2023 IEEE 6th International Conference on Multimedia Information Processing and Retrieval (MIPR)

2023 IEEE 6th International Conference on Multimedia Information Processing and Retrieval (MIPR) | 979-8-3503-0781-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/MIPR59079.2023.00032

A Multi-stage Bias Reduction Framework for Eye Gaze Detection

Omkar N. Kulkarni[1], Shashank Arora[1], Aryan Mishra[2], Vivek K. Singh[3], and Pradeep K. Atrey[1]
[1]Albany Lab for Privacy and Security, University at Albany, State University of New York, NY, USA
[2]Westview High School, Portland, OR, USA, [3] Rutgers University, NJ, USA
Email: {onkulkarni, sarora3, patrey}@albany.edu, aryanspmishra@gmail.com, v.singh@rutgers.edu

Abstract

Eye gaze detection is an important component of many multimedia processing algorithms, including user authentication, online education, and medical diagnostics. Hence, it is important to ensure that gaze detection works equally well for different sections of society. For instance, if such algorithms work well for men and not women, this would amplify existing societal biases to provide more security, education, and medical functionalities to men than women. Here, we audit one of the state-of-the-art gaze detection algorithms for gender bias. Audit results suggest that the algorithm performs better for the male group than the female group, indicating a gender bias. To tackle this challenge, we propose a multi-stage bias reduction framework that considers multiple subtasks performed at different stages during the course of the gaze detection algorithm. Like many multimedia algorithms, the decisions made at each stage can impact the performance of the next algorithm stage. Hence, we have designed a framework that finds optimal algorithmic parameters to support high fairness and accuracy by holistically considering multiple stages. The results suggest that the proposed approach yields promising results in terms of fairness and accuracy, thus yielding a path toward accurate and fair eye gaze detection.

1. Introduction

Human emotions and expressions are complex, and determining them is an important multimedia processing task in any non-verbal communication. The human face, and especially the eyes, are one of the most important factors in it. The position of the pupil signifies where the person intends to look, called the gaze of that person.

Gaze detection algorithms aim to estimate the direction of a person's gaze from an image or a video of their face. They have many applications in various domains, such as authentication [5], education [19] etc. However, designing accurate and robust gaze detection algorithms is challenging due to factors such as head pose variation, eye occlusion, illumination changes, camera quality, etc.

Recent results in multimedia research [3,20] have shown that multiple face analysis algorithms are susceptible to bias (systematic differences in classifier performance) with respect to different social groups such as race, gender, age, etc. For instance, if such algorithms yield high accuracy for men and not women, the underlying societal bias toward males would be aggravated. Therefore, it is important to audit gaze detection algorithms for any potential bias and try to mitigate it, if possible. In this paper, we address the following two research questions:

1. Are state-of-the-art gaze detection algorithms susceptible to gender bias?

2. How can we design a gaze detection algorithm that is fair and accurate for different genders?

We answer the first question by auditing the performance of the gaze detection algorithm proposed in [11], which builds upon the Retinaface [6] algorithm for face and pupil detection and Dlib-ml [7] for eye detection across genders using a large public data set (CelebA [12]). This algorithm is selected for auditing in order to check and rectify the gaze detection algorithm we proposed previously. We tackle the second question by proposing a novel framework that considers multiple subtasks and stages of the gaze detection process and optimizes the algorithmic parameters holistically across different stages. Our framework is motivated by the need to ensure that gaze detection algorithms work equally well for different sections of society and do not amplify existing biases or discrimination.

The key contributions of this paper are:

1. Audit of the gaze detection algorithm for gender bias.

2. A new framework for multi-stage bias reduction in the gaze detection algorithm.

The remainder of this paper is organized as follows: In Section 2 we describe the related literature. Section 3 describes

2770-4319/23/$31.00 ©2023 IEEE
DOI 10.1109/MIPR59079.2023.00032

some background on the gaze detection method that is audited in this paper. In Section 4, we explain the dataset and discuss the results for auditing the algorithm for bias. In Section 5, we describe the framework for reducing bias in the gaze detection algorithm and discuss its results. Finally, in Section 6, we conclude the paper along with future work.

2 Related Work

Recent works in the gaze tracking area include work from Shrivastava et al. [16] where the authors proposed an adversarial training methodology for detecting gaze on unsupervised as well as simulated images with improved accuracy. Also, in [21], Zhang et al. proposed an evaluation of three deep appearance-based approaches for gaze estimation on a self-curated dataset using real-world images annotated with gazes. Some work such as [8, 13] have also proposed methods of gaze estimation in group images with more than one person. Various authors have summarized different ways of gaze tracking in the recent past. In [14], Pathirana et al. reviewed deep learning-based approaches to gaze tracking. Similarly, in [1], Akinyelu and Blignaut have summarized various CNN-based approaches to do the same. Furthermore, in [4], the authors have summarized various gaze-tracking methodologies for shape and appearance-based gaze detection for various use cases. Recently, Kulkarni et al. [10] introduced the concept of gaze uniformity in group images. They applied this work to Apple's live photos for selecting a gaze-aware representative image from them [9].

In the past few years, many researchers have worked toward auditing different multimedia analysis algorithms [20]. Recently, Singh et al. [17] audited the image-searching algorithm for gender bias in digital marketing platforms. Similarly, in [2], Alasadi et al. have audited bias in a face-matching algorithm. In this work, the authors have also introduced a method using adversarial networks to overcome such bias. They have also discussed a fairness-aware framework that uses multimodal fusion to counter the bias that can be found in cyberbullying detectors. Also, in [3], Boulamwini and Gebru assessed bias in commercial face recognition systems on two social aspects, namely skin tone, and gender. Furthermore, Kulkarni et al. [11] audited a popular OpenCV algorithm implementation for pupil detection and found it biased toward a particular gender. Overall, despite many works related to the auditing of multimedia analysis tasks, to the best of our knowledge, this is the first work to audit a gaze detection algorithm.

3 Background

In this section, first, for the sake of clarity, the assumptions made in this work are stated in Section 3.1. Then, in Section 3.2, the gaze detection algorithm audited for gender bias in this paper, is described.

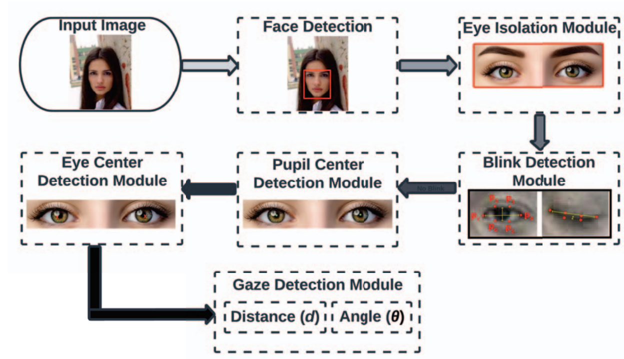

Figure 1: A detailed workflow of the gaze detection method

3.1 Assumptions

In this work, we make the following assumptions:

A1 Each face is free of any disrupting accessories like sunglasses or hats.

A2 Both eyes are visible for each person in the image.

A3 The gaze directions are with respect to the viewer and hence, opposite to the person's point of view.

A4 The gender of the subject is known beforehand.

Note that, in this work, the audit is limited to only binary genders (i.e., male and female) as obtained from the CelebA dataset.

3.2 Gaze Detection Method

Figure 1 depicts the workflow of the gaze detection method under audit. The method starts by applying the face detection algorithm [6] on a given input image. We then proceed to detect the landmarks on the face using a landmark detection algorithm [7] that detects multiple landmarks for the detected face in the image. These landmarks are then used to separate out the eye region from the entire face so as to focus on the gaze of the person. This eye frame is extracted and passed onto the blink detection module. The blink detection algorithm proposed by Soukupova and Cech in [18] uses certain landmarks detected and calculates the Eye-Aspect Ratio for each eye. If the ratio falls below a set threshold, the person is assumed to have blinked, meaning that the eye was closed or was too small to gather any more information about their gaze. If a blink is detected, we assign (0,0) to the gaze value for that person and move on. If a blink is not detected, we move ahead to detect the pupil centers. Further, eye centers are detected for the eye region. The eye center detection algorithm [6] provides the landmark position of the center of the eye whereas the pupil detection algorithm [7] provides the coordinates of the center of the pupil. These coordinates are then passed on to the gaze detection module to determine the gaze of the person.

We calculate gaze as a measure of the distance d and the angle θ between the eye center and pupil center. The distance d was used to classify the gaze to be center or non-center whereas the angle θ was used to determine the direction of gaze from one of the following eight directions: Top (T), Right (R), Left (L), Bottom (B), Top Right (TR), Top Left (TL), Bottom Right (BR), and Bottom Left (BL), as depicted in Figure 2.

We use the following steps to find the gaze for each person in the image:

Input: Image with one face, keeping our assumptions mentioned in Section 3.1 in mind.

Step 1. Apply the face detection algorithm on the input image to detect the face in it.

Step 2. Apply the landmark detection algorithm and isolate the eye region.

Step 3. Apply blink detection algorithm on the eye region to detect if there is a blink.

If yes, then set the pupil center coordinates (X_p, Y_p) of that face to $X_p = 0$ and $Y_p = 0$. Algorithm ends.

Else, move to Step 4.

Step 4. Pupil detection algorithm is applied to the eye region to find the pupil center coordinates (X_p, Y_p) for the eye frame.

Step 5. Get eye center coordinates (X_e, Y_e) for the eye frame using the eye detection algorithm.

Step 6. Calculate the distance between the eye center and pupil center as:

$$d = \sqrt{(X_e - X_p)^2 + (Y_e - Y_p)^2} \tag{1}$$

If the distance is less than initial $d_{thresh} = 1\%$ of eye width then gaze = "center". Algorithm ends.

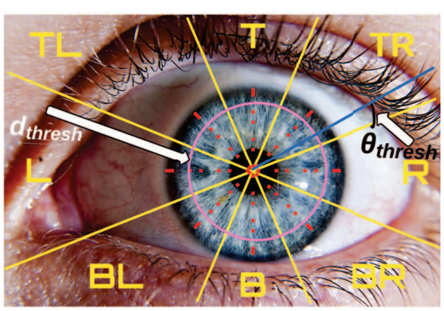

Figure 2: Distance and angle thresholding

Else, calculate the angle (in degrees) between the eye center and pupil center as:

$$\theta = \text{atan2}(X_e - X_p, Y_e - Y_p) \tag{2}$$

And map each angle θ to one of the eight directions based on the 45-degree sectors as shown in Figure 2. For instance, for Right (R), θ is between -22.5 to 22.5 degrees.

Output: The distance d and angle θ between the eye center and pupil center along with a letter direction based on the angle θ.

4 Auditing Gaze Detection

In this section, we describe the dataset and then provide the results for bias auditing.

4.1 Dataset

The gaze detection method described in Section 3 was tested on the CelebA dataset [12]. The dataset consists of 202599 aligned and cropped face images, annotated with 5 landmarks and over 40 attributes per image. We selected a subset of this dataset for testing with a total of 67234 images, out of which 28020 were male and 39214 were female. This selection was based on the successful outcome of the face detection algorithm so as to maintain its perfect accuracy. The ground truth for the dataset consists of each person's gender, pupil center coordinates, eye center coordinates, and a letter gaze direction, annotated manually.

4.2 Auditing Approach

Our primary goal in the paper was to audit the gaze detection algorithm for bias and then reduce the said bias (if found). Following extant literature, bias is defined here as the difference in the performance of the algorithm for different demographic groups [15]. In particular, we follow the literature to use Δaccuracy, i.e., the difference in accuracy for two groups, as our primary metric [2, 15]. The accuracy scores were calculated by comparing the results of ground truth values to the detected values.

Following past literature, we audit for bias by comparing the mean of the accuracy score using a Z-test for the two demographic groups (here, male and female) [2, 11]. Formally, we undertake the following hypothesis:

Null Hypothesis H_0: The bias between the male and female faces is statistically insignificant.

Alternate Hypothesis H_1: The bias between the male and female faces is statistically significant.

For the Z-test, we performed bootstrapping, where 50000 random samples from the dataset were selected to calculate the accuracy 1000 times.

Table 1: Results of auditing gaze detection algorithm

	Male	Female	Total
Number of faces	28020	39214	67234
Correct detection	25781	34470	60251
Incorrect detection	2239	4744	6983
Accuracy (α) %	92.01	87.90	89.61

Table 2: Results of auditing pupil detection algorithm

	Male	Female	Total
Number of faces	28020	39214	67234
Correct detection	25691	35731	61422
Incorrect detection	2329	3483	5812
Accuracy (α) %	91.69	91.12	91.36

4.3 Audit Results

While auditing gender bias for the gaze detection algorithm (described in Section 3) on the CelebA dataset, 4.1% bias (i.e., higher accuracy) was found to favor the male group. The accuracy for images with male faces was 92.01% and that for female faces was 87.90%. The results have been summarized in Table 1. For the statistical validation, a Z-value of -4.076 was obtained with a standard error of 0.0116, which in turn generated a p-value less than the significance level set at 0.05. Hence, the null hypothesis H_0 was rejected, and the alternative hypothesis H_1 was accepted. This indicates that the bias of 4.1% toward male faces was statistically significant.

Looking back at the steps described in Section 3.2, we note that the gaze detection algorithm consists of multiple steps which work with different component algorithms. The outcomes of Steps 1-3 effectively curtail the dataset used for computing accuracy and fairness, hence, the observed bias results cannot be attributed to them. We note the use of a pupil detection algorithm in Step 4, eye center detection in Step 5, and the setting of multiple thresholds for classification in Step 6. Hence, we proceed to audit these components for bias.

4.3.1 Bias Auditing for Pupil Detection Algorithm

We audited the pupil detection implementation using the Dlib-ml [7] for fairness over the CelebA dataset. The algorithm showed a bias of 0.57% toward male faces, with accuracy for male and female faces at 91.69% and 91.12%, respectively (see Table 2). For statistical validation, a Z-value of 0.9157 was obtained, which in turn generated a p-value that was greater than the significance level set at 0.05. Hence, we fail to reject the null hypothesis H_0 and the alternative hypothesis H_1 was rejected. This suggested that the bias toward male faces in the pupil detection algorithm was not statistically significant and unlikely to be the primary reason for the statistically significant eye gaze bias discussed above.

4.3.2 Bias Auditing for Eye Detection algorithm

Along with the pupil detection algorithm, we also audited the Retinaface algorithm's [6] eye detection implementation. When tested on the CelebA dataset, the algorithm yielded 1.05% better performance for male faces with an accuracy of 90.74% as compared to female faces with an accuracy of 89.69%. The results are summarized in Table 3. For statistical validation, a Z-value of 1.637 was obtained, which in turn, generated a p-value that was greater than the significance level set at 0.05. Hence, the null hypothesis H_0 was accepted, rejecting the alternative hypothesis H_1. This suggested that the bias of 1.05% bias toward male faces in the eye detection algorithm was also not statistically significant.

This suggests that the computation and thresholding in Step 6 could indeed be the important contributors to the observed significant bias levels.

5 Multi-stage Bias Reduction Framework

Based on the analysis in the previous section, we now focus our attention on analyzing and reducing the bias in Step 6 of the gaze detection algorithm. We consider these to be a two-stage process, where, in the first stage, we calculate the distance d between the eye center and the pupil center and distinguish the center gaze from the other eight directions. Note that, in this stage, we used precision over accuracy as a metric to determine center versus non-center classifications. This is because in this stage we focus only on the images that should have been assigned a "center" label and not passed onto the next stage. Any sample that is not as-

Table 3: Results of auditing eye detection algorithm

	Male	Female	Total
Number of faces	28020	39214	67234
Correct detection	25425	35171	60596
Incorrect detection	2595	4043	6638
Accuracy (α) %	90.74	89.69	90.13

signed a center direction in stage one, moves on to stage two. In stage two, we first calculate the angle θ between the eye center and pupil center. Then we use a mapping sub-stage where each angle is assigned a letter gaze direction which can be from one of the eight directions mentioned previously. Thus, there are two tunable parameters, distance (d) and angle (θ), which, if chosen optimally, can provide us with maximum accuracy and fairness. Distance thresholding (d_{thresh}) was used to distinguish the gaze from center versus non-center, and angle thresholding (θ_{thresh}) was used as an offset to move the sectors marking the eight non-center directions. As can be seen in Figure 2, initially the eye region is divided into 8 equal quadrants of 45 degrees each, which can be offset by (θ_{thresh}) as needed. We begin by calculating the loss function at each stage and at each thresholding level with the objective to find the ArgMin of the cumulative loss for every combination of distance and angle thresholding.

5.1 Loss Function to Minimize Bias

We formulate the composite Loss function for the gaze detection algorithm as follows:

- Let $\mathcal{T}$ be a gaze detection task that is composed of two sub-tasks, $\mathcal{T} = \{T_1, T_2\}$, where T_1 is the stage 1 task of calculating distance d and classifying the gazes with center versus non-center directions and T_2 is the stage 2 task of calculating the angle θ and mapping it to one of the eight-letter directions. Any image that is assigned the "center" label in T_1 is excluded from further processing, i.e., T_2.

- Also, let L_{T_i}, be the Loss function of the corresponding i^{th} sub-task, which is defined as follows:

$$L_{T_i} = L^{\alpha}_{T_i} + L^{\beta}_{T_i} \tag{3}$$

  Where, $L^{\alpha}_{T_i}$ denotes the loss in accuracy for stage T_i and $L^{\beta}_{T_i}$ denotes the bias at stage T_i.

- Loss in accuracy L^{α} is calculated as follows:

$$L^{\alpha} = 100 - \alpha \tag{4}$$

  Where $\alpha \in [0, 100]$ represents the accuracy of the task.

- Bias L^{β}, is defined as the difference in accuracy between privileged and unprivileged groups:

$$L^{\beta} = \alpha_{priv} - \alpha_{unpriv} \tag{5}$$

  where, $\alpha_{prev} \in [0, 100]$ and $\alpha_{unprev} \in [0, 100]$ denote the accuracy for privileged (male) and unprivileged (female) groups respectively, and $-100 \leq L^{\beta} \leq 100$ where, $L^{\beta} = -100$ and 100 implies a 100% bias toward unprivileged and privileged groups respectively, while $L^{\beta} = 0$ denotes no bias.

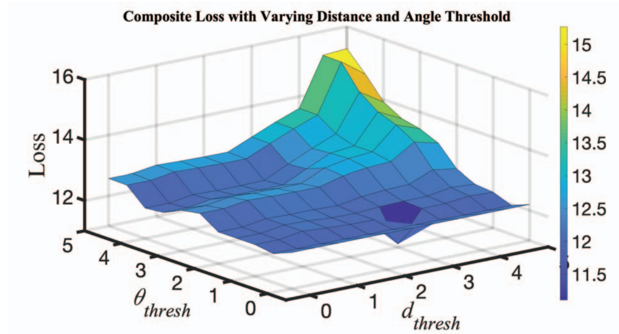

Figure 3: Composite loss with varying distance and angle thresholding

- The objective is to determine the function $L_{\mathcal{T}}$ to calculate the composite loss for the gaze detection algorithm, represented as:

$$\underset{d,\theta}{\text{argmin}}\, L_{\mathcal{T}} = \text{PR}(L_{T_1}) + (1 \text{ - PR})(L_{T_2}) \tag{6}$$

  Where, PR is the positive rate, defined as the fraction of positive detections over all the detections. This is because once an image is assigned a positive label, it is excluded from processing and label assignment from any subsequent stages.

5.2 Results of Bias Reduction

The thresholds were incremented empirically in a step of 0.5 units from 0 to 10, where the distance thresholding was incremented with respect to the eye width in order to incorporate the varying face sizes in the dataset.

Figure 3 represents the loss function values for varying distance and angle thresholding combinations (shown from 0 to 5, for aiding visual presentation). As can be seen from the figure, the minimum loss incurs when the distance threshold (d_{thresh}) is set at 3 units, whereas the angle threshold (θ_{thresh}) is set to 1 unit.

We then computed the accuracy and fairness of the gaze detection method based on the new configuration parameters identified above. The results are presented in Table 4. As can be seen from the table, it was possible to achieve elevations in accuracy and fairness for the groups in the dataset. The algorithm with new configurations yielded an overall accuracy of 93.59%, a 3.98% improvement over the earlier result. The algorithm also achieved a 1.36% bias reduction from 4.1% to 2.74% toward the male group.

Next, we employed a Z-test to compare the mean level of bias observed before and after the abovementioned bias reduction process. The test yielded a Z-value of 34.63 which in turn returned a p-value $< .05$. Thus we reject the null hypothesis and accept the alternate hypothesis implying that the reduction in bias is, in fact, statistically significant.

Table 4: Results of gaze detection after bias reduction

	Male	**Female**	**Total**
Number of faces	28020	39214	67234
Correct detection	26672	36253	62925
Incorrect detection	1348	2961	4309
Accuracy (α) %	95.19	92.45	93.59

6 Conclusion

This paper audits the gaze detection algorithm for gender bias. As eye and pupil detection are the base algorithms behind gaze detection, we also audit them for gender bias. By testing the framework over the CelebA dataset, it was observed that the gaze detection algorithm presents a statistically significant bias toward the male group, whereas the eye and pupil detection algorithms do not present a statistically significant bias. We then propose a multi-stage framework for reducing bias at multiple stages by computing the optimal parameters that yield minimum cumulative loss and in turn, yield promising results for both fairness and accuracy.

References

[1] A. Akinyelu and P. Blignaut. Convolutional neural network-based methods for eye gaze estimation: A survey. *IEEE Access*, 8:142581–142605, 2020.

[2] J. Alasadi, A. Al Hilli, and V. K. Singh. Toward fairness in face matching algorithms. In *the 1st International Workshop on Fairness, Accountability, and Transparency in MultiMedia*, pages 19–25, Nice, France, 2019.

[3] J. Buolamwini and T. Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *ACM Conference on Fairness, Accountability and Transparency (FAccT)*, pages 77–91, New York, USA, 2018.

[4] D. Cazzato, M. Leo, C. Distante, and H. Voos. When I look into your eyes: A survey on computer vision contributions for human gaze estimation and tracking. *Sensors*, 20(13):3739, 2020.

[5] H.-R. Chou, J.-H. Lee, Y.-M. Chan, and C.-S. Chen. Data-specific adaptive threshold for face recognition and authentication. In *IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 153–156, San Jose, CA, USA, 2019.

[6] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5203–5212, Virtual, 2020.

[7] D. King. Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, 10:1755–1758, 2009.

[8] Y. Kodama, Y. Kawanishi, T. Hirayama, D. Deguchi, I. Ide, H. Murase, H. Nagano, and K. Kashino. Localizing the gaze target of a crowd of people. In *Asian Conference on Computer Vision (ACCV)*, pages 15–30, Perth, Australia, 2018.

[9] O. Kulkarni, S. Arora, and P. K. Atrey. GARGI: selecting gaze-aware representative group image from a live photo. In *IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 115–120, Singapore, 2022.

[10] O. Kulkarni, V. Patil, S. Parikh, S. Arora, and P. K. Atrey. Can you all look here? Towards determining gaze uniformity in group images. In *IEEE International Symposium on Multimedia (ISM)*, pages 100–103, Naples, Italy, 2020.

[11] O. Kulkarni, V. Patil, V. K. Singh, and P. K. Atrey. Accuracy and fairness in pupil detection algorithm. In *IEEE Seventh International Conference on Multimedia Big Data (BigMM)*, pages 17–24, Taichung, Taiwan, 2021.

[12] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015.

[13] J. O'Reilly, A. Khan, Z. Li, J. Cai, X. Hu, M. Chen, and Y. Tong. A novel remote eye gaze tracking system using line illumination sources. In *IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 449–454, San Jose, CA, USA, 2019.

[14] P. Pathirana, S. Senarath, D. Meedeniya, and S. Jayarathna. Eye gaze estimation: A survey on deep learning-based approaches. *Expert Systems with Applications*, 199:116894, 2022.

[15] D. Pessach and E. Shmueli. Algorithmic fairness. *arXiv preprint arXiv:2001.09784*, 2020.

[16] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb. Learning from simulated and unsupervised images through adversarial training. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2107–2116, Honolulu, HI, USA, 2017.

[17] V. K. Singh, M. Chayko, R. Inamdar, and D. Floegel. Female librarians and male computer programmers? Gender bias in occupational images on digital media platforms. *Journal of the Association for Information Science and Technology*, 71(11):1281–1294, 2020.

[18] T. Soukupova and J. Cech. Real-time eye blink detection using facial landmarks. In *The 21st Computer Vision Winter Workshop (CVWW)*, pages 1–8, Rimske Toplice, Slovenia, 2016.

[19] Y. Tao, E. Coltey, T. Wang, M. Alonso, M.-L. Shyu, S.-C. Chen, H. Alhaffar, A. Elias, B. Bogosian, and S. Vassigh. Confidence estimation using machine learning in immersive learning environments. In *IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 247–252, Virtual, 2020.

[20] B. H. Zhang, B. Lemoine, and M. Mitchell. Mitigating unwanted biases with adversarial learning. In *the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, pages 335–340, New York, NY, USA, 2018.

[21] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling. Mpiigaze: Real-world dataset and deep appearance-based gaze estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1):162–175, 2017.