Exploiting Unlabeled Data to Improve Detection of Visual Anomalies in Soft Fruits

Taeyeong Choi and Xin Liu

Department of Computer Science, University of California, Davis
1 Shields Ave, Davis, CA 95616, USA
{taechoi, xinliu}@ucdavis.edu

Abstract

Recently, self-supervised learning methods have been proposed to learn a useful representation for visual detection of anomalous, unhealthy crops while a neural network classifies augmented images of normal instances, which are relatively easy-to-obtain. Their pipelines are largely designed within the one-class classification paradigm, in which training samples are *all* of normal (negative) class, considering the severe scarcity of anomalous (positive) observations in realistic scenarios. In this paper, we study whether this "homogeneity" of training set is necessary to boost up the performance of learned detector because otherwise "unlabeled" data newly gathered from the field could simply be utilized during training without the need of expensive human annotation. To be specific, we first explore the scenarios treating every unlabeled instance as a normal one as the proportion of anomalous samples in the unlabeled set varies. Also, we introduce an *iterative* training procedure for "negative-unlabeled" learning, in which the unlabeled data are incrementally labeled based on predictions to train an one-class classifier with samples regarded to potentially be normal. Our experiments use CH-Rand—a state-of-the-art method for learning useful representations for anomaly detection from fruit images—on the Riseholme-2021 dataset, which includes a number of healthy and unhealthy strawberry images collected under realistic conditions. Specifically, the results show that using an unlabeled set as normal data can lead to the 8.7% performance improvement without any effort for labeling, though 4% in the set are of anomalous strawberry. In addition, our iterative training can benefit trained anomaly detectors by automatically filtering out unlabeled anomalies to reduce the overall anomaly ratio in the unlabeled data from 6% to 4.3%.

Introduction

Automated detection of unhealthy crops is crucial to realize precision agriculture, in which specialized treatments can be selectively applied to targeted individuals so as to not only increase overall productivity but also save available resources and protect the natural environments. Deep learning methods thus have been actively studied to adapt their great success in computer vision and field robotics to practical applications to natural outdoor environments in agriculture (Salazar-Gomez et al. 2022; Gao et al. 2021). In particular, one-class classification has been widely adopted as a useful approach for building anomaly detectors because of its realistic assumption that those can be generated only from observations of normal class (e.g., healthy fruits) while samples of anomalous class (e.g., unhealthy fruits) are too rare to be available during training (Choi et al. 2022).

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

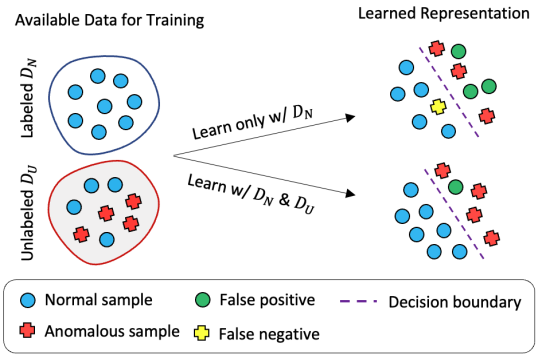


Figure 1: Hypothetical scenario in which an unlabeled dataset $\mathcal{D}_U$ can benefit learning a better representation for classification between normal and anomalous samples.

In this paper, we hypothesize that as illustrated in Fig. 1, such anomaly detectors could be *enhanced* by involving "unlabeled" yet easy-to-obtain datasets in training, although some anomalous (positive) samples may be contained along with normal (negative) instances. Inspired by (Lee et al. 2022), we first evaluate detection performance under various levels of anomaly ratio in the unlabeled set to investigate if simply treating them as normal data would tend to deteriorate the models. Also, we propose an *iterative* algorithm to systematically select unlabeled samples to utilize for subsequent training based on estimated anomaly scores. A relevant approach is positive-unlabeled learning in (Mu et al. 2021; Zhang et al. 2017). However, we adopt an one-class classifier instead of a binary model, and our task is called "negative-unlabeled" (NU) learning due to the anomalous class being of positive target.

Specifically, we use Channel Randomisation (CH-Rand) (Choi et al. 2022)—a state-of-the-art method for learning useful representations for anomaly detection from fruit images—on the Riseholme-2021 dataset, which includes a number of healthy and unhealthy strawberry im-

ages collected under realistic conditions. To the best of our knowledge, we are the first demonstrating classification of anomalies with agricultural data for NU learning. In particular, our study offers crucial perspectives from empirical results in order to fully exploit unlabeled datasets:

- Simple combination of unlabeled data into the training set of normal class can outperform the model not using it, even when the anomaly ratio is non-zero.
- Negative impact from a higher anomaly ratio can be mitigated by providing a larger size of unlabeled set.
- Anomaly ratio can be reduced by repeatedly selecting samples predicted to be normal, which benefits a final classifier.

Preliminary: Channel Randomisation

In this section, we briefly review the CH-Rand technique, which was proposed to learn useful representations of visual anomalies in fruits only from normal examples (Choi et al. 2022). Unlike other state-of-the-art methods such as CutPaste (Bergmann et al. 2021) designed to identify structural normality, CH-Rand is to focus more on patterns of color for agricultural applications since those generally change dramatically depending on the health status of crops.

In particular, Choi et al. (2022) have discovered that informative representations g_θ can be obtained within a neural network f_Θ during the *pretext* task shown in Fig. 2—i.e., it is to distinguish original images $\mathcal{I} \in \mathbb{R}^{W\times H\times C}$ from channel-randomized ones $\mathcal{A} \in \mathbb{R}^{W\times H\times C}$, where W and H are the width and the height, respectively, and C denotes the number of channels, which is set to 3 in the case of RGB space.

An application of CH-Rand can be regarded as adopting a random permutation of colour channels with a possibility of repetition over the entire image $\mathcal{I}$. More formally, for a new input $\mathcal{I}$, an arbitrary function $\pi : \chi \rightarrow \chi'$ is drawn for a random permutation, where $\chi = \{1, 2, ..., C\}$ is the set of channel indices, and $\chi' \in \mathcal{P}(\chi) \setminus \varnothing$, as $\mathcal{P}$ denotes the powerset of input. The function π then determines each pixel $a^c_{w,h}$ in $\mathcal{A}$ as follows:

$$a^c_{w,h} = i^{\pi(c)}_{w,h}. \tag{1}$$

Note that the *same* π is applied for every w, h, and c to keep the fixed channel arrangement over the image. In fact, we avoid the result to be $\mathcal{A} = \mathcal{I}$ by continuing drawing a new π until $\exists c \in \chi, c \neq \pi(c)$. In other words, 26 possible channel sequences can be generated by this augmentation for a typical 3-channel color space, such as RGB. Several examples of augmentation are visualized in Fig. 3

For representation learning, a classifier is employed to learn to infer if input images contain any artefacts in color generated by CH-Rand. As in (Li et al. 2021; Gidaris, Singh, and Komodakis 2018), the loss function is set up as below to train a deep neural network f_Θ with images $\mathcal{I}$ from a training dataset $\mathcal{D}_N$:

$$\mathcal{L} = \mathbb{E}_{\mathcal{I}\in\mathcal{D}_N}\Big[H(f_\Theta(\mathcal{I}), 0) + H(f_\Theta(CHR(\mathcal{I})), 1)\Big], \tag{2}$$

where CHR returns the output of CH-Rand, and H is the binary cross entropy function to consider prediction errors.

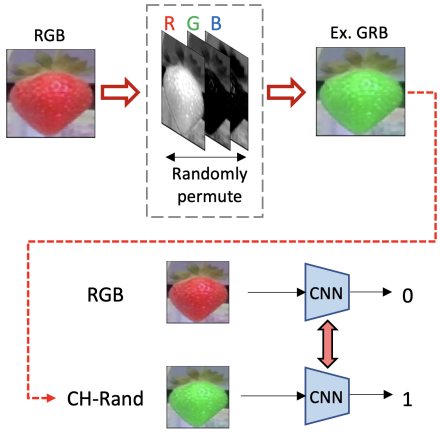


Figure 2: Pretext task, in which a convolutional neural network (CNN) is set to learn classification between original images in RGB and channel-randomized ones (e.g., GRB).

Choi et al. (2022) suggested taking a random batch $\mathcal{D}' \subseteq \mathcal{D}_N$ at each iteration applying CH-Rand only to a half of it.

Anomaly Score Prediction

According to (Choi et al. 2022), once a neural network f_Θ has been trained, the outputs of an intermediate layer are employed as feature representations g_θ of input images. Similarly to (Perera and Patel 2019), the anomaly score s for an input image $\mathcal{I}'$ can then be estimated using the average Euclidean distance to the k nearest neighbors $\mathcal{M} \subseteq \mathcal{D}_N$ on the space of g_θ—i.e., $s(\mathcal{I}') = (1/k)\sum_{\mathcal{I}\in\mathcal{M}} \delta\big(g_\theta(\mathcal{I}), g_\theta(\mathcal{I}')\big)$, where δ is the function of Euclidean distance. Hence, $\mathcal{I}'$ is predicted to be either anomalous if s is larger than a certain threshold τ or normal otherwise.

Problem Description

Instead of one-class classification, we here explore "NU-learning" (Ding et al. 2022) scenarios, in which an unlabeled dataset is available along with a labeled set of normal class (cf. Fig. 1). To be specific, while the data set D_N is known to be of normal class (e.g., healthy crop), we also assume that a disjoint set D_U is available, but samples therein are *not* annotated at all. In other words, D_U may contain some instances of anomalous class (e.g., diseased crop), and so, simply involving it in training of one-class classifier might cause an *unexpected* impact on the overall performance. In this work, we thus investigate the utility of such unlabeled datasets in the NU-learning setting and discuss a useful strategy to maximize the positive effect from using those.

For empirical exploration, we mainly use CH-Rand to build detectors of anomalous images of soft fruit (i.e., strawberry). More formally, our goal is to gain useful representations g_θ from training data Λ, which is the union of $\mathcal{D}_N$ and $\mathcal{D}_U$ Therefore, our new loss function can be reformulated from Equation 2 by replacing D_N with Λ:

$$\mathcal{L} = \mathbb{E}_{\mathcal{I}\in\Lambda}\Big[H(f_\Theta(\mathcal{I}), 0) + H(f_\Theta(CHR(\mathcal{I})), 1)\Big]. \tag{3}$$

Algorithm 1: Iterative NU Learning

Input: $\mathcal{D}_N, \mathcal{D}_U, m$
Output: Classifier f_Θ

$\mathcal{D}'_N \leftarrow \mathcal{D}_N$
while $|\mathcal{D}_U| \geq m$ **do**
  Initialize f_Θ
  Train f_Θ on $\mathcal{D}'_N$ using CH-Rand
  Compute anomaly score s_j for each $\mathcal{I}_j \in \mathcal{D}_U$
  $\mathcal{D}'_U \leftarrow \{\mathcal{I}_t : s_t \leq m\text{-th lowest socre } s^*\}$
  $\mathcal{D}'_N \leftarrow \mathcal{D}'_N \cup \mathcal{D}'_U$
  $\mathcal{D}_U \leftarrow \mathcal{D}_U - \mathcal{D}'_U$
end while
return f_Θ

In experiments, various conditions on $\mathcal{D}_U$ are considered while learning representations g_θ to examine the reliability of anomaly scores computed based on them.

Iterative Negative-Unlabeled Learning

Inspired by (Mu et al. 2021), we build an iterative framework to gradually merge some of informative samples from the unlabeled set $\mathcal{D}_U$ into the labeled $\mathcal{D}_N$. In particular, our approach is to prioritize the ones predicted to be most likely to be of normal, negative class so that the final selection of data can be utilized for training a one-class classifier.

Algorithm 1 shows that initially only the labeled dataset $\mathcal{D}_N$ is used for learning a representation in f_Θ with CH-Rand, but at each iteration, m images from $\mathcal{D}_U$ with the lowest anomaly scores are offered a pseudo-label as normal to move from $\mathcal{D}_U$ to the training set $\mathcal{D}'_N$ for the next iteration. This process repeats until at least m images remain in the unlabeled set $\mathcal{D}_U$.

An alternative approach could be to use such an iterative procedure to gather both positive and negative samples from the unlabeled dataset to learn a binary classifier (Zhang et al. 2017). However, the selection size m' for the positive class needs to be carefully chosen because anomaly ratios are usually unknown. Furthermore, the binary classifier might suffer from the severe class imbalance problem if the ratio is significantly small, which would motivate to use an one-class classifier.

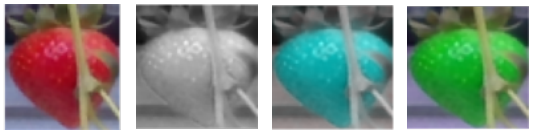
Experiments

Here, we first describe technical information of experimental configurations. To be specific, the following sections offer the details of Riseholme-2021 dataset, which is used throughout the experiments, as well as the network structures and hyperparameters for training sessions. Primary results of each experiment are then discussed.

Riseholme-2021 Dataset

For experiments, we utilize the Riseholme-2021 dataset[1] since it contains several thousands of strawberry images under three normal categories—such as `Ripe`, `Unripe`, and `Occluded`—and `Anomalous` in natural environments.

[1]https://github.com/ctyeong/Riseholme-2021



(a) (b) (c) (d)

Figure 3: Examples of CH-Rand when applied to (a) RGB image of strawberry: (b) RRR, (c) GRR, and (d) GRG.

Similar to (Choi et al. 2022), we split the data into `Train`, `Val`, and `Test` sets, but the `Train` set consists of a labeled set D_N of normal classes and an unlabeled set D_U, where $|\mathcal{D}_N|$ is fixed to 300 while $|\mathcal{D}_U|$ can vary in each experiment. In addition, the `Test` set contains 674 images from all normal categories as well as 99 instances of anomalous class, and the `Val` set has 337 normal examples.

Furthermore, general steps for image preprocessing follow, in which each image is resized to 64×64 pixels, and conventional augmentations[2] including heuristic color jitters and horizontal/vertical flips are performed before application of CH-Rand. Also, each pixel value is normalized to be between -1 and 1.

Implementation Details

We adopt the same structure of a deep network as in (Choi et al. 2022) using the publicly available code[3]. To be specific, the network has 5 ConvLayers and 2 DenseLayers, where the number of 3×3 convolutional filters increments at each layer—i.e., $64, 128, 256, 512$, and 512—and the DenseLayers employ 256 and 1 output nodes, respectively. Also, the output from each ConvLayer is processed by a BatchNorm layer and a 2×2 MaxPool layer, and every activation is computed with the LeakyReLU function except the sigmoid function at the last DenseLayer.

In particular, the learned representation g is taken from the outputs at the first DenseLayer, on which anomaly scores can be estimated with $k = 1$ nearest neighbor in training data $\mathcal{D}_N$ since $k = 1$ was found to be most effective as all normal categories are considered (Choi et al. 2022). Note here that the unlabeled set $\mathcal{D}_U$ is not considered for distance calculation to avoid the case in which the distance to its anomalies may lead to inaccurate prediction.

Evaluation Metrics

For evaluation, we compute the Area Under the Curve (AUC) of the Precision-Recall (PR) curve to take into account the highly skewed class distribution in the `Test` set (674 normal vs. 99 anomalous) (Davis and Goadrich 2006). Also, every resulting score here is the average of three separate runs to mitigate the stochastic effects from training neural networks. Moreover, each run saves the model achieving the highest validation accuracy until either training is performed for 1.2K epochs or a validation accuracy of $> .95$ has been achieved in average

[2]https://pytorch.org/vision/stable/transforms.html
[3]https://github.com/ctyeong/CH-Rand

Δ	It. 1	It. 2	It. 3	It. 4	It. 5	Total
450	0.3 ± 0.6	2.0 ± 1.0	5.3 ± 1.5	9.7 ± 2.1	N/A	17.3
900	2.7 ± 0.6	4.0 ± 1.0	5.0 ± 2.6	9.0 ± 1.0	13.7 ± 0.6	34.3

Table 1: Numbers of anomalous samples chosen from $\mathcal{D}'_U$ at iteration i to be in $\mathcal{D}'_L$ at the next iteration as $\alpha = 6\%$.

for the past five checkpoints while each checkpoint is performed every 30 epochs.

Results

This section discusses observed impacts on final performance of anomaly detector under different settings.

Various Compositions of $\mathcal{D}_U$: Here, we examine the impact on the overall performance depending on the configuration of the unlabeled data set $\mathcal{D}_U$. To be specific, we here control two parameters:

- Size of $\mathcal{D}_U$: $\Delta \in \{0, 450, 900\}$
- Ratio of anomalies in $\mathcal{D}_U$: $\alpha = \{0\%, 4\%, 6\%\}$

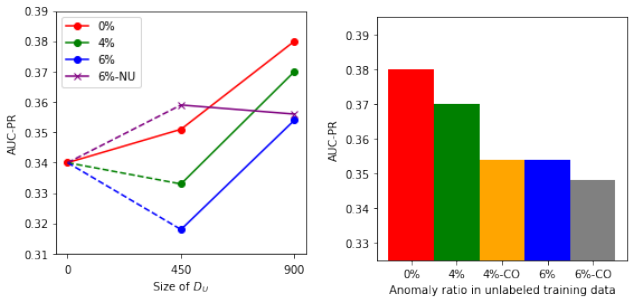
In particular, the range of α is based on the anomaly ratio of 4.3% in the whole Riseholme-2021 dataset.

Figure 4a provides average performance for each controlled setting. Specifically, for $\alpha = 0\%$, as expected, more data in $\mathcal{D}_U$ can provide improved detection performance since the neural network can gain a better quality representation from diverse normal samples and their channel randomisation. In particular, the improvement can be enhanced due to no interference of anomalies in the set.

On the other hand, any case with some level of anomaly ratio offers poor performance particularly when $\Delta = 450$, which is even worse than the case not using it ($\Delta = 0$). This implies that even though $\mathcal{D}_U$ fed more than 422 additional normal samples to the neural network, those small percentages of anomalies significantly hindered it from learning an appropriate representation of normal and anomalous images. Consequently, the 6% ratio led to a worse quality representation than the case of 4% ratio.

However, a dramatic change has been observed in the performance for those non-zero-ratio models when more data (i.e., $\Delta = 900$) were involved. To be specific, 4% and 6% ratios lead to 8.7% and 3.9% performance improvements over the scenario only using $\mathcal{D}_N$ (i.e., $\Delta = 0$). Though those perform not as well as the model of $\alpha = 0\%$ using the same amount of data, the 4% ratio can show a 5.5% better result than the 0%-ratio model with $\Delta = 450$.

This successful result can be surprising since the number of involved anomalies increases while Δ becomes doubled because of the fixed ratio α. From this observation, we can discover that unlabeled datasets can be highly useful without any annotation, *if the total size of data is sufficiently large*. Although some anomalies can be present in proportion to interfere with representation learning, a much larger number of normal samples can compensate for the potential negative effect during training.



(a) (b)

Figure 4: (a): Detection performance under various values of Δ and α, and (b): Performance comparison at $\Delta = 900$ when samples in $\mathcal{D}_U$ are not used as the RGB examples.

Iterative NU Learning: We apply the iterative algorithm for NU learning to the case of 6% anomaly ratio. For $\Delta = 450$, $m = 100$ image samples were selected from $\mathcal{D}_U$ at each iteration whereas $m = 160$ instances were considered as $\Delta = 900$. That is, the former finally adds 400 images to the training set while the latter 800 instances.

Table 1 implies that every case eventually *reduced* the number of anomalies in the training set with the anomaly ratio around 4.3%. Also, the detector obtained from the filtered dataset has shown significant performance improvement. In particular, with $\Delta = 450$, the model performed 13% better than the case in which the unlabeled dataset was all simply merged into the training set. This result suggests that using such a iterative training method can be particularly beneficial when the size of available unlabled data is relatively small.

Weaker Learning from $\mathcal{D}_U$: One of the key reasons that the anomalies in $\mathcal{D}_U$ may decrease the quality of learned representation is that those samples can also be used as the RGB input to the neural network (cf. Fig. 2), which is actually designed to learn features for normal visuals. An alternative, safer method could thus be to feed the instances from $\mathcal{D}_U$ only by CH-Rand to avoid this issue but still be able to provide the signals of unnatural color patterns.

Figure 4b shows the AUC-PR at $\Delta = 900$ for comparative evaluation between the conventional learning and this "CH-Rand Only" (CO) approach. Interestingly, for each α, CO learning led a worse performance, which may be because the design for weaker learning signal can safely separate anomalies from learning of normal images, but normal instances $\mathcal{D}_U$ are *segregated* as well.

Conclusion & Future Work

With real strawberry images, we have shown that unlabeled datasets can be utilized to improve the performance of anomaly detectors. Our results imply that if the unlabeled set is sufficiently large, it can compensate for the negative impact from a relatively high ratio of anomaly in it. Also, we have introduced a iterative algorithm for negative-unlabeled learning to further boost up the detection performance.

References

Bergmann, P.; Batzner, K.; Fauser, M.; Sattlegger, D.; and Steger, C. 2021. The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection. *International Journal of Computer Vision*, 129(4): 1038–1059.

Choi, T.; Would, O.; Salazar-Gomez, A.; and Cielniak, G. 2022. Self-supervised Representation Learning for Reliable Robotic Monitoring of Fruit Anomalies. In *2022 International Conference on Robotics and Automation (ICRA)*, 2266–2272. IEEE.

Davis, J.; and Goadrich, M. 2006. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning*, 233–240.

Ding, Y.; Rao, A.; Song, H.; Willett, R.; and Hoffmann, H. H. 2022. NURD: Negative-Unlabeled Learning for Online Datacenter Straggler Prediction. *Proceedings of Machine Learning and Systems*, 4: 190–203.

Gao, J.; Westergaard, J. C.; Sundmark, E. H. R.; Bagge, M.; Liljeroth, E.; and Alexandersson, E. 2021. Automatic late blight lesion recognition and severity quantification based on field imagery of diverse potato genotypes by deep learning. *Knowledge-Based Systems*, 214: 106723.

Gidaris, S.; Singh, P.; and Komodakis, N. 2018. Unsupervised Representation Learning by Predicting Image Rotations. In *International Conference on Learning Representations*.

Lee, C.-Y.; Li, C.-L.; Yoon, J.; Sohn, K.; Arik, S.; and Pfister, T. 2022. Self-supervise, Refine, Repeat: Improving Unsupervised Anomaly Detection. *Transactions on Machine Learning Research (TMLR)*.

Li, C.-L.; Sohn, K.; Yoon, J.; and Pfister, T. 2021. CutPaste: Self-Supervised Learning for Anomaly Detection and Localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9664–9674.

Mu, H.; Sun, R.; Yuan, G.; and Shi, G. 2021. Positive unlabeled learning-based anomaly detection in videos. *International Journal of Intelligent Systems*, 36(8): 3767–3788.

Perera, P.; and Patel, V. M. 2019. Learning deep features for one-class classification. *IEEE Transactions on Image Processing*, 28(11): 5450–5463.

Salazar-Gomez, A.; Darbyshire, M.; Gao, J.; Sklar, E.; Parsons, S.; et al. 2022. Beyond mAP: Towards practical object detection for weed spraying in precision agriculture.

Zhang, J.; Wang, Z.; Yuan, J.; and Tan, Y.-P. 2017. Positive and unlabeled learning for anomaly detection with multi-features. In *Proceedings of the 25th ACM international conference on Multimedia*, 854–862.