Robotics and
Visual Intelligence Lab

PixelSynth: Generating a 3D-Consistent Experience from a Single Image

ICCV 2021

Chris Rockwell David F. Fouhey Justin Johnson
University of Michigan

Gyeongsu Cho @ UNIST

Lab Seminar

2022.03.03 (Thr)

Introduction

Network

Experiments

Conclusion

Introduction

View Synthesis?

Introduction

View Synthesis?

Introduction

View Synthesis?

Introduction

Single Image View Synthesis

Introduction

Single Image View Synthesis

Introduction

Single Image View Synthesis

Translation: Step Forward
Rotation: Turn left

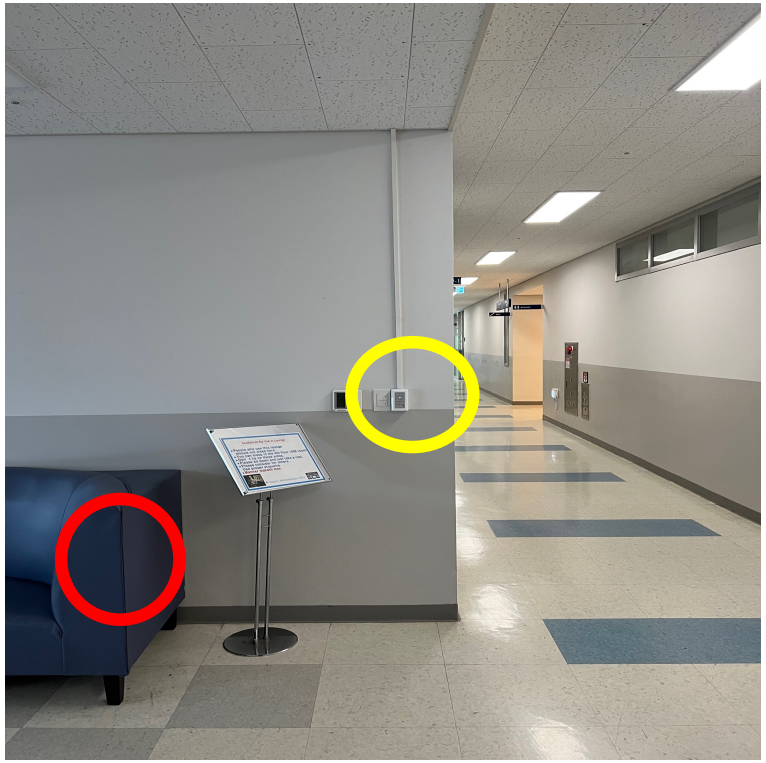
Introduction

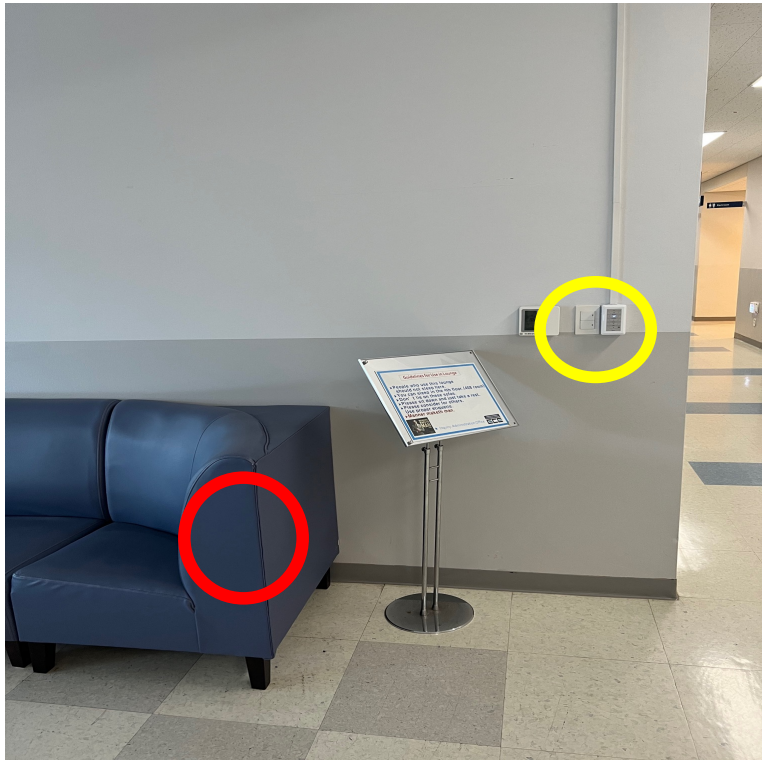
Single Image View Synthesis

Translation: Step Forward
Rotation: Turn left

Introduction

Challenges of Single Image View Synthesis





Challenge 1: Need to know depth

Introduction

Challenges of Single Image View Synthesis

Challenge 2: Inpainting missing regions

Introduction

Challenges of Single Image View Synthesis

Challenge 3: Outpainting missing regions consistently

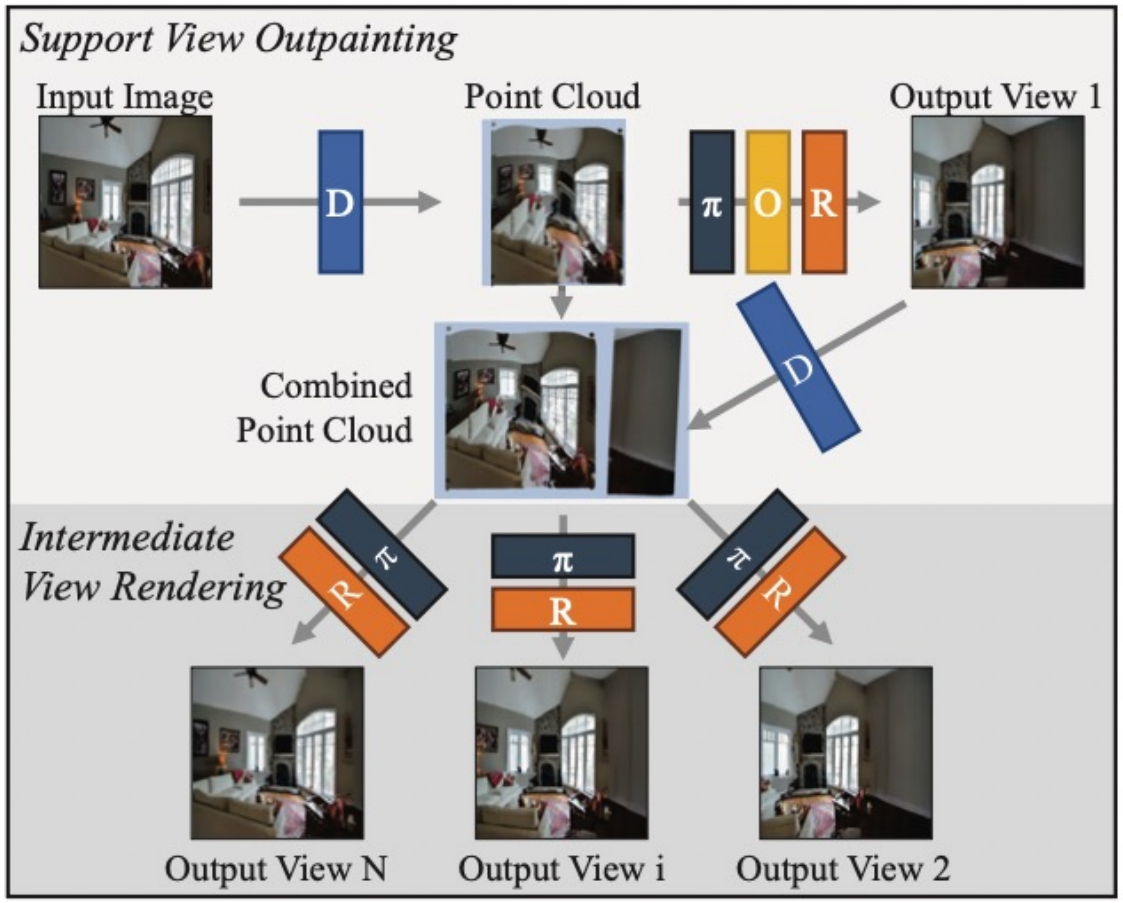
Introduction

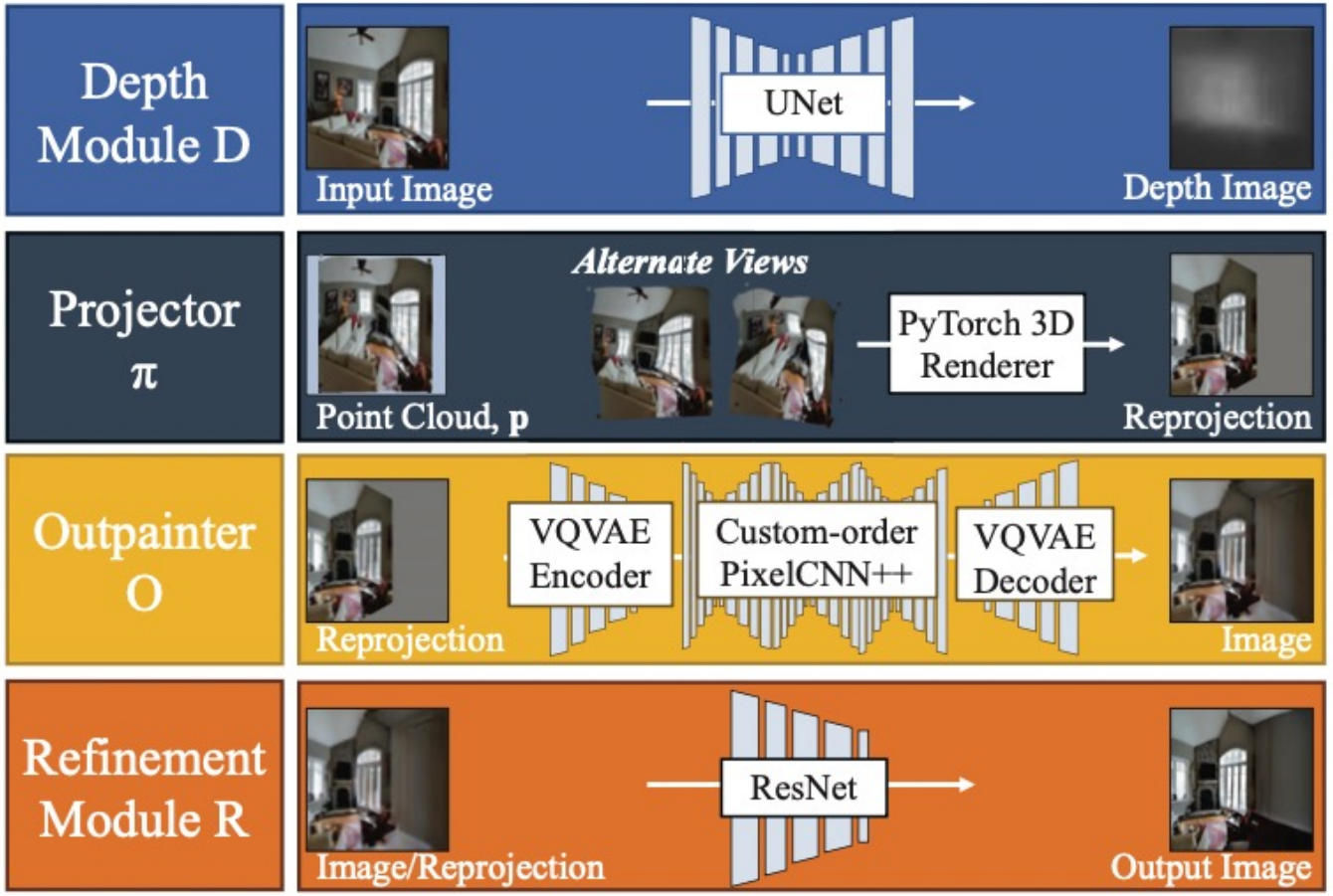
Network

Experiments

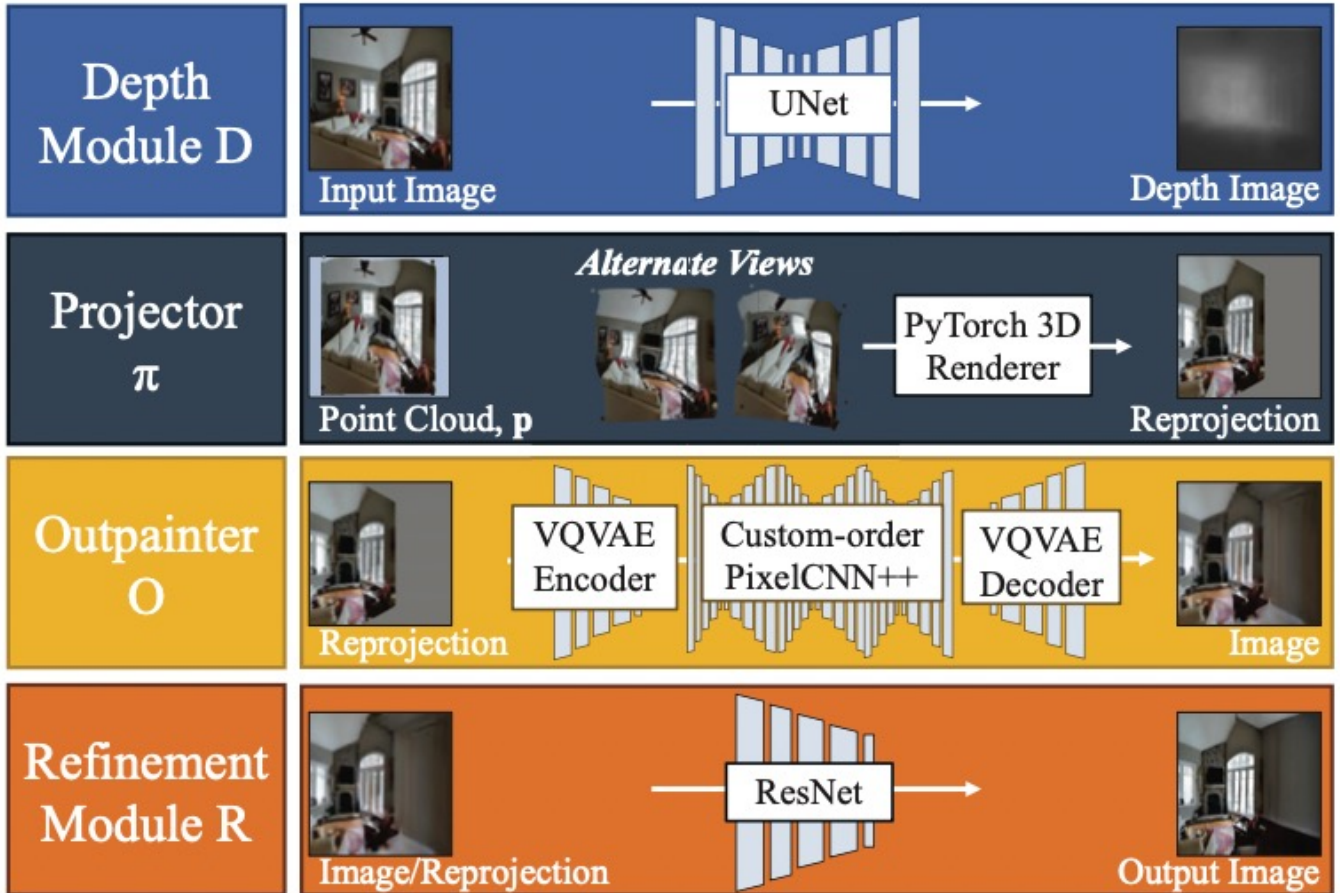
Conclusion

Network

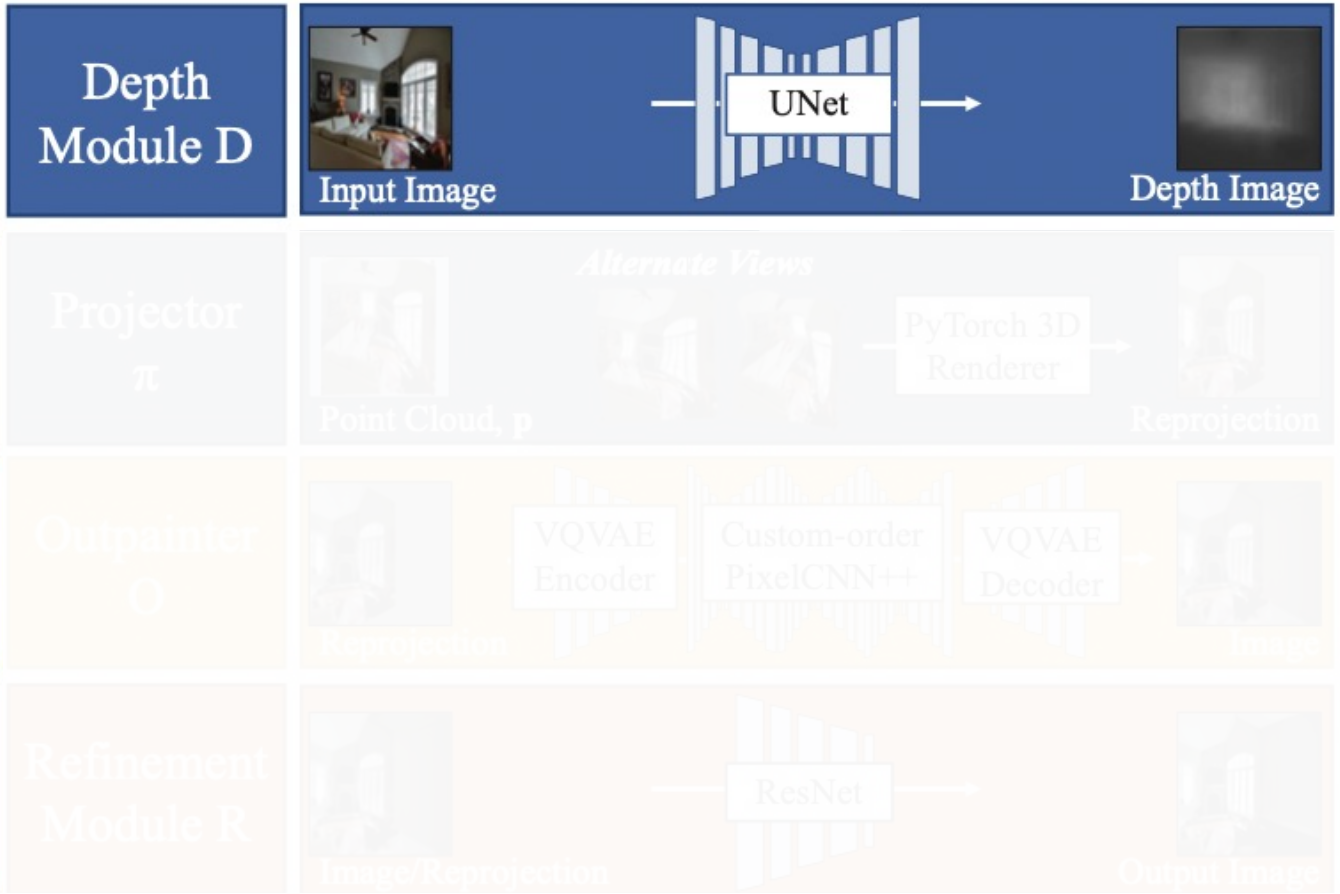

Support View Outpainting
Input Image
D
Point Cloud
π
O
R
Output View 1
Combined
Point Cloud
D
Intermediate
View Rendering
R
π
π
R
π
R
Output View N
Output View i
Output View 2


Depth
Module D
Input Image
UNet
Depth Image
Projector
π
Alternate Views
Point Cloud, p
PyTorch 3D
Renderer
Reprojection
Outpainter
O
Reprojection
VQVAE
Encoder
Custom-order
PixelCNN++
VQVAE
Decoder
Image
Refinement
Module R
Image/Reprojection
ResNet
Output Image

Network


Depth Module D
Input Image
UNet
Depth Image
Projector π
Alternate Views
Point Cloud, p
PyTorch 3D Renderer
Reprojection
Outpainter O
Reprojection
VQVAE Encoder
Custom-order PixelCNN++
VQVAE Decoder
Image
Refinement Module R
Image/Reprojection
ResNet
Output Image

Network



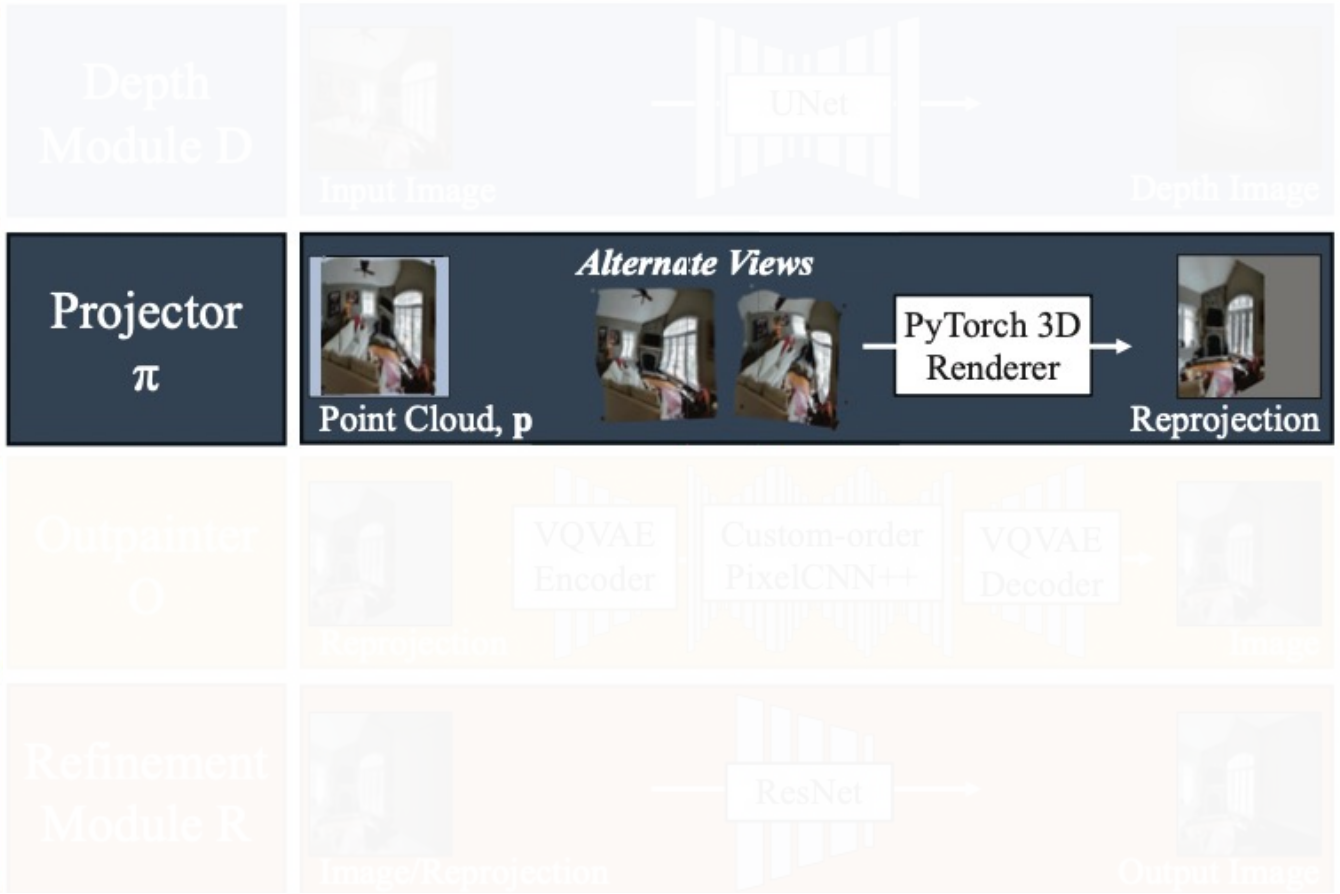
$$C = D(I)$$

Point Cloud

Depth Module

Image

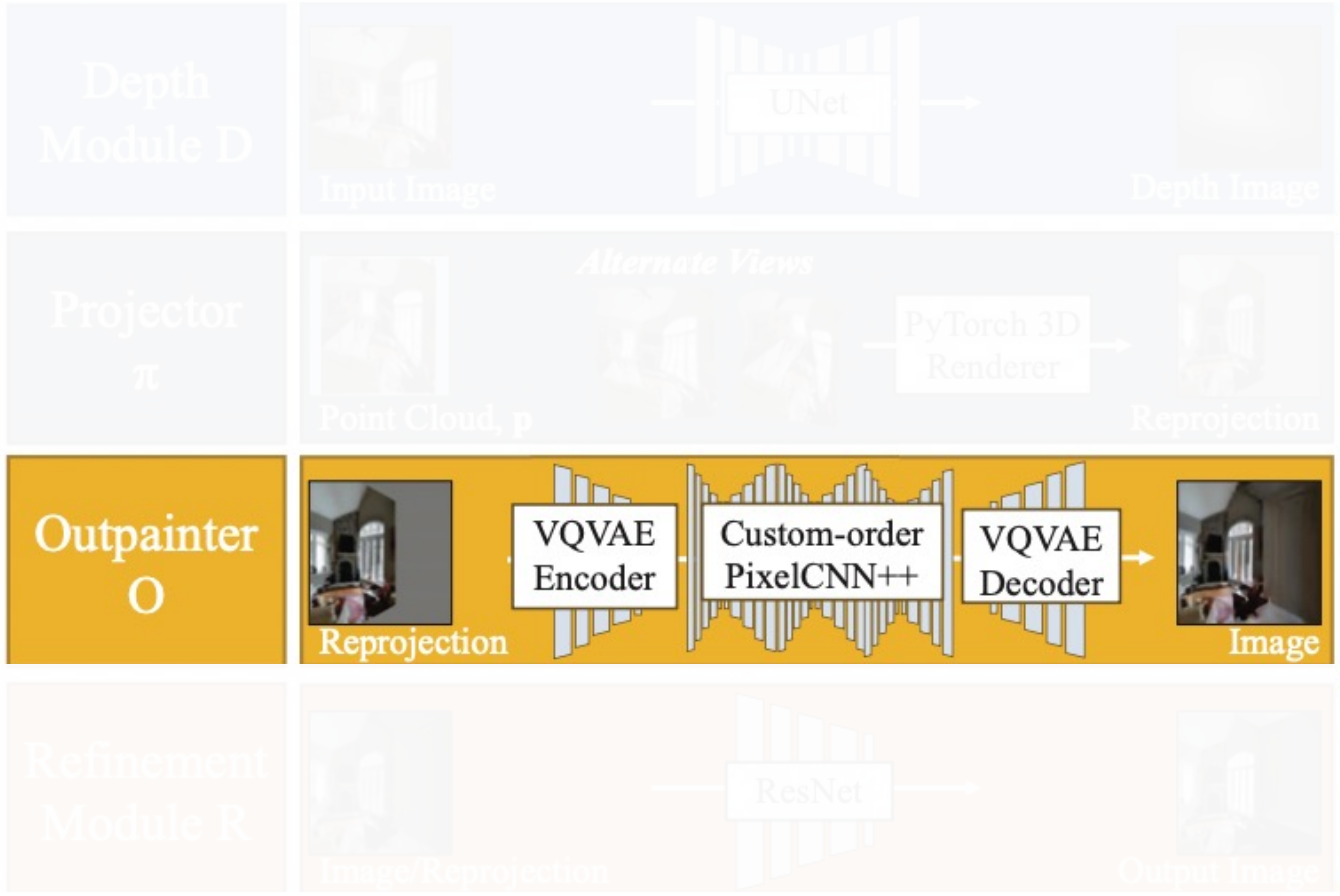
Network



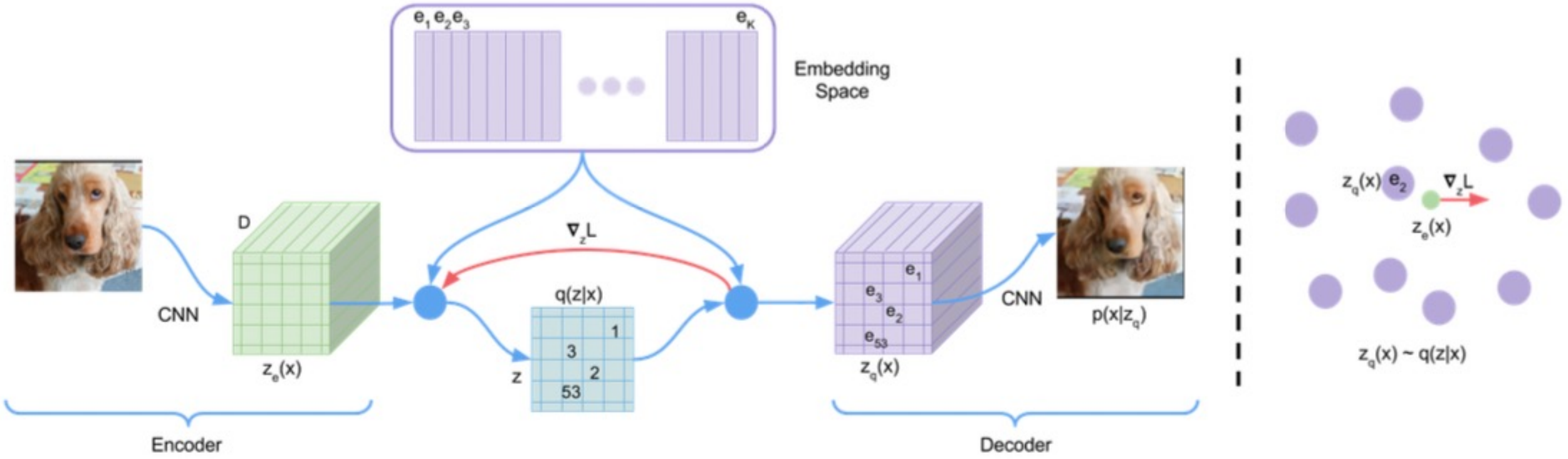
$$I = \pi(C, \mathbf{p})$$

Image — Projector — Point Cloud — Target pose

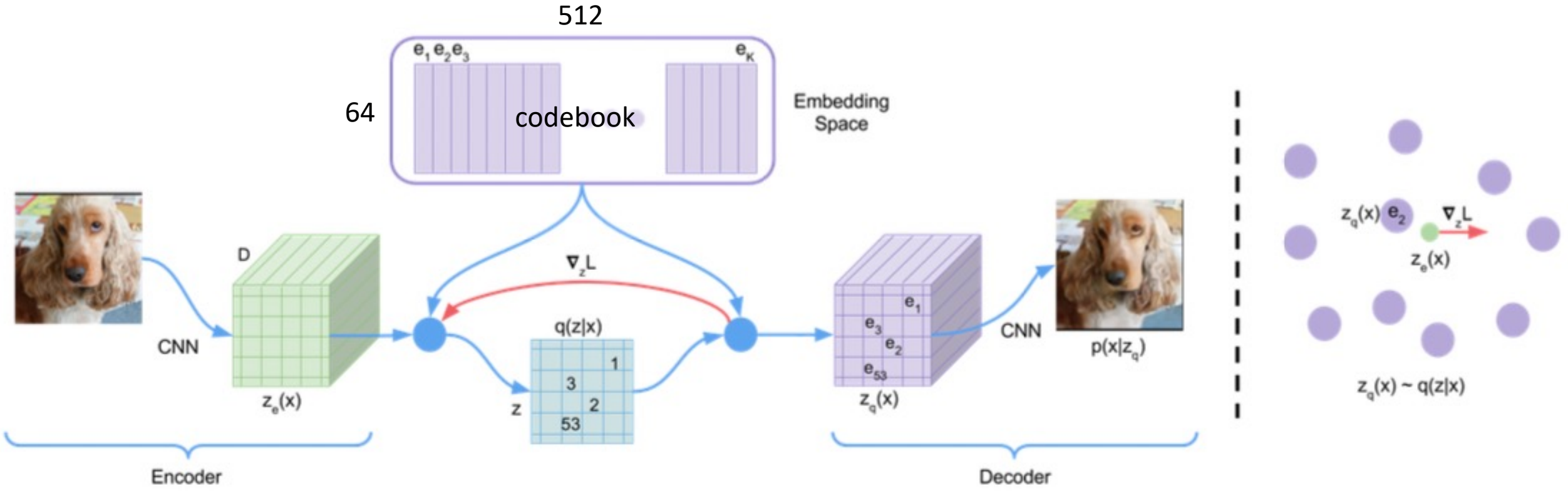
Network


Depth Module D
Input Image
UNet
Depth Image
Projector π
Alternate Views
Point Cloud, p
PyTorch 3D Renderer
Reprojection
Outpainter O
Reprojection
VQVAE Encoder
Custom-order PixelCNN++
VQVAE Decoder
Image
Refinement Module R
Image/Reprojection
ResNet
Output Image

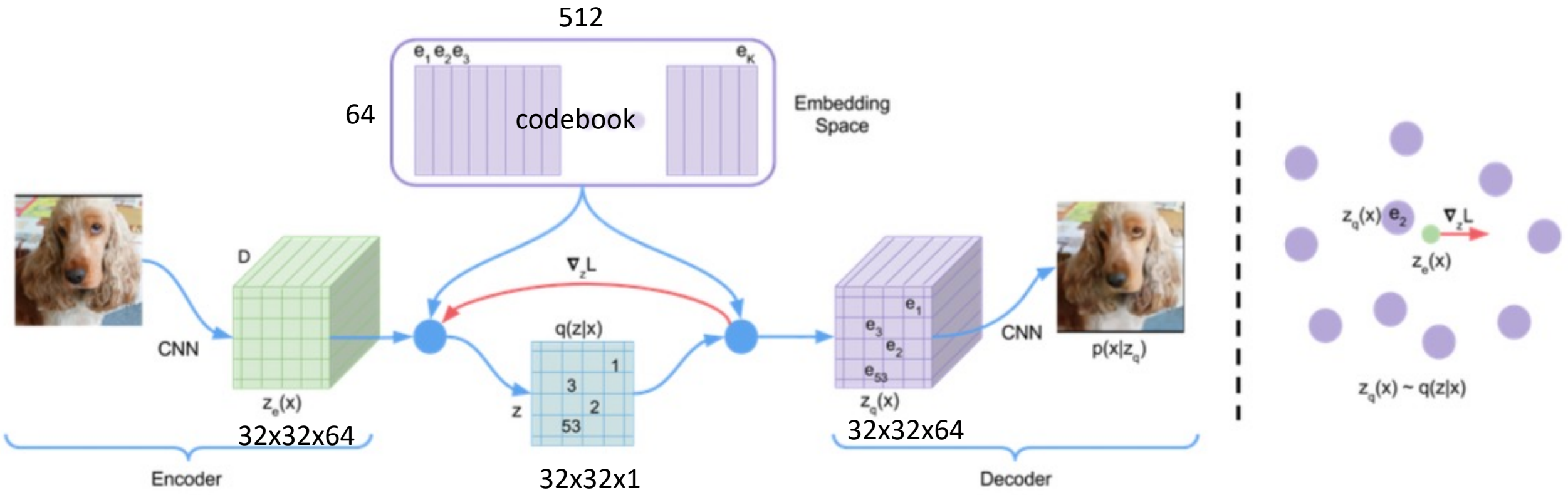
Network-VQVAE



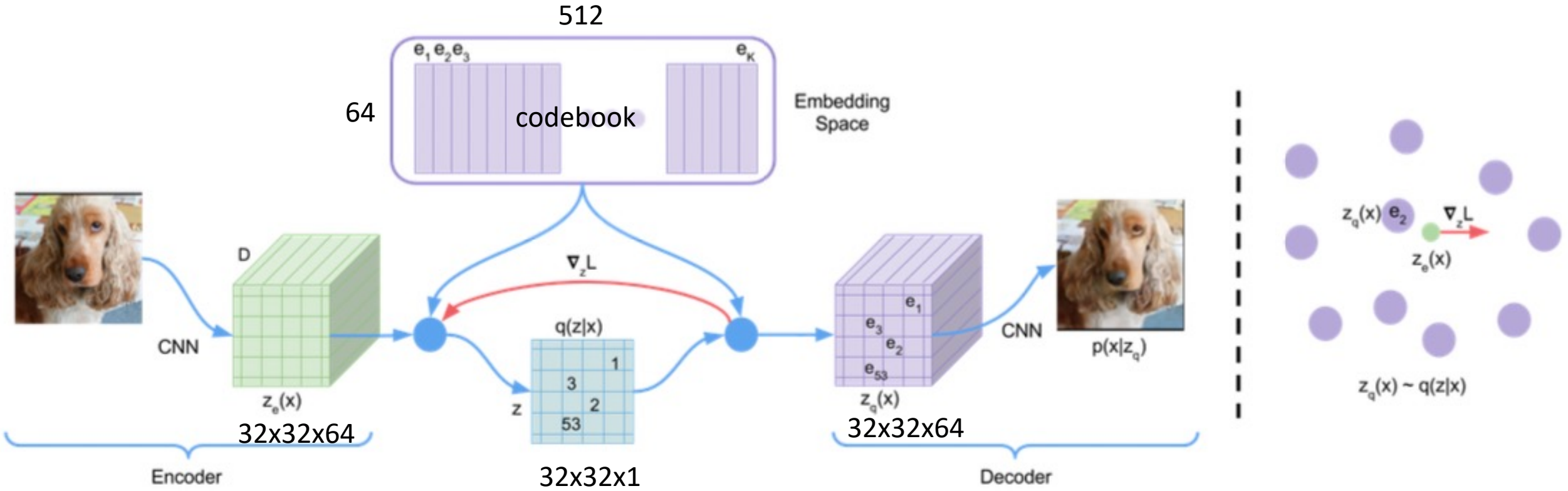
Network-VQVAE



Network-VQVAE



Network-VQVAE

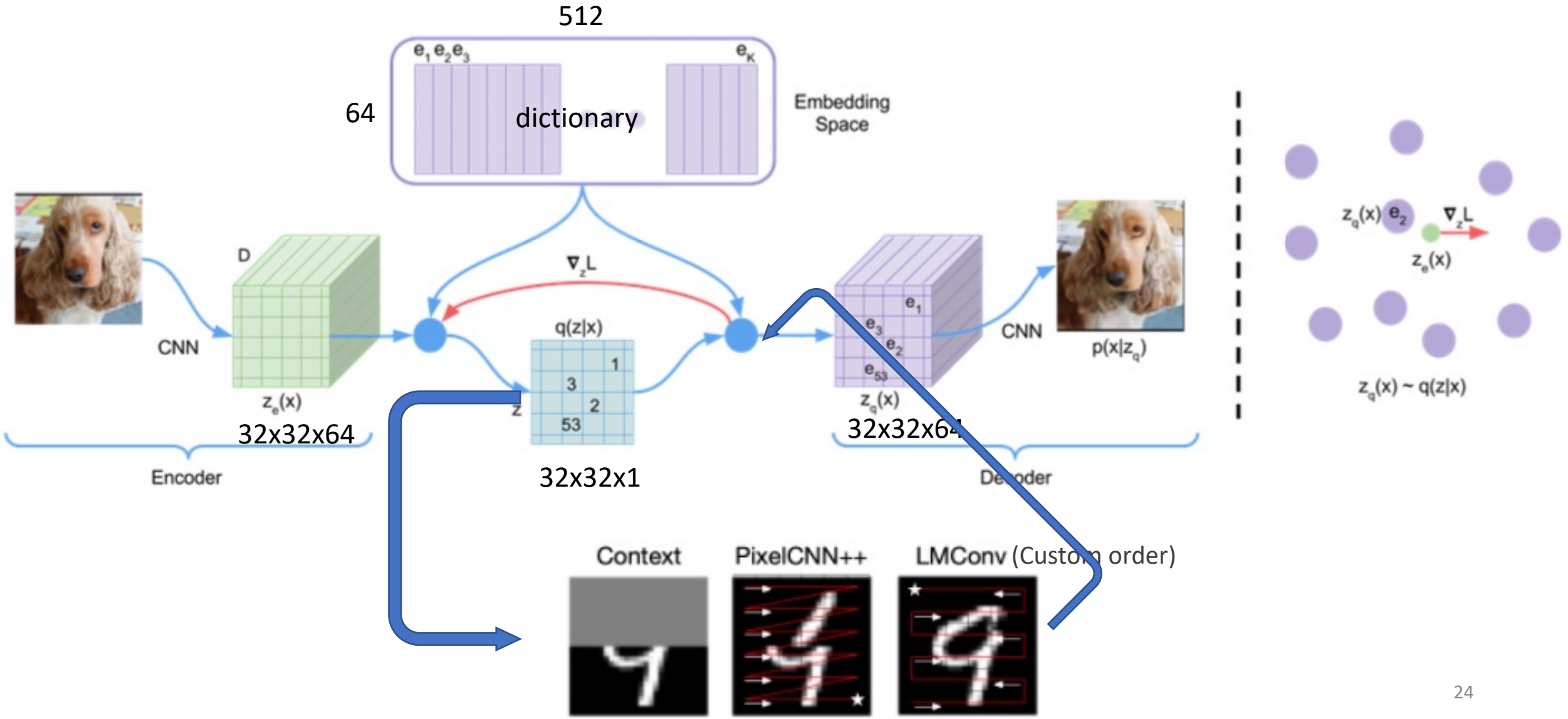


$$L = \log p(x|z_q(x)) + \|\text{sg}[z_e(x)] - e\|_2^2 + \beta\|z_e(x) - \text{sg}[e]\|_2^2,$$

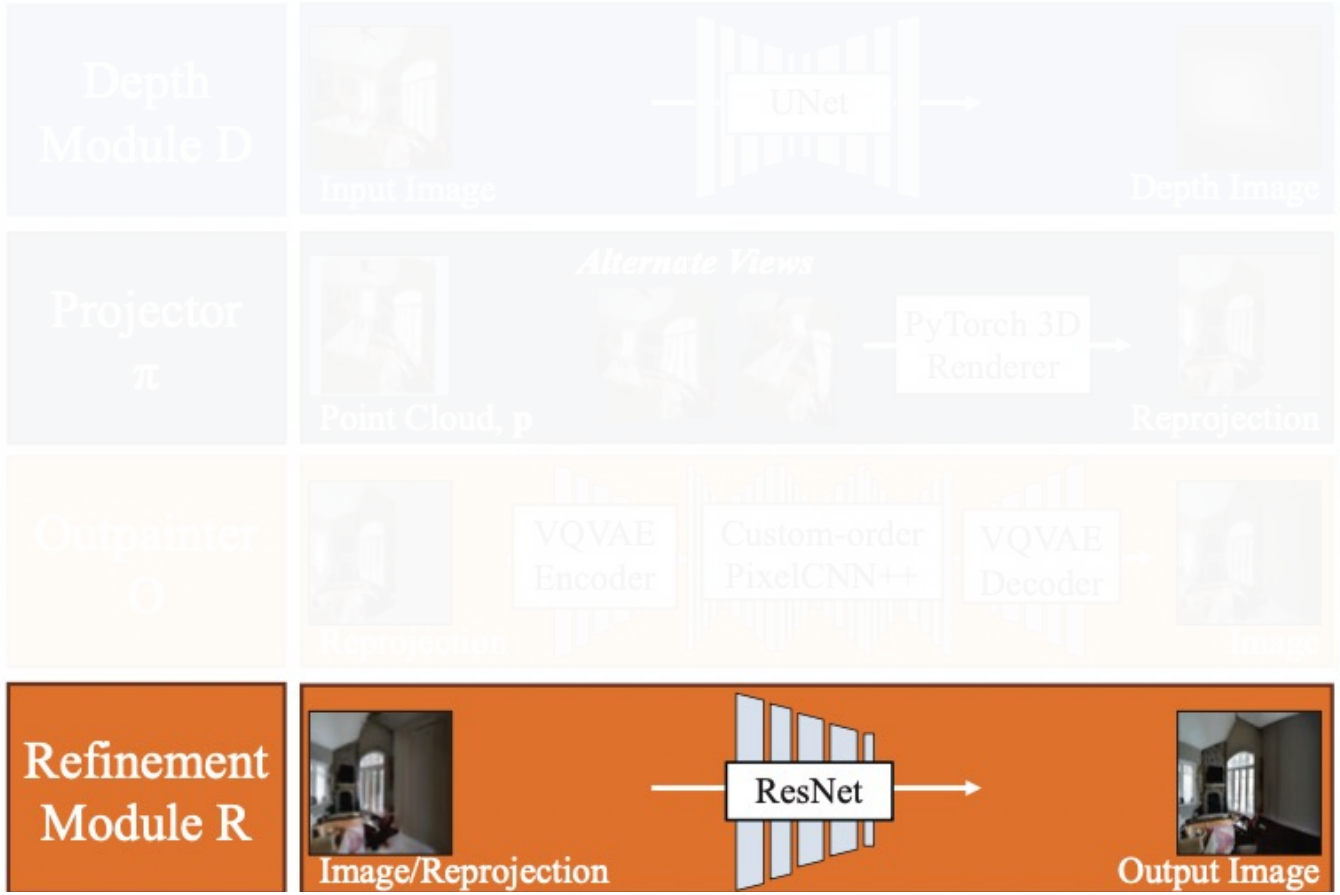
$$= ||\mathbf{x} - D(\mathbf{e})||_2^2 + ||sg[E(\mathbf{x})] - \mathbf{e}||_2^2 + \beta||sg[\mathbf{e}] - E(\mathbf{x})||_2^2$$

reconstruction codebook commitment

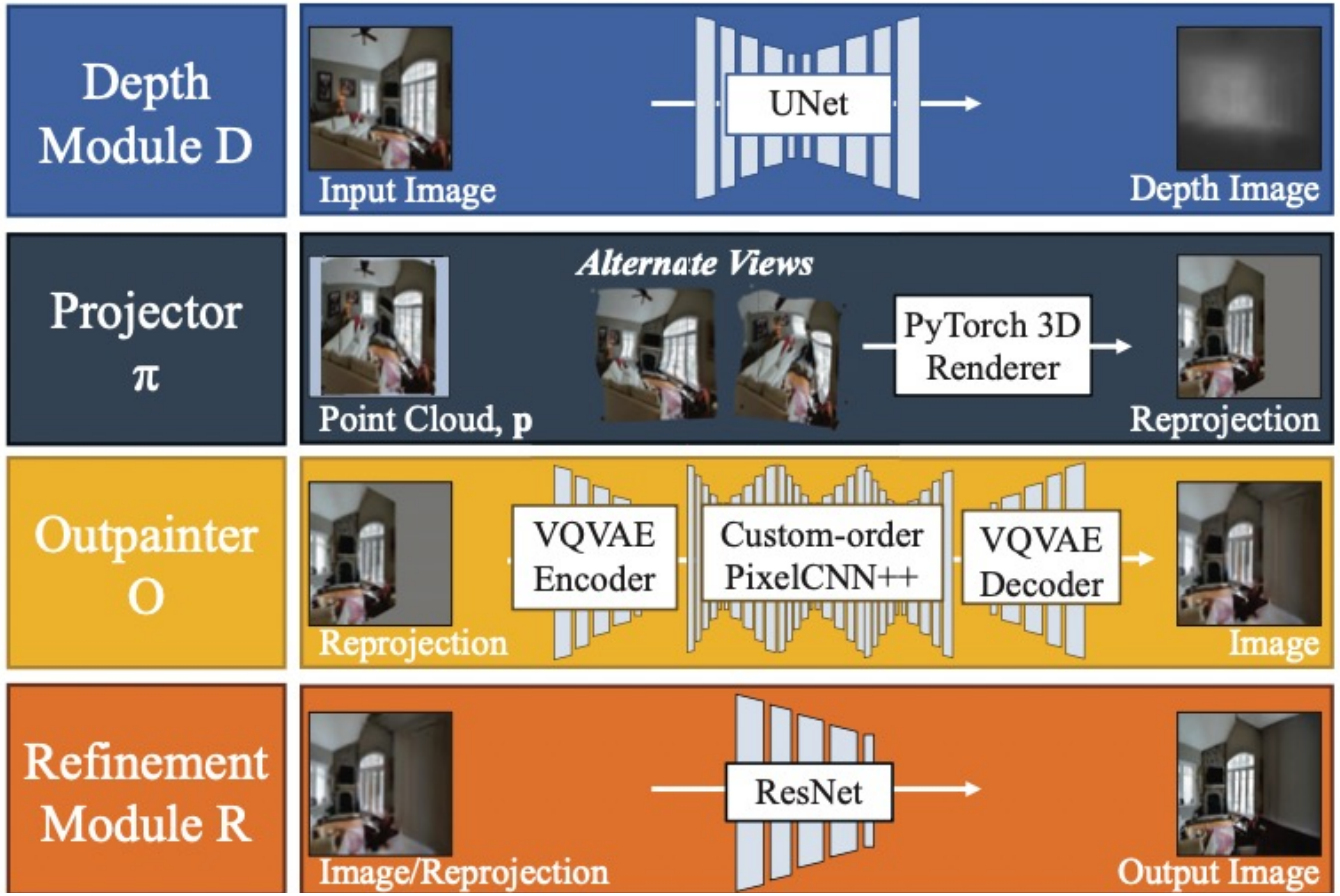
Network-Custom order pixelcnn++



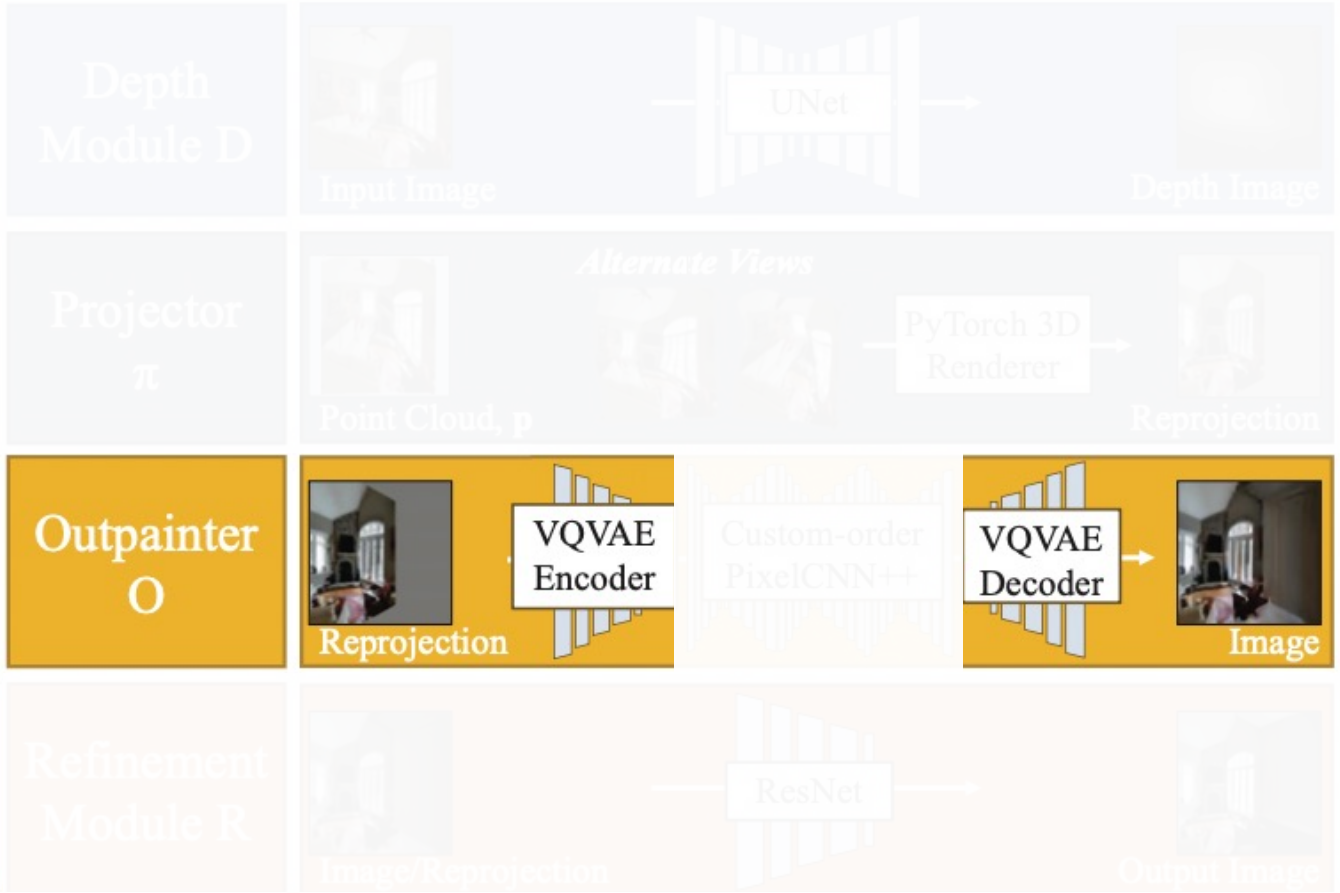
Network


Depth Module D
Input Image
UNet
Depth Image
Projector π
Alternate Views
Point Cloud, p
PyTorch 3D Renderer
Reprojection
Outpainter O
Reprojection
VQVAE Encoder
Custom-order PixelCNN++
VQVAE Decoder
Image
Refinement Module R
Image/Reprojection
ResNet
Output Image

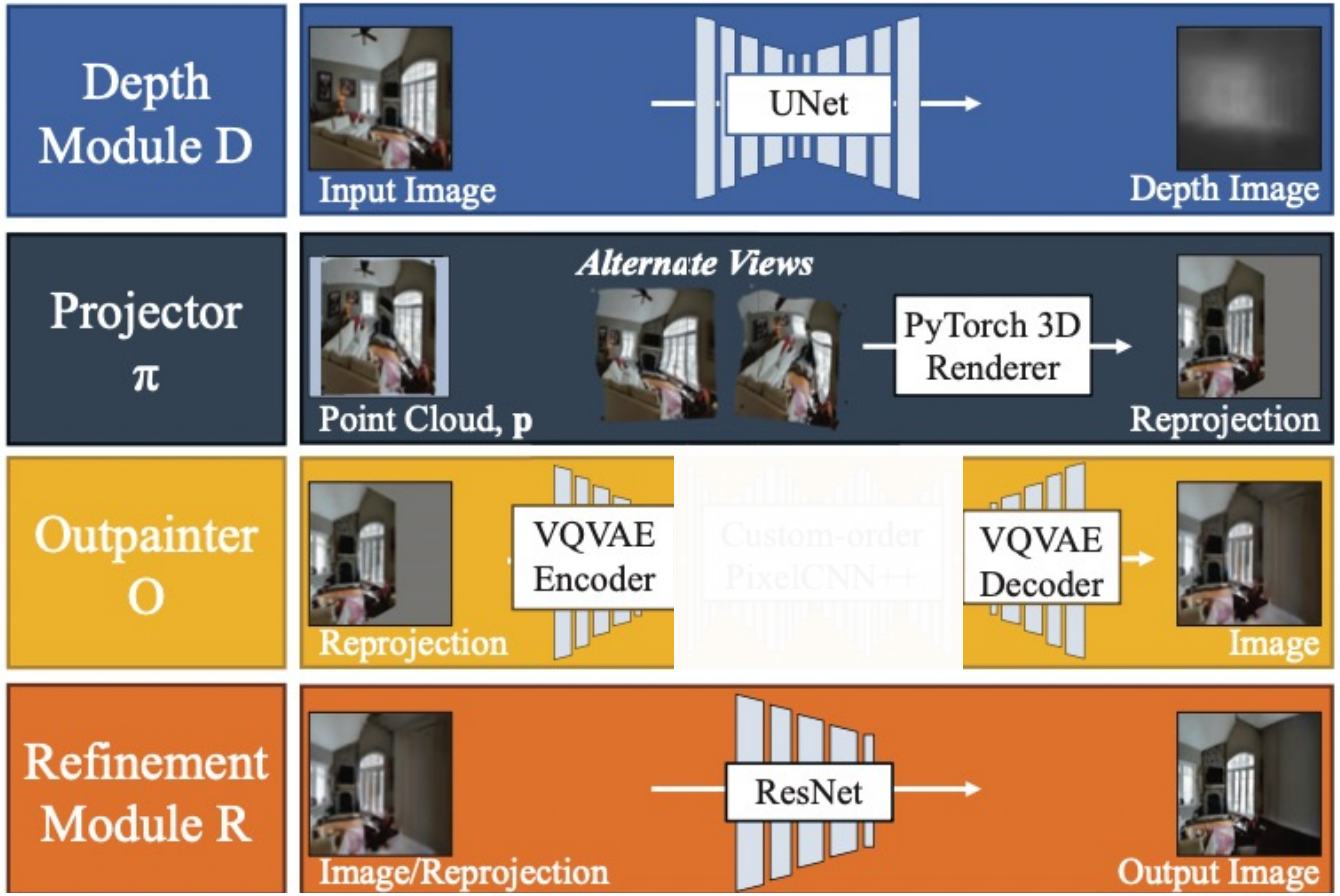
Network-Training



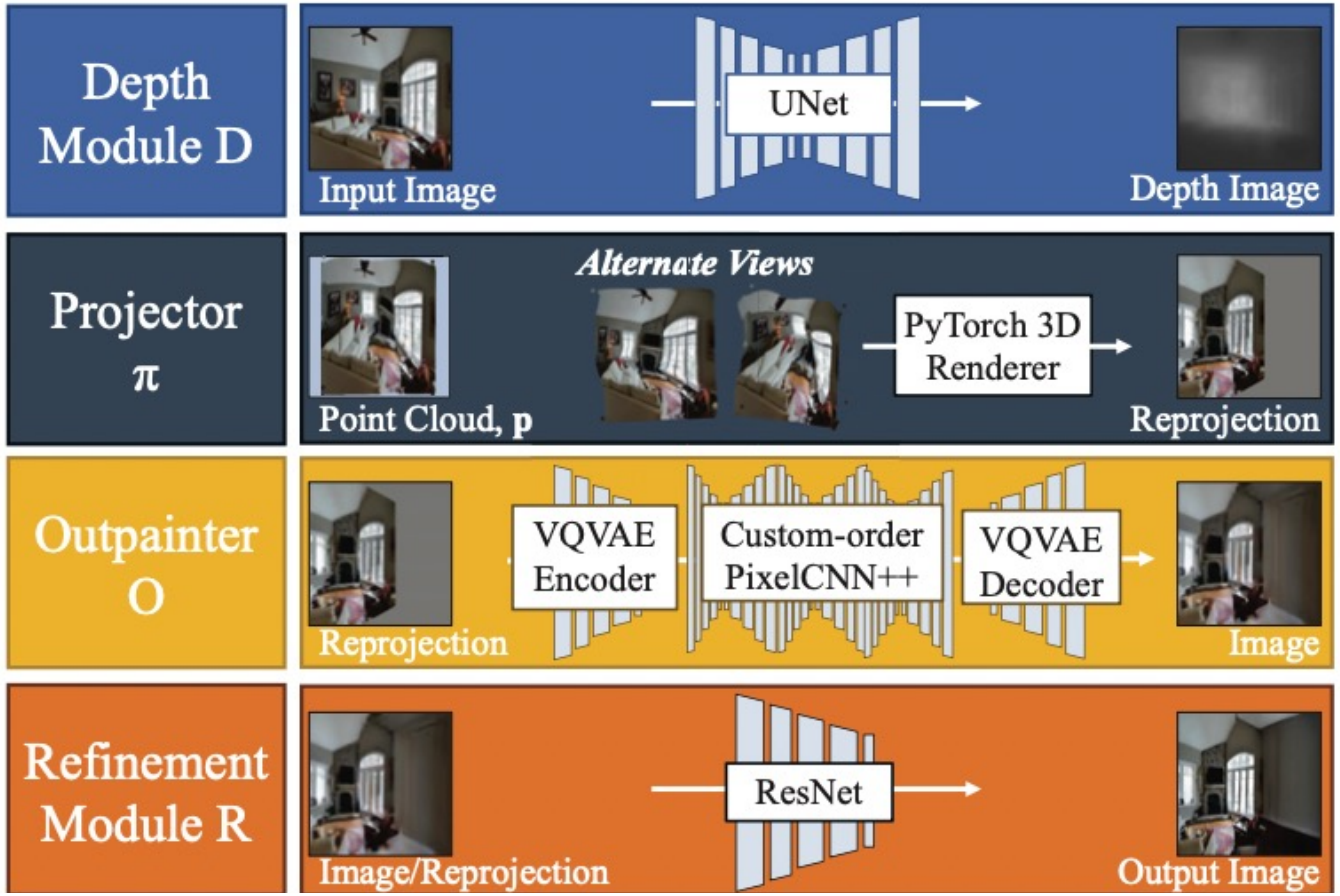
Network-Training



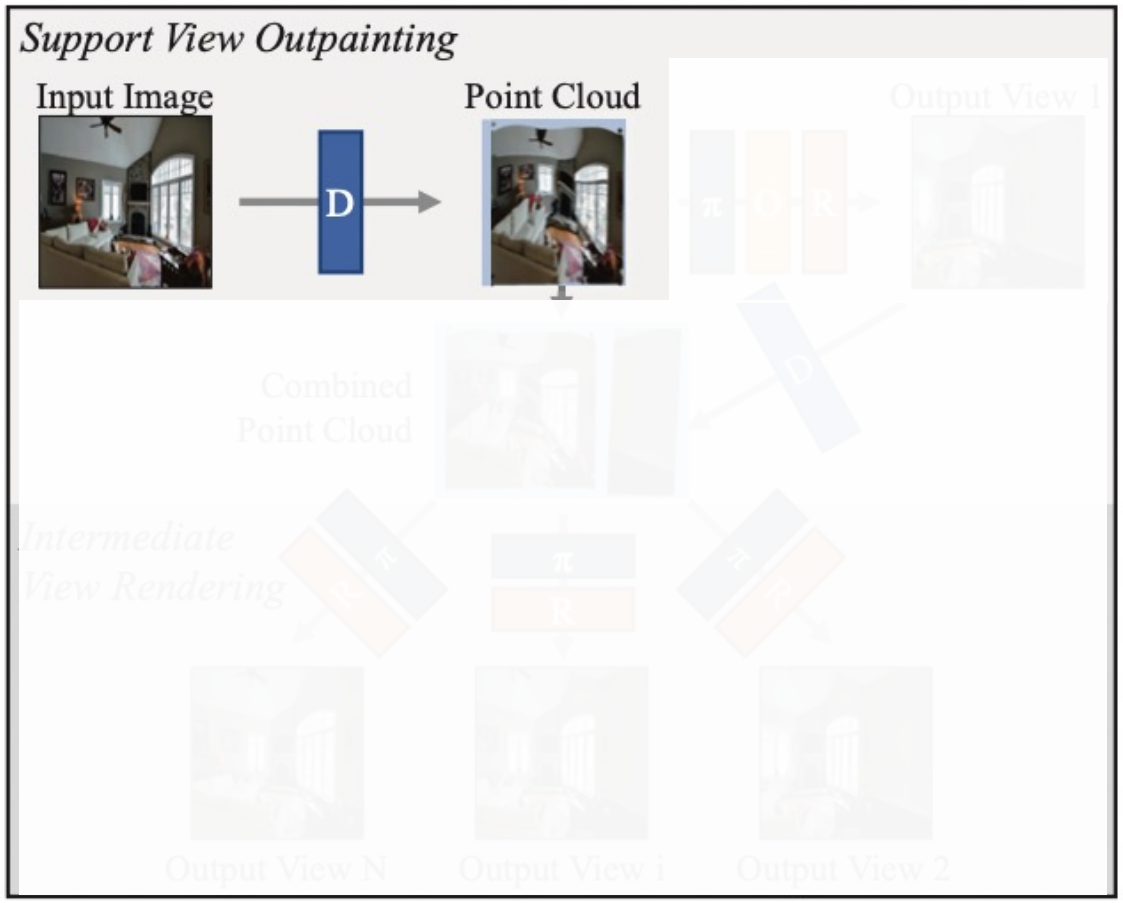
Network-Training



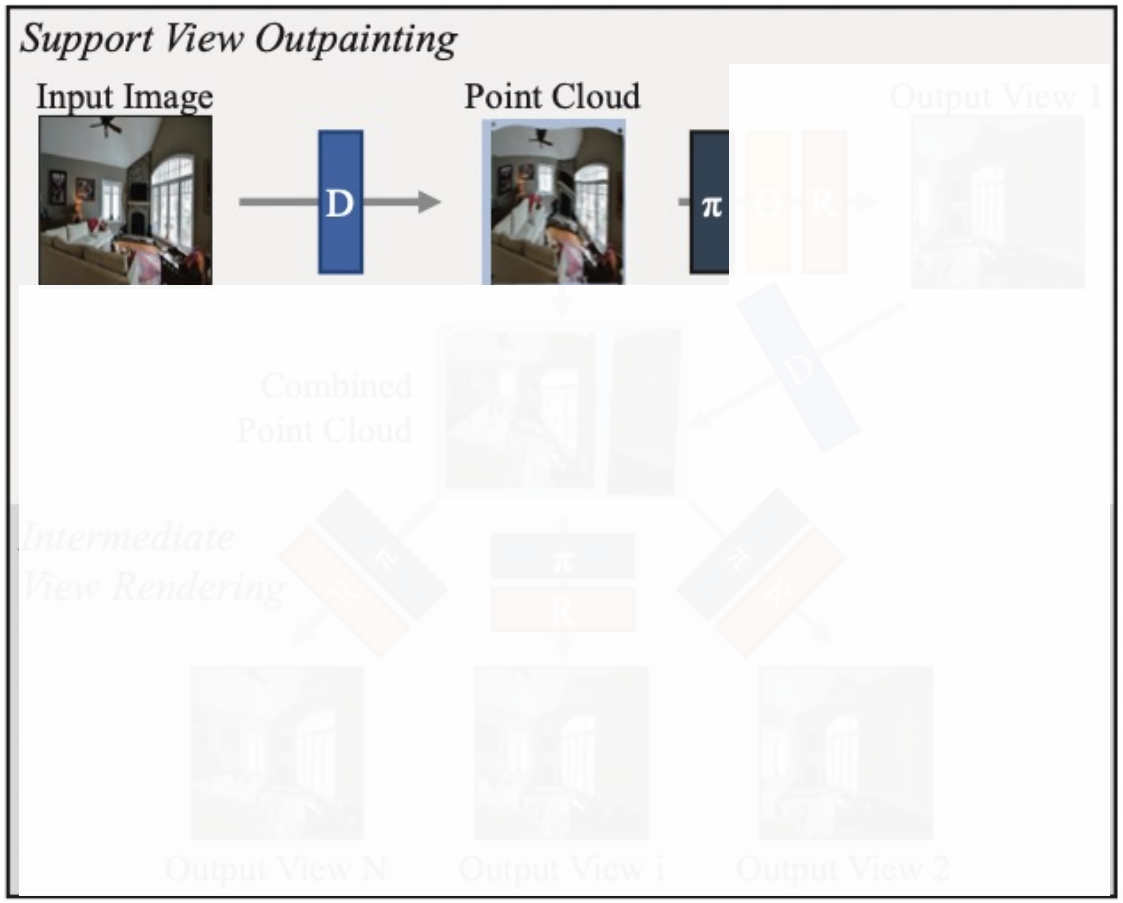
Network-Training



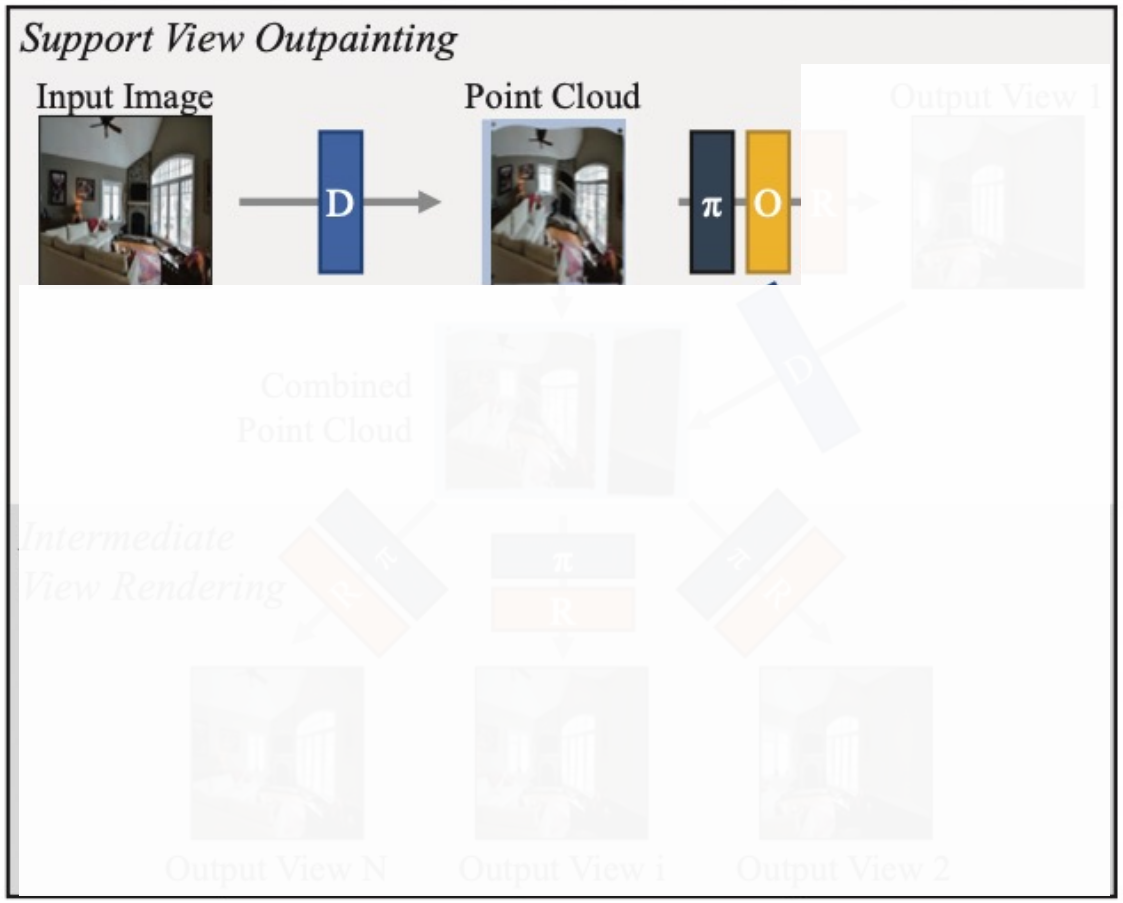
Network-Inference

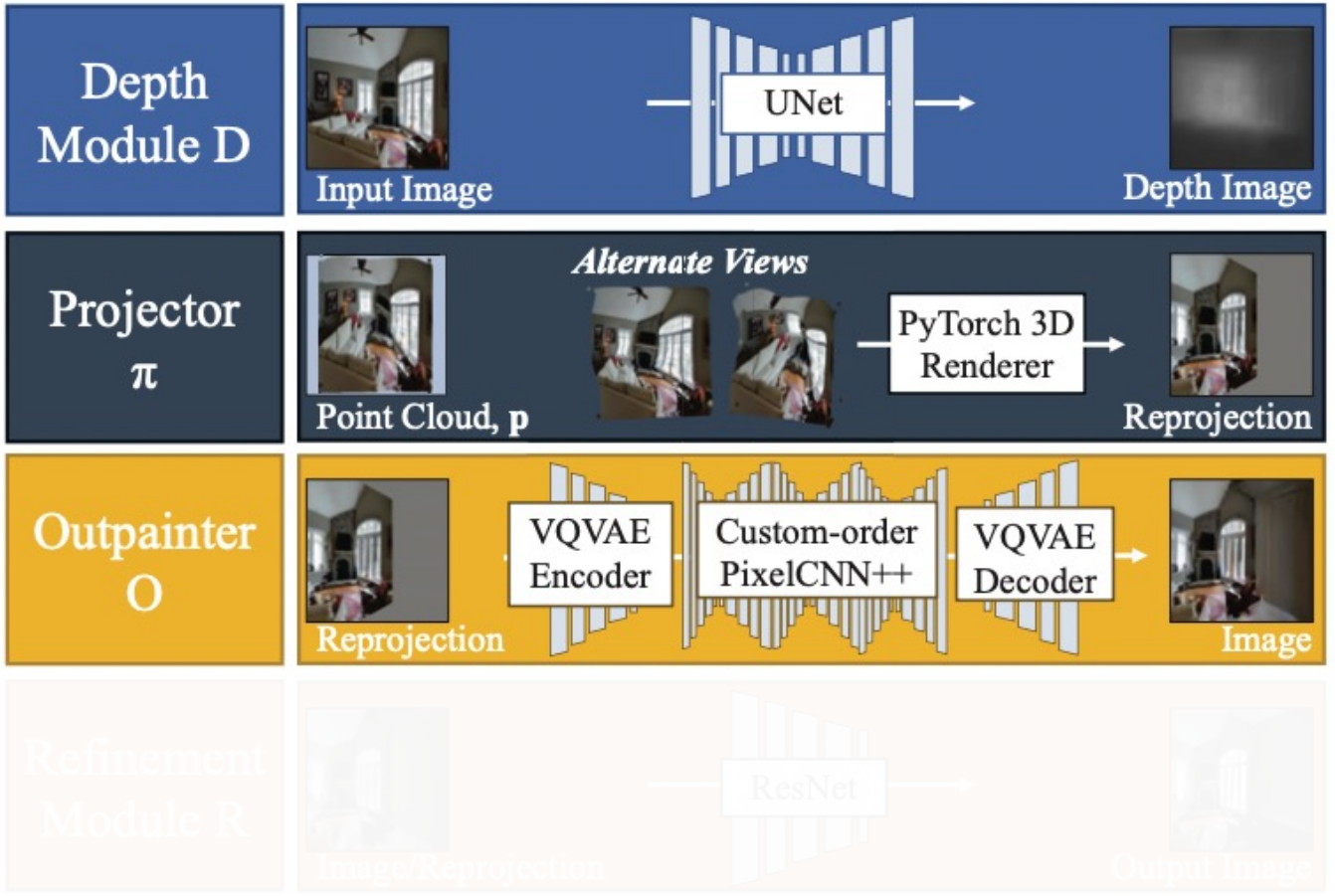


Network-Inference

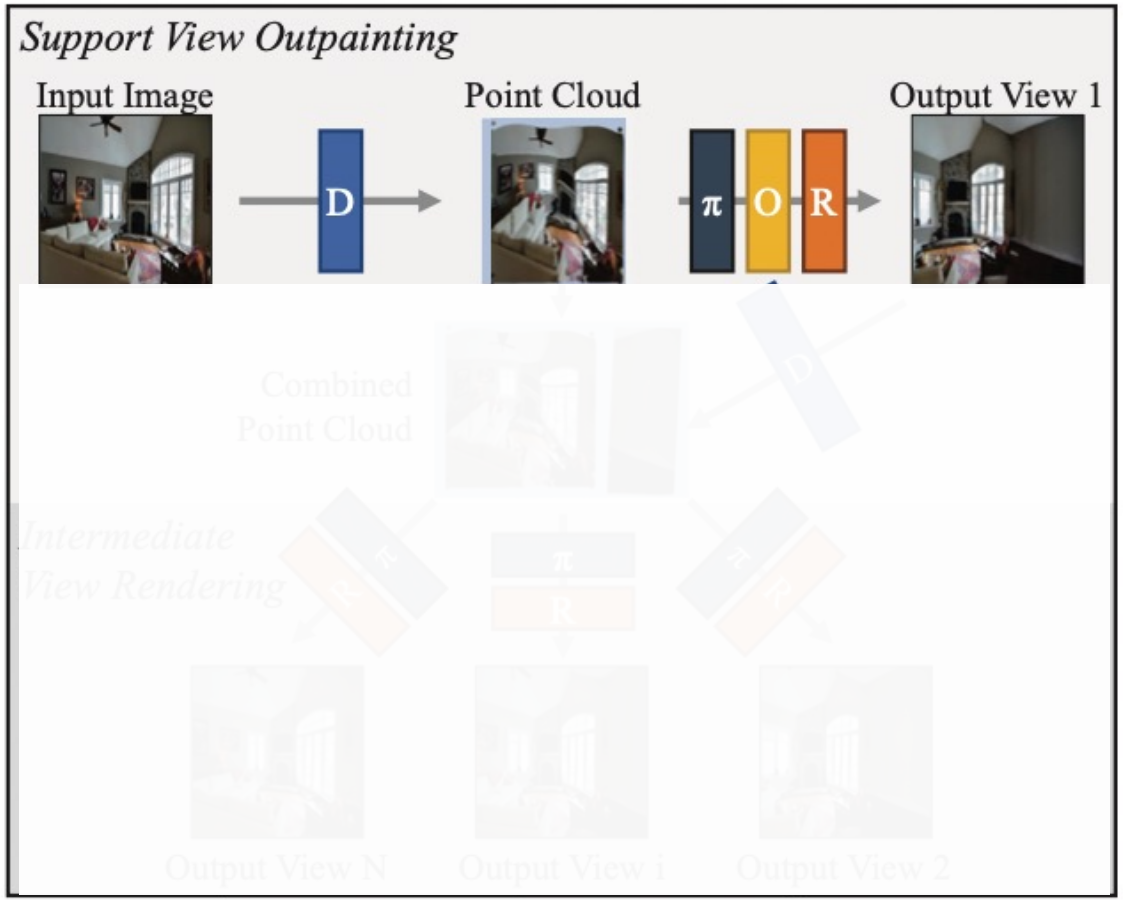


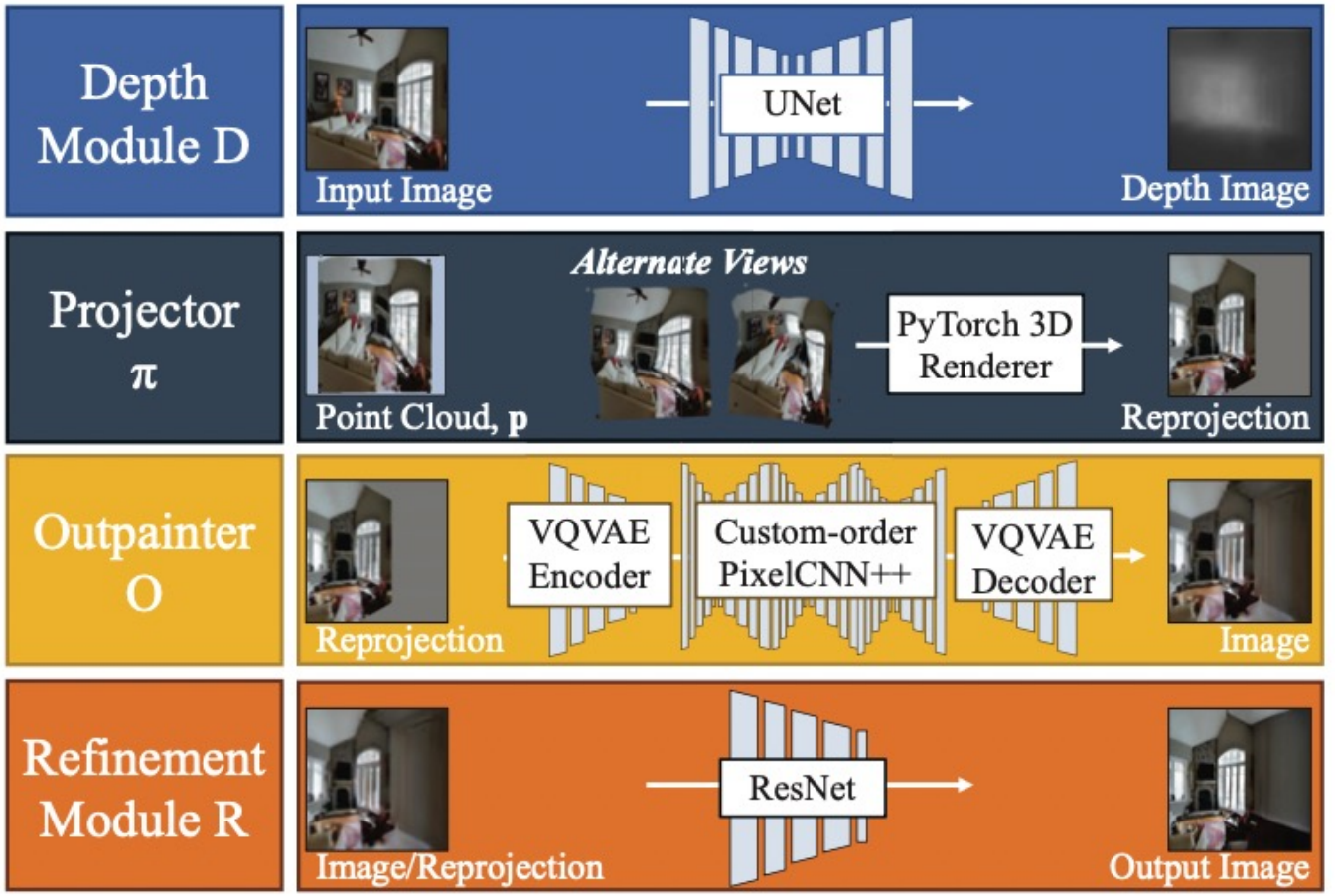
Network-Inference



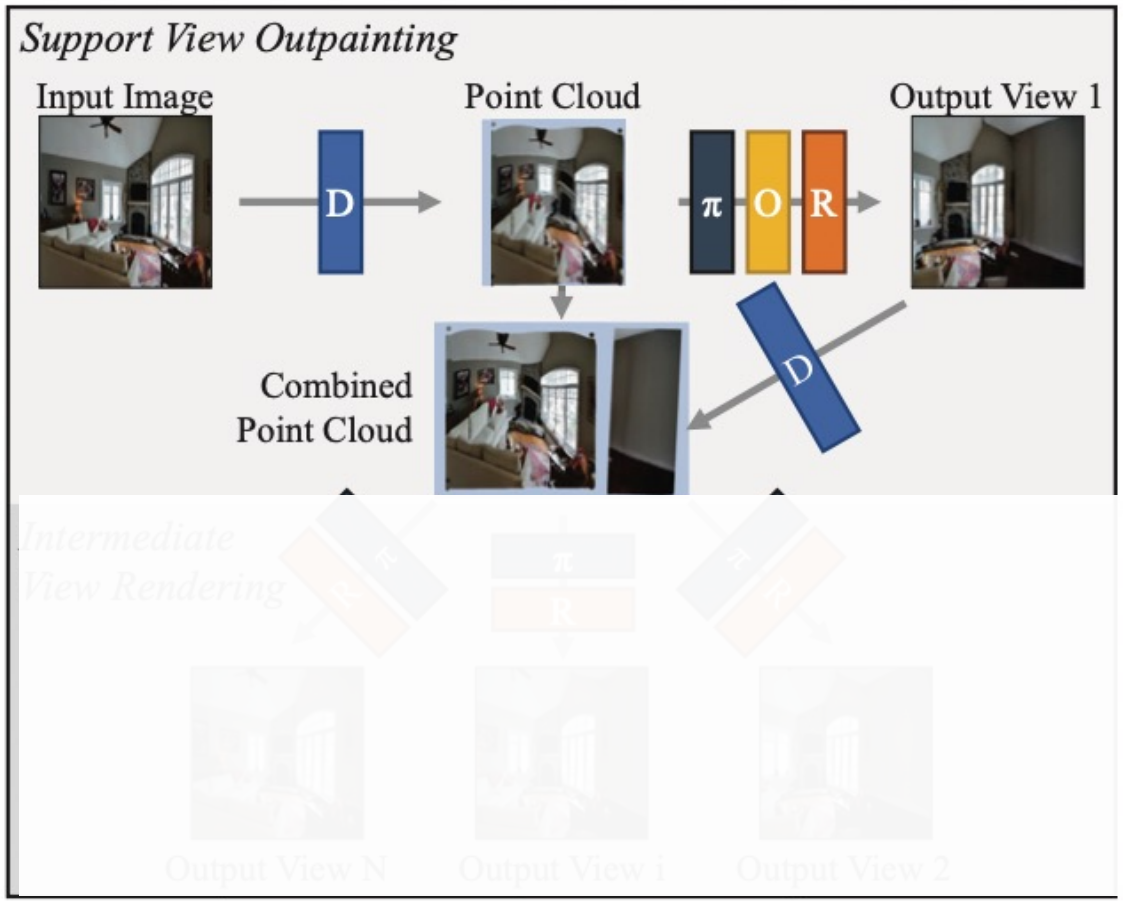


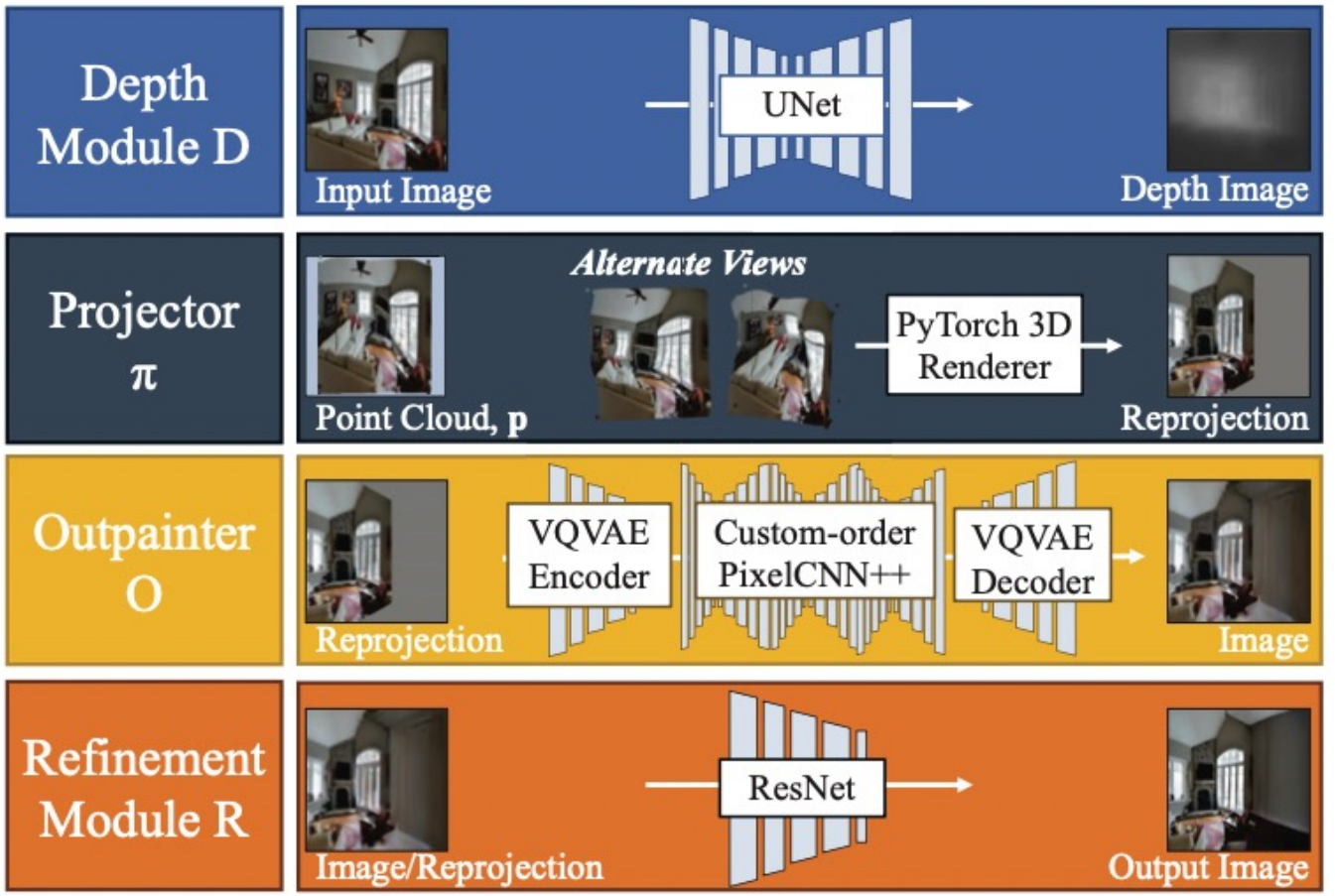
Network-Inference



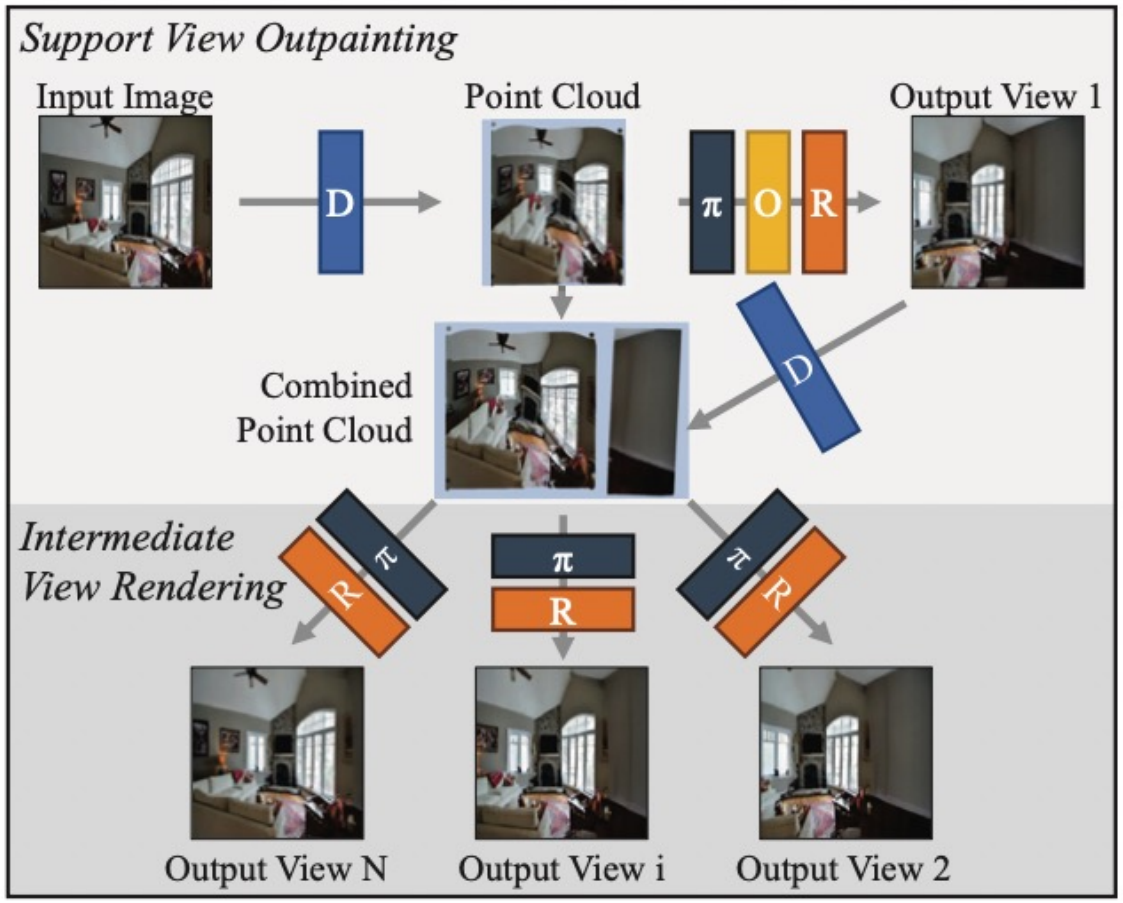


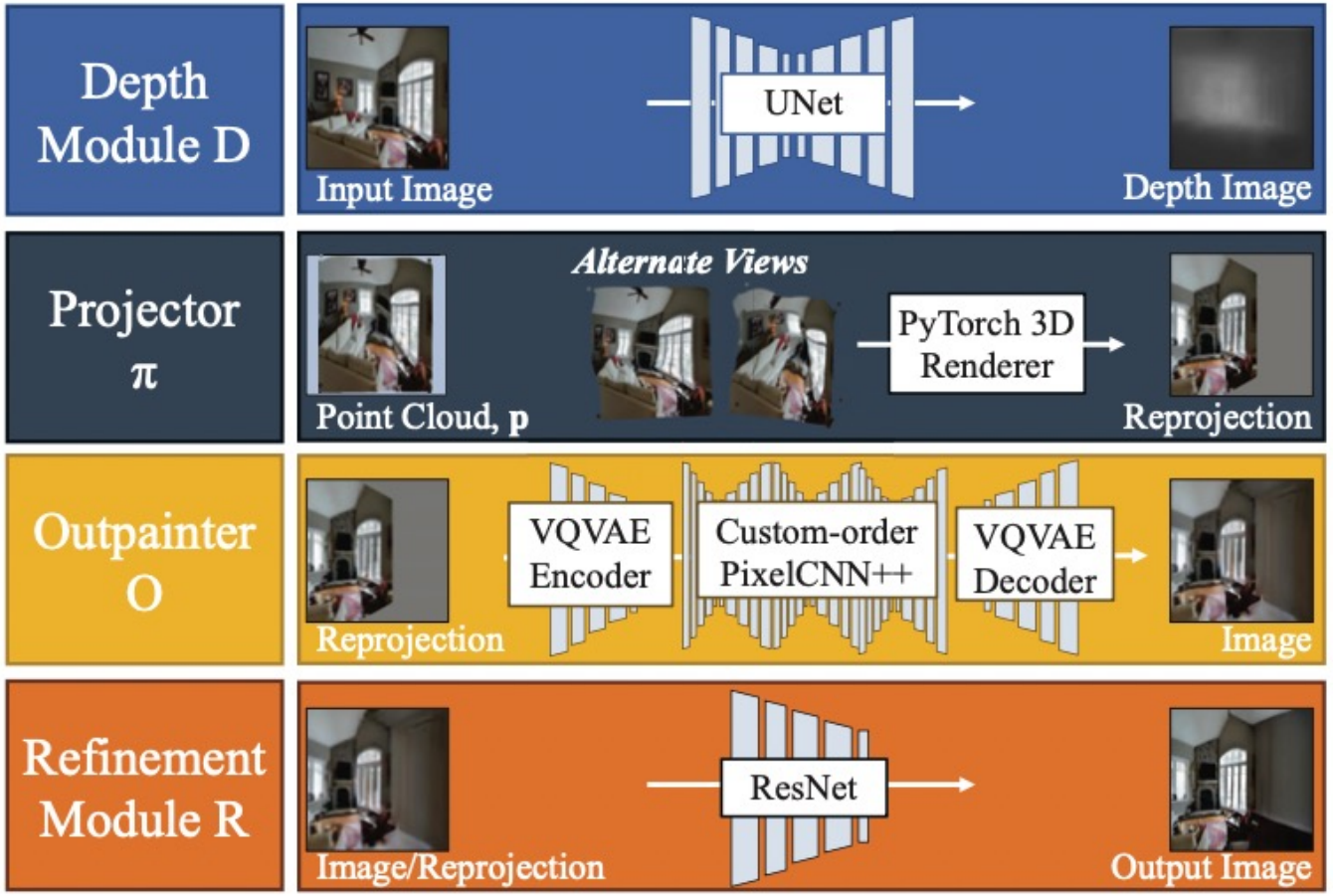
Network-Inference





Network-Inference





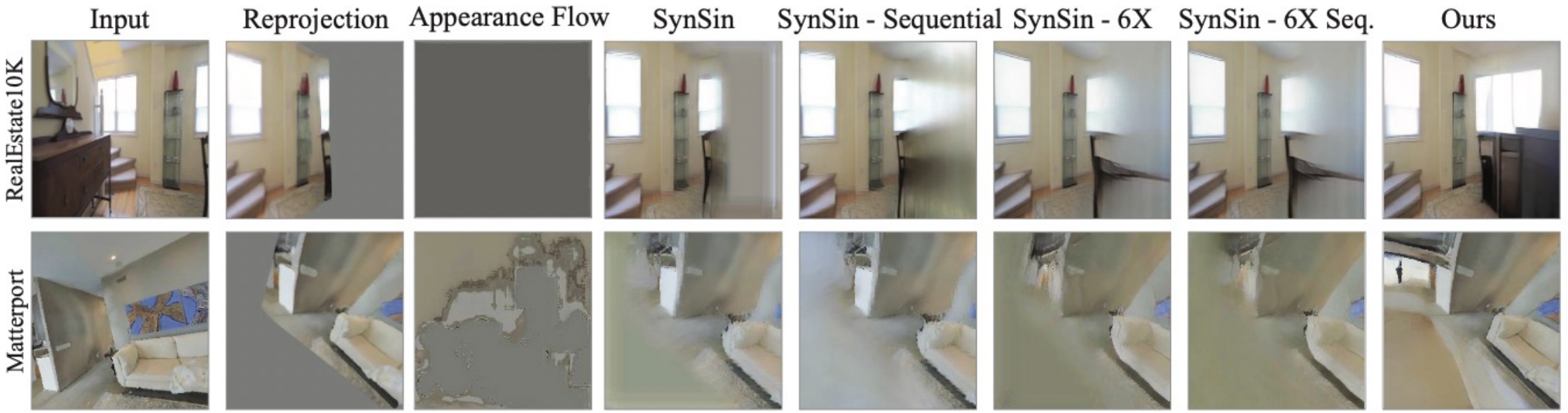
Introduction

Network

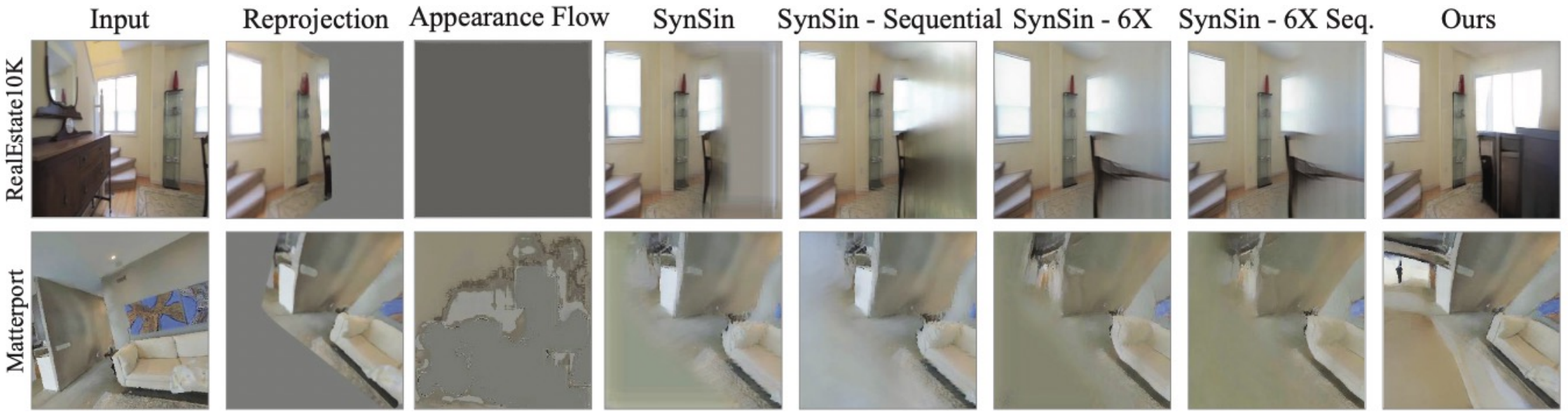
Experiments

Conclusion

Experiments

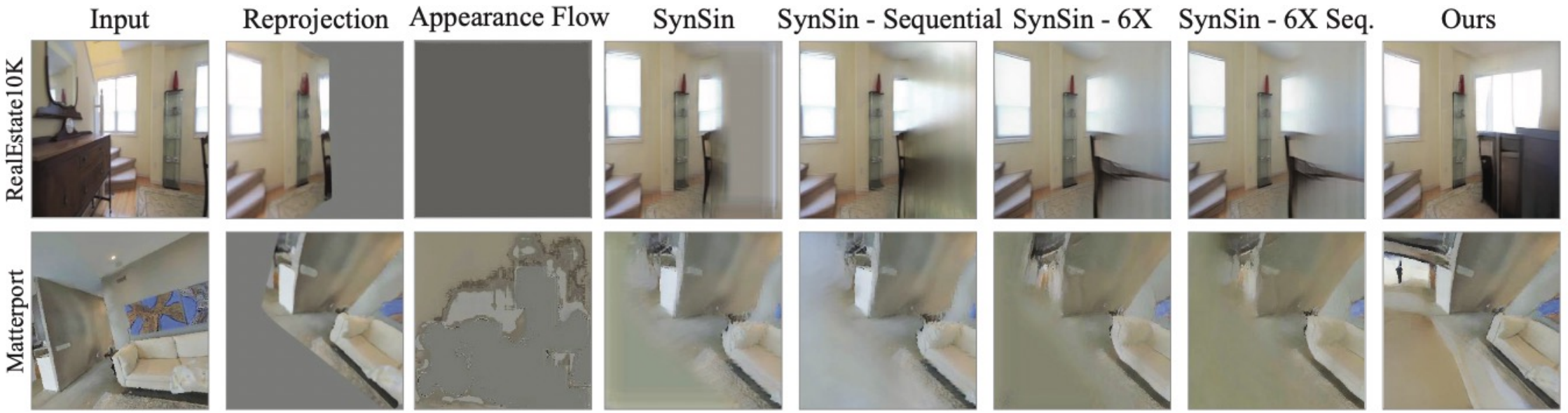


Experiments



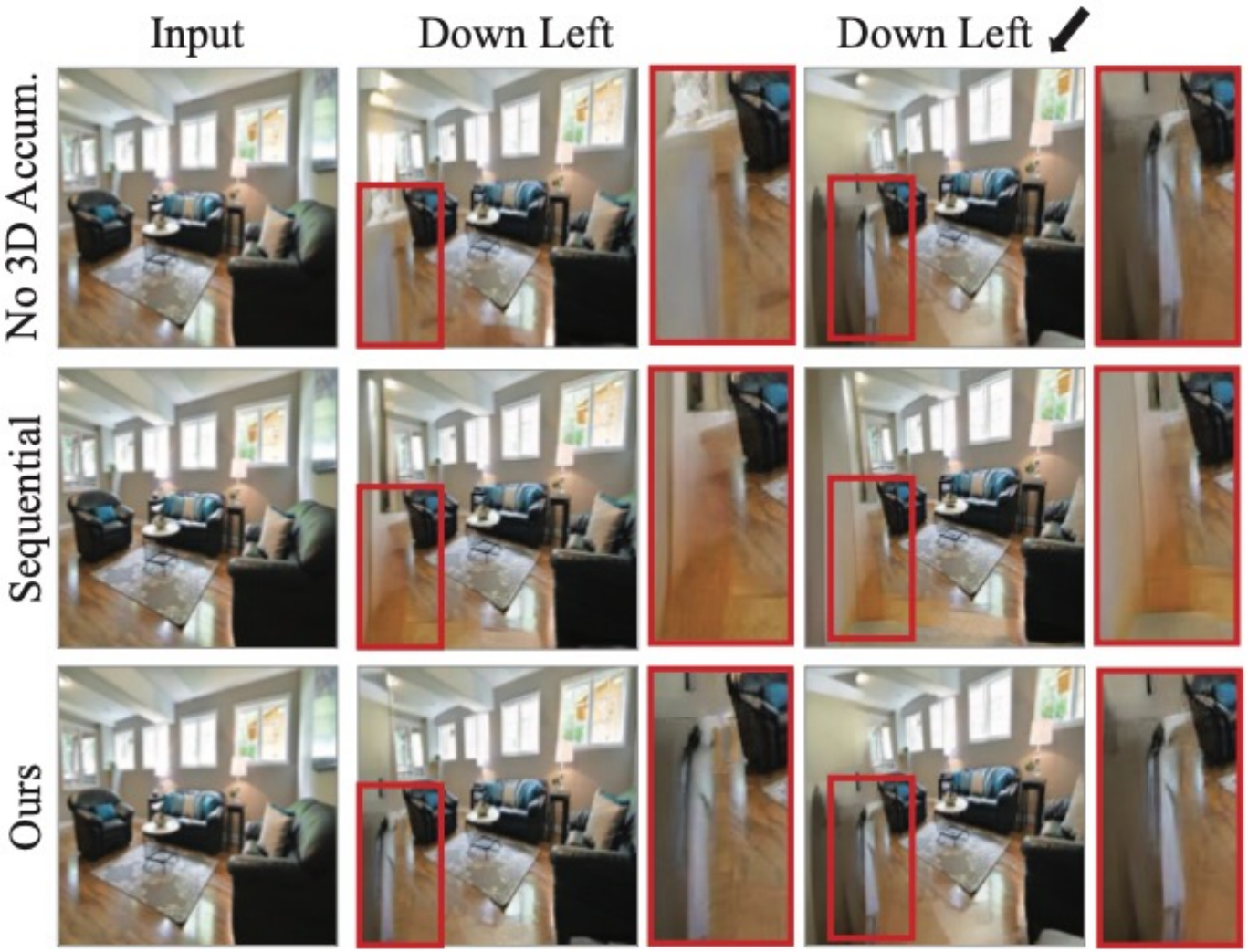
Method	Matterport		RealEstate	
	A/B ↑	FID ↓	A/B ↑	FID ↓
Tatarchenko *et al.* [45]	0.0%	427.0	0.0%	256.6
Appearance Flow [64]	19.8%	95.8	1.9%	248.3
Single-View MPI [47]	-	-	2.7%	74.8
SynSin [53]	14.8%	72.0	5.8%	34.7
SynSin - Sequential	19.5%	77.8	11.5%	34.9
SynSin - 6X	27.3%	70.4	22.0%	27.9
SynSin - 6X, Sequential	21.2%	79.3	14.4%	33.1
Ours	-	56.4	-	25.5

Experiments



Method	Matterport		RealEstate10K	
	PSNR ↑	Perc Sim ↓	PSNR ↑	Perc Sim ↓
Tatarchenko *et al.* [45]	13.72	3.82	10.63	3.98
Appearance Flow [64]	13.16	3.68	11.95	3.95
Single-View MPI [47]	-	-	12.73	3.45
SynSin - 6X, Sequential	15.61	3.17	14.21	2.73
Ours	14.60	3.17	13.10	2.88

Experiments



Method	A/B vs. Ours ↑ Matterport	RealEstate10K
No 3D Accumulation	22.6%	7.5%
Sequential Generation	44.0%	36.2%
Ours	-	-

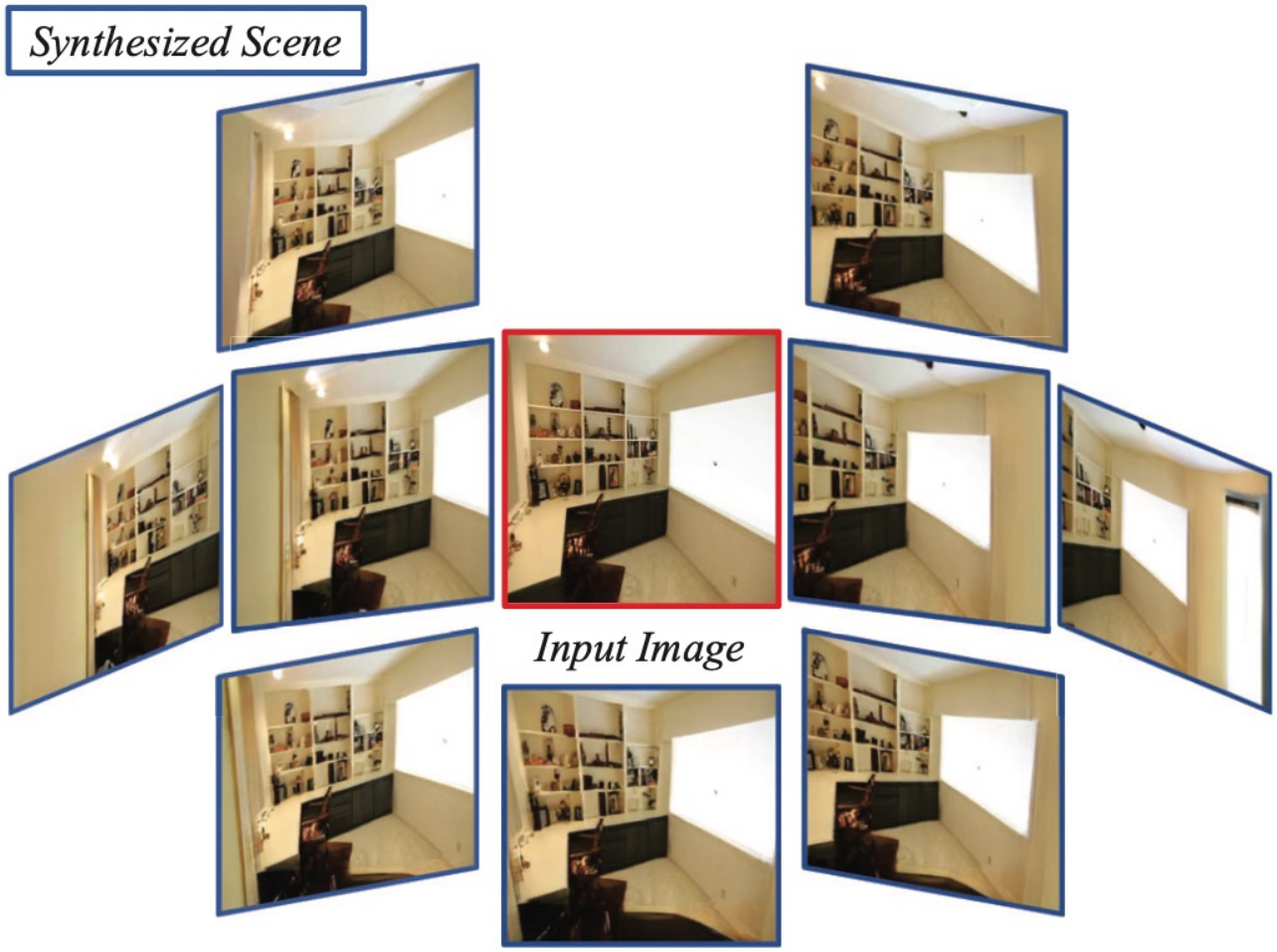
Introduction

Network

Experiments

Conclusion

Conclusion



Chris at al., PixelSynth: Generating a 3D-Consistent Experience from a Single Image ,ICCV 2021