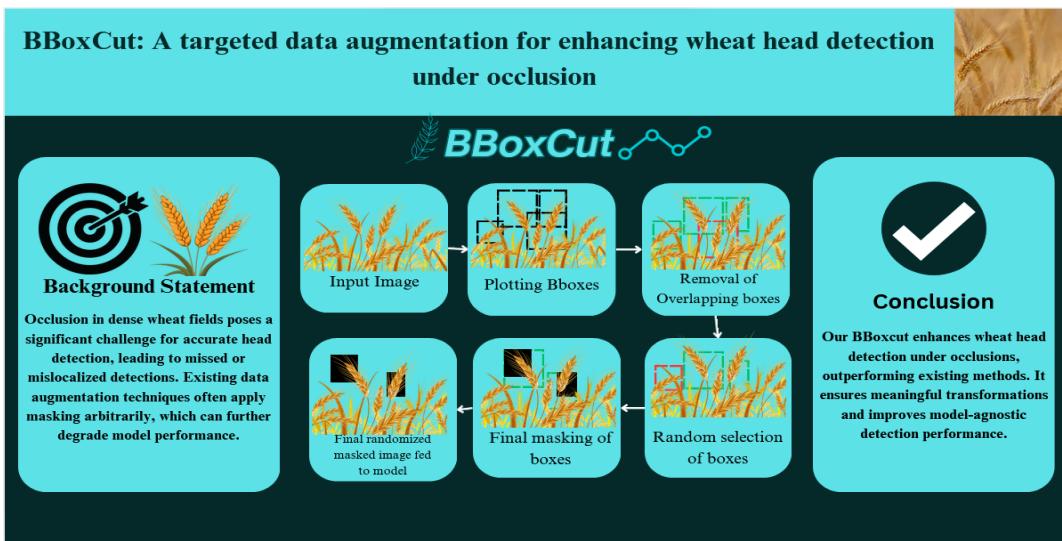


Graphical Abstract

BBoxCut: A Targeted Data Augmentation Technique for Enhancing Wheat Head Detection Under Occlusions

Yasashwini Sai Gowri P, Karthik Seemakurthy, Andrews Agyemang Opoku, Sita Devi Bharatula



Highlights

BBoxCut: A Targeted Data Augmentation Technique for Enhancing Wheat Head Detection Under Occlusions

Yasashwini Sai Gowri P, Karthik Seemakurthy, Andrews Agyemang Opoku, Sita Devi Bharatula

- Proposed a random localised masking based data augmentation technique for improved occluded wheat detection.
- Improved detection performance for two scenarios: a) wheat heads that overlap each other and b) wheat heads that are occluded by leaves.
- The effectiveness of the proposed approach was evaluated across three different object detection architectures.

BBoxCut: A Targeted Data Augmentation Technique for Enhancing Wheat Head Detection Under Occlusions

Yasashwini Sai Gowri P^a, Karthik Seemakurthy^{b,1}, Andrews Agyemang Opoku^c, Sita Devi Bharatula^a

^a*Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Department of Electronics and Communications Engineering, Chennai, India*

^b*Hydronium Energies Limited, London, London, UK*

^c*School of Computer Science, University of Lincoln, Lincoln, UK*

^d*Lincoln Institute of Agri-Food Technology, University of Lincoln, Lincoln, UK*

Abstract

Wheat plays a critical role in global food security, making it one of the most extensively studied crops. Accurate identification and measurement of key characteristics of wheat heads are essential for breeders to select varieties for cross-breeding, with the goal of developing nutrient-dense, resilient, and sustainable cultivars. Traditionally, these measurements are performed manually, which is both time-consuming and inefficient. Advances in digital technologies have paved the way for automating this process. However, field conditions pose significant challenges, such as occlusions of leaves, overlapping wheat heads, varying lighting conditions, and motion blur. In this paper, we propose a novel data augmentation technique, BBoxCut, which uses random localized masking to simulate occlusions caused by leaves and neighboring wheat heads. We evaluated our approach using three state-of-the-art object detectors and observed mean average precision (mAP) gains of 2.76, 3.26, and 1.9 for Faster R-CNN, FCOS, and DETR, respectively. Our augmentation technique led to significant improvements both qualitatively and quantitatively. In particular, the improvements were particularly evident in scenarios involving occluded wheat heads, demonstrating the robustness of our method in challenging field conditions.

Keywords: Wheat Detection, Occlusion, Data Augmentation, Masking

1. Introduction

Wheat is a staple crop for more than three billion people around the world and serves as a critical source of essential nutrients. To meet the increasing demands of a growing global population, agricultural production must double by 2050 [1]. Given its status as the most widely grown crop, accurate monitoring of wheat yield is crucial to ensure global food security [2]. Estimating wheat yield and making informed decisions about which wheat varieties to cross-breed for more nutritious, resilient, and high-yield cultivars are dependent on accurate detection of wheat head. Traditionally, yield prediction has relied on expert forecasts and manual counting of wheat heads, which are both time consuming and prone to inaccuracies [3].

The Global Wheat Head Detection (GWHD) 2021 dataset [4] was introduced to facilitate the development of automated systems capable of measuring the visual traits of wheat heads. Despite significant advances in object detection models [5], detecting wheat heads in unconstrained real-world field conditions presents considerable challenges, as illustrated in Fig. 1. Fig. 1 (a) shows occlusions caused by leaves, while Fig. 1 (b) highlights wheat heads that overlap one another. In both cases, state-of-the-art object detectors often struggle to accurately locate wheat heads. This performance degradation is likely due to the lack of diverse training examples that effectively represent occlusion scenarios. Furthermore, Fig. 2 reveals that the test domains exhibit higher wheat head densities compared to the training set, further increasing the likelihood of occlusions caused by leaves and neighboring wheat heads, as shown in Fig. 1. Such occlusions are expected to significantly affect the average performance of detectors across unseen test domains.



Figure 1: Samples from GWHD 2021 dataset. (a) Wheat heads partially occluded by leaves, (b) Wheat heads occluding each other. Occlusions are indicated by the red circles.

We propose BBoxCut, a data augmentation technique that applies random localized masking to simulate realistic occlusions in wheat fields. Unlike

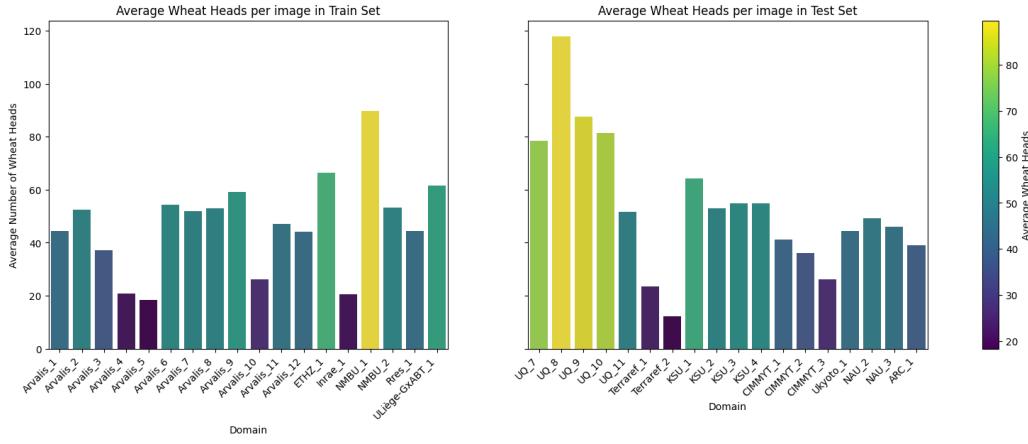


Figure 2: Average wheat head count per image in GWHD 2021 dataset.

Cutout [6], which randomly masks square regions throughout the image, BBoxCut specifically masks portions of the bounding boxes, mimicking partial occlusions by leaves or neighboring wheat heads. By incorporating BBoxCut into the data augmentation pipeline for the training of object detection models (Faster R-CNN, FCOS, and DETR) we observed significant improvements in mean Average Precision (mAP), particularly under occlusion scenarios. This enhancement is crucial to accurately predict wheat yields, directly impacting farmer profitability. The following is a list of our contributions proposed in this paper.

- We proposed a new data augmentation technique that improves wheat head localization in the presence of occlusions.
- We propose using a histogram-based dominant color estimation for mask generation.
- We have shown the validity of our approach on three different types of detectors such as Faster R-CNN, FCOS, and DETR.
- Our comparison with state-of-the-art masking-based data augmentation approaches demonstrates improved detection capability using the proposed technique.

2. Related work

In this section, we provide an overview of related works, taking a broad look at deep learning-based object detection, recent efforts to improve wheat detection accuracy, and studies that explicitly model occlusion in the context of object detection. We also highlight how our approach differs from existing techniques.

2.1. Deep Learning for object detection

Deep neural networks have been adopted for various computer vision tasks, such as object detection, due to their exceptional performance in feature extraction [7]. Most object detection networks can be classified into two categories: single-stage [8, 9, 10, 11, 12, 13] and two-stage [14, 15]. Although most existing approaches use convolutional neural network (CNN) architectures, transformer-based architectures [16, 17] have been gaining importance. Despite significant improvements in speed and accuracy, deep learning-based object detection networks generally do not explicitly model occlusion [18], which is common in many real-world scenarios [4, 19].

2.2. Wheat Head Detection

The importance of wheat for global food security has driven researchers to apply various detection techniques to detect wheat heads. Earlier work used image processing and machine learning techniques for wheat head detection. [20] used color-based features to count wheat heads. However, wheat ears, leaves, and stalks in the wheat field share similar color and texture characteristics. In addition, as the wheat plant matures, the color of different parts of the plant changes. As a result, using color cues to reliably identify wheat ears proved challenging. [21] successfully detected and counted wheat ears using the support vector machine approach. [22] utilized the k-means approach to partition the wheat head and obtain identification. To segment and count wheat heads, the twin support vector machine segmentation approach was used in [23]. [24] segmented wheat heads from backgrounds using Fourier filtering and Fourier transform.

Due to the limitations of traditional object detection methods, researchers in recent years have adopted deep learning-based wheat head localization techniques. [25] used R-CNN to detect and analyze wheat heads. Similarly, [26] applied ablation in color, affine variation, and random cutout on the Global Wheat Head Detection (GWHD) data set and trained the model on

a pre-trained YOLOv5 model. Although the performance of the model was not overly impressive, the random cutout showed an increase in performance compared to the ablation of color and the affine variation. [27] successfully trained a ResNet-50 model to detect and classify wheat head diseases, but no mechanisms for handling occlusion were mentioned. [28] proposed a domain-independent feature representation approach that performed well when tested with wheat head data from several locations, but the problem of missed detection due to occlusion persisted.

2.3. Occlusion Handling in Object Detection

The presence of occlusions that hinder object visibility is one of the key factors contributing to the decrease in object detection network performance. Recently, numerous works have summarized and highlighted the need to modify occlusions to advance the state-of-the-art object detection performance [29, 30, 31, 32, 33]. The latest work [33] summarized the occlusion handling approaches into three different categories: Generative techniques, deep learning strategies, and alternative approaches like graphical models and data augmentation techniques. The key advantage of data augmentation approaches over other occlusion modelling techniques is that they do not introduce additional trainable parameters, making them well-suited for resource-constrained scenarios. The proposed approach falls under the data augmentation category. In the next subsection, we describe related works that also belong to this category.

2.3.1. Data Augmentation

The quantity and diversity of training data are critical factors in determining a model’s ability to generalize effectively. Data augmentation is a key strategy to generate additional useful data from existing datasets, enabling models to achieve optimal performance. For instance, [34] conducted experiments with and without data augmentation to evaluate object detection algorithms, demonstrating that models trained with augmented data consistently outperformed those trained without it. Existing data augmentation approaches can be broadly categorized into three types: spatial transformation [7], color distortion [35], and information dropping [6, 36]. Spatial transformation involves fundamental augmentation techniques such as random scaling, cropping, flipping, and rotation, which are widely used during training. Color distortion adjusts properties such as brightness, color balance, and saturation to mimic real-world variations [35]. Both spatial transformations

and color distortions aim to enhance model robustness by modifying specific aspects of the training data to better reflect real-world scenarios. Information dropping techniques introduce diversity by varying visual features across training samples and are particularly effective for simulating occlusion conditions in computer vision tasks, such as visual tracking, where occlusions often result from object interactions or dynamic environments [37, 38]. These techniques can also help models ignore irrelevant features and focus on constraint cues, as demonstrated by [39], Augmentation by Information Dropping (AID) improved performance in crowded scenes by reducing the importance of appearance features.

Cutout [38] and Random Erasing [37] are two widely-used methods. Cutout masks random regions with zero pixel values using a square matrix, improving model robustness, especially when combined with other regularization techniques like dropout. Random Erasing applies random-sized masks with pixel values between 0 and 255 at random locations, conditionally applied with specified probabilities and aspect ratios to simulate partial occlusion. Region-aware Random Erasing [40] extends this idea to object detection by avoiding the deletion of important regions, such as bounding box areas.

More recent approaches, such as Hide-and-Seek [41], partition images into grids and delete content from random cells, leaving the rest with zero pixel values. A common limitation of these methods is their random nature, which may delete informative regions and degrade performance. To address this, KeepAugment [42] uses saliency maps to identify and preserve informative areas while augmenting less important regions. Similarly, GridMask [43] organizes image regions into grids and selectively masks portions, improving performance compared to AutoAugment. GridCut [44] further adapts grid masks to image-specific characteristics for more nuanced augmentation, while FenceMask [45] applies finer, wire-like grid patterns to reduce information loss and improve generalization, especially in fine-grained recognition tasks.

In this work, we propose a new data augmentation technique, BBoxCut, which is primarily motivated by masking-based techniques. Majority of the existing masking-based approaches in the literature randomly identify regions to be masked. However, this random mask placement can significantly impact the performance of detection networks and may or may not effectively simulate the real-world masking that occurs in wheat head scenarios. In our approach, we first identify wheat heads that do not have significant overlaps with adjacent wheat heads. This is followed by the random selection of wheat heads for mask placement, with the mask color determined through histogram

analysis. We demonstrate that our approach outperforms existing mask based data augmentation strategies designed to model occlusion scenarios for object detection application.

3. Proposed Approach

To address the problem of improving wheat head detection under occlusions, we propose a data augmentation strategy that includes a series of steps starting from removal of overlapping bounding boxes, histogram-based dominant color estimation followed by random masking for realistic occlusion simulation.

3.1. Identification of non-overlapping boxes

The first step of our approach is to identify the bounding boxes where masks should be applied. Given the nature of occlusions, as seen in Fig. 1, selecting appropriate candidate bounding boxes for masking is challenging. This is crucial because applying masks to randomly selected bounding boxes may result in masking regions that are already occluded, which could severely impact detection performance. Therefore, it is essential to identify bounding boxes where the likelihood of the occlusions being present is very minimal. To achieve this, we use the intersection-over-union (IoU) metric to detect overlapping bounding boxes and exclude them from our analysis, minimizing the adverse impact on detection performance.

Let B be the set of all the bounding boxes in an image. For each pair of bounding boxes $b_i, b_j \in B$, we compute the intersection over Union (IoU) as:

$$\text{IoU}(b_i, b_j) = \frac{\text{area}(b_i \cap b_j)}{\text{area}(b_i \cup b_j)}. \quad (1)$$

Bounding boxes with significant overlap, where $\text{IoU}(b_i, b_j) > \tau_{\text{IoU}}$, are excluded for the masking step. The resulting subset $B_{\text{non-overlap}} \subseteq B$ satisfies the following:

$$\forall b_i, b_j \in B_{\text{non-overlap}}, \quad i \neq j \implies \text{IoU}(b_i, b_j) \leq \tau_{\text{IoU}}. \quad (2)$$

3.2. Dominant Color Estimation

Before the actual masking step, it is important to determine the mask color that will effectively simulate real world occlusions. Most masking-based techniques [37, 38] use a fixed color mask for all training images. However,

as seen in Fig. 1, the color of the real-world occlusions is region-specific and depends on the environment in which the image is captured. We propose using the dominant color of the image as the mask color, as the occluders in the GWHD 2021 dataset are most likely to have colors that are dominant in the respective images.

The dominant color of the image is identified using histogram analysis, which determines the most frequently occurring intensity values across the RGB channels. Let I represent an image with N pixels. The intensity distribution of each channel $c \in \{\text{R}, \text{G}, \text{B}\}$ is represented as a histogram H_c , where $H_c[k]$ denotes the number of pixels with intensity k in channel c :

$$H_c[k] = \sum_{i=1}^N \mathbb{1}(I_c[i] = k), \quad k \in \{0, 1, \dots, 255\}, \quad (3)$$

where $\mathbb{1}(\cdot)$ is the indicator function that outputs 1 if the condition is true and 0 otherwise. The dominant intensity $k_{\text{dom},c}$ for each channel is defined as the intensity value with the maximum frequency:

$$k_{\text{dom},c} = \arg \max_k H_c[k]. \quad (4)$$

The dominant color C_{dom} is then expressed as:

$$C_{\text{dom}} = (k_{\text{dom},\text{R}}, k_{\text{dom},\text{G}}, k_{\text{dom},\text{B}}). \quad (5)$$

This dominant color represents the objects most likely to cause occlusions in the image and is used to create the mask, enhancing the robustness of wheat head detection in real-world scenarios.

3.3. Masking

After identifying the mask color and removing the list of overlapping bounding boxes, a second probabilistic sampling is applied to $B_{\text{non-overlap}}$, where each bounding box b_i is retained with probability p_m , producing a refined subset B_{masked} :

$$B_{\text{masked}} = \{b_i \in B_{\text{non-overlap}} : v_i \leq p_m\}, \quad (6)$$

where $v_i \sim \text{Uniform}(0, 1)$. For each bounding box $b_i = (x_i, y_i, w_i, h_i) \in B_{\text{masked}}$, a random region $b_{i,\text{mask}}$ within b_i is sampled for masking. The parameters (x'_i, y'_i, w'_i, h'_i) of $b_{i,\text{mask}}$ are randomly chosen to ensure that they are fully contained within b_i :

$$w'_i \sim \text{Uniform}(0, \alpha_w \cdot w_i), \quad h'_i \sim \text{Uniform}(0, \alpha_h \cdot h_i). \quad (7)$$

$$x'_i \sim \text{Uniform}(x_i, x_i + w_i - w'_i), \quad y'_i \sim \text{Uniform}(y_i, y_i + h_i - h'_i), \quad (8)$$

where α_w and α_h determine the percentage of areas to be sampled. The masking operation M occludes the sampled region $b_{i,\text{mask}}$ by modifying the pixel values in that region to the dominant color C_{dom} :

$$M(I, b_{i,\text{mask}}) = C_{\text{dom}} \cdot \mathbb{1}_{\text{mask}(b_{i,\text{mask}})} + I \cdot (1 - \mathbb{1}_{\text{mask}(b_{i,\text{mask}})}), \quad (9)$$

where $\mathbb{1}_{\text{mask}(b_{i,\text{mask}})}$ is an indicator function that assigns the dominant color to pixels inside $b_{i,\text{mask}}$ and leaves the rest of the image unchanged. The augmented image I_{aug} is generated as:

$$I_{\text{aug}} = M(I, b_{i,\text{mask}}) \quad \text{for all } b_i \in B_{\text{masked}}. \quad (10)$$

This probabilistic sampling framework, coupled with dominant color masking, systematically creates occlusion-rich training samples. The use of probabilities p_s and p_m allows fine-grained control over the augmentation process, generating diverse occlusion scenarios to improve the robustness of wheat head detection models under real-world conditions. Fig. 3 presents a visual demonstration of our proposed approach. Note that the black mask was used for a visual demonstration. However, we use the approach described in Sec. 3.2 to mask the randomly selected regions of the selected bounding boxes. The algorithm 1 summarizes the proposed approach using a step-by-step procedure.

4. Experimental Results

In this section, we present the quantitative and qualitative analysis of the proposed BBoxCut algorithm on the GWHD 2021 dataset. We show that our data augmentation technique can aid the wheat head localisation performance across three different kinds of object detection architectures (Faster R-CNN, FCOS, DETR). We perform ablation studies to understand how the parameters of BBoxCut affect the performance of the output model. We also show an improvement in the ability of the trained model to detect in the presence of occlusions.

Algorithm 1 Dominant Color-Based Masking and Augmentation

Require: Training set $\mathcal{D} = \{I_1, I_2, \dots, I_M\}$, Bounding boxes $B = \{b_1, b_2, \dots, b_N\}$, probabilities p_m , area percentage α_w, α_h , IoU threshold τ_{IoU} , augmentation probability p_{aug} .

Ensure: Augmented dataset \mathcal{D}_{aug} .

```
1: for each  $I_i \in \mathcal{D}$  do
2:   Sample  $r_i \sim \text{Uniform}(0, 1)$ 
3:   if  $r_i \leq p_{\text{aug}}$  then
4:     Remove overlapping boxes from  $B$  using IoU and threshold  $\tau_{\text{IoU}}$  as
       in Eq. (2).
5:     Compute histograms for each RGB channel  $H_R, H_G, H_B$  of  $I_i$  as
       in Eq. (3).
6:     Estimate dominant color  $C_{\text{dom}}$  from histograms as in Eq. (5).
7:     Apply probabilistic sampling to  $B_{\text{non-overlap}}$  with probability  $p_m$ ,
       resulting in  $B_{\text{masked}}$  as in Eq. (6).
8:     for each  $b_m \in B_{\text{masked}}$  do
9:       Compute masked region  $b_{m,\text{mask}}$  as in Eqs. (8) & (7).
10:      Apply masking  $M(I_i, b_{m,\text{mask}})$  using dominant color  $C_{\text{dom}}$  as in
        Eq. (9).
11:     end for
12:      $\mathcal{D}_{\text{aug}} \leftarrow \mathcal{D}_{\text{aug}} \cup I_{i,\text{aug}}$ .
13:     continue
14:   end if
15:    $\mathcal{D}_{\text{aug}} \leftarrow \mathcal{D}_{\text{aug}} \cup I_i$ .
16: end for
17: Output  $\mathcal{D}_{\text{aug}}$ .
```

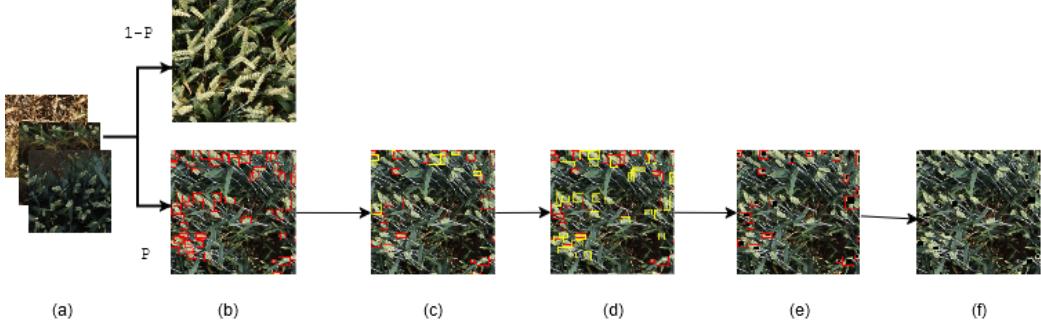


Figure 3: BboxCut Augmentation. (a) Batch of images from training set. (b) Randomly sample the images from training set to which BboxCut is to be applied. (c) Identify non-overlapping boxes using IoU metric. (d) Randomly sample onto which BboxCut needs to be applied. (e) Randomly select the region of masking based on the percentage. (f) Augmented image. Color codes: Selected bounding boxes ■, Unselected bounding boxes ■■, BBoxcut mask ■■■.

4.1. Dataset

The Global Wheat Head Detection (GWHD) 2021 dataset [4] is the largest dataset proposed for multiple phenotyping tasks. It contains 6,000 RGB images with a resolution of 1024×1024 pixels, collected between 2015 and 2020 by 16 institutions in 11 countries, covering genotypes from Europe, Africa, Asia, Australia, and North America. The dataset includes images taken under diverse lighting conditions and at various developmental stages, such as post-flowering and ripening, which are clustered into domains or sessions. The training set consists of 3,657 images in 18 domains. The validation set comprises data from 11 domains, totaling 1,476 images, while the test set includes data from 18 domains, totaling 1,382 images.

4.2. Network Architecture

We evaluate the performance of our data augmentation on diversified detector types namely Faster R-CNN, FCOS and DETR. FCOS and Faster R-CNN are single-stage and two-stage object detection architectures, respectively, while DETR is a single-stage transformer-based object detection architecture. All networks are initialized with COCO pretrained weights.

4.3. Training Details

For all detectors, we used an early stop criterion with a patience of 10 epochs. Adam (learning rate = 0.00001, batchsize=16, weight decay =

0.0001) was used as an optimizer for all experiments. ReduceLRonPlateau schedulers [46] (factor=0.1, patience=5, threshold = 0.0001, min-lr = 0, eps = 1e-08) have been used as a learning rate scheduler. The models for Faster R-CNN and FCOS are imported from the torchvision library, while the DETR was used from the META’s official repository. All models were trained on the Colab platform. We used mean average precision (mAP) as a metric to compare performance across various detectors. From empherical analysis, we found that the best values for the constants used in our approach are as follows: $p_{aug} = 0.3$, $p_m = 0.3$, $\alpha_w = 0.3$, $\alpha_h = 0.3$, $\tau_{IoU} = 0.5$.

4.4. Quantitative Analysis

In this subsection, we compare the quantitative performance of the proposed approach with state-of-the-art masking-based data augmentation techniques. Although there are several masking-based approaches in the literature as described in Sect. 2.3.1, we pick the most relevant techniques that specifically aim to model occlusions for comparison. Table 1 presents the performance of our approach in three different object detection architectures, in comparison with existing masking-based augmentation methods. The results demonstrate that our approach consistently outperforms competitive techniques, highlighting the importance of using constrained masking in complex datasets like GWHD 2021 to effectively simulate occlusion scenarios.

The most comparable method is that proposed by [40], but our approach demonstrates superior performance in all objects detection models tested. This can be attributed to the fact that [40] does not consider the nature of bounding boxes, increasing the likelihood of masking already occluded wheat heads. This masking can further degrade the model’s ability to detect wheat heads correctly, leading to suboptimal performance. Furthermore, our method shows a significant improvement over the technique introduced by [38], which often applies masks randomly throughout the image, including to the background areas. This random masking does not accurately simulate the real occlusion conditions encountered in complex phenotyping scenarios like those in GWHD 2021. In contrast, our approach leverages information about the spatial distribution of wheat heads, selectively applying masks to specific regions to better mimic realistic occlusion events. In doing so, our method encourages the detection network to learn more robust features for wheat head detection, resulting in improved performance across various object detection architectures.

	Faster R-CNN	FCOS	DETR
Baseline	51.14	57.18	57.40
CutOut [38]	51.29	57.87	58.20
Region-aware Random Erasing [40]	52.23	58.82	58.50
BBoxCut (Ours)	53.90	60.44	59.30

Table 1: Quantitative Analysis for GWHD 2021 dataset (mAP scores).

4.5. Qualitative Analysis

In this subsection, we present a qualitative analysis of the proposed approach compared to competitive techniques. Fig. 4 shows the detection results from vanilla Faster R-CNN, Cutout, Region-Aware Random Erasing, and the proposed approach in the first, second, third, and fourth columns, respectively. In the first row of Fig. 4, the baseline and competitive techniques either mislocalize the overlapping wheat heads or produce duplicate detections, whereas the proposed approach accurately localizes them. In the second row, the proposed method successfully detects all six wheat heads, while the other techniques miss some detections. In the third row, a small wheat head partially occluded by a leaf is correctly localized using the proposed approach, whereas other methods fail to detect it. The superior performance of our method can be attributed to its controlled and adaptive masking strategy. By identifying overlapping wheat heads and strategically selecting the mask color based on the dominant image region, our approach minimizes the risk of masking already occluded wheat heads. This targeted augmentation enhances the model’s ability to learn discriminative features, resulting in more accurate localization.

4.6. Influence on the choice of mask color

One of the key contributions of the proposed approach is the adaptive masking strategy to model occlusions. In this subsection, we analyze the impact of the chosen mask color on the performance of the detector. Some of the popular mask colors include black, white, gray, and random. However, these are generic choices and may not accurately capture the nature of occlusions in real-world scenarios. We demonstrate that an adaptive color mask performs better than existing mask color choices. Table 2 presents the performance of the Faster R-CNN detector with different mask color choices. The results show that the proposed approach, in the last two rows

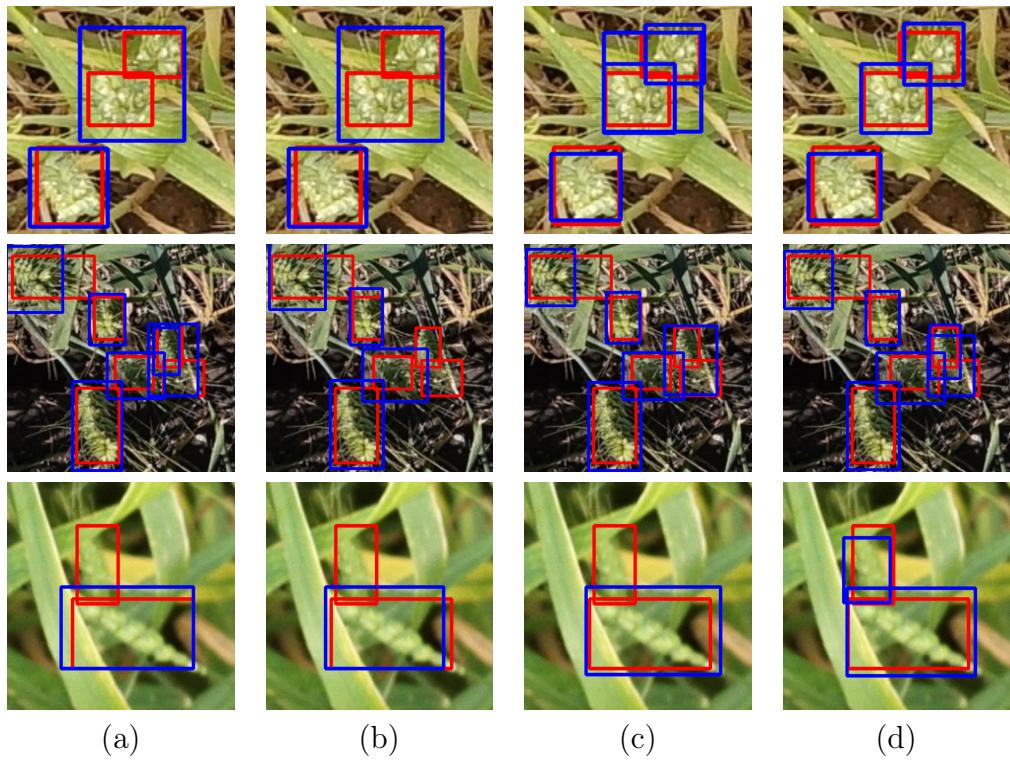


Figure 4: Qualitative Analysis for Faster R-CNN detector. Detections of: (a) Baseline, (b) Cutout, (c) Region-aware Random Erasing [40], and (d) Proposed approach. Red box: Ground truth. Blue Box: Prediction.

Mask Color	mAP
Black	48.75
Gray	50.51
White	52.75
Random	52.70
Global Dominant Color (Ours)	53.90

Table 2: Faster R-CNN performance (mAP) on the choice of mask color.

of Table 2, outperforms the other mask color choices, demonstrating that adaptive masking leads to improved detection accuracy by mitigating the limitations of static masks and providing a more realistic training scenario. The primary reason for this improvement is that adaptive color masks more effectively model real-world occlusions by dynamically selecting colors that blend naturally with the surrounding region. In contrast, static mask colors introduce a distribution shift between training and real-world data, as they do not accurately resemble actual occlusions such as objects partially covering one another in complex scenes. This distribution shift can lead to unnatural artifacts that hinder detection performance. In addition, generic masks can obscure valuable contextual information, making it harder for the model to learn robust features for occluded objects. By adapting to the context, the proposed approach ensures that the occlusions resemble real-world scenarios, preserving critical cues and improving generalization. Furthermore, adaptive masking enhances robustness by exposing the detector to a diverse range of occlusions, reducing its reliance on specific patterns and improving its ability to handle varying lighting conditions, textures, and object interactions.

5. Conclusions

In this study, we proposed a constrained masking-based data augmentation technique to improve wheat head detection under occlusion conditions in the GWHD 2021 dataset. The method outperformed state-of-the-art masking-based augmentation techniques across multiple object detection architectures, both qualitatively and quantitatively, indicating that our approach is agnostic to the detection network. This also demonstrates that our method effectively simulates realistic occlusion scenarios by carefully identifying wheat heads with minimal overlap and selectively applying masks based on histogram analysis, ensuring meaningful data transformations. In general, this work

highlights the importance of incorporating domain-specific constraints into data augmentation strategies to improve object detection performance. In future research, we plan to extend our approach to other datasets and explore adaptive, multistage augmentation techniques for further improvements in occlusion modeling and domain generalization.

Acknowledgments

This work has been partially supported by Research England (Lincoln Agri-Robotics) as part of the Expanding Excellence in England (E3) Program and the School of Computer Science, University of Lincoln, UK. The views expressed in this publication are solely those of the authors and not necessarily those of the funders.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT and Writefull to improve the readability and language of the work. After using this tool/service, the author(s) reviewed and edited the content as needed and takes (s) full responsibility for the content of the publication.

References

- [1] J. A. Foley, N. Ramankutty, K. A. Brauman, E. S. Cassidy, J. S. Gerber, M. Johnston, N. D. Mueller, C. O’Connell, D. K. Ray, P. C. West, et al., Solutions for a cultivated planet, *Nature* 478 (7369) (2011) 337–342.
- [2] S. Chakraborty, A. C. Newton, Climate change, plant diseases and food security: an overview, *Plant pathology* 60 (1) (2011) 2–14.
- [3] S. Madec, X. Jin, H. Lu, B. De Solan, S. Liu, F. Duyme, E. Heritier, F. Baret, Ear density estimation from high resolution rgb imagery using deep learning technique, *Agricultural and forest meteorology* 264 (2019) 225–234.
- [4] E. David, M. Serouart, D. Smith, S. Madec, K. Velumani, S. Liu, X. Wang, F. Pinto, S. Shafiee, I. S. Tahir, et al., Global wheat head detection 2021: an improved dataset for benchmarking wheat head detection methods, *Plant Phenomics* 2021 (2021).

- [5] Z. Zou, K. Chen, Z. Shi, Y. Guo, J. Ye, Object detection in 20 years: A survey, *Proceedings of the IEEE* 111 (3) (2023) 257–276.
- [6] T. DeVries, G. W. Taylor, Improved regularization of convolutional neural networks with cutout, *arXiv preprint arXiv:1708.04552* (2017).
- [7] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: *Advances in Neural Information Processing Systems* 25, Curran Associates, Inc., 2012, pp. 1097–1105.
- [8] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [9] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, *arXiv preprint arXiv:1804.02767* (2018).
- [10] B. Alexey, C. Wang, H. Mark Liao, Optimal speed and accuracy of object detection, *arXiv preprint arXiv:2004.10934* (2020).
- [11] G. Jocher, et al., ultralytics/yolov5: v6. 2-yolov5 classification models, apple m1, reproducibility, clearml and deci. ai integrations, Zenodo. org (2022).
- [12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, Ssd: Single shot multibox detector, in: *European conference on computer vision*, Springer, 2016, pp. 21–37.
- [13] Z. Tian, C. Shen, H. Chen, T. He, Fcos: A simple and strong anchor-free object detector, *IEEE transactions on pattern analysis and machine intelligence* 44 (4) (2020) 1922–1933.
- [14] R. Girshick, Fast r-cnn, in: *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.
- [15] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [16] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-end object detection with transformers, in: *European conference on computer vision*, Springer, 2020, pp. 213–229.

- [17] A. Gupta, S. Narayan, K. Joseph, S. Khan, F. S. Khan, M. Shah, Ow-detr: Open-world detection transformer, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 9235–9244.
- [18] A. Ayvaci, Occlusions and their role in object detection in video, University of California, Los Angeles, 2012.
- [19] P. M. Blok, R. Barth, W. Van Den Berg, Machine vision for a selective broccoli harvesting robot, IFAC-PapersOnLine 49 (16) (2016) 66–71.
- [20] P. R. Narkhede, A. V. Gokhale, Color image segmentation using edge detection and seeded region growing approach for cielab and hsv color spaces, in: 2015 International Conference on Industrial Instrumentation and Control (ICIC), IEEE, 2015, pp. 1214–1218.
- [21] Y. Zhu, Z. Cao, H. Lu, Y. Li, Y. Xiao, In-field automatic observation of wheat heading stage using computer vision, Biosystems Engineering 143 (2016) 28–41.
- [22] X. Xu, H. Li, F. Yin, L. Xi, H. Qiao, Z. Ma, S. Shen, B. Jiang, X. Ma, Wheat ear counting using k-means clustering segmentation and convolutional neural network, Plant Methods 16 (1) (2020) 1–13.
- [23] C. Zhou, D. Liang, X. Yang, H. Yang, J. Yue, G. Yang, Wheat ears counting in field conditions based on multi-feature optimization and twsvm, Frontiers in plant science 9 (2018) 1024.
- [24] J. A. Fernandez-Gallego, S. C. Kefauver, N. A. Gutiérrez, M. T. Nieto-Taladriz, J. L. Araus, Wheat ear counting in-field conditions: high throughput and low-cost approach using rgb images, Plant methods 14 (1) (2018) 1–12.
- [25] M. M. Hasan, J. P. Chopin, H. Laga, S. J. Miklavcic, Detection and analysis of wheat spikes using convolutional neural networks, Plant Methods 14 (1) (2018) 1–13.
- [26] A. Lad, M. Raval, Resolving issues with wheat head detection: A use case of xai in agriculture scenario, Data Science for Agriculture and Natural Resource Management (12 2021).

- [27] A. G. Singh, K. Singh, S. A. Sampson, Wheat head detection and crop health classification system, in: 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), IEEE, 2021, pp. 01–05.
- [28] K. Seemakurthy, C. Fox, E. Aptoula, P. Bosilj, Domain generalisation for object detection, arXiv preprint arXiv:2203.05294 (2022).
- [29] K. Saleh, S. Szénási, Z. Vámossy, Occlusion handling in generic object detection: A review, in: 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI), IEEE, 2021, pp. 000477–000484.
- [30] K. Saleh, S. Szénási, Z. Vámossy, Generative adversarial network for overcoming occlusion in images: A survey, Algorithms 16 (3) (2023) 175.
- [31] J. Ruan, H. Cui, Y. Huang, T. Li, C. Wu, K. Zhang, A review of occluded objects detection in real complex scenarios for autonomous driving, Green energy and intelligent transportation 2 (3) (2023) 100092.
- [32] S. Gilroy, E. Jones, M. Glavin, Overcoming occlusion in the automotive environment—a review, IEEE Transactions on Intelligent Transportation Systems 22 (1) (2019) 23–35.
- [33] Z. Ouardirhi, S. A. Mahmoudi, M. Zbakh, Enhancing object detection in smart video surveillance: A survey of occlusion-handling approaches, Electronics 13 (3) (2024) 541.
- [34] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, F. Shen, Image data augmentation for deep learning: A survey, arXiv preprint arXiv:2204.08610 (2022).
- [35] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, CoRR abs/1409.4842 (2014).
URL <http://arxiv.org/abs/1409.4842>
- [36] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, Random erasing data augmentation, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 34, 2020, pp. 13001–13008.

- [37] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, Random erasing data augmentation, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 34, 2020, pp. 13001–13008.
- [38] T. DeVries, G. W. Taylor, Improved regularization of convolutional neural networks with cutout. arxiv, arXiv preprint arXiv:1708.04552 10 (2017).
- [39] J. Huang, Z. Zhu, G. Huang, D. Du, Aid: Pushing the performance boundary of human pose estimation with information dropping augmentation, arXiv preprint arXiv:2008.07139 (2020).
- [40] Z. Yang, Z. Wang, W. Xu, X. He, Z. Wang, Z. Yin, Region-aware random erasing, in: 2019 IEEE 19th International Conference on Communication Technology (ICCT), IEEE, 2019, pp. 1699–1703.
- [41] K. K. Singh, H. Yu, A. Sarmasi, G. Pradeep, Y. J. Lee, Hide-and-seek: A data augmentation technique for weakly-supervised localization and beyond, arXiv preprint arXiv:1811.02545 (2018).
- [42] C. Gong, D. Wang, M. Li, V. Chandra, Q. Liu, Keepaugment: A simple information-preserving data augmentation approach, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 1055–1064.
- [43] P. Chen, S. Liu, H. Zhao, X. Wang, J. Jia, Gridmask data augmentation, arXiv preprint arXiv:2001.04086 (2020).
- [44] S. Feng, S. Yang, Z. Niu, J. Xie, M. Wei, P. Li, Grid cut and mix: flexible and efficient data augmentation, in: Twelfth International Conference on Graphics and Image Processing (ICGIP 2020), Vol. 11720, SPIE, 2021, pp. 656–662.
- [45] P. Li, X. Li, X. Long, Fencemask: a data augmentation approach for pre-extracted image features, arXiv preprint arXiv:2006.07877 (2020).
- [46] PyTorch: ReduceLROnPlateau — PyTorch 1.9.0 documentation, Accessed: 2022-09-01 (2022).
URL https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html