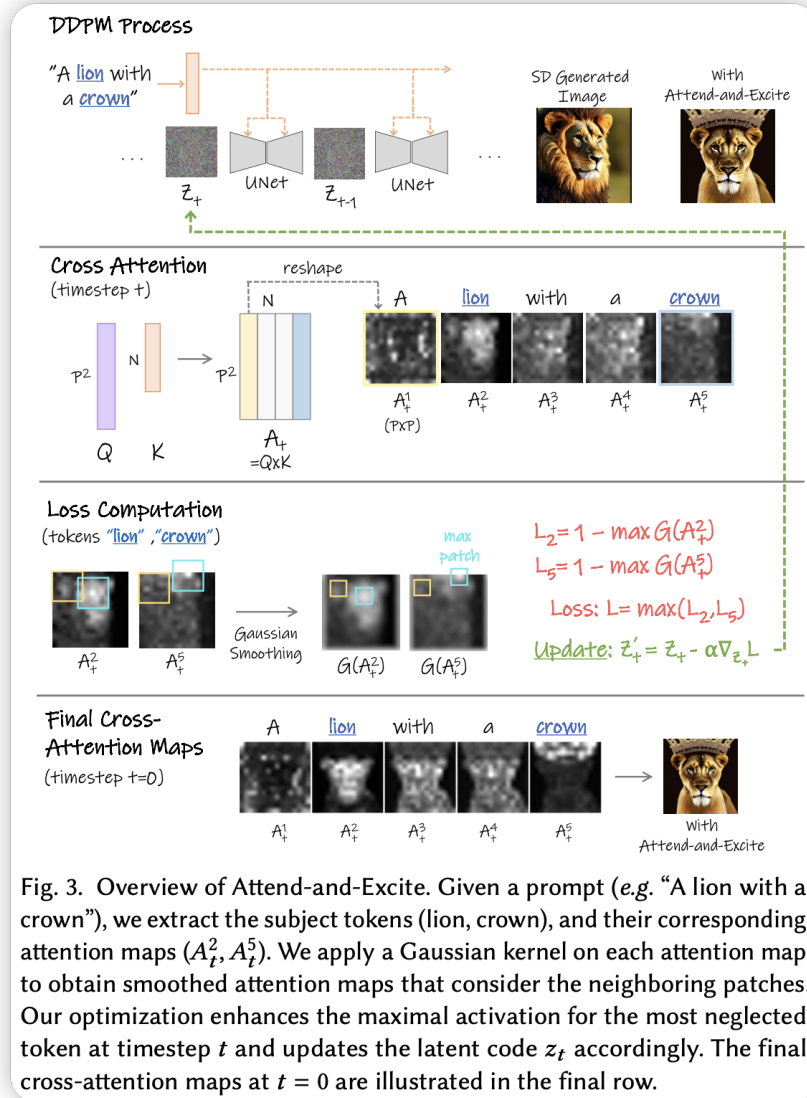




home
papers
study
life
About me

Attend-and-Excite

- Attend-and-Excite: Attention-Based Semantic Guidance for Text-to-Image Diffusion Models
- <https://arxiv.org/abs/2301.13826>
- TOG 2023



prompt prompt
Attend-and-Excite
token attention map

Algorithm 1 A Single Denoising Step using Attend-and-Excite

Input: A text prompt $\mathcal{P}$, a set of subject token indices $\mathcal{S}$, a timestep t , a set of iterations for refinement $\{t_1, \dots, t_k\}$, a set of thresholds $\{T_1, \dots, T_k\}$, and a trained Stable Diffusion model SD .
Output: A noised latent z_{t-1} for the next timestep

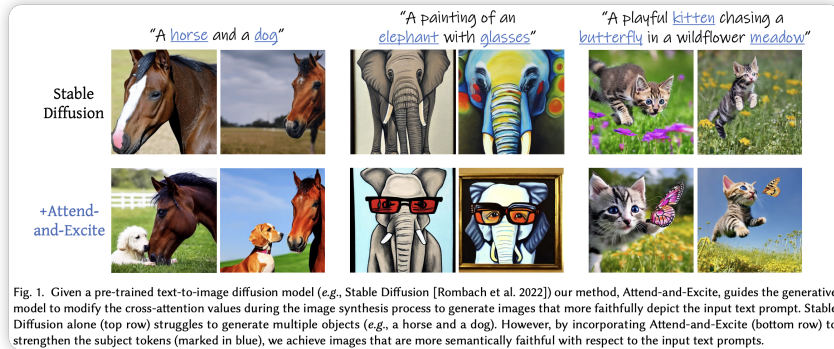
- $_, A_t \leftarrow SD(z_t, \mathcal{P}, t)$
- $A_t \leftarrow \text{Softmax}(A_t - \langle \text{so} \rangle)$
- for** $s \in \mathcal{S}$ **do**
- $A_t^s \leftarrow A_t[:, :, s]$
- $A_t^s \leftarrow \text{Gaussian}(A_t^s)$
- $\mathcal{L}_s \leftarrow 1 - \max(A_t^s)$
- end for**
- $\mathcal{L} \leftarrow \max_s(\mathcal{L}_s)$

```

9:  $z'_t \leftarrow z_t - \alpha_t \cdot \nabla_{z_t} \mathcal{L}$ 
10: if  $t \in \{t_1, \dots, t_k\}$  then ▷ If performing iterative refinement at  $t$ 
11:     if  $\mathcal{L} > 1 - T_t$  then
12:          $z_t \leftarrow z'_t$ 
13:         Go to Step 1
14:     end if
15: end if
16:  $z_{t-1, -} \leftarrow SD(z'_t, \mathcal{P}, t)$ 
17: Return  $z_{t-1}$ 

```

Page 6



- `python benchmark`
- `CLIP image-text` `CLIP text-text` `prompt` `caption` `BLIP2`
- `python`
- <https://yuval-alaluf.github.io/Attend-and-Excite/>

Newer

Older

2024-10-16

Mix-of-Show

2024-10-13

FaceCaption-15M

leicheng © 2022-2025

[Archive](#) [RSS feed](#) [GitHub](#) [Email](#) [QR Code](#)

Made with [Montaigne](#) and [bigmission](#)