

Projeto de Inteligência Computacional

Diffusion For Text To Image Generation : InstaFlow

João Cláudio Paco, M13709

Estagiário de Investigação em Universidade Kimpa-Vita em Angola

2024

INTRODUÇÃO

Os modelos de difusão revolucionaram a geração de texto para imagem com sua qualidade e criatividade excepcionais. No entanto, sabe-se que seu processo de amostragem em múltiplas etapas é lento, muitas vezes exigindo dezenas de etapas de inferência para obter resultados satisfatórios. Tentativas anteriores de melhorar a velocidade de amostragem e reduzir custos computacionais através da destilação não tiveram sucesso na obtenção de um modelo funcional de one-step. Neste projeto, exploramos um método recente chamado :RectifiedFlow, que, até agora, só foi aplicado a pequenos conjuntos de dados.

A escolha deste tema se fundamenta na importância crescente da síntese de imagens baseada em texto. Foi escolhido o método InstaFlow :

One Step is Enough for High-Quality Diffusion-Based Text-to-Image Generation super rápido que a técnica de última geração anterior, destilação progressiva, por uma margem significativa (em FID), notavelmente, o treinamento do InstaFlow custa apenas 199 dias de GPU A 100.

MÉTRICAS DE AVALIAÇÃO

1. Tabela 1 pontuação de(a) FID e CLIP no MS COCO 2017

Method	Inf. Time	FID-5k	CLIP
SD 1.4(25 step)[70]	0.88s	22.8	0.315
(1)(Pre)2-RF(25 step)	0.88s	22.1	0.313
PD(1 step)[58]	0.09s	37.2	0.275
SD 1.4+Distill	0.09s	40.9	0.255
(Pre)2-RF(1 step)	0.09s	68.3	0.252
(2)(Pre)2-RF+Distill	0.09s	31.0	0.285

Tabela 1: (a) MS COCO 2017

2. Tabela 2 pontuação de (b) FID no MS COCO 2014

Method	Inf. Time	FID-30k
SD*[70]	0.2.9s	9.62
(3)(Pre)2-RF(25 step)	0.88s	13.4
SD 1.4+Distill	0.09s	34.6
((4)Pre)2-RF+Distill	0.09s	20.0

Tabela 2: (b) MS COCO 2014

1. Legenda:

- **FID (Frechet Inception Distance)** :
- Uma métrica de avaliação de qualidade de imagens em comparação com um conjunto de dados de referência. Valores menores de FID indicam uma melhor qualidade das imagens geradas.
- **CLIP (Contrastivel Language-ImagePretrainig)** :
- Uma técnica de aprendizado de máquina que associa imagens e texto
- **SD: (Stable Diffusion)** :
- Um método ou modelo no contexto de geração de imagens a partir do texto.
- **Inf. Timee (Inference Time)** :
- Tempo necessário para realizar uma inferência ou uma previsão com o modelo.
- **PD (ProgressiveDistillation)**

2. Resultados Experimentais

(1) O (Pre) 2-RectifiedFlow :

Pode gerar(5k->5000) imagens realistas que produzem FID semelhante de (22,1- 22.3) com SD 1.4 usando 25 step dentro de uma Inf. Time de 0.88s.

(2) O (Pre) 2-Rectified Flow+Distill:

Obtém um FID de 31.0, com um SD 1.4+Distill, superando o melhor modelo SD de uma etapa anterior(FID=37.2) da Progressive Distillation com muito menos custo de treinamento(1 step), dentro de um Inf. Time de 0.9s.

(3) O (Pre) 2 Rectified Flow+Distill:

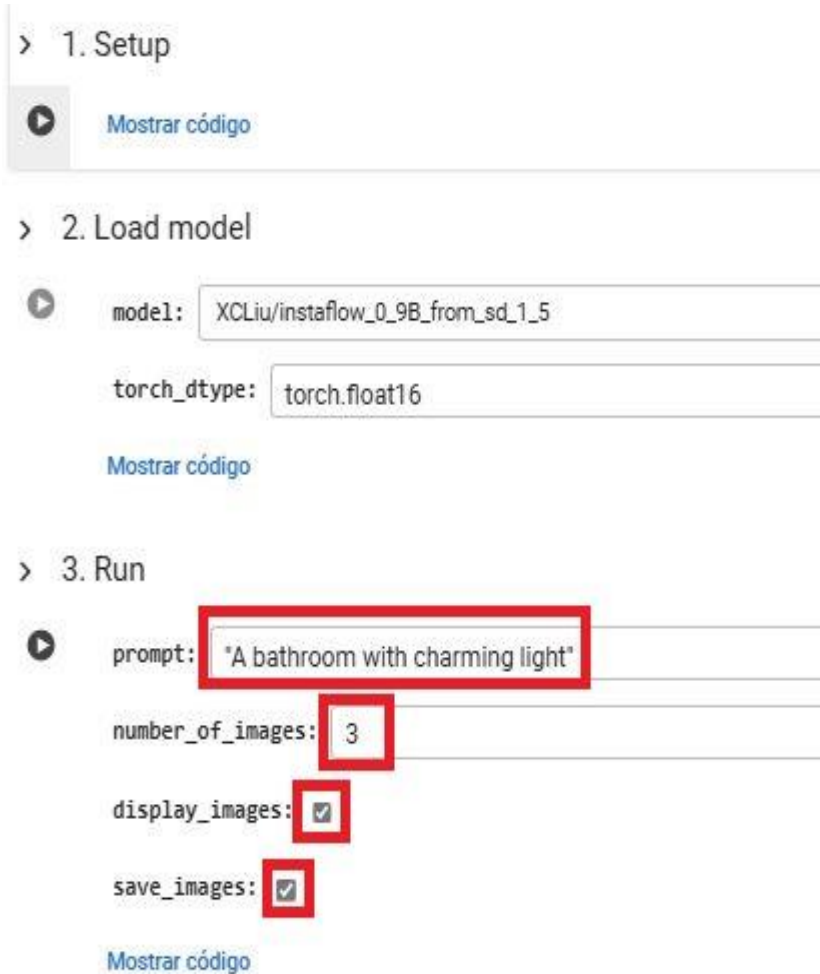
Tem vantagem notável(FID = 20.0) em comparação com Directed Distillation SD 1.4 + Distill (FID=34.6

(4) O (Pre) - Rectified Flow :

Tem pior desempenho(FID=13.4) do que o SD original(FID=9.62) devido a insuficiência de treinamento, indicando a eficácia da operação de reFlow.

AVALIAÇÃO COMPARATIVA

1. Interface do InstaFlow-0.98 2. Teste experimentais



The screenshot displays the InstaFlow-0.98 web interface. It is organized into three main sections: '1. Setup', '2. Load model', and '3. Run'. Each section has a play button icon and a 'Mostrar código' (Show code) link. In the '3. Run' section, the 'prompt' field contains the text 'A bathroom with charming light'. The 'number_of_images' field is set to 3. The 'display_images' and 'save_images' checkboxes are both checked. The '1. Setup' and '2. Load model' sections are partially visible, showing a 'model' field with the value 'XCliu/instafLOW_0_9B_from_sd_1_5' and a 'torch_dtype' field set to 'torch.float16'.

```
> 1. Setup
▶ Mostrar código

> 2. Load model
▶ model: XCLiu/instafLOW_0_9B_from_sd_1_5
  torch_dtype: torch.float16
  Mostrar código

> 3. Run
▶ prompt: "A bathroom with charming light"
  number_of_images: 3
  display_images: ☒
  save_images: ☒
  Mostrar código
```

Correndo o código dos métodos estudados InstaFlow e Text2im usando os mesmos dados em ambos aplicativos, para gerar as imagens partindo do texto fornecido, com os parâmetros seguintes :

Prompt : "A **big dog**"

Número of image: 2

Display images : **ativado**

Save images: **ativado**

E obtivemos os resultados seguintes:

100%  1/1 [00:00<00:00, 10.84it/s]
time: 0.42400693893432617 s




/content/imgs/image0.png

100%  1/1 [00:00<00:00, 10.11it/s]

time: 0.30594754219055176 s



/content/imgs/image1.png

100%  27/27 [00:09<00:00, 3.03it/s]



3. Os detalhes sobre os resultados experimentais:

(GPU: T4) de back-end do Google Colab Compute Engine em Python 3 RAM do sistema: 3.4 / 12.7 GB RAM da GPU : 4.6 / 15.0 GB Disco : 32.8 / 78.2 GB	(GPU: T4) de back-end do Google Colab Compute Engine em Python 3 RAM do sistema: 3.4 / 12.7 GB RAM da GPU : 4.6 / 15.0 GB Disco : 32.8 / 78.2 GB
Aplicativo : InstaFlow	Aplicativo : Text2im
Imagem nº1 :	Imagem nº1 e 2 :
Indicador de progresso: 00:00;00:00 Iterações : 10.84 it por segundo Tempo : 0.42400693893432617 s	Indicador de progresso: 00:09;00:00 Iterações : 3.03 it por segundo Tempo :
Imagem nº2 :	Imagem nº :
Indicador de progresso: 00:00;00:00 Iterações : 10.11 it por segundo Tempo : 0.38594754219055176 s	

4. Discussão dos resultados obtidos:

- O **InstaFlow**, gera as duas imagens uma por uma, e cada uma com as suas métricas enquanto o **Text2im** gera as duas imagens simultaneamente;

- Em **InstaFlow** o indicador de progresso para as ambas imagens marca 00:00<00:00, não atinge 1 segundo: 0.42400693893432617 s para a 1ª imagem e 0.38594 para a 2ª imagem enquanto o **Text2im** que indica 00:09<00:00, significa fez 9 segundos para gerar as duas imagens;
- O **InstaFlow** gera a 1ª e 2ª imagem em 10.84 e 10.11 iterações enquanto **Text2im** gera as duas imagens simultaneamente em 3.03 iterações

5. Conclusão sobre discussão dos resultados:

- Por gerar imagens separadamente e cada uma com as suas métrica, o **InstaFlow** tem mais performance comparando ao **Text2im**;
- Em termos da latência, o **Text2im** gera simultaneamente as duas imagens em 00:09<00:00 enquanto o **InstaFlow** gera duas imagens uma por uma em 00:00<00:00, tem assim mais performance;
- Em termos de número de iterações, o **InstaFlow** gera a 1ª e 2ª imagem em 10.84 e 10.11 iterações enquanto **Text2im** gera as duas imagens simultaneamente em
- 3.03 iterações, tem assim mais performance.

IMPLEMENTAÇÃO DE ALGUMAS ALTERAÇÕES AOS CÓDIGOS

1. Códigos originais:

2. Load model

```
#@title 2. Load model
import torch

model = 'XCLiu/instafLOW_0_9B_from_sd_1_5' # @param ["XCLiu/instafLOW_0_9B_from_sd_1_5", "XCLiu/instafLOW_0_9B_from_sd_1_5"]
torch_dtype = torch.float16 # @param [torch.float16, torch.float32]{type:"r

from pipeline_rf import RectifiedFlowPipeline

pipe = RectifiedFlowPipeline.from_pretrained("XCLiu/instafLOW_0_9B_from_sd_1_5")
### switch to torch.float32 for higher quality
pipe.requires_safety_checker = False
pipe.safety_checker = None
pipe.to("cuda") ### if GPU is not available, comment this line

clear_output()
```

1.1 Descrição :

Disponibiliza unicamente o dispositivo GPU como o tipo de tempo de execução.

1.2 Problema:

Assim sendo caso tiver CPU configurado como o tipo de tempo de execução :

Alterar tipo de tempo de execução

Tipo de tempo de execução

Python 3

Acelerador de hardware ?



CPU



T4 GPU



A100 GPU



L4 GPU



V100 GPU (deprecated)



TPU (deprecated)



TPU v2

Quer aceder a GPUs premium? [Compre unidades de computação adicionais](#)

O InstaFlow gera o erro :

```
> 3. Run

prompt: "A big dog"

number_of_images: 2

display_images: ☒

save_images: ☒

Mostrar código

-----
NameError                                Traceback
<ipython-input-1-de378b3181a1> in <cell line: 13>()
    13 for i in range(0,number_of_images):
    14     time_0=time.time()
--> 15     images = pipe(prompt=prompt,
    16                     num_inference_steps=1,
    17                     guidance_scale=0.0).images

NameError: name 'pipe' is not defined
```

2. Códigos alterados:

2.1 Códigos :

2. Load model

```
#@title 2. Load model
import torch

model = 'XCLiu/instafLOW_0_9B_from_sd_1_5' # @param ["XCLiu/instafLOW_0_9B_from_sd_1_5", "XCLiu/instafLOW_0_9B_from_sd_1_5"]
torch_dtype = torch.float16 # @param [torch.float16,torch.float32]{type:"r

from pipeline_rf import RectifiedFlowPipeline

pipe = RectifiedFlowPipeline.from_pretrained("XCLiu/instafLOW_0_9B_from_sd_1_5", torch_dtype=torch_dtype)
### switch to torch.float32 for higher quality
pipe.requires_safety_checker = False
pipe.safety_checker = None
#pipe.to("cuda") ### if GPU is not available, comment this line
#Verificar se o GPU está disponível sanão usar o CPU
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
# Mover o pipeline para dispositivo adequado
pipe.to(device)

clear_output()
```

2.2 Descrição :

- O código verifica se há uma GPU disponível
- Se houver, define o dispositivo como GPU(cuda)
- Senão houver, define o dispositivo como CPU
- Move o pipeline do modelo para o dispositivo adequado, garantindo que as operações subsequentes sejam executadas no dispositivo correto (GPU ou CPU).

2.3 Resultados:

Aproveita-se o poder computacionais da GPU (se disponível). Se uma GPU nã tiver disponível. o modelo será movido para CPU (embora tem tempo de execução mais longo comparado ao GPU), garantindo que o código funcione em qualquer máquina, independentemente da presença de uma GPU, **\textbf{com isso erradica-se o erro gerado anteriormente devido a indisponibilidade da GPU}**. Isso é uma prática comum para aproveitar o hardware disponível da melhor forma possível, otimizando a performance do aplicativo.

CONCLUSÃO

Neste projeto, exploramos a geração de imagens a partir de texto utilizando o método de descrição textuais no contexto de aplicações criativas e práticas. Em nossa análise, apresentamos dois (2) métodos recentes e inovadores : InstaFlow e Text2im na área de difusão para geração de imagens.

Realizamos uma avaliação comparativa desses métodos, considerando as suas abordagens, vantagens e limitações.

Adicionalmente, implementamos uma alteração ao código existente do InstaFlow, baseada nas percepções obtidas durante a análise dos métodos recentes. Essa modificação, teve como objetivo aprimorar os tipos de tempo de execução entre GPU e CPU de modo a se aplicar alternativamente.

Os resultados experimentais e comparativos obtidos entre os dois modelos distintos InstaFlow e Text2im mostraram que as técnicas de difusão de imagens a partir de descrições textuais são ambas eficazes, mas com eficiência relativa em termos de (1) latência em geração de imagens, (2) o número de iterações} para tal geração (2) e (3) a técnica utilizada.

Concluimos que o uso de métodos de difusão no InstaFlow oferece um avanço significativo em relação às abordagens tradicionais