Weakly supervised 3D Reconstruction with Adversarial Constraint

JunYoung Gwak*†, Christopher B Choy*†, Manmohan Chandraker‡, Animesh Garg†, Silvio Savarese†
*Indicates equal contribution †Stanford University ‡NEC Laboratories America, Inc., UCSD

1. Motivation

Goal: Learning 3D reconstruction from weak supervision of 2D masks

Previous works:

- **Full 3D supervision**[1][2][3]: 3D model is a very **expensive** label for practical use such as real image reconstruction.
- **2D mask supervision**[4]: Limited by visual hull. No concavity, symmetry, stability, etc.

Proposed method:

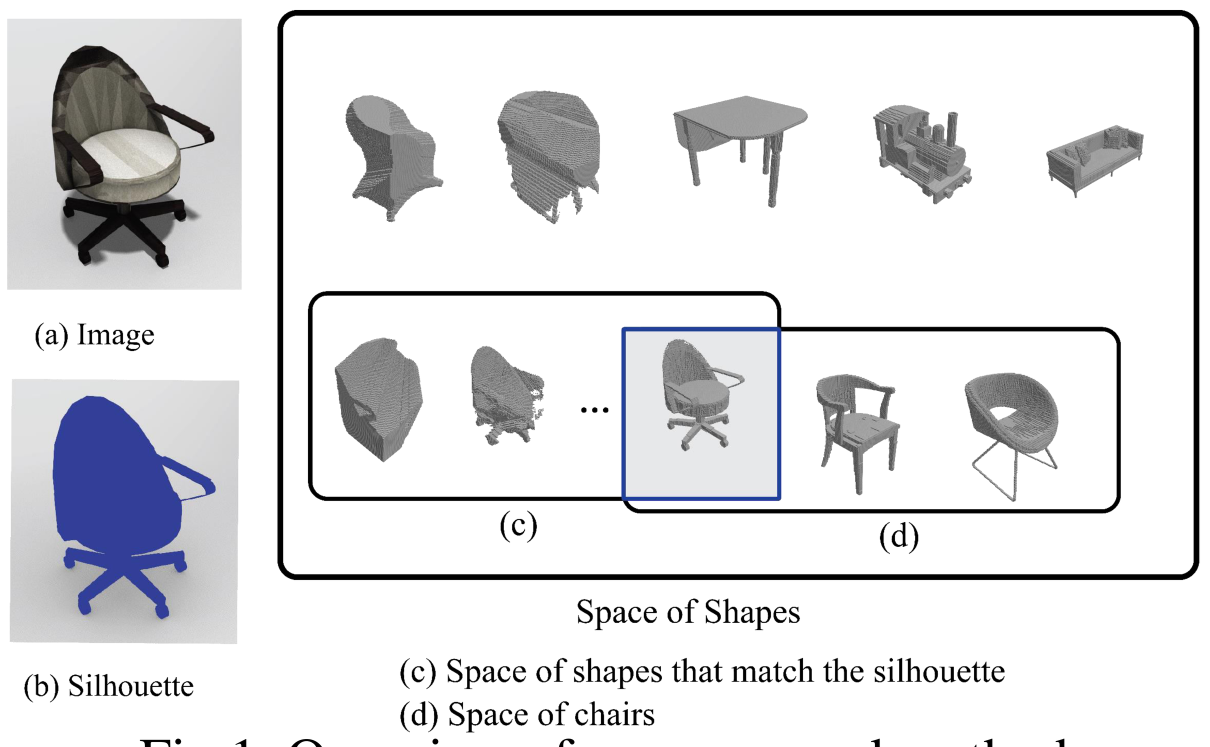

Fig 1. Overview of our proposed method

Solving constrained optimization

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad \text{ReprojectionError}(x) \\ &\text{subject to} \quad \text{Reconstruction } x \text{ to be a valid chair} \end{aligned} \tag{1}$$

- **ReprojectionError** resembles **2D mask supervision** [4] and Fig 1 (c)
- **The constraint** resembles Fig 1 (d)
- Together learns correct 3D reconstruction

2. Adversarial Constraint

1. Equation (1) can be re-written as

$$\underset{x}{\text{minimize}} \quad \text{ReprojectionError}(x) - \frac{1}{t}\log g(x) \tag{2}$$

using **log barrier method** where $g(x) = 1$ iff reconstruction x is realistic and 0 otherwise

2. Ideal **discriminator** of GAN $g^{\star}(x)$, which outputs $g^{\star}(x) = 1$ iff reconstruction x is realistic, is analogous to the penalty function $g(x)$

3. Therefore, we can train $g(x)$ as discriminator

$$\underset{g}{\text{minimize}} \quad \mathop{\mathbb{E}}_{x^{\star}\sim p} \log g(x^{\star}) + \mathop{\mathbb{E}}_{\hat{x}\sim q} \log(1 - g(\hat{x})) \tag{3}$$

3. Raytrace Pooling

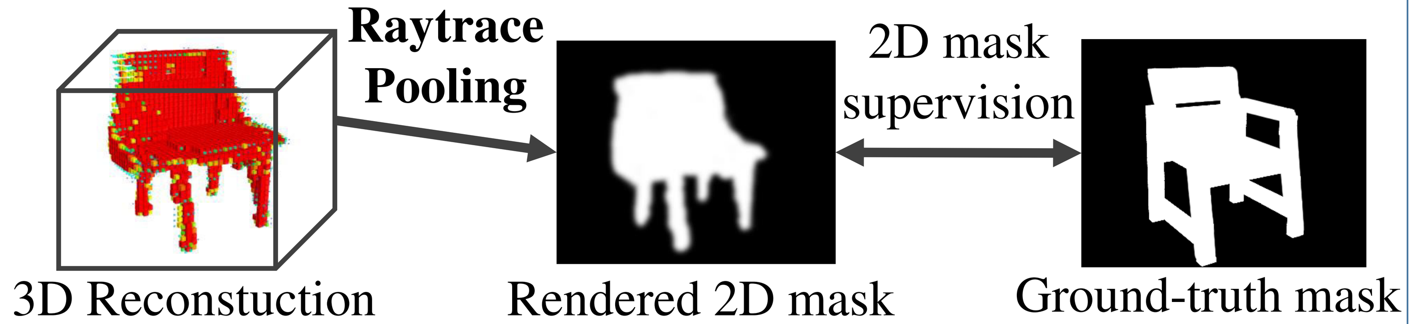

Fig 2. Overview of raytrace pooling and ReprojectionError

- Bridge the gap between the target 3D reconstruction and **2D mask supervision**
- Takes reconstruction x and camera parameter of the ground-truth mask as input
- Renders mask of x through ray-voxel hit test
- Does not suffer from sampling artifacts as compared to [4]

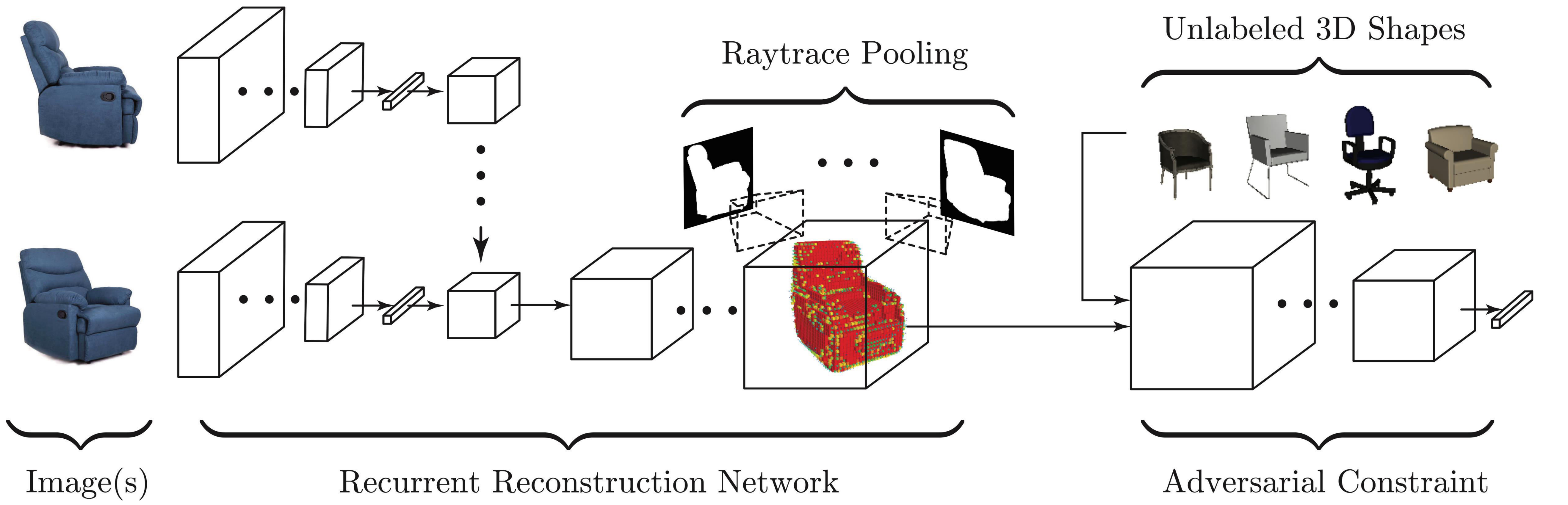

Fig 3. Overview of our proposed network architecture

4. Experiments

1. Ablation study on ShapeNet

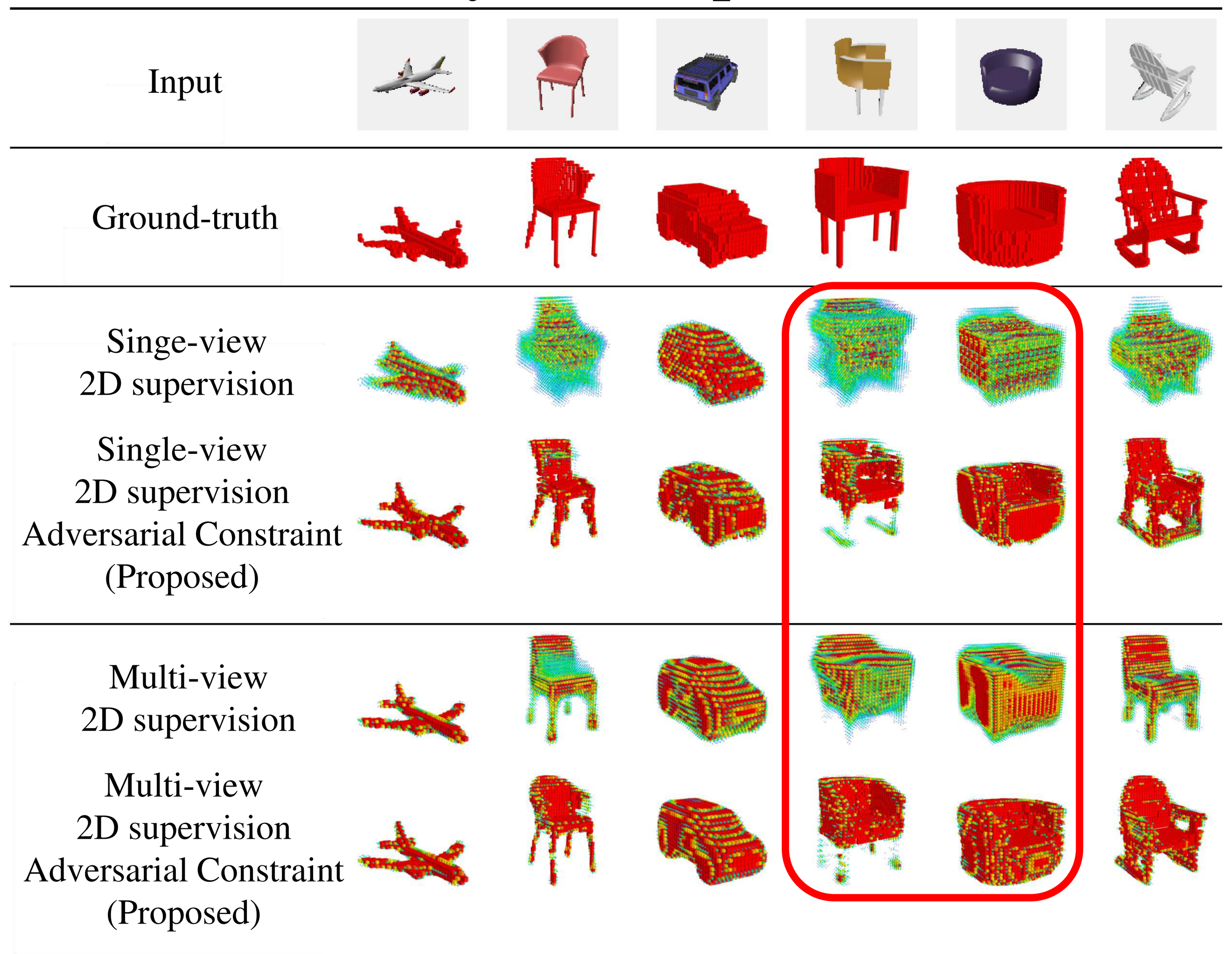

Fig 4. Qualitative results on ShapeNet

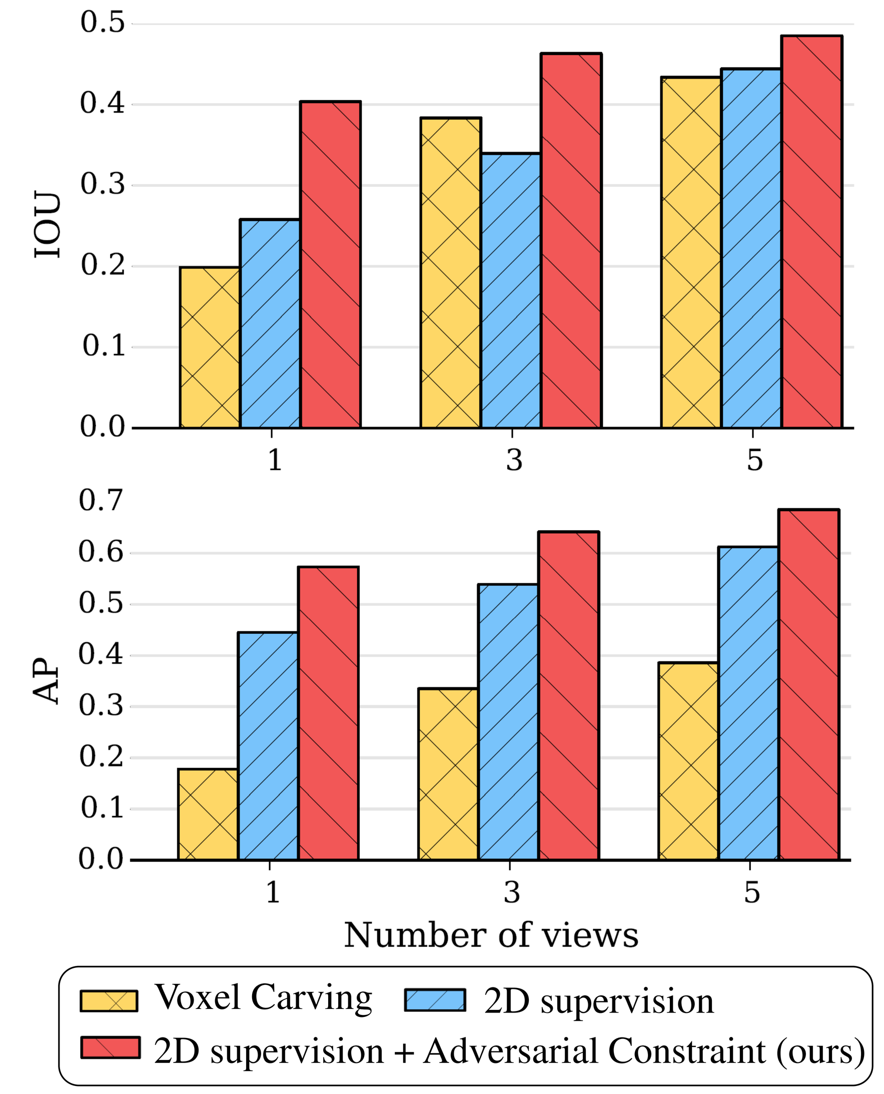

Fig 5. Quantitative results on ShapeNet

Our proposed method reconstructed a reasonable 3D shape from weak 2D supervision including **concavity**(red box in Fig 4). It is also worth nothing that the adversarial constraint gives a noticeable performance boost especially on weak single-view supervision as shown in Figure 5

2. Real image reconstruction

Fig 6. Single-view reconstruction on ObjectNet3D

Fig 7. Multi-view reconstruction on Stanford Online Product

3. Representation analysis

Fig 8. Linear interpolation of hidden variables of two images

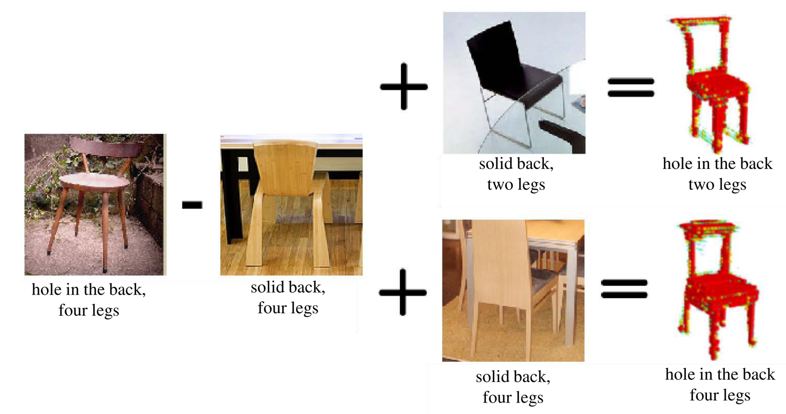

Fig 9. Semantic feature arithmetic

References

- [1] C. B. Choy, et. al. 3DR2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction. In ECCV, 2016
- [2] R. Girdhar, et. al. Learning a predictable and generative vector representation for objects. In ECCV, 2016
- [3] J. Wu, et. al. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In NIPS, 2016
- [4] X. Yan, et. al. Learning volumetric 3d object reconstruction from single-view with projective transformations. In NIPS, 2016

Acknowledgement

We acknowledge the support of Nvidia and Toyota (1186781-31-UDARO) to make this work possible.