Trade-off Between Accuracy, Robustness, and Fairness of Deep Classifiers

Philipp Benz*
pbenz@kaist.ac.kr

Chaoning Zhang*
chaoningzhang1990@gmail.com

Soomin Ham
smham@kaist.ac.kr

Adil Karjauv
mikolez@gmail.com

Gyusang Cho
gyusang.cho@kaist.ac.kr

In So Kweon
iskweon77@kaist.ac.kr

Korea Advanced Institute of Science and Technology (KAIST)

Abstract

Deep Neural Networks (DNNs) have achieved great success, however, their vulnerability to adversarial examples remains an open issue. Among numerous attempts to increase the robustness of deep classifiers, mainly adversarial training has stood the test of time as a useful defense technique. It has been shown that the increased model robustness comes at the cost of decreased accuracy. At the same time, deep classifiers trained on balanced datasets exhibit a class-wise imbalance, which is even more severe for adversarially trained models. This work aims to highlight that the fairness of classifiers should not be neglected when evaluating DNNs. To this end, we propose a class-wise loss re-weighting to obtain more fair standard and robust classifiers. The final results suggest, that fairness as well comes at the cost of accuracy and robustness, suggesting that there exists a triangular trade-off between accuracy, robustness, and fairness.

1. Introduction

The vulnerability of Deep Neural Networks (DNNs) [19] to adversarial examples [34, 11] hinders their application in security-sensitive applications. Various defense techniques have been proposed, to increase the robustness of convolutional neural networks (CNNs) [30, 12, 42, 24, 22]. While most of these defense techniques are ineffective against more advanced adversaries or due to inherent design flaws [7, 1, 37], adversarial training [11, 22] remains as one of the reliable defenses against adversarial examples. While adversarial training increases the robustness of CNNs, it comes at the cost of decreasing the overall accuracy [38]. While this trade-off is well known in the machine learning community, in this work we identify fairness as an additional trade-off to accuracy and robustness. Commonly, only a single value is used to describe the standard accuracy or robust accuracy. A single accuracy value does not reflect insights about the fairness properties of a classifier. For example, a critical class-wise accuracy could be very low, while the class-wise accuracy for another relatively less important class might be higher. Such an imbalanced class-wise distribution might have additional security implications and the average accuracy might give a false sense of fairness.

We first revisit the phenomenon of class-wise imbalance of CNNs. Specifically, for models trained on a balanced dataset, a class-wise imbalance can already be identified for standard training. This class-wise imbalance is even more pronounced for adversarially trained models. This phenomenon is somewhat related to the topic of long-tailed recognition [16], where class-wise imbalances naturally occur due to an imbalance in the dataset distribution. Hence, similar techniques from the area of long-tail recognition can be applied to improve the naturally occurring class-wise imbalance. To this end, we propose a class-wise loss re-weighting scheme. Specifically, we re-weight the class-wise losses by a class-specific weight, which is adapted according to the underlying class-wise accuracy distribution. In detail, when a class-wise accuracy is above the average, its class-wise weight should be decreased and vice versa. Our experimental results on CIFAR10 show that our fair training scheme results in more fair classifiers. However, the increased fairness comes at the cost of decreased accuracy. Additionally, our fair training methodology can further be incorporated into the adversarial training procedure, where it results in significantly more fair classifiers, compared to the initially more "unfair" robust classifier accuracy distribution. However, the increased fairness does also lead to a decrease in robustness. Due to these observations next to the proposed fair training strategy, another of our contributions is the identification of the triangular trade-off between accuracy, robustness, and fairness for deep learning classifiers.

*Equal contribution

2. Related Work

Adversarial machine learning. Deep Neural Networks are vulnerable to adversarial examples [34, 11]. This phenomenon has attracted numerous image-dependant attack methods [34, 11, 26, 22], universal attack methods [25, 27, 31, 21, 45, 44, 2], and defense techniques [30, 12, 42, 24, 4]. However, most defense methods were later shown to provide a false sense of security [7, 1, 37], mainly due to gradient obfuscation. Among the defense methods, adversarial training [11, 22] has stood the test of time and is widely adopted. Numerous works have been proposed to improve the performance of adversarial training [39, 46, 32, 29, 40, 41, 20]. However, it has later been suggested that the robustness may be at odds with accuracy [38]. Tsipras *et al.* further argue that robust classifiers learn fundamentally different feature representations than standard classifiers, which has been further investigated by [14]. Inspired by the accuracy robustness conflict, the adversarial training method TRADES had been introduced to trade-off adversarial robustness against accuracy [47]. Friendly adversarial training [48] tackles the robustness-accuracy issue by searching for weaker adversaries during adversarial training.

Long-tailed recognition. Most classification datasets come with a balanced label distribution [17, 18]. However, in real-world classification problems often long-tailed label distributions are exhibited, where a greater amount of labels is only associated with a few classes [16]. Neglecting such imbalance can have detrimental effects on the model performance [15, 5]. The long-tail problem has been extensively studied in the literature and can traditionally be divided into two streams, re-sampling [8, 15, 13, 33, 5], and re-weighting [49, 9, 6, 35]. Due to our observed class imbalance, this topic is closely related to long-tailed recognition, and concepts from long-tailed recognition can be borrowed.

Fairness in adversarial machine learning With the wide adoption of Deep Learning models, their fairness properties have also been studied [23, 10]. In this work, we mainly focus on the fairness properties of standard and robust deep classifiers trained on balanced datasets in the context of their class-wise accuracies. Robustness bias of DNNs has been discussed in [28], which refers to certain subgroups of classes that exhibit a decreased adversarial robustness and might therefore be at a disadvantage. The phenomenon of class-wise accuracy has also been discovered by other works concurrently to ours [43, 36, 3]. Similar to our approach the fair robust learning framework [43] also attempts to train robust models with a balanced accuracy and robustness performance. Inspired by the class-wise imbalance phenomenon [36] proposed the temperature-PGD attack which exploits the robustness disparity among classes.

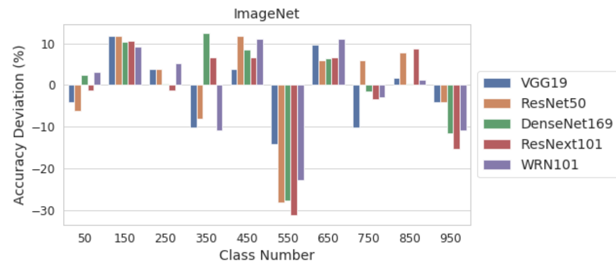

Figure 1: Class-wise accuracies for different model architectures trained on the ImageNet dataset

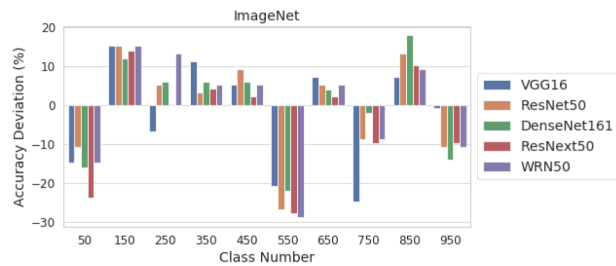

Figure 2: Class-wise accuracies for different model architectures adversarially trained on the ImageNet dataset

3. Class-imbalance Phenomenon.

In this section we revisit the class-imbalance phenomenon for different models trained on the ImageNet dataset. Figure 1 and Figure 2 show the class-wise accuracy deviations from the mean accuracy for standard trained and adversarially trained model architectures. Two major observations can be made. First, the class-wise accuracies for standard and adversarially trained models are very similar among different model architectures. Second, overall the class-wise accuracy deviations from the mean are more severe for adversarially trained models than for standard trained models. This phenomenon can be observed more pronounced for the CIFAR10 model architectures shown in Figure 3 and Figure 4.

4. Methodology

Our proposed solution to mitigate the class-imbalance occurring in standard and robust DNNs is based on a class-specific re-weighting strategy. Given a C-classification dataset $\mathcal{X}$ composed of samples $x \in \mathbb{R}^{w \times h \times 3}$ and their corresponding ground truth $y \in [1, C]$, a classifier F_θ, parameterized through the weights θ (from here on omitted) is commonly trained via mini-batch stochastic gradient descent (SGD). After training a classifier, one common indicator for the performance is the average accuracy $\zeta \in \mathbb{R}$. In this work, we additionally focus on the class-wise ac-

Algorithm 1: Fair Training Algorithm

Input: Training Dataset $\mathcal{X}$, Classifier F_θ, Loss function $\mathcal{L}$, Optimizer `Optim`, Hyperparameters α, β

Output: Fair classifier F_{fair}

1 $\gamma \leftarrow \mathbf{1}$ ▷ `Initialization`
2 $\tau \leftarrow \mathbf{0}$ ▷ `Initialization`
3 $\mathcal{X}_{\text{train}}, \mathcal{X}_{\text{val}} \leftarrow \mathcal{X}$ ▷ `Split the data`
4 **for** $i, \ldots, I$ **do**
5 $\quad x_{\text{train}}, y_{\text{train}} \sim \mathcal{X}_{\text{train}}$
6 $\quad x_{\text{val}}, y_{\text{val}} \sim \mathcal{X}_{\text{val}}$
7 $\quad$ **for** $c, \ldots, C$ **do**
8 $\quad\quad g^c \leftarrow \nabla \gamma^c \mathcal{L}(F, x^c_{\text{train}}, y^c_{\text{train}})$
9 $\quad\quad \tau^c \leftarrow \tau^c + \beta \cdot \text{CWAD}(F, x^c_{\text{val}}, y^c_{\text{val}})$
10 $\quad$ **end**
11 $\quad \theta \leftarrow$ `Optim`(g) ▷ `Update weights`
12 $\quad \gamma \leftarrow \gamma \cdot (1 + \alpha \cdot \text{sign}(\tau))$ ▷ `Update`
13 **end**

curacy, which we indicate by $\phi \in \mathbb{R}^C$. From these two metrics, the class-wise accuracy deviation can be calculated as $\text{CWAD}(F, x, y) = \phi - \zeta$. A common choice for the loss function $\mathcal{L}$ is the cross-entropy loss. The objective of the proposed fair training strategy is a class-wise weighting of the losses based on the class-wise accuracies. Simply speaking, when the accuracy deviation for a class is lower than one, the weight for the loss for samples should be increased, and vice versa.

The fair training algorithm is described in Algorithm 1. We introduce a class-wise weighting vector $\gamma \in \mathbb{R}^C$, where each entry corresponds to the respective class-wise weight. The values of γ will be updated based on the moving average of the class-wise accuracy deviations τ. To obtain the moving average of the class-wise accuracy deviations we use a small portion of the original training dataset as a validation dataset. The loss values of the samples from a certain class are multiplied with their respective weight γ^c as shown in line 8 of the algorithm. By calculating the class-wise accuracy deviation from the hold-out validation data the moving average of the class-wise accuracy deviations γ can be updated. The classifier weights are updated with the gradients calculated from the weighted losses. Finally, in line 12 of the algorithm, the class-wise weights are updated, by multiplying the current γ by a multiplier, based on the accuracy deviation. Here the sign function fulfills the purpose to indicate the update direction, while $\alpha \in [0, 1]$ indicates the update step. Note, that for $\alpha = 0$, the algorithm equals standard training.

5. Experiments

5.1. Experimental Setup

We evaluate the fair training algorithm for a ResNet56 and Wide-ResNet-28-10 architectures on CIFAR10 for standard and adversarial training. We set the two hyperparameters to $\alpha = 0.001$ and $\beta = 0.1$ for all our experiments if not otherwise mentioned. To obtain robust models, we use adversarial training with l_∞-PGD attack [22]. We use an allowed perturbation magnitude $\epsilon = 8/255$ for images in range $[0, 1]$, a number of 7 steps and a step size calculated as $2.5\frac{\epsilon}{\#\text{steps}}$. To evaluate adversarial robustness we also use the same l_∞ attack but with an increase in the number of steps to 10.

Table 1: Comparison of fair training with standard training. The values in the worst and best column indicate the worst and best class-wise accuracy, respectively together with the corresponding accuracy deviation.

	Avg. Acc	Worst	Best
ResNet56 (Std.)	92.8	86.2 / -6.6	96.9 / 4.1
ResNet56 (Fair)	90.3	87.8 / -2.5	92.1 / 1.8
WRN-28-10 (Std.)	95.4	89.0 / -6.4	98.0 / 2.6
WRN-28-10 (Fair)	92.9	89.0 / -3.9	94.6 / 1.7

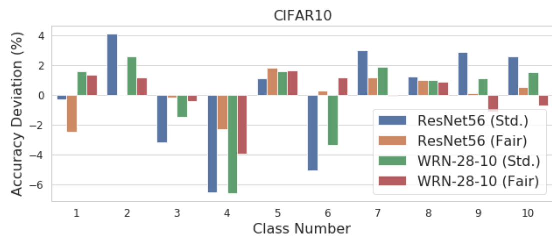

Figure 3: Comparison of accuracy deviations of the standard and fair training.

5.2. Standard Training

Table 1 compares the performance of the proposed fair training with standard training. The average accuracies show that the fair training strategy leads to slightly inferior performance. For example, the standard accuracy of the ResNet56 accuracy decreases by 2.5% with fair training. However, the class-wise deviation from the mean is significantly decreased. The class-wise deviation for the ResNet56 and Wide-ResNet are decreased by 4.1% and 2.5%, respectively, resulting in fairer classifiers. This result suggests a trade-off between standard accuracy and fairness. In Figure 3 the class-wise accuracies of the compared models are shown. Interestingly, with standard and fair training class number 4 ("cat") stays the worst class. Further, class number 6 ("dog") changed from being a relatively weak class to a class with a positive accuracy deviation.

Table 2: Comparison of fair training with adversarial training. The values in the worst and best column indicate the worst and best class-wise accuracy, respectively, together with the corresponding accuracy deviation (left). The worst and best class-wise robustness accuracies under adversarial attack are shown on the right.

	Avg. Acc	Worst Clean	Best Clean	Avg. Adv.	Worst Adv.	Best Adv.
ResNet56 (Adv.)	79.8	51.9 / -27.9	94.0 / 14.2	47.8	16.3 / -31.5	65.8 / 18.0
ResNet56 (Adv. Fair)	76.5	70.5 / -6.0	83.4 / 6.9	39.2	27.0 / -12.2	51.8 / 12.6
WRN-28-10 (Adv.)	86.3	74.9 / -11.4	94.5 / 8.2	43.8	17.5 / -26.3	68.2 / 24.4
WRN-28-10 (Adv. Fair)	80.7	77.7 / -3.0	84.7 / 4.0	34.3	20.9 / -13.4	46.5 / 12.2

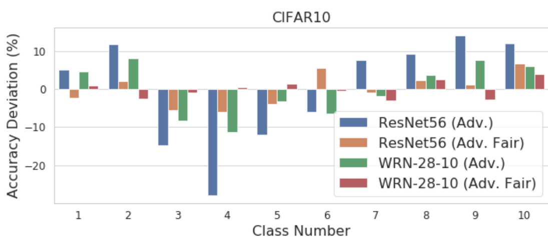

Figure 4: Comparison of accuracy deviations of the adversarial training and adversarial fair training.

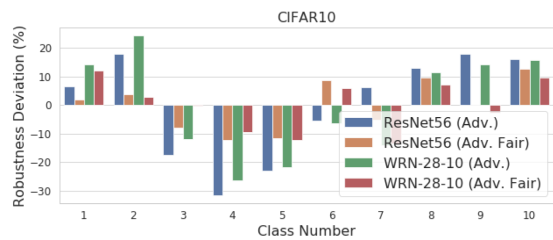

Figure 5: Comparison of robustness deviations of the adversarial training and adversarial fair training under adversarial attack.

5.3. Adversarial Training

Table 3 shows the results for adversarial training and fair adversarial training on CIFAR10. It can be observed that adversarially trained models are more "unfair", as indicated by their high class-wise accuracy deviations. For example, the lowest class-wise accuracy of ResNet56 is 51.7%, which deviates by 27.9% from the average accuracy of 79.8%. Adversarial fair training can significantly decrease this class-wise accuracy discrepancy to -6% for ResNet56. Table 2 additionally shows the evaluation of the adversarially trained models under adversarial attack. Interestingly, there also exists a class-wise robustness discrepancy for adversarially trained models, with the worst class-wise accuracy achieving robustness of only 16.3% and the best having a class-wise robustness of 65.8%. Adversarial fair training also increases the robustness fairness. For example, the robustness of the worst robust class was increased from 16.3% to 27.0% for ResNet56. Figure 4 presents the detailed class-wise accuracies for the adversarially trained models. The proposed fair training strategy results in a more fair robust classifier, with the class-wise accuracy deviations being closer to zero for all classes. The robustness deviations for the adversarially trained models are shown in Figure 5, where a similar fairness trend can be observed.

5.4. Ablation

In Figure 6 the mean accuracy, as well as the highest and lowest class-wise accuracy are shown over various α values. The accuracy-fairness trade-off can be well observed for increasing α values. While the band around the mean value decreases, indicating a more fair classifier, the mean accuracy decreases, for increasing α values. Among the examined α values the fairest classifier is obtained for $\alpha = 0.001$, which motivated us to choose this value for our experiments.

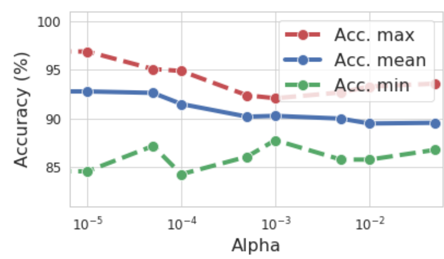

Figure 6: Ablation on Alpha

6. Conclusion

This work investigated the fairness of standard and robust classifiers and proposes a class-wise loss re-weighting strategy to increase fairness. The results suggest that fairness comes at the cost of accuracy or robustness and adds a third dimension to the already explored accuracy-robustness trade-off, suggesting a triangular trade-off between accuracy, robustness, and fairness. We leave the mitigation of the triangular trade-off open for future work.

References

[1] Anish Athalye, Nicholas Carlini, and David Wagner. Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples. In *ICML*, 2018. 1, 2

[2] Philipp Benz, Chaoning Zhang, Tooba Imtiaz, and In So Kweon. Double targeted universal adversarial perturbations. In *ACCV*, 2020. 2

[3] Philipp Benz, Chaoning Zhang, Adil Karjauv, and In So Kweon. Robustness may be at odds with fairness: An empirical study on class-wise accuracy. *NeurIPS workshop*, 2020. 2

[4] Philipp Benz, Chaoning Zhang, Adil Karjauv, and In So Kweon. Universal adversarial training with class-wise perturbations. *ICME*, 2021. 2

[5] Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 2018. 2

[6] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. In *NeurIPS*, 2019. 2

[7] Nicholas Carlini and David Wagner. Adversarial examples are not easily detected. In *ACM Workshop on Artificial Intelligence and Security-AISec'17*, 2017. 1, 2

[8] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 2002. 2

[9] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *CVPR*, 2019. 2

[10] Mengnan Du, Fan Yang, Na Zou, and Xia Hu. Fairness in deep learning: A computational perspective. *Intelligent Systems*, 2020. 2

[11] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *ICLR*, 2015. 1, 2

[12] Chuan Guo, Mayank Rana, Moustapha Cisse, and Laurens Van Der Maaten. Countering adversarial images using input transformations. *arXiv preprint arXiv:1711.00117*, 2017. 1, 2

[13] Haibo He and Edwardo A Garcia. Learning from imbalanced data. *Transactions on knowledge and data engineering*, 2009. 2

[14] Andrew Ilyas, Shibani Santurkar, Dimitris Tsipras, Logan Engstrom, Brandon Tran, and Aleksander Madry. Adversarial examples are not bugs, they are features. In *NeurIPS*, 2019. 2

[15] Nathalie Japkowicz and Shaju Stephen. The class imbalance problem: A systematic study. *Intelligent data analysis*, 2002. 2

[16] Maurice George Kendall et al. The advanced theory of statistics. *The advanced theory of statistics.*, (2nd Ed), 1946. 1, 2

[17] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. Cifar-10 (canadian institute for advanced research). *URL http://www.cs. toronto. edu/kriz/cifar. html*, 2010. 2

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. In *NeurIPS*, 2012. 2

[19] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 2015. 1

[20] Saehyung Lee, Hyungyu Lee, and Sungroh Yoon. Adversarial vertex mixup: Toward better adversarially robust generalization. In *CVPR*, 2020. 2

[21] Hong Liu, Rongrong Ji, Jie Li, Baochang Zhang, Yue Gao, Yongjian Wu, and Feiyue Huang. Universal adversarial perturbation via prior driven uncertainty approximation. In *ICCV*, 2019. 2

[22] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. In *ICLR*, 2018. 1, 2, 3

[23] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. *arXiv preprint arXiv:1908.09635*, 2019. 2

[24] Jan Hendrik Metzen, Tim Genewein, Volker Fischer, and Bastian Bischoff. On detecting adversarial perturbations. In *ICLR*, 2017. 1, 2

[25] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. Universal adversarial perturbations. In *CVPR*, 2017. 2

[26] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. Deepfool: a simple and accurate method to fool deep neural networks. In *CVPR*, 2016. 2

[27] Konda Reddy Mopuri, Utkarsh Ojha, Utsav Garg, and R. Venkatesh Babu. Nag: Network for adversary generation. In *CVPR*, 2018. 2

[28] Vedant Nanda, Samuel Dooley, Sahil Singla, Soheil Feizi, and John P Dickerson. Fairness through robustness: Investigating robustness disparity in deep learning. In *ACM Conference on Fairness, Accountability, and Transparency*, 2021. 2

[29] Tianyu Pang, Xiao Yang, Yinpeng Dong, Kun Xu, Jun Zhu, and Hang Su. Boosting adversarial training with hypersphere embedding. In *NeurIPS*, 2020. 2

[30] Nicolas Papernot, Patrick McDaniel, Xi Wu, Somesh Jha, and Ananthram Swami. Distillation as a defense to adversarial perturbations against deep neural networks. In *SP*, 2016. 1, 2

[31] Omid Poursaeed, Isay Katsman, Bicheng Gao, and Serge Belongie. Generative adversarial perturbations. In *CVPR*, 2018. 2

[32] Ali Shafahi, Mahyar Najibi, Mohammad Amin Ghiasi, Zheng Xu, John Dickerson, Christoph Studer, Larry S Davis, Gavin Taylor, and Tom Goldstein. Adversarial training for free! In *NeurIPS*, 2019. 2

[33] Li Shen, Zhouchen Lin, and Qingming Huang. Relay backpropagation for effective learning of deep convolutional neural networks. In *ECCV*, 2016. 2

[34] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013. 1, 2

[35] Jingru Tan, Changbao Wang, Buyu Li, Quanquan Li, Wanli Ouyang, Changqing Yin, and Junjie Yan. Equalization loss for long-tailed object recognition. In *CVPR*, 2020. 2

[36] Qi Tian, Kun Kuang, Kelu Jiang, Fei Wu, and Yisen Wang. Analysis and applications of class-wise robustness in adversarial training. *arXiv preprint arXiv:2105.14240*, 2021. 2

[37] Florian Tramer, Nicholas Carlini, Wieland Brendel, and Aleksander Madry. On adaptive attacks to adversarial example defenses. In *NeurIPS*, 2020. 1, 2

[38] Dimitris Tsipras, Shibani Santurkar, Logan Engstrom, Alexander Turner, and Aleksander Madry. Robustness may be at odds with accuracy. In *ICLR*, 2019. 1, 2

[39] Jianyu Wang and Haichao Zhang. Bilateral adversarial training: Towards fast training of more robust models against adversarial attacks. In *ICCV*, 2019. 2

[40] Eric Wong, Leslie Rice, and J Zico Kolter. Fast is better than free: Revisiting adversarial training. *ICLR*, 2020. 2

[41] Cihang Xie, Mingxing Tan, Boqing Gong, Alan Yuille, and Quoc V Le. Smooth adversarial training. *arXiv preprint arXiv:2006.14536*, 2020. 2

[42] Cihang Xie, Jianyu Wang, Zhishuai Zhang, Zhou Ren, and Alan Yuille. Mitigating adversarial effects through randomization. *ICLR*, 2018. 1, 2

[43] Han Xu, Xiaorui Liu, Yaxin Li, and Jiliang Tang. To be robust or to be fair: Towards fairness in adversarial training. *arXiv preprint arXiv:2010.06121*, 2021. 2

[44] Chaoning Zhang, Philipp Benz, Tooba Imtiaz, and In-So Kweon. Cd-uap: Class discriminative universal adversarial perturbation. In *AAAI*, 2020. 2

[45] Chaoning Zhang, Philipp Benz, Tooba Imtiaz, and In-So Kweon. Understanding adversarial examples from the mutual influence of images and perturbations. In *CVPR*, 2020. 2

[46] Dinghuai Zhang, Tianyuan Zhang, Yiping Lu, Zhanxing Zhu, and Bin Dong. You only propagate once: Accelerating adversarial training via maximal principle. In *NeurIPS*, 2019. 2

[47] Hongyang Zhang, Yaodong Yu, Jiantao Jiao, Eric P Xing, Laurent El Ghaoui, and Michael I Jordan. Theoretically principled trade-off between robustness and accuracy. In *Proceedings of Machine Learning Research*, 2019. 2

[48] Jingfeng Zhang, Xilie Xu, Bo Han, Gang Niu, Lizhen Cui, Masashi Sugiyama, and Mohan Kankanhalli. Attacks which do not kill training make adversarial learning stronger. In *ICML*. PMLR, 2020. 2

[49] Xiao Zhang, Zhiyuan Fang, Yandong Wen, Zhifeng Li, and Yu Qiao. Range loss for deep face recognition with long-tailed training data. In *ICCV*, 2017. 2