

基于深度神经网络的GUI测试异构场景图像相似度分析

蒋林呈^{1,2}, 姜磊^{1,2}, 陈纪龙^{1,2}

(1. 四川大学, 成都 610065;
2. 工程数值模拟基础算法与模型全国重点实验室, 成都 610207)

摘要: 在软件测试中, 判断GUI异构场景的图像相似度是提高测试效率和质量的关键。文章提出了一种基于深度神经网络的图像相似度分析方法。其中, 首先对GUI图像进行了预处理并通过VGG-16提取了图像特征, 同时引入了OCR模块获取文本信息; 在相似度计算阶段, 分别计算了图像特征和文本的相似度, 通过交叉熵损失函数和字符串匹配算法得到了相似度分数, 最终进行了加权求和。经验证, 该方法能有效抵抗设备差异导致的噪声和变形干扰, 提高了GUI测试的效率和质量, 为测试人员提供了准确、全面的结果和决策依据。

关键词: 软件测试; OCR; 图像相似度

中图分类号: TP311 **文献标识码:** A

Analysis of image similarity in heterogeneous scenarios of GUI testing based on deep neural networks

JIANG Lincheng^{1,2}, JIANG Lei^{1,2}, CHEN Jilong^{1,2}

(1. Sichuan University, Chengdu 610065, China;
(2. National Key Laboratory of Fundamental Algorithms and Models for Engineering Simulation, Chengdu 610207, China)

Abstract: In software testing, determining the image similarity of GUI heterogeneous scenes is the key to improving testing efficiency and quality. The article proposes an image similarity analysis method based on deep neural networks. Firstly, the GUI image was preprocessed and image features were extracted using VGG 16, while an OCR module was introduced to obtain text information. In the similarity calculation stage, the similarity between image features and text was calculated separately. The similarity scores were obtained through cross entropy loss function and string matching algorithm, and finally weighted sum was performed. After verification, this method can effectively resist noise and deformation interference caused by device differences, improve the efficiency and quality of GUI testing, and provide accurate and comprehensive results and decision making basis for testers.

Key words: software testing, OCR, image similarity

1 引言

在软件测试中, 判断GUI异构场景的图像相似度是提高测试效率和质量的关键, 准确的图像相似度判断有利于发现界面差异。在此基础上, 可准确识别潜在的问题和缺陷, 从而加快测试速度并提高可靠性和用户体验。在GUI异构场景中, 不同平台、设备和环境下的图像差异会影响测试的准确性和效率, 因此需要优化图像相似度计算方法。

现有的图像相似度判断方法主要包括传统算法和基于深度神经网络的算法。逐像素比较、直方图比较、SSIM^[1]和哈希算法等传统算法分别通过像素值、直方图数据、图像亮度对比度和结构相似性等评估图像相似度, 但在应对GUI测试异构场景中的大幅度图像变化(如旋转、缩放、偏移、色彩变化等)时表现并不理想。

随着深度学习的发展, 基于深度神经网络的图像相似度分析日益重要。孪生和三支网络利用深度

学习和余弦相似度处理复杂图像变化,具有优异的性能。然而,GUI测试异构场景中的图像元素复杂多变,现有方法难以排除干扰。因此,开发一种能准确判断复杂异构图像相似度的方法至关重要。

2 方法

针对GUI测试异构场景中的图像多样性,本文提出了一种增强的图像相似度判断方法。该方法通过VGG-16网络提取图像关键特征,引入特征增强注意力模块以丰富和增强这些特征,从而得到更准确的图像相似度分数。基于此,该方法可结合OCR技术识别图像中的文字,并计算文字相似度分数,最终将图像和文字相似度分数加权结合,以形成综合相似度判别分数。这种方法可实现包含复杂图形和文字的高度复杂图形对的相似度对比,增强了模型在GUI测试异构场景中的精确性和鲁棒性。

2.1 基于孪生网络的图像相似度判断

2.1.1 图像特征提取

在使用孪生网络对GUI图像提取关键特征时,本文将VGG-16^[2]作为孪生网络的骨干网络。具体来说,所构建的VGG-16由Stem层及若干CBR层组成。其中,Stem层用于调整输入特征图的尺寸和通道数;各CBR层由卷积层、BatchNorm层、ReLU层组成,用于完成图像关键特征(包含图像的轮廓特征、颜色特征以及纹理特征)提取。包含骨干网络VGG-16的孪生网络结构如图1所示。

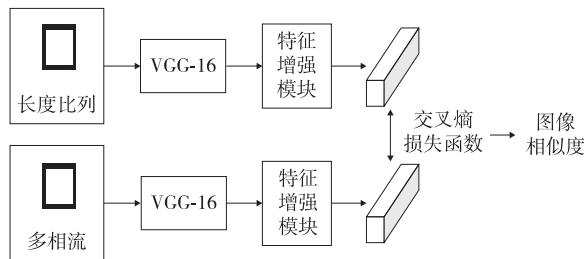


图1 孪生网络结构示意图

将用于输入GUI图像特征提取的孪生网络定义为 $Network_{siamese}$,它使用的2个骨干网络VGG-16分别为 $Backbone_{VGG-16}^1$ 和 $Backbone_{VGG-16}^2$;将输入孪生网络的GUI图像定义为 x ,输入至孪生网络的2个骨干网络后分别得到2个图像关键特征 f^1 和 f^2 ,用如下公式表示:

$$f^1, f^2 = Network_{siamese}(x) \quad (1)$$

2.1.2 特征增强模块

如图2所示,本文引入了2个特征增强注意力模块,旨在对提取的图像特征进行语义丰富和增强。针对VGG-16骨干网络所提取的图像特征向量,首先进行层归一化操作以确保彼此的尺度一致性;随后利用自注意力机制对这些特征向量之间的关系进行建模,

从而获得更加精细的特征表示;将这一过程得到的结果与原始特征向量进行残差连接,以保留原有的信息;随后再次对残差结果进行层归一化操作,以进一步提升特征的稳定性;接着将这一结果输入多层感知机,以引入非线性变换并捕捉更高级的语义信息;最后将多层感知机的输出与原始残差结果再次进行残差连接,旨在获得更加丰富和高质量的特征表示。

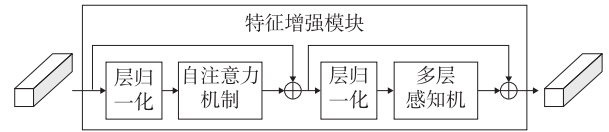


图2 特征增强模块

2.1.3 损失函数

通过绝对差函数和线性层计算得到图像 x_1 和 x_2 的图像相似度分数 P_i :

$$P_i = Linear(Abs(x_1, x_2)) \quad (2)$$

其中, $Abs(\cdot)$ 表示计算2个特征向量之差的绝对值; $Linear$ 表示线性层。

采用交叉熵损失函数对特征相似性进行监督学习,得到图像的相似度分数。具体来说,为了比较2个图像特征向量之间的相似程度,首先需定义一组样本对,每个样本对由2个图像及其相似性标签组成,相似性标签为二元制(相似/不相似);然后使用这些样本对来训练模型。对于每个样本对,通过骨干网络和特征增强模块提取特征向量,然后利用交叉熵损失函数计算相似度,并经由反向传播梯度来更新网络参数。通过最小化交叉熵损失^[3],可以训练模型以更好地拟合实际的特征相似性。具体地,基于交叉熵损失函数 L ,采用多个图像对进行图像相似度分数的评估训练:

$$L = -\frac{1}{N} \sum_{i=1}^N (y_i \log(p_i) + (1-y_i) \log(1-p_i)) \quad (3)$$

式中, N 为参与计算损失的图像对的总数; i 为图像对的序号; y_i 为第 i 对图像的真实标签,若 $y_i=0$,则图像对不相似,若 $y_i=1$,则图像对相似。基于此,多次迭代至交叉熵损失函数 L 收敛或达到预设迭代次数。

2.2 基于OCR的图像相似度判断

将用于比较相似度的一对GUI图像定义为 x_1 和 x_2 ,并利用开源的文本识别技术分别提取 x_1 和 x_2 中的文本信息 T_1 和 T_2 ,用如下公式表示:

$$T_1, T_2 = OCR(x_1, x_2) \quad (4)$$

其中, OCR ^[4]表示开源的百度文本识别技术,通过本地部署来提取待比较GUI图像的文本信息。用OCR技术提取出的文本信息 T_1 和 T_2 中包含的字符定义如下:

$$T_1 = (w_1^1, w_1^2, \dots, w_1^m), T_2 = (w_2^1, w_2^2, \dots, w_2^n) \quad (5)$$

其中, w_1^i 为 T_1 包含的某个字符; w_2^i 为 T_2 包含的某个字符; m 为 T_1 中的字符数量; n 为 T_2 中的字符数量。根据

T_1 和 T_2 中包含的字符,通过字符匹配得到 T_1 和 T_2 相同字符数量 C :

$$C = \text{Find}(T_1, T_2) \quad (6)$$

其中, Find 是找寻 2 个文本 T_1 和 T_2 的相同字符数量 C 的函数。当然,该判定值可根据实际情况做适应性的调整。计算图像 x_1 和 x_2 的文本相似度分数 V :

$$V = \frac{2C}{m+n} \quad (7)$$

若 $V \geq 0.99$, 则判定文本 T_1 和 T_2 相似, 否则不相似。同时, 可以通过比较相同字数与预设数量的大小判定图像 x_1 和 x_2 的文本是否相似:

$$U = \max(m, n) \times k \quad (8)$$

其中, U 为预设相似次数; k 为预设文本相似比例系数, 一般为 0.99。

对文本相似度分数和图像相似度分数进行加权求和。可根据数据集中含有文本的数量占比设置权重比例, 若含有文本的数量为 70%, 可将文本相似度分数权重设置为 0.7, 并将图像相似度分数权重设置为 0.3, 最后将这 2 个分数乘以各自的权重并相加以形成最终的相似度分数, 具体过程可如表示为:

$$\text{MixScore} = P_i * RI + V * RO \quad (9)$$

其中, MixScore 为综合相似度分数; R_i 为图像相似度分数权重; RO 为文本相似度分数权重。

在完成以上计算后, 需设置相似度阈值, 若最后得到的相似度分数大于等于该阈值, 则模型认为这 2 张图像相似; 若最后得到的相似度分数小于该阈值, 则模型认为这 2 张图像不相似。该阈值可根据实际情况调整, 一般设置为 0.99。

3 数据预处理与实验

3.1 数据扩充与数据增强

本文采用的数据集主要源自 2 个未公开的软件, 涵盖纯文字、纯图形及图文混合三类内容。该数据集被进一步细分为单张图像与 405 对相似图像(共 810 张)两部分。为丰富数据量, 每张原始图像经裁剪得到 4 张新图像, 裁剪点位于图像的 4 个角落, 各占图像长宽的 1/25。后续数据增强步骤如下: 随机以 0.5 的概率进行图像反转, 实施色域变换, 添加灰条以模拟多余图像部分, 应用 padding 处理, 并统一调整所有图像至 105×105 像素大小以适应训练。

经过上述处理, 最终得到 1 472 张图片, 将其整合为 300 个类别集合。同一集合内的图像属于相同类别

(即相似图像对), 不同集合间图像则属于不同类别。最终, 训练集和验证集的划分为 9:1, 即 1 324 对图片用于训练, 148 对图片用于验证。

3.2 实验设置

模型训练的 epoch 次数设为 200; 由于设置的图像分辨率较小, batchsize 设为 128; 使用的优化器为 SGD, 学习率下降方式为 cos, 初始学习率设置为 1e-2; 每 40 个 epoch 保存 1 次模型权重, 使用的损失函数为二值交叉熵损失函数。

3.3 实验结果

实验结果如表 1 所列, 本文设置的相似阈值为 0.99, 即模型给的分值大于 0.99 则认为这对图相似, 若小于 0.99 则认为不相似。目前, 基于图像相似度对比任务的模型较少, 故本文仅列出 LPIPS^[5] 模型在这 3 个数据集上的准确率。

表 1 实验结果

方法	数据集		
	只包含文字 准确率	只包含图形 准确率	既包含文字又包含 图形准确率
Ours	0.982 3	0.991 4	0.986 7
LPIPS	0.976 8	0.987 9	0.981 5

由表 1 可知, 本文模型在 3 个数据集上的准确率都比 PLIPS 模型高约 1%。

参考文献:

- [1] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.
- [2] KAREN S. Very deep convolutional networks for large-scale image recognition[J]. arxiv preprint arxiv: 1409.1556, 2014.
- [3] WANG Y, MA X, CHEN Z, et al. Symmetric cross entropy for robust learning with noisy labels [C] // Proceedings of the IEEE/CVF international conference on computer vision, 2019: 322-330.
- [4] DU Y, LI C, GUO R, et al. PP-OCR: A practical ultra lightweight OCR system. arxiv 2020[J]. arxiv preprint arxiv: 2009.09941, 2009.
- [5] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [C] // Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 586-595.

作者简介:

蒋林呈(2001—), 硕士, 研究方向: 图像处理。