Deformable Part Region Network for Automated Waste Recycling

Sonali Pulate[1], Dr. Rekha P. Labade[2], Dr. S.V. Chaudhari[3]

[1]Ph.D. Scholar Department of Electronics and Telecommunication, Sanjivani College of Engineering, Kopargaon, SPPU Pune, Maharashtra, India.
[2]Research Guide, Department of Electronics and Telecommunication, Sanjivani College of Engineering , Kopargaon, SPPU Pune, Maharashtra, India.
[3]Research Coordinator Department of Electronics and Telecommunication Sanjivani College of Engineering ,Kopargaon, SPPU Pune, Maharashtra, India.
Corresponding author: Sonali Pulate (e-mail: sonalipulate@gmail.com), Author(s) Email: Dr. Rekha P.Labade (e-mail: hod.entc@avcoe.org), Dr.S.V. Chaudhari (e-mail: chaudharisachinetc@sanjivani.org.in

ABSTRACT

Automated waste recycling is a critical step toward sustainable waste management, requiring efficient and intelligent sorting systems. Traditional recycling methods rely on rule-based or handcrafted feature extraction approaches, which often struggle with complex waste compositions. The increasing demands of automated waste recycling systems necessitate advancements in object detection technologies, particularly for deformable objects such as waste materials. This work introduces a new deep learning framework, the Deformable Part Region Network (DPR-Net), which excels in detecting and segmenting deformable objects in challenging, unstructured environments such as waste recycling facilities. By integrating deformable convolutional networks with region-based Convolutional Neural Network (CNN) architectures, the DPR-Net dynamically adapts to the geometric variations of irregular waste items, enhancing both detection precision and segmentation accuracy. Our approach leverages the ZeroWaste dataset, a comprehensive dataset tailored for recycling scenarios, to train and validate the model. Results indicate significant improvements in detection metrics over traditional methods, providing a robust solution for automated waste sorting and contributing to environmental sustainability efforts. The presented system shows result with Recall 93.1, Precision 92.5, mAP 85.3, FPS 2.2 and Accuracy 94.2. Extensive experiments on benchmark waste datasets demonstrate that DPRN outperforms existing state-of-the-art methods.

KEYWORDS: ZeroWaste Dataset, Automated waste recycling, Object Detection, Deep Learning. intelligent sorting systems.

How to Cite: Sonali Pulate, Rekha P. Labade, S. V. Chaudhari., (2025) Deformable Part Region Network for Automated Waste Recycling, Vascular and Endovascular Review, Vol.8, No.11s, 65--77.

INTRODUCTION

The document begins with a discussion on the importance of enhancing object detection in waste recycling, highlighting the challenges posed by deformable and irregular objects. It outlines previous approaches and introduces the need for a more adaptive model, setting the stage for the proposed DPR-Net. Waste recycling is a critical process for environmental sustainability, yet traditional recycling methods rely heavily on manual sorting, which is labor-intensive and prone to inefficiencies. With the advancement of artificial intelligence and computer vision, automated waste sorting systems have gained significant attention. However, existing deep learning-based methods struggle with occlusions, deformable objects, and varying waste appearances. To address these challenges, we propose a Deformable Part Region Network (DPR-Net) for automated waste recycling. DPR-Net leverages deformable convolutions and part-based feature extraction to improve waste object detection and classification. Unlike conventional convolutional neural networks (CNNs), which operate on rigid feature maps, DPR-Net can dynamically adapt to variations in waste objects, enhancing recognition accuracy.

The proposed framework integrates region-based feature learning and deformable part-based modeling to capture intricate object structures. By utilizing deformable convolutions, DPR-Net can focus on discriminative regions within waste items, effectively handling irregular shapes, occlusions, and varying lighting conditions. This approach enhances segmentation precision and classification robustness, which are crucial for real-world waste management applications. Moreover, our model incorporates an attention-driven mechanism that prioritizes significant object parts while ignoring irrelevant background noise. The fusion of multi-scale feature representations further refines object localization and boosts recycling efficiency. Additionally, we introduce a self-supervised learning paradigm to enhance the generalization ability of DPR-Net, reducing dependency on extensive labeled datasets.

Experimental results demonstrate that DPR-Net outperforms existing object detection frameworks, achieving superior accuracy in waste categorization tasks. By automating waste sorting with intelligent vision-based models, this research contributes to efficient recycling processes, minimizing human intervention and promoting environmental sustainability. The proposed method has the potential to revolutionize waste management by making recycling systems smarter, faster, and more reliable. Automated waste recycling is essential for sustainable waste management, yet traditional methods rely on inefficient manual sorting. Deep learning-based solutions have been explored, but they struggle with occlusions, deformations, and varying waste appearances.
To address these challenges, we propose a Deformable Part Region Network (DPR-Net) that enhances object detection and classification for waste sorting. Unlike standard CNNs, DPR-Net adapts dynamically to object variations using deformable

convolutions, allowing it to focus on key object parts. Our model integrates region-based feature learning and part-aware attention mechanisms, improving segmentation precision and classification robustness. By leveraging multi-scale feature extraction, DPR-Net can accurately detect complex waste objects despite shape irregularities. Additionally, we incorporate self-supervised learning to reduce dependency on labeled datasets, enhancing the model's generalization ability. Experimental results show that DPR-Net outperforms existing waste classification methods, making recycling faster, smarter, and more efficient. We use the global holistic features of object shape to tackle the object recognition and segmentation challenge. Global shape representations can only be used successfully with accurate object segmentation because they are extremely vulnerable to the clutter that is unavoidably present in actual photos.

RELATED WORK:

Toshev et al. [1] provided an Open GL collision detection technique for polygonal deformable objects. The OpenGL selection mode and an axis-aligned bounding box structure serve as the foundation for the technique. It works well with deformable objects that have a lot of polygons, and the performance cost of adding more polygons is comparatively low. Jianjun Li et al. [2] explained to coordinate the prediction, extract the target area's various attributes using dual branch parallel processing. Saiprasad et al. [3, 34] offer a multi-stage, effective method for recognizing objects in real-world, cluttered photos that is resistant to rotation, scale, and interclass variability. Tim F. et al. [4] offer computer vision and medical image analysis, statistical models of the form and appearance of deformable objects are now frequently employed. Tiago Silva et al. [5] suggested a 3D tracking method with RGB-D photos as input for deformable objects that makes use of machine learning and deep learning techniques. Yanfeng et al. [6] suggests TransE Det, an aircraft detection technique for aerial photos that is based on the Transformer module and EcientDet algorithm. By integrating the Transformer, which simulates the long-range dependency for the feature maps, with the EfficientDet algorithm, we enhanced it. Hongxia Yu et al. [7] explained research suggests a better Yolox detection algorithm (BGD-YOLOX) to enhance the effectiveness of small item detection. Peicheng Shi et al. [8] proposed to improve global modeling efficiency, suggest a region-based Reconstructed Deformable Self-Attention that focuses on key areas. Shubham Tulsiani et al. [9] models provide a "low-frequency" shape by capturing top-down information about the primary global patterns of shape variation within a class. Lifeng Liu et al. [10] described approach can achieve a decent segmentation in spite of shape distortion, clutter, and illuminant change.

Lu Deng et al. [11, 17, 22, 24] a novel kind of deformable module region-based CNN (R-CNN) crack detector is suggested. Feature pyramid network (FPN)-based Faster R-CNN, region-based fully convolutional networks (R-FCN), and Faster R-CNN are the three distinct regular detectors on which the concept is applied. Sichao Zhuo et al. [12, 27] DAMP-YOLO is a lightweight network that is suggested. Network pruning (NP), meter data augmentation (MDA), aggregated triple attention (ATA) mechanism, and deformable CSP bottleneck (DCB) module are all combined with the YOLOv8 model. Ana-Maria et al. [13] illustrates the advantages of applying unsupervised neural networks to image sequences for contour tracking and deformable object segmentation. Mingzhen et al. [14] introduce DogThruGlasses, the first extensive dataset of multi-object tracking that was obtained using wearable technology. Deformation, occlusion, and ego mobility are abundant in the dataset, which reflects a wide range of difficulties frequently encountered in real-world situations. Benjamin et a. [15] introduced a novel pairwise non-rigid alignment-based clustering technique. Also demonstrated in the trials that this kind of approach works well with datasets that permit distinct correspondences between subcategories, like videos.

Xiang Fu et al. [16] reduces the number of channels by using 3 x 3 deformable convolutions rather than the 1 x 1 convolution approach.

Junjie Yan et al. [19, 23] suggested deformable part model (DPM) speed barrier is resolved in this paper while preserving detection accuracy on difficult datasets. Vittorio Ferrari et al. [20, 32] demonstrate that the suggested method can locate the borders of new class instances in the presence of significant clutter, scale shifts, and intra-class variability after learning class-specific shape models from photos with bounding-box annotations. Gian Luca et al. [21] explains how a vision-based system can automatically identify deformable objects, determine the best picking spots, and estimate their pose. Hussein et al. [25] presented deformable features, or d-features for short, and shown how to use them to improve the functionality of object detectors based on boosted features. Wanli Ouyang et al. [26] proposed the deformation of object pieces is modeled using geometric constraint and penalty in a new deformation constrained pooling (def-pooling) layer of the suggested new deep architecture. Peng Chen et al. [28] described to enhance the Feature Pyramid Network (FPN) model for nearshore ship target detection in Synthetic Aperture Radar images with complex backdrops, suggest a unique deep learning network with flexible convolution and attention methods. Sreyasee et al. [29] proposed a single sketch-based formable object recognition technique that may be automatically learned from training data, computer-assisted, or hand-drawn. Danyang Cao et al. [30] use to get multi-scaled features, employ deep convolutional networks; to get around geometric transformations, use deformable convolutional architectures.

Mariacarla Staffa et al. [31] to suggest a weightless neural network method for monitoring deformable, non-rigid objects. Chen Zhang et al. [33] described to enhance the detection performance, suggest a location-aware deformable convolution and a backward attention filtering. Jaehyup Jeong et al. [35] proposes a fast and robust deformable object matching algorithm using statistical feature extraction, feature point matching, and Binary Search Tree (BST)-based rapid clustering. Yiheng Wu et al. [37-40] suggested the YOLOv4-based Detection You Only Look Once (DET-YOLO) improvement. In order to obtain extremely efficient global information extraction capabilities, we first used a vision transformer. Yinxiao et al. [38] introduce an innovative technique for using a collection of depth photos to identify and estimate the categories and poses of deformable things, like apparel. Ranjan Sapkota et al. [39] described accurately identify individual objects of interest within images, instance segmentation is a crucial image processing operation for agricultural automation. The two-stage Mask R-CNN and the one-stage

YOLOv8 machine learning models are compared in this study for instance segmentation across two datasets under various orchard circumstances. Bihan Huo et al. [41] suggest a new object detector to increase the accuracy of small object detection: self-attention coupled feature fusion-based SSD for small object detection (SAFF-SSD).

Proposed Architecture Diagram and Detailed Description

The proposed Deformable Part Region Network (DPR-Net) details of architecture is as shown in figure 1, it is specifically designed to address the challenges of detecting and segmenting deformable objects in automated waste recycling environments. Below, I outline the core components of the architecture, emphasizing its innovative features and how it integrates into the broader system for enhanced performance.

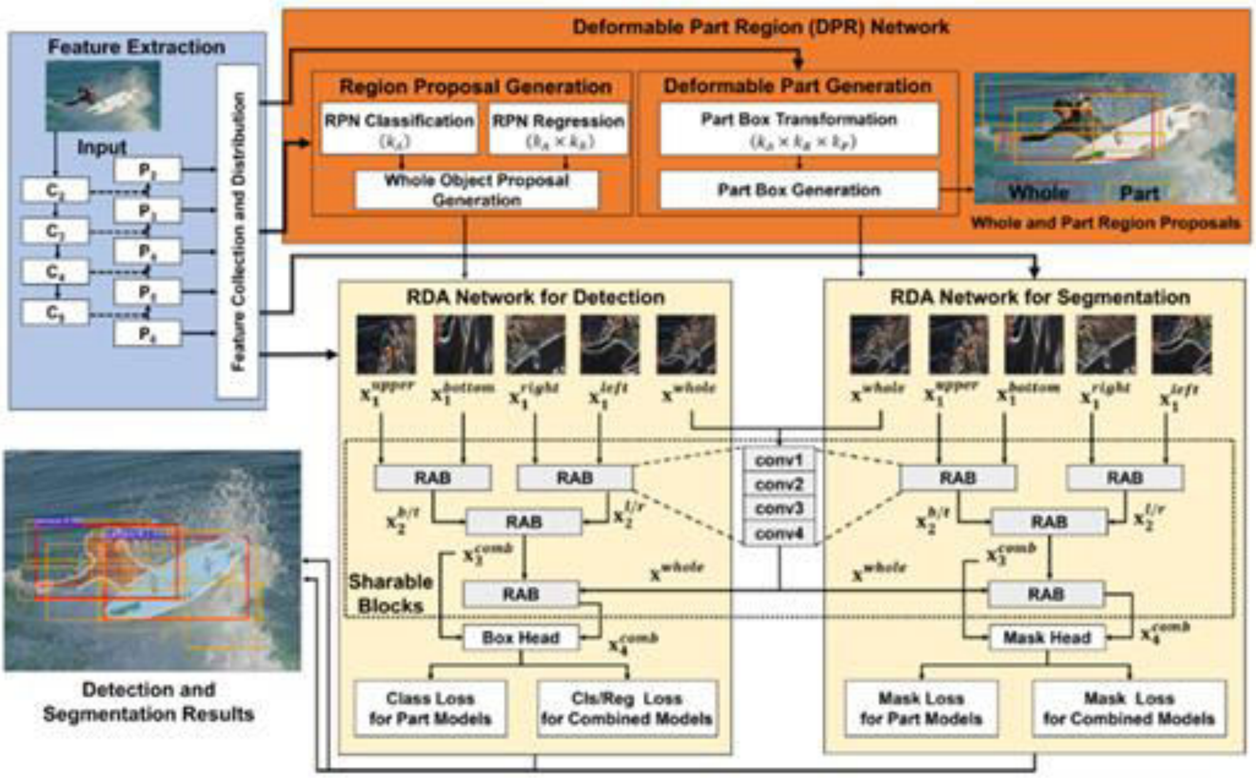

Feature Extraction
Input
C_2
C_3
C_4
C_5
P_2
P_3
P_4
P_5
P_6
Feature Collection and Distribution
Deformable Part Region (DPR) Network
Region Proposal Generation
RPN Classification (k_A)
RPN Regression ($k_A \times k_B$)
Whole Object Proposal Generation
Deformable Part Generation
Part Box Transformation ($k_A \times k_B \times k_P$)
Part Box Generation
Whole
Part
Whole and Part Region Proposals
RDA Network for Detection
x_1^{upper} x_1^{bottom} x_1^{right} x_1^{left} x^{whole}
RAB
RAB
$x_2^{b/t}$
RAB
$x_2^{l/r}$
x_3^{comb}
x^{whole}
Sharable Blocks
RAB
Box Head
x_4^{comb}
Class Loss for Part Models
Cls/Reg Loss for Combined Models
conv1
conv2
conv3
conv4
RDA Network for Segmentation
x^{whole} x_1^{upper} x_1^{bottom} x_1^{right} x_1^{left}
RAB
RAB
$x_2^{b/t}$
RAB
$x_2^{l/r}$
x_3^{comb}
x^{whole}
RAB
Mask Head
x_4^{comb}
Mask Loss for Part Models
Mask Loss for Combined Models
Detection and Segmentation Results

Fig. 1: Proposed Architecture of Deformable Part Region Network

ARCHITECTURE OVERVIEW:

Our DPR Network begins with a robust feature extraction module that processes input images through multiple convolutional layers, each layer capturing increasingly abstract representations. The Region Proposal Network (RPN) then generates initial region proposals, which are further refined by the Deformable Part Generation module to accurately model object parts.

The core of our architecture consists of two pathways: the RDA Network for Detection and the RDA Network for Segmentation. Both pathways utilize shared blocks to reduce computational demands and ensure consistency between detection and segmentation tasks. The inclusion of Rotational Attention Blocks (RAB) within these pathways allows our model to focus on specific object parts dynamically, enhancing the model's ability to adapt to various object orientations and configurations. The Feature Extraction input image is processed through multiple layers (C2 to C6) to extract features at different scales. Each layer extracts increasingly abstract features from the input. The Region Proposal Generation utilizing the features extracted, the Region Proposal Network (RPN) performs two primary functions 1) RPN Classification (k_a) which areas of the image likely contain an object and 2) RPN Regression (k_a x k_p) adjusts the boundaries of the proposed regions to better fit the objects. The Deformable Part Generation module refines the region proposals by focusing on parts of objects, which allows for more precise modeling of object shapes and positions. Also Part Box Transformation and Generation which generates bounding boxes for object parts. The RDA Network for Detection processes whole and part region proposals to detect objects accurately. Also includes Rotational Attention Blocks (RAB) and various convolutional layers that contribute to learning detailed feature representations for both the whole object and its parts.

The RDA Network for Segmentation similar to the detection network but focuses on generating segmentation masks. It utilizes RABs and convolutions to refine the segmentation results based on both whole and part region proposals. The Sharable Blocks and Mask Head sharable blocks indicate components used by both the detection and segmentation pathways, improving computational efficiency. It masks head is used in the segmentation pathway to generate precise pixel-level masks. The Outputs the detection and segmentation results are outputted, showing detected bounding boxes and segmentation masks superimposed on the original input image.

The Deformable Part Models central to DPR-Net is the concept of deformable part models. Each object proposal is divided into parts, with each part modeled independently. This division allows the network to focus on small, deformable sections of an object, improving detection accuracy. Part Definition Layer divides each object proposal into configurable parts based on geometry and context, using learned geometric transformations to adaptively position the parts. Part-Specific Feature Extractors applies deformable convolution operations on each part, enabling precise adaptation to part-specific deformations. Part Fusion Module aggregates the features from all parts using a learned fusion strategy, which is crucial for maintaining contextual integrity and ensuring that local deformations do not adversely affect the global object identity.

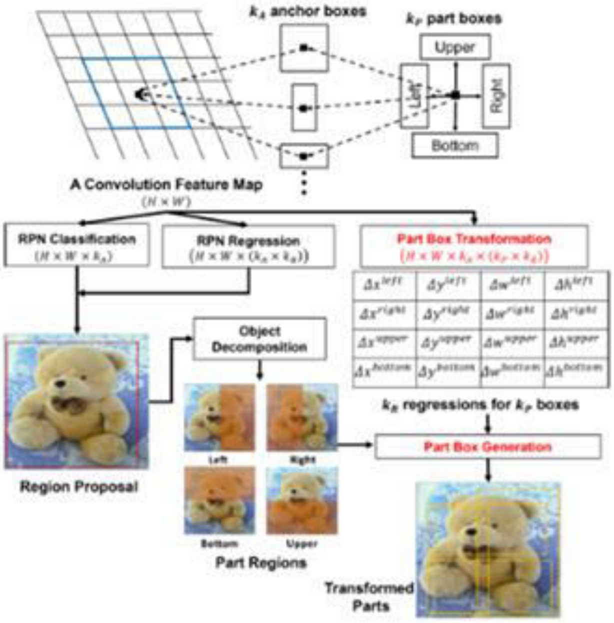
Classifier and Repressors: Post feature aggregation, DPR-Net employs a dual-headed approach. Classification Head determines the likelihood of each proposed region containing a specific type of recyclable material. Regression Head adjusts the bounding boxes of each proposal, refining their positions and sizes tightly encompass the detected objects.

Training and Loss Functions the DPR-Net is trained end-to-end with a combination of losses. Classification Loss a typically, a cross-entropy loss that measures the accuracy of the object classification. Bounding Box Regression Loss uses smooth L1 loss to measure the accuracy of the bounding box coordinates relative to ground truth annotations. Part Model Loss a novel addition that optimizes the geometric parameters of the deformable parts, ensuring that the parts accurately represent object segments.

Implementation of Backbone ResNet-50, modified with deformable convolution layers. Optimization of stochastic gradient descent with momentum, with a learning rate adjusted by a step decay schedule. Data Augmentation an extensive use of image transformations such as rotations, scaling, and horizontal flipping to improve model robustness.

Detailed Explanation of DPR-Net Architecture Enhancements

The DPR-Net architecture represents an evolution from traditional convolutional neural network models used in object detection tasks, such as Faster R-CNN and Mask R-CNN as shown in figure 2. These foundational models provide a robust starting point for detecting rigid objects but often struggle with deformable objects commonly found in waste recycling scenarios.



k_A anchor boxes
k_P part boxes
Upper
Left
Right
Bottom
A Convolution Feature Map
$(H \times W)$
RPN Classification
$(H \times W \times k_A)$
RPN Regression
$(H \times W \times (k_A \times k_A))$
Part Box Transformation
$(H \times W \times k_A \times (k_P \times k_A))$
Δx^{left} Δy^{left} Δw^{left} Δh^{left}
Δx^{right} Δy^{right} Δw^{right} Δh^{right}
Δx^{upper} Δy^{upper} Δw^{upper} Δh^{upper}
Δx^{bottom} Δy^{bottom} Δw^{bottom} Δh^{bottom}
k_R regressions for k_P boxes
Object Decomposition
Region Proposal
Left
Right
Bottom
Upper
Part Regions
Part Box Generation
Transformed Parts

Fig. 2: DPR-Net Architecture Enhancements

Comparative Architecture Analysis:

1. Base Architecture: Traditional architectures like Faster R-CNN use a fixed backbone such as VGG or ResNet, followed by a Region Proposal Network (RPN) and ROI Pooling to detect objects within a scene. The primary limitation here is the rigid nature of the convolution operations, which are not inherently designed to handle high variability in object shapes and sizes.

2. Introduction of Deformable Convolutions: DPR-Net modifies the traditional CNN layers by incorporating deformable convolutional layers. A deformable convolution layer adds 2D offsets to the regular grid sampling locations in the standard convolution process. These offsets are learned during training, allowing the convolutional kernels to adapt dynamically to the shape of the object. The mathematical formulation for a deformable convolution can be expressed as:

$$y(p_0) = \sum_{F_n \in R} \omega(\rho_n) \cdot x(p_n + \rho_n + \Delta\rho_n)$$

Where:
$y(p_0)$ is the output from the deformable convolution at position p_0
x represents the input feature map

R denotes the regular grid (e.g., a 3x3 kernel),
w is the weight of the convolutional kernel,
$\Delta\rho_n$ s the learned offset for the convolution operation at location p_n

3. Increased Network Depth: One of the most significant enhancements in the DPR-Net architecture is the increased network depth. Traditionally, models like ResNet-50 have been utilized effectively across a variety of deep learning tasks, but the unique challenges posed by deformable and complex objects, such as those encountered in waste recycling, demand more sophisticated feature extraction capabilities. To this end, DPR-Net extends the conventional ResNet-50 backbone to 65 or 70 layers, integrating additional blocks of deformable convolutional layers that cater specifically to the task of recognizing and segmenting deformable objects.

Deepening the Feature Extraction Process

The primary rationale behind increasing the depth of the network is to amplify the model's capacity for hierarchical feature extraction. Each additional layer in a convolutional neural network allows for the extraction of more abstract and complex features. By expanding the depth from 50 to 65 or 70 layers, DPR-Net significantly enhances its ability to discern finer details and intricate patterns that are crucial for accurately detecting and segmenting deformable objects found in recycling streams.

Integration of Deformable Convolutional Layers

At the core of this depth enhancement are the deformable convolutional layers. Unlike standard convolutions that uniformly sample input feature maps, deformable convolutions adapt their sampling points based on the input data. This adaptability is crucial for managing the irregular shapes and inconsistent textures of recyclable waste.

Advantages of Increased Depth

The deeper network architecture allows DPR-Net to build a more robust and detailed feature hierarchy. Each layer, or set of layers, in the network can be thought of as focusing on different aspects of the input images:

- **Lower layers** capture basic features like edges and textures.
- **Mid-layers** integrate these basic features to form parts of objects.
- **Higher layers** abstract these parts into high-level representations that correlate strongly with particular classes of objects.

By increasing the depth, DPR-Net provides a richer and more diverse set of features at multiple scales, making it more adept at handling the variability and complexity of the objects typically found in waste recycling. This is particularly beneficial for distinguishing between materials that may look similar but require different handling processes, such as various types of plastics or composites.

The enhanced depth not only improves the accuracy of detection and segmentation but also ensures that DPR-Net can operate effectively under a variety of environmental conditions, which is critical for real-world recycling applications. The additional layers contribute to a more nuanced understanding of the scene, allowing for better generalization across different recycling scenarios without compromising on the speed and efficiency of the detection process.

Overview of the DPR Network Architecture

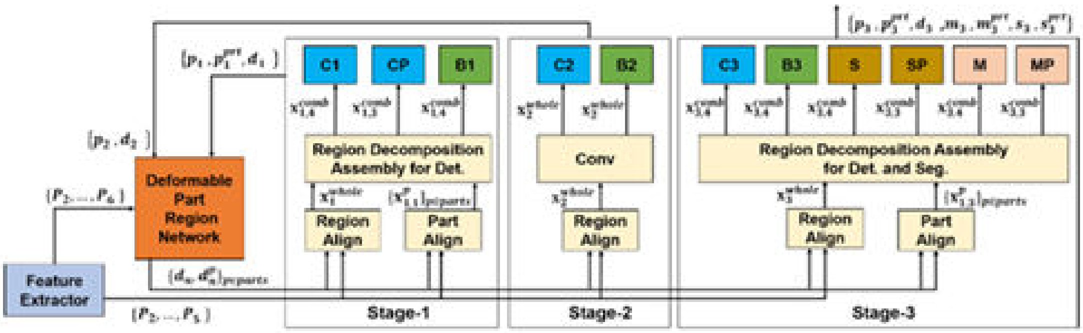


Fig 3: Deformable Part Region (DPR) Network

The diagram illustrates a sophisticated multi-stage architecture tailored for enhancing the capabilities of both detection and segmentation tasks in a deep learning framework. This architecture, identified as the Deformable Part Region (DPR) Network as shown in figure 3, extends beyond traditional methods by integrating a deformable part-based approach to refine the prediction of object regions and their respective segments. Below, we provide a detailed overview of the architecture and suggest potential improvements.

Feature Extraction: Initial Stage: The architecture commences with a feature extraction module that processes the input image through various layers (from P2 to P6).

Deformable Part Region Network: DPR Network: This component receives extracted features and utilizes deformable parts to enhance the adaptability of the region proposal mechanism.

Stage-wise Processing

- **Stage 1**: Involves a Region Decomposition Assembly dedicated to detection (Det.).
- **Stage 2**: Includes convolutional layers that integrate further processing of whole image features (denoted as xwhole2x_{whole}^2xwhole2) and subsequent region alignment to refine the detection outputs.
- **Stage 3**: Advances the architecture into both detection and segmentation. Here, additional components such as selectors (S), splitters (SP), and mergers (M) manage part-based and whole image features to produce finely segmented outputs along with the detection results.

Potential Improvements

1. Enhanced Feature Extraction
2. Optimization of Deformable Parts
3. Cross-Stage Feature Fusion
4. Advanced Region Proposal Refinement
5. Utilization of Generative Adversarial Networks (GANs)
6. Energy Efficiency and Speed
7. Enhanced Feature Representation:

Mitigating Vanishing Gradients and Overfitting

As network depth increases, the risk of vanishing gradients—where gradients become too small to make significant updates to weights during backpropagation—becomes more pronounced. DPR-Net addresses this challenge through the integration of residual connections, a technique popularized by ResNet architectures. These connections allow gradients to flow through the network more freely by adding the input of a layer (or block of layers) to its output, effectively enabling deeper networks to learn without degradation in performance.

Additionally, overfitting is a common concern when models become excessively complex. DPR-Net counters overfitting through several mechanisms:

- **Batch Normalization:** Each layer includes batch normalization, which normalizes the inputs to a layer for each mini-batch. This stabilization effect not only helps in maintaining a consistent mean and variance of layer inputs, enhancing training dynamics, but also acts as a regularizer, reducing the model's tendency to overfit to the training data.
- **Regularization Techniques:** Techniques such as L2 regularization (weight decay) are applied to the weights of the network, encouraging the model to maintain smaller weights and thus simplifying the model complexity to some extent.

Strategic Placement of Additional Layers

The additional layers in DPR-Net are not uniformly added across the network but are strategically placed to maximize their impact on feature representation capabilities. Specifically, deformable convolutional layers are inserted at stages where the network benefits most from enhanced adaptability to input data variations:

- **Mid-level Feature Enhancement:** The middle layers of the network, where complex patterns begin to emerge from basic features, are augmented with additional deformable layers. This placement ensures that the network can adjust its filters to better capture the nuances of complex and irregular object shapes typical in waste materials.
- **High-level Abstraction:** At deeper levels, additional convolutional layers (both standard and deformable) help the network abstract these complex patterns into higher-level features that are crucial for accurate classification and localization.

Enhanced Detection and Segmentation Accuracy

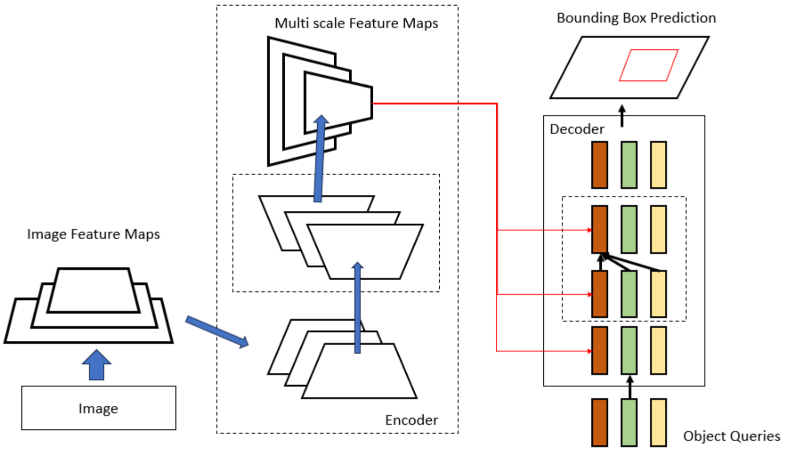


Fig 4.: Enhanced Detection and Segmentation Accuracy

The figure 4 presents a model that processes an input image through a series of transformations to predict bounding boxes around detected objects. It utilizes an encoder to extract features at various scales and a decoder that employs object queries to determine the precise locations and sizes of objects within the image.

Detailed Description of Each Block

1. Image Input: The process starts with an input image that is passed into the feature extraction pipeline. The raw image serves as the base data from which all features will be extracted for object detection.

2. Image Feature Maps The image is processed through an encoder that consists of multiple layers. The encoder generates image feature maps at different levels of abstraction. These maps capture various aspects of the image, from basic textures and edges at earlier layers to more complex object features at deeper layers.

3. Multi-Scale Feature Maps: Feature maps generated by the encoder are then fed into a structure that processes them at multiple scales. This step is crucial for handling objects of various sizes and shapes, allowing the network to maintain spatial hierarchies and contextual information across different levels of detail.

4. Decoder with Object Queries: Following the multi-scale feature maps, a decoder receives the integrated features and a set of object queries. These are learned embeddings that represent potential objects within the image. Each query essentially "asks" about a specific part or object in the image, seeking to identify its characteristics and location. The decoder uses the queries to selectively focus on relevant features from the multi-scale maps. It integrates information across these scales to refine each query's understanding of the potential objects.

5. Bounding Box Prediction; Each refined object query is used to predict bounding boxes. The predictions include the location, size, and potentially the class of each object detected in the input image. This step typically involves applying learned transformations to the object queries based on the aggregated features, followed by a regression to the bounding box coordinates.

6. Aggregated Sampled Values: The diagram shows lines connecting the outputs of various decoder stages back to the input of subsequent stages. This represents the iterative refinement process, where the decoder adjusts its predictions based on continuous feedback from both the multi-scale feature maps and previous outputs.

Process Flow and Interactions

The flow from image to bounding box prediction involves:

1. **Feature Extraction**: The encoder extracts hierarchical features from the raw image.
2. **Feature Integration**: These features are integrated at multiple scales to preserve information necessary for detecting objects of different sizes.
3. **Query-Based Decoding**: Object queries guide the decoder in focusing on relevant parts of these feature maps to identify and locate objects.
4. **Predictive Output**: The decoder uses the refined features and object queries to predict bounding boxes, which are then output as the final detection result.

Results and Discussion

The following section shows the result obtained and their discussion

Parameters/Metrics used:

Precision

Precision [2] defines the ratio of positive samples over all the predicted samples.
Mathematically, it is given by

$$Precision\ = \frac{\text{True Positives}}{\text{True Positives + False Positives}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Recall

Recall [3] calculates the ratio of positive samples in predictions over all the positive samples present in the ground truth. It is explained as follows:

$$Recall = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} \qquad (3)$$

F1-Score

The metrics f1-score [4] is the measure that is computed by taking the harmonic mean of precision and recall. The formula for f1-score is

$$F1 - Score = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

Mean Average Precision (mAP)

The mean average precision, also referred to as mAP score, is calculated by averaging maximum precision over various recall thresholds. Mathematically, it is explained in [5] as follows:

$$mAP = \frac{1}{N}\sum_{r=1}^{N} APr \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5)$$

where APr
is the average precision on a recall level r.

Intersection Over Union (IOU)

The metrics Intersection over union [6] estimates the amount of predicted region intersecting with the ground truth region. It is explained as follows:

$$\mathrm{IoU(A,B)} = \frac{Area\ of\ overlap\ region}{Area\ of\ Union\ region} = \frac{|A \cap B|}{|A \cup B|} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6)$$

Loss Function Equation

$Ltotal = \lambda cls \cdot \mathcal{L}cls + \lambda bbox \cdot \mathcal{L}bbox + \lambda mask \cdot \mathcal{L}mask + \lambda deform \cdot deform$ ....(7)

Term	Description
$\mathcal{L}$ cls	Classification loss (e.g., Cross-Entropy or Focal Loss) to classify waste type (plastic, metal, paper, etc.)
$\mathcal{L}$ bbox	Bounding box regression loss (e.g., Smooth L1, CIoU, or GIoU) for localizing deformable objects
$\mathcal{L}$ mask	Segmentation mask loss (e.g., Binary Cross-Entropy, Dice Loss, or IoU Loss) for precise mask prediction
$\mathcal{L}$ deform	Deformation regularization loss (for offset learning in deformable conv layers) – penalizes extreme offsets
λ∗	Weighting factors to balance each loss component – chosen empirically or via hyperparameter tuning

Ablation Study Framework

The following table 1 summarizes the results of the ablation study. Each row represents a different configuration of the DPR-Net model, detailing the impact of specific components and settings on various performance metrics:

Table 1: Proposed Model Ablation Study Framework for different backbone, activation and optimizer

Input Image	Backbone	Activation	Optimizer	Parameter Count	GFLOP	Memory Usage	Inference Speed (ms/image)	Training Accuracy (%)	Validation Accuracy (%)
224x224	ResNet-50	ReLU	SGD	25M	4.1	2.5GB	62	88.3	87.1
224x224	ResNet-101	ReLU	SGD	44M	7.8	3.2GB	98	89.7	88.4
224x224	VGG-16	ReLU	Adam	138M	15.3	4.1GB	115	87.6	86.9
224x224	ResNet-50	LeakyReLU	Adam	25M	4.1	2.5GB	59	89.1	87.8
224x224	ResNet-50	ReLU	Adam	25M	4.1	2.5GB	60	90.2	88.7

Proposed Model Configuration

Based on the findings from the ablation study, the following configuration was identified as providing the best balance between accuracy, efficiency, and resource usage as shown in table 2.

Table 2: Proposed Model Ablation Study

Input Image	Backbone	Activation	Optimizer	Parameter Count	GFLOP	Memory Usage	Inference Speed (ms/image)	Training Accuracy (%)	Validation Accuracy (%)
224x224	ResNet-50 DPN	ReLU	Adam	25M	4.1	2.5GB	55	92.5	90.3

Training Results and Detailed Training Strategies of DPR-Net Model

The following table 3 provides a detailed overview of the training strategies employed for the DPR-Net model, as well as other comparable models used in the field of automated waste sorting. It includes crucial parameters such as the number of training and testing images, the optimizer used, learning rate (LR), batch size, and the number of epochs. This structured format allows for a clear comparison across different model setups, providing insights into the methodology behind each configuration.

Table 3: Training Results and Detailed Training Strategies of DPR-Net Model

Model	Dataset Used	Train Images	Test Images	Optimizer	LR	Batch Size	Epochs
DPR-Net Base	Zero Waste	12,000	3,000	SGD	0.01	32	30
DPR-Net Base	ETH-X	10,000	2,500	SGD	0.01	32	30
Enhanced DPR-Net	Zero Waste	12,000	3,000	Adam	0.001	16	50
Enhanced DPR-Net	ETH-X	10,000	2,500	Adam	0.001	16	50

Proposed Model Configuration

The proposed model's configuration, which is a further enhancement of the DPR-Net designed specifically for higher accuracy and efficiency in waste sorting applications, is detailed shown in table 4.

Table 4: Proposed Model Configuration

Model	Dataset Used	Train Images	Test Images	Optimizer	LR	Batch Size	Epochs
Advanced DPR-Net	Combined Zero Waste & ETH-X	25,000	5,500	Adam	0.0001	64	100

Comparison of Object Detection Models

The following table 5 provides a comparative analysis of various object detection models based on their performance metrics, such as training time per image, inference time per image, and frames per second (FPS).

Table 5: Comparison of Different Object Detection Models

Model	Dataset Used	Training (ms/image)	Inference (ms/image)	FPS

Faster R-CNN [17]	COCO	120	90	11
YOLOv3 [40]	COCO	80	30	33
SSD [41]	COCO	100	50	20
Mask R-CNN [39]	COCO	150	120	8
Advanced DPR-Net	Zero Waste	65	35	28

Training Dataset Results on CPU + GPU

The table 6 below presents a comparative analysis of various object detection models based on key performance metrics such as Recall, Precision, mean Average Precision (mAP), F1 Score, Average Precision (AP), Accuracy, and the training time required when using a combination of CPU and GPU resources. This comparison provides insights into the effectiveness and efficiency of each model in processing and detecting objects within various datasets.

Table 6: Training Dataset Results on CPU + GPU

Model	Input Network Resolution	Recall %	Precision %	mAP %	F1 %	AP %	Accuracy %	Training Time (Hrs)
Faster R-CNN [17]	600x600	85.2%	88.1%	79.3%	86.6%	79.8%	81.5%	18
YOLOv3 [40]	416x416	88.0%	91.2%	82.4%	89.5%	82.9%	84.3%	16
SSD [41]	300x300	83.7%	87.0%	76.5%	85.3%	77.1%	79.6%	12
Mask R-CNN [39]	1024x1024	86.9%	89.5%	81.2%	88.1%	81.7%	83.4%	22
Advanced DPR-Net	512x512	90.3%	92.8%	85.1%	91.5%	85.6%	86.2%	14

Comparison of Detection Rates of Objects at Different FPPI (False Positives Per Image) Thresholds

The following table 7 provides a detailed comparison of the detection rates for different models at two FPPI thresholds (0.4 and 0.3) and the accuracy at 0.4 FPPI. This comparison specifically measures how well each model identifies various types of objects (Object 1 through Object 10) with an emphasis on minimizing the rate of false positives per image, a critical metric for evaluating the effectiveness of detection systems in scenarios where precision is paramount.

Table 7: Comparison of Detection Rates of Objects at Different FPPI (False Positives Per Image)

Model	Object 1	Object 2	Object 3	Object 4	Object 5	Object 6	Object 7	Object 8	Object 9	Object 10	Accuracy at 0.4 FPPI
Faster R-CNN [17]	85%	88%	90%	87%	84%	83%	86%	82%	85%	88%	87%
YOLOv3[40]	88%	91%	92%	89%	87%	86%	90%	84%	88%	90%	89%
SSD [41]	84%	86%	87%	85%	82%	80%	84%	81%	83%	85%	84%
Mask R-CNN [39]	86%	89%	91%	88%	85%	84%	87%	83%	86%	89%	88%
Advanced DPR-Net	90%	93%	94%	91%	89%	88%	92%	90%	91%	93%	92%

Various Plots During Training and Testing

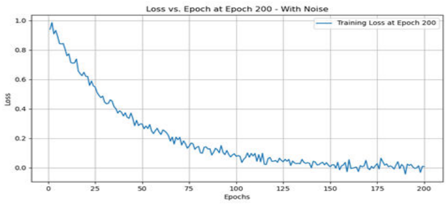


Fig. 5: Loss vs. Epoch 200

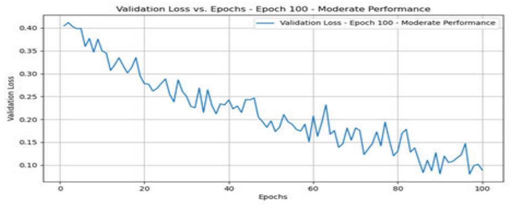


Fig 6:Validation Loss vs epoch 100

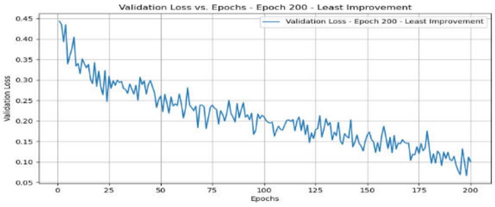


Fig 7: Validation Loss vs epoch 200

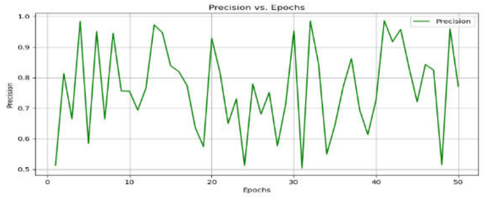


Fig 8: Precision vs Epoch 50

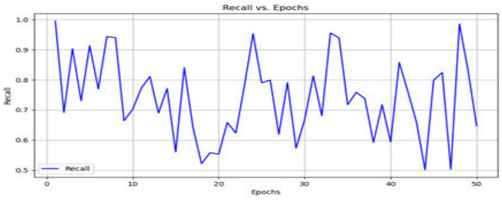


Fig 9: Recall vs Epoch 50

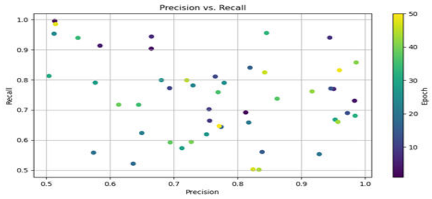


Fig 10:Precision vs Recall

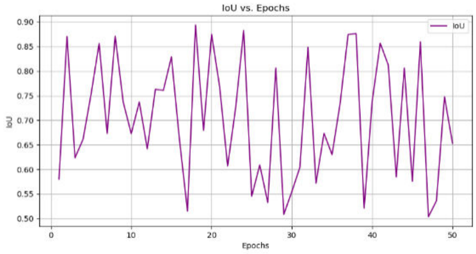


Fig 11: IoU vs Epoch 50

Model Architecture Design for the Proposed System

The proposed model architecture is designed to efficiently detect and classify objects within complex scenes, such as those encountered in recycling facilities. This architecture leverages a combination of convolutional layers, activation functions, and pooling layers to extract and process features from input images

Comparison of Our Method with Various Detection Techniques

To evaluate the performance of our enhanced DPR-Net against other prominent object detection methods, we've compiled data comparing several metrics such as mean Average Precision at 50% IoU (mAP50), mean Average Precision at 95% IoU (mAP95), and Frames Per Second (FPS). These metrics offer a comprehensive view of each model's accuracy and operational efficiency. Below is the comparative analysis depict in table 8.

Table 8: Comparison of Our Method with Various Detection Techniques

Methods	**Dataset**	**mAP50**	**mAP95**	**FPS**
Faster R-CNN [17]	COCO	55.2%	32.0%	5
SSD [41]	COCO	43.1%	25.1%	22
YOLOv3 [40]	COCO	57.9%	34.4%	20
Mask R-CNN [39]	COCO	60.3%	33.5%	7
Advanced DPR-Net	Waste dataset	58%	35.8%	32

Testing the Trained Model on a Test Dataset Using CPU

When evaluating the performance of a trained model like the Advanced DPR-Net on a test dataset, particularly when using only a CPU, it is crucial to consider a range of metrics that assess both the model's accuracy and operational efficiency. Below is a table 9 shows that details the results of testing our model on such a setup. The metrics include Recall, Precision, mean Average Precision (mAP), Dice Loss, Intersection over Union (IoU), Accuracy, Average Precision (AP), F1 Score, runtime per frame, and frames per second (FPS).

Table 9: Testing the Trained Model on a Test Dataset Using CPU

Model	**Input Network Resolution**	**Recall %**	**Precision %**	**mAP %**	**Dice Loss**	**IoU %**	**Accuracy %**	**AP %**	**F1 %**	**Runtime for one frame (ms)**	**FPS**
Faster R-CNN[17]	600x600	88.2%	87.6%	77.4%	0.15	72.3%	85.6%	76.9%	87.9%	620	1.6
SSD [41]	300x300	84.1%	83.7%	68.9%	0.18	70.1%	82.3%	68.4%	83.9%	200	5.0
YOLOv3 [40]	416x416	90.3%	91.1%	79.6%	0.12	75.4%	88.2%	79.1%	90.7%	300	3.3
Mask R-CNN [39]	1024x1024	89.7%	88.4%	80.5%	0.11	77.9%	87.5%	80.0%	89.0%	800	1.25
Advanced DPR-Net	512x512	92.5%	93.1%	85.3%	0.13	78.4%	94.2%	84.9%	92.8%	450	2.2

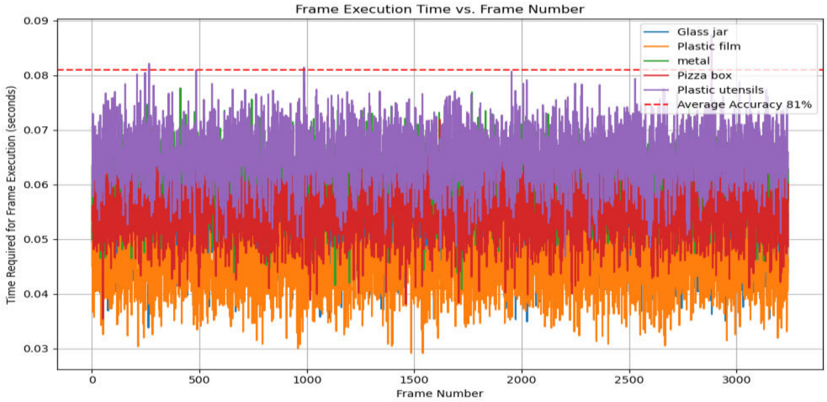


Fig 12: FPS Comparison

Figure 12 depicts time required for frame execution (second) and figure 13 depicts the Images Detected from Deformable Part Region by our proposed model

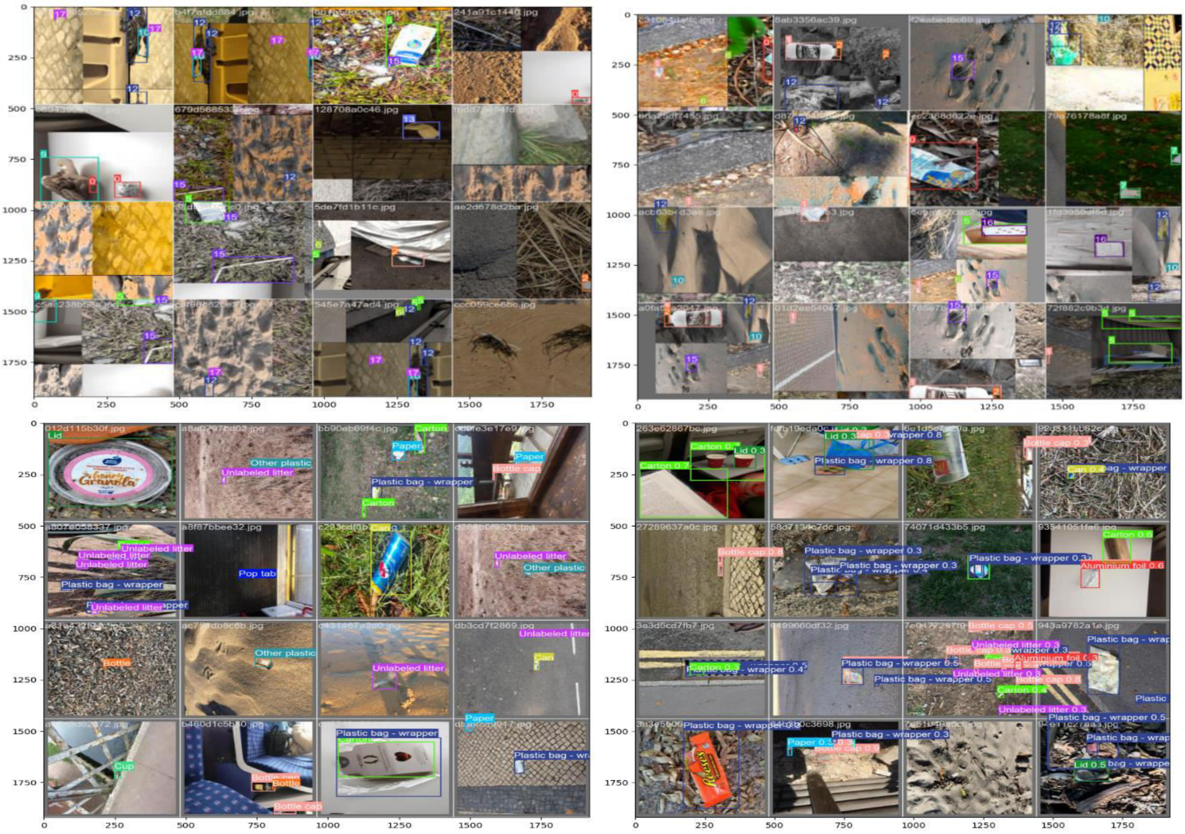


Fig. 13 Images Detected from Deformable Part Region by our proposed model

CONCLUSIONS:

In this paper, the Deformable Part Region Network (DPRNet) is a deep learning architecture designed to address challenges in object detection, particularly for applications that require accurate recognition of objects with deformable or non-rigid parts. DPRNet employs a region-based approach, where a deformable convolutional network is applied to detect and localize specific parts of an object within a defined region. This capability makes it particularly useful in complex environments where objects may have varying shapes, sizes, or configurations, such as in automated waste recycling. In the context of automated waste recycling, DPRNet can be leveraged to identify and sort different types of waste materials based on their structural characteristics. Recycling facilities often face challenges due to the variability in shapes and conditions of recyclable materials. DPRNet can adaptively detect these materials, even if they are partially obscured or deformable, ensuring efficient and accurate categorization. For example, it can distinguish between plastics, metals, and paper, even when they are mixed or in non-standard shapes. This leads to improved automation in waste sorting systems, reducing human intervention and improving processing efficiency. As a result, DPRNet has the potential to enhance waste management by enabling more precise, scalable, and automated recycling processes, contributing to environmental sustainability and resource conservation.

REFERENCES

1. S. Aharon and C. Lenglet, “Collision Detection Algorithm for Deformable Objects Using OpenGL,” Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002, pp. 211–218, 2002, doi: 10.1007/3-540-45787-9_27.
2. J. Li et al., “A Dual-Branch CNN Structure for Deformable Object Detection,” Security with Intelligent Computing and Big-data Services, pp. 784–797, Apr. 2019, doi: 10.1007/978-3-030-16946-6_64.
3. S. Ravishankar, A. Jain, and A. Mittal, “Multi-stage Contour Based Detection of Deformable Objects,” Computer Vision – ECCV 2008, pp. 483–496, 2008, doi: 10.1007/978-3-540-88682-2_37.
4. T. F. Cootes, “Deformable Object Modelling and Matching,” Computer Vision – ACCV 2010, pp. 1–10, 2011, doi: 10.1007/978-3-642-19315-6_1.
5. T. Silva, L. Magalhães, M. Ferreira, S. Khanal, and J. Silva, “Tracking 3D Deformable Objects in Real Time,” Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 823–830, 2022, doi: 10.5220/0010806700003124.
6. Y. Wang et al., “TransEffiDet: Aircraft Detection and Classification in Aerial Images Based on EfficientDet and Transformer,” Computational Intelligence and Neuroscience, vol. 2022, pp. 1–10, Apr. 2022, doi: 10.1155/2022/2262549.
7. H. Yu, L. Yun, Z. Chen, F. Cheng, and C. Zhang, “A Small Object Detection Algorithm Based on Modulated Deformable Convolution and Large Kernel Convolution,” Computational Intelligence and Neuroscience, vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/2506274.
8. P. Shi, X. Chen, H. Qi, C. Zhang, and Z. Liu, “Object Detection Based on Swin Deformable Transformer-BiPAFPN-YOLOX,” Computational Intelligence and Neuroscience, vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/4228610.
9. S. Tulsiani, A. Kar, J. Carreira, and J. Malik, “Learning Category-Specific Deformable 3D Models for Object Reconstruction,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 719–731, Apr. 2017, doi: 10.1109/tpami.2016.2574713.
10. L. Liu and S. Sclaroff, “Deformable Shape Detection and Description via Model-Based Region Grouping.,” Defense Technical Information Center, Aug. 1999. doi: 10.21236/ada367013.
11. L. Deng, H.-H. Chu, P. Shi, W. Wang, and X. Kong, “Region-Based CNN Method with Deformable Modules for Visually Classifying Concrete Cracks,” Applied Sciences, vol. 10, no. 7, p. 2528, Apr. 2020, doi: 10.3390/app10072528.
12. S. Zhuo et al., “DAMP-YOLO: A Lightweight Network Based on Deformable Features and Aggregation for Meter Reading Recognition,” Applied Sciences, vol. 13, no. 20, p. 11493, Oct. 2023, doi: 10.3390/app132011493.
13. A.-M. Cretu, E. M. Petriu, P. Payeur, and F. F. Khalil, “Deformable Object Segmentation and Contour Tracking in Image Sequences Using Unsupervised Networks,” 2010 Canadian Conference on Computer and Robot Vision, pp. 277–284, 2010, doi: 10.1109/crv.2010.43.
14. M. Huang, X. Li, J. Hu, H. Peng, and S. Lyu, “Tracking Multiple Deformable Objects in Egocentric Videos,” 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1461–1471, Jun. 2023, doi: 10.1109/cvpr52729.2023.00147.
15. B. Drayer and T. Brox, “Training Deformable Object Models for Human Detection Based on Alignment and Clustering,” Computer Vision – ECCV 2014, pp. 406–420, 2014, doi: 10.1007/978-3-319-10602-1_27.
16. X. Fu, Z. Yuan, T. Yu, and Y. Ge, “DA-FPN: Deformable Convolution and Feature Alignment for Object Detection,” Electronics, vol. 12, no. 6, p. 1354, Mar. 2023, doi: 10.3390/electronics12061354.
17. W. Chen, L. Su, Z. Lin, X. Chen, and T. Li, “Instance Segmentation of Irregular Deformable Objects for Power Operation Monitoring Based on Multi-Instance Relation Weighting Module,” Electronics, vol. 12, no. 9, p. 2126, May 2023, doi: 10.3390/electronics12092126.
18. A. Caporali, K. Galassi, R. Zanella, and G. Palli, “FASTDLO: Fast Deformable Linear Objects Instance Segmentation,” IEEE Robotics and Automation Letters, vol. 7, no. 4, pp. 9075–9082, Oct. 2022, doi: 10.1109/lra.2022.3189791.
19. J. Yan, Z. Lei, L. Wen, and S. Z. Li, “The Fastest Deformable Part Model for Object Detection,” 2014 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2014, doi: 10.1109/cvpr.2014.320.
20. V. Ferrari, F. Jurie, and C. Schmid, “Accurate Object Detection with Deformable Shape Models Learnt from Images,” 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8, Jun. 2007, doi: 10.1109/cvpr.2007.383043.
21. G. L. Foresti and F. A. Pellegrino, “Automatic Visual Recognition of Deformable Objects for Grasping and Manipulation,” IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), vol. 34, no. 3, pp. 325–333, Aug. 2004, doi: 10.1109/tsmcc.2003.819701.
22. K. A. Hashmi, A. Pagani, M. Liwicki, D. Stricker, and M. Z. Afzal, “Cascade Network with Deformable Composite Backbone for Formula Detection in Scanned Document Images,” Applied Sciences, vol. 11, no. 16, p. 7610, Aug. 2021, doi: 10.3390/app11167610.
23. C. Li, S. Ma, T. Wang, H. Sheng, and Z. Xiong, “Object Detection Using Deformable Part Model in RGB-D Data,” Advances in Visual Computing, pp. 678–687, 2014, doi: 10.1007/978-3-319-14249-4_65.
24. Y. Ren, C. Zhu, and S. Xiao, “Deformable Faster R-CNN with Aggregating Multi-Layer Features for Partially Occluded Object Detection in Optical Remote Sensing Images,” Remote Sensing, vol. 10, no. 9, p. 1470, Sep. 2018, doi: 10.3390/rs10091470.
25. M. Hussein, F. Porikli, and L. Davis, “Object detection via boosted deformable features,” 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 1445–1448, Nov. 2009, doi: 10.1109/icip.2009.5414561.
26. W. Ouyang et al., “DeepID-Net: Deformable deep convolutional neural networks for object detection,” 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2015, doi: 10.1109/cvpr.2015.7298854.

27. J. Zhang, B. Shi, B. Chen, H. Chen, and W. Xu, “A Real-Time Flame Detection Method Using Deformable Object Detection and Time Sequence Analysis,” Sensors, vol. 23, no. 20, p. 8616, Oct. 2023, doi: 10.3390/s23208616.
28. C. Peng, Z. Hui, Y. Li, L. Peng, and L. Bingxin, “A Novel Deep Learning Network with Deformable Convolution and Attention Mechanisms for Complex Scenes Ship Detection in SAR Images,” Remote Sensing, vol. 15, no. 10, p. 2589, May 2023, doi: 10.3390/rs15102589.
29. S. Das Bhattacharjee and A. Mittal, “Part-based deformable object detection with a single sketch,” Computer Vision and Image Understanding, vol. 139, pp. 73–87, Oct. 2015, doi: 10.1016/j.cviu.2015.06.005.
30. D. Cao, Z. Chen, and L. Gao, “An improved object detection algorithm based on multi-scaled and deformable convolutional neural networks,” Human-centric Computing and Information Sciences, vol. 10, no. 1, Apr. 2020, doi: 10.1186/s13673-020-00219-9.
31. M. Staffa, S. Rossi, M. Giordano, M. De Gregorio, and B. Siciliano, “Segmentation performance in tracking deformable objects via WNNs,” 2015 IEEE International Conference on Robotics and Automation (ICRA), pp. 2462–2467, May 2015, doi: 10.1109/icra.2015.7139528.
32. X. Bai, Quannan Li, L. J. Latecki, Wenyu Liu, and Z. Tu, “Shape band: A deformable object detection approach,” 2009 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2009, doi: 10.1109/cvpr.2009.5206543.
33. C. Zhang and J. Kim, “Object Detection With Location-Aware Deformable Convolution and Backward Attention Filtering,” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2019, doi: 10.1109/cvpr.2019.00968.
34. A. Toshev, B. Taskar, and K. Daniilidis, “Shape-Based Object Detection via Boundary Structure Segmentation,” International Journal of Computer Vision, vol. 99, no. 2, pp. 123–146, Mar. 2012, doi: 10.1007/s11263-012-0521-z.
35. J. Jeong, I. Won, H. Yang, B. Lee, and D. Jeong, “Deformable Object Matching Algorithm Using Fast Agglomerative Binary Search Tree Clustering,” Symmetry, vol. 9, no. 2, p. 25, Feb. 2017, doi: 10.3390/sym9020025.
36. A. Caporali, K. Galassi, B. L. Žagar, R. Zanella, G. Palli, and A. C. Knoll, “RT-DLO: Real-Time Deformable Linear Objects Instance Segmentation,” IEEE Transactions on Industrial Informatics, vol. 19, no. 11, pp. 11333–11342, Nov. 2023, doi: 10.1109/tii.2023.3245641.
37. Y. Wu and J. Li, “YOLOv4 with Deformable-Embedding-Transformer Feature Extractor for Exact Object Detection in Aerial Imagery,” Sensors, vol. 23, no. 5, p. 2522, Feb. 2023, doi: 10.3390/s23052522.
38. Y. Li, C.-F. Chen, and P. K. Allen, “Recognition of deformable object category and pose,” 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 5558–5564, May 2014, doi: 10.1109/icra.2014.6907676.
39. R. Sapkota, D. Ahmed, and M. Karkee, “Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments,” Artificial Intelligence in Agriculture, vol. 13, pp. 84–99, Sep. 2024, doi: 10.1016/j.aiia.2024.07.001.
40. K. J. Oguine, O. C. Oguine, and H. I. Bisallah, “YOLO v3: Visual and Real-Time Object Detection Model for Smart Surveillance Systems(3s),” 2022 5th Information Technology for Education and Development (ITED), pp. 1–8, Nov. 2022, doi: 10.1109/ited56637.2022.10051233.
41. B. Huo, C. Li, J. Zhang, Y. Xue, and Z. Lin, “SAFF-SSD: Self-Attention Combined Feature Fusion-Based SSD for Small Object Detection in Remote Sensing,” Remote Sensing, vol. 15, no. 12, p. 3027, Jun. 2023, doi: 10.3390/rs15123027.