

home papers

study life

About me

BLIP-Diffusion

- DDDDDBLIP-Diffusion: Pre-trained Subject Representation for Controllable Text-to-Image Generation and Editing
- DDDDDhttps://arxiv.org/abs/2305.14720
- NIPS 2024

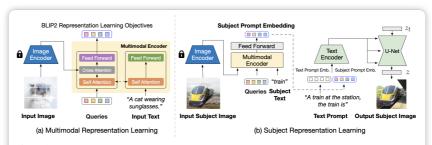


Figure 2: Illustration of the two-staged pre-training for BLIP-Diffusion. Left: in the multimodal representation learning stage, we follow prior work [12] and pretrain BLIP-2 encoder to obtain text-aligned image representation. Right: in the subject representation learning stage, we synthesize input images by composing subjects with random background. The BLIP-2 is then tasked to produce subject prompt embedding, which is later used by the latent diffusion model to generate output subject image. The image encoder remains frozen during pre-training.



Figure 1: Leveraging the pre-trained subject representation, BLIP-Diffusion enables subject-driven generation under efficient fine-tuning or zero-shot setups. The model can also serve as a foundation subject-driven text-to-image generation model and supports applications such as controlled generation and image editing, when combined with techniques such as ControlNet [1] and prompt-to-prompt [2].

- DDDLAION, COCO, Visual Genome, Conceptual Captions, OpenImage-V6
- DDDDINODCLIP-IDCLIP-T
- 0016 A100/bs16
- DDDhttps://github.com/salesforce/LAVIS/tree/main/projects/blip-diffusion

 Newer
 Older

 2024-09-23
 2024-09-19

 COW
 MasterWeaver

leicheng © 2022-2025

Archive RSS feed GitHub Email QR Code

Made with Montaigne and bigmission