clarifai

Machine Learning Approaches to Face Detection and Recognition

Introduction

Face detection and recognition are two different categories of problems in computer vision applications. Face detection algorithms identify the presence of a human face in an image or a video, while face recognition confirms the identity of a person by analyzing facial features. Thus face detection is a significant first part of the face recognition process (Chi, Zhang, and Chen 2017).

Both face detection and recognition have several applications in image and video analysis ranging from security surveillance to simple unlocking your phone. The use cases include quicker airport boarding, event celebrity recognition, smarter advertising, finding the missing person, VIP recognition, and public safety, etc.
Face detection is a complex task for machines due to the variability in human faces such as the presence of glasses, the orientation of the face, presence of facial hair, differences in lighting conditions, and image resolution. Face recognition is an even more complex task because of interpreting inherent facial features, aging, occlusion, and facial expressions, etc.

Shortcomings of conventional methods used for face detection and recognition

The working of face detection algorithms follows a unified approach. The images are first filtered by applying accurate classifiers. One of the famous conventional face detection algorithms in the pre-machine learning era is Viola Jones developed by Paul Viola Michael J. Jones (Viola and Jones 2001). The algorithm can be trained to detect a variety of objects; however, it was primarily developed for face detection. The algorithm has four stages of detection i.e. Haar feature selection, creation of integral features, Ada boost training, and cascading classification. Viola Jones framework is robust with a very high true-positive rate. The other variations such as orientation, translation, and scaling are also dealt with in the face detection step (Dang and Sharma 2017). The next step is the prediction of the approximate location of facial features (eyes, nose, mouth, etc.) by using anthropometric data set based systems (Gupta, Markey, and Bovik 2010). The geometric combinations generate dedicated anchor points during face detection after which starts the actual process of face recognition. There are a plethora of conventional face recognition algorithms based on holistic features and local features. The holistic group is further divided into linear and nonlinear projection approaches. Linear approaches include independent component analysis, principal component analysis, linear regression classifier, and linear discriminant analysis. The linear approaches fail to conceive variations caused by illuminations and facial expressions. Non-linear approaches are used to deal with the aforementioned issues. Most non-linear approaches use kernel techniques. Kernel-based techniques have a strong theoretical foundation; however, they do not produce significant improvement in face recognition application when compared with linear methods (Képešiová and Kozák 2018).

Machine learning approaches to face detection and recognition

The hybrid approaches for face recognition offer the best of feature-based and holistic methods. However, the major limitation is a selection of the good features that can fully extract the useful information required for face detection. Deep learning systems provide the function to train end-to-end systems that can learn a large number of features useful for optimal face detection and recognition tasks. The researchers have proposed different variations of deep learning approaches, one such method was Deep Dense Face Detector.

This method uses a single model to detect faces in a wide range of orientations (Kumar, Kaur, and Kumar 2019).

A commonly used approach to detect the face is the Viola-Jones algorithm which is based on four steps i.e. feature selection, feature evaluation, feature learning, and cascading classifier (Filali et al. 2018). In order to identify a face, it is first necessary to specify what features of the face should be used to train the model. During training, the face portion of the image is used to extract the features. Once the features are extracted, the next step is to train the model using extracted features of previously identified faces. A support vector machine is the commonly used training tool. Two-class recognition is achieved using support vector machines which can be extended for implementing multi-class face recognition (Nasr et al. 2017).

A comparative study to evaluate the Haar-AdaBoost, Support Vector Machine with Grouped Features, Gabor filter Neural Networks, and local binary pattern-AdaBoost. It was found that Haar-AdaBoost was best in terms of detection rate and accuracy (Zhao et al. 2019). The comparison is shown in the following figure:

The comparison of the detection rate of each algorithm is shown in the following graph:

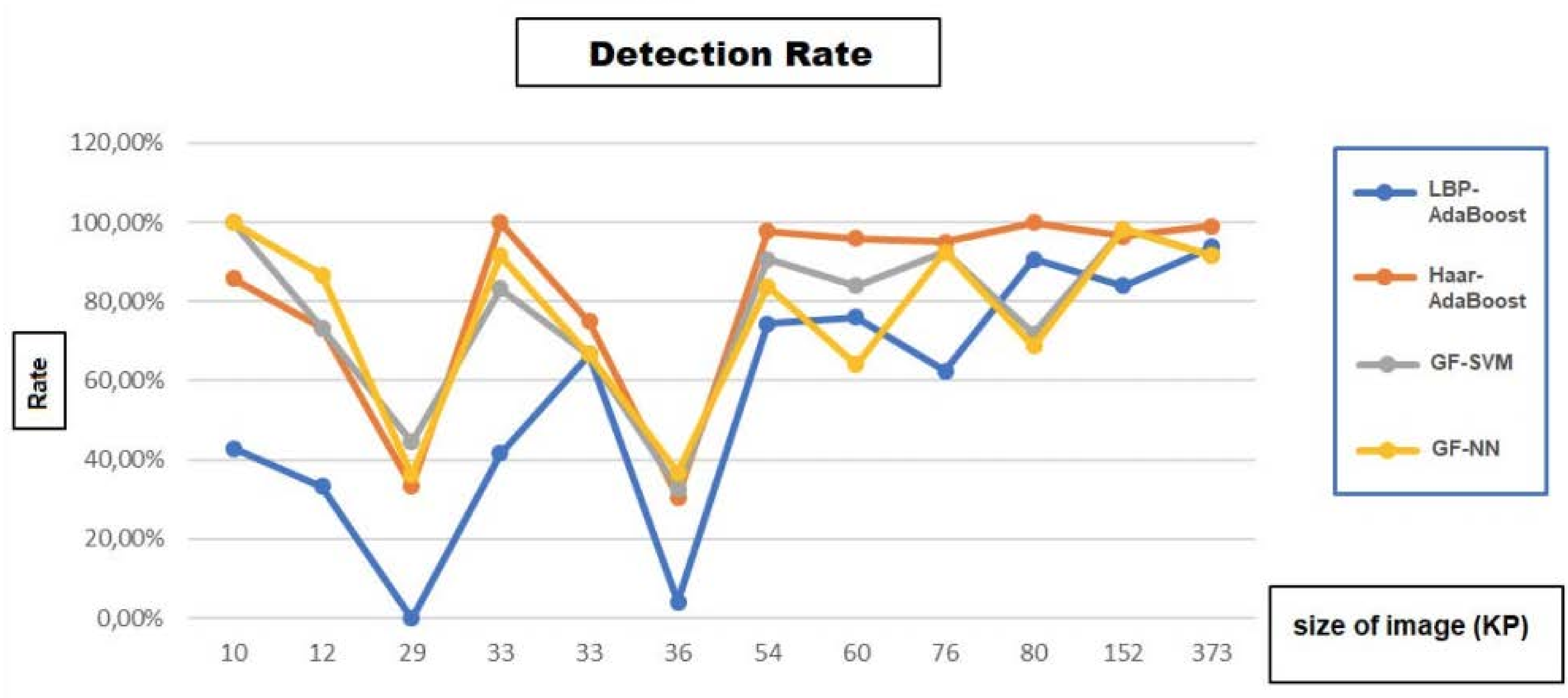

Figure 1: Comparison of different algorithms used for face detection

Machine learning approaches to face recognition

The convolutional neural networks have been found to be the most useful and accurate type of deep learning method for face recognition. The benefit of using neural networks in face recognition applications is that it can reduce the dimensionality and can be trained as a classifier. The bottleneck features extracted during training are used to detect and recognize the objects not previously seen by the machine. The convolutional neural network architectures for face recognition are inspired by the ImageNet Large Scale Visual Recognition Challenge. The common architectures used for CNN are VGGnet and GoogleNet. Both these architectures achieve the nearly same level of face recognition accuracy; however, VGGNet used 20 times more parameters than GoogleNet (Daniel Sáez Trigueros, Li Meng, and Margaret Hartnett 2018). Some researchers also proposed residual networks for face recognition applications. The concept is to learn a residual mapping using a building block as shown in the following figure:

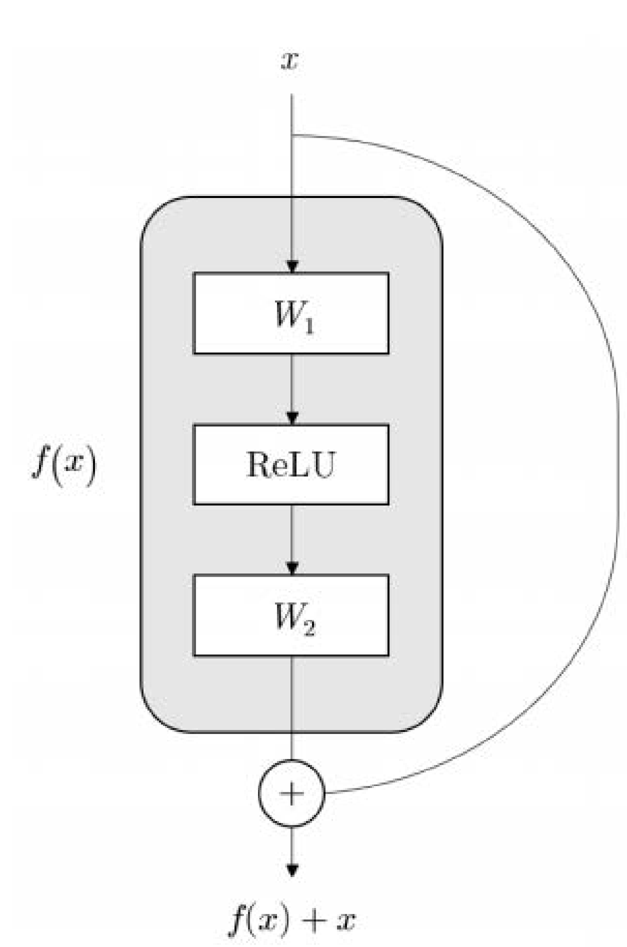

Figure 2: Residual Networks for face recognition

Most of the CNNs are unable to achieve ground-breaking results because of the low capacity of networks used and the scarcity of datasets available for training. Facebook used a CNN-based approach, named as DeepFace, for face recognition across its platform. Facebook used a high capacity model for this purpose and achieved an accuracy of above 97.35% on the Labeled Faces in the Wild benchmark. In other studies, researchers have trained the CNN using softmax loss2 with the help of 4.4 million faces. The study used an effective facial alignment system based on 3D modeling and a CNN architecture with locally connected layers. The benefit of connected layers is that it can learn different features from all regions of an image. Similarly, another approach to face recognition systems used 60 different CNNs to train on patches of ten regions, greyscale, and RGB. During testing, 160 bottleneck features were extracted from each patch.

Further research in the field of face detection and recognition is still in progress. There is a need for a large-scale open-access image database so that the newer machine learning models can be trained on it.

About Clarifai

Clarifai is the leading AI platform for computer vision, natural language processing and audio recognition. We help enterprises and public sector organizations transform unstructured images, video, text and audio data into structured data, significantly faster and more accurately than humans would be able to do on their own. Founded in 2013 by Matt Zeiler, Ph.D., Clarifai has been a market leader in computer vision AI since winning the top five places in image classification at the 2013 ImageNet Challenge. Clarifai is headquartered in Delaware with more than 100 employees and offices in New York City, San Francisco, Washington, D.C., and Tallinn, Estonia. For more information, please visit www.clarifai.com.

References:

Chi, Liying, Hongxin Zhang, and Mingxiu Chen. 2017. "End-to-End Face Detection and Recognition." *ArXiv Preprint ArXiv:1703.10818.*

Dang, Kirti, and Shanu Sharma. 2017. "Review and Comparison of Face Detection Algorithms." In 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, 629–33. IEEE.

Daniel Sáez Trigueros, Li Meng, and Margaret Hartnett. 2018. *"Face Recognition: From Traditional to Deep Learning Methods." ArXiv.*

Filali, Hajar, Jamal Riffi, Adnane Mohamed Mahraz, and Hamid Tairi. 2018. "Multiple Face Detection Based on Machine Learning." 2018 International Conference on Intelligent Systems and Computer Vision, ISCV 2018 2018-May: 1–8. https://doi.org/10.1109/ISACV.2018.8354058.

Gupta, Shalini, Mia K. Markey, and Alan C. Bovik. 2010. "Anthropometric 3D Face Recognition." International Journal of Computer Vision 90 (3): 331–49. https://doi.org/10.1007/s11263-010-0360-8.

Képešiová, Zuzana, and Štefan Kozák. 2018. "An Effective Face Detection Algorithm." *In 2018 Cybernetics & Informatics (K&I), 1–6. IEEE.*

Kumar, Ashu, Amandeep Kaur, and Munish Kumar. 2019. "Face Detection Techniques: A Review." Artificial Intelligence Review 52 (2): 927–48. https://doi.org/10.1007/s10462-018-9650-2.

Nasr, Salah, Kais Bouallegue, Muhammad Shoaib, and Hassen Mekki. 2017. "Face Recognition System Using Bag of Features and Multi-Class SVM for Robot Applications." *In 2017 International Conference on Control, Automation and Diagnosis (ICCAD), 263–68. IEEE.*

Viola, Paul, and Michael Jones. 2001. "Robust Real-Time Object Detection." Robust Real-Time Object Detection.

Zhao, Zhong-Qiu, Peng Zheng, Shou-tao Xu, and Xindong Wu. 2019. "Object Detection with Deep Learning: A Review." *IEEE Transactions on Neural Networks and Learning Systems 30 (11): 3212–32.*