International Journal of Scientific Research & Engineering Trends
Volume 8, Issue 6, Nov-Dec-2022, ISSN (Online): 2395-566X

Automated Product Recognition for Retail Shopping from Video Imaging Using Machine Learning

Sanghita Datta, Ankita Sah, Upamita Das, Debmitra Ghosh, Aman Malhotra
datttasangy@gmail.com, as6729274@gmail.com, dasupamita143@gmail.com, debmitra.ghosh@jisuniversity.ac.in, amanmalhotraa5@gmail.com
Computer Science and Engineering
JIS University, Kolkata, India

Abstract- The key factor to increase the profit in grocery stores now-a-days is the availability of items on the shelf. The growing market of computer vision has made it possible for the grocery stores to grow in various aspects. To tackle is growing market of on shelf detector, our model has been designed where the products kept on the shelf would be scanned and their recognition would be done in the computer screen using machine learning for the training of data onto the model. This study examines the creation of a real-time, video-based action recognition system for removing items from shelves and putting them back. In order to prevent the two classification components from operating continually, the system also includes a detector component. The action classification component of the system is evaluated to have an accuracy of 80 percent, and the object identification component of the system to have an accuracy of 70 percent.

Index Terms-Machine Learning, Dataset, Video Imaging, GAN.

I. INTRODUCTION

The goal of product identification is to improve customer shopping experiences by streamlining the management of retail inventory. The most prevalent technology nowadays is barcode recognition, not just in research but also in industries that use automatic commodity identification. A barcode is printed on every product package, and it can be scanned to help with product administration. It is common for almost every product on the market to have a corresponding barcode. However, it frequently requires time to manually identify the barcode and assist the device in recognising it at the checkout counter due to the ambiguity of the barcode's printing location.

Companies are concentrating more and more on how to use artificial intelligence technology to revolutionise the ecology of the retail business and merge online and offline experiences as the sector rapidly changes. Retailers will spend 12billiononAIservicesgloballyin2023, upfrom3.6 billion in 2019, according to a Juniper Research study. In other words, the future of retail may fully materialise thanks to artificial intelligence technologies. Furthermore, when

living standards grow, shoppers and supermarket personnel are greeted by an excess of retail products. [15]
Machine Learning has become the key behind the main profits of many fields in today's world. One of such field is the shelf detecting of products. In this the profit has doubled due to the detection of the products as many a times the items available in the store is not available on the shelf, whereas in the presence of such a model this issue is rare to take place. Whenever a product is not available on the shelf the item is shown to be "out of stock", which usually causes a negative impact on the customers. It has been seen in the research of Corsten and Gruen that 31 percent of the consumers tend to buy items from other stores on not finding the product in one store. For this increasing issue the shelf item detector has been modeled and has helped the retailers to increase their profits. Moreover, it also helps the workers in the retail shop to keep a track of the items in much easier way. It also requires human efforts as they need to keep a constant track of the computer screen where the items are been scanned and the output is shown, thus giving employment to lots of people. Additionally, when living standards rise, supermarket em- ployees and patrons are met with an abundance of retail goods. Throughout this instance, managing the goods required a large quantity of human labour and a sizable fraction of the effort to identify the things. In addition, the quantity of digital resources for product images is growing significantly every day as a result of the abundance of photo-taking electronic devices. How to effectively analyse and process the enormous amount of visual data while simultaneously being able to recognise and categorise the products found in supermarkets presents a key research challenge in the field of product recognition.

The retail industry is significantly impacted by the imple- mentation of automatic product recognition using photographs in grocery stores. It will first aid in the compliance of the products on the shelf with the planogram. By recognising which products are missing from the shelf, for example, product detection enables

International Journal of Scientific Research & Engineering Trends
Volume 8, Issue 6, Nov-Dec-2022, ISSN (Online): 2395-566X

store staff to be alerted to replenish the products as soon as possible. The evidence indicates that when an optimum planogram is 100percent matched, sales will grow by 7.8percent and profits by 8.1percent. By integrating automatic self-checkout systems with image-based commodity identification, checkout procedures can be made more user- friendly. Worldwide SCO shipments have steadily increased from 2014 to 2019. SCOs are being used more commonly to reduce costs for retailers and enhance customer service. According to the research, the length of time customers must wait for checkout procedures has a negative impact on the how pleasant their shopping is, thus integrating computer vision- based product recognition into SCOs is beneficial for both merchants and customers. Lastly, item recognition technology enables visually challenged individuals to shop independently, enhancing their social connectivity. Because it can be challeng- ing for a person who is visually impaired to recognise things by their visual qualities (e.g., price, brand, and due date), buying selections can be challenging, traditional shopping methods typically require assistance from a sighted person.

One difficult example of problems with image categorization and item detection in general is problems with merchandise recognition. Particularly in the area of computer vision, deep learning has emerged during the past 10 years as the go- to technique for image categorization and object detection. Deep learning may directly learn features from picture data rather than depending on manually constructed features, which distinguishes it from traditional methods for pattern identifi- cation. Another element that contributes to deep learning's outstanding ability is its deeper layers, which can extract more precise characteristics than traditional neural networks. Deep learning approaches can offer unique solutions to various important computer vision problems, such as picture separa- tion and major point recognition, according to the advantages outlined above. A few recent projects in the retail industry have produced ground-breaking results.

Numerous studies have been done to automate shelf item monitoring. These papers examine the topic from several an- gles. One of them suggests using radio frequency identification (RFID) tagging to keep track of how many products are on the shelf, however this method is not practical financially to use the technology and incorporate it into current systems. Some of them employed deep learning techniques to identify objects on the shelves, while others used more conventional image processing methods to determine whether a product was present or not. There are benefits and drawbacks to these methodologies when all the prior works are considered. Tradi- tional image processing techniques, including the histogram of oriented gradient for feature extraction and the support vector machine for a classifier, have limited performance even on big data sets, and it is difficult to improve their capabilities.
[1] Additionally, misclassification can result from the aesthetic resemblance of the many products from the same brand. On the other hand, high accuracy might be reached when deep learning techniques like recurrent convolutional neural network or you only look once were utilised. The largest

and most well-known computer vision datasets do not contain annotation for shop products, hence a significant amount of work is required to manually identify products on photos in order to train the DL algorithms, which require annotated images. [9].

II.O BJECTIVE

Machine learning has turned to become the key component of most of the applications now. This model is one such device following the norms of machine learning . The shelf item detector is a device that keeps a track of the items kept on the shelf and has aided the retailers in boosting their earnings. Additionally, it makes it much simpler for the retail store employees to keep track of the merchandise. It also demands human effort because workers must constantly monitor the computer screen where the objects are scanned and the results are shown, creating jobs for many individuals. [1] [10] From a new angle, clients must appreciate and have confidence in the built-in detector model, which is a form of artificial intelligence. Additionally, these apps must be handled in accordance with their specifications if their needs change. The first explainable artificial intelligence system on a shelf item detection system is proposed in our study.

III.M ATERIALS AND METHODS

1.Product Recognition Based on Machine Learning:
Machine learning-based product recognition studies on ob- ject detection has advanced rapidly as a result of machine learning. In this work, we view the research question of product recognition as being specifically connected to object detection. Although computer vision is being used extensively, its application for item image recognition is still far from flawless. An object detector was generally used to acquire a set of bounding boxes as local proposals. The original image, which features many products, is then cropped to create several photos of a single item. The recognition of the products is now an image classification task since each cropped image may be fed into the classifier.

2.Data set used in training:
Simply described, a data set in machine learning is a group of data points that can be evaluated and predicted by a computer as a single entity. This means that since machines don't see data in the same way that people do, the data collected should be standard and intelligible. For

this, it's crucial to preprocess the data by cleaning, filling it out, and annotating it with relevant tags that a computer can understand once it has been collected. A decent data set should also meet specific quality and quantity requirements. We ensured our data set is pertinent and evenly distributed in order to facilitate quick and easy training. [3] However, in this project we have used a dataset of various biscuits that can be available in the usual retail shops for selling. This collection of data will thus help to classify among a variety of biscuits that are available on the shelf of the retail shop.

Fig. 1 Example of the data set collected.

1.Generative Adversarial Network model

A Generator and a Discriminator are the two agents in a typical GAN framework that are in competition with one another. Different networks, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), or simply Regular Neural Networks (ANNs or RegularNets), may be used to design them. CNNs are more appropriate for the task because we will generate images. As a result, we will use convolutional neural networks to create our bots. [5]

1.GAN Model- In effect, we will instruct the generator to produce images without receiving any more information. In parallel, we will retrieve the already-existing photos and give them to the discriminator, asking it to determine whether the images produced by the Generator are real or not. The Generator will first produce poor photos that the Discriminator will immediately identify as false. Due to the decreased diversity from the real photos, the Generator will eventually learn to mislead the Discriminator after receiving enough data from it. As a result, we will get a very good generative model that can produce results that are really realistic. [5].

2.GAN in shelf item detector- For the following generation of images, the GAN model used in this shelf item detector is DCGAN (Deep Convolutional Generative Adverserial Net- work).

3.DCGAN- Convolutional and convolutional-transpose layers are utilised by the DCGAN's generator and discriminator, respectively. The discriminator in this case is made up of the batch normalisation layers, strided convolution layers, and LeakyRelu as the activation function. The provided image is 3x64x64 in size. Convolutional-transpose layers, ReLU activations, and batch normalisation layers are all included in the generator. The end result will be a 3x64x64 RGB image. [6]

4.DCGAN Architecture- The Generator and Discriminator models are combined to form a Generative Adversarial Net- work (GAN), which turns them into competitors in a zero- sum game. To trick the Discriminator into thinking they are real, the Generator fabricates images. The Discriminator is simultaneously learning the key characteristics of the im- ages in order to distinguish between authentic and fraudulent samples. [6] The interaction between the two models inside the DCGAN architecture is shown in the diagram below.

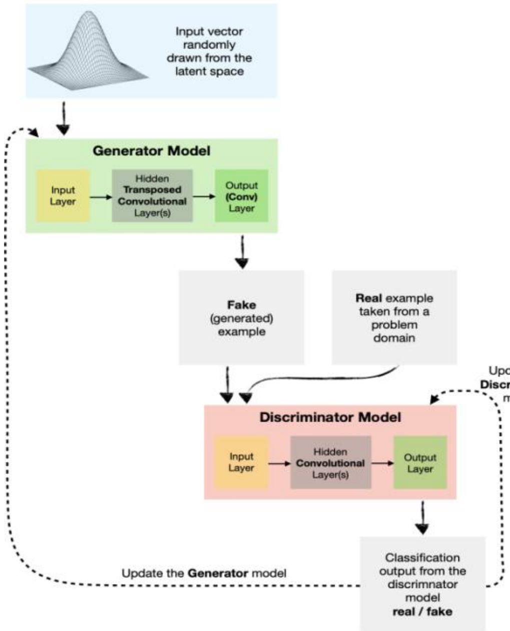

Fig 2 DCGAN Model Architecture.

A convolutional classification model is all that the dis- criminator model is. The Generator model, in comparison, is more complicated since it develops the ability to use regular and transposed convolutions to turn latent inputs into actual images. [6]

5. Proposed Generator- Starting with a 100-node latent vec- tor, we reshape it to 8 x 8 x 128 before connecting it to the Dense layer's 8192 nodes. The data is then upsampled to an output size of 64 × 64 by passing it via Transposed Convolutions. Because we limit the number of filters from 512 to just 3, which represent the various colour channels, observe that we also employ ordinary Convolution in the output layer. [6]

```
Model: "Generator"
_________________________________________________________________
 Layer (type)                 Output Shape              Param #
=================================================================
 Generator-Hidden-Layer-1 (D  (None, 8192)              827392
 ense)

 Generator-Hidden-Layer-Resh  (None, 8, 8, 128)         0
 ape-1 (Reshape)

 Generator-Hidden-Layer-2 (C  (None, 16, 16, 128)       262272
 onv2DTranspose)

 Generator-Hidden-Layer-Acti  (None, 16, 16, 128)       0
 vation-2 (ReLU)

 Generator-Hidden-Layer-3 (C  (None, 32, 32, 256)       524544
 onv2DTranspose)

 Generator-Hidden-Layer-Acti  (None, 32, 32, 256)       0
 vation-3 (ReLU)

 Generator-Hidden-Layer-4 (C  (None, 64, 64, 512)       2097664
 onv2DTranspose)

 Generator-Hidden-Layer-Acti  (None, 64, 64, 512)       0
 vation-4 (ReLU)

 Generator-Output-Layer (Con  (None, 64, 64, 3)         38403
 v2D)

=================================================================
Total params: 3,750,275
Trainable params: 3,750,275
Non-trainable params: 0
```

Fig .3 Proposed Generator.

Proposed Discriminator: We will see that the Discriminator model performs the opposite of what the Generator model does. In other words, a 64 x 64 image is reduced to a binary classification output of "real" or "fake" by being processed through several Convolutional layers. The DCGAN is then produced by combining the two models. We make the Discriminator model untrainable, which is a vital component of the code following. We do this action in order to train the Discriminator independently utilising a mixture of genuine and fictitious (made) data. [6]

```
Model: "Discriminator"
_________________________________________________________________
 Layer (type)                 Output Shape              Param #
=================================================================
 Discriminator-Hidden-Layer-  (None, 32, 32, 64)        3136
 1 (Conv2D)

 Discriminator-Hidden-Layer-  (None, 32, 32, 64)        0
 Activation-1 (LeakyReLU)

 Discriminator-Hidden-Layer-  (None, 16, 16, 128)       131200
 2 (Conv2D)

 Discriminator-Hidden-Layer-  (None, 16, 16, 128)       0
 Activation-2 (LeakyReLU)

 Discriminator-Hidden-Layer-  (None, 8, 8, 128)         262272
 3 (Conv2D)

 Discriminator-Hidden-Layer-  (None, 8, 8, 128)         0
 Activation-3 (LeakyReLU)

 Discriminator-Flatten-Layer  (None, 8192)              0
  (Flatten)

 Discriminator-Flatten-Layer  (None, 8192)              0
 -Dropout (Dropout)

 Discriminator-Output-Layer   (None, 1)                 8193
 (Dense)

=================================================================
Total params: 404,801
Trainable params: 404,801
Non-trainable params: 0
```

Fig .4 Proposed Discriminator.

5. Video Imaging for Object Detection

Technology researchers and programmers have taken the decision to develop Video Object Detection (or VOD) appli- cations, which provide machines the ability to study images and detect the objects present in them. These applications were inspired by the function of the human visual cortex. They originally conceived procedures and processes designed to exclusively work with photos. Today, though, video images are the norm. [2] The way object detection works in video is very similar to how it operates in photos. Such a tool would enable the computer to find, recognise, and categorise things visible in the supplied moving images. [12] [13]

The machine needs to be fed with reference data in the beginning. A database of images and video must be compiled. Video datasets are accessible for download on a variety of plat- forms, including Keras and Tensorflow, or other APIs, much like it is for images (Application Programming Interface). Python is the most widely used language in AI nowadays, therefore users who want to create their own VOD application may wish to start with it because it is simple to understand and practise on. Upon downloading the entire dataset, the new user must assign class labels to the objects visible in the moving photos in order to categorise them.

Bounding boxes will need to be manually drawn around the things we want the computer to recognise in order for the application to be able to locate them. [4] [12] Some already-set up object recognition techniques can be obtained via various platforms after having found and impacted items with matching labels. In contrast to image processing, video on demand analyses the data it must process using a combination of image detection and video tracking. [14] The algorithm needs to be trained on the labelled data, which was entered at the very beginning of the procedure, before being validated. To make sure that the precision is sufficient, this step is crucial. After the training phase is through, the programme is prepared for evaluation and validation on a fresh set of data that it has never encountered before, before being made public. [4]

IV. IMPLEMENTATION

1. Model used in identification

The model used in the Shelf detection model is DenseNet 121 and a self collected dataset has been saved in a file and the link of which is used in the model for training the model. DenseNet 121: By leveraging shorter connections between the layers, the DenseNet (Dense Convolutional Network) de- sign aims to increase the depth of deep learning networks while also improving training efficiency. A convolutional neural network called DenseNet connects every layer to every other layer below it. [7] [8] For example, the first layer is connected to the second, third, fourth, and so on levels, and the second

International Journal of Scientific Research & Engineering Trends
Volume 8, Issue 6, Nov-Dec-2022, ISSN (Online): 2395-566X

layer is connected to the third, fourth, fifth, and so on layers.

In order to maximise information flow between network tiers, this is done. Each layer receives input from all the layers that came before it and transmits its own feature maps to all the layers that will follow it in order to maintain the feed-forward nature. Contrary to Resnets, it concatenates the features rather than combining them by summation. As a result, the ”ith” layer contains I inputs and is made up of feature maps from all the convolutional blocks that came before it. All of the subsequent ”I-i” layers receive its own feature map information. Instead of only ’I’ connections as in conventional deep learning architectures, this adds ’(I(I+1))/2’ connections to the network. As a result, it needs fewer parameters than conventional convolutional neural networks because no unimportant feature maps need to be learned. [7] [8]

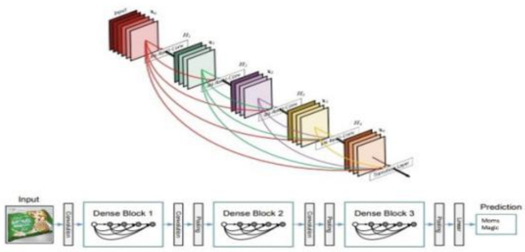

Fig. 5 DenseNet 121.

2. Training Procedure

To help us with the sampling and generation of data for the two models, we will develop three straightforward functions. In the first, genuine images are taken as samples from the training data, in the second, random vectors are drawn from the latent space, and in the third, latent variables are fed into a generator model to produce fictitious examples.

The last two functions will aid in model training and interval outcomes evaluation.

1. The model performance evaluation function is first created.
2. The training function is created by us.

As previously noted, we train the Discriminator independently by running a batch that contains 50percent real and 50percent created samples. A hybrid DCGAN model is used for the Generator training in the meantime.

3. Working of the Model-

The working of the model follows the following steps:

1. The model begins with collection of data set manually as we have used real time images of biscuit packets as data set. There were about 550 images of various biscuits were collected.
2. The collected 550 images of biscuit packets were then trained into the GAN model where the DCGAN (Deep Convolutional Generative Adverserial Network) was used for generating more than 5000 images from the trained images. 500 images of every kind of biscuit packets were generated by the DCGAN model at 30,000 epochs, which took about 96hours of duration. More epochs was ran for the trained images, more was the accuracy.
3. The images of biscuit packets were then processed into the model for feature extraction, classification and localization. When an algorithm analyses photos to find data insights or support automated operations in computer vision use cases, this is image processing.
4. The trained image of biscuit packets then underwent feature extraction where the features of the biscuit packet(like color, size, shape, etc.) was checked and sent for classification purpose. The technique of turning raw data into numerical features that can be handled while keeping the information in the original data set is feature extraction.
5. The collected data set of biscuit packets were then fit into the densenet 121 model for image localization and classification. By leveraging shorter connections between the layers, the DenseNet (Dense Convolutional Network) design aims to increase the depth of deep learning networks while also improving training efficiency. Thus, giving a higher accurate result.
6. Lastly, on executing the model, the output of the recognition of biscuit packet was 97 percent accurate. The model could successfully detect which company the biscuit packet belonged to. Following is the flowchart of the proposed model.

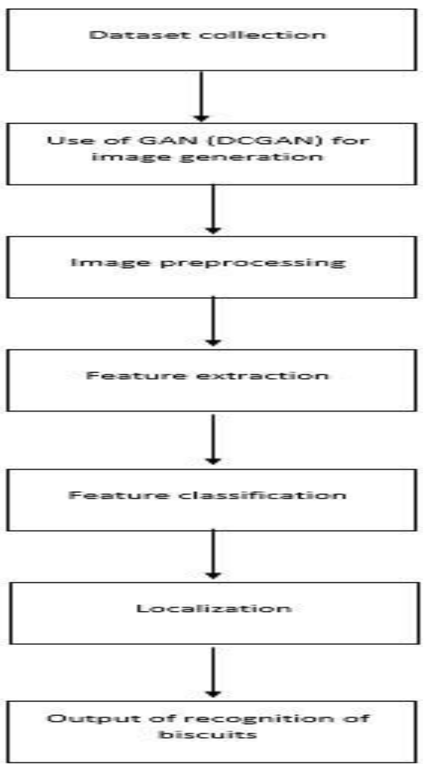

Fig 6: Flowchart of the model.

International Journal of Scientific Research & Engineering Trends
Volume 8, Issue 6, Nov-Dec-2022, ISSN (Online): 2395-566X

V.RESULTS

Firstly, we used a system of 4 GB GPU, 8GB RAM and a descent CPU @ 2.50GHz clock-speed for generating the images. The biscuit images collected by us was trained into the GAN model for generating 500 images of each kind of biscuits. We chose 11 different kinds of biscuits, example Oreo, Parle-G, Marie gold, Cream Cracker, etc. These biscuits were trained one at a time and the model individually worked on each kind and generated 500 images of size 64x64 for each biscuit type, thus generating 5500 images in total for training the image detection model. It took approximately 96hours for generating the images at 30,000 epochs. After every biscuit was trained and the 500 images were generated, the model refreshed itself for the new type of biscuit and for generating the respective images. Thus, more epochs being used in the model means more accuracy of the images thus generated. Below are the images of one of the biscuits at 1000 epochs and 30,000 epochs respectively.

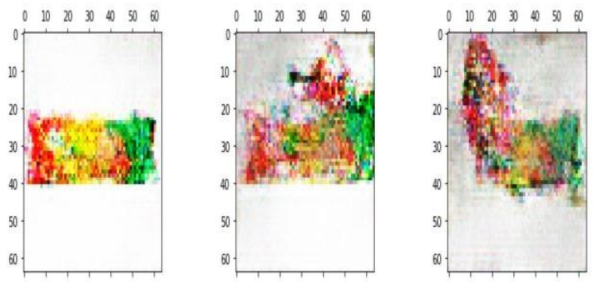

Fig. 7 Biscuit image at 1000 epochs.

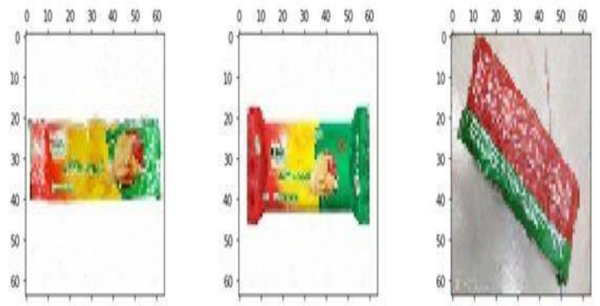

Fig. 8 Biscuit image at 30,000 epochs.

After generating the images, a dataset of 5500 images were trained into the Densenet 121 model for identification of the biscuit. The output thus given was at 97 percent accuracy. In this the model recognises which image belongs to which company (like oreo, marie gold, cream cracker, etc.) and gives an output respective of the serial number of the biscuit type list. Below is the output of the Densenet 121 model recognising the type of biscuit.

```
for img in os.listdir(DIRECTORY_train):
#label = 0 means parleG
#label = 1 means Mom's Magic
#label = 2 means Oreo
#label = 3 means Rich Marie
#label = 4 means Cream Crack
#label = 5 means Good Day
#label = 6 means Happy Happy
#label = 7 means Marie Gold
#label = 8 means Malkist
#label = 9 means Nutrichoice
#label = 10 means Sugar Free
```

Fig. 9 List of Types of Biscuits

```
1/1 [==============================] - 0s 53ms/step
[4.9719867e-01 7.5928801e-05 9.9986774e-01 3.5411620e-03 2.0064792e-01
 9.9611884e-01 1.7447805e-02 2.7977332e-01 5.5641842e-01 8.4585719e-02
 6.7067993e-01]
2
```

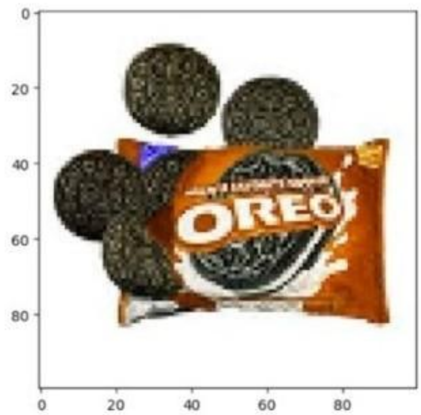

Fig. 10 The biscuit is Oreo (serial no. 2)

After classification of the images the localization is also performed on a collection of biscuits where the various colours of boxes appear on the respected areas and localizing them as their respective brand names and if the image does not match any of the trained images then the box is labelled as 'Invalid'.

Fig. 11 Localization of the biscuits.

VI. SUMMARY AND FUTURE WORK

The broad topic of product recognition technologies is covered in this article. In a world where customers are under more pressure to use up their available time and cost margins are getting tighter, product recognition will become more and more crucial. We advance the subject by making research in this area more approachable to new scholars by summarising the relevant literature. It is crucial that this field tackles these four difficult issues: Large-scale categorisation, data constraints, intraclass variance, and flexibility are among the issues. One of the primary strategies for boosting revenues in grocery stores is to offer high on-shelf availability.

The tagged picture data is used to develop a classifier in the conventional Shelf item detection apps for this purpose. But much of the data created in real life is unnamed, and manually classifying the items on the photos is a time-consuming and expensive procedure. This research offers a novel solution to this issue that relies on the densenet model and actively employs GAN for image production. The automatic integration of the avail- able unlabeled picture data into the model using the suggested way solves the issue. Our suggested solution uses a small quantity of labelled data and a huge amount of unlabeled data to identify empty and nearly empty shelves depending on each product type. The experiment that was conducted by us gave an accuracy of 100 percent, which shows that our model is better than many other proposed solutions for the similar problem statement mentioned. [11]

Presently, the built Shelf item detection software application explains the results of the suggested approach using rule-based interpretation. In the future, a deep learning-based translation may be created in place of rule-based translation. There are a number of areas that need more research, including data gen- eration with deep neural networks, graph neural networks with deep learning, cross-domain recognition with transfer learning, two feature learning from text information on packaging, incremental learning with the CNN, and regression-based object detection techniques for retail product recognition.
Additionally, some virtual things connected to the goods recognised on the shelves can be rendered using augmented reality technology. Also, using video imaging, this software might eventually be developed into a piece of hardware for usage in malls or retail stores.

ACKNOWLEDGMENT

We wish to express our sincere appreciation to all those who have contributed to this thesis and supported us in one way or the other during this amazing journey. We express our heartfelt thanks to Dr. Indranil Sengupta (on leave professor of IIT KGP), Vice Chancellor, JIS University, Kolkata, India, for giving us this opportunity to do this project. Special thanks to Dr. M Bhattacharya, an ISI professor, for helping us with our project. We also thank Aman Malhotra, CEO, Proxie Studio Pvt. Ltd., for being a constant guide throughout our project.

REFERENCES

[1]Y. Ramiz, B. Derya, "Shelf Auditing Based on Image Classification Using Semi-Supervised Deep Learning to Increase On-Shelf Availability in Grocery Stores," MDPI Publishers, Sensors 2021, 21, 327.

[2]H. W. Irfan, D. H. Widyantoro, "Development Of Video Based Action Recognition System For Item Taking From Shelf," 2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA), Electronic ISBN:978-1- 6654-1743-3, Print on Demand(PoD) ISBN:978-1-6654-1744-0, DOI:10.1109/ICAICTA53211.2021.9640283, 16 December 2021.

[3]S. Iryna , "What Is a Dataset in Machine Learning: Sources, Features, Analysis", Editor-at-Large, April 4, 2021.

[4]Anonymous, "Video Object Detection: AI's New Challenge", Kili Arti- cles, February, 2022.

[5]G. Y. Orhan, "Image Generation in 10 Minutes with Generative Adver- sarial Networks", Towards Data Science, Sep 18, 2020.

[6]D. Saul, "Deep Convolutional GAN — How to Use a DCGAN to Generate Images in Python", Towards Data Science, Jul 14.

[7]S. Arjun , "Creating DenseNet 121 with TensorFlow", Towards Data Science, Jul 11, 2020

[8]H. Gao, L. Zhuang, M. D. V. Laurens, W. Q. Kilian, "Densely Connected Convolutional Networks", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Electronic ISBN:978-1-5386-0457-1, Print on Demand(PoD) ISBN:978-1-5386-0458-8, ISSN: 1063-6919, 09 November 2017.

[9]S. Gurpreet, "Grocery product Recognition with Supervise.ly and Mask R-CNN", Medium publishers, Jul 7,2020.

[10]T. Alessio, S. Eugenio, S. D. Luigi, "A deep learning pipeline for product recognition on store shelves", 2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS), DOI:10.1109/IPAS.2018.8708890, December 2018.

[11]C. Chao, "On-device Supermarket Product Recognition", Google Re- search, AUGUST 11, 2020.

[12]P. Marina, S. Mirco, M. Adriano, Z. Primo, F. Emanuele, "Mobile robot for retail surveying and inventory using visual and textual analysis of monocular pictures based on deep learning", 2017 European Confer- ence on Mobile Robots (ECMR), DOI: 10.1109/ECMR.2017.8098666, Electronic ISBN:978-1-5386-1096-1, Print on Demand(PoD) ISBN:978- 1-5386-1097-8, 09 November 2017.

[13]D. Praveen, ”Robots in Retail Marketing: A Timely Opportu- nity”, Global Disclosure of Economics and Business 9(2):97-106, DOI:10.18034/gdeb.v9i2.527, November 2020.

[14] De Gauquier, L., Brengman M. Willems K. (2020), “The Rise of Service Robots in Retailing: Literature Review on Success Factors and Pitfalls”, Retail Futures: The Good, the Bad and the Ugly of the Digital Transformation. Pantano, E. (ed.). First edition ed. Emerald Group Publishing Ltd., pp. 15-35.

[15] Hampshire, “AI spending by retailers to reach 12 billion dol- lars by 2023, driven by the promise of improved margins,” April 2019, https://www.juniperresearch.com/press/ press-releases/ai- spending-by-retailers-reach-12-billion-2023.