Exploration into Anomaly Detection on Environmental Images

Christopher Lee

Introduction

Argonne National Lab has developed a physical device consisting of a NX Xavier ARM SOC capable of GPU acceleration along with various sensors for imaging, temperature, rainfall, humidity, and other modular attachments to expand its capabilities. This device is officially referred to as a Sage Node and colloquially as a Waggle Node.

Currently, there are 124 nodes, with plans to install more. The majority are located across the United States and may reside in, or near rural regions with poor cellular infrastructure. Thus facilitating network bandwidth limitations (although it should be mentioned that we are exploring the usefulness of starlink as an alternative). As the number of nodes is expected to grow, and bandwidth is currently constrained, the idea is to execute AI computation at the edge and forward the final results back to a central server (called the beehive). At a high level, this workflow acts to compression the data at each time interval thereby shifting the balance from spatial to temporal.

One use case that exploits both imaging and edge computing while satisfying the constraints is automated anomaly detection. This is evermore relevant due to recent concern events of wildfires and floods. Investigation into anomaly detection for these nodes was also, in part, inspired by coincidently capturing a volcanic eruption on one of the sensors.

Dataset

The dataset consists of colored 2560 by 1920 images taken from the Sage/Waggle Node on the Big Island overlooking Hawai'i Volcanoes National Park. They are sampled every hour from 06/01/2024 at 8:00 am to 06/21/2024 at 8:00 pm. However, there may be occasional gaps in the data, where some images were not captured, stemming from known sensor reliability issues. Figure 1 depicts a normal state example of what the environment contains.

Figure 1: normal state example

Model

Exploration:

There were various popular techniques [1], However I've lightly explored the following: PCA with (MAE, MSE, Relative reconstruction errors, semantic segmentations), SIFTs, KNN, isolation forest, DinoV2 embeddings for similarity comparison, fully connected VAE and CNN VAE. The appendix figures contain a very small subset of example cases explored. PCA and VAE models had center focus due to training data only requiring examples of normal states. This idea works extremely well as affordances of real world anomalous examples are rare and unpredictable. Furthermore, since PCA and VAEs are only exposed to normal conditions during training; at inference they are incapable of reconstructing anomalous artifacts. Therefore the concept is that, feeding an anomalous image in, will remove the anomaly on reconstruction. Thereafter, append an effective metric for reconstruction error. KNN, and isolation forest were also experimented with as they detect outliers, and these outliers are considered the anomalies.

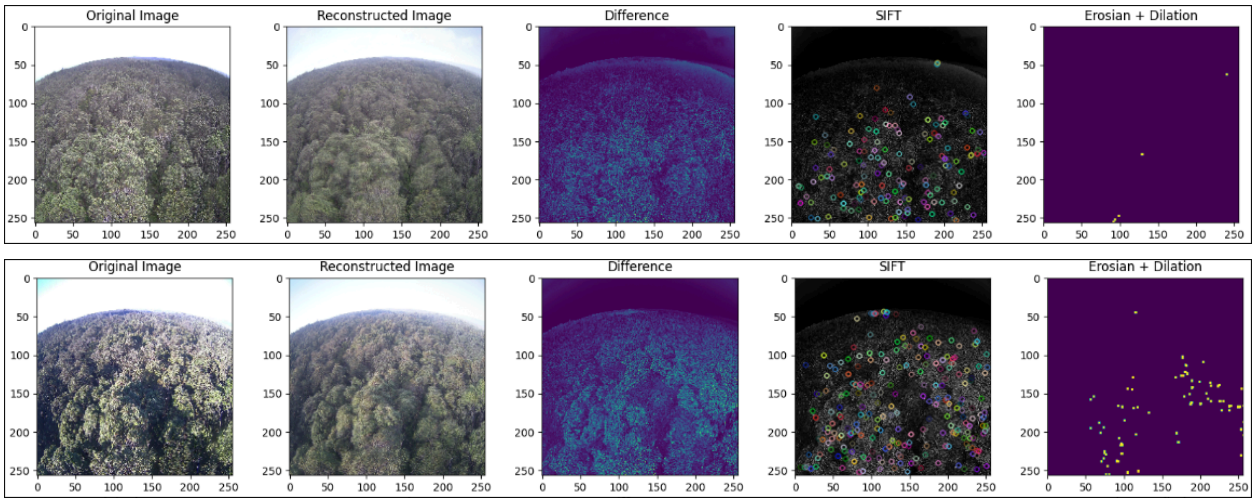

Figure 2: Pure PCA solution with difference map, and image post processing. Input images are of normal states

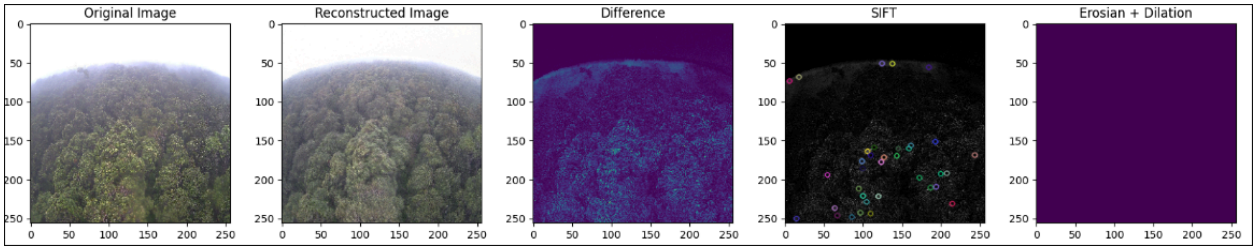

Figure 3: Input/Original image has been artificially augmented with a small indication of smoke in the top left region (approx. at coords 75, 75). Smoke not detected in difference map based on color.

Model Used:

While in theory, these techniques are promising; in practice they do not produce the desired results due to aleatoric uncertainty. The primary issue originates from the environment depicted in the dataset. Because the model captures pixel level variations, the trees (as seen in figure 2's difference map) are often flagged as anomalous due to their swaying from the wind. While it is true that reducing the resolution will mitigate this, as PCA reconstruction will not be as blurry, we are also reducing the chances of picking up on smaller anomalies. These anomalies could theoretically be a small patch of smoke starting to form (figure 3 original image top left). Additionally, smaller anomalies may have reduced impact/ weight for certain reconstruction metrics. Lastly, a simple solution may be masking out the tree's range of motion, which are the edges of the leaves, but it may also hide anomalous events.

Utilizing an idea mentioned in the spade paper [2], a better approach would be to first pass the data through Resnet and examine the outputs at the average pool layer to be used as a feature map. This feature map could then be passed to a KNN for discrimination. While this paper's [2] accompanying code performed well on the MVTecAD dataset [3], its performance suffered when applied to our task (more details in appendix). My alternative solution took the output from the last layer of Resnet18 in the fourth block as the feature map and those features were passed to PCA with all components. This approach ideally should capture higher level concepts in the latent space rather than purely pixel level deduction, which hopefully should improve invariance to anomaly size and irreverent environmental changes such as tree sway.

Features and Pre-Processing

As anomalies may manifest with colors, such as gray smoke over green trees. The images were kept in color (3 channels) and only resized to 224 by 224 to fit into a pretrained Resnet18, which handled the automated feature extraction.

Data Splits and Augmentation

There are a total of 473 images.

Train and Validation:

8 canonicalized images that capture a unique example of normal scenery in various lighting conditions are selected to be the training set. Another 8 images were selected to be the

validation set, these 8 images follow a similar distribution to the training set. As an example, for each night time image in the training set, there will be a very similar night time image in the validation set.

Test Sets:

The first (full) test set contained a total of 457 images with 342 examples of normal conditions and 115 other images depicting various anomalous events such as volcanic eruptions, rain droplets on the camera lenses or fog. The model was trained on the 8 normal images training set, validated with the 8 normal images validation set and evaluated on the test set.

A second (extreme) test set was created as some of the anomalous images could be vague. Such is the case with dense fog versus fog. Where does the distinction get drawn? Therefore this second test set, being identical to the full tests set in all regards besides the anomaly class. A subset of the anomaly class, filtered from the original test set, containing only images that are without a doubt abnormal, was created via manual selection. This new extreme anomalies set contains 81 samples of abnormal images and the same 342 normal examples.

Other:

A third, but non evaluated anomaly dataset was created through data augmentation to view the effects of the model of small glimmers and obvious signs of smoke. This dataset only contains four augmented images and is purely used for investigation of the model.

Hyperparameter Search Space and Optimization

Decisions were made based on understanding of the model architecture, then thereafter performance was evaluated with the test set. This is to avoid the pitfall of being influenced by the anomalous data in the test set. For Resnet18, I believe that the last layer of the fourth block provided high level features more akin to semantic features. Therefore that layer was used as the feature map.

Selecting the best reconstruction method from MSE, MAE, and Relative based on the best AUROC became a challenge due to needing examples of abnormal images which are not in the validation and training set. This is because the goal for this system is to be easily trainable without the need for anomalous images as they should be theoretically rare and unpredictable. To correct this, augmented images created via image editing could be applied to introduce various defects or objects in the image to simulate an anomaly. This could be smoke in a forest image or cones in a traffic image.

Regarding the threshold value for the reconstructed errors; the maximum error generated from the differences when running the validation set through the system was used as the threshold value.

Evaluation

The final model chosen for evaluation was Resnet18 (at last layer of fourth block) fed into a PCA and extracting reconstruction error.

Evaluation 1 Full Dataset:

Evaluation was based on AUROC to determine the model's performance. Because we cannot perform model selection using the test set, the AUROCs for each reconstruction method is listed as follows:

Reconstruction Method	AUROC	J Statistic Optimal Threshold	Threshold from Validation Set
MSE/ L2	0.842258	0.356212	0.409046
MAE/ L1	0.831325	0.384021	0.374319
Relative	0.871726	0.496795	0.542224

Table 1: AUROC and from reconstruction distances. J-Statistic optimal thresholds versus thresholds derived from validation set

Reconstruction Method	J-Stat Anomaly Accuracy	J-Stat Nominal Accuracy	Validation Set Anomaly Accuracy	Validation Set Nominal Accuracy	Anomaly Biased Threshold Nominal Accuracy
MSE/ L2	0.739130	0.915205	0.617391	0.976608	0.035088
MAE/ L1	0.678261	0.961988	0.686957	0.938596	0.011696
Relative	0.800000	0.877193	0.678261	0.956140	0.052632

Table 2: Anomaly and Nominal accuracy for J-statistic, and validation set. Anomaly biased represents the nominal accuracy based on a 100% perfect anomaly accuracy.

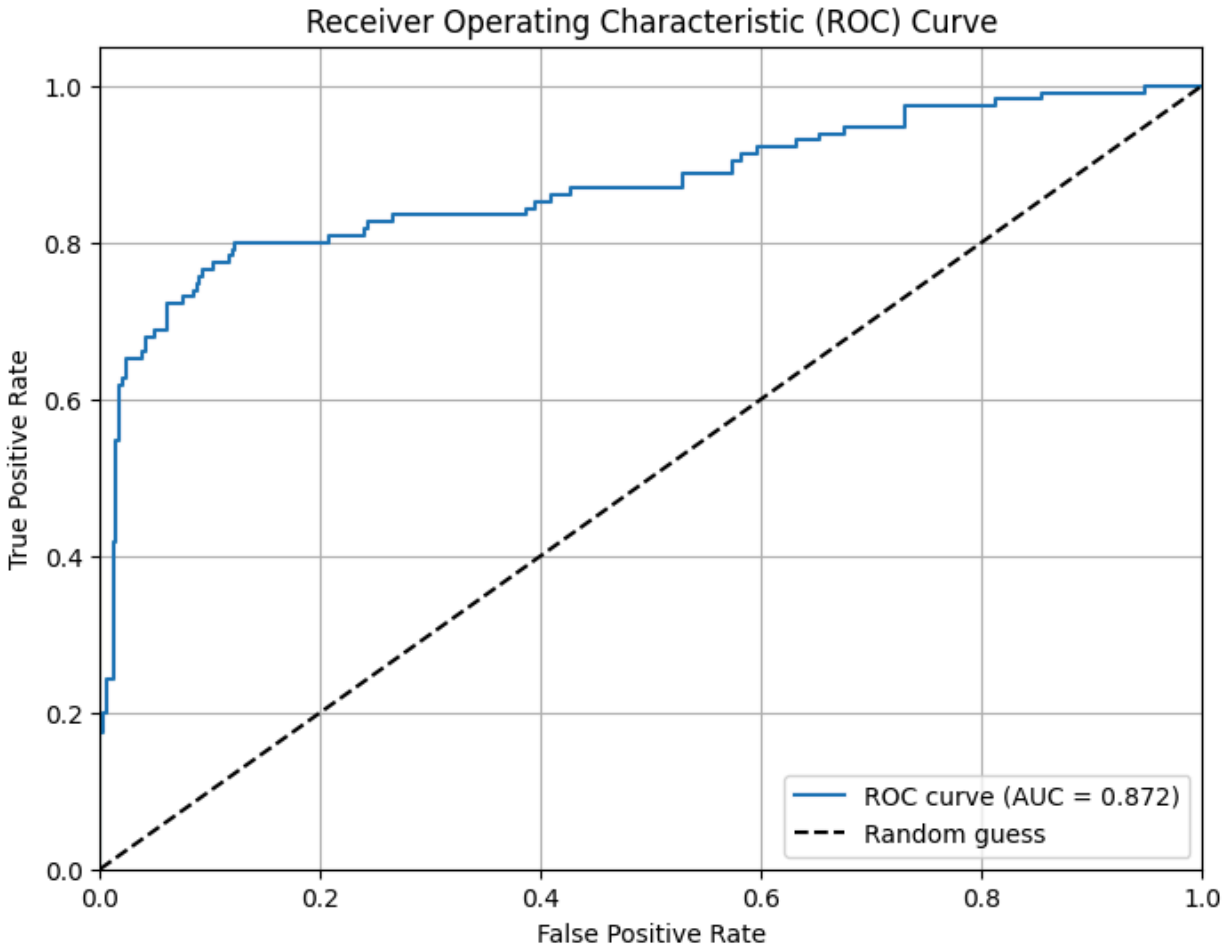

Figure 4: AUROC of Resnet18 + PCA w/ Relative Error model on first (full) test set

The AUROC with relative error performed the best, however as it was done on the test set, results remain inconclusive. While accuracies for the J statistic and validation set threshold values, performed decently, it heavily favored accurate detection of normal states, allowing abnormal states to suffer. For the application of first detection, it is critical to prefer a near perfect accuracy on anomaly detection while allowing leniency of false positives. Overall, when considering the model's performance based on AUROC, it is decent, but should be improved for critical applications.

Evaluation 2 Extreme Dataset:

Another evaluation was performed on the second (extreme) test set to evaluate performance on certain anomalies.

Reconstruction Method	AUROC	J Statistic Optimal Threshold	Threshold from Validation Set
MSE/ L2	0.892824	0.393074	0.409026

MAE/ L1	0.881741	0.384037	0.374300
Relative	0.929066	0.567767	0.542211

Table 3: AUROC and from reconstruction distances. J-Statistic optimal thresholds versus thresholds derived from validation set

Reconstruction Method	J-Stat Anomaly Accuracy	J-Stat Nominal Accuracy	Validation Set Anomaly Accuracy	Validation Set Nominal Accuracy	Anomaly Biased Threshold Nominal Accuracy
MSE/ L2	0.790123	0.964912	0.740741	0.976608	0.192982
MAE/ L1	0.777778	0.961988	0.777778	0.777778	0.190058
Relative	0.827160	0.976608	0.827160	0.956140	0.146199

Table 4: Anomaly and Nominal accuracy for J-statistic, and validation set. Anomaly biased represents the nominal accuracy based on a 100% perfect anomaly accuracy.

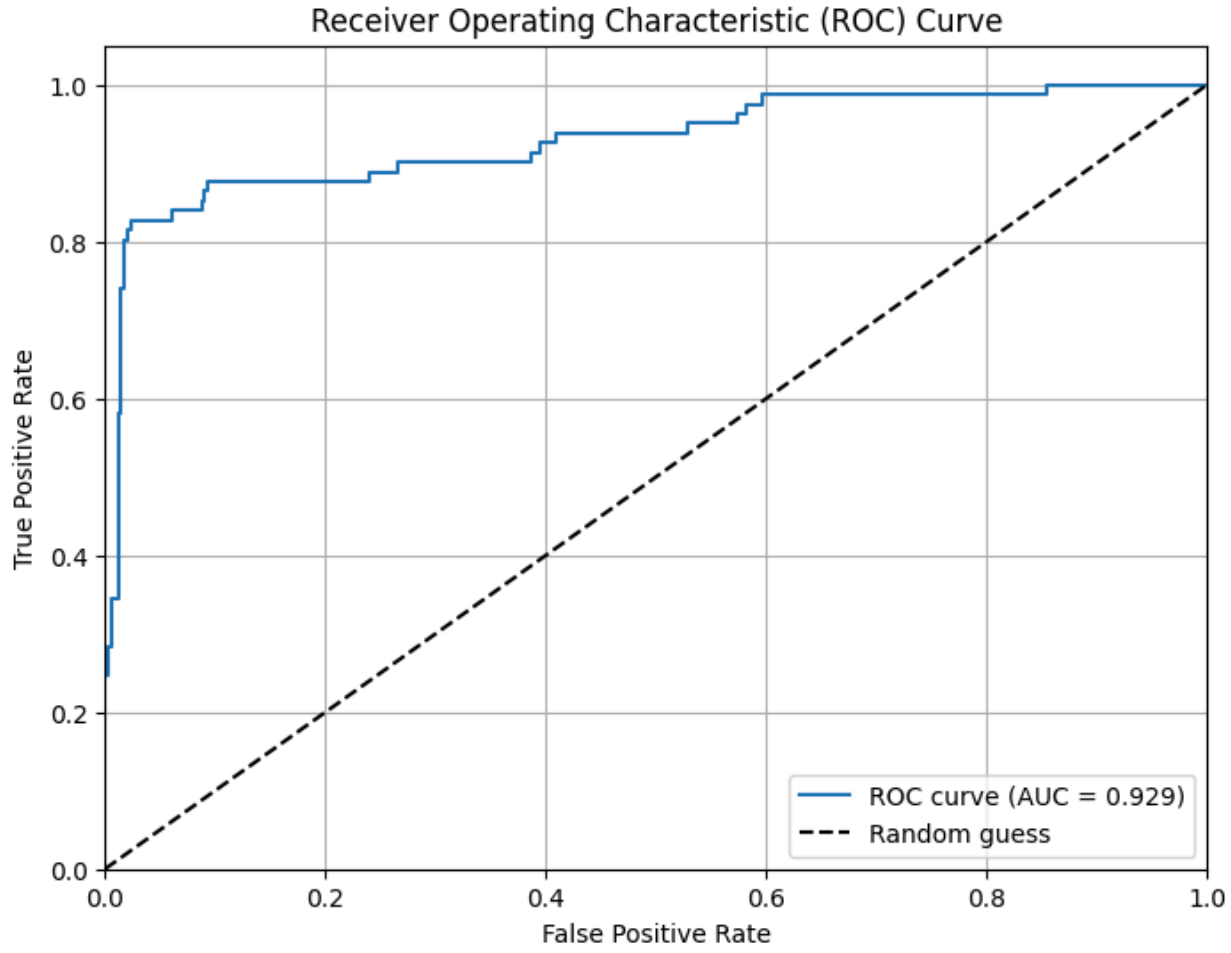

Figure 5: AUROC of Resnet18 + PCA w/ Relative Error model on second (extreme) test set

Based on the performance of the model on the extreme test set, the model struggles to distinguish data on boundary lines effectively. Deeper investigation into viewing the miss classified images could give additional clues for next steps to take to improve the model. Brief investigation on the augmented images of smoke suggests that the Restnet model detected some variations of smoke that the pure PCA solution failed to extract. As such there may be hope for an ensemble approach to boost performance.

Data Drifts

WIthout multiple test sets depicting various environments, it may be challenging to determine with certainty that this model can be applied to various other environments. The hope is that, if the model can ignore irrelevant movements such as tree sway or other pixel level changes and focus more on semantic differences; there is a strong possibility that this may fare well with environments that have variations in snow levels or beaches with active changes in wave heights and textures. Unfortunately, due to the lack of a saliency map, model explainability suffers, although there are solutions proposed in the literature [2].

Conclusion

PCA on images itself is great due to being able to calculate the difference from the reconstructed to the original, to generate a saliency map of pixel level area changes. However, pixel level variations do not necessarily capture semantics of the image in a meaningful way to discriminate between abnormal and normal. Therefore Resnet was introduced as a feature extraction step. Unfortunately, we lose some of the benefits of a pure PCA solution due to the feature obfuscation in the resnet's latent layers. Furthermore, the lack of explainability may lead to lackluster threshold selection, thereby introducing a point of human error.

As I've only started looking into anomaly detection within the past month, there is a lot more trial and error, literature reviewing, etc. that needs to be done to improve the pipeline. Thus far, the approach that feels most promising in accurately depicting anomalies should contain some information of latent features. These features should be invariant to the scale of anomalies and small environmental changes such as tree sway or camera shake. This semantic image representation should aid in distinguishing based on objects, textures, etc. and not solely at a pixel level. As I am unsatisfied with the model thus far, I will continue to find alternative solutions that can provide perfect anomaly accuracy with admissible normal accuracy.

Next Steps

- Reorganizing the dataset, creating additional datasets from a different node.
- Try single class classifiers like One-Class SVMs
- Try Resnet + VAE instead of Resnet + PCA
- Expand date range of data, get more data!

- Exploration of vision language models.
- It should be noted that brief experimentations with photoshopping images with anomalous activities such as some were briefly explored. Furthermore, utilizing a diffusion model with inpainting to generate more life-like representations may be a promising path. However, these augmented images were not included in the dataset.
- Split image into patches and analyze
- It may be beneficial to explore additional manual pre-processing steps prior to PCA. This could be various edge detections, blur, sharpening, color histogram, reducing color space from 255 to 16, etc.

Appendix

It should be noted that the spade paper [2] is primarily about how they would extract a saliency map based on a feature pyramid. Regarding performance, the spade paper's [2] Wide Resnet50 + KNN received a 0.661 AUROC on the (first) test set and a 0.799 AUROC on the second (extreme) test set. The model was slightly modified to ignore saliency map generation aspects and only utilized the first half the inference process.

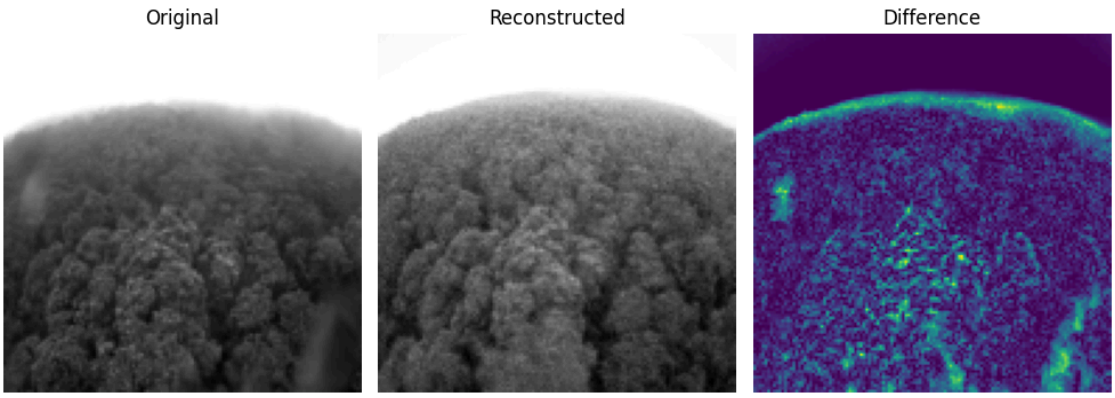

Figure A1: Results from FC VAE reconstruction with grayscale with 128 x 128 and 400 latent dim

Figure A2: Result from CNN VAE, reconstruction of seven input images

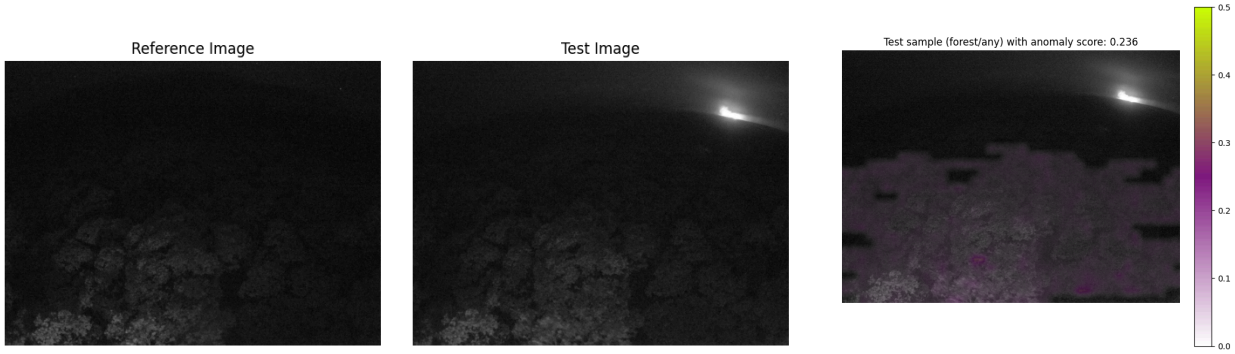

Figure A3: AnomalyDino saliency map (right most) generated from reference and test image

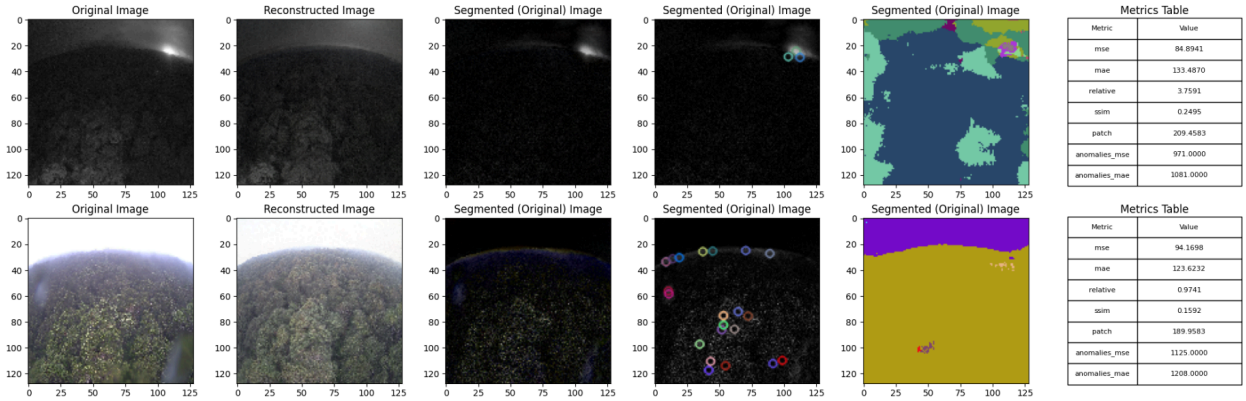

Figure A4: Depicting trials with semantic segmentation of differenced image to attempt to extract anomaly objects

References

[1] Yang, J., Xu, R., Qi, Z., & Shi, Y. (2021). Visual Anomaly Detection for Images: A Survey. *CoRR*, *abs/2109.13157*. Retrieved from https://arxiv.org/abs/2109.13157

[2] Cohen, N., & Hoshen, Y. (2020). Sub-Image Anomaly Detection with Deep Pyramid Correspondences. *CoRR*, *abs/2005.02357*. Retrieved from https://arxiv.org/abs/2005.02357

[3] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, Carsten Steger: **The MVTec Anomaly Detection Dataset: A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection**; in: *International Journal of Computer Vision* 129(4):1038-1059, 2021, DOI: 10.1007/s11263-020-01400-4. Retrieved from https://www.mvtec.com/company/research/datasets/mvtec-ad