A REPORT ON

LEATHER DEFECT DETECTION

By

PRANAV MANDLIK	2020A7PS1392G	B.E. COMPUTER SCIENCE
YASH KISAN DABHADE	2020A7PS0974G	B.E. COMPUTER SCIENCE

Prepared in partial fulfillment of the
Practice School-II Course No.
BITS C412/BITS G639

At

CSIR-CEERI Chennai

A Practice School –I Station of

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI (Rajasthan)
June 2022

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude towards our instructor in charge of this project, Dr. Amalin Prince, Department of Electrical and Electronics Engineering, BITS Pilani - Goa campus, for constantly motivating and encouraging us to do this project. Through this project, we have been introduced to one of the most rapidly developing technologies based on Image processing and Deep learning techniques. We would like to thank our project mentor, Mr. J Suriya Prakash, for providing us with the opportunity to work on this project. This might have been an incomplete project without his constant guidance. We also like to thank our college administration BITS Pilani, for providing us with this opportunity through the Practise School -1 program.

PS-1 Organization Profile

Central Electronics Engineering Research Institute (CEERI) is a part of the Council of Scientific and Industrial Research (CSIR) and is committed to research and Development in electronics to meet the requirements of its customers. CEERI has been continually identifying the thrust areas of key National/Industrial relevance and has been developing competitive technologies with a goal to achieve excellence and self-reliance in these areas.

CEERI Centre in Chennai, a pioneering institution in the field of Quality Control Instrumentation for process industries, is a regional centre of Central Electronics Engineering Research Institute (CEERI), Pilani, Rajasthan. The centre has a rich experience and expertise in this field and has developed many online monitoring systems for various process on indigenous development of special sensors and systems suited to the online measurement and control and these involve varying technologies such as Near Infra Red (NIR) Gauging, Beta Gauging, Optical, Electromechanical methods and the currently emerging Image Processing techniques. The centre has been extensively contributing its expertise in Lab VIEW for online apple sorting. The Centre has also unveiled a system for sorting plastics through NIR spectroscopy.

CEERI Centre has a well-equipped Optics laboratory for the development of electro-optical systems, a DSP and FPGA lab and a rapid prototyping environment of electronic systems hardware and firmware and work stations with advanced imaging and computing software/hardware to develop machine vision technologies for online inspection and grading. It also has CAD and drafting facilities to cater for the needs of R & D activities

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI-(Rajasthan)
Practice School Division

Station: Central Electronics Engineering Research Institute (CSIR-CEERI)

Centre: Chennai

Duration: 8 weeks

Date of Start: 30th May 2022 **Date of Submission:** 22nd July 2022

Title of the Project: Leather defect detection

Participants:

S.No.	**Name**	**ID**	**Discipline**
1.	PRANAV MANDLIK	2020A7PS1392G	B.E. COMPUTER SCIENCE
2.	YASH KISAN DABHADE	2020A7PS0974G	B.E. COMPUTER SCIENCE

Name and designation of the expert(s): Mr. Suriya Prakash, Scientist, CSIR-CEERI Chennai

Name(s) of the PS Faculty: Dr. Amalin Prince, Associate Professor, BITS Pilani, Goa Campus

Key Words: Leather defects, Image Processing, Deep Learning

Project Areas: Leather defect detection

Abstract: The leather industry in India accounts for around 13% of the world's leather production. Also, India is one of the major exporters of leather in global markets. Leather Defect Detection is the most important step before its sale and further processing. This industry needs more advanced and accurate defect detection techniques to meet the increasing demand to improve leather quality. Herein, we propose a model that will accurately identify if the given sample is that of defected leather or not, using state of art Visual Geometry Group(VGG -16) Deep CNN architecture and further count the number of defects in it using image processing techniques like blob detection.

Signature(s) of Student(s)

1.	**Pranav Mandlik**	
2.	**Yash Kisan Dabhade**	

Signature of PS Faculty: **Date:**

TABLE OF CONTENTS

1. INTRODUCTION

The Indian Leather Products and Footwear Industry is very important to the Indian economy. This industry accounts for high export earnings. During the 2020-21 fiscal year, India exported $3.68 billion in footwear, leather, and leather products. The leather industry is a high-employment sector, employing over 4.42 million people, most of whom come from lower-income families. India is the second largest exporter of leather garments and the 4th largest exporter of leather goods in the world. [1] Modernisation and technology up-gradation of the leather industry will prove beneficial not only for the sector but also for the Indian economy.

The Indian leather quality control industry has an urgent need for automated leather inspection. Utilising cutting-edge Artificial Intelligence techniques, CSIR-CEERI has been a pioneer in creating automated leather quality and grading systems in recent years. Leather Inspection System offers high-quality output and remarkable performance using the computer vision method with machine learning/deep learning techniques.

Several processing procedures must be performed on leather before it may be shipped as a manufactured good. Preparation for tanning, tanning, and finishing are the general phases of leather processing. The leather is then graded for quality, which is the process of classifying leather according to the surface defects discovered during the inspection.

Some of the typical leather defects are depicted in *Fig. 1.* Line defects are produced by damage to the hides and mainly occur during the production process. Stains are natural irregularities in the texture of leather. Wears are caused due to friction. Knots are strictly biological defects, while wear is caused both by natural causes and manufacturing errors. [2]

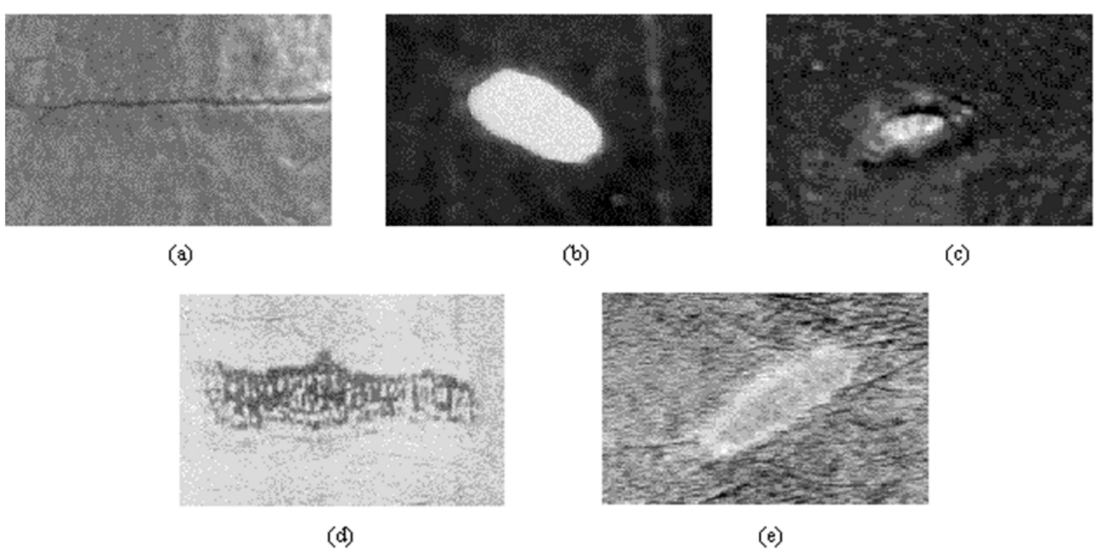

Fig. 1. Defects on leather. (a) Line (b) Hole (c) Knot (d) Stain (e) Wear.

The process of discovering surface defects and grading is carried out manually. The existing techniques for detecting defects in industrial settings are extremely time-consuming and labour-intensive. Our goal is to automate the process, maximize resource utilization, eliminate human error, and make production more efficient and effective. There are many challenges to the automation process, as the defects vary vastly in size and shape. Nevertheless, leather defect detection is necessary not only for accurate grading but also to locate the source of the defect so that corrective action can be taken to avoid future defects.

Since most of our report revolves around Open-CV and Convolutional Neural Networks (CNN), to fully understand the project, the following is a brief introduction to both of them.

1.1. OpenCV

OpenCV, or Open Source Computer Vision Library, is free for use as it is released under a BSD license. It has C++, Python and Java interfaces and supports Windows, Linux, Mac OS, iOS and Android. OpenCV is written in C/C++ and thus can use parallel computing to utilize all the cores of the system, thus being very useful for real-time applications.

OpenCV has a staggering 47 thousand people in its user community and downloads exceeding 14 million. It is used worldwide for various tasks as diverse as a collaborative art to cutting-edge robotics. The OpenCV library has more than 2500 optimised algorithms, including a comprehensive set of classic and state-of-the-art computer vision and machine learning algorithms.

These algorithms can be used for facial recognition, object detection, data augmentation, real-time filters, and so much more. OpenCV is a must for any practical image processing-related work. The powerful library is the basis for the first half of our project, including facial recognition, object detection, and object classification using SVMs.

1.2. Convolutional Neural Networks:

Convolutional neural networks are neural networks that are used in image recognition and classification. There are three main operations performed in a convolutional neural network

- Convolution
- Non-linearity
- Pooling or sub-sampling

1.2.1 Convolution

Any grayscale image can be represented as a 2D matrix with the entries corresponding to the pixel intensities. An RGB image can be visualized as a set of three 2D matrices stacked on top of each other. A convolution's primary purpose is to extract an image's features by preserving the spatial relation between pixels. The input image is convolved with a kernel (filter) to produce the output image, also called a Feature Map or Activation Map. An example of convolution is shown below in *Fig. 2.*

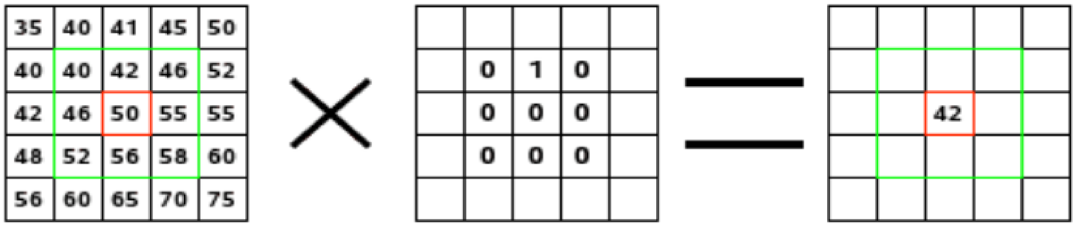

Fig. 2. Example of convolution

Different values of the filter matrix will produce different feature maps of the same image. In the table shown above, we see the effect of convolution with different filters. As observed in *Fig. 3,* different filters can detect different features of an image, like edges, curves, contrast variation, etc.

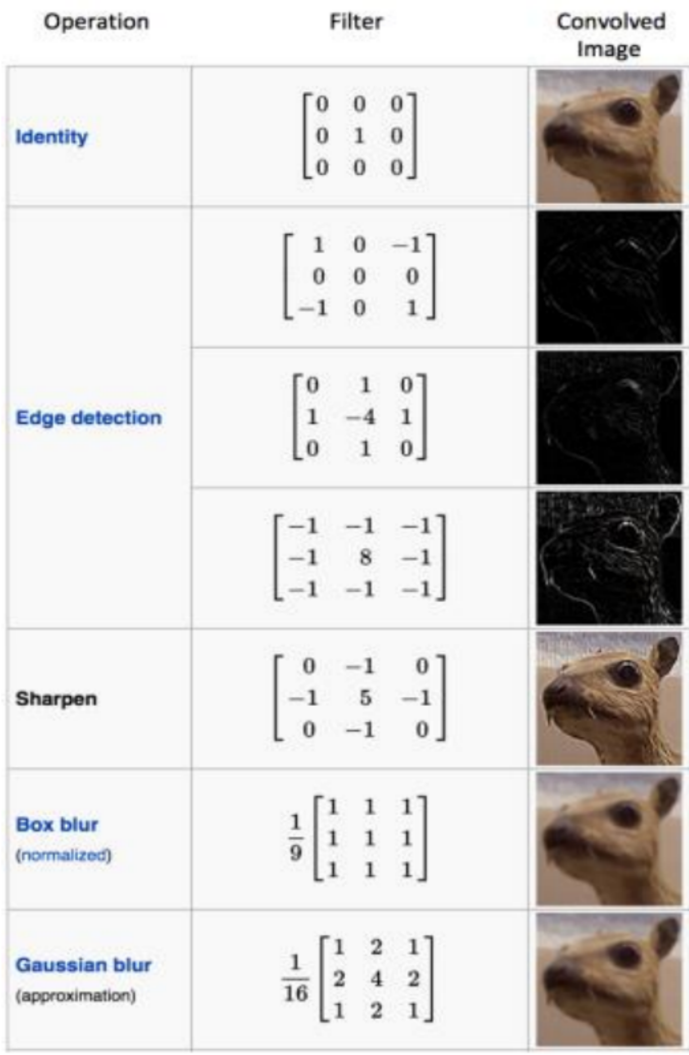

Fig. 3. Effect of convolution on the image after applying different filters

CNNs learn the values of these filters on their own during the training process. The more filters we have, the more features will be extracted, and the network becomes better at recognizing images.

1.2.2 Non-linearity

A non-linear layer is mainly applied to the output image from convolution because most of the real-world data we deal with are non-linear. Some of the non-linear functions include Tanh, sigmoid and ReLU(Rectified Linear Unit)

1.2.3 Pooling

Pooling (also called subsampling or downsampling) reduces the dimensionality of the feature map but retains the most important feature, as shown in *Fig. 4*. Pooling can be of different types: Max, Average, Sum, etc.

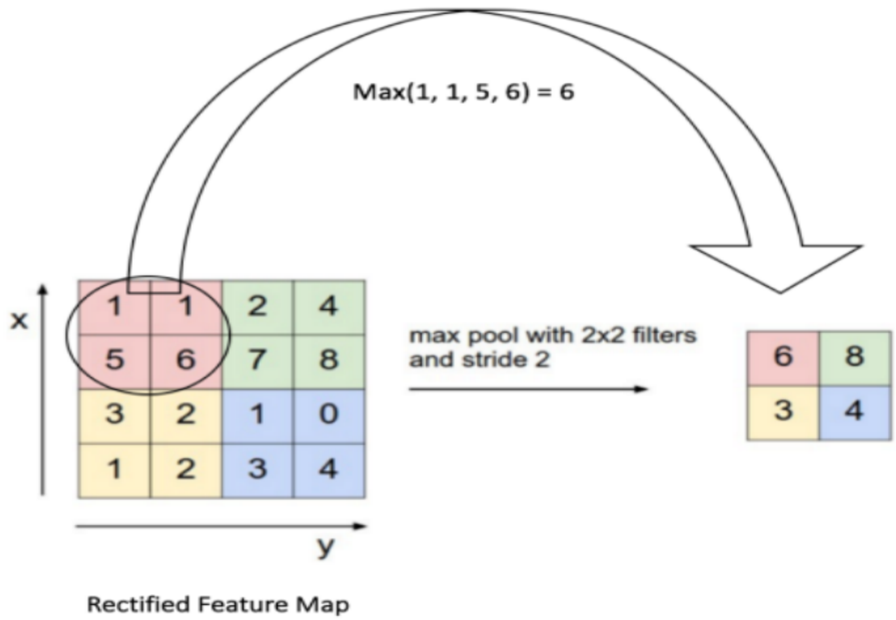

Fig. 4. Example of MaxPooling

The whole process of using a CNN model is shown in *Fig. 5*.

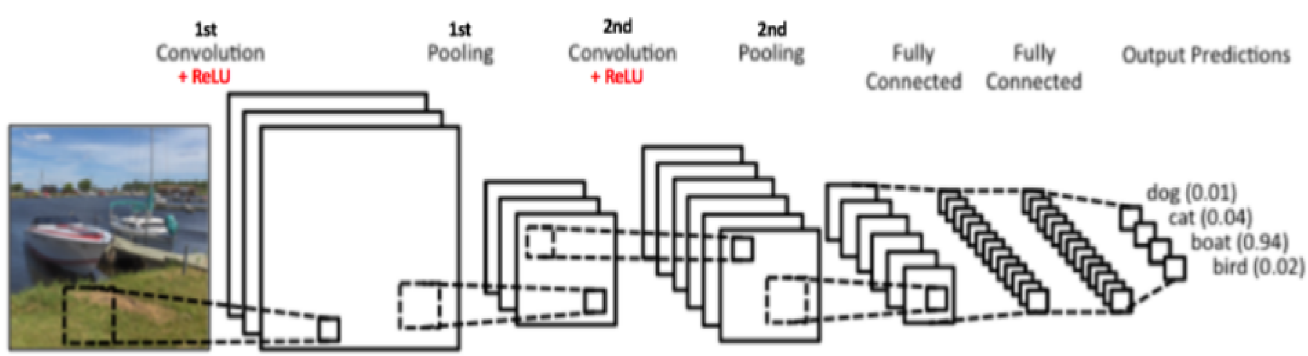

Fig. 5. Working of a Convolutional Neural Network

2. METHODOLOGY

2.1 Using Convolutional Neural Networks

2.1.1 Data preparation

For training our CNN model, we used MVTEC Anomaly Detection Dataset [3], which is distributed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (can only be used for research purposes). The dataset has around 400 images of leather (defected and non-defected leather), each labelled as either 'Good' or 'Anomaly'.

For counting the number of defects found on leather, images were randomly selected from the above dataset and web and used for further analysis. Some images of our dataset are shown in *Fig. 6*.

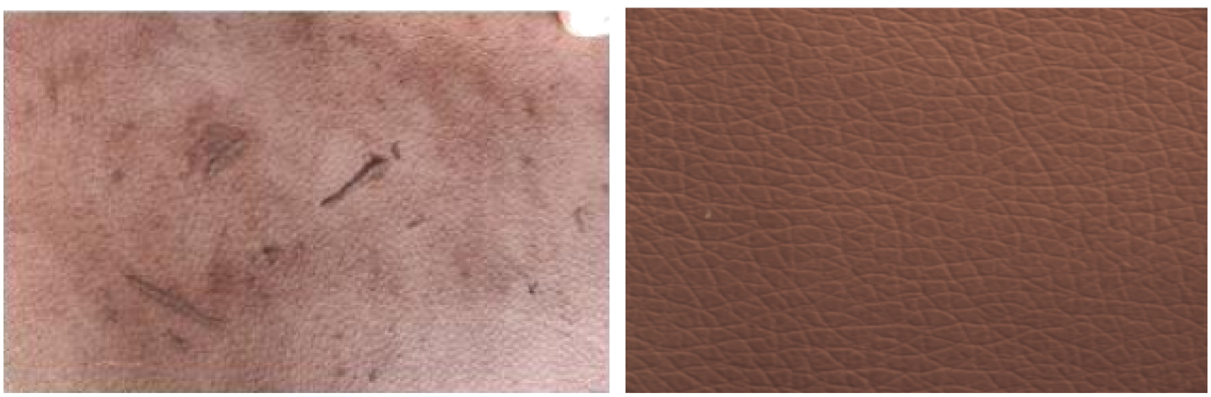

Fig. 6. Some images from dataset

2.1.2. Data Preprocessing

The first step in an image processing chain is to perform preprocessing on the images. Preprocessing can speed up the rest of the chain and reduce noise in the images. Detection of defects in leather is a difficult process due to the complex, non-homogenous nature of texture patterns on leather. A common defect in leather includes fine lines, stains, holes, and wear, all of which would result in a change in texture pattern around the defect.

To preprocess the data, we resized images to 224×224 pixels to speed up training. Images in the dataset are of size 1024×1024, but as defects are also of a large size, we may resize the image to a lower resolution without sacrificing model accuracy.

To clearly detect defective regions, bilateral filtering was also used. Without going into the mathematical details, which are beyond the scope of this report, Bilateral Filtering is a way to smoothen out the textural information without affecting the edges. As shown in *Fig. 7*, these pre-processing methods make blobs' edges more prominent, improving Blob detection.

Fig. 7. Effect of Bilateral filtering

We preferred to have a train, validation, and test parts - to train a model, tune hyperparameters and evaluate model accuracy, respectively. The entire dataset was split at an 80:20 ratio for training and testing data. For small datasets, we also made a provision for 5-Fold cross-validation to make sure that evaluation results are robust.

2.1.3 Defect detection

The traditional Visual Geometry Group(VGG) CNN architecture takes input as an image of dimensions (224, 224, 3). The first two layers have 64 channels of a 3*3 filter size and the same padding. Then after a max pool layer of stride (2, 2), two layers have convolution layers of 128 filter size and filter size (3, 3). This is followed by a max-pooling layer of stride (2, 2) which is the same as the previous layer. Then there are two convolution layers of filter size (3, 3) and 256 filters. After that, there are 2 sets of 3 convolution layers and a max pool layer. Each has 512 filters of (3, 3) size with the same padding. This image is then passed to the stack of two convolution layers. In these convolution and max-pooling layers, the filters we use are of the size 3*3 instead of 11*11 as in AlexNet and 7*7 as in ZF-Net. In some of the layers, it also uses 1*1 pixel, which is used to manipulate the number of input channels. There is a padding of 1 pixel (same padding) done after each convolution layer to prevent the spatial feature of the image [4].

We modified the traditional VGG-16 pre-trained on ImageNet and changed its classification head by replacing Flattening and Dense layers with Global Average Pooling and a single Dense layer.

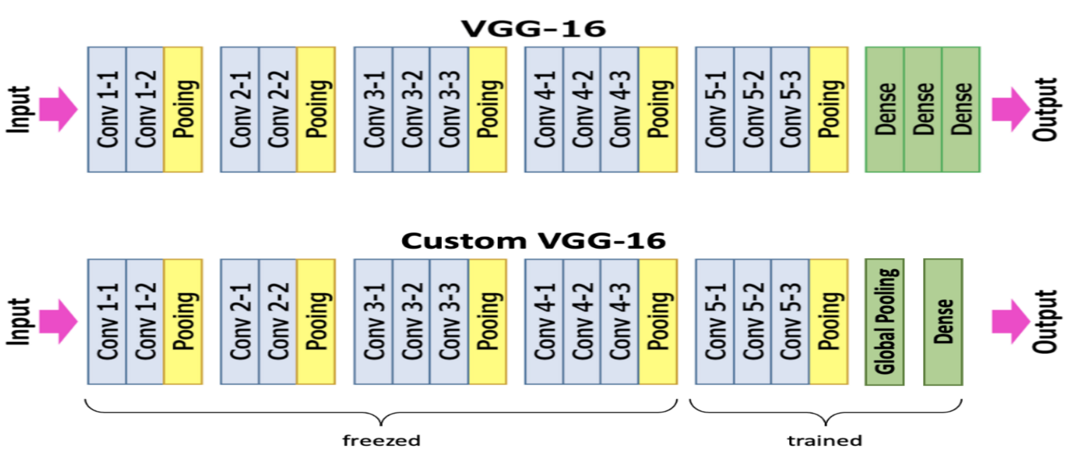

Fig. 8. Various layers used in VGG-16 and Custom VGG-16

We made our model in such a way that it not only helps us to classify images into ‘Good’ / ‘Anomaly’ classes but also to get a bounding box for the defect if the image is classified as an ‘Anomaly’. For this reason, we made the model in inference mode to output class probabilities and the heatmap, which will later be processed into the bounding box. Heatmap was created from the feature maps from deep layers.

To get a bounding box around the defect from the heatmaps, we normalised the heatmap in the range [0,1]. A threshold value was selected so all pixels above the threshold were converted to 1s and smaller ones to 0. Finally, a bounding box is generated around regions concentrated with more 1s, as shown in *Fig. 9.*

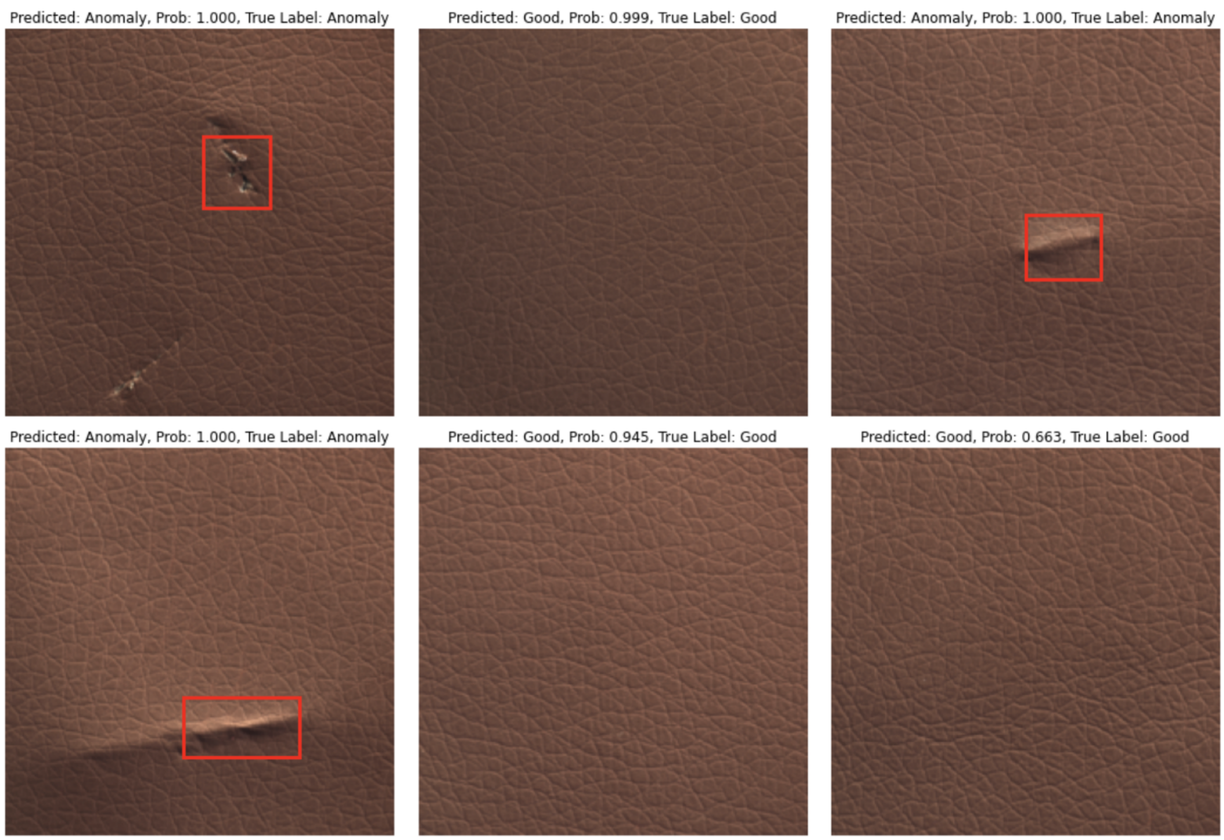

Fig. 9. Result of Custom VGG-16 model on MVTEC Anomaly Detection Dataset

2.2 Using Blob Detection

2.2.1 Blob Detectors

Blob detection techniques in computer vision are used to find areas in an image that are different from surrounding areas in terms of brightness or colour. A blob is a portion of a picture where some characteristics are constant or almost constant; all the points in a blob may be thought of as somewhat similar. Convolution is the approach used most frequently for blob detection.[7]

There are two basic kinds of blob detectors, based on detecting the local maxima and minima of the function and given some property of interest represented as a function of position on the picture. The differential techniques and the local extrema methods. These detectors may also be referred to as interest point operators or interest area operators, depending on the more modern nomenclature in use in the field.

Types of Blob Detectors used in the project.

1. **Laplacian of Gaussian**

   Given an input image f(x,y), this image is convolved by a Gaussian kernel

   $g(x, y, t) = \frac{1}{2\pi t} e^{-\frac{x^2+y^2}{2t}}$ at a certain level t to give a scale space representation

   $L(x, y; t) \ = g(x, y, t) * f(x, y)$. Then the result of applying the Laplacian Operator

   $\nabla^2 L = L_{xx} + L_{yy}$ is computed, which results in strong positive responses for dark blobs of radius $r = \sqrt{2t}$ for 2D images. It also produces a strong negative response for bright blobs or anomalies of similar size. The operator response is strongly dependent on the relationship between the size of the blob in the image domain and the size of the Gaussian Kernel [8].

2. **Difference of Gaussians**

   Difference of Gaussian (DoG), a feature-enhancement approach used in imaging science, involves subtracting one version of the original picture that has been blurred using a Gaussian blur from another version that has been less blurred.

   From the fact that the scale-space representation L(x,y,t) satisfies the given equation for diffusion, $\partial_t L = \frac{1}{2} \nabla^2 L$.

   Therefore, the difference between two Gaussian-smoothed pictures may also be calculated as the limit case of the Laplacian of the Gaussian operator (scale space representations)

   $$\nabla^2_{\text{norm}} L(x, y; t) \approx \frac{t}{\Delta t} \left(L(x, y; t + \Delta t) - L(x, y; t) \right)$$

3. **Determinant of the Hessian**
   Here we consider the scale normalization of the determinant of the Hessian.
   $$\det H_{\text{norm}} L = t^2 \left(L_{xx} L_{yy} - L_{xy}^2\right)$$
   When scale-space maxima of this operator are detected, one obtains another simple differential blob detector with automated scale selection that also reacts to saddles. Here, H L indicates the Hessian Matrix of the scale space representation L.

2.2.2 Image Processing and Blob Detection

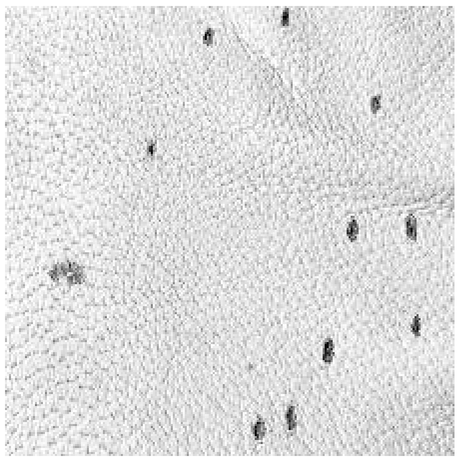

Fig. 10. Sample image from dataset

In image preprocessing, we applied a bilateral filter to the image so as to keep defects intact while reducing the texture noise of leather images which can be misjudged to defects after thresholding.
Gaussian filters of size 3x3 were also applied to further smoothen the image without distorting the defect. The modified image is shown in *Fig. 11.*

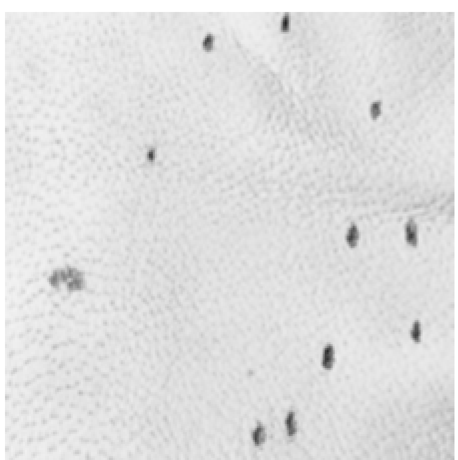

Fig. 11. Output Image after preprocessing input image

Adaptive Gaussian Thresholding was used so that defected regions are clearly identified in a normalized image. This image is then inverted as Blob detectors like LoG identifies lighter shades more effectively. The modified image is shown in *Fig. 12.*

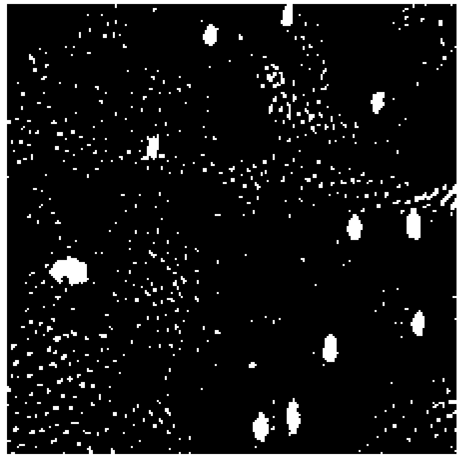

Fig. 12. Output image after Thresholding and Inversion

This image is then taken as input for Blob detectors LoG, DoG and DoH and the number of blobs(defects) are counted. The final output is shown in *Fig. 13.*

```
Number of Defects in LoG= 12
Number of Defects in DoF= 16
Number of Defects in DoH= 11
```

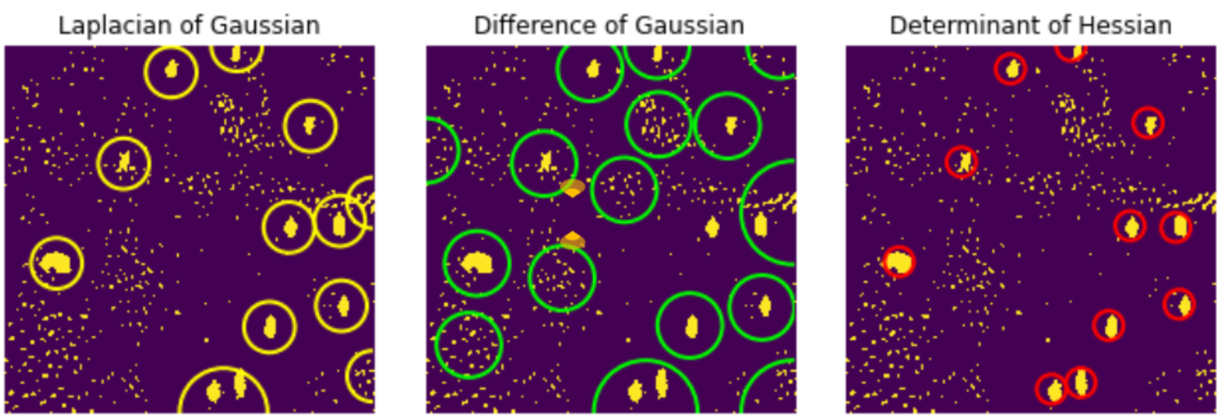

Fig. 13. Result of Blob Detection

3. CONCLUSION

Leather Defect detection is not easy as it involves identifying defects that don't have a regular shape or size. We initially looked for various Deep Learning models to detect the presence of anomalies and came across the implementation of Convolutional Neural Networks, which gave very good results in object detection.

We used a modified version of VGG which plots heatmaps of the images by comparing good images from the dataset with the defective images and then constructing bounding boxes around a defect (if any).VGG could identify the defect in leather with ~**93% accuracy.**

To improvise this model we simultaneously applied a Blob Detection model on pre-processed images of leather to count the number of defects in a specified area of the leather . This model gave us the number of defects with **~85.3% accuracy.**

We can use the above-mentioned techniques to automate the detection of defects in leather. This will enhance the process of sorting leather through machines by filtering the goods having defects above a certain threshold value from the goods with very less or no defects.

4. REFERENCES.

[1] Basu, S., Mukhopadhyay, S., Karki, M., DiBiano, R., Ganguly, S., Nemani, R., & Gayaka, S. (2018). *Deep neural networks for texture classification.*

[2] P. L. Bartlett, W. Maass, *Vapnik-chervonenkis dimension of neural nets (2003)*

[3] *MVTec Anomaly Detection Dataset - MVTec AD*. (n.d.). MVTec Software. Retrieved July 20, 2022, from https://www.mvtec.com/company/research/datasets/mvtec-ad - *MVTec AD, MVTec Software.*

[4] *Explainable Defect Detection Using Convolutional Neural Networks: Case Study*. (n.d.). Towards Data Science. Retrieved July 20, 2022, from https://towardsdatascience.com/explainable-defect-detection-using-convolutional-neural-networks-case-study-284e57337b59 *Case Study*

[5] *INDIAN LEATHER INDUSTRY | Council For Leather Exports*. (n.d.). Council for Leather Exports. Retrieved June 21, 2022, from https://leatherindia.org/indian-leather-industry/

[6] Jawahar M. (2016). Compression of leather images for automatic leather grading system using multiwavelet. *2016 IEEE international conference on computational intelligence and computing research (ICCIC)*..

[7] *Blob detection*. (n.d.). Wikipedia. Retrieved July 20, 2022, from https://en.wikipedia.org/wiki/Blob_detection

[8] *Lindeberg, T. Scale Selection Properties of Generalized Scale-Space Interest Point Detectors. J Math Imaging Vis 46, 177–210 (2013). https://doi.org/10.1007/s10851-012-0378-3*. (n.d.).

5. GLOSSARY

Convolutional Neural Network (**CNN** or **ConvNet**) is a class of Artificial Neural networks (**ANN**) most commonly applied to analyze visual imagery

Kernel is a small matrix used for blurring, sharpening, embossing, and edge detection by doing a convolution between the kernel and the image

Resolution refers to the number of pixels displayed per inch of an image and is usually described in PPI

Image Filtering is a technique through which size, colours and other characteristics of an image are altered. An image filter is used to transform the image using different graphical editing techniques

Hyperparameter is a parameter whose value is used to control the learning process

Cross-validation is a resampling method that uses different portions of the data to test and train a model on different iterations

Padding refers to the number of pixels added to an image when it is being processed by the kernel of a Convolution Neural network

Heat map represents coefficients to visualize the strength of correlation among variables and transforms the correlation matrix into colour coding.

Thresholding is a type of image segmentation, where we change the pixels of an image to make the image easier to analyze.

Support Vector Machine(SVM) A machine learning algorithm used for both regressions and classification

Vapnik-Chervonenkis Dimensions (VC) can be defined as the cardinality of the largest set of points that the algorithm can shatter, which means the algorithm can always learn a perfect classifier for any labelling of at least one