

Deep implicit network for single-view textured 3D reconstruction

Student : Kuan-Hsun Wu

Advisor : Shun-Cheng Wu

Supervisor : Federico Tombari

Outline

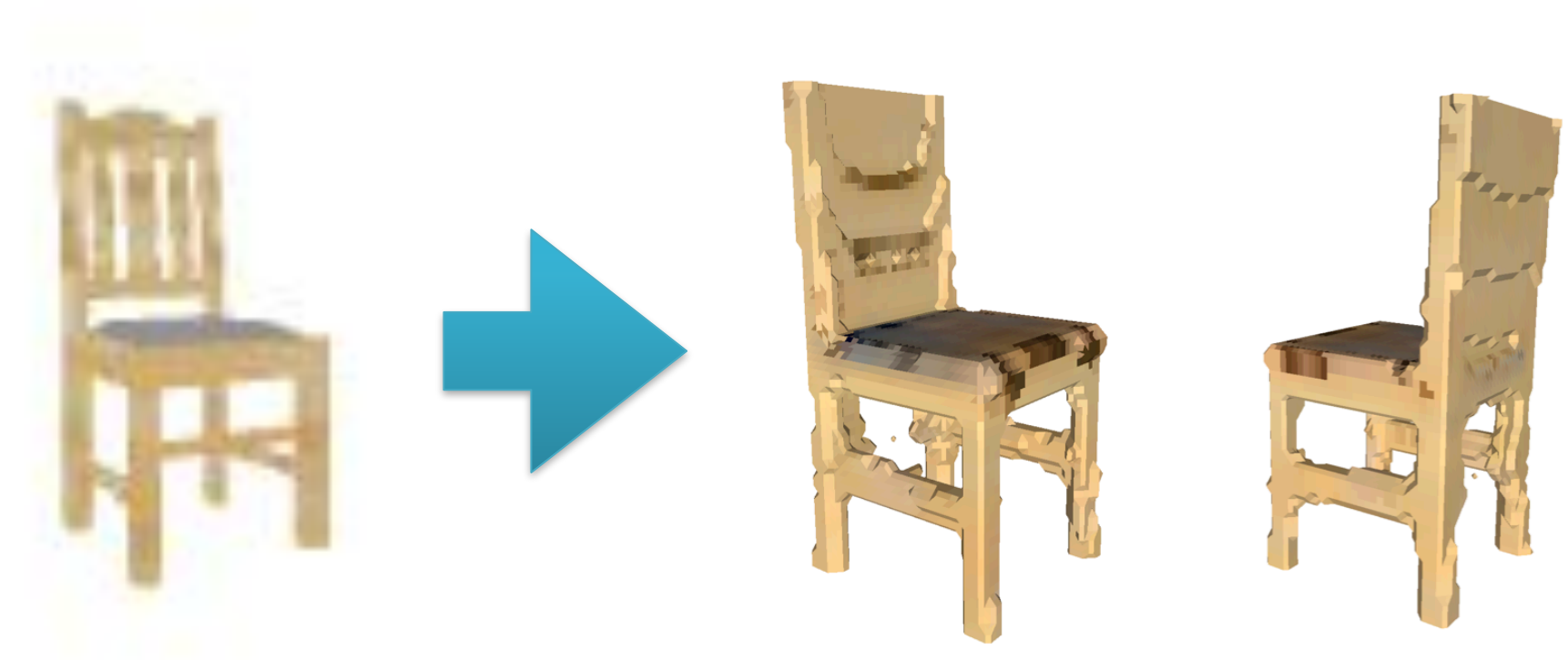
- Introduction
- Related Work
- Approach
- Experiments
- Conclusion



Introduction

Single-view textured 3D object reconstruction

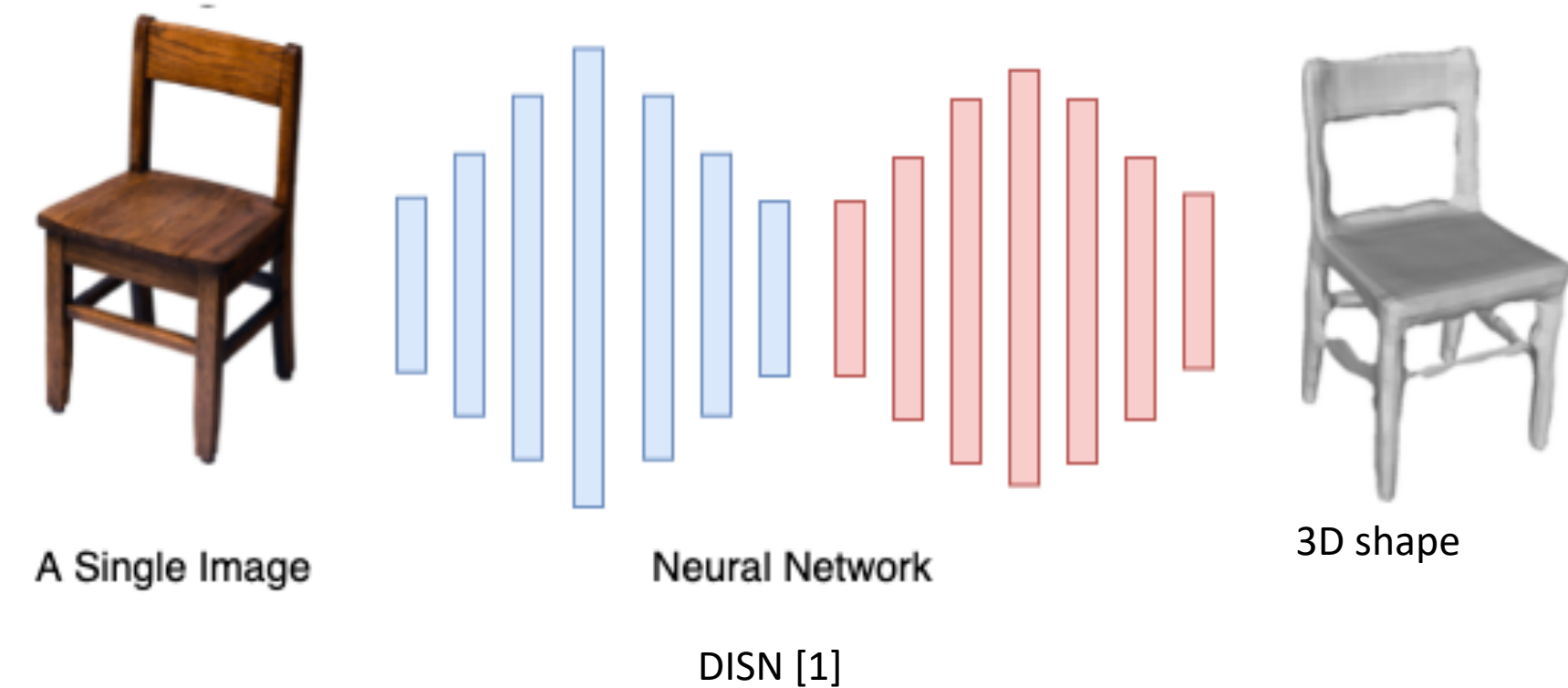
- Input : A single image, camera matrix
- Output : Textured 3D object



Related Work

Most of 3D reconstruction methods only focus on 3D geometry

- DISN [1]
- Occupancy networks [2]



[1]: Xu, Q., Wang, W., Ceylan, D., Mech, R., & Neumann, U. (2019). Disn: Deep implicit surface network for high-quality single-view 3d reconstruction. *arXiv preprint arXiv:1905.10711*.

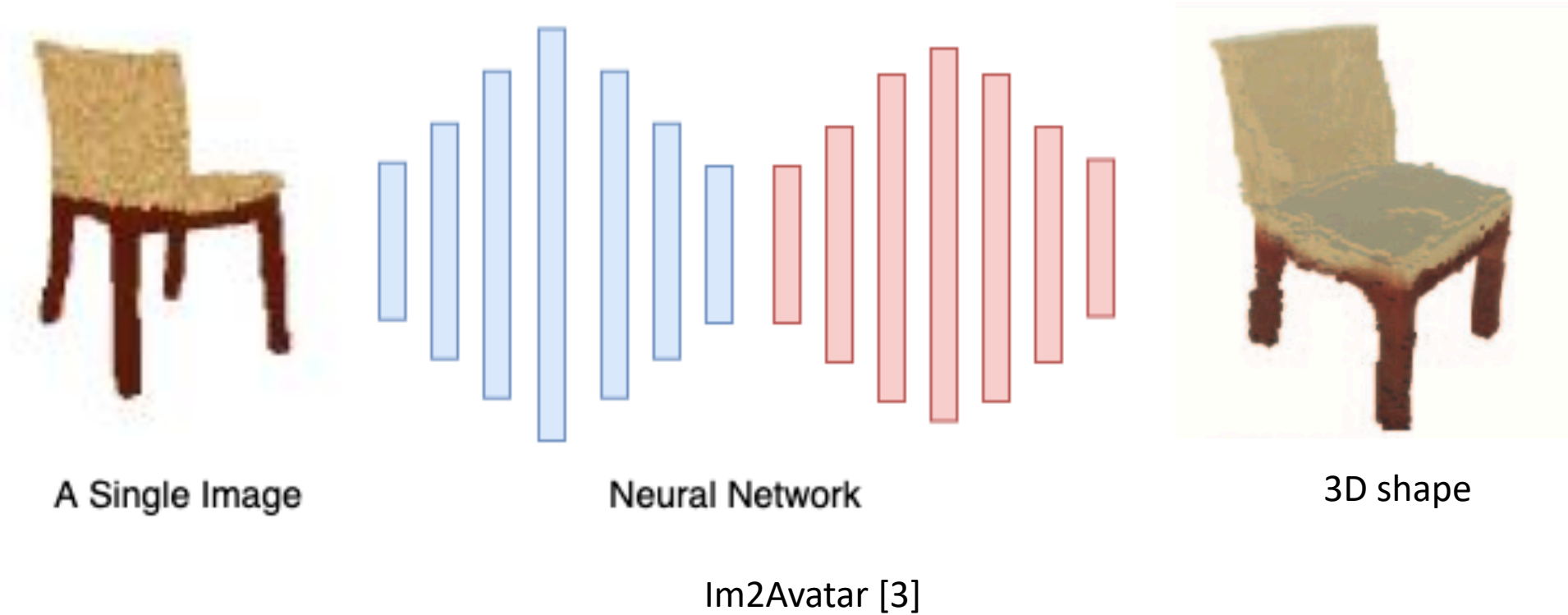
[2]: Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., & Geiger, A. (2019). Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4460-4470).



Related Work

Some of them try to estimate color information

- Im2Avatar [3]
- PIFu [4]



[3]: Sun, Y., Liu, Z., Wang, Y., & Sarma, S. E. (2018). Im2avatar: Colorful 3d reconstruction from a single image. *arXiv preprint arXiv:1804.06375*.

[4]: Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., & Li, H. (2019). Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 2304-2314).



Related Work

PIFu [4]

- requires two stages (shape first, then using the result to predict colors)
- only focus on the human

Im2Avatar [3]

- applies one stage inference
- uses 2 models for shape and color texture
- needs to train different models for different classes.

[3]: Sun, Y., Liu, Z., Wang, Y., & Sarma, S. E. (2018). Im2avatar: Colorful 3d reconstruction from a single image. *arXiv preprint arXiv:1804.06375*.

[4]: Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., & Li, H. (2019). Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 2304-2314).



Approach

Inspired by

SPSG [5]

- Completes the textured 3D scene in a single end-to-end model
- Apply implicit function (TSDF)

Im2Avatar [3]

- Jointly predicting color and shape is accessible, but it deteriorates performance slightly

[3]: Sun, Y., Liu, Z., Wang, Y., & Sarma, S. E. (2018). Im2avatar: Colorful 3d reconstruction from a single image. *arXiv preprint arXiv:1804.06375*.

[4]: Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., & Li, H. (2019). Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 2304-2314).

[5]: Dai, A., Siddiqui, Y., Thies, J., Valentin, J., & Nießner, M. (2020). Spsg: Self-supervised photometric scene generation from rgb-d scans. *arXiv preprint arXiv:2006.14660*.

c



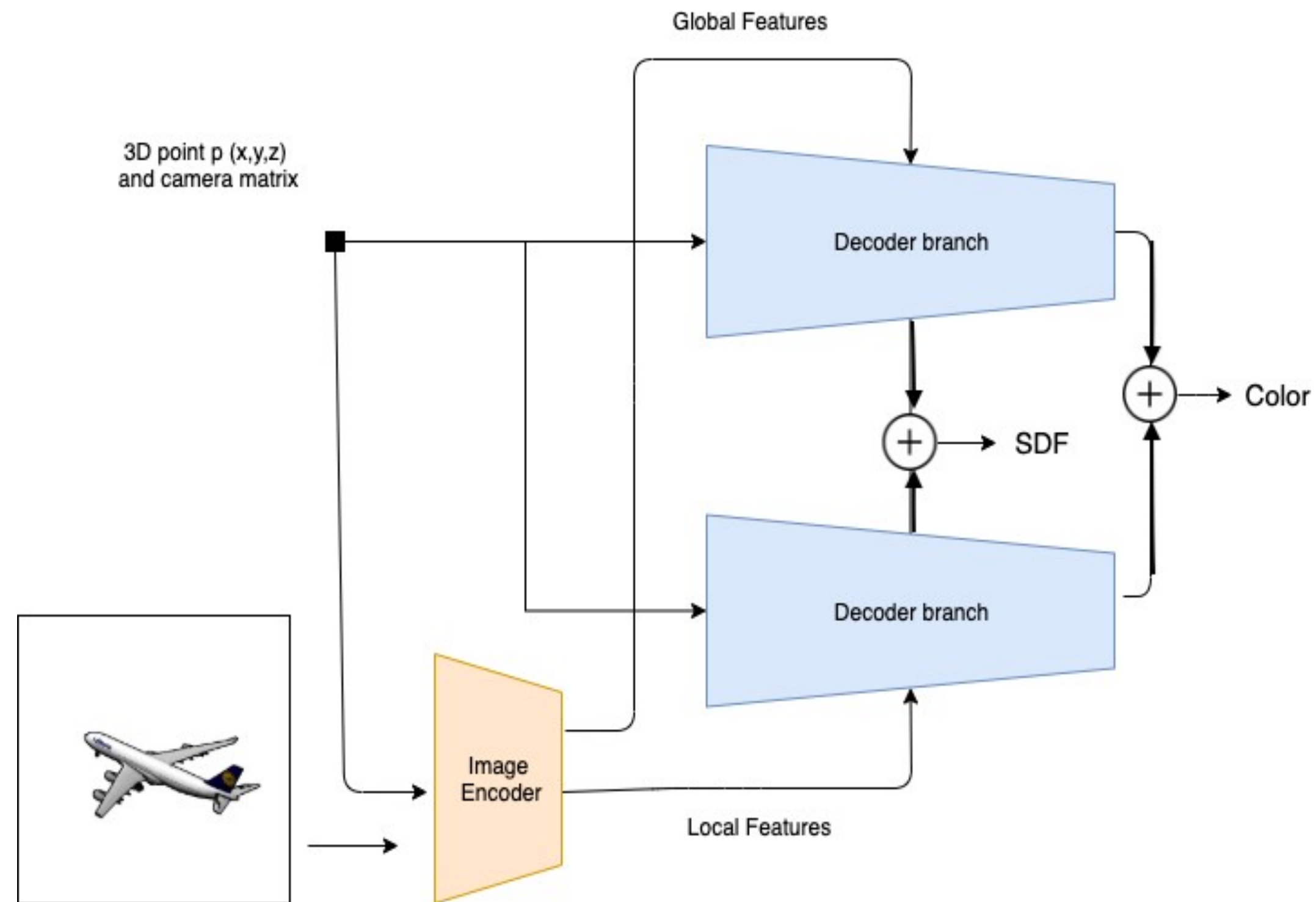
Approach

Therefore, we propose a deep implicit network for single-view textured 3D reconstruction

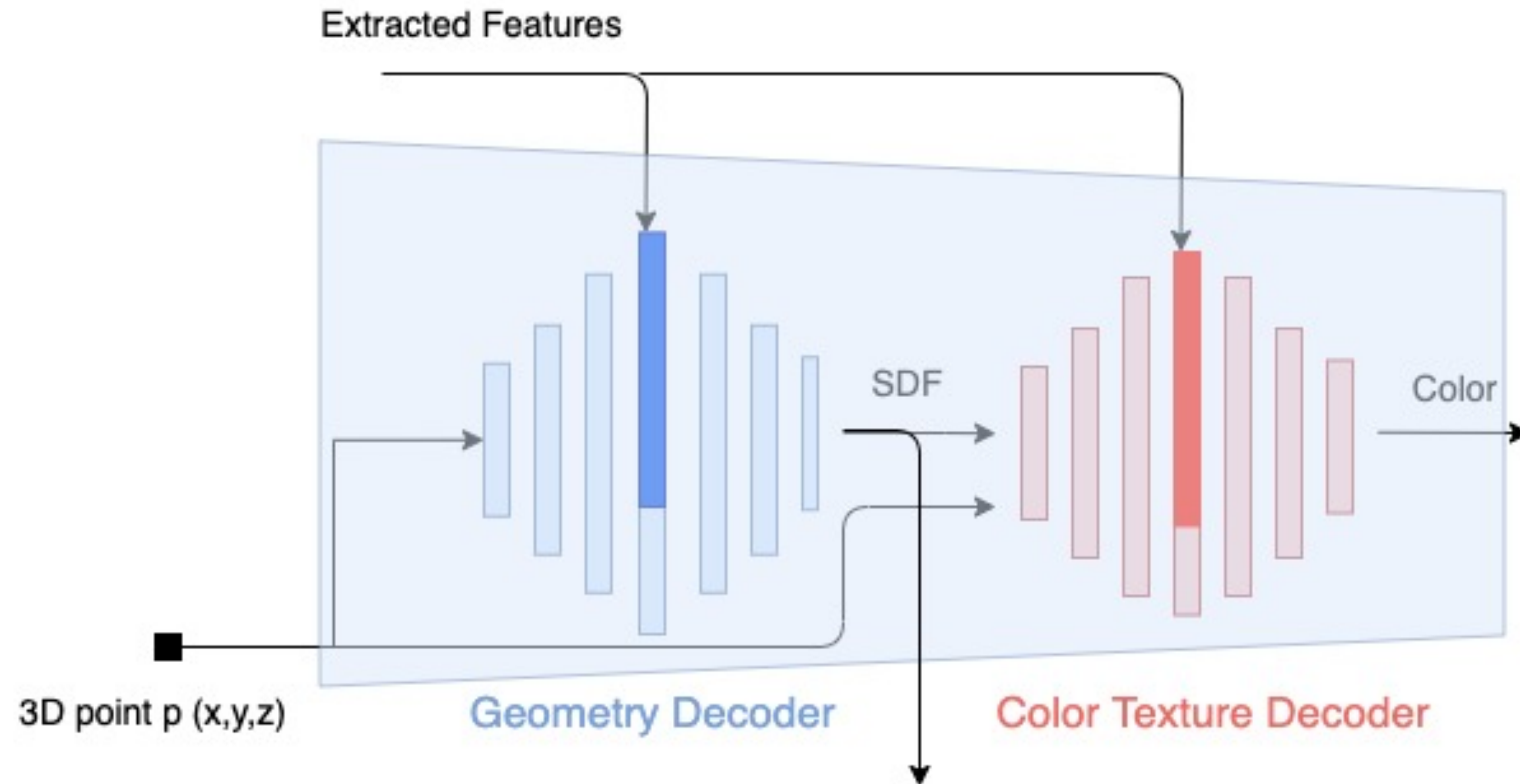
- Inherit DISN [1] structure
- Output to a colorful mesh at any resolution
- Jointly recovers color and shape within a single inference
- Multi-classes recovery
- Does not sacrifice the precision of geometry



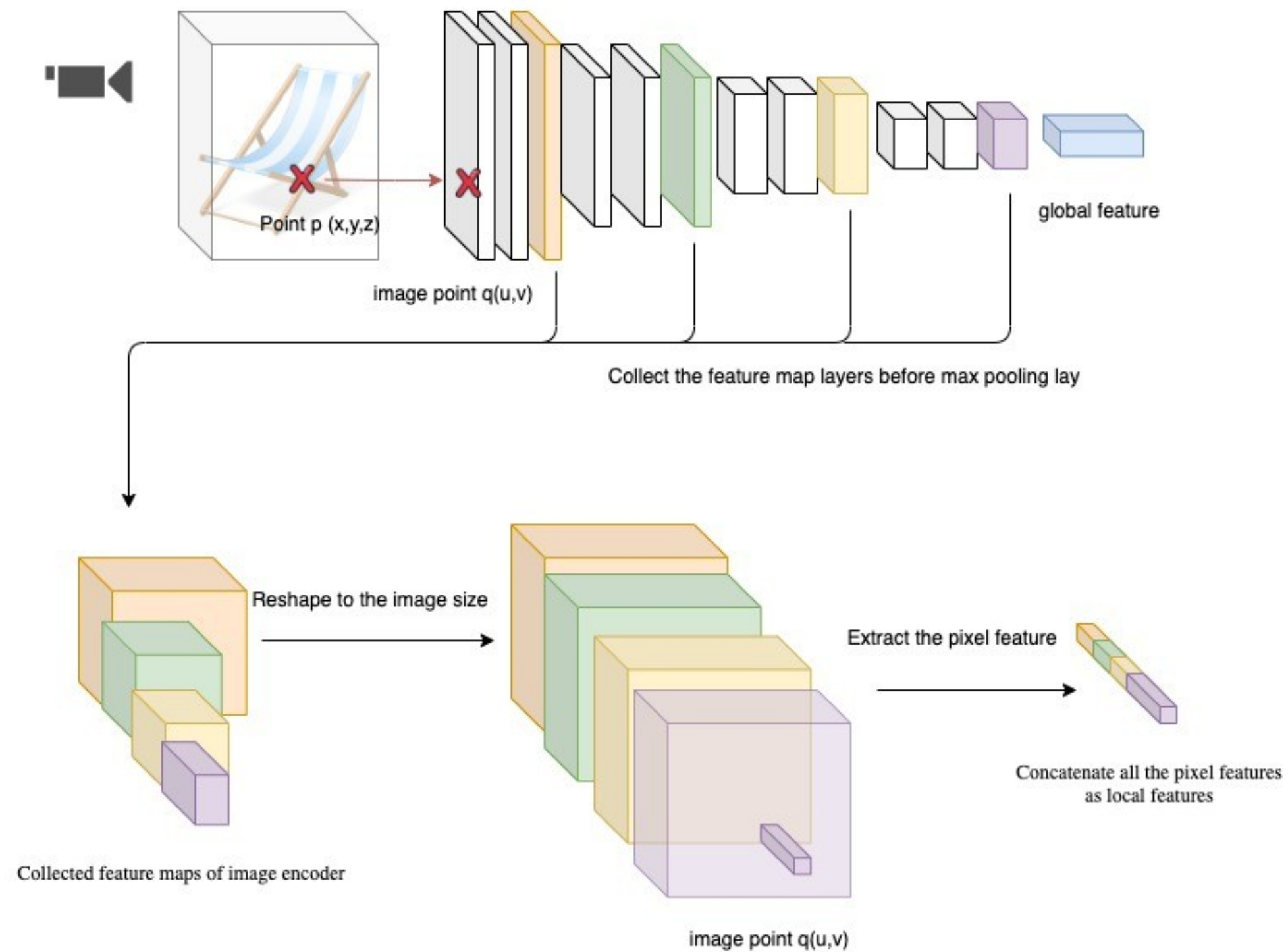
Approach



Decoder Branch



Local Feature Extraction



Experiments

- Competing methods : Im2Avatar [3], adapted R2N2 [6]
- Dataset : 4 categories (car, chair, table and guitar) Shapenet [7]
- Metrics : IoU (shape), PSNR (color)





















[3]: Sun, Y., Liu, Z., Wang, Y., & Sarma, S. E. (2018). Im2avatar: Colorful 3d reconstruction from a single image. *arXiv preprint arXiv:1804.06375*.

[6]: Choy, C. B., Xu, D., Gwak, J., Chen, K., & Savarese, S. (2016, October). 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In *European conference on computer vision* (pp. 628-644). Springer, Cham.

[7]: Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., ... & Yu, F. (2015). Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*.























Qualitative results

Input view	Adapted R2N2	Im2Avatar	Ours
			
			
			
			
			















Qualitative results

Input view	Adapted R2N2	Im2Avatar	Ours
			
			
			
			
			















Qualitative results

Input view	Adapted R2N2	Im2Avatar	Ours
			
			
			







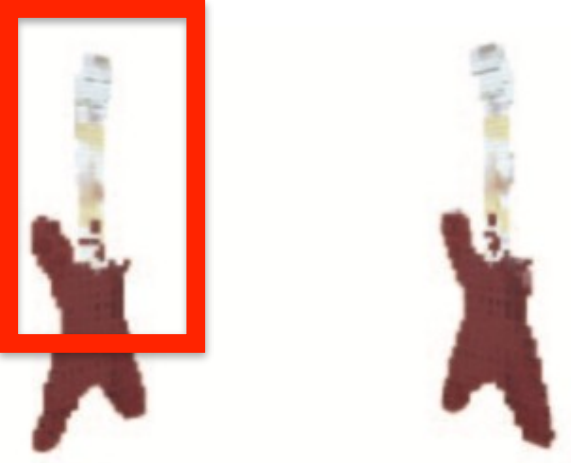







Qualitative results

Input view	Adapted R2N2	Im2Avatar	Ours
			
			
			



Qualitative results

Input view	Adapted R2N2	Im2Avatar	Ours
			
			
			



Shape Analysis

	car	table	chair	guitar	avg
Adapted R2N2	0.238	0.216	0.179	0.308	0.235
Im2Avatar	0.395	0.2381	0.180	0.374	0.298
Ours	0.419	0.314	0.252	0.376	0.340
Ours (only shape)	0.415	0.325	0.262	0.376	0.344

Table 5.2: Mean IoU on testing samples



Shape Analysis

	car	table	chair	guitar	avg
Adapted R2N2	0.238	0.216	0.179	0.308	0.235
Im2Avatar	0.395	0.2381	0.180	0.374	0.298
Ours	0.419	0.314	0.252	0.376	0.340
Ours (only shape)	0.415	0.325	0.262	0.376	0.344

Table 5.2: Mean IoU on testing samples



Shape Analysis



Illustration of examples in chair and table classes. Especially for the legs of chairs and tables, the cases are too diverse to reconstruct by the general shape



Textured Color Analysis

	car	table	chair	guitar	avg
Adapted R2N2	6.502	6.901	7.063	5.668	6.533
Im2Avatar	10.881	19.728	18.252	10.617	14.870
Ours	10.90	17.822	15.710	10.820	13.813
Ours (only chair)	- -	- -	18.514	- -	- -

Table 5.3: Mean PSNR on testing samples



Angle Analysis

	car	table	chair	guitar
90°	0.421	<u>0.298</u>	0.260	0.372
120°	0.421	0.316	0.259	0.381
150°	0.420	0.318	0.253	0.379
180°	<u>0.403</u>	0.314	<u>0.234</u>	0.376
210°	0.419	0.318	0.252	0.378
240°	0.420	0.318	0.257	0.381
270°	0.420	<u>0.300</u>	0.261	<u>0.370</u>
300°	0.423	0.316	0.257	0.382
330°	0.423	0.318	0.251	0.372
0°	<u>0.406</u>	0.313	<u>0.241</u>	<u>0.369</u>
30°	0.423	0.316	0.249	0.374
60°	0.423	0.318	0.256	0.381

Table 5.4: IoU under 12 different viewpoints

Underlined number indicates the worst two predictions of testing case



Angle Analysis



(a) Car: 180°



(b) Table: 0°



(c) Chair: 180°



(d) Guitar: 270°



(e) Car: 0°



(f) Table: 270°



(g) Chair: 0°



(h) Guitar: 0°

Illustration of the viewpoints of the worst cases in each category. Each label indicates the name of category and the angle of the viewpoint.



Angle Analysis

	car	table	chair	guitar
90°	11.02	17.73	15.62	10.86
120°	11.00	17.82	15.73	11.08
150°	10.89	17.80	15.72	10.76
180°	<u>10.45</u>	17.77	<u>15.55</u>	<u>10.74</u>
210°	10.89	18.06	15.64	10.90
240°	11.02	17.94	15.75	10.95
270°	11.05	<u>17.65</u>	<u>15.61</u>	10.87
300°	11.06	17.85	15.77	10.80
330°	10.96	17.76	15.79	10.76
0°	<u>10.44</u>	<u>17.71</u>	15.88	<u>10.55</u>
30°	10.94	17.83	15.78	10.75
60°	11.06	17.95	15.69	10.83

Table 5.5: PSNR under 12 different viewpoints

Underlined number indicates the worst two predictions of testing case



Failure Analysis



(a) Chair



(b) Table



(c) Guitar



(d) Car



(e) Chair



(f) Table



(g) Guitar



(h) Car

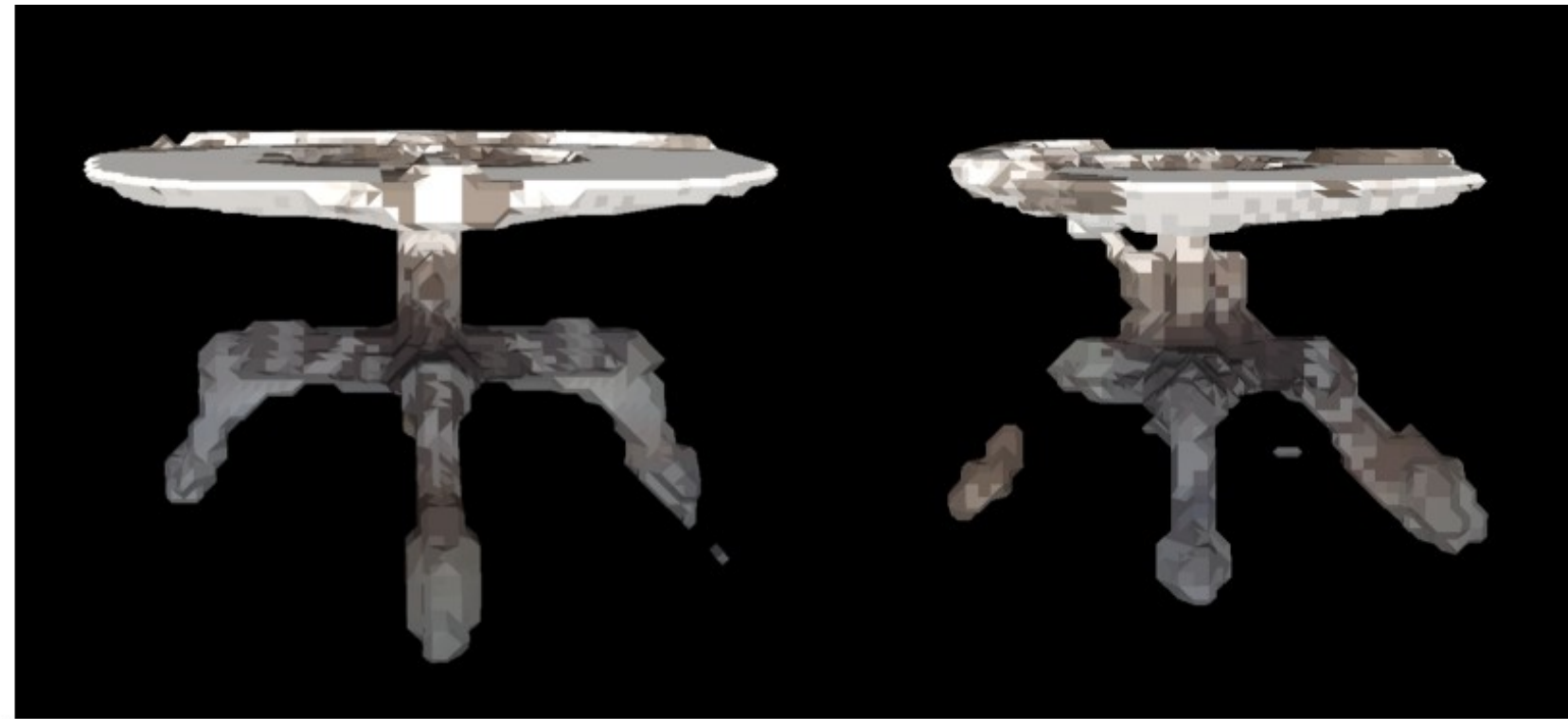
Example of white objects. In each category, there are white objects which are too difficult to identify, so we put the label to every example image



Failure Analysis





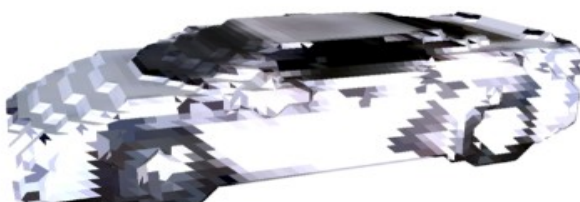



(a) Input view



(b) Result

Visualization of failed shape prediction. Figure (a) is a white chair as the input image. Figure (b) is the output of our model.

Failure Analysis

Input view	Results	
		
		



Conclusion

- Jointly recovers color and shape
- An end-to-end network
- Multi-classes recovery

Our method shows the results with smoother shape and plausible color recovery. We hence believe that our network is a useful tool which can be applied to a wide variety of 3D tasks.



Future Work

- Increase resolution of input images
- Address the problem of white object
- Increase the shape precision by sample the points with SDF distance value instead of binary SDF



Q & A



Citation

- [1]: Xu, Q., Wang, W., Ceylan, D., Mech, R., & Neumann, U. (2019). Disn: Deep implicit surface network for high-quality single-view 3d reconstruction. *arXiv preprint arXiv:1905.10711*.
- [2]: Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., & Geiger, A. (2019). Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4460-4470).
- [3]: Sun, Y., Liu, Z., Wang, Y., & Sarma, S. E. (2018). Im2avatar: Colorful 3d reconstruction from a single image. *arXiv preprint arXiv:1804.06375*.
- [4]: Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., & Li, H. (2019). Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 2304-2314).
- [5]: Dai, A., Siddiqui, Y., Thies, J., Valentin, J., & Nießner, M. (2020). Spsg: Self-supervised photometric scene generation from rgb-d scans. *arXiv preprint arXiv:2006.14660*.



Citation

- [6]: Choy, C. B., Xu, D., Gwak, J., Chen, K., & Savarese, S. (2016, October). 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In *European conference on computer vision* (pp. 628-644). Springer, Cham.
- [7]: Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., ... & Yu, F. (2015). Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*.

