

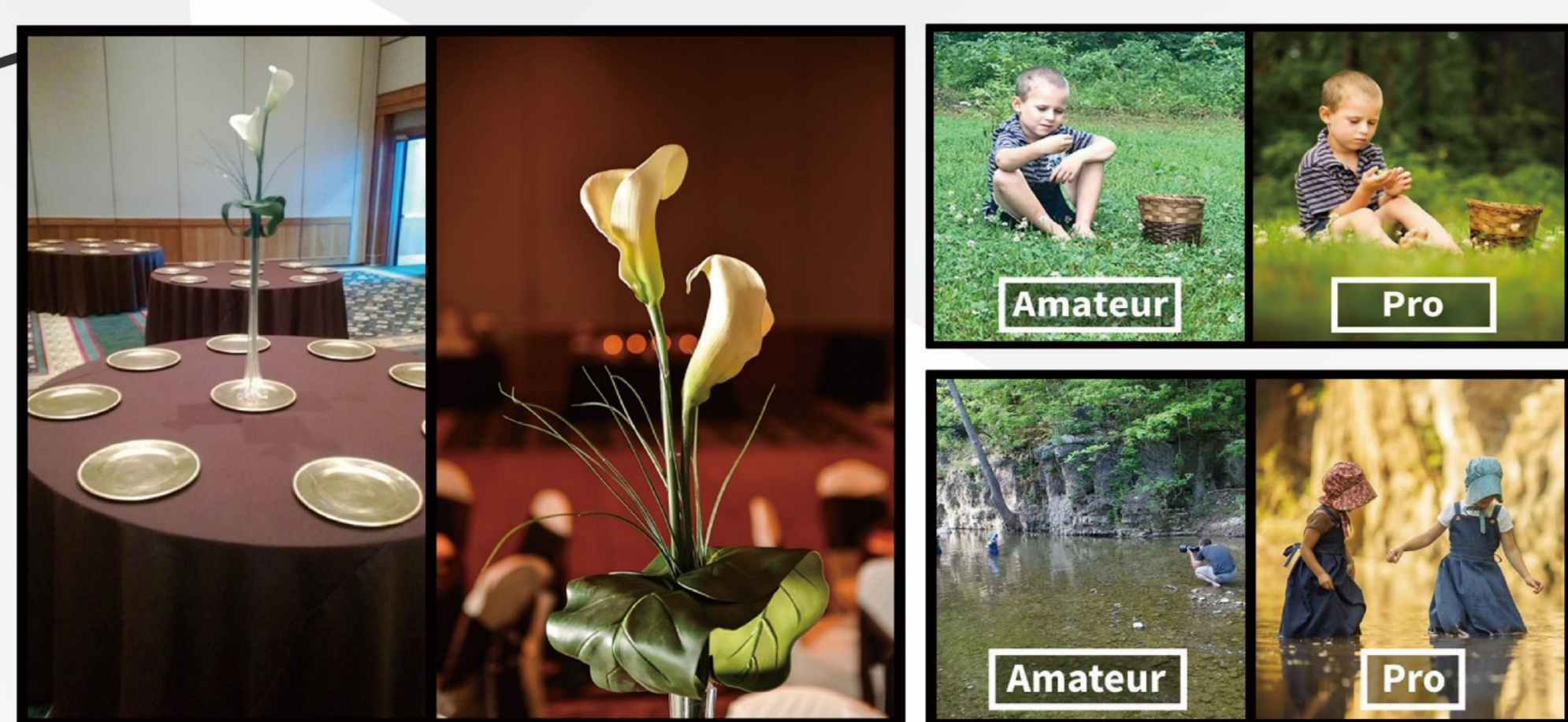
SmartEye

SmartEye: Assisting Instant Photo Taking via Integrating User Preference with Deep View Proposal Network.

马帅, 田丰, 范向民, 魏子钧, 黄进, Jianming Zhang, Radomir Mech, Dimitris Samaras, 王宏安
主要联系人: 马帅 手机号: 13126967288 邮箱: mashuai171@mails.ucas.ac.cn

ACM CHI 2019 Honorable mention

问题产生——日常生活中的拍照构图需求 (Photo Composition)



良好的构图对一张照片的质量来说十分重要, 同样的景物, 不同的拍摄手法会导致截然不同的效果但是非专业的人很难掌握构图技巧, 拍摄出高质量的照片, 现有的一些构图方法存在着几个突出的问题:

- ①一些后期处理的方法不具有实时性 (例如美图APP, PS 等, 常规的方法都是先拍照, 后修图)
- ②构图评估算法无法提供与人的交互 (一些自动的算法没有合适的交互手段来供用户使用)
- ③构图是一件十分主观的任务 (每个人的审美偏好不同, 如何为用户提供符合他口味的构图推荐)

如何解决上述三个基本问题?



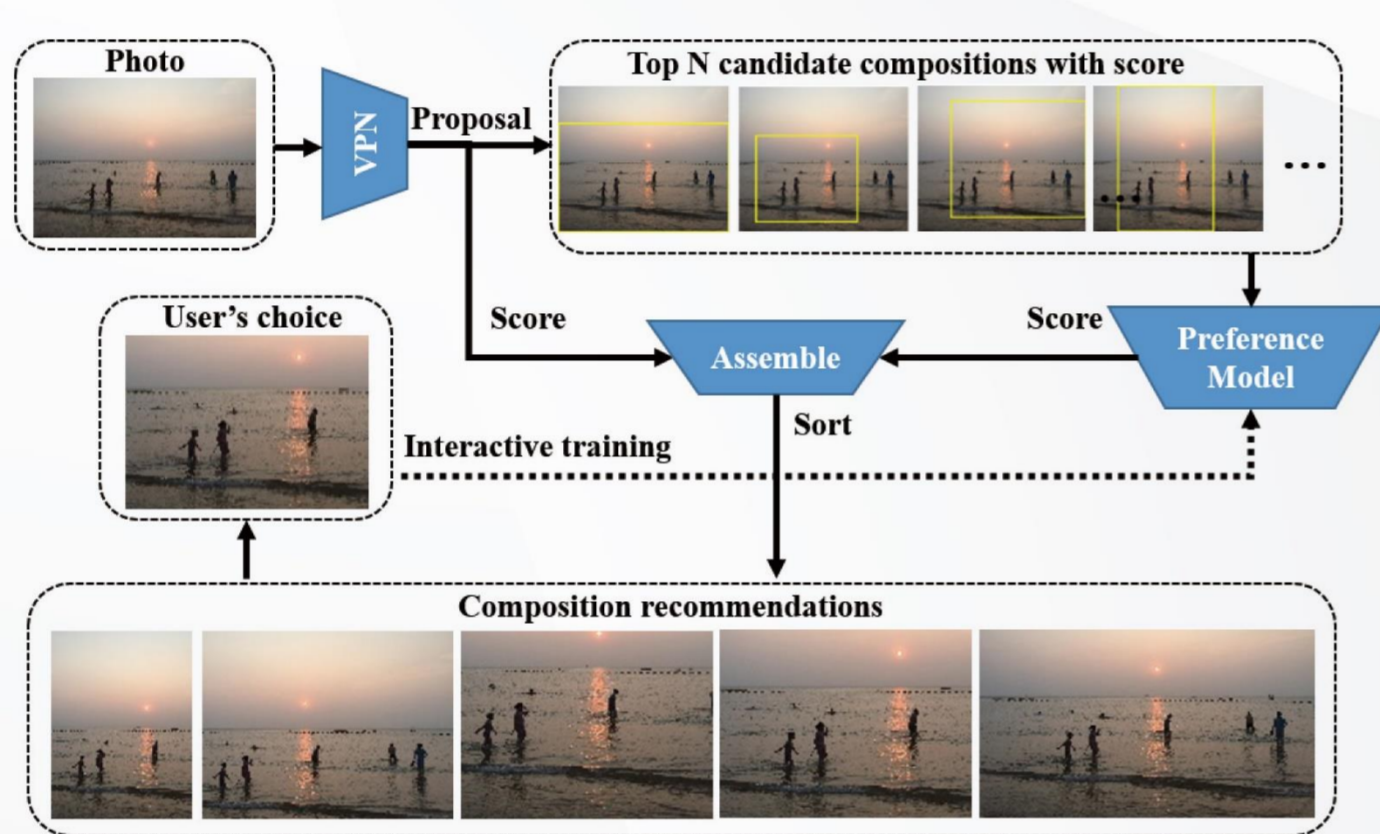
左) 构图推荐算法为当前镜头生成多种构图推荐 右) 我们的系统交互式地逐渐掌握用户偏好

1. 利用深度学习技术, 将构图推荐算法提升到一个实时的、可在线的、高质量的水平, 解决现有的两步走离线构图问题
2. 设计一个新颖的交互界面, 将后台的构图推荐算法通过可视化的形式展示在移动端, 与用户进行无缝友好的交互
3. 引入用户偏好学习模块, 推测用户意图, 交互式地不断学习用户的构图喜好
 - (a) 构图推荐网络给定一张原始照片, 给出最好的N的推荐构图
 - (b) 我们将算法部署到移动端, 并融入用户偏好学习模块, 交互式的为用户推荐越来越合口味的高质量构图

SmartEye系统介绍——算法部分

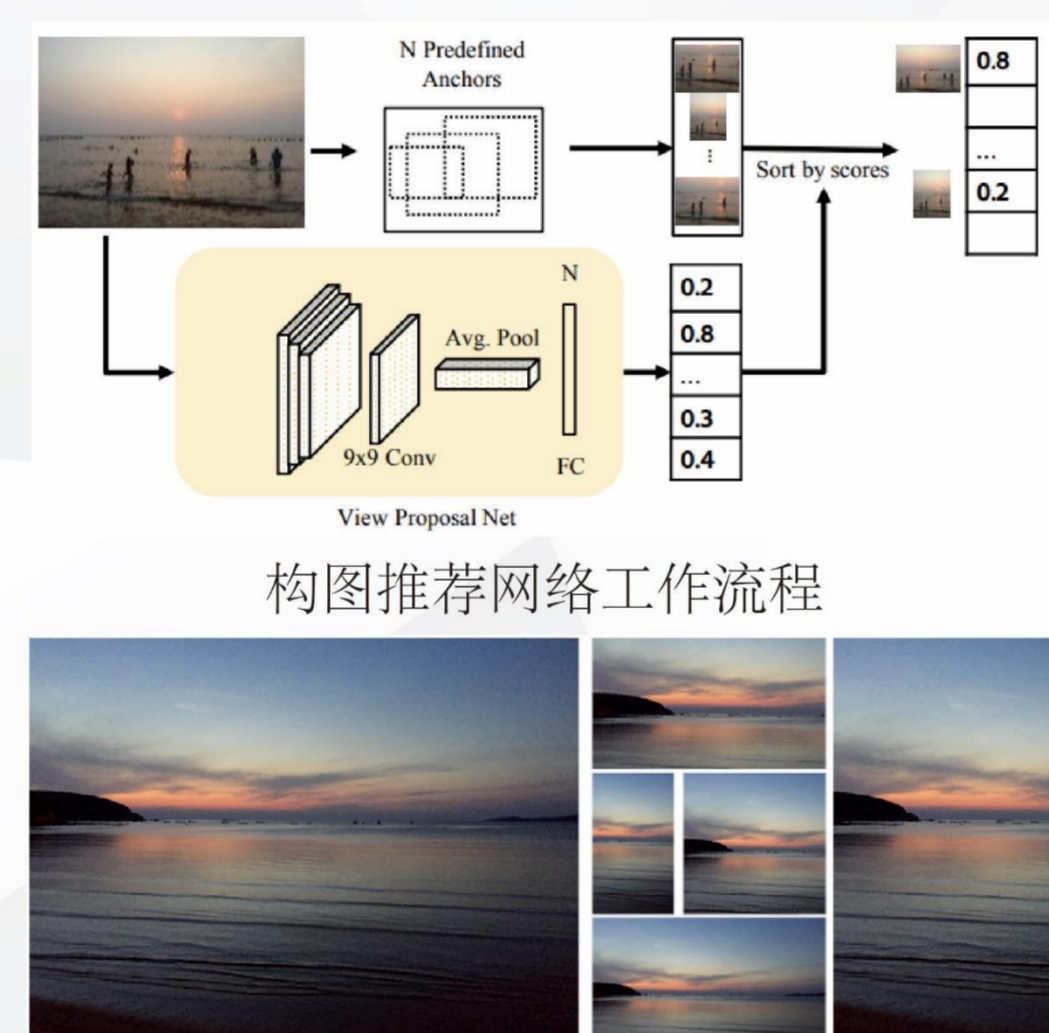
算法流程

- 给定一张原始照片, 通过VPN(View Proposal Network)可以得到最好的几种构图推荐以及每一个构图的得分。
- Preference Module对VPN给出的推荐构图进行打分
- P-Module和VPN二者的打分根据置信度(confidence value)进行动态的融合, 最后按得分高低对推荐构图排序输出
- 利用用户最终满意的构图作为训练样本交互式的训练P-Module



View Proposal Network(构图推荐网络)

VPN是由百万张带标注图片数据训练而来的深度神经网络, 功能是输入一张照片, 推荐出N种构图方式, 并且对每种构图进行打分, 最终输出排名前N的构图及其得分。但是基于大量数据训练出来的深度学习模型只能符合人类的一种“共性”的喜好。萝卜青菜各有所爱, 构图好不好看并不是一个能够用规则来衡量的事情, 所以如何为不同的用户搭建适合自己的构图推荐模型成了我们最需要考虑的问题。



左) 给定一张照片

右) VPN会给出得分最高的5个推荐构图

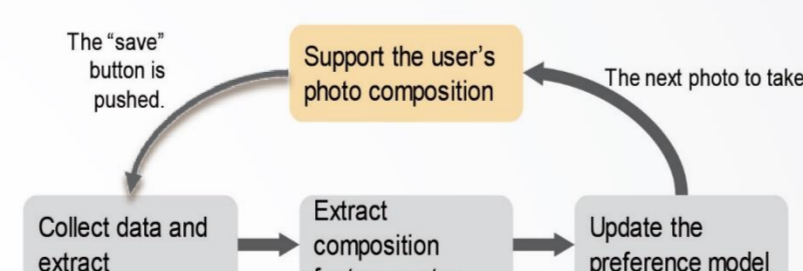
P-Module (用户偏好学习模块)

Number	Type	Feature
1	1	Position of subject
2	1	Position of subject
3	1	Position of subject
4	1	Position of subject
5	1	Position of subject
6	1	Position of subject
7	1	Position of subject
8	1	Position of subject
9	1	Position of subject
10	1	Position of subject
11	1	Position of subject
12	1	Position of subject
13	1	Position of subject
14	1	Position of subject
15	1	Position of subject
16	1	Position of subject
17	1	Position of subject
18	1	Position of subject
19	1	Position of subject
20	1	Position of subject
21	1	Position of subject
22	1	Position of subject
23	1	Position of subject
24	1	Position of subject
25	1	Position of subject
26	1	Position of subject
27	1	Position of subject
28	1	Position of subject
29	1	Position of subject
30	1	Position of subject
31	1	Position of subject
32	1	Position of subject

P-Module是基于Logistic Regression的交互式机器学习模型, 我们通过用户调查的方式收集了用户对于照片构图的看法和建议, 结合专业知识, 设计了4类32维特征来表征用户对照片构图的喜好。

训练方法: 用户每选择一个满意的构图按下保存按钮, 就默认为一个正样本, 并从未被选择的构图中随机选取一个作为负样本, 随着用户使用, 不断地获取新的样本对模型进行再训练。功能: 为每一个构图进行打分

$$S_i^P = \frac{1}{1 + e^{W\phi(x_i, I) + b}}$$



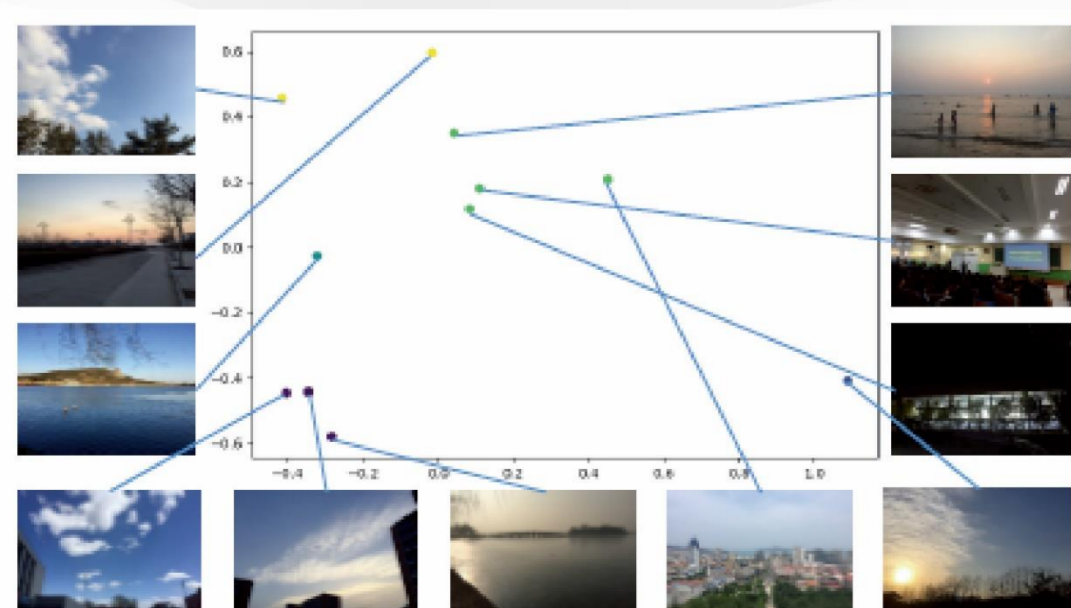
对VPN和P-Module进行融合

我们通过计算置信度 (Confidence Value) 的方式, 动态的调整二者的权重。

$$S_i = (1 - c_i)S_i^{VPN} + c_iS_i^P$$

置信度计算方法:

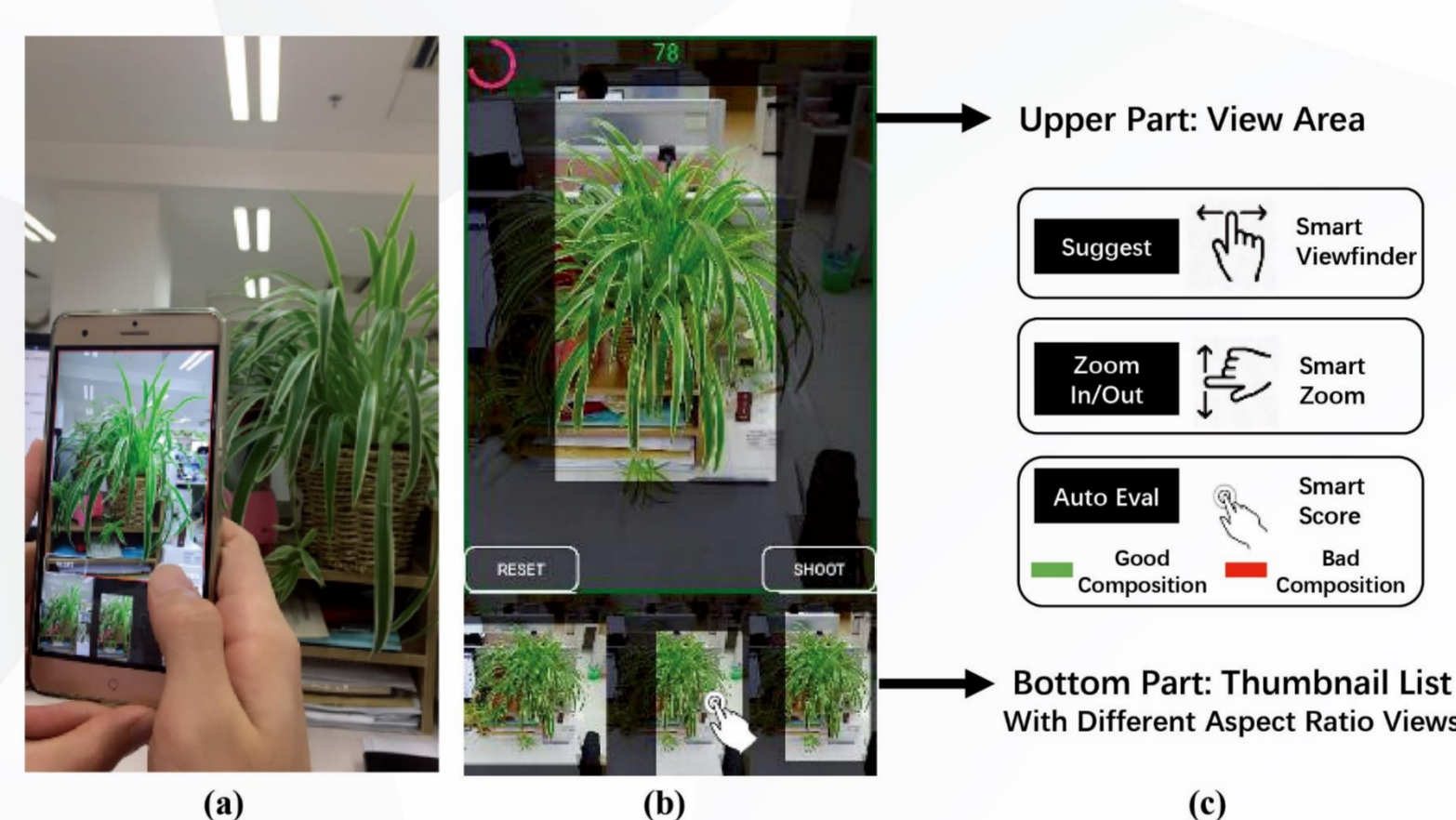
我们基于一个假设: 如果一个相似的图片已经作为知识 (训练样本) 被P-Module学习过了, 我们有理由相信最终的得分应该更依赖于P-Module的打分。因此, 我们通过计算当前照片与已处理过的照片的构图距离来计算置信度。当前照片与已有照片越相似, 置信度越高, P-Module的打分在最终打分中占的比重就越大。



$$c_i = \frac{1}{(1 + \min_j d(\phi(x_i), \phi(x_j)))^\beta}$$

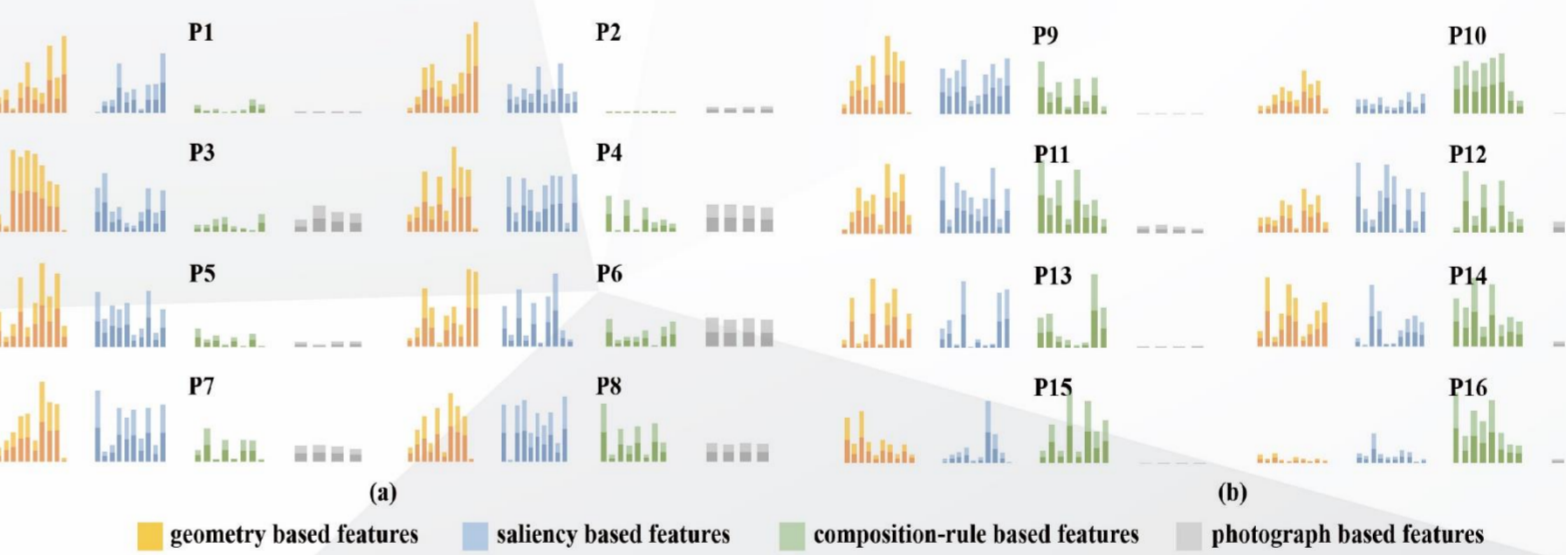
构图距离计算方法: 我们从设计的特征空间中选取23维有效的表征图片构图的特征, 通过计算欧式距离的方法, 来计算照片之间的相似度。

交互式界面设计



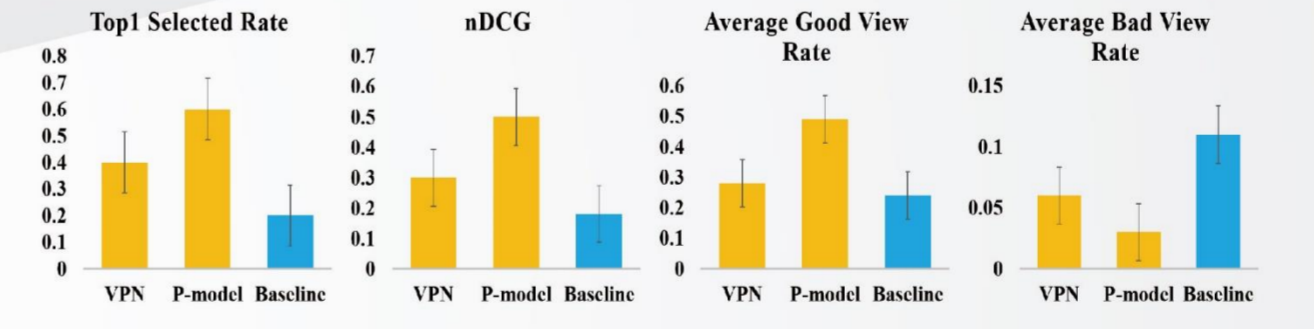
我们设计了一个完整的系统, 称为SmartEye。在前端, 我们设计了新颖的交互式界面, 包含了一些强大的功能, 比如SmartViewfinder, 能够实时的为用户推荐得分TopN的构图; SmartZoom, 智能地对镜头进行缩放; SmartScore, 智能地对镜头进行实时打分……此外还支持几种手势操作, 方便用户调用不同的功能。

特征相关分析 & 方法评估

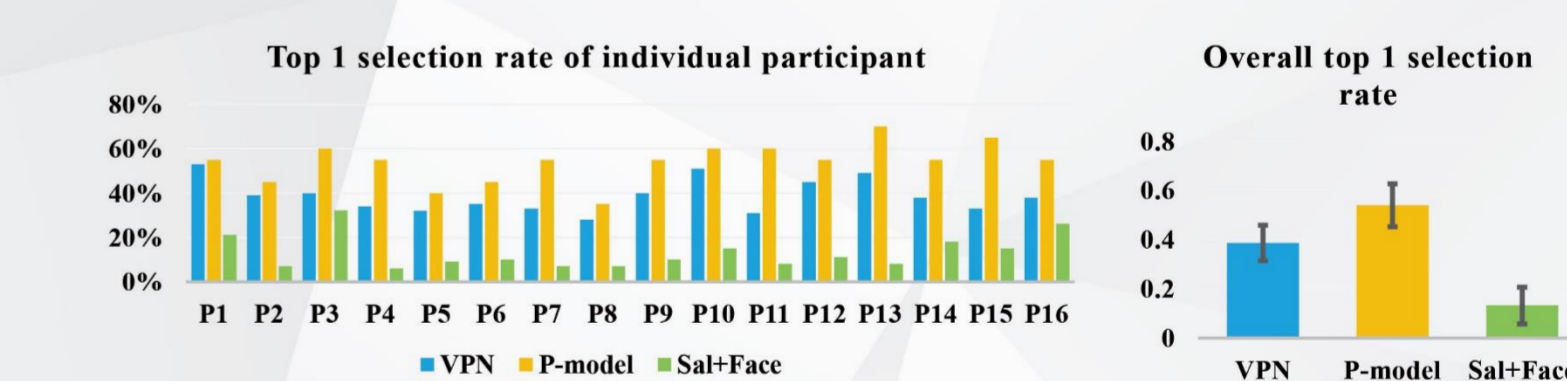


我们邀请了16位被试进行用户实验, 通过对4类32维特征计算相关性 (Pearson and Spearman), 得到如下结论:
① 在构图任务中, 不同特征的重要性有所不同
② 不同用户对构图有着不同的侧重和偏好, 这也印证了我们引入用户偏好学习的必要性

我们设计了一个评估指标: Top 1 selection rate, 即我们希望展示在最前面的构图推荐正是用户最喜欢的那一个, 所以我们计算首位构图被用户选择的比例来评估不同方法的优劣。



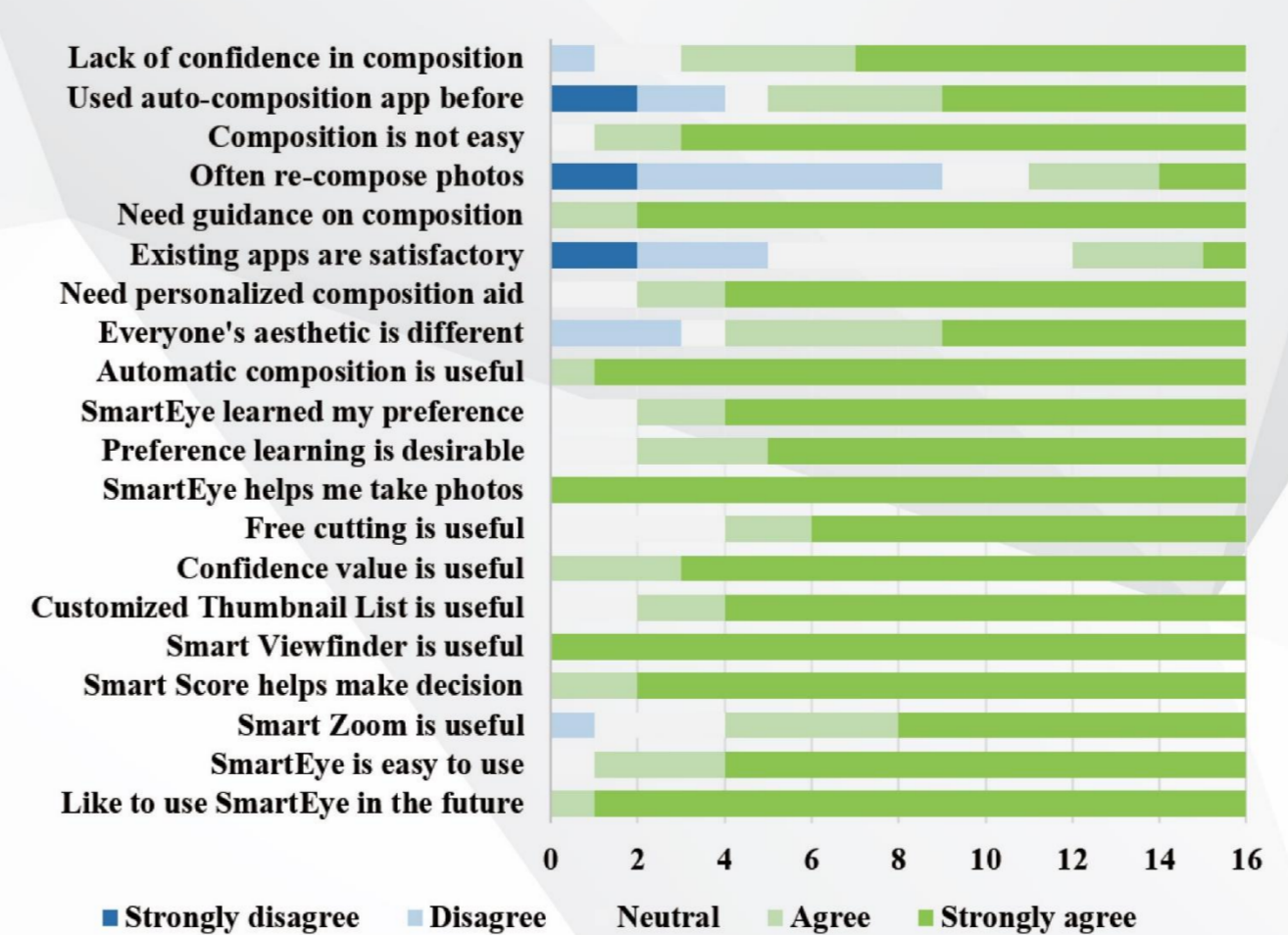
此外我们也从其他指标上对我们的模型进行评估, 均得到了理想的结果。



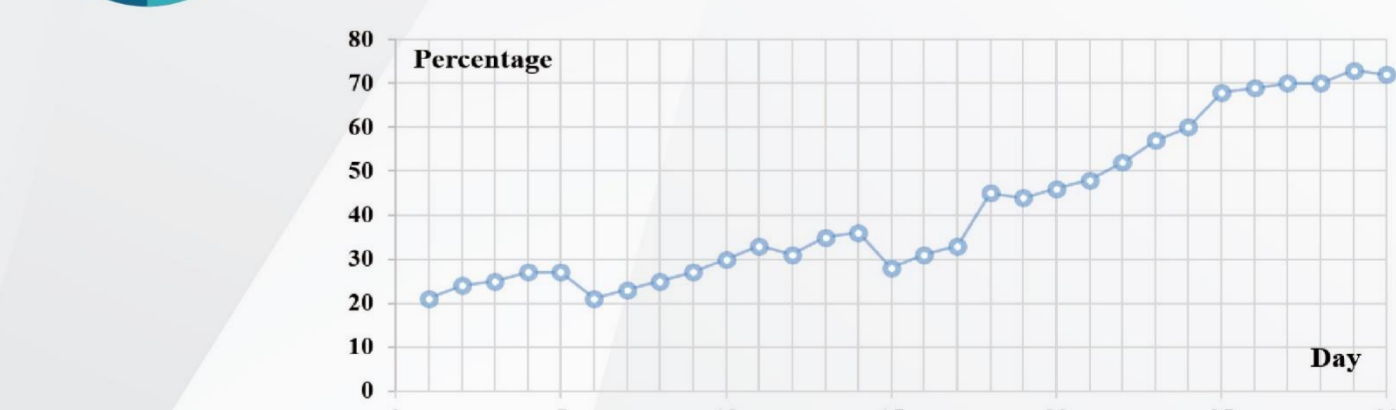
我们进行了对比实验, 分别对纯VPN, P-model (我们的方法), Sal+Face (作为Baseline) 进行比较, 可以发现P-model比另外两张方法都要好, 通过statistics test得出p<0.001, 统计学具有显著性。

主观问卷

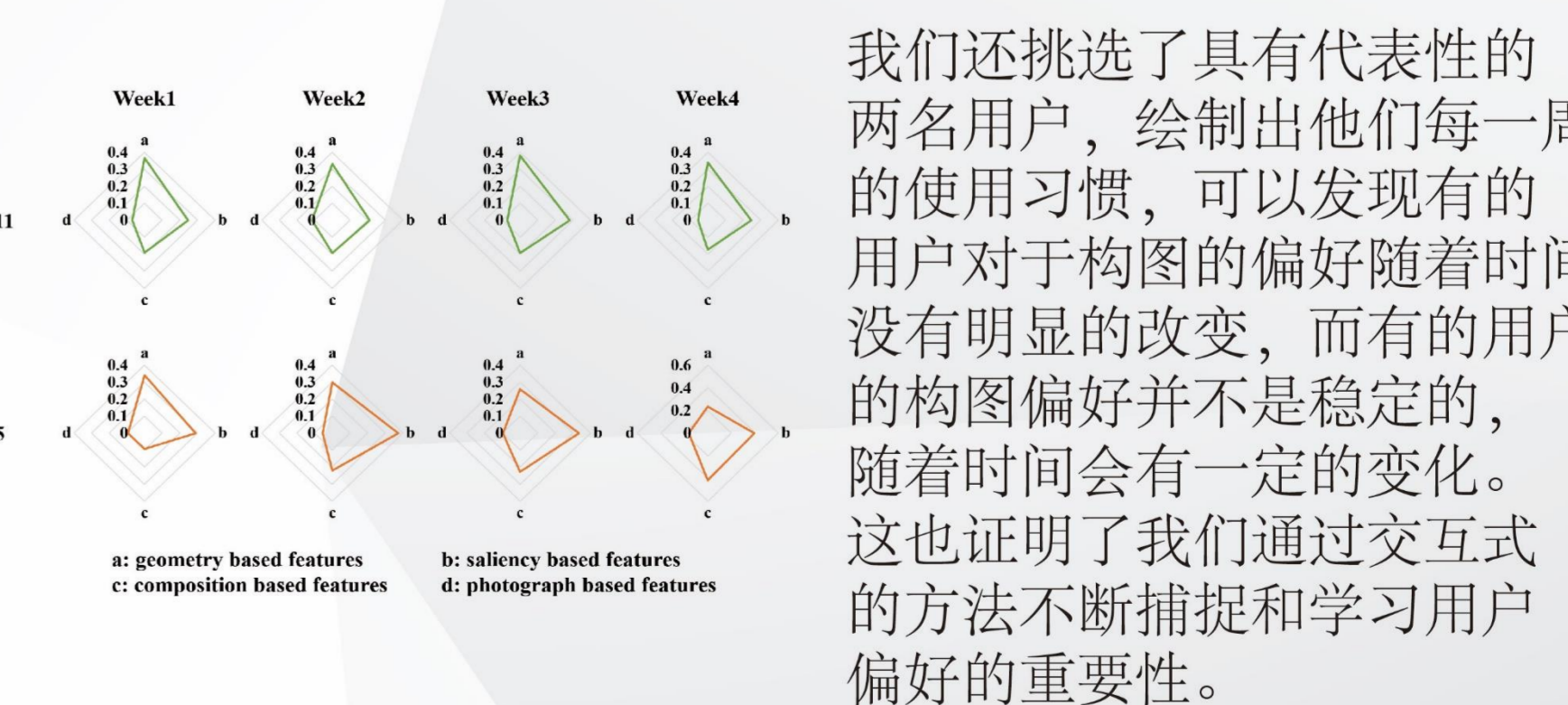
我们在用户实验最后进行了用户问卷调查, 可以发现用户对我们的系统满意度非常高, 而且认为我们的系统可以学习到他们的个人喜好。



其他结论



我们将P-model部署到移动端, 让用户进行了为期一个月的使用, 并且记录下每一天的top1 selection rate的数值变化, 可以发现我们的系统能够逐渐的学习到用户的偏好, 并且随着使用可以越来越好的对用户偏好进行建模。



我们还挑选了具有代表性的两名用户, 绘制出他们每一周的使用习惯, 可以发现有的用户对于构图的偏好随着时间没有明显的改变, 而有的用户的构图偏好并不是稳定的, 随着时间会有一定的变化。这也证明了我们的方法不断捕捉和学习用户偏好的重要性。