



基于 Stable Diffusion 的蜜雪冰城 “雪王” IP 表情包生成研究

汇报人：邱晨宸 指导老师：龙思薇 靳聪

广告学（计算广告双学士学位复合型人才培养项目）



目录 CONTENTS

- 1 研究背景与意义
- 2 研究思路与方法
- 3 研究内容与实施过程
- 4 研究结论与展望



1

研究背景与意义

Research Background and Significance

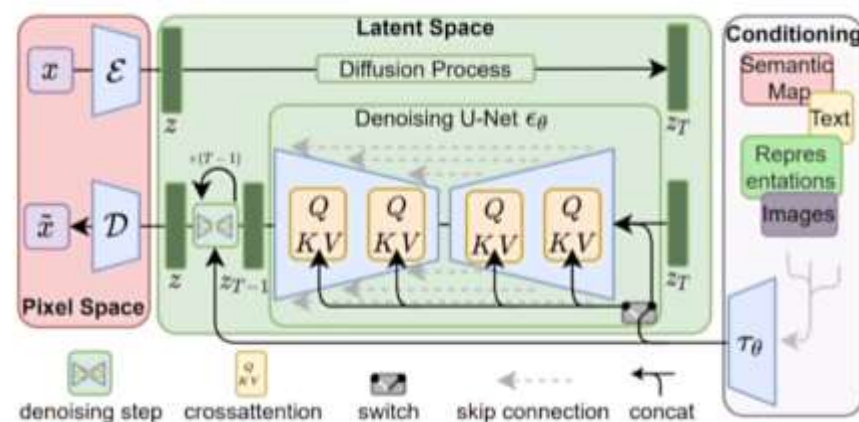
- :: 研究背景
- :: 研究意义

:: 研究背景与意义



行业背景

- 近年来，新茶饮品牌IP化趋势日益显著，消费者对品牌拟人化形象接受度上升。
- 表情包作为情绪沟通工具，在社交媒体中的传播力强，是品牌用户黏性构建的重要媒介。
- 传统IP表情包创作依赖专业人员设计，创意成本高，难以满足品牌大规模传播需求。



技术背景

- 在营销创意领域，以 “Stable Diffusion” 为代表的文生图大模型展现出巨大潜力。能依据文本描述快速生成大量创意图像。
- 借助 LoRA 、 Controlnet等技术，还可针对特定风格或主题微调生成结果，增强对画面风格和细节的控制，使生成图像更贴合营销需求。



2

研究思路与方法

Research Ideas and Methods

- :: 研究思路
- :: 研究方法

基于 Stable Diffusion 的蜜雪冰城 “雪王” IP 表情包生成研究

“雪王” IP营销研究

“雪王” IP发展状况分析

- 蜜雪冰城品牌基本情况
- “雪王” IP的出现与发展

“雪王” IP营销策略分析

- “雪王” IP的战略定位
- “雪王” IP的内容生产
- “雪王” IP的生态运营

“雪王” IP LoRa 模型训练

“雪王” IP表情包设计问卷调查

- 问卷内容设计
- 问卷数据分析

“雪王” IP LoRa 模型训练

- 数据集构建与处理
- 训练超参数组合调试

LoRa 模型效果评估实验

- 评价维度设计
- 实验设计
- 结果分析

“雪王” IP 表情包生成

LoRa+Controlnet 生成表情包

- 基于 LoRa 模型的提示词工程与图像生成
- 引入 Controlnet 优化生成图像

表情包图像质量评估实验

- 实验设计
- 结果分析

研究方法

01

案例分析法

研究“雪王”IP发展及营销案例，尤其是表情包营销案例，总结“雪王”IP特征，为后续实证提供理论支持

02

实证研究法

展开“雪王”IP LoRa模型训练与表情包生成，并在2个环节各自进行效果评估实验

03

问卷调查法

通过问卷调查对“雪王”IP LoRa模型效果与表情包效果进行评估

技术工具

01

数据标注工具

1. WD 1.4 Tagger 自动标注系统
2. Dataset Tag Editor 标签编辑系统

02

模型训练工具

1. Stable Diffusion 预训练模型
2. kohya-ss LoRA 训练脚本



3

研究内容与实施过程

Research content and implementation process

- ❖ “雪王” IP 营销研究
- ❖ “雪王” IP LoRa模型训练与评估
- ❖ “雪王” IP 表情包生成与评估

“

3.1 “雪王” IP 营销研究

”

:: “雪王” IP发展情况



2018年推出“雪王” IP

- 2018年，蜜雪冰城与华与华合作推出“雪王” IP，以头戴皇冠、手握冰淇淋权杖的雪人形象为核心，赋予其独特的性格和世界观，成为品牌的代言人。



多渠道打造“雪王” IP 视觉体系

- 线下门店：**基于免物流费、免空间设计费、免宣传物料费的加盟政策与完善的饮品材料供应链体系，“雪王” IP形象快速覆盖线下45000+家门店。
- 线上渠道：**APP、官网、社交账号等多个场景广泛运用“雪王” IP形象



推出多类“雪王” IP 衍生内容

- 围绕“雪王”推出了主题曲、绘本、动画、短剧、表情包及周边文创等衍生内容。
- 截至2024年12月27日，“雪王”相关话题的总曝光量达到191亿次，“蜜雪冰城主题曲”的曝光量为95亿次，动画《雪王驾到》的播放量已突破2亿次。

:: “雪王” IP营销策略分析



战略定位

- IP形象与品牌强绑定，高频展示核心视觉元素以构建品牌心智
- **核心视觉特征**：雪人造型、红色披风、白色皇冠、冰淇淋权杖
- **核心人设特征**：“热情” “搞怪” “社牛”



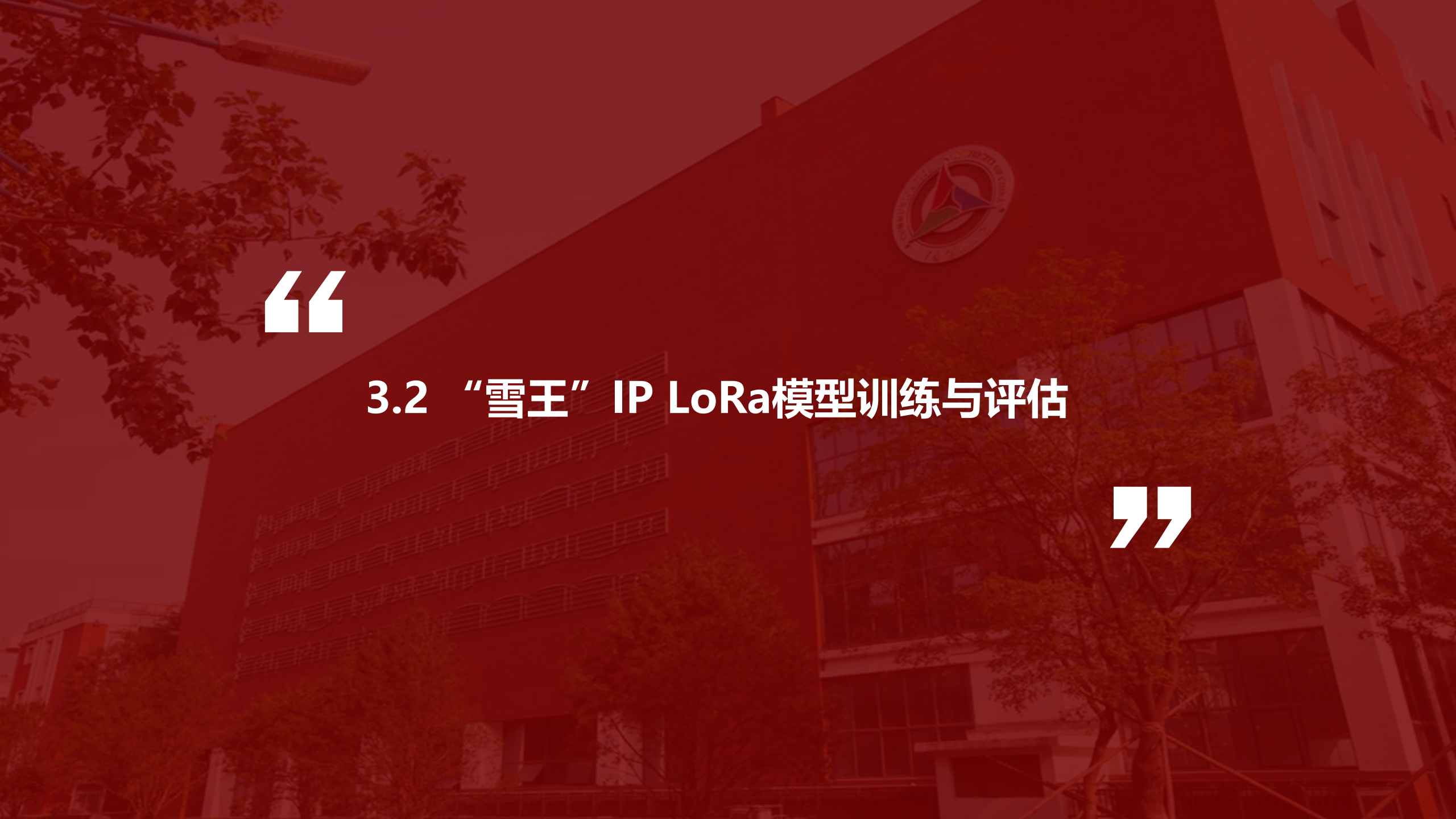
内容生产

- **内容生产矩阵**:
 - 基底层：表情包、短视频等碎片化内容
 - 腰部层：主题动画、绘本等周边衍生内容
 - 头部层：音乐节、快闪店、IP节日等热点内容
- **账号矩阵**：品牌官方号+门店号+IP人设号



生态运营

- PGC 生态：基于内容生产矩阵与账号矩阵构建“雪王”文化符号体系，为UGC提供创作原型
- UGC 生态：基于官方素材进行二创，内容主力为表情包、短视频等碎片化内容



“

3.2 “雪王” IP LoRa模型训练与评估

”

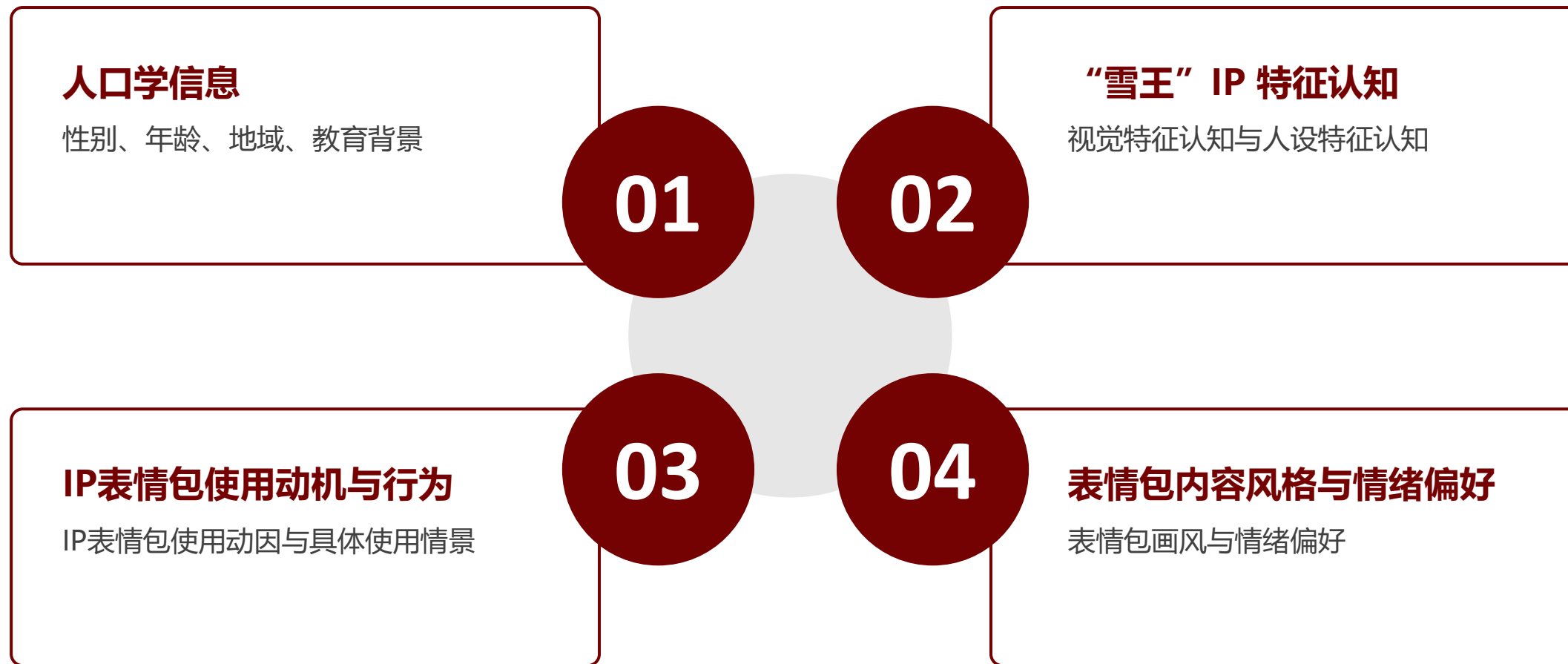


“

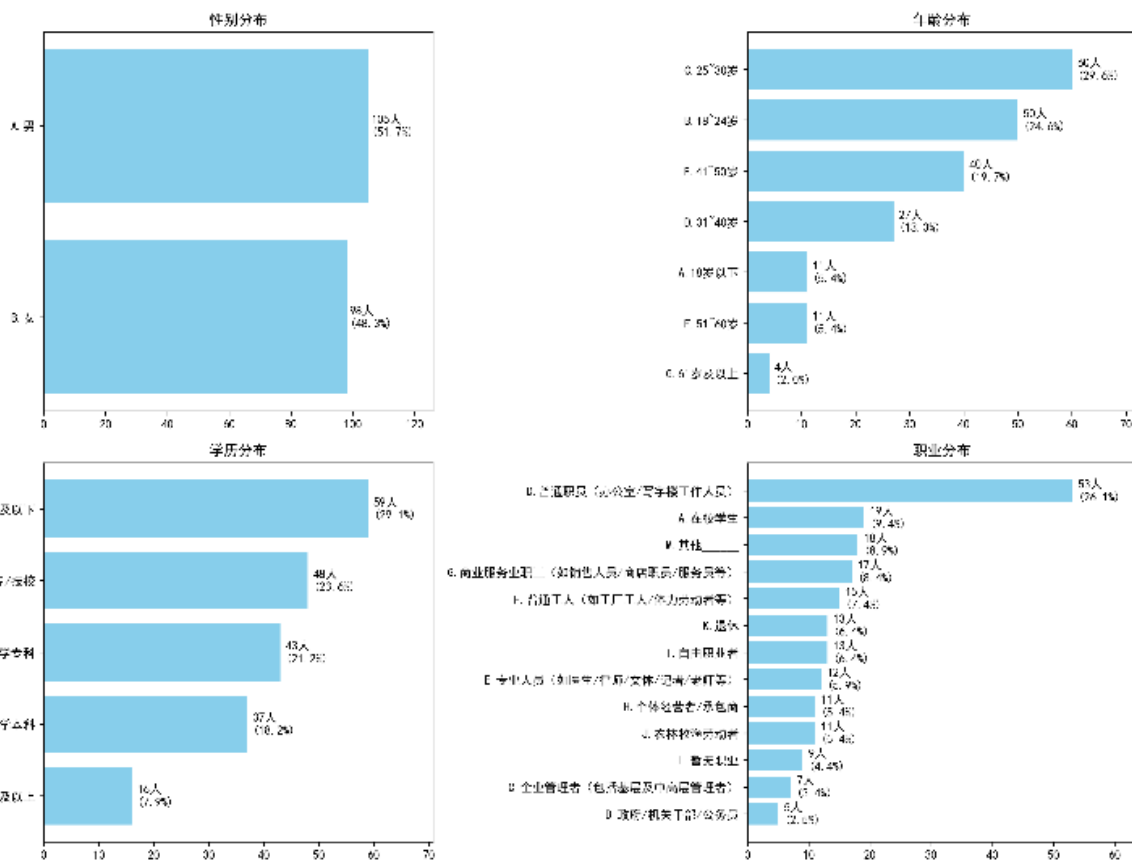
3.2.1 “雪王” IP表情包设计问卷调查

”

:: “雪王” IP表情包设计问卷调查



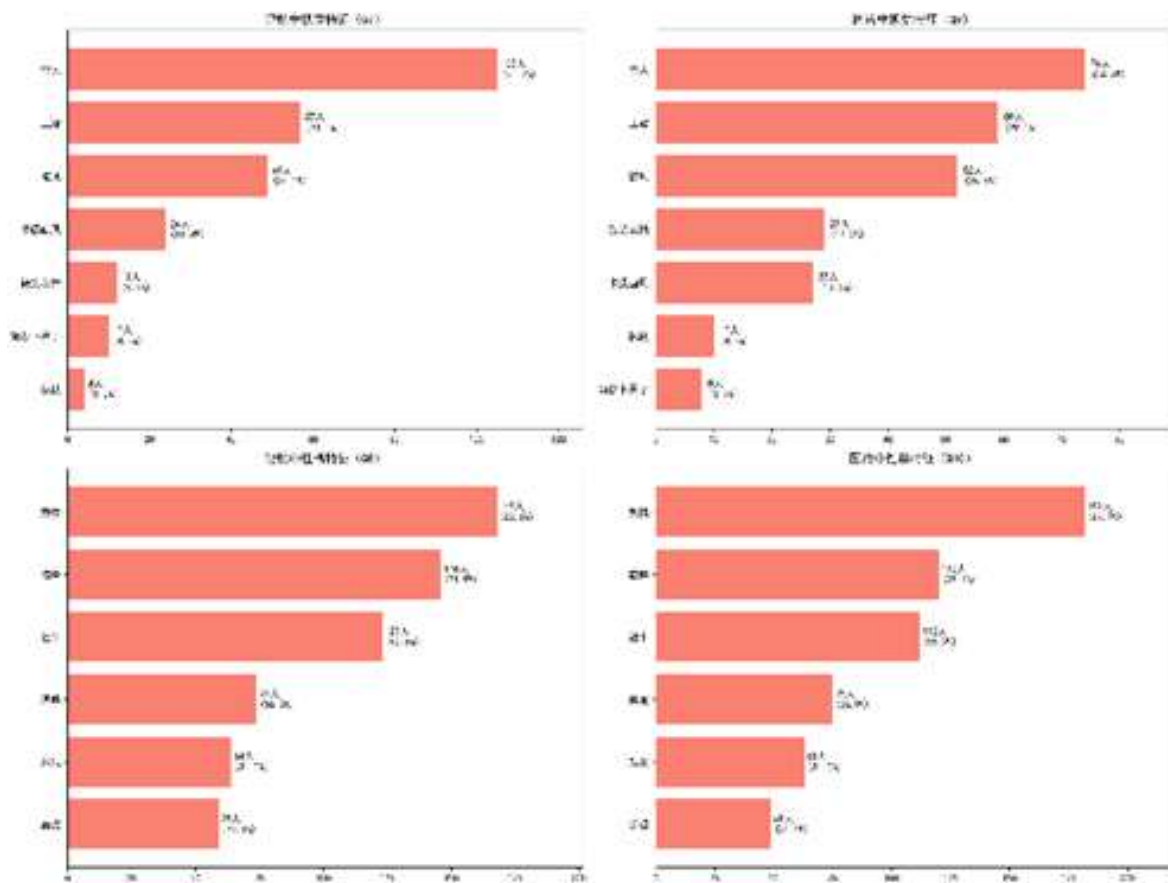
:: “雪王” IP表情包设计问卷调查--描述性统计分析



人口学统计数据柱状图

人口学信息

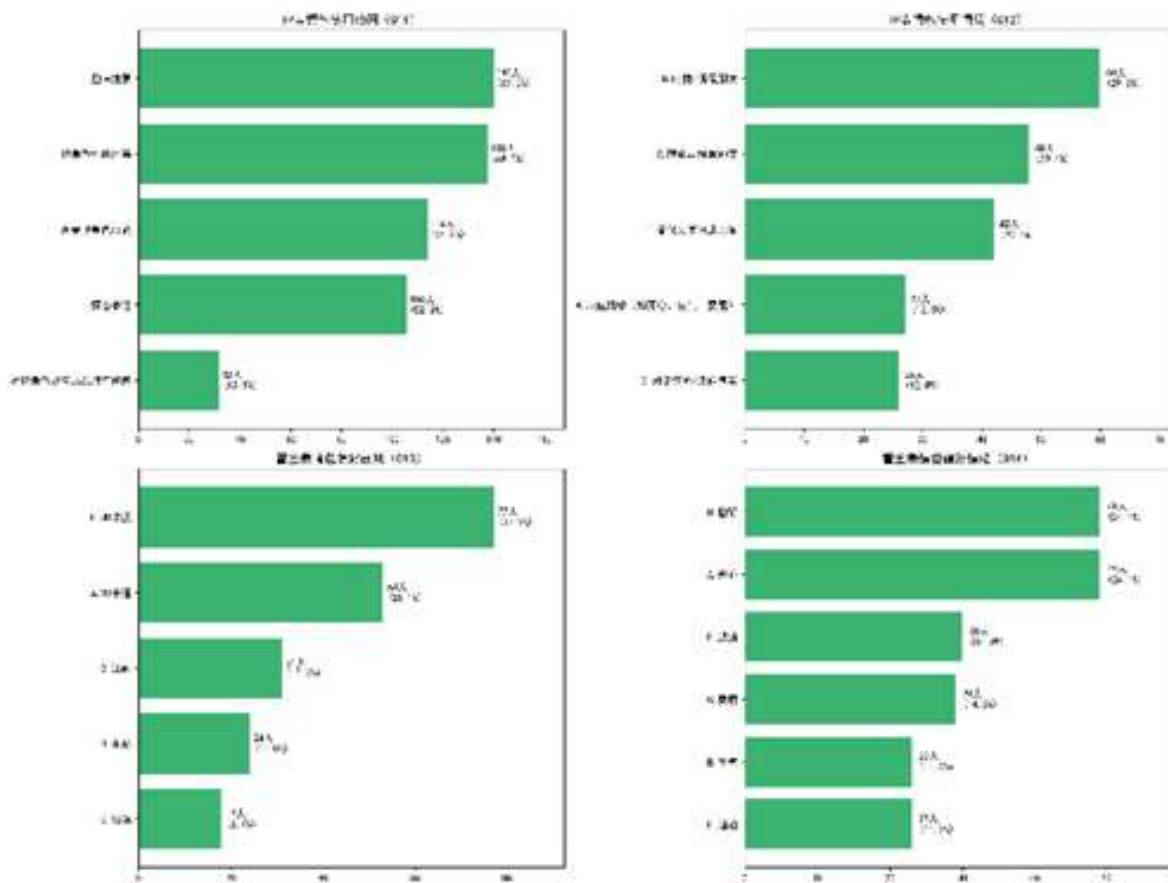
调查问卷通过“腾讯问卷”平台搭建并发放，样本招募采用非概率滚雪球抽样法，问卷链接在微信、微博、小红书等社交媒体渠道进行传播，并引导用户匿名自填。共获得有效样本203份。其中男性占比51.7%，女性占比48.3%。主要集中于年龄18至40岁之间的年轻用户群体，占比67.5%。覆盖全国21个省市地区，其中以东部一二线省市为主。教育背景以高中及以上为主，共占比70.9%。符合蜜雪冰城消费主力人群画像。



“雪王” IP 特征识别数据柱状图

IP特征识别

- 视觉特征认知：80.8%的用户在无图片条件下准确回忆出“雪王”的典型元素（雪人、王冠、披风）；在呈现标准形象后，同样集中于“王冠”“雪人”“披风”三项，在选择行为中的总占比达到71.4%。
- 性格特征认知：“热情”“搞怪”“社牛”是最被认同的三类性格关键词。在呈现标准形象后，占比分别达到82.8%，71.9%以及60.6%



IP表情包使用行为数据柱状图

使用行为与偏好

- 使用动因：选择“趣味性强”“IP角色性格讨喜”“喜爱IP角色形象”的用户比例分别为69%、68%、56.2%
- 使用场景：以吐槽、调侃朋友”“萌化或缓解冲突”“替代文字快速回复”三类为主。
- 生成表情包风格偏好：“3D卡通”“2D卡通”“油画”风格最受欢迎，比例分别为37.9%、26.1%、15.3%。
- 生成表情包情绪偏好：“开心”“撒娇”“鼓励”得票率领先，分别为“24.1”“24.1”以及“14.8”

“雪王”视觉锚点清晰

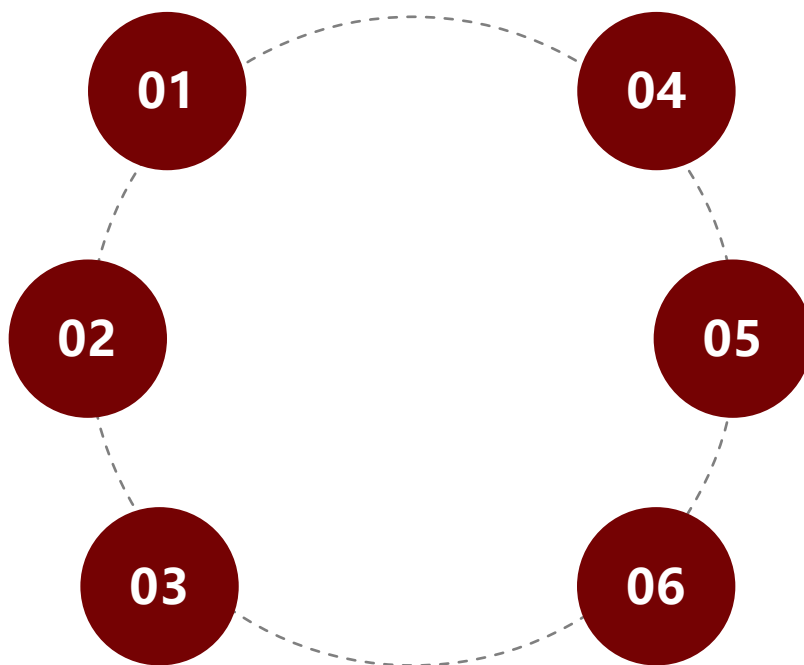
超过80%的用户能在无图条件下准确回忆“雪人、王冠、披风”三要素。

“雪王”性格认知偏正向

性格认知集中于“热情”“搞怪”“社牛”三类，显示出“雪王”亲民、外向、逗趣的形象在用户心智中高度统一。

IP表情包使用动因聚焦 “趣味+认同”双功能

“趣味性”与“角色讨喜”是主因，占比近7成。“喜欢该IP角色”同样驱动用户主动传播表情包。



主要使用场景为非正式交流

“吐槽朋友”“萌化气氛”“快速回复”等非正式语境为高频应用场景。表明IP表情包在日常社交中承担情绪调节与幽默传播作用。

图像风格偏好集中于“3D卡通”和“2D卡通”

用户对表情包视觉风格偏好集中，3D卡通最受欢迎，2D卡通次之。

情绪表现偏好以“开心”“撒娇”“鼓励”为主

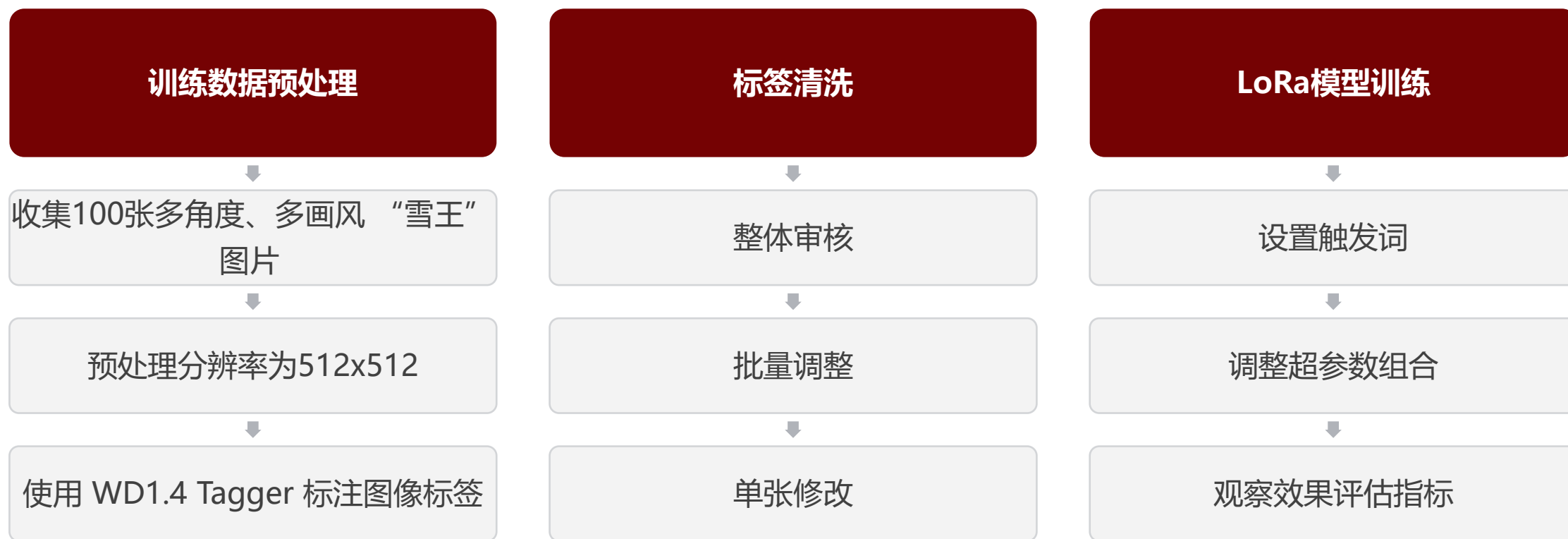
分别获24.1%、24.1%、14.8%的选择，凸显用户倾向具象、情感浓度高的表情形象。



“

3.2.2 “雪王” IP LoRa模型训练

”

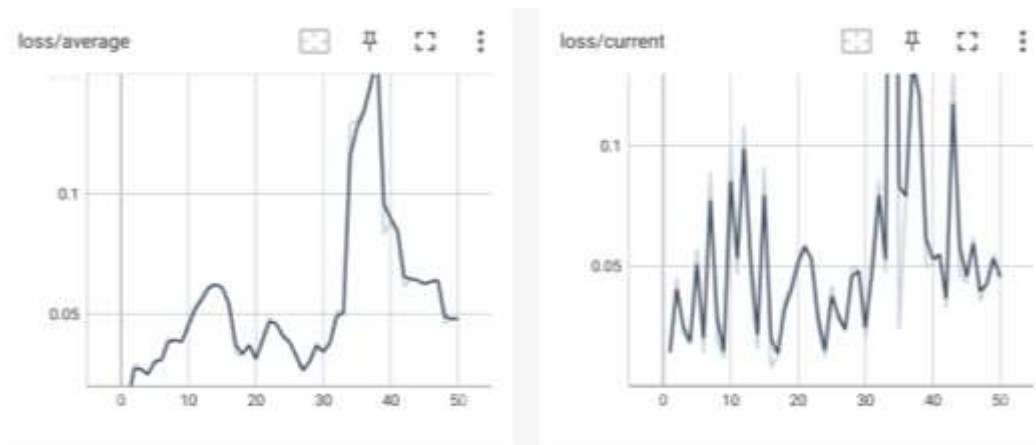


本研究共尝试了42种超参数组合，最终使用的主要LoRa训练参数方案如表3.1

参数类型	具体数据
Learning Rate	1e-4
Batch Size	1
Epoch	12
Dimension	64
Network Alpha	32
optimizer	AdamW

表3.1 LoRa模型训练主要参数表

:: LoRa 模型训练流程



Loss 曲线

- 通过TensorBoard界面观察使用该参数配置下的损失曲线，可以分析得到，训练损失均值与当前训练损失分别稳定在0.044与0.048附近，证明该模型在训练集上收敛，且拟合度较好。



LoRa 模型预览图像

- 在训练过程中，使用该参数配置下训练的“雪王”IP LoRa模型生成预览图片，并与训练集图像进行对比，可以观察到，生成图像中“雪王”IP的标志性元素均得以保留，说明该模型满足生成IP表情包所需的视觉一致性。



“

3.2.3 “雪王” IP LoRa模型效果评估

”

评价维度设计

主观维度

图像自然度：评价图像是否存在明显畸变、模糊、颜色失真等问题，关注图像整体的技术质量与视觉可接受性

特征一致性：评价生成图像是否还原“雪王”IP的核心视觉特征，如白色雪人造型、白色王冠、红色披风等

描述准确性：衡量图像内容与输入 prompt 在语义上的一致性，反映模型对提示词的理解与转化能力

情绪传达度：判断角色表情与肢体动作是否准确表达了提示词设定的情绪状态（如开心、疑惑、委屈等）

客观维度

SSIM (Structural Similarity Index)：衡量生成图像与训练集中参考图像在亮度、对比度和结构层面的视觉一致性。

FID (Fréchet Inception Distance)：基于 Inception-V3 网络评估生成图像与真实图像在深度特征空间中的分布距离，数值越低代表质量越高。

CLIP Score：利用 CLIP 模型计算生成图像与其 prompt 之间的语义相似度，得分越高代表文本-图像匹配越强。

实验组选取

实验组别	Learning Rate	Batch Size	Epoch	Dimension	Network Alpha	Optimizer
基线组	1e-4	1	12	64	32	AdamW
对照组1	2e-4	1	12	64	32	AdamW
对照组2	1e-4	2	12	64	32	AdamW
对照组3	1e-4	1	12	128	32	AdamW

表3.2 实验组 LoRa 模型训练主要参数表

实验设计

用户主观评估问卷调查

各模型训练时使用相同prompt生成唯一预览图

将4组模型的预览图与“雪王”标准形象进行对照，组织用户打分

回收数据，分析评分结果

客观指标计算

采用1.4 Tagger 对训练集图像反推出 100组 prompt

每组模型使用100组 prompt 生成100张图像，与训练集图像配对，计算 SSIM 并取均值

计算每组生成图像的 FID 值

计算每张图像的CLIP Score，取均值

最终采用归一化加权融合的综合评分方法整合主客观维度评分

:: “雪王” IP表情包图像质量评估问卷调查



人口学信息

性别、年龄、地域、教育背景

01

02

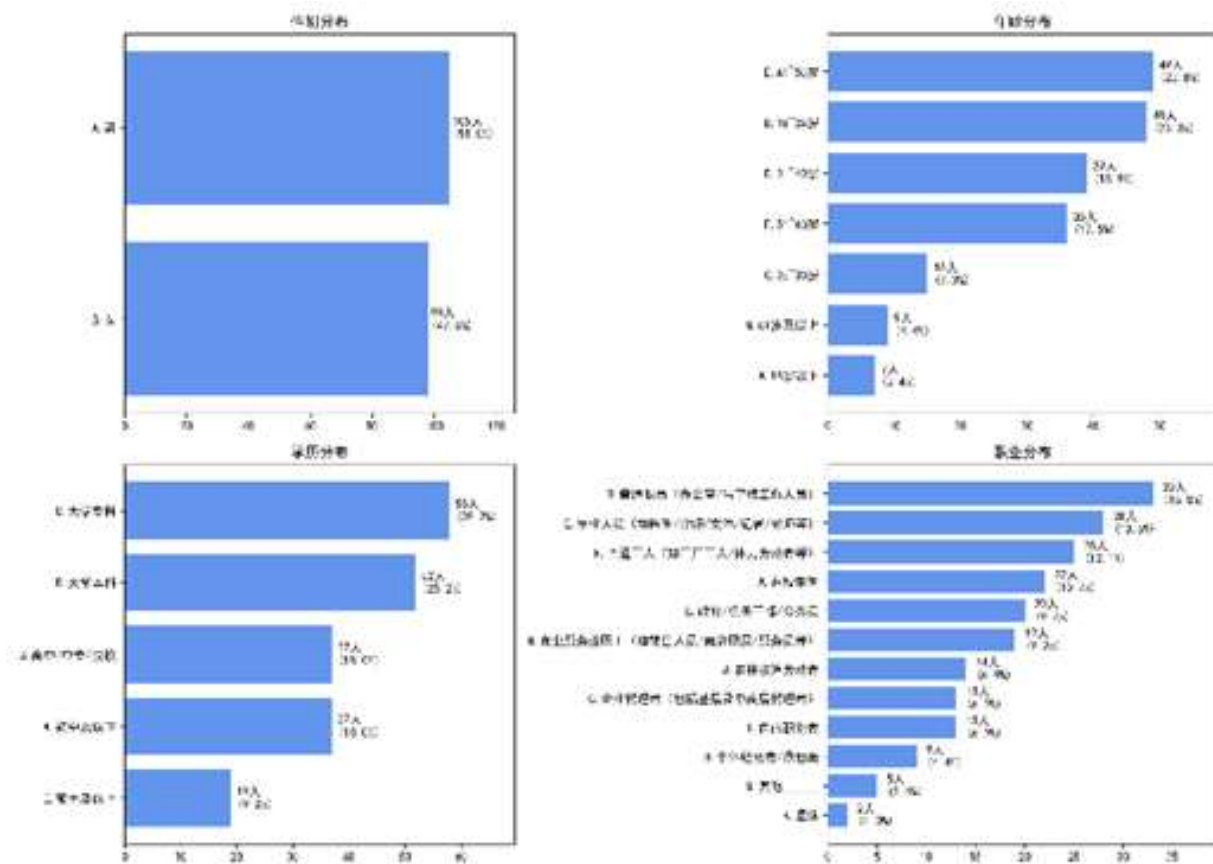
LoRa 模型预览图质量评估

图像自然度、特征一致性、描述准确性、情绪传达度

03

“雪王” IP表情包图像质量评估

图像自然度、特征一致性、描述准确性、情绪传达度、使用意愿



人口学统计数据柱状图

人口学信息

调查问卷通过“腾讯问卷”平台搭建并发放，样本招募采用非概率滚雪球抽样法，问卷链接在微信、微博、小红书等社交媒体渠道进行传播，并引导用户匿名自填。剔除逻辑矛盾、答题时间过短、缺失严重等无效问卷，共获得有效样本203份。其中男性占比51.7%，女性占比48.3%。主要集中于年龄20至40岁之间的年轻用户群体，占比50.2%。覆盖全国20个省市地区，其中以东部一、二线省市为主。教育背景以大学专科、本科、硕士及硕士以上为主，共占比81.8%。



克隆巴赫 α 系数= 0.642，具备统计分析基础

实验组别	图 像 自 然 度 (平均 $\pm$ SD)	特 征 一 致 性 (平均 $\pm$ SD)	描 述 准 确 性 (平均 $\pm$ SD)	情 绪 传 达 度 (平均 $\pm$ SD)	图 像 自 然 度 (平均 $\pm$ SD)
基线组	2.35 $\pm$ 1.16	2.44 $\pm$ 1.14	2.62 $\pm$ 1.17	2.48 $\pm$ 1.16	2.35 $\pm$ 1.16
对照组1	2.10 $\pm$ 1.08	2.28 $\pm$ 1.14	2.29 $\pm$ 1.12	2.35 $\pm$ 1.17	2.10 $\pm$ 1.08
对照组2	2.03 $\pm$ 0.99	2.46 $\pm$ 1.22	2.44 $\pm$ 1.25	2.37 $\pm$ 1.30	2.03 $\pm$ 0.99
对照组3	1.96 $\pm$ 1.05	2.11 $\pm$ 1.16	1.78 $\pm$ 0.99	1.99 $\pm$ 1.11	1.96 $\pm$ 1.05

表3.3 实验组 LoRa 模型训练主要参数表

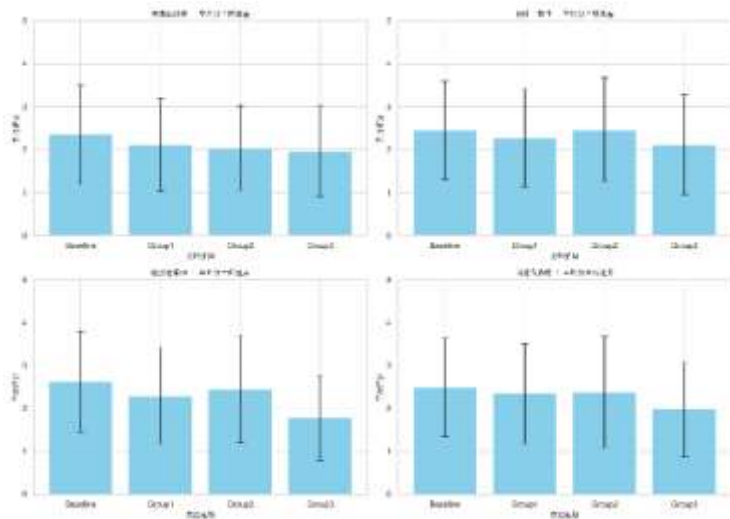


四个维度p值<0.05，均存在显著组间差异，尤其是描述准确性维度差异最为显著

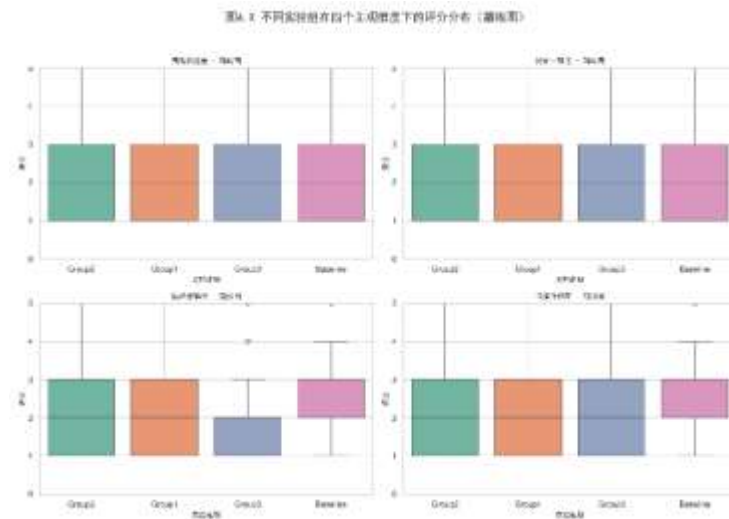
评分维度	F value	Pr(>F)
图像自然度	5.0827	1.7207e-03
特征一致性	3.9256	8.4885e-03
描述准确性	20.5706	7.4984e-13
情绪传达度	6.7549	1.6756e-04

表3.4 LoRa 模型质量评估实验单因素方差分析结果

:: LoRa 模型效果评估



各组评分箱线图



各组评分均值-标准差图

- 基线组在箱线图和均值-标准差图中数据分布较为集中，标准差较小，具有明显优势。这表明，在其他训练参数不变的前提下，适中的学习率（ $1e-4$ ）、单张批处理（batch size=1）与中等网络维度（dim=64）配置能在LoRA模型训练中取得更优的图像生成感知效果。

主观评分采用等权重法，最终得分为四项评分的算术平均值，每项权重为25%，最终综合得分计算公式如下：

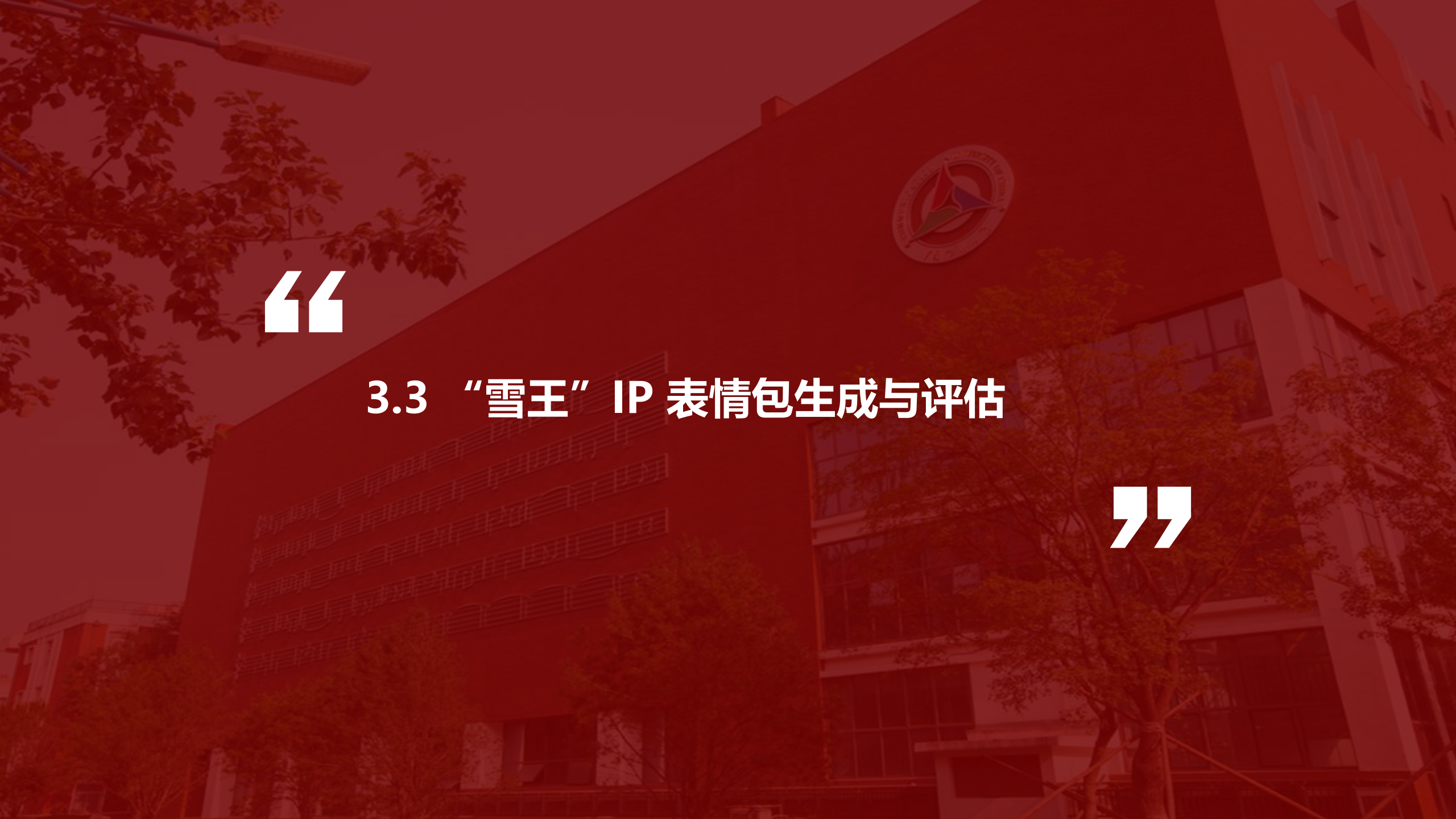
$$Final\ Score = 0.6 \cdot S_{subjective} + (1 - 0.4) \cdot \left(\frac{1}{3} \cdot SSIM + \frac{1}{3} \cdot (1 - FID) + \frac{1}{3} \cdot CLIP \right)$$

其中，主观部分权重设为0.6，反映表情包场景中用户感知的重要性；客观部分各项指标均等加权。

实验结论表明，基线组最终得分最高，评分显著优于对照组1、对照组2与对照组3。因此，本文选定基线组作为后续“雪王”表情包生成实验的主要模型。

实验组别	综合评分
基线组	0.9064
对照组1	0.5976
对照组2	0.5939
对照组3	0.0717

表3.5 各LoRa模型综合评分



“

3.3 “雪王” IP 表情包生成与评估

”

:: “雪王” IP表情包生成流程



:: “雪王” IP 表情包生成流程



选择合适的底模与权重

- 选用了 3 种不同画风的 Stable Diffusion 模型：anything-v5（二次元平涂风）、NEW ERA（复古赛璐璐风）、ONE FOR ALL（兼容多风格）。通过调整LoRA权重（0~1），观察生成图表现。
- 结果表明，ONE FOR ALL 模型在权重为0.8时效果最优，能准确还原“雪王”舌头等特征。

撰写提示词，生成图片

- 基于用户调研和LoRA测试结果，以 ONE FOR ALL为底模撰写高质量提示词，生成“雪王”疑惑表情图像。
- 结果表明，尽管在形象特征和风格控制方面表现良好，但在“playful expression”等表情类提示词表达上仍显不足，说明单靠提示词工程与 LoRA模型对图像内容控制能力有限。

引入 Controlnet 优化生成结果

- 为提升情绪表现力，引入ControlNet结构控制模块，采用SoftEdge预处理器提取边缘特征图，将其作为辅助条件输入生成流程中。
- 结果显示，融合ControlNet后，图像的情感表达更加丰富明确，有效弥补 LoRA模型在表情细节上的不足。

评价维度设计

主观维度

图像自然度：评价图像是否存在明显畸变、模糊、颜色失真等问题，关注图像整体的技术质量与视觉可接受性

特征一致性：评价生成图像是否还原“雪王”IP的核心视觉特征，如白色雪人造型、白色王冠、红色披风等

使用意愿：用户在实际社交中主动使用该表情包的倾向程度。

描述准确性：衡量图像内容与输入 prompt 在语义上的一致性，反映模型对提示词的理解与转化能力

情绪传达度：判断角色表情与肢体动作是否准确表达了提示词设定的情绪状态（如开心、疑惑、委屈等）

实验设计

用户主观评估问卷调查

设置三组生成模型策略组：Stable Diffusion v1.5底模组、底模+ “雪王” IP LoRa组、底模+ “雪王” IP LoRa + Controlnet组以及一组官方 “雪王” 表情包图像组

使用 WD1.4 tagger 根据官方 “雪王” 表情包图像组反推，并人工清洗错误标签后，得到10个 prompt 模版。每个对照组使用10个 prompt 模板各生成10张图像

将4组图像进行对照，组织用户打分

回收数据，分析评分结果

最终采用归一化加权融合的综合评分方法整合主客观维度评分

:: 表情包图像质量评估



克隆巴赫 α 系数= 0.964，问卷内部一致性优秀，具备统计分析基础

评分维度	图像自然度	特征一致性	描述准确性	情绪传达度	使用意愿
底模组	2.08±0.49	2.10±0.47	2.10±0.51	2.06±0.46	2.01±0.51
LoRa模型组	2.28±0.57	2.26±0.58	2.24±0.53	2.30±0.59	2.18±0.55
LoRa模型+Controlnet组	2.23±0.55	2.33±0.58	2.31±0.60	2.27±0.59	2.25±0.55
官方表情包组	2.25±0.56	2.31±0.53	2.27±0.59	2.26±0.58	2.28±0.60

表3.6 表情包图像质量评估实验各组主观评分表

:: 表情包图像质量评估

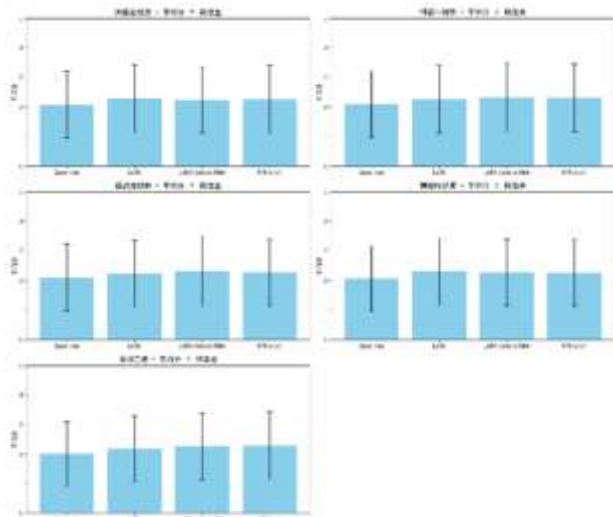


五个维度在四组之间均存在显著差异（ $p < 0.05$ ），表明生成图片时是否使用LoRa模型与Controlnet对生成结果的主观体验存在显著影响。

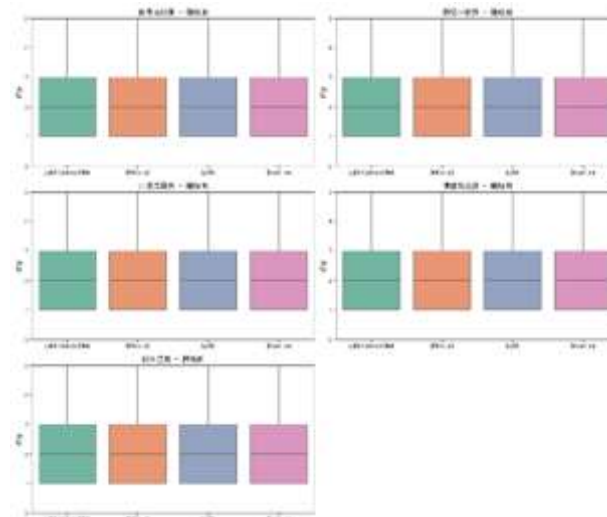
评分维度	F value	Pr(>F)
图像自然度	12.7088	2.7725e-08
特征一致性	15.7624	3.2248e-10
描述准确性	13.5150	8.5684e-09
情绪传达度	19.0435	2.6550e-12

表3.7 表情包图像质量评估实验各维度单因素方差分析结果

:: 表情包图像质量评估



各组评分箱线图



各组评分均值-标准差图

- 从箱线图可以看出各组得分的分布范围和中位数位置。LoRA及LoRA+ControlNet组在多个维度的上四分位表现优于底模组，表明使用LoRa模型不仅提升了平均水平，也扩大了用户对生成结果的接受区间。
- 此外，均值图中，LoRA+ControlNet组与官方表情包组在五个维度中整体表现优于底模组，LoRA组也较底模组有明显提升，验证了LoRA微调对图像质量与表达效果的积极影响。特别地，LoRA+ControlNet在“描述准确性”与“特征一致性”两个维度上得分高于其他组，显示了Controlnet技术对语义还原及IP还原能力的强化。

:: 表情包图像质量评估



主观评分采用等权重法，最终得分为四项评分的算术平均值，每项权重为25%，最终综合得分计算公式如下，其中 S_i 分别代表五项指标的平均评分：

$$S_{subjective} = \frac{1}{5} \sum_{i=1}^5 S_i$$

基于上述评分公式，对五项维度的平均值进行等权重合成，得到各实验组的综合主观评分，结果如表3.8所示。

主观评价结果显示 LoRA+ControlNet 模型在多个关键维度上表现优异，验证了本研究提出的工作流在表情包图像生成任务中的有效性，具备良好的实际应用潜力。因此，本文最终选定 LoRA模型+ControlNet 作为 ‘雪王’ IP表情包图像生成的核心技术方案。

实验组别	综合评分
底模组	2.07
LoRa模型组	2.25
LoRa 模型 +Controlnet组	2.28
官方表情包组	2.27

表3.8 各LoRa模型综合评分

:: 表情包图像质量评估



图3.1 采用LoRa模型+Controlnet生成的3D
风格表情包图像



图3.2 采用LoRa模型+Controlnet生成的2D
风格表情包图像及其过程图像



4

结论与展望

Conclusions and prospects

- 研究结论
- 改进方向

基于 Stable Diffusion 的蜜雪冰城 “雪王” IP 表情包生成研究

研究结论

LoRa+Controlnet 的品牌IP表情包生成 workflow 效果优于官方表情包

本研究围绕 “雪王” IP，从0开始构建了完整的AI生成工作流程，包括数据标注与清洗、LoRA训练与效果评估、提示词工程与图像生成、生成图像优化4个步骤。

- 同时通过设计表情包图像质量评估实验，证明所提出 workflow 产出的内容优于现有官方内容。

研究不足

训练数据尚需补充高复杂度样本

- 面对复杂的场景、多人物互动或跨模态条件下，模型可能仍存在泛化能力不足的问题。

模型调参存在权衡限制

- LoRA模型涉及多个超参数维度（学习率、秩值、优化器等），使得当前最优模型可能并非全局最优解。

模型效果评估体系解释力需提高

- 客观指标本身具有不同量纲与敏感性，有可能会产生与人类感知差异较大的结果，影响最终评估

改进方向

进一步拓展数据源

- 拓展数据源，提升样本数量与场景多样性

引入自动化调参工具

- 引入自动化调参工具，通过贝叶斯优化等算法自动探索最优超参数组合

进一步精细化评估体系

- 在主观评分体系中引入“用户类型建模”机制，构建更具解释力的分层评分模型。
- 采用多指标决策方法优化客观评分指标权重体系，提高评估体系适应性与解释力。



汇报完毕

恳请各位老师批评指正