DETECTION OF ANOMALOUS GRAPEVINE BERRIES USING ALL-CONVOLUTIONAL AUTOENCODERS

Laurenz Strothmann 1, Uwe Rascher1, Ribana Roscher2

1 Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Forschungszentrum Jülich, Germany
2 Institute of Geodesy and Geoinformation, University of Bonn, Germany

ABSTRACT
A regular monitoring of plants is inevitable to ensure an effective production and to reduce yield losses, for example, caused by different diseases. Infected plants show a visual effect shortly after inoculation. These effects can be understood as anomalies, which do not occur in healthy plant stocks. For automation of harvesting or spraying it is important to recognize anomalies to ensure an on-time reaction by the farmer or breeder. However, these anomalies differ largely in their appearance and a representative model is generally too complex to be learned. Our main objective is reconstruction-based anomaly detection by all-convolutional autoencoder (all-CAE), which combines convolutions with the architecture of an autoencoder (AE). To achieve our objective, we use an hourglass all-convolutional encoder-decoder architecture to create a highly compressed representation in the middle layer. Moreover, we compare different types of noise as regularizer. In our experiments, the method is tested on images of grapes acquired in a vineyard. We show that all-CAE are suitable for anomaly detection and that unnatural noise (salt) shows the best results.

Index Terms— Autoencoder, anomaly detection, grapevine

1. INTRODUCTION
Plant diseases are responsible for a major part of damage and yield losses in agriculture. An automatic detection of plant diseases has a high potential to reduce costs and negative environmental impacts. The grapevine production, for example, consumes about 60% of the yearly used fungicides in Europe [1]. Therefore, it is important to separate fungi-infected individuals in order to prevent a further spread. Furthermore, a combination of autonomous harvesting machine with sensors for an automated detection of diseased berries can ensure a high quality of the harvest.

A challenge for such a detection task is that plant individuals occur in many varieties in a single field and especially across different fields. Collecting representative training data for a supervised classification covering the whole range of variability is time-consuming and difficult. Thus, supervised learning methods are often limited to specific conditions and often suffer from sub-optimal conditions like overlapping objects, image degradation due to weather, and change of physiological state over time [2]. Anomaly detection states a suitable alternative for plant disease detection, since only the target class of healthy plants has to be modelled and data that does not fit to the model is assumed as anomaly [3, 4].

Recently, also deep learning (DL) approaches have been used for agricultural application [5]. So far, however, there are only a few works on anomaly detection. For example, [6] propose DeepAnomaly which uses background subtraction and pre-trained weights of state-of-the-art convolutional neural network (CNN) to identify anomalies.

Our contribution is the development of a novel method for anomaly detection called deep all-convolutional autoencoder (all-CAE). In our experiments, we explore all-CAEs with different types of noises and analyze their performance for the detection of anomalous berries in the vineyard.

2. DATA

Fig. 1. Example image patches from the data set showing healthy berries from the training set, and healthy and anomalous berries from the test set.
An AE is a neural network (NN) architecture that is capable of reducing the input dimension to a smaller representation and scale it back to the original input dimension. Thus, an AE can be described as an encoder function \( f \) with an input \( I \) that performs a stochastic mapping \( P_{\text{encoder}}(f(I) | I) \), followed by a decoder function \( g(f(I)) \), which maps back to the input with \( P_{\text{decoder}}(I | f(I)) \) [8].

In order to ensure that the network learns an useful representation in the middle layer and to prevent an identity mapping, various approaches have been presented. One approach is to constrain the output size of the encoder function, which forces the network to only learn the most salient features [8]. This dimensional reduction can be also understood as regularization, aiming for a high generalization ability of the network. A denoising AE is another way to avoid an identity mapping from input to output by adding noise to the input. Gaussian noise can be understood as natural noise, as it shifts the values into both directions of the spectral range, but the mean over larger areas remains the same [9]. Salt noise can be considered as turning-off information and thus as a less natural noise [9].

CAE combine AE and the concept of convolutions. We use an all-convolutional architecture, as proposed in [10], which consists of convolutional layers only (see Fig. 2). Instead of the commonly used max-pooling layer, an CAE uses convolutions with increased stride > 1 to accomplish downsampling, and transposed convolution as method for the decoder to increase the dimensions.

**Fig. 2.** Hourglass encoder-decoder architecture of all-CAE.

The network consists of convolutional layers only, where each connecting branch represents one convolutional layer. The downsampling is managed by strides > 1 and the upsampling is done by transposed convolution.

For anomaly detection, we follow a reconstruction-based approach. If a all-CAE learns a certain stochastic mapping and a sample is exposed to the trained network that does not belong to the trained distribution, it shows a higher loss than the one caused by a training sample. An increase can be interpreted as anomaly [11]. We use the following loss in this study:

\[
J(I, \hat{I}) = 2 \left( \frac{1}{N} \sum_{x=1}^{X} \sum_{y=1}^{Y} |I_{x,y} - \hat{I}_{x,y}| \right) + \left( \frac{1}{M} \sum_{x=S-1}^{X-S} \sum_{y=S-1}^{Y-S} \text{SSIM}(I_{x,y}, \hat{I}_{x,y}) \right) - 1,
\]

where \( I \) denotes the input image and \( \hat{I} \) denote the reconstructed image. The structural similarity index (SSIM) defines a distance between two image patches of size \((2S - 1) \times (2S - 1)\), where \( x \) and \( y \) define the pixel position in the image around which the image patch is defined [12]. The number of rows and columns are defined as \( X \) and \( Y \), respectively, the number of pixels in both images are denoted by \( N \), and \( M \) is the number of valid pixel positions inside the images without padding. We calculate the loss for each image band and use the average for further processing.

### 4. EXPERIMENTAL SETUP

#### 4.1. Data Set

We selected 300 patches of healthy berries of size 300 × 300, and selected 30 different patches with healthy berries and 30 patches with diseased damaged berries as anomalies (see Fig. 1). We chose random flipping for data augmentation of the training data. For the test and anomaly set, we did not use any augmentations.

#### 4.2. Compared Architectures

We compare three different all-CAE architectures for the task of anomaly detection: (1) all-CAE, (2) Gaussian denoising all-CAE, and (3) salt denoising all-CAE.

**All-convolutional autoencoder.** As the baseline architecture, we use a symmetric all-CAE with four convolutional encoder and decoder layers. We use \( \ell_1 \) regularization for both weights and biases. As optimizer, we use Adam [13] with a learning rate of 0.5 \( \cdot 10^{-4} \). Weights are initialized by the Glorot-uniform-initializer [14]. As loss, we use (1). The convolutional kernel size is 9 × 9 for every layer with a stride of 1-2-2-3 and a number of feature maps of 16-18-20-22. As activation function, we use leaky RELU. We use a mini-batch size of eight for training in order to regularize the network. The input size of our network is 200 × 200 × 3 pixels. Considering the input size, the network reduces the dimensions to 12 × 12 × 22, which is roughly 2.6% of the original data points. We carefully selected the hyperparameters to match our aim of reducing the output of the encoder.

**Gaussian denoising all-convolutional autoencoder.** We use a Gaussian distribution with zero mean and a standard deviation of 0.5. After applying noise, the spectral range of the
image is restricted to $[0, 1]$. We choose the same hyperparameters as for the noiseless all-CAE.

**Salt denoising all-convolutional autoencoder.** For this autoencoder, we select 30% of all pixels and set their values in all bands to zero.

### 4.3. Anomaly Detection of Patches

We train the network for 28 epochs with 750 steps per epoch and estimate the loss for all images in the test set. We compute histograms of the losses and calculate the area under curve of the receiver-operating-characteristic (ROC-AUC-score) for different thresholds for the losses.

### 4.4. Pixel-wise Anomaly Detection

We utilize a sliding window approach to evaluate the performance of a pixel-wise anomaly detection. For our study we use an image containing both anomalous berries and non-anomalies. The image borders are extended by 100 pixels using reflective padding. We apply our proposed all-CAE architectures and calculate the loss for each pixel position in the extended image in a sliding window manner.

### 5. RESULTS

Figure 3 shows a histogram of the losses computed for all non-anomaly (blue) and anomaly (red) test image patches. We can see a clear distinction between both groups for every all-CAE architecture. Despite some overlaps, the bulk part is concentrated around two peaks for all used architectures. Regarding the ROC-AUC-score and the number of overlaps, the salt denoising all-CAE shows the best results. It has a ROC-AUC-score of 0.95 and eleven overlaps. Whereas the Gaussian Denoising all-CAE has a slightly lower ROC-AUC-score, but also eleven overlaps. In comparison, the baseline all-CAE has a similar ROC-AUC-score of 0.94, but the most overlaps of 14.

As mentioned above, we can consider Gaussian noise as natural noise and salt noise as ‘turning-off’ features. If we interpret Gaussian noise as a moderate regularization and salt noise as a higher and more aggressive regularization to prevent identity mapping, we can observe a small tendency, that a higher regularization shows better results, as the ROC-AUC-score increases.

Figure 4 shows the results of the pixel-wise anomaly detection. All three architectures correctly show one bright hot spot in the center of the anomaly (250, 200). Furthermore, the bulk parts of the two berry clusters show significantly lower losses than the anomaly, which means that they were correctly classified as non-anomaly. Nevertheless, the bottom left corner (500, 550) and the top left corner have slightly increased values, which could be incorrectly classified as anomaly if a threshold is set too high. This indicates that non-berry objects show a higher loss than healthy berries. We assume that our training set is not yet representative enough for these objects, and that an extension by characteristic objects present in the vineyard will lead to an improvement of the results. Furthermore, if we look closer at the bottom left side, the berry at (100, 450) has a small cut and is damaged. This is neither recognized by the baseline all-CAE, nor by the denoising architectures. This can be contributed to the size of the input, because the anomaly is to insignificant for a $200 \times 200$ patch. A potential extension would by an all-CAEs that works on different scales.

In sum, all analyzed architectures show only slight differences, which is in accordance with the ROC-AUC-score in Figure 3. We conclude, that our networks are able to detect larger anomalies, whereas small anomalies are still suppressed by surrounding non-anomalies. A possible way to improve the detection of small anomalies is the additional use of a smaller input to assess different scales of anomalies.

### 6. CONCLUSION

Anomaly detection is a relevant task for agricultural data. We proposed a convolutional autoencoder architecture which learns a representation of the target class of healthy berries,
and is able to distinguish them from anomalous berries. We showed that adding noise has the tendency to improve the results, where salt noise showed the best performance. Furthermore, we showed that our approach works best for large anomalous structures. However, different scales of our input will be investigated in order to assess anomalies on smaller size. In addition, an automatic determination of an appropriate threshold will be developed, such as trade-off between sensitivity and specificity to improve the detection.

7. ACKNOWLEDGEMENT

We would like to thank Dr. Anna Kicherer and the Julius Kühn-Institut, Federal Research Centre for Cultivated Plants, Siebeldingen, Germany for providing the data, which was collected during the novosys project (FKZ031A349E). Furthermore, we thank the Bosch Rexroth AG for funding the project.

8. REFERENCES


