图片审核方案

由 杨智帆创建, 最后修改于四月 09, 2019 00:03

0. 版本历史

序号	版本	修改人	修改时间	备注
1	V0.1	杨智帆	2019.03.25	初版

1. 背景

供应商在商家平台上传图片过程中可能存在图片内容不合规问题，如非配件图片被上传、配件图片若包含敏感词、竞品水印等敏感信息，则一方面会使企业面临违法违规经营风险，另一方面也影响企业形象。目前供应商上传的配件图片类型众多且数量较大，靠人工审核，费时费力。

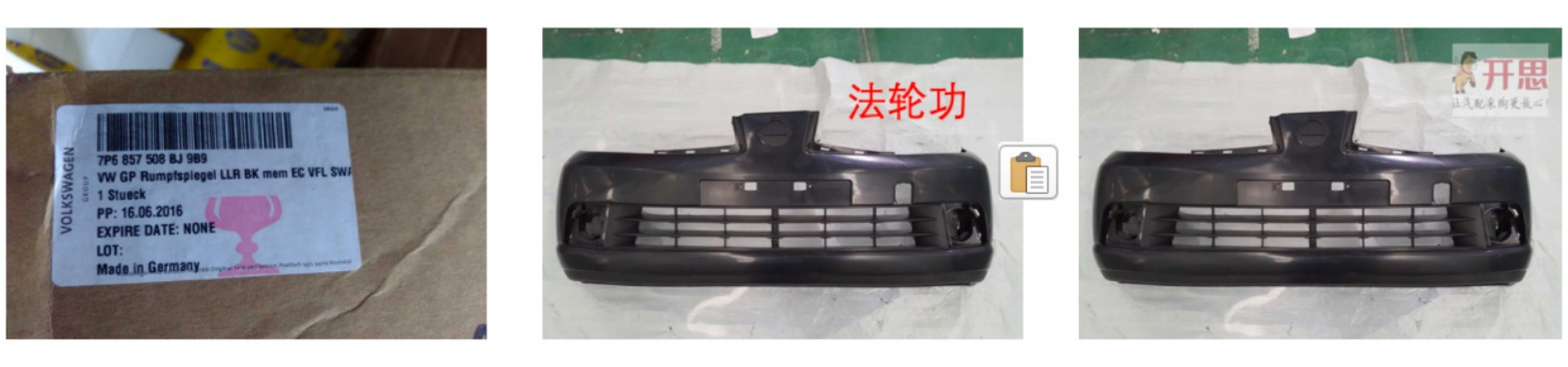


图1 不合规图片

2. 需求

对图片进行智能审核，避免内容不合规图片进入平台中

3. 方案简介

首先对图片进行分类，将其分类为配件图片和非配件图片，其次对配件图片进行敏感信息识别，筛选出含有敏感信息的图片，然后将不合规图片（非配件图+含敏感信息图）退回，要求供应商重新上传，最后将合规图片上传至商家平台中。图1为方案流程图。

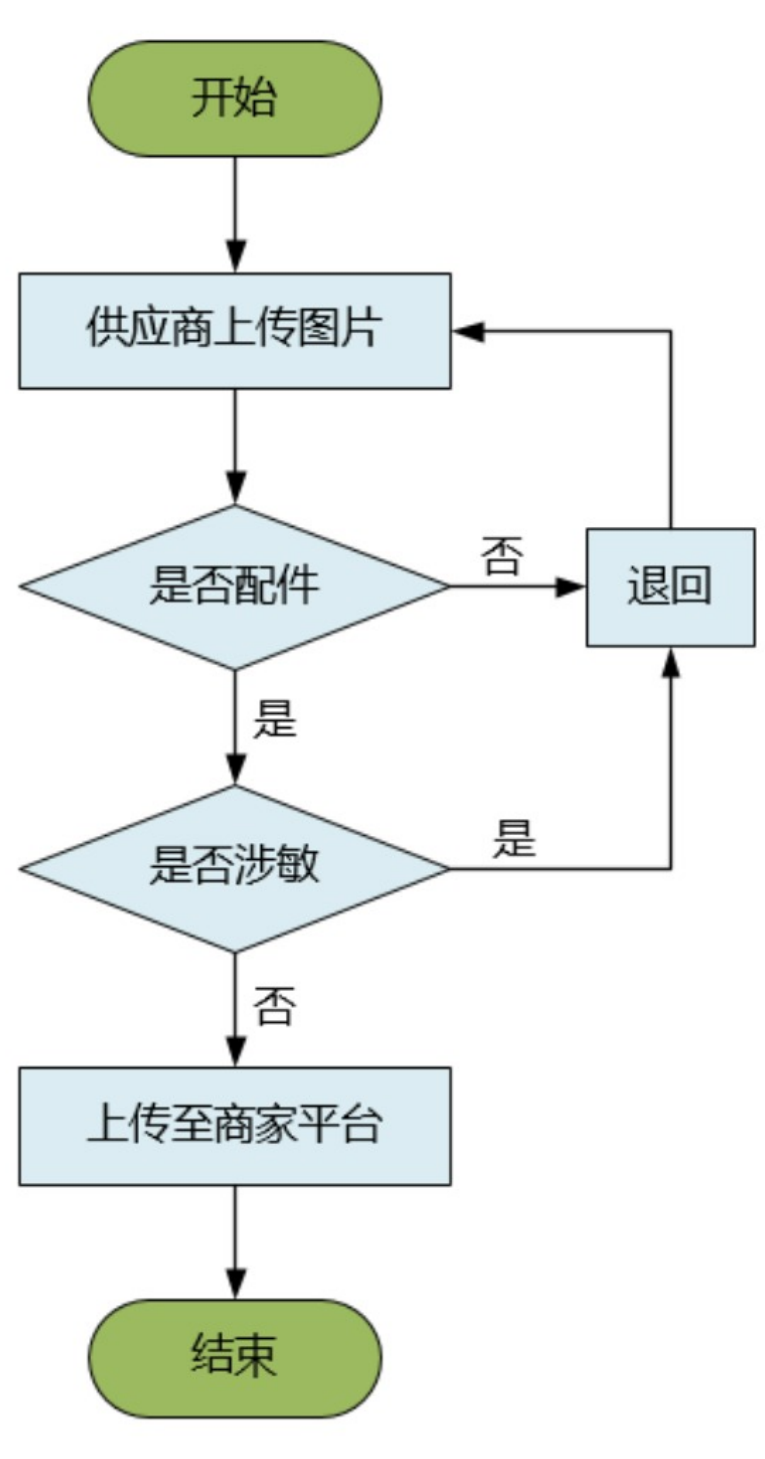


图2 图片审核流程

4. 算法开发

本方案的实现需要开发图片分类算法和敏感信息判别算法，前者判断图片是否配件，后者判断配件图片是否涉敏。

4.1 图片分类算法

目前，基于**深度卷积神经网络**（Deep Convolutional Neural Networks, **CNNs**）建立的分类模型，可以组合低层特征形成更加抽象的高层特征/属性表示，分类准确率高，在图像识别领域中应用广泛。这里，我们基于CNNs建立分类模型，通过训练使其能够对图片进行识别，将图片分为“配件”与“非配件”两类，从而实现图片分类算法，其算法开发流程如下：

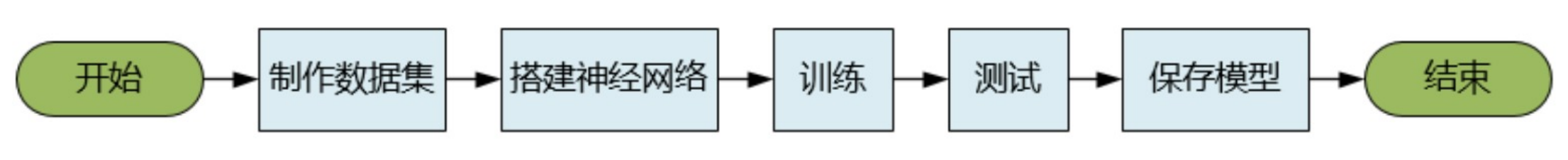


图3 图片分类算法开发流程

4.1.1 数据集

4.1.1.1 组成

数据集不仅需要**标准配件图片**让模型学习汽车配件的特征，也需要**非配件图片**提高模型的泛化能力，图片数量越多，质量(图片内容单一且突出、轮廓清晰等)越高，越有利于提高图片分类算法正确率。随机分为训练集、验证集与测试集，比例为6 : 2 : 2。

4.1.1.2 来源

配件图片：供应商上传的配件图片(4500张)

含敏感信息的配件图片(自制，500张)

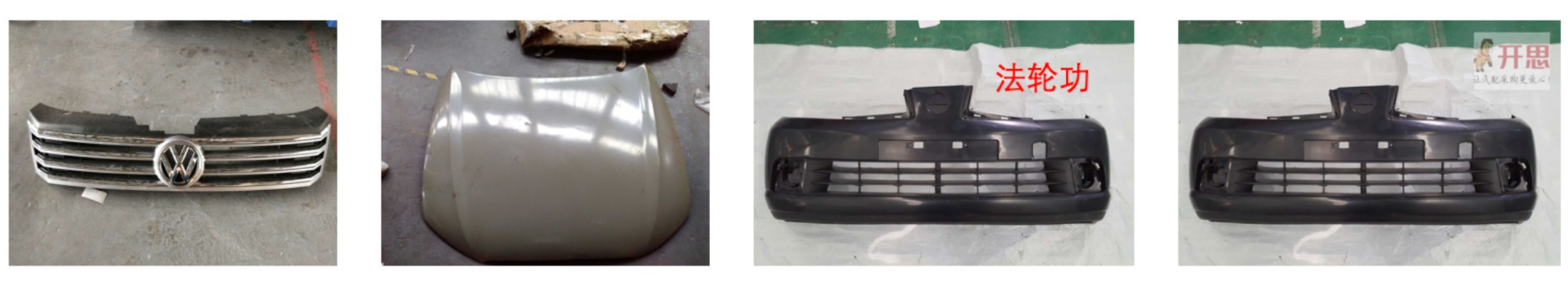


图4 配件图片

非配件图片：供应商上传图片中筛选出的不合规图片(包裹图)(500张)

ImageNet图库中采集(4500张)

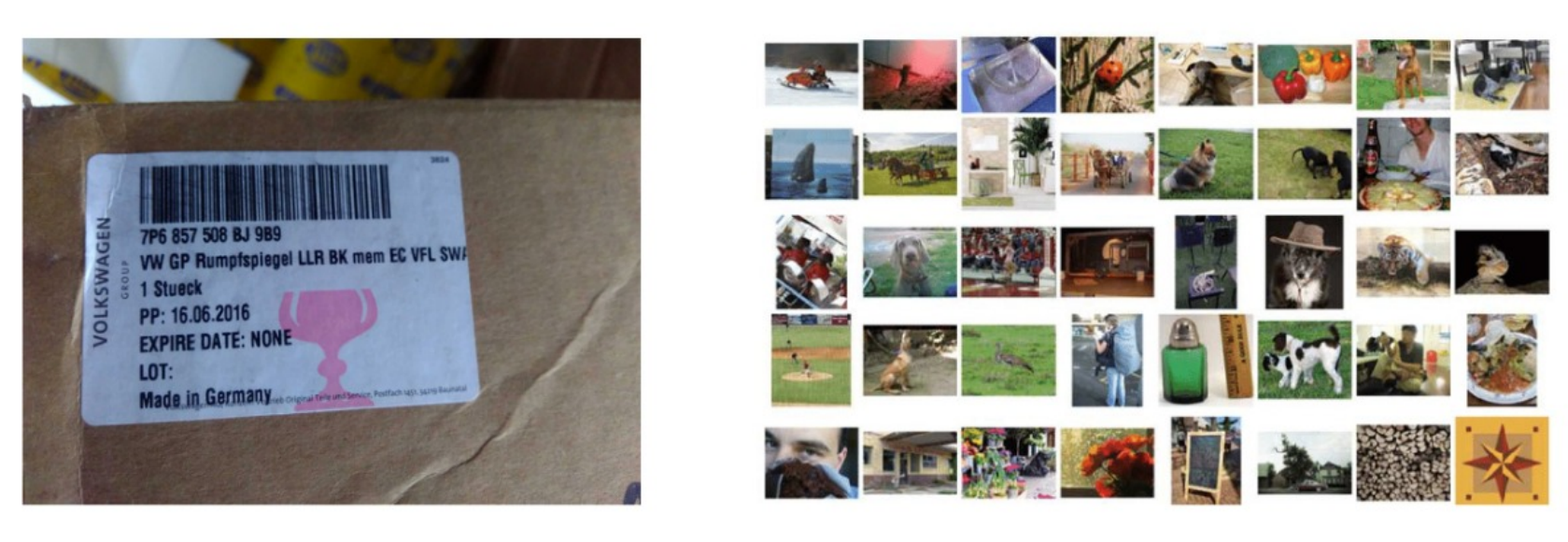


图5 非配件图片

4.1.1.3 制作

其制作步骤瑞如下：

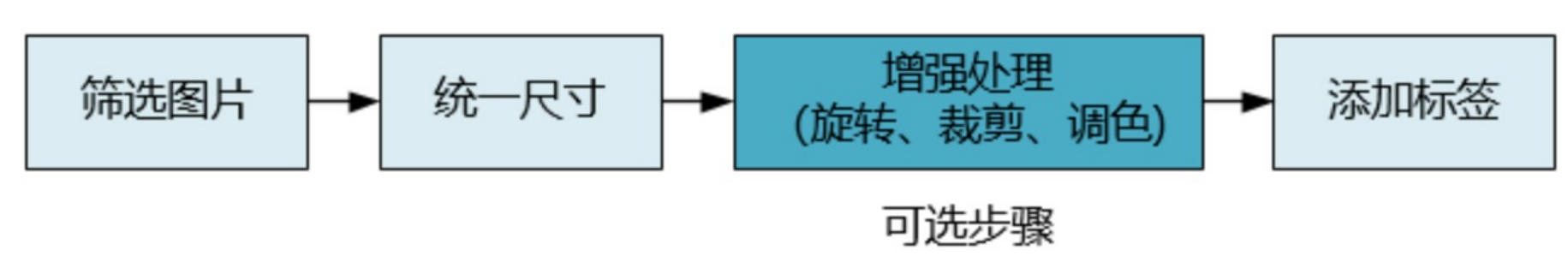


图6 数据集制作流程

步骤说明：

1. **筛选**：选取内容单一且突出、轮廓清晰的配件图片；
2. **统一图片尺寸**：比如224 x 224；
3. **增强处理**：通过增强处理（旋转、翻转、调色等）提高数据量（1万张→10万张）；
4. **添加标签**：配件图片添加标签为”1“，非配件图片添加标签为”0“。

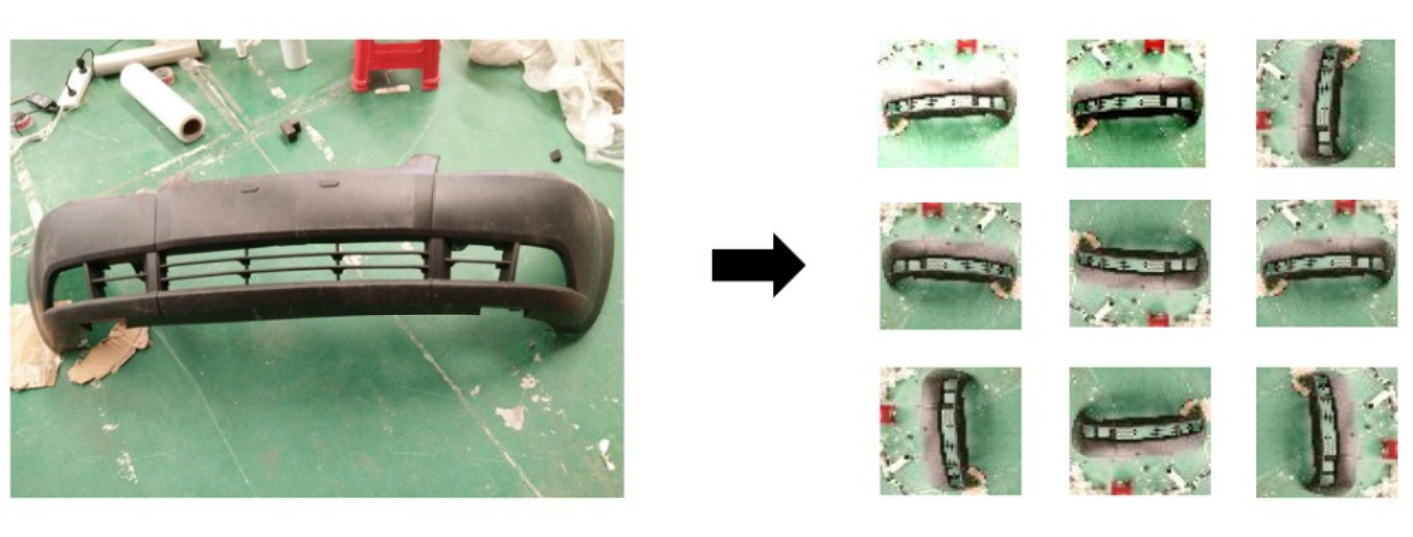


图7 数据增强

4.1.2 模型选择

按照CNNs层数与模型复杂度不同，列出以下两种CNNs模型，综合评估指标F1值和模型大小进行选择。

1. FusionNet(8层，模型简单，复杂度低，训练速度快)
2. ResNet18(18层，模型复杂，分类准确率高)

4.1.3 训练

CNNs建立分类模型的主要工作包括在训练集上进行训练、在验证集上与进行调参、在测试集上对模型泛化能力进行评估。

训练：搭建网络模型，使用自制数据集作为输入训练模型，选择带动量SGD作为优化方法，加入Batch Normalization方法加速训练，采用Dropout方法防止模型过拟合。

- **优化**：训练的目标是使模型的预测值与真实值的差达到最小，这在数学中属于最优化问题。机器学习中常用优化方法有SGD、带动量SGD、Adam等。本方案选用带动量SGD作为优化方法，因为它收敛速度快。
- **带动量SGD**：SGD(Stochastic Gradient Descent)意为“随机梯度下降”，可理解为下山，每一次都沿着最陡峭的地方往下走一步。SGD可能会得到局部最优解，带动量的SGD可以跳出局部最小区域，避免局部最优解问题。

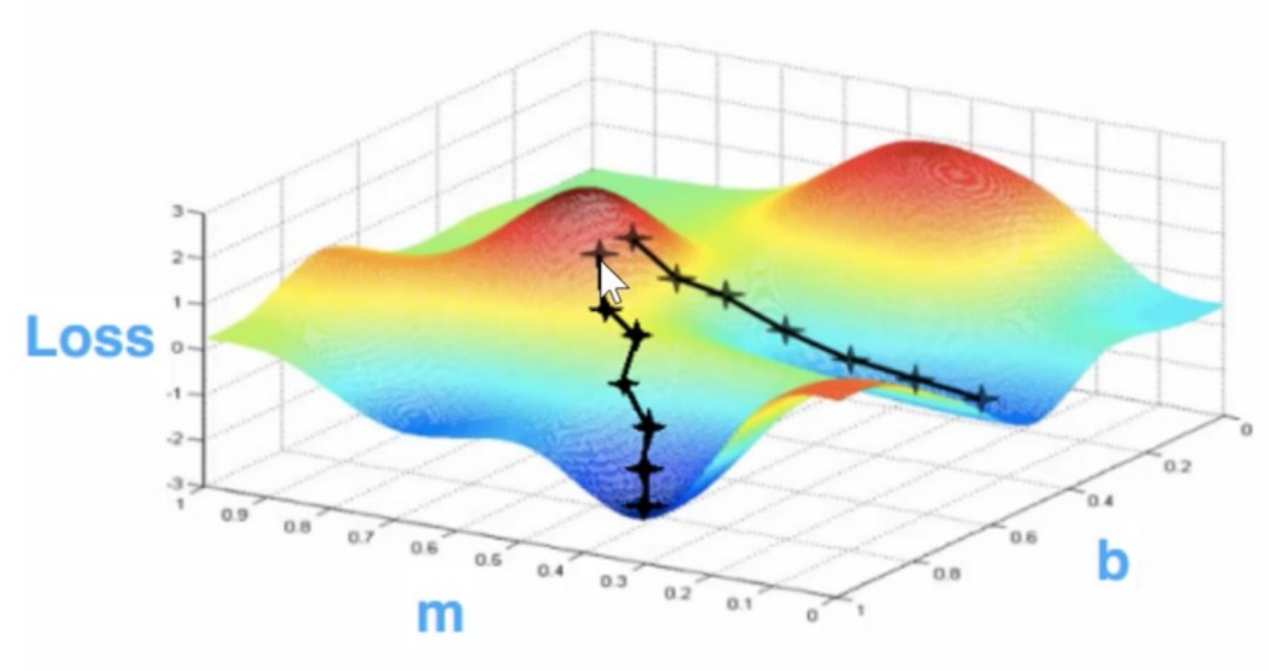


图8 梯度下降法

- **Batch Normalization**: 即“批归一化”，将隐藏层输入分布调整回标准正态分布，使得较小的输入变化有较大的梯度，避免梯度消失问题，而较大的梯度则意味着收敛速度快，有利于提高训练速度；

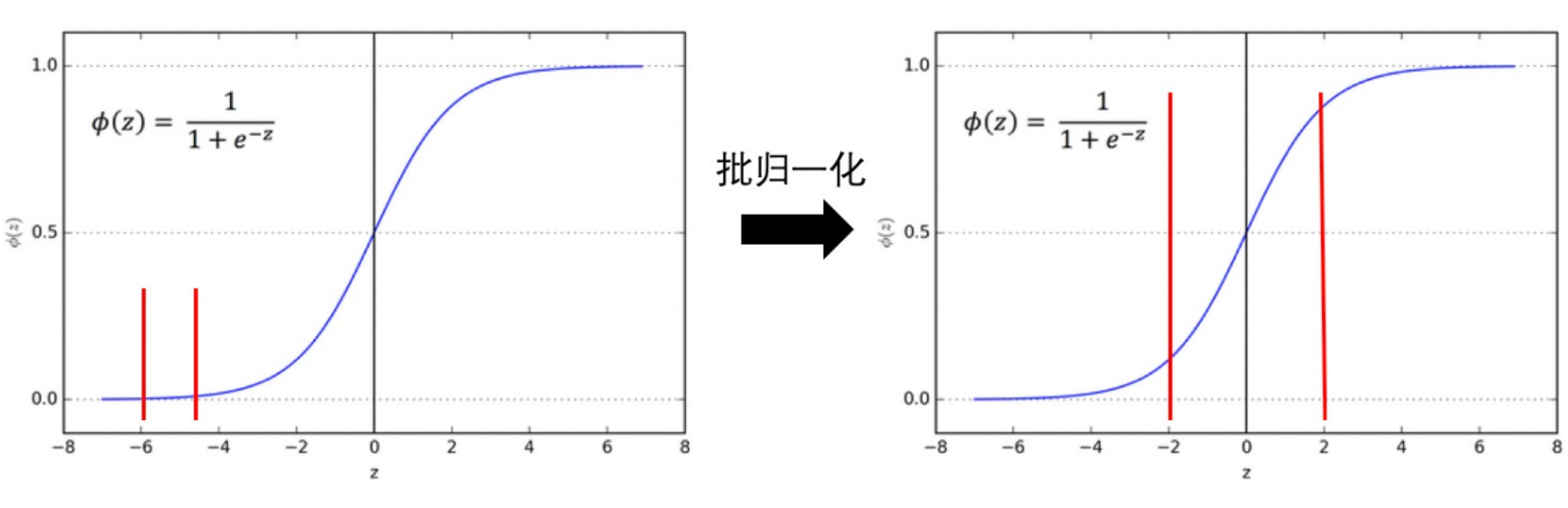


图9 批归一化

- **Dropout**: 原意为“中途退学”，这里指“随机失活”(随机使部分神经元的权重或输出设为0)，该技巧可以降低模型对某些特定特征的敏感程度，学习更具代表性的特征，降低过拟合风险，提高其泛化能力。

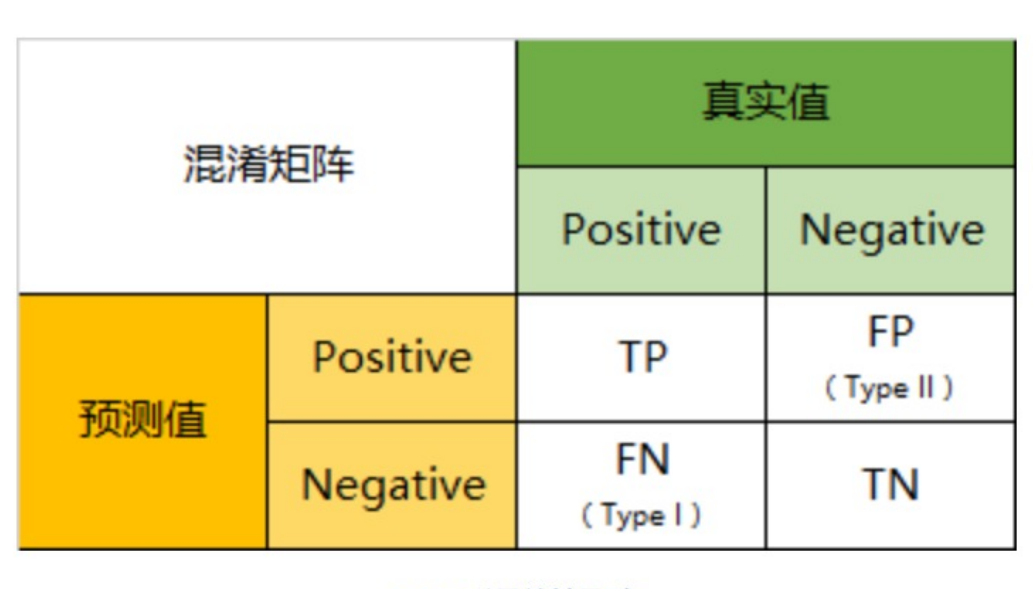
4.1.4 调参

机器学习涉及算法参数(又称“超参数”，如学习率)和模型参数(如权重参数)两类参数，两者不同之处在于前者由人工设定/调节，后者可由模型通过训练自学习得。

调参：对学习率等超参数进行fine-tune，然后再进行训练；

4.1.5 性能度量

性能度量：计算模型在测试集上精度、准确率、覆盖度及F1值等评估指标，度量模型泛化能力。



混淆矩阵		真实值	
		Positive	Negative
预测值	Positive	TP	FP (Type II)
	Negative	FN (Type I)	TN

图10 混淆矩阵

准确率：亦称查准率，表示模型预测类别为Positive的结果中，真实类别也为Positive所占比重，其计算公式为 Precision = TP / (TP + FP)

召回率：亦称查全率，表示模型真实类别为Positive的结果中，预测类别也为Positive所占比重，其计算公式为 Recall = TP / (TP + FN)

F1值：基于准确率和召回率的调和平均，综合考虑了准确率和召回率，其计算公式为 F1 = 2 * P * R / (P + R)

备注：分类任务还常用精度作为性能度量指标，应注意区分精度与准确率，精度的计算公式为 Accuracy = (TP + TN) / (TP + FP + FN + TN)

4.1.6 预期目标

分类精度：95%，F1值：0.95

4.1.7 demo效果

- FusionNet:

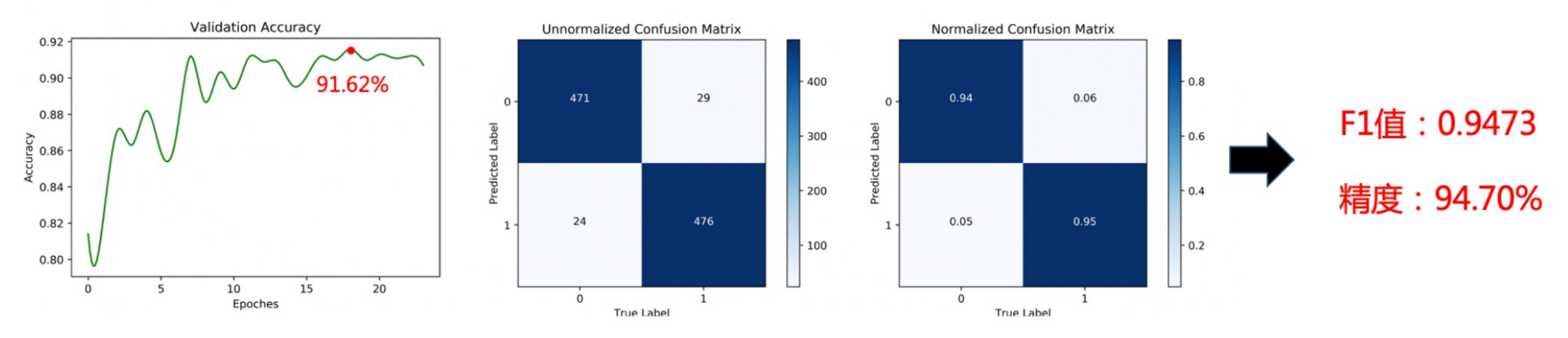


图11 FusionNet效果

- ResNet18:

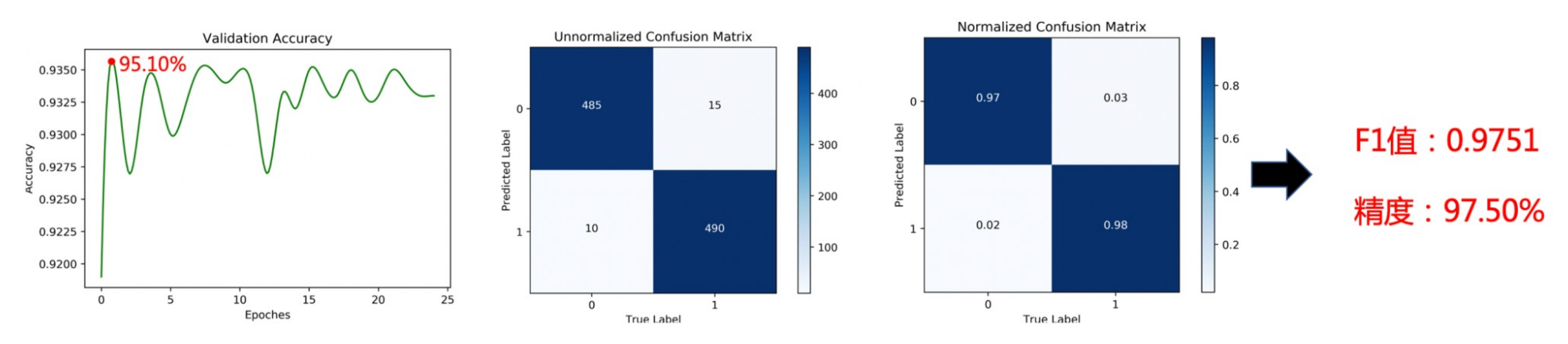


图12 ResNet18效果

4.2 敏感信息识别算法

一般地，敏感信息以字符形式存在。该算法利用OCR(Optical Character Recognition, 光学字符识别)技术对图片进行字符识别处理，若识别出的字符存在于敏感信息库(敏感词汇库+自定义敏感词汇)，则将该图片判定为不合规，需将其退回。敏感信息识别算法流程如下：

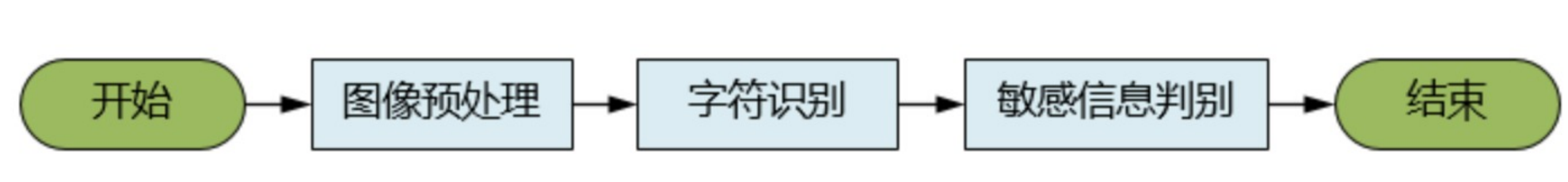


图13 敏感信息识别流程

4.2.1 图像预处理

图片在进行字符识别前需要对其进行预处理，以便后续进行特征提取，预处理步骤为：原始图像→灰度化→二值化→降噪等，各步骤处理效果如下：

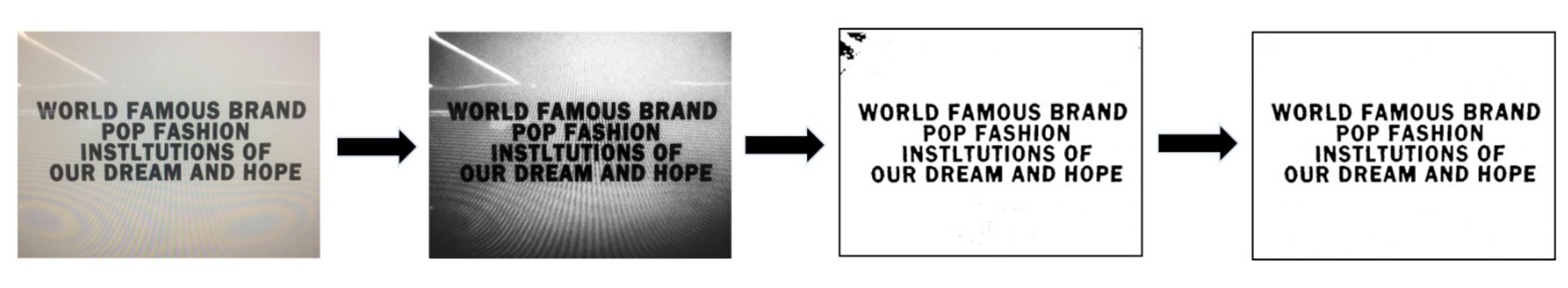


图14 图像预处理

4.2.2 字符识别

字符识别是一个特征提取然后进行分类的过程，本方案我们采用谷歌开源OCR项目中的Tesseract作为识别引擎，该引擎识别准确率高于其余识别引擎。

4.2.3 敏感信息判别

在开源词汇库(中英文敏感词汇)的基础上，添加自定义词汇(如竞品名称)，建立敏感信息库，用于判别被识别出的字符是否属于敏感信息。

4.2.4 预期目标

识别准确率95%

5. 风险评估

风险分类	风险内容	风险级别	规避措施
算法准确率	需要自制数据集，图片分类准确率与数据集质量、数量正相关	低	尽可能提高数据量，同时保证图片主体内容突出且轮廓清晰
算法覆盖度	图片被错误分类	中	1. 使用F1值评价模型分类能力 2. 可申请人工审核，避免误审、漏审

6. 时间安排

项目评审时间：2019年4月09日

开发开始时间：2019年4月10日

开发结束时间：2019年4月26日

赞 成为第一个赞同者 无标签

编写评论...