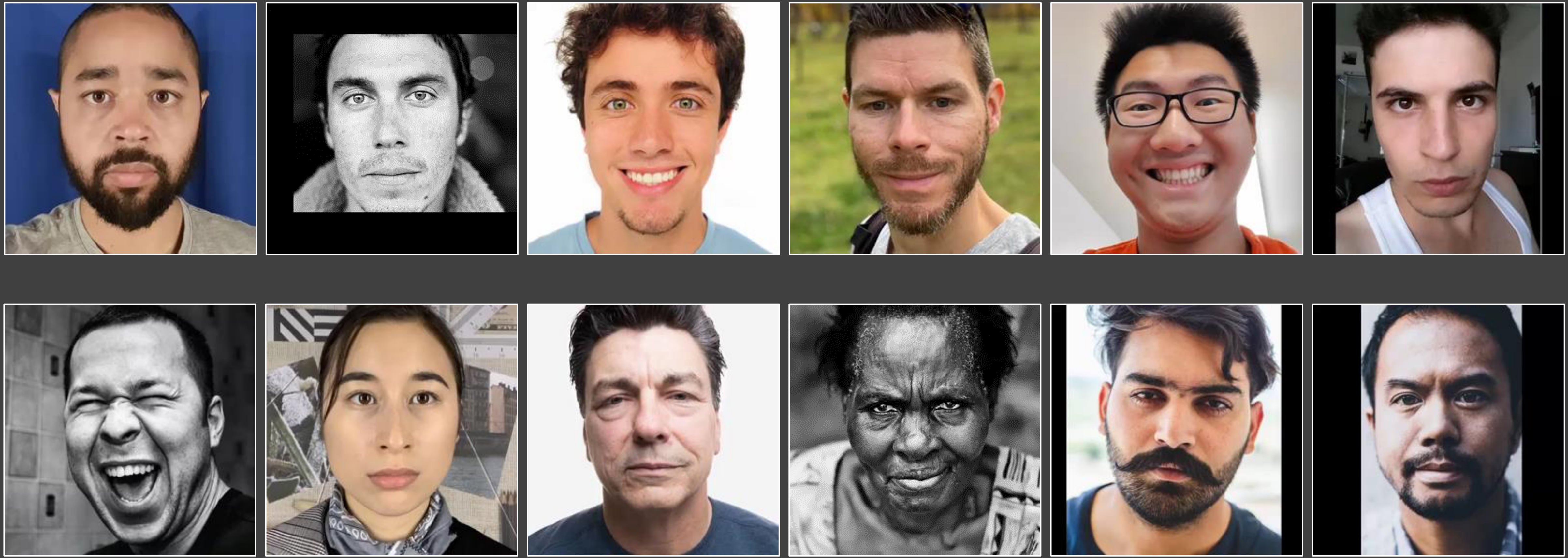
Other GAN inversion methods

Results

Input

Output

Results – Continuous Manipulation



Evaluation – Setup

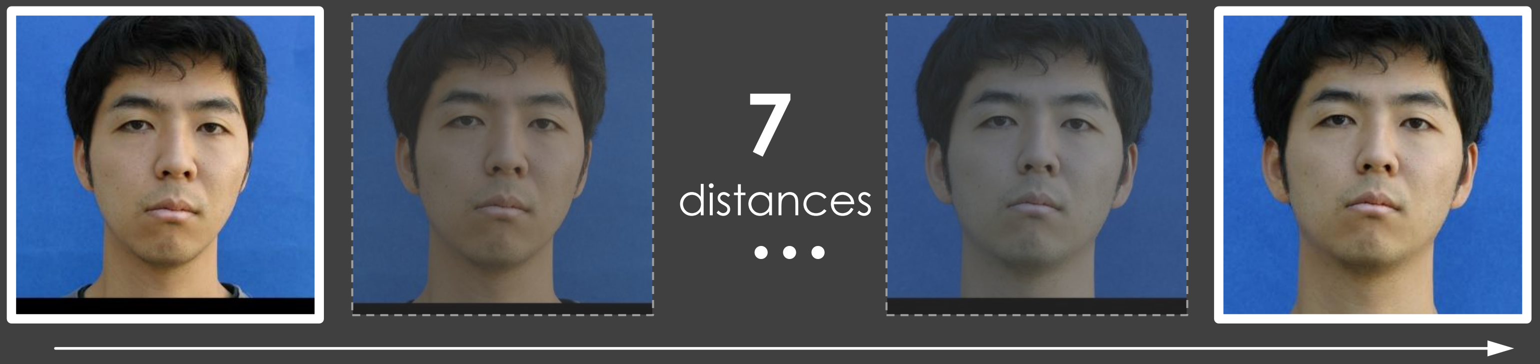
0

image used for training

Quantitative Evaluation

Input

Reference



60 cm

90 cm

360 cm

480 cm

CMDP **51** faces

Qualitative Evaluation

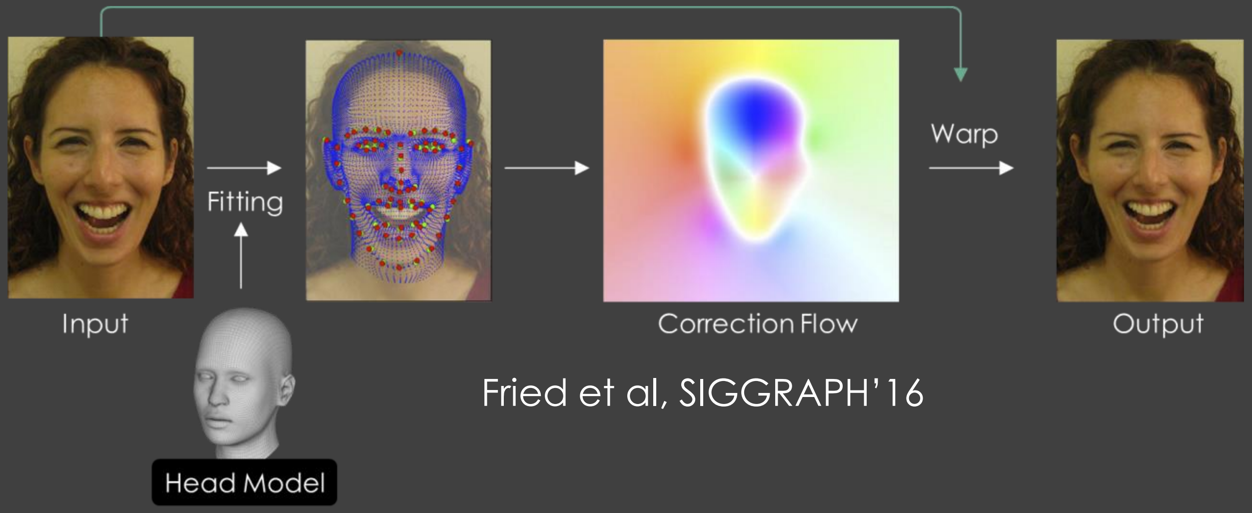
<< 60 cm, **severe** distortion

In-the-wild data

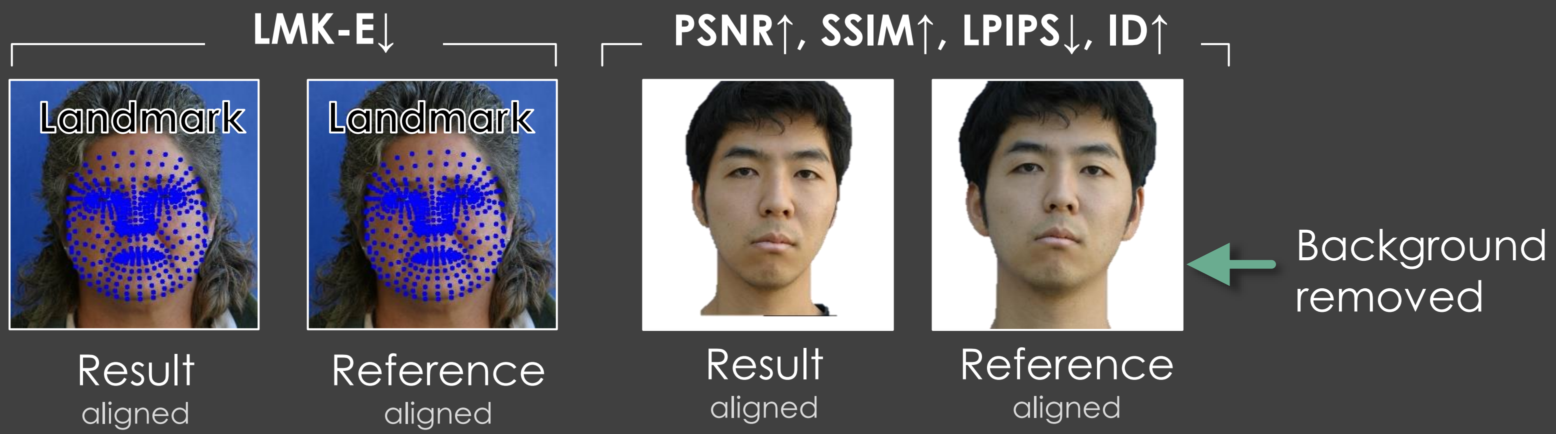
Burgos-Artizzu et al, CMDP, 2014

Evaluation – Setup

- Competed methods
  - **Warping-based**
    - Fried et al, SIGGRAPH'16
    - ~~Zhao et al, ICCV'19~~
      - No code, no training data
- Metrics
  - Landmark error (**LMK-E↓**)
  - **PSNR↑**
  - **SSIM↑**
  - **LPIPS↓**
  - Identity (**ID↑**)



Fried et al, SIGGRAPH'16



Evaluation – Results

	Method	LMK-E↓	PSNR↑	SSIM↑	LPIPS↓	ID↑
prior	Fried's	0.175	15.41	0.724	0.188	**0.893**
	*Fried's	0.165	14.41	0.716	0.208	0.860
ML + prior	Ours	**0.138**	**17.52**	**0.747**	**0.167**	0.859

Ours achieves **highest score** for most metrics

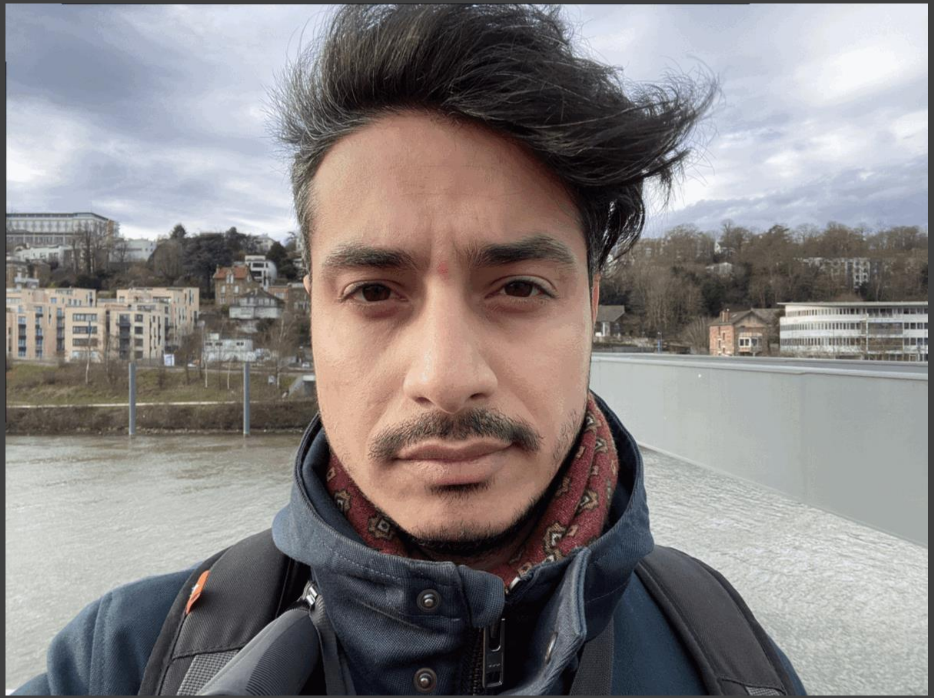
Fried et al, SIGGRAPH'16

*: Our implemented

Evaluation – Results

Stretch-like

3D **geometric consistent**



Fried et al, SIGGRAPH'16

Ours

Evaluation – Results

Input

Fried et al, SIGGRAPH'16

Ours

Evaluation – Results

Input

Fried et al, SIGGRAPH'16

Ours

Evaluation – Results

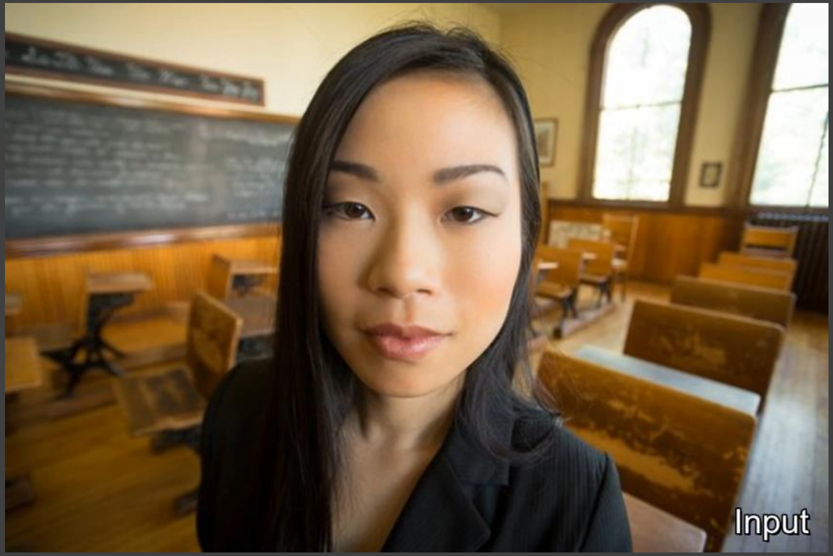
Input

Fried et al, SIGGRAPH'16

Ours

Dolly Zoom

Dolly Zoom



Summary – Portrait Distortion Correction

- We introduce the pre-trained 3D face GAN as a real-world prior
- We change the warping-based learning structure into rendering-like
- We develop strategies to reduce optimization ambiguity
- We develop a real-world system for full-frame images

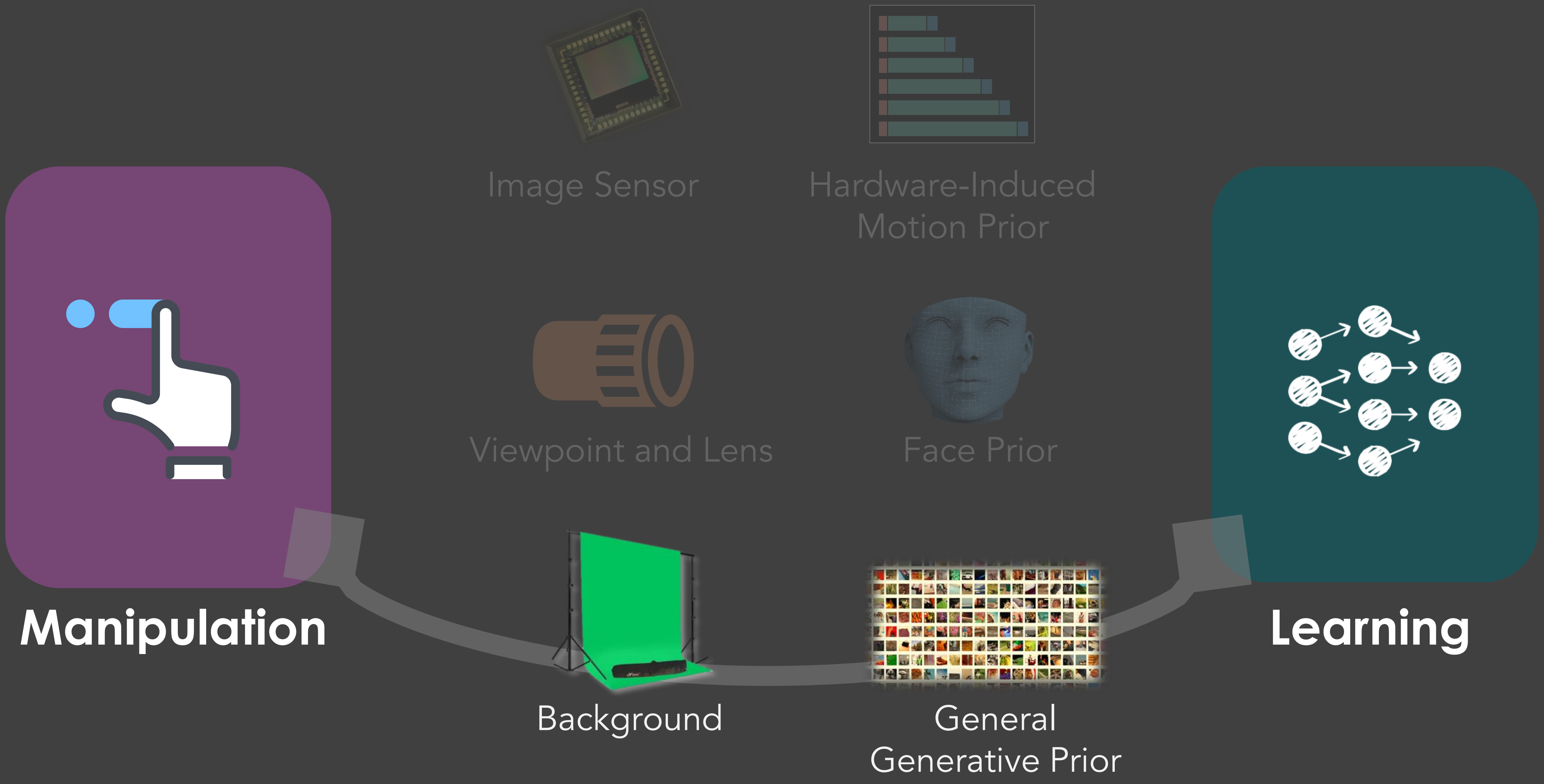
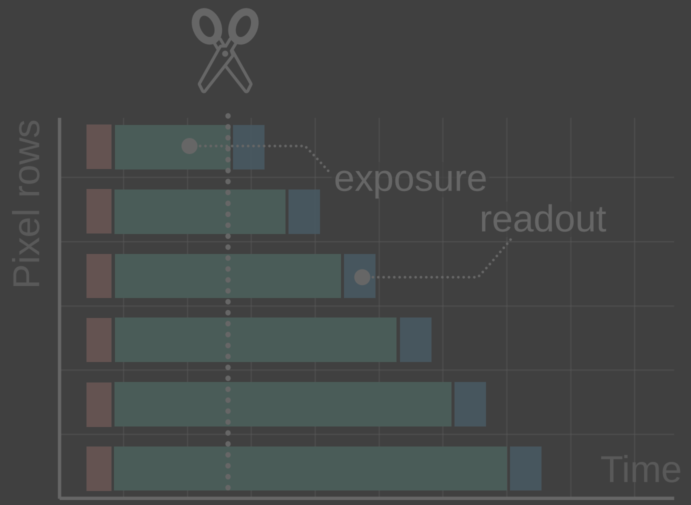


Image Sensor
Hardware-Induced Motion Prior
Viewpoint and Lens
Face Prior
Manipulation
Learning
Background
General Generative Prior

Image Sensor



Neural Global Shutter

Wang et al, CVPR 2022

Viewpoint + Lens

Portrait Distortion Correction

Wang et al, IJCV 2024

Background

Matting by Generation

Wang et al, SIGGRAPH 2024

New Learning Structures

Regression → Generation

AAAI'22	MODNet
MM'22	P3M
IJCV'23	ViTAE-S

Ours

Do You Want the Background?

Image credit: [url]

Manipulate Background

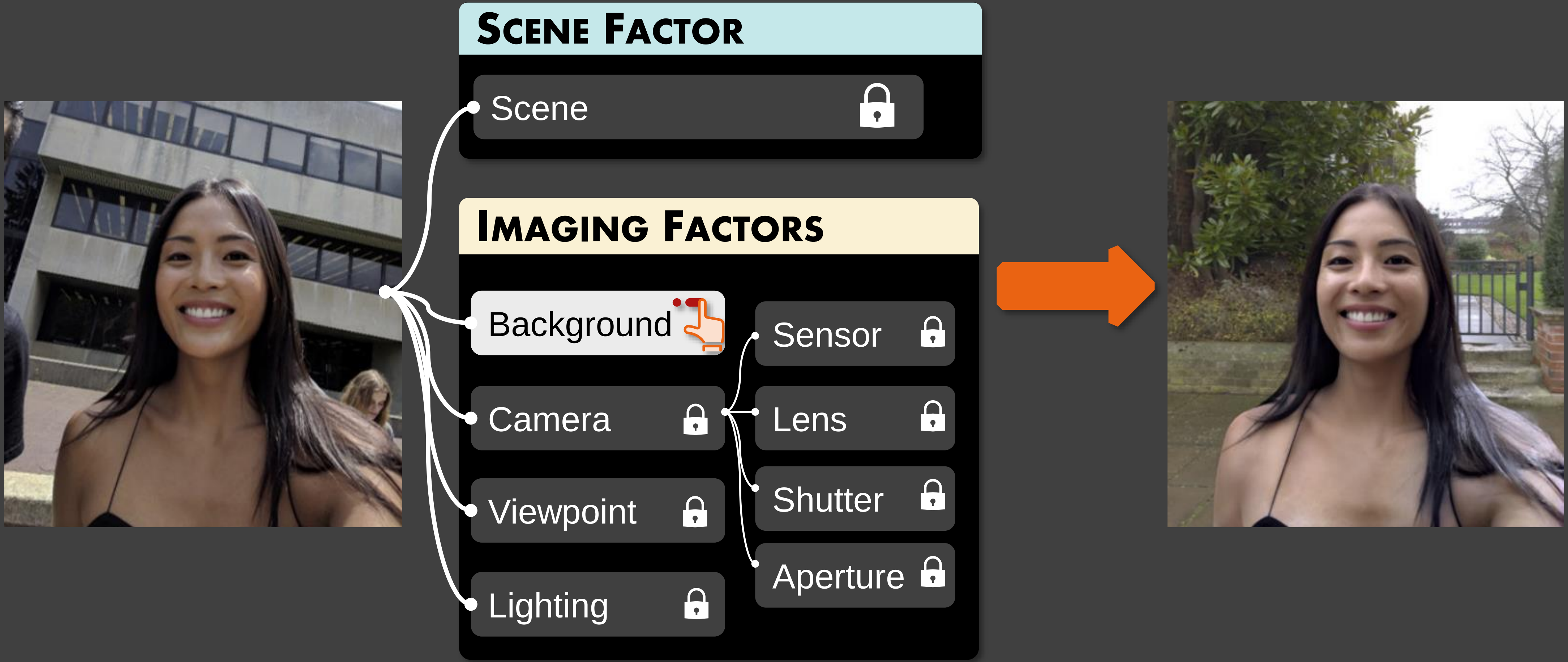


Image credit: total relighting

Composition Equation

Foreground

Background

$$C = \alpha F + (1 - \alpha)B$$

Slides credit: Yung-Yu Chuang

Image Matting

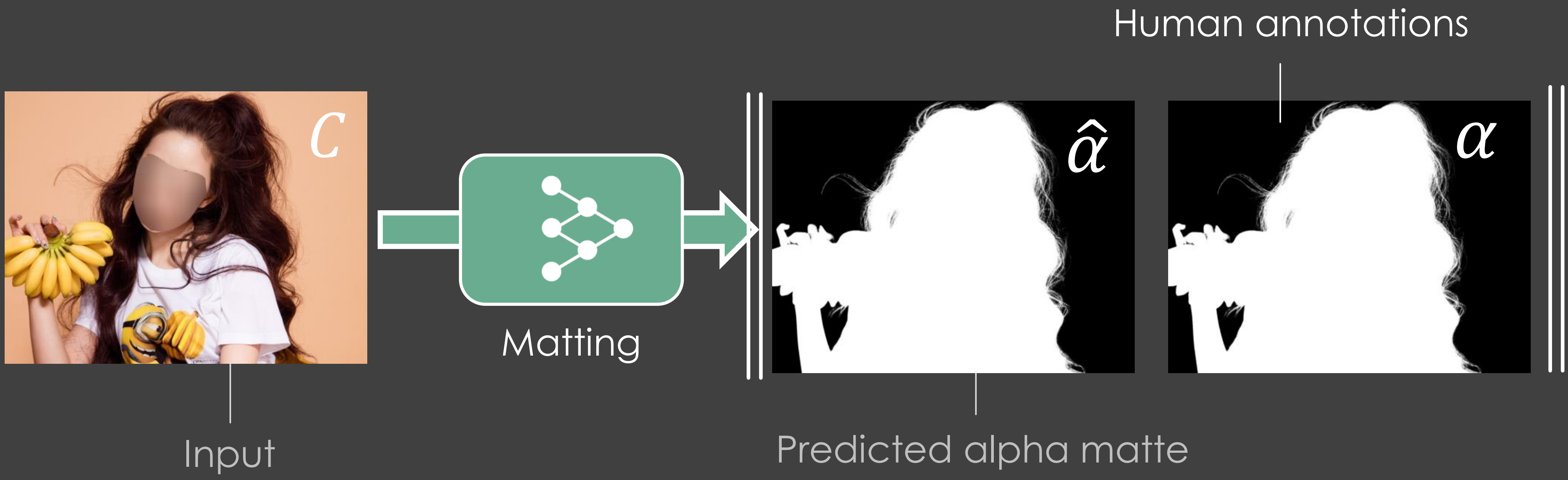
Foreground

Background

$$C = \alpha F + (1 - \alpha)B$$

Slides credit: Yung-Yu Chuang

Existing Method – Regression-based



Ke et al, MODNet, AAAI'22
Li et al, P3M, MM'22
Ma et al, ViTAE-S, IJCV'23

Limitations of Existing Methods

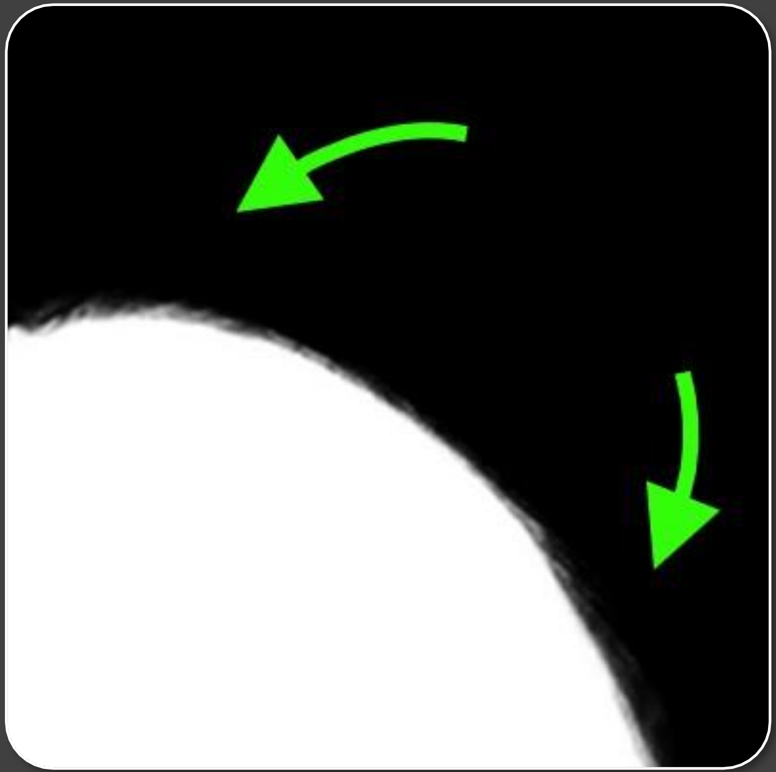
Ke et al, MODNet, AAAI'22

Limitations of Existing Methods

Annotation is **challenging**

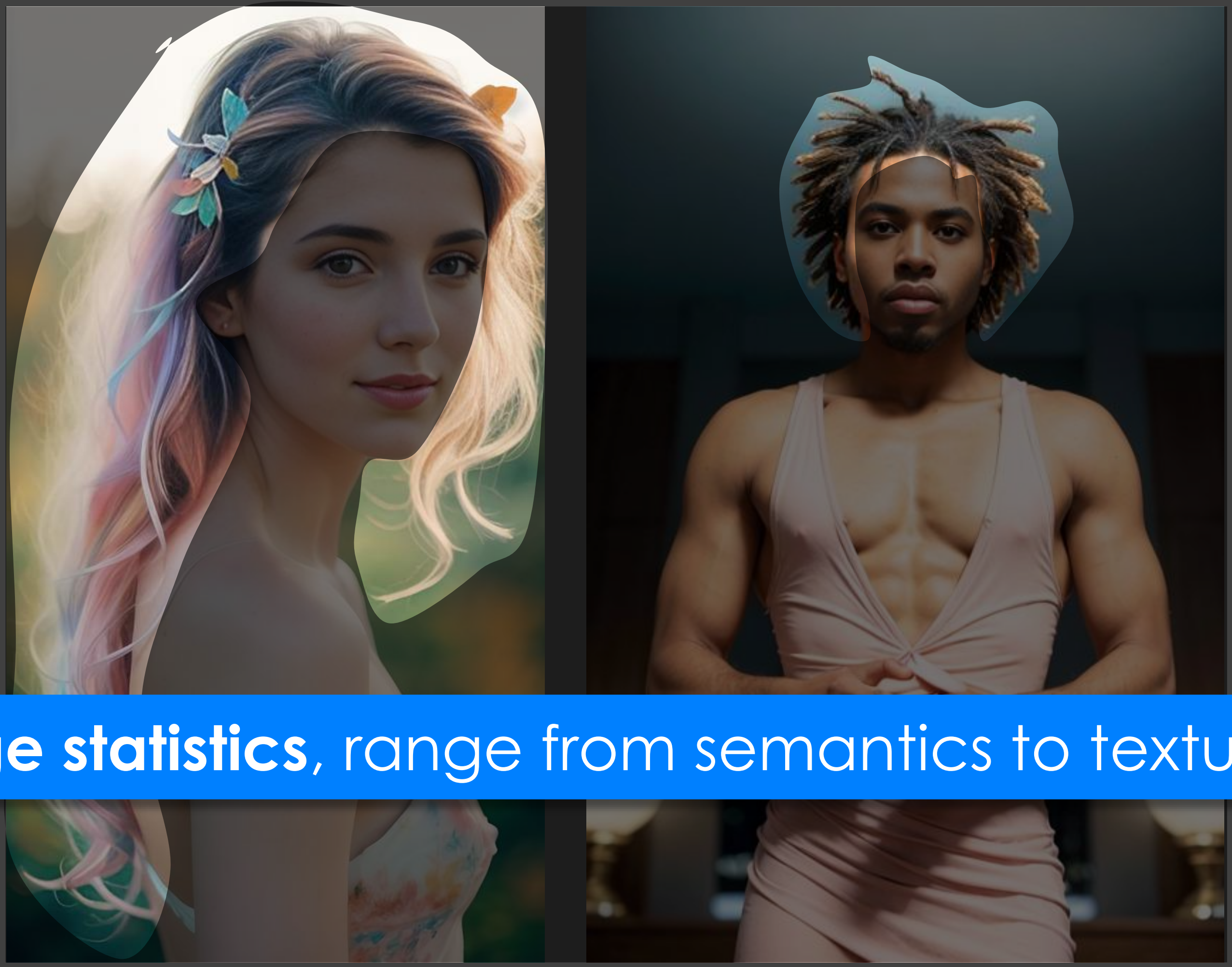
Poor label quality

Input image



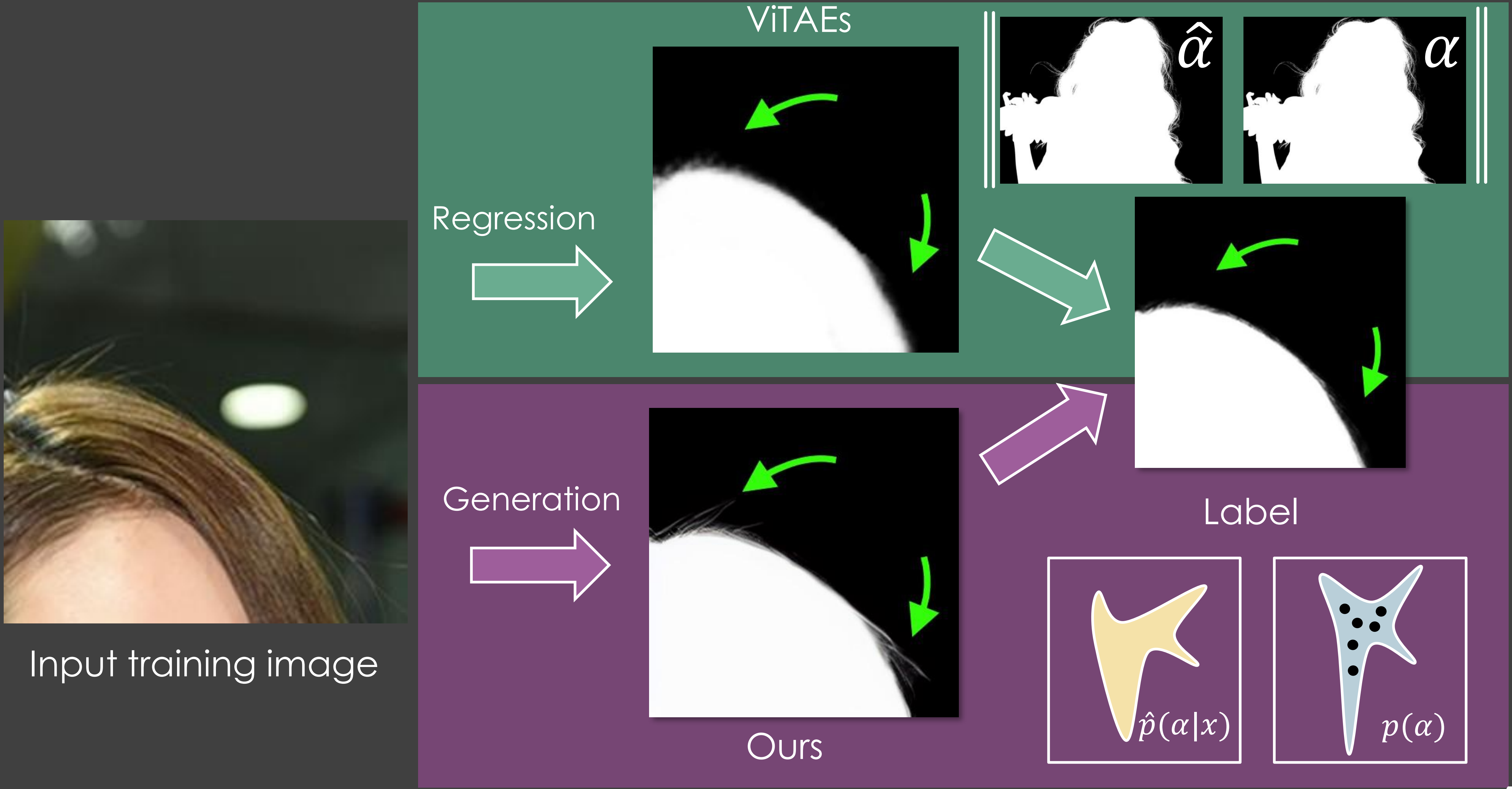
Label

Pre-trained Diffusion Models as Prior



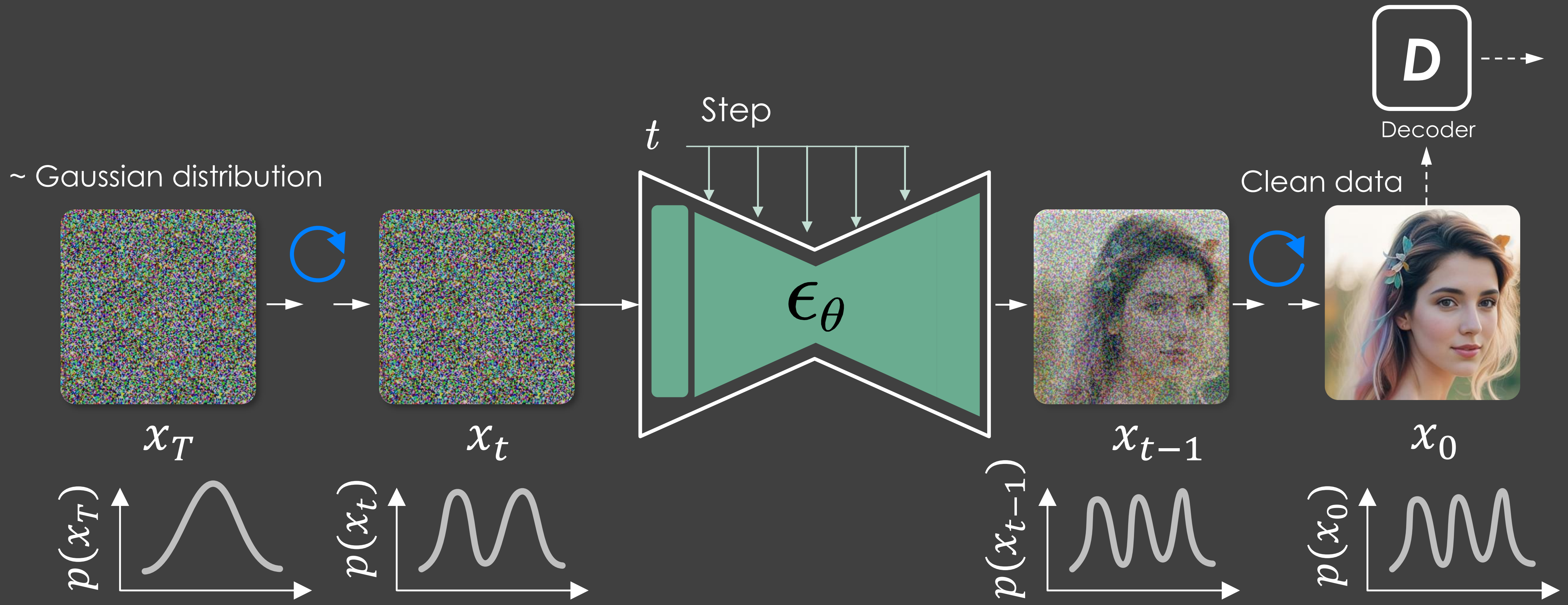
Rich image statistics, range from semantics to texture details

New Learning Structure

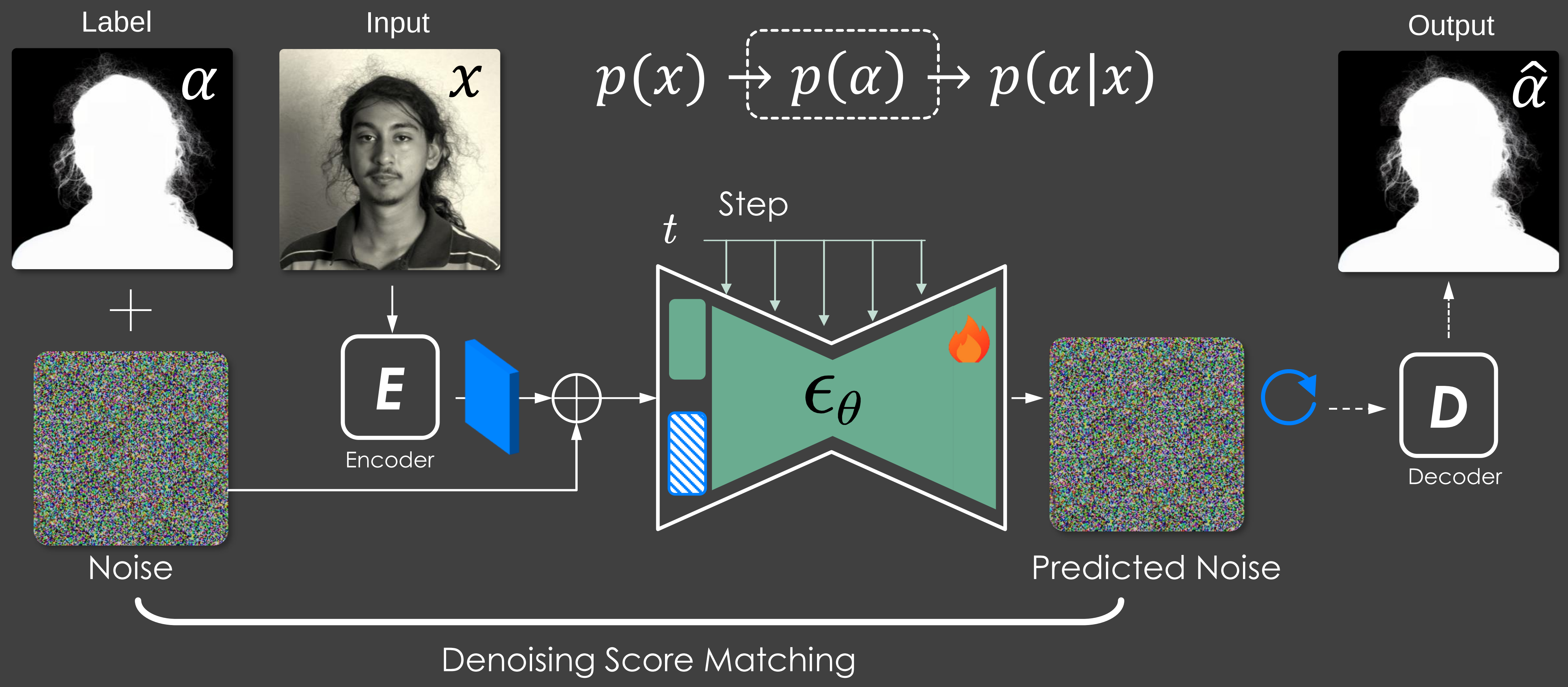


Repurposing Latent Diffusion Model

Latent diffusion model models $p(x_0)$



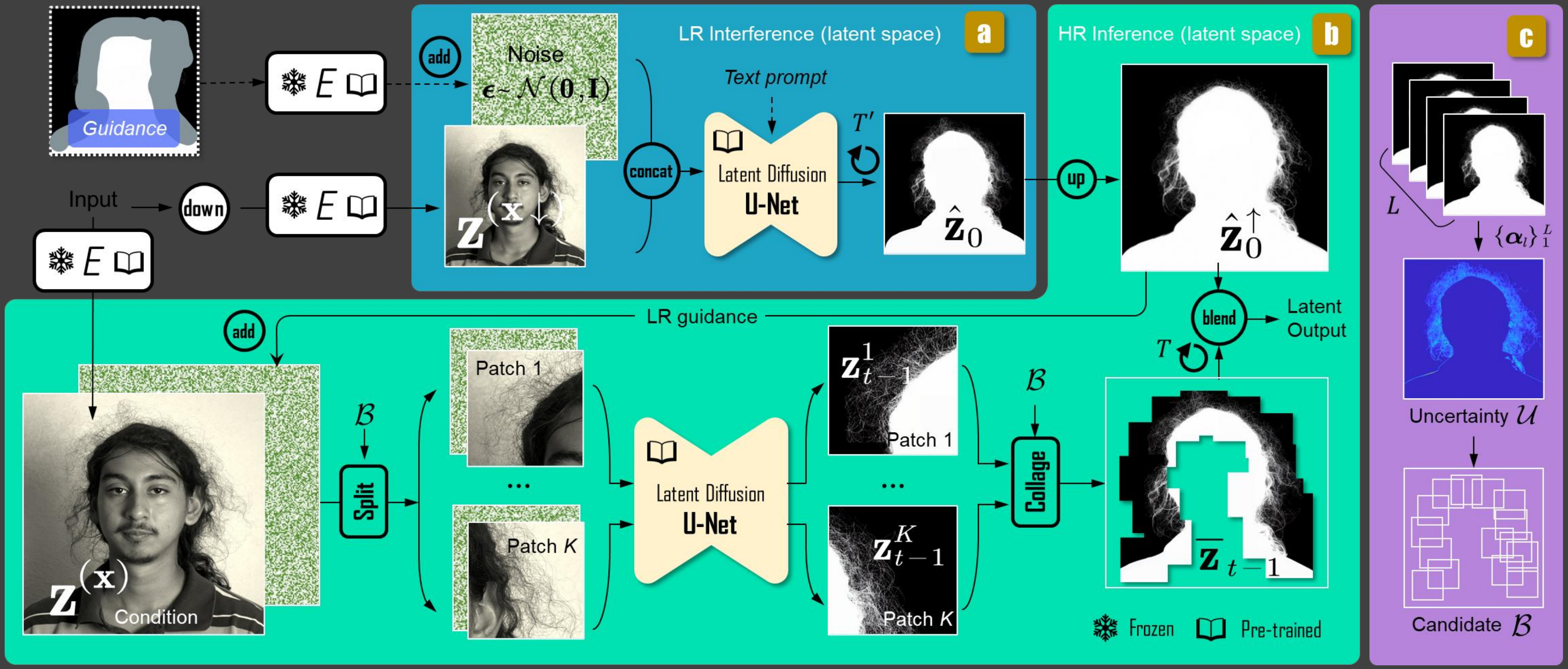
Repurposing Latent Diffusion Model



System for Real-world Applications

w/ or w/o users' hint

(a-c) Efficient Inference for high-resolution images



Evaluation – Setup

P3M-10K

9,421 images
Corse-grained labels

Training

PPM-100

100 images

Quantitative Evaluation

Qualitative Evaluation

RVP

636 images

Qualitative Evaluation

Evaluation – Setup

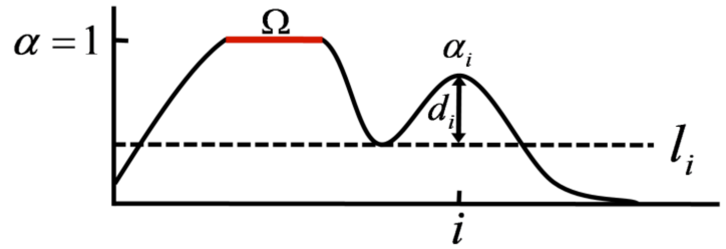
- Competed methods
  - Regression-based
    - MODNet
    - P3M
    - ViTAE-S
- Metrics
  - Mean Squared Error: **MSE↓**
  - Mean Absolute Difference: **MAD↓**
  - Sum of Absolute Differences: **SAD↓**
  - Connectivity: **Conn↓**

$$\sum_i \left(\varphi(\alpha_i, \Omega) - \varphi(\alpha_i^*, \Omega)\right)^p$$

Degree of Connectivity

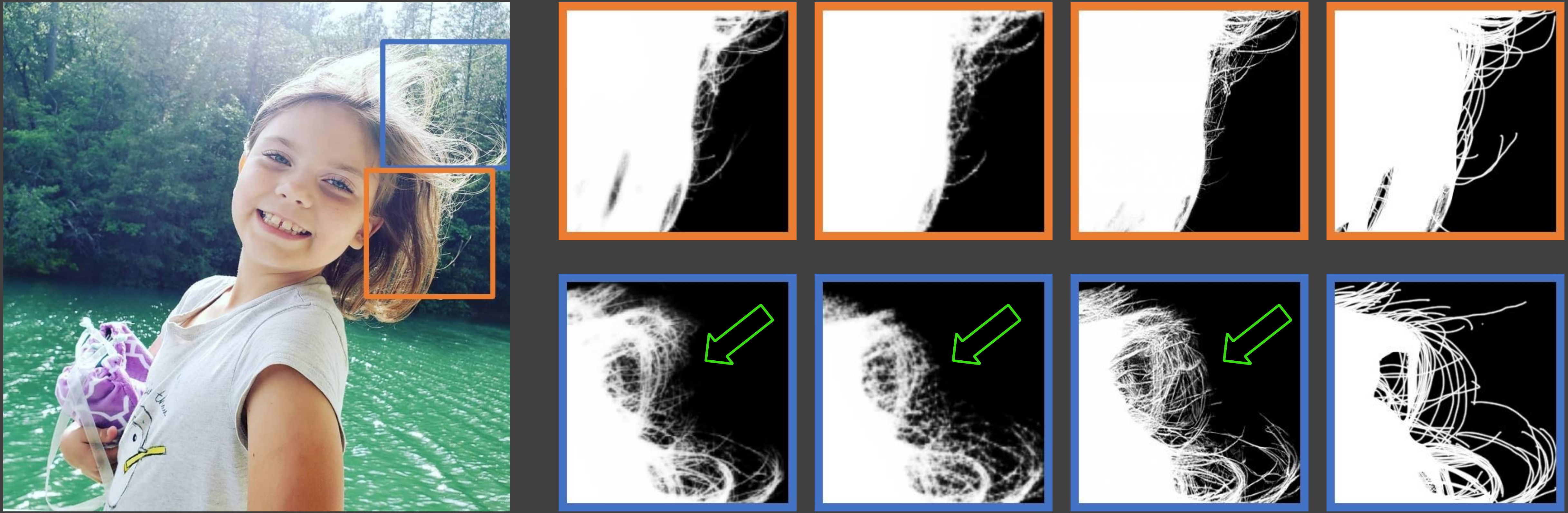
$$\varphi\left(\alpha_i, \Omega\right) = 1 - \left(\lambda_i \cdot \delta\left(d_i \geq \theta\right) \cdot d_i\right).$$

$$d_i = \alpha_i - l_i \qquad \lambda_i = \frac{1}{|K|} \sum_{k \in K} dist_k(i)$$



Ke et al, MODNet, AAAI’22
Li et al, P3M, MM’22
Ma et al, ViTAE-S, IJCV’23

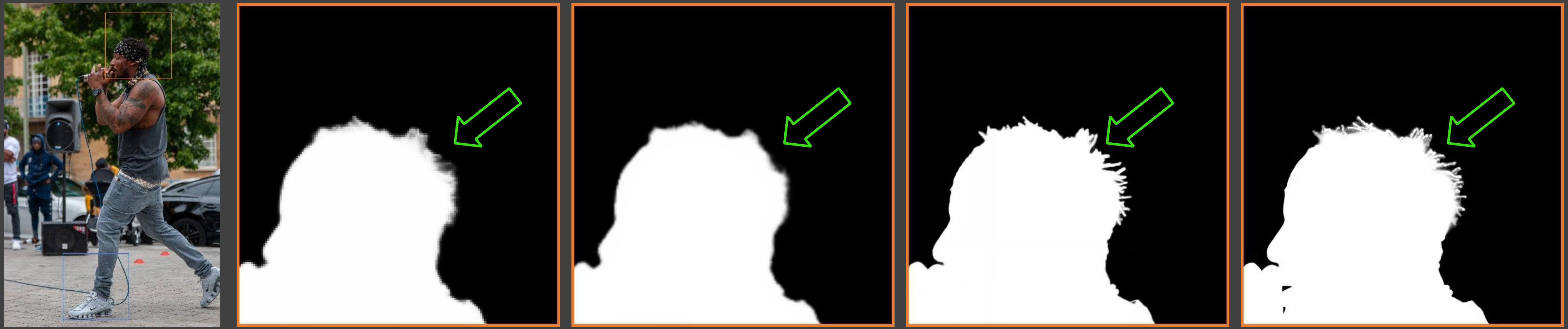
Evaluation – Results



MODNet | ViTAE-S | Ours | Reference

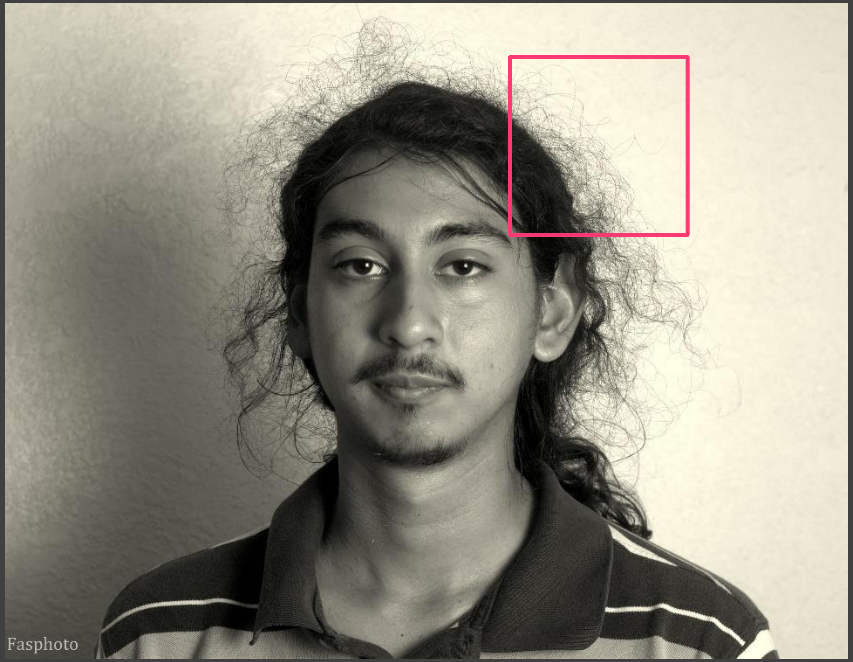
Ke et al, MODNet, AAAI’22
Li et al, P3M, MM’22
Ma et al, ViTAE-S, IJCV’23

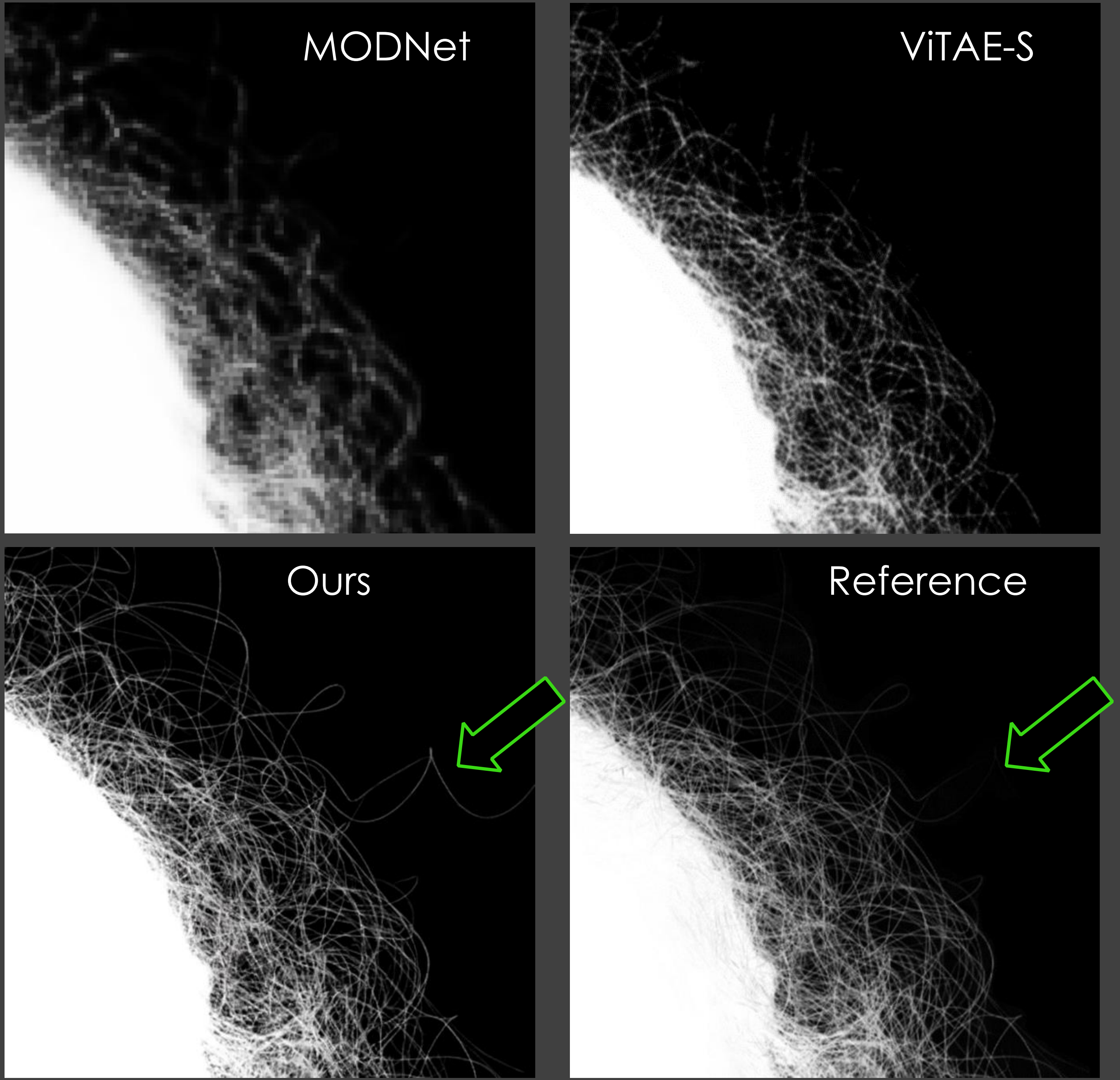
Evaluation – Results



MODNet ViTAE-S Ours Reference

Ke et al, MODNet, AAAI'22
Li et al, P3M, MM'22
Ma et al, ViTAE-S, IJCV'23





Ke et al, MODNet, AAAI'22
Li et al, P3M, MM'22
Ma et al, ViTAE-S, IJCV'23

Evaluation – Results

Method	MSE↓	MAD↓	SAD↓	Conn↓
MODNet	4.5	10.1	96.0	81.1
P3M	5.8	9.6	93.3	96.1
ViTAE-S	3.4	6.5	62.6	59.3
Ours	**2.5**	**6.3**	**56.9**	**54.0**

Ours achieves **highest score** for all metrics

Ke et al, MODNet, AAAI'22
Li et al, P3M, MM'22
Ma et al, ViTAE-S, IJCV'23

DiffMat

Ours

Human Annotation


Input
Ours
Human Annotation

Input

Ours

Human Annotation

ViTAE-S

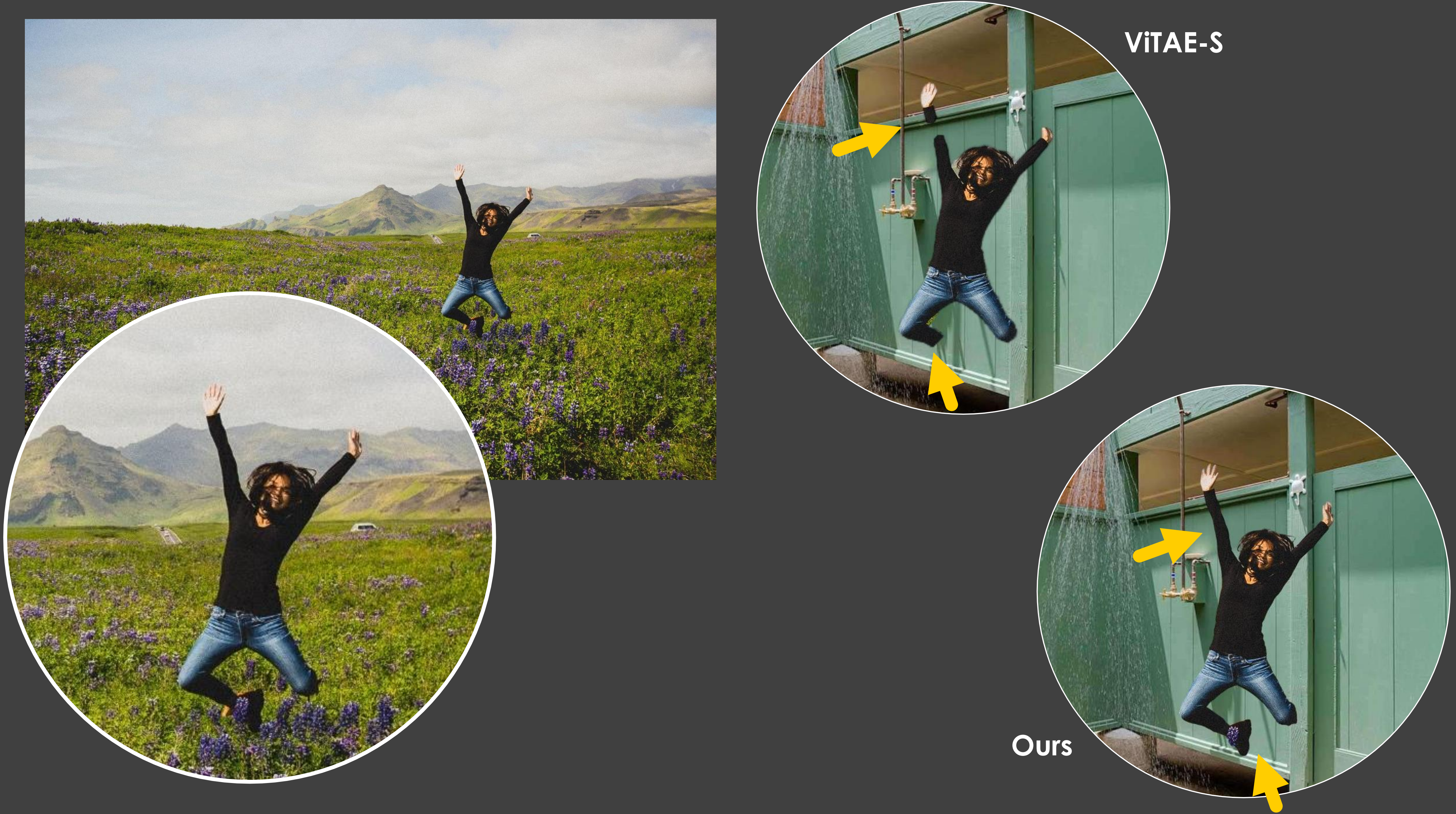
Ours

Input

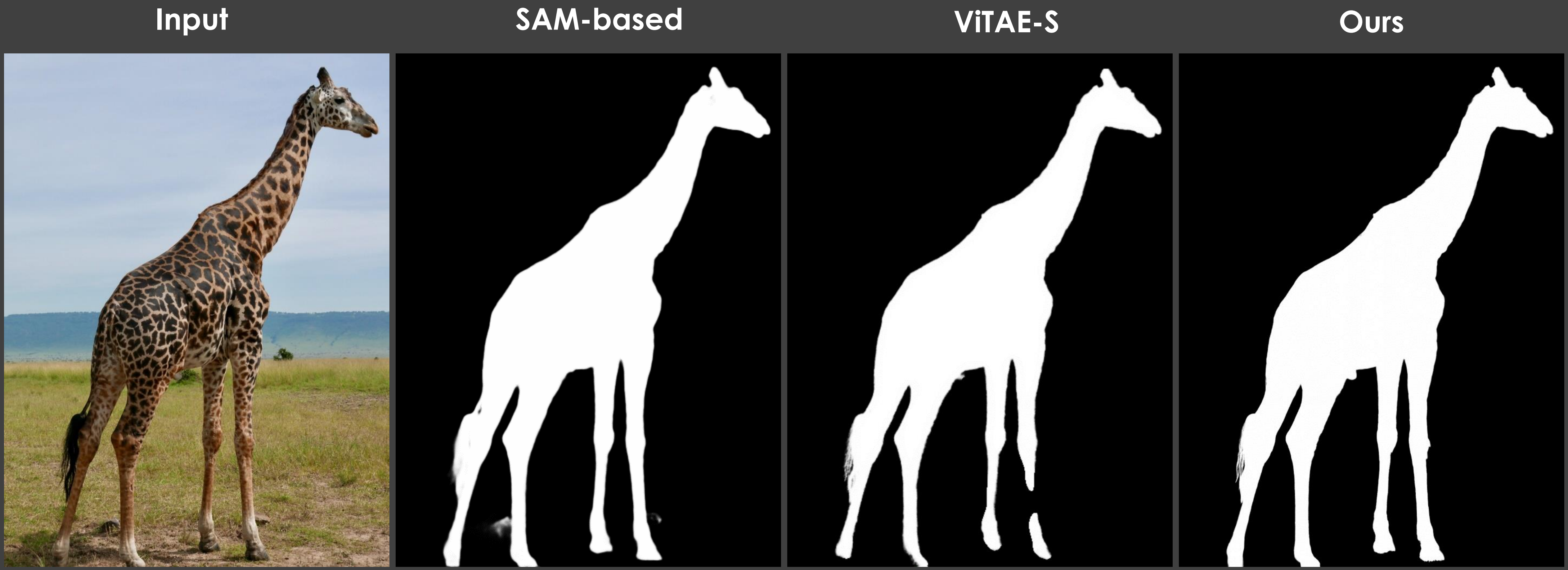
ViTAE-S

Ours


Input
PUSH
BUTTON
FOR
ViTAE-S
Ours


ViTAE-S
Ours

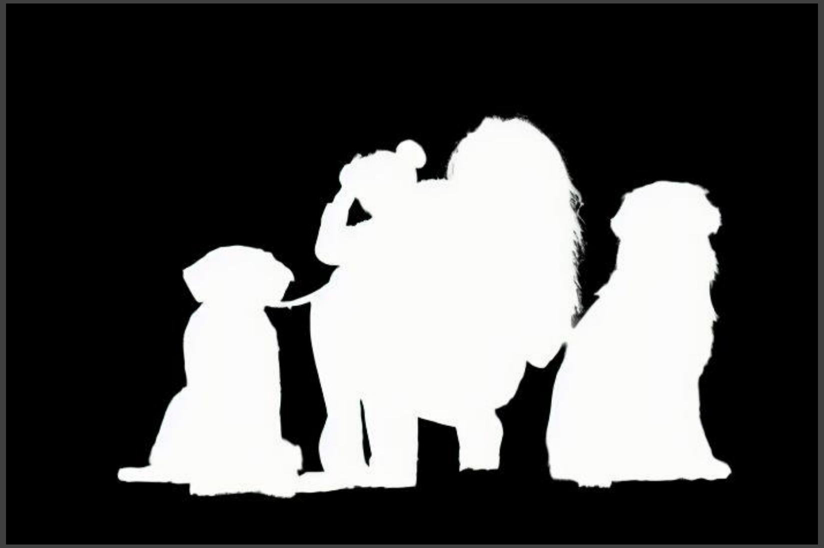
Out-of-Distribution Matting



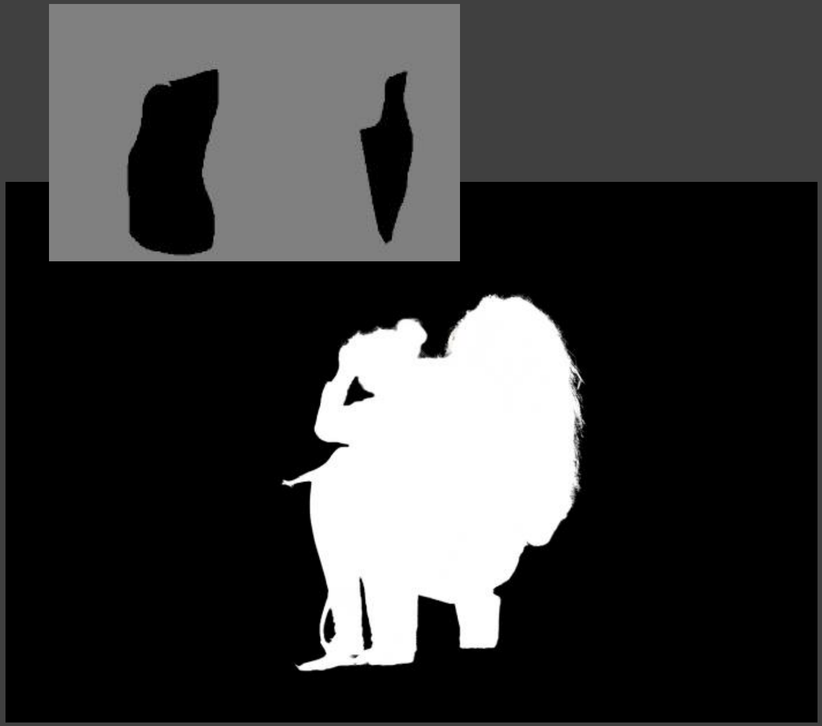
Matting with Additional Guidance

guidance

Input



w/o guidance



w/ guidance

Summary – Matting by Generation

- We introduce pre-trained generative diffusion model as a real-world prior
- With the pre-trained generative model, we convert the regression problem into a conditional generation problem
- We develop a system to efficiently process high-resolution images and leverage users' inputs

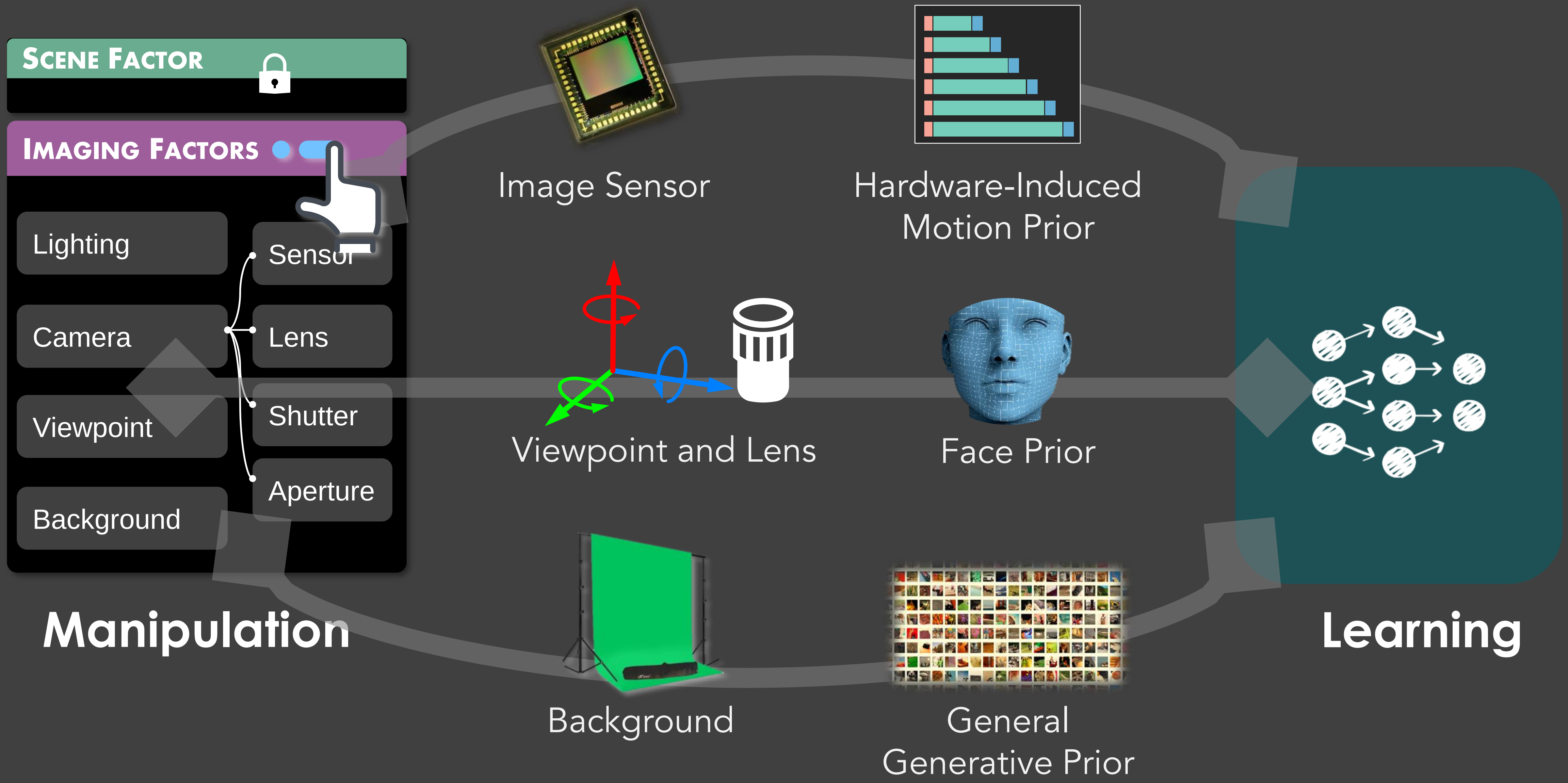
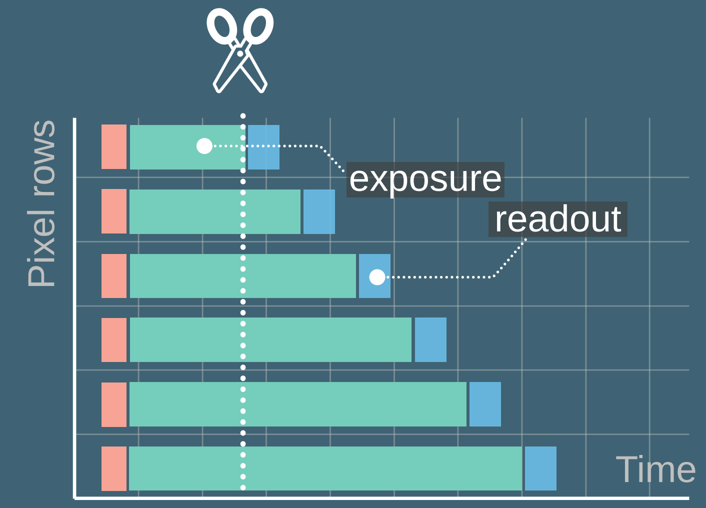

Scene Factor
Imaging Factors
Lighting
Camera
Viewpoint
Background
Sensor
Lens
Shutter
Aperture
Image Sensor
Hardware-Induced Motion Prior
Viewpoint and Lens
Face Prior
Background
General Generative Prior
Manipulation
Learning

Image Sensor



Neural Global Shutter

Wang et al, CVPR 2022

Viewpoint + Lens

Portrait Distortion Correction

Wang et al, IJCV 2024

Background

Matting by Generation

Wang et al, SIGGRAPH 2024

New Learning Structures

Warping → Deblurring

Warping → Rendering

Regression → Generation

Contributions

- We propose to combine ML learning approaches with real-world prior for imaging factor manipulation
- We introduce three new real-world priors and change the learning structure in the conventional problems
- These new learning methods show significant advantages
- We propose systems to make the new approaches work for real-world applications

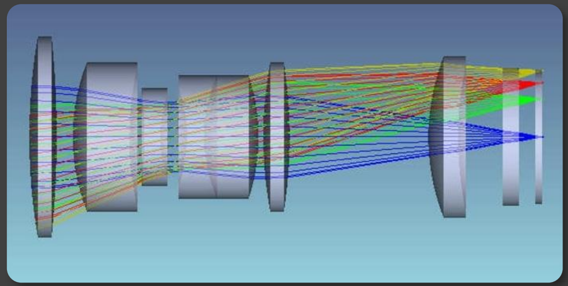
Research Goal:
Towards Model the Physical World

- Capture → Recreate → Re-render

Applications

Environment for Agent Learning

and for Camera Design



Digital twins

Cinematic Filming

Limitations and Future Work

An Unified Model

- Manipulating one factor need one system
    - Future: an all-in-one system that takes arbitrary priors
- Generative prior and hardware induced prior are separately used
    - Future: combination of both

Explore Additional Factors

- Image Quality
- Lighting

Explore Other Priors

- Generative Video Models
- Physical Principle

Publications

Included in this thesis

1. Zhixiang Wang, Xiang Ji, Jia-Bin Huang, Shin'ichi Satoh, Xiao Zhou, and Yinqiang Zheng, Neural Global Shutter: Learn to Restore Video from a Rolling Shutter Camera with Global Reset Feature, CVPR, 2022

2. Zhixiang Wang, Yu-Lun Liu, Jia-Bin Huang, Shin'ichi Satoh, Sizhuo Ma, Guru Krishnan, and Jian Wang, DisCO: Portrait Distortion Correction with Perspective-Aware 3D GANs, IJCV, 2024

3. Zhixiang Wang, Baiang Li, Jian Wang, Yu-Lun Liu, Jinwei Gu, Yung-Yu Chuang, and Shin'ichi Satoh, Matting by Generation, SIGGRAPH, 2024

Publications

Other papers

1. Xianzheng Ma, Zhixiang Wang, Yacheng Zhan, Yinqiang Zheng, Zheng Wang, Dengxin Dai, and Chia-Wen Lin, Both Style and Fog Matter: Cumulative Domain Adaptation for Semantic Foggy Scene Understanding, CVPR, 2022, Oral
2. Xiang Ji, Zhixiang Wang, Shin'ichi Satoh, and Yinqiang Zheng, Single Image Deblurring with Row-dependent Blur Magnitude, ICCV, 2023
3. Xiang Ji, Zhixiang Wang, Zhihang Zhong, and Yinqiang Zheng, Rethinking Video Frame Interpolation from Shutter Mode Induced Degradation, ICCV, 2023
4. Caoyuan Ma, Yu-Lun Liu, Zhixiang Wang, Wu Liu, Xinchen Liu, and Zheng Wang, HumanNeRF-SE: A Simple yet Effective Approach to Animate HumanNeRF with Diverse Poses, CVPR, 2024
5. Hui Wei*, Zhixiang Wang*, Xuemei Jia, Yinqiang Zheng, Hao Tang, Shin'ichi Satoh, and Zheng Wang, HotCold Block: Fooling Thermal Infrared Detectors with a Novel Wearable Design, AAAI, 2023 co-first author

Publications

Other papers

6. Zhijing Wan, Zhixiang Wang, Yuran Wang, Zheng Wang, Hongyuan Zhu, and Shin'ichi Satoh, Contributing Dimension Structure of Deep Feature for Coreset Selection, AAAI, 2024
7. Kejun Lin, Zhixiang Wang^, Zheng Wang^, Yinqiang Zheng, and Shin'ichi Satoh, Beyond Domain Gap: Exploiting Subjectivity in Sketch-Based Person Retrieval, ACM Multimedia, 2023 ^: corresponding author
8. Hui Wei, Hanxun Yu, Kewei Zhang, Zhixiang Wang, Jianke Zhu, and Zheng Wang, Moiré Backdoor Attack (MBA): A Novel Trigger for Pedestrian Detectors in the Physical World, ACM Multimedia, 2023
9. Zhijing Wan, Zhixiang Wang, CheukTing Chung, and Zheng Wang, A Survey of Dataset Refinement for Problems in Computer Vision Datasets, ACM Computing Surveys, 2023
10. Zhijing Wan, Xin Xu, Zheng Wang, Zhixiang Wang, and Ruimin Hu, From Multi-source Virtual to Real: Effective Virtual Data Search for Vehicle Re-Identification, IEEE Transactions on Intelligent Transportation Systems, 2023
11. Hui Wei, Hao Tang, Xuemei Jia, Zhixiang Wang, Hanxun Yu, Zhubo Li, Shin'ichi Satoh, Luc Van Gool, and Zheng Wang, Physical Adversarial Attack Meets Computer Vision: A Decade Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence (Accepted)

Thank you!