


Article

Detection of Forestry Pests Based on Improved YOLOv5 and Transfer Learning

Dayang Liu , Feng Lv, Jingtao Guo, Huiting Zhang and Liangkuan Zhu *

College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China; liudayang@nefu.edu.cn (D.L.)

* Correspondence: zhulk@nefu.edu.cn

Abstract: Infestations or parasitism by forestry pests can lead to adverse consequences for tree growth, development, and overall tree quality, ultimately resulting in ecological degradation. The identification and localization of forestry pests are of utmost importance for effective pest control within forest ecosystems. To tackle the challenges posed by variations in pest poses and similarities between different classes, this study introduced a novel end-to-end pest detection algorithm that leverages deep convolutional neural networks (CNNs) and a transfer learning technique. The basic architecture of the method is YOLOv5s, and the C2f module is adopted to replace part of the C3 module to obtain richer gradient information. In addition, the DyHead module is applied to improve the size, task, and spatial awareness of the model. To optimize network parameters and enhance pest detection ability, the model is initially trained using an agricultural pest dataset and subsequently fine-tuned with the forestry pest dataset. A comparative analysis was performed between the proposed method and other mainstream target detection approaches, including YOLOv4-Tiny, YOLOv6, YOLOv7, YOLOv8, and Faster RCNN. The experimental results demonstrated impressive performance in detecting 31 types of forestry pests, achieving a detection precision of 98.1%, recall of 97.5%, and $mAP@.5:.95$ of 88.1%. Significantly, our method outperforms all the compared target detection methods, showcasing a minimum improvement of 2.1% in $mAP@.5:.95$. The model has shown robustness and effectiveness in accurately detecting various pests.

Keywords: forestry pest; detection; transfer learning; deep learning



Citation: Liu, D.; Lv, F.; Guo, J.; Zhang, H.; Zhu, L. Detection of Forestry Pests Based on Improved YOLOv5 and Transfer Learning. *Forests* **2023**, *14*, 1484. <https://doi.org/10.3390/f14071484>

Academic Editor: Young-Seuk Park

Received: 31 May 2023

Revised: 14 July 2023

Accepted: 17 July 2023

Published: 20 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The maintenance of ecological balance is contingent upon the support of forestry, as economic development and forestry are closely intertwined. However, forestry development faces significant challenges due to pests and diseases, leading to substantial financial losses [1]. Therefore, the exploration of effective methods for identifying and detecting forestry pests holds great significance for the advancement of the forestry industry. Conventional pest identification approaches primarily rely on insect researchers to visually analyze insect characteristics, resulting in a slow and impractical process for large-scale implementation [2]. With advancements in machine vision and pattern recognition technology, there has been a growing interest in leveraging these technologies for pest identification and detection.

The majority of pest identification tasks rely on the utilization of machine learning frameworks. Ebrahimi et al. [3] conducted a study on employing the support vector machines (SVM) method with region index and reinforcement as the color index to detect pests. The machine vision technique was introduced, which aims to detect insects in pictures using multivariate analysis [4]. Li et al. [5] utilized multiple fractal analysis to segment small-sized insect pests based on local singularities and global image features. Wang et al. [6] combined artificial neural networks (ANN) with morphological features and SVM to develop an insect species classification system. However, it is important to note

that detection and classification tasks based on conventional machine learning methods are susceptible to variations in lighting conditions and environmental factors, resulting in limited algorithm portability.

With the continuous advancements in deep learning models and the remarkable progress of graphics processing units, CNNs have revolutionized image recognition and target detection. Consequently, the field of pest detection has also witnessed significant advancements in sophistication. For instance, Liu et al. [7] proposed an innovative approach that combines saliency maps and deep CNNs for the classification and localization of agricultural pests. Zhu et al. [8] presented a leaf black rot detection method based on super-resolution image enhancement and a convolutional neural network, which achieved a detection accuracy of 95.79%, effectively solving the small target detection task. Xia et al. [9] produced a dataset based on greenhouse pests, and a segmentation procedure was designed to identify some common greenhouse pests. Regarding target detection algorithms, two-stage methods like the R-CNN series utilize region proposals in the initial stage and subsequently classify each proposed region in the subsequent stage [10,11]. On the other hand, single-stage target detection algorithms such as the SSD [12] and YOLO [13–17] series offer real-time monitoring and rapid detection capabilities. However, it is worth noting that the current research on forestry pests often lacks a consideration of real-world environments and encounters limitations due to the scarcity of real environmental samples [18–20]. Specifically, there is a dense distribution, similarity between pests in categories, and an obscuration of pests in pest aggregation areas, leading to difficulties in distinguishing individual pests and the exact location of locating frames. In addition, there are variations in the size of the same pests in different pictures, leading to difficulties in feature identification. Finally, lighting conditions and complex backgrounds can also have an impact on detection. Therefore, pest detection still involves great challenges [20].

To solve the above problems, a forestry pest detection method based on improved YOLOv5 and transfer learning was proposed in this paper. The objectives of this work were to (1) introduce C2f and DyHead modules and build a high-accuracy pest monitoring model (YOLOv5_C2fD) to address the shortcomings of existing models for forestry pest detection; and (2) transfer the detection model trained by agricultural pests to forestry pest detection and improve the recognition performance of the model.

2. Materials and Methods

2.1. Introduction of Pest Datasets

Liu et al. [21] provided a comprehensive dataset containing 31 classes of forestry pests. This dataset offers pest images that fulfill the detection needs of natural environments. During the label transformation process, a partial error in label conversion occurred, resulting in a training set consisting of 5860 images with 13,530 target tags, and a test set comprising 651 images with 1512 target tags. Detailed information on each category is shown in Table 1.

The dataset was obtained through manual filtering and LabelImg annotation after conducting a comprehensive search on an image search engine. In order to address the discrepancy in the number of available images across different pest categories, various techniques such as brightness transformation, noise addition, and rotation were employed to ensure a balanced representation of images within each category. A visual representation of the diverse pest specimens encompassed in the dataset is provided in Figure 1. As shown in Figure 1, the pests exhibit both common characteristics across different classes and substantial variations within classes across different stages of their life cycles, posing a significant challenge in accurately classifying them.

Table 1. Details of forestry pest dataset.

Class Index	Name	Train Target Tags	Test Target Tags
0	Drosicha_contrahens_female	541	53
1	Drosicha_contrahens_male	183	17
2	Chalcophora_japonica	116	11
3	Anoplophora_chinensis	369	36
4	Psacotheta_hilaris (Pascoe)	198	16
5	Apriona_germari (Hope)	246	38
6	Monochamus_alternatus	147	18
7	Plagiodera_versicolora (Laicharting)	390	49
8	Latoia_consocia_Walker	245	25
9	Hyphantria_cunea	392	37
10	Cnidocampa_flavescens (Walker)	253	26
11	Cnidocampa_flavescens (Walker_pupa)	229	43
12	Erthesina_fullo	276	23
13	Erthesina_fullo_nymph-2	2143	385
14	Erthesina_fullo_nymph	185	25
15	Spilarctia_subcarnea (Walker)	171	16
16	Psilogramma_menephron	189	15
17	Seriginus_montela	342	45
18	Seriginus_montela_larvae	268	60
19	Clostera_anachoreta	251	26
20	Micromelalopha_troglodyta (Graeser)	161	21
21	Latoia_consocia_Walker_larvae	555	38
22	Plagiodera_versicolora (Laicharting)_larvae	826	47
23	Plagiodera_versicolora (Laicharting)_ovum	2716	227
24	Spilarctia_subcarnea (Walker)_larvae	161	19
25	Cerambycidae_larvae	392	20
26	Micromelalopha_troglodyta (Graeser)_larvae	142	14
27	Cerambycidae_larvae	370	27
29	Micromelalopha_troglodyta (Graeser)_larvae	342	48
29	Hyphantria_cunea_larvae	395	42
30	Hyphantria_cunea_pupa	336	45

**Figure 1.** Various pests in the dataset.

Despite efforts to collect substantial images in [21], the sample quantity is still insufficient. IP102 [22] is an agricultural pest dataset that covers 102 different species of pests and their multiple developmental stages. In terms of target detection tasks, the IP102 dataset comprises a total of 18,974 images with 22,253 annotated target tags. Further information about the IP102 dataset can be found in [21]. It is worth noting that the identification of forestry pests is challenging with complex environments and backgrounds, such as different lighting conditions and insects hidden underneath the background. In addition to this, there is a wide variety of images, including insects of varying age groups, colors, sizes, and shapes. Given these considerations, using the IP102 dataset in place of ImageNet or COCO as a pre-training dataset proved to be a valuable option [23].

During the pretest phase of this study, it was observed that specific categories exhibited low detection accuracy. Upon a closer examination of the labeling process for the original dataset, it was discovered that some instances were incorrectly labeled, making it almost impossible to solve the detection accuracy problem through model optimization. In light of this, samples from the IP102 dataset were selected as the agricultural dataset for pre-training by the results of the pre-experimental detection of each category. Finally, 7737 samples were selected for pre-training.

2.2. Composition of YOLOv5 Model

The YOLO family is a single-stage target detection algorithm that solves detection as a regression task with the output category and localization [24]. YOLOv5 is one of the effective deep learning algorithms for target detection based on YOLOv1–YOLOv4, with variations in network width and depth [25]. The network architecture of YOLOv5 consists of four main components, as illustrated in Figure 2. (1) Input: The input image is typical with a size of $640 \times 640 \times 3$ and undergoes preprocessing steps such as Mosaic data augmentation, adaptive image scaling, and adaptive anchor frame calculation. (2) Backbone: This component extracts feature maps of various sizes from the input image through multiple convolutional and pooling layers. The critical aspect of the backbone network is the feature extraction module, which includes the Conv module, C3 module, SPPF module, and others. (3) Neck: YOLOv5 adopts the PANet structure as its neck component. The shallow layers of the network contain detailed information but lack semantic understanding, while the deep layers provide rich semantic information but lack fine-grained details. The PANet structure effectively fuses feature maps from different levels to enhance detection performance. This is achieved through a feature pyramid structure that combines feature maps of varying resolutions, generating a new feature representation. (4) Head: The head component of YOLOv5 is responsible for predicting image features, constructing bounding boxes, and outputting feature maps for object detection. It completes the final stage of the detection process.

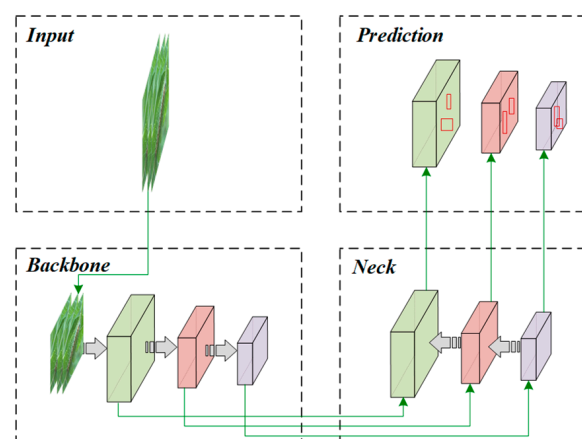


Figure 2. Composition of YOLOv5 model.

2.3. Improvements to the YOLOv5 Model

This section presents a comprehensive analysis of the prevailing target detection frameworks utilized for identification and localization tasks. The primary objective of this research is to propose a highly accurate and efficient detection method tailored explicitly for forestry pest identification and localization. Given the challenges posed by complex environmental backgrounds and limited training samples, we introduce the following vital modifications in this study: (1) adding the DyHead detection module to achieve size awareness, task awareness, and space awareness; (2) replacing the partial C3 modules with C2f modules to improve the detection capability for small targets. The network structure shown in Figure 3 illustrates the proposed improvements in the model's architecture. Our enhanced model, named YOLOv5_C2fD, integrates these modifications to achieve a superior performance.

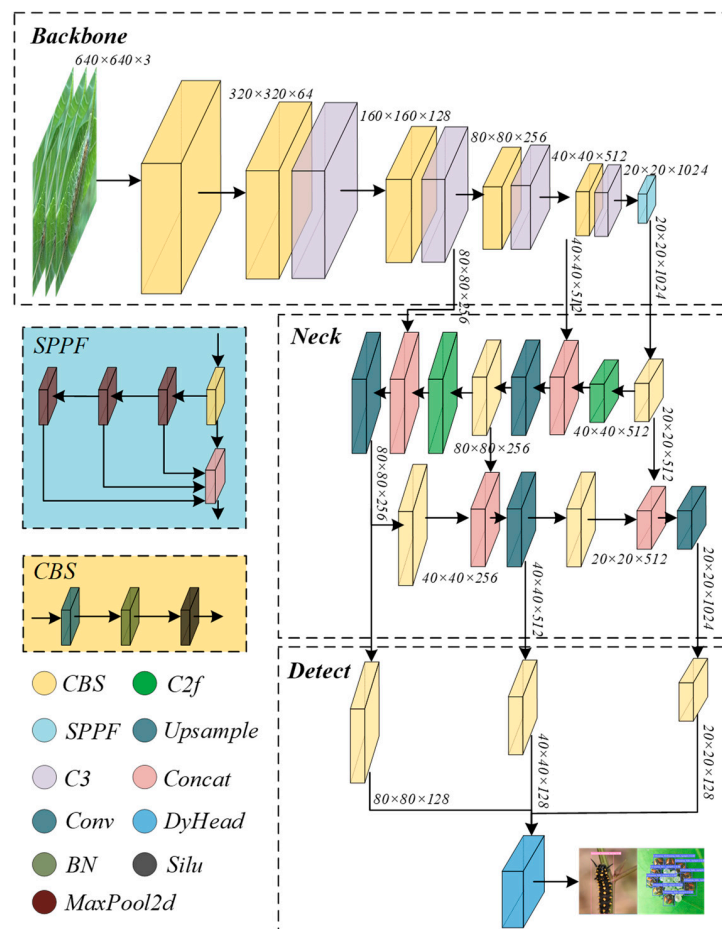


Figure 3. The network structure.

2.3.1. DyHead Module

The forestry pest dataset utilized in this study presents a complex background comprising numerous pest species at varying developmental stages and scales. Consequently, it is crucial for the detection algorithm to possess full-scale perception capabilities. Moreover, the targets in the feature element maps extracted from the model neck feature pyramid exhibit diverse spatial locations and shapes, necessitating the detection algorithm to capture spatial information effectively. To address these challenges, the DyHead detection module was introduced, which offers simultaneous size, task, and spatial awareness, making it well-suited for pest detection tasks [26]. By incorporating DyHead into the detection head component, the detection capabilities were enhanced while striving to optimize computational efficiency. The structure diagram of DyHead is shown in Figure 4.

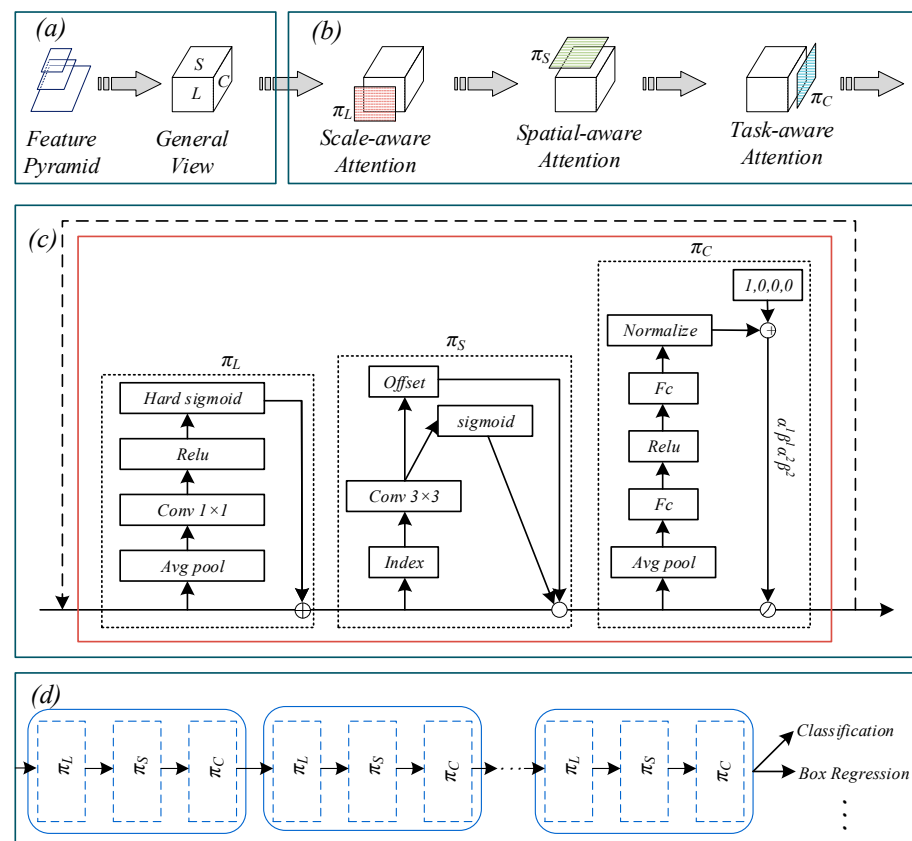


Figure 4. DyHead. (a) The input feature tensor of DyHead. (b) An illustration of the DyHead approach. (c) Details of a DyHead block. (d) Specific application of DyHead.

As shown in Figure 4a, the feature map is transformed into a three-dimensional tensor, where L represents the number of layers, C denotes the number of channels, and S represents the height and width of the feature map. Three attention mechanisms are sequentially employed to enhance the detection performance, as depicted in Figure 4b, with their respective structures detailed in Figure 4c. The specific application of DyHead is demonstrated in Figure 4d, showcasing its role within the detection framework.

DyHead achieves the integration of the target detection head by effectively combining multiple self-attentive mechanisms across feature layers for scale perception, spatial locations for spatial perception, and output channels for task perception. Size perception focuses on the scale size of the target pest, enabling the learning of the relative importance between feature layers, and enhancing features on the appropriate layer. Spatial perception aims to understand the spatial differences among different pests within the pest image features. Task perception processes the feature data of pests across channels, guiding different feature channels to identify various pests separately.

2.3.2. C2f Module

Although the YOLO series algorithm has evolved up to YOLOv8, its development is still in an immature stage. Therefore, we adopt YOLOv5 in this study while drawing inspiration from the advanced modules introduced in YOLOv8. Specifically, the C2f module is designed by referencing the C3 module and incorporating the concept of ELAN. This design allows the model to capture rich gradient flow information, thereby enhancing the detection capability for small targets. Figure 5 illustrates C3 and C2f modules.

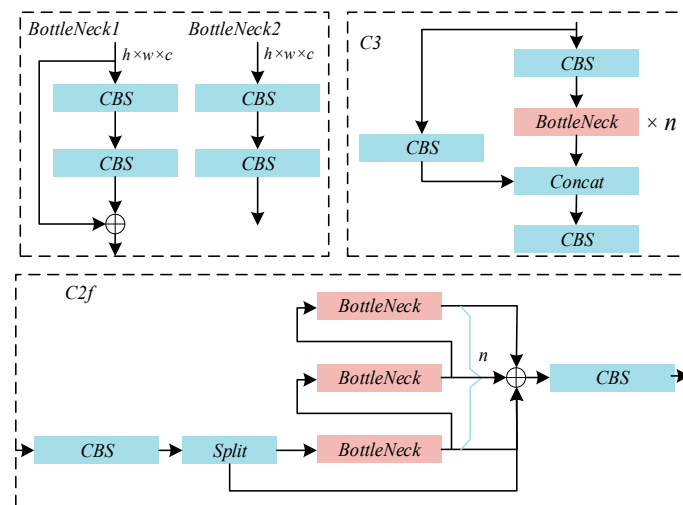


Figure 5. C2f module.

As shown in Figure 5, the C3 module is composed of three convolutional modules, which leverage the concept of early streaming from CSPNet and integrate the residual structure idea. The C2f module enhances the gradient flow by parallelizing multiple branches of gradient streams, thus incorporating richer gradient information for improved performance.

2.4. Transfer Learning

Transfer learning, which focuses on transferring knowledge, is a promising research method [27,28]. Sharing learned knowledge from a specific domain to other domains with similar characteristics can improve the generalization capability of the model [29]. Deep learning networks typically require a large number of training samples. However, in realistic scenarios, the availability of training data is often limited, making it challenging to enhance the model's robustness. In such cases, applying transfer learning to deep learning enables the efficient and rapid parameter optimization of prediction networks, leading to the development of highly robust models. In this study, we initially trained the model using an agricultural dataset and subsequently applied it to the forestry pest dataset. The network training process based on transfer learning is shown in Figure 6.

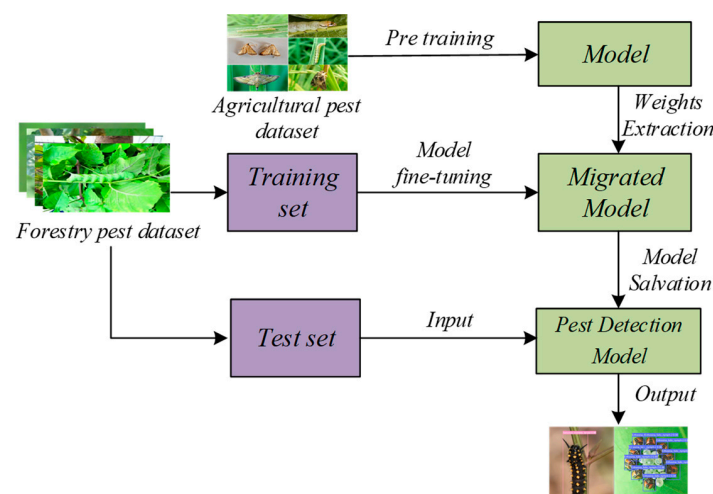


Figure 6. Network training process based on transfer learning.

The specific process involves training the model on the agricultural pest dataset, where the initial network weights are replaced with random initialization. Subsequently,

transfer learning is performed on the forestry pest dataset using the pre-trained model. The resulting model is then used for the target detection task on the test set.

2.5. Introduction of Correlation Models

2.5.1. Faster RCNN

The Faster R-CNN algorithm [30] is a well-known two-stage target detection algorithm. In the first stage, a convolutional layer is employed to extract features and determine the bounding box of the target object. In the second stage, a fully connected layer and bounding box regression are utilized to localize and classify the detected target precisely.

2.5.2. YOLOv4-Tiny

YOLOv4-Tiny [31] is a compact iteration of YOLOv4 [32], renowned for its accelerated training and detection performance. YOLOv4-Tiny features fewer detection heads and employs distinct activation functions and training convolution layers compared to YOLOv4. It incorporates multi-task functionality, end-to-end training, attention mechanisms, and multi-scale features, enabling it to handle diverse detection tasks effectively.

2.5.3. YOLOv6

YOLOv6 [33] is a single-stage object detector that combines efficient design principles with high-performance capabilities. It introduces a redesigned backbone network called EfficientRep and a new neck architecture named Rep-PAN. These changes in the backbone and neck design aim to further enhance the efficiency and effectiveness of the YOLOv6 model.

2.5.4. YOLOv7

YOLOv7 [34] is a highly advanced target detection algorithm that has gained significant attention. Its architecture consists of two main components: the backbone and head networks. The backbone network is responsible for feature extraction and is fed with preprocessed input images. On the other hand, the head network utilizes the extracted features to refine them further and fuse them for accurate detection. An integral feature of the YOLOv7 architecture is the implementation of the ELAN structure within the backbone network. This innovative structure efficiently controls the gradient paths, ensuring effective learning and convergence throughout the deep network. This results in improved efficiency and enhanced performance during training.

2.5.5. YOLOv8

The most recent and advanced target detection algorithm in the YOLO series is YOLOv8. It offers several variations, including YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x, each optimized for object detection tasks. The variations in the YOLO family models are determined by factors such as the depth and width multiplier.

2.6. Model Evaluation Metrics

In this study, the evaluation metrics chosen for assessing the performance of the target detection model are precision, recall, and average precision. The average precision (AP) is calculated as the average precision values across different recall levels. The mean average precision (mAP) is then derived by calculating the average of the AP values. mAP is considered a crucial metric for assessing the overall precision of the target detection model, making it a reliable indicator of its performance [35]. These metrics are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

$$AP = \int_0^1 P(R) dR \quad (3)$$

$$mAP = \frac{1}{M} \sum_{k=1}^M AP(k) \quad (4)$$

where TP is the true positive, FP is the false positive, TN is the true negative, and FN is the false negative. M is the number of pest species.

The model was quantitatively evaluated using the mAP at the threshold ranging from 0.5 to 0.95 with an interval of 0.05, denoted as $mAP@.5:.95$. This metric calculates the average precision at different thresholds and provides a comprehensive assessment of the model's performance in object detection.

3. Results and Discussion

3.1. Experimental Environment and Setting

The experiments presented in this paper were performed on an Ubuntu system, leveraging GPU acceleration to enhance computational speed. The specific parameters utilized for the experiments are detailed in Table 2.

Table 2. Experimental environment parameters and setting.

	Name	Parameter
Hardware	CPU	Intel (R) Xeon (R) Silver 4215 CPU @ 2.50 GHz
	Memory	252 G
	GPU	GeForce RTX 3090 × 2
Software	Graphics card	24 G × 2
	Operation system	Ubuntu 20.04
	Deep Learning framework	Pytorch 1.13.0
	Programming languages	Python 3.8
Algorithms	CUDA	11.6
	Iterations	300
	Batch size	128
	Picture size	640 × 640
	Learning rate	0.01
	Momentum	0.937
	Weight decay	0.0005

3.2. Ablation Experiment

The constructed model for forestry pest detection comprised the YOLOv5s baseline module, DyHead module, and C2f module. Ablation experiments were performed to analyze the contribution of each component, and the results are reported in Table 3. In this study, the DyHead module was first added alone to the baseline model, resulting in a 0.9% increase in $mAP@.5:.95$. This is likely because the DyHead detection module provides synchronized size, task, and spatial awareness, making the model more suitable for the pest detection task. Subsequently, adding the C2f module alone showed a better effect than DyHead, leading to a 2% increase in $mAP@.5:.95$. This may be related to the fact that the C2f module incorporates the concept of ELAN for the design so that the model captures rich gradient flow information and thus enhances the detection of small targets. Lastly, the results of multi-scale training were analyzed. As shown in Table 3, it can be observed that multi-scale training achieved the highest precision, recall, and $mAP@.5:.95$, indicating that it can significantly enhance the accuracy of forestry pest detection.

Table 3. Results of the ablation experiment.

YOLOv5	DyHead	C2f	Precision (%)	Recall (%)	mAP@.5:.95 (%)
✓			97.6	95.6	82.7
✓	✓		96.8	97.0	83.6
✓		✓	97.7	96.2	84.7
✓	✓	✓	97.8	97.2	85.4

3.3. Analysis of Transfer Learning Results

3.3.1. Analysis of Whole Detection Situation

The influence of transfer learning on the results was compared in terms of precision, recall, and $mAP@.5:.95$, and the comparison graph is presented in Figure 7. The results show that precision, recall, and $mAP@.5:.95$ have all increased to varying degrees, with $mAP@.5:.95$ experiencing the highest increase, reaching 2.7%. This indicates that transfer learning has a positive impact on improving the performance of the model.

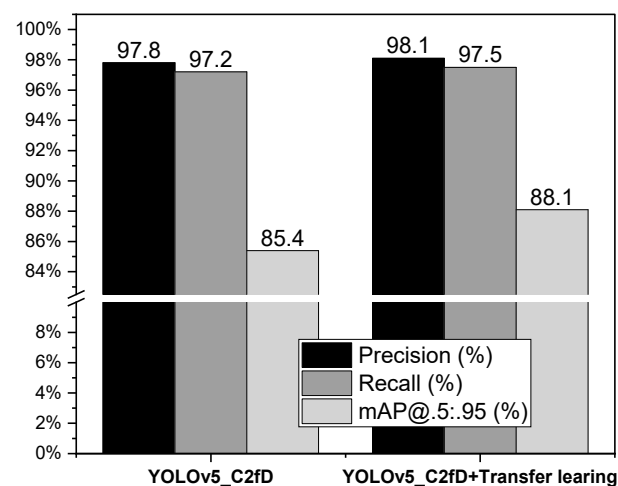
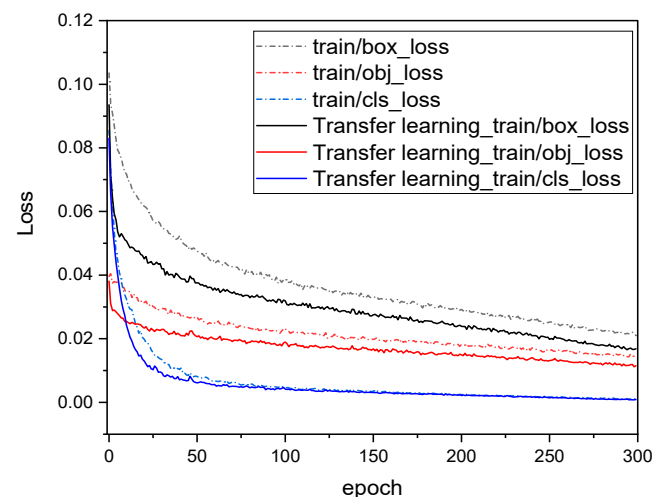
**Figure 7.** Detection before and after transfer learning.

Figure 8 illustrates the declining trend of the three losses during training before and after applying transfer learning. The comparison reveals that transfer learning has a significant impact on accelerating the training process, particularly during the initial stages. This leads to faster model convergence and expedites the overall training process.

**Figure 8.** Comparison of loss.

3.3.2. Analysis of Single Type Pest

The dataset used in this experiment consists of 31 categories, representing various types of pests with distinct body states and specific shared attributes. Figure 9 showcases the detection results of each pest category by YOLOv5_C2fD, comparing the performance with and without transfer learning.

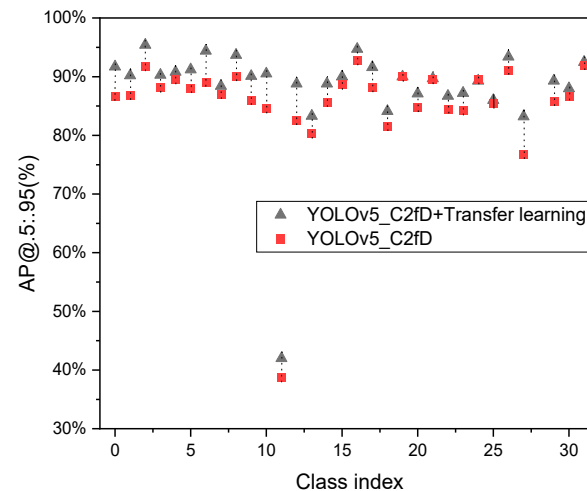


Figure 9. Comparison of AP@.5:.95 for each category of pests.

Figure 9 illustrates the impact of transfer learning on the detection accuracy of different pest categories. The results indicate that after applying transfer learning, 29 pest categories exhibited an improved detection accuracy, one category showed a decreased detection accuracy, and one category remained unchanged. These findings demonstrate the potential of transfer learning in enhancing the detection of forestry pests, particularly in complex backgrounds.

3.4. Confusion Matrix

Confusion matrix is the most intuitive and simple method to measure the accuracy of classification models. For the test set of forestry pests, the confusion matrix of the YOLOv5_C2fD model based on transfer learning is shown in Figure 10.

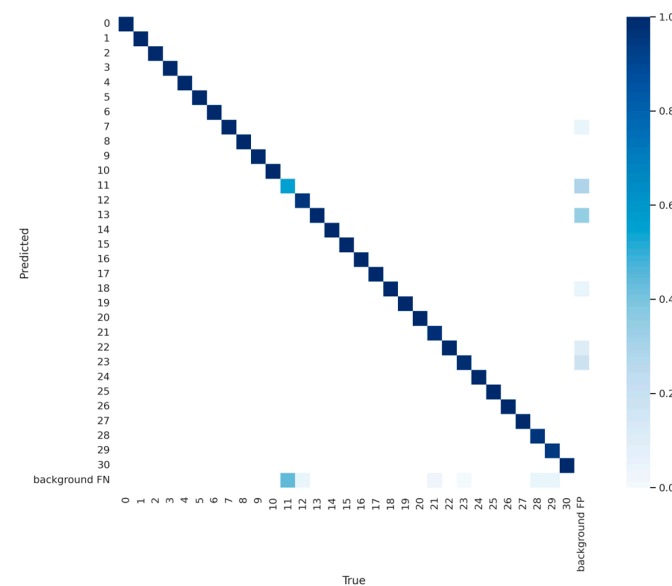


Figure 10. The confusion matrix of the YOLOv5_C2fD model based on transfer learning.

In Figure 10, the diagonal of the confusion matrix represents the proportion of correct detections in each category. The higher the accuracy of the detection, the deeper the color of the diagonal. The color of the diagonal of one class of pests is obviously lighter than that of other pests, which may be caused by the inaccurate labeling of the position of the label box. The accuracy of each type of pest detection is closely related to the size and shape of the pest itself, as well as the sample amount of that type of pest.

3.5. Comparison with Existing Target Detection Methods

3.5.1. Detection Results of Different Models

To assess the performance of our method, we conducted a comparative analysis with several existing target detection methods using the forestry pest dataset. Specifically, we compared our method with Faster RCNN, YOLOv4-Tiny, YOLOv6, YOLOv7, and YOLOv8, which are representative mainstream target detection methods. The evaluation results can be found in Table 4.

Table 4. Performance comparison of the models.

Models	Precision (%)	Recall (%)	mAP@.5:.95 (%)
Faster RCNN	92.3	90.4	58.7
YOLOv4-Tiny	96.0	93.0	64.3
YOLOv6	98.0	97.4	84.0
YOLOv7	97.7	98.1	83.8
YOLOv8	98.0	96.0	86.0
Ours	98.1	97.5	88.1

The results obtained from the comparison experiments presented in Table 4 reveal that our proposed method surpasses the selected target detection method YOLOv8 by achieving a 2.1% higher detection accuracy. Additionally, our method outperforms all the other selected target detection methods in terms of detection accuracy, with a remarkable maximum improvement of 29.4% in *mAP@.5:.95*. These compelling experimental results demonstrate that our method excels in pest detection performance, delivering a high detection accuracy.

3.5.2. Pest Detection Visualization Comparison

The comprehensive analysis of the aspects mentioned above shows that the YOLOv5_C2fD combined with the transfer learning model exhibits a superior detection performance. To further validate its effectiveness in real-world scenarios, a set of test data consisting of small targets, dense scenes, and occlusions was selected to compare the detection results of each model. The detection outcomes are illustrated in Figure 11.

Figure 11 shows that the proposed model in this paper demonstrates a good detection performance in dense scenes and complex backgrounds. In comparison, YOLOv4-Tiny and YOLOv7 exhibit incorrect detections or fail to detect objects in complex backgrounds. Additionally, YOLOv6, YOLOv8 and Faster RCNN show cases of missed detections in dense scenes. These results clearly illustrate the superior detection capability of the proposed model in complex and lush environments. The improved performance outperforms existing models, making it a valuable tool for forestry pest detection and management.



Figure 11. Detection outcomes of six methods.

3.6. The Realization of the Model Deployed on a Laptop

To verify that the model can achieve real-time detection after deployment, this paper deployed the model on a Lenovo Xiaoxin Pro14 laptop (Lenovo (Beijing) Co., Ltd., Beijing, China) with an external cell phone as a camera for easy mobility. Figure 12 presents the results of real-time detection after deployment. As can be seen in Figure 12, the pests were identified and localized.

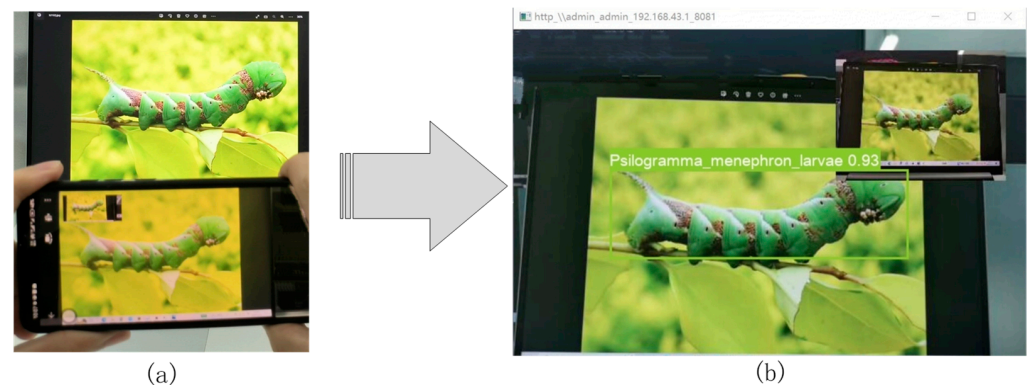


Figure 12. Real-time detection results after deployment. (a) Pest monitoring. (b) Detection results.

In order to better promote the development of forestry pest identification, our next step was to advance the issue of model deployment to enable real-time monitoring such as for drones or robots.

4. Conclusions

Forestry is facing significant challenges due to the threats posed by pests and diseases, resulting in substantial economic losses. The accurate identification of forestry pests is crucial for timely preventive measures and for minimizing financial impacts. To address the limitations of conventional methods, this paper proposes the YOLOv5_C2fD model based on transfer learning, which enhances pest feature extraction, extracts meaningful features, reduces redundancies, and improves pest identification and localization in complex backgrounds. The C2f module, which parallelizes gradient flow branches, captures richer gradient flow information, and the DyHead module exhibits a simultaneous awareness of size, task, and spatial factors, making it suitable for pest detection. The model is initially trained on an agricultural dataset and then transferred to a forestry pest dataset, facilitating rapid parameter optimization, and improving the detection of pest targets. The experimental results demonstrated an impressive performance in detecting 31 types of forestry pests, achieving a detection precision of 98.1%, recall of 97.5%, and $mAP@.5:.95$ of 88.1%. Significantly, our method outperforms all the compared target detection methods, showcasing a minimum improvement of 2.1% in $mAP@.5:.95$. The improved performance outperforms existing models, making it a valuable tool for forestry pest detection and management. Quantitative experiments were conducted, demonstrating the superiority of the proposed method over other detectors in terms of pest localization and classification. Although the current model achieves good results in regard to forest pest identification, small target identification is still a challenge. We will subsequently optimize and improve the model to further enhance its ability to detect small targets. In addition, reducing the impact of missing labels on the model of the pest dataset will be the next direction of our research in our future work.

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

Funding: This research was funded by the National Natural Science Foundation of China (No. 32202147), the China Postdoctoral Science Foundation (No. 2021M690573), and the Fundamental Research Funds for the Central Universities (No. 2572020BF05).

Data Availability Statement: The data presented in this study are available upon request from the corresponding authors.

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

References

- Hurley, B.P.; Slippers, J.; Wingfield, M.J.; Dyer, C.; Slippers, B. Perception and knowledge of the Sirex woodwasp and other forest pests in South Africa. *Agric. For. Entomol.* **2012**, *14*, 306–316. [\[CrossRef\]](#)
- Hiary, H.; Ahmad, S.B.; Reyalat, M.; Braik, M.; Al-Rahamneh, Z. Fast and Accurate Detection and Classification of Plant Diseases. *Int. J. Comput. Appl.* **2011**, *17*, 31–38.
- Ebrahimi, M.A.; Khoshtaghaza, M.H.; Minaei, S.; Jamshidi, B. Vision-based pest detection based on SVM classification method. *Comput. Electron. Agric.* **2017**, *137*, 52–58. [\[CrossRef\]](#)
- Zayas, I.Y.; Flinn, P.W. Detection of insects in bulk wheat samples with machine vision. *Trans. Am. Soc. Agric. Eng.* **1998**, *41*, 883–888. [\[CrossRef\]](#)
- Li, Y.; Xia, C.; Lee, J. Detection of small-sized insect pest in greenhouses based on multifractal analysis. *Opt.-Int. J. Light Electron Opt.* **2015**, *126*, 2138–2143. [\[CrossRef\]](#)
- Wang, J.; Lin, C.; Ji, L.; Liang, A. A new automatic identification system of insect images at the order level. *Knowl.-Based Syst.* **2012**, *33*, 102–110. [\[CrossRef\]](#)
- Liu, Z.; Gao, J.; Yang, G.; Zhang, H.; He, Y. Localization and Classification of Paddy Field Pests using a Saliency Map and Deep Convolutional Neural Network. *Sci. Rep.* **2016**, *6*, 20410. [\[CrossRef\]](#)
- Zhu, J.J.; Cheng, M.; Wang, Q.F.; Yuan, H.B.; Cai, Z.J. Grape Leaf Black Rot Detection Based on Super-Resolution Image Enhancement and Deep Learning. *Front. Plant Sci.* **2021**, *12*, 695749. [\[CrossRef\]](#)
- Xia, C.; Chon, T.-S.; Ren, Z.; Lee, J.-M. Automatic identification and counting of small size pests in greenhouse conditions with low computational cost. *Ecol. Inform.* **2015**, *29*, 139–146. [\[CrossRef\]](#)
- Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
- Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving into High Quality Object Detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
- Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
- Chen, Q.; Wang, Y.; Yang, T.; Zhang, X.; Cheng, J.; Sun, J. You Only Look One-level Feature. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13034–13043.
- Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
- Roy, A.M.; Bhaduri, J. Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4. *Comput. Electron. Agric.* **2022**, *193*, 106694. [\[CrossRef\]](#)
- Lawal, M.O. Tomato detection based on modified YOLOv3 framework. *Sci. Rep.* **2021**, *11*, 1447. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, D.; Wang, R.; Xie, C.; Liu, L.; Zhang, J.; Li, R.; Wang, F.; Zhou, M.; Liu, W. A Recognition Method for Rice Plant Diseases and Pests Video Detection Based on Deep Convolutional Neural Network. *Sensors* **2020**, *20*, 578. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sun, Y.; Liu, X.; Yuan, M.; Ren, L.; Wang, J.; Chen, Z. Automatic in-trap pest detection using deep learning for pheromone-based *Dendroctonus valens* monitoring. *Biosyst. Eng.* **2018**, *176*, 140–150. [\[CrossRef\]](#)
- Hong, S.-J.; Kim, S.-Y.; Kim, E.; Lee, C.-H.; Lee, J.-S.; Lee, D.-S.; Bang, J.; Kim, G. Moth Detection from Pheromone Trap Images Using Deep Learning Object Detectors. *Agriculture* **2020**, *10*, 170. [\[CrossRef\]](#)
- Jiao, L.; Li, G.Q.; Chen, P.; Wang, R.J.; Du, J.M.; Liu, H.Y.; Dong, S.F. Global Context-Aware-Based Deformable Residual Network Module for Precise Pest Recognition and Detection. *Front. Plant Sci.* **2022**, *13*, 895944. [\[CrossRef\]](#)
- Liu, B.; Liu, L.; Zhuo, R.; Chen, W.; Duan, R.; Wang, G. A Dataset for Forestry Pest Identification. *Front. Plant Sci.* **2022**, *13*, 857104. [\[CrossRef\]](#)
- Wu, X.; Zhan, C.; Lai, Y.K.; Cheng, M.M.; Yang, J. IP102: A Large-Scale Benchmark Dataset for Insect Pest Recognition. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 8779–8788.
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2012**, *60*, 84–90. [\[CrossRef\]](#)
- Fu, L.; Yang, Z.; Wu, F.; Zou, X.; Lin, J.; Cao, Y.; Duan, J. YOLO-Banana: A Lightweight Neural Network for Rapid Detection of Banana Bunches and Stalks in the Natural Environment. *Agronomy* **2022**, *12*, 391. [\[CrossRef\]](#)
- Xiang, Q.; Huang, X.; Huang, Z.; Chen, X.; Cheng, J.; Tang, X. Yolo-Pest: An Insect Pest Object Detection Algorithm via CAC3 Module. *Sensors* **2023**, *23*, 3221. [\[CrossRef\]](#)
- Dai, X.Y.; Chen, Y.P.; Xiao, B.; Chen, D.D.; Liu, M.C.; Yuan, L.; Zhang, L.; IEEE Comp, S.O.C. Dynamic Head: Unifying Object Detection Heads with Attentions. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021, Nashville, TN, USA, 20–25 June 2021; pp. 7369–7378.
- Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2021**, *109*, 43–76. [\[CrossRef\]](#)
- Fraiwani, M.; Faouri, E.; Khasawneh, N. Classification of Corn Diseases from Leaf Images Using Deep Transfer Learning. *Plants* **2022**, *11*, 2668. [\[CrossRef\]](#) [\[PubMed\]](#)

29. Han, L.; Zhao, Y.Y.; Chen, H.A.; Chandrasekar, V. Advancing Radar Nowcasting through Deep Transfer Learning. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–9. [[CrossRef](#)]
30. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2015**, arXiv:1506.01497. [[CrossRef](#)] [[PubMed](#)]
31. Ji, H.; Hu, C.; Zhang, S.; Zhang, L.; Yang, X. BiO(OH)_xI_{1-x} solid solution with rich oxygen vacancies: Interlayer guest hydroxyl for improved photocatalytic properties. *J. Colloid Interface Sci.* **2022**, *605*, 1–12. [[CrossRef](#)]
32. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [[CrossRef](#)]
33. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv* **2022**, arXiv:2209.02976. [[CrossRef](#)]
34. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696. [[CrossRef](#)]
35. Khasawneh, N.; Fraiwan, M.; Fraiwan, L. Detection of K-complexes in EEG signals using deep transfer learning and YOLOv3. *Clust. Comput.* **2022**. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.