進捗報告

1　やろうとしていたこと

ダ鳥獣ギ画 [1] のカエルのような画風の 3D モデルを Text to 3D 手法を用いて生成する.

2　用いた Text to 3D 手法

DreamFusionc[2] を用いた. 深層学習を用いて様々な画像から 3D シーンを生成する NeRF [3], Text to Image 手法の Imagen[4] に加えて拡散モデルを利用している. 図 1 に Dream Fusion の概略を示す.

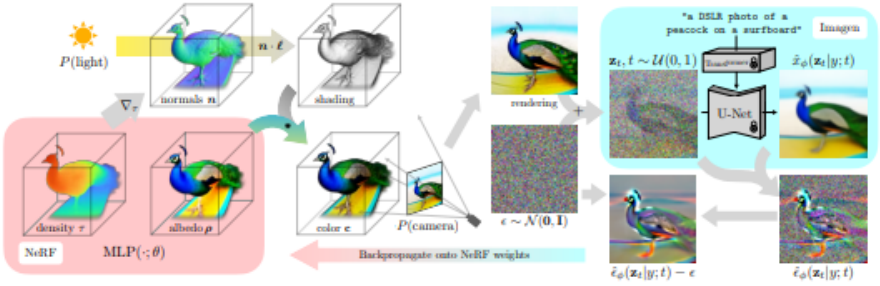

図 1: DreamFusion 概略図

3　下準備

ダ鳥獣ギ画のカエルの画風の Text to Image モデルを用意する必要がある. 今回は, Stable Diffusion のファインチューニング手法として Text Inversion[5] を用いた. 図 2 にファインチューニング前のモデルに ”an illustration of flog” というプロンプトを与えた際に生成された画像, 図 3 にファインチューニング後のモデルに ”an <giga-style >illustration of flog” というプロンプトを与えた際に生成された画像を示す. ダ鳥獣ギ画っぽい画風で生成できていることが分かる.

4　Text to 3D

README に従って 100 エポックほど学習しようとしたところ, 実験で用いていた aiserv では 20 エポック時点で cuda out of memory error が出てしまい, 十分に学習が進まなかった. 20 エポック時点 (学習が止まった段階) の動画があるため, それを流します. FT 前はもちろん, FT 後の拡散モデルを用いた場合でもカエルっぽい概形はできているが, FT 後の拡散モデルを用いて生成したモデルがダ鳥獣ギ画の画風を反映しているとは現時点では言えない.

図 2: DreamFusion 概略図

5 課題

- 研究の方向性
- マシンスペック

  DreamFusion のレポジトリの issue を見ていると GPU のメモリが最低 12GB 必要らしい．aiserv の GPU(Geforce RTX 3070) は 8GB だったため実験が途中で終わったと考えられる．

参考文献

[1] ダ鳥獣戯画. `https://chojugiga.com/`.

[2] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion, 2022.

[3] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *CoRR*, Vol. abs/2003.08934, , 2020.

[4] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho, David J Fleet, and Mohammad Norouzi. Photorealistic text-to-image diffusion models with deep language understanding, 2022.

図 3: DreamFusion 概略図

[5] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H. Bermano, Gal Chechik, and Daniel Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion, 2022.