Enhance Prompt Understanding In Text-to-Image Diffusion Model2

2024.12.27

이진우

DMQA Open Seminar

발표자 소개

❖ **이진우 (Jin Woo Lee)**

- **학력**
  - **대학** 고려대학교 산업경영공학과 (2016.09 ~ 2023.02)
  - **대학원** 고려대학교 산업경영공학과 (2023.03 ~ present)
  - **연구실** Data Mining & Quality Analytics Lab
  - **지도교수** 김성범 교수님
- **연구 분야** 이미지 생성, 시계열 예측
- **E-mail** dlwlsdn0225@korea.ac.kr

목차

Introduction
Background

❖ 이미지 생성 모델

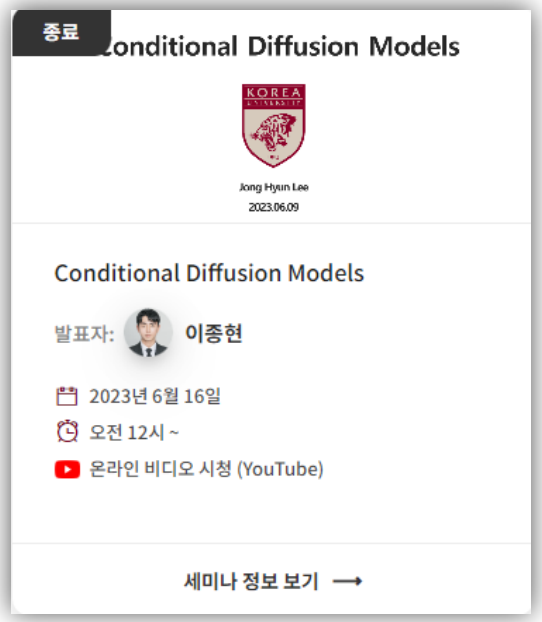
- 상상을 현실로 그려내는 흥미로운 분야
  - **Text to Image generation**: 사용자가 프롬프트를 주면 원하는 이미지를 생성하는 방식으로 작동
- 프롬프트를 통해 사용자 의도를 반영한 고퀄리티 이미지 생성하는 디퓨젼 모델이 각광받음

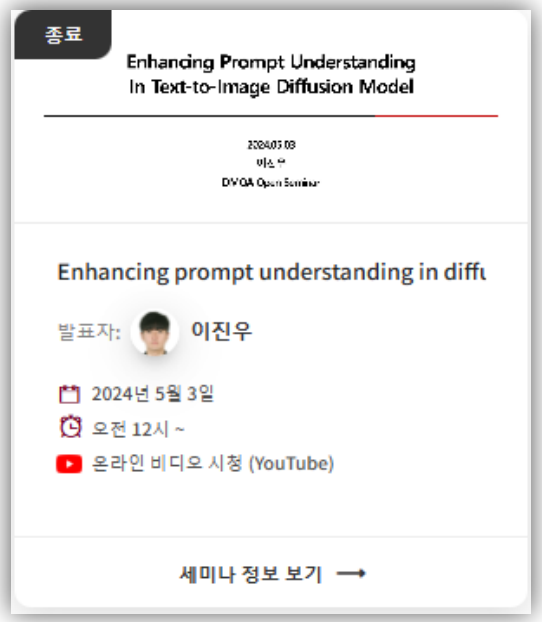
열대 벽지 배경으로 울퉁불퉁한 흰색 외부와 푹신한 내부를 가진 리치에서 영감 받은 구형 의자 사진

DALLE-3

Introduction
Text to Image Generation

- **Improving Sampling Speed of Diffusion Model :** 디퓨젼 모델 기본 원리와 근간이 되는 DDPM과 DDIM 소개
- **Conditional Diffusion Model :** 디퓨젼 모델에 컨디션을 부여하는 Classifier Guidance, Classifier Free Guidance 그리고 Latent Diffusion 소개
- **Enhance Prompt Understanding:** 디퓨젼 모델의 프롬프트 반영도를 높인 Structure Diffusion, Attend-and-Excite 소개

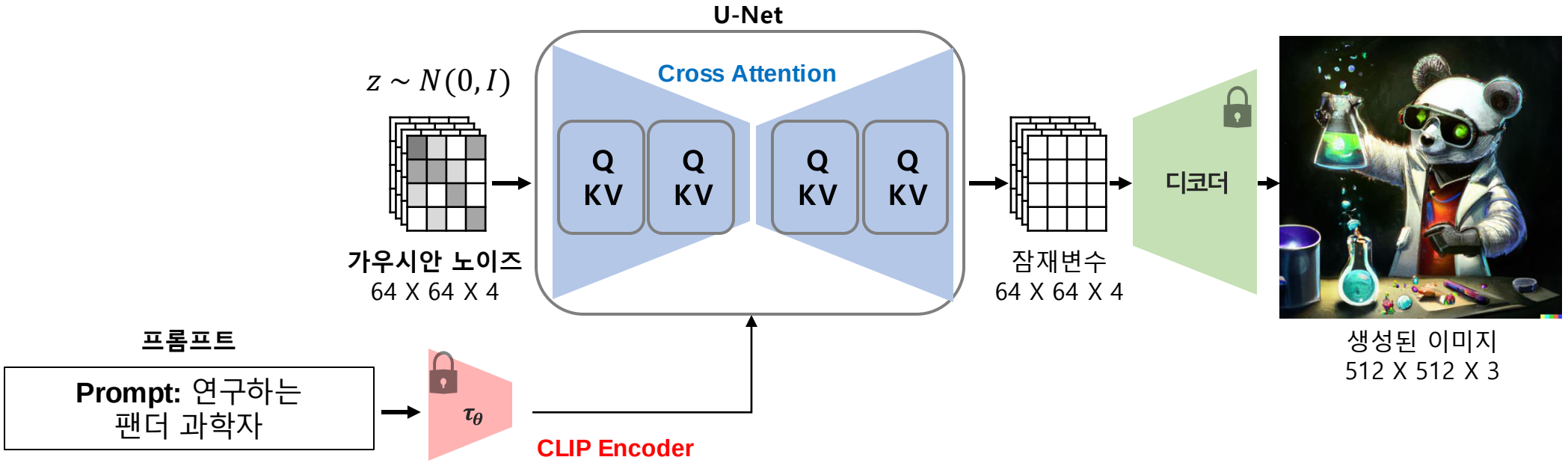




Introduction
Text to Image Generation

❖ 입력 조건에 맞는 이미지 생성: Latent Diffusion Model

- **Text to Image Generation** : LDM은 가우시안 노이즈와 프롬프트를 입력받아 원하는 이미지를 생성
  - **CLIP Encoder** : 텍스트와 이미지의 joint embeddding을 구하는 **텍스트 인코더**
  - **Cross-Attention**: 프롬프트로 주어진 **condition을 이미지에 반영**
- 많은 연구들이 **Stable Diffusion**을 기반으로 진행



R. Rombach, et al. "High-Resolution Image Synthesis with Latent Diffusion Models" CVPR (2022)
https://www.filipcvetko.com/dalle2/

Introduction

Text to Image Generation

디퓨젼 모델은

사용자가 입력한 '**프롬프트**'를

완벽하게 반영하는가?

Introduction
Text to Image Generation

❖ **Text to Image 디퓨전 모델에서 발생하는 문제점**

- 사용자가 입력하는 프롬프트를 **완벽하게 반영**하기 어려움

Prompt: " A bear and a **turtle** "

< Missing object >

Prompt: " A **pink** sunflower and a **yellow** flamingo "

< Attribute binding>

Introduction
Text to Image Generation

❖ **Text to Image 디퓨전 모델에서 발생하는 문제점**

- 사용자가 입력하는 프롬프트를 **완벽하게 반영**하기 어려움
- **객체가 생성되지 않는** (missing object), **객체에 잘못된 속성이 부여되는** (attribute binding) 문제 발생

Prompt: " A bear and a **turtle** "

< Missing object >

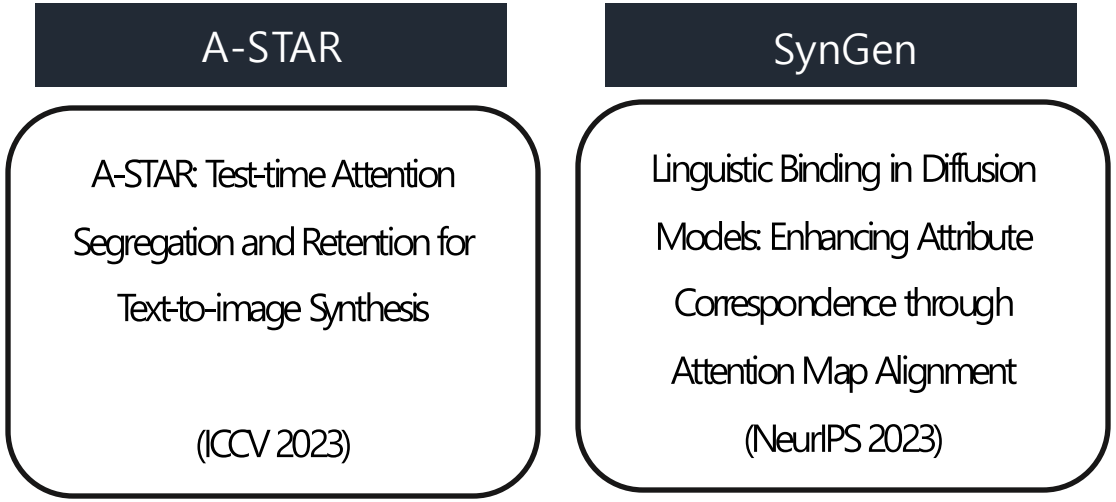
Prompt: " A **pink** sunflower and a **yellow** flamingo "

< Attribute binding>

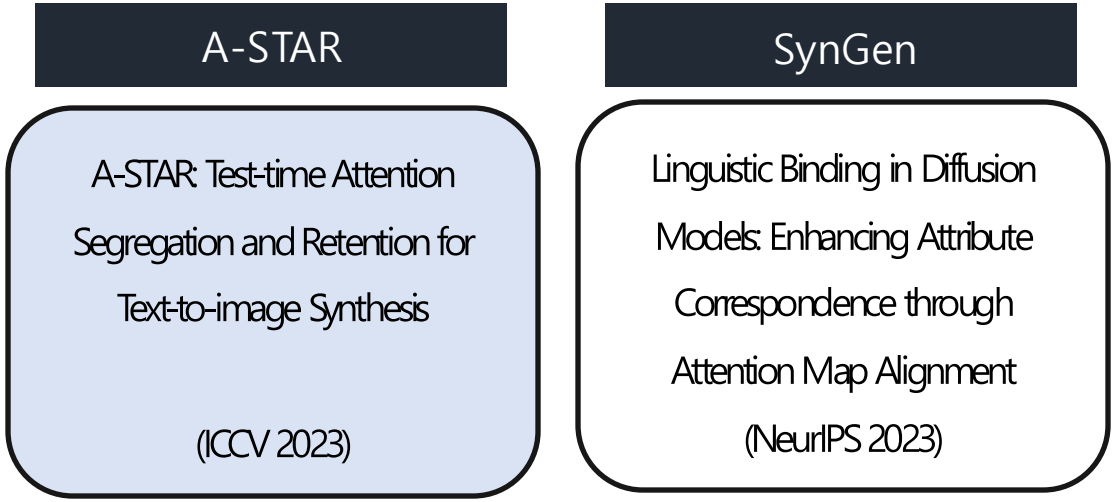
Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Rassin, Royi, et al. "Linguistic binding in diffusion models: Enhancing attribute correspondence through attention map alignment." *Advances in Neural Information Processing Systems* 36 (2024).

Enhancing Prompt Understanding



Enhancing Prompt Understanding



Enhancing Prompt Understanding
A-STAR

❖ **A-STAR:** Test-time **A**ttention **S**egrega**T**ion **A**nd **R**etention for Text-to-image Synthesis **(ICCV 2023)**

- **Motivation:** 객체들이 생성되지 않는 문제 - Missing object
- 프롬프트 내 객체 토큰들이 ①영향을 끼치는 영역이 **겹침** ②영향력이 **감소**

Prompt: " A turtle and a **bear** "

< Missing object >

Prompt: " A bear and a **fish**"

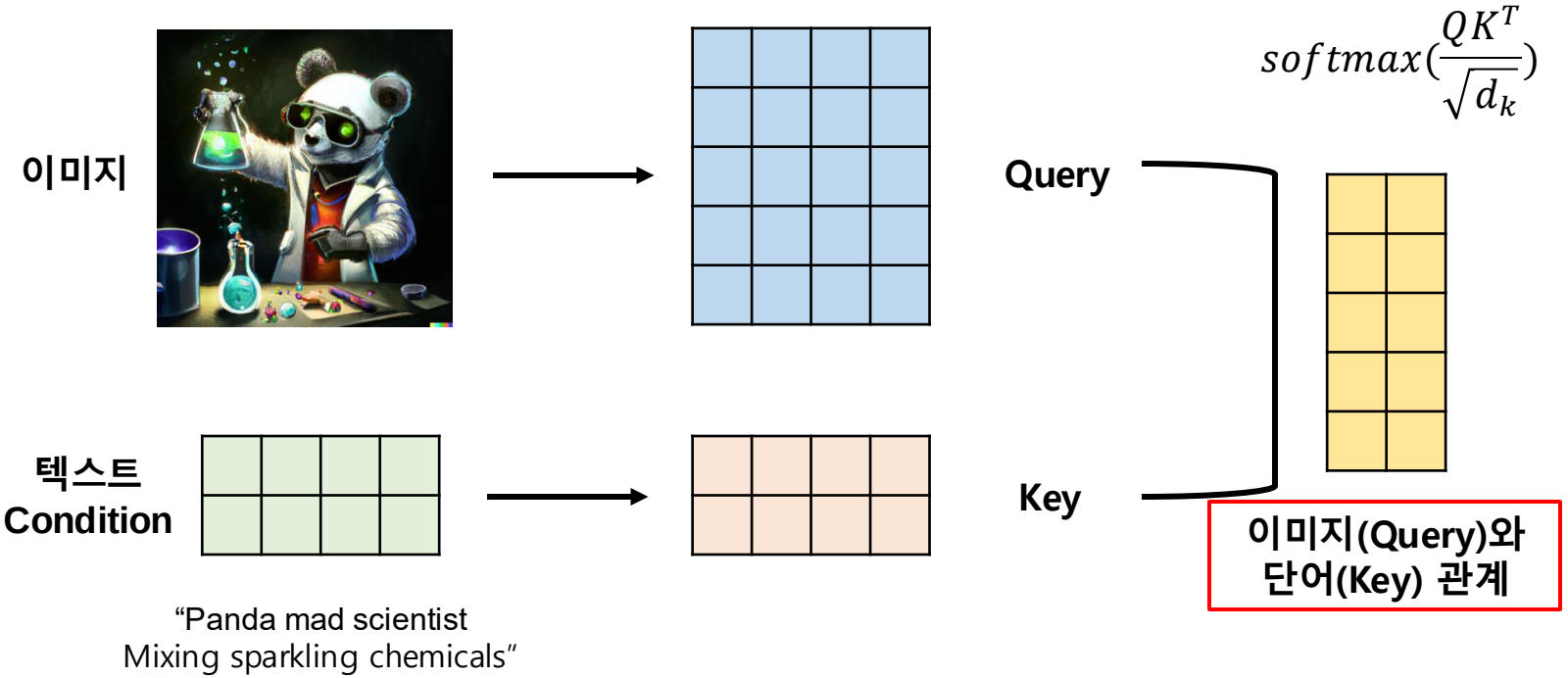
< Missing object >

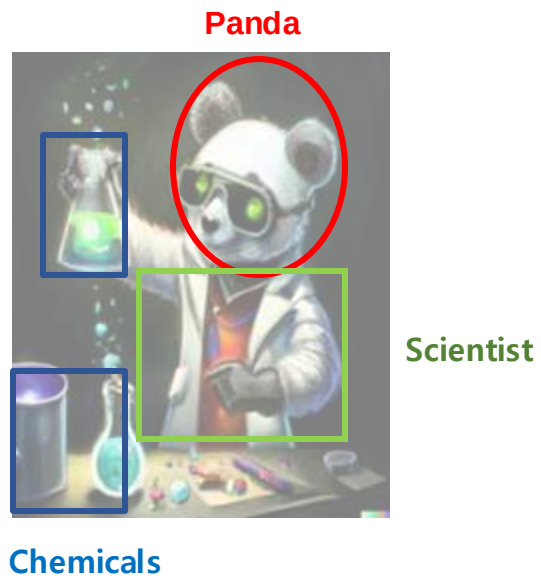
Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Enhancing Prompt Understanding

❖ Cross attention map

- Cross-Attention: 프롬프트로 주어진 condition을 이미지에 반영
- **Cross attention map**은 각 프롬프트내 **단어가 이미지 어느 부분과 연관있는지** 나타냄
- 이를 통해 생성과정에서 각 토큰이 어떤 영향을 주는지 파악 가능





Enhancing Prompt Understanding
A-STAR

❖ Missing object 발생 원인

- Cross attention map을 통해 각 단어가 이미지 어느 부분에 영향을 주는지 확인 가능 (**빨간색**일수록 강한 영향력)
  - ① Overlap: 프롬프트 내 토큰들이 **영향을 끼치는 영역이 겹치는 현상**
  - ② Attention Decay: 프롬프트 내 토큰의 **영향력 감소**

Prompt: " A turtle and a **bear**"

생성된 이미지

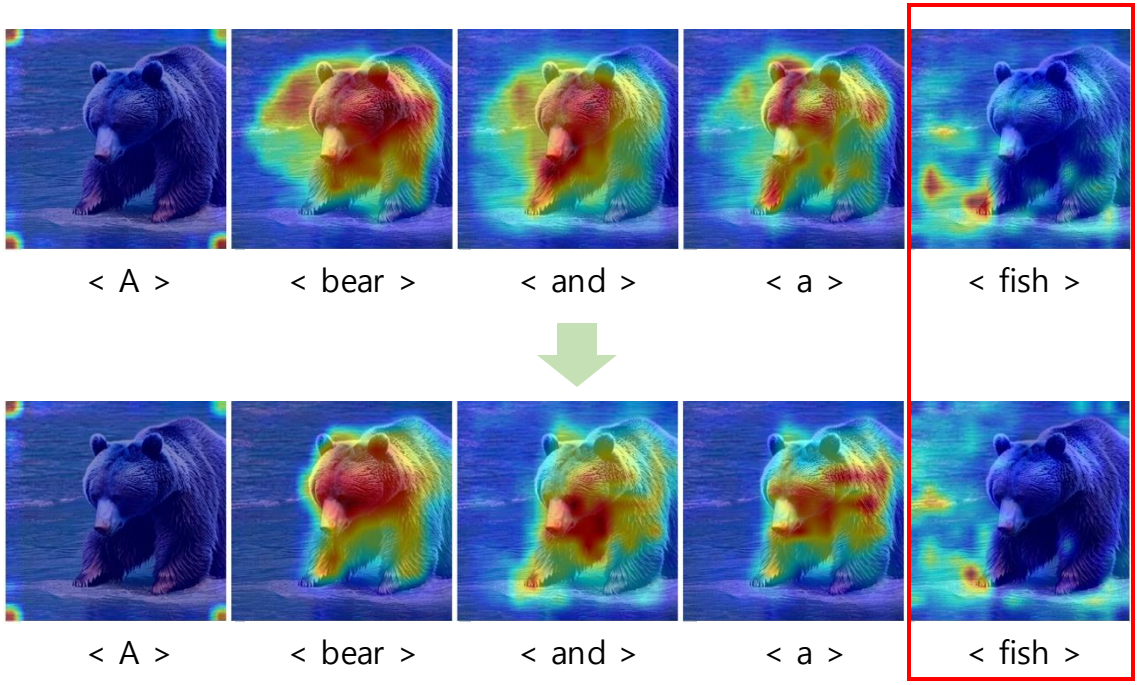
Enhancing Prompt Understanding
A-STAR

❖ Missing object 발생 원인

- Cross attention map을 통해 각 단어가 이미지 어느 부분에 영향을 주는지 확인 가능 (**빨간색**일수록 강한 영향력)
  - ① Overlap: 프롬프트 내 토큰들이 영향을 끼치는 영역이 겹치는 현상
  - ② Attention Decay: 프롬프트 내 토큰의 **영향력 감소**

Prompt: " A bear and a **fish** "

생성된 이미지



Cross attention map at time step 50

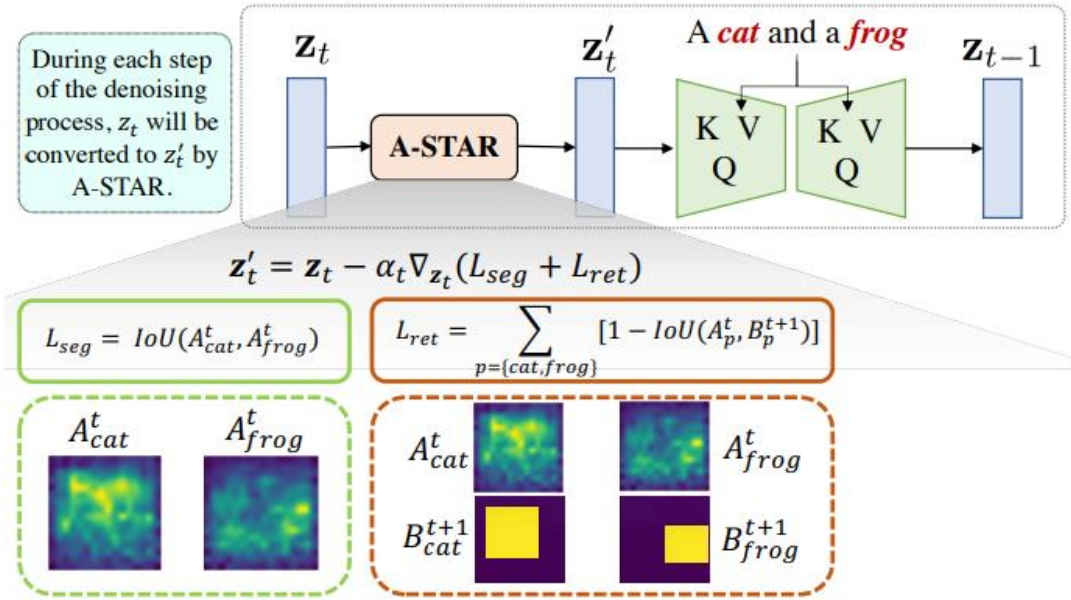
Cross attention map at time step 40

Enhancing Prompt Understanding
A-STAR

❖ **A-STAR:** Test-time **A**ttention **S**egrega**T**ion **A**nd **R**etention

- **Segregation Loss:** 서로 다른 객체 토큰이 다른 부분에 영향을 주도록 **cross attention overlap** 방지
- **Retention Loss:** 특정 객체 토큰의 cross attention 값을 유지하여 **attention decay** 방지

Segregation loss



Retention loss

Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

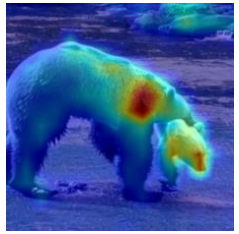
Enhancing Prompt Understanding
A-STAR

❖ **Missing object 문제 해결방법1 – Segregation loss**

- 각 객체 토큰 Cross attention이 겹치는 **Overlapping 최소화**
- Cross attention이 겹치지 않고 분리되도록 유도하는 segregation loss 제안

Prompt: " A **bear** and a **turtle** "

A_t^{bear}



A_t^{turtle}

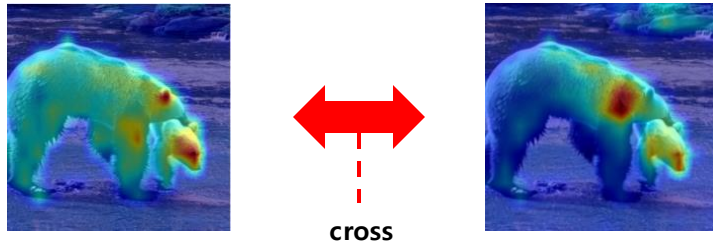
Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Enhancing Prompt Understanding
A-STAR

❖ **Missing object 문제 해결방법1 – Segregation loss**

- 각 객체 토큰 Cross attention이 겹치는 **Overlapping 최소화**
- Cross attention이 겹치지 않고 분리되도록 유도하는 segregation loss 제안

Prompt: " A **bear** and a **turtle** "



A_t^{bear}

cross attention이 겹치지 않도록

A_t^{turtle}

$$\mathcal{L}_{\text{seg}} = \sum_{\substack{m,n \in \mathcal{C} \\ \forall m > n}} \left[\frac{\sum_{ij} \min([\mathbf{A}_t^m]_{ij}, [\mathbf{A}_t^n]_{ij})}{\sum_{ij} ([\mathbf{A}_t^m]_{ij} + [\mathbf{A}_t^n]_{ij})} \right]$$

m: bear
n: turtle
t: time step
i,j: 위치

Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Enhancing Prompt Understanding
A-STAR

❖ **Missing object 문제 해결방법1 – Segregation loss**

- 각 객체 토큰 Cross attention이 겹치는 **Overlapping 최소화**
- Cross attention이 겹치지 않고 분리되도록 유도하는 segregation loss 제안

Prompt: " A **bear** and a **turtle** "

$$\mathcal{L}_{\text{seg}} = \sum_{\substack{m,n \in \mathcal{C} \\ \forall m > n}} \left[\frac{\sum_{ij} \min([\mathbf{A}_t^m]_{ij}, [\mathbf{A}_t^n]_{ij})}{\sum_{ij} ([\mathbf{A}_t^m]_{ij} + [\mathbf{A}_t^n]_{ij})} \right]$$

m: bear
n: turtle
t: time step
i,j: 위치

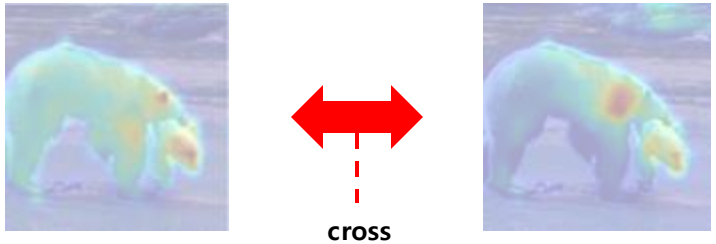
Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Enhancing Prompt Understanding
A-STAR

❖ **Missing object 문제 해결방법1 – Segregation loss**

- 각 객체 토큰 Cross attention이 겹치는 **Overlapping 최소화**
- Cross attention이 겹치지 않고 분리되도록 유도하는 segregation loss 제안

Prompt: " A **bear** and a **turtle** "



A_t^{bear}

cross attention이 겹치지 않도록

A_t^{turtle}

$$\mathcal{L}_{\text{seg}} = \sum_{\substack{m,n \in \mathcal{C} \\ \forall m > n}} \left[\frac{\sum_{ij} \min([\mathbf{A}_t^m]_{ij}, [\mathbf{A}_t^n]_{ij})}{\sum_{ij} ([\mathbf{A}_t^m]_{ij} + [\mathbf{A}_t^n]_{ij})} \right]$$

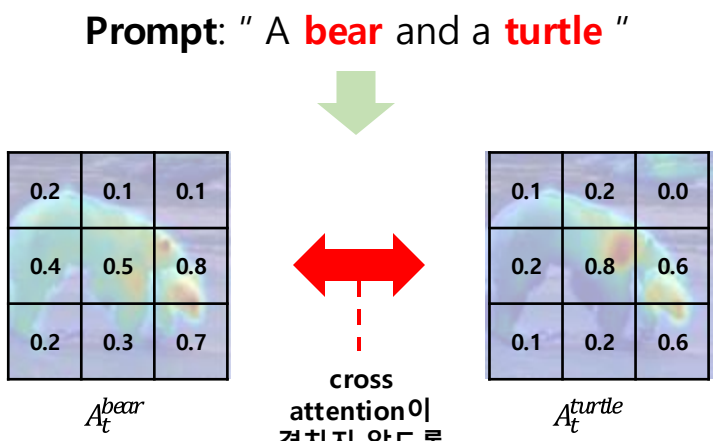
m: bear
n: turtle
t: time step
i,j: 위치

Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Enhancing Prompt Understanding
A-STAR

❖ **Missing object 문제 해결방법1 – Segregation loss**

- 각 객체 토큰 Cross attention이 겹치는 **Overlapping 최소화**
- Cross attention이 겹치지 않고 분리되도록 유도하는 segregation loss 제안



$$\mathcal{L}_{\text{seg}} = \sum_{\substack{m,n \in \mathcal{C} \\ \forall m > n}} \left[\frac{\sum_{ij} \min([\mathbf{A}_t^m]_{ij}, [\mathbf{A}_t^n]_{ij})}{\sum_{ij} ([\mathbf{A}_t^m]_{ij} + [\mathbf{A}_t^n]_{ij})} \right]$$

m: bear
n: turtle
t: time step
i,j: 위치

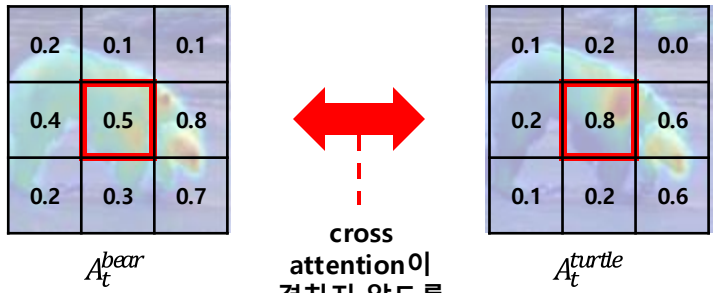
Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Enhancing Prompt Understanding
A-STAR

❖ **Missing object 문제 해결방법1 – Segregation loss**

- 각 객체 토큰 Cross attention이 겹치는 **Overlapping 최소화**
- Cross attention이 겹치지 않고 분리되도록 유도하는 segregation loss 제안

Prompt: " A **bear** and a **turtle** "



$$\mathcal{L}_{\text{seg}} = \sum_{\substack{m,n \in \mathcal{C} \\ \forall m > n}} \left[\frac{\sum_{ij} \min([\mathbf{A}_t^m]_{ij}, [\mathbf{A}_t^n]_{ij})}{\sum_{ij} ([\mathbf{A}_t^m]_{ij} + [\mathbf{A}_t^n]_{ij})} \right]$$

m: bear
n: turtle
t: time step
i,j: 위치

i = 2, j = 2

$$\frac{min\,(0.5,0.8)}{0.5+0.8} = \frac{0.5}{1.3}$$

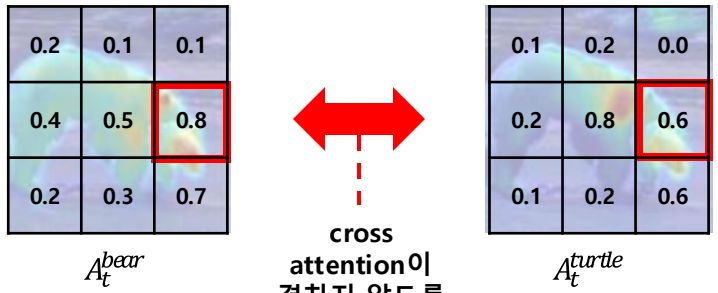
Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Enhancing Prompt Understanding
A-STAR

❖ Missing object 문제 해결방법1 – Segregation loss

- 각 객체 토큰 Cross attention이 겹치는 **Overlapping 최소화**
- Cross attention이 겹치지 않고 분리되도록 유도하는 segregation loss 제안

Prompt: " A **bear** and a **turtle** "



$$\mathcal{L}_{\text{seg}} = \sum_{\substack{m,n \in \mathcal{C} \\ \forall m > n}} \left[\frac{\sum_{ij} \min([\mathbf{A}_t^m]_{ij}, [\mathbf{A}_t^n]_{ij})}{\sum_{ij} ([\mathbf{A}_t^m]_{ij} + [\mathbf{A}_t^n]_{ij})} \right]$$

m: bear
n: turtle
t: time step
i,j: 위치

i = 2, j = 2 $\quad \frac{min\,(0.5,0.8)}{0.5+0.8} = \frac{0.5}{1.3}$

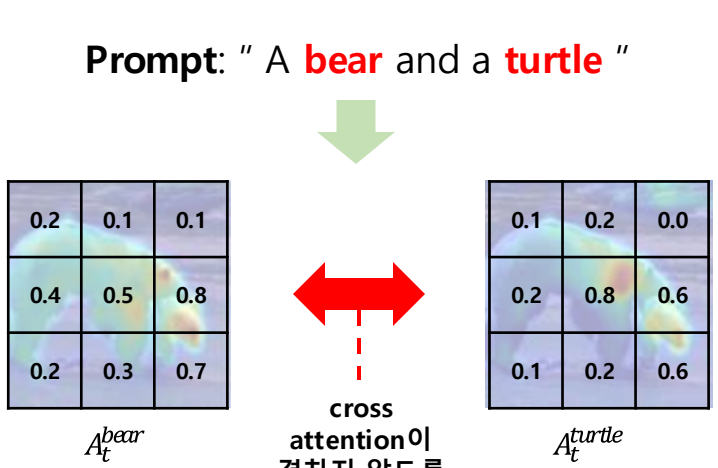
i = 2, j = 3 $\quad \frac{min\,(0.8,0.6)}{0.8+0.6} = \frac{0.6}{1.4}$

Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Enhancing Prompt Understanding
A-STAR

❖ Missing object 문제 해결방법1 – Segregation loss

- 각 객체 토큰 Cross attention이 겹치는 **Overlapping 최소화**
- Cross attention이 겹치지 않고 분리되도록 유도하는 segregation loss 제안



$$\mathcal{L}_{\text{seg}} = \sum_{\substack{m,n \in \mathcal{C} \\ \forall m > n}} \left[\frac{\sum_{ij} \min([\mathbf{A}_t^m]_{ij}, [\mathbf{A}_t^n]_{ij})}{\sum_{ij} ([\mathbf{A}_t^m]_{ij} + [\mathbf{A}_t^n]_{ij})} \right]$$

m: bear
n: turtle
t: time step
i,j: 위치

i = 2, j = 2 $\quad \frac{min\,(0.5,0.8)}{0.5+0.8} = \frac{0.5}{1.3}$

i = 2, j = 3 $\quad \frac{min\,(0.8,0.6)}{0.8+0.6} = \frac{0.6}{1.4}$

겹치는 부분에서 cross attention 값이 작은 객체는 생성되지 않도록 유도
→ Overlapping 최소화

Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Enhancing Prompt Understanding

A-STAR

❖ **Missing object 문제 해결방법2 – Retention loss**

- 객체 토큰의 Cross attention이 **영향력을 유지**하여 attention decay 방지
- 이전 시점 Cross attention A_m^{t+1} 과 현재 시점 Cross attention A_m^t 이 유사하도록 하는 Retention loss 제안

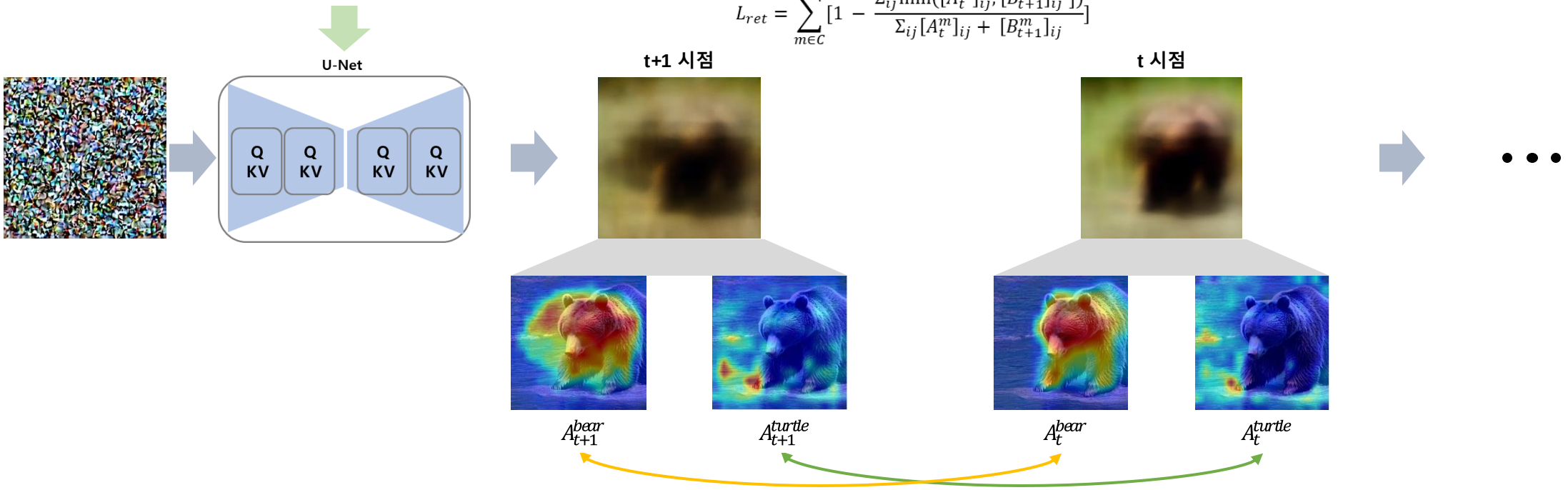
Enhancing Prompt Understanding
A-STAR

❖ **Missing object 문제 해결방법2 – Retention loss**

- 객체 토큰의 Cross attention이 **영향력을 유지**하여 attention decay 방지
- 이전 시점 Cross attention A_m^{t+1} 과 현재 시점 Cross attention A_m^t 이 유사하도록 하는 Retention loss 제안

Prompt: " A **bear** and a **fish** "

$$L_{ret} = \sum_{m \in C} [1 - \frac{\Sigma_{ij} \min([A_t^m]_{ij}, [B_{t+1}^m]_{ij})}{\Sigma_{ij} [A_t^m]_{ij} + [B_{t+1}^m]_{ij}}]$$



Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

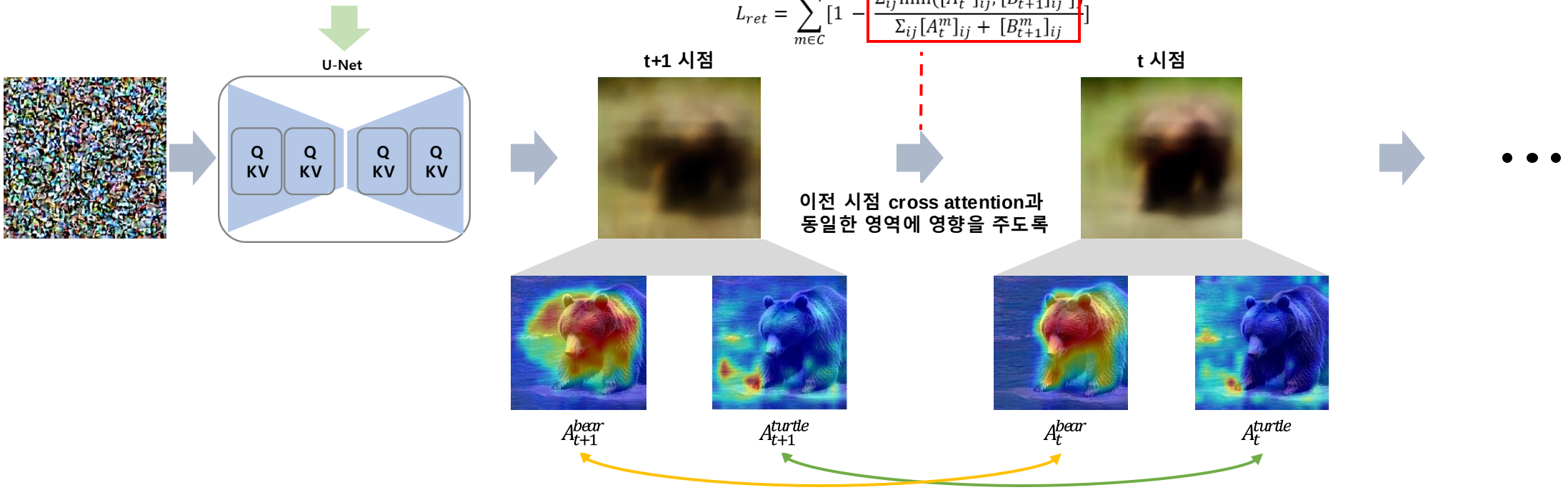
Enhancing Prompt Understanding
A-STAR

❖ **Missing object 문제 해결방법2 – Retention loss**

- 객체 토큰의 Cross attention이 **영향력을 유지**하여 attention decay 방지
- 이전 시점 Cross attention A_m^{t+1} 과 현재 시점 Cross attention A_m^t 이 유사하도록 하는 Retention loss 제안

Prompt: " A **bear** and a **fish** "

$$L_{ret} = \sum_{m \in C} [1 - \frac{\Sigma_{ij} \min([A_t^m]_{ij}, [B_{t+1}^m]_{ij})}{\Sigma_{ij} [A_t^m]_{ij} + [B_{t+1}^m]_{ij}}]$$



Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

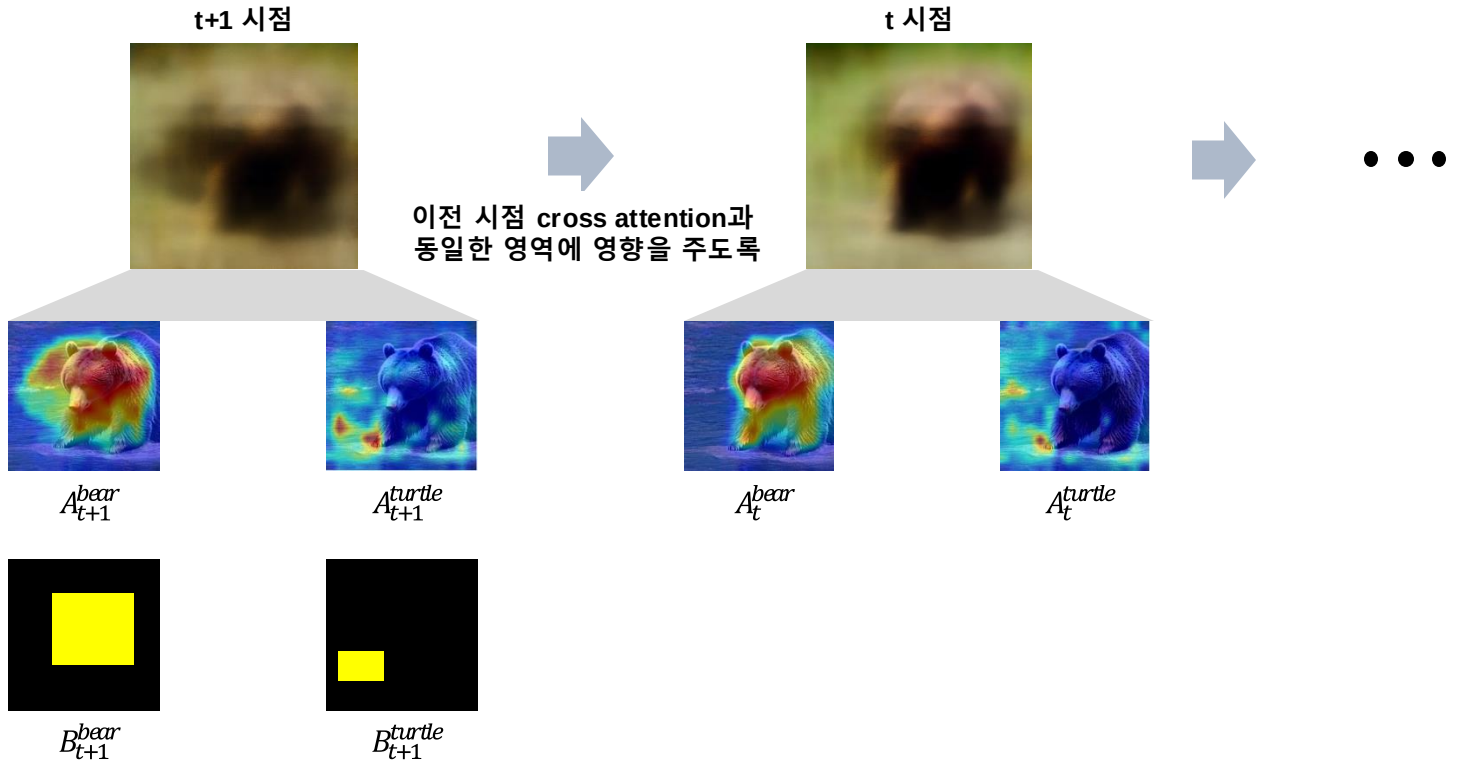
Enhancing Prompt Understanding
A-STAR

❖ Missing object 문제 해결방법2 – Retention loss

- 객체 토큰의 Cross attention이 **영향력을 유지**하여 attention decay 방지
- 이전 시점 Cross attention A_m^{t+1} 과 현재 시점 Cross attention A_m^t 이 유사하도록 하는 Retention loss 제안

$$L_{ret} = \sum_{m \in C} [1 - \frac{\Sigma_{ij} \min([A_t^m]_{ij}, [B_{t+1}^m]_{ij})}{\Sigma_{ij} [A_t^m]_{ij} + [B_{t+1}^m]_{ij}}]$$

m: 특정 객체 토큰
B: Cross attention binary mask



Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

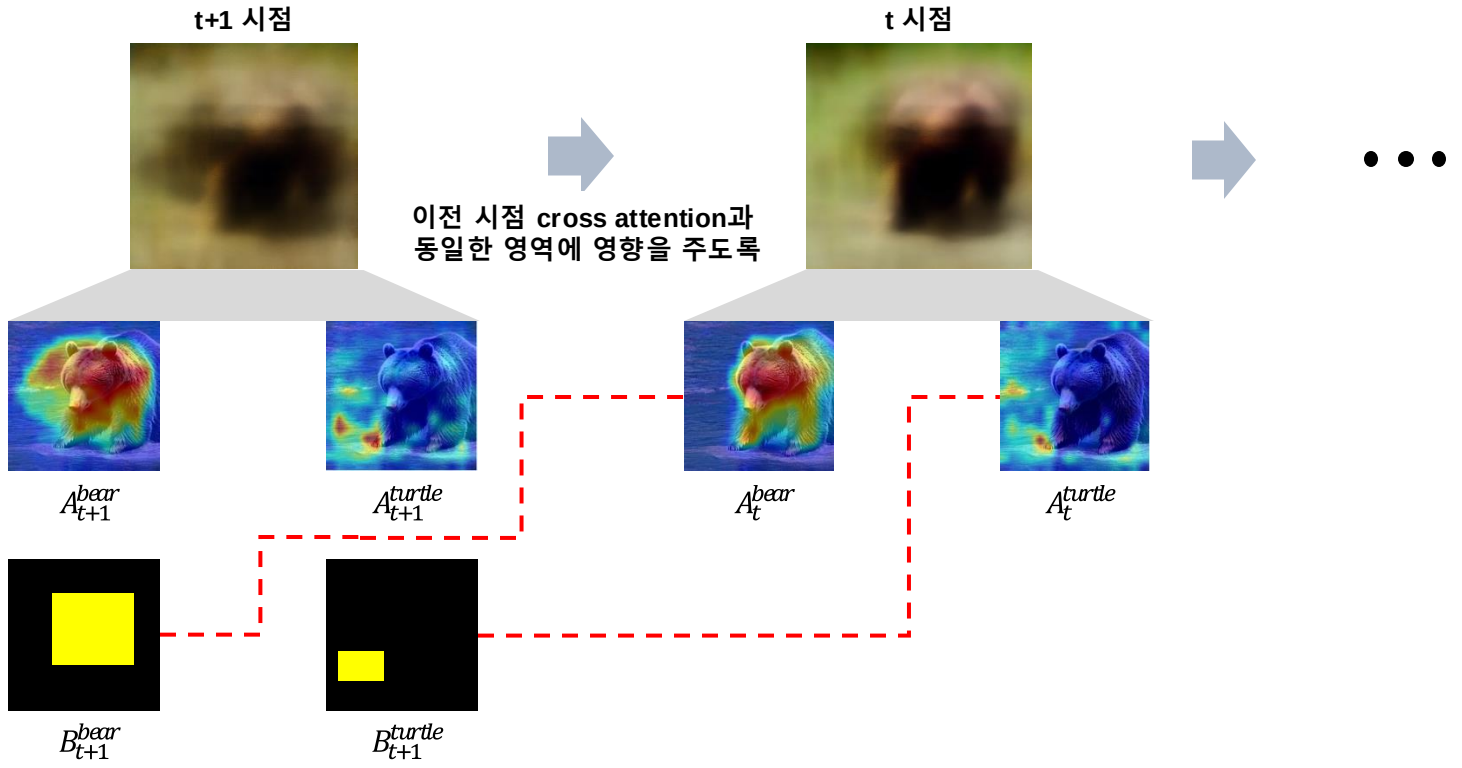
Enhancing Prompt Understanding
A-STAR

❖ Missing object 문제 해결방법2 – Retention loss

- 객체 토큰의 Cross attention이 **영향력을 유지**하여 attention decay 방지
- 이전 시점 Cross attention A_m^{t+1} 과 현재 시점 Cross attention A_m^t 이 유사하도록 하는 Retention loss 제안

$$L_{ret} = \sum_{m \in C} [1 - \frac{\Sigma_{ij} \min([A_t^m]_{ij}, [B_{t+1}^m]_{ij})}{\Sigma_{ij} [A_t^m]_{ij} + [B_{t+1}^m]_{ij}}]$$

m: 특정 객체 토큰
B: Cross attention binary mask



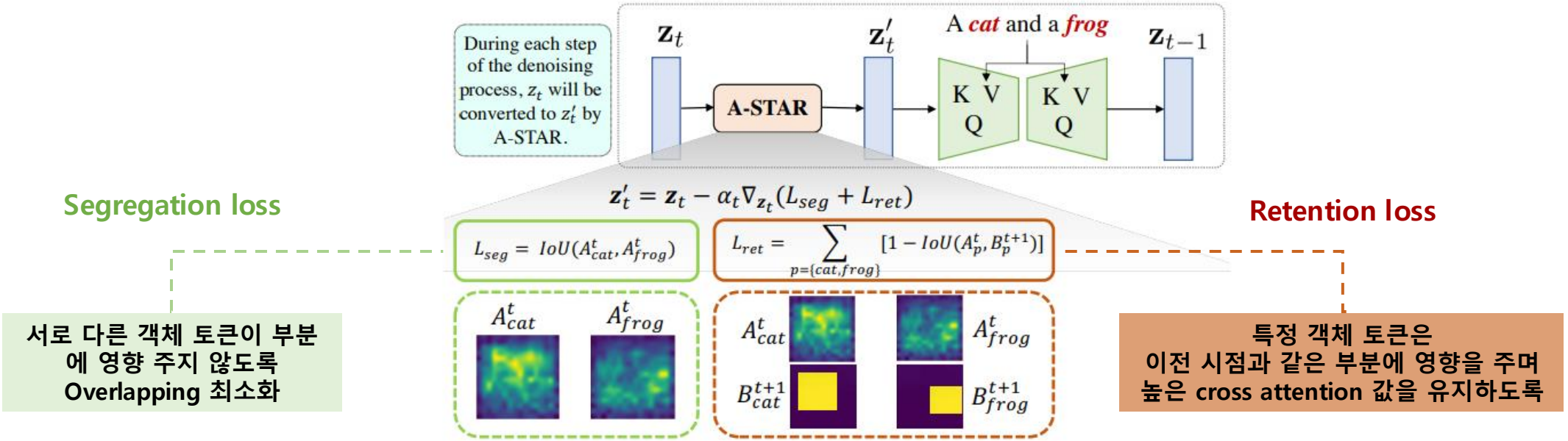
➢ 현재 시점과 이전 시점 cross attention 이 같은 부분에 영향 주도록
➢ cross attention 값을 높일 수 있도록

Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Enhancing Prompt Understanding
A-STAR

❖ **A-STAR:** Test-time **A**ttention **S**egrega**T**ion **A**nd **R**etention

- **Segregation Loss:** 서로 다른 객체 토큰이 다른 부분에 영향을 주도록 **cross attention overlap** 방지
- **Retention Loss:** 특정 객체 토큰의 cross attention 값을 유지하여 **attention decay** 방지



Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Enhancing Prompt Understanding
A-STAR

❖ 실험결과

- Stable Diffusion에 A-STAR 활용 여부에 따라 생성된 이미지를 비교
- A-STAR 활용시 Missing Object 문제 완화

Stable Diffusion | **With A-STAR**

Prompt: A bear and a **turtle**


Stable Diffusion | **With A-STAR**

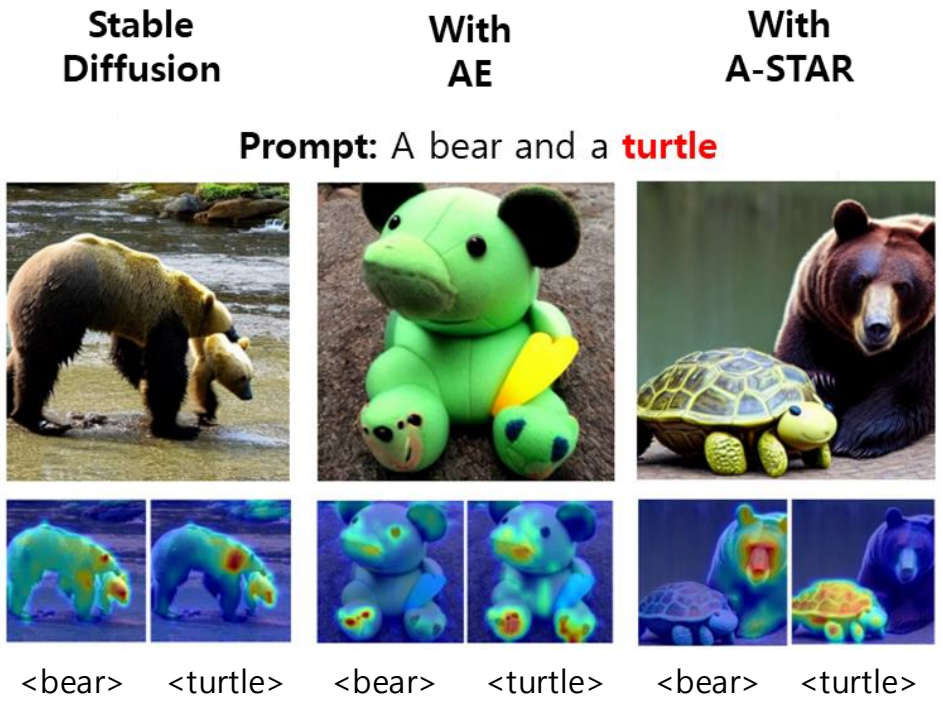
Prompt: A cat and a **frog**


Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Enhancing Prompt Understanding
A-STAR

❖ 실험결과

- 각 객체 토큰의 cross attention이 잘 분리되었으며, 각 객체 토큰은 높은 cross attention을 유지
- Stable Diffusion은 물론 Attend-and-Excite 보다 우수한 성능



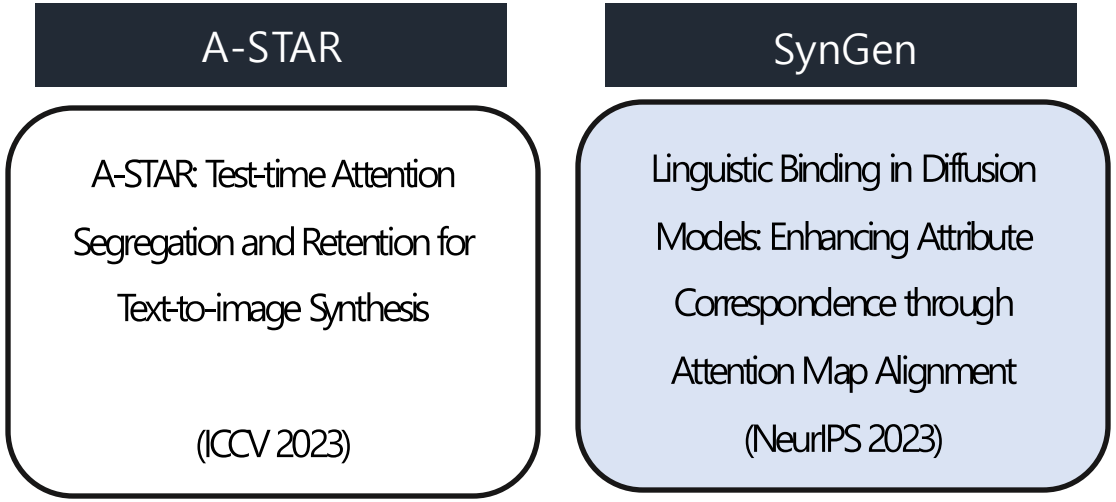
Method	Animal - Animal	Animal - Object	Object - Object
Stable [22]	0.76 (-7.9%)	0.78 (-7.7%)	0.77 (-6.5%)
Composable [13]	0.69 (-18.9%)	0.77 (-9.1%)	0.76 (-7.9%)
Structure [3]	0.76 (-7.9%)	0.78 (-7.7%)	0.76 (-7.9%)
Attend-Excite [2]	0.80 (-2.5%)	0.82 (-2.4%)	0.81 (-1.2%)
A-STAR	**0.82**	**0.84**	**0.82**

Table 1: Text-text similarities between the text prompts and BLIP-generated captions over the generated images.

입력 프롬프트와
생성된 이미지를 image captionining
모델을 사용해 만든 문장을 비교

Agarwal, Aishwarya, et al. "A-star: Test-time attention segregation and retention for text-to-image synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Enhancing Prompt Understanding



Enhancing Prompt Understanding

SynGen

❖ **SynGen: Syn**tax-Guided **Gen**eration **(NeurIPS 2023)**

- **Motivation:** 객체에 잘못된 속성이 부여되는 문제 – Attribute binding
- 객체에 해당되는 속성이 영향을 주지 못하거나, 잘못된 객체에 영향을 줌

Prompt: " A frog and a **brown** apple"

Prompt: " A **pink** sunflower and a **yellow** flamingo "

Rassin, Royi, et al. "Linguistic binding in diffusion models: Enhancing attribute correspondence through attention map alignment." Advances in Neural Information Processing Systems 36 (2024).

Enhancing Prompt Understanding

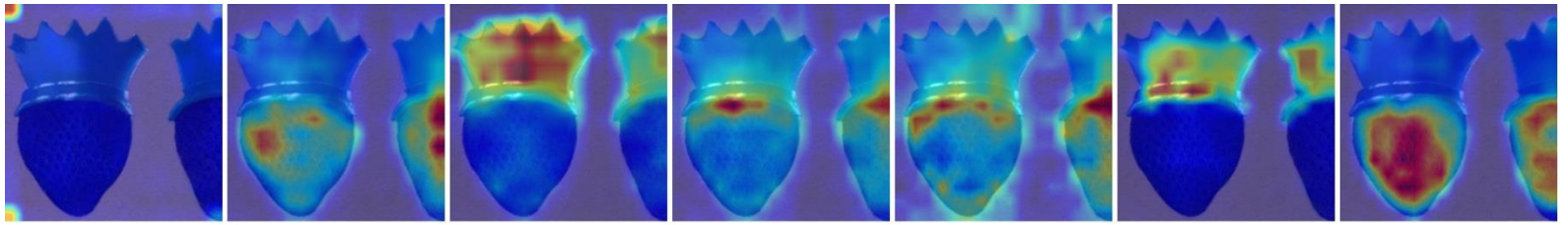
SynGen

❖ Attribute binding 발생 원인

- Cross attention map을 통해 각 단어가 이미지 어느 부분에 영향을 주는지 확인 가능 (**빨간색**일수록 강한 영향력)
- 속성이 **잘못된 객체에 영향**을 끼침 →디퓨젼 모델이 프롬프트내 **언어적 구조를 완벽하게 파악하지 못함**

Prompt: " A **red crown** and a **golden strawberry**"

생성된 이미지



< A > **< red >** **< crown >** < and > < a > **< golden >** **< strawberry >**

Enhancing Prompt Understanding

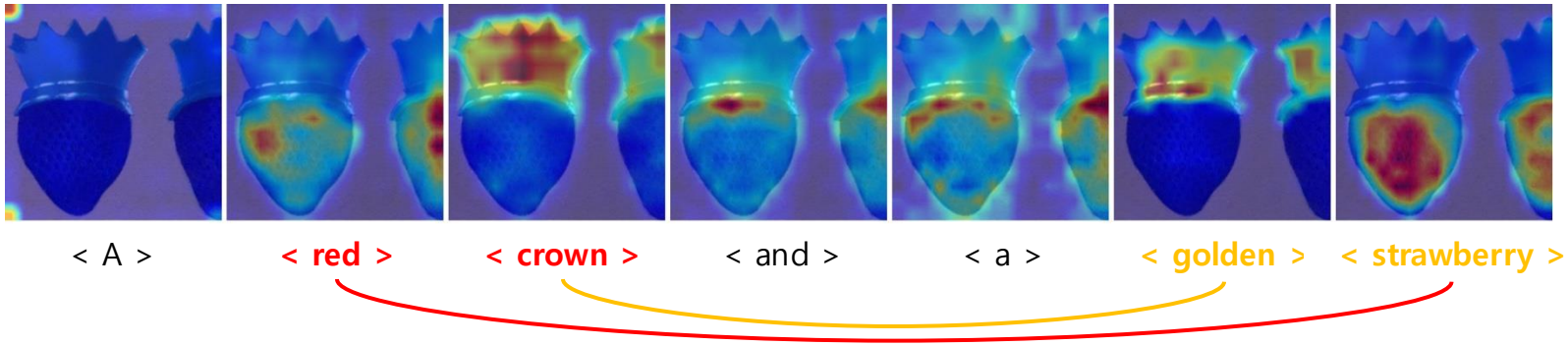
SynGen

❖ Attribute binding 발생 원인

- Cross attention map을 통해 각 단어가 이미지 어느 부분에 영향을 주는지 확인 가능 (**빨간색**일수록 강한 영향력)
- 속성이 **잘못된 객체에 영향**을 끼침 →디퓨젼 모델이 프롬프트내 **언어적 구조를 완벽하게 파악하지 못함**

Prompt: " A **red crown** and a **golden strawberry**"

생성된 이미지



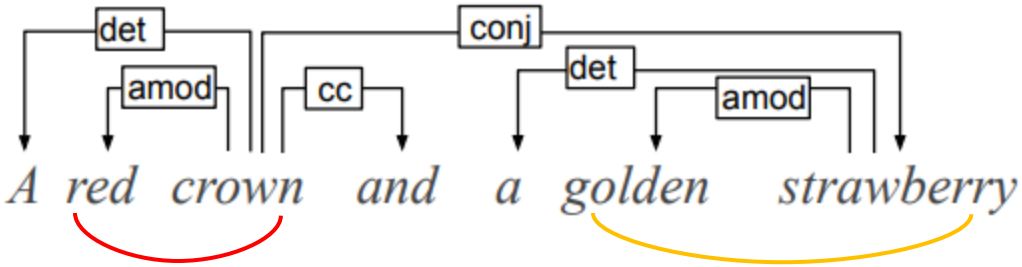
Cross attention map

Enhancing Prompt Understanding

SynGen

❖ 언어적 구조 이해: Parser

- 문장내 **객체(명사)와 속성(수식어)간 관계**를 식별 및 구분
- Parsing: 문장을 단어 조각들로 분해하고, 단어간의 관계를 파악하여 문법적 구성을 이해
- 언어적 구조를 파악하기 위해 transformer기반 spaCy parser 활용



< Parser를 사용해 문장 관계 파악 >

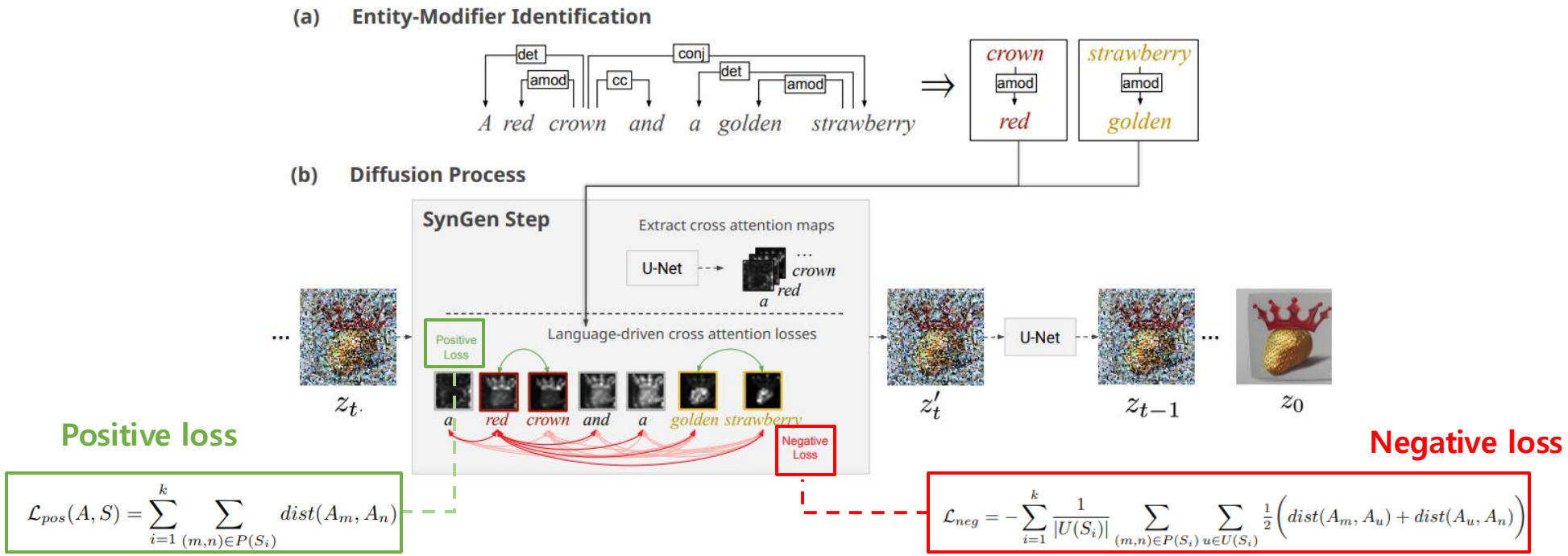
Matthew Honnibal and Ines Montani. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear, 2017.

Enhancing Prompt Understanding

SynGen

❖ **SynGen: Syn**tax-Guided **Gen**eration

- **Positive loss: 같은 세트**내 객체와 속성은 이미지내 **같은 부분에 영향**을 주도록
- **Negative loss: 다른 세트** 토큰들과는 이미지내 **다른 부분에 영향**을 주도록



Rassin, Royi, et al. "Linguistic binding in diffusion models: Enhancing attribute correspondence through attention map alignment." Advances in Neural Information Processing Systems 36 (2024).

Enhancing Prompt Understanding
SynGen

❖ **Attribute binding 해결방법1 – Positive loss**

- **같은 세트**내 객체와 속성은 **같은 부분에 영향**을 주도록
- 같은 세트내 토큰들간 cross attention **유사도를 최대화**하는 positive loss 제안

Prompt: " A **red crown** and a **golden strawberry**"

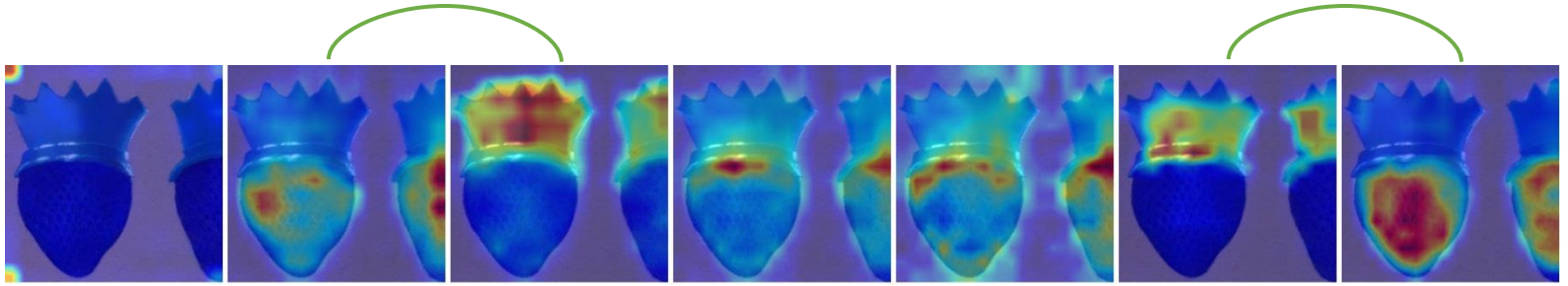
생성된 이미지

같은 세트 속성 토큰 cross attention

같은 세트 객체 토큰 cross attention

Positive loss

$$\mathcal{L}_{pos}(A,S) = \sum_{i=1}^{k} \sum_{(m,n)\in P(S_i)} dist(A_m, A_n)$$



< A > < red > < crown > < and > < a > < golden > < strawberry >

Enhancing Prompt Understanding
SynGen

❖ **Attribute binding 해결방법2 – Negative loss**

- **다른 세트** 토큰들과는 이미지내 **다른 부분에 영향**을 주도록
- 다른 세트 토큰들간 cross attention **유사도를 최소화**하는 negative loss 제안

같은 세트 속성 토큰 cross attention

다른 세트 토큰 cross attention

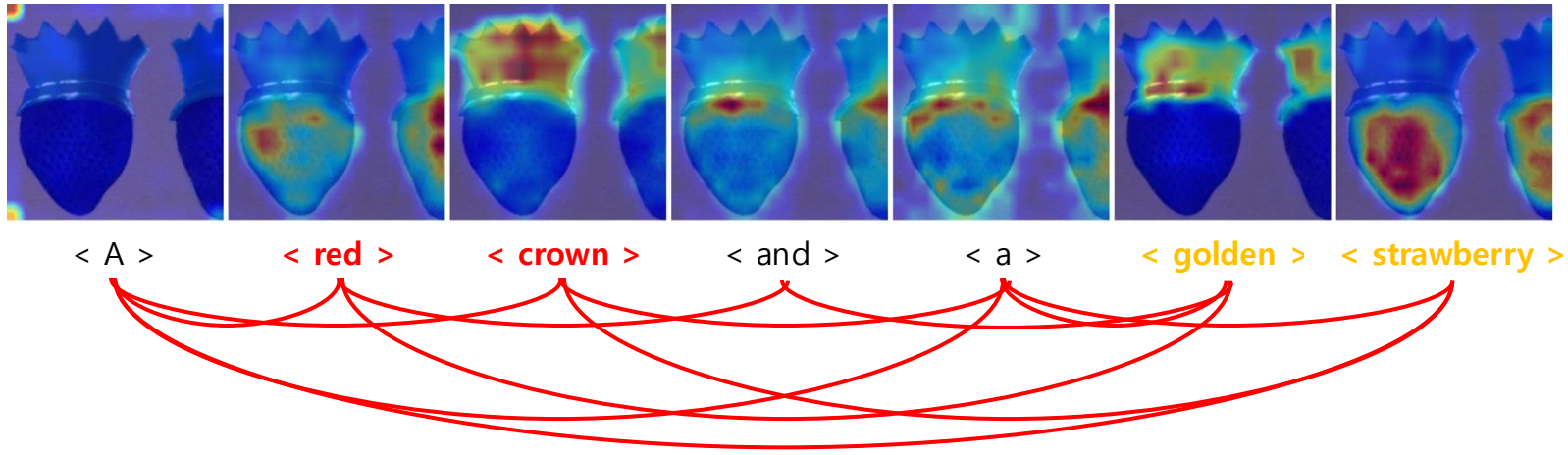
같은 세트 객체 토큰 cross attention

Prompt: " A **red crown** and a **golden strawberry**"

생성된 이미지

Negative loss

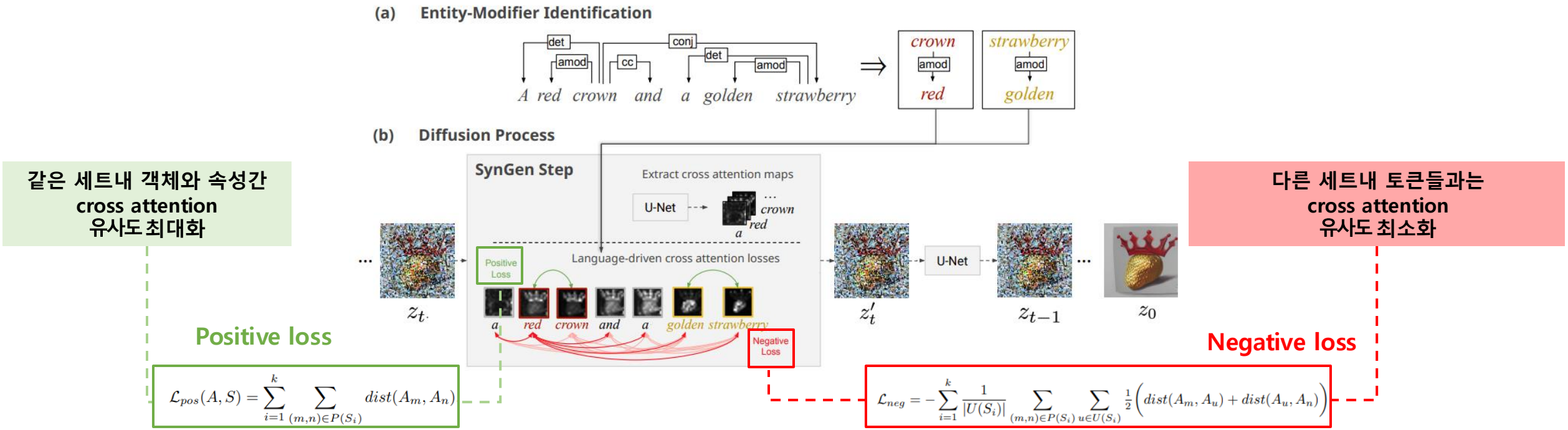
$$\mathcal{L}_{neg} = -\sum_{i=1}^{k} \frac{1}{|U(S_i)|} \sum_{(m,n)\in P(S_i)} \sum_{u\in U(S_i)} \frac{1}{2}\Big(dist(A_m, A_u) + dist(A_u, A_n)\Big)$$



Enhancing Prompt Understanding
SynGen

❖ **SynGen: Syn**tax-Guided **Gen**eration

- **Positive loss: 같은 세트**내 객체와 속성은 이미지내 **같은 부분에 영향**을 주도록
- **Negative loss: 다른 세트** 토큰들과는 이미지내 **다른 부분에 영향**을 주도록



Rassin, Royi, et al. "Linguistic binding in diffusion models: Enhancing attribute correspondence through attention map alignment." Advances in Neural Information Processing Systems 36 (2024).

Enhancing Prompt Understanding

SynGen

❖ 실험결과

- Stable Diffusion에 Syngen 활용 여부에 따라 생성된 이미지를 비교
- Syngen 활용시 Attribute binding 문제 완화

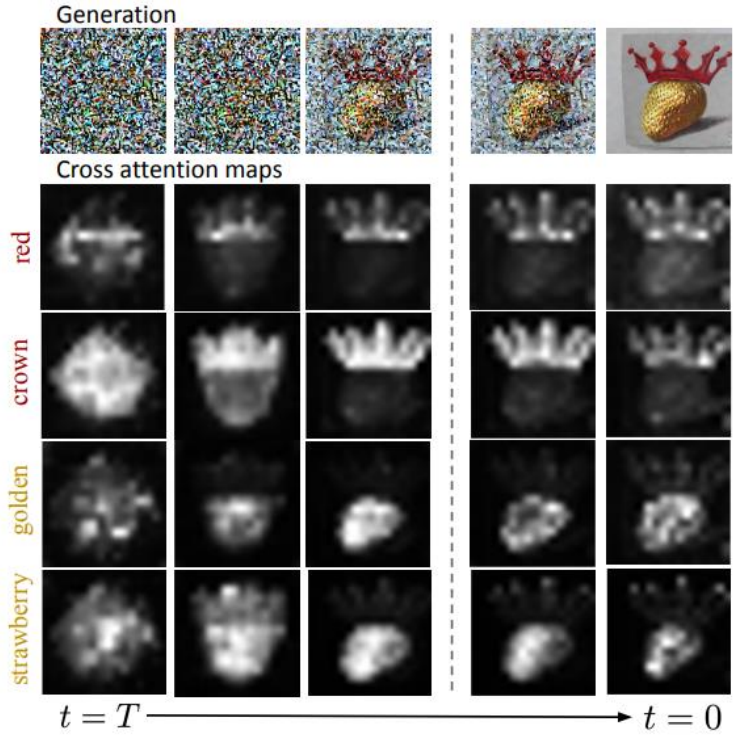
Rassin, Royi, et al. "Linguistic binding in diffusion models: Enhancing attribute correspondence through attention map alignment." Advances in Neural Information Processing Systems 36 (2024).

Enhancing Prompt Understanding

SynGen

❖ 실험결과

- Cross attention을 통해 객체와 속성이 이미지내 어느 부분에 영향을 주는지 파악
- 같은 세트내 단어(객체와 속성)들이 이미지상 같은 부분에 영향을 줌



Rassin, Royi, et al. "Linguistic binding in diffusion models: Enhancing attribute correspondence through attention map alignment." Advances in Neural Information Processing Systems 36 (2024).

Enhancing Prompt Understanding

SynGen

❖ 실험결과

- 세 가지 데이터 셋에 대해 디퓨젼 기반 방법론들과 정량적인 성능 비교
- Attribute binding 문제에서 우수한 성능을 보이며, missing object 문제에서도 준수

		객체에 알맞는 속성 매칭	객체에 잘못된 속성이 매칭	객체가 생성되지 않음
		Proper Binding ↑	Improper Binding ↓	Entity Neglect ↓
Dataset	Model			
A&E	SynGen (ours)	**94.76**	**23.81**	02.82
	A&E	81.90	63.81	**01.41**
	Structured Diffusion	55.71	67.62	21.13
	Stable Diffusion	59.05	68.57	20.56
DVMP (challenge set)	SynGen (ours)	**74.90**	**19.49**	16.26
	A&E	52.47	31.64	**10.77**
	Structured Diffusion	48.73	30.57	28.46
	Stable Diffusion	47.80	30.44	26.22
ABC-6K	SynGen (ours)	**63.68**	**14.37**	34.41
	A&E	56.26	26.43	**33.18**
	Structured Diffusion	51.47	29.52	34.57
	Stable Diffusion	52.70	27.20	36.57

Rassin, Royi, et al. "Linguistic binding in diffusion models: Enhancing attribute correspondence through attention map alignment." Advances in Neural Information Processing Systems 36 (2024).

Enhancing Prompt Understanding
Summary

❖ **Stable Diffusion의 문제점**

- 프롬프트를 완벽하게 반영하기는 어려움
- 이에 따라 missing object와 attribute binding 문제 발생

❖ **A-STAR**

- **객체가 생성되지 않는** missing object 문제를 해결하기 위해 고안된 방법론
- 다른 객체끼리는 cross attention이 **overlap 하지 않고, 높은 attention 값을 유지**하도록 latent update

❖ **SynGen**

- **객체와 속성이 잘못 매칭**되는 attribute binding 문제 해결하기 위해 고안된 방법론
- **같은 세트끼리는 유사하게, 다른 세트는 멀어지도록** latent update

감사합니다.