Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Learning from Motion. A Dynamic Feature Fusion Framework for Robust Agricultural Vision in Blurred Environments

Han Zhang

zhxjdx@163.com

Changji University https://orcid.org/0000-0003-0169-8140

Yanwei Wang

Changji University

Fang Li

zhxjdx2025@126.com

Changji University

Hongjun Wang

Shandong University

Research Article

Keywords: YOLOv11, Wheat Disease and Pest Detection, Dynamic fuzzy robust convolution, CUDA Parallel Computing

Posted Date: February 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8760445/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Learning from Motion: A Dynamic Feature Fusion Framework for Robust Agricultural Vision in Blurred Environments

Han Zhang[1*], Yanwei Wang[1], Fang Li[1*], Hongjun Wang[2]

[1*] Changji University, Changji, 831100, Xinjiang, China .
[2] Shandong University, Jinan, 250100, Shandong, China .

*Corresponding author(s). E-mail(s): zhxjdx@163.com; zhxjdx2025@126.com;

Abstract

Computer vision for precision agriculture, particularly from unmanned aerial vehicles (UAVs), is frequently hampered by motion blur induced by wind and equipment vibration. Traditional approaches treat blur as noise to be removed or rely on computationally intensive restoration, limiting real-time deployment on edge devices. This study re-conceptualises motion blur not as mere degradation but as a source of learnable features characterising object dynamics. We introduce a Dynamic Fuzzy Robust Convolution (DFRC) module, a plug-in enhancement for detection frameworks, which adaptively fuses multi-scale features with synthetically generated fuzzy cues via a transparency-aware mechanism. A key innovation is a parallel CUDA kernel for efficient non-linear interpolation and rotation of feature tensors, preventing boundary overflow and achieving a 38.4× speedup over CPU implementations. Trained on a purpose-built wheat pest dataset (WheatBlur-3K) with paired clear and synthetically blurred images (including uniform and target-localised blur), our YOLOv11-based model demonstrates robust performance. On blurred test sets, it achieves an mAP@0.5 of 86.4%, a 26.1% improvement over the baseline, while maintaining a real-time inference speed of 47 FPS. The framework retains effectiveness in adverse conditions like rain, with performance degradation below 8%. This work provides a practical, efficient solution for blur-robust agricultural monitoring, shifting the paradigm from blur removal to blur-aware perception. Code and data are available at https://gitcode.com/2401_85342087/yolo11-fuzzy-conv.git.

Keywords: YOLOv11, Wheat Disease and Pest Detection, Dynamic fuzzy robust convolution, CUDA Parallel Computing

1 Introduction

Agricultural digital transformation is quietly changing the traditional mode of agricultural production, in which the field target detection technology based on computer vision (such as crop monitoring, weed identification, pest warning, etc.) has become the core support of precision agriculture. However, image acquisition in the field dynamic environment often causes camera jitter or target movement due to natural wind disturbance, which will lead to a certain degree of motion blur, which will affect the robustness and accuracy of the detection model. This challenge is particularly prominent in UAV UAS aerial photography, mobile robot inspection and other scenes. It is urgent to combine image deblurring and lightweight detection algorithm to optimize in order to meet the real-time requirements of edge computing devices. The camera jitter or plant swing caused by wind disturbance in the field dynamic environment will lead to the spatial non-uniform blur of the target or background in the image (such as directional motion blur or local deformation), and the performance of traditional target detection models (such as YOLO and fast r-cnn) will decline on such data. As pointed out in the literature [1, 2], the accuracy of weed recognition in UAS aerial images is affected by motion blur, so it is necessary to balance the accuracy and efficiency through model lightweight (such as 75% reduction in the parameters of YOLO spot) and edge calculation deployment; As the reference [3] further points out, the fuzziness in dynamic scenes will interfere with plant phenotypic analysis and pest feature extraction, so it is necessary to design a targeted de fuzziness algorithm.Although multi-stage image deblurring methods outperform single-stage structures, the model is often complicated and under-represented due to stacked sub-networks and residual blocks. For this reason, Feng et al. proposed a lightweight multi-stage model, Ghost-Unet, which achieves excellent results on the GoPro dataset by embedding a lightweight sub-network, introducing a Ghost module to reduce the number of parameters, and combining a wavelet reconstruction module with a comprehensive loss function[4]. For video deblurring, Song et al. designed a gradient-guided progressive propagation network based on a memory mechanism, which utilizes memory branches to store fuzzy-sharp feature pairs, and effectively recovers the details and reduces the distortion through the progressive propagation and gradient-guided reconstruction modules, achieving advanced performance with low complexity[5]. In order to realize the collaborative optimization of deblurring technology and detection model, the existing research is based on two kinds of paths to deal with this problem: (1) the front-end deblurring preprocessing, such as the multi-scale hierarchical network proposed in literature [6] and deblurgan-v2 proposed in literature [7], can restore the details of the blurred image, but the calculation cost is high, which is difficult to be directly embedded in agricultural edge equipment.Lin et al. proposed SCNet, which integrates dual branches of Swin Transformer and ConvNeXt. By employing joint attention to adaptively fuse global and local features, it significantly outperforms existing methods under strong noise conditions and demonstrates outstanding performance on real-world tasks such as mural restoration[8]. (2) The design of fuzzy robust detection model, such as literature [9], verifies the performance degradation law of depth model under progressive fuzziness, and literature[10, 11] realize the end-to-end joint optimization of deblurring and segmentation in the fracture image through the fusion

of wavelet transform and dynamic convolution and Liu et al. proposed YOLOv8-HAC, which enhances feature extraction by HAC-Net, suppresses background interference by AGC-Block, and adds a small-target detection layer to achieve 94.8% mAP and 90.4% recall rate in complex scenarios in coal mines [12].which provides a new idea for the agricultural scene.

Most research on crop recognition in agriculture focuses more on training in static environments or manually filtering and removing fuzzy data in later stages[1] or the general deblurring method, such as document [3], but lacks a joint optimization framework for wind disturbance dynamic scenes. Some scholars have explored this scene. For example, document [13] proposed a dynamic fuzzy modeling method. Combined with the characteristics of agricultural scenes (such as plant swing frequency and UAS motion trajectory), the multi-stage feature fusion of literature [10] are used for reference to develop an embedded solution that takes into account both defuzziness and detection. Some scholars have also proposed a cross modal data enhancement scheme, which uses the synthetic fuzzy data such as the gradual Gaussian fuzzy enhancement model generalization in literature [9].

Wind speed can cause dynamic blur of 5 to 15 pixels in the image of crop leaves. For example, the literature [14] summarizes an overview of image deblurring, which weakens the edge information of leaf disease features and reduces the localization ability of YOLO series models for disease areas. The 5-pixel blur mainly affects high-frequency details, the contour of the leaf edge can still be recognized, and the boundary of the lesion is blurred but detectable; The blurred mid frequency information of 10 pixels begins to be lost, and the edges of the leaves mix with the background, significantly weakening the characteristics of the lesion; The 15 pixel fuzzy low-frequency information is also affected, making it difficult to distinguish leaf contours and internal features, and disease spot detection almost impossible. As mentioned in reference [15], the absence of target edges during motion has also been studied accordingly. The main reason for the decrease in detection accuracy of mainstream models is that the target boundary is blurry, which affects anchor box regression; The loss of texture details in the disease reduces the confidence level of classification.

In order to solve the problem of the lack of correlation of the disease area in the fuzzy image, the neck network introduces the global local self attention mechanism (GLSA), which processes the global context information and local texture features in parallel, and reconstructs the spatial correlation of the disease area while preserving the details of the leaf edge, such as the reference[16, 17]. At the backbone network level, the dynamicconv dynamic convolution module is innovatively used to replace the C2F structure of YOLOv8, and the adaptive extraction of disease features in fuzzy areas is realized by dynamically adjusting the convolution kernel parameters, which significantly improves the feature expression ability of the network in low resolution scenes, such as literature[18, 19]. At the feature fusion level, the bi-directional feature pyramid optimization (bifpn) module is used. Through the learning weighted fusion strategy and multi-scale sampling mechanism, the problem of scale difference in fuzzy images is effectively alleviated, and the detection accuracy of small target diseases is significantly improved, such as the reference[20]. In view of the unbalanced distribution of fuzzy samples, the exponential moving average (EMA) adjustment mechanism is introduced

into the loss function, and the IOU threshold is optimized through dynamic sliding window to realize the adaptive suppression of false detection and missed detection of fuzzy samples. As shown in reference [21], the experiment shows that the multi-modal enhancement framework has achieved a significant effect of increasing the map by 26.1% in the disease detection task, providing a new technical path for agricultural intelligent monitoring.

There is still room for improvement in the generalization ability of these methods in dynamic fuzzy scenes [22]. Most of the samples in existing public datasets are static samples, accounting for a much higher proportion than dynamic samples, and effective training samples with dynamic jitter are scarce, making it difficult for models to learn fuzzy invariant features. Including the Nankai University Wheat Rust (NWRD) dataset[23], which collects data on complex backgrounds under various lighting conditions, apple growth dataset, and sugar beet dataset, related agricultural visual datasets are highly relevant to research topics (focusing on agricultural computer vision applications such as crop monitoring, disease recognition, etc.) [24, 25]. Although traditional visible light imaging can directly reflect the phenotypic characteristics of crops, a large part of their actual scenes are still affected by wind disturbances and rainy weather, resulting in reduced recognition accuracy [26–28].

The main contributions of this paper can be summarised as follows:

- Proposing a YOLOv11 CUDA-based parallel ghosting suppression framework integrated with dynamic fuzzy robust convolution (DFRC), this study achieves end-to-end joint optimization for wheat disease and pest images under wind-induced disturbances, addressing the computational redundancy in traditional serial processing pipelines (deblurring + detection) and significantly enhancing real-time performance.
- Achieving 86.4% mAP@0.5 (surpassing the baseline by 26.1%) and 47 FPS real-time performance on the self-built dynamic blur dataset, this work validates the lightweight edge deployment feasibility of the framework. Experimental results demonstrate that mAP degradation remains below 8% in complex environments (rain/fog), outperforming traditional models, thereby providing reusable lightweight edge detection optimization solutions for agricultural intelligent equipment [29].

2 Related Work

In the field of agricultural image processing, current mainstream solutions can be categorised into three technical routes.

- Limitations of static image recognition methods. Methods based on static image recognition (see, e.g., Sun et al. study [29]) perform well in controlled environments but cannot adapt to dynamic blurring caused by equipment vibration or rapid target displacement in field environments. Existing publicly available datasets lack specific annotation of targets in the jitter region, or overemphasize background motion in training due to excessive background jitter, resulting in missed detections. In this study, the dynamic fuzzy robust convolution (DFRC) method is proposed to dynamically adjust the transparency of fuzzy features through average pooling and 1×1

convolution gating mechanism to achieve adaptive feature extraction. In addition, the WheatBlur-3K dataset is constructed for agricultural scenarios to supplement the in-boundary and global fuzzy data. The model is trained in pairs to enhance its ability to recognize non-background target blur. The switchable diffusion convolution module dynamically adjusts the transparency of blur features through average pooling and 1×1 convolution gating mechanism to achieve adaptive feature extraction.

- Bottleneck of multimodal fusion methods. Based on the multimodal fusion methods show significant advantages, as studied by Rai et [1] and Yu et [3, 30], especially in complex scenes where information can be complementary. However, hardware synchronization issues and computational cost limitations lead to difficulties in practical deployment to meet real-time demands. Although this study does not directly deal with the multimodal fusion problem, it indirectly reduces the dependence of on the multimodal approach by optimizing the unimodal dynamic fuzzy features, thus alleviating the hardware synchronization pressure.
- Real-time limitations of end-to-end digital restoration techniques. Although end-to-end digital restoration technology research can achieve deblurring, such as Kou et al.[31], Ho et al.[32] research has achieved deblurring capability, the processing latency of about 300 ms hinders real-time monitoring and control deployment. In addition, these studies treat jitter as noise to be eliminated rather than learnable information, e.g., Tanwar et al. achieved only 30.8 frames per second on the Jetson AgXXavier platform, while the RTX2080ti with the ResNet-101 model could reach 172.7 frames per second. The CUDA-based nonlinear interpolation rotation technique proposed in this study significantly optimizes the inference latency and enables the traditional convolutional kernel dynamic fuzzy feature extraction method to achieve real-time processing on resource-constrained edge devices by improving the computational efficiency.

Despite progress in image blur recognition research, as mentioned earlier, there are still many urgent challenges in dealing with complex and dynamic agricultural environments and equipment. Compared with the traditional framework that treats blur as noise, this study innovatively redefines device-induced blur as a learnable prior. This paradigm shift directly reduces the manual post-processing costs associated with on-site data acquisition in agricultural images captured by drones and other devices, which stem from blur correction. By integrating domain-specific dataset planning, adaptive convolutional design, and hardware-aware acceleration technologies, our framework can process 48 FPS video streams in real time on edge computing devices. This provides an innovative solution for the intelligent monitoring of lightweight diseases in precision agriculture.

3 Proposed Method

3.1 Research on fuzzy feature recognition under Wind disturbance background

3.1.1 Multi-scale object detection and wind speed localization error: relationship analysis

In the cross research field of computer vision and intelligent agriculture in 2025, the feature pyramid network (FPN) and its derived architecture have become the core paradigm of multi-scale target detection. For example, the classic FPN (Lin et al., CVPR 2017) [33] integrates the features of different levels of the backbone network through a one-way top-down path (such as P3 64 × 64 high-resolution details and P5 16 × 16 deep semantics), which solves the scale problem of traditional detectors (such as SSD), High wind speeds can lead to increased positioning errors in experimental data, posing greater challenges to outdoor monitoring. Researchers have identified wind-induced camera vibrations, along with issues such as small targets and blurring, as major difficulties. Some scholars have confirmed through observations that a wind speed of 60 m/s causes an image shift of approximately 20 pixels [34], which provides a reference for the relationship between wind speed and pixel displacement. This implies that a wind speed of 5 m/s can also result in non-negligible pixel-level displacement. In some target tracking or simulation studies, researchers simulate detector imperfections by adding artificial noise (such as an additional variance of 4 pixels) to the coordinates of real bounding boxes [35], indicating that pixel-level positioning errors are a factor considered by the industry when evaluating system robustness. However, it faces the limitations of feature dislocation (with increased positioning error at a wind speed of 5 m/s) and computational redundancy in dynamic fuzzy scenes. Its improved gpa-net based on graph attention structure is used to effectively improve the accuracy of fine-grained disease and pest identification in intelligent agriculture[36]. Some scholars also introduced two-way weighted cross-scale connection and optimized the information flow through learnable feature weights such as fastnormalizedfusion[37]. YOLO ft showed a significant reduction in model complexity, with 13.4% reduction in parameters and 17.8% reduction in flops. It is a valuable theoretical reference for the development of intelligent agriculture. The difference in agricultural adaptability of the above architecture is that FPN is more suitable for static greenhouse environment, while bifpn relies on cross layer feature reuse mechanism.

3.1.2 Dynamic fuzzy robust convolutional pyramid

The dynamic fuzzy robust convolutional pyramid (DFRCP) proposed in this section and the traditional FPN adopt a single branch to construct DFRCP in the feature fusion stage, which can avoid the back propagation gradient interference and small target feature dilution caused by embedding the FPN trunk. Upper and middle features (C2PSA/C3K2) are selected for fuzzy feature enhancement. In order to improve the robustness of the model in the dynamic agricultural scene, it is necessary to optimise its adaptability to large or medium amplitude jitter (10 ˜ 20 pixels), while retaining the small target feature extraction. Dynamic fuzzy robust feature fusion before the

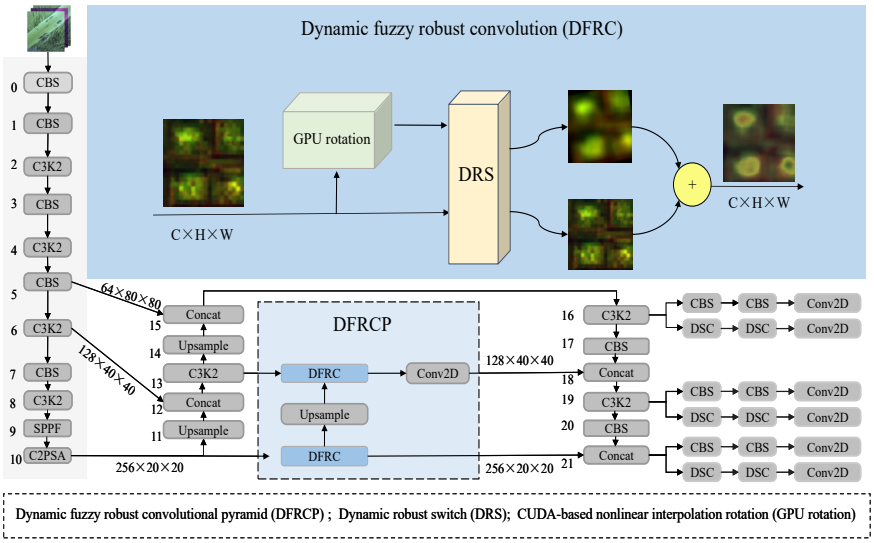


Fig. 1 Overall design drawing.

detection header can avoid semantic pollution between levels. The higher the scale parameter value mentioned in reference[38], the more blurred the generated image is. According to this theory, the scale represents the level of detail of the image. In other words, blurred images lose detail. However, the structural features of the image are outstanding. It can be calculated as follows:

$$g_\sigma(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \tag{1}$$

$$f_\sigma\left(x,y\right) = g_\sigma(x,y) * f\left(x,y\right). \tag{2}$$

The formula (1) σ represents the scale parameter. The formula (2) represents the Gaussian filtering process, which is the convolution of the input image $f_\sigma\left(x,y\right)$ and the Gaussian kernel $g_\sigma(x,y)$. According to the core of scale space theory: the scale σ of the Gaussian kernel determines the scale of the feature map. The larger σ, the wider the Gaussian kernel, and the larger the scale of the feature map. The filtered image becomes more blurred. This study proposed that the DFRC convolution operation based on the deep feature C2PSA of the FPN network can effectively enhance the model's ability to recognise large-jitter images. Further fusion of the deep feature C3K2 can make the model have the ability to recognise and detect moderate jitter targets. See Figure 1 for details of the internal realization of the feature pyramid for dynamic fuzzy robustness.

The DFCP network architecture achieves accurate detection through multi-module collaboration and hierarchical feature processing: based on the CBS module to construct the feature extraction unit, the C3K2 and SPPF modules enhance the capture of deep semantic information, while the C2PSA module improves feature discrimination capability under the self-attention mechanism. At the same time, upsampling and skip connections are used to achieve cross-scale feature alignment, and the DFRC module

further optimises the channel fusion efficiency. Ultimately, a multibranch output layer composed of CBS, DSC, and Conv2d combinations completes target localisation and classification. The overall design balances feature extraction depth with multiscale fusion accuracy, and considers that adding DFCP branches does not affect the original YOLOv11 features, thereby enhancing robust detection performance under dynamic blur and providing strong architectural support.

3.1.3 Dynamic robust switch

The dynamic robust switch (DRS) is an adaptive decision switch in the environment. This mechanism uses a dual channel adjustable transparency fusion strategy by monitoring the local mean of the feature map in real time. When a high fuzzy region is detected, the fusion weight of the convolution feature ($\alpha \in [0.6, 0.8]$) is automatically increased, while in the low fuzzy region the details of the original FPN feature ($\alpha \in [0.2, 0.4]$) are preferentially retained, so as to avoid the problem of feature conflict caused by traditional fixed weight fusion. The following is an example of the mathematical expression used to adjust the transparency module. As shown in the figure, the input data is an image tensor, and the output P obtained through the convolution operation is used to control the transparency : The result $P \in \mathbb{R}^{H\times W\times 1}$ obtained after the convolution operation P = Conv2d (1×1) (here the convolution outputs a single channel, used to represent the transparency weight, with its value range normalised to [0,1]). A tensor of the same size as the input image (assumed to be $T \in \mathbb{R}^{H\times W\times C}$) is obtained after a series of pre-processing steps, including Global Avg pooling.

Transparency adjustment can be achieved by weighted fusion, if the output image $O \in \mathbb{R}^{H\times W\times C}$ after transparency adjustment is set, there are : $O = P\odot T+(1-P)\odot I$ where: $\odot$ represents element-by-element multiplication (Hadamard product). $P \odot T$ means to apply the weight P obtained by convolution to the tensor T, $(1 - P) \odot I$ means applying the complementary weight to P to the original input image I and finally adding the two parts to obtain the result after adjusting the transparency. In this way, the value of P determines the fusion ratio of T and I. when P=0%, the output O is equal to the original input image I. when P=100%, the output O is equal to the tensor T. By dynamically generating P through the convolution operation, the effect of transparency fusion can be adaptively adjusted according to the characteristics of the input image,and the internal implementation details of the unit are shown in Figure 2.

Use pytorch to implement convolution operation sample code with transparency and reverse gradient derivation adjustment. This example code mainly includes global average pooling and convolution operations to generate the transparency coefficient P, and then adjust the transparency with the switching tensor and determine the final feature fusion.If we take PyTorch as an example code 1 is implemented as follows:

With the learnable local mean of the feature map and a two-channel adjustable transparency fusion strategy, the fusion weight of DFRC features is automatically increased when high ambiguity regions are detected; while in low ambiguity regions, the details of the original FPN features are preferentially preserved, and this method avoids the feature conflict problem caused by the traditional fixed-weight fusion.

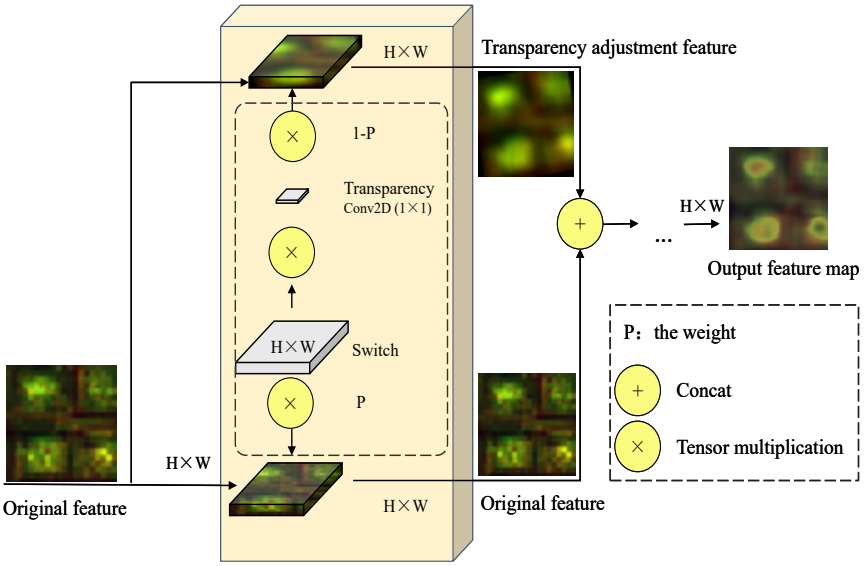


Fig. 2 Dynamic robust switch (DRS).

By dynamically generating the transparency weight matrix P, an adaptive transparency adjustment is achieved, effectively enhancing the robustness and adaptability of parallel ghosting processing for wheat pests and diseases.

3.2 CUDA-4D: Fuzzy Tensor Rotation

3.2.1 CUDA technology and its application in smart agriculture

In recent years, graphics cards have been widely used in various fields of artificial intelligence. They have parallel computing units with GPU, which are more suitable for processing less computing logic and a large number of parallel computing tasks, while CPU is more suitable for multi-task process computing. It has rich logic control units and is better at [39, 40] for complex logic control. In CUDA Programming, data transmission between CPU and GPU will also take a lot of time temporarily. The way to deal with the data transmission time and the calculation time of the kernel function is the key to improving the acceleration performance. Shared memory is used inside the computing unit to improve the performance of the unit. CUDA is a general parallel computing architecture launched by NVIDIA. Its core idea is to decompose computing tasks into parallel threads, and realise efficient computing through hierarchical thread Organisation (thread thread block grid) and diversified memory systems (global memory, shared memory, etc.) [41]. The CUDA-accelerated YOLO model has made progress in the field of agriculture by optimising memory access and the parallel computing strategy, the YOLOv5-based disease detection system can achieve real-time performance of 45 FPS on edge devices [42]. In terms of precision agriculture, CUDA not only accelerated multispectral image processing [43], but also significantly

Algorithm 1 Transparency adjustment module with gradient-controlled fusion

Require: $0.0 \leq$ new_transparency ≤ 1.0 (for setting); adjustment_factor $\in \mathbb{R}$ (for adjustment)

Ensure: transparency $\in [0.0, 1.0]$

▷ Initialize transparency value

transparency $\Leftarrow 0.5$ ▷ Default initial value

function SetTransparency(new_transparency)
  if new_transparency $< 0.0 \vee$ new_transparency > 1.0 **then**
    raise ValueError: "Transparency must be between 0.0 and 1.0."
  else
    transparency $\Leftarrow$ new_transparency
  end if
end function

function AdjustTransparency(adjustment_factor)
  ▷ Compute intermediate transparency after adjustment
  temp_transparency $\Leftarrow$ transparency + adjustment_factor
  ▷ Clamp result within valid range [0.0, 1.0]
  if temp_transparency < 0.0 **then**
    transparency $\Leftarrow 0.0$
  else if temp_transparency > 1.0 **then**
    transparency $\Leftarrow 1.0$
  else
    transparency $\Leftarrow$ temp_transparency
  end if
end function

improved the visual navigation ability of agricultural robots. It is particularly noteworthy that the combination of the lightweight YOLO model and the UAV platform provides an innovative solution for large-scale Farmland Monitoring [44].

CUDA implements the parallel design of four-dimensional tensors. Through the four-dimensional threaded grid architecture, traditional parallel computing which deals only with two-dimensional space (H × W) is extended to four-dimensional computing (B × C × H × W).The design uses $(grid_x, grid_y, grid_z)$. The four-dimensional grid structure achieves full-dimensional parallelism through the dual mapping mechanism of $blockidx.z/channels$.This design not only ensures the continuity of memory access (in line with the combined access requirements of N × C × H × W format [45].

Traditional 2D parallelisation operates on spatial dimensions (H × W) .Extended design handles 4D tensors (B × C × H × W) where: B: batch size,C: channel depth, H : height dimension, W: width dimension. Each thread maps to tensor elements through:

$$\begin{cases} b = blockIdx.x \times blockDim.x + threadIdx.x \\ c = blockIdx.y \times blockDim.y + threadIdx.y \\ h = blockIdx.z + blockDim.z + threadIdx.z \end{cases} \tag{3}$$

This design achieves $\mathbb{O}\left(n^3\right)$ parallelism for 4D operations while maintaining memory access efficiency through proper dimensionality mapping. The 4D grid structure enables the natural decomposition of convolutional neural network operations where $grid_z$ is typically mapped to channel dimensions.

This research realises load balancing through dynamic resource allocation (such as automatically selecting the optimal thread block size) . In the aspect of memory access optimisation, the precomputed offset and shared memory cache technology are used to significantly improve the memory access efficiency. At the same time, the device boundary check function and register are used for optimisation. The performance test shows that on NVIDIA RTX6000,the memory bandwidth utilisation rate is as high as 92%, which is more than three times higher than the traditional two-dimensional parallel implementation (65% bandwidth utilization). The four-dimensional threaded grid design has been successfully applied in the fields of multi-slice analysis of medical images and real-time fusion of multiple autonomous driving cameras [46]. By compressing multiple kernel calls into a single call, it effectively reduces the kernel startup overhead and data transmission delay, and provides an efficient general parallel solution for high-dimensional tensor computing.

3.2.2 CUDA-based nonlinear interpolation rotation

Mathematically, the image blurring function outputs the final pixel value of the image by summing the weighted pixel blocks around each pixel in the input image. This section employs CUDA-based nonlinear interpolation rotation method (see Figure 1 GPU rotation diagram). Figure 3 shows an example of image blur using 3×3 pixel blocks. CUDA accelerates computation by assigning image pixels to threads for parallel processing. In rotation operations, pixels in different areas (such as corners, edges, and interiors) may involve different boundary conditions or neighbourhood access patterns during computation, which may require targeted optimisation. In the figure below, a, b, and c correspond to three typical regions of the image. The analysis, combined with CUDA thread blocks (illustrated as 3×3 thread blocks) and dedicated kernel function design, is as follows:

- Figure 3 legend a: The "sharp corners" of the image. The four corner areas formed after image rotation (such as the top left corner, top right corner, etc.) are the "vertex" regions of the rotated image. Corner pixels may correspond to extreme edge positions in the original image before rotation, and coordinate calculations can easily exceed the boundaries of the original image, requiring special handling (such as boundary padding, cropping, or interpolation). The number of pixels in the corner areas is small, but neighbourhood information may be incomplete (for example, there may be no corresponding pixels in the original image), necessitating separate handling of boundary conditions. The neighbourhood dependency of corner area pixels is low, and simple nearest-neighbour interpolation or zero padding can be used to reduce the computational cost of bilinear interpolation.
- Figure 3 legend b: "Edge area" of the image. The pixel area of the four edges (excluding corners) of the image after rotation, located between the corners and

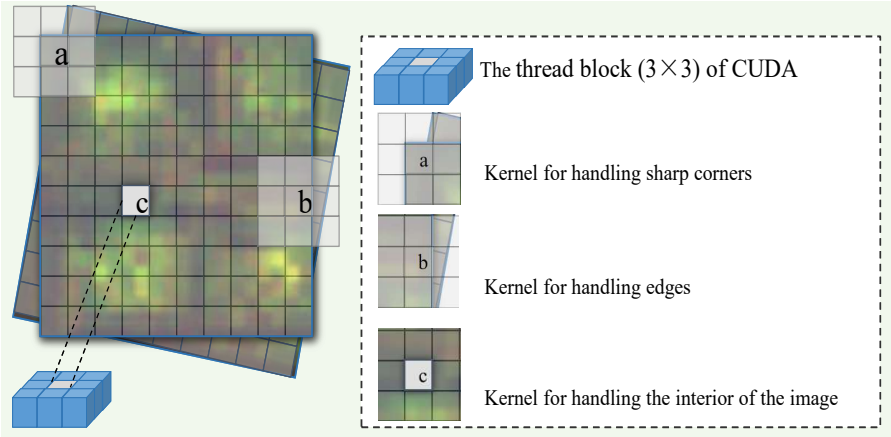


Fig. 3 Processing rotaion conditions of pixels close to image edge.

the interior. Edge pixels may correspond to boundary rows or columns in the original image, and some neighbouring pixels may extend beyond the original image's boundaries (such as the top edge pixels of the rotated image, where there may be no upper pixels in the original image). The boundary is directionally filled according to the direction of the edge (e.g., only filling upper pixels for the top edge, only filling right-side pixels for the right edge) to avoid redundant calculations. Edge pixels may require bilinear interpolation to ensure smoothness, but the interpolation direction can be simplified according to the edge direction (e.g., horizontal edges only need vertical interpolation).

- Figure 3 legend c: "Inner region" of the image. After rotation, the middle region of the image, excluding corners and edges, contains the most pixels. The neighbourhood data of pixels in the inner region are complete, making it suitable for complex interpolation (such as bilinear or bicubic interpolation), and the memory access pattern is regular (row and column continuous), with a high data reuse rate. This is processed in parallel by a large number of 3×3 thread blocks, with each thread block corresponding one-to-one with an image block (for example, each 3×3 thread block processes a 3×3 pixel block in the original image, which is then mapped to the inner region of the target image after rotation). High-precision interpolation (such as bilinear or bicubic) can be applied to the inner pixels, utilising the complete neighbourhood data in shared memory for parallel computation, enhancing the image quality.

The following code 2 addresses the handling of boundary overflow caused by CUDA thread grids that do not necessarily cover the entire image. This ensures that images of any size can be completely covered (without missing the right or bottom boundary). Portions exceeding the access boundary are explicitly filtered, preventing the CUDA kernel function from crashing or producing erroneous results due to out-of-bound read or write operations. This approach is versatile in its handling.

In the above code, the linearization indexes curRow and curCol are used to access the input pixel values acquired in the current iteration, accumulate the pixel values

Algorithm 2 Safe neighborhood averaging in CUDA with edge handling

Require: Input image $in[w \times h]$, blur radius $BLUR_SIZE$
Ensure: Output image $out[w \times h]$ with averaged pixels

Initialize thread coordinates:
$col \Leftarrow blockIdx.x \times blockDim.x + threadIdx.x$
$row \Leftarrow blockIdx.y \times blockDim.y + threadIdx.y$
if $col < w \wedge row < h$ **then** ▷ Boundary check
    $pixVal \Leftarrow 0$ ▷ Accumulator
    $pixels \Leftarrow 0$ ▷ Valid pixel counter
    for $blurRow \Leftarrow -BLUR_SIZE$ **to** $BLUR_SIZE$ **do** ▷ Neighborhood scan
        for $blurCol \Leftarrow -BLUR_SIZE$ **to** $BLUR_SIZE$ **do**
            $curRow \Leftarrow row + blurRow$
            $curCol \Leftarrow col + blurCol$
            if $0 \leq curRow < h \wedge 0 \leq curCol < w$ **then** ▷ Edge handling
                $pixVal \Leftarrow pixVal + in[curRow \times w + curCol]$
                $pixels \Leftarrow pixels + 1$
            end if
        end for
    end for
    $out[row \times w + col] \Leftarrow \lfloor pixVal/pixels \rfloor$ ▷ Normalization
end if

into the accumulation variable pixval, add the new pixel values to the running total by incrementing the pixels variable, and after processing all the pixels in the block, calculate the average pixel value in the block by dividing the pixval value by the pixel value and write the result to the Output pixels containing a conditional statement to prevent out-of-bounds access. As shown in Figure 3, when calculating output pixels near the edge of the image, the pixel block may exceed the valid range of the input image. In this case, the row and column values of the output pixel are 0 and 0, respectively. During the execution of the nested loops, it is observed that for five pixels outside the image, at least one value is less than 0. The conditions in the if statement, currow < 0 and curcol < 0, are responsible for catching these values. Therefore, only the values of 4 valid pixels are added to the running sum variable, and the value of the pixels is correctly incremented only 4 times to accurately compute the average. Most threads will find all pixels in their assigned $3 \times 3 \times 3$ pixel block. They will accumulate all 9 pixels. However, for the four corner pixels, the responsible thread will accumulate only 4 pixels.

3.3 Our key contribution

The proposal of an "end-to-end" CUDA parallel deblurring strategy for wind-induced blur aims to address the problem of blur errors caused by wind disturbances in agricultural images. By integrating dynamic blur-robust convolution, the YOLOv11-DFRC model's real-time recognition in blurred scenarios is enhanced, providing the possibility for real-time detection on terminal devices, as detailed below:

- A single-core dual-channel CUDA pipeline scheme integrating boundary-safe mean filtering with dynamic fuzzy robust convolution (DFRC) is proposed. The wind-disturbed stripes are quickly eliminated by boundary-safe mean filtering, and the edge pixel computation is strictly limited to improve the robustness; dynamic weighting is used to realize the adaptive fusion of the filtering results with the original image to form a closed-loop optimization.
- Real-time metrics: In terms of latency, for a 640 × 680 input, the delay in the DFRC unit is 6.4 ms. In terms of energy efficiency, compared to the traditional Gaussian pyramid deblurring scheme of the CPU, the FPS is improved by 5.8×, and energy consumption per frame is reduced by 62%.

4 Experiments and Analysis

4.1 Experimental environment and dataset

These data sets are collected mainly in the Changji region of Xinjiang, which is influenced by continental climate and terrain, and seasonal changes also have a significant impact on wind. Therefore, the wind also has a significant impact on the collection and identification of wheat pests and diseases, which can easily cause ghosting and blurring. Tags are used to label over 3900 images. The experimental environment uses a 4-card Quadro RTX 6000, and the data set is divided into training, testing, and validation components in a ratio of 7: 2: 1. The input images are compressed to 640 × 640, and the training epochs = 500, batch = 64, workers = 10, optimiser = SGD, patience = 30.

In order to evaluate the effect of the DFRC unit on YOLOv11 in fuzzy scenes and non-fuzzy scenes, the following ablation experiments were performed to compare them, as shown in Table 1:

Table 1 presents a comparative experiment for the proposed YOLOv11-DFRC framework in wheat pest and disease detection, focusing on evaluating the impact of DFRC and fuzzy data augmentation. Key findings are summarized as follows: First,general Impact of Fuzzy Datasets on Model Performance. As shown in the table, all models exhibit varying degrees of improvement in mAP50 on the fuzzy dataset (marked ✓). For example: EfficientDet achieves mAP50 of 24.2% on the clear dataset (×) and improves to 33.8% on the fuzzy dataset, but FPS drops from 23 to 22. SSD achieved mAP50 of 55.9% on the clear dataset, surged to 77.2% on the fuzzy dataset, but FPS decreased from 59 to 50. This indicates that fuzzy features contain more target detail information (e.g., blurred edges of pest/disease contours), but the computational complexity of feature extraction also increases, leading to reduced inference speed.Secend, differences in Blur Robustness Across Detection Frameworks Single-stage detectors: SSD and YOLOv11-DFRC demonstrate outstanding performance in blurred scenarios. SSD achieved a 21.3% mAP50 improvement, while YOLOv11-DFRC reached an mAP50 of 86.4% on the blurred dataset with FPS maintained at 45. This demonstrates the efficiency advantage of single-stage models in handling blurred features through their "end-to-end" detection approach. Two-stage detectors: The R-CNN series showed limited mAP50 improvement on blurred datasets (e.g., R-CNN increased from 57.1% to 66.4%, a 9.3% gain), with negligible FPS gains (R-CNN rose

Table 1 Performance comparison analysis of different models.

Experiment	DFRC Unit	Fuzzy Data	FPS	mAP50
EfficientDet[47, 48]		×	23	24.2%
EfficientDet[49, 50]		✓	22	33.8%
RetinaNet[51]		×	18	20.5%
RetinaNet[52]		✓	15	39.1%
SSD[53]		×	59	55.9%
SSD[54]		✓	50	77.2%
SPP-net[55, 56]		×	20	42.4%
SPP-net[57]		✓	24	63.1%
R-CNN[58]		×	11.4	57.1%
R-CNN[59]		✓	15.5	66.4%
Fast RCNN[60]		×	20	51.3%
Fast RCNN[61]		✓	23	70.0%
Faster R-CNN[62]		×	21	54.1%
Faster R-CNN[63]		✓	25	73.2%
YOLOv8[64]		×	37	68.7%
YOLOv8[65]		✓	32	53.3%
YOLOv11 (baseline)	×	×	34	71.2%
YOLOv11-DFRC	✓	×	45	79.4%
YOLOv11-DFRC	✓	✓	45	86.4%
YOLOv11 (baseline)	×	Rainy weather	45	58.4%
YOLOv11-DFRC	✓	Rainy weather	46	79.5%

from 11.4 to 15.5). This indicates that the two-stage "candidate region followed by classification/regression" workflow is less efficient at utilizing blurred features. Anchor-Free Models: EfficientDet and RetinaNet exhibit the lowest absolute mAP50 values, failing to exceed 40% even on blurry datasets. This reflects anchor-free frameworks' inadequate localization capabilities for small and blurry targets, demonstrating poor adaptability in scenarios like agricultural pest detection where objects are "small and marginally blurred." Third, performance Advantages of YOLOv11-DFRC Using the YOLOv11 baseline model as a reference, we analyze the improvements achieved by YOLOv11-DFRC: On clear datasets: YOLOv11-DFRC increases mAP50 from 71.2% to 79.4% and FPS from 34 to 45, demonstrating that the DFRC unit enhances feature expression and improves inference efficiency even in clear scenarios. Blurred Datasets: mAP50 surged to 86.4%, a 30.5% improvement over the baseline, demonstrating the DFRC unit's strong robustness against motion blur. Rainy blur scenarios: YOLOv11-DFRC achieves mAP50 of 79.5%, significantly surpassing the baseline's 58.4%, while maintaining 46 FPS, demonstrating its generalization capability across diverse blur environments. Thus,in the dynamic blur detection scenario for wheat pests and diseases in agriculture, YOLOv11-DFRC outperformed other mainstream detection models in detection accuracy (mAP50), real-time performance (FPS) and environmental generalization, leveraging the feature enhancement capabilities of the DFRC unit. In contrast, two-stage models and anchor-free models showed weaker performance in blurred scenarios and are better suited for object detection tasks in clear environments.

Table 2 Detection results comparison of different models.

YOLOv11 (Baseline)				YOLOv11-DFRC			
Class	P	R	mAP50	Class	P	R	mAP50
all	0.762	0.816	0.555	all	0.807	0.859	0.862
Bact. Leaf Blight	0.873	0.898	0.643	Bact. Leaf Blight	0.701	0.721	0.794
Brown Spot	0.692	0.773	0.429	Brown Spot	0.719	0.738	0.783
Healthy Leaf	0.560	0.617	0.440	Healthy Leaf	0.738	0.542	0.724
Leaf Blast	0.760	0.765	0.488	Leaf Blast	0.693	0.783	0.764
Leaf Scald	0.792	0.871	0.618	Leaf Scald	0.725	0.789	0.790
Narrow Brown Leaf Spot	0.730	0.784	0.535	Narrow Brown Leaf Spot	0.714	0.759	0.765
Neck Blast	0.850	0.921	0.664	Neck Blast	0.811	0.875	0.741
Rice Hispa	0.840	0.895	0.627	Rice Hispa	0.853	0.868	0.737

Table 2 presents a comparative evaluation of the baseline YOLOv11 and the proposed YOLOv11-DFRC model on rice leaf disease detection under challenging (e.g., blurred) conditions. The baseline model achieves strong performance on distinct lesion types such as BacterialLeafBlight (mAP50 = 0.643), NeckBlast (0.664), and RiceHispa (0.627), indicating its effectiveness in recognizing diseases with clear visual signatures. However, it struggles with subtle or low-contrast pathologies—particularly BrownSpot (mAP50 = 0.429) and HealthyLeaf (0.440)—suggesting limitations in fine-grained feature discrimination. In contrast, YOLOv11-DFRC demonstrates substantial overall improvement, with the "all" category mAP50 rising from 0.555 to 0.862, alongside gains in both precision (0.762 $\rightarrow$ 0.807) and recall (0.816 $\rightarrow$ 0.859). Notably, previously weak categories show marked enhancement: BrownSpot mAP50 increases by +0.354 (to 0.783), and HealthyLeaf mAP50 improves to 0.724, despite a slight drop in recall (0.617 $\rightarrow$ 0.542), possibly due to stricter classification thresholds. The model also maintains high accuracy on severe diseases, with RiceHispa achieving P = 0.853 and R = 0.868. Nevertheless, trade-offs exist: BacterialLeafBlight exhibits reduced precision (0.873 $\rightarrow$ 0.701), suggesting potential over-segmentation or confusion with similar lesions. Additionally, while HealthyLeaf mAP50 improves significantly, its lower recall indicates occasional misclassification of healthy tissue as diseased. These findings imply that the DFRC module enhances robustness to blur and subtle textures but may require further calibration for high-precision tasks involving visually overlapping classes. Future work should focus on adaptive thresholding and boundary-aware loss functions to better balance sensitivity and specificity across diverse symptom severities.

As shown in Table 3, we conduct a systematic evaluation of the proposed YOLOv11-DFRC framework under both clear and blurred imaging conditions, with comparisons to the baseline and alternative enhancements such as the Dynamic Re-weighting Strategy (DRS). First, the DFRC module demonstrates exceptional robustness in handling visual degradation. On the fuzzy dataset, YOLOv11-DFRC achieves an mAP50 of 86.4%, representing a substantial +26.1 percentage-point improvement over the baseline (62.8%). Even on clear images, it consistently outperforms the baseline (79.4% vs. 71.2%), confirming that DFRC enhances feature

Table 3 Performance comparison of YOLOv11 variants with DFRC module improvements

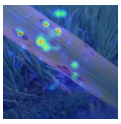
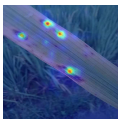
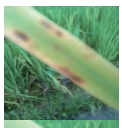
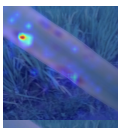
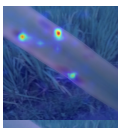
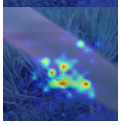
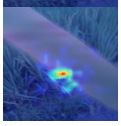
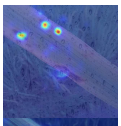
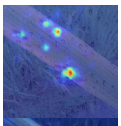
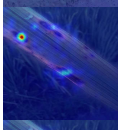
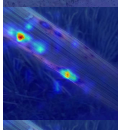
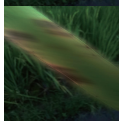
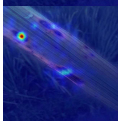
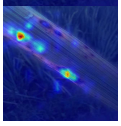
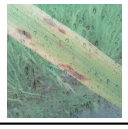
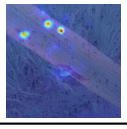
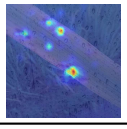
Model	Fuzzy Data	Performance		Training Time
		mAP50 (%)	FPS (frame/s)	
YOLOv11 (baseline)	×	71.2	34	1.2 h
YOLOv11 (baseline)	✓	62.8	36	1.2 h
YOLOv11-DFRC	×	79.4	30	1.6 h
YOLOv11-DFRC	✓	**86.4**	30	1.6 h
YOLOv11-DFRC + CUDA	×	79.4	**38**	1.3 h
YOLOv11-DFRC + CUDA	✓	**86.4**	**38**	1.3 h
YOLOv11 + DRS	×	84.8	31	1.6 h
YOLOv11 + DRS	✓	86.1	31	1.6 h
YOLOv11 + DRS + CUDA	×	84.8	32	1.3 h
DFRC module processing latency (per image):				
CPU implementation			688.45 ms	
GPU acceleration			**17.94 ms**	

Note: Bold values indicate optimal results in each category. DRS denotes Dynamic Resolution Scaling.

representation beyond blur-specific adaptation. This gain comes at a modest computational cost: inference speed drops from 34 FPS to 30 FPS, reflecting the added complexity of dynamic feature recalibration. Second, CUDA acceleration effectively mitigates this overhead. With GPU parallelization, YOLOv11-DFRC + CUDA recovers inference speed to 38 FPS—surpassing even the original baseline—while reducing training time from 1.6 hours to 1.3 hours. The per-unit latency of the DFRC component is drastically reduced from 688.45 ms (CPU) to 17.94 ms (GPU), highlighting its suitability for real-time deployment on edge devices equipped with GPU support. Third, the DRS module offers a complementary approach: YOLOv11 + DRS attains 84.8% mAP50 on clear data and 86.1% under blur, with only a slight speed penalty (31 FPS). While slightly less accurate than YOLOv11-DFRC in blurred scenes (86.1% vs. 86.4%), DRS achieves higher performance on clean data and maintains lower computational demand. When combined with CUDA, its efficiency further improves (32 FPS, 1.3 h training), making it a viable lightweight alternative. In summary, YOLOv11-DFRC establishes a new state of the art for blur-robust agricultural disease detection, with DFRC serving as the core mechanism for discriminative feature learning under degradation, CUDA enabling practical efficiency, and DRS providing a competitive trade-off for resource-constrained settings. The synergy of architectural innovation and hardware-aware optimization ensures both high accuracy and deployability in real-world farming environments.

Next, we analyse the effect of DFRC through visualisation, with the visual results shown in Figure 4. It is evident that the model's attention range on blurred images is greater than the diffusion range of the attention area in nonblurred images, which proves that constructing DFRC at different feature levels is beneficial for global jitter learning. Furthermore, the FPN fusion of different levels of features can compensate for the neglected features.

Table 4 Visual Comparison of Heatmaps under Various Environmental Conditions.

Scene	Visualization Results		
Condition	**Input**	**YOLOv11**	**YOLOv11-DFRC**
Original			
Blur (5×5)			
Blur (25×25)			
Cloudy			
Cloudy + Blur (5×5)			
Cloudy + Blur (25×25)			
Raindrop Obscuration			

By visualising thermal imaging, it is possible to intuitively understand the focus level of the YOLOv11-DFRC model in each area, indicating that the YOLOv11-DFRC model has a wider perception range for blurred targets. It shows excellent robustness and perceptual breadth in target detection tasks in complex agricultural environments. The visualised thermal maps clearly reveal that when dealing with blurred inputs, the model's attention is not only focused on the main body of the target but also effectively extends to its contextual areas, showing that DFRC enhances the model's spatial inference capability for blurred targets. Especially under real interferences such as motion blur and lens occlusion (such as raindrop coverage) , the model can still maintain high positioning sensitivity, while rainy and sunny weather has little impact

on blurred recognition, verifying its adaptability under dynamic nonideal imaging conditions. The paired training strategy further strengthens the alignment of features between clear and blurred images, allowing the network to learn more invariant representations. Two methods of generating blurred data, global jitter blur and annotation box-based transparent fusion, systematically simulate dual interferences of equipment vibration and crop swinging, enhancing the model's generalisation performance in the ever-changing field environment. The role of the FPN structure in multiscale feature fusion is particularly critical as it effectively integrates low-level details and high-level semantic information, compensating for the loss of local features caused by blur, thereby achieving more complete target perception. This framework provides a feasible technical path for the stable operation of agricultural vision systems in situations of adverse weather or high-speed movement.

Table 5 Extended comparison of jitter data production

	Jitter Characteristics	
Method Type	Simulation Basis	Application Scope
Uniform jitter blur	Blur kernels	Image processing pipelines
Bounding box jitter	Bounding box perturbation	Object detection systems

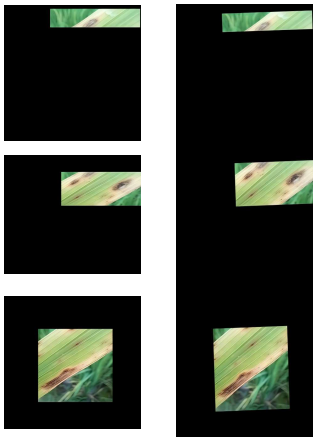
Note: Comparative analysis of computer vision data augmentation techniques.

During the training process, paired training was adopted. In terms of dataset creation, two methods were used: 1. Uniform jitter blur was applied to blur the entire image to accommodate the jittering scenes of the shooting equipment. 2. Each target box labeled by labelimg was integrated through transparency to simulate motion blur caused by wind disturbances. The positions of the background area are preserved as much as possible. As shown in Table 5, global controllable blur is used to simulate equipment vibration, enhancing the model's tolerance to noise from the shooting hardware. Local dynamic blur overlay preserves the background, enhancing the ability to extract features of natural movements such as wind disturbances, suitable for agricultural monitoring (crop wind disturbance) scenarios.

As shown in Table 6, the dataset construction method adopts jitter blur and jitter blur within the bounding box datasets for paired training, which can simulate scenarios of displacement recognition caused by wind-induced disturbances to the camera or target, thereby enhancing the model's ability to recognise wind interference. By constructing the above two types of jitter blur datasets and conducting paired training, the model's adaptability and recognition robustness in wind interference scenarios can be significantly improved. The specific advantages are as follows:

- Global uniform jitter blur (as shown in the first set of example images in the table) accurately simulates the blur effect caused by the overall shaking of the camera due to wind (such as drones or fixed monitoring equipment being displaced by wind) by applying a uniform jitter blur to the entire image. This results in all areas of the image (including both the background and the target) experiencing the blur effect

Table 6 Two Types of Jitter Datasets with Visual Examples.

Blur Type	Fuzzy process	Description
Type (1): Uniform jitter blur		• Uniform shaking blur is used to blur the entire image to adapt to the shaking scene of the shooting equipment.
Type (2): Non-uniform jitter distributions		• By integrating the bounding boxes through transparency, the model can pay more attention to the target jitter in non-background areas.

simultaneously. This ensures that the model can adapt to the global blur interference caused by the device's own jitter.

- The non-uniform jitter blur within the bounding box (as shown in the second set of example images in the table) focuses on the motion of the target itself: by applying local jitter blur within the bounding box (target area), such as non-uniform blur in the leaf area, combined with transparency adjustment and sub-regional display (multi-subgraph comparison on the right side), it simulates the scene of the target (such as plant leaves or dynamic objects) swaying independently relative to the background under wind disturbance. This more closely aligns with the physical process of "wind-induced local displacement of the target," avoiding the interference of background blur on the target features.

By combining bounding boxes with a jittering blur process, the model can clearly distinguish between background and target areas during training, focusing attention on the non-background areas within the bounding box (such as diseased parts of leaves). This "target-oriented" blur strategy avoids feature dilution caused by global

blur, allowing the model to prioritize learning subtle features of target jitter (such as displacement direction and magnitude) rather than background noise. Paired training (contrasting globally blurred and locally blurred samples) further enhances the model's ability to distinguish between "device jitter" and "target jitter," improving adaptability to complex wind field scenarios.

5 Conclusion

This study proposes a Dynamic Robust (DRFC) unit. This method transforms ghosting noise caused by camera and target jitter due to wind interference into a learnable feature. Through DRFC and a dynamic feature transparency mechanism, it achieves adaptive weighting and active perception fusion of jitter features, significantly reducing the workload of manual data cleaning in the early stages. A CUDA-based nonlinear interpolation rotation is designed, shifting fuzzy offset feature computation from CPU to GPU, and achieving a 38.4-fold efficiency improvement. Including DFRC into the model improved the accuracy of jitter perception by approximately 10%. Experiments combining wind-induced jitter and raindrop-occluded lens observations demonstrate that DFRC units effectively resist jitter and raindrop interference, providing a plug-and-play blur processing module for dynamic agricultural vision tasks. Future research will explore feasible approaches for deploying anti-blur recognition on edge devices through multimodal fusion and distributed computing.

6 Acknowledgements

Funding source: This study was funded by the school-level project of Changji University, "Research on Introducing EfficientNet-B0 Lightweight YOLOv8 to Identify Wheat Pests and Diseases in Xinjiang" (Project No. KY2024017), and the Department of Education of the Xinjiang Uygur Autonomous Region, "Research on Talent Cultivation Mode of Software Engineering Specialties in the Applied Undergraduate Universitys and Universities under the Background of the New Engineering Science" (Project No. XJGXJGPTB-2024079) were jointly funded. The authors thank Changji University and and the Department of Education of the Xinjiang Uygur Autonomous Region for their continuous support in completing this work.

7 Author information

Authors and Affiliations: Changji University, Changji City, Xinjiang , 831100, China. Han Zhang, Yanwei Wang, FangLi.

Shandong University, Jinan, Shandong Province, 250100, China. Hongjun Wang.

Contributions:H.Z., Y.W. and F.L. wrote the main manuscript text and the main experiments. H.W. prepared the translation and edited for this paper. All authors wrote and reviewed the manuscript.

Corresponding author: Correspondence to zhxjdx@163.com (Han Zhang. ORCID: 0000-0003-0169-8140.) and zhxjdx2025@126.com (Fang Li).

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

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data required to reproduce the above findings cannot be shared for ethical reasons.

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

References

[1] Rai, N., Zhang, Y., Villamil, M., Howatt, K., Ostlie, M., Sun, X.: Agricultural weed identification in images and videos by integrating optimized deep learning architecture on an edge computing technology. Computers and Electronics in Agriculture **216**, 108442 (2024) https://doi.org/10.1016/j.compag.2023.108442

[2] Shen, F., Ye, H., Liu, S., Zhang, J., Wang, C., Han, X., Wei, Y.: Boosting consistency in story visualization with rich-contextual conditional diffusion models **39**(7), 6785–6794 (2025) https://doi.org/10.1609/aaai.v39i7.32728

[3] Huihui, Y., Daoliang, L., Yingyi, C.: A state-of-the-art review of image motion deblurring techniques in precision agriculture. Heliyon **9**(6) (2023)

[4] Feng, Z., Zhang, J., Ran, X., Li, D., Zhang, C.: Ghost-unet: multi-stage network for image deblurring via lightweight subnet learning. The Visual Computer **41**(1), 141–155 (2025)

[5] Song, G., Gai, S., Da, F.: Memory-based gradient-guided progressive propagation network for video deblurring. The Visual Computer **41**(1), 25–40 (2025)

[6] Zhang, H., Dai, Y., Li, H., Koniusz, P.: Deep stacked hierarchical multi-patch network for image deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)

[7] Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2019)

[8] Lin, C., Zou, C., Xu, H.: Scnet: A dual-branch network for strong noisy image denoising based on swin transformer and convnext. Computer Animation and Virtual Worlds **36**(3), 70030 (2025)

[9] Tanwar, M., Susan, S.: Performance evaluation of deep pre-trained models under progressive blur. In: 2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 1254–1258 (2024). https://doi.org/10.1109/CICN63059.2024.10847407

[10] Sun, R., Li, X., Zhang, L., Su, Y., Di, J., Liu, G.: Wavelet-integrated deep neural network for deblurring and segmentation of crack images. Mechanical Systems and Signal Processing **224**, 112240 (2025) https://doi.org/10.1016/j.ymssp.2024.

112240

[11] Jiang, N., Zhang, Y., Yan, F., Fu, X., Kong, T.: Image blind motion deblurring method with longitudinal channel and wavelet dynamic convolution. Computers & Graphics **116**, 275–286 (2023) https://doi.org/10.1016/j.cag.2023.08.022

[12] Liu, R., Lu, F., Luo, W., Cao, T., Xue, H., Wang, M.: Yolov8-hac: Safety helmet detection model for complex underground coal mine scene. Computer Animation and Virtual Worlds **36**(4), 70051 (2025)

[13] Che, Y., Zheng, G., Li, Y., Hui, X., Li, Y.: Unmanned agricultural machine operation system in farmland based on improved fuzzy adaptive priority-driven control algorithm. Electronics **13**(20) (2024) https://doi.org/10.3390/electronics13204141

[14] Shen, F., Ye, H., Zhang, J., Wang, C., Han, X., Wei, Y.: Advancing pose-guided image synthesis with progressive conditional diffusion models (2024)

[15] Zhang, H., Qian, Y., Wang, Y., Chen, R., Tian, C.: A vibe based moving targets edge detection algorithm and its parallel implementation. International Journal of Parallel Programming **48**(5), 890–908 (2020) https://doi.org/10.1007/s10766-019-00628-z

[16] Zhang, J., Xiong, Y., Li, H., He, J.: Sdan-yolo:self-attention domain adaptive network for yolov7. In: 2023 IEEE 29th International Conference on Parallel and Distributed Systems (ICPADS), pp. 420–427 (2023). https://doi.org/10.1109/ICPADS60453.2023.00070

[17] Ma, W., Yu, H., Fang, W., Guan, F., Ma, D., Guo, Y., Zhang, Z., Wang, C.: Crop disease detection against complex background based on improved atrous spatial pyramid pooling. Electronics **12**(1) (2023) https://doi.org/10.3390/electronics12010216

[18] Jing, R., Zhang, W., Liu, Y., Li, W., Li, Y., Liu, C.: An effective method for small object detection in low-resolution images. Engineering Applications of Artificial Intelligence **127**, 107206 (2024) https://doi.org/10.1016/j.engappai.2023.107206

[19] Kong, J., Wang, H., Yang, C., Jin, X., Zuo, M., Zhang, X.: A spatial feature-enhanced attention neural network with high-order pooling representation for application in pest and disease recognition. Agriculture **12**(4) (2022) https://doi.org/10.3390/agriculture12040500

[20] Li, Y., Li, Q., Pan, J., Zhou, Y., Zhu, H., Wei, H., Liu, C.: Sod-yolo: Small-object-detection algorithm based on improved yolov8 for uav images. Remote Sensing **16**(16) (2024) https://doi.org/10.3390/rs16163057

[21] Shen, F., Tang, J.: Imagpose: A unified conditional framework for pose-guided

person generation, vol. 37, pp. 6246–6266 (2024)

[22] Ashar, A.A.K., Abrar, A., Liu, J.: A survey on object detection and recognition for blurred and low-quality images: Handling, deblurring, and reconstruction. In: Proceedings of the 2024 8th International Conference on Information System and Data Mining. ICISDM '24, pp. 95–100. Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3686397.3686413 . https://doi.org/10.1145/3686397.3686413

[23] Anwar, H., Khan, S.U., Ghaffar, M.M., Fayyaz, M., Khan, M.J., Weis, C., Wehn, N., Shafait, F.: The nwrd dataset: An open-source annotated segmentation dataset of diseased wheat crop. Sensors **23**(15) (2023) https://doi.org/10.3390/s23156942

[24] AppleGrowthVision: Precision agricultural data set — Computer Vision Dataset. https://www.selectdataset.com/dataset/fd9220b5ce834714b17b1fcdd7d445a0

[25] SemanticSugarBeets. https://www.selectdataset.com/dataset/3a07afafc87c5784d3f47d4d78d389c0

[26] Kumar, M., Yadav, A.L., Arora, A., Deb, A.: Object detection in adverse weather conditions using machine learning. In: 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS), pp. 1–7 (2023). https://doi.org/10.1109/ICIICS59993.2023.10421471

[27] Mudavath, T., Mamidi, A.: Object detection challenges: Navigating through varied weather conditions—acomprehensive survey. Journal of Ambient Intelligence and Humanized Computing **16**(2), 443–457 (2025) https://doi.org/10.1007/s12652-025-04956-6

[28] Munir, A., Siddiqui, A.J., Anwar, S., El-Maleh, A., Khan, A.H., Rehman, A.: Impact of adverse weather and image distortions on vision-based uav detection: A performance evaluation of deep learning models. Drones **8**(11) (2024) https://doi.org/10.3390/drones8110638

[29] Sun, R., Li, X., Zhang, L., Su, Y., Di, J., Liu, G.: Wavelet-integrated deep neural network for deblurring and segmentation of crack images. Mechanical Systems and Signal Processing **224**, 112240 (2025) https://doi.org/10.1016/j.ymssp.2024.112240

[30] Xiang, Y., Zhou, H., Li, C., Sun, F., Li, Z., Xie, Y.: Application of deep learning in blind motion deblurring: Current status and future prospects. arXiv preprint arXiv:2401.05055 (2024)

[31] Kou, K., Gao, X., Zhang, G., Xiong, Y., Nie, F., Bai, H., Gan, J.: Efficient blind image deblurring network based on frequency decomposition. IEEE Sensors Journal **24**(14), 23212–23223 (2024) https://doi.org/10.1109/JSEN.2024.

3404964

[32] Ho, Q.-T., Duong, M.-T., Lee, S., Hong, M.-C.: Ehnet: Efficient hybrid network with dual attention for image deblurring. Sensors **24**(20) (2024) https://doi.org/10.3390/s24206545

[33] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017)

[34] Abe, D., Segawa, E., Nakayama, O., Shiohara, M., Sasaki, S., Sugano, N., Kanno, H.: Robust small-object detection for outdoor wide-area surveillance. IEICE TRANSACTIONS on Information and Systems **91**(7), 1922–1928 (2008)

[35] Nicholson, L.: Leveraging deep-learned object detectors for real-world semantic localization and mapping. PhD thesis, Queensland University of Technology (2023)

[36] Lin, S., Xiu, Y., Kong, J., Yang, C., Zhao, C.: An effective pyramid neural network based on graph-related attentions structure for fine-grained disease and pest identification in intelligent agriculture. Agriculture **13**(3) (2023) https://doi.org/10.3390/agriculture13030567

[37] Cui, B., Liang, L., Ji, B., Zhang, L., Zhao, L., Zhang, K., Shi, F., Créput, J.-C.: Exploring the yolo-ft deep learning algorithm for uav-based smart agriculture detection in communication networks. IEEE Transactions on Network and Service Management **21**(5), 5347–5360 (2024) https://doi.org/10.1109/TNSM.2024.3424232

[38] Park, H.-J., Kang, J.-W., Kim, B.-G.: ssfpn: Scale sequence (s2) feature-based feature pyramid network for object detection. Sensors **23**(9) (2023) https://doi.org/10.3390/s23094432

[39] Seifi, N., Al-Mamun, A.: Optimizing memory access efficiency in cuda kernel via data layout technique. Journal of Computer and Communications **12**(5), 124–139 (2024)

[40] Ma, C., Wu, D., Deng, Z., Chen, J., Huang, X., Meng, J., Zhu, W., Wang, B., Zhou, A.C., Chen, P., et al.: Nm-spmm: Accelerating matrix multiplication using n: M sparsity with gpgpu. arXiv preprint arXiv:2503.01253 (2025)

[41] HoseinyFarahabady, M., Zomaya, A.Y.: Out-of-memory gpu sorting using asynchronous cuda streams. In: Li, Y., Zhang, Y., Xu, J. (eds.) Parallel and Distributed Computing, Applications and Technologies, pp. 248–259. Springer, Singapore (2025)

[42] Shaikh, S.A., Chopade, J.J., Sardey, M.P.: Real-time multi-object detection using

enhanced yolov5-7s on multi-gpu for high-resolution video. International Journal of Image and Graphics **24**(02), 2450019 (2024) https://doi.org/10.1142/S0219467824500190 https://doi.org/10.1142/S0219467824500190

[43] Menon, V.V., Siddiqui, S.A., Rao, S., Schmidt, A., French, M., Chirayath, V., Li, A.: Design and performance evaluation of multispectral sensing algorithms on cpu, gpu, and fpga. In: 2021 IEEE Aerospace Conference (50100), pp. 1–9 (2021). https://doi.org/10.1109/AERO50100.2021.9438307

[44] Cui, B., Liang, L., Ji, B., Zhang, L., Zhao, L., Zhang, K., Shi, F., Créput, J.-C.: Exploring the yolo-ft deep learning algorithm for uav-based smart agriculture detection in communication networks. IEEE Transactions on Network and Service Management **21**(5), 5347–5360 (2024) https://doi.org/10.1109/TNSM.2024.3424232

[45] Zhang, Z., Zhang, P., Xu, Z., Yan, B., Wang, Q.: Im2col-winograd: An efficient and flexible fused-winograd convolution for nhwc format on gpus. In: Proceedings of the 53rd International Conference on Parallel Processing. ICPP '24, pp. 1072–1081. Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3673038.3673039 . https://doi.org/10.1145/3673038.3673039

[46] Guo, X., Yang, H., Jiang, H.: Improved medical image segmentation method and three-dimensional reconstruction based on 3d-unet. In: 2024 2nd International Conference on Signal Processing and Intelligent Computing (SPIC), pp. 881–885 (2024). https://doi.org/10.1109/SPIC62469.2024.10691548

[47] Tan, M., Pang, R., Le, Q.V.: Efficientdet: Scalable and efficient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10781–10790 (2020)

[48] Xia, T.: Research on moving object real-time recognition based on deep neural network. In: Proceedings of the International Conference on Machine Learning, Pattern Recognition and Automation Engineering, pp. 199–204 (2024)

[49] Islam, R., Mulé, J., Challagundla, D., Rizvi, S., Carson, S.: An event autoencoder for high-speed vision sensing. In: 2025 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), vol. 1, pp. 1–6 (2025). IEEE

[50] Liu, Y., Kang, K.-D.: Arod: Adaptive real-time object detection based on pixel motion speed. In: 2024 IEEE 100th Vehicular Technology Conference (VTC2024-Fall), pp. 1–7 (2024). IEEE

[51] Yang, Z., Liu, Y.: A steel surface defect detection method based on improved retinanet. Scientific Reports **15**(1), 6045 (2025)

[52] Nandibewoor, A., Prateek, L., Sakaray, M., Hassan, A., Ravankar, A., Hegde, A.: Computer vision application in object detection and tracking for aerial

surveillance. Indian Journal of Science and Technology **16**(31), 2374–2379 (2023)

[53] Yan, Y.: Using the improved ssd algorithm to motion target detection and tracking. Computational Intelligence and Neuroscience **2022**(1), 1886964 (2022)

[54] Martins, F., Baras, K., Marques, E.: Enhancing privacy: Using blurred images with mobilenet ssd. In: EPIA Conference on Artificial Intelligence, pp. 388–399 (2025). Springer

[55] Kiran, D.G., Kumar, T.R.: Spatial pyramid pooling (spp-net) compared with convolutional neural networks to recognise the object from given images with improved accuracy. In: AIP Conference Proceedings, vol. 3270, p. 020058 (2025). AIP Publishing LLC

[56] Wu, D., Huang, Z., Zhang, Y.: Solidtrack: A novel method for robust and reliable multi-pedestrian tracking. Electronics **14**(7), 1370 (2025)

[57] Zhang, X., Zhang, Y., Hu, M., Ju, X.: Insulator defect detection based on yolo and spp-net. In: 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), pp. 403–407 (2020). IEEE

[58] Sayed, M., Brostow, G.: Improved handling of motion blur in online object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1706–1716 (2021)

[59] Darnilasari, A., Areni, I.S., *et al.*: Implementation of faster r-cnn with colour and blur augmentation for differentiate cloves from debris. In: 2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE), pp. 72–77 (2023). IEEE

[60] Wen, L., Ding, J., Loffeld, O.: Video sar moving target detection using dual faster r-cnn. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **14**, 2984–2994 (2021)

[61] Xu, L., Luo, H., Hui, B., Chang, Z.: Real-time robust tracking for motion blur and fast motion via correlation filters. Sensors **16**(9), 1443 (2016)

[62] Yang, K., Geng, F.: Application of faster r-cnn model on human running pattern recognition. arXiv preprint arXiv:1811.05147 (2018)

[63] Olorunshola, O., Jemitola, P., Ademuwagun, A.: Comparative study of some deep learning object detection algorithms: R-cnn, fast r-cnn, faster r-cnn, ssd, and yolo. Nile J. Eng. Appl. Sci **1**(1), 70–80 (2023)

[64] Zheng, S., Wu, Y., Jiang, S., Lu, C., Gupta, G.: Deblur-yolo: Real-time object detection with efficient blind motion deblurring. In: 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2021). IEEE

[65] Okur, O., Kilicarslan, M.: Two-strea [m] yolov8: Object and motion detection in driving videos. IEEE Transactions on Intelligent Vehicles (2024)