

# CLIPasso: Semantically-Aware Object Sketching

## Abstract

动机：计算机如何“理解”图片抽象的含义（Could a computer convert a photograph from a concrete depiction to an abstract one），本文巧妙地通过简笔画的数量来衡量抽象的程度

## 1. Introduction

草图sketch，由一系列笔画strokes组成，四个控制点control points构成了笔画stroke

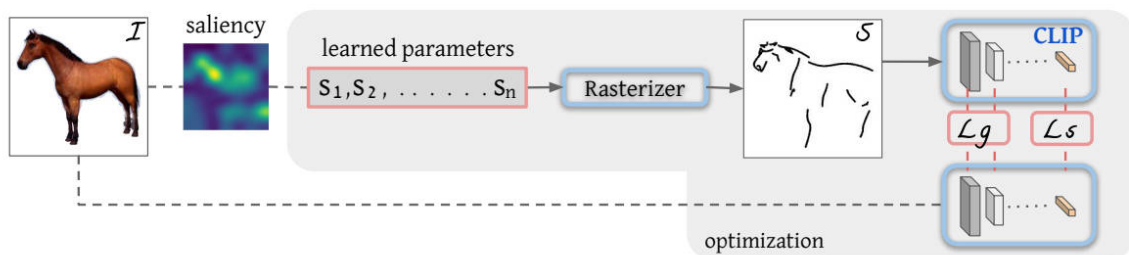
## 2. Related Work

前人研究的缺陷：

固定数据集，data-driven，风格和总类固定

要么只重视geometric representation，要么只重视semantic representation

## 3. Method



CLIP得到semantic信息

Image自身得到geometric信息

为增加robustness，改进了initialization，基于saliency的方法，利用Vision Transformer得到显著区域，在显著区域中初始贝塞尔曲线

带有背景的图片，利用U2-Net先将主体“抠出来”

### 3.1 Loss Function

loss function：兼重几何特征——几何形状、笔画位置 ( $L_g$ ) 和语义特征 ( $L_s$ )

### 3.2 Optimization

初始化：基于saliency（Image=>Vision Transformer=>去最后一层Multi-Head Self-Attention的加权平均=>Saliency Map=>得到显著区域，然后在这些显著区域上去采点）

### 3.3 Strokes Initialization

loss function是non-convex的，所以初始化至关重要

基于saliency的方法

后处理：从3张图片中，选择1张具有最低loss的做为最终的输出结果

## 4. Results

### 4.1 Qualitative Evaluation

### 4.2 Comparison with Existing Methods

### 4.3 Quantitative Evaluation

## 5. Limitations

① 图像有背景时，使用U2-Net，将带背景图片中的主体部分“抠出来”，但这样一来就变成了two step，而在深度学习时代，这并不是最优的结构（end-to-end）

② 笔画是同时生成的，不是依序生成的

③ 需要提前制定笔画数量，这就成一种超参数了，把它改为一种可学习的参数，自动适配图片需要的抽象程度

## 6. Conclusion

对不常见物体也能生成简笔画；

能达到任意程度的抽象（笔画数量）；

能实现不同的风格（控制一个笔画的控制点的数量）；