Autonomous Pallet Counting for Mixed Storage Configurations in Low-Light Warehouses

Yonglan Qi, Chenyang Wang, Chenru Wu, Xinchang Tong
Khoury College of Computer Science
Northeastern University
Vancouver, BC, Canada
{*qi.yon, wang.chenyan, wu.chenru, tong.xin*}*@northeastern.edu*

Michal Aibin
Department of Computing
British Columbia Institute of Technology
Vancouver, Canada
maibin@bcit.ca

***Abstract*—This paper introduces a novel autonomous cycle counting method for warehouse inventory management, specifically designed for double-deep and bulk storage configurations. The proposed methodology processes 2D RGB images captured by autonomous robots navigating warehouse aisles to accurately count pallets within bin locations. Our two-stage approach first performs pallet detection using Faster R-CNN and YOLO models enhanced for low-light conditions, then applies configuration-specific counting algorithms. For bulk storage, we developed a dimension-based KNN classifier that exploits the inverse relationship between pallet depth and apparent size. For double-deep storage, we implemented a color-based KNN classifier leveraging RGB attenuation with depth. Faster R-CNN achieved 97.2% average precision, outperforming both standard YOLO (83.8%) and YOLO with dark enhancement (90.8%). The counting algorithms achieved 86% accuracy for bulk storage and 100% accuracy for optimized double-deep storage. Our experimental results demonstrate that combining enhanced detection models with configuration-specific counting algorithms provides an effective solution for autonomous warehouse inventory management, though challenges remain in handling imbalanced training data and generalizing to diverse warehouse environments.**

***Index Terms*—object detection, object counting, computer vision, inventory management**

I. Introduction

Automated inventory management systems are critical for optimizing modern warehouse operations. Recent advances in AI technologies have demonstrated nearly 100% accurate demand forecasting for business applications [1], driving a transformation in traditional warehouse inventory management through computer vision and artificial intelligence. The maturity of AI-based autonomous systems for industrial applications has been comprehensively surveyed by [2], demonstrating readiness for deployment in complex environments like warehouse inventory management. Traditional inventory counting methods are labor-intensive, error-prone, and lack real-time updates. This research addresses the specific challenge of automatically counting pallets in warehouse storage locations using images captured by autonomous robots. While existing computer vision systems can reliably detect pallets [3], [4], they struggle to provide accurate counts in complex scenarios involving stacked or double-deep storage configurations, particularly under variable lighting conditions [5].

Our primary objective is to develop a computer vision system that automatically counts pallets in specific bin locations based on images from autonomous cycle counting robots. The implementation allows enhanced counting accuracy, reduced labor costs, and real-time monitoring capabilities.

We introduce an intelligent inventory approach combining a specialized pallet detection model for low-light environments with precise quantity determination algorithms. Building on prior work, we broaden our scope to diverse warehouse lighting scenarios and introduce novel counting algorithms that calculate pallet quantities within each bin using prior knowledge of storage configurations.

The paper is organized as follows: Section II reviews related work. Section III presents our problem formulation, data sources, and performance metrics. Section IV describes the proposed detection models and counting algorithms. Section V presents experimental results, and Section VI concludes with limitations and future work.

II. Literature Review

A. Object Detection

Two-stage detection algorithms like Faster R-CNN [6] provide superior accuracy through region proposal networks, while one-stage methods like YOLO [7] offer real-time performance. In warehouse applications, [5] successfully applied Faster R-CNN with Swin Transformer for pallet surface detection, achieving high accuracy under optimal lighting. However, their approach struggles in the low-illumination conditions that characterize many warehouse environments. To address this limitation, [8] proposed a night-vision detector based on RFB-Net, incorporating feature pyramid networks to compensate for feature loss in dark conditions. Recent work has demonstrated successful computer vision applications in challenging conditions. Aibin et al. [9] developed advanced object detection techniques for UAV imagery in variable lighting, while Sharma [10] achieved efficient detection near railway tracks using similar approaches. These works validate the potential of modern detection algorithms with lighting challenges.

B. Object Counting

Traditional counting approaches that detect individual objects often struggle with occlusions and stacked items. Authors of [11] introduced regression-based Counting CNN for density estimation rather than individual detection, offering a promising alternative. For warehouse-specific applications, [12] proposed KAZE feature-based counting that demonstrates robustness to lighting variations, while [13] employed SURF features for inventory control. Recent work by [5] pioneered machine vision for warehouse inventory, though their approach required favorable lighting to achieve reliable results.

C. Research Gap

While these studies demonstrate the potential for computer vision in warehouse management, we identify significant gaps that prevent practical deployment. Current methods achieve reasonable performance under controlled conditions but fail to address the realities of warehouse operations. Specifically, accurate detection under variable warehouse lighting remains problematic—particularly in deep storage areas where illumination decreases dramatically. Models trained on well-lit images suffer substantial performance degradation when deployed in darker warehouse sections. Furthermore, precisely counting stacked or double-deep pallets presents unique challenges that existing work has not adequately addressed. Standard detection algorithms cannot differentiate between single pallets and multiple stacked units when viewing from the limited angles available to warehouse robots. Perhaps most critically, current approaches fail to leverage the domain-specific knowledge readily available in warehouse settings—standardized storage configurations, known pallet dimensions, and established placement rules that could significantly improve counting accuracy. To our knowledge, no existing work provides a complete pipeline from low-light detection to configuration-specific counting algorithms, highlighting the need for specialized solutions tailored to autonomous warehouse inventory management.

III. Problem Statement

The current robot has the ability to capture images of pallets from specific locations in the warehouse. With these images, we can retrieve the barcode of the pallets and cross-reference it with the Item Master Report, which records inventory data of the warehouse operation. This allows us to have live information on the dimensions of bins and single pallets. By using object detection techniques in computer vision, we can determine the edges of the actual pallets in the image. With information on the dimensions of single pallets, the model can then estimate the number of pallets within the bin location. However, there are difficulties we need to tackle, such as the extreme lighting conditions in the warehouse that could lead to low-quality images, and the limitation of the angle at which the robot can capture images. To summarize, the existing technology enables the robot to identify the type of product in a specific location. This research aims to develop a model that estimates the number of pallets at that location.

A specific implementation that combines Faster R-CNN and Swin Transformer for box surface detection has gained popularity in the literature [5]. This approach customizes the detection model to generate the most suitable result which is the crucial parameter for the subsequent counting algorithm. However, it's important to highlight that the detection model was trained using images captured under optimal lighting conditions, a scenario quite distinct from our source images. Faster R-CNN cannot perform optimally under low-illumination conditions due to the loss of valuable information in deeper layers. This occurs because parts of the object are often merged into the dark background during convolution operations. Instead of enhancing the images used for training detection model, a night version detector based on the representative state-of-the-art detector RFB-Net [14] was proposed. Authors of [8] also introduced feature pyramid network into detection layers and added context information into the detection backbone to compensate for the loss of features.

To fulfill the research goal, we propose a customized pallet detection model designed to accurately detect pallets and generate tailored results for the sequencing pallet counting algorithm. A brief overview of the process is illustrated in Fig. 1. In the proposed pallet detection model, we enhance performance under low-illumination conditions compared to the model proposed by [5]. For the pallet counting algorithm, we combine the results of the detection model with prior knowledge to create a calculating algorithm, enabling accurate pallet quantity estimation.

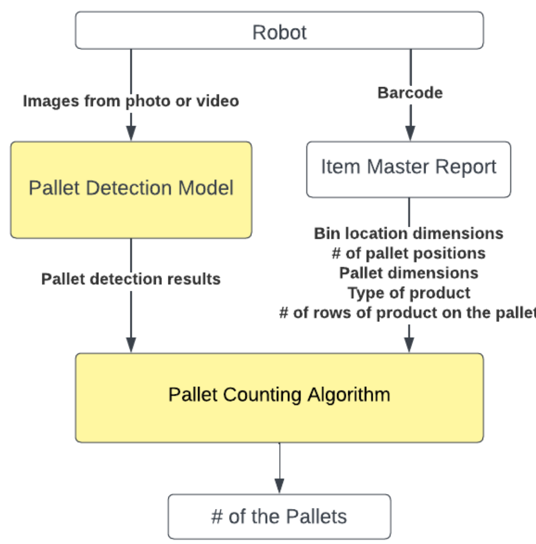

Fig. 1: An overview of the proposed process

By processing the captured image, the pallet detection model will output the pallet location x_i in the image, as shown in Fig. 2. Through the pallet counting algorithm, which will be discussed in the later section, the number of pallets N_p could be estimated based on the detection model output. The Item Master Report contains a number of pallets N_r recorded during warehouse operation and will be used to evaluate the accuracy of our approach according to the performance metrics discussed later in this section.

A. The Source of Data

The Autonomous Cycle Counting Robot provided by In-Dro Robotics captures the source images in the warehouse. The robot fulfills scheduled missions within specified aisles, capturing images or videos of each bin location along the

Fig. 2: Pallet Detection Model result

way. The robot is restricted to following paths around the bins and cannot move through the bins. Additionally, the robot has the capability to capture pictures from various heights. Fig. 3 shows some of the sample images taken by the robot. We can observe from the sample images that the light conditions in the middle of the bin are not as good as those at the front of the bin. To ensure accurate pallet detection, a model designed for low-illumination will be introduced and discussed in detail in the following sections.

(a) (b)

Fig. 3: Sample images taken by the robot

Operators in the warehouse always put items from back to front while placing them on shelves. This means that if a pallet is located in the front position, there is also a pallet in the back position. The Barcode software is utilized to identify the barcode attached on the bin, determining the product type in the location, and also to identify the bin location. The dimensions of the bin and the pallet can be obtained from other models. Therefore, in this paper, we consider the location of the bin, the dimensions of bin and pallets as prior knowledge.

B. Performance Metrics

In our research, we aim to assess two key components: the performance of our pallet detection model and the effectiveness of our pallet counting algorithm.

For the detection model, we employ Average Precision (AP) metrics to evaluate its accuracy and robustness, which collectively provides a comprehensive overview of the model's ability to identify and localize pallets in images. AP serves as a pivotal indicator, quantifying the accuracy of the model by taking into account two fundamental metrics: precision and recall [15]. Precision gauges the model's ability to accurately detect pallets, measuring the ratio of true positives to all positive predictions. In contrast, recall assesses the model's effectiveness in capturing all actual positive instances [16]. AP adeptly balances these two metrics by generating a precision-recall curve and subsequently computing the area under this curve. This calculated area encapsulates the model's overall performance, with a higher AP score signifying enhanced accuracy in pallet detection.

We complement the standard AP metrics with specialized AP variants, including AP50 and AP70, to provide a comprehensive assessment of our model's performance across IoU thresholds [4].

For the pallet counting algorithm, we employ Mean Absolute Error (MAE) and Mean Squared Error (MSE) metrics, which are standard in regression tasks. The formulas for these metrics are:

$$MAE = \frac{1}{M} \sum_{i=1}^{M} |N_r - N_p|$$

$$MSE = \frac{1}{M} \sum_{i=1}^{M} (N_r - N_p)^2$$

where M is the sample size, N_r is the actual number of pallets, and N_p is the predicted number of pallets.

IV. Algorithms

A. Pallet Detection Models

1) Faster R-CNN for Low-Light Conditions: We use Faster R-CNN with ResNet-50 Feature Pyramid Network (R50-FPN) as our backbone for pallet detection. The R50-FPN combines ResNet-50's strong feature extraction capabilities with FPN's multi-scale feature maps, effectively addressing scale variations when pallets appear at different depths in warehouse images. The architecture operates through a two-step region-based approach. First, the network scans the image for objects without considering their class, then classifies them in the second step. The feature map generated by R50-FPN feeds into the Region Proposal Network (RPN) to predict region proposals. After the Region of Interest (ROI) Align extracts fixed-size features from these proposals, a classifier determines whether each region contains a pallet or a hole.

Images captured by the robot are annotated using Labelme [17], classifying them into two categories: 'pallet' and 'hole'. These annotated images are divided into training and validation sets. During training, we carefully tuned several crucial parameters to optimize performance. Our experiments showed best results with a learning rate of 0.0001 and 1000 iterations. The model outputs bounding box coordinates for both pallets and holes, which serve as vital inputs for our counting algorithms.

2) YOLO with Dark Enhancement: YOLOv7 adopts a one-step, end-to-end approach, dividing the input image into a grid where each cell predicts object bounding boxes and their associated class probabilities. However, YOLOv7 struggles in the poor lighting conditions found deep within warehouse bins.

To address this, we utilize a pre-trained Convolutional Neural Network [18] called DarkEnvModel as a preprocessor in the detection pipeline.

The DarkEnvModel features multiple convolutional layers that transform the input image's features to improve visibility. The model applies operations like max pooling and upsampling to maintain important information while reducing image size. ReLU activation [19] adds non-linearity, enabling the network to learn complex patterns. We use a set of specialized loss functions serving distinct purposes: enhancing color, improving spatial clarity, adjusting exposure, regulating variation, preserving saliency, and enhancing perception. Together, these loss components guide the model towards generating images well-suited for enhanced object detection in dark environments. This approach aligns with recent findings by Yang et al. [20], who demonstrated that optimizing color spaces beyond standard sRGB significantly improves object detection performance in challenging visual conditions, achieving better precision in environmental assessment tasks with similar lighting constraints.

B. Configuration-Specific Counting Algorithms

1) Dimension-Based Counting for Bulk Storage: For bulk storage locations, the maximum storage stacks are 3 high, 2 wide, and 4 deep, meaning we have 6 fixed possible positions when looking from the front. We noticed that the number of pallets in each position is discrete—it can only be 0, 1, 2, 3, or 4 pallets. Our key insight is that for each fixed position, the fewer pallets stacked, the further the first pallet appears, and consequently the smaller it looks in captured images.

This observation allows us to convert the counting problem into a classification task where the five class labels represent the number of pallets, and the input consists of the dimensions of the first visible pallet. We use K-Nearest Neighbors as our classifier, with the input being the length and height of the pallet bounding box obtained from the detection step.

For position (i, j) with observed dimensions (l, h), we calculate the Euclidean distance to each training point t in S_{train} and store the results with labels in list D. After sorting D by distance in ascending order, we find the majority label among the first K elements. The total count is $n = \sum_{positions} label_{majority}$.

To determine the best K value, we use cross-validation, iterating through the range (k_{min}, k_{max}) and evaluating performance using $p_{Cal}(S_{train}, S_{test}, k)$. The algorithm selects the k value that optimizes our performance metrics.

The main challenges we face include dimension differences becoming smaller between 1, 2, or 3 pallets, making the data points more concentrated and harder to distinguish. Additionally, we have limited training data for these intermediate cases since warehouse positions are either empty (0) or full (4).

2) Color-Based Counting for Double-Deep Storage: We developed a color-based supervised learning classification approach specifically for double-deep storage locations. This method leverages two key observations: indoor warehouse maintains relatively stable artificial lighting. However, illumination decreases significantly in bin depths, and warehouse operating procedures specify placing items from back to front.

RGB data (r, g, b, a) from the central pixel is extracted from the detection model and labeled with the corresponding number of pallets. Since RGB colors tend to cluster together for a given pallet count, we use this color data as our primary feature. We discovered that extracting pixels from multiple locations—center, upper, and right positions—and computing their average significantly improves accuracy compared to single-pixel measurements, particularly when handling damaged pallets.

For each test data point x_{test}, we find the Euclidean distance to all training data points x_{train}, sort these distances, choose the nearest K points, and assign y_{pred} based on the majority class among these neighbors. We apply KNN classification with $K = 7$ (determined empirically) to predict pallet counts based on averaged RGB values. Interestingly, we found that excluding the blue channel and using only RG values provides more stable results under warehouse LED lighting conditions.

Potential sources of error include varying illumination near aisles where areas receive more light, and damaged pallets with abnormally darkened pixels that could skew predictions.

C. Experimental Setup

This study uses Detectron2 [21] for implementing and training Faster R-CNN. Detectron2 is an open-source computer vision library developed by Facebook AI Research, built on the PyTorch deep learning framework. We annotated 71 images from robot video footage using Labelme [17], converting the output to COCO format for Detectron2 compatibility. We acknowledge that 71 images represents a preliminary dataset. To mitigate overfitting risks, we employed pre-trained models where available and focused on simple classifiers (KNN) that are less prone to overfitting on small datasets.

For YOLO implementation, both DarkEnvModel and YOLO leverage Google Colab with GPU support. The datasets are partitioned into training, validation, and test sets. We adjust key hyperparameters, including learning rate, batch size, and epochs, through systematic fine-tuning to optimize performance based on precision, recall, F1 score, and specialized Average Precision variants. As reported in Section IV, optimal performance was achieved with a learning rate of 0.0001 and 1000 iterations for Faster R-CNN.

For bulk storage counting, we use bounding box outputs from the detection models as input. We evaluate the range of $K = 7$ values for the KNN classifier and test three distance-calculating algorithms: brute-force search, BallTree, and KDTree.

For double-deep storage, classification distinguishes between 0, 1, and 2 pallets, where 0 indicates no bounding box detected. We test the algorithm's robustness with edge cases, including damaged pallets and extreme lighting conditions.

V. Results

In our study, we conducted a thorough comparison of the performance of two proposed pallet detectors, focusing on crucial metrics such as Average Precision (AP) and Average Precision at IoU thresholds of 50% (AP50) and 70% (AP70), as elaborated in Section III. The findings are presented in Table I, providing a detailed overview of the pallet detection models utilizing the proposed approach across a consistent dataset of sample images.

The results presented derive from the optimal performances of the models after hyperparameter tuning. Specifically, as illustrated in Fig. 4 and Table I, YOLO in Dark enhances detection performance in low-illumination environments and outperforms standard YOLO in pallet detection, achieving higher average precision scores across various metrics. The improvement of 7% in AP demonstrates the effectiveness of the DarkEnvModel preprocessing step, particularly crucial for warehouse depths where lighting degrades significantly.

In our pallet detection case, the two-stage object detection model Faster R-CNN outperformed the one-stage object detection model YOLO across all metrics, as visible in Fig. 5. This superior performance can be attributed to Faster R-CNN's region proposal mechanism, which provides more accurate localization in complex warehouse scenes with multiple overlapping pallets.

Fig. 4: Comparison of YOLO and YOLO in Dark

Fig. 5: The detection result of the faster R-CNN

From the detection model, we obtain the location of pallet bounding boxes. The robot then identifies the storage location type: either bulk storage or double deep, applying the appropriate counting algorithm. While controlled comparison requires identical test datasets unavailable from prior work, we provide qualitative context for our results. Detection-only approaches like [5] achieve high accuracy under optimal lighting but lack counting mechanisms for stacked configurations. Our dimension-based and color-based counting algorithms represent novel contributions that exploit domain knowledge (storage configurations, pallet dimensions) unavailable to pure detection methods. For the counting step, alternative approaches could include regression models (linear, logistic) or support vector machines; however, KNN was selected for its simplicity, interpretability, and resistance to overfitting on small datasets.

TABLE I: Performance for pallet detection models

Metric	Faster R-CNN	YOLO	YOLO in Dark
Pallet AP	**97.2**	83.8	90.8
Pallet AP50	**92.2**	81.8	87.2
Pallet AP70	**95.3**	74.5	79.1
Hole AP	**82.8**	N/A	N/A
Hole AP50	**80.5**	N/A	N/A
Hole AP70	**77.4**	N/A	N/A

A. Bulk Storage Counting

For bulk storage cases, we created a specific counting algorithm based on dimension information from the detection model. We applied KNN as the classifier to determine the number of pallets at each position. The results for bulk storage counting are presented in Table II.

The accuracy score of 0.86 falls short of the 0.90 threshold we targeted. This limitation stems primarily from the imbalanced training dataset—warehouse operations typically maintain positions either empty (0 pallets) or completely full (4 pallets), resulting in insufficient training samples for intermediate cases (1, 2, or 3 pallets). This scarcity of intermediate cases makes the KNN classifier struggle to accurately distinguish between similar dimension ranges. Additionally, we did not implement position-specific weights, which could account for the varying reliability of predictions at different stack positions where occlusion patterns differ.

TABLE II: Performance for KNN counting model

Storage Type	Accuracy Score	MAE	MSE
Bulk Storage	0.86	0.14	0.14

B. Color Classification for Double Deep Storage

The color classification approach is specifically designed for double deep storage locations. Using bounding boxes from the pallet detection step, we extract pixels for our classification model. Initially, we tested using RGB values from only the center pixel as input, resulting in the baseline performance shown in Table III.

Upon investigating the inaccuracies, we discovered that single-pixel sampling proved unreliable due to surface variations on pallets—damage, dirt, or printing could cause any individual pixel to be unrepresentative. To address this, we optimized our model to extract pixels from three locations: center, upper (20% above center), and right (20% to the right of center). Through empirical testing, we found that excluding the blue channel improved stability under warehouse LED lighting, as blue values showed higher variance. Computing the mean of red and green values from these three pixels provided a more robust feature for classification.

TABLE III: Performance for color-based classification

Metric	First Experiment	Optimized
Accuracy	0.86	1.00
Precision	1.00	1.00
Recall	0.67	1.00
F1 Score	0.80	1.00

The large improvement to 100% accuracy in our optimized model validates the multi-pixel averaging approach. The perfect precision in both experiments indicates our model never falsely predicted pallets where none existed, while the recall improvement from 0.67 to 1.00 shows the optimized approach successfully detected all pallets that were present. This performance improvement demonstrates that spatial averaging effectively handles the noise introduced by surface irregularities on warehouse pallets.

VI. Conclusion

We successfully developed a two-pronged approach combining enhanced detection models for low-light conditions with configuration-specific counting algorithms, achieving 97.2% AP for Faster R-CNN detection and up to 100% accuracy for optimized color-based counting in double-deep storage. However, certain limitations constrain broader implementation.

First, the algorithm may occasionally misinterpret stacked goods as pallets when there are similarities in patterns or shapes. Second, our method remains specific to the tested warehouse configuration. While successfully categorizing storage into "Bulk" and "Double Deep" types, the approach has not been generalized to diverse warehousing scenarios. The limited dataset of 71 training images with insufficient samples for intermediate pallet counts (1-3 pallets) constrains the model's ability to handle edge cases effectively.

Despite these limitations, our work demonstrates the feasibility of configuration-specific counting algorithms that leverage domain knowledge about warehouse operations. The dimension-based approach for bulk storage and color-based method for double-deep storage both exploit physical properties inherent to warehouse environments, suggesting that similar domain-specific insights could improve automation in other industrial settings. The significant performance difference between standard YOLO (83.8% AP) and our enhanced models highlights the importance of addressing environmental conditions in real-world deployments.

Future work should focus on three key areas: (1) expanding the dataset with diverse warehouse environments and more representative scenarios through varied camera settings and robot routes; (2) exploring alternative CNN architectures and machine learning classifiers through comparative analysis; and (3) extending the methodology to various warehouse configurations, layouts, and storage systems. Additionally, integrating temporal information from video sequences rather than single frames could improve counting accuracy by providing multiple viewpoints of ambiguous cases. These improvements would enable a more robust and universally applicable solution for autonomous pallet counting, advancing warehouse inventory automation across diverse operational contexts.

References

[1] R. Dash *et al.*, "Application of artificial intelligence in automation of supply chain management," *Journal of Strategic Innovation and Sustainability*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:201148169

[2] M. Aibin *et al.*, "Survey of rpas autonomous control systems using artificial intelligence," *IEEE Access*, vol. 9, pp. 167 580–167 591, 2021.

[3] R. Varga *et al.*, "Robust pallet detection for automated logistics operations." in *VISIGRAPP (4: VISAPP)*, 2016, pp. 470–477.

[4] M. Zaccaria *et al.*, "A comparison of deep learning models for pallet detection in industrial warehouses," in *2020 IEEE 16th international conference on intelligent computer communication and processing (ICCP)*. IEEE, 2020, pp. 417–422.

[5] H. Yin *et al.*, "A vision-based inventory method for stacked goods in stereoscopic warehouse," *Neural Computing and Applications*, vol. 34, no. 23, pp. 20 773–20 790, Dec 2022. [Online]. Available: https://doi.org/10.1007/s00521-022-07551-4

[6] S. Ren *et al.*, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[7] J. Redmon *et al.*, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[8] Y. Xiao *et al.*, "Making of night vision: Object detection under low-illumination," *IEEE Access*, vol. 8, pp. 123 075–123 086, 2020.

[9] M. Aibin *et al.*, "Advancing forest fire risk evaluation: An integrated framework for visualizing area-specific forest fire risks using uav imagery, object detection and color mapping techniques," *Drones*, vol. 8, no. 2, p. 39, 2024.

[10] R. Sharma *et al.*, "Aerial footage analysis using computer vision for efficient detection of points of interest near railway tracks," *Aerospace*, vol. 9, no. 7, p. 370, 2022.

[11] D. Onoro-Rubio *et al.*, "Towards perspective-free object counting with deep learning," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14*. Springer, 2016, pp. 615–629.

[12] T. Sharma *et al.*, "Object counting using KAZE features under different lighting conditions for inventory management," in *2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. IEEE, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/9174578/

[13] N. K. Verma *et al.*, "Vision based object counting using speeded up robust features for inventory control," in *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, pp. 709–714. [Online]. Available: http://ieeexplore.ieee.org/document/7881432/

[14] S. Liu *et al.*, "Receptive field block net for accurate and fast object detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[15] A. Tharwat, "Classification assessment methods," *Applied computing and informatics*, vol. 17, no. 1, pp. 168–192, 2020.

[16] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.

[17] K. Wada, "labelme: Image polygonal annotation with python," https://github.com/wkentaro/labelme, 2018.

[18] S. Albawi *et al.*, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6.

[19] V. Nair *et al.*, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[20] Z. Yang *et al.*, "Beyond srgb: Optimizing object detection with diverse color spaces for precise wildfire risk assessment," *Remote Sensing*, vol. 17, no. 9, p. 1503, 2025.

[21] Y. Wu *et al.*, "Detectron2," https://github.com/facebookresearch/ detectron2, 2019.