Camila Lightfoot

CS 480 State of the Art Project Proposal: Evaluating Facial Recognition Models

Members

- Camila Lightfoot

Introduction

Facial recognition is one of the most important and rapidly advancing application areas in modern Artificial Intelligence (AI), with real-world usage in biometric authentication, surveillance systems, automated access control, border security, and consumer technologies, such as Face ID and smart cameras. However, facial recognition performance heavily depends on a variety of factors, such as image quality, lighting conditions, pose variations, occlusions and demographic diversity. In the last decade, researchers have developed increasingly sophisticated deep-learning methods to improve the accuracy and precision of facial recognition software, while addressing the challenges posed by the aforementioned performance factors.

Problem Description

This project explores state of the art facial recognition models and compares their performance on publicly available benchmark datasets. Specifically, we are going to analyze and evaluate three influential models: FaceNet [1], ArcFace [2], and AdaFace [3]. These models represent three generations of innovation.

FaceNet, developed by Google, introduced the concept of learning facial embeddings through a triplet loss mechanism that encourages large separation between embeddings of different individuals while bringing embeddings of the same individual closer together. Although groundbreaking at the time and influential on subsequent methods, FaceNet exhibits limitations in training stability and sensitivity to triplet selection. These limitations become even more pronounced when evaluating performance on unconstrained images. ArcFace builds upon the embedding paradigm but introduces an additive angular margin loss function that significantly improves discriminative power by maximizing the angular distance between classes. This modification enables ArcFace to achieve state-of-the-art performance on widely recognized benchmarks. More recently, AdaFace extends the ArcFace framework by incorporating a quality-adaptive margin strategy, enabling the model to dynamically adjust the contribution of individual samples based on their estimated image quality. This makes AdaFace more robust to low-resolution, blurred, or occluded faces, conditions that commonly occur in natural, uncontrolled environments.

Topic Significance

This topic is significant because there are still persistent challenges that facial recognition systems face in real-world conditions. This is because while models often perform well on curated datasets, performance can degrade significantly under practical variability, such as changes in lighting, presence of masks, aging, and poor image resolution and other aforementioned factors. Evaluating how these models handle these challenges provides critical insight into the strengths and weaknesses of current approaches and highlights areas where further research is needed. Furthermore, these models have been published in leading computer vision conferences such as the Conference on Computer Vision and Pattern Recognition (CVPR) [1] [2] [3] [6], have been trained on large-scale datasets such as MS1M [4] and VGGFace2 [5], and have demonstrated competitive placement on standardized leaderboards [6] [7] [8]. Their significance, therefore, aligns closely with the objectives of the CS480 project requirement to examine recent advances in AI/ML that represent state-of-the-art practice and theoretical development.

Datasets Used for Evaluation

This project will evaluate the above models using publicly available benchmark datasets. These include the widely used Labeled Faces in the Wild (LFW) dataset [7], which supports standard face verification protocols, and datasets such as CFP-FP [8] and AgeDB-30 [9], which provide controlled tests for robustness to pose variations and age progression. We will also use the CelebA dataset [10], which contains attribute-rich annotations that allow deeper analysis of model behavior under specific conditions, such as partial occlusion or lighting variation. All datasets are openly accessible and have well-established splits and verification protocols, ensuring that the evaluation can be conducted in a systematic and reproducible manner.

Evaluation Metrics

We will evaluate the models by first preprocessing the datasets using established face detection and alignment tools, such as MTCNN [11] or RetinaFace [12], followed by extraction of face embeddings using each of the selected models. Verification performance will be assessed by computing cosine similarity or Euclidean distance between pairs of embeddings; classification accuracy will be determined using standard thresholding approaches. To evaluate the models' abilities more comprehensively, the project will compute receiver operating characteristic (ROC) curves, estimate the area under the curve (AUC), and calculate the Equal Error Rate (EER). These standard benchmarks should determine how well each model generalizes images from the aforementioned datasets.

Expected Outcomes

The expected outcome of this project is a multidimensional performance comparison that clearly demonstrates the advantages and limitations of each model. We expect ArcFace will achieve the highest performance on standard datasets due to its strong angular margin optimization, while AdaFace will outperform both ArcFace and FaceNet under conditions of decreased image quality due to its adaptive margin mechanism. We expect FaceNet to yield the lowest accuracy and weakest robustness among the three models due to its lack of features when compared to ArcFace and AdaFace. The final report will present quantitative findings, visualizations of embedding distributions, and an analysis of error cases to illustrate where and why each model succeeds or fails. These insights will support conclusions regarding the current state of the art facial recognition models and hopefully identifies potential directions for improvement.

Works Cited

[1] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2015, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.

[2] J. Deng et al., “ArcFace: Additive angular margin loss for deep face recognition,” Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 4690–4699, doi: 10.1109/CVPR.2019.00482.

[3] M. Kim, A. K. Jain, and X. Liu, “AdaFace: Quality adaptive margin for face recognition,” Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 18750–18759, doi: 10.1109/CVPR52688.2022.01822.

[4] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, “MS-Celeb-1M: A dataset and benchmark for large-scale face recognition,” *Microsoft Research*, Aug. 2016. Available: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/08/MSCeleb-1M-a.pdf.

[5] Q. Cao, L. Shen, W. Xie, O. M. Parkhi and A. Zisserman, “VGGFace2: A dataset for recognising faces across pose and age,” *Proc. 13th IEEE Int. Conf. Automatic Face & Gesture Recognition (FG)*, Xi’an, China, May 2018, pp. 67-74, doi:10.1109/FG.2018.00020.

[6] K. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, “The MegaFace benchmark: 1 million faces for recognition at scale,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 4873–4882, doi: 10.1109/CVPR.2016.527.

[7] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments,” Univ. Massachusetts Amherst, Tech. Rep. 07-49, 2007.

[8] C. Whitelam *et al.*, “IARPA Janus Benchmark–C: Face dataset and protocol,” *Proc. IEEE Int. Conf. Biometrics Theory, Applications and Systems (BTAS)*, 2017, pp. 1–9, doi: 10.1109/BTAS.2017.8272710.

[9] S. Moschoglou et al., “AgeDB: The first manually collected, in-the-wild age database,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2017, pp. 1997–2005, doi: 10.1109/CVPRW.2017.250.

[10] Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep learning face attributes in the wild,” *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 3730–3738, doi: 10.1109/ICCV.2015.425.

[11] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” IEEE Signal Process. Lett., vol. 23, no. 10, pp. 1499–1503, 2016, doi: 10.1109/LSP.2016.2603342.

[12] J. Deng et al., “RetinaFace: Single-shot multi-level face localisation in the wild,” Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 5203–5212, doi: 10.1109/CVPR42600.2020.00525.