

Addressing Challenges in Existing 3D Datasets: Insights from Mondial3D.AI (working-Paper)

Advancements in Natural Language Processing and 2D/3D Models

In recent years, the field of natural language processing and 2D vision models has experienced remarkable progress, largely attributed to the increase in training datasets. In the changing AI landscape, the importance of extensive datasets is crucial. Large-scale models have transformed various fields, from language comprehension to image creation. Substantial advancements in NLP are evident with models such as GPT-4 - llama, highlighting the impact of utilizing abundant web-crawled data. Similarly, within computer vision, ImageNet's adaptation to billions of images was a significant breakthrough that resulted in enhanced visual recognition. Transformer-based language models have reshaped natural language processing and become the standard paradigm for most NLP downstream tasks. Now, these models are rapidly advancing to other domains such as computer vision. With the integration of natural language processing and computer vision, the synergy between these domains has led to innovations such as image captioning and visual question answering ([Zhou, 2015](#)). This progress has enabled significant improvements in tasks such as image recognition, object detection, and language understanding. Moreover, advancements in artificial intelligence powered by deep artificial neural networks have significantly impacted foundational visual tasks such as object recognition and scene generation from natural language descriptions ([Ren & Bashivan, 2023](#)). It is important to note that deep learning techniques, particularly CNN models, have demonstrated remarkable performance in handling 2D and 3D images ([Polat, 2022](#)).

Furthermore, the success of machine learning, especially deep neural networks like CNNs, VAEs, and GANs in computer vision and natural language processing, has paved the way for applying deep learning models in geometric modeling and 3D voxel grid generation ([Zhang et al., 2020](#)). Continuing with this idea, the emergence of Mondial3D Dataset, aka **M3D Dataset**, presents a promising solution to two of the most pressing challenges (if not more) faced in 3D design/vision tasks - the scarcity of high-quality and diversity of 3D data for training and testing. The availability of large 3D object datasets and advances in deep learning has given rise to new methods for multi-view 3D reconstruction, often employing an encoder-decoder architecture to map 2D images into 3D volumes ([Yagubbayli et al., 2021](#)). The M3D Dataset

is a significant advancement in the realm of 3D vision tasks, as it offers a comprehensive collection of over 20 million 3D objects, which is growing rapidly with the help of [our virtual 3D studio](#).

The Challenge of 3D Data Collection

Recently, there has been an increase in the use of 3D model datasets for research purposes in computer vision and computer graphics due to advancements in deep learning algorithms. These include CAD model datasets like ModelNet, ShapeNet, COSEG Dataset, and ABC. Many varied 3D deep-learning tasks are conducted on these databases. One prominent benchmark dataset used is called ModelNet, while ShapeNet provides a vast repository of clean mesh models covering multiple categories organized under WordNet taxonomy. Supporting further progress is important because understanding objects in three dimensions remains challenging due to limited real-world data sets compared to two-dimensional tasks.

Collecting high-quality 3D data has been a bottleneck in the progression of 3D vision tasks. Traditional 3D datasets, created through labor-intensive processes such as manual 3D scanning or human annotation, are limited in scale and diversity. The complexities involved in gathering high-caliber 3D data have been a prominent obstacle. The introduction of large-scale 3D model datasets such as ShapeNet and ModelNet has been a crucial step in supporting deep learning models to learn complex 3D representations. The utilization of 3D deep learning techniques has been widely explored in interdisciplinary fields such as architecture, computer graphics, and computer vision ([Keshavarzi et al., 2020](#)). Large-scale datasets often lack the diversity and high quality required to train LLMs, leading to less reliable 3D results.

In recent years, with the remarkable performance of Convolutional Neural Network in target detection, natural language processing, and image recognition, there has been a growing interest in adapting these models to handle 3D data. Particularly, CNN models have demonstrated remarkable performance in handling 2D and 3D images. Expanding on the challenge of 3D data collection, the difficulty in acquiring high-quality large and diverse 3D datasets has been a significant hurdle facing researchers attempting to advance the capabilities of deep learning models in 3D vision tasks.

The transition from 2D image synthesis to the application of machine learning on 3D voxel grids has presented its own challenges, as it requires robust 3D datasets for pre-training such models. Continuing with this idea, the emergence of Mondial3D Dataset, aka M3D Dataset,

presents a promising solution to two of the most pressing challenges faced in 3D design/vision tasks - the scarcity of high-quality and diversity of 3D models for training and testing.

Introduction to M3D Dataset

In the realm of 3D Design, M3D stands as a pivotal dataset. As an AI-driven online 3D studio and a platform for 3D model exchange, we have harnessed the potential of user-generated 3D models. Extensively curated by our users, our collection of 3D data captures each object from various angles within a spherical environment. This careful approach enables us to produce detailed multi-view renderings of over 20 million unique objects in various dimensions. Each image comes with precise metadata encompassing object types, positions, rotations, and camera coordinates. The M3D dataset includes high-quality material maps with BSDF and BRDF data, ensuring the assets are re-lightable and adaptable to various lighting scenarios. However, the M3D dataset offers pre-rendered 359+1 high-resolution images, capturing the essence of each detailed 3D model, thus enabling researchers to concentrate on dataset training without additional processing. Such innovative methodology in data aggregation transcends conventional practices in crafting a dataset, distinguished by its unparalleled scale and intricacy.

In widely explored interdisciplinary fields such as architecture, computer graphics, and computer vision, large-scale datasets often lack the necessary diversity and high quality required to train complex machine learning models effectively. To address this limitation, the M3D Dataset introduces a diverse and expansive collection of 3D objects sourced from various origins. Drawing on the challenges mentioned in the sources, Mondial3D Dataset (M3D) comes as a game-changing resource for researchers and practitioners in the field of 3D vision tasks.

Creation & Curation of the M3D Dataset

Our position as an AI-powered virtual 3D studio, a 3D model marketplace, and an online AI scene generator has made the creation of the M3D dataset possible. By offering complimentary tokens to access our AI tools, we provide users with the opportunity to create new 3D objects and scenes. This approach, as well as our in-house developed and tested algorithms, have helped us collect one of the largest live datasets of high-quality 3D objects ever created. We maintain exceptional quality through human oversight within our studio platform on the admin panel before uploading in the M3D dataset. Designers have the option to modify or enhance their uploaded 3D objects in our studio without losing ownership rights but granting us

permission for training purposes upon confirmation at upload time. Our careful vetting process, which includes both automated checks and human supervision adhering strictly to high standards, ensures that only valuable resources are contributed back into this community while eliminating subpar or unnecessary items rigorously.

The extensive scope of the M3D dataset notably enhances the capabilities of 3D models. Utilizing our dataset to test/train models such as Zero123 for novel view synthesis demonstrates remarkable zero-shot generalization abilities. This improvement is evident across diverse 3D tasks, highlighting the significant potential of our expansive and detailed 3D data. The complexities involved in gathering high-caliber 3D data have been a prominent obstacle. However, with the M3D Dataset, this barrier has been effectively overcome and tested.

Leveraging Existing Resources in 3D Computer Vision

The advancements in 3D data capturing and processing techniques have significantly increased the accessibility of large-scale 3D datasets, thereby revolutionizing the field of 3D computer vision. This has led to the emergence of multifunctional datasets, such as the M3D dataset, designed for benchmarking multiple problems, including 3D mesh reconstruction, object instance recognition, and categorization. Access to such rich resources is instrumental in driving innovative research applications in various domains, such as gaming, virtual reality, and generative AI. The M3D dataset is positioned as an invaluable resource in fueling the development of cutting-edge 3D deep learning techniques, enabling researchers to explore novel pathways in tasks like 3D object recognition, semantic segmentation, and shape analysis. By incorporating the extensive collection of over 20 million 3D objects from diverse sources, the M3D dataset stands as one of the most comprehensive and diverse resources available for training deep learning algorithms in 3D computer vision.

Conclusion

In conclusion, the Mondial3D Dataset (M3D Dataset) represents a significant breakthrough in the field of 3D computer vision. The meticulous curation process, which includes both automated checks and human supervision adhering strictly to high standards, ensures that only valuable resources are contributed back into eliminating subpar or unnecessary items rigorously.

The extensive scope of the M3D dataset notably enhances the capabilities of 3D models. Utilizing our dataset to test/train models such as Zero123 for novel view synthesis demonstrates remarkable zero-shot generalization abilities.

The emergence of multifunctional datasets, such as the M3D dataset, designed for benchmarking multiple problems, including 3D mesh reconstruction, object instance recognition, and categorization, is indicative of a promising direction for the field of 3D computer vision.

This dataset is expected to play a significant role in overcoming the existing challenges in the field of 3D reconstruction and evaluation, thereby facilitating the widespread adoption of standardized datasets for 3D tasks in computer vision.

References

- Zhou, B. (2015, January 1). Simple Baseline for Visual Question Answering.
<https://scite.ai/reports/10.48550/arxiv.1512.02167>
- Ren, Y., & Bashivan, P. (2023, May 3). How well do models of visual cortex generalize to out of distribution samples?. <https://scite.ai/reports/10.1101/2023.05.03.539191>
- Polat, H. (2022, December 31). Time-Frequency Complexity Maps for EEG-Based Diagnosis of Alzheimer's Disease Using a Lightweight Deep Neural Network.
<https://scite.ai/reports/10.18280/ts.390623>
- Zhang, Y., Chan, W O., Zheng, J., & Lie, S. (2020, September 1). Creative Corbel Modeling Using Evolution Principle. <https://scite.ai/reports/10.1109/cw49994.2020.00010>
- Yagubbayli, F., Wang, Y., Tonioni, A., & Tombari, F. (2021, January 1). LegoFormer: Transformers for Block-by-Block Multi-view 3D Reconstruction.
<https://scite.ai/reports/10.48550/arxiv.2106.12102>
- Keshavarzi, M., Afolabi, O., Caldas, L., Yang, A Y., & Zakhor, A. (2020, January 1). GenScan: A Generative Method for Populating Parametric 3D Scan Datasets.
<https://scite.ai/reports/10.48550/arxiv.2012.03998>