图文生成进展及其应用

王浩帆

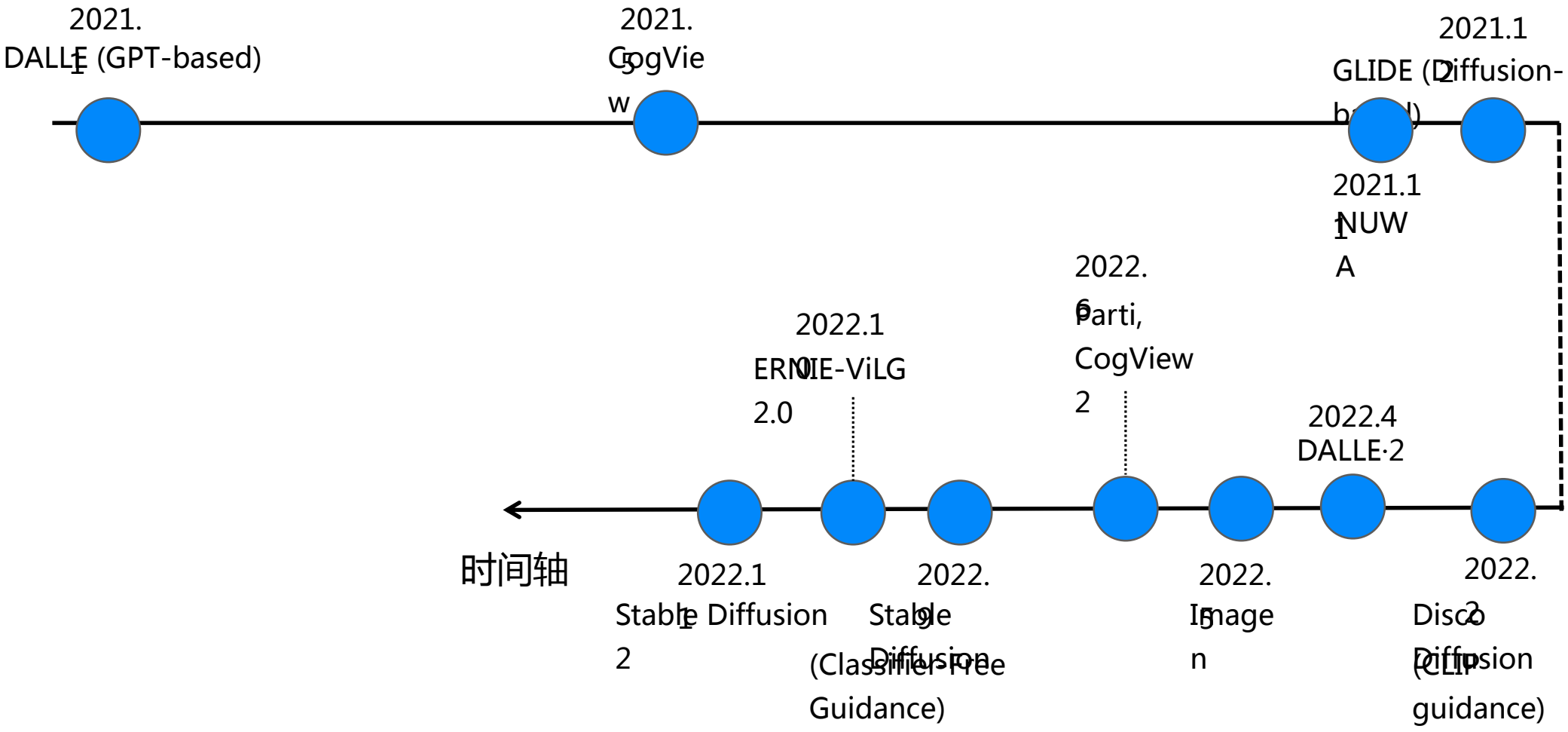
2022年12月30日

haofanwang.github.io

概要

- 发展轨迹
- 技术方案
- 相关应用
- 当前挑战

文本-图像生成模型

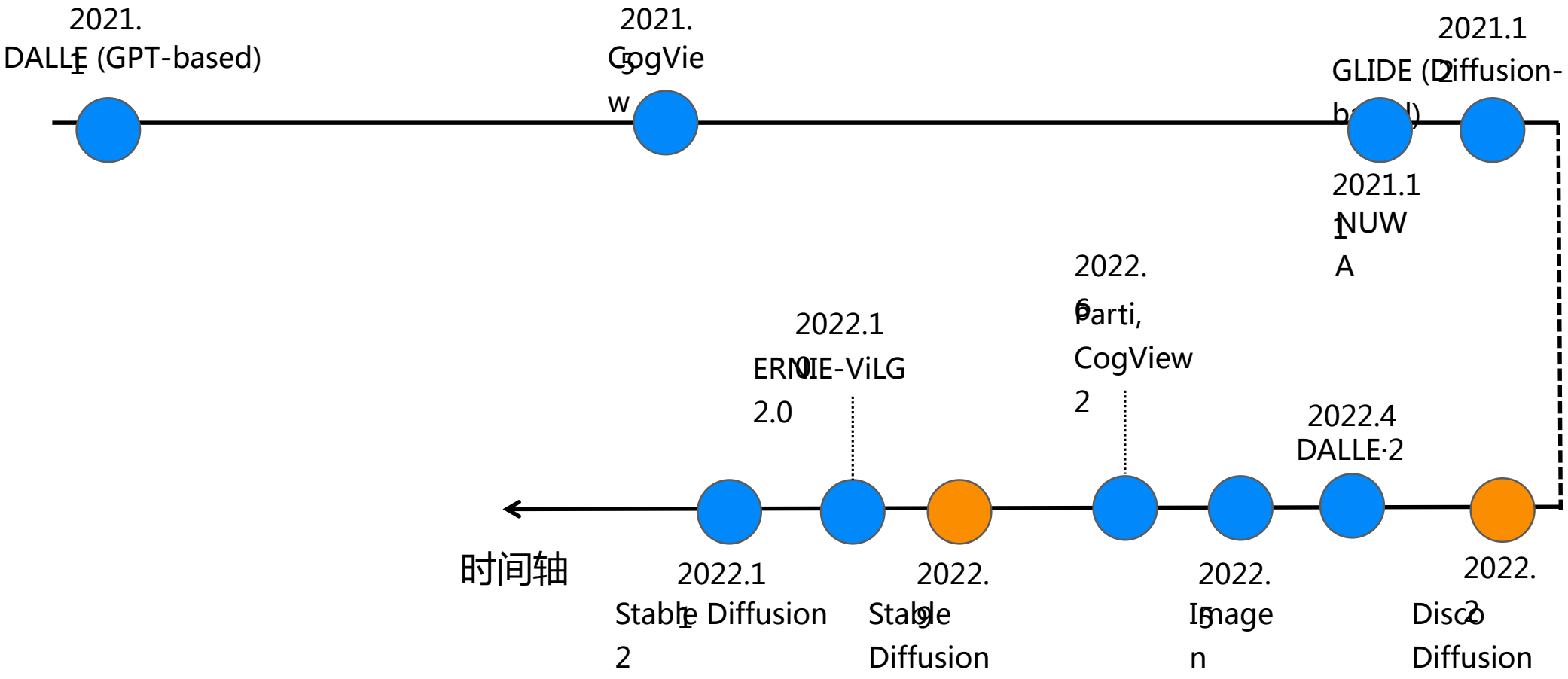


Google Brain: Imagen, Parti

OpenAI: DALLE, GLIDE, DALLE·2

StabilityAI: Stable Diffusion, Stable Diffusion 2

两个重要节点



Google Brain: Imagen, Parti

OpenAI: DALLE, GLIDE. DALLE·2

StabilityAI: Stable Diffusion, Stable Diffusion 2

闭源与开源

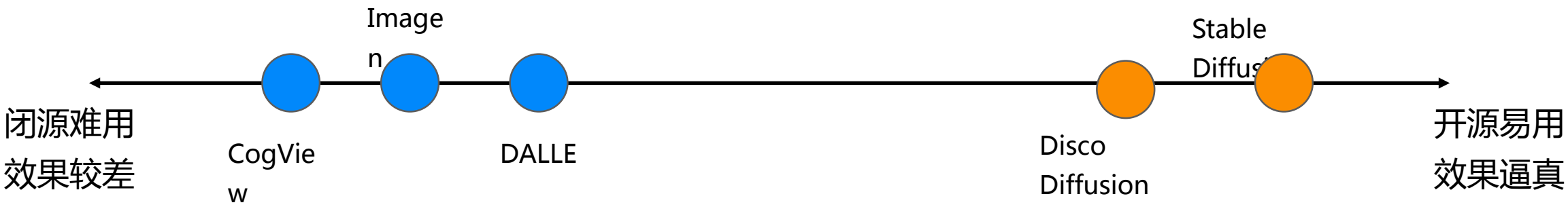
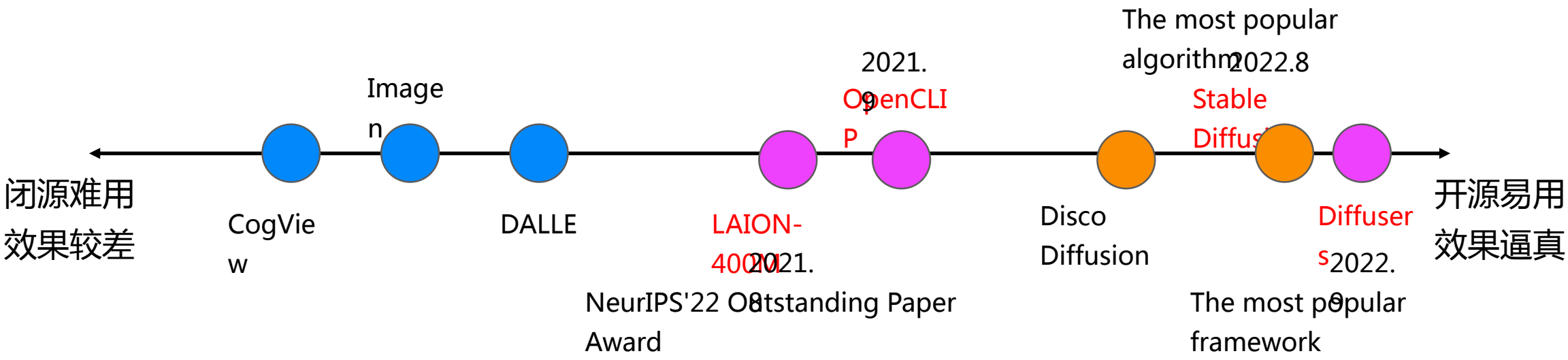


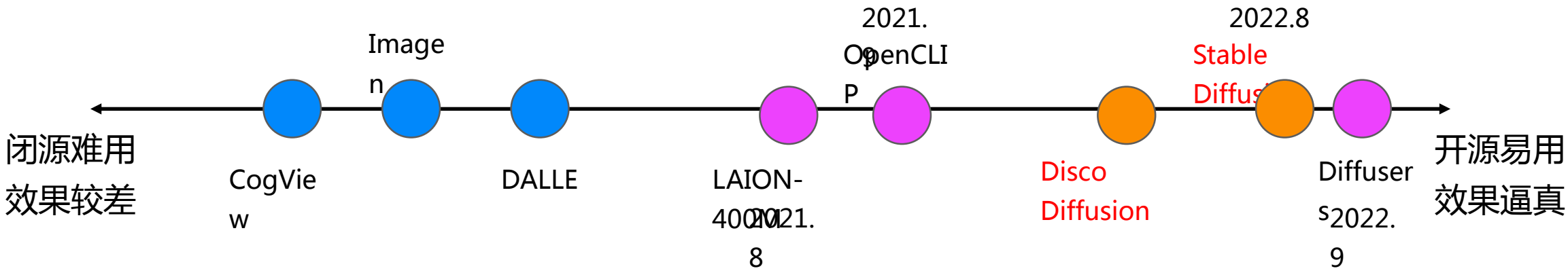
Imagen
Stable Diffusion
闭源难用
效果较差
CogView
DALLE
Disco Diffusion
开源易用
效果逼真

闭源与开源



大数据（LAION）+ **高效率大模型**（SD）+ **完善的框架和活跃社区**（Diffusers/HuggingFace）

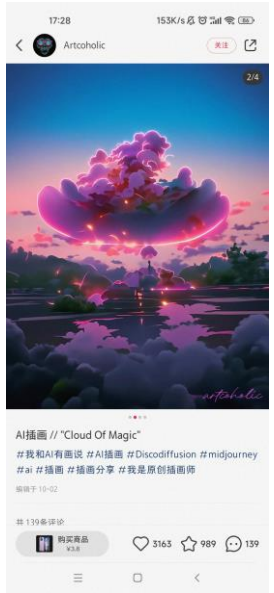
闭源与开源



大数据（LAION）+ **高效率大模型**（SD）+ **完善的框架和活跃社区**（Diffusers/HuggingFace）

Disco Diffusion

2022.3 -- 2022.8



社交平台上出圈的AI绘画热潮

Stable Diffusion

2022.9 --

Disco Diffusion与Stable Diffusion有什么差别

生成效果　　用户体验　　技术方案

Disco Diffusion与Stable Diffusion有什么差别

生成效果　　用户体验　　技术方案

Prompt: “a boy looks outside his bedroom window to see the beautiful cosmos, trending on artstation”

More colorful, aristic

Disco Diffusion

High coherence, quality

Stable Diffusion

Prompt: “Interstellar and inception, 4k resolution incredible digital illustration trending on artstation”

Disco Diffusion

Stable Diffusion

Prompt: “a small space orbited by moons and a larger planet with colorful rings, painted in acrylic on canvas”

Disco Diffusion

Stable Diffusion

易用性/可玩性

Colab --> HuggingFace，参数暴露 --> 参数隐藏

Disco与Stable Diffusion体验上有什么差别

生成效率/速度

5分钟以上 --> 30秒以下

效果/稳定性

有氛围但不够精细 --> 精致、清晰、逼真

文案设计较难 --> 普通文案也可以取得不错效果

开源程度

预训练模型 --> 训练代码

Disco Diffusion与Stable Diffusion有什么差别

生成效果　　用户体验　　**技术方案**

本质上均为Guided Diffusion Model

Classifier guidance

使用用外部分类器（如GAN, CLIP）的输出作为引导条件（梯度）

DALLE, Disco Diffusion

Classifier-free Guidance

把引导条件作为condition也作为输入的一部分（Classifier-Free）

GLIDE, DALLE·2, Stable Diffusion

Disco Diffusion技术方案

- 底层模型：CLIP-guided Diffusion
  - 引入CLIP loss
    - 使用CLIP对文本、生成的图片分别进行编码后，计算编码距离

```
loss = clip_losses.sum() * clip_guidance_scale + tv_losses.sum() * tv_scale + range_losses.sum() * range_scale
```

像素图像空间采样，生成速度5-30分钟

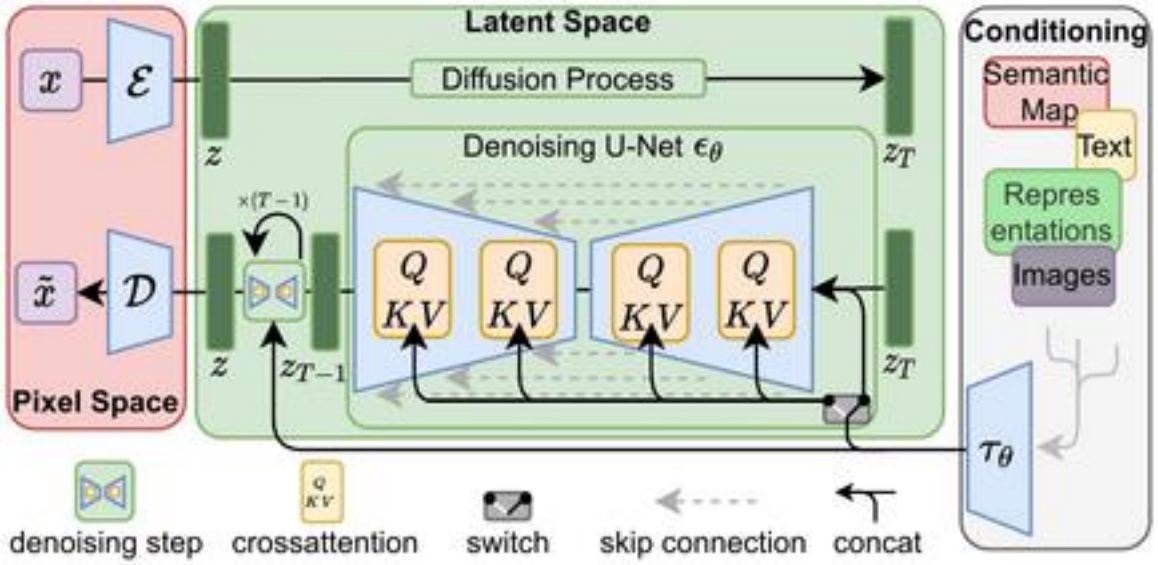
```
def spherical_dist_loss(x, y):
    x = F.normalize(x, dim=-1)
    y = F.normalize(y, dim=-1)
    return (x - y).norm(dim=-1).div(2).arcsin().pow(2).mul(2)


def tv_loss(input):
    """L2 total variation loss, as in Mahendran et al."""
    input = F.pad(input, (0, 1, 0, 1), 'replicate')
    x_diff = input[..., :-1, 1:] - input[..., :-1, :-1]
    y_diff = input[..., 1:, :-1] - input[..., :-1, :-1]
    return (x_diff**2 + y_diff**2).mean([1, 2, 3])


def range_loss(input):
    return (input - input.clamp(-1, 1)).pow(2).mean([1, 2, 3])
```

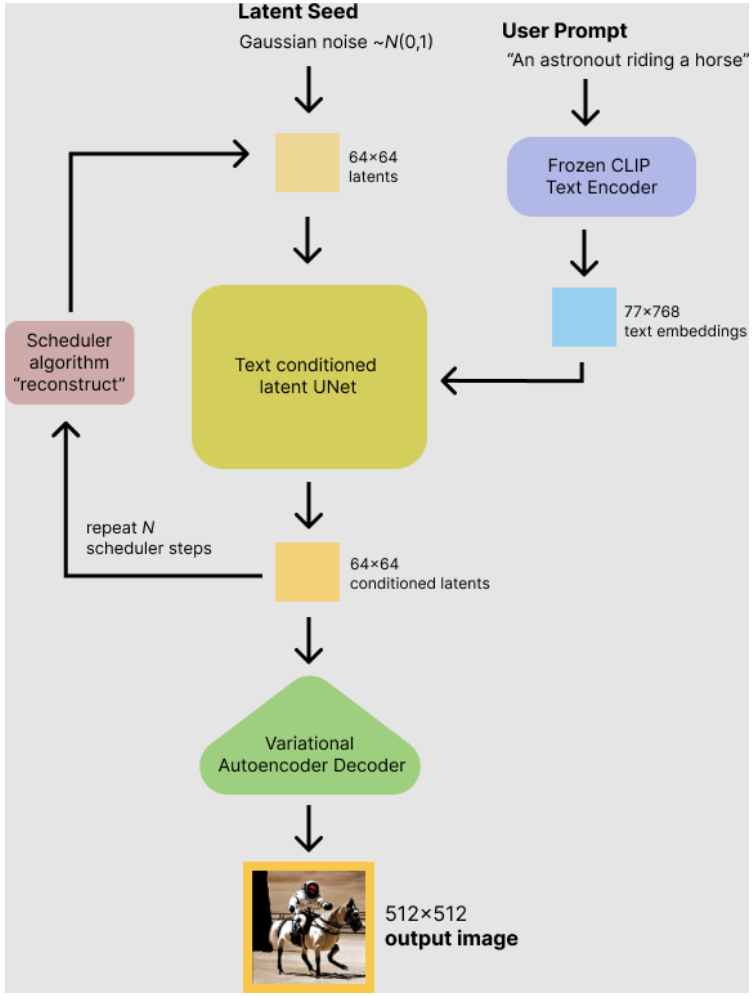
Stable Diffusion技术方案

- 底层模型：Latent Diffusion
  - 首次提出在隐空间（VAE）进行加噪和去噪
  - 扩散空间（UNet）相比原图像空间大幅度减小，效率显著提升
  - CLIP仅用于对文本编码（Text Encoder），投影变换后，通过cross attention来支持各种类型的条件嵌入



近期中文Stable Diffusion技术方案

- 太乙（IDEA-Research）
  - 基于OpenCLIP自行训练的CLIP作为文本编码器
  - 英文Stable Diffusion初始化，仅训练文本编码器，其余冻结
  - 数据采用Wukong中文数据（0.2亿），有使用自训CLIP分数过滤
- PAI-Diffusion（阿里云）
  - 与太乙类似，但仅微调Unet部分，其余冻结
  - 数据采用Wukong中文数据（0.2亿）
  - 进一步下特定数据集微调，包括古诗、食物
- AltDiffusion（智源）
  - 支持多语言，未提供训练细节



近期中文Stable Diffusion技术方案

- 未开源训练代码
- 训练细节、参数均未详细描述
- 效果一般，不及英文Stable Diffusion
- 会针对特定场景/风格进行微调，如PAI-Diffusion





自研中文Stable Diffusion技术方案

- 训练方案（参考现有工作）
  - 使用英文Stable Diffusion进行初始化
  - 训练部件
    - 仅训练text encoder（Taiyi中文版本）
    - 仅训练unet（PAI-Diffusion）
    - 先训练text encoder，再训练unet（Taiyi双语版本）
    - 同时训练text encoder和unet（Stable Diffusion不同版本间迭代）
  - 数据
    - Wukong全量1亿数据，暂未进行清洗
    - 站内亿级规模中文数据

应用场景

- 艺术创作
- 故事配图
- Prompt售卖

Jason Allen的作品《太空歌剧院》，使用Midjourney完成，
在美国科罗拉多州博览会艺术比赛获得一等奖

PS已经集成自动集成“一键补全”

应用场景

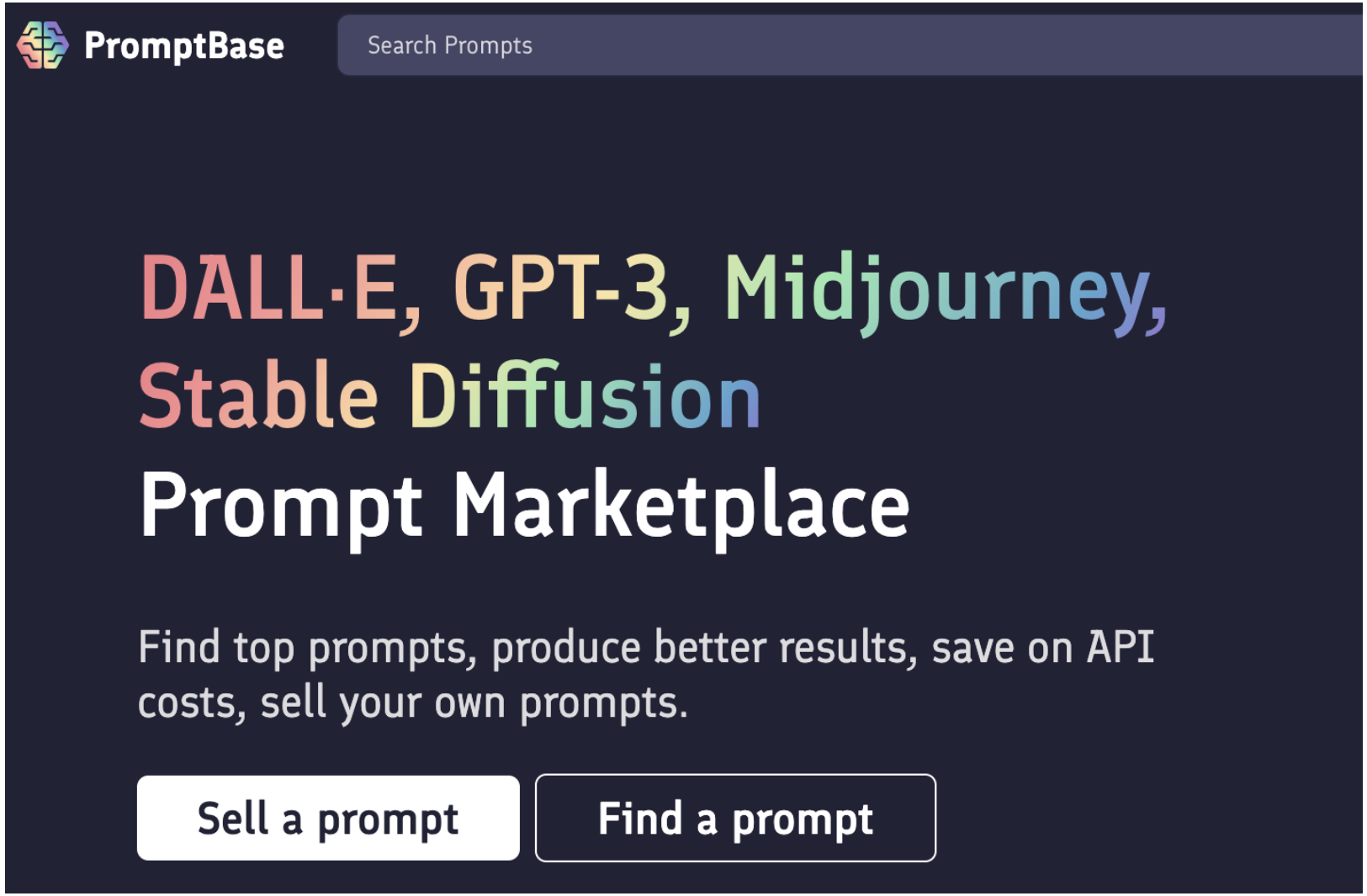
- 艺术创作
- 故事配图
- Prompt售卖

百度AI作画平台文心一格创作的AI绘本《外星超能战队》

小学教材中经常涉及卡通背景、人物的创作

应用场景

- 艺术创作
- 故事配图
- Prompt售卖



正在招聘

Prompt研发实习生

北京/海淀区·300–400元/天·在校/应届

职位详情

职位描述

1. 负责基于大语言模型做Prompt Engineering，创造更高效的Prompt

职位要求：

1.985或top200计算机专业，优秀本科或硕士生，对于大模型有所了解并感兴趣。

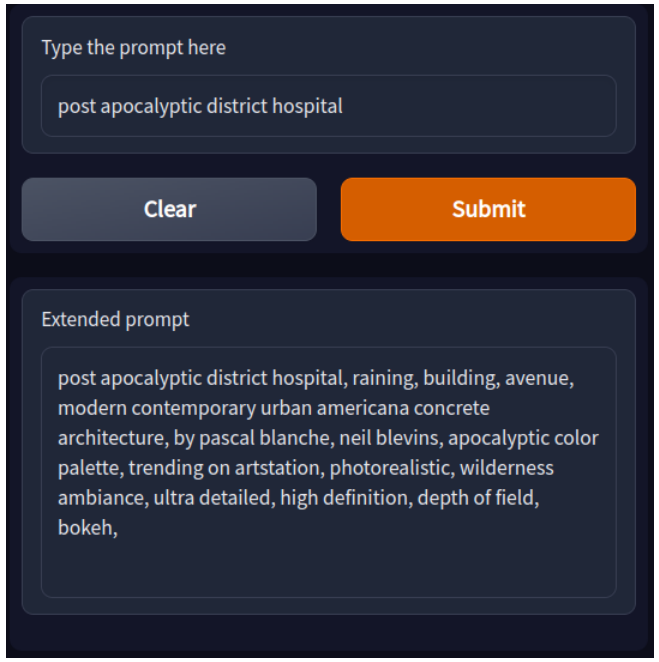
2.逻辑思维强，具有良好的分析和解决问题的能力

3.英文出色，具备较强的读写能力

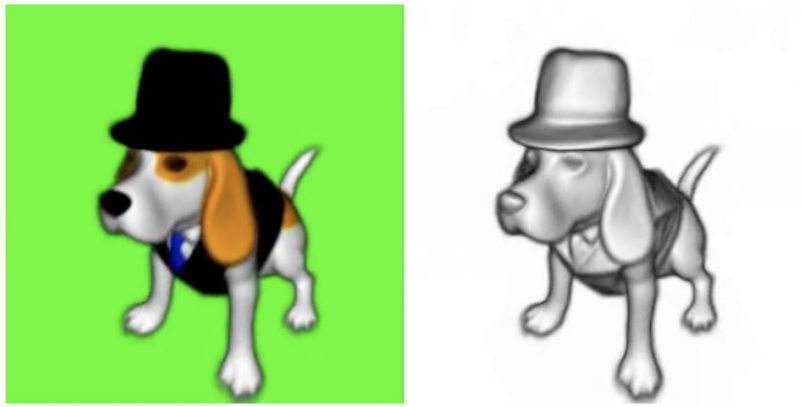
4.有一定的编程基础，具有良好的代码风格和软件工程思维

除了图片外，Prompt Engineering也逐渐成为一种工作模式

当前挑战



- 需要反复调试提示词
- 较难生成连贯、一致的的内容
- AIGC在视频/3D/音乐等领域离图片生成效果差距较远



a beagle in a detective's outfit

Thanks