



南京大學
NANJING UNIVERSITY



iQIYI VID FACE 2019挑战赛

答辩汇报

一个名字团队

2019年6月20日

- 1 团队简介**
- 2 数据理解**
- 3 算法流程**
- 4 实验结果**
- 5 结论建议**

- 1 团队简介**
- 2 数据理解
- 3 算法流程
- 4 实验结果
- 5 结论建议

团队简介:

组织



董传奇

代码编写 多模态融合



顾峥

视频特征融合



黄中豪

模型集成



季雯

数据分析



- 1 团队简介
- 2 数据理解**
- 3 算法流程
- 4 实验结果
- 5 结论建议

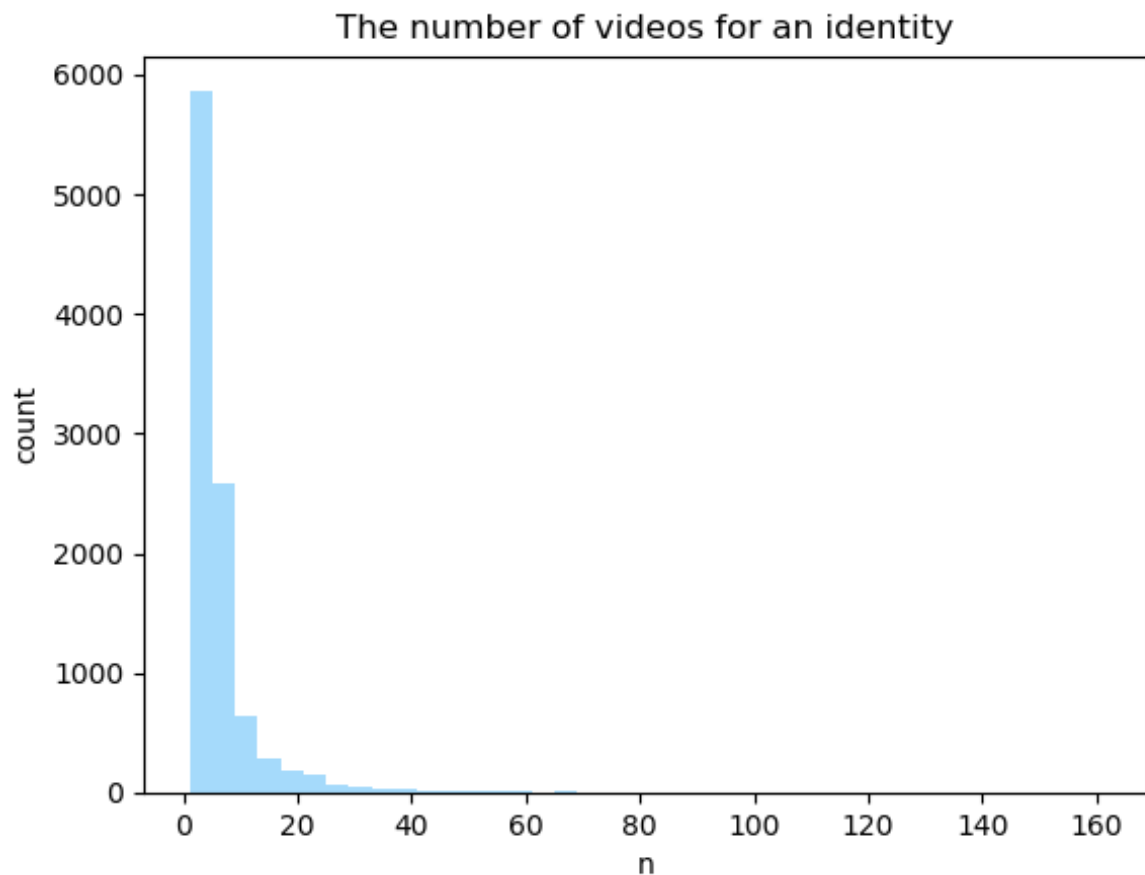
基本信息

- 共有10034人
- 平均每人37段视频
- 训练集共有60566段视频,
验证集共有76013段视频,
其中噪声30541条

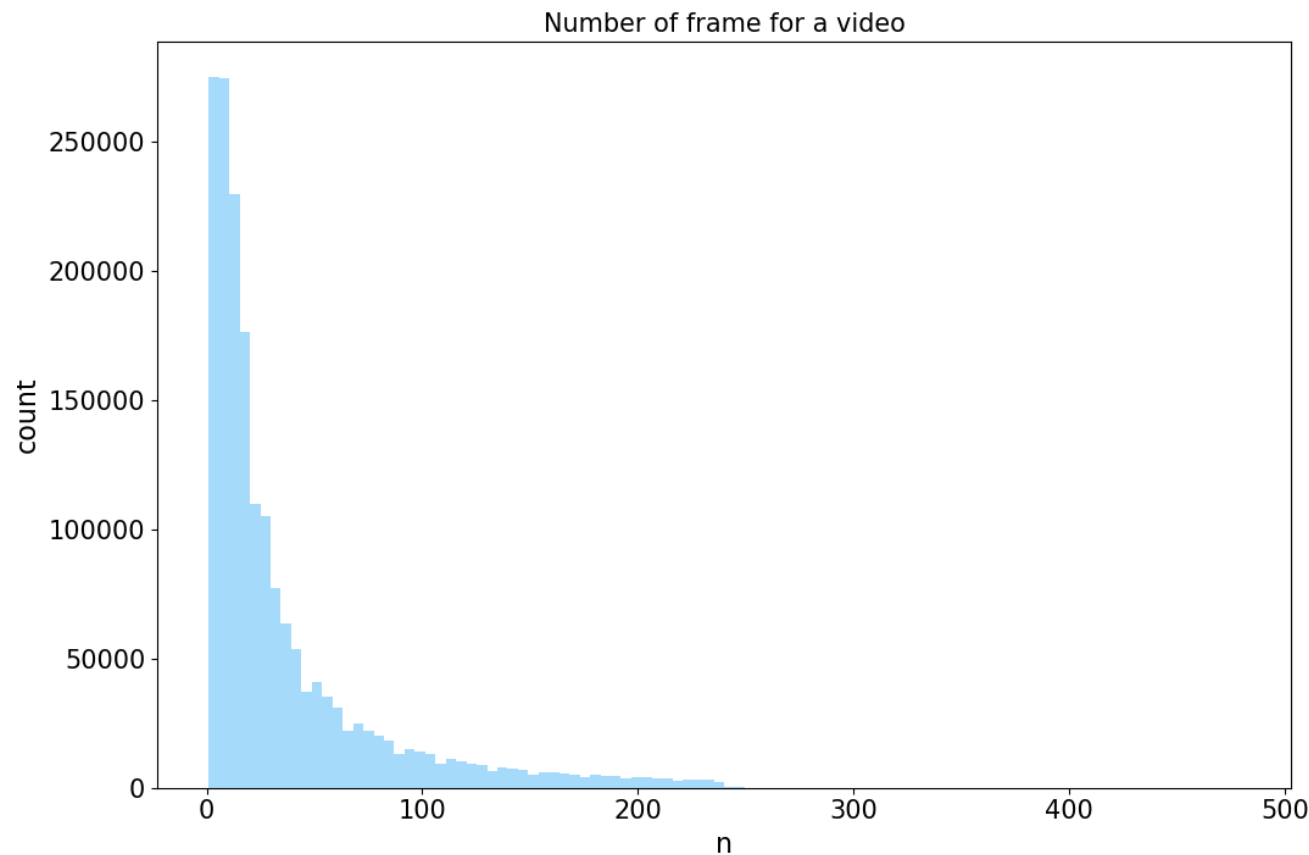
挑战

- 数据分布不平衡
- 质量分数差异大
- 视频内部有噪音

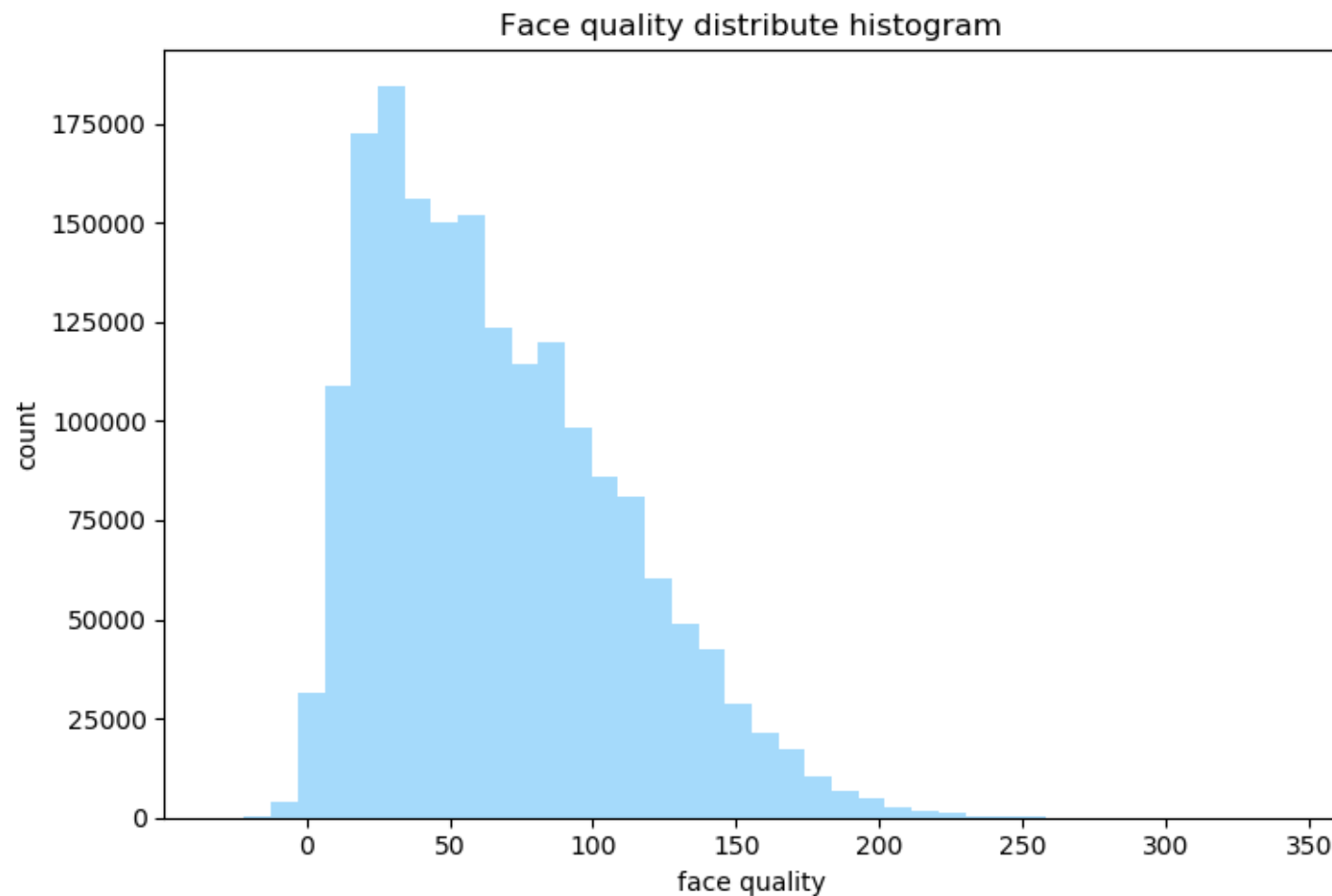
VIDEO/ID 变化多样



FRAME/VIDEO 变化多样



质量分数变化多样，且总体偏低



视频内部存在噪音



- 1 团队简介
- 2 数据理解
- 3 算法流程**
- 4 实验结果
- 5 结论建议

算法流程：概要设计

两类特征

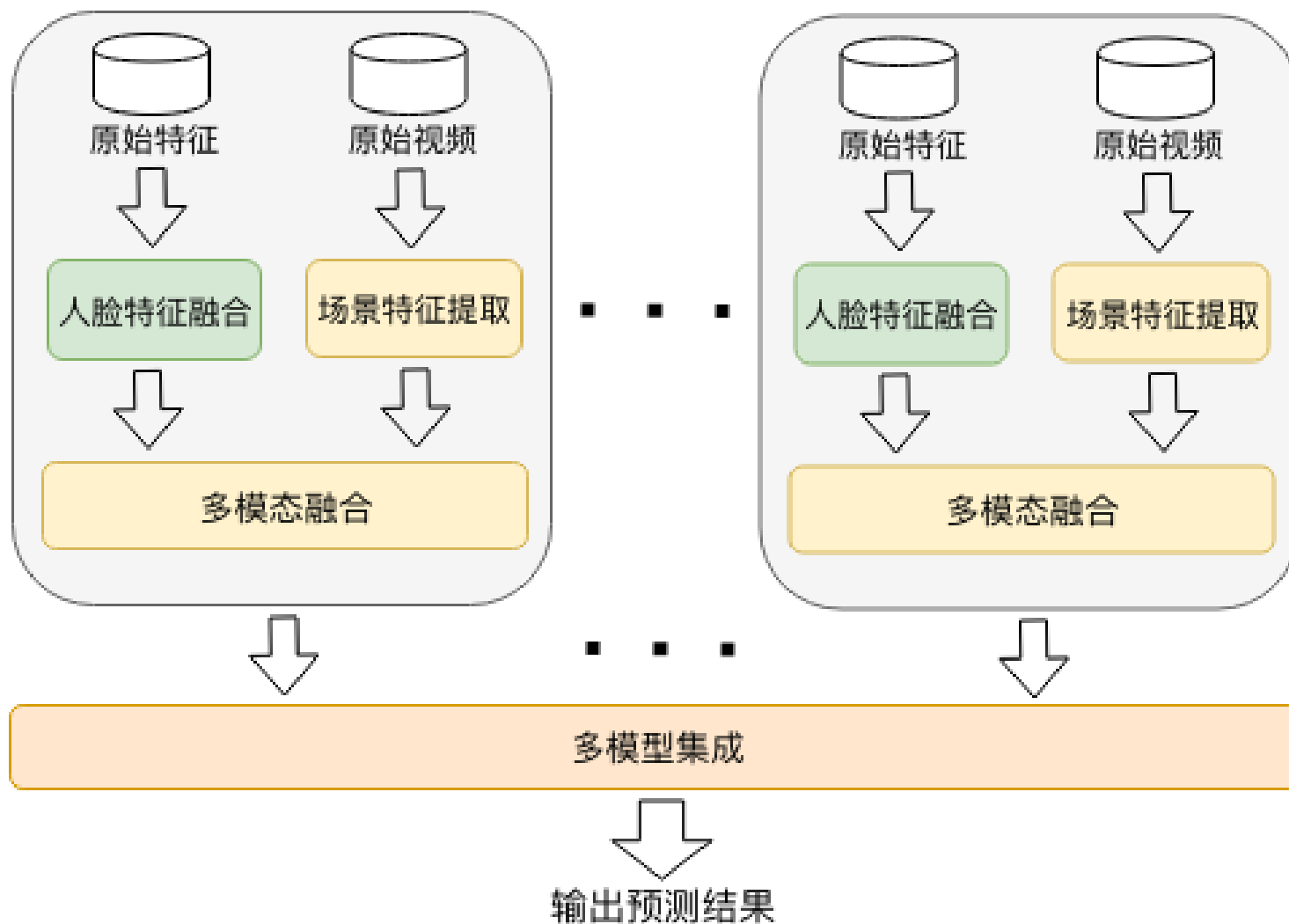
- 人脸特征
- 场景特征

三大模块

- 视频特征融合
- 多模态特征融合
- 模型集成

六项技术

- Dual Attention
- NAN
- ArcFace
- Focal Loss
- Multi Model
- Multi View



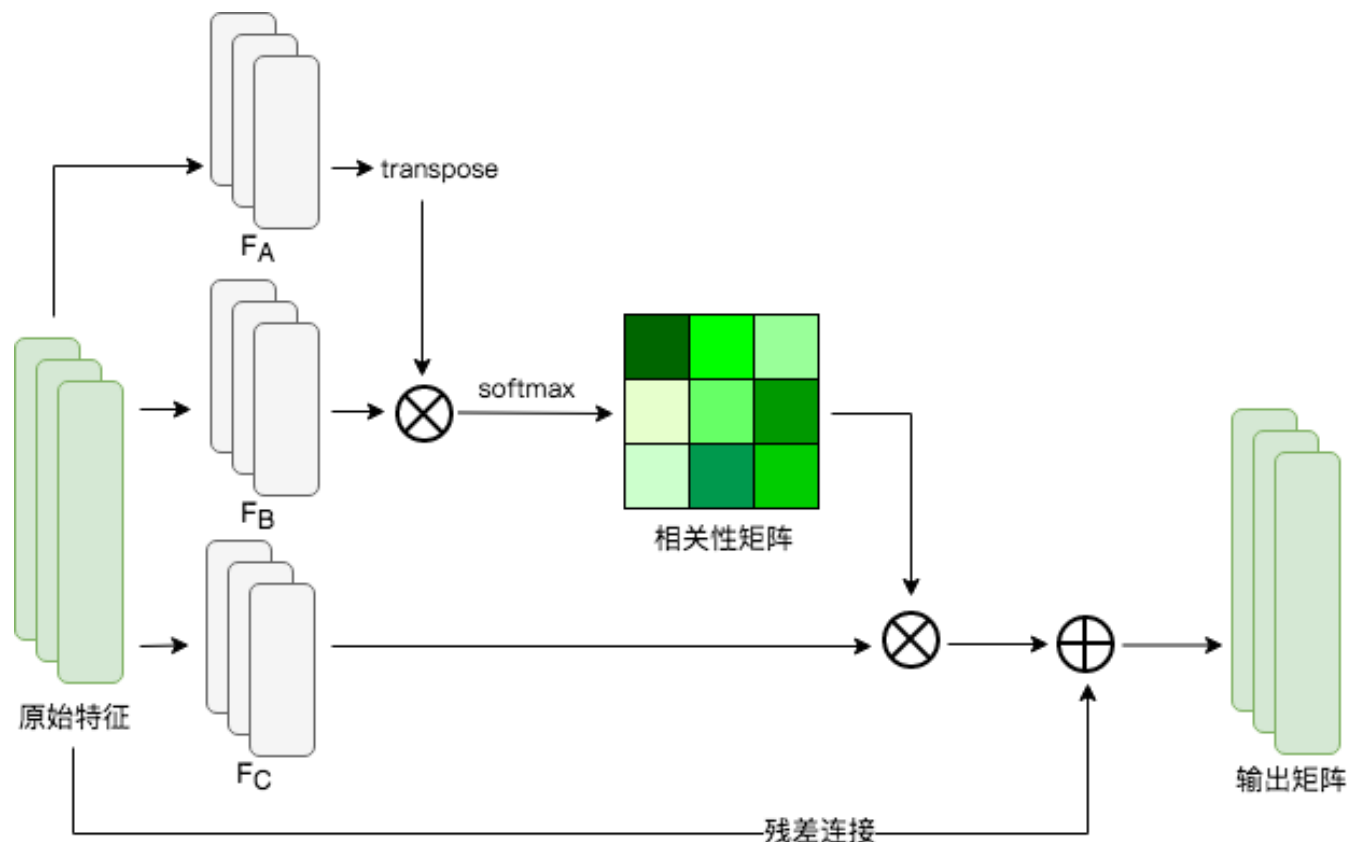
算法流程一：视频特征融合

存在问题

- 视频中存在噪音特征

解决方法

- 度量特征之间的相似性
- 增强相似特征
- 抑制离群特征



算法流程一：视频特征融合

存在问题

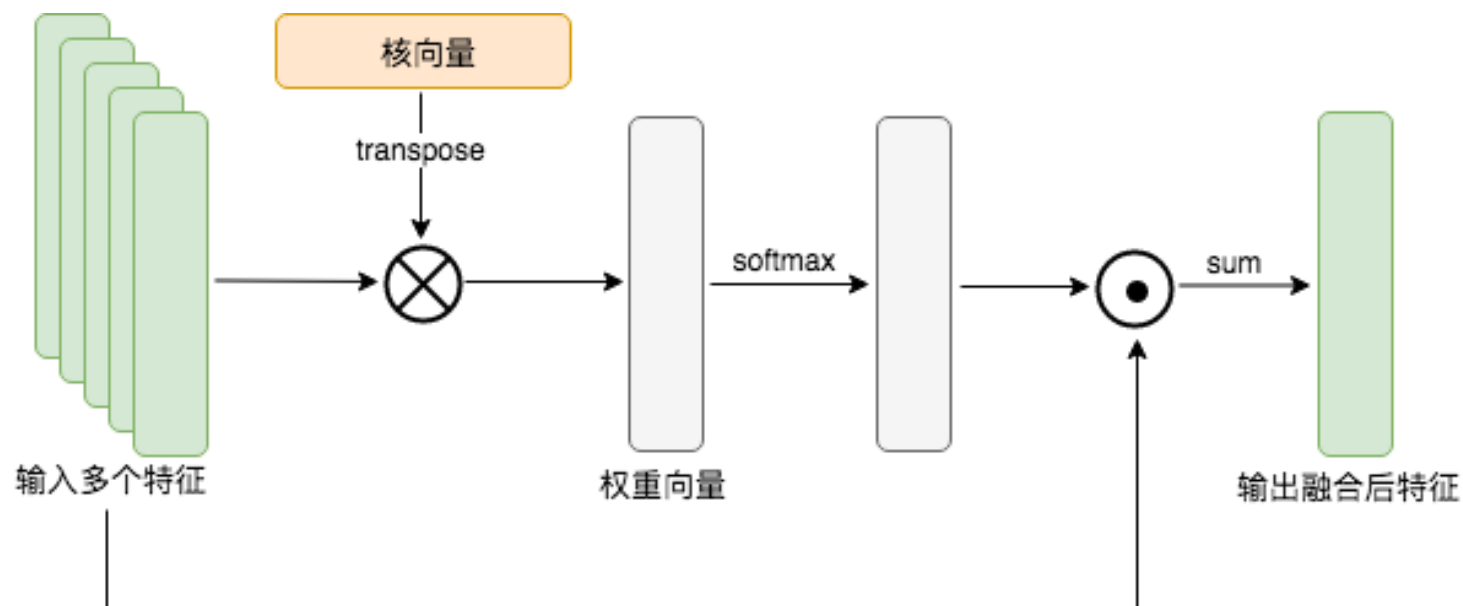
- 如何从多帧中提取有效表示

可能的方法

- 均值/加权均值：
启发式、次优解
- 循环神经网络：
时间分布差异

我们的方法

- NAN融合



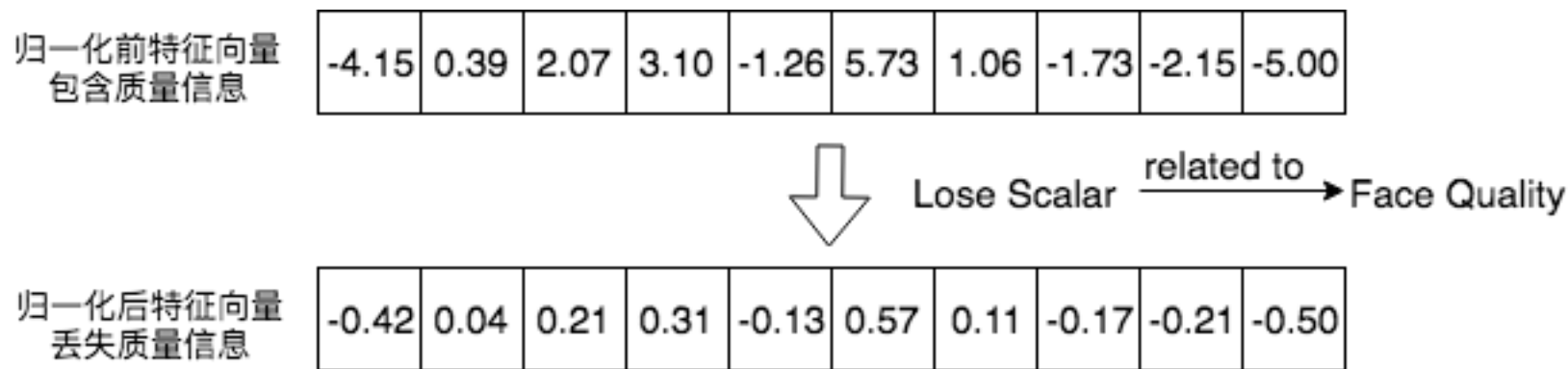
损失度量

- ArcFace：最大化不同特征在角度空间的差异

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}}$$

思考：ArcFace真的契合这个任务吗？

- 特征归一化丢失质量信息！



算法流程一：视频特征融合

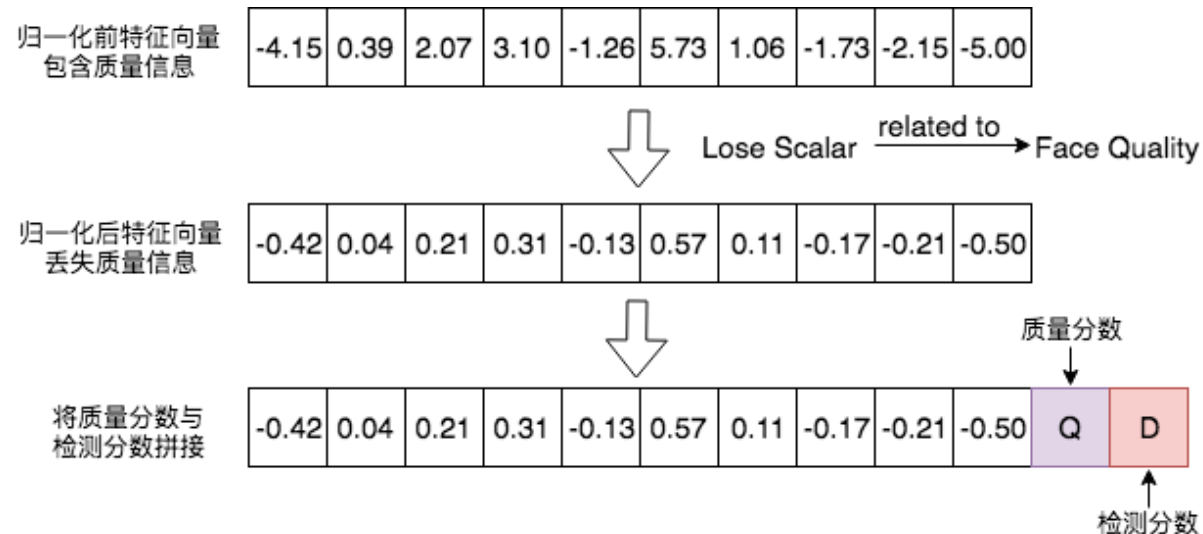
解决方法

- 归一化后再次引入质量信息

损失函数

- Focal Loss
- 注重困难样本（优于MEAN）、兼顾简单样本（优于OHEM）

$$FL(p_t) = -\alpha_t (1 - p_1)^\gamma \log(p_1)$$



算法流程二：多模态特征融合

存在问题

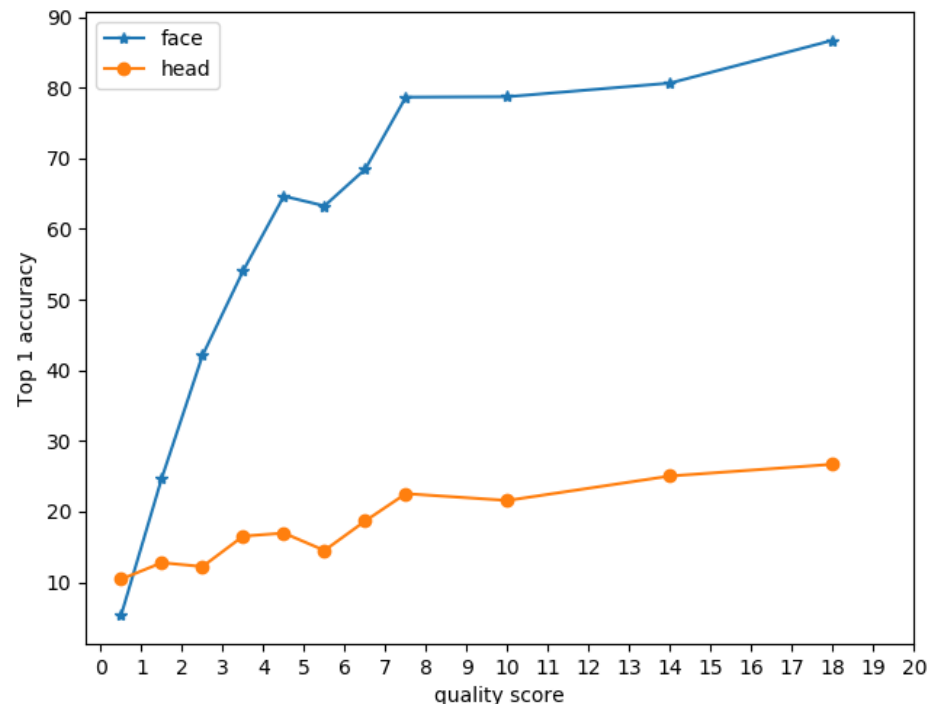
- 人脸、头部等特征对低质量视频的识别率低

解决方法

- 提取视频的场景特征帮助判断

基本模型

- SE-ResNeXt: Attention + Cardinality



[1] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[2] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7132–7141, 2018.

算法流程二：多模态特征融合

存在问题

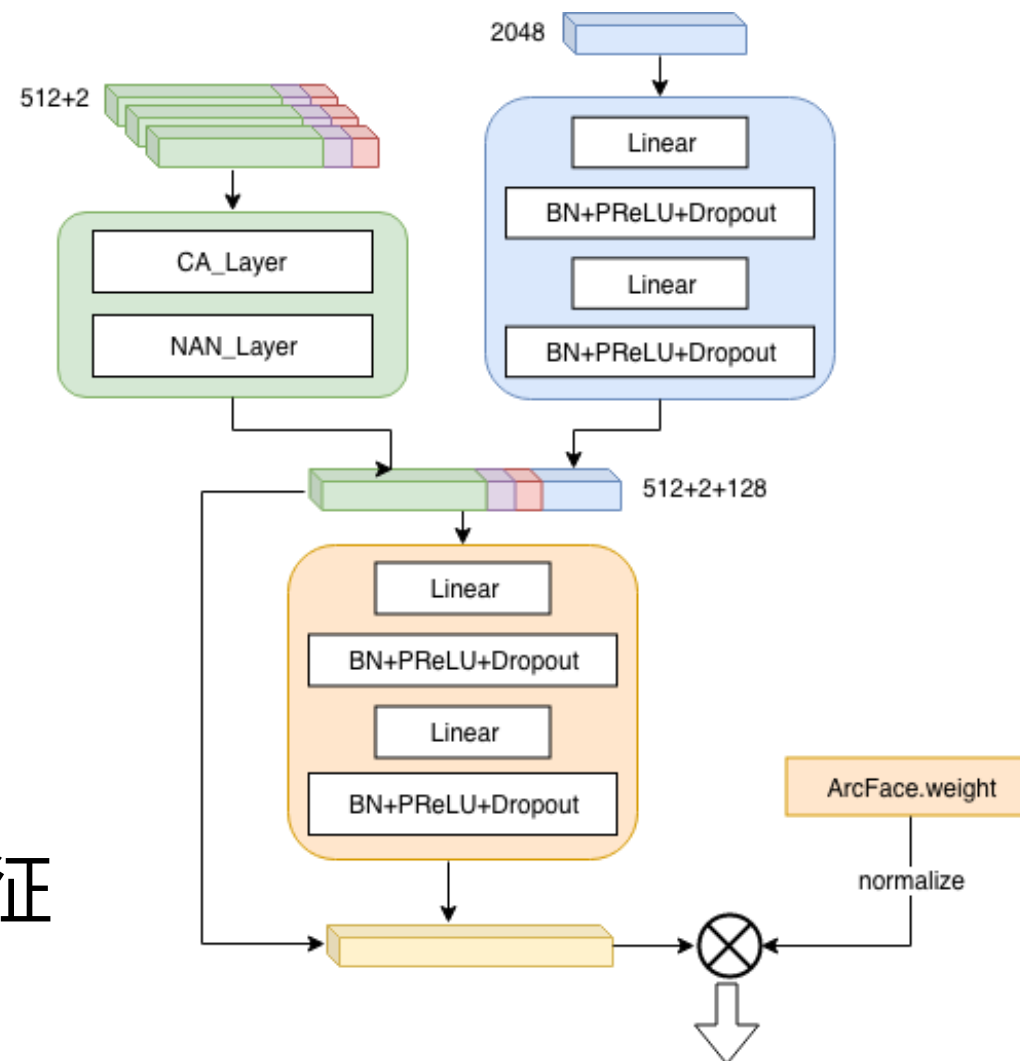
- 人脸、头部等特征对低质量视频的识别率低

训练过程

- 每段视频采样一帧进行训练
- 使用余弦退火训练20个epoch

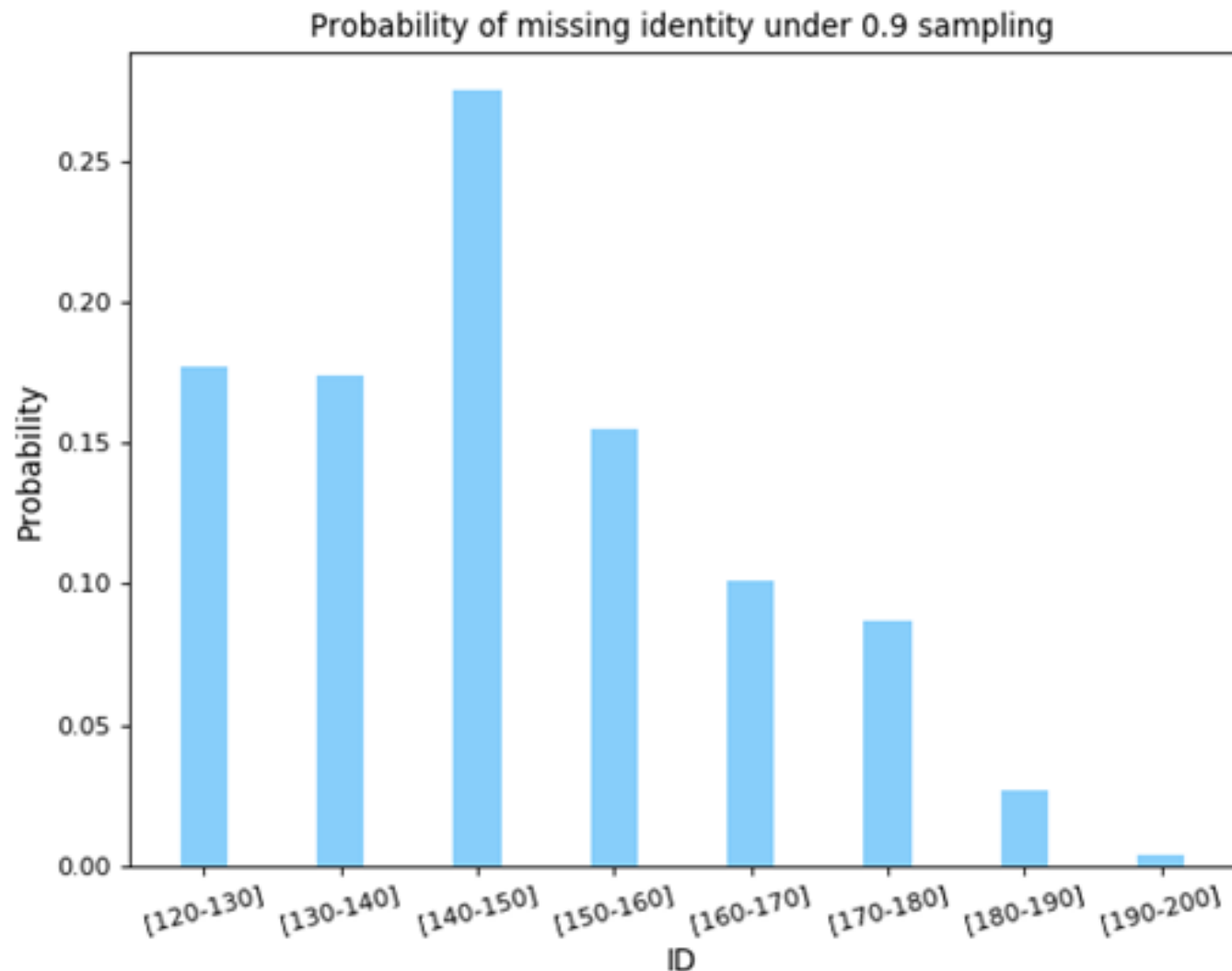
测试过程

- 每段视频采样一帧作为视频特征



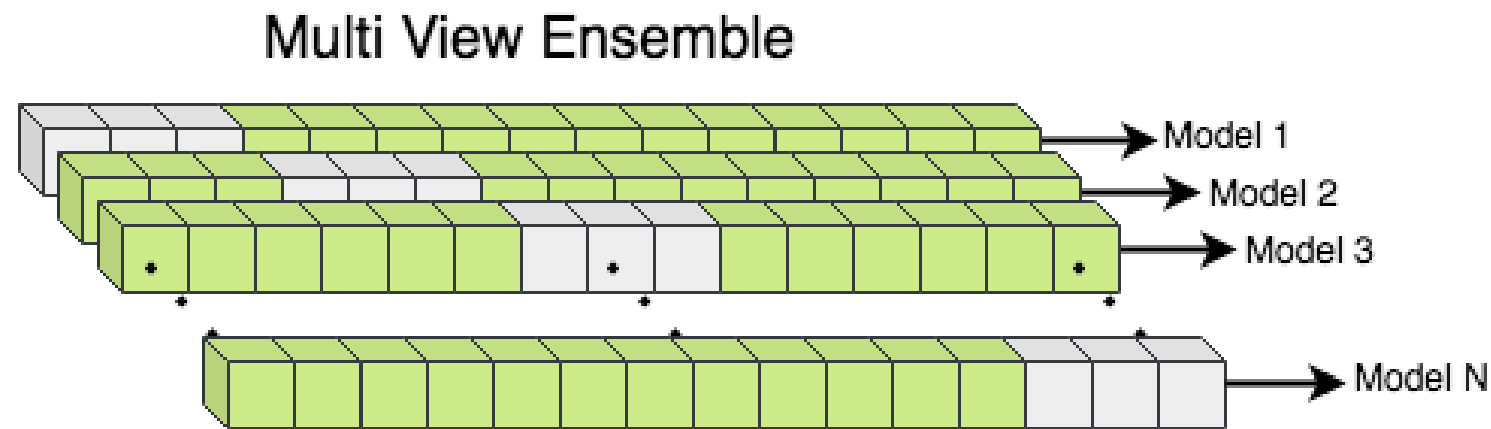
现有方法

- Bagging 导致ID缺失
- Boosting 对离群点敏感
- Stacking 过于复杂



我们的方法

- 借鉴协同训练的思想
- 对特征进行分割处理



其他情况

- 对无人脸的视频，采用场景特征预测
- 对质量分数最低的1%视频，采用场景、人脸加权预测

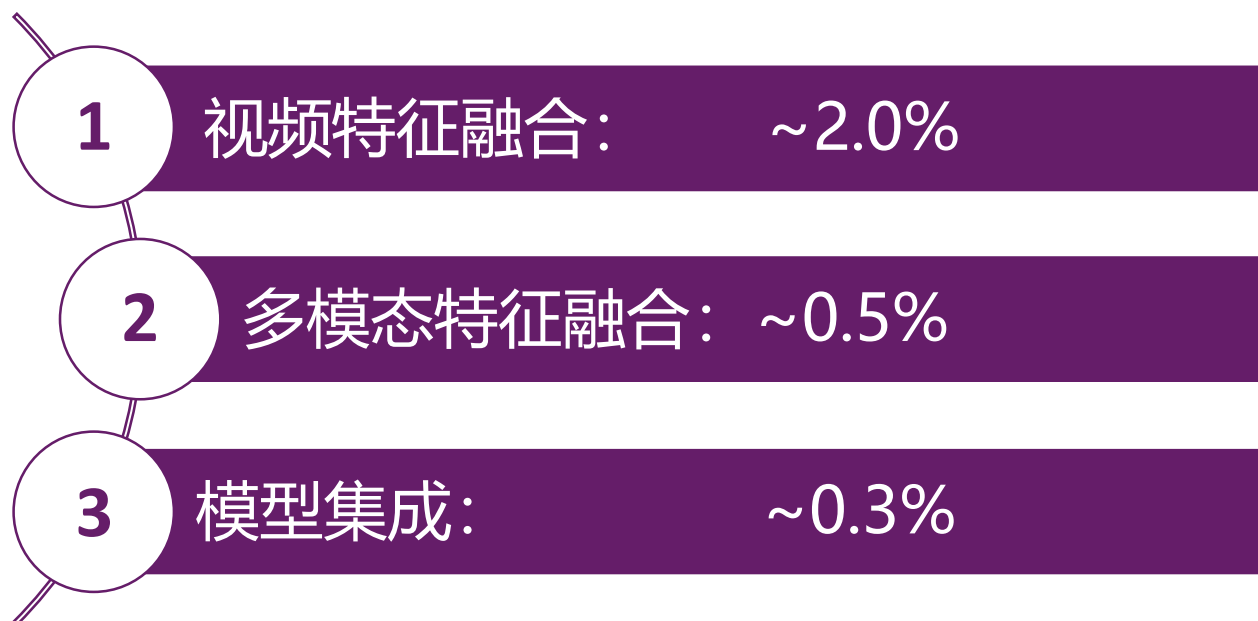
- 1 团队简介
- 2 数据理解
- 3 算法流程
- 4 实验结果**
- 5 结论建议

实验结果:

最终结果:

- mAP: 0.8986

模块提升:





- 1 团队简介
- 2 数据理解
- 3 算法流程
- 4 实验结果
- 5 结论建议**

其他可能的改进

- 多模态融合：两阶段→端到端
- 多视图划分：启发式→自适应
- 重新提取人脸特征
- 超参调优、re-rank...

其他

- 多模态与多视图的数据表示仍存在值得探究的问题
- 大规模人脸视频数据的其他可能应用场景：
 - 表情识别、人脸属性识别、视频生成...

- [1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009.
- [2] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. arXiv preprint arXiv:1801.07698, 2018.
- [3] J. Fu, J. Liu, H. Tian, Z. Fang, and H. Lu. Dual attention network for scene segmentation. arXiv preprint arXiv:1809.02983, 2018.
- [4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.
- [5] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7132–7141, 2018.
- [6] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision, pages 2980–2988, 2017.
- [7] I. Loshchilov and F. Hutter. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983, 2016.
- [8] J. Yang, P. Ren, D. Zhang, D. Chen, F. Wen, H. Li, and G. Hua. Neural aggregation network for video face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4362–4371, 2017.
- [9] 周志华. 半监督学习中的协同训练风范. 机器学习及其应用, pages 259–275, 2007.

汇报完毕 恳请指正

Presented by

一个名字团队