Re-exploring the Kayseri Culture Route by Using Deep Learning for Cultural Heritage Image Classification

Cultural Heritage Image Classification by Using Deep Learning: Kayseri Culture Route

Özlem Kevseroğlu
Abdullah Gül University, Department of Architecture,
38080, Kayseri, Türkiye
ozlem.kevseroglu@agu.edu.tr

Rifat Kurban
Abdullah Gül University, Department of Computer Engineering, 38080, Kayseri, Türkiye
rifat.kurban@agu.edu.tr

ABSTRACT

The categorization of images captured during the documentation of architectural structures is a crucial aspect of preserving cultural heritage in digital form. Dealing with a large volume of images makes this categorization process laborious and time-consuming, often leading to errors. Introducing automatic techniques to aid in sorting would streamline this process, enhancing the efficiency of digital documentation. Proper classification of these images facilitates improved organization and more effective searches using specific terms, thereby aiding in the analysis and interpretation of the heritage asset. This study primarily focuses on applying deep learning techniques, specifically SqueezeNet convolutional neural networks (CNNs), for classifying images of architectural heritage. The effectiveness of training these networks from scratch versus fine-tuning pre-existing models is examined. In this study, we concentrate on identifying significant elements within images of buildings with architectural heritage significance of Kayseri Culture Route. Since no suitable datasets for network training were found, a new dataset was created. Transfer learning enables the use of pre-trained convolutional neural networks to specific image classification tasks. In the experiments, 99.8% of classification accuracy have been achieved by using SqueezeNet, suggesting that the implementation of the technique can substantially enhance the digital documentation of architectural heritage.

CCS CONCEPTS

• **Machine learning**; • **Supervised learning by classification**; • **Neural networks**;

KEYWORDS

Convolutional neural networks, Deep learning, SqueezeNet, Cultural Heritage, Image Classification

ACM Reference Format:
Özlem Kevseroğlu and Rifat Kurban. 2024. Re-exploring the Kayseri Culture Route by Using Deep Learning for Cultural Heritage Image Classification: Cultural Heritage Image Classification by Using Deep Learning: Kayseri Culture Route. In *Cognitive Models and Artificial Intelligence Conference (AICCONF '24), May 25, 26, 2024, İstanbul, Turkiye*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3660853.3660913

AICCONF '24, May 25, 26, 2024, İstanbul, Turkiye

ACM ISBN 979-8-4007-1692-8/24/05
https://doi.org/10.1145/3660853.3660913

1 INTRODUCTION

Cultural heritage comprehends a collection of distinctive traditions, items, locations, principles, and artistic expressions that have developed over time across various countries and regions. Each culture has a distinct cultural heritage that is continuously passed down from one generation to the next, facilitating learning and sharing. Digital archiving of cultural heritage collections plays a crucial role in preserving and providing access to cultural heritage materials [1]. Classification is highlighted as essential in organizing and managing digital archives for educational purposes [2]. Information retrieval tools and classification systems are emphasized in supporting conservation efforts and managing cultural heritage information effectively [3]. Challenges of digital preservation in cultural heritage institutions likely touch upon the importance of classification in preserving and maintaining access to digital cultural heritage materials [4]. Since digital technologies are constantly evolving, so too are the opportunities they present for developing and expanding. When it comes to cultural legacy, digital technologies allow for its easy accessibility, preservation, and even recreation.

Cultural heritage comes in many forms and may be found all around the world. Some examples are historical monuments, records, photos, and archaeological sites. Information about heritage sites is required for their preservation, upkeep, and rehabilitation. Because of the volume and quality of the material, it is difficult to archive, record, and disseminate the knowledge of these cultural valuables. Therefore, these new technologies can be an effective instrument for developing a new approach and enhancing traditional standards of heritage measurement and documentation. Integration of many information types is necessary for digital heritage documentation, including historical documents, 3D models, photos, thermographs, and multispectral photographs. Documentation of cultural heritage must, of course, take into account more than simply the raw data; relevant metadata and paradata must also be taken into account [5, 6]. The emergence of networking and crowd technology, smart devices, machine learning, and high-performance computing facilities has made it feasible to digitize the processes involved in heritage site documentation, preservation, and information availability. In most circumstances, there is no way to tell people when or where the information is already available, thus every building has hundreds or even thousands of

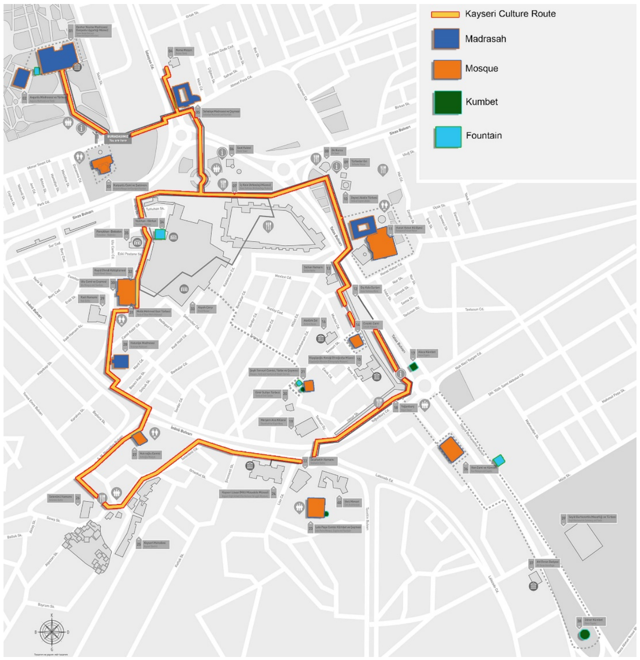

Figure 1: Kayseri Culture Route [7].

images (including ones from historical archives). Frequently, the same information has been submitted at least twice.

One of the important crossroads of Anatolia, Kayseri, one of the most important cities in the world, with a history dating back nearly six thousand years, Old Bronze, Hittite, Phrygian, Hellenistic, Roman Byzantine, Seljuk, Ottoman and Republican periods has a rich cultural heritage consisting of artefacts. Kayseri is a developing city with deep historical roots and traces. The city is being shaped, populated, and expanded under the influence of sectors such as industry, services, sports and tourism. In addition, being one of the limited examples of the concept of a planned city and with its strong economic structure, it has always maintained and preserved its characteristic of being a dominant and central city. The *Cultural Route* project, which was selected as the sample area as shown in Fig. 1, is a rare urban project study and implementation work to ensure that the cultural richness and historical buildings in Kayseri city center are better perceived by local and foreign users and to facilitate their visit.

As Kayseri's urban identity is approaching a tourism city with the investments made in addition to being an industrial city, such a project is also important to reveal the multi-layered historical structure of the city. Therefore, it is aimed to provide a rethinking and a different perspective on the historical urban fabric and the relationship between the buildings themselves and their surroundings by combining closely and distantly related buildings on a walking route. In the context of these objectives, an inventory list has been prepared to determine the architectural and historically important buildings in and near the center. Following this list, the structures that must be visited were determined and a route connecting these structures was prepared. This route covers the cultural heritage buildings located in the center of Kayseri. These artefacts are classified as *mosques, madrasahs, fountains* and *kumbets*. In this context, this classification study using deep learning method is expected to contribute to the existing cultural route sample area in terms of recognition, inventory creation and time saving.

Classification of cultural heritage images is a significant area of research that involves various techniques and methodologies. Several studies have contributed to this field, focusing on different aspects such as deep learning, image feature extraction, and metadata schema evaluation. Wang et al. discussed the use of deep semantic annotation for cultural heritage images, emphasizing the importance of vocabularies Union List of Artists Names for image representation [8]. Jankovic compared machine learning models for cultural heritage image classification, highlighting the use of public datasets for this purpose [9]. Furthermore, [10] explored the application of deep learning techniques, particularly convolutional neural networks (CNNs), for image classification and annotation tasks in cultural heritage. Belhi et al. emphasized the reliability and cost-effectiveness of visual recognition techniques in classifying and annotating cultural heritage images [11]. Grilli et al. utilized CNNs for the classification of heritage point clouds, demonstrating the use of machine and deep learning strategies in this context [12]. Additionally, Qin et al. presented a statistical system based on image feature extraction technology for cultural heritage tourism information, showcasing the integration of computer science in heritage studies [13]. Overall, these studies collectively contribute to the advancement of classification techniques for cultural heritage images, highlighting the importance of deep learning, image feature extraction, and metadata evaluation in this domain.

The utilization of CNNs in image classification tasks has significantly impacted various domains. CNNs have played a crucial role in revolutionizing image classification, leading to breakthroughs in recognizing objects, faces, handwritten digits, and more [14]. These advancements underscore the practical relevance of CNNs in real-world applications, demonstrating their effectiveness in accurately identifying and categorizing visual content [15]. By employing CNNs in image classification tasks, automatic feature learning is facilitated without the need for manual feature extraction, thereby streamlining the process and improving efficiency [16]. Furthermore, deep learning techniques, including CNNs, have shown remarkable success in fields such as computer vision, medical imaging, and image classification [17]. Overall, the application of deep learning techniques, particularly CNNs, in the classification of cultural images not only enhances the accuracy and efficiency of the classification process but also paves the way for innovative solutions in preserving and analyzing cultural heritage through digital means [18].

In this paper, an image dataset is generated for Kayseri Culture route consisting of a total 268 images containing four different types of cultural heritage, classified successfully by using SqueezeNet deep learning techniques for the first time.

2 DEEP LEARNING WITH SQUEEZENET CONVOLUTIONAL NEURAL NETWORKS

Inspired by the architecture and operation of the human brain, artificial neural networks (ANNs) are the main focus of deep learning,

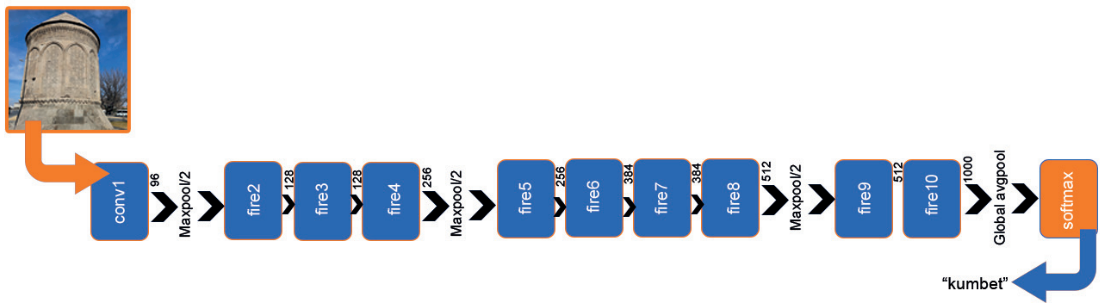

Figure 2: Architectural block diagram of SqueezeNet CNN deep learning model.

a branch of machine learning. It requires creating sophisticated, multi-layered structures that let computers learn from enormous volumes of data and, devoid of explicit programming, make wise choices or predictions. Deep learning models continuously improve their performance by iteratively adjusting their parameters to reduce the error between expected to and actual outputs through a process called backpropagation. With its astounding performance in tasks like image classification, language translation, and speech synthesis, deep learning has transformed a number of fields, including computer vision, natural language processing, and speech recognition. It is an effective tool for solving complicated problems in a variety of domains due to its capacity to automatically extract sophisticated features from raw data and recognize intricate patterns. Deep learning's applications are growing as research and processing power increase, providing previously unheard-of chances for creativity and artificial intelligence improvement.

An advanced class of ANNs called CNNs is made expressly to process and analyze structured grid data, like photographs. The distinct design of these networks, which combines fully connected, pooling, and convolutional layers, is what makes them distinctive. CNNs can efficiently train and extract features at many levels of abstraction by making use of the spatial locality and hierarchical structure of visual input. The network can identify patterns and features like edges, textures, and forms thanks to the convolutional layers, which apply filters or kernels to the input data. The dimensionality of the feature maps is decreased by subsequent pooling layers, improving computing efficiency and translation invariance. To create predictions or classifications, fully linked layers incorporate the high-level information that was extracted by preceding layers. CNNs are now the mainstay of cutting-edge methods for image analysis and recognition because of their outstanding performance in a variety of computer vision tasks, such as object detection, semantic segmentation, and image classification. The reason for their effectiveness is that they can automatically acquire hierarchical representations from unprocessed data, which makes them suitable for a variety of visual identification tasks in both academic and industrial contexts.

Through transfer learning, SqueezeNet, a pre-trained CNN for image classification, provides adaptability for different classification tasks. Fig. 2 shows the architecture of SqueezeNet [19]. This CNN was trained on an extensive dataset with more than a million photos; thus, it is ready for customization. Images that are indicative of the target classification domain are fed into the CNN architecture throughout the adaption phase. The dataset is purposefully restricted to highlight important architectural elements in order to accelerate learning. One possible way to optimize the learning process is to freeze the parameters of the network's earlier layers and set their learning rates to zero. Deep learning techniques like transfer learning make use of pre-existing networks to accelerate learning on novel problems. It is easier and faster to fine-tune a pre-trained network than to build a network from scratch with randomly initialized weights. Even with a limited set of training images, this method enables the quick transfer of acquired features to new duties. Specifically, SqueezeNet requires that the input image data have dimensions of 227x227x3, which stand for width, height, and color channels (RGB), in that order.

3 KAYSERI CULTURE ROUTE DATASET

The architectural spatial elements on the Kayseri Culture Route, which was selected as the case study area, are in the city center within the framework of a specific route that includes important historical buildings, most of which are located within the outer castle walls. Starting from Gevher Nesibe Madrasah, the route offers visitors 40 stops that reveal the historical and cultural heritage of Kayseri. Starting from the square in front of the Gevher Nesibe Madrasah and including Cumhuriyet Square, Hunat, Yoğunburç, Kayseri Quarter and Camiikebir, the route includes traces of the Roman Empire, the Byzantine Empire, the Seljuk Empire, the Ottoman Empire and the Republic of Türkiye. Since Kayseri is one of the most important provinces in Türkiye with the artefacts of the Seljuk Civilization, of which Kayseri was the second capital, the buildings from the 12th century and onwards stand out on the Culture Route.

Among the 40 artefacts on the trail, architectural space elements with the same typologies were classified into 4 different groups as *mosques, madrasahs, fountains* and *kumbets.* As mosques, Kurşunlu

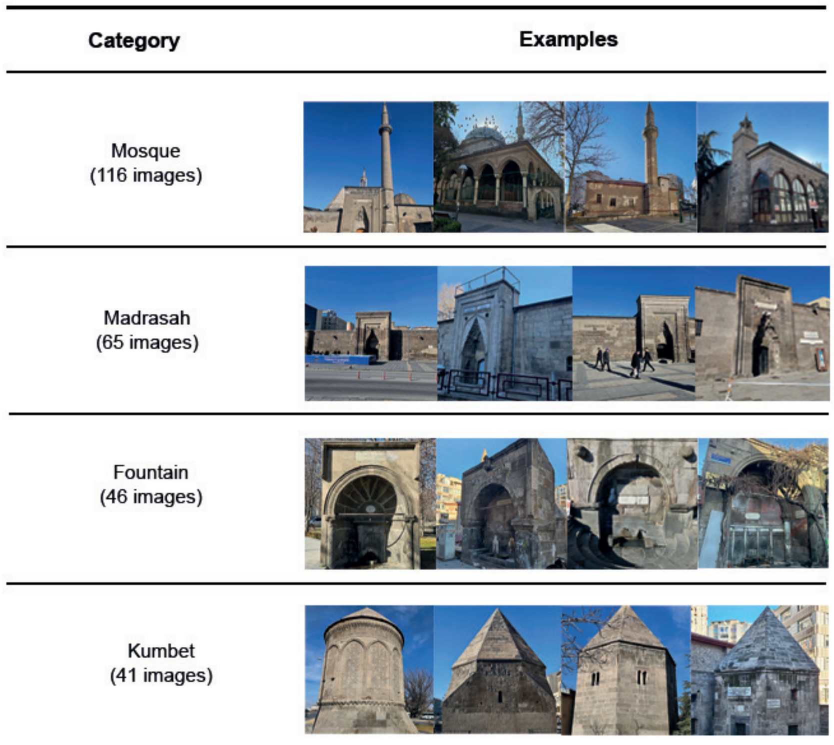

Figure 3: Sample images from Kayseri Culture Route image dataset.

Mosque, Ulu Mosque, Hunat Mosque, CıncıklıMosque, Şeyh Tennuri Mosque, Lala Paşa Mosque, Han Mosque, Hatıroğlu Mosque, HacıKılıç Mosque were identified. Fountains; Şeyh Müyessef Fountain, Şeyh Tennuri Fountain, Avgunlu Fountain. Kumbets; Alaca Kumbet, Döner Kumbet, Lala Pasha Kumbet, Şeyh Tennuri Kumbet.

The photos were taken with Nikon D700 camera with a Tamron f/2.8 28-75 mm lens, Samsung Galaxy Z Fold 5 and iPhone 14 cameras on 7 February 2024 in sunny weather. The photographs were then converted to 227x227 pixel size and processed for pretrained deep-learning models. The dataset contains a total of 268 images in four categories. Fig. 3 shows some samples for each category.

4 EXPERIMENTAL RESULTS

In this study, the classification of four different types of cultural heritage images was realized by using a pre-trained deep learning model called SqueezeNet. The number of training samples used in a single optimization algorithm iteration to update the neural network's weights is referred to as the mini-batch size. When a neural network is being trained, an epoch is one full run of the training dataset. The effect of mini batch size and epoch number was analyzed in the experiments. 70% percent of the dataset was used in training stage and the remaining 30% was used for testing the model. CNNs are frequently trained using an adaptive learning rate optimization approach called Root Mean Square Propagation (RMSprop). By dynamically altering the learning rate for each parameter in response to the magnitude of previous gradients, RMSprop, overcomes the drawbacks of conventional gradient descent techniques. All the experiments were repeated 30 times due to the random nature of the algorithms. Mini batch size numbers tested were 8, 16 and 32, respectively. Epoch numbers evaluated were 4, 8, and 16, respectively.

Accuracy in deep learning-based classification is the fraction of correctly classified examples in the dataset relative to the total number of examples. It is among the measures that are most frequently used to assess how well categorization models are performing:

$$Accuracy\ (\%) = \frac{Number\ of\ correctly\ classified\ samples}{Total\ number\ of\ samples} \times 100$$

Experiments were conducted on a personal computer with Intel i7-4790K @ 4 GHz processor and 16GB of RAM by using Matlab 2023a software.

Table 1 shows the classification accuracy results of the cultural images test dataset by using SqueezeNet trained by the help of transfer learning technique. Experiments were repeated 30 times for each case and thus, the values in the table are the average accuracy values for each case. The standard deviations are given

Table 1: Classification accuracy of cultural images by SqueezeNet deep learning model (test dataset).

Accuracy (%)		Epoch Number		
		4	8	16
Mini Batch Size	8	95.11 (10.26)	98.83 (2.14)	99.15 (2.69)
	16	93.99 (6.71)	98.30 (5.02)	99.79 (0.67)
	32	85.48 (10.53)	94.20 (5.29)	98.67 (2.55)

Table 2: CPU time consumption of the training stage for cultural image classification by using SqueezeNet deep learning model.

CPU Time (sec)		Epoch Number		
		4	8	16
Mini Batch Size	8	91.73	160.39	293.16
	16	88.01	164.00	270.15
	32	86.31	145.74	259.65

in the parenthesis as well. As it is seen from Table 1, better values were obtained by using higher epoch numbers. However, higher mini batch size gives worse results. The standard deviation is also higher with low epoch numbers which indicates the instability of the model. Overall best result was obtained when the mini batch size is 16 and epoch number is 16. As a result, the classification of the cultural images of Kayseri Culture Route is successfully realized with an accuracy of 99.8%.

Table 2 shows the CPU time consumption of training stage of each case. Using 4 epochs, the training stage was completed in ~90 seconds. For 8 and 16 epochs, the training of SqueezeNet model was completed in ~160 and ~280 seconds, respectively. CPU time consumption is related with the complexity of the model and data size. Lower CPU times are preferable in many real-life cases. In Fig. 4, training progress of a sample experiment case is visualized. As can be seen from the figure, training accuracy and validation (test) accuracy increases, while the loss function decreases during the epochs.

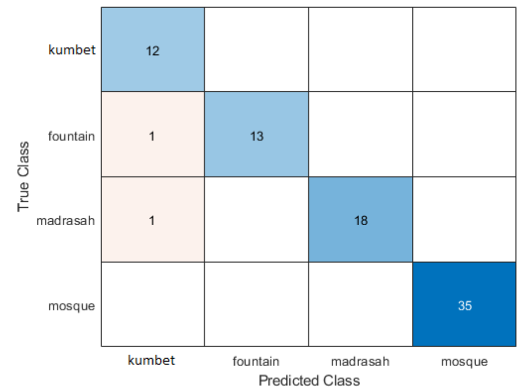

Figure 5: A sample confusion matrix from the classification experiments.

A table that is frequently used to assess how well a classification model is performing is called a confusion matrix. By combining the model's predictions with the actual labels in the dataset, it enables an in-depth examination of the predictions. A sample confusion matrix from one of the experiments is shown in Fig. 5.

5 CONCLUSION

This research concludes by highlighting how crucial architectural structure image categorization is to the digital preservation of cultural assets. The tedious and error-prone nature of arranging a

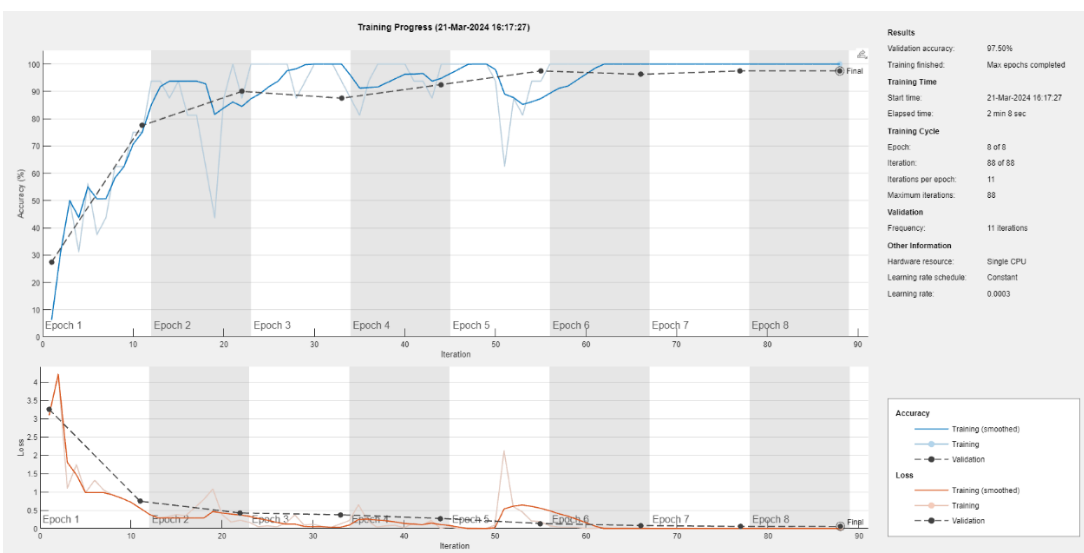

Figure 4: A sample of SqueezeNet CNN deep learning training progress.

large number of pictures makes the use of automatic procedures necessary to expedite the process. This paper shows how deep learning, and more specifically SqueezeNet convolutional neural networks (CNNs), can be used to improve the effectiveness of digital documentation procedures by classifying images.

This paper focuses on the recognition of important cultural heritage monuments of Kayseri Culture Route. To encourage more research in this area, a new dataset was developed. The study emphasizes the value of transfer learning, which is the process of adapting CNN models that have already been trained to categorization problems. An accuracy of 99.8% is achieved especially when SqueezeNet CNNs are used, highlight how effective this strategy is at raising the quality of architectural heritage documentation.

Overall, the outcomes point to the possibility of improving image classification and analysis within the framework of architectural heritage preservation through the incorporation of deep learning techniques. Research, interpretation, and conservation efforts in the field of cultural heritage are facilitated by these tools, which make it possible to organize and retrieve digital data more efficiently.

REFERENCES

[1] Tiurmenko, I., Bozhuk, L., Struk, I., & Syerov, Y. (2022). Digital documentary collections of national cultural heritage on the ukrainian regional state archives websites., 449-470. https://doi.org/10.1007/978-3-030-97008-6_20

[2] Georges, A., Logghe, S., & Schuurman, D. (2015). Developing an audiovisual cultural heritage platform for educational purposes: a case study of teacher-involvement using a living lab approach. International Journal of Heritage in the Digital Era, 4(1), 87-102. https://doi.org/10.1260/2047-4970.4.1.87

[3] Kara, G. (2021). Developing a sustainable cultural heritage information system. Library Hi Tech News, 38(6), 17-20. https://doi.org/10.1108/lhtn-08-2021-0053

[4] Evens, T. and Hauttekeete, L. (2011). Challenges of digital preservation for cultural heritage institutions. Journal of Librarianship and Information Science, 43(3), 157-165. https://doi.org/10.1177/0961000611410585

[5] Apollonio, F.I.; Giovannini, E.C. (2015). A paradata documentation methodology for the Uncertainty Visualization in digital reconstruction of CH artifacts. SCIRES-IT 2015, 5, 1–24. http://dx.doi.org/10.2423/i22394303v5n1p1

[6] Di Giulio, R.;Maietti, F.; Piaia, E.;Medici,M.; Ferrari, F.; Turillazzi, B. Integrated Data Capturing Requirements for 3d Semantic Modelling of Cultural Heritage: The INCEPTION Protocol. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2017, XLII-2/W3, 251–257. https://doi.org/10.5194/isprs-archives-XLII-2-W3-251-2017

[7] Kayseri Metropolitan Municipality (2024). Kayseri Culture Road Map. Available: https://www.kayseri.bel.tr/kesfet-listeleme/kultur-yolu

[8] Wang, X., Song, N., Liu, X., & Xu, L. (2021). Data modeling and evaluation of deep semantic annotation for cultural heritage images. Journal of Documentation, 77(4), 906-925. https://doi.org/10.1108/jd-06-2020-0102

[9] Jankovic, R. (2019). Machine learning models for cultural heritage image classification: comparison based on attribute selection. Information, 11(1), 12. https://doi.org/10.3390/info11010012

[10] Karterouli, K. and Batsaki, Y. (2021). Ai and cultural heritage image collections: opportunities and challenges.. https://doi.org/10.14236/ewic/eva2021.33

[11] Belhi, A., Bouras, A., & Foufou, S. (2018). Leveraging known data for missing label prediction in cultural heritage context. Applied Sciences, 8(10), 1768. https://doi.org/10.3390/app8101768

[12] Grilli, E., Özdemir, E., & Remondino, F. (2019). Application of machine and deep learning strategies for the classification of heritage point clouds. The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences, XLII-4/W18, 447-454. https://doi.org/10.5194/isprs-archives-xlii-4-w18-447-2019

[13] Qin, L., Chen, S., Huang, J., & Lin, H. (2022). Statistical system of cultural heritage tourism information based on image feature extraction technology. Mathematical Problems in Engineering, 2022, 1-12. https://doi.org/10.1155/2022/5250853

[14] Meena, S., Dhaka, V., Sinwar, D., Kavita, .., Ijaz, M., & Woźniak, M. (2021). A survey of deep convolutional neural networks applied for prediction of plant leaf diseases. Sensors, 21(14), 4749. https://doi.org/10.3390/s21144749

[15] Yu, Q. (2022). Animal image classifier based on convolutional neural network. SHS Web of Conferences, 144, 03017. https://doi.org/10.1051/shsconf/202214403017

[16] Liu, L. (2021). Image classification in htp test based on convolutional neural network model. Computational Intelligence and Neuroscience, 2021, 1-8. https://doi.org/10.1155/2021/6370509

[17] Adige, S., Kurban, R., Durmuş, A., & Karaköse, E. (2023). Classification of apple images using support vector machines and deep residual networks. Neural Computing and Applications, 35(16), 12073-12087. https://doi.org/10.1007/s00521-023-08340-3

[18] Llamas, J., M. Lerones, P., Medina, R., Zalama, E., & Gómez-García-Bermejo, J. (2017). Classification of architectural heritage images using deep learning techniques. Applied Sciences, 7(10), 992. https://doi.org/10.3390/app7100992

[19] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360. https://doi.org/10.48550/arXiv.1602.07360