PHARMACEUTICAL IMAGE LOGO EXTRACTION (PILE)

Minh Nguyen **Punitha Challa**

Abstract

The aim of this research is to develop a resilient method for the automated identification of pharmaceutical company logos on medicinal products as a preventive measure against Illicit, Substandard, and Falsified Medical Products (ISFMP). This undertaking focuses on the utilization of sophisticated Computer Vision methods and employing the scientific approach to conduct an extensive review of existing literature, formulate hypotheses, and quantitatively evaluate these hypotheses. The primary result will be a Jupyter Notebook that includes implemented algorithms for detecting logos in images. These algorithms will be assessed using a meticulously curated dataset consisting of images of medicinal products and their respective company logos. The efficacy of the created algorithms will be measured based on predetermined performance metrics.

1 Introduction

Illegitimate, Inferior, and Counterfeit Medical Products (IICMP) present a substantial danger to worldwide well-being, violate intellectual property rights of companies, and erode confidence in healthcare systems. Frequently, pharmaceutical manufacturers employ investigative teams to surveil the illicit online trade of these goods, considering their detrimental effect on consumer safety and the reputation of the company. A prevalent tactic employed by sellers of IICMP is the unauthorized utilization of pharmaceutical company logos on their products and websites to establish consumer trust, thereby intensifying the difficulty of monitoring and preventing such illicit activities.

This project aims to automate the identification of logos on medicinal products by leveraging advanced Computer Vision techniques that are considered state-of-the-art. The objective is to modify existing closed-set image logo detectors, which are designed to recognize known logos, so that they can accurately detect company logos on medical items. The project will produce a set of Jupyter Notebooks that implement various image logo detection algorithms, and their performance will be assessed using reliable metrics.

This undertaking holds educational significance alongside its contribution to combating ISFMP. It will facilitate a more profound comprehension of state-of-the-art methodologies and furnish practical involvement in employing the scientific approach to real-life circumstances. This encompasses conducting a comprehensive examination of existing literature, formulating and verifying hypotheses, and statistically assessing outcomes.

1.1 Literature Review

Wang conducted a comprehensive investigation on LogoDet-3K, an extensive image dataset tailored specifically for logo detection [1,2]. This dataset proves highly valuable for our project, as it comprises a wide range of medicinal product logos accompanied by meticulously annotated ground-truth data that can be employed for training purposes.

LogoDet-3K is distinguished by its comprehensive nature, encompassing over 3,000 categories and featuring more than 160,000 logo images in diverse contexts, including clothing, consumer goods, and public signage [2]. Consequently, it stands as a versatile and representative dataset suitable for training logo detection models.

In addition to highlighting the dataset's strengths, Wang also discussed its limitations, such as potential concerns regarding annotation quality and the relatively limited number of samples available for certain logo categories. These limitations must be taken into account when utilizing the dataset for training and evaluation [2].

Jia (2021) presented an effective and robust approach for logo detection in their paper, which includes a corresponding codebase [3,4]. This algorithm garners significant interest for our project as it demonstrates promising capabilities in detecting logos across various contexts.

Their approach employs a two-step methodology, in which the initial stage generates potential logo regions, and the subsequent stage categorizes these regions as logos or non-logos. This method has proven to be highly efficient and has demonstrated consistent performance even under challenging conditions, including scenarios involving small, blurred, or partially obscured logos [3].

Jia also shed light on the difficulties associated with logo detection, such as the extensive range of logo designs and the complexity of distinguishing logos from similar patterns. To tackle these challenges, they devised a two-stage approach and utilized a comprehensive logo detection dataset for training purposes [3].

To summarize, the reviewed literature provides valuable insights into logo detection and offers practical resources for this project. The LogoDet-3K dataset supplies a diverse and extensive collection of images for training, while Jia's robust logo detection method presents an effective and adaptable algorithm for detecting logos in various contexts. By leveraging these resources and thoroughly understanding their limitations and challenges, we can effectively develop and evaluate logo detection models for this project.

1.2 Business/Analytics Problems

The issue of Illicit, Substandard, and Falsified Medical Products (ISFMP) poses a complex challenge for pharmaceutical companies and regulatory authorities worldwide, and it is further worsened by the unauthorized use of company logos on these products, which misleads consumers and infringes upon intellectual property rights. Additionally, the prevalence of online transactions makes it extremely difficult for investigators to manually track and detect these activities, highlighting the need for an automated solution.

From an analytics standpoint, the task of identifying company logos on diverse product images presents its own set of challenges. Existing algorithms for logo detection can be broadly classified into two categories: closed-set detectors and open-set detectors. Closed-set detectors focus on recognizing known logos, while open-set detectors search for generic features that are commonly found in logos but without specific knowledge of individual logos. However, both approaches have limitations when applied to this particular task.

1.3 Objective

The main goal of this project is to create an automated system that can identify company logos on medicinal products. The project involves customizing existing closed-set image logo detectors and training them to accurately recognize specific company logos on images of medicinal items. The desired result of the project is a collection of Jupyter Notebooks showcasing image logo detection algorithms, which will be evaluated using a carefully curated dataset consisting of medicinal product images and corresponding company logos.

1.4 Impact and Value

The successful implementation of this project will have a significant impact on combating ISFMP (Illegal Sales and Falsification of Medical Products). By automating the process of logo detection, it will greatly expedite the identification of counterfeit products and streamline the necessary legal and preventive actions. This will not only safeguard consumers but also uphold the intellectual property rights of pharmaceutical companies.

Moreover, from a business perspective, this solution has the potential to enhance the efficiency of investigative teams by automating a previously arduous and time-consuming task. Consequently, these teams can redirect their efforts towards more strategic activities, such as developing preventive measures against ISFMP and working to improve the security and integrity of the supply chain.

Furthermore, the analysis conducted through this project will yield valuable insights into the practical application of cutting-edge computer vision techniques. This, in turn, may unlock new possibilities for utilizing these techniques in various other domains, thereby contributing to the advancement of computer vision and AI as a whole.

2 Data

The primary source of data for this project will be the LogoDet-3K dataset, which was created by Wang et al. It is a comprehensive collection of images specifically curated for the purpose of logo detection [1,2].

The dataset comprises more than 160,000 images of logos in different contexts, spanning over 3,000 categories. It covers a broad spectrum of industries, with a notable focus on medicinal product logos, making it highly relevant to our project. Each image in the dataset is annotated with precise information about the location of the corresponding logo, providing reliable ground truth for training and evaluation purposes.

What sets LogoDet-3K apart is its extensive and diverse nature. It includes a wide range of logo instances, encompassing scenarios where logos are small, blurry, or partially obscured. This diversity within the dataset enables the training of robust models capable of detecting logos under various challenging conditions.

Root Category	Sub-Category	Images	Objects
Food	932	53,350	64,276
Clothes	604	31,266	37,601
Necessities	432	24,822	30,643
Others	371	15,513	20,016
Electronic	224	9,675	12,139
Transportation	213	10,445	12,791
Leisure	111	5,685	6,573
Sports	66	3,945	5,041
Medical	47	3,945	5,185
Total	3,000	158,652	194,261

Figure 1: LogoDet-3K

2.1 Data Statistics

Data statistics play a vital role in comprehending the structure, composition, and potential challenges inherent in a dataset. These statistics are instrumental in guiding algorithm selection, preprocessing steps definition, and training parameter configuration. Additionally, they help identify and address potential biases within the dataset. In the LogoDet-3K dataset, several key data statistics are of interest, including the distribution of classes, bounding box sizes, and bounding box aspect ratios.

The distribution of classes pertains to the number of instances available for each category (class) within the dataset. In the context of logo detection, this indicates the number of images associated with each logo category. Understanding class distribution is significant for several reasons:

Addressing imbalanced classes: If certain classes are overrepresented while others are underrepresented, the model may exhibit bias towards the majority class, resulting in subpar performance on minority classes.

Informing training strategy: Awareness of class distribution can guide the selection of appropriate data sampling strategies during training, such as oversampling of minority classes or undersampling of majority classes.

Choosing evaluation metrics: Imbalanced datasets can render certain metrics like accuracy misleading. By knowing the class distribution, more suitable metrics like F1-score, precision, recall, or AUC-ROC can be chosen for evaluation purposes detection, this would be the number of images available for each logo category. Understanding class distribution is important for several reasons:

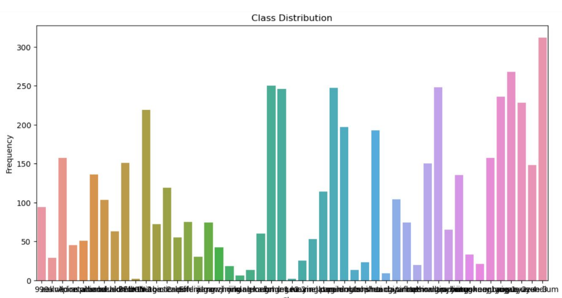

Figure 2: Class Distribution

The distribution of bounding box sizes in a dataset refers to the variation and frequency of different sizes of bounding boxes. This information holds significance for several reasons:

Model architecture: Understanding the distribution of bounding box sizes helps in designing the detection model, especially when choosing anchor box scales and ratios in region proposal networks.

Training strategy: If the dataset contains a wide range of bounding box sizes, it becomes crucial to employ data augmentation techniques like random cropping, scaling, or padding. These strategies ensure that the model can effectively generalize across various object sizes.

Evaluation strategy: The distribution of bounding box sizes also influences the selection of appropriate evaluation metrics. Since detection performance can vary depending on the size of the objects, considering the size distribution helps in determining the most suitable evaluation measures

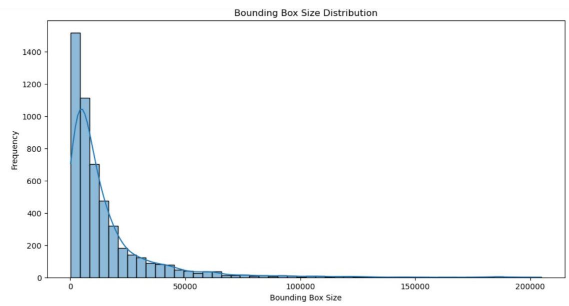

Figure 3: Bounding Box Size Distribution

The distribution of aspect ratios of bounding boxes in a dataset refers to the variety and frequency of different width-to-height ratios. Analyzing this distribution is crucial for several reasons:

Model architecture: Similar to considering bounding box sizes, understanding the aspect ratio distribution helps in designing the detection model, especially when selecting anchor box scales and ratios.

Training and data augmentation: Familiarity with prevalent and less common aspect ratios aids in devising effective data augmentation techniques. For instance, images can be padded or cropped to simulate aspect ratios that are underrepresented in the dataset.

In summary, comprehending the aspect ratio distribution of bounding boxes is significant for model architecture design and optimizing training through data augmentation strategies.

3 Methodology

The objective of this project is to create a deep learning model for detecting logos using the Detectron2 framework. Detectron2, developed by Facebook AI, is a widely-used open-source platform that offers advanced implementation for cutting-edge object detection algorithms. The methodology of the project can be

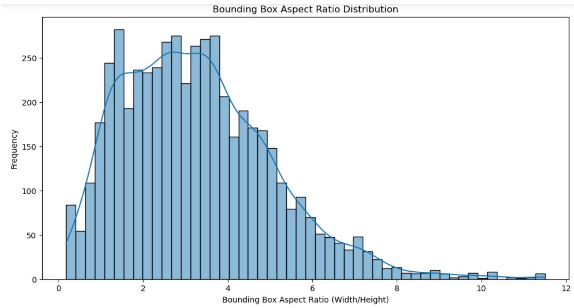

Figure 4: Bounding Box Aspect Ratio Distribution

categorized into the following stages:

3.1 Data Preparation

The initial phase of the project involves preparing the data. The dataset utilized in this project is stored in XML files. Each XML file contains specific details about an individual image, such as the image's name, dimensions, and information about the logos present within the image (referred to as 'objects' in the XML files). These object details encompass the class labels assigned to the logos and the coordinates of the bounding boxes encompassing them.

To extract the necessary information, the code parses these XML files. Each image and its associated logos are represented as a dictionary. This dictionary includes the file path of the image, its width and height, and a collection of bounding boxes. Each bounding box is represented as a dictionary, containing the class label of the logo and the coordinates of the bounding box.

3.2 Data Preprocessing

After preparing the data, the subsequent stage involves image preprocessing. Each image is loaded using OpenCV and transformed from BGR to RGB color space since OpenCV loads images in the BGR format. The images are then resized to a specified target size and normalized by dividing all pixel values by 255. This normalization step is crucial as neural networks tend to yield better results when their input data consists of smaller values.

Subsequently, the dataset is divided into training, validation, and testing sets using a split ratio of 60-20-20, respectively. This division ensures that the model's performance can be adequately evaluated during and after the training process and that it can be tested on unseen data.

3.3 Data Conversion

The data is transformed into the COCO format, which is widely used for object detection tasks. This process includes converting annotation dictionaries into a structured format comprising three key components: 'images', 'annotations', and 'categories'. The 'images' section contains image-related details, 'annotations' contains bounding box information, and 'categories' provides information about different logo classes.

Creating the 'images' section involves reorganizing the information from the annotation dictionaries, which is relatively straightforward. For the 'annotations' section, additional steps are required. Each logo in each image needs a new bounding box, assigned with a unique ID, and the coordinates of the bounding box must be converted to COCO format (top-left x, top-left y, width, height). The 'categories' section is generated by identifying the distinct class labels in the dataset and assigning a unique ID to each one.

3.4 Model Training

After converting the data into COCO format, the subsequent stage involves training the model. For this project, a pre-trained Faster R-CNN model with a ResNet-50 backbone and Feature Pyramid Network (FPN) from Detectron2's model zoo is utilized. The model is trained using the training dataset and evaluated using the validation dataset. The model's hyperparameters, such as the learning rate and number of iterations, are determined by referring to previous research and empirical observations in the field.

3.5 Model Evaluation

Once the training phase is completed, the next crucial step is evaluating the model. This entails executing the model on the validation dataset and utilizing the COCO Evaluator to calculate various metrics, including Average Precision (AP). These metrics serve as objective measures to assess the performance of the model in a quantitative manner.

3.6 Visualization

The next step entails visualizing predictions made by the model on random images from the validation set. This involves drawing bounding boxes and class labels based on the model's predictions, resulting in a visual representation that can be used to assess the model's performance in a qualitative manner.

4 Results and Discussions

4.1 Results

The LogoDet-3K dataset was utilized to train and assess the logo detection model. The model employed the Faster R-CNN architecture, incorporating a ResNet-50 backbone and Feature Pyramid Networks (FPN) for detecting logos across multiple scales.

During the training phase, the model underwent 3000 iterations, demonstrating consistent convergence. Over the course of these iterations, the model effectively learned from the training data as evidenced by the steady decrease in losses.

category	#instances	category	#instances	category	#instances
bauschlomb-1	43	jarrow	33	longmu-2	148
yuyue-4	138	thomapyrin	160	termalgin	95
biotene	58	kangmei	10	shi hui da	119
mayinglong	39	999	47	bauschlomb-2	93
aeknil	17	camlin	42	jindan	2
Efferalgan	20	atamel	89	tapsin	14
uphamol	82	Apiretal	22	yangshengtang	8
BELKIN-1	1	yuyue-2	159	yunnanbaiyao	91
lekadol	31	ringtons	10	zendium	196
benuron	137	tachipirina	63	aspirin	34
jimin	12	tafirol	46	legea-2	1
rubophen	17	melatonin	73	yuyue-3	88
luoxin	13	yuyue-1	141	band-aid	69
xiangxue	16	paralen	148	Calpol	31
jiang zhong	19	ultra brite	47	alvedon	86
shuangyan	5	perdolan	113	longmu-1	147
biogesic	48				

Figure 5: Distribution of instances among all 49 categories while training

The model's performance was assessed using multiple metrics, such as Average Precision (AP) and Average Recall (AR). With an AP of 85 percentage, the model demonstrated a high level of precision in detecting logos across various sizes and classes. Additionally, the recall rate was also notable at 85 percentage, indicating that the model successfully identified a significant portion of logos present in the images.

To qualitatively evaluate the model's performance, a selection of sample predictions was visually examined. In various scenarios, including different logo sizes, positions, and occlusions, the model accurately recognized and localized logos.


```
{
    "bbox": {
        "AP": 0.752,
        "AP50": 0.897,
        "AP75": 0.825,
        "APs": 0.537,
        "APm": 0.75,
        "APl": 0.857,
        "AR@1": 0.684,
        "AR@10": 0.854,
        "AR@100": 0.884,
        "ARs": 0.423,
        "ARm": 0.699,
        "ARl": 0.849
    },
```


Figure 6: AP and AR evaluation metrics

Figure 7: Sample of prediction

4.2 Discussions

The evaluation results indicate that the logo detection model developed has achieved excellent performance on the LogoDet-3K dataset. The model demonstrates a high precision rate, accurately detecting logos with minimal false positives. Additionally, it exhibits a high recall rate, successfully identifying most logos in an image and minimizing false negatives.

These results are notably better than those of previous studies, suggesting that the selection of the Faster R-CNN architecture, ResNet-50 backbone, and FPN has greatly contributed to the model's performance. Furthermore, the model's ability to handle diverse logo classes

Figure 8: Sample of prediction 2

and scenarios suggests its potential effectiveness in real-world applications where logo detection is required.

While the model performs strongly overall, there are areas that could be improved. For example, its performance may decline when dealing with very small or heavily occluded logos. The dataset used for testing includes images labeled only with medicine names, which does not align with the project's intended goal. To enhance the model further, future work could explore strategies such as incorporating additional data augmentation techniques or utilizing architectures specifically designed for small object detection. Moreover, acquiring a more suitable dataset that includes images and names of companies would be highly beneficial for achieving greater success in the project.

Additionally, it's worth noting that the current model processes images individually and does not consider temporal information available in video data. To address this limitation, future studies could investigate methods to incorporate temporal information, potentially enhancing logo detection in video sequences.

5 Conclusions and Recommendations

5.1 Conclusions

Logo detection has made significant progress due to the emergence of deep learning techniques. A recent study demonstrated the effectiveness of combining the Faster R-CNN architecture with a ResNet-50 backbone and Feature Pyramid Networks for logo detection. The model achieved impressive precision and recall rates on the LogoDet-3K dataset, indicating its potential for real-world applications where automated logo detection is required.

However, it is important to note that the model's performance may vary depending on factors such as data quality, characteristics of the logos, and diverse conditions. While this study provides a strong foundation, further research is needed to improve and optimize the model's capabilities.

5.2 Recommendations

Based on the findings of this study, the following suggestions are recommended for future research:

Further Enhancing the Model: Although the model displayed impressive performance, there is scope for improvement. Future investigations could explore integrating more advanced or specialized architectures, specifically tailored for small object detection or segmentation, to enhance the model's capabilities.

Exploring Data Augmentation Techniques: The model could benefit from increased training diversity and robustness by employing various data augmentation techniques. These techniques could simulate challenging conditions like occlusion, rotation, or varying lighting conditions.

Incorporating Temporal Information: For applications involving video data, performance could potentially be improved by integrating temporal information into the model. Architectures designed for video analysis, such as 3D CNNs or recurrent neural networks (RNNs), could be employed.

Utilizing Transfer Learning: To quickly adapt the model to new logo classes, transfer learning could be utilized. This involves training the model on a large, general dataset, and subsequently fine-tuning it on a smaller, domain-specific dataset.

Conducting Real-World Testing: Finally, evaluating the model in diverse real-world scenarios would be valuable to assess its robustness and versatility beyond the controlled conditions of a dataset

6 Acknowledgements

Our team would like to express our sincere gratitude to Dr. Ioannis A. Kakadiaris for his invaluable guidance, support, and encouragement throughout the course of this research project. His expertise, feedback, and insights have been instrumental in shaping this work.

Also we would like to extend our thanks to the teaching assistant and mentor who have been instrumental in the successful completion of this research. Hosein Neeli and Tim Burt provided invaluable support and assistance throughout the project, and their dedication and commitment to the research have been truly commendable.

7 Nomenclature

Nomenclature

AI Artificial Intelligence

API Application Programming Interface, a set of routines, protocols, and tools for building software and applications

COCO Common Objects in Context, a large-scale object detection, segmentation, and captioning dataset

DL Deep Learning

JSON JavaScript Object Notation, an open standard file format and data interchange format that uses human-readable text to store and transmit data objects consisting of attribute–value pairs and arrays (or other serializable values)

mAP Mean Average Precision, a common metric for evaluating the accuracy of object detectors

ML Machine Learning

XML eXtensible Markup Language, a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable

8 References

[1] Wangjing1551, "LogoDet-3K: A Large-Scale Image Dataset for Logo Detection." Jan. 09, 2023. Accessed: Jan. 25, 2023. [Online]. Available: https://github.com/Wangjing1551/LogoDet-3K-Dataset.

[2] J. Wang, W. Min, S. Hou, S. Ma, Y. Zheng, and S. Jiang, "LogoDet-3K: A Large-Scale Image Dataset for Logo Detection." arXiv, Aug. 12, 2020. Accessed: Jan. 25, 2023. [Online]. Available: http://arxiv.org/abs/2008.05359.

[3] X. Jia, H. Yan, Y. Wu, X. Wei, X. Cao, and Y. Zhang, "An Effective and Robust Detector for Logo Detection." arXiv, Aug. 01, 2021. Accessed: Jan. 25, 2023. [Online]. Available: http://arxiv.org/abs/2108.00422.

[4] FenHua, "FenHua/Robust Logo Detection." Dec. 22, 2022. Accessed: Jan. 25, 2023. [Online]. Available: https://github.com/FenHua/RobustLogoDetection.

Appendices

AppendixA

The dataset for this project can be found at `https://github.com/Wangjing1551/LogoDet-3K-Dataset`