Abstract

This paper explores the integration of graph-based conceptual knowledge into 3D shape reconstruction from single-view images, addressing challenges in accuracy and detail that plague current methodologies. Through experimental validation on standard datasets, we demonstrate enhancements in reconstruction fidelity and propose future directions for research in this area.

Contents

Chapter 1

Introduction

In this paper, we analyze the work presented by Sun et al. [31], exploring its significant contributions to the field of knowledge graphs while also addressing its limitations. We discuss potential avenues for future research and propose solutions to overcome the identified challenges. This exploration sets the stage for understanding the balance between theoretical advancements and practical applications in 3D shape reconstruction.

1.1 Importance of 3D Shape Reconstruction

Reconstructing accurate three-dimensional (3D) objects from single-view images is a foundational pillar in computer vision that directly impacts an array of other high-level vision applications like shape retrieval, object localization, shape retrieval and virtual reality (VR). Due to its under-constrained nature, traditional reconstruction methods rely on silhouettes [9] or shading [27], but these heavy assumptions limit practical use. Recent advances in deep learning have significantly improved 3D shape reconstruction from single-view images using deep neural networks [19].

1.2 Challenges in Current Methods

Current approaches to single-view 3D reconstruction largely employ deep learning models structured around encoder-decoder frameworks, where 2D convolutional layers encode the image into feature representations and 3D deconvolutional layers decode these features to construct the 3D shape. Some methods extend such frameworks by incorporating a refinement step using 3D convolutional layers, which aim to enhance the details of the reconstructed shapes. However, these methods still struggle with creating realistic and de-

tailed shapes, often producing models that lack the intricacies of smaller, more complex parts. This limitation is primarily due to the dependence on basic, elementary assumptions derived from the 2D input images, which do not capture the full complexity of 3D structures [26].

1.3 Proposed Solution

Sun et al. [31] propose a novel approach that integrates graph-based conceptual knowledge with deep learning to overcome the limitations of current 3D shape reconstruction methods. Their framework combines a conceptual knowledge graph, including object categories and conceptual part labels, with a deep learning pipeline consisting of an image encoder, a conceptual knowledge encoder, a volume decoder, and a volume refiner. This setup allows for the fusion of high-level conceptual information with image-derived features, leading to the reconstruction of 3D shapes that are not only geometrically consistent with the input images, but also enriched with conceptual accuracy. By employing concept classifiers at both the input and output stages, the model ensures that the reconstructed shapes adhere to the visual and conceptual fidelity of the target objects. A concept loss function, reinforcing consistency between the reconstructed shape and its associated conceptual knowledge, enhances the quality and realism. The presented experimental results, conducted across multiple datasets including ShapeNet [5], Pascal3D+ [41], and Pix3D [33], demonstrate the superiority of this approach over existing state-of-the-art methods in terms of accuracy and realism [31].

Chapter 2

Literature Review

2.1 Traditional 3D Reconstruction Techniques

Reconstructing 3D shapes from single-view images is challenging due to its underdetermined nature. Traditional methods relied on extracting features from multiple images using techniques like Structure from Motion (SfM) [38] and Simultaneous Localization and Mapping (SLAM) [3], which are computationally demanding and error-prone. More recent methods use additional cues like shading [27], texture [39] and silhouettes [9], but are hindered by strong assumptions and limited practical applicability, especially when images of all object surfaces are needed. Kar et al. [18] introduced a category-specific approach that manipulates a mean shape for specific object categories, offering a more feasible but still restricted solution.

2.2 Deep Learning Approaches

The evolution from geometric models to neural networks in 3D shape reconstruction marks a significant technological shift. Initially, 3D reconstruction relied on manual-intensive geometric models, which evolved with the introduction of deep learning and large datasets like ShapeNet. This transition to neural networks, including innovations like 3D-R2N2 [10] and LSM [17], enhanced reconstruction efficiency by utilizing techniques such as recurrent neural networks and max pooling for feature fusion from images. Recent developments have focused on improving accuracy and speed, leveraging advanced methods like adversarial models [15] and consistency-based frameworks to reconstruct 3D shapes from single and multiple images with higher fidelity and realism.

2.3 Hybrid Methods

The field of 3D shape completion and generation tries to create realistic shapes from feature vectors and incomplete models. Wu et al. [40]introduced 3D-GAN and 3D-VAE-GAN models to produce category-specific shapes from a high-dimensional vector. Chen and Zhang's [8] IM-NET model employs an implicit decoder that combines feature vectors and coordinates to create high-resolution shapes. Other approaches leverage neural networks to determine the global structure of partial shapes, refining them locally [16]. Dai et al. [11] begin with a rough shape, enhancing it with geometric priors. Wang et al. [36] applied 3D encoder-decoder GANs and LSTM for detailed completion from low-resolution inputs. Additionally there are weakly supervised methods to streamline shape completion [30]. Unlike these methods, this approach [31] controls the generated shapes precisely using conceptual knowledge.

2.4 Knowledge Graphs in 3D Reconstruction

The integration of knowledge graphs in 3D reconstruction introduces innovative methods to enhance model retrieval and automate animation processes. Utilizing "geometric words" to represent basic structural components of 3D models, models connect simplified components through a 3D shape knowledge graph [24]. This facilitates efficient retrieval across diverse databases and modalities without extensive training, addressing challenges in both cross-domain and cross-modal retrieval. Notable methods leverage graph embedding techniques to capture structural similarities, improving accuracy in tasks like retrieval from standard datasets such as ModelNet40 [32] and ShapeNet [24].

Dynamic Knowledge Graphs (DKGs) further extend the utility of knowledge graphs by encoding spatial and temporal data to automate the generation of dynamic 3D scene animations from single images [37]. This system enables the creation of realistic animations that reflect plausible interactions within a given environment, tailored specifically to virtual and augmented reality applications. The innovations in contextual encoding and DKG-driven variational auto-encoders show significant potential, enhancing automation and realism in 3D animation across varied contexts.

By linking geometric components with shape categories and attributes, these methodologies provide a unified framework for 3D model retrieval and animation, reducing reliance on large training datasets and simplifying the processing of complex models [6]. This integration demonstrates superior performance in handling complex retrieval and animation tasks, setting a new standard for efficiency and accuracy in 3D reconstruction technologies.

Innovative applications of knowledge graphs in the realm of 3D design

and scene interpretation are advancing the capabilities of current technologies significantly. One notable development is the Graph Bridging Network (GB-Net) [45], which enhances scene graph generation from images by bridging commonsense knowledge graphs with scene graphs. This method matches scene graph entities and predicates to commonsense classes, iteratively refining the connections to improve accuracy and interpretability, as demonstrated in evaluations on the Visual Genome dataset [45].

In the area of 3D clothing design, a new framework utilizes a multimodal clustering network to construct a knowledge graph that addresses the challenges in 3D clothing visualization such as camera positioning and collector shape variability. By integrating advanced feature extraction and style template integration, this system enhances realism and customization in 3D clothing models [47].

Furthermore, the integration of a multi-relational and multi-hierarchical knowledge graph in 3D product design and manufacturing using large CAD model repositories has shown substantial advancements [2]. This knowledge graph combines 3D model metadata with assembly-part hierarchies and shape-based clustering to enable sophisticated search functionalities, including Approximate Nearest Neighbor search and rule-based inference for assembly similarity. This approach not only improves 3D shape retrieval and design reuse but also facilitates collaborative filtering for multimodal search of manufacturing processes, organizing large datasets into structured groups that enhance retrieval systems in design applications.

A novel approach employs a dual-layer knowledge graph structure for 3D semantic modeling of Chinese grottoes [44], integrating a schema layer and a data layer using ChgOnto—a new ontology. This framework leverages established ontologies like CIDOC CRM and GeoSPARQL, enhancing interoperability and precision in 3D modeling. Applied effectively to the Dazu Rock Carvings, this method underscores the importance of structured knowledge for robust knowledge sharing and improved 3D model utility in cultural heritage conservation.

Addressing dataset biases in scene graph generation, Gu et al. incorporate commonsense knowledge from external databases like ConceptNet and an image reconstruction loss to refine predictions of object relationships [14]. By integrating knowledge-based feature refinement and image-level supervision, this model improves scene graph accuracy and generalizability on benchmark datasets such as Visual Relationship Detection and Visual Genome. This method mitigates the effects of biased data and enhances the reliability of detected relationships.

The development of a Semantic 3D City Database integrates a dynamic geospatial knowledge graph to support The World Avatar project, focusing on

urban sustainability [4]. Utilizing an enhanced CityGML ontology and Blazegraph™ for scalable data storage, this system transforms static city models into dynamic repositories of urban data, facilitating complex queries and real-time updates essential for urban planning and management.

Building on the foundational work in deep learning and hybrid methods for 3D reconstruction, knowledge graphs offer a structured approach to further enhance model accuracy and applicability across different domains.

Chapter 3

Methodology

3.1 Framework Overview

The proposed framework integrates graph-based conceptual knowledge of object categories with deep neural networks for 3D shape reconstruction from a single RGB image. The framework includes the Pix2Vox-A architecture [42] and incorporates a conceptual knowledge encoder, image-based concept classifier, and volume-based concept classifier.

3.2 Component Descriptions

The image-based concept classifier predicts object category and conceptual part labels, which are used as input for the conceptual knowledge encoder. The proposed conceptual knowledge encoder extracts features from the input conceptual knowledge represented as 3D prototype volumes, derived from the mean shape of volumetric conceptual parts based on the concept graph. VGG16, a convolutional neural network architecture, is widely used due to its effectiveness in capturing hierarchical patterns in images. In this setup, the VGG16 module, which is a convolutional neural network architecture [22], serves as the backbone. In conjunction, the model incorporates three fully connected layers with 512, 512, and N channels respectively. The value of N corresponds to the total number of unique object categories and part labels targeted to classify. The first two 512-channel layers end with leaky ReLU activation functions. Leaky ReLU is particularly beneficial over standard ReLU in scenarios where negative input values may carry information, as it allows a small, non-zero gradient when the unit is inactive and helps maintain a flow of gradients during the backpropagation process [12].

$$L_{\text{rec}} = \frac{1}{N} \sum_{i=1}^{N} (p'_i \log(p_i) + (1 - p'_i) \log(1 - p_i)) \tag{3.1}$$

$$L_{\text{cpt}} = \frac{1}{M} \sum_{i=1}^{M} (s_i \log(s_i) + (1 - s_i) \log(1 - s_i)) \tag{3.2}$$

The volume-based concept classifier is used to detect conceptual parts from the reconstructed 3D shapes, contributing to network training based on the differences between the reconstructed 3D shape and the predicted conceptual knowledge. The framework is trained using reconstruction loss 3.1, involving the mean cross-entropy between reconstructed volume and ground-truth volume, to ensure the generated 3D shape is similar to the ground truth at the voxel level. Ground truth refers to the original, accurate 3d mesh data used as a reference standard for comparison with data produced by an experimental test or model [25]. Additionally, the concept loss shown in 3.2, which is the mean cross-entropy between predicted concept labels and ground-truth concept labels of input conceptual knowledge, encourages the network to generate a 3D shape similar to the input conceptual knowledge at the concept level.

The structure of the conceptual knowledge encoder involves the extraction of features from the input volume using 3D convolutional layers and fully connected layers, with subsequent feeding into the volume decoder and volume refiner. Preliminary experiments validate the effectiveness of this structure. The volume-based concept classifier comprises of 3D convolutional layers and a fully connected layer and detects conceptual parts from the reconstructed 3D shapes for network training.

3.3 Conceptual Knowledge Integration

The authors emphasize that the proposed modules are pipeline agnostic and can be applied to most 3D reconstruction pipelines. The proposed framework is capable of category-agnostic and category-specific training, with the latter involving two losses - reconstruction loss and concept loss, the latter being composed of the mean cross-entropy between predicted concept labels and ground-truth concept labels of input conceptual knowledge. The authors conducted experiments and established the weight of the concept loss λ as 0.01.

Chapter 4

Results and Analysis

4.1 Relevance of Datasets

The proposed method for 3D shape reconstruction using knowledge graphs is evaluated across synthetic and real-world datasets. The datasets include ShapeNet [5], which contains over 50,000 unique 3D models across 55 categories, with experiments conducted on 13 major categories comprising over 40,000 models. Models are rendered from multiple views to create synthetic RGB images. The Pascal3D+ dataset [41] provides real images with 3D annotations from various categories, aligned with those in ShapeNet, to fine-tune the models. The Pix3D dataset [33], specifically the chair category, is directly tested with the model trained on ShapeNet data.

4.2 Experimental Design - configuration and methodologies

The study integrates conceptual annotations from PartNet [23], and missing annotations were added manually. Detailed statistics on conceptual parts for various categories in the ShapeNet dataset are listed, including numbers of different components like wings on airplanes or seats on chairs.

For evaluating the reconstruction quality, the Intersection-over-Union (IoU) metric shown in 4.1 is used, which is standard in 3D object reconstruction. This metric measures the overlap between predicted and actual 3D volumes, with higher values indicating better accuracy.

The implementation details reveal that the system uses TensorFlow [1] with the Pix2Vox configuration for consistency across datasets. The models are optimized using the Adam algorithm, with adjustments in learning rates and epochs depending on whether the models are category-specific or

category-agnostic. The authors note that the code and processed datasets will be publicly released. This could unfortunately not be verified after thorough research.

$$\text{IoU} = \frac{\sum_{i,j,k} I(p(i,j,k) > t) I(p'_{i,j,k})}{\sum_{i,j,k} \left[I(p(i,j,k) > t) + I(p'_{i,j,k}) \right]} \tag{4.1}$$

4.3 Results - comparisons to the state-of-the-art

Sun et al. [31] detail various methods, comparisons, and improvements in performance metrics. In the synthetic data context, the research compares several state-of-the-art approaches like 3D-R2N2 [10], OGN [34], PSGN [13], Matryoshka [28], Voxel-Tube [28], Pix2Vox [42], and Pix2Vox++ [43] on the ShapeNet dataset. Pix2Vox++ upgrades the Pix2Vox by using a more advanced ResNet50 for image feature extraction instead of VGG16, alongside an improved fusion module. The study evaluates both category-agnostic and category-specific models. Notably, category-specific models are fine-tuned on different object categories. Additionally, the impact of concept loss on model training is analyzed by comparing models trained with and without this concept.

Table 4.1: Single-View 3D Object Reconstruction on ShapeNet with Volume Size 64^3 Using IoU (in %) Evaluation Metric. Adapted from [31]

Method	airplane	bench	cabinet	car	chair	display	lamp	speaker	rifle	sofa	table	telephone	vessel	all
3DensiNet	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	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	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
This	**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'	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
This'	**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**

The last two methods are category-specific models, while the others are category-agnostic models.

4.4 Analysis

Performance metrics reported in the study highlight that the researchers' method outperformed other approaches significantly. On category-agnostic

models, their method exceeded the baseline Pix2Vox in 11 out of 13 categories by a margin of +1.5% in IoU. The performance was still superior to Pix2Vox++ by +0.6%, despite Pix2Vox++ having a stronger feature extraction backbone. For category-specific models, the improvement was even more pronounced, with their approach exceeding Pix2Vox in all categories and showing a +2.4% improvement over their own variant without concept loss. Visual results in 4.1 showcase that their method preserves more details and reconstructs plausible 3D shapes on challenging data, such as chairs and boats, with clear preservation of complex details like airplane tails and chair arms.

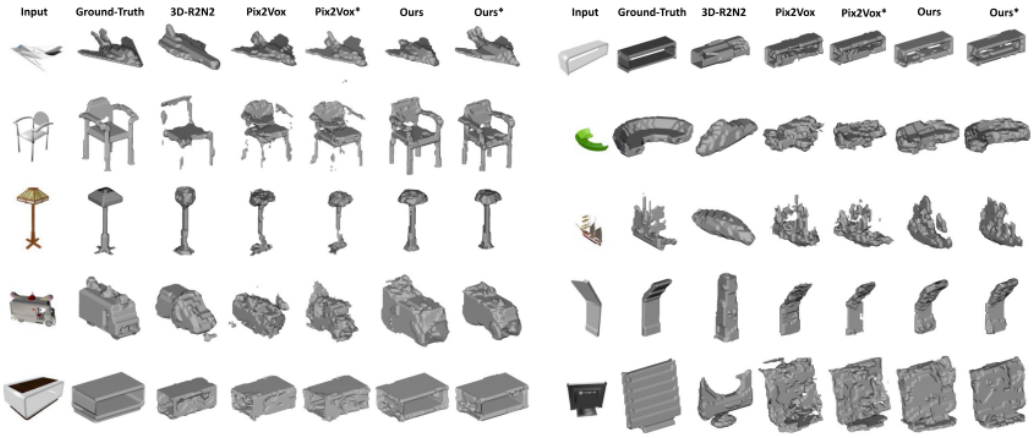

Figure 4.1: reconstruction results on ShapeNet [31]

Furthermore, the researchers modified existing frameworks like 3DensiNet [35] and Pix2Vox by adding 3D deconvolutional layers to enhance the resolution of 3D shape reconstruction. This adaptation led to better detail preservation in the reconstructed models, specifically in challenging parts like the legs and arms of chairs, where previous methods like Pix2Vox failed.

Transitioning to real-world data, the paper discusses comparisons on the Pascal3D+ and Pix3D datasets, as shown in 4.2, which contain real rather than synthetic images. On Pascal3D+, the new method outperformed existing approaches significantly across all categories, with improvements of +7.3% on category-agnostic and +4.9% on category-specific models compared to Pix2Vox. The better performance on Pascal3D+ was partly attributed to the dataset characteristics, where multiple images share the same ground-truth model, simplifying the reconstruction task. On the Pix3D dataset, the methodology included preprocessing images with BlendMask [7] to isolate object silhouettes before reconstruction, leading to superior results over previous methods by +1.7%. Visual examples highlighted the ability to accurately reconstruct 3D

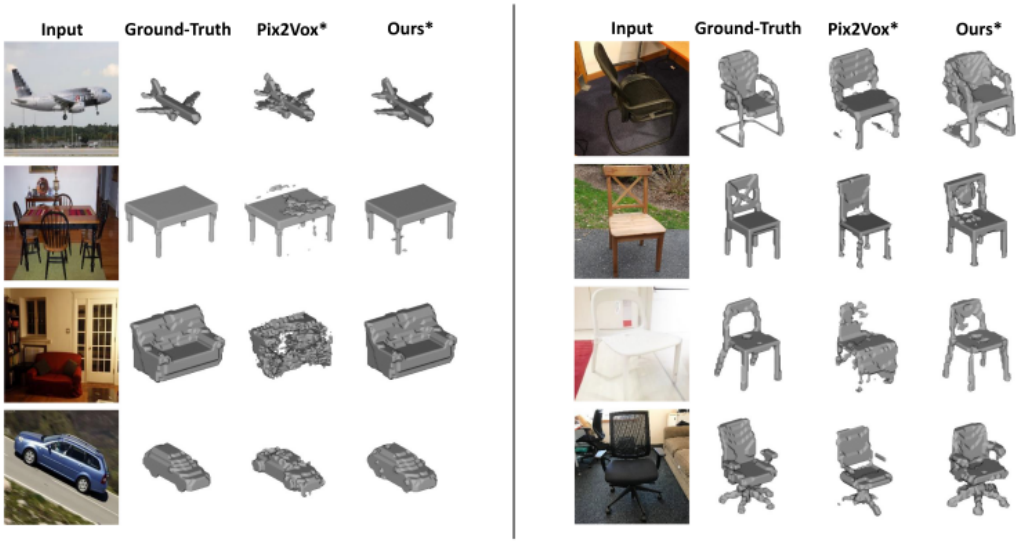

Figure 4.2: reconstruction results on Pascal3D+ (left) and Pix3D (right) [31]

shapes from challenging real-world images, like a chair with a hole in its back and maintaining topological consistency.

Overall, the results outline significant advancements in 3D shape reconstruction from both synthetic and real-world images, highlighting enhanced accuracy, detail preservation, and the effectiveness of incorporating concept loss and advanced neural network architectures in improving reconstruction.

4.4.1 Evaluating the Impact of Conceptual Knowledge

We validate this pipeline-agnostic conceptual knowledge framework for 3D shape reconstruction, including its prototype volume-based representation.

The effectiveness of the conceptual knowledge framework is tested by integrating it into existing 3D reconstruction pipelines including 3DensiNet, VoxelTube, and Pix2Vox. The modification involves concatenating encoded conceptual features with image features prior to the volume decoder/refiner module. The results show a significant improvement across all tested pipelines, with increases in performance ranging from +1.5% to +2.8% [31], affirming the framework's effectiveness.

The prototype volume-based conceptual knowledge representation, is compared against a baseline method that uses semantic labels (object category labels and object part labels) for knowledge representation. For example, in the Chair category of the ShapeNet dataset, conceptual knowledge is typically encoded using a one-hot vector for category and another vector for parts of the chair. The study employs a 6-layer MLP network to extract features from

these vectors for 3D reconstruction, using Pix2Vox as the baseline. The findings indicate that the prototype volume-based method, which includes more detailed information like coarse shapes and relative locations of object parts, outperforms the semantic label-based method by a margin of +1.5% compared to +0.7% [31], demonstrating its superior utility in guiding the reconstruction process.

Chapter 5

Discussion

The introduction of three modules—concept classifier (image), conceptual knowledge encoder, and concept classifier (volume)—into the Pix2Vox baseline architecture markedly enhanced 3D reconstruction performance across all evaluation datasets, with improvements ranging from +1.5% to +7.3%. This enhancement stems from the inclusion of conceptual knowledge, which offers a holistic understanding of objects and influences the learning of high-level structural features during training. This approach differs from traditional bottom-up 3D reconstruction methods that rely solely on voxel probability predictions from image features and are limited by a loss function that only addresses voxel dissimilarity.

The integration of conceptual knowledge with image features through a volume decoder and refiner enables more effective and simpler reconstruction from fused multi-modal features, unlike methods that use only singular modal features. Additionally, the capacity to associate one image with multiple ground-truth 3D shapes provides a form of data augmentation, enhancing network generality and reducing overfitting [31].

However, the research acknowledges limitations such as increased training complexity and time due to the addition of novel modules, which collectively add more parameters (161M vs. 114M) and extend training duration (45h vs. 25h). Future work would benefit from addressing these challenges by exploring more memory-efficient 3D shape representations like octrees [46] and point clouds [21].

5.1 Expanding Applications: Shape Creation and Semantic Reconstruction

The use of knowledge graphs introduces additional applications of the framework. One application is concept-assisted shape creation, where modifying predicted conceptual knowledge allows for the creation of novel 3D shapes from images. By manipulating the conceptual knowledge graph, different parts of an object can be included or excluded in the reconstruction process, such as creating a chair without arms by removing 'arms' from the graph. This method ensures that the generated 3D shapes are consistent with both the geometry and concept of the input images.

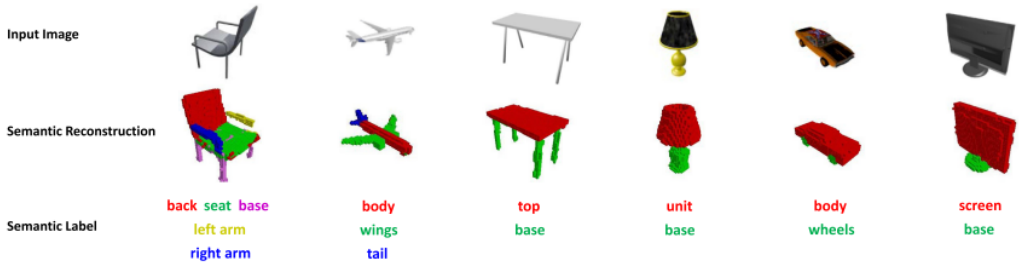

Figure 5.1: Reconstructing semantic 3D shapes [31]

Another application is concept-assisted semantic shape reconstruction. This approach treats each concept as a semantic part, reconstructing semantic parts of a 3D shape through subtraction of different conceptual reconstructions. For instance, to isolate the back of a chair, the shape reconstructed using knowledge of 'base and seat' is subtracted from that using 'base, seat, and back'. The performance of this approach is quantitatively compared with three other methods: a baseline encoder-decoder architecture , Shi's part-level reconstruction [29], and AICNet [20], a semantic scene reconstruction approach. Results show that this framework [31] outperforms the others in reconstructing larger parts and classifying semantic labels, though it may underperform on smaller or more detailed parts compared to Shi's method.

Overall, the additional study substantiates the utility of the conceptual knowledge framework in enhancing the accuracy and detail of 3D reconstructions across different implementations and applications.

5.2 Limitations and Future Work

Framework Complexity and Training Requirements The integration of graph-based conceptual knowledge with deep neural networks adds significant complexity to the model architecture. This complexity can lead to increased computational demand, higher memory usage, and longer training times. The framework incorporates advanced neural architectures which, while effective, may not be easily replicable in environments with limited computational resources.

Data Dependency and Model Generalizability The results highlight the framework's performance across specific datasets like ShapeNet, Pascal3D+, and Pix3D. While impressive, this may not fully indicate the model's performance on datasets with different characteristics or poorer annotations. The dependency on high-quality, well-annotated data for training and validation could limit the applicability in scenarios where such data is not available.

Scalability and Real-Time Processing The scalability of the method to larger datasets or real-time applications isn't discussed extensively. Given the detailed architecture and the computational overhead, it would be useful to address the framework's ability to scale and its suitability for real-time processing applications in fields such as augmented reality or robotics.

Comparative Analysis with Other Methods While this paper discusses the superiority of the proposed method over existing techniques, a very thorough comparative analysis could be beneficial. This includes not only comparisons in terms of accuracy but also in aspects like efficiency, ease of deployment, and performance under varying conditions.

Impact of Conceptual Knowledge Integration The novel aspect of integrating conceptual knowledge is a key strength of the work by Sun et al. [31]. However, discussing potential limitations related to the preciseness and extent of this integration could be insightful. For instance, how does the variability in conceptual knowledge quality affect the reconstruction results? Exploring this could highlight areas for further enhancement of the knowledge encoding process.

Adaptability to New Categories The adaptability of the model to new object categories or unforeseen scenarios is another area worth exploring. Discussing how the model performs when faced with categories not present in the training set could give insights into its flexibility and adaptability.

Chapter 6

Conclusion

This work provides a critical analysis of the integration of graph-based conceptual knowledge with 3D shape reconstruction from single-view images, as presented in the work by Sun et al [31]. The enhancements in reconstruction fidelity and conceptual accuracy achieved through this integration mark significant advancements over traditional methods. These improvements underscore the potential of knowledge graphs to augment the geometric consistency and conceptual detail of 3D models.

Despite these advancements, the research identifies several notable limitations. The approach introduces increased computational complexity and resource demands, which may inhibit its deployment in environments with limited computational capabilities. Additionally, the dependency on highly annotated datasets restricts the model's generalizability across different scenarios where such data may not be available. It is recommended that future investigations focus on optimizing computational efficiency and enhancing model robustness to ensure broader applicability and scalability.

In summation, while the application of knowledge graphs in 3D shape reconstruction shows considerable promise in enhancing methodological outcomes, critical attention must be given to its existing constraints to encourage its development and adoption in diverse applications within the field of computer vision.

Bibliography

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. {TensorFlow}: a system for {Large-Scale} machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pages 265–283, 2016.

[2] Akshay Bharadwaj and Binil Starly. Knowledge graph construction for product designs from large cad model repositories. *Advanced Engineering Informatics*, 53:101680, 08 2022.

[3] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, and John J Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on robotics*, 32(6):1309–1332, 2016.

[4] Arkadiusz Chadzynski, Nenad Krdzavac, Feroz Farazi, Mei Lim, Shiying Li, Ayda Grisiute, Pieter Herthogs, Aurel von Richthofen, Stephen Cairns, and Markus Kraft. Semantic 3d city database — an enabler for a dynamic geospatial knowledge graph. *Energy and AI*, 6:100106, 07 2021.

[5] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

[6] Rihao Chang, Yongtao Ma, Tong Hao, and Weizhi Nie. 3d shape knowledge graph for cross-domain 3d shape retrieval. 2022.

[7] Hao Chen, Kunyang Sun, Zhi Tian, Chunhua Shen, Yongming Huang, and Youliang Yan. Blendmask: Top-down meets bottom-up for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8573–8581, 2020.

[8] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5939–5948, 2019.

[9] KMG Cheung, Simon Baker, and Takeo Kanade. Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 1, pages I–I. IEEE, 2003.

[10] Christopher B Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14*, pages 628–644. Springer, 2016.

[11] Angela Dai, Charles Ruizhongtai Qi, and Matthias Nießner. Shape completion using 3d-encoder-predictor cnns and shape synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5868–5877, 2017.

[12] Arun Kumar Dubey and Vanita Jain. Comparative study of convolution neural network's relu and leaky-relu activation functions. In Sukumar Mishra, Yog Raj Sood, and Anuradha Tomar, editors, *Applications of Computing, Automation and Wireless Systems in Electrical Engineering*, pages 873–880, Singapore, 2019. Springer Singapore.

[13] Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 605–613, 2017.

[14] Jiuxiang Gu, Handong Zhao, Zhe Lin, Sheng Li, Jianfei Cai, and Mingyang Ling. Scene graph generation with external knowledge and image reconstruction, 2019.

[15] JunYoung Gwak, Christopher B Choy, Manmohan Chandraker, Animesh Garg, and Silvio Savarese. Weakly supervised 3d reconstruction with adversarial constraint. In *2017 International Conference on 3D Vision (3DV)*, pages 263–272. IEEE, 2017.

[16] Xiaoguang Han, Zhen Li, Haibin Huang, Evangelos Kalogerakis, and Yizhou Yu. High-resolution shape completion using deep neural networks

for global structure and local geometry inference. In *Proceedings of the IEEE international conference on computer vision*, pages 85–93, 2017.

[17] Abhishek Kar, Christian Häne, and Jitendra Malik. Learning a multi-view stereo machine. *Advances in neural information processing systems*, 30, 2017.

[18] Abhishek Kar, Shubham Tulsiani, Joao Carreira, and Jitendra Malik. Category-specific object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1966–1974, 2015.

[19] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[20] Jie Li, Kai Han, Peng Wang, Yu Liu, and Xia Yuan. Anisotropic convolutional networks for 3d semantic scene completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3351–3359, 2020.

[21] Chen-Hsuan Lin, Chen Kong, and Simon Lucey. Learning efficient point cloud generation for dense 3d object reconstruction. In *proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[22] Sheldon Mascarenhas and Mukul Agarwal. A comparison between vgg16, vgg19 and resnet50 architecture frameworks for image classification. In *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*, volume 1, pages 96–99, 2021.

[23] Kaichun Mo, Shilin Zhu, Angel X Chang, Li Yi, Subarna Tripathi, Leonidas J Guibas, and Hao Su. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 909–918, 2019.

[24] Weizhi Nie, Ya Wang, Dan Song, and Wenhui Li. 3d model retrieval based on a 3d shape knowledge graph. *IEEE Access*, PP:1–1, 08 2020.

[25] M. Nielsen, Hans Andersen, David Slaughter, and Erik Granum. Ground truth evaluation of computer vision based 3d reconstruction of synthesized and real plant images. *Precision Agriculture*, 8:49–62, 01 2007.

[26] Yun-he Pan. On visual knowledge. *Frontiers of Information Technology Electronic Engineering*, 20:1021–1025, 08 2019.

[27] Stephan R. Richter and Stefan Roth. Discriminative shape from shading in uncalibrated illumination. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1128–1136, 2015.

[28] Stephan R. Richter and Stefan Roth. Matryoshka networks: Predicting 3d geometry via nested shape layers, 2018.

[29] Dingfeng Shi, Yifan Zhao, and Jia Li. Reconstructing part-level 3d models from a single image. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2020.

[30] David Stutz and Andreas Geiger. Learning 3d shape completion under weak supervision. *Int. J. Comput. Vision*, 128(5):1162–1181, may 2020.

[31] Guofei Sun, Yongkang Wong, Mohan S. Kankanhalli, Xiangdong Li, and Weidong Geng. Enhanced 3d shape reconstruction with knowledge graph of category concept. *ACM Trans. Multimedia Comput. Commun. Appl.*, 18(3), mar 2022.

[32] Jiachen Sun, Qingzhao Zhang, Bhavya Kailkhura, Zhiding Yu, Chaowei Xiao, and Z Morley Mao. Modelnet40-c: A robustness benchmark for 3d point cloud recognition under corruption. In *ICLR 2022 Workshop on Socially Responsible Machine Learning*, volume 7, 2022.

[33] Xingyuan Sun, Jiajun Wu, Xiuming Zhang, Zhoutong Zhang, Chengkai Zhang, Tianfan Xue, Joshua B. Tenenbaum, and William T. Freeman. Pix3d: Dataset and methods for single-image 3d shape modeling. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2974–2983, 2018.

[34] Maxim Tatarchenko, Alexey Dosovitskiy, and Thomas Brox. Octree generating networks: Efficient convolutional architectures for high-resolution 3d outputs. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2107–2115, 2017.

[35] Meng Wang, Lingjing Wang, and Yi Fang. 3densinet: A robust neural network architecture towards 3d volumetric object prediction from 2d image. In *Proceedings of the 25th ACM International Conference on Multimedia*, MM '17, page 961–969, New York, NY, USA, 2017. Association for Computing Machinery.

[36] Weiyue Wang, Qiangui Huang, Suya You, Chao Yang, and Ulrich Neumann. Shape inpainting using 3d generative adversarial network and recurrent convolutional networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2317–2325, 2017.

[37] Song Wenfeng, Xinyu Zhang, Yuting Guo, Shuai Li, Aimin Hao, and Hong Qin. Automatic generation of 3d scene animation based on dynamic knowledge graphs and contextual encoding. *International Journal of Computer Vision*, 131:1–29, 07 2023.

[38] Matt Westoby, James Brasington, Neil Glasser, Michael Hambrey, and John Reynolds. Structure-from-motion photogrammetry: a novel, low-cost tool for geomorphological applications. *Geomorphology*, pages 936–, 04 2012.

[39] Andrew P. Witkin. Recovering surface shape and orientation from texture. *Artif. Intell.*, 17(1–3):17–45, aug 1981.

[40] Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T. Freeman, and Joshua B. Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 82–90, Red Hook, NY, USA, 2016. Curran Associates Inc.

[41] Yu Xiang, Roozbeh Mottaghi, and Silvio Savarese. Beyond pascal: A benchmark for 3d object detection in the wild. In *IEEE Winter Conference on Applications of Computer Vision*, pages 75–82, 2014.

[42] Haozhe Xie, Hongxun Yao, Xiaoshuai Sun, Shangchen Zhou, and Shengping Zhang. Pix2vox: Context-aware 3d reconstruction from single and multi-view images. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2690–2698, 2019.

[43] Haozhe Xie, Hongxun Yao, Shengping Zhang, Shangchen Zhou, and Wenxiu Sun. Pix2vox++: Multi-scale context-aware 3d object reconstruction from single and multiple images. *Int. J. Comput. Vision*, 128(12):2919–2935, dec 2020.

[44] Su Yang and Miaole Hou. Knowledge graph representation method for semantic 3d modeling of chinese grottoes. *Heritage Science*, 11, 12 2023.

[45] Alireza Zareian, Svebor Karaman, and Shih-Fu Chang. Bridging knowledge graphs to generate scene graphs, 2020.

[46] Ming Zeng, Fukai Zhao, Jiaxiang Zheng, and Xinguo Liu. Octree-based fusion for realtime 3d reconstruction. *Graphical Models*, 75(3):126–136, 2013. Computational Visual Media Conference 2012.

[47] Jia Zheng and Wei Hong. Construction of knowledge graph of 3d clothing design resources based on multimodal clustering network. *Computational Intelligence and Neuroscience*, 2022:1–12, 06 2022.