A Comprehensive Analysis of AI Biases in Deepfake Detection With Massively Annotated Databases

Ying Xu, *Student Member, IEEE*, Philipp Terhöst, *Member, IEEE*, Kiran Raja, *Senior Member, IEEE*, Marius Pedersen, *Member, IEEE*

***Abstract*—In recent years, image and video manipulations with Deepfake have become a severe concern for security and society. Many detection models and datasets have been proposed to detect Deepfake data reliably. However, there is an increased concern that these models and training databases might be biased and, thus, cause Deepfake detectors to fail. In this work, we investigate the bias issue caused by public Deepfake datasets by (a) providing large-scale demographic and non-demographic attribute annotations of 47 different attributes for five popular Deepfake datasets and (b) comprehensively analysing AI-bias of three state-of-the-art Deepfake detection backbone models on these datasets. The investigation analyses the influence of a large variety of distinctive attributes (from over 65M labels) on the detection performance, including demographic (age, gender, ethnicity) and non-demographic (hair, skin, accessories, etc.) information. The results indicate that investigated databases lack diversity and, more importantly, show that the utilised Deepfake detection backbone models are strongly biased towards many investigated attributes. The Deepfake detection backbone methods, which are trained with biased datasets, might output incorrect detection results, thereby leading to generalizability, fairness, and security issues. We hope that the findings of this study and the annotation databases will help to evaluate and mitigate bias in future Deepfake detection techniques. The annotation datasets are publicly available[1]**

***Index Terms*—Deepfake, Deepfake detection, Databases, Bias, Fairness, Image manipulation, Video manipulation**

I. Introduction

Deepfake refers to a deep learning-based technique that is able to create fake videos/images by swapping the face of a person with the face of another. Due to the harmful usage of such fake content, such as fake news, fake pornography, or financial fraud, Deepfake has become a great concern for security and society [50]. Moreover, the availability of large-scale public face datasets and the development of strong generative artificial intelligence (AI), and especially deep learning techniques, such as Autoencoder or Generative Adversarial Networks (GAN) [15], [42] have strongly increased the realism of Deepfake. Various open-source and mobile applications [4], [2], [3] further allow to create highly realistic Deepfake videos or images without any expert knowledge and thus, make it possible for everyone to automatically manipulate images of videos with Deepfake technology.

Consequently, many works have developed detection methods capable of detecting such face manipulations [57]. Previous studies, however, pointed out some bias issues with these detection methods for different factors such as age, gender, and ethnicity [19], [51], [38]. The main reasons for bias in such AI models have been pointed out to originate from unbalanced training databases [19], [51], [38]. Biased decisions from detection approaches have a significant impact on both security and society if, for example, images from a certain group of people are always scrutinised as Deepfake.

To develop methods that are antagonistic to different factors leading to biases and report correct performance, there is a need for annotated datasets and consequently balanced performance metrics. This work therefore presents two main contributions to analyse bias-causing factors in Deepfake.

1) We provide massive and diverse annotations for five widely-used Deepfake detection datasets. Existing Deepfake detection datasets contain none or only sparse annotations restricted to demographic attributes, as shown in Table II. This work provides over 65.3M labels using 47 different attributes for five popular Deepfake detection datasets (Celeb-DF [30], DeepFakeDetection (DFD) [1], FaceForensics++ (FF++) [41], DeeperForensics-1.0 (DF-1.0) [22] and Deepfake Detection Challenge Dataset (DFDC) [12]).
2) We comprehensively analyse detection bias in three state-of-the-art Deepfake detection backbone models with respect to various demographic and non-demographic attributes regarding to four current Deepfake datasets. Previous investigations restricted their analysis to a maximum of four demographic attributes on a single dataset. Contrarily, we analyse detection bias on a much larger scale of distinctive attributes on four widely-used Deepfake datasets[2].

For the first contribution, five annotated datasets are created in the direction of earlier work using the MAAD-Face principle [47]. By computing a reliability score from the predictions of the MAAD classifier, we consider high-confidence predictions for labelling process to ensure a high annotation correctness. While the annotations from previous works at most contain demographic information like age, gender, and

Ying Xu and Philipp Terhöst contributed equally to this work.
e-mail:{ying.xu@ntnu.no}

[1]https://github.com/pterhoer/DeepFakeAnnotations

[2]For the analysis, we do not consider DF-1.0 data as the detection methods did not produce enough errors (Details in Table XIVTable XVTable XVI) on this dataset to analyse biased behaviours.

ethnicity, the annotations in this work are highly diverse and include attributes such as hair-color and -style, skin, face geometry, mouth, noise, and various accessories. We assert that these rich annotations will allow future works to evaluate the role of each attribute and use it to train better detection models that can mitigate bias issues.

The second contribution of our work is a detailed analysis of detection bias in Deepfake detection approaches by comparing the differential outcomes of three state-of-the-art Deepfake backbone networks (EfficientNetB0 [44], Xception [8], and Capsule-Forensics-v2 [36]) on four of the proposed Deepfake annotation datasets with respect to 31 demographic and non-demographic attributes[3].

The results indicate that the investigated datasets are highly imbalanced leading to highly biased detection backbone models when trained on such databases for a large variety of demographic and non-demographic attributes.

The observed bias in the detection backbone models can further explain the low generalisability of current Deepfake detectors [10], [53] across different attributes. Interestingly, the effect of the imbalanced attributes often differs in detection performance if the attribute is observed on a pristine (nonfake) image or a Deepfake. The results indicate that the detection backbone models learn several questionable factors that require a deeper investigation. For example, a person smiling or wearing a hat is strongly detected as a Deepfake despite being a pristine image. Depending on the application, these factors can lead to biases and, subsequently, strong fairness issues when a pristine video of a smiling woman is detected as Deepfake data. Conversely, a biased detection backbone model deciding a manipulated video as an unaltered video may lead to security implications. A complete list of such findings from our work alongwith the recommendations for future work is provided in Section VI.

II. Related Work

A. Deepfake Detection

There are two main approaches that are used to detect manipulated media. One focuses on the spatial features extracted from frames of a video. The other utilises temporal features among frames to capture falsified clues.

- **Spatial features:** Most of the early efforts to detect Deepfake have been made using spatial features extracted from video frames. Researchers have been working on detecting artifacts using unnatural facial features [28], [33], blending traces [29], [27], CNN-generated/GAN-generated fingerprints [16], [7]. Some studies have also been conducted in the frequency domain in order to detect artificial image contents [14], [40].
- **Temporal features:** Instead of individual frames, temporal features across frames have also been used recently, for example, unsynchronised color [34], [25], and phoney heartbeats appearing on faces [9], [39] and inconsistent facial information [54], [17], [59], [11].

[3]For experiments, we neglected attributes that are not frequently occurring to avoid wrong conclusions caused by limited testing data. Details can be found in the Appendix and Table IV

Most works have focused on developing Deepfake detection solutions tailored to available datasets. However, these solutions can be imbalanced, leading to bias and low generalisability across different demographic factors. We analyse four Deepfake detection approaches to demonstrate the biased performances for different demographic factors.

B. Deepfake Datasets

Table II shows seven popular Deepfake datasets that are popularly used for the development and evaluation of reliable Deepfake detection backbone models. DeepfakeTIMIT [**?**] and FFW [23] were published in 2018, followed by DFD [1] and FF++ [41], [26] in 2019. DFDC [12], Celeb-DF [30] and DF-1.0 [22] were released in 2020. Over the years, the size of the datasets has increased in terms of manipulations and the total number of images/videos. However, there are limited efforts to create more balanced datasets for gender and ethnicity. Both Celeb-DF and DF-1.0 maintain parity between males and females. Celeb-DF has a more extensive range of ages, while DF-1.0 holds balanced skin types. Despite these efforts, only a few databases provide additional annotations that could be utilised for developing Deepfake detection algorithms or testing these for influences of demographic factors. In contrast to previous works, we provide high-quality annotations for five popular databases for 47 demographic and non-demographic attributes. We hope to enable the development and evaluation of balanced and less-biased Deepfake detectors.

TABLE I: **Comparison of previous bias investigations in Deepfake detection** - this work provides a more comprehensive bias analysis involving more datasets up to 4 and more investigated attributes reaching the quantity of 47.

	Attributes	Models	Datasets
Hazirbas *et al.* [19]	4	2	1
Loc and Yan [51]	2	**3**	1
Pu *et al.* [38]	2	1	1
This work	**47**	**3**	**4**

C. Analysing Bias in Deepfake Detection

Internal representations of neural network models preserve attribute-related information of the training data even if it is not directly needed for the model objective [45], [46]. These encoded attribute patterns are reported to lead to biased performance in AI models [49]. Although there are many works on studying fairness in AI [35], [13], [49], only a few works analyse biases in the Deepfake detection field. Hazirbas *et al.* [19] measured the robustness of Deepfake detection backbone models across four primary dimensions: age, gender, apparent skin type, and lighting. They analysed the top five winners of the Deepfake Detection Challenge [43], [18], [37], [55], [21] for these attributes and concluded that all methods are biased towards lighter skin tones and fail in subjects with darker skin. Loc and Yan [51] measured the predictive performance of popular Deepfake detectors, MesoInception-4 [5], Xception [8] and Face X-Ray [27] on racially balanced datasets for gender and race. Significant

TABLE II: **Overview of popular Deepfake datasets and the proposed annotations databases** - While previous databases lack diverse annotations, the five proposed annotation databases close this gap and provide the resources needed to comprehensively analyse and mitigate bias in Deepfake detection backbone models.

			Number of videos		Number of frames		Annotated
	Dataset	Identities	Pristine	Forged	Pristine	Forged	Attributes
Previous works	DeepfakeTIMIT [**?**]	32	320	620	34.0k	68.0k	-
	FFW [23]	150	-	150	-	53k	-
	DeepFakeDetection (DFD) [1]	28	363	3,068	315.4k	2.2M	-
	FaceForensics++ (FF++) [41], [26]	1000	1,000	5,000	300k	1.5M	-
	Deepfake Detection Challenge Dataset (DFDC) [12]	960	23,654	104,500	7M	31M	-
	Celeb-DF [30]	59	590	5639	225.4k	2.1M	-
	DeeperForensics-1.0 (DF-1.0) [22]	100	50,000	10,000	2.9M	14.7M	1
	KoDF [24]	403	62,166	175,776	135M	65.9M	2
This work	**A-DFD**	28	363	3068	10.8k	89.6k	**47**
	A-FF++	1000	1000	5000	29.8k	149.1k	**47**
	A-DFDC	960	23,654	104,500	54.5k	52.6k	**47**
	A-Celeb-DF	59	590	5639	26.3k	166.5k	**47**
	A-DF-1.0	100	50000	10000	870.3k	321.5k	**47**

disparities were found in predictive performances across races and large representation bias in widely used FF++ [41]. Pu *et al.* [38] used a subset of the Face2Face dataset in FF++ and investigated MesoInception-4 to verify the existence of gender bias. Studying bias in Deepfake detection so far is limited to a few demographic factors such as gender and ethnicity. In contrast, this work analyses bias of three state-of-the-art Deepfake detection methods on four widely-used Deepfake datasets considering 31 demographic and non-demographic attributes as shown in Table I. With this work, we provide up to 47 attribute annotations on 4 popular Deepfake datasets. This work makes it possible to study the bias problem in a more comprehensive and reliable manner.

III. Methodology

To analyse different attributes, we first create large-scale annotations of 47 demographic and non-demographic attributes for five Deepfake detection databases. Following this, we conduct a comprehensive bias analysis of the state-of-the-art Deepfake detection methods on these annotated databases. In the following section, the process for creating the large-scale annotations is described, and methodology for measuring bias is presented.

A. Annotating Deepfake Databases

MAAD-Face classifier [47] makes use of a novel annotation-transfer technique that transfers the attribute annotations from several source databases to target databases. We make use of MAAD-Face classifier [47] classifier trained on LFW [20] and CelebA [31] as source databases. Our work, therefore, makes use of a novel annotation-transfer technique that transfers the attribute annotations from several source databases to target databases. We annotate five current Deepfake detection databases (DFD [1], FF++ [41], DFDC [12], Celeb-DF [30], and DF-1.0 [22]) in this work.

In the annotation process, each image is assigned with one of three possible labels for an attribute, positive (1), negative (-1), or undefined (0). A positive annotation for attribute a of an image means that the face in the image has attribute a. For instance, a face with a positive annotation for 'Young' represents a face of an young individual. In contrast, a negative annotation for attribute a of an image means that the face in the image does not possess attribute a. e further enforce a confidence-driven threshold to assert if an attribute cannot be classified. The confidence score is calculated based on the reliability measure from [48] and aims at preventing error-prone annotations. Specifically, if the classifier produces a decision for an attribute with a confidence below 90%, we annotate the attribute as undefined (0). We apply this methodology on five Deepfake detection datasets (DFD, FF++, DFDC, Celeb-DF, DF-1.0), resulting in the annotation datasets A-DFD (4.7M labels), A-FF++ (8.5M labels), A-DFDC (4.6M labels), A-Celeb-DF (9.2M labels) and A-DF-1.0 (38.3M labels), shown in Table II. These provide annotations for 47 attributes including information on demographics, skin, hair, beard, face geometry, mouth, nose, and accessories.

B. Measuring Bias on Unbalanced Data Testing

In this study, we assess the bias of a detection backbone model to an attribute a by comparing its performance when the attribute is present versus absent. However, there is a potential issue of an unbalanced distribution of positive and negative labelled testing samples during the experiments. To avoid inaccurate results caused by this imbalance, we introduce a corrected performance measure using control groups of positive and negative samples. We adopt the methods of creating two control groups for each attribute a by randomly selecting N samples from the testing data from [49], where N is the number of samples with or without attribute a. By doing so, we ensure that each control group has the same number of samples as their counterparts in the real data, thus making the positive and negative control groups independent of individual sample properties and drawn from the same distribution.

Comparing the classification performance of the positive and negative control groups for an attribute a allows us to

measure the effect of data imbalance on performance. If the performances of the negative and positive control groups are similar, the distribution of testing data does not significantly impact the performance. Contrarily, if the performances of the negative and positive control groups are dissimilar, the unbalanced testing data affects the classification performance. To measure the bias effect of an attribute a on the performance, we adopt the *relative performance (RP)* measure from [49]

$$RP_{type}(a) = 1 - \frac{err^{(+)}_{type}(a)}{err^{(-)}_{type}(a)}, \tag{1}$$

with $type = \{data, control\}$. $RP_{type}(a)$ measures the performance differences for an attribute a based on the error rates for a positive $err^{(+)}_{type}(a)$ and a negative $err^{(-)}_{type}(a)$ group. If the error rates are the same, $RP(a) = 0$ and, thus, attribute a does not affect performance. Positive RP values refer to lower error rates for the positive class (samples with this attribute). Contrarily, negative RP values refer to lower error rates for the negative class.

To correct this bias in the relative performance measure originating from the unbalanced testing data, we propose the *corrected relative performance* (CRP)

$$CRP(a) = RP_{data}(a) - RP_{control}(a) \tag{2}$$

which describes the difference between the relative performance of the real data RP_{data} and the relative performance of the control groups $RP_{control}$. The CRP measure simplifies to

$$CRP(a) = \frac{err^{(+)}_{control}(a)}{err^{(-)}_{control}(a)} - \frac{err^{(+)}_{data}(a)}{err^{(-)}_{data}(a)}, \tag{3}$$

and aims at removing the influence of the testing data distribution from the performance measure. If biased performance comes only from unbalanced test data, RP_{data} and $RP_{control}$ will be equal, and thus the corrected relative performance CRP will be zero. We use the $CRP(a)$ to measure the influence of the presence of attribute a on the performance and thus, to measure bias independently of the testing data parity.

IV. Experimental Setup

A. Database and Considered Attributes

For the experiments, we choose five widely-used Deepfake detection datasets, DFD [1], FF++ [41], [26], DFDC [12], Celeb-DF [30] and DF-1.0 [22]. Details for the different databases are provided in Table II. 30 frames are extracted from the first 300 frames of each video using a 10-frame interval. The faces are detected and aligned using MTCNN [56] for each of the frames. To ensure that enough data is available for analysing bias originating from specific attributes, we ignore attributes where a minimum of 100 positive or negative labelled images are unavailable. Out of the 47 attributes available in the annotated databases, only 31 were included in the bias analysis due to such a curation process. The specific details of this process can be found in Appendix Table IV.

B. Deepfake Detection BackBone Models

For the experiments, we choose three well used Deepfake detection backbone models, EfficientNetB0 [44], Xception [8], and Capsule-Forensics-v2 [36]. These three networks have been used frequently as backbone networks in the Deepfake detection [58], [27], [52], [32], [**?**], [40]. Therefore, we consider it reasonable to use them for the bias analysis. Furthermore, we have trained and evaluated the three backbone networks with a subject-exclusive train/dev/test for all the attributes. Due to the lack of a standardised protocol for all datasets, we spilt the datasets with a $60\%/20\%/20\%$ proportion for train/val/test respectively.

- **Xception** uses depth-wise separable convolutions to reduce the computational cost of traditional convolutions while maintaining high accuracy. This is achieved by performing spatial and channel convolutions separately, allowing for more efficient image feature processing. It is a highly effective deep learning architecture for image recognition tasks that require high accuracy and computational efficiency.
- **EfficientNet** is a model scaling method that uses a simple yet highly effective compound coefficient to scale up CNNs in a more structured manner, balancing the network's depth, width, and resolution to optimize its performance on a given resource budget. The architecture includes several novel features, including a mobile inverted bottleneck block, squeeze-and-excitation optimisation, and stochastic depth regularisation, further improving its performance. In our paper, we select the most lightweight version of EfficientNet, EfficientNetB0, to showcase its effectiveness.
- **Capsule-Forensics-v2** uses capsules to extract facial features and their spatial relationships from the input image to detect discrepancies. It incorporates a novel loss function that encourages disentangled representations, improving forgery detection accuracy. The model has demonstrated high effectiveness in detecting image manipulations, including copy-move, splicing, and face morphing.

C. Evaluation Metrics

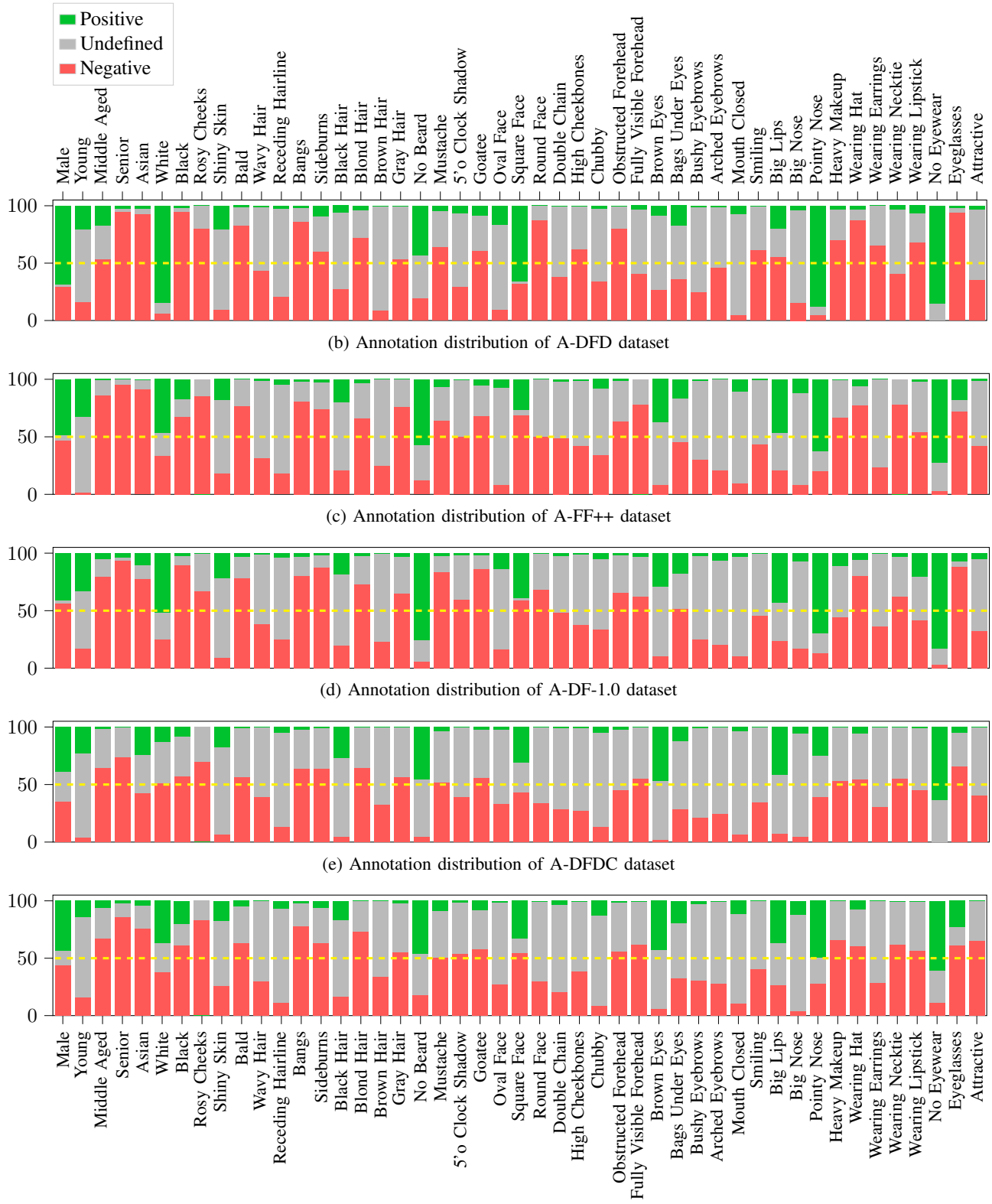
Previous work on Deepfake detection has reported its results based on a simple accuracy measure [51], [38]. However, dealing with unbalanced testing data is the norm, and a simple accuracy measure is vulnerable to this. We further notice many attributes being unbalanced in terms of the positive/negative labels from Figure 1. We, therefore, make use of a balanced accuracy measure, which computes the arithmetic mean of the sensitivity and specificity and is more robust to unbalanced data [6]. More precisely, we report the performances in terms of error rates (1-balanced accuracy) since this work investigates bias issues driven by inaccurate predictions.

V. Results

This section presents our findings on the presence of bias in Deepfake detection datasets utilising our proposed annotations. We analysed the relationship among various variables

Fig. 1: **Annotation distribution of the annotated Deepfake detection datasets** - The distributions of the proposed dataset annotations are shown with the y-axis presenting percentage. For each attribute, green indicates the percentage of positive annotations, red indicates the percentage of negatively annotations, and grey represents the percentage of images that have an undefined annotation for the attribute. The distributions show that these databases are highly unbalanced concerning these attributes.

(a) Annotation distribution of A-Celeb-DF dataset

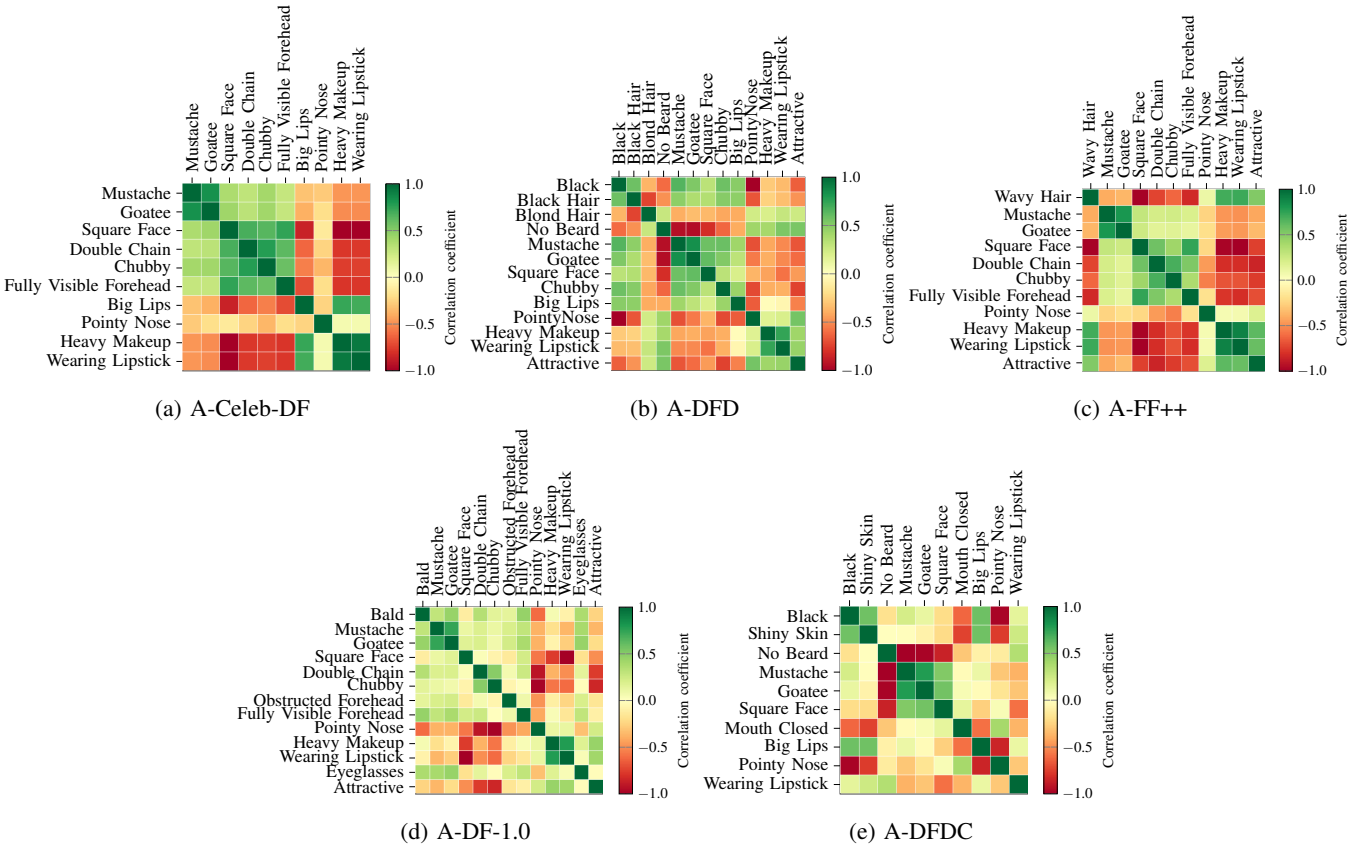


(b) Annotation distribution of A-DFD dataset

(c) Annotation distribution of A-FF++ dataset

(d) Annotation distribution of A-DF-1.0 dataset

(e) Annotation distribution of A-DFDC dataset



(a) A-Celeb-DF (b) A-DFD (c) A-FF++

(d) A-DF-1.0 (e) A-DFDC

Fig. 2: **Attribute annotation correlations of the Deepfake detection databases** - The 20 most positive and negative (Pearson) correlations are shown for each of the five databases. Green indicates positive correlations, while red indicates negative correlations. For working with these databases, the highly-correlating attributes should be considered to prevent misinterpretations.

regarding RP-vs-CRP and PDRP-vs-DDRP and used 24 plots to visualize these relationships. As we discuss these findings, it is crucial to keep in mind the concepts of causality and correlation in research and statistics. Causality refers to the relationship between cause and effect, while correlation measures the degree to which two or more variables are associated. It is important to note that correlation does not necessarily imply causality. Therefore, our results will focus on explaining correlation, and we will leave an in-depth exploration of causality for future studies.

A. Analysing Database Annotations

1) Attribute Statistics: Figure 1 shows the annotation distribution of the five annotated Deepfake detection datasets. For each attribute, green indicates the percentage of positive annotations, red indicates the percentage of negative annotations, and grey represents the percentage of images that have an undefined annotation for the attribute. According to the data given by Celeb-DF [30], this dataset contains male of 56.8% and female 43.2%, and in Figure 1a, the percentage of male (positive) is 70.15% and the percentage of female (negative) is 29.85%. The reason of the increased gap between gender asymmetry is that Celeb-DF only generates Deepfake videos using the same gender, so the number of differences between male and female are enlarged among the synthesised videos. We notice that most databases are quite balanced for the gender attribute.

However, there is a big imbalance with respect to skin colour, especially in Celeb-DF where people with white skin tones occupy the vast majority of this dataset. The big gap between numbers corresponds to the number disparity of Celeb-DF [30] (5.1% Asian, 6.8% African, and 88.1% Caucasian) which clarifies the high accuracy of the MAAD-classifier. The DFDC dataset appears to have a noticeable underrepresentation of individuals of Asian descent. Furthermore, there is a prevalence of white individuals in both the DFD and FF++ datasets. The variations in skin color distribution across different datasets may result in biases in the Deepfake detection system.

To conclude, it is clearly visible that the investigated Deepfake detection databases (DFD, FF++, DF-1.0, and DFDC) are strongly imbalanced with respect to most analysed attributes. Future work should consider creating balanced datasets to prevent any potential biases in Deepfake detection algorithms when such datasets are used for training.

2) Attribute Correlations: We present 20 most positive and negative pairwise attribute correlations in Figure 2 to understand the quality of the labels and potential biases in the attribute space. For instance, we notice in Figure 2a that

the attributes of Mustache and Goatee are highly correlating with each other. A high correlation also occurs between Heavy Makeup and Wearing Lipstick. This is easy to understand as the former attributes are mainly associated with males, and the latter ones mainly are with females. In contrast, Mustache and Goatee are negatively correlated to Heavy Makeup, and Wearing Lipsticks. Similar patterns are observable across different attribute correlations. The presence of negative correlations, such as the inverse relationship between *No Beard* and *Mustache*, as well as *Goatee*, highlights the quality of the annotations. It still should be noted that some correlations might also origin from the MAAD-classifier. Most of these correlations can be explained with background knowledge of the databases. For instance, the Celeb-DF dataset contains mainly images of celebrities in which these are presenting themselves to the camera. Therefore, a high correlation between Heavy Makeup and Wearing Lipstick is observed which may not necessarily represent real-world Deepfake of non-celebrities.

To conclude, our investigation has identified attribute pairs within the databases that exhibit strong correlations. It is imperative that future studies utilising these datasets and annotations consider these attribute correlations to avoid any misinterpretations that may result in biases. By acknowledging and accounting for these correlations, we can enhance the accuracy and fairness of any analysis or application of these databases.

3) Annotation Correctness: To evaluate the effectiveness of the proposed annotations, we have adopted Table III from the MAAD-classifier [47]. This table verifies the accuracy of the MAAD-classifier for the attributes utilised in our study. This table originates from [47] and shows the attribute correctness of the classifier with respect to three human evaluators. For each attribute, 100 images with and 100 images without this attribute were chosen randomly and shown to the evaluators to determine the true attribute label for each image. If the evaluators disagreed on an attribute, majority voting was used to decide on a label. Then, the accuracy, precision, and recall of the classifier predictions are calculated based on the ground truth provided by the human evaluators. The results are shown in Table III. For most attributes, the classifier agrees with human evaluators, resulting in an average accuracy of 92%, precision of 90%, and recall of 94%. Compared to similar facial annotation databases, such as LFW [20] (72% accuracy, 61% precision, 84% recall) and CelebA [31] (85% accuracy, 83% precision, 89% recall) [47], the proposed annotations are of high correctness.

The annotations provided in this work are of higher quality than the annotations provided for previous databases and we assert them to be suitable for analysing bias in Deepfake detection. Future works can make use of these attributes for developing and analysing bias-mitigating approaches in Deepfake detection.

TABLE III: **Annotation Correctness Study** - Annotation correctness of the utilized annotation generator is compared with the annotations of three human evaluators[47]. Compared to similar large-scale facial annotation classifiers used for databases, such as LFW [20] (72% annotation accuracy) and CelebA [31] (85% annotation accuracy), the proposed annotations are of high correctness [47] (92% annotation accuracy).

Category	Attribute	Accuracy	Precision	Recall
Demographics	Male	0.99	0.98	1.00
	Young	0.99	1.00	0.98
	Asian	0.90	0.88	0.92
	White	0.89	1.00	0.82
	Black	0.94	0.90	0.98
Skin	Shiny Skin	0.77	0.84	0.74
Hair	Bald	0.96	0.92	1.00
	Wavy Hair	0.99	1.00	0.98
	Receding Hairline	0.77	0.54	1.00
	Bangs	0.98	0.96	1.00
	Black Hair	0.98	0.96	1.00
	Blond Hair	1.00	1.00	1.00
Beard	No Beard	0.98	1.00	0.96
	Mustache	0.98	0.98	0.98
	Goatee	0.95	0.90	1.00
Face Geometry	Oval Face	0.81	0.90	0.76
	Square Face	0.80	0.78	0.81
	Double Chin	0.94	0.88	1.00
	Chubby	0.94	0.88	1.00
	Obstructed Forehead	0.91	0.94	0.89
	Fully Visible Forehead	0.80	0.75	1.00
Mouth	Mouth Closed	0.84	0.80	0.87
	Smiling	0.95	1.00	0.91
	Big Lips	0.70	0.50	0.83
Nose	Big Nose	0.97	0.98	0.96
	Pointy Nose	0.88	0.88	0.88
Accessories	Heavy Makeup	0.98	0.98	0.98
	Wearing Hat	0.92	0.84	1.00
	Wearing Lipstick	0.95	0.90	1.00
	No Eyewear	0.98	0.98	0.98
	Eyeglasses	0.90	0.80	1.00
Other	Attractive	1.00	1.00	1.00
		0.92	0.90	0.94

B. Analysing Bias in Deepfake Detection

To understand the bias in Deepfake detection, we will first study the general detection performance in presence of several potentially imbalanced attributes and secondly, analysing the detection performance in presence of these attributes separately on pristine and fake data. We exclude DF-1.0 dataset as the detection methods did not produce high enough errors necessary to analyse biased behaviours. The detailed results are shown in Appendices Table XIV, Table XV, and Table XVI due to page limits.

1) Investigating General Bias Issues: This section analyses the general bias issues in Deepfake detection based on RP-vs-CRP plots as shown in Figure 3. In these plots, the relative performance (RP) for each attribute is shown with respect to the corrected relative performance (CRP). As mentioned in Section III-B, RP describes the ratio of the performance for images with a certain attribute versus the performance without this attribute. Consequently, $RP(a) = -100\%$ for an attribute a means that the error is twice as high if the image has this attribute than without it. Since the testing data is imbalanced for many attributes, the CRP was introduced

in Section III-B to remove the influence of data imbalance. Consequently, attributes that lie in the top area (green) of the RP-vs-CRP plots indicate an increased detection performance and, contrarily, attributes that lie in the bottom (red) indicate increased detection errors. Moreover, each plot contains a bisectrix line where the attributes close to this line are less affected by imbalanced testing data than attributes away from it.

The RP-vs-CRP plots in Figure 3 are shown for three models on four Deepfake detection databases. The plots show strong influences of most of the investigated attributes on the performance, indicating strongly biased Deepfake detectors. For instance, the analysis of EfficientNetB0 on Celeb-DF shows that having a big nose/big lips/or being black or chubby leads to more than twice the detection errors compared to images without these attributes. This shows serious fairness issues of these models when these are applied in real-world applications for specific category of people. In general, most attributes can be observed as strong factors leading to unfair performance differences in DeepFake detection.

Moreover, we have observed that training Deepfake detection backbone models on various datasets results in significant variations in the influence of attributes on detection performance. For instance, the misclassification of the pattern *Obstructed Forehead* is observed in the Celeb-DF and DFDC datasets. This finding suggests that both the selection of Deepfake detection backbone networks and the choice of datasets may significantly impact bias in the system.

To conclude, the experimental results demonstrate that the analysed Deepfake detection backbone models are strongly biased against a variety of demographic and non-demographic attributes. The variation of the biased performances across the models and databases indicates that this bias originates from several sources such as unbalanced training data, the utilised network, and their training process. The observed attribute-related variation in performances shows a strong need for mitigating bias in Deepfake detection models.

2) Investigating Bias Issues in Pristine and Fake Data: To investigate the bias issues in DeepFake detection in more detail, we conduct another analysis for pristine and fake data individually in this section. The results of this analysis is shown in Figure 4, for three Deepfake detection backbone models on four databases. The pristine data relative performance ($PDRP$) refers to the CRP that is only evaluated on pristine data and, analogous, the Deepfake data relative performance ($DDRP$) refers to the CRP that is calculated on fake data only. For an attribute a, a negative CRP on the pristine data means that people having this attribute are more likely to be falsely detected as fakes than people without these attributes. A negative CRP on fake data means that fake images that are generated with such an attribute are less likely to be detected as fake and thus, demonstrate weak points that attackers are likely to exploit. Each plot also contains bisectrix line where attributes that lie close to this line have a similar affect on pristine data than on Deepfake. Attributes placed above this line have a higher CRP on the Deepfake than on the pristine data. Conversely, attributes below this line have a higher CRP on prisitne data than on the Deepfake.

The results clearly show that the effect of the investigated attributes on the detection performance strongly differ between pristine and fake data since most attributes lie far away from the bisectrix line. Analysing the four quadrants of the plots shows that in most cases the attributes are distributed in all four. Attributes in quadrant I (top right), indicate that the attributes have the same positive effect on the performance, while attribute in quadrant III (bottom left), have the same negative effect on the detection performance. Observed performance in these areas indicates a similar biased effect on the decision. Attributes in quadrant II (top left) and IV (bottom right) show the opposite effect on the detection performance on pristine and Deepfake data. Consequently, for attributes in these areas, the model learnt the critical assumptions that the presence of the attribute is an indicator for Deepfake detection decision. For instance, the analysis of EfficientNetB0 on Celeb-DF for attribute wearing hat, shows a positive $PDPR \approx 100\%$ and a negative $DDPR \approx -75\%$. Consequently, if a real person is wearing a hat the model comes twice as often to the right decision than if the person is not wearing a hat. Conversely, if a Deepfake image with hat is analysed by the model it leads to nearly twice as many errors as without a hat. The model in this case has seemingly learnt the presence of a hat as a strong indicator for the pristine data. Such observations point to questionable assumptions learned by the network and can result in biased detection performance.

In general, the observations reveal similar trends and patterns corresponding to the investigation from Section V-B1. The biased performances for the different attributes vary across the utilised models and training databases. If a real person has a goatee, a big nose, is chubby, male, or black, the probability that the model (EfficientNetB0 on Celeb-DF) comes to a wrong decision is doubled compared to persons without these attributes. The detection models therefore show strong biases leading to fairness issues in real-life applications if deployed without considering the attribute distribution. It should also be kept in mind that this analysis limits its investigation to the influence of single attributes on the detection performance. The analysis of multiple attributes can be asserted to lead to an exponential increase in its bias effects. However, this aspect is not considered in this work.

To conclude, the impact of biased performance for the analysed attributes on detection accuracy varies significantly between pristine and fake data for several attributes. The results suggest that the models learn several questionable assumptions that the presence of a certain attribute, such as if the person is smiling or wears a hat, is an strong indicator for Deepfake detection decision. Lastly, the investigated Deepfake detection backbone models have demonstrated unfair behavior, with a significant increase in the probability of making incorrect decisions when presented with specific attributes such as having a big nose or belonging to a particular gender or race. This bias in current Deepfake detectors affects their accuracy and limits their generalisability. In other words, these biases may cause the Deepfake detectors to perform well on certain datasets or scenarios, but may fail to perform effectively in others, especially those where such attributes are different or not present. Therefore, addressing these biases and improving

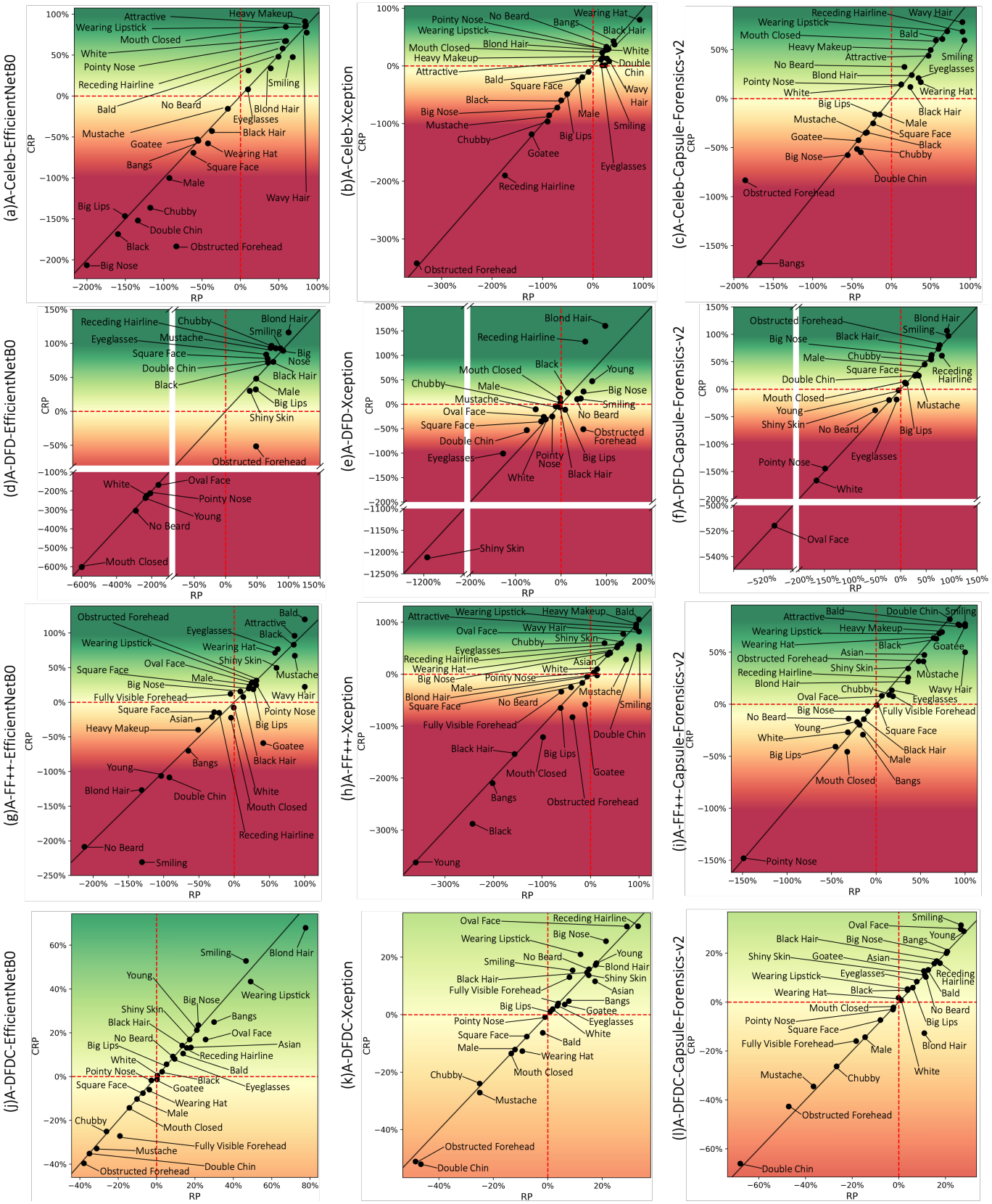


Fig. 3: **Bias analysis** - The relative performance RP is reported with respect to the corrected relative performance CRP using three Deepfake detection backbone models, EfficientNetB0 [44], Xception [8], and Capsule-Forensics-v2 [36] on four annotated databases, A-Celeb-DF, A-DFD, A-FF++, and A-DFDC. Many attributes strongly influence the detection performance.

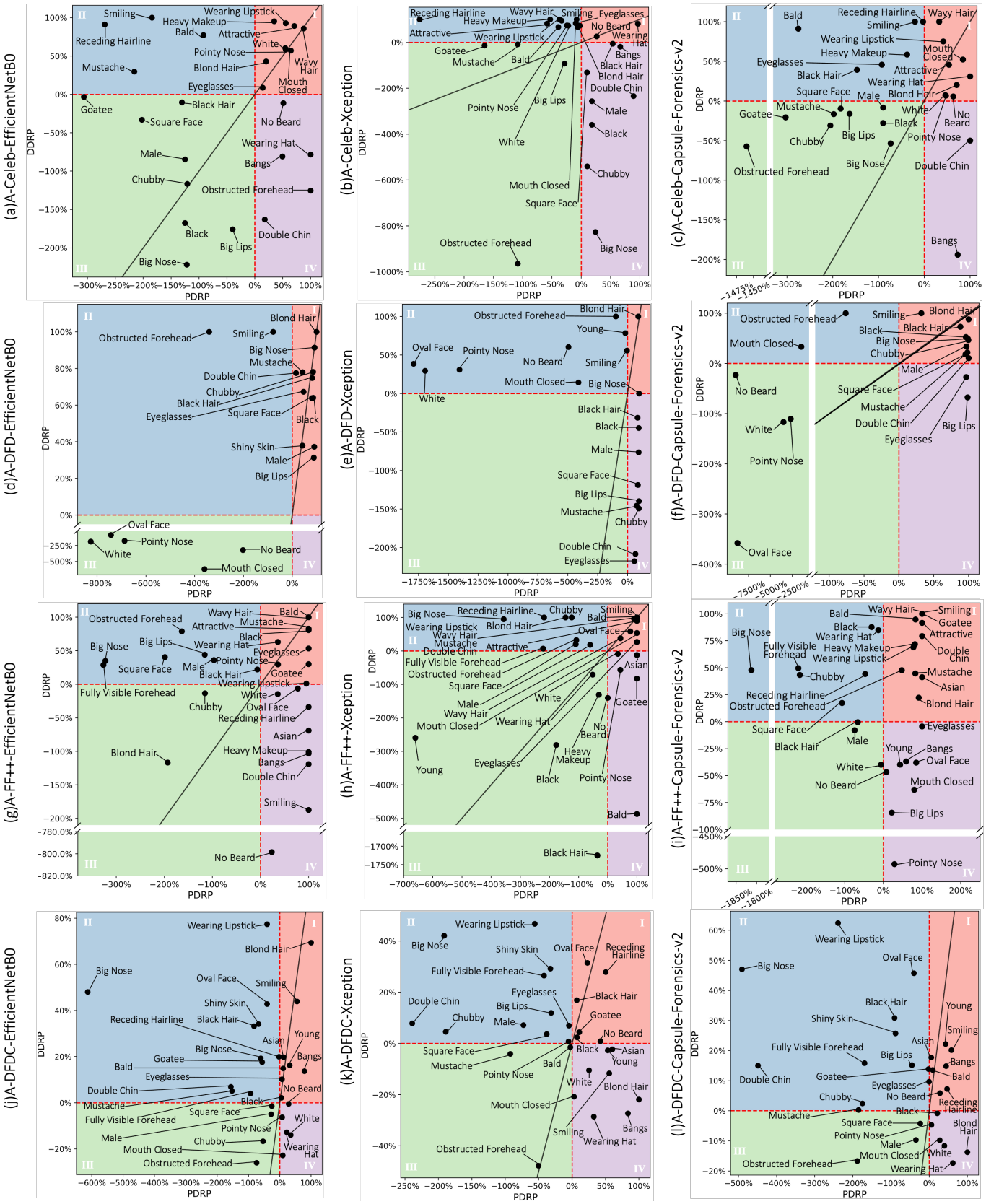


Fig. 4: **Bias analysis on Pristine Data and Deepfake Data** - The CRP on the pristine data (PDRP) is reported with respect to the CRP on Deepfake data (DDRP) using three Deepfake detection backbone models, EfficientNetB0 [44], Xception [8], and Capsule-Forensics-v2 [36] on our four annotated databases, A-Celeb-DF, A-DFD, A-FF++, and A-DFDC. Many of the attributes strongly influence the detection performance.

the generalisability of Deepfake detectors is crucial to ensure their robustness and reliability in real-world applications.

VI. Key Findings & Recommendations For Future Works

In the following, we summarise our key findings from our bias investigation of three Deepfake detection backbone models in four databases with respect to 31 demographic and non-demographic attributes:

- **Deepfake detection databases and strong imbalance** - The investigated Deepfake detection databases (Celeb-DF, DFD, FF++, and DFDC) lack diversity with respect to most analysed attributes. Future work should aim to provide more unbiased, balanced, and divers datasets to prevent the development of potential biased Deepfake detection algorithms.
- **Current Deepfake detection databases contain some strongly correlating attribute pairs** - Future works using these databases (or our annotations) should take into account that some attributes show strong pairwise correlations to prevent misinterpretations in their results.
- **Deepfake detection backbone networks and demographic/non-demographic attributes** - The results demonstrate that the analysed Deepfake detection backbone models are strongly biased for a variety of demographic and non-demographic attributes. The variation of the biased performances across the models and databases indicates that this bias might origin from several sources such as imbalanced training data, the utilised network, and their training process. These omnipresent biases might also be the reason for low generalisability of current Deepfake detection methods. This shows a strong need for developing bias-mitigating Deepfake detection solutions for future works.
- **Bias in attributes for pristine and Deepfake data** - For many of the investigated attributes, the biased performance similarly affects the pristine and Deepfake data. However, also the strong opposite behaviour was observed for many attributes leading the models to learning potentiality wrong assumptions.
- **Deepfake detection backbone models and questionable assumptions** - The results suggest that the model tend to learn questionable assumptions where the presence of a certain attribute, such as if the person is smiling or wears a hat, is a strong indicator for Deepfake. Although this may have originated from the training data, our analysis is limited and indicates it as a potential topic in future works to enhance the reliability of these systems.
- **Deepfake detection backbone models and societal security** - The presence of a certain attribute in a Deepfake image resulted in an increased error rate, several times higher than for a Deepfake without this attribute. Attackers will likely exploit these issues to increase their chances to overcome Deepfake detection if unaddressed. On the other hand, the strong performance differences based on the presence of an attribute show a strong unfairness of these models. Future works therefore should focus on mitigating bias problems for Deepfake detection for the sake of security and society.

Based on the key observations of the three backbone networks analysed, there appears to be a significant research gap in developing Deepfake detection methods suitable for real-world applications. However, further analysis of additional methods may be necessary to make a more definitive statement. Our analysis points to a need for more diverse and richly annotated databases for training and testing, as well as developing bias-mitigating Deepfake detection approaches.

VII. Conclusion

In this work, we provided large-scale annotations for five popular Deepfake detection datasets and used these to comprehensively analyse bias in Deepfake detection. While existing Deepfake detection databases are only sparsely annotated, we closed this gap by making over 65.3M annotations of 47 different attributes for five Deepfake detection datasets publicly available. Based on these datasets, we comprehensively analyse bias-causing factors in Deepfake detection purely from an attribute perspective. The results indicated that both the datasets as well as the state-of-the-art AI-based Deepfake detectors trained on this data, demonstrate strong bias issues for many demographic and non-demographic attributes. Depending on the use case, the biased performance can result in serious societal fairness and security problems. Moreover, imbalanced attributes in these datasets can further lead to generalization problems across different attributes in current Deepfake detection algorithms. Our findings from the study and proposed publicly-available annotations are expected to help future works to effectively evaluate and mitigate bias issues in Deepfake detection and thus, to develop reliable Deepfake detectors.

Acknowledgment

Parts of this work was carried out during the tenure of an ERCIM 'Alain Bensoussan' Fellowship Programme.

References

[1] Contributing data to deepfake detection research. https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html. 24 September, 2019.
[2] Faceapp: face editor. https://www.faceapp.com/. Accessed: 2022-08-02.
[3] Faceswap: Open source multi-platform deepfakes software. https://faceswap.dev/. Accessed: 2022-08-02.
[4] Reface: face swapp videos. https://hey.reface.ai/. Accessed: 2022-08-02.
[5] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen. Mesonet: a compact facial video forgery detection network. In *2018 IEEE international workshop on information forensics and security (WIFS)*, pages 1–7. IEEE, 2018.
[6] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann. The balanced accuracy and its posterior distribution. In *2010 20th international conference on pattern recognition*, pages 3121–3124. IEEE, 2010.
[7] L. Chai, D. Bau, S.-N. Lim, and P. Isola. What makes fake images detectable? understanding properties that generalize. In *European Conference on Computer Vision*, pages 103–120. Springer, 2020.
[8] F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.
[9] U. A. Ciftci, I. Demir, and L. Yin. Fakecatcher: Detection of synthetic portrait videos using biological signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
[10] D. Cozzolino, M. Nießner, and L. Verdoliva. Audio-visual person-of-interest deepfake detection. *CoRR*, abs/2204.03083, 2022.

[11] D. Cozzolino, A. Rössler, J. Thies, M. Nießner, and L. Verdoliva. Id-reveal: Identity-aware deepfake video detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15108–15117, 2021.

[12] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer. The deepfake detection challenge (dfdc) dataset. *arXiv preprint arXiv:2006.07397*, 2020.

[13] M. Du, F. Yang, N. Zou, and X. Hu. Fairness in deep learning: A computational perspective. *IEEE Intelligent Systems*, 36(4):25–34, 2020.

[14] R. Durall, M. Keuper, F.-J. Pfreundt, and J. Keuper. Unmasking deepfakes with simple features. *arXiv preprint arXiv:1911.00686*, 2019.

[15] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 2672–2680, 2014.

[16] L. Guarnera, O. Giudice, and S. Battiato. Deepfake detection by analyzing convolutional traces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 666–667, 2020.

[17] A. Haliassos, K. Vougioukas, S. Petridis, and M. Pantic. Lips don't lie: A generalisable and robust approach to face forgery detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5039–5049, 2021.

[18] W. Z. Hanqing Zhao, Hao Cui. https://github.com/cuihaoleo/kaggle-dfdc.

[19] C. Hazirbas, J. Bitton, B. Dolhansky, J. Pan, A. Gordo, and C. C. Ferrer. Towards measuring fairness in ai: the casual conversations dataset. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2021.

[20] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[21] I. P. James Howard. https://github.com/jphdotam/DFDC/.

[22] L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy. DeeperForensics-1.0: A large-scale dataset for real-world face forgery detection. In *CVPR*, 2020.

[23] A. Khodabakhsh, R. Ramachandra, K. Raja, P. Wasnik, and C. Busch. Fake face detection methods: Can they be generalized? In *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–6. IEEE, 2018.

[24] P. Kwon, J. You, G. Nam, S. Park, and G. Chae. Kodf: A large-scale korean deepfake detection dataset. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10744–10753, 2021.

[25] H. Li, B. Li, S. Tan, and J. Huang. Identification of deep network generated images using disparities in color components. *Signal Processing*, 174:107616, 2020.

[26] L. Li, J. Bao, H. Yang, D. Chen, and F. Wen. Faceshifter: Towards high fidelity and occlusion aware face swapping. *arXiv preprint arXiv:1912.13457*, 2019.

[27] L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen, and B. Guo. Face x-ray for more general face forgery detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5001–5010, 2020.

[28] Y. Li, M.-C. Chang, and S. Lyu. In ictu oculi: Exposing ai generated fake face videos by detecting eye blinking. *arXiv preprint arXiv:1806.02877*, 2018.

[29] Y. Li and S. Lyu. Exposing deepfake videos by detecting face warping artifacts. *arXiv preprint arXiv:1811.00656*, 2018.

[30] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu. Celeb-df: A large-scale challenging dataset for deepfake forensics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3207–3216, 2020.

[31] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015.

[32] Y. Luo, Y. Zhang, J. Yan, and W. Liu. Generalizing face forgery detection with high-frequency features. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16317–16326, 2021.

[33] F. Matern, C. Riess, and M. Stamminger. Exploiting visual artifacts to expose deepfakes and face manipulations. In *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, pages 83–92. IEEE, 2019.

[34] S. McCloskey and M. Albright. Detecting gan-generated imagery using saturation cues. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 4584–4588. IEEE, 2019.

[35] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6):1–35, 2021.

[36] H. Nguyen, J. Yamagishi, and I. Echizen. Use of a capsule network to detect fake images and videos. arxiv 2019. *arXiv preprint arXiv:1910.12467*.

[37] NTech-Lab. https://github.com/NTech-Lab/deepfake-detection-challenge.

[38] M. Pu, M. Y. Kuan, N. T. Lim, C. Y. Chong, and M. K. Lim. Fairness evaluation in deepfake detection models using metamorphic testing. *arXiv preprint arXiv:2203.06825*, 2022.

[39] H. Qi, Q. Guo, F. Juefei-Xu, X. Xie, L. Ma, W. Feng, Y. Liu, and J. Zhao. Deeprhythm: Exposing deepfakes with attentional visual heartbeat rhythms. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 4318–4327, 2020.

[40] Y. Qian, G. Yin, L. Sheng, Z. Chen, and J. Shao. Thinking in frequency: Face forgery detection by mining frequency-aware clues. In *European Conference on Computer Vision*, pages 86–103. Springer, 2020.

[41] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner. Faceforensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[42] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

[43] S. Seferbekov. https://github.com/selimsef/dfdc_deepfake_challenge.

[44] M. Tan and Q. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.

[45] P. Terhörst, D. Fährmann, N. Damer, F. Kirchbuchner, and A. Kuijper. Beyond identity: What information is stored in biometric face templates? In *2020 IEEE International Joint Conference on Biometrics, IJCB 2020, Houston, TX, USA, September 28 - October 1, 2020*, pages 1–10. IEEE, 2020.

[46] P. Terhörst, D. Fährmann, N. Damer, F. Kirchbuchner, and A. Kuijper. On soft-biometric information stored in biometric face embeddings. *IEEE Trans. Biom. Behav. Identity Sci.*, 3(4):519–534, 2021.

[47] P. Terhörst, D. Fährmann, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper. Maad-face: A massively annotated attribute dataset for face images. *IEEE Transactions on Information Forensics and Security*, 16:3942–3957, 2021.

[48] P. Terhörst, M. Huber, J. N. Kolf, I. Zelch, N. Damer, F. Kirchbuchner, and A. Kuijper. Reliable age and gender estimation from face images: Stating the confidence of model predictions. In *10th IEEE International Conference on Biometrics Theory, Applications and Systems, BTAS 2019, Tampa, FL, USA, September 23-26, 2019*, pages 1–8. IEEE, 2019.

[49] P. Terhörst, J. N. Kolf, M. Huber, F. Kirchbuchner, N. Damer, A. M. Moreno, J. Fierrez, and A. Kuijper. A comprehensive study on face recognition biases beyond demographics. *IEEE Transactions on Technology and Society*, 3(1):16–30, mar 2022.

[50] R. Tolosana, R. Vera-Rodríguez, J. Fiérrez, A. Morales, and J. Ortega-Garcia. Deepfakes and beyond: A survey of face manipulation and fake detection. *Inf. Fusion*, 64:131–148, 2020.

[51] L. Trinh and Y. Liu. An examination of fairness of ai models for deepfake detection. *arXiv preprint arXiv:2105.00558*, 2021.

[52] Y. Xu, K. Raja, L. Verdoliva, and M. Pedersen. Learning pairwise interaction for generalizable deepfake detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 672–682, 2023.

[53] Y. Xu, K. B. Raja, and M. Pedersen. Supervised contrastive learning for generalizable and explainable deepfakes detection. In *IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, WACV - Workshops, Waikoloa, HI, USA, January 4-8, 2022*, pages 379–389. IEEE, 2022.

[54] X. Yang, Y. Li, and S. Lyu. Exposing deep fakes using inconsistent head poses. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8261–8265. IEEE, 2019.

[55] T. E. years old. https://github.com/Siyu-C/RobustForensics.

[56] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23(10):1499–1503, 2016.

[57] T. Zhang. Deepfake generation and detection, a survey. *Multim. Tools Appl.*, 81(5):6259–6276, 2022.

[58] H. Zhao, W. Zhou, D. Chen, T. Wei, W. Zhang, and N. Yu. Multi-attentional deepfake detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2185–2194, 2021.

[59] Y. Zheng, J. Bao, D. Chen, M. Zeng, and F. Wen. Exploring temporal

coherence for more general video face forgery detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 15044–15054, 2021.

APPENDIX

Reporting the results of our comprehensive bias investigation resulted in a large amount of results. The main paper focused on reporting these results in a condensed manner (e.g. in terms of relative performance measures) that allow a more easy interpretation and distillation of knowledge. Since the detailed results of our analysis might be interesting for the research community, we will use this supplementary material to report the results in more detail.

To analyse bias originating from specific attributes, we employed a filtering process that excluded attributes with fewer than 100 positive or negative labelled images. As a result, only 31 out of the 47 available attributes in the annotated databases were included in the bias analysis. A detailed account of this process can be found in Table IV, which lists all 47 attributes across 5 annotated datasets and denotes those with less than 100 labeled images using the symbol "x". Attributes that were deemed insignificant for more than two datasets were eliminated from the analysis, resulting in the selection of 31 attributes for further study. Furthermore, due to their low annotation accuracy of 0.68 as indicated in Table III, we decided to exclude the attributes of *Brown Eyes* and *Bags Under Eyes*.

More precisely, we report the balanced detection error as well as the error on the pristine and Deepfake data for each of the 47 attributes on five datases using three state-of-the-art Deepfake detection backbone models. As explained in Section III-B, the performance of the positive and negative groups of the data is reported as well as of the control groups. This leads to a large amount of results that we report 15 pages of tables in this supplementary. These are shown from Table V to Table XIX. Each table refers to the combination of one (out of three) models with one (out of five) databases. For attributes that were neglected for the experiments due to the low number of labels (see Section IV-A) no results are shown to avoid unreliable statements. In some cases, the detector predicted all samples wrongly (shown as "all wrong") or correctly (shown as "all correct"). The latter case happens constantly for the DF-1.0 database. This can be seen in Tables V,XV, and XVI. Since the detectors are barely not making any errors, not enough information was provided to study bias on this databases (it was too easy). Therefore, we neglected this database from our experiments as described in Section I.

TABLE IV: **Metrics for identifying pertinent attributes.** The table includes all the 47 attributes in 5 annotated datasets and marks those that have less than 100 positive or negative labelled images with **x**. Attributes not deemed valuable for more than two datasets are excluded from the analysis, resulting in the selection of 31 attributes for further study. In addtion, attributes Brown Eyes and Bags Under Eyes are excluded due to low annotation accuracy.

	A-Celeb-DF	A-DFD	A-FF++	A-DF-1.0	A-DFDC
Male					
Young	x				
Middle Aged	x	x		x	
Senior	x	x		x	
Asian	x	x			
White				x	
Black					
Rosy Cheeks	x	x	x	x	x
Shiny Skin	x				
Bald		x		x	
Wavy Hair		x		x	x
Receding Hairline					
Bangs					
Sideburns	x		x	x	
Black Hair				x	
Blond Hair				x	
Brown Hair	x	x	x	x	x
Gray Hair	x	x		x	
No Beard				x	
Mustache				x	
5'o Clock Shadow	x	x	x	x	
Goatee		x		x	
Oval Face	x			x	
Square Face					
Round Face	x	x	x	x	
Double Chin				x	
High Cheekbones	x	x	x	x	
Chubby					
Obstructed Forehead					
Fully Visible Forehead	x	x			
Brown Eyes				x	
Bags Under Eyes					
Bushy Eyebrows	x	x		x	
Arched Eyebrows	x	x		x	
Mouth Closed					
Smiling					
Big Lips					
Big Nose					
Pointy Nose					
Heavy Makeup		x		x	x
Wearing Hat		x			
Wearing Earrings	x	x	x	x	x
Wearing Necktie	x	x		x	
Wearing Lipstick		x		x	
No Eyewear	x	x		x	
Eyeglasses					
Attractive		x		x	x

TABLE V: EfficientNetB0/Celeb-DF - Experiments with EfficientNetB0 on the Celeb-DF dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	15.01%	9.34%	6.87%	2.94%	23.16%	16.76%
		Negative	7.79%	10.09%	3.05%	3.59%	12.53%	16.59%
		Rel. Perf.	-92.70%	7.41%	-125.36%	18.07%	-84.76%	-1.01%
	Young	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Asian	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	White	Positive	7.94%	10.10%	3.12%	3.59%	12.76%	16.55%
		Negative	19.52%	9.35%	6.87%	2.94%	32.17%	17.50%
		Rel. Perf.	59.33%	-7.92%	54.63%	-21.88%	60.34%	5.42%
	Black	Positive	21.32%	9.30%	6.87%	2.94%	35.77%	18.03%
		Negative	8.21%	10.19%	3.05%	3.59%	13.37%	16.56%
		Rel. Perf.	-159.72%	8.78%	-125.36%	18.07%	-167.56%	-8.86%
Skin	Shiny Skin	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Hair	Bald	Positive	4.62%	10.07%	5.77%	3.53%	3.47%	20.14%
		Negative	9.13%	10.21%	3.00%	3.59%	15.27%	16.57%
		Rel. Perf.	49.41%	1.33%	-92.09%	1.89%	77.25%	-21.54%
	Wavy Hair	Positive	2.14%	8.57%	0.83%	6.67%	3.45%	20.69%
		Negative	15.59%	9.39%	6.81%	2.95%	24.37%	16.43%
		Rel. Perf.	86.27%	8.78%	87.77%	-125.94%	85.85%	-25.91%
	Receding Hairline	Positive	4.72%	9.78%	7.78%	5.00%	1.67%	18.33%
		Negative	10.49%	9.49%	2.11%	3.21%	18.88%	17.55%
		Rel. Perf.	54.99%	-3.09%	-269.11%	-55.93%	91.17%	-4.46%
	Bangs	Positive	15.98%	10.24%	1.95%	3.75%	30.00%	18.33%
		Negative	10.23%	9.91%	3.86%	3.55%	16.59%	16.67%
		Rel. Perf.	-56.24%	-3.35%	49.38%	-5.39%	-80.81%	-10.00%
	Black Hair	Positive	10.90%	9.31%	8.18%	3.48%	13.61%	19.72%
		Negative	7.94%	9.85%	3.55%	3.38%	12.32%	16.82%
		Rel. Perf.	-37.32%	5.41%	-130.74%	-3.16%	-10.44%	-17.24%
	Blond Hair	Positive	7.15%	9.38%	3.05%	3.16%	11.24%	17.80%
		Negative	11.77%	9.92%	3.83%	3.37%	19.71%	16.65%
		Rel. Perf.	39.26%	5.37%	20.14%	6.36%	42.97%	-6.89%
Beard	No Beard	Positive	8.05%	10.11%	3.05%	3.59%	13.06%	16.87%
		Negative	8.99%	8.37%	6.24%	3.40%	11.73%	18.25%
		Rel. Perf.	10.40%	-20.76%	51.15%	-5.57%	-11.28%	7.58%
	Mustache	Positive	9.44%	10.27%	9.64%	3.39%	9.24%	19.89%
		Negative	8.10%	10.17%	3.05%	3.59%	13.16%	16.47%
		Rel. Perf.	-16.48%	-0.94%	-216.06%	5.80%	29.76%	-20.76%
	Goatee	Positive	13.79%	10.23%	12.33%	3.67%	15.25%	18.86%
		Negative	8.90%	10.19%	3.03%	3.58%	14.77%	16.68%
		Rel. Perf.	-54.94%	-0.37%	-306.48%	-2.43%	-3.25%	-13.12%
Face Geometry	Oval Face	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Square Face	Positive	14.65%	9.39%	9.18%	3.20%	20.12%	16.68%
		Negative	9.07%	10.18%	3.03%	3.64%	15.11%	16.73%
		Rel. Perf.	-61.41%	7.78%	-202.47%	12.07%	-33.10%	0.33%
	Double Chin	Positive	19.53%	8.27%	2.22%	5.00%	36.84%	22.81%
		Negative	8.36%	10.12%	2.70%	3.57%	14.01%	16.59%
		Rel. Perf.	-133.72%	18.28%	17.82%	-40.06%	-162.97%	-37.44%
	Chubby	Positive	16.96%	8.12%	5.81%	3.20%	28.12%	18.59%
		Negative	7.80%	10.02%	2.63%	3.62%	12.98%	16.66%
		Rel. Perf.	-117.42%	18.97%	-121.13%	11.51%	-116.67%	-11.64%
	Obstructed Forehead	Positive	16.67%	0.00%	0.00%	0.00%	33.33%	0.00%
		Negative	9.07%	10.20%	3.36%	3.61%	14.79%	16.73%
		Rel. Perf.	-83.67%	100.00%	100.00%	100.00%	-125.37%	100.00%
	Fully Visible Forehead	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Mouth	Mouth Closed	Positive	4.34%	9.93%	1.78%	3.57%	6.90%	21.72%
		Negative	10.57%	7.91%	5.03%	2.97%	16.11%	20.67%
		Rel. Perf.	58.94%	-25.59%	64.55%	-19.85%	57.19%	-5.11%
	Smiling	Positive	2.94%	7.94%	5.88%	6.86%	0.00%	18.75%
		Negative	9.18%	9.97%	2.07%	3.36%	16.29%	16.55%
		Rel. Perf.	67.97%	20.34%	-184.50%	-104.34%	100.00%	-13.31%
	Big Lips	Positive	14.36%	9.95%	2.97%	3.27%	25.75%	16.74%
		Negative	5.73%	9.58%	2.13%	2.90%	9.34%	17.23%
		Rel. Perf.	-150.50%	-3.96%	-39.67%	-12.53%	-175.76%	2.83%
Nose	Big Nose	Positive	21.32%	9.30%	6.87%	2.94%	35.77%	18.03%
		Negative	7.10%	9.94%	3.09%	3.40%	11.11%	16.73%
		Rel. Perf.	-200.22%	6.44%	-122.09%	13.47%	-221.97%	-7.78%
	Pointy Nose	Positive	8.14%	10.21%	3.05%	3.58%	13.22%	16.68%
		Negative	19.30%	9.35%	6.87%	2.94%	31.74%	17.50%
		Rel. Perf.	57.86%	-9.14%	55.55%	-21.67%	58.35%	4.66%
Accessories	Heavy Makeup	Positive	2.11%	10.09%	3.21%	3.61%	1.01%	18.59%
		Negative	13.23%	9.43%	4.93%	3.30%	21.53%	16.47%
		Rel. Perf.	84.06%	-6.97%	34.79%	-9.66%	95.33%	-12.91%
	Wearing Hat	Positive	12.22%	8.59%	0.00%	5.00%	24.44%	22.22%
		Negative	8.58%	10.16%	3.46%	3.63%	13.71%	16.49%
		Rel. Perf.	-42.37%	15.47%	100.00%	-37.77%	-78.31%	-34.75%
	Wearing Lipstick	Positive	2.33%	9.44%	3.06%	3.31%	1.60%	18.72%
		Negative	14.73%	9.37%	6.87%	2.94%	22.58%	16.54%
		Rel. Perf.	84.18%	-0.82%	55.45%	-12.40%	92.92%	-13.16%
	Eyeglasses	Positive	9.06%	9.97%	2.87%	4.60%	15.25%	18.64%
		Negative	10.04%	10.13%	3.35%	3.61%	16.73%	16.60%
		Rel. Perf.	9.71%	1.60%	14.22%	-27.43%	8.81%	-12.33%
Other	Attractive	Positive	2.33%	9.56%	2.01%	3.61%	2.65%	22.12%
		Negative	15.65%	9.45%	6.86%	3.07%	24.44%	16.42%
		Rel. Perf.	85.10%	-1.17%	70.75%	-17.81%	89.14%	-34.76%

TABLE VI: Xception/Celeb-DF - Experiments with Xception on the Celeb-DF dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	12.54%	12.13%	13.09%	15.65%	11.98%	7.81%
		Negative	9.67%	11.76%	15.98%	15.57%	3.36%	7.52%
		Rel. Perf.	-29.63%	-3.11%	18.09%	-0.54%	-256.25%	-3.85%
	Young	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Asian	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	White	Positive	10.30%	11.76%	16.28%	15.57%	4.32%	7.71%
		Negative	14.75%	11.91%	13.09%	15.65%	16.41%	8.59%
		Rel. Perf.	30.15%	1.28%	-24.41%	0.48%	73.65%	10.16%
	Black	Positive	16.61%	12.12%	13.09%	15.65%	20.14%	8.45%
		Negative	10.18%	11.72%	15.98%	15.57%	4.38%	7.77%
		Rel. Perf.	-63.20%	-3.40%	18.09%	-0.54%	-359.44%	-8.77%
Skin	Shiny Skin	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Hair	Bald	Positive	11.94%	11.50%	21.79%	14.42%	2.08%	12.50%
		Negative	10.96%	11.65%	15.66%	15.60%	6.26%	7.78%
		Rel. Perf.	-8.93%	1.33%	-39.13%	7.52%	66.70%	-60.71%
	Wavy Hair	Positive	9.17%	9.24%	18.33%	15.83%	0.00%	13.79%
		Negative	13.46%	12.13%	13.36%	15.50%	13.55%	8.15%
		Rel. Perf.	31.88%	23.81%	-37.24%	-2.12%	100.00%	-69.32%
	Receding Hairline	Positive	20.28%	10.68%	40.56%	16.11%	0.00%	15.00%
		Negative	7.38%	12.59%	10.79%	15.30%	3.98%	8.78%
		Rel. Perf.	-174.63%	15.21%	-275.94%	-5.30%	100.00%	-70.93%
	Bangs	Positive	6.85%	12.04%	5.37%	15.15%	8.33%	15.00%
		Negative	11.52%	11.79%	16.04%	15.63%	7.01%	7.73%
		Rel. Perf.	40.53%	-2.13%	66.50%	3.06%	-18.93%	-93.98%
	Black Hair	Positive	5.66%	11.64%	7.42%	14.85%	3.89%	9.17%
		Negative	9.85%	12.31%	16.02%	15.59%	3.68%	7.89%
		Rel. Perf.	42.57%	5.45%	53.65%	4.77%	-5.68%	-16.25%
	Blond Hair	Positive	9.00%	12.61%	15.65%	15.37%	2.34%	8.78%
		Negative	11.89%	12.22%	15.27%	15.57%	8.50%	7.47%
		Rel. Perf.	24.30%	-3.12%	-2.51%	1.27%	72.46%	-17.54%
Beard	No Beard	Positive	9.91%	11.75%	15.98%	15.57%	3.84%	7.53%
		Negative	13.46%	11.01%	21.70%	15.32%	5.21%	8.57%
		Rel. Perf.	26.37%	-6.73%	26.39%	-1.61%	26.32%	12.06%
	Mustache	Positive	19.05%	11.96%	33.33%	15.36%	4.76%	9.24%
		Negative	10.18%	11.77%	15.98%	15.57%	4.39%	7.59%
		Rel. Perf.	-87.09%	-1.64%	-108.65%	1.29%	-8.56%	-21.81%
	Goatee	Positive	24.62%	12.11%	42.00%	14.00%	7.24%	8.79%
		Negative	11.10%	11.72%	15.83%	15.56%	6.36%	7.66%
		Rel. Perf.	-121.81%	-3.26%	-165.28%	10.02%	-13.68%	-14.65%
Face Geometry	Oval Face	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Square Face	Positive	12.34%	11.97%	14.23%	15.66%	10.45%	7.98%
		Negative	10.16%	11.75%	15.79%	15.69%	4.52%	7.45%
		Rel. Perf.	-21.47%	-1.90%	9.93%	0.19%	-131.19%	-7.13%
	Double Chin	Positive	6.97%	9.99%	1.67%	16.11%	12.28%	13.16%
		Negative	9.61%	11.72%	15.55%	15.55%	3.67%	7.32%
		Rel. Perf.	27.44%	14.77%	89.28%	-3.61%	-234.28%	-79.84%
	Chubby	Positive	18.12%	11.03%	14.01%	15.52%	22.22%	8.84%
		Negative	9.54%	11.80%	15.60%	15.66%	3.47%	7.43%
		Rel. Perf.	-90.02%	6.54%	10.17%	0.92%	-540.44%	-19.10%
	Obstructed Forehead	Positive	50.00%	12.50%	33.33%	0.00%	66.67%	0.00%
		Negative	11.12%	11.66%	15.99%	15.64%	6.26%	7.67%
		Rel. Perf.	-349.46%	-7.21%	-108.47%	100.00%	-965.05%	100.00%
	Fully Visible Forehead	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Mouth	Mouth Closed	Positive	10.11%	11.52%	17.47%	15.51%	2.76%	10.69%
		Negative	12.35%	10.99%	16.50%	15.37%	8.21%	10.03%
		Rel. Perf.	18.13%	-4.81%	-5.86%	-0.91%	66.39%	-6.57%
	Smiling	Positive	7.84%	9.53%	15.69%	17.65%	0.00%	18.75%
		Negative	10.30%	12.31%	14.54%	15.59%	6.06%	7.45%
		Rel. Perf.	23.86%	22.60%	-7.89%	-13.20%	100.00%	-151.59%
	Big Lips	Positive	12.48%	12.55%	13.62%	15.42%	11.34%	8.03%
		Negative	8.24%	12.23%	10.60%	15.63%	5.89%	8.43%
		Rel. Perf.	-51.42%	-2.62%	-28.59%	1.29%	-92.48%	4.74%
Nose	Big Nose	Positive	16.61%	12.12%	13.09%	15.65%	20.14%	8.45%
		Negative	9.72%	12.29%	17.26%	15.56%	2.17%	8.03%
		Rel. Perf.	-70.94%	1.39%	24.20%	-0.59%	-826.48%	-5.22%
	Pointy Nose	Positive	10.23%	11.76%	16.00%	15.56%	4.45%	7.68%
		Negative	14.75%	11.91%	13.09%	15.65%	16.41%	8.59%
		Rel. Perf.	30.67%	1.26%	-22.30%	0.58%	72.89%	10.53%
Accessories	Heavy Makeup	Positive	9.34%	11.93%	18.67%	15.86%	0.00%	10.05%
		Negative	11.55%	12.54%	12.20%	15.43%	10.90%	7.45%
		Rel. Perf.	19.16%	4.88%	-53.09%	-2.83%	100.00%	-34.87%
	Wearing Hat	Positive	0.83%	10.20%	0.56%	16.11%	1.11%	13.33%
		Negative	10.99%	11.64%	16.18%	15.67%	5.79%	7.61%
		Rel. Perf.	92.41%	12.36%	96.57%	-2.80%	80.81%	-75.30%
	Wearing Lipstick	Positive	9.08%	12.13%	17.47%	15.20%	0.68%	8.90%
		Negative	12.56%	12.08%	13.09%	15.65%	12.03%	7.62%
		Rel. Perf.	27.70%	-0.38%	-33.53%	2.90%	94.31%	-16.92%
	Eyeglasses	Positive	9.19%	9.63%	17.24%	15.52%	1.13%	10.73%
		Negative	11.21%	11.62%	15.76%	15.65%	6.66%	7.51%
		Rel. Perf.	18.04%	17.11%	-9.43%	0.83%	83.04%	-42.99%
Other	Attractive	Positive	11.67%	11.58%	20.68%	15.86%	2.65%	13.27%
		Negative	13.80%	12.14%	13.05%	15.55%	14.54%	8.32%
		Rel. Perf.	15.43%	4.63%	-58.46%	-2.02%	81.74%	-59.53%

TABLE VII: Capsule-Forensics-v2/Celeb-DF - Experiments with Capsule-Forensics-v2 on the Celeb-DF dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	13.73%	12.25%	3.65%	2.18%	23.81%	22.83%
		Negative	11.96%	12.43%	1.92%	2.07%	22.00%	22.63%
		Rel. Perf.	-14.79%	1.38%	-89.98%	-5.39%	-8.22%	-0.91%
	Young	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Asian	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	White	Positive	12.36%	12.39%	1.97%	2.07%	22.74%	22.85%
		Negative	14.11%	12.13%	3.65%	2.18%	24.57%	22.61%
		Rel. Perf.	12.42%	-2.19%	46.09%	4.99%	7.41%	-1.07%
	Black	Positive	15.91%	12.07%	3.65%	2.18%	28.17%	21.97%
		Negative	11.98%	12.34%	1.92%	2.07%	22.04%	22.99%
		Rel. Perf.	-32.79%	2.19%	-89.98%	-5.39%	-27.80%	4.43%
Skin	Shiny Skin	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Hair	Bald	Positive	4.57%	11.88%	7.05%	1.92%	2.08%	20.14%
		Negative	12.90%	12.35%	1.88%	2.08%	23.91%	23.00%
		Rel. Perf.	64.58%	3.77%	-275.07%	7.57%	91.29%	12.42%
	Wavy Hair	Positive	1.25%	10.52%	2.50%	1.67%	0.00%	24.14%
		Negative	12.42%	11.94%	3.70%	2.15%	21.14%	23.03%
		Rel. Perf.	89.94%	11.86%	32.46%	22.33%	100.00%	-4.79%
	Receding Hairline	Positive	1.39%	9.55%	2.78%	2.78%	0.00%	23.33%
		Negative	14.06%	12.19%	2.70%	2.11%	25.41%	22.24%
		Rel. Perf.	90.12%	21.71%	-2.78%	-31.82%	100.00%	-4.89%
	Bangs	Positive	32.83%	12.28%	0.65%	2.44%	65.00%	23.33%
		Negative	12.27%	12.26%	2.41%	2.07%	22.12%	22.90%
		Rel. Perf.	-167.63%	-0.19%	72.95%	-17.87%	-193.82%	-1.90%
	Black Hair	Positive	8.17%	10.78%	4.39%	2.27%	11.94%	24.17%
		Negative	10.73%	12.26%	1.78%	2.14%	19.68%	23.07%
		Rel. Perf.	23.88%	12.01%	-147.06%	-6.38%	39.32%	-4.74%
	Blond Hair	Positive	10.21%	12.18%	0.86%	2.07%	19.56%	22.83%
		Negative	13.74%	12.38%	2.94%	2.09%	24.54%	22.74%
		Rel. Perf.	25.71%	1.61%	70.77%	1.02%	20.32%	-0.42%
Beard	No Beard	Positive	11.93%	12.39%	1.92%	2.07%	21.94%	22.84%
		Negative	14.26%	10.69%	5.25%	2.13%	23.28%	22.72%
		Rel. Perf.	16.35%	-15.95%	63.36%	2.73%	5.75%	-0.53%
	Mustache	Positive	15.61%	11.98%	5.73%	2.08%	25.49%	24.37%
		Negative	11.90%	12.39%	1.92%	2.07%	21.88%	22.82%
		Rel. Perf.	-31.17%	3.31%	-197.93%	-0.67%	-16.51%	-6.79%
	Goatee	Positive	17.92%	12.27%	7.67%	2.00%	28.17%	22.74%
		Negative	12.62%	12.34%	1.90%	2.06%	23.35%	22.86%
		Rel. Perf.	-41.92%	0.51%	-302.90%	2.89%	-20.65%	0.54%
Face Geometry	Oval Face	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Square Face	Positive	14.75%	12.12%	5.30%	2.10%	24.21%	23.23%
		Negative	11.98%	12.36%	1.88%	2.11%	22.08%	22.63%
		Rel. Perf.	-23.20%	1.94%	-182.77%	0.12%	-9.64%	-2.68%
	Double Chin	Positive	17.54%	10.39%	0.00%	2.78%	35.09%	21.05%
		Negative	12.63%	12.32%	1.84%	2.05%	23.41%	22.45%
		Rel. Perf.	-38.96%	15.69%	100.00%	-35.27%	-49.87%	6.21%
	Chubby	Positive	17.88%	11.36%	5.61%	2.20%	30.16%	22.22%
		Negative	12.44%	12.29%	1.84%	2.07%	23.04%	22.83%
		Rel. Perf.	-43.77%	7.61%	-205.23%	-6.46%	-30.90%	2.67%
	Obstructed Forehead	Positive	33.33%	25.00%	33.33%	0.00%	33.33%	0.00%
		Negative	11.67%	12.36%	2.14%	2.09%	21.20%	22.89%
		Rel. Perf.	-185.66%	-102.31%	-1460.20%	100.00%	-57.22%	100.00%
	Fully Visible Forehead	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Mouth	Mouth Closed	Positive	6.13%	11.29%	0.53%	2.50%	11.72%	23.79%
		Negative	14.01%	10.94%	3.40%	2.27%	24.62%	24.62%
		Rel. Perf.	56.25%	-3.18%	84.27%	-10.12%	52.38%	3.36%
	Smiling	Positive	0.98%	8.24%	1.96%	0.98%	0.00%	31.25%
		Negative	13.05%	12.30%	1.63%	2.10%	24.46%	22.99%
		Rel. Perf.	92.49%	33.04%	-20.29%	53.32%	100.00%	-35.93%
	Big Lips	Positive	15.18%	12.62%	2.16%	2.11%	28.20%	23.24%
		Negative	12.56%	12.02%	0.82%	1.98%	24.30%	22.12%
		Rel. Perf.	-20.89%	-5.00%	-163.07%	-6.26%	-16.08%	-5.04%
Nose	Big Nose	Positive	15.91%	12.07%	3.65%	2.18%	28.17%	21.97%
		Negative	10.23%	12.32%	2.11%	2.06%	18.36%	23.31%
		Rel. Perf.	-55.51%	2.10%	-73.50%	-6.10%	-53.45%	5.74%
	Pointy Nose	Positive	12.34%	12.39%	1.93%	2.07%	22.76%	22.79%
		Negative	14.05%	12.13%	3.65%	2.18%	24.46%	22.61%
		Rel. Perf.	12.17%	-2.13%	47.27%	4.94%	6.93%	-0.79%
Accessories	Heavy Makeup	Positive	6.73%	12.11%	3.41%	2.41%	10.05%	18.59%
		Negative	13.38%	12.12%	2.48%	2.11%	24.29%	22.63%
		Rel. Perf.	49.70%	0.15%	-37.62%	-14.38%	58.62%	17.84%
	Wearing Hat	Positive	7.78%	9.86%	0.00%	2.78%	15.56%	22.22%
		Negative	12.38%	12.38%	2.18%	2.10%	22.57%	22.98%
		Rel. Perf.	37.16%	20.34%	100.00%	-32.41%	31.08%	3.28%
	Wearing Lipstick	Positive	4.04%	12.02%	2.14%	2.14%	5.94%	22.37%
		Negative	13.78%	12.27%	3.65%	2.18%	23.90%	22.95%
		Rel. Perf.	70.70%	2.05%	41.55%	2.10%	75.16%	2.52%
	Eyeglasses	Positive	8.23%	10.80%	4.02%	2.30%	12.43%	18.64%
		Negative	12.53%	12.46%	2.09%	2.10%	22.98%	22.98%
		Rel. Perf.	34.36%	13.37%	-92.82%	-9.62%	45.90%	18.86%
Other	Attractive	Positive	7.00%	11.63%	1.61%	2.41%	12.39%	21.24%
		Negative	13.17%	12.00%	3.48%	2.08%	22.86%	22.71%
		Rel. Perf.	46.88%	3.07%	53.89%	-15.84%	45.81%	6.49%

TABLE VIII: EfficientNetB0/DFD - Experiments with EfficientNetB0 on the DFD dataset.

Category	Attribute	Class	Balanced Error		Real Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	0.71%	2.85%	0.28%	1.67%	6.77%	8.58%
		Negative	4.13%	2.78%	2.88%	1.67%	10.80%	9.04%
		Rel. Perf.	82.79%	-2.34%	90.43%	-0.29%	37.30%	5.11%
	Young	Positive	3.07%	2.97%	1.06%	1.79%	12.26%	8.91%
		Negative	3.24%	3.67%	3.98%	1.86%	0.00%	4.65%
		Rel. Perf.	5.16%	19.18%	73.44%	3.43%	All wrong	-91.46%
	Asian	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	White	Positive	4.82%	2.87%	3.36%	1.71%	15.25%	8.66%
		Negative	0.97%	2.82%	0.36%	1.64%	5.29%	9.45%
		Rel. Perf.	-398.34%	-1.58%	-826.33%	-4.09%	-188.37%	8.39%
	Black	Positive	1.17%	2.87%	0.36%	1.64%	5.29%	9.45%
		Negative	3.78%	2.76%	2.57%	1.66%	14.68%	9.20%
		Rel. Perf.	69.13%	-4.02%	86.19%	1.00%	63.98%	-2.66%
Skin	Shiny Skin	Positive	1.07%	2.86%	0.54%	1.66%	4.14%	8.67%
		Negative	2.40%	3.60%	0.92%	1.62%	6.67%	7.33%
		Rel. Perf.	55.48%	20.50%	41.81%	-2.52%	37.90%	-18.19%
Hair	Bald	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wavy Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Receding Hairline	Positive	1.62%	3.05%	0.00%	1.67%	6.30%	9.26%
		Negative	19.21%	3.39%	23.13%	2.72%	0.00%	3.33%
		Rel. Perf.	91.56%	9.92%	100.00%	38.59%	All wrong	-177.78%
	Bangs	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Black Hair	Positive	0.77%	2.86%	0.38%	1.65%	3.22%	8.92%
		Negative	3.81%	2.88%	2.07%	1.80%	12.75%	8.68%
		Rel. Perf.	79.80%	0.63%	81.83%	8.48%	74.76%	-2.70%
	Blond Hair	Positive	0.00%	4.04%	0.00%	2.51%	0.00%	3.45%
		Negative	2.28%	2.64%	1.38%	1.62%	10.01%	8.96%
		Rel. Perf.	100.00%	-52.85%	100.00%	-54.62%	100.00%	61.50%
Beard	No Beard	Positive	3.67%	2.67%	2.57%	1.63%	10.24%	9.04%
		Negative	1.02%	2.80%	0.85%	1.82%	2.43%	8.74%
		Rel. Perf.	-258.15%	4.83%	-202.14%	10.68%	-321.89%	-3.46%
	Mustache	Positive	1.61%	2.96%	1.43%	1.74%	2.43%	8.74%
		Negative	3.67%	2.67%	2.50%	1.63%	11.01%	9.08%
		Rel. Perf.	56.14%	-10.73%	42.58%	-6.88%	77.95%	3.72%
	Goatee	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Face Geometry	Oval Face	Positive	10.46%	3.15%	8.80%	1.82%	14.87%	9.29%
		Negative	2.40%	2.88%	1.04%	1.64%	8.00%	8.62%
		Rel. Perf.	-335.79%	-9.30%	-743.44%	-10.75%	-85.87%	-7.87%
	Square Face	Positive	0.62%	2.84%	0.42%	1.76%	3.86%	8.58%
		Negative	3.34%	2.52%	2.15%	1.60%	10.68%	9.66%
		Rel. Perf.	81.53%	-12.76%	80.49%	-9.51%	63.85%	11.12%
	Double Chin	Positive	2.87%	3.55%	2.69%	1.57%	3.42%	6.85%
		Negative	4.68%	2.87%	3.18%	1.68%	15.25%	8.66%
		Rel. Perf.	38.58%	-23.76%	15.40%	6.67%	77.54%	20.88%
	Chubby	Positive	0.93%	3.07%	0.65%	1.91%	3.38%	8.65%
		Negative	6.29%	2.87%	4.50%	1.75%	15.46%	8.60%
		Rel. Perf.	85.15%	-7.07%	85.54%	-8.85%	78.11%	-0.52%
	Obstructed Forehead	Positive	5.88%	0.00%	6.25%	0.00%	0.00%	0.00%
		Negative	2.47%	2.57%	1.42%	1.61%	10.69%	9.68%
		Rel. Perf.	-138.55%	100.00%	-340.83%	100.00%	100.00%	100.00%
	Fully Visible Forehead	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Mouth	Mouth Closed	Positive	3.90%	3.57%	1.67%	1.85%	28.57%	2.04%
		Negative	0.80%	3.03%	0.36%	1.85%	3.97%	8.20%
		Rel. Perf.	-390.15%	-17.77%	-360.00%	-0.22%	-620.00%	75.12%
	Smiling	Positive	2.13%	3.72%	2.45%	3.07%	0.00%	0.00%
		Negative	3.53%	2.90%	1.38%	1.67%	17.46%	7.94%
		Rel. Perf.	39.81%	-28.57%	-78.18%	-83.42%	100.00%	100.00%
	Big Lips	Positive	1.15%	2.65%	0.44%	1.64%	6.56%	8.90%
		Negative	4.77%	2.88%	3.86%	1.81%	9.56%	8.53%
		Rel. Perf.	75.79%	7.90%	88.71%	9.08%	31.38%	-4.41%
Nose	Big Nose	Positive	0.54%	3.02%	0.45%	1.73%	1.40%	8.15%
		Negative	9.60%	2.94%	5.36%	1.83%	16.33%	8.50%
		Rel. Perf.	94.38%	-2.70%	91.65%	5.63%	91.40%	4.18%
	Pointy Nose	Positive	3.91%	2.78%	2.66%	1.64%	14.68%	9.20%
		Negative	1.12%	2.84%	0.34%	1.63%	5.29%	9.45%
		Rel. Perf.	-248.80%	2.09%	-687.68%	-0.86%	-177.61%	2.59%
Accessories	Heavy Makeup	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wearing Hat	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wearing Lipstick	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Eyeglasses	Positive	1.17%	2.90%	0.95%	1.62%	3.42%	6.85%
		Negative	2.83%	2.58%	1.78%	1.61%	10.47%	9.80%
		Rel. Perf.	58.57%	-12.11%	46.66%	-0.84%	67.29%	30.10%
Other	Attractive	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-

TABLE IX: Xception/DFD - Experiments with Xception on the DFD dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	6.41%	5.75%	0.41%	3.23%	12.42%	6.32%
		Negative	6.32%	5.49%	5.60%	3.21%	7.04%	7.28%
		Rel. Perf.	-1.43%	-4.73%	92.76%	-0.50%	-76.35%	13.18%
	Young	Positive	4.50%	5.92%	3.46%	3.43%	5.55%	7.01%
		Negative	14.25%	7.52%	2.92%	3.18%	25.58%	4.65%
		Rel. Perf.	68.40%	21.25%	-18.52%	-7.64%	78.31%	-50.66%
	Asian	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	White	Positive	6.88%	5.85%	6.79%	3.18%	6.98%	6.85%
		Negative	5.14%	5.67%	0.38%	3.17%	9.90%	6.75%
		Rel. Perf.	-33.98%	-3.16%	-1694.66%	-0.27%	29.52%	-1.46%
	Black	Positive	5.12%	5.88%	0.33%	3.13%	9.90%	6.75%
		Negative	6.02%	5.41%	5.20%	3.17%	6.84%	7.09%
		Rel. Perf.	15.03%	-8.67%	93.60%	1.15%	-44.70%	4.80%
Skin	Shiny Skin	Positive	2.98%	5.80%	0.40%	3.16%	5.56%	6.86%
		Negative	0.23%	7.28%	0.46%	3.23%	0.00%	3.33%
		Rel. Perf.	-1191.61%	20.39%	12.72%	2.33%	All correct	-105.69%
Hair	Bald	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wavy Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Receding Hairline	Positive	5.74%	7.42%	0.00%	3.73%	11.48%	4.07%
		Negative	12.24%	4.24%	24.49%	3.40%	0.00%	3.33%
		Rel. Perf.	53.12%	-75.00%	100.00%	-9.59%	All correct	-22.22%
	Bangs	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Black Hair	Positive	5.17%	5.79%	0.40%	3.18%	9.94%	7.02%
		Negative	5.02%	5.97%	2.48%	3.46%	7.56%	6.86%
		Rel. Perf.	-2.96%	3.04%	83.88%	8.28%	-31.45%	-2.26%
	Blond Hair	Positive	0.21%	8.58%	0.42%	2.93%	0.00%	1.72%
		Negative	7.35%	5.26%	3.85%	3.08%	10.85%	6.53%
		Rel. Perf.	97.15%	-63.10%	89.13%	4.75%	100.00%	73.61%
Beard	No Beard	Positive	6.08%	5.24%	5.05%	3.12%	7.12%	7.28%
		Negative	9.41%	6.98%	0.85%	3.40%	17.96%	2.91%
		Rel. Perf.	35.30%	24.89%	-494.78%	8.25%	60.36%	-149.95%
	Mustache	Positive	9.60%	7.64%	1.23%	3.79%	17.96%	2.91%
		Negative	6.15%	5.25%	5.00%	3.13%	7.31%	7.14%
		Rel. Perf.	-55.93%	-45.58%	75.39%	-21.28%	-145.68%	59.22%
	Goatee	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Face Geometry	Oval Face	Positive	13.51%	7.42%	15.50%	3.77%	11.52%	4.09%
		Negative	9.79%	6.61%	0.82%	3.43%	18.77%	4.92%
		Rel. Perf.	-37.97%	-12.23%	-1791.34%	-10.01%	38.60%	16.94%
	Square Face	Positive	8.23%	5.89%	0.58%	3.38%	15.88%	3.00%
		Negative	5.72%	5.42%	4.19%	3.08%	7.26%	7.26%
		Rel. Perf.	-43.75%	-8.68%	86.23%	-9.61%	-118.72%	58.62%
	Double Chin	Positive	12.08%	7.24%	2.24%	3.36%	21.92%	3.42%
		Negative	6.88%	5.90%	6.65%	3.16%	7.11%	6.85%
		Rel. Perf.	-75.59%	-22.73%	66.30%	-6.36%	-208.44%	49.99%
	Chubby	Positive	9.45%	6.47%	0.48%	3.43%	18.42%	4.14%
		Negative	8.45%	6.01%	9.51%	3.38%	7.39%	6.85%
		Rel. Perf.	-11.84%	-7.71%	94.98%	-1.51%	-149.19%	39.67%
	Obstructed Forehead	Positive	3.13%	0.00%	6.25%	0.00%	0.00%	0.00%
		Negative	6.13%	5.35%	3.16%	3.03%	9.10%	7.36%
		Rel. Perf.	49.00%	100.00%	-97.92%	100.00%	100.00%	100.00%
	Fully Visible Forehead	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Mouth	Mouth Closed	Positive	3.80%	7.25%	1.48%	3.15%	6.12%	2.04%
		Negative	3.72%	6.31%	0.29%	3.51%	7.14%	5.29%
		Rel. Perf.	-2.30%	-14.94%	-411.11%	10.42%	14.29%	61.43%
	Smiling	Positive	3.23%	3.97%	2.45%	3.07%	4.00%	4.00%
		Negative	5.69%	5.81%	2.34%	3.27%	9.05%	6.98%
		Rel. Perf.	43.31%	31.59%	-5.03%	6.22%	55.79%	42.73%
	Big Lips	Positive	5.76%	5.25%	0.37%	3.09%	11.15%	6.84%
		Negative	6.33%	6.57%	8.02%	3.42%	4.65%	5.17%
		Rel. Perf.	9.00%	20.08%	95.34%	9.71%	-139.78%	-32.39%
Nose	Big Nose	Positive	5.67%	6.00%	0.39%	3.40%	10.96%	5.06%
		Negative	11.19%	7.82%	11.42%	3.67%	10.96%	6.26%
		Rel. Perf.	49.33%	23.23%	96.61%	7.26%	0.06%	19.28%
	Pointy Nose	Positive	6.12%	5.54%	5.41%	3.18%	6.84%	7.09%
		Negative	5.13%	5.87%	0.36%	3.10%	9.90%	6.75%
		Rel. Perf.	-19.39%	5.57%	-1406.73%	-2.42%	30.89%	-5.04%
Accessories	Heavy Makeup	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wearing Hat	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wearing Lipstick	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Eyeglasses	Positive	12.05%	6.79%	1.49%	3.52%	22.60%	3.42%
		Negative	5.29%	5.35%	3.46%	3.03%	7.11%	7.45%
		Rel. Perf.	-127.79%	-27.10%	56.98%	-16.21%	-217.72%	54.03%
Other	Attractive	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-

TABLE X: Capsule-Forensics-v2/DFD - Experiments with Capsule-Forensics-v2 on the DFD dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	7.89%	11.90%	0.19%	3.49%	15.58%	20.99%
		Negative	14.85%	12.10%	6.17%	3.44%	23.52%	20.96%
		Rel. Perf.	46.88%	1.59%	96.84%	-1.47%	33.78%	-0.16%
	Young	Positive	16.37%	11.20%	3.11%	3.68%	29.64%	20.58%
		Negative	15.70%	10.97%	0.00%	3.18%	31.40%	17.44%
		Rel. Perf.	-4.29%	-2.17%	All correct	-15.69%	5.61%	-18.01%
	Asian	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	White	Positive	18.95%	12.01%	7.41%	3.49%	30.49%	20.80%
		Negative	7.13%	12.09%	0.21%	3.42%	14.06%	20.81%
		Rel. Perf.	-165.68%	0.65%	-3514.42%	-1.90%	-116.85%	0.04%
	Black	Positive	7.14%	11.71%	0.22%	3.49%	14.06%	20.81%
		Negative	17.65%	12.06%	5.70%	3.38%	29.60%	21.14%
		Rel. Perf.	59.54%	2.89%	96.10%	-3.30%	52.50%	1.58%
Skin	Shiny Skin	Positive	10.65%	11.81%	0.09%	3.52%	21.22%	20.83%
		Negative	8.67%	11.44%	0.00%	3.23%	17.33%	14.00%
		Rel. Perf.	-22.92%	-3.23%	All correct	-8.75%	-22.40%	-48.77%
Hair	Bald	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wavy Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Receding Hairline	Positive	2.17%	10.89%	0.26%	3.86%	4.07%	18.15%
		Negative	11.22%	13.53%	22.45%	3.40%	0.00%	16.67%
		Rel. Perf.	80.71%	19.49%	98.85%	-13.37%	All correct	-8.89%
	Bangs	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Black Hair	Positive	4.60%	11.76%	0.28%	3.53%	8.92%	20.61%
		Negative	17.78%	11.81%	2.37%	3.79%	33.19%	20.45%
		Rel. Perf.	74.13%	0.46%	88.10%	6.89%	73.13%	-0.81%
	Blond Hair	Positive	0.86%	13.78%	0.00%	3.35%	1.72%	17.24%
		Negative	9.29%	11.90%	4.45%	3.31%	14.12%	20.44%
		Rel. Perf.	90.72%	-15.78%	100.00%	-1.17%	87.79%	15.66%
Beard	No Beard	Positive	14.99%	12.10%	5.51%	3.33%	24.48%	20.96%
		Negative	9.98%	10.83%	0.06%	3.88%	19.90%	15.05%
		Rel. Perf.	-50.21%	-11.75%	-8979.91%	14.18%	-23.00%	-39.28%
	Mustache	Positive	10.00%	10.79%	0.10%	4.30%	19.90%	15.05%
		Negative	15.48%	12.08%	5.50%	3.34%	25.46%	20.76%
		Rel. Perf.	35.40%	10.67%	98.14%	-28.86%	21.83%	27.50%
	Goatee	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Face Geometry	Oval Face	Positive	36.68%	10.77%	19.83%	4.05%	53.53%	18.22%
		Negative	5.96%	10.79%	0.22%	3.87%	11.69%	18.46%
		Rel. Perf.	-515.68%	0.14%	-8771.69%	-4.53%	-357.84%	1.33%
	Square Face	Positive	9.12%	11.54%	0.21%	3.70%	18.03%	17.17%
		Negative	13.30%	12.11%	4.62%	3.30%	21.99%	20.89%
		Rel. Perf.	31.47%	4.65%	95.46%	-12.11%	18.01%	17.82%
	Double Chin	Positive	13.70%	11.49%	0.00%	3.36%	27.40%	14.38%
		Negative	19.06%	11.91%	7.49%	3.47%	30.62%	20.80%
		Rel. Perf.	28.12%	3.56%	100.00%	3.16%	10.53%	30.85%
	Chubby	Positive	8.48%	11.64%	0.04%	3.78%	16.92%	18.42%
		Negative	21.07%	12.34%	10.55%	3.69%	31.59%	20.70%
		Rel. Perf.	59.75%	5.69%	99.59%	-2.27%	46.44%	11.00%
	Obstructed Forehead	Positive	3.13%	12.50%	6.25%	0.00%	0.00%	0.00%
		Negative	13.32%	12.03%	3.54%	3.17%	23.10%	20.72%
		Rel. Perf.	76.55%	-3.87%	-76.33%	100.00%	100.00%	100.00%
	Fully Visible Forehead	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Mouth	Mouth Closed	Positive	10.01%	11.50%	5.74%	3.15%	14.29%	16.33%
		Negative	10.90%	11.05%	0.36%	3.84%	21.43%	19.05%
		Rel. Perf.	8.10%	-4.08%	-1484.44%	18.03%	33.33%	14.29%
	Smiling	Positive	1.23%	12.59%	2.45%	3.07%	0.00%	16.00%
		Negative	19.19%	12.15%	3.62%	3.62%	34.76%	20.48%
		Rel. Perf.	93.61%	-3.61%	32.12%	15.15%	100.00%	21.86%
	Big Lips	Positive	9.83%	11.91%	0.16%	3.30%	19.49%	20.43%
		Negative	10.97%	11.93%	10.31%	3.86%	11.63%	19.12%
		Rel. Perf.	10.41%	0.22%	98.43%	14.51%	-67.65%	-6.85%
Nose	Big Nose	Positive	12.15%	11.24%	0.15%	3.67%	24.16%	18.26%
		Negative	30.91%	11.04%	14.39%	3.81%	47.43%	21.25%
		Rel. Perf.	60.68%	-1.85%	98.96%	3.64%	49.06%	14.09%
	Pointy Nose	Positive	17.75%	12.20%	5.90%	3.41%	29.60%	21.14%
		Negative	7.14%	11.70%	0.21%	3.48%	14.06%	20.81%
		Rel. Perf.	-148.72%	-4.29%	-2693.07%	2.15%	-110.53%	-1.61%
Accessories	Heavy Makeup	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wearing Hat	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wearing Lipstick	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Eyeglasses	Positive	13.77%	10.81%	0.14%	3.93%	27.40%	14.38%
		Negative	12.71%	12.06%	3.87%	3.17%	21.54%	20.74%
		Rel. Perf.	-8.35%	10.41%	96.50%	-23.97%	-27.17%	30.64%
Other	Attractive	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-

TABLE XI: EfficientNetB0/FF++ - Experiments with EfficientNetB0 on the FF++ dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	1.63%	1.45%	1.27%	0.78%	3.37%	4.26%
		Negative	1.42%	1.44%	0.65%	0.84%	5.28%	4.56%
		Rel. Perf.	-14.66%	-0.78%	-96.99%	7.17%	36.12%	6.55%
	Young	Positive	1.12%	1.47%	0.43%	0.74%	4.55%	4.24%
		Negative	0.41%	1.55%	0.00%	0.61%	2.44%	4.89%
		Rel. Perf.	-172.49%	4.66%	All right	-21.58%	-86.06%	13.31%
	Asian	Positive	1.25%	1.66%	0.00%	0.85%	6.83%	4.40%
		Negative	1.62%	1.48%	1.15%	0.86%	4.04%	4.42%
		Rel. Perf.	22.93%	-12.69%	100.00%	1.26%	-68.96%	0.37%
	White	Positive	1.53%	1.47%	0.95%	0.84%	4.49%	4.56%
		Negative	1.92%	1.59%	1.49%	0.73%	3.92%	4.50%
		Rel. Perf.	20.46%	7.28%	35.75%	-15.96%	-14.63%	-1.29%
	Black	Positive	0.13%	1.67%	0.00%	1.52%	0.83%	2.50%
		Negative	1.44%	1.49%	0.79%	0.86%	4.58%	4.40%
		Rel. Perf.	91.07%	-11.61%	100.00%	-76.91%	81.82%	43.23%
Skin	Shiny Skin	Positive	0.89%	1.52%	0.44%	0.59%	2.97%	4.76%
		Negative	1.14%	1.59%	0.00%	0.94%	8.48%	3.33%
		Rel. Perf.	21.59%	4.12%	All right	36.71%	65.04%	-42.86%
Hair	Bald	Positive	0.00%	1.68%	0.00%	0.86%	0.00%	3.33%
		Negative	1.78%	1.48%	1.19%	0.88%	4.70%	4.42%
		Rel. Perf.	100.00%	-13.81%	100.00%	1.33%	100.00%	24.51%
	Wavy Hair	Positive	0.00%	0.98%	0.00%	3.39%	0.00%	0.00%
		Negative	1.87%	1.44%	1.64%	0.80%	3.03%	4.26%
		Rel. Perf.	100.00%	32.03%	100.00%	-325.31%	100.00%	100.00%
	Receding Hairline	Positive	1.38%	1.67%	0.00%	0.93%	6.33%	3.67%
		Negative	1.28%	1.57%	0.48%	0.68%	4.73%	4.55%
		Rel. Perf.	-7.94%	-6.10%	100.00%	-35.61%	-33.76%	19.33%
	Bangs	Positive	1.58%	1.48%	0.00%	1.20%	8.94%	2.79%
		Negative	1.60%	1.48%	1.02%	0.87%	4.40%	4.38%
		Rel. Perf.	1.36%	0.05%	100.00%	-38.07%	-103.17%	36.22%
	Black Hair	Positive	2.07%	1.50%	1.90%	0.74%	2.89%	4.78%
		Negative	2.13%	1.54%	1.78%	0.65%	3.71%	4.62%
		Rel. Perf.	2.79%	2.52%	-6.72%	-14.03%	22.16%	-3.37%
	Blond Hair	Positive	4.70%	1.54%	2.96%	1.18%	9.29%	3.41%
		Negative	1.58%	1.47%	1.01%	0.89%	4.29%	4.51%
		Rel. Perf.	-198.43%	-4.40%	-193.12%	-32.84%	-116.75%	24.50%
Beard	No Beard	Positive	1.45%	1.47%	0.91%	0.88%	4.28%	4.44%
		Negative	1.05%	1.49%	1.18%	1.08%	0.48%	3.33%
		Rel. Perf.	-38.10%	1.52%	23.43%	18.25%	-798.42%	-33.15%
	Mustache	Positive	0.18%	1.40%	0.00%	1.78%	0.83%	2.50%
		Negative	1.66%	1.48%	1.00%	0.87%	4.89%	4.41%
		Rel. Perf.	89.41%	5.36%	100.00%	-105.35%	82.97%	43.36%
	Goatee	Positive	0.67%	0.00%	0.00%	4.17%	3.33%	0.00%
		Negative	1.53%	1.49%	0.88%	0.86%	4.78%	4.42%
		Rel. Perf.	56.53%	100.00%	100.00%	-383.86%	30.20%	100.00%
Face Geometry	Oval Face	Positive	1.75%	1.59%	0.79%	0.76%	6.36%	4.55%
		Negative	3.94%	1.49%	3.50%	0.74%	5.99%	4.81%
		Rel. Perf.	55.64%	-6.57%	77.57%	-2.73%	-6.30%	5.47%
	Square Face	Positive	1.77%	1.44%	1.47%	0.79%	3.22%	4.36%
		Negative	1.30%	1.45%	0.49%	0.84%	5.42%	4.57%
		Rel. Perf.	-35.93%	0.42%	-198.99%	6.83%	40.53%	4.61%
	Double Chin	Positive	2.13%	1.47%	0.00%	1.67%	10.67%	2.67%
		Negative	1.42%	1.44%	0.69%	0.78%	4.87%	4.37%
		Rel. Perf.	-50.39%	-1.61%	100.00%	-112.50%	-119.02%	39.01%
	Chubby	Positive	2.17%	1.69%	1.60%	0.90%	4.85%	3.33%
		Negative	1.39%	1.49%	0.74%	0.74%	4.28%	4.23%
		Rel. Perf.	-56.40%	-13.51%	-115.57%	-20.72%	-13.35%	21.11%
	Obstructed Forehead	Positive	2.21%	1.53%	2.56%	1.71%	0.84%	2.52%
		Negative	1.49%	1.47%	0.97%	0.87%	3.99%	4.41%
		Rel. Perf.	-48.22%	-4.12%	-163.92%	-95.97%	78.94%	42.89%
	Fully Visible Forehead	Positive	2.20%	1.74%	2.01%	0.97%	3.15%	3.50%
		Negative	1.13%	1.47%	0.47%	0.84%	4.42%	4.42%
		Rel. Perf.	-94.55%	-17.65%	-326.88%	-15.57%	28.87%	20.97%
Mouth	Mouth Closed	Positive	1.15%	1.50%	0.29%	1.47%	4.28%	2.67%
		Negative	0.64%	1.70%	0.00%	0.84%	4.37%	4.57%
		Rel. Perf.	-81.82%	11.86%	All wrong	-74.43%	2.19%	41.53%
	Smiling	Positive	1.14%	0.00%	0.00%	6.33%	11.11%	0.00%
		Negative	1.41%	1.44%	0.95%	0.80%	3.87%	4.24%
		Rel. Perf.	19.42%	100.00%	100.00%	-692.09%	-187.38%	100.00%
	Big Lips	Positive	1.68%	1.44%	1.40%	0.78%	3.03%	4.22%
		Negative	1.39%	1.57%	0.65%	0.76%	5.42%	4.59%
		Rel. Perf.	-20.44%	8.24%	-116.20%	-2.93%	44.16%	7.93%
Nose	Big Nose	Positive	3.00%	1.58%	2.58%	0.96%	5.01%	4.53%
		Negative	1.92%	1.52%	0.61%	0.74%	7.70%	5.01%
		Rel. Perf.	-56.48%	-3.93%	-322.90%	-30.54%	34.91%	9.48%
	Pointy Nose	Positive	1.44%	1.48%	0.98%	0.89%	3.81%	4.48%
		Negative	2.18%	1.60%	1.54%	0.78%	5.42%	4.72%
		Rel. Perf.	33.88%	7.97%	36.22%	-13.88%	29.62%	5.14%
Accessories	Heavy Makeup	Positive	1.28%	1.72%	0.00%	0.86%	8.33%	4.55%
		Negative	1.84%	1.44%	1.37%	0.80%	4.15%	4.39%
		Rel. Perf.	30.30%	-19.23%	100.00%	-7.66%	-100.99%	-3.49%
	Wearing Hat	Positive	0.82%	1.64%	0.69%	0.90%	1.62%	3.56%
		Negative	1.67%	1.48%	1.07%	0.87%	4.40%	4.42%
		Rel. Perf.	50.93%	-10.60%	35.96%	-2.89%	63.19%	19.37%
	Wearing Lipstick	Positive	0.73%	1.54%	0.07%	0.61%	4.02%	4.93%
		Negative	1.87%	1.44%	1.42%	0.79%	4.07%	4.36%
		Rel. Perf.	61.12%	-6.38%	95.38%	22.54%	1.09%	-12.97%
	Eyeglasses	Positive	0.41%	1.69%	0.00%	0.93%	2.16%	4.52%
		Negative	1.62%	1.49%	1.01%	0.85%	4.64%	4.42%
		Rel. Perf.	74.47%	-13.13%	100.00%	-8.55%	53.46%	-2.23%
Other	Attractive	Positive	0.22%	1.65%	0.00%	0.94%	0.92%	3.36%
		Negative	2.11%	1.49%	1.59%	0.76%	4.79%	4.09%
		Rel. Perf.	89.79%	-10.81%	100.00%	-23.61%	80.83%	17.71%

TABLE XII: Xception/FF++ - Experiments with Xception on the FF++ dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	1.54%	1.24%	1.52%	1.06%	1.63%	1.74%
		Negative	1.12%	1.23%	0.95%	1.13%	2.00%	1.91%
		Rel. Perf.	-37.21%	-1.45%	-60.77%	6.58%	18.44%	8.59%
	Young	Positive	0.67%	1.21%	0.63%	1.03%	0.88%	2.01%
		Negative	0.11%	1.33%	0.08%	0.91%	0.24%	2.12%
		Rel. Perf.	-509.93%	9.19%	-660.29%	-14.04%	-259.44%	4.86%
	Asian	Positive	0.42%	1.41%	0.00%	1.15%	2.28%	1.82%
		Negative	1.58%	1.31%	1.49%	1.15%	2.03%	1.81%
		Rel. Perf.	73.67%	-7.76%	100.00%	-0.09%	-12.03%	-0.55%
	White	Positive	1.33%	1.25%	1.22%	1.14%	1.89%	1.89%
		Negative	1.84%	1.32%	1.86%	1.05%	1.74%	2.06%
		Rel. Perf.	27.78%	5.68%	34.44%	-8.74%	-8.67%	8.31%
	Black	Positive	3.33%	1.41%	2.73%	1.67%	6.67%	0.00%
		Negative	1.11%	1.30%	0.99%	1.17%	1.75%	1.79%
		Rel. Perf.	-199.05%	-8.40%	-176.85%	-43.03%	-280.82%	100.00%
Skin	Shiny Skin	Positive	1.00%	1.31%	0.92%	0.90%	1.41%	2.03%
		Negative	0.85%	1.22%	0.00%	1.22%	6.36%	1.82%
		Rel. Perf.	-17.62%	-7.49%	All wrong	26.28%	77.92%	-11.63%
Hair	Bald	Positive	0.00%	1.47%	0.00%	1.04%	0.00%	2.22%
		Negative	1.49%	1.32%	1.43%	1.17%	1.76%	1.80%
		Rel. Perf.	100.00%	-11.17%	100.00%	11.42%	100.00%	-23.36%
	Wavy Hair	Positive	0.00%	1.96%	0.00%	3.39%	0.00%	0.00%
		Negative	1.91%	1.23%	2.04%	1.08%	1.27%	1.80%
		Rel. Perf.	100.00%	-59.42%	100.00%	-212.91%	100.00%	100.00%
	Receding Hairline	Positive	1.31%	1.45%	1.67%	1.11%	0.00%	2.00%
		Negative	1.02%	1.32%	0.52%	1.02%	3.16%	2.08%
		Rel. Perf.	-28.42%	-9.56%	-218.67%	-9.26%	100.00%	4.00%
	Bangs	Positive	1.48%	1.48%	0.00%	1.32%	8.38%	1.12%
		Negative	1.36%	1.32%	1.35%	1.15%	1.43%	1.80%
		Rel. Perf.	-8.74%	-12.12%	100.00%	-14.31%	-487.29%	38.04%
	Black Hair	Positive	2.48%	1.36%	2.46%	1.02%	2.56%	2.11%
		Negative	1.51%	1.28%	1.81%	0.96%	0.14%	2.03%
		Rel. Perf.	-63.93%	-6.30%	-35.59%	-6.21%	-1724.67%	-3.95%
	Blond Hair	Positive	2.22%	1.37%	3.07%	1.30%	0.00%	1.86%
		Negative	1.44%	1.33%	1.25%	1.19%	2.30%	1.85%
		Rel. Perf.	-54.91%	-3.20%	-145.03%	-9.45%	100.00%	-0.45%
Beard	No Beard	Positive	1.17%	1.31%	1.18%	1.17%	1.14%	1.83%
		Negative	1.05%	1.40%	1.18%	1.29%	0.48%	1.43%
		Rel. Perf.	-11.61%	6.76%	0.16%	9.17%	-140.22%	-28.12%
	Mustache	Positive	0.18%	1.40%	0.00%	1.78%	0.83%	0.00%
		Negative	1.40%	1.31%	1.31%	1.16%	1.80%	1.86%
		Rel. Perf.	87.43%	-7.19%	100.00%	-52.82%	53.80%	100.00%
	Goatee	Positive	0.67%	1.33%	0.00%	4.17%	3.33%	0.00%
		Negative	1.29%	1.30%	1.18%	1.17%	1.83%	1.83%
		Rel. Perf.	48.17%	-2.66%	100.00%	-257.43%	-82.35%	100.00%
Face Geometry	Oval Face	Positive	0.94%	1.44%	0.81%	1.00%	1.56%	1.95%
		Negative	3.60%	1.30%	3.57%	1.03%	3.73%	2.16%
		Rel. Perf.	73.81%	-10.15%	77.21%	3.04%	58.21%	9.77%
	Square Face	Positive	1.68%	1.23%	1.70%	1.07%	1.61%	1.80%
		Negative	1.01%	1.22%	0.81%	1.14%	2.02%	1.89%
		Rel. Perf.	-67.07%	-0.68%	-110.08%	5.89%	20.09%	5.08%
	Double Chin	Positive	0.00%	1.33%	0.00%	1.67%	0.00%	0.00%
		Negative	1.11%	1.24%	0.95%	1.06%	1.88%	1.74%
		Rel. Perf.	100.00%	-7.69%	100.00%	-56.74%	100.00%	100.00%
	Chubby	Positive	1.48%	1.48%	1.79%	1.15%	0.00%	1.82%
		Negative	0.97%	1.22%	0.80%	1.06%	1.72%	2.03%
		Rel. Perf.	-52.88%	-21.83%	-123.94%	-9.16%	100.00%	10.63%
	Obstructed Forehead	Positive	2.21%	1.36%	2.56%	1.71%	0.84%	0.00%
		Negative	1.23%	1.30%	1.23%	1.18%	1.25%	1.90%
		Rel. Perf.	-79.31%	-4.50%	-108.01%	-44.17%	32.61%	100.00%
	Fully Visible Forehead	Positive	1.91%	1.56%	2.01%	1.11%	1.40%	2.10%
		Negative	0.77%	1.28%	0.62%	1.15%	1.50%	1.87%
		Rel. Perf.	-147.72%	-21.91%	-222.10%	3.70%	6.95%	-12.08%
Mouth	Mouth Closed	Positive	1.39%	1.39%	0.88%	1.62%	3.21%	1.07%
		Negative	0.78%	1.41%	0.68%	1.15%	1.39%	2.19%
		Rel. Perf.	-77.78%	2.04%	-30.82%	-41.09%	-130.56%	51.09%
	Smiling	Positive	0.00%	1.14%	0.00%	6.33%	0.00%	0.00%
		Negative	1.18%	1.23%	1.03%	1.10%	1.95%	1.76%
		Rel. Perf.	100.00%	7.74%	100.00%	-473.76%	100.00%	100.00%
	Big Lips	Positive	1.74%	1.22%	1.69%	1.07%	2.02%	1.76%
		Negative	1.12%	1.30%	1.11%	1.06%	1.18%	2.02%
		Rel. Perf.	-55.45%	5.46%	-51.82%	-1.46%	-70.81%	12.85%
Nose	Big Nose	Positive	2.42%	1.25%	2.88%	1.21%	0.24%	2.15%
		Negative	1.35%	1.32%	0.63%	1.03%	4.55%	2.23%
		Rel. Perf.	-78.71%	5.16%	-356.90%	-17.78%	94.75%	3.52%
	Pointy Nose	Positive	1.35%	1.33%	1.24%	1.19%	1.95%	1.88%
		Negative	2.02%	1.46%	2.18%	1.03%	1.25%	1.67%
		Rel. Perf.	33.01%	9.36%	43.20%	-14.76%	-56.30%	-12.94%
Accessories	Heavy Makeup	Positive	0.06%	1.43%	0.03%	1.17%	0.19%	2.08%
		Negative	1.72%	1.23%	1.61%	1.11%	2.28%	1.79%
		Rel. Perf.	96.61%	-15.68%	97.86%	-5.80%	91.71%	-16.26%
	Wearing Hat	Positive	0.82%	1.32%	0.79%	1.22%	0.97%	1.94%
		Negative	1.43%	1.32%	1.40%	1.16%	1.57%	1.85%
		Rel. Perf.	42.85%	0.05%	43.43%	-4.78%	38.05%	-4.73%
	Wearing Lipstick	Positive	0.15%	1.33%	0.16%	0.91%	0.08%	2.13%
		Negative	1.78%	1.23%	1.67%	1.07%	2.35%	1.80%
		Rel. Perf.	91.54%	-7.80%	90.14%	15.24%	96.50%	-18.89%
	Eyeglasses	Positive	0.26%	1.31%	0.00%	1.20%	1.38%	2.16%
		Negative	1.40%	1.30%	1.30%	1.16%	1.88%	1.83%
		Rel. Perf.	81.23%	-0.80%	100.00%	-3.90%	26.99%	-18.24%
Other	Attractive	Positive	0.07%	1.44%	0.00%	1.13%	0.31%	1.83%
		Negative	2.10%	1.26%	1.92%	1.05%	3.04%	1.94%
		Rel. Perf.	96.59%	-13.54%	100.00%	-7.51%	89.94%	5.62%

TABLE XIII: Capsule-Forensics-v2/FF++ - Experiments with Capsule-Forensics-v2 on the FF++ dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	4.39%	4.21%	2.07%	1.67%	6.71%	6.22%
		Negative	3.70%	4.33%	1.19%	1.69%	6.21%	6.25%
		Rel. Perf.	-18.57%	2.85%	-74.32%	1.05%	-7.91%	0.35%
	Young	Positive	3.11%	4.20%	0.65%	1.65%	5.58%	6.66%
		Negative	2.57%	3.61%	1.14%	1.54%	3.99%	6.76%
		Rel. Perf.	-21.32%	-16.38%	43.16%	-7.18%	-39.80%	1.42%
	Asian	Positive	2.05%	3.84%	0.00%	1.78%	4.10%	6.98%
		Negative	4.47%	4.10%	1.97%	1.63%	6.98%	6.23%
		Rel. Perf.	54.20%	6.34%	100.00%	-9.36%	41.29%	-12.08%
	White	Positive	4.68%	4.34%	1.82%	1.72%	7.54%	6.19%
		Negative	3.55%	3.82%	1.96%	1.62%	5.14%	6.17%
		Rel. Perf.	-31.73%	-13.49%	7.44%	-5.90%	-46.70%	-0.35%
	Black	Positive	1.30%	2.16%	1.76%	0.98%	0.83%	5.00%
		Negative	4.03%	3.98%	1.35%	1.61%	6.71%	6.41%
		Rel. Perf.	67.79%	45.67%	-30.55%	38.94%	87.59%	21.96%
Skin	Shiny Skin	Positive	4.17%	3.61%	0.60%	1.56%	7.73%	6.64%
		Negative	6.52%	3.26%	0.00%	1.93%	13.03%	5.15%
		Rel. Perf.	36.06%	-10.69%	All correct	19.13%	40.69%	-28.81%
Hair	Bald	Positive	0.33%	2.81%	0.30%	2.28%	0.37%	4.81%
		Negative	4.38%	4.11%	1.75%	1.65%	7.01%	6.22%
		Rel. Perf.	92.37%	31.76%	83.00%	-38.64%	94.71%	22.55%
	Wavy Hair	Positive	0.00%	2.10%	0.00%	0.00%	0.00%	3.70%
		Negative	4.89%	4.23%	2.38%	1.68%	7.41%	6.10%
		Rel. Perf.	100.00%	50.36%	100.00%	100.00%	100.00%	39.30%
	Receding Hairline	Positive	2.72%	2.56%	1.11%	2.00%	4.33%	4.33%
		Negative	4.26%	3.68%	0.76%	1.49%	7.77%	6.12%
		Rel. Perf.	36.10%	30.51%	-47.02%	-34.00%	44.20%	29.24%
	Bangs	Positive	4.57%	2.50%	0.76%	1.67%	8.38%	6.15%
		Negative	3.98%	4.03%	1.84%	1.64%	6.12%	6.18%
		Rel. Perf.	-14.77%	37.93%	58.69%	-2.02%	-36.84%	0.61%
	Black Hair	Positive	4.50%	3.76%	2.66%	1.64%	6.33%	6.67%
		Negative	3.95%	3.77%	1.60%	1.50%	6.30%	6.51%
		Rel. Perf.	-13.78%	0.14%	-66.13%	-9.08%	-0.49%	-2.37%
	Blond Hair	Positive	2.70%	2.77%	0.14%	1.86%	5.26%	5.26%
		Negative	4.19%	4.20%	1.62%	1.66%	6.77%	6.20%
		Rel. Perf.	35.50%	34.14%	91.16%	-12.03%	22.21%	15.14%
Beard	No Beard	Positive	3.40%	4.16%	1.48%	1.66%	5.33%	6.22%
		Negative	2.60%	2.99%	1.39%	1.81%	3.81%	6.19%
		Rel. Perf.	-30.94%	-39.04%	-6.30%	7.93%	-39.93%	-0.52%
	Mustache	Positive	1.82%	2.34%	0.30%	0.91%	3.33%	5.00%
		Negative	3.88%	4.00%	1.74%	1.61%	6.03%	6.28%
		Rel. Perf.	53.15%	41.35%	82.55%	43.55%	44.68%	20.32%
	Goatee	Positive	0.00%	0.68%	0.00%	0.00%	0.00%	3.33%
		Negative	3.98%	3.98%	1.56%	1.60%	6.40%	6.38%
		Rel. Perf.	100.00%	82.78%	100.00%	100.00%	100.00%	47.75%
Face Geometry	Oval Face	Positive	5.11%	3.98%	0.60%	1.76%	9.61%	7.14%
		Negative	5.46%	3.81%	3.95%	1.66%	6.97%	7.07%
		Rel. Perf.	6.44%	-4.41%	84.77%	-6.50%	-37.93%	-1.09%
	Square Face	Positive	4.00%	4.20%	2.22%	1.68%	5.79%	6.19%
		Negative	4.03%	4.36%	1.07%	1.71%	7.00%	6.15%
		Rel. Perf.	0.78%	3.81%	-107.02%	1.62%	17.28%	-0.74%
	Double Chin	Positive	0.67%	2.16%	0.00%	0.83%	1.33%	5.33%
		Negative	3.88%	4.26%	1.32%	1.68%	6.43%	6.22%
		Rel. Perf.	82.80%	49.28%	100.00%	50.49%	79.28%	14.27%
	Chubby	Positive	3.08%	3.21%	2.52%	2.11%	3.64%	5.15%
		Negative	3.61%	4.10%	0.80%	1.65%	6.42%	6.57%
		Rel. Perf.	14.65%	21.62%	-216.00%	-27.87%	43.33%	21.62%
	Obstructed Forehead	Positive	2.06%	2.21%	0.76%	0.76%	3.36%	5.04%
		Negative	3.94%	4.20%	1.46%	1.64%	6.43%	6.28%
		Rel. Perf.	47.72%	47.51%	47.54%	53.32%	47.76%	19.77%
	Fully Visible Forehead	Positive	2.85%	3.03%	2.55%	2.36%	3.15%	4.55%
		Negative	3.52%	4.33%	0.80%	1.66%	6.24%	6.41%
		Rel. Perf.	19.03%	30.05%	-219.78%	-41.94%	49.56%	29.08%
Mouth	Mouth Closed	Positive	5.53%	2.54%	0.36%	1.27%	10.70%	6.42%
		Negative	4.17%	3.60%	1.79%	1.74%	6.56%	7.36%
		Rel. Perf.	-32.46%	29.50%	79.73%	27.13%	-63.02%	12.76%
	Smiling	Positive	0.00%	0.88%	0.00%	0.00%	0.00%	0.00%
		Negative	4.77%	4.27%	1.54%	1.68%	8.00%	6.12%
		Rel. Perf.	100.00%	79.46%	100.00%	100.00%	100.00%	100.00%
	Big Lips	Positive	4.86%	4.25%	1.92%	1.73%	7.81%	6.17%
		Negative	3.34%	3.86%	2.44%	1.67%	4.24%	6.53%
		Rel. Perf.	-45.68%	-9.88%	21.40%	-3.63%	-84.24%	5.61%
Nose	Big Nose	Positive	3.62%	3.19%	3.90%	1.95%	3.34%	5.97%
		Negative	3.30%	3.76%	0.20%	1.63%	6.40%	6.86%
		Rel. Perf.	-9.67%	15.20%	-1813.01%	-19.56%	47.80%	13.08%
	Pointy Nose	Positive	4.57%	4.18%	1.73%	1.65%	7.41%	6.24%
		Negative	1.84%	3.97%	2.42%	1.69%	1.25%	7.08%
		Rel. Perf.	-148.91%	-5.19%	28.52%	2.25%	-492.80%	11.87%
Accessories	Heavy Makeup	Positive	1.26%	3.67%	0.43%	1.75%	2.08%	7.39%
		Negative	4.84%	4.26%	2.31%	1.68%	7.38%	6.18%
		Rel. Perf.	74.07%	13.92%	81.44%	-4.68%	71.77%	-19.44%
	Wearing Hat	Positive	1.43%	3.23%	1.90%	2.02%	0.97%	4.85%
		Negative	4.04%	4.04%	1.68%	1.62%	6.40%	6.31%
		Rel. Perf.	64.51%	20.13%	-12.97%	-24.57%	84.84%	23.04%
	Wearing Lipstick	Positive	1.40%	3.59%	0.49%	1.56%	2.30%	6.73%
		Negative	4.80%	4.27%	2.19%	1.69%	7.40%	6.19%
		Rel. Perf.	70.90%	15.87%	77.54%	7.56%	68.94%	-8.71%
	Eyeglasses	Positive	3.34%	3.20%	0.00%	1.95%	6.68%	7.27%
		Negative	4.06%	3.97%	1.72%	1.59%	6.40%	6.38%
		Rel. Perf.	17.70%	19.42%	100.00%	-22.72%	-4.42%	-13.97%
Other	Attractive	Positive	0.31%	2.82%	0.00%	2.07%	0.61%	5.20%
		Negative	4.85%	4.16%	2.47%	1.64%	7.23%	6.58%
		Rel. Perf.	93.69%	32.32%	100.00%	-26.67%	91.54%	20.99%

TABLE XIV: EfficientNetB0/DF-1.0 - Experiments with EfficientNetB0 on the DF-1.0 dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	0.07%	0.06%	0.03%	0.00%	0.11%	0.10%
		Negative	0.11%	0.06%	0.00%	0.00%	0.22%	0.11%
		Rel. Perf.	39.32%	-12.10%	-821.51%	-31.64%	51.23%	7.24%
	Young	Positive	0.10%	0.05%	0.00%	0.01%	0.20%	0.11%
		Negative	0.00%	0.06%	0.00%	0.00%	0.00%	0.12%
		Rel. Perf.	All correct	22.42%	All correct	All correct	All correct	9.04%
	Asian	Positive	0.08%	0.06%	0.00%	0.01%	0.15%	0.10%
		Negative	0.00%	0.05%	0.00%	0.00%	0.00%	0.12%
		Rel. Perf.	All correct	-7.14%	All correct	All correct	All correct	23.23%
	White	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Black	Positive	0.00%	0.08%	0.00%	0.00%	0.00%	0.13%
		Negative	0.08%	0.06%	0.01%	0.01%	0.15%	0.10%
		Rel. Perf.	100.00%	-38.45%	100.00%	100.00%	100.00%	-33.71%
Skin	Shiny Skin	Positive	0.00%	0.05%	0.00%	0.00%	0.00%	0.12%
		Negative	0.00%	0.06%	0.00%	0.00%	0.00%	0.18%
		Rel. Perf.	All correct	16.62%	All correct	All correct	All correct	33.23%
Hair	Bald	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wavy Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Receding Hairline	Positive	0.00%	0.09%	0.00%	0.00%	0.00%	0.32%
		Negative	0.14%	0.05%	0.00%	0.00%	0.28%	0.11%
		Rel. Perf.	100.00%	-89.27%	All correct	All correct	100.00%	-206.71%
	Bangs	Positive	0.00%	0.05%	0.00%	0.00%	0.00%	0.11%
		Negative	0.09%	0.06%	0.01%	0.01%	0.17%	0.09%
		Rel. Perf.	100.00%	21.64%	100.00%	100.00%	100.00%	-15.89%
	Black Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Blond Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Beard	No Beard	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Mustache	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Goatee	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Face Geometry	Oval Face	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Square Face	Positive	0.08%	0.06%	0.03%	0.00%	0.13%	0.10%
		Negative	0.08%	0.06%	0.00%	0.00%	0.15%	0.10%
		Rel. Perf.	-1.07%	-5.43%	-751.57%	-21.65%	14.02%	-1.67%
	Double Chin	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Chubby	Positive	0.11%	0.05%	0.00%	0.00%	0.22%	0.10%
		Negative	0.00%	0.05%	0.00%	0.00%	0.00%	0.12%
		Rel. Perf.	All correct	5.11%	All correct	All correct	All correct	12.53%
	Obstructed Forehead	Positive	0.07%	0.05%	0.14%	0.00%	0.00%	0.11%
		Negative	0.07%	0.06%	0.00%	0.01%	0.14%	0.10%
		Rel. Perf.	7.55%	28.21%	-4496.49%	100.00%	100.00%	-13.03%
	Fully Visible Forehead	Positive	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Negative	0.05%	0.06%	0.01%	0.01%	0.09%	0.10%
		Rel. Perf.	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Mouth	Mouth Closed	Positive	0.00%	0.08%	0.00%	0.00%	0.00%	0.12%
		Negative	0.00%	0.06%	0.00%	0.00%	0.00%	0.11%
		Rel. Perf.	All correct	-26.56%	All correct	All correct	All correct	-8.26%
	Smiling	Positive	0.00%	0.10%	0.00%	0.00%	0.00%	0.42%
		Negative	0.10%	0.06%	0.03%	0.00%	0.16%	0.09%
		Rel. Perf.	100.00%	-74.66%	100.00%	100.00%	100.00%	-340.83%
	Big Lips	Positive	0.05%	0.06%	0.00%	0.00%	0.11%	0.09%
		Negative	0.00%	0.04%	0.00%	0.00%	0.00%	0.12%
		Rel. Perf.	All correct	-36.99%	All correct	All correct	All correct	22.88%
Nose	Big Nose	Positive	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Negative	0.00%	0.05%	0.00%	0.00%	0.00%	0.12%
		Rel. Perf.	All correct	100.00%	All correct	All correct	All correct	100.00%
	Pointy Nose	Positive	0.00%	0.08%	0.00%	0.00%	0.00%	0.00%
		Negative	0.08%	0.05%	0.02%	0.01%	0.14%	0.10%
		Rel. Perf.	100.00%	-37.90%	100.00%	100.00%	100.00%	100.00%
Accessories	Heavy Makeup	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wearing Hat	Positive	0.00%	0.04%	0.00%	0.00%	0.00%	0.10%
		Negative	0.08%	0.06%	0.00%	0.01%	0.16%	0.09%
		Rel. Perf.	100.00%	23.67%	100.00%	100.00%	100.00%	-6.09%
	Wearing Lipstick	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Eyeglasses	Positive	0.00%	0.07%	0.00%	0.00%	0.00%	0.20%
		Negative	0.08%	0.05%	0.02%	0.01%	0.14%	0.10%
		Rel. Perf.	100.00%	-19.98%	100.00%	100.00%	100.00%	-110.81%
Other	Attractive	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-

TABLE XV: Xception/DF-1.0 - Experiments with Xception on the DF-1.0 dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	0.01%	0.01%	0.02%	0.02%	0.00%	0.00%
		Negative	0.00%	0.00%	0.01%	0.01%	0.00%	0.00%
		Rel. Perf.	-163.29%	-118.88%	-163.29%	-31.64%	All correct	All correct
	Young	Positive	0.00%	0.00%	0.01%	0.02%	0.00%	0.00%
		Negative	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Rel. Perf.	All correct	All correct	All correct	All correct	All correct	All correct
	Asian	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	White	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Black	Positive	0.00%	0.01%	0.00%	0.01%	0.00%	0.00%
		Negative	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Rel. Perf.	All correct	All correct	All correct	All correct	All correct	All correct
Skin	Shiny Skin	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Hair	Bald	Positive	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Negative	0.01%	0.01%	0.01%	0.01%	0.00%	0.00%
		Rel. Perf.	100.00%	100.00%	100.00%	100.00%	All correct	All correct
	Wavy Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Receding Hairline	Positive	0.06%	0.00%	0.12%	0.00%	0.00%	0.00%
		Negative	0.01%	0.00%	0.02%	0.02%	0.00%	0.00%
		Rel. Perf.	-560.87%	All correct	-560.87%	100.00%	All correct	All correct
	Bangs	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Black Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Blond Hair	Positive	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Negative	0.02%	0.00%	0.04%	0.02%	0.00%	0.00%
		Rel. Perf.	100.00%	All correct	100.00%	100.00%	All correct	All correct
Beard	No Beard	Positive	0.17%	0.00%	0.34%	0.00%	0.00%	0.00%
		Negative	0.00%	0.01%	0.01%	0.01%	0.00%	0.00%
		Rel. Perf.	-4554.11%	100.00%	-4554.11%	100.00%	All correct	All correct
	Mustache	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Goatee	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Face Geometry	Oval Face	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Square Face	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Double Chin	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Chubby	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Obstructed Forehead	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Fully Visible Forehead	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Mouth	Mouth Closed	Positive	0.01%	0.01%	0.02%	0.02%	0.00%	0.00%
		Negative	0.00%	0.01%	0.01%	0.01%	0.00%	0.00%
		Rel. Perf.	-143.31%	-63.02%	-143.31%	-21.65%	All correct	All correct
	Smiling	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Big Lips	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Nose	Big Nose	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Pointy Nose	Positive	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Negative	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Rel. Perf.	All correct	All correct	All correct	All correct	All correct	All correct
Accessories	Heavy Makeup	Positive	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Negative	0.00%	0.01%	0.01%	0.01%	0.00%	0.00%
		Rel. Perf.	100.00%	100.00%	100.00%	100.00%	All correct	All correct
	Wearing Hat	Positive	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Negative	0.00%	0.01%	0.01%	0.02%	0.00%	0.00%
		Rel. Perf.	100.00%	100.00%	100.00%	100.00%	All correct	All correct
	Wearing Lipstick	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Eyeglasses	Positive	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		Negative	0.00%	0.00%	0.01%	0.01%	0.00%	0.00%
		Rel. Perf.	100.00%	All correct	100.00%	100.00%	All correct	All correct
Other	Attractive	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-

TABLE XVI: Capsule-Forensics-v2/DF-1.0 - Experiments with Capsule-Forensics-v2 on the DF-1.0 dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	-35.40%	All correct	All correct	All correct	All correct
	Young	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	All correct	All correct	All correct	All correct	All correct
	Asian	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	All correct	All correct	All correct	All correct	All correct
	White	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Black	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	100.00%	All correct	All correct	All correct	100.00%
Skin	Shiny Skin	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	All correct	All correct	All correct	All correct	All correct
Hair	Bald	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wavy Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Receding Hairline	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	100.00%	All correct	All correct	All correct	All correct
	Bangs	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	100.00%	100.00%	All correct	All correct	100.00%	100.00%
	Black Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Blond Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Beard	No Beard	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Mustache	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Goatee	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
Face Geometry	Oval Face	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Square Face	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	-35.96%	All correct	All correct	All correct	All correct
	Double Chin	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Chubby	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	All correct	All correct	All correct	All correct	All correct
	Obstructed Forehead	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	100.00%	All correct	All correct	All correct	100.00%
	Fully Visible Forehead	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	100.00%	All correct	All correct	All correct	100.00%
Mouth	Mouth Closed	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	All correct	All correct	All correct	All correct	All correct
	Smiling	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	100.00%	All correct	All correct	All correct	All correct
	Big Lips	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	All correct	All correct	All correct	All correct	All correct
Nose	Big Nose	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	All correct	All correct	All correct	All correct	All correct
	Pointy Nose	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	100.00%	100.00%	All correct	All correct	100.00%	100.00%
Accessories	Heavy Makeup	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wearing Hat	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	100.00%	All correct	All correct	All correct	100.00%
	Wearing Lipstick	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Eyeglasses	Positive	0	0	0	0	0	0
		Negative	0	0	0	0	0	0
		Rel. Perf.	All correct	100.00%	All correct	All correct	All correct	100.00%
Other	Attractive	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-

TABLE XVII: EfficientNetB0/DFDC - Experiments with EfficientNetB0 on the DFDC dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	24.12%	23.66%	13.00%	11.89%	35.25%	35.52%
		Negative	21.88%	23.66%	10.18%	11.83%	33.58%	35.68%
		Rel. Perf.	-10.26%	-0.01%	-27.77%	-0.52%	-4.95%	0.46%
	Young	Positive	17.66%	23.57%	9.25%	10.93%	26.06%	35.93%
		Negative	22.33%	23.50%	13.55%	11.54%	31.11%	35.76%
		Rel. Perf.	20.92%	-0.29%	31.72%	5.27%	16.22%	-0.48%
	Asian	Positive	19.52%	22.66%	10.30%	11.23%	28.73%	35.58%
		Negative	23.76%	23.75%	11.71%	11.83%	35.81%	35.67%
		Rel. Perf.	17.86%	4.59%	12.04%	5.08%	19.76%	0.23%
	White	Positive	23.46%	23.69%	9.07%	11.84%	37.85%	35.57%
		Negative	23.55%	23.68%	13.89%	11.77%	33.21%	35.63%
		Rel. Perf.	0.40%	-0.01%	34.75%	-0.67%	-13.97%	0.15%
	Black	Positive	22.52%	23.57%	10.74%	11.70%	34.31%	35.76%
		Negative	23.19%	23.73%	11.32%	11.74%	35.06%	35.47%
		Rel. Perf.	2.88%	0.67%	5.17%	0.27%	2.14%	-0.82%
Skin	Shiny Skin	Positive	20.71%	23.45%	13.94%	11.54%	27.49%	36.00%
		Negative	25.01%	23.51%	8.30%	11.77%	41.72%	35.77%
		Rel. Perf.	17.19%	0.25%	-67.85%	1.97%	34.11%	-0.67%
Hair	Bald	Positive	19.86%	23.01%	10.28%	10.84%	29.44%	35.17%
		Negative	23.04%	23.80%	11.51%	11.74%	34.57%	35.51%
		Rel. Perf.	13.83%	3.30%	10.73%	7.64%	14.86%	0.95%
	Wavy Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Receding Hairline	Positive	16.47%	23.12%	7.52%	11.34%	25.42%	35.52%
		Negative	19.54%	23.74%	7.34%	10.78%	31.75%	35.77%
		Rel. Perf.	15.72%	2.61%	-2.46%	-5.19%	19.93%	0.71%
	Bangs	Positive	16.37%	22.57%	2.52%	11.05%	30.22%	35.83%
		Negative	23.37%	23.77%	11.73%	11.83%	35.00%	35.70%
		Rel. Perf.	29.96%	5.07%	78.54%	6.63%	13.67%	-0.36%
	Black Hair	Positive	19.26%	23.63%	13.98%	11.12%	24.54%	35.87%
		Negative	22.21%	23.42%	7.66%	11.63%	36.75%	35.77%
		Rel. Perf.	13.27%	-0.93%	-82.36%	4.36%	33.21%	-0.30%
	Blond Hair	Positive	5.23%	21.46%	0.00%	5.26%	10.47%	33.72%
		Negative	23.39%	23.77%	12.57%	11.75%	34.21%	35.69%
		Rel. Perf.	77.63%	9.72%	100.00%	55.20%	69.41%	5.51%
Beard	No Beard	Positive	21.02%	23.67%	10.34%	11.81%	31.69%	35.56%
		Negative	22.96%	23.50%	14.38%	11.48%	31.54%	35.83%
		Rel. Perf.	8.48%	-0.72%	28.13%	-2.88%	-0.47%	0.74%
	Mustache	Positive	28.22%	23.41%	25.04%	10.85%	31.40%	35.71%
		Negative	21.50%	23.79%	9.91%	11.75%	33.09%	35.56%
		Rel. Perf.	-31.25%	1.61%	-152.66%	7.69%	5.12%	-0.41%
	Goatee	Positive	22.48%	23.37%	16.56%	11.28%	28.39%	35.77%
		Negative	22.52%	23.72%	10.59%	11.78%	34.45%	35.55%
		Rel. Perf.	0.19%	1.48%	-56.39%	4.23%	17.58%	-0.62%
Face Geometry	Oval Face	Positive	19.45%	21.53%	15.20%	9.20%	23.70%	35.89%
		Negative	26.12%	23.56%	10.80%	11.74%	41.44%	35.42%
		Rel. Perf.	25.54%	8.59%	-40.75%	21.64%	42.81%	-1.32%
	Square Face	Positive	24.04%	23.66%	13.22%	11.74%	34.86%	35.67%
		Negative	22.45%	23.77%	10.53%	11.74%	34.36%	35.61%
		Rel. Perf.	-7.10%	0.47%	-25.53%	0.02%	-1.45%	-0.17%
	Double Chin	Positive	28.60%	23.47%	27.91%	10.88%	29.29%	35.63%
		Negative	21.18%	23.50%	10.90%	11.35%	31.47%	35.66%
		Rel. Perf.	-35.02%	0.13%	-156.12%	4.07%	6.92%	0.09%
	Chubby	Positive	24.90%	23.63%	15.47%	10.85%	34.32%	35.82%
		Negative	19.74%	23.38%	10.08%	11.00%	29.40%	35.53%
		Rel. Perf.	-26.10%	-1.04%	-53.50%	1.36%	-16.71%	-0.81%
	Obstructed Forehead	Positive	31.07%	23.34%	19.31%	11.13%	42.82%	35.45%
		Negative	22.53%	23.75%	11.08%	11.75%	33.98%	35.59%
		Rel. Perf.	-37.89%	1.73%	-74.26%	5.25%	-26.02%	0.40%
	Fully Visible Forehead	Positive	26.21%	21.83%	20.21%	9.57%	32.20%	32.20%
		Negative	22.00%	23.72%	10.45%	11.78%	33.55%	35.57%
		Rel. Perf.	-19.14%	7.97%	-93.50%	18.71%	4.01%	9.46%
Mouth	Mouth Closed	Positive	23.17%	23.63%	9.77%	11.31%	36.58%	35.55%
		Negative	20.29%	23.64%	10.79%	11.32%	29.79%	35.72%
		Rel. Perf.	-14.21%	0.04%	9.49%	0.13%	-22.80%	0.46%
	Smiling	Positive	13.63%	25.16%	5.69%	5.69%	21.58%	34.02%
		Negative	25.53%	23.68%	12.59%	11.83%	38.47%	35.63%
		Rel. Perf.	46.59%	-6.22%	54.79%	51.91%	43.91%	4.51%
	Big Lips	Positive	23.19%	23.66%	13.82%	11.74%	32.56%	35.64%
		Negative	24.46%	23.56%	8.60%	11.82%	40.32%	35.66%
		Rel. Perf.	5.20%	-0.45%	-60.65%	0.70%	19.24%	0.04%
Nose	Big Nose	Positive	21.16%	23.62%	15.35%	10.98%	26.98%	35.84%
		Negative	27.04%	23.21%	2.15%	11.31%	51.93%	35.70%
		Rel. Perf.	21.73%	-1.76%	-613.31%	2.89%	48.04%	-0.39%
	Pointy Nose	Positive	23.22%	23.80%	10.70%	11.76%	35.73%	35.58%
		Negative	22.60%	23.56%	11.57%	11.76%	33.62%	35.44%
		Rel. Perf.	-2.74%	-0.98%	7.51%	-0.02%	-6.26%	-0.39%
Accessories	Heavy Makeup	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wearing Hat	Positive	23.10%	23.27%	8.83%	10.68%	37.37%	35.68%
		Negative	22.25%	23.78%	11.36%	11.73%	33.14%	35.51%
		Rel. Perf.	-3.84%	2.14%	22.24%	8.94%	-12.78%	-0.50%
	Wearing Lipstick	Positive	12.77%	22.46%	16.92%	5.38%	8.61%	34.72%
		Negative	25.03%	23.79%	12.02%	11.76%	38.05%	35.58%
		Rel. Perf.	49.00%	5.59%	-40.84%	54.21%	77.37%	2.41%
	Eyeglasses	Positive	21.29%	23.50%	10.81%	11.73%	31.77%	35.61%
		Negative	23.49%	23.80%	11.61%	11.74%	35.37%	35.47%
		Rel. Perf.	9.36%	1.28%	6.89%	0.03%	10.17%	-0.40%
Other	Attractive	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-

TABLE XVIII: Xception/DFDC - Experiments with Xception on the DFDC dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	28.83%	27.98%	21.35%	17.42%	36.31%	38.65%
		Negative	25.73%	28.00%	12.36%	17.31%	39.10%	38.72%
		Rel. Perf.	-12.05%	0.09%	-72.74%	-0.62%	7.13%	0.20%
	Young	Positive	22.95%	27.40%	9.56%	17.16%	36.34%	38.91%
		Negative	27.93%	27.38%	20.47%	17.28%	35.40%	38.37%
		Rel. Perf.	17.84%	-0.07%	53.31%	0.74%	-2.67%	-1.41%
	Asian	Positive	22.70%	26.32%	6.87%	17.70%	38.53%	38.50%
		Negative	27.46%	27.92%	17.25%	17.00%	37.67%	38.87%
		Rel. Perf.	17.32%	5.73%	60.18%	-4.14%	-2.30%	0.96%
	White	Positive	27.23%	28.06%	14.20%	17.39%	40.26%	38.74%
		Negative	27.72%	28.02%	19.01%	17.36%	36.44%	38.63%
		Rel. Perf.	1.77%	-0.16%	25.30%	-0.15%	-10.50%	-0.29%
	Black	Positive	26.51%	27.93%	15.45%	17.51%	37.56%	38.62%
		Negative	27.57%	27.92%	16.67%	16.97%	38.47%	38.77%
		Rel. Perf.	3.85%	-0.03%	7.29%	-3.17%	2.35%	0.39%
Skin	Shiny Skin	Positive	24.95%	27.50%	17.94%	17.43%	31.97%	38.44%
		Negative	29.35%	27.86%	13.54%	17.40%	45.17%	38.69%
		Rel. Perf.	14.99%	1.28%	-32.46%	-0.21%	29.21%	0.65%
Hair	Bald	Positive	27.88%	26.70%	17.18%	17.72%	38.58%	38.40%
		Negative	27.37%	27.95%	16.73%	16.98%	38.01%	38.82%
		Rel. Perf.	-1.85%	4.47%	-2.70%	-4.35%	-1.48%	1.09%
	Wavy Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Receding Hairline	Positive	17.22%	26.67%	5.95%	17.64%	28.50%	38.27%
		Negative	25.75%	27.32%	11.99%	17.17%	39.52%	38.49%
		Rel. Perf.	33.12%	2.38%	50.40%	-2.76%	27.87%	0.59%
	Bangs	Positive	24.87%	27.09%	2.80%	18.04%	46.95%	39.37%
		Negative	26.96%	27.93%	17.04%	17.00%	36.88%	38.95%
		Rel. Perf.	7.73%	3.00%	83.58%	-6.13%	-27.31%	-1.07%
	Black Hair	Positive	22.69%	27.42%	12.48%	17.22%	32.90%	38.61%
		Negative	26.50%	27.35%	13.42%	17.50%	39.58%	38.42%
		Rel. Perf.	14.39%	-0.27%	7.01%	1.59%	16.89%	-0.50%
	Blond Hair	Positive	22.67%	27.87%	0.00%	19.30%	45.35%	38.37%
		Negative	27.48%	27.92%	17.75%	16.96%	37.21%	38.94%
		Rel. Perf.	17.49%	0.20%	100.00%	-13.82%	-21.87%	1.47%
Beard	No Beard	Positive	24.92%	27.95%	11.55%	17.27%	38.30%	38.67%
		Negative	29.36%	27.77%	20.09%	17.25%	38.64%	38.67%
		Rel. Perf.	15.13%	-0.68%	42.53%	-0.15%	0.88%	0.01%
	Mustache	Positive	31.75%	27.40%	23.14%	17.51%	40.35%	38.19%
		Negative	25.40%	27.99%	12.03%	17.10%	38.77%	38.59%
		Rel. Perf.	-25.01%	2.11%	-92.44%	-2.44%	-4.09%	1.05%
	Goatee	Positive	24.75%	27.18%	12.98%	17.56%	36.53%	38.16%
		Negative	26.38%	27.92%	14.55%	17.05%	38.21%	38.69%
		Rel. Perf.	6.17%	2.64%	10.81%	-2.98%	4.41%	1.37%
Face Geometry	Oval Face	Positive	22.35%	28.30%	14.00%	17.20%	30.70%	38.60%
		Negative	31.48%	27.83%	18.17%	17.53%	44.80%	38.58%
		Rel. Perf.	29.00%	-1.69%	22.93%	1.89%	31.47%	-0.06%
	Square Face	Positive	28.34%	27.96%	19.84%	17.38%	36.83%	38.75%
		Negative	26.30%	27.92%	14.38%	17.07%	38.22%	38.60%
		Rel. Perf.	-7.74%	-0.14%	-37.96%	-1.85%	3.63%	-0.40%
	Double Chin	Positive	37.62%	26.28%	38.36%	17.46%	36.87%	38.40%
		Negative	25.65%	27.78%	11.31%	17.07%	40.00%	38.80%
		Rel. Perf.	-46.63%	5.40%	-239.30%	-2.30%	7.83%	1.03%
	Chubby	Positive	29.86%	27.43%	21.09%	17.22%	38.64%	38.43%
		Negative	23.88%	27.14%	7.29%	17.91%	40.47%	38.11%
		Rel. Perf.	-25.04%	-1.07%	-189.14%	3.82%	4.53%	-0.83%
	Obstructed Forehead	Positive	39.22%	27.25%	25.53%	17.35%	52.91%	37.90%
		Negative	26.39%	27.94%	16.98%	17.13%	35.80%	38.63%
		Rel. Perf.	-48.60%	2.49%	-50.33%	-1.29%	-47.79%	1.88%
	Fully Visible Forehead	Positive	24.17%	29.35%	20.21%	16.67%	28.14%	39.32%
		Negative	26.25%	27.91%	14.23%	17.04%	38.28%	38.68%
		Rel. Perf.	7.92%	-5.14%	-42.08%	2.21%	26.50%	-1.67%
Mouth	Mouth Closed	Positive	26.45%	27.39%	14.12%	17.00%	38.77%	38.15%
		Negative	23.31%	27.41%	14.53%	17.10%	32.10%	37.88%
		Rel. Perf.	-13.44%	0.06%	2.79%	0.59%	-20.78%	-0.70%
	Smiling	Positive	26.89%	29.80%	8.13%	15.45%	45.64%	38.59%
		Negative	29.57%	28.02%	18.28%	17.41%	40.86%	38.68%
		Rel. Perf.	9.08%	-6.34%	55.53%	11.27%	-11.70%	0.24%
	Big Lips	Positive	27.66%	28.01%	18.74%	17.38%	36.59%	38.71%
		Negative	27.90%	27.96%	14.25%	17.49%	41.54%	38.78%
		Rel. Perf.	0.84%	-0.18%	-31.44%	0.67%	11.92%	0.18%
Nose	Big Nose	Positive	28.28%	27.39%	18.61%	17.22%	37.96%	38.26%
		Negative	35.97%	26.28%	6.38%	17.63%	65.55%	38.93%
		Rel. Perf.	21.36%	-4.20%	-191.43%	2.31%	42.09%	1.72%
	Pointy Nose	Positive	27.03%	27.97%	16.50%	17.11%	37.57%	38.61%
		Negative	26.76%	27.91%	15.65%	17.42%	37.86%	38.48%
		Rel. Perf.	-1.03%	-0.21%	-5.39%	1.78%	0.77%	-0.33%
Accessories	Heavy Makeup	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wearing Hat	Positive	28.78%	27.01%	11.20%	17.47%	46.35%	38.64%
		Negative	26.32%	27.94%	16.58%	16.95%	36.05%	38.81%
		Rel. Perf.	-9.36%	3.32%	32.46%	-3.03%	-28.60%	0.45%
	Wearing Lipstick	Positive	25.66%	30.47%	30.77%	14.62%	20.56%	39.44%
		Negative	29.16%	27.96%	19.75%	17.11%	38.57%	38.69%
		Rel. Perf.	11.99%	-8.99%	-55.80%	14.58%	46.70%	-1.94%
	Eyeglasses	Positive	26.26%	27.84%	16.89%	17.45%	35.64%	38.69%
		Negative	27.21%	27.95%	16.13%	16.97%	38.30%	38.80%
		Rel. Perf.	3.49%	0.39%	-4.75%	-2.79%	6.95%	0.27%
Other	Attractive	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-

TABLE XIX: Capsule-Forensics-v2/DFDC - Experiments with Capsule-Forensics-v2 on the DFDC dataset.

Category	Attribute	Class	Balanced Error		Pristine Error		Fake Error	
			Data	Control	Data	Control	Data	Control
Demographics	Male	Positive	27.64%	26.80%	12.24%	10.91%	43.05%	43.09%
		Negative	24.15%	26.76%	9.06%	10.86%	39.23%	43.27%
		Rel. Perf.	-14.48%	-0.18%	-35.12%	-0.49%	-9.71%	0.41%
	Young	Positive	19.74%	27.38%	8.67%	10.32%	30.82%	43.61%
		Negative	27.38%	27.08%	15.15%	10.98%	39.61%	43.82%
		Rel. Perf.	27.88%	-1.10%	42.76%	5.95%	22.19%	0.46%
	Asian	Positive	22.82%	27.00%	10.30%	11.36%	35.35%	42.94%
		Negative	26.90%	26.83%	10.86%	10.75%	42.95%	42.92%
		Rel. Perf.	15.16%	-0.62%	5.09%	-5.63%	17.70%	-0.06%
	White	Positive	26.10%	26.86%	7.76%	11.00%	44.45%	43.25%
		Negative	26.35%	26.91%	12.89%	10.82%	39.82%	43.15%
		Rel. Perf.	0.94%	0.20%	39.83%	-1.65%	-11.65%	-0.24%
	Black	Positive	25.28%	27.15%	8.63%	11.21%	41.93%	43.39%
		Negative	26.26%	26.83%	10.95%	10.68%	41.58%	42.90%
		Rel. Perf.	3.74%	-1.17%	21.14%	-4.89%	-0.84%	-1.15%
Skin	Shiny Skin	Positive	24.61%	27.03%	13.35%	11.01%	35.87%	43.75%
		Negative	27.68%	27.00%	7.08%	11.05%	48.27%	43.39%
		Rel. Perf.	11.09%	-0.09%	-88.46%	0.30%	25.70%	-0.83%
Hair	Bald	Positive	22.65%	27.02%	9.77%	11.10%	35.53%	42.70%
		Negative	25.92%	26.87%	10.74%	10.69%	41.10%	42.89%
		Rel. Perf.	12.63%	-0.59%	9.03%	-3.81%	13.57%	0.44%
	Wavy Hair	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Receding Hairline	Positive	20.98%	27.03%	7.12%	10.73%	34.85%	43.36%
		Negative	25.48%	27.50%	13.38%	10.47%	37.58%	43.74%
		Rel. Perf.	17.64%	1.72%	46.79%	-2.48%	7.26%	0.86%
	Bangs	Positive	20.56%	26.80%	5.87%	11.05%	35.24%	42.81%
		Negative	25.95%	26.83%	10.52%	10.75%	41.38%	42.96%
		Rel. Perf.	20.79%	0.11%	44.19%	-2.83%	14.84%	0.33%
	Black Hair	Positive	21.69%	27.15%	11.84%	10.44%	31.54%	43.55%
		Negative	25.90%	27.00%	6.20%	11.10%	45.59%	43.74%
		Rel. Perf.	16.24%	-0.55%	-90.95%	5.91%	30.82%	0.42%
	Blond Hair	Positive	23.26%	20.51%	0.00%	7.02%	46.51%	41.86%
		Negative	26.10%	26.85%	11.32%	10.60%	40.88%	42.97%
		Rel. Perf.	10.91%	23.62%	100.00%	33.80%	-13.77%	2.58%
Beard	No Beard	Positive	23.50%	26.77%	9.93%	10.84%	37.06%	43.09%
		Negative	26.58%	27.12%	13.78%	10.91%	39.37%	43.58%
		Rel. Perf.	11.59%	1.30%	27.93%	0.70%	5.87%	1.13%
	Mustache	Positive	32.91%	27.42%	27.29%	11.16%	38.54%	43.57%
		Negative	24.10%	26.87%	9.56%	10.79%	38.64%	43.04%
		Rel. Perf.	-36.57%	-2.07%	-185.51%	-3.38%	0.26%	-1.22%
	Goatee	Positive	22.72%	27.33%	10.24%	10.85%	35.21%	43.59%
		Negative	25.46%	26.80%	10.05%	10.69%	40.88%	42.95%
		Rel. Perf.	10.75%	-1.98%	-1.93%	-1.51%	13.86%	-1.48%
Face Geometry	Oval Face	Positive	22.04%	27.92%	18.80%	8.80%	25.28%	41.31%
		Negative	30.00%	27.04%	13.41%	11.05%	46.60%	43.28%
		Rel. Perf.	26.54%	-3.27%	-40.16%	20.33%	45.74%	4.55%
	Square Face	Positive	27.12%	26.99%	11.94%	10.98%	42.30%	43.27%
		Negative	25.12%	26.84%	9.67%	10.74%	40.58%	43.10%
		Rel. Perf.	-7.93%	-0.57%	-23.46%	-2.27%	-4.24%	-0.40%
	Double Chin	Positive	40.56%	27.68%	47.41%	11.42%	33.70%	42.62%
		Negative	24.14%	27.15%	8.64%	10.80%	39.65%	43.40%
		Rel. Perf.	-67.97%	-1.93%	-448.88%	-5.72%	15.02%	1.80%
	Chubby	Positive	29.46%	27.42%	21.00%	10.40%	37.92%	43.83%
		Negative	23.25%	27.31%	7.64%	10.96%	38.87%	43.24%
		Rel. Perf.	-26.69%	-0.41%	-174.96%	5.12%	2.44%	-1.36%
	Obstructed Forehead	Positive	36.91%	28.04%	25.70%	11.62%	48.12%	41.27%
		Negative	25.07%	26.82%	8.90%	10.78%	41.24%	43.04%
		Rel. Perf.	-47.22%	-4.56%	-188.69%	-7.83%	-16.69%	4.12%
	Fully Visible Forehead	Positive	29.18%	27.42%	24.47%	9.22%	33.90%	40.00%
		Negative	24.67%	26.80%	9.08%	10.70%	40.27%	42.93%
		Rel. Perf.	-18.28%	-2.34%	-169.62%	13.82%	15.82%	6.83%
Mouth	Mouth Closed	Positive	24.95%	27.45%	7.12%	10.65%	42.78%	43.25%
		Negative	24.39%	27.42%	9.82%	10.55%	38.96%	43.16%
		Rel. Perf.	-2.30%	-0.09%	27.50%	-0.96%	-9.80%	-0.20%
	Smiling	Positive	20.91%	28.15%	4.07%	5.69%	37.76%	40.66%
		Negative	28.50%	26.86%	9.72%	10.87%	47.29%	43.22%
		Rel. Perf.	26.63%	-4.79%	58.17%	47.66%	20.15%	5.92%
	Big Lips	Positive	26.10%	27.03%	12.28%	10.81%	39.92%	43.26%
		Negative	27.76%	27.05%	8.45%	10.90%	47.06%	43.38%
		Rel. Perf.	5.98%	0.08%	-45.26%	0.86%	15.18%	0.28%
Nose	Big Nose	Positive	26.24%	27.42%	19.28%	10.33%	33.21%	43.74%
		Negative	32.97%	27.53%	3.26%	10.69%	62.69%	43.22%
		Rel. Perf.	20.41%	0.40%	-491.07%	3.32%	47.03%	-1.20%
	Pointy Nose	Positive	26.36%	26.84%	10.05%	10.75%	42.67%	43.08%
		Negative	25.71%	26.97%	10.65%	11.00%	40.76%	43.23%
		Rel. Perf.	-2.53%	0.51%	5.63%	2.27%	-4.67%	0.36%
Accessories	Heavy Makeup	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-
	Wearing Hat	Positive	25.41%	27.38%	4.21%	10.91%	46.60%	42.87%
		Negative	25.37%	26.88%	11.05%	10.69%	39.70%	42.92%
		Rel. Perf.	-0.14%	-1.88%	61.84%	-2.12%	-17.39%	0.11%
	Wearing Lipstick	Positive	26.93%	27.30%	36.92%	5.38%	16.94%	40.28%
		Negative	27.97%	26.85%	10.89%	10.76%	45.05%	42.97%
		Rel. Perf.	3.70%	-1.66%	-239.21%	49.98%	62.39%	6.25%
	Eyeglasses	Positive	24.33%	27.06%	10.46%	11.07%	38.19%	43.35%
		Negative	26.35%	26.87%	10.41%	10.70%	42.28%	42.91%
		Rel. Perf.	7.66%	-0.71%	-0.49%	-3.50%	9.67%	-1.02%
Other	Attractive	Positive	-	-	-	-	-	-
		Negative	-	-	-	-	-	-
		Rel. Perf.	-	-	-	-	-	-