remote sensing

MDPI

Article

Benchmarking Anchor-Based and Anchor-Free State-of-the-Art Deep Learning Methods for Individual Tree Detection in RGB High-Resolution Images

Pedro Zamboni [1,*], José Marcato Junior [1], Jonathan de Andrade Silva [2], Gabriela Takahashi Miyoshi [3], Edson Takashi Matsubara [2], Keiller Nogueira [4] and Wesley Nunes Gonçalves [1,2]

1 Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande 79070-900, Brazil; jose.marcato@ufms.br (J.M.J.); wesley.goncalves@ufms.br (W.N.G.)
2 Faculty of Computer Science, Federal University of Mato Grosso do Sul, Campo Grande 79070-900, Brazil; jonathan.andrade@ufms.br (J.d.A.S.); edsontm@facom.ufms.br (E.T.M.)
3 Department of Cartography, São Paulo State University (UNESP), Presidente Prudente 19060-900, Brazil; gabriela.t.miyoshi@unesp.br
4 Computing Science and Mathematics Division, Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA, UK; keiller.nogueira@stir.ac.uk
* Correspondence: pedro.zamboni@ufms.br

Abstract: Urban forests contribute to maintaining livability and increase the resilience of cities in the face of population growth and climate change. Information about the geographical distribution of individual trees is essential for the proper management of these systems. RGB high-resolution aerial images have emerged as a cheap and efficient source of data, although detecting and mapping single trees in an urban environment is a challenging task. Thus, we propose the evaluation of novel methods for single tree crown detection, as most of these methods have not been investigated in remote sensing applications. A total of 21 methods were investigated, including anchor-based (one and two-stage) and anchor-free state-of-the-art deep-learning methods. We used two orthoimages divided into 220 non-overlapping patches of 512 × 512 pixels with a ground sample distance (GSD) of 10 cm. The orthoimages were manually annotated, and 3382 single tree crowns were identified as the ground-truth. Our findings show that the anchor-free detectors achieved the best average performance with an AP_{50} of 0.686. We observed that the two-stage anchor-based and anchor-free methods showed better performance for this task, emphasizing the FSAF, Double Heads, CARAFE, ATSS, and FoveaBox models. RetinaNet, which is currently commonly applied in remote sensing, did not show satisfactory performance, and Faster R-CNN had lower results than the best methods but with no statistically significant difference. Our findings contribute to a better understanding of the performance of novel deep-learning methods in remote sensing applications and could be used as an indicator of the most suitable methods in such applications.

Keywords: object detection; convolutional neural network; remote sensing

Citation: Zamboni, P.; Junior, J.M.; de Andrade Silva, J.; Miyoshi, G.T.; Matsubara, E.T.; Nogueira, K.; Gonçalves, W.N. Benchmarking Anchor-Based and Anchor-Free State-of-the-Art Deep Learning Methods for Individual Tree Detection in RGB High-Resolution Images. *Remote Sens.* **2021**, *13*, 2482. https://doi.org/10.3390/rs13132482

Academic Editor: Antonio Pertusa

Received: 18 May 2021
Accepted: 15 June 2021
Published: 25 June 2021

1. Introduction

The urban population is expected to grow at the highest rates in human history in the next decades, with an increase of 1.2 billion urban residents worldwide by 2030 [1]. Densely populated areas are hotspots of numerous environmental problems, including air pollution [2,3] and hydrological disturbance [4,5], and are also linked to mental illness and health. [6,7]. Global climate change affects climate patterns, and the increase of surface temperature has led to more frequent, longer, and more severe heatwaves [8] and, likewise, increased the occurrence of floods [9]. In this scenario, urban forests could play an important role in mitigating some of these threats [10] and filling the gap between sustainable and livable cities. These systems are important assets to achieve urban sustain-

ability, which was established as one of the Sustainable Development Goals (SDG 11) by the United Nations [11].

Urban forests are composed of trees in urban areas, from individuals to clusters of trees, and publicly accessible green spaces [12]. These systems provide an array of ecosystem services and help to mitigate the ills caused by the concentrated poverty and sickness that often occur in cities [13]. However, the proper management of these systems requires accurate data on the quantity and distribution of individual trees throughout cities. Individual tree detection is key information for multiple applications, including resource inventory and the management of hazards and stress [14].

The urban tree inventory is increasing due to rapid urbanization, climate change, and global trade [15]. Nonetheless, there is a lack of information in urban tree inventories due to the costs associated with tree mapping and monitoring [16]. Individual tree detection is still an open challenge that is especially difficult since there are different vegetation canopies, crown sizes, and density as well as the overlapping of crowns, among other situations [17]. Further, due to the proximity and overlapping of tree crowns, it is not always possible to conduct segmentation as a strategy to detect each tree individually.

In this sense, methods that can do this task on RGB images could unlock data at larger scales [18] and provide insights for policymakers and the community. Developing approaches to detect trees automatically is decisive to building more sustainable cities, by helping with the planning and management of urban forests and trees. Most urban tree inventory is done manually, which is slow, costly, and difficult to map large areas and follow the temporal evolution of these assets.

To that end, remote sensing has been seen as a less time-consuming alternative to tree field surveys [19]. The recent development of these platforms has allowed researchers to collect data with higher spatial, temporal, and spectral resolutions, unlocking new scales of Earth observation. High-resolution images are recommended for individual tree detection [20], especially in urban areas where images are heterogeneous and complex. Nonetheless, this increase in data has made it difficult to process it manually. As an alternative, object detection methods based on deep learning have been applied successfully in remote sensing applications, mainly using convolutional neural networks [21–26].

Object detection methods can be divided into anchor-based and anchor-free detectors. Anchor-based methods first build an extensive number of anchors on the image, predict the category, refine each anchor's coordinates, and then output the refined anchor as the prediction. These types of techniques can be categorized into two groups: one and two-stage detectors. One-stage detectors commonly use a direct fully convolution architecture, while two-stage detectors first filter the region that may contain the object and feed a convolution network with this region [27].

Usually, one-stage methods (e.g., Yolo and SSD) provide high inference speed while two-stage methods (e.g., Faster R-CNN) present high localization and accuracy [28]. In contrast to the anchor-based detectors, anchor-free methods directly find objects beyond the anchor using a key or center point or region of the object and have similar performance to anchor-based methods [29]. The design of object detection methods using deep learning often uses natural images and, with the constant development of new methods, it is essential to evaluate the latest methods in remote sensing.

Within remote sensing and deep learning, many data sources have been investigated for individual tree detection. Multi and hyperspectral images, LiDAR (Light Detection And Ranging) data, and their combinations have all been investigated [30–36]. However, these data sources are costly and could be problematic to process due to their high dimensionality. Alternatively, studies were conducted that combined RGB images with other data sources [18,37,38]. Compared to other data sources, RGB sensors are cheaper, and RGB imagery is easier to process with absence of three-dimensional information about the tree crown shape [18]. However, few studies tackled this task using only remote sensing RGB data [19,39–42].

Santos at al. [19] applied three deep-learning methods (Faster R-CNN, YOLOv3, and RetinaNet) to detect one tree species, *Dipteryx alata* Vogel (Fabaceae), in Unmanned Aerial Vehicle (UAV) high-resolution RGB images. The authors found that RetinaNet achieved the best results. Culman et al. [39] implemented RetinaNet to detect palm trees in aerial high-resolution RGB images achieving a mean average precision of 0.861. Further, Oh et al. [40]used YOLOv3 to count cotton plants. Roslan et al. [41] applied RetinaNet for this task in super-resolution RGB images in a tropical forest. For a tropical forest again, [42] evaluated RetinaNet. However, most of the research on this field has been done using methods, such Faster R-CNN and RetinaNet, both being dated before 2018. With the constant development of new methods, there is a need to assess the performance of these methods in remote sensing applications.

Despite these initial efforts, there is a lack of studies assessing the performance of the novel deep-learning methods for individual tree crown detection, regarding tree species or size, in urban areas. This task is challenging in the urban context due to the heterogeneity of these scenes [20], with different tree types and sizes combined with overlap between objects, shades, and other situations. Our objective is to benchmark anchor-based and anchor-free detectors for tree crown detection in high-resolution RGB images in urban areas.

To the best of our knowledge, our study is the first to present a large assessment of novel deep learning detection methods for individual tree crown detection in urban areas. Further, we also provide an analysis covering the main lines of research in computer vision for anchor-based methods (one and two-stages) and anchor-free methods. Different from previous studies, our focus is to detect all trees, regarding tree species or size in an urban environment. Thus, our study intends to fill the gap and demonstrate the performance of the most advanced object detection methods in remote sensing applications.

Two high-resolution RGB orthoimages were manually annotated and split into non-overlapping patches. We evaluate 21 novel deep-learning methods for the proposed task, covering the main directions in object detection research. We present a quantitative and qualitative analysis of the performance for each method and for each main types of detectors. The dataset is publicly provided for further investigation in: https://github.com/pedrozamboni/individual_urban_tree_crown_detection (accessed on 21 June 2021).

2. Material and Methods

2.1. Image Dataset

We used two RGB high-resolution orthoimages with 5619 × 5946 pixels with a ground sample distance (GSD) equal to 10 cm of Campo Grande urban area, Mato Grosso do Sul state, Brazil (Figure 1). These are airborne images collected in 2013 by the city hall of Campo Grande. Campo Grande has 96.3% of urban households on public roads with the afforestation being recognized [43], in 2019, as a Tree City of the World by the Food and Agriculture Organization of the United Nations and the Arbor Day Foundation (Figure 1). A total of 161 plant species were identified on the streets of the municipality totaling more than 150 thousand trees [44]. *Licania tomentosa* is the most abundant species, representing 18.35%, followed by *Ficus benjamina* with 18.18%, and 66 species presented only one individual [44].

We manually annotated the orthoimages with rectangles (bounding boxes) in QGIS software. Since the object detection inputs are patches of images; the orthoimages were split into 220 non-overlapping patches of 512 × 512 pixels (51.20 × 51.20 m), which represents an area of 2621.44 m^2 per patch. The manually annotated polygons were converted into bounding boxes (Figure 2) where 3382 trees were identified as ground-truth. The object detection methods were trained to learn and predict the bounding box coordinates in the images given the ground-truth data. For our experiments, we divided the patches into training (60%), validation (20%), and test (20%) sets (Table 1). The validation set was used during an intermediate phase in training to select the model hyperparameters.

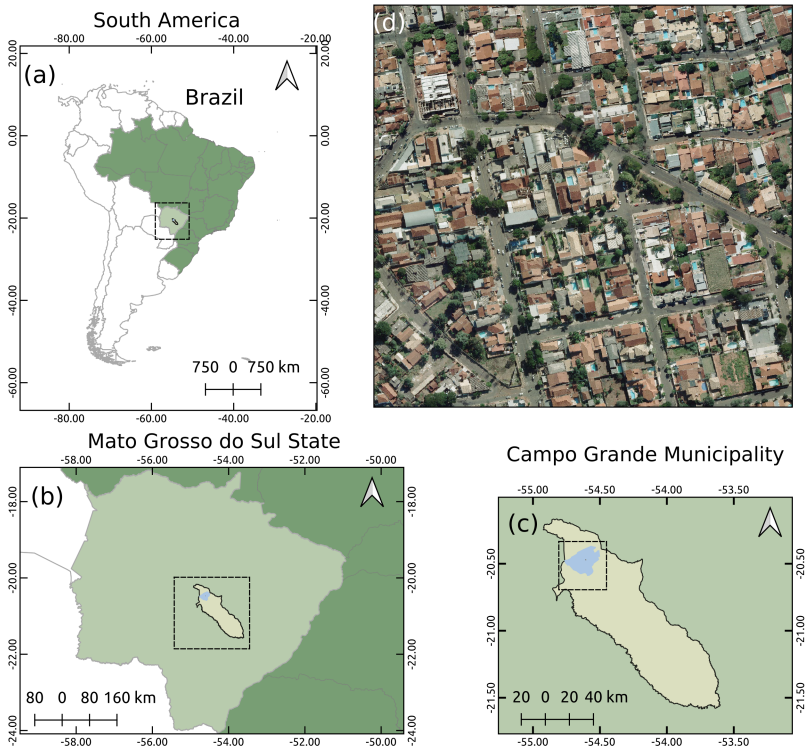


Figure 1. Study area in (**a**) South America and Brazil, (**b**) Mato Grosso do Sul, (**c**) Campo Grande, and (**d**) an example of an orthoimage used in this study.

Figure 2. Example of an annotated patch. The bounding boxes for each tree considered as ground−truth are represented in white.

Table 1. The number of image patches and trees annotated as ground-truth in each set. The training, validation, and test sets comprised 60%, 20%, and 20% of the image patches, respectively.

Set	nº of Patches	nº of Instances
Train	132	2124
Validation	44	582
Test	44	676
Total	220	3382

2.2. Individual Tree Crown Detection Approach

Our experiment was divided into two parts (Figure 3). First, 21 state-of-the-art algorithms were evaluated in this task. These methods cover the most diverse approaches currently used to detect objects, including anchor-based (one and two-stage) and anchor-free (Table 2). Second, we selected the best five methods in terms of AP_{50}. Faster R-CNN and RetinaNet were also included (among the best ones) since these methods are present as a standard baseline in the remote sensing literature.

We evaluated seven (top five + Faster R-CNN + RetinaNet) methods using hold out repeated four times to obtain a more robust evaluation due to bias-variance tradeoffs. In the holdout procedure with four repetitions, we randomly shuffled and split the data into three disjoint sets: training, validation, and test sets. Then, the average and standard deviation values for the AP_{50} considered the five repetitions for each method. We also performed One-Way ANOVA with the Holm–Bonferroni post hoc test to assess if these AP_{50} averages were statistically different. We used the methods implemented in the MMDetecion project source code proposed by Multimedia Laboratory [45]. MMDetecion is an open-source project available online on: https://github.com/open-mmlab/mmdetection (accessed on 21 June 2021).

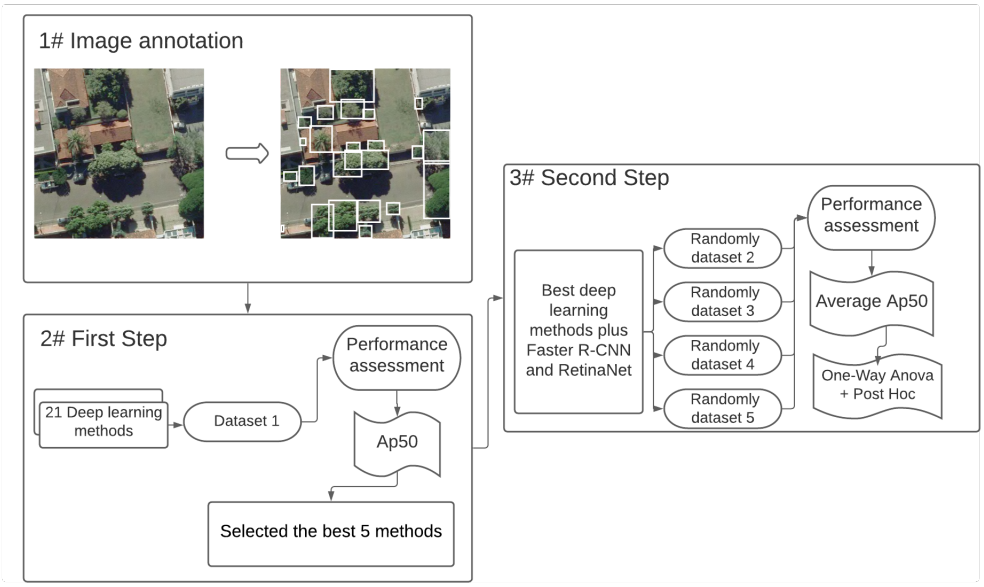


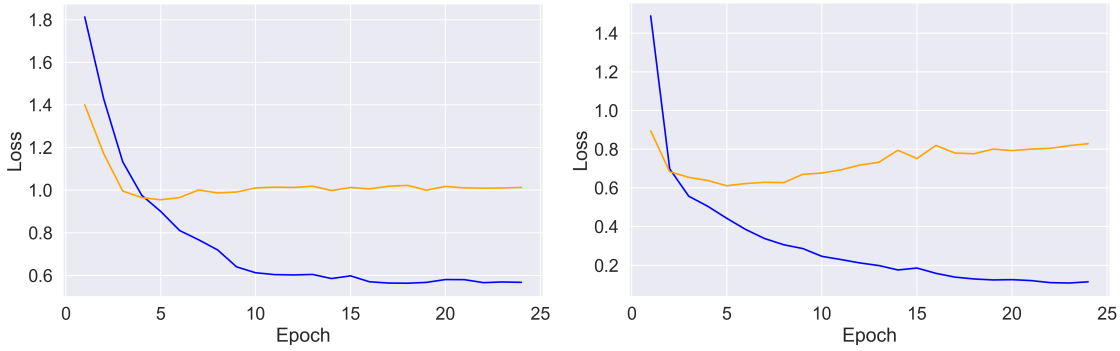
Figure 3. The workflow for individual tree crown detection. Initially, the images were annotated with bounding boxes. In the first step, 21 deep-learning methods were trained, and the best methods were selected based on the value of the third quartile plus Faster R-CNN and RetinaNet. In the second step, the selected methods were trained four more times with randomly shuffled datasets.

For the training, the backbone of all methods was initialized with pre-trained weights from the well-known ImageNet dataset. A stochastic gradient descent optimizer with a momentum of 0.9 and weight decay of 0.0001 was applied. The initial learning rate was empirically set to 0.00125. All the models were trained over 24 epochs. Figure 4 illustrates the training and validation loss curves. The training loss decreased rapidly after a few epochs and stabilized at the end. This indicates that the number of epochs was sufficient and that the learning rate was adequate. The training and testing procedures were conducted in Google Colaboratory with GPU.

Table 2. The object detection methods used in this study, including backbone, year of publication, reference, and type of method.

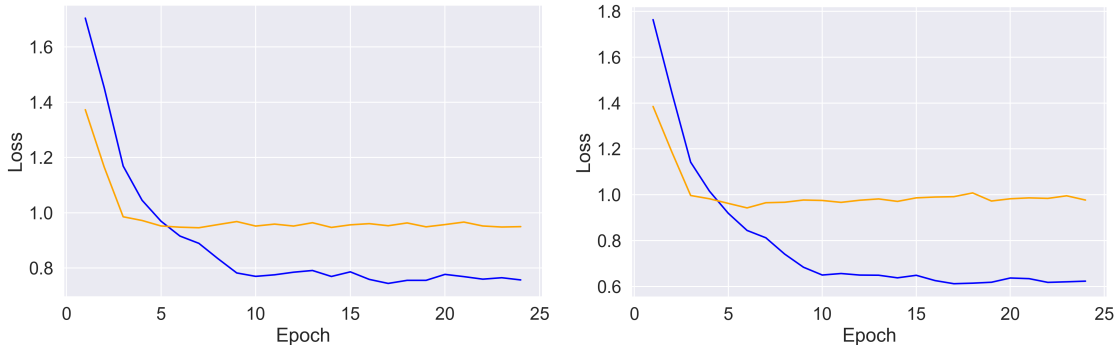
Method	Backbone	Year	Reference	Type
Faster R-CNN	X-101-64x4d-FPN-2x	2017	[46]	AB-TS
RetinaNet	X-101-64x4d-FPN-2x	2017	[47]	AB-OS
Mixed precision training	Faster R-CNN-R50-FPN-FP16-1	2017	[48]	AB-TS
Deformable ConvNets v2	Faster R-CNN X101-32x4d-FPN-dconv-c3-c5-1x	2018	[49]	AB-TS
YoloV3	DarkNet-53	2018	[50]	AB-OS
ATSS	R-101-FPN-1x	2019	[29]	AF
Weight Standardization	Faster R-CNN-X101-32x4d-FPN-gn-ws-all-1x	2019	[51]	AB-TS
CARAFE	Faster R-CNN-R50-FPN-_carafe-1x	2019	[52]	AB-TS
FSAF	X101-64x4d-FPN-1x	2019	[53]	AF
NAS-FPN	RetinaNet-R-50-NASFPN-crop640_50e	2019	[54]	AB-OS
FoveaBox	R-101-FPN-gn-head-mstrain-640-800-4x4-2x	2019	[55]	AF
Double Heads	dh-Faster R-CNN-R-50-FPN-1x	2019	[56]	AB-TS
Gradient Harmonized Single-stage Detector	X-101-64x4d-FPN-1x	2019	[57]	AB-OS
Empirical Attention	Faster R-CNN-R-50-FPN-attention-1111-dcn-1x	2019	[58]	AB-TS
DetectoRS	rcnn-R-50-1x	2020	[59]	AB-MS
VarifocalNet (1)	R-101-FPN-1x	2020	[60]	AF
VarifocalNet (2)	X-101-64x4d-FPN-mdconv-c3-c5-mstrain-2x	2020	[60]	AF
SABL	cascade rcnn-r101-FPN-1x	2020	[61]	AB-OS
Generalized Focal Loss	X-101-32x4d-FPN-dconv-c4-c5-mstrain-2x	2020	[62]	AB-OS
Probabilistic Anchor Assignment	R-101-FPN-2x	2020	[63]	AB-OS
Dynamic R-CNN	R-50-FPN-1x	2020	[64]	AB-TS

AF: anchor-free; AB-OS: anchor-based one-stage; AB-TS: anchor-based two-stage; and AB-MS: anchor-based multi-stage.



(**a**) Faster R-CNN

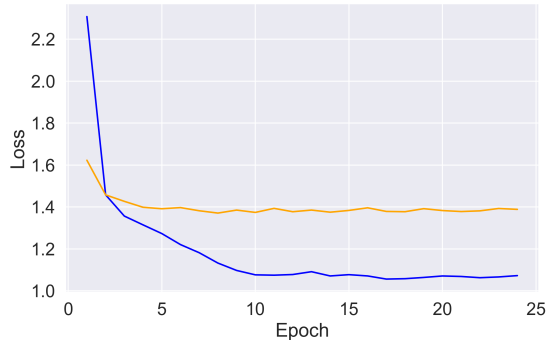
(**b**) RetinaNet



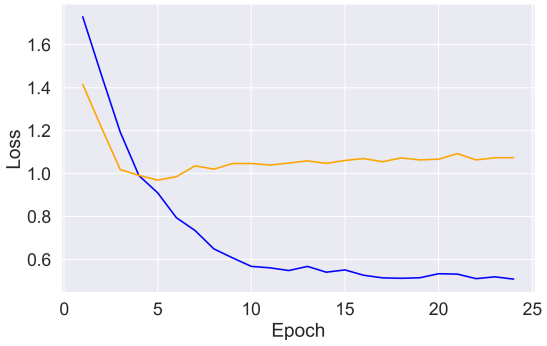
(**c**) Mixed precision trainin

(**d**) Deformable ConvNets v2

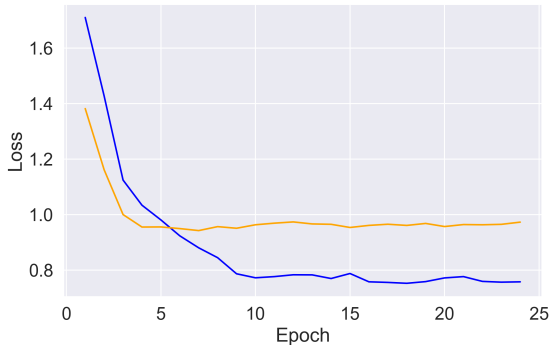
Figure 4. *Cont.*



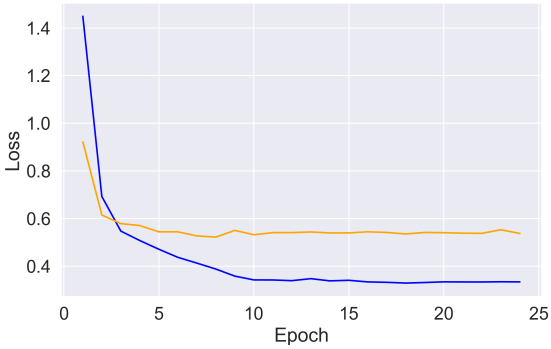
(**e**) ATSS



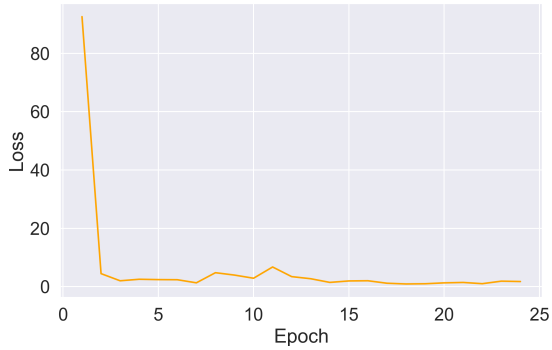
(**f**) Weight Standardization



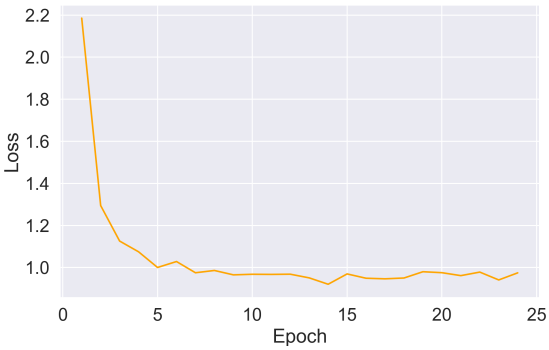
(**g**) CARAFE



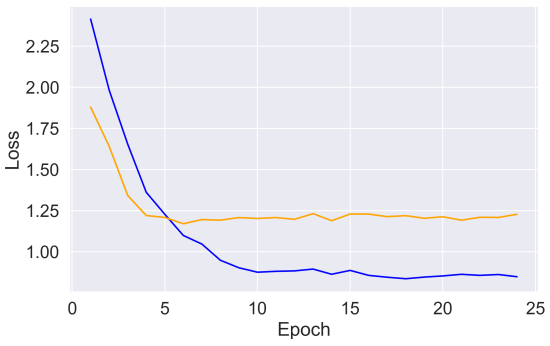
(**h**) FSAF



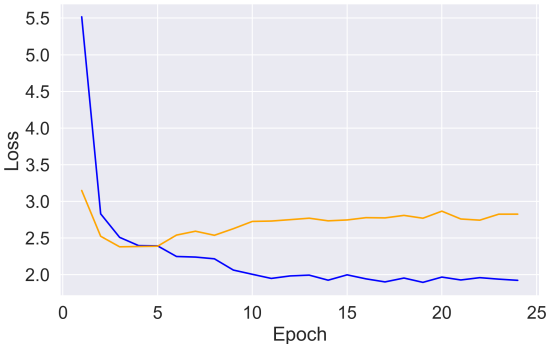
(**i**) NAS-FPN



(**j**) FoveaBox

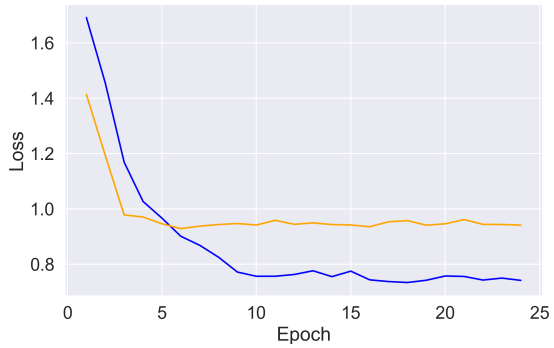


(**k**) Double Head

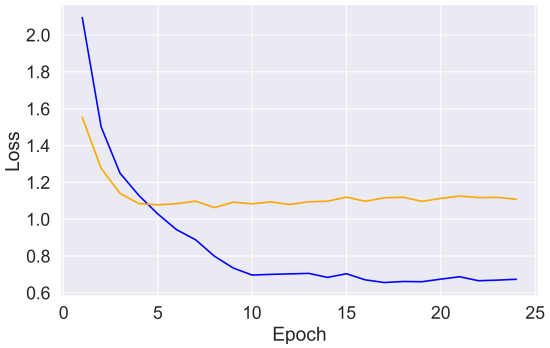


(**l**) Gradient Harmonized Single-stage Detector

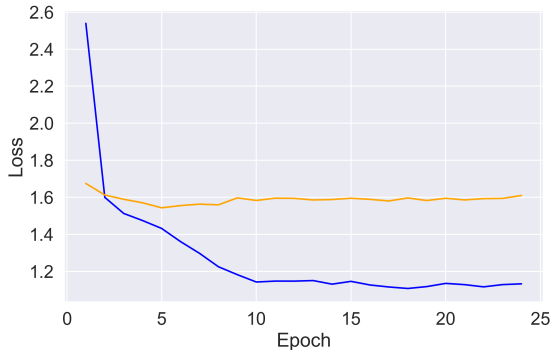
Figure 4. *Cont.*



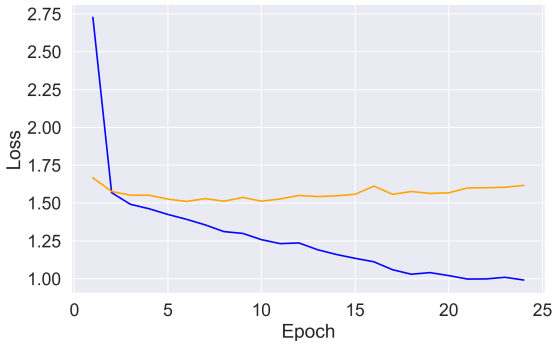
(**m**) Empirical Attention



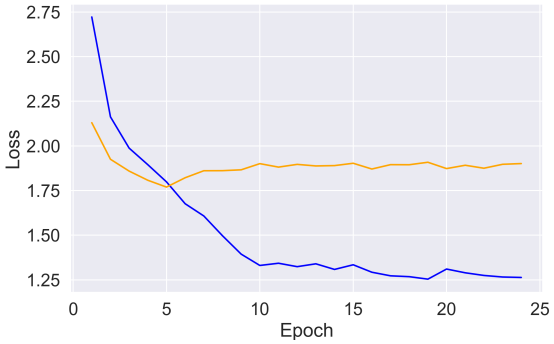
(**n**) DetectoRS



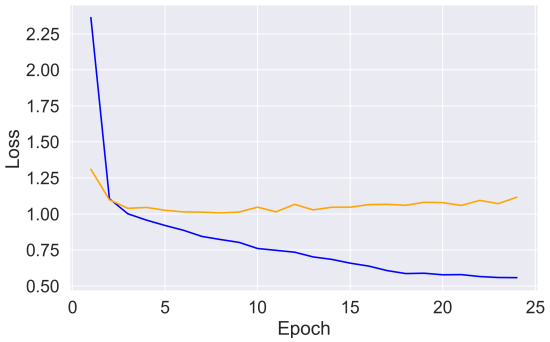
(**o**) VarifocalNet (1)



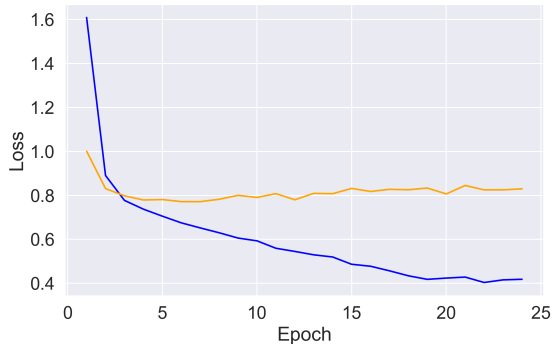
(**p**) VarifocalNet (2)



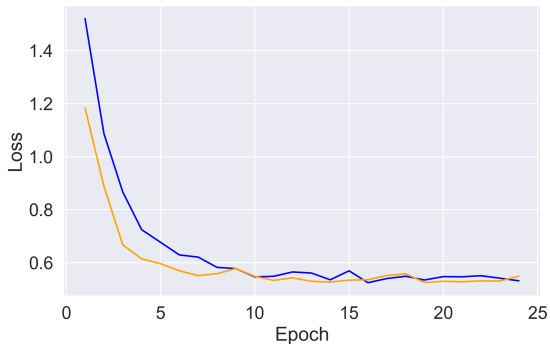
(**q**) SABL



(**r**) Generalized Focal Loss

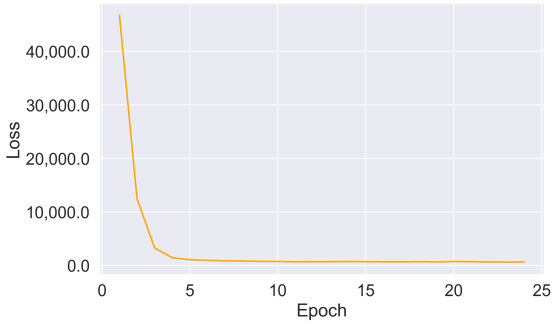


(**s**) Probabilistic Anchor Assignment



(**t**) Dynamic R-CNN

Figure 4. *Cont.*



(**u**) YoloV3

Figure 4. Loss curves for training (blue) and validation (orange) for each object detection method. For YoloV3, NAS-FPN, and FoveaBox, we only show the validation curves since the log for these two methods did not return the training loss.

2.3. Performance Evaluation

We assessed the overall performance of the methods using the Average Precision (AP). The AP is the area under the precision–recall curve. The precision and recall were estimated using Equations (1) and (2). To obtain the precision and recall values, we defined the Intersection Over Union (IoU). The IoU is the relation between the overlapping area and the union area between the predicted and ground-truth bounding box. When a predicted bounding box reaches a greater IoU value than the threshold, the prediction is classified as true positive (TP). On the other hand, if the IoU value is below the threshold, the prediction is a false positive (FP). Further, if a ground-truth bounding box is not detected by any prediction, it is considered a false negative (FN). We used IoU thresholds of 0.5 (AP_{50}), the most common IoU thresholds used in computer vision.

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

2.4. Statistical Analysis

We performed the Shapiro–Wilk test to check the normality of the data. All samples reported *P*-values greater than 0.05; therefore, we cannot reject the null hypothesis that the samples were normally distributed. We also conducted Bartlett's test for equal variances. The *p*-values were greater than 0.05, failing to reject the null hypothesis, and we can, thus, assume that the samples had equal variance. As for data independence, the samples were randomly obtained from the set.

An ANOVA test with the Holm–Bonferroni post hoc test was performed in order to assess if the means of best methods (top five + Faster R-CNN + RetinaNet) were statistically different. For the ANOVA, the *P*-value was compared to the significance level ($\alpha = 0.05$) to assess the null hypothesis. If the p-value was equal to or less than the significance level, there were statistically significant differences between the means. However, the ANOVA test did not identify differences between pairs but indicated that not all AP_{50} means were equal. Therefore, after rejecting the null hypothesis using ANOVA, the evaluation proceeded using the post hoc test to identify the differences between pairs of algorithms. We used the Holm–Bonferroni as a Post hoc to run the assessment of the experiment.

3. Results

Here, we present the results of our experiments. First, we performed a quantitative and qualitative analysis for all 20 methods. Therefore, the results were separated by the type of method, i.e., anchor-based (AB-OS: one-stage; AB-TS: two-stage; and AB-MS: multi-stage) and anchor-free (AF). In the quantitative analysis, we evaluated the methods using the IoU threshold of 0.5 (AP_{50}). The qualitative analysis was conducted to identify in which

situations the models had good and bad performance over different conditions, such as shadow and overlap by other objects.

Later, we present the results for the second part of the experiments with the top five models, Faster R-CNN, and RetinaNet. The images presented in this section were from the test set; therefore, the images provide a better indication of the performance of the models. Even though different areas (with different tree species, tree crown sizes, and distributions) were used to train and test the model, the two images are from the same city. Thus, it is not possible to comment on the capacity for generalizability of these models on different datasets.

3.1. Anchor-Based (AB) Detectors

In this section, we discuss the performance of the one, two, and multi-stage anchor-based detectors. For one-stage methods, the average AP_{50} was 0.657 ± 0.032. Table 3 shows the test set results for the one-stage methods. We observed that the Gradient Harmonized Single-stage Detector outperformed all the others in AP_{50}. The increase in performance ranged from 1.4% to 10%. RetinaNet, NAS-FPN, and SABL provided similar results. Probabilistic Anchor Assignment and Generalized Focal Loss presented similar performances, and YoloV3 was the worst method.

Table 4 shows the performance for the two-stage and multi-stage (DetectoRS) methods. During the test, on average, the two-stage and multi-stage methods reached 0.669 ± 0.023 for AP_{50}. The Double Heads method achieved the best performance for these methods when analyzing the AP_{50}, outperforming the others from 0.2% to 6.8%. The CARAFE and Empirical Attention methods obtained performances similar to Double Heads in terms of the AP_{50}. Faster R-CNN, DetectoRS, Deformable ConvNets v2, and Dynamic R-CNN reached similar performances, and Weight Standardization provided the worst results.

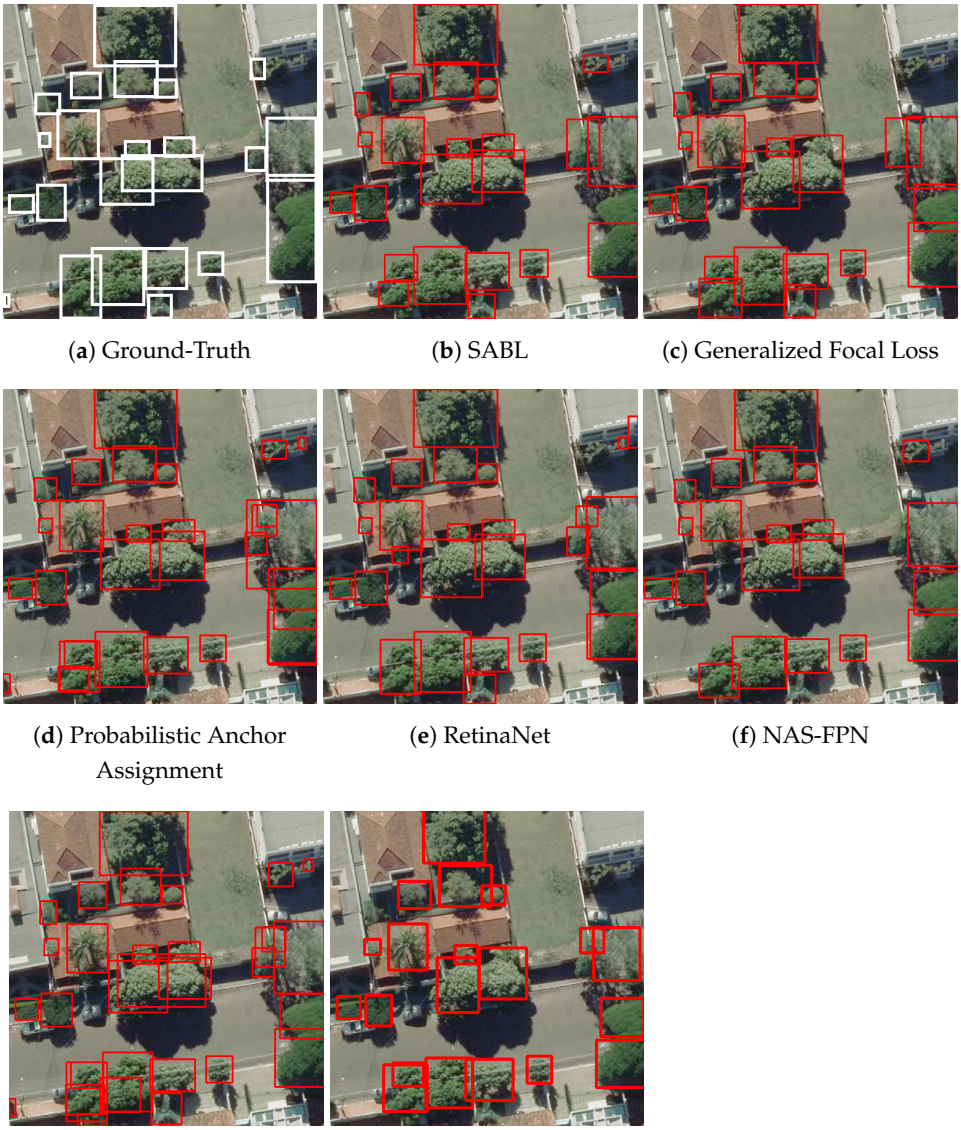
Figures 5 and 6 show the tree detection achieved using the one-stage methods. As we can see in Figure 5, for smaller tree crowns and even medium-sized ones, the one-stage methods had good assertiveness. However, for larger crowns (Figure 6), we observed a decrease in the performance, with Probabilistic Anchor Assignment being the unique method with good performance. For more irregular trees, where the crown did not have a circular shape, the methods usually detected more than one bounding box for a given ground-truth annotation. In areas where there were large agglomeration of trees, the methods did not detect the trees or detected only a part.

Table 3. Performance of the one-stage methods for the test set using AP_{50}.

Model	Test Set AP_{50}
SABL	0.661
Generalized Focal Loss	0.677
Probabilistic Anchor Assignment	0.677
RetinaNet	0.650
NAS-FPN	0.658
YoloV3	0.591
Gradient Harmonized Single-stage Detector	0.691

Table 4. Performance of the two-stage and multi-stage (DetectoRS) methods for the test set using AP_{50}.

Model	Test Set AP_{50}
Faster R-CNN	0.660
DetecoRS	0.651
Weight Standardization	0.631
Deformable ConvNets v2	0.657
CARAFE	0.697
Dynamic R-CNN	0.655
Double Heads	0.699
Mixed precision training	0.679
Empirical Attention	0.690



(**a**) Ground-Truth (**b**) SABL (**c**) Generalized Focal Loss

(**d**) Probabilistic Anchor Assignment (**e**) RetinaNet (**f**) NAS-FPN

(**g**) Gradient Harmonized Single-stage Detector (**h**) YoloV3

Figure 5. Examples of tree detection by the one-stage methods.

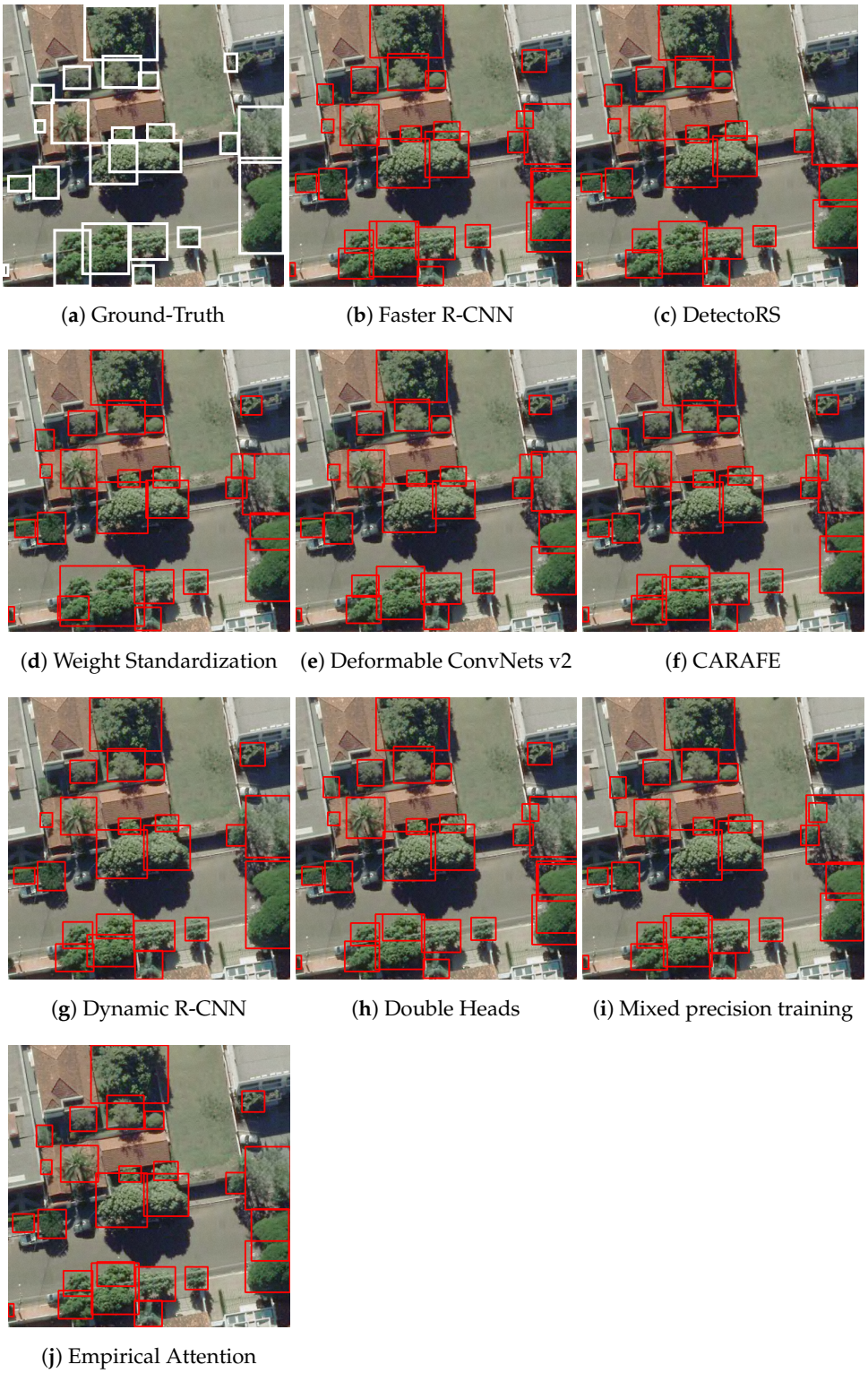
(**a**) Ground-Truth (**b**) SABL (**c**) Generalized Focal Loss

(**d**) Probabilistic Anchor Assignment (**e**) RetinaNet (**f**) NAS-FPN

(**g**) Gradient Harmonized Single-stage Detector (**h**) Gradient Harmonized Single-stage Detector

Figure 6. Examples of tree detection in areas with high density using the one-stage methods.

Figures 7 and 8 present the detection for two-stage methods. Similar to the one-stage methods, the two-stage methods presented good performance in detecting smaller and medium-sized tree crowns. For larger ones and in areas with a greater agglomeration of objects (Figure 8), the two-stage methods performed substantially better than the one-stage methods. Thus, these methods appeared to generalize the problem better with better assertiveness in detecting the tree crowns in more complex scenes. Further, we observed that the presence of shadow did not cause a great decrease in the detection. We observed that the main challenge was to detect single trees with larger crowns and areas where the limits of each object were not clear. In such cases (Figures 6 and 9), even for the human eye, it is difficult to separate the trees from each other.



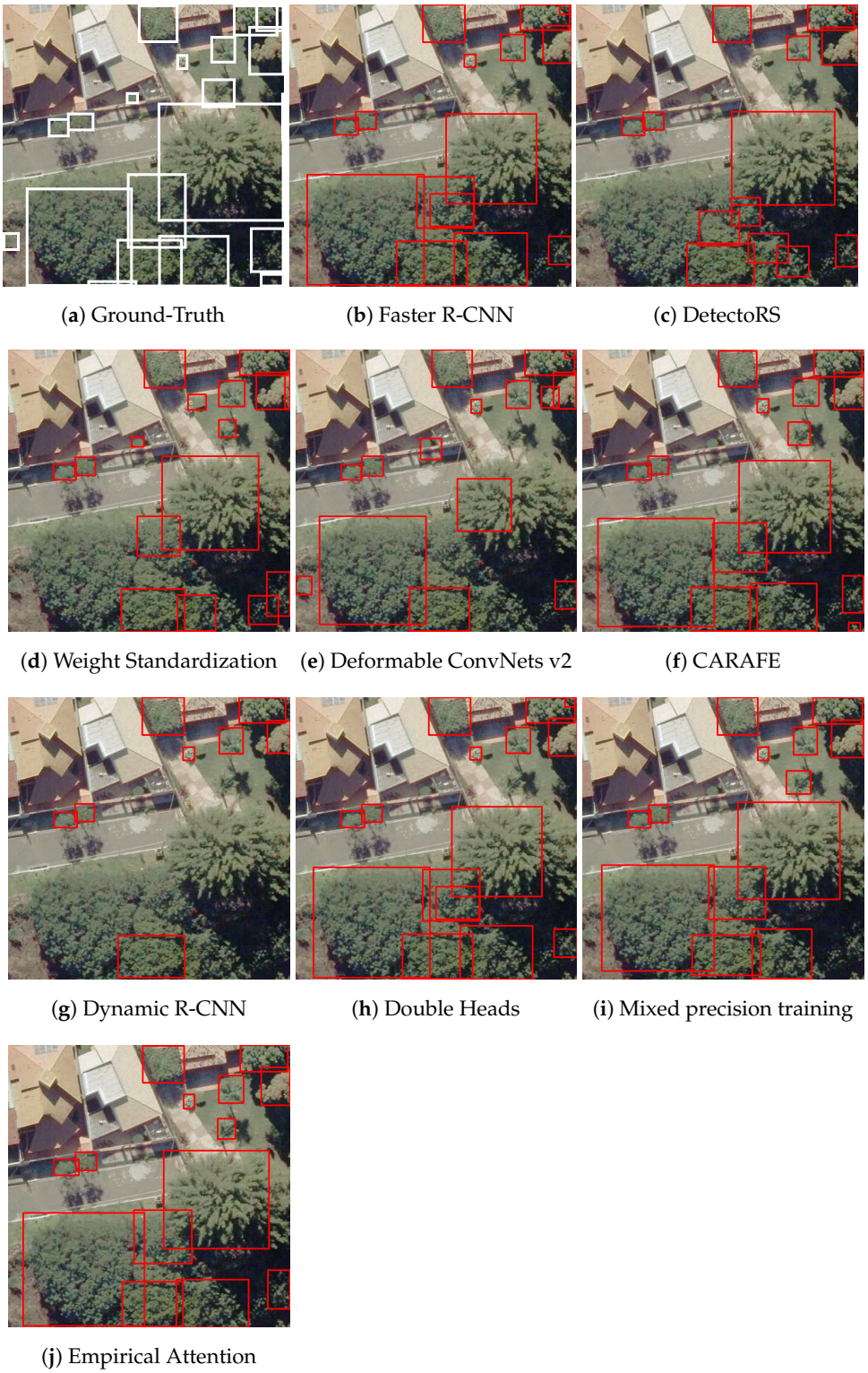
(**a**) Ground-Truth (**b**) Faster R-CNN (**c**) DetectoRS

(**d**) Weight Standardization (**e**) Deformable ConvNets v2 (**f**) CARAFE

(**g**) Dynamic R-CNN (**h**) Double Heads (**i**) Mixed precision training

(**j**) Empirical Attention

Figure 7. Examples of tree detection using the two-stage methods.



(**a**) Ground-Truth (**b**) Faster R-CNN (**c**) DetectoRS

(**d**) Weight Standardization (**e**) Deformable ConvNets v2 (**f**) CARAFE

(**g**) Dynamic R-CNN (**h**) Double Heads (**i**) Mixed precision training

(**j**) Empirical Attention

Figure 8. Examples of tree detection in areas with high density using the two-stage methods.

(**a**) Ground-Truth (**b**) ATSS (**c**) VarifocalNet (1)

(**d**) VarifocalNet (2) (**e**) FSAF (**f**) FoveaBox

Figure 9. Examples of tree detection in areas with high density using the anchor-free methods.

3.2. Anchor-Free (AF) Detectors

The results obtained for anchor-free (AF) methods are described in Table 5. In the test, the anchor-free methods achieved an average performance of 0.686 ± 0.014. FSAF reached the best performance in terms of the AP_{50} with 0.701. This demonstrated a superior performance over the others, ranging from 0.9% to 3.7%. FoveaBox, ATSS, and VarifocalNet (2) had similar results in terms of the AP_{50}, and VarifocalNet (1) had the worst in performance.

Anchor-free methods demonstrated similar behavior when compared with the one-stage methods. These models performed well for small trees (Figure 10). For occluded objects and more irregular tree crowns, we observed a decrease in the performance, with multiple detections and the detection of only part of the object. For areas with larger tree crowns and more agglomerations, the performance also decreased. VarifocalNet (2) was the only method that managed to produce relatively good detection in the most complex images (Figure 9). This highlights that these areas with larger canopies and more agglomerations are the main challenges for the methods.

Table 5. Performance of the anchor-free (AF) methods on the test set using AP_{50}.

Model	Test Set AP_{50}
ATSS	0.692
VarifocalNet (1)	0.664
VarifocalNet (2)	0.683
FSAF	0.701
FoveaBox	0.692

(**a**) Ground-Truth (**b**) ATSS (**c**) VarifocalNet (1)

(**d**) VarifocalNet (2) (**e**) FSAF (**f**) FoveaBox

Figure 10. Example of tree detection using anchor-free methods.

3.3. Analysis of the Best Methods

Here, we present the best five methods considering AP_{50}, which were FSAF, Double Heads, CARAFE, ATSS, and FoveaBox. We also included Faster R-CNN and RetinaNet, since these two are commonly used in remote sensing. We noticed that none of the five best were a one-stage method. As seen in the previous sections, the anchor-free methods showed better average performance compared with the one and two-stage methods in terms of the AP_{50}. Figure 11 shows the box plot for the methods. Figures 12–15 show some results for the best methods.

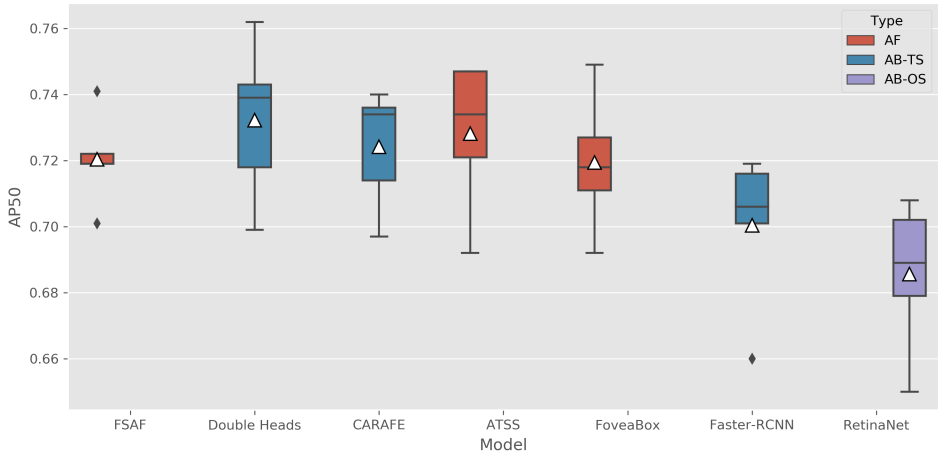


Figure 11. Boxplot for the best five methods plus Faster R-CNN and RetinaNet.

We observed that Double Heads, a two-stage method, achieved the best average AP_{50} (0.732), with differences ranging from 0.4%, when compared to ATSS, and 4.6%, when compared to RetinaNet. ATSS and CARAFE achieved similar values with averages AP_{50} of 0.728 and 0.724, respectively, which were close to Double Heads. FSAF and FoveaBox had slightly worse performances with average AP_{50} values of 0.720 and 0.719. Faster

R-CNN (average AP_{50} of 0.700) and RetinaNet (average AP_{50} of 0.686) obtained the worst average results.

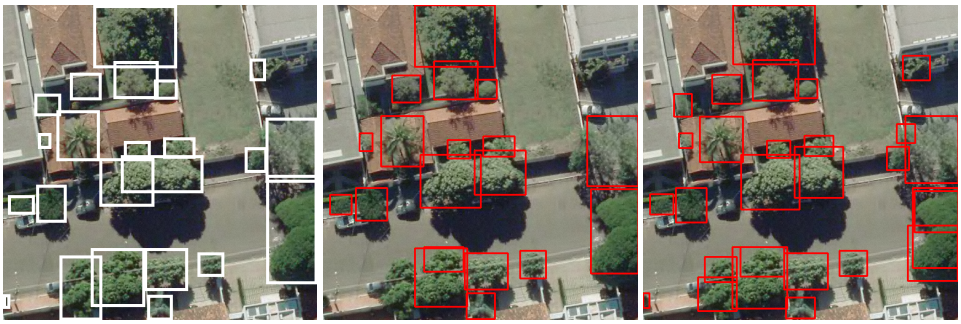
Despite the performance analysis conducted using the AP_{50}, we performed One-Way ANOVA to assess if the averages of the AP_{50} values of the best methods were significantly different. One-Way ANOVA for the top five, Faster R-CNN, and RetinaNet indicated a *P*-value of 0.019, which is less than the significance level ($\alpha = 0.05$). Therefore, we can reject the null hypothesis that the results were similar. We continued the evaluation using a post hoc test to identify differences between pairs of algorithms.

A simple strategy in multiple comparisons is to use $\frac{\alpha}{m}$ to evaluate the *P*-value, which is the Bonferroni correction. However, this value is rigorous and can lead to the rejection of a true null hypothesis (Type I error). Holm–Bonferroni adjusts the rejection criteria for each comparison reducing the chance of a Type I error. The Holm–Bonferroni sorts the *p*-values in increasing order creating a rank of $P_1, ..., P_k, ..., P_m$ and compares them with $\frac{\alpha}{m+1-k}$ where k is the ranking order in the comparison. When $P_k < \frac{\alpha}{m+1-k}$ is false, the procedure stops, and we cannot reject the null hypothesis of the subsequent P_k. Table 6 shows the results of the Holm–Bonferroni test. For simplicity, the column *P*-value corr represents this comparison, and, when its value is lower than 0.05, we can reject the null hypothesis.

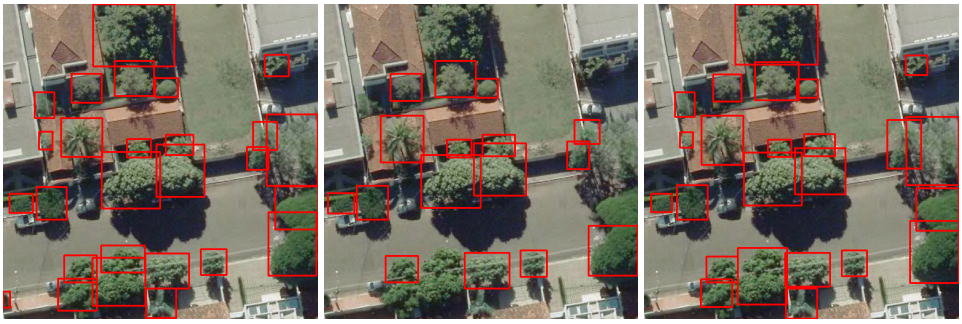
The results indicate that the results of RetinaNet were significantly different from ATSS, CARAFE, and Double Heads. Further, a comparison between the other methods showed no statistically significant differences. The test indicates that RetinaNet, among the tested models, was not indicated for the proposed task.

Table 6. Multiple comparison Holm–Bonferroni test (FWER = 0.05, alphacSidak = 0.00, and alphacBonf = 0.002).

Method 1	Method 2	Stat.	*p*-Value	*p*-Value Corr	Reject
ATSS	CARAFE	1.2172	0.2904	1.0	False
ATSS	Double Heads	−0.9589	0.3919	1.0	False
ATSS	FSAF	1.2161	0.2908	1.0	False
ATSS	Faster R-CNN	5.0387	0.0073	0.1166	False
ATSS	FoveaBox	1.2631	0.2752	1.0	False
ATSS	RetinaNet	37.9511	0.0	0.0001	True
CARAFE	Double Heads	−2.2274	0.0899	0.8987	False
CARAFE	FSAF	0.6059	0.5773	1.0	False
CARAFE	Faster R-CNN	3.8948	0.0176	0.2643	False
CARAFE	FoveaBox	0.6554	0.548	1.0	False
CARAFE	RetinaNet	9.7542	0.0006	0.0124	True
Double Heads	FSAF	1.2605	0.276	1.0	False
Double Heads	Faster R-CNN	3.7234	0.0204	0.2654	False
Double Heads	FoveaBox	1.2241	0.2881	1.0	False
Double Heads	RetinaNet	8.9582	0.0009	0.0163	True
FSAF	Faster R-CNN	3.2573	0.0312	0.3739	False
FSAF	FoveaBox	0.2654	0.8038	1.0	False
FSAF	RetinaNet	6.0	0.0039	0.0699	False
Faster R-CNN	FoveaBox	−3.8623	0.0181	0.2643	False
Faster R-CNN	RetinaNet	2.737	0.0521	0.5727	False
FoveaBox	RetinaNet	5.4165	0.0056	0.0957	False



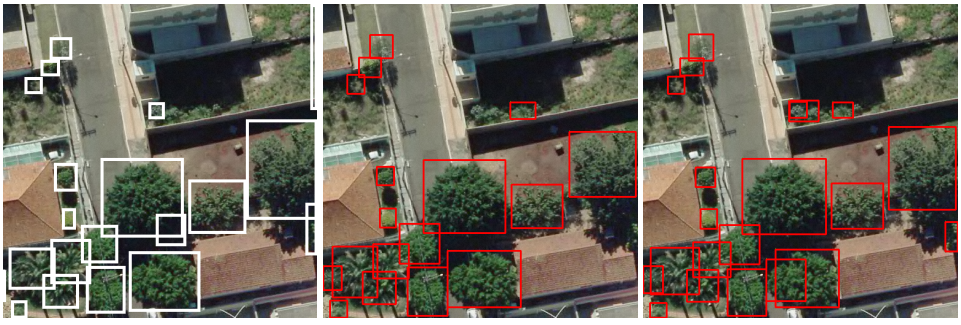
(**a**) Ground-Truth (**b**) FSAF (**c**) Double Heads



(**d**) CARAFE (**e**) ATSS (**f**) FoveaBox

(**g**) Faster R-CNN (**h**) RetinaNet

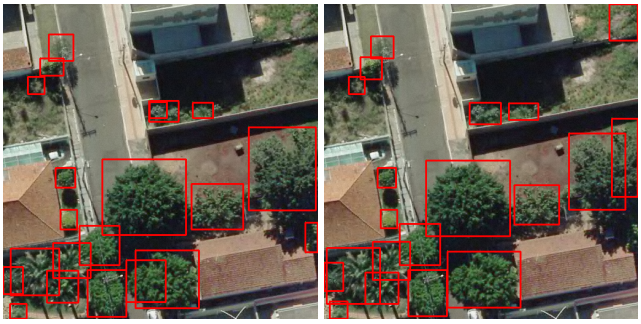
Figure 12. Tree detection with the top five methods, RetinaNet, and Faster R-CNN: performance with small and medium trees from several species.



(**a**) Ground-Truth (**b**) FSAF (**c**) Double Heads

Figure 13. *Cont.*

(**d**) CARAFE (**e**) ATSS (**f**) FoveaBox



(**g**) Faster R-CNN (**h**) RetinaNet

Figure 13. Tree detection with the top five methods, RetinaNet, and Faster R-CNN: performance in a high density scenario.

(**a**) Ground-Truth (**b**) FSAF (**c**) Double Heads

(**d**) CARAFE (**e**) ATSS (**f**) FoveaBox

Figure 14. *Cont.*

(**g**) Faster R-CNN (**h**) RetinaNet

Figure 14. Tree detection with the top five methods, RetinaNet, and Faster R-CNN: performance considering big trees.

(**a**) Ground-Truth (**b**) FSAF (**c**) Double Heads

(**d**) CARAFE (**e**) ATSS (**f**) FoveaBox

(**g**) Faster R-CNN (**h**) RetinaNet

Figure 15. Tree detection with the top five methods, RetinaNet, and Faster R-CNN: performance in a challenging illumination scenario with shadows.

4. Discussion

Anchor-based one-stage methods achieved the worst average precision (0.657). Gradient Harmonized Single-Stage Detector was the best with an AP_{50} of 0.691, and YoloV3 was the worst with 0.591 precision. The commonly used RetinaNet had an AP_{50} of 0.650, being the second worst one-stage method.

A previous study [39] implemented RetinaNet to detect Phoenix palms with the best AP value of 0.861; however, the authors split the dataset only into training and validation sets and used a score threshold of 0.2 and an IoU of 0.4. This may be lead to better performance. They also considered only one tree species as the target. Roslan et al. [41] used RetinaNet to detect individual trees and achieved superior results with a precision of 0.796, and similar results were found by [42]. Refs. [41,42] utilized images of non-urban areas (tropical forests). In our experiments, RetinaNet provided less accurate results among the one-stage methods.

Anchor-based two-stage methods had the second best highest average AP_{50} of 0.669. Double Heads had the best performance among these methods, and DetecoRS had the worst. Faster R-CNN and RetinaNet (baseline) had similar results. Santos et al. [19] investigated both methods and concluded that RetinaNet outperformed Faster R-CNN and YOLOv3 in the detection of a single tree species, achieving an AP_{50} higher than 0.9. Wu et al. [65] proposed a model that used Faster R-CNN as a detector in a hybrid model to detect and segment apple tree crowns in UAV imagery.

In the detection section, the authors achieved high-precision for the task. However, these authors considered only one tree species and used images with higher resolution with small variation in scale. These factors may lead to better performance for the methods. In the other hand, the anchor-free methods had the best average precision with 0.686. FSAF, ATSS, and FoveaBox stood out among others. The results for the anchor-free methods corroborate with the study of Gomes et al. [23], where ATSS also outperformed Faster R-CNN and RetinaNet by about 4%.

Previous studies [28,29] also reported that two-stage methods had higher performance over the one-stage methods, which corroborates our findings. We found that anchor-free methods performed similarly to anchor-based two-stage methods. This behavior has already been reported in the literature. The advantage of anchor-free detectors is the removal of the hyper-parameters associated with the anchors, implying potentially better generalization [29]. RetinaNet (one-stage) and Faster R-CNN (two-stage) showed relatively poor results when compared with the top five methods selected. It is important to note that these two methods have been reported in the literature as having superior performance in other remote sensing applications [19,24].

As previously presented, our experiment aimed to detect all the trees in urban scenes. Compared to the previous work that only targeted a single tree species, our objective was considerably more challenging. First, urban scenes are more complex and heterogeneous. Second, our dataset presented various tree species and tree crown sizes with overlap between objects, shadows, and other situations. In the Campo Grande city urban area, there are 161 tree species and more than 150 thousand trees. Thus, this complexity in the task led to better performance of the two-stage anchor-based methods, especially in more challenging images as can be seen in Figure 15. These methods first filter the region that may contain an object and then they eliminate most negative regions [27]. Comparatively, [66] proposed the identification of trees in urban areas using street-level imagery and Mask-RCNN [67]. They found an AP_{50} between 0.620 and 0.682.

5. Conclusions

Here, we presented a large assessment of the performance of novel deep-learning methods to detect single tree crowns in urban high-resolution aerial RGB images. We evaluated a total of 21 object detection methods, including anchor-based (one, two, and multi-stage) and anchor-free detectors in a remote sensing relevant application. We provided a quantita-

tive and qualitative analysis of each type of method. We also provided a statistical analysis of the best methods as well as RetinaNet and Faster R-CNN.

Our results indicate that the anchor-free methods showed the highest average AP_{50}, followed by anchor-based two-stage and anchor-based one-stage. Our findings suggest that the best methods for the current task were the two-stage anchor-based and anchor-free detectors. For the one-stage anchor-based detectors, only the Gradient Harmonized Single-stage Detector performed slightly worse than the best methods. This may be an indication that one-stage methods are not recommended for the proposed task. Meanwhile, the two-stage (Double Heads and CARAFE) and anchor-free (FSAF, ATSS, and FoveaBox) detectors achieved superior performance, which is the study's suggestion for urban single tree crown detection.

Our experimental results demonstrated that RetinaNet, one of the most used methods in remote sensing, did not have satisfactory performance for the proposed task and unperformed several of the best methods (ATSS, CARAFE, and Double Heads). This may indicate that this method is not suitable for the proposed task. Faster R-CNN had slightly inferior results compared with the best methods; however, no statistically significant difference was found. However, it is worth mentioning that research aimed at detecting single trees in an urban environment is still incipient, and further investigation regarding the most appropriate techniques is needed. In our work, we set out to detect all tree crowns in an urban environment. This task is considerably more complex than detecting specific species or types of trees since there will be a greater variety of trees. Likewise, images from the urban environment are more complex and challenging than rural environments as they present a more heterogeneous environment.

Our work demonstrates the potential of the existing techniques based on deep learning by leveraging the application of different methods for remote sensing data. This study may contribute to innovations in remote sensing based on deep-learning object detection. The majority of the research applying deep learning in remote sensing was done using methods dated before 2018 (e.g., Faster R-CNN and RetinaNet), and, with the development of new methods, it is essential to evaluate their performance in these tasks. The development of techniques capable of accurately detecting trees using RGB images is essential in preserving and maintaining forest systems. These tools are essential for cities, where accelerated population growth and climate change are becoming significant threats. Future works will focus on developing a method capable of working with high density objects. We also intend to increase the size of the dataset with images from different cities in order to obtain models with better generalization capabilities.

Author Contributions: Conceptualization, P.Z., J.M.J. and W.N.G.; methodology, W.N.G., J.d.A.S. and J.M.J.; software, J.d.A.S., W.N.G. and P.Z.; formal analysis, P.Z., J.M.J.; resources, J.M.J., J.d.A.S. and W.N.G.; data curation, P.Z., G.T.M.; writing—original draft preparation, P.Z. and J.M.J.; writing—review and editing, J.M.J., J.d.A.S., E.T.M., G.T.M., K.N. and W.N.G.; supervision, project administration, and funding acquisition, J.M.J. and W.N.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by CNPq (p: 433783/2018-4, 303559/2019-5 and 304052/2019-1), CAPES Print (p: 88881.311850/2018-01) and Fundect (p: 59/300.066/2015).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this work and the source code used for training and evaluation of the models are available at https://github.com/pedrozamboni/individual_urban_tree_crown_detection (accessed on 30 May 2020).

Acknowledgments: The authors acknowledge the support of the UFMS (Federal University of Mato Grosso do Sul) and CAPES (Finance Code 001).

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

References

1. McDonald, R.I.; Mansur, A.V.; Ascensão, F.; Colbert, M.; Crossman, K.; Elmqvist, T.; Gonzalez, A.; Güneralp, B.; Haase, D.; Hamann, M.; et al. Research gaps in knowledge of the impact of urban growth on biodiversity. *Nat. Sustain.* **2020**, *3*, 16–24. [CrossRef]
2. Ke, J.; Zhang, J.; Tang, M. Does city air pollution affect the attitudes of working residents on work, government, and the city? An examination of a multi-level model with subjective well-being as a mediator. *J. Clean. Prod.* **2021**, *265*. [CrossRef]
3. Khomenko, S.; Cirach, M.; Pereira-Barboza, E.; Mueller, N.; Barrera-Gómez, J.; Rojas-Rueda, D.; de Hoogh, K.; Hoek, G.; Nieuwenhuijsen, M. Premature mortality due to air pollution in European cities: A health impact assessment. *Lancet Planet. Health.* **2021**. [CrossRef]
4. Abass, K.; Buor, D.; Afriyie, K.; Dumedah, G.; Segbefi, A.Y.; Guodaar, L.; Garsonu, E.K.; Adu-Gyamfi, S.; Forkuor, D.; Ofosu, A.; et al. Urban sprawl and green space depletion: Implications for flood incidence in Kumasi, Ghana. *Int. J. Disaster Risk Reduct.* **2020**, *51*. [CrossRef]
5. The Human Cost of Weather Related Disasters (1995–2015): Center For Research on the Epidemiology of Disasters (CRED). 2015. Available online: https://www.unisdr.org/2015/docs/climatechange/COP21_WeatherDisastersReport_2015_FINAL.pdf (accessed on 8 April 2021).
6. Li, H.; Zhang, S.; Qian, Z.; Xie, X.H.; Luo, Y.; Han, R.; Hou, J.; Wang, C.; McMillin, S.E.; Wu, S.; et al. Short-term effects of air pollution on cause-specific mental disorders in three subtropical Chinese cities. *Environ. Res.* **2020**, *191*. . [CrossRef] [PubMed]
7. Heinz, A.; Deserno, L.; Reininghaus, U. Urbanicity, social adversity and psychosis. *World Psychiatry* **2013**, *12*, 187–197. [CrossRef] [PubMed]
8. IPCC. *Summary for Policymakers. Climate Change 2013: The Physical Science Basis Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2013.
9. IPCC. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*; Field, C.B., Ed.; Cambridge University Press: Cambridge, UK, 2012.
10. Fasihi, H.; Parizadi, T. Analysis of spatial equity and access to urban parks in Ilam, Iran. *J. Environ. Manag.* **2020**, *15*. [CrossRef] [PubMed]
11. (UN), U.N. *Transforming Our World: The 2030 Agenda for Sustainable Development*; Cambridge University Press: Cambridge, UK, 2015.
12. Roy, S.; Byrne, J.; Pickering, C. A systematic quantitative review of urban tree benefits, costs, and assessment methods across cities in different climatic zones. *Urban For. Urban Green.* **2012**, *11*, 351–363. [CrossRef]
13. Endreny, T.A. Strategically growing the urban forest will improve our world. *Nat. Commun.* **2018**, *9*. [CrossRef]
14. Fassnacht, F.E.; Latifi, H.; Stereńczak, K.; Modzelewska, A.; Lefsky, M.; Waser, L.T.; Straub, C.; Ghosh, A. Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* **2016**, *186*, 64–77. [CrossRef]
15. Padayaahce, A.; Irlich, U.; Faulklner, K.; Gaertner, M.; Proches, S.; Wilson, J.; Rouget, M. How do invasive species travel to and through urban environments? *Biol. Invasions* **2017**, *19*, 3557–3570.
16. Nielsen, A.; Ostberg, J.; Delshammar, T. Review of Urban Tree Inventory Methods Used to Collect Data at Single-Tree Level. *Arboric. E Urban For.* **2014**, *40*, 96–111.
17. Wagner, F.; Ferreira, M.; Sanchez, A.; Hirye, M.; Zortea, M.; Glorr, E.; ans Carlos Souza Filho, O.P.; Shimabukuro, Y.; Aragão, L. Individual tree crown delineation in a highly diverse tropical forest using very high resolution satellite images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 362–377. [CrossRef]
18. Weinstein, B.G.; Marconi, S.; Bohlman, S.; Zare, A.; White, E. Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. *Remote Sens.* **2019**, *11*, 1309. [CrossRef]
19. dos Santos, A.A.; Junior, J.M.; Araújo, M.S.; Martini, D.R.D.; Tetila, E.C.; Siqueira, H.L.; Aoki, C.; Eltner, A.; Matsubara, E.T.; Pistori, H.; et al. Assessment of CNN-Based Methods for Individual Tree Detection on Images Captured by RGB Cameras Attached to UAVs. *Sensors* **2019**, *19*, 3595. [CrossRef]
20. Torres, D.L.; Feitosa, R.Q.; Happ, P.N.; Rosa, L.E.C.L.; Junior, J.M.; Martins, J.; Bressan, P.O.; Gonçalves, W.N.; Liesenberg, V. Applying Fully Convolutional Architectures for Semantic Segmentation of a Single Tree Species in Urban Environment on High Resolution UAV Optical Imagery. *Sensors* **2020**, *20*, 563. [CrossRef]
21. Osco, L.P.; dos Santos de Arruda, M.; Gonçalves, D.N.; Dias, A.; Batistoti, J.; de Souza, M.; Gomes, F.D.G.; Ramos, A.P.M.; Jorge, L.A.C.; Liesenberg, V.; et al. A CNN approach to simultaneously count plants and detect plantation-rows from UAV imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *174*, 1–17. [CrossRef]
22. Biffi, L.J.; Mitishita, E.; Liesenberg, V.; dos Santos, A.A.; Gonçalves, D.N.; Estrabis, N.V.; de Andrade Silva, J.; Osco, L.P.; Ramos, A.P.M.; Centeno, J.A.S.; et al. ATSS Deep Learning-Based Approach to Detect Apple Fruits. *Remote Sens.* **2021**, *13*, 54. . [CrossRef]
23. Gomes, M.; Silva, J.; Gonçalves, D.; Zamboni, P.; Perez, J.; Batista, E.; Ramos, A.; Osco, L.; Matsubara, E.; Li, J.; et al. Mapping Utility Poles in Aerial Orthoimages Using ATSS Deep Learning Method. *Sensors* **2020**, *20*, 6070. [CrossRef]
24. Santos, A.; Junior, J.M.; de Andrade Silva, J.; Pereira, R.; Matos, D.; Menezes, G.; Higa, L.; Eltner, A.; Ramos, A.P.; Osco, L.; et al. Storm-Drain and Manhole Detection Using the RetinaNet Method. *Sensors* **2020**, *20*, 4450. [CrossRef]
25. Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 296–307. [CrossRef]

26. Courtrai, L.; Pham, M.T.; Lefèvre, S. Small Object Detection in Remote Sensing Images Based on Super-Resolution with Auxiliary Generative Adversarial Networks. *Remote Sens.* **2020**, *12*, 3152. [CrossRef]
27. Lu, X.; Li, Q.; Li, B.; Yan, J. MimicDet: Bridging the Gap Between One-Stage and Two-Stage Object Detection. 2020. Available online: http://xxx.lanl.gov/abs/2009.11528 (accessed on 8 April 2021).
28. Licheng, J.; Fan, Z.; Fang, L.; Shuyuan, Y.; Lingling, L.; Zhixi, F.; Rong, Q. A Survey of Deep Learning-Based Object Detection. *IEEE Access* **2019**, *7*, 128837–128868. [CrossRef]
29. Zhang, S.; Chi, C.; Yao, Y.; Lei, Z.; Li, S.Z. Bridging the Gap Between Anchor-based and Anchor-free Detection via Adaptive Training Sample Selection. *arXiv* **2019**, arXiv:1912.02424.
30. Chen, X.; Jiang, K.; Zhu, Y.; Wang, X.; Yun, T. Individual Tree Crown Segmentation Directly from UAV-Borne LiDAR Data Using the PointNet of Deep Learning. *Forests* **2021**, *12*, 131. [CrossRef]
31. Miyoshi, G.T.; dos Santos Arruda, M.; 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]
32. Ampatzidis, Y.; Partel, V.; Meyering, B.; Albrecht, U. Citrus rootstock evaluation utilizing UAV-based remote sensing and artificial intelligence. *Comput. Electron. Agric.* **2019**, *164*. [CrossRef]
33. Ampatzidis, Y.; Partel, V. UAV-Based High Throughput Phenotyping in Citrus Utilizing Multispectral Imaging and Artificial Intelligence. *Remote Sens.* **2019**, *11*, 410. [CrossRef]
34. Hartling, S.; Sagan, V.; Sidike, P.; Maimaitijiang, M.; Carron, J. Urban Tree Species Classification Using a WorldView-2/3 and LiDAR Data Fusion Approach and Deep Learning. *Sensors* **2019**, *19*, 1284. [CrossRef]
35. 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]
36. Li, W.; Fu, H.; Yu, L.; Cracknell, A. Deep Learning Based Oil Palm Tree Detection and Counting for High-Resolution Remote Sensing Images. *Remote Sens.* **2017**, *9*, 22. [CrossRef]
37. Nezami, S.; Khoramshahi, E.; Nevalainen, O.; Pölönen, I.; Honkavaara, E. Tree Species Classification of Drone Hyperspectral and RGB Imagery with Deep Learning Convolutional Neural Networks. *Remote Sens.* **2020**, *12*, 70. . [CrossRef]
38. Pleșoianu, A.I.; Stupariu, M.S.; Șandric, I.; Pătru-Stupariu, I.; Draguț, L. Individual Tree-Crown Detection and Species Classification in Very High-Resolution Remote Sensing Imagery Using a Deep Learning Ensemble Model. *Remote Sens.* **2020**, *12*, 2426. [CrossRef]
39. Culman, M.; Delalieux, S.; Tricht, K.V. Individual Palm Tree Detection Using Deep Learning on RGB Imagery to Support Tree Inventory. *Remote Sens.* **2020**, *12*, 3476. [CrossRef]
40. Oh, S.; Chang, A.; Ashapure, A.; Jung, J.; Dube, N.; Maeda, M.; Gonzalez, D.; Landivar, J. Plant Counting of Cotton from UAS Imagery Using Deep Learning-Based Object Detection Framework. *Remote Sens.* **2020**, *12*, 2981. [CrossRef]
41. Roslan, Z.; Long, Z.A.; Ismail, R. Individual Tree Crown Detection using GAN and RetinaNet on Tropical Forest. In Proceedings of the 2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM), Seoul, Korea, 4–6 January 2021; pp. 1–7. [CrossRef]
42. Roslan, Z.; Awang, Z.; Husen, M.N.; Ismail, R.; Hamzah, R. Deep Learning for Tree Crown Detection In Tropical Forest. In Proceedings of the 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM), Taichung, Taiwan, 3–5 January 2020, pp. 1–7. [CrossRef]
43. Afforestation of Public Roads: IBGE, 2010 Population Census. Available online: https://cidades.ibge.gov.br/brasil/ms/campo-grande/panorama (accessed on 30 March 2021).
44. Campo Grande Urban Arborization Master Plan: Campo Grande City Hall. 2010. Available online: http://www.campogrande.ms.gov.br/semadur/canais/arborizacao-urbana-plano-diretor/ (accessed on 30 March 2021).
45. Chen, K.; Wang, J.; Pang, J.; Cao, Y.; Xiong, Y.; Li, X.; Sun, S.; Feng, W.; Liu, Z.; Xu, J.; et al. MMDetection: Open MMLab Detection Toolbox and Benchmark. *arXiv* **2019**, arXiv:1906.07155.
46. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**. [CrossRef]
47. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
48. Micikevicius, P.; Narang, S.; Alben, J.; Diamos, G.; Elsen, E.; Garcia, D.; Ginsburg, B.; Houston, M.; Kuchaiev, O.; Venkatesh, G.; others. Mixed precision training. *arXiv* **2017**, arXiv:1710.03740.
49. Zhu, X.; Hu, H.; Lin, S.; Dai, J. Deformable ConvNets v2: More Deformable, Better Results. *arXiv* **2018**, arXiv:1811.11168.
50. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
51. Qiao, S.; Wang, H.; Liu, C.; Shen, W.; Yuille, A. Weight Standardization. *arXiv* **2019**, arXiv:1903.10520.
52. Wang, J.; Chen, K.; Xu, R.; Liu, Z.; Loy, C.C.; Lin, D. CARAFE: Content-Aware ReAssembly of FEatures. In Proceedings of the The IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
53. Zhu, C.; He, Y.; Savvides, M. Feature Selective Anchor-Free Module for Single-Shot Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 840–849.
54. Ghiasi, G.; Lin, T.Y.; Le, Q.V. Nas-fpn: Learning scalable feature pyramid architecture for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7036–7045.

55. Kong, T.; Sun, F.; Liu, H.; Jiang, Y.; Shi, J. FoveaBox: Beyond Anchor-based Object Detector. *arXiv* **2019**, arXiv:1904.03797.
56. Wu, Y.; Chen, Y.; Yuan, L.; Liu, Z.; Wang, L.; Li, H.; Fu, Y. Rethinking Classification and Localization for Object Detection. *arXiv* **2019**, arXiv:1904.06493.
57. Li, B.; Liu, Y.; Wang, X. Gradient Harmonized Single-stage Detector. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019.
58. Zhu, X.; Cheng, D.; Zhang, Z.; Lin, S.; Dai, J. An Empirical Study of Spatial Attention Mechanisms in Deep Networks. *arXiv* **2019**, arXiv:1904.05873.
59. Qiao, S.; Chen, L.C.; Yuille, A. DetectoRS: Detecting Objects with Recursive Feature Pyramid and Switchable Atrous Convolution. *arXiv* **2020**, arXiv:2006.02334.
60. Zhang, H.; Wang, Y.; Dayoub, F.; Sünderhauf, N. VarifocalNet: An IoU-aware Dense Object Detector. *arXiv* **2020**, arXiv:2008.13367.
61. Wang, J.; Zhang, W.; Cao, Y.; Chen, K.; Pang, J.; Gong, T.; Shi, J.; Loy, C.C.; Lin, D. *Side-Aware Boundary Localization for More Precise Object Detection*; ECCV 2020. Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2020; Volume 12349. [CrossRef]
62. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. *arXiv* **2020**, arXiv:2006.04388.
63. Kim, K.; Lee, H.S. *Probabilistic Anchor Assignment with IoU Prediction for Object Detection*; ECCV 2020. Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2020; Volume 12370. [CrossRef]
64. Zhang, H.; Chang, H.; Ma, B.; Wang, N.; Chen, X. Dynamic R-CNN: Towards High Quality Object Detection via Dynamic Training. *arXiv* **2020**, arXiv:2004.06002.
65. Wu, J.; Yang, G.; Yang, H.; Zhu, Y.; Li, Z.; Lei, L.; Zhao, C. Extracting apple tree crown information from remote imagery using deep learning. *Comput. Electron. Agric.* **2020**, *174*. [CrossRef]
66. Lumnitz, S.; Devisscher, T.; Mayaud, J.; Radic, V.; Coops, N.; Griess, V. Mapping trees along urban street networks with deep learning and street-level imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 144–157. [CrossRef]
67. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.