

基于FG-ECVG的细粒度情感可控图像生成算法

Enhanced FG-ECVG for Fine-Grained Emotional Image Generation

作者：刘申扬洋

指导老师：卫青蓝、吴殿义

研究总结

研究背景：

图像生成算法在艺术创作、广告设计、影视制作等领域展现出巨大的潜力。然而，传统的图像生成模型往往侧重于内容的逼真性和多样性，对生成图像的情感表达缺乏精细控制。这种局限性导致生成的图像难以精准传递特定的情感，无法满足用户对情感化内容的需求。

现有方法及问题：

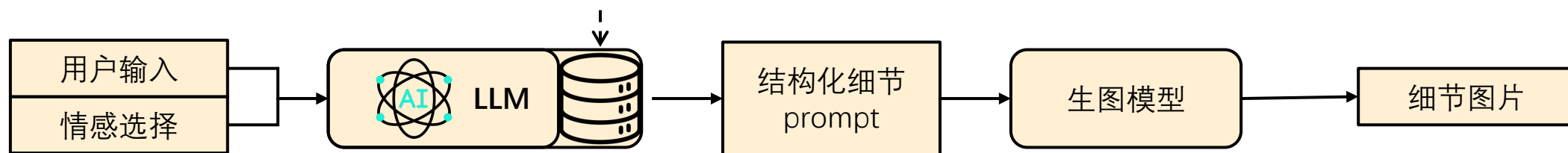
Qinglan Wei等人针对大模型生成的视觉内容情感属性模糊、交互性弱的问题, 提出一种基于文本指令优化的视频生成算法FG-ECVG, 可以实现文本指令到视频内容的高情感可控、强交互性自动化生成。但这种方法存在两点问题：1.细节控制与氛围控制的分离：FG-ECVG将视觉细节和整体氛围分开处理，缺乏统一的框架来实现两者的协同优化；2.可控性不足：生成图片的情感特征主要依赖于生图AI的性能，不能实现二次校正。

我的解决方法（创新点）：

在生图任务的基础上，引入图的优化任务，设置一个氛围校准模块，输入为细节图片和氛围引导词，输出为基于氛围引导词进行氛围校准后的细节图片。该模块能增强生成图片整体氛围的可控性，使生成结果更加符合用户的情感表达需求。

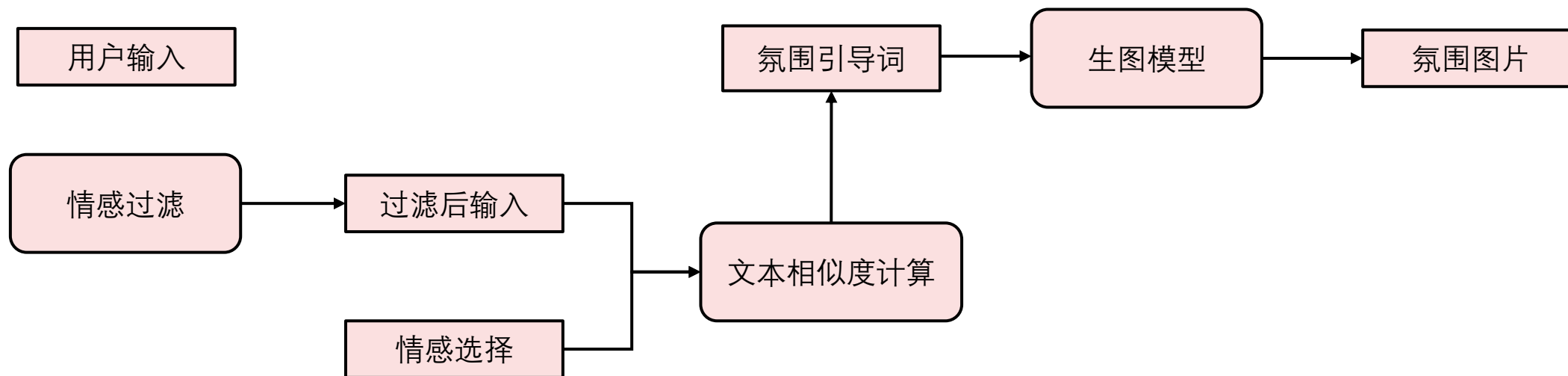
该模块需要满足两个目标：1）校准模块应当准确理解输入文本中的情感，并在输入图像中反应这些情感；2）校准模块应当保证内容一致性——输出图片应当与输入图片具备一致的整体结构和视觉细节。因此，设计了一个基于多模态Transformer的整体氛围校准模型并设计三类损失函数进行模型训练。

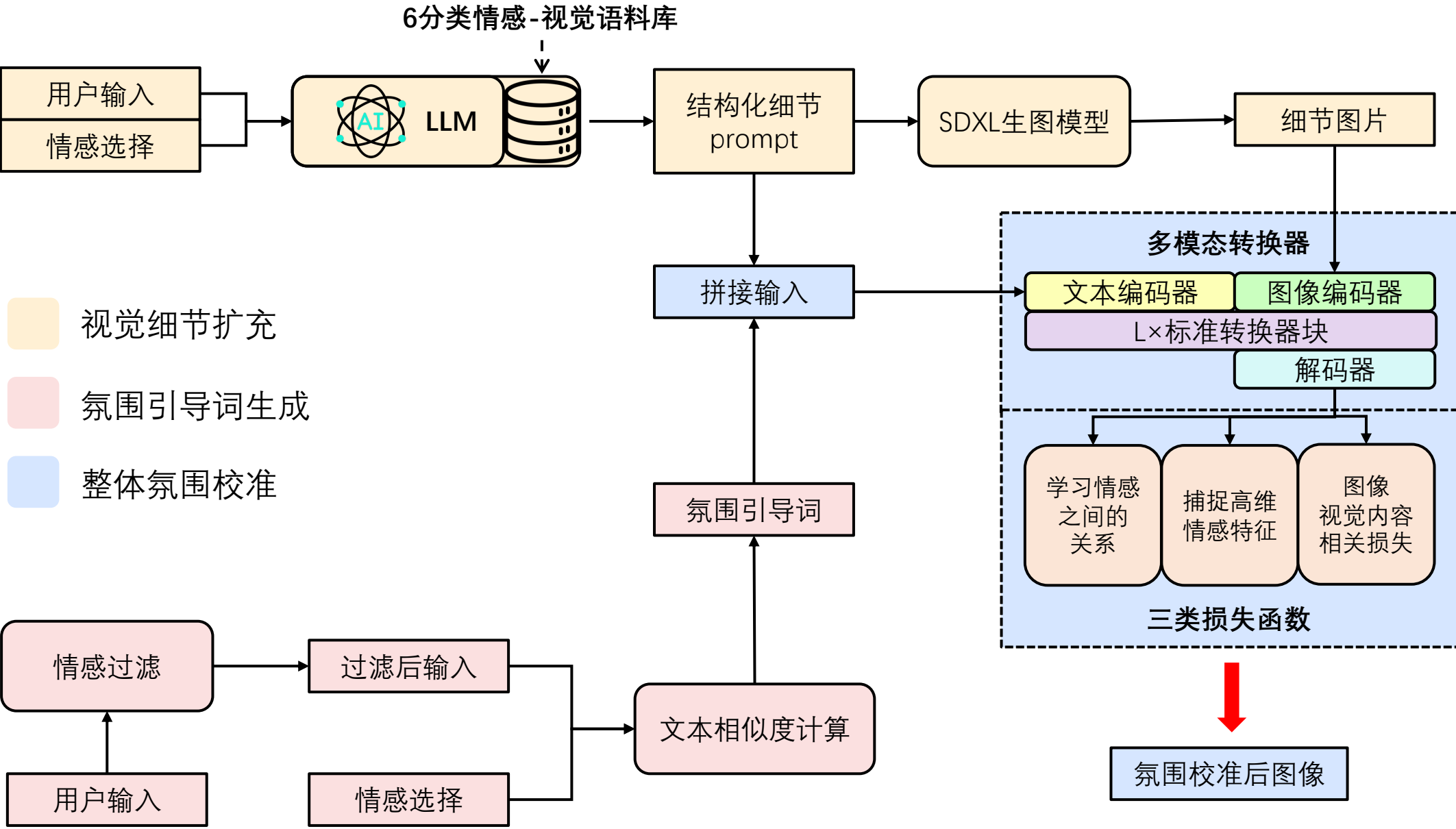
6分类情感-视觉语料库



细节控制

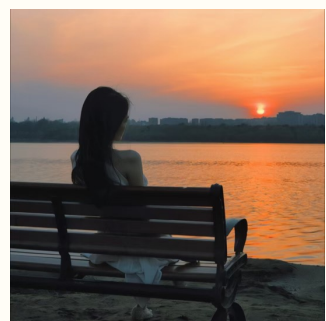
氛围控制



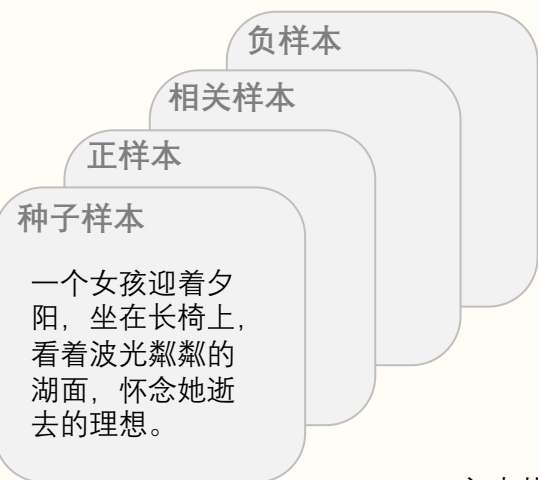


核心技术

(a) 输入



内容图像



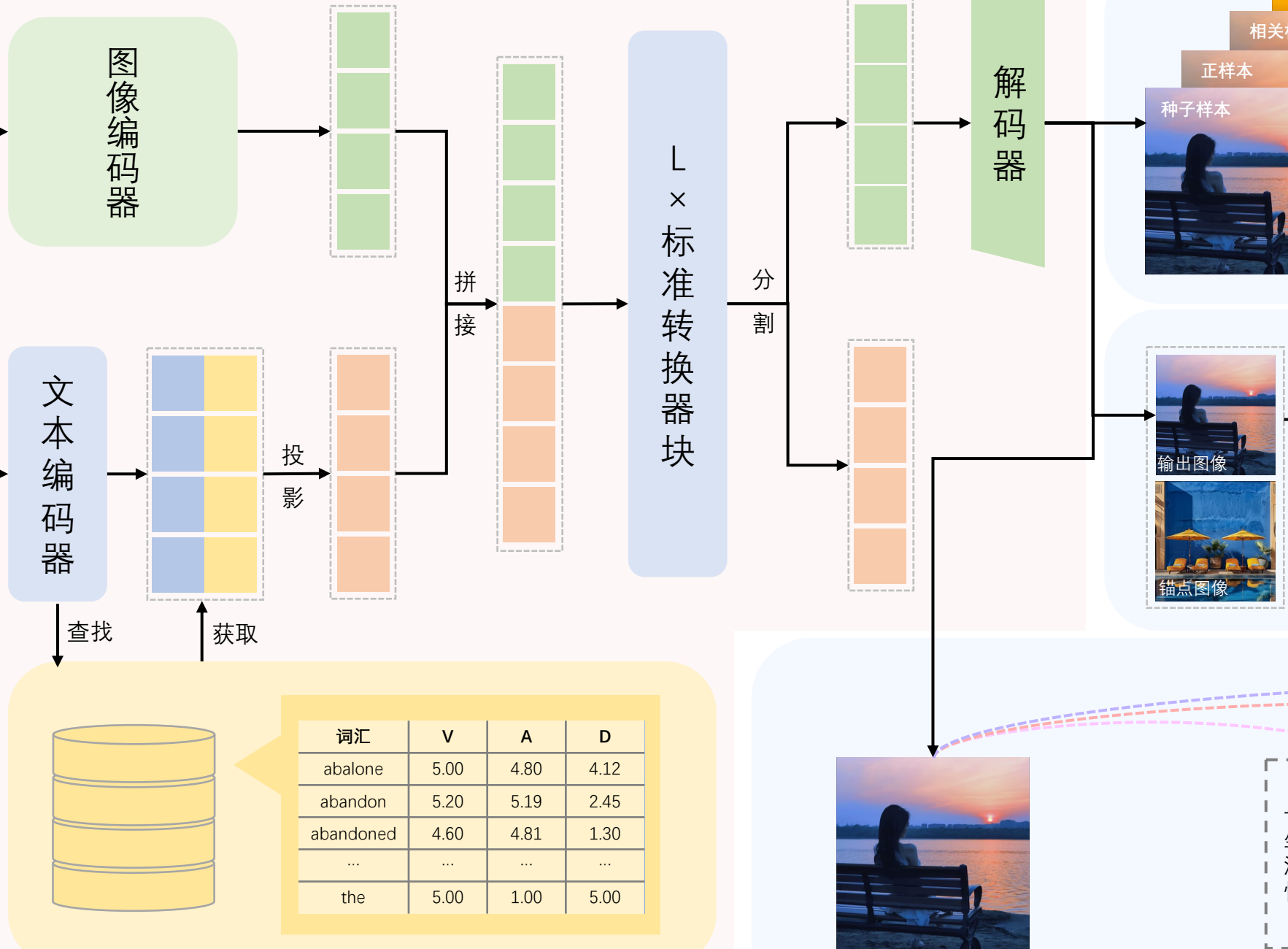
文本描述

样本

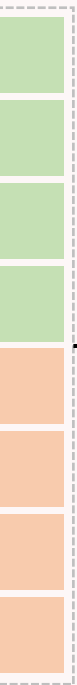


在Mikel轮盘中的相对位置

(b) 多模态转换器



生成图像

 $L \times$ 标准转换器块

分割

解码器

种子样本

正样本

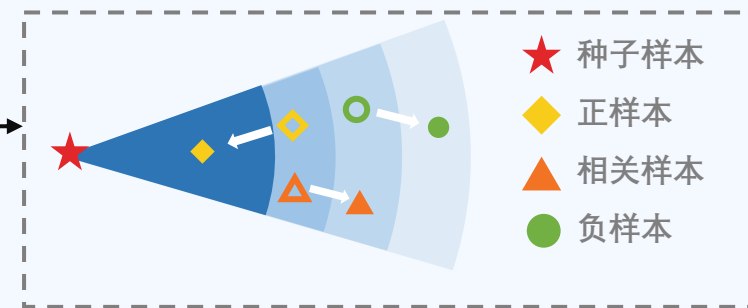
相关样本

负样本

输出图像

(c) 学习情感之间的关系

情感提取器



(d) 捕捉高维情感特征

情感提取器

分布估计器

生成值

真实值

锚点向量
输出向量

A	D
4.80	4.12
5.19	2.45
4.81	1.30
...	...
1.00	5.00

(e) 图像视觉内容相关损失

保证内容一致性

文本-图像对齐

锚定颜色与纹理

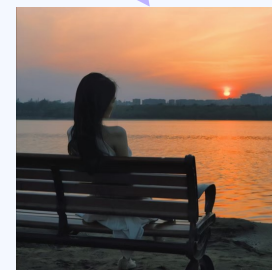
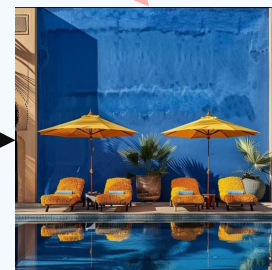
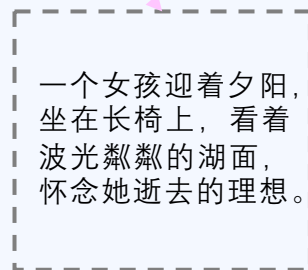
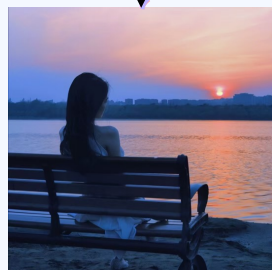
参考

生成图像

文本描述

锚点图像

内容图像



实验与评价

表 2.1 在 2 种场景上的与基础算法的定性对比结果

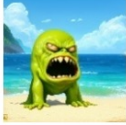
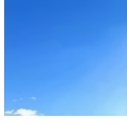
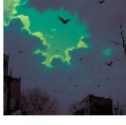
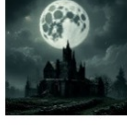
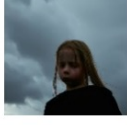
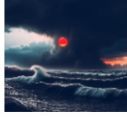
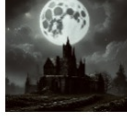
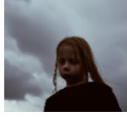
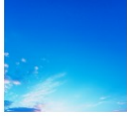
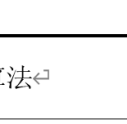
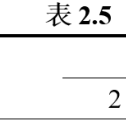
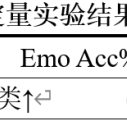
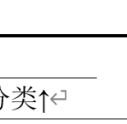
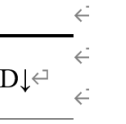
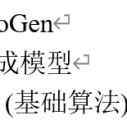
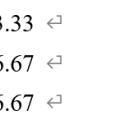
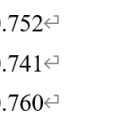
场景	实验设置	情感类别					
		anger	disgust	fear	sadness	joy	surprise
seaside	仅生成模型						
	FG-ECVG						
	基础算法						
sky	仅生成模型						
	FG-ECVG						
	基础算法						
	FG-ECVG						
	+氛围校准						

表 2.5 定量实验结果

算法	Emo Acc%		LPIPS↑	FID↓
	2 分类↑	6 分类↑		
AIF	53.33	16.67	0.746	388.54
EmoGen	66.67	23.33	0.752	429.25
仅生成模型	73.33	26.67	0.741	339.91
FG-ECVG (基础算法)	100.00	46.67	0.760	311.41
FG-ECVG +氛围校准	100.00	50.00	0.779	301.15

表 2.2 与 EmotiCrafter^[35]的定性对比结果

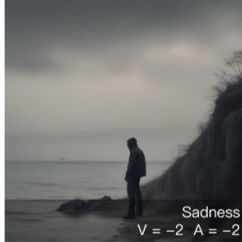
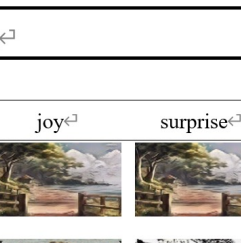
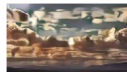
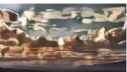
	实验设置	(a) A person is standing over the ocean (sadness)	(b) A cartoon shows children playing football (amusement)	(c) A man is running fast (anger)
EmotiCrafter ^[35]				
		Sadness V = -2 A = -2	Amusement V = 2.5 A = 2	Anger V = -2 A = 2.5
FG-ECVG				
	+氛围校准			

表 2.4 在 2 种场景上 3 种算法的定性对比结果

场景	实验设置	情感类别					
		anger	disgust	fear	sadness	joy	surprise
seaside	AIF ^[53]						
	EmoGen ^[14]						
	FG-ECVG						
sky	AIF ^[53]						
	EmoGen ^[14]						
	FG-ECVG						
	+氛围校准						

广告应用

典型品类的广告的情感可控生成



(a) 越野车-surprise



(b) SUV-joy

图 3.1 汽车广告中的情感可控生成



(a) 能量饮料-surprise



(b) 巧克力-joy

图 3.3 食品广告中的情感可控生成

特殊节点的广告的情感可控生成



(a) 圣诞节-joy



(b) 万圣节-fear

图 3.4 万圣节 VS 圣诞节的情感可控生成



(a) 开学季-joy



(b) 毕业季-sadness

图 3.6 开学季 VS 毕业季的情感可控生成

前端



情感控制图片生成系统

细节描述

细节生成

氛围描述

氛围增强

请输入情感

请输入简单描述

请输入第一步生成的细节描述

请输入细节图片的名字(物体_情感)

提交

