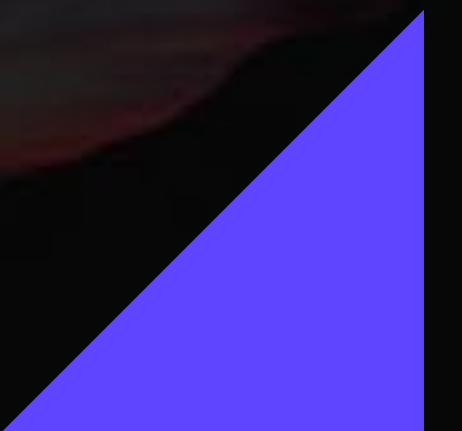


移动直播浪潮下的智能内容审核

胡易

腾讯优图实验室

高级研究员





个人介绍

2016年硕士毕业后，加入腾讯优图实验室，进行计算机视觉方面应用研究，负责研发DeepEye图像内容识别系统，用于腾讯内部业务并通过腾讯云服务大量外部客户。

内容大纲

背景——当前移动直播概况及隐患

传统的解决方案

图像识别技术的发展

DeepEye图像内容识别系统

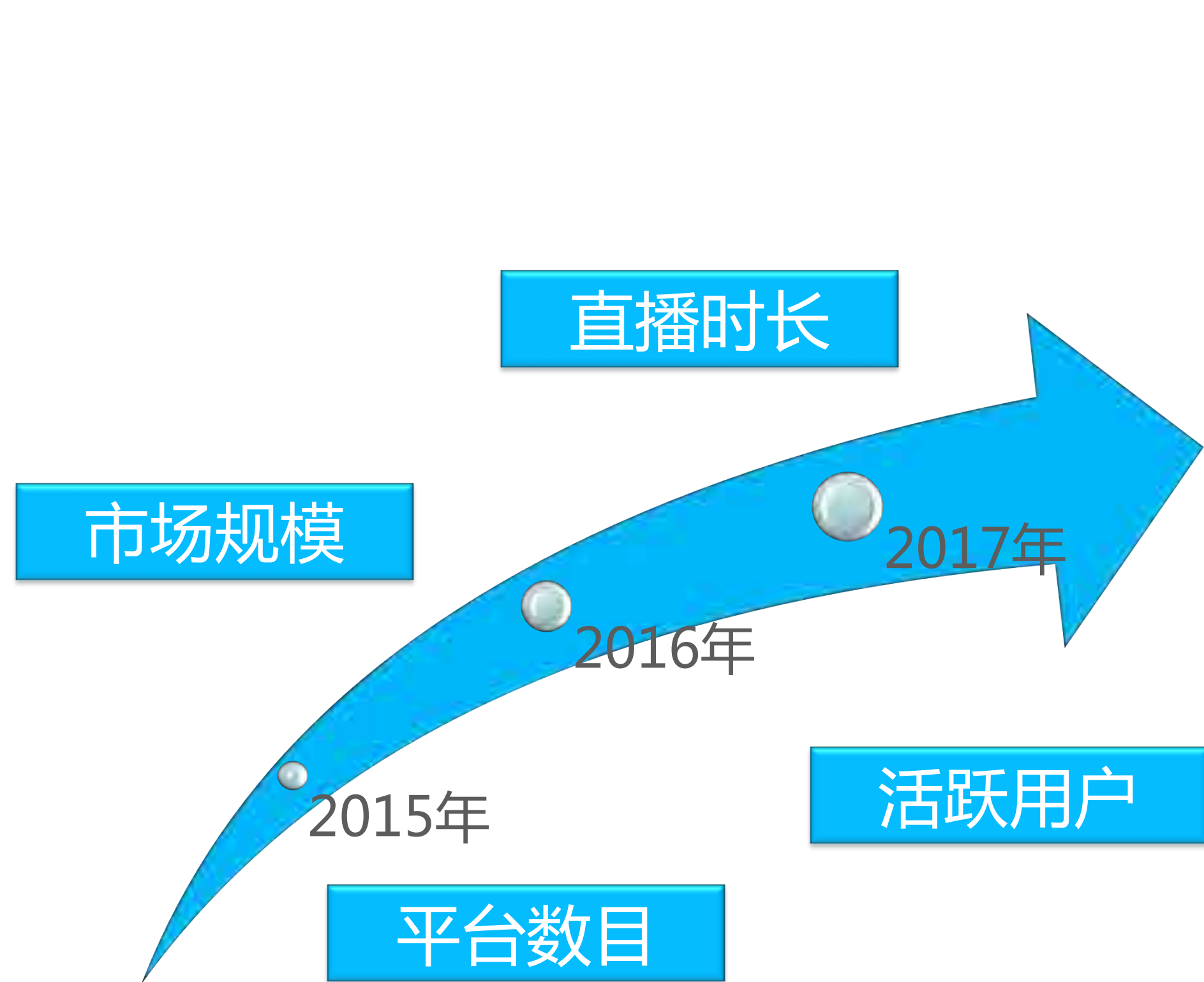
DeepEye系统产生的价值

优图实验室简介

背景——当前移动直播概况



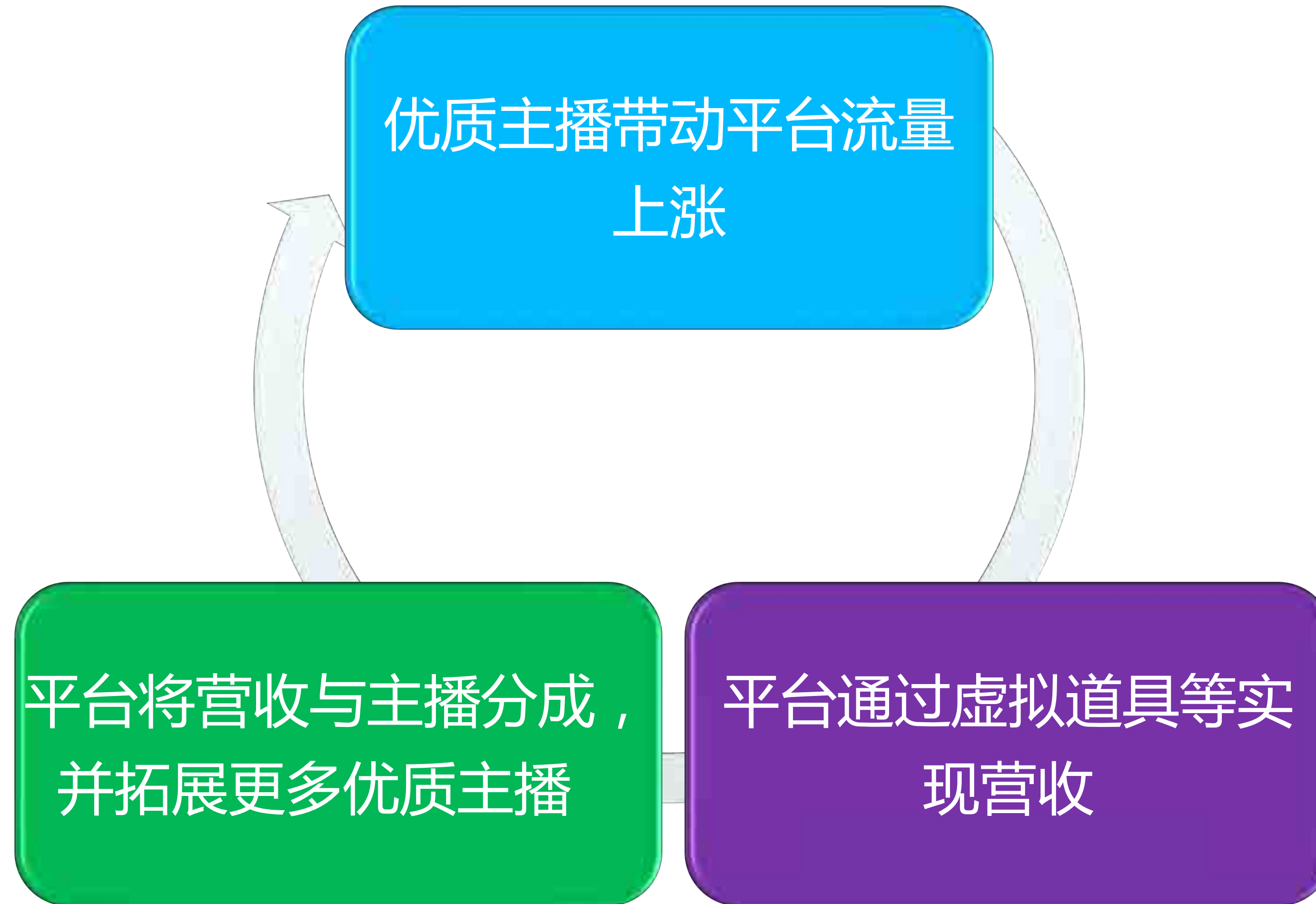
背景——当前移动直播概况及隐患



移动直播平台



背景——当前移动直播概况及隐患



背景——当前移动直播概况及隐患

几起比较恶劣直播事件

直播造人事件

- 2016年1月份，在某直播平台发生了主播直播造人事件，虽然仅仅播出2分钟就被平台掐断，但仍然造成了很大的影响。

直播自杀事件

- 2016年10月，江苏泰州某女主播因患有抑郁症在网上直播自杀，过程异常血腥，幸好警察及时赶到，制止了自杀行为。

直播醉酒飚车


- 2015年12月，某平台著名游戏主播与另外几名主播直播驾车，由于车速过快造成严重交通事故，造成三人受伤。驾车主播被警方刑事拘留。

虚假慈善事件

- 2016年11月，在某直播平台上，有两名男子在直播给云南大凉山地区村名发钱，但在直播结束时又将钱收回，如此虚假慈善行为引起了网民们的愤怒，最终两名主播以诈骗罪被判处三年有期徒刑。

背景——当前移动直播概况及隐患

暴力，色情等违法违规的直播内容不仅挑战社会的公序良俗，也直接损害直播平台本身的利益。

 新浪科技

新浪科技 > 互联网 > 正文

文化部关停12家网络表演平台，

2017年06月29日 11:59 央视新闻

传统的解决方案

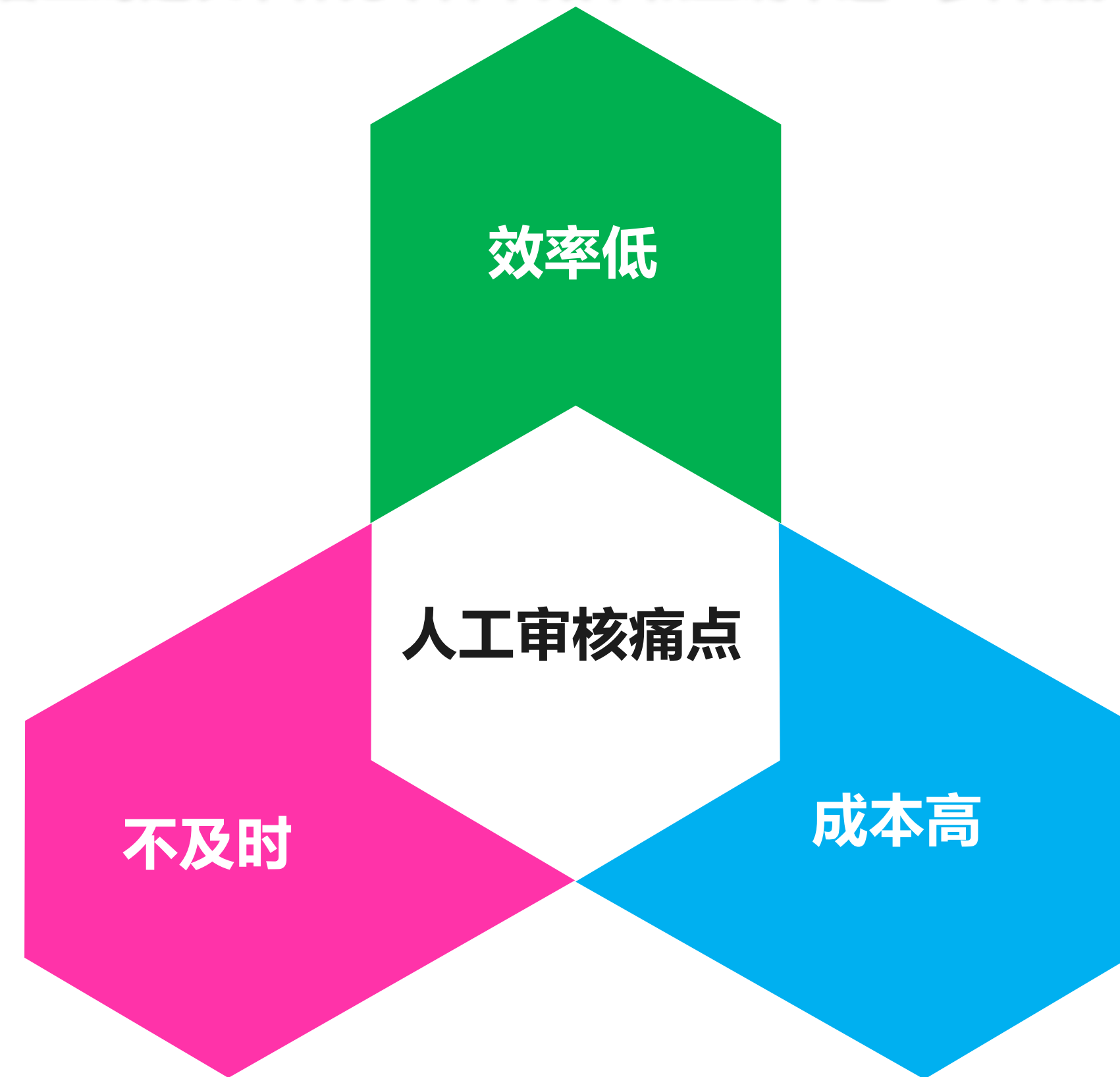


传统的解决方案

人工审核



人工处理图片、视频信息的能力本就远不如机器，随着工时延长，体力下降，效率和正确率进一步降低。



人工很难做到24小时不间断监控，即使发现色情信息，处理也需要时间，无法实时反应。

由于人力市场成本的自然增长和鉴黄师职业的特殊性，人工鉴黄的成本居高不下。

传统的解决方案

基于肤色的色情图像检测



高效

相比人工审核，机器审核效率高，可以24小时不间断审核

效果差

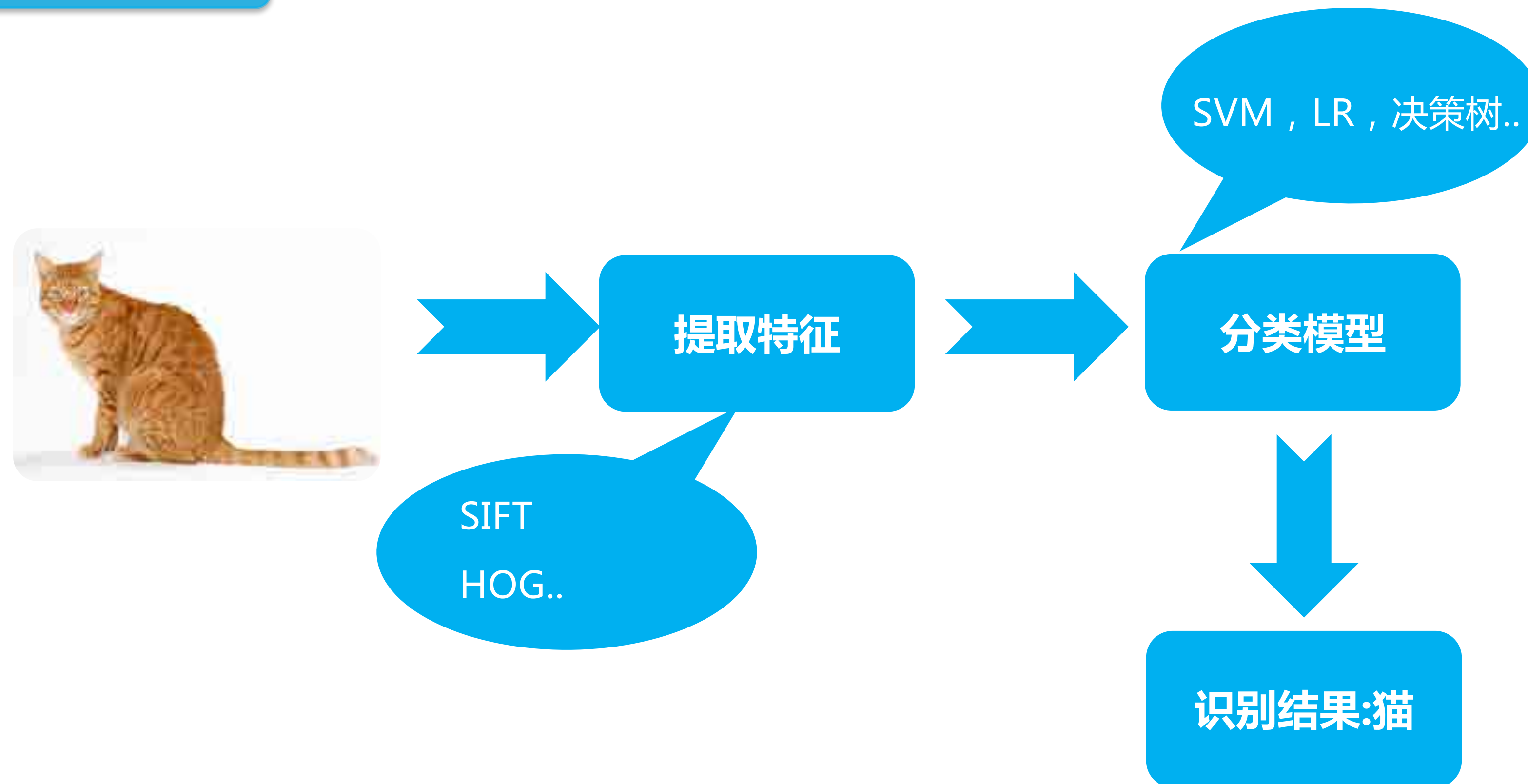
基于肤色的色情图像检测算法效果较差，对于对抗样本处理不好，无法达到实用

图像识别技术的发展



图像识别技术的发展

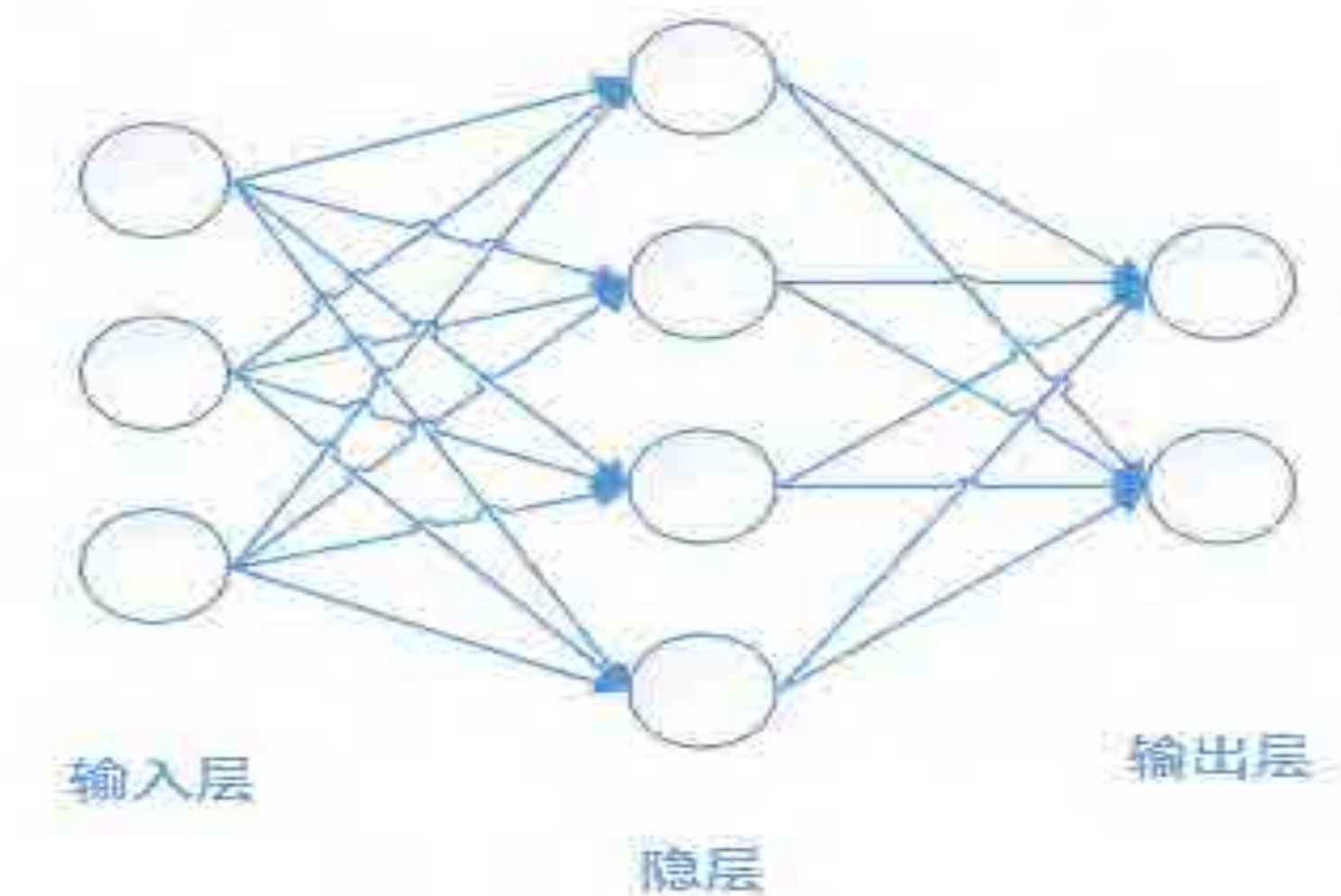
图像识别基本流程



图像识别技术的发展

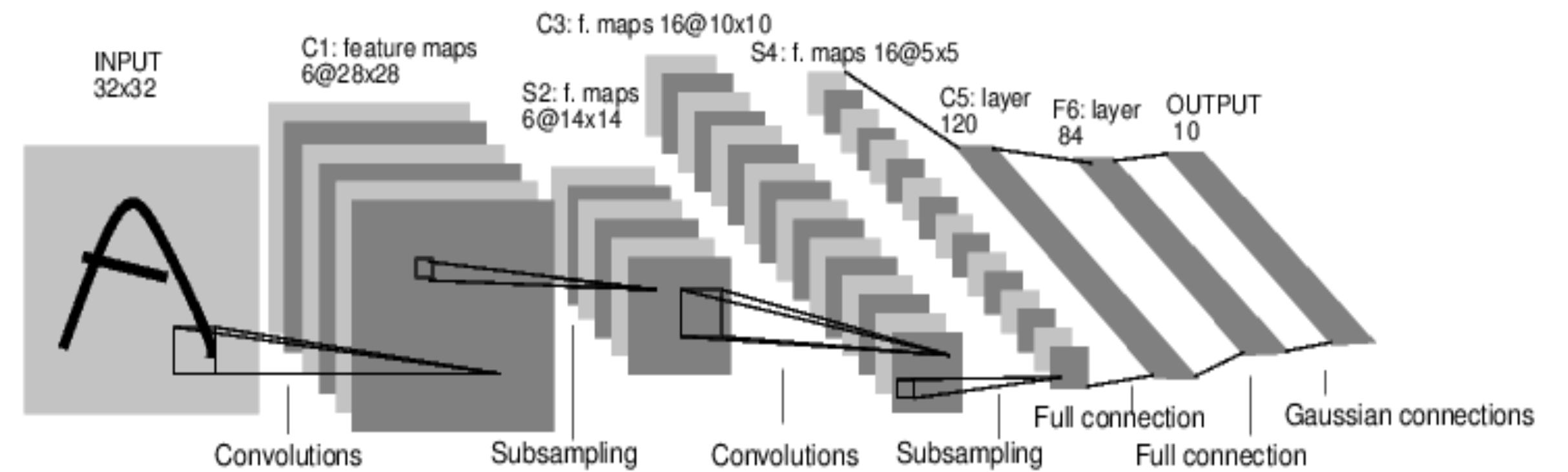
深度学习

深度学习的核心是神经网络，根据不同的任务需要，又分为全连接神经网络(Fully connected neural network)卷积神经网络(Convolutional Network, CNN)循环神经网络(Recurrent Network, RNN)等。



神经网络基本结构

早在上世纪90年代，卷积神经网络就已经应用于图像识别。著名的AI学者Yann LeCun设计出LeNet，用于光学字符识别，取得了很好的效果。但由于计算能力的限制，卷积神经网络在更大规模的图像识别上表现不是很好，一度被学术界放弃。



LeNet网络结构

图像识别技术的发展

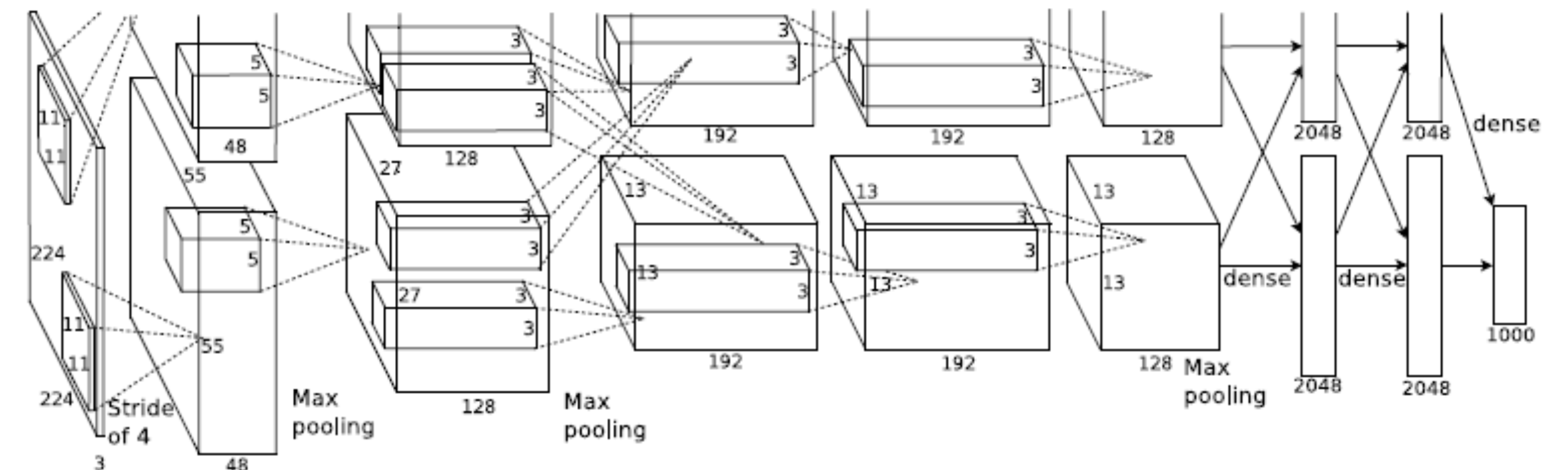
ImageNet图像识别比赛

ImageNet图像识别比赛是由华人AI学者Li Feifei组织发起的，该比赛的主要任务是对1000个图像类别进行分类，它提供了超过一百万张图片的训练集，参赛者基于这一百万张训练图片训练分类模型，然后在测试集上测试模型效果，按top5错误率从小到大排名



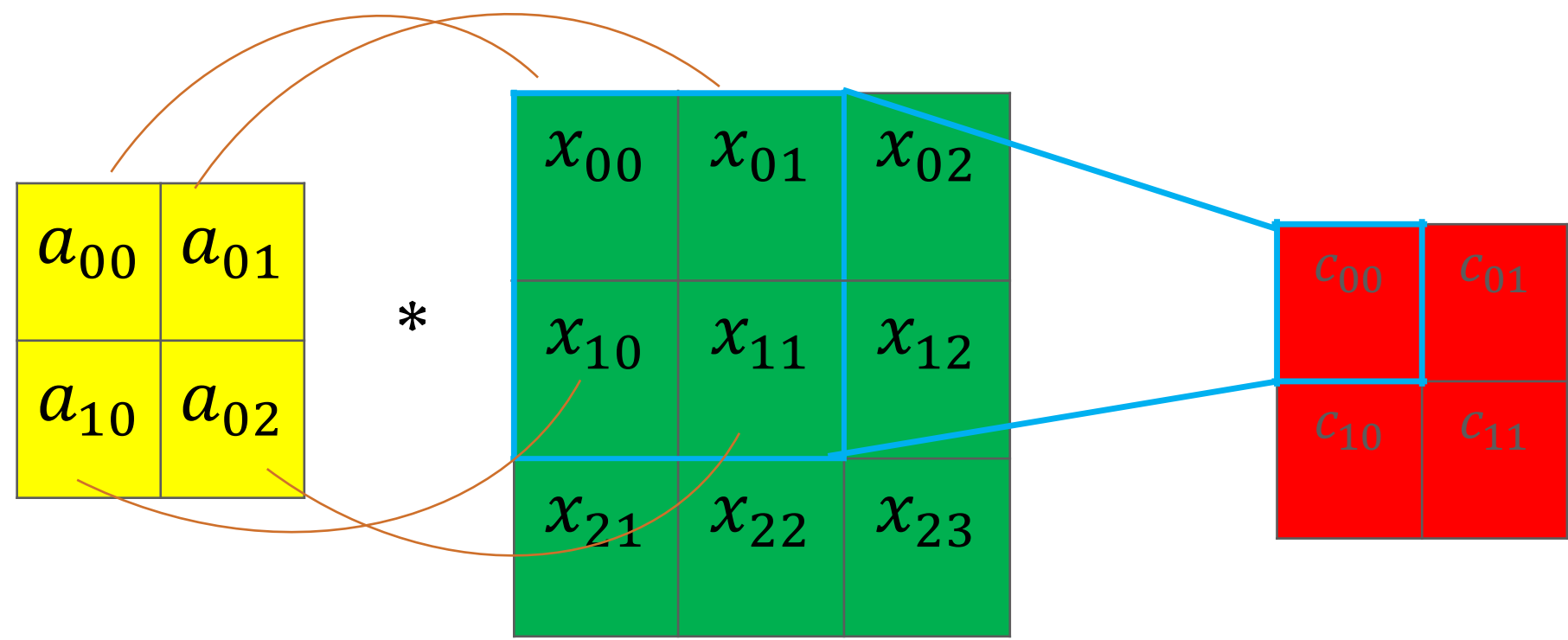
卷积神经网络的重新崛起

2012年，多伦多大学Hinton教授和他的学生Alex在ImageNet大规模图像识别比赛中，使用了深度卷积网络AlexNet，最终以16.4%的top5错误率获得第一名，震惊了学术界。



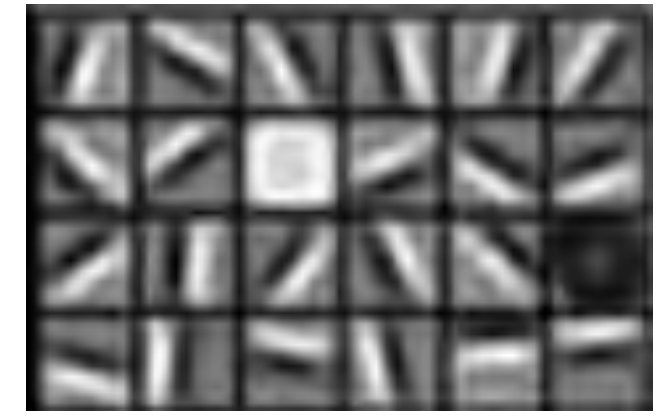
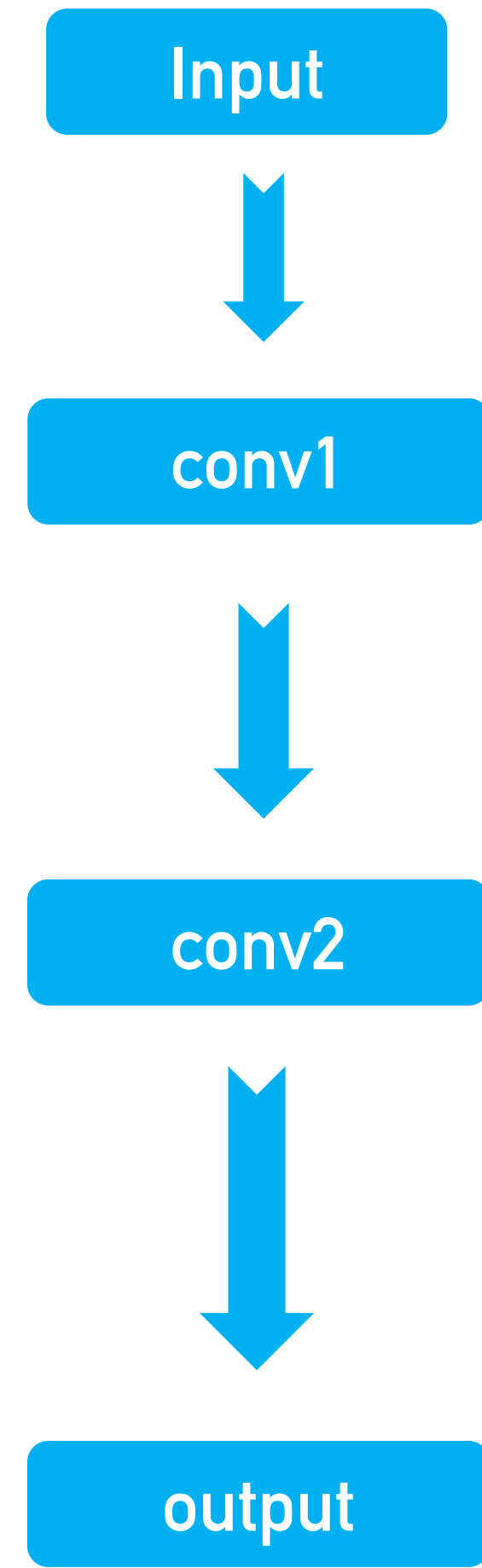
图像识别技术的发展

卷积神经网络基本原理



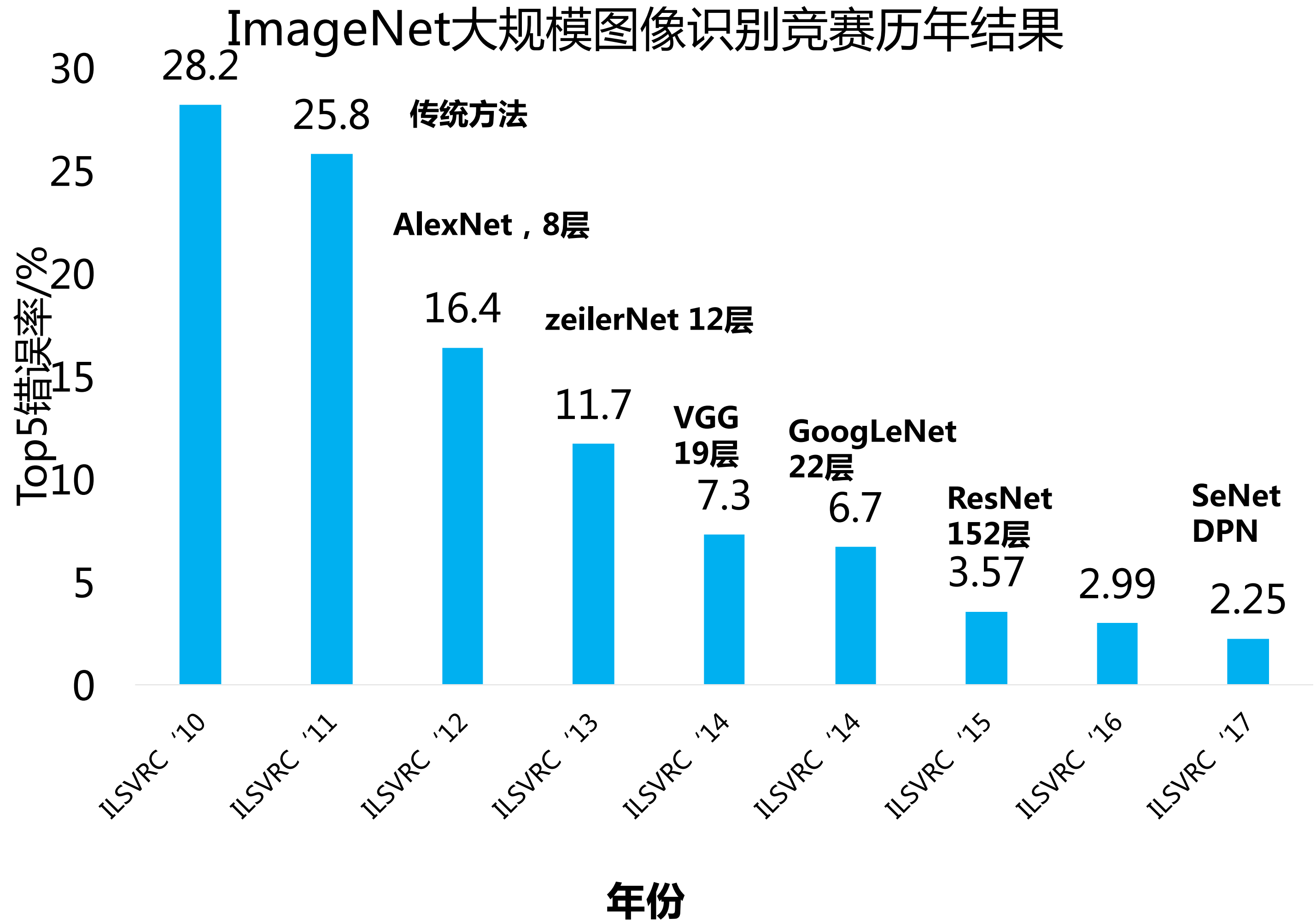
$$c_{00} = a_{00}x_{00} + a_{01}x_{01} + a_{10}x_{10} + a_{11}x_{11}$$

卷积操作示意图



图像识别技术的发展

近几年卷积网络发展历程



图像识别技术的发展

卷积网络结构演变



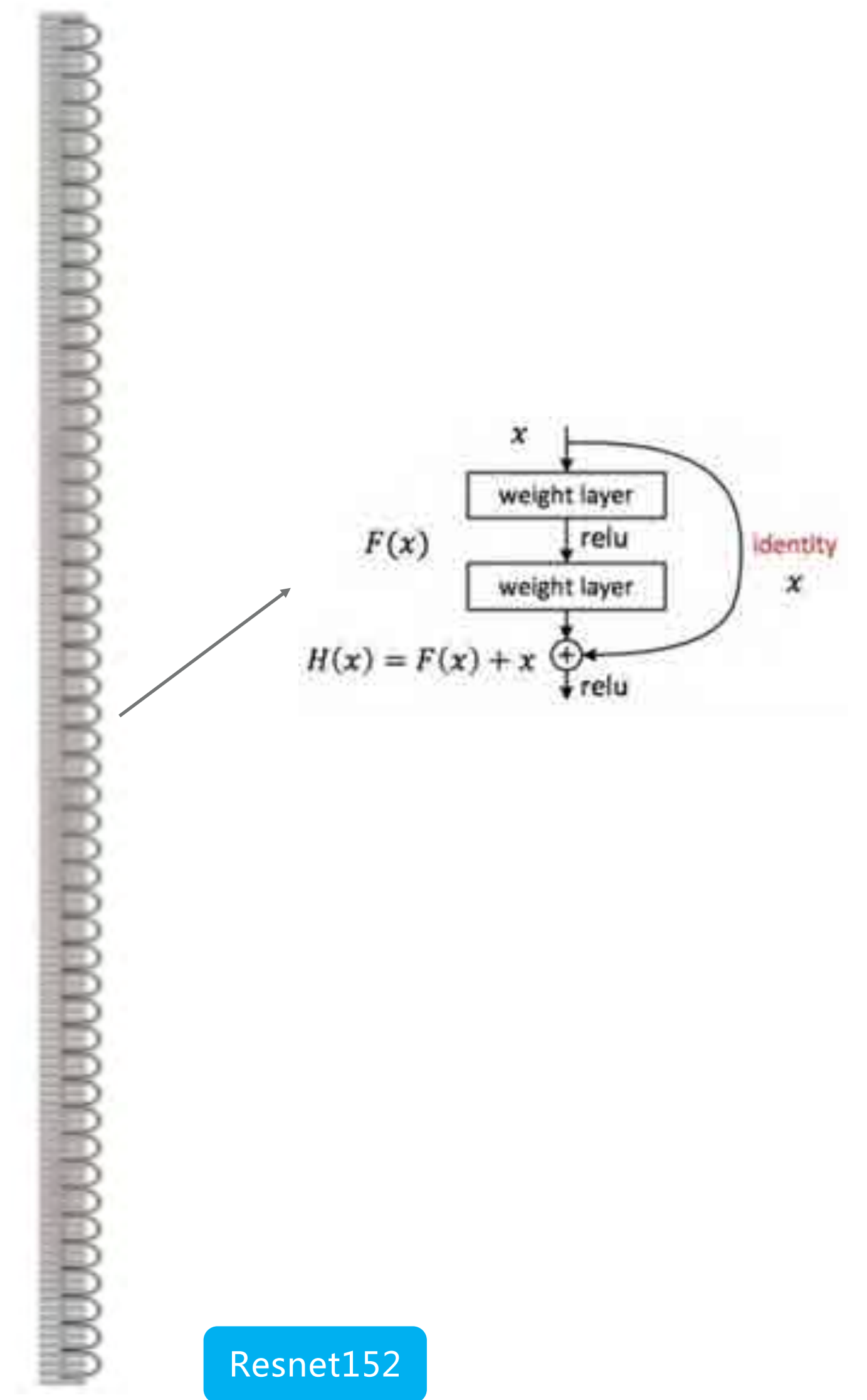
Alexnet



VGG19



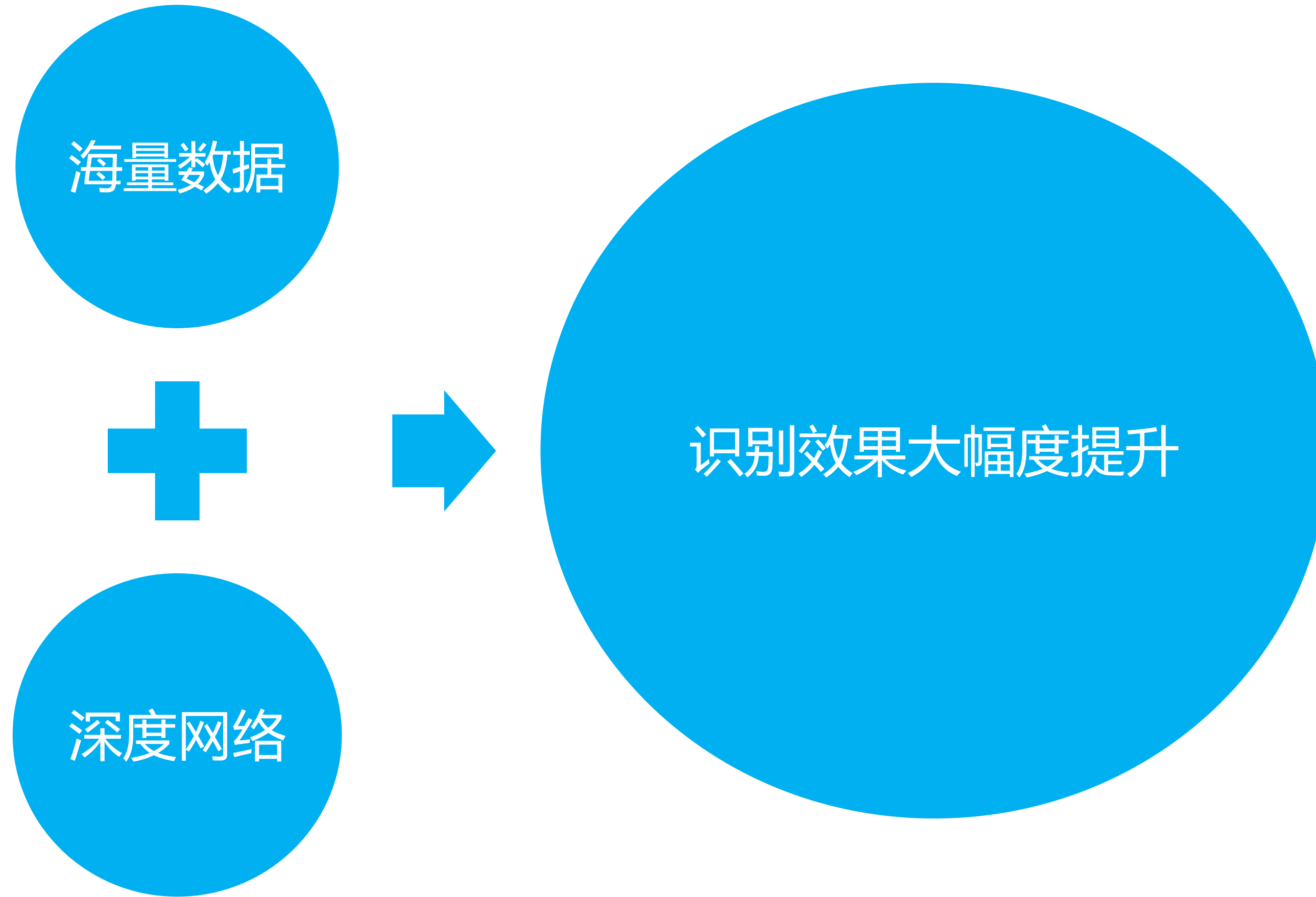
GoogleNet



Resnet152

图像识别技术的发展

深度学习



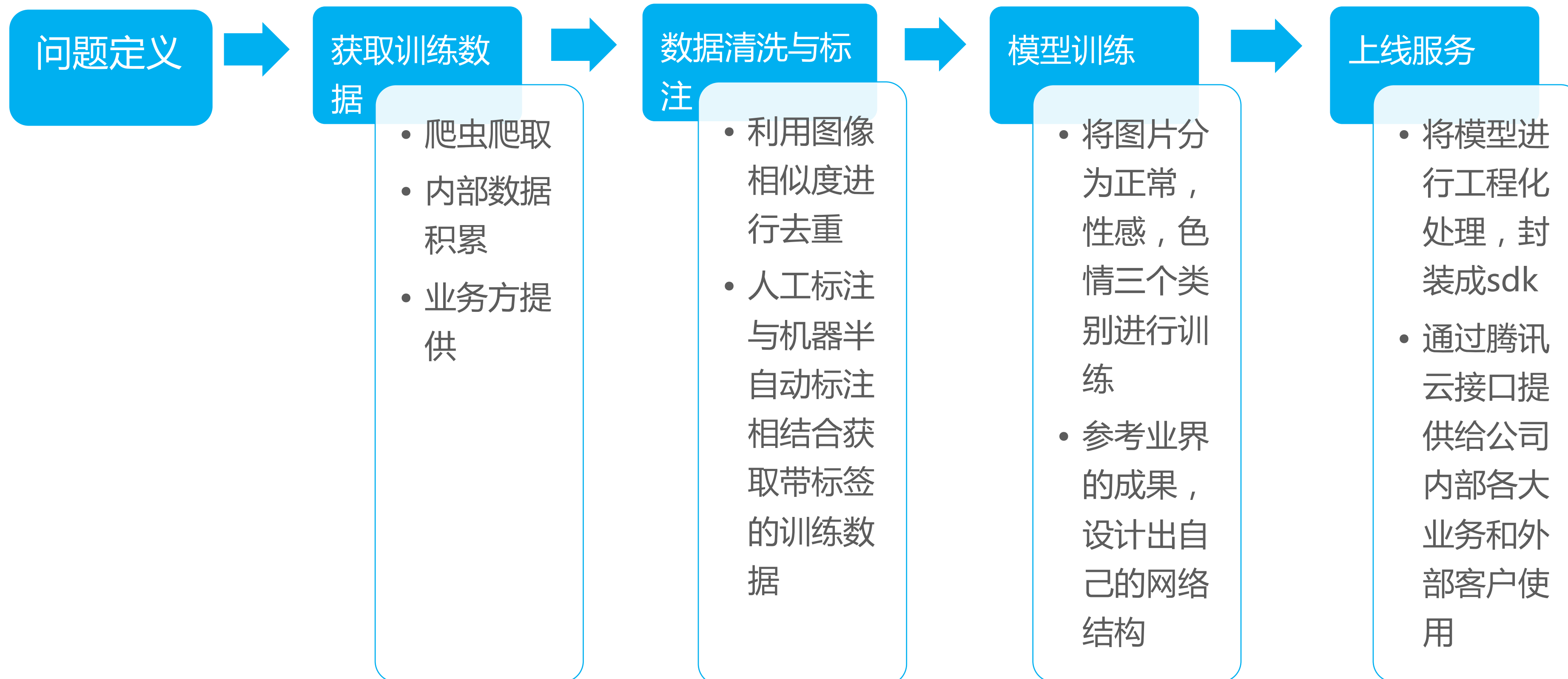
Deepeye图像内容识别系统



Deepeye图像内容识别系统

从鉴黄入手

为应对网络上日益增长的有害图片，优图实验室联合腾讯内部其他相关部门开发了Deepeye图像内容识别系统，从鉴黄这一老大难问题入手，再逐步扩展出其他能力，缓解图片审核压力。



Deepeye图像内容识别系统

问题定义

将图片鉴黄定义为一个分类问题，将图片分为正常，性感，色情三个类别，收集训练数据训练卷积网络，实现对图片的分类



色情



性感



正常

Deepeye图像内容识别系统

数据准备

多渠道采集



- 爬虫工具从网络上爬取图片
- 内部积累的图片
- 业务方提供的图片

数据清洗



- 去除重复相似图片
- 去重特殊格式图片

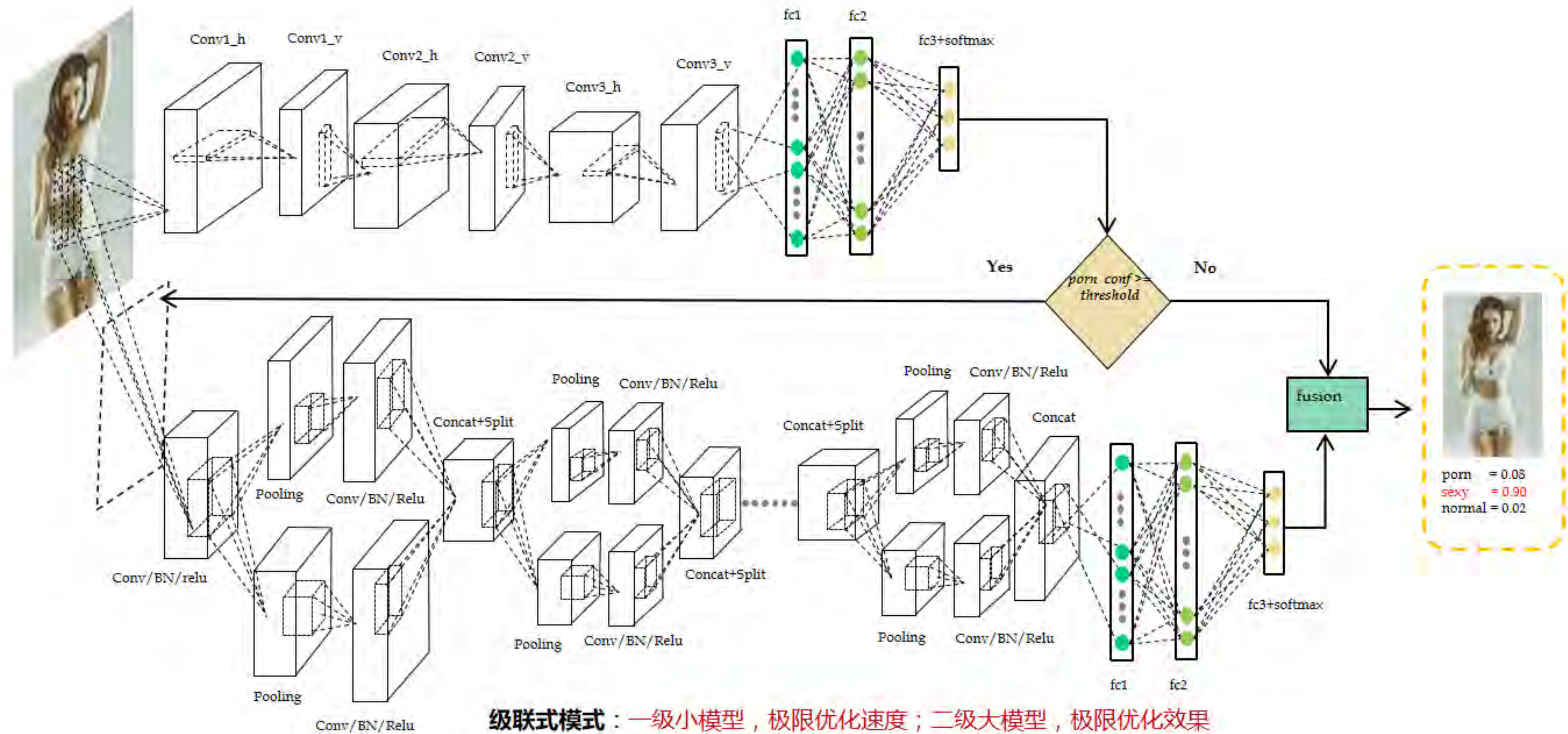
数据标注

- 制定标注规则，人工进行标注
- 依靠少量数据训练得到的模型对未标注图像进行预测，半自动标注

Deepeye图像内容识别系统

两级级联结构

采用两级级联结构，兼顾速度与效果

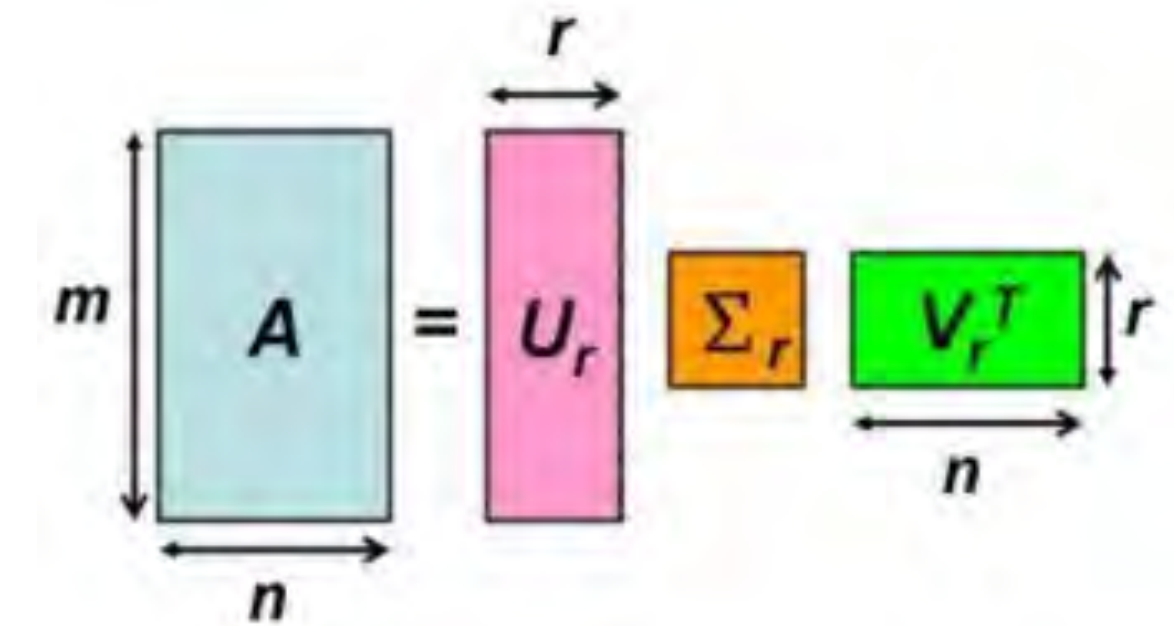
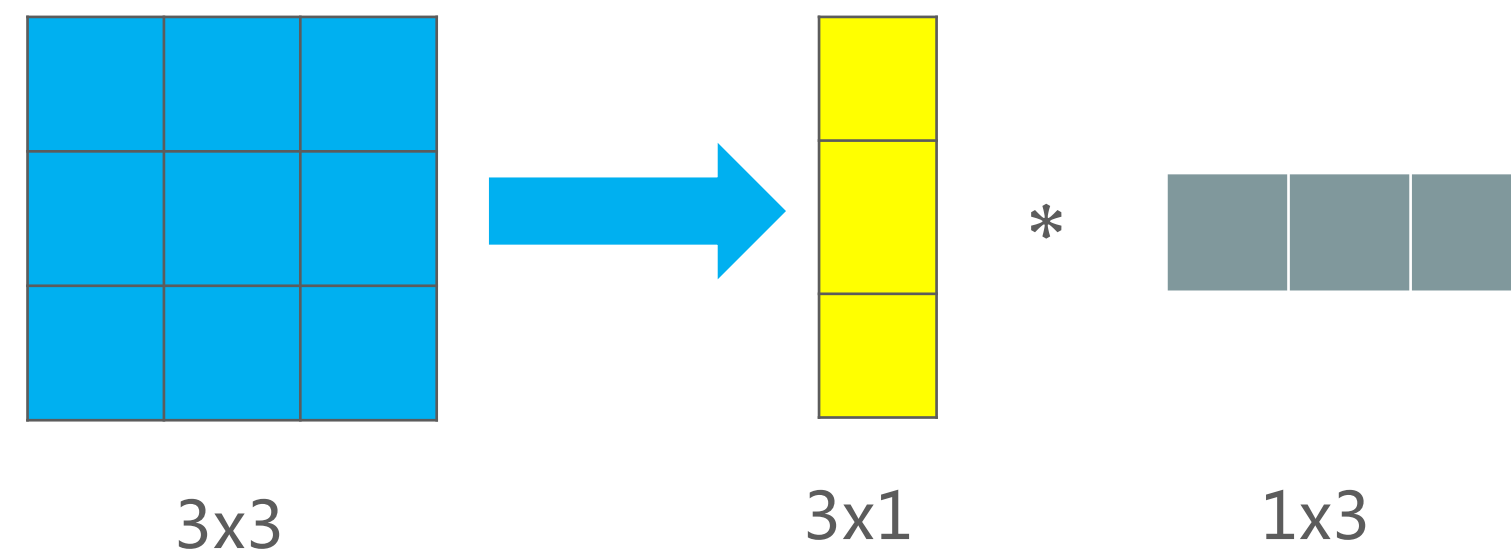


Deepeye图像内容识别系统

性能优化

- 设计更高效的第一级模型Deepsmart，用于快速拒绝非色情图片

- 卷积核拆分，全连接层SVD分解

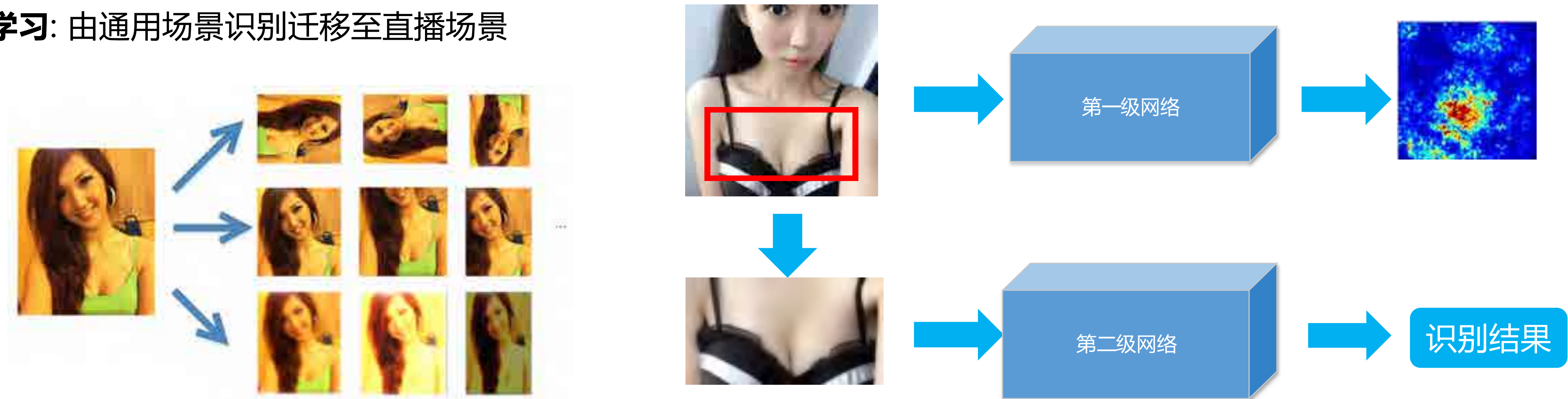


- 模型压缩：对网络进行剪枝，去除冗余节点，与厦门大学合作，产出论文《Towards Convolutional Neural Networks Compression via Global Error Reconstruction》发表于IJCAI2016

Deepeye图像内容识别系统

效果优化

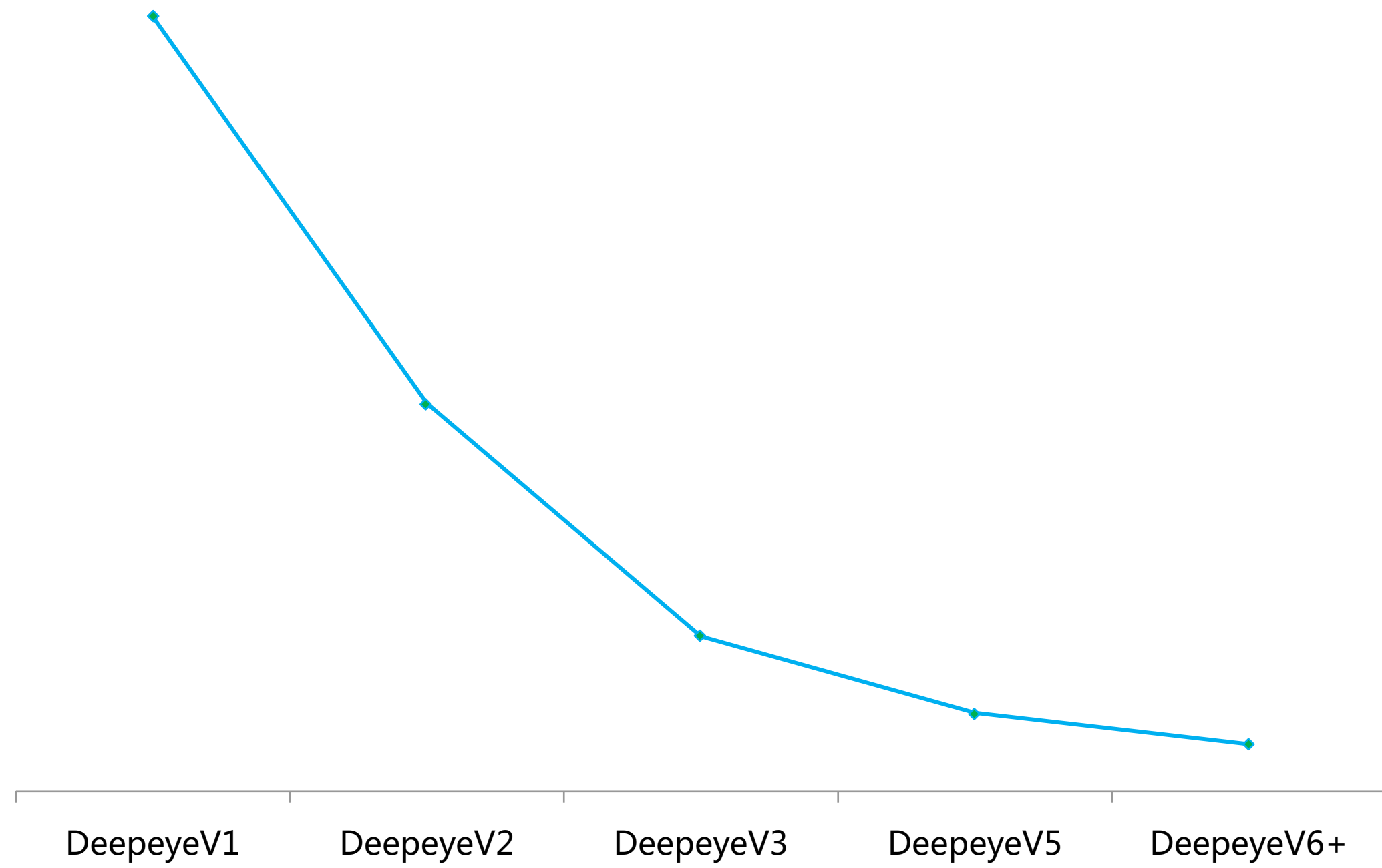
- **数据扩增:** 随机裁剪, 旋转, 模糊处理, 提升模型鲁棒性
- **注意力机制:** 第一级模型发现潜在色情区域, 第二级模型针对潜在区域进行精准识别
- **针对性补充数据:** 根据业务方反馈的数据, 分析其特点, 对训练集进行针对性补充。
- **迁移学习:** 由通用场景识别迁移至直播场景



DeepEye图像内容识别系统

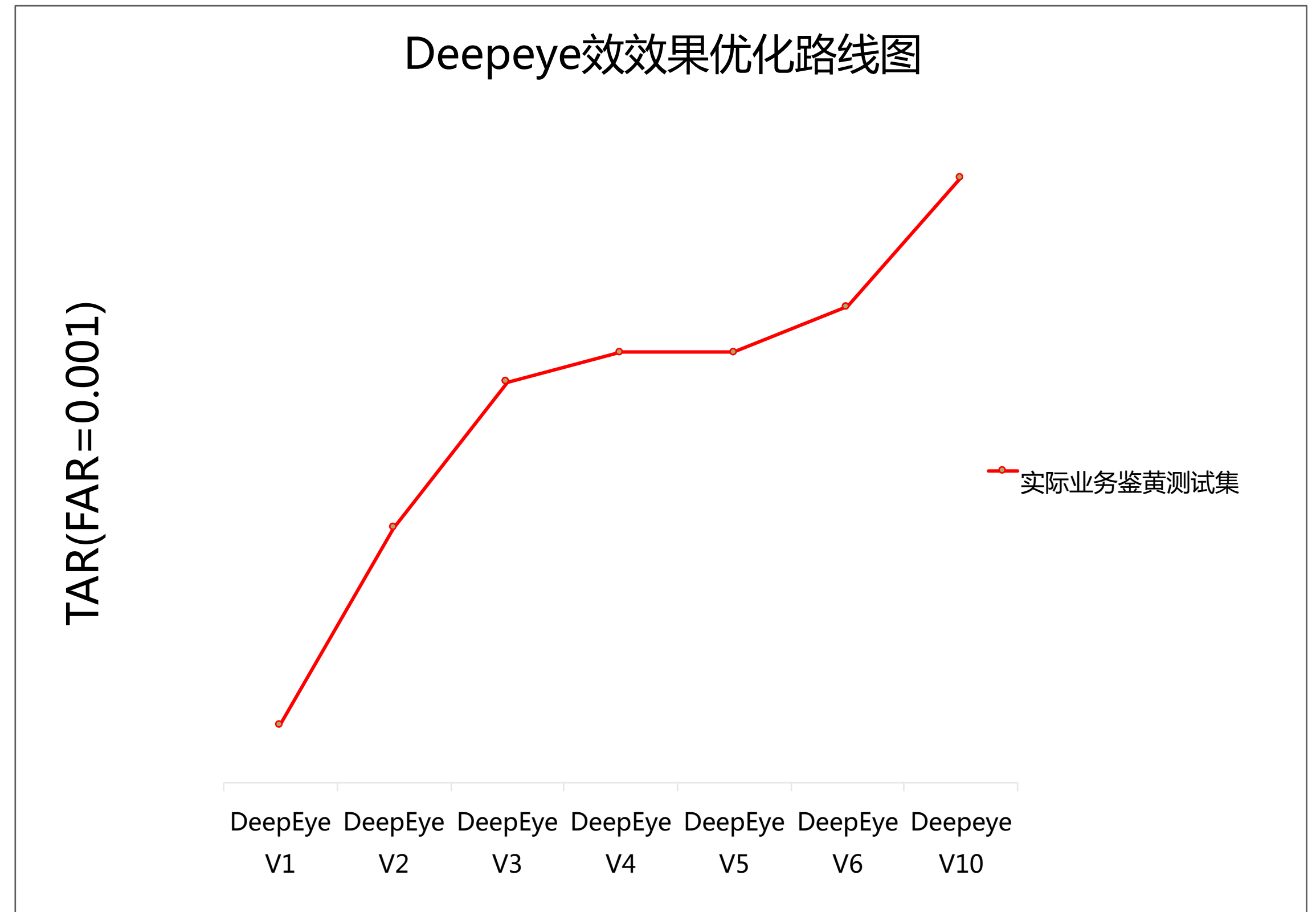
时间性能和效果优化成果

DeepEye性能优化路线图



时间优化曲线

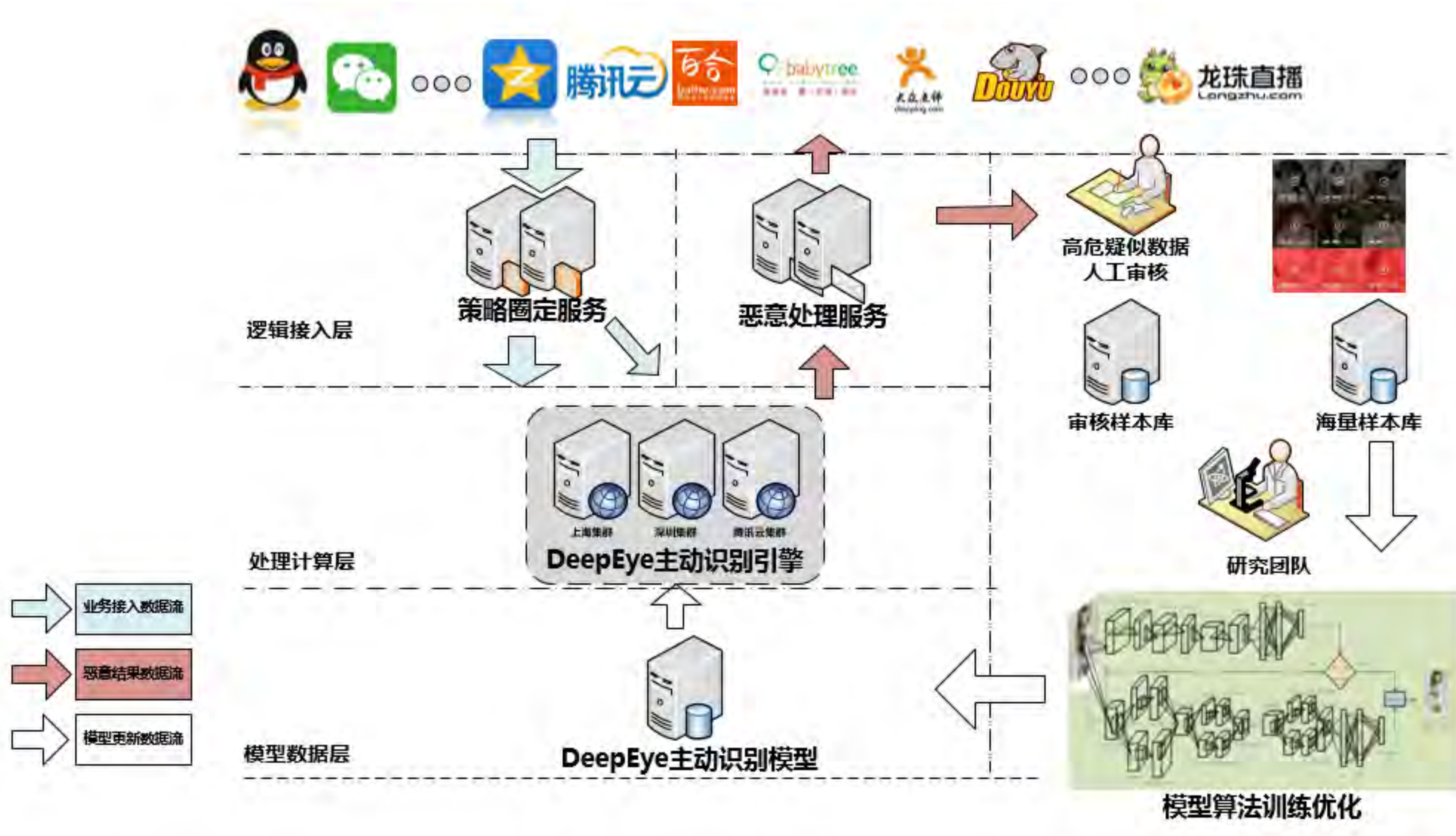
DeepEye效果优化路线图



效果优化曲线

Deepeye图像内容识别系统

Deepeye系统架构



Deepeye图像内容识别系统

能力扩展——细化类别

将色情和性感图片类别更加细化，方便业务方更加灵活的制定打击策略，采用多分支联合训练，提升效果。

凸显胸部



凸显腿部



半身裸露

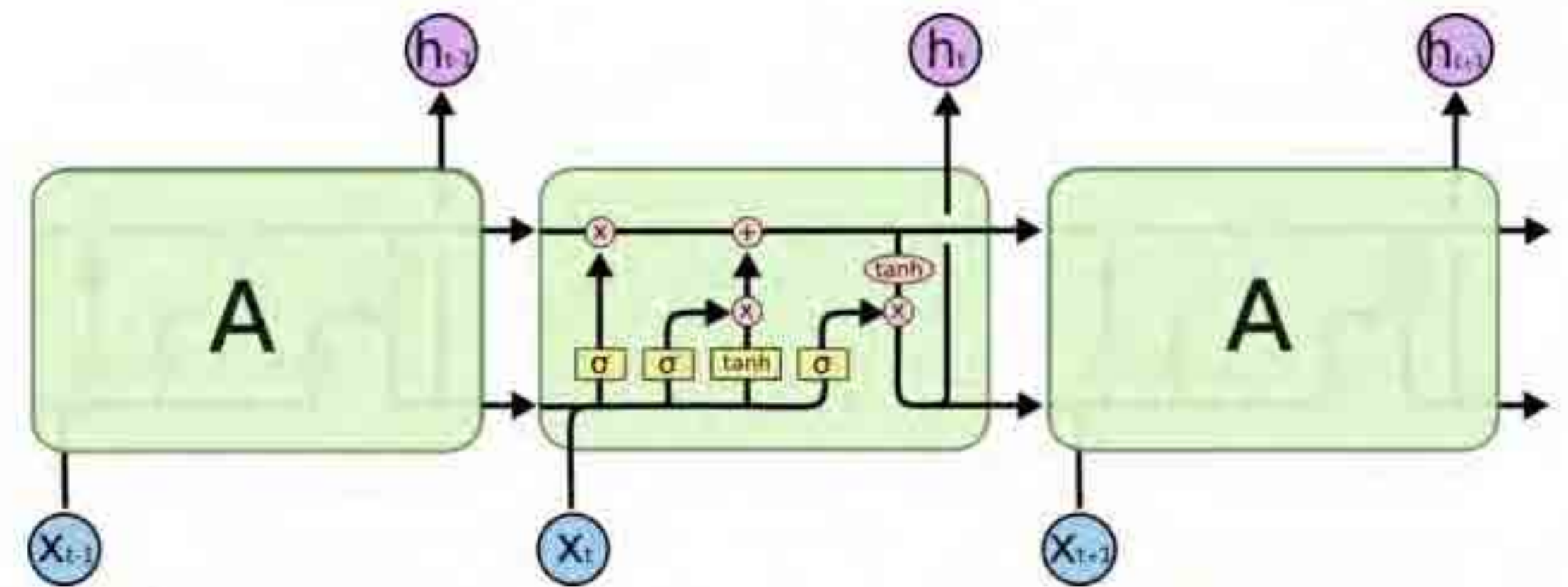
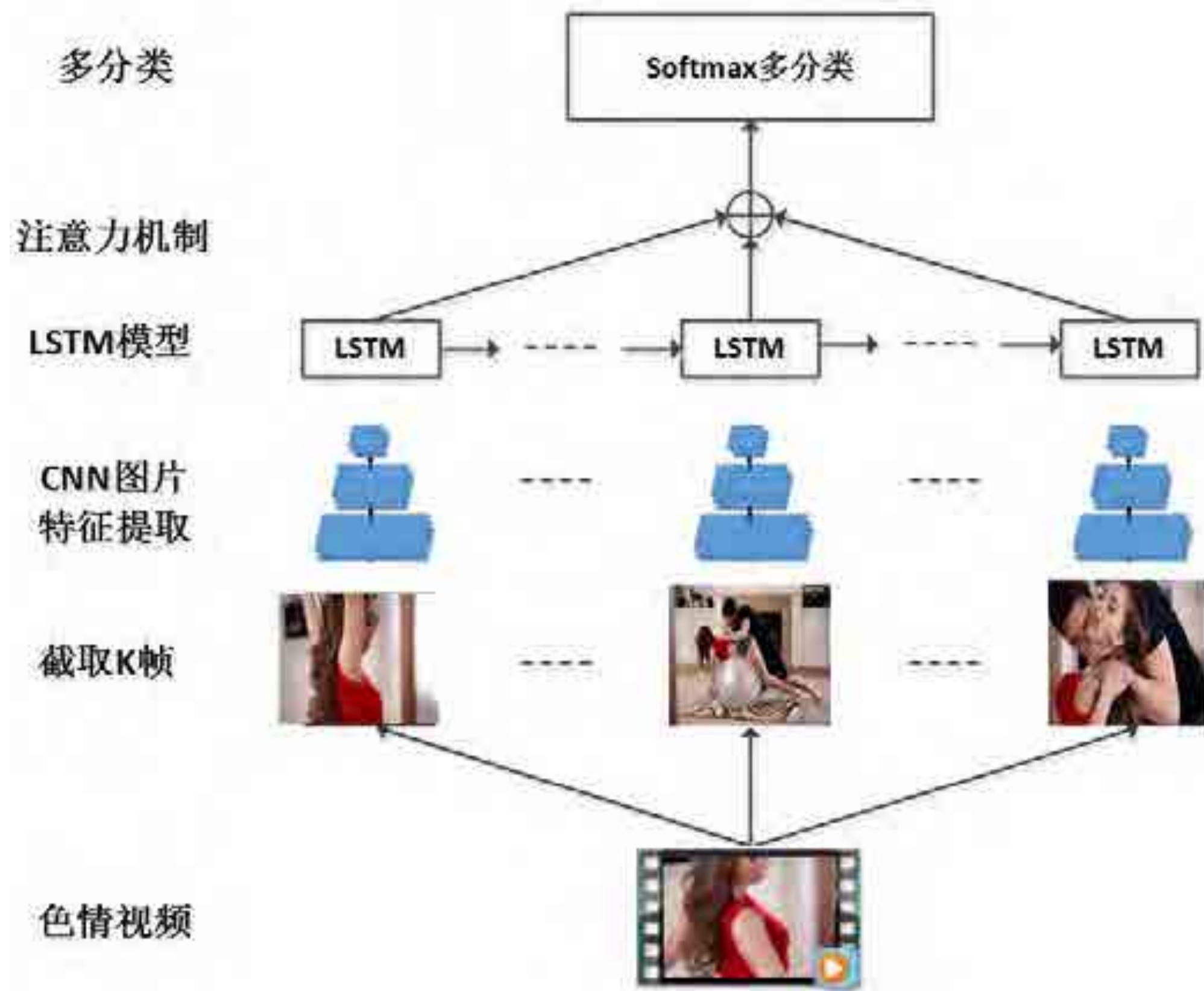


非真实人物画面



Deepeye图像内容识别系统

能力扩展——视频内容识别



Deepeye图像内容识别系统

能力扩展——暴恐内容识别

采用多标签分类，识别暴恐类型图片，一张图片可以有多个标签，具体包括以下类别



武装分子



刀具



枪支



火灾



血腥



极端主义旗帜



人群聚集

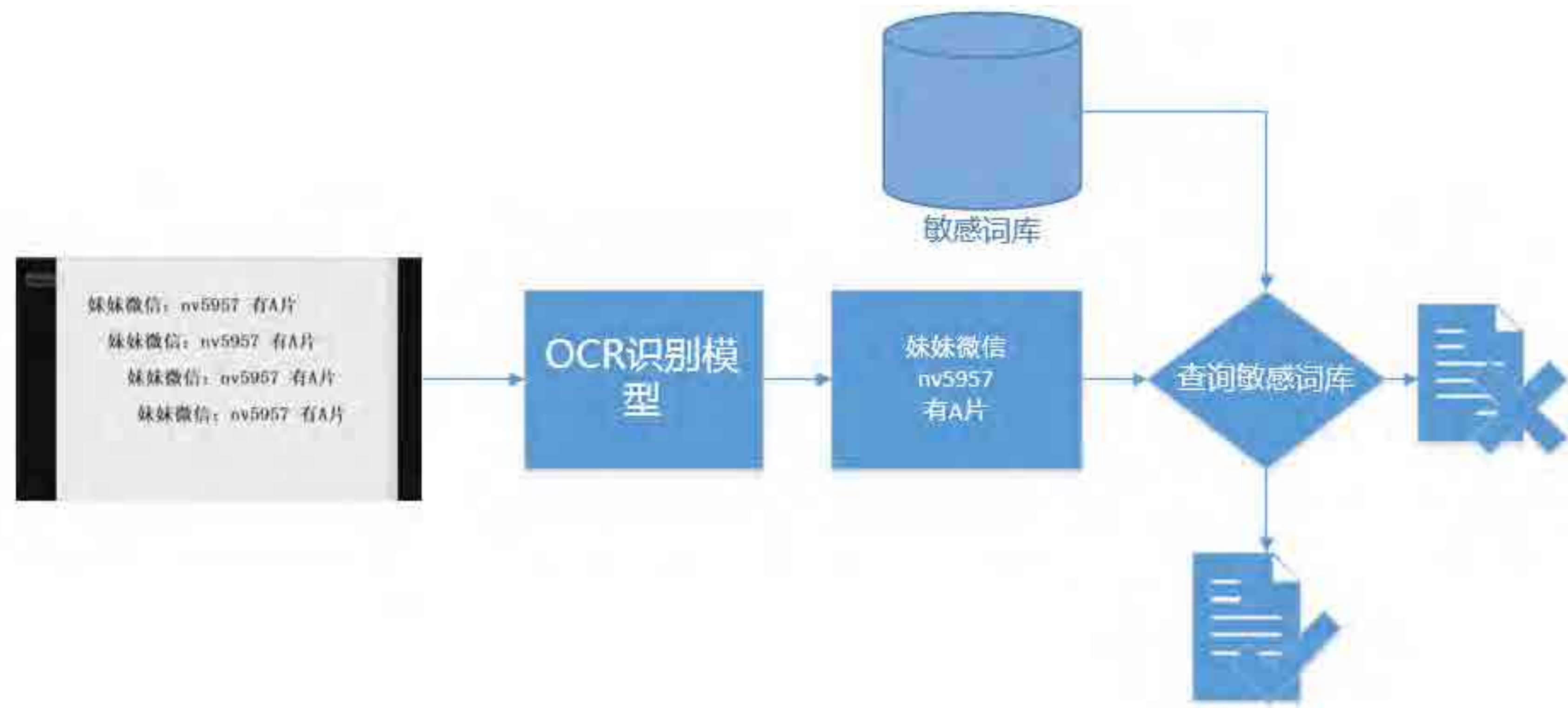


大型武器装备

Deepeye图像内容识别系统

能力扩展——OCR敏感词识别

识别图片中的文字，可用于打击招嫖广告，政治敏感内容



Deepeye图像内容识别系统

能力扩展——涉政人物识别

识别图片中的人物是否为政治敏感人物，降低监管风险



Deepeye图像内容识别系统

能力扩展——主播吸烟识别

识别出主播是否在吸烟，给予警告或封禁

