Article

Rock Particle Segmentation Using Mask R-CNN

Fiyinfoba Ogunkeye [*] , Liqun Shan , Xiali Hei

Article

Rock Particle Segmentation using Mask R-CNN

Fiyinfoba Ogunkeye *, , Liqun Shan and Xiali Hei

Centre for Advanced Computing Studies, School of Computing and Informatics, University of Louisiana at Lafayette, LA 70503, USA

*   Correspondence: fobaogunkeye@gmail.com.

Abstract: Identification and segmentation of rock particles is very important in geology, mining, engineering construction, and environmental monitoring. Traditional methods for recognizing the rock particle characteristics are time-consuming, labor-intensive, and limited by the geographical scope of the sampling sites, which makes it more difficult to represent the spatial distribution of the rock particles accurately. Recent advancements in deep learning techniques offer promising solutions to automate this process. This study aims at utilizing state-of-the-art (SOTA) Mask R-CNN deep learning algorithm from the Detectron2 framework, to capture global features of rocks that will enable the recognition of rock particles more efficiently, just from having the rock images. These images are subject to image processing techniques such as a Gaussian filter, to denoise and smoothen the image, and an Illumination Adaptive Transformer (IAT) framework, to adjust the lighting exposure; all of which add to improve the quality of the image dataset. Afterward, the Mask R-CNN and Detectron2 models are utilized in training the images using transfer learning. Experimental results from evaluating the performance of the proposed algorithms showcase the effectiveness of the approach, thereby highlighting its potential to revolutionize rock particle segmentation across various domains.

Keywords: rock segmentation; mask R-CNN; Detectron2; instance segmentation; deep learning

1. Introduction

Computer vision has advanced greatly in recent years with deep learning approaches pushing the boundaries of image analysis and object recognition [1]. These approaches have created opportunities for the segmentation of complex and irregular objects such as rock particles in geological images. Accurate identification and segmentation of rock particles is vital and holds high significance in geology [2], and other industries such as engineering construction [3–5], mining [6,7], and the petroleum industry [8], just to mention a few. In geology for example, geological analysis has shown to be essential to understanding the Earth's history, processes, and composition, thus proving vital in lithology, mineral prospecting, environmental studies, and hazard assessment [9]. Before carrying out this analysis, Geologists carefully study rock formations where they can identify geological markers, from microscopic grains to large slabs of rock, that can provide essential information to understanding ancient ecosystems, geological transitions, and current geological dynamics [10]. By examining the shapes, sizes, and textures of these rock particles, Geologists can utilize the information to reveal key characteristics of the Earth's geology, minerals, sedimentary processes, and climate, thereby providing a stronger foundation for explaining the Earth's past and predicting its future [11]. However, carrying out any form of statistical analysis to extract useful information from these rock particles is only possible after the accurate segmentation of the particles.

Traditional methods of rock segmentation have required manually identifying and classifying the rocks either physically or from photographs. This human-subjective particle recognition process is laborious, time-consuming, and prone to standardization errors that may emerge from the growing data volumes, thus rendering the process inefficient and unreliable. This brought forth the incorporation of image segmentation techniques like thresholding, watershed, region and edge-based segmentation, and clustering [12,13], to address the limitations of the traditional method. However, these segmentation techniques have been found to be ineffective when working on a batch of images, most specifically when these images have occluded objects or irregular shapes of different sizes. This is so, because, they expect the objects in the images to exhibit specific characteristic features based on their background,

lighting, size, or shape. They also require manual tuning of their parameters and fail to produce accurate segmentations when the objects don't have clear boundaries, as they are more useful with high-contrast images.

Matthew [14] implemented a multi-Otsu threshold algorithm [15] to segment a synthetic rock image dataset generated from sandstone and shale rock. For the greyscale rock objects with high noise levels, the algorithm was ineffective, while also failing completely at segmentations for non-greyscale (textural contrast) images. Qinpeng et al [5] proposed an adaptive watershed algorithm for segmenting blasted rock particles based on the rock contour solidity threshold. The algorithm was able to reduce errors from blurry edges and overlapping particles. Still, it wasn't robust enough as it required different solidity thresholds for different rock types which had to be manually set up. Haibo and Jialing [16] combined characteristics of the Fuzzy C-means clustering (FCM) algorithm with a semi-supervised Support Vector Machine (SVM) for segmenting rock images by taking into account the spatial information of the image pixels. Amankwah and Aldrich [17] came up with a novel approach to the watershed algorithm for segmenting rocks, solving the watershed over-segmentation problem by leveraging the adaptive threshold technique to generate object markers. This ensured the watershed segmentation lines correctly matched the object boundaries. Now, although exponential development has occurred with the introduction of these techniques in comparison with prior manual human segmentation methods, they are not robust enough for rock particles of varying features - sizes, and shapes.

Consequently, classical machine learning algorithms like the supervised learning algorithms: K-nearest neighbors (KNN), ensemble learning - Adaboost, Random Forest, SVM, and the unsupervised learning: K-Means Clustering, have shown prospects in recognizing and processing rock particles from digital images [18–21]. These algorithms require a diverse dataset containing tons of images with particles along with their corresponding labels so the model built from the algorithm can learn the intricate features of the data. Junxing and Roman [22] utilized the Adaboost ensemble algorithm with varying aspect ratios and sliding window orientation angles for recognizing sand particles. However, although the authors used a large training dataset of 85,000 images, the proposed method still failed to detect particles with irregular surfaces and elongated sizes. Therefore, there is a need for a more robust and adaptable system that can learn hierarchical representations captured in complex particle features with varied object boundaries and lightning conditions, thus leading to the implementation of Deep Learning systems [4,23–25].

In recent years, Deep Learning computer vision systems have paved the way in capturing global features of objects thus pushing the boundaries of image analysis and object recognition [1], and creating opportunities for the segmentation of complex and irregular objects such as rock particles in geological images. These systems offer robust capabilities in recognizing complex object features by passing in the training data with pixel-wise labels, so the model can learn the unique mappings of the rock objects [26]. Karolina and Sebastian [20] proposed a multi-textural approach in segmenting rock grains using the SVM, KNN, and a fully connected Artificial Neural Network (ANN) with the Levenberg-Marquardt back propagation method. Although the ANN classifier performed the best at a 79% accuracy score, the authors identified the need for post-processing techniques to improve the border thinning and automatic closing of delineated contours of the rock particles. With remarkable achievements being made from the emergence of Convolutional Neural Networks (CNN) [27], the efficiency and accuracy of segmentation techniques enhanced through Semantic segmentation architectures like the Fully Convolutional Network (FCN) and U-Net [7,24,28–32], providing a means of classifying every image pixel into a class label. Huang et al [28] utilized an FCN for the segmentation of cracks and leakage defects in tunnels, resulting in an improved performance in comparison to traditional techniques like Adaptive thresholding and region growing algorithm. In the mining industry, Yang et al [7] leveraged depth-separable convolutions with feature-depth concatenation to improve the rock particle segmentation accuracy, using an enhanced U-Net model architecture. Zhou et al [24] proposed a dual U-Net with multi-scale inputs and side-output for the automatic

segmentation of muck images. Even though the U-Net algorithm was improved upon to produce more accurate results, it encountered limitations when segmenting overlapping rock objects which led to under-segmentation and over-segmentation. Additionally, the proposed algorithm did not bear in mind the regional features of the rocks such as the edges, and the rock surface. Liang et al [29] developed a deep convolutional neural network, by leveraging the U-Net architecture to extract particle projects from raw images. However, the authors highlighted the lack of generalization with the models' predictions when attempting to segment rock types that were different from what was used in the training dataset.

Due to its design, semantic segmentation encounters limitations when the objects are densely packed with overlapping particles, leading to under-segmentation or over-segmentation [24]. Beyond semantic segmentation where segmentation identifies and labels every image pixel in the same class under the same category, instance segmentation has been shown to accurately distinguish overlapping instances by delineating object boundaries, making it easier to carry out further independent actions on the individual rock particles. For example, size calculation, shape identification, or any form of counting/tracking application would be more straightforward. These instance segmentation algorithms - Mask RCNN [33], in the Detectron2 [34] framework, are better suited to identify each individual object instance, making it more adaptable and efficient for rock recognition. Fan et al [4] utilized the Mask R-CNN model to segment rockfill particles by leveraging the ResNeXt101 backbone with the squeeze and excitation block, enhancing the feature extraction capability. Trong et al [35] developed an automatic means of measuring blast fragments in open-pit mines using the Mask R-CNN algorithm, although the authors observed a decline in segmentation accuracy due to poor spatial resolution of the image. As the application of instance segmentation methods to rock recognition is fairly novel, this research study provides a stepping stone for further research.
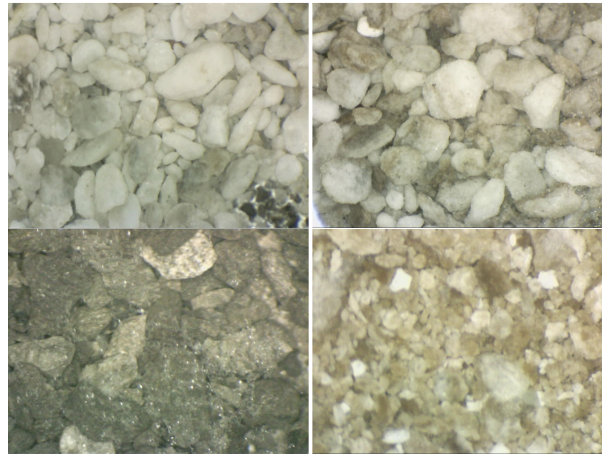
In this study, Mask R-CNN and Detectron2 models are implemented for recognizing rock particles by leveraging the transfer learning technique, which saves time and delivers excellent results with limited training data [36]. To start off, the quality of the training data is improved through a Gaussian filter and an Illumination Adaptive Transformer (IAT) [37], to denoise the image and to adjust the lighting exposure respectively. Then, the models are trained using the annotated preprocessed data with different hyperparameters, after which the performance of the proposed algorithms is evaluated using the standard Average Precision (AP) evaluation metric on a test dataset with diverse environments, by comparing the identification ability with similar research works. The rest of the paper is organized as follows: **Section 2** covers the proposed methodology, introducing the dataset with the segmentation algorithms. **Section 3** describes the training process implemented on the preprocessed data as well as the steps involved in evaluating the developed models. In **section 4**, the results from the training experiments are discussed alongside identified drawbacks. Finally, **Section 5** concludes the paper with recommendations for future work.

2. Methodology

2.1. Data Collection

The dataset comprises of 69 images containing multiple rock particles, each obtained from a drilling company. These images, as seen in Figure 1 exhibit diverse environmental conditions, spanning underwater and outdoor settings, with variations in lighting, humidity, and surface wetness. Encompassing a wide spectrum of particle sizes, shapes (from gravel to cobbles), and colors, the dataset reflects real-world scenarios, thus making it suitable for robust analysis.

Figure 1. Sample of Dataset.

However, it should be noted that the raw images were beset with certain challenges that required rectification for reliable analysis:

- Blurriness: the resolution of several images was unclear, hampering the image clarity.
- Indistinct Boundaries: Distinguishing particle boundaries or edges in the images proved to be difficult.
- Limited data size: The dataset was small, comprising only 69 images to work with.

2.2. Data Quality Enhancement and Preprocessing

To mitigate the data issues highlighted, 2 key techniques were carried out:
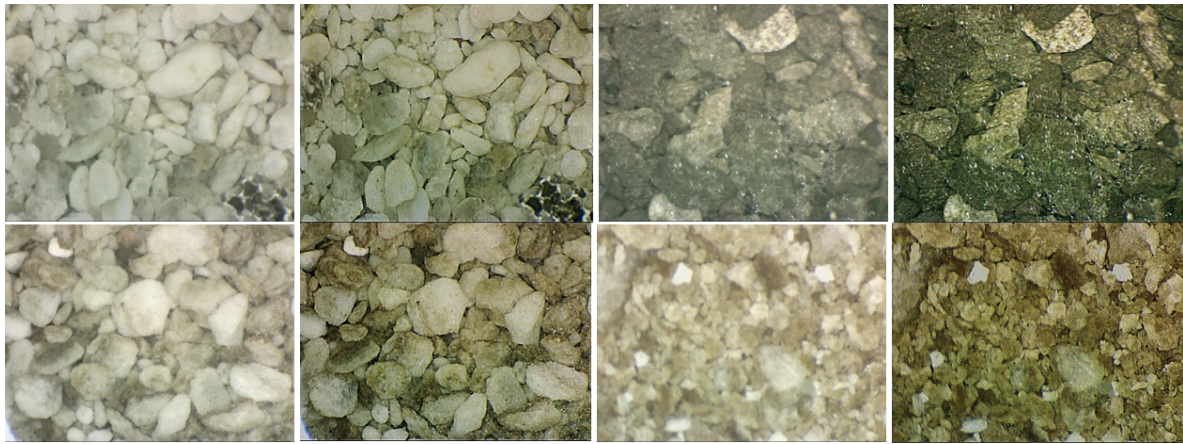
2.2.1. Image Denoising

Due to the amount of noise and blur in the dataset, a Gaussian filter was utilized to denoise and smoothen the images. The filter uses the weighted mean of the pixels surrounding the central pixel to replace the central pixel - the pixels further from the central pixel will contribute less weight to the average while those closer would have a more weighted effect. This helps us better preserve the edges of the rock objects in the images. Equation 1 shows the equation for a Gaussian function in two directions (x and y axis), where x and y are the respective distances to the horizontal and vertical center of the kernel, and σ is the standard deviation of the Gaussian kernel.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

Afterward, the image was sharpened by blending the blurry and denoised images through a weighted sum of the weights from the blurry image with the weights of the denoised image.

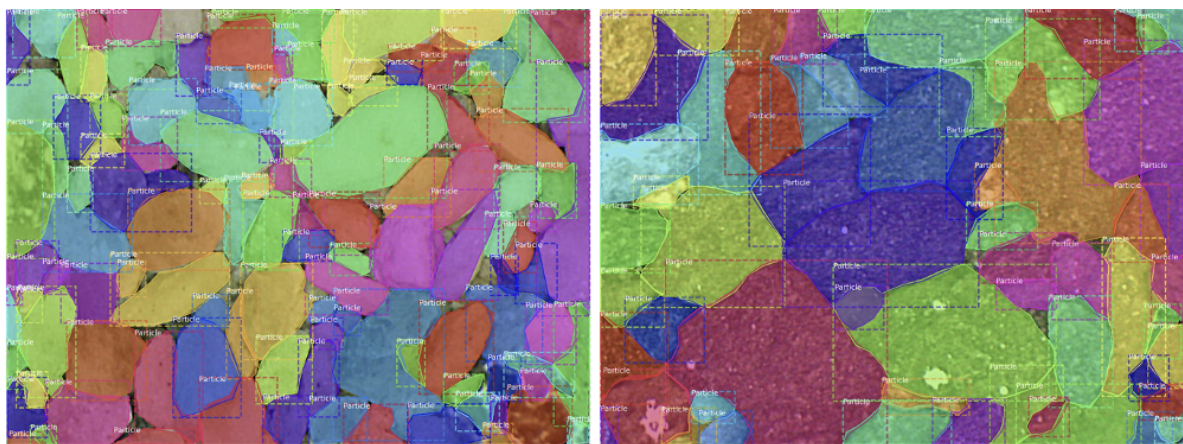2.2.2. Exposure Correction

To adjust the lighting exposure on some of the images that had either too bright or too low a contrast, a transformer network called IAT, which uses attention queries to control the image signal processing (ISP) parameters like color transform matrix, and gamma value, was employed thereby improving the overall image quality. The resulting preprocessed data is shown in Figure 2.

Figure 2. Raw data vs Preprocessed data.

2.3. Data Annotation

For the final preparation of the dataset, the data was annotated using the makesense.ai labeling tool to obtain the ground truth mask labels and bounding box coordinates which the model will use to learn, as seen in Figure 3. A dedicated script was developed to parse and load the exported JSON files from the labeling tool. This script was instrumental in overlaying the masks and bounding boxes on the images, ensuring the accuracy and correctness of the labeling process.

Figure 3. Samples of annotated data.

2.4. MASK R-CNN

Mask R-CNN is a multi-stage instance segmentation technique that detects objects in an image and generates a mask to segment each object instance. It extends the Faster R-CNN algorithm with an additional branch to output a binary mask for each Region of Interest (ROI) in parallel with the object class and bounding box prediction as shown in Figure 4. ROIs are boxes drawn to tell us where an object might be present in the image. They are extracted through the Region Proposal Network (RPN) which was initially introduced in the Faster R-CNN framework as a neural network with attention. It takes in feature maps of raw images extracted from a Feature Pyramid Network (FPN) with the ResNet backbone while leveraging anchor boxes to allow the ROI to be detected at different scales and aspect ratios. The final stage involves using the Fully Connected (FC) layers to predict the bounding box coordinates of the object with the category of the object class, and a mask branch which uses an FCN leveraged from semantic segmentation, to extract and maintain the spatial layout of the mask unlike the FC layers from Faster R-CNN that collapses the spatial structure into a vector representation. With the FCN, the mask is generated using fewer parameters and more accurately.
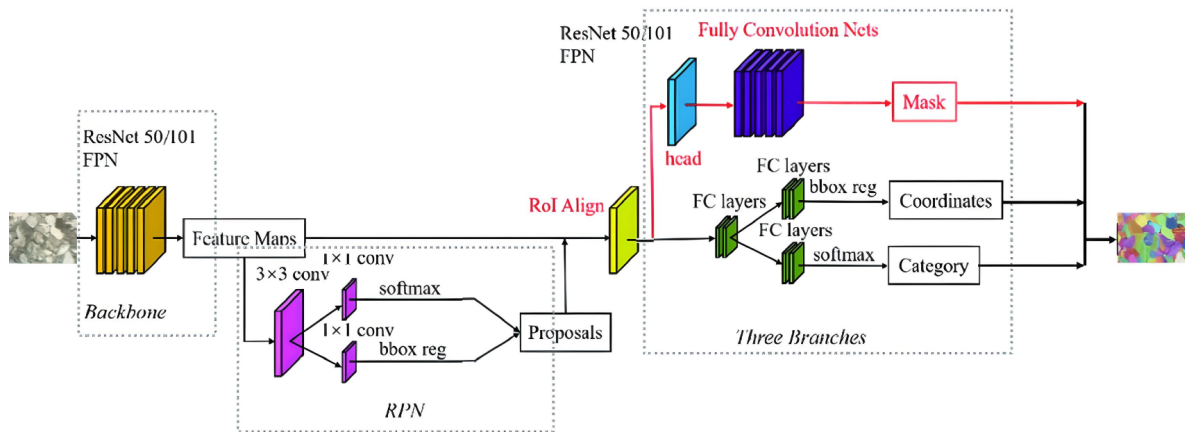
Figure 4. Mask R-CNN Architecture [38].

When Mask R-CNN was introduced, it was developed to solve some of the drawbacks of the Faster R-CNN - the ROIPool layer which introduced misalignments between the ROI and the extracted features that affected the accuracy of the predicted pixel masks. These misalignments were a result of the quantization of the ROI boundaries/bins while extracting the features, such that a continuous coordinate x becomes [x/16] due to quantization. To resolve this, the ROIAlign layer was introduced to preserve the spatial dimensions for pixel-to-pixel alignment of the features. As the Mask R-CNN has 3 branches, it has a multi-task loss function (2), for each of the branches - the classifier loss, bounding box loss, and the mask loss.

$$L = L_{cls} + L_{box} + L_{mask} \tag{2}$$

For every ROI with ground truth, k, the mask loss is defined on only the K_{th} mask (loss is independent of other mask outputs). Masks are generated for every object without competition using the class predicted to select the mask rather than how prior systems were built where the classification depends on the mask predictions, or using a per-pixel sigmoid and binary loss rather than a per-pixel softmax and multinomial cross-entropy loss from previous systems.

2.5. Detectron2

Detectron2 framework was built and leveraged from the Mask R-CNN architecture. It was developed by Facebook AI Research for its application in object detection and instance segmentation tasks. With similarity to the Mask R-CNN, it has support for more feature extraction backbones from the FPN, ResNet, and ResNeXt than Mask R-CNN; in addition to its modular and extensible design. This makes Detectron2 invaluable for this study as it can be utilized in extracting essential morphological features of individual rock particles, thereby producing instance segmentation masks. The utilization of the Detectron2 framework permitted us the opportunity to ensure that our system was robust since we could evaluate and compare the different algorithm's performance.

3. Implementation Details

Before the model training began, the dataset was split into training, validation, and test sets to make it easy to evaluate the mask r-cnn and detectron2's performance not just on the data it has learned from (training and validation data), but on data yet to be seen (test data). This approach is a standard method recognized as it facilitates model generalization on the patterns present in the dataset, preventing potential cases of false predictions.

3.1. Evaluation Metrics

To evaluate the models, the Average Precision (AP) standard COCO evaluation metric [33] was used at different thresholds, where this threshold is determined by the Intersection over Union (IoU) value - how close the ground truth mask/bounding box values match with the predicted

masks/bounding boxes generated by the model. In this study, the AP is evaluated at the default IoU threshold of 0.5 - **AP50**, 0.75 IoU threshold - **AP75**, and across all 10 IoU thresholds (0.5 to 0.95 with increments of 0.05) - **AP**. The closer the value is to 1 (100%), the more accurate the segmentation.

3.2. Transfer Learning

In order to adapt to the limited data size that the drilling company provided us with, the transfer learning technique was leveraged in the model training of both the Mask R-CNN and Detectron2 models. This technique prevents training the models from scratch which would be a daunting task especially when the dataset size is low, as it can lead to poor performance since the network expects tons of images to learn and generalize from. Instead, the knowledge garnered from feature representations of models that have been trained on large volumes of data for solving different problems was leveraged. By doing this, the knowledge can be transferred to the specific use case with a limited data size, while getting good results. In this paper, the pre-trained mask r-cnn model built to recognize 80 different classes from the COCO dataset such as cars, dogs, traffic lights, etc. [33,39], is leveraged in recognizing the rock particles.

3.3. Training the Mask R-CNN Model

Several hyperparameters were experimented with by altering the backbones, learning rate, detection confidence threshold, and more, to finalize at this model with the following features.

ResNet101 performed better than ResNet50 [33] so it was selected as the backbone network, with a learning rate of 0.003 after experimenting with values ranging from the default - 0.001 to 0.00025. The confidence threshold for an ROI detected to be accepted was increased from the default - 0.7 to 0.9. Although this helped reduce false positive ROI instances, it increased the false negative occurrence rate. Then, an early stopping strategy was also added by implementing a terminating callback that will stop the training when validation losses plateaued, thereby preventing overfitting and ensuring model generalization. Finally, to actualize transfer learning for convergence and optimal training, a four-phase training approach was explored:

- Phase 1: Training the network head for 40 epochs with an increased learning rate at 0.006 (learning rate * 2)
- Phase 2: Fine-tuning layers 3+ (ResNet stage 3 and up) for 120 epochs
- Phase 3: Fine-tuning layers 4+ (ResNet stage 4 and up) for 160 epochs
- Phase 4: Fine-tuning all layers for 200 epochs with a reduced learning rate at 0.0003 (learning rate / 10)

3.4. Training the Detectron2 Model

Here, five different baseline models were experimented with, all having the ResNet+FPN backbone, and 3-times schedule: ResNet50, ResNet101, ResNeXt101-32x8d trained with caffe2, ResNet101 using Large-Scale Jitter and Longer Training Schedule [40], and Cascade R-CNN - ResNet50 with Cascade ROI heads. All models utilized a learning rate of 0.0025, batch size of 512, and maximum iterations of 10,000. They were all evaluated against the test data using the AP metric.
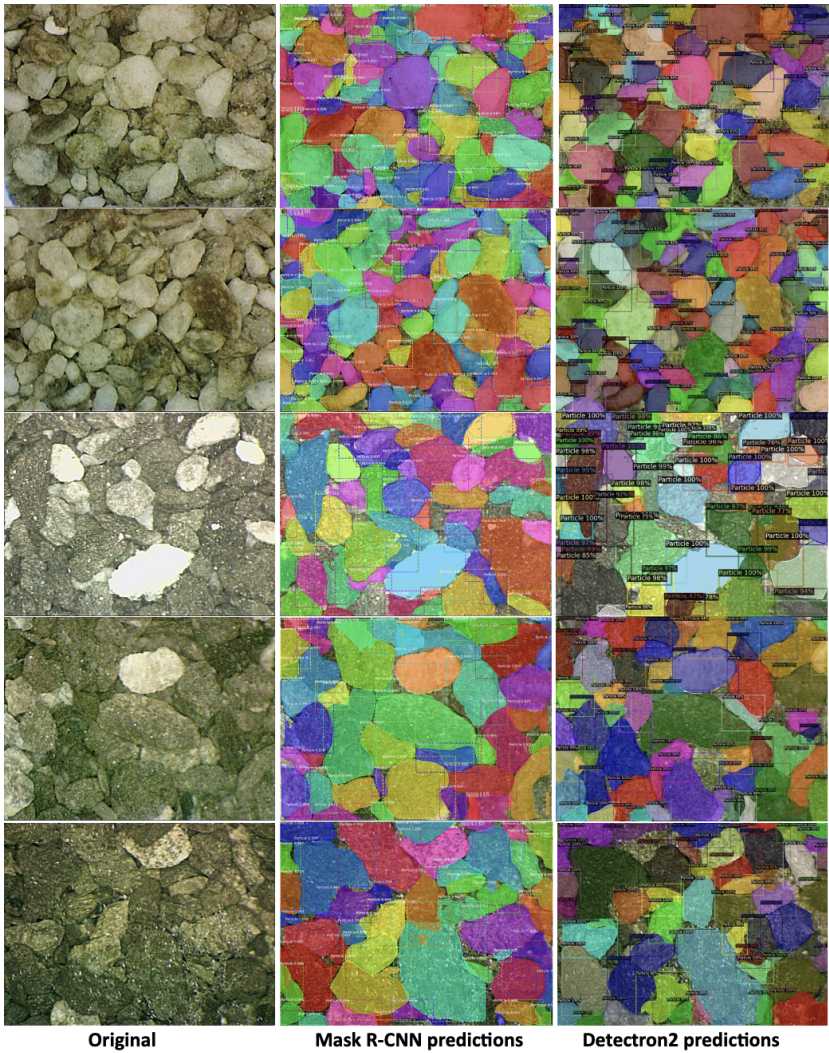
4. Results

An Nvidia Quadro RTX 8000 GPU Linux server with 48GB memory was used as the training environment for the developed models. The GPU provided us with an environment to experiment with different hyperparameters more effectively and efficiently. Table 1 reports the metric evaluation of the trained Mask R-CNN model with the ResNet101 backbone against the Detectron2 models with 5 different backbones, on the test dataset.

Table 1. Performance comparison of Mask R-CNN and Detectron2.

Models	AP	AP50	AP75
Mask R-CNN: ResNet101	**27.5**	**55.5**	**24.8**
Detectron2: ResNet50	19.8	42.8	17.9
Detectron2: ResNet101	15.5	34.1	13.3
Detectron2: ResNeXt101-32x8d	18.6	38.9	15.9
Detectron2: ResNeXt101-LSJ	17.3	36.9	15.6
Detectron2: Cascade R-CNN	17.5	38	13.8

Based on the results across all metrics especially the AP since it accounts for all IoU thresholds, the Mask R-CNN model had the best performance overall while the Detectron2 with the ResNet50 backbone performed the best across the Detectron2 models. The Cascade R-CNN model and the Resnet 101 with Large Scale Jitter, had the worst performances across all the metrics. Looking at the predictions visually in Figure 5, it can be seen that although the model was able to do a decent job recognizing the individual rock particles regardless of the low AP values, there are few cases of over-segmentation and under segmentation which unsurprisingly, was seen less in the mask r-cnn predictions.

Figure 5. Inferences from Mask R-CNN and Detectron2 models on validation (top 2 rows) and test (bottom 3 rows) images.

It was also observed that the models experienced more difficulty in recognizing the particles for images that had several particles in comparison to the ones with few particles, images with tiny rocks in comparison to the ones with huge rock slabs, and images that weren't very clear. Owing to the complexities and variations of the rock particle sizes, and the laborious nature of annotating data, it was recognized that the quality and accuracy of the ground truth labels used to train the models might diminish since the mask annotations were labeled by a human manually drawing the polygons on each rock particle. This poses a huge problem as it can affect the performance of any model trained with the data. Furthermore, the model performance could be further improved by expanding the dataset via data augmentation [41–43]. By transforming the limited data, the model is able to generalize and adapt better to new rock images, making it more robust to imaging scenarios from different geological contexts, and less prone to overfitting [44].

5. Conclusions

In this paper, a framework for recognizing rock particles from digital images was presented using the pre-trained Mask R-CNN and Detectron2 deep learning algorithms. These algorithms were trained and validated on an annotated preprocessed dataset using the transfer learning technique, providing a more efficient and optimal training approach. Thereafter, the segmented particles were evaluated using the Average Precision standard coco metric across several IoU thresholds. The Mask R-CNN model with the ResNet101 backbone performed the best with an AP value of 55.5 for the 0.5 IoU threshold, on the test dataset. This shows its potential for use in segmentation tasks and applications across several engineering domains. In light of the results obtained, expansion of the dataset either by means of a re-collection from the drilling company or by employing data augmentation methods such as image rotation, translation, flipping, cropping, brightness control, saturation control, etc., can provide more features for the algorithm to learn from, thereby improving its performance. This can be applied in future works alongside experimenting with other training hyperparameter values that can produce a more robust recognition capability.

Author Contributions: Conceptualization, L.S. and X.H.; methodology, F.O.; software, F.O.; validation, F.O., L.S. and X.H.; formal analysis, F.O.; investigation, L.S. and X.H.; data curation, F.O. and L.S.; writing—original draft preparation, F.O.; writing—review and editing, F.O., L.S. and X.H.; supervision, L.S. and X.H.; project administration, F.O.; funding acquisition, X.H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data is not available due to the confidentiality of datasets.

Conflicts of Interest: The authors declare no conflicts of interest

References

1. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E.; others. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience* **2018**, *2018*.
2. Wang, W. Rock particle image segmentation and systems. *Pattern recognition techniques, technology and applications* **2008**, pp. 197–226.
3. Ai, D.; Jiang, G.; Lam, S.K.; He, P.; Li, C. Automatic pixel-wise detection of evolving cracks on rock surface in video data. *Automation in Construction* **2020**, *119*, 103378.
4. Fan, H.; Tian, Z.; Xu, X.; Sun, X.; Ma, Y.; Liu, H.; Lu, H. Rockfill material segmentation and gradation calculation based on deep learning. *Case Studies in Construction Materials* **2022**, *17*, e01216.
5. Guo, Q.; Wang, Y.; Yang, S.; Xiang, Z. A method of blasted rock image segmentation based on improved watershed algorithm. *Scientific Reports* **2022**, *12*, 7143.

6. Thurley, M.J. Automated image segmentation and analysis of rock piles in an open-pit mine. 2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE, 2013, pp. 1–8.

7. Yang, Z.; Wu, H.; Ding, H.; Liang, J.; Guo, L. Enhanced U-Net model for rock pile segmentation and particle size analysis. *Minerals Engineering* **2023**, *203*, 108352.

8. Liu, H.; Ren, Y.L.; Li, X.; Hu, Y.X.; Wu, J.P.; Li, B.; Luo, L.; Tao, Z.; Liu, X.; Liang, J.; others. Rock thin-section analysis and identification based on artificial intelligent technique. *Petroleum Science* **2022**, *19*, 1605–1621.

9. Elvis Nkioh, N. Geological characterization of rock samples by LIBS and ME-XRT analytical techniques, 2022.

10. Rong, G.; Liu, G.; Hou, D.; Zhou, C.b.; others. Effect of particle shape on mechanical behaviors of rocks: a numerical study using clumped particle model. *The Scientific World Journal* **2013**, *2013*.

11. Rodriguez, J. Importance of the particle shape on mechanical properties of soil materials. PhD thesis, Luleå tekniska universitet, 2013.

12. Zaitoun, N.M.; Aqel, M.J. Survey on image segmentation techniques. *Procedia Computer Science* **2015**, *65*, 797–806.

13. Zhang, Z.; Yang, J.; Ding, L.; Zhao, Y. Estimation of coal particle size distribution by image segmentation. *International Journal of Mining Science and Technology* **2012**, *22*, 739–744.

14. Andrew, M. A quantified study of segmentation techniques on synthetic geological XRM and FIB-SEM images. *Computational Geosciences* **2018**, *22*, 1503–1512.

15. Otsu, N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics* **1979**, *9*, 62–66.

16. Liang, H.; Zou, J. Rock image segmentation of improved semi-supervised SVM–FCM algorithm based on chaos. *Circuits, Systems, and Signal Processing* **2020**, *39*, 571–585.

17. Amankwah, A.; Aldrich, C. Rock image segmentation using watershed with shape markers. 2010 IEEE 39th Applied Imagery Pattern Recognition Workshop (AIPR). IEEE, 2010, pp. 1–7.

18. Reinhardt, M.; Jacob, A.; Sadeghnejad, S.; Cappuccio, F.; Arnold, P.; Frank, S.; Enzmann, F.; Kersten, M. Benchmarking conventional and machine learning segmentation techniques for digital rock physics analysis of fractured rocks. *Environmental Earth Sciences* **2022**, *81*, 71.

19. Chauhan, S.; Rühaak, W.; Khan, F.; Enzmann, F.; Mielke, P.; Kersten, M.; Sass, I. Processing of rock core microtomography images: Using seven different machine learning algorithms. *Computers & Geosciences* **2016**, *86*, 120–128.

20. Nurzynska, K.; Iwaszenko, S. Application of texture features and machine learning methods to grain segmentation in rock material images. *Image Analysis & Stereology* **2020**, *39*, 73–90.

21. Harvey, A.; Fotopoulos, G. Geological mapping using machine learning algorithms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2016**, *41*, 423–430.

22. Zheng, J.; Hryciw, R.D. Identification and characterization of particle shapes from images of sand assemblies using pattern recognition. *Journal of Computing in Civil Engineering* **2018**, *32*, 04018016.

23. Liu, Y.; Zhang, Z.; Liu, X.; Wang, L.; Xia, X. Efficient image segmentation based on deep learning for mineral image classification. *Advanced Powder Technology* **2021**, *32*, 3885–3903.

24. Zhou, X.; Gong, Q.; Liu, Y.; Yin, L. Automatic segmentation of TBM muck images via a deep-learning approach to estimate the size and shape of rock chips. *Automation in Construction* **2021**, *126*, 103685.

25. Liu, C.; Li, M.; Zhang, Y.; Han, S.; Zhu, Y. An enhanced rock mineral recognition method integrating a deep learning model and clustering algorithm. *Minerals* **2019**, *9*, 516.

26. Hao, S.; Zhou, Y.; Guo, Y. A brief survey on semantic segmentation with deep learning. *Neurocomputing* **2020**, *406*, 302–321.

27. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems* **2021**.

28. Huang, H.w.; Li, Q.t.; Zhang, D.m. Deep learning based image recognition for crack and leakage defects of metro shield tunnel. *Tunnelling and underground space technology* **2018**, *77*, 166–176.

29. Liang, Z.; Nie, Z.; An, A.; Gong, J.; Wang, X. A particle shape extraction and evaluation method using a deep convolutional neural network and digital image processing. *Powder Technology* **2019**, *353*, 156–170.

30. Liu, X.; Zhang, Y.; Jing, H.; Wang, L.; Zhao, S. Ore image segmentation method using U-Net and Res_Unet convolutional networks. *RSC advances* **2020**, *10*, 9396–9406.

31. Duan, J.; Liu, X.; Wu, X.; Mao, C. Detection and segmentation of iron ore green pellets in images using lightweight U-net deep learning network. *Neural Computing and Applications* **2020**, *32*, 5775–5790.

32. Jin, C.; Wang, K.; Han, T.; Lu, Y.; Liu, A.; Liu, D. Segmentation of ore and waste rocks in borehole images using the multi-module densely connected U-net. *Computers & Geosciences* **2022**, *159*, 105018.

33. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.

34. Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.Y.; Girshick, R. Detectron2. https://github.com/facebookresearch/detectron2, 2019.

35. Vu, T.; Bao, T.; Hoang, Q.V.; Drebenstetd, C.; Hoa, P.V.; Thang, H.H. Measuring blast fragmentation at Nui Phao open-pit mine, Vietnam using the Mask R-CNN deep learning model. *Mining Technology* **2021**, *130*, 232–243.

36. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *Journal of Big data* **2016**, *3*, 1–40.

37. Cui, Z.; Li, K.; Gu, L.; Su, S.; Gao, P.; Jiang, Z.; Qiao, Y.; Harada, T. Illumination adaptive transformer. *arXiv preprint arXiv:2205.14871* **2022**.

38. Jiao, L.; Zhao, J. A survey on the new generation of deep learning in image processing. *Ieee Access* **2019**, *7*, 172231–172263.

39. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer, 2014, pp. 740–755.

40. Ghiasi, G.; Cui, Y.; Srinivas, A.; Qian, R.; Lin, T.Y.; Cubuk, E.D.; Le, Q.V.; Zoph, B. Simple copy-paste is a strong data augmentation method for instance segmentation. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 2918–2928.

41. Qin, H.; Zhang, D.; Tang, Y.; Wang, Y. Automatic recognition of tunnel lining elements from GPR images using deep convolutional networks with data augmentation. *Automation in Construction* **2021**, *130*, 103830.

42. Liu, Y.; Wang, X.; Zhang, Z.; Deng, F. Deep learning based data augmentation for large-scale mineral image recognition and classification. *Minerals Engineering* **2023**, *204*, 108411.

43. Li, D.; Zhao, J.; Liu, Z. A novel method of multitype hybrid rock lithology classification based on convolutional neural networks. *Sensors* **2022**, *22*, 1574.

44. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *Journal of big data* **2019**, *6*, 1–48.