Face Verification with Veridical and Caricatured Images using Prominent Attributes

Jayam Sutariya
Computer Science & Engineering
University of Nevada, Reno
Reno, USA
jsutariya@unr.edu

Cooper Flourens
Computer Science & Engineering
University of Nevada, Reno
Reno, USA
cooperflourens@unr.edu

Nathan Thom
Computer Science & Engineering
University of Nevada, Reno
Reno, USA
nthom@unr.edu

Emily Hand
Computer Science & Engineering
University of Nevada, Reno
Reno, USA
emhand@unr.edu

***Abstract*—Caricatures, with their exaggerated features, offer an efficient means for individuals to recognize each other compared to veridical (real) images. However, matching veridical images to caricatures remains a challenging task in machine learning. This difficulty stems from poor quality caricature datasets lacking clear labels, inadequate labeling in widely used veridical image datasets like CelebA, and a shift away from attribute-based representations due to the rise of neural networks. These issues significantly impact the accuracy of face verification tasks between caricatures and veridical images. To address these challenges, this paper introduces a classification protocol for prominent facial feature recognition and a verification protocol for matching celebrity veridical images to their caricatures. We utilize CarVer, a recently curated dataset comprising both veridical and caricature images accompanied by detailed prominent attribute labels. Our approach aims to develop a set of prominent facial attributes that can effectively represent both real and caricatured images, enabling improved face verification across these modalities. This research has potential applications across various industries where robust cross-modal face recognition is crucial.**

***Index Terms*—face verification, facial features, caricatures, AI explainability, multi-label learning**

I. Introduction

Facial recognition technology is ubiquitous, finding usability in various applications such as unlocking smartphones and strengthening security measures. Face Verification, a task in the broader domain of face recognition involves verifying whether two images belong to the same person or not. Near perfect accuracy on face verification have been achieved with current facial recognition systems that utilize state-of-the-art (SOTA) deep learning (DL) models [34]. However, despite these advancements, a significant challenge persists: accurately matching veridical images to their caricatured counterparts. Caricatures present a unique set of obstacles due to their exaggerated features and artistic interpretations, posing challenges distinct from those encountered with veridical images. Addressing these challenges is crucial for developing more versatile facial recognition systems applicable across various industries.

In the realm of machine learning (ML), the terms 'attributes' and 'features' are sometimes used interchangeably, but they can have distinct meanings in certain contexts. Features often refer to internal representations generated by ML models during the learning process. Attributes, on the other hand, typically denote predefined characteristics used to describe visual data, such as age or gender. Traditionally, attributes have served as effective tools for outlining various characteristics of objects, with applications in tasks like medical diagnosis and image processing [35]. Attributes have also been used in facial recognition systems to describe faces. Examples of these attributes would be wide eyes, pointy ears, and dark hair. The human brain processes faces much differently than other objects [36]. Our brain relies on configural information, the spatial relationship between components of the face, along with featural information, the shape and size of components of the face [37]. This emphasizes the importance of facial attributes which describe the shape and size of facial features. In fact, facial features like the eyes, eyebrows, nose, mouth, and jaw profoundly influence our perception and judgment of traits such as attractiveness, masculinity, personality, and mood, with different features playing more prominent roles for certain judgments [38].

Besides facial attributes, there are additional tools that can be utilized to emphasize the importance of featural information in a face. Caricatures, with a rich history spanning centuries, progressed from tools of political satire to beloved forms of entertainment [6]. Beyond their artistic appeal, caricatures also serve as invaluable tools for solving the complex challenges of facial recognition. With their exaggerated features and simplified representations, caricatures could potentially be used to leverage the featural information in a face. Moreover, research in the field of Psychology and Neuroscience has found that people accurately identify the person quicker when looking at a caricature than a veridical photo [1] [2] [3] [8]. The inherent recognizability of caricatures suggests a promising direction for innovation in facial recognition, potentially enhancing accuracy and efficiency in diverse applications.

At the core of facial recognition technology lies the task of face verification. Face Verification using predefined attributes has been performed on face image datasets like CelebA and LFW. Researchers were able to achieve about 85% accuracy on attribute-based face verification [16] [39], however, achieved over 99% accuracy using SOTA DL models [34] that did not rely on attribtues. While face verification algorithms have attained impressive accuracy in matching veridical images [5], challenges emerge when extending these algorithms to match veridical images to caricatures. The distortions inherent in caricatures, combined with the shortcomings of existing datasets and labeling inconsistencies present significant challenges in achieving accurate face verification between caricatures and real images.

The intersection of caricatures, attributes, and face verification presents a complicated problem with profound implications for facial recognition technology. This paper explores new methods to improve the accuracy of matching real photos to caricatures by examining the relationships between facial features in both types of images. We establish a solid foundation for future research and innovation in the field of facial recognition by examining the nuances ingrained within caricatures, attribute-based representations, and face verification.

The central focus of this research lies in developing protocols for classification of prominent facial features and verification, specifically for matching veridical images to caricatures. By exploring these relationships, we aim to enhance the accuracy and efficiency of facial recognition systems, establishing a foundation for future innovations in cross-modal face recognition.

II. Related Works

This research explores the intersection of caricatures, attributes, and face verification. Understanding existing literature in these domains is crucial for contextualizing our research efforts and outlining directions for innovation.

A. Caricatures

Caricatures, artistic representations emphasizing and distorting prominent features, have a long history dating back to ancient civilizations [6]. They play a unique role in facial recognition technology, highlighting distinctive facial features used by the human brain for recognition [36], [37]. Research has shown that caricatures can enhance recognition accuracy and efficiency, with studies indicating that people often identify individuals more quickly from caricatures than from veridical photos [1]–[3], [7], [8], [11].

Early computational studies, such as the Caricature Generator program [9], laid the groundwork for caricature research in facial recognition. This program amplified differences between the face to be caricatured and a comparison face, simulating the visualization process in the caricaturist's imagination.

More recent efforts have focused on creating datasets and developing frameworks for caricature face recognition. The WebCaricature dataset [10] was introduced to facilitate research in this area. However, it suffers from significant limitations, including demographic imbalances (approximately 72% male and 75% Caucasian) and low-quality images [13]. These issues are illustrated in Figure 1, which showcases examples of poor-quality caricatures from the dataset. WebCariA was introduced as an extension of WebCaricature, incorporating CelebA-like image-level attribute labels. However, it still includes the low-quality caricatures from the original WebCaricature dataset, thus not fully resolving the image quality issues [40].

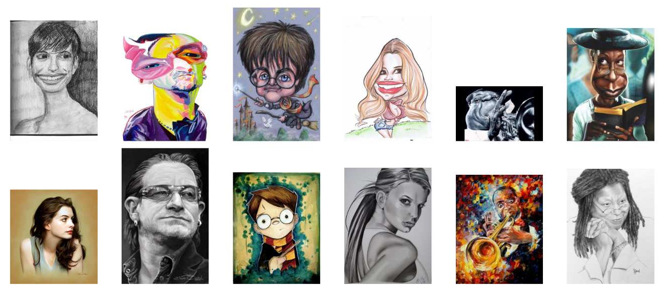

Fig. 1. Sample of poor quality caricatures from the WebCaricature dataset as presented in [13].

The limitations of existing datasets, with the exception of CarVer [13], pose significant challenges for caricature-based facial recognition research. These include:

1) Perceptual distortions inherent in caricatures, which can deviate significantly from the original appearance.
2) Limited dataset quality and diversity, hampering the development and evaluation of robust algorithms.

Despite these challenges, recent research has shown promise in improving caricature recognition. For instance, Zheng et al. proposed a novel approach for few-shot caricature recognition, demonstrating improved performance across multiple caricature datasets [12].

Recent advancements in Generative Adversarial Networks (GANs) have shown promise in addressing some of these challenges. Researchers have developed models like CariGAN [41], StyleCariGAN [42], and WarpGAN [43] for caricature generation and matching. However, the quality of generated caricatures still needs improvement, and the challenge of accurately matching photos to caricatures persists [44] [45] [46].

B. Attribute-based Representations

Attribute-based representations in facial recognition date back to early computer vision research. Significant contributions include the SIFT algorithm [15], which revolutionized feature detection and description, and the FERET program [14], which established benchmarks for evaluating attribute-based recognition systems.

Popular face image datasets like LFW and CelebA offer predefined attributes for facial attribute recognition and verification [24] [47]. Methods such as Simile Classifiers and the MOON classifier achieved notable accuracies for their time using these attribute-based approaches [16] [39].

However, attribute-based representations face several challenges:

1) The dominance of deep learning architectures, particularly convolutional neural networks (CNNs), which learn complex representations directly from raw pixel data.
2) Subjectivity in attribute definitions. Attributes like ”attractive” or ”smiling” in datasets like CelebA can vary in interpretation across individuals or cultural contexts [13] [24].
3) Limited discriminative power in capturing fine-grained facial characteristics.
4) Data imbalance and label noise. For example, in CelebA, images labeled as ”bald” were contradictorily labeled as having bangs, receding hairline, straight hair, or wavy hair 33.3% of the time [48].

Despite these challenges, recent advancements have revitalized interest in attribute-based representations. Integration with deep learning architectures and generative adversarial networks (GANs) has shown promise [4] [19]. For instance, Lu et al. developed an approach for attribute-guided face generation using conditional CycleGAN [18], allowing seamless control over facial appearance based on input attributes [17].

Furthermore, self-supervised learning techniques offer potential benefits for attribute-based facial recognition, enabling models to learn from unlabeled data and reducing reliance on manually annotated datasets [20].

C. Face Verification

Face verification, a core task in facial recognition technology, involves determining whether two facial images correspond to the same individual. It faces challenges due to variations in pose, expression, lighting conditions, and environmental factors [49] [50] [51] [55].

Early face verification systems relied on handcrafted features and shallow learning algorithms, such as Eigenfaces, Fisherfaces, and Local Binary Patterns (LBP) encoding [21] [22] [23]. However, these methods were limited in their ability to handle real-world variations.

Recent deep learning approaches, particularly convolutional neural networks (CNNs), have revolutionized the field. Landmark works like DeepFace introduced deep neural network architectures capable of directly mapping facial images to compact feature representations [5]. Subsequent advancements, including ArcFace, have further improved performance and generalization ability [34].

Face verification has diverse applications across sectors such as identity authentication, security, healthcare, and entertainment. For example, Wu et al. (2019) proposed identity authentication frameworks using face verification that outperform state-of-the-art methods [52]. In healthcare, multi-factor authentication methods incorporating face recognition have been proposed to enhance security in e-health systems [53].

Recent developments focus on improving accuracy, robustness, and efficiency through advanced deep learning architectures, novel evaluation metrics, and enhanced model training techniques. Transfer learning and adversarial training have contributed to improved generalization capabilities [54]. Additionally, explainable AI techniques, such as attribute-specific balanced accuracy and decision trees, have been applied to enhance the transparency and interpretability of face verification systems [32].

D. Summary

Our review highlights the need for diverse methodologies in facial attribute recognition research. While deep learning has revolutionized the field, attributes remain relevant, particularly for explainability and semantic richness. There is a clear need for more comprehensive and diverse datasets, especially for caricatures, to address current limitations and biases. By building upon existing literature and addressing these challenges, we aim to contribute to the advancement of facial attribute recognition technology, particularly in the context of matching veridical images to caricatures.

III. Experiments

Our research investigates the efficacy of attribute-based representations in facial recognition, focusing on two key aspects: classification of prominent facial features and verification of veridical images against caricatures. We utilize the CarVer dataset [13], which offers a comprehensive collection of caricatures and veridical images of celebrities.

A. Data Collection and Preparation

The CarVer dataset forms the cornerstone of our research, featuring 229 distinct identities, each represented by at least 5 images and often more. This dataset stands out for its demographic diversity, with approximately 58% male identities and only about 54% Caucasian, offering a more balanced representation compared to datasets like CelebA.

Fig. 2. Example veridical images of Barack Obama in the CarVer dataset [13].

Fig. 3. Example caricatures of Barack Obama in the CarVer dataset [13].

TABLE I
LIST OF ALL MAIN ATTRIBUTES AND THEIR SUBLABEL DESCRIPTORS FOR CARVER.

Attribute	Descriptors
cheekbones	high, sharp
cheeks	chubby/full, dimples, thin/hollow
chin	cleft, crooked, double chin, forward, pointed, rounded, scar, square, strong jawline, weak jawline
ears	big, flat, high, low, pierced, pointy, small, stick out
eyebrows	arched, bushy, curved down, far apart, flat, furrowed, light, long, scar, short, slanted down, thick, thin, unibrow
eyelids	drooping, hooded, puffy, receded
eyes	almond, bags under eyes, crows feet, deep-set, glasses, lazy eye, light colored, long eyelashes, narrow, narrow-set, round, slanted down, slanted up, small, stick out, wide, wide-set, wide-x
facial hair	beard, goatee , handlebar, messy, mustache, sideburns, soul patch, stubble, thick, thin, trimmed, white
forehead	big, narrow, scar, small, wide, wrinkled
hair	bald, bangs, big, black, blond, curly, dreads, hat, long, receding hairline, red, short, slicked back, white, white streaks, widows peak
head	big, long, round, small, square, wide
lips	downturned,large, medial cleft, pouty/full, red lipstick, thick lower, thin, thin upper, upturned
mouth	big/wide, crooked, small
neck	Adam's apple, lines, tattoos, thick
nose	bulbous, button, cleft, crooked, dorsal hump, flared nostrils, flat, hooked, long, pointy, rounded tip, short, small, small nostrils, thin, thin bridge, upturned, v-shaped, well-defined tip, wide, wide bridge, wide nostrils, wide tip
skin	freckles, mole, pale, rough, smooth
teeth	big, buck, crooked, gap, overbite, small, straight, white

A key strength of CarVer is its inclusion of high-quality caricatures alongside veridical images (Fig. 2, 3). Unlike some caricature datasets that may feature simplistic drawings, CarVer's caricatures are characterized by their detailed and realistic portrayal of facial features, significantly enriching the dataset's diversity and applicability.

Recently, the CarVer dataset underwent a significant update, introducing identity-level labels (attributes) that offer detailed insights into the prominent facial features of each identity [32]. The attribute labels are structured uniquely, providing comprehensive information about various facial attributes. Each identity is associated with a set of main labels representing prominent facial features such as eyes, cheeks, facial hair, chin, and forehead. These main labels further encompass multiple sub-labels, offering nuanced descriptions of specific characteristics.

In total, the CarVer dataset comprises 17 main attributes and 151 sub-attributes, reflecting the complexity and diversity of facial features captured within the dataset. This elaborate labeling scheme enables a detailed characterization of each identity, strengthening the dataset's utility for facial recognition and analysis tasks. A list of all of the main and sub attributes is provided in Table I.

The images within the CarVer dataset undergo preprocessing steps such as alignment, rotation, and cropping to ensure consistency and quality across the dataset. Notably, these preprocessing techniques were applied during the dataset creation process itself, eliminating the need for additional preprocessing steps during our research.

B. Model Architecture and Implementation

Our research utilizes the ResNet18 architecture as the backbone for our models [25]. ResNet18 strikes a balance between depth and computational efficiency, making it suitable for our multi-label classification task. Pretrained on the ImageNet1K dataset, ResNet18 captures high-level features from diverse visual data, providing a solid foundation for our specific task [26].

Modifications are made to ResNet18 to tailor it to our dataset and task requirements. We adjust the model's output dimensions to match the number of attributes in our dataset, whether it be the 17 main attributes or the 151 sub-attributes. This customization ensures that the model's output aligns with our classification task, allowing it to predict the presence or absence of each attribute for a given input image.

C. Training Procedure

1) Hyperparameters: The models are trained using Binary Cross Entropy with Logits as the loss function, chosen for its suitability in multi-label binary classification tasks. This loss function allows for the efficient optimization of models for recognizing multiple facial attributes simultaneously. Stochastic Gradient Descent (SGD) serves as the optimization algorithm, initialized with a learning rate of 0.1 and a momentum of 0.9.

Weight decay of 0.01 is added when training a model that includes main attributes, aiding in preventing overfitting. Additionally, a Reduce Learning Rate on Plateau scheduler is employed, with a patience parameter of 5. This scheduler monitors the validation loss and reduces the learning rate by a factor of 0.1 when the validation loss stops improving for a specified number of epochs. The training process is terminated when there is no improvement in the validation loss after 10 consecutive epochs, ensuring optimal convergence and preventing overfitting.

2) Dataset Splitting: We employ two distinct splitting strategies to partition the dataset into training, validation, and testing sets. We call these image split and identity split.

Image split: In this approach, every celebrity's image data is available in training and validation. Three randomly selected images per celebrity are allocated to the training set, one image per celebrity is dedicated to the validation set, while one additional image is reserved for testing. This strategy ensures balanced training and evaluation, especially for identities with a limited number of images.

Identity split: This protocol involves shuffling all identities in the dataset. 24 out of 229 identities are reserved for testing, while the remaining identities are subjected to 5-fold cross-validation. Each fold utilizes 5 images per identity for training and validation. The best performing model from the cross-validation folds is retained for further evaluation, ensuring robustness and generalization across diverse identities and facial attributes.

3) Classification: Multiple model variants are trained to accommodate the hierarchical structure of the attribute labels. Separate models are trained to recognize both main attributes and sub-attributes, reflecting the intricate details of facial features captured in the dataset. Additionally, models are trained using caricatures only, veridical images only, and a combined set containing both caricatures and veridical images. In total, 12 models are trained to evaluate each combination of labels and input data.

4) Verification: Verification tasks are executed by utilizing the test set derived from both the identity split and the image split. In the image split scenario, each veridical image within the test set is paired with corresponding caricatures, all sourced from the same image split test set. This process generates one correct pair per veridical image, matched with the caricature belonging to the identical identity, and one incorrect pair per image, paired with a caricature from a distinct identity within the same test set. This procedure yields a total of 458 pairs across 229 identities.

In the identity split, each test identity contributes to the creation of verification pairs. This involves pairing five veridical images with their corresponding caricatures from the same identity, alongside five incorrect pairs formed by matching the veridical images with caricatures from random identities within the identity split test set. This results in 50 pairs per identity, culminating in a total of 1200 pairs across 24 identities.

Leveraging the saved models from the preceding classification phase, embeddings are generated for each image pair, facilitating the computation of cosine similarities. Subsequently, the True and False positives are employed to construct Receiver Operating Characteristic (ROC) curves, from which the Area Under the Curve (AUC) is computed to evaluate the verification performance for both primary and subsidiary attributes.

TABLE II
ACCURACY ACROSS DIFFERENT EXPERIMENT PARAMETERS (DATA SPLIT, ATTRIBUTE CATEGORY, AND TEST SET)

Data Split	Attribute Type	Veridical	Caricature	Combined
Image	Main	77.4	74.5	78.3
	Sub	93.3	93.1	93.4
Identity	Main	82.5	84.9	83.8
	Sub	94.5	95.1	95.2

D. Evaluation Metrics

For the classification tasks aimed at recognizing prominent facial attributes, accuracy is used as a primary metric to evaluate model performance. Furthermore, attribute-specific accuracy, along with the number of occurrences of each attribute, are calculated and plotted to provide a deeper understanding of the model's performance on specific attribute categories.

Additionally, confusion matrices are generated to visualize the performance of the models across different attributes. These matrices provide valuable insights into the distribution of true positive, true negative, false positive, and false negative predictions, facilitating a detailed analysis of model errors and misclassifications.

For the verification tasks aimed at matching veridical images with corresponding caricatures, Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) are utilized to evaluate model performance. The ROC curve provides a graphical representation of the trade-off between true positive and false positive rates, offering insights into the discriminatory power of the model. The AUC quantifies the overall performance of the verification model by calculating the area under the ROC curve.

E. Experimental Setup

Our experiments were conducted on a high-performance computing cluster equipped with Nvidia GeForce RTX 2080 GPUs. PyTorch served as the primary deep learning framework, complemented by additional libraries including Scikit-learn for machine learning tasks, Matplotlib and Seaborn for visualization, Weights and Biases for metric tracking, and Pandas for data manipulation [27] [28] [29] [30] [31] [33].

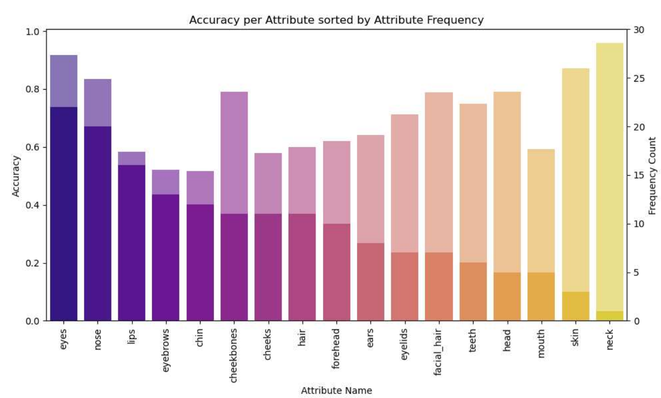

Fig. 4. Accuracy per attribute sorted by attribute occurrences on combined test set using main attributes with identity split (Best viewed in color)

IV. RESULTS

This section presents the results of our experiments evaluating model performance in two key tasks: classification (facial feature recognition) and verification (matching veridical images to caricatures).

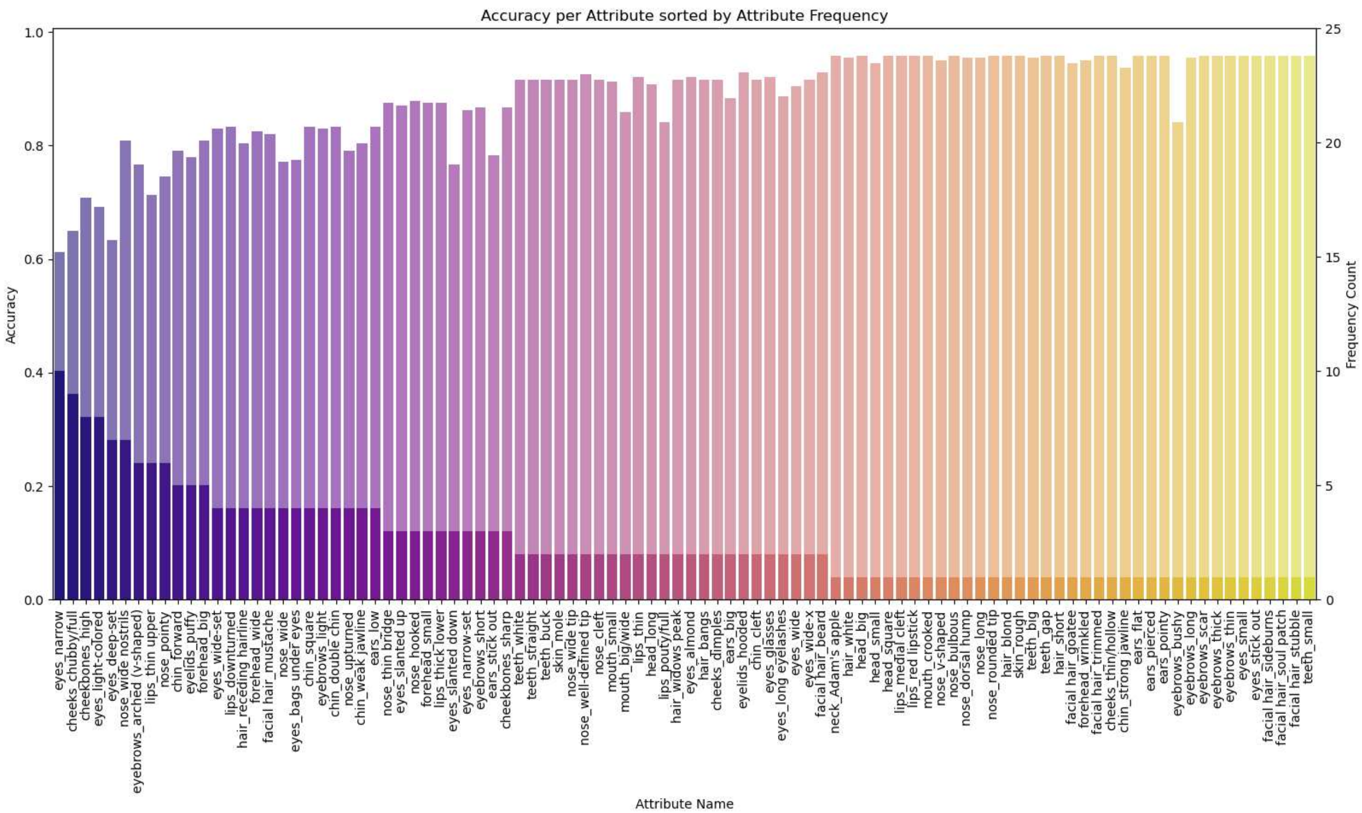

Fig. 5. Accuracy per attribute sorted by attribute occurrences on combined test set using sub attributes with identity split (Best viewed in color)

A. Prominent Feature Recognition

We assess the performance of developed models in accurately identifying and classifying various facial attributes, providing insights derived from confusion matrices and attribute-specific accuracies.

1) Model Performance: Table II summarizes the accuracy of our models on the test sets across different experimental parameters.

Models trained on sub attributes consistently outperform those trained on main attributes, indicating the importance of detailed labeling schemes in capturing nuanced facial characteristics. The identity split approach demonstrates higher accuracy compared to the image split, suggesting improved model performance with a more balanced distribution of identities in the training and validation sets.

The combined dataset consistently yields the highest accuracy scores, followed by the veridical dataset, and then the caricature dataset. This highlights the importance of dataset diversity in enhancing model generalization capabilities and the potential challenges associated with caricature recognition.

2) Attribute Analysis: We focus on a detailed analysis of individual attributes to illustrate model performance in recognizing specific facial features. Our examination comprises attribute-specific accuracy plots and attribute-specific confusion matrices. Based on the superior performance observed in the previous section, we present results for the identity split using the combined dataset, which consistently demonstrated the best performance across experimental conditions.

In the accuracy per attribute plots (Fig. 4 and 5), attributes are sorted based on the number of occurrences in the test set with the darker bars indicating number of occurrences and the lighter accuracies. Main attributes exhibit varying accuracies with no clear correlation to the number of occurrences, while sub attributes display a discernible trend of decreasing performance with higher occurrence counts. This indicates that attributes with fewer occurrences generally yield higher accuracies.

The attribute confusion matrices provide a detailed examination of model performance in terms of true positives, false positives, false negatives, and true negatives for each attribute.

For main attributes such as eyes and neck, which exhibited the highest accuracies in the previous plots, the confusion matrices reveal high numbers of true positives or true negatives, corresponding to their occurrence frequency (Fig. 6). Attributes like nose and skin, with slightly lower accuracies, demonstrate similar patterns. However, for other attributes, no clear patterns emerge, underscoring the need for further investigation into the factors influencing model performance for these attributes.

The confusion matrices for sub attributes reflect the mixed performance observed in the attribute-specific accuracy plots. The five most frequent attributes display an even distribution of true positives, true negatives, false positives, and false negatives, aligning with their relatively lower accuracies (Fig. 7). The five least frequent attributes, often present in only one identity, exhibit distinct patterns in the confusion matrices (Fig. 8). These attributes are typically predicted as absent, leading to a high number of true negatives but incorrect predictions of absence for the one present instance. These trends highlight the consistent performance patterns of the models across different attribute frequencies.

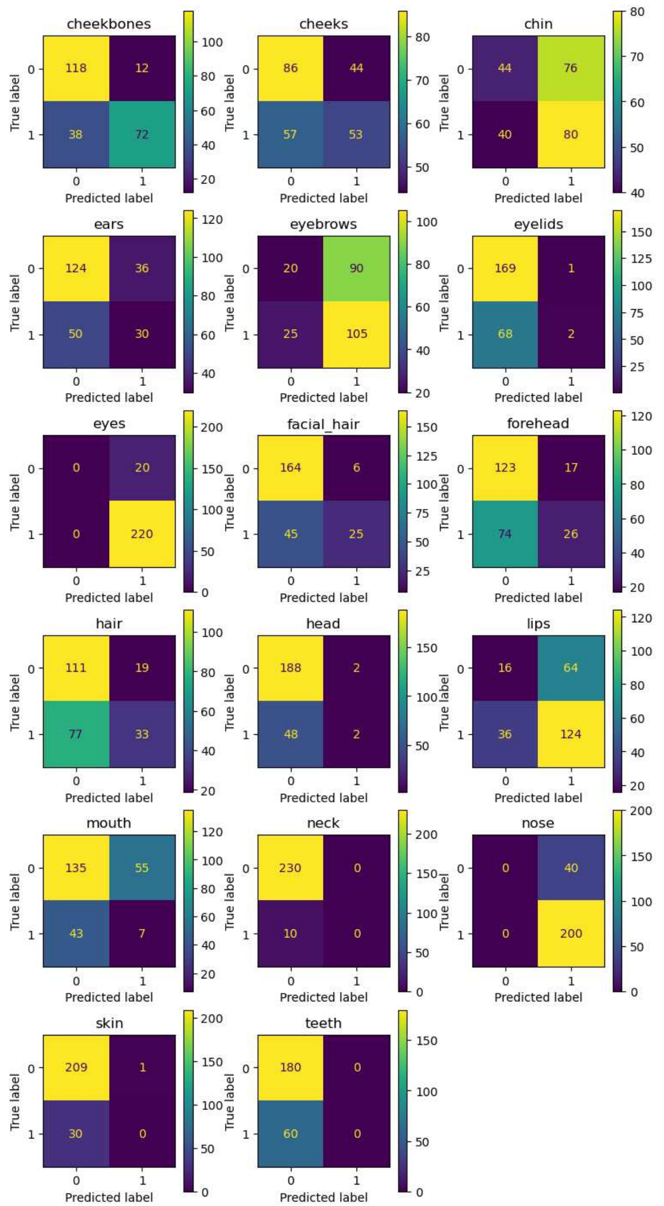

Fig. 6. Confusion Matrices for all main attributes on the combined test set with identity split

B. Verification with Caricatures

This subsection focuses on the verification task of matching veridical images to corresponding caricatures. We assess the effectiveness of our models in accurately identifying matching pairs using ROC curves and AUC scores.

In the image split scenario, models trained on sub attributes outperform those trained on main attributes, with AUC scores of 0.77 and 0.73, respectively (Fig. 9). This suggests that the finer-grained information captured by sub attributes contributes to enhanced verification performance. In the identity split scenario, sub attributes continue to exhibit superior performance, with an AUC score of 0.78 compared to 0.75 for main attributes (Fig. 10). This difference highlights the effectiveness of the identity split in promoting model generalization, leading to slightly improved verification performance.

An analysis of images with the highest rates of misprediction, using the identity split with sub-attribute results, revealed that all fifteen of the most mispredicted images were caricatures. This finding highlights the inherent challenge of face verification tasks involving caricatures, likely due to the varying styles of caricature depictions and artistic preferences in exaggerating prominent facial features.

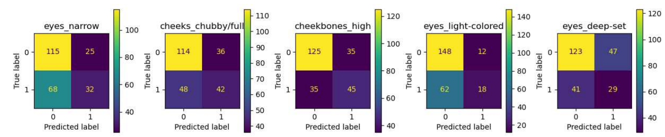

Fig. 7. Confusion Matrices for 5 most present sub attributes on the combined test set with identity split

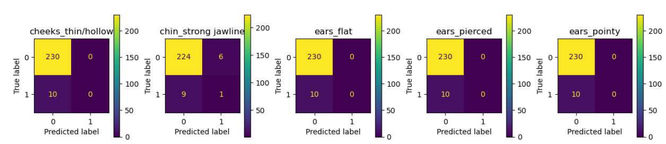

Fig. 8. Confusion Matrices for 5 least present sub attributes on the combined test set with identity split

C. Summary

Our investigation into prominent facial attribute recognition and veridical to caricature verification has yielded valuable insights into model performance across various experimental conditions.

For prominent facial attribute recognition, main attributes exhibited variable performance with no clear correlation to attribute occurrence counts, while sub attributes demonstrated decreasing performance with higher occurrence counts. Attributes with fewer occurrences generally yielded higher accuracies, highlighting the influence of attribute prevalence on model performance.

In veridical to caricature verification, models trained on sub attributes consistently outperformed those trained on main attributes, with higher AUC scores observed across both image and identity splits. The identity split scenario resulted in improved verification performance across attribute types, demonstrating better model generalization.

These findings contribute to the advancement of facial attribute recognition and verification tasks, providing insights into model performance, attribute analysis, and verification effectiveness. They highlight the importance of robust and discriminative facial attribute recognition models in various applications.

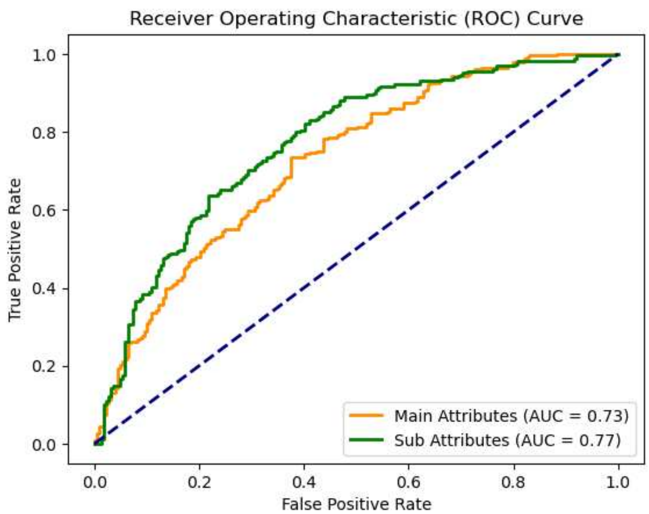

Fig. 9. Main and sub attribute ROC Curves for the image split

V. Conclusions and Future Work

This research has established a foundation for advancing facial recognition technology, focusing on identity-level attribute-based veridical-to-caricature face verification. Our protocol employs a two-stage pipeline: first classifying images according to their prominent facial attributes using state-of-the-art deep learning models, then comparing output attribute vectors for verification using cosine similarity. We utilized the CarVer dataset, comprising both veridical images and caricatures along with identity-level attributes.

Despite our accomplishments, several limitations were identified in our study. The scarcity of images (five per celebrity) and limited number of identities in our dataset posed challenges for comprehensive analysis and generalizability. Furthermore, the uneven distribution of attributes across identities, with some attributes present in only a few or even single identities, highlights the need for a more balanced attribute representation. Inherent limitations of identity-level labels were also observed, as not all images belonging to an identity consistently conformed to the assigned labels. Labeler biases could have further contributed to inconsistencies in prominent feature identification. Caricature-specific challenges emerged as well, with varying artistic styles leading to inconsistencies in prominent feature exaggeration and causing difficulties with image alignment methods.

Looking ahead, several promising avenues for future research present themselves. Enhancing our models through the integration of traditional machine learning techniques with deep learning methodologies could refine accuracy and robustness. Developing a unique loss function to align the feature space of caricatures and veridical images could significantly improve the training procedure. Exploring a hierarchical architecture that better resembles the nature of the labels could yield valuable insights. Additionally, combining featural information (attributes) with configural information (landmark-based labels) could more accurately replicate the brain's face processing mechanisms, potentially leading to improved performance.

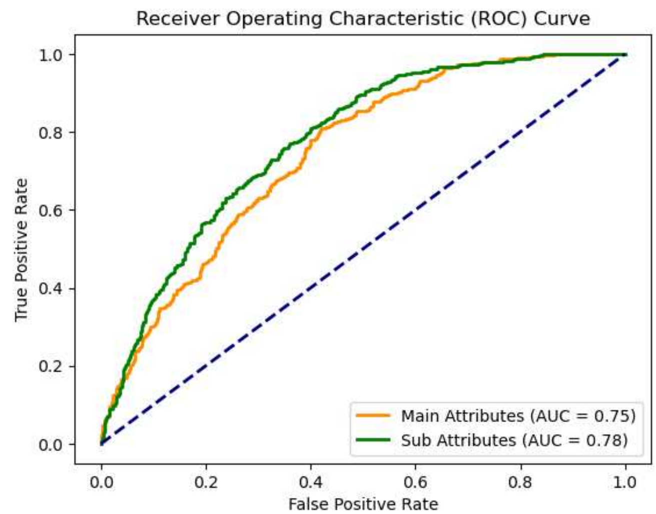

Fig. 10. Main and sub attribute ROC Curves for the identity split

Expanding the dataset to include a wider range of images and identities with balanced attribute representation would bolster the generalizability and effectiveness of our protocols. Comparative analyses, such as benchmarking our results against other datasets like WebCariA, would provide valuable insights and help identify areas for improvement. The exploration of attribute-conditioned caricature generation presents an intriguing opportunity for further investigation, with potential applications across various domains. Moreover, investigating the impact of pre-processing techniques and data augmentation strategies on model performance could yield valuable insights for improving facial recognition systems.

In conclusion, our research lays a solid foundation for future exploration and innovation in facial recognition, particularly in addressing the challenges of cross-modal face matching. This work has potential applications in enhancing accessibility for individuals with visual impairments and those on the autism spectrum, as well as improving security and entertainment technologies. By addressing the identified limitations and leveraging these opportunities, we aim to drive progress in developing more versatile and effective facial recognition solutions. The intersection of caricatures, attributes, and face verification presents a rich landscape for further research, holding promise for significant advancements that could expand the capabilities and applications of facial recognition technology across various domains.

REFERENCES

[1] R. Mauro and M. Kubovy, "Caricature and face recognition," *Memory & Cognition*, vol. 20, pp. 433–440, 1992.

[2] G. Rhodes, S. Brennan, and S. Carey, "Identification and ratings of caricatures: Implications for mental representations of faces," *Cognitive Psychology*, vol. 19, pp. 473–497, 1987. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0010028587900168

[3] Benson, P. & Perrett, D. Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal Of Cognitive Psychology*. **3**, 105-135 (1991)

[4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. Generative adversarial nets. *Advances In Neural Information Processing Systems*. **27** (2014)

[5] Taigman, Y., Yang, M., Ranzato, M. & Wolf, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition (CVPR)*. (2014,6)

[6] Lynch, B. A history of caricature. (Faber,1927)

[7] Davis, S. & Hand, E. Improving Face Recognition Using Artistic Interpretations of Prominent Features: Leveraging Caricatures in Modern Surveillance Systems. *Intelligent Video Surveillance*. (2022), https://doi.org/10.5772/intechopen.106073

[8] Lee, K., Byatt, G. & Rhodes, G. Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science*. **11**, 379-385 (2000)

[9] Brennan, S. & Generator, C. The dynamic exaggeration of faces by computer. *Leonardo*. **18**, 170-178 (1985)

[10] Huo, J., Li, W., Shi, Y., Gao, Y. & Yin, H. WebCaricature: a benc

[11] J. Halberstadt and G. Rhodes, "It's not just average faces that are attractive: Computer-manipulated averageness makes birds, fish, and automobiles attractive," *Psychonomic Bulletin & Review*, vol. 10, pp. 149–156, 2003.

[12] Zheng, W., Yan, L., Gou, C., Zhang, W. & Wang, F. A relation network embedded with prior features for few-shot caricature recognition. *2019 IEEE International Conference On Multimedia And Expo (ICME)*. pp. 1510-1515 (2019)

[13] Davis, S., Lingenfelter, B., McElhinney, K., Sengupta, S. & Hand, E. CarVer: Setting the Standard for Face Verification with Caricatures. *2023 IEEE International Joint Conference On Biometrics (IJCB)*. pp. 1-10 (2023)

[14] P. Rauss, J. Phillips, M. Hamilton, and A. DePersia, "FERET (Face Recognition Technology) program," in *25th AIPR Workshop: Emerging Applications Of Computer Vision*, vol. 2962, pp. 253–263, 1997. [Online]. Available: https://doi.org/10.1117/12.267831

[15] Lowe, D. Object recognition from local scale-invariant features. *Proceedings Of The Seventh IEEE International Conference On Computer Vision*. **2** pp. 1150-1157 (1999)

[16] Kumar, N., Berg, A., Belhumeur, P. & Nayar, S. Attribute and simile classifiers for face verification. *2009 IEEE 12th International Conference On Computer Vision*. pp. 365-372 (2009)

[17] Lu, Y., Tai, Y. & Tang, C. Attribute-guided face generation using conditional cyclegan. *Proceedings Of The European Conference On Computer Vision (ECCV)*. pp. 282-297 (2018)

[18] Zhu, J., Park, T., Isola, P. & Efros, A. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings Of The IEEE International Conference On Computer Vision*. pp. 2223-2232 (2017)

[19] Schmidt, R. Recurrent neural networks (rnns): A gentle introduction and overview. *ArXiv Preprint ArXiv:1912.05911*. (2019)

[20] Gui, J., Chen, T., Zhang, J., Cao, Q., Sun, Z., Luo, H. & Tao, D. A survey on self-supervised learning: Algorithms, applications, and future trends. *ArXiv Preprint ArXiv:2301.05712*. (2023)

[21] Kshirsagar, V., Baviskar, M. & Gaikwad, M. Face recognition using Eigenfaces. *2011 3rd International Conference On Computer Research And Development*. **2** pp. 302-306 (2011)

[22] Belhumeur, P., Hespanha, J. & Kriegman, D. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions On Pattern Analysis And Machine Intelligence*. **19**, 711-720 (1997)

[23] Hadid, A. The local binary pattern approach and its applications to face analysis. *2008 First Workshops On Image Processing Theory, Tools And Applications*. pp. 1-9 (2008)

[24] Liu, Z., Luo, P., Wang, X. & Tang, X. Deep Learning Face Attributes in the Wild. *Proceedings Of International Conference On Computer Vision (ICCV)*. (2015,12)

[25] He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition*. pp. 770-778 (2016)

[26] Deng, J., Dong, W., Socher, R., Li, L., Li, K. & Fei-Fei, L. Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference On Computer Vision And Pattern Recognition*. pp. 248-255 (2009)

[27] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L. & Others Pytorch: An imperative style, high-performance deep learning library. *Advances In Neural Information Processing Systems*. **32** (2019)

[28] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal Of Machine Learning Research*. **12** pp. 2825-2830 (2011)

[29] J. Hunter, "Matplotlib: A 2D graphics environment," *Computing In Science & Engineering*, vol. 9, pp. 90–95, 2007.

[30] Waskom, M. seaborn: statistical data visualization. *Journal Of Open Source Software*. **6**, 3021 (2021), https://doi.org/10.21105/joss.03021

[31] Biewald, L. Experiment Tracking with Weights and Biases. (2020), https://www.wandb.com/, Software available from wandb.com

[32] Cooper Flourens. *Identity Level Attributes for Explainable Facial Verification*. Master's thesis, University of Nevada, Reno, May 2024. Address: Reno, NV.

[33] W. McKinney, "Data structures for statistical computing in Python," in *Proceedings of the 9th Python in Science Conference*, S. van der Walt and J. Millman, Eds., 2010, pp. 56–61, doi: 10.25080/Majora-92bf1922-00a.

[34] Deng, J., Guo, J., Xue, N. & Zafeiriou, S. Arcface: Additive angular margin loss for deep face recognition. *Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition*. pp. 4690-4699 (2019)

[35] Kononenko, I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence In Medicine*. **23**, 89-109 (2001)

[36] Tsao, D. & Livingstone, M. Mechanisms of face perception. *Annu. Rev. Neurosci.*. **31** pp. 411-437 (2008)

[37] Cabeza, R. & Kato, T. Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science*. **11**, 429-433 (2000)

[38] Diego-Mas, J., Fuentes-Hurtado, F., Naranjo, V. & Alcañiz, M. The influence of each facial feature on how we perceive and interpret human faces. *I-Perception*. **11**, 2041669520961123 (2020)

[39] Rudd, E., Günther, M. & Boult, T. Moon: A mixed objective optimization network for the recognition of facial attributes. *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V 14*. pp. 19-35 (2016)

[40] Ji, W., He, K., Huo, J., Gu, Z. & Gao, Y. Unsupervised domain attention adaptation network for caricature attribute recognition. *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*. pp. 18-34 (2020)

[41] Li, W., Xiong, W., Liao, H., Huo, J., Gao, Y. & Luo, J. CariGAN: Caricature generation through weakly paired adversarial learning. *Neural Networks*. **132** pp. 66-74 (2020)

[42] Jang, W., Ju, G., Jung, Y., Yang, J., Tong, X. & Lee, S. StyleCariGAN: caricature generation via StyleGAN feature map modulation. *ACM Transactions On Graphics (TOG)*. **40**, 1-16 (2021)

[43] Shi, Y., Deb, D. & Jain, A. Warpgan: Automatic caricature generation. *Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition*. pp. 10762-10771 (2019)

[44] Khan, M., Hanif, M., Talib, R. & Others Caricature Face Photo Facial Attribute Similarity Generator. *Complexity*. **2022** (2022)

[45] Garg, J., Peri, S., Tolani, H. & Krishnan, N. Deep cross modal learning for caricature verification and identification (cavinet). *Proceedings Of The 26th ACM International Conference On Multimedia*. pp. 1101-1109 (2018)

[46] Abaci, B. & Akgul, T. Matching caricatures to photographs. *Signal, Image And Video Processing*. **9** pp. 295-303 (2015)

[47] Huang, G., Ramesh, M., Berg, T. & Learned-Miller, E. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. (University of Massachusetts, Amherst,2007,10)

[48] Lingenfelter, B., Davis, S. & Hand, E. A Quantitative Analysis of Labeling Issues in the CelebA Dataset. *International Symposium On Visual Computing*. pp. 129-141 (2022)

[49] Li, H., Hua, G., Lin, Z., Brandt, J. & Yang, J. Probabilistic elastic matching for pose variant face verification. *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition*. pp. 3499-3506 (2013)

[50] Heusch, G., Cardinaux, F. & Marcel, S. Lighting normalization algorithms for face verification. (IDIAP,2005)

[51] Yang, L., Ma, J., Lian, J., Zhang, Y. & Liu, H. Deep representation for partially occluded face verification. *EURASIP Journal On Image And Video Processing*. **2018**, 143 (2018)

[52] Wu, X., Xu, J., Wang, J., Li, Y., Li, W. & Guo, Y. Identity authentication on mobile devices using face verification and ID image recognition. *Procedia Computer Science*. **162** pp. 932-939 (2019)

[53] Alghamdi, A. A verification system for multi-factor authentication for E-healthcare architectures. *Arab Journal For Scientific Publishing (AJSP)*. **2663** pp. 5798 (2021)

[54] Yin, X., Yu, X., Sohn, K., Liu, X. & Chandraker, M. Feature transfer learning for face recognition with under-represented data. *Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition*. pp. 5704-5713 (2019)

[55] Tsitsoulis, A. & Bourbakis, N. A Methodology for Detecting Faces from Different Views. *2012 IEEE 24th International Conference On Tools With Artificial Intelligence*. **1** pp. 238-245 (2012)