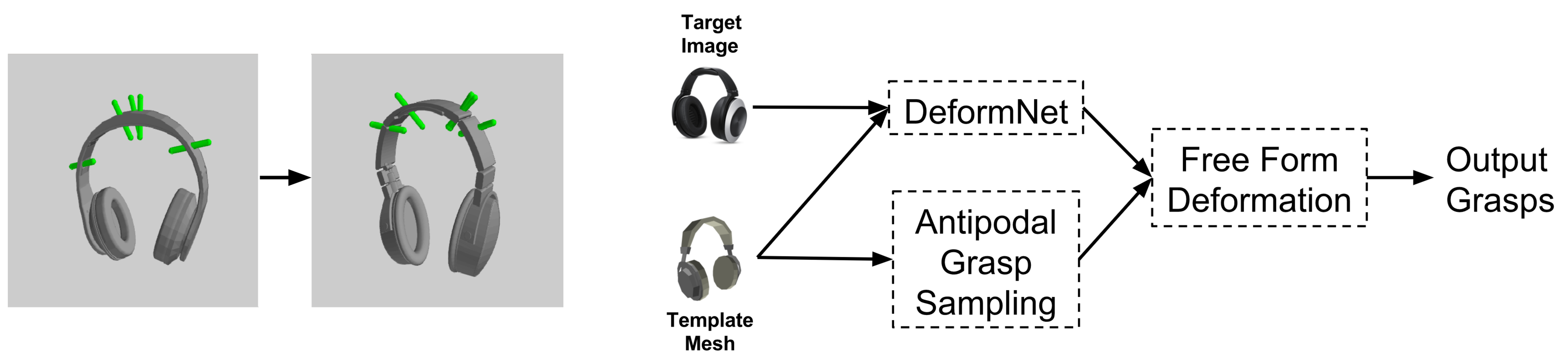
Towards Grasp Transfer Using Shape Deformation

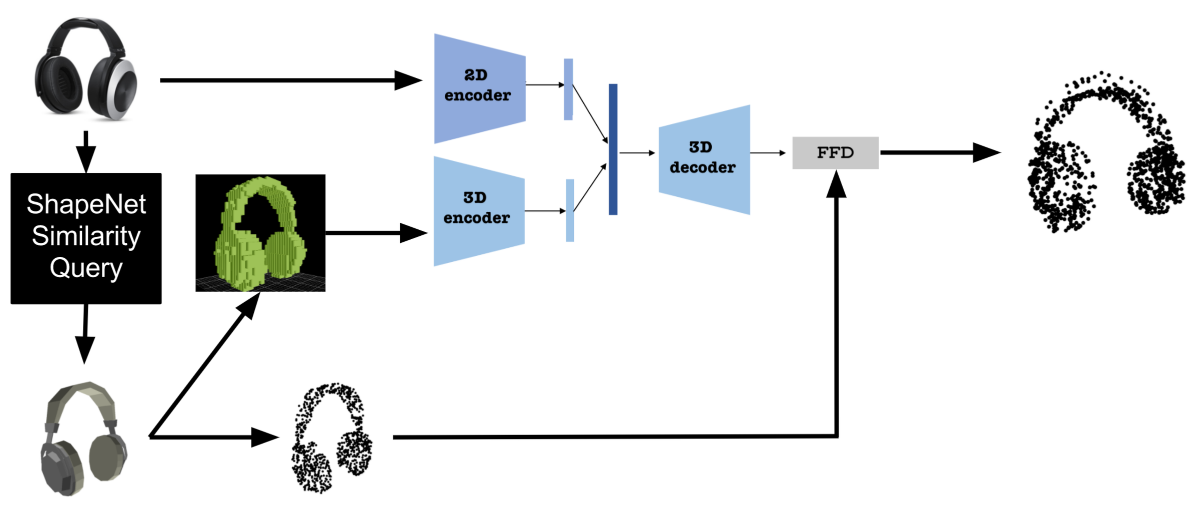
Andrey Kurenkov*, Viraj Mehta*, Jingwei Ji, Animesh Garg, Silvio Savarese

Overview

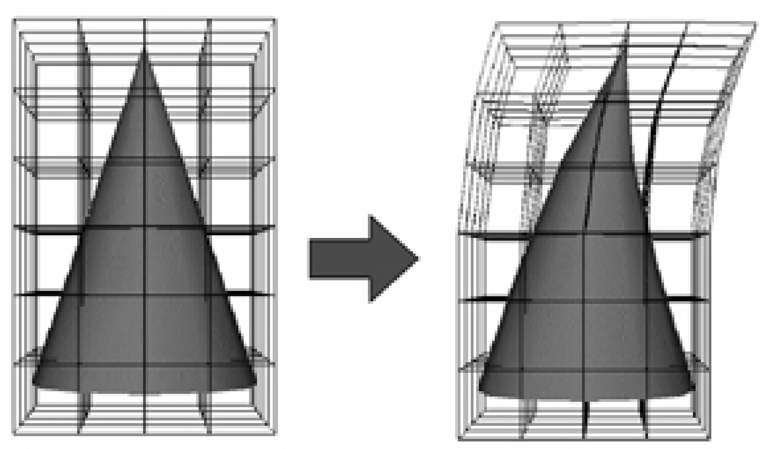
- **Problem: Robotic Grasp Planning**. Robust grasping of novel objects still unsolved.
- **Prior methods**: Deep Learning is robust but limited and inefficient, traditional grasp transfer can extend to semantic grasping and requires less data but is less usable.
- **Approach: Deep Learning Enabled Grasp Transfer**. Deform template object and its grasps towards target object, given an image of the target object and the mesh of the template.



Model



- **Two CNN encoders:** a 2D CNN for the target image and a 3D CNN for the template voxel.
- Encoder outputs are combined by stacking the final fully connected layer activation to the last 3D CNN activation along the channel.
- **3D Deconvolution Decoder:** The final output of the decoder is a vector field $V = \{v_i\}_{i=1,\dots,N^3}, v \in \mathbb{R}^3$ - offsets for the N^3 control points in the Free-Form Deformation Layer
- **FFD**[2]: offset of control points in grid controls deformation. **Deform object along with its grasps through a point cloud representation.**



Training

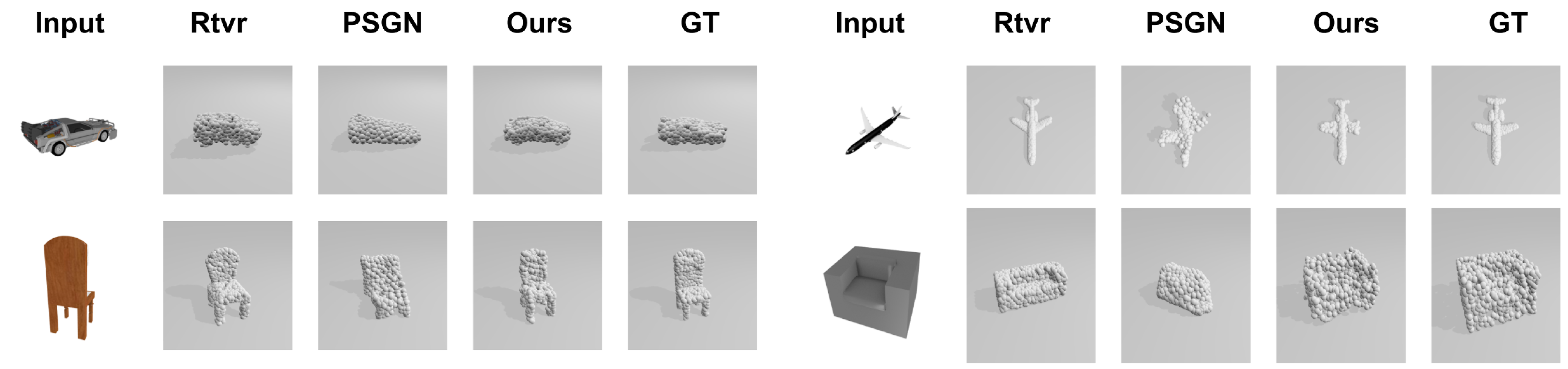
- **Dataset: ShapeNet** [3]. We select 9 categories for a total of 24,324 shape models, and render synthetic train and test data. Training/testing split is 80%/20%.

- **Loss: Chamfer distance** - for each point in one point set, finds distance to nearest point in the other set. Formally, CD is $d_{CD}(S_1, S_2) =$

$$\sum_{p_1 \in S_1} \min_{p_2 \in S_2} \|p_1 - p_2\|_2^2 + \sum_{p_2 \in S_2} \min_{p_1 \in S_1} \|p_1 - p_2\|_2^2$$

- **Regularization**: L1 loss over all point cloud offsets to force the network to deform the template as little as possible, and L2 loss over the difference between neighboring control point offsets to promote smooth deformation.
- **Implementation:** TensorFlow, Adam optimizer, initial learning rate of 5e-4 and goes down to 5e-5 after 20k iterations, batch size of 16 for training, leaky ReLU, point clouds with 1024 points, N=4 as the number of control points in each dimension that points are computed with . We use $\lambda = 0.05$ for regularization on control point offsets.

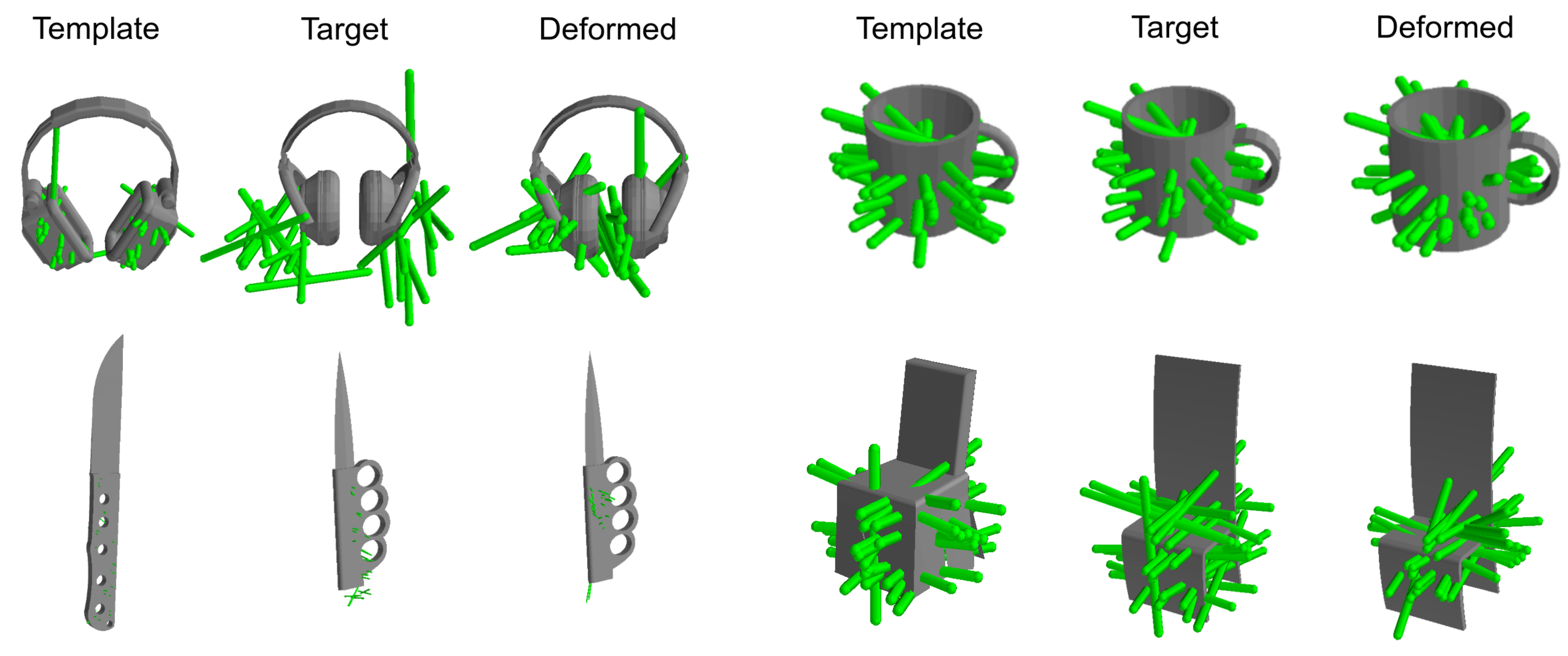
DeformNet Evaluation



Visual results - input, retrieval, point-set generation network, DeformNet, and ground truth.

- DeformNet quantitatively **matches or outperforms** SOA benchmarks by significant margins. Qualitatively, it benefits from strong template retrieval and improves upon it with deformation.

Grasp Transfer Results



Transfer of grasps (green lines) without and with deformation, respectively.

- Despite flawed performance for graspable categories, deformation improves grasp transfer according to Ferrari-Canny metrics due to more grasps being on object and near its center[1].

	Template	**Deformed**	GT
Top 50	0.02	**0.03**	0.05
Top 5	0.12	**0.15**	0.22
Top 1	0.18	**0.21**	0.27

Discussion

- The application of a deep learning transfer-through-deformation approach to grasp transfer shows promising results.
- Further work is needed for accurate deformation and usability with real world images.

References

[1] Jonathan Weisz and Peter K Allen. Pose error robust grasping from contact wrench space metrics. In *Robotics and Automation (ICRA), 2012 IEEE Int'l Conf. on*, pages 557–562. IEEE, 2012.

[2] Thomas W Sederberg and Scott R Parry. Free-form deformation of solid geometric models. *ACM SIGGRAPH computer graphics*, 20(4):151–160, 1986.

[3] Angel X. Chang and et al. ShapeNet: An Information-Rich 3D Model Repository. Technical report, 2015.