3D TISSUE RECONSTRUCTION AND GENERATION FOR SINGLE-CELL SPATIAL TRANSCRIPTOMICS USING NEURAL RADIANCE FIELDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Single-cell spatial transcriptomics (scST) is a groundbreaking technique that allows for the exploration of gene expression patterns, cell-cell interactions, and tissue organization at the single-cell level. Traditional approaches in scST reconstruction mainly focus on assigning two-dimensional (2D) coordinates to individual cells within a pre-established region. This often requires a large amount of 2D slice data, such as ssDNAs images, which escalates both costs and the complexity involved in studying and reconstructing the tissue's three-dimensional (3D) organization. Here, we introduce a novel method for scST reconstruction, which is a Neural Radiance Fields (NeRF)-based 3D-aware generative model termed STscan, that aims to reconstruct a 3D scST scene using a minimal amount from 2D images (fewer than 10). Additionally, STscan can identify cell types and their expression levels within this 3D environment. To the best of our knowledge, STscan is the first NeRF-based method specifically designed for single-cell ST reconstruction, and it is the first end-to-end solution capable of directly reconstructing in vitro cell-cell environments from ssDNA images. This approach has the potential to significantly reduce both the complexity and cost associated with scST studies.

1 INTRODUCTION

The growing scientific consensus posits that cellular spatial positioning has a profound influence on gene transcriptional activity, ultimately affecting physiological processes. Situated within complex three-dimensional microenvironments, cells occupy specific spatial coordinates and execute specialized functions(Li et al., 2022; Palla et al., 2022). For example, in the human brain, neurons in the hippocampus are spatially organized in a way that allows them to effectively process and store memories. This unique arrangement ensures that incoming signals are relayed through a specific network of neurons, optimizing the brain's ability to encode, store, and retrieve information(Piwecka et al., 2023). Advances in Single-Cell Spatial Transcriptomics (scST) technologies have opened new avenues for cellular observation(Longo et al., 2021). These technologies, which enable the simultaneous measurement of gene expression profiles at single-cell or even subcellular resolutions while preserving spatial information, were recognized as Nature Methods' Method of the Year for 2020(Marx, 2021).

Within the scope of scST technologies, methods can be divided into two primary categories: In Situ Hybridization (ISH) methods and Spatial Barcoding techniques. The raw data output of scST for each respective methodology amalgamates both imaging data (including, Staining Images are often referred to as ssDNA, Sequecial images) and localized sequencing information Figure 1, thereby facilitating the identification of cell types and aiding in the understanding of their expression patterns to gain insights into biological processes(Kleino et al., 2022; Dries et al., 2021; Tian et al., 2023).

While there have been rapid advances in technologies for handling standard dimensional reduction, clustering, and differential expression tools in spatial transcriptomic data, methods that effectively leverage the most crucial feature of profiling—space itself—have lagged far behind(Burgess, 2019; Bressan et al., 2023). To enable more general spatial profiling, some researchers employ a particular method: they produce serial thin sections from a biological sample, process each section through

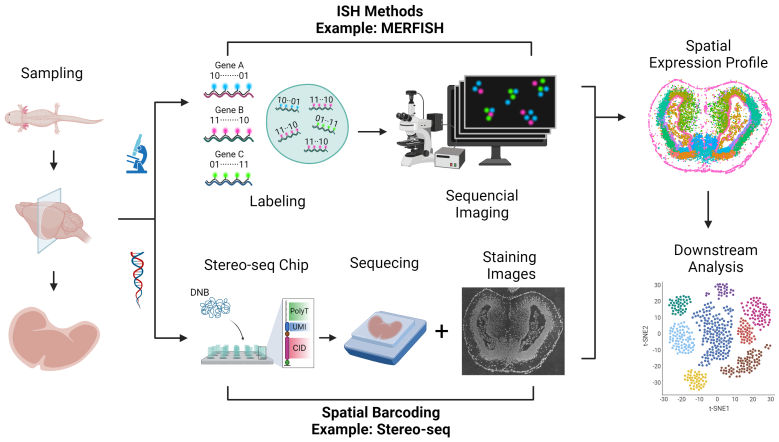

Figure 1: **The scope of scST methods.** After initial sample processing, both methods require the integration of sequencing with microscopic imaging for cellular data spacial extraction. The ISH approach often involves imaging sequencing results through fluorescent markers on the sequence. In contrast, the spatial barcoding technique presents a staining image (i.e., ssDNA) in the RNAscope procedures and decodes through 2D information. Ultimately, both methods aim to map cell types to their respective spatial expression levels, laying the foundation for downstream analysis.

2D imaging, and then use computational methods to realign the data and produce a 3D cube(Qiu et al., 2022; Xu et al., 2023). For example, Chen et al. (2023) used Stereo-Seq to reconstruct the monkey brain. They utilized 86 continuous slides in the monkey brain and recognized 143 macaque cortical regions. From this, they obtained a comprehensive atlas of 264 transcriptome-defined cortical cell types and mapped their spatial distribution across the entire cortex. They also discovered a relationship between the regional distribution of various cell types and the region's hierarchical level in the visual and somatosensory systems. However, this approach is very expensive and difficult to reproduce. Meanwhile, the loss of spatial information during the experimental process is an issue that stereo seq has yet to overcome.

With the goal of cost-effectively modeling the 3D structure of single-cell spatial transcriptomics (scST), our study employs neural radiance fields (NeRFs) (Mildenhall et al., 2021) as the 3D representation, termed the method **STscan**. NeRFs have recently made significant strides in view synthesis, they represent a 3D scene as a continuous radiation field parameterized by a neural network as inputs are coordinates and view directions. Despite showing promising results in common scenes and macroscopic objects, the problem of reconstructing microenvironments with NeRF remains largely unexplored. Therefore, we choose to apply them to the biomedical realms, specifically, we introduce a NeRF-based generator that renders and synthesizes novel views for single-cell spatial transcriptomics from a set of unposed ssDNA images.

Due to the data scarcity stemming from collecting ssDNA images through RNAscope procedures, we contend with a constrained quantity of training samples. To resolve this, we choose to jointly encode cell-type information and expression patterns along with appearance and geometry, which yield a joint distribution of ssDNA and corresponding semantics. This joint distribution affords us an enhanced understanding of the internal structure of scST and enables the modeling of the spatial distribution of cell type and expression, thereby assisting subsequent analyses of biological significance. By incorporating two discriminators with a differentiable data augmentation technique, we are able to synthesize high-resolution images while training our model solely on unposed 2D images. We systematically analyze our approach using raw data from the Stereo-Seq experiment(Wei et al., 2022). The experimental results demonstrate the efficacy of our model as a potent tool for scST image synthesis. Our model not only facilitates the reconstruction and generation of single-cell spatial transcriptomics (scST) from 2D images but also adds cell type and expression data simultaneously.

2 RELATED WORKS

Application Insights in ISH and Barcoding Techniques, in Situ Hybridization (ISH) represents a set of techniques specifically designed for tagging RNA molecules using fluorescent probes through complementary hybridization, which is subsequently visualized via fluorescence microscopy(Moses & Pachter, 2022). Among the available methodologies, three stand out based on their spatial resolution capabilities: MERFISH(Moffitt & Zhuang, 2016), seqFISH+(Eng et al., 2019), listed in descending order of resolution. Notably, MERFISH, a pioneering technique in this domain, was innovated by Professor Zhuang. Presently, it serves as a dominant technique in ISH applications, facilitating intricate studies on neurological structures(Zhang et al., 2021), oncological developments(Chen & Teichmann, 2021), and other related biomedical realms(Fang et al., 2022). In contrast, the realm of Spatial Barcoding techniques is characterized by the presence of unique barcoded DNA primers within each pixel, enabling the precise localization of a pixel in bidimensional representations. Noteworthily, this technique witnesses a broader application spectrum compared to ISH. Two of the paramount platforms in this field include 10x Visium(Galeano Niño et al., 2022) and BGI-Genomics's Stereo-Seq(Xia et al., 2022). However, given the limitation of 10x Visium in capturing single-cell granularity(Dong & Zhang, 2022), Stereo-Seq can capture in the subcellular emerging as a preferred choice, especially in contemporary research spanning genetics, oncology, and developmental biology(Koch, 2022). In light of these insights and the evolving research landscape, our current experiment harnesses data generated from the Stereo-Seq platform for reconstruction purposes.

Cell Type and Expression Profiler, in the realm of single-cell spatial transcriptomics, cell type identification, and cellular expression level analysis are two pivotal applications, laying the foundation for subsequent research and services(Cable et al., 2022). There are mainly two methods to estimate the cellular type composition: firstly, by assessing the enrichment level of cell type-specific markers in expressed genes, such as Leiden clustering(Traag et al., 2019); secondly, through deconvolution techniques aimed at precisely estimating the proportion of different cell types at each location, including SPOTlight(Elosua-Bayes et al., 2021) and DSTG(Song & Su, 2021). However, with the rapid advancement of spatial transcriptomics technology, the challenge of effectively integrating single-cell information with spatial transcriptome data has become increasingly prominent. Several groups are working on modeling spatial patterns of gene expression based on predefined processes(Bressan et al., 2023), like spatialDE(Svensson et al., 2018), whereas some methods mainly focus on spatial continuity. Incorporating continuous spatial expression information into spatial transcriptomic data also remains a challenge.

NeRF, short for Neural Radiance Field (Mildenhall et al., 2021), is a groundbreaking neural network architecture for 3D scene representation and reconstruction from 2D images. Since the inception of NeRF, many works have been proposed to enhance its quality and efficiency. For example, Yu et al. (2021); Xu et al. (2022a); Deng et al. (2022) aimed at diminishing the number of training views and enhancing generalization by utilizing image features or depth supervision. Barron et al. (2021; 2022) employ an integrated positional encoding of conical frustums to achieve anti-aliasing. Besides, some work combines NeRF with other tasks, such as Zhi et al. (2021); Fu et al. (2022) exploring incorporating semantic parsing with NeRF. Yariv et al. (2021); Oechsle et al. (2021); Wang et al. (2022); Li et al. (2023) have integrated NeRF with signal distance function and achieve both surface reconstruction and volume rendering using a single model. Furthermore, a series of studies have been conducted with the aim of augmenting the representation capabilities of NeRF through grid-based (Müller et al., 2022; Fridovich-Keil et al., 2022)and point-based (Xu et al., 2022b) architecture. Due to NeRF's generalizability, it has found applications across various domains, encompassing text-guided generation (Poole et al., 2022; Lin et al., 2023), human body modeling (Xu et al., 2021; Zhao et al., 2022), and even in the realm of medicine (Corona-Figueroa et al., 2022; Petkov, 2023). In this paper, we employ NeRF within the domain of microscopic medicine, specifically focusing on single-cell spatial transcriptomics. Our investigation is centered on the exploration of 3D reconstruction of scST with semantic information.

Generative Modeling endeavors to generate novel samples that manifest similar statistical properties to the training samples through learning the underlying distribution. During the early stage of deep learning, the variational autoencoder (VAE) (Kingma & Welling, 2013; Rezende et al., 2014; Kusner et al., 2017; Vahdat & Kautz, 2020) emerged as a popular generative model, which comprises an encoder network designed to map data into a latent space and a decoder network to reconstruct

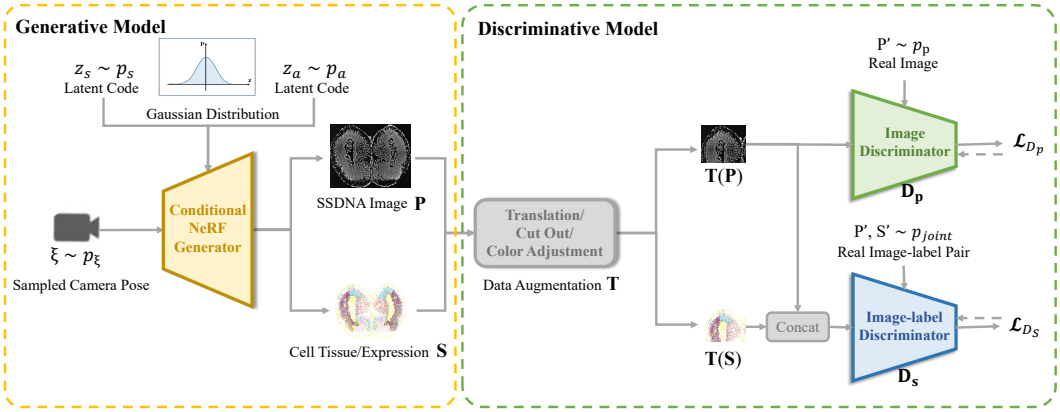

Figure 2: **The illustration of the overall network architecture.** Given a sampled camera pose ξ, the NeRF generator which is conditioned on two latent codes z_s, z_a synthesizes an ssDNA image and corresponding cell type or expression segmentation. The synthesized results are first transformed by a differentiable transformation T for data augmentation and then fed into two discriminators $\mathrm{D_p}$ and $\mathrm{D_s}$, and the gradients of loss $\mathcal{L}_{D_p}$ and $\mathcal{L}_{D_s}$ will be backpropagated into the generator for optimization.

the data. Later, Generative Adversarial Networks (GAN) (Goodfellow et al., 2014) have become the predominant framework for generative modeling. It consists of a generator and a discriminator, which are trained in a competitive manner and have been applied to various applications including image synthesis (Brock et al., 2018; Qin et al., 2020), video generation (Tulyakov et al., 2018; Chu et al., 2020), style transfer (Azadi et al., 2018; Karras et al., 2019). Recently, diffusion models (Sohl-Dickstein et al., 2015; Rombach et al., 2022) have been proposed and achieved great success in image synthesis, they generate new samples by gradually denoising a normally distributed variable. However, there is a problem with all the above methods, they cannot generate 3D consistent scenes due to the lack of 3D modeling. Since NeRF has become one of the most popular 3D representations, many methods have been proposed to combine 2D generative models with NeRF and generate 3D assets. For example, Poole et al. (2022); Lin et al. (2023) incorporate NeRF with diffusion model and generate 3D objects using text prompt, while Graf (Schwarz et al., 2020) achieves 3d-aware image synthesis by integrating NeRF into GAN-based framework. In this paper, we adopt a GAN-styled framework based on Graf (Schwarz et al., 2020) and aim to generate new single-cell spatial transcriptomics under limited training data, to resolve which we introduce a joint distribution of scST and semantic labels and a data augmentation technique.

3 METHOD

Conventional 3D reconstruction methods for single-cell spatial transcriptomics (scST) are not only costly but also, due to their technical limitations, unable to recover lost information in the spatial domain. In our study, we seek to address the challenges of scST reconstruction and generation by employing computer vision algorithms, offering a cost-effective solution that can facilitate further research in the field of scST. In particular, we introduce a NeRF-based generative model, **STscan** that consists of three components: a conditional NeRF generator for synthesizing ssDNA images along with their corresponding semantic information, two discriminators for gradient backpropagation, and a data augmentation technique to address data scarcity. The overall network architecture is illustrated in Figure 2.

3.1 DATASETS GENERATION

We conduct **STscan** experiments on the brain of *Axolotl Telencephalon* using Stereo-Seq(Wei et al., 2022). During the utilization of this data, we preprocessed the cell types and their expression patterns. Current alignment methods cannot directly deal with images, so we employed a novel approach for **STscan**.

Cell Type Images Generation. For our study, processed Stereo-seq data was downloaded from ARTISTA, which encompassed the segmented cell bin matrix, cell coordinates, and annotation metadata. Key elements such as cell coordinates and annotations were extracted from this dataset. Following extraction, the data was using the ggplot2(Wickham, 2011), which is a plotting library. Finally, spatial maps for each section were crafted to depict the spatial distribution of cell types.

Expression Pattern Images Generation. Expression values for each gene were then stratified based on peak expression values observed in each section, resulting in categorization into six distinct levels: 0, Low, Below Average, Average, Above Average, and High. Following this stratification process, the processed data utilize the ggplot2(Wickham, 2011), and spatial maps for each section were generated, illustrating the spatial expression patterns of the selected genes.

Affine Image Alignment. To align ssDNA images and the images of cell types or expression patterns, we employed an affine transformation based on the optimization of key point alignments. A 2D affine transformation is represented as:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{1}$$

Our goal was to determine the optimal transformation parameters a, b, c, d, e, f that minimize the Euclidean distance between transformed source points and target points. This optimization was achieved using the SGD gradient descent method(Qian et al., 2015), and the derived transformation was subsequently applied to the images.

3.2 Generative Model

We employ a NeRF as the representation for capturing the 3D structure of single-cell spatial transcriptomics inspired by Graf (Schwarz et al. (2020)). Besides, we leverage a discriminator D_ϕ. to provide feedback to the generator and improve the realism of the synthesized ssDNA images.

Conditional NeRF Generator. NeRF is a neural network-based approach that has shown great success in 3D scene representation. It is designed to capture complex 3D scenes by learning a continuous radiance field parameterized by a neural network function, denoted as G_θ, where θ represents the network's parameters. Specifically, given a 3D coordinate x and a viewing direction d:

$$\mathrm{G}_\theta(\mathrm{x}, \mathrm{d}) = (\mathrm{c}, \sigma), \tag{2}$$

where σ is a volume density and c is the corresponding RGB color value c. Let r(t) = o + td denote a camera ray, the expected color $\mathrm{C}(\mathrm{r})$ of the ray with near and far bounds t_n and t_f is:

$$\mathrm{C}(\mathrm{r}) = \int_{\mathrm{t_n}}^{\mathrm{t_f}} \mathrm{T}(\mathrm{t})\sigma(\mathrm{r}(\mathrm{t}))\mathrm{G}_\theta(\mathrm{r}(\mathrm{t}), \mathrm{d})\mathrm{dt}, \tag{3}$$

$$\mathrm{T}(\mathrm{t}) = \exp(-\int_{\mathrm{t_n}}^{\mathrm{t_f}} \sigma(\mathrm{r}(\mathrm{s}))\mathrm{ds}). \tag{4}$$

With a sampled camera pose $\xi \sim p_\xi$, G_θ can generate a corresponding image patch P. Subsequently, the discriminator D_ϕ is employed to evaluate the synthesized patch P in comparison to an authentic patch P' extracted from the training dataset. During the training process, a 2D sampling pattern is applied to produce image patches at a resolution of K ×K for computational efficiency.

To achieve controllable generation, two latent codes are introduced, one of which is to model shape, denoted as $z_s \sim p_s$, and the other is to model appearance, denoted as $z_a \sim p_a$. Both z_s and z_a are sampled from Gaussian distributions. In particular, z_s exerts control over shape by modulating the density σ, while z_a operates on appearance. The formulation of the conditional NeRF is:

$$\mathrm{x}, z_s \rightarrow \sigma, \tag{5}$$

$$\mathrm{x}, \mathrm{d}, z_s, z_a \rightarrow \mathrm{c}, \tag{6}$$

$$\mathrm{G}_\theta(\mathrm{x}, \mathrm{d}, z_s, z_a) = (\mathrm{c}, \sigma) \tag{7}$$

Semantic branch. In contrast to usual macroscopic objects, single-cell spatial transcriptomics is a collection of microstructures, each falling into distinct cell types and different expressions. The

conventional NeRF formulation, primarily designed for capturing appearance and geometry, cannot model the inherent property distribution of cells. To address this limitation, **STscan** model a joint distribution of ssDNA and semantic labels by introducing a semantic branch that predicts cell type or expression labels and the formulation can be expressed as:

$$f_s(\mathrm{x}, z_s, z_a) = s, \tag{8}$$

$$\mathrm{S(r)} = \int_{\mathrm{t_n}}^{\mathrm{t_f}} \mathrm{T(t)}\sigma(\mathrm{r(t)})\mathrm{f_s(r(t)},z_s, z_a)\mathrm{dt} \tag{9}$$

where s and $\mathrm{S(r)}$ are predicted semantic values, $\mathrm{f_s}$ is the branch for cell type or expression. The overall formulation for generative NeRF can be expressed as:

$$\mathrm{G}_\theta(\mathrm{x}, \mathrm{d}, z_s, z_a) = (\mathrm{c}, \mathrm{s}, \sigma) \tag{10}$$

The inclusion of semantic branches allows us to simultaneously acquire cell information during the synthesis of ssDNA images. This dual objective not only contributes to the reconstruction of ssDNA images but also improves the performance of generation by the acquisition of insights into cell type and expression distributions, which will be presented in the next section.

3.3 Discriminative Model

Utilizing the NeRF-based generator $\mathrm{G}(z_a, z_s) \rightarrow (\mathrm{P}, \mathrm{S})$, we achieve effective modeling of the joint distribution of ssDNA image patch P and the associated semantic attributes S, which encompass cell type or expression labels. These semantic labels inherently capture the intrinsic properties of ssDNA, therefore the joint distribution definitely enhances the generator's capacity to comprehend the 3D structure of ssDNA. In this section, we delve into a detailed exploration of strategies for leveraging the semantic attributes to provide feedback to the generator.

Data augmentation. As mentioned before, considering the difficulty of ssDNA data acquisition, **STscan** use a small set of training samples and focus on few-shot generation. However, the effectiveness of GAN is heavily dependent on the abundance of training data and the performance tends to deteriorate given a paucity of data. To resolve this, we introduce a data augmentation module following y DiffAugment (Zhao et al. (2020)). This module enhances data efficiency by applying a variety of differentiable augmentations to both authentic and synthetic samples. Specifically, we apply a random differentiable transformation T on both synthetic patch and real patch before feeding them into the discriminator, and T is a composition of three simple transformations including translation, cut out, and color adjustment. Building upon the data augmentation, we address the issue of inadequate training data while simultaneously mitigating concerns about discriminator overfitting.

Discriminator. We introduce two discriminators denoted as $\mathrm{D_P} : \mathrm{P} \rightarrow \mathbb{R}$ and $\mathrm{D_S} : \mathrm{concat(P, S)} \rightarrow \mathbb{R}$ respectively. Both discriminators are implemented with ResNet blocks. $\mathrm{D_P}$ plays the same role as the discriminator in most GAN models, it compares the predicted ssDNA patch P and the ground truth patch P', and the loss gradient is backpropagated to the generator thus facilitating the synthesis of more realistic images. p_p denotes distribution over image patches of trainset, $\mathrm{f_D}$ means hinge loss and the loss function for $\mathrm{D_P}$ with data augmentation T is:

$$\mathcal{L}_{\mathrm{D_P}} = \mathop{\mathbb{E}}_{\mathrm{P'\sim p_p}}[\mathrm{f_D}(-\mathrm{D_P}(\mathrm{T(P')}))] + \mathop{\mathbb{E}}_{\mathrm{P=G(z_a,z_s,\xi),z_a\sim p_a,z_s\sim p_s,\xi\sim p_\xi}}[\mathrm{f_D}(\mathrm{D_P}(\mathrm{T(P)}))] \tag{11}$$

To leverage the semantic information, we also implement a discriminator $\mathrm{D_S}$ for identifying the semantic labels. $\mathrm{D_S}$ is dedicated to optimizing image-label pairs, whose input is the concatenation of P and S. This naturally ensures alignment between synthesized images and corresponding semantic labels, as the non-aligned image-label pairs can be easily classified into fake. The loss function for $\mathrm{D_S}$ is as follows:

$$\mathcal{L}_{\mathrm{D_S}} = \mathop{\mathbb{E}}_{P', S'\sim p_{joint}}[\mathrm{f_D}(-\mathrm{D_S}(\mathrm{concat}(\mathrm{T(P')}, \mathrm{T(S')})))] \tag{12}$$

$$+ \mathop{\mathbb{E}}_{\mathrm{P,S=G(z_a,z_s,\xi),z_a\sim p_a,z_s\sim p_s,\xi\sim p_\xi}}[\mathrm{f_D}(\mathrm{D_S}(\mathrm{concat}(\mathrm{T(P)}, \mathrm{T(S)})))] \tag{13}$$

where p_{joint} is the joint distribution of ssDNA and semantic labels. In addition, to generate better semantic results, we compute the L1 loss with respect to feature maps produced by semantic labels

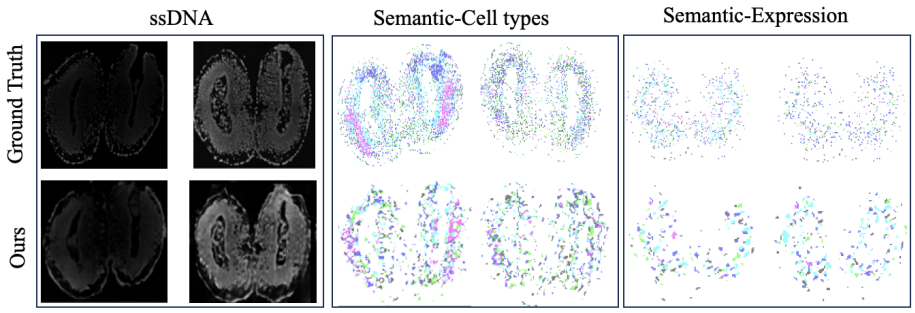

Figure 3: **Results of construct images.** The results of the 3D image generation process produced a set of images. These images were categorized into three distinct groups: ssDNA images, and two different sets of semantic images.

P and generated semantic P' in some hidden layers of $\mathrm{D_S}$, and the objective of the generator is formulated as:

$$\mathcal{L}_{\mathrm{G}} = \mathop{\mathbb{E}}_{\mathrm{P=G(z_a,z_s,\xi),z_a\sim p_a,z_s\sim p_s,\xi\sim p_\xi}} [\mathrm{f_e(T(D_P(P)))}] \tag{14}$$

$$+ \mathop{\mathbb{E}}_{\mathrm{P,S=G(z_a,z_s,\xi),z_a\sim p_a,z_s\sim p_s,\xi\sim p_\xi}} [\mathrm{f_e(T(D_S(concat(P,S))))}] \tag{15}$$

$$+ \sum_{i\in L} \mathrm{L_1(D_{S_i}(concat(P,S)) - D_{S_i}(concat(P',S')))} \tag{16}$$

Therefore, the total objective of the training procedure is defined as:

$$\mathcal{L} = \mathcal{L}_{\mathrm{G}} + \mathcal{L}_{\mathrm{D_S}} + \mathcal{L}_{\mathrm{D_P}} \tag{17}$$

During the inference process, we randomly sample z_s, z_a and camera pose ξ, and predict value and corresponding semantic information for all pixels in the image.

4 Experiment

4.1 Reconstructs The Brain of Axolotl Telencephalon From ssDNA images

We embarked on our investigation by applying the **STscan** to axolotl telencephalon datasetsWei et al. (2022), considering both paired cell type information and developmental expression patterns. Our primary findings, as summarized in Table 1, the ssDNA reconstruction relied on two crucial metrics: the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM), and the annotation including cell type and expression use mIoU metric. The generative loss function we employed showcased a notable perception-distortion trade-off in the renderings. Consistent alignment was observed between these renderings and both the cell type annotations and expression patterns, especially when viewed from continuous perspectives in comparison to the benchmark ground truth Figure3. For a more detailed evaluation, we noticed a consistency in the distance metrics associated with cell annotations and expressions. This consistency was further validated by comparing the cellular composition within pixels to their predecessor data. We specifically employed two metrics for this comparison: The Jaccard similarity coefficient is used to assess the similarity between cell types and their expression patterns in the original images compared to those in the reconstructed results, with a calculated result of 88.1%, and the Pearson Correlation Coefficient (PCC) is used to measure the similarity between cells in the reconstruction and their spatial distribution in the original images, with a calculated result of 90% Figure 6A. Both metrics manifested exemplary performance, suggesting that our model adeptly captures additional spatial information.

Semantic type	ssDNA PSNR	ssDNA SSIM	Semantic mIoU
Expression	25.0294	0.6972	0.8456
Cell type	27.3329	0.6115	0.7792

Table 1: Quantitative results on reconstruction.

Method	FID	KID
w/ DiffAug	82.4411	0.0544
w/o DiffAug	224.1103	0.1944
w/o Semantic	188.0024	0.1722
w/Cell type	81.2233	0.0609
w/Expression	79.1333	0.0588

Table 2: **Ablation study** on the effectiveness of semantic branch and data augmentation.

4.2 Ablation Study

To validate the effectiveness of our proposed model, we conducted two sets of ablation studies: one focusing on the semantic branch and the other on data augmentation. The quantitative results are presented in Table 2.

As depicted in the table, the absence of the semantic branch results in a noticeable increase in both FID and KID scores for the generated samples (lower FID and KID values are indicative of better performance). This observation underscores the significance of incorporating a joint distribution encompassing both image and semantic information in enhancing the synthesis of ssDNA.

Additionally, we conducted an ablation study on data augmentation by simply removing the transformation T. The omission of data augmentation led to a substantial increase in both FID and KID metrics. This can be attributed to the limited size of our training dataset. In the absence of data augmentation, there is an elevated risk of discriminator overfitting, which adversely affects the quality of the synthesized samples. The introduction of diffAugmentation effectively mitigated this issue, resulting in improved generation results. Furthermore, Figure4 provides qualitative results to further support the above analysis.

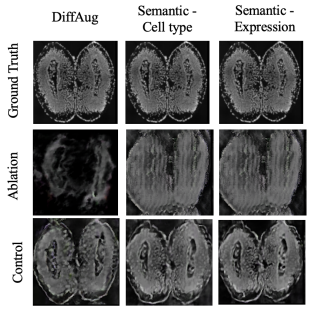

Figure 4: **The results of ablation experiments.** The control group consists of images generated by jointly encoding cell-type information and expression patterns, as well as applying data augmentation.

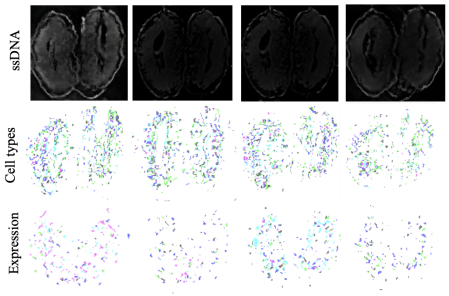

Figure 5: **Qualitative Results of Generative Reconstruction.** By leveraging the parameters z_s and z_a for inference control, STscan can generate a series of results, including both the ssDNA outcome and its corresponding cell type information or cellular expression.

Method	FID	KID
w/Expression	80.5758	0.0667
w/ Cell type	76.3329	0.0552

Table 3: Quantitative results on generation.

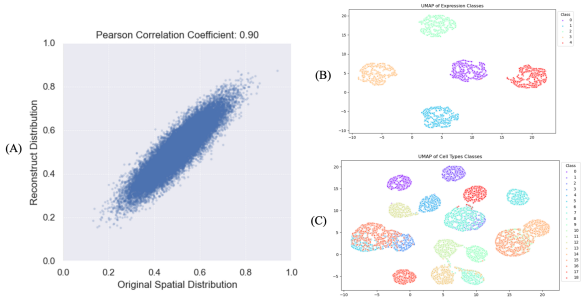

Figure 6: **The results of investigating biology meaning.** A represents PCC of cell distance between the reconstruction and 2D images. B and C summarize the UMAP results for cell type and expression. These results demonstrate a class type corresponding to the original classes.

4.3 Evaluating Generative Reconstruction Results

Our approach learns to disentangle continuous volumetric scenes which can be controlled via offering an inference control via parameters z_s and z_a. Leveraging these adjustments, we engaged in synthesizing novel views within the scST datasets, as elucidated in Table II. The evaluation, grounded on the FID and KID metrics, and Figure5 provides qualitative results that support it.

Furthermore, to benchmark the consistency of our generative framework, especially in the biological domain, we implemented a cellular clustering approach on the synthesized scene outputs. By using UMAP for clustering the generated views, we found that our generated data maintained a high consistency with previous data in terms of both cell type count and expression categories Figure 6B,C. This attests to the robustness and precision of our proposed model in the realm of generative tasks.

5 Conclusion

In this study, we introduce a generative model based on Neural Radiance Fields (NeRF), **STscan** aimed at learning a continuous 3D representation for ssDNA images. To obtain additional biological information in a three-dimensional environment, such as cell types and expression profiles, we incorporated a semantic component into the model. Experimental validation revealed that the model achieves favorable results in both qualitative and quantitative reconstructions. Notably, we are the first team to introduce NeRF in reconstructing biological environments from a limited number of single-cell spatial transcriptomes, paving a new avenue in the field of biology and reducing the complexity and cost associated with scST research.

References

Samaneh Azadi, Matthew Fisher, Vladimir G Kim, Zhaowen Wang, Eli Shechtman, and Trevor Darrell. Multi-content gan for few-shot font style transfer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7564–7573, 2018.

Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields.

In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5855–5864, 2021.

Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5470–5479, 2022.

Dario Bressan, Giorgia Battistoni, and Gregory J Hannon. The dawn of spatial omics. *Science*, 381 (6657):eabq4964, 2023.

Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.

Darren J Burgess. Spatial transcriptomics coming of age. *Nature Reviews Genetics*, 20(6):317–317, 2019.

Dylan M Cable, Evan Murray, Vignesh Shanmugam, Simon Zhang, Luli S Zou, Michael Diao, Haiqi Chen, Evan Z Macosko, Rafael A Irizarry, and Fei Chen. Cell type-specific inference of differential expression in spatial transcriptomics. *Nature methods*, 19(9):1076–1087, 2022.

Ao Chen, Yidi Sun, Ying Lei, Chao Li, Sha Liao, Juan Meng, Yiqin Bai, Zhen Liu, Zhifeng Liang, Zhiyong Zhu, et al. Single-cell spatial transcriptome reveals cell-type organization in the macaque cortex. *Cell*, 186(17):3726–3743, 2023.

Song Chen and Sarah A Teichmann. Completing the cancer jigsaw puzzle with single-cell multi-omics. *Nature Cancer*, 2(12):1260–1262, 2021.

Mengyu Chu, You Xie, Jonas Mayer, Laura Leal-Taixé, and Nils Thuerey. Learning temporal coherence via self-supervision for gan-based video generation. *ACM Transactions on Graphics (TOG)*, 39(4):75–1, 2020.

Abril Corona-Figueroa, Jonathan Frawley, Sam Bond-Taylor, Sarath Bethapudi, Hubert PH Shum, and Chris G Willcocks. Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single x-ray. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 3843–3848. IEEE, 2022.

Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. Depth-supervised nerf: Fewer views and faster training for free. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12882–12891, 2022.

Kangning Dong and Shihua Zhang. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature communications*, 13(1):1739, 2022.

Ruben Dries, Jiaji Chen, Natalie Del Rossi, Mohammed Muzamil Khan, Adriana Sistig, and Guo-Cheng Yuan. Advances in spatial transcriptomic data analysis. *Genome research*, 31(10):1706–1718, 2021.

Marc Elosua-Bayes, Paula Nieto, Elisabetta Mereu, Ivo Gut, and Holger Heyn. Spotlight: seeded nmf regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic acids research*, 49(9):e50–e50, 2021.

Chee-Huat Linus Eng, Michael Lawson, Qian Zhu, Ruben Dries, Noushin Koulena, Yodai Takei, Jina Yun, Christopher Cronin, Christoph Karp, Guo-Cheng Yuan, et al. Transcriptome-scale super-resolved imaging in tissues by rna seqfish+. *Nature*, 568(7751):235–239, 2019.

Rongxin Fang, Chenglong Xia, Jennie L Close, Meng Zhang, Jiang He, Zhengkai Huang, Aaron R Halpern, Brian Long, Jeremy A Miller, Ed S Lein, et al. Conservation and divergence of cortical cell organization in human and mouse revealed by merfish. *Science*, 377(6601):56–62, 2022.

Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5501–5510, 2022.

Xiao Fu, Shangzhan Zhang, Tianrun Chen, Yichong Lu, Lanyun Zhu, Xiaowei Zhou, Andreas Geiger, and Yiyi Liao. Panoptic nerf: 3d-to-2d label transfer for panoptic urban scene segmentation. In *2022 International Conference on 3D Vision (3DV)*, pp. 1–11. IEEE, 2022.

Jorge Luis Galeano Niño, Hanrui Wu, Kaitlyn D LaCourse, Andrew G Kempchinsky, Alexander Baryiames, Brittany Barber, Neal Futran, Jeffrey Houlton, Cassie Sather, Ewa Sicinska, et al. Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature*, 611 (7937):810–817, 2022.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410, 2019.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Iivari Kleino, Paulina Frolovaitė, Tomi Suomi, and Laura L Elo. Computational solutions for spatial transcriptomics. *Computational and structural biotechnology journal*, 2022.

Linda Koch. A panoramic view of mouse organogenesis. *Nature Reviews Genetics*, 23(7):393–393, 2022.

Matt J Kusner, Brooks Paige, and José Miguel Hernández-Lobato. Grammar variational autoencoder. In *International conference on machine learning*, pp. 1945–1954. PMLR, 2017.

Bin Li, Wen Zhang, Chuang Guo, Hao Xu, Longfei Li, Minghao Fang, Yinlei Hu, Xinye Zhang, Xinfeng Yao, Meifang Tang, et al. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nature methods*, 19(6):662–670, 2022.

Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. Neuralangelo: High-fidelity neural surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8456–8465, 2023.

Chen-Hsuan Lin, Jun Gao, Luming Tang, Towaki Takikawa, Xiaohui Zeng, Xun Huang, Karsten Kreis, Sanja Fidler, Ming-Yu Liu, and Tsung-Yi Lin. Magic3d: High-resolution text-to-3d content creation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 300–309, 2023.

Sophia K Longo, Margaret G Guo, Andrew L Ji, and Paul A Khavari. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nature Reviews Genetics*, 22 (10):627–644, 2021.

Vivien Marx. Method of the year: spatially resolved transcriptomics. *Nature methods*, 18(1):9–14, 2021.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

Jeffrey R Moffitt and Xiaowei Zhuang. Rna imaging with multiplexed error-robust fluorescence in situ hybridization (merfish). In *Methods in enzymology*, volume 572, pp. 1–49. Elsevier, 2016.

Lambda Moses and Lior Pachter. Museum of spatial transcriptomics. *Nature Methods*, 19(5):534–546, 2022.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022.

Michael Oechsle, Songyou Peng, and Andreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5589–5599, 2021.

Giovanni Palla, David S Fischer, Aviv Regev, and Fabian J Theis. Spatial components of molecular tissue biology. *Nature Biotechnology*, 40(3):308–318, 2022.

Kaloian Petkov. Guided training of nerfs for medical volume rendering. In *ACM SIGGRAPH 2023 Posters*, pp. 1–2. 2023.

Monika Piwecka, Nikolaus Rajewsky, and Agnieszka Rybak-Wolf. Single-cell and spatial transcriptomics: deciphering brain complexity in health and disease. *Nature Reviews Neurology*, pp. 1–17, 2023.

Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. *arXiv preprint arXiv:2209.14988*, 2022.

Qi Qian, Rong Jin, Jinfeng Yi, Lijun Zhang, and Shenghuo Zhu. Efficient distance metric learning by adaptive sampling and mini-batch stochastic gradient descent (sgd). *Machine Learning*, 99: 353–372, 2015.

Zhiwei Qin, Zhao Liu, Ping Zhu, and Yongbo Xue. A gan-based image synthesis method for skin lesion classification. *Computer Methods and Programs in Biomedicine*, 195:105568, 2020.

Xiaojie Qiu, Daniel Y Zhu, Jiajun Yao, Zehua Jing, Lulu Zuo, Mingyue Wang, Kyung Hoi Min, Hailin Pan, Shuai Wang, Sha Liao, et al. Spateo: multidimensional spatiotemporal modeling of single-cell spatial transcriptomics. *BioRxiv*, pp. 2022–12, 2022.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pp. 1278–1286. PMLR, 2014.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Katja Schwarz, Yiyi Liao, Michael Niemeyer, and Andreas Geiger. Graf: Generative radiance fields for 3d-aware image synthesis. *Advances in Neural Information Processing Systems*, 33:20154–20166, 2020.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. PMLR, 2015.

Qianqian Song and Jing Su. Dstg: deconvoluting spatial transcriptomics data through graph-based artificial intelligence. *Briefings in bioinformatics*, 22(5):bbaa414, 2021.

Valentine Svensson, Sarah A Teichmann, and Oliver Stegle. Spatialde: identification of spatially variable genes. *Nature methods*, 15(5):343–346, 2018.

Luyi Tian, Fei Chen, and Evan Z Macosko. The expanding vistas of spatial transcriptomics. *Nature Biotechnology*, 41(6):773–782, 2023.

Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):5233, 2019.

Sergey Tulyakov, Ming-Yu Liu, Xiaodong Yang, and Jan Kautz. Mocogan: Decomposing motion and content for video generation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1526–1535, 2018.

Arash Vahdat and Jan Kautz. Nvae: A deep hierarchical variational autoencoder. *Advances in neural information processing systems*, 33:19667–19679, 2020.

Jiepeng Wang, Peng Wang, Xiaoxiao Long, Christian Theobalt, Taku Komura, Lingjie Liu, and Wenping Wang. Neuris: Neural reconstruction of indoor scenes using normal priors. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII*, pp. 139–155. Springer, 2022.

Xiaoyu Wei, Sulei Fu, Hanbo Li, Yang Liu, Shuai Wang, Weimin Feng, Yunzhi Yang, Xiawei Liu, Yan-Yun Zeng, Mengnan Cheng, et al. Single-cell stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science*, 377(6610):eabp9444, 2022.

Hadley Wickham. ggplot2. *Wiley interdisciplinary reviews: computational statistics*, 3(2):180–185, 2011.

Keke Xia, Hai-Xi Sun, Jie Li, Jiming Li, Yu Zhao, Lichuan Chen, Chao Qin, Ruiying Chen, Zhiyong Chen, Guangyu Liu, et al. The single-cell stereo-seq reveals region-specific cell subtypes and transcriptome profiling in arabidopsis leaves. *Developmental cell*, 57(10):1299–1310, 2022.

Dejia Xu, Yifan Jiang, Peihao Wang, Zhiwen Fan, Humphrey Shi, and Zhangyang Wang. Sinnerf: Training neural radiance fields on complex scenes from a single image. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXII*, pp. 736–753. Springer, 2022a.

Hao Xu, Jun Lin, Shuyan Wang, Minghao Fang, Songwen Luo, Chunpeng Chen, Siyuan Wan, Rirui Wang, Meifang Tang, Tian Xue, et al. Spacel: characterizing spatial transcriptome architectures by deep-learning. 2023.

Hongyi Xu, Thiemo Alldieck, and Cristian Sminchisescu. H-nerf: Neural radiance fields for rendering and temporal reconstruction of humans in motion. *Advances in Neural Information Processing Systems*, 34:14955–14966, 2021.

Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. Point-nerf: Point-based neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5438–5448, 2022b.

Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34:4805–4815, 2021.

Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4578–4587, 2021.

Meng Zhang, Stephen W Eichhorn, Brian Zingg, Zizhen Yao, Kaelan Cotter, Hongkui Zeng, Hongwei Dong, and Xiaowei Zhuang. Spatially resolved cell atlas of the mouse primary motor cortex by merfish. *Nature*, 598(7879):137–143, 2021.

Fuqiang Zhao, Wei Yang, Jiakai Zhang, Pei Lin, Yingliang Zhang, Jingyi Yu, and Lan Xu. Humannerf: Efficiently generated human radiance field from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7743–7753, 2022.

Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. Differentiable augmentation for data-efficient gan training. *Advances in neural information processing systems*, 33:7559–7570, 2020.

Shuaifeng Zhi, Tristan Laidlow, Stefan Leutenegger, and Andrew J Davison. In-place scene labelling and understanding with implicit scene representation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 15838–15847, 2021.

A Appendix

You may include other additional sections here.