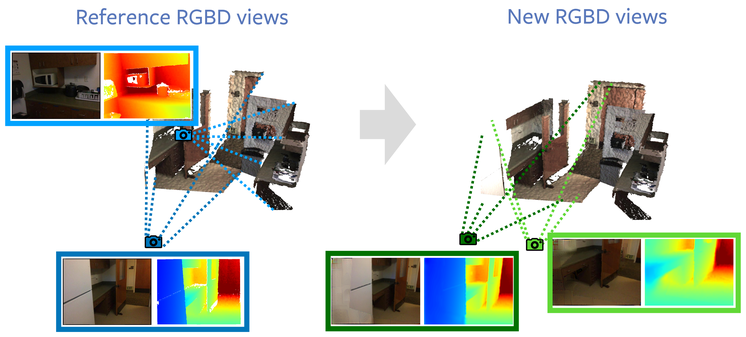
facebook Artificial Intelligence Research

David Novotny, Benjamin Graham, Jeremy Reizenstein

PerspectiveNet

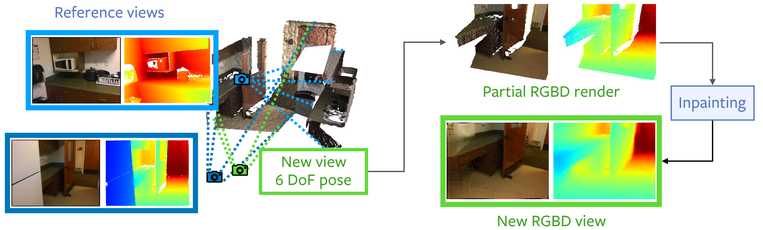
A Scene-consistent Image Generator for New View Synthesis in Real Indoor Environments

New view synthesis in real indoor environments



Existing approaches

a) Render-inpaint methods



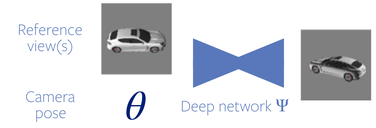
=> our method follows the render-inpaint paradigm

b) Learning-based methods

Works well for isolated objects (Ψ is relatively simple)

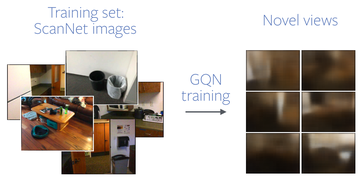
Object-centric setup = new views of cars, chairs

Tatarchenko et al.: *Multi-view 3D Models from Single Images with a Convolutional Network*
Dosovitskiy et al.: *Learning to Generate Chairs, Tables and Cars with Convolutional Networks*
Esteves et al.: *Cross-Domain 3D Equivariant Image Embeddings*
Eslami et al.: *Neural scene representation and rendering*

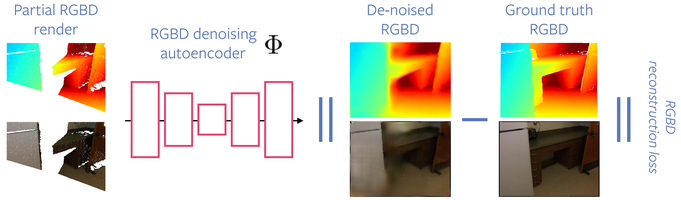


Fails for unconstrained environments (Ψ too complex)

Learning GQN on ScanNet



Naive solution: Inpainting with RGBD denoising autoencoder

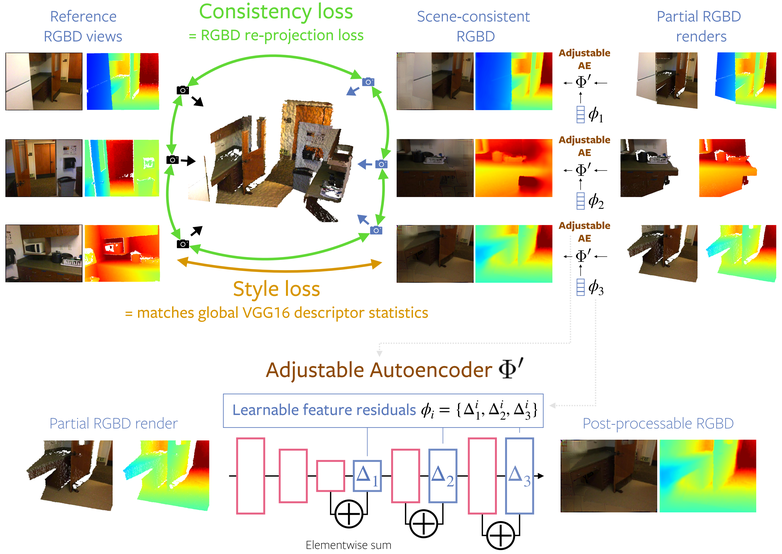


1) Multiple possible in paintings for a single partial render
=> Leads to blurry new views (average over possible solutions)
2) Independent inpainting for each partial render
=> New views of a single scene not geometrically consistent

=> PerspectiveNet
= local scene-specific optimization for scene-consistent solution

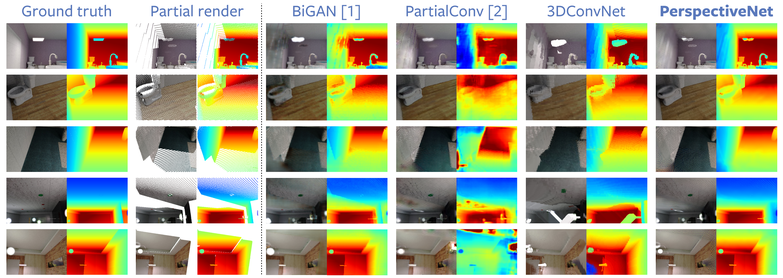
PerspectiveNet

Scene-level optimization: $\underset{\phi_1,\phi_2,\phi_3}{\arg\min}$ **Consistency loss + Style loss**

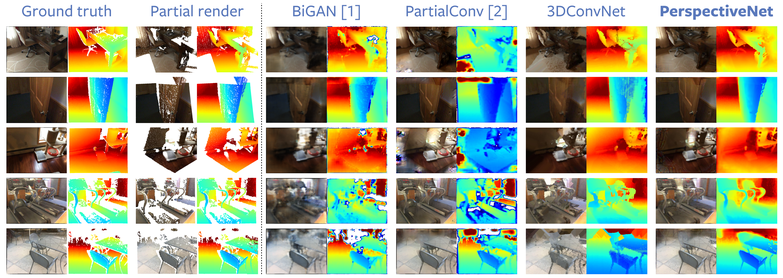


Experiments

SceneNet dataset



ScanNet dataset



Quantitative evaluation

ScanNet

Method	$\ell_1^{RGB}\downarrow$	PSNR ↑	LPIPS ↓	$\ell_1^D[m]\downarrow$	$\delta_1\uparrow$	$\delta_2\uparrow$	$\delta_3\uparrow$
	Color metrics			Depth metrics			
PerspectiveNet	68.022	13.762	**0.422**	**0.115**	**0.352**	**0.411**	**0.471**
PerspectiveNet w/o opt	**66.511**	**13.986**	0.426	0.120	0.188	0.230	0.279
PartialConv	93.604	11.374	0.461	0.750	0.194	0.236	0.283
3DConvNet	78.590	12.190	0.531	0.138	0.301	0.359	0.426
BiGAN	77.313	12.742	0.523	0.215	0.169	0.212	0.265

SceneNet

Method	$\ell_1^{RGB}\downarrow$	PSNR ↑	LPIPS ↓	$\ell_1^D[m]\downarrow$	$\delta_1\uparrow$	$\delta_2\uparrow$	$\delta_3\uparrow$
	Color metrics			Depth metrics			
PerspectiveNet	49.698	15.687	**0.424**	**0.219**	**0.366**	**0.431**	**0.494**
PerspectiveNet w/o opt	**48.521**	**16.324**	0.442	0.227	0.101	0.125	0.155
PartialConv	76.470	12.377	0.481	1.846	0.008	0.010	0.013
3DConvNet	75.942	12.614	0.570	0.653	0.040	0.050	0.062
BiGAN	55.815	15.112	0.485	0.249	0.319	0.375	0.431

Conclusions

- One of first systematic evaluations of new view synthesis in real indoor environments
- Proposed PerspectiveNet, a local optimization scheme for obtaining scene-consistent solutions
- Qualitative results show better performance than baselines on 2 datasets

References

[1] Zhu et al.: *Toward multimodal image-to-image translation*, NIPS 2017
[2] Liu et al.: *Image Inpainting for Irregular Holes Using Partial Convolutions*, ECCV 2018