A New Approach for Classification and Detection of World Cultural Heritages with YOLOv3

Mert Çeçen
University of Fırat
Faculty of Computer Engineering
Elazığ, Türkiye
mert.cecen23@gmail.com

Mehmet Karaköse
University of Fırat
Faculty of Computer Engineering
Elazığ, Türkiye
mkarakose@firat.edu.tr

2023 International Conference on Sustaining Heritage: Innovative and Digital Approaches (ICSH) | 979-8-3503-9828-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSH57060.2023.10482838

***Abstract*—Since the formation of the world, many civilizations have lived and these civilizations have left important information about their social and architectural structures by creating a cultural heritage infrastructure in order to transfer their experiences to future generations. In this process, civilizations have left many cultural heritage works, both tangible and intangible. The concept of digital heritage has emerged for the protection and promotion of cultural heritage sites and works that still exist today, and studies have been carried out in the literature for the digitization of the works of certain regions. In this study, apart from literature studies, the images of 200 different artifacts included in the World Heritage List, which contain images of various structures from different regions created by the United Nations Educational, Scientific and Cultural Organization (UNESCO), are manually collected, labeled and classified and necessary pre-processing is carried out. By performing the steps, it was primarily turned into a data set. Then, with the created dataset, a training process was carried out using the You Only Look Once version 3 (YOLOv3) technology, which detects objects in images using Convolutional Neural Networks, one of the deep learning techniques, and the process of modeling is explained. Our main goal in creating the model has been to classification these works, which have scientific data characteristics, in the digital environment, to identify them, to be recognized by future generations, to transfer them in a safe way, and to ensure that researchers who will work in this field can use them as a source. The performance of the model was tested by using different images to verify with the data for the designed model. As a result of the performance tests carried out, an accuracy rate of 97% was obtained.**

Keywords—Deep Learning, Cultural Heritage, Convolutional Neural Networks, YOLOv3

I. INTRODUCTION

Cultural heritage works have an important role in transferring and preserving the unique characteristics and ethnic structure of civilizations to future generations. Cultural heritage works are examined under two separate headings as tangible and intangible cultural heritage works. Intangible cultural heritage works usually consist of social traditions that contain the unique characteristics of nations, lifestyles and etiquette. Concrete cultural heritage works represent structures that have come into existence both by civilizations and spontaneously from the formation of the world and the beginning of human history to the present day. The United Nations Educational, Scientific and Cultural Organization (UNESCO) was established as a result of the adoption of the Convention on the Protection of the World Cultural and Natural Heritage on 16 November 1972, as a result of a conference, in order to identify and help protect these cultural heritage works. 1154 heritage sites declared as World Heritage Sites were added to the World Heritage List, which was created according to the Convention for the Protection of the World Cultural and Natural Heritage and determined by the World Heritage Committee. 897 of the works in this list have been determined as cultural, 218 as natural and 39 as mixed (natural and cultural) heritage. These cultural heritage works, both tangible and intangible, are characterized as elements of historical and scientific value. In this article, we have made a model by identifying and classifying cultural heritage works in the digital environment with our studies using deep learning methods. In this way, we tried to protect the cultural heritage artifacts and make it easier for future generations to learn about the cultural heritage artifacts.

A. Literature Review

These artifacts, which are protected by the United Nations Educational, Scientific and Cultural Organization (UNESCO), face the possibility of being endangered due to natural wear, human neglect or wars. There are literature studies in various and different fields by using artificial intelligence techniques at the point of intelligent protection and digitalization of these cultural heritage works[1-16]. Shaimaa Fahad Rashid and others gathered information from various social networks, Google Maps and various encyclopedias in order to protect and promote the cultural heritage of Nineveh, one of the important cities of ancient times ruled by the Assyrians, using deep learning and cloud computing technologies together. Cloud computing techniques, Docker Containers he mentioned that the digitalization of cultural heritage works by using various computer technologies, such as virtual machines provides benefits for the protection of these works and for different users to use digital resources easily. They explained how the collected data can be queried using Docker Containers and virtual machines by performing the necessary cleaning and configuration processes[2].

B. Deep Learning and Cultural Heritage

In the literature, different deep learning methods have been used for cultural heritage works within the scope of these studies. Deep learning is a subset of machine learning that uses classification operations and learning methods to represent data in specific formats[3]. The preprocessing steps significantly strengthen the inputs and suppress irrelevant variations significantly when classifying [4]. Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Deep Belief Network, Deep Boltzmann Machine and Deep Auto-Encoder techniques are among the main techniques of deep learning. Convolutional Neural Networks (CNN) method, which is one of the deep learning methods in image classification, has been proven to provide good performance [5]. Although neural networks are not a new technology, they are based on Alexnet[5] and Imagenet[6] technologies used in the classification of large-scale data. Other studies in the field of artificial intelligence with deep learning have proven that deep learning methods can be used in data-critical studies[7]. You Only Look Once (YOLO) takes the entire image at once and estimates the bounding box coordinates and class probabilities. In this study, YOLOv3 technology using Convolutional Neural Networks was used. YOLOv3 consists of 106 layers in total[8]. In general, its architecture consists of 3 different layer forms. This architecture is shown in Fig. 1. The first layer is the residual layer and is created during activation. The outputs formed here enter the detection layer, which is the second layer, and enter the detection process in 3 different stages, and at this stage, the size of the grids is increased for detection. In the last and third layer, the sampling layer, the spatial resolution is increased. The pink blocks in Fig. 1 are the residual layers, the orange ones are the detection layers, and the green ones are the sampling layers.

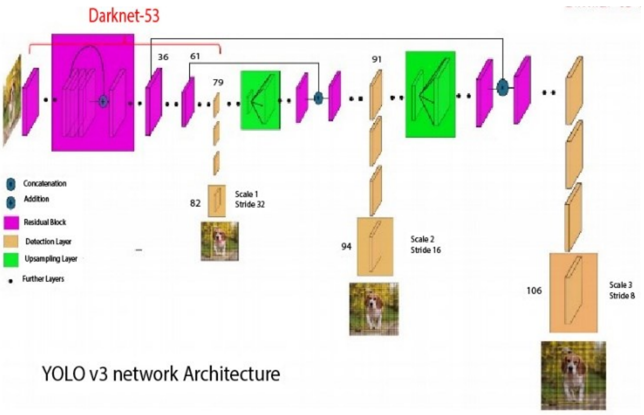

Fig. 1. Architecture of YOLOv3 [8]

YOLOv3 uses 3 different forecasting scales while performing the forecasting process. The detection layer is used to detect feature maps of three different dimensions, consisting of 32, 16, and 8 steps, respectively. This shows that the detections are made with the input of 416 x 416 by fitting them to 13 x 13, 26 x 26 and 52 x 52 scales[8]. In line with all this information, the cultural heritage images included in the dataset we will use in the study were resized to 416 x 416 dimensions. In order to train the model quickly and with high accuracy, the training process was carried out using YOLOv3 technology. In the rest of the article, within the scope of the studies, in the second part, the dataset creation, model training and research experiment processes are explained in detail and the main goal of the study is mentioned. In the third part of the article, the variables to be used during the study, the training environment and the parameter values that need to be adjusted are explained and the experimental results of the system obtained as a result of the training are explained. In the last section, the results of the studies carried out are drawn.

II. Proposed Method

The United Nations Educational, Scientific and Cultural Organization (UNESCO) was established by the United Nations in 1946, examining many cultural structures in our world and taking these structures under protection for transfer to future generations. In this article, a study has been carried out to transfer and protect cultural heritage works in the World Heritage List prepared by the United Nations Educational, Scientific and Cultural Organization (UNESCO). Images of 200 different tangible cultural heritage works in various structures from different regions in this list were manually collected, labeled, classified and turned into a dataset. The generated dataset is divided into training, validation and test data. While the validation data was used to measure the accuracy of the model, the test data were used to test the model after the model was created. The operations performed are shown in the block diagram in Fig. 2.

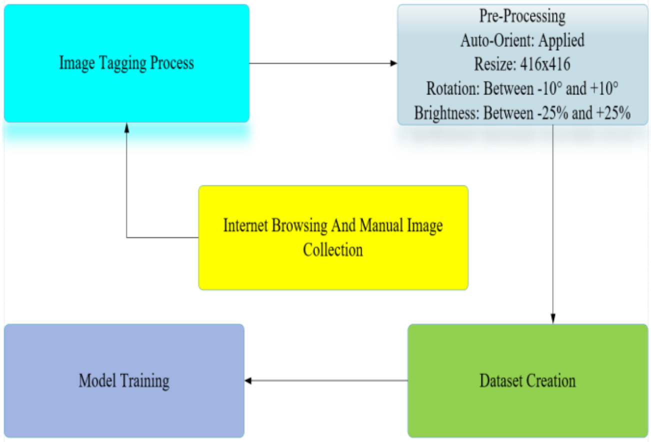

Fig. 2. Model training block diagram.

In the block diagram in Fig. 2, manual image collection, image tagging process, preprocessing steps and the development of the model are shown step by step to create the dataset. The operations performed at the stages in the block diagram created for the studies carried out during the model training are explained below.

A. Dataset Creation and Preprocessing

2050 images were manually collected in picture format for cultural heritage works included in the United Nations Educational, Scientific and Cultural Organization (UNESCO) World Heritage List. The collected images were then labeled with the bounding box labeling method. In order for the data to be available and enriched in the model, an automatic orientation process has been applied to the existing images. All image data has been resized in 416x416 size, which is the input format of YOLOv3 technology. Then, 10 degrees of rotation and 25% brightness adjustment were performed on the images for better recognition of each image. After the operations, the number of images of the dataset was updated to 4100. The aggregated datasets are divided into 70% training data, 10% validation data, and 20% test data.

B. Model

In order to use the YOLOv3 technology used in object detection using Convolutional Neural Networks (CNN) with the created dataset, a Jupyter Notebook was created in the Colaboratory cloud environment created by the Google company, and the necessary code blocks, parameter values, technologies and files were transferred to the cloud environment and the training process was started. With the studies carried out during the education process, a new method has been proposed for the digitization, preservation, detection, classification and easy use of cultural heritage works by researchers and users.

C. Research Experiment

In this study, the United Nations Educational, Scientific and Cultural Organization (UNESCO) World Heritage cultural heritage dataset was used. During the creation of this dataset, different image data collected manually were brought together. The gathered data were put into the training process with YOLOv3 technology using Convolutional Neural Networks (CNN), as a result, necessary research was carried out for the classification and detection of 200 different works in the list, and a training was carried out as a result of these researches. With the model created as a result of the experiments and the proposed method, it is aimed to introduce cultural heritage works, to turn them into digital heritage works in order to take precautions against the threats they may encounter, and to help researchers who will work in the field of cultural heritage works and digital heritage by providing resources.

III. Experimental Results

A. Dataset and Experimental Environment

In order to show the accuracy of the data set we created with deep learning and the model we trained, the results we obtained are included in this section. In order to get better results in the studies, the Graphics Processing Unit (GPU), which enables clearer graphics to be created by accessing the computing resources provided by Google Colaboratory technology, and Jupyter Notebook for cloud use. It was created by NVIDIA Company and implemented using CUDA (Compute Unified Device Architecture) technology. While creating the dataset, labeling operations were carried out with the bounding box method and Roboflow technology was used for this work. After applying the necessary preprocessing steps to the created dataset, the lost values of the Training Graphs are shown in detail in Fig. 3.

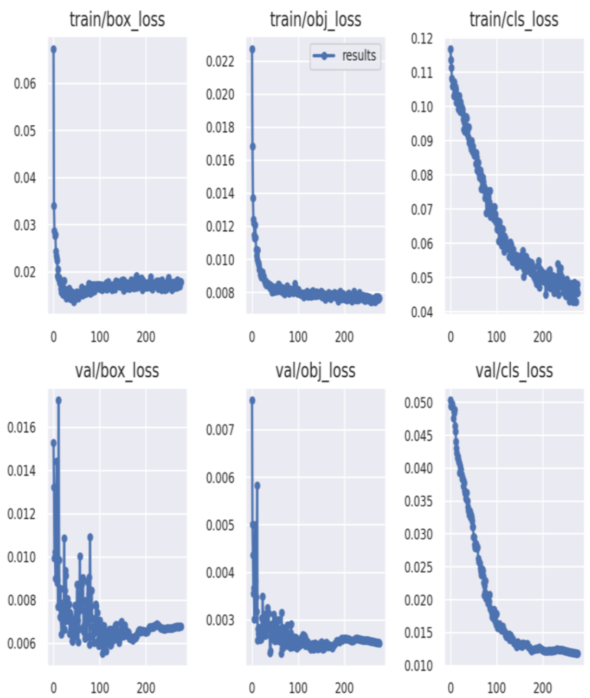

Fig. 3. Training Graphs loss values.

The train values in the graph in Fig. 3 represent the measurements of the values of the training set, while the val values represent the measurements of the values of the validation set. The box_loss value focuses on the measure of the loss rate that will occur after the application of the bounding box technique used in labeling operations. The low values in the graph show that the model has improved for generalization and the dataset is more well labeled. Whereas the cls_los value shows the measure of the loss rate that occurs as a result of the classification. The decrease in the value in the graph shows that a better classification process is performed.

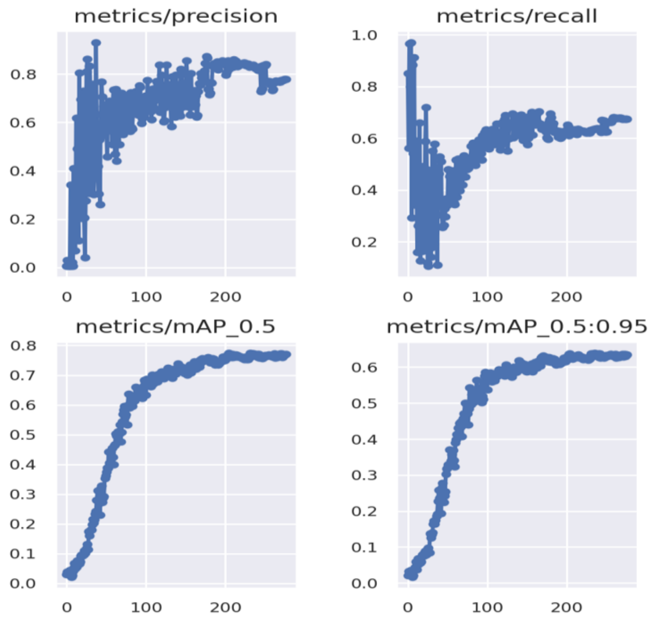

Fig. 4. Training Graphs metrics values.

The Precision value in Fig. 4 shows the precision of the model at the estimated time, while the Recall value determines the performance measure of the system. The value of mAP_0.5 in the other graph indicates the average precision, and the value of mAP_0.5:0.95 indicates the average precision. From the results in the graph, it is observed that the labeling process of the model has been successful and a good dataset has been created. While training with the created dataset, the cloud environment provided by Google Colaboratory technology was used and the model was trained by accessing the Graphics Processing Unit (GPU), which is one of the computing resources.

B. Experiment Study and Results Analysis

Darknet technology is generally used to print the confidence value of the objects detected by the model. In order to perform the model training process, the necessary codes were written first, and the Darknet technology was made available and the cloning process was carried out into the cloud environment. Then, the CUDA technology offered by the NVIDIA company, which we will use to perform the training, is uploaded to the cloud.

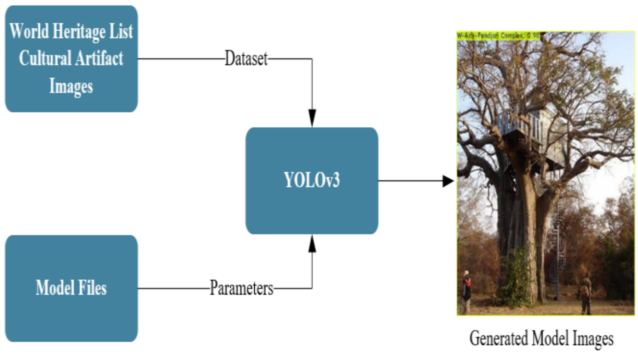

Fig. 5. YOLOv3 Training Architecture.

As seen in Fig. 5, he created a folder of Obj files and brought together the information in the dataset to manually collect and upload the dataset to the cloud. The created folder was transferred to Google Drive and copied to the cloud by writing the necessary codes. Next is the internal manual configuration of the model in the cloud to use the required parameter. The required parameter values for the configuration results created have been adjusted to get the most appropriate result. Again, Objnames file was created and added class names, Objdata files was created, and the class number was assigned as 200. Inside the Google Drive account, the YOLOv3 files create a file and place the Backup results in a folder. Finally, the training process of the model was started by creating the Train file, which is required for training, including the model files, using the python programming language and running it on the cloud. After performing the training with YOLOv3 technology, it is ensured that the weight values obtained as a result of the training are entered by entering the Backup file, which is opened by entering the Google Drive account by changing the necessary directory. Deep learning average loss value was obtained as 0.208852 according to the weight values obtained, the images in the model dataset were entered and results over 97% were obtained. A few of the study results are reflected in Fig. 6.

Fig. 6. The detection rate of a few of the cultural artifact images included in the UNESCO World Heritage List.

Fig. 6 shows the accuracy percentages obtained as a result of detecting a few images from the United Nations Educational, Scientific and Cultural Organization (UNESCO) World Heritage List cultural artifacts. In order to obtain these values, the necessary information was given to the model from the Backup file in which the weights values obtained as a result of the use of YOLOv3 technology were recorded, and these ratios were printed with the Darknet technology. The results show that the images in the dataset were detected with high accuracy.

C. Rates obtained in Cultural Heritage Artifacts Deep Learning Studies

Many studies have been carried out in the literature using different datasets for the preservation, digitization and classification of cultural heritage works. In this article, a table showing the accuracy rates obtained from studies conducted using deep learning techniques and the accuracy rates obtained from other studies in the literature has been created. Comparison Results With Studies In The Literature are shown below.

TABLE I. COMPARISON RESULTS WITH STUDIES IN THE LITERATURE

Referance	*Dataset*	*Accuracy Rates*
[9]	Indian Heritage Infrastructure	%98,75
[12]	Indonesian Cultural Heritage Site	%77
This Study	UNESCO World Heritage List	%97

In the first study in Table 1, the authors aimed to create a website by creating a Deep Neural Network (DNN) that classifies using deep learning technologies in order to help protect the data of the Indian heritage infrastructure. In addition, they explained the system to achieve 98.75% accuracy for the heritage domain dataset created under a MobileNet architecture learned by transfer in the proposed framework with the crowdsourced dataset. In the second study in the table, the authors classified the data of the Indonesian cultural heritage area using the Convolutional Neural Networks (CNN) technique, which is one of the deep learning techniques. As a result of this classification, they observed that the model created from the image data was 77% successful, using the test data of the dataset they used. Finally, in this study, a dataset was created using the images of UNESCO World Heritage List cultural heritage works, and this dataset was transformed into a model containing 200 different classes using YOLOv3 technology using Convolutional Neural Network (CNN). This model, which was created as a result of the studies, was classification and detection 97% successful.

IV. CONCLUSION

The concept of cultural heritage is used as sources that provide important information in obtaining historical and scientific information about the social and ethnic structure of the period in which civilizations lived. In the literature, there are studies on transforming cultural heritage works of certain regions into digital heritage. In this article, a study describing the process of protecting, classifying and detecting 200 different artifacts in different regions that have been included in the World Heritage List by the United Nations Educational, Scientific and Cultural Organization (UNESCO) from problems that may occur over time is explained. During this study, a dataset was created with the bounding box labeling method using Roboflow technology. Then, preprocessing steps were applied to the created dataset. With the use of YOLOv3 technology, which uses Convolutional Neural Networks (CNN), one of the deep learning methods, a training was conducted to identify the cultural heritage sites in the list and a model was created. After the model training, it was observed that 97% successful results were obtained in the classification and identification of the cultural heritage works in the dataset. With these studies, a comprehensive study has been carried out for the protection, recognition and learning of these cultural heritage works that are important for our world and future generations and have scientific value. In future studies, it is aimed to enlarge the dataset and integrate new technologies into our model and get better results. In this context, images of other works on the United Nations Educational, Scientific and Cultural Organization (UNESCO) World Heritage List heritage works will be manually collected and added to the dataset in order to enlarge the dataset. Then, as a result of training the expanded dataset in different versions of YOLO technology, the rates obtained in these versions will be compared. As a result of the studies, the version that gives the best results will be selected and it will be provided to help researchers who will work in this field.

REFERENCES

[1] Q. Li and G. Qinzhou, "Intelligent Intangible Cultural Heritage Innovation Platform Under the Background of Big Data and Virtual Systems .", 535000, China.

[2] S. Rashid and R, Qasha, "Cloud-Based Big Data Approach for Managing Multi-Types Data of Cultural Heritage. ", 2022 IEEE.

[3] Y. Lecun and Y. Bengio, "Deep learning - review."

[4] N. Srivastava and R. Salakhutdinov, "Multimodal Learning with Deep Boltzmann Machines," *Adv. neural Inf. Process. Syst.*, pp. ,2012.

[5] Z A. Krizhevsky, I. Sutskever, and H. Geoffrey E., "ImageNet Classification with Deep Convolutional Neural Networks," *Adv.Neural Inf. Process. Syst. 25*, pp. 1–9, 2012.

[6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255: IEEE.

[7] A. Belhi, H. Gasmi, A K. Al-Ali, A. Bouras, S. Foufou, X. Yu and H. Zhang, "Deep Learning and Cultural Heritage: The CEPROQHA Project Case Study," 2019 IEE.

[8] S.V. Viraktamath, M. Yavagal and R. Byahatti, " Object Detection and Classification using YOLOv3", International Journal of Engineering Research & Technology, Dharwad, India, 2021.

[9] U. Kulkarni, Meena S M, P. Joshua, K. Rodrigues and S V. Gurlahosur, "Integrated Crowdsourcing Framework Using Deep Learning for Digitalization of Indian Heritage Infrastructure. ", 2020 IEEE.

[10] M. Casillo, F. Colace, Brij B. Gupta, A. Lorusso, F. Marongiu, D. Santaniello, "A Deep Learning Approach to Protecting Cultural Heritage Buildings Through IoT-Based Systems", 2022 IEEE.

[11] M. Shimoni, T. Croonenborghs, P.-Y Declercq, A. Drougkas, E. Verstrynge, F-P. Hocquet, R. Hayen and K. Van Balen , "Advanced Processıng Of Remotely Sensed Bıg Data For Cultural Heritage Conservation", Brussels, Belgium 2019 IEEE.

[12] R A. Kambau, Z A. Hasibuan, M O. Pratama , " Classification for Multiformat Object of Cultural Heritage using Deep Learning", IEEE.

[13] M. Hess, V. Petrovic, D. Meyer, D. Rissolo and Falko Kuester, " Fusion of Multimodal Three-dimensional Data for Comprehensive Digital Documentation of Cultural Heritage Sites", San Diego, California 92093 2015 IEEE.

[14] S. Alam, K. Islam, N. Sharmila, Z R. Sovon, R M. Rahman, " Image Captioning- Bangladesh's Heritage Perspective Using Deep Learning", Dhaka1229, Bangladesh, 2022 IEEE.

[15] K. Madani, A M. Rinaldi and C. Russo, " Combining Linked Open Data and Multimedia Knowledge Base for Digital Cultural Heritage Robotic Applications", 2021 IEEE.

[16] M. Li, Y. Wang and Y.Q Xu, "Computing for Chinese culturalheritage". Visual Informatics, vol 6, no 1, 2022, pp.1-13.

[17] A. Ramachandran, S. Vasudevan, and V. Naganathan, "Deep Learningfor Music Era Classification," pp. 1–8, 2016.

[18] S. Mccann and D. G. Lowe, "Local Naive Bayes Nearest Neighbor forImage Classification_TR-2011-11_1.pdf," *IEEE Comput. Vis. Pattern Recognit.*, pp. 3650–3656, 2012.

[19] S. Hershey et al., "Cnn Archıtectures For Large-Scaleaudıo Classıfıcatıon," IEEE Int. Conf. Acoust. Speech Signal Process., pp. 131–135, 2017.

[20] F. Stanco, S. Battiato, and G. Gallo, " Digital imaging for cultural heritage preservation: Analysis, restoration, and reconstruction of ancient artworks. ", CRC Press, 2011.

[21] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, " Deep learning. ", MIT press Cambridge, 2016.

[22] A. Belhi and A. Bouras, "Towards a multimodal classification of cultural heritage," in *Qatar Foundation Annual Research Conference Proceedings*, 2018, vol. 2018, no. 3, p. ICTPD1010: HBKU Press Qatar.

[23] S. Banerji and A. Sinha, "Painting Classification Using a Pretrained Convolutional Neural Network," in International Conference on Computer Vision, Graphics, and Image processing, 2016, pp. 168-179.

[24] S. Fern, A. Graves, and T. U. Munich, "Sequence Labelling in Structured Domains with Hierarchical Recurrent Neural Networks,"*Proceeding 20th Int. Join Conf. Artif. Intell. IJCAI*, pp. 774–779, 2007.

[25] M. Karaköse and H. Yetis "A New Framework Containing Convolution and Pooling Circuits for Image Processing and Deep Learning Applications with Quantum Computing Implementation," International Information and Engineering Technology Association, Vol. 39, No. 2, April, 2022, pp. 501-512.

[26] P. Tang, F. Chen, X. Zhu and W. Zhou, "Monitoring cultural heritage sites with advanced multi-temporal InSAR technique: The case study of the Summer Palace," Remote Sensing 8 (5),432, 2016.

[27] D. Ocón, "Digitalising endangered cultural heritage in SoutheastAsian cities: preserving or replacing?", International Journal of Heritage Studies, vol.27,No.10, Oct 2021, pp.975-990.

[28] N. A. Khan, S. M. Shafi and H. Ahangar, "Digitization of cultural heritage: Global initiatives, opportunities and challenges", Journal of Cases on Information Technology, vol 20, no 4, 2018, pp.1-16.

[29] Y. Gao, L. Wang, H. Zhang, and R. Liu,"Protection and activation of cultural heritage based on cloud computing platform", In IOP Conference Series: Materials Science and Engineering 2020, Vol. 750, No. 1, p. 012216, DOI 10.1088/1757-899X/750/1/012216.

[30] F. Amato, V. Moscato, A. Picariello, F. Colace, M. D. Santo, F. A.Schreiber and L. Tanca, "Big data meets digital cultural heritage: Design and implementation of scrabs, a smart context-aware browsing assistant for cultural environments", Journal on Computing and Cultural Heritage, vol 10, no 1, 2017, pp.1-23.

[31] W. R. Tan, C. S. Chan, H. E. Aguirre, and K. Tanaka, "Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification," in 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 3703-3707.