

---

# ViewPoint: Panoramic Video Generation with Pretrained Diffusion Models

---

Anonymous Author(s)

Affiliation

Address

email

1 In this appendix, we discuss the usage of computational resources, including VRAM consumption  
2 and inference latency in appendix A. In appendix B, we demonstrate that our method can support  
3 text-driven panoramic video generation, further expanding its application scenarios. Afterwards, we  
4 present the limitations of our method in appendix C. Finally, we discuss the potential societal impact  
5 in appendix D. We also provide more visual results, please refer to [https://anonymouser00.  
6 github.io/](https://anonymouser00.github.io/).

## 7 A Computational Resources

### 8 A.1 Representation Comparison

9 We compare our method with two mainstream panorama representations in terms of area. As shown  
10 in Fig. 1, assuming that the side length of a perspective anchor with a field-of-view (FoV) of  $90^\circ \times 90^\circ$   
11 is  $r$ , the area of the equirectangular representation is  $8r^2$ , and the area of the 2D grid in the cubemap is  
12  $12r^2$ , among which the valid representation area is  $6r^2$ . In contrast, our representation occupies an  
area of  $8r^2$ .

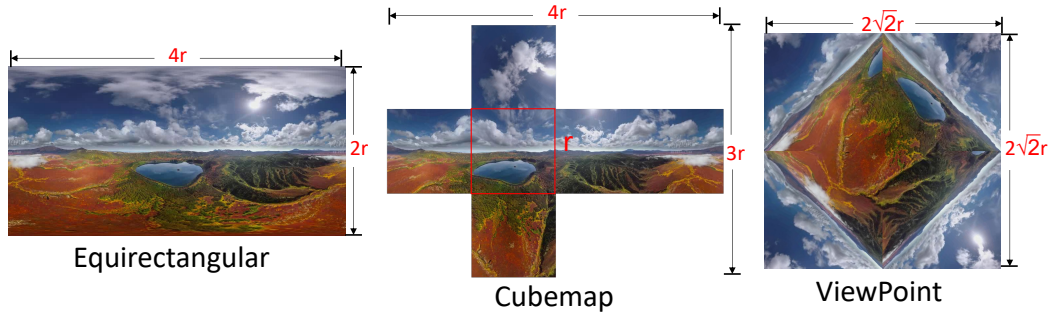


Figure 1: Comparison of different representations in terms of area.

13

### 14 A.2 Methods Comparison

15 We compare our method with 360DVD [4], Follow-Your-Canvas [2], and Imagine360 [3] in terms of  
16 model complexity and inference-time latency on a single NVIDIA A800(80GB). As shown in Tab. 1,  
17 in the case of generating  $512 \times 1024$  49-frame equirectangular videos (with the exception of 360DVD,  
18 which is limited to 16-frame video generation), our method demonstrates superior performance in  
19 terms of inference latency, model parameters, and peak VRAM consumption compared to prior  
20 approaches. Note that we follow the original settings of the compared methods. Among them,  
21 Follow-Your-Canvas [2] exhibits extremely long inference latency due to the progressive outpainting  
22 design. Imagine360 [3], on the other hand, shows significantly high model complexity and a large

number of parameters, resulting from its dual-branch architecture. In contrast, our method achieves excellent model efficiency, generating high-quality results with significantly fewer computational resources.

Table 1: Comparison of computational resource usage with previous methods. Note that 360DVD only supports the generation of 16-frame videos, while the other three methods are evaluated in the 49-frame scenario.

	360DVD(16 frames) [5]	Follow-Your-Canvas [2]	Imagine360 [3]	Ours
Latency↓	1min30s	29min19s	4min44s	1min47s
Peak VRAM↓	17.44GB	32.57GB	40.45GB	20.77GB
Parameters↓	1.33B	1.42B	5.94B	1.42B

## B More Application

Our model not only supports video input, but also pure text input. Therefore, our method can also be applied to text-driven panoramic video generation. As shown in Fig. 2, our approach is capable of generating high-quality panoramic video using only textual guidance.

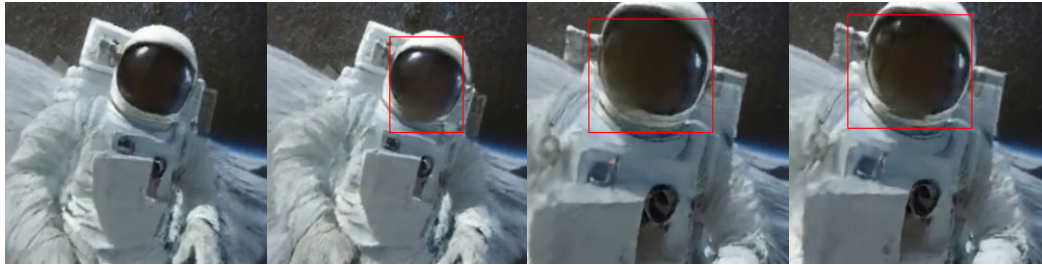


Figure 2: Text-driven generation. Our method can generate high-quality panoramic video using only textual guidance.

## C Limitations

As shown in Fig. 3, the panoramic videos generated by our model contain watermarks, exhibit mosaic artifacts on human faces, and occasionally reveal photographic equipment such as selfie sticks in the  $\mathcal{D}$  regions of the video, which can degrade the immersive experience. These limitations can be attributed to the presence of watermarks and specialized post-processing in our training data [1, 5]. Specifically, 360 camera manufacturers typically embed their logos into the recorded videos, leading to an abundance of watermarked content within the dataset. In addition, some works [1] choose to apply a mosaic effect to human faces in the datasets for ethical and moral considerations. We acknowledge and respect the ethical considerations taken by the creators of the datasets. We argue that these limitations can be addressed by using higher-quality datasets, where high quality refers to,

40 for example, removing watermarks or hiding photographic equipment through algorithms, as well as  
filtering out video clips containing human faces.



Mosaic effect



Watermarks



Photographic equipment

Figure 3: Caption

41

## 42 **D Potential Societal Impact**

43 Panoramic video generation has the potential to bring significant positive societal impact. It can  
44 empower non-expert users to create high-quality immersive content, thereby fostering innovation  
45 in digital art, virtual tourism, education, and cultural heritage preservation. This technology also  
46 enhance remote experiences and contributes to accessibility, enabling broader participation in virtual  
47 environments for individuals with physical or economic limitations. Furthermore, it offers valuable  
48 support in scientific visualization and educational applications, especially in fields such as geography  
49 and urban planning. However, generating virtual panoramas also pose several challenges. Concerns  
50 around privacy and ethics arise when real-world data containing personal identities are used for  
51 training. The potential for misuse, including the creation of deepfakes and misleading visual content,  
52 raises serious issues regarding misinformation and public trust. Additionally, copyright infringement  
53 and labor displacement in creative industries are growing concerns. To mitigate these risks, it  
54 is essential to implement robust regulatory frameworks, ethical guidelines, and public awareness  
55 campaigns that promote responsible use and transparency in AI-driven content generation.

56 **References**

- 57 [1] Hao Chen, Yuqi Hou, Chenyuan Qu, Irene Testini, Xiaohan Hong, and Jianbo Jiao. 360+x: A panoptic  
58 multi-modal scene understanding dataset. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2024.
- 59 [2] Qihua Chen, Yue Ma, Hongfa Wang, Junkun Yuan, Wenzhe Zhao, Qi Tian, Hongmei Wang, Shaobo Min,  
60 Qifeng Chen, and Wei Liu. Follow-your-canvas: Higher-resolution video outpainting with extensive content  
61 generation. *arXiv preprint arXiv:2409.01055*, 2024.
- 62 [3] Jing Tan, Shuai Yang, Tong Wu, Jingwen He, Yuwei Guo, Ziwei Liu, and Dahua Lin. Imagine360:  
63 Immersive 360 video generation from perspective anchor. *arXiv preprint arXiv:2412.03552*, 2024.
- 64 [4] Jionghao Wang, Ziyu Chen, Jun Ling, Rong Xie, and Li Song. 360-degree panorama generation from few  
65 unregistered nfov images. In *ACM Int. Conf. Multimedia*, pages 6811–6821, 2023.
- 66 [5] Qian Wang, Weiqi Li, Chong Mou, Xinhua Cheng, and Jian Zhang. 360dvd: Controllable panorama video  
67 generation with 360-degree video diffusion model. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages  
68 6913–6923, 2024.

69 **NeurIPS Paper Checklist**

70 **1. Claims**

71 Question: Do the main claims made in the abstract and introduction accurately reflect the  
72 paper’s contributions and scope?

73 Answer: [\[Yes\]](#)

74 Justification: The abstract and introduction clearly state the claims we made.

75 Guidelines:

- 76 • The answer NA means that the abstract and introduction do not include the claims  
77 made in the paper.
- 78 • The abstract and/or introduction should clearly state the claims made, including the  
79 contributions made in the paper and important assumptions and limitations. A No or  
80 NA answer to this question will not be perceived well by the reviewers.
- 81 • The claims made should match theoretical and experimental results, and reflect how  
82 much the results can be expected to generalize to other settings.
- 83 • It is fine to include aspirational goals as motivation as long as it is clear that these goals  
84 are not attained by the paper.

85 **2. Limitations**

86 Question: Does the paper discuss the limitations of the work performed by the authors?

87 Answer: [\[Yes\]](#)

88 Justification: We discuss the limitations in the supplemental material.

89 Guidelines:

- 90 • The answer NA means that the paper has no limitation while the answer No means that  
91 the paper has limitations, but those are not discussed in the paper.
- 92 • The authors are encouraged to create a separate "Limitations" section in their paper.
- 93 • The paper should point out any strong assumptions and how robust the results are to  
94 violations of these assumptions (e.g., independence assumptions, noiseless settings,  
95 model well-specification, asymptotic approximations only holding locally). The authors  
96 should reflect on how these assumptions might be violated in practice and what the  
97 implications would be.
- 98 • The authors should reflect on the scope of the claims made, e.g., if the approach was  
99 only tested on a few datasets or with a few runs. In general, empirical results often  
100 depend on implicit assumptions, which should be articulated.
- 101 • The authors should reflect on the factors that influence the performance of the approach.  
102 For example, a facial recognition algorithm may perform poorly when image resolution  
103 is low or images are taken in low lighting. Or a speech-to-text system might not be  
104 used reliably to provide closed captions for online lectures because it fails to handle  
105 technical jargon.
- 106 • The authors should discuss the computational efficiency of the proposed algorithms  
107 and how they scale with dataset size.
- 108 • If applicable, the authors should discuss possible limitations of their approach to  
109 address problems of privacy and fairness.
- 110 • While the authors might fear that complete honesty about limitations might be used by  
111 reviewers as grounds for rejection, a worse outcome might be that reviewers discover  
112 limitations that aren’t acknowledged in the paper. The authors should use their best  
113 judgment and recognize that individual actions in favor of transparency play an impor-  
114 tant role in developing norms that preserve the integrity of the community. Reviewers  
115 will be specifically instructed to not penalize honesty concerning limitations.

116 **3. Theory assumptions and proofs**

117 Question: For each theoretical result, does the paper provide the full set of assumptions and  
118 a complete (and correct) proof?

119 Answer: [\[NA\]](#)



Justification: This paper is not theoretical.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

#### 4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [\[Yes\]](#)

Justification: Ours results is reproducible, and we will release the codes.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

#### 5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: We will release the codes and weights, but not now as we are still in the process of organizing them.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

## 6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The paper specify all the training and test details.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

## 7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: The methodologies employed in this work involve computational models that yield consistent and repeatable outputs without variability under the same conditions. These tests are based on established simulations that deterministically produce the same results each time they are run, provided the input parameters remain unchanged.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

## 8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Sufficient information on the computer resources is provided in the paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

## 9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: We understand the importance of ethical standards in research and take this matter seriously.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

## 10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: We discuss the potential societal impacts of our work in the supplemental material.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.



- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

## 11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [\[Yes\]](#)

Justification: We commit to continuously monitoring the usage of our released models and codes and will take action to restrict access or provide additional guidance if we identify concerning patterns of misuse.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

## 12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [\[Yes\]](#)

Justification: All the assets we used have been properly cited.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, [paperswithcode.com/datasets](https://paperswithcode.com/datasets) has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

### 13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

### 14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

### 15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

#### 16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this research does not involve LLMs.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.