Indonesian Journal of Electrical Engineering and Computer Science
Vol. 27, No. 1, July 2022, pp. 198~205
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v27.i1.pp198-205

Image anomalies detection using transfer learning of ResNet-50 convolutional neural network

Zaid Taher Omer, Amel Hussein Abbas
Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq

Article Info

Article history:

Received Jul 21, 2021
Revised Apr 12, 2022
Accepted May 26, 2022

Keywords:

Artificial
Convolutional neural network
Deep learning
ResNet-50
Supervised

ABSTRACT

With the quick advancement of keen fabricating, information-based blame determination has pulled in expanding attention. As one of the foremost prevalent strategies of diagnosing errors, deep learning has accomplished exceptional comes about. Be that as it may, due to the truth that the estimate of the seeded tests is little in diagnosing mistakes, the profundities of the deep learning (DL) models for fault conclusion are shallow compared to the convolutional neural network in other regions (including ImageNet), which limits the accuracy of the final prediction. In this paper, ResNet-50 with a 25 convolutional layer depth has been proposed to diagnose anomalous images. Trained ResNet-50 applies ImageNet as a feature extractor to diagnose errors. It was proposed on three sets of data which are the bottle, the spoon, and the carton, and the proposed method was achieved. The prediction accuracy of the data set was 99%, 95% and 90%, respectively.

Corresponding Author:

Zaid Taher Omer
Department of Computer Science, College of Science, Mustansiriyah University
Baghdad, Iraq
Email: ZaidAlradi@uomustansiriyah.edu.iq

1. INTRODUCTION

A common requirement when analyzing sets of information in the real world is to recognize which cases stand out as distinct from others. Such cases are known as anomalies, and the aim of detecting anomalies is to distinguish between all of these cases in a data dependent manner [1]. Abnormal characteristics can be caused by fatal errors within the information but every now and then they hint at a modern process that was not already known [2]-[4]. One of the main variables in manufacturing improvement is the automatic detection of defects, which makes it possible to anticipate generation errors, and in this way it can improve the quality and produce economic edges to the plant. A common refinement for peculiarity discovery within the industry is for a machine to judge picture obtained through progressed camera or sensor. This can basically be a problem in detecting anomalies in the image that is looking for different designs than the normal photos [5]. People can handle this task easily by paying attention to typical designs, but this can be somewhat annoying for machines [6].

Deep learning enhances traditional machine learning by adding additional "depth" (complexity) to the model and changing the data using various functions that allow data representation in a hierarchical manner, through many levels of abstraction. Some strategies more often than not consider peculiarity discovery a one-class issue that first identifies typical information as a basis and then assesses whether or not the test information has a place for that baseline, by the degree of contrast from the baseline [7]. Within the early surface applications deformation location, such as ceramics, nails, spoons, and other industrial products, the foundation was designed regularly by planning carefully grouped highlights on impeccable information.

Liu *et al*. [8] utilized support vector data description (SVDD) to distinguish abandons in thin-film transistor liquid crystal display (TFT-LCD) cluster pictures. They portrayed an picture fix with four highlights counting entropy, vitality, differentiate, and homogeneity and prepared an SVDD demonstrate utilizing ordinary picture patches [6]. On the off chance that a include vector lies exterior super set by SVDD all through testing, the picture fix appreciate this include vector is taken under consideration irregular.

Abdel-Qader *et al*. [9] In this project, bridge deck images were examined for the purpose of automating crack detection using a PCA-based framework using three different algorithms in an effort to improve the accuracy of the results. The three methods are, i) principal component analysis (PCA) is performed on the data directly, ii) PCA is performed on the data after linear features are detected, and finally, iii) PCA is performed with only features detected on a small block of data. A set of 10 bridge deck images, 5 cracked and 5 non-fractured, were trained. Forty other images in the database were used as test images. These include 20 cracked images and 20 non-cracked images. PCA results alone are 30% false-negative, 12.5% false-positive, and 57.5% an overall correct identification. The use of linear structure detectors prior to PCA improved overall true identification to 60%, reduced false positives to 0%, and reduced false negatives to 20%. Applying local processing in the algorithm increased the overall true determination to 73%, reduced false negatives to 12.5%, and increased false positives to 15%. Napoletano *et al*. [10] utilized a pre-trained ResNet-18 to extricate highlight vectors from the filtering magnifying lens (SEM) pictures to build a lexicon. Within the forecast section, a tried picture is taken into account abnormal if the typical geometrician separate between its highlight and its m closest neighbors within the lexicon is on top of the edge.

Ruff *et al.* [11] in this work. Our method, deep SVDD, jointly trains a deep neural network while optimizing the supershell that contains the data in the output space. Through this deep SVDD extracts the common factors of difference from the data. We have demonstrated theoretical properties of our method such as the property that allows incorporation of a prior assumption regarding the number of outliers present in the data. Our experiments demonstrate both quantitatively and qualitatively the audio performance of Deep SVDD, on modified national institute of standards and technology database (MNIST) and CIFAR-10 image benchmark datasets as well as on the detection of adversarial examples of GTSRB stop signs.

Sinha and Fieguth [12] introduced 2 crack detectors for characteristic crack items in buried concrete pipes, then a linking and improvement operation were afterward performed to attach crack items. Iyer and Sinha took advantage of the linear property of crack options and planned morphology-based filters with linear structuring components to discover cracks [13]. Chae and patriarch relied on a neural network for classifying pipe defects [14], wherever image information were directly input to retrieve the attributes of defects.

Yang *et al*. [15] have anticipated a picture investigation technique to capture thin breaks and minimize the require for write checking in ferroconcrete basic tests. They require utilized the thinks about like split profundity expectation [16], revision in discovery whereas not picture enlistment, split design acknowledgment upheld manufactured neural systems [17], applications to micro-cracks of rocks [18], and conservative sub-pixel measurement mensuration [19]. Stereo triangulation strategy was the embraced strategy backed barrel equation estimation and picture correction. Once they require the rectified yield, the surface of the decided locales may be unfurled and given in a really plane picture for taking after uprooting and distortion investigation. From that the break location was analyzed.

Rodríguez-Martín *et al*. [20] have planned associate degree infrared (IR) diagnostic procedure technique supported IR image rectification with the extraction of Isotherms that permits the detection of cracks moreover because the geometric characterization and orientation of the crack to help the prediction of the direction of propagation of the crack through the fabric. It permits the quick and straightforward assessment of the morphology of various cracks (toe crack and longitudinal crack). the applying of think about with IR camera and afterward picture correction that was utilized in their proposition licenses the geometric characterization of the surrenders encouraging their classification in step with the measures [21]. The detection of the crack victimisation the notches within the inconsistencies was planned by Broberg *et al*. [22]. Here, victimisation the IR diagnostic procedure picture rectification strategy, they need identified supported notches which is able to disagree counting on the temperature.

2. METHOD

We usually increase the number of layers in deep learning to increase the accuracy of the results, but unfortunately, experiments have proven that a large increase in layers leads to the emergence of a problem in the training and testing processes called Vanishing/Exploding gradient, which leads to wrong results. For this reason, an architecture called the residual network (ResNet) was built [23]. ResNet-50 is a

50-layer residual network in abbreviated form. ResNet-50 is similar to VGG-16, with the exception that it adds an extra identity mapping feature. Figure 1 depicts this approach.

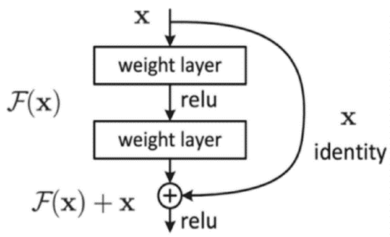

Figure 1. ResNet-50

ResNet forecasts the delta necessary to get from one layer to the next and arrive at the final prediction. ResNet solves the vanishing gradient problem by enabling gradient to flow along an alternative shortcut path. ResNet's identity mapping allows the model to skip a convolutional neural network (CNN) weight layer if the current layer isn't required. This helps to prevent the problem of over fitting the training set. ResNet-50 is a 50-layer network. Below we explain the general methodology of the proposed method.

- Stack pictures: Transfer the dataset with image data store to assist you supervises the information. Since image data store works on picture recording regions, making it suitable for utilize with extended picture collections. An illustration of an picture from one of the categories included in the entire data set. The photo is shown by Mario. Where count each label is used to summarize the number of pictures for each class. We make the pictures equal so that the number of photos within the brew set is adjusted.
- Load pre trained network: There are a few pre-tested systems out there that are popping up everywhere. Most of them are prepared on ImageNet dataset, that contains 1,000 question category and 1.2 million setup pictures [24]. The "ResNet-50" is one of the most popular of these systems. Other major ImageNet based systems incorporate AlexNet, GoogLeNet, VGG-16 and VGG-19 [25], that can be stacked with AlexNet, GoogLeNet, VGG-16 and VGG-19 from the profound learning tool kit ™ program. Utilize plot to imagine thegrid. Since this can be an expanded organization, change the show window to appear fair in the most segment. In the primary layer, we define the input measurements. Each CNN includes the various input measurement necessities. The shape used in this illustration requires the input of a 224x224x3 image. The midway layers form the bulk of the CNN. This is an arrangement of convolutional layers, mixed with modified direct units (ReLU) and maximal collector layers [26]. After these layers are three completely associated layers. The ultimate layer is the classification layer and its characteristics classification depends on errand. In this case, the CNN software stacked was set up to reveal a 1000 way classification issue. Hence, the classify layer contains 1000 class is from ImageNet dataset.
- Get ready preparing and test picture sets: As previously specified, the grid can prepare red green blue (RGB) pictures at 224x224. To maintain a strategic distance from all pictures being saved in this arrangement, use augmented image data store to size any grayscale pictures and alter them to RGB on the fly. Augmented image data store will be used to also augment extra information when used to train a network.
- Extricate preparing highlights utilizing CNN: each layer of CNN produces a reaction, or operation, to an input picture. Be that as it may, there are many layers within CNN that are suitable for extracting the features of an image. The layers at the beginning of the grid capture the primary feature of the image, such as edges and points. To see this, imagine the arrangement of the channel weights from the primary convolutional layer. This may provide offer assistance construct an instinct about why take down highlights from CNNs work thus well for picture acknowledgment assignments. Note that highlighting from more profound layer weights can be done utilizing deep dream image from the profound learning tool kit™. Notice how the base layer learned to arrange channels to capture edge and blob highlights. This is “primitive” highlights are at this point prepared by more profound layers, which combine early highlights to bring out a higher level picture. Those higher-level highlights are way better suited for acknowledgment assignments since they mix all primitive highlights in a wealthier picture representation [25]. You'll be able to effortlessly pull out highlights from one of the deepest layers with an operations strategy. Deciding which of the deeper layers to select can be a

choice of plan, but more often than not beginning with the appropriate layer sometime recently the classification layer can be a great start. In a network, this layer is called "fc1000". Let is break out the preparation for the highlights using that layer. Note that legislation works normally to employ the Graphics processing unit (GPU) to prepare in case it is accessible, or if the central processing unit (CPU) is utilized up. Within the code over, "MiniBatchSize" is set to 32 to guarantee that the CNN information and the picture are contained in the GPU memory. You should reduce "MiniBatchSize" in case your GPU is out of memory. Moreover, the yield of the legislation is organized as columns. This makes a difference in speeding up the next multi-class direct SVM setup.

- Prepare a multiclass support-vector machine (SVM) classifier utilizing CNN highlights: next, utilize the CNN picture highlights to prepare a multiclass SVM classifier. A quick Stochastic Slope Plummet solver is utilized for preparing by setting the Fitcecoc function is 'Learner' parameter to “Linear”. This makes a difference speed-up the preparing when working with high-dimensional CNN highlight vectors.
- Assess classifier: rehash the strategy utilized before removing the picture highlights from the test set. The take a look at options may be passed at that point to the classify to the degree of accuracy of the prepared classifier in Figure 1. Figure 2 shows the pipelines of the proposed framework, where the top is the training stage, and the bottom is the forecast stage.

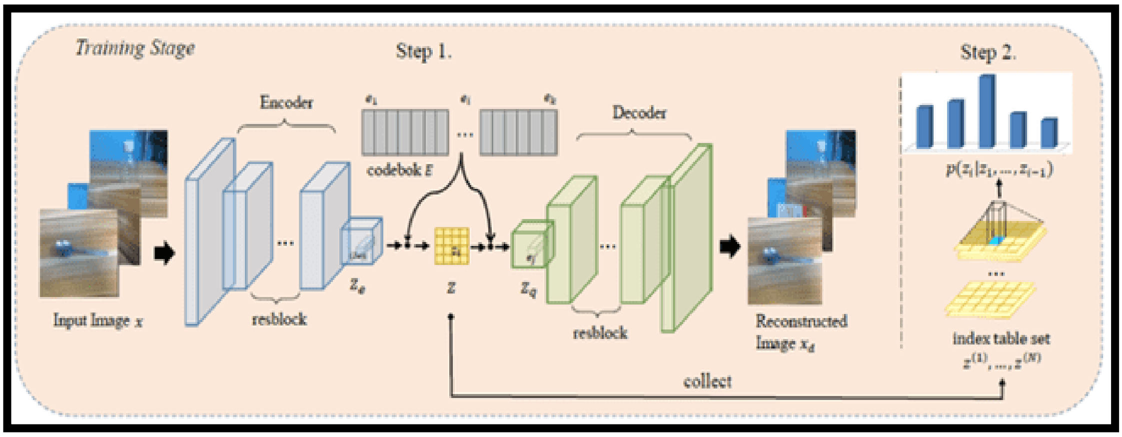

Figure 2. The pipelines of the proposed framework

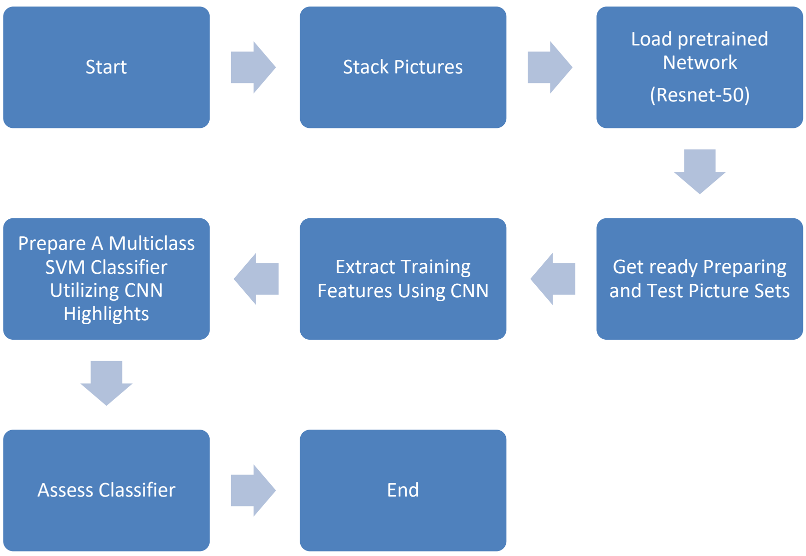

Figure 3. shows the basic parts of the proposed system

2.1. Dataset

We assess the implementation of the planned strategy on the overall normal picture irregularity site dataset: a dataset consisting of three types of industrial elements namely: bottle ,carton and spoon. Where the data of the bottle is made into two parts (typical and atypical), where the typical contains 116 pictures and the atypical contain four classes, the first contains 22 pictures, the second 19 pictures, the third 26 pictures, the last 26 pictures, so the total number of pictures becomes Atypical 93 images. The cartoon consists of two parts (typical and atypical), where the model contains 116 pictures. Atypical is divided into five categories, the first contains 24 pictures, the second 25 pictures, the third 35 pictures, the fourth 20 pictures, and the last contains 30 pictures, so the total number of atypical images becomes 134 pictures. The spoon consists of two parts (typical and atypical) as it contains 106 typical and atypical images divided into five parts, the first contains 20 pictures, the second 23 pictures, the third 32 pictures, the fourth 24 pictures and the last 28 atypical pictures, so it becomes. The total number of atypical images is 127, as shown in Table 1, where the pictures were taken locally by Apple's versatile 12-megapixel camera. The image size is 4032*3024 with a depth of 24 bit, the image type is jpg and the images are RGB color. An overview of the dataset is shown in Figure 4, where top is normal and button is anomaly.

Figure 4. Proposed system dataset

3. RESULTS AND DISCUSSION

The data set is not compressed in the used image file, as the system gives the first instance of images for each of the three data sets. As shown in Figure 5, the training images that were used in the proposed system. Table 1 also shows the number of anomalies.

ResNet-50 is trained on more than a million images and can classify images into 1,000 feature. Analyze the network architecture. The first layer, the image input layer, requires input images of size 224-by-224-by-3, where 3 is the number of color channels. As Figure 6 illustrates:

Figure 5. Training image

Table 1. Number of anomaly data

Spoon		Bottle		Carton	
Count	Label	Count	Label	Count	Label
Good	20	Good	22	Good	24
Scratched	23	A little bit dulled	19	Irregular	25
Crooked	24	Braided	26	Another shape	35
More Crooked	32	Crooked more	26	Spoiled	20
broken	28				

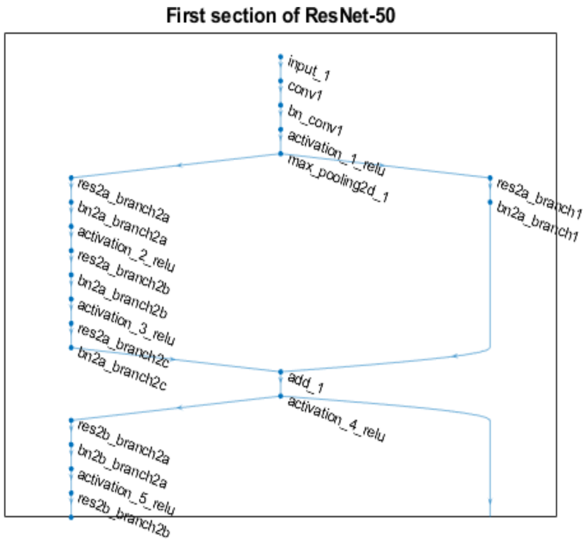

Figure 6. First section of ResNet-50

The first layer is defined as the input dimensions. Where the original size of the input images is changed to 224x224x3, as shown in Figure 7. The last layer is the classification layer and its properties depend on the classification task. In this system the CNN model is trained to solve the problem of classifying 1000 features of the images. Thus, the classification layer contains 1,000 classes of the ImageNet dataset. The number of category names for the ImageNet classification errand is counted is 1,000. Augmented image data store is created from the training and test sets for resizing the images of the desired size from the grid. Obtain the organize weights for the moment convolutional layer, and change the weights of the network for visualization. A montage of grid weights is shown in Figure 8, as there are 96 person groups of weights within the to begin with layer.

```
ImageInputLayer with properties:

                   Name: 'input_1'
              InputSize: [224 224 3]

 Hyperparameters
       DataAugmentation: 'none'
          Normalization: 'zerocenter'
 NormalizationDimension: 'auto'
                   Mean: [224×224×3 single]
```

Figure 7. The first layer in model

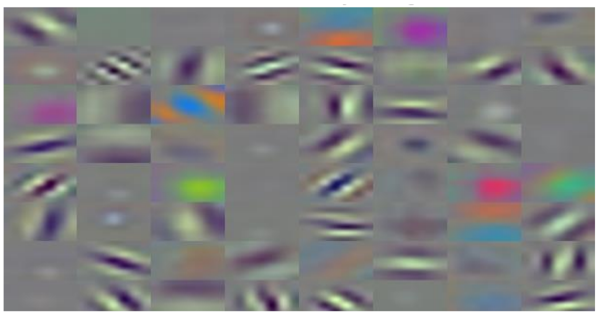

Figure 8. First convolutional layer weights

Training posters are obtained from the training set. We train a multi-class SVM classifier employing a quick direct analyzer, as it maps notes into columns to capture the identical features in the training. The test features are extracted using CNN to pass the image features to the trained classifier and obtain the known nomenclature. Display a table of results using the confusion matrix that converts the number into percentage form and displays the average accuracy. As Table 2 shows.

Table 2. Accuracy

Carton		Bottle		Spoon	
accuracy	95%	accuracy	99.11%	accuracy	90%

4. CONCLUSION

In this work we propose a component for classifying anomalous pictures using training imagery and deep learning methods. The information gotten within the ImageNet classification assignment can be effectively exchanged to the anomalous image classification assignments. In our tests, the highlight extractor that ResNet-50 learned performed very well. Our results confirmed that the anomalous images could be detected accurately. Where we evaluated our approach on three types of data, namely the spoon, the carton and the Bottle, where in the spoon an average accuracy of 95% was obtained. And the carton has obtained an average accuracy of 90%. And bottle gets an average accuracy of 99%. While many solutions rely on handcrafted highlight extractors, our approach does not require any include building, utilizing crude pixel values for spoon, cartoon, or bottle images to represent basic anomalies. In the future we will apply the proposed method to all metal and paper objects.

REFERENCES

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, Jul. 2009, doi: 10.1145/1541880.1541882.
[2] R. Chalapathy, A. K. Menon, and S. Chawla, "Anomaly detection using one-class neural networks," 2018, *arXiv:1802.06360*.
[3] R. Chalapathy, A. K. Menon, and S. Chawla, "Robust, deep and inductive anomaly detection," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10534, 2017, pp. 36–51, doi: 10.1007/978-3-319-71249-9_3.
[4] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," *arXiv preprint*, Jan. 2019, [Online]. Available: http://arxiv.org/abs/1901.03407.
[5] T. Ehret, A. Davy, J. M. Morel, and M. Delbracio, "Image anomalies: A review and synthesis of detection methods," *Journal of Mathematical Imaging and Vision*, vol. 61, no. 5, pp. 710–743, Jun. 2019, doi: 10.1007/s10851-019-00885-0.
[6] L. Wang, D. Zhang, J. Guo, and Y. Han, "Image anomaly detection using normal data only by latent space resampling," *Applied Sciences (Switzerland)*, vol. 10, no. 23, pp. 1–19, Dec. 2020, doi: 10.3390/app10238660.
[7] M. Markou and S. Singh, "Novelty detection: A review - Part 1: Statistical approaches," *Signal Processing*, vol. 83, no. 12, pp. 2481–2497, Dec. 2003, doi: 10.1016/j.sigpro.2003.07.018.
[8] Y. H. Liu, S. H. Lin, Y. L. Hsueh, and M. J. Lee, "Automatic target defect identification for TFT-LCD array process inspection using kernel FCM-based fuzzy SVDD ensemble," *Expert Systems with Applications*, vol. 36, no. 2 PART 1, pp. 1978–1998, Mar. 2009, doi: 10.1016/j.eswa.2007.12.015.
[9] I. Abdel-Qader, S. Pashaie-Rad, O. Abudayyeh, and S. Yehia, "PCA-Based algorithm for unsupervised bridge crack detection," *Advances in Engineering Software*, vol. 37, no. 12, pp. 771–778, Dec. 2006, doi: 10.1016/j.advengsoft.2006.06.002.
[10] P. Napoletano, F. Piccoli, and R. Schettini, "Anomaly detection in nanofibrous materials by CNN-based self-similarity," *Sensors (Switzerland)*, vol. 18, no. 1, p. 209, Jan. 2018, doi: 10.3390/s18010209.
[11] L. Ruff *et al*., "Deep one-class classification," in *35th International Conference on Machine Learning, ICML 2018*, 2018, vol. 10, pp. 6981–6996.
[12] S. K. Sinha and P. W. Fieguth, "Automated detection of cracks in buried concrete pipe images," *Automation in Construction*, vol. 15, no. 1, pp. 58–72, Jan. 2006, doi: 10.1016/j.autcon.2005.02.006.
[13] S. Iyer and S. K. Sinha, "Segmentation of pipe images for crack detection in buried sewers," *Computer-Aided Civil and Infrastructure Engineering*, vol. 21, no. 6, pp. 395–410, Aug. 2006, doi: 10.1111/j.1467-8667.2006.00445.x.
[14] M. J. Chae and D. M. Abraham, "Neuro-fuzzy approaches for sanitary sewer pipeline condition assessment," *Journal of Computing in Civil Engineering*, vol. 15, no. 1, pp. 4–14, Jan. 2001, doi: 10.1061/(asce)0887-3801(2001)15:1(4).
[15] Y. Sen Yang, C. M. Yang, and C. W. Huang, "Thin crack observation in a reinforced concrete bridge pier test using image processing and analysis," *Advances in Engineering Software*, vol. 83, pp. 99–108, May 2015, doi: 10.1016/j.advengsoft.2015.02.005.
[16] R. S. Adhikari, O. Moselhi, and A. Bagchi, "Image-based retrieval of concrete crack properties for bridge inspection," *Automation in Construction*, vol. 39, pp. 180–194, Apr. 2014, doi: 10.1016/j.autcon.2013.06.011.
[17] B. Y. Lee, Y. Y. Kim, S.-T. Yi, and J.-K. Kim, "Automated image processing technique for detecting and analysing concrete surface cracks," *Structure and Infrastructure Engineering*, vol. 9, no. 6, pp. 567–577, Jun. 2013, doi: 10.1080/15732479.2011.593891.
[18] A. Arena, C. Delle Piane, and J. Sarout, "A new computational approach to cracks quantification from 2D image analysis: Application to micro-cracks description in rocks," *Computers and Geosciences*, vol. 66, pp. 106–120, May 2014, doi: 10.1016/j.cageo.2014.01.007.
[19] H. N. Nguyen, T. Y. Kam, and P. Y. Cheng, "An automatic approach for accurate edge detection of concrete crack utilizing 2D geometric features of crack," *Journal of Signal Processing Systems*, vol. 77, no. 3, pp. 221–240, Dec. 2014, doi: 10.1007/s11265-013-0813-8.

[20] M. Rodríguez-Martín, S. Lagüela, D. González-Aguilera, and J. Martínez, “Thermographic test for the geometric characterization of cracks in welding using IR image rectification,” *Automation in Construction*, vol. 61, pp. 58–65, Jan. 2016, doi: 10.1016/j.autcon.2015.10.012.
[21] R. Bendikiene, S. Baskutis, J. Baskutiene, A. Ciuplys, and T. Kacinskas, “Comparative study of TIG welded commercially pure titanium,” *Journal of Manufacturing Processes*, vol. 36, pp. 155–163, Dec. 2018, doi: 10.1016/j.jmapro.2018.10.007.
[22] P. Broberg, “Surface crack detection in welds using thermography,” *NDT and E International*, vol. 57, pp. 69–73, Jul. 2013, doi: 10.1016/j.ndteint.2013.03.008.
[23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2016, vol. 2016-December, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[24] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2010, pp. 248–255, doi: 10.1109/cvpr.2009.5206848.
[25] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Sep. 2015, [Online]. Available: http://arxiv.org/abs/1409.1556.
[26] J. Donahue et al., “DeCAF: A deep convolutional activation feature for generic visual recognition,” in *31st International Conference on Machine Learning, ICML 2014*, Oct. 2014, vol. 2, pp. 988–996, [Online]. Available: http://arxiv.org/abs/1310.1531.

BIOGRAPHIES OF AUTHORS

Zaid Taher Omer he received B.Sc degree in computer from Al-rafidain University, College of Science, Iraq in 2012 and 2016 respectively. He received his M.sc degree in 2022 from Al-Mustansiriya University, in the field of image processing. Research interests include image processing and image classification with python programming. He can be contacted at email: zaidalradi@uomustansiriyah.edu.iq.

Amel Hussein Abbas she received B.Sc and M.Sc degree in Physics from Al-Mustansiriya University, College of Science, Iraq in 1988 and 1996 respectively. She received his Ph.D degree in 2003 from the University of Baghdad in the field of image processing. Research interests include image processing and image classification with Matlab programming. She worked as a lecturer in Mustansiriyah University. Skills Image Processing Pattern Recognition Computer Vision MATLAB. She can be contacted at email: dr.amelhussein2017@uomustansiriyah.edu.iq.