Archeological Sites Classification Through Partial Imaging and Convolutional Neural Networks

https://doi.org/10.3991/ijoe.v19i07.39045

Yaser Saleh(✉), Muhanna Muhanna
Princess Sumaya University for Technology, Amman, Jordan
y.saleh@psut.edu.jo

Abstract—In this paper, a novel approach for classifying archeological sites using publicly available images through the use of Convolutional Neural Networks (CNNs) is presented. To surmount the problem of having a limited amount of data to use in training and testing the CNNs, our approach employs the technique of fine tuning. We conducted an experiment with four popular CNN architectures: VGG-16, VGG-19, ResNet50, and InceptionV3. The results show that our models achieved an impressive accuracy of up to 98% using the VGG-16 and InceptionV3 models and up to 97% using the ResNet50 model, while the VGG-19 model produced results with an accuracy of 95%. The results of this study demonstrate the effectiveness of our proposed approach in classifying archeological sites using publicly available images and highlight the potential of deep learning techniques for archeological site classification.

Keywords—archeological sites classification, deep learning, fine tuning, partial recognition

1 Introduction

With the improvement in image capturing technology, like smartphone and dashboard cameras, ordinary individuals are now able to efficiently capture and share information about their experiences, events, and journeys. This creates vast amounts of new data that require analysis and examination in order to uncover new insights and findings.

Archeological sites classification using deep learning is an emerging field that aims to use the power of deep learning algorithms to classify and identify archeological sites. The goal is to automate the process of identifying and classifying these sites, which can be time-consuming and labor-intensive when done manually. The use of deep learning in this field has the potential to significantly improve the efficiency and accuracy of archeological site classification.

Deep learning algorithms, particularly convolutional neural networks (CNNs), have been successfully applied in various image classification tasks, including natural images and satellite images. These algorithms are able to learn hierarchical representations of images, which allows them to identify and classify objects within the images.

This makes them well-suited for the task of archeological site classification, as they can be trained to identify the unique characteristics of archeological sites in satellite or aerial images.

One of the main challenges in archeological site classification using deep learning is the limited availability of labeled training data. However, recent advancements in transfer learning and data augmentation techniques have made it possible to use pre-trained networks and generate synthetic data to overcome this challenge.

Overall, the use of deep learning in archeological site classification is a promising field with the potential to greatly improve the efficiency and accuracy of identifying and classifying these important historical sites.

The organization of this paper is as follows: The introduction is presented in Section 1, while related work in the field is discussed and analyzed in Section 2. Section 3 provides a detailed explanation of the models used, including the process of collecting the data and fine-tuning. The results of the study are then presented and discussed in Section 4. Finally, the conclusion and suggestions for future work are summarized in Section 5.

2 Related work

This section explores and discusses relevant research, including various machine learning and deep learning approaches, as well as the classification of archeological sites.

Deep learning has been widely used in image classification in recent years. Convolutional neural networks (CNNs) have been particularly successful in this task, due to their ability to learn hierarchical representations of images.

One early success of CNNs in image classification was the AlexNet, which won the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) by a large margin [1]. Since then, many other CNN architectures have been proposed and have shown to be effective in image classification, such as VGGNet [2], GoogleNet which later was renamed the Inception model [3], and ResNet [4].

One of the key advantages of CNNs is their ability to learn features from images directly, rather than relying on hand-designed features as in traditional image classification methods. This has led to improved performance on a wide range of image classification tasks [5].

Another important aspect of CNNs for image classification is their ability to be fine-tuned for specific tasks using transfer learning. This allows for the use of pre-trained CNNs as a starting point for image classification tasks with limited data, which can greatly improve performance [6].

Fig. 1. Samples of the original images for three of the six sites used [7]

In recent years, there has been a significant increase in the use of deep learning in medical imaging. CNNs have shown great promise in this field, with applications including tumor detection [8], segmentation [9], and diagnosis [10].

Archeological sites are important cultural and historical resources that provide valuable insights into past human societies and civilizations. The identification and documentation of these sites are crucial for their preservation and protection, but the task can be challenging, especially in large and remote areas where manual surveys are not feasible. In recent years, image classification techniques, particularly deep learning, have been applied to overcome this challenge.

Deep convolutional neural networks (CNNs) have been shown to be effective in identifying and classifying archeological sites. [11] used CNNs for the remote sensing investigation of looting at the archeological site of Al-Lisht, Egypt, and demonstrated the potential of CNNs in identifying looting damage. Similarly, [12] applied CNNs for the automated recognition of historical courtyard houses in Yazd, Iran, aiming at the recognition of historical and non-historical buildings employing airborne and satellite imagery, where the model achieved an accuracy of 98%.

In addition to CNNs, other deep learning techniques have been applied to the detection of archeological structures. [13] reviewed the application of deep learning for the detection of archeological structures and found that Region-based Convolutional Neural Network (R-CNN) and Mask Region-based Convolutional Neural Network (MR-CNN) have been the most suitable techniques. They also highlighted the potential of deep learning in improving the efficiency and accuracy of archeological site detection.

However, the application of deep learning techniques in archaeology also has some limitations. [14] studied the potential and limitations of designing a deep learning model for discovering new archeological sites and found that the model's performance depends on the quality and quantity of the data used for training.

Another application of deep learning in archaeology is the detection of specific types of archeological sites, such as tombs. [15] applied CNNs for the detection of tombs in satellite images and achieved an accuracy of 98%. This study highlights the potential of deep learning in identifying specific types of archeological sites, which can be useful in targeted surveys and excavations.

Fig. 2. Samples of the patches taken from different archeological sites

In conclusion, deep learning techniques, particularly CNNs, have been shown to be effective in identifying and classifying archeological sites in satellite imagery, aerial photography, and UAV images. These techniques have the potential to significantly improve the efficiency and cost-effectiveness of archeological site identification and documentation. However, further research is needed to address the limitations of deep learning in archaeology and to explore its application to different types of images and specific types of archeological sites.

3 Methodology

The methods and models utilized in this study are outlined in this section. Additionally, the data collection process, including the patch-based strategy and fine-tuning procedures, will be detailed.

3.1 Utilized models

A convolutional neural network (CNN) works by inputting the data (images) and going through the layers of the network from first to last. The layers of any CNN normally consist of convolution, pooling and fully connected layers, where at the final layer the network would be able to give a result of a specific class from the group of all probable classes for the entered image.

Taking a look at some of the most popular networks that have been used for image classification, the VGGNet architecture which had achieved the first place in the ImageNet Challenge, has a unique advantage by having less layers in comparison to other state-of-the-art, while yet providing considerably excellent results when trained on the ImageNet dataset with an error rate of around 7% on the accuracy [2]. Another strength of the VGGNet network is that it's distributed and widely used through different artificial intelligence frameworks, making it the textbook model to start experimenting with. Perceiving that convolutional layers with are symbolized with their parameters as "con-quantity of channels" the architectures used (16 and 19 layers) can be defined as follows starting with the 16 layers' model:

- input (224 × 224 RGB image)
- con-64; con-64; maxpool
- con-128; con-128; maxpool
- con-256; con-256; con-256; maxpool
- con-512; con-512; con-512; maxpool
- con-512; con-512; con-512; maxpool
- Fully Connected (FC); FC; FC; soft-max

While the 19 layers' model structure can be explained as follows:

- input (224 × 224 RGB image)
- con-64; con-64; maxpool
- con-128; con-128; maxpool
- con-256; con-256; con-256; con-256; maxpool
- con-512; con-512; con-512; con-512; maxpool
- con-512; con-512; con-512; con-512; maxpool
- Fully Connected (FC); FC; FC; soft-max

At the end, with the purpose of adding more networks to the experiments, the InceptionV3 [3], and the ResNet50 model is included [4]. The Inception model provided a new technique in constructing a convolutional neural network, where the model presented an inception architecture, which contained three convolutions with filters of 1×1, 3×3, and 5×5 in size. The construction of the model can be described like so (the number of filters would be included in parentheses after the inception):

- input (224 × 224 RGB image)
- con-32; con-32; maxpool
- con-64; con-80; con-192
- inception (inc) (35, 35, 288); inc (17, 17, 768); inc (8, 8, 1280); maxpool
- Global pool layer; soft-max

Examining the ResNet50 model, a similar approach to the VGG models was used including 50 layers with layers distributed as follows:

- input (224 × 224 RGB image)
- con-64; con-64; con-256; maxpool
- con-64; con-64; con-256; maxpool

- con-64; con-64; con-256; maxpool
- con-128; con-128; con-512; maxpool
- con-128; con-128; con-512; maxpool
- con-128; con-128; con-512; maxpool
- con-265; con-265; con-1024; maxpool
- con-265; con-265; con-1024; maxpool
- con-265; con-265; con-1024; maxpool
- con-512; con-512; con-2048; maxpool
- con-512; con-512; con-2048; maxpool
- con-512; con-512; con-2048; Global pool layer; soft-max

3.2 Data collection

As mentioned previously, Google image search engine and Google Street View were used to collect a set of images for six different famous archeological sites in Jordan, the images had to have high quality in order to be able to create patches from the originals, this resulted of a problem which was the lack of high-quality images resulting in only 54 images collected. The sites were categorized into the following: Petra Treasury, Jerash Ruins, Ajloun Castle, Amra Castle, Kerak Castle, and Wadi Mujib (a sample is shown in Figure 1).

After the process of collecting the data, each image was split into smaller patches while disregarding any surrounding areas of the site such as the sky or roads, this process created a dataset of around 1200 patches. Each cut-out was taken as a square with a width and height of 224 pixels, and nominal enlarging was required. Samples can be seen in Figure 2.

3.3 Fine-tuning

The previous description outlined the use of two VGG models, the InceptionV3 model, and the ResNet50 model for the training and fine-tuning process. These models were chosen due to their established performance. The fine-tuning process allowed the network to avoid the need for a large dataset and reduced the number of epochs required, making it suitable for both large and small datasets.

To start the process, matching models were created, and the pre-trained weights were loaded into the different networks. A soft-max layer was added in the last part of the network as the final layer, next to the last fully connected layer, to perform the classification. This prepared the models for fine-tuning, which was accomplished by running additional training iterations using new data.

The models were then trained using a patch dataset of images, with 80% of the data designated for the training set and the remaining 20% set aside for testing. The training involved 60 epochs, a complete pass through the network using the training dataset. In each epoch, the dataset was divided into groups and forwarded through the network, and the function for the loss is computed using categorical cross-entropy using the prediction and target labels. A Stochastic Gradient Descent with Nesterov momentum was applied using a learning rate of 0.0001, momentum of 0.9, and a learning rate of 0.0001, which controlled the size of the update steps.

Described in [16], the "nesterov_momentum" function, where it would generate update expressions with the following forms:

$$\text{velocity} \coloneqq \text{momentum} \times \text{velocity-learning\_rate} \times \text{gradient} \tag{1}$$

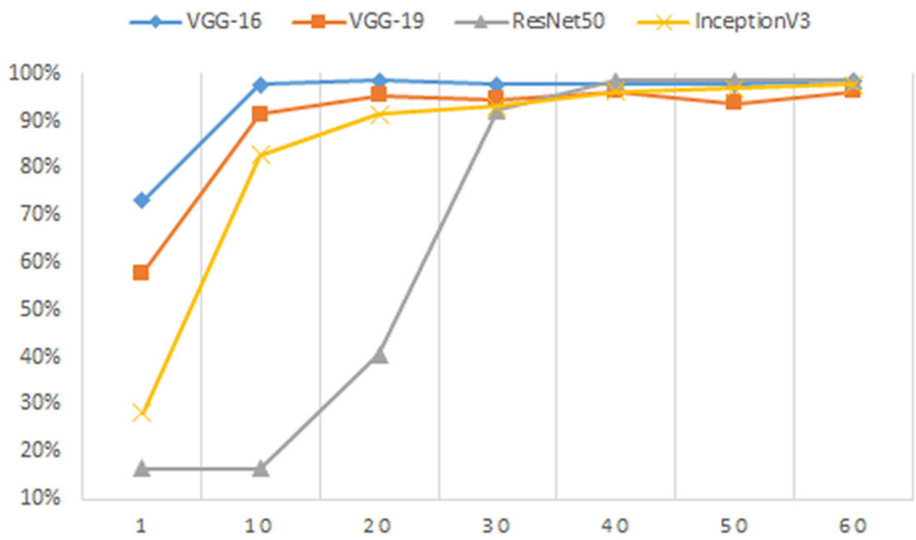

Fig. 3. The accuracy progress of the used models

As explained before, the loss was measured using a cross-entropy function. Considering p = targets tensor and q = predictions tensor, the function of cross-entropy is described as:

$$H(p,q) = -\sum_{x} p(x) log(q(x)) \tag{2}$$

After each epoch, results were displayed that included the runtime, the total loss function, and the validation accuracy. These results tracked the development of the training procedure during the epochs run, and showed when the maximum conceivable accuracy was achieved.

4 Results and analysis

The fine-tuning process for the four models (VGG-16, VGG-19, InceptionV3 and ResNet50) was carried out as described previously, and the first training step involved running 60 epochs on the VGG-16 network. The highest validation accuracy of 98.4% was reached at the 19th epoch, as shown in Table 1.

For the VGG-19 model, the highest accuracy of 95.3% was reached after running 60 epochs, at the 5th epoch, as shown in Table 2. The ResNet50 model achieved an accuracy of 98.4% at the 40th epoch after running 60 epochs, as seen in Table 3. Finally, the InceptionV3 model produced an accuracy of 97.6% after the 59th epoch, as presented in Table 4, and an overview of the accuracy growth during the training process can be examined in Figure 3.

Figures 4–7 depict the confusion matrices which illustrate the accuracy of the models in the process of classification and validation in the models. It was observed that the VGG-19 model showed the lowest level of confidence, particularly in the case of classifying the Mujib archeological site.

Table 1. Results of epochs 15–19 using the VGG-16 network, showing the validation loss and validation accuracy after each epoch

Epoch #	Validation Loss	Validation Accuracy
15	0.1390	0.9609
16	0.1272	0.9766
17	0.1112	0.9766
18	0.1361	0.9766
19	0.1282	0.9844

Table 2. Results of epochs 1–5 using the VGG-19 network, showing the validation loss and validation accuracy after each epoch

Epoch #	Validation Loss	Validation Accuracy
1	1.3390	0.5781
2	0.8207	0.7734
3	0.4776	0.8750
4	0.4776	0.8906
5	0.2610	0.9531

Table 3. Results of epochs 36–40 using the ResNet50 network, showing the validation loss and validation accuracy after each epoch

Epoch #	Validation Loss	Validation Accuracy
36	0.1021	0.9766
37	0.0990	0.9766
38	0.0958	0.9766
39	0.0941	0.9766
40	0.0917	0.9844

Table 4. Results of epochs 55–59 using the InceptionV3 network, showing the validation loss and validation accuracy after each epoch

Epoch #	Validation Loss	Validation Accuracy
55	0.1721	0.9688
56	0.1709	0.9688
57	0.1676	0.9688
58	0.1664	0.9688
59	0.1643	0.9766

5 Conclusion

In this research, as shown in previous work [17], [18], [19], an innovative method was produced which should aid in growing the research on the usage of deep learning methodologies on partially captured images of archeological sites, which will provide improvements on applicable areas such as security, surveillance, location detection, and much more.

Our primary aim was to demonstrate the effectiveness of deep learning techniques in processing images captured by common sensors, such as smartphone cameras. However, acquiring a large enough dataset to train a deep learning network was a major challenge. To overcome this challenge, a dataset was created and utilized effectively through the use of fine-tuning. Transfer learning, specifically fine-tuning, was deemed the most viable solution due to the requirement for a large dataset to train a Convolutional Neural Network (CNN). Additionally, patches of the original images were generated to create the dataset, providing a novel approach that allowed the models to recognize archeological sites with only a limited view of the location.

The VGG-16, VGG-19, InceptionV3, and ResNet50 networks were built and fine-tuned with the dataset created by taking pieces of the full images, where validation accuracy provided reached 98.4% and 95.3% for the VGG-16 and VGG-19 respectively, while the ResNet50 presented results that reached up to 98.4%, and the InceptionV3 with accuracy reaching 97.6%. Looking at the mentioned results, and also at the development and evolution of the different deep learning structures and models produced every year, it can be seen that future applications and uses for such accurate models is now possible.

In conclusion, the approach outlined in this study showcases the incredible potential of using images for archeological sites classification and highlights the benefits of using fine-tuning to overcome the challenges of a limited dataset. This innovative method of using patches of images to create the dataset not only avoids the need for a large collection of images, but also allows the models to recognize the sites with minimal view of the location, effectively reducing the need for high-quality images. The results of this study demonstrate the effectiveness of this approach, providing a clear comparison between well-known models and highlighting their strengths and limitations. This study underscores the power of using deep learning techniques in the field of archeology and provides a solid foundation for future research in this area. By utilizing fine-tuning, it is possible to train models on smaller datasets, enabling the use of deep learning techniques in areas where collecting large amounts of data may be difficult or impractical. This research opens up new opportunities for exploring the use of images in archeology and demonstrates the viability of using deep learning techniques for this purpose.

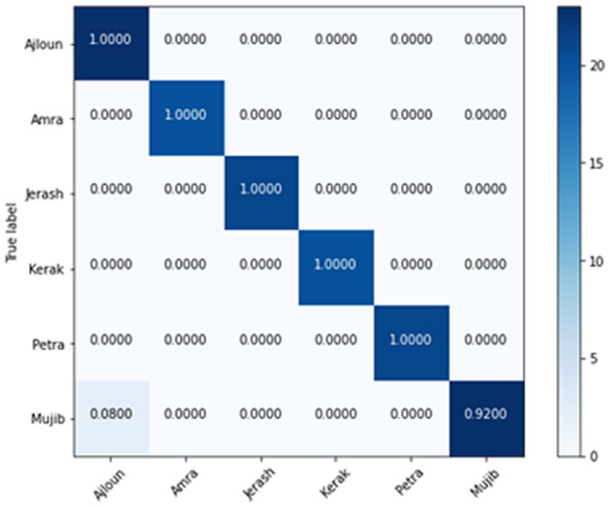

Fig. 4. Confusion matrix for the VGG-16 model

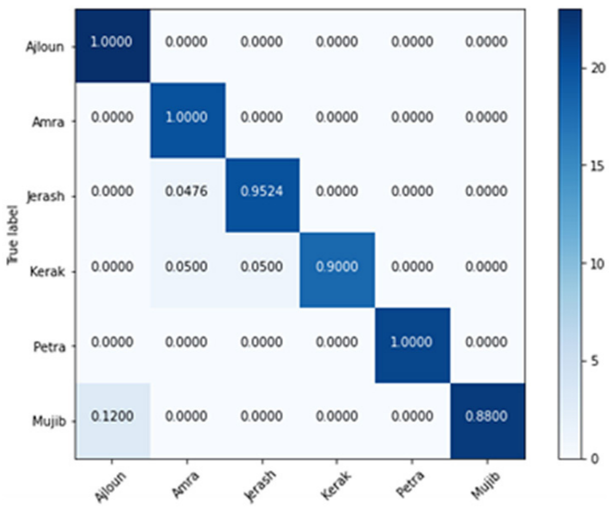

Fig. 5. Confusion matrix for the VGG-19 model

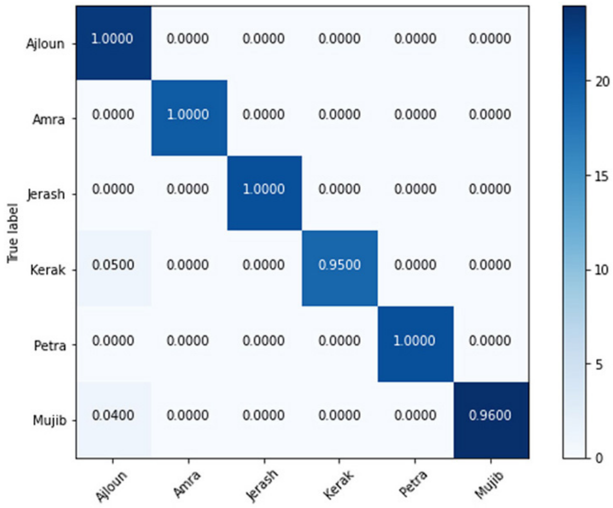

Fig. 6. Confusion matrix for the ResNet50 model

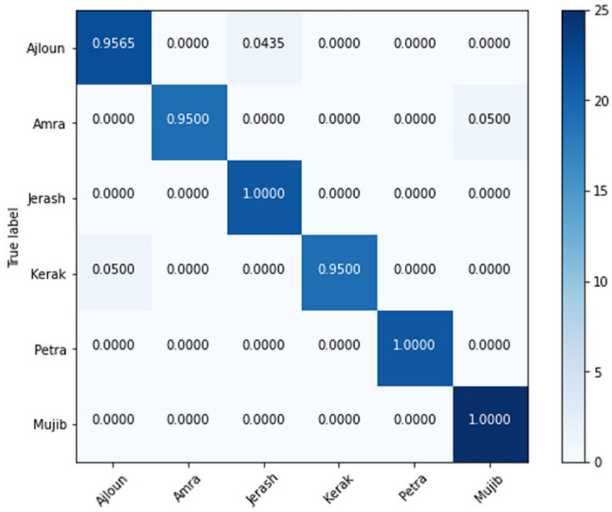

Fig. 7. Confusion matrix for the InceptionV3 model

6 References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Communications of the ACM, 2017, vol. 60, no. 6, pp. 84–90. https://doi.org/10.1145/3065386

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in arXiv preprint arXiv:1409.1556, 2014.

[3] C. Szegedy et al., "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594
[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90
[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015. https://doi.org/10.1038/nature14539
[6] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," Adv Neural Inf Process Syst, vol. 27, 2014.
[7] V. Jordan, "Attractions." 2023. [Online]. Available: https://international.visitjordan.com/Where-to-go
[8] W. Zhang et al., "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," Neuroimage, vol. 108, pp. 214–224, 2015. https://doi.org/10.1016/j.neuroimage.2014.12.061
[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, 2015, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
[10] G. Litjens et al., "A survey on deep learning in medical image analysis," Med Image Anal, vol. 42, pp. 60–88, 2017. https://doi.org/10.1016/j.media.2017.07.005
[11] T. Woolf, "Deep Convolutional Neural Networks for Remote Sensing Investigation of Looting of the Archeological Site of Al-Lisht, Egypt," 2018.
[12] H. Yazdi, S. Sad Berenji, F. Ludwig, and S. Moazen, "Deep learning in historical architecture remote sensing: Automated historical courtyard house recognition in Yazd, Iran," Heritage, vol. 5, no. 4, pp. 3066–3080, 2022. https://doi.org/10.3390/heritage5040159
[13] A. H. Jamil, F. Yakub, A. Azizan, S. A. Roslan, S. A. Zaki, and S. A. Ahmad, "A review on deep learning application for detection of archaeological structures," Journal of Advanced Research in Applied Sciences and Engineering Technology, vol. 26, no. 1, pp. 7–14, 2022. https://doi.org/10.37934/araset.26.1.714
[14] M. Roccetti, L. Casini, G. Delnevo, V. Orrù, and N. Marchetti, "Potential and limitations of designing a deep learning model for discovering new archaeological sites: a case with the mesopotamian floodplain," in Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good, 2020, pp. 216–221. https://doi.org/10.1145/3411170.3411254
[15] G. Caspari and P. Crespo, "Convolutional neural networks for archaeological site detection–Finding 'princely' tombs," J Archaeol Sci, vol. 110, p. 104998, 2019. https://doi.org/10.1016/j.jas.2019.104998
[16] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in International conference on machine learning, 2013, pp. 1139–1147.
[17] S. T. Ahmed and S. M. Kadhem, "Using machine learning via deep learning algorithms to diagnose the lung disease based on chest imaging: A survey," Int. J. Interact. Mob. Technol., vol. 15, no. 16, 2021. https://doi.org/10.3991/ijim.v15i16.24191
[18] S. Puttinaovarat, A. Saeliw, S. Pruitikanee, J. Kongcharoen, S. Chai-Arayalert, and K. Khaimook, "Flood damage assessment geospatial application using geoinformatics and deep learning classification," Int. J. Interact. Mob. Technol., vol. 16, no. 21, 2022. https://doi.org/10.3991/ijim.v16i21.34281
[19] M. A. Ali, A. J. Hussain, and A. T. Sadiq, "Deep learning algorithms for human fighting action recognition," Int. J. Online & Biomed. Eng., vol. 18, no. 2, 2022. https://doi.org/10.3991/ijoe.v18i02.28019

7 Authors

Yaser Saleh is an Assistant Professor at Princess Sumaya University for Technology in Jordan. He holds a Ph.D. degree in Computer Science from Loughborough University. His research interests include Artificial Intelligence, Deep learning Computer Vision, And Game Development. (email: y.saleh@psut.edu.jo)

Muhanna Muhanna is the Director of the Quality Assurance and Accreditation Center at Princess Sumaya University for Technology. He is also currently serving as an Assistant Professor in the Department of Computer Graphics and Animation. Dr. Muhanna earned his Ph.D. degree in Computer Science and Engineering (2011) from the University of Nevada, Reno, US. When it comes to academic research, Dr. Muhanna's main interest is in Human-Computer Interaction (HCI). Particular topics of study include mobile interaction design, Arabic interfaces, and heuristic evaluation. (email: m.muhanna@psut.edu.jo)

Article submitted 2023-02-21. Resubmitted 2023-03-27. Final acceptance 2023-03-29. Final version published as submitted by the authors.