Published by AI and R&D Department, Orel IT – 2024

Comprehensive Analysis of Segmentation models for Multiclass Segmentation

Kavitha Pathmanathan, Sasadara Adikari*

AI and R&D Department, OREL IT, Sri Lanka

Abstract

This paper investigates the application of image segmentation for automating planogram compliance in retail stores. The main objective of this study is to identify the best suitable segmentation model for accurately segmenting the inner and outer shelves in a planogram zone of the supermarket rack. To identify the optimal segmentation model architecture for our task, we trained and evaluated seven segmentation architectures, including U-Net, LinkNet, FPN (Feature Pyramid Network), PAN (Pyramid Attention Network), PSPnet (Pyramid Scene Parsing Network), MAnet (Multi-Attention-Network) and DeeplapV3. Our dataset consists of 100 annotated supermarket shelf rack images. We utilized data augmentation techniques to increase the training data and prevent overfitting. The models were evaluated using Intersection over Union (IoU) score, accuracy, training time, and model size. All evaluated models achieved high performance in object segmentation (IoU > 0.977) and pixel-wise classification accuracy (accuracy > 0.992). U-Net and LinkNet achieved marginally higher average IoU scores (0.986 and 0.987, respectively). However, U-Net emerged as the optimal model due to its balance between segmentation accuracy, training time (around 4 hours), and model size (97 MB). Additionally, we demonstrated the potential for using segmentation masks to detect corner points of the inner shelves and the planogram zone.

Keywords: *Computer vision, Deep learning, Multiclass Segmentation, Supermarket rack detection, Planogram compliance*

1. Introduction

In today's competitive retail market, maximizing customer engagement and optimizing product placement are essential for success. Planograms, regulate how products should be placed on shelves of a retail store which play a vital role in achieving these goals [1]. Ensuring adherence to planograms, known as planogram compliance, directly impacts factors like product discoverability, brand visibility and ultimately sales. However, it is a time-consuming and resource-intensive task for manually verifying planogram compliance across numerous stores [2]. This is where image segmentation techniques come into play. We can automate the process of planogram compliance assessment by effectively segmenting the planogram area within supermarket rack sections.

Automated planogram compliance can be addressed through various technologies including integration of sensor networks, internet of things, computer vision, machine learning, and data analysis [2]. While various computer vision and pattern recognition methods address planogram compliance control and object detection in shelf images [3], achieving high accuracy depends on the ability to accurately identify the planogram area within the supermarket rack image.

This paper explores the application of image segmentation for supermarket rack planogram zones. We evaluated the performance of various deep learning architectures, including U-Net, LinkNet, FPN, PAN, PSPnet, MAnet, and DeepLabV3 [6-9], in accurately segmenting the inner and outer shelves within the designated planogram region (figure 1). As illustrated in figure 1, yellow color area represents the planogram zone of the rack, red color mask indicates the inner shelves, and the blue color mask indicates the outer shelves of a supermarket rack section. Accurate segmentation of these regions is essential for detecting and classifying each product in the planogram zone.

The following sections detail the previous studies that address planogram compliance (section2), methodology used for training and evaluation (Section 3), present the results of our experiments (Section 4), and discuss the findings (Section 5).

Figure 1. supermarket rack with merged annotated mask

2. Literature Review

In this section, we discuss some relevant studies that address planogram compliance and image segmentation using image processing techniques and computer vision.

Paper	Highlight	Deficiencies
Planogram Compliance Checking Based on Detection of Recurring Patterns [1]	Proposed a novel method for planogram compliance checking which doesn't require product template images for training. It leverages recurring pattern detection to identify product layouts within the created planogram zone.	This approach may struggle with significant deviations from the planned layout or highly dynamic product placements
Embedded Planogram Compliance Control System [2]	Proposed a complete embedded system to planogram compliance control The object detection block was based on YOLOv5 as the deep learning method and local feature extraction.	The specific performance and limitations of the chosen deep learning model weren't addressed.

Planogram compliance control via object detection, sequence alignment, and focused iterative search [3]	Proposed a method for planogram compliance that uses object detection, sequence alignment, and focused iterative search	Not directly focus on image segmentation, it denotes the importance of accurate identification of objects within the planogram zone for achieving high accuracy in compliance
Computer Vision Based Planogram Compliance Evaluation [4]	Focuses on evaluating planogram compliance using computer vision techniques. Proposed a new metric for segmentation accuracy	Didn't mention the specific models or architectures used for segmentation

Table 1. Summary of previous studies

Existing research offers valuable insights for planogram compliance using computer vision. However, some limitations are identified:

- Limited focus on image segmentation: While some studies use computer vision for planogram compliance and object detection, they didn't explicitly address image segmentation as the core technique.
- Lack of comparative analysis: Direct comparisons between different segmentation model architectures in this context might be missing

Our task builds upon these existing works by focusing specifically on image segmentation for supermarket rack planogram compliance. We assess the performance of various segmentation model architectures for accurate segmentation of the planogram zone within the rack image.

3. Methodology

3.1. Data Preprocessing

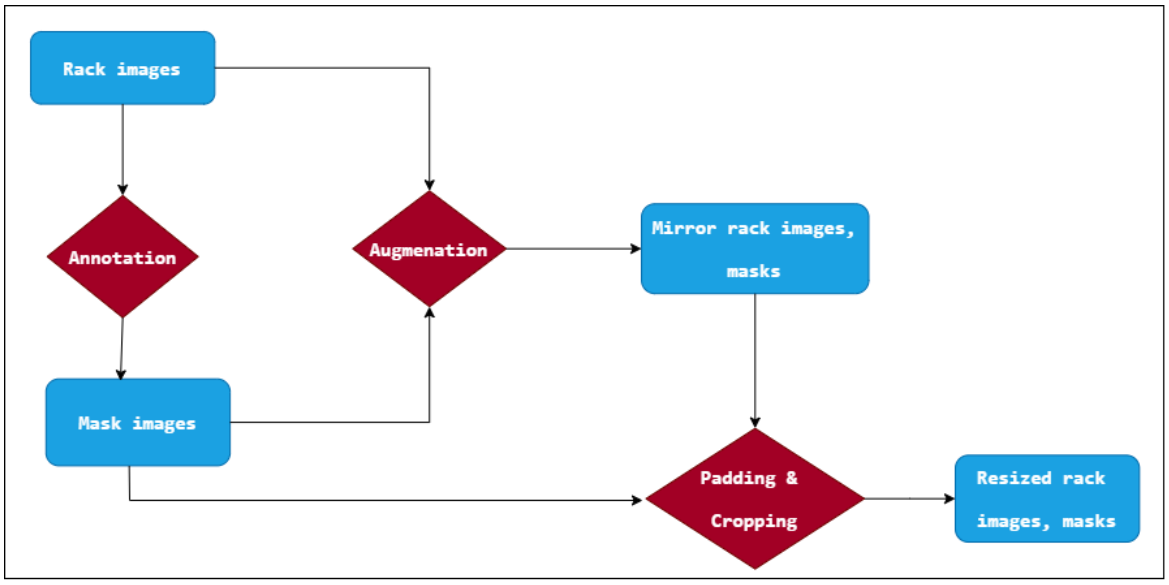

Figure 2. Data preprocessing workflow

Our dataset consists of 100 rack images. However, the images have varying dimensions (width and height). For this segmentation task, resizing them could lead to a loss of important information. Therefore, we followed a strategy of fixing the image size to 2048 x 4016 pixels. Images with different dimensions were padded and cropped to match this size, as illustrated in Figure 3. We used the CVAT annotation tool to label the images in COCO format. The annotations included three categories: background, inner shelves, and outer shelves. Each class was assigned a distinct pixel value. After the annotation, we created single channel mask and assigned distinct pixel values to represent different categories for each image.

- Background: 0
- Inner shelves: 1
- Outer shelves: 2
- Padding: 3

Due to the limited size of our dataset (100 images), we used data augmentation to artificially increase the training data and prevent the model overfitting. One technique we used was horizontal flipping, which was applied synchronously to both the images and their corresponding masks. This process effectively generated a mirrored dataset.

Figure 3. Padded input image with relevant padded single channel mask.

To ensure that the model has learnt in a generalizes way, we carefully selected a subset of the dataset for training, validation, and testing. The training set comprised 70 original images and their corresponding mirrored counterparts which was generated through data augmentation, along with their respective masks. This aim of this selection process is to incorporate a diverse range of image variations within the training data. Similarly, 20 images and their mirrors were allocated for validation, and 10 images and mirrors were allocated for testing.

3.2. Model Training

3.2.1. Architectures

We explored seven different architectures which leverage a pre-trained ResNet-34 encoder using ImageNet weights: U-Net, FPN, LinkNet, PAN, PSPnet, MAnet and DeepLapv3 for our task. ResNet34 is a deep neural network pre-trained by using ImageNet dataset specific for image classification tasks. It has learnt high level features from a large collection of labeled images. This pre-trained backbone provides a powerful base for these segmentation models, as it can effectively capture important image features that can be further refined for the specific task of rack image segmentation. Each of these chosen architectures provides unique strength and specific structure to extract the features from the images and ultimately generate the mask.

3.2.2. Parameters

Input size 512 x 1024 pixels was used to preserve the loss of dimensionality information within the rack images and potentially improve segmentation accuracy. The input images are in RGB format and masks are single-channel image, where each pixel value corresponds to a specific object class. Dice Loss was used as loss function for this segmentation tasks. It performs well when dealing with multiple object classes, as in our case with background, inner shelves, and outer shelves. Adam optimizer was used as optimizer to update the weight during the backward propagation. due to its effectiveness in adapting learning rates and facilitating convergence. We trained each model for 500 epochs to ensure sufficient learning and potentially avoid underfitting.

$$DiceLoss(y, \bar{p}) = 1 - \frac{2y\bar{p}+1}{y+\bar{p}+1} \qquad (1)$$

Here, y represents the true segmentation of the image, and $\bar{p}$ represents the predicted segmentation generated by the machine learning model.

3.3. Model Evaluation

Evaluating the performance of the trained segmentation models is an important step for selecting the most suitable architecture for our specific objective. We used two main metrics to evaluate the effectiveness of the trained models. IoU measures the area of overlap between the predicted segmentation mask and the ground truth mask. The IoU value range is between 0: no overlap and 1: perfect overlap. A higher IoU score denotes better model performance in perfectly detecting object boundaries. Our segmentation dataset has imbalanced class distributions where some classes have many more pixels than others. The accuracy matric calculates the overall proportion of pixels that are correctly classified in the predicted segmentation mask compared to the ground truth mask.

$$IoU\ score = \frac{TP}{TP+FP+FN} \qquad (2) \qquad\qquad Accuracy = \frac{(TP+TN)}{TP+FP+FN+TN} \qquad (3)$$

where TP is the number of true positives, TN is the number of true negatives FP is the number of false positives, and FN is the number of false negatives [4]

4. Result and Discussion

This section describes the evaluation metrics results of the segmentation models. The testing set was used to evaluate the performance of the models. We analyzed and compared the models based on their average IoU score, accuracy, training time, and the size of the models.

4.1. IoU and Accuracy Analysis

Table 2 summarizes the average IoU scores, all models achieved very similar average IoU scores, which shows the strong overall performance of the models in detecting object boundaries accurately over four classes. However, U-Net (0.986) and LinkNet (0.987) provided marginally higher average IoU scores, but the differences are relatively small.

Segmentation Model	*IoU score*	*Accuracy*	*Training time*	*Model size (MP)*
Unet	0.9861	**0.9957**	3:52:17	97
FPN	0.9829	0.9938	3:21:02	92
LinkNet	**0.9867**	0.9954	3:21:46	87
PAN	0.9817	0.9937	3:38:25	86
PSPnet	0.9790	0.9926	**2:52:18**	**86**
MAnet	0.9779	0.9911	4:00:20	127
DeepLapV3	0.9838	0.9940	7:38:42	104

Table 2. Summary of the Evaluation results

4.2. Training Time and Model Size

When consider the training time over 500 epochs and size of the best models, LinkNet, PAN and PSPnet have shorter training times compared to other models. DeepLabV3 has the longest training time by a considerable margin. Similarly, PAN and PSPnet has the smallest model size, followed by LinkNet, U-Net and FPN.

4.3. Selecting the best model

Based on the evaluation results and our specific focus on accurate conner point detection, U-Net emerges as the optimal model. U-Net achieved a very high average IoU score of 0.9861, proving the excellent performance of the model in accurately identifying object boundaries. High accuracy score (0.9957) of U-Net model further demonstrates its effectiveness in pixel-wise classification. The training time of around 4 hours on a powerful Nvidia GPU (6GB) PC and a model size of 97 MB, U-Net appears manageable to our requirements.

4.4. Segmentation and Corner Point Detection

The U-Net model emerged as the most suitable architecture for our rack image segmentation task based on the evaluation results. We used the selected Unet model to predict segmentation masks for unseen images. These masks effectively described the expected object classes such as background, inner shelves, outer shelves within the images. To detect corner points of the inner shelves class based on the predicted segmentation masks, we employed a popular computer vision library called OpenCV.

The results of the segmentation and corner point detection using the Unet model are shown in Figure 4. Image 1 is the predicted segmentation mask where each pixel value corresponds to a specific object class (background, inner shelves, outer shelves). Image 2 is the combination of the

original image with the predicted segmentation mask potentially with transparency to visualize both the image content and the segmentation results simultaneously.

Image 3 is the detected corner points visualization of inner shelves. These points represent the predicted corners of the inner shelves' region within the image. Image 4 is the overall four corner polygon of the inner shelves area. The conner points of the inner shelves after processing the detected corner points from Image 3.

Figure 4. Sample of predicted mask and the conner point detection result.

5. Conclusion

In this work, we trained and evaluated various segmentation models for perfectly segmenting inner and outer shelves in supermarket rack images. Seven pre-trained architectures (U-Net, LinkNet, FPN, PAN, PSPnet, MAnet, and DeepLabV3) were investigated using a dataset of 100 images with data augmentation. The models were evaluated based on their average IoU score, accuracy, training time, and model size.

Key Findings

- All models achieved high overall performance in segmenting objects IoU > 0.977 and pixel-wise classification accuracy > 0.992.
- U-Net and LinkNet provided marginally higher average IoU scores 0.986 and 0.987, respectively.
- Training time and model size varied considerably across the models. PSPnet had the shortest training time, while PAN and PSPnet had the smallest model sizes.

We conclude that U-Net is the optimal segmentation model for our task based on the comprehensive evaluation and requirement of our project goals. This model segmenting the object conners accurately while maintaining acceptable training time and model size constraints.

Future directions

We plan to explore several promising future directions to further enhance the applicability and robustness of our approach:

- Significantly expand our dataset of annotated supermarket rack images. This will encompass a wider variety of rack layouts and angles, lighting conditions, and product types.
- The current work utilizes a ResNet-34 encoder as the backbone for the segmentation models. Future investigations will involve experimenting with alternative encoder architectures.
- Hence Unet achieved better accuracy, we will also consider architecture like: Unet++, EfficientUNet+, ResUnet and ResUnet++ to improve the performance of the model furthermore.

6. References

[1] S. Liu, W. Li, S. J. Davis, C. Ritz, and H. Tian, "Planogram Compliance Checking Based on Detection of Recurring Patterns," IEEE MultiMedia, vol. 23, no. 2, pp. 54–63, Apr. 2016, doi: https://doi.org/10.1109/mmul.2016.19

[2] A. Preprint, M. Erkin, Y. R&d, M. Ticaret, S. Topaloğlu, and C. Ünsalan, "Embedded Planogram Compliance Control System Embedded Planogram Compliance Control System," 2024. Accessed: Jun. 03, 2024. [Online]. Available: https://arxiv.org/pdf/2401.06690

[3] M. Erkin Yücel and Cem Ünsalan. Planogram compliance control via object detection, sequence alignment, and focused iterative search. Multimedia Tools and Applications, 2023

[4] J. Laitala and L. Ruotsalainen, "Computer Vision Based Planogram Compliance Evaluation," Applied sciences, vol. 13, no. 18, pp. 10145–10145, Sep. 2023, doi: https://doi.org/10.3390/app131810145.

[5] "Welcome to segmentation_models_pytorch's documentation! — segmentation_models_pytorch 0.1.0 documentation," segmentation-modelspytorch.readthedocs.io. https://segmentation-modelspytorch.readthedocs.io/en/latest/

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation." Available: https://arxiv.org/pdf/1505.04597v1

[7] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," 2017 IEEE Visual Communications and Image Processing (VCIP), Dec. 2017, doi: https://doi.org/10.1109/vcip.2017.8305148

[8] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic Feature Pyramid Networks," arXiv.org, Apr. 10, 2019. https://arxiv.org/abs/1901.02446

[9] H. Li, P. Xiong, J. An, and L. Wang, "Pyramid Attention Network for Semantic Segmentation," arXiv:1805.10180 [cs], Nov. 2018, Available: https://arxiv.org/abs/1805.10180

[10] R. Li et al., "Multiattention Network for Semantic Segmentation of Fine-Resolution Remote Sensing Images," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–13, 2022, doi: https://doi.org/10.1109/tgrs.2021.3093977

[11] M. Hamdaan, "Multi-Class Semantic Segmentation with U-Net & PyTorch," Medium, Jul. 22, 2021. https://medium.com/@mhamdaan/multi-class-semantic-segmentation-with-u-net-pytorch-ee81a66bba89