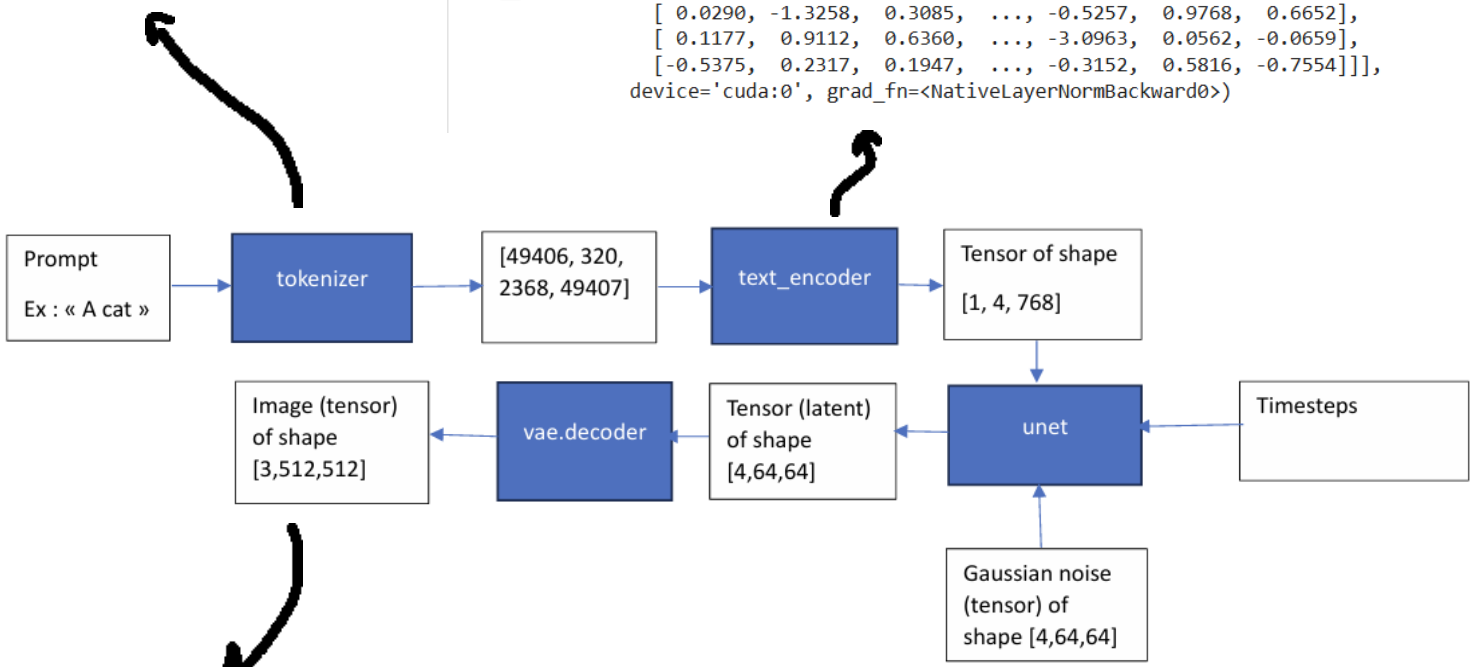
TokenCompose: Diving into multi-categories generation.

By Raphael Bernas & Maxime Corlay

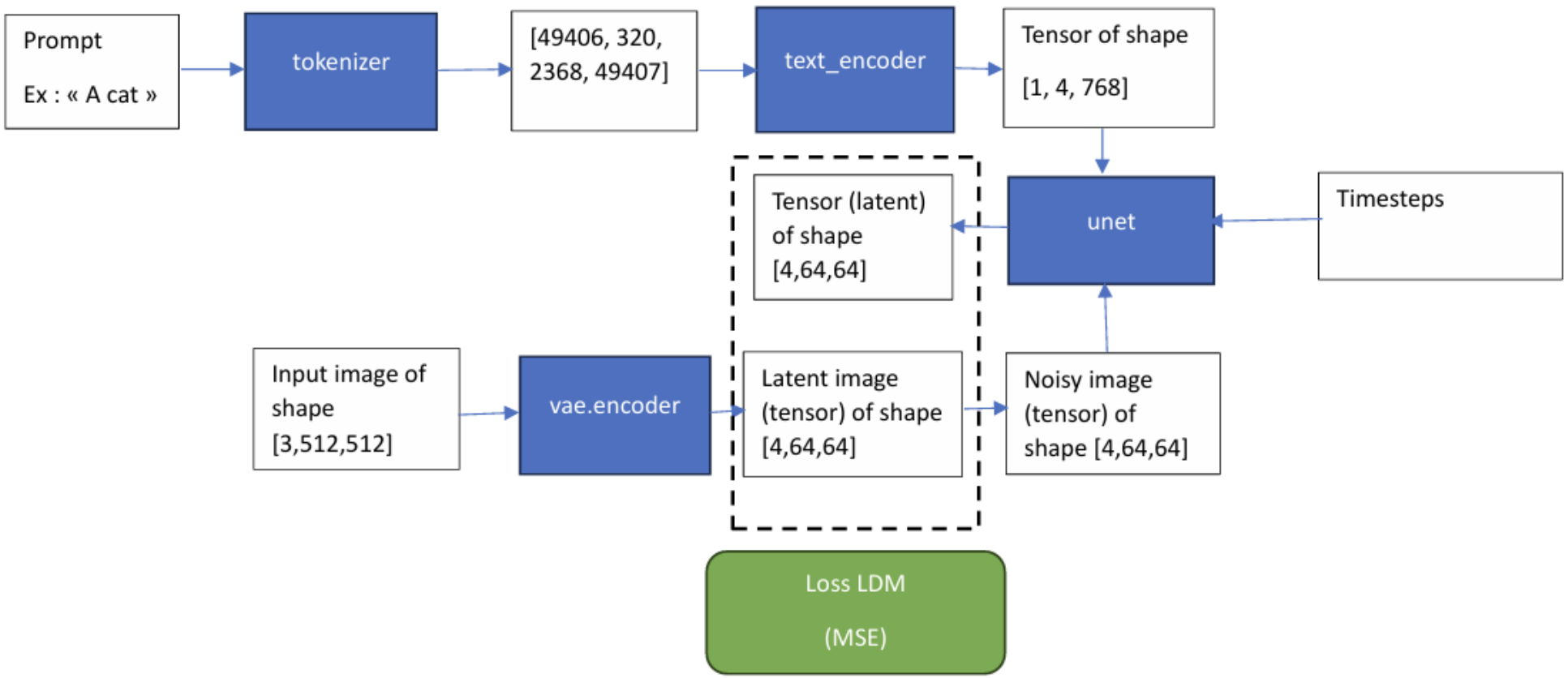
I.1/ Latent Diffusion Model

```
[14] tok=pipe.components['tokenizer'](prompt,return_tensors='pt')
     tok=tok.to(device)
     print(tok['input_ids'])

tensor([[49406,   320,  2368, 49407]], device='cuda:0')
```

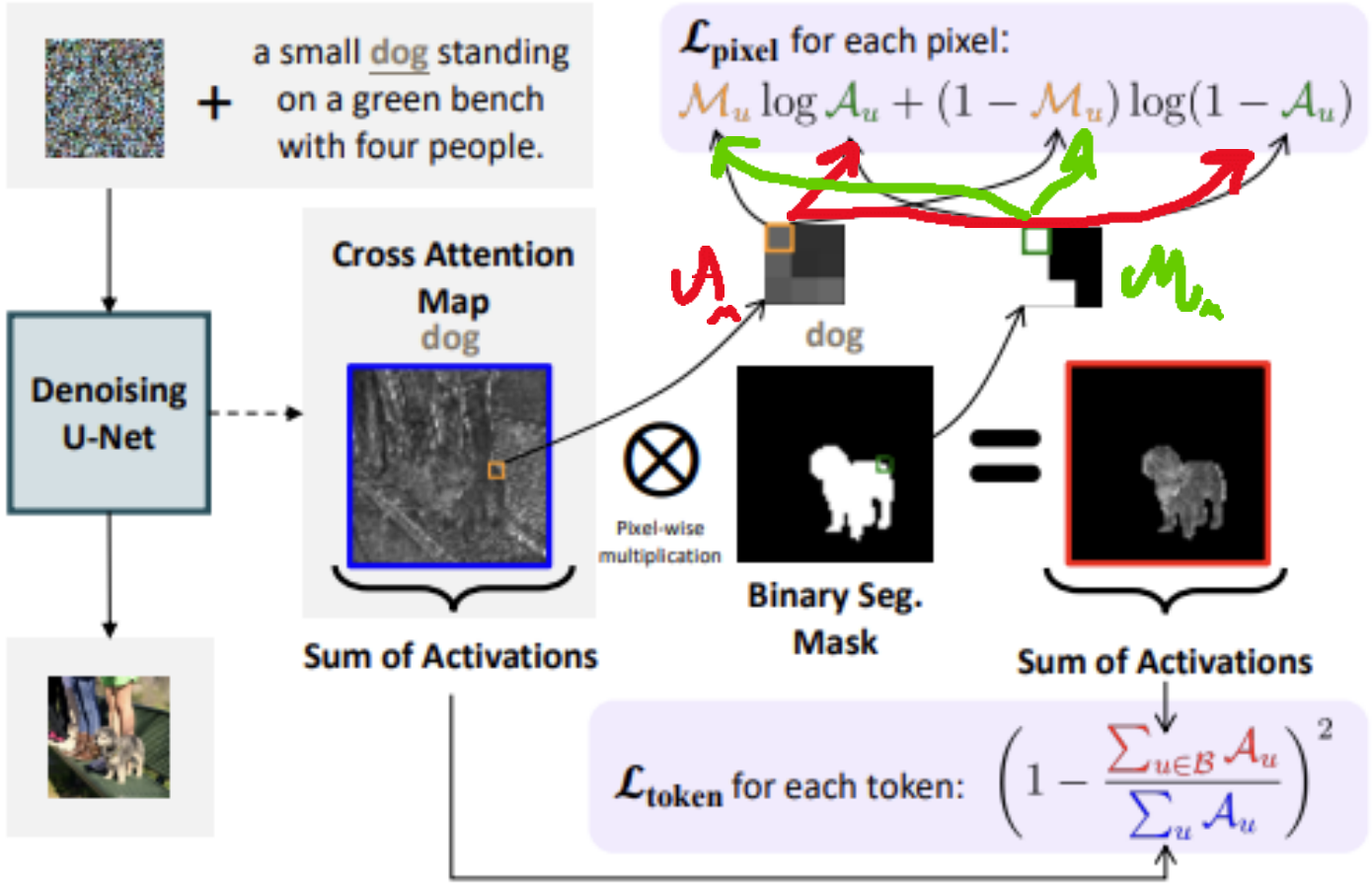
```
[15] pipe.components['text_encoder'](tok["input_ids"]).last_hidden_state

tensor([[[-0.3884,  0.0229, -0.0522,  ..., -0.4899, -0.3066,  0.0675],
         [ 0.0290, -1.3258,  0.3085,  ..., -0.5257,  0.9768,  0.6652],
         [ 0.1177,  0.9112,  0.6360,  ..., -3.0963,  0.0562, -0.0659],
         [-0.5375,  0.2317,  0.1947,  ..., -0.3152,  0.5816, -0.7554]]],
       device='cuda:0', grad_fn=<NativeLayerNormBackward0>)
```



During training...



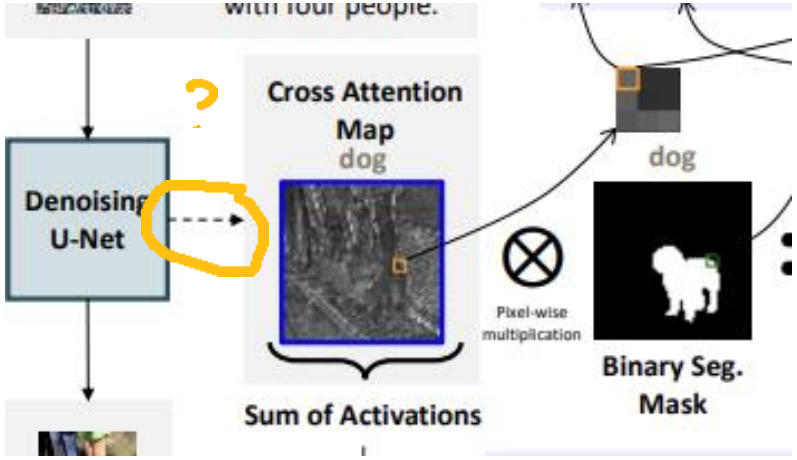
I.2/ Token Compose : A new training



Source : [2]

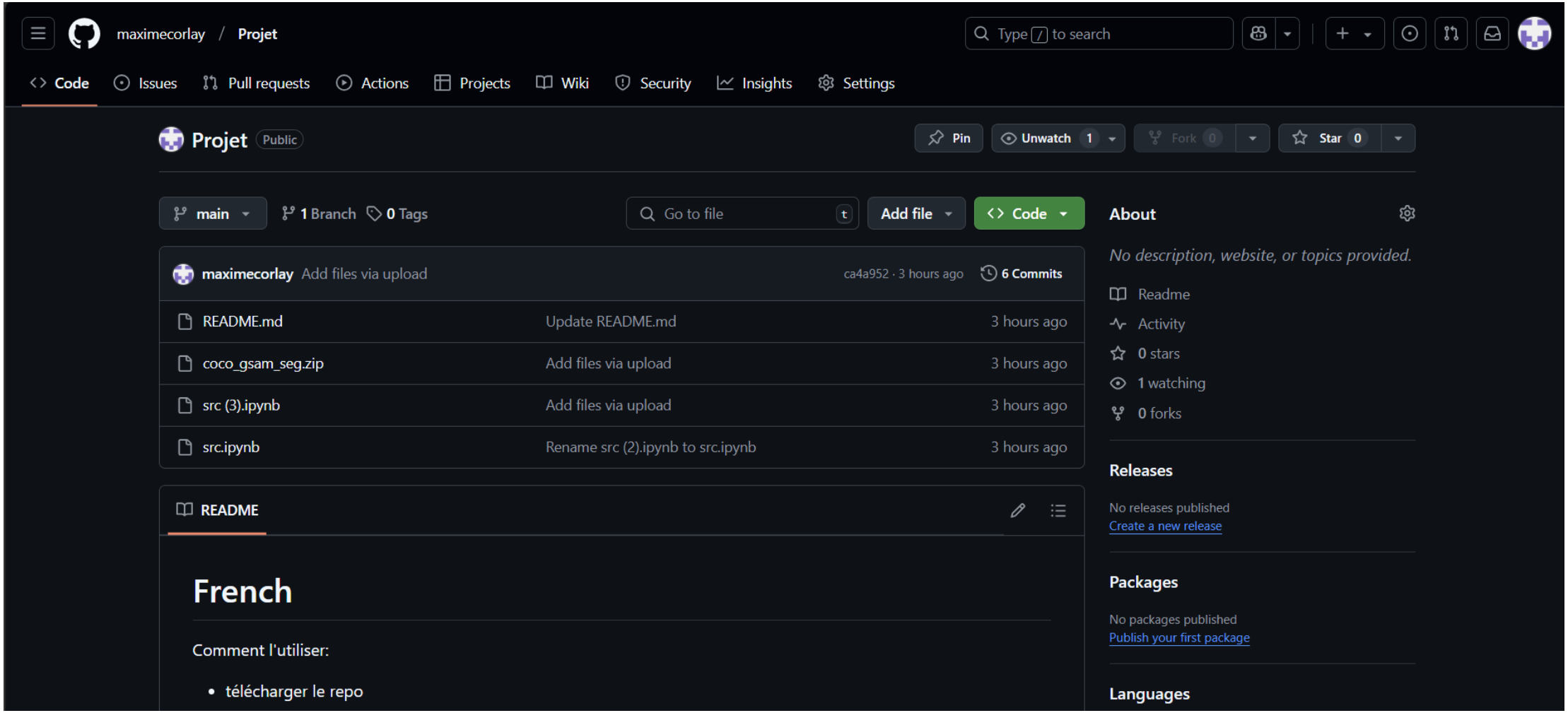
My contribution

- Understand what happens precisely in the model and in the paper
- Try to recode a simple version of TokenCompose based on my knowledge of the Hugging Face libraries and my reading of the paper
- Particularly, try to understand what happens here:



Source : [2]

My code


maximecorlay / Projet
Type / to search
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
Insights
Settings
Projet
Public
Pin
Unwatch 1
Fork 0
Star 0
main
1 Branch
0 Tags
Go to file
t
Add file
Code
maximecorlay Add files via upload
ca4a952 · 3 hours ago
6 Commits
README.md
Update README.md
3 hours ago
coco_gsam_seg.zip
Add files via upload
3 hours ago
src (3).ipynb
Add files via upload
3 hours ago
src.ipynb
Rename src (2).ipynb to src.ipynb
3 hours ago
README
French
Comment l'utiliser:
télécharger le repo
About
No description, website, or topics provided.
Readme
Activity
0 stars
1 watching
0 forks
Releases
No releases published
Create a new release
Packages
No packages published
Publish your first package
Languages

My conclusions

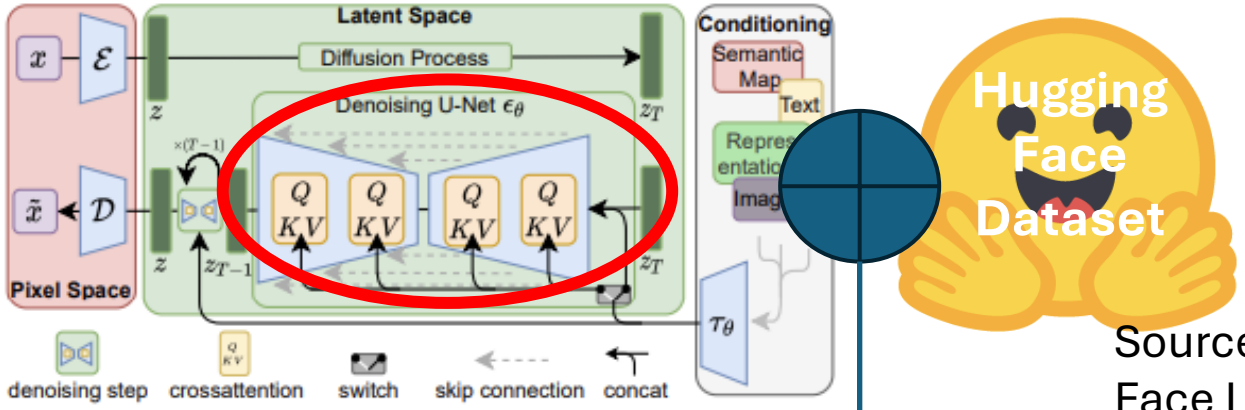
What I learned about TokenCompose:

- well-written paper except for a small error on figure 3 (the arrows are not connected correctly).
- I have realised that the real difficulty they have had to face is recovering the attentions (because I have not succeeded so far)... but I have understood that theoretically with a `hook` it should be possible (obj 1.2 well advanced)

General lessons:

- I learned how to manipulate Hugging Face (the platform).
- I realised that Hugging Face is designed for models to be used, but not necessarily fine-tuned. For example, SD14 has no forward function, only a __call__ function.
- You can easily access the template code on Hugging Face, but it is difficult to read because there are so many classes nested inside each other.

II.1/ Fine-Tuning Stable Diffusion

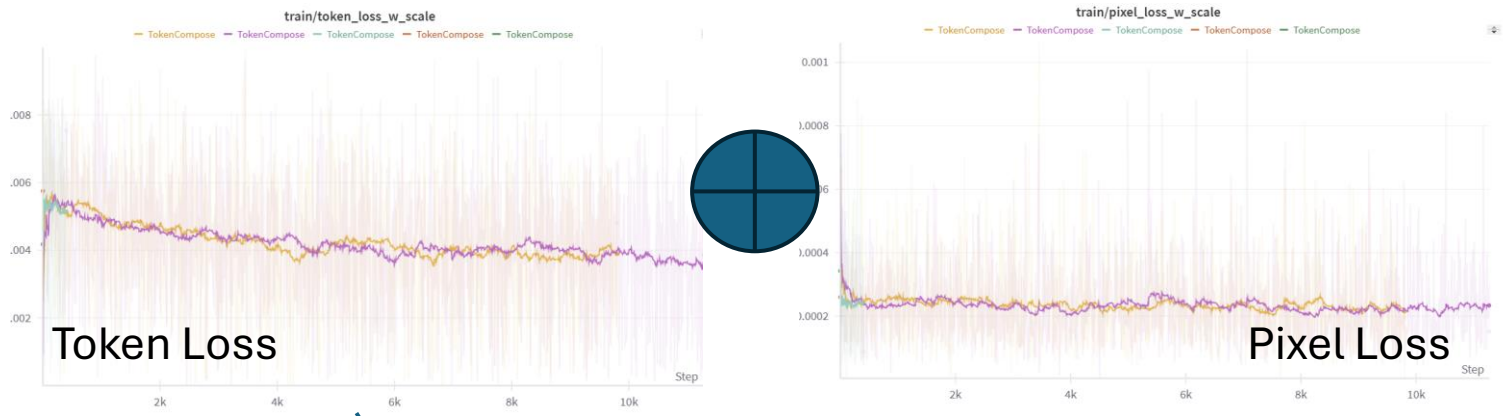


Source : [1]

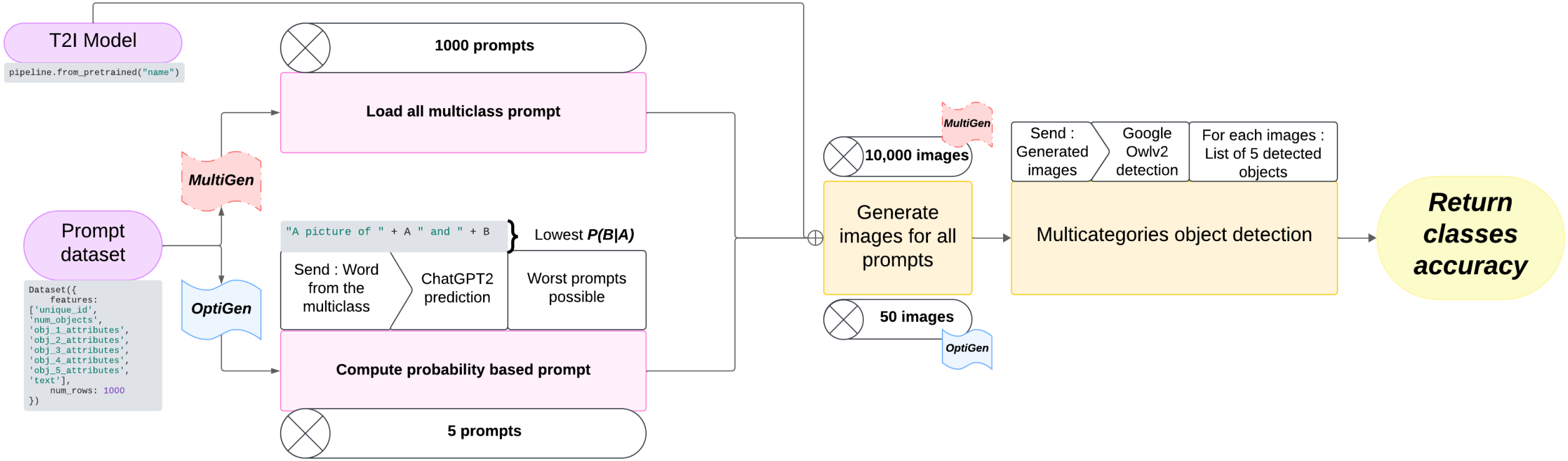
Source : Hugging Face Logo

'a picture of a man standing at the top of a mountain in romanticism style'

II.2 / Training with Token Compose



$$\min_{\text{w.r.t } \mathcal{L}_{Pixel}=Cst} \mathcal{L}_{DM} + \mathcal{L}_{Token}$$



II.3 / On the Benchmarking

Introducing : OptiGen !

II.3/ On the Benchmarking

Original

Model	MG2	MG3	MG4
SD1.4	0.90	0.55	0.05
SD1.4TK	0.95	0.60	0.20
SD1.4TK2000	0.95	0.45	0.25
SD3.5 (large)	1.00	1.00	1.00

0.91 0.51 0.11
0.98 0.76 0.28

Table 1: MultiGen Benchmarking on four models.

Not very useful

Model	Acc
SD1.4	0.52
SD1.4TK	0.96
SD1.4TK2000	0.74
SD3.5 (large)	0.90

More accurate,
Less data dependant

Table 2: OptiGen Benchmarking on four models.

Reproducible results !

Raphael Bernas
Raphael-Bernas

Currently a master's student in machine learning field. My master 2 is MVA (Mathematiques, Vision, Apprentissage) at Ecole Normale Supérieur de Paris-Saclay.

Edit profile

Training_TokenCompose Public

A notebook to test on COLAB the code from "TOKENCOMPOSE: Text-to-Image Diffusion with Token-level Supervision" by Zirui Wang , Zhizhou Sha , Zheng Ding Yilin Wang and Zhuowen Tu (https://github.com...

Jupyter Notebook · Apache License 2.0 · Updated 10 hours ago

Star

OptiGen_benchmark Public

OptiGen is a MultiGen complementary and computationally-free benchmark for text-to-image generation models. Its main objective is to test a T2I model in the worst condition possible for multi-categ...

Jupyter Notebook · Apache License 2.0 · Updated 13 hours ago

Star

Stable-Diffusion-Fine-Tuning-UNet- Public

Notebook on how to fine tune Stable Diffusion for almost any Hugging Face dataset

Jupyter Notebook · Updated last week

Star

[1] : “ TOKENCOMPOSE: Text-to-Image Diffusion with Token-level Supervision”; Zirui Wang, Zhizhou Sha et al.

[2] : “High-Resolution Image Synthesis with Latent Diffusion Models”; Robin Rombach, Andreas Blattmann et al.

Thanks !