VISUAL COMPUTING Lab

KAIST

SIGGRAPH 2021

DEEPFORMABLETAG: END-TO-END GENERATION AND RECOGNITION OF DEFORMABLE FIDUCIAL MARKERS

MUSTAFA B. YALDIZ,
ANDREAS MEULEMAN,
HYEONJOONG JANG, HYUNHO HA,
MIN H. KIM

THE PREMIER **CONFERENCE** & **EXHIBITION** IN
COMPUTER GRAPHICS & INTERACTIVE TECHNIQUES

Motivation

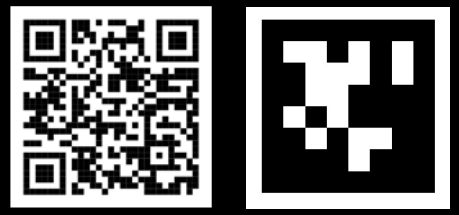
Digital information

101010101001101
001101011000101
010010100100010

Barcode

1-dimensional

QR codes
and AR-tags

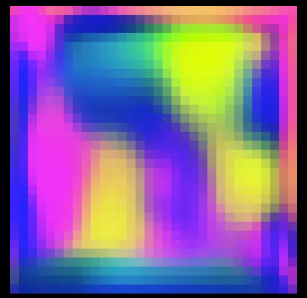


2-dimensional

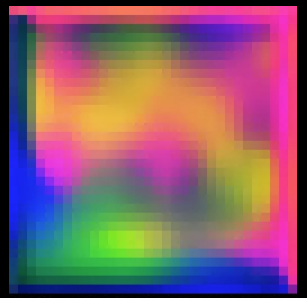
?

Scanning image1: http://leanmedia.org/why-i-dont-partner-with-other-amazon-sellers/
Scanning image2: https://www.carsguide.com.au/car-news/the-qr-code-that-could-save-your-life-28435

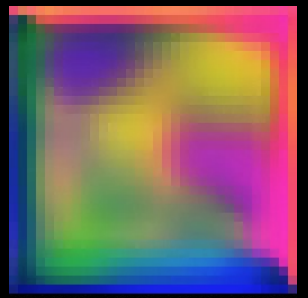
DeepFormableTag



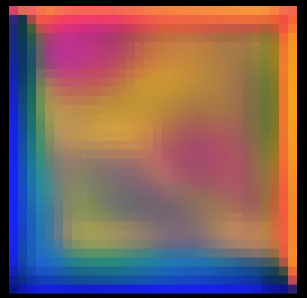
0000000000000000000
0000000000000000001



1111100000000000000
0000000000011111111



1000000000000000000
0001111111111111111

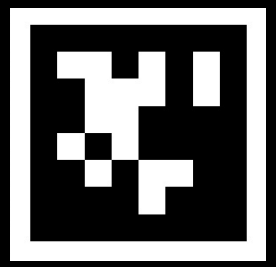


1111111111000000111
1111111111111111111

Previous Work

Binary-square

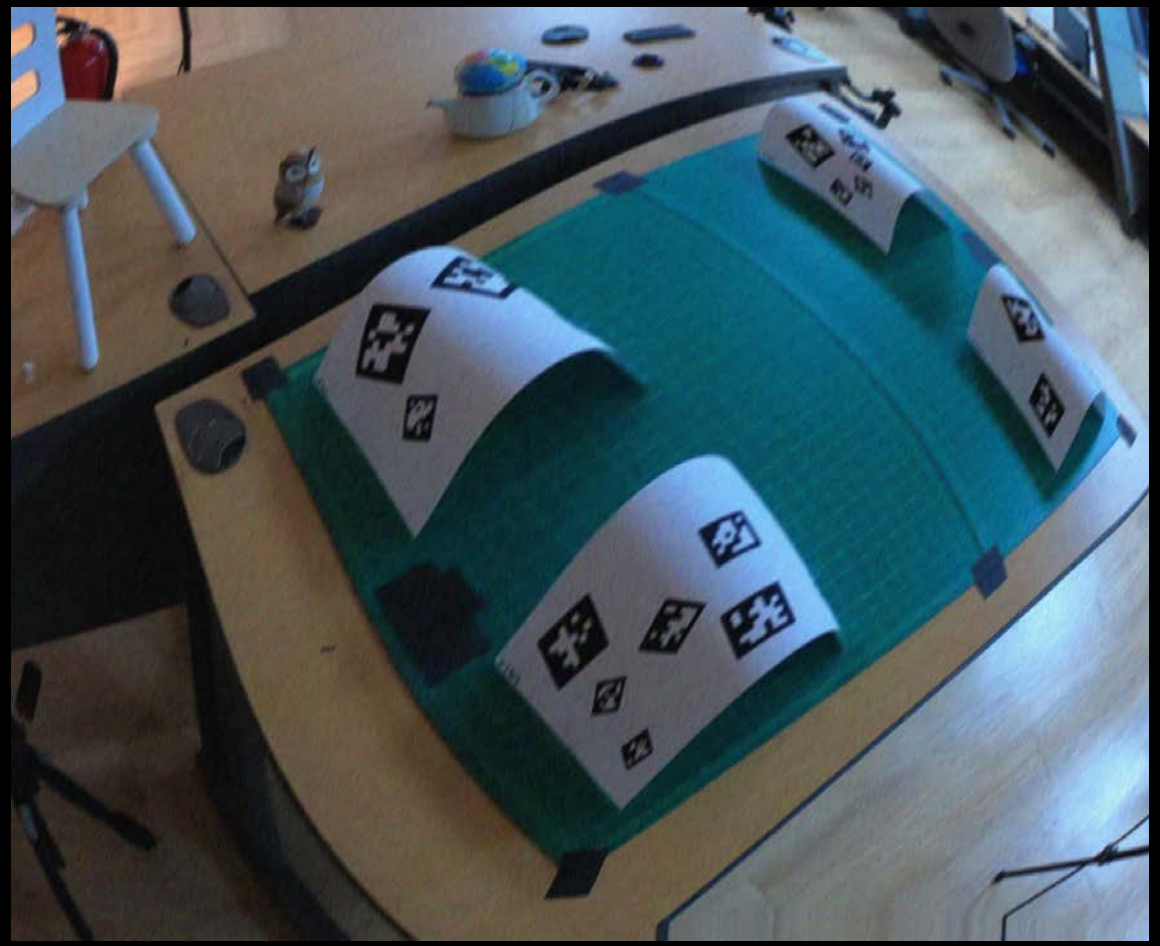
- QR code [Denso Wave 1994],
- ARToolKit [Kato and Billinghurst1999], ARTag [Fiala 2005]
- AprilTag [Olson 2011], ArUco [Munoz-Salinas 2012]



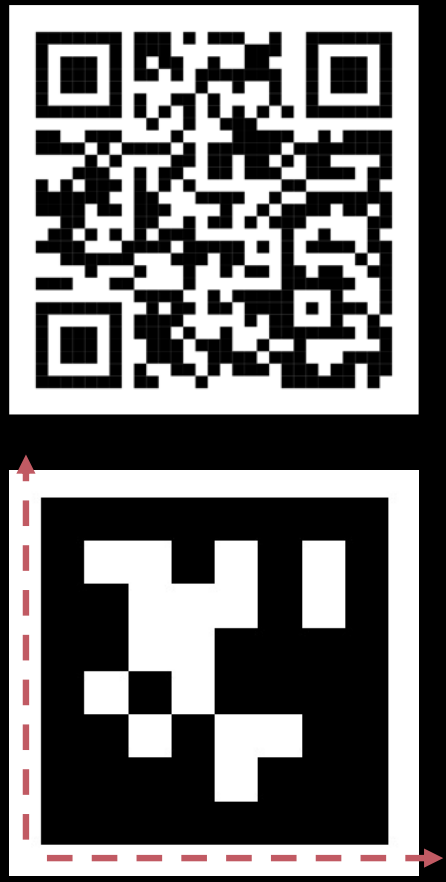
[Images are from corresponding reference papers]

Previous Work

Planar and rigid surface assumption



Deformed surfaces, distortion, etc.

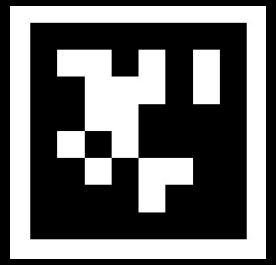


Message embedding capacity

Previous Work

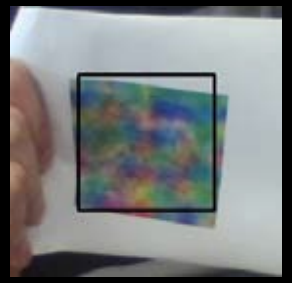
Binary-square

- QR code [Denso Wave 1994],
- ARToolKit [Kato and Billinghurst1999], ARTag [Fiala 2005]
- AprilTag [Olson 2011], ArUco [Munoz-Salinas 2012]

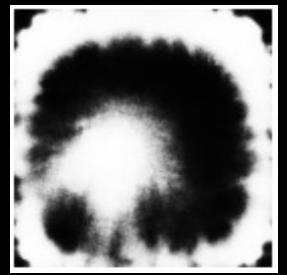


Learning-based

- Learnable visual markers [Grinchuk 2016]



- E2E-Tag [Peace 2020]

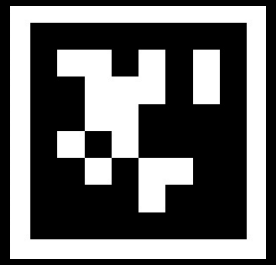


[Images are from corresponding reference papers]

Previous Work

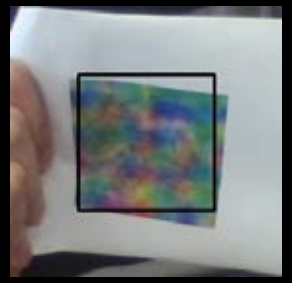
Binary-square

- QR code [Denso Wave 1994],
- ARToolKit [Kato and Billinghurst1999], ARTag [Fiala 2005]
- AprilTag [Olson 2011], ArUco [Munoz-Salinas 2012]



Learning-based

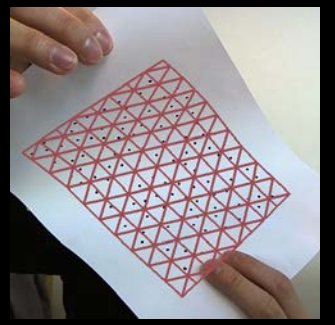
- Learnable visual markers [Grinchuk 2016]



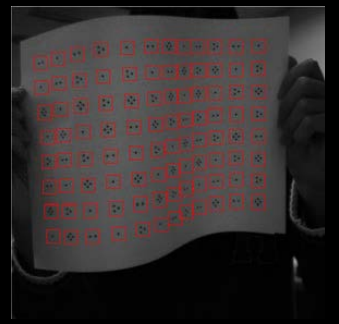
- E2E-Tag [Peace 2020]

Dot pattern

- DRDM [Uchiyama and Marchand 2011]

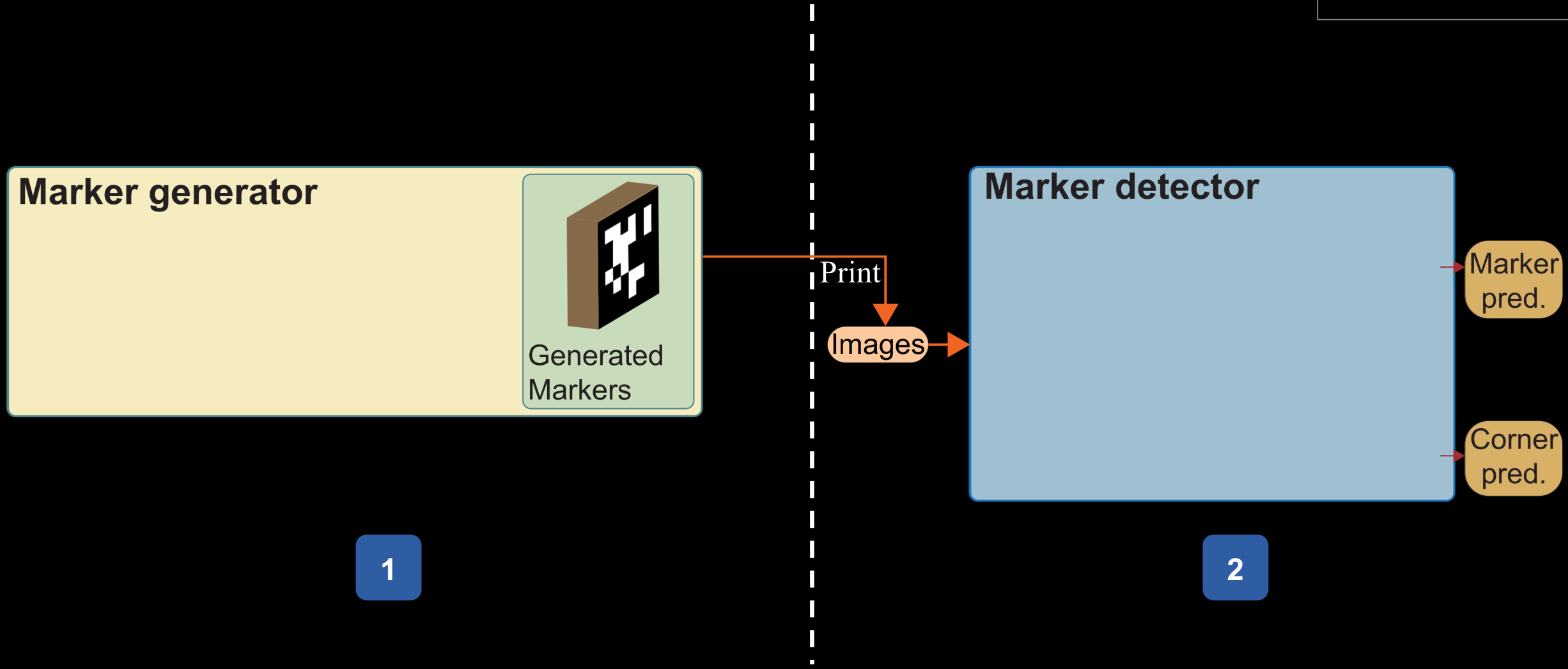


- DDCM [Narita 2016]

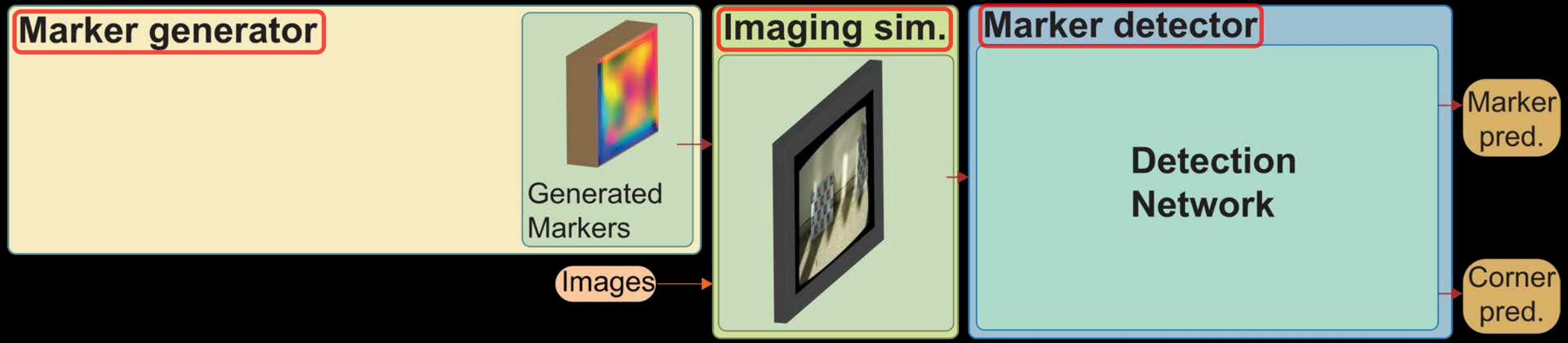


[Images are from corresponding reference papers]

Classical Marker Systems

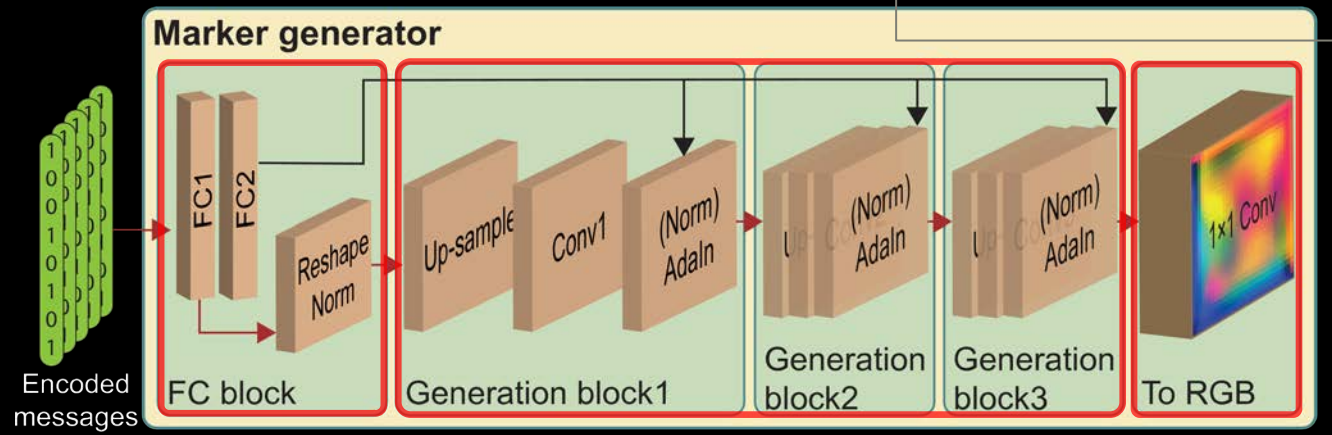


DeepFormableTag - Overview

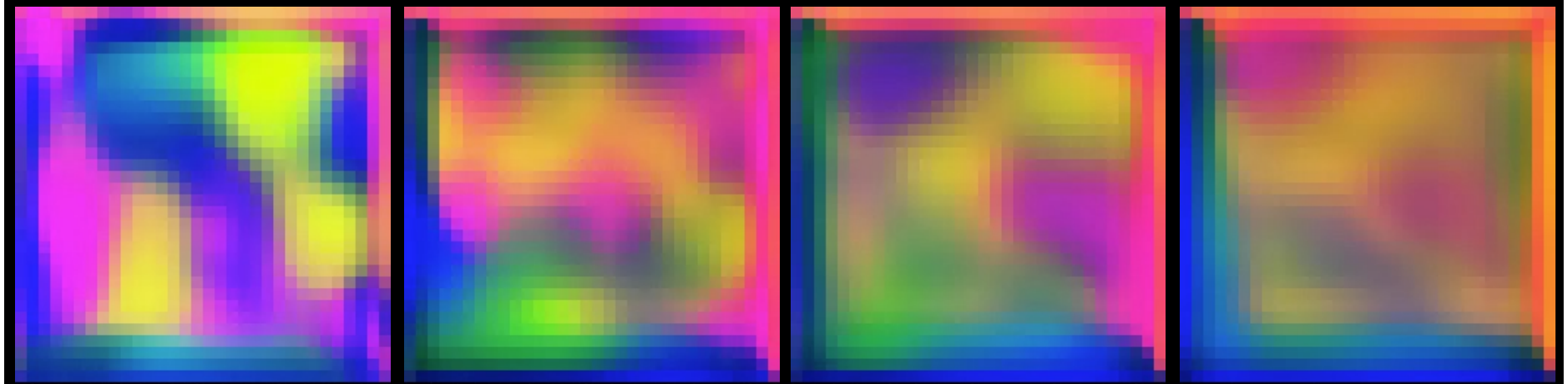


Marker Generator

- Rich appearance
- Marker similarity (localization)
- Marker uniqueness (classification/decoding)

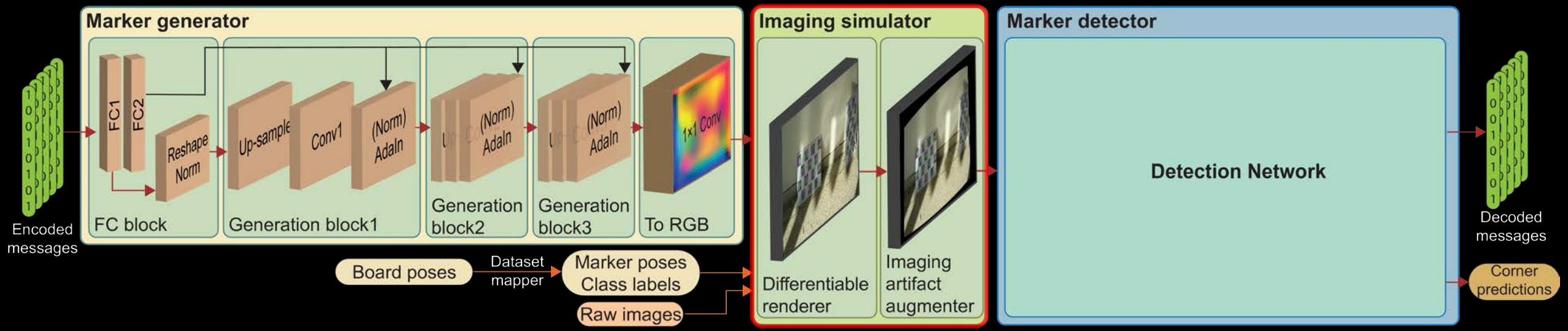


Convolutional generator



Imaging Simulator

- Differentiable
- High photorealism
- Imaging artifacts



Imaging Simulator: Rendering (1/5)

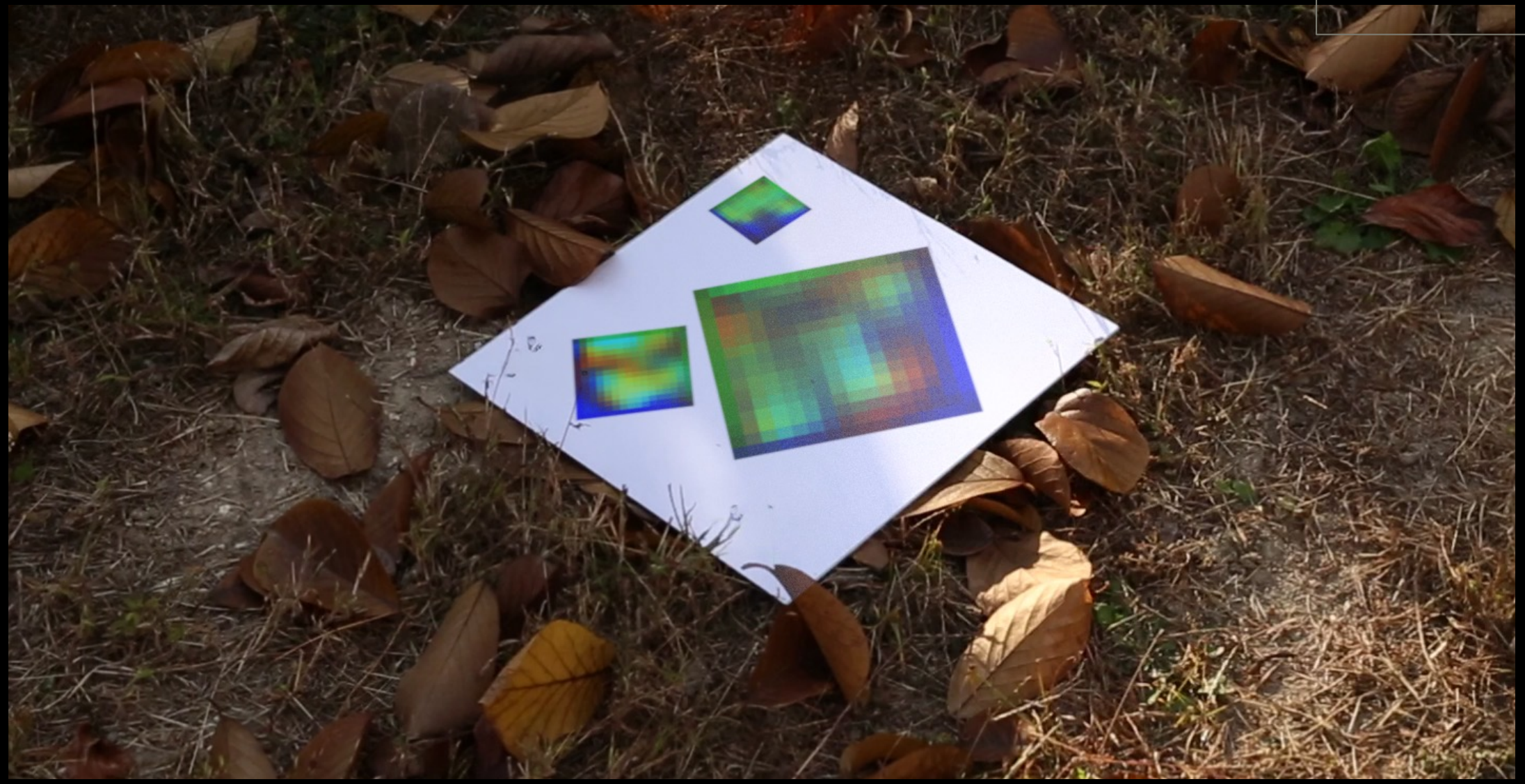
- Fast rendering
- Small domain gap with real-world

(a) Random superimposed rendering

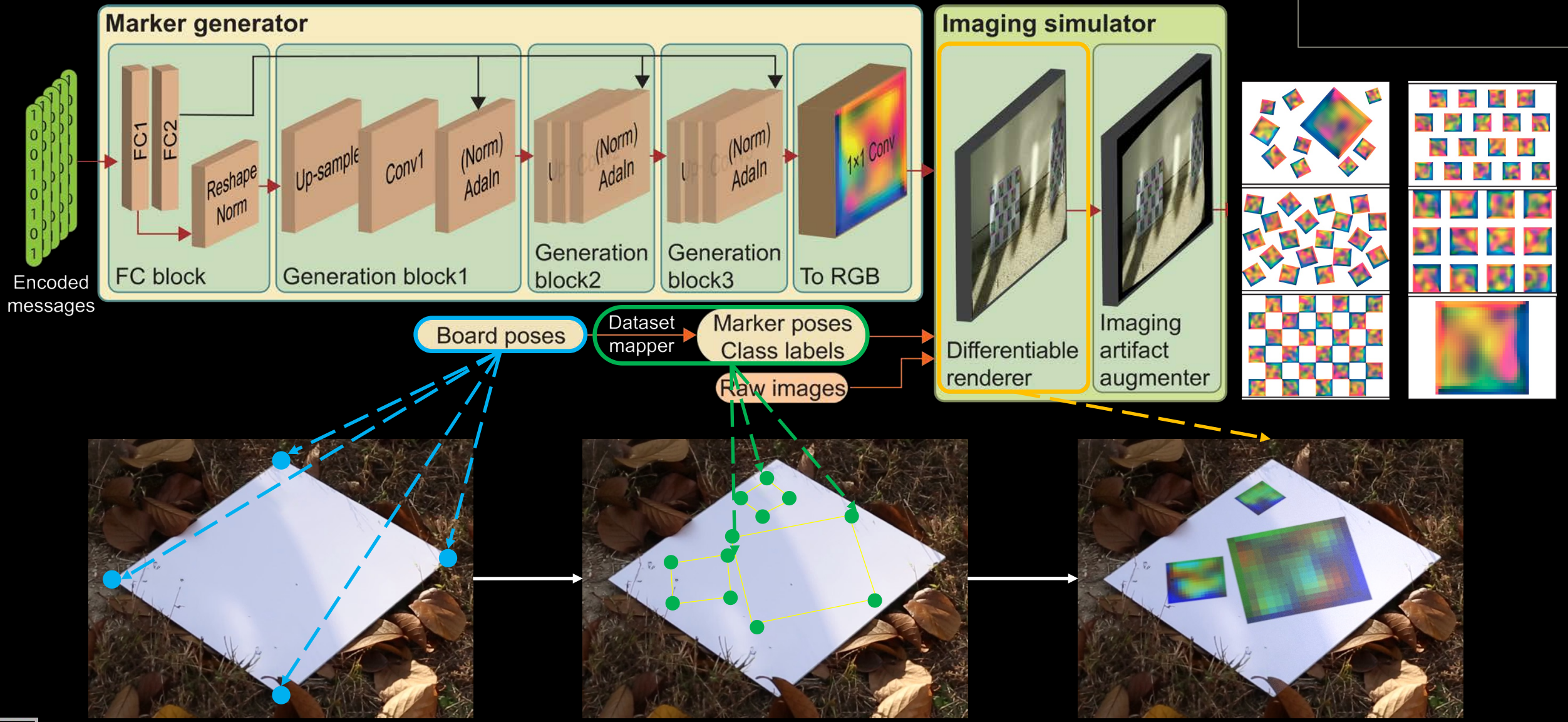
(b) Our photorealistic rendering

[Original image courtesy of the Mennonite Church USA Archives]

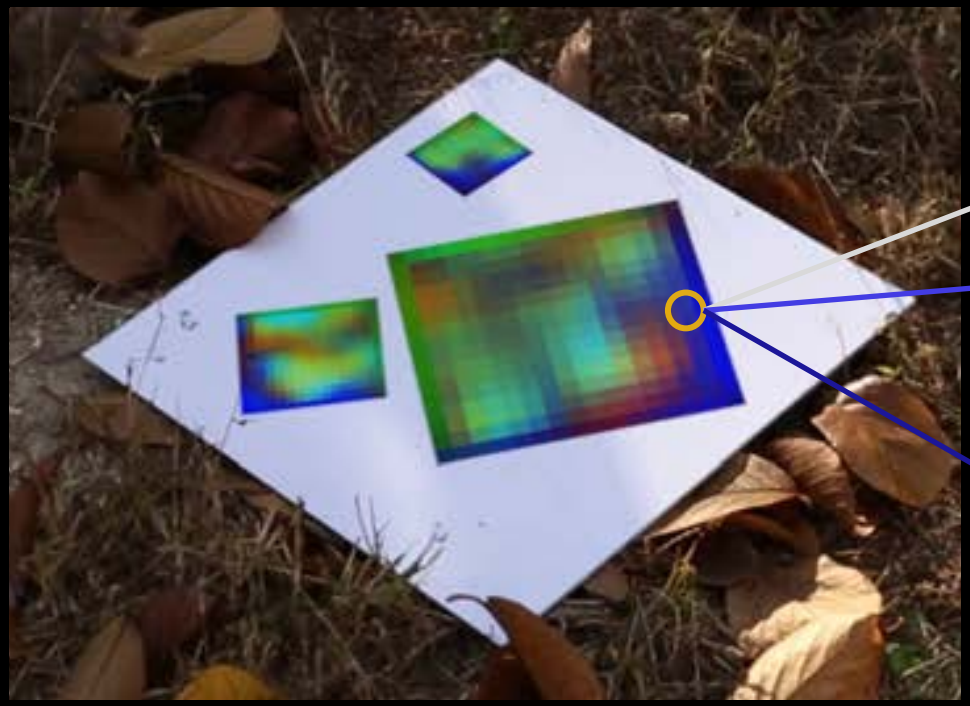
Imaging Simulator: Rendering (2/5)



Imaging Simulator: Rendering (3/5)



Imaging Simulator: Rendering (4/5)



Under Lambertian assumption:

$$f_{\mathbf{x},\vec{n},m_p}(\vec{w},\vec{v}) = \frac{\rho_p}{\pi}$$: Board's BRDF

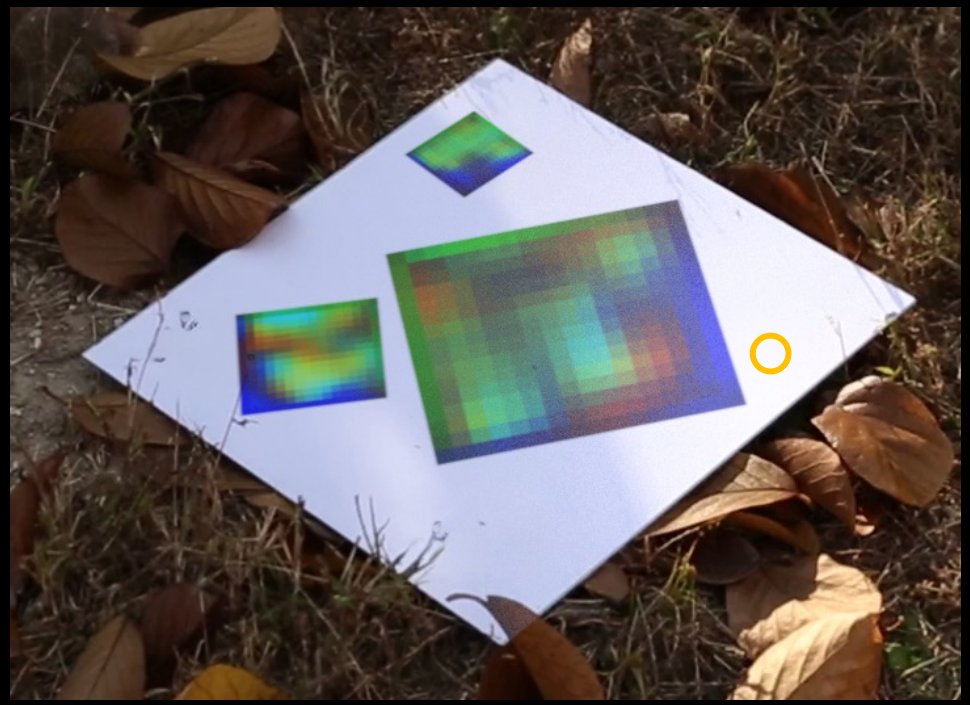
$$f_{\mathbf{x},\vec{n},m_t}(\vec{w},\vec{v}) = \frac{\rho_t}{\pi}$$: Target diffuse

$$L(\mathbf{x},\vec{v}) = \frac{\rho}{\pi}\int_{\Omega} L(\vec{w},\mathbf{x})cos(\theta)d\omega$$

$$L_{m_t}(\mathbf{x},\vec{v}) = L_{m_p}(\mathbf{x},\vec{v})\frac{\rho_t}{\rho_p}$$

Final radiance

Imaging Simulator: Rendering (5/5)



Cook-Torrance specular term:

$$\frac{F}{\pi}\frac{DG}{(\vec{n}\cdot\vec{l})(\vec{n}\cdot\vec{v})}$$

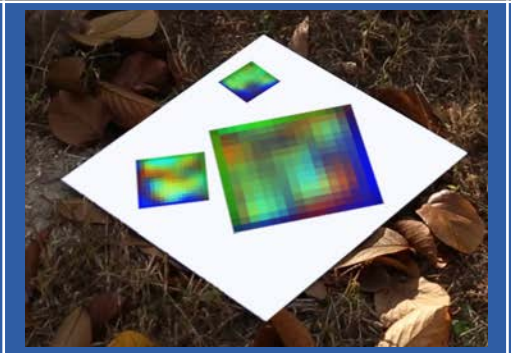
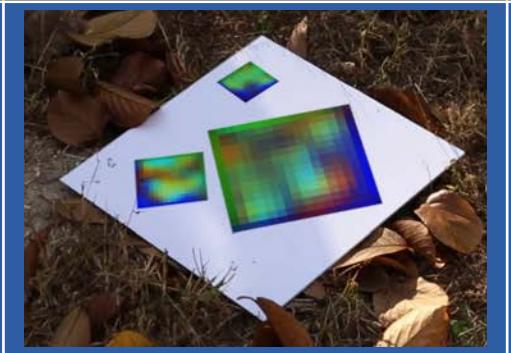
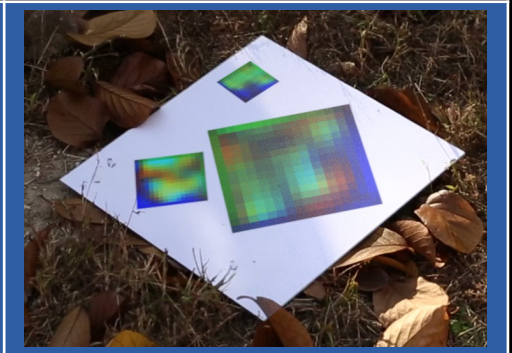
F: Fresnel term

D: Facet distribution function

G: Geometric attenuation factor

Imaging Simulator: Ablation Study

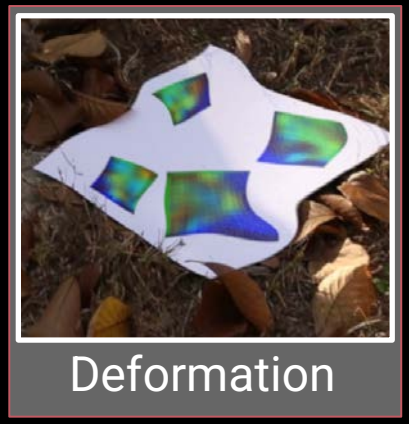
- Trained with specified rendering method, tested on real-world

Method	COCO Superimposed	w/o shading	Diffuse	Diffuse+Specular
				
AP Score	35.59	52.30	80.52	**87.66**

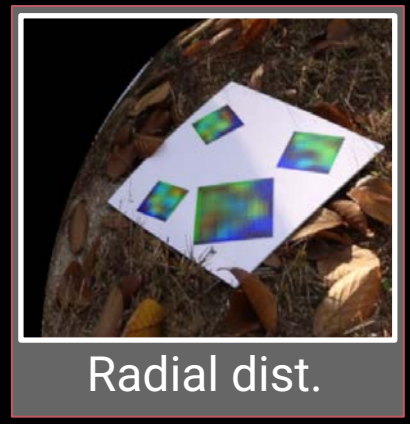
Higher is better

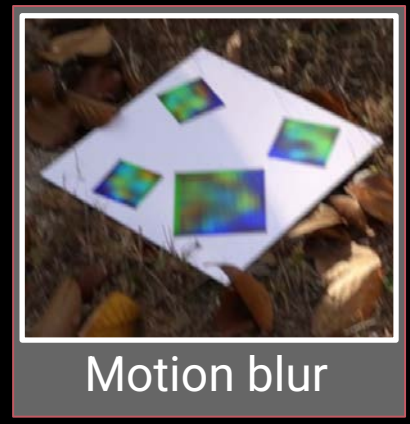
Imaging Simulator: Imaging Artifact Augmenter

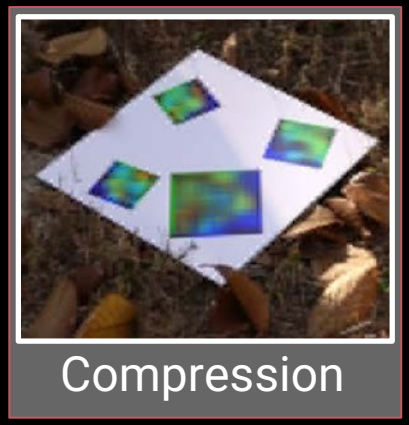
- Robustness against combination of various edge conditions

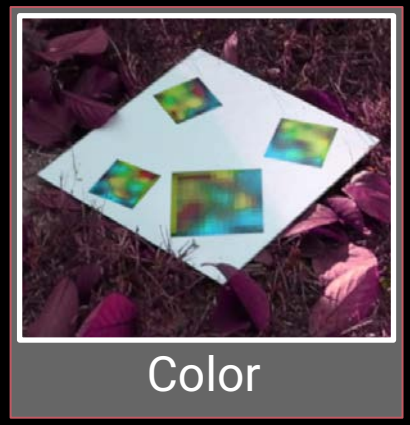

Deformation

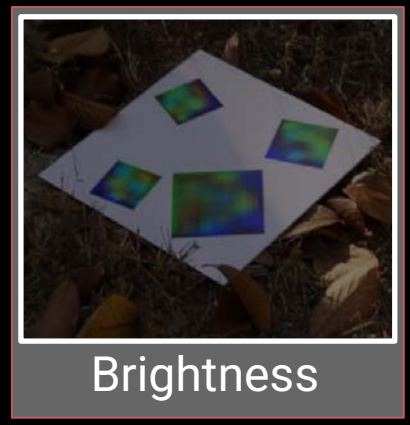

Perspective

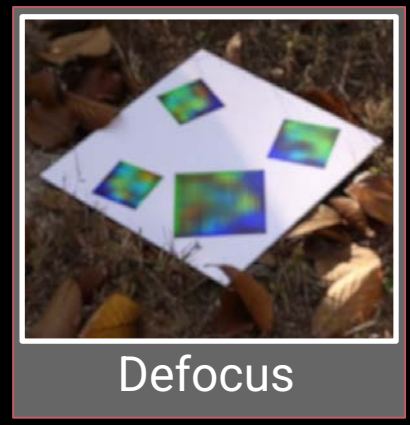

Radial dist.


Motion blur


Compression


Color

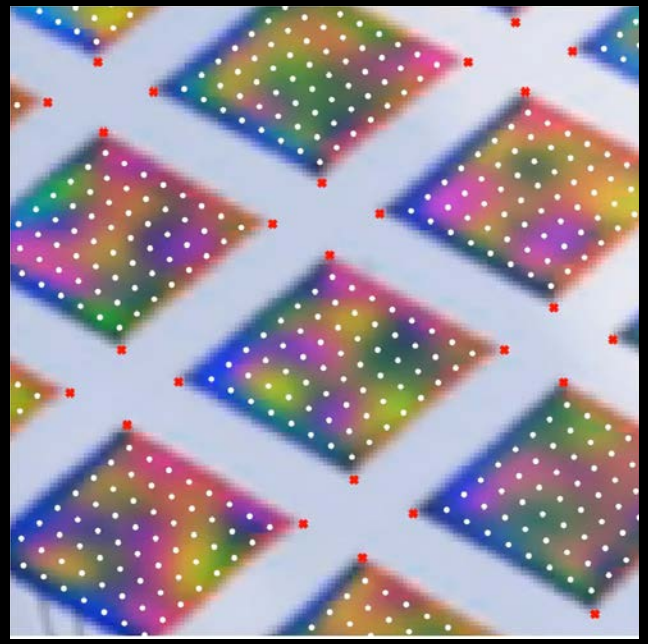

Brightness


Defocus

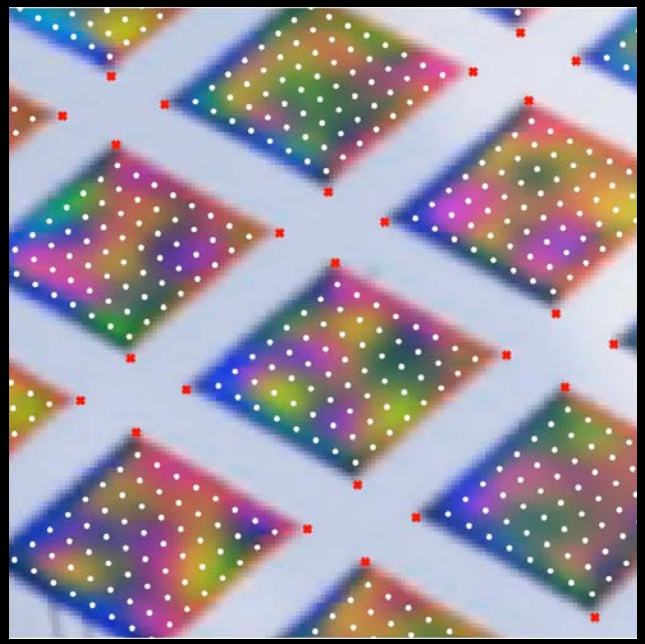
Imaging Simulator: Geometric Distortions

- Modified corners
- Internal sampling points

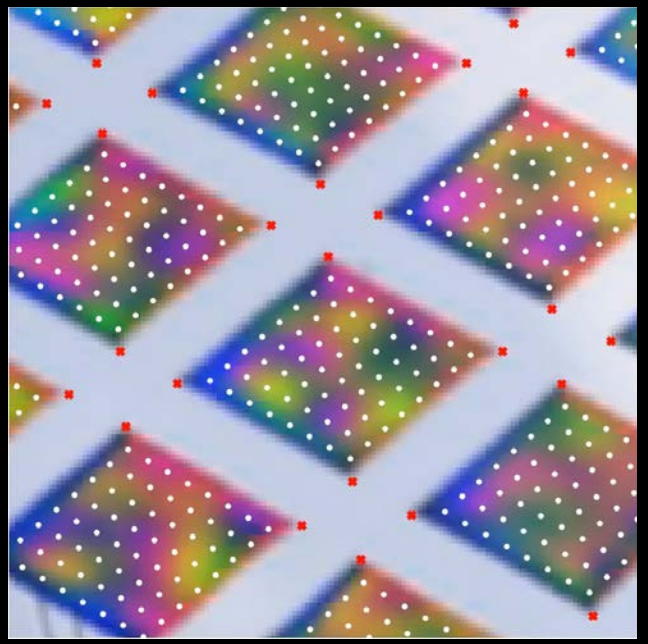
Deformation



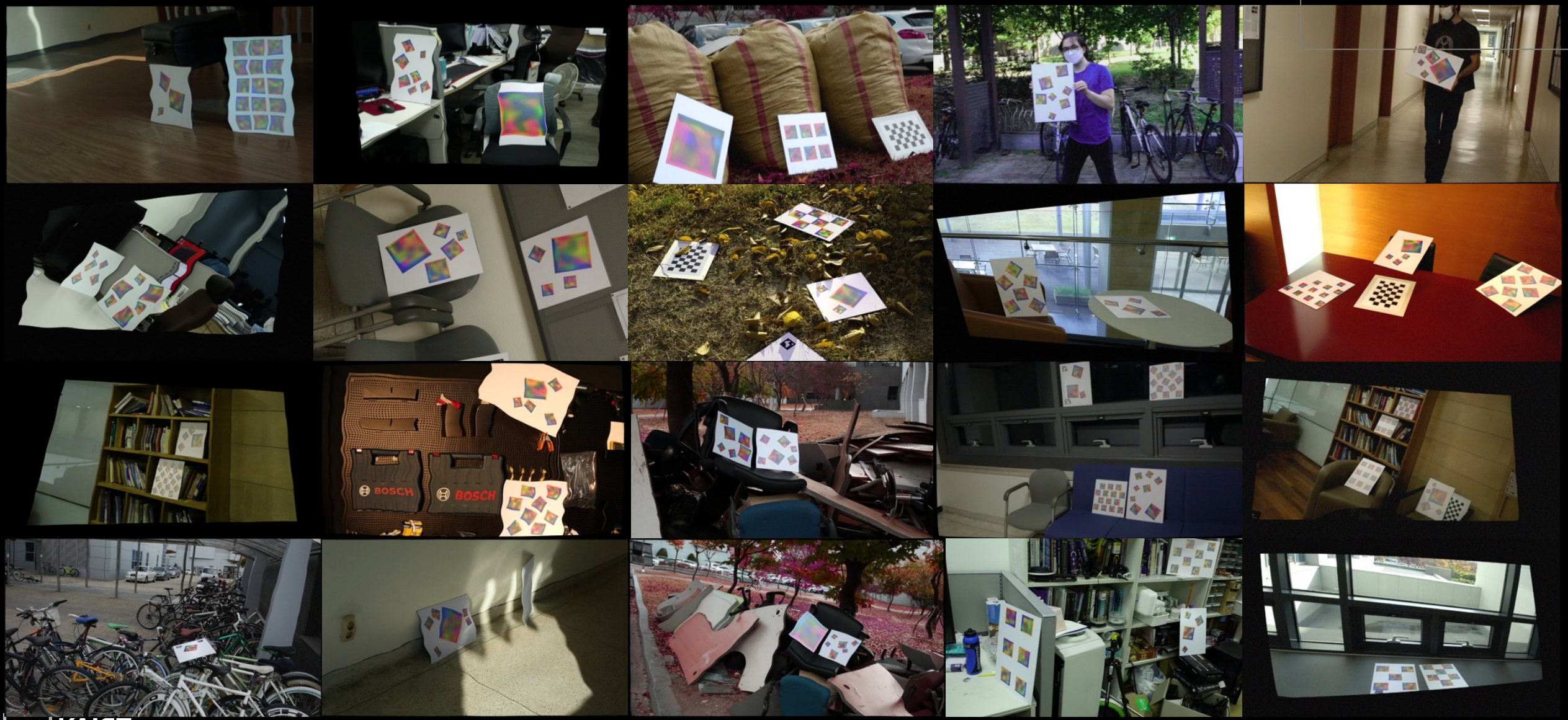
Radial distortion



Perspective distortion

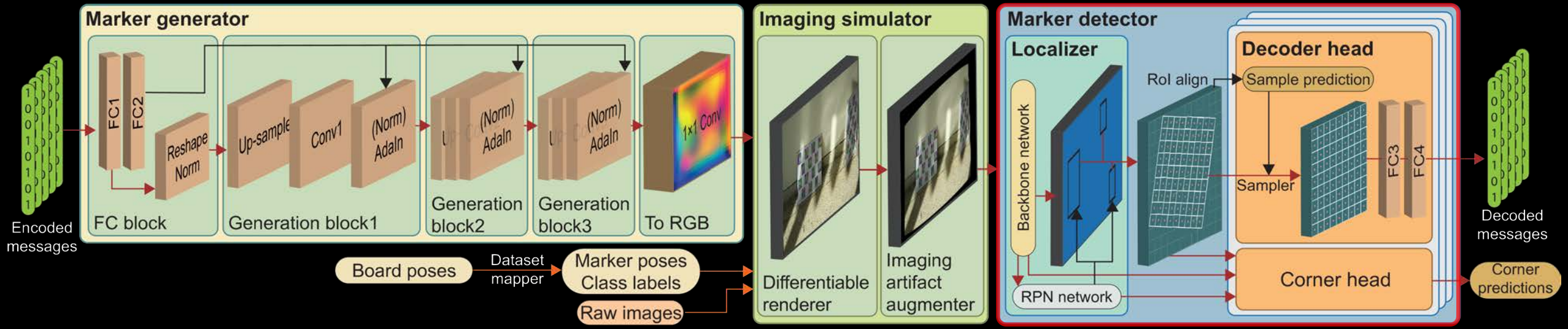


Training Dataset Rendered with Augmentations

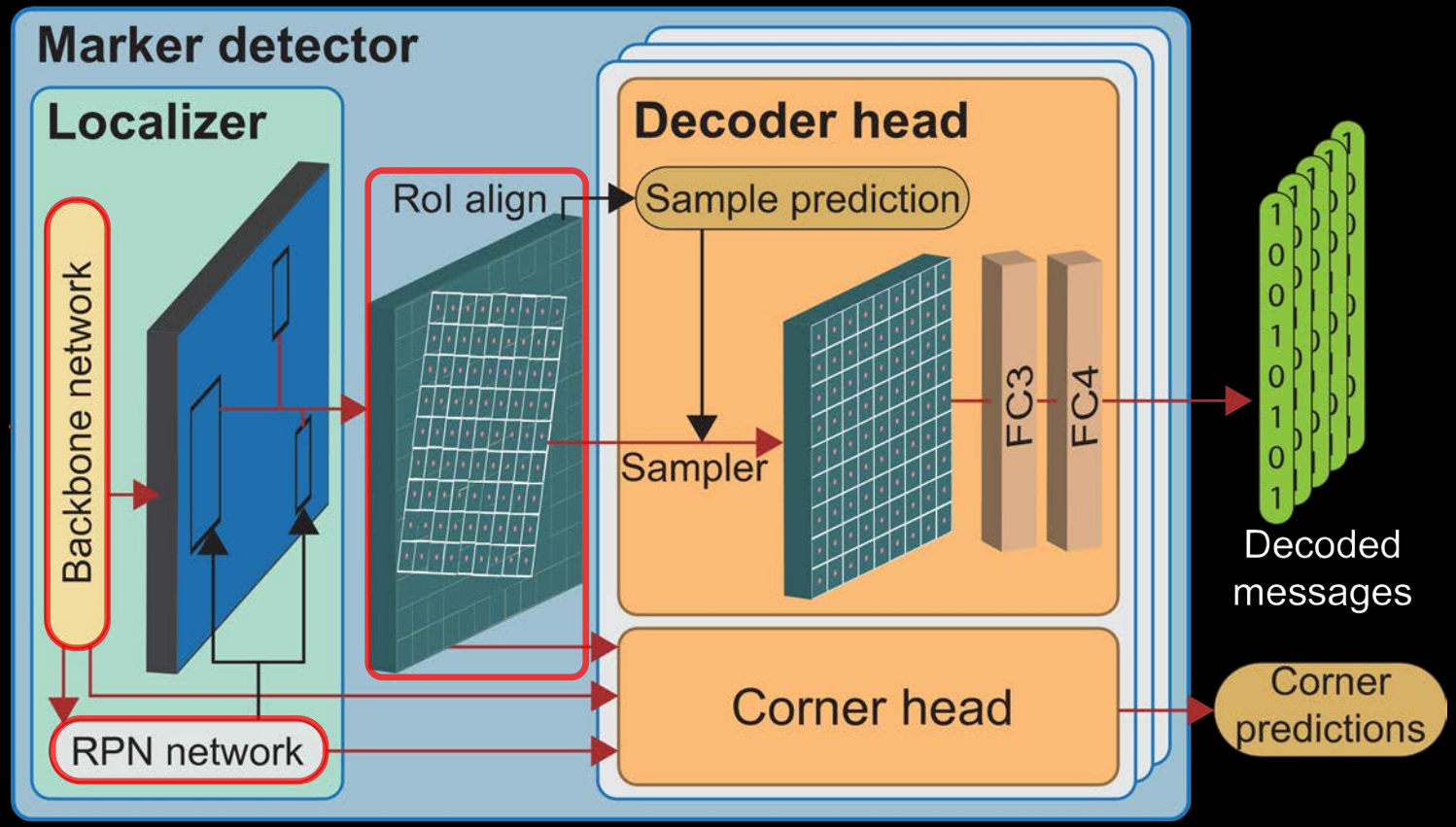


Marker Detector (1/5)

- Efficiency
- Geometric invariance



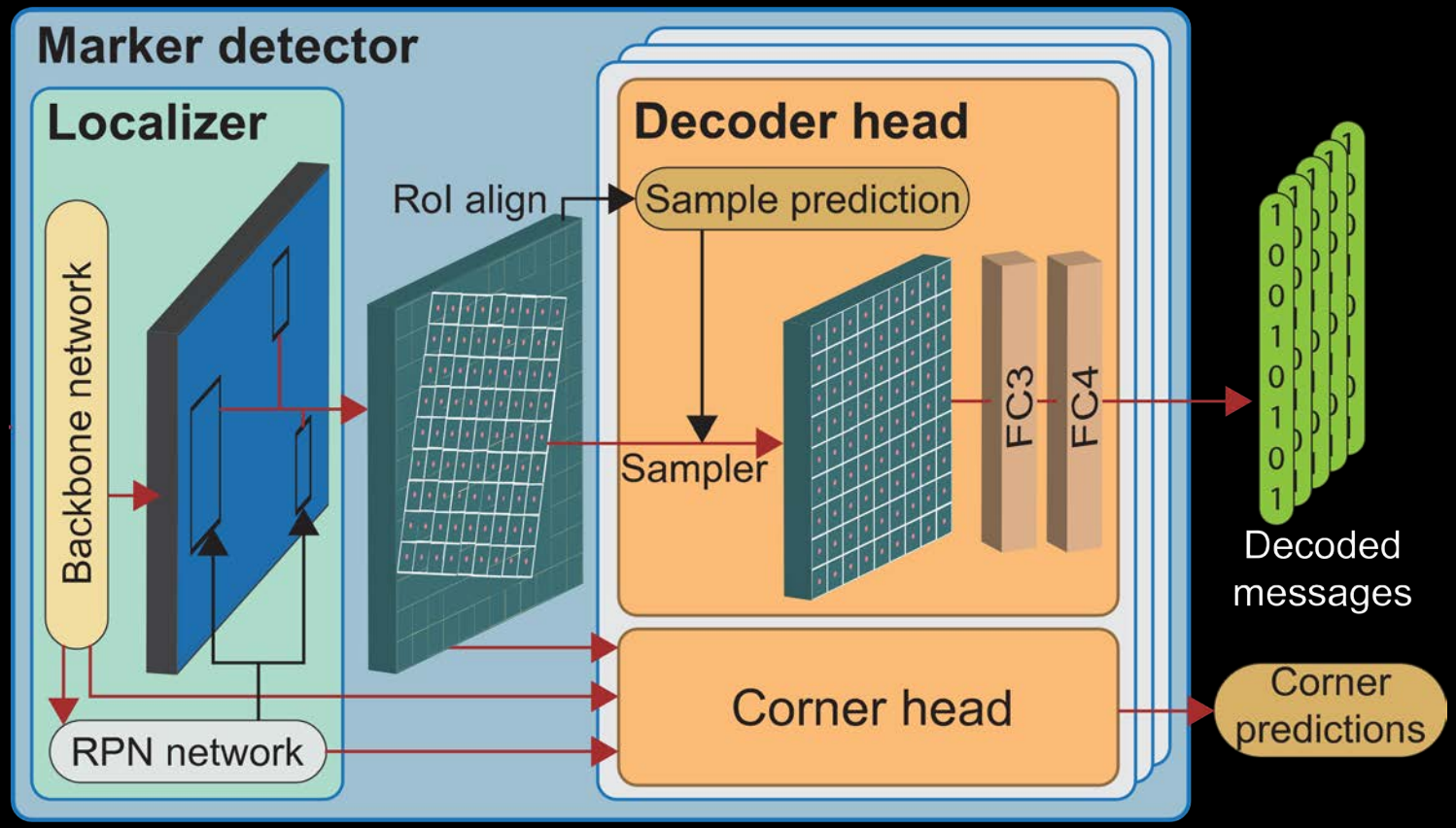
Marker Detector: Backbone (2/5)

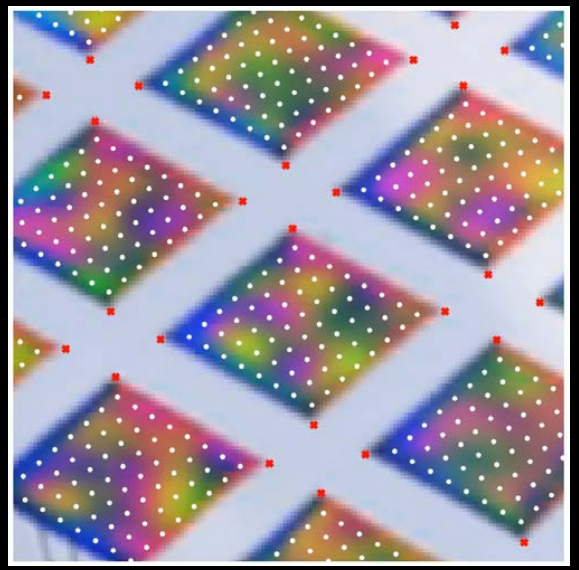


- Two-stage Faster RCNN
- VoVNet19-FPNLite

https://github.com/facebookresearch/detectron2

Marker Detector: Decoder Head (3/5)

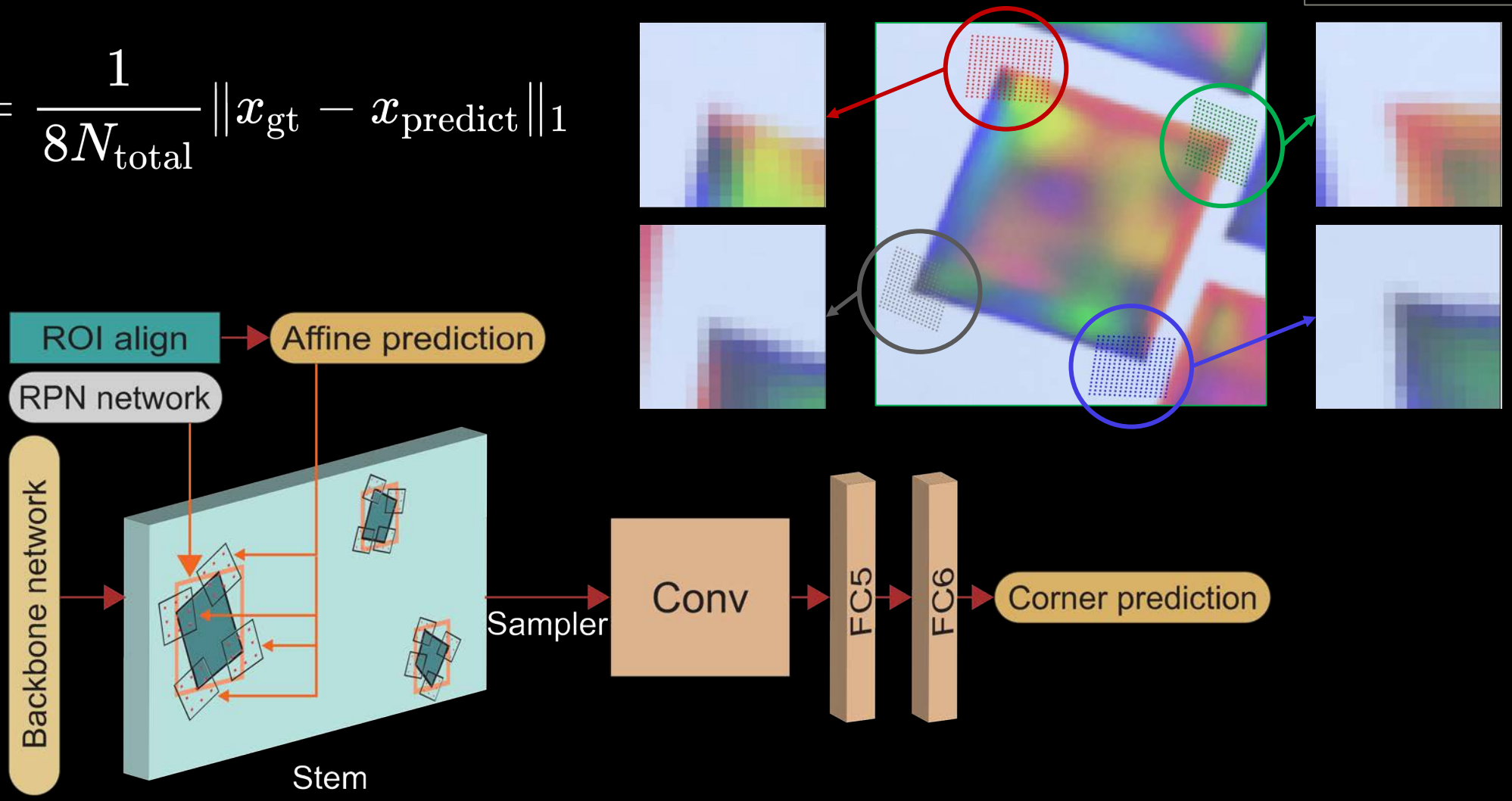




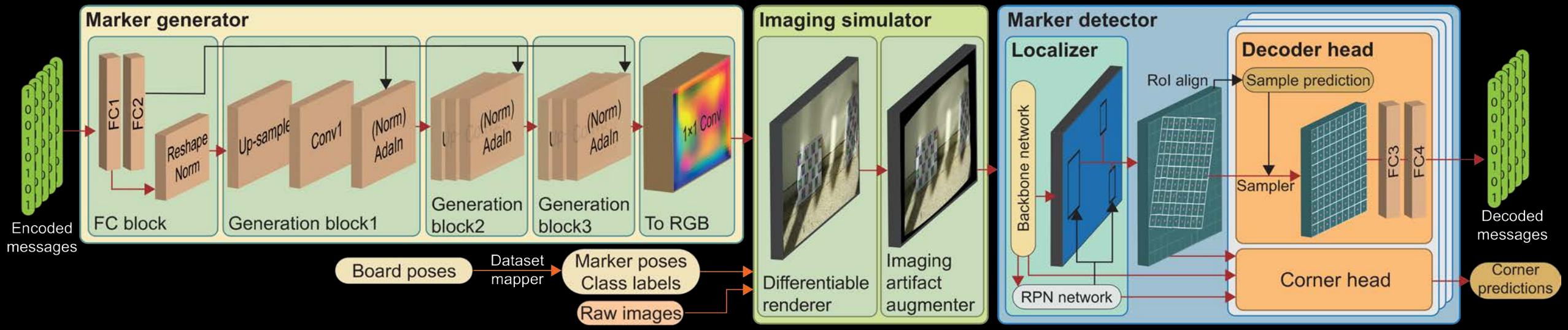
$$\mathcal{L}_{\text{sample}} = \frac{1}{2N_{\text{total}}N_{\text{sample}}} \left\| x_{\text{gt}} - x_{\text{predict}} \right\|_1$$

Marker Detector: Corner Head (4/5)

$$\mathcal{L}_{\text{corner}} = \frac{1}{8N_{\text{total}}} \|x_{\text{gt}} - x_{\text{predict}}\|_1$$



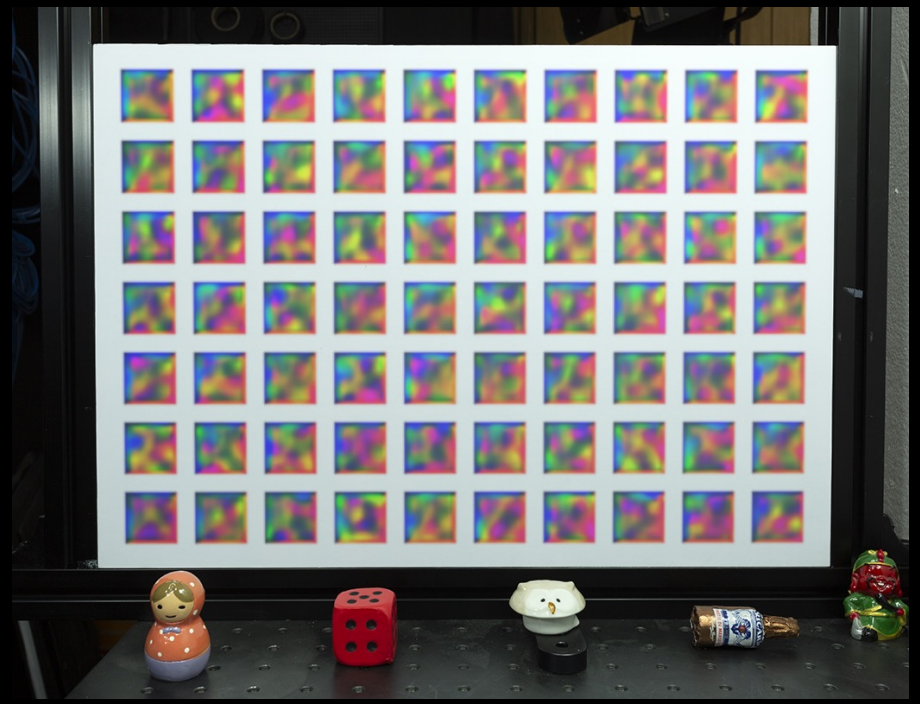
Marker detector: Loss terms (5/5)



$$\mathcal{L}_{\text{total}} = \underbrace{(\mathcal{L}_{\text{rpn}_{\text{class}}} + \mathcal{L}_{\text{rpn}_{\text{loc}}})}_{\text{Region proposal network loss}} + \underbrace{\mathcal{L}_{\text{sample}}}_{\text{Sampling loss}} + \underbrace{0.1\mathcal{L}_{\text{corner}}}_{\text{Corner loss}} + \underbrace{(0.5\mathcal{L}_{\text{obj}} + 10.0\mathcal{L}_{\text{decode}})}_{\text{Objectness and decoding loss}}$$

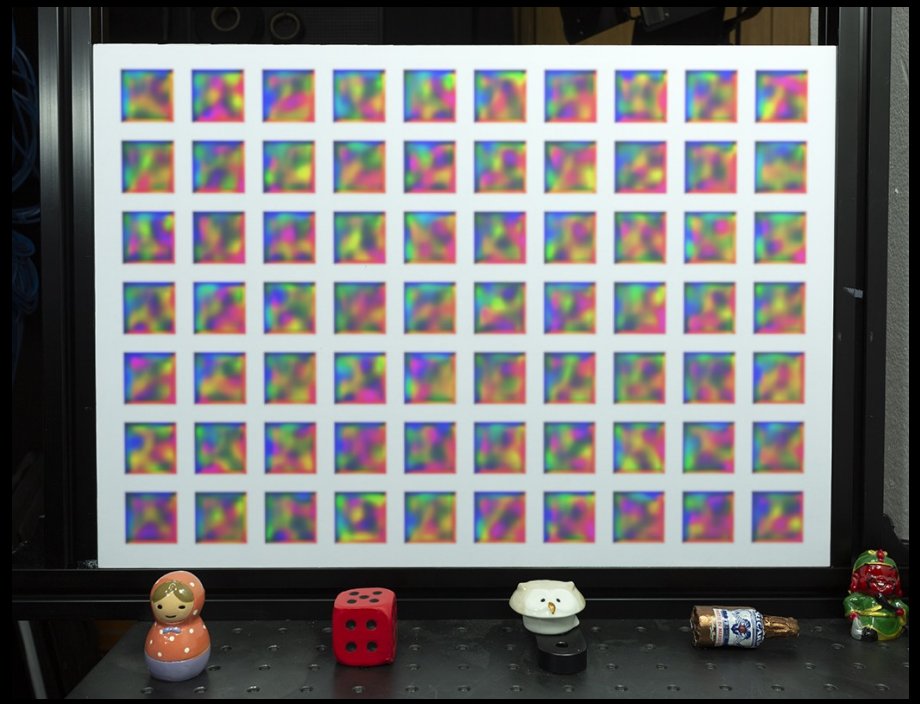
Results: Message Decoding Capability

Model	Mean decoding accuracy	Standard deviation	0-bit error	1-bit error
16 bits	99.998%	0.1143	99.97%	0.03%
36 bits	99.921%	0.7605	98.51%	0.75%
64 bits	99.558%	1.0787	80.39%	14.0%



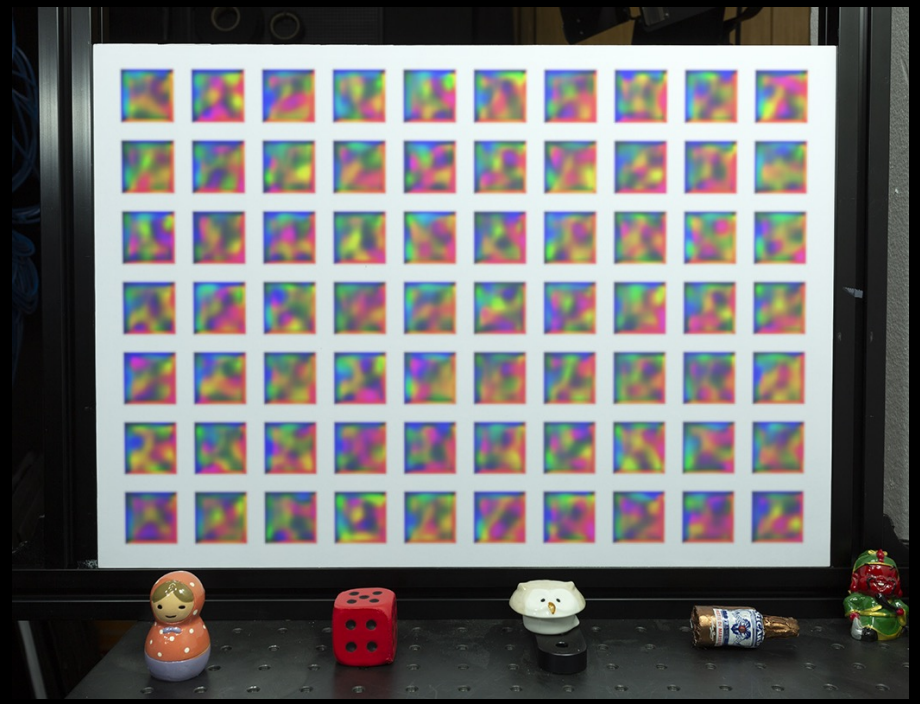
Results: Message Decoding Capability

Model	Mean decoding accuracy	Standard deviation	0-bit error	1-bit error
16 bits	99.998%	0.1143	99.97%	0.03%
36 bits	99.921%	0.7605	**98.51%**	0.75%
64 bits	99.558%	1.0787	80.39%	14.0%

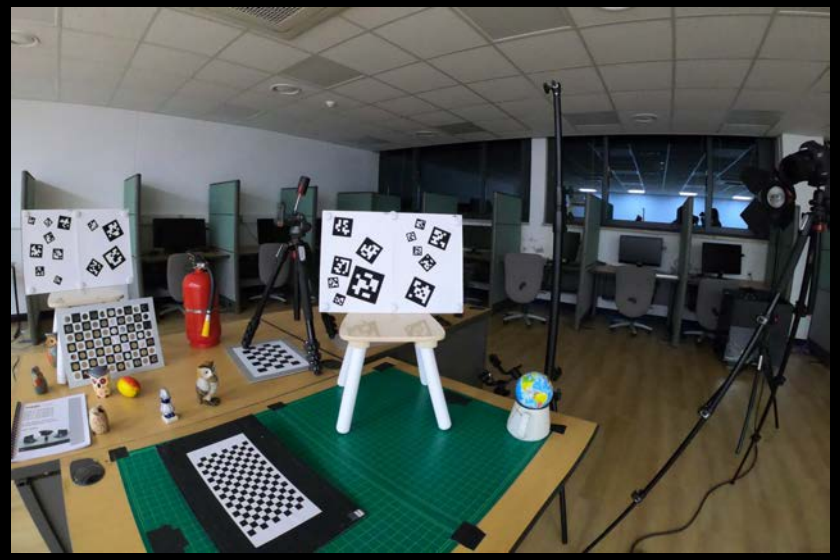


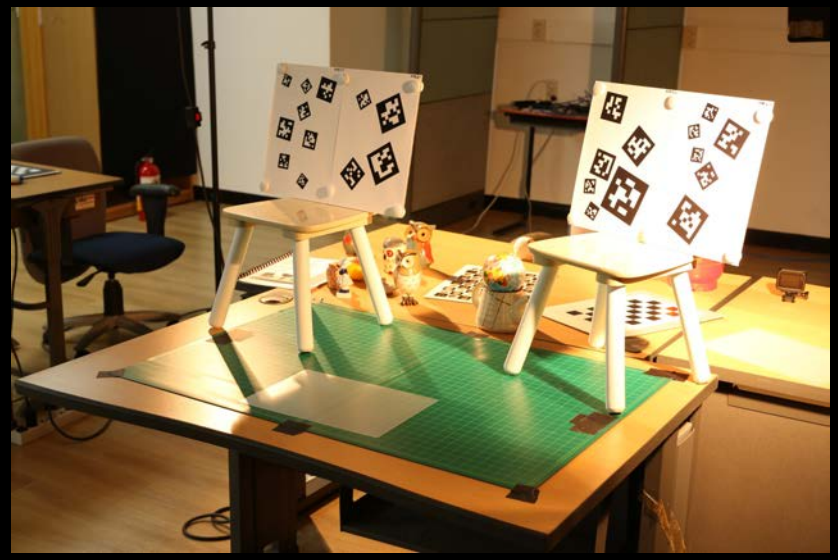
Results: Message Decoding Capability

Model	Mean decoding accuracy	Standard deviation	0-bit error	1-bit error
16 bits	**99.998%**	0.1143	99.97%	0.03%
36 bits	**99.921%**	0.7605	98.51%	0.75%
64 bits	**99.558%**	1.0787	80.39%	14.0%



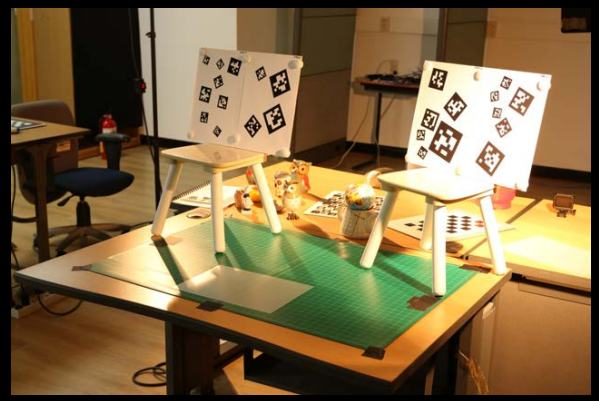
Results: Comparison on Flat Surfaces





Results: Comparison on Flat Surfaces

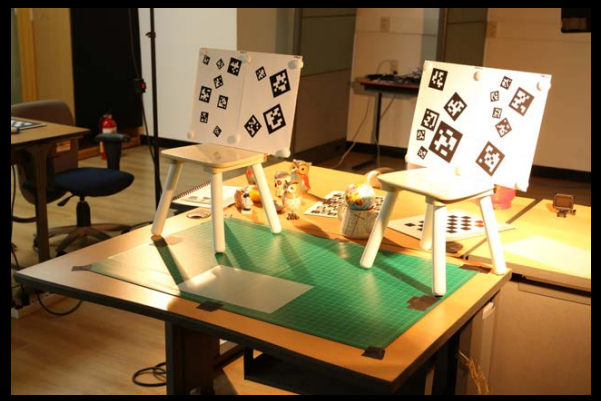


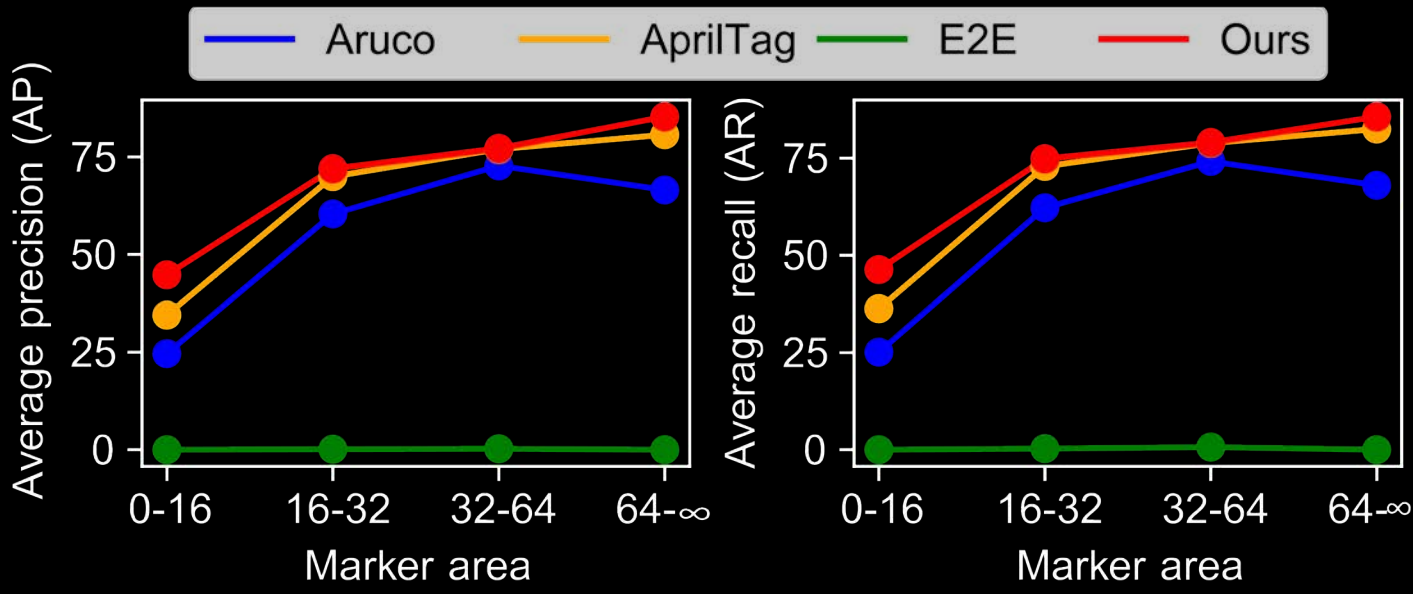
Model	AP	FP-Rate	FPS
ArUco [2012]	50.19	0.0000	31
AprilTag [2016]	57.58	0.0000	19
E2ETag [2020]	00.04	0.8625	13
Ours	**60.84**	0.0000	29

Results: Comparison on Flat Surfaces



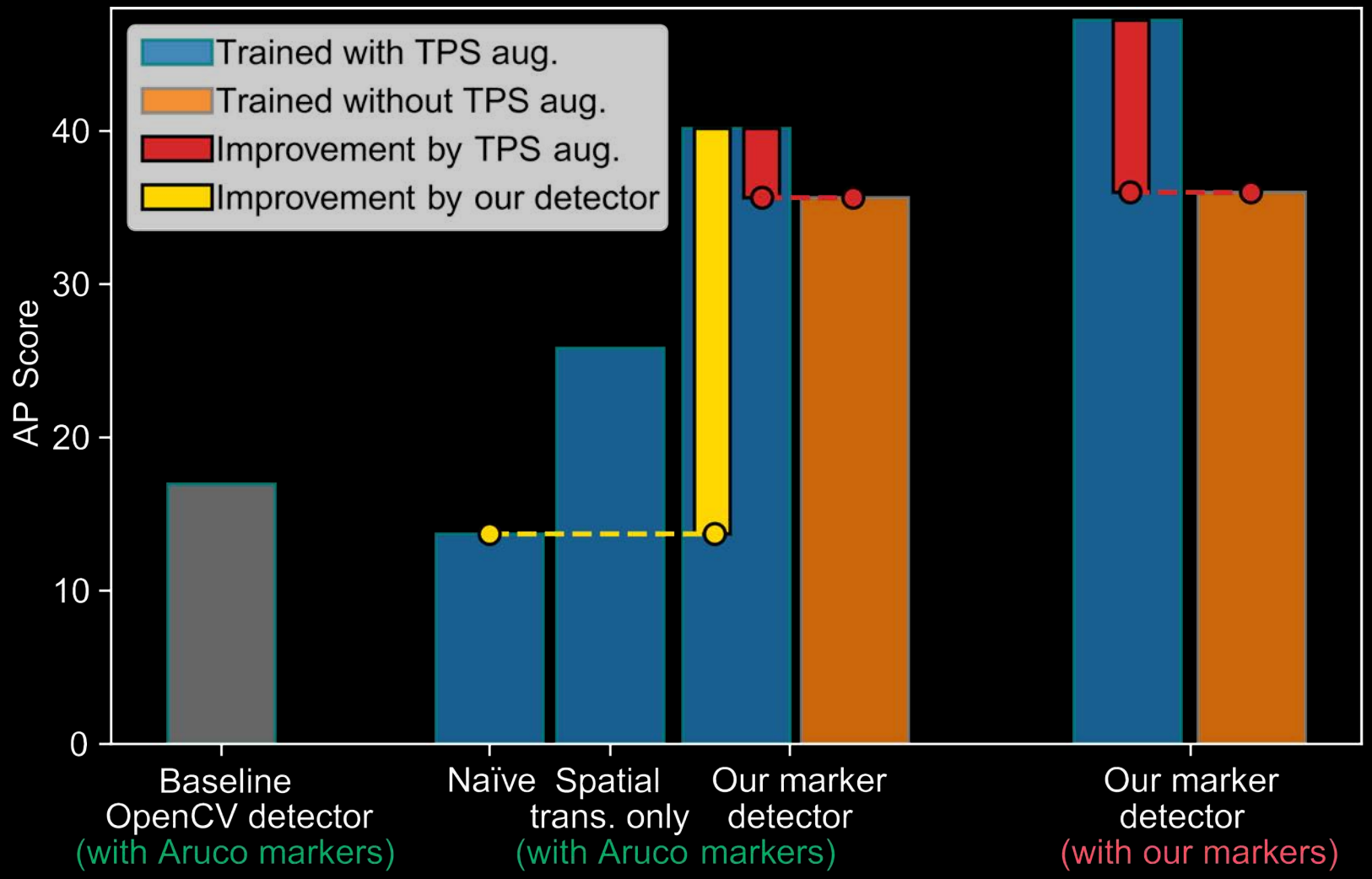
Model	AP	FP-Rate	FPS
ArUco [2012]	50.19	0.0000	31
AprilTag [2016]	57.58	0.0000	19
E2ETag [2020]	00.04	0.8625	13
Ours	**60.84**	0.0000	29

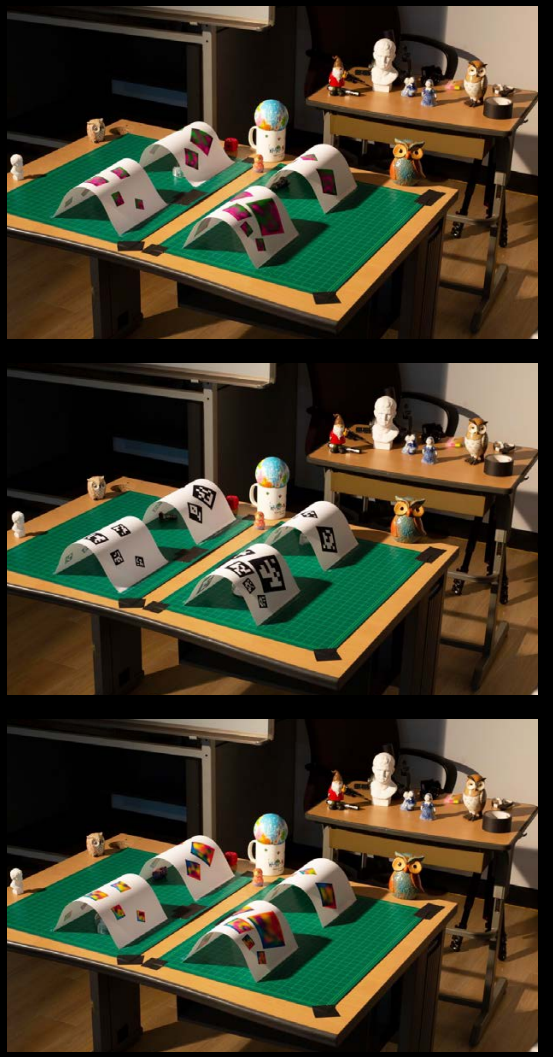


Results: Deformation (1/4)



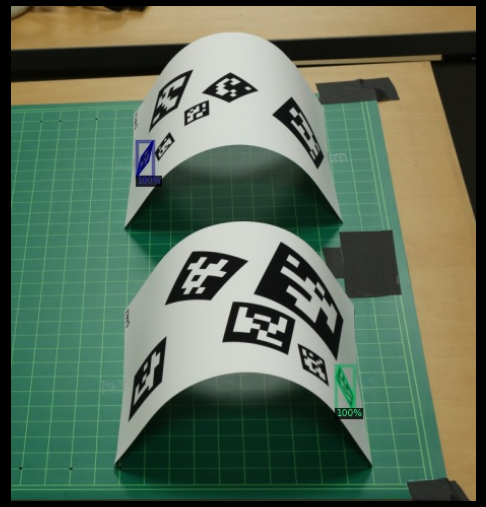
Results: Deformation (1/4)



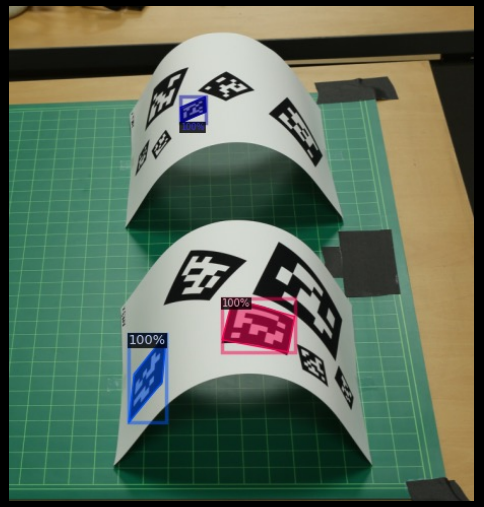


Results: Deformation (2/4)

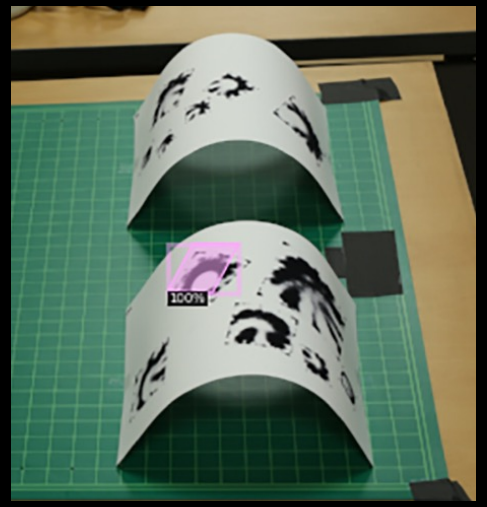
ArUco



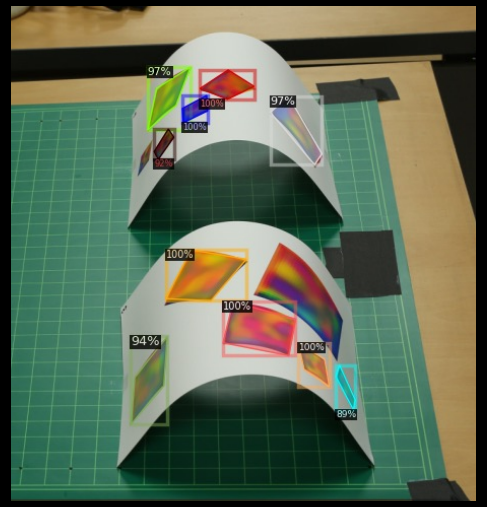
AprilTag

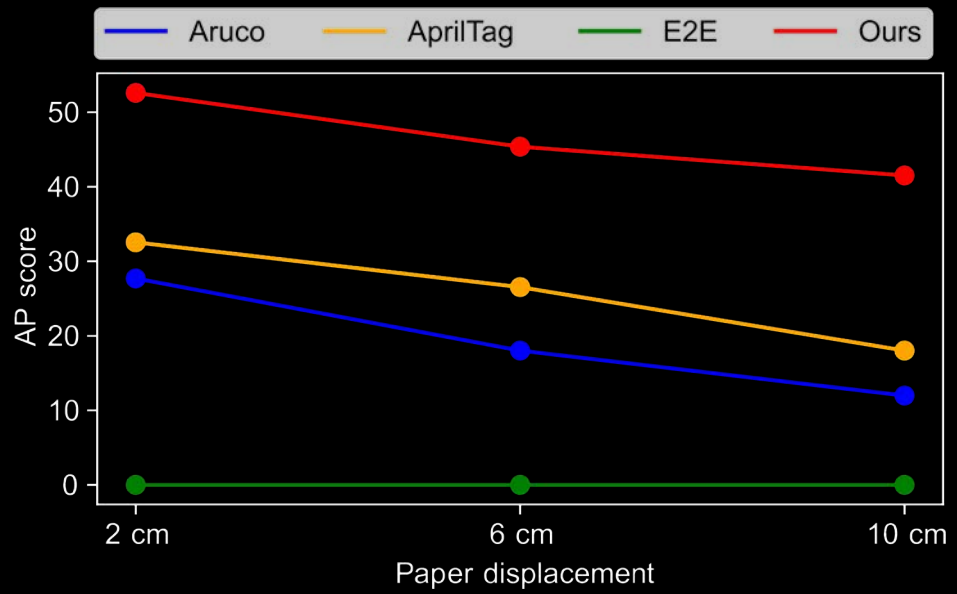


E2ETag



Ours

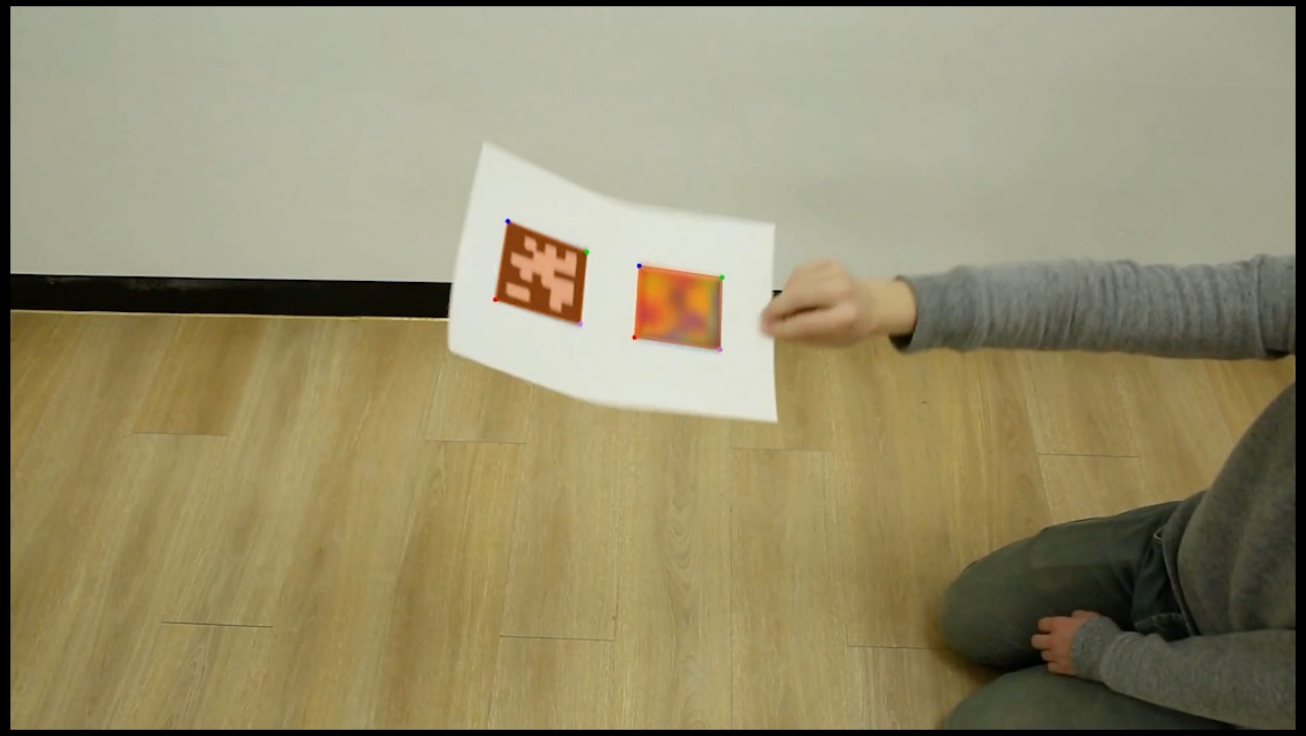




KAIST VISUAL COMPUTING Lab

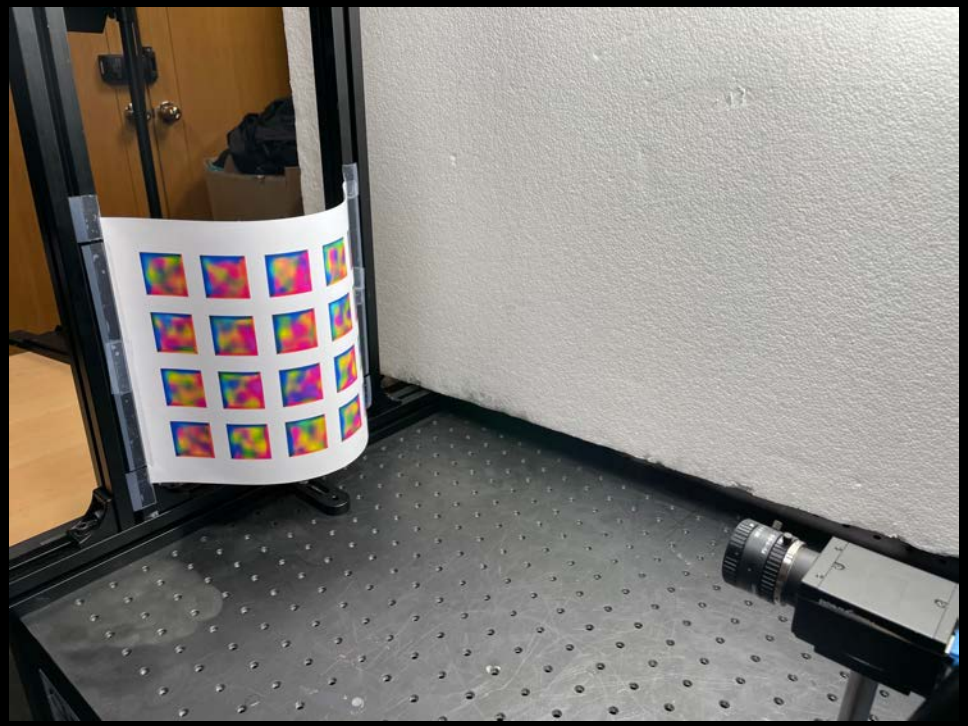
Results: Deformation (3/4)

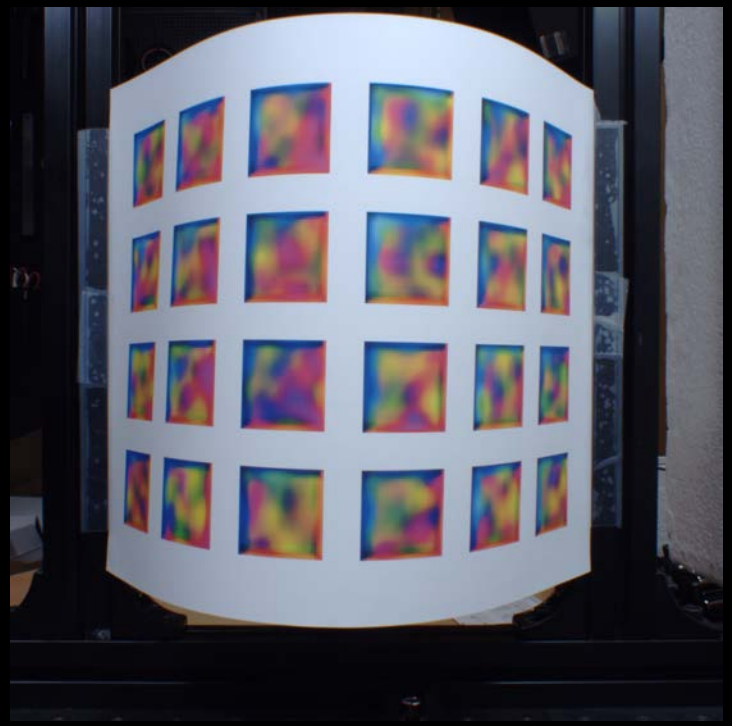
Deformation with motion blur (0.5x speed)

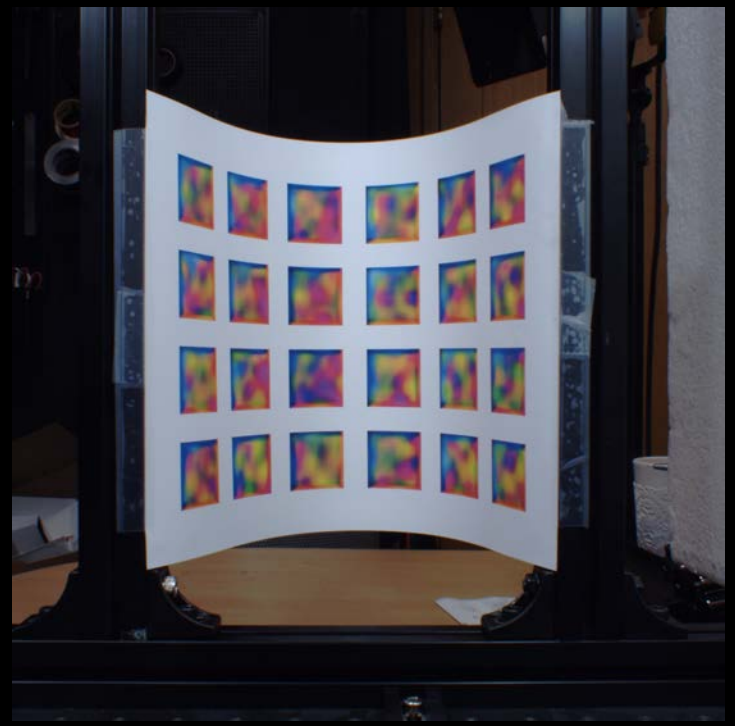


Marker detection results (AprilTag / ours)

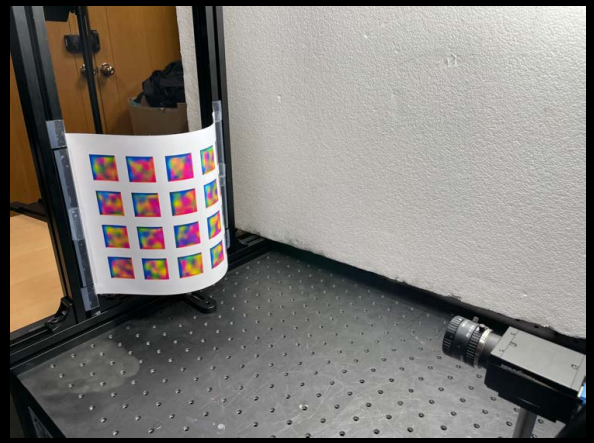
Results: Deformation (4/4)

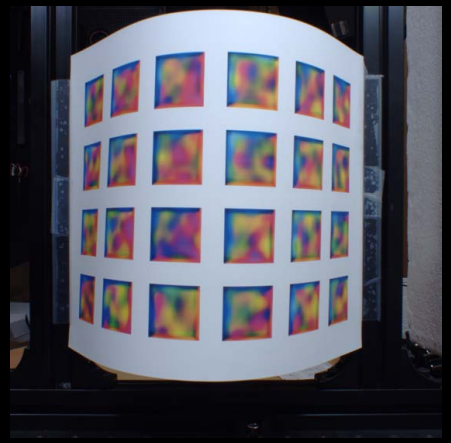


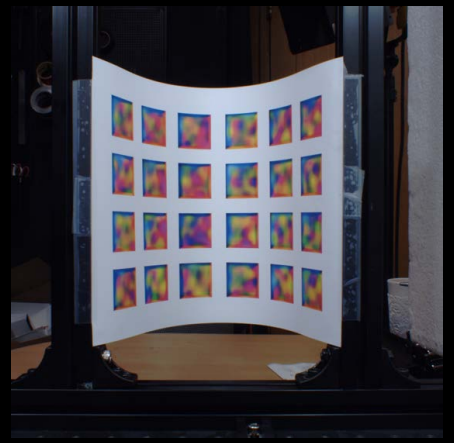


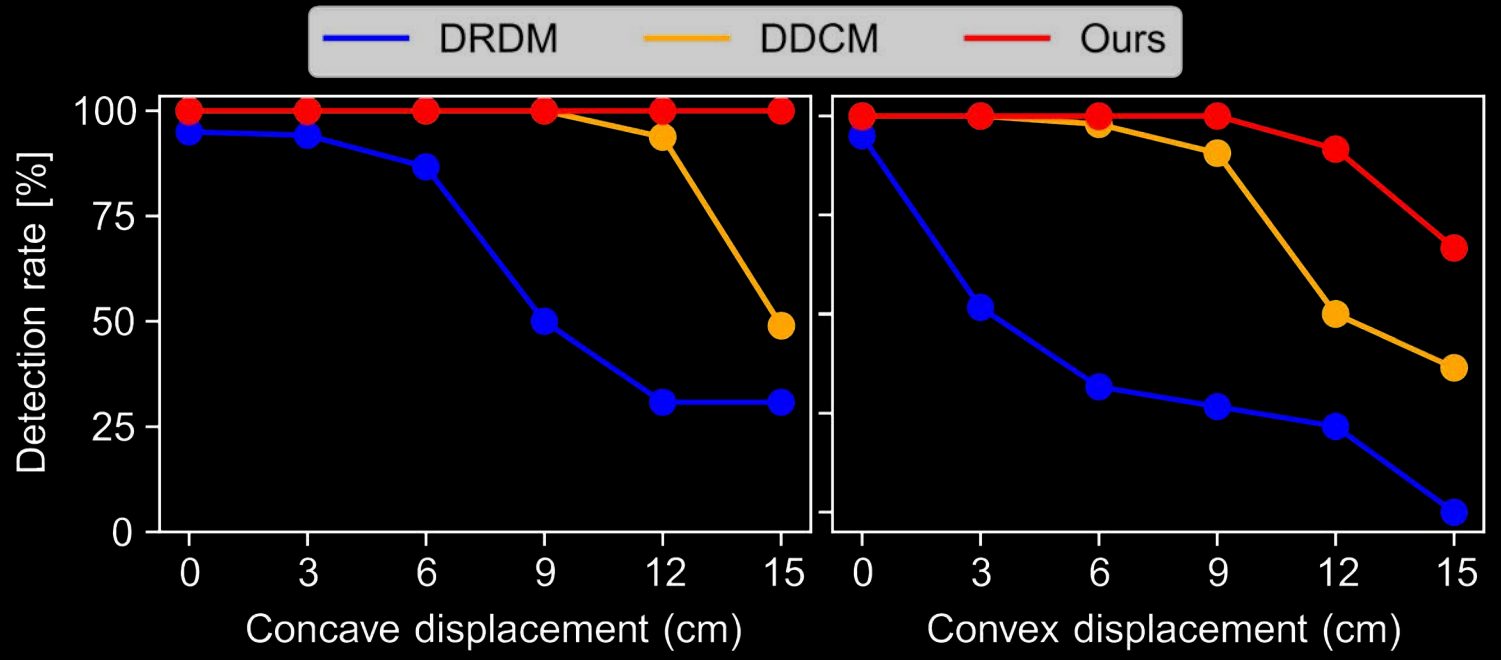


Results: Deformation (4/4)



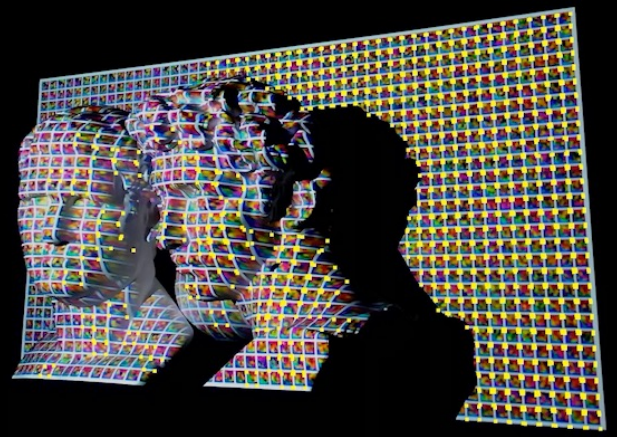






Applications: Structured Light 3D Imaging

Structured light 3D imaging with camera motion

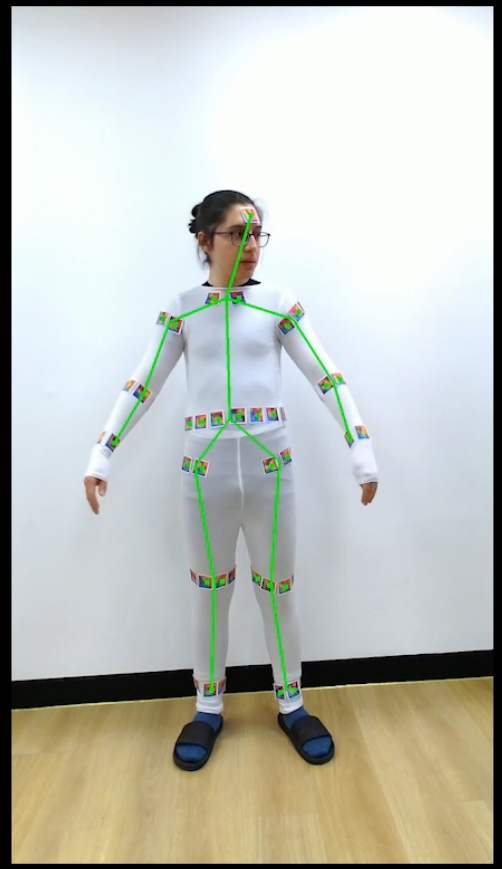


Input frames with our markers

3D points

Applications: Motion Capture

Our marker detection



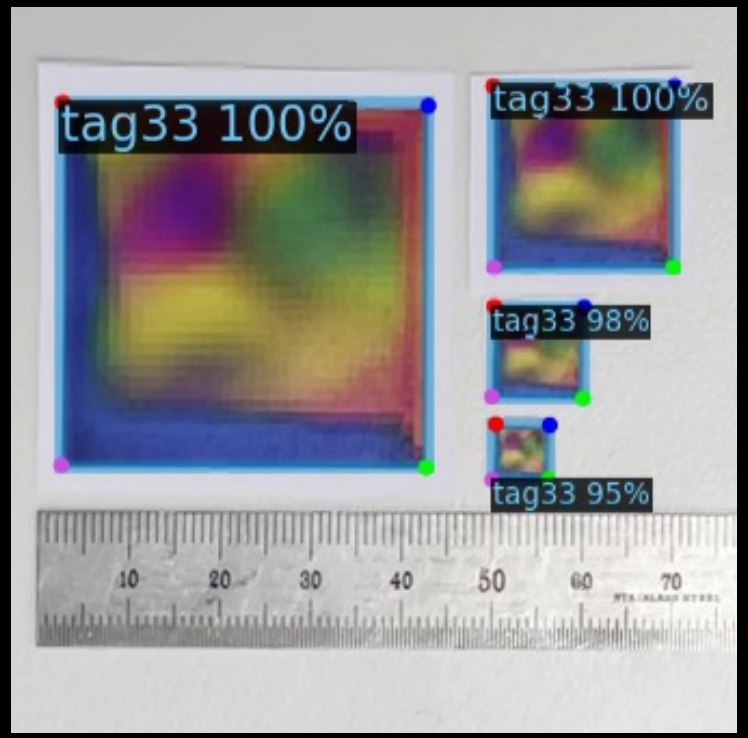
Human motion capture

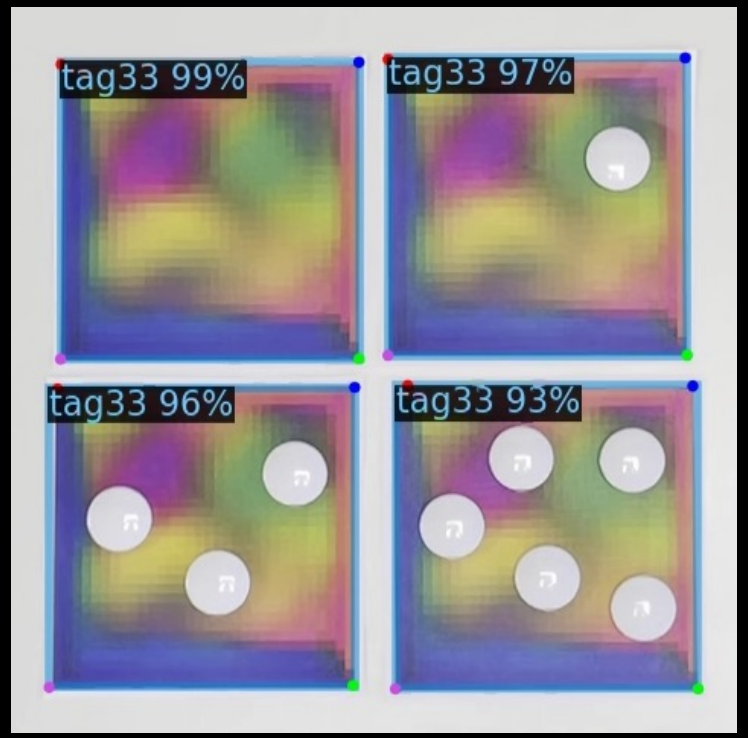
Applications: Augmented Reality

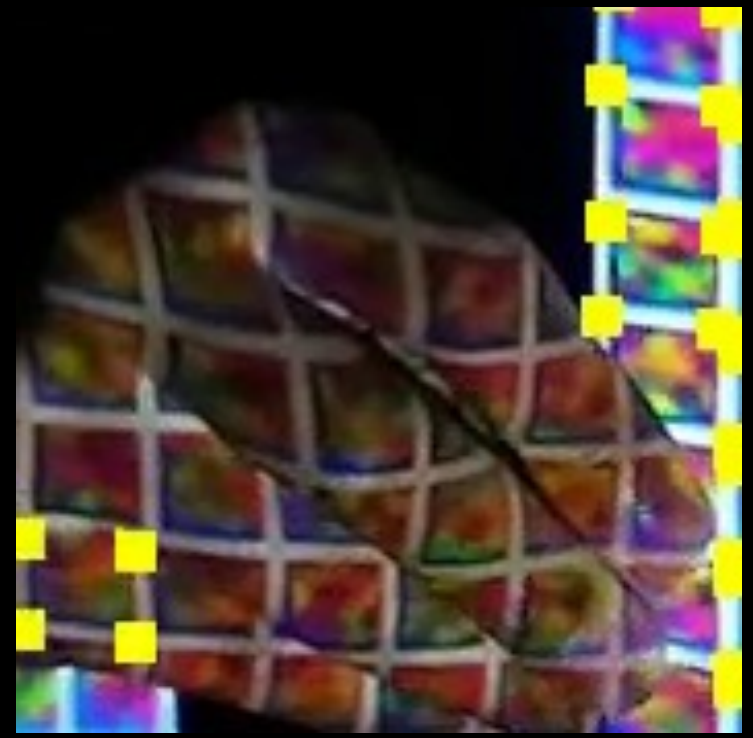
Smooth camera motion

Marker detection results (AprilTag / ours)

Limitations and Future Work







Conclusions

- **Deformable fiducial marker system**
  - End-to-end optimization of the marker generator and detector networks via photorealistic differentiable rendering
  - Deformed fiducial markers with strong motion blur
  - Large number of messages can be embedded

- Various applications demonstrated
  - Structured light 3D imaging
  - Human motion capture
  - Augmented reality rendering

SIGGRAPH 2021

THANK YOU