

REFERENCES

- [1] U. N. Strategy, P. Of, A. On, and H. Speech, “United nations strategy and plan of action on hate speech,” no. May, pp. 1–5, 2019.
- [2] M. Westerdal, A. Rights, and I. Copyright, *Confronting Accountability for Hate Speech in Sri Lanka*, vol. 73. 2010.
- [3] University of Alabama at Birmingham, “Social Media Risks,” *UAB Collat School of Business*, 2022. [Online]. Available: <https://businessdegrees.uab.edu/blog/social-media-risks/#:~:text=Risks of Personal Usage of Social Media&text=Cyberbullying,Social Information overload.>
- [4] Mazin Hussain, “The deadly Facebook cocktail: Hate Speech & Misinformation in Sri Lanka By Mazin Hussain -May 14, 2019,” *README*, 2019. [Online]. Available: <https://www.readme.lk/facebook-hate-speech-misinformation-sri-lanka/>. [Accessed: 30-Jul-2019].
- [5] X. Zhou *et al.*, “EAST: An efficient and accurate scene text detector,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2642–2651, 2017, doi: 10.1109/CVPR.2017.283.
- [6] S. V. Seeri, J. D. Pujari, and P. S. Hiremath, “Text Localization and Character Extraction in Natural Scene Images using Contourlet Transform and SVM Classifier,” *Int. J. Image, Graph. Signal Process.*, vol. 8, no. 5, pp. 36–42, 2016, doi: 10.5815/ijigsp.2016.05.02.
- [7] Chucai Yi and YingLi Tian, “Text String Detection From Natural Scenes by Structure-Based Partition and Grouping,” *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2594–2605, 2011, doi: 10.1109/tip.2011.2126586.
- [8] V. Rajan and S. Raj, “Text detection and character extraction in natural scene images using fractional poisson model,” *Proc. Int. Conf. Comput. Methodol. Commun. ICCMC 2017*, vol. 2018-Janua, no. Iccmc, pp. 1136–1141, 2018, doi: 10.1109/ICCMC.2017.8282651.
- [9] R. Ranjan, P. Venugopal, P. S, and P. N S, “Text Extraction from Images with Complex Background,” *Commun. Comput. Inf. Sci.*, vol. 815, no. 21, pp. 3–11,

2018, doi: 10.1007/978-981-10-8108-8_1.

- [10] C. Silva and C. Kariyawasam, "Segmenting Sinhala Handwritten Characters," *Int. J. Conceptions Comput. Inf. Technol.*, vol. 2, no. 4, pp. 2345–9808, 2014.
- [11] U. K. A. U. R. K.A.K.N.D. Dharmapala, W.P.M.V. Wijesooriya, C.P. Chandrasekara, L. Ranathunga, "Sinhala Handwriting Recognition Mechanism Using Zone Based Feature Extraction," *ITRU Res. Symp. 2015 Fac. Inf. Technol. Univ. MORATUWA, SRI LANKA*, pp. 10–15, 2015.
- [12] K. S. A. Walawage and L. Ranathunga, "Segmentation of Overlapping and Touching Sinhala Handwritten Characters," *2018 3rd Int. Conf. Inf. Technol. Res.*, pp. 1–6, 2018, doi: 10.1109/ICITR.2018.8736129.
- [13] A. M. Zeki, "The segmentation problem in arabic character recognition the state of the art," *Proc. 1st Int. Conf. Inf. Commun. Technol. ICICT 2005*, vol. 2005, pp. 11–26, 2005, doi: 10.1109/ICICT.2005.1598538.
- [14] U. Pal and S. Datta, "Segmentation of Bangla Unconstrained Handwritten Text," *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 2003-Janua, no. Icdar, pp. 1128–1132, 2003, doi: 10.1109/ICDAR.2003.1227832.
- [15] N. Priyanka, S. Pal, and R. Mandal, "Line and Word Segmentation Approach for Printed Documents," *IJCA Spec. Issue Recent Trends Image Process. Pattern Recognition-RTIPPR*, pp. 30–36, 2010.
- [16] U. Pal, A. Belaid, and C. Choisy, "Water Reservoir Based Approach for Touching Numeral Segmentation," *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 2001-Janua, pp. 892–896, 2001, doi: 10.1109/ICDAR.2001.953915.
- [17] M. Kumar, M. K. Jindal, and R. K. Sharma, "Segmentation of Isolated and Touching Characters in Offline Handwritten Gurmukhi Script Recognition," *Int. J. Inf. Technol. Comput. Sci.*, vol. 6, no. 2, pp. 58–63, 2014, doi: 10.5815/ijitcs.2014.02.08.
- [18] B. Thakral and M. Kumar, "Devanagari handwritten text segmentation for overlapping and conjunct characters- A proficient technique," *Proc. - 2014 3rd Int. Conf. Reliab. Infocom Technol. Optim. Trends Futur. Dir. ICRITO 2014*,

2015, doi: 10.1109/ICRITO.2014.7014746.

- [19] F. Kurniawan, M. S. M. Rahim, D. Daman, A. Rehman, D. Mohamad, and S. M. Shamsuddin, "Region-based touched character segmentation in handwritten words," *Int. J. Innov. Comput. Inf. Control*, vol. 7, no. 6, pp. 3107–3120, 2011.
- [20] B. M. and J. E.-S. Ahmad Droby, Berat Kurar Barakat, Raid Saabni, Reem Alaasam, "Understanding Unsupervised Deep Learning for Text Line Segmentation," *Appl. Sci.*, 2022.
- [21] I. Rabaev and M. Litvak, *Recent advances in text line segmentation and baseline detection in historical document images : a systematic review*. Springer Berlin Heidelberg, 2025.
- [22] R. Ptak, B. Zygałło, and O. Unold, "Projection-based text line segmentation with a variable threshold," *Int. J. Appl. Math. Comput. Sci.*, vol. 27, no. 1, pp. 195–206, 2017, doi: 10.1515/amcs-2017-0014.
- [23] S. Chanda, S. Pal, and U. Pal, "Word-wise Sinhala Tamil and English script identification using Gaussian kernel SVM," *Proc. - Int. Conf. Pattern Recognit.*, no. February 2015, 2008, doi: 10.1109/icpr.2008.4761823.
- [24] S. Gunawardhana and L. Ranathunga, "Segmentation and identification of presence of Sinhala characters in Facebook Images," *2018 13th Int. Conf. Ind. Inf. Syst. ICIIS 2018 - Proc.*, no. 978, pp. 77–82, 2018, doi: 10.1109/ICIINFS.2018.8721400.
- [25] Z. Zhang, H. Mamat, X. Xu, and A. Aysa, "FAS-Res2net : An Improved Res2net-Based Script Identification Method for Natural Scenes," *Proc. Int. Conf. Pattern Recognit. Artif. Intell.*, 2023.
- [26] U. P. and T. M. Shivakumara Palaiahnakote, "A New Hybrid Wavelet Decomposition-based Networks for Script Identification in Scene Images," *Artif. Intell. Appl.*, vol. 00, no. June 2024, pp. 1–8, 2025, doi: 10.47852/bonviewAIA52023569.
- [27] V. Naosekpam and N. Sahu, "A Hybrid Scene Text Script Identification Network for Regional Indian Languages," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 23, no. 8, 2024, doi: 10.1145/3649439.

- [28] P. T. C. Madushanka, R. Bandara, and L. Ranathunga, "Sinhala handwritten character recognition by using enhanced thinning and curvature histogram based method," *2017 IEEE 2nd Int. Conf. Signal Image Process. ICSIP 2017*, vol. 2017-Janua, pp. 46–50, 2017, doi: 10.1109/SIPROCESS.2017.8124503.
- [29] G. I. Gunarathna, M. A. P. Chamikara, and R. G. Ragel, "A fuzzy based model to identify printed Sinhala characters," *2014 7th Int. Conf. Inf. Autom. Sustain. "Sharpening Futur. with Sustain. Technol. ICIAfS 2014*, 2014, doi: 10.1109/ICIAFS.2014.7069553.
- [30] M. Kumar, M. K. Jindal, R. K. Sharma, and S. R. Jindal, "Offline handwritten pre-segmented character recognition of gurmukhi script," *Mach. Graph. Vis.*, vol. 25, no. 1–4, pp. 45–55, 2016, doi: 10.22630/mgv.2016.25.1.5.
- [31] T. N. Kumara and R. Ragel, "A systematic feature selection process for a Sinhala character recognition system," *2016 IEEE Int. Conf. Inf. Autom. Sustain. Interoper. Sustain. Smart Syst. Next Gener. ICIAfS 2016*, pp. 1–6, 2016, doi: 10.1109/ICIAFS.2016.7946523.
- [32] M. M. Rahman, M. A. H. Akhand, S. Islam, P. Chandra Shill, and M. M. Hafizur Rahman, "Bangla Handwritten Character Recognition using Convolutional Neural Network," *Int. J. Image, Graph. Signal Process.*, vol. 7, no. 8, pp. 42–49, 2015, doi: 10.5815/ijigsp.2015.08.05.
- [33] J. Mariyathas, V. Shanmuganathan, and B. Kuhaneswaran, "Sinhala Handwritten Character Recognition using Convolutional Neural Network," *Proc. ICITR 2020 - 5th Int. Conf. Inf. Technol. Res. Towar. New Digit. Enlight.*, pp. 41–45, 2020, doi: 10.1109/ICITR51448.2020.9310914.
- [34] M. A. Wibowo, M. Soleh, W. Pradani, A. N. Hidayanto, and A. M. Arymurthy, "Handwritten javanese character recognition using descriminative deep learning technique," in *Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017*, 2018, vol. 2018-Janua, pp. 325–330, doi: 10.1109/ICITISEE.2017.8285521.
- [35] F. P. Such, D. Peri, F. Brockler, H. Paul, and R. Ptucha, "Fully Convolutional Networks for Handwriting Recognition," *Proc. Int. Conf. Front. Handwrit.*

Recognition, ICFHR, vol. 2018-Augus, pp. 86–91, 2018, doi: 10.1109/ICFHR-2018.2018.00024.

- [36] C. Wigington, S. Stewart, B. Davis, B. Barrett, B. Price, and S. Cohen, “Data Augmentation for Recognition of Handwritten Words and Lines using a CNN-LSTM Network,” *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 1, pp. 639–645, 2017, doi: 10.1109/ICDAR.2017.110.
- [37] R. R. Nair, N. Sankaran, B. U. Kota, S. Tulyakov, S. Setlur, and V. Govindaraju, “Knowledge transfer using Neural network based approach for handwritten text recognition,” *Proc. - 13th IAPR Int. Work. Doc. Anal. Syst. DAS 2018*, pp. 441–446, 2018, doi: 10.1109/DAS.2018.75.
- [38] Wikipedia, “Tesseract (software),” 2020. [Online]. Available: [https://en.wikipedia.org/wiki/Tesseract_\(software\)](https://en.wikipedia.org/wiki/Tesseract_(software)). [Accessed: 20-Feb-2020].
- [39] H. Sharifi and A. Shahbahrami, “A comparative study on different license plate recognition algorithms,” *Commun. Comput. Inf. Sci.*, vol. 167 CCIS, no. PART 2, pp. 686–691, 2011, doi: 10.1007/978-3-642-22027-2_57.
- [40] S. Rice, F. Jenkins, and T. Nartker, “The fourth annual test of OCR accuracy,” *1995 Annu. Rep. ISRI, ...*, vol. 1, no. April, pp. 1–39, 1995.
- [41] M. A. Hasnat, M. R. Chowdhury, and M. Khan, “An open source Tesseract based optical character recognizer for Bangla script,” *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, pp. 671–675, 2009, doi: 10.1109/ICDAR.2009.62.
- [42] A. Hasnat, M. R. Chowdhury, and M. Khan, “Integrating Bangla Script Recognition Support in Tesseract OCR,” pp. 108–112, 2009.
- [43] S. Rakshit, A. Kundu, M. Maity, S. Mandal, S. Sarkar, and S. Basu, “Recognition of handwritten Roman Numerals using Tesseract open source OCR engine,” pp. 1–5, 2010.
- [44] Q. Li, W. An, A. Zhou, and L. Ma, “Recognition of offline handwritten Chinese characters using the tesseract open source OCR engine,” *Proc. - 2016 8th Int. Conf. Intell. Human-Machine Syst. Cybern. IHMSC 2016*, vol. 2, pp. 452–456, 2016, doi: 10.1109/IHMSC.2016.239.
- [45] C. Vasantharajan, L. Tharmalingam, and U. Thayasivam, “Adapting the

Tesseract Open-Source OCR Engine for Tamil and Sinhala Legacy Fonts and Creating a Parallel Corpus for Tamil-Sinhala-English,” *Int. Conf. Asian Lang. Process.*, pp. 143–149, 2022.

- [46] T. Enstad, T. Trosterud, M. I. Røsok, Y. Beyer, and M. Roald, “Comparative analysis of optical character recognition methods for Sámi texts from the National Library of Norway,” *Proc. NoDaLiDa/Baltic-HLT 2025*, pp. 98–108, 2025.
- [47] doxygen, “Feature Matching,” *doxyden*. [Online]. Available: https://docs.opencv.org/4.x/dc/dc3/tutorial_py_matcher.html. [Accessed: 10-Mar-2022].
- [48] doxygen, “ORB (Oriented FAST and Rotated BRIEF).” [Online]. Available: https://docs.opencv.org/3.4/d1/d89/tutorial_py_orb.html. [Accessed: 10-Mar-2022].
- [49] doxygen, “Introduction to SIFT (Scale-Invariant Feature Transform).” [Online]. Available: https://docs.opencv.org/4.x/da/df5/tutorial_py_sift_intro.html. [Accessed: 10-Mar-2022].
- [50] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, “Image Matching from Handcrafted to Deep Features: A Survey,” *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 23–79, 2021, doi: 10.1007/s11263-020-01359-2.
- [51] Y. Jin *et al.*, “Image Matching Across Wide Baselines: From Paper to Practice,” *Int. J. Comput. Vis.*, vol. 129, no. 2, pp. 517–547, 2021, doi: 10.1007/s11263-020-01385-0.
- [52] T. E. de Campos, B. R. Babu, and M. Varma, “Character Recognition in Natural Images,” *Proc. Int. Conf. Comput. Vis. Theory Appl. (VISAPP), Lisbon, Port.*, doi: 10.1109/ICICT.2015.7469575.
- [53] Doxygen, “Morphological Transformations,” *OpenCV*, 2019. [Online]. Available: https://docs.opencv.org/master/d9/d61/tutorial_py_morphological_ops.html. [Accessed: 25-Aug-2019].
- [54] doxygen, “Image Thesholding,” *OpenCV*, 2019. [Online]. Available:

- https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html. [Accessed: 05-Aug-2019].
- [55] OpenCV, “Contours : Getting Started,” *OpenCV*, 2020. [Online]. Available: https://docs.opencv.org/master/d4/d73/tutorial_py_contours_begin.html. [Accessed: 05-Jan-2020].
- [56] A. Antonacopoulos and D. Karatzas, “Document Image Analysis for World War II Personal Records,” *Proc. - First Int. Work. Doc. Image Anal. Libr. - DIAL* 2004, no. February, pp. 336–341, 2004, doi: 10.1109/DIAL.2004.1263263.
- [57] Wikipedia, “Gaussian blur,” *Wikipedia*, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Gaussian_blur. [Accessed: 05-Aug-2019].
- [58] Harald Scheidl, “WordSegmentation,” 2020 *GitHub, Inc.*, 2020. [Online]. Available: <https://github.com/githubharald/WordSegmentation>. [Accessed: 23-Apr-2020].
- [59] doxygen, “Image Filtering,” 2021. [Online]. Available: https://docs.opencv.org/4.x/d4/d86/group__imgproc__filter.html.
- [60] K. S. A. Walawage and L. Ranathunga, “Devising a Distinguishable Feature Set for Sinhala and English Script Separation on Social Media Images,” 2020 *From Innov. to Impact*, 2020, doi: 10.1109/FITI52050.2020.9424904.
- [61] OpenCV, “Adding borders to your images,” *OpenCV*, 2020. [Online]. Available: https://docs.opencv.org/3.4/dc/da3/tutorial_copyMakeBorder.html. [Accessed: 30-Apr-2020].
- [62] TensorFlow, “An end-to-end platform for machine learning,” *TensorFlow*, 2020. [Online]. Available: <https://www.tensorflow.org/>. [Accessed: 05-Sep-2020].
- [63] B. and Sphinx, “Skeletonize,” *The scikit-image development team*, 2021. [Online]. Available: https://scikit-image.org/docs/dev/auto_examples/edges/plot_skeleton.html. [Accessed: 10-Aug-2021].
- [64] “Sinhala English OCR Web Application,” 2023. [Online]. Available:

<http://192.248.12.8:5500/>. [Accessed: 12-Dec-2023].

- [65] “Sinhala English Handwritten Character Recognition Web Application,” 2023. [Online]. Available: <http://192.248.12.8:8000/>. [Accessed: 12-Dec-2023].
- [66] D. G. Low, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, pp. 91–110, 2004.
- [67] TensorFlow, “Convolutional Neural Network (CNN),” *TensorFlow*, 2020. [Online]. Available: <https://www.tensorflow.org/tutorials/images/cnn>. [Accessed: 05-Sep-2020].
- [68] Google, “Detect text in images,” *Cloud, Google*, 2020. [Online]. Available: <https://cloud.google.com/vision/docs/ocr>. [Accessed: 05-Feb-2020].