Facial Attribute Recognition

1st Prathamesh Joshi, 2nd Amruta Koshe, 3rd Nidhi Vanjare
[1,2,3,]*Master of Engineering Systems and Technology*
McMaster University
Hamilton, ON Canada
1st joship14@mcmaster.ca, 2st koshea@mcmaster.ca, 3st vanjaren@mcmaster.ca

***Abstract*—Facial attribute recognition is the process of detecting and analyzing facial features to identify specific attributes such as gender, age, ethnicity, and emotions. The process involves capturing images or videos of faces and using machine learning algorithms to analyze and classify facial features. Facial attribute recognition has become increasingly popular and useful in a variety of industries, such as security, surveillance, marketing, and entertainment. For example, security and surveillance systems can use facial attribute recognition to detect and identify potential threats or suspects, while marketing companies can use it to target specific demographics with their advertising campaigns. In entertainment, facial attribute recognition can be used to create more realistic and immersive virtual reality experiences. In this report, we propose a CNN architecture as well as some pre-trained models to predict 40 facial attributes of an image. Our results show that ResNet50 model achieved the highest accuracy for detecting facial attributes with the fastest training time. The ResNet50 model was further used for real-time facial attribute recognition using OpenCV.**

***Index Terms*—Adam, VGG19, ResNet50, InceptionV3**

I. Introduction

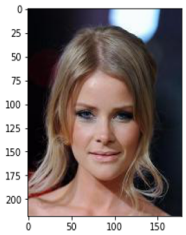
A key challenge in computer vision known as facial attribute recognition involves the identification and categorization of a variety of face traits, such as age, gender, expression, ethnicity, and facial hair. It is beneficial in many different fields, including security, entertainment, healthcare, and marketing. Facial attribute recognition is used in biometric identification in a variety of contexts, including face detection and recognition, emotion analysis, virtual try-on, and personalized advertising. In recent years, state-of-the-art outcomes have been attained on this challenge thanks in large part to the superior performance of convolutional neural networks [5](CNNs) and conditional generative adversarial networks (CGAN). Face attribute identification is a challenging task because of the variations in illumination, location, occlusion, ethnicity, privacy, and ethical concerns.

As a result, effective and reliable facial attribute identification methods are required that can handle both these issues and the emerging applications for this technology. This paper presents a comprehensive analysis of CNN's face attribute recognition system and proposes a novel approach to improve the task's robustness and accuracy.

II. Methodology

A. Dataset

For any categorization issue, having access to extensive and well-designed databases is essential. For our task of Facial Attribute Recognition, we make use of the CelebA dataset obtained from Kaggle [1]. CelebA is a massive collection of celebrity photos with a lot of background noise and different poses. There are a total of 202,599 face photos, representing 10,777 individuals.




```
Index(['5_o_Clock_Shadow', 'Arched_Eyebrows', 'Attractive', 'Bags_Under_Eyes',
       'Bald', 'Bangs', 'Big_Lips', 'Big_Nose', 'Black_Hair', 'Blond_Hair',
       'Blurry', 'Brown_Hair', 'Bushy_Eyebrows', 'Chubby', 'Double_Chin',
       'Eyeglasses', 'Goatee', 'Gray_Hair', 'Heavy_Makeup', 'High_Cheekbones',
       'Male', 'Mouth_Slightly_Open', 'Mustache', 'Narrow_Eyes', 'No_Beard',
       'Oval_Face', 'Pale_Skin', 'Pointy_Nose', 'Receding_Hairline',
       'Rosy_Cheeks', 'Sideburns', 'Smiling', 'Straight_Hair', 'Wavy_Hair',
       'Wearing_Earrings', 'Wearing_Hat', 'Wearing_Lipstick',
       'Wearing_Necklace', 'Wearing_Necktie', 'Young'],
      dtype='object')
--------------------------------------------------------------------------------
[0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0,
0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1]
```

Fig. 1. 228 x 228 example from the Dataset

The dataset has a partition file which specifies the partitions required for training, validation and test sets. 80% of the images belong to the training set, 10% to the validation set and 10% to the testing set. Each image is marked with 40 different facial attributes. Some of these attributes include arched eyebrows, smiling, brown hair, wearing eyeglasses etc. Each of these attributes is a binary label and has the value of either "0" or "1", where "1" signifies the presence of the associated facial attribute whereas "0" signifies its absence. The final aim of our project is to accurately predict the binary values of all 40 facial attributes for a new unseen image. An example is shown in Figure 1.

B. Exploratory Data Analysis

Exploratory Data Analysis is a method of analyzing and summarizing datasets to gain a better understanding of the data and identify patterns, relationships, and anomalies. The importance of EDA lies in its ability to reveal patterns and relationships that may not be evident from a simple statistical summary or raw data. By exploring the data in detail, we can identify potential errors, outliers, or missing values, which can affect the quality and validity of the analysis.

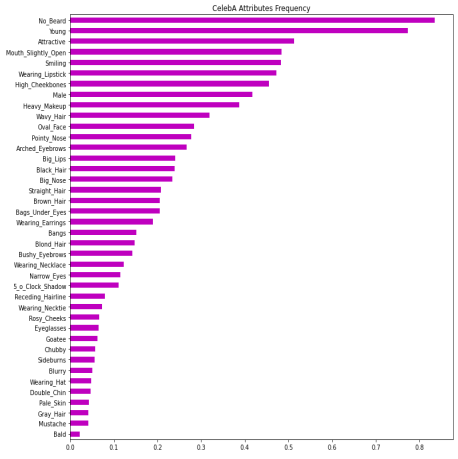


Fig. 2. CelebA Attribute Frequency

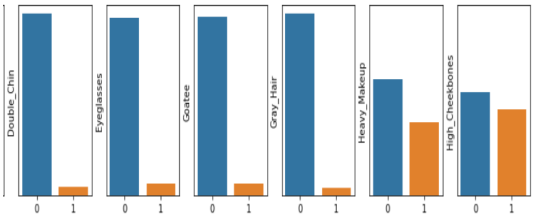


Fig. 3. Distribution of some of the features in the Dataset

Figure 2 represents the frequencies of the 40 attributes as present in the dataset. It can be observed that “No_Beard” is the most frequent attribute whereas “Bald” is the least frequent attribute in the dataset. Such high differences in the frequencies of the attributes lead to an imbalanced dataset and can further lead to biased predictions on the testing dataset.

Figure 3 depicts the distribution of values of each of the 40 attributes. With a significant disparity in the number of ”0” and ”1” values for many attributes, it is clear that the majority of the attributes are unbalanced.

A typical method for decreasing the number of features or variables in a dataset while keeping the most crucial data is PCA (Principal Component Analysis) feature reduction. PCA captures the most variation in the data by converting the original characteristics into a set of new, uncorrelated features called principle components. In order to study the correlation among the 40 facial attributes, a heatmap was plotted as illustrated in Figure 4. However, it is clear that there is a minimum correlation between the 40 attributes, with the attributes ”Heavy_Makeup” and ”Wearing_Lipstick” showing the strongest correlation. Therefore, we decided against using PCA as it would not make any significant difference to the results obtained.

C. Data Preprocessing

Data preprocessing refers to the techniques and processes used to prepare raw data for analysis. It involves a series of steps, including data cleaning, data transformation, and data

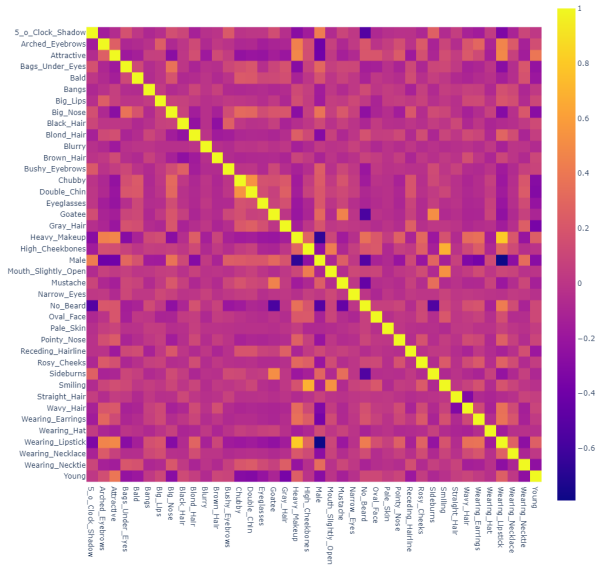


Fig. 4. Distribution of each feature example

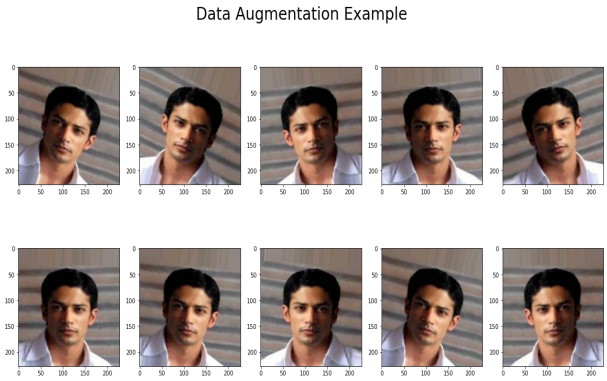


Fig. 5. Data Augmentation sample image

reduction, aimed at improving the quality and usefulness of the data. The importance of data pre-processing lies in its ability to improve the quality and accuracy of the analysis. By removing errors, inconsistencies, and missing values and normalizing the data [6], we can reduce the likelihood of biased results and improve the effectiveness of models.

Since the CelebA dataset is perfectly filled with no missing or wrong values, data cleaning was not required. Nonetheless, we replaced the ”-1” values from the attributes with ”0” for a better understanding of whether those attributes are present or not.

In order to tackle the imbalances present in the dataset, Data Augmentation was performed [7]. Data augmentation is a technique to artificially increase the size of a dataset by creating new examples from existing data. It involves applying a set of transformations or operations to the existing data to create new data that is similar to the original but not identical as shown in Figure 5.

The goal of data augmentation is to increase the variety

and diversity of the data, which can help to improve the accuracy and robustness of models. Under Data augmentation, we implemented rescaling, rotating the images, horizontal flipping, and filling pixels with the nearest pixel value.

D. Model training

Model training refers to the process of feeding a machine learning algorithm with labeled data, known as the training data, and using it to learn patterns and relationships between the input and output variables. The goal is to create a model that can accurately predict the output variable for new input data.

We propose a basic Convolutional Neural Network model [5] for this dataset, that consists of some Conv2D, MaxPool2D, Flatten and Dropout layers, with a total of approximately 6 million trainable parameters. An input image of 228x228x3 is passed through the model with a kernel size of 3x3 and pooling of 2x2 in the max pool layers. We propose dropping 50% of neurons before flattening the image to avoid overfitting. The model was compiled with binary cross-entropy loss function and Adam optimizer with a learning rate of 0.001. The accuracy metric used was binary accuracy [8] since there are 40 binary attributes to predict as the output.

Apart from our proposed model, we implemented Transfer Learning with 3 pre-trained models namely InceptionV3 [2], Vgg19 [4] and ResNet50 [3]. Transfer learning is a technique where a pre-trained model is used as a starting point for a new task instead of training a new model from scratch. This approach leverages the knowledge gained from training on a large dataset to improve the performance of a model on a smaller, more specific dataset. In transfer learning, the pre-trained model is typically trained on a large dataset, such as ImageNet, that contains millions of images with various labels. The model learns to recognize high-level features, such as edges, textures, and shapes. The weights of the pre-trained model are then fine-tuned on a smaller dataset that is specific to the target task.

Transfer learning can offer several advantages, such as faster convergence, reduced overfitting, and improved accuracy, especially when the target dataset is small or similar to the original dataset.

All the transfer learning models were fine-tuned i.e. they were first trained by freezing all the layers and then by keeping some layers trainable. Fine-tuning can reduce the time and resources required for training a new model from scratch, and can often achieve better performance than training a new model from scratch. The models with some of the layers kept trainable had increased in accuracy as discussed in the results section.

III. Results

Table 1 compares the performance of our proposed model and the 3 transfer learning models with the training parameters along with their accuracy. It is evident that ResNet-50 after fine-tuning gave the most accurate results on the testing dataset with an accuracy of 90%. Figure 5 describes the training curve for the ResNet-50 model with training and validation accuracy curves.

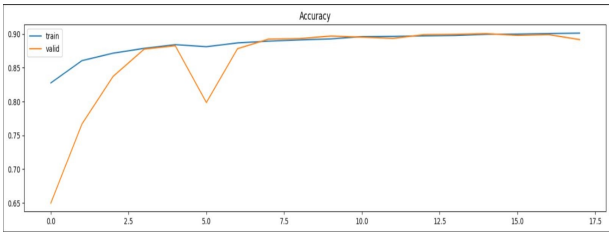


Fig. 6. Training curve for ResNet50 pre-trained model

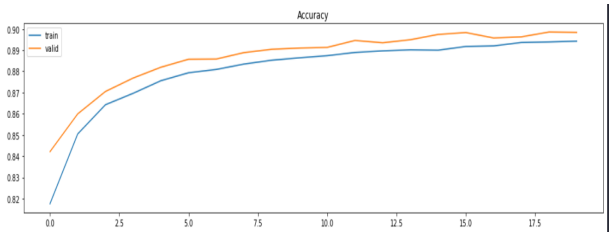


Fig. 7. Training curve for Our Proposed CNN model

One unexpected result was that despite having less training parameters, our suggested model outperformed the other two transfer learning models, namely Inception V3 [2] and Vgg-19 [4]. Figure 5 represents the training curve for our proposed model with the training and validation accuracy. As is well-known, there is often a relationship between the number of training parameters and the accuracy of a deep learning model. Generally, models with more parameters are capable of capturing more complex patterns in the data, which can lead to higher accuracy.

However, this relationship is not always straightforward, and there are other factors that can affect model performance as well. For example, models with too many parameters can suffer from overfitting, as was the case of Inception v3. Overfitted models become too specialized to the training data and perform poorly on new data. Additionally, models with too few parameters may not have enough capacity to capture all the relevant patterns in the data, leading to underfitting and reduced accuracy.

Therefore, it's important to strike a balance between model complexity and performance by tuning the number of training parameters and other hyperparameters to optimize performance on a validation set or through other evaluation methods.

Figure 8 depicts one image from the testing dataset and the 40 facial attributes as predicted by our model for the image along with their true values. As is visible from the figure, a majority of the facial attributes are predicted correctly and are displayed in green. However, some facial attributes like "Heavy_makeup" and "High_Cheekbones" have been mispredicted and are displayed in red.

Our model was also tested on real-world images by cap-

Fig. 8. Model Predictions on testing data

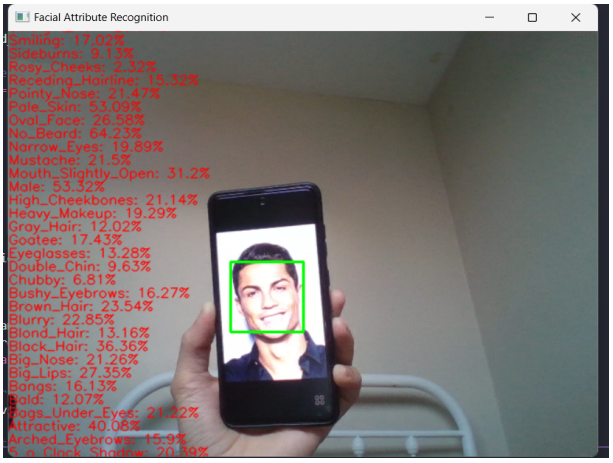


Fig. 9. Live Demo of our working model

turing them through a laptop camera of 1080p resolution. The model performs satisfactorily under normal lighting conditions. We hypothesize that it will perform better with an accuracy of around 70-80% if the proper lighting scenarios are taken care of. The live demo can be seen from Figures 9 & 10.

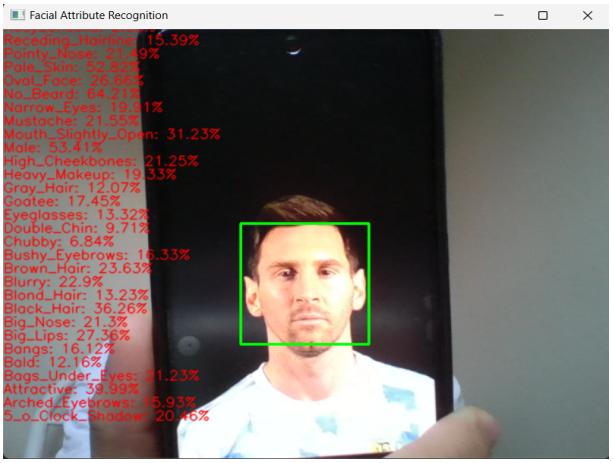


Fig. 10. Live Demo of our working model

TABLE I
MODEL PERFORMANCE COMPARISON

Model	Training Parameters	Accuracy
Proposed model	6,518,440	89.4%
Inception V3	38,899,880	84%
Vgg-19	19,945,512	88.47%
ResNet-50	25,722,984	90%

IV. DISCUSSION

Some of our key takeaways from this project are; To output a probability distribution over the many classes, a convolutional neural network (CNN) for multi-label classification [9]uses a softmax layer. However, in our case each label being binary in nature we use a sigmoid activation in the final layer of the model. Each label here denotes a potential class to which an input image may belong 0 or 1.

For binary classification problems, CNN frequently uses binary cross-entropy as a loss function. It calculates the discrepancy between the actual binary classification task probability distribution and the projected probability distribution. This is a common option for binary classification problems as it is comparatively simple to calculate, optimize, and generally performs well.

As the model size increases, the training time generally increases, especially in our case as there were more parameters that need to be learned. However, increasing the model size does not necessarily guarantee an improvement in performance, and in some cases, it can even lead to overfitting or diminishing returns. For example, our pre-trained InceptionV3 model was overfitting to the training dataset and hence performed poorly on the testing set.

The size and variety of the training data are increased using augmentation approaches in CNN, which can enhance a model's generalization capabilities. However, not every augmentation strategy works as well for every dataset and model, and in certain circumstances, it might even result in performance degradation. For example, we did not use other augmentation techniques like randomized cropping, zooming, etc since it led to a decrease in the accuracy in some models.

Some of the future prospects include using EfficientNetV2 as a pre-trained model which might work better, while also performing some more data augmentation to increase the size of training data. Some of the model's performance also depends on compute resources as it takes a significant amount of time to train models like Inceptionv3 and EfficientNetV2.

V. CONCLUSION

Facial attribute recognition is a challenging but important task in computer vision, with applications in various domains such as security, entertainment, and healthcare [10]. In conclusion, our project involved the usage of CelebA faces dataset from Kaggle, annotating them with labels for specific attributes such as age, gender, and facial expression, and training a deep learning model to accurately predict these attributes on new

images. ResNet-50 provided the best results on the test data with an accuracy of 90%.

One of the key challenges in facial attribute recognition is dealing with variations in lighting, pose, and occlusion, which can affect the accuracy of the model's predictions. Another important consideration is the ethical and privacy implications of facial recognition technology, particularly with regard to potential biases and the need for informed consent.

Overall, facial attribute recognition has the potential to provide valuable insights and improve decision-making in a variety of contexts, but it is important to approach this technology with caution and awareness of its limitations and potential risks.

REFERENCES

[1] https://www.kaggle.com/datasets/jessicali9530/celeba-dataset.

[2] Szegedy, Christian, et al. "Rethinking the Inception Architecture for Computer Vision." ArXiv.org, 11 Dec. 2015, https://arxiv.org/abs/1512.00567.

[3] He, Kaiming, et al. "Deep Residual Learning for Image Recognition." ArXiv.org, 10 Dec. 2015, https://arxiv.org/abs/1512.03385.

[4] Simonyan, Karen, and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." ArXiv.org, 10 Apr. 2015, https://arxiv.org/abs/1409.1556.

[5] O'Shea, Keiron, and Ryan Nash. "An Introduction to Convolutional Neural Networks." ArXiv.org, 2 Dec. 2015, https://arxiv.org/abs/1511.08458.

[6] Ioffe, Sergey, and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." ArXiv.org, 2 Mar. 2015, https://arxiv.org/abs/1502.03167.

[7] Perez, Luis, and Jason Wang. "The Effectiveness of Data Augmentation in Image Classification Using Deep Learning." ArXiv.org, 13 Dec. 2017, https://arxiv.org/abs/1712.04621.

[8] Balayla, Jacques. "Prevalence Threshold and Bounds in the Accuracy of Binary Classification Systems." ArXiv.org, 25 Dec. 2021, https://doi.org/10.48550/arXiv.2112.13289.

[9] Tawiah, Clifford A., and Victor S. Sheng. "A Study on Multi-Label Classification." SpringerLink, Springer Berlin Heidelberg, 1 Jan. 1970, https://link.springer.com/chapter/10.1007/978-3-642-39736-3_11.

[10] Chen, Zhenghao, et al. "Improving Facial Attribute Recognition by Group and Graph Learning." ArXiv.org, 28 May 2021, https://doi.org/10.48550/arXiv.2105.13825.