What Do Single-view Reconstruction Networks Learn?

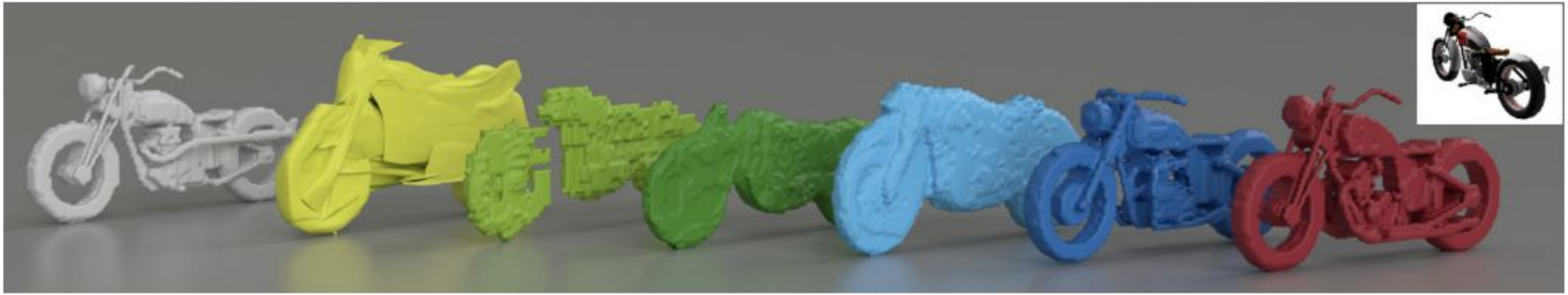
Maxim Tatarchenko, Stephan R. Richter, Rene Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox

Group 6 : Kshitij Karnawat, Arshad Shaik, Rishikesh Jadhav, Andrew Zhong, Sandeep Thalapanane, Guru Appiya, Vineet Singh

Scribe - Vineet Singh

Problem statement

1. The paper explores the techniques learnt by 3D reconstruction networks.
2. Analysis of the drawbacks of current techniques.
3. Deviations in the reconstruction done the networks and the role of training data in shaping the results of the network. .



Methodology:

1. Use 3 state of the art networks for analysis.
    a. AtlasNet, Octree Generating Networks (OGN), Matrayoshka Networks.
2. Designed 2 simple recognition baselines for comparison.
    a. Clustering - of training shapes in conjunction with image classifier
    b. Retrieval - learns to embed images and shapes in a joint space
3. Oracle Nearest Neighbor : For each of the test 3D shapes, find the closest shape from the training set in terms of IoU. Gives upper Bound of recognition estimation.
4. Used mean Intersection over Union (IoU) method for results comparison.

Mean IoU comparison results

1. The recognition baseline performed better than the models.
2. All the models performed worse w.r.t. Oracle NN.
3. Shows that IoU not sufficient, need other methods for Analysis.

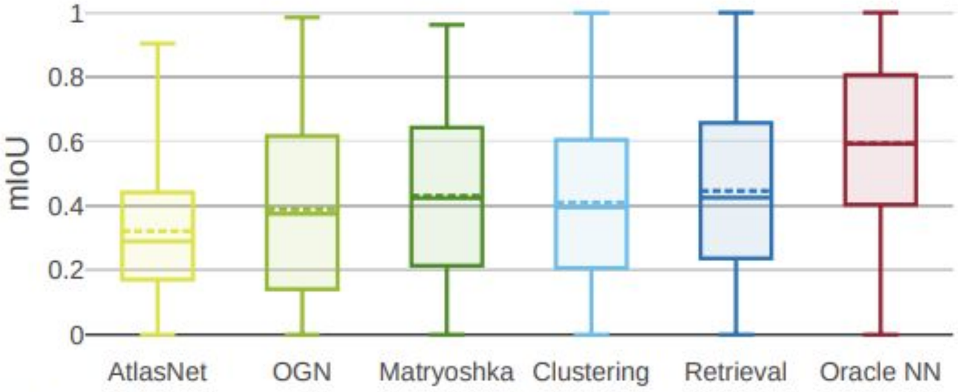


Figure 2: Comparison by mean IoU over the dataset. The box corresponds to the second and third quartile. The solid line in the box depicts the median; the dashed line the mean. Whiskers mark the minimum and maximum values, respectively.

More Rigorous Analysis

1. **Per class analysis :** showed that there is no correlation between the number of samples in a class and its mean IoU score, and its coefficient is close to zero. So, the process of selecting only 13 out of 55 classes has no technical justification.

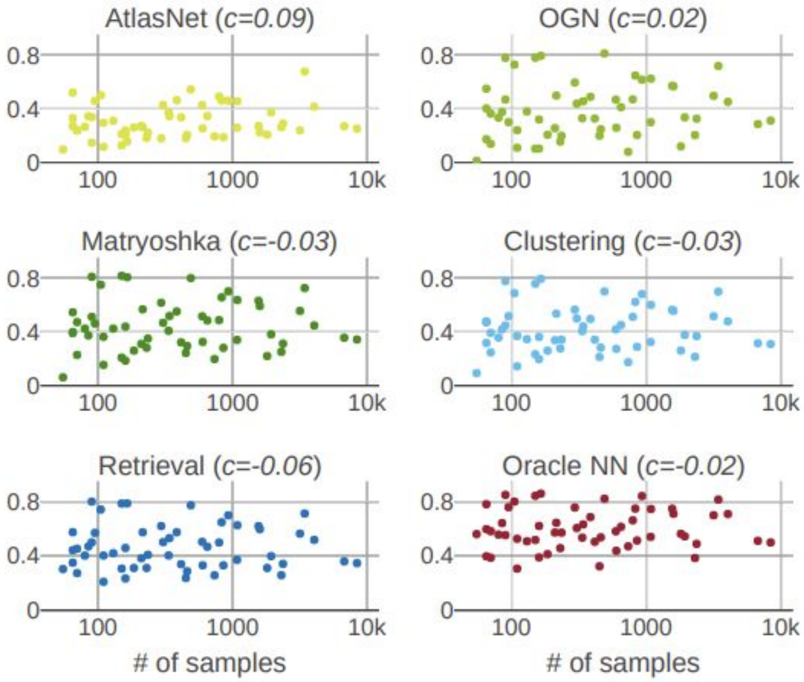


Figure 4: mIoU versus number of training samples per class. We find no correlation between the number of samples within a class and the mIoU score for this class. The correlation coefficient c is close to zero for all methods.

Per class mIoU Comparison Results

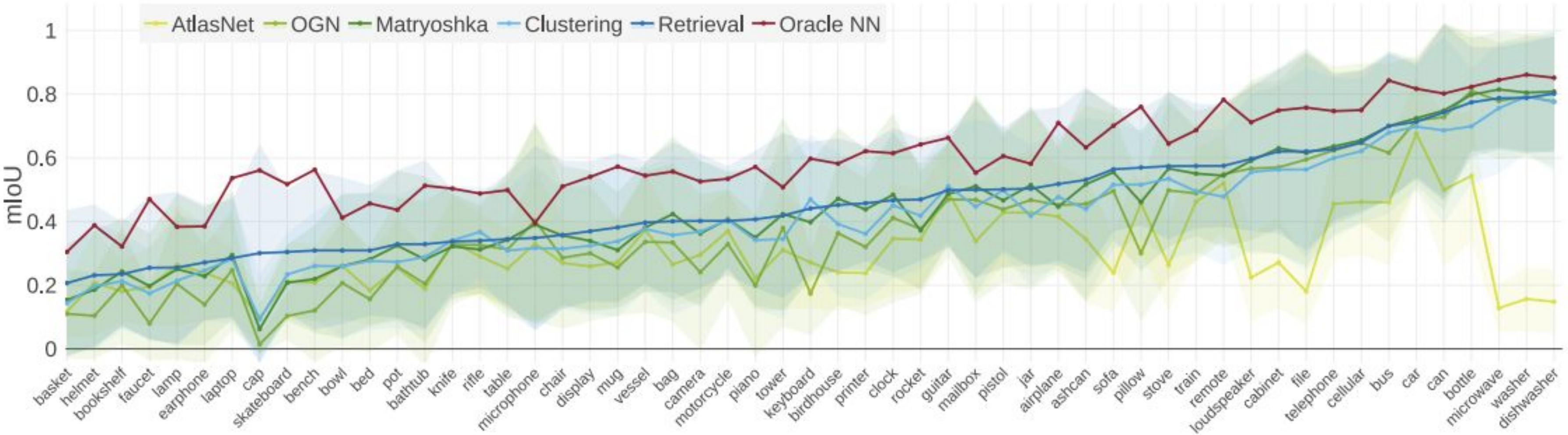
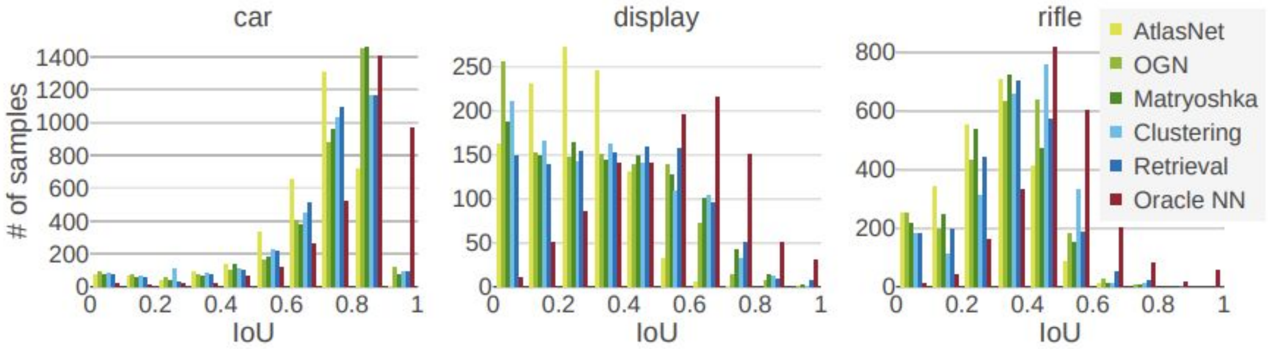


Figure 3: Comparison by mIoU per class. Overall, the methods exhibit consistent relative performance across different classes. The retrieval baseline produces the best reconstructions for the majority of classes. The variance is high for all classes and methods.

Statistical evaluation

1. **Statistical evaluation:** visualized the histograms of IoU scores for individual object classes. The results pointed out that within classes, distributions of decoder-based methods and recognitions baselines were very similar.
2. These results show that networks rely heavily on recognition Instead of reconstruction.
3. The methods used currently are not sufficient and need different evaluation metrics.



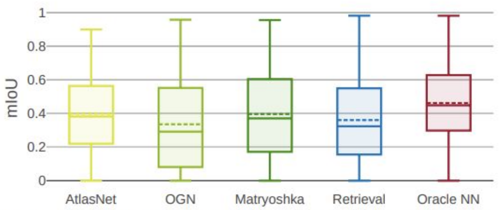


Figure 7: Mean IoU in viewer-centered mode. The retrieval baseline does not perform as well in this mode.

Analysis suggestions

1. **Choice of Coordinate system:** Changing the object-centered view system to a viewer-centered coordinate system as it leads to better generalization.
2. **F-score:** Use F-score instead of IoU or CD. It explicitly evaluates the distance between object surfaces and is defined as the harmonic mean between precision and recall and its strictness is controlled by varying the distance threshold d.
3. **Training Data:** The training data to be changed to viewer-centered.
4. Contamination of testing and validation data.

$$F(d) = \frac{2P(d)R(d)}{P(d) + R(d)}.$$

Quantitative results

The analysis results with above suggestions show better performance quantitatively.

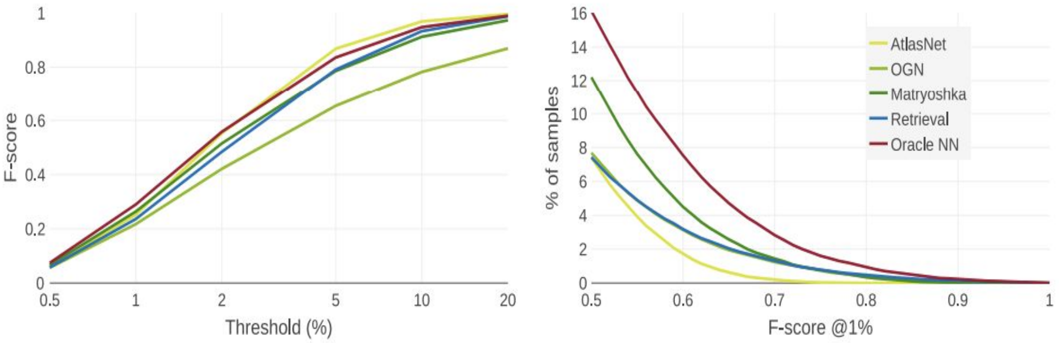


Figure 10: F-score statistics in viewer-centered mode. Left: F-score for varying distance thresholds. Right: percentage of reconstructions with F-score above a value specified on the horizontal axis, with a distance threshold $d = 1\%$.

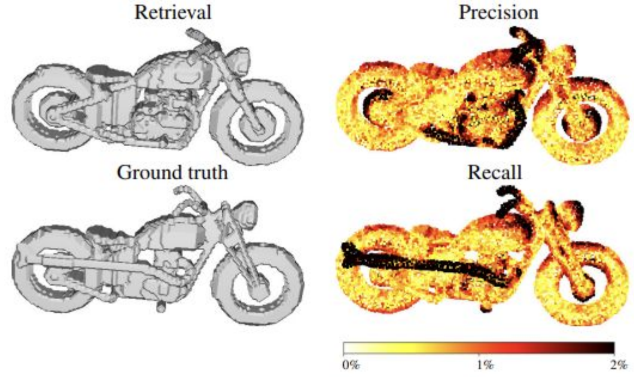


Figure 12: Visualizing precision and recall provides detailed information about which object parts were reconstructed correctly. Colors encode the normalized distance between shapes (as used for the distance threshold).

Strengths

- Scrutinizing evaluation metrics
- Pointing out Datasets Drawbacks
- Highlighting that most networks perform recognition

Weakness

- Comparing with less no of state of the art methods
- Not providing proper solutions for the questions raised

Backup Slides

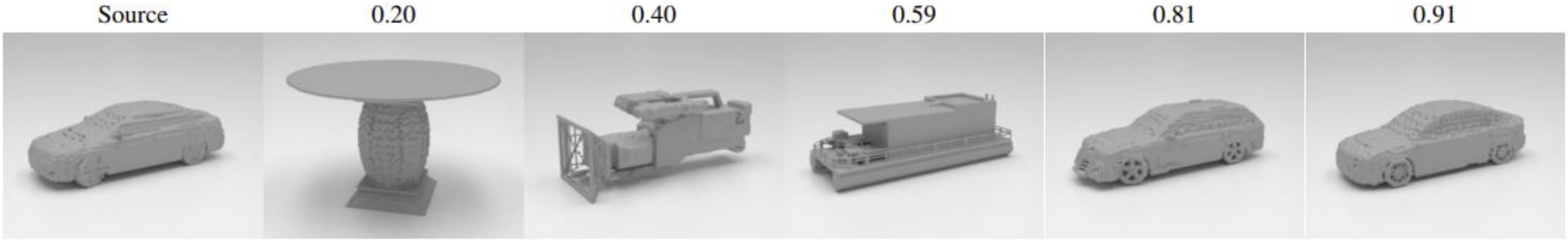


Figure 8: IoU between a source shape and various target shapes. Low to mid-range IoU values are a poor indicator of shape similarity.

Figure 9: The Chamfer distance is sensitive to outliers. Compared to the source, both target shapes exhibit non-matching parts that are equally wrong. While the F@1% is 0.56 for both shapes, the Chamfer distance differs significantly.

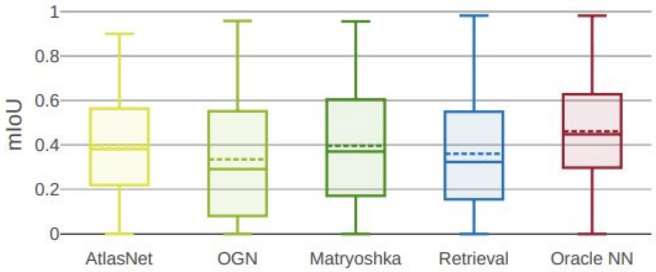


Figure 7: Mean IoU in viewer-centered mode. The retrieval baseline does not perform as well in this mode.