A REPORT

ON

ANOMALY DETECTION IN X-RAY BAGGAGE SECURITY IMAGES USING NORMAL CLASS DATA

BY

Preetika Verma 2019A7PS0088P

AT

Central Electronics Engineering Research Institute, Pilani

A Practice School-1 Station of

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

June, 2021

A REPORT

ON

ANOMALY DETECTION IN X-RAY BAGGAGE SECURITY IMAGES USING NORMAL CLASS DATA

BY

Preetika Verma 2019A7PS0088P Computer Science

Prepared in partial fulfillment of the
Practice School -1 course nos.
BITS C221/BITS C231/BITS C241

AT

Central Electronics Engineering Research Institute, Pilani

A Practice School-1 Station of

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

June, 2021

ACKNOWLEDGEMENT

I want to express my deep and sincere gratitude to my instructors, Dr. Sandeep Joshi, who allowed us to work on this project on the topic ‘Anomaly detection in X-Ray images using normal class data,’ which strengthened my knowledge and understanding of this topic and the entire field in general. I am grateful to Dr. Dhiraj Sangwan who is guiding us for this project and helping us with the research.

Further, it allowed me to do much research as a result of which I learned many new things. I also have to thank people who have previously contributed to this field. Their work gave me better insight into the path I followed for completion of this project. It gave valuable guidance. Lastly, I would thank my family and friends for their kind cooperation and encouragement, which helped me in the completion of this project.

With regards,
Preetika Verma

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)

Practice School Division

Station : CEERI   Centre : Pilani
Duration : 7 weeks   Date of start : 1st June, 2021
Data of submission : 23rd June, 2021

Title of the project : Anomaly detection in X-ray security images using normal class data

ID No. : 2019A7PS0088P
Name : Preetika Verma
Discipline: Computer Science

Name and designation of the expert: Dr. Dhiraj Sangwan, Senior scientist, CEERI

Name of the PS faculty : Dr. Sandeep Joshi

Key words: Anomaly Detection, Deep Learning
Project Areas: Image Processing, Deep Learning

Abstract: Deep learning for anomaly detection aims at learning anomaly scores via neural networks for the sake of anomaly detection. The dataset is highly biased towards one class (normal) due to insufficient images with abnormal data or anomalies. We are trying to recreate the results of the paper - 'A New GAN-Based Anomaly Detection (GBAD) Approach for Multi-Threat Object Classification on Large-Scale X-Ray Security Images'.

Signature of Student: Signature of PS faculty

Preetika Verma   Dr. Sandeep Joshi

Date: 20/07/2021   Date: 20/07/2021

Table of Contents

ABOUT CEERI

Central Electronics Engineering Research Institute (CEERI) is a national laboratory established first in Pilani, Rajasthan and then subsequently in Chennai and Jaipur for the advancement of Research and Development in the field of Electronics. It has made immense contributions to the growth of electronics and pioneered the steps of R&D in the country. It has established state-of-the-art infrastructure and encouraged innovations in various areas such as Plasma Tubes, MEMS and Microsensors, Microwave Tubes, VLSI Design, Nano Structures, Power Electronics, Embedded Systems and many more. It has dedicated many projects towards the nation and has always worked towards implementing the latest technologies and helping the country keep up with the changing tech trends in the field of Electronics in the world

Our mentor for this project is Dr. Dhiraj Sangwan, who is a Senior Scientist at CEERI Pilani centre. He has tremendous experience in the field of Machine Learning, Computer Vision, Deep Learning to name some and he has worked on several projects under this field. He has worked as a Co-Principal Investigator on several deep learning projects, and he has also been invited to scientific talks and seminars such as “Deep Learning for Computer Vision”: Expert talk in 3rd International Conference on Intelligent Information Technology (ICIIT), CEG, Anna University, Chennai on December, 13,2018.

INTRODUCTION

Problem statement –

Recognizing threat objects in x-ray security images. Screening the images for security threats used to be done manually but locating and identifying all threats can be a challenging task. Research in this field has been limited by lack of sufficient data. The research paper we have studied adapts a novel GAN-based anomaly detection approach using multi-label classification to solve this problem. The datasets of x-ray security images that are available have skewed distribution of samples.

CNNs have been widely used for object detection and they have become deeper and wider in recent years thus giving higher accuracy. However, it is difficult using them for these security purposes since datasets have very few images of anomalies in comparison to images that are normal. In the paper, a GAN based anomaly detection approach has been used along with CNN and SVM classifier.

LITERATURE REVIEW

For the GBAD approach, the training was divided into three phases –

1. **Phase 1 -** This was done using the ideal dataset, i.e. the dataset was balanced by balancing from negative samples the same number of data that is in the positive samples. An ideal dataset is used for CNN so that it learns strong representations of target threat objects and can accurately discriminate between positive and negative samples. The negative images are used as background class (label - [0 0 0 0 0]) and reduce false positives. The subnetwork used in this phase has a CNN backbone which is used for classification and feature extraction. L1, L2, L3 and L4 are the convolutional layers of FPN. FC1, FC2, FC3 and FC4 are the fully connected layers pooled from each of the convolutional layers of FPN. FC5 is the 5 dimensional (SIXRay dataset has 5 types of anomaly) label that is output from the model. Ideal dataset is used for this subnetwork to prevent the CNN from being distracted by a large number of negative samples since the data distribution is skewed. The training function is the binary cross-entropy loss function. p is the predicted output, *yi* is the label for class i and N is the number of classes.

$$\begin{aligned} &L(p, y) \\ &= -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot \log(p(y_i)) + (1 - y_i)\log(1 - p(y_i)) \end{aligned}$$

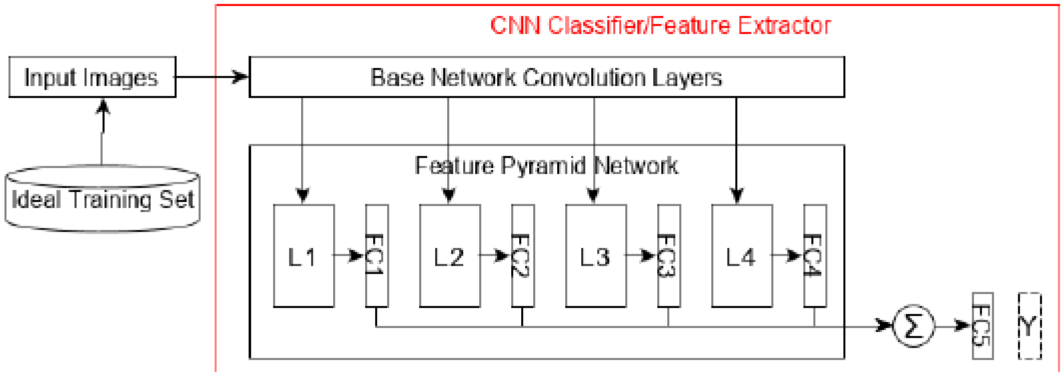


Figure 1 :Training phase I

2. **Phase II -** For the second phase of the training, a Bi-GAN is trained only on features extracted from negative samples. 50,000 images are used in this phase. The network learns normal class data, i.e. images without any anomaly. The objective function for this subnetwork is –

$$\begin{aligned}&\min_{G,E}\max_{D} V(D,E,G)\\&\quad= \mathbb{E}_{x\sim p_X}\left[\mathbb{E}_{z\sim p_E(\cdot|x)}\left[\log D(x,z)\right]\right]\\&\qquad+ \mathbb{E}_{z\sim p_z}\left[E_{x\sim p_G(\cdot|z)}\left[1-\log D(x,z)\right]\right]\end{aligned}$$

Reconstruction loss is also added to the objective function. This loss helps in generating similar features. fd are the features extracted from the fully connected penultimate layer of the discriminator network.

$$\begin{aligned}R_{loss} &= \|x - G(E(x))\|_1\\FM_{loss} &= \|f_D(x, E(x)) - f_D(G(E(x)), E(x))\|_2\end{aligned}$$

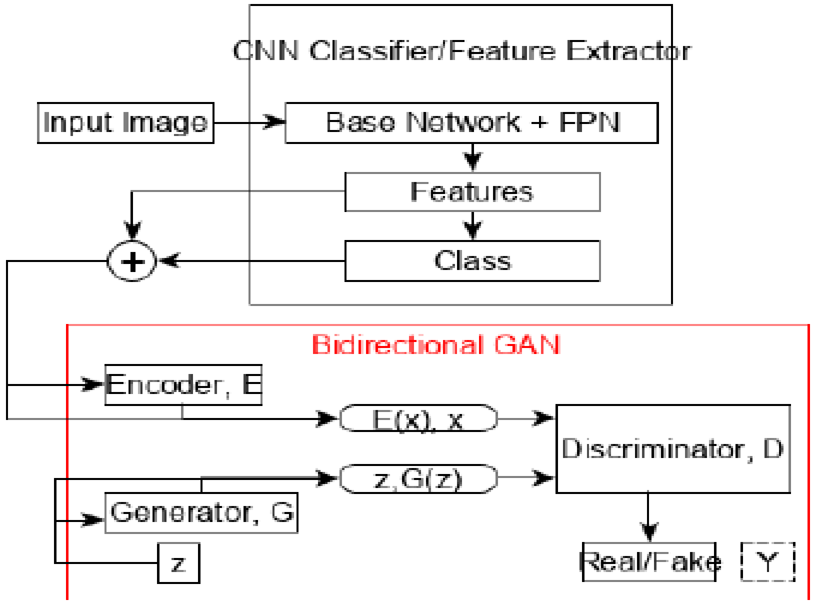


Figure 2 : Training phase - II

A BiGAN, or Bidirectional GAN, is a type of generative adversarial network where the generator not only maps latent samples to generated data, but also has an inverse mapping from data to the latent representation. The motivation is to make a type of GAN that can learn rich representations for us in applications like unsupervised learning.

In addition to the generator G from the standard GAN framework, BiGAN includes an encoder E which maps the data x to latent representations z. The BiGAN discriminator D discriminates not only in data space (x versus G(z)), but jointly in data and latent space (tuples (x,E(x)) versus (G(z),z)), where the latent component is either an encoder output E(x) or a generator input z.

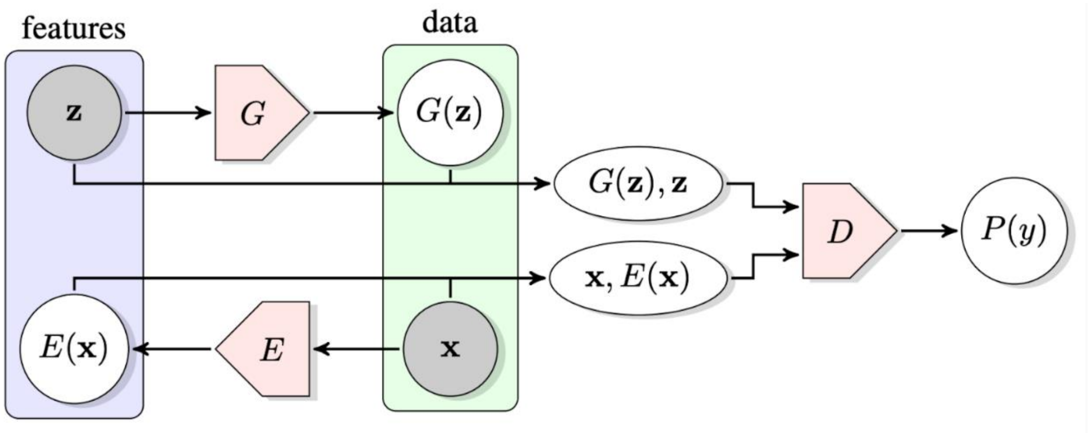


Figure 3 : Structure of Bi-GAN

3. **Phase III -** The third subnetwork has an SVM. The input features used here are the reconstruction and feature matching losses from all the positive training samples and 50000 samples that were used in the second phase. SVM is optimized to find the best separation of normal class data and anomalous images.

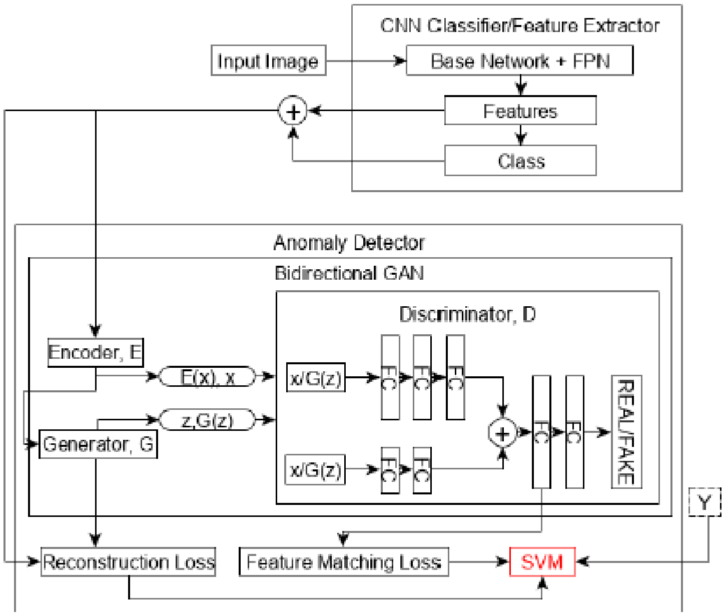


Figure 4 : Training Phase III

Dataset and evaluation metrics –

The dataset used in the paper is the SIXRay dataset, the largest public dataset on X-ray security images. A subset of this, SIXray10 is used, in which 10% of the samples are positive.

All the test data is ranked and average precision for each class is calculated. Average of this precision for all classes tells how well the model performs. This approach is much better than the previous approaches that used CNNs since the class imbalance problem is eliminated.

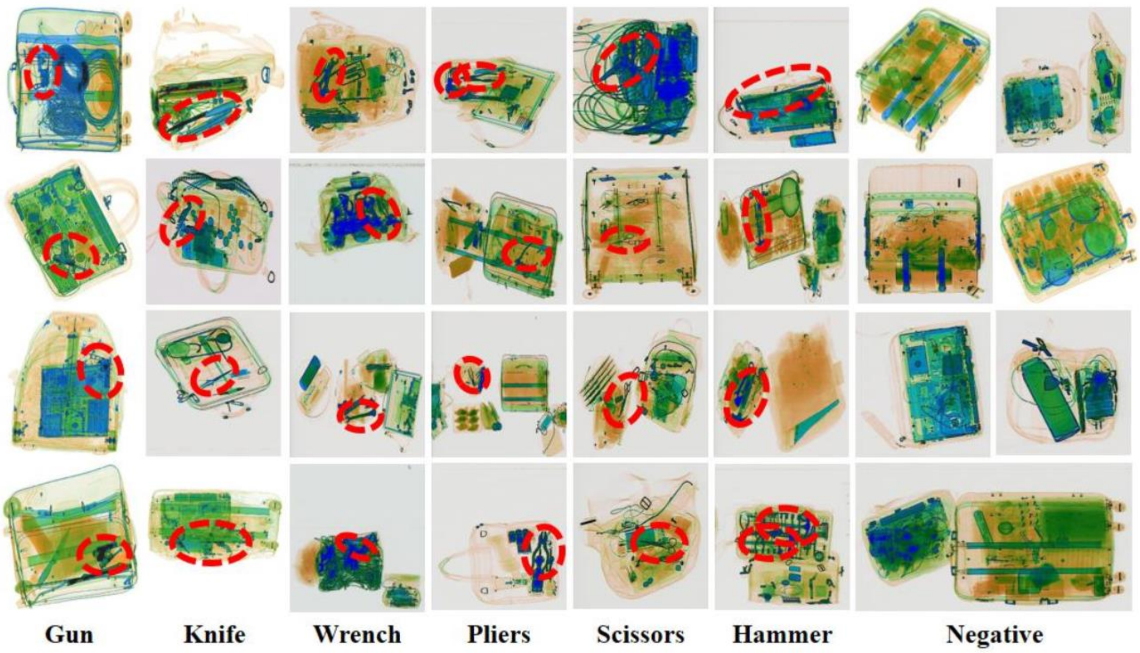


Figure 5 : A few examples from the SIXRay dataset

MULTI-LABEL CLASSIFICATION

Augmentation –

The annotations present in the dataset were the names of the specific objects present in the positive images. For the purpose of multi-label classification we created a label vector for each positive image denoting what objects were present in it. The following label vector was produced for each image –

Labels : [

0/1 : 1 if “Gun" is present, 0 otherwise

0/1 : 1 if “Knife" is present, 0 otherwise

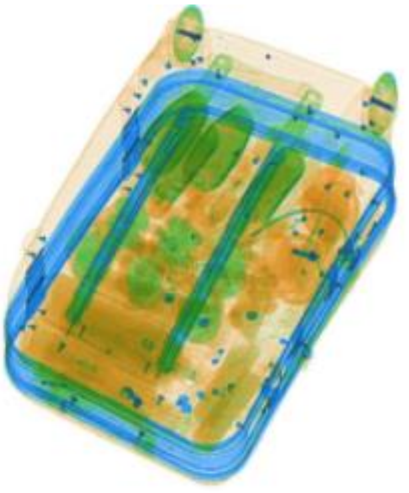
0/1 : 1 if “Wrench" is present, 0 otherwise

0/1 : 1 if “Pliers" is present, 0 otherwise

0/1 : 1 if “Scissors" is present, 0 otherwise

]

For example, the label vector for the following image containing a knife would be-



Label : [0 1 0 0 0]

Images containing multiple objects had multiple 1’s in their label vectors. The training images were resized into size 256*256 and then normalized. For the test images, we resized them, added random horizontal flips, added Random Affine and normalized them.

PHASE 1 -

Training –

Training dataset - 8129 positive images from SIXRay 10 subset have been used. With them, we have used 8129 negative images, training the CNN on an ‘ideal’ dataset. The images and labels were in a csv file. We converted these into json data, writing a Python script for it.

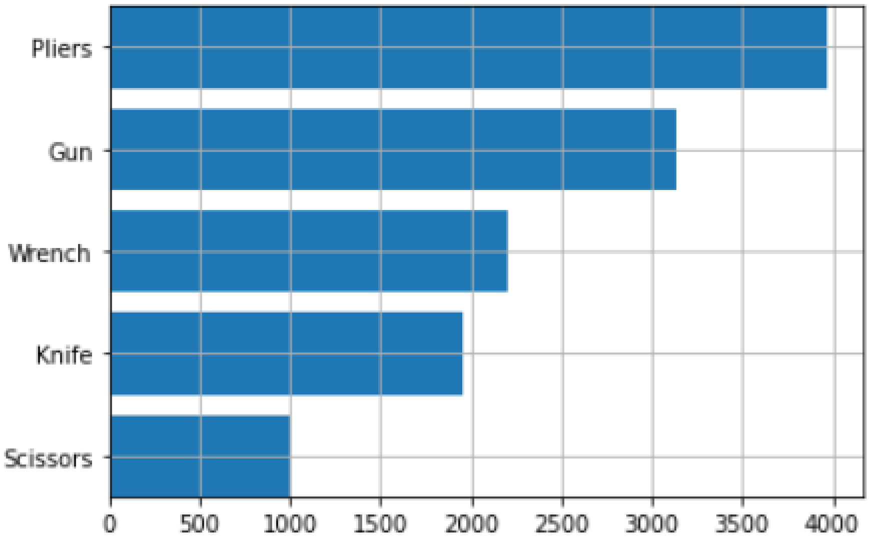


Figure 6 : Label Distribution

The count of various anomalous objects in the distribution was as follows –

Pliers – 3963

Gun – 3130

Wrench – 2204

Knife – 1953

Scissors – 995

Our focus was not to equally balance the various classes, but to train the model on equal number of positive (containing at least one anomalous object) and negative (containing no anomalous objects) samples.

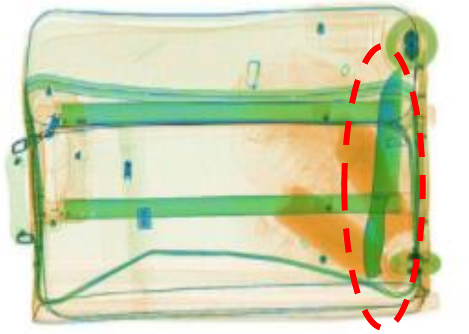


Figure 7 : Knife

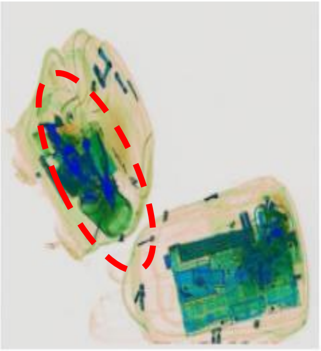


Figure 8 : Plier

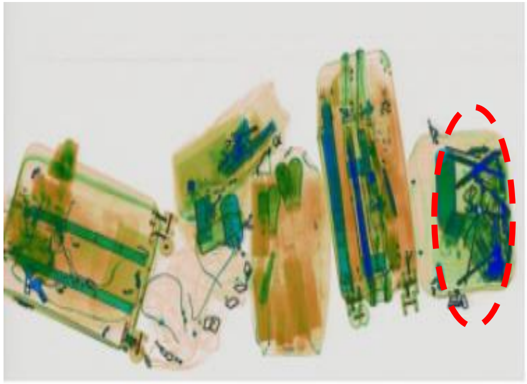


Figure 9 : Wrench

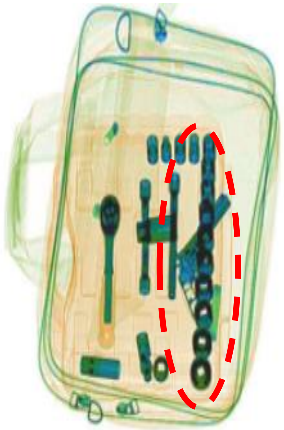
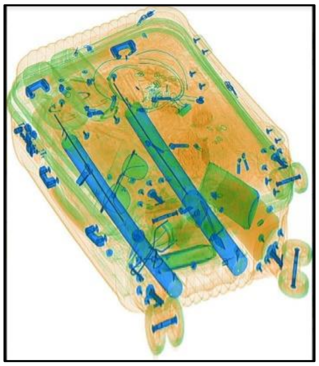


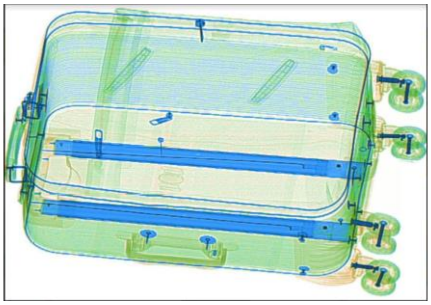
Figure 10 : Gun

The loss used in this project is Binary Cross Entropy loss in a simple multi-class classification case when the target labels are independent. We used this loss instead of SoftMax as we wanted a separate output, true or false for each class. Using SoftMax would have meant that the labels assigned must be mutually exclusive which is not our case. During the loss computation, only the logit corresponding to the truth target label and how large it is compared to other labels were taken care of. The network was trained for ~15 epochs after which the model started to overfit. The model used was ResNext50. The final layer was modified. There are 5 classes - Gun, Knife, Wrench, Pliers and Scissors. For each of these, there will be a true or false value. For negative images, output will be a five-dimensional vector of zeros. Presence of one would mean that the image is anomalous.



Predicted labels – Wrench

Ground truth labels – (negative image)



Predicted labels – (negative image)

Ground truth labels – (negative image)

RESULTS -

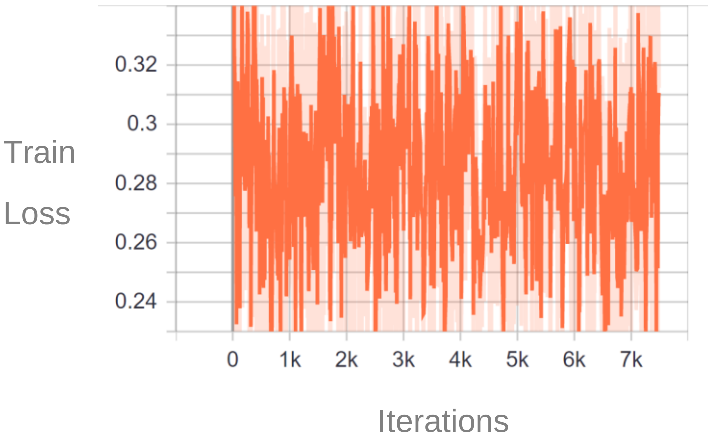


Figure 11 : Train Loss

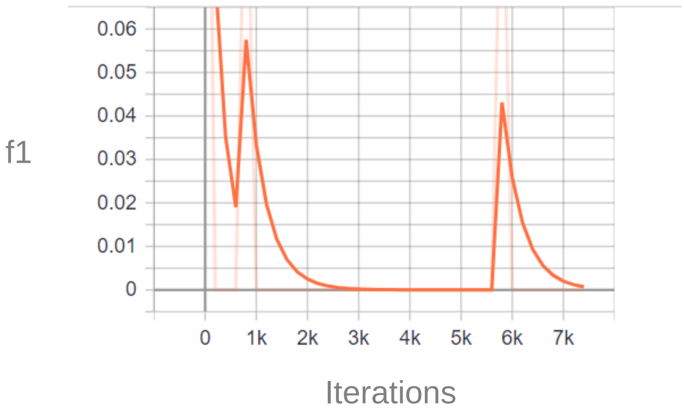


Figure 12 : Macro f1

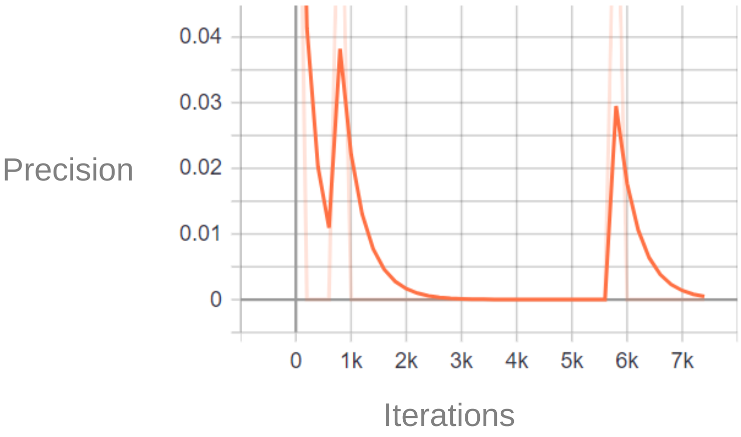


Figure 13 : Macro Precision

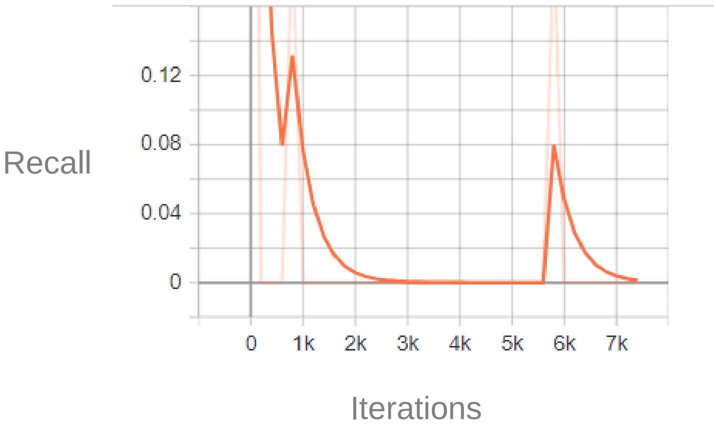


Figure 14 : Macro Recall

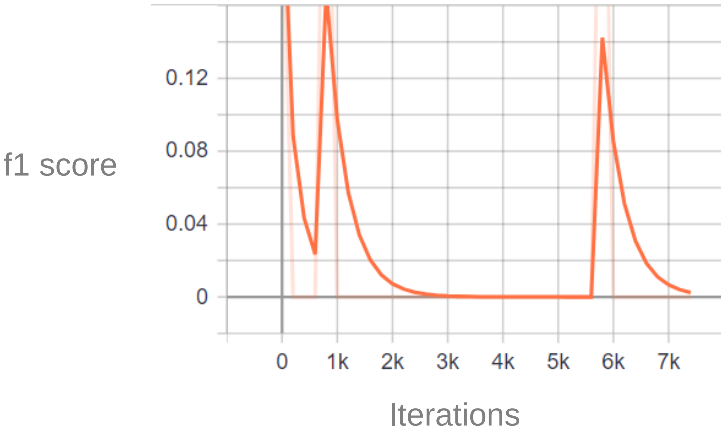


Figure 15 : Micro f1

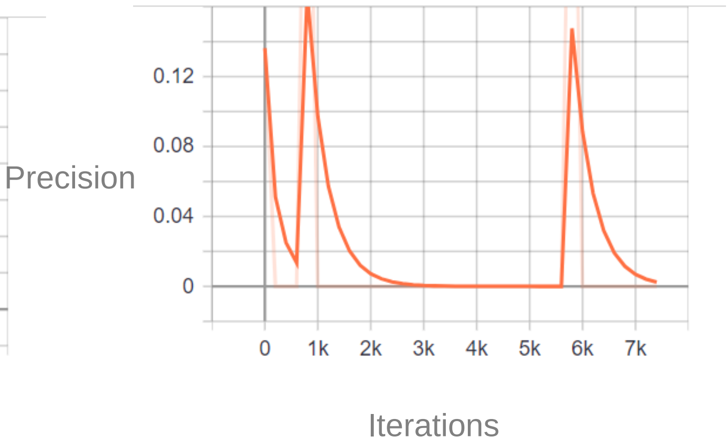


Figure 16 : Micro Precision

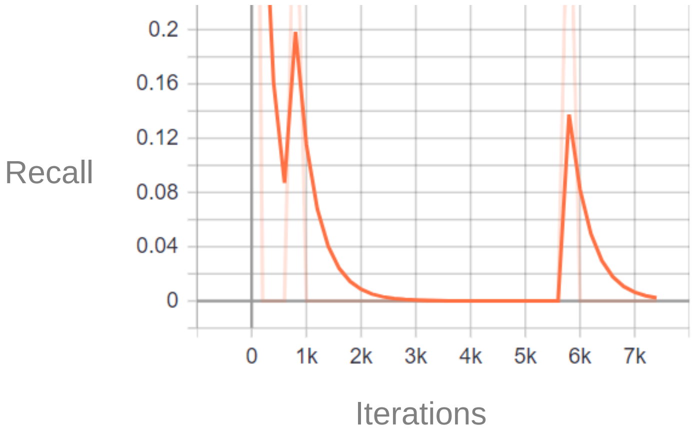


Figure 17 : Micro Recall

Test Sample Results -

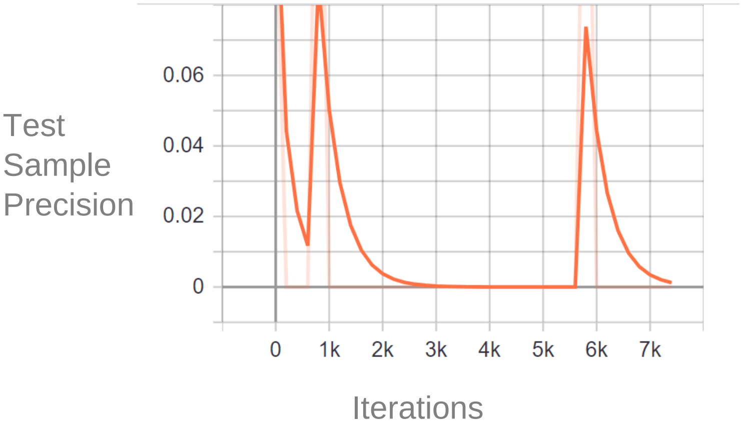


Figure 18 : Test Sample Precision

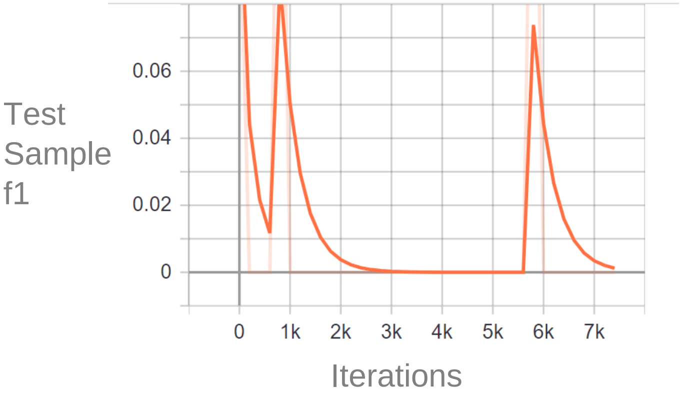


Figure 19 : Test Sample Precision

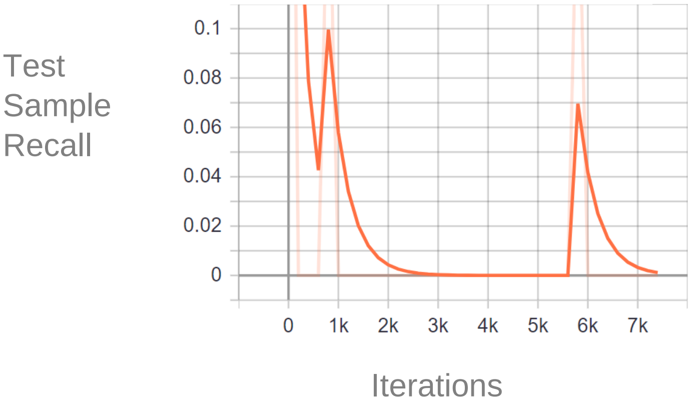


Figure 20 : Test Sample Recall

We trained our model for 15 epochs and the average training loss was 0.282. The difference in the macro and micro scores is explained by the intra-class imbalance in our dataset. The macro averages have each of the metrics calculated for each class separately and then the average is taken, hence considering all classes equally. Micro averages aggregate the contributions of all classes.

CONCLUSION

We are trying to tackle extreme class imbalance using multilabel multiclass classification. The first phase of the training gives us the classification labels output by the base CNN architecture. Compared to naïve CNN architectures, this approach is much better as it eliminates the class imbalance problem and suppresses the false positives.

Future Work –

- Adding FPN on the backbone CNN architecture to get more enhanced Feature Maps.
- Adding a Bi-GAN, to which, only the feature maps of negative images would be fed. The aim is to make the Generative Model learn the underlying distribution of only the normal class data.
- Using SVM classifier to learn the optimal separation between the normal class data and the anomalous images.
- Trying other CNN architectures instead of those tested by the authors, such as NASNET to increase the prediction accuracies.

REFERENCES

1. Dumagpi, Joanna Kazzandra, Woo-Young Jung, and Yong-Jin Jeong. "A new GANbased anomaly detection (GBAD) approach for multi-threat object classification on largescale x-ray security images." IEICE TRANSACTIONS on Information and Systems 103.2 (2020): 454-458.
2. Akcay, Samet, Amir Atapour-Abarghouei, and Toby P. Breckon. "Ganomaly: Semisupervised anomaly detection via adversarial training." Asian conference on computer vision. Springer, Cham, 2018.
3. Akçay, Samet, Amir Atapour-Abarghouei, and Toby P. Breckon. "Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection." 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019.
4. Davletshina, Diana, et al. "Unsupervised anomaly detection for x-ray images." arXiv preprint arXiv:2001.10883 (2020).
5. Nakao, Takahiro, et al. "Unsupervised Deep Anomaly Detection in Chest Radiographs." Journal of Digital Imaging (2021): 1-10.
6. Pang, Guansong, et al. "Deep learning for anomaly detection: A review." arXiv preprint arXiv:2007.02500 (2020).
7. Federico Di Mattia et al. “A Survey on GANs for Anomaly Detection”. In: CoRR abs/1906.11632 (2019). arXiv: 1906.11632.
8. learnopencv/PyTorch-Multi-Label-Image-Classification-Image-Tagging at master · spmallick/learnopencv · GitHub
9. Multi-Label Image Classification with PyTorch: Image Tagging | Learn OpenCV

GLOSSARY

- **Class Imbalance Problem -** An example of a classification problem where the distribution of samples across the known classes is biased or skewed. The distribution can vary from a slight bias to a severe imbalance where there is one example in the minority class for hundreds, thousands, or millions of examples in the majority class or classes.

- **Feature Extraction -** Feature extraction involves reducing the number of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

- **Precision -** In pattern recognition, information retrieval and classification (machine learning), precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances.

- **Recall -** In pattern recognition, information retrieval and classification (machine learning), recall (also known as sensitivity) is the fraction of relevant instances that were retrieved.

- **BiGAN -** Has an encoder along with the generator-discriminator GAN architecture

- **AnoGAN -** Deep convolutional GAN architecture proposed in 2017 for anomaly detection.

- **AUC-ROC curve -** This is used for checking or visualizing the result of a multi-class classification problem. AUC plots the entire area under the ROC curve. ROC curve plots the True positive rate on y-axis and the false positive rate on x-axis at different classification thresholds. Higher the value of AUC, the better the model is at distinguishing between positive and negative classes.