

基于蜕变关系模式的光学字符识别系统软件评估方法

金铃子, 牛佳*

浙江理工大学信息技术中心, 浙江 杭州

收稿日期: 2025年4月17日; 录用日期: 2025年6月10日; 发布日期: 2025年6月18日

摘要

光学字符识别系统通过扫描图像, 使得原本只能以图像形式存在的文字信息可以被计算机处理和分析, 已在各领域广泛应用。光学字符识别技术的质量是重点, 如何系统地对其质量进行评估仍是一项具有挑战性的工作, 因为它的评估方法依赖于大量的标记数据, 同时也依赖于人工判断。在本文中, 从用户的角度对光学字符识别系统进行评估, 帮助用户更好地了解这些系统的功能, 进而选择适合自己的光学字符识别系统, 以满足他们的特定需求。为了更好地评估光学字符识别系统, 光学字符识别系统的特征, 从用户角度出发, 我们定义了两类蜕变关系模式。在此基础上, 定义了六个抽象的蜕变关系, 并选取了光学字符识别的广泛应用场景车牌识别来评估系统的质量。为保证在没有先验标注的情况下评估OCR系统的性能以及鲁棒性, 本文提出了此方法。

关键词

光学字符识别, 蜕变测试, 质量评估方法

Software Evaluation Method of Optical Character Recognition System Based on Metamorphic Testing

Lingzi Jin, Jia Niu*

Information Technology Centre, Zhejiang Sci-Tech University, Hangzhou Zhejiang

Received: Apr. 17th, 2025; accepted: Jun. 10th, 2025; published: Jun. 18th, 2025

*通讯作者。

文章引用: 金铃子, 牛佳. 基于蜕变关系模式的光学字符识别系统软件评估方法[J]. 软件工程与应用, 2025, 14(3): 585-595. DOI: 10.12677/sea.2025.143051

Abstract

Optical Character Recognition (OCR) systems enable computers to process and analyze text information that originally exists only in image form through scanning, and have been widely adopted across various domains. The quality of OCR technology is of paramount importance, yet systematically evaluating its performance remains challenging as current assessment methods heavily rely on large volumes of labeled data and manual judgment. In this paper, we propose evaluating OCR systems from the user perspective to help users better understand system capabilities and select appropriate OCR solutions that meet their specific needs. To facilitate comprehensive evaluation, we define two categories of metamorphic relation patterns based on key OCR characteristics from the user perspective. Furthermore, we establish six abstract metamorphic relations and validate them through license plate recognition—a prevalent application scenario for OCR technology. The proposed method enables performance and robustness evaluation of OCR systems without requiring prior annotated data.

Keywords

Optical Character Recognition, Metamorphic Testing, Quality Assessment Method

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在各个行业中, 纸质文档或者是图像形式的符号常常需要被记录成电子文档以便于提高工作效率和保存。光学字符识别(Optical Character Recognition, OCR)能够将纸质资料、图像或是手写稿件, 转换为便于计算机处理的数字文本。数字文本大幅提升数据处理的效率和便捷性, 也便于存储和分析。它使得信息管理和检索更加高效, 广泛应用于各个行业以减少手动输入和提高工作效率。在各种 OCR 系统中, 办公自动化[1]、车牌识别[2]和文档管理[3]问题在具体应用中非常广泛。

光学字符识别系统的应用场景广泛, 但车牌识别由于大多数是移动的户外场景, 受到多因素的干扰, 对于机器的准确识别带来了巨大的挑战。由此看来, 对 OCR 系统的车牌识别进行鲁棒性评估具有重要的现实意义和研究意义。目前, 在该领域还缺乏相关研究。光学字符识别模型大多依赖于深度学习等模型[4], 由于深度学习模型缺乏可解释性, 如何提高模型的鲁棒性仍是当前人们研究的热点和挑战。车牌识别面临复杂环境适应性技术挑战, 不同的光照、天气、车牌污损等情况对识别精度有较大影响。并且, 在车辆快速移动的情况下, 仍需保证识别准确度和效率。中国车牌有多个种类和格式, 识别系统需要有兼容性能够准确识别各种车牌。

车牌识别 OCR 技术推动交通管理自动化发展, 同时为构建智慧城市奠定基础, 增强智慧治安防控能力。随着技术的应用与改善, 车牌识别 OCR 的准确率和适用范围也会提升。对 OCR 系统的评价通常限制于先验标注的数据集。这些数据集通常依赖人工标注, 且场景单一, 涉及扰动因素不多。与此同时, 由于车辆车牌种类的新增, 如新能源车牌颜色和字体的变化, 导致原有的数据集不够全面。这些因素都导致了我们所评估维度和指标较为单一, 无法真正体现 OCR 系统的性能。

针对 OCR 技术应用中的关键问题, 本研究创新性地采用蜕变关系模式, 设计了面向 OCR 系统的

软件评估框架。该方法从用户角度出发, 重点分析 OCR 系统的实际表现与用户需求之间的匹配程度。通过这种评估方式, 不仅能够帮助用户更全面地认识不同 OCR 系统的性能特点, 还能为其在实际应用中选择最适合自身需求的解决方案提供科学依据。因此, 我们使用了基于属性的测试技术蜕变测试[5](metamorphic testing, MT)技术, 它在软件验证和系统理解[6][7]、图像生成[8]等领域已经被证明其有效。

蜕变关系(metamorphic relations, MRs)是蜕变测试的关键组成部分, 通过定义多个相关输入和输出之间的关系, 来判定系统的属性。最初, 蜕变测试的出现是为了解决传统测试无法覆盖的软件测试问题。近年来, 它已成功地扩展到故障检测[9][10]仿真与建模系统和优化或加密程序[11][12]。

在这项研究中, 本研究针对 OCR 系统开发了两类通用蜕变模式及六组蜕变关系(MRs)。为验证方法的实用性, 特别选取车牌识别这一典型应用场景进行验证性测试。通过多维度评估框架, 该方法能够全面展示 OCR 系统的性能特征, 为用户提供系统选型参考。

本文提出应用蜕变测试技术首次提出两个蜕变关系模式, 并将蜕变关系模式具体化, 形成了可以具体应用的 MRs。

2. 相关工作

在光学字符识别(OCR)领域, 大量的研究都集中在鲁棒性问题上。而在构建输入数据方面, 研究包括大型车牌人工标注数据[13]以及对抗性样本生成[14]等。在评价系统的性能方面, 研究人员通过评估文本识别性能的指标: 字符识别准确率(character-recognition accuracy, CRA) [15]和单词识别准确率(word-recognition accuracy, WRA) [16]来作为 OCR 算法识别率的评价指标。其中, CRA 反映正确识别的字符占比, WRA 则衡量正确识别的单词比例, 但二者均依赖标注数据的支持。

当输入数据包含噪声时, 即使是人工无法识别的微小扰动也可能显著影响识别置信度, 甚至造成完全错误的识别结果。而针对 OCR 领域中的输入图像噪声模拟, Xu 等人[17]提出了 OCR 噪声模拟, 通过基于规则的模拟、攻击的模拟。基于模型的仿真来对 OCR 文本加入噪声, 通过多源噪声模拟提升模型对噪声的鲁棒性, 并引入稳定性损失优化特征表示。Todorov 等人[18]通过实验验证了 OCR 噪声对语言模型的影响, 发现噪声会导致模型分布偏离纯净数据, 且噪声强度与模型性能呈负相关。并使用量化依据字符错误率(CER)和语义表示差异来评估噪声模型。

针对目前已有的蜕变测试对光学字符识别领域的应用中, 社媒图像分析[19]和手写字符识别[20]等领域, 没有针对车牌识别领域有说服力的蜕变测试方法。由于车牌的复杂性和多样性, 系统的质量评估难度更大。基于此, 本文主张从用户实际需求出发, 建立能够有效评估模型语义理解能力和抗干扰性能的新指标体系。因此, 我们从用户角度出发, 提出了将 MT 作为监测系统的方法。

3. 蜕变测试

在软件工程中, 测试环节是一项重要且必不可少的工作。软件测试的目的是发现错误、缺陷、漏洞或不足, 并确保软件产品满足既定的需求和规格。在传统软件测试中, 我们通常设定测试用例中凤有预期的输出结果, 通过比较程序预期输出和实际的输出之间的关系来判定缺陷。但是对于某些程序, 我们难以区分程序的实际输出结果与期望结果之间的关系, 也就是 oracle 问题[21]。为了解决这类问题, 蜕变测试的概念被提出[22]。近年来, 已有大量研究阐明了蜕变测试技术的合理性, 并检验了它在解决 Oracle 问题和先验标注问题。蜕变测试并不聚焦于被测软件(software under test, SUT)的期望输出, 而是对属性做出检验。它会多次执行被测软件, 进而验证输入和输出之间的关系, 这种关系被称为蜕变关系。我们所定义的 MR 需为 SUT 必须满足的属性, 如果某些测试用例违反了 MR, 那么说明 SUT 必有缺陷。

3.1. 蜕变关系的定义

在 MT 中, 我们根据 SUT 的属性和功能来定义输入和输出的期望关系, 也就是 MRs, SUT 根据预定义的 MRs 进行检查。蜕变关系(MRs)规定了软件应遵循的功能规则, 如果在测试中发现某些用例违反 MR, 则被测系统(SUT)必然存在缺陷。以函数 f 为例, 被测软件 P 需对输入序列 $(x_1, \dots, x_n)$ 处理并输出对应结果 $(f(x_1), \dots, f(x_n))$ 。我们用一个关系 (I, I') 来表示源输入和后续输入的关系, 那么 $(x_1, \dots, x_n)$ 在满足关系 (I, I') 时, $(f(x_1), \dots, f(x_n))$ 满足关系 (O, O') 。我们可以称 $r(I, I', O, O')$ 是被测软件 P 的蜕变关系。若被测软件 P 满足上述 MR, 则系统执行正确。

以经典函数 $f(x) = e^x$ 为例, 假设被测软件 P 被用来执行此目标函数, 对于不可验证结果或是较难验证结果的随机输入 x , 基于 $f(x)$ 函数法人特性定义蜕变关系以产生蜕变测试用例, 如下所示:

$$e^x \times e^{-x} = 1$$

例如: $x_1 = 0.2$, 衍生测试用例为 $x_2 = -0.2$, 验证蜕变关系 $e^{0.2} \times e^{-0.2} = 1$ 是否通过。若结果不满足蜕变关系, 则说明被测软件存在一定的缺陷。

3.2. 蜕变关系模式

在蜕变测试领域, 由于是从用户的使用角度出发, 往往有特定的研究问题, 研究人员提出具有针对性的蜕变关系(MRs), 这些 MRs 具有较强的特殊性。为了能够使研究方法更有普适性, Zhou 等人提出了一般蜕变关系的概念[21], 并验证了基于这种抽象概念所构建的具体模型, 在故障检测方面的有效性。此外, Zhou 等人[23]尝试将该概念应用于搜索引擎的测试故障检测中。

在此基础上, 为了避免被测试软件局限于执行单一特定任务的类型, 即仅围绕搜索操作的软件展开研究这一问题, Segura 团队[5]创新性地提出了蜕变关系输出模式(metamorphic relation output pattern, MROP)这一概念, 在更高层面抽象出了源输出与后续输出间的关联性。这一理论突破为蜕变关系(MR)的普适性研究带来了新的视角。与此同时, Zhou 等学者[24]从形式化定义了蜕变关系模式(MRP), 构建能够概括描述多个蜕变关系的抽象框架, 并进一步提出蜕变关系输入模式(metamorphic relation input pattern, MRIP), 为源输入向后续输入的转换提供了方法论指导。这些概念的提出不仅简化了蜕变关系的识别过程, 还为自动化推导相关蜕变关系提供了理论依据。

然而, 在光学字符识别(OCR)这一特定领域, 现有的蜕变测试方法仍存在明显局限性。目前的蜕变测试方法, 如社媒图像分析[18]和手写字符识别[20]等领域, 均局限于特定任务场景, 缺乏普适性。针对这一现状, 本研究基于 MRP 的理论框架, 提出了两种适用于 OCR 领域的通用蜕变关系模式, 为相关领域的蜕变测试研究提供了新的思路和方法。

4. 蜕变测试过程

蜕变测试要求根据用户需求和不同系统特点, 构造测试方法来评估系统。对于光学字符识别, 我们分析输入的重要属性, 以此来构建蜕变关系。我们使用的输入通常是图片, 而图片需要进行预处理, 来构造我们需要的蜕变关系。

图 1 概述了该方法。给定一组源输入(即车牌)和 MRs 列表, 生成一组相应的衍生输入数据。通过使用与单个 MR 相关的源输入和衍生输入执行 OCR 系统, 可以得到它们的源结果和衍生结果。

5. 光学字符识别系统的蜕变关系

本研究为光学字符识别(OCR)系统构建了一套基于蜕变关系的评估框架。通过分析用户对 OCR 系统的核心需求特性, 我们设计的蜕变关系可通过通过率与违反率的量化指标, 直观揭示系统性能边界。本文创新性地提出两种蜕变模式, 共包含 6 个 MRs。本节将详细阐述各 MRs 的技术实现方案, 并辅以典型

应用案例进行说明。

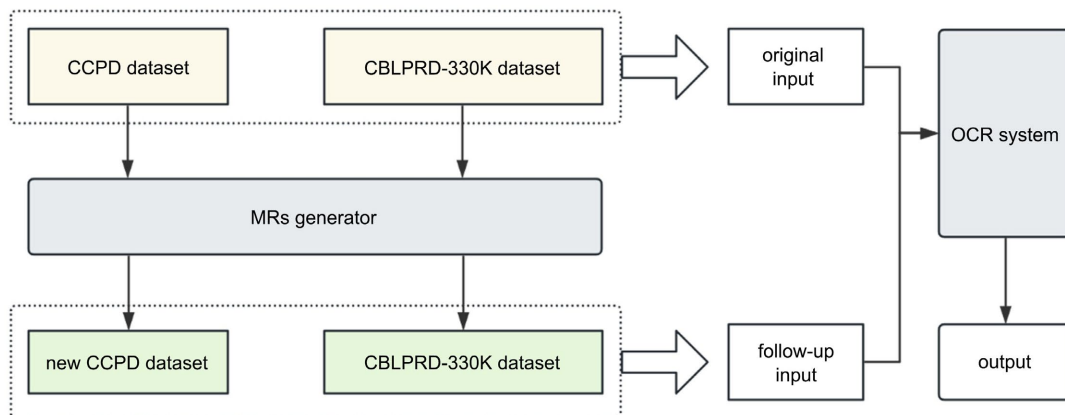


Figure 1. Evaluation method for optical character recognition system based on metamorphic testing

图 1. 基于蜕变测试的光学字符识别系统评价方法

经过对光学字符识别系统的分析，我们定义了两类蜕变关系模式，我们将在接下来两小节中详细描述。

5.1. 噪点蜕变关系模式

基于图像退化模型根据不同的退化因素，可以分为多种类型，例如运动模糊模型(物体运动导致的模糊拍摄)、高斯模糊模型(高斯滤波器处理)、散焦模糊模型(相机对焦不准确或景深限制)。Jin 等人[25]生成对抗网络(TSR-WGAN)涡流的湍流介质在波传播过程中会引起折射率的波动，模拟现实物体二维成像；蒋等人[26]提出模拟图像退化方案，采用数据增强的方式生成大量模拟退化图像供训练网络。

基于此，我们设计的第一类为噪点蜕变关系模式，该模式的核心思想是通过在原始输入数据中引入环境干扰因素，对比系统在原始输入和衍生输入下的输出结果，从而检测系统可能存在的缺陷。我们定义一个函数来代表替换关系：

$$P_n = Add_E + P$$

P 表示给定的源图像， E 为添加的环境因素噪点， P_n 为加入噪点后的衍生图像。基于系统特性和用户需求，我们为该模式定义了六种蜕变关系，这些关系可根据具体应用场景进行调整，如下：

MR1: 给定源输入图像为 P_1 ，通过添加天气相关的噪声 $E-w$ 生成衍生图像输入 P_n ，其表达式为 $P_n = Add_{E-w} + P$ 。随后，运用 OCR 系统分别对 P_1 和 P_n 进行处理，得到初始输出 O_1 和衍生输出 O_n 。初始输出结果 O_1 和衍生输出结果 O_n 需满足一致性。

MR2: 给定源输入图像为 P_1 ，通过添加光线相关的噪声 $E-L$ 生成衍生图像输入 P_n ，其表达式为 $P_n = Add_{E-L} + P$ 。随后，运用 OCR 系统分别对 P_1 和 P_n 进行处理，得到初始输出 O_1 和衍生输出 O_n 。初始输出结果 O_1 和衍生输出结果 O_n 需满足一致性。

MR3: 给定源输入图像为 P_1 ，通过添加脏污相关的噪声 $E-D$ 生成衍生图像输入 P_n ，其表达式为 $P_n = Add_{E-D} + P$ 。随后，运用 OCR 系统分别对 P_1 和 P_n 进行处理，从而得到初始输出结果 O_1 和衍生输出结果 O_n ，初始输出结果 O_1 和衍生输出结果 O_n 需满足一致性。

5.2. 空间蜕变关系模式

除了环境与天气因素，距离和空间也对 OCR 系统的识别有着不可忽视的影响。张等人[27]通过使用

输入图像的 VGGNet 特征测量输入图像和训练图像的距离来验证基于 DNN 的系统的输入图像对自动驾驶的影响。

基于此, 我们涉及的第二类为空间蜕变关系模式, 主要用于研究构成源输入的空间结构变化对系统造成的影响。该模式通过对比原始输入与经过空间变换的衍生输入的处理结果, 以发现系统可能存在的漏洞。我们定义的蜕变关系如下:

MR4: 设源输入图像为 P_1 , 将 P_1 中进行字符完整切割, 打乱顺序后拼接成完整图像生成衍生输入 P_n , 分别对 P_1 和 P_n 进行光学字符识别, 得到初始输出 O_1 和衍生输出 O_n 。 O_1 和 O_n 需满足一致性。

MR5: 设源输入图像为 P_1 , 将 P_1 旋转一定角度生成衍生输入 P_n , 分别对 P_1 和 P_n 进行光学字符识别, 得到初始输出 O_1 和衍生输出 O_n 。 O_1 和 O_n 需满足一致性。

MR6: 设源输入图像为 P_1 , 将 P_1 中的图像进行裁剪(不裁剪字符)生成衍生输入 P_n , 分别对 P_1 和 P_n 进行光学字符识别, 得到初始输出 O_1 和衍生输出 O_n 。 O_1 和 O_n 需满足一致性。

6. 车牌识别

我们把蜕变关系模式, 以车牌识别任务为例, 进一步说明。车牌识别 OCR 技术是一种利用图像识别将车辆牌照信息转化为计算机可接受的信息的技术。它是智能交通系统(ITS)的重要组成部分, 被普遍运用于道路监管、安全保障、停车系统运营等诸多方面。

6.1. 蜕变关系实现

我们通过定义的蜕变关系来评估 OCR 系统。一些特定的 MR 实现流程如图 2 所示。

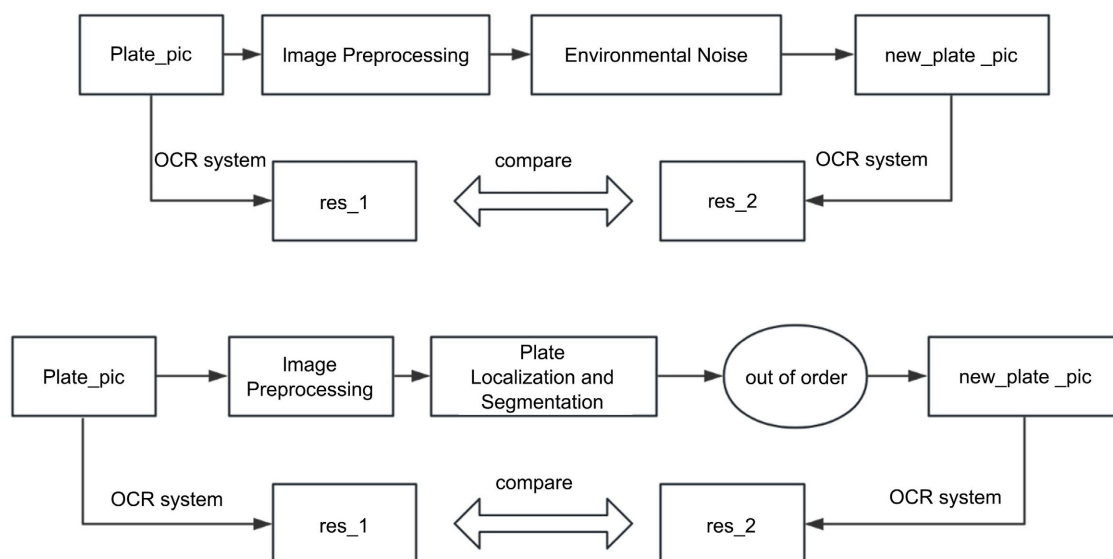


Figure 2. Metamorphic relations implementation process

图 2. 蜕变关系实现流程

在噪点蜕变关系模式下, 可以采用 CycleGAN 对抗网络[28]来进行噪点加入。他的原理是加入一些不影响人工判断的噪点进入图片以检测系统的鲁棒性。通过对源输入进行分析, 我们从进行修改, 然后比较源输出和衍生输出是否满足我们定义的蜕变关系。

在空间蜕变关系模式下, 我们首先采用中值滤波器去噪声, 灰度化, 去斜和校正区域划定 RRPN 模型字符分割, 交换字符位置, 然后以切割为完整的文字进行空间处理, 生成新的衍生输入文本, 然后比

较源输出和衍生输出是否满足我们定义的蜕变关系。

6.2. 蜕变关系模式

为了更好地阐述蜕变关系模式，我们以车牌识别为例，将上述定义的通用性蜕变关系模式具体化，得到一组的蜕变关系。给定的一组源输入 P_1 ，通过光学字符识别工具对车牌进行识别。通过我们定义的 MRs 处理图像生成衍生输入 P_n ，对比源输入的结果和衍生输入的结果是否符合我们期望的结果，以此来判断是否符合蜕变关系。

噪点蜕变关系模式主要用于检验 OCR 系统基于噪声鲁棒性评估的能力。我们采用多种方式利用源输入图像构建生成衍生输入图像，具体流程如下：给定初始源输入为 P_1 ， P_n 为通过蜕变关系生成的衍生输入。将 P_1 和 P_n 分别输入待测 OCR 系统，得到对应输出结果 O_1 和 O_n 。基于蜕变关系理论， O_1 和 O_n 之间应满足特定的关系约束。根据蜕变测试原理，这两个输出应当符合预定义的逻辑约束。本文详细阐述了六种针对性的蜕变关系(MRs)，通过多维度测试来系统评估 OCR 技术的稳定性和可靠性，具体实现方案如下所述。

(噪点蜕变关系模式) MR1.1~MR3.1：定义原始输入图像为 P_1 ，通过对 P_1 添加噪声分量 E 构造扰动衍生样本 P_n ，即 $P_n = Add_E + P$ 。再将原始图像 P_1 和衍生样本 P_n 通过光学字符识别系统得到基准输出 O_1 和衍生输出 O_n ，根据系统稳定性要求，二者应当满足等价关系： $O_1 = O_n$ 。

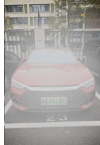
(空间蜕变关系模式) MR4.1：定义原始输入图像为 P_1 ，首先对其包含的字符单元进行完整分割，然后打乱顺序后拼接成完整图像生成衍生图像 P_n ，再将 P_1 和 P_n 通过光学字符识别系统得到原始识别结果 O_1 和衍生识别结果 O_n ，根据系统稳定性要求，二者应当满足等价关系： $O_1 = Order(O_n)$ 。

MR5.1：定义原始输入图像为 P_1 ，将 P_1 旋转一定角度生成衍生图像 P_n ，再将 P_1 和 P_n 通过光学字符识别系统得到原始识别结果 O_1 和衍生识别结果 O_n ，根据系统稳定性要求，二者应当满足等价关系： $O_1 = O_n$ 。

MR6.1：定义原始输入图像为 P_1 ，将 P_1 中的图像进行裁剪(不裁剪字符)生成衍生输入 P_n ，再将 P_1 和 P_n 通过光学字符识别系统得到原始识别结果 O_1 和衍生识别结果 O_n ，根据系统稳定性要求，二者应当满足等价关系： $O_1 = O_n$ 。

表 1 对上述 MR 给出了简单的解释和示例。

Table 1. License plate recognition MR explanation and simple example
表 1. 车牌识别 MR 解释和简单示例

MRs	Ori	MR1.1	MR2.1	MR3.1	MR4.1	MR5.1	MR6.1
违反 MR 的解释说明	原图	OCR 系统对天气因素具有高敏感性	OCR 系统对环境具有高敏感性	OCR 系统对脏污具有高敏感性	OCR 系统对字符乱序具有高敏感性	OCR 系统对字符角度具有高敏感性	OCR 系统对图像大小具有高敏感性
用于构造衍生输入的操作	原始	加入天气噪点 (模拟下雨)	加入环境噪点 (模拟夜晚)	加入脏污噪点 (模拟镜头脏污)	打乱字符顺序	旋转图片至一定角度	放大或缩小图片
源输入/衍生输入							
源输出/衍生输出	浙 AFR1397	浙 AFR1397	浙 AFR1397	浙 AFR1397	浙 AFR7391	浙 AFR1397	浙 AFR1397

7. 实验设计与结果分析

7.1. 目标系统

根据 IDC 发布的《中国公有云服务市场研究报告(2024H1)》[29], 见图 3, 阿里云市场份额在中国公有云服务市场长期稳居榜首, 市场占有率高, 并且在细分 OCR 领域也有不俗的表现和较高的市场份额。阿里云、腾讯云和华为云是最受市场认可的前三云服务。因此, 从为贴近用户实际使用场景, 实验对象为当前主流的三大系统。

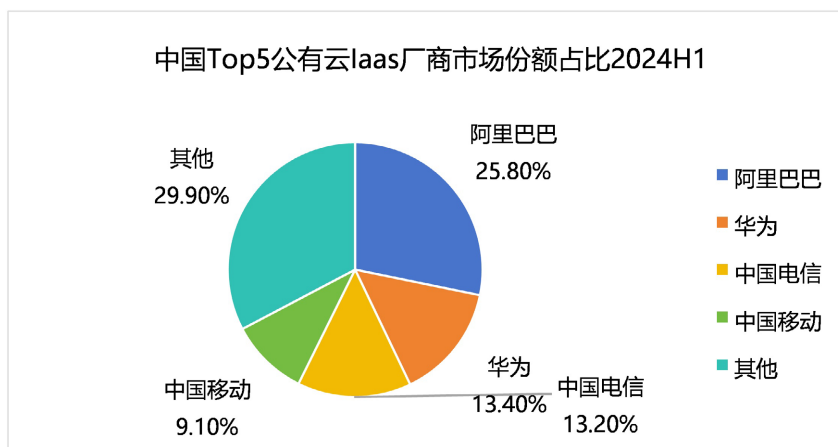


Figure 3. Market share of AI public cloud services in China in H1 2024
图 3. 2024H1 中国 AI 公有云服务市场份额

- 1) 华为云车牌识别: 支持车牌信息、车牌颜色识别, 支持双行车牌识别, 支持单张图片内多个车牌识别。
- 2) 腾讯云车牌识别: 提供了强大的车牌识别 API, 可以自动定位和识别中国大陆机动车的车牌, 返回车牌号码、颜色及其他相关信息。该 API 支持多种图片格式, 并且可以通过 URL 或 Base64 编码的方式上传图片。
- 3) 阿里云车牌识别: 可以准确识别出图像中车牌位置, 输出车牌位置坐标、车牌类型、车牌号码、车牌号码置信度、车牌置信度。

7.2. MR 的数据集和源输入

我们使用了 CCPD 数据集[13]和 CBLPRD-330K [14]数据集用于准备车牌的源输入。CCPD 数据集是大型车牌开源数据集, 其中包括了绿牌车和蓝牌车的 25 万多幅中国城市车牌图像和车牌检测与识别信息的标注。CBLPRD-330K 数据集通过生成对抗网络(GAN)技术来生成图片, 确保了图像质量的同时, 也保证了不同类型车牌分布均衡, 解决了实际采集中常见的样本不平衡问题。它被广泛应用于交通监控系统、自动驾驶汽车和安防领域多个场景, 总计 33 万余张图像。

对于我们定义的两类蜕变关系模式下的六种蜕变关系, 每个 MR 都被应用于产生相关的后续输入。请注意, MR 不能用于所有的源输入, 因为某些输入可能不满足用于构造后续输入的条件, 例如雨雪天被遮挡的车牌, 因此在以上数据集中仅采用基础高质量数据集, 以便进行后续的噪声及空间处理。

7.3. 实验结果分析

本节对三个 OCR 系统产生的蜕变测试结果进行了分析, 以此评估系统性能。我们使用违反率(violation

rate, VR)作为评估指标, 违反率可以计算如下:

$$VR = \frac{V}{P_c}$$

其中 P_c 是总输入数, V 是系统违反相关 MR 的数量。对于给定的 MR, VR 揭示了系统违反该 MR 的程度。VR 取值范围为从 0 到 1, $(1 - VR)$ 可以反应系统的准确度相似信息。较高的 VR 值意味着 OCR 系统在此 MR 下表现欠佳, 违反了我们定义的 MR。这意味着我们可以从用户的角度去发现软件缺陷, 判断它在何种程度上满足用户需求, 加强对系统的理解从而对系统进行更好的利用和完善; 反之, VR 值越低, 说明它能够更好地满足用户的需求。VR 值为 0 意味着在被测系统没有显示违规, 可能在此方面具有鲁棒性, 不受噪声或扰动的干扰。

表 2 展示了 MT 方法与基于数据集的评估方法的比较。在车牌识别方面, 三种主流 OCR 系统在 6 个蜕变关系下的表现: 腾讯云复杂光照方面展现出优势, 华为云针对天气变化的抗干扰能力较强, 而阿里云更适用距离不可靠的场景。基于数据集的评估方法依赖于数据集的真实结果来确认 OCR 任务输出的正确性, 我们使用了 7.2 节中介绍的数据集提供的输入和相应的标注对比结果。

Table 2. Accuracy of MT and dataset based evaluation method
表 2. MT 的准确率和基于数据集的评估方法

(%)	MT (1 - VR)						基于数据集的评估方法(VR)					
	MR1.1	MR2.1	MR3.1	MR4.1	MR5.1	MR6.1	MR1.1	MR2.1	MR3.1	MR4.1	MR5.1	MR6.1
AliOCR	90.79	91.20	94.37	90.95	95.32	96.58	99.50	99.45	98.77	97.43	99.64	99.32
TencentOCR	90.23	89.60	92.86	89.91	94.72	97.54	98.66	99.84	96.59	98.27	98.62	99.46
HuaweiOCR	92.75	90.68	90.83	90.75	94.85	97.79	99.46	99.03	97.64	98.45	99.73	99.21

MT 的准确度值是通过 $1 - VR$ 获得的。从表 2 中可以发现, 对于每个被评估的 OCR 任务, 与基于数据集的评估方法相比, MT 报告的准确度值较低。这些结果表明, 与评估方法相比, 我们的方法可以揭示系统的故障。

8. 总结

随着技术发展, 光学字符识别(OCR)技术作为一项关键的智能化信息处理技术, 已在众多行业实现规模化应用, 并成为不可或缺的工具。然而, 现有研究通过多种技术手段对 OCR 系统进行评估, 暴露了其在多个方面存在的问题。本文聚焦于将蜕变测试技术引入 OCR 系统的评估中, 结合用户对 OCR 系统的多样化需求, 提出了两种蜕变关系模式及七种蜕变关系, 从用户视角揭示了 OCR 系统在语言理解能力上的表现。

当前聚焦车牌识别场景, 因其格式标准化程度高且错误容忍度低, 可为 MRs 设计提供清晰验证路径。未来将扩展至票据、街景文本等场景, 重点围绕以下维度深入探索: 首先, 研究者可以进一步优化蜕变测试的流程, 以提升评估的准确性和全面性, 例如探索更有效的替换方法或蜕变关系, 以检测系统在不同层面的缺陷。其次, 可以改进测试方法, 增强其通用性和适用性, 使其能够适应更多场景和需求。

参考文献

[1] Du, Y., Li, C., Guo, R., *et al.* (2020) Pp-ocr: A Practical Ultra Lightweight OCR System.
[2] Agbemenu, A.S., Yankey, J. and Addo, E.O. (2018) An Automatic Number Plate Recognition System Using OpenCV

- and Tesseract OCR Engine. *International Journal of Computer Applications*, **180**, 1-5. <https://doi.org/10.5120/ijca2018917150>
- [3] Ugale, M.K., Patil, S.J. and Musande, V.B. (2017) Document Management System: A Notion Towards Paperless Office. 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM), Aurangabad, 5-6 October 2017, 217-224. <https://doi.org/10.1109/icisim.2017.8122176>
 - [4] Siek, M. and Soeharto, R. (2021) Developing Automated Optical Character Recognition System Using Machine Learning Algorithm to Solve Payment Verification Issues. 2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS), Makasar, 25-26 October 2021, 1-6. <https://doi.org/10.1109/icoris52787.2021.9649514>
 - [5] Segura, S., Fraser, G., Sanchez, A.B. and Ruiz-Cortes, A. (2016) A Survey on Metamorphic Testing. *IEEE Transactions on Software Engineering*, **42**, 805-824. <https://doi.org/10.1109/tse.2016.2532875>
 - [6] Zhou, Z.Q., Xiang, S. and Chen, T.Y. (2016) Metamorphic Testing for Software Quality Assessment: A Study of Search Engines. *IEEE Transactions on Software Engineering*, **42**, 264-284. <https://doi.org/10.1109/tse.2015.2478001>
 - [7] Zhou, Z.Q., Sun, L., Chen, T.Y. and Towey, D. (2020) Metamorphic Relations for Enhancing System Understanding and Use. *IEEE Transactions on Software Engineering*, **46**, 1120-1154. <https://doi.org/10.1109/tse.2018.2876433>
 - [8] Deng, Y., Zheng, X., Zhang, T., Lou, G., Liu, H. and Kim, M. (2020) RMT: Rule-Based Metamorphic Testing for Autonomous Driving Models.
 - [9] Cao, Y., Zhou, Z.Q. and Chen, T.Y. (2013) On the Correlation between the Effectiveness of Metamorphic Relations and Dissimilarities of Test Case Executions. 2013 13th International Conference on Quality Software, Naging, 29-30 July 2013, 153-162. <https://doi.org/10.1109/qsic.2013.43>
 - [10] Zhou, Z.Q. (2010) Using Coverage Information to Guide Test Case Selection in Adaptive Random Testing. 2010 IEEE 34th Annual Computer Software and Applications Conference Workshops, Seoul, 19-23 July 2010, 208-213. <https://doi.org/10.1109/compsacw.2010.43>
 - [11] Barus, A.C., Chen, T.Y., Grant, D., Kuo, F. and Lau, M.F. (2011) Testing of Heuristic Methods: A Case Study of Greedy Algorithm. In: Huzar, Z., Koci, R., Meyer, B., Walter, B. and Zendulka, J., Eds., *Software Engineering Techniques*, Springer, 246-260. https://doi.org/10.1007/978-3-642-22386-0_19
 - [12] Chen, T.Y., Kuo, F., Liu, H. and Wang, S. (2009) Conformance Testing of Network Simulators Based on Metamorphic Testing Technique. In: Lee, D., Lopes, A. and Poetzsch-Heffter, A., Eds., *Formal Techniques for Distributed Systems*, Springer, 243-248. https://doi.org/10.1007/978-3-642-02138-1_19
 - [13] Xu, Z., Yang, W., Meng, A., Lu, N., Huang, H., Ying, C., et al. (2018) Towards End-to-End License Plate Detection and Recognition: A Large Dataset and Baseline. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision-ECCV 2018*, Springer International Publishing, 261-277. https://doi.org/10.1007/978-3-030-01261-8_16
 - [14] China-Balanced-License-Plate-Recognition-Dataset-330K (2023). <https://github.com/SunlifeV/CBLPRD-330k>
 - [15] Liu, W., Chen, C., Wong, K.Y.K., Su, Z. and Han, J. (2016) STAR-Net: A Spatial Attention Residue Network for Scene Text Recognition.
 - [16] Baek, J., Kim, G., Lee, J., Park, S., Han, D., Yun, S., et al. (2019) What Is Wrong with Scene Text Recognition Model Comparisons? Dataset and Model Analysis. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, 27 October-2 November 2019, 4714-4722. <https://doi.org/10.1109/iccv.2019.00481>
 - [17] Xu, G., Ding, W., Fu, W., Wu, Z. and Liu, Z. (2021) Robust Learning for Text Classification with Multi-Source Noise Simulation and Hard Example Mining. *Applied Data Science Track: European Conference, ECML PKDD 2021*, Bilbao, 13-17 September 2021, 285-301. https://doi.org/10.1007/978-3-030-86517-7_18
 - [18] Todorov, K. and Colavizza, G. (2022) An Assessment of the Impact of OCR Noise on Language Models. *Proceedings of the 14th International Conference on Agents and Artificial Intelligence*, Volume 2, 674-683. <https://doi.org/10.5220/0010945100003116>
 - [19] Wang, W., Huang, J., Huang, J., Chen, C., Gu, J., He, P., et al. (2023) An Image Is Worth a Thousand Toxic Words: A Metamorphic Testing Framework for Content Moderation Software. 2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE), Luxembourg, 11-15 September 2023, 1339-1351. <https://doi.org/10.1109/ase56229.2023.00189>
 - [20] Wildandyawan, A. and Nishi, Y. (2020) Object-Based Metamorphic Testing through Image Structuring.
 - [21] Barr, E.T., Harman, M., McMinn, P., Shahbaz, M. and Yoo, S. (2015) The Oracle Problem in Software Testing: A Survey. *IEEE Transactions on Software Engineering*, **41**, 507-525. <https://doi.org/10.1109/tse.2014.2372785>
 - [22] Chen, T.Y., Kuo, F., Liu, H., Poon, P., Towey, D., Tse, T.H., et al. (2018) Metamorphic Testing: A Review of Challenges and Opportunities. *ACM Computing Surveys*, **51**, 1-27. <https://doi.org/10.1145/3143561>
 - [23] Zhou, Z.Q., Tse, T.H., Kuo, F.-C. and Chen, T.Y. (2007) Automated Functional Testing of Web Search Engines in the Absence of an Oracle. Department of Computer Science, The University of Hong Kong, Tech. Rep. TR-2007-06.

-
- [24] Zhou, Z.Q., Sun, L., Chen, T.Y. and Towey, D. (2020) Metamorphic Relations for Enhancing System Understanding and Use. *IEEE Transactions on Software Engineering*, **46**, 1120-1154. <https://doi.org/10.1109/tse.2018.2876433>
- [25] Jin, D., Chen, Y., Lu, Y., Chen, J., Wang, P., Liu, Z., *et al.* (2021) Neutralizing the Impact of Atmospheric Turbulence on Complex Scene Imaging via Deep Learning. *Nature Machine Intelligence*, **3**, 876-884. <https://doi.org/10.1038/s42256-021-00392-1>
- [26] 蒋一纯, 刘云清, 詹伟达, 朱德鹏. 基于图像退化模型的红外与可见光图像融合方法[J]. 电子与信息学报, 2022, 44(12): 4405-4415.
- [27] Zhang, M., Zhang, Y., Zhang, L., Liu, C. and Khurshid, S. (2018) Deeproad: Gan-Based Metamorphic Testing and Input Validation Framework for Autonomous Driving Systems. *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*, Montpellier, 3-7 September 2018, 132-142. <https://doi.org/10.1145/3238147.3238187>
- [28] Zhu, J.-Y., *et al.* (2020) Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. <https://arxiv.org/abs/1703.10593>
- [29] IDC: AI 助力 2024 上半年中国公有云市场回暖[Z/OL]. <https://my.idc.com/getdoc.jsp?containerId=prCHC52690424>, 2024-10-29.