



基于深度学习的人脸表情识别

学生：胡含哲 指导老师：崔锦实 副教授，张建 教授



胡含哲 151120043

南京大学物理学院

2019-06-03



目录 | CONTENT



- 1 研究背景**
- 2 基于集成学习的静态表情识别**
- 3 基于生成对抗网络的表情识别**
- 4 总结与展望**



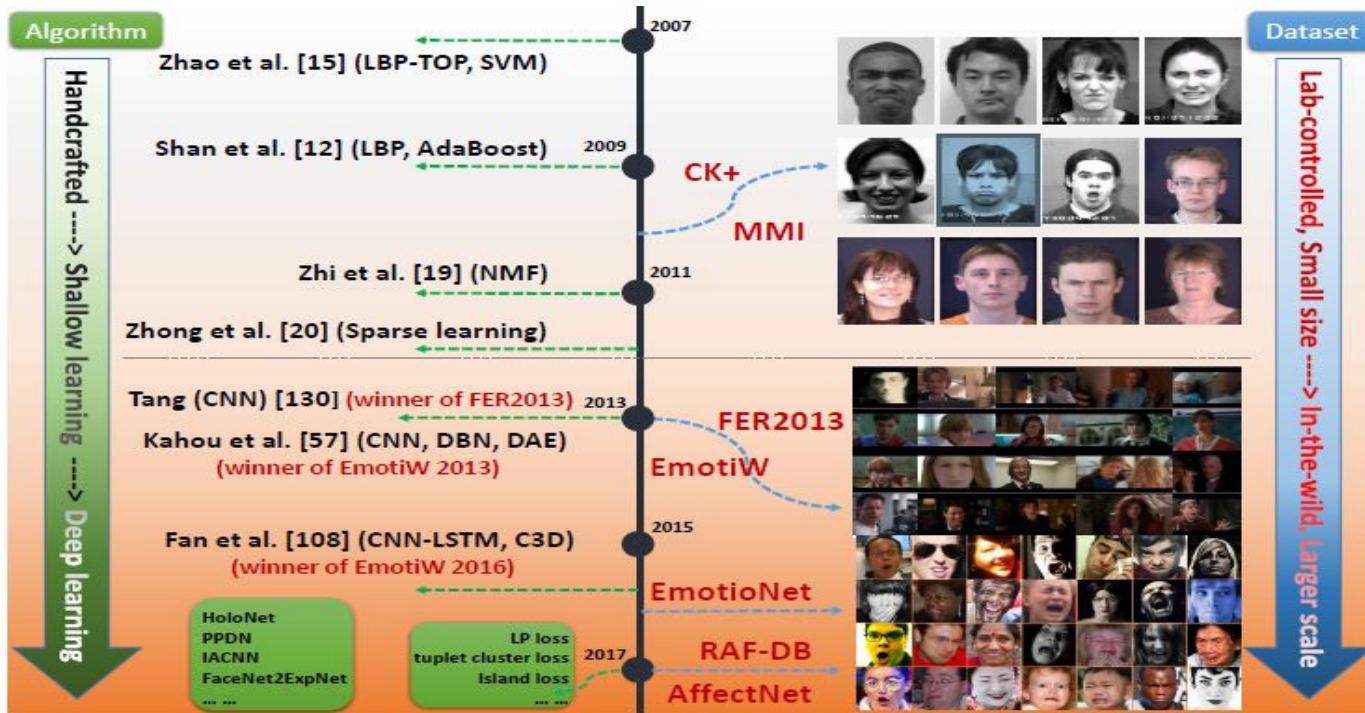
01

研究背景

研究背景

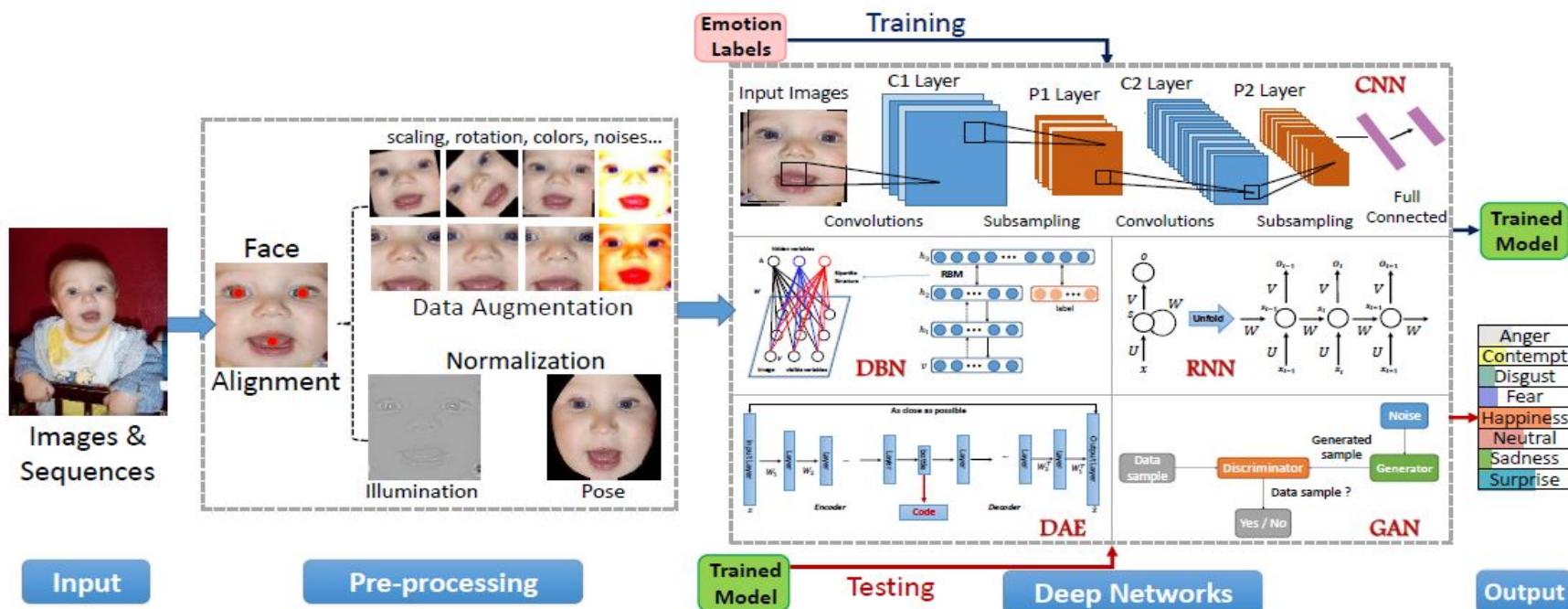
人脸表情是人们表达情感状态和传递意图最有力且自然的方式。因为其在机器人系统、医疗系统、驾驶行为监测等很多机交互方面的广阔应用，出现了很多人脸表情识别的相关研究。

六种基本表情被定义出来：快乐、惊讶、伤心、生气、害怕、厌恶，由于其概括性与简洁性，此表情模型被广泛应用。



研究背景

人脸识别系统主要由人脸图像的获取、人脸检测、特征提取和特征分类四部分组成。使用深度学习方法，可以将整个系统做成端到端模式。





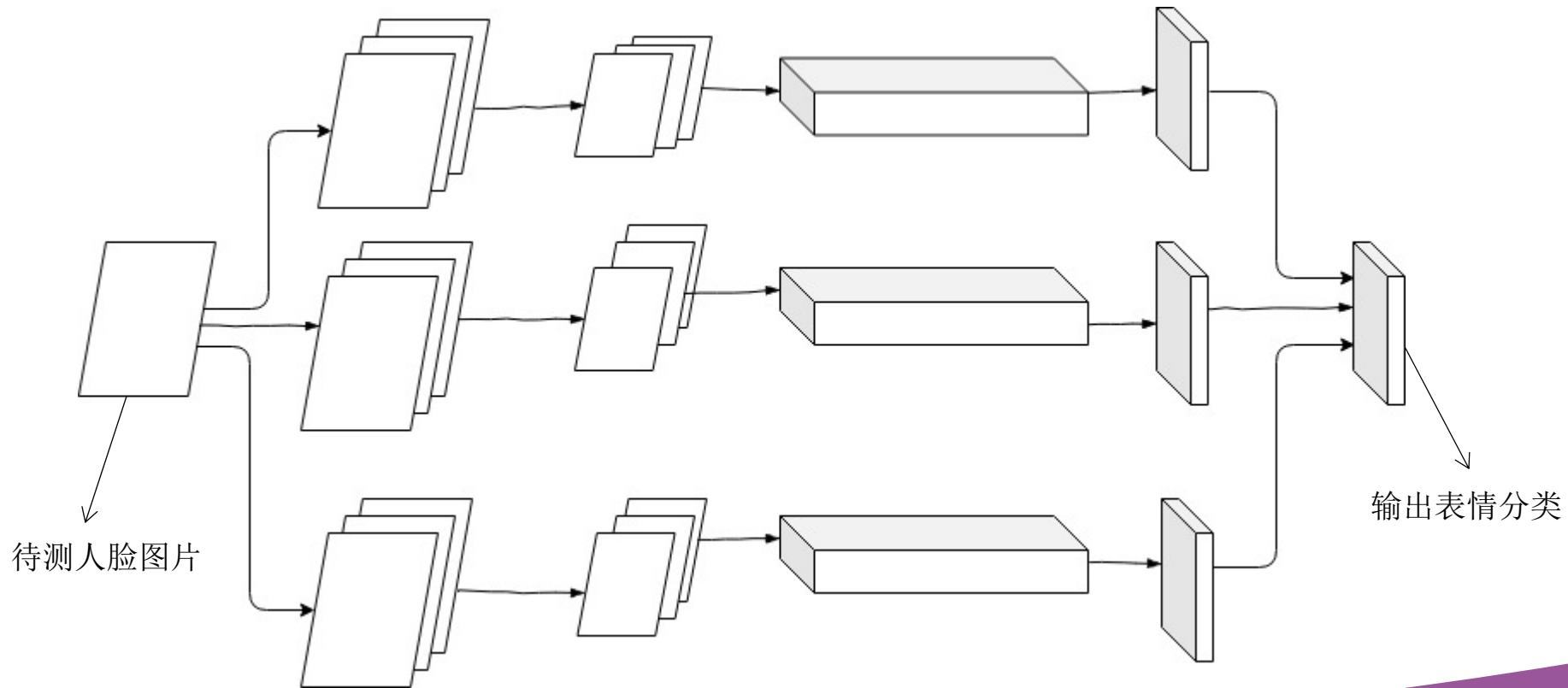
02

基于集成学习的 静态表情识别



CNN集成示例

三个分支各自充当一个分类器，并将各自输出集成在一起作为最终输出





网络细节

网络层	CNN1	CNN2	CNN3
输入	-	-	-
conv1	3×3	5×5	7×7
Max pooling1	2×2	3×3	2×2
conv2	3×3	4×4	7×7
Max pooling 2	2×2	3×3	2×2
conv3	3×3	5×5	7×7
Max pooling 3	2×2	3×3	2×2
FC	FC3072, dropout=0.8		
	FC2048, dropout=0.5		
	FC1024, dropout=0.5		
	FC512, dropout=0.5		
输出	7	7	7



实验结果

FER2013数据集

表情	生气	厌恶	恐惧	开心	伤心	惊讶	中性	总计
标签	0	1	2	3	4	5	6	-
数量	4953	547	5121	8989	6077	4002	6198	35877

实验结果

网络	准确率
CNN1	63.45%
CNN2	63.27%
CNN3	65.89%
VGG19	68.98%
ResNet18	69.65%
ResNet34	68.23%
集成	71.79%



与其他方法对比

方法	准确率
Ensembled CNNs	71.79%
CPC[2018]	71.35%
ICML FER2013 1st	71.16%
DNNRL[2016]	70.60%
FC3072[2016]	70.58%
ICML FER2013 2nd	69.27%
ICML FER2013 3rd	68.82%

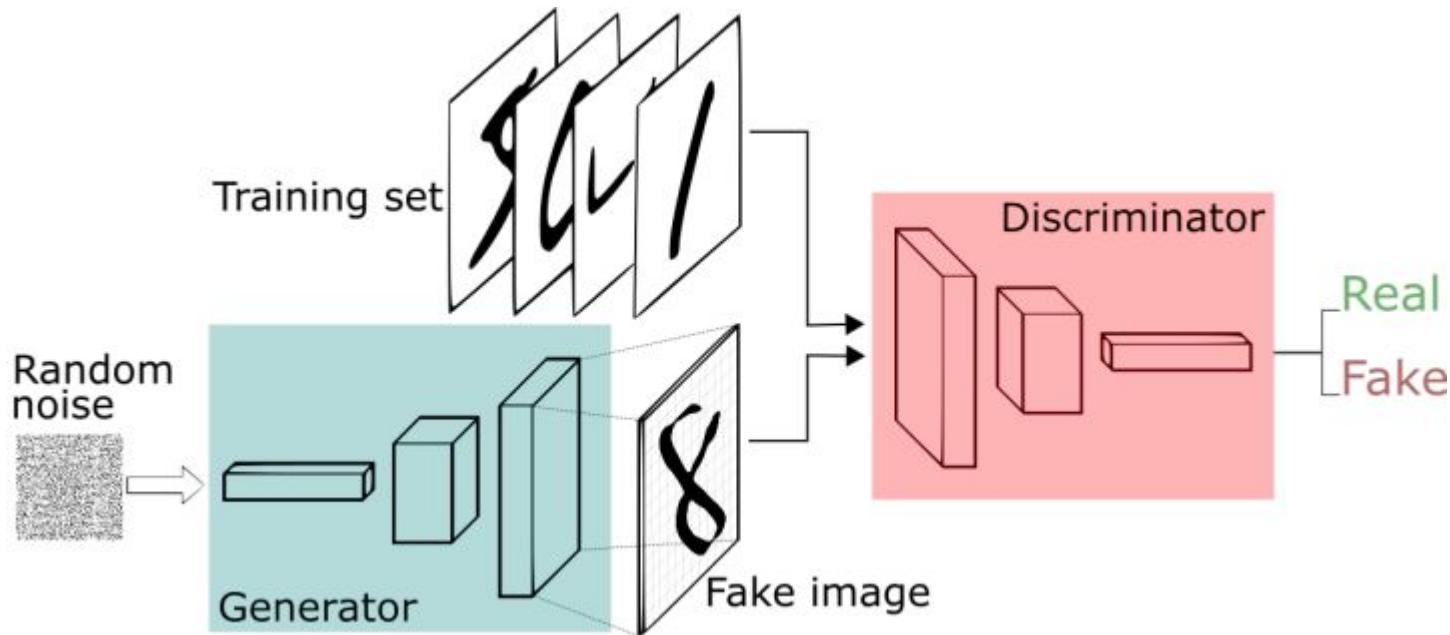


03

基于生成对抗网 络的表情识别

生成对抗网络 (GAN)

生成器G和判别器D的相互博弈，达到一种动态平衡。

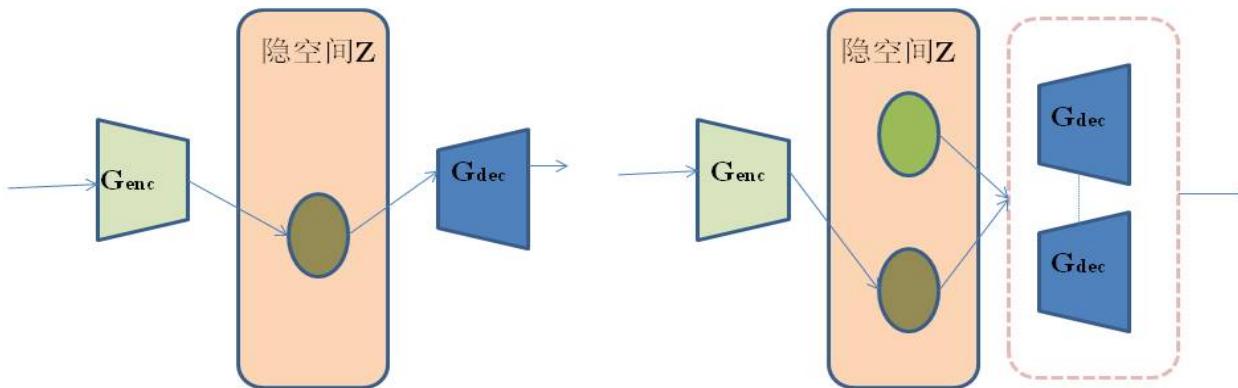


$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$



表征学习

表征信息学习不完整问题

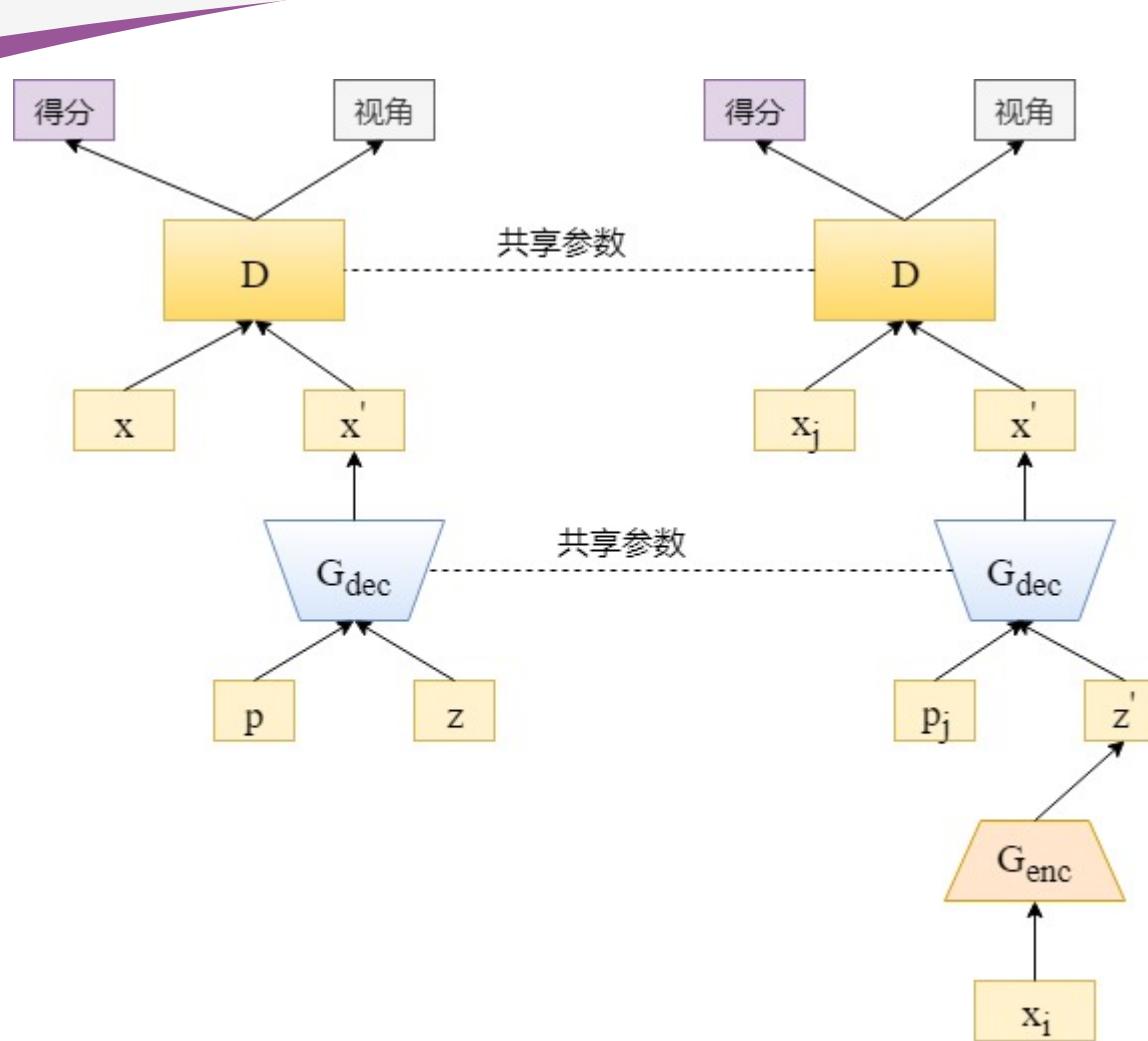


普通的编码器-解码器结构

改进的编码器-解码器结构



模型结构





损失函数设计

生成路径：训练解码器 G_{dec} 和判别器D

$$D: \quad E_{z \sim P_z} [D_s(G_{dec}(v, z))] - E_{x \sim P_x} [D_s(x)] - \alpha_1 E_{x \sim P_x} [P(D_p(x) = p)]$$

$$G_{dec}: \quad E_{z \sim P_z} [D_s(G_{dec}(p, z))] + \alpha_2 E_{z \sim P_z} [P(D_p(G_{dec}(p, z)) = p)]$$

重构路径：训练编码器 G_{enc} ，优化判别器D

$$D: \quad E_{x_i, x_j \sim P_x} [D_s(\bar{x}_j) - D_s(x_i)] - \beta_1 E_{x_i \sim P_x} [P(D_p(x_i) = p_i)]$$

$$G_{enc}: \quad E_{x_i, x_j \sim P_x} [D_s(\bar{x}_j) + \beta_2 P(D_p(\bar{x}_j) = p_j) - \beta_3 L_1(\bar{x}_j, x_j) - \beta_4 L_p(E_p(x_i), p_i)]$$



整体算法流程

算法 5.1 双路径训练模型

输入:

有标签的训练图片 X , 最大步数 T , 批数量 m

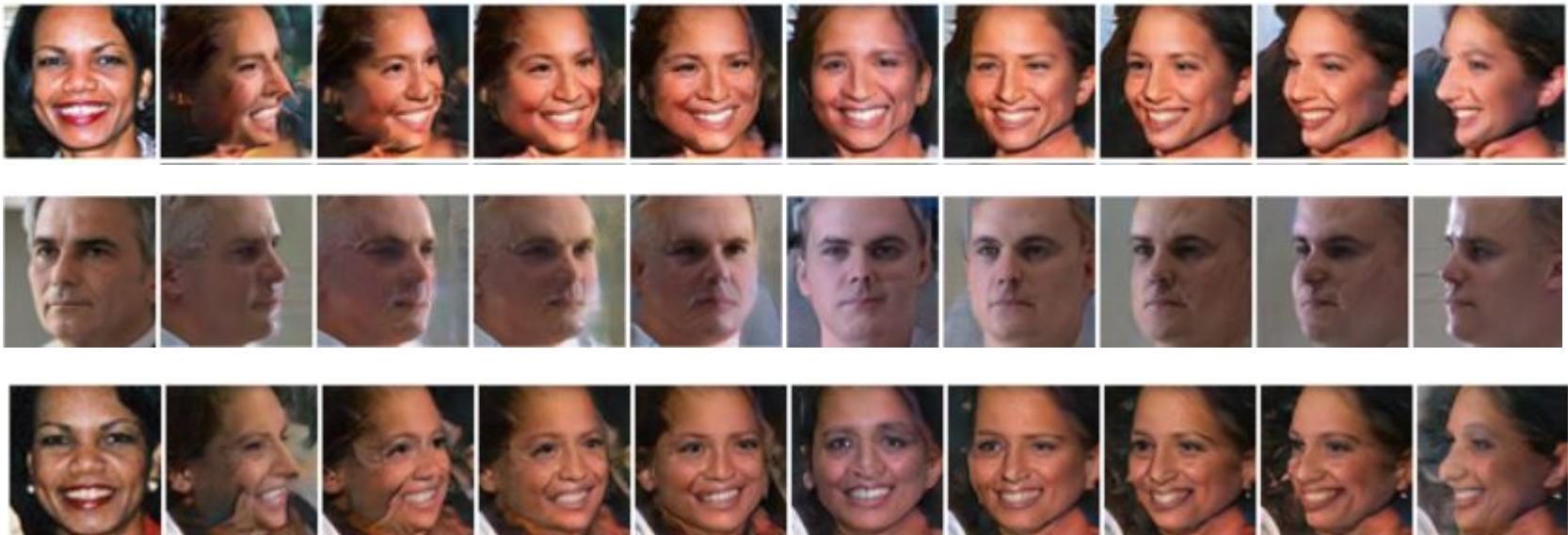
输出:

训练好的 G_{enc}, G_{dec}, D .

```
1: for  $t = 1$  to  $T$  do
2:   for  $i = 1$  to  $m$  do
3:     取样  $z \sim P_z, x_i \sim P_x$ , 姿态标签为  $p_i$ ;
4:      $\bar{x} = G_{dec}(p_i, z)$ ;
5:     根据对应优化函数更新判别器  $D$  和解码器  $G_{dec}$ ;
6:     取样姿态标签为  $p_j$  的  $x_j \sim P_x(x_i)$  和  $x_j$  为同一人的不同头部姿态的图
      片);
7:      $\bar{z} = E(x_i)$ ;
8:      $\bar{x}_j = G_{dec}(p_j, \bar{z})$ ;
9:     根据对应优化函数更新判别器  $D$  和编码器  $G_{enc}$ 
10:    end for
11:  end for
```

实验效果

可以生成不同角度的人脸





实验结果

在Multi-PIE 数据集上不同视角不同表情对应准确率

准确率	0度	15度	30度	45度	60度	75度	90度	平均
中性	91%	92%	94%	90%	90%	86%	69.5%	87.5%
微笑	93%	96.5%	99%	98%	94%	96%	82%	94.1%
眯眼	79%	76.5%	77%	77%	75%	78%	74%	76.64%
惊讶	99%	96%	97%	98%	97%	91%	82%	94.3%
厌恶	77%	75%	77%	73%	74%	68%	70%	73.4%
尖叫	97%	97%	97%	96%	97%	91%	85%	94.3%
全部	89.3%	88.8%	90.2%	88.7%	87.8%	85.0%	77.1%	86.71%



04

总结展望



总结与展望

- 本文采用的集成学习思想，很好地说明了集成学习的强大，未来可以改进子学习器的结构以达到更好的识别精度。
- 对于人脸矫正问题，本文提出的方法可以很好地还原正面人脸，并且保留其主要特征。未来也可以从人脸关键点入手，并且为表情信息定义新的关键点，从而防止转正过程中出现信息丢失。
- 新型的GAN结构，可以更好地提取特征，未来可以深入研究GAN的内部原理，构造更加完备的生成结构。



学生：胡含哲
学号：151120043

谢谢聆听

Thank You

