



有限样本条件下的相机来源 鉴别方法

谭跃, 王波, 赵美娟, 孔祥维, 郭艳卿, 李明

大连理工大学信息安全研究中心

yue985@mail.dlut.edu.cn

- **1. 研究背景**
- **2. 半监督学习算法&相机来源鉴别**
- **3. 实验结果**
- **4. 总结&展望**

军事



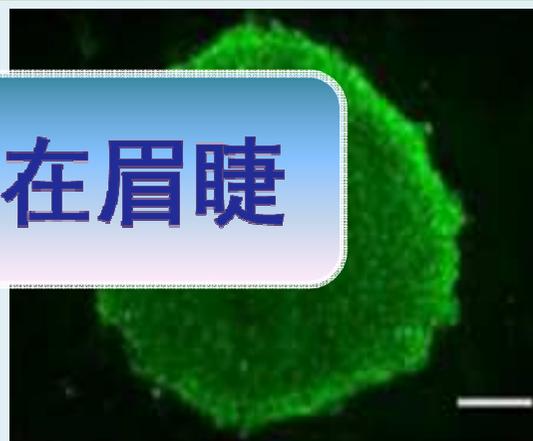
政治



社会



科学



数字
图像

数字图像取证迫在眉睫

- 《Nature》（2007）
- 美国“桑迪”飓风（2012）
- 周老虎（2006）



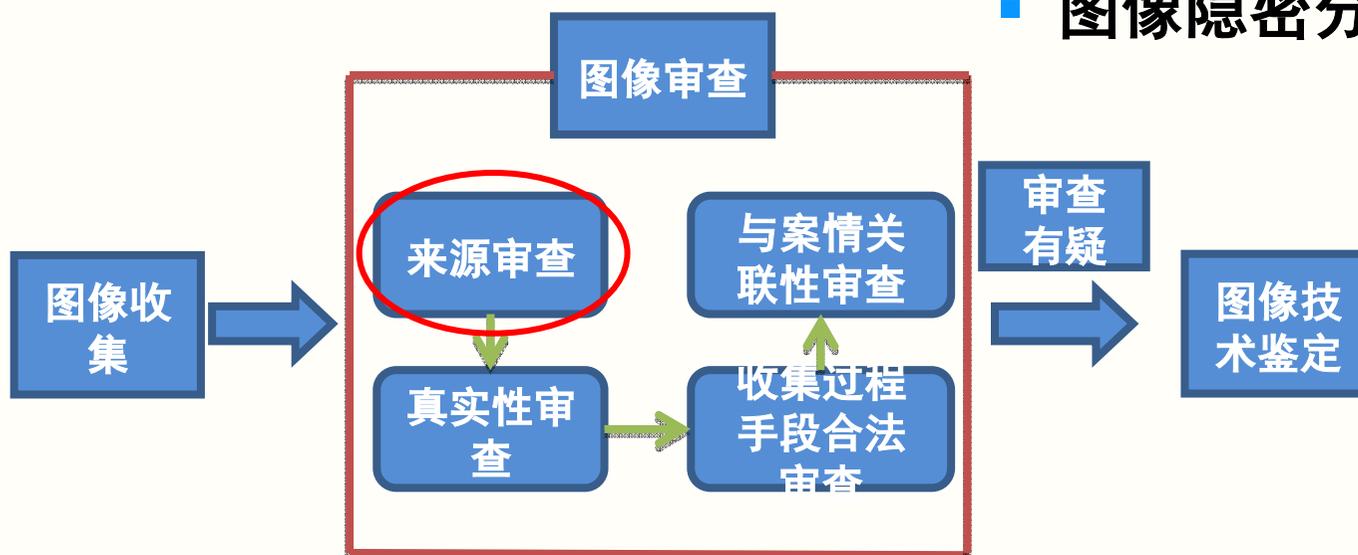
有图未必有真相
眼见不再为实！



➤ 数字图像取证技术

目的
分类

- 图像来源认证技术
- 图像篡改检测技术
- 图像隐密分析检测技术



数字图像作为证据的举证流程

➤ 数字图像取证技术

手段
分类

- 主动水印技术
- 被动盲取证技术

不需要事先的预处理，实施性更强

来源取证的三个层次

设备类型



设备型号



Nikon D3200



Sony DSC-RX1



Sony DSC-HX50

.....

设备个体



目前相机来源取证存在的问题

——训练样本数量不足时，算法性能下降显著

文献	每个型号相机训练样本数量
Kharrazi M, Sencar H, Memon N. Blind source camera identification [C]. in Proc. IEEE International Conference on Image Processing (ICIP), 2004: 709-712.	150
Xu G, Shi Y Q. Camera model identification using local binary patterns [C]. in Proc. IEEE International Conference on Multimedia and Expo (ICME), Melbourne, 2012: 392-397.	150-300
王波. 利用成像引入特征的数字图像被动盲取证研究 [D]. 大连: 大连理工大学, 2010.	300
孙雪辉. 利用多特征和软件成像的手机图像盲取证技术[D]. 大连: 大连理工大学, 2010.	80

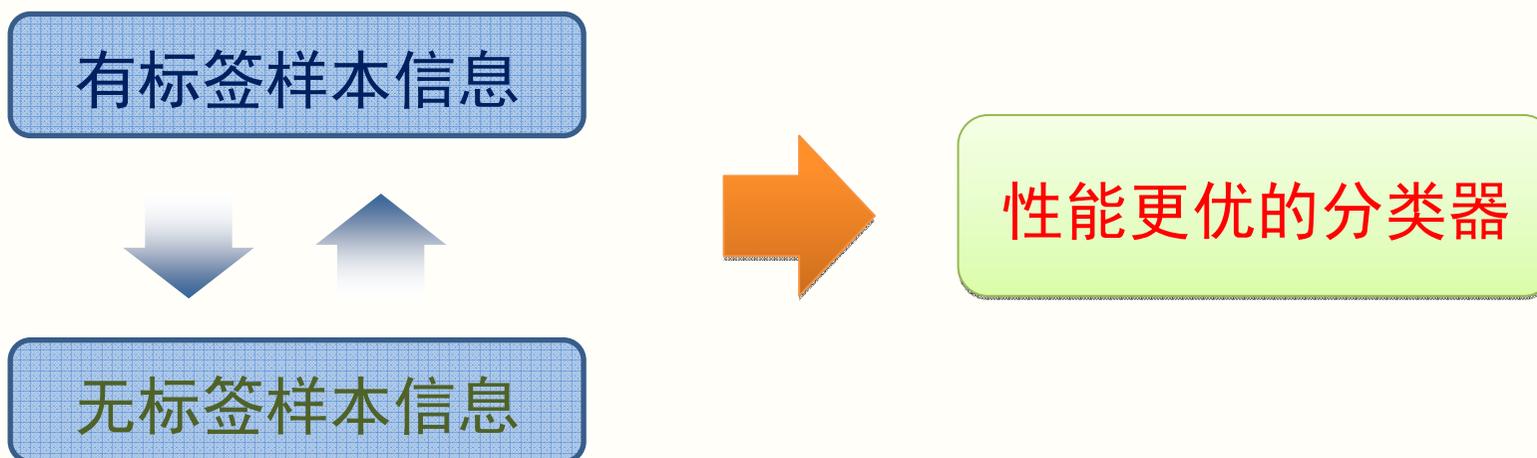


- 1. 研究背景
- 2. 半监督学习算法&相机来源鉴别
- 3. 实验结果
- 4. 总结&展望

半监督学习

- 自学习
- 协同训练
- 传导支持向量机
- 基于图
- 基于生成模型

半监督学习的基本思想



自学习算法

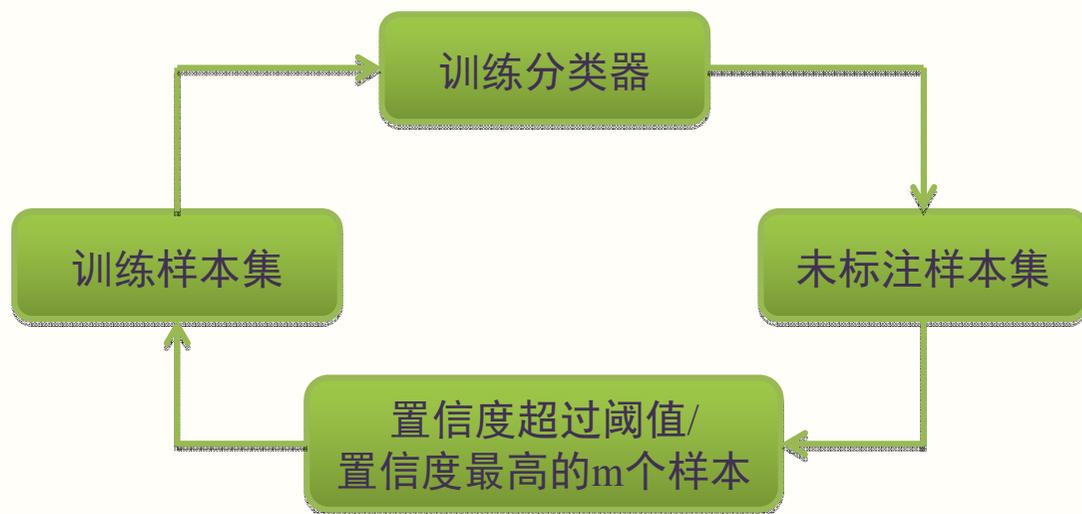


图1 自学习模型

协同训练算法

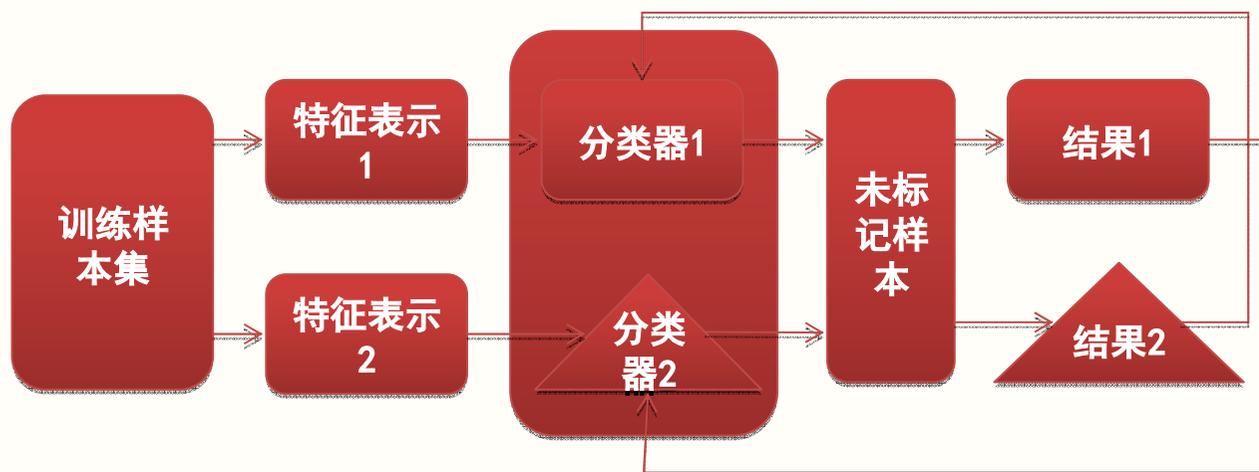


图2 协同训练模型

本文基本思想:

半监督学习

引入



经典的相机来源鉴别特征

经典的相机来源鉴别特征



局部二进制模型
(LBP)

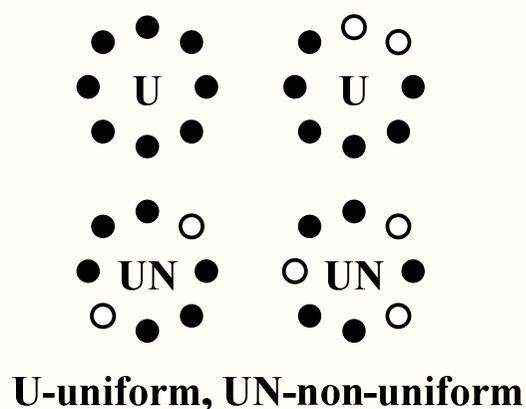
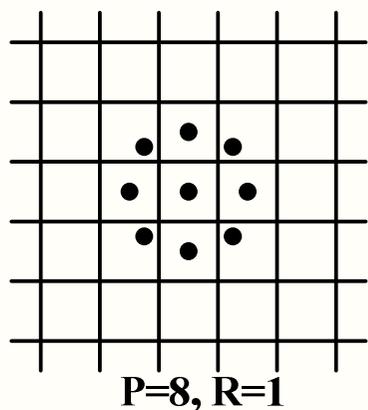


图像质量特征
(IQM)



色彩滤波阵列插值
(CFA)

➤ 局部二进制模型 (LBP)



$$LBP_{P,R}^{u2} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

共有 $2^8=256$ 种pattern, 但仅考虑“uniform-pattern”则为59种

空域图像

经过预测误差处理的图像

经过一阶小波变换的图像



$(59+59+59) \times 2=354$ 维



➤ 图像质量特征 (IQM)

类别	彩色滤波阵列插值 (CFA)	类别	图像质量特征
基于像素差的特征	均方误差: Q1	基于频谱的特征	频谱幅值误差: Q8
	平均绝对误差: Q2		均值系数240维
	修正无穷范数: Q3		频谱相位误差: Q9
			方差系数240维
基于相关性的特征	图像的保真度: Q4	图像块频谱幅值误差: Q11	
	规范化的交叉相关: Q5	图像块频谱相位误差: Q12	
	Czekonowski相关: Q6	图像块频谱相位-幅值误差: Q13	
	平均角度相似性: Q7		



- 1. 研究背景
- 2. 半监督学习算法&相机来源鉴别
- 3. 实验结果
- 4. 总结&展望

实验数据库：5部相机，5种不同品牌

图像包括风景、建筑、人物、室外、室内、白天、晚上等。

序号	相机型号	分辨率	图像数量
Cam1	Casio_EX_Z150	3264×2448	350
Cam2	Olympus_mju	3648×2736	350
Cam3	Nikon_D200	3872×2592	350
Cam4	Pentax_OptioA40	4000×3000	350
Cam5	Canon_Ixus70	3072×2304	350

参数设置：

- y —每一个型号相机有标签样本数量
- m —迭代步长
- n —迭代次数

自学习与协同训练实验结果比较

- 有标签样本 y 为30
- 迭代次数为10

	I			II			III		
	文献 [4]	自学 习	协同 训练	文献 [9]	自学 习	协同 训练	文献 [10]	自学 习	协同 训练
1	89.4	50.0	33.7	0.0	72.1	25.3	58.1	98.9	94.9
2	66.3	59.7	82.0	18.1	77.1	19.6	100	100	100
3	25.9	78.9	80.1	61.6	15.2	75.7	35.0	90.5	100
4	4.1	32.4	22.2	10.0	31.4	84.8	0.0	56.7	97.1
5	67.8	81.9	77.3	27.5	47.6	20.5	51.9	100	98.2
平均	50.7	60.6	59.1	24.3	48.7	45.2	49.1	89.2	98.1

准确率提升
10%

准确率提升
25%

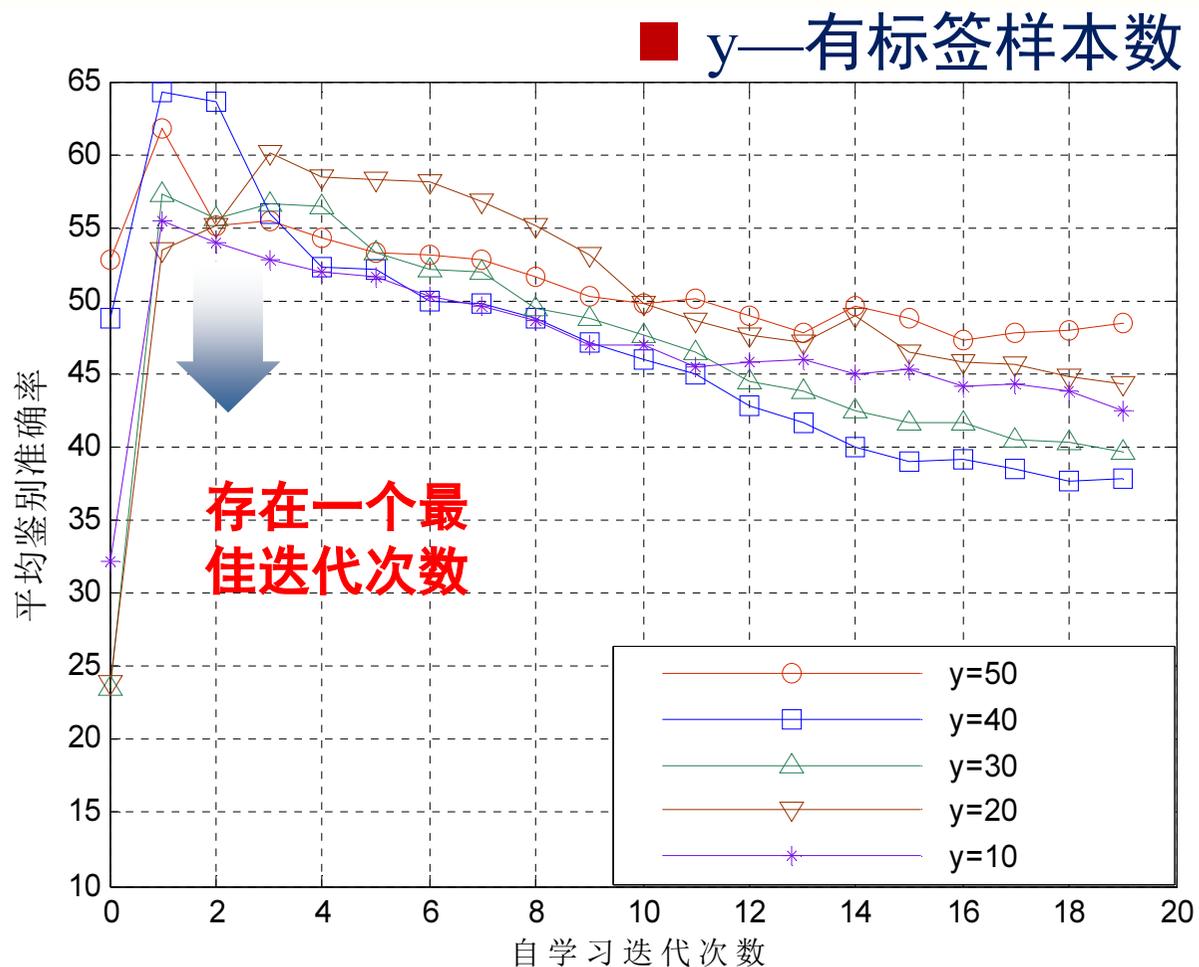
准确率提升40-
50%

[4] Xu G, Shi Y, et al. Parallel identification of local binary patterns in Proc. IEEE International Conference on Multimedia and Expo (ICME), Melbourne, 2012: 392-397.

[9] 孙雪辉. 利用多特征和软件成像的手机图像盲取证技术[D]. 大连: 大连理工大学, 2010.

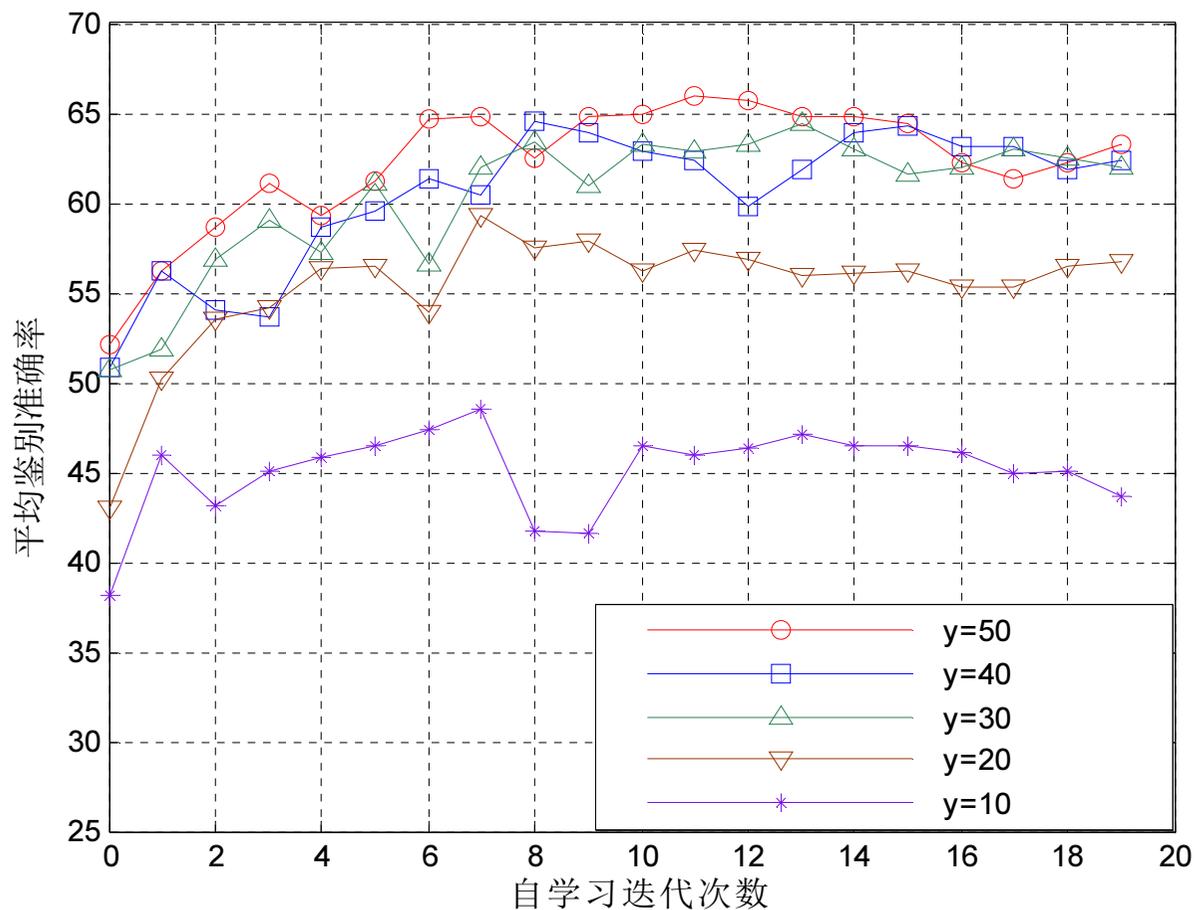
[10] 王波. 利用成像引入特征的数字图像被动盲取证研究 [D]. 大连: 大连理工大学, 2010.

基于自学习的相机来源鉴别



采用LBP特征的自学习方法准确率变化曲线

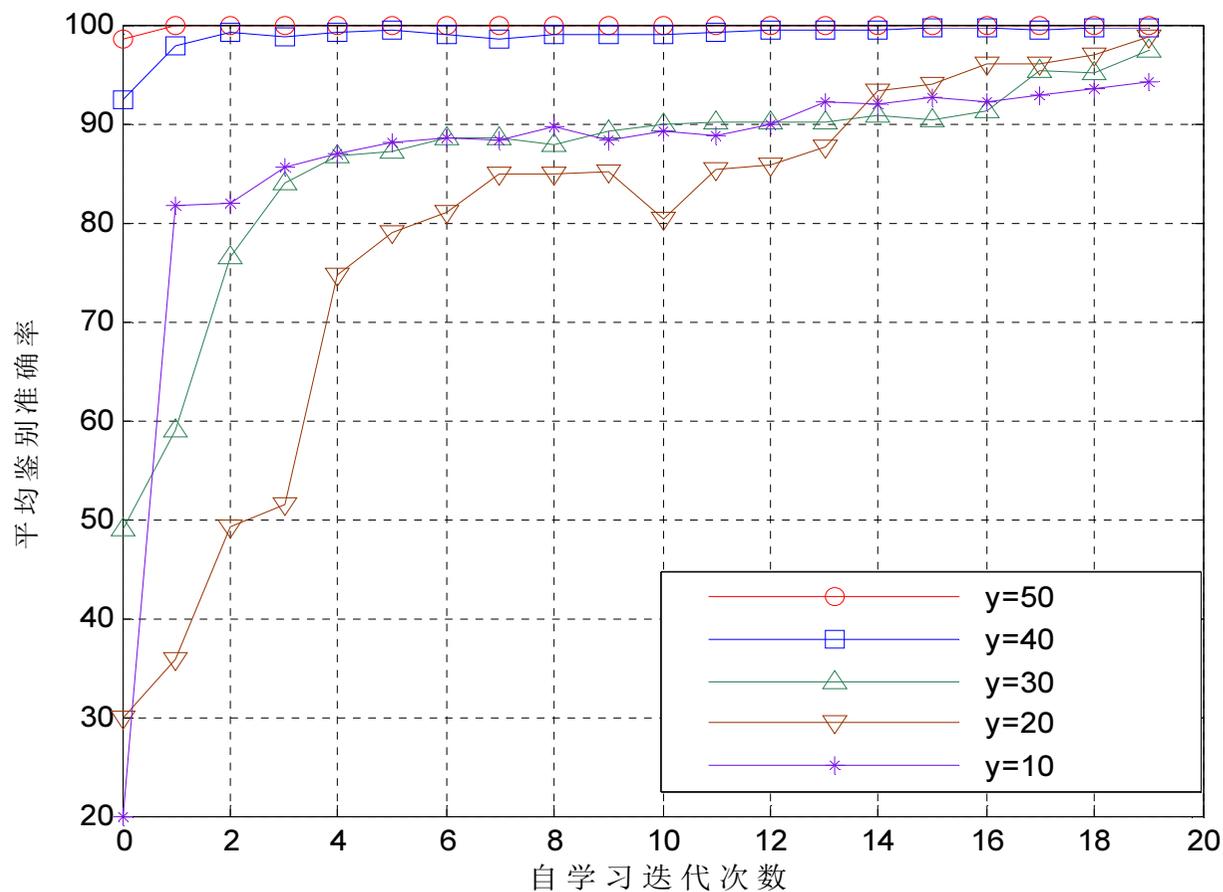
基于自学习的相机来源鉴别



采用IQM特征的自学习方法准确率变化曲线

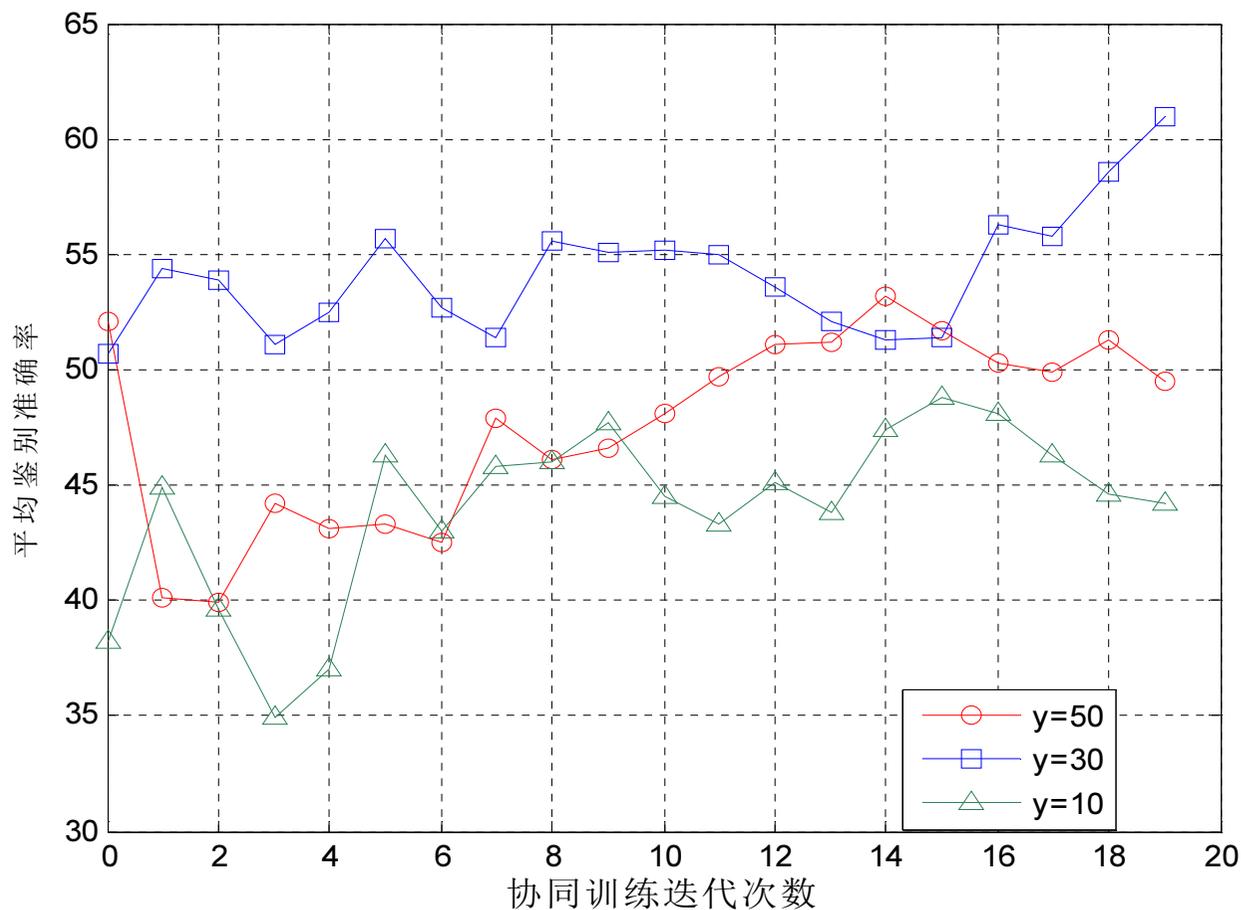
基于自学习的相机来源鉴别

✓ 不同特征最佳迭代次数不同



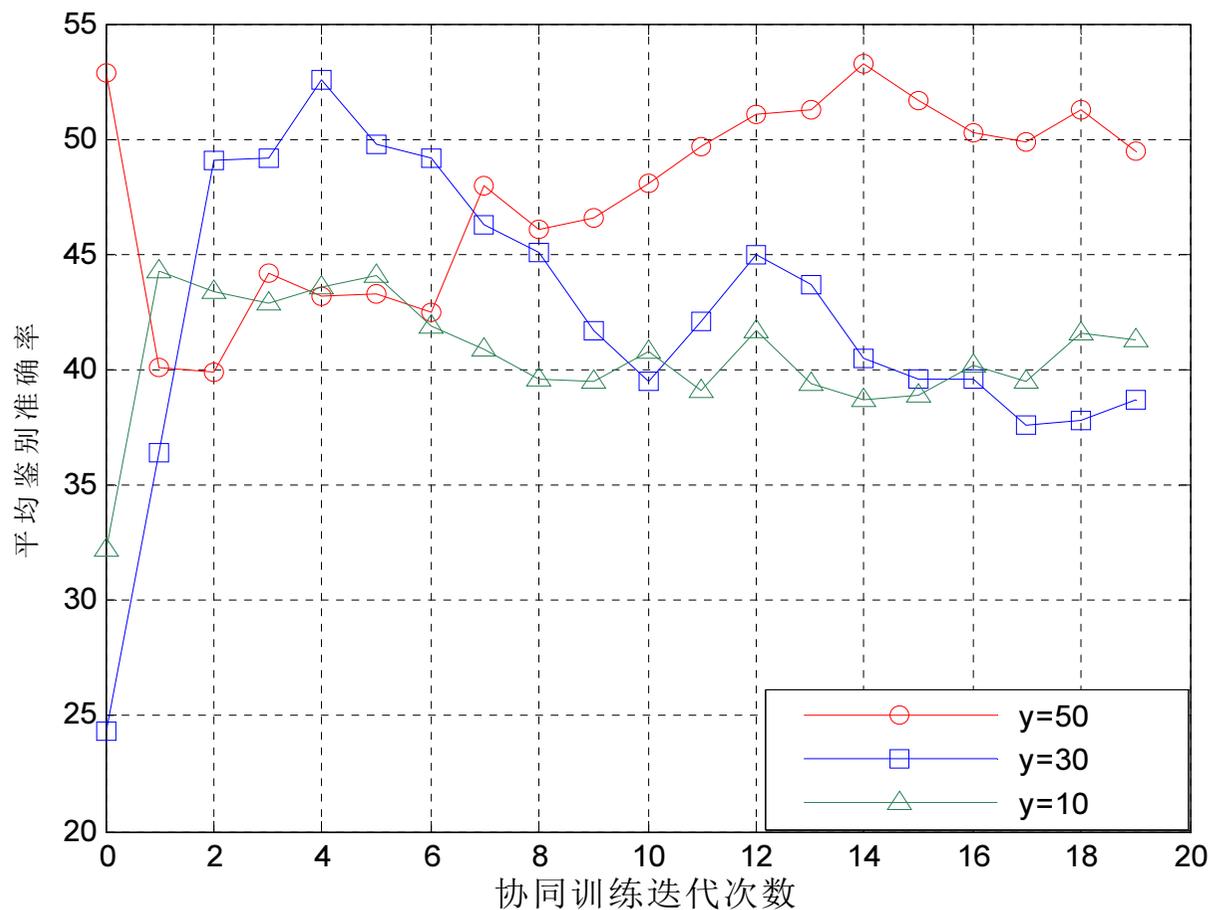
采用CFA特征的自学习方法准确率变化曲线

基于协同训练的相机来源鉴别



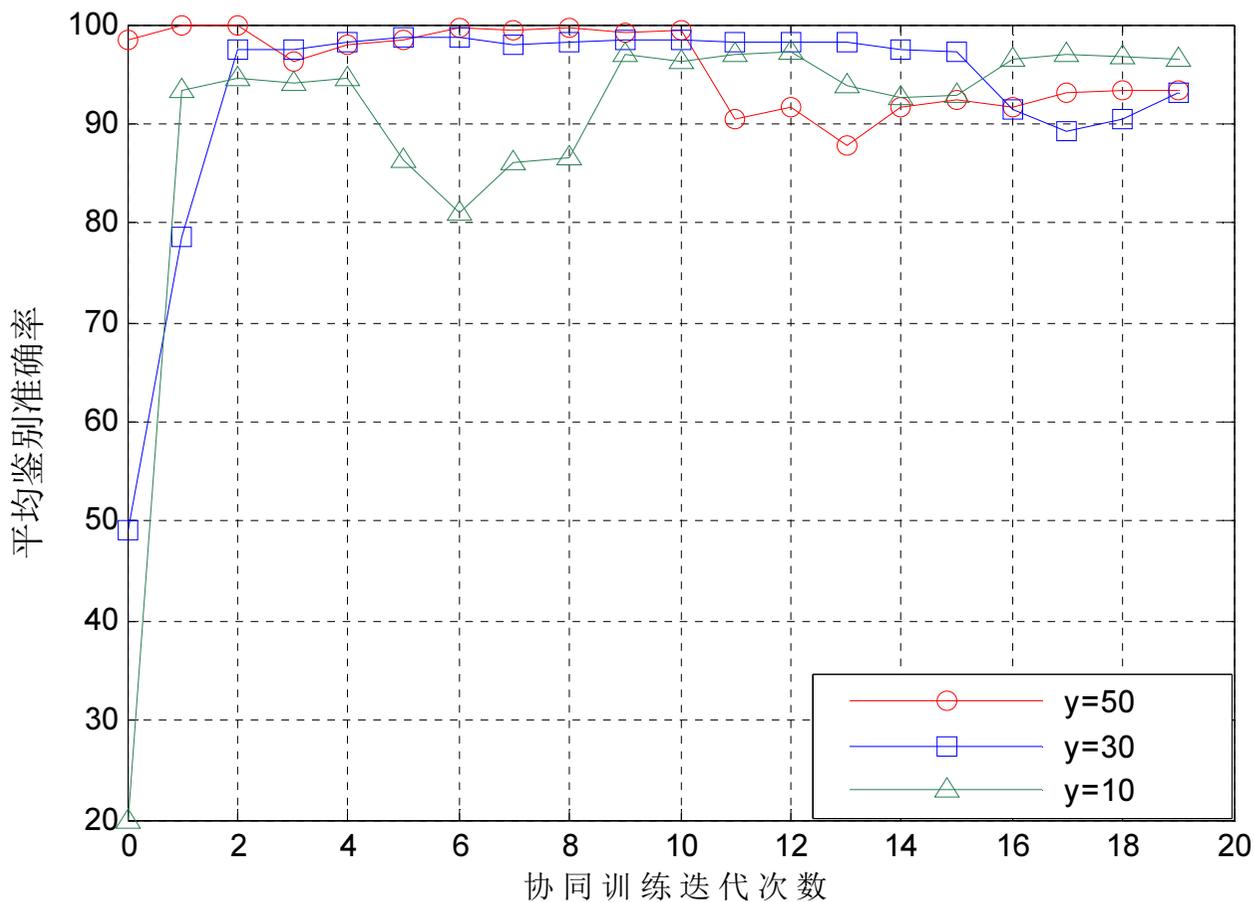
采用LBP特征的协同训练学习方法准确率变化曲线

基于协同训练的相机来源鉴别



采用IQM特征的协同训练学习方法准确率变化曲线

基于协同训练的相机来源鉴别



采用CFA特征的协同训练学习方法准确率变化曲线

总结：

- 本文针对训练样本数不足的情况下相机来源鉴别问题给出了一种新的解决办法
- 将半监督学习模型引入到经典的来源鉴别方法中，实现了对鉴别准确的大幅度提升
- 其中将CFA特征与协同训练相结合的方法在本文中获得了最高的鉴别准确率。

展望：

- 未来可以考虑将现有的半监督学习算法与经典的相机来源鉴别方法进行更为多样的组合，从而寻找出较为稳定且适用于相机来源鉴别的最佳鉴别方法。
- 可以考虑在样本数更为有限的条件下，这种基于半监督学习的相机来源鉴别方法是否依然有效。



Thank You !

Any Question ?