Letter

Deep Learning Applied to Phenotyping of Biomass in Forages with UAV-Based RGB Imagery

Wellington Castro [1], **José Marcato Junior [2,*]**, **Caio Polidoro [1]**, **Lucas Prado Osco [3]**, **Wesley Gonçalves [2]**, **Lucas Rodrigues [1]**, **Mateus Santos [4]**, **Liana Jank [4]**, **Sanzio Barrios [4]**, **Cacilda Valle [4]**, **Rosangela Simeão [4]**, **Camilo Carromeu [4]**, **Eloise Silveira [2]**, **Lúcio André de Castro Jorge [5]** and **Edson Matsubara [1]**

[1] Faculty of Computer Science, Federal University of Mato Grosso do Sul, Campo Grande 79070900, MS, Brazil; wellingtonvcastro@gmail.com (W.C.); caio.polidoro@aluno.ufms.br (C.P.); lucas.rodrigues@ifms.edu.br (L.R.); edsontm@facom.ufms.br (E.M.)

[2] Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande 79070900, MS, Brazil; wesley.goncalves@ufms.br (W.G.); eloiseambiental@gmail.com (E.S.)

[3] Faculty of Engineering, Architecture and Urbanism, University of Western São Paulo, Presidente Prudente 19067175, SP, Brazil; lucasosco@unoeste.br

[4] Embrapa Beef Cattle, Brazilian Agricultural Research Corporation, Campo Grande 79106550, MS, Brazil; mateus.santos@embrapa.br (M.S.); liana.jank@embrapa.br (L.J.); sanzio.barrios@embrapa.br (S.B.); cbdovalle@gmail.com (C.V.); rosangela.simeao@embrapa.br (R.S.); camilo.carromeu@embrapa.br (C.C.)

[5] Embrapa Instrumentation, São Carlos 13560970, SP, Brazil; lucio.jorge@embrapa.br

* Correspondence: jose.marcato@ufms.br

check for updates

Abstract: Monitoring biomass of forages in experimental plots and livestock farms is a time-consuming, expensive, and biased task. Thus, non-destructive, accurate, precise, and quick phenotyping strategies for biomass yield are needed. To promote high-throughput phenotyping in forages, we propose and evaluate the use of deep learning-based methods and UAV (Unmanned Aerial Vehicle)-based RGB images to estimate the value of biomass yield by different genotypes of the forage grass species *Panicum maximum* Jacq. Experiments were conducted in the Brazilian Cerrado with 110 genotypes with three replications, totaling 330 plots. Two regression models based on Convolutional Neural Networks (CNNs) named AlexNet and ResNet18 were evaluated, and compared to VGGNet—adopted in previous work in the same thematic for other grass species. The predictions returned by the models reached a correlation of 0.88 and a mean absolute error of 12.98% using AlexNet considering pre-training and data augmentation. This proposal may contribute to forage biomass estimation in breeding populations and livestock areas, as well as to reduce the labor in the field.

Keywords: Convolutional Neural Network; biomass yield; data augmentation; phenotyping

1. Introduction

Monitoring crop parameters like nutrient content, biomass, and plant height is essential for yield prediction and management optimization [1]. In situ measurements of these parameters can be a time-consuming, expensive, and biased task. To assist in plant breeding programs [2] as well as in precision agriculture practices [3], remote sensing technologies have been used in multiple approaches [4–6], and, lately, this has expanded with the implementation of UAV (Unmanned Aerial Vehicle) based-data. In recent years, UAV-based images, in conjunction with robust and intelligent data

processing methods, are being used as an alternative to the human-visual inspection of agricultural landscapes [7].

The use of machine learning methods [8], such as RF (Random Forest) and SVM (Support Vector Machine) in remote sensing applications has been increasing in the last few years, even in precision farming approaches [9–11]. Deep learning is a type of machine learning technique that automatically extracts features, and uses a deeper neural network with hierarchical data representation [12]. This technique uses different structures to perform multiple tasks, and, recently, Convolutional Neural Networks (CNNs) [13] are considered as one of the most important architectures to image and pattern recognition, especially regarding remote sensing related tasks [14,15].

Deep learning was applied in diverse scenarios with crops such as citrus-trees [11,16], canola [17], rice [18], corn [19], bean and spinach [20], and others. There is still a lack of deep learning-based applications in forage crops, specifically using remote sensing data. Although many deep neural networks were proposed to object detection and segmentation in different agriculture practices, few studies implemented this evaluation for biomass estimate for pasture fields. Biomass yield is the main trait evaluated in forage breeding programs, and selection gains can be increased by strategies that allow high-throughput phenotyping in larger breeding populations. In livestock farms, accurately measuring biomass is relevant since the profitability can be strongly affected by it [21].

In agriculture-related problems, Osco et al. [4] estimated nitrogen content in citrus-trees using multispectral UAV-based imagery. A study [22] was able to model water-stress in a vineyard using hyperspectral UAV-based imaging. Another approach [9] proposed a new technique with UAV-based hyperspectral imagery to detect citrus canker disease. Regarding plant detection, a study [23] proposed a method to identify tobacco plants in UAV-RGB-based images automatically. Most of these practices were beneficiated from the usage of machine learning [4,9,22] and deep learning [23] methods combined with UAV data, mainly because of the relatively low cost of the equipment and a high capacity to map plantation fields with very high spatial-resolutions.

Regarding biomass estimation, Bendig et al. [1] proposed an approach adopting CSMs (Crop Surface Models), which are estimated based on the difference of the DSM (Digital Surface Model) and the DTM (Digital Terrain Model), both obtained with UAV Photogrammetry. Going forward, Bendig et al. [24] compared CSMs and VIs (Vegetation Indices) approaches to determine biomass. The authors used three visible band (RGB) VIs: the green red vegetation index (GRVI), the modified GRVI (MGRVI), and the red-green-blue VI (RGBVI). They concluded that GRVI and RGBVI are promising as data input, but the CSMs approach based on plant height returned more accurate results. Ballesteros et al. [25] used canopy cover, crop height, and canopy volume to estimate onion biomass. Batistoti et al. [26] also investigated CSM models to estimate biomass in pasture species, achieving satisfactory results.

Strategies based on plant height are limited because they require DTMs, which are 3D models of the terrain (i.e., without the plants covering the surface). The generation of a DTM can be performed using a GNSS (Global Navigation Satellite System)-based survey [26], which is time-consuming, especially in large areas, or flight-based approaches before initializing the planting process [1]. Although it is not possible to provide more accurate results compared to the CSMs approach, RGB VI-based approaches already demonstrated promising results, and Bendig et al. [24] pointed out that further investigation in this matter is necessary.

Nasi et al. [27] investigated the use of RGB and hyperspectral UAV-based data to estimate biomass. They used the machine learning method RF (Random Forest) combining 3D information, RGB, and hyperspectral imageries. The authors achieved the best performance by integrating hyperspectral and 3D features. However, they highlighted that the integration of RGB images and the 3D features also provided excellent results. Li et al. [28] conducted experiments using the RF learner by joining 3D information, RGB, and hyperspectral data for estimating potato plantations biomass. Unfortunately, hyperspectral sensors are expensive, and, in a general sense, they are not available for a wide range of applications.

Previous researches related to our proposal used conventional machine learning methods, which require handcraft features [29]. Although some of them returned good accuracies, there is still a need for more robust models to be proposed in this task. Based on a review analysis, Kamilaris and Prenafeta-Boldú [29] verified that deep learning methods outperformed traditional machine learning in several agriculture applications. In this sense, the estimation of biomass using deep learning is still scarce in the literature. In previous work, Ma et al. [30] assessed a deep learning architecture based on VGGNet to predict the above-ground biomass of winter wheat. The authors used an RGB camera on a terrestrial tripod, which somewhat limits the application in larger areas. Nevertheless, this experiment demonstrated the potential of CNN to ascertain this task and stated that the proposal of novel methods on this theme is still necessary.

To the best of our knowledge, no literature focused on investigating deep learning-based biomass estimation methods using UAV-RGB orthoimages in tropical forages. Approaches using RGB orthoimages seem to be an interesting practice since they have higher spatial resolution compared to images from other types of sensors, and it does not rely on the 3D information of the canopy, reducing the amount of data necessary to perform the said task. The contribution of this study is to propose a deep learning approach to estimate biomass in forage breeding programs and pasture fields using only UAV-RGB imagery and AlexNet and ResNet deep learning architectures. We also compared the results with VGGNet, used in previous work [30] on biomass estimation. The rest of the paper is organized as follows. Section 2 presents both the materials and methods implemented in this study. Section 3 presents and discusses the results obtained in the experimental analysis. Finally, Section 4 summarizes the main conclusions of our approach.

2. Materials and Methods

2.1. Study Area and Dataset

The dataset was formed by images obtained by the UAV DJI Phantom 4 PRO embedded with an RGB digital camera with an image resolution of 5472×3648. The area of the experiment is located in Brazilian Cerrado at the Experimental Station of Embrapa Beef Cattle, Campo Grande, Mato Grosso do Sul, Brazil (Figure 1—latitude $20°26'46''$ S, longitude $54°43'16''$ W) and altitude 535 m. The flight was carried out on 23 January 2019, at around 9 a.m. with a relative height of 18 m, resolving 0.5 cm/pixel. The photos were taken with a frontal overlap of 81% and a lateral overlap of 61%. The orthoimage (Figure 2b) was generated using the Pix4D software based on the SfM (structure-from-motion) and MVS (multi-view stereo) techniques.

The experiment was composed of 110 genotypes representing a high genetic diversity of the species *Panicum maximum* (syn. *Megathyrsus maximus*), an important tropical forage grass for livestock production [31]. These genotypes are grouped by 86 full-sib progenies, ten sexual and ten apomictic progenitors along with four commercial cultivars (Mombaça, MG12 Paredão, BRS Quenia, and BRS Tamani). All the genetic material, with exception of MG12 Paredão, was developed by the *P. maximum* Breeding Program of Embrapa. A randomized complete block design was used with three replications, totaling 330 plots. Each plot consisted of two rows of 2.0 m with 0.5 m apart. Each row contained five plants spaced by 0.5 m between plants, totaling ten plants per plot. Plots were 1.0 m apart, representing an area of 4.5 m². We evaluated total green matter yield trait as ground truth data. Each plot was harvested 0.2 m from the soil, and the green material was weighted in kg·plot^{-1} using a field dynamometer and converted in kg·ha^{-1} on 25 January 2019. For term simplicity, this trait will be named later only as biomass yield.

Figure 2 shows the plot's definition procedure. We developed a python script tool (https://github.com/wvmcastro/tiffviewer) named field plot cropper (FPLOTCROPPER). The inputs of the tool are: the orthomosaic, the number of blocks in the image, the number of lines, and columns within each block, and the user-defined rectangular polygons (Figure 2b). The result can be seen in Figure 2c.
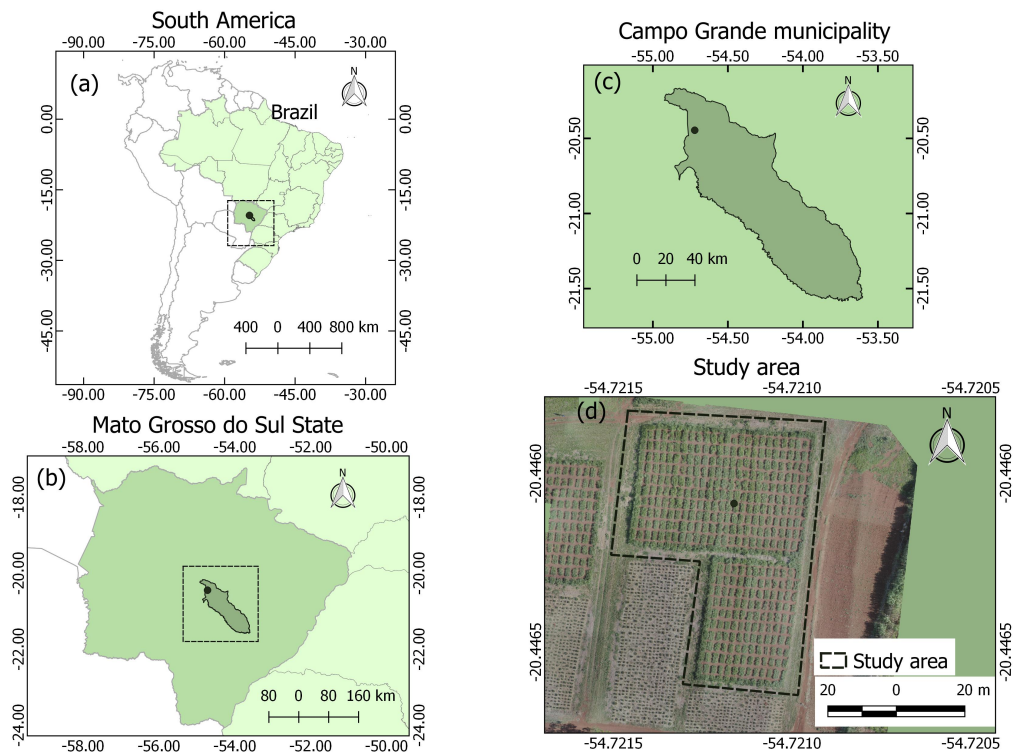
Figure 1. Dataset location considering WGS-84 reference system: (**a**) South America/Brazil; (**b**) Mato Grosso do Sul State; (**c**) Campo Grande municipality and; (**d**) Study area.

Dealing with orthoimages can prove to be a challenging task from the computational point of view, as images of this type can go beyond gigabytes in size. In this regard, the use of specific software to reduce these images is necessary. For this, FPLOTCROPPER uses Matplotlib [32] packages for viewing the orthoimages and capturing mouse events, which the user produces when defining the four corners of each block. FPLOTCROPPER also uses Rasterio [33] for reading and writing of the images, along with the Python 3 programming language [34].

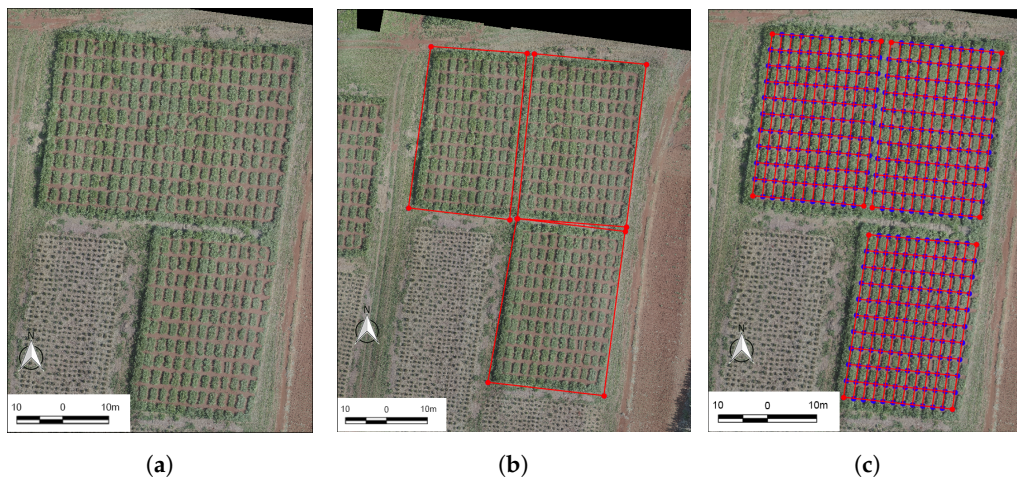(**a**)　　　　　　　　　　(**b**)　　　　　　　　　　(**c**)

Figure 2. Plots identification procedure: (**a**) Orthomosaic; (**b**) user defined experiment field in red and; (**c**) plots defined using our Python script.

The presented proposal uses biomass yield as a class attribute y. Figure 3 plots the histograms of the y data distribution of the 330 plots of the experimental station.
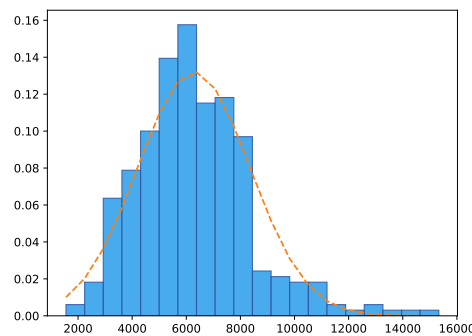
Figure 3. Class attribute y distribution-biomass in kg·ha^{-1}.

After this pre-processing step, where plots were correctly cropped and identified, we then proceeded with the experimental evaluation described in the next sections.

2.2. Deep Learning Approach and Experimental Setup

Conventional machine learning techniques require considerable domain expertise and careful engineering to extract meaningful features to train the models. According to LeCun [35], deep learning methods can learn convolutional filters that can obtain those meaningful features from the images. The advantage is that the learning process incorporates the learning of convolutional filters that replaces the feature extractors designed by human engineers. This approach requires minimal engineering by hand and has achieved a state of the art results in many areas of machine intelligence [35]. Therefore the filters to emphasize and extract essential features are now inside the model, allowing the model to process images in its raw format.

However, the training of CNNs requires large datasets, such as ImageNet, with 14 million images. In the forage literature, to the best of our knowledge, there is no such dataset. In green biomass estimation, the sample size is relatively small for deep learning standards. This limitation relates to obtaining the ground truth of the field biomass, which requires expensive laboring work to harvest and weigh each plot's biomass.

In our study, we have only 330 plots. Training CNNs with a small sample size is a challenging task. The literature points to fine-tuning approaches using a pre-trained model [36], learn smaller models [37], and perform data augmentation [38]. We address the learning problem by trying these three approaches. We selected AlexNet [39] (8 layers), a popular and relatively small convolutional neural network, and ResNet [40] (18 layers), both with and without a pre-trained model, and with and without data augmentation. We designed the experimental evaluation, not only to show the individual accuracy rate of each model but also to measure how effective is each of these three approaches. As a baseline we use VGGNet [41] (11 layers), because it was used for the same purpose in other species [30].

We used AlexNet, ResNet18 and VGGNet from PyTorch [42]. The last fully connected layer architectures was changed so that the models were adapted to a regression problem. The settings for each experiment can be seen in Table 1. In all experiments, the Adam [43] optimization method was used, with the descending gradient algorithm, in a fixed learning rate of 0.001, constant $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The number of epochs was defined empirically using early stopping evaluated every 100 epochs. The pre-trained models are PyTorch pre-trained model on ImageNet (see https://pytorch.org/docs/stable/torchvision/models.html), where we loaded these pre-trained weights and fine-tuned these models using the training set. Models without pre-training were fully trained using the training set.

Considering the relatively restricted number of 330 plots (examples), we performed all experiments using ten-fold cross-validation, since cross-validation produces better generalization error than hold-out [44,45]. To pursue more robust models, we trained them using the data augmentation

technique. We named *augmented horizontally* (`augmented h`) for regular data augmentation, where the images were flipped from left to right. We named *augmented horizontally and vertically* (`augmented hv`) for data augmentation, where the images were also flipped from top to bottom. The models without data augmentation were named `original` in Table 1.

Table 1. Experimental setup.

#Experiment	Model	Batch Size	Data-Set	Epochs
1	AlexNet	256	`original`	400
2	AlexNet	256	`augmented h`	400
3	AlexNet	256	`augmented hv`	500
4	Resnet18	128	`original`	500
5	Resnet18	128	`augmented h`	500
6	Resnet18	128	`augmented hv`	500
7	AlexNet Pre-Trained	256	`original`	200
8	AlexNet Pre-Trained	256	`augmented h`	200
9	AlexNet Pre-Trained	256	`augmented hv`	200
10	ResNet18 Pre-Trained	128	`original`	500
11	ResNet18 Pre-Trained	128	`augmented h`	400
12	ResNet18 Pre-Trained	128	`augmented hv`	400
13	VGGNet11 Pre-Trained	64	`augmented hv`	400

All models were trained to adopt the MSE (mean square error) as the loss function. The mean square error evaluated values between the actual value of biomass, in kg·ha^{-1} (y_i), and the value predicted by the model was used ($\hat{y}_i$). The formula is in the Equation (1),

$$\varepsilon = \frac{1}{n} \sum_{i=0}^{n} (y_i - \hat{y}_i)^2 \tag{1}$$

where n is the number of examples, y_i is the true score, $\hat{y}_i$ is the predicted score, and $y_i \in [1556, 15, 333]$, which means that the y_i varied from 1556 kg·ha^{-1} to 15,333 kg·ha^{-1}. We used a desktop with a NVIDIA Titan X GPU (12 GB), Intel i7-6800K 3.4 GHz CPU, and 64 GB of RAM.

We assess the regression problem using MAE (mean absolute error), MAPE (mean absolute percentage error), R (Pearson correlation). However, these metrics can hide predictions biased towards higher or lower values than the true prediction. The same can occur with the graphs in Section 3, where high-density regions can overlap points. This limitation motivates the evaluation of the results under RROC (Receiver Operating Characteristic Curves from Regression) [46]. We also evaluate the results using histograms (Section 3.3). The number of *bins* was determined using the elbow rule [47] from the partitioning of the samples' real values.

For visual inspection of the model activations, the last convolutional layers of CNNs retain spatial information and high-level semantics of the CNN [48]. Looking at these layers, it is possible to highlight the class-discriminative regions of the images. One of the first methods to emphasize the discriminative areas of the images was the CAM (class activation mapping) [48]. However, CAM works only on CNNs without fully connected layers architectures. One year later, Grad-CAM [49] was proposed, enabling the use of fully connected layers without architectural changes or re-training. We applied Grad-CAM and displayed some results in Section 3.4.

3. Experimental Results Evaluation

We divided our evaluation into four parts: (1) evaluation of the results using standard metrics MAE, MAPE, and R, and the graphs of predicted versus real values; (2) a visual representation of the error by ROC Regression curves; (3) the histograms of the predictions; and (4) the heat map of the feature map activation for visual inspection.

3.1. Standard Evaluation: MAE, MAPE, R, and Graph of Predicted Versus Real

Table 2 shows the mean and standard deviation of mean absolute error, mean absolute percentage error, and Pearson Correlation of ten-fold cross-validation. The results indicate four groups of outcomes. The best first group represents AlexNet pre-training, Experiment #7, #8, and #9. These results are among the top 3 results showing an average MAE lower than 768.75 kg·ha^{-1}. The second best group is also from AlexNet, but without pre-training, Experiment #1, #2, and #3 with average MAE lower than 924.48. The third and fourth groups are the ResNet18 results, where the MAE is higher than 1000. Overall and surprisingly, the results of AlexNet are better than ResNet18. The VGGNet (baseline) was considered only with pre-training and data augmentation because we verified that this contributed to the improvement of the other models. The AlexNet also outperformed VGGNet11, which presented an average MAE of 825.94 kg·ha^{-1}. Experiment #9 presented the best absolute result.

Table 2. Experimental results. Experiment #9's results present the lowest MAE (mean absolute error) and highest correlation. It is good to remind the reader that y_i, ranging from 1556.00 kg·ha^{-1} to 15,333.00 kg·ha^{-1}, therefore MAE of 730 represents a variation of 730 kg·ha^{-1} in this range of values.

#Experiment	Model	Mean Absolute Error	Mean Absolute Percentage Error	Correlation (*r*)
1	AlexNet	837 ± 106	14.58 ± 2.52	0.84 ± 0.03
2	AlexNet h	880 ± 202	15.11 ± 3.24	0.83 ± 0.06
3	AlexNet hv	924 ± 143	15.48 ± 2.30	0.82 ± 0.05
4	ResNet18	1086 ± 219	17.70 ± 3.41	0.74 ± 0.06
5	ResNet18 h	1046 ± 107	19.01 ± 2.77	0.74 ± 0.06
6	ResNet18 hv	1031 ± 153	18.76 ± 4.28	0.75 ± 0.06
7	AlexNet Pre-Trained	759 ± 102	13.23 ± 2.23	0.87 ± 0.05
8	AlexNet Pre-Trained h	768 ± 123	13.54 ± 2.88	0.87 ± 0.03
9	AlexNet Pre-Trained hv	730 ± 59	12.98 ± 2.18	0.88 ± 0.04
10	ResNet18 Pre-Trained	1206 ± 233	19.46 ± 5.15	0.73 ± 0.04
11	ResNet18 Pre-Trained h	1205 ± 194	23.16 ± 4.80	0.71 ± 0.07
12	ResNet18 Pre-Trained hv	1012 ± 128	18.58 ± 2.34	0.77 ± 0.05
13	VGGNet11 Pre-Trained	825 ± 152	13.89 ± 3.09	0.84 ± 0.04

We then proceed with the experimental evaluation performing a one-way ANOVA test. We obtained an F-statistic of 9.81, and with a *p*-value of $5.16 \times 10^{-13} < 0.05$, we can reject the null hypothesis that the models have are equal performance using MAE. We continue the evaluation with Tukey's pos-hoc test to find the differences among the experiments. Figure 4 shows the Tukey's HSD.
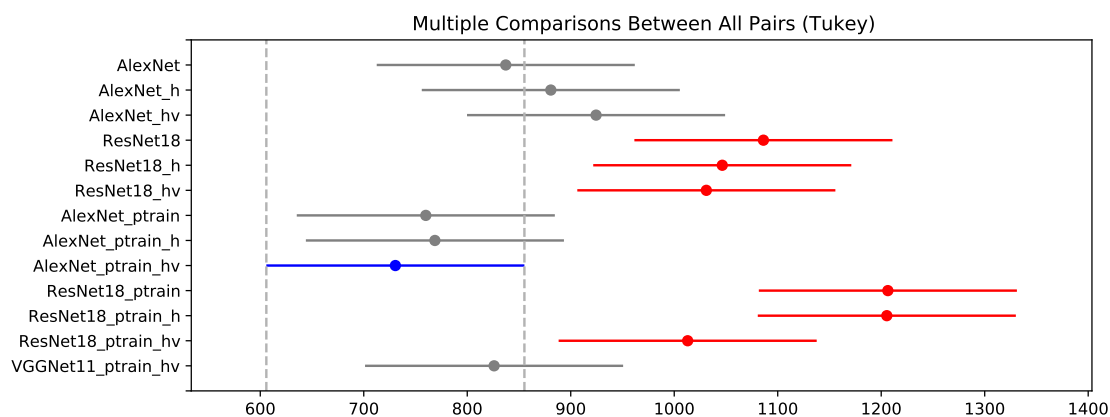

Figure 4. Mean and the 95% confidence interval of a pos-hoc Tukey's HSD test performed on MAE results.

The results that differ significantly with AlexNet_ptrain_hv are red, the insignificant difference are gray. Therefore the results indicate that AlexNet_ptrain_hv has a significant difference with all

ResNet18 results. The pre-trained models showed a slight improvement in the AlexNet result but not in ResNet. The ResNet results, except for ResNet18_ptrain_hv, the pre-training deteriorated the results. For data augmentation hv, we can see a small improvement in pre-trained models of AlexNet and ResNet. For the AlexNet without pre-training, the use of data augmentation worsens the results.

The mean and standard deviations of MAE, MAPE, and R assume Gaussian distributions. Plotting the real values and the prediction in the same graph can be a more precise visualization of the results, such as point cloud density and outliers.The graphs presented in Figure 5 show exactly these visualizations, where points are pair of $(y, \hat{y})$. In a situation of perfect predictions, these graphs would be perfect lines (1:1).

(**a**) Experiment #1　　　　　　(**b**) Experiment #2　　　　　　(**c**) Experiment #3

(**d**) Experiment #4　　　　　　(**e**) Experiment #5　　　　　　(**f**) Experiment #6

(**g**) Experiment #7　　　　　　(**h**) Experiment #8　　　　　　(**i**) Experiment #9

(**j**) Experiment #10　　　　　　(**k**) Experiment #11　　　　　　(**l**) Experiment #12
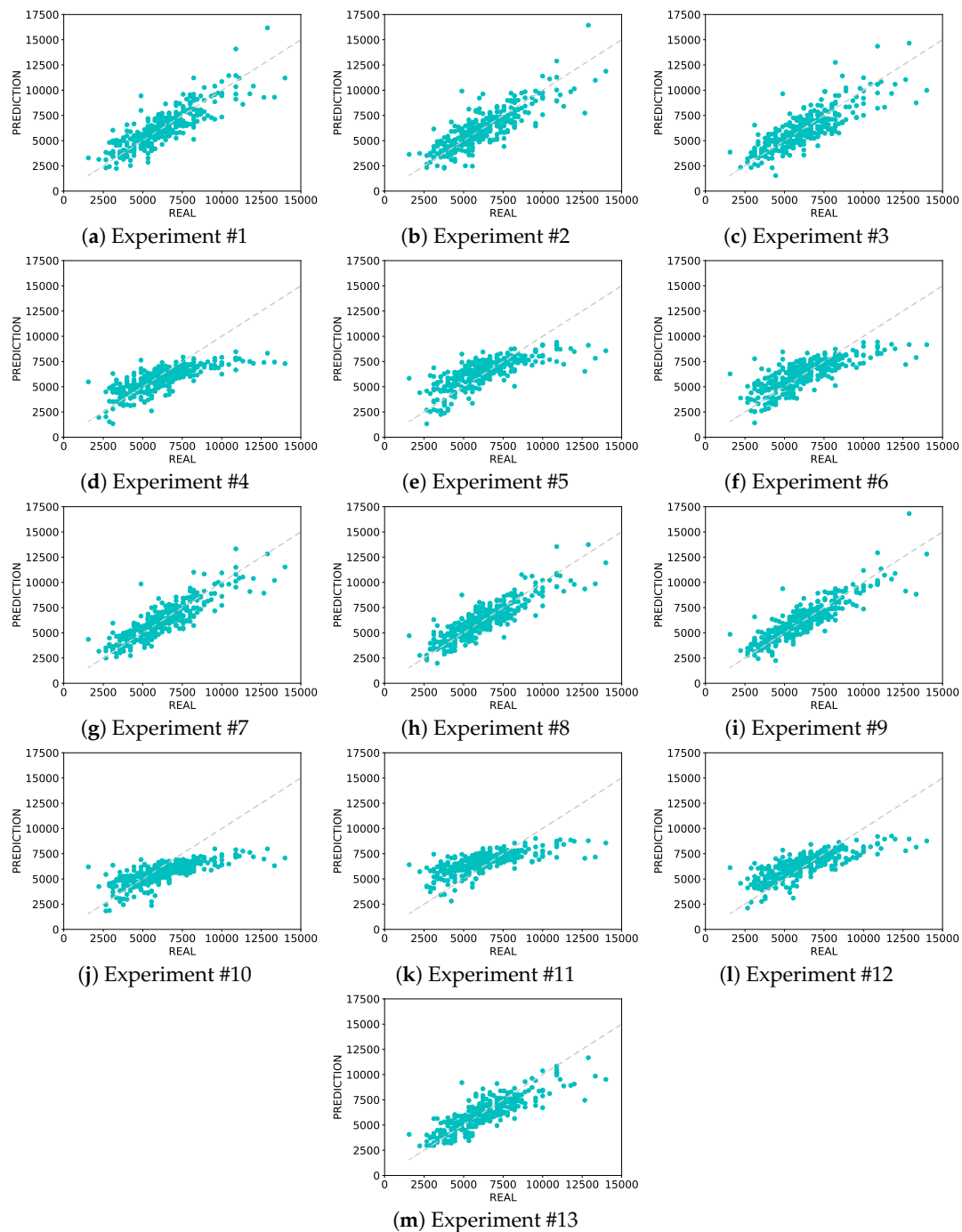
(**m**) Experiment #13

Figure 5. Predicted vs. real plots.

In all graphs, the models show more spread results on the top right and a high concentration of points in the center of the diagram, which corroborates with the real class y (Figure 3) where the numbers of biomass higher than 11,000 kg·ha^{-1} are scarce, and the average value of y is close to 6000 kg·ha^{-1}.

When comparing the graphs of AlexNet and ResNets, we can see that all ResNet results (Experiments #4, #5, #6, #10, #11, and #12,) show a point cloud lying more to the bottom right, indicating that ResNet has difficulties predicting values higher than 10,000 kg·ha^{-1} than AlexNet. The pre-trained AlexNet (Experiments #7, #8, and #9) shows a narrow corridor of points close to the ascending diagonal (dotted line). The narrowest point cloud and closer to the dotted line seems to be the Experiment #9, which confirms the results of the absolute number of MAE, MAPE, and R.

3.2. ROC Regression

Figure 6 shows RROC, where the closer to the $(0,0)$ point, called RROC heaven, the better the model. Closer the results to the dotted line $UNDER + OVER = 0$, the less biased the model is to predict values below or above the true regression values.
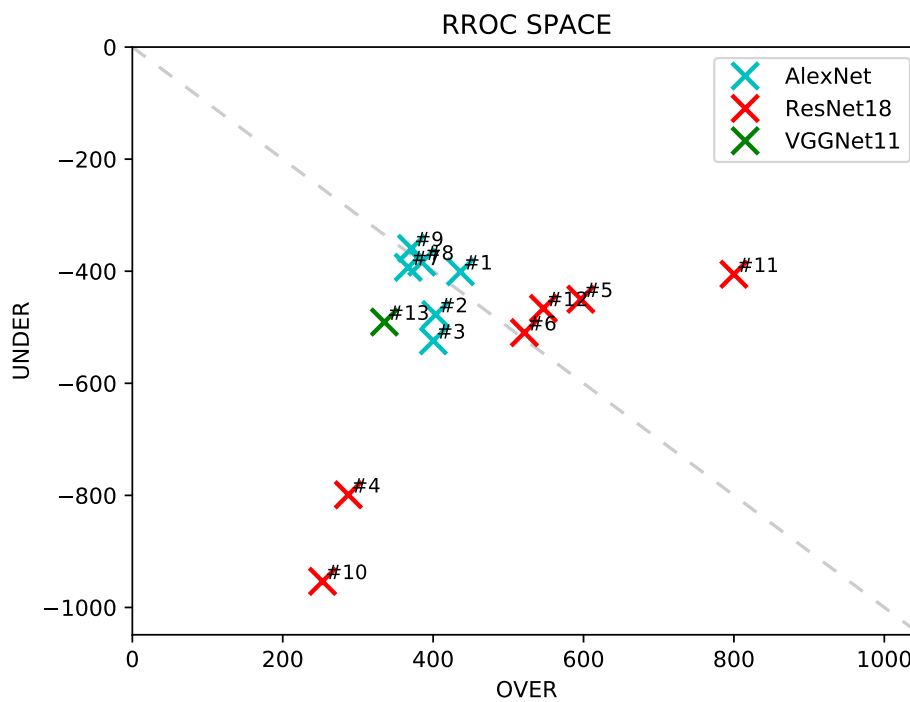

Figure 6. ROC (Receiver Operating Characteristic) for regression. Points closer to (0,0) mean better results.

Among the results with pre-training (Figure 6), AlexNet `original`, AlexNet `augmented h`, and AlexNet `augmented hv` (#7, #8, and #9) are closer to (0,0). ResNet18 pre-trained `augmented h` (Experiment #11) is above the dotted line, indicating better results for UNDER the prediction than OVER the prediction. In practice, this is equivalent to an average prediction of $\hat{y}$ slightly higher than the true prediction y for Experiment #11. This interpretation may be counter-intuitive when we look at Figure 5k due to the points on the middle right of the graph, however looking closely, we can see high-density points in the middle left, that corroborates with the RROC result. The other way around occurs with ResNet18 `original` (Experiment #4) and VGGNet11(Experiment #13), where the results are below the dotted line.

When comparing AlexNet (light blue) and ResNet (red), we can see that AlexNet points are closer to the (0,0). All ResNet points are further away from the AlexNet points and spread over the graph, showing a barrier of points close to the ascendant diagonal.

3.3. Histograms

The histogram graphs of Figure 7 show the intersection between the distribution of the real data and the distribution of the predictions of each experiment. The addition of new groups did not significantly increase the representativeness of the data with more than 20 bins, so the number of *bins* was set to 20. The intersection areas between the distributions were calculated for each experiment and are presented in Table 3.
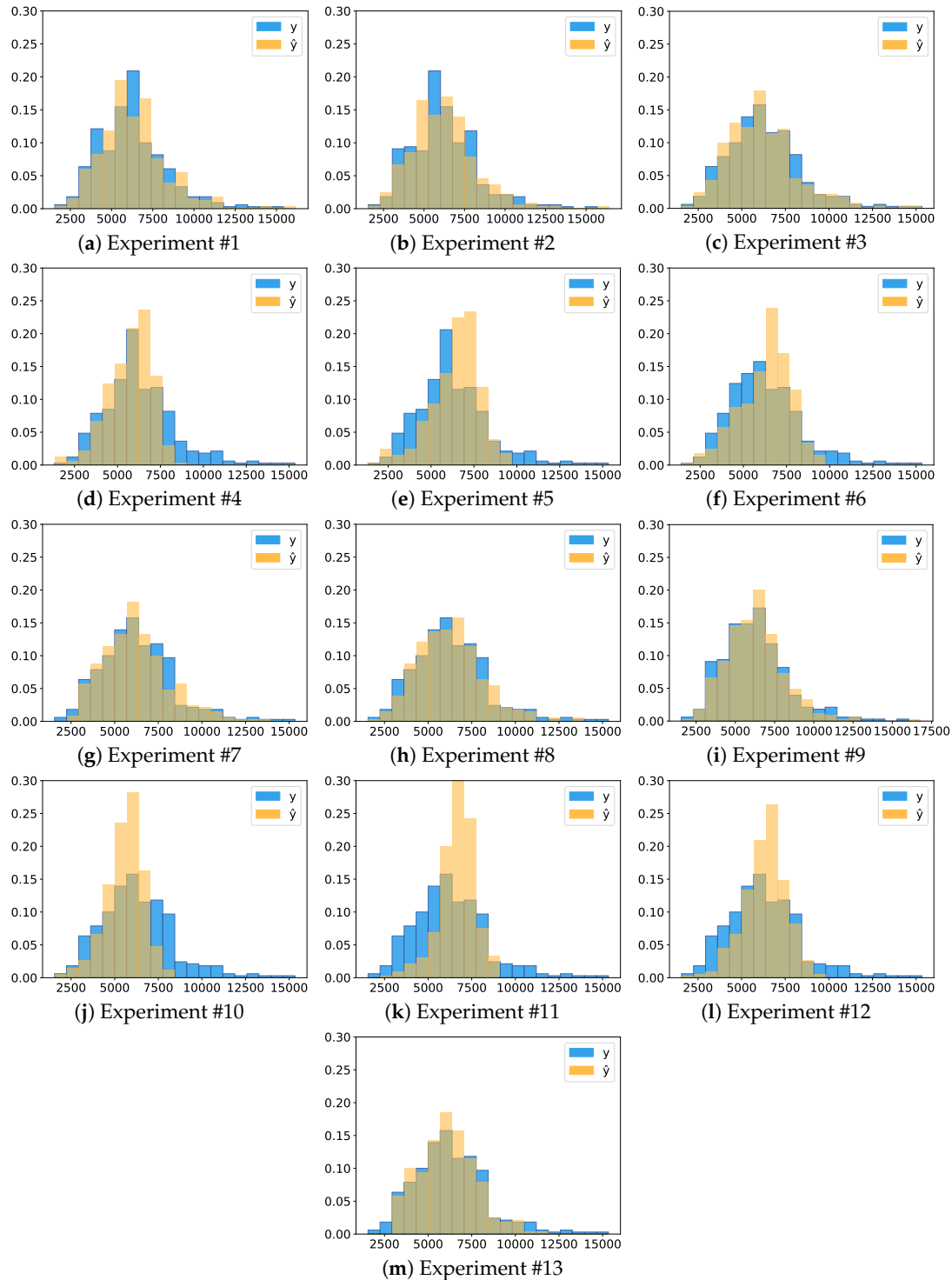

Figure 7. Comparison of prediction $\hat{y}$ vs. real y data distribution. Larger intersecting areas between histograms indicates better prediction.
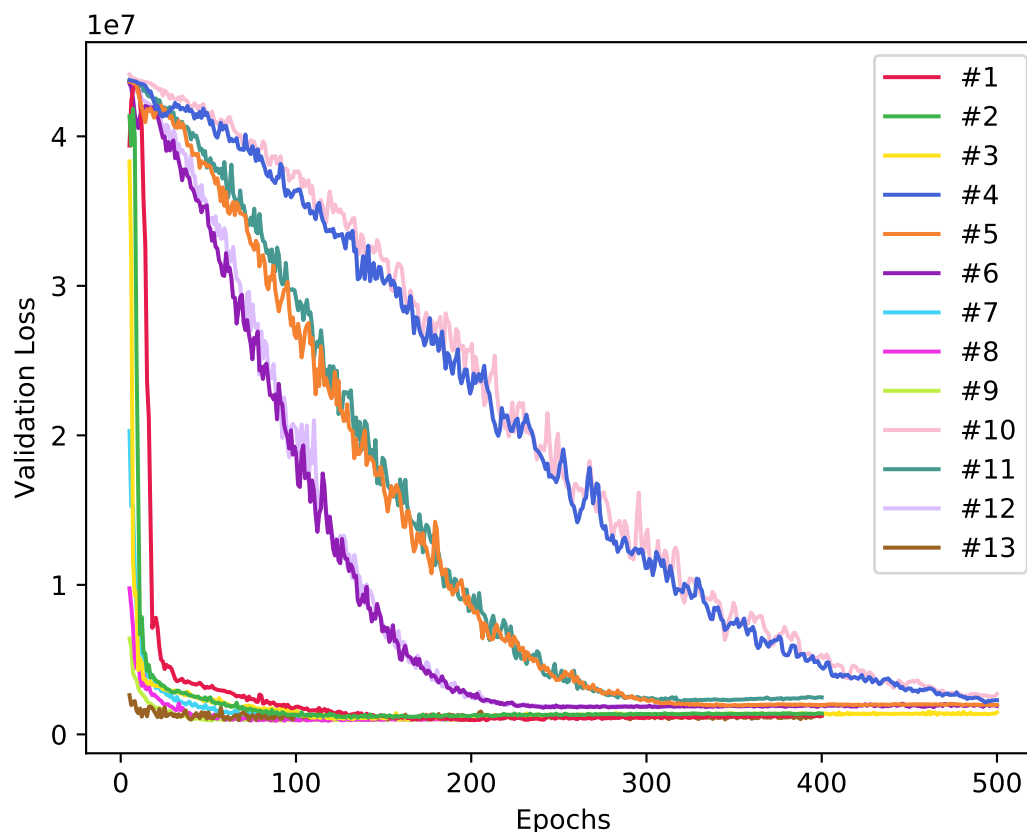
Experiment	1	2	3	4	5	6	7	8	9	10	11	12	13
Intersection Area	0.83	0.83	0.91	0.78	0.72	0.78	0.89	0.89	0.92	0.68	0.62	0.76	0.90

Comparing the ResNet and AlexNet, we can see a peek of orange values on x-axis surrounding the 7500 kg·ha^{-1} in all ResNet experiments (Experiment #4, #5, #6, #10, #11, and #12). The AlexNet histograms go along the true class (blue) histogram. Experiment #1 and #2 are the only histograms where the highest bin is blue.

Through the Table 3 and Figure 7, it is possible to conclude the superiority of the results of Experiment #9, where the training set used was at least twice as large concerning any other experiment due to the data augmentation.

From the plot of the validation loss given the number of epochs during training presented in Figure 8, it can be observed that Experiments #1, #2, #3, #7, #8, #9, and #13 were the ones to converge the fastest getting to a plateau at around 100 epochs. Experiments #5 and #11 converged at approximately 230 epochs. Experiments #6 and #12 converged after about 300 epochs. Finally, Experiments #4 and #10 were the ones that took the longest to reach the plateau, taking approximately 500 epochs. From this analysis, it is possible to conclude that the AlexNet and VGGNet11 models were able to consistently present better results earlier when compared to experiments using the ResNet18 model.

Figure 8. Validation loss over epochs for all experiments.

3.4. Visual Inspection

Figures 9 and 10 show the heatmaps from Grad-CAM on Experiment #9. Warm colors (red) indicate a more class-discriminative region to the prediction, while cold colors (blue) represent the lower class-discriminative region.
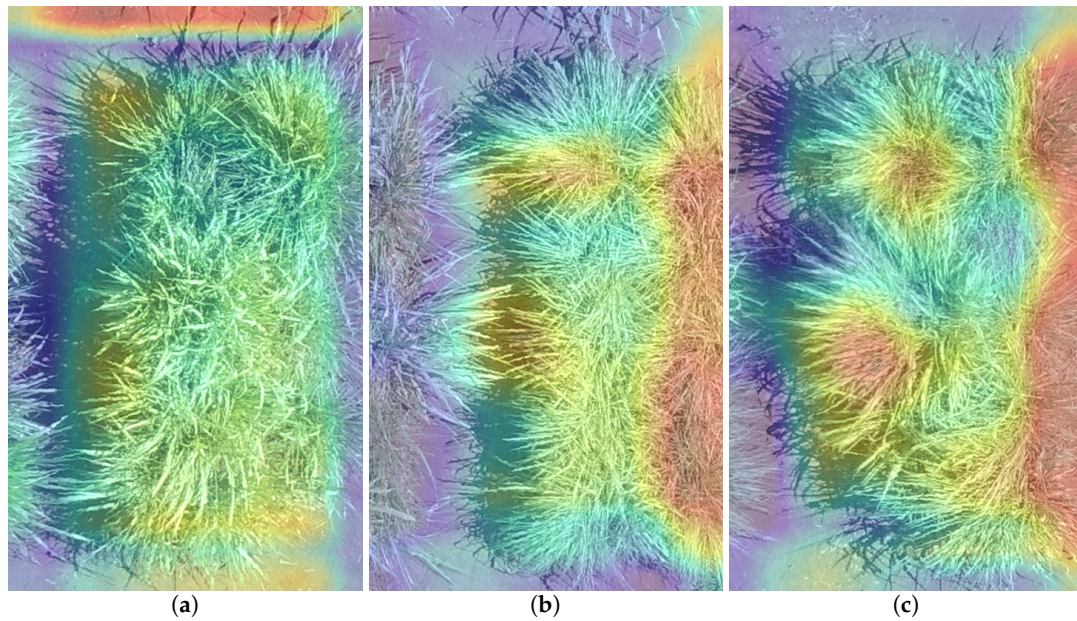
Figure 9. Heatmaps of the top 3 best predictions made by Experiment #9 (Pre-trained AlexNet Model with hv data augmentation): (**a**) First best prediction; (**b**) Second best prediction; (**c**) Third best prediction.
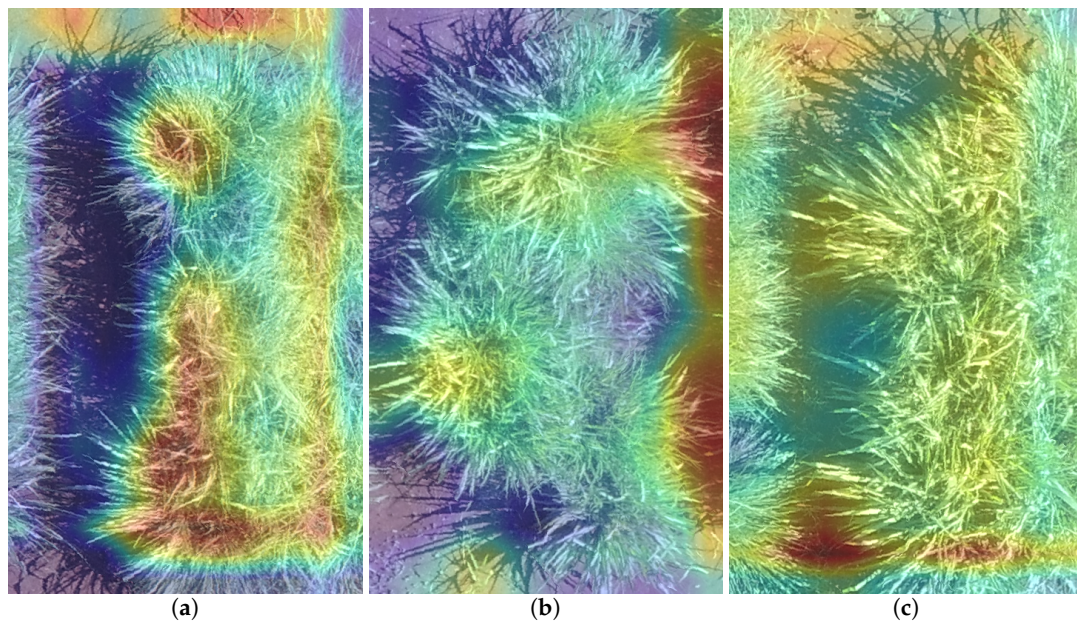

Figure 10. Heatmaps of the top 3 worst predictions made by Experiment #9 (Pre-trained AlexNet Model with hv data augmentation): (**a**) First worst prediction; (**b**) Second worst prediction; (**c**) Third worst prediction.

Figure 9 shows the three best predictions, where no strong red or blue regions are highlighted. Figure 10 shows the three worst predictions, where extreme values in the neurons, strong red and blues, occur. We believe that these extreme values create a broader range of values that make the regression problem more difficult to solve, worsening the results' MAE.

3.5. Training and Test Time

Finally, we also evaluated the training and test time of one fold of the ten-fold cross-validation procedure. The test set has fix size of 33 examples, and the training set 300, 600, and 900 examples for original, augmented h and augmented hv, respectively. Table 4 shows the results.

Table 4. Training and test time on one-fold of the cross-validation procedure.

#Experiment	Model	Training Time (min)	Test Time (s)
1	AlexNet	35.8	0.39
2	AlexNet h	95.6	0.46
3	AlexNet hv	122.2	0.40
4	ResNet18	60.2	0.47
5	ResNet18 h	131.3	0.54
6	ResNet18 hv	133.7	0.53
7	AlexNet Pre-Trained	15.8	0.37
8	AlexNet Pre-Trained h	36.2	0.44
9	AlexNet Pre-Trained hv	43.7	0.50
10	ResNet18 Pre-Trained	58.1	0.47
11	ResNet18 Pre-Trained h	102.2	0.67
12	ResNet18 Pre-Trained hv	104.4	0.43
13	VGGNet11 Pre-Trained hv	372.2	0.81

The training time needs to be analyzed with the respective number of training steps. It is interesting to see that although the h and hv have two and three times more examples, their training time was not multiplied by a factor of two and three on ResNet. While for the AlexNet, these factors are consistent. The test time stays between 0.39 and 0.67 for AlexNet and ResNet. The testing time of VGG was the highest among the tested results.

4. Discussion

Our study focused on a deep learning-based approach to estimate biomass yield in forage fields. Furthermore, we investigated the impact of the data augmentation and pre-training steps on the estimation results. CNN is a state-of-the-art method to evaluate imagery data. We also considered different genotypes of a forage species, which contribute to the heterogeneity of our dataset. For assessing the experiments, we presented regression analysis results with a high correlation between predicted and measured yields (Table 2 and Figure 4), an RROC space comparing the deep networks implemented (Figure 5), an intersection analysis between each experiment (Table 3 and Figure 6) and qualitative information such as heat maps illustrating the best and worst predictions of our data (Figures 8 and 9).

When comparing against other methods to estimate biomass, our approach differentiates from a methodological point-of-view. Up to the point, most approaches considered shallow learners (i.e., conventional machine learning methods) and RGB imagery combined with DSM and DTM models, LiDAR or even vegetation spectral indices from multi and hyperspectral data [1,24–28,50]. As an advantage of adopting deep neural networks is the oversimplification of remote sensing data variety necessary to conduct this estimation. Although the production of input data such as DTM, DSM, spectral indices, and multiple bands is a relatively easy task in remote sensing, the costs to obtain such products, specifically LiDAR and hyperspectral data, are highly-priced when comparing against RGB only imagery. Our deep networks, trained with RGB inputs, can perform similarly or even better than most traditional methods and data previously described. This is an important advance for agricultural remote sensing approaches. The caveat of the proposal and other CNN-based approaches are the requirements of computational intensive procedures, often requiring the use of GPUs.

Compared to the previous study that used 3D information to estimate the biomass for the same species [26], we achieved more accurate results. Batistoti et al. [26] achieved an R2 of 0.74 considering a limited number of plots (four in total). Another study [51] estimated yield with proximal sensing equipment in a heterogeneous sward structure of grasslands and applied a MLPSR (Multiple Partial Least Square Regression) approach, which returned an R2 of 0.69. In estimating biomass from legume-grass, Wachendorf et al. [52] produced high accuracies with a similar approach as Moeckel et al. [51], returning accuracies up to 0.95 for specific models. Another study [53] considered machine learning techniques to estimate grassland biomass in spectral data, with an RMSE equal to 71.2 t·ha^{-1}. This indicates that our UAV-based RGB dataset, in conjunction with more robust methods (i.e., deep neural networks), can be compared against expensive measurement remote systems, considering even terrestrial measurements.

In this study, we implemented two convolutional networks (AlexNet and ResNet18) and tested as a baseline another network (VGGNet11), as stated in the previous sections. To evaluate the impact of samples in our approach we tested whether pre-trained models and data augmentation produced significant improvements in its accuracy. Our results indicated that the AlexNet method performed better. A possible explanation for this is that the ResNet18 method, although being a deeper network than the implemented AlexNet, was unable to represent the pre-trained problem with its convolutional filters properly. In other words, it was not able to modify its layers with enough precision. Results without pre-training steps were also not sufficient. This demonstrates how the lack of data for training impacted its performance. Nevertheless, the evaluation of different pre-processing steps (with and without data augmentation and pre-training) resulted in essential implications for integrating agronomic measurements collected in the field with these robust methods in remote sensing RGB imagery.

The VGGNet11 performance was calculated to be compared against the accuracy obtained in a previous paper [30]. In this paper [30], more traditional approaches were also compared against this deep learning method. These approaches were related to spectral vegetation indices and 3D models as standalone data to estimate biomass yield. Even so, it was demonstrated that the VGGNet11 outperformed these traditional approaches. Here, our proposal focused mostly on evaluations with the AlexNet and ResNet18 networks throughout the experiment. We noticed that in both methods, data augmentation improved the overall performance in estimating biomass yield. As a result of this outcome, we also implemented data augmentation in the VGGNet11 network. Nonetheless, our analysis (Figure 5) demonstrated that the AlexNet method was superior to the deep learning method implemented in [30], even with data augmentation. This may be an indicator that this type of approach with RGB imagery performs better with shallow architectures.

It is possible to observe that the models used were able to return a high correlation between the RGB images of the plots and the real yield value (Table 2 and Figure 4). This study is the first approximation of its kind. We believe that many aspects to be evaluated in future research. Although we stated the importance of RGB data in an economic point-of-view, we do not disregard the impact of other types of remote sensing methods to increase the accuracy of deep learning-based neural networks. The use of 3D reconstruction data from point-clouds can also be explored in the future since biomass has a strong relationship with the volume of the plant. This data insertion could assess whether there are performance gains in the prediction and infer the density of the plant, a significant characteristic for researchers in the area. Regardless, the results utilizing only RGB and the cross-validation method indicate the capability of the proposed method in this approach.

5. Conclusions

Until this moment, this paper's proposed approach is the first research that implemented and evaluated a CNN-based architecture, combined with high-resolution UAV RGB images, for the prediction of biomass yield considering different forage genotypes.

Two regression models based on CNNs (Convolutional Neural Networks) named AlexNet and ResNet18 were evaluated, and compared to VGGNet—adopted in previous work in the same thematic for other grass species. The predictions returned by the models reached a correlation of 0.88 and a mean absolute error of 12.98% using AlexNet considering pre-training and data augmentation. Comparing the achieved results to a previous study that was based on 3D information to estimate the biomass for the same species [26], we achieved more accurate results.

In conclusion, the models used were able to establish a high correlation between the images and the biomass value measured in the field. This demonstrates how feasible the proposed approach is to predict forage yield at highly detailed RGB imagery, producing accuracy comparable to more expensive approaches with both aerial and proximal remote sensing.

Since this is the first study of its kind, there are many aspects to be evaluated in future research. It is worth noting that the models developed here are not yet ready to be deployed in commercial production. Although cross-validation has been used in all experiments, the dataset is still considered small. However, the results obtained are strong indications of the method's future success in more significant variations of forage crop datasets. Experiments using datasets from different locations and weather conditions are essential to provide more generalized models, and we intend to conduct this in future work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bendig, J.; Bolten, A.; Bennertz, S.; Broscheit, J.; Eichfuss, S.; Bareth, G. Estimating Biomass of Barley Using Crop Surface Models (CSMs) Derived from UAV-Based RGB Imaging. *Remote Sens.* **2014**, *6*, 10395–10412. [CrossRef]
2. Gebremedhin, A.; Badenhorst, P.E.; Wang, J.; Spangenberg, G.C.; Smith, K.F. Prospects for measurement of dry matter yield in forage breeding programs using sensor technologies. *Agronomy* **2019**, *9*, 65. [CrossRef]
3. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, 111402. [CrossRef]
4. Osco, L.P.; Ramos, A.P.M.; Pereira, D.R.; Moriya, É.A.S.; Imai, N.N.; Matsubara, E.T.; Estrabis, N.; de Souza, M.; Junior, J.M.; Gonçalves, W.N.; et al. Predicting canopy nitrogen content in citrus-trees using random forest algorithm associated to spectral vegetation indices from UAV-imagery. *Remote Sens.* **2019**, *11*, 2925. [CrossRef]
5. D'Oliveira, M.V.N.; Broadbent, E.N.; Oliveira, L.C.; Almeida, D.R.A.; Papa, D.A.; Ferreira, M.E.; Zambrano, A.M.A.; Silva, C.A.; Avino, F.S.; Prata, G.A.; et al. Aboveground Biomass Estimation in Amazonian Tropical Forests: a Comparison of Aircraft- and GatorEye UAV-borne LiDAR Data in the Chico Mendes Extractive Reserve in Acre, Brazil. *Remote Sens.* **2020**, *12*, 1754. [CrossRef]

6. Miyoshi, G.T.; Arruda, M.D.S.; Osco, L.P.; Junior, J.M.; Gonçalves, D.N.; Imai, N.N.; Tommaselli, A.M.G.; Honkavaara, E.; Gonçalves, W.N. A novel deep learning method to identify single tree species in UAV-based hyperspectral images. *Remote Sens.* **2020**, *12*, 1294. [CrossRef]

7. Leiva, J.N.; Robbins, J.; Saraswat, D.; She, Y.; Ehsani, R. Evaluating remotely sensed plant count accuracy with differing unmanned aircraft system altitudes, physical canopy separations, and ground covers. *J. Appl. Remote Sens.* **2017**, *11*, 036003. [CrossRef]

8. Liu, T.; Abd-Elrahman, A.; Morton, J.; Wilhelm, V.L. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GISci. Remote Sens.* **2018**, *55*, 243–264. [CrossRef]

9. Abdulridha, J.; Batuman, O.; Ampatzidis, Y. UAV-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning. *Remote Sens.* **2019**, *11*. [CrossRef]

10. Feng, P.; Wang, B.; Liu, D.L.; Yu, Q. Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. *Agric. Syst.* **2019**, *173*, 303–316. [CrossRef]

11. Osco, L.P.; Ramos, A.P.M.; Pinheiro, M.M.F.; Moriya, É.A.S.; Imai, N.N.; Estrabis, N.; Ianczyk, F.; de Araújo, F.F.; Liesenberg, V.; de Castro Jorge, L.A.; et al. A machine learning framework to predict nutrient content in valencia-orange leaf hyperspectral measurements. *Remote Sens.* **2020**, *12*, 906. [CrossRef]

12. Ghamisi, P.; Plaza, J.; Chen, Y.; Li, J.; Plaza, A.J. Advanced Spectral Classifiers for Hyperspectral Images: A review. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–32. [CrossRef]

13. Al-Saffar, A.A.M.; Tao, H.; Talab, M.A. Review of deep convolution neural network in image classification. In Proceedings of the 2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), Jakarta, Indonesia, 23–24 October 2017; pp. 26–31.

14. Alshehhi, R.; Marpu, P.R.; Woon, W.L.; Mura, M.D. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 139–149. [CrossRef]

15. Khamparia, A.; Singh, K.M. A systematic review on deep learning architectures and applications. *Expert Syst.* **2019**, *36*, 1–22,. [CrossRef]

16. Csillik, O.; Cherbini, J.; Johnson, R.; Lyons, A.; Kelly, M. Identification of Citrus Trees from Unmanned Aerial Vehicle Imagery Using Convolutional Neural Networks. *Drones* **2018**, *2*, 39. [CrossRef]

17. Hassanein, M.; Khedr, M.; El-Sheimy, N. Crop row detection procedure using low-cost uav imagery system. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *42*, 349–356. [CrossRef]

18. Wu, L.; Zhu, X.; Lawes, R.; Dunkerley, D.; Zhang, H. Comparison of machine learning algorithms for classification of LiDAR points for characterization of canola canopy structure. *Int. J. Remote Sens.* **2019**, *40*, 5973–5991. [CrossRef]

19. Kitano, B.T.; Mendes, C.C.T.; Geus, A.R.; Oliveira, H.C.; Souza, J.R. Corn Plant Counting Using Deep Learning and UAV Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, 1–5. [CrossRef]

20. Dian Bah, M.; Hafiane, A.; Canals, R. Deep learning with unsupervised data labeling for weed detection in line crops in UAV images. *Remote Sens.* **2018**, *10*, 1–20. [CrossRef]

21. Legg, M.; Bradley, S. Ultrasonic Arrays for Remote Sensing of Pasture Biomass. *Remote Sens.* **2019**, *12*, 111. [CrossRef]

22. Loggenberg, K.; Strever, A.; Greyling, B.; Poona, N. Modelling water stress in a Shiraz vineyard using hyperspectral imaging and machine learning. *Remote Sens.* **2018**, *10*, 202. [CrossRef]

23. Fan, Z.; Lu, J.; Gong, M.; Xie, H.; Goodman, E.D. Automatic Tobacco Plant Detection in UAV Images via Deep Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 876–887. [CrossRef]

24. Bendig, J.; Yu, K.; Aasen, H.; Bolten, A.; Bennertz, S.; Broscheit, J.; Gnyp, M.L.; Bareth, G. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *39*, 79–87. [CrossRef]

25. Ballesteros, R.; Ortega, J.F.; Hernandez, D.; Moreno, M.A. Onion biomass monitoring using UAV-based RGB imaging. *Precis. Agric.* **2018**, *19*, 840–857. [CrossRef]

26. Batistoti, J.; Marcato Junior, J.; Ítavo, L.; Matsubara, E.; Gomes, E.; Oliveira, B.; Souza, M.; Siqueira, H.; Salgado Filho, G.; Akiyama, T.; et al. Estimating Pasture Biomass and Canopy Height in Brazilian Savanna Using UAV Photogrammetry. *Remote Sens.* **2019**, *11*, 2447. [CrossRef]

27. Näsi, R.; Viljanen, N.; Kaivosoja, J.; Alhonoja, K.; Hakala, T.; Markelin, L.; Honkavaara, E. Estimating Biomass and Nitrogen Amount of Barley and Grass Using UAV and Aircraft Based Spectral and Photogrammetric 3D Features. *Remote Sens.* **2018**, *10*, 1082. [CrossRef]

28. Li, B.; Xu, X.; Zhang, L.; Han, J.; Bian, C.; Li, G.; Liu, J.; Jin, L. Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 161–172. [CrossRef]

29. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]

30. Ma, J.; Li, Y.; Chen, Y.; Du, K.; Zheng, F.; Zhang, L.; Sun, Z. Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network. *Eur. J. Agron.* **2019**, *103*, 117–129. [CrossRef]

31. Jank, L.; Barrios, S.C.; do Valle, C.B.; Simeão, R.M.; Alves, G.F. The value of improved pastures to Brazilian beef production. *Crop Pasture Sci.* **2014**, *65*, 1132–1137. [CrossRef]

32. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]

33. Gillies, S.; Ward, B.; Petersen, A.S. Rasterio: Geospatial Raster I/O for Python Programmers. 2013. Available online: https://github.com/mapbox/rasterio (accessed on 17 August 2020)

34. Oliphant, T.E. Python for scientific computing. *Comput. Sci. Eng.* **2007**, *9*, 10–20. [CrossRef]

35. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]

36. Lu, J.; Behbood, V.; Hao, P.; Zuo, H.; Xue, S.; Zhang, G. Transfer learning using computational intelligence: A survey. *Knowl.-Based Syst.* **2015**, *80*, 14–23. [CrossRef]

37. Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; Vinyals, O. Understanding deep learning requires rethinking generalization. *arXiv* **2016**, arXiv:1611.03530.

38. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]

39. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. 2012. Available online: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networ (accessed on 10 August 2020).

40. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

41. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.

42. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An imperative style, high-performance deep learning library. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 8024–8035.

43. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

44. Weiss, S.M.; Kapouleas, I. An empirical comparison of pattern recognition, neural nets, and machine learning classification methods. *IJCAI* **1989**, *89*, 781–787.

45. Blum, A.; Kalai, A.; Langford, J. Beating the hold-out: Bounds for k-fold and progressive cross-validation. In Proceedings of the Twelfth Annual Conference on Computational Learning Theory, Santa Cruz, CA, USA, 7–9 July 1999; pp. 203–208.

46. Hernández-Orallo, J. ROC curves for regression. *Pattern Recognit.* **2013**, *46*, 3395–3411. [CrossRef]

47. Thorndike, R.L. Who belongs in the family. *Psychometrika* **1953**, *18*, 267–276. [CrossRef]

48. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.

49. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.

50. Viljanen, N.; Honkavaara, E.; Näsi, R.; Hakala, T.; Niemeläinen, O.; Kaivosoja, J. A novel machine learning method for estimating biomass of grass swards using a photogrammetric canopy height model, images and vegetation indices captured by a drone. *Agriculture* **2018**, *8*, 70. [CrossRef]

51. Moeckel, T.; Safari, H.; Reddersen, B.; Fricke, T.; Wachendorf, M. Fusion of ultrasonic and spectral sensor data for improving the estimation of biomass in grasslands with heterogeneous sward structure. *Remote Sens.* **2017**, *9*, 98. [CrossRef]

52. Wachendorf, M.; Fricke, T.; Möckel, T. Remote sensing as a tool to assess botanical composition, structure, quantity and quality of temperate grasslands. *Grass Forage Sci.* **2018**, *73*, 1–14. [CrossRef]

53. Marabel, M.; Alvarez-Taboada, F. Spectroscopic determination of aboveground biomass in grasslands using spectral transformations, support vector machine and partial least squares regression. *Sensors* **2013**, *13*, 10027–10051. [CrossRef] [PubMed]