

# Detection and Segmentation of Iron Ore Green Pellets Using Computational Vision

Caio Mario Carletti Vilela Santos<sup>1</sup>, Gustavo Maia de Almeida<sup>1</sup>, Marco Antônio de Souza Leite Cuadros<sup>1</sup>, Raphael Mendonça Sepulcri<sup>1</sup>, Ricardo Olympio de Freitas<sup>2</sup>, Bruno Meschiatti Vasconcellos<sup>1</sup>, Ramyson de Araujo Nascimento<sup>1</sup>.

<sup>1</sup>Dept. Engenharia de Automação do Instituto Federal do Espírito Santo – IFES. - ES. Rodovia ES-010 - Km 6,5 - Manguinhos, 29173-087 - Serra - ES, Brasil

[caiomario95@gmail.com](mailto:caiomario95@gmail.com), [gmaia@ifes.edu.br](mailto:gmaia@ifes.edu.br), [marcoantonio@ifes.edu.br](mailto:marcoantonio@ifes.edu.br), [rmsepulcri@gmail.com](mailto:rmsepulcri@gmail.com), [brunomeschiatti@gmail.com](mailto:brunomeschiatti@gmail.com), [ramyson.nascimento@gmail.com](mailto:ramyson.nascimento@gmail.com).

<sup>2</sup>VALE – Complexo Tubarão, Av Dante Michelini, 5500, Jardim Camburi, Vitória, ES, Brasil, [ricardo.olympio@vale.com](mailto:ricardo.olympio@vale.com).

**Abstract.** In the industry, iron ore pellets, agglomerates with a diameter ranging from 6 to 16mm, composed mainly of fine iron oxide particles, are one of the essential inputs used in the global production of steel, where sphericity and strength of the pellets are necessary for the process, as well as the correct diameter, called the particle size range. To provide mechanical resistance to the newly formed pellets, a characteristic that prevents the pellets from breaking and turning into fines, a firing process is carried out where the thermal efficiency of this pellet is intrinsically linked to the ideal diameter and humidity of the pellets. In the work, two neural models of deep learning were presented and compared among themselves in the segmenting, and then measuring the diameter of each of the iron ore pellets, they are Mask R-CNN and YOLACT. Such work makes possible improvements in the controllers of the pelletizing discs, improving the quality of the pellets as a whole, as well as, a greater precision in the desired granulometric range of the pellets. It was seen that the two networks mentioned had excellent results, however, the Mask R-CNN proved to be more costly in processing compared to YOLACT.

**Keywords:** Deep Learning, Mask R-CNN, YOLACT, Computational Vision, Pelletizing.

## 1 Introduction

Disc pelletizer is widely used in the steel manufacturing industry in the agglomeration process to form powdered iron ore into iron ore green pellets. The raw material is fed into the disc and sprayed by water. With continuous rotation of the disc, fine particles are gathered and formed into larger pellets in the stable area, finally falling out of the disc as green pellets. The green pellets are then heat hardened to get the required properties in the induration process and finally be fed into the furnace. The pellet size in the disc should be constantly monitored to guarantee that it is in a desired size range, otherwise, the thermal process in the furnace will be greatly affected, leading to low heat transfer efficiency or bad product quality, Liu, et al [1].

Manual measurement of PSD has many limitations. For instance, the pellet samples are limited and the sampling and sieving process are rather time consuming. Human error also has a significant impact on the measuring results, Liao and Tarng [2]. Optical imaging is a good method for online monitoring of pellet size, Facco, Santomaso and Barolo, [3]. It has advantages in simple hardware configuration, wide view field and abundant choice of image processing algorithms. A typical optical imaging system is shown in Fig. 1. The images of the pellets are usually taken by a camera and then processed to detect individual pellets and to calculate the pellet size.

Traditional image processing framework for pellet size measurement includes filtering, thresholding, clustering and segmentation. Pellet detection and segmentation are crucial prerequisites in image processing procedures, the performance of which greatly affects the measuring accuracy of pellet size Heydari, et al [4].

However, the noisy background, uneven illumination, harsh light reflectance at the pellet centers and pellet overlapping are the key problems in image segmentation. Traditional algorithms, like watershed-based algorithms, Subramanyam, Patra and Singh [5] and multi-threshold segmentation algorithm, Roozbahani, Borela and Frost [6], are limited in solving such problems. For instance, watershed may result in over segmentation, meanwhile, multi-threshold algorithms require redundant setting of thresholds to achieve satisfactory results and it may turn out to be tedious work to adapt to the local pixel value distribution, Budzan and Pawełczyk [7]. Zhang et al. [8] proposed an effective approach for particle segmentation based on combining the background difference method and the graph cut-based local threshold method. Building the background model and performing subtraction between particle image and background model are performed to eliminate the droplets, while the local threshold method is performed further to eliminate the influences of particle shadow.

Deep neural networks could be a powerful solution to solve the above-mentioned problems in image segmentation, because it has been successfully applied in computational vision, He et al [9] and Bolya et al [10]. Despite the high processing required to implement convolutional neural networks, we see that the cited works of Fast/Faster R-CNN brought an improvement in data processing speed, allowing an application to achieve real-time speed rates, and later, allowing the advancement of computer vision techniques for Mask R-CNN (one of the main neural network models for object segmentation). In parallel to Mask R-CNN, YOLACT comes as a model for instance segmentation with a proposal to be even faster and yet with segmentation accuracy equivalent to that of Mask R-CNN.

In this work, two deep neural networks for pellet segmentation, Mask R-CNN, He et al [9] and YOLACT Bolya et al [10], will be used and compared as a basis for detecting iron ore pellets, allowing the measurement and tracking of pellet diameter growth. To implement the neural models, the algorithms for each of the networks were prepared in Python language, and the datasets, a bank of annotated images used to train the models, were created for Mask R-CNN and YOLACT. The results of the tests show a closeness in the accuracy of the two models, and a superiority of YOLACT in processing speed.

## **2 Materials and Methods**

This chapter will discuss the theoretical bases used in the development of the work. Such bases are necessary for the best use of the work and the understanding of the methods and steps described.

Initially, the Mask R-CNN and YOLACT networks used in this work to perform the segmentation of the pellets will be presented. Finally, the pelletizing process will be briefly presented.

### **2.1 Mask R-CNN**

Before describing the Mask R-CNN it is important to mention some frameworks that boosted advances in object detection and semantic segmentation as Fast R-CNN presented by Girshick [11], Faster R-CNN presented by Ren et al. [12]. They are intuitive, flexible and robust frameworks, in addition to fast training and inference time.

One of the main stages of Mask R-CNN is semantic segmentation, which combines classic computer vision tasks in object detection. Its objective is to classify individual objects and locate each one using a bounding box and semantic segmentation, where the objective is to classify each pixel in a fixed set of categories without differentiating object instances. This challenging task requires a complex method to obtain good results.

Mask R-CNN extends Fast R-CNN by adding a branch to predict segmentation masks in each region of interest (RoI), parallel to the existing branch for classification and boundary box regression (Figure 1). The mask branch is a small FCN - Full Connection Network applied to each RoI, providing a segmentation mask pixel by pixel. Mask R-CNN is simple to implement and train, given the structure of the Faster CNN framework, which facilitates a wide variety of flexible architecture projects. In addition, the mask branch adds only a small computational overhead, allowing for a quick system and quick experimentation.

In order to preserve the exact spatial locations, Mask R-CNN proposes a simple layer of and without quantization, called RoiAlign. This change, although simple, improves the accuracy of the mask by 10% to 50%, showing greater gains with more rigorous location metrics. Another change was to decouple the mask and class prediction and predict a binary mask for each class independently, without competition between classes. To predict the category, the RoI classification branch of the network is used. On the other hand, FCNs usually perform a categorization of several classes per pixel, which combines segmentation and classification, and based on our

experiments it works poorly, for example, segmentation.

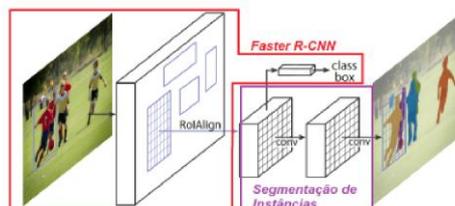


Figure 1. The MASK R-CNN structure for instance segmentation.

The Mask R-CNN surpasses all previous results of a unique model of the art in the task of instance segmentation in the competition using the COCO dataset, data set with the objective of advancing the state of the art in object recognition, presented by Lin et al. [13], including the highly designed entries from the 2016 competition winner. As a by-product, our method also excels in the COCO object detection task. In ablation experiments, we evaluate several basic instantiations, which allows us to demonstrate their robustness and analyze the effects of the main factors.

Mask R-CNN models can run on a GPU for about 200ms per frame, and training using the COCO dataset takes one to two days on a machine with 8 GPUs. The fast training and testing speeds, together with the flexibility and precision of the structure, will benefit and facilitate future research on instance segmentation. The system proposed in this article, its application and the results obtained are described further ahead.

## 2.2 YOLACT

Analogous to Mask R-CNN to Faster R-CNN, YOLACT aims to add a mask branch to the existing one-stage detector to achieve the purpose of instance segmentation, but it is not desirable to introduce feature positioning steps in this process. YOLACT accomplishes this task by adding two parallel branches: the first branch uses FCN to generate a series of prototype masks independent of a single instance; the second branch adds additional headers to the detection branch to predict the mask coefficients to use to encode the representation of an instance in the prototype mask space. Finally, after the NMS step, the final prediction result is obtained by linearly combining the output results of the two branches. The network structure of YOLACT is shown in Figure 2:

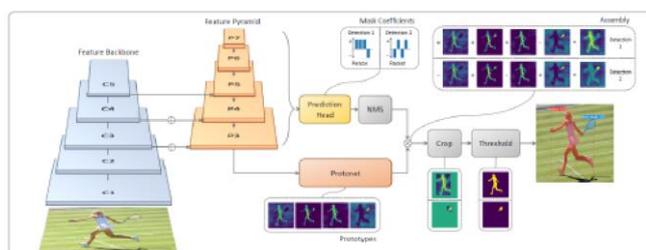


Figure 2: YOLACT's network structure

Because the goal of the segmentation task is to get the mask, and the feature of the mask is that there is a natural spatial connection, YOLACT adopts the above organization form. From the perspective of NN, the Conv layer naturally uses spatial correlation, but the FC layer does not. This leads to a problem, because most One-stage detectors predict box parameters and categories through the FC layer. Two-stage retains spatial information through feature positioning steps such as ROI Align, and uses the Conv layer to output the mask, but these operations must wait for the RPN to complete, which greatly affects efficiency. In YOLACT, the FC layer is responsible for predicting semantic tags, and the Conv layer is responsible for predicting the prototype mask and mask coefficients. The two branches are in parallel, and finally assembled by matrix multiplication, which not only preserves the spatial correlation, but also maintains the One-stage model structure, which is extremely fast.

Loss is composed of classification loss, box regression loss and mask loss. The classification loss and box regression loss are the same as SSD, and the mask loss is the pixel-by-pixel binary cross entropy of the predicted mask and ground truth mask. In order to improve the segmentation effect of small targets, YOLACT cuts the

Mask. During inference, it will first cut according to the detection frame and then threshold. During training, the ground truth box is used for cropping, and the loss scale is balanced by dividing by the area of the corresponding ground truth box.

Because predicting a set of prototype masks and mask coefficients is a relatively difficult task and requires richer and more advanced features, YOLACT hopes to balance speed and feature richness in network design. Therefore, the design of YOLACT's backbone detector follows the idea of Retina-Net, while paying more attention to speed. YOLACT uses ResNet-101 combined with FPN as the default backbone network, and the default input image size is 550×550. Compared with the original Retina-Net, the design of the YOLACT detection head (as shown below) is lighter and faster.

### 2.3 Pelletizing Process

Pelletizing is an agglomeration process with the objective of aggregating the ore fines (smaller than 0.15 mm) into balls with appropriate granulometry and quality - the pellets - for direct use in the steelmaking process.

In order to produce uniform and good quality pellets, it is necessary to take into account the wide variety of properties of the ores, such as their mineralogy, particle size and shape, crystal habit and chemical composition. Although, at present, the different properties of the ores can be compensated for, the parameters of the pelletizing process must be varied and selected according to the type of ore involved.

In the pelletizing stage it is necessary to obtain a particle size that meets customer needs, along with other quality characteristics. Such characteristics can be abrasion, compression, ore content, among others.

According to Fonseca e Campos [14], with greater particle size homogeneity, there tends to be lower fuel consumption, which reduces operating costs.

According to Matos [15], heterogeneous particle size negatively affects the permeability of the kiln.

To adjust the granulometry, it is necessary to manipulate some component variables of the process and pay attention to the impacts that may occur in other characteristics, i.e., the search for improved granulometry may interfere with other quality characteristics, which can be harmful to the process and the product. Therefore, one must understand the behavior of the process and manipulate each variable correctly.

## 3 Methodology

A prototype of Computer Vision was developed to perform online the detection, measurement, validation and analysis of the pellets inside the pelletizing disc as shown in Figure 3.

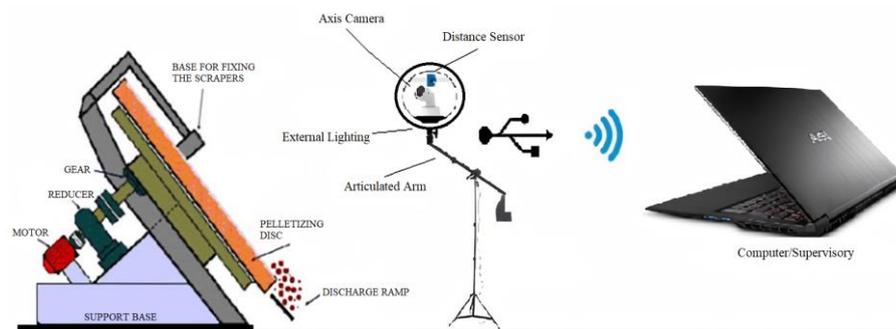


Figure 3: Computer Vision System Implemented in Pellet Disk.

The prototype developed has a fixing structure with its own lighting and adaptation to an articulated arm with camera and distance sensor. The communication between the computer and the distance sensor takes place via serial and with the camera via TCP/IP in its own network. The computer runs the supervision service to manage the segmentation of the pellets on the disk and thus measure the diameters.

The steps for test implementation of the neural models used were: Algorithm development; Preparation of data sets (datasets); Conducting the training of models; Validation and testing of models. Within these steps, it is worth emphasizing the importance of the dataset preparation step. For a network to have a good precision in object segmentation a robust data is needed, with thousands of images, in addition to intense training, which can take a lot of time and consume a lot of processing. Therefore, the transfer of learning technique was used, where, instead

of training the network from scratch, the training is performed from previously generated weights to another dataset, in this case, the COCO dataset (LIN [16]).

## 4 Results

The comparison between the accuracy values achieved in the models can be seen in Table 1. It shows that for low IoU values YOLACT has higher accuracy values, but for higher IoU values Mask R-CNN generally does better.

Table 1: mAP presented by Mask R-CNN an Yolact

Model	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95	Total
Mask R-CNN	85.18	84.67	83.26	82.12	77.34	64.97	43.16	14.20	1.28	0.00	53.62
Yolact	89.4	88.31	86.79	82.81	74.86	58.81	32.46	10.42	1.32	0.01	52.53

As a way of comparing the two networks, a set of images was defined where the pellets must be segmented and, based on these detections, measure the diameter (in pixels) of each one of them. Thus, the three images were chosen for a comparison between the proposed models, identified in Figure 4.

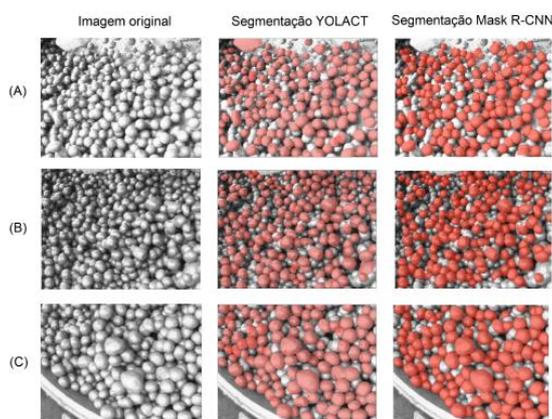


Figure 4: Result of image segmentations A, B and C for Yolact and Mask R-CNN

It can be seen from Figure 4 that both YOLACT and Mask R-CNN were able to perform the measurements satisfactorily, maintaining very close results.

To prove the effectiveness in measuring the pellets against the actual size in millimeters, a sample of pellets was separated (Figure 5) and with the help of a digital caliper, manual measurement (in millimeters) of these pellets was performed. To perform the pixel/millimeter conversion, a first generation 25 cent coin with a known size of 23.5mm was used, so it is sufficient to count the diameter of the coin in pixel in the image used (425 pixels in diameter) and, from there, calculate the ratio of millimeters per pixel (Rel) in the image, according to Equation 1:

$$Rel = \frac{\text{Medida em mm}}{\text{Medida em pixel}} = \frac{23,5}{425} = 0.05529 \quad (1)$$



Figure 5: Separate pellets for measurement identified by numbering 1-12.

In addition to this the measurement from the Mask R-CNN and Yolact pixel segmentations was also performed (Figure 6).

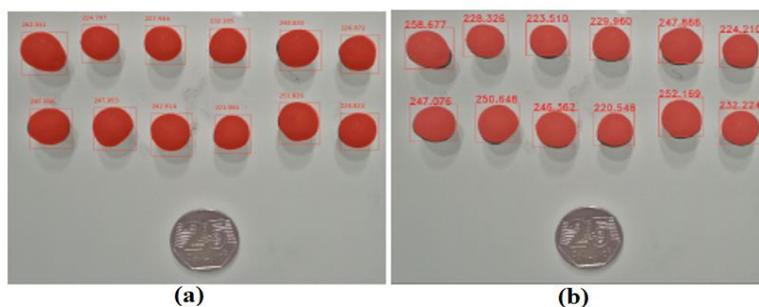


Figure 6: Pixel measurements on the image of the separated pellets.(a) Mask R-CNN, (b)Yolact

With the ratio in hand, a simple multiplication of the diameter obtained in pixel by the ratio will result in the diameter of each pellet measured by the neural models, in millimeters. Table 2 shows the comparison of the results obtained by manually measuring and imaging the pellets.

Table 2: Comparison of Measurement Results

	Manual Measurement	Mask R-CNN	Yolact
Average diameter (mm)	13.4891	13.1841	13.1833
Standard Deviation	0.7521	0.6879	0.7011
Relative error		2.2610%	2.2672%

Analogously to the image containing 12 pellets tested earlier, another image with a larger number of pellets was separated and segmented, this time the pellets were arranged in a way to simulate a real arrangement of the pellets on a disk, i.e., with pellets overlapping each other, where only the pellets in the foreground should be segmented (Figure 7). Table 3 shows a higher error in the measurement of the individual pellets, however, when looking at the average value of the segmentations the error remains low with respect to the real measurements. However, it is now seen that Mask R-CNN loses in accuracy to YOLACT.

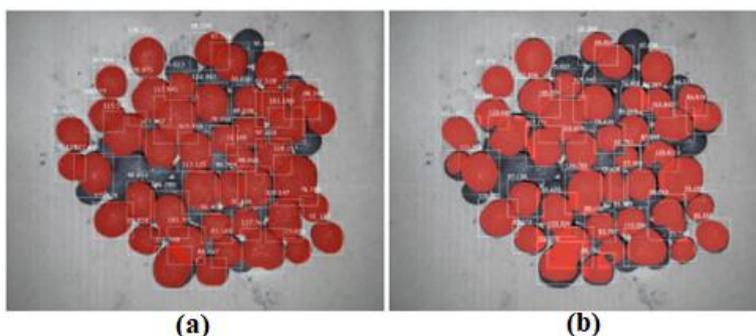


Figure 7: Pixel measurement on the image of the overlapping pellets, (a) Mask R-CNN, (b) Yolact

Table 3: Comparison of Measurement Results (overlapping pellets)

	Manual Measurement	Mask R-CNN	Yolact
Average diameter (mm)	13.53	13.09	13.19
Standard Deviation	2.0003	1.8722	2.3752
Relative error		3.2563%	2.5376%

Naturally, the speed of the network segmentation varies for each image depending on the number of object instances recognized in them. Table 4 presents the average speeds for prediction and segmentation in the models for images with 1 and 200 segmented instances, using 800x600 resolution images on a GTX 1660 GPU.

Table 4: Speed of segmentation.

Model	1 object	200 objects
Mask R-CNN	1.57 FPS	0.585 FPS
Yolact	8 FPS	5 FPS

## 5 Conclusions

Particle size measurement in the formation of iron ore disk pellets is of great importance to the pelletizing process. It is from this sensing that the controllers make the necessary decisions to control the process, varying attributes such as disc rotation speed, inclination, material injection, etc.

Two artificial neural network models for object detection and segmentation (YOLACT and Mask R-CNN) were implemented to perform grain size measurement of iron ore pellets. The results obtained show that YOLACT is able to achieve measurements very close to those obtained by Mask R-CNN with a significant gain in speed. This gain is very important for the implementation of the algorithm in real time.

On an NVIDIA GeForce GTX 1660 GPU, while the model known as Mask R-CNN achieves an average speed of 0.585 frames per second for images with 200 pellet instances and 800x600 resolution, YOLACT achieves an average speed of 5 frames per second for the same images, a gain of more than 8x in processing speed.

The Mask R-CNN model showed a total mAP value of 53.62% and an error in the measurement in millimeters on the test pellets of 2.2610% relative to manual measurement, while the YOLACT showed a total mAP of 52.53% and error relative to manual measurement of 2.2672%. The closeness in the measurements of the two models is evident.

For future work, it is suggested to study in depth the effects of the various variables of the pelletizing process on pellet diameter and, in this way, implement a specific controller for the process using the measurements from the models presented.

## References

- [1] X. Liu, C. Mao, W. Sun and X. Wu, "Image-based method for measuring pellet size distribution in the stable area of disc pelletizer". ISIJ International, Advance Publication by J-Stage, 2018.
- [2] C.W. Liao, Y.S. Tarnq, "On-line automatic optical system for coarse particle size distribution". Powder Technol 189:508-513, 2009.
- [3] P. Facco, A.C. Santomaso and M. Barolo, "Artificial vision system fro particle size characterization from bulk materials". Chem Eng. Sci 164:246-257, 2017.
- [4] M. Heydari, R. Amirfattahi, B. Nazari and P. Rahimi, "An industrial image processing-based approach for estimation of iron ore green pellet size distribution". Powder Technol, 303:260-268, 2016.
- [5] V. Subramanyam, P. Patra and M.K. Singh. "Automatic image processing based size characterization of green pellets". Int J Autom Smart 7(3):85-91, 2017.
- [6] M.M. Roozbahani, R. Borela and J.D. Frost. "Pore size distribution in granular material microstructure". Materials 10(11):1237-1257, 2017.
- [7] S. Budzan and M. Pawelczyk. (2018). "Grain size determination and classification using adaptive image segmentation with grain shape information for milling quality evaluation". Diagnostyka 19(1):41-48, 2018.
- [8] B. Zhang B, A. Abbas and J.A. Romagnoli. "Multi-resolution fuzzy clustering approach for image-based particle characterization for particle systems". Chemom Intell Lab Syst 107(1):155-164, 2011.
- [9] K He, G. Gkioxari, P. Dollár and R. Girshick. "Mask R-CNN". In: The IEEE International Conference on Computer Vision (ICCV). [S.l.: s.n.], 2017.
- [10] D. Bolya, C. Zhou, F. Xiao and Y.J. Lee. "Real-time Instance Segmentation". The IEEE International Conference on Computer Vision (ICCV). [S.l.: s.n.], 2019.
- [11] R. Girshick, Ross. "Fast R-CNN". In: The IEEE International Conference on Computer Vision (ICCV). [S.l.: s.n.], 2015.
- [12] S. Ren, K. He, R. Girshick and J. Sun. "Faster R-CNN": Towards Real-Time Object Detection with Region Proposal Networks. 2015.
- [13] T.Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie. "Feature Pyramid Networks for Object Detection". 2016
- [14] V.O. Fonseca, A.R. Campos, "Aglomeración: Pelotização". In: Tratamento de Minérios. 6. ed. Rio de Janeiro: CETEM/MCTIC, cap. 15.2, p. 677-724, 2018.
- [15] A.P. MATOS, "Influência da Temperatura, Pressão, Produção e Granulometria no Processo de Secagem das Pelotas Cruas". Dissertação De Mestrado, REDEMAT, Ouro Preto, 2007.
- [16] LIN, Tsung-Yiet al. Microsoft COCO: Common Objects in Context. 2014.