

Classification of Graphomotor Impressions using Convolutional Neural Networks: An Application to Automated Neuro-psychological Screening Tests

Haris Bin Nazar*, Momina Moetesum[†], Shoaib Ehsan[‡], Imran Siddiqi[†]
Khurram Khurshid*, Nicole Vincent[§] & Klaus D. McDonald-Maier[‡]

* Institute of Space Technology, Islamabad, Pakistan

[†] Bahria University, Islamabad, Pakistan

[‡] University of Essex, England

[§] Paris Descartes University, Paris, France

Correspondence: momina.moetesum@bui.edu.pk, imran.siddiqi@bahria.edu.pk

Abstract—Graphomotor impressions are a product of complex cognitive, perceptual and motor skills and are widely used as psychometric tools for the diagnosis of a variety of neuro-psychological disorders. Apparent deformations in these responses are quantified as errors and are used as indicators of various conditions. Contrary to conventional assessment methods where manual analysis of impressions is carried out by trained clinicians, an automated scoring system is marked by several challenges. Prior to analysis, such computerized systems need to extract and recognize individual shapes drawn by subjects on a sheet of paper as an important pre-processing step. The aim of this study is to apply deep learning methods to recognize visual structures of interest produced by subjects. Experiments on figures of Bender Gestalt Test (BGT), a screening test for visuo-spatial and visuo-constructive disorders, produced by 120 subjects, demonstrate that deep feature representation brings significant improvements over classical approaches. The study is intended to be extended to discriminate coherent visual structures between produced figures and expected prototypes.

I. INTRODUCTION

Neuro-psychological assessments are performance-based methods which are designed to assess various cognitive abilities such as memory, attention, reasoning and problem-solving of an individual [1]. The results obtained from these assessments are used for the purposes of early detection and diagnosis of various neuro-psychological disorders, rehabilitation and to correlate with other clinical findings. Most of these tests involve ‘pencil-and-page’ based tasks which require subjects to reproduce a stimulus either by copying it or by memory. Trained assessors inspect these graphomotor responses and score deviations. The extent of deformation indicates the practitioners to determine various aspects of the neuro-psychological state of the subject. These tests are non-intrusive and easy to administer and give an overall picture of the working of the brain; hence are increasingly gaining popularity in clinical neuro-psychology.

Graphomotor skills are directly linked with handwriting and sketches. These graphomotor impressions are a product of complex cognitive, perceptual and motor skills [2]. They can give an insight into the ability of a person to interpret,

memorize and reproduce, what he sees and thus are considered useful indicators in neuro-psychology. Writing and drawing tests are widely used as psychometric tools for the diagnosis of a variety of neuro-psychological disorders such as dyspraxia, visuo-spatial neglect and Parkinson. In addition to being perceived as screening tools for the presence of any neuro-psychological disorder, handwriting and hand drawn impressions have also been linked to the behavior, emotion and personality of the writer in psychology. Popular tests involving graphomotor tasks include Bender Gestalt Visual Motor Test [3], Clock Draw Test [4] and Rey Osterrieth Complex Figure Test [5] as shown in Figure 1.

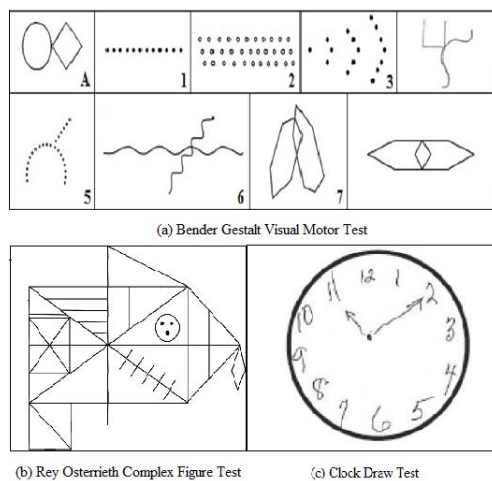


Fig. 1. Examples of Neuro-psychological Graphomotor Tests

Bender Gestalt Visual Motor Test [3] is one such popular screening test that evaluates the visual-motor maturity and perceptual distortions associated with various neuro-psychological disorders. The test requires the participants to copy nine different shapes shown to them. The produced figures are then analyzed and errors in copying shapes (like

rotation, simplification, inability to overlap etc.) are scored to assess the neurological state of the subject. Conventional assessment of this test involves statistical examination of the completed patient responses by trained assessors. Manual identification and scoring of errors in these responses is a time consuming activity. Despite using standard guidelines for scoring, the presence of human biasness cannot be ruled out. Studies [6], [7], [8], [9] suggest that development of a computer-based framework to facilitate analysis of these tests will not only allow standardization but will also facilitate the practitioners to focus more on diagnosis and future test development.

Nevertheless despite its advantages, designing and developing a reliable and robust computer-based analysis system for such tests is a challenging task. The first challenge in such an automation is the localization, segmentation and classification of the objects of interest from within an off-line, scanned image of the sample. Despite being a simple task for human beings, classification of the subject's intended figure is fairly complex.

This paper presents the findings of applying Convolutional Neural Networks (CNN) as feature extractor for recognition of the Bender Gestalt drawings. Transfer learning with pre-trained CNNs is employed for feature extraction while classification is carried out using Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA). The high classification rates realized using the hybrid CNN-SVM and CNN-LDA models support the use of deep learning in such systems.

The paper is organized as follows. We first briefly discuss current state-of-the-art in computerized analysis of hand drawn figures in the next section. Section III describes the proposed methodology followed by a discussion on the realized results in Section IV. Finally, we conclude the paper with a discussion on our further work on this problem.

II. RELATED WORK

Over the past few decades, computerized analysis of handwriting has been performed for a variety of tasks like handwriting recognition [10], writer identification [11], signature verification [12] and prediction of writer demographics from handwriting [13]. Despite the advancements in technology and maturation of computerized algorithms, prospects of automated analysis of handwriting and hand-drawn shapes for health and behavior interpretation of a writer have not been fully explored. A major factor limiting research in this area was the hesitation of neuro-psychologists to embrace applied computer technologies in their work. Nevertheless, recent years have seen a shift in trends where computer-based technologies have moved from emergent status to current acceptance amongst practitioners. According to studies [1], [7], clinical practitioners agree that computer-based neuro-psychological assessments have inherent features that are absent in traditional forms. This encouraged computer scientists to initiate research ventures in this domain.

Among computerized systems for analysis of neurological disorders, Remi et al. [6] present a study for the detection

of graphomotor difficulties present in hand drawn samples of children with learning difficulties. The authors employ handwritten sentences and a set of geometrical shapes as stimuli. Straight lines, circles and squares are identified using Hough transform and the effectiveness of various features in analyzing these primitives is discussed. Authors in [8] propose a framework that involves data acquisition, analysis of features and compilation of results covering a number of drawing and handwriting based test batteries. Two clinical conditions i.e. visuo-spatial neglect and dyspraxia, are discussed as case study, and a generic structured approach to system implementation is demonstrated. In a similar study [9], analysis of online geometrical sketches is carried out for visuo-spatial classification. For each shape, a set of interest points is identified and local descriptors in neighborhood of interest points are calculated. Dissimilarity between a drawn sketch and expected prototype is then calculated. Experiments on geometrical sketches (5 classes) produced by 19 different individuals reported promising classification rates.

A pilot study on partial automation of the Rey Osterrieth Complex Figure (ROCF) is presented in [14] where the authors focus on scoring the parts of drawing including triangles, rectangles, diamonds and simple lines. These geometrical sketches are located and identified using fuzzy metrics based on Gestalt laws. Analysis on 31 ROCF drawings reported correct localization of 99% of features of interest in the drawings. Recently, the problem of automatic assessment of Clock Draw Test (CDT) drawings attracted significant attention of the research community. CDT is considered an effective measure for early detection of dementia in the elderly. Studies like [15], [16] present image analysis based systems to analyze the CDT samples. Contrary to other tests which involve drawings and sketches, CDT involves recognition of handwritten digits which represents a mature area of research [17].

In addition to neuro-psychological tests, classification of hand-drawn sketches has also been investigated for other applications. These include generalized sketch based retrieval systems [18], [19], forensic applications [20] and recognition of domain specific sketches like UML diagrams [21], architectural drawings [22], circuit diagrams [23] and flow charts [24]. The key difference between recognition of general hand-drawn sketches and those in produced by subjects in neurological tests is the fact that the samples collected during these tests come from both healthy and unhealthy participants. While the healthy subjects produce drawings which are likely to be similar to the expected models, those produced by unhealthy subjects may deviate from the anticipated pattern to a great extent. This results in very high intra-class variation making classification of these drawings a challenging task. Deep learning based techniques, therefore, represent an attractive choice for classification of these visual structures and make the subject of our study as explained in the following section.

III. PROPOSED METHODOLOGY

This section presents the details of the proposed hybrid CNN-SVM and CNN-LDA models to recognize the BGT

drawings. We first briefly outline the segmentation technique that extracts individual drawings from the image produced by a subject. We then present an overview of CNNs and transfer learning followed by a discussion on the different pre-trained models employed in our study.

A. Segmentation of Drawings

The samples produced by subjects comprise drawings on a single sheet of paper. Since the subjects are provided with blank sheets without any markers, they may draw the figures at arbitrary positions on the paper. Consequently, the drawings first need to be segmented before they can be fed to the recognition engine. The segmentation technique used in our study relies on Gestalt laws of Perception [25], [26] and is presented in detail in [27]. For completeness, we briefly outline the employed segmentation technique. It can be observed from the BGT drawings (Figure 1-a) that the figures produced by subjects can be categorized into three classes [27] based on rules of Gestalt theory.

- 1) **Group A:** Enclosed shapes
- 2) **Group B:** Shapes formed by solid lines
- 3) **Group C:** Shapes formed by dots or small circles/lines

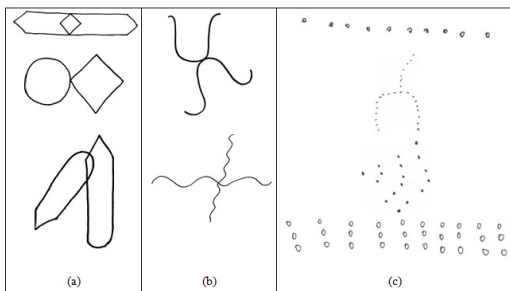


Fig. 2. (a): Group A - Enclosed shapes (b): Group B - Shapes formed by solid lines (c): Group C - Shapes formed by dots or small circles/lines

Different segmentation approaches are applied to images in each of the three groups. Drawings in Group-A represent enclosed shapes and are distinguished from other drawings on the basis of enclosed area which is much greater as compared to other shapes. Group B comprises drawings with solid lines which are distinguished from Group C using the foreground (drawing) area while Group C drawings are segmented using a k-means clustering on the spatial coordinates of the connected components. Complete details on the segmentation scheme can be found in [27].

The key contribution of the present study is investigation of CNN based feature extractors in classification of these drawings. The next section provides an overview of the CNNs.

B. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) [28], [29], though introduced in the 1990s, attracted the attention of computer vision community once the CNN based classification [30] outperformed conventional techniques by a significant margin in

the ImageNet Large Scale Visual Recognition Challenge [31]. Since then, CNN based techniques have been applied to a variety of recognition tasks and have reported high classification rates on a number of challenging datasets [32], [33], [34], [35], [36].

A Convolutional Neural Network [29] typically comprises a series of convolutional and sub-sampling layers followed by one or more fully connected layers. Convolutional layers, in most cases, are followed by a layer of ReLU activation functions. The convolutional, ReLU and the down sampling layers together serve as feature extractors while the fully connected layers represent a trainable classifier similar to a standard multi-layer neural network. In some cases, the fully connected layers (performing classification) are removed and the output of the feature extractor layers is fed to another classifier. The combination CNN-SVM, for instance, has been investigated in a number of studies [37], [38], [36], [39] and outperforms each of the individual classifiers.

Feature extraction is the most critical step in any pattern classification task. For an effective classification, discriminative characteristics are sought which tend to cluster objects of same class in the feature space. In most cases, the choice of features is driven by domain knowledge where the domain expert identifies the key discriminating factors among objects of different classes for a given problem. These attributes are then mapped to computational features and fed to a learning algorithm. Identification and extraction of these 'hand-crafted' features is not only a tedious task but does not work on raw images either. Machine extracted features, on the contrary, not only provide a generalized framework but also allow extraction of features from raw images directly.

Two issues with CNNs include the huge amount of training data and the computational time to train the network. Consequently, in practice, it is quite rare to train the network from scratch. Transfer learning is generally employed to address the aforementioned issues. Transfer learning exploits a pre-trained model and adapts it for a given problem using any of the following strategies.

- **Using a pre-trained model as a fixed feature extractor:** In this scheme, the last fully connected layer is removed and the weights of the feature extraction layers are left unchanged. A classifier is then trained on the outputs of the remaining layer using the dataset under study.
- **Fine-tuning a pre-trained model:** In this case, a pre-trained model is tuned on the dataset under study by continuing back propagation.

In our study, we employ transfer learning using a pre-trained model as a fixed feature extractor. The architectures employed in our work are discussed in the next section.

C. Architectures Employed

A number of convolutional neural network architectures have gained popularity over the years. These include LeNet [29], AlexNet [30], ZFNet [40], GoogLeNet [41] and VGGNet [42] and ResNet [43]. In our study, we have employed AlexNet and VGG (16 and 19 layer models) as fixed

feature extractors while classification is carried out using SVM and LDA. We briefly outline each of the two architectures in the following.

1) *AlexNet*: AlexNet [30] is considered to be one of the pioneer works that lead to the current popularity of CNNs. The AlexNet architecture comprises of 5 convolution layers, max-pooling layers, dropout layers, and 3 fully connected layer and is trained on 1.2 Million images (with 1000 different classes) of the ImageNet dataset.

2) *VGGNet*: The Visual Geometry Group (VGG) at University of Oxford released two of their best-performing architectures, VGG-16 and VGG-19 having 16 and 19 weight layers respectively [42]. The architecture comprises of 3x3 filters with stride and pad of 1, along with 2x2 max pooling layers with stride 2.

In the next section, we present the classification results of the proposed deep learning based recognition of BGT drawings.

IV. EXPERIMENTS AND RESULTS

We first present the data set employed in our evaluations followed by a discussion on the realized classification rates.

A. Dataset

To the best of our knowledge, no existing data set is available for most of the neuro-psychological tests. Due to this limitation, creating and labeling a dataset is one of the vital steps. For this study, we collected samples of 120 subjects varying from 16 to 66 years of age. All the samples were acquired under expert supervision of trained psychologists and were later scored. Approximately 65% of the samples are contributed by healthy subjects while 35% by subjects with some potential neurological disorders. Since the present study targets only the recognition part, the scoring will be vital once the complete system has been developed.

Drawings of 60 subjects were used as training while those of 60 subjects as the test set. This gives a total of 540 (60 × 9) images with 9 classes in each of the training and test sets. The training and test sets comprise an equal distribution of healthy and unhealthy subjects.

B. Results

The classification rates using the feature extraction layers of the pre-trained AlexNet, VGG16 and VGG19 are summarized in Table I. As discussed earlier, classification is carried out using SVM and LDA. It can be seen that all three architectures and both the classifiers realized more or less similar classification rates. A highest classification rate of 93.52% is realized using VGG16 architecture with LDA.

It should be noted that since the present study targets exploration of deep learning for classification of BGT drawings, segmentation results are not presented or discussed. From the view point of a practical system, the output of segmentation module is fed to the classification engine. Consequently,

some of the classification errors could be due to incorrect segmentation of drawings.

TABLE I
CLASSIFICATION RESULTS ON DIFFERENT PRE-TRAINED MODELS

Model	SVM	LDA
AlexNet	92.50%	91.50%
VGG-16	93.33%	93.52%
VGG-19	92.96%	93.15%

We also present the confusion matrix of the best performing combination (VGG16-LDA). It can be seen from Figure 3 that most of the classification errors result from the confusion between Drawing 1 and Drawing 2. This is very much natural as the two drawings have very similar visual appearance. In an attempt to study the stability of system performance, we carried out 5-fold cross validation using 80% of drawings in the training set and 20% in the test set. Using Alex-Net with SVM classifier, an overall average accuracy of 92.88% is reported which is comparable to the values in Table I indicating the stability of results to different combinations of training and test sets.

We also implemented the shape context descriptor feature [44] employed in [27] for classification of same drawings. Using the same experimental protocol, shape context descriptor realizes a classification rate of 86%. Significantly improved classification rates using CNN validate the idea that deep learning based feature extraction techniques outperform state-of-the-art shape matching descriptors for recognition of visual structures.

	A	1	2	3	4	5	6	7	8
A	55	2	1	0	2	0	0	0	0
1	0	55	5	0	1	0	0	0	0
2	2	2	54	0	2	0	0	0	0
3	0	0	0	59	0	0	1	0	0
4	0	0	3	0	56	1	0	0	0
5	0	0	0	2	0	58	0	0	0
6	0	0	0	1	0	1	55	1	2
7	0	0	0	0	0	0	4	56	0
8	0	0	0	2	0	0	1	0	57

Fig. 3. Confusion Matrix corresponding to the highest classification rate of 93.52%(VGG16+LDA)

V. CONCLUSION

Investigation of deep learning based feature extraction for recognition of free-hand drawings is presented using the figures of Bender Gestalt Test (BGT) as a case study. The hybrid CNN-SVM and CNN-LDA models are explored for classification. Evaluations on samples contributed by 120 subjects reported high classification rates using AlexNet and VGG architectures.

The presented work is a step towards development of a complete clinical decision support system to facilitate neuro-psychologists. In our further work, we intend to model the drawings and develop techniques to measure deformation

from the expected prototype. This will allow automatically identifying errors and scoring the drawings. This may require designing and training a novel CNN architecture from scratch. The major challenge in developing such a system is to have training data with instances of all possible errors that the subjects can make. This may require generating synthetic data and applying data augmentation techniques to allow sufficient training data. We also plan to propose a generalized framework that is able to model and measure deviations of hand produced impressions from expected patterns.

ACKNOWLEDGMENT

This work was carried out as a part of a project supported by the Franco-Pakistani Hubert Curien Partnership (PHC) - PERIDOT, implemented in Pakistan by the Higher Education Commission (HEC) and in France, by the Ministry of Foreign Affairs (MAE). We would also like to express our gratitude to Dr. Uzma Masroor, Department of Professional Psychology, Bahria University, Islamabad, Pakistan, for her guidance at different stages of the research.

REFERENCES

- [1] A. M. Poreh, *The quantified process approach to neuropsychological assessment*. Psychology Press, 2012.
- [2] J. Ziviani and M. Wallen, *The development of graphomotor skills*. Mosby Elsevier, 2006.
- [3] R. Walrath, "Bender visual motor gestalt test," in *Encyclopedia of Child behavior and Development*. Springer, 2011, pp. 233–234.
- [4] B. J. Mainland and K. I. Shulman, "Clock drawing test," in *Cognitive Screening Instruments*. Springer, 2013, pp. 79–109.
- [5] M.-S. Shin, S.-Y. Park, S.-R. Park, S.-H. Seol, and J. S. Kwon, "Clinical and empirical applications of the rey-osterrieth complex figure test," *Nature protocols*, vol. 1, no. 2, pp. 892–899, 2006.
- [6] C. Rémi, C. Frélicot, and P. Courtellemont, "Automatic analysis of the structuring of children's drawings and writing," *Pattern Recognition*, vol. 35, no. 5, pp. 1059–1069, 2002.
- [7] P. Schatz and J. Browndyke, "Applications of computer-based neuropsychological assessment," *The Journal of head trauma rehabilitation*, vol. 17, no. 5, pp. 395–410, 2002.
- [8] M. C. Fairhurst, T. Linnell, S. Glenat, R. Guest, L. Heutte, and T. Paquet, "Developing a generic approach to online automated analysis of writing and drawing tests in clinical patient profiling," *Behavior Research Methods*, vol. 40, no. 1, pp. 290–303, 2008.
- [9] N. Renau-Ferrer and C. Remi, "A method for visuo-spatial classification of freehand shapes freely sketched," *arXiv preprint arXiv:1305.1520*, 2013.
- [10] R. Plamondon and S. N. Srihari, "Online and off-line handwriting recognition: a comprehensive survey," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 1, pp. 63–84, 2000.
- [11] A. Bensefia, T. Paquet, and L. Heutte, "A writer identification and verification system," *Pattern Recognition Letters*, vol. 26, no. 13, pp. 2080–2092, 2005.
- [12] K. Cpałka, M. Zalasinski, and L. Rutkowski, "New method for the on-line signature verification based on horizontal partitioning," *Pattern Recognition*, vol. 47, no. 8, pp. 2652–2661, 2014.
- [13] I. Siddiqi, C. Djeddi, A. Raza, and L. Souici-Meslati, "Automatic analysis of handwriting for gender classification," *Pattern Analysis and Applications*, vol. 18, no. 4, pp. 887–899, 2015.
- [14] R. Canham, S. L. Smith, and A. M. Tyrrell, "Automated scoring of a neuropsychological test: the rey osterrieth complex figure," in *Euromicro Conference, 2000. Proceedings of the 26th*, vol. 2. IEEE, 2000, pp. 406–413.
- [15] H. Kim, Y. S. Cho, and E. Y.-L. Do, "Computational clock drawing analysis for cognitive impairment screening," in *Proceedings of the fifth international conference on tangible, embedded, and embodied interaction*. ACM, 2011, pp. 297–300.
- [16] Z. Harbi, Y. Hicks, and R. Setchi, "Clock drawing test digit recognition using static and dynamic features," *Procedia Computer Science*, vol. 96, pp. 1221–1230, 2016.
- [17] M. Diem, S. Fiel, A. Garz, M. Keglevic, F. Kleber, and R. Sablatnig, "Ic-dar 2013 competition on handwritten digit recognition (hdc 2013)," in *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*. IEEE, 2013, pp. 1422–1427.
- [18] C. Alvarado and R. Davis, "Sketchread: a multi-domain sketch recognition engine," in *Proceedings of the 17th annual ACM symposium on User interface software and technology*. ACM, 2004, pp. 23–32.
- [19] L. B. Kara and T. F. Stahovich, "An image-based, trainable symbol recognizer for hand-drawn sketches," *Computers & Graphics*, vol. 29, no. 4, pp. 501–517, 2005.
- [20] Y. Zhang, C. McCullough, J. R. Sullins, and C. R. Ross, "Hand-drawn face sketch recognition by humans and a pca-based algorithm for forensic applications," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 3, pp. 475–485, 2010.
- [21] T. Hammond and R. Davis, "Tahuti: A geometrical sketch recognition system for uml class diagrams," in *ACM SIGGRAPH 2006 Courses*. ACM, 2006, p. 25.
- [22] S. Ahmed, M. Weber, M. Liwicki, C. Langenhan, A. Dengel, and F. Petzold, "Automatic analysis and sketch-based retrieval of architectural floor plans," *Pattern Recognition Letters*, vol. 35, pp. 91–100, 2014.
- [23] M. Angadi and R. Lakshman Naika, "Handwritten circuit schematic detection and simulation using computer vision approach," *International Journal of Computer Science and Mobile Computing*, vol. 3, no. 6, pp. 754–761, 2014.
- [24] W. Szwoch, "Recognition, understanding and aestheticization of free-hand drawing flowcharts," in *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, vol. 2. IEEE, 2007, pp. 1138–1142.
- [25] F. Attneave, "Some informational aspects of visual perception," *Psychological review*, vol. 61, no. 3, p. 183, 1954.
- [26] J. Wagemans, J. H. Elder, M. Kubovy, S. E. Palmer, M. A. Peterson, M. Singh, and R. von der Heydt, "A century of gestalt psychology in visual perception: I. perceptual grouping and figure-ground organization," *Psychological bulletin*, vol. 138, no. 6, p. 1172, 2012.
- [27] M. Moetesum, I. Siddiqi, U. Masroor, N. Vincent, and F. Cloppet, "Segmentation and classification of offline hand drawn images for the bgt neuropsychological screening test," in *Eighth International Conference on Digital Image Processing (ICDIP 2016)*. International Society for Optics and Photonics, 2016, pp. 100 334N–100 334N.
- [28] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Handwritten digit recognition with a back-propagation network," in *Neural Information Processing Systems (NIPS)*, 1989.
- [29] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [31] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 248–255.
- [32] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Convolutional neural network committees for handwritten character classification," in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*. IEEE, 2011, pp. 1135–1139.
- [33] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, and G. Penn, "Applying convolutional neural networks concepts to hybrid nn-hmm model for speech recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 4277–4280.
- [34] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.
- [35] A. T. Lopes, E. de Aguiar, and T. Oliveira-Santos, "A facial expression recognition system using convolutional networks," in *Graphics, Patterns and Images (SIBGRAPI), 2015 28th SIBGRAPI Conference on*. IEEE, 2015, pp. 273–280.

- [36] X.-X. Niu and C. Y. Suen, "A novel hybrid cnn-svm classifier for recognizing handwritten digits," *Pattern Recognition*, vol. 45, no. 4, pp. 1318–1325, 2012.
- [37] M. Szarvas, A. Yoshizawa, M. Yamamoto, and J. Ogata, "Pedestrian detection with convolutional neural networks," in *Intelligent vehicles symposium, 2005. Proceedings. IEEE*. IEEE, 2005, pp. 224–229.
- [38] K. Mori, M. Matsugu, and T. Suzuki, "Face recognition using svm fed with intermediate output of cnn for face detection." in *MVA*, 2005, pp. 410–413.
- [39] F. Lauer, C. Y. Suen, and G. Bloch, "A trainable feature extractor for handwritten digit recognition," *Pattern Recognition*, vol. 40, no. 6, pp. 1816–1824, 2007.
- [40] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*. Springer, 2014, pp. 818–833.
- [41] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [43] 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.
- [44] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 4, pp. 509–522, 2002.