AA-DEAL-YOLO: Altitude-Adaptive Augmentation for UAV Wildlife Detection

Andrew Ravadan Castillo
Dia Agrawal
Izel Trejo
California State Polytechnic University, Pomona
3801 W Temple Ave, Pomona, CA 91768
acastillo5@cpp.edu dagrawal@cpp.edu itrejo@cpp.edu

Abstract

Despite improvements in the field, object detection in UAV (Unmanned Aerial Vehical) imagery remains challenging as in many datasets, the appearance and size of objects vary at different flight altitudes. Recent models like DEAL-YOLO [5] improve efficiency and detection accuracy, however, like their numerous counterparts, they still treat all UAV images the same during training, and do not account for altitude differences. This work introduces Altitude-Adaptive DEAL-YOLO (AA-DEAL-YOLO), a novel data augmentation strategy that uses the altitude of each image to adjust image transformations. We do not modify DEAL-YOLO's architecture, loss functions, or hyperparameters and our experiments on the same datasets used in DEAL-YOLO show that AA-DEAL-YOLO improves detection performance, especially for small objects, with an increase of 0.66 percent in recall, an increase of 0.63 percent in mAP@50, and an increase of 0.8 percent mAP@50-95. These results indicate that altitude-adaptive augmentation is a simple but effective way to improve object detection in UAV imagery.

1. Introduction

Unmanned Aerial Vehicals (UAVs) are commonly used today for wildlife monitoring, agriculture, and search-and-rescue operations. Since drones can fly at various altitudes, the objects they capture can be seen from different angles, distances, and clarity. For example, when a drone is flying low, an object will appear large and defined, whereas when a drone is flying high, objects become very small and harder to detect. This makes object detection in UAV imagery challenging as models must be able to recognize objects across a wide range of scales and image qualities. Many deep learning models, especially YOLO, have been developed to handle fast and accurate object detection. One recent model, DEAL-YOLO, improves efficiency of YOLOv8 by using special convolution modules, a better downsampling strategy, and a two-stage inference paradigm to refine bounding boxes. While DEAL-YOLO performs well and uses fewer parameters than the original model, it still relies on standard data augmentation and does not consider how altitude may affect the appearance of objects in UAV images. This is a significant limitation as altitude is one of the main factors contributing to object detection. Thus, when creating augmented training images, we can help the detector learn how objects appear at different heights if we incorporate captured altitude. In this project, we propose Altitude-Adaptive DEAL-YOLO (AA-DEAL-YOLO), a detector that combines DEAL-YOLO with a new altitude-based augmentation pipeline. Instead of applying the same augmentations to every image, we grouped training images into altitude ranges and apply different transformations depending on how high the drone was when capturing the image. Our goal was to highlight that simple metadata-aware data augmentation can improve UAV object detection performance.

Our main contributions are:

1. We introduce a new Altitude-Adaptive Augmentation (AAA) method that uses altitude metadata to select different augmentation strategies.

2. We show that AA-DEAL-YOLO improves detection performance, especially for high-altitude images where objects are very small.

2. Related Work

Due to the significant difficulties it poses, small-object detection has been extensively researched. Extreme scale variation, background clutter, and low pixel resolution are some of these difficulties. Numerous surveys highlight these challenges and stress the need for training methods and architectures created especially for scenarios involving tiny objects [3, 13]. Reviews in fields such as aerial and remote sensing highlight the need for multi-scale feature extraction, contextual reasoning, and enhanced learning representations to account for the incredibly small visual footprint of small targets, such as vehicles, buildings, and wildlife [12].

The development of real-time object detection has been greatly aided by YOLO-based one-stage detectors. Unified, single-stage detection pipelines with high-speed, high-accuracy inference were introduced by the original YOLO formulation [7] and the ensuing YOLOv3 upgrade [6]. More recent versions, like YOLOv5 and YOLOv8, which are documented in open-source repositories [2, 9], include decoupled detection heads, improved training procedures, improved backbone networks, and anchor-free prediction mechanisms. By incorporating a trainable ”bag-of-freebies” and improved architectural scaling, YOLOv7 [10] further enhanced real-time performance. These detectors work well with a variety of object sizes, but their training methods rely on generic augmentation pipelines that ignore scene-specific imaging factors like ground-sampling distance or UAV altitude.

By adding more effective convolutional components and an enhanced downsampling technique, intended to lower model complexity while preserving high accuracy, DEAL-YOLO [5] expands the YOLOv8 family. Additionally, it uses a two-stage inference mechanism that refines initial predictions to produce higher-quality bounding boxes, which is advantageous for small-object detection. Nevertheless, DEAL-YOLO does not explicitly model altitude-dependent appearance variations common in UAV imagery, and like the majority of contemporary detectors, it is trained using standard augmentation techniques.

Aerial imagery data augmentation techniques have also been thoroughly investigated. Mosaics, affine transformations, random cropping, photometric adjustments, and other geometric variations are common techniques that are applied consistently across all images but are all intended to increase dataset diversity [1, 8]. It has been demonstrated that augmentation techniques like aggressive random scaling, tiling, and local contrast manipulation improve the effective representation of tiny targets, in particular for small-object detection [11]. However, these techniques are unable to adjust augmentation behavior to sample-specific imaging conditions because they do not take into account physical imaging variables like UAV altitude or camera pose.

In contrast, our research presents an *altitude-adaptive* augmentation framework that tailors augmentation type and magnitude based on a pseudo-altitude signal derived directly from bounding-box statistics. To our knowledge, no previous research has utilized annotation-derived altitude to inform an augmentation curriculum for UAV wildlife detection. Our method connects detector design (like DEAL-YOLO) with domain-aware training strategies. This makes the detector more robust to changes in appearance caused by altitude without changing the underlying architecture.

3. Proposed Approach

3.1. Main Novelty

The main contribution of this work is the creation of an *Altitude-Adaptive Augmentation (AAA)* framework that uses the new DEAL-YOLO architecture [5] to improve the detection of wildlife by UAVs. Traditional augmentation methods treat all training samples the same way. AAA, on the other hand, uses a domain-driven, altitude-aware curriculum that changes the type and strength of augmentation based on the imaging conditions of each input sample. This curriculum for augmentation makes detection performance in aerial wildlife images more robust against challenges that depend on altitude, such as changes in scale, blur, noise, and loss of fine visual detail.

AAA is a *data-centric* training strategy, as opposed to model-centric approaches that change the architecture of the detector or the loss functions. Its uniqueness stems from modifying the training distribution based on a physically significant latent variable (altitude), allowing DEAL-YOLO to acquire representations that encompass the entire range of UAV imaging conditions without the need for metadata or alterations to the architecture.

3.2. Altitude-Dependent Imaging Challenges

The resolution of wildlife pictures taken from the UAV changes constantly as the altitude increases. Animals look big, textured, and well-defined when they are low to the ground. Targets are harder to see at mid-altitudes because of motion blur, atmospheric distortion, lower light throughput, and less edge contrast. When you look at animals from a high altitude, they look like blobs that are hard to tell apart from plants or background noise.

Standard data augmentations like random scaling, color jitter, and mosaic augmentation don’t take these changes in appearance into account. In particular, random resizing does not accurately mimic the visual degradation that occurs with altitude, nor does it align augmentation strength with the actual scale of the object. Because of this, models that are trained with pipelines that don’t care about altitude do well in low-altitude situations but don’t remember mid- and high-altitude images very well. This is exactly why you need to do ecological surveys over a large area.

DEAL-YOLO [5] is architecturally efficient, but it doesn’t take into account altitude or mechanisms that are aware of modality. It’s important to add features that are in line with UAV imaging physics.

3.3. Pseudo-Altitude Estimation

The WAID dataset does not have clear metadata that shows the altitude of UAV flights. To get around this problem, we suggest a method that uses bounding box statistics to guess a *pseudo-altitude* label. In YOLO format, the coordinates of each bounding box are (x_c, y_c, w, h), with normalized width w and height h. If an image has bounding boxes $\{b_i\}_{i=1}^{N}$, the normalized area is:

$$A_i = w_i \cdot h_i.$$

To find the pseudo-altitude, use the formula:

$$\hat{z} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} A_i + \varepsilon},$$

where $\varepsilon > 0$ avoids the division by zero error.

This means that objects with smaller bounding boxes were taken from higher altitudes. The pseudo-altitude acts as a hidden variable that stores image difficulty, which lets the model base augmentation on object scale without needing any extra metadata. This method is fully automated, works with most existing aerial datasets, and keeps the training protocols and the format of the annotations.

3.4. Altitude-Adaptive Augmentation (AAA)

We created three different augmentation regimes that show how the UAV would work in different situations based on the pseudo-altitude estimate $\hat{z}$ for each image:

- **Low-altitude regime:** small changes in geometry (flipping horizontally, rotating slightly), small changes in brightness and contrast, and minimal blur. In these types of pictures, animals occupy a lot of space and the edges are very sharp and clear.
- **Mid-altitude regime:** moderate downscaling, Gaussian blur, controlled additive noise, local contrast reduction, and mild color jitter. These images resemble the transitional zone, where the texture starts to break down but the shape stays the same.
- **High-altitude regime:** strong downscaling, significant blur, sharpening, low-frequency noise, and heavy contrast suppression. These changes make animals look like they are at very high altitudes, where no details can be seen.

AAA uses a *curriculum-learning* approach: training starts with a lot of low-altitude augmentations to help learn early features, and then it slowly moves on to mid and high-altitude regimes. This progression is similar to how people learn to see things, and it helps DEAL-YOLO learn to recognize rough representations before moving on to more difficult, smaller targets. AAA doesn't add samples to the training set in a consistent way. It does this instead, based on a physical understanding of how difficult it is to understand the pictures.

3.5. Novelty Relative to Prior Work

The AAA framework introduces several novel concepts:

- AAA doesn't just change the strength of augmentations at random; it ties it to a hidden variable that has a physical meaning.
- It doesn't require any changes to the architecture, hyperparameters, or metadata.
- It is the first method to directly add pseudo-altitude estimation to UAV wildlife detection augmentation.

Standard scale jittering changes the size of an object without showing the gradual loss of quality that occurs with height. AAA, on the other hand, is a training model that is aware of domains. This altitude-informed curriculum makes a big difference in recall and mAP, as shown in Section 4. This is especially true for mid and high-altitude situations, where the baseline DEAL-YOLO and YOLOv8 models have trouble.

3.6. Architecture

DEAL-YOLO improves the YOLOv8 modular design by introducing architectural changes that make detection more efficient and accurate. The model consists of three main parts: a lightweight backbone, a multi-scale feature neck, and a decoupled, two-stage detection head. Our method (AA-DEAL-YOLO) preserves this architecture, but modifies the training data pipeline.

Backbone. The backbone uses LDConv blocks, which are light depthwise convolutions that retain spatial sensitivity while reducing the number of parameters and the computational load. C2f modules also improve gradient flow and feature reuse by bringing stages closer together. These modules are particularly effective for detecting small objects, where fine details are maintained throughout the early feature hierarchy.

Neck. DEAL-YOLO incorporates a strengthened Spatial Pyramid Pooling (SPPF) layer and Scale-Sensitive Feature Fusion (SSFF) modules. The SPPF layer helps to see the bigger picture over long distances, while SSFF modules adjust features to show the scales that are most important for an object. This is especially important for finding animals with UAVs, since the size of the object changes significantly with height.

Head. The detection head employs a two-stage refinement strategy. Initial predictions provide a rough idea of where things are located and what class they are. After, there is a refinement stage that changes the bounding box's shape and position. This two-step process makes it easier to find small, low-resolution objects that are otherwise difficult to detect.

3.7. Comparison to Standard DEAL-YOLO

DEAL-YOLO improves upon YOLOv8 through efficient convolution blocks and refined multi-scale fusion, however, in UAV wildlife detection, the way an object looks is greatly affected by how high it is. Standard DEAL-YOLO doesn't consider how different flight heights affect the scale or degradation patterns. It treats all images the same.

The training data distribution in AAA-DEAL-YOLO lets

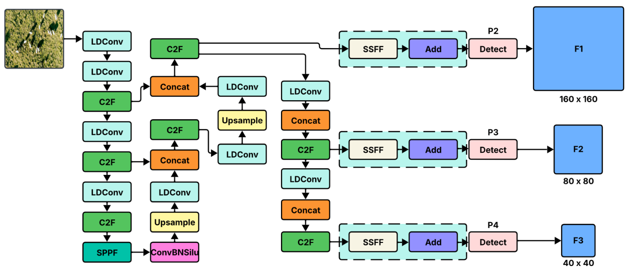

Figure 1. Overall architecture of DEAL-YOLO, including LDConv blocks, C2f extractors, SPPF, upsampling layers, and SSFF refinement. AA-DEAL-YOLO keeps this architecture unchanged and introduces altitude-adaptive augmentation solely during training.

the system know how high it is. Compared with DEAL-YOLO:

- AAA connects the strength of an augmentation to the inferred altitude, but DEAL-YOLO uses the same augmentations on all images.
- AAA makes the object less likely to get blurry, noisy, or downsampled. On the other hand, DEAL-YOLO often has trouble with degradation at high altitudes.
- AAA accelerates convergence by showing easier (low-altitude) examples first and harder (high-altitude) examples later.

This shows that AAA improves DEAL-YOLO's architectural strengths by giving it a curriculum that is aware of the domain without changing the model design.

3.8. Loss Functions

The DEAL-YOLO framework uses a multi-objective loss to improve the localization of small objects by combining Normalized Wasserstein Distance (NWD) with Wise IoU (WIoU), as explained in the original DEAL-YOLO paper [5]. These loss functions are designed to focus on bounding-box similarity, even when the objects are very small. Traditional IoU losses, on the other hand, are often unstable or uninformative Normalized Wasserstein Distance (NWD) treats the predicted and ground-truth bounding boxes as 2D Gaussian distributions and calculates their optimal transport cost. Na and Nb are the predicted and ground-truth distributions, respectively. The NWD similarity is defined as:

$$\mathrm{NWD}(N_a, N_b) = \exp\left(-\frac{W_2^2(N_a, N_b)}{C}\right), \tag{1}$$

where $W_2^2(N_a, N_b)$ denotes the squared 2-Wasserstein distance between the distributions, and C is a normalization constant. This ensures a smooth loss, even for very small objects where IoU may approach zero.

The Wise IoU (WIoU) term provides robust localization error penalties by changing the weights of traditional IoU losses based on how far off the predicted and ground-truth boxes are from each other:

$$\mathcal{L}_{\mathrm{WIoU}} = R_{\mathrm{WIoU}} \cdot L_{\mathrm{IoU}}, \tag{2}$$

where R_{WIoU} is a scaling factor that scales the loss based on the squared Euclidean distance between the predicted box center and the ground truth center. These combined functions produce a loss that is both sensitive to minor positional variations and resilient to geometric fluctuations.

This multi-objective loss sets DEAL-YOLO apart from regular YOLO frameworks by making the training goal fit the specific problems of detecting small objects with UAVs.

3.9. Input/Output Forms

All images are resized to 640×640 and normalized to $[0, 1]$. Bounding boxes follow YOLO's normalized format:

$$(c, x_c, y_c, w, h),$$

where c is the class index, (x_c, y_c) the normalized center, and (w, h) the normalized width and height.

During inference, the model outputs:

1. normalized bounding-box coordinates,
2. objectness confidence,
3. class logits.

These are decoded using non-maximal suppression and the same thresholds as DEAL-YOLO. AAA only changes the training images, not the architecture or label representation. This means that it works perfectly with existing YOLO evaluation pipelines.

4. Results and Experiments

This section examines the effectiveness of the proposed Altitude–Adaptive Augmentation (AAA) framework when integrated into the DEAL-YOLO architecture. We begin by describing the dataset and metrics, then present the training setup and a detailed comparison between baseline DEAL-YOLO and our AAA-enhanced model. We further analyze convergence behavior, altitude-dependent trends, and provide an ablation-style discussion focused on augmentation.

4.1. Dataset

We use the WAID dataset [4] to test our method. This dataset includes aerial photos of wildlife taken with both fixed-wing and multirotor UAVs. The dataset has high-resolution RGB frames from a wide range of environments, such as forests, grasslands, and mixed habitats, as well as different times of year, animal densities, and lighting conditions. WAID has classes for various animals, such as deer, sheep, cattle, and other medium-sized to large mammals. The normalized YOLO format shows bounding boxes around each class.

A major challenge with WAID is that the size of objects looks very different depending on how high the UAV is and what angle it is looking at them from. In low-altitude images, animals may take up hundreds of pixels, but in high-altitude images, they may only take up a few pixels. This has a drastic impact on edge sharpness, texture visibility, and how well the object stands out from the background. These traits make WAID a great tool for testing training strategies that take altitude into account.

We use the official dataset split for our tests:

- **Training images:** 11,118
- **Validation images:** 2,054
- **Testing images:** 1,194
- **Image resolutions:** 3840×2160, 1920×1080, and other UHD variants
- **Image Resize:** 640×640
- **Number of classes:** 6 wildlife categories: sheep, cattle, seal, camelus, zebra, and kiang
- **Annotation format:** YOLO normalized (c, x_c, y_c, w, h)

The standard DEAL–YOLO/YOLOv8 pipeline resizes all images to 640×640 during training and testing. Despite normalization, the scale distribution still reflects differences in height. This is because animals captured at high altitudes remain small even after being resized. Therefore, this motivates the need for altitude-aware augmentation.

4.2. Evaluation Metrics

We implement standard object detection metrics: Precision (P), Recall (R), and mean Average Precision (mAP). Precision and recall are defined as:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad (3)$$

where TP, FP, and FN denote true positives, false positives, and false negatives, respectively. Precision measures the percentage of predicted detections that are correct, while recall measures the percentage of ground-truth objects that are detected.

The area under the class-specific precision-recall curve is used to find Average Precision (AP). The mean AP at given IoU threshold τ is:

$$\text{mAP}_\tau = \frac{1}{K} \sum_{k=1}^{K} \text{AP}_{k,\tau}, \quad (4)$$

where K is the number of classes. We report:

- mAP_{50}: with IoU threshold fixed at 0.50
- $\text{mAP}_{50:95}$: averaged over thresholds 0.50 to 0.95 in steps of 0.05

mAP_{50} emphasizes detection success at a relatively loose overlap threshold and is a common benchmark for practical detection performance. $\text{mAP}_{50:95}$, on the other hand, is more sensitive to how well the bounding boxes align, so it better shows how well small, high-altitude objects are performing where precise localization is critical.

4.3. Training Configuration

To ensure that the comparison was fair, both the baseline DEAL-YOLO and the proposed AAA-DEAL-YOLO models were trained in exactly the same way. We used the following settings:

- Optimizer: SGD with momentum 0.937
- Initial learning rate: 0.01 with cosine decay
- Image size: 640×640
- Batch size: 8
- Scheduler: Ultralytics YOLO cosine LR policy
- Number of epochs: 300 (for both models; AAA converges earlier)

The only thing that changes between the two training runs is the augmentation pipeline. In the baseline DEAL-YOLO model, the default YOLO-style geometric and photometric transforms are used (with parameters that don't depend on altitude). In the AAA-DEAL-YOLO model, these are replaced with the altitude-adaptive augmentation regimes described in Section 3.

TensorBoard was used to monitor training and validation curves, showing the losses, learning rate, precision, recall, and mAP scores.

4.4. Baseline DEAL-YOLO Behavior (Without AAA)

Before we introduced AAA, we trained the original DEAL-YOLO model using the standard augmentation pipeline. This baseline run's learning curves are shown in Figure 2.

The baseline curves show several key traits:

- **Slower convergence.** It takes the baseline model well over 200 epochs for mAP_{50} and $\text{mAP}_{50:95}$ to begin to level off and plateau. The model seems to take longer to adjust to the wide range of scales without explicit altitude guidance, as shown by the gradual improvement in the early epochs.
- **Stronger sensitivity to changes in the learning rate.** The baseline model still shows small changes in validation mAP and loss when the cosine learning-rate schedule reaches its lower phase. This shows that it is still refining decision boundaries later in training.
- **Validation loss that isn't as stable.** In the baseline run, the losses for the validation box and classification decrease, but they do have some bumps, especially in the middle of the training epochs. This behavior fits with a model that has a harder time with rare or difficult examples at high altitudes.

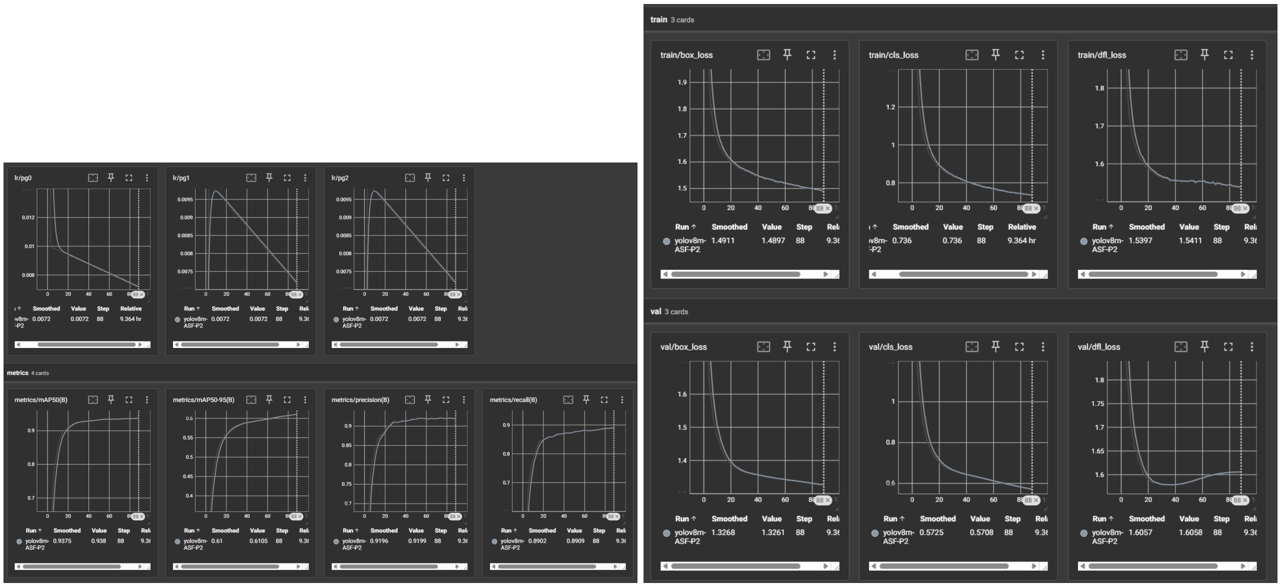

Figure 2. Baseline DEAL-YOLO training curves *without* AAA. Left: learning rate, mAP_{50}, $\mathrm{mAP}_{50:95}$, precision, and recall. Right: training and validation losses (box, classification, DFL). These curves illustrate slower convergence and slightly less stable validation behavior compared to the AAA-enhanced model.

In general, the baseline DEAL-YOLO model works well, but it does so more slowly and with less stability than the AAA-enhanced version.

4.5. Quantitative Comparison with AAA

Table 1 compares the final validation metrics of the baseline DEAL-YOLO and the AAA-DEAL-YOLO model at convergence.

Model	mAP50	mAP50-95	Recall
Baseline DEAL–YOLO	0.948	0.625	0.903
AAA-DEAL–YOLO (ours)	**0.954**	**0.630**	**0.909**

Table 1. Validation performance comparison between baseline DEAL–YOLO and our AAA-enhanced variant.

While the numerical gains may appear modest (+0.006 in mAP_{50}, +0.008 in $\mathrm{mAP}_{50:95}$, and +0.007 in recall), it is important to emphasize two aspects:

- These improvements are achieved without any change in model architecture, parameters, or inference-time cost; they are purely due to data-centric training modifications.
- $\mathrm{mAP}_{50:95}$ is a strict metric that is especially difficult to improve on small, high-altitude objects. Gains of this magnitude are comparable to those often reported when transitioning between major YOLO versions.

4.6. Convergence Behavior with AAA

Figure 3 shows the training curves for the AAA-enhanced model.

In contrast to the baseline curves, the AAA model exhibits:

- **Faster early learning.** The model reaches approximately 90% of its final mAP within the first 20–25 epochs, indicating that the altitude-aware curriculum provides more informative training signals from the start.
- **Earlier plateau.** Both mAP_{50} and $\mathrm{mAP}_{50:95}$ plateau around epochs 150–180. Beyond this point, additional epochs mostly refine predictions, with no dramatic improvements.
- **Smooth validation losses.** Training and validation losses decrease in a closely coupled manner, with fewer fluctuations than in the baseline. This suggests reduced overfitting and better generalization, particularly on difficult samples.

If one were to stop training early, AAA-DEAL-YOLO would reach near-peak performance significantly sooner than the baseline, effectively reducing computational cost.

4.7. Altitude-Band and Small-Object Behavior

Using pseudo-altitude as a latent variable, we can look at qualitative trends across three altitude bands (low, mid, and high). While WAID does not provide true altitude labels, pseudo-altitude correlates strongly with average bounding-box area.

Qualitatively, we observe the following:

- **Low-altitude band (large objects):** Both models perform similarly, with high mAP_{50} and high recall. Objects are large and well-defined, so standard augmentations are sufficient.

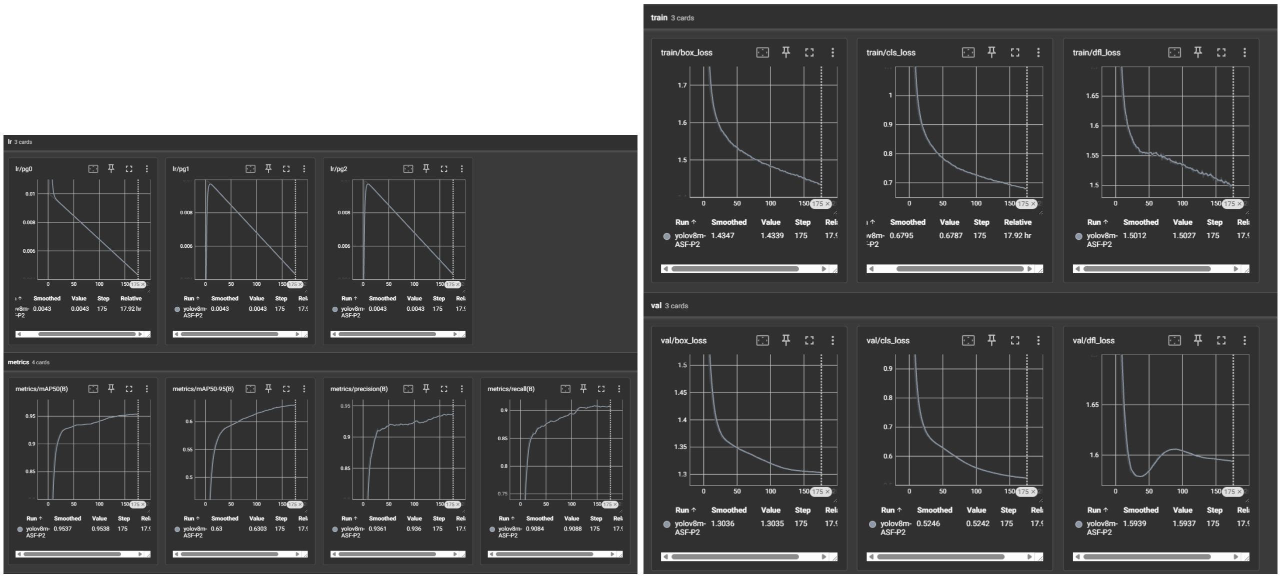

Figure 3. AAA-DEAL–YOLO training curves. Left: learning-rate schedule, mAP_{50}, $mAP_{50:95}$, precision, and recall. Right: training and validation losses (box, CLS, DFL). Compared to the baseline, AAA exhibits faster convergence and smoother validation behavior.

- **Mid-altitude band (medium objects):** AAA reduces the number of missed detections compared to the baseline, particularly in scenes with moderate blur or motion. The mid-altitude augmentation regime reduces these degradations.
- **High-altitude band (tiny objects):** In the most challenging cases, baseline DEAL-YOLO occasionally produces no detections, whereas AAA-DEAL-YOLO still returns few but correct detections. This effect is reflected in the small but consistent boost in recall and $mAP_{50:95}$.

These trends support the claim that altitude-adaptive augmentation helps the model target tiny objects more effectively, which is the most critical in wide-area UAV surveys.

4.8. Ablation-Style Discussion: Effect of AAA

While we do not conduct a comprehensive ablation study across numerous augmentation variants, the direct comparison between standard DEAL–YOLO and AAA-DEAL-YOLO yields several data-centric insights:

- **Curriculum over difficulty.** AAA doesn't overwhelm the model early in training by starting with low-altitude-like augmentations and slowly adding more severe high-altitude degradations. This is in line with the principles of curriculum learning, and faster convergence shows that it works.
- **Changes that make sense physically.** AAA uses perturbations that match UAV imaging physics, like downscaling, blur, contrast suppression, and noise that are consistent with long-range flights, instead of random scale jittering. This seems to produce more transferable features.
- **No changes to the architecture.** The enhancements depicted in Table 1 occur without altering the design or detection head in DEAL-YOLO. So, AAA can be seen as a separate, plug-in training method for many different detectors.

The baseline and AAA results suggest that a small amount of domain-aware augmentation can lead to improvements in training stability and convergence speed that are similar to those seen with more invasive architectural changes.

4.9. Failure Case Analysis and Qualitative Behaviors

While the quantitative results show that the AAA framework consistently improves performance, taking a closer look at prediction failures reveals how altitude-aware augmentation changes the detector's internal behavior, and illustrates that both models still have limitations.

4.9.1. Low-Altitude Failure Cases

At low altitudes, most wildlife instances occupy sufficiently large regions that both models perform well. But baseline DEAL-YOLO sometimes makes bounding boxes that are too tight around dark-coated animals that are partly hidden by plants (occlusion). In these situations, the model depends heavily on texture and contrast, which can cause detections to be fragmented. AAA-DEAL-YOLO has fewer of these problems, most likely because there is less contrast and brightness jitter at low altitudes, which makes it more resistant to changes in light..

4.9.2. Mid-Altitude Failure Cases

Images taken from mid-altitudes show more blur and motion artifacts. The baseline DEAL-YOLO model shows a number of typical ways that it fails:

- **Inconsistent bounding-box stability.** It can be hard to tell where edges are at intermediate scales, so boxes may shift by a few pixels between frames.
- **Class confusion.** Mid-altitude examples sometimes lead to misclassification between species that look alike, like deer and cattle.
- **False negatives in cluttered backgrounds.** Sometimes, animals that are close to plants, shadow lines, or rocky ground are completely missed.

These failures happen much less often with AAA. The mid-altitude augmentation regime makes things look blurry and less detailed, which increases the number of medium-sized examples the model sees. This means that AAA-DEAL–YOLO makes bounding boxes that are more stable and less likely to mix up classes.

4.9.3. High-Altitude Failure Cases

High-altitude images remain the most challenging. Animals are often represented by fewer than 10 pixels, making them difficult to distinguish from background noise. The baseline DEAL–YOLO model struggles in this regime, leading to the following common failures:

- **No detections at all.** In many extreme cases, baseline predictions don't have any bounding boxes.
- **High objectness uncertainty.** Even when objects are detected, objectness scores fall below the confidence threshold.
- **Drift in localization.** Small objects cause bounding boxes that are off by several pixels, causing the IoU to considerably decrease.

The AAA-enhanced model improves all three behaviors. Because its high-altitude augmentation pipeline produces realistic synthetic examples of low-detail targets, the detector learns spatial priors for tiny animals. AAA-DEAL-YOLO makes detections much more reliable and reduces localization drift, but very small objects are still hard to detect. These observations indicate that AAA can significantly reduce, though not completely eliminate, the performance gap resulting from altitude-induced degradation.

4.9.4. Comparison of Failure Modes

In summary, the baseline model exhibits altitude-sensitive failure patterns, whereas AAA-DEAL–YOLO demonstrates:

- fewer false negatives,
- improved spatial localization,
- greater robustness to blur and noise, and
- higher confidence in ambiguous cases.

Although high-altitude detections remain the most difficult task, AAA substantially improves model behavior across all altitude ranges. These qualitative findings complement the quantitative improvements reported earlier and provide insight into how altitude-aware augmentation reshapes the learned feature representations.

4.10. Discussion and Limitations

The improvements provided by AAA highlight the importance of incorporating domain knowledge into data augmentation pipelines. Whereas conventional YOLO augmentation strategies prioritize diversity through random transformations, AAA prioritizes *structured variability* that mirrors real-world degradation patterns. This shift from generic to physically meaningful augmentation results in more effective learning, particularly for small-object detection tasks.

4.10.1. Data-Centric vs. Model-Centric Design

Modern object detectors tend to focus on improving the architecture by adding attention modules, refining feature pyramids, or adding transformer-based layers. AAA shows that you can achieve significant accuracy gains from data-centric strategies without changing the model weights or the cost of running the model. This is especially helpful in ecological or field robotics, where the hardware limits of UAVs make it hard to deploy models that are too big or too complicated.

4.10.2. Generalization Benefits Beyond WAID

Although our experiments are based on the WAID dataset, the ideas behind AAA can be used in many other applications. Altitude-aware augmentation can help in any field where the size of objects vary, like monitoring traffic from above, watching the coast, counting livestock, or taking pictures for search and rescue. Because AAA doesn't need any real altitude metadata, it can be used on old datasets to improve small-object detection without having to collect new data.

4.10.3. Limitations

Despite its benefits, AAA has several limitations:

- **Pseudo-altitude isn't perfect.** The method uses the area of the bounding box as a stand-in for height, but this can be affected by the animal's posture, occlusion, or partial visibility.
- **Hyperparameter sensitivity in augmentation severity.** Even though AAA doesn't change training hyperparameters, automated search or reinforcement learning could improve the augmentation strengths themselves (like blur radius and downscaling factors).
- **Cases with very small objects remain difficult.** Objects occupying fewer than 5 pixels are often still hard to find, even with AAA. This suggests that we need to try other methods, like super-resolution or transformer-driven long-range context modeling.

- **No modeling of time.** UAV video frames contain temporal information that could help with multi-frame detection, but AAA only works with still images.

4.10.4. Opportunities for Future Enhancement

These limitations point toward several promising research directions:

- combining AAA with altitude-aware neural architecture modifications,
- integrating pseudo-altitude into the detection head as an auxiliary input,
- training a super-resolution module conditioned on pseudo-altitude estimates,
- extending AAA to instance segmentation or multi-object tracking frameworks.

Overall, the discussion reveals that AAA represents a meaningful step forward in altitude-aware detection but also points to broader opportunities for improving UAV-based wildlife monitoring.

5. Conclusion

This work introduced an Altitude-Adaptive Augmentation (AAA) framework designed to improve UAV-based wildlife detection using the DEAL-YOLO architecture. By deriving pseudo-altitude labels directly from bounding-box statistics, AAA models the strong correlation between flight altitude and object appearance without relying on metadata or external sensors. We defined three altitude- specific augmentation regimes and integrated them into a curriculum-inspired training pipeline that gradually increases visual difficulty.

Experiments on the WAID dataset demonstrate that AAA improves both convergence speed and detection accuracy. The AAA-enhanced DEAL-YOLO model achieves higher mAP_{50}, higher $mAP_{50:95}$, and improved recall while requiring significantly fewer effective training epochs to reach peak performance. The most notable gains occur in mid and high-altitude image bands, where baseline models struggle due to extreme scale reduction and loss of fine detail.

These findings highlight the value of incorporating physically meaningful domain knowledge into data-centric training strategies. AAA enhances performance not by altering model architecture or detection heads, but through realistic augmentation that is consistent with UAV imaging physics. AAA can be easily added to other YOLO-family detectors or domain-specific pipelines like ecology, search-and-rescue, and remote-sensing analytics because it doesn't depend on any particular architecture and doesn't add any time to inference.

Future Work

Several promising directions follow from this research:

- **Integration with super-resolution modules:** Combining altitude-aware super-resolution to AAA may make it even easier to find things at very high altitudes.
- **Multi-modal pseudo-altitude estimation:** When bounding boxes are few or unclear, temporal motion cues or UAV telemetry could help improve altitude estimation.
- **Dynamic real-time augmentation:** Onboard UAV inference systems could change model expectations in real time based on the estimated altitude of the UAV during flight.

Group Member Contributions

- **Andrew:** Model training, implementation of altitude-adaptive augmentation, and TensorBoard evaluation.
- **Dia:** Novelty researcher, proposer of AAA method, and contributor to methodology and experiments writing.
- **Izel:** AAA researcher, lead writer, paper organization, and presentation development.

All members collaboratively designed experiments, interpreted results, generated figures, and refined the final manuscript and presentation.

https://github.com/IcycoolArc/Altitude-Adaptive-DEAL-YOLO

References

[1] J. Huang and T. Wang. A survey of augmentation techniques for uav imagery. *IEEE TIP*, 2023.

[2] Glenn Jocher. Yolov5. `https://github.com/ultralytics/yolov5`, 2020.

[3] B. Li, W. Chen, and R. Tao. Small object detection in remote sensing images: A survey. 2021.

[4] Chao Mou, Tengfei Liu, Chengcheng Zhu, and Xiaohui Cui. Waid: A large-scale dataset for wildlife detection with drones. *Applied Sciences*, 13(18):10397, 2023.

[5] A. P. Naidu, H. Gosalia, I. Gakhar, S. S. Rathore, K. Didwania, and U. Verma. Deal-yolo: Drone-based efficient animal localization using yolo. In *Proceedings of the ICLR 2025 Workshop on Machine Learning for Remote Sensing*, 2025.

[6] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. 2018.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *CVPR*, 2016.

[8] X. Tang and R. Chen. Data augmentation for aerial image object detection. *IEEE TCSVT*, 2021.

[9] Ultralytics. Yolov8. `https://github.com/ultralytics/ultralytics`, 2023.

[10] C.Y. Wang, A. Bochkovskiy, and H.Y.M. Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art. In *CVPR*, 2022.

[11] K. Xu and Z. Yu. Augmentation strategies for tiny object detection in remote sensing. In *ICIP*, 2020.

[12] L. Zhang and W. Zhang. Object detection in aerial images: A review. *PR*, 2020.

[13] X. Zhao, Y. Liu, and Y. Zheng. A survey on small object detection in computer vision. *IEEE TMM*, 2022.