

과학상상화 Generator

Team | 컴통이들 (김채은, 김창현, 최민석, 황원준)

CONTENTS

01

Introduction

- 주제 선정 동기
- 주제 소개

02

Related Work & Method

- Stable Diffusion
- LoRA
- Multi-LoRA

03

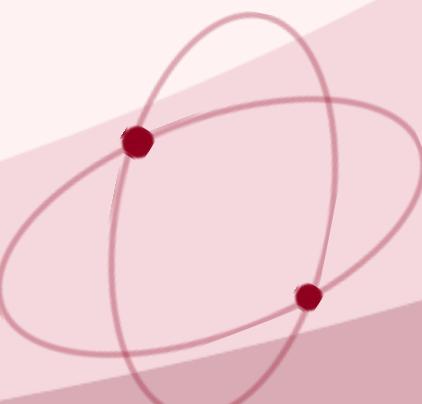
Experiments

- baseline
- 아이 그림체 LoRA
- 아이 그림체 + 수채화
- 아이 그림체 + 수채화+ 과학

04

Analysis

- 각 LoRA의 영향 분석
- LoRA 선형 결합



01. Introduction

01. 주제 선정 동기



- 과연 Image Generation Model로 과학 상상화의 창의성과 작품성을 구현할 수 있을까?

01. 주제 소개

• 과학상상화 특징

- 어린아이가 그린 그림체
- 수채화, 크레파스를 활용한 비비드한 색감
- 우주선, 잠수함 등의 과학적 대상 중심



해저 시험관 아기



환경 미화 로봇



문어 잠수함

01. 주제 소개



child LoRA

+
watercolor
LoRA

+
science
LoRA

Stable
Diffusion

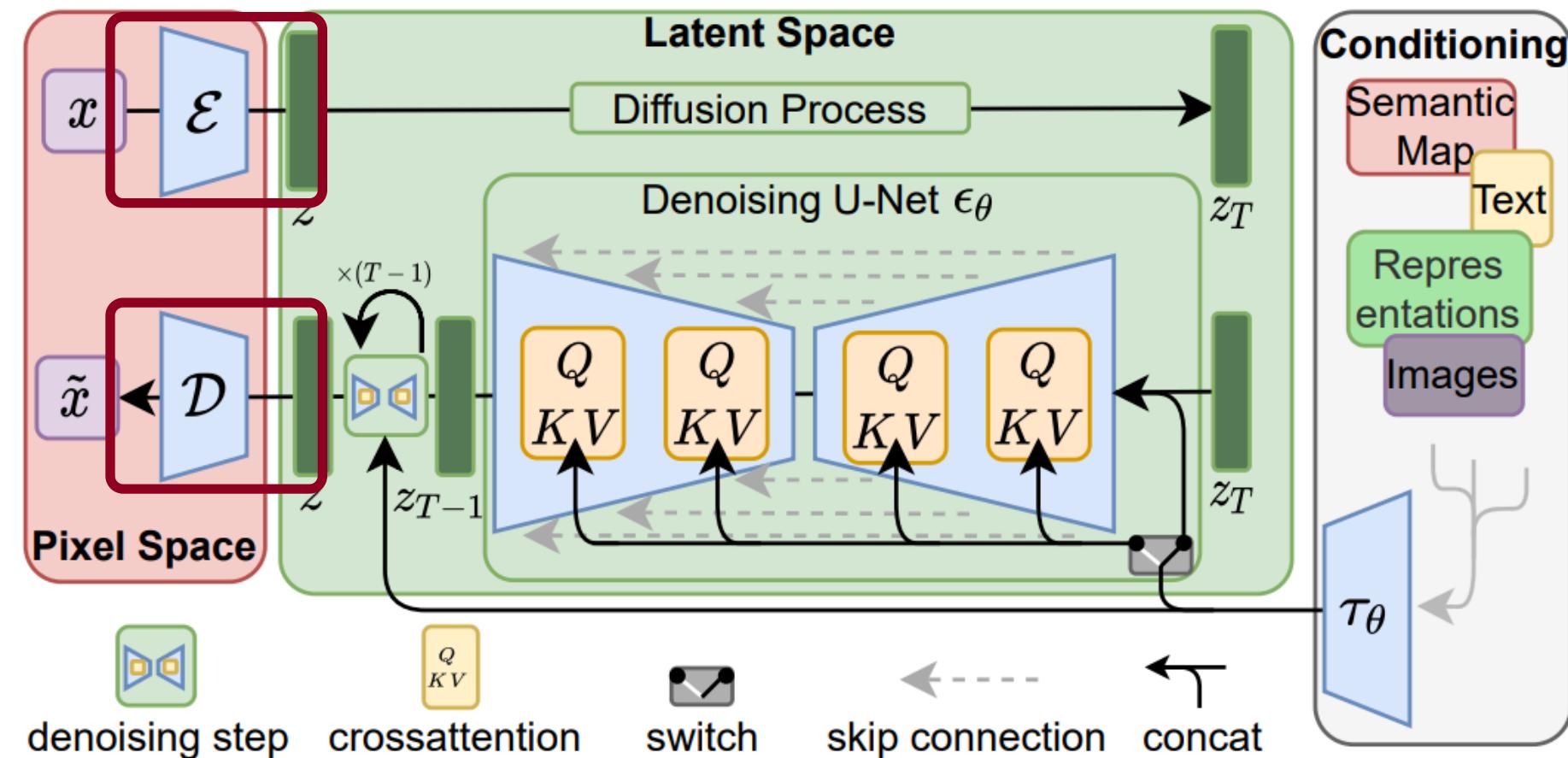


02. Related Work & Method

02. Stable Diffusion

- LDM : Latent Diffusion Model

- 학습된 Autoencoder를 통해 input을 latent vector space로 투영하고, latent vector space에서 Diffusion을 수행해 연산의 효율성을 높인 모델



02. Stable Diffusion

- **Stable Diffusion** : 기존의 LDM 모델에 CLIP의 text encoder를 이용해 conditioning을 준 모델
 - 텍스트 y 가 주어졌을 때 잠재공간의 분포 $p(z_0|y)$ 에서 z_0 을 샘플링하고, $x \sim D(z_0)$ 로 복원

1. 텍스트 임베딩: 프롬프트 $y \rightarrow$ CLIP 텍스트 인코더 \rightarrow 토큰 임베딩 $E(y)$

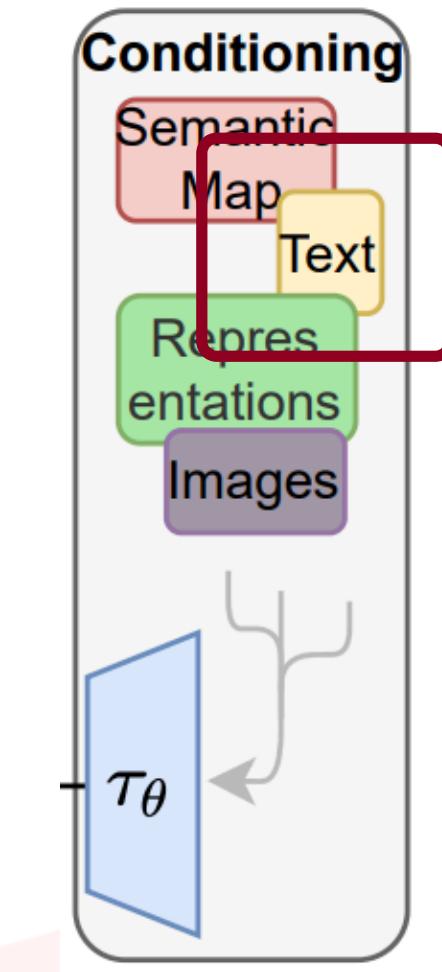
- 이 임베딩이 U-Net 내부의 cross-attention에서 K/V로 쓰인다

2. Cross-Attention in U-Net

- $Q = \text{U-Net latent space}$, $K/V = \text{CLIP 토큰 임베딩}$
- U-Net 각 블록에서 해당 연산을 통해 텍스트 조건을 latent space에 주입

$$Q = \phi(z_t) W_Q, \quad K = E(y) W_K, \quad V = E(y) W_V.$$

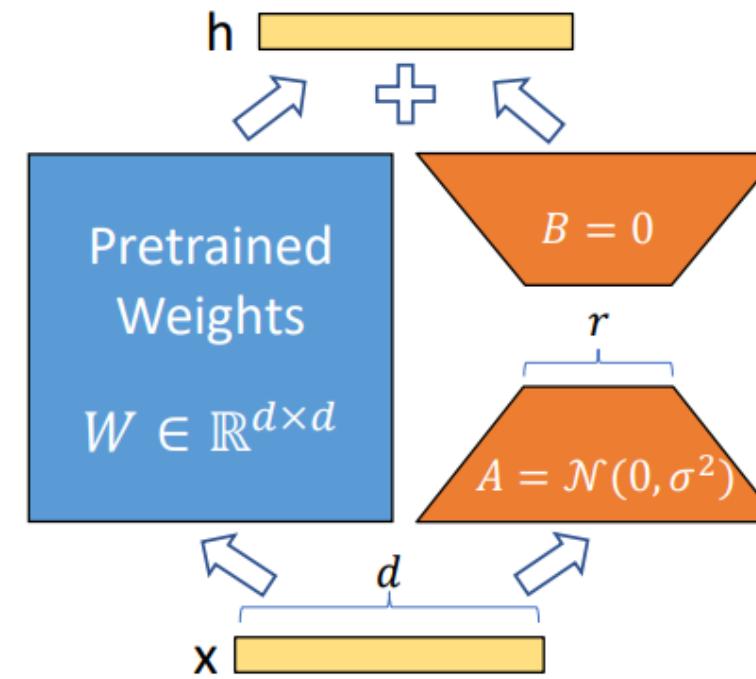
$$\text{Attn}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right) V.$$



02. LoRA

- **LoRA : Low-Rank Adaptation**

- pre-trained weight matrix는 freeze 하고
업데이트 항을 low-rank 행렬곱으로 파라미터화한 후 training



$$h = Wx + sBAx \quad (s = \alpha/r)$$

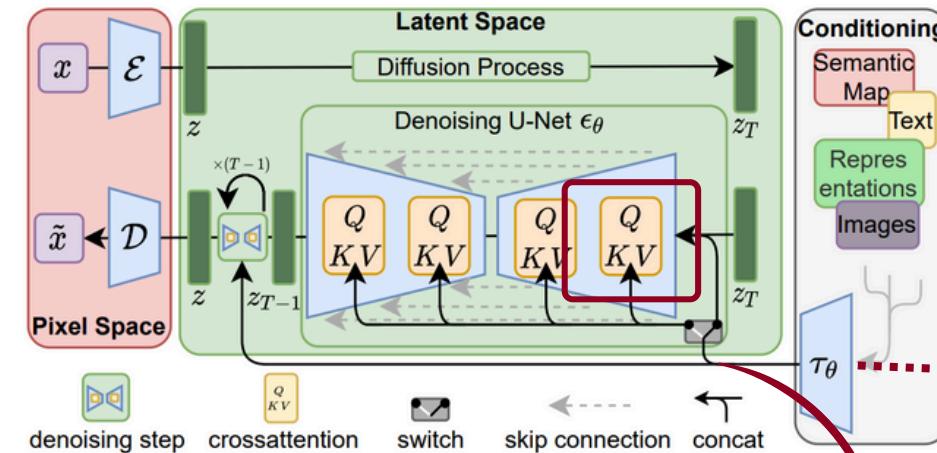
• r = LoRA rank
• α = LoRA alpha(스케일 하이퍼파라미터)

$$\Delta W = sBA, \quad A \in \mathbb{R}^{r \times k}, B \in \mathbb{R}^{d \times r}$$

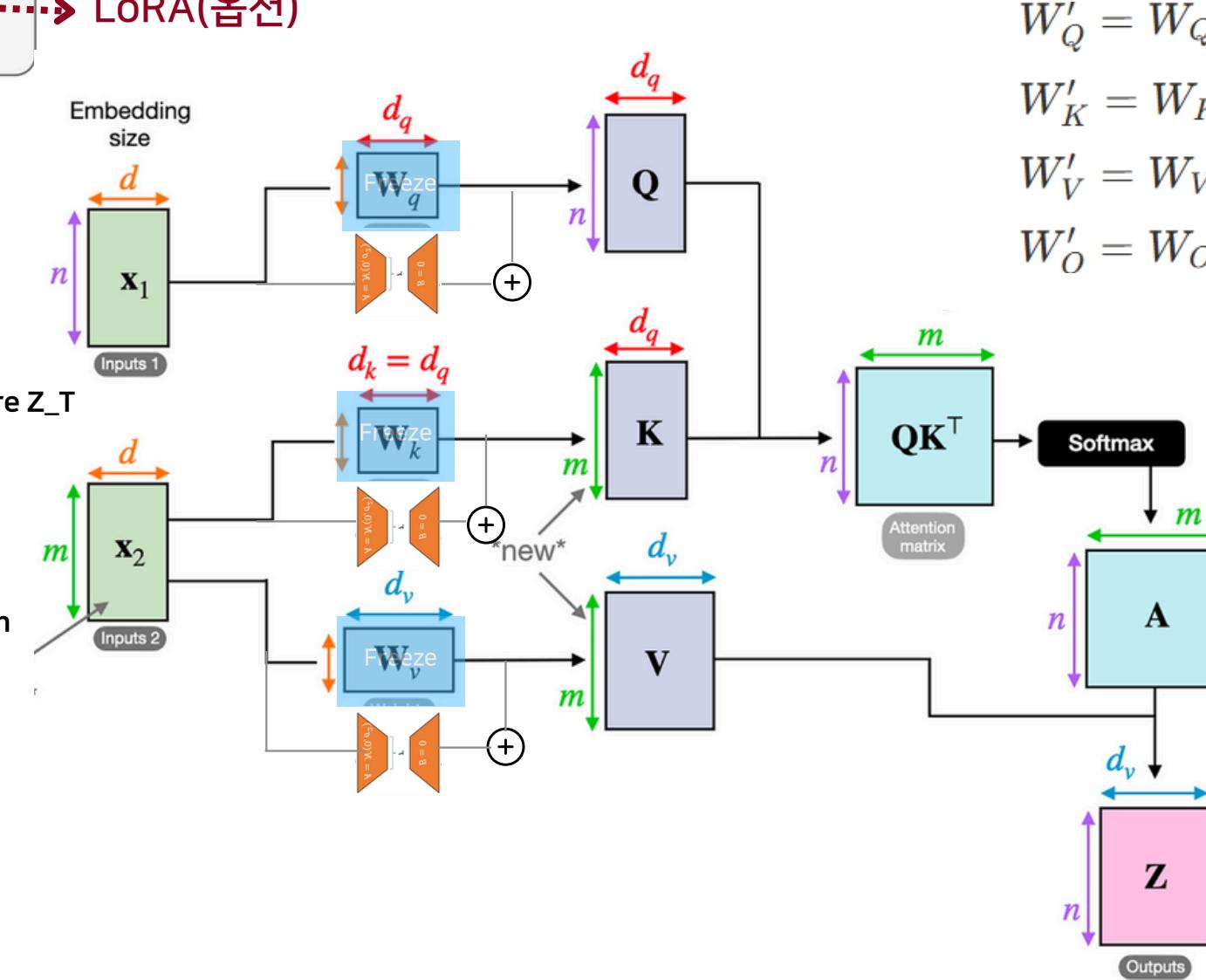
- 메모리 효율성
 - full finetuning 시 parameter 개수 = dk
 - LoRA로 학습되는 parameter 개수 = $r(d+k)$

$$d \times k \rightarrow r(d+k), r \ll d, k$$

02. LoRA



U-net latent feature Z_T
 $\tau(y)$
 "a children's drawing-style cat"



• LoRA adapter 위치

- U-net 내부 cross attention → 스타일 학습
- text Encoder → 트리거 토큰 안정화, 새 개념 학습

$$W'_Q = W_Q + \Delta W_Q = W_Q + sB_Q A_Q$$

$$W'_K = W_K + \Delta W_K = W_K + sB_K A_K$$

$$W'_V = W_V + \Delta W_V = W_V + sB_V A_V$$

$$W'_O = W_O + \Delta W_O = W_O + sB_O A_O$$

02. Multi-LoRA

- Diffusers 라이브러리의 Merge - set_adapters
 - 각각의 LoRA에 가중치를 주어 선형적으로 결합
 - 동일한 버전의 Stable Diffusion에서 학습된 LoRA는 같은 네트워크 형태
 - 동일 레이어에 여러 LoRA 델타를 병렬로 더하기 (runtime multi-LoRA) → 깊이 증가 x
- 사용 시
 - `pipeline.set_adapters(["A", "B"], adapter_weights=[lambda1,lambda2])`

$$\Delta W_{A,B} = \lambda_1 \Delta W_A + \lambda_2 \Delta W_B$$

$$\underbrace{y'}_{\text{총 출력}} = Wx + \sum_k s_k B_k(A_k x) = \left(W + \sum_k s_k B_k A_k \right) x$$

02. Our Method



trained
separately

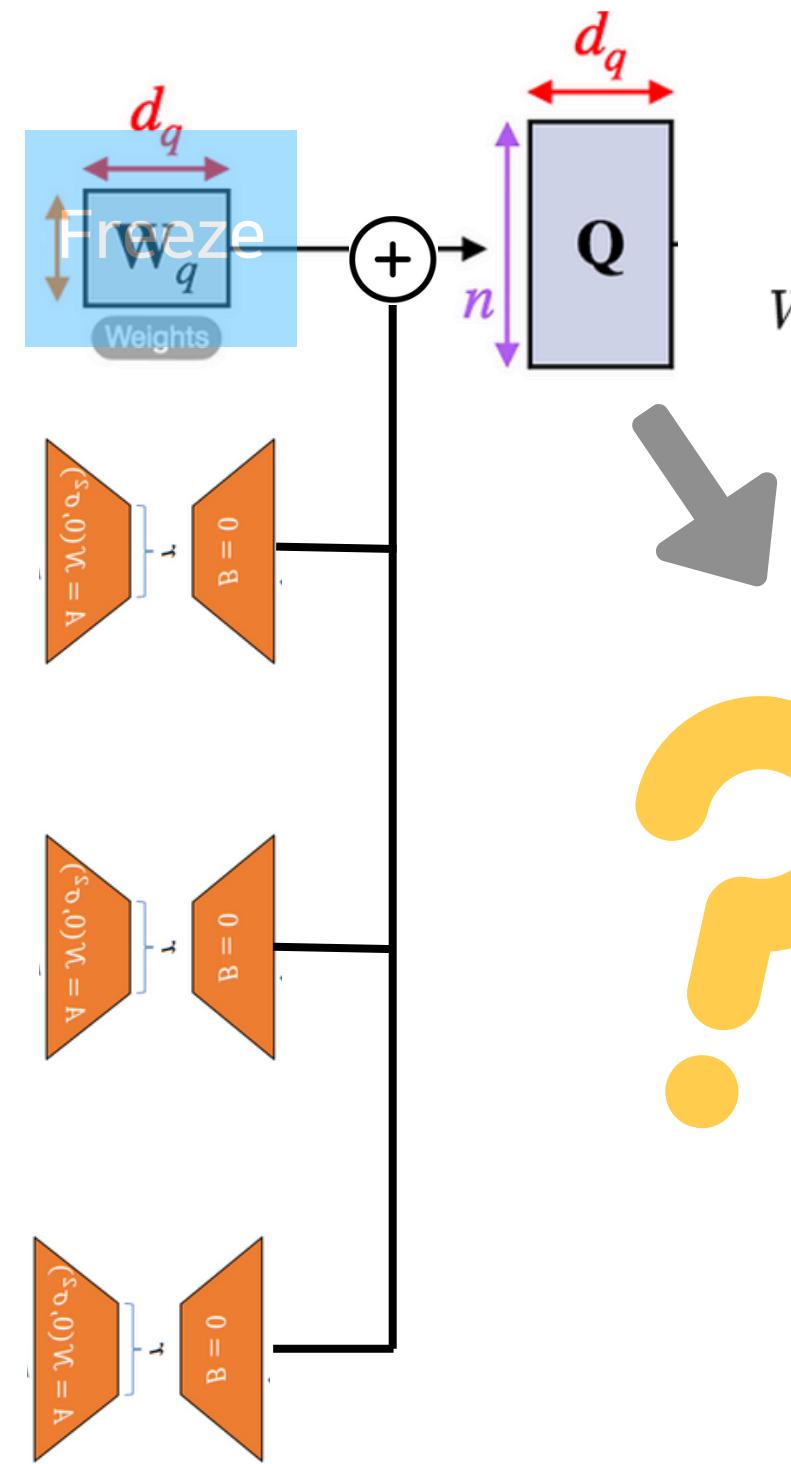
child



watercolor



science



run-time mixing

$$W'_q = W_q + \lambda_c \Delta W_q^{(child)} + \lambda_w \Delta W_q^{(water)} + \lambda_s \Delta W_q^{(sci)}.$$

target



03. Experiments

03. Settings

- Baseline
 - SD v1.5
- LoRA
 - child
 - watercolor
 - science fiction
- Sampling settings
 - seed: 777
 - CFG: 6
 - step: 35
- Prompt
 - 과학 상상화 대표 주제에 대해 GPT 작성
 - 2가지 공통 프롬프트로 비교
- sampling setting 조정하며 추가 실험

03. Baseline (SD v1.5)

L. 미래 교통수단, 하늘을 나는 자동차 과학 상상화

"A colorful science fantasy illustration of futuristic transportation, flying creative cars soaring through the sky, small personal aircraft gliding above futuristic skyscrapers, and underground ultra-high-speed transit systems with glowing tunnels. Bright, playful, and imaginative children's art style, vibrant watercolor textures, cartoon-like, detailed and whimsical, imaginative science fair artwork"

R. 외계 행성 과학 상상화

"A colorful science fantasy illustration of a futuristic space colony and massive space station orbiting an alien planet, astronauts exploring the strange landscape, glowing alien plants and bioluminescent forests, playful and imaginative children's art style, vibrant watercolor textures, highly detailed yet whimsical, imaginative science fair artwork."



03. + 아이 그림체 LoRA w=0.65



- **개선점**
 - SD가 생성한 이미지에 비해 확실히 “아이가 그린” 과학 상상화 스타일
- **문제점**
 - 자동차, 드론, 비행기, 외계인 등 객체들이 많이 뭉개짐
 - 구조물들이 너무 단순

03. + 아이 그림체 LoRA

• 과잉 단순 구조물



• 뭉개진 객체



• 알아볼 수 없는 객체



03. + 수채화 그림체 LoRA w=0.55



- 개선점
 - 객체 윤곽이 훨씬 뚜렷
 - 너무 단순했던 구조물에 디테일이 추가 → 표현력 증가
- 문제점
 - 과학 관련 구조물과 객체 표현은 살짝 미흡 => 개선의 여지

03. + 수채화 그림체 LoRA



03. + 공상과학 구조물 LoRA $w=0.5$



- 개선점

- child+watercolor LoRA 경우에서 구도가 많이 바뀌긴 했으나
- 과학 관련 구조물 디테일이 증가하고 표현력도 증가 → 창의성 증가

03. + 공상과학 구조물 LoRA

아이들이 그린 과학 상상화



Stable Diffusion



LoRA finetuning



Result



03. Summary

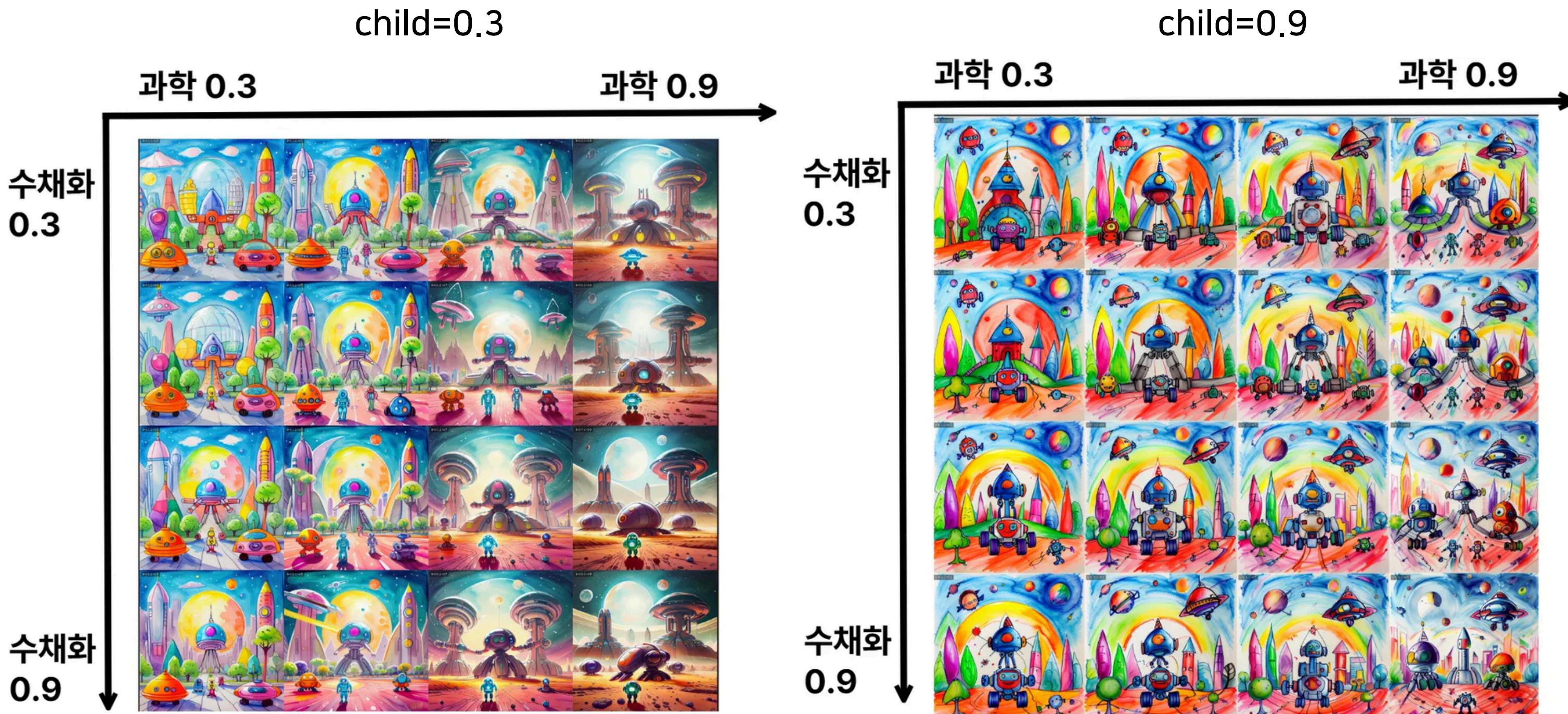
- 미래 교통 수단 (하늘을 날아다니는 자동차)



- 외계 행성



03. 추가 실험 - 가중치 조절



04. 추가 실험 - LoRA의 영향 분석

- 특정 LoRA 가중치 높이기 → 메인 스타일을 명확히 잡고 부가적인 요소를 추가하는 양상



- 모든 LoRA 가중치 높이기 [0.9,0.9,0.9] → 밸런스가 오히려 무너짐



03. 추가 실험 - 해상도

512x512



1024x1024

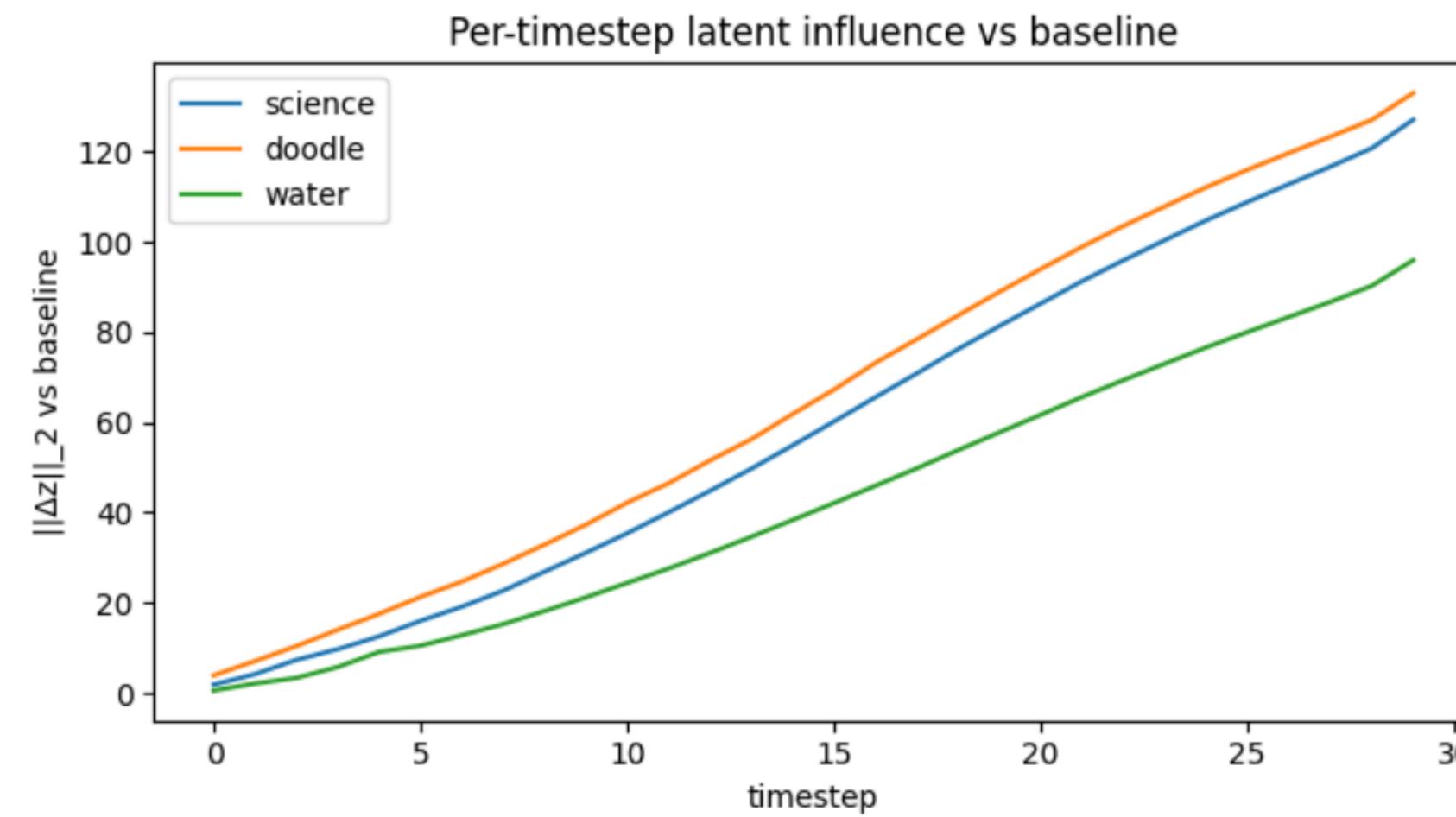


빈 공간을 채우기 위해 특정 요소가 반복되거나 과밀집하는 현상

04. Analysis

04. 각 LoRA 영향 분석

- Per timestep latent influence 곡선- LoRA의 구간별 latent 변화량

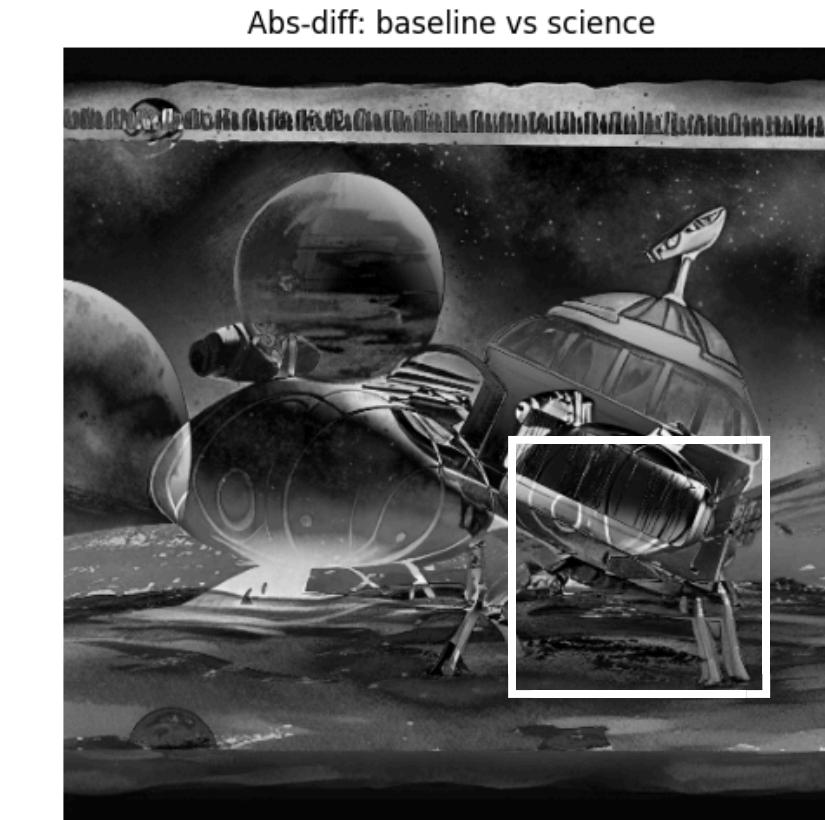
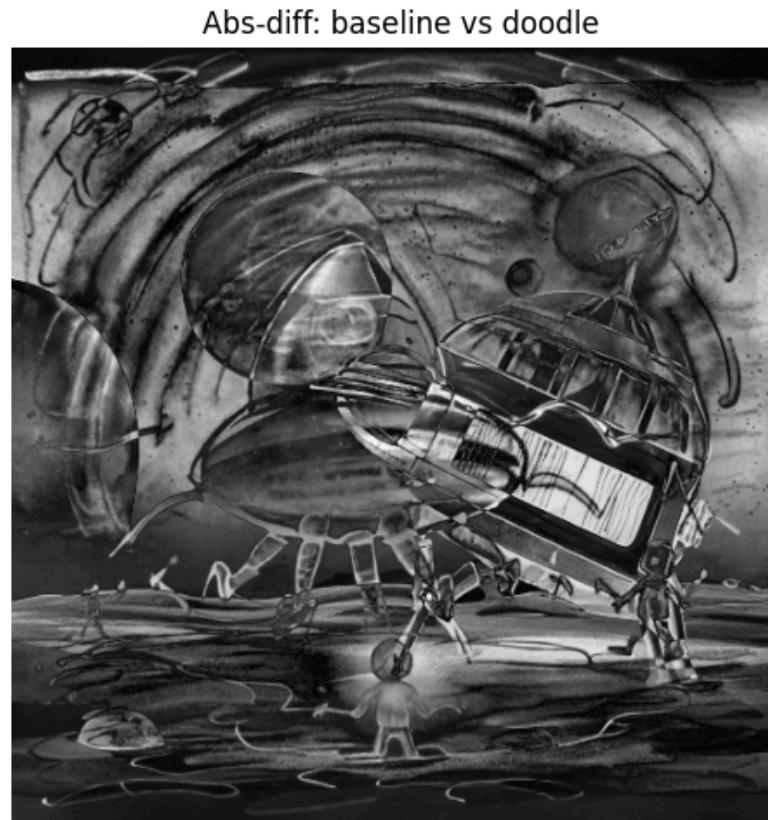


$$\|z_t^{\text{LoRA}} - z_t^{\text{Base}}\|_2$$

- 아이들 그림체 : 그림 전반적인 구도, 질감에 가장 큰 영향.
- 공상과학: 초중반에 형태 및 구조 형성에 영향
- 수채화 : Timestep이 커짐에 따라 다른 LoRA와의 차이가 ↑ . 초반부 채색, 질감에 상대적으로 큰 영향

04. 각 LoRA 영향 분석

- 베이스라인과 각 LoRA별 이미지 픽셀 차이 히트맵
 - 밝을수록 베이스라인 대비 변화가 큼



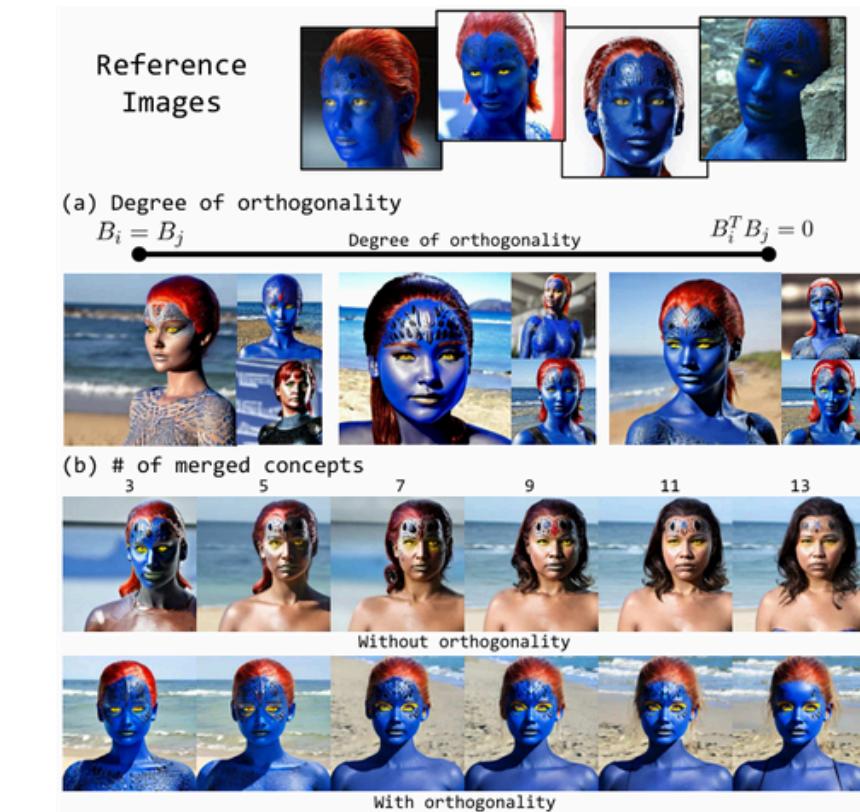
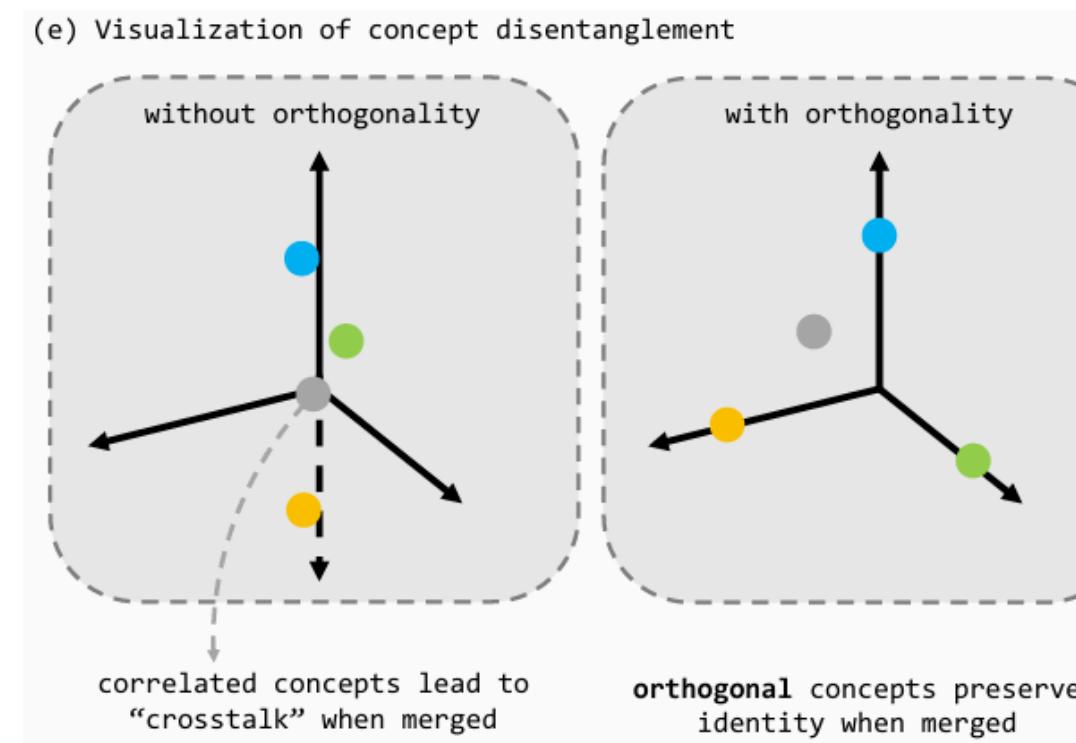
child: 전체적인 질감 및 구도

watercolor : 매끈한 텍스처, 톤 science: 우주선 윤곽 등 구조 및 형태

04. LoRA 선형 결합

- 중첩 원리(Superposition principle)

- A,B LoRA 2개를 합성해도 성능이 A+B로 유지될 것인가?
- 직교성(Orthogonality)은 간섭을 줄여주고 A,B 각각의 정체성을 보존해주는 역할



04. LoRA 선형 결합

- LoRA 선형 결합이 효과적인 이유

- 가중치의 선형 결합이 출력 결과에도 선형적으로 영향(Talyor 근사)

$$f_{W(\alpha)}(x) \approx f_{W_0}(x) + J_{W_0}(x) \left(\sum_i \alpha_i \Delta W_i \right)$$

- 손실 함수에서 직교성이 성립하면, 간섭이 줄어드는 효과

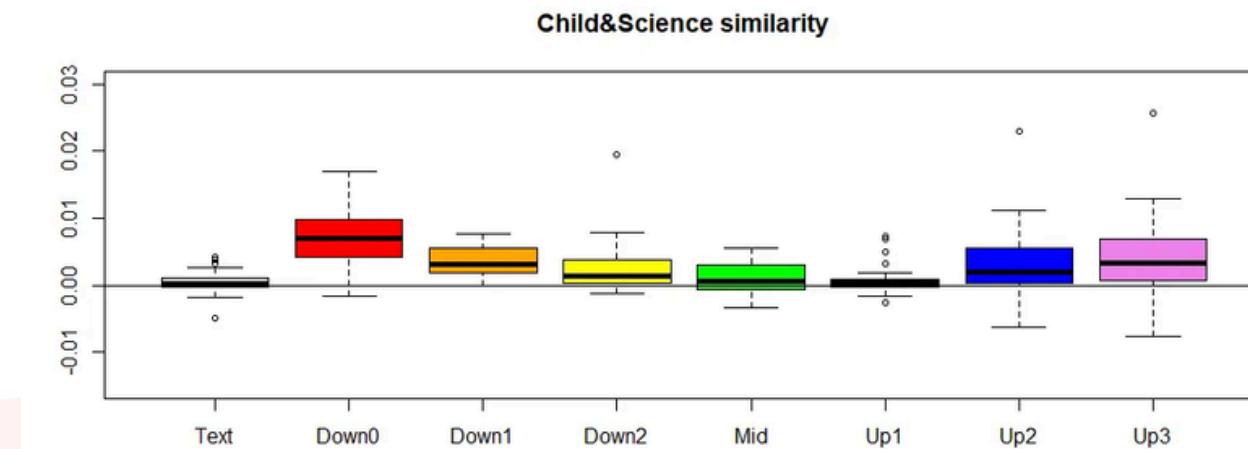
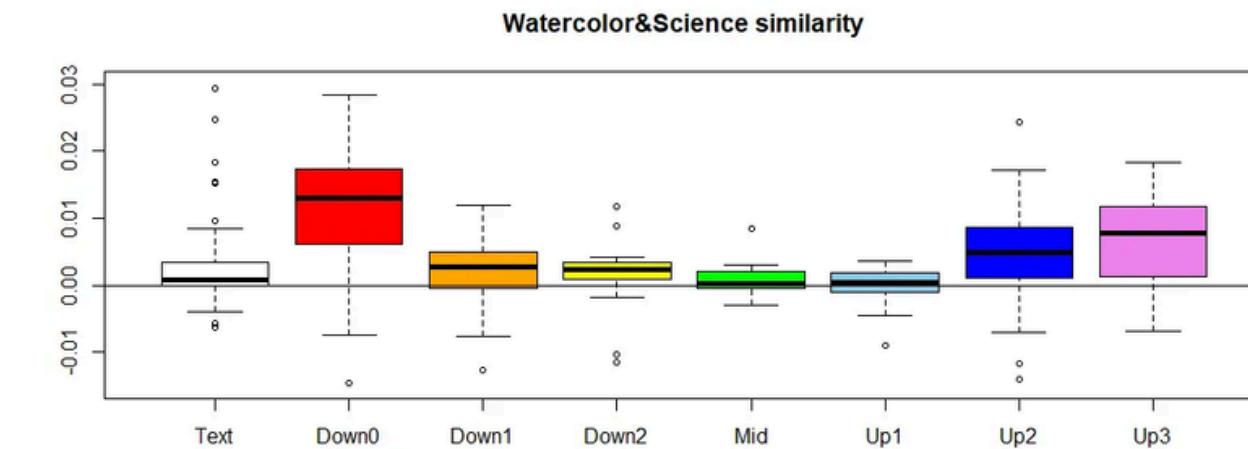
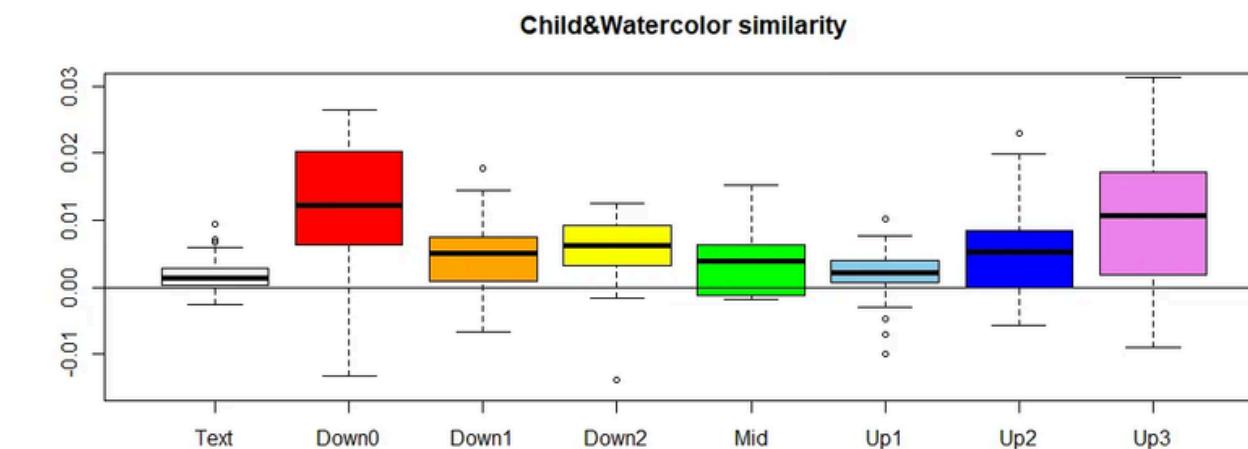
$$L_i(W_0 + \Delta) \approx L_i(W_0) + \nabla L_i^\top \Delta + \frac{1}{2} \Delta^\top H_i \Delta$$

- 결합할 수 있는 LoRA의 개수가 Upper bounded.
 - rank가 임의의 r보다 작거나 같은 n개의 행렬에 대해, 모든 행렬을 더한 결과의 rank는 최대가 존재

>> 특정 최대치까지 덧셈으로 중첩 원리 구현 가능

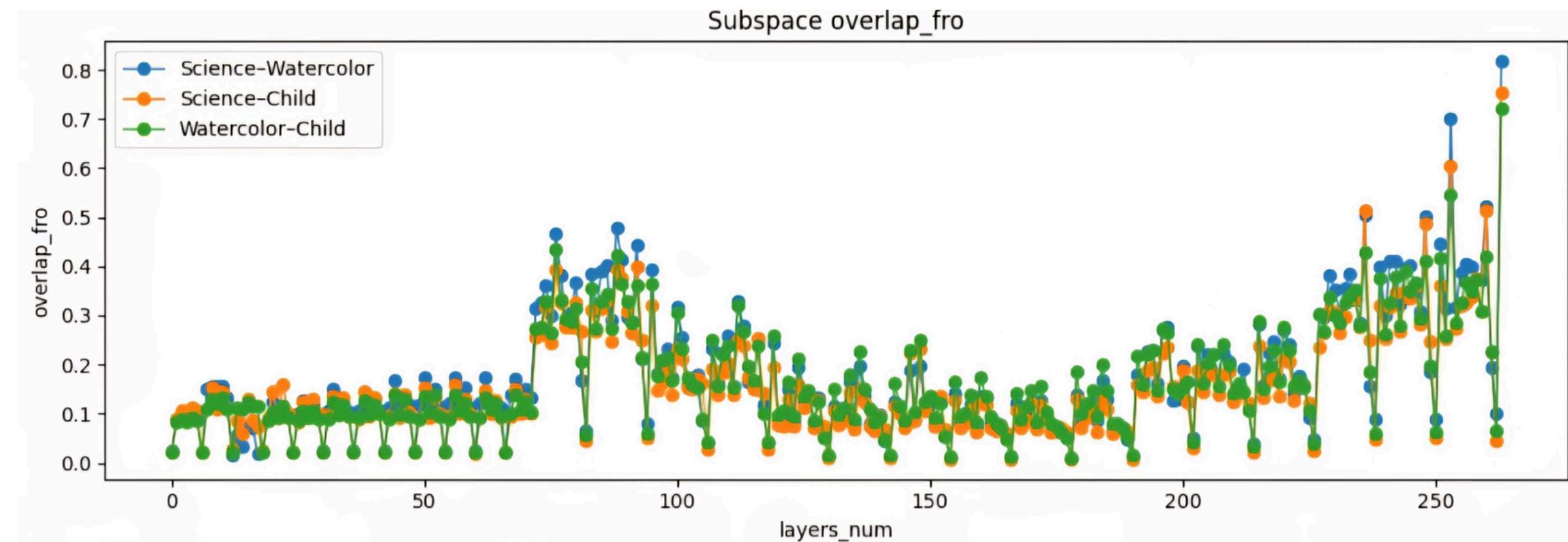
04. LoRA 선형 결합

- 각 LoRA 레이어별 Cos 유사도 분석
 - 업데이트의 주방향의 내적 계산
 - 직교할수록 0에 수렴
 - 전체적으로 U자형 구조
 - 모든 Mid block 단계에서 0에 수렴
→ Main feature를 선형으로 표현 가능
 - Child & Watercolor / Science
 - 상대적으로 child와 watercolor는 유사
 - science와는 거의 직교에 가까움



04. LoRA 선형 결합

- 레이어별 유사도 측정



$$\text{overlap} = \frac{\|U_1^\top U_2\|_F^2}{k}$$

- 공상과학, 수채화, 아이 그림체에 대한 SVD(특이값분해)를 바탕으로 Overlap score 측정.
- 각각의 유사도 분포가 전체적으로 비슷함.
- 중간, 끝 부분 레이어 제외하고, 모든 LoRA 합성에 대해 전반적으로 유사도 0에 수렴(직교)

>> 세 LoRA의 결합으로 다양한 결과물 도출 가능성 높아짐.

04. 한계 및 보완점

- LoRA 결합 면에서
 - 완벽한 직교가 아님. Projection 을 통한 방식으로도 다중 LoRA 성능 향상 기대
 - 다중 LoRA를 DARE, TIE 등의 방식으로 보다 task-aware한 병합 가능
- 결과 해석 면에서
 - 각 LoRA별 영향에 대한 정량 평가 어려움
 - 결과 그림 및 그래프 해석에서 연구자의 주관이 개입할 여지가 우려됨.
 - 다른 LoRA method 활용 or PCA, t-SNE 등으로 다양한 정량 분석 기대



Thank You