Muddy underwater Image Generation

Prepared by BackPropagators

Shashank Aggarwal - 244101050
Ashwin Iyengar - 244101007
Soumajit Roy - 244101057
Shwetank Tiwari -244101055
Shiva Ram - 244101053

Content

OUTLINE OF TOPICS

Problem statement
Methodology
- Diffusion Model
- LoRA

Proposed Solution
Results
- Qualitative
- Quantitative

Limitations
Future Scope

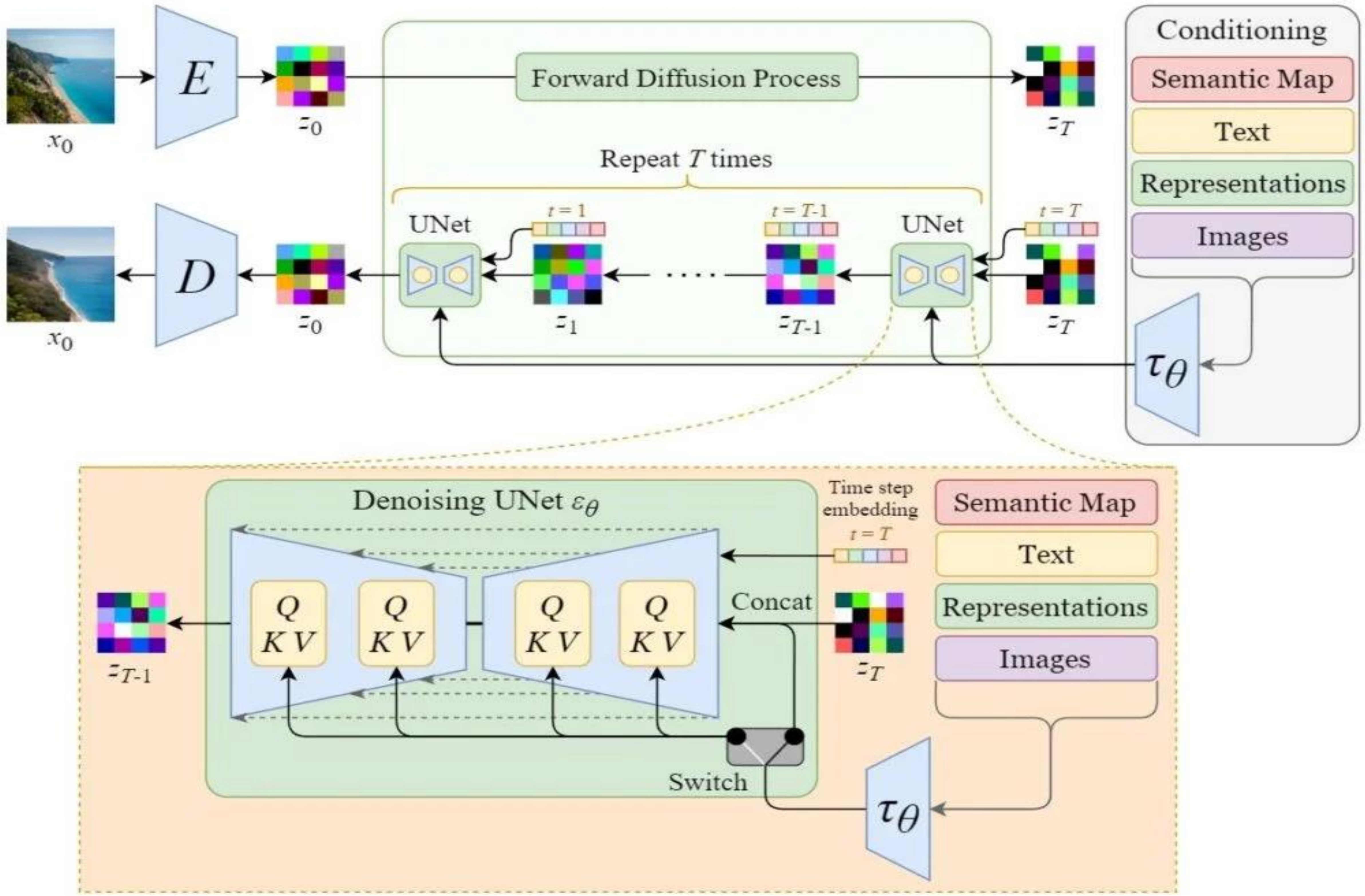
PROBLEM STATEMENT

Generation of synthetic underwater images for turbid environment.

- The main challenge in obtaining underwater images is the high cost, as it demands specialized equipment, deep-sea operations, and ongoing system maintenance.
- Limited dataset access hinders the development of accurate, robust models involving enhancement of underwater images for Indian rivers.

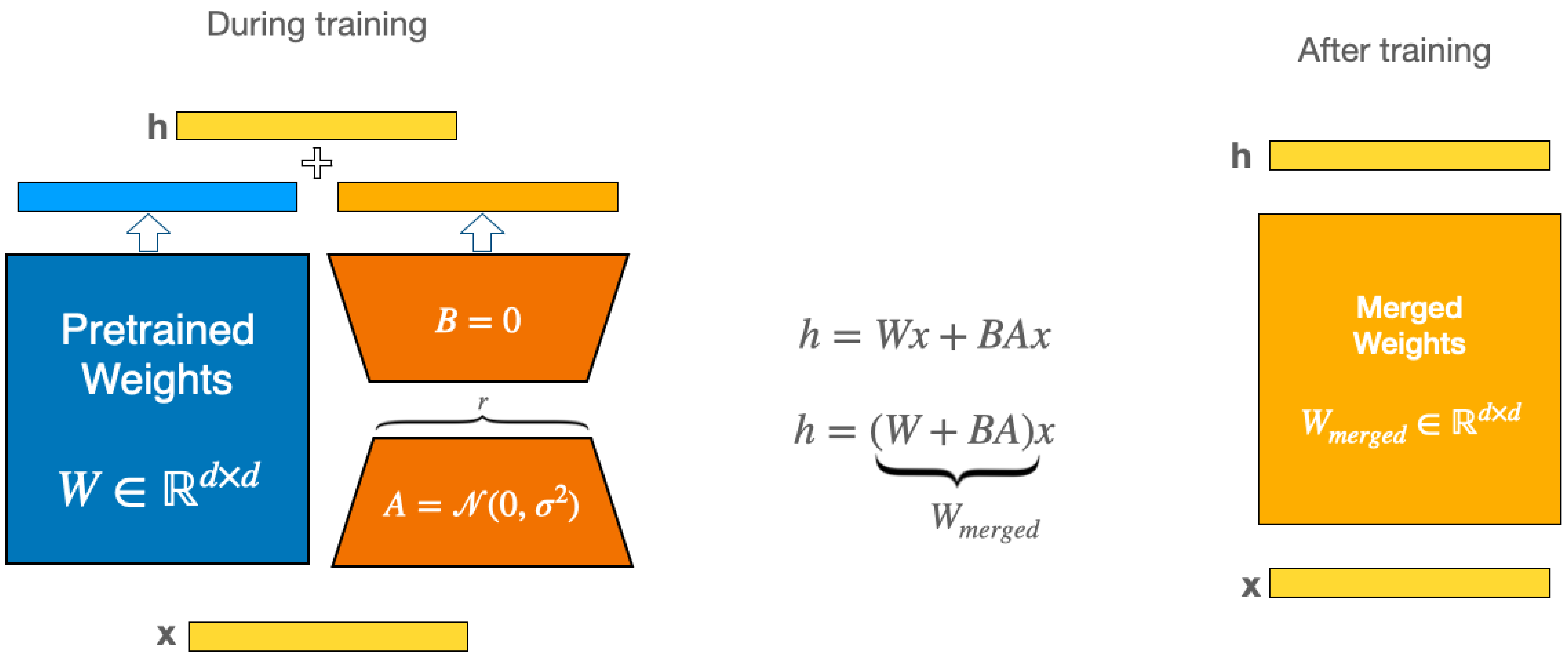
Stable Diffusion

- The model is based on a diffusion process, where an image is gradually corrupted by noise, starting from a clean image and progressively adding random noise at each step.
- The idea is that, through this process, the model learns to reverse the diffusion process, i.e., to go from noisy images back to clean, coherent images. During training, it learns how to "denoise" images step-by-step to recover original data.



LoRA

- LoRA (Low-Rank Adaptation) is a technique in deep learning used to efficiently fine-tune large models.
- It works by reducing the rank of weight matrices during training, which means fewer parameters need to be updated, significantly lowering computational costs.
- Instead of modifying the entire model, LoRA introduces low-rank matrices to capture essential updates, allowing the core model to remain largely unchanged.
- This makes LoRA ideal for fine-tuning on limited hardware or when adapting models to specific tasks without extensive resources.



Proposed Solution

Dataset

muddy underwater style, green underwater with diver in the background surrounded by scattered aquatic plants

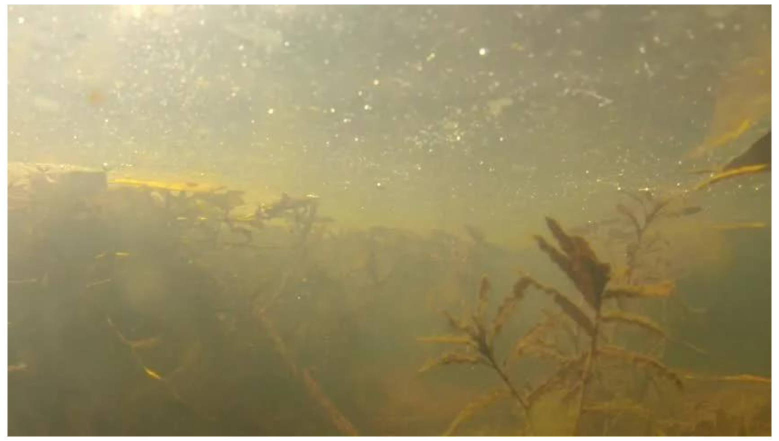
muddy underwater style, seabed with scattered rocks and coral formations.

muddy underwater style, low visibility diffused sunlight across scattered rocks and aquatic plants.

muddy underwater style, sandy seabed with aquatic plants

muddy underwater style, soft green tint underwater background.

muddy underwater style, A sandy and debris covered rocky bottom with fish swimming underwater.



muddy underwater style, a yellowish tint and underwater scene with an aquatic plant and rocks in the background

muddy underwater style, shark with green underwater background.

muddy underwater style, Fish swimming underwater with a yellowish background.

muddy underwater style, school of fish swimming over a rocky bed covered in algae and aquatic plants.

Proposed Solution

Finetune SDXL 1.0 Base (Method 1) – Single Prompt

- All the images were trained on a single prompt "muddy underwater style"
- Epochs: 500
- Learning Rate: 1e-4
- Image size: 1024 x 1024
- Inference Steps: 50
- Training Time: 3 hrs

Finetune SDXL 1.0 Base (Method 2) – Multiple Prompts

- All the images were trained on there individual prompts which includes "muddy underwater style" along with additional descriptions. To make dataset large we duplicated each image 20 times.
- Epochs: 10
- Learning Rate: 0.75
- Batch size: 4
- Optimizer: Prodigy
- Training Time: 3 hrs

Proposed Solution

Training our custom defined diffusion model based on SD 1.5 (Method 3)

- All the images were trained on a single prompt "muddy underwater style"
- While fine tuning the diffusion parameters were frozen
- Epochs: 15
- Learning Rate: 1e-4
- Image size: 512 x 512
- Inference Steps: 80
- Training Time: 2 hrs

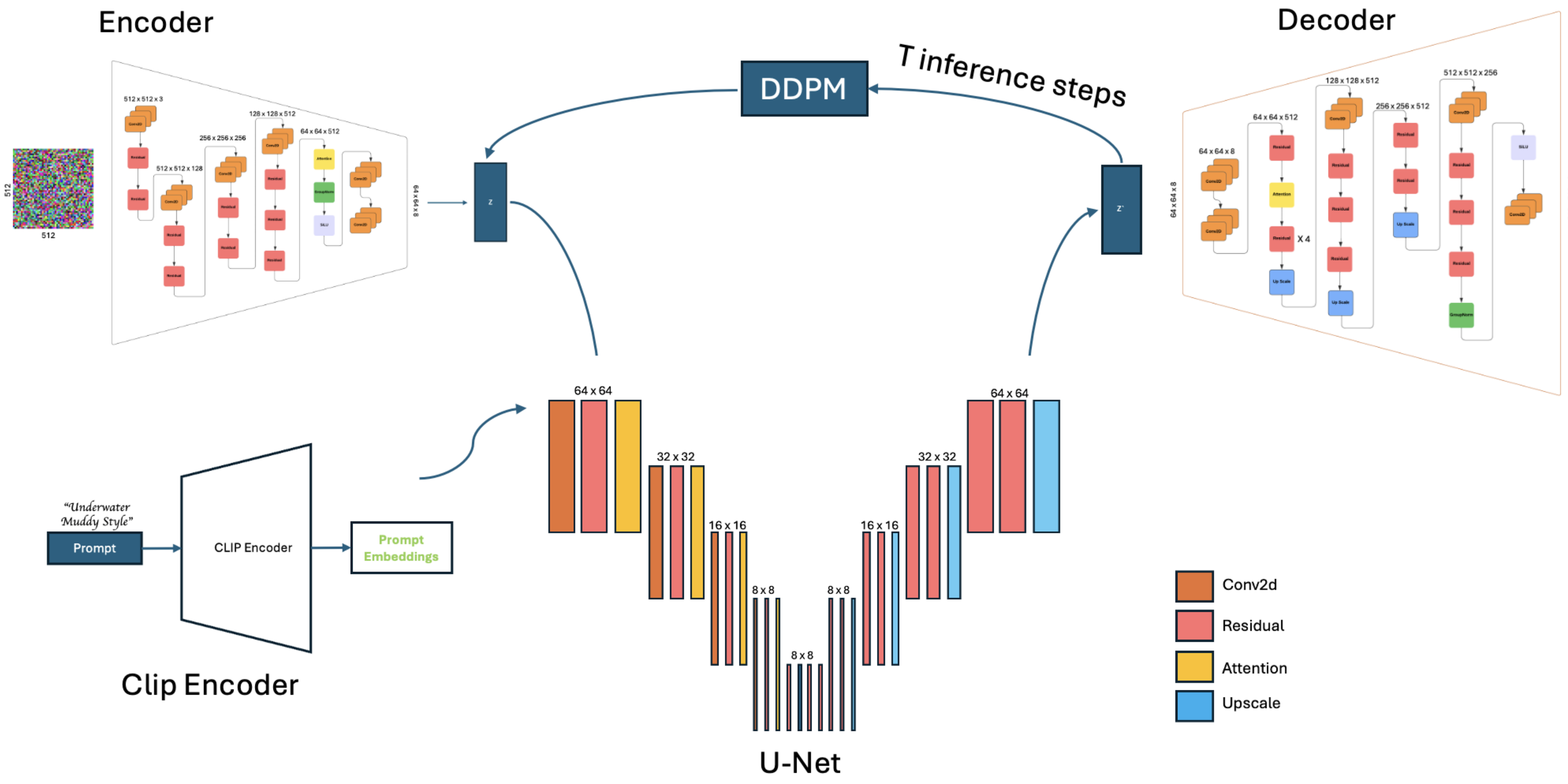
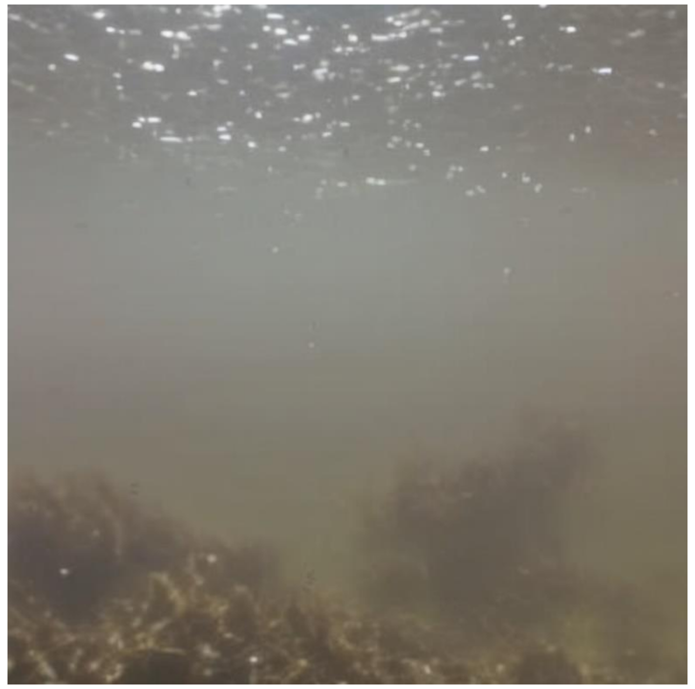


Fig 1. Our Customized Stable Diffusion Architecture

Results

Finetune SDXL 1.0 Base (Method 1) – Single Prompt

- Results are quite good on prompt “muddy underwater images”, but it gets overdriven when some larger prompts are given



muddy underwater image

Coral reef scene with vibrant marine life and fish

Results

Finetune SDXL 1.0 Base (Method 2) – Multiple Prompts

- Results are quite satisfactory, but model tends to deviate if unknown objects and prompts with negative words to our style such as 'colorful', 'vibrant', etc. are given.

muddy underwater image, Coral reef scene with vibrant marine life and fish.

muddy underwater image, Sunlight filtering through crystal-clear water with sea turtles.

muddy underwater image, Whale gliding through calm, deep ocean with rays of light.

muddy underwater image, Underwater volcano with lava illuminating dark ocean..

muddy underwater image, Serene underwater cave filled with plants.

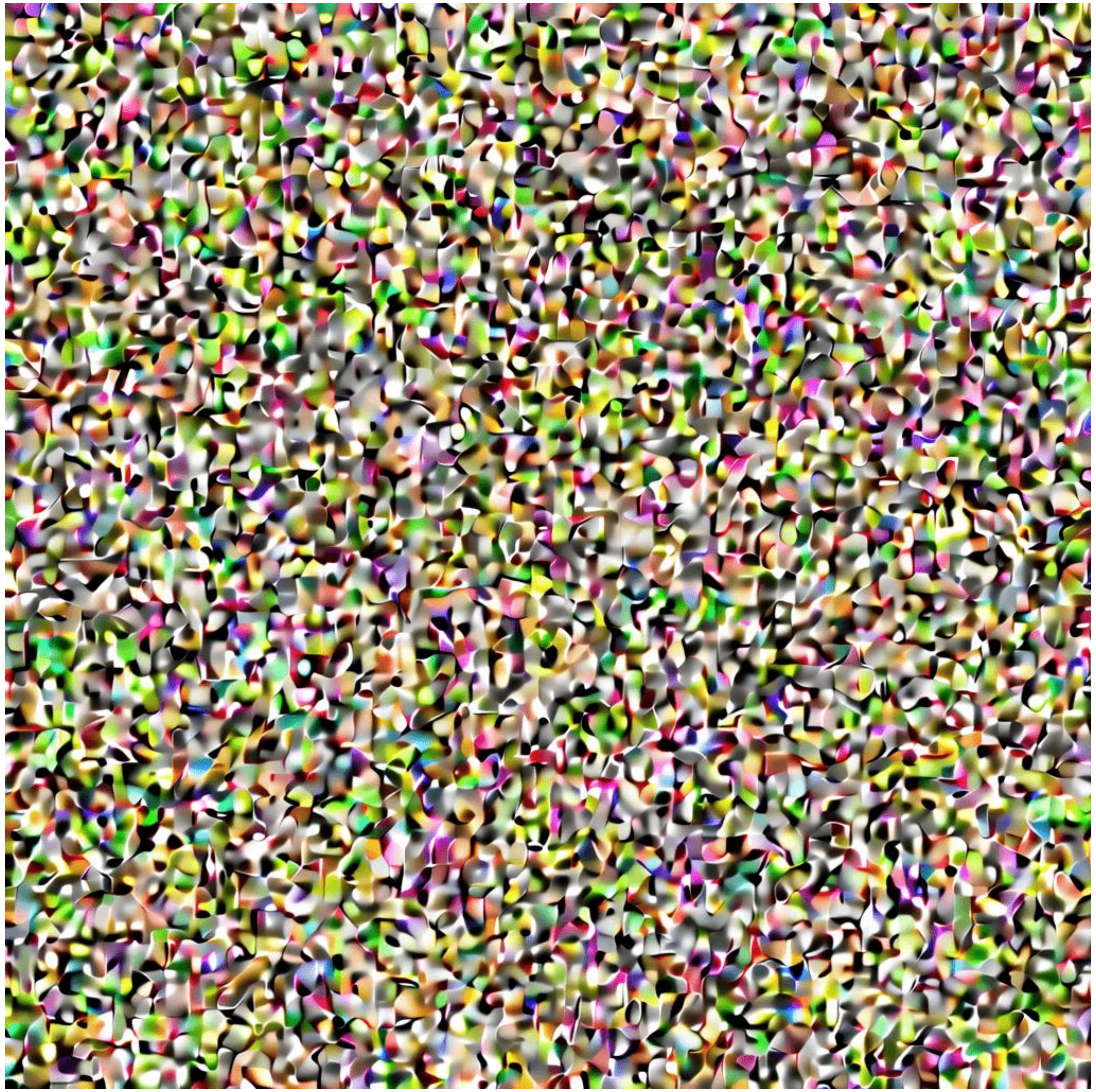
muddy underwater image, Stingray swimming above sandy ocean floor near coral reefs,yellow tint.

Results

Finetune SDXL 1.0 Base (Method 2) – Multiple Prompts

- **Prompt used:** Prompt 1
- **Given Prompt:** Coral reef scene with vibrant marine life and colorful fish

- The gif shows, the output after every inference steps of diffusion process.



Results

Custom defined diffusion model based on SD 1.5 (Method 3)

muddy underwater image

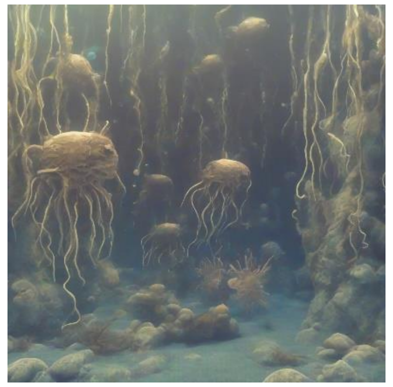
Coral reef scene with vibrant marine life and fish

Qualitative Evaluation

- Prompt : muddy underwater style, Ancient shipwreck surrounded by fish and flowing seaweed.

- Prompt : muddy underwater style, Dense kelp forest with schools of fish swimming around, yellow tint.

- Prompt : muddy underwater style, Deep ocean trench with mysterious underwater creatures.



Generated Images after FineTuning (Model 2)

Generated Image of SD XL 1.0

Quantitative Evaluation

- We have used TypeScore and StyleFidelity as quantitative metrices .
- TypeScore
  - For our second approach (Multiple Prompts) we were able to produce an average TypeScore of 0.336504282 for prompt 6..
  - Baseline StableDiffusion XL has a Type score of 0.238
  - We saw an improvement in TypeScore after FineTuning

- StyleFedelity
  - We were able to reproduce the StyleFedelity of 0.25(SD XL) on our model after FineTuning.

MODEL	***TYPE SCORE***	***STYLE FIDELITY***
Our fine-tuned model	0.3147965	0.240925175
Stable Diffusion XL	0.238 $\pm$0.013	0.25

Limitations and Future scope

- Limitations
  1. Model’s response to prompts outside its training scope, such as diverse object representations and longer prompt tokens, which impact the coherence of generated outputs.
  2. Keywords conveying contrasting styles, like "colorful" or "vibrant," also result in unsatisfactory images, highlighting the model’s restricted adaptability to variations that deviate from its muddy underwater theme.

- Future scope.
  1. Future improvements could involve expanding the dataset to accommodate a wider variety of objects and prompt complexities. Such enhancements could enhance the model's robustness and versatility, making it better suited for diverse applications in synthetic data generation for underwater scenes

Thank you