REACTION PAPER

In an era of generative AI, 3D generative modeling would be an extrapolation to capture the widely growing market. In 2D images and in video the end goal is usually to be pleasing to the eye and not hallucinate, however 2D Gen AI has crossed to 'AI hallucination' quite a few times where the generated image is nowhere close to reality. It is quite evident that 3D Gen needs to be take the necessary precautions against following the same path. Compare a bad image with no depth perception to an 3D figurines ad like the Airbnb with tables and chairs having holes and defects. This rightly puts the emphasis on significance of the research presented. With the research such as the one presented it was quite evident that the perspective geometry which was not caught on by the humans plays an important role in 3D Gen. Physical stability which was quite emphasized in the presentation has many practical applications such as in 3D printing. If nuts and bolts of an aircraft is 3D printed it needs to be stable. The research also emphasis on topology which needs to be smooth for consumption and not have holes or discontinuity. Many of the 3D Gen models are used in fashion design industry where the right topology will play a very important role. The last two papers which presented the work on deformable shape generator would play a major role in the cinematography.

The flow of the presentation did go in the right direction. For the first 4 papers which the speaker mentioned he goes on to explain the issue with the existing technique and how the current work fixes the issue. This is aided by the quantitative comparative analysis with the baseline frameworks. It was also supplemented with visual analysis/comparison which was the right approach as it is the humans who are at the receiving end of this technology. The fifth paper presentation however did not meet the expectation as that of the first 4. With Ads using human figurines in 3D such as with Airbnb, it would be the right approach to spend more time on deformable shape generators.

The concepts were explained in simple terms spending very less time on the technical aspects which was the right approach as it would cover a lot wider audience, than limiting it to computer vision. The first three paper made it easier for the presenter to explain the intuition behind the mathematics, example how setting the hyperparameters for the training would improve the outcome, defining the weights as to how much of the training data is to be used and how much of the quality checked data to be used for training. However, an in-depth technical understanding would be the need of the hour to dwell into deformable shape generators. Geodesic interpolation was something that fascinated me the most.

In terms of the visual presentation, there were tabular data which rightly conveyed that the work being presented was doing better than the state-of-the-art technologies. The geometry figures rightly pointed to the improvement in shape and stability highlighted by the red dotted lines. The lines drawn on the OpenAI SORA video was convincing enough to prove that human perception system was not very sensitive to perspective geometry. There could be better images to explain how two shapes could be interpolated which could work, but does not preserve deformable structure. Since the work being presented deals majorly with sight, animation videos would have done better justice than images. The diffusion model in 2D images was presented very well. Some of the mathematical expressions were well explained through plotting the functions such as the Persistent diagram. The improvement in the loss function was well explained but could be visualized better through comparison of outputs.

REACTION PAPER

The presenter took a critical analysis perspective where he mentions the state of the art work in 2D generative AI and how it performs in the 3D generative modelling. He differentiates between the two domains which helps to communicate the idea perfectly to the audience. He also challenges the current notion that having more training data will lead to better models and connects the concept to human perception which was quite intuitive. Mentioning an idea and countering it with another idea provides a comparative analysis which was well executed when the presenter mentioned worst case structural analysis and comparing it with use of simulation which does not optimize for the worst-case forces. The presenter initially mentions new ideas which were incorporated using the EDM framework such as setting the hyperparameters for the training and concludes by pinpointing how much improvement in percentage (30%) was caused by hyperparameter optimization. The mapping between improvement percentage to the idea gives a detailed report on the outcome. This approach nudges the audience to concentrate more on fields which are bound to give better outcomes. The idea of presenting a technology, pointing out the issues with it and then providing the solution sits right with the flow of thought. The author also does not shy away from mentioning the challenges associated with the field such as to incorporate 2DgenAI to 3D there is a need to solve the representation problem and encoding data into standard forms like vectors, arrays, graphs which is yet to be completely optimized. He gives a timeline example to help the audience understand the complexity of the field. The tools being generated in the computer graphics field needs to be 100% robust while a few years ago 90% was well appreciated. This rightly conveys the need of the hour, for a technology to have a market space there is a need for it be robust and easily incorporated into learning. This would rightly fit to ChatGPT which approximately has 80% of the market share in transformer market, while other such as Gemini, Mistral are yet to evolve completely.

The research presented showed that Machine learning and Computer graphics techniques synchronize which rightly fits into my research interest. Having worked on dictionary learning for image processing the next right step was towards computer vision and graphics. While my work on dictionary learning in image processing worked with the idea that more the training samples better the results which aligns with the ideology of most of the machine learning algorithms, a whole new idea was presented where more the training data might not lead to best results in 3D Gen modeling which might lead to new field of study where appropriating the right training data would optimize the output.