



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

文娱大脑的内容及用户理解浅谈 --Cognitive Framework在QOE里应用

阿里巴巴大文娱 蔡龙军（牧己）
2019.8.23-24

出品: LiveVideoStack
——音视频技术社区——

CSDN



深圳
2019

遨游“视”界 做你所想
Explore World, Do What You Want

LiveVideoStackCon 2019 深圳

2019.12.13-14



出品: **LiveVideoStack**
—— 音视频技术社区 ——

成为讲师: speaker@livevideostack.com

成为志愿者: volunteer@livevideostack.com

赞助、商务合作: kathy@livevideostack.com



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

Agenda

- 内容质量(QOE)
 - 内容越来越多&圈层分众分化是大势所趋
 - 复杂&没有完整衡量理论基础
- 内容认知框架（Cognitive Framework for QOE）
 - 基础认知进展介绍
- 内容认知框架应用及QOE优化
 - 基于流量结果的内容受众预测
 - 基于行为动作的内容质量挖掘
 - 基于用户情绪的内容质量检测



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

这个夏天的一半属于你们

年度剧集 NO1



经典款



男性高知群体



年轻款



年度综艺 NO1





北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

内容的趋势思考：多，从商业驱动到消费驱动

人类需求

生存

生活

生孩子

当前社会

物质前所未
有丰富

信息前所未
有繁多

非人智能前
所未有强大

社会变化

胖子越来越
多

商业越来越
复杂

AI时代开启



内容蓬勃发展，过剩生产，分级消费



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

分层、分级消费：90后30了，物联网/AI时代播放技术更多样

从全民爆款

这世界很酷

圈层爆款



人人喜欢 -> 各有所爱





北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

困难及思考

内容特点

商业需要

① 内容具有动态的不确定性和复杂性

低风险，确定性

② 平台内容吞吐量

啥是爆款

VS

③ 长视频有“延迟满足”特性

谁喜欢啥，咋办 (5WH)

④ 长视频属于信息非完备问题

.....

流行

社会

文化

人性

基因





北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

内容的复杂性决定了量化的困难

长安的一些数据



从前在死而作战，却不是为了赢，
久了，杀戮点死伤无数，便不知以命取功，意义何在。
所谓保家卫国，只觉一纸空文，直到在长安见到他们。
长安好吃，好喝，好生活，当然让人留恋下，可让我觉得值得留恋的，就是在长安城里，再普通不过的这些人。
大家没读过什么书，没有多少钱，也说不上有什么了不得的前途，可他们，不，可我们，都在尽心做自己的事。



故事：延迟满足&信息不完备

技术：NLP/CV/语音的语义理解 & KG

非群演800-1000人

群演300-1500人

筹备7个月，拍摄217天

涉及工种极多的复杂系统工程

技术：不确定性问题的衡量&计算



专业技能 VS 流量商业价值

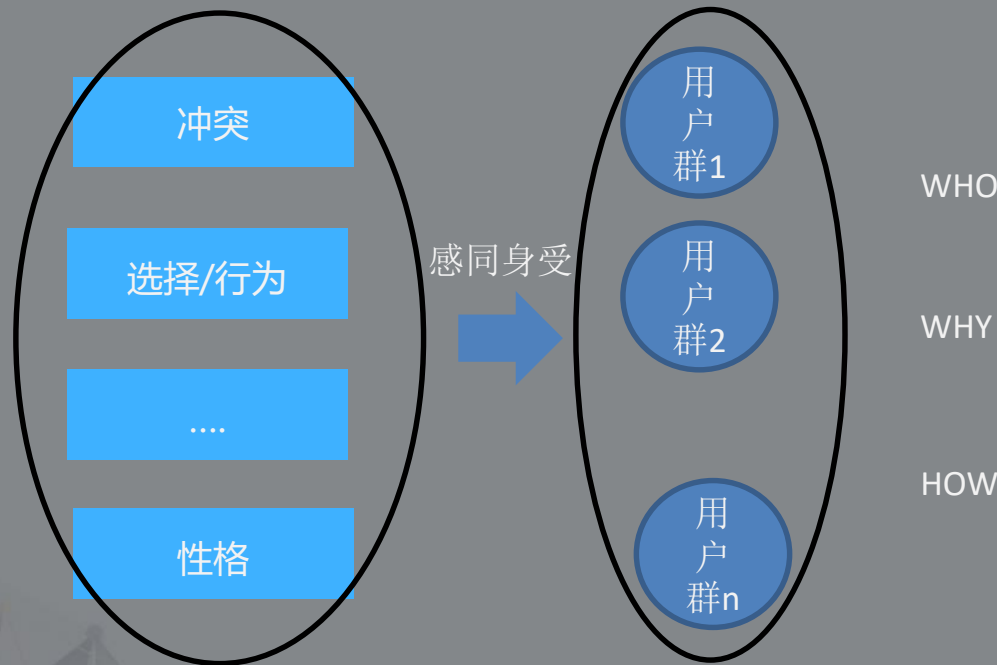
技术：用户理解+心理学



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

高级感：这个时代，大家都很难



00:04:34	甄嬛传的宁嫔啊
00:04:34	檀琪你敢对你老公这么横!

研究完《长安十二时辰》的望楼系统,我魔怔了_36氪

00:04:34	婢女是甄嬛传里养虎那个么
----------	--------------

易烊千玺的发簪为啥是这个方向

唐朝的大数据平台 - 大案牍术

粉丝
。
。
。

悬疑推理爱好者

历史军事爱好者

电视剧爱好者

内容创作

多种多样用户

?

平台

文娱大脑：北斗星，QOE是基础



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

采

投

制

宣

发

放



$P \sim V$

图像语音



剧本开发

内容

明星选
角

内容制
作

剧集/电影

电商

音乐

体育

表达

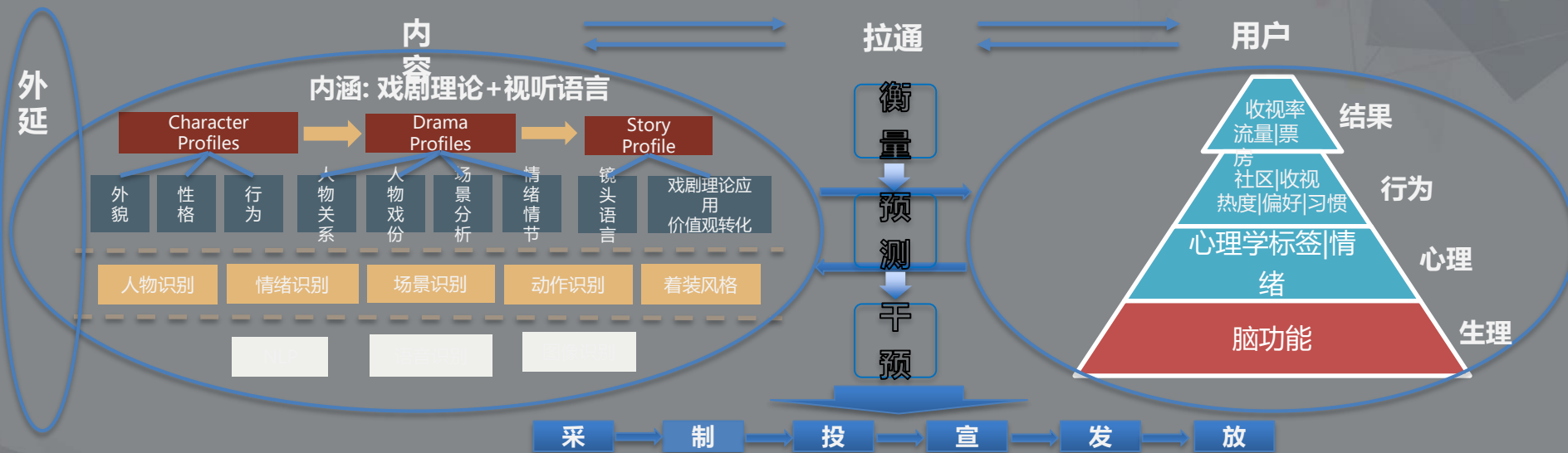
用户



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

内容认知框架(Cognitive Framework)





北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

基础研究

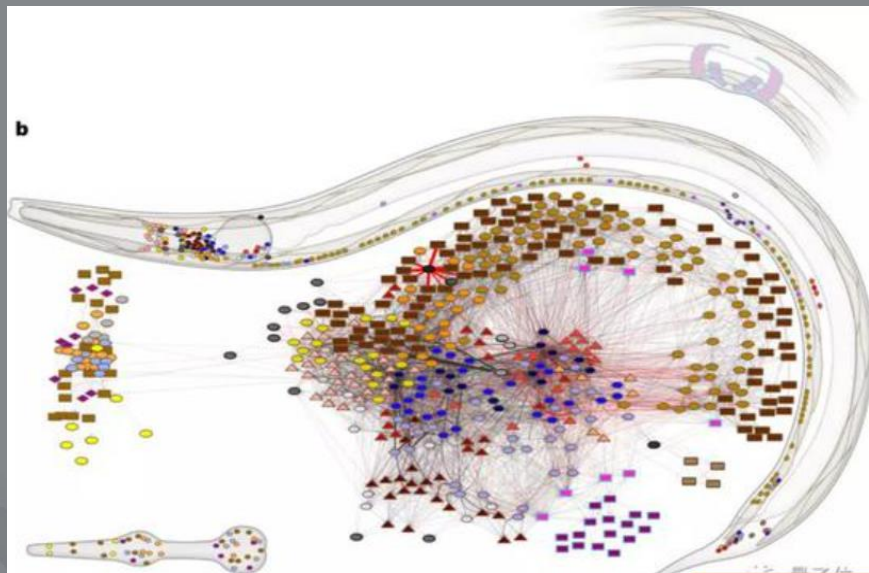


北京
2019

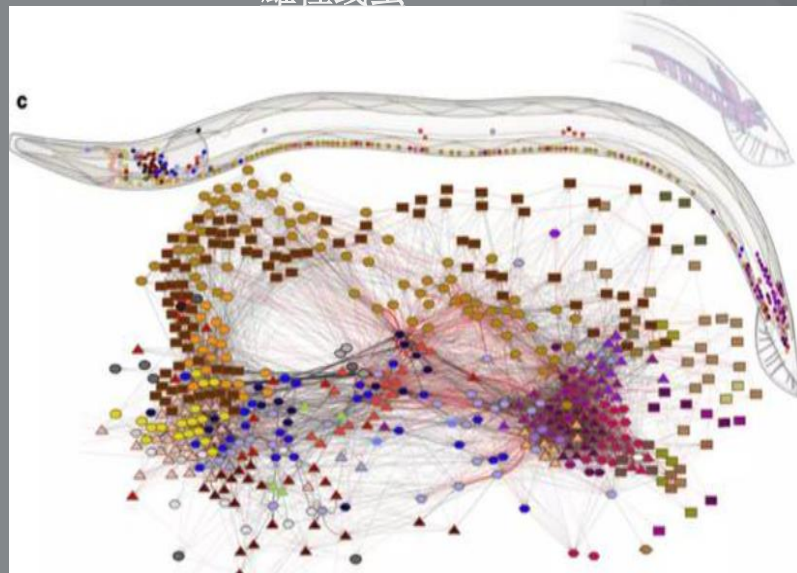
遨游“视”界 做你所想
Explore World, Do What You Want

雌雄脑回路不一样

雌性线虫



雄性线虫



1000个神经元，7000个连接，人有百亿计！

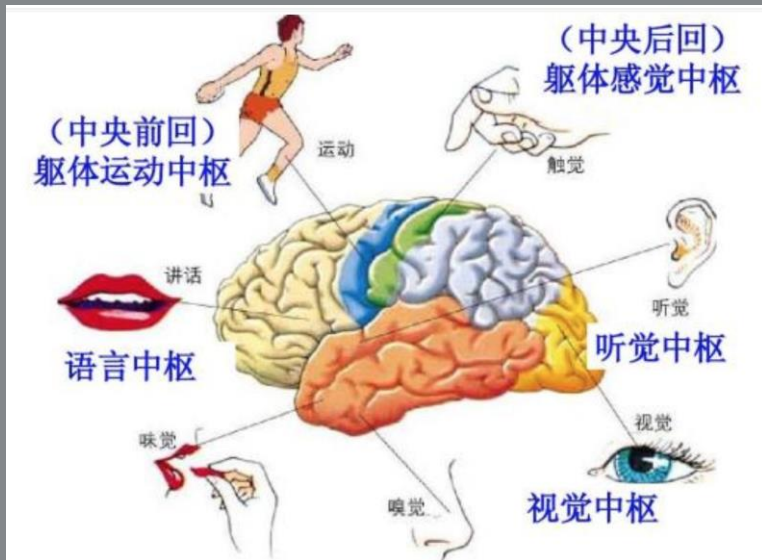
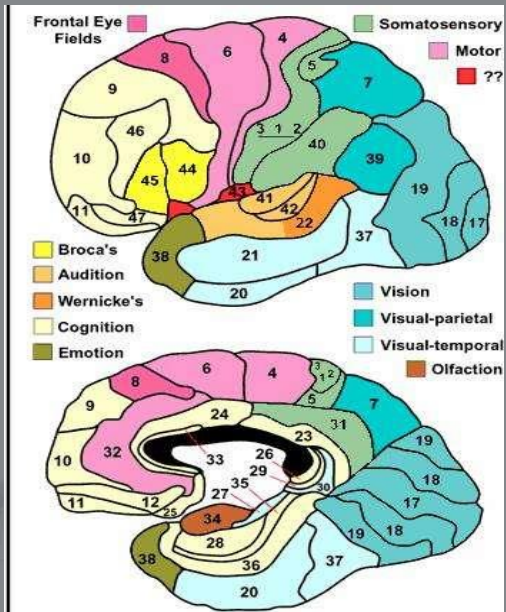
• 大数据下心理学的基础研究



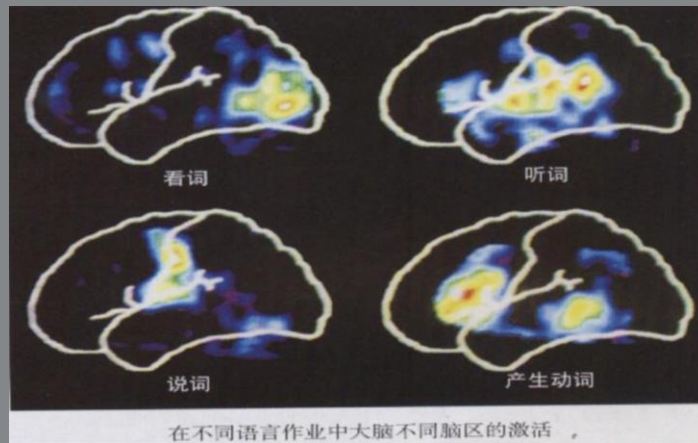
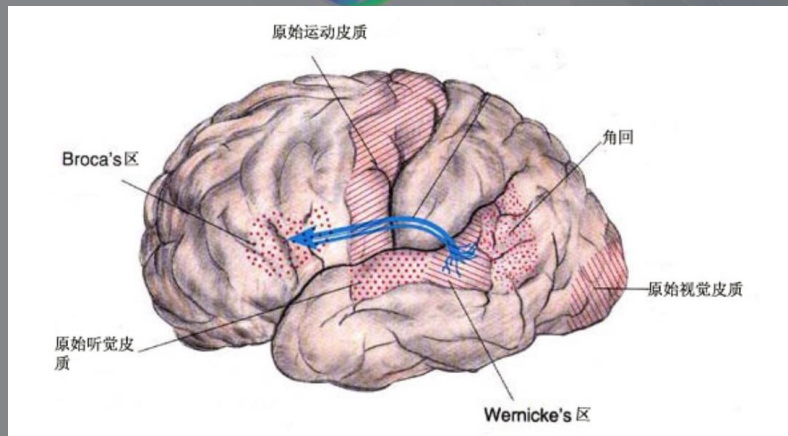
北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

- 布鲁德曼分区（Brodmann）分区系统
— 52个分区



- 语言机制
- 布洛卡区
 - 形成语言
 - 损伤导致“表达性失语症”
- 韦尼克区
 - 分辨语音，形成语义
 - 损伤导致“接受性失语症”
- 角回
 - 视觉和听觉通道、视觉记忆
 - 损伤导致“失读症”



典型测量方法

- 脑损伤研究法
 - 传统方法
 - 有损伤



➤ 脑电图 (EEG) / 脑电皮层图 (EcoG)

- ✓ 高时间分辨率
- ✓ 低空间分辨率



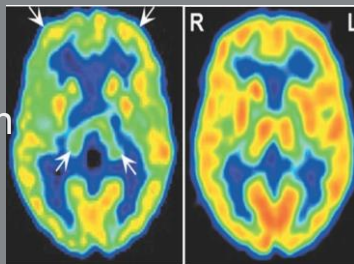
➤ 脑磁图 (MEG)

- ✓ 高时间分辨率
- ✓ 低空间分辨率



➤ 正电子发射断层扫描 (PET)

- ✓ 低时间分辨率
- ✓ 高空间分辨率cm





北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

• fMRI: 何为“爆”款

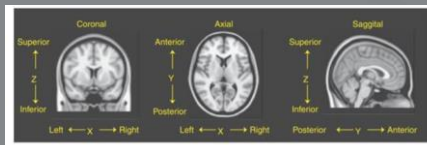
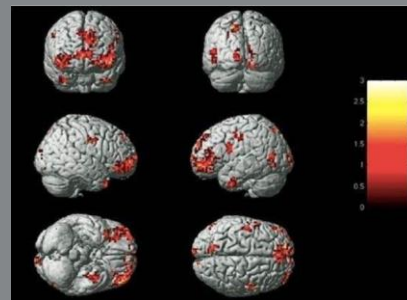
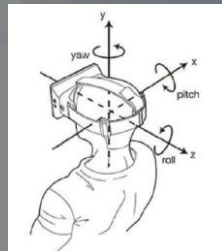
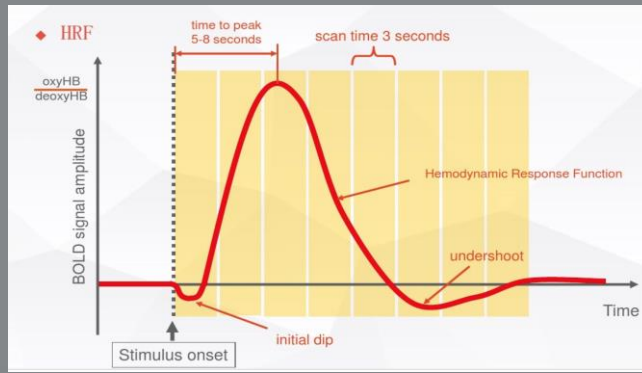
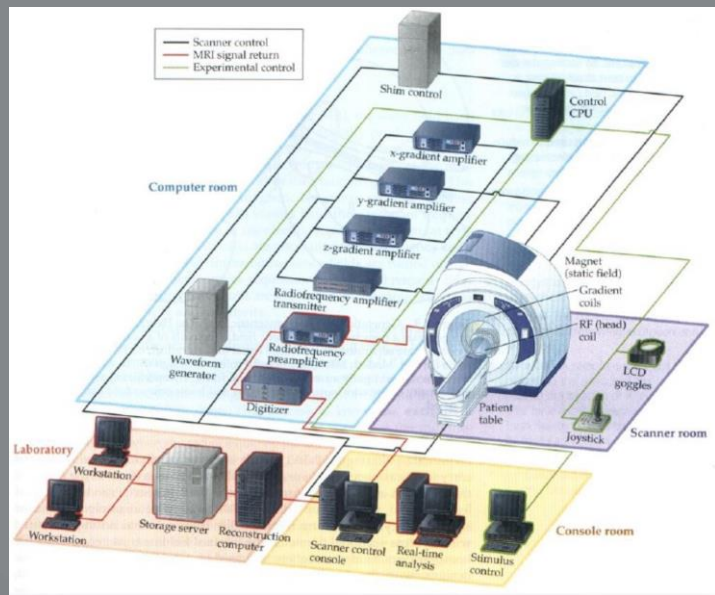


Figure 2.2. A depiction of the three main axes used in the standard coordinate space for MRI; images taken directly from an MRI scanner may have different axis orientations.



功能项: Hb血
红
BOLD血氧水平
依赖度
结构像: H原子
核



定位精度:
1mm
时间分辨率:
1s

情感相关反应区



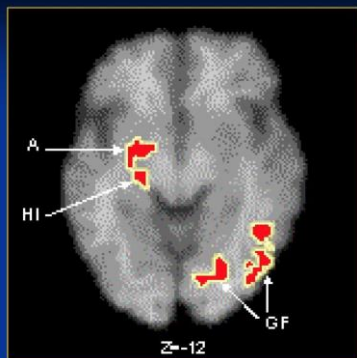
北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

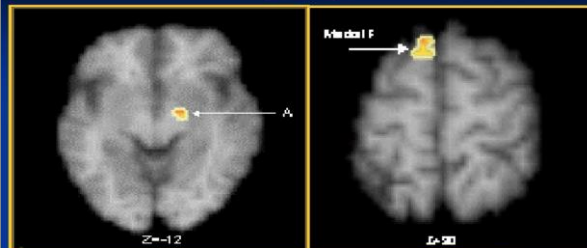
恐怖刺激

边缘系统的杏仁核
和海马激活

枕叶激活可能与刺激以视觉形式传入有关，也可能与感受恐惧时的视觉调整有关

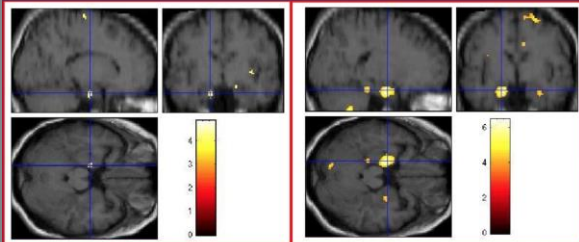


恐惧



杏仁核与额叶同时激活，说明产生恐惧的同时存在身体反应（如出汗）

音乐中的情绪反应



轻音乐-音阶，左侧杏仁核激活

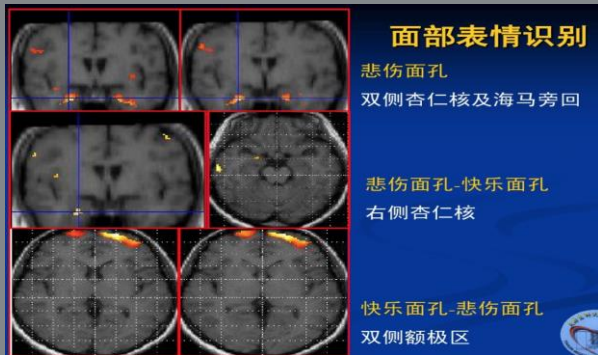
恐怖音乐-音阶，双侧杏仁核激活更明显

面部表情识别

悲伤面孔
双侧杏仁核及海马旁回

悲伤面孔-快乐面孔
右侧杏仁核

快乐面孔-悲伤面孔
双侧额极区





北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

视觉：从好不好看开始，世界杯球员颜值最高与颜值最低

排名1-5



排名710-715



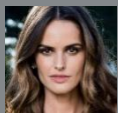


北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

热门球员（前）女友/太太 颜值排行

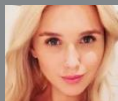
(1-10)



特拉普（德国）/ 伊莎贝尔-古
特拉（维密超模）：9



贝赫拉米（瑞士）/ 埃雷娜：9



罗伊斯（德国）/ 斯嘉丽：9



卡塞米罗/Anna Mariana Casemiro：8



奥乔亚（墨西哥）/ Karla Mora：8



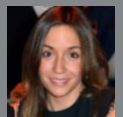
埃尔南德斯（墨西哥）/ 莱迪：8



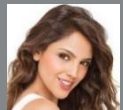
伊瓜因（阿根廷）/ 卢西亚娜：8



博格巴（法国）/ Chantel Jeffries，8



伊涅斯塔（西班牙）/ 安娜·奥尔蒂斯，8



c·罗纳尔多（葡萄牙）/ 乔治娜·罗德里格斯，8



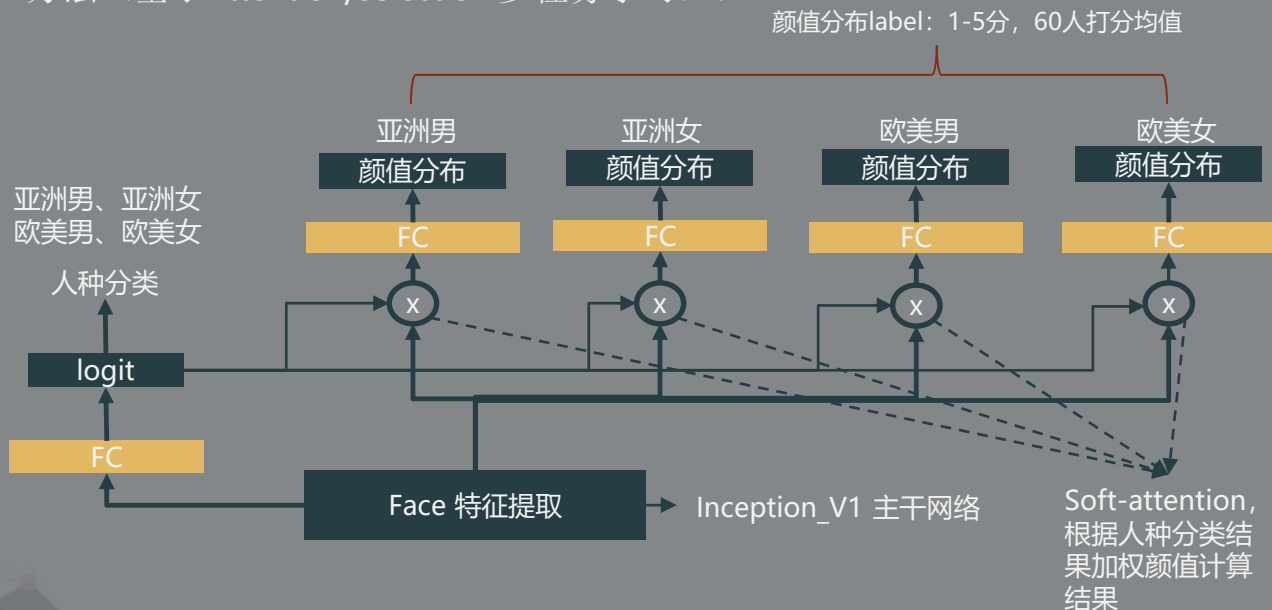
北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

明星基础视觉特征

根据相关人脸图片，识别人物性别、人种、颜值信

方法（基于Attention/selection 多任务学习）：





北京
2019

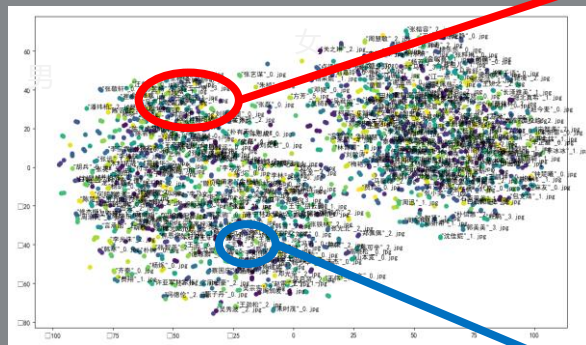
遨游“视”界 做你所想
Explore World, Do What You Want

明星基础视觉特

撞脸与相似明星选角



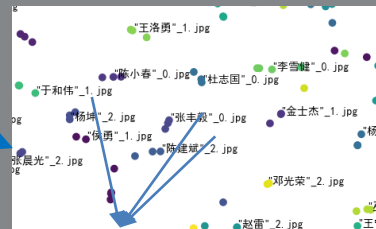
明星长相分布 (t-sne)



帅哥鲜肉



中年大叔



均饰演过曹



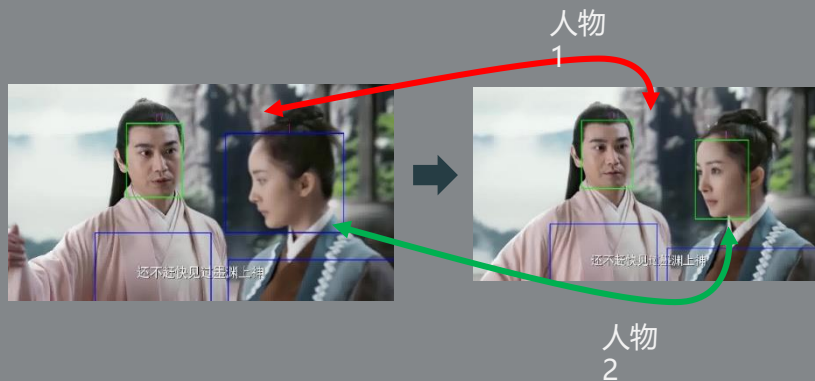
北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

明星视频曝光分析框架

当前数值测试结果（人工标定测试）

	Tracklet识别	单帧图片识别
mAPprecision	93%	88.3%
mARecall	88%	67%



- 1.openPose/DeepLabV3修改版可以在侧脸等诸多情况下检测到人物
2. KM匹配算法，可将侧脸人物通过下一帧信息可识别



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

视频剧情表达与人物感受

理解视频内容

— <人物>, 穿着<服装风格>, 在<场景>, <感情> 地<行为>

e.g. <孙红雷>穿着<嘻哈的衣服>在<广场>上<快乐>地<跳舞>

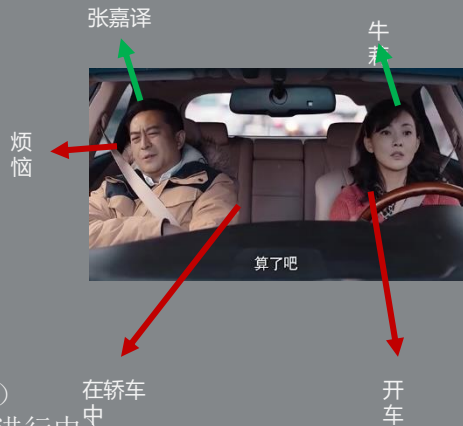
— 应用场景

- 选角：人物特定风格检索，表演风格评估
- 人设分析

— 方法

- 自底而上

- 人物识别：采用人物曝光分析框架（进行中）
- 服装风格：human parsing + 服装语义空间（进行中）
- 场景识别：Places365 场景识别（进行中）
- 情感识别：表情识别（进行中）
- 动作识别：openPose，骨骼点检测(进行中)





北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

认知框架的应用及QOE优化

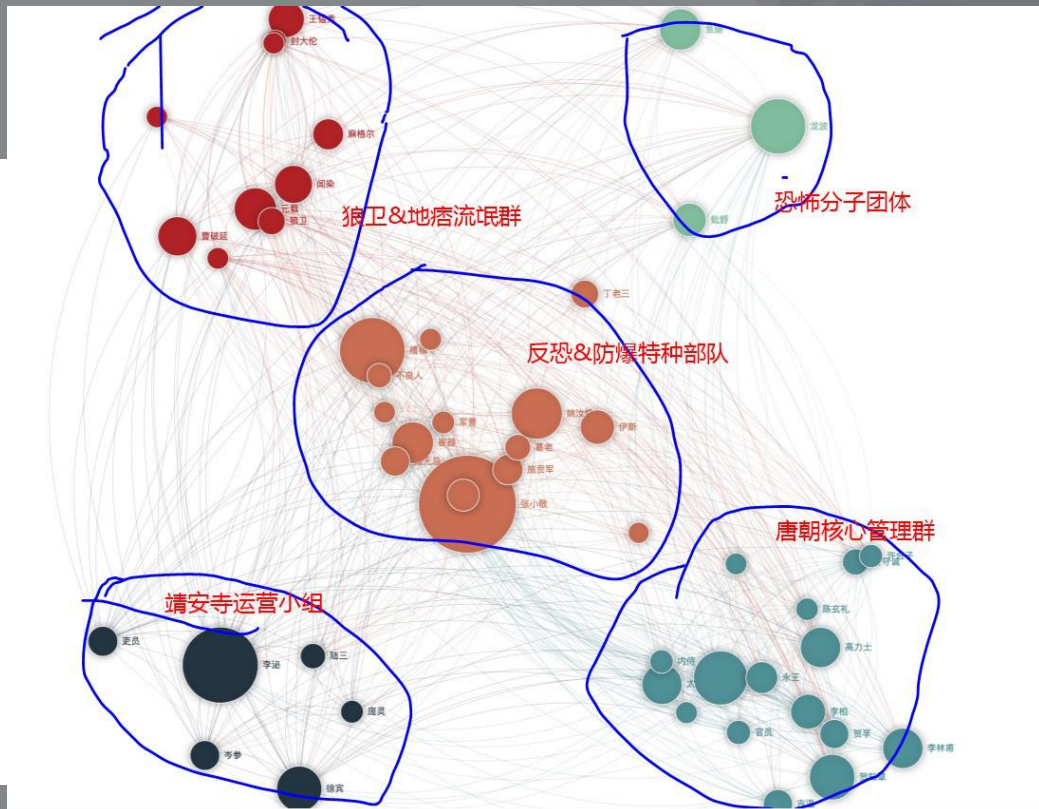
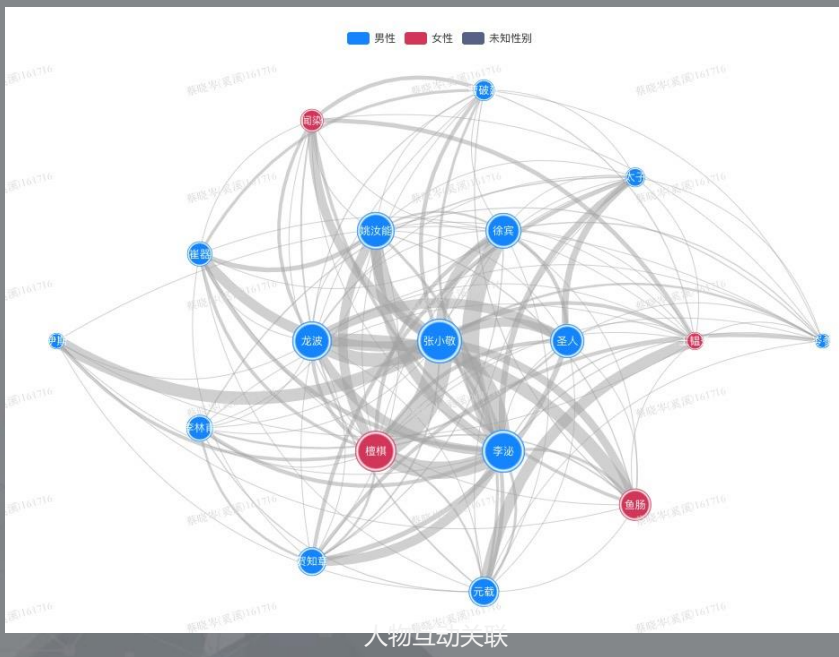
内容创作理解 - IP/剧本分析



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

- 人物互动关联 & 人物社团关系：
快速定位剧情人物关系设定



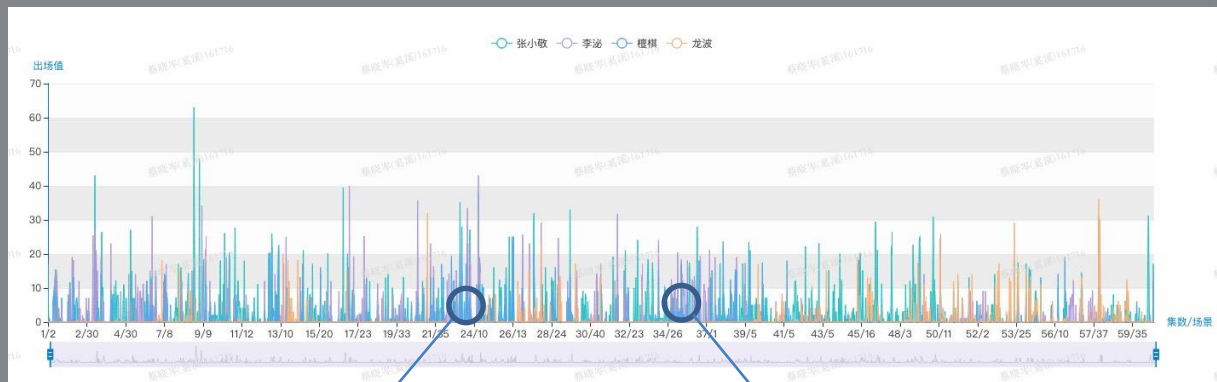
内容创作理解 – IP/剧本分析



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

- 人物出场分布&出镜率—快速定位角色场次、判断角色戏份
- 各场次 & 全局人物热词—判断各场次和全剧的核心线索，人物设定



张小敬	15%
李泌	10%
檀棋	5%
龙波	4%
姚汝能	3%
徐宾	3%
圣人	3%
鱼肠	3%
元载	2%
贺知章	2%
李林甫	2%
崔器	2%
闻染	2%
曹破延	2%
太子	2%
高力士	2%
王韞秀	2%
伊斯	2%
吉温	1%
岑参	1%



张小敬



李泌



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

《药神》用户情绪VA

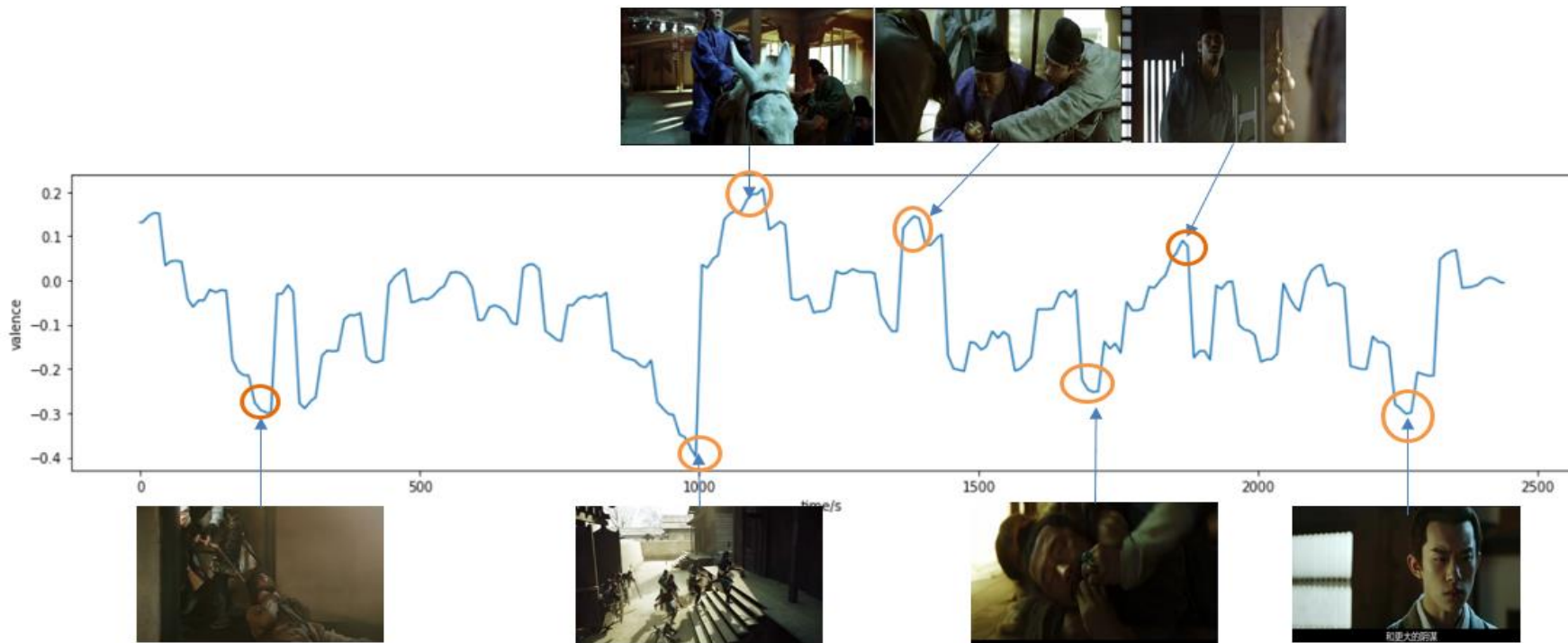


成片情绪挖掘



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want





北京
2019

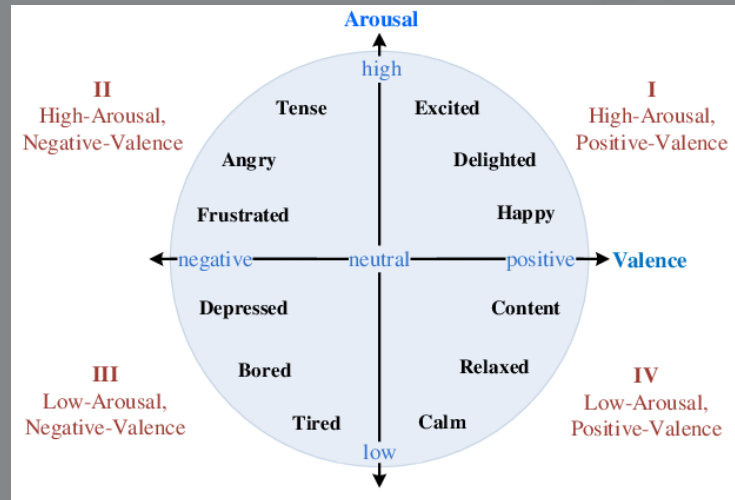
遨游“视”界 做你所想
Explore World, Do What You Want

基于VA的情感模型

业内通用的情感模型:

Valence: 情绪正负向。-1 到 +1 之间, -1 表示负向情感, 如悲伤, +1表示正向情感, 如高兴

Arousal: 情绪的强烈程度。-1表示情绪最不强烈, 如困乏平静, +1 表示轻度最强烈, 如激动兴奋。

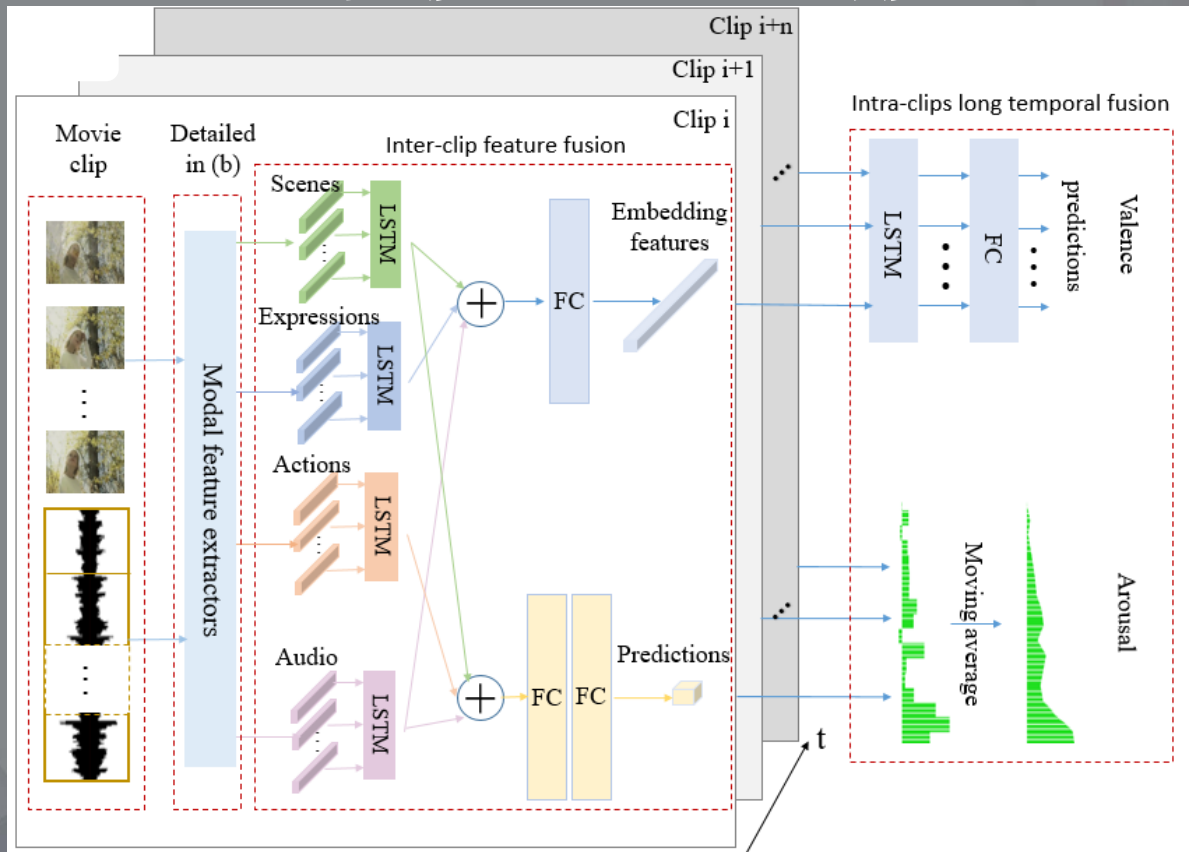




北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

多模态的VA识别模型





北京
2019

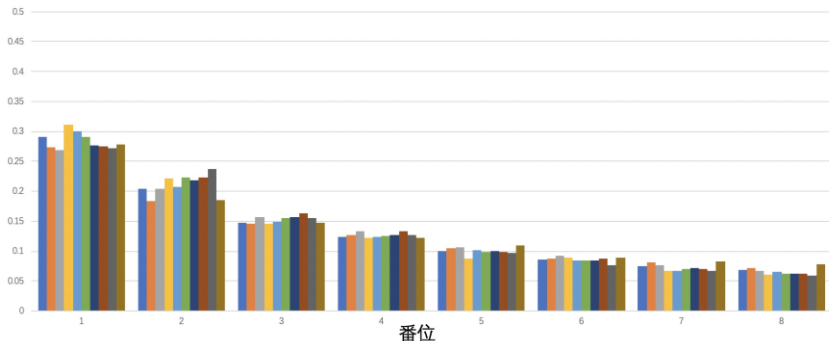
遨游“视”界 做你所想
Explore World, Do What You Want

内容创作理解 – 成片分析

演员出镜率benchmark & 出镜率打分—判断剧集成片角色出镜合理性

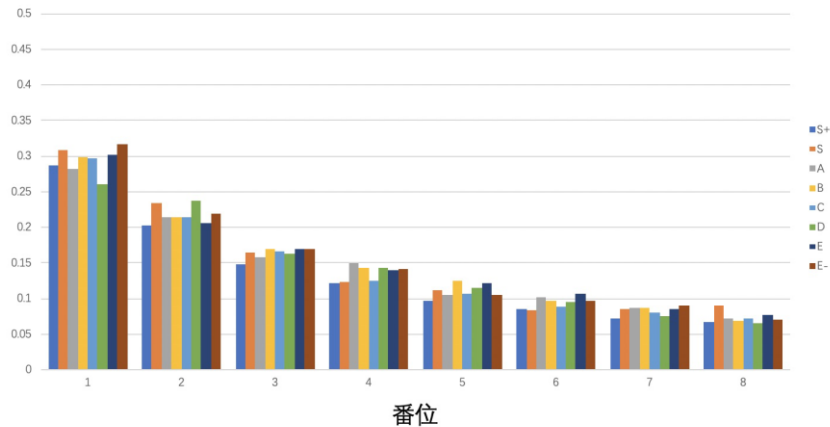
演员出镜率benchmark

电视剧出镜率benchmark



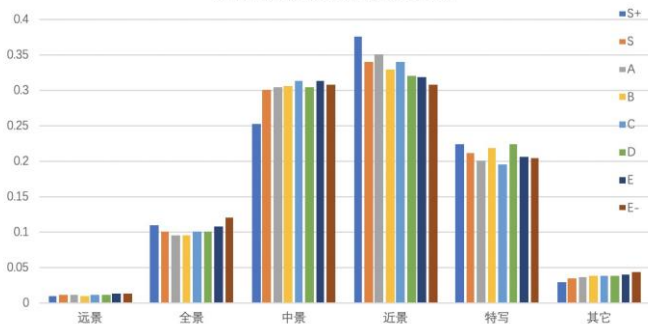
题材-出镜率benchmark

电视剧出镜率benchmark 按级别



播放级别-出镜率
benchmark

电视剧景别benchmark按级别



A bar chart showing the proportion of different camera shot types (S+, S, A, B, C) across three categories: Close-up (近景+特写), Medium/Long Shot (中远景), and Other (其他). The Y-axis represents the proportion, ranging from 0 to 0.8. The legend indicates the following colors for the shot types: S+ (Blue), S (Red), A (Green), B (Purple), and C (Cyan).

Category	S+	S	A	B	C
近景+特写	0.60	0.55	0.55	0.55	0.55
中远景	0.38	0.40	0.40	0.40	0.40
其他	0.02	0.02	0.02	0.02	0.02

Scenario	Proportion of 'Yes' responses
L S+	0.20
M S+	0.23
S S+	0.57
L S	0.16
M S	0.24
S S	0.60
L A	0.19
M A	0.23
S A	0.58
L B	0.19
M B	0.25
S B	0.56
L C	0.20
M C	0.25
S C	0.55
L D	0.17
M D	0.24
S D	0.59
L E	0.20
M E	0.24
S E	0.55
L E-	0.37
M E-	0.22
S E	0.40

Clinical Stage	Number of Cases
S+	10.5
S	9.5
A	9.0
B	8.5
C	8.0
D	11.5
E	8.5
E-	6.0

两个中长镜头之间短镜头使用的个数



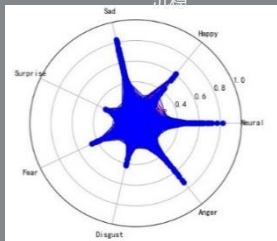
北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

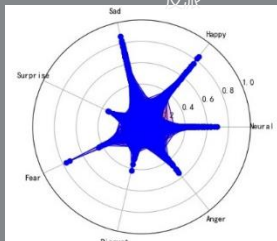
人物性格理解

反贪风暴 (2004)

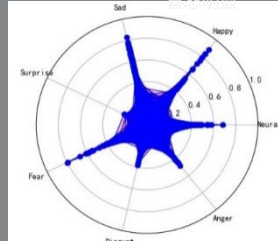
古天乐
沉稳



骆应钧
反派

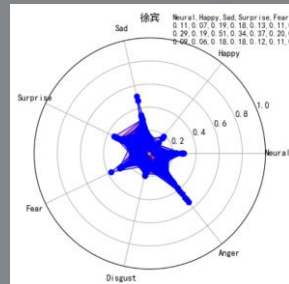


王敏德
警察

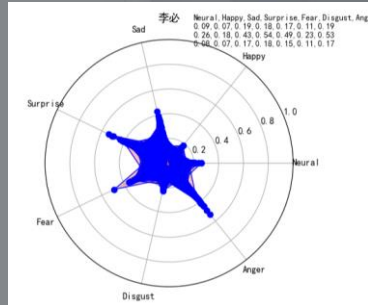


长安十二时辰 (2019)

徐宾

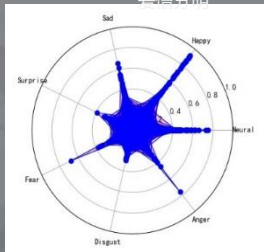


李必

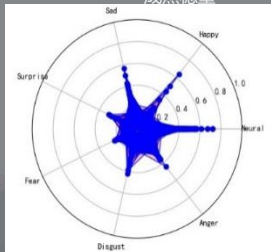


北京遇上西雅图 (2013)

汤唯
爱憎分明

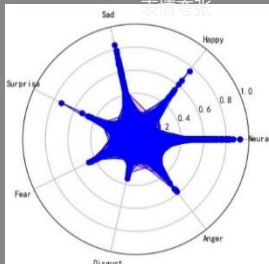


吴秀波
成熟稳重

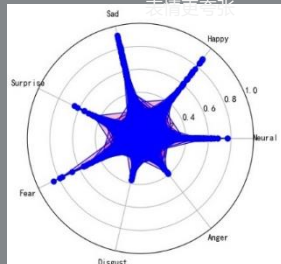


羞羞的铁拳 (2017)

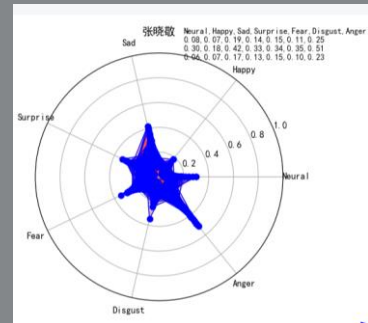
艾伦
表情夸张



马丽
表情更夸张



张晓敬



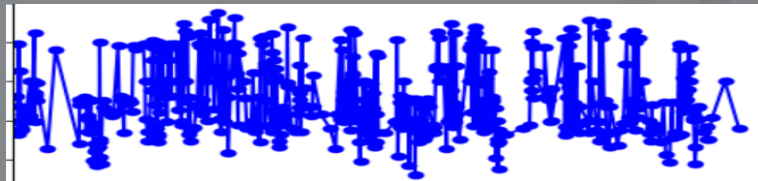
情绪曲线



北京
2019

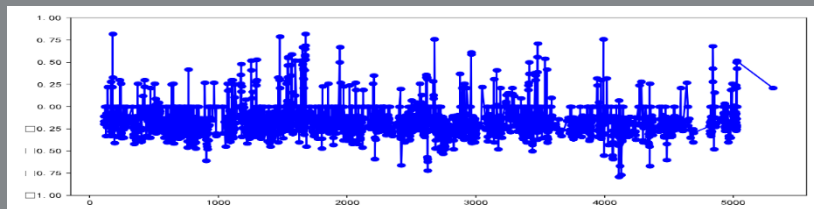
遨游“视”界 做你所想
Explore World, Do What You Want

长安十二时辰 (2019)



正面

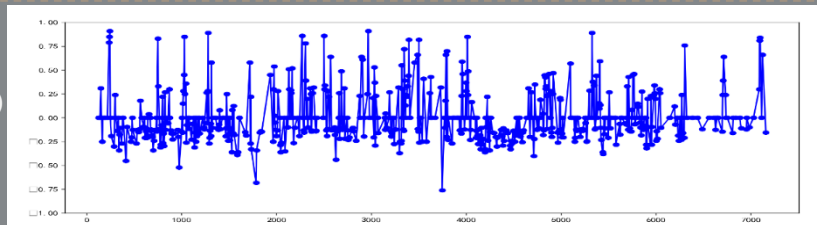
反贪风暴 (2004)



反面

正面

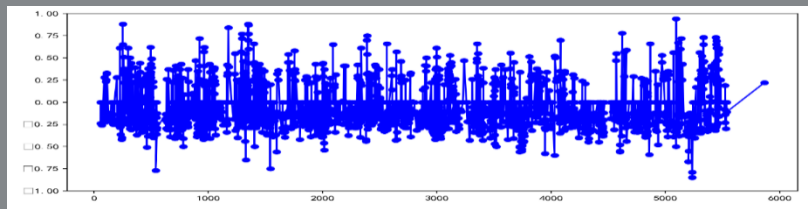
北京遇上西雅图 (2013)



正面

反面

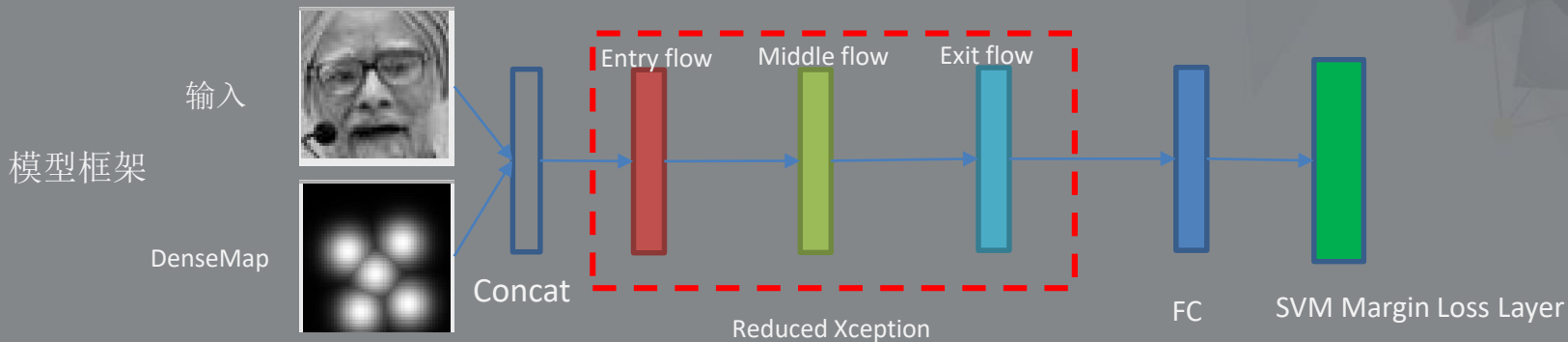
羞羞的铁拳 (2017)



正面

反面

• 改进模型 (Reduced Xception with Margin Loss)



➤ 输入：引入人脸关键点densemap

✓ 原理：精确判断人脸表情需重点关注五官如眼睛，鼻子，嘴的区域信息

✓ 关键点检测模型 (MTCNN)

✓ Densemap 计算

$$I(i, j) = \max_{k=1}^K e^{-\frac{(i-x_k)^2 + (j-y_k)^2}{2\sigma^2}}$$

$I(i, j)$ 为 densmap 在像素 i, j 处的强度， x_k, y_k 关键点 k 的坐标， $k \in \{1, 2, \dots, K\}$



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

情感识别：图片表情识别

- 模型提升 (Reduced Xception with Margin Loss)

- Reduced Xception*

- ✓ 使用可分离卷积 (deepwise 卷积 + pointwise 卷积)
- ✓ Entry flow, middle flow, exit flow 各缩减至2层卷积

- SVM Margin Loss Layer**

$$\min_w \frac{1}{2} w^T w + C \sum_{n=1}^N \max(1 - w^T x_n t_n, 0)$$



$$\begin{aligned} \min_{w, \xi_n} & \frac{1}{2} w^T w + C \sum_{n=1}^N \xi_n, \\ \text{s.t. } & w^T x_n t_n \geq 1 - \xi_n \quad \forall n \\ & \xi_n \geq 0 \quad \forall n \end{aligned}$$

$x_n \in R^D$ 表示上一层输出, $t_n \in \{-1, 1\}$ 为分类标签, w 表示本层参数

*Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions[J]. 2016:1800-1807

** Yichuan Tang, Deep Learning using Linear Support Vector Machines, ICML 2013

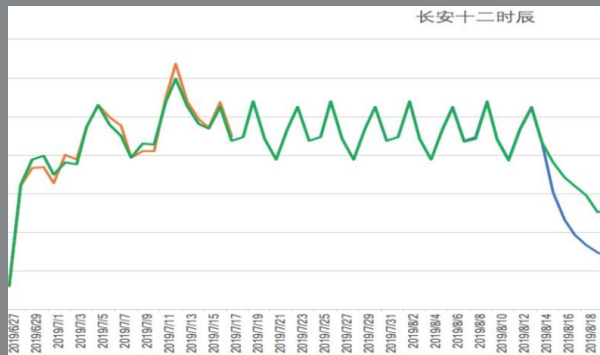


北京
2019

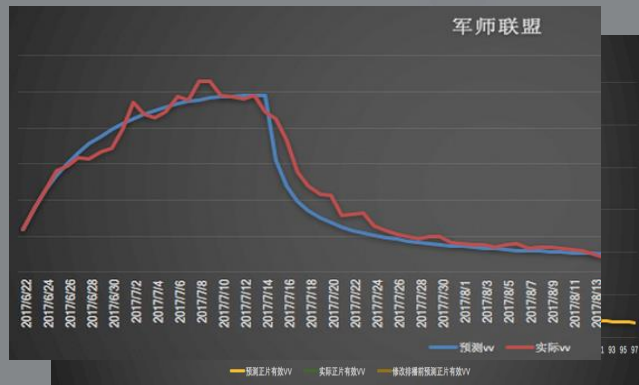
遨游“视”界 做你所想
Explore World, Do What You Want

采制阶段：预测能力建设

行业现状：不信、统计分析、简单模型



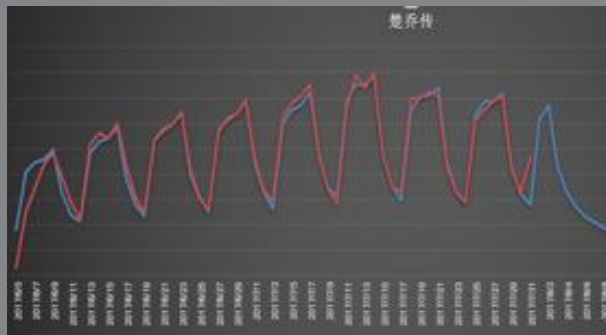
准确率92%



准确率90%+



准确率90%



准确率92%



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

总结

在强人工智能尚遥远的情形下，如何结合机器AI和人工经验是个永恒的主题，

- 如何更好的结合符号学派智能和链接学派的智能是非常重要的，决策引擎的建设，包括结合人工逻辑规则和可学习数据AI，不确定性分析框架和经久不衰的贝叶斯因果决策，以及神经元化的混合智能计算框架。
- 量化的心理学研究也越来越重要了，如何结合大数据应用价值非常大



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

Thank you



出品: LiveVideoStack CSDN
—— 音视频技术社区 ——