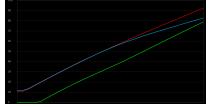
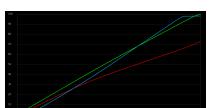
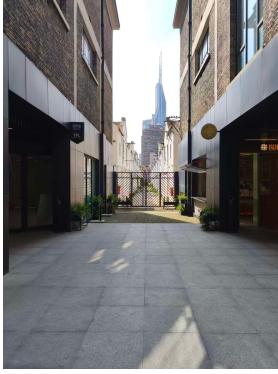
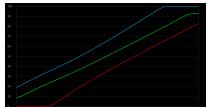


CLIP-DLUT

自然语言驱动的生成式风格化调色引擎。使用基于 CLIP 语义引导的零样本图像风格化框架，通过构建可微分的 3D LUT 演化场，实现从自然语言到对应高维色彩流形的映射。

效果展示

本框架能够精准捕捉自然语言中的色彩情绪。以下结果均为 **Zero-shot**（零样本）生成，无需针对特定风格进行微调。

原始图像 (Original)	引导词 (Prompt)	风格化结果 (Stylized)	3D LUT 可视化 (IRE Chart)
	"赛博朋克，霓虹灯光，冷蓝与玫紫对比"		
	"深秋森林，枯叶橙，低沉的电影感"		
	"高斯冷调，冷漠与寒意"		

产品价值

我们认为，当前大众化图像处理市场（手机修图 App、短视频、图片分享平台）存在较为显著的供需不匹配矛盾：平台提供的预设风格化滤镜虽然数量不断增加，但相较于海量用户对于对图片进行风格

化调色的实际需求相比仍然存在着巨大的鸿沟，而平台提供的参数调整器作为专业后期工作流的简化下放，对于大多数不具备色彩科学知识的用户来说门槛依然偏高。这导致了这导致了用户创作表达的严重趋同与体验断层。

而我们的产品为用户提供了一种全新的交互范式：用自然语言描述你想要的画面风格，系统自动生成对应的专业级调色方案。用户只需上传一张图片，输入提示词，系统即可在无需用户任何色彩学知识的前提下，自动生成符合语义的风格化调色结果，并输出工业标准的 .cube 3D LUT 文件，可直接导入 Adobe Photoshop、Premiere Pro、DaVinci Resolve 等专业工具或任何支持 LUT 的平台进行复用、分享。这意味着创作门槛的进一步降低。用户无需经历从感性风格到理性参数的翻译过程，也无需掌握专业的调色知识就能够进行更加个体化的审美表达。同时，由于产出物是标准化的工业资产，用户创造的审美价值不会被锁死在单一生态内，具备分享，交易和复用的潜力。

交互设计

为快速展示项目，我们使用 Gradio 构建了一个可视化的操作界面（图-placeholder）。用户在界面可以上传待调色的图片，设计调色 prompt，并能预览实时迭代的调色效果和下载 .cube 文件。

同时，我们还正在基于 Flutter 进行跨平台应用的进一步开发，将调色功能集成到相机内，同时引入 Agent 自动撰写 prompt 实现图像增强，进一步提高智能化水平。我们希望构建的是一个智能的调色相机 App。用户可以在该 App 内体验到图像拍摄->智能调色->平台分享的全流程。在未来还将引入 LUT 分享社区，为构建门槛更低的调色生态打下基础。

使用技术

本系统构建了一个基于 CLIP 语义引导的零样本图像风格化深度学习框架，通过将色彩变换参数化为可微分的三维查找表（3D LUT）并结合多目标约束优化，实现了从自然语言到高维色彩流形的端到端映射。在模型架构层面，我们采用预训练的 ChineseCLIP（OFA-Sys/chinese-clip-vit-large-patch14-336px）作为多模态语义编码器，该模型基于 ViT-Large/14 架构并在大规模中文图文对上训练，能够将图像和文本投影到共享的 768 维语义空间。与传统的风格迁移方法不同，我们并不直接优化图像像素，而是将色彩变换显式建模为一个 $33 \times 33 \times 33$ 的三维查找表 $\mathcal{L} : [0, 1]^3 \rightarrow [0, 1]^3$ ，该表通过三线性插值算子 \mathcal{T} 作用于输入图像 I ，即 $I' = \mathcal{T}(I, \mathcal{L})$ ，这种参数化方式不仅将优化维度从数百万像素压缩至约 10 万个 LUT 顶点，更关键的是能够保证变换的空间一致性并直接输出工业标准的 .cube 格式 LUT 文件。

在优化流程的设计上，我们借鉴 Langevin 动力学的思想构建了一个可微分的 LUT 演化场：通过一个四层全连接网络 $\phi_\theta : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ 作为梯度预测器（Gradient Predictor），该网络将 LUT 的每个三维色彩坐标映射为一个梯度向量，随后通过动力学更新规则 $\mathcal{L}_{t+1} = \mathcal{L}_t + \frac{\epsilon}{2} \phi_\theta(\mathcal{L}_t)$ 迭代演化 LUT 流形，其

中 ϵ 为演化步长。这种设计将 LUT 的离散采样点转化为连续场的采样，使得优化过程更加稳定且具备理论上的平滑性保证。在损失函数设计上，我们同时使用了一个方向性语义损失（Directional CLIP Loss）和传统的余弦相似度损失：记原始图像特征为 \mathbf{f}_{orig} ，目标文本特征为 $\mathbf{f}_{\text{target}}$ ，原始文本特征为 \mathbf{f}_{src} ，我们首先计算语义风格方向 $\mathbf{v}_{\text{style}} = \frac{\mathbf{f}_{\text{target}} - \mathbf{f}_{\text{src}}}{\|\mathbf{f}_{\text{target}} - \mathbf{f}_{\text{src}}\|_2}$ ，再计算风格化后图像的编辑方向 $\mathbf{v}_{\text{edit}} = \frac{\mathbf{f}_{\text{edit}} - \mathbf{f}_{\text{orig}}}{\|\mathbf{f}_{\text{edit}} - \mathbf{f}_{\text{orig}}\|_2}$ ，最终的 CLIP 损失定义为 $\mathcal{L}_{\text{CLIP}} = 10.0 \cdot (1 - \cos(\mathbf{v}_{\text{edit}}, \mathbf{v}_{\text{style}})) + \alpha \cdot (1 - \cos(\mathbf{f}_{\text{edit}}, \mathbf{f}_{\text{target}}))$ ，其中第一项强制编辑方向与语义风格方向对齐，第二项保证目标语义的绝对匹配， α 为内容权重系数设定为 0.7，这种方向性约束能够更好地捕捉“从 A 到 B”的风格化本质而非简单的特征相似性。

为确保生成的 LUT 在色彩科学层面的合理性与视觉连贯性，我们设计了一组在感知均匀色彩空间 CIELAB 上的统计约束与几何约束。首先，单调性约束通过惩罚 LUT 在 RGB 三个维度上的负梯度 $\mathcal{L}_{\text{mono}} = \sum_{d \in \{R, G, B\}} \text{ReLU}(-\nabla_d \mathcal{L})$ 防止出现颜色反转现象，同时通过二阶差分约束 $\mathcal{L}_{\text{smooth}} = \sum_d \|\nabla_d^2 \mathcal{L}\|_2^2$ 惩罚 LUT 网格的曲率以避免局部锯齿和不自然的色彩跳变。

其次，我们在实验过程中发现，CLIP 的风格化引导往往粗暴地将全局色彩向纯色化推动，这显然和大众普遍认知的风格化色彩存在巨大差异，于是我们在 LAB 色彩空间中引入了色彩体积保持约束：将图像在 LAB 空间的像素分布视为三维点云并计算其协方差矩阵 $\Sigma = \frac{1}{N} \sum_{i=1}^N (\mathbf{p}_i - \bar{\mathbf{p}})(\mathbf{p}_i - \bar{\mathbf{p}})^\top$ ，通过约束 $\mathcal{L}_{\text{vol}} = \text{ReLU}(\log |\Sigma_{\text{orig}}| - \log |\Sigma_{\text{edit}}|)$ 确保风格化后的图像色彩分布不会过度塌缩，同时通过最小特征值约束 $\mathcal{L}_{\text{eigen}} = \text{ReLU}(\lambda_{\min}(\Sigma_{\text{orig}}) - \lambda_{\min}(\Sigma_{\text{edit}}))$ 防止色彩空间在某一维度上退化。此外，色彩均值偏移损失 $\mathcal{L}_{\text{shift}} = \|\bar{\mathbf{p}}_{\text{edit}} - \bar{\mathbf{p}}_{\text{orig}}\|_2$ 约束图像的整体色调偏移幅度，防止出现不自然的全局色偏。最终的总损失函数为各项加权和： $\mathcal{L}_{\text{total}} = w_1 \mathcal{L}_{\text{CLIP}} + w_2 \mathcal{L}_{\text{mono}} + w_3 \mathcal{L}_{\text{smooth}} + w_4 \mathcal{L}_{\text{vol}} + w_5 \mathcal{L}_{\text{shift}} + w_6 \mathcal{L}_{\text{eigen}}$ ，其中各权重通过实验调优设定为 (1.0, 1.0, 1.0, 1.0, 20.0, 10000.0)，高权重的色彩偏移和特征值损失能够有效抑制色彩空间的病态退化。

在实现细节上，整个流水线被设计为端到端可微分：图像预处理模块通过可微分的双三次插值（Bicubic Interpolation）和缩放裁剪实现 CLIP 输入的标准化，LUT 应用模块通过 PyTorch 的 grid_sample 算子实现高效的三线性插值并支持梯度回传，LAB 色彩空间转换通过分段函数和幂运算的平滑近似保证数值稳定性。优化器采用 AdamW 配合余弦退火学习率调度（初始学习率 2×10^{-4} ，最小学习率 10^{-6} ），典型迭代次数设定为 1000 步，整个优化过程在单张 NVIDIA RTX 4060 GPU 上约需 2~3 分钟。

在工程架构层面，我们基于 FastAPI 构建 RESTful API 服务，通过 Celery 分布式任务队列与 Redis 消息中间件实现异步非阻塞的任务调度，支持多任务并发处理与实时进度反馈，确保在生产环境下的高可用性与可扩展性。最终系统输出除了风格化图像外，还包括标准格式的 .cube 3D LUT 文件以及 PNG 格式的 LUT 可视化条带图，可直接导入 Adobe Premiere Pro、DaVinci Resolve 等专业调色软件或通过 OpenCV、FFmpeg 等工具链应用于视频流处理，实现了从语义描述到工业级色彩资产的完整闭

环。

未来展望

在未来，我们希望能够开发这一项目：在算法层进一步提高计算效率的同时，能够进一步优化调色的指令遵循度，同时引入区域遮罩 LUT 机制，实现更灵活的调色。在应用层，我们将引入具有视觉功能的大语言模型 Agent 进行调色自主决策，提高自动化水平，同时构建跨平台的调色相机程序和 LUT 分享平台，为用户提供更好的体验。