Deep Shape from a Low Number of Silhouettes

Xinhan Di(✉), Rozenn Dahyot, and Mukta Prasad

School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland
{dixi,Rozenn.Dahyot,prasadm}@tcd.ie

Abstract. Despite strong progress in the field of 3D reconstruction from multiple views, holes on objects, transparency of objects and textureless scenes, continue to be open challenges. On the other hand, silhouette based reconstruction techniques ease the dependency of 3d reconstruction on image pixels but need a large number of silhouettes to be available from multiple views. In this paper, a novel end to end pipeline is proposed to produce high quality reconstruction from a low number of silhouettes, the core of which is a deep shape reconstruction architecture. Evaluations on ShapeNet [1] show good quality of reconstruction compared with ground truth.

Keywords: Deep 3D reconstruction · End to end architecture · Silhouettes

1 Introduction

3D geometry reconstruction techniques have made significant progress during this two decades from theoretical development to software implementation. The aim of both contributions are to build high-quality 3D reconstruction of scenes and objects from 2D or 2.5D source information in terms of image pixels, depth and other source data. Current progress in this filed involves wider application of 3D reconstruction towards real world application including video data application, large-scale scene reconstruction, light field reconstruction and other applications. However, several open questions remain challenging for 3D reconstruction such as holes, wrinkles, coarse region and other unwanted artifacts in the 3D rebuilt world for the reconstruction of transparent objects, textureless scenes and other challenging objects and scenes (Fig. 1).

Among the two most popular 3D reconstruction technique groups including the increment multiview reconstruction and volume-based reconstruction, the first technique framework is based on 2D image source information. Lots of techniques and theories have been developed to produce high-quality 3D reconstruction. Furthermore, for the 3D reconstruction of challenging objects such as transparent objects and objects containing textureless parts, lots of priors including surface normals, object-specific shape priors and other priors are applied and integrated in the reconstruction systems. These priors are demonstrated to improve the quality of 3D reconstruction in order to avoid of holes, wrinkles, and other artifacts.

G. Hua and H. Jégou (Eds.): ECCV 2016 Workshops, Part III, LNCS 9915, pp. 251–265, 2016.
DOI: 10.1007/978-3-319-49409-8_21

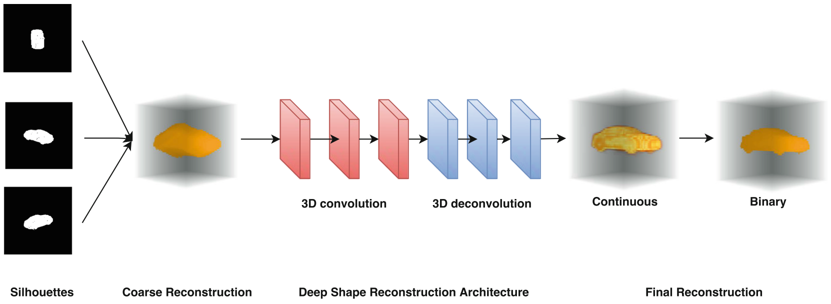

Fig. 1. Two stage direct 3D reconstruction pipeline

The second technique groups build the 3D real world in the grid space consists of voxels. Priors such as connectivity priors, surface orientation priors are represented as data constraint terms. These terms are calculated in a mathematical framework such as multiple label convex framework or MRF pipelines to provide solutions to the opening challenges. Besides, sufficient information of viewpoints including camera parameter matrix of each viewpoint, a large number of silhouettes and 2D image pixels are the base of quality of 3D reconstruction in the grid space.

The proposed pipeline is a deep reconstruction pipeline consisting of two reconstruction stages. The first stage is shape coarse reconstruction stage. It takes a small number of silhouettes as input with known camera parameters of the associated viewpoints, and produces a coarse visual hull. The second stage is deep shape reconstruction stage, it works based on a deep shape reconstruction architecture. This proposed 3D convolution networks (3D-CNNs) architecture reconstructs good quality shapes from coarse shapes. The currently proposed pipeline is designed for reconstructing category-specific object shapes.

Two contributions are made. First, this pipeline produces high quality of 3D reconstruction based on a low number of silhouettes. It is not dependent on images pixels, depth data and et al. Therefore, this silhouettes based technique is considered as a potential solution for 3D reconstruction of transparent objects or objects with textureless parts. Second, compared with techniques relying on a large number of images taken in different viewpoints, the proposed pipeline reduces the number of views.

This paper is organized as follows. We review related work in Sect. 2. We formulate the problem in Sect. 3. Both the solution to the formulated problem and the proposed reconstruction pipeline is presented in Sect. 4. Evaluation details including the dataset, traning, test and results are shown in Sect. 5. Finally, conclusion and future work for our pipeline is discussed in Sect. 6.

2 Related Work

End to end deep learning architectures have been applied successfully to a variety of vision problems such as segmentation [2–5], edge prediction [6], classification [7], optical flow prediction [8], depth prediction [9], keypoint prediction [4] and feature learning [10,11]. Also, fully convolutional networks prove their efficiency in one dimensional input strings [12] extended from LeNet [13], two dimensional detection [14,15] with learning and inference, three dimensional representation [16], volumetric 3D shapes classification and interpolation [17].

Convolutional Neural Networks(CNNs) have been proven their efficiency in improving 3D reconstruction. For example, CNNs is developed for the task of prediction of surface normals from a single image [18]. A combined framework for viewpoint estimation and local keypoint prediction is proposed through application of a convolutional neural network architectures [19]. A convolutional neural network is built to perform extremely well for stereo matching [20].

Unlike the above methods, end to end deep architectures are built for 3D reconstruction. For example, convolution network is applied to build 3D models from single images [21], 3D recurrent reconstruction neural network (3D-R2N2) is built to unify both single and multi-view 3D object reconstruction [22], Semantic deformation flows are learned with 3D convolution networks for improving 3D reconstruction [23], 3D volumetric reconstruction is learned from single-view with projective transformations [24]. However, these 3D reconstruction deep architectures are dependent on image pixels, transformation and etc. In contrast, a novel reconstruction pipeline is proposed directly from a small number of silhouettes input end to 3D reconstruction end.

3 Problem Formulation

We aim at building a function V of Visual Hull H_k (inferred from k silhouettes [25]) that reconstructs a shape as close as possible to the Ground Truth (GT) shape:

$$V(H_k) \simeq GT \tag{1}$$

It is commonly understood that the visual hull improves as the number k of silhouettes increases such that:

$$\lim_{k \to \infty} H_k = H \tag{2}$$

where the limit visual hull H is the best approximation possible of the object shape estimated from silhouettes. H itself is often far away from the true shape GT as concave areas fail to be recovered from Shape-from-Silhouettes techniques [26].

To avoid these artifacts, we propose an improvement on visual hull inferred by shape-from-silhouettes techniques using a Bayesian like framework where we aim at using prior information about the object category to design a function $V(H_k) = V_k$ that is as close as possible to GT. The function V is designed using deep neural network and is trained using ShapeNets dataset [1]. Our formulation is tested for $k = 2, 3, 4, 5$ silhouettes available as input of the pipeline.

4 Two-Stage Deep Shape Reconstruction Pipeline

An end to end deep shape reconstruction pipeline is proposed. The input end is a small number of silhouettes with known camera parameters and the output end is reconstructed 3D shape in the form of volume. The volume grid consists of voxels whose values are binary. 0 represents the empty space and 1 represents the occupied space of the shape. To be noted, both space on the surface of the shape and inside the shape are represented as 1. The proposed shape reconstruction pipeline is split in two stages including coarse shape reconstruction stage and deep shape reconstruction stage. The first stage is to produce a 3D shape of object H_k with known silhouettes and corresponding camera parameters. The second stage is to reconstruct a good quality of 3D shape V_k with known H_k through a deep shape reconstruction architecture. This architecture works as a solver to the formulated problem. H_k is produced through a common method of intersection of known silhouette cones [25]. And the architecture of the second stage is represented as follows.

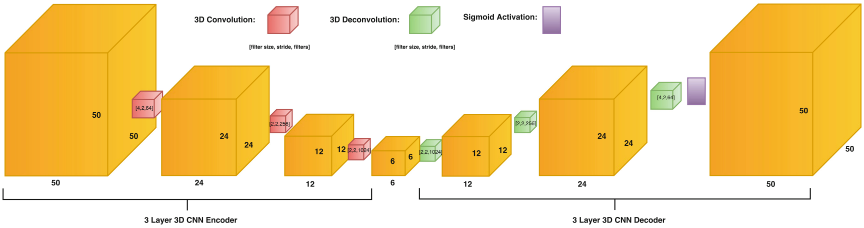

Fig. 2. Deep shape reconstruction architecture. Details of the deep shape reconstruction architecture are represented including the number of 3D convolution layers and 3D deconvolution layers, size of filters, the umber of filters and parameters of the stride

4.1 Architecture

Both the input end H_k and the output end V_k are in the forms of volume space consisting of voxels. The values of the voxels are binary. The key components of the deep learning architecture we built for the deep shape reconstruction are the convolutional encoding layers(recognition network) and convolutional decoding layers(generative network). As shown in Fig. 2, there are three 3D convolutional layers and three 3D deconvolutional layers in our deep shape reconstruction architectures. From the input end to the output end of the deep shape architecture, we use different size of filters. From the beginning layer of the recognition network, the filter size changes from $4 \times 4 \times 4$ to $2 \times 2 \times 2$, the number of filters increase from 64 to 1024. In contrast, from the beginning layer of the generative network, the filter size changes from $2 \times 2 \times 2$ to $4 \times 4 \times 4$, and the number of filters reduces from 1024 to 64.

4.2 Convolution and Deconvolution

CNNs are good at extracting high-level abstract information of data by interleaving convolutional and deconvolutional layers, pooling and spatially shrinking. For both convolutional and deconvolutional layers, learnable filters are important components for the stronger learning ability of these layers. A filter of a trainable convolutional layer acts as a learnable local down-sampling unit and the filter of a trainable deconvolutional layer acts as a learnable local up-sampling unit. In 3D convolution, 3D input signals are convolved by the kernel filter and the values are placed on the output 3D grid. Conversely, 3D deconvolution takes the values of the input 3D grid and the result values are got through multiplying the values by weights in the filters. If one 3D filter has size $s \times s \times s$, it generates a $s \times s \times s$ output matrix for each voxel input. The output matrices can be stored overlapping and the amount of the output overlap depends on the output stride. If the amount of output stride of the convolution filter is bigger than 1, then the convolution layer produces an output with size smaller than the input and works as down-sampler. While, if the amount of input stride of the deconvolution filter is bigger than 1, the deconvolution layer produces an output with size bigger than the input and works as up-sampler.

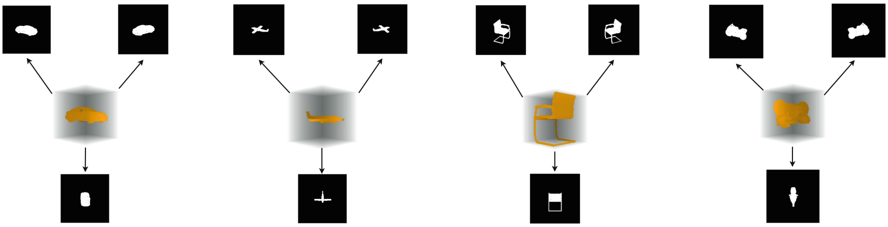

Fig. 3. Triple silhouettes and viewpoints for training deep 3D reconstruction model. For the training of 4 object categories including cars, planes, motorbikes and chairs, both silhouettes and views used to produce coarse shapes are represented

5 Evaluation

Evaluation are conducted for the proposed pipeline. First, training of a deep shape model is conducted through usage of the ShapeNet [1] dataset. Second, test is carried on to reconstruct 3D shapes from a low number of silhouettes. Finally, comparison between the reconstructed 3D shapes and the ground truth shape is made and the reconstruction errors are calculated. We did 4 experiments. The goal of each experiment is to reconstruct the 3D shape of instances of a specific object category. For conducting each experiment, category-specific deep shape model is trained and then the evaluation on the reconstruction accuracy is calculated.

5.1 Dataset

ShapeNet is presented as a richly-annotated, large-scale repository of shapes containing 3D CAD models of objects for a rich number of object categories. ShapeNets contains more than $3,000,000$ models, and $220,000$ models of which are classified into $3,135$ categories. And ShapeNets has been used in a range of deep 3D shape research work [27–29]. Here, we use a subset of ShapeNets to train our category-specific deep shape reconstruction model. For example, we use 372 car CAD models for training our car deep shape reconstruction network, and then use 110 car CAD models for testing the performance of our pipeline. The number of CAD models used in the training for other three categories including planes, motorbikes and chairs are 372, 277, 315 respectively. The number of CAD models for testing are 107, 164 and 62 respectively.

5.2 Training

In the training, both the ground truth shape and the coarse shape of each CAD model are used. The coarse shapes for training a category-specific deep shape reconstruction model are produced through applying triple silhouettes. These silhouettes are taken from three different views around each CAD model. The three views are 0°, 120°, and 240° around each CAD model. Figure 3 visualizes the fixed three views and triple silhouettes for the training. Then, with known triple silhouettes and camera parameters, the coarse shapes are produced simply from the intersection of three silhouette cones [25]. And both the ground truth shape and the coarse shape are represented in the form of volume consisting of binary voxels. Figure 4 presents ground truth shape and their corresponding H_3 used for training. To be noted, for both training and testing, the volume size of each shape is $50 \times 50 \times 50$. The category-specific reconstruction network is trained end to end from scratch and with pure stochastic gradient decent. And the learning rate is $1e^{-5}$ for 200 epoches. The value of momentum is 0.9 and the implementation is based on the open source library Torch.

5.3 Testing

In the testing, the category-specific 3D shape reconstruction pipeline is evaluated for 4 object categories including cars, planes, chairs and motorbikes. For each object category, we test the 3D deep shape reconstruction pipeline for 4 different number of views. First, we test the 3D deep shape reconstruction pipeline with input of 2 silhouettes. These 2 silhouettes are produced from 2 fixed views, 0° and 180° around a ground truth CAD model. The other three tests evaluate the performance of the deep shape reconstruction pipeline for 3 fixed views, 4 fixed views and 5 fixed views respectively. Also, the views chosen for these three tests are $[0°, 120°, 240°]$, $[0°, 90°, 180°, 270°]$ and $[0°, 72°, 144°, 216°, 288°]$ respectively. Figure 5 visualizes the four arrangements of views for the test.

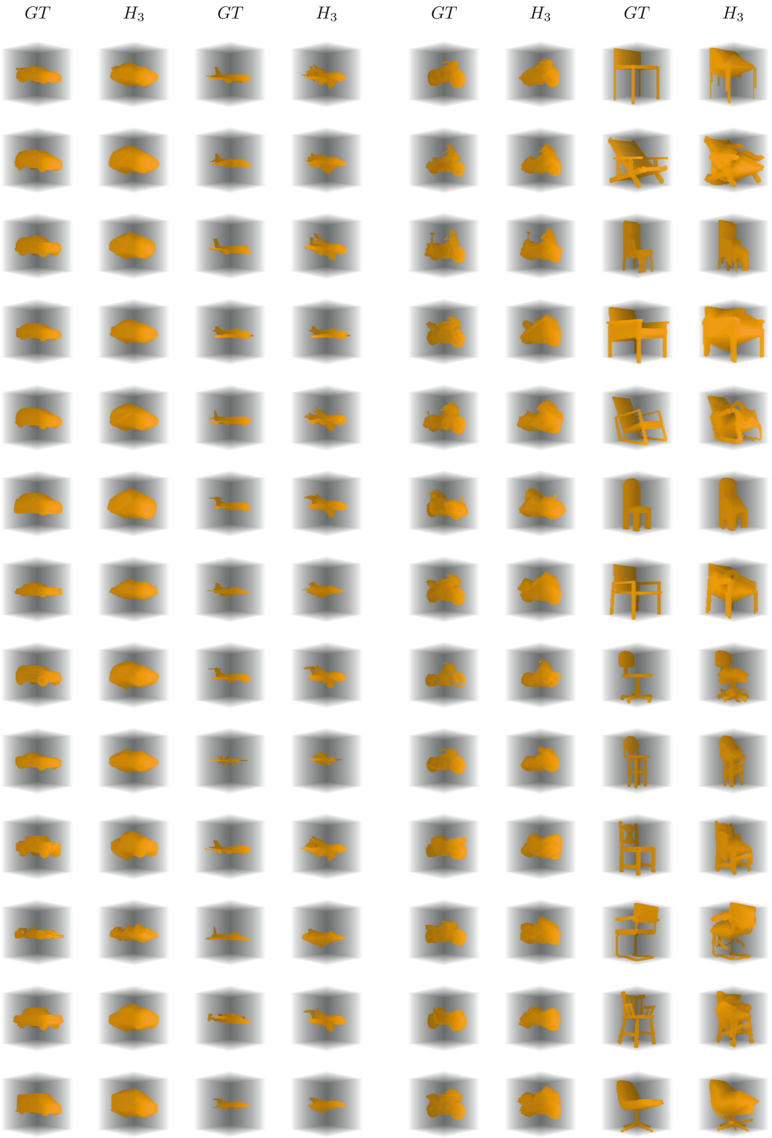

Fig. 4. Sample shapes for training deep shape reconstruction architecture. Both ground truth (GT) shapes and coarse shapes H_3 are alternatively presented from the left to right columns. Both two kinds of shapes are used for training category-specific deep shape reconstruction architecture. Samples of the shapes for training 4 object categories are all represented

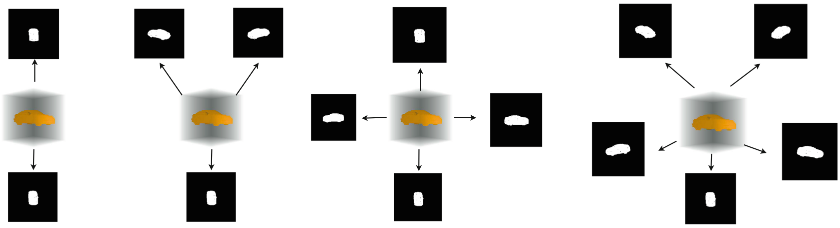

Fig. 5. Four view arrangements of silhouettes for test. The view arrangements for the test of the car object category are shown

5.4 Results

The reconstructed shape in the two stages of our deep 3D reconstruction pipeline are represented. For the test of the 4 view arrangements and 4 object categories, both the coarse shapes and final shapes are shown qualitatively and quantitatively. Both the qualitative and quantitative results show that the deep 3D reconstruction pipeline is capable of reconstructing category-specific 3D shapes with a small number of silhouettes as input. It also demonstrates that the network improves the reconstruction after coarse shape stage. The reconstruction error between V_k and GT is shown to be smaller than the reconstruction error between H_k and GT, proving that the deep shape reconstruction architecture reconstructs v_k which is better than H_k.

Qualitative Results. Figures 6, 7, 8, 9 visualize the shape reconstruction of four object categories including cars, planes, motorbikes and chairs. Ground truth shapes, shapes reconstructed in both the coarse shape reconstruction stage and the deep shape reconstruction stage are all represented in the figures.

Quantitative Results. In order to get quantitative measures of the reconstructed shape, we evaluate the 3D reconstruction in two ways. The first evaluation is the mean square error between a 3D voxel reconstruction before thresholding and its ground truth voxelized model. The second evaluation is the voxel Intersection-over-Union(IoU) between the 3D voxel reconstruction and the ground truth model. More formally,

$$Reconstruction\,Error = \frac{\sum_{i=1}^{n} |vp_i - Gvp_i|^2}{n} \tag{3}$$

where vp_i represents the final output at voxel i in a grid space before thresholding, $vp_i \in [0,1]$. And let the corresponding ground truth occupancy be Gvp_i, $Gvp_i \in \{0,1\}$. Lower error indicates better reconstruction. To be noted, we train and test in a $50 \times 50 \times 50$ grid space so the total number of voxels is determined as $n = 50 \times 50 \times 50$.

$$Voxel\,IoU = \frac{\sum_{i=1}^{n}[I(vp_i > t)\ I(Gvp_i)]}{\sum_{i=1}^{n} I(I(vp_i > t) + I(Gvp_i))} \tag{4}$$

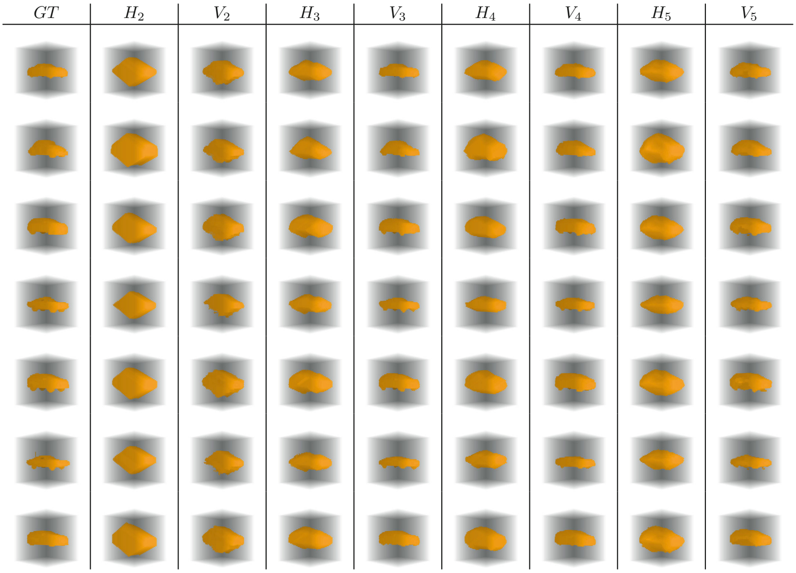

Fig. 6. Qualitative car reconstruction results. Samples of reconstructed car shapes are represented, the GT column represents the ground truth shapes, the other eight columns represent reconstructed shapes from the coarse stage and the 3D deep reconstruction stage. H_2 column represents coarse reconstructed shapes in the 2-view arrangement. V_2 column represents final reconstructed shapes in the 2-view arrangement. Shapes in other columns are represented in a similar way

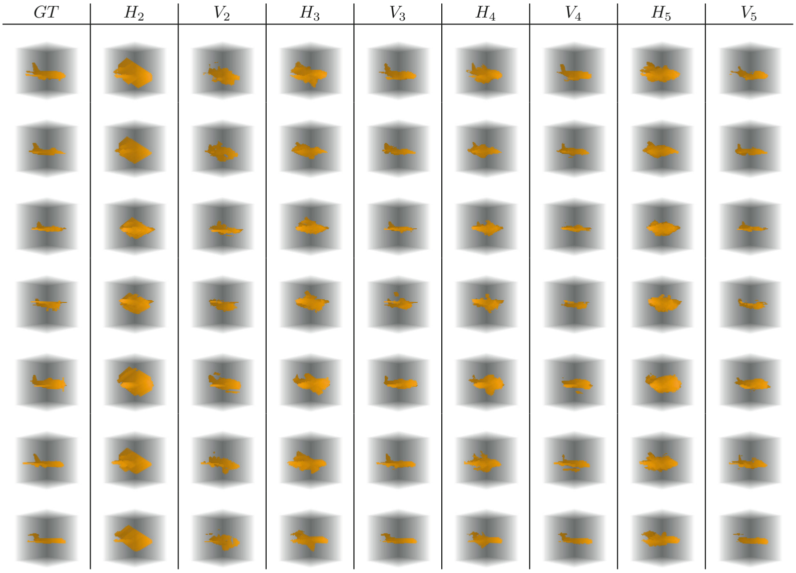

Fig. 7. Qualitative plane reconstruction results

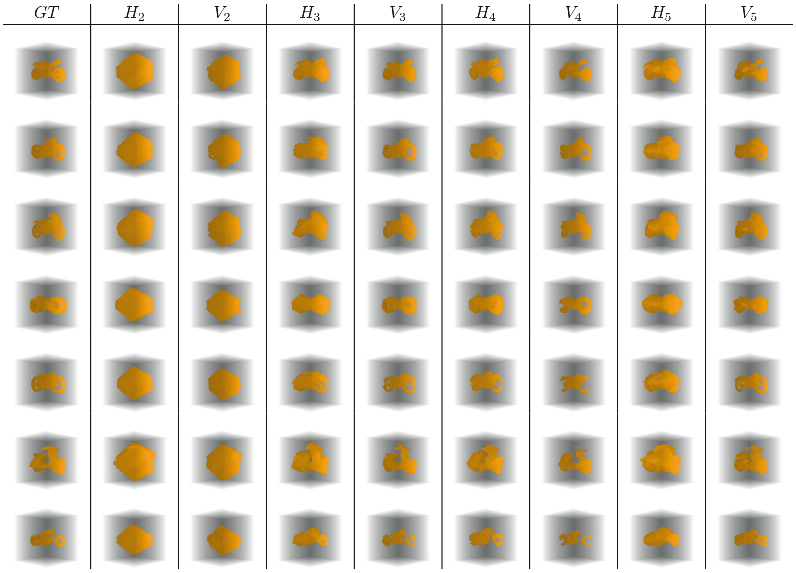

Fig. 8. Qualitative motorbike reconstruction results

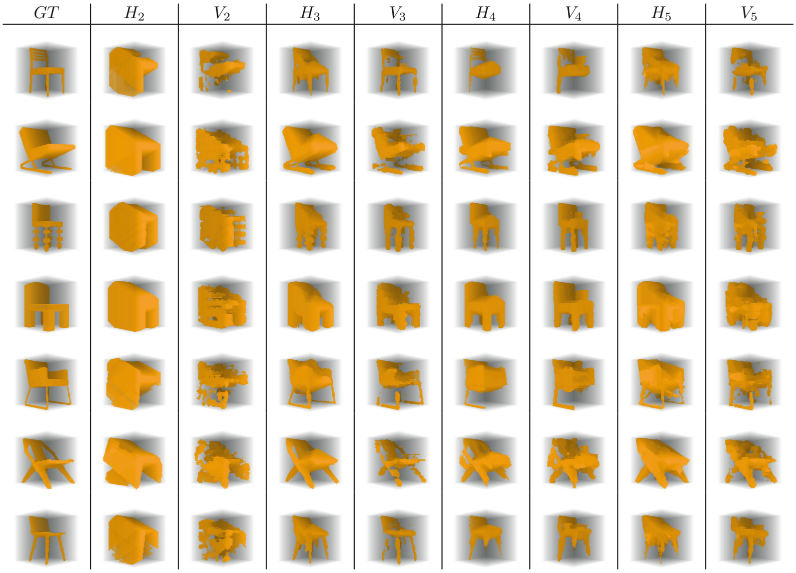

Fig. 9. Qualitative chair reconstruction results

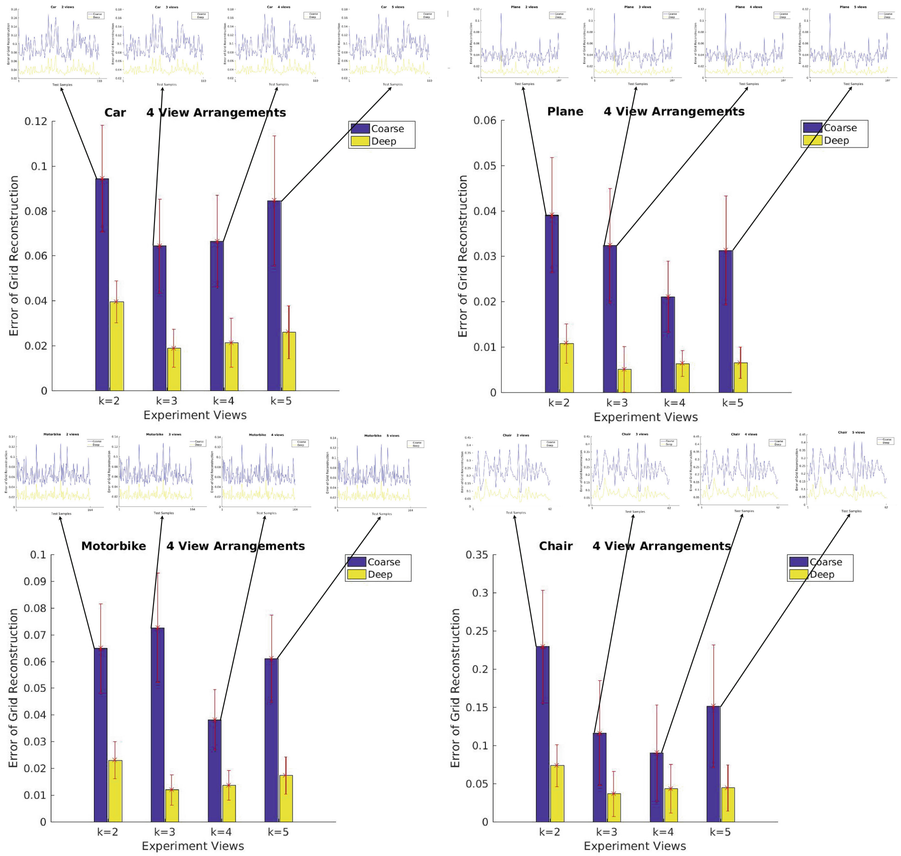

Fig. 10. Reconstruction errors for four object categories. For each object category, reconstruction errors of reconstructed shapes from both the 2 stages including coarse stage (in blue curves/bars) and the deep stage (in yellow curves/bars) are shown. The errors of each test sample of 4 view arrangements are shown on the top 4 sub-figures, the average error for each view arrangement is shown in the bar figures. Variance of errors is plot in red lines (Color figure online)

where $I(.)$ is an indicator function and $t \in [0, 1]$ is a voxelization threshold. Higher IoU values indicates better reconstruction. In the test, the value of threshold is set to $t = 0.5$ for cars, planes and motorbikes. $t = 0.3$ is chosen for chairs.

Figure 10 visualizes the reconstruction errors of all tests. Error of each test sample, the average error and variance of error for each object ategory and each view arrangement are all shown. Figure 11 visualizes the voxel IoU of all tests. Voxel IoU of each test sample, the average voxel IoU and variance of voxel IoU for each object category and each view arrangement are all shown.

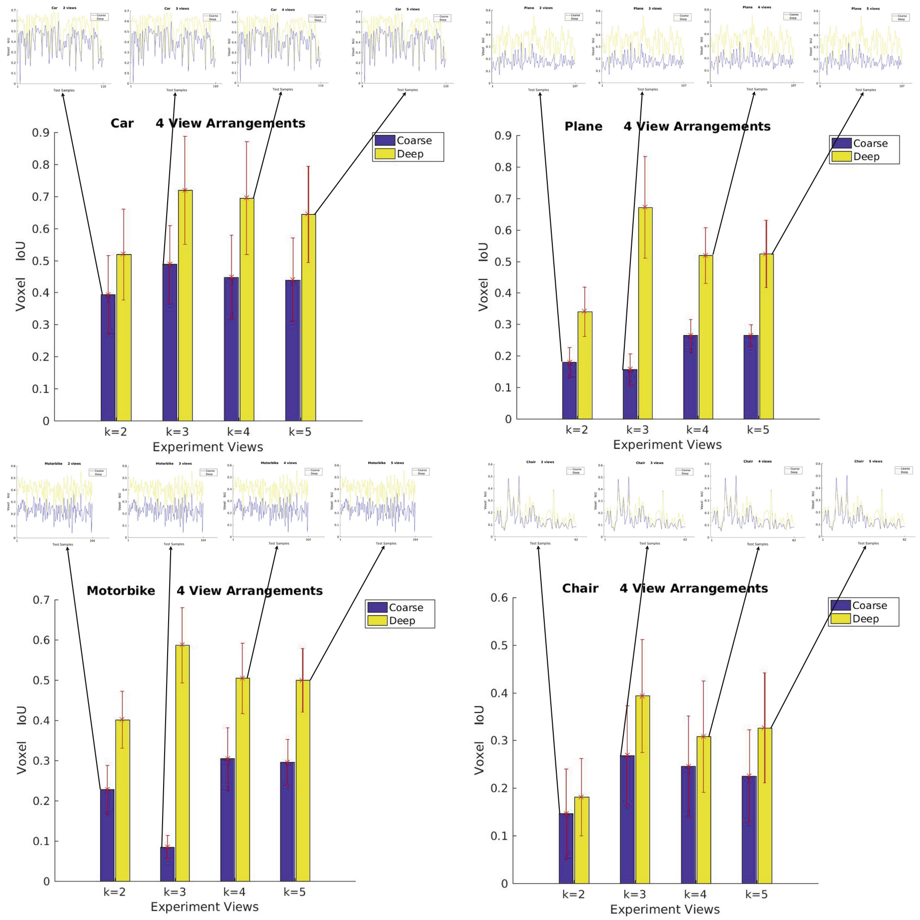

Fig. 11. Voxel IoU for four object categories. For each object category, voxel IoU of reconstructed shapes from both the 2 stages including coarse stage (in blue curves/bars) and the deep stage (in yellow curves/bars) are shown. Voxel IoU of each test sample of 4 view arrangements are shown on the top 4 sub-figures, the average Voxel IoU for each view arrangement is shown in the bar figures. Variance of each voxel IoU is plot in red lines (Color figure online)

6 Conclusion and Future Work

Our 3D reconstruction pipeline using 3D CNNs networks has been trained end to end. The input of the pipeline are a small number of silhouettes with corresponding camera parameter matrix and the output is the reconstruction of category-specific 3D shapes. Our approach proves its efficiency in tackling the complexity of the shapes considered where the object categories contain instances with large non-linear shape variations. The proposed pipeline works in two stages including coarse shape reconstruction stage and deep shape reconstruction stage.

Our 3D reconstruction pipeline works from a low number of silhouettes given as inputs, to reconstruct good-quality 3D category-specific shapes. This reconstruction pipeline is independent on pixel values, feature matches and other forms of data. It provides a potential solution to opening challenges in current 3D reconstruction field including reconstruction failures for objects containing textureless, transparent parts and low-quality reconstruction due to insufficient dense feature correspondences. Furthermore, this pipeline is practical to use because it depends on a low number of silhouettes inputs as opposed to providing a large number of images/silhouettes from multiple views as needed for reconstruction.

However, some limitations of the reconstruction pipeline exists. First, the proposed pipeline is not capable of reconstructing good shapes from two silhouettes or a single silhouette. Second, the proposed pipeline is demonstrated to produce reconstruction for a range of selected view arrangements: the selected silhouettes were taken from evenly spaced locations on a circle around the object. Third, the reconstruction pipeline relies on camera parameters matrix to be available. Finally, quality of the final reconstruction is not very good for chairs that have legs broken. Moreover, our current shape resolution is only $50 \times 50 \times 50$ (inputs, and outputs) and this is an open computational challenge to address high resolution volumetric shape in 3D reconstruction.

Therefore, our future work will improve the reconstruction quality of our pipeline in four aspects. First, we will explore to produce good reconstruction from two silhouettes or a single silhouette. Second, more work is planed to produce good reconstruction from random views. Third, in order to let our pipeline to work more automatically, we will improve our pipeline to reduce the input to only silhouettes without the knowledge of their camera parameter matrix. Finally, the improvement on both higher resolution of final reconstruction and less failure such as broken legs will also be made.

Acknowledgments. This work has been supported by a scholarship from Trinity College Dublin (Ireland), and partially supported by EU FP7-PEOPLE-2013-IAPP GRAISearch grant (612334).

References

1. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F.: ShapeNet: an information-rich 3D model repository. Technical report [cs.GR], Stanford University – Princeton University – Toyota Technological Institute at Chicago (2015). arXiv:1512.03012
2. Farabet, C., Couprie, C., Najman, L., LeCun, Y.: Learning hierarchical features for scene labeling. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1915–1929 (2013)
3. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)

4. Hariharan, B., Arbeláez, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 447–456 (2015)
5. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
6. Ganin, Y., Lempitsky, V.: N^4-fields: neural network nearest neighbor fields for image transforms. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9004, pp. 536–551. Springer, Heidelberg (2015). doi:10.1007/978-3-319-16808-1_36
7. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: Large-scale video classification with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1725–1732 (2014)
8. Dosovitskiy, A., Fischery, P., Ilg, E., Hazirbas, C., Golkov, V., van der Smagt, P., Cremers, D., Brox, T., et al.: Flownet: learning optical flow with convolutional networks. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 2758–2766. IEEE (2015)
9. Eigen, D., Puhrsch, C., Fergus, R.: Depth map prediction from a single image using a multi-scale deep network. In: Advances in Neural Information Processing Systems, pp. 2366–2374 (2014)
10. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
11. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3D convolutional networks. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 4489–4497. IEEE (2015)
12. Matan, O., Burges, C.J., LeCun, Y., Denker, J.S.: Multi-digit recognition using a space displacement neural network. In: NIPS, pp. 488–495 (1991)
13. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Comput. **1**(4), 541–551 (1989)
14. Wolf, R., Platt, J.C.: Postal address block location using a convolutional locator network. In: Advances in Neural Information Processing Systems, p. 745 (1994)
15. Ning, F., Delhomme, D., LeCun, Y., Piano, F., Bottou, L., Barbano, P.E.: Toward automatic phenotyping of developing embryos from videos. IEEE Trans. Image Process. **14**(9), 1360–1371 (2005)
16. Dosovitskiy, A., Tobias Springenberg, J., Brox, T.: Learning to generate chairs with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1538–1546 (2015)
17. Sharma, A., Grau, O., Fritz, M.: VConv-DAE: deep volumetric shape learning without object labels. arXiv preprint (2016). arXiv:1604.03755
18. Wang, X., Fouhey, D., Gupta, A.: Designing deep networks for surface normal estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 539–547 (2015)
19. Tulsiani, S., Malik, J.: Viewpoints and keypoints. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1510–1519. IEEE (2015)
20. Luo, W., Schwing, A.G., Urtasun, R.: Efficient deep learning for stereo matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5695–5703 (2016)

21. Tatarchenko, M., Dosovitskiy, A., Brox, T.: Multi-view 3D models from single images with a convolutional network. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 322–337. Springer, Heidelberg (2016). doi:10.1007/978-3-319-46478-7_20
22. Choy, C.B., Xu, D., Gwak, J., Chen, K., Savarese, S.: 3D–R2N2: a unified approach for single and multi-view 3D object reconstruction. arXiv preprint (2016). arXiv:1604.00449
23. Yumer, M.E., Mitra, N.J.: Learning semantic deformation flows with 3D convolutional networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 294–311. Springer, Heidelberg (2016). doi:10.1007/978-3-319-46466-4_18
24. Yan, X., Yang, J., Yumer, E., Guo, Y., Lee, H.: Learning volumetric 3D object reconstruction from single-view with projective transformations. In: Neural Information Processing Systems (NIPS 2016) (2016)
25. Laurentini, A.: The visual hull concept for silhouette-based image understanding. IEEE Trans. Pattern Anal. Mach. Intell. **16**(2), 150–162 (1994)
26. Kim, D., Ruttle, J., Dahyot, R.: Bayesian 3D shape from silhouettes. Digit. Signal Proc. **23**(6), 1844–1855 (2013)
27. Su, H., Qi, C.R., Li, Y., Guibas, L.J.: Render for CNN: viewpoint estimation in images using CNNs trained with rendered 3D model views. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2686–2694 (2015)
28. Maturana, D., Scherer, S.: Voxnet: a 3D convolutional neural network for real-time object recognition. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 922–928. IEEE (2015)
29. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3D shapenets: a deep representation for volumetric shapes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1912–1920 (2015)