

**DEEP LEARNING PROJECT**

**CSE 4006**

**LITERATURE REVIEW**

**Project Title: INVISIBLE MAN USING MASK- RCNN**

**Team:  
1.Kashish(22BCE8559)  
2.Vanama Dhhanush Krishna Kumor(22BCE7763)  
3.Parepalli Kesava Naga Sai Durga Vardhan(22BCE20530)**

**Paper 1:**

Title: “[Deep learning based object detection **using mask RCNN**](https://ieeexplore.ieee.org/abstract/document/9489152/)”Year: 2021

1. Proposes a deep learning solution for automatic object detection using Mask R-CNN, addressing viewpoint changes and occlusion issues.
2. Object detection aims to identify all instances of known object classes (e.g., people, vehicles, faces) in an image, using bounding boxes, class labels, and masks to mark them.
3. Implements pixel-level instance segmentation with bounding boxes, class labels, and masks.
4. Explores several RCNN approaches for object detection.
5. In order to achieve even greater performance in object detection tasks, the paper discusses the possibility of making further advancements to strengthen the Mask R-CNN model's robustness and reliability.

Karunakaran, V. (2021, July). Deep learning based object detection using mask RCNN. In *2021 6th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1684-1690). IEEE.

**Paper 2:**

Title: “[A **Mask**-**RCNN**based object detection and captioning framework for industrial videos](https://search.proquest.com/openview/4e6107516f5ac61c2a0fba89732258e3/1?pq-origsite=gscholar&cbl=2037694)”Year: 2022

1. Automate surveillance video analysis, especially for industrial videos, to improve productivity and reduce human effort.
2. Utilizes uniform sampling to extract frames, turning video captioning into image captioning. Mask R-CNN is used for object detection.
3. Creation of template sentences and template-based video frame caption generation utilizing the detected objects.
4. Achieved 95.62% accuracy with an average confidence score of 0.8975 for object detection.
5. Useful for reporting, remote monitoring, and work analysis in labor-intensive industries.

Namjoshi, M., & Khurana, K. (2021). A Mask-RCNN based object detection and captioning framework for industrial videos. *International Journal of Advanced Technology and Engineering Exploration*, *8*(84), 1466.

**Paper 3:**

Title: “ [Detection and segmentation of mature green tomatoes based on **mask R**-**CNN with**automatic image acquisition approach](https://www.mdpi.com/1424-8220/21/23/7842)”Year: 2023

1. The paper uses the Mask R-CNN algorithm for the accurate detection and segmentation of mature green tomatoes in a greenhouse environment.
2. The CNNs based fruit detection method can roughly calculate the position of the fruits using the bounding box, while cannot accurately extract contour and shape information.
3. ResNet50-FPN is chosen for feature extraction during the Mask R-CNN training process.
4. By pooling feature maps into a fixed size based on the preselection box, the ROI (Region of Interest) is computed using bilinear interpolation through ROIAlign, allowing for precise segmentation.
5. When Intersection over Union (IoU) is set to 0.5, the best performance is achieved, with F1-Scores for both bounding boxes and mask regions reaching 92.0%.

Zu, L., Zhao, Y., Liu, J., Su, F., Zhang, Y., & Liu, P. (2021). Detection and segmentation of mature green tomatoes based on mask R-CNN with automatic image acquisition approach. *Sensors*, *21*(23), 7842.

**Paper 4:**

1. Title: “[Extract and Merge: Merging extracted humans from different images utilizing **Mask R**-**CNN**](https://arxiv.org/abs/1908.00398)”Year: 2021
2. The application proposed in the paper uses mask segmentation and object detection with Mask R-CNN to extract human instances from pictures or videos.
3. The application runs at 5 frames per second, it does not add significant computational overhead to Mask R-CNN, making it effective for real-time usage.
4. It can handle human extraction from multiple images and videos, merging them seamlessly into a single scene or new background.
5. The code is designed to handle videos of varying lengths, where the output video’s length is set equal to the longest input video.
6. The application automates the most time-consuming aspects of photo editing, such as object selection and background merging, allowing editors to focus on creative design work.

Minkesh, A., Worranitta, K., & Taizo, M. (2019). Extract and Merge: Merging extracted humans from different images utilizing Mask R-CNN. *arXiv preprint arXiv:1908.00398*.

**Paper 5:**

Title: “[Semantic and instance segmentation of room features in floor plans **using Mask R**-**CNN**](https://www.diva-portal.org/smash/record.jsf?pid=diva2:1352780)”

1. Year: 2022
2. The paper evaluates the performance of machine learning techniques, particularly Mask R-CNN, for processing and segmenting bitmap images of floor plans, a task that traditionally required manual human intervention.
3. Mask R-CNN model for instance segmentation, is used to detect objects and generate high-quality segmentation maps within floor plans, providing complete segmentation with a single network.
4. The paper uses a limited size and challenging dataset of floor plans, making the task more complex due to noise and variations in blueprint designs.
5. Mask R-CNN performs well on smaller blueprints with minimal noise, yielding almost completely segmented floor plans.
6. For large blueprints with more noise, the model produces decent but not completely segmented floor plans, indicating room for improvement on noisier datasets.

Sandelin, F. (2019). Semantic and instance segmentation of room features in floor plans using Mask R-CNN.

**Paper 6:**

Title: “[Double **Mask R**‐**CNN**for Pedestrian Detection in a Crowd](https://onlinelibrary.wiley.com/doi/abs/10.1155/2022/4012252)”Year: 2023

1. The paper introduces the Double Mask R-CNN, an improved version of Mask R-CNN with Feature Pyramid Network (FPN).
2. Aims to address the challenges of feature extraction and limitations of Nonmaximum Suppression (NMS) in crowded pedestrian detection.
3. To improve edge feature extraction of crowded pedestrians, a Semantic Feature Pyramid Network (SFPN) is added to the network. This improves the model's ability to detect pedestrians in crowded scenes.
4. Unlike other methods that apply masks to all detected images (increasing detection time), this approach significantly reduces detection time by only masking and reprocessing images with low pedestrian visibility.
5. The experimental results demonstrate that Double Mask R-CNN achieves a 13.12% lower log-average miss rate (MR) than other mainstream networks on the CrowdHuman dataset

Liu, C., Wang, H., & Liu, C. (2022). Double Mask R‐CNN for Pedestrian Detection in a Crowd. *Mobile Information Systems*, *2022*(1), 4012252.

**Comparison Table:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Paper** | **Title of the Paper** | **Methodology** | **Datasets used** | **Performance Metrics** | **Advantages** | **Disadvantages** |
| **1** | **Deep Learning based Object Detection using Mask RCNN** | **Mask R-CNN with visual tracking algorithms; ResNet-101 + FPN** | **Plasmodium vivax dataset** | **The Mask R-CNN model obtained 94% mAP** | **High accuracy in dynamic scenes** | **High computational cost, limited real-time performance** |
| **2** | **A Mask-RCNN based object detection and captioning framework for industrial videos** | **Real-time Mask R-CNN segmentation in AR using TensorRT** | **Custom-created dataset comprises 216 training images with 13 different class selections** | **95.62% accuracy with an average confidence score of 0.8975** | **Real-time performance, multi-person segmentation** | **Poor generalization in cluttered environments** |
| **3** | **Detection and Segmentation of Mature Green Tomatoes Based on Mask R-CNN with Automatic Image Acquisition Approach** | **Integrated spatial-temporal context, hybrid loss function** | **Custom dataset with the help of a robot** | **F1-Score of bounding box and mask region all achieve 92.0%** | **Improved occlusion handling** | **High false positives on non-pedestrian datasets** |
| **4** | **Extract and Merge: Merging extracted humans from different images utilizing Mask R-CNN** | **Mask R-CNN for background subtraction; motion compensation module** | **MS COCO dataset** | **Not Mentioned** | **Effective in dynamic and static backgrounds** | **Struggles with similar textures to the background** |
| **5** | **Semantic and Instance Segmentation of Room Features in Floor Plans using Mask R-CNN** | **Mask R-CNN combined with deep metric learning for person re-identification** | **CVC-FP [11],** **R-FP [37],** **Rent3D [31],** **CubiCasa5k [22]** | **Models AP50 is 22.6 and Prec.50 is 26.4 with pixel accuracy of 76%** | **High precision in crowded areas, effective across camera views** | **Sensitive to changes in appearance** |
| **6** | **Double Mask R-CNN for Pedestrian Detection in a Crowd** | **Mask R-CNN for blurring/masking humans in video; fine-tuned for indoor environments** | **CrowdHuman** | **Doble mask R\_CNN AP50 is 86.80 and** | **Preserves privacy in controlled indoor environments** | **L** **imited effectiveness outdoors, additional post-processing needed** |

**References:**

1. Karunakaran, V. (2021, July). Deep learning based object detection using mask RCNN. In *2021 6th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1684-1690). IEEE.
2. Namjoshi, M., & Khurana, K. (2021). A Mask-RCNN based object detection and captioning framework for industrial videos. *International Journal of Advanced Technology and Engineering Exploration*, *8*(84), 1466.
3. Zu, L., Zhao, Y., Liu, J., Su, F., Zhang, Y., & Liu, P. (2021). Detection and segmentation of mature green tomatoes based on mask R-CNN with automatic image acquisition approach. *Sensors*, *21*(23), 7842.
4. Minkesh, A., Worranitta, K., & Taizo, M. (2019). Extract and Merge: Merging extracted humans from different images utilizing Mask R-CNN. *arXiv preprint arXiv:1908.00398*.
5. Sandelin, F. (2019). Semantic and instance segmentation of room features in floor plans using Mask R-CNN.
6. Liu, C., Wang, H., & Liu, C. (2022). Double Mask R‐CNN for Pedestrian Detection in a Crowd. *Mobile Information Systems*, *2022*(1), 4012252.