Contrastive Learning for Out-of-Distribution Image Detection in AI-Based Pest Management App

Austin Nguyen [*1] **Eric Helmold** [*1] **Erin Tomlinson** [*1] **Weiwei Pan** [1] **Jerome White** [2] **Soma Dhavala** [2]

Abstract

We seek to help with out-of-distribution (OOD) detection in images submitted on a mobile platform (CottonAce) in applications for pest management. Previous work in OOD detection for pest management are limited because the scope of OOD is too wide and the models built are too large to load onto a mobile device at inference time. **Our goal is to explore contrastive learning, namely contrasting shifted instances (CSI), as an potential technique for OOD detection.** We focus our work on better understanding the hyperparameter space of CSI to discern how it can produce effective latent representations for the downstream task of OOD detection in a self-supervised manner.

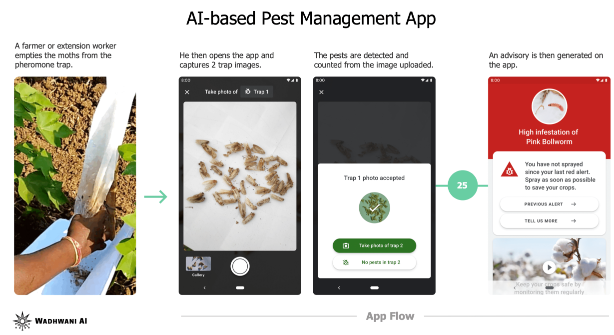

Figure 1. Example user-flow for the CottonAce pest management app. Graphic created by app development team at Wadhwani AI.

1. Introduction

Wadhwani AI is an independent, nonprofit institute developing AI-based solutions for underserved communities in developing countries. One important area of focus is pest management for cotton farmers. Cotton is the most important fiber and a cash crop for India, providing about 6 million farmers with a direct livelihood and 40-50 million people work in the cotton trade. Small-holder farmers, contributing 75% of aggregate production, struggle with uncertainty in crop yields and income. Cotton is exceptionally vulnerable to pest attacks, with bollworms responsible for an estimated 70% of all pest damage (White et al., 2022b; Dalmia et al., 2020).

Wadhwani AI has developed a mobile phone application called CottonAce that helps cotton farmers manage bollworm infestations in their fields. Bollworms are a pernicious pest, requiring consistent monitoring and expert decision making to properly address. The CottonAce app uses an AI algorithm to identify and count bollworms in photos, and then makes customized treatment recommendations to the farmers based on what it has found. This process is illustrated in Figure 1.

A key challenge is gracefully handling photos that are outside of the expected domain. The app instructs users to submit photos of bollworms on top of a blank, white sheet of paper that fills the entire frame (White et al., 2022b). Images that follow these guidelines are referred to as in-distribution (ID), while those that do not are considered out-of-distribution (OOD). Examples of OOD images may include both photos of bollworms taken under a relaxed interpretation of the guidelines (e.g., on soiled paper that does not fill the frame), and photos that contain no bollworms at all (e.g., test images taken by new users experimenting with the app's functionality). In either case, the AI algorithm is liable to make a mistake, and the app should abstain from making a recommendation.

The app's ability to detect OOD images and abstain from making a recommendation is critically important. From a user-design perspective, farmers can simply be prompted to retake and resubmit the photo. From a broader perspective, however, such actions have the potential to build farmers' trust and dramatically reduce false infestation alerts, which have been linked to several negative outcomes, such as excessive pesticide use, emotional distress, and early app

[*]Equal contribution [1]Harvard University, Institute for Applied Computational Science (IACS) [2]Wadhwani AI. Correspondence to: Austin Nguyen <austinnguyen@g.harvard.edu>, Erin Tomlinson <erintomlinson@g.harvard.edu>, Eric Helmold <ehelmold@g.harvard.edu>.

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021.

abandonment (White et al., 2022b).

Our project goal is to explore contrastive learning as a potential approach for OOD detection for pest management. We focus on exploration on two aspects (1) choosing a reasonable set of distribution-shifting augmentations for our application of OOD detection for pest management and (2) modifying the weight hyperparameter in the multi-objective CSI loss function. Our contributions are three-fold: (1) an empirical and applied example of a contrastive learning for agricultural pest OOD detection, (2) a more robust understanding of the loss function unique to CSI, and (3) an examination of the performance of the CSI score function on a non-CSI derived latent embedding.

2. Contributions

Previous literature in OOD detection has focused on density-based, reconstruction-based, one-class classification, and self-supervised approaches to model representations that encode normality and define a detection score (Yang et al., 2021; Ruff et al., 2021). Recent work (Tack et al., 2020) explores a new approach by taking ideas from computer vision and audio processing — contrastive learning — to extract strong inductive bias from multiple views of a sample by letting them attract each other and repelling them from other samples. The authors propose a simple method called contrasting shifted instances (CSI) specifically for OOD detection. This unique contribution to representation learning helps to learn a more discriminative representation for detecting OODs and designing a score function that utilizes this new representation form. Their method has two goals: (1) discriminating between in-and out-of-distribution as well as (2) discriminating within in-distribution.

In addition to contrasting a sample with other instances in a batch consistent with traditional contrastive learning approaches, the authors additionally contrast the sample with versions of itself with augmentations that shift the distribution towards being OOD. The authors build on past literature in two ways: (1) a novel training method which contrasts samples with distribution-shifting augmentations applied to the sample with itself and other instances and (2) a scoring function which utilizes the representation learned by contrastive learning. Additionally, the CSI method was shown to improve confidence calibration for conventional classifier models. The authors conduct experiments that demonstrate the superior performance of CSI across multiple domains (unlabeled one-class, unlabeled multi-class, and labeled multi-class settings). Additional details about these domains can be found in the original article (Tack et al., 2020).

3. Background

The CSI concept builds on SimCLR: a simple contrastive framework for learning visual representations (Chen et al., 2020). SimCLR is a powerful, but generic contrastive learning framework that is not specifically optimized for the OOD detection task. It was originally developed for unsupervised pretraining of image classifiers. The authors of CSI recognized, however, that it could be adapted for OOD detection if they could eliminate its need for negative samples and train it solely on ID images.

3.1. Loss function

In SimCLR, if we let x be a query image, $\{x_+\}$ and $\{x_-\}$ be sets of positive and negative samples, and $\mathrm{sim}(z, z') := z \cdot z' / \|z\| \|z'\|$ be the cosine similarity between two representations z and z', then the simple contrastive loss function can be written as

$$\mathcal{L}_{\text{con}}(x, \{x_+\}, \{x_-\}) :=$$
$$\frac{1}{|\{x_+\}|} \log \frac{\sum_{x' \in \{x_+\}} \exp(\mathrm{sim}(z(x), z(x'))/\tau)}{\sum_{x' \in \{x_+\} \cup \{x_-\}} \exp(\mathrm{sim}(z(x), z(x'))/\tau)},$$

where $|\cdot|$ denotes cardinality, $z(x)$ is the representation for image x, and τ is a temperature hyperparameter. The representation $z(x)$ can come from the output of the encoder network f_θ, i.e, $z(x) = f_\theta(x)$, or it can come from an additional projection layer g_ϕ, i.e., $z(x) = g_\phi(f_\theta(x))$. The authors specifically examine the SimCLR objective based on the task of instance discrimination. For example, if $\tilde{x}_i^{(1)}$ and $\tilde{x}_i^{(2)}$ are independent augmentations of x_i from a predefined set of transformations $\mathcal{T}$, then the SimCLR objective can be defined in terms of the contrastive loss as

$$\mathcal{L}_{\text{SimCLR}}(\mathcal{B}; \mathcal{T}) :=$$
$$\frac{1}{2B} \sum_{i=1}^{B} \mathcal{L}_{\text{con}}(\tilde{x}_i^{(1)}, \tilde{x}_i^{(2)}, \tilde{\mathcal{B}}_{-i}) + \mathcal{L}_{\text{con}}(\tilde{x}_i^{(1)}, \tilde{x}_i^{(2)}, \tilde{\mathcal{B}}_{-i}),$$

where batch $\mathcal{B} := \{x_i\}_{i=1}^B$, $\tilde{\mathcal{B}} := \{\tilde{x}_i^{(1)}\}_{i=1}^B \cup \{\tilde{x}_i^{(2)}\}_{i=1}^B$, and $\tilde{\mathcal{B}}_{-i} := \{\tilde{x}_j^{(1)}\}_{j \neq i} \cup \{\tilde{x}_j^{(2)}\}_{j \neq i}$.

The SimCLR authors carried out a study to understand the set of augmentations $\mathcal{T}$ that would generate good representations when used in the SimCLR training objective. They found that that some augmentations, such as rotation, can degrade the discriminative power of SimCLR when it is paired with a linear classifier. In contrast, Tack et al. find that the very augmentations that hurt the SimCLR discriminative performance can be useful for OOD detection by considering these examples as negative pairs instead of positive pairs. Tack et al. thus define a family of augmentations $\mathcal{S}$, called distribution-shifting transformations or shifting transformations, that generate better representations for OOD detection.

Contrasting shifted instances. The authors consider a set $\mathcal{S}$ of K different random or deterministic transformations that include the identity I, i.e., $S := \{S_0 = I, S_1, ..., S_{K-1}\}$. Vanilla SimCLR considered augmented images as positive samples. In contrast, the CSI authors define them as negative samples if the augmentation is from $\mathcal{S}$. The authors define the contrasting shifted instances (con-SI) loss function as follows:

$$\mathcal{L}_{\text{con-SI}} := \mathcal{L}_{\text{SimCLR}} \left(\bigcup_{S \in \mathcal{S}} \mathcal{B}_S; \mathcal{T} \right),\ \mathcal{B}_S := \{S(x_i)\}_{i=1}^{B}\,.$$

In this regime, we consider each shifted sample wherein $S \neq I$ as OOD with respect to original sample. This formulation allows con-SI to discriminate an in-distribution ($S = I$) sample from other OOD samples.

Classifying shifted instances. The authors consider an auxiliary task to predict which shifting transformation $y^S \in \mathcal{S}$ is applied for input x to allow f_θ to discriminate between each shifted example. Specifically, they add a linear layer at the output of f_θ to produce a softmax classifier $p_{\text{cls}-\text{SI}}(y^S|x)$, and train it using the classifying shifted instances (cls-SI) loss as follows:

$$\mathcal{L}_{\text{cls}-\text{SI}} := \\ \frac{1}{2B}\frac{1}{K}\sum_{S\in\mathcal{S}}\sum_{\tilde{x}_S\in\tilde{\mathcal{B}}_S} -\log p_{\text{cls}-\text{SI}}(y^S = S|\tilde{x}_S)\,.$$

This auxiliary task is motivated by research that shows that by training a discriminative classifier to distinguish between transformed images and non-transformed images, the classifier must learn salient geometrical features that are likely unique to a single class of features (Golan & El-Yaniv, 2018). For example, discriminator of transformed and normal images encourages learning of features that are useful for detecting novelties, which overlaps with our task of detecting anomalies in our pest management data.

CSI objective. The final objective for the CSI model is a weighted combination of the two previously considered loss functions

$$\mathcal{L}_{\text{CSI}} = \lambda\mathcal{L}_{\text{con}-\text{SI}} + \mathcal{L}_{\text{cls}-\text{SI}}\,,$$

where λ is a tunable hyperparameter which they set equal to one in their experiments[1].

[1] In the CSI paper, the hyperparameter λ appears on the second term in the loss objective, rather than the first. However, λ appears on the first term in their software implementation, so we adopt that convention here for consistency with our later experiments.

3.2. Score function

From the image representations $z(\cdot)$ produced by the CSI training objective, two features were found to be highly effective for OOD detection: (a) the cosine similarity to the nearest training sample and (b) the norm of the representation. By definition, the contrastive loss increases the norm of the in-distribution samples because it is an easy way to minimize cosine similarity of identical samples by increasing the denominator in L_{con}. Thus, the authors combine these two features into a baseline OOD detection score:

$$s_{\text{con}}(x; \{x_m\}) := \max_m \text{sim}(z(x_m), z(x)) \cdot \|z(x)\|\,.$$

The authors additionally reduce computational cost and memory load by selecting a representative coreset of ID samples to use for the nearest-neighbor calculation above.

Incorporating information from shifted transformations. The authors further improve the detection score for s_{con} by incorporating shifted transformations $\mathcal{S}$ from the component objectives L_{con-SI} and L_{cls-SI}:

$$s_{\text{con}-\text{SI}}(x; \{x_m\}) := \sum_{S\in\mathcal{S}} \lambda_S^{\text{con}} s_{\text{con}}(S(x); \{S(x_m)\})\,,$$

where λ_S^{con} is a balancing term that scales scores of each shifting transformation for M training samples, and

$$s_{\text{cls}-\text{SI}}(x) := \sum_{S\in\mathcal{S}} \lambda_S^{\text{cls}} W_S f_\theta(S(x))\,,$$

where λ_S^{cls} is a separate balancing term and W_s is weight vector from the linear layer of $p(y^S|x)$ for $s \in \mathcal{S}$.

The combined OOD detection score for CSI is as follows:

$$s_{\text{CSI}}(x; \{x_m\}) := s_{\text{con}-\text{SI}}(x; \{x_m\}) + s_{\text{cls}-\text{SI}}(x)\,.$$

Ensemble over random augmentations. The authors find additional improvements by ensembling scores over random distribution-preserving augmentations $T(x)$ for $T \sim \mathcal{T}$. For each instance, an ensembled CSI score is defined by $s_{\text{CSI}-\text{ens}}(x) := E_{T\sim\mathcal{T}}[s_{\text{CSI}}(T(x))]$.

4. Data

Our analysis is based on images from the Wadhwani AI Pest Management Open Data repository (White et al., 2022a). This repository provides a dataset consisting of 9,727 images captured by farmers using the CottonAce app during a period of its initial deployment. The images are annotated with bounding boxes around each bollworm, which support the primary pest-counting task. To support the task of OOD detection, our team sorted the images into three categories based on how closely they adhered to the app's image quality guidelines, example images from each category are shown in Figure 2. In addition to the aforementioned ID and OOD categories, we made the decision to introduce a third category for edge case (EC) images that fell somewhere in-between. These images lie in the difficult-to-define boundary between ID and OOD images and, as such, present a unique challenge that any OOD detection algorithm will ultimately have to resolve.

Figure 2. Our group organized the data into three categories based on how well the images aligned with the quality guidelines provided by the CottonAce mobile app (White et al., 2022b).

5. Experiments

For our project, we conducted a series of experiments intended to (a) assess the effectiveness of CSI for OOD detection in our pest management dataset and (b) explore the impact that model design choices and hyperparameters have on its performance. We downloaded and modified the official PyTorch implementation to work with our data. CSI models take a long time to train (the authors recommended 1000 epochs), so we installed the software on Harvard's FASRC supercomputer cluster and ran our experiments on a large number of dedicated, high-performance GPU nodes.

5.1. Distribution-shifting augmentations

As discussed in Section 2, distribution-maintaining augmentations are transformations that we could apply to ID images that would leave them within the ID category (e.g., noise, blur, etc.). Distribution-shifting augmentations, on the other hand, are severe enough to flip ID images into the OOD category (e.g., color inversion, block permutation, etc.). Considered augmentations are shown in Figure 3; which of these will be most effective in each category will be dataset dependent. For example, the CSI authors use rotation as a distribution-shifting augmentation for ImageNet because nearly all of its images are right-side up. Our dataset consists of pictures of bollworms on white sheets of paper

Figure 3. Distribution-shifting image augmentations explored by the CSI authors (Tack et al., 2020).

and, as such, we hypothesized that it would be better to treat rotations as distribution-maintaining.

To test this hypothesis, we trained two CSI models on our ID images: one using rotation as its distribution-shifting augmentation, and the other using block permutation. In each case, we fit a logistic regression model with balanced class weights to the resulting CSI scores to find an effective threshold for OOD detection in our dataset. The results of this experiment are presented in Table 1. As we suspected, it is clear from these results that block permutation is a more appropriate distribution-shifting augmentation for our dataset than rotation.

5.2. Tuning λ hyperparameter

As discussed in Section 3.1, the CSI loss function is a weighted sum of two terms: $\mathcal{L}_{\text{con-SI}}$ which measures how well the network is satisfying the contrasting representation objective, and $\mathcal{L}_{\text{cls-SI}}$ which measures how well it is performing the auxiliary task of classifying shifted instances of the query image. We repeat this loss function below for convenience

$$\mathcal{L}_{\text{CSI}} = \lambda \mathcal{L}_{\text{con-SI}} + \mathcal{L}_{\text{cls-SI}} .$$

The authors set the hyperparameter $\lambda = 1$ in their experiments. However, by changing its magnitude, we can increase or decrease the importance of the contrastive representation objective relative to the classifying shifted instances objective. Of particular interest is the case where λ is large enough to make the auxiliary objective irrelevant.

To explore the impact of this parameter on OOD detection performance in our bollworms dataset, we first trained 20 CSI models using evenly spaced λ values between 0.1 and 2 (intentionally centered around the authors choice of 1). For each CSI model, we fit a logistic regression model with balanced class weights to the resulting CSI scores to find an effective threshold for OOD detection, and recorded their macro-averaged F1 scores. The results of this experiment are shown in Figure 4. Small values of λ de-emphasize the primary contrasting representation objective, so it makes sense that the OOD detection performance suffers toward the left end of the plot. It would also appear that $\lambda = 1$ is not optimal on our dataset, and that larger λ values should be explored.

To that end, we ran an additional coarse-grained tuning

Shifting Augmentation	AUROC	Macro-F1	Acc. (ID)	Acc. (OOD)
Rotation	0.91	0.77	0.92	0.70
Permutation	0.95	0.83	0.96	0.73

Table 1. Results from distribution-shifting augmentations experiment described in Section 5.1.

experiment where we trained 11 CSI models using evenly spaced λ values between 0 and 5. To reduce noise, we trained each model 3 times from random initializations and computed their average macro-F1 scores; the results are shown in Figure 5. Here we see an interesting phenomenon: the OOD detection performance on our dataset appears to plateau around a λ value of 2. Increasing the value of λ beyond this does not impact the macro-F1 score. If the auxiliary task of classifying shifted instances were important, then we would expect the performance to decrease once λ becomes large enough to overshadow its contribution to the loss function. The fact that we don't see this behavior suggests that the auxiliary objective may not be necessary at all in our case.

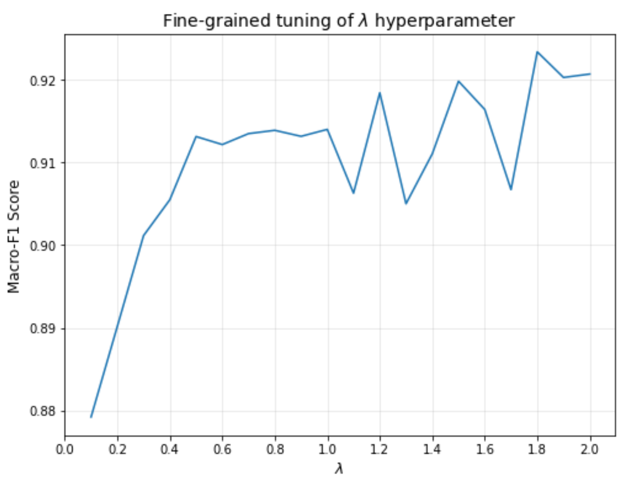

Figure 4. Results from fine-grained hyperparameter tuning experiment described in Section 5.2. Each data point in this plot represents a single measurement.

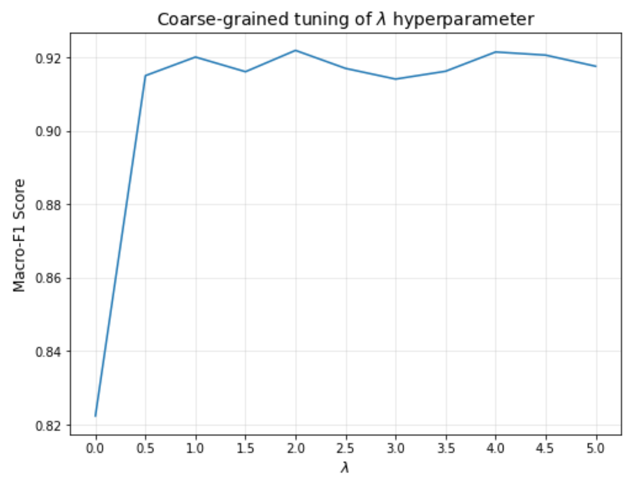

Figure 5. Results from coarse-grained hyperparameter tuning experiment described in Section 5.2. Each data point in this plot represents the average of three measurements.

5.3. CSI score with non-contrastive representation

As a comparison to the results obtained in our experiments using the full CSI framework, we also considered the performance of a logistic classifier trained on the score function of the CSI model applied to the latent embeddings generated by a convolutional autoencoder (CAE). Our experimentations in this area were motivated by an interest in examining how well the CSI score function works on a set of non-contrastive latent vector representations that we obtained using a conventional CAE. Applying the CSI score function to a different set of latent vector embedding generated by a non-contrastive learning approach helps us better understand the importance the score function has in generating the strong performance of the CSI model.

For our analysis, we developed a customized CAE model and trained it to compress ID images by specifically training on train ID images. Deep CAE models can be difficult to train due to vanishing gradients, and are particularly susceptible to getting stuck in sub-optimal regions of the loss function (Goodfellow et al., 2016). To mitigate these issues, we implemented a greedy, layer-wise training procedure (Bengio et al., 2006). With this procedure, sometimes referred to as "unsupervised pretraining", we add pairs of compression and decompression layers to the network one at a time, training it from the outside-in. The results are shown in Figure 6. This staged training procedure also gives us the flexibility to experiment with latent embeddings of varying sizes since the CAE at each encoder/decoder paired depth is itself a standalone CAE. Therefore, in the process of training a CAE with 8 encoder blocks, we have essentially trained 8 auto-encoders with differing latent dimension sizes.

We then evaluate the utility of applying the score function used by the CSI model to the task of OOD detection using the latent representations of input images from various layers in the CAE by applying logistic regression to the aggregate set of image scores for each ID and OOD image in the training set. This score function is repeated here for convenience:

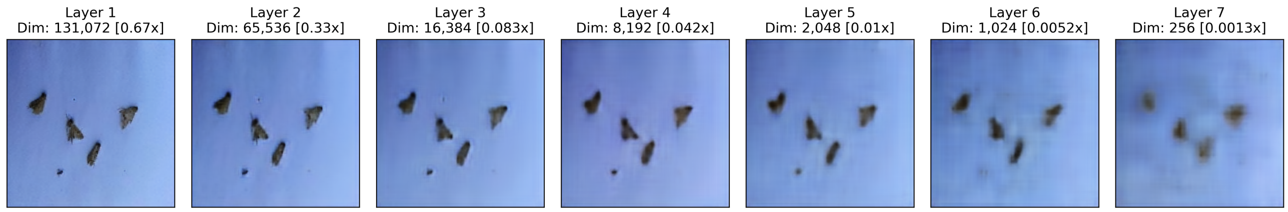

Figure 6. Results from greedy, layer-wise training of our CAE on an example ID image from the validation set. Above each image, we provide the (flattened) latent-space dimension and the corresponding image compression factor. The network has 8 layers, but layer 0 is omitted from the graphic because its latent-space representation is larger than the original image.

$$s_{\text{con}}(x; \{x_m\}) := \max_m \text{sim}(z(x_m), z(x)) \cdot \|z(x)\| \,.$$

The results at each layer are shown in in in Figure 7. Encoder blocks 0 and 1 were not included in our analysis since they were not viewed as producing sufficiently high image compression. We find that this score function formulation performs well for the task of OOD detection even on the CAE latent representations for layers that are less deep with a maximal performance at encoder block 3 as evidenced by its validation set accuracy in each class. This demonstrates the effectiveness of the CSI score function to meaningfully transform latent embedding vectors into largely distinct distributions for ID and OOD images. We regard the logistic regression model trained on the CSI score function applied to the embeddings of encoder block 3 as our best model and find that on the test set of bollworm images it is able to achieve an AUROC of 0.91, a macro F1 of 0.73, an ID accuracy of 0.85 and an OOD accuracy of 0.83. This minimal out-of-sample performance decay evidences the robustness of this formulation.

6. Discussion

In our analysis we find that contrastive learning (CSI) applied to the bollworms pest management image data set achieves competitive results out-of-the-box using the model configuration provided by the authors. We see this as a strong indication that this technique warrants further exploration and could perform even better for this specific task with a dedicated hyper-parameter search and the potential addition of other distribution-maintaining and distribution-altering permutations. We find that adjusting the λ tuning parameter on the CSI loss function does not impact the performance of the model beyond a value of 2, thus indicating that the auxiliary task of classifying shifted instances of the query images is minimally important to OOD detection in this application. In our experiments, we also find that the score function used by the CSI model is also a useful technique for transforming latent image embeddings into minimally overlapping ID and OOD score distributions even for non-contrastive learning encoder models. However, the best performance we find comes when a contrastive learning embedding is paired with the score function analyzed in our experiments.

7. Broader Impact

Improving the robustness of the app through accurate out-of-domain detection will improve the overall user experience, increase user confidence in the app, and increase mobile app user retention, thus amplifying the effectiveness of Wadhwani AI's solution to the pest management problem. Helping farmers protect their crops from pests will naturally improve their crop yield, reduce their usage of (and exposure to) toxic pesticides, and will lead to better mental health outcomes by alleviating unnecessary stressors and anxiety. Since cotton pest infestation is a common issue that farmers around the world are facing, a further social benefit could be a scalable solution that would be able to help farmers around the world protect their crops and increase farmland productivity. Our work could also potentially serve as a useful reference for the broader machine learning community by demonstrating how a solution to the out-of domain detection problem has been deployed in a social good setting.

One potential negative social impact from automating cotton farm pest detection that we can foresee is its threat to the jobs of farm extension workers who are not technologically savvy, but were hired to deal with the pests. A more effective version and widespread adoption of this app could create short-term unemployment within the farming community. If our OOD detection model is not effective, it can also lead to an overuse of pesticides. Given that a false negative is more concerning than a false positive in the case of bollworms, farmers would be incentivized to use pesticides whenever possible to prevent future infestation and mitigate the risk of losing their season's crops. A high usage of pesticides can have negative effects on the local biome, especially for organisms living in the soil. It can also potentially lead to contamination of drinking water and increased pest resistance overtime rendering them less and less effective for future seasons. There could also potentially be the risk of

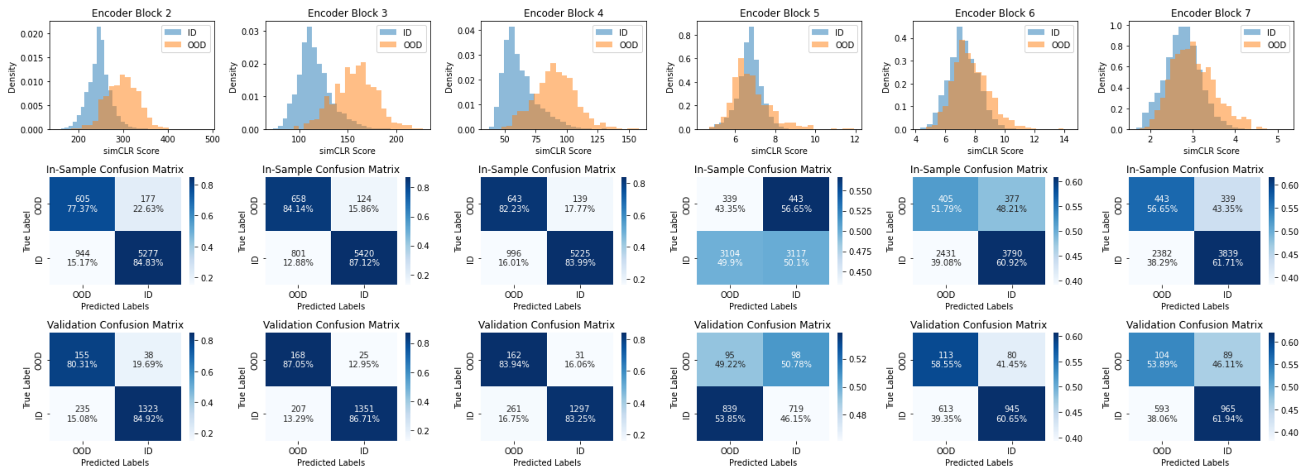

Figure 7. Results from applying class-weighted linear regression to the score values produced by the CSI model score function applied to the latent embeddings of a CAE across varying sizes of dimensional reduction.

this technology being maliciously exploited by hackers to manipulate pesticides usage on cotton farms.

References

Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19, 2006.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PMLR, 2020.

Dalmia, A., White, J., Chaurasia, A., Agarwal, V., Jain, R., Vora, D., Dhame, B., Dharmaraju, R., and Panicker, R. Pest management in cotton farms: An ai-system case study from the global south. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3119–3127, 2020.

Golan, I. and El-Yaniv, R. Deep anomaly detection using geometric transformations. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pp. 9781–9791, Red Hook, NY, USA, 2018. Curran Associates Inc.

Goodfellow, I., Bengio, Y., and Courville, A. *Deep learning*. MIT press, 2016.

Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., Dietterich, T. G., and Müller, K.-R. A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE*, 109(5):756–795, 2021.

Tack, J., Mo, S., Jeong, J., and Shin, J. CSI: Novelty detection via contrastive learning on distributionally shifted instances. *Advances in neural information processing systems*, 33:11839–11852, 2020.

White, J., Madaan, P., Shenoy, N., Agnihotri, A., Sharma, M., and Doshi, J. Wadhwani AI pest management open data. `https://github.com/WadhwaniAI/pest-management-opendata`, 2022a.

White, J., Madaan, P., Shenoy, N., Agnihotri, A., Sharma, M., and Doshi, J. A case for rejection in low resource ML deployment. *arXiv preprint arXiv:2208.06359*, 2022b.

Yang, J., Zhou, K., Li, Y., and Liu, Z. Generalized out-of-distribution detection: A survey. *arXiv preprint arXiv:2110.11334*, 2021.