FairQuantize: Achieving Fairness Through Weight Quantization for Dermatological Disease Diagnosis

Yuanbo Guo[1], Zhenge Jia[1], Jingtong Hu[2], and Yiyu Shi[1]

[1] University of Notre Dame, Notre Dame, IN 46556, USA
{yguo6,zjia2,yshi4}@nd.edu
[2] University of Pittsburgh, Pittsburgh, PA 15260, USA
jthu@pitt.edu

Abstract. Recent studies have demonstrated that deep learning (DL) models for medical image classification may exhibit biases toward certain demographic attributes such as race, gender, and age. Existing bias mitigation strategies often require sensitive attributes for inference, which may not always be available, or achieve moderate fairness enhancement at the cost of significant accuracy decline. To overcome these obstacles, we propose FairQuantize, a novel approach that ensures fairness by quantizing model weights. We reveal that quantization can be used not as a tool for model compression but as a means to improve model fairness. It is based on the observation that different weights in a model impact performance on various demographic groups differently. FairQuantize selectively quantizes certain weights to enhance fairness while only marginally impacting accuracy. In addition, resulting quantized models can work without sensitive attributes as input. Experimental results on two skin disease datasets demonstrate that FairQuantize can significantly enhance fairness among sensitive attributes while minimizing the impact on overall performance.

Keywords: Fairness · Quantization · Deep Learning · Dermatological Disease Diagnosis.

1 Introduction

In conventional computer-aided diagnosis (CAD) system design, essential features and detection criteria are first derived from clinical trials and then transformed into a program deployed on medical devices. Considerable expertise is required to optimize the extracted feature set, detection criteria, and programmable parameters. Deep learning (DL) provides an alternative solution to reduce the demand for domain expertise in the method design. DL-based CAD approaches are generally designed to achieve higher detection accuracy. To maximize accuracy performance, the DL model would leverage information that is present in some data but absent in other data during training. However, such trained DL model can result in discrimination towards certain demographics such

as skin tone or gender. For example, studies [6,20] show that dermatological disease classification models trained on two publicly available dermatology datasets (ISIC 2019 Challenge and Fitzpatrick-17k) have identified significant bias across different skin tones. When biased models are implemented in real-world systems, they can have negative impacts on both individuals and society. For instance, these models may misdiagnose individuals from certain demographic groups, resulting in greater healthcare disparities.

Many bias mitigation methods have been studied and proposed for fairness. One of the most widely used bias mitigation methods is adversarial training [11,18,4,1,21]. But directly excluding features linked to sensitive attributes for both privileged and unprivileged groups might undermine classification accuracy by omitting crucial information, consequently lowering prediction precision [19]. Fairness through explanation is another technique for bias mitigation [14,9,16]. This approach requires fine-grained feature-level annotation as the domain knowledge to train the model to only focus on bias-unrelated features in the original input. However, such suppression of information about sensitive attributes increases the potential to miss useful features, greatly degrading the prediction performance. These state-of-the-art (SOTA) methods usually sacrifice considerable accuracy on both groups to improve fairness. The most recent work FairPrune [20] and its succeeding work ME-FairPrune [2], are proposed to achieve better fairness via model pruning. Model pruning-based methods for fairness involve toggling individual connections on or off, which can remove vital diagnostic information and impact model performance. A more balanced approach would be moderating, instead of eliminating, certain information to maintain both fairness and performance, allowing for partial data flow.

To address these challenges, we develop FairQuantize, which achieves fairness by using quantization, diverging from its traditional use for reducing size and speeding up inference. We identify that specific weights within trained models disproportionately benefit certain demographic groups, causing imbalances and biased results. FairQuantize addresses this by adjusting the computation precision of these pivotal weights through quantization, thereby balancing accuracy and fairness between different demographic groups. By employing a Taylor series approximation [12], we pinpoint and quantize weights that significantly impact demographic disparities. This approach allows for customizable balances between fairness and accuracy, meeting diverse user needs. Our evaluations on two skin lesion datasets demonstrate that FairQuantize outperforms existing SOTA fairness methods by achieving greater fairness with minimal accuracy loss. To make our research more reproducible, the code for FairQuantize is publicly available at https://github.com/guoyb17/FairQuantize.

2 Method

2.1 Problem Definition

Consider a dataset $D = \{x_i, y_{0i}, c_i\}$, $i \in \{1, ..., N\}$, where x_i is the input image, y_{0i} is the class label, c_i is the sensitive attribute (skin tone, gender, etc.). $y =$

$F(\Theta, x)$ is a pre-trained classification model with weights Θ that maps the input x_i to the prediction $y_i = F(\Theta, x_i)$. Our target is to reduce the bias of the model $F(\Theta, x)$ between groups with different values of a sensitive attribute c by only modifying some of the weights. In this paper, only binary sensitive attributes (i.e., $c_i \in \{0, 1\}$) are considered, based on which all data can be divided into two groups, named as unprivileged and privileged groups. The unprivileged group represents the group with lower performance, while the privileged group represents the one with higher performance.

2.2 Weight-wise Fairness Score

Weights hold different importance for a model. Typical neural networks have millions or even billions of weights, so it is essential to decide which weights to compress first. One direct impact for certain weights to be quantized is that the calculation results of corresponding neurons are changed due to the change of these weights. If we change Θ for $\Delta\Theta$ to represent the difference between weights after and before quantization, then with the approximation of Taylor series [12], we can modify the calculation of this neuron to

$$\begin{aligned} F(\Theta) &= F(\Theta_0 + \Delta\Theta) \\ &= F(\Theta_0) + \Delta F(\Theta_0)\Delta\Theta + \frac{1}{2}\Delta\Theta^{\mathrm{T}} H(\Theta_0)\Delta\Theta + O(\Delta\Theta^3), \end{aligned} \tag{1}$$

where $H(\Theta_0)$ is the Hessian matrix of $F(\Theta)$ at Θ_0, and the third term $O(\Delta\Theta^3)$ can be neglected. For a pre-trained model, it is assumed that it has converged and that the objective function is at its local minimum, thus we can ignore the first term part containing $\Delta F(\Theta_0)$ which is close to zero. Then, it can be found that the critical part that decides how much the change of Θ impacts the results falls to the $\frac{1}{2}\Delta\Theta^{\mathrm{T}} H(\Theta_0)\Delta\Theta$ part. The bigger this part is, the more significant changes of Θ may influence $F(\Theta)$. Therefore, based on the above formulas, we define the importance of a weight θ_i of all weights Θ as $Importance(\theta_i) = \frac{1}{2} h_{ii} {\Delta\theta_i}^2$, where h_{ii} is the corresponding element from the Hessian matrix $H(\Theta)$, which is actually the second derivative with respect to θ_i.

In our work, the quantization takes an easy-to-understand one, power-of-2 [22]. By converting a weight into the format of a power of 2, say, $\theta_i = 2^n$, the weight θ_i can be quantized by selecting the nearest integer n_q to replace the original n, and it can be easily restored to $\theta_{iq} = 2^{n_q}$ if needed. Therefore, the change of the weight is $\Delta\theta_i = |2^{log_2(abs(\theta_i))} sign(\theta_i) - \theta_i|$, where $log_2(abs(\theta_i))$ is the integer that represents the power-of-2 quantized version of the weight θ_i.

Scoring is based on difference between importance. The idea of our work stems from our empirical observation that in a pre-trained model, the significance of certain weights can vary dramatically for inputs with different attributes, treating these attributes as distinct groups. Specifically, some weights might be less important for one group while being crucial for another. To pinpoint weights suitable for quantization, we introduce a fairness-aware importance scoring method. This approach identifies weights that are less significant

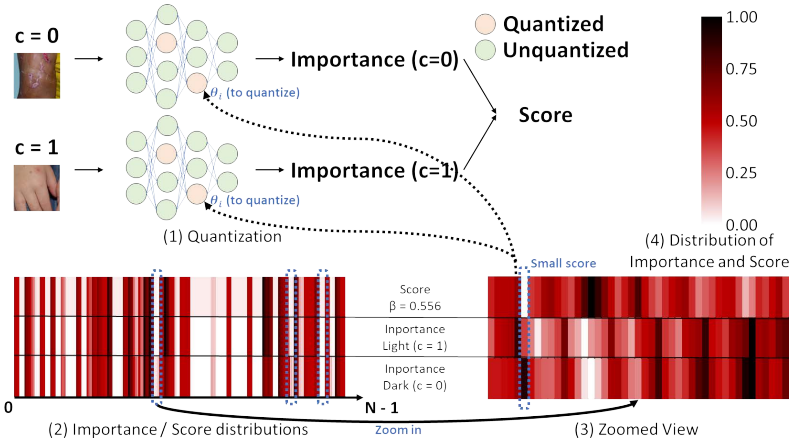


Fig. 1. The illustration of (1) quantization, (2) importance / score distributions, (3) zoomed viewing of distributions, and (4) the color map of distributions. This case is performed on a partly quantized model produced by our method on Fitzpatrick-17k. In (2) and (3), the x-axis represents the index of all weights and areas of selected weights accordingly. Blue boxes mark some typical areas of weights distributions with low scores, which are basically because of low importance of the unprivileged (dark) and high importance of the privileged (light). Score and importance values are normalized to [0, 1] range for the simplicity of illustration, as shown in (4).

for the unprivileged group but crucial for the privileged group. We quantify each weight's relevance to fairness by combining the importances for both the unprivileged and privileged groups, adjusted by a negative scalar. That is to say, for a weight θ_i, its fairness-related importance, or score, can be defined as:

$$\begin{aligned} Score(\theta_i) &= Importance^{unprivileged}(\theta_i) - \beta Importance^{privileged}(\theta_i) \\ &= \frac{1}{2}\Delta{\theta_i}^2(h^u_{ii} - \beta h^p_{ii}), \end{aligned} \tag{2}$$

where h^u_{ii} and h^p_{ii} are the second derivative calculated for the unprivileged group and the privileged group, and β is a positive-ranged hyper-parameter that balances the importance values for different groups. β represents how much FairQuantize tends to lose performance on the privileged group in exchange for performance on the unprivileged group. If β is too small (close to zero), the score will be basically irrelevant to fairness; but if β is too big (close to infinite), the score will mostly care about the performance on the unprivileged group alone, which probably will make the model unfair in the other way. Ablation study in Sec. 3 introduces more about the usage of β.

Algorithm 1 illustrates the steps of FairQuantize. Quantization ratio is defined as the number of quantized weights divided by the total number of weights. In incremental network quantization (INQ) based on [22], which is adopted by our algorithm, each iteration increases quantization ratio by a step q, as the

Algorithm 1 FairQuantize

Input: Pre-trained model M_0 and weight number N, training set T, scoring sets S^u and S^p, quantization ratio step q, re-train epoch E, hyper-parameter β.
Output: Quantized models M_i and their quantization ratios Q_i, where $i = 1, 2, ..., n$.

$n \Leftarrow 1$
while $n \times q \leq 1.0$ **do**
  $M_n \Leftarrow M_{n-1}$, $j \Leftarrow 0$
  for $\{s^u, s^p\}$ in $\{S^u(unprivileged), S^p(privileged)\}$ **do**
    Infer with $M_n(s^u)$ and $M_n(s^p)$ to get Hessian of weights H^u and H^p
    $Score_j \Leftarrow H^u - \beta \times H^p$
    Set scores of quantized weights in $Score_j$ to arbitrarily small numbers
    $j \Leftarrow j + 1$
  end for
  $Order, Sorted\_Score \Leftarrow Sort(Average(Score_1, Score_2, ..., Score_{j-1}))$
  $Q_n \Leftarrow min(1.0, n \times q)$
  $M_n \Leftarrow quantization(M_n)$ on $Q_n \times N$ weights of the lowest scores
  $M_n \Leftarrow M_n$ with re-training on T for E epochs
  $n \Leftarrow n + 1$
end while

main loop of Algorithm 1 goes. For each iteration, the model infers on two scoring sets of two groups, and generates hessian matrices with back propagation, with which scores are calculated as defined above. Fig. 1 also illustrates this workflow, especially about how scoring works and determines which weights to quantize. It is noted that scores of previously quantized weights are assigned significantly low values to ensure they remain smaller than regular score values. This is because after sorting, the algorithm selects weights with smallest scores for quantization, which should include weights that have already been quantized. This procedure ensures that quantized weights are not changed unexpectedly. Then, weights that are not quantized yet may receive certain re-training to adapt to the performance loss due to quantization.

3 Experiments and Results

Datasets and Pre-processing. The experiments are performed on two dermatology datasets for disease classification, including the Fitzpatrick-17k [5,6] and ISIC 2019 challenge [3,17] datasets. Fitzpatrick-17k dataset contains 16,577 clinical images of 114 dermatological conditions. Images are categorized into 6 types of skin tones (marked as 1 to 6), from light to dark. We group them into light (1 to 3) and dark (4 to 6) groups, and use this binary skin tone attribute as the sensitive attribute. ISIC 2019 contains 25,331 dermoscopic images of 9 diagnostic categories. It does not have skin color information, but it has the binary sex label (male or female) for each sample, so we use it as the sensitive attribute for this dataset. More details can be found in the public repository provided above.

Table 1. Results on Fitzpatrick-17k using VGG-11. For EOpp0, EOpp1, EOdd, and Diff., the lower the better. "Diff." stands for difference (absolute values of differences between light and dark metrics), while "Avg." stands for average (metrics on the entire test set). The dark skin tone group is privileged, and the light skin tone group is unprivileged. The best result of each metric is shown in bold.

Method	Skin Tone	Accuracy			Fairness		
		Precision	Recall	F1-Score	EOpp0 ↓	EOpp1 ↓	EOdd ↓
Vanilla	Light	0.482	0.495	0.473	0.0013	0.361	0.182
	Dark	0.563	0.581	0.546			
	Avg. ↑	0.523	**0.538**	0.510			
	Diff. ↓	0.081	0.086	0.073			
MFD [10]	Light	0.489	0.469	0.457	0.0011	0.334	0.166
	Dark	0.514	0.545	0.503			
	Avg. ↑	0.502	0.507	0.480			
	Diff. ↓	0.025	0.076	0.046			
FairPrune [20]	Light	0.496	0.477	0.459	**0.0008**	0.330	0.165
	Dark	0.567	0.519	0.507			
	Avg. ↑	0.531	0.498	0.483			
	Diff. ↓	0.071	0.042	0.048			
ME-FairPrune [2]	Light	0.542	0.535	0.522	0.0012	0.305	0.152
	Dark	0.564	0.529	0.523			
	Avg. ↑	**0.553**	0.532	0.522			
	Diff. ↓	0.022	0.006	0.001			
FairQuantize	Light	0.519	0.493	0.493	0.0012	**0.269**	**0.135**
	Dark	0.592	0.537	0.537			
	Avg. ↑	0.551	0.517	**0.524**			
	Diff. ↓	0.073	0.044	0.044			

Pre-training Details. We use VGG-11 [15] and ResNet18 [8] as the backbone models. More details can be found in the public repository provided above.

Baselines. Standard training without interference of any method is denoted as *Vanilla*. *FairPrune* [20] achieves fairness by pruning weights that are important for the privileged group but unimportant for the unprivileged group to reduce the performance gap. *ME-FairPrune* [2] applies multi-exit (ME) training framework to FairPrune, achieveing better fainess performance. *AdvConf* [1] and *AdvRev* [21] are two of adversarial training based de-biasing methods. *HSIC* [13] achieves fairness by masking sensitive areas in the input images. *DomainIndep* [19] leverages multiple classifiers, one classifier for each group to explicitly split group information. *MFD* [10] improves fairness via knowledge distillation.

Fairness Metrics. Equalized opportunity (EOpp) and equalized odds (EOdd) [7] metrics are used to evaluate the fairness of the methods. The EOpp0 is the True Negative Rate difference between the two groups. The EOpp1 is the True Positive Rate difference between the two groups, while the EOdd is the summation of the True Positive Rate difference and False Positive Rate difference.

Results on Fitzpatrick-17k dataset using VGG-11. Table 1 shows the results of various methods on Fitzpatrick-17k dataset using VGG-11. The results of some methods (proposed before 2020) are demonstrated in the supplemental material. As shown by Table 1, FairQuantize (with a quantization ratio of 20% and $\beta = 1.0$) achieves the best fairness performance in terms of EOpp1 and EOdd as well as the highest F1 score over SOTA methods. Compared with

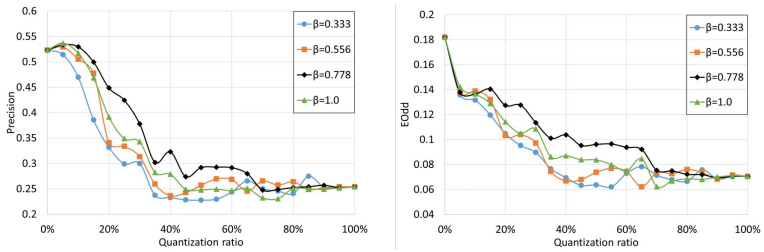
Table 2. Results on ISIC 2019 using ResNet18. The female group is privileged, and the male group is unprivileged. The best result of each metric is shown in bold.

Method	Sex	Accuracy			Fairness		
		Precision	Recall	F1-Score	EOpp0 ↓	EOpp1 ↓	EOdd ↓
Vanilla	Female	0.793	0.721	0.746	0.006	0.044	0.022
	Male	0.731	0.725	0.723			
	Avg. ↑	0.762	0.723	0.735			
	Diff. ↓	0.063	0.004	0.023			
MFD [10]	Female	0.770	0.697	0.726	0.005	0.051	0.024
	Male	0.772	0.726	0.744			
	Avg. ↑	0.771	0.712	0.735			
	Diff. ↓	0.002	0.029	0.018			
FairPrune [20]	Female	0.776	0.711	0.734	0.007	0.026	0.014
	Male	0.721	0.725	0.720			
	Avg. ↑	0.748	0.718	0.727			
	Diff. ↓	0.055	0.014	0.014			
ME-FairPrune [2]	Female	0.770	0.723	0.742	0.006	0.020	**0.010**
	Male	0.739	0.728	0.730			
	Avg. ↑	0.755	0.725	0.736			
	Diff. ↓	0.032	0.005	0.011			
FairQuantize	Female	0.845	0.835	0.834	**0.003**	**0.019**	0.012
	Male	0.863	0.849	0.856			
	Avg. ↑	**0.857**	**0.843**	**0.850**			
	Diff. ↓	0.018	0.014	0.017			

the vanilla baseline, our method improves EOpp0, EOpp1, and EOdd by 7.7% (0.0013 →0.0012), 25.5% (0.361 →0.283), and 25.8% (0.182 →0.135), while maintaining the accuracy performance. In contrast to the performances of other SOTA methods, ours improves fairness more with much less accuracy trade-off.

Results on ISIC 2019 dataset using ResNet18. Table 2 shows the results of various methods on ISIC 2019 dataset using ResNet18. The results of some methods (proposed before 2020) are demonstrated in the supplemental material. As shown by Table 2, FairQuantize (with a quantization ratio of 80% and $\beta = 0.778$) outperforms SOTA methods in terms of almost all accuracy and fairness metrics. It again indicates that FairQuantize could improve the fairness of the deep model while effectively maintaining its accuracy via quantization.

Analysis. In comparison with ISIC 2019, the vanilla baseline on the Fitzpatrick 17k dataset shows lower accuracy and poorer fairness due to its limited number of images that span a wide range of different classes. However, FairQuantize addresses these challenges effectively, outperforming SOTA methods in fairness scores. While the fairness improvement on ISIC 2019 is less pronounced, FairQuantize significantly enhances accuracy, benefiting from adequate re-training. Additionally, performance comparisons in Table 1 and Table 2 demonstrate that FairQuantize surpasses methods like FairPrune and ME-FairPrune, mainly due to its fine-grained tuning capability. Unlike the binary nature of pruning (it either sets a weight to zero or not), FairQuantize employs power-of-2 quantization, allowing more nuanced adjustments and providing greater flexibility in weight modification.



(a) Precision results for different β and quantization ratios

(b) EOdd results for different β and quantization ratios

Fig. 2. Ablation study on hyper-parameter β and quantization ratio for precision and EOdd on Fitzpatrick-17k using VGG-11.

Ablation Study. In this section, we explore how the hyper-parameters (β and quantization ratio) affect the outcomes of FairQuantize. Our method incrementally applies quantization to a pre-trained model, suggesting that fine-tuning could help the model adjust to the quantization changes. Although quantized weights remain fixed, other weights can still be re-trained. However, to avoid additional uncertainties unrelated to quantization, we omit the re-training phase in our ablation study experiments. Therefore, FairQuantize is applied to the pre-trained VGG-11 model on Fitzpatrick-17k, with different β settings, and without re-training. The quantized models at quantization ratios from 5% to 100%, with a step size of 5%, are all saved for comparison.

Fig. 2(a) and Fig. 2(b) illustrate the relationship between precision (accuracy) and EOdd (fairness) as affected by quantization levels, revealing a trade-off: as quantization intensifies, precision typically decreases while fairness improves. The effect varies with different β values; for instance, in the case of this experiment, a β of 0.778 shows the most conservative impact, minimally affecting accuracy and fairness, whereas a β of 0.333 is more aggressive. This indicates that adjustments in β and quantization ratio can fine-tune the balance between accuracy and fairness. However, the presence of performance "overlaps" suggests that optimal β settings may need to be determined individually for each task, a complexity further compounded by the process of re-training.

4 Conclusion

In this paper, we propose FairQuantize, a fairness-aware weight quantization framework to optimize model fairness and to preserve classification accuracy. The framework introduces a fairness score to observe the impact of weights on fairness of a given model. On top of that, it applies incremental quantization on weights selected by scores with proper re-training to improve fairness, which goes beyond conventional thoughts that quantization is merely a model compression method. We evaluate our framework with two datasets and two backbone models,

compared with multiple SOTA fairness-aware methods. Experimental results show that FairQuantize outperforms the existing methods in terms of maximizing fairness performance while minimizing accuracy loss. The framework has the potential to enable the development of more ethical and equitable AI systems.

Acknowledgments. This project is supported in part by NIH grant R01EB033387.

Disclosure of Interests. The authors have no competing interests in the paper.

References

1. Alvi, M., Zisserman, A., Nellåker, C.: Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops. pp. 0–0 (2018)
2. Chiu, C.H., Chung, H.W., Chen, Y.J., Shi, Y., Ho, T.Y.: Toward Fairness Through Fair Multi-Exit Framework for Dermatological Disease Diagnosis, p. 97–107. Springer Nature Switzerland (2023), http://dx.doi.org/10.1007/978-3-031-43898-1_10
3. Combalia, M., Codella, N.C., Rotemberg, V., Helba, B., Vilaplana, V., Reiter, O., Carrera, C., Barreiro, A., Halpern, A.C., Puig, S., et al.: Bcn20000: Dermoscopic lesions in the wild. arXiv preprint arXiv:1908.02288 (2019)
4. Elazar, Y., Goldberg, Y.: Adversarial removal of demographic attributes from text data. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. pp. 11–21 (2018)
5. Groh, M., Harris, C., Daneshjou, R., Badri, O., Koochek, A.: Towards transparency in dermatology image datasets with skin tone annotations by experts, crowds, and an algorithm. arXiv preprint arXiv:2207.02942 (2022)
6. Groh, M., Harris, C., Soenksen, L., Lau, F., Han, R., Kim, A., Koochek, A., Badri, O.: Evaluating deep neural networks trained on clinical images in dermatology with the fitzpatrick 17k dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1820–1828 (2021)
7. Hardt, M., Price, E., Srebro, N.: Equality of opportunity in supervised learning. Advances in neural information processing systems **29** (2016)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
9. Hendricks, L.A., Burns, K., Saenko, K., Darrell, T., Rohrbach, A.: Women also snowboard: Overcoming bias in captioning models. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 771–787 (2018)
10. Jung, S., Lee, D., Park, T., Moon, T.: Fair feature distillation for visual recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12115–12124 (2021)
11. Kim, B., Kim, H., Kim, K., Kim, S., Kim, J.: Learning not to learn: Training deep neural networks with biased data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9012–9020 (2019)
12. LeCun, Y., Denker, J., Solla, S.: Optimal brain damage. Advances in neural information processing systems **2** (1989)

13. Quadrianto, N., Sharmanska, V., Thomas, O.: Discovering fair representations in the data domain. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8227–8236 (2019)
14. Rieger, L., Singh, C., Murdoch, W., Yu, B.: Interpretations are useful: penalizing explanations to align neural networks with prior knowledge. In: International conference on machine learning. pp. 8116–8126. PMLR (2020)
15. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2015)
16. Singh, K.K., Mahajan, D., Grauman, K., Lee, Y.J., Feiszli, M., Ghadiyaram, D.: Don't judge an object by its context: learning to overcome contextual bias. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11070–11078 (2020)
17. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data **5**(1), 1–9 (2018)
18. Wang, T., Zhao, J., Yatskar, M., Chang, K.W., Ordonez, V.: Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5310–5319 (2019)
19. Wang, Z., Qinami, K., Karakozis, I.C., Genova, K., Nair, P., Hata, K., Russakovsky, O.: Towards fairness in visual recognition: Effective strategies for bias mitigation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8919–8928 (2020)
20. Wu, Y., Zeng, D., Xu, X., Shi, Y., Hu, J.: Fairprune: Achieving fairness through pruning for dermatological disease diagnosis. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part I. pp. 743–753. Springer (2022)
21. Zhang, B.H., Lemoine, B., Mitchell, M.: Mitigating unwanted biases with adversarial learning. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. pp. 335–340 (2018)
22. Zhou, A., Yao, A., Guo, Y., Xu, L., Chen, Y.: Incremental network quantization: Towards lossless cnns with low-precision weights. arXiv preprint arXiv:1702.03044 (2017)