Escaping Plato’s Cave: 3D Shape From Adversarial Rendering

Philipp Henzler, Niloy J. Mitra, Tobias Ritschel

Presented by Arjun Karanam

Plato’s Allegory of the Cave

An allegory to the power of reason, and how living life without reason traps us in viewing just a small part of the world.

Motivation

Current limitation for generative models - Availability of training data

This problem is amplified in 3D, especially for edge cases

- “ShapeNet has lots of chairs, but no chantarelle”

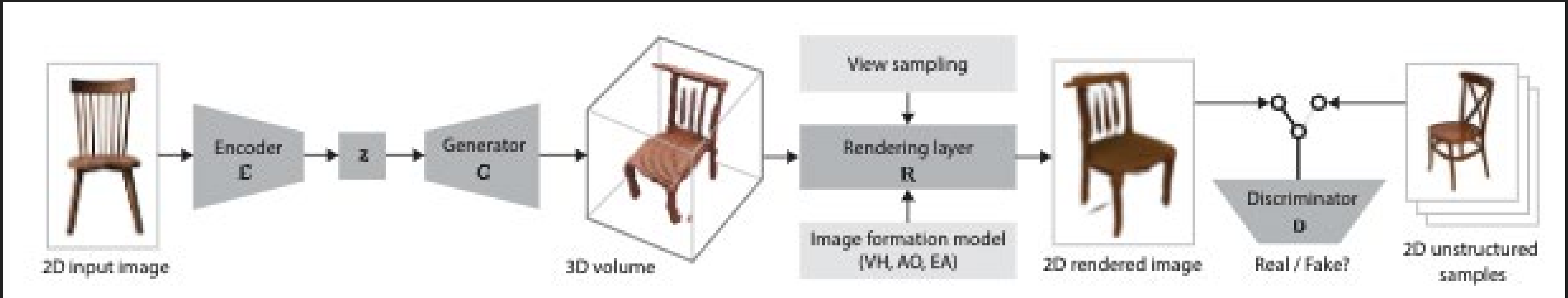
Solution

Generate 3D structure of an object class from an ***unstructured*** collection of 2D images

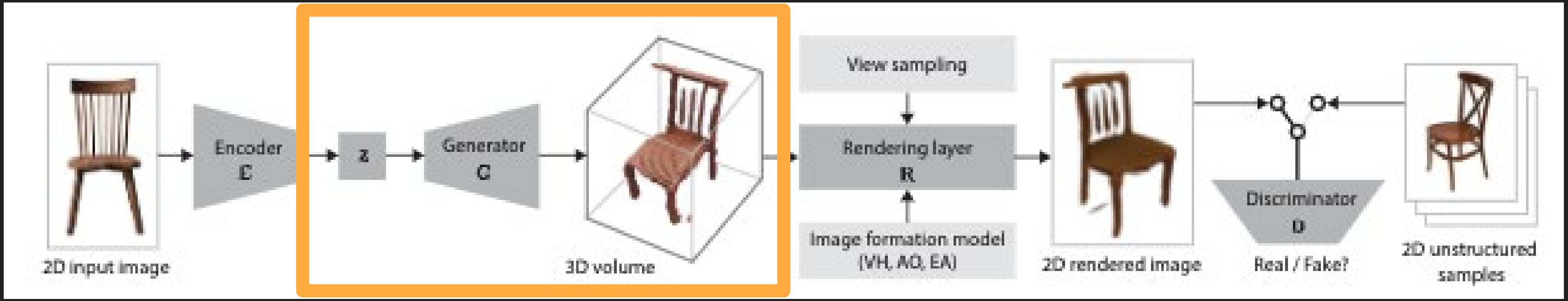
What is an Unstructured Collection?

- Annotation free
- Random Pose
- Just Single Views of an object

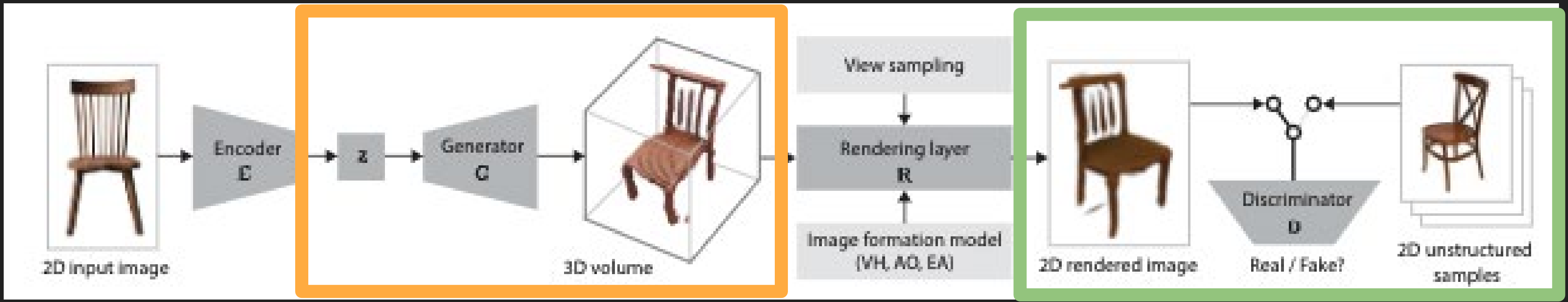
How? PlatonicGAN



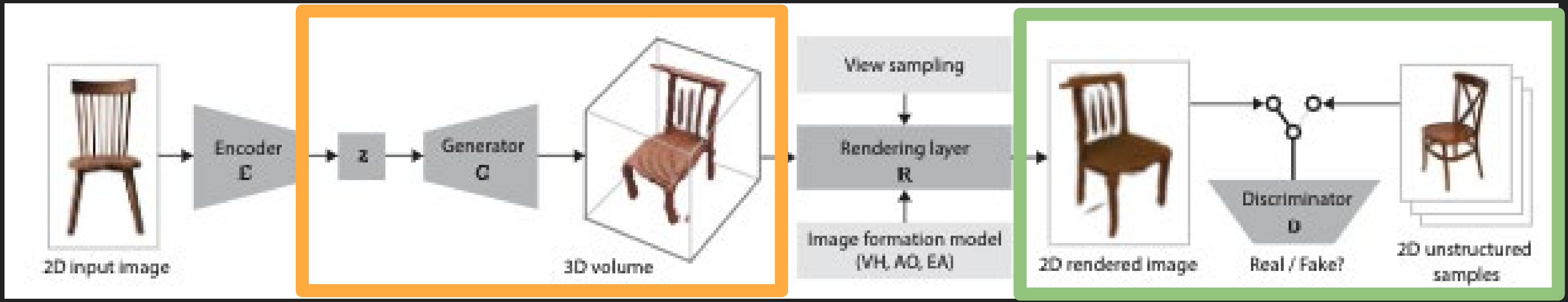
How? PlatnonicGAN



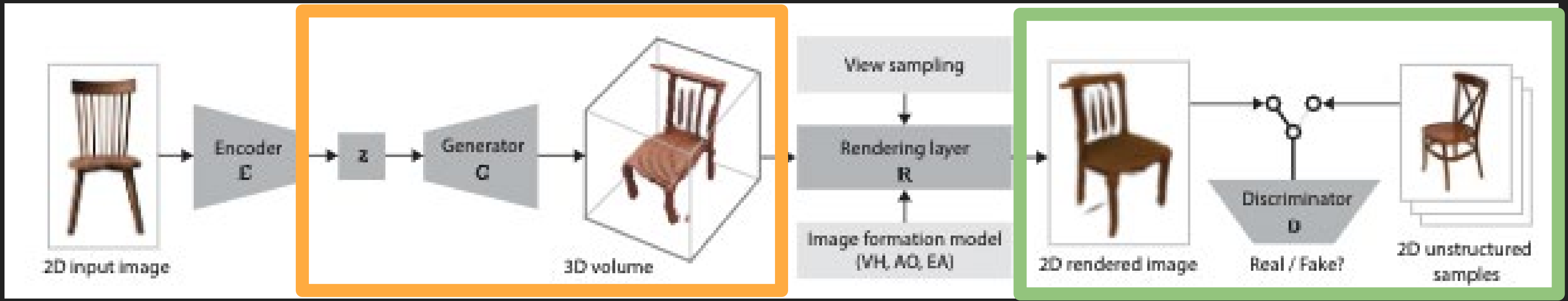
How? PlatonicGAN



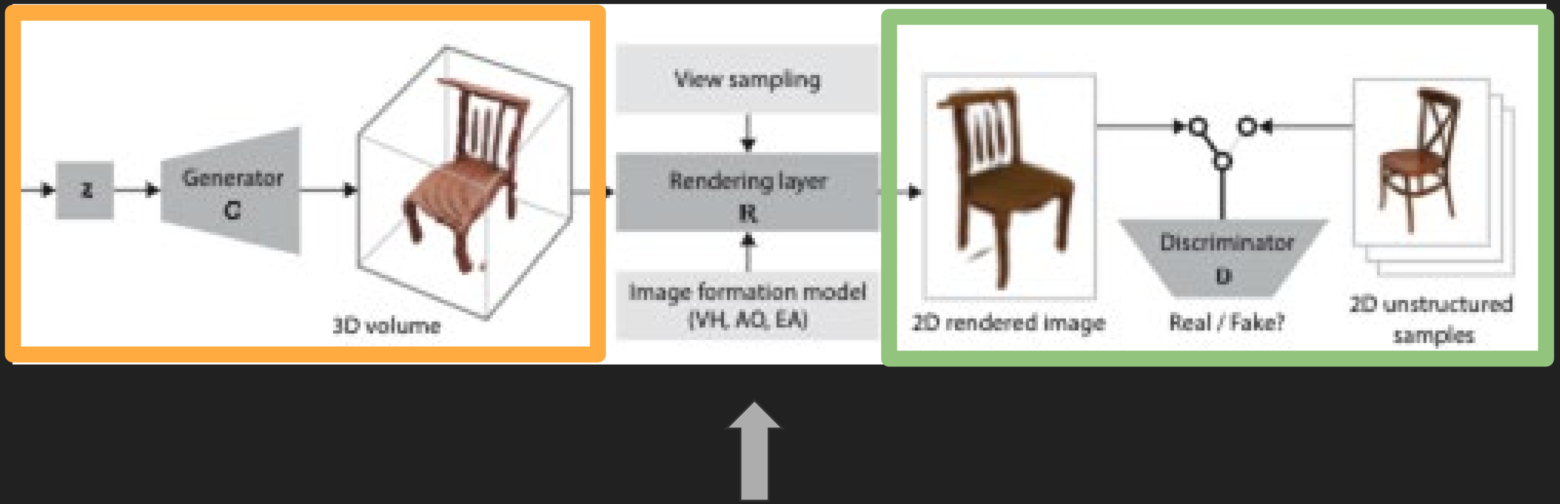
How? PlatonicGAN



How? PlatonicGAN



The Rendering Layer



TLDR - Projects the 3D Object down into 2D

How? - Comparing to a “normal” GAN

GAN

$$c_{\mathrm{Gen}'}(\Theta) = \mathbb{E}_{p_{\mathrm{Gen}}(\mathbf{z})}[\log(1 - D_\Psi(G_\Theta(\mathbf{z}))]$$

PlatonicGAN

$$c_{\mathrm{Gen}}(\Theta) = \mathbb{E}_{p_{\mathrm{Gen}}(\mathbf{z})}\mathbb{E}_{p_{\mathrm{View}}(\omega)}[\log(1 - D_\Psi(R(\omega, G_\Theta(\mathbf{z})))]$$

How? - Comparing to a “normal” GAN

GAN

$$c_{\mathrm{Gen}'}(\Theta) = \mathbb{E}_{p_{\mathrm{Gen}}(\mathbf{z})}[\log(1 - D_\Psi(G_\Theta(\mathbf{z}))]$$

PlatonicGAN

$$c_{\mathrm{Gen}}(\Theta) = \mathbb{E}_{p_{\mathrm{Gen}}(\mathbf{z})}\mathbb{E}_{p_{\mathrm{View}}(\omega)}[\log(1 - D_\Psi(R(\omega, G_\Theta(\mathbf{z})))]$$

$$R(\omega, \mathbf{v}) := \rho(\mathsf{T}(\omega)\mathbf{v})$$

Looking closer into the Rendering Layer

$$R(\omega, \mathbf{v}) := \rho(\mathrm{T}(\omega)\mathbf{v})$$

ω - View (Randomly Sampled)

T - Transformation Matrix

ρ - Image Transformation Function

Explored variants of ρ

Visual Hull

- A binary value indicating if any voxel blocked the ray

Absorption Only (AO)

Emission-Absorption (EA)

Figure 4. Different image formation models visual hull (VH), absorption-only (AO) and emission-absorption (EA).

Explored variants of ρ

Visual Hull

- A binary value indicating if any voxel blocked the ray

Absorption Only (AO)

- Visual Hull, but allows for “softer” attenuation of rays

Emission-Absorption (EA)

Figure 4. Different image formation models visual hull (VH), absorption-only (AO) and emission-absorption (EA).

Explored variants of ρ

Visual Hull

- A binary value indicating if any voxel blocked the ray

Absorption Only (AO)

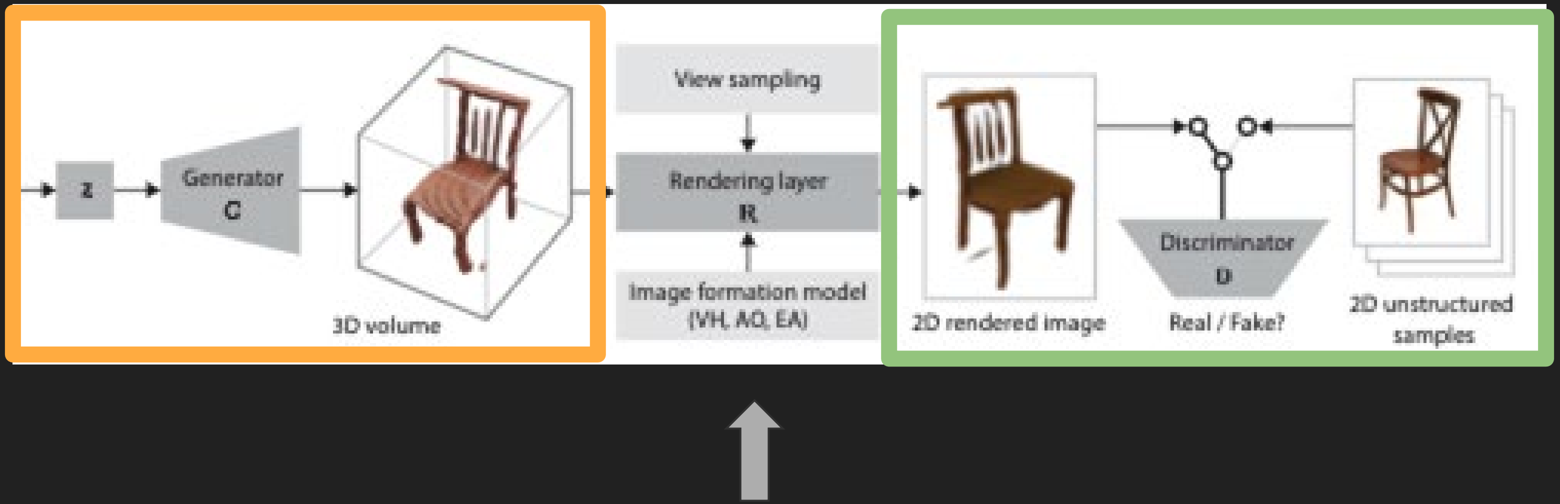
- Visual Hull, but allows for “softer” attenuation of rays

Emission-Absorption (EA)

- Adds emission, which allows for occlusion modelling

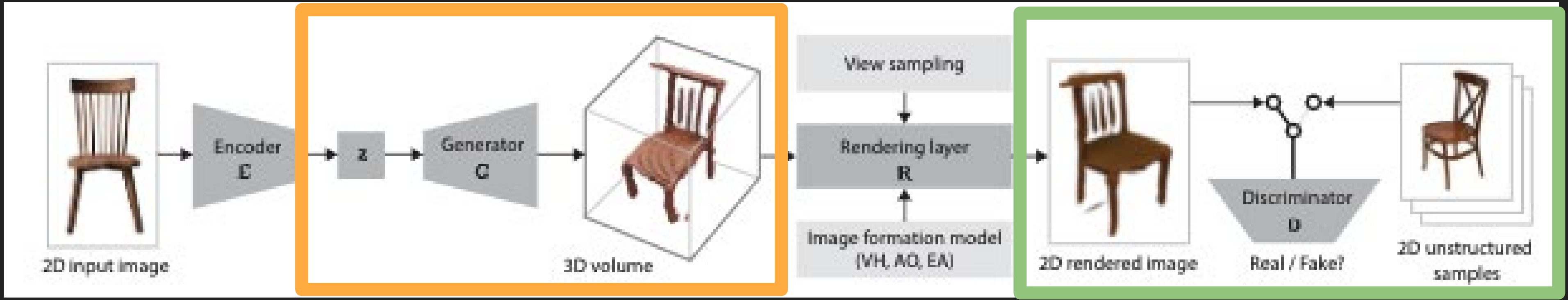
Figure 4. Different image formation models visual hull (VH), absorption-only (AO) and emission-absorption (EA).

The Rendering Layer



TLDR - Projects the 3D Object down into 2D

How?


2D input image
Encoder
E
z
Generator
G
3D volume
View sampling
Rendering layer
R
Image formation model
(VH, AO, EA)
2D rendered image
Discriminator
D
Real / Fake?
2D unstructured
samples

How?

Step 1 - Add an Encoder (with associated input images)

How?

Step 1 - Add an Encoder (with associated input images)

Step 2 - Add the cost of reconstruction to your min-max

$$\min_{\Psi} \max_{\Theta,\Phi} \; c_{\mathrm{Disc}}(\Psi) + c_{\mathrm{Gen}}(\Theta,\Phi) + \lambda c_{\mathrm{Rec}}(\Theta,\Phi),$$

How?

Step 1 - Add an Encoder (with associated input images)

Step 2 - Add the cost of reconstruction to your min-max

$$\min_{\Psi} \max_{\Theta,\Phi} \; c_{\text{Disc}}(\Psi) + c_{\text{Gen}}(\Theta,\Phi) + \lambda c_{\text{Rec}}(\Theta,\Phi),$$

Step 3 - Modify your cost of generation

$$c_{\text{Gen}}(\Theta) = \mathbb{E}_{p_{\text{Gen}}(\mathbf{z})}\mathbb{E}_{p_{\text{View}}(\omega)}[\log(1 - D_{\Psi}(R(\omega, G_{\Theta}(\mathbf{z}))))],$$

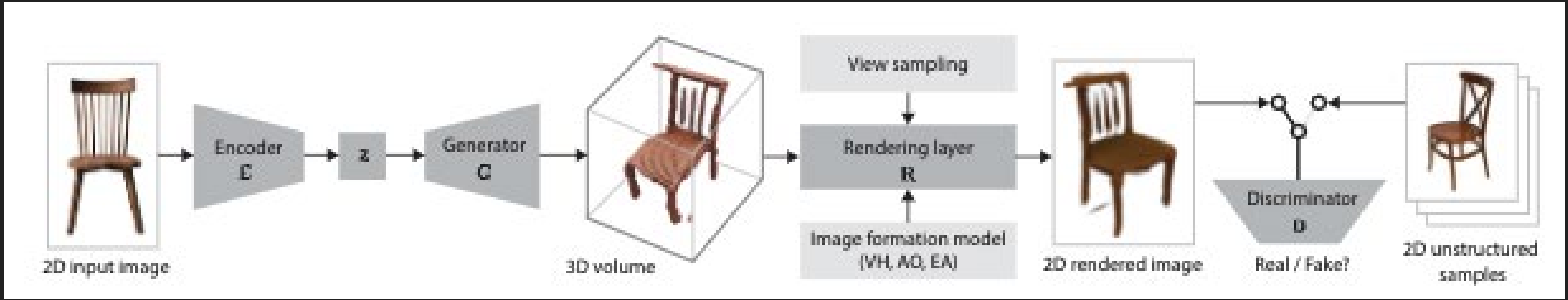
$$c_{\text{Gen}}(\Theta,\Phi) = \mathbb{E}_{p_{\text{Dat}}(\mathbf{I})}\mathbb{E}_{p_{\text{View}}(\omega)}[\log(1 - D_{\Psi}(R(\omega, G_{\Theta}(E_{\Phi}(\mathbf{I})))))].$$

How? (Continued)

Step 4 - Define the cost of Reconstruction

$$c_{\mathrm{Rec}}(\Theta, \Phi) = \|\mathbf{y} - R(\omega_0, G_\Theta(E_\Phi(\mathbf{I})))\|_2^2$$

PlatonicGAN

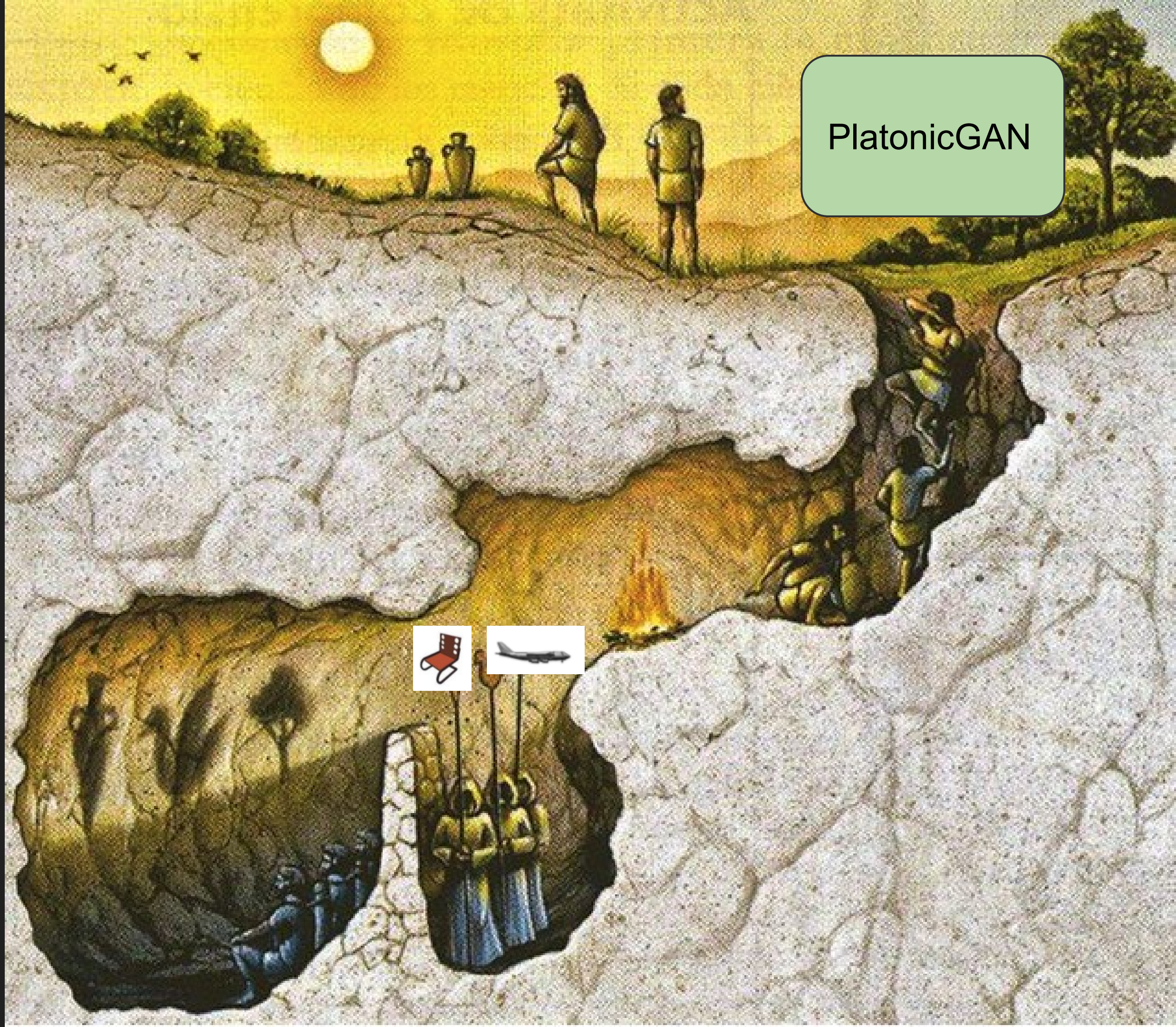


Plato's Allegory of the Cave (but AI)

How can we take unstructured 2D Images and use that to learn about the real world (the world outside the cave?)

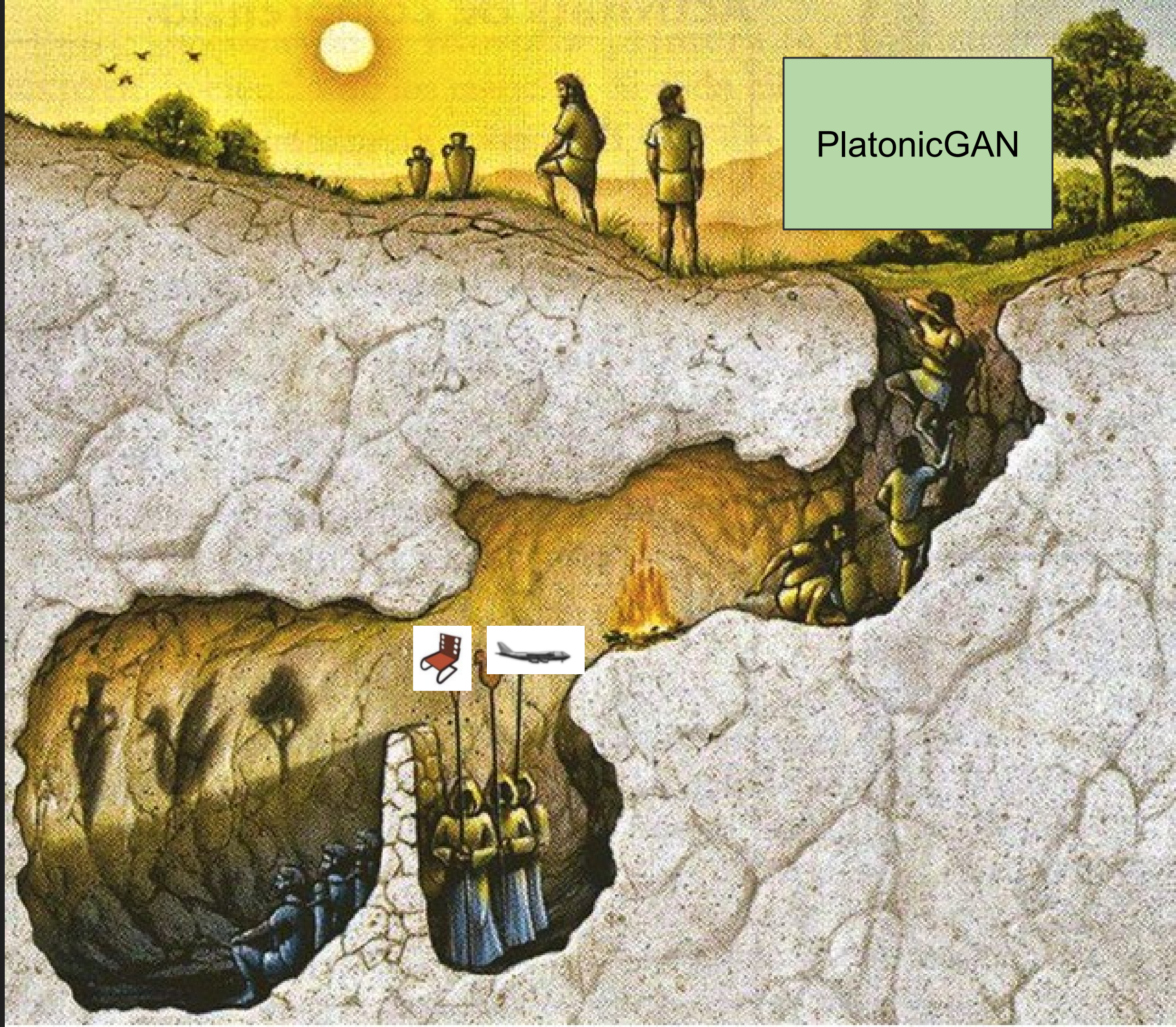
Plato’s Allegory of the Cave (but AI)

How can we take unstructured 2D Images and use that to learn about the real world (the world outside the cave?)



Plato's Allegory of the Cave (but AI)

How can we take unstructured 2D Images and use that to learn about the real world (the world outside the cave?)



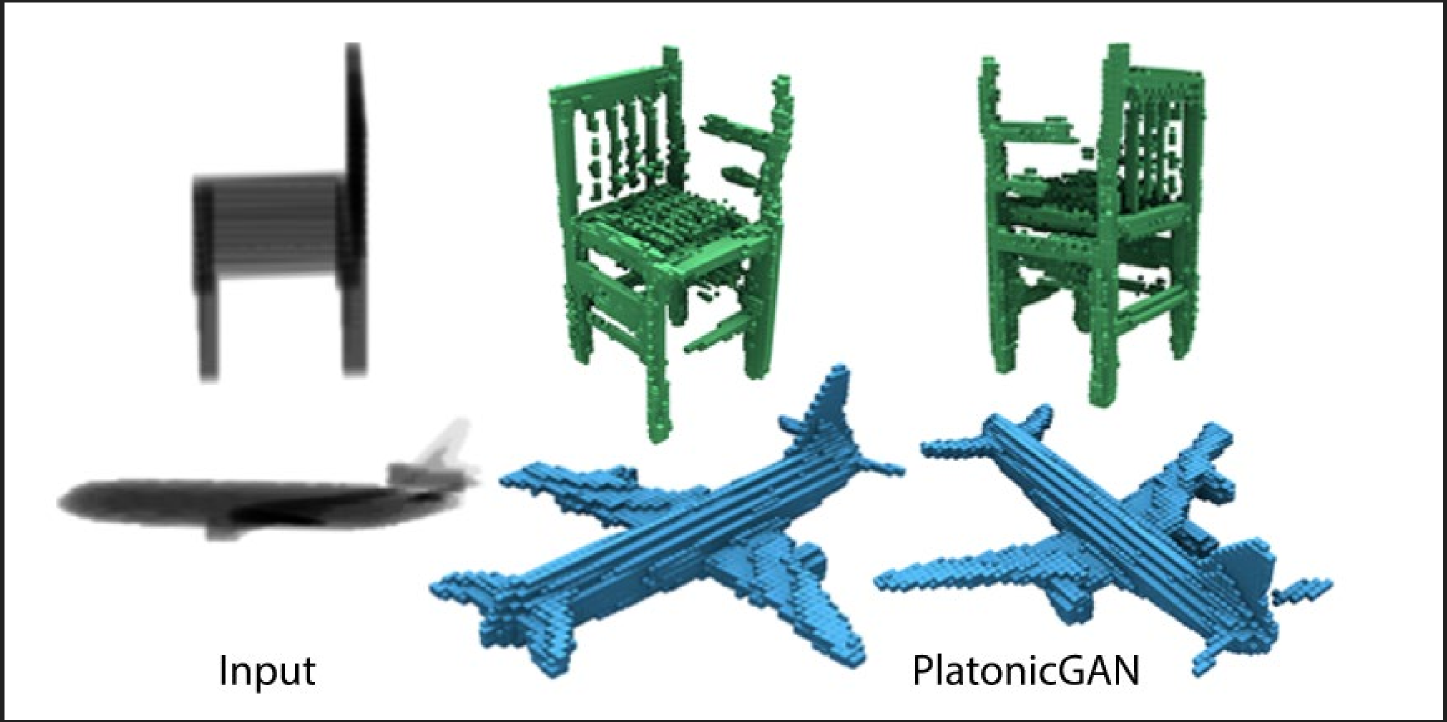
Results

Method	IF	Superv. 2D	Superv. 3D	VH DSSIM	VH VGG	AO DSSIM	AO VGG	EA DSSIM	EA VGG	VOX DSSIM	VOX VGG	ISO DSSIM	ISO VGG	3D Volume RMSE	3D Volume IoU	3D Volume CD	FID EA
PrGAN [11]	VH	✓	×	1.55	6.57	1.37	**4.85**	1.41	**4.63**	1.68	5.41	1.83	6.15	**7.46**	0.11	**0.22**	207
Ours	VH	✓	×	**1.14**	**5.37**	**1.16**	4.93	**1.12**	4.68	**1.33**	**5.22**	**1.28**	**5.96**	9.16	**0.20**	0.55	**55**
Mult.-View [38]	VH	✓	×	0.87	4.89	0.80	4.31	0.90	4.07	1.38	4.83	1.21	5.56	5.37	0.36	**0.29**	155
3DGAN [36]	VH	✓	×	0.83	5.01	**0.75**	4.02	0.86	**3.83**	1.30	4.73	1.17	5.82	**4.97**	**0.46**	0.48	111
Ours 3D	VH	✓	×	**0.81**	**4.82**	0.77	**3.98**	**0.83**	**3.83**	**1.18**	**4.59**	**1.09**	**5.50**	5.20	0.44	0.42	**98**
PrGAN [11]	AO	✓	×	1.41	6.40	1.27	4.80	1.27	4.52	1.53	5.32	1.63	6.00	7.11	0.09	**0.16**	190
Ours	AO	✓	×	**0.94**	**5.35**	**0.93**	**4.46**	**0.91**	**4.26**	**1.11**	**4.96**	**1.09**	**5.75**	**5.70**	**0.27**	0.36	**90**
Mult.-View [38]	AO	✓	×	0.95	4.99	0.78	4.23	0.91	4.01	1.51	4.92	1.29	5.39	**4.89**	0.34	**0.28**	165
3DGAN [36]	AO	✓	×	0.67	4.37	0.69	3.77	0.72	3.57	0.99	**4.25**	0.97	**4.92**	5.08	**0.43**	0.50	**58**
Ours 3D	AO	✓	×	**0.66**	**4.36**	**0.66**	**3.73**	**0.70**	**3.52**	**0.98**	4.28	**0.96**	4.94	5.17	0.37	0.53	64
PrGAN [11]	EA	✓	×	**1.31**	**6.22**	**1.15**	**4.77**	**1.16**	**5.37**	**1.36**	**6.71**	**1.47**	**7.07**	**6.80**	0.08	**0.12**	196
Ours	EA	✓	×	2.18	6.53	1.99	5.38	1.89	6.00	2.21	7.43	2.36	7.92	14.13	**0.13**	1.24	**181**
Mult.-View [38]	EA	✓	×	1.62	6.21	1.53	4.58	1.63	5.48	1.95	6.97	1.94	7.41	15.05	0.12	2.52	172
3DGAN [36]	EA	✓	×	0.89	5.28	**0.78**	**3.93**	0.98	4.79	1.29	6.76	1.30	7.09	**5.24**	**0.46**	**0.47**	110
Ours 3D	EA	✓	×	**0.82**	**4.71**	0.82	3.96	**0.97**	**4.77**	**1.12**	**6.12**	**1.16**	**6.47**	7.43	0.04	1.10	**73**

Results (but prettier)

Figure 5. Visual results for 3D reconstruction of three classes (`airplane`, `chair`, `rifle`) from multiple views.

Results - Failure Cases



What's next?

Address failure cases (object-space vs. view-space reconstruction)

Explore different types of differential rendering (more options for ρ)

Apply to shape completion?

Apply to different dimensions (4D to 3D, where 4D is motion)

Questions?