Data Augmentation Technique to Expand Road Dataset Using Mask RCNN and Image Inpainting

Plabon Kumar Saha[1], Sinthia Ahmed[2], Tajbiul Ahmed[3], Hasidul Islam[4], Al Imran[5],
A. Z. M. Tahmidul Kabir[6], Al Mamun Mizan[7]
Department of Computer Science and Engineering[1-5] ,Department of Electrical and Electronic Engineering[6-7]
American International University-Bangladesh
Dhaka-1229, Bangladesh
pkumarsaha71@gmail.com[1], ahmed.sinthia97@gmail.com[2], ahmedtajbiul@gmail.com[3], ihasidul@gmail.com[4], asq24i@gmail.com[5], tahmidulkabir@gmail.com[6], almamuneee15@gmail.com[7]

***Abstract*—A popular method for training a machine learning model is to use a data-driven approach. This research contributes to expanding the dataset of urban road images without vehicles. Using this method will benefit a variety of existing and new research projects that require an empty road to train their data-driven model. To achieve the desired result, the method combines image segmentation and image inpainting. The model detects the vehicle with Mask RCNN and removes the detected object with image inpainting. Morphological transformation was used to improve the method's efficiency. Using dilation operation of morphological transformation, the mask generated from mask RCNN is enlarged. The results of the experiment support the method's efficacy.**

Keywords—Mask RCNN, Image inpainting, Data augmentation, Image processing, Object detection.

I. INTRODUCTION

Artificial Intelligence breakthroughs in machine learning and deep learning are causing a paradigm shift in nearly every sector of the technology industry. From autonomous vehicles to virtual environments and urban city planning, AI is taking its place through different research domains. Many approaches are used in various research domains, one of which is the data-driven approach [1]. Although the Big Data revolution has made it easier to obtain real-world data from the Internet, creating a large dataset remains a time-consuming and laborious task [2]. Many research domains, such as Internet of Things (IoT), Image processing [4, 5], machine learning, etc rely on data for practical implementations [3]. Augmented autonomous driving simulation was proposed in the study[1] who has used a data-driven approach, with the data being real-world road images collected from the ApolloScape, Kitti, and CityScape datasets to create photorealistic simulation images and renderings. Extracted road networks are required for urban mapping and image-based vehicle navigation. However, due to a lack of prior road datasets, methods for extracting road networks from images face a number of challenges [6]. True virtual cities necessitate realistic 3D models of urban built stature. A sufficient amount of real city data is required for realistic 3D models in virtual environments [7]. Moving obstacles are not ideal for creating 3D models, and there is insufficient data that can be found without obstacles such as vehicles. It is difficult to detect road signs such as lanes, crosswalks, stop lines, and so on when there are vehicles on the road, so vehicles tend to be a constraint to road and lane detection research [8]. Quality training data is the most important aspect of machine learning. Training data comes in a variety of formats, reflecting the numerous potential applications of machine learning algorithms. The accuracy and performance of a machine learning model are determined by the quality and quantity of training data. Image augmentation is considered when creating a diverse image dataset. Image augmentation is the artificial generation of training images through the use of various processing methods that are typically required to improve the performance of AI models [9].

The proposed methodology in this paper combines image segmentation and image inpainting to enhance existing road images and create a road dataset that is free of obstacles (vehicles). Vehicles are segmented and masks are created using the Mask RCNN method. Mask-RCNN produced cutting-edge results for object detection and instance segmentation on the MSCOCO [10] dataset [11]. MSCOCO's labeled data explicitly identifies features. In this study, vehicles are identified from the images, and that pattern trains the algorithm to recognize the same pattern in unlabeled data. Mask RCNN generates masks of the segmented vehicles, which are then removed from the images using the image inpainting method. Inpainting has a wide range of applications, from the restoration of damaged paintings and photographs to the removal/replacement of specific objects [12]. To remove the vehicle from the road images, the inpainting method detects adjacent pixels in the masked area and fills them in. Finally, the similarity ratio between the method's input image and the output image is computed.

II. BACKGROUND STUDY

In surveillance camera systems, the paper [13] concentrated on recognizing unattended items. Two timely updated backgrounds are obtained via foreground blob extraction. The proposed approach recognizes a removed or forgotten object by a human from the previous frame based on the method's succession. Despite the method's high accuracy in object detection, it only worked with persons who were standing motionless. In this paper [14] the authors have described a strategy for eliminating natural items. A target detection technique based on contour transformation was proposed in this study. In this paper [15], the authors have described several methods for carrying out the picture inpainting process. Here, the many techniques to completing the inpainting process were identified and their shortcomings were compared using parameters such as performance, output, and efficiency. Although the paper shows how inpatient methods are used in various projects, it lacks the data or actual visualization of output comparisons needed to

comprehend the picture inpainting techniques presented fully. This study [16] proposed a technique for deleting undesired targets or elements from an image input. They employed the viola jones object detection approach for facial identification, which let them identify the primary topic of the frame, and the suggested model uses SVM to crop the image only, keeping the main subject. If the uninvited object is in the corner of the image, this strategy is quite effective. When employing the crop approach to remove an item from the middle of an image, the model was not particularly effective. In this paper [17], the researchers proposed an algorithm for eliminating rain from still pictures. They employed edge detection to extract a mask that captures the nuances of the rain removal image after decomposing the input image. They then processed the rain using a defogging algorithm. To specify a robot's target component capture, this study [18] presented an improved version of the mask RCNN method. To expand RCNN subnet and ROI wrapping, they replaced Mask RCNN with a Light –Head Mask RCNN network. This helps the system to reduce complexity and achieve slightly higher detection results than Mask RCNN, but it does so at the expense of the Mask RCNN's real-time detection speed and efficiency. In this paper [19], the authors have created a technique for reading meter values from an image automatically. Mask RCNN has been tweaked to read a wide range of digital values from an image. A mask RCNN implementation was investigated in this paper [20]. The researcher utilized Mask RCNN to accomplish nucleus segmentation and object localization. They used ResNet-101-FPN to modify the RCNN mask on the BBBC038v1 picture, which contains tiny nuclei images. Another mask RCNN implementation was carried out by the authors of the following paper [21]. They have discovered how to automatically identify dents, convex, hole, damage, and distortion in containers using Fmask-RCNN. They used Res2Net101 structure, flip fuzzy data, fusion unsampling, and other techniques to modify the mask of detected container damages, allowing them to successfully identify various types of damages in containers using just a single image frame. In this study [22] the researchers have proposed a model that employs Mask RCNN and stereo vision to recognize numerous things from varying distances, such as bananas, apples, and oranges. The approach might name the thing that is being identified in addition to identifying it. The COCO17 dataset was used to create the image set. To construct an inpainting method, the authors [23] proposed a deep neural network model in their paper. In their research, Inpainting has been divided into two categories: structure reconstruction and texture production. The first component fills in the gaps in the framework, while the second part creates the texture. In this paper [24] the authors have recommended a method for inpainting objects from a single video frame in their research. The video is divided into a number of frames using this way. The algorithm then employs graph-based region segmentation following foreground extraction. Finally, they completed the inpainting method and converted the frames into videos using their proposed method. In this study [25], the authors have proposed a method for eliminating snow or rain from a single still picture in their paper. They extracted the rain and snow portion of the image, as well as other details, using image decompositioning dictionary learning. In this paper [26], the authors have shown the importance of data augmentation in a deep learning image classifier model in their work. They demonstrated the absence of training data for picture classification as well as a comparison of existing data augmentation strategies. Later, they presented their own data augmentation strategy, as well as the usage of the created output to train a deep learning model. In this following paper [27], the researchers have improved a small data collection for model training by combining it with a larger data set at their work. They've successfully created minuscule object masks and used data augmentation to create photorealistic images of blood cells. In this study [28], the authors used data augmentation to train an intelligent disease detection algorithm in their research. They've identified the problems with open-source data sets and proposed a strategy for developing their own data collection for model training. Crop, rotation, image scaling, zoom, channel shift, and other minor augmentation techniques were applied. In this paper [29] the authors have discussed the limitations of medical-related intelligence due to a lack of data in their paper. As a result, they offered image synthesis as a way to expand the data set. They created a binary mask using two-stage generative adversarial networks (GAN) and then used GAN for conditional production of the synthesized picture in a subsequent step. They later used these photos to create additional training datasets. In this study [30], researchers have proposed an image inpainting library in their research. They've used their image augmentation method on a variety of photos, including map images, animal photographs, landscape images, and other regularly used images. This method is based on a deep neural networking model that has been strengthened by image augmentation. In a malware scenario, they used picture augmentation to discover malware families. Using image inpainting, this research [32] suggested a learnable bidirectional attention mappings (LBAM) method. Their proposed LBAM method is effective in adapting to uneven holes and propagation of convolution layers since the introduction of learnable attention maps. Reverse attention maps are shown to allow the U-Net decoder to focus solely on filling in gaps. In this paper [33] the authors have presented a coarse-to-fine generative picture inpainting framework, as well as our baseline and full models with a novel contextual attention module. By learning feature representations for explicitly matching and attending to appropriate backdrop patches, the contextual attention module greatly enhanced image inpainting outcomes. Based on the current state-of-the-art generative image inpainting network, they have offered numerous strategies to increase training stability and speed, including inpainting network upgrades, global and local WGANs, and spatially discounted reconstruction loss. As a result, they were able to train the network in a week instead of two months. An adaptive Mask RCNN technique for recognizing multi-class objects in remote sensing images was proposed in this study [34]. To overcome the lack of labeled remote sensing imagery, they used transfer learning, fine-tuning, and augmentation techniques such as rotation, scaling, and illumination conditions. In terms of average precision, computation time, Intersection over Union (IOU), and Precision-Recall Curves, the study compares the proposed method to baseline deep object recognition algorithms (PRC). The method used in the paper [35] is Tozero thresholding, which is applied to the image and converts the pixels at the thresholding value to black and the others to lighter shades. The image undergoes morphological transformation in order to eliminate noise and identify specific parts. The morphological closing technique utilized here will dilate the image first, removing the object image's

noise before closing the discontinuous sections. After that, depending on the kernel used, erosion reduces the white noise by removing pixels near the boundary. Another study [36] based on morphological reconstruction proposed an improved image segmentation algorithm. Catchment basins for all sizes of objects were morphologically defined and re-shaped using erosion and dilation operations, ensuring watershed transformation in the final stage segmentation of the image appropriately. To achieve an acceptable effect, the approach proposed in that paper uses different thresholds to search for the catchment basin marker in distinct grey areas and overcomes the problem of un-matching different-sized froths at a single threshold.

III. METHODOLOGY

To generate the output image the project strictly follows the steps mentioned in Figure 1. The operation begins with getting the input image and applying the mask RCNN to it, which then returns the masks of the detected vehicles in the image. The generated masks are then dilated, applying which the original input image is inpainted. Hence, generating a new image without any vehicles.

Fig. 1. Flow chart of image inpainting.

A. *Mask RCNN:*

MASK R-CNN (Region-Based Convolutional Neural Network) is a framework for object instance segmentation that utilizes Feature Pyramid Network (FPN) and a ResNet101 backbone. Mask R-CNN extends Faster R-CNN by adding the implementation of instance segmentation and replacing ROIPooling from Faster R-CNN to ROIAlign. ROIAlign is capable of representing fractions of a pixel, thus producing a much more definitive mask than its predecessor. Mask RCNN performs a dual-stage operation on its input images where in the first stage, RPN, a light neural network is used to propose regions that might be comprised of objects. A set of boxes with preordained coordinates and scales relative to that of the original input image called Anchors are utilized to bind bounding boxes to the locations of interest. Based on an IoU(Intersection over Union) value, each anchor is assigned to a bounding box that then uses the anchors to estimate the object's size and adjust itself accordingly. In the second stage, the proposed regions from the first stage are used to generate bounding boxes with a similar process to that of the first stage but here anchors are replaced with ROIAlign which then constructs masks for each detected object at the pixel level. Figure 2a shows an input image in which Mask RCNN detects multiple objects and assigns each of them a bounding box. After that, a mask is generated off of the instance segmentation of the vehicle as shown in Figure 3.

a b

Fig. 2. (a)Inputted image, (b) Generating bounding box after object detection Inputted image.

Fig. 3. Mask of the Car object.

B. *Morphological transformation & image inpainting:*

The generated masks were used for inpainting the input images. After evaluating the resulting image as shown in figure 5, it can be concluded that the image inpainting did

not do a very good job as there are some unwanted parts left in the image. The reason for this was that the mask generated by Mask RCNN did not compensate for the real-world implications of objects in that environment such as the shadow of the car. Hence when inpainting the image with that mask it does not envelope both the vehicle and its shadow generating a faulty outcome.

To rectify the above-mentioned issues morphological transformation was introduced to the generated masks. Morphological transformation is used on binary images to manipulate image shapes using a kernel that dictates the nature of the manipulation. The generated masks were fed to the morphological transformation function with the kernel of 4 x 7 for dilating the masks. The resulting dilated mask is shown in figure 4b.

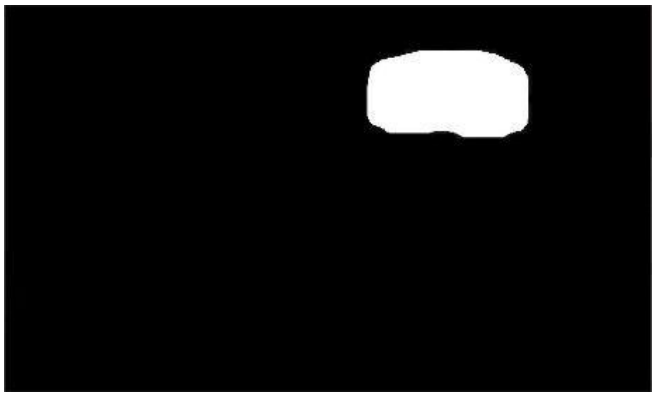

Fig. 4. Mask of the car after dilation.

With the dilated masks image inpainting was performed on the original input image. As the dilated mask compensated for the environmental elements, it composed a much better result than without dilation, as shown in figure 6. Image inpainting is an image rebuilding method through which an image is reconstructed. It can also be used for removing unwanted elements such as noise, marks, texts, or in some cases even damaged parts to restore an image. The basic methodology of image inpainting is to remove unwanted pixels and then replace them with adjacent pixels.

Fig. 5. Image inpainting without dilation.

Fig. 6. The output of inpainting after using the dilated mask.

C. Output Validation:

A function that calculates the similarity of two images was introduced as a validation function. It was used to calculate the similarity between the output image from inpainting (Figure 6) and the image of the same frame without the vehicle(Figure 7). The validation function works by using openCV's SIFT(Scale Invariant Feature Transform) to extract key points from the images to be compared. Then openCV's KNNMatch is used to match the descriptors of both images and return the good matches. The final result is compiled by using a ratio of the match points and key points. The resulting similarity index of the validation function will dictate how similar the resulting image is to the image taken without the car. The higher the similarity percentage, the higher the success rate of the proposed model.

Fig. 7. Image without car taken from the same frame.

IV. Result And Analysis

The process proposed in this paper produces a mask with dilation operation from the input image after detecting the vehicles of the image. The white color indicates the car. Then applying the inpainting method with dilation on the image makes the car disappear from the image. Using RCNN in this method to detect objects and later to generate masks shows its usefulness. The performance of RCNN in detection proves that it is a good choice if there is a task of object detection and mask generation. The method explained in this paper used Mask RCNN to detect cars, and the confidence turned out to be 98.7%. This result speaks for RCNN's viability in object detection.

The morphological transformation was achieved by doing image dilation in this proposed method. Using Morphological transformation showed its influence on the result. The side-by-side comparison between Morphological transformation and without Morphological transformation makes it evident. It can be visually compared from images 5 and 6, after impainting the images using both methods shows which one performed better.

After comparing the figures(Figures 5&6), it can be seen that figure 6 generates a better result compared to figure 5.Between these two images, the first image(figure 5) does not use Morphological transformation and the second image(figure 6) does. Because of the usage of a transformed mask in the second image (figure 6), there are no vehicles visible in the image where figure 5 has some vehicle sections visible. While analyzing methods, it can be seen that if the generated mask is used in the model for image inpainting the model accuracy rate comes around 93.73%. But using a transformed mask changes the story. Using a transformed mask increases the accuracy rate of the model. And the result of the accuracy is visible in the resulting photos in Figures 5 and 6. Usage of the transformed mask made the portion of the car visible in figure 5 disappear in figure 6. The reason behind this kind of behavior can be found out if some characteristics of Mask RCNN are analyzed. Here the shadow of the object is one of the key reasons for inaccuracy. Because if Mask RCNN is analyzed it can be seen that it excludes shadow of the objects in the image while running the segmentation process. In the image inpainting method, neighboring pixels have a good impact in

the inpainting method. Normally the neighboring pixels of the object are substituted to do the image inpainting. Because of that, the shadow of the object create a problem in the inpainting process. This problem is solved by enlarging the mask area that covers the shadows. Covering the shadow while masking the object leaves no trace of the shadow of the object. Because of that, a better result is generated with dilation compared to without dilation. The method using dilation while inpainting provides an accuracy of 95.6%, which is a great improvement over the results generated by the method without using dilation. So it can be seen that the dilation of images plays a great role in improving the method and getting the most optimal output. While not using it fails to generate optimal results. In figure 10 the difference between the two methods is marked using a red mask. Red marks in figure 10 provide information on how much the dilation method is helping to improve the performance of this proposed method. More Outputs are given below in figure 11.

```
G:\. bs\results>python SimilarImagePercentage.py
Key points in first image :319
Key points in second image :395
The matching percentage of the images : 93.73040752351098
Good matches: 299
```

Fig. 8. Accuracy of inpainting without dilation

```
Key points in first image :2394
Key points in second image :2271
The matching percentage of the images : 95.59665345662704
Good matches: 2171
```

Fig. 9. Accuracy of inpainting with dilation

Fig. 10. Difference between input and output picture

V. The Novelty of the Work

The process described in this paper can be used to expand any dataset that has images of roadways with vehicles. If a model requires road image data without vehicles for training, image data with cars can be converted into images of roads without vehicles using this method. Self-driven vehicle models are frequently taught in a virtual environment. Scaling up can improve those self-driving vehicles in simulated environments. Data-driven algorithms can scale up and enhance their model by creating new data and delivering more data to the model. It has been demonstrated that combining large-scale driving datasets with statistical models can increase efficiency by a factor of ten thousand [30]. However, a new study will require roughly 8.8 billion driving miles to present adequate evidence to compare the safety of autonomous vehicles with human driving based on logged data if they want this appealing outcome[30]. As a result, researchers can use our proposed algorithm to expand the amount of image data in an existing dataset. This strategy can be used by researchers who do not have access to a huge amount of data to train their models.

VI. Conclusion

After analyzing and reviewing the method, it is clear that it is a useful viable method for generating augmented data by removing vehicles from any road data set. It is a sturdy way of road data generation. Using this, better training data can be generated without going through the hassle of collecting it from the field. Also, from the results and analysis section, the viability of the method is proven. This provided method performed well in both object detection and image inpainting. The accuracy of the results proves the competence of the explained method. This method can be of great use for independent researchers and small computer vision-based software. Comparatively, it is more costly to collect data from the field compared to generating it from an existing dataset. The proposed method can be used to make a new dataset or increase the existing dataset. It will open a new door to training data-driven models.

References

[1] Li, W., Pan, C., Zhang, R., Ren, J., Ma, Y., Fang, J., Yan, F., Geng, Q., Huang, X., Gong, H., Xu, W., Wang, G., Manocha, D. and Yang, R., 2019. AADS: Augmented autonomous driving simulation using data-driven algorithms. Science Robotics, 4(28), p.eaaw0863.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] H. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 343-347, doi: 10.1109/ICIP.2014.7025068.

[3] A. Z. M. Tahmidul Kabir, A. M. Mizan, N. Debnath, A. J. Ta-sin, N. Zinnurayen and M. T. Haider, "IoT Based Low Cost Smart Indoor Farming Management System Using an Assistant Robot and Mobile App," *2020 10th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)*, Malang, Indonesia, 2020, pp. 155-158, doi: 10.1109/EECCIS49483.2020.9263478.

[4] A. M. Mizan, A. Z. M. Tahmidul Kabir, N. Zinnurayen, T. Abrar, A. J. Ta-sin and Mahfuzar, "The Smart Vehicle Management System for Accident Prevention by Using Drowsiness, Alcohol, and Overload Detection," *2020 10th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)*, Malang, Indonesia, 2020, pp. 173-177, doi: 10.1109/EECCIS49483.2020.9263429.

[5] M. E. Raihan, U. Rafin Akther, S. Afrin, F. M. Chowdhury and M. Rawnak Sarker, "Toddlers Working Memory Development via Visual Attention and Visual Sequential-Memory," *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICCIT48885.2019.9038580.

[6] B. Chen, W. Sun and A. Vodacek, "Improving image-based characterization of road junctions, widths, and connectivity by leveraging OpenStreetMap vector map," 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 2014, pp. 4958-4961, doi: 10.1109/IGARSS.2014.6947608.

[7] Chen, F. Huang and Y. Fang, "Integrating virtual environment and GIS for 3D virtual city development and urban planning," 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 2011, pp. 4200-4203, doi: 10.1109/IGARSS.2011.6050156.

[8] J. Kim, J. Yoo and J. Koo, "Road and Lane Detection Using Stereo Camera," 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, China, 2018, pp. 649-652, doi: 10.1109/BigComp.2018.00117.

[9] M. C. Olgun, Z. Baytar, K. M. Akpolat and O. Koray Sahingoz, "Autonomous Vehicle Control for Lane and Vehicle Tracking by

Using Deep Learning via Vision," 2018 6th International Conference on Control Engineering & Information Technology (CEIT), Istanbul, Turkey, 2018, pp. 1-7, doi: 10.1109/CEIT.2018.8751764.

[10] Lin TY. et al. (2014) Microsoft COCO: Common Objects in Context. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham. https://doi.org/10.1007/978-3-319-10602-1_48.

[11] Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick; Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2961-2969.

[12] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. 2000. Image inpainting. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques (SIGGRAPH '00). ACM Press/Addison-Wesley Publishing Co., USA, 417–424. DOI:https://doi.org/10.1145/344779.344972.

[13] L. H. Jadhav and B. F. Momin, "Detection and identification of unattended/removed objects in video surveillance," *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, India, 2016, pp. 1770-1773, doi: 10.1109/RTEICT.2016.7808138.

[14] L. Chonglun, "Removing Natural Objects from the Sea Surface Background Image Based on Contour Map and Local-Hausdorff Distance," *2016 3rd International Conference on Information Science and Control Engineering (ICISCE)*, Beijing, China, 2016, pp. 519-526, doi: 10.1109/ICISCE.2016.118.

[15] M. Mahajan and P. Bhanodia, "Image inpainting techniques for removal of object," *International Conference on Information Communication and Embedded Systems (ICICES2014)*, Chennai, India, 2014, pp. 1-4, doi: 10.1109/ICICES.2014.7034008.

[16] N. Shan, D. S. Tan, M. S. Denekew, Y. -Y. Chen, W. -H. Cheng and K. -L. Hua, "Photobomb Defusal Expert: Automatically Remove Distracting People From Photos," in *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 5, pp. 717-727, Oct. 2020, doi: 10.1109/TETCI.2018.2865215.

[17] J. Liu, S. Teng and Z. Li, "Removing Rain from Single Image Based on Details Preservation and Background Enhancement," *2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP)*, Weihai, China, 2019, pp. 322-326, doi: 10.1109/ICICSP48821.2019.8958586.

[18] J. Shi, Y. Zhou and W. X. Q. Zhang, "Target Detection Based on Improved Mask Rcnn in Service Robot," *2019 Chinese Control Conference (CCC)*, Guangzhou, China, 2019, pp. 8519-8524, doi: 10.23919/ChiCC.2019.8866278.

[19] A. Azeem, W. Riaz, A. Siddique and U. A. K. Saifullah, "A Robust Automatic Meter Reading System based on Mask-RCNN," *2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications(AEECA)*, Dalian, China, 2020, pp. 209-213, doi: 10.1109/AEECA49918.2020.9213531.

[20] Johnson J.W. (2020) Automatic Nucleus Segmentation with Mask-RCNN. In: Arai K., Kapoor S. (eds) Advances in Computer Vision. CVC 2019. Advances in Intelligent Systems and Computing, vol 944. Springer, Cham. https://doi.org/10.1007/978-3-030-17798-0_32.

[21] Li X., Liu Q., Wang J., Wu J. (2020) Container Damage Identification Based on Fmask-RCNN. In: Zhang H., Zhang Z., Wu Z., Hao T. (eds) Neural Computing for Advanced Applications. NCAA 2020. Communications in Computer and Information Science, vol 1265. Springer, Singapore. https://doi.org/10.1007/978-981-15-7670-6_2.

[22] M. Songhui, S. Mingming and H. Chufeng, "Objects detection and location based on mask RCNN and stereo vision," *2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, Changsha, China, 2019, pp. 369-373, doi: 10.1109/ICEMI46757.2019.9101563.

[23] Y. Ren, X. Yu, R. Zhang, T. H. Li, S. Liu and G. Li, "StructureFlow: Image Inpainting via Structure-Aware Appearance Flow," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 181-190, doi: 10.1109/ICCV.2019.00027.

[24] D. J. Tuptewar and A. Pinjarkar, "Robust exemplar based image and video inpainting for object removal and region filling," *2017 International Conference on Intelligent Computing and Control (I2C2)*, Coimbatore, India, 2017, pp. 1-4, doi: 10.1109/I2C2.2017.8321964.

[25] Y. Wang, S. Liu, C. Chen and B. Zeng, "A Hierarchical Approach for Rain or Snow Removing in a Single Color Image," in *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3936-3950, Aug. 2017, doi: 10.1109/TIP.2017.2708502.

[26] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, Świnoujście, Poland, 2018, pp. 117-122, doi: 10.1109/IIPHDW.2018.8388338.

[27] O. Bailo, D. Ham and Y. M. Shin, "Red Blood Cell Image Generation for Data Augmentation Using Conditional Generative Adversarial Networks," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, 2019, pp. 1039-1048, doi: 10.1109/CVPRW.2019.00136.

[28] Gorad, B. & Kotrappa, D. S. Novel Dataset Generation for Indian Brinjal Plant Using Image Data Augmentation *IOP Conference Series: Materials Science and Engineering*, *IOP Publishing*, 2021, *1065*, 012041.

[29] Pandey, S.; Singh, P. R. & Tian, J. An image augmentation approach using two-stage generative adversarial network for nuclei image segmentation *Biomedical Signal Processing and Control*, 2020, *57*, 101782.

[30] Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albumentations: Fast and Flexible Image Augmentations. Information 2020, 11, 125. https://doi.org/10.3390/info11020125.

[31] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?" Transportation Research Part A: Policy and Practice, vol. 94, pp. 182–193, 2016.

[32] C. Xie *et al*., "Image Inpainting With Learnable Bidirectional Attention Maps," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 8857-8866, doi: 10.1109/ICCV.2019.00895.

[33] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu and T. S. Huang, "Generative Image Inpainting with Contextual Attention," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5505-5514, doi: 10.1109/CVPR.2018.00577.

[34] A. Mahmoud, S. Mohamed, R. El-Khoribi, and H. AbdelSalam, "Object Detection Using Adaptive Mask RCNN in Optical Remote Sensing Images," *International Journal of Intelligent Engineering and Systems*, vol. 13, pp. 65–76, 2020.

[35] J. Harikrishnan, A. Sudarsan, A. Sadashiv and R. A. S. Ajai, "Vision-Face Recognition Attendance Monitoring System for Surveillance using Deep Learning Technology and Computer Vision," *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, 2019, pp. 1-5, doi: 10.1109/ViTECoN.2019.8899418.

[36] Y. Wu, X. Peng, K. Ruan, and Z. Hu, "Improved image segmentation method based on morphological reconstruction," *Multimed. Tools Appl.*, vol. 76, no. 19, pp. 19781–19793, 2017.

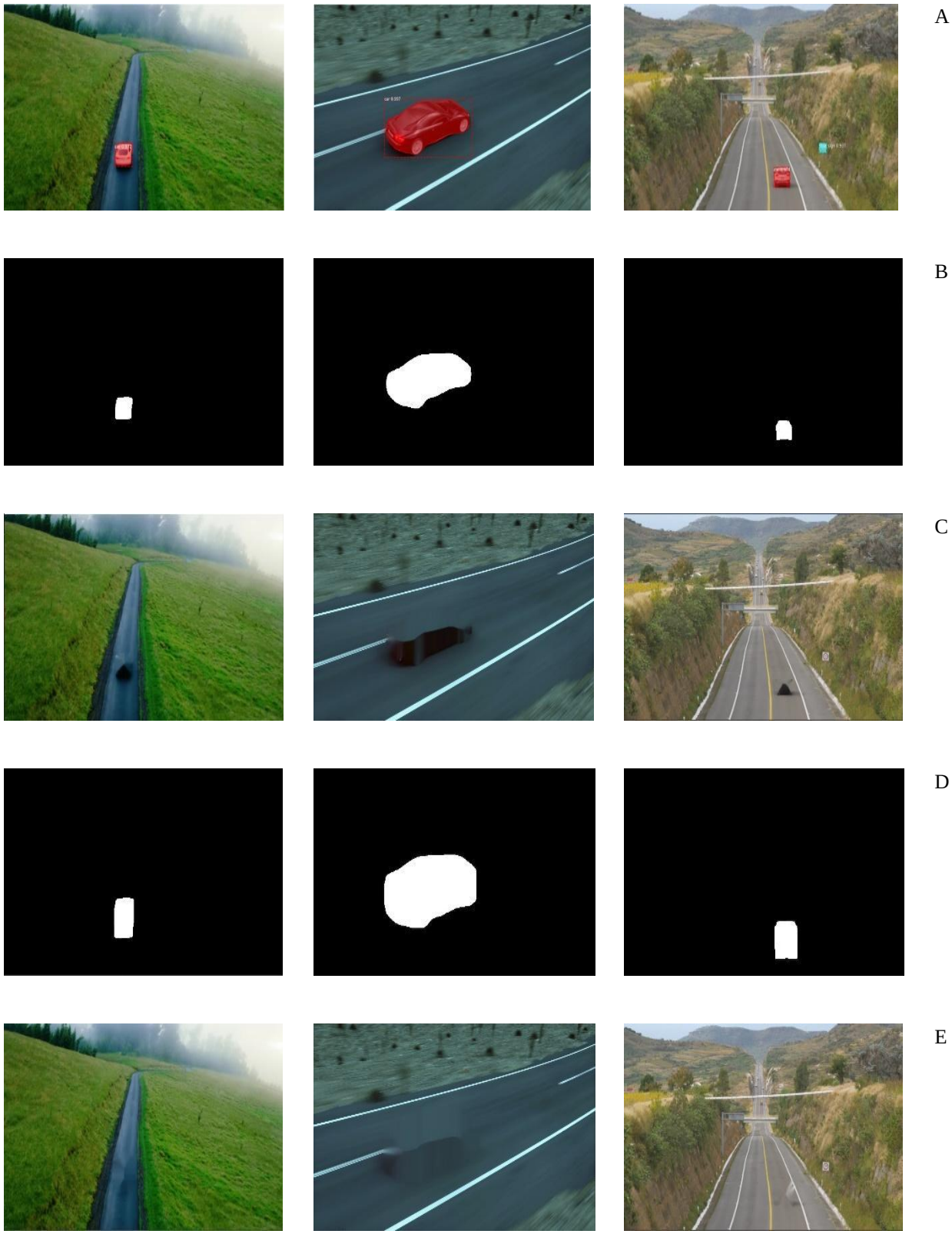

Fig. 11. A. confidence rate of identifying the vehicle to remove, B.Mask of the car without dilation, C. Output after inpainting using mask without dilation, D. Mask of the vehicle after dilation, E. output after inpainting using mask with dilation.