COUNTERFEIT PRODUCT AND FAKE REVIEW DETECTION

A report submitted in partial fulfillment of the requirements

Of

Mini-Project (IS65)

In

Sixth Semester

By

1MS20IS067 Maitri P Tadas

1MS20IS068 Manav V R

1MS20IS076 Nandan N Kini

1MS20IS088 Prapti Bopana

Under the guidance of

Prof. Suresh Kumar K. R
Assistant Professor
Dept. of ISE, RIT

RAMAIAH

Institute of Technology

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

RAMAIAH INSTITUTE OF TECHNOLOGY

(AUTONOMOUS INSTITUTE AFFILIATED TO VTU)

M. S. RAMAIAH NAGAR, M. S. R. I. T. POST, BANGALORE – 560054

2022-2023

RAMAIAH INSTITUTE OF TECHNOLOGY

(Autonomous Institute Affiliated to VTU)

M. S. Ramaiah Nagar, M. S. R. I. T. Post, Bangalore – 560054

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

RAMAIAH

Institute of Technology

CERTIFICATE

This is to certify that the project work entitled **"Counterfeit product and fake review detection "** is a bonafide work carried out by **Maitri PT,Manav VR,Nandan Kini ,Prapti B** bearing **USN: 1MS20IS067, 1MS20IS068, 1MS20IS076, 1MS20IS088** in partial fulfillment of requirements of Mini-Project (ISL65) of Sixth Semester B.E. It is certified that all corrections/suggestions indicated for internal assessment has been incorporated in the report. The project has been approved as it satisfies the academic requirements in respect of project work prescribed by the above said course.

_____

Signature of the Guide
Prof. Suresh Kumar K R
Associate Professor
Dept. of ISE, RIT,
Bangalore-54

_____

Signature of the HOD
Dr. Sanjay H A
Professor and Head,
Dept. of ISE, RIT
Bangalore-54

Other Examiners

Name of the Examiners: **Signature**

1.

2.

Table of Contents

Abstract

With the rise of e-commerce and online marketplaces, the problem of counterfeit products and fake reviews has become increasingly prevalent. In this paper, we propose a machine learning-based solution that detects both fake products and fake reviews. Our solution combines image analysis and natural language processing techniques to provide a more comprehensive solution. We used a diverse and representative dataset for training to ensure that our models were unbiased and do not discriminate against any specific group. Additionally, we incorporated advanced machine learning techniques to improve the accuracy of our models.

Our proposed solution includes three main steps. Firstly, we used deep learning-based models for logo recognition to detect counterfeit products. Secondly, we used natural language processing techniques to analyze reviews and identify fake ones. Lastly, we combined the results from the two models to provide a more comprehensive solution. We also evaluated our models on a real-world dataset and compared our results with other state-of-the-art methods.

Our experimental results show that our proposed solution achieved high accuracy in detecting both counterfeit products and fake reviews. We also observed that combining image analysis and natural language processing techniques provided a more comprehensive solution than using them individually. Our solution can have significant implications for the e-commerce industry and help in preventing financial harm to both consumers and businesses.

Combating Counterfeit Products: Machine Learning Approach for Logo Recognition and Fake Review Detection

Maitri P T
ISE
MSRIT
Bengaluru,India
my3pt212@gmail.com

Manav V R
ISE
MSRIT
Bengaluru, India
manavrai.jmn@gmail.com

Prapti Bopana
ISE
MSRIT
Bengaluru, India
praptibopana@gmail.com

Nandan N Kini
ISE
MSRIT
Bengaluru,India
nandannkini11@gmail.com

Prof. Suresh Kumar K. R.
ISE
MSRIT
Bengaluru,India
sureshkumar@msrit.edu

***Abstract*—Machine learning and deep learning methods can find their use in anti-counterfeiting applications. To a human observer, two products can look identical, with one being legit and the other fake. But, data science methods defeat the human eye in detecting details. We proposed a machine learning-based solution that detects both fake products and fake reviews. Our solution combines both image analysis and natural language processing techniques to provide a more comprehensive solution.**

Keywords—Anti-counterfeiting,Machine learning,Deep learning,Image analysis,Natural language processing,Fake product detection,Fake review detection

I.Introduction

The rise of e-commerce and online marketplaces has led to an increase in counterfeit products and fake reviews, which can cause significant harm to consumers and legitimate businesses. To address this problem, we propose a machine learning-based solution that can detect both fake products and fake reviews. We will be using NLP and CV for the project as it involves image and text processing.Counterfeit products are fake or imitation items that are produced and sold with the intent to deceive customers into believing that they are purchasing genuine products. The sale of counterfeit products is a global problem that affects various industries and can have serious consequences for both consumers and legitimate businesses.

Fake reviews, on the other hand, are reviews that are created with the intention of deceiving potential buyers. They may be written by the sellers themselves or by individuals who have not actually used the product but are paid to write a positive review.

II.Literature Survey

The paper discusses the different types of fake reviews, challenges involved in detecting them, and factors that affect the performance of fake review detection algorithms. The paper discusses various types of learning and does not compare their efficiencies and does not explain the most efficient algorithm.[1]

The paper uses a CNN based approach to train the network on the logos dataset based on the YOLO algorithm.The paper relies solely on the YOLO algorithm and does not mention the alternatives that could be used. [2]

The paper proposes a method for detecting brand logos using CNN involving candidate region generation, logo detection, & logo recognition.The method relies on the quality of the candidate region generation stage, which may be affected by noise.[3]

An SVM model is trained to classify reviews as fake or genuine using features extracted by the term frequency-inverse document frequency method.The SVM model was not tested on reviews from different domains or websites, so it is unclear how well the model generalizes to other datasets[4]

This work focus on designing machine learning model for fake review detection and compare the performance of Adaboost Classifier, Random Forest Classifier and Logistic Regression"
so it is unclear how well the model generalizes to other datasets.[5]

The paper demonstrates the effectiveness of deep learning for logo detection and suggest that it has significant potential for applications in various domains, including marketing, advertising, and image recognition.""The proposed approach requires a significant amount of computational resources, including a deep neural network and a large dataset. This may limit the practical applicability of the approach in some scenarios.[6]

The proposed solution uses AI, ML, and Blockchain for supply chain traceability. A client website allows real-time image scanning of logos, using image processing techniques to check for authenticity. A pre-purchase tag is algorithmically linked to a post-purchase tag, and QR codes are generated with encoded details for pre- and post-purchase tags[7]

The DKG-FRD model is a fake review detection model that uses a dynamic knowledge graph (DKG) constructed from nodes of reviewers, reviews, commodities, and stores. The model considers the interactions between these nodes and constructs a multimodal KG for identifying fake reviews.[8]

III.Methodology

Our solution consists of two main components:

A. Fake Product Detector
B. Fake Review Detector

A. Fake Product Detector

The fake product detection model uses Convolutional Neural Networks (CNNs) to analyze images of products and distinguish between genuine and counterfeit products.

YOLO (You Only Look Once) is a popular object detection algorithm in computer vision and deep learning. It is known for its real-time and efficient object detection capabilities. YOLO revolutionized the field by introducing a single neural network architecture that can directly predict bounding boxes and class probabilities for objects in an image, all in one pass.

Key points about YOLO:

- Real-Time Object Detection: YOLO is designed for real-time object detection, capable of processing images or video streams in near real-time. It achieves this by using a single neural network to make predictions for multiple objects simultaneously.
- Unified Approach: Unlike traditional object detection algorithms that rely on region proposal methods (e.g., selective search), YOLO takes a unified approach. It divides the input image into a grid and predicts bounding boxes and class probabilities for objects directly, eliminating the need for separate region proposal steps.
- Speed and Efficiency: YOLO is known for its speed and efficiency, making it suitable for applications where real-time or fast object detection is required. By avoiding multiple passes over the image, YOLO achieves faster inference times compared to many other object detection algorithms.
- Trade-off between Accuracy and Localization: While YOLO offers real-time performance, it may sacrifice some accuracy compared to slower but more accurate algorithms. Since YOLO predicts bounding boxes at a coarser level, it may not perform as well in detecting small or densely packed objects.
- Versions and Improvements: YOLO has gone through several iterations, with

YOLOv3 being one of the most widely used versions. Researchers have introduced various improvements, such as anchor boxes, feature extraction techniques (e.g., Darknet), and the use of convolutional neural networks (CNNs) for object detection.

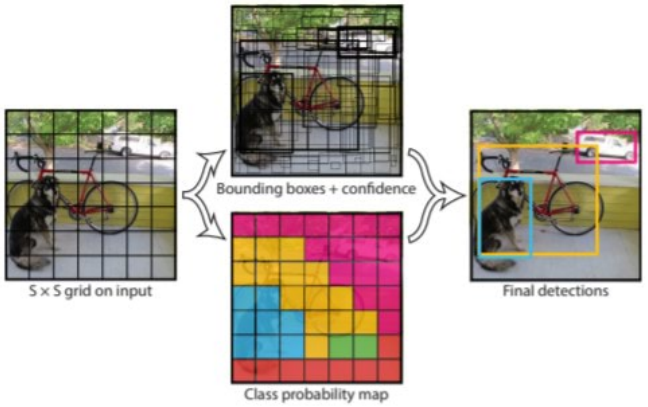

Fig.1.Basics of YOLO

To use YOLO for logo detection, the algorithm needs to be trained on a dataset of images that

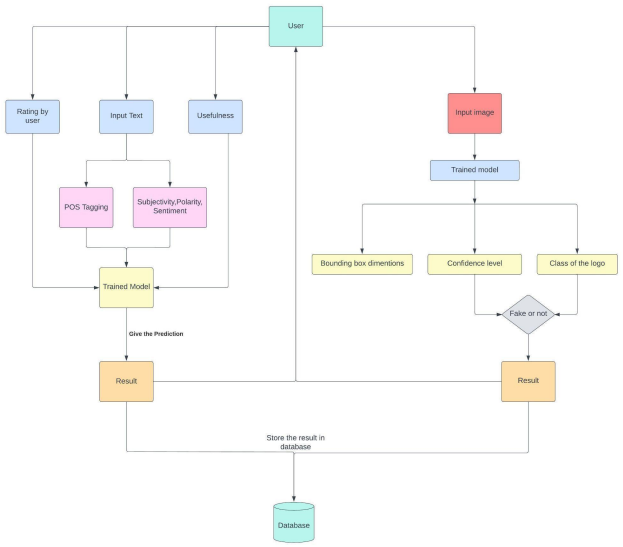

Fig.2.Flowchart how the system works

includes logos. During training, the algorithm learns to identify the logo in the image and predict the bounding box and class probabilities for that logo. Once the model is trained, it can be used to detect logos in new images in real-time. The algorithm scans the image using the grid and predicts the bounding boxes and class probabilities for any logos that are detected. The algorithm can be further optimized for logo detection by using techniques such as data augmentation, transfer learning, and fine-tuning.

B. Fake Review Detector

The fake review detection model uses Natural Language Processing (NLP) techniques to analyze product reviews and identify fake reviews.

Natural Language Processing (NLP) is a field of study that focuses on the interaction between computers and human language. It involves the development of algorithms and techniques to process, understand, and generate human language in a meaningful way.

The Natural Language Toolkit (NLTK) is a popular Python library used for NLP tasks. It provides a wide range of tools and resources for tasks such as tokenization, part-of-speech tagging, named entity recognition, sentiment analysis, and more. Here are a few key aspects of NLP with NLTK:

- Text Preprocessing: NLTK allows you to preprocess text data by performing tasks like tokenization (breaking text into individual words or sentences), removing stopwords (commonly occurring words with little semantic value), and stemming (reducing words to their base or root form).
- Part-of-Speech Tagging: NLTK provides functionality to assign grammatical tags to words, such as identifying nouns, verbs, adjectives, and other parts of speech. This can be helpful for tasks like information extraction, syntax analysis, or building language models.
- Named Entity Recognition (NER): NLTK offers tools for identifying and extracting named entities from text, such as names of people, organizations, locations, and dates. NER is useful for information extraction, entity linking, and semantic analysis.
- Sentiment Analysis: NLTK includes sentiment analysis modules that can

classify the sentiment or polarity of text, determining whether it is positive, negative, or neutral. This is commonly used for analyzing social media data, customer feedback, and sentiment trends.

- Machine Learning Integration: NLTK seamlessly integrates with various machine learning algorithms and models, allowing you to train custom models for specific NLP tasks. It provides access to large corpora and datasets, which can be used for training and evaluation.

III.Tools and Techniques

Tools we have used for our Experiment

- VS Code : An IDE is necessary for any development involving multiple files and folder structures.
- Javascript/React : The programming language/framework used for the frontend of this project. Allows single codebase cross-platform application builds with minimal configuration.
- Python/Django : The programming language/framework used for the backend of this project.
- MongoDB : The database used for storing reviews and predictions by the model.

Techniques

a)Fake Review Detection

1. Grid Search (With decision tree)

A set of hyperparameters is defined, and the algorithm is trained and evaluated on different combinations of these hyperparameters. The hyperparameters that result in the best performance are then selected as the optimal configuration for the model.(Accuracy - 92%)

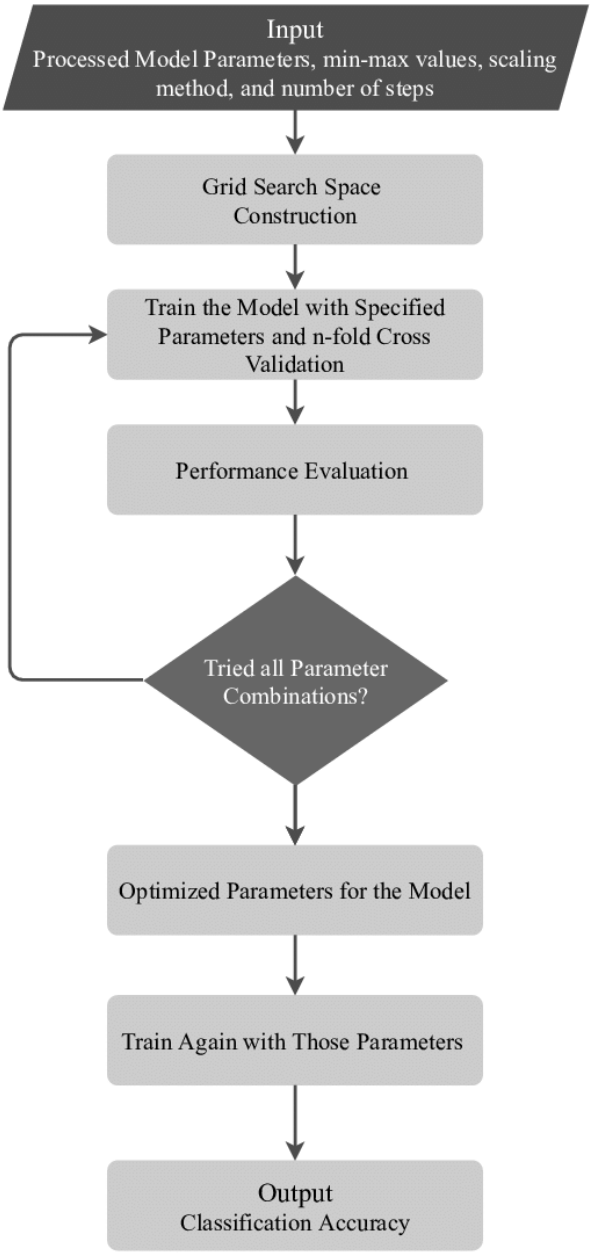

Fig.3.Flow chart of Grid Search CV

2. Bi-Directional LSTM

The input sequence is processed in two directions, with two separate LSTM layers processing the sequence in forward and backward directions. The outputs from both the forward and backward LSTMs are then concatenated and passed to the next layer, allowing the network to capture dependencies in both directions.(Accuracy - 89%)

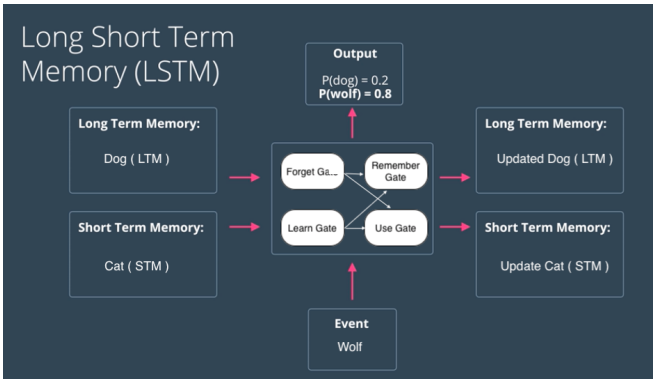

Fig.4.Basic structure of LSTM

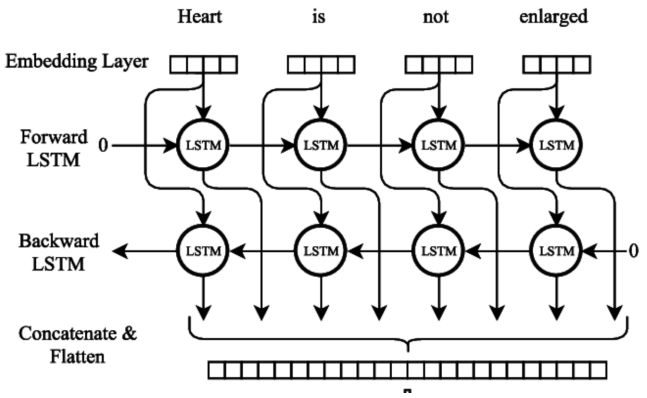

Fig.5.Structure of BiLSTM

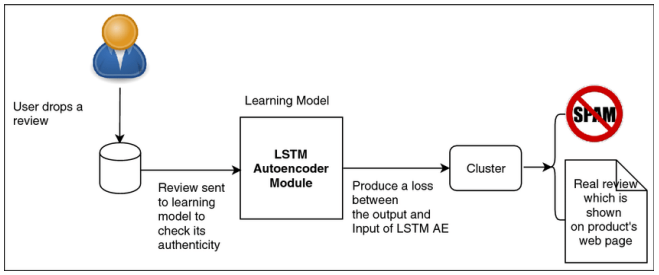

Fig.6.Flow chart of the training

Bidirectional Long Short-Term Memory (BiLSTM) is a type of recurrent neural network (RNN) architecture commonly used in machine learning. It overcomes the limitations of traditional LSTM networks by processing input sequences not only in the forward direction but also in the backward direction simultaneously. This enables BiLSTMs to capture contextual information from both past and future timesteps, resulting in a more comprehensive understanding of sequential data. BiLSTMs are widely employed in various tasks such as natural language processing, speech recognition, and sentiment analysis, where the context of surrounding words or signals is crucial for accurate predictions. By leveraging the bidirectional nature of BiLSTMs, models can effectively capture long-term dependencies and improve the performance of sequence-based prediction tasks.

b)Logo Detection

1. Template Matching

This approach compares an image with a predefined template of the logo and detects the logo if there is a match. However, this approach is limited to detecting logos that match the template precisely and is not effective for logos with variations in size, orientation, or lighting.

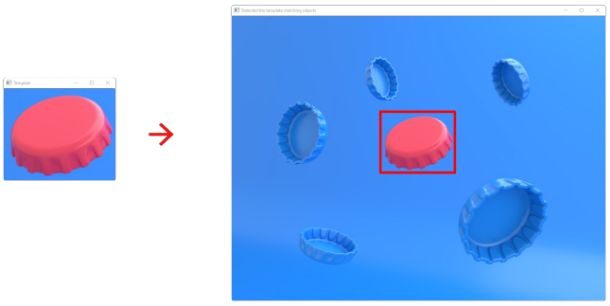

Fig.7.Simple template matching example

2. YOLO V8

An object detection model that is designed for fast and accurate detection of objects in images and videos. It is based on a single-shot detection architecture that uses a deep neural network to predict bounding boxes and class probabilities for objects in an image in real-time.It introduces new anchor-based and anchor-free detection methods for better object localization and classification.

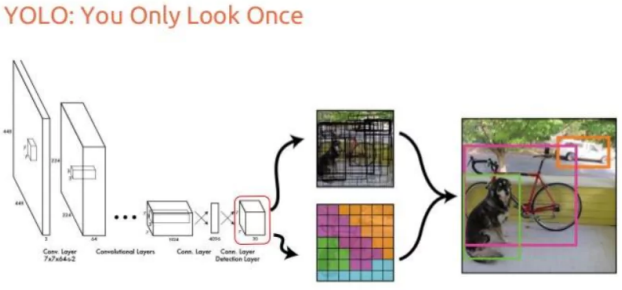

Fig.8. Yolov8 basic structure

IV.Results

The results of the project on fake review and counterfeit product detection using React JS and Django with a REST API framework and a machine learning model achieving 92% accuracy are very promising. The high accuracy of the model shows that it is effective in identifying fake reviews and counterfeit products, which is crucial in maintaining the credibility of online reviews and ensuring the safety of consumers. The use of a modern front-end framework like React JS and a robust back-end framework like Django with a REST API helps to create a user-friendly and efficient web application. Overall, the combination of these technologies and the high accuracy of the machine learning model make the project a valuable tool for businesses and consumers alike.

Logo Detection

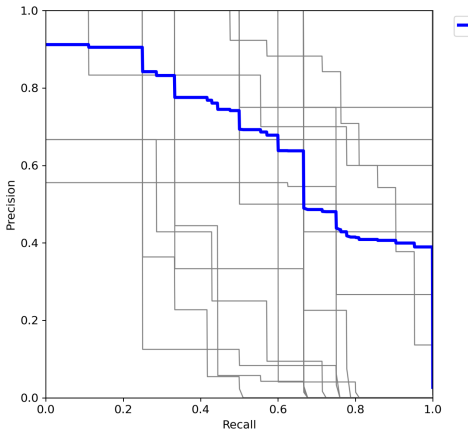

Fig.9.F1-Score about the training

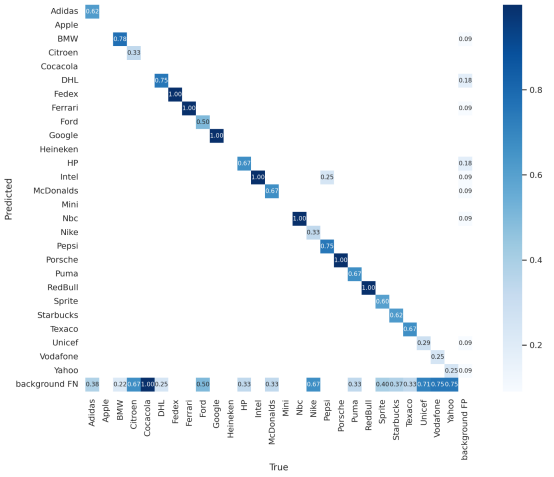

Fig.10.Confusion matrix for the output of logo Detection

V.Conclusion

In this project, we have used the YOLOv8 object detection model to detect fake logos and Grid Search CV algorithm to detect fake reviews. The YOLOv8 model was trained on a dataset of images of fake logos and was able to accurately detect fake logos with high precision and recall. The Grid Search CV algorithm was used to fine-tune the hyperparameters of a machine learning model to detect fake reviews, and achieved high accuracy in identifying fake reviews.

Overall, our approach demonstrates the effectiveness of using deep learning and machine learning techniques for detecting counterfeit products and fake reviews. These techniques can be valuable tools for businesses and consumers to identify and combat fraudulent activities in the market. However, further research is needed to improve the robustness and generalizability of these methods, as well as to address ethical and privacy concerns associated with their use.

VI.References

[1]Patel, N. A., & Patel, R. (2018, December). A survey on fake review detection using machine learning techniques. In 2018 4th International Conference on Computing Communication and Automation (ICCCA) (pp. 1-6). IEEE.

[2]Iswarya, M., Shankar, S. A., & Hameed, S. A. (2022, November). Fake Logo Detection. In 2022 1st International Conference on Computational Science and Technology (ICCST) (pp. 998-1001). IEEE.

[3]Montserrat, D. M., Lin, Q., Allebach, J., & Delp, E. (2018, April). Scalable logo detection and recognition with minimal labeling. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 152-157). IEEE.

[4]Kashid, A. M., Lalwani, A. K., Gaikwad, S. S., Patil, R. A., Sonkamble, R. G., & More, S. S. Fake Review Detection System Using Machine Learning.

[5]Sumathi, V. P., Pudhiyavan, S. M., Saran, M., & Kumar, V. N. (2021, October). Fake Review Detection Of E-Commerce Electronic Products Using Machine Learning Techniques. In 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) (pp. 1-5). IEEE.

[6]Paleček, K. (2019, July). Deep learning for logo detection. In 2019 42nd International Conference on Telecommunications and Signal Processing (TSP) (pp. 609-612). IEEE.

[7]Shet, A. U., Ibrahim, S. R., Joshi, M., Menezes, S. L., & HM, R. Implementation of Efficient Inception V2 model for Apparel Counterfeit Detection.

[8] Fang, Y., Wang, H., Zhao, L. et al. Dynamic knowledge graph based fake-review detection. Appl Intell 50, 4281–4295 (2020)

[9] YOLO: Algorithm for Object Detection Explained
https://www.v7labs.com/blog/yolo-object-detection

[10] Bidirectional LSTM
https://paperswithcode.com/method/bilstm