DmifNet: 3D Shape Reconstruction Based on Dynamic Multi-Branch Information Fusion

Lei Li, Suping Wu *

School of Information Engineering, Ningxia University

lliicnxu@163.com wspg123@163.com

Main task

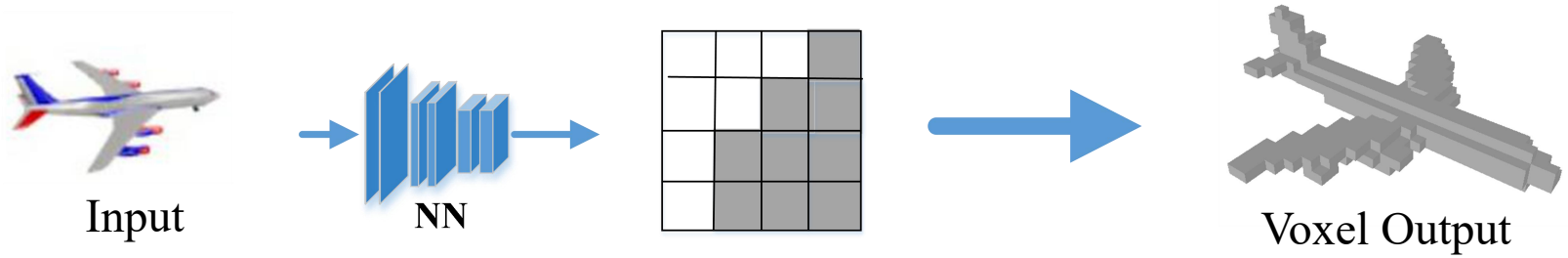




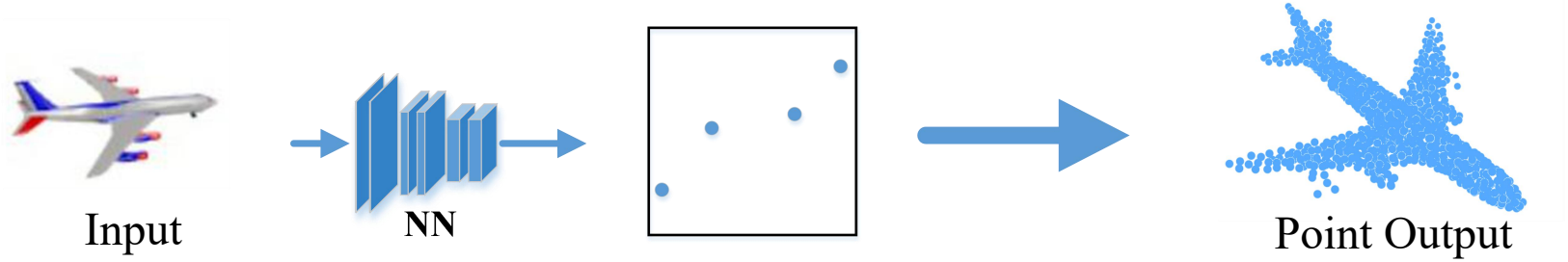




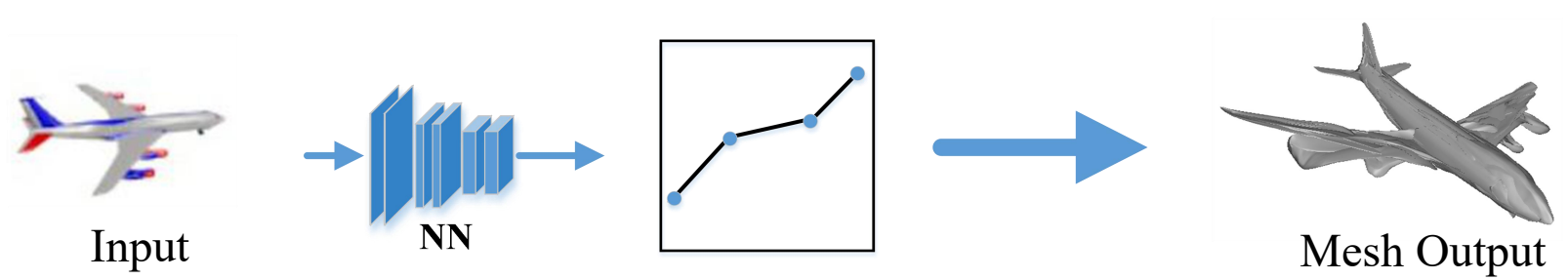

Related Work



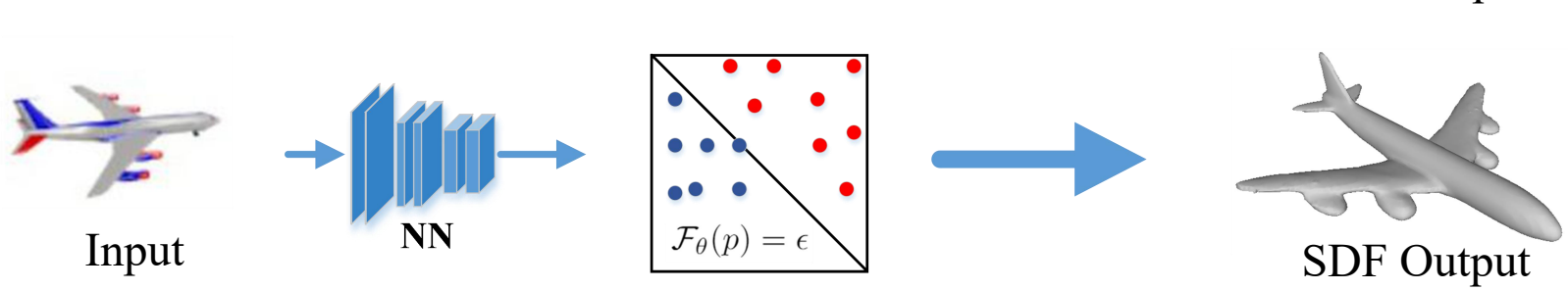
Choy et al. *ECCV* 2016



Fan et al. *CVPR* 2017



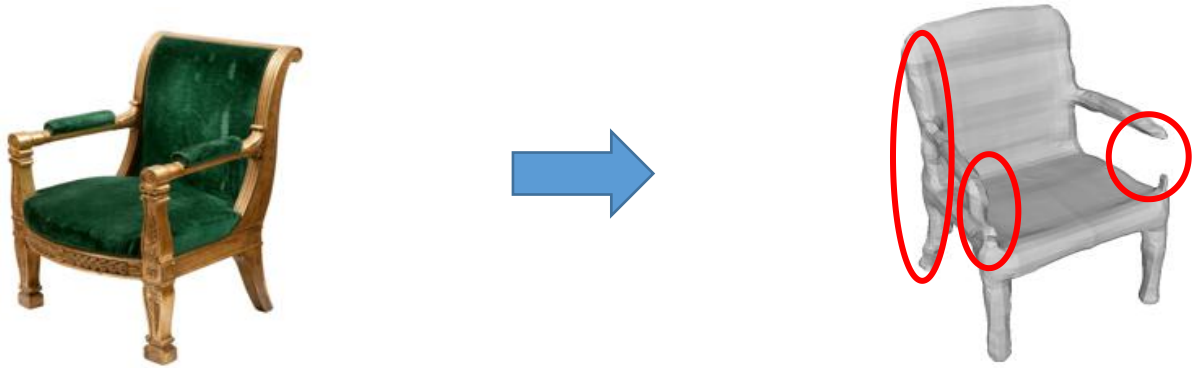
Wang et al. *ECCV* 2018



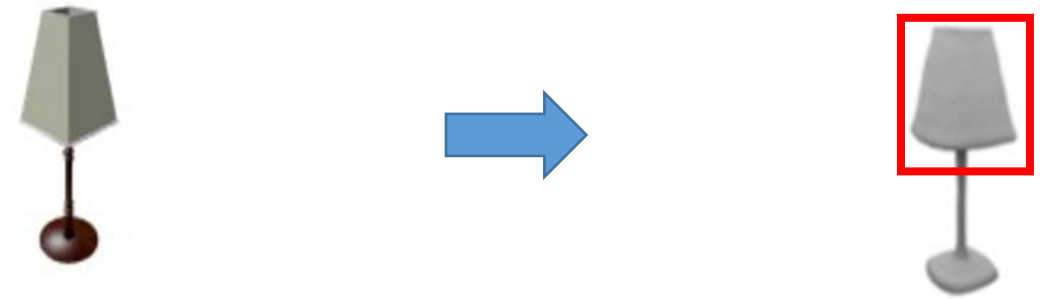
Mescheder et al. *CVPR* 2019

Issue

1. Objects with complex topology are difficult to accurately reconstruct.



2. High-frequency edge information is easy to be smoothed.



Our Motivation : To address these problems

Main Contribution of Our Method

(1) We use DoG to process input images to extract edge geometry and corners information. Because we realize the object edge geometry and corners information are important for neural network to capture complex topology.

(2) We design side branches from the intermediate layers of our neural network, so each side branch produces more diverse representations along its own pathway.

(3) We Unlike previous methods computing the average value or fixed weight of all branches predicted probability, we dynamically fuse the predicted probability of all branches to obtain the final predicted probability

(4) Extensive evaluation on a large-scale publicly available dataset ShapeNet demonstrates our method can achieve higher evaluation results than the state-of-the-art methods.

Proposed Framework

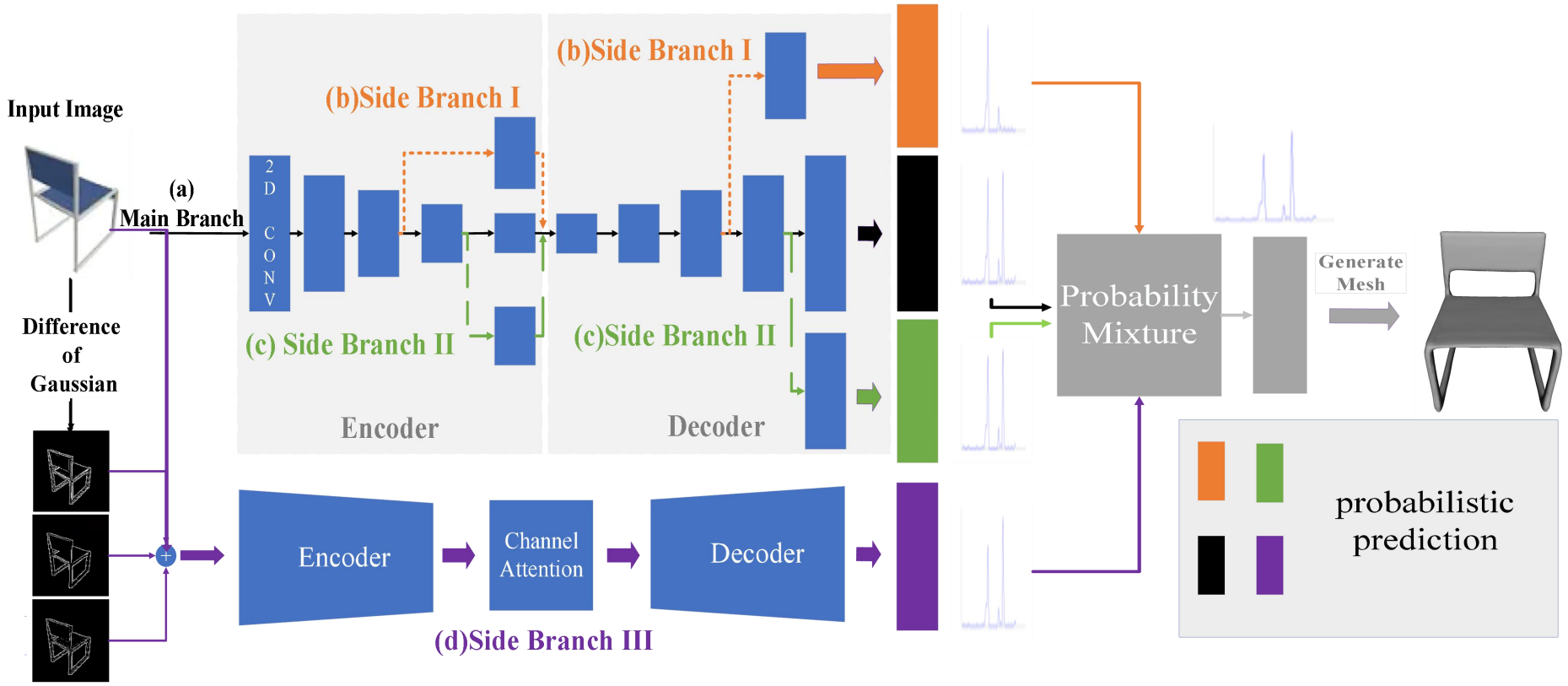
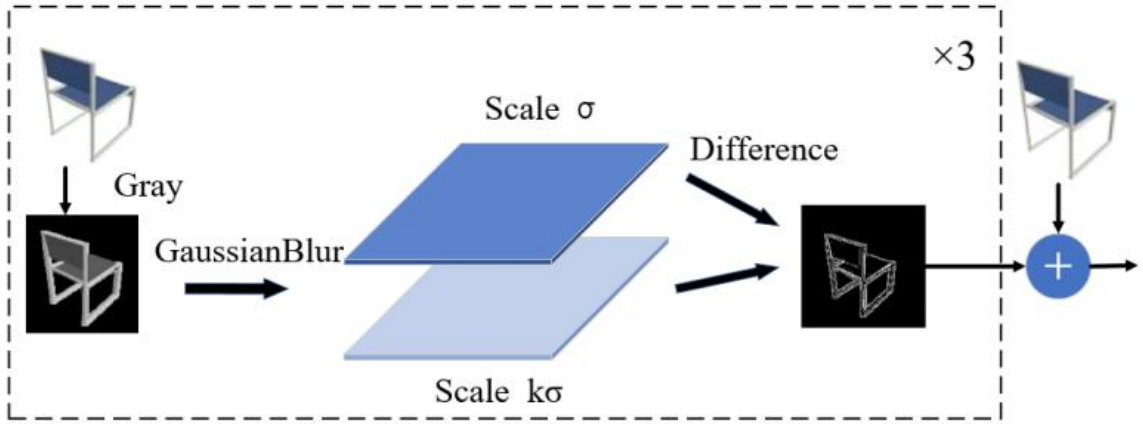


Fig. 1 The workflow of the proposed DmifNet framework. Our model has a main branch and three side branches: (a) The main branch uses the autoencoder to process the sample data and get the prediction results. (b)(c)Branches I and II process data by exploiting sub-branches from different intermediate layers of the main branch. (d)Branch III first uses the DoG to process the samples to obtain the Gaussian difference map, then we concatenate the original input image and Gaussian difference map as input information to predict result. Finally, we dynamically fuse the prediction results of the main branch and the side branches to get final prediction results.

Difference of Gaussians

1. The process of side branch III preprocessing input image



2. Formulation

$$
\begin{aligned}
Diff &= F_{i+1}(x,y) - F_i(x,y) \\
&= (G_{\sigma_{i+1}}(x,y) - G_{\sigma_i}(x,y)) f(x,y) \\
&= \frac{1}{2\pi^{\frac{d}{2}}} \left(\frac{1}{\sigma_{i+1}^d} exp(-\frac{r^2}{2\sigma_{i+1}^2}) - \frac{1}{\sigma_i^d} exp(-\frac{r^2}{2\sigma_i^2}) \right) f(x,y)
\end{aligned}
$$

Multi-branch Consistent Optimization

$$Loss = \frac{1}{|B|}\sum_{i=1}^{|B|}\sum_{j=1}^{K} L_{M_{CE}}(f_{\theta;I_{\Theta_n}}(p_{ij}, x_i), o_{ij})$$
$$+ \frac{1}{|B|}\sum_{n=1}^{N}\sum_{i=1}^{|B|}\sum_{j=1}^{K} L_{S_{CE}}(f_{\Theta n;I_\theta}(p_{ij}, x_i), o_{ij})$$

we use a consistent optimization goal to optimize the multi-branch network. By the objective function equation, we not only directly collect the classification loss of each branch to optimize the network, but also pay attention to the different representations of each branch in its pathway.

Quantitative Results on ShapeNet

Tab1. Quantitative evaluations on the ShapeNet under IoU, Normal consistency and Chamfer distance. We observe that our method approach outperforms other state-of-the-art learning based methods in Normal consistency and IoU.

IoU ↑	Airplane	Bench	Cabinet	Car	Chair	Display	Lamp	Loudspeaker	Rifle	Sofa	Table	Telephone	Vessel	Mean
3D-R2N2 [2] ECCV'16	0.426	0.373	0.667	0.661	0.439	0.440	0.281	0.611	0.375	0.626	0.420	0.6118	0.482	0.493
Pix2Mesh [10] ECCV'18	0.420	0.323	0.664	0.552	0.396	0.490	0.323	0.599	0.402	0.613	0.395	0.661	0.397	0.480
AtlasNet [16] CVPR'18	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ONet [17] CVPR'19	0.571	0.485	0.733	0.737	0.501	0.471	0.371	0.647	0.474	0.680	0.506	0.720	0.530	0.571
Our	**0.603**	**0.512**	**0.753**	**0.758**	**0.542**	**0.560**	**0.416**	**0.675**	**0.493**	**0.701**	**0.550**	**0.750**	**0.574**	**0.607**
Normal Consistency ↑	Airplane	Bench	Cabinet	Car	Chair	Display	Lamp	Loudspeaker	Rifle	Sofa	Table	Telephone	Vessel	Mean
3D-R2N2 [2] ECCV'16	0.629	0.678	0.782	0.714	0.663	0.720	0.560	0.711	0.670	0.731	0.732	0.817	0.629	0.695
Pix2Mesh [10] ECCV'18	0.759	0.732	0.834	0.756	0.746	0.830	0.666	0.782	0.718	0.820	0.784	0.907	0.699	0.772
AtlasNet [16] CVPR'18	0.836	0.779	0.850	0.836	0.791	0.858	0.694	0.825	0.725	0.840	0.832	0.923	0.756	0.811
ONet [17] CVPR'19	0.840	0.813	0.879	0.852	0.823	0.854	0.731	0.832	0.766	0.863	0.858	0.935	0.794	0.834
Our	**0.853**	**0.821**	**0.885**	**0.857**	**0.835**	**0.872**	**0.758**	**0.847**	**0.781**	**0.873**	**0.868**	**0.936**	**0.808**	**0.846**
Chamfer-L_1 ↓	Airplane	Bench	Cabinet	Car	Chair	Display	Lamp	Loudspeaker	Rifle	Sofa	Table	Telephone	Vessel	Mean
3D-R2N2 [2] ECCV'16	0.227	0.194	0.217	0.213	0.270	0.314	0.778	0.318	0.183	0.229	0.239	0.195	0.238	0.278
Pix2Mesh [10] ECCV'18	0.187	0.201	0.196	0.180	0.265	0.239	0.308	0.285	0.164	0.212	0.218	0.149	0.212	0.216
AtlasNet [16] CVPR'18	**0.104**	**0.138**	0.175	**0.141**	0.209	**0.198**	**0.305**	**0.245**	**0.115**	**0.177**	0.190	0.128	**0.151**	**0.175**
ONet [17] CVPR'19	0.147	0.155	0.167	0.159	0.228	0.278	0.479	0.300	0.141	0.194	0.189	0.140	0.218	0.215
Our	0.131	0.141	**0.149**	0.142	**0.203**	0.220	0.351	0.263	0.135	0.181	**0.173**	**0.124**	0.189	0.185

[a] The Bold-faced numbers represent the best results.

Qualitative Results on ShapeNet

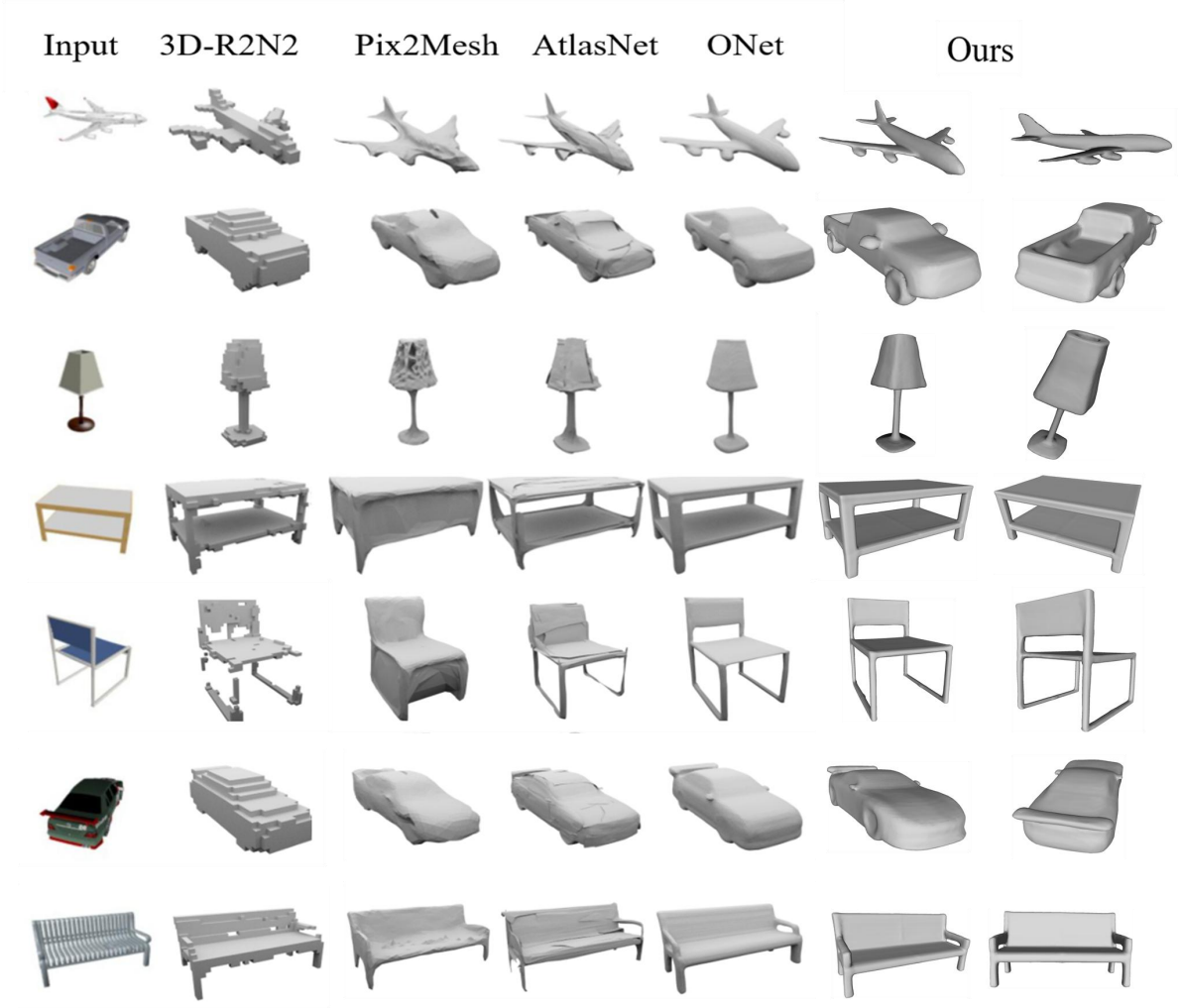
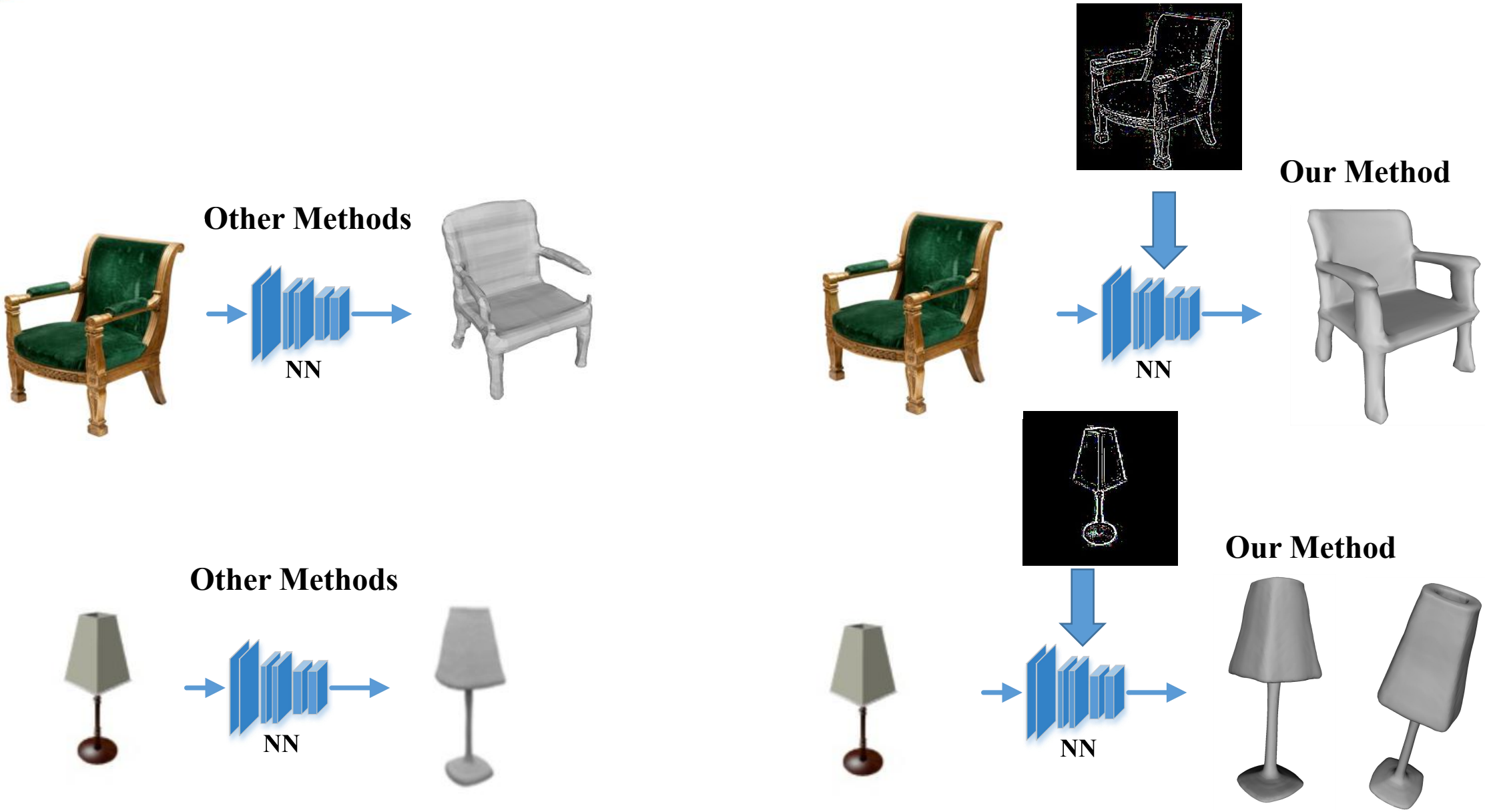


Fig. 2 Single Image 3D Reconstruction on ShapeNet. The first column is the input 2D images, the last two columns are the results of our method. The other columns show the results for various methods.

Qualitative Results on ShapeNet



Qualitative Results on Real Data

Fig.3 Single Image 3D Reconstruction on Real Data. The first column is the input 2D images, the other columns are the reconstructed results of our method in different viewpoints.

Reference

[1] Jiajun Wu, Chengkai Zhang, Xiuming Zhang, Zhoutong Zhang,William T Freeman, and Joshua B Tenenbaum, "Learning shape priors for single-view 3d completion and reconstruction," in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 646–662.

[2] Haoqiang Fan, Hao Su, and Leonidas J Guibas, "A point set generation network for 3d object reconstruction from a single image," in Proceedings of the IEEE conference on computer vision and pattern recognition,2017, pp. 605–613.

[3] Christopher B Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese, "3d-r2n2: A unified approach for single and multiview 3d object reconstruction," in European conference on computer vision. Springer, 2016, pp. 628–644.

[4] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang, "Pixel2mesh: Generating 3d mesh models from single rgb images," in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 52–67.

[5] Qiangeng Xu, Weiyue Wang, Duygu Ceylan, Radomir Mech, and Ulrich Neumann, "Disn: Deep implicit surface network for high-quality singleview 3d reconstruction," in Advances in Neural Information Processing Systems, 2019, pp. 492–502.

[6]Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger, "Occupancy networks: Learning 3d reconstruction in function space," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 4460–4470.

[7]Duo Li and Qifeng Chen, "Dynamic hierarchical mimicking towards consistent optimization objectives," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 7642–7651.


1958
NINGXIA UNIVERSITY


ICPR2020
25th INTERNATIONAL CONFERENCE
ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Thank you !