Osagie, Efosa ORCID logoORCID: https://orcid.org/0009-0004-3462-7175, Ji, Wei and Helian, Na (2024) Medical Image Character Recognition Using Attention-Based Siamese Networks for Visually Similar Characters with Low Resolution. In: Daimi, Kevin and Al Sadoon, Abeer, (eds.) Proceedings of the Third International Conference on Innovations in Computing Research (ICR'24). Springer Nature, pp. 110-119

Downloaded from: https://ray.yorksj.ac.uk/id/eprint/12900/

The version presented here may differ from the published version or version of record. If you intend to cite from the work you are advised to consult the publisher's version: https://doi.org/10.1007/978-3-031-65522-7_10

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. Institutional Repositories Policy Statement

RaY

Research at the University of York St John

For more information please contact RaY at ray@yorksj.ac.uk

Medical Image Character Recognition using Attention-based Semantic Similarity Learning for Visually Similar Characters with Low Resolution and Limited Samples.

Efosa Osagie
Department of Computer Science
University of Hertfordshire,
College Ln, Hatfield AL10 9AB,
Hatfield, Hertfordshire, United Kingdom
e.osagie @ herts.ac.uk

Wei Ji
Department of Computer Science
University of Hertfordshire,
College Ln, Hatfield AL10 9AB,
Hatfield, Hertfordshire, United Kingdom
w.1.ji @ herts.ac.uk

Na Helian
Department of Computer Science
University of Hertfordshire,
College Ln, Hatfield AL10 9AB,
Hatfield, Hertfordshire, United Kingdom
n.helian @ herts.ac.uk

Abstract* — The emergence of optical character recognition (OCR) has been adopted in many domains in automating various tasks. Still, recognising visually similar characters (VSCs) remains a challenging problem in the general OCR domain. Applying conventional class probability predictions by deep learning techniques may be difficult due to the limited datasets in some domains, such as medical imaging modalities. VSCs recognition becomes more complicated with the problem of low resolution and background interference in the image. With advancements in computing power and numerical methods, techniques such as the few-shot method have been proposed to tackle the limited sample problems in training deep learning models. Still, very little work has been done regarding designing an OCR solution to deal with tiny textual data on low-resolution images with background interference while training on limited samples per class. In this study, we propose an Attention-based Siamese Network to accurately recognise VSCs by efficiently learning the semantic similarities between the extracted embeddings from the input images. The learned similarities and attention-focused feature extraction layer enable the proposed model to discriminate between different character classes efficiently, with only limited samples available. Bayesian optimisation is used to determine optimal network parameters. We further aim to set a benchmark for the performance of the Siamese network in OCR in medical image character recognition in terms of reduced parameter size and accuracy at a determined sample size.

Keywords— Medical Image Character Recognition, Siamese network, Burned-in Textual data recognition, Channel-wise attention, Similarity learning.

I. INTRODUCTION

OCR is an important computer vision application in converting text on images into easily accessible forms. It is widely used in numerous applications, such as industrial, medical, and educational institutions, mainly in automating data entry and other database-driven processes. Numerous cases, such as medical imaging modalities, may share similar textual data. These similarities may include low resolution, character distortion, text overlapping and background interference. These can be caused either by the mode of acquisition in the case of the image or the mode of storage in the case of historical document images. The textual data are usually burned-in on the medical imaging modalities but poorly printed or handwritten on historical document images.

Due to distortion, poor image quality, background noise and low resolution, certain characters may appear visually similar in their structure and appearance [1]. These can be termed visually similar characters (VSC). Recognising these characters may become more challenging due to the nature of the images by conventional OCRs.

Even with the rapid growth in the application of deep learning techniques in the field of OCR for pattern recognition and character classification, the problem of recognising VSC remains unsolved, resulting in various research to find a solution [2]. This is because conventional classification using deep learning techniques relies on a large and equally distributed dataset to achieve good performance. However, collecting a large dataset in certain domains, such as burned-in textual data on medical images, requires a lot of resources, such as privacy permissions [3], and it is time-consuming.

Hence, it become important to develop an OCR solution that can learn highly discriminative features from low-resolution images with background interference to classify VSC with only a limited dataset available. This will enable further adoption of OCR in low-resource domains where datasets are highly limited. This study proposes a few-shot learning-based method to deal with the limited dataset while leveraging the advantage of the Siamese network architecture and channel attention mechanism. The Siamese neural network (SNN) is a major component of few-shot learning methods, as

seen in notable and recent [4,5,6,7]. The SNN can learn semantic similarities between classes of images by minimising the metric distance between the same class and maximising the metric distance between different classes. However, using the concept to define a fine-tuned classification decision boundary for VSC on these complex images is a major challenge when the issue of tiny text, low resolution, and background interference must be considered along with the limited samples available. This is because the complex nature of the character images may affect the extracted feature embeddings to be compared during the learning process. Hence, combining a channel-wise attention mechanism will enable the SNN to focus on the image's critical discriminative region by exploiting the features' inter-channel relationship. Since each channel of a feature map can be considered a feature detector, an SNN with a channel-wise attention mechanism focuses on the meaningful aspect of an input image that sets it apart for effective representation learning. Human perception and intelligence often seek to combine additional clues from associated characters around these VSCs to accurately interpret the entire word and overlook any structural similarities. Examples of this are “STOP” and “5TOP”. Even though advanced OCR solutions have incorporated post-correction features, these remain limited in low-resource domains where textual data are not entirely defined, and there is no available all-inclusive vocabulary. This makes it necessary to accurately recognise the individual characters independently while paying attention to VSCs. Little research has been done on recognising VSCs in low-resolution images under the limited sample constraint.

The main contributions of this paper are.

- We propose a Siamese neural network to learn semantic similarities between extracted embeddings of image pairs in metric space in the presence of a limited dataset and low image resolution with background interference. The resulting model can discriminate between visually similar characters by learning a fine-tuned decision boundary.

- We propose a Channel attention mechanism combined with a Siamese neural network to learn meaningful parts of an input image that discriminate when compared to a visually similar image. The problem of recognising medical image character was presented in this study.

- Provide a benchmark on using similarity learning in Medical Image Character Recognition (MICR). After an extensive literature review in the past 10 years, no previous work has been done regarding MICR and SNN with and without channel attention mechanisms. We aim to justify this, by comparing our architecture’s performance with that of related past works on OCR with Siamese network, based on a medical image character dataset.

Related work is presented in section II. The proposed method is provided in section III. Section IV presents the experiments and analysis. Section V provides our conclusion and future works regarding this study.

II. RELATED WORKS

In a classification problem with limited training data and similarity problems in the dataset samples, using an SNN has become a popular concept. This is also known as metric learning, where a model learns the distance of a few-shot class representation in metric space, and the difference in the distance between these samples can be used for classification tasks. In this section, the study will review related works on applying SNN in the general field of OCR because extensive reviews have shown that Siamese networks have not been applied in the MICR.

SNN and K-Nearest Neighbour classification algorithms to classify similar text were done by [8]. An evaluation was done on machine-printed and handwritten text, and they reported an accuracy of 99.5%. The authors [8] used a large dataset containing over 188,526-character images. A combined loss function was used, which caused difficulty during training, and the dataset was of high quality. Good accuracy of 97%, 79% and 89% were reported on three datasets, but this method will not be efficient in situations where the dataset is much more limited.

With more focus on leveraging the advantages of the feature extraction capabilities on the radical-level composition of characters, [9] proposed a radical aggregation network for few-shot recognition of handwritten character recognition. Their network used a convolutional block, ResNet, and an attention module. It performed an efficient radical feature selection using a radical mapping encoder to map the input into a radical representation sequence, where each representation is a high-dimensional feature vector. A distance metric is calculated between these radical representations and radical prototypes, and a character analysis decoder does transcription to a character. A 96.97% accuracy on the CASIA-HWDB character dataset was obtained using only 6,391 training samples. Although the accuracy was good, the representation mapping of distorted characters and VSCs was poor based on comparison with human performance, meaning that the radical representation learned by the network is still

ineffective. It was complex and highly resource-demanding, more than twice the baseline CNN-based classifier model used in the study. Another study that leverages the use of a prototype was done by [10] to compute an N-dimensional representation of each class through an embedding function with learnable parameters. Each prototype is the mean vector of the embedded support points belonging to its class. Their proposed method obtained an accuracy of 49.42% on only 1623 samples of handwritten characters with 50 classes and 68.20% on training with 5 samples per class. However, the study did not propose any defined architecture; there was no consideration for low resolution and background interference in these characters. According to the authors [10], episodic training was done to simplify the training. However, this depends on representing each class by the mean of its samples in a representation space learned by a neural network. This technique becomes inefficient when characters are blurred, distorted, or degraded.

III. PROPOSED METHOD

Model architecture

The study proposed a Siamese network to learn semantic similarity between limited samples of visually similar characters, which are low-resolution with background interference. The network is two CNNs that are joined at the end. Before being joined, each CNN has 5 layers (3 convolutional and 2 dense layers). Then, a Euclidean distance layer merges both CNNs with a single output. The study used the Bayesian optimisation based on a Tree-Structured Parzen Estimator to find the optimal combination of hyper-parameters, which includes the number of layers, units, dropout percentage and other relevant parameters. The study will not focus on the optimisation technique but on its application to achieve CNN's architecture. The aim was to find the maximum value of the objective function, which is the similarity score between a query image and a set of support images. See CNNs' Dimensions below in Table 1:

Table 1: CNN's Layerwise Summary

Nth Layer	Input size	Filters	Kernel	Max pooling	Activation function
1	48 x 48	16	3x3	Stride of 2	ReLU
2	16 filters of 3x3	32	3x3	Stride of 2	ReLU
3	32 filters of 3x3	64	3x3	Stride of 2	ReLU
4	128 dense layers	None	None	None	ReLU
5	254 dense layers	None	None	None	ReLU

There are twin CNNs designed according to Table 1.

- All CNN layers, except the dense layers, are defined with a fixed stride of 2 and a padding value of 'valid' to enable the small kernel to traverse only within the image.
- All parameters that were not mentioned used the default Tensorflow 2.2.2 Keras values.
- After the last dense layer, a lambda layer is created to merge both CNNs, thus creating the Siamese network. The activation function of this merging layer is a Sigmoid function due to the binary classification problem to be learned.
- Total params: 187406, Trainable params: 187404 and non-trainable params: 2. The resulting network is smaller in terms of the number of layers and trainable parameters when compared with the notable past works based on a classifier approach in the general OCR domain.

A visual representation of the Siamese is shown in Fig 1 below after Bayesian optimisation of over 600 iterations. This model will be referred to as SIAM-MICR for ease of reference.

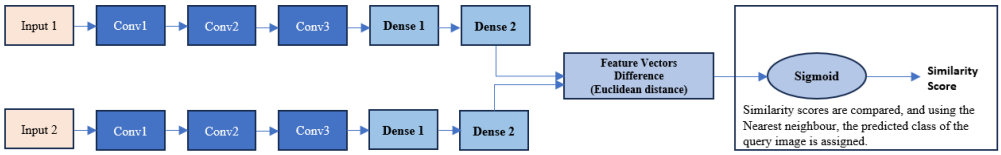

Fig 1: SIAM-MICR

The CNNs were designed as a pair, and training was achieved using the two parallel CNNs with shared weights, trained on matched and unmatched character image pairs. Each image is fed through one

branch of the CNN, generating a d-dimensional embedding for the image. These embeddings optimise a loss function (Contrastive Loss) rather than the images themselves. Contrastive Loss aims to predict relative distances between model inputs when projected onto a hyperspace. The embeddings between the pairs are used to calculate the Euclidean distance to be used as a measure of similarity. In the Siamese architecture, the Lambda layer computes the Euclidean distances between the outputs of the two parallel CNNs. The major challenge was getting to the right depth of the CNN architecture to get the optimal semantic similarity learning with the limited samples of the training dataset and the low resolution of the character image patches. This was solved using Bayesian optimisation techniques for the optimal depth of the CNNs.

Model + Attention Mechanism

The study proposed using a channel attention mechanism as motivated by notable works by [11, 12] to improve the previously designed SIAM-MICR further. The channel attention mechanism in each CNN generates channel-wise responses by using global average pooling to aggregate spatial information [13]. Given the aggregated features obtained from the global average pooling (GAP), a fast 1D convolution of kernel size, k, is performed to generate the output channel weights. k is the kernel size of the 1d convolutional layer. It represents the coverage of local cross-channel interactions, that is, the number of pixel neighbours taking part in the output of one channel map. The use of a 1D convolution is to avoid dimension reduction and allow efficient learning across the channel for significant and discriminate features of the input images for the Siamese network. Much investigation via experiments was carried out to determine the optimal position for the attention module. The optimal for the channel-wise attention module was only after each CNN's first or third convolutional layer. This setup improved the network's ability to focus on learning weights for more primitive and discriminative features, such as curves, lines, and edges, which may appear similar across the character classes. The channel attention mechanism used in the SIAM-MICR is motivated by a notable work by [11]. The input tensor to the module is the output of a convolutional layer and has a 4-D shape of B, C, H, and W, where B is the batch size, C is the number of channels, and H and W are the dimensions of each feature map. The output of the attention module is also a 4-D tensor of the same shape. Fig 2 shows the SIAM-MICR with the attention module.

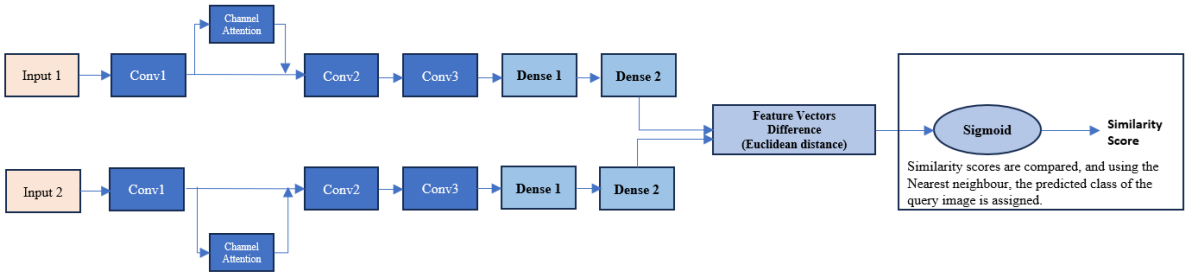

Fig 2: SIAM-MICR + Attention

IV. EXPERIMENT, RESULT AND ANALYSIS

Dataset Description and Training Strategy

The medical image dataset used to test our proposed MICR model is the Medpix medical image collection. Medpix medical image dataset is open-source and contains 60,613 image collections of ultrasounds, X-rays, MRIs, and CTs. The dataset contains burned-in textual data representing various medical interpretations of the images acquired using standard medical imaging practices. This study manually created a character dataset from the dataset for 62-character classes ['A-Z', 'a-z' and '0-9'] with a dimension of [28,28,3]. The Python Image Pillow library was used to check the resolution of the character images patch, and the result gave a tuple of (96, 96), which is 96 dpi. This means the sample images are of a low resolution of 96 dpi.

Training

To train and evaluate the Siamese network, the 62-class classification was changed into a binary classification problem created by making a new dataset of pairs, where match images are given a label of 0.0 and unmatched images are given a label of 1.0. RMSprop optimiser was used, as supported by studies on adaptative-based optimisers by [14,15], because of its advantage in fast convergence speed over a few iterations, which is required for complex models such as Siamese models where training time may be a critical concern [16]. The

training pairs were formed randomly and were balanced. The dataset split was 80% and 20% for training and testing, respectively. Samples of the training data are shown below in Fig 3.

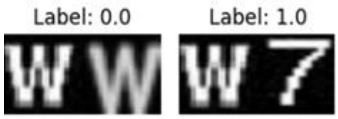

Fig 3: Pairing of images for training.

Result

This study investigated the accuracy of the Siamese network base model with and without channel attention on a limited sample size of 25 samples and 20 samples. The study could not take more than 25 sample sizes due to the sample availability in the dataset. This is presented in Table 2

Table 2: Comparison of Model's Accuracy with/without channel attention at 100 epochs.

Dataset size	**Accuracy-SIAM-MICR (%)**	**Accuracy - SIAM-MICR + Attention (%)**
25 samples per class	87.73±0.92	90.77±0.80
20 samples per class	85.73±0.61	87.58±0.45
15 samples per class	82.35±0.56	85.67±0.78

The standard deviation is represented as ±SD in Table 2 to show the average dispersion of the results relative to the mean. The results from Table 2 show that adding the channel attention mechanism on the base Siamese network improved the accuracy by approximately 3.04%. The study investigated the optimal layer for the attention module insertion on the Siamese network architecture, and the results showed the accuracy of the SIAM-MICR + Attention of 90.35%, 88.62%, and 90.46% at CNN layers 1, 2 and 3, respectively, averaged at 20 runs. Similarly, the study investigated the optimal batch size based on the best-performing sample size of 25 samples per class, and the results showed the accuracy of the SIAM-MICR + Attention of 89.68%, 90.25%, 89.83% and 90.66% at batch sizes of 8,16,32, and 64, respectively, averaged at 20 runs.

Performance Analysis on AUC - ROC Curve

This is a performance measurement for the classification problems at various threshold settings and tells how much the models could distinguish the classes based on the predicted Euclidian distance for similar and dissimilar character classes. The higher the AUC value, the better the model can distinguish whether the actual Euclidean distance is 0 or 1. The study's experiments on 25 sample sizes, as shown in Fig 4, show a 98.2% AUC value for the SIAM-MICR + Attention and a 94.9% AUC value for the SIAM-MICR. From these results, it is agreeable that the model has a good measure of separability since the AUC values are closer to 1 than 0.

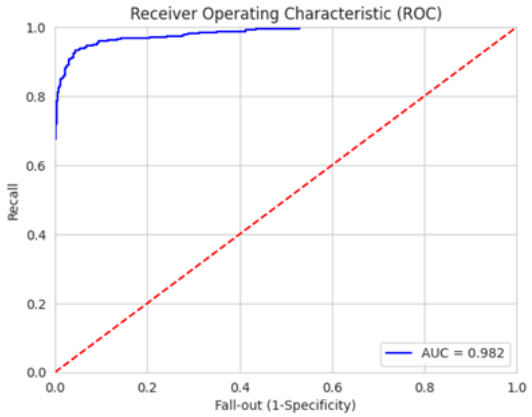

Fig. 4: SIAM-MICR + Attention ROC AUC

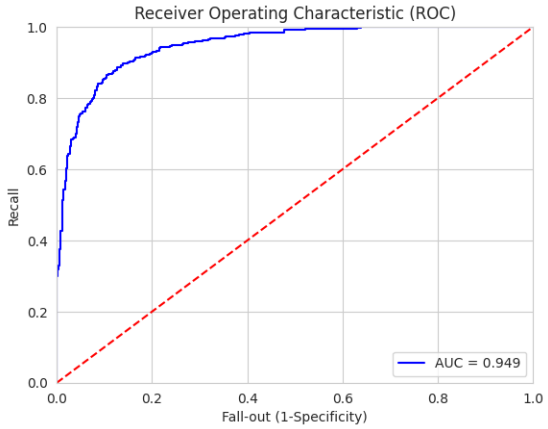

Fig. 5: SIAM-MICR + Attention ROC AUC

Performance Analysis using feature map visualisation.

The intuition here is that the channel attention module inserted after the first convolutional layer acts as a masking matrix that identifies and locates the prominent regions that contain significant morphological characteristics and passes these reinforced identified representations to the subsequent layers of the network for better representation learning. This enables the network to focus only on a certain part of the feature maps that is more prominent and, therefore, more discriminating. This leads to the network achieving a lower loss and learning a finer decision boundary between character classes. This is demonstrated visually in Fig 7 below, where the feature map

visualisation of the output of the second convolutional layers with the channel attention module shows that more prominent regions of the characters are densely populated with pixels when compared with the output of the second convolutional layers without the attention module in Fig 6, where these prominent regions' pixels are missing or very limited. Fig 6 and Fig 7 are shown below.

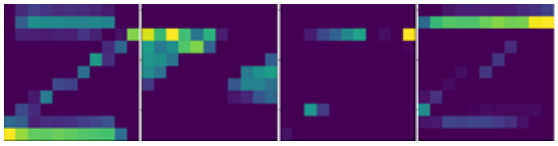

Fig 6. SIAM-MICR's Second layer output for Character "Z."

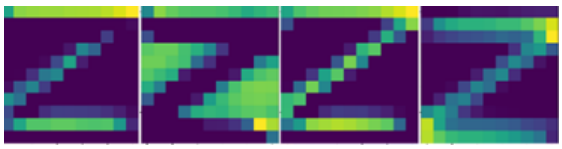

Fig 7. SIAM-MICR + Attention's Second layer output after channel attention module insertion - Character "Z."

It has been shown that a prominent data representation improves performance compared to a poor data representation, as the DL algorithm is highly dependent on the integrity of the input-data representation [17]. Hence, the DL algorithm in SIAM-MICR + Attention, which has its feature maps visualised after the attention module in Fig 7, will better perform distinguishing characters during recognition.

Quantitative Analysis with Related Works

To further support setting a benchmark on using Siamese for MICR, the study investigated notable past works on OCR using a Siamese network on limited samples. The Medpix dataset was used for the quantitative analysis with the network architectures in [18,19], with 25 samples per class, and the results are presented in Table 3 below.

Table 3: Comparison with related works on the Medpix Dataset on 62-ways 25-shot learning. Averaged on 30 runs.

Works	**Trainable Parameters**	**Accuracy**
[18]	50,184,000	83.12%
[19]	10,234,502	80.24%
SIAM-MICR	187,406	87.36%
SIAM-MICR + Attention	187,409	90.58%

The results shown in Table 3 show that our models require fewer parameters and faster model training due to better convergence during the training phase than existing Siamese networks from the past works compared in Table 3. Therefore, it can be agreed that our models are more efficient than [18] and [19]. Hence, our model is more memory efficient and less computational power is an advantage, therefore setting a benchmark for Siamese network on VSCs for medical image character recognition with limited samples.

V. CONCLUSION

In this paper, we proposed a channel attention-based Siamese neural network, optimised using the Bayesian optimisation technique for its hyperparameters, suited for metric learning to adequately learn a discriminative pattern of individual classes for MICR, with the existing problems of low resolution and background interference. Our experiments revealed that the channel attention module, with an adaptive kernel size of 3 and CNN configuration of 3 X 3 filters, is able to perform better when compared to the Siamese neural network without an attention module. Our proposed model learns a sharp decision boundary, as evident on the AUC-ROC curve, as seen in Fig 4, to differentiate different characters, which are visually similar where the problem of low resolution and background interference exists. There is an overall increase in accuracy, especially with limited samples per class, by achieving good accuracy and reduced training parameters compared to related past works. Hence, our model achieves good character recognition with less computational resources. Furthermore, the architecture of our proposed model is generic. It can be applied for any few-shot learning task, where there are cases of limited samples per class, visually similar images, and problems of low resolution with background interference.

In future work, we will consider multi-scale modelling techniques in metric learning with Siamese neural networks; it may help learn more features at different scales from low-resolution images, as seen in [20,21].

REFERENCE

[1] F. Röhrbein, P. Goddard, M. Schneider, G. James, and K. Guo, 'How does image noise affect actual and predicted human gaze allocation in assessing image quality?', Vision Research, vol. 112, pp. 11–25, Jul. 2015, doi: 10.1016/j.visres.2015.03.029.

[2] P. Inkeaw, J. Bootkrajang, S. Marukatat, T. Gonçalves, and J. Chaijaruwanich, 'Recognition of

similar characters using gradient features of discriminative regions', Expert Systems with Applications, vol. 134, pp. 120–137, Nov. 2019, doi: 10.1016/j.eswa.2019.05.050.
[3] S. T. Padmapriya and S. Parthasarathy, 'Ethical Data Collection for Medical Image Analysis: a Structured Approach', ABR, Apr. 2023, doi: 10.1007/s41649-023-00250-9.
[4]. K. He, N. Pu, M. Lao, and M. S. Lew, 'Few-shot and meta-learning methods for image understanding: a survey', Int J Multimed Info Retr, vol. 12, no. 2, p. 14, Dec. 2023, doi: 10.1007/s13735-023-00279-4.
[5]. T. Müller, G. Pérez-Torró, and M. Franco-Salvador, 'Few-Shot Learning with Siamese Networks and Label Tuning', 2022, doi: 10.48550/ARXIV.2203.14655.
[6]. S. Dey, A. Dutta, J. I. Toledo, S. K. Ghosh, J. Llados, and U. Pal, 'SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification', 2017, doi: 10.48550/ARXIV.1707.02131.
[7]. Q. Cao, Y. Ying and P. Li, "Similarity Metric Learning for Face Recognition," 2013 IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 2013, pp. 2408-2415, doi: 10.1109/ICCV.2013.299.
[8]. E. Hosseini-Asl and A. Guha, 'Similarity-based Text Recognition by Deeply Supervised Siamese Network', 2015, doi: 10.48550/ARXIV.1511.04397.
[9]. T. Wang, Z. Xie, Z. Li, L. Jin, and X. Chen, 'Radical aggregation network for few-shot offline handwritten Chinese character recognition', Pattern Recognition Letters, vol. 125, pp. 821–827, Jul. 2019, doi: 10.1016/j.patrec.2019.08.005.
[10]. Snell, J, Swersky, K, Zemel, Prototypical Networks for Few-Shot Learning. In Proceedings of the 31st International Conference on Neural Information Processing Systems 2017 (pp. 4080–4090). Curran Associates Inc..
[11] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W. and Hu, Q. (2019) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. DOI:10.48550/ARXIV.1910.03151.
[12] Z. Shen, M. Zhang, H. Zhao, S. Yi, and H. Li, "Efficient Attention: Attention with Linear Complexities." arXiv, 2018.
[13] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks." arXiv, 2017.
[14] I. Kandel, M. Castelli, and A. Popovič, "Comparative Study of First Order Optimizers for Image Classification Using Convolutional Neural Networks on Histopathology Images," Journal of Imaging, vol. 6, no. 9. MDPI AG, p. 92, 08-Sep-2020.
[15] E. Hassan, M. Y. Shams, N. A. Hikal, and S. Elmougy, "The effect of choosing optimizer algorithms to improve computer vision tasks: a comparative study," Multimedia Tools and Applications, vol. 82, no. 11. Springer Science and Business Media LLC, pp. 16591–16633, 28-Sep-2022.
[16] H. Lee, J. Lee, Y. Kwon, J. Kwon, S. Park, R. Sohn, and C. Park, "Multitask Siamese Network for Remote Photoplethysmography and Respiration Estimation," Sensors, vol. 22, no. 14. MDPI AG, p. 5101, 07-Jul-2022.
[17] 5. L. Alzubaidi et al., 'Review of deep learning: concepts, CNN architectures, challenges, applications, future directions', J Big Data, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
[18] Q. Wang and Y. Lu, "Similar Handwritten Chinese Character Recognition Using Hierarchical CNN Model," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 2017, pp. 603-608, doi: 10.1109/ICDAR.2017.104.
[19] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese Neural Networks for One-shot Image Recognition," 2015.
[20] N. K. Mishra, M. Dutta, and S. K. Singh, "Multiscale parallel deep CNN (mpdCNN) architecture for the real low-resolution face recognition for surveillance," Image and Vision Computing, vol. 115. Elsevier BV, p. 104290, Nov-2021.
[21] Q. Yuan, Y. Wei, X. Meng, H. Shen and L. Zhang, "A Multiscale and Multidepth Convolutional Neural Network for Remote Sensing Imagery Pan-Sharpening," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 3, pp. 978-989, March 2018, doi: 10.1109/JSTARS.2018.2794888.