Project work ML4CV

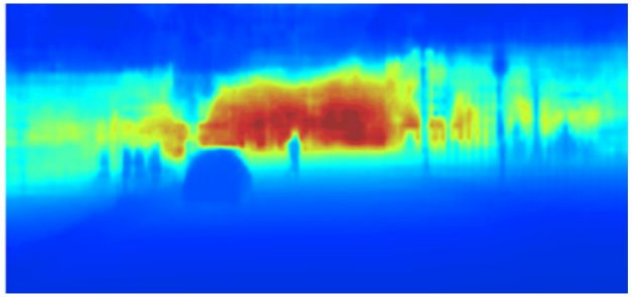
SIDE + NVS

Giovanni Minelli
giovanni.minelli2@studio.unibo.it

SIDE task

1 Image

Depth estimation



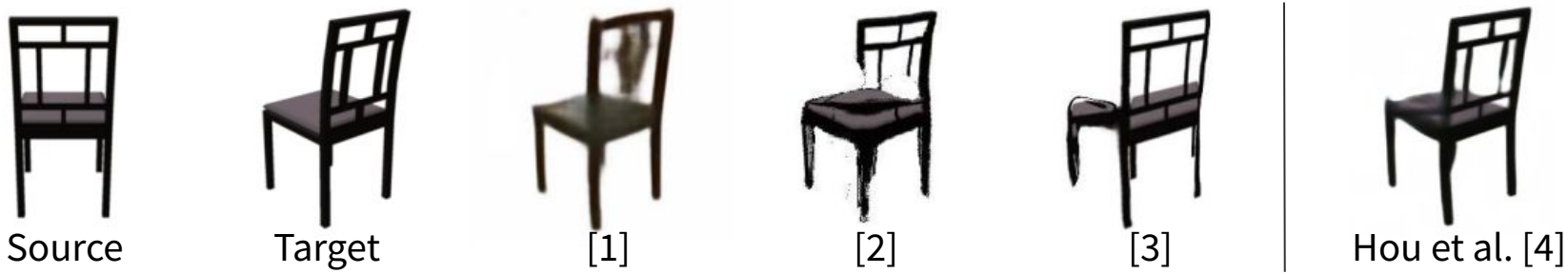
”Single image depth estimation: An overview”

1 View

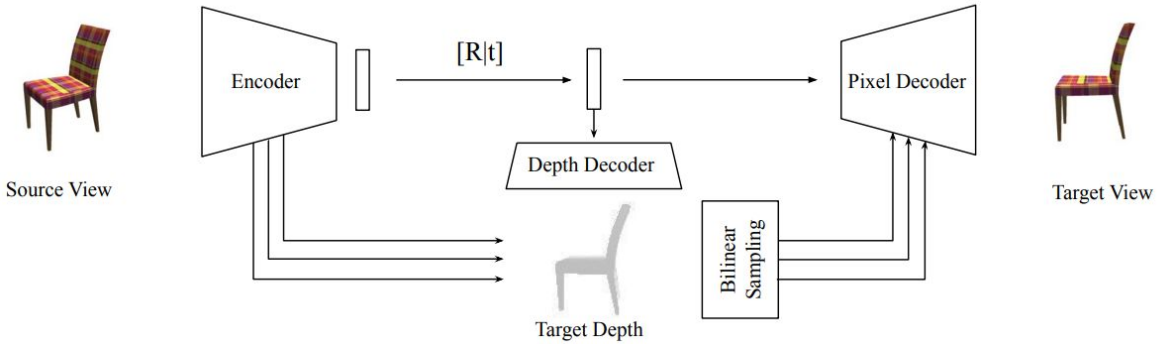
NVS task

Arbitrary view

“Multi-view to Novel view: Synthesizing novel views with Self-Learned Confidence”



- Problems with wide angles
- Blurred predictions, holes in images and missing details



- Very long time of convergence
- Use just as side task: not reliable, inconsistent depth

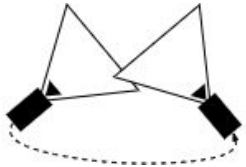
[1] Multi-view 3d models from single images with a convolutional network;
[2] Multi-view to novel view: Synthesizing novel views with self-learned confidence;
[3] Monocular neural image based rendering with continuous view control;
[4] Hou; Novel View Synthesis via Depth Guided Skip Connections; WACV 2021

Pipeline

Inputs

S

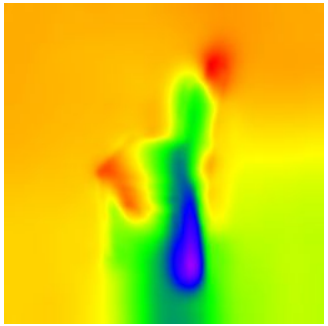
[R|t]



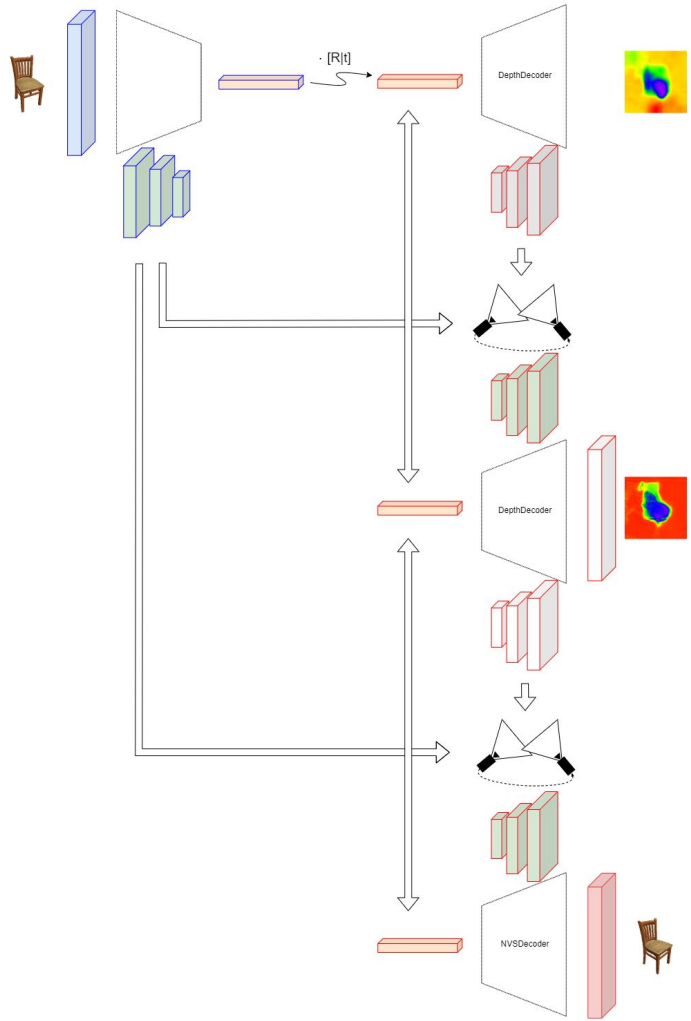
T

256x256

Predictions

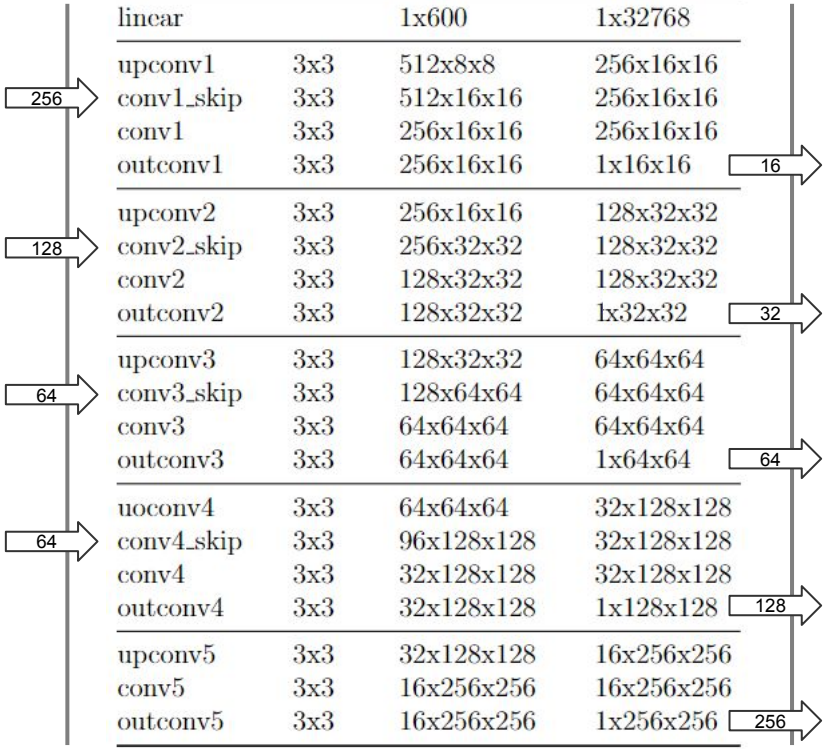


https://github.com/Xharlie/ShapenetRender_more_variation



Improvements

Layer (type)	Kernel	Output Shape	Input Shape
linear		1x600	1x32768
upconv1	3x3	512x8x8	256x16x16
conv1_skip	3x3	512x16x16	256x16x16
conv1	3x3	256x16x16	256x16x16
outconv1	3x3	256x16x16	1x16x16
upconv2	3x3	256x16x16	128x32x32
conv2_skip	3x3	256x32x32	128x32x32
conv2	3x3	128x32x32	128x32x32
outconv2	3x3	128x32x32	1x32x32
upconv3	3x3	128x32x32	64x64x64
conv3_skip	3x3	128x64x64	64x64x64
conv3	3x3	64x64x64	64x64x64
outconv3	3x3	64x64x64	1x64x64
uoconv4	3x3	64x64x64	32x128x128
conv4_skip	3x3	96x128x128	32x128x128
conv4	3x3	32x128x128	32x128x128
outconv4	3x3	32x128x128	1x128x128
upconv5	3x3	32x128x128	16x256x256
conv5	3x3	16x256x256	16x256x256
outconv5	3x3	16x256x256	1x256x256



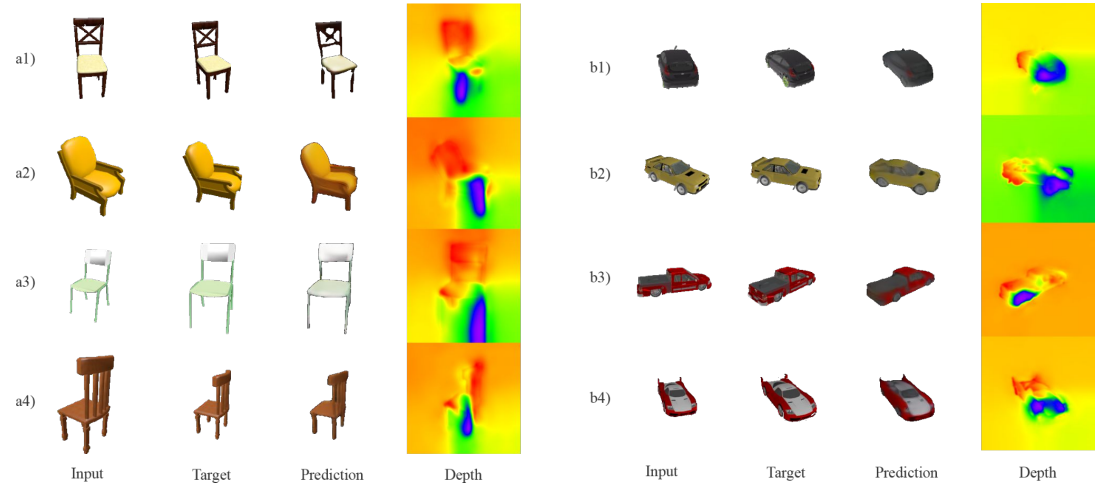
- Better encoder: ResNet-18 pretrained
- Revised decoders: UNet like structure
- Focused supervision loss

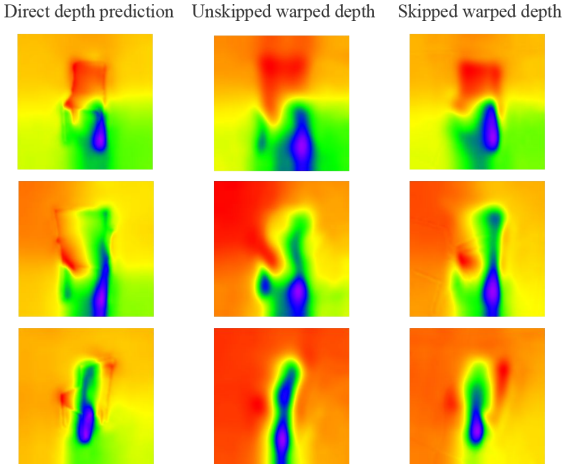
$$L_{tot} = \alpha L_{recon} + \beta L_{VGG} + \gamma L_{smooth} + \delta L_{skip}$$

- Faster convergence with modified training
- Hyper-parameters tuning

Results

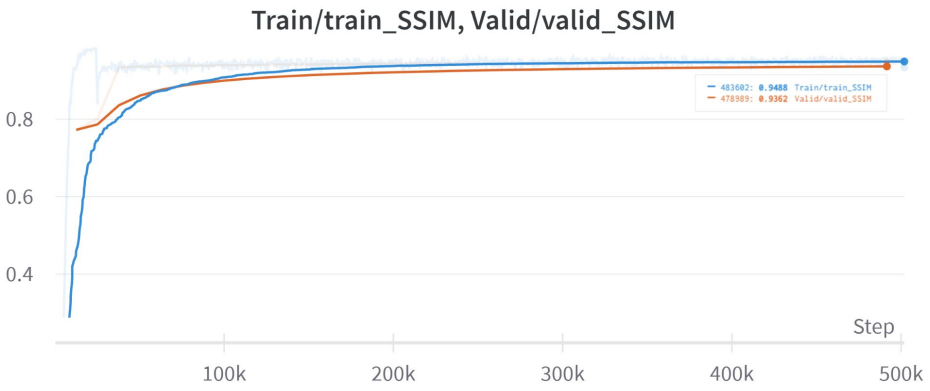
Methods	SSIM(↑)	L1(↓)	SILog(↓)	Abs Rel(↓)	Sq Rel(↓)	RMSE(↓)	Log RMSE(↓)	$\sigma \leq 1.25$(↑)	$\sigma \leq 1.25^2$(↑)	$\sigma \leq 1.25^3$(↑)
CHAIR										
Hou [2]	**0.906**	0.136	21.075	0.110	0.035	0.197	0.218	0.818	0.940	0.973
Ours	0.888	**0.075**	**10,100**	**0,067**	**0,011**	**0,114**	**0,101**	**0,956**	**0,992**	**0,999**
CAR										
Hou [2]	**0.930**	0.109	36.074	0.268	0.100	0.327	0.399	0.496	0.704	0.942
Ours	0,905	**0,059**	**9,629**	**0.037**	**0.010**	**0,110**	**0,097**	**0,952**	**0,988**	**0,997**

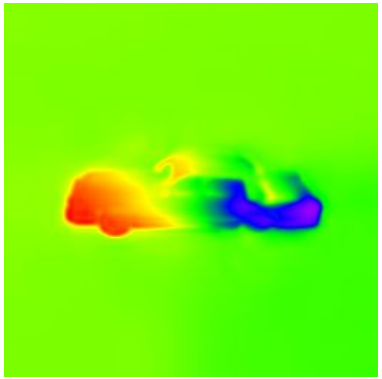




Weak points

- BatchNorm (not needed?)
- No regularization (low model capacity?)
- Small range of rotation
- NVS results blurry, decolorized, missing texture details
- Depth on parallel planes
- Not tested on real world data





Thanks