Multimedia Tools and Applications (2020) 79:34129–34139
https://doi.org/10.1007/s11042-020-09232-7

GUEST EDITORIAL

Artificial intelligence in deep learning algorithms for multimedia analysis

Gwanggil Jeon[1] · Marco Anisetti[2] · Ernesto Damiani[3] · Burak Kantarci[4]

Published online: 16 July 2020

1 Introduction

Deep learning has gained a lot of research interest in artificial intelligence (AI) in many applications, such as image understanding, object detection, feature extraction, audio/video processing, image demosaicking and denoising, overhead views in industrial applications. In addition, exploitation of deep learning in the field of data science, particularly in big data analytics focuses on high-level feature extraction and abstraction as data representation based on the hierarchical learning process. Moreover, deep learning is also designed to tackle supervised learning problems for a wide variety of tasks. How to reliably solve unsupervised tasks with a similar degree of success is an important issue to address. Such studies are investigated based on the adoption of parallel computing, i.e., GPUs and CPUs clusters.

Various algorithms are already applied to achieve the desired goals. For instance, convolutional neural network has established superior performance on large-scale image and video classification. The supervised learning (semi/weakly) methods have expressively enhanced the performance when only a small amount of annotated data is available. In addition, correlation analysis, transfer learning, multi-tasking have proven

✉ Marco Anisetti
marco.anisetti@unimi.it

Gwanggil Jeon
gjeon@inu.ac.kr

Ernesto Damiani
ernesto.damiani@kustar.ac.ae

Burak Kantarci
burak.kantarci@uottawa.ca

[1] Incheon National University, Incheon, South Korea

[2] Università Degli Studi Di Milano, Milan, Italy

[3] Khalifa University, Abu Dhabi, UAE

[4] University of Ottawa, Ottawa, Canada

their effective contribution in integrating heterogeneous data. Moreover, clustering and sparse techniques are examined in demosaicking and denoising raw data. A large amount of audio, video, and text data being generated by machines that require efficient deep learning algorithms in terms of accuracy and efficiency. Another key aspect is to work on smaller datasets, which focuses on how we can get the gains of un-labelled instances with few labelled samples. Deep agents can play a vital role in the decision system where other deep learning techniques are used to focus on bridging the data gap between data and the application decision.

At this juncture, several challenges need to be addressed to reduce capability gaps. Indeed, increasing depth and width of audio/video and the emerging phenomenon of big dimensionality of data render the shortfalls of ensemble deep learning, thus increasing the number of challenges. Methods and techniques are needed to address the aforementioned constraints in terms of enhancing accuracy and efficiency, reducing the complexity, removal of noise, and to be more specific, novel frameworks are required that focus on the deep learning aspect of artificial intelligence while keeping the quality within the bounds in a real-time and context aware fashion.

2 Themes of this special issue

Starting from the above considerations, we aimed to provide a forum for researchers with an interest in efficiency to examine challenging research questions, showcase the state-of-the-art, and share breakthroughs. This edition of the special issue is focused primarily on artificial intelligence in deep learning algorithms for multimedia analysis. This issue is intended to provide a highly recognized international forum to present recent advances in *Multimedia Tools and Applications*. We welcomed both theoretical contributions as well as papers describing interesting applications. Papers were invited for this special issue considering aspects of this problem, including:

- Compact representation of text and speech data using AI in deep learning
- Fast and/or space efficient algorithms for analysis of audio/video using AI in deep learning
- Information retrieval using AI in deep learning
- Information retrieval algorithms for audio/video through the use of specialized hardware such as GPUs, FPGAs and quantum computers
- Parallel algorithms for AI in deep learning
- Deep learning applications in any aspects of perceptual tasks (e.g., object/face detection and recognition, image/video understanding, and audio/speech recognition)
- AI in deep multimodal learning
- AI in deep learning for data indexing and retrieval
- Multi-view and cross-view deep learning based visual content analysis
- Mathematical foundations of AI in deep learning
- Domain adaptation and transfer learning with AI in deep learning

After review, a total of 23 papers out of 100 submissions (23.0%) have been accepted for publication in this issue.

2.1 Models

Many applications require action recognition skills, from human-machine interaction to intelligent video surveillance. Action recognition in video sequences cannot be based on simply processing raw color images or optical flow fields. Color images provide appearance information of moving objects, but lack motion features. They are also very sensitive to variations due to clothing and camera pose that badly affect the action recognition accuracy. In turn, raw optical flow measures instantaneous motion, not the overall dynamics of actions, and is sensitive to noise. More robust and meaningful motion features and classifiers are thus required for action recognition to be reliable. The contribution by Rashwan et al. "Action representation and recognition through temporal co-occurrence of flow fields and convolutional neural networks" proposes a new action recognition technique based on a deep convolutional neural network (CNN) fed with histograms of optical flow co-occurrence (HOF-CO) motion features [14]. HOF-CO is a robust motion representation previously proposed by the authors to encode the relative frequency of pairs of optical flow directions computed at each image pixel. Their simulation results show that this approach outperforms state-of-the-art action recognition methods on three different public datasets KTH, UCF-11 Youtube and HOLLYWOOD2.

The contribution by Qu et al. "Scattering of aerosol by a high-order Bessel vortex beam for multimedia information transmission in atmosphere" studies essence of wireless communication and multimedia information transmission, which is the propagation of electromagnetic waves in the atmosphere [13]. Within the framework of Generalized Lorenz Mie theory, and combining the vector wave theory with the generalized multi-spheres Mie theory, the analytical solution to the scattering of the high-order Bessel vortex beam (HOBVB) by aerosol aggregation in atmosphere is investigated. The angle distributions of the scattered field of soot, silicate and nitrate aerosol cluster particles illuminated by a HOBVB are numerically discussed. The examples are selected to illustrate the effects of aggregation configuration, mean value, particle number, topological charge and half-cone angle of the beam on the angle distribution of scattered field. It is noticed that the angle distribution of scattered field is sensitive to the configuration of the cluster for the multiple refraction and interactive scattering. The variation of the mean value of radius of the aerosol aggregation will result in different scattering characteristics and different transmission efficiency. The integration of the scattering algorithm and deep learning can be used in inversion of the shape and components of the aerosol clusters and in the improvement of transmission efficiency of multimedia information in the atmosphere.

Ocean waves are complex systems with the contributions of wind waves and swells. The study on interaction mechanism between electromagnetic wave and actual sea surface is of significant importance in ocean remote sensing and engineering application, which is also helpful in the prediction and inversion of wave information. In the contribution by Wu et al. "Deep learning for inversion of significant wave height based on actual sea surface backscattering coefficient model", an efficient model for estimating backscattering coefficient is built, considering the characteristics of the wind-wave regime based on the inverse wave age [22]. The backscattering coefficient results have been verified by comparing with the data collected in Lingshan Island during the period of October and November 2014 at low grazing angles and the Ku-band measurements at moderate grazing angles. Their results indicate perfect agreement (within about 2 dB) with field data. Deep learning is an excellent method that can be used not only for classification but also for inversion and fitting of non-linear functions. In order to

simulate the application of actual radar detection and inversion technology, the inversion of significant wave height from actual sea surface backscattering coefficients training data sets has been performed by using deep learning technology. The accuracy of 99.01% has been achieved under the condition of three hidden layers and iterating 100 times. The root mean square errors of the test data sets are less than 0.10, which indicates that deep learning is available in the inversion of significant wave height.

Overfitting is one of the most challenging problems in deep neural networks with a large number of trainable parameters. To prevent networks from overfitting, the dropout method, which is a strong regularization technique, has been widely used in fully-connected neural networks. In several state-of-the-art convolutional neural network architectures for object classification, however, dropout was partially or not even applied since its accuracy gain was relatively insignificant in most cases. Also, the batch normalization technique reduced the need for the dropout method because of its regularization effect. In the contribution by Lee et al. "Revisiting spatial dropout for regularizing convolutional neural networks", authors show that conventional element-wise dropout can be ineffective for convolutional layers [6]. They found that dropout between channels in the CNNs can be functionally similar to dropout in the FCNNs, and spatial dropout can be an effective way to take advantage of the dropout technique for regularizing. To prove their points, they conducted several experiments using the CIFAR-10 and CIFAR-100 databases. For comparison, they only replaced the dropout layers with spatial dropout layers and kept all other hyperparameters and methods intact. DenseNet-BC with spatial dropout showed promising results (3.32% error rates with CIFAR-10, 3.0 M parameters) compared to other existing competitive methods.

Information of red blood cell (RBC) morphology, obtained by analyzing RBC images, is regularly requested by veterinarians to diagnose anemic dogs. Machine learning techniques have been exploited to speed up the image classification. Recently, many researchers used deep learning techniques for classification; however, a large quantity of labelled data is necessary to extract performance with them. A lack of annotated data, due to time and costs for the pathologist and their limited numbers, has become a difficulty. This limits the amount of annotated data and leads to a large number of unannotated data, preventing traditional deep learning algorithms from being effective. The authors of "Semi-supervised learning with deep convolutional generative adversarial networks for canine red blood cells morphology classification" show that a semi-supervised learning method, using the generative adversarial networks (GANs) for canine RBC morphology classification, can solve the lack of labelled data, when they want to train a deep learning classifier [11]. Their semi-supervised GAN can use both labelled and un-labelled data and showed that they can achieve the same level of performance as a traditional convolutional neural network, with a smaller number of labelled images. Furthermore, they showed that augmenting the limited numbers of labelled images enhanced the overall performance. A key benefit of their method is reduced pathologist cost and time to annotate cell images for developing a deep learning classifier.

With the process of economic globalization and political multi-polarization accelerating, it is especially important to predict policy change in the United States. While current research has not taken advantage of the rapid advancement in natural language processing and the relationship between news media and policy change, authors of "Learning to predict U.S. policy change using New York Times corpus with pre-trained language model" propose a BERT-based model to predict policy change in the United States, using news published by the New York Times [23]. Specifically, they propose a large-scale news corpus from the New York Times covering the period from 2006 to 2018. Then they use the corpus to fine-tune the

pre-trained BERT language model to determine whether the news is on the front page, which corresponds to the policy priority. They propose a BERT-based policy change index (BPCI) for the United States to predict the policy change in the future short period of time. Their simulation results in the New York Times corpus demonstrate the validity of the proposed method.

2.2 Performance improvements

Wireless sensor network (WSN) is composed of numerous tiny smart sensor nodes integrated with internet of things (IoT) and plays a crucial role in many applications. The IoT connects physical devices to form a network which consist of software, sensor for exchange of information. Clustering is the most common technique for efficient energy utilization in WSN. Sensor nodes when they have data, forward it to cluster head (CH) and CH transfers the received data from the sensor nodes to the sink. When the sink nodes are far away from CH, long-haul transmission consumes higher power. In the contribution by Seema et al. "Efficient data transfer in clustered IoT network with cooperative member nodes", authors propose efficient data transfer mechanism for clustered IoT network through the cooperation of member nodes [16]. First, they use greedy algorithm to select cooperative sensor nodes to act as relay for long distance transmission. Then, to encourage sensor nodes in data forwarding, cluster head uses priority buffers to prioritize assisting sensor nodes data. Their simulation results show that the proposed approach conserves energy and increases the life-time of clustered IoT network.

Ensemble clustering techniques have improved in recent years, offering better average performance between domains and data sets. Benefits range from finding novelty clustering which are unattainable by any single clustering algorithm to providing clustering stability, such that the quality is little affected by noise, outliers or sampling variations. The main clustering ensemble strategies are: to combine results of different clustering algorithms; to produce different results by resampling the data, such as in bagging and boosting techniques; and to execute a given algorithm multiple times with different parameters or initialization. Often ensemble techniques are developed for supervised settings and later adapted to the unsupervised setting. Recently, Blaser and Fryzlewicz proposed an ensemble technique to classification based on resampling and transforming input data. Specifically, they employed random rotations to improve significantly random forests performance. In the contribution by Rodrigues et al. "An empirical evaluation of random transformations applied to ensemble clustering", authors empirically studied the effects of random transformations based in rotation matrices, Mahalanobis distance and density proximity to improve ensemble clustering [15]. Their experiments considered 12 data sets and 25 variations of random transformations, given a total of 5580 data sets applied to 8 algorithms and evaluated by 4 clustering measures. Statistical tests identified 17 random transformations that are viable to be applied to ensembles and standard clustering algorithms, which had positive effects on cluster quality. In their results, the best performing transforms were Mahalanobis-based transformations.

In the contribution by Shan et al. "A parallel sliding-window belief propagation algorithm for Q-ary LDPC codes accelerated by GPU", a parallel sliding-window belief propagation algorithm to decode Q-ary low-density-parity-codes is proposed [17]. This algorithm is accelerated by taking advantage of high parallel features of GPU, and applied to video compression under distributed video coding framework. The experiment results show that their parallel algorithm achieves 2.3× to 30.3× speedup ratio under 256 to 2048 codeword

length and 69.21× to 78.31× speedup ratio under 16,384 codeword length than sequential algorithm.

Activity recognition in smart environments is essential for ensuring the wellbeing of older residents. By tracking activities of daily living (ADLs), a person's health status can be monitored over time. Nonetheless, accurate activity classification must overcome the fact that each person performs ADLs in different ways and in homes with different layouts. One possible solution is to obtain large amounts of data to train a supervised classifier. Data collection in real environments, however, is very expensive and cannot contain every possible variation of how different ADLs are performed. A more cost-effective solution is to generate a variety of simulated scenarios and synthesize large amounts of data. Nonetheless, simulated data can be considerably different from real data. The contribution by Ortiz-Barrios et al. "Complementing real datasets with simulated data: a regression-based approach" proposes the use of regression models to better approximate real observations based on simulated data [10]. To achieve this, ADL data from a smart home were first compared with equivalent ADLs performed in a simulator. Such comparison was undertaken considering the number of events per activity, number of events per type of sensor per activity, and activity duration. Then, different regression models were assessed for calculating real data based on simulated data. The results evidenced that simulated data can be transformed with a prediction accuracy R2 = 97.03%.

The computer-aided analysis of indirect-immunofluorescence (IIF) images is important for the differential diagnosis of several autoimmune diseases. A fully automatic approach consists in segmentation of individual cells in IIF images and subsequently its classification into various pattern types. The contribution by Ul Islam et al. "Towards the automatic segmentation of HEp-2 cells in indirect immunofluorescence images using an efficient filtering based approach" explores the segmentation of HEp2 cells in IIF images through the use of a filtering based approach [20]. Their algorithm is based on a local convergence filter named as sliding band filter (SBF). They propose a modified SBF that is capable of handling the low contrast, noise and illumination variations peculiar to IIF images. In addition, they follow a simple algorithmic pipeline and achieve better accuracy as compared to several state-of-the-art segmentation algorithms on standard HEp2 image dataset.

Traditional iris recognition methods, which are still preferred against artificial intelligence (AI) approaches in practical applications, are often required to capture high-grade iris samples by an iris scanner for accurate subsequent processing. To reduce the system cost for mass deployment of iris recognition, pricey scan devices can be replaced by the average quality cameras combined with additional processing algorithm. In the contribution by Lin et al. "Fast iris localization using Haar-like features and AdaBoost algorithm", authors propose a Haar-like-feature-based iris localization method to quickly detect the location of human iris in the images captured by low-cost cameras for the ease of post-processing stages [8]. The AdaBoost algorithm was chosen as a learning method for training a cascade classifier using Haar-like features, which was then utilized to detect the iris position. The experimental results have shown acceptable accuracy and processing speed for this novel cascade classifier. This achievement stimulates them to implement this novel capturing device in their iris recognition.

Access control is used to prevent data from access of unauthorized users. Over the years, several access control models have been proposed to meet requirements of various applications and domains. Role-based access control model is one such model which enforces security based on the roles. However, role-based access control model is static in nature and does not provide the dynamism of collaboration required in the multi-domain environment. The

contribution by Aslam et al. "OBAC: towards agent-based identification and classification of roles, objects, permissions (ROP) in distributed environment" presents an ontology-based access control (OBAC) model, which provides a solution by using an ontology-based approach [1]. In OBAC model, agents are used for the identification and classification of roles, objects and permissions (ROP) in distributed environment. The proposed method exploits the ontology-based approach, where agent learns and adapts changes to identify roles, objects and permissions from a given dataset and classifies them into ontology according to rules and policies. The proposed ontology also provides extensibility and reusability. Moreover, they simulated their technique on datasets of two different domains. The first dataset is related to the university environment and the second one is about hospital domain. The promising experimental results indicate the effectiveness of proposed approach.

A single sensor camera uses color filter array (CFA) to capture single color information at each pixel. Thus, to estimate the missing color samples and then to reconstruct an original image is known as CFA interpolation or demosaicking. Despite remarkable improvements made in the last decade, a fundamental issue remains to be addressed, i.e., how to assure the visual quality of an image in the presence of noise. Hence, the CFA images without denoising leads to the demosaicking artifacts that eventually reduce the image quality. Therefore, based on the aforementioned constraints, the contribution by Din et al. "Lightweight deep dense demosaicking and denoising using convolutional neural networks" presents a novel approach for demosaicking and denoising based on the convolutional neural network [3]. The proposed technique is using convolutional neural networks, and consists of four phases. In the first phase, the image is organized. In the second phase, the demosaicking is performed using the deep dense convolutional neural network, which gives them the demosaicked image. In the third phase, denoising is performed and passes this image to the final phase. Finally, in the fourth phase, the image passes to the final post-processing phase producing a higher quality image. To test the feasibility of the proposed scheme, Python language is used. The proposed scheme outperforms several existing methods in terms of throughput delay, latency, accuracy.

2.3 Applications

State-of-the-art methods for handwriting recognition are based on long short term memory (LSTM) recurrent neural networks (RNN), which now provide very impressive character recognition performance. Character recognition is generally coupled with a lexicon-driven decoding process which integrates dictionaries. Unfortunately, these dictionaries are limited to hundreds of thousands of words for the best systems, which prevents them from having good language coverage, and therefore limits global recognition performance. In the contribution by Stuner et al. "Handwriting recognition using Cohort of LSTM and lexicon verification with extremely large lexicon", authors propose an alternative to the lexicon-driven decoding process based on a lexicon verification process and a new method to obtain hundreds of complementary LSTM RNN that are extracted from a single training, called cohort, coupled in different combination systems [19]. Their first combination is a cascade made of a large number of complementary LSTM RNN for isolated handwritten word recognition. The proposed cascade achieves new state-of-the art performance on the Rimes and IAM datasets. The second contribution extends the idea of cohort and lexicon verification in a ROVER combination for handwriting line recognition and achieves state-of-the-art results on the Rimes dataset. Dealing with gigantic lexicon of 3 million words, the method also demonstrates interesting performance with a fast decision stage.

Integrating and analyzing a large amount of data extracted from different sources can be considered a key asset for businesses, organizations, research institutions that also deal with the Cultural Heritage domain. In the last decade, internet of things (IoT) technologies and the massive use of mobile devices contributed to generate an enormous flow of multimedia data, whose collection, analysis and interpretation allows for real-time analysis related to the behaviors, preferences and opinions of users. In the contribution by Piccialli et al. "Unsupervised learning on multimedia data: a Cultural Heritage case study", authors present and discuss an unsupervised learning approach on multimedia features of a dataset coming from an IoT framework [12]. The main research objective of this work is to assess how the collection of behavioral IoT data coming from the Cultural Heritage domain can be opportunely exploited by means of unsupervised learning techniques in order to produce useful insights for the stakeholders, especially considering the multimedia features of such data. The presented experimental results, executed in a real case study, assess how the Cultural Heritage domain, and the related stakeholders, can benefit from these kinds of services and applications.

In the contribution by Chen et al. "A multiscale dilated residual network for image denoising", a more effective Gaussian denoiser is designed to enhance the resulting image quality [7]. The authors propose a novel image denoising method using a multiscale dilated residual network, named MDRNet. The proposed method is based on two main strategies. First, they adopt dilated convolutions in their network to enlarge the receptive field while requiring fewer parameters. The hybrid dilation rate pattern (HDP) is implemented such that each pixel in the pattern contributes similarly to the receptive field, allowing their network to learn the image details equally. Second, they employ a contextualized structure to take advantage of the low-level features which are mainly concentrated in the first two layers. Their method achieves competitive denoising performance and requires fewer parameters compared to existing denoising methods that use convolutional networks. Through comprehensive experiments, they show that the denoising performance of their method is competitive with the state-of-the-art methods in terms of both quantitative and qualitative evaluation.

To solve the complex computation, unstable network and slow learning speed problems of a generative adversarial network for image super-resolution (SRGAN), Zeng et al. proposed a single image super-resolution reconstruction model called the Res_WGAN based on ResNeXt in their contribution "Single image super-resolution reconstruction based on the ResNeXt network" [9]. The generator is constructed by the ResNeXt network, which reduced the computational complexity of the model generator to 1/8 that of the SRGAN. The discriminator was constructed by the Wasserstein GAN (WGAN), which solved the SRGAN's instability. By removing the normalization operation in the residual network, the learning rate is improved. The experimental results from the Res_WGAN demonstrated that the proposed model achieved better performance in the subjective and objective evaluations using four public data sets compared with other state-of-the-art models.

Action temporal detection is a derivative task of action recognition which needs researchers to predict temporal intervals and specific categories in untrimmed videos. Aiming at the problem of too many proposed segments and insufficient filtering effect in multi-stage networks, Song et al. propose an action temporal detection method using confidence curve analysis to generate proposal segments in their contribution "Action temporal detection method based on confidence curve analysis" [18]. Fixed step window sliding is adopted to generate candidate segments in a video, and they adjust a training mode in segment network. The proposal segments are generated by analyzing the confidence curve of candidate segments, finally proposal segments are input into localization network to classify and adjust

confidence level. Their extensive experiments performed on THUMOS2014 benchmark show that the proposed method performs significantly better than the original multi-stage convolutional network that mAP increase from 19.0% to 26.4% with 252% accelerating.

Matching video clips of people across non-overlapping surveillance cameras (video-based person re-identification) is of significant importance in many real-world applications. In the contribution by Cheng et al. "Local and global aligned spatiotemporal attention network for video-based person re-identification", authors address the video-based person re-identification by developing a local and global aligned spatiotemporal attention (LGASA) network [2]. Their LGASA network consists of five cascaded modules, including 3D convolutional layers, residual block, spatial transformer network (STN), multi-stream recurrent network and multiple-attention module. Specifically, the 3D convolutional layers are used to capture local short-term fast-varying motion information encoded in multiple adjacent original frames. The residual block is used to extract mid-level feature maps. STN is applied to align the mid-level feature maps. The multi-stream recurrent network is designed to exploit the useful local and global long-term temporal dependency from the aligned mid-level feature maps. The multiple-attention module is designed to aggregate feature vectors of the same body part (or global) from different frames within each video into a single vector according to their importance. Their experimental results on three video pedestrian datasets verify the effectiveness of the proposed local and global aligned spatiotemporal attention network.

The wood species classification is an essential field of investigation that can help to combat illegal logging, then providing the timber certification and allowing the application of correct timber taxing. Today, the wood classification relies on highly qualified professionals that analyze texture patterns on timber sections. However, these professionals are scarce, costly, and subject to failure. Therefore, the automation of this task using computational methods is promising. Deep learning has proven to be the ultimate technique in computer vision tasks, but it has not been much exploited to perform timber classification due to the difficulty of building large databases to train such networks. In the contribution by Geus et al. "An analysis of timber sections and deep learning for wood species classification", authors introduced the biggest data set of wood timber microscope images to date, with 281 species, having three types of timber sections: transverse, radial, and tangential [4]. They investigated the use of transfer learning from pre-trained deep neural networks for wood species classification and compared their results with a state-of-the-art pre-designed feature method. Their simulation results show that traverse section images using a densely connected network achieved 98.7% of correct classification against 85.9% of standard pre-design features.

Acquiring all-in-focus images is significant in the multi-media era. Limited by the depth-of-field of the optical lens, it is hard to acquire an image in which all targets are clear. One possible solution is to merge the information of a few complementary images in the same scene. In the contribution by Wen et al. "Multifocus image fusion using convolutional neural network", authors employ a two-channel convolutional network to derive the clarity map of source images [21]. Then, the clarity map is smoothed by using morphological filtering. Finally, the fusion image is constructed via merging the clear parts of source images. Their experimental results prove that their approach has a better performance on both visual quality and quantitative evaluations than many previous fusion approaches.

Numerous patients die every year due to the leading causes of deaths all over the world and burn injuries are one of them. Burn injury cases are most viewed in low and middle-income countries (LMIC). Researchers show great interest to classify the burn into different depths through digital means. In Pakistan, at provisional level, it is a really significant issue to

categorize the burn and its depths due to the non-availability of expert doctors and surgeons; hence the decision for the correct first treatment cannot be made, so this may cause a serious issue later on. The main objective of the contribution by Khan et al. "Computer-aided diagnosis for burnt skin images using deep convolutional neural network" is to segment the burn wounds and classify burn depths into 1st, 2nd and 3rd degrees respectively [5]. A real-time dataset of burns patients has been collected from the burn unit of Allied Hospital Faisalabad, Pakistan. The dataset used for this research task contains 450 images of all the three levels of burn depths. Segmentation of the burnt area was done by the use of Otsu's method of thresholding and feature vector was obtained through the use of statistical methods. They have used the deep convolutional neural network (DCNN) to estimate the burn depths. The network was trained by 65% of the images and the remaining 35% images were used for testing the accuracy of the classifier. The maximum average accuracy obtained by using the DCNN classifier is reported around 79.4% and these results are the best if they compare them with previous results. From the obtained results of this research work, non-expert doctors will be able to apply the correct first treatment for the quality evaluation of burn depths.

3 Conclusion

The articles presented in this special issue provide insights in fields related to multimedia analysis using artificial intelligence in deep learning algorithms, including models, performance evaluation and improvements, and application developments. We wish the readers can benefit from insights of these papers, and contribute to these rapidly growing areas. We also hope that this special issue would shed light on major developments in the area of *Multimedia Tools and Applications* and attract attention by the scientific community to pursue further investigations leading to the rapid implementation of these technologies.

Acknowledgements We would like to express our appreciation to all the authors for their informative contributions and the reviewers for their support and constructive critiques in making this special issue possible. Finally, we would like to express our sincere gratitude to Professor Borko Furht, the Editor-in-Chief, for providing us with this unique opportunity to present our works in *Springer Multimedia Tools and Applications*.

References

1. Aslam, S., Ahmed, M., Ahmed, I., Khan A., Ahmad A., Imran M., Anjum A., Hussain S. (2020). OBAC: towards agent-based identification and classification of roles, objects, permissions (ROP) in distributed environment. Multimed Tools Appl https://doi.org/10.1007/s11042-020-08764-2
2. Cheng, L., Jing, X., Zhu, X. et al. (2020). Local and global aligned spatiotemporal attention network for video-based person re-identification. Multimed Tools Appl https://doi.org/10.1007/s11042-020-08765-1
3. Din, S., Paul, A., and Ahmed, A. (2020). Lightweight deep dense demosaicking and denoising using convolutional neural networks. Multimed Tools Appl
4. Geus, A. R., da Silva, S. F., Gontijo, A. B. et al (2020). An analysis of timber sections and deep learning for wood species classification. Multimed Tools Appl
5. Khan, F. A., Butt, A. U. R., Asif, M., Ahmad W., Nawaz M., Jamjoom M., Alabdulkreem E. (2020). Computer-aided diagnosis for burnt skin images using deep convolutional neural network. Multimed Tools Appl https://doi.org/10.1007/s11042-020-08768-y
6. Lee, S., Lee, C. (2020). Revisiting spatial dropout for regularizing convolutional neural networks. Multimed Tools Appl

7. Li, D., Chen, H., Jin, G. et al. (2020). A multiscale dilated residual network for image denoising. Multimed Tools Appl
8. Lin, Y., Hsieh, T., Huang, J. et al. (2020). Fast Iris localization using Haar-like features and AdaBoost algorithm. Multimed Tools Appl https://doi.org/10.1007/s11042-020-08907-5
9. Nan, F., Zeng, Q., Xing, Y., Qian Y. (2020). Single image super-resolution reconstruction based on the ResNeXt network. Multimed Tools Appl https://doi.org/10.1007/s11042-020-09053-8
10. Ortiz-Barrios, M. A., Lundström, J., Synnott, J., Järpe E., Sant'Anna A. (2020). Complementing real datasets with simulated data: a regression-based approach. Multimed Tools Appl https://doi.org/10.1007/s11042-019-08368-5
11. Pasupa, K., Tungjitnob, S. and Vatathanavaro, S. (2020). Semi-supervised learning with deep convolutional generative adversarial networks for canine red blood cells morphology classification. Multimed Tools Appl https://doi.org/10.1007/s11042-020-08767-z
12. Piccialli, F., Casolla, G., Cuomo, S., Giampaolo F., Prezioso E., di Cola V. S. (2020). Unsupervised learning on multimedia data: a cultural heritage case study. Multimed Tools Appl https://doi.org/10.1007/s11042-020-08781-1
13. Qu, T., Li, H., Wu, Z., Shang Q., Wu J., Kong W. (2020). Scattering of aerosol by a high-order Bessel vortex beam for multimedia information transmission in atmosphere. Multimed Tools Appl https://doi.org/10.1007/s11042-020-08773-1
14. Rashwan, H. A., Garcia, M. A, Abdulwahab, S., and Puig, D. (2020) Action representation and recognition through temporal co-occurrence of flow fields and convolutional neural networks. *Multimed Tools Appl*
15. Rodrigues, G. D., Albertini, M. K., and Yang, X. (2020). An empirical evaluation of random transformations applied to ensemble clustering. Multimed Tools Appl
16. Seema, B., Yao, N., Carie, A., Shah S.B.H. (2020). Efficient data transfer in clustered IoT network with cooperative member nodes. Multimed Tools Appl https://doi.org/10.1007/s11042-020-08775-z
17. Shan, B., Chen, S. and Fang, Y. (2020) A parallel sliding-window belief propagation algorithm for Q-ary LDPC codes accelerated by GPU. Multimed Tools Appl https://doi.org/10.1007/s11042-020-08738-4
18. Song, H., Tian, L. and Li, C. (2020). Action temporal detection method based on confidence curve analysis. Multimed Tools Appl https://doi.org/10.1007/s11042-020-08771-3
19. Stuner, B., Chatelain, C. and Paquet, T. (2020). Handwriting recognition using cohort of LSTM and lexicon verification with extremely large lexicon. Multimed Tools Appl
20. Ul Islam, I., Ullah, K., Afaq, M., Iqbal J., Ali A. (2020). Towards the automatic segmentation of HEp-2 cells in indirect immunofluorescence images using an efficient filtering based approach. Multimed Tools Appl https://doi.org/10.1007/s11042-020-08651-w
21. Wen, Y., Yang, X., Celik, T., Sushkova O., Albertini M. K. (2020). Multifocus image fusion using convolutional neural network. Multimed Tools Appl https://doi.org/10.1007/s11042-020-08945-z
22. Wu, T., Cao, Y., Wu, Z. et al. (2019) Deep learning for inversion of significant wave height based on actual sea surface backscattering coefficient model. Multimed Tools Appl https://doi.org/10.1007/s11042-019-07967-6
23. Zhang, G., Wu, J., Tan, M. et al. (2020). Learning to predict U.S. policy change using New York Times corpus with pre-trained language model. *Multimed Tools Appl* https://doi.org/10.1007/s11042-020-08946-y