ISBN: 978-981-18-5852-9
WCSE 2022 Spring Event: 2022 9th International Conference on Industrial Engineering and Applications
doi: 10.18178/wcse.2022.04.069

Multi-dimensional Metamorphic Testing Method in Vehicle Intelligent Identification System

Jiaze Sun[1,2], Yanyan Zhang[1,2] , Shuyan Wang[1,2] and Wei Wu[1,2]

[1]School of Computer Science and Technology

[2]Xi'an University of Posts and Telecommunications

Email: sunjiaze@xupt.edu.cn, 18893722140@163.com, wsylxj@126.com, 939498966@qq.com

Abstract. The image recognition systems, like other Deep Learning-driven systems, face the problem of Oracle Problem. In order to solve the problem of a single metamorphic relation in the image recognition system testing, a multi-dimensional metamorphic testing method is proposed. First, the multi-dimensional metamorphic relations of general image scene and transformation image background region is constructed; Then Then the derived test images are generated by image transformation technology, and the authenticity and naturalness of the derived test images are detected by introducing structural similarity. Finally, the test results are analyzed by Mean Absolute Error. Experiments conducted on Baidu AI service platform, Jingdong AI open platform and Aliyun visual intelligence open platform show that the multi-dimensional metamorphic testing method detects 3724, 2432 and 2747 inconsistencies in the identification results of the three platforms, and each metamorphic relation in the multi-dimensional metamorphic testing method can detect hundreds of inconsistent behaviors in recognition results. The image recognition system is tested more comprehensively in different image scenes, and the defects of the system are exposed more easily.

Keywords: metamorphic testing; metamorphic relation; image recognition systems; oracle problem.

1. Introduction

In recent years, image recognition systems have been widely used in people's daily lives. Reasonable verification and evaluation of the accuracy of image recognition systems can maximize the security and stability of image recognition products, similar to other deep learning-driven systems, The image recognition system also has the problem of Oracle Problem [1]. In order to solve the problem of Oracle Problem, Chen et al. [2] of the Chinese University of Hong Kong proposed a metamorphic testing method to judge the correctness of the tested software through whether the program attributes are satisfied. In recent years, the application of metamorphosis test in metamorphosis relation construction and other fields has made great progress., including Web services [3], Embedded Systems [4] and other research fields. In the development practice, some real program defects have been found by the metamorphic testing, such as GCC and LLVM compiler [5], Google, Bing search engine [6]. Inspired by the metamorphic testing paradigm of traditional software engineering [7], a new metamorphic testing technique is proposed for deep learning-driven system software that takes images as input. For example, Dwarakanath et al. [8] proposed a metamorphic test method for image classification systems. They constructed metamorphic relations by changing the order of input image channels, normalized test data, and scaled test data. However, such metamorphic relationships, which have good test performance in other fields, have not been reused in image recognition system testing work. Based on GAN, Jiang Jingjie et al. [9] transformed the original data into rainy scenes, snow scenes and night scenes to construct a metamorphic relationship. However, the inconsistent behavior of the image recognition system is only detected in three scenarios, and its ability to detect code defects is limited. DeepTest [10] is a metamorphic testing method for autonomous driving systems. This method generates derived test images through brightness adjustment, scaling, translation and other transformations of marked driving scene images, and then detects whether the operation of automatic driving system is consistent between the original scene and the transformed scene. Although these methods can detect systematic errors to a certain extent, most of the transformation techniques used to generate derived test images work on the entire image and do not consider part of the image transformation, such as changing only the background area of the image. Therefore, It is impossible to measure the influence of image background transformation on image recognition results.

This paper proposes a multi-dimensional metamorphic testing method for image recognition systems, which starts from multiple dimensions and automatically generates a variety of derived test images based on multi-dimensional metamorphic relations to detect inconsistent behaviors of image recognition systems. From the overall test process, the influence of the multi-dimensional metamorphic testing method on the recognition results of the image recognition system is evaluated.

2. Metamorphic Tesing Method of Image Recognition System

2.1. Construction of multi-dimensional metamorphic relations

In this paper, combined with the basic requirements for the construction of effective metamorphic relations and the geometric properties of images in the literature [11], three types of metamorphic relations, namely general, transformed image scene and transformed image background region, are constructed. The core content is: while maintaining the recognition target Under the condition that it does not change, the corresponding image transformation is performed on the image according to the semantic information of the image. The image transformation here can be adding fog, adding rain, adding sunlight, or transforming other image information irrelevant to the recognition target. If the image and the transformed image obtain the same target recognition result, the image transformation satisfies the metamorphic relations. Based on three dimensions, this paper establishes the metamorphic relation construction method suitable for image recognition systems:

(1) General metamorphic relationship

In the field of image, image processing methods such as rotation, scaling and disruption of RGB channel order are often used to construct metamorphic relations [13]. These general metamorphic relations have good test effect in image classification systems, so they are used in image recognition systems. and expand it to construct general metamorphic relation.

(2) The metamorphic relations of the transformed image scene

The image recognition system based on the deep learning model only relies on the general metamorphic relations, and it is difficult to achieve a more adequate test. In order to detect whether the data input of the image recognition system can correctly identify the target in different scenes, it is necessary to simulate the real scene to generate test images, and transform different image scenes for the original test images. Construct the transformed image scene metamorphic relations.

(3) The metamorphic relations of the background area of the transformed image

Most transformation techniques used to generate derived test images only work on the entire image, without considering partial transformations of the image (e.g., changing only the background regions of the image). Therefore, it is impossible to measure the effect of changes in the image background area on the image recognition results.

Table 1. Metamorphic relations and corresponding derived image datasets

Type	Metamorphic Relation	Content	Derived Image Dataset
General metamorphic relationship	MR1	Changed BGR channel ordering.	BRG、RBG、RGB、GBR、GRB
	MR2	Add Gaussian blur.	blur
	MR3	Rotate by a certain angle.	rotate
	MR4	horizontal flip	flip
Transform image scene	MR5	Image scene transition to foggy day	fog
	MR6	Image scene transition to rainy day	rain
	MR7	Add sunlight effect	sun
Transform image background area	MR8	Increase the brightness of the background area	B_up
	MR9	Reduce the brightness of background areas	B_down
	MR10	Add Gaussian blur to background area	B_blur

	MR11	Add Gaussian noise to the background area	B_noise

2.2. Generate Derived Test Images

After the metamorphic relations is defined, the derived test images corresponding to each metamorphic relations can be generated in batches through the original test images. The constructed general metamorphic relations, when generating the derived test image, uses the related functions of the OpenCV in Python to realize the image input RGB channel order change, Gaussian blur, rotation and horizontal flip effects. Derived test images are generated by transforming image scenes, MUNIT [12] is used as the generative model for constructing images, and the CycleGAN [13] model is used for image style transfer. When transforming the image background area, it is necessary to separate the recognition target in the image from the image background. The real-time instance segmentation model YOLACT [14] is used, which mainly realizes instance separation through two parallel sub-networks and can separate object and object accurately and quickly

2.3. Verification of test results

According to the defined metamorphic relations, it is judged whether the original test image and the derived test image have the same recognition results in the same image recognition system. In this paper, the IoU evaluation standard for detection and recognition accuracy in the field of target detection is introduced [15]. The calculation formula is shown in formula (1):

$$IoU = \frac{A \cap B}{A \cup B} \tag{1}$$

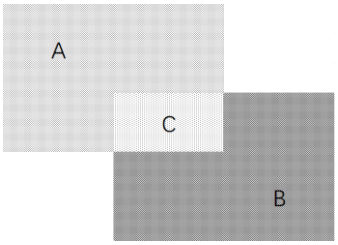

Fig 1 Schematic diagram of IOU

A is the area where the original test image identifies the target in the image recognition system, B is the area where the derived test image identifies the target in the image recognition system, and C represents the intersection of the two. IoU describes the degree of overlap between two boxes. When IoU is 0, the two boxes do not overlap and have no intersection. When IOU is 1, the two boxes overlap completely. The value of IOU ranges from 0 to 1, indicating the degree of overlap between two enclosures. A higher value indicates a higher degree of overlap. In this paper, if the IoU of the target identified in the original test image is greater than 0.8 after identification in the derived test image, it is considered to meet the metamorphic relationsship.

Mean Absolute Error

The number of inconsistent behaviors identified in the test results of the image recognition system can directly reflect the probability of software error. If the image recognition system violates the metamorphic relations, the number of inconsistent behaviors of the system is increased by 1. Furthermore, MAE of the test results of the image recognition system in the derived test data set and the original test data set, is calculated to determine the recognition ability of the image recognition system. The calculation formula is shown in Formula (2):

$$MAE = \frac{\sum_{i=1}^{N} |r_i - s_i|}{N} \tag{2}$$

where r_i represents the test result of the i-th image in the original test data set, s_i represents the test result of the i-th image in the derived test data set, there are N test images in the data set, the absolute value of the sum difference is calculated r_i and s_i summed; The higher the probability of wrong recognition by the image recognition system, the worse the recognition ability of the image recognition system for this type of image transformation;

3. Experiment and Result Analysis

Vehicle recognition is the most widely used in the field of image recognition, and plays an important role in automatic driving systems, intelligent transportation systems, and automatic parking management systems. The experimental data selects the open driving dataset BDD100K [16] released by the University of Berkeley as the original test image. In order to evaluate the performance of the multi-dimensional metamorphic testing method in detecting the inconsistency of the image recognition system, for the whole test process, we start from the following three research questions, and perform corresponding analysis experiments to verify the effectiveness of the multi-dimensional metamorphic testing method:

Question 1: Does the multi-dimensional metamorphic test method generate high-quality and natural-looking test images?

Question 2: Can the multidimensional metamorphic testing method test real-life image recognition systems? How did the test work?

Question 3: Compared with the GAN-based image recognition system metamorphic testing method, does the multi-dimensional metamorphic testing method conduct a more comprehensive test on the test object?

3.1. Experimental process

This paper uses the following process to evaluate the accuracy of the image recognition system.

(1) Data preparation. As there are fewer vehicles in some original test images, this experiment aims to identify the vehicles in the images, so the original test images are cleaned and the images with more vehicles are reserved as the original test images.

(2) The recognition results of original test images are obtained by running the image recognition system, and the number of target vehicles recognized in each image is counted automatically.

(3) According to the image transformation technique described in Section 2.3, combined with the constructed metamorphic relation, a derived test image is generated.

(4) The derived test image is taken as the input of the image recognition system, and the image recognition system is re-run to verify whether the output of the derived test image is consistent with the expected result, and calculate the MAE.

(5) Combined with the test results of three tested image recognition platforms, the change of recognition results and recognition indicators is analyzed.

3.2. Analysis of experimental results

3.2.1. Image structure similarity analysis

Question 1: Does the multi-dimensional metamorphic test method generate high-quality and natural-looking test images?

In order to evaluate the quality of derived test images, structural similarity is used to evaluate the quality and naturalness of derived test images. The structural similarity is measured from the three key features of brightness, contrast and structure, and the value is between 0 and 1. The structural similarity of the derived test image and the corresponding original test image is calculated. The closer to 1, the higher the similarity of the two images is, and the higher the quality of the composite image is, the more natural it is. After comparing each derived test data set with the original test data set, the obtained structural similarity is shown in Table 2:

Table 2. Derived Test Dataset Structural Similarity

Derived Image Dataset	SSIM	Derived Image Dataset	SSIM
BRG	0.99	rainy	0.73
GBR	0.98	flip	0.43
GRB	0.99	sunshine	0.76
RBG	0.96	B_up	0.93
RGB	0.98	B_down	0.86

blur	0.88	B_blur	0.78
foggy	0.78	B_noise	0.84

The table 2 shows that 13 derived test data set structure similarity is very high (more than 0.7 and is close to 1), which the structural similarity of flip derived test data set is 0.43, the data set is based on the original test image flip horizontal. Therefore, the content structure of the image changes, resulting in a low structural similarity value. Rotate Derived test data set is obtained by rotating the original test image. Therefore, the image will be reduced in a certain proportion, resulting in the size difference between the derived test image and the original test image. When calculating the structural similarity of two images, SSIM requires that the two images have the same number of channels and the same size. Therefore, SSIM values of two images cannot be calculated. In general, the derived test images generated in this experiment are of high quality and natural.

3.2.2 Analysis of Intelligent Platform Test Results

Question 2: Can the multidimensional metamorphic testing method test real-life image recognition systems? How did the test work?

In order to test whether the proposed test method can test the image recognition system in real life, the vehicle recognition application is taken as an experimental case. The original test data set consists of 200 images with more than 5 vehicles in BDD100K, and the test object is selected from the domestic intelligent open platform: Baidu AI service platform [17], Jingdong AI open platform [18] and Aliyun visual intelligence open platform [19] (The Baidu Platform, Jingdong Platform and Aliyun Platform are respectively referred to below). If the image recognition system violates the metamorphic relation, the number of inconsistent behaviors of the system increases by 1. The number of inconsistent behaviors and MAE in the identification results are automatically counted to evaluate the testing effect of multidimensional disintegration testing method. Input the original test image and the derived test image into the test object respectively, and count the test results. Table 3 shows the number of behaviors identified inconsistency in the test results of the three intelligent platforms detected.

Table 3. The number of inconsistent behaviors in the test results of each platform

MR		The number of inconsistent behaviors in the test results (%)		
		Baidu Platform	Jingdong Platform	Aliyun Platform
General metamorphic relationship	MR1	160(9.7%)	201(16.0%)	312(19.2%)
	MR2	242(14.3%)	163(12.6%)	185(11.8%)
	MR3	418(**24.5%**)	316(**24.4%**)	395(**24.2%**)
	MR4	128(7.1%)	193(14.9%)	175(10.5%)
Transform image scene	MR5	316(18.5%)	116(8.66%)	373(**22.9%**)
	MR6	420(**25.1%**)	219(16.5%)	203(12.4%)
	MR7	235(13.7%)	95(7.0%)	237(14.2%)
Transform image background area	MR8	303(17.9%)	204(15.7%)	153(9.3%)
	MR9	427(**25.1%**)	216(16.5%)	152(9.3%)
	MR10	502(**29.9%**)	434(**33.8%**)	365(**22.3%**)
	MR11	573(**34.1%**)	275(**21.2%**)	197(11.8%)

.

The MAE of the general metamorphic relation among the test subjects is shown in Fig. 2:

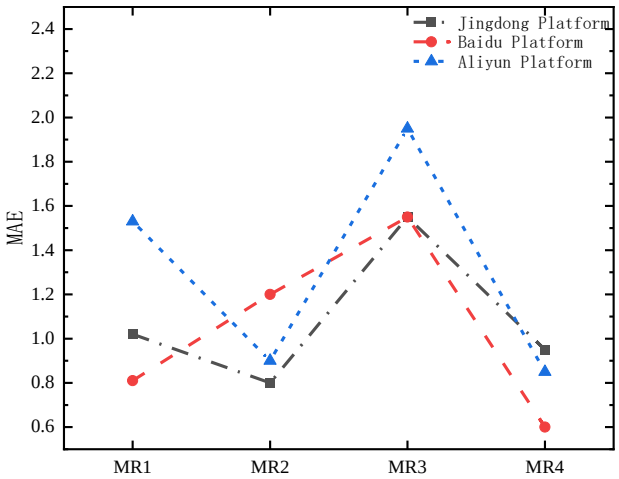

Fig. 2: General metamorphic relationship

It can be seen from Figure 2 that the MAE of Baidu platform fluctuates between 0.5 and 1.05, jingdong platform between 0.75 and 1.1, and Ali Cloud platform between 0.85 and 2.2. In other words, under the general metamorphic relationship, the recognition results of all platforms have large errors. Among them, in MR3, the MAE of Aliyun platform, Jingdong platform and Baidu platform are 2.2, 1.55 and 1.55 respectively, and a total of 1129 inconsistent recognition results were detected.

The MAE of the metamorphic relation constructed by transforming the image scene in the test object is shown in Fig. 3.

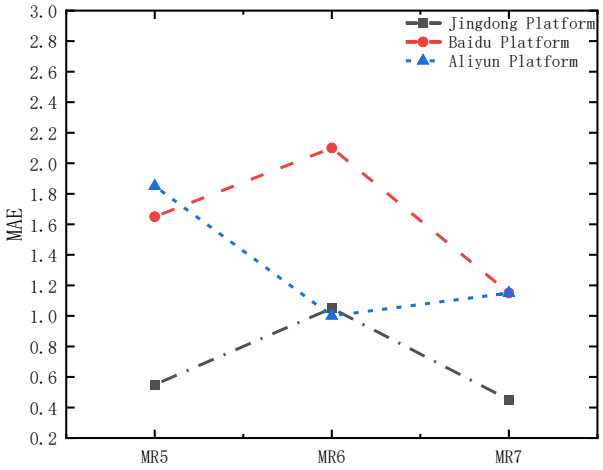

Fig. 3: Transform image scene

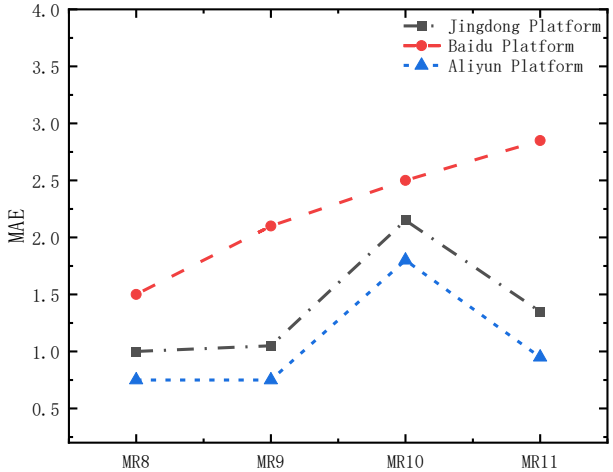

Fig. 4: Transform image background area

It can be seen from Figure 3 that under the metamorphic relations constructed by the transformed image scene, the MAE of the Baidu platform recognition results is 2.1 at the highest and 1.15 at the lowest. Inconsistent behaviors detected by Baidu platform, Alibaba Cloud platform and JD platform were 971, 813 and 430 respectively.

The MAE of the metamorphic relation constructed by transforming the background region of the image in the test object is shown in Fig. 4.

As can be seen from Figure 4, in the metamorphic relations constructed by transforming the background area of the image, the MAE of Baidu platform is between 1.5 and 2.85, the MAE r of Jingdong platform is between 1 and 2.15, and the MAE of Aliyun platform is between 0.75 and 1.95, the number of inconsistent behaviors detected by the distribution is 1805, 1129 and 867.

Question 3: Compared with the GAN-based image recognition system metamorphic testing method, does the multi-dimensional metamorphic testing method conduct a more comprehensive test on the test object?

The experimental results and conclusions of the multi-dimensional metamorphic testing method and the GAN-based image recognition system metamorphic testing method are compared and analyzed to judge whether the multi-dimensional metamorphic testing method is more comprehensive. Jiang Jingjie et al. [6] constructed the metamorphic relationship based on GAN to generate rainy day scene, snow scene and night scene. According to the experimental results, Baidu platform and Jingdong platform have the same recognition ability. However, the experimental results of the multi-dimensional metamorphic testing method show that different platforms perform differently under different metamorphic relations. For example, compared with the other two intelligent platforms, the Baidu platform has the smallest MAE of the recognition results in the general metamorphic relations, However, the MAE is the largest in the metamorphic relations constructed in the background region of the transformed image, that is, the number of inconsistent behaviors detected in the recognition results is more. To sum up, the method in this paper detects the vehicle detection system of three intelligent platforms from multiple test dimensions in a more comprehensive way.

4. Conclusion

Aiming at the problem of a single metamorphic relationship in the image recognition system testing work, this paper proposes a multi-dimensional metamorphic test method, and constructs a total of 11 metamorphic relationships from three aspects. The experimental results show that the method can effectively detect the inconsistent behavior of the image recognition system in different image changing scenes, and the accuracy test of the image recognition system is more comprehensive. In the future work, we will continue to study the testing of system software driven by deep learning, and apply the established metamorphic relations to image recognition systems in other fields to further optimize the authenticity of the derived test images.

5. References

[1] Weyuker E J. On testing non-testable programs[J]. The Computer Journal, 1982, 25(4): 465-470.

[2] Chen T Y. Metamorphic testing: A simple method for alleviating the test oracle problem[C]//2015 IEEE/ACM 10th International Workshop on Automation of Software Test. IEEE, 2015: 53-54.

[3] Sun C A,Wang G ,Mu B, et al. Metamorphic Testing for Web Services: Framework and a Case Study[C]// IEEE International Conference on Web Services. Washington D. C., USA: IEEE Press, IEEE, 2011:283-290.

[4] Jiang M, Chen T Y, Kuo F C, et al. Testing central processing unit scheduling algorithms using metamorphic testing[C] //2013 IEEE 4th International Conference on Software Engineering and Service Science. IEEE, 2013: 530-536.

[5] Le V, Afshari M, Su Z. Compiler validation via equivalence modulo inputs[J]. ACM Sigplan Notices, 2014, 49(6): 216-226.

[6] Zhou Z Q, Xiang S, Chen T Y. Metamorphic Testing for Software Quality Assessment:A Study of Search Engines[J]. IEEE Transactions on Software Engineering, 2016, 42(3):264-284.

[7] Zhang J M, Harman M, Ma L, et al. Machine Learning Testing: Survey, Landscapes and Horizons[J]. IEEE Transactions on Software Engineering, 2020, PP(99):1-1.

[8] Dwarakanath A, Ahuja M, Sikand S, et al. Identifying implementation bugs in machine learning based image

classifiers using metamorphic testing[C]// Proceedings of the 27th ACM SIGSOFT International Symposium on Software Testing and Analysis. 2018: 118-128.

[9] Jiang J J,Xu L,Li Ning. GAN-based image recognition system metamorphic testing method[J]. Computer And Modernization. 2021(02):24-29.

[10] Tian Y, Pei K, Jana S, et al. Deeptest: Automated testing of deep-neural-network-driven autonomous cars[C]// Proceedings of the 40th international conference on software engineering. 2018: 303-314.

[11] Chen T Y, Tse T H, Zhou Z Q . Fault-based testing without the need of oracles[J]. Information and Software Technology, 2003, 45(1):1-9.

[12] Huang X, Liu M Y, Belongie S, et al. Multimodal unsupervised image-to-image translation[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 172-189.

[13] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]// Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.

[14] Bolya D, Zhou C, Xiao F, et al. Yolact: Real-time instance segmentation[C]// Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9157-9166.

[15] Sun S Q, Zuo H W, ZHAO L T, et al. Research on Detection Box Optimization of Joint Feature Similarity Measurement and Intersection over Union[J]. Computer Knowledge and Technology, 2019,15(29):190-193.

[16] Yu F, Chen H, Wang X, et al. Bdd100k: A diverse driving dataset for heterogeneous multitask learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 2636-2645.

[17] Baidu AI Open Platform. [EB/OL].[2022-02-12]. https://ai.baidu.com/tech/vehicle/detect.

[18] Jingdong AI Open Platform. [EB/OL]. [2022-02-12]. https://neuhub.jd.com/gwtest/init/266

[19] Aliyun Visual Intelligence Open Platform. [EB/OL].[2022-02-12]. https://vision.aliyun.com/.