Segmentation and Classification of Criollo Horses using Deep Learning

Guilherme Veiga Santos Pinto
Universidade do Vale do Itajaí, Brasil
guilhermeveigarj@edu.univali.br

Douglas Rossi Melo
Universidade do Vale do Itajaí, Brasil
drm@univali.br

Eros Comunello
Universidade do Vale do Itajaí, Brasil
eros.com@univali.br

Anita Maria Rocha Fernandes
Universidade do Vale do Itajaí, Brasil
anita.fernandes@univali.br

Marcelo Dornbusch Lopes
Universidade do Vale do Itajaí, Brasil
marcelo@univali.br

ABSTRACT

The Criollo horse, one of the leading horse breeds in the south of Brazil, has been moving the market with the demand for sports horses. The rise of the breed in recent years was motivated by the Brazilian Association of Criollo Horse Breeders (ABCCC), with the mission of preserving and spreading the breed in the country in conjunction with competitions such as the Freio de Ouro. The demand for breed optimization follows the animal's morphological balance, determined by its entire bone and muscle structure characteristics that directly impact its health and performance as an athlete horse. Computer vision has been solving several current problems with the deep training of convolutional neural networks. Within this context, this work proposed to evaluate the average precision of the Mask R-CNN model for the detection and segmentation of Criollo horses through training this neural network model with images collected from Google's search engine. The best results achieved up to 99.4% average precision for detecting horses in the images and 89.1% accuracy for the segmentation task.

KEYWORDS

Criollo Horse, Segmentation, Classification, Deep Learning

1 INTRODUCTION

The Criollo horse originated with the Andalusian and Berber breeds, brought from the Iberian Peninsula by the settlers in the 16th century. These horses spread throughout the different regions of South America and underwent adverse temperature and feeding conditions, resulting in striking characteristics that define today the Criollo breed [1].

The Criollo breed has been on a constant rise in recent years. The Brazilian Association of Criollo Horse Breeders (ABCCC) [1] preserves and spreads the breed in the country along with important competitions such as the Freio de Ouro [2], which evaluates the morphological and functional optimization of the Criollo horse.

However, there are several morphological characteristics to be evaluated in these horses. One of these characteristics is its balance, based on bone and muscle structure that directly affects the functional movements of the horse, impacting its health as an athlete horse [2].

This feature can benefit from techniques in computer vision, such as Deep Learning in Artificial Neural Networks, allowing solutions that recognize and classify objects, people, and animals. These networks can propose solutions to the most diverse areas, such as facial recognition [3], autonomous cars [4], and even the detection of cancer [5]. In the equine area, some studies make the recognition of different breeds [6], periocular recognition [7], real-time recognition [8], and even zoometric measurements [9].

This work sought to evaluate the Average Precision of the Mask R-CNN model for detecting and segmenting Criollo horses using segmentation and deep learning.

This paper is organized as follows. Section 1 presented an introduction to the topic. In Section 2, the background and related works are presented. Section 3 presents the materials and methods used to develop this work. The results obtained are presented in Section 4. Finally, Section 5 presents the final considerations.

2 BACKGROUND

The background of this work is related to the Criollo horse and its characteristics, morphological balance, and computer vision techniques, such as deep learning in deep neural networks.

2.1 Criollo Horse and Morphological Balance

According to the Criollo Breeder's Manual [1], for a horse to be considered as belonging to this breed, it must meet a measurement standard. This standard for males consists of a height between 1.40 and 1.50m, a minimum chest perimeter of 1.68m, and a minimum shin perimeter of 0.18m. For females, these measurements are height between 1.38 and 1.50m, minimum chest perimeter of 1.70m, and minimum shin perimeter of 0.17m.

In competitions, such as the Freio de Ouro, during the morphological tests, the judges assess the competing horses according to the main characteristics that make up the Criollo breed seal, in addition to an assessment concerning their morphological balance [2].

One of the ways to assess the horse's balance is the Trapezoid Theory. According to the American Quarter Horse Association (AQHA) [10], the Trapezoid Theory consists of visualizing the figure of an isosceles trapezoid in the horse's silhouette, as can be seen in Figure 1. For this visualization, it is necessary to carry out the following steps: position the horse on a flat level, where the distribution of weight between the forelegs and hindquarters is uniform and without inclinations; imagine a line parallel to the ground running down the horse's back, connecting the withers and the croup; imagine another line parallel to the ground extending from the tip of the horse's chest to its hindquarters; and complete a symmetrical isosceles trapezoid over his body, connecting one line from the withers to the tip of the shoulder and another from the loin to his back.

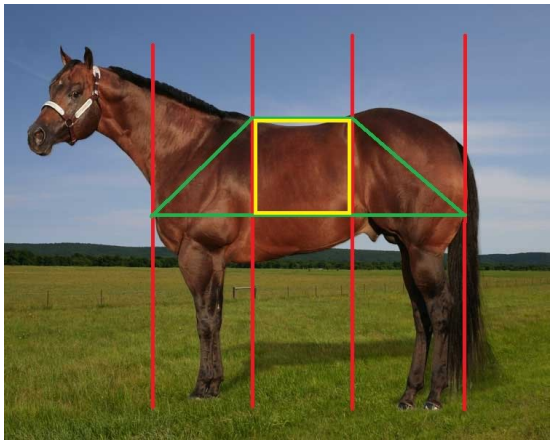

Figure 1: Trapezoid on the body of a horse [11]

The Trapezoid Theory concerns how a specific part of the horse relates to the general structure of its body. When parallel to the ground, the line of the horse's back can indicate whether the horse has higher withers, a higher rump, or is appropriately leveled. In addition, the length of the dashed line on the horse's back should be approximately one-third or half of the line on the chest [12].

According to the AQHA [10], the trapezoid also helps to assess the inclination of the horse's shoulder and hip. The angle formed between the line along the shoulder and the body line should be the same as the angle formed by the pelvis and the body line, ideally with a 45° inclination.

The trapezoid helps to check that the horse is similar in size and slopes between the shoulder and hip and has the correct back length with its body. These points contribute to the horse's stronger upper line, facilitating balance. The more balanced the animal is, the greater its potential for athletic ability and good movement [10].

2.2 Semantic Segmentation

According to Wang et al. [13], semantic segmentation consists of assigning a categorical label to each pixel found in an image, playing an essential role in systems for detecting and classifying people or objects in an image. The recent success of convolutional neural networks has provided remarkable progress for pixel-wise semantic segmentation tasks due to advanced hierarchical features and trainable end-to-end structures.

Most systems that use semantic segmentation have three main components: a fully convolutional network (FCN) that replaces the last layers of a network with convolutional layers; conditional random fields (CRFs) for capturing local and long-range dependencies within an image to refine the network's prediction map; and dilated convolution (DC), responsible for increasing the resolution of intermediate resource maps to generate more accurate predictions while maintaining the same computational cost [13].

Semantic segmentation can be implemented using image annotation tools, such as Labelme [14], which allows one to demarcate regions in images and classify them with a label, thus creating a segmentation mask.

2.3 Deep Learning

Deep learning can be defined as a machine learning technique that uses deep neural networks. This technique was used to solve problems that previous neural networks had, such as gradient dissipation, overfitting, and the high computational load for training. These problems were solved through deep learning with the ReLU (Rectified Linear Unit) activation function, and with GPUs for algorithm training, due to their high capacity to perform several calculations [15].

Convolutional neural networks (CNNs) are deep neural networks that use deep learning and resemble how the visual cortex of the human brain processes and recognizes images. The functioning of these networks is done through a convolution layer and another pooling layer. Upon receiving an input image, the convolution layer passes a filter through the entire image and generates a feature map, accentuating the original image's unique features. The pooling layer is responsible for reducing the image size by combining the neighboring pixels of a specific region into a single representative value selected by a square matrix that goes through the entire image. This process is beneficial to alleviate the computation load and avoid overfitting [16].

2.4 Related Work

We looked for works that addressed Criollo horses or horses in general. We also sought to find works related to equine morphology, segmentation, and deep learning in neural networks.

Atabay [6] presented a solution for detecting different horse breeds using deep learning in convolutional neural networks for image classification. The work uses a dataset containing 1,693 images of 6 different horse breeds and uses the learning transfer in architectures with 16 and 19 layers, InceptionV3, ResNet50, and Xception. The architecture with the best results was the ResNet50, reaching an average accuracy of 95.90% confidence.

The work [7] used semantic segmentation and deep convolutional neural networks for periocular and iris recognition in Arabian racehorses. With a dataset of approximately 2,000 images and training in HorseNet-4 architecture, the validation was performed in two tests: one comparing the results in two types of samples and another comparing unilateral recognition with bilateral eye recognition. The first dataset showed the best results with an error rate of 12.7% against 14.4% for the full dataset. In the second test, the bilateral recognition dataset obtained results with an error rate of 10.9%. In the end, a combination was made between the two datasets with the best results, reaching an error rate of only 9.5%.

Delgado [8] developed a tool for real-time detection of horses in stables. The convolutional neural network was trained using the YOLO (You Only Look Once) framework for object detection using a dataset containing 10,000 images of horses in several positions. As a result, tests on the Darknet53 and Yolov3-tiny architectures showed an accuracy of up to 95% confidence.

In [9], an approach to obtain zoometric measurements of a horse's body based on digital three-dimensional modeling is proposed. The data capture uses a LiDAR sensor, in which 16 laser beams fully scan the horse, making a 3D reconstruction of its entire side. The work presented a correspondence of 82.5% of the evaluated characteristics.

Freitas et al. [17] present an intelligent system that classifies and segments foods in an image to monitor the diet and nutritional intake. A comparison between FCN, ENet, Segnet, DeepLabV3+, and Mask RCNN image classification and segmentation algorithms was shown. With a dataset containing nine classes and 1,250 images of Brazilian foods, the models were evaluated using the intersection metrics on union, sensitivity, specificity, balanced precision, and positive preset value.

Table 1 presents the comparison between related works. This work focuses on generating an image base containing Criollo horses and horses of other breeds with similar characteristics, performing tests on this base using segmentation and deep learning in convolutional neural networks.

Table 1: Comparative of related work

Work	Uses Segmentation	Trains CNNs	Horse Theme
[6]	No	Yes	Race recognition
[7]	Yes	Yes	Periocular recognition
[8]	No	Yes	Real-time detection
[9]	No	No	Zoometric measurement
[17]	Yes	Yes	No
This Work	Yes	Yes	Detection and segmentation

3 MATERIALS AND METHODS

In this work, the dataset was created and the image segmentation and model training were performed. The Python language was used with the Detectron2 platform [18] together with the PyTorch library [19], which provides a straightforward way to create a variety of object detection and segmentation models using deep learning. The model implemented in this work was processed in a Google Colab environment [20], which allows writing code in Python directly through the browser, with processing in the cloud.

3.1 Dataset Generation

The procedure to generate the dataset was based on how the morphological evaluations of Criollo horses in equine morphology events and tests are performed. It was defined that the dataset should contain images of horses on the side for the photographer, preferably in a plane with no inclinations and entirely still.

The collection of images for the formation of the dataset was carried out in Google's search engine, using search terms: Criollo Horse (Brazilian, Argentinean, Chilean, and Uruguayan); Cattles of Criollo Horses; Morphological Images of Criollo Horses; Criollo Horse Auctions; Expointer Morphology [21]; Best Criollos and Champions of Morphology; Freio de Ouro. Although the search terms focus on Criollo horses, a selection process with the help of an expert was not carried out to determine which horses in the images belong to this breed.

The images were collected using a script responsible for downloading all the resulting images from the search. In the next step, a manual filtering procedure was performed, seeking to remove repeated images, images that did not have a horse in the lateral position in the photograph, and images that were not part of the theme or context of this work.

After this procedure, the images were then separated into two sets for training the algorithm, namely: a dataset called High Resolution, which contains images with dimensions greater than or equal to 912 x 417; and a dataset called Medium Resolution, containing images with smaller dimensions or images with lower resolutions. The criteria to define the dataset the image would fit was according to its width and height dimensions, file size, and resolution. In the end, 275 images were totaled for the High Resolution dataset and 655 for the Medium Resolution dataset.

3.2 Horse Segmentation and Ground Truth

The image annotations were made with the Labelme tool [14], where the regions that contained horses in each image were manually marked. As for rules for the notes, the manes on some occasions, the animal's tail, and eventually a hind limb that was not visible were ignored.

These choices were due to the morphological analysis focusing on the horse's body, seeking to assess its bone and muscle structure. Following this criterion, the annotated images provide a more focused view of these characteristics, as shown in Figure 2.

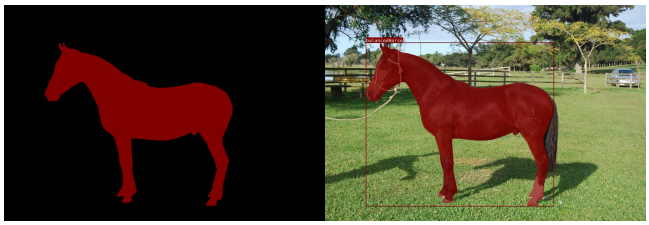

Figure 2: Annotated image segmentation mask

For each annotated image, a JSON (JavaScript Object Notation) file was generated containing the coordinates of the regions that were demarcated, together with a single class (or label), naming and determining what the annotation means. The label defined was *balancedHorse*, where there were one or more occurrences in each image. Then, the JSON files underwent a conversion to COCO (Common Objects In Context) standard dataset through a Python script. It was necessary so that the neural network model could receive the dataset images with their annotations as input.

3.3 Model Training

The model was trained on Google Colab using the Detectron2 platform and the PyTorch library. The neural network used was the Detectron2 Model Zoo Mask [18], which makes use of a 50-layered ResNet [22] with a feature extraction pyramid [23] in the backbone. The training environment was configured to use GPUs in its processing, automatically allocating NVIDIA Tesla P100 and NVIDIA Tesla V100-SXM2 cards with 16GB of memory.

The training was carried out to evaluate three scenarios: training only on the High Resolution dataset; training on Medium Resolution dataset only, and training with the union of the two datasets (Mixed).

We decided to separate the dataset using a configuration of 80% of the images for the training stage, 10% for testing, and 10% for algorithm validation. This separation was done randomly using the Python split-folders library [24], resulting in 219 images for training, 28 for testing, and 27 for validation in the first dataset. For the second dataset, the result was 524 images for training, 66 for testing, and 65 for validation. For the third scenario, with the two datasets mixed, 724 images were obtained for training, and 93 for testing and validation, respectively.

The network was configured to recognize only one class, which is called *balancedHorse*, in one or more occurrences in each image. The training was also carried out using the default configuration of Detectron2 for the batch size. 50,000 and 100,000 iterations were configured for each of the three scenarios.

The Tensorboad tool was used to observe how the model behaved during training. As shown in Figure 3, as the training was carried out, the Mask RCNN model started to converge according to the accuracy. As it approached the final number of iterations, it started to stabilize.

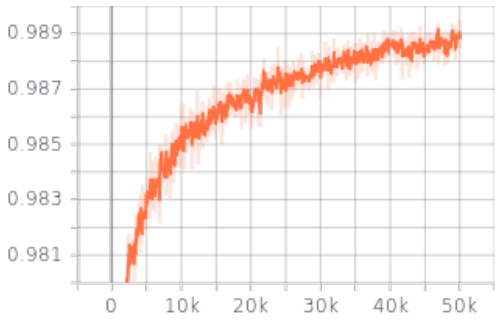

Figure 3: Accuracy of the RCNN Mask in Scenario 1 with 50,000 iterations

Compared with the graph of the same scenario for the training of 100,000 iterations, its convergence behavior remains similar, as seen in Figure 4. For Scenarios 2 and 3, with 50 and 100 thousand iterations, the same convergence of results was observed.

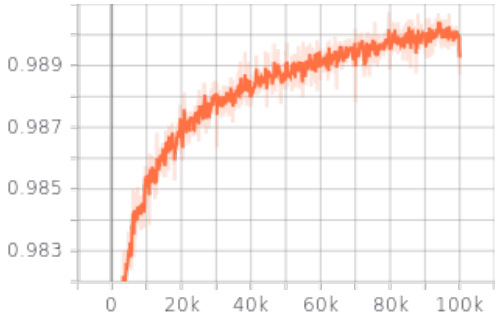

Figure 4: Accuracy of the RCNN Mask in Scenario 1 with 100,000 iterations

The training stage in each scenario had an average duration of 5 to 7 hours in the training of 50,000 iterations using the Tesla P100 GPU. For training with 100,000 iterations, the average duration of training was between 10 and 12 hours with the same GPU model. Only in Scenario 2, with 100,000 iterations, the training had a shorter duration, with a training time of 5 hours and 43 minutes, due to the Tesla V100-SXM2 GPU automatically allocated by Google Colab.

4 RESULTS

For each scenario, the performance of the trained model was evaluated using the metric offered by Detectron2. This evaluation considers the bounding box's accuracy and the segmentation's accuracy. A square is drawn around the object to detect the bounding box. A mask (or bubble) is drawn for the segmentation, marking the pixels detected as belonging to the object. The metric used to evaluate these tasks consists of the Intersection over Union (IoU). An Average Precision (AP) or average recognition accuracy of the algorithm is returned as a result.

As seen in Table 2, regarding training with 50 thousand iterations, Scenario 1 stood out among the others for detecting the horse, where only the High Resolution dataset was used for training. With an Average Precision of 99.3% of accuracy, a difference of 1.8% was obtained in relation to Scenario 2, with the dataset of Medium Resolution, and 0.1% difference in relation to Scenario 3, with the Mixed dataset. As for horse segmentation, Scenario 3 showed better results with an Average Precision of 89.1% accuracy, with a difference of 1.1% for Scenario 1 and 2.1% for Scenario 2.

Table 2: Results for training with 50,000 iterations

	AP (Bounding Box)	AP (Segmentation)
Scenario 1	99.4%	88.1%
Scenario 2	97.6%	87.1%
Scenario 3	99.3%	89.2%

Regarding training with 100,000 iterations, as seen in Table 3, Scenario 3 (Mixed dataset) was highlighted for horse detection, with an Average Precision of 99.4%. We obtained 1.8% for Scenario 1 (High Resolution) and 1.8% difference for Scenario 2 (Medium Resolution). Scenario 3 also remained highlighted in the horse segmentation task with an Average Precision of 89.1%, with a difference of 1.1% for Scenario 1 and 2.4% for Scenario 2.

Table 3: Results for training with 100,000 iterations

	AP (Bounding Box)	AP (Segmentation)
Scenario 1	97.6%	88.1%
Scenario 2	97.6%	86.8%
Scenario 3	99.4%	89.2%

When comparing Tables 2 and 3, it is observed that Scenario 3, trained with 100 thousand iterations, presented better results for the detection of horses in the images, with a difference of 0.1% to Scenario 1, trained with 50 thousand iterations. Scenario 3 was again highlighted for horse segmentation when trained with only 50 thousand iterations, with a difference of 0.1% more accuracy than when trained with twice as many iterations.

For model inference, a new image is returned containing a bounding box around the detected horse in the separate images for testing, together with a segmentation mask drawn across its silhouette. As shown in Figure 5, the model's detection and segmentation prediction showed differences when trained with 50,000 iterations and when trained with 100,000 comparisons.

Figure 5: Original image [25] (High Resolution) in conjunction with Scenario 1 prediction in 50,000 and 100,000 iterations

5 CONCLUSION

The Criollo horse has been quite prominent in the search for more balanced animals and high competitiveness in morphological tests. This work sought to use techniques from the field of computer vision to automate the task of evaluating these horses by training in convolutional neural networks using deep learning.

As a contribution, this work evaluated the Average Precision of the Mask R-CNN model for detecting and segmenting Criollo horses. We tried to detect in which dataset the recognition of these animals would be more accurate from evaluating three training scenarios. For training with 50 thousand iterations, Scenario 1 (High Resolution) presented better results for horse detection, while Scenario 3 (Mixed) presented better results in the segmentation task. In training with 100,000 iterations, Scenario 3 was highlighted for both tasks, presenting the best results for the horse's detection and segmentation.

For future work, we intend to perform further tests on these scenarios, modifying the model's training settings, such as the number of batch sizes and total iterations. It is also interesting to look for a solution to publicize this image base and submit it to an equine specialist for segmented regions analysis.

ACKNOWLEDGMENTS

This project was supported by the University of Vale do Itajaí (UNIVALI), and the National Council for Scientific and Technological Development (CNPq).

REFERENCES

[1] ABCCC. *Manual do criador: Raça Crioula.* Associação Brasileira de Criadores de Cavalos Crioulos, 2016. URL http://www.cavalocrioulo.org.br/admin/assets/upload/manuais/manual.pdf.

[2] Anelise Maria Hammes Pimentel. Associação da biometria no desempenho morfo funcional no cavalo crioulo participante do freio de ouro, 2016.

[3] Guosheng Hu, Yongxin Yang, Dong Yi, Josef Kittler, William Christmas, Stan Z. Li, and Timothy Hospedales. When face recognition meets with deep learning: An evaluation of convolutional neural networks for face recognition. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, pages 142–150, Parque Araucano, Las Condes, Chile, December 2015. IEEE.

[4] Mike Daily, Swarup Medasani, Reinhold Behringer, and Mohan Trivedi. Self-driving cars. *Computer*, 50(12):18–23, 2017.

[5] Zilong Hu, Jinshan Tang, Ziming Wang, Kai Zhang, Ling Zhang, and Qingling Sun. Deep learning for image-based cancer detection and diagnosis- a survey. *Pattern Recognition*, 83:134–149, 2018.

[6] Habibollah Agh Atabay. Deep learning for horse breed recognition. *The CSI Journal on Computer Science and Engineering*, 15(1):45–51, 2017.

[7] Mateusz Trokielewicz and Mateusz Szadkowski. Iris and periocular recognition in arabian race horses using deep convolutional neural networks. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 510–516, Denver, CO, USA, 2017. IEEE, IEEE.

[8] Alejandro Juan Delgado Sanchis. Detecting and tracking horses using deep neural networks. Master's thesis, Universitat Politècnica de València, 2019. URL https://riunet.upv.es/handle/10251/123523.

[9] Manuel Pérez-Ruiz, D Tarrat-Martín, María José Sánchez-Guerrero, and M Valera. Advances in horse morphometric measurements using lidar. *Computers and Electronics in Agriculture*, 174:105510, 2020.

[10] AQHA. The trapezoid theory: This tool can help you evaluate a horse's balance., 2018. URL https://www.aqha.com/pt/-/the-trapezoid-theory.

[11] Frederick County 4-H. Horse judging, 2019. URL https://www.frederickco4h.com/horse-judging.html.

[12] Kylee Jo Duberstein. Evaluating horse conformation. *UGA Cooperative Extension Bulletin*, 2012. URL https://extension.uga.edu/publications/detail.html?number=B1400&title=Evaluating%20Horse%20Conformation.

[13] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell. Understanding convolution for semantic segmentation. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1451–1460, Los Alamitos, CA, USA, mar 2018. IEEE Computer Society.

[14] Kentaro Wada. labelme: Image Polygonal Annotation with Python. https://github.com/wkentaro/labelme, 2016.

[15] Phil Kim. Deep learning. In *MATLAB Deep Learning*, pages 103–120. Springer, Berkeley, CA, USA, 2017.

[16] Phil Kim. Convolutional neural network. In *MATLAB deep learning*, pages 121–147. Springer, Berkeley, CA, USA, 2017.

[17] Charles NC Freitas, Filipe R Cordeiro, and Valmir Macario. Myfood: A food segmentation and classification system to aid nutritional monitoring. In *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 234–239, Los Alamitos, CA, USA, 2020. IEEE, IEEE Computer Society.

[18] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019.

[19] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32:8026–8037, 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[20] Google. Welcome to colaboratory - google research. URL https://colab.research.google.com/notebooks/.

[21] ABCCC. Tudo o que você precisa saber sobre a morfologia da expointer 2020, September 2020. URL https://www.cavalocrioulo.org.br/noticias/detalhes/135834/tudo-o-que-voc-precisa-saber-sobre-a-morfologia-da-expointer-2020.

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, Caesars Palace, Las Vegas Valley, Nevada, USA, 2016. IEEE.

[23] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, Hawaii Convention Center, Honolulu, Hawaii, USA, 2017. IEEE.

[24] Johannes Filter, Marius Mézerette, and Marc P. Rostock. split-folders. https://github.com/jfilter/split-folders, 2018.

[25] Crioulo Remates. Leilão ano 10: Cavalo crioulo na vitrine do dc, January 2010.