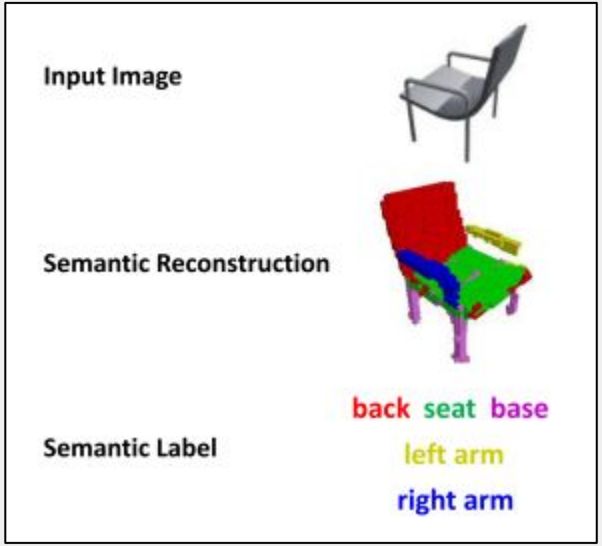
“Enhanced 3D Shape Reconstruction With Knowledge Graph of Category Concept”

Machine Learning with Knowledge Graphs 2024 - Nicolas Mogicato

Why is this paper interesting?

- **Reconstructing 3D objects from 2D** images is crucial in computer graphics/vision
- Traditional methods struggle to **capture realistic essence** of object categories
- Innovative approach incorporates **graph-based conceptual knowledge** into deep learning frameworks
- Structured **knowledge graphs** + neural networks = geometrically and conceptually consistent 3D shapes



Key Findings

01 Enhanced Performance

- Knowledge graph integration outperforms current methods
- Accurate and realistic reconstructions

02 Robust Representation

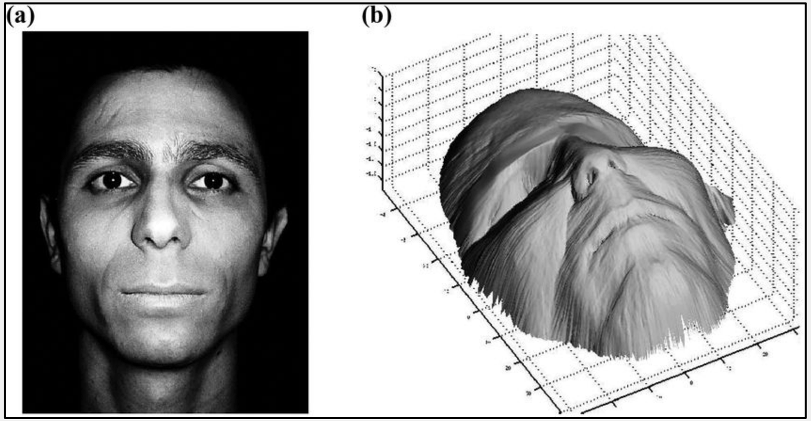
- Volume-based conceptual knowledge aids in creating detailed 3D models

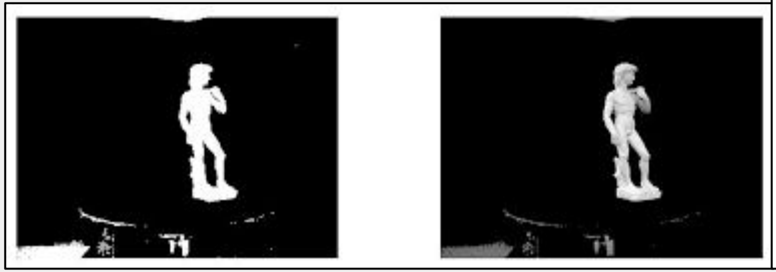
03 Versatility

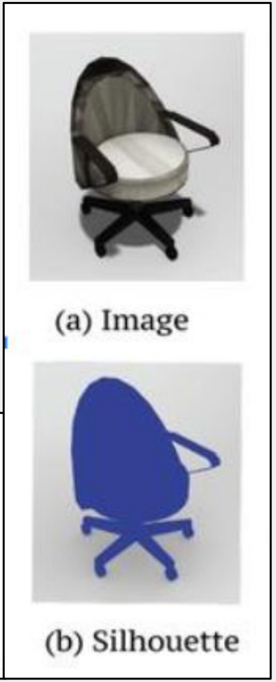
- Existing 3D reconstruction pipelines can be extended with different knowledge graphs

3D Shape Reconstruction

- Under-constrained problem
- Traditional methods reliant on silhouettes or shading often inadequate for real-world applications
- Significant progress leveraging deep learning, particularly with encoder-decoder(-refiner) frameworks, integrating view and shape info to increase accuracy







Novel Multimodal Framework

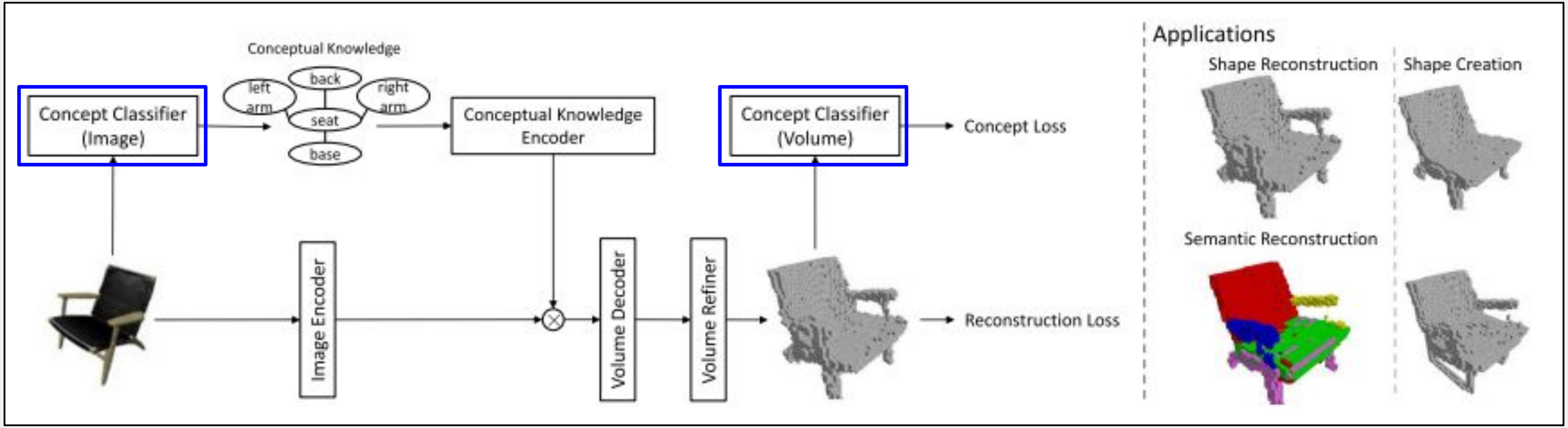
Conceptual knowledge about object parts into opens up **real-world applications**

Graph-based conceptual knowledge to enhance 3D shape reconstruction from a single RGB image

Approach rooted in intuitive **human understanding** of object structure to enhance perception

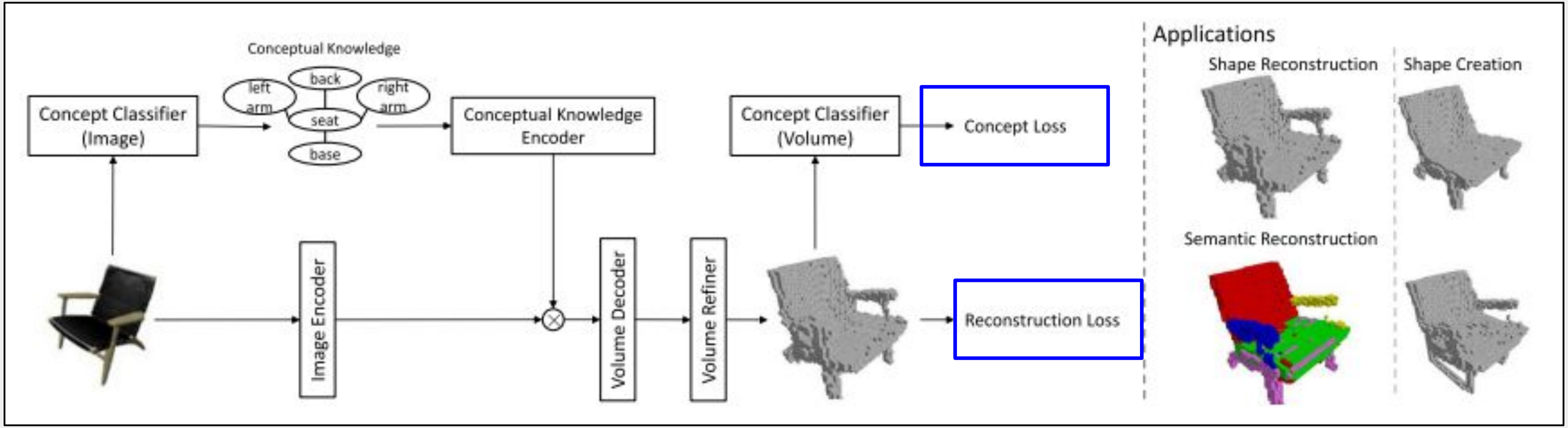
Framework Design

- Dual Concept Classifiers
- Loss Functions to verify quality
- Mirroring human cognitive abilities in machine learning contexts



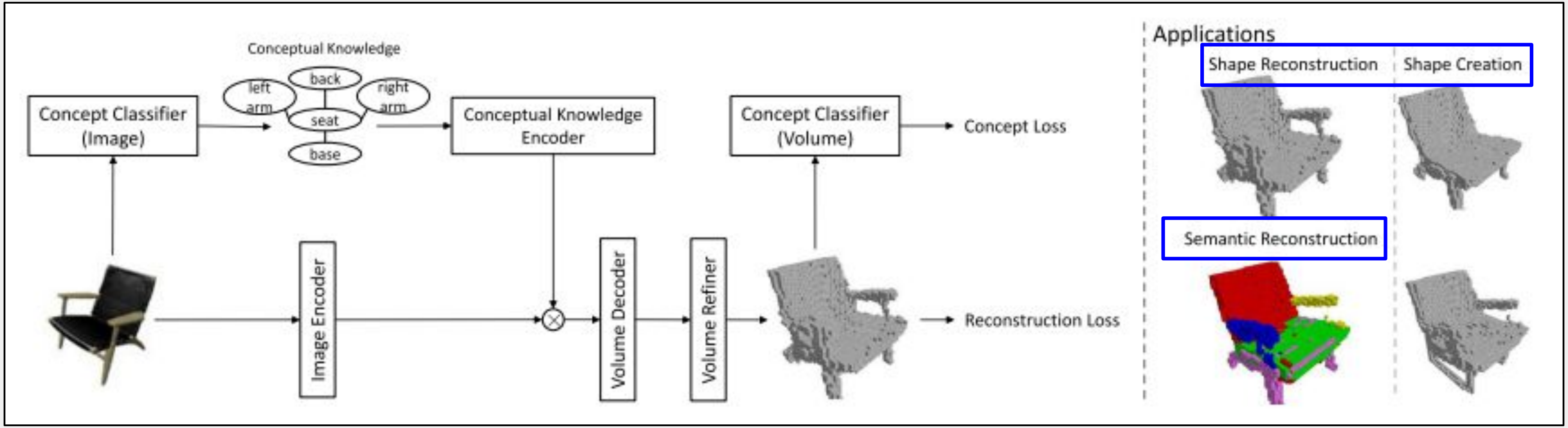
Framework Design

- Dual Concept Classifiers
- Loss Functions to verify quality
- Mirroring human cognitive abilities in machine learning contexts

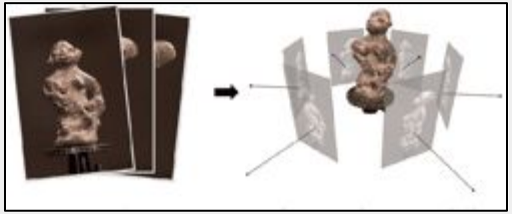


Framework Design

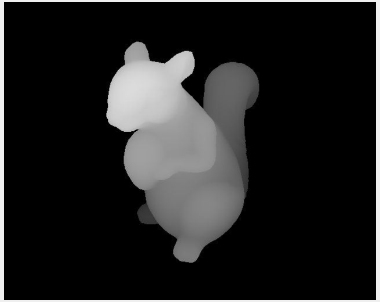
- Dual Concept Classifiers
- Loss Functions to verify quality
- Mirroring human cognitive abilities in machine learning contexts



3D shape reconstruction methodologies



01 Traditional methods

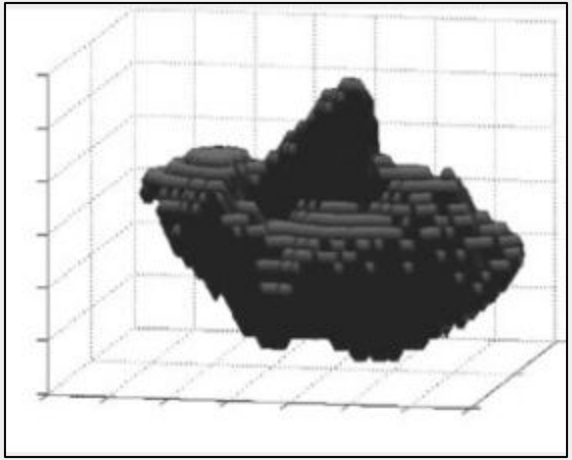


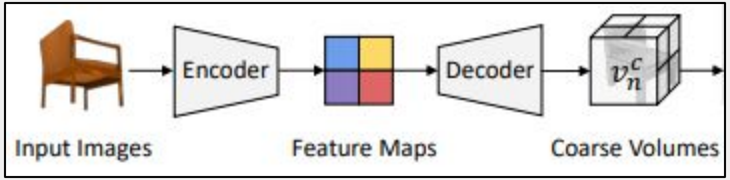
02 Deep learning-based reconstruction

03 Deep learning-based generation and completion

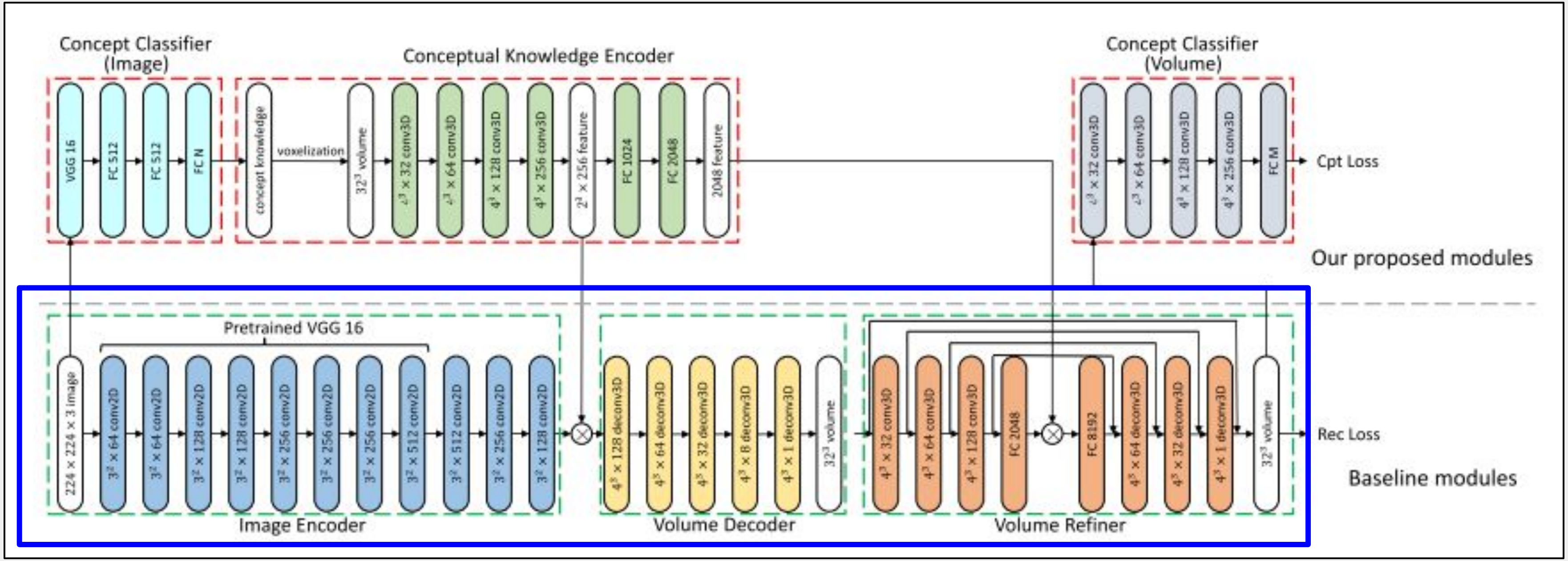
Framework - Implementation Overview

- Pix2Vox-A architecture + graph-based conceptual knowledge
- Binary occupancy volume

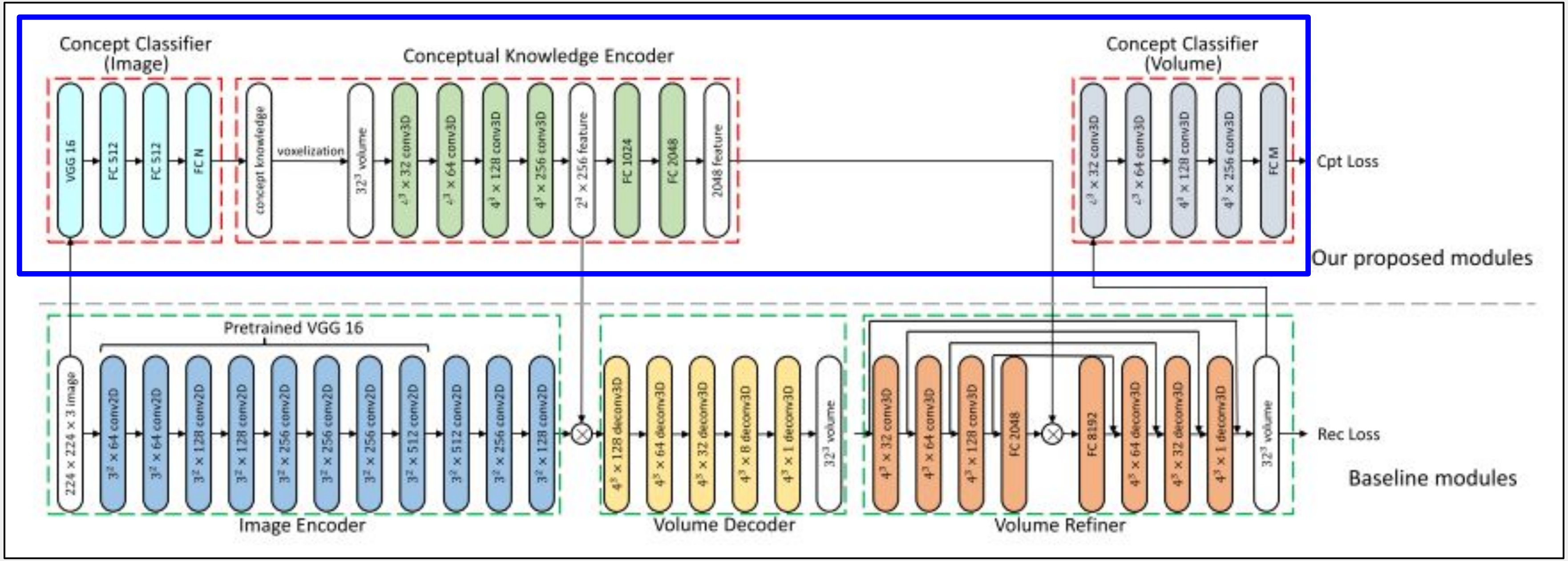




Framework - Baseline Architecture



Framework - Proposed Enhancements



Framework - Proposed Enhancements

Modules are pipeline agnostic, can be integrated into existing frameworks

High-level conceptual info avoids need of direct visual data

Advanced loss functions during training to optimize realism and accuracy

Framework - Loss Functions

Reconstruction Loss

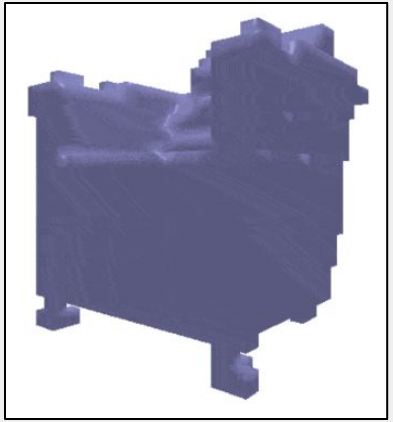
Concept Loss

Combined Loss Function

- Used to train the proposed framework, optimization
- Helps the network reconstruct the geometry accurately
- Adherence to the conceptual model of the object shapes

Reconstruction Loss

- Mean cross-entropy between predicted voxel values of the reconstructed volume and ground-truth voxel values
- Does the generated 3D shape match the actual structure of the object at voxel level ?
- Binary voxel occupation guiding the network



$$L_{rec} = \frac{1}{N}\sum_{i=1}^{N}\left(\boxed{p_i'}log(p_i) + (1 - p_i')log(1 - \boxed{p_i})\right),$$

Concept Loss

- Mean cross-entropy between predicted conceptual part labels of reconstructed shape and the ground-truth labels from input
- Does the reconstructed shape align with the conceptual understanding of the object ?
- Only used in category-specific training
- Using concept loss in category-agnostic setting could result in misleading training signals

$$L_{cpt} = \frac{1}{M}\sum_{i=1}^{M}(s_i' log(\boxed{s_i}) + (1 - \boxed{s_i'}) log(1 - s_i)),$$

Experiments

- Evaluate 3D shape reconstruction framework using synthetic and real datasets
- Study focused on 13 major categories with more than 40,000 models

Experiments - Evaluation Metric

- Intersection-over-Union (IoU) to assess the quality of the reconstruction
- Higher IoU values indicate better reconstruction quality

$$\mathrm{IoU} = \frac{\sum_{i,j,k} I(p_{(i,j,k)} > t) I(p'_{i,j,k})}{\sum_{i,j,k} I[I(p_{(i,j,k)} > t) + I(p'_{i,j,k})]},$$

Experiment - Synthetic Datasets

Table 3. Comparison of Single-View 3D Object Reconstruction on ShapeNet with Volume Size 64^3 Using IoU (in %) Evaluation Metric

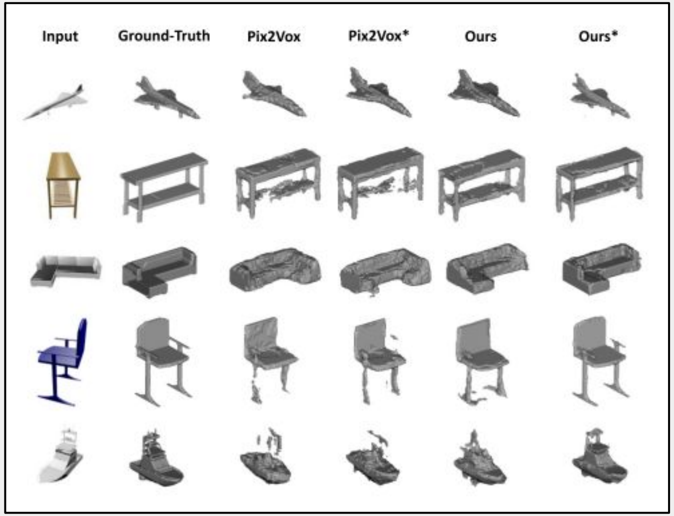
	airplane	bench	cabinet	car	chair	display	lamp	speaker	rifle	sofa	table	telephone	vessel	all
3DensiNet [56]	54.6	38.5	72.1	79.7	45.5	43.6	33.3	67.3	47.1	64.5	49.6	75.3	50.8	55.5
Voxel-Tube [44]	60.2	46.3	74.4	81.6	51.3	**49.6**	36.9	68.2	50.9	67.4	52.9	75.4	53.5	59.1
Pix2Vox [67]	61.4	46.4	72.9	80.6	50.6	47.3	**37.9**	66.8	51.6	67.2	52.2	74.3	53.7	59.7
Ours	**63.6**	**48.5**	**75.8**	**82.1**	**52.9**	49.3	36.7	**68.5**	**52.5**	**68.9**	**54.0**	**77.1**	**54.8**	**61.4**
Pix2Vox* [67]	63.2	48.4	73.9	81.3	50.8	48.0	38.8	67.2	54.2	67.8	51.9	76.8	54.9	60.4
Ours*	**68.1**	**53.2**	**77.1**	**83.1**	**53.6**	**50.8**	**38.9**	**69.8**	**54.7**	**69.7**	**54.2**	**79.4**	**56.7**	**62.9**

Experiment - Real-World Datasets

	aeroplane	boat	car	chair	diningtable	sofa	tvmonitor	all
3D-R2N2 [10]	54.4	56.0	69.9	28.0	-	33.2	57.4	-
DCT [22]	55.5	52.3	63.5	25.0	-	46.2	54.9	-
3DensiNet [56]	-	32.6	60.7	25.9	-	57.4	60.6	-
Pix2Vox [67]	60.4	67.3	79.4	41.1	39.8	66.6	54.9	58.5
Ours	**65.7**	**73.6**	**81.0**	**54.8**	**45.7**	**75.9**	**63.8**	**65.8**
Pix2Vox* [67]	71.2	76.3	82.3	57.8	46.3	80.6	64.7	68.5
Ours*	**75.6**	**84.7**	**85.1**	**64.0**	**46.5**	**84.3**	**73.3**	**73.4**

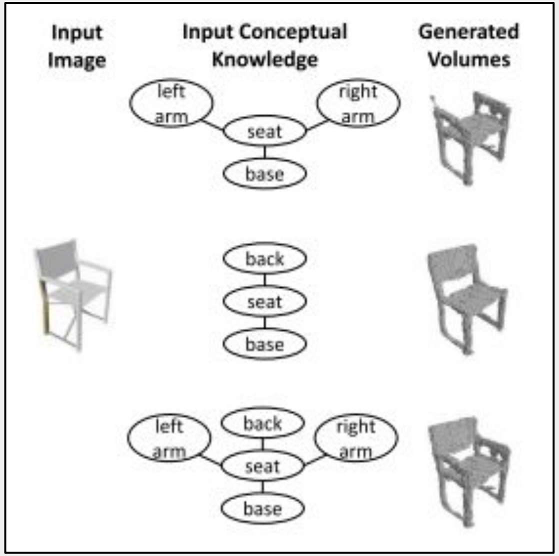
Experiments - Visual Improvements

- Incorporation of conceptual knowledge enhances accuracy and detail of 3D shape reconstructions, outperforming existing state-of-the-art models

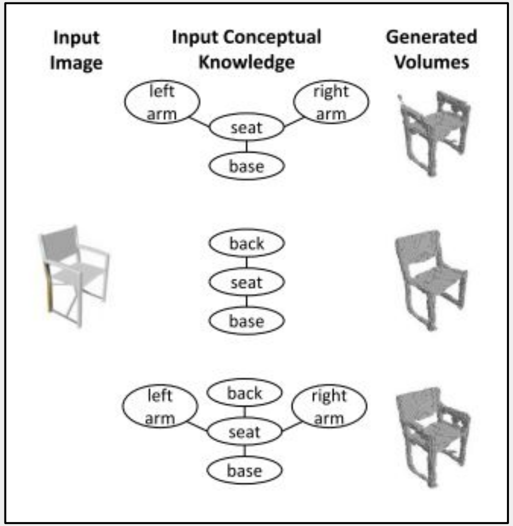


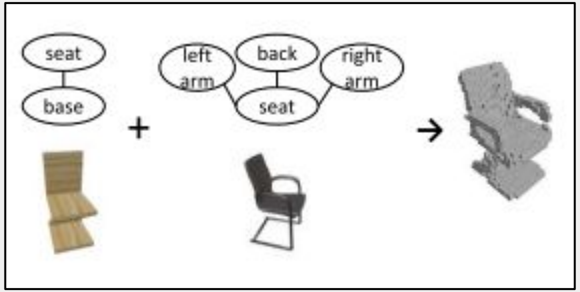
Experiments - Conceptual Knowledge Framework

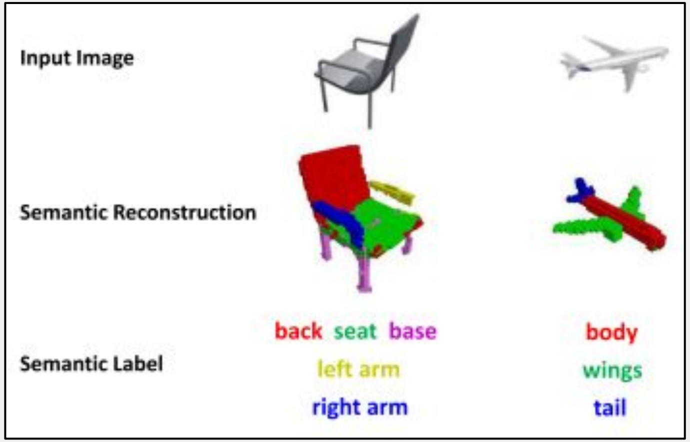
	airplane	bench	cabinet	car	chair	display	lamp	speaker	rifle	sofa	table	telephone	vessel	all
baseline	68.4	**61.6**	79.2	85.4	56.7	53.7	44.3	71.4	61.5	70.9	60.1	77.6	59.4	66.1
baseline + semantic label	68.5	60.8	78.6	85.0	58.1	55.0	**47.5**	71.7	63.5	72.3	59.5	**82.6**	60.5	66.8
baseline + prototype volume (ours)	**69.2**	60.7	**79.4**	85.4	**59.2**	**56.0**	47.4	**71.8**	**63.6**	**72.9**	**61.4**	80.8	**60.6**	**67.6**



Concept-Assisted Shape Creation







Discussion

Benefits

- Improved Understanding and Performance
- Structural Emphasis
- Multi-modal Fusion
- Data Augmentation

Limitations

- Increased Complexity and Training Time
- Memory and Efficiency Concerns

Summary

Novel multimodal framework

- Combines knowledge graph conceptual knowledge with deep neural networks
- Precise 3D shape reconstruction from single RGB images

Promising Results

- Tested on benchmark datasets
- Framework surpasses existing methods
- Ability to generate novel 3D shapes

Future Potential

- Expanding the framework to utilize other shape representations for more efficient high-resolution 3D reconstruction

Sources

https://www.researchgate.net/publication/363153397_Enhanced_3D_Shape_Reconstruction_With_Knowledge_Graph_of_Category_Concept

https://github.com/KKeishiro/Shape-from-Silhouettes

https://paperswithcode.com/task/3d-reconstruction

https://www.researchgate.net/publication/52006563_Three-Dimensional_Scene_Reconstruction_A_Review_of_Approaches

https://www.semanticscholar.org/paper/Multi-View-Stereo%3A-A-Tutorial-Furukawa-Hern%C3%A1ndez/a9d8591b30c579470edbece62201604b368887f4

https://www.researchgate.net/publication/329482025_A_Region-Based_Gauss-Newton_Approach_to_Real-Time_Monocular_Multiple_Object_Tracking

https://www.researchgate.net/publication/3414425_Representation_Plurality_and_Fusion_for_3-D_Face_Recognition

https://github.com/hzxie/Pix2Vox

https://www.andrew.cmu.edu/course/16-889/projects/rohansax/proj2/

https://cvgl.stanford.edu/projects/pascal3d.html

https://arxiv.org/abs/1901.11153