Comparing Deep Learning Techniques against Simple Machine Learning Algorithms for Image Classification

David Tang*, Waley Chen*, Karmveer Sidhu*, Mohammad Noaeen*, Abdullah Sarhan†,
Zahra Shakeri Hossein Abad‡
*Department of Electrical and Computer Engineering, University of Calgary, Canada
{david.tang, wachen, ksidhu, mohammad.noaeen} @ucalgary.ca
† Department of Computer Science, University of Calgary, Canada, asarhan@ucalgary.ca
‡ Department of Community Health Sciences, University of Calgary, Canada, zshakeri@ucalgary.ca

***Abstract*—Advancements in machine learning, particularly deep learning has paved the way for widespread adoption of facial recognition technology. However, it is still difficult to estimate age and gender from images due to high variance in facial features and image quality. The increased performance in predicting gender in images is due to leaps in deep learning methods using Convolutional Neural Networks (CNNs). Large technology companies such as Amazon have released their own image processing tools that employ pre-trained deep learning models. This paper utilizes a large labeled data set of facial images scraped from the Internet Movie Database (IMDb) to determine the effectiveness of Amazon Rekognition (AR). Specifically, it aims to determine the accuracy of predicting gender while using it's APIs. Furthermore, the CNN was compared against easier to implement algorithms such as a Support Vector Machine (SVM) and Logistic Regression (LR) and the predictive accuracy's were approximately 90%, 77%, and 80% respectively. Various image pre-processing and feature extraction techniques were used to increase prediction accuracy and increased the LR model accuracy from 60% to 80%. Pre-processing and feature extraction brought these algorithms within 10% of Amazon's CNN. However, AR maintained high accuracy with low quality images. Overall, Deep Learning algorithms are ideal choices for image classification.**

***Index Terms*—Gender classification, sex classification, image processing, machine learning, Amazon Rekognition, support vector machine, logistic regression, feature extraction**

I. Introduction

Facial recognition and analysis are some of the most important topics in computer vision. Automatic analysis of facial images have become increasingly important and are being used in areas like social media, video surveillance, business intelligence, security, and autonomous vehicles. However, gender recognition are challenging problems within the field of computer vision as they rely on interpretation of facial features [1]. Many facial images exhibit high variance even within the same age range. These variances include personal traits of each individual, nature of aging process, different race and genders, different pose variations, bad illuminations, effect of cosmetics, haircuts, and image quality [2]. Even humans can only give a rough estimation of someone's age by looking at their face. Due to these image variations, it is hard to gather complete and sufficient data to train accurate models.

Within the last few years, Convolutional Neural Networks (CNNs) within the field of deep learning have proven to be effective for computer vision [3]. CNNs are widely used for both gender recognition [4], age estimation [5] and have significantly improved object detection and recognition [6], and image captioning [7]. These CNNs can be trained through the use of TensorFlow and Keras Application Programming Interfaces (API). TensorFlow is a library that is typically employed in Python and used for machine learning applications [8]. Keras is a popular high-level neural network API that runs on top of TensorFlow [9].

Amazon also offer image analysis through their own proprietary model: Rekognition. Their model is pre-trained and offer label detection and image attributes [10]. In this paper, an analysis was conducted using their APIs to determine the accuracy of these models for gender recognition. The main contributions of this work will be as follows:

- Use Amazon Rekognition (AR) APIs for gender prediction and determine the effect of image quality on the CNN
- Compare the performance of this model with a Support Vector Machine (SVM) and Logistic Regression (LR) to determine which model predicted gender more accurately and how comparable they are
- The impact of various pre-processing and feature extraction techniques on LR model accuracy

The remainder of the sections are structured as follows: II covers similar work done in this field. III outlines the research questions aimed to be answered with the study and how the data will be processed and analyzed. Section IV will present the results of the questions asked in the Methodology. V will discuss the results found in the previous section, future work, and recommendations for other studies. Finally, section VI will conclude this paper.

II. Related Work

Classic age estimation methods involve a two-stage pipeline: local binary patterns are determined for feature extraction and then machine learning algorithms like Support Vector Machine are used to predict the age [4]. Early studies found that age estimation can be substantially improved by incorporating other facial traits like gender and race information; furthermore, the age estimation error can be reduced by 20% if trained separately on male and female [2].

Currently, most image recognition models rely on deep learning technology through the use of CNNs. Neural networks use algorithms that are layered next to each other and makes each algorithm contingent on the outcome of other surrounding algorithms [11]. A neural network uses convolution by merging multiple sets of information, pooling them together to create an accurate representation of an image. After pooling, the neural network uses the data to make predicts about the image [11]. CNNs incorporate the aforementioned two-stage pipeline into one step: the network extracts the best features and classifies these features into age categories or performs age regression.

The ChaLearn Looking at People Workshop in 2015 challenged teams to perform age estimation on a large data set [1]. The organizers of ChaLearn provided one of the largest data sets known to date of images with age and gender annotations. Of the top ten teams, nine used CNN models for age estimation. Rothe et al. [5] employed a CNN network that uses VGG-16 architecture (16 layered model created by Visual Geometry Group) and pre-trained on ImageNet for image classification.

VGG models have been further improved since 2015. Xing et al. [2] incorporated a newly proposed deep multi-task learning architecture which resulted in a high-accuracy race and gender classification. Rodriguez et al. [4] proposed a novel feed-forward attention mechanism that is able to discover the most information and reliable parts of a given face for improving age and gender classification.

Google Cloud Vision (GCV) and AR also employ CNNs using their own deep learning algorithms. Disney uses the GCV AutoML technology to build vision models to annotate products with Disney characters, product categories, and colors [12]. Motorola Solutions uses the video analytic features of AR to help with search of historical and real time video for persons-of-interest [10].

III. Methodology

A. Research Questions

The research was aimed at addressing the following Research Questions (RQs).

RQ1– *Can image quality affect the accuracy of AR's CNN when predicting gender?*
This question aims to explore how much image quality can be lost before Amazon's model has a significant drop in confidence in determining the correct gender from the celebrity's face.

RQ2– *How do simpler algorithms such as a SVM and LR compare to a CNN?*
Deep Learning has rightfully taken over majority of image processing in machine learning. However, their implementation is typically more complicated than simpler algorithms. By what degree do Neural Networks outclass a SVM and LR.

RQ3– *How well can various pre-processing and feature extraction techniques improve model accuracy?*
Extracting features from images can be challenging as there is no relation between how the brain discerns gender and how a computer would. Applying various techniques can better allow machines determine genders from a facial image.

B. Data Collection and Preparation

For this analysis, a large data set of facial images are used that have been scraped from the Internet Movie Database (IMDb) [13] for another study and have been since made available [5]. The data set contains 461,871 images with the following provided information:

- **dob**: Date of Birth.
- **photo_taken**: Date of Photo Taken. Used to calculate the subjects age when pictured.
- **full_path**: Path to File.
- **gender**: Gender of Subject. 0 for female, 1 for male, *NaN* for unknown.
- **name**: Name of Subject.
- **face_location**: Location of Face within Image.
- **face_score**: Score of Detector from DEX project [5]. *Inf* for no found face
- **second_face_score**: Detector Score for Secondary Face. *NaN* if only single face present.
- **celeb_names**: Subject Name.
- **celeb_id**: Subject ID.

For the purposes of this study, the only relevant information are the file path, gender, face location, and face score.

Majority of the data was handled with Apache Spark [14] as it greatly reduced image transformation and model training and testing times. Spark runs on the University of Calgary's Advanced Research Computing (ARC) cluster [15]. The exact information of the hardware used is: Lattice Partition which has 8 cores per node with each core running at 2.27 GHz and all 8 cores share 12GB of RAM. Additionally, due to the larger size of image files, they were modified and only the primary face was extracted. This reduced needed disk space and processing time throughout the entire process.

To use the same data set for varying image resolutions, Python Image Library (PIL) [16] was utilized to reduce the resolution of the images to 100x100, 80x80, and 40x40 pixels where 100x100 is the default for all tests unless specified. All tests were also carried out with grey scale images.

C. Data Analysis and Procedures

To test the Amazon CNN, their Representational State Transfer (REST) API's will be utilized which requires image bytes as base64 encoded (for the technique implemented in this study) and will return a response in Java-Script Object Notation (JSON) format with various attributes such as age, gender, facial landmarks, and various emotions. Each of these attributes (except age) will also return a confidence rating from 0.0–100.0.

The pre-processing techniques selected were histogram equalization and median filtering. Histogram equalization enhances the contrast and equalizes the image's histogram, as the name suggests [17]. Median filtering is used to reduce image noise while preserving edges [18]. For the feature extraction techniques, DAISY, canny, histogram of orientated gradients (HOG), and entropy were used. These were selected as they were utilized in similar studies [19]–[22]. Scikit-image was utilized to perform all the pre-processing and feature extraction techniques [23]. To analyze the data, the actual and predicted classes were recorded for each trial with the file titles referring to the specific techniques used.

The models selected were a radial SVM and LR. Scikit-learn was used to implement the SVM [24]. Ideally Spark would've been utilized but from preliminary testing, it was quickly determined that a linear SVM is not able to classify gender from images and switching to radial fitted the data set significantly better. The binary LR from Spark's Machine Learning Library was utilized for large scale tests [14]. LR was utilized for all tests for pre-processing and feature extraction.

Results from all models are parsed, their metrics calculated, and visualized.

IV. Results

A. Data Description

The following section presents the results from the tests performed to answer the research questions introduced in III.

Figure 1 are the results from comparing the SVM and CNN by prediction accuracy. It also contains the accuracy of the CNN for varying resolutions.

Figure 2 display the results of prediction confidence rating of the CNN at image resolutions of 100x100, 80x80, and 40x40 pixels.

Figure 3 presents the overall predictive accuracy of the Spark LR model for the original images versus the pre-processed.

Figure 4 splits the data from Figure 3 into their separate classes.

Figure 5 is the confusion matrix for the LR predictions for the original versus pre-processed images. The actual labels are represented on the left hand axis and the bottom represents the predicted.

Figures 6, 7, and 8 contain labels corresponding to Table I.

TABLE I
Abbreviations of Feature Extraction Techniques

Feature Extraction Technique	Abbreviation
Original	O
Pre-processed	P
DAISY	D
Canny	C
HOG	H
Entropy	E

Figure 6 are the overall results from the LR model for various combinations of applied feature extraction techniques.

Figure 7 are the gender separated (descending order for males) results from the LR model for various combinations of applied feature extraction techniques.

Figure 8 are the gender separated (descending order for females) results from the LR model for various combinations of applied feature extraction techniques.

Figure 9 contains the overall prediction accuracy for the training and testing splits for varying sample sizes.

Figure 10 contains the class specific recall and precision for the training and testing splits for varying sample sizes.

B. Data Visualization

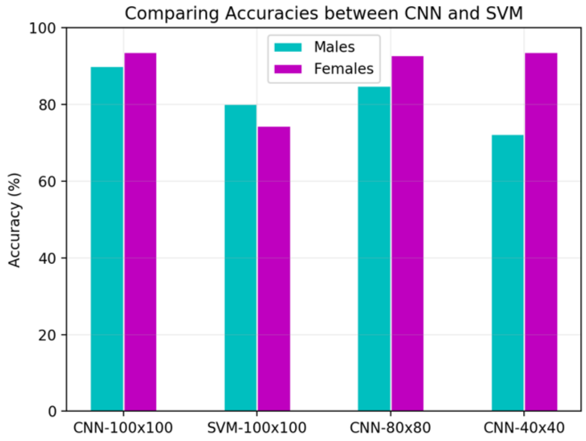

Fig. 1. Comparing Accuracy between CNN and SVM

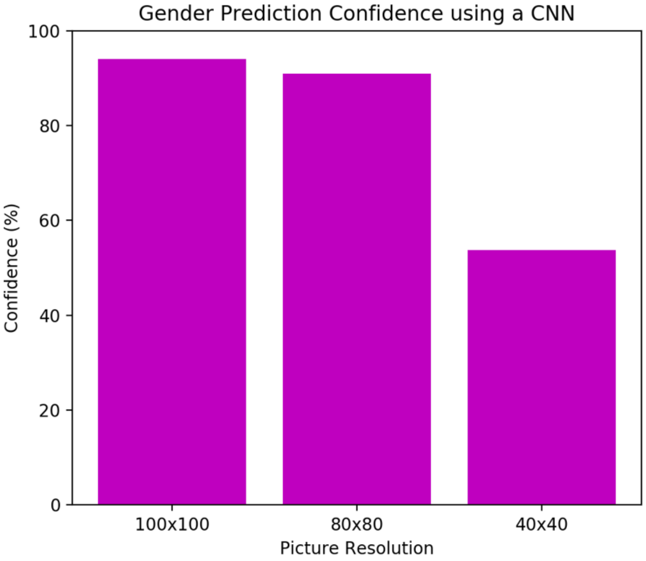

Fig. 2. CNN Gender Prediction Confidence

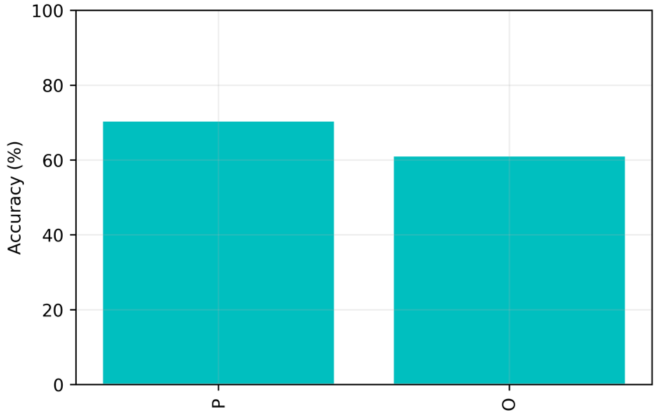

Fig. 3. Overall Prediction Accuracy of Pre-processed vs Original Images

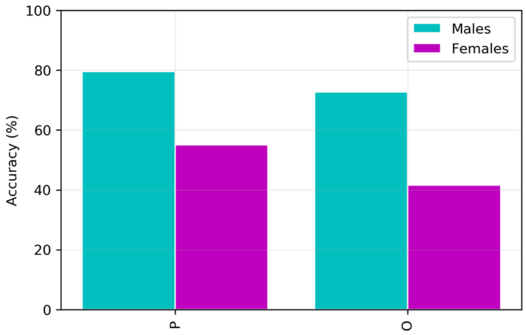

Fig. 4. Gender Specific Prediction Accuracy of Pre-processed vs Original Images

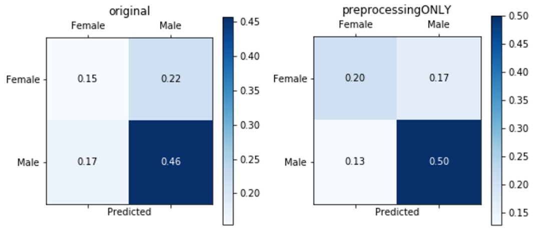

Fig. 5. Normalized Confusion Matrix for Original and Pre-processed Images

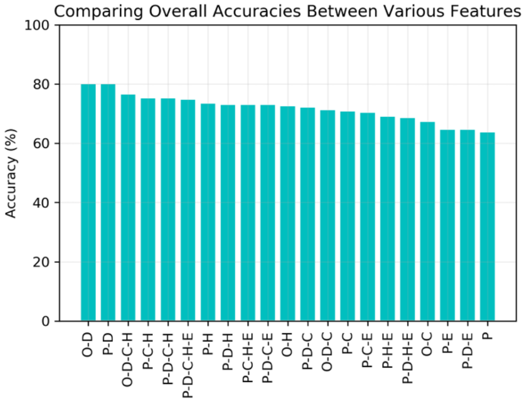

Fig. 6. Overall Predictive Accuracy for Various Feature Extractions

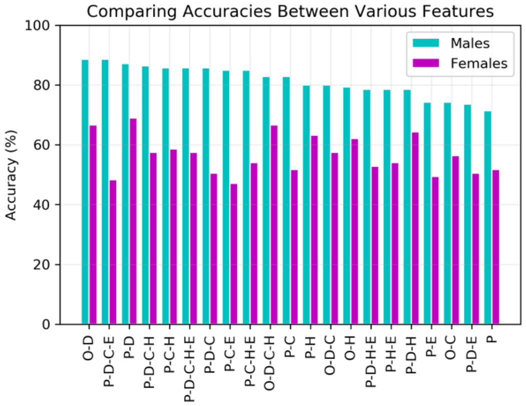

Fig. 7. Male Sorted Predictive Accuracy for Various Feature Extractions

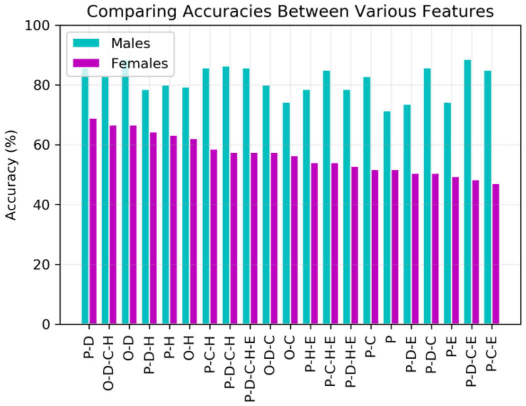

Fig. 8. Female Sorted Predictive Accuracy for Various Feature Extractions

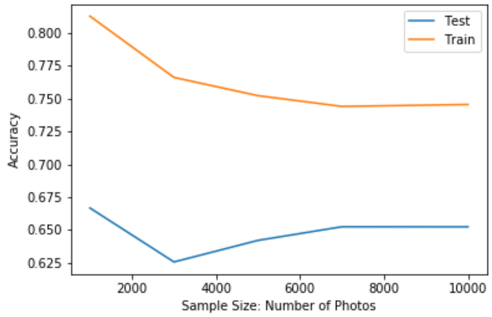

Fig. 9. Radial SVM Accuracy for Image Sample Size

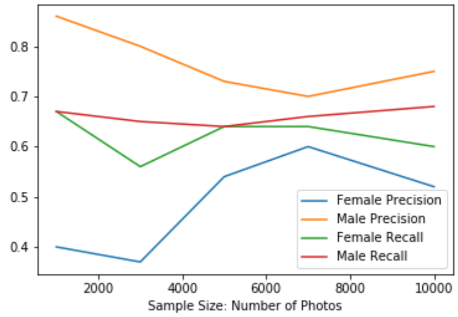

Fig. 10. Radial SVM Precision & Recall for Image Sample Size

V. Discussion

A. Key Findings

RQ1 – Based on Figure 1 it can be observed that as the resolution of the images fed to the CNN were reduced, there was a drop in overall accuracy. With the gender split plot, it can be seen that only the male accuracy dropped with respect to picture resolution. This could indicate that Amazon's CNN model has certain feature extraction dependencies for determining males that do not do a sufficient job at the reduced resolutions. When looking at Figure 2, it can be seen that at 40x40 pixels, there was a significant drop in prediction confidence from Amazon's model. Impressively, it still maintained to outperform both the SVM and LR models with a 55% prediction confidence. This low confidence was expected as Amazon outlines that there would be such a drop in confidence after 80x80 pixels [10].

RQ2 – The CNN, radial SVM, and LR accuracy was approximately 90%, 77%, and 80% respectively. The accuracy between the CNN and simpler models ranging between 10–15%. Although the error isn't that significant and would be acceptable in some situations, a huge drawback of using the SVM or LR models are testing times. CNN's typically require significant training time but are able to test very quickly. This is opposite to the SVM and LR models as they are quick to train but testing takes much more time. This leaves very little room for practical applications as most users except quick turnaround. Even with parallel computing with Spark, testing time was over 60% of the total process time. This time can be significantly reduced by increasing the number of nodes where 5 and 13 nodes have a total model training and testing time of 14.7 and 8 minutes for a total of 10,000 images.

RQ3 – As seen from the figures relating to pre-processing and feature extraction, the applied techniques were successful in increasing predictive accuracy for the LR model. The pre-processing techniques of median filtering and histogram equalization increased overall predictive accuracy by approximately 10%. When looking at benefits on a class level, female accuracy was increased by 13.5% where as male classification accuracy improved by 6%.

Observing specific examples of the pre-processing steps in Figures 11 and 12, it can be seen that there were cases of improving and reducing image quality. This results in a more normalized data set which contributes for improved predictions from the model.

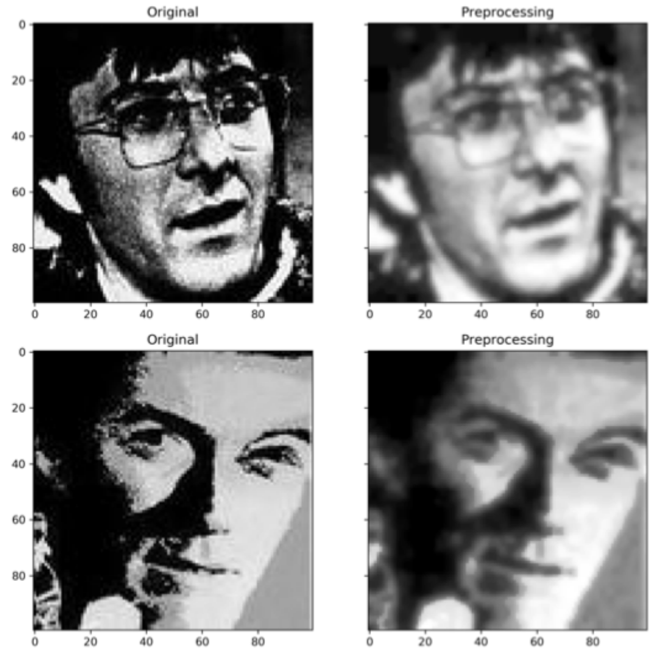

Fig. 11. Improving Image Quality with Pre-processing Step

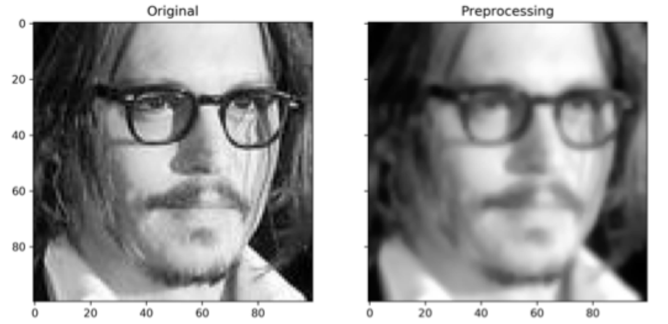

Fig. 12. Reducing Image Quality with Pre-processing Step

By using feature extraction techniques, the overall accuracy improved by approximately 15%. For gender based, feature extraction techniques improved male and female accuracy approximately by 10% and 14% respectively.

For the feature extraction techniques, it can be seen that DAISY provided the best results. Interestingly, the original and pre-processed data nets in similar accuracy's. It can be noted that when utilizing DAISY feature extraction, median filtering and histogram equalization have minimal impact to gender classification. For using multiple feature extraction techniques, it can be observed that DAISY and Canny are consistently very successful at improving accuracy.

For every combination of feature extractors, male prediction accuracy improved where as certain combinations for females caused a drop in accuracy.

B. Future Work & Recommendations

Comparing the radial SVM against the LR, their accuracy's were within 3–5%. This is notable as the radial SVM used only the original 100x100 grey scale images where as the LR model utilized multiple pre-processing and feature extraction techniques. Testing the potential of the radial SVM with the same data could potentially create a model that has comparable accuracy to the CNN if the SVM follows a similar trend as the LR. A potential challenge for this work would be moving away from the parallel computing power of Spark as there are no Spark libraries for SVM's other than linear.

Another potential project would consist of extracting facial features such as facial hair and eye locations to better characterize an image. This could either be implemented manually or utilizing AR as it's API returns facial landmarks. This can better quantify the benefit of extracting these facial features for a gender classification model.

One of the largest errors introduced early on this study was the imbalance in the male and female subjects in the data set. As seen from Figure 5, the data set contains 63% males and 37% females. This has created a significant bias in all of the test results. Future tests must be carried out on a more balanced data set for meaningful results.

VI. Conclusion

Image processing is a powerful tool that is being utilized in various fields such as security, surveillance, and autonomous vehicles. Deep learning advancements have made it possible to perform age estimation and gender classification on sets of images. However, this can be difficult due to the variance in an image data set. A large labeled data set similar to the one analyzed in this paper will improve the predictions of any model, regardless if it is a CNN, SVM, or LR.

Deep Learning is currently one of the best tools for image processing in machine learning but from the results of this study, it can be noted that many techniques can be employed to better improve simpler algorithms. Pre-processing and feature extraction brought these algorithms within 10% of Amazon Rekognition. However, AR showed strengths in maintaining a high level of accuracy with very low quality data, even though the prediction confidence might be low.

CNN's like Amazon Rekognition excel with image data and will continue being the leading implementation for gender classification and image processing due to it's ability to output test results quickly, versatility to image quality, and compatibility with large, unstructured data.

References

[1] S. Escalera, J. Fabian, P. Pardo, X. Baró, J. Gonzalez, H. J. Escalante, D. Misevic, U. Steiner, and I. Guyon, "Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 1–9.

[2] J. Xing, K. Li, W. Hu, C. Yuan, and H. Ling, "Diagnosing deep learning models for high accuracy age estimation from a single image," *Pattern Recognition*, vol. 66, pp. 106–116, 2017.

[3] I. Gruber, M. Hlaváč, M. Železný, and A. Karpov, "Facing face recognition with resnet: Round one," in *International Conference on Interactive Collaborative Robotics*. Springer, 2017, pp. 67–74.

[4] P. Rodríguez, G. Cucurull, J. M. Gonfaus, F. X. Roca, and J. Gonzalez, "Age and gender recognition in the wild with deep attention," *Pattern Recognition*, vol. 72, pp. 563–571, 2017.

[5] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image," in *The IEEE International Conference on Computer Vision (ICCV) Workshops*, December 2015.

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[7] G. Antipov, M. Baccouche, S.-A. Berrani, and J.-L. Dugelay, "Effective training of convolutional neural networks for face-based gender and age prediction," *Pattern Recognition*, vol. 72, pp. 15–26, 2017.

[8] TensorFlow, "TensorFlow." [Online]. Available: https://www.tensorflow.org/

[9] Keras, "Keras." [Online]. Available: https://keras.io

[10] Amazon Web Services, "Amazon Rekognition." [Online]. Available: https://aws.amazon.com/rekognition/

[11] Exxact, "Beginner's guide: Image recognition and deep learning." [Online]. Available: hhttps://blog.exxactcorp.com/how-does-image-recognition-work-deep-learning-basics/

[12] Google Cloud, "Cloud Vision." [Online]. Available: https://cloud.google.com/vision/

[13] Internet Movie Database, "IMDb." [Online]. Available: https://www.imdb.com/

[14] Apache, "Apache Spark." [Online]. Available: https://spark.apache.org/

[15] University of Calgary, "Advanced Reserach Computing." [Online]. Available: https://hpc.ucalgary.ca/

[16] "Python image library." [Online]. Available: https://pillow.readthedocs.io

[17] R. E. W. Rafael C. Gonzalez, *Digital Image Processing*. Prentice Hall, 2008.

[18] J. S. Lim, *Two-Dimensional Signal and Image Processing*. Prentice Hall, 1998.

[19] V. L. Engin Tola and P. Fua, "Daisy: An efficient dense descriptor appliedto wide-baseline stereo," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 32, pp. 815–830, 2010.

[20] V. K. R. Ramesha K, K B Raja and L. M. Patnaik, "Feature extraction based face recognition, gender and age classification," *International Journal on Computer Science and Engineering*, vol. 2, pp. 14–23, 2010.

[21] W.-Y. Y. E. S. Jian-Gang Wang, Jun Li, "Boosting dense sift descriptors and shape contexts of face images for gender recognition," *International Conference on Pattern Recognition*, pp. 96–102, 2010.

[22] R. A. Saeed Mozaffari, Hamid Behravan, "Gender classification using single frontal image per person:," *International Conference on Pattern Recognition*, pp. 1192–1195, 2010.

[23] Scikit-Image, "scikit-image: Image processing in python." [Online]. Available: https://scikit-image.org/

[24] Scikit-Learn, "scikit-learn: machine learning in python." [Online]. Available: https://scikit-learn.org/