EMPLOYING MACHINE LEARNING AND AN OCR VALIDATION TECHNIQUE TO IDENTIFY PRODUCT CATEGORY BASED ON VISIBLE PACKAGING FEATURES

Takorn Prexawanprasut
SCHOOL OF SCIENCE AND TECHNOLOGY
SUKHOTHAI THAMMATHIRAT OPEN UNIVERSITY, THAILAND
takorn.pre@stou.ac.th

Lalita Santiworarak*
SCHOOL OF SCIENCE AND TECHNOLOGY
SUKHOTHAI THAMMATHIRAT OPEN UNIVERSITY, THAILAND
lalita.san@stou.ac.th

Piyaporn Nurarak
SCHOOL OF SCIENCE AND TECHNOLOGY
SUKHOTHAI THAMMATHIRAT OPEN UNIVERSITY, THAILAND
piyaporn.nur@stou.ac.th

Poom Juasiripukdee
SCHOOL OF SCIENCE AND TECHNOLOGY
SUKHOTHAI THAMMATHIRAT OPEN UNIVERSITY, THAILAND
poom.jua@stou.ac.th

ABSTRACT

Customs clearance is a challenging and time-consuming process that must be completed in the sphere of international trade. As a result, the cargo is frequently delayed at the port. If the personnel know the initial number of items, they may be able to continue with other procedures even when they are not physically present at the location. Image processing is helpful in this area since it allows for the prediction of the type of goods based on the appearance of the package. This allows for the determination of the quantity of each type of product prior to the arrival of the employees at the site. Three distinct import-export companies contributed 5,675 photos, and a machine learning approach was used to create a model that can predict the types of things that fall into one of five categories. Also, the researchers made an OCR-based classification algorithm with the goal of making machine learning work better for certain types of things that have trouble learning.

CCS CONCEPTS

• **Computing methodologies**; • **Machine learning**; • **Machine learning approaches**;

KEYWORDS

Image processing, Prediction, Machine learning, OCR-based classification

ACM Reference Format:
Takorn Prexawanprasut, Lalita Santiworarak*, Piyaporn Nurarak, and Poom Juasiripukdee. 2023. EMPLOYING MACHINE LEARNING AND AN OCR VALIDATION TECHNIQUE TO IDENTIFY PRODUCT CATEGORY BASED ON VISIBLE PACKAGING FEATURES. In *2023 The 6th International Conference on Machine Vision and Applications (ICMVA) (ICMVA 2023), March 10–12, 2023, Singapore, Singapore.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3589572.3589589

ICMVA 2023, March 10–12, 2023, Singapore, Singapore

ACM ISBN 978-1-4503-9953-1/23/03…$15.00
https://doi.org/10.1145/3589572.3589589

1 INTRODUCTION

It was uncovered that the founders of three import/export start-ups established their own businesses after leaving their former company. Two key factors led to their departure from their previous workplaces. First, they believed they might obtain a larger market share in the import/export industry due to its constant growth. Second, they considered that their former employers had disregarded a number of smaller client groups. Old corporations were compelled to serve their most distinguished customers first because they were relatively huge. As a result, some consumers with lower ranks were ignored. Consequently, a chunk of the lower-tier clientele had moved to a competitor that offered superior service. During the first stages of establishing their businesses, the number of packages sent was extremely low. As the business grew, however, the number of packages increased. The business owners need computer solutions, such as a workflow management system, to manage their company's delivery. One of the most difficult components of cargo handling at the port is determining whether specific items are included in a shipment and where they are situated within a container. However, the price of this software system is quite high and may not be appropriate for startups, mandating the use of alternative problem-solving techniques.

Object recognition refers to a series of interrelated computer vision problems involving the identification of objects in digital images. The mechanism begins with the definition of the ontology, or the category of detectable objects. Then, classification and tagging identify what is in the image and the associated confidence level. Tagging can recognize many object classes inside an image, but classification can only identify a single class. In other words, the computer will simply remember the existence of objects when categorizing, disregarding all previous classifications. In contrary, while tagging an image, it will attempt to acquire all applicable classes. The algorithm will then identify the presence of objects in

an image, show their location with a bounding box, and calculate the classes of the identified objects.

This work aims to provide a technique for identifying products based on their package images. There are five preset categories, which include medical products, construction equipment, documents, electrical equipment, and additional products. Five thousand, six hundred and seventy-five parcel photos from three startup companies were chosen for the initial feature extraction and machine learning examination. Each image is labeled with a class by humans to indicate the composition of the objects within it. CNN, the fundamental technique utilized in the process of picture classification, was employed to tackle this issue. It is a kind of Neural Networks that is primarily used for image and speech recognition applications. Its built-in convolutional layer minimizes the high dimensionality of images without sacrificing their data. The experiment also employs the 10-fold cross-validation procedure, which aids in confirming the experimental outcomes.

The remainder of the paper will be presented below. The section 2 literature review. In Section 3, the technique is proposed. The fourth section contains the experimental results. The section 5 conclusion and discussion are followed by the section 6 acknowledgements.

2 LITERATURE REVIEW

It is difficult to construct a traditional convolutional neural network with a fully connected output layer when the length of the output layer is not predetermined due to the presence of a large number of bounding boxes representing diverse things of interest inside the image. A straightforward solution to this issue would be to extract many zones of interest from the image and classify the presence of the object within each zone using a CNN. The problem with this strategy is that the objects of interest inside the image may have variable spatial positions and aspect ratios. To rapidly recognize these instances, algorithms such as R-CNN, YOLO, etc. have been developed.

The R-CNN [1] is an object localization, detection, and segmentation application based on convolutional neural networks. Three modules make up their proposed R-CNN model: Generate and extract category-independent region proposals, such as potential bounding boxes. Then, extract features from each possible region, for instance using a convolutional neural network with deep layers. Classify characteristics as belonging to one of the known classes, for instance using a linear SVM classifier model. Nonetheless, R-CNN requires the development and operation of three distinct models. In addition, training a deep CNN on numerous area suggestions is a lengthy process. Consequently, the model is inefficient when applied to a very big dataset.

Fast R-CNN [1] is a single model that immediately trains and outputs regions and classifications in order to overcome the limitations of R-CNN and provide a quicker object recognition system. An picture and a set of suggested region borders are used in the model's creation as input to a deep convolutional neural network. Utilizing a CNN with prior training, such as VGG-16, is used for feature extraction. A special layer known as a Region of Interest Pooling Layer, or RoI Pooling, is the final layer of a deep CNN and it extracts features unique to a certain candidate input region. However, the model still needs a list of candidate areas to be suggested along with each input image, despite being substantially faster to train and provide predictions. Region suggestions therefore constitute performance bottlenecks in the Fast R-CNN algorithm.

Despite being a single, unified model, the Faster R-CNN [5] consists of two parts. Initially, a Convolutional neural network for suggesting areas and the form of item to consider in each region. Then, a convolutional neural network is utilized to extract features from the suggested areas and to output the bounding box and class labels. Both modules utilize the same deep CNN output. The region suggestion network functions as an attention mechanism for the Fast R-CNN network, instructing it where to look or pay attention. Both sub-networks are trained at the same time, using an alternate training approach. This permits the settings of the deep CNN feature detector to be simultaneously adjusted or fine-tuned for both tasks. This Faster R-CNN architecture continues to deliver near-state-of-the-art performance on object identification tasks. A further extension includes picture segmentation capability, as detailed in Mask R-CNN [4, 6].

The YOLO model [2, 3] consists of a single neural network trained from beginning to end that accepts an image as input and immediately predicts bounding boxes and class labels for each bounding box. Each cell is responsible for predicting a bounding box if the center of a bounding box falls within its cell. Each grid cell predicts a bounding box consisting of the x and y coordinates, width and height, and confidence. Additionally, a class prediction is based on each cell. The R-CNN models may be more accurate in general, but the YOLO family of models is quick, considerably faster than R-CNN, and achieves real-time object recognition.

Similar to Faster R-CNN, YOLOv2 [5] utilizes anchor boxes, which are predefined bounding boxes with useful shapes and sizes that are customized during training. The selection of the image's bounding boxes is pre-processed using k-means analysis on the training dataset. Importantly, the anticipated representation of the bounding boxes is modified to reduce the impact of modest changes on the predictions, resulting in a more stable model. Rather of explicitly forecasting location and size, offsets for moving and reshaping pre-defined anchor boxes relative to a grid cell are anticipated and attenuated by a logistic function. Further improvements to the YOLOv2 were proposed by YOLOv3 [7]. The enhancements included a deeper network of feature detectors and modest representational adjustments.

There are numerous approaches to combining image detection and classification techniques using traditional neural networks. To distinguish between images of normal and abnormal blood cells, Thanh et al. [8] provide a solution based on a convolutional neural network. To expand the quantity of data currently available, the suggested solution uses data augmentation techniques such as histogram equalization, picture translation, reflection, and rotation. For evaluation objectives, the researchers compared their strategy with traditional statistical properties. When used on a dataset that has been significantly improved, the CNN-generated features perform better than the conventional statistical features.

Modern computer vision should enable effective online processing of massive volumes of high-dimensional data without the need for any specialized pre-processing. Without the need for sophisticated domain-specific feature extraction methods, I. Mrazova and M. Kukackal [9] offer a model known as "Growing Hierarchical

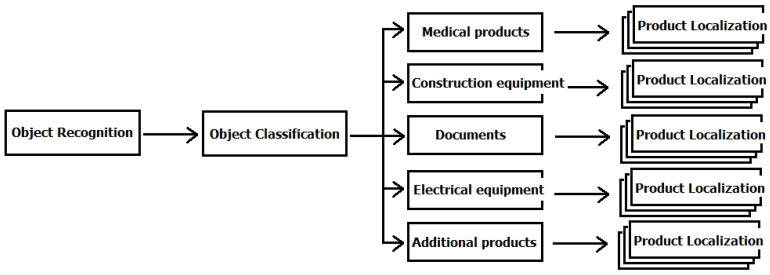

Figure 1: The process of classifying goods based on images gathered from unloading port containers.

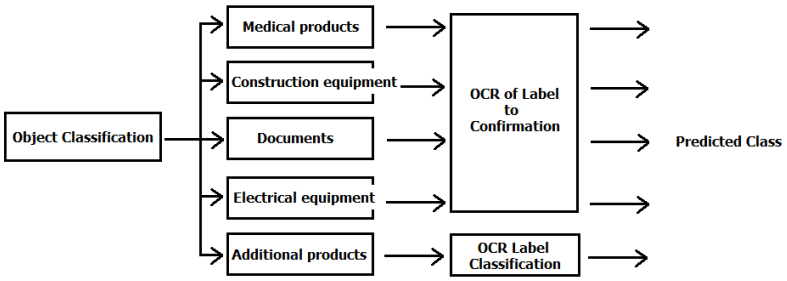

Figure 2: Classes validation process by using OCR label reading.

Neural Networks (GHNN)" that can be used for picture categorization. The recently developed GHNN is a potent multi-layered self-organizing architecture that can automatically create a suitable feature detector from the processed data. Similar to previous techniques, a proper topology that avoids duplicate processing has been shown to produce better performance on CBCL face data and MNIST handwritten digit data.

A important characteristic that can be used to improve the categorization capabilities is the gradient histogram. Each pixel in an image has a gradient value assigned to it. Then, for effective spam classification, these collected features are normalized. For the feed forward back propagation neural network (BPNN) model, M. Soranamageswari and C. Meena [10] employed normalized features as input. Using MATLAB, a feature point extraction approach based on a gradient histogram is developed for an image spam classification system. Gradient histogram is used as an image feature in the experimental system to report on a new picture spam categorization algorithm.

3 METHODOLOGY

Typically, specific logos are imprinted on the shipments that the three companies are responsible for. Experts determine the type of product based on the size, shape, and seals of the packaging. Therefore, the standard procedure for classifying objects cannot be used to enable the computer to determine the classification of the product category. A novel classification method is intended to improve the learning efficiency of this project. It begins with the Object Recognition procedure, which identifies the target area likely to contain an object. Each object is then classified into one of five classes using machine learning techniques such as CNN, R-CNN, Fast R- CNN, Faster R-CNN, and the YOLO family technique, as described before in the study. The objects are then re-counted in the image to establish the exact quantity of each parcel. All of these procedures will aid officials in estimating the quantity of items that will require customs clearance at the port, even though they have not yet arrived at the worksite.

The method of categorizing products based on photographs collected from unloading containers at the port is depicted in figure1. The procedure commences with image recognition to identify the objects in the image. The system then uses the identified objects to classify the objects into the following five categories: Medical product, Construction equipment, Documents, Electrical. equipment, and Additional product, which are the miscellaneous things added to the four previous ordinal types. The final step in the process is product localization, which involves redrawing the original image's frame to locate various types of objects. In addition, as previously mentioned, experts frequently classify products based on the characteristics of the packaging and the label imprinted on it. Consequently, the researcher devised a validation procedure employing Optical Character Recognition (OCR), a technique for reading characters from photographs.

This method certifies the product type once machine learning has identified it. It will only be done for products in the first four classes, and only for those whose packaging includes the letter type logo. For products that are classified as type 5, the OCR Classification process will be used to classify Clustering (The unsupervised learning model). Figure 2 demonstrates the validation procedure by using OCR label reading.

Table 1: Styles available in the Word template

Object Detection Technique	Overall accuracy	
	without OCR validation	with OCRvalidation
R-CNN	79.23	79.87
Fast R-CNN	82.35	83.45
Faster R-CNN	83.24	82.14
YOLOv2	**85.12**	**86.78**
YOLOv3	85.03	85.12
Average	82.994	83.472

4 EXPERIMENTAL RESULTS

It was discovered, while attempting to write a program in accordance with the concept using Python programming with the TensorFlow and Keras library, that the classification of objects inside container has an accuracy of 82.994 percent on average, with the YOLOv2 technique having the highest accuracy at 85.12 percent. With the additional OCR validation process, the experiment showed a slight improvement, with accuracy going up to an average of 83.472 percent and to 86.78 percent with the YOLOv2 method. The total accuracy of the item categorization can be seen in Table1, which was produced using R-CNN, Fast R-CNN, Faster R-CNN, YOLOv2 and YOLOv3 models.

Because the total accuracy value does not represent the performance of the model when it is applied to each class, it is misleading. Therefore, in order to demonstrate the correctness and completeness of the categorization, the researchers carried out performance measurements making use of Precision and Recall values. A higher Precision value suggests a more accurate filtering of sample units that belong to that class, while a higher Recall value indicates a more comprehensive collection of sample units that belong to that class. Both values are measured in percentages.

Table 2 shows the Accuracy (Ac), Precision (Pr), and Recall (Re) values for the Medical products (Medical) and Construction equipment (Construction) classes, whereas Table 3 shows the Accuracy, Precision, and Recall values for the Documents (Documents) and Electrical equipment (Electrical) classes. It was discovered that the Medical class was the only class that was experiencing the problem of low Recall, which indicates that the model was not able to collect sample units as well as it should have for that class. The average Recall of the Medical class was just 55.30 percent, as demonstrated. As can be seen in Table 4, the introduction of the OCR validation procedure appears to have resolved the issues that had been plaguing the Medical class, as the Recall value for that class increased from 55.288 to 67.324 percent on average. Note that "Pr woOCR" refers to Precision without performed OCR Validation, "Pr wOCR" to Precision with Validation, "Re woOCR" to Recall without OCR validation, and "Re wOCR" to Recall with OCR validation. Only with the Medical class did we observe an increase in Recall value when verifying answers using the OCR approach. This strategy does not have any effect on the Recall when applied to other classes. As seen in Figure 3, Recall increased during OCR validation. The last bar set, labeled Re wOCR, has a larger value than the second bar set, Re woOCR, but this tendency does not occur in Figure 4, which depicts the Recall values for other classes.

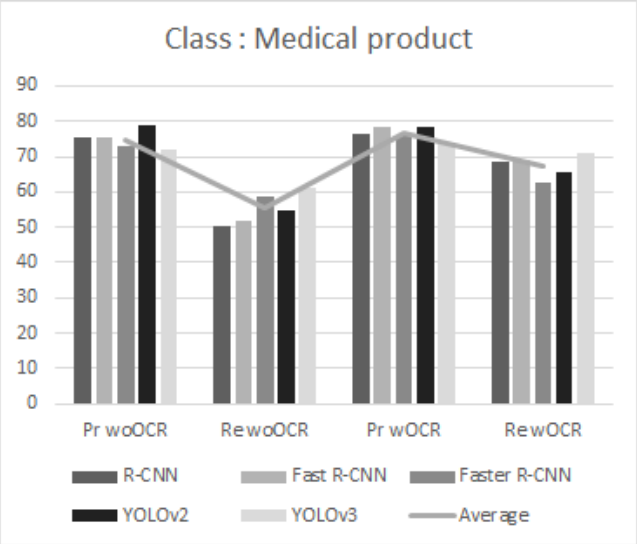

Figure 3: Recall increased during OCR validation in Medical class.

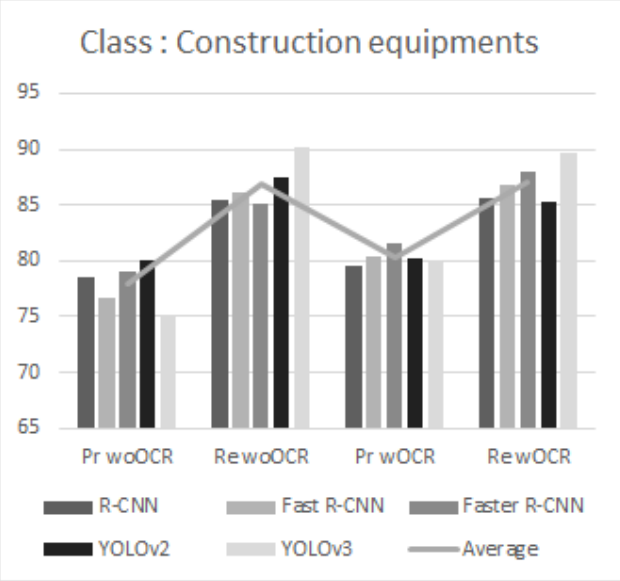

Figure 4: The other class does not show any improvement in recall.

5 CONCLUSION AND DISCUSSION

Although there is not a significant difference in the overall performance of classification when using various machine learning techniques, when each class is considered separately, it was discovered that the Medical class had the lowest classification efficiency.

Table 2: Accuracy, Precision and Recall of Medical products and Construction equipment class

	Class					
	Medical			Construction		
	Ac	Pr	Re	Ac	Pr	Re
R-CNN	68.75	75.23	50.12	85.59	78.56	85.45
Fast R-CNN	70.65	75.64	52.01	87.23	76.74	86.12
Faster R-CNN	71.28	72.78	58.74	87.53	78.96	85.19
YOLOv2	72.55	78.77	54.55	89.54	79.98	87.56
YOLOv3	73.21	72.04	61.02	89.08	75.12	90.12
Average	**71.3**	**74.9**	**55.3**	**87.8**	**77.9**	**86.9**

Table 3: Accuracy, Precision and Recall of Documents and Electrical equipment class

	Class					
	Documents			Electric		
	Ac	Pr	Re	Ac	Pr	Re
R-CNN	83.25	79.92	84.56	80.45	74.56	75.32
Fast R-CNN	86.19	78.14	88.64	85.24	75.23	84.57
Faster R-CNN	86.29	79.72	88.05	88.38	78.47	88.89
YOLOv2	88.73	75.25	89.45	90.45	74.26	85.64
YOLOv3	89.54	78.74	90.92	88.24	78.95	82.14
Average	**86.8**	**78.4**	**88.3**	**86.6**	**76.3**	**83.3**

Table 4: Precision and Recall of Medical class with and without OCR validation

	Class : Medical products			
	Pr woOCR	Pr wOCR	Re woOCR	Re wOCR
R-CNN	75.23	76.59	50.12	68.45
Fast R-CNN	75.64	78.23	52.01	69.21
Faster R-CNN	72.78	75.82	58.74	62.54
YOLOv2	78.77	78.29	54.55	65.55
YOLOv3	72.04	73.67	61.02	70.87
Average	**74.892**	**76.52**	**55.288**	**67.324**

It's possible that this is because the shape and texture of the packaging for this category is unclear. As a direct consequence of this, machine learning algorithms are unable to learn important features that are used to classify them. Moreover, when OCR validation was applied to such a class, it was discovered that such a technique could boost the classification effectiveness by improving the Recall value of this class. Without changing the Precision value, the OCR validation process collects sample units that diverged from classification using the normal classifier. This experiment's findings are advantageous since, under normal conditions, boosting Recall tends to decrease Precision, as the two tend to be inversely proportional.

However, these strategies cannot optimize other classes whose performance with machine learning algorithms is already excellent. In other words, reading OCR from a product label is unnecessary for the Construction, Documents, and Electric classes. This result may be caused by the fact that the packing of these three classes is sufficiently distinct that the classifier is able to differentiate between them. Therefore, using the OCR validation method does not result in a more comprehensive collection of sample units. On the other hand, it can cause the precision of the initial data to suffer, which would result in the data being less accurate overall.

Data pre-processing techniques such histogram equalization, picture translation, image reflection, image rotation, and grayscale implementation are being carefully studied in to improve performance in future works. Together with the attributes of the label on the packaging, a wide range of optical character recognition techniques should be researched and reported on. To compare the categorization effectiveness of various product classes, the more recent iterations of the YOLO detector should be employed.

ACKNOWLEDGMENTS

The researcher would like to thank the senior executives of the three import-export service companies. In particular, Khun Nutthicha Plonjan provided the information and images used in this research. The researcher is aware that the information collected for this study is confidential and will be used only for research reasons.

REFERENCES

[1] Girshick, R., Donahue, J., Darrell, T., and Malik, J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, Columbus, OH, USA, 580-587. https://doi.org/ 10.1109/CVPR.2014.3

[2] Ren, S., He, K., Girshick, R., and Sun, J.. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks, Vol. 28. Advances in neural information processing systems, ISBN: 9781510825024

[3] He, K., Zhang, X., Ren, S., and Sun, J. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. In Proceedings of the IEEE transactions on pattern analysis and machine intelligence. IEEE, 1904-1916. https://doi.org/10.1109/TPAMI.2015.2427753

[4] He, K., Gkioxari, G., Dollár, P., and Girshick, R. 2017. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision. IEEE, Venice, Italy, 2961-2969. https://doi.org/10.1109/ICCV.2017.322

[5] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. 2016. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, Las Vegas, NV, USA, 779-788. https://doi.org/10.1109/CVPR.2016.91

[6] Redmon, J., and Farhadi, A. 2017. YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, Honolulu, HI, USA, 7263-7271. https://doi.org/10.1109/CVPR.2017.690

[7] Redmon, J., and Farhadi, A. 2018. Yolov3: An incremental improvement. arXiv preprint arXiv. arXiv, 1804.02767.

[8] T. T. P. Thanh, Caleb Vununu, Sukhrob Atoev, Suk-Hwan Lee, and Ki-Ryong Kwon: Leukemia Blood Cell Image Classification Using Convolutional Neural Network. Vol.10. International Journal of Computer Theory and Engineering, https://doi.org/10.77.63/IJCTE

[9] Iveta Mrazova and Marek Kukacka: Image Classification with Growing Neural Networks. Vol.5. International Journal of Computer Theory and Engineering, https://doi.org/10.77.63/IJCTE

[10] M.Soranamageswari, Dr.C.Meena: A Novel Approach towards Image Spam Classification. Vol.3. International Journal of Computer Theory and Engineering, https://doi.org/10.77.63/IJCTE