Proc. of the 2017 IEEE International Conference on Signal and Image Processing Applications (IEEE ICSIPA 2017), Malaysia, September 12-14, 2017

Application of image retrieval for aesthetic evaluation and improvement suggestion of isolated Bangla handwritten characters

Mithun Biswas*, Rafiqul Islam†, Gautam Kumar Shom‡, Nabeel Mohammed§, Sifat Momen¶, Nafees Mansoor‖, Anowarul Abedin**
*†‡¶‖**Department of Computer Science and Engineering
University of Liberal Arts Bangladesh, Satmasjid Road, Dhaka, Bangladesh
§Department of Electrical and Computer Engineering
North South University, Bashundhara, Dhaka, Bangladesh.
Email: nabeel.mohammed@northsouth.edu, sifat.momen@ulab.edu.bd

***Abstract*—Bangla is one of the most widely used languages worldwide. This paper presents an application of image retrieval techniques to automatically judge the aesthetic quality of handwritten Bangla isolated characters. Retrieval techniques are also adapted to give improvement suggestions, with a plan to incorporate the methods in applications which can assist in learning/teaching handwriting. The proposed method borrows key concepts from content-based image retrieval. Our method was tested on the BanglaLekha-Isolated data set, which contains images of 84 Bangla characters, with nearly 2000 samples per character. The data set contains evaluation of the aesthetic quality of the handwriting judged on a scale of 1 - 5. For this work, the dataset was partitioned into a test set of 400 images and a database-set of ≈ 1600 images, per Bangla character. Assuming that a scoring difference of 1 is acceptable, the proposed method achieves an accuracy of 77.39% when using features extracted by a convolutional neural network based autoencoder. Experiments were also done with the popular HOG feature. However, the autoencoder-based results showed clear superiority compared the HOG-based results. Our proposed method for improvement suggestions also shows that it is possible to shows samples from the dataset which will help users improve their handwriting while requiring small changes to their own handwriting.**

***Index Terms*—Bangla handwriting Aesthetic Evaluation Image Retrieval Autoencoder HOG**

I. Introduction

While Optical Handwriting Recognition (OHR) has received a lot of interests over the past few years, the evaluation of its aesthetic quality still remains as a largely ignored area. Good handwriting at an early age has been found to play positive role on the success of a person. For instance in [7], Hughes and colleagues reported that there is a strong positive correlation between the marks received in assessment items with that of the quality of handwriting. This finding takes on even more importance when the compounding effect of good marks on a students confidence is understood. Students who receive positive feedback on their work early in their academic life tend to go forth with better confidence, compared to students who are faced with initial setbacks and/or disappointments.While Optical Handwriting Recognition (OHR) has received a lot of interests over the past few years, the evaluation of its aesthetic quality still remains as a largely ignored area. Good handwriting at an early age has been found to play positive role on the success of a person. For instance in [7], Hughes and colleagues reported that there is a strong positive correlation between the marks received in assessment items with that of the quality of handwriting. This finding takes on even more importance when the compounding effect of good marks on a students confidence is understood. Students who receive positive feedback on their work early in their academic life tend to go forth with better confidence, compared to students who are faced with initial setbacks and/or disappointments.Bangladesh is a country of over 160 million people with Bangla as the official language. Out of these 160 million people, 55 million of them go to schools and universities with majority of them going to Bangla medium schools where Bangla is the key language used to understand, comprehend and write. Therefore, for a country like Bangladesh, it is important to nurture handwriting quality from an early age. However, due to the shortage of resources such as teachers with specialized training to teach handwriting, this is not always possible. Bangladesh Government, recognizing this shortcoming, has invested in using IT as a tool to mitigate the gap. This paper reports on the findings of the application of image retrieval (IR) techniques to the task of evaluating the aesthetic quality of Bangla isolated characters.In particular, this paper presents two contributions the first being a computerized mimicry of a handwriting evaluation method proposed by Marlow [5] through the application of image retrieval and the second contribution being an extension of the first to devise a method to suggest improvements to handwritten samples. As mentioned previously, the paper reports on a IR-based approach to the tasks, specifically one which leverages Content-based Image Retrieval (CBIR) like techniques. To this end, this study reports results on the BanglaLekha-Isolated dataset, which contains over $1,60,000$ samples of 84 different handwritten Bangla characters. The dataset also contains numeric marks (in a scale of 1-5) indicating a coarse aesthetic quality of handwriting. We partitioned the dataset into a testing set and a database-set to evaluate the efficacy of our proposed IR-based methods. Handwriting score ascription and improvement suggestion was done using the HOG feature and also a convolutional neural network based features extracted from a low dimensional representation of an autoencoder trained on Bangla letters in an unsupervised manner.

II. BACKGROUND

Studies, since the 1920s, consistently demonstrated that there exists strong positive correlation between good handwriting and getting high scores in essay writing [7].This has important implications for children in particular, as a series of poor grades at a young age can quite severely dent the confidence of a child, who otherwise could have been a high achieving student. Furthermore, recent studies have also shown that good handwriting skills re mediate certain reading-related learning disabilities [16]. Assessing the quality of handwriting is a complex task. While many criteria to assess handwriting have been proposed for the English language [6, 12], there is virtually none for Bangla. In [5], Marlow argues for the case of using an established criterion for such assessment, but then also points out that for early age students, the teachers examples and assessments are essentially the most effective technique. In order to achieve higher objectivity in marking, Marlow proposed a rubric-based approach to assess the handwritings of students[5]. In this approach, he suggested to arrange the specimens from low quality to high quality. Teachers may assess each persons handwriting by moving the sample under consideration to the location where it best matches with the quality of the previously arranged specimens in the rubric. Marlows method entails the judgment of visual similarity between a sample and other already marked samples. Inspired by this approach, we present the use of Content-Based Image Retrieval (CBIR) techniques as a method to judge similarity of handwriting samples with already-scored handwritten samples of a database. Content Based Image Retrieval is a at its core a search system where the query presented is in the form of a digital image, instead of text [15]. Such systems usually have a database of search-able images, and the query may be presented using many different methods, e.g. image selection, drawing etc. Figure 1 shows a basic scheme of a stripped-down CBIR, to highlight the parts relevant to this study. The system extracts image features from the searchable database and stores them. When a query image is presented, the same feature is extracted and the features are compared. The task of a CBIR system is to present the user with a list of images sorted in descending order of feature similarity. The hypothesis is that similarity in feature space will translate to similarity in image space. The feature distances can be measured using many different methods such as Euclidean distance, L1 distance, Cosine distance etc. Quite evidently, image features play an important role in this process. This study reports results using two different image features which are discussed below.

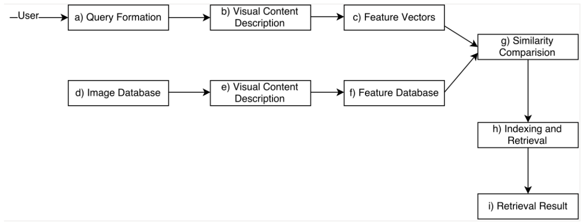

Fig. 1. CBIR schematic diagram

A. Image Features

An image feature is any information collected from an image, which can be useful in certain applications. There have been a great many different useful features pro- posed in the literature, however most of them can be thought of as a vector of some high dimensional space representing some information about the image. The encoding of features as vectors is also the reason we can easily use the distance measures mentioned previously. This study reports results using the popular Histogram of Oriented Gradients (HOG) feature and an Autoencoder-based feature. This is discussed below.

1) Histogram of Oriented Gradients(HOG): Histogram of Oriented Gradients (HOG) is a feature initially proposed by [2] in a human identification task, however since then it has been successfully applied to other applications [4][11]. It is a simple yet effective features based on accumulating pixel-wise gradients in block-based gradient histograms. The original image is first subdivided in difference cells of a configurable size. The gradients of pixels in each cell is calculated and then accumulated in cell-level histograms. The numbers of bins are also configurable, but [2] used 8 bin histograms. The histograms of each cell are then concatenated together to form an image descriptor. This feature descriptor is useful because it ignores very small deviations in image gradients (due to cell-wise division), but overall captures the shape in a particular image.

2) Autoencoder-based feature: For this paper we will limit the discussion to convolutional neural network (CNN [8] based autoencoders[9]. Although most CNNs are followed by one or more fully connected layers for classification, it is possible to arrange a CNN architecture so that the it receives its input in the form of a digital image and then through successive convolutional and pooling [14] operations create a low dimensional

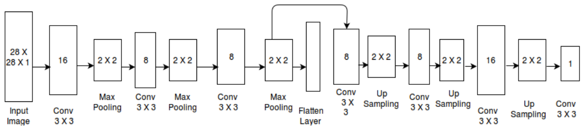

Fig. 2. Example Autoencoder architecture whose input is an image and the expected output is the same image

representation of the image. This representation can then be used by successive convolution and up-sampling operations to re-create an image with the dimensionality of the original image. Such a scheme is represented in Figure 2.

An autoencoder is simply such a network which is asked to reproduce, as its output, the original input image. This is not perfectly possible in most cases, as the low dimensional representation causes some information loss. However, when trained, the low dimensional representation learns a representation of the original image which can be useful to reproduce the original image to a great degree of accuracy. The hypothesis is that if the network is successful in recreating the image, albeit a lossy one, then the low dimensional representation is capturing some important information about the input. In some literature, this representation is called a laten representation. For the purposes of this study which concentrates on the beauty of a character, the encoded representation can be a useful feature.

B. Proposed Method

The method outlined by Marlow in [5] can be summarized by the following steps:

1) For each score of 0 5, find handwriting samples which receive the score.
2) When an unmarked handwriting sample is found, compare it with all the samples in (1).
3) Find the sample from (1) which is closest visually to the unmarked sample.
4) Assign the mark of the closest sample to the unmarked sample.

Our proposed method attempts to be an automated mimicry of the method mentioned above. Step 1 is implemented by collecting handwriting samples from 2000 people, each containing 84 Bangla characters. Each sample was then assessed by handwriting experts in a scale of 0 5, with 0 being the worst and 5 being a good handwriting sample.Steps 2,3 and 4 can be reduced to an image search problem, where the query image is the unmarked handwriting sample and the database is the collection of marked samples. If we can identify the character, then it is possible to find the character closest to the unmarked sample and ascribe the corresponding mark to the sample.

More formally, we define the mark of an unmarked sample s of character l, as

$$M(s) = M(\delta_f(s, D_l)) \tag{1}$$

Where $M(x)$ is the mark ascribed to an image x containing a character. is the search function which searches the database D_l using the feature f to return the image closest to s, where D_l is the subset of the entire database D containing all the images of the character l. It is assumed that $M(x), \forall_x \epsilon D$ is known apriori.

1) Improvement Suggestion: This then also allows us to define a simple method to suggest improvements. For this initial work our proposed method for suggestion improvement is finding the image t of the character l from D_l , most similar (in feature space) to the test sample s with $M(t) > M(s)$. This proposed method relies on the presence of a good database of images D, to ascribe marks to a character and also to suggest improvements. The finding is then a simple re-utilisation of the search function

$$t = \delta_f(s, D'_l) \tag{2}$$

Where D_l is a subset of the entire database D, which only contains images of the character l and where $M(i) = M(s) + 1, i \epsilon D_l$. t is then the suggested improved version of the character drawn in s. This method assumes that the character drawn in s has been identified. Previous work done by authors of this paper and other distinguished researchers has exclusively concentrated on Bangla character recognition, and is not included in the scope of this paper.

C. The BanglaLekha-Isolated dataset

Although there exists many widely used Bangla handwritten isolated character datasets [13][1], we faced two problems with them. Firstly, the data sets do not have a good distribution of handwriting, in terms of aesthetic quality. Secondly, the data for different types of Bangla characters, e.g. numerals, compound characters, vowels etc., were collected separately. The BanglaLekha-Isolated dataset [18] to be a multi-purpose large dataset of Bangla isolated handwritten characters. The dataset includes $1, 66, 105$ samples of 84 different Bangla characters collected from a variety of locations within Bangladesh. The handwriting samples were assessed for aesthetic quality by three literate native Bangla speakers following the criteria set by a nationally recognised Bangla handwriting expert. Each of three scored the samples between $0 - 5$, with 5 indicating clear and good handwriting. The dataset has also been pre-proccessed to make its presentation more like the popular MNIST dataset[3], with the view of facilitating more consistent research finding.

- Consistent Size and format
- Clear and easy to read
- Consistent style
- Proper dimension
- Correctness.

For each Bangla character, the dataset provides $\approx$ 2000 images. The marks given To test the efficacy of the system, we randomly selected 400 images/character to be part of the testing set, and the rest were made part of the database D_l

III. Experiments

A. Feature extraction

1) HOG Feature Extraction: HOG features were extracted with an 8 8 cell size. The gradients in each cell was accumulated in an 8 bin histogram, giving a total image descriptor of 72 dimensions. Using a cell size of 8 does not give coverage on the edges of the image, but that is not important as the edges rarely ever contain parts of the drawn character.

2) Autoencoder feature: An autoencoder with the architecture shown in Figure 2 was trained using the 1,34,400 characters of the database. The autoencoder uses convolution layers and maxpooling layers to create a lower dimensional representation of the image. This lower dimensional representation is a 3 dimensional box, which is then further convolved and upsampled to re-create the image. The layer marked Flatten is not actually a part of the network, but is the vector form (unrolled) of the low dimensional representation.
The network is trained with images as input and the same image given as expected output. We used the Adadelta [17] optimiser with the binary cross entropy loss function.

B. Distance measures

The proposed method depends on finding an image from D_l , which is most similar to the query/test image of a character l. The HOG and autoencoder-based features were used with the the Euclident, L1 and Cosine distance measures shown in Equations 3-5, to determine the image in D_l with the shortest distance. In this equations, A and B are two feature vectors of n dimensions. $|A^i|$ is the absolute value of the i^{th} element of vector A and $|A|$ is the norm of vector A.

$$EuclideanDistance(A,B) = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2} \quad (3)$$

$$L1Distance(A,B) = \sum_{i=1}^{n} |A_i - B_i| \quad (4)$$

$$CosineDistance(A,B) = \frac{A.B}{|A||B|} \quad (5)$$

C. Performance measurement

1) Scoring accuracy: Although we have employed image retrieval techniques,the applications is strictly not presented to the user as a retrieval-based application. Therefore, classical performance image retrieval measures, such as precision and recall [10], are not applicable in this context. For each test image, the score suggested by the proposed method is compared to the expert-given score of the same image. If the scores have a difference of 1 or less, then we assume that the predicted score is accurate, otherwise not. However, when reporting results we will show a breakdown of how many predicted score differ by 0, 1, 2, 3 and 4. We use the term score difference (S.D.) to refer to this difference in mark. The threshold of 1 S.D. was chosen because the average standard deviation of the expert scores was 1.08.

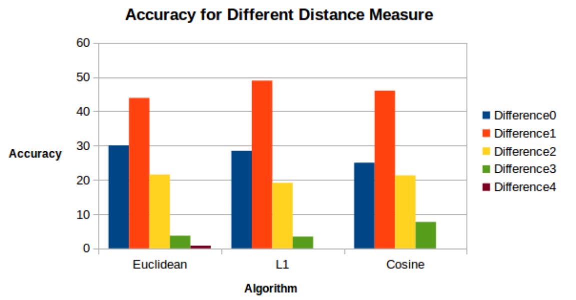

Fig. 3. Score prediction performance using different distance measures with the autoencoder feature

2) Improvement Suggestion: Unfortunately there is no quantitative method of measuring whether the suggested improvement actually improve the sample. This is because what makes a character better is still largely a subjective human evaluation. In this paper, suggested improvements will be shown side-by-side with the test character and it is up to the reader to make a judgment on whether the suggestions actually "improve" the test characters or not.

IV. Results

Table 1 and 2 tabulates the percentage of times the score predictions from the proposed method matched the expert-given scores (difference of 0 and 1) and otherwise. Figures 3 and 4 shows the graphical representation of the same information. The most effective results were obtained when using the autoencoder features with the L1 distance function, achieving a S.D. 0 accuracy of 28.45% and a S.D. 1 accuracy of 48.93%, for a combined accuracy of 77.39%. The extra effort invested in training the Autoencoder is worthwhile as the best performing result when using HOG is significantly lower at 74.46%.

TABLE I
Handwriting Mark Accuracy using HOG feature

	Euclidean(%)		L1(%)		Cosine(%)	
Difference0 Difference1	29.78 44.14	73.93	29.52 44.94	74.46	30.58 43.35	73.93
Difference2	19.68	19.68	18.61	18.61	20.21	20.21
Difference3	6.382	6.38	6.12	6.12	5.85	5.85
Difference4	0	0	0.79	0.79	0	0
Total(%)	100		100		100	

TABLE II
Handwriting Mark Accuracy using CNN feature

	Euclidean(%)		L1(%)		Cosine(%)	
Difference0 Difference1	30.05 43.88	73.93	28.45 48.93	77.39	25 46.01	71.01
Difference2	21.54	25.54	19.14	19.14	21.27	21.27
Difference3	3.72	3.72	3.457	3.45	7.71	7.71
Difference4	0.79	0.79	0.0	0.0	0.0	0.0
Total(%)	100		100		100	

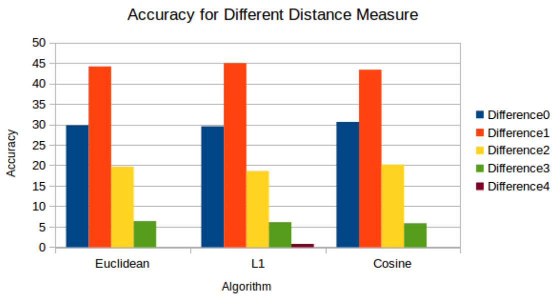

Fig. 4. Score prediction performance using different distance measures with the HOG feature

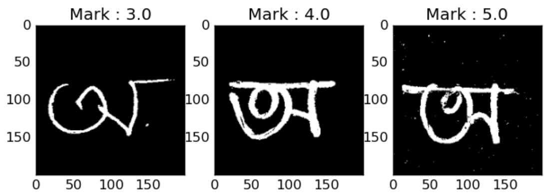

Fig. 5. Sample query image (left), suggested improved version (right)

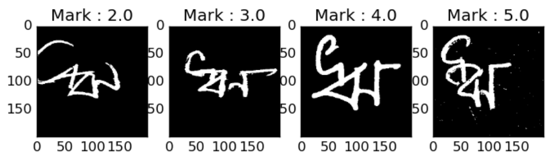

Fig. 6. Sample query image (left), suggested improved version (right)

A. Improvement Suggestion

As mentioned previously, the suggested improvement for the proposed system is a handwriting sample with a score higher than the one received by the query image, where there suggested image is most similar to the query. This method has been proposed to facilitate a scenario where the user needs to modify his/her handwriting minimally to achieve a better output.

Figures 5 - 7 shows three examples of the suggestions made by the proposed improvement suggestion method. In all three cases, the suggested improvement closely resembles the query image, in terms of shape and dimension, with the example shown in Figure 6 requiring the greatest amount of change from the query image. The other two (Figures 5 and 7) shows suggestions which are easily achievable through small tweaks to the original query images. A subjective comparison with the ideal characters show that in all three cases, the suggestion actually brings the query closer to the ideal case.

Figure 8 shows the suggestion made for a Bangla compound character. It is debatable whether the shape suggested in the suggestion is actually more aesthetically pleasing. However, the two different characters making up the compound character are more easily distinguishable in the suggestion, compared

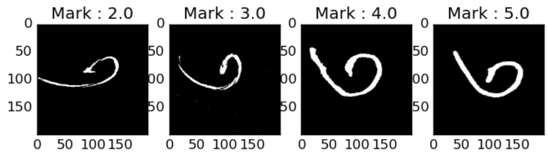

Fig. 7. Sample query image (left), suggested improved version (right)

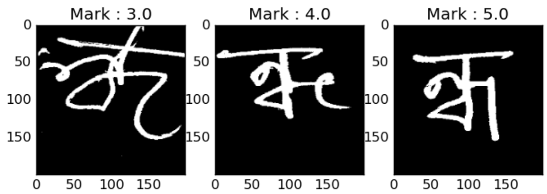

Fig. 8. Sample query image (left), suggested improved version (right)

to the original query image. Also, it is interesting to note that the suggested image also has the elongated matra (top line), similar to the query image, again demonstrating that the proposed method can suggest improvements which require little adjustments on the part of the user.

V. Conclusion

In this paper we proposed an application of image retrieval techniques to the task of evaluating the quality of handwritten Bangla isolated characters and suggesting improvements. We collected handwriting samples of 84 isolated Bangla characters from 2000 individuals.Each form of 84 characters were assessed by 5 handwriting experts, who evaluated the quality in a scale of 0 to 5, with 5 being the best. The isolated characters were extracted after the samples were digitised. For each character we separated 400 images as test cases and the remaining 1600 formed the database. As for our evaluation criteria, we checked whether the score ascribed to a test sample matches the expert-given score exactly, or had a score difference of 1. In these situations, we accepted the scoring to be accurate. Score difference of 2 or more were judged to be inaccurate. Experimental results show that our proposed method, which is meant to be an automatic mimicry the method proposed by Marlow[6], achieved an accuracy of 77.39% when using features extracted from an autoencoder coupled with the L1 distance measure. Using the same arrangement, the suggested improved versions of the test characters also demonstrate that the proposed method can give suggestions which will required small changes to the original test characters.

This is the first of a series of planned experiments and methods relating to this application area. Future work will include expanding the dataset, more fine-grained expert evaluation, classification-based approaches to scoring handwriting quality and suggesting improvements and finally incorporating.

Acknowledgment

This project is supported by the ICT Division, Ministry of ICT, Government of the Peoples Republic of Bangladesh (Project ID: 56.00.0000.028.33.066.16-731)

References

[1] Bhattacharya, U., Chaudhuri, B.B.: Handwritten nu- meral databases of indian scripts and multistage recog- nition of mixed numerals. IEEE transactions on pattern analysis and machine intelligence 31(3), 444457 (2009)
[2] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05), vol. 1, pp. 886893. IEEE (2005)
[3] Deng, L.: The mnist database of handwritten digit images for machine learning research. IEEE Signal Processing Magazine 29(6), 141142 (2012)
[4] D eniz, O., Bueno, G., Salido, J., De la Torre, F.: Face recognition using histograms of oriented gradients. Pat- tern Recognition Letters 32(12), 15981603 (2011)
[5] Ediger, M.: Teaching reading successfully. Discovery Publishing House (2000)
[6] Ediger, M.: Assessing handwriting achievement. Reading Improvement 39(3), 103 (2002)
[7] Hughes, D.C., Keeling, B., Tuck, B.F.: Effects of achieve- ment expectations and handwriting quality on scoring es- says. Journal of Educational Measurement 20(1), 6570 (1983)
[8] LeCun, Y., Bengio, Y.: Convolutional networks for im- ages, speech, and time-series. In: MIT Press (1995)
[9] Masci, J., Meier, U., Cire san, D., Schmidhuber, J.: Stacked convolutional auto-encoders for hierarchical fea- ture extraction. In: International Conference on Artificial Neural Networks, pp. 5259. Springer (2011)
[10] Muller, H., Muller, W., Squire, D.M., Marchand-Maillet, S., Pun, T.: Performance evaluation in content-based image retrieval: overview and proposals. Pattern Recognition Letters 22(5), 593601 (2001)
[11] Newell, A.J., Griffin, L.D.: Multiscale histogram of ori- ented gradient descriptors for robust character recogni- tion. In: 2011 International Conference on Document Analysis and Recognition, pp. 10851089. IEEE (2011)
[12] Phelps, J., Stempel, L., Speck, G.: The childrens hand- writing scale: A new diagnostic tool. The Journal of Ed- ucational Research 79(1), 4650 (1985)
[13] Roy, R.: Cmaterdb 1.1. 1. preparation and ground truthing of handwritten bangla text line database. Ph.D. thesis (2010)
[14] Scherer, D., Muller, A., Behnke, S.: Evaluation of pooling operations in convolutional architectures for object recog- nition. In: International Conference on Artificial Neural Networks, pp. 92101. Springer (2010)
[15] Veltkamp, R., Burkhardt, H., Kriegel, H.P.: State-of-the-art in content-based image and video retrieval, vol. 22.Springer Science & Business Media (2013)
[16] Young, R.A., Rose, R.V., Nelson, R.: Teaching fluent handwriting remediates many reading-related learning disabilities. Creative Education 6(16), 1752 (2015)
[17] Zeiler, M.D.: Adadelta: an adaptive learning rate method. arXiv preprint arXiv:1212.5701 (2012)
[18] Biswas, M., Islam, R.I., Shom, G.K., Shopon, M., Mohammed, N., Momen, S., Abedin, A.: BanglaLekha-Isolated: A multi-purpose comprehensive dataset of Handwritten Bangla Isolated characters. Data In Brief. doi: http://dx.doi.org/10.1016/j.dib.2017.03.035