The primary research inquiry within the experimental framework is to determine the feasibility of training a global model for detecting anomalies in all sheet metal parts with one threshold. The problem at hand is characterized by the presence of sheet metal parts with cracks or anomalies as well as the parts that do not have anomalies at all. Currently, the system only provides anomaly detection for specific parts that have anomalies, whereas the ideal objective is to extend its functionality to encompass all possible parts.

A potential approach to address this issue is to employ an unsupervised model that can be seamlessly applied to other parts, provided that the same conditions and setup are maintained. This includes factors such as consistent camera configurations, lighting conditions, environmental settings, color consistency, and similar parameters. As a result, the unsupervised model could be trained on subsets of 60 or 70% of parts with anomalies in order to get the average threshold and then make an evaluation of the left parts that were not used in setting the threshold. Finally, a summary and comparison of the results with the help of the existing anomalies in semisupervised settings could be done. Specifically in terms of the Area Under the Curve (AUC) after training in order to compare and see the effectiveness. The hyperparameters of the models are then to be fine-tuned in order to optimize their performance and achieve the most accurate anomaly detection outcomes.

The choice of the abovementioned model is based on the results and observation of the experimentation performed in the master’s thesis of Andrei Perov “Leveraging Deep Learning Methods for Visual Anomaly Detection in Sheet Metal Forming Inspection”. It can be seen that Convolutional Autoencoder (CAE) is already a powerful approach for an existing problem. The most important result of Andrei’s work is that AE-GMM (Autoencoder and Gaussian Mixture Model) framework helped to avoid local minimum and the CLLV (Correction of Log-Likelihood LossVectors ) helped to avoid false positives. However, the combination of GMM and AE imposes challenges, such as:

* + The convergence problem of the model makes it difficult to replicate.
  + Hard to avoid trivial solutions and singularity problems.
  + In the encoder part, Andrei used GMM, so in order to gain good log-likelihood the model can generate strange features which spoil the performance of CAE. So CAE needs to be optimized as well as the latent space size for the CAE model to overcome weird features.

Thus in order to generalize the model to other parts in a time and efforts saving manner the experimentation was decided to be performed on pure CAE.

Overall, 5 tools 2, 5, 115, 117, and 120 were taken for experimentation, all of them have anomalous samples except for tool 120. Each tool consists of several cameras with the view from different angles. The periods taken are from July 6, 2020, to December 7, 2022, the last 4 production runs. First, we train the models separately to see how it performs on each tool for further analysis, except for tool 5 since the samples are mostly identical to 2. The results are described below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tool #** | **AUC** | **Threshold** | **TPR** | **FPR** |
| Tool 2 | 0.96 | 0.33528766 | 0.968750 | 0.107009 |
| Tool 115 | 0.41 | 0.22674602 | 0.953488 | 0.907068 |
| Tool 117 | 0.68 | 0.02672278 | 0.986486 | 0.838164 |
| Tool 120 | N/A | N/A | N/A | N/A |

From the results we can see that Tool 2 demonstrates a high AUC value of 0.96, indicating strong performance in anomaly detection. The selected threshold of ~ 0.34 achieves a high True Positive Rate (TPR) of ~0.97, meaning that the tool successfully identifies a large majority of true anomalies. The False Positive Rate (FPR) is relatively low at ~0.11, suggesting that the tool maintains a good balance between accurate detection and minimizing false alarms. With an AUC value of ~0.68, Tool 117 demonstrates moderate performance in anomaly detection. However, the FPR is relatively high at 0.838164, indicating a notable number of false positives. Tool 115 performs relatively poorly due to its high FPR.

Initially, in order to reach primary research inquiry, the idea was to start training one tool and gradually add the extra tool only after we got notably good AUC and other parameters. However, due to the time constraint, the focus shifted towards evaluating the performance of the global model when applied to all parts simultaneously. The results of this evaluation can be observed below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tool #** | **AUC** | **Threshold** | **TPR** | **FPR** |
| All tools | 0.72 | 0.024134 | 0.994505 | 0.760382 |

Given the objective of achieving an AUC of at least 0.80 for the global model, with a FPR of less than 0.15, subsequent steps were taken to enhance the results of tools 115 and 117 due to their comparatively worse performance:

The notion on which improvement steps are based is emphasizing the importance of clean and organized data. It is well known that even the most advanced algorithms can struggle when confronted with poor-quality data. To address this, a thorough review of the patches was conducted, eliminating and adjusting complex and unnecessary elements based on the analysis of false positives and false negatives. The results were as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tool #** | **AUC** | **Threshold** | **TPR** | **FPR** |
| Tool 115 | 0.87 | 1.8197951 | 0.8364 | 0.163718 |
| Tool 117 | 0.75 | 2.3361623 | 0.8018 | 0.346739 |

Both Tool 115 and Tool 117 demonstrate improvements in AUC compared to their previous results. Also, Tool 117 shows a slight decrease in the false positive rate, which may impact its overall effectiveness. Tool 115 exhibits a visible decrease in the false positive rate, indicating a more refined performance.

However, the evaluation results of the global model based on the improvements done unexpectedly showed slightly worse results than it was before:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tool #** | **AUC** | **Threshold** | **TPR** | **FPR** |
| All tools | 0.69 | 0.01804204 | 0.99 | 0.81 |

Based on the results of the experimentation deployment was run on Tool 2 since it showed the best result, however, model predictions were 95% of the time wrong. After some analysis, a strong relation to normalization indicators such as the mean and standard deviation of the data was observed in defining the results.

In the subsequent phase of the experimentation, our objective was to explore the feasibility of applying the aforementioned global model with distinct thresholds for each tool. To assess this scenario, we employed the parameters of the global model to evaluate and test Tool 2 once more. Additionally, the testing and evaluation phases were divided into each camera in the abovementioned tool, and normalization parameters were set individually to get more accurate results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Camera #** | **AUC** | **Threshold** | **TPR** | **FPR** |
| Cam 1 | 0.23 | 0.9177122 | 0.941176 | 0.789174 |
| Cam 3 | 1.0 | 4491.3193 | 0.941176 | 0.0 |

The obtained results exhibited inconsistency and lacked meaningful information, leading to the abandonment of the concept of validating a single global model with multiple distinct thresholds.

Based on the comparative analysis of the results, it can be concluded that the most effective approach is to train individual models for each tool, employing individual thresholding.

As a result, in order to set the threshold for Tool 120 which is without anomalous samples the anomaly score distribution graph was analyzed. According to the observation of the values of 95% and 99%, it shows huge threshold numbers as 27 which means that we accept most of the anomalies as normal. In order to get more meaningful results threshold was set based on visual analysis of the anomaly score distribution, the area where distribution is declining visually was set and the number 4 was selected which showed the result of FPR 0,13.

Overall, the high dependency of the model on normalization parameters causes complications in setting the one threshold in one global model. Since the spread of data is high due to lighter regions in one tool and darker regions in another in general it is hard to define one average normalization indicator for one global model. It can be assumed that the false positives of one tool would be given up for the sake of the other. Additionally, while training the global model the influence of easy shapes was observed. It was noticed that the presence of simplistic shapes, such as basic lines, could hinder the model's ability to learn complex structures. To counteract this, the overall proportion of such easy shapes could be reduced within the pool of tools. However, in this case, the distortion of the probability distribution of the data will be under question. Another, approach to reaching effective results in the global model could be changing the loss function from MSE (Mean Squared Loss) to SSIM (Structural Similarity Index Measurement). Since MSE calculates the mean square error between each pixel for the two images we are comparing. Whereas SSIM does the opposite and looks for similarities within pixels; i.e. if the pixels in the two images line up and or have similar pixel density values. This idea is based on the observation of the illumination within the bending machine area generated by a series of lamps positioned alongside the machine. However, the lighting conditions undergo significant variations due to different settings employed during various production periods. The placement of the lamps is constrained by the geometry of the bending machine and its movement paths. This limited positioning, combined with the dynamic nature of the machine's moving parts, introduces inconsistent lighting patterns and changing shadows, creating challenging conditions for detecting anomalies effectively.

On precision, 193.174.44.180 :

Parameters & Results for singular tools F:\Meerim\5\_tools\Tool2F:\Meerim\5\_tools\Tool5 etc.

Parameter & Results for the global model: F:\Meerim\5\_tools\Global\_model

The data for the singular & global model(s): F:\Meerim\5\_tools\Final\_data\multiple\_patches