

# 10/07-10/30 Research Track

Jintong Yang

In this group research project, I was responsible for creating data visualizations of the cross-attention maps in Stable Diffusion.

10/07-10/14

## a. Learning the Basics of Stable Diffusion

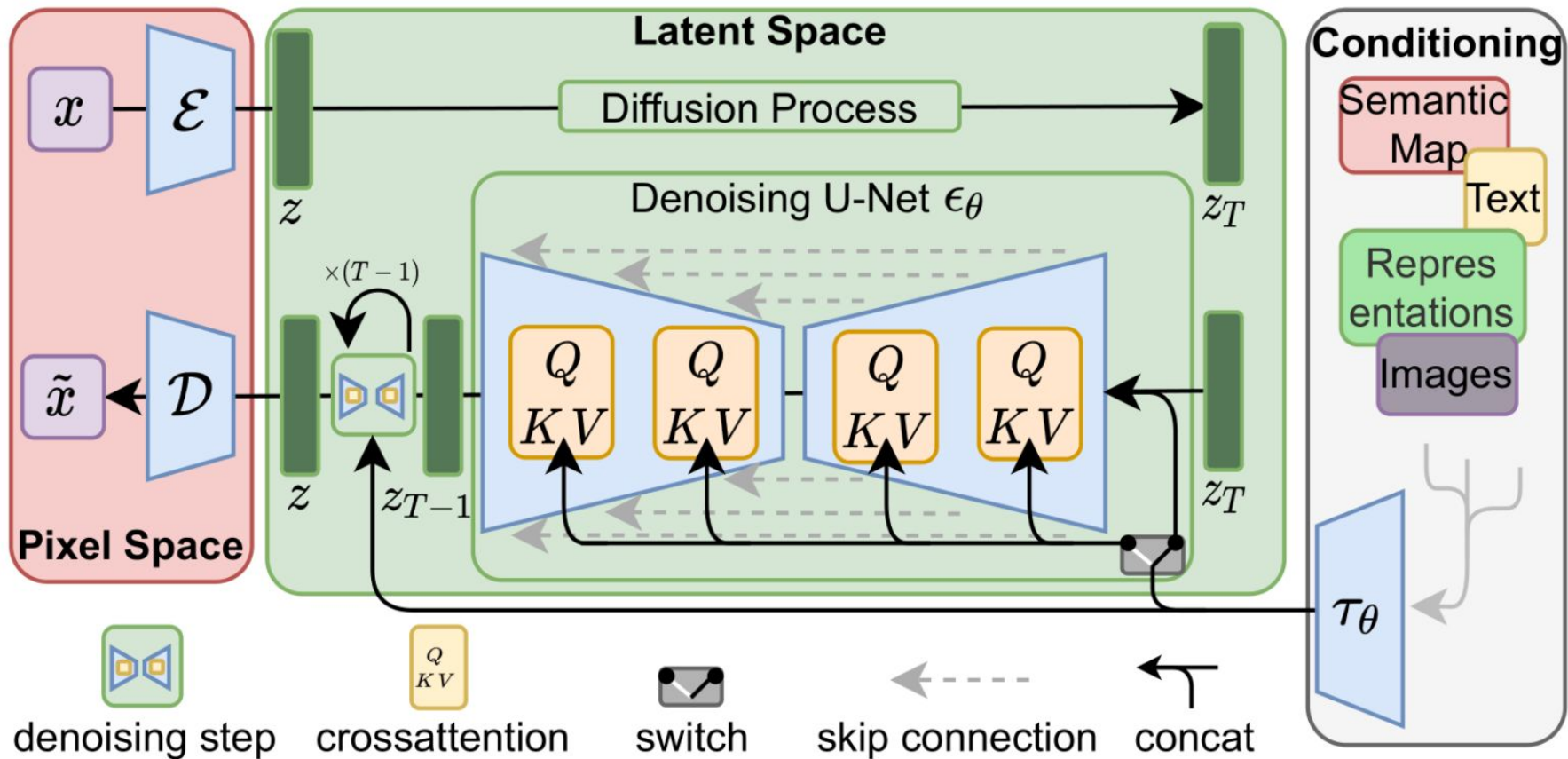
- **Observation:** Stable Diffusion was not originally designed for image generation.
- **Generation Process of Stable Diffusion:**  
Pixel Space → **Encoding (VAE Encoder)** →  
**Latent Space** with Gaussian Sampling (a noise canvas) →  
**Reverse Diffusion** (iterative denoising process) →  
**U-Net** (spatial features) + **Cross-Attention** (semantic features) →  
**Decoding (VAE Decoder)** back to Pixel Space.
- The generative process like a black box because **cross-attention weights change at every step** of the denoising process.

10/07-10/14

a. The basics of Stable Diffusion

b. Literature Review (two key papers)

- ***What the DAAM: Interpreting Stable Diffusion Using Cross Attention (ACL 2023)***
  - DAAM = Diffusion Attentive Attribution Maps
  - Visualizes word-to-pixel links for nouns, verbs, adjectives
  - Cohyponyms (e.g., *giraffe*, *zebra*) show strong overlap
- ***Attend-and-Excite: Attention-Based Semantic Guidance for Text-to-Image Diffusion Models (TOG 2023)***
  - limitations of text-to-image models: *catastrophic neglect* and *attribute binding errors* – *Generative Semantic Nursing*
  - **Attend-and-Excite** to make models attend to all subjects without retraining
  - Works effectively for **multi-object, color, and attribute prompts.**



**10/07-10/14**

- a. The basics of Stable Diffusion**
- b. Literature Review (two key papers)**
- c. Analysis of attention map data structure**

# Data Structure – 4D tensor

- apple\_step\_1
- apple\_step\_2
- apple\_step\_3
- apple\_step\_4
- apple\_step\_5
- apple\_step\_6
- apple\_step\_7
- apple\_step\_8
- apple\_step\_9
- apple\_step\_10
- apple\_step\_11
- apple\_step\_12
- apple\_step\_13
- apple\_step\_14
- apple\_step\_15
- apple\_step\_16
- apple\_step\_17
- apple\_step\_18
- apple\_step\_19
- apple\_step\_20
- apple\_step\_21
- apple\_step\_22
- apple\_step\_23
- apple\_step\_24
- apple\_step\_25
- apple\_step\_26
- apple\_step\_27
- apple\_step\_28
- apple\_step\_29
- apple\_step\_30



attention weights [rows]

	A	B	C	D	E	F	G	H	I	J	K
1	0.00548	0.00548	0.00548	0.00548	0.00556	0.00573	0.00589	0.00606	0.00622	0.00639	0.00655
2	0.00548	0.00548	0.00548	0.00548	0.00556	0.00573	0.00589	0.00606	0.00622	0.00639	0.00655
3	0.00548	0.00548	0.00548	0.00548	0.00556	0.00573	0.00589	0.00606	0.00622	0.00639	0.00655
4	0.00548	0.00548	0.00548	0.00548	0.00556	0.00573	0.00589	0.00606	0.00622	0.00639	0.00655
5	0.00545	0.00545	0.00545	0.00545	0.00553	0.00568	0.00583	0.00599	0.00614	0.00629	0.00645
6	0.00539	0.00539	0.00539	0.00539	0.00545	0.00558	0.00571	0.00584	0.00597	0.0061	0.00623
7	0.00533	0.00533	0.00533	0.00533	0.00538	0.00549	0.00559	0.0057	0.0058	0.00591	0.00601
8	0.00527	0.00527	0.00527	0.00527	0.00531	0.00539	0.00547	0.00555	0.00563	0.00571	0.00579
9	0.00521	0.00521	0.00521	0.00521	0.00523	0.00529	0.00535	0.00541	0.00546	0.00552	0.00558
10	0.00514	0.00514	0.00514	0.00514	0.00516	0.00519	0.00523	0.00526	0.00529	0.00533	0.00536
11	0.00508	0.00508	0.00508	0.00508	0.00509	0.0051	0.00511	0.00511	0.00512	0.00513	0.00514
12	0.00502	0.00502	0.00502	0.00502	0.00501	0.005	0.00498	0.00497	0.00495	0.00494	0.00492
13	0.00501	0.00501	0.00501	0.00501	0.005	0.00497	0.00495	0.00492	0.0049	0.00487	0.00485
14	0.00505	0.00505	0.00505	0.00505	0.00504	0.00502	0.005	0.00498	0.00496	0.00494	0.00492
15	0.00509	0.00509	0.00509	0.00509	0.00508	0.00506	0.00505	0.00503	0.00501	0.005	0.00498
16	0.00513	0.00513	0.00513	0.00513	0.00512	0.00511	0.0051	0.00508	0.00507	0.00506	0.00505
17	0.00517	0.00517	0.00517	0.00517	0.00516	0.00515	0.00515	0.00514	0.00513	0.00512	0.00511
18	0.00521	0.00521	0.00521	0.00521	0.0052	0.0052	0.0052	0.00519	0.00519	0.00518	0.00518
19	0.00525	0.00525	0.00525	0.00525	0.00525	0.00525	0.00525	0.00525	0.00525	0.00525	0.00525
20	0.00529	0.00529	0.00529	0.00529	0.00529	0.00529	0.0053	0.0053	0.0053	0.00531	0.00531

steps

tokens

attention weights [columns]

- apple
- sunflower
- table
- vase

10/07-10/14

- a. The basics of Stable Diffusion**
- b. Literature Review (two key papers)**
- c. Analysis of attention map data structure**
- d. Data visualization research questions**
  - How to visualize attention maps in an artistic way?
  - How can visualization explain why input words influence image locations?
- e. Inspiration for visualizing attention maps**
  - From noise to image — from chaos to order — a “universe”

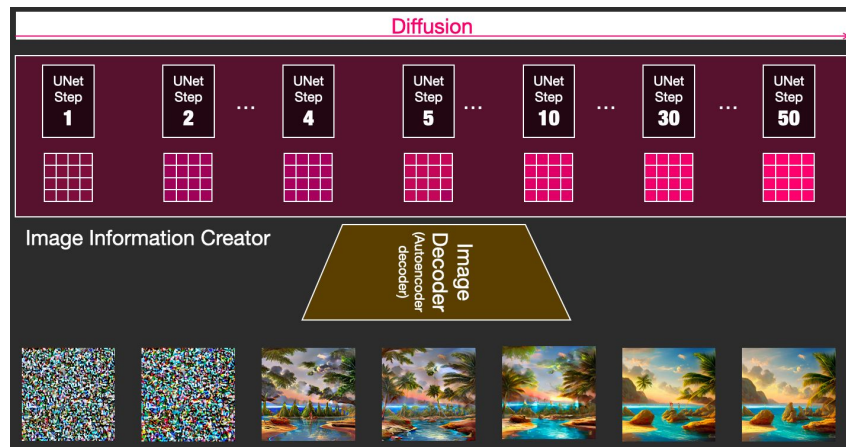
10/15-10/22

a. Propose data visualization possibilities & experiments

■ “Flow Field Diffusion”

To echo the diffusion generative process:

- Particle systems
- A flowing energy field
- Data mapping targets:
  - Position, size, rotation
  - Velocity
  - Lifespan
  - Brightness or hue
- Each token as a “force”

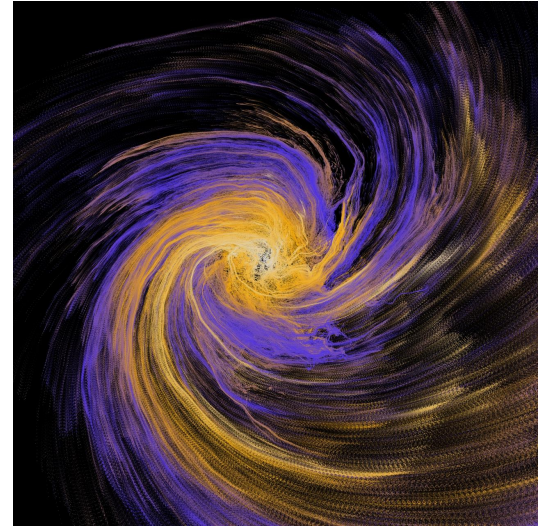
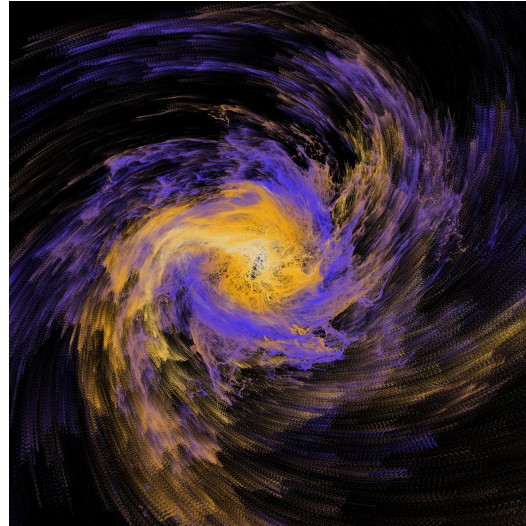
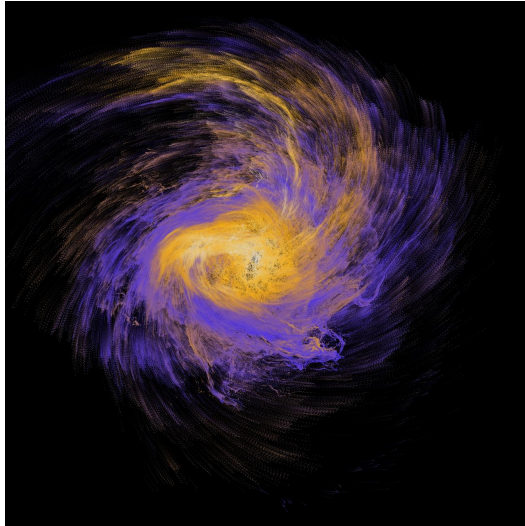


10/15-10/22

a. Propose data visualization possibilities & experiments

■ **“Flow Field Diffusion”**

Experiment:



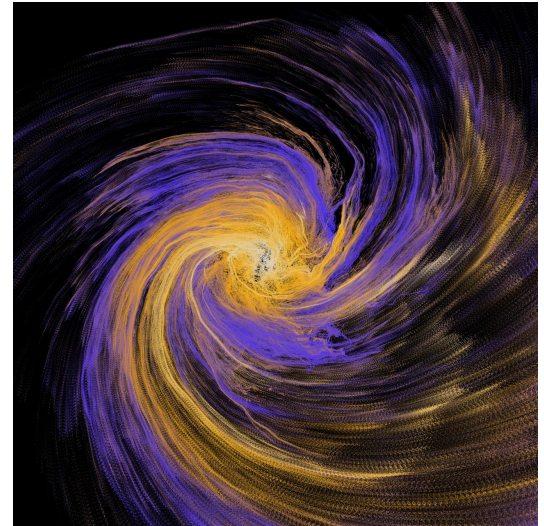
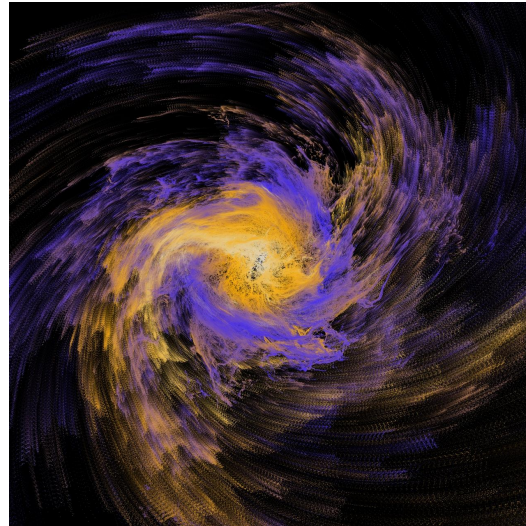
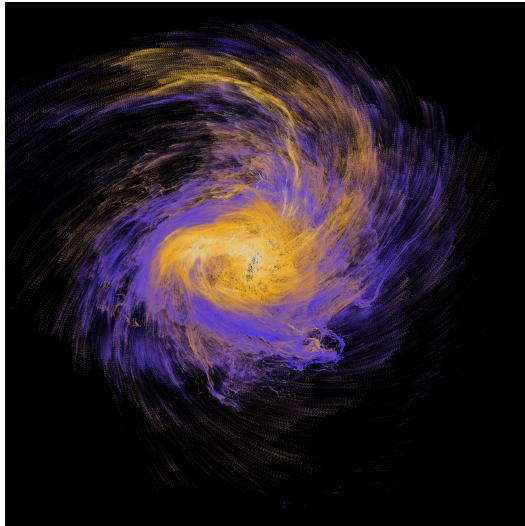
10/15-10/22

a. Propose data visualization possibilities & experiments

■ “Flow Field Diffusion”

Problems:

- Large datasets caused lag in 3D particle rendering
- Visualization too abstract: hard to see clear data mapping



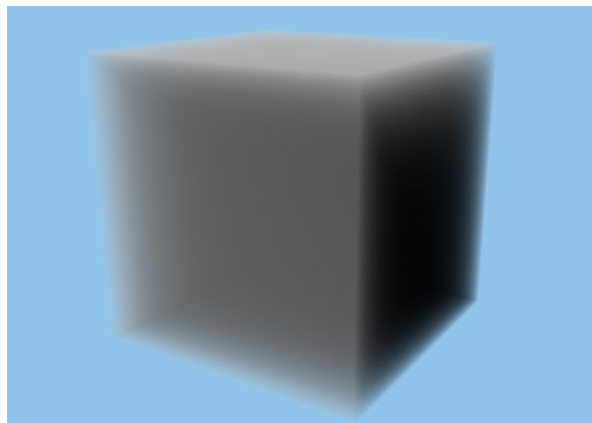
10/15-10/22

a. Propose data visualization possibilities & experiments

■ “Flow Field Diffusion”

■ “Interactive Latent Observatory”

- A 3D volume cloud ( $512 \times 512 \times 512$ )
- Time (Steps) – Space (attention weight pixels) – Semantics (token)
- By clicking each token, its corresponding region in the cube emits light step by step



10/15-10/22

a. **Propose data visualization possibilities & experiments**

■ **“Flow Field Diffusion”**

■ **“Interactive Latent Observatory”**

**Two technical directions:**

● **Touchdesigner 3D Instancing**

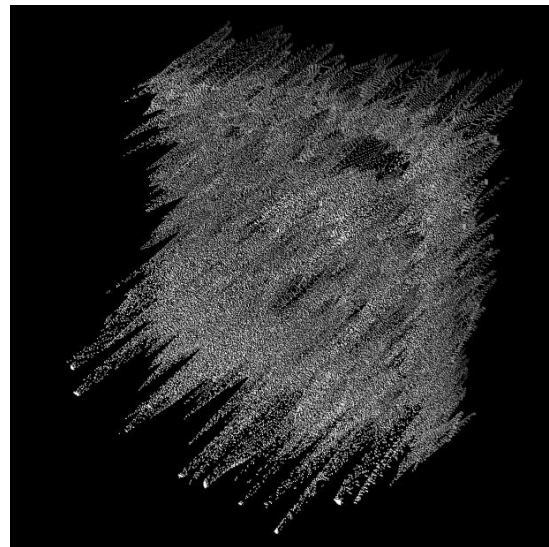
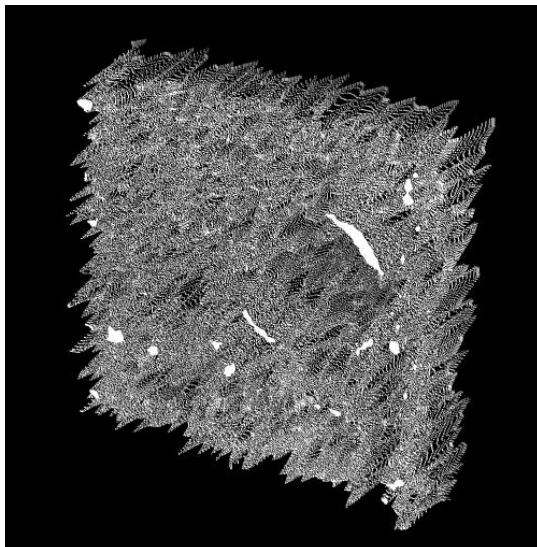
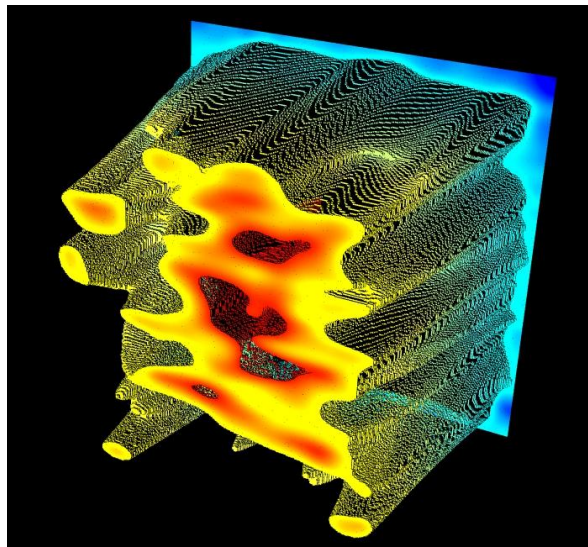
- GLSL
- Point cloud

● **3D Gaussian Splatting**

- <https://github.com/graphdeco-inria/gaussian-splatting>
- <https://github.com/MrNeRF/awesome-3D-gaussian-splatting>
- <https://github.com/hustvl/4DGaussians>

10/23-10/30

- a. Learning GLSL & Point Cloud in Touchdesigner
- b. Experiment with the “Interactive Latent Observatory” solution



# 10/31-11/27 Research Plan

<b>10/31-11/06</b>	Experiment with Shaw's sketch
<b>11/07-11/13</b>	Learn GLSL & 3D Point Cloud
	Literature review for data visualization solutions
	Refine the test
<b>11-14-11/20</b>	Learn GLSL & 3D Point Cloud
	Literature review for data visualization solutions
	Refine test with improved visual effects (possibly using 3D software)
<b>11/21-11/27</b>	Learn GLSL & 3D Point Cloud
	Refine test using real-time input attention map