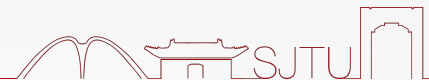




生成式人工智能应用实践

- 基于机器学习

2024 Fall





概述

- 课程背景：AI生成内容（AIGC）的发展如火如荼，逐渐在各方面对人们的生活和工作模式产生越来越大的影响。
- 课程目标：通过本课程中设计的实践任务，探讨数据收集，数据清洗，模型训练，模型测试等全流程开发，对AIGC基于机器学习的方法论获得宏观的认知和理解。
- 课程任务：10~12人一个小组，每个小组从四个项目中选择一个，强制选题均匀分布（先到先得）
 - AI辅助矢量LOGO设计
 - AI辅助艺术画生成与编辑
 - 媒体/时尚工业中的人体编辑技术
 - 数字文旅中的三维物体生成编辑技术



成绩评定

- 每个阶段提交结果，独立评分，互不影响
 - 数据收集：十月中旬提交收集到的数据（评分占比30%）
 - 数据清洗：十一月初提交清洗后的数据（评分占比10%）
 - 模型训练与测试：十二月底提交代码与模型（评分占比30%）
 - 项目报告与项目答辩（线上组织）：期末提交（评分占比30%）



项目1：虚拟试穿体验

我们以动虚拟试穿为例，研究视觉大模型在时尚领域的应用

输入模特图与平铺图，商品的图像与模特的身体图像融合，来模拟穿衣后的效果。

静态试穿



现有多多种开源的预训练的模型。

1. 但是由于公开数据集的数据**bias** (T恤居多)，在很多品类上表现不佳。
2. 现有模型由于数据集限制，通常只支持输入平铺服饰去给模特穿衣，不支持输入人身上的衣服换到模特上。即garment-person与person-person

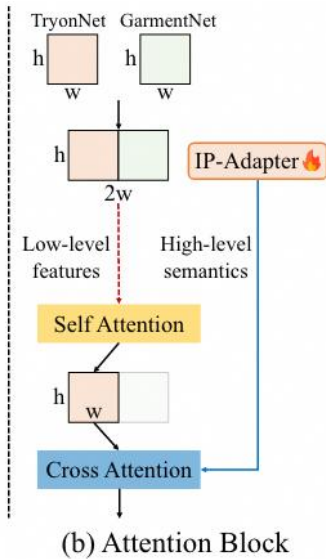
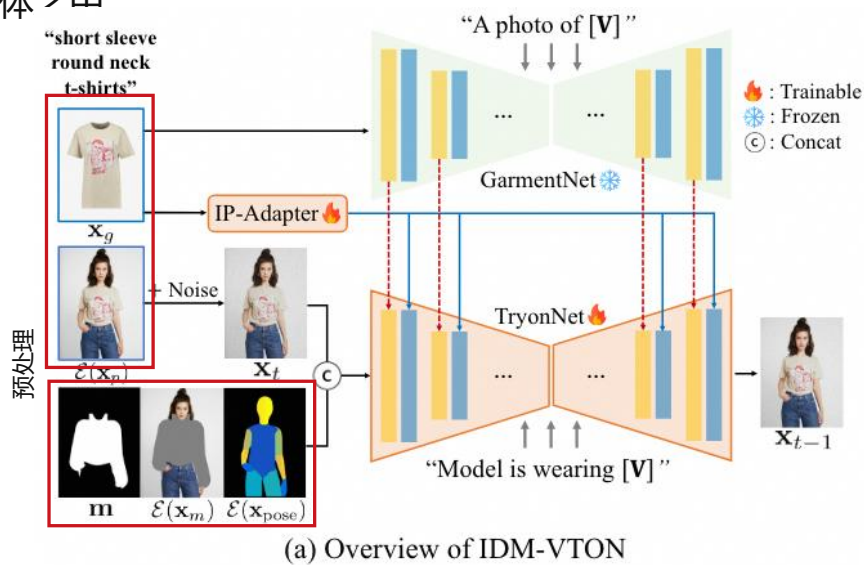
在本项目中，我们需要收集这两方面的数据，并针对性的提升开源模型在这两方面的结果。





一阶段：静态试穿模型管线

以Stable Diffusion模型为骨架，采用双Unet架构，利用Reference Unet进行衣服特征的编码并注入人体之中

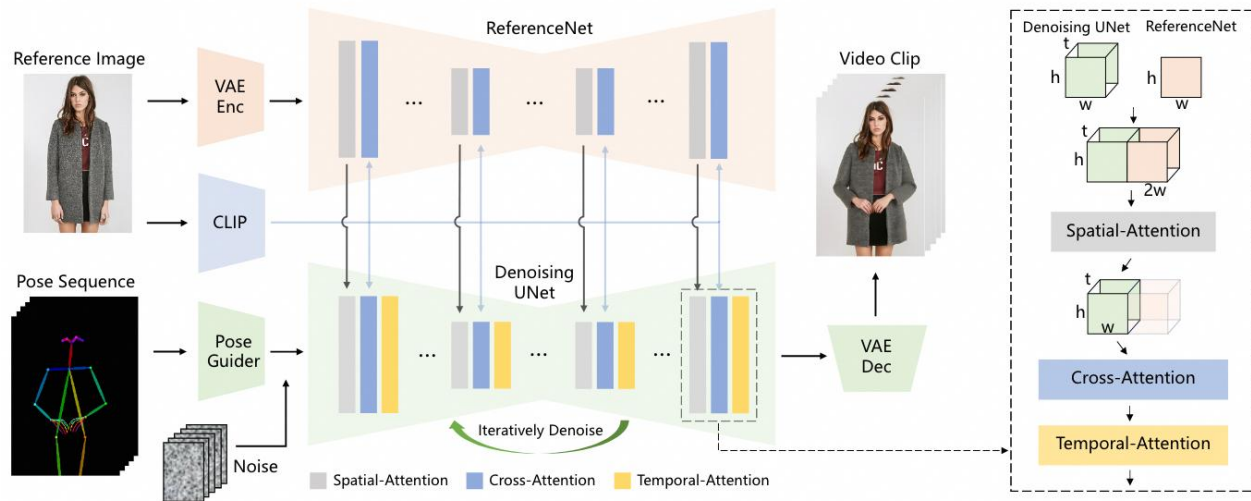




二阶段：动态驱动

动态驱动需要两种输入：参考图（即静态试穿结果图）以及人体骨架序列。
有预训练模型，实现一系列动作驱动。是否适合同学们进行 *fine tune* 正在评估。

同样采用双Unet架构，利用Reference Unet进行参考图的编码并注入人体之中。骨架序列通过轻量模型进行引导。



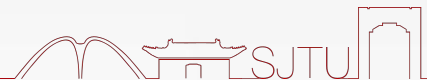


作业内容:

虚拟试衣的关键在于高质量数据的覆盖度，同学们使用高质量数据*fine tune*以后可以轻松达到超过*baseline*的效果。

具体内容:

- 1.【收集多品类的*garment-person*数据】每个小组分到一些品类，使用爬虫脚本，按照品类收集一批高质量的衣服对（平铺衣服以及模特上身图）。
- 2.【通过视频收集*person-person*数据】每个小组分到一些视频，利用视频分割工具，切出几秒至十几秒的*clip*，*clip*中只有一个占比适中的主体人物。并通过姿态检测模型，标注出一帧最标准的模特穿着衣服代表性的图片。
- 3.【微调试衣模型】利用以上的*garment-person*数据和*person-person*数据，对现有*sota*模型进行训练（训练的模型的实验指标不作要求，因为同学们训练资源不同。仅要求经过完整的训练过程）
- 4.【试衣结果动态化】【可选项，不计分，想体验的提供指导】完成一项最时尚的动态试衣！利用人体驱动模型对试衣结果进行驱动，形成时尚视频。





④ 动态虚拟试穿

- 算法推荐: OOTDiffusion, Animate Anyone: Consistent and Controllable Image-to-Video Synthesis for Character Animation
- 代码资源: <https://github.com/levihsu/OOTDiffusion>, <https://humanaigc.github.io/animate-anyone/>



项目2：感兴趣物品单图三维重建



➤ 大作业目标

- 以单图三维重建为例，探究大模型在三维重建中的应用
- 了解二维大模型如何应用于三维视觉领域
- 收集自己的数据集，并基于此尝试大模型的微调



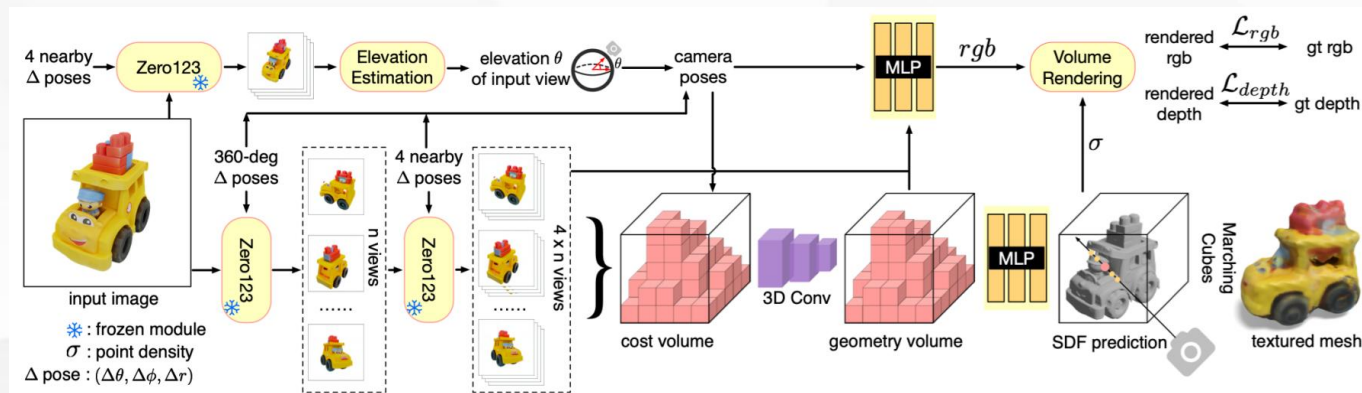


➤ 单图三维重建

- 根据用户提供的单张图片重建出完整的三维模型
- 在大模型时代前，这是一个难以完成的任务，因为单图提供的信息量太少
- 借助大模型学到的丰富先验知识，补全单图的信息缺失，实现单图重建

➤ One-2-3-45 (NIPS23)

- 借助大模型 Zero-1-to-3 (ICCV23) 将单图拓展为多视角图片
- 根据拓展出的多视角图片重建出目标图片中的物体





- 代码下载&环境配置
 - <https://github.com/One-2-3-45/One-2-3-45>
 - 阅读代码，读懂代码逻辑
- 收集自己的三维数据集
 - 利用爬虫（不强制）爬取数据，并对爬下来的数据进行清洗，删除不符合主题或者低质量的数据
 - 根据代码需要的数据格式生成并整理数据
- 基于自己的数据集微调模型
- 成绩评定
 - 数据收集与清洗（40%）
 - 模型微调与测试（30%）
 - 提交期末报告（30%）



- 数据要求：
 - 数据类别定在：动物、植物、文玩、游戏形象、动漫形象共五类中
 - 每个组内自行分工，每个类别的数据量不少于500个
 - 将数据以以下格式提交：
 - 组号_组长学号.zip
 - |--- 下载数据
 - |--- 清洗后的数据
 - |--- 渲染结果
- 数据源网站（推荐）
 - <https://sketchfab.com/>
 - <https://3d.si.edu/>
 - <https://poly.cam/explore>
- 数据收集工具
 - 提供爬虫的例子以及代码开发指导

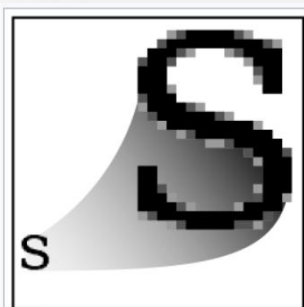


项目3：矢量设计图形LOGO生成



➤ 项目目标

- 了解矢量图形的意义以及大语言模型、多模态大模型在该领域的应用；
- 熟悉大语言模型训练、微调及部署全过程；
- 掌握通用计算机视觉项目开发基础技能；



Raster
GIF, JPEG, PNG



Vector
SVG

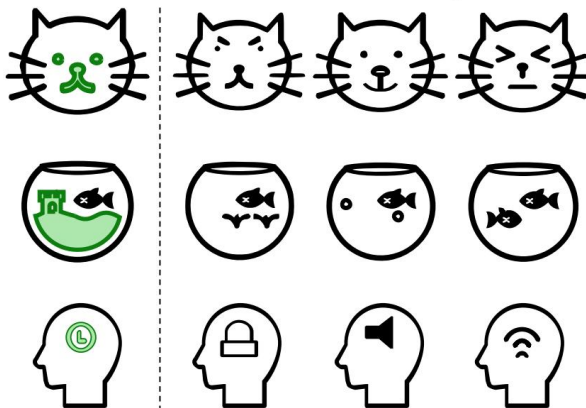
```
<svg width="402pt" height="147pt" viewBox="0 0 401.58 147.08" xmlns="http://www.w3.org/2000/svg">
  <svg width="402pt" height="147pt" viewBox="0 0 401.58 147.08" xmlns="http://www.w3.org/2000/svg">
    <svg width="402pt" height="147pt" viewBox="0 0 401.58 147.08" xmlns="http://www.w3.org/2000/svg">
      <xlink:href="http://www.w3.org/1999/xlink">
        <g id="graph" class="graph" transform="scale(1 1) rotate(0) translate(4 143)">
          <poly fill="white" stroke="transparent" points="-4.4 -4, -143 397.5, -143 397.5, 4 -4" />
          <g id="model" class="node">
            <poly fill="none" stroke="black" points="137.5, -120.5 137.5, -138.5 237.5, -138.5 237.5, -120.5 137.5, -120.5" />
            <text text-anchor="middle" x="187.5" y="-126.5" font-family="Helvetica, sans-serif" font-size="18.00">src/loss_sort_hv/text</text>
          </g>
        </g>
      </xlink:href>
    </svg>
  </svg>
</svg>
```

- Small File Size
- Scalability
- Easy to edit

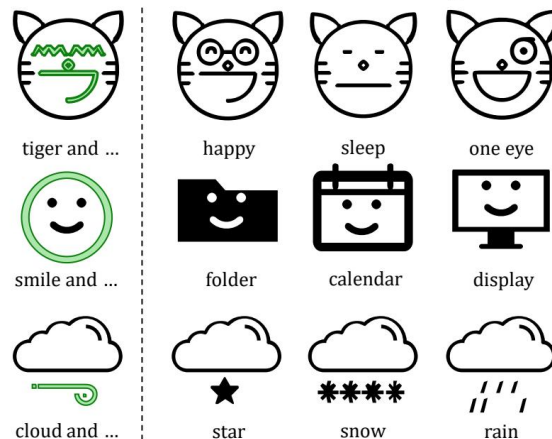
"cake, dessert,
sweet, food"



Random Editing



Text-Guided Editing

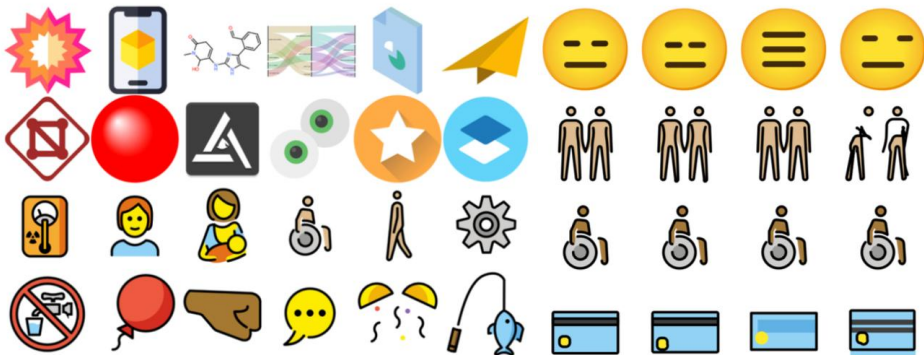




项目内容--阶段一



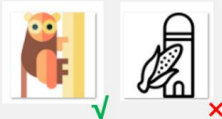
➤ 按照示例程序收集项目相关数据;



1. 数据样本与类别名直接关联，且表达语义信息清晰明了。

一个类别的涵盖范围是狭义的，数据样本反映出的信息必须直指类别名称。以“Animal”类别为例，下图中只有动物本身被筛选，而与动物间接有关的数据如“农场”要剔除。以“Building”类别为例，下图中只有房屋建筑被筛选，而房屋中的“灯饰”要剔除。以“Clothing”类别为例，宇航员头盔从感官上不是很直接（不能一眼看出来），且语义信息有歧义（即是否是衣服），所以要剔除。

Animal



Building



Clothing



2. 数据样本要通过图形信息表现一个具体物体或活动事件。

以“Animal”为例，下图左侧，通过十二地支来表征“猴子”，这是用文字（或字母）来传递语义信息，应该剔除。下图中右侧在动物的身上加了斜杠，代表这是一个用于指示禁止性的图标，并不是直接表示动物这一具体物体，应该剔除。

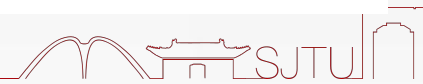
Animal



3. 数据样本表征的是单一的对象，可以有其他具有空间关联性的装饰物（如花草在真实空间中会在一起，具有关联性），但不可喧宾夺主。

以“Household item”为例，下图左侧一个svg文件中表示了4个不同种类，没有空间关联性的对象，应该剔除。

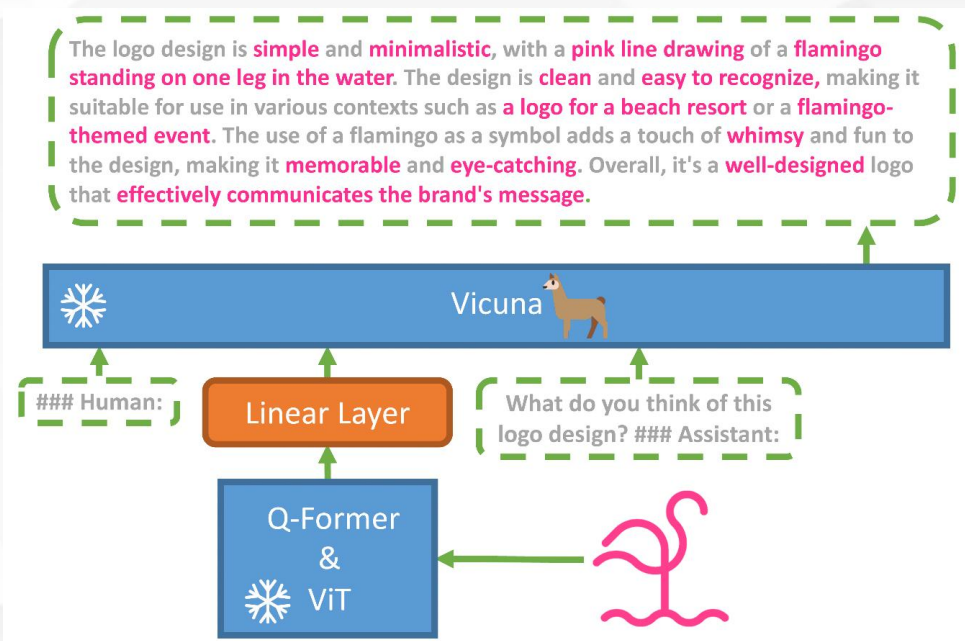
Household item





- 利用大模型对SVG进行标注;

将SVG光栅化成图像，并使用预训练模型为其生成文本描述



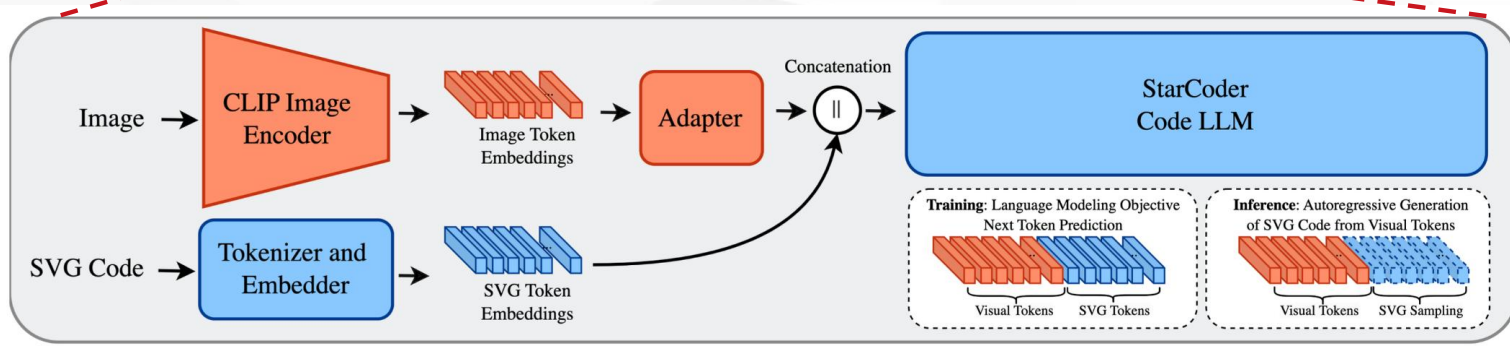
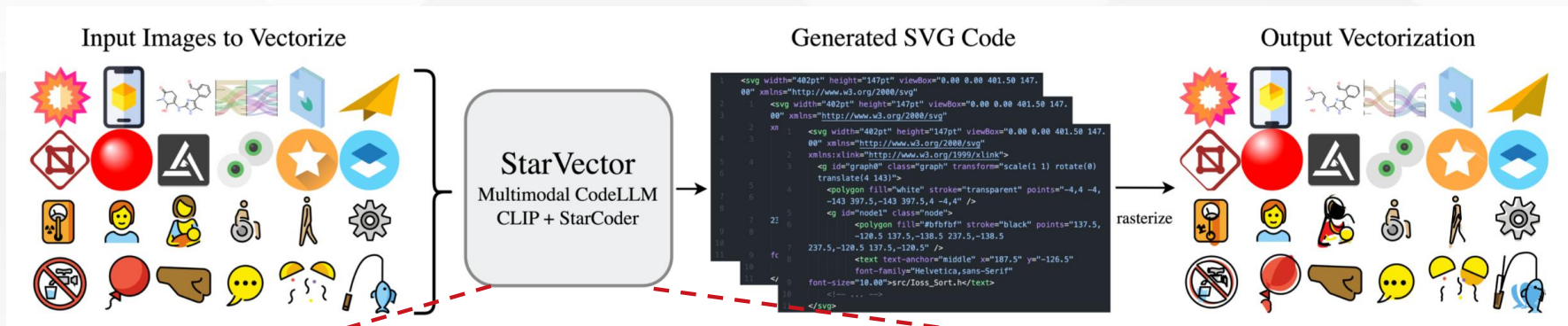
需设计量化指标对文本标注进行refine，修改或删除低质量标注



项目内容--阶段三



- 使用推荐的生成模型或自定义模型在收集的数据上部署训练;





④ 矢量图形生成

• 算法推荐:

- StarVector: Generating Scalable Vector Graphics Code from Images
- Vector Graphics Generation via Mutually Impulsed Dual-domain Diffusion
- Llama3.1 <https://llama.meta.com/>

• 工具推荐:

- ChatGPT 4o (可自行尝试其zero-shot SVG生成能力)



项目4：艺术图像编辑与生成



➤ 大作业目标

- 探究大模型对艺术风格图像的编辑与生成能力
- 收集艺术图像数据集，并基于此尝试大模型微调，提升大模型垂域能力



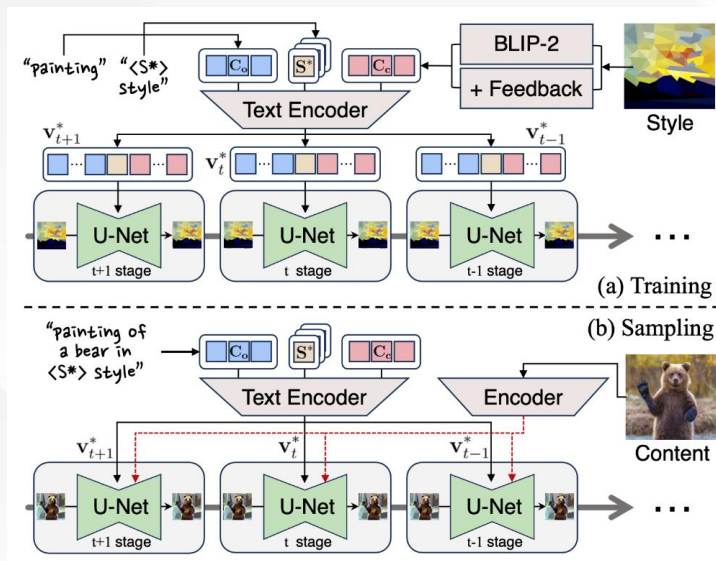


方法简介

➤ 基于Textual Inversion的定制化Text-to-Image技术;

Textual Inversion:

通过学习特定文本提示的嵌入向量, 使模型生成与这些嵌入向量相关的特定概念或风格的图像。





- 收集自己的艺术图像数据集
 - 利用爬虫（不强制）爬取数据，并对爬下来的数据进行清洗，删除低质量（如清晰度较低）的数据
 - 对数据进行归类整理（如按照艺术风格分类，每一类对应一种风格名称）
- 基于自己的数据集微调模型
 - 测试Stable Diffusion zero-shot地生成艺术图像的能力；
 - 基于收集的数据集，使用Textual Inversion技术来增强模型的艺术生成能力；
- 成绩评定
 - 数据收集与清洗（40%）
 - 模型微调与测试（30%）
 - 提交期末报告（30%）



④ 艺术图像生成

• 算法推荐:

- DreamStyler: Paint by Style Inversion with Text-to-Image Diffusion Models
- StableDiffusion: High-Resolution Image Synthesis with Latent Diffusion Models

• 开源代码推荐:

- <https://nmhkahn.github.io/dreamstyler/>
- <https://github.com/CompVis/latent-diffusion>



分组登记





加入组织



群聊：动态虚拟试穿作业群



该二维码7天内(9月25日前)有效，重新进入将更新



群聊：三维物体生成



该二维码7天内(9月26日前)有效，重新进入将更新



群聊：矢量图形生成



该二维码7天内(9月26日前)有效，重新进入将更新



群聊：艺术图像生成



该二维码7天内(9月26日前)有效，重新进入将更新