Out-of-distribution algorithms for robust insect classification

Mojdeh Saadati[1], Shivani Cheeranjeevi[2], Aditya Balu[2], Talukder Jubery[2], Asheesh K Singh[3], Soumik Sarkar[2], Arti Singh[3], Baskar Ganapathysubramanian[2]

[1] Department of Computer Science, Iowa State University
[2] Department of Mechanical Engineering, Iowa State University
[3] Department of Agronomy, Iowa State University
*baskarg@iastate.edu

Abstract

Plants are exposed to various useful and harmful insect pests during their growth cycle. Accurate identification of these pests is critical for deciding on a timely and appropriate mitigation strategy with significant economic and environmental implications. Recent progress in deep learning-based approaches has resulted in insects exhibiting good accuracy. However, deploying them in the wild is still problematic due to the fact that input images that are wildly out of the distribution (for e.g., non-insect images like vehicles, animals, or a blurred image of an insect or insect class that is not yet trained on) can still produce insect classification. To counter this, methods that ensure that a model abstains from making predictions are needed. To address this issue, we leverage the out-of-distribution detection concept that showed promising results in detecting out-of-distribution data in dermatology tasks (Roy et al. 2022). In our work, we evaluate the performance of state-of-the-art out-of-distribution (OOD) algorithms on insect detection classifiers. These algorithms represent a diversity of methods of approaching an OOD problem. Additionally, we focus on extrusive algorithms – i.e., algorithms that wrap around a pre-trained classifier without the need for additional co-training. We choose three OOD detection algorithms: (i) Maximum Softmax Probability (MSP), commonly referred to as the baseline algorithms (Hendrycks and Gimpel 2016), (ii) Mahalanobis distance-based algorithm which solves the problem using a generative classification approach (Lee et al. 2018; Ren et al. 2021), and (iii) Energy-Based Model OOD detection algorithm, which exhibits SOTA for OOD detection (Liu et al. 2020). We perform an extensive series of evaluations of these OOD algorithms across two performance axes: (a) how the accuracy of the classifier impacts OOD performance, and (b) how the degree of out-of-domain impacts OOD performance. The result of our analysis shows OOD detection algorithms can significantly improve from abstaining classification across different settings of models' structures and datasets. Thus, our OOD-robust classifier improves user trust in using the application for insect-pests classification.

Introduction

Insect pest-related diseases in crops and plants can be observed at all stages of their growth, negatively affecting the quality and quantity of yields. Therefore, accurate detection of insects is imperative for the decision-making of strategies. The insect detection task was initially handled with traditional machine-learning algorithms. In these algorithms, first, a set of features such as color and texture are extracted from images, and then an object detector creates a mapping from the feature space to their corresponding label. SVM is one of the most common examples of these algorithms used frequently by researchers in insect detection problems (Ebrahimi et al. 2017; Kasinathan, Singaraju, and Uyyala 2021). Such approaches require extensive domain knowledge about the input data for feature extraction and choice of classifiers. However, this issue has been tackled with the emergence of deep learning algorithms. In classifying 13 soybean pests, the performance of 5 models(Inception-v3, Resnet-50, VGG-16, VGG-19, and Xception) was compared across the dataset of size 5000 samples in (Tetila et al. 2020). (Li et al. 2020) leveraged the GoogLeNet and achieved 98 percent accuracy in a 10-class insect classification task with a manually collected dataset. Manual labeling of the large dataset requires extensive hours of experts, which is not always accessible. Self-supervised algorithms eliminate the need for labeling by utilizing specific parts of the image to predict other parts. (Kar et al. 2021) offered the BYOL, a self-supervised pest classifier algorithm, with up to 93 percent accuracy. While these algorithms achieved high accuracy in classifications, neither of them could claim the certainty of their decisions or abstain from classification in the case of uncertainty.

One of the initial algorithms proposed for handling out-of-distribution data is maximum-softmax-probability (MSP) (Hendrycks and Gimpel 2016). This algorithm relies on the assumption that deep learning models are more confident in the classification of in-distribution data rather than out-of-distribution (OOD) ones. The paper used the softmax value as a metric to measure the confidence of predictions. Due to its simplicity and good performance, this algorithm has been prevalent in addressing OOD detection. However, it has been shown in practice that MSP has a high false positive rate in OOD detection. Liu et al. (2020) has proven theoretically and analytically that an energy-based model can be a great substitute for MSP as they are aligned with the probability density of in-distribution data.

Another group of OOD algorithms proposed introducing

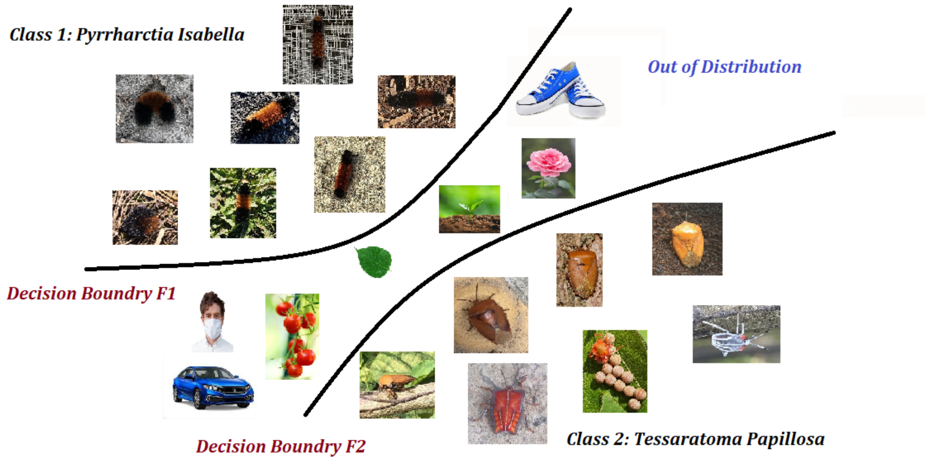

Figure 1: Out-of-distribution visualization in insect classification

part of the OOD data during the training process. Hendrycks, Mazeika, and Dietterich (2018) modify the classifier's cross-entropy loss function and add an extra term to handle OOD data so that the softmax distribution for OOD data will be uniform. Researchers also incorporated OOD detection into the classifier's structure and designed a hierarchical outlier detection algorithm (HOD) to identify outlier dermatological conditions (Roy et al. 2022). Fort, Ren, and Lakshminarayanan (2021) and Ren et al. (2019) also adjust the classifier by adding an extra class to classify the OOD data during the prediction.

While discriminative algorithms try to find the best decision boundaries, generative models focus on estimating the in-distribution density. Scientists used this feature of generative models in solving the OOD problem (Choi and Jang 2018; Nalisnick et al. 2018; Ren et al. 2019; Serrà et al. 2019). Among generative model-based OOD algorithms, OOD detection based on Mahalanobis distance is the most popular. Denouden et al. (2018) solves the OOD problem in the context of auto-encoder architecture, using the fact that auto-encoder is ineffective in encoding and reconstructing the OOD data in comparison to in-distribution ones. They distinguish OOD data from in-distribution by defining a threshold based on the Mahalanobis distance metric on reconstruction error. Moreover, Lee et al. (2018) extracts class conditional gaussian distributions of deep learning features based on Gaussian discriminant analysis, leading to a Mahalanobis distance-based confidence score. Ren et al. (2021) commented that the latter algorithm suffered from near-OOD detection and offered an adjustment to the previous algorithm.

In this paper, for the first time, we impose the idea of out-of-distribution detection in the agricultural domain. Also, we evaluate the OOD algorithm's performance in a large insect dataset of iNaturalists. This differs from previous works, which conduct their analysis on benchmark datasets such as CIFAR10, CIFAR100, and SVHN, which have relatively smaller data sizes. We evaluate the performance of the three OOD methods (MSP, Mahalanobis distance, and energy-based model) for the insect detection tasks on various fine-tuned insect classifiers. Mainly, we answer the following questions:

- How the model accuracy is affecting the OOD detection performance?
- Does the OOD detection algorithms' performance depend on the distribution of the dataset?

The paper is organized as follows. Section 3 explains the problem definition and each OOD detection method we evaluate. In section 4, we illustrate the evaluation results. In section 5, we will conclude the research paper and the potential future work.

Methods

To answer the above-mentioned questions, we first prepared the datasets, next trained the insect classifiers, and last applied the OOD algorithms on the combination of insect classifiers and dataset and measured their performance. In the following section, we will go through each of the processes in more detail.

Datasets

We curated two series of datasets, one for in-distribution (ID) and one for OOD. We used ID data for three main purposes: (i) training the insect classifier, (ii) training the generative OOD algorithm (Mahalanobis distance), and (iii) evaluating the performance of OOD algorithms with respect to distinguishing the ID data from OOD data. For these objectives, we curated an insect dataset consisting of the top 142 agriculturally relevant species (in terms of economic impact in North America). This dataset is a subset of the publicly available iNaturalist dataset and consists of 2 million insect pictures. We split the dataset into train and validation sets with a ratio of 7 to 3 and then obtain the accuracy of each classifier. We then divide the validation folder into two equally sized smaller datasets. One was used for training the

OOD model (only for the generative OOD model), and the other was used for evaluating the performance of OOD algorithms.

For OOD data, we utilized four datasets with different degrees of similarity to ID data. We briefly describe each of these datasets.

- ImageNet (Russakovsky et al. 2015) (far OOD): We downloaded the ImageNet 2012 classification data with 1000 object categories. We then excluded all insect-related objects from it.
- Human Face (Wang et al. 2020) (far OOD): We downloaded the data from the face mask recognition dataset of the Kaggle competition[1], which includes pictures of human faces with and without masks.
- NonInsecta (Van Horn et al. 2021) (near OOD): this dataset is a subset of iNaturalist [2], where we exclude all the Insecta images from it.
- OODInsect (near OOD): This dataset includes all insect pictures that do not belong to any of the 142 classes of ID data. This dataset is also collected from the iNatrualist website.

Insect Classifiers Methods

- **Resnet50** ResNet is a popular CNN model that was proposed by He et al. (2016), and it has proven to produce high classification accuracies for computer vision and image classification tasks. The success was attributed to the presence of skip connections in its residual blocks that overcome the diminishing or exploding gradients. We use a variant of this model, ResNet-50, which is 50 layers deep in our paper.
- **RegnetY32** RegNet is an optimized design space developed by Radosavovic et al. (2020) where they explore a diverse set of parameters of a network structure like width, depth, groups, etc (commonly called as AnyNet, an initial space of unconstrained models which uses models like ResNet (He et al. 2016) as its base). By conducting many experiments of trying different parameter values for the design space, they arrived at the optimized RegNetX or RegNetY models. In this paper, we use the RegNetY32 model for our experiments.

Out-of-distribution Methods

- **Maximum Softmax Probability(MSP) (Hendrycks and Gimpel 2016)** This algorithm, simply by utilizing the maximum/predicted class probability as a confidence score, distinguishes between ID vs OOD data.
- **Mahalanobis distance-based algorithm (Lee et al. 2018)** In this algorithm, they solve the problem of OOD binary classification with the help of a generative classifier. This classifier is created under Gaussian discriminative analysis. For a given input data X and a classifier with a range of labels, $\{1, ..., C\}$, the algorithm assumes that the class conditional distributions of predictions are from a multivariate Gaussian distribution.

  Using the output of the penultimate layer of the classifier (denoted as $f(x)$), they assume $P(f(x)|y = c) = \mathcal{N}(f(x)|\mu_c, \hat{\Sigma})$ where μ_c and $\hat{\Sigma}$ are consecutively mean and covariance of the multivariate Gaussian distributions. They calculate μ_c and $\hat{\Sigma}$ for a given OOD training sample $\{(x_1, y_1), ..., (x_N, y_N)\}$ with the following formula:

$$\hat{\mu_c} = \frac{1}{N_c} \sum_{i:y_i=c} f(x_i) \tag{1}$$

$$\hat{\Sigma} = \frac{1}{N} \sum_{c} \sum_{i:y_i=c} (f(x_i) - \hat{\mu_c})(f(x_i) - \hat{\mu_c})^T \tag{2}$$

  They introduce Mahalanobis distance-based(MAH) confidence score, M(x) as a distance from a sample x to the closets class-conditional Gaussian distribution:

$$M(x) = \max_c -(f(x) - \hat{\mu_c})^T \hat{\Sigma}^{-1}(f(x) - \hat{\mu_c}) \tag{3}$$

- **Energy-based models (EBM) (Liu et al. 2020)** The energy model is a function from $E(x) : R^D \rightarrow R$ where each input value is mapped into a non-probabilistic Energy value. Energy values can be converted to probability density through Gibbs distribution:

$$p(y|x) = \frac{e^{-E(x,y)/T}}{\int_{y'} e^{-E(x,y)/T}} \tag{4}$$

  The Helmholtz free energy function $E(x)$ for $x \in R^D$ is expressible based on the denominator with the following formula:

$$E(x) = -Tlog \int_{y'} e^{-E(x,y')/T} \tag{5}$$

  Based on the similarity of Equation 4 with the softmax formula, we can replace the energy parameter $E(x, y)$ with the logit value of the classifier $f(x) : R^D \rightarrow R^K$ with $-f_y(x)$ and define the energy function for a given classifier and input data as bellow:

$$E(x) = -Tlog \sum_{i}^{K} e^{f_i(x)/T} \tag{6}$$

  In the above formula, the parameter T is referred to as temperature which we set $T = 1$ for the purpose of our paper. We utilize the value returned from the energy function as a confidence score; we expect the in-distribution data to return a lower value for the energy than OOD.

Results

We present our results in terms of answers to questions that we had posed in the introduction:

RQ1. How does the model accuracy affect the OOD detection performance? To answer this question, we first trained the insect classifiers based on two different architectures for the classifier(ResNet50 and RegNetY32) for 50 epochs each. We trained our classifiers with the settings of

[1]https://www.kaggle.com/datasets/ashwingupta3012/human-faces

[2]https://github.com/visipedia/inat$_{c}omp/tree/master/$2021

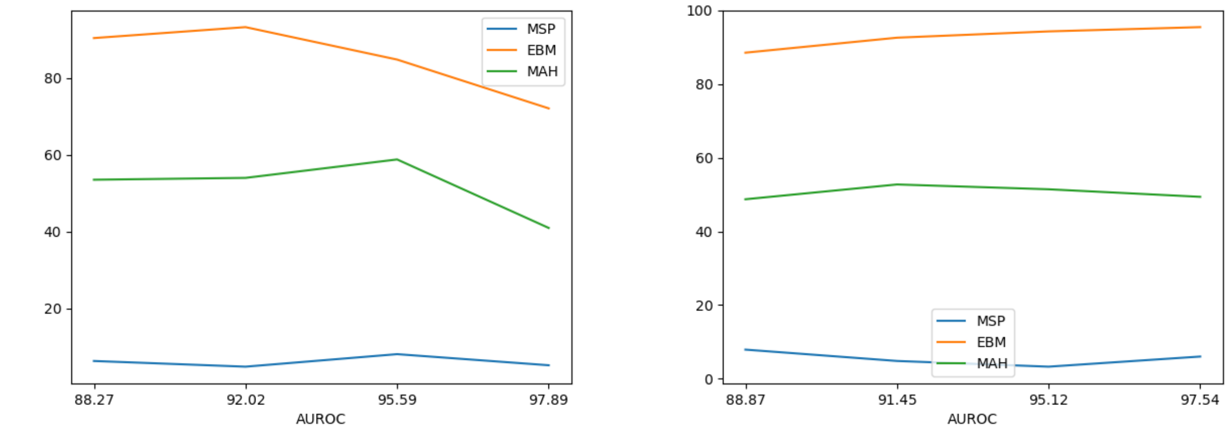

Figure 2: The trend of OOD detection with respect to increment in accuracy

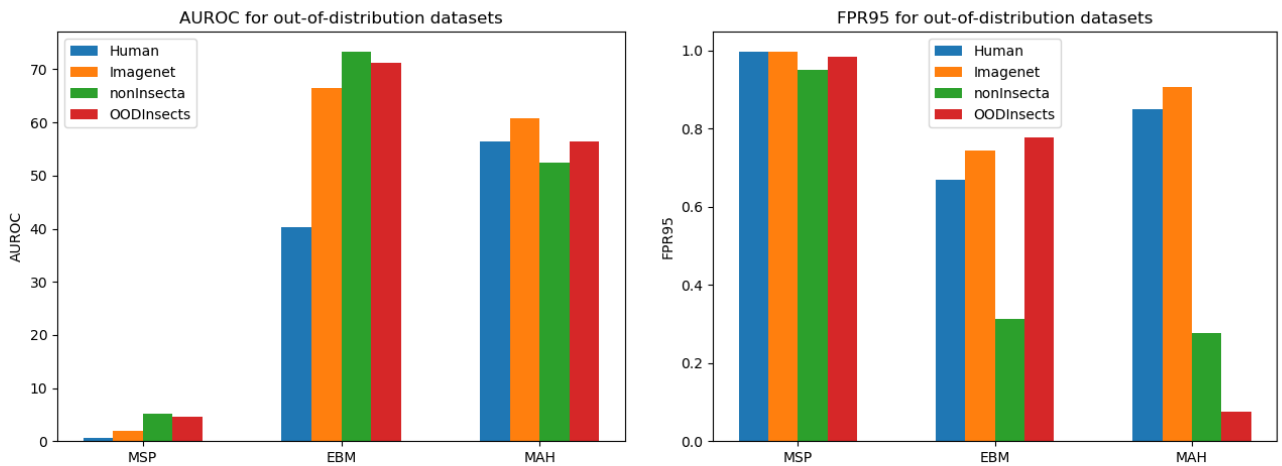

Figure 3: Comparison between OOD datasets with different degrees of similarity to ID on OOD detection

the cross-entropy loss function and AdamW optimizer with a start learning rate of $1e-3$. For loading the data, we utilize the batch size of 256. We then choose epochs $\{0, 1, 10, 49\}$ to evaluate their accuracy on the validation set. Next, we run all three OOD algorithms on the 4 chosen epochs of Resnet50 and Regnet32. The result of our analysis is shown in Figure 2 (left) and Figure 2 (right).

These figures illustrate the consistently better performance of EBM in comparison to the other two approaches. Furthermore, we also note that for the ResNet50 architecture, better accuracy of the classifier leads to better OOD detection, but this is not consistent in RegNetY32. This contradicts the common notion that a good classifier will always lead to good OOD Detection (Vaze et al. 2021). Our results indicate more stability of OOD Detection in ResNet50 than its counterpart, RegNetY32.

RQ2. Does the OOD detection algorithms' performance depend on the distribution of the out-of-distribution dataset? To explore this question, we select 4 sets of data with a wide range of degrees of similarity to ID data, as explained above. Then for the trained RegnetY32 model, we compare the in-distribution data to the four OOD dataset: Imagenet, Human Face, NonInsecta, and OODInsect. Our results shown in Figure 3 also confirm our observations from RQ1 that EBM has the best performance on all 4 OOD datasets. Also, the figures endorse the claim in (Liu et al. 2020) about MSP having the highest FPR95. It is also noticeable that MSP is not affected by the level near OOD or far OOD. Despite MSP, EBM shows a significant improvement in detecting near OOD. Moreover, we observe that FPR95 for near OOD is lower in EBM and MAH in comparison to far OOD.

Conclusions

Automated insect pest detection is an economically critical agricultural task. It is important that well trained models, when deployed in the wild, abstain from making predictions when encountering data that is out of their training distribution. We explore and quantify the performance of several OOD approaches applied to insect pest classification. We expect this study to ensure enhanced trust worthiness of deployed models.

References

Choi, H.; and Jang, E. 2018. Generative ensembles for robust anomaly detection.

Denouden, T.; Salay, R.; Czarnecki, K.; Abdelzad, V.; Phan, B.; and Vernekar, S. 2018. Improving reconstruction autoencoder out-of-distribution detection with Mahalanobis distance. *arXiv preprint arXiv:1812.02765*.

Ebrahimi, M.; Khoshtaghaza, M. H.; Minaei, S.; and Jamshidi, B. 2017. Vision-based pest detection based on SVM classification method. *Computers and Electronics in Agriculture*, 137: 52–58.

Fort, S.; Ren, J.; and Lakshminarayanan, B. 2021. Exploring the limits of out-of-distribution detection. *Advances in Neural Information Processing Systems*, 34: 7068–7081.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hendrycks, D.; and Gimpel, K. 2016. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*.

Hendrycks, D.; Mazeika, M.; and Dietterich, T. 2018. Deep anomaly detection with outlier exposure. *arXiv preprint arXiv:1812.04606*.

Kar, S.; Nagasubramanian, K.; Elango, D.; Nair, A.; Mueller, D. S.; O'Neal, M. E.; Singh, A. K.; Sarkar, S.; Ganapathysubramanian, B.; and Singh, A. 2021. Self-Supervised Learning Improves Agricultural Pest Classification. In *AI for Agriculture and Food Systems*.

Kasinathan, T.; Singaraju, D.; and Uyyala, S. R. 2021. Insect classification and detection in field crops using modern machine learning techniques. *Information Processing in Agriculture*, 8(3): 446–457.

Lee, K.; Lee, K.; Lee, H.; and Shin, J. 2018. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Advances in neural information processing systems*, 31.

Li, Y.; Wang, H.; Dang, L. M.; Sadeghi-Niaraki, A.; and Moon, H. 2020. Crop pest recognition in natural scenes using convolutional neural networks. *Computers and Electronics in Agriculture*, 169: 105174.

Liu, W.; Wang, X.; Owens, J.; and Li, Y. 2020. Energy-based out-of-distribution detection. *Advances in Neural Information Processing Systems*, 33: 21464–21475.

Nalisnick, E.; Matsukawa, A.; Teh, Y. W.; Gorur, D.; and Lakshminarayanan, B. 2018. Do deep generative models know what they don't know? *arXiv preprint arXiv:1810.09136*.

Radosavovic, I.; Kosaraju, R. P.; Girshick, R.; He, K.; and Dollár, P. 2020. Designing network design spaces. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10428–10436.

Ren, J.; Fort, S.; Liu, J.; Roy, A. G.; Padhy, S.; and Lakshminarayanan, B. 2021. A simple fix to mahalanobis distance for improving near-ood detection. *arXiv preprint arXiv:2106.09022*.

Ren, J.; Liu, P. J.; Fertig, E.; Snoek, J.; Poplin, R.; Depristo, M.; Dillon, J.; and Lakshminarayanan, B. 2019. Likelihood ratios for out-of-distribution detection. *Advances in neural information processing systems*, 32.

Roy, A. G.; Ren, J.; Azizi, S.; Loh, A.; Natarajan, V.; Mustafa, B.; Pawlowski, N.; Freyberg, J.; Liu, Y.; Beaver, Z.; et al. 2022. Does your dermatology classifier know what it doesn't know? detecting the long-tail of unseen conditions. *Medical Image Analysis*, 75: 102274.

Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; Berg, A. C.; and Fei-Fei, L. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3): 211–252.

Serrà, J.; Álvarez, D.; Gómez, V.; Slizovskaia, O.; Núñez, J. F.; and Luque, J. 2019. Input complexity and out-of-distribution detection with likelihood-based generative models. *arXiv preprint arXiv:1909.11480*.

Tetila, E. C.; Machado, B. B.; Astolfi, G.; de Souza Belete, N. A.; Amorim, W. P.; Roel, A. R.; and Pistori, H. 2020. Detection and classification of soybean pests using deep learning with UAV images. *Computers and Electronics in Agriculture*, 179: 105836.

Van Horn, G.; Cole, E.; Beery, S.; Wilber, K.; Belongie, S.; and Mac Aodha, O. 2021. Benchmarking Representation Learning for Natural World Image Collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12884–12893.

Vaze, S.; Han, K.; Vedaldi, A.; and Zisserman, A. 2021. Open-set recognition: A good closed-set classifier is all you need. *arXiv preprint arXiv:2110.06207*.

Wang, Z.; Wang, G.; Huang, B.; Xiong, Z.; Hong, Q.; Wu, H.; Yi, P.; Jiang, K.; Wang, N.; Pei, Y.; et al. 2020. Masked face recognition dataset and application. *arXiv preprint arXiv:2003.09093*.