Advanced International Journal for Research (AIJFR)

E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

# Understanding Image Resolution Sensitivity in Modern YOLO Architectures

**Ketan Kanjiya [1], Piyush Sonani [2], Upendrasinh Zala [3]**

[1] Chief Research Officer, Information Technology Department, Kshatrainfotech
[2] Chief Technology Officer, Information Technology Department, Kshatrainfotech
[3] Chief Executive Officer, Information Technology Department, Kshatrainfotech

## Abstract

Input image resolution plays a critical yet often underexplored role in the performance and efficiency of modern object detection systems. While YOLO architectures support flexible input sizes, models are typically trained at fixed resolutions, making resolution selection a key deployment decision. This paper presents a systematic investigation of image resolution sensitivity in YOLOv11 across multiple model scales. Using the Aerial Sheep dataset, five YOLOv11 variants (Nano to Extra Large) are fine-tuned at three training resolutions 320×320, 640×640, and 1280×1280 under identical training conditions. Detection performance is evaluated using mAP@50, mAP@50-95, precision, and recall, alongside a detailed analysis of inference latency. Results demonstrate that input resolution is a dominant factor influencing detection accuracy, often exceeding the impact of model scaling. Substantial performance gains are observed when increasing resolution from 320×320 to 640×640, while improvements beyond 640×640 show diminishing returns for coarse detection metrics. Inference analysis reveals that model size and training resolution primarily govern runtime, with inference time resolution and image content exerting secondary effects. These findings provide practical guidance for balancing accuracy and efficiency in real-world YOLO deployments.

**Keywords:** Yolo, Object Detection, Image Resolution, Deep Learning, Image Resolution Sensitivity

## 1. Introduction

Object detection is a foundational task in computer vision, enabling a wide range of applications including autonomous driving, medical image analysis, intelligent surveillance, retail analytics, and human computer interaction [1-4]. Recent advances in deep learning have significantly improved detection accuracy and efficiency, particularly through one-stage architectures such as YOLO and SSD, which offer real-time inference with competitive performance [5-7]. These gains are largely attributed to advances in convolutional neural networks, feature pyramid representations, and optimized training strategies.

Despite this progress, the influence of input image resolution on object detection performance remains insufficiently characterized. In real-world deployment scenarios, input images originate from heterogeneous sources such as smartphones, CCTV systems, embedded cameras, and aerial platforms, each producing images with different native resolutions, compression characteristics, and noise profiles [8,9]. To satisfy bandwidth, memory, or latency constraints, images are frequently resized prior to inference [10,11]. Although YOLO architectures permit variable input sizes, models are typically trained at a fixed resolution, requiring all incoming data to be rescaled regardless of its original fidelity.

Image resizing directly affects visual characteristics such as texture granularity, edge sharpness, and spatial detail, which are critical for convolutional feature extraction [12]. Higher resolutions may preserve fine-grained information but incur greater computational cost, while aggressive downsampling can eliminate small objects entirely or distort their spatial structure [13]. As a result, selecting an appropriate image resolution becomes a trade-off between detection accuracy, robustness, and inference efficiency. While prior studies have examined resolution effects alongside factors such as compression artifacts, camera distance, or image quality degradation [14], these confounding variables make it difficult to isolate the pure effect of spatial resolution alone.

Modern YOLO architectures further complicate this relationship through extensive use of multi-scale feature extraction mechanisms, including Feature Pyramid Networks (FPN) and Path Aggregation Networks (PAN). These structures are designed to capture objects across multiple spatial scales by fusing low-level and high-level features. However, their effectiveness depends implicitly on the resolution of the input image. Reduced resolutions may weaken fine-scale features critical for small object detection, while excessively high resolutions may introduce diminishing returns relative to the increased computational burden.

Motivated by these considerations, this paper presents a controlled experimental study of image resolution sensitivity in modern YOLO architectures. Identical models are fine-tuned on the same dataset resized to multiple spatial resolutions, with all other training parameters held constant. Detection performance is evaluated using precision, recall, mAP@50, mAP@50-95, and inference time to determine how accuracy and efficiency scale with resolution. By isolating resolution as the sole variable, this work provides a clear empirical understanding of how YOLO models respond to changes in input fidelity and identifies resolution ranges that offer an effective balance between performance and computational cost.

## 2. Dataset

The experiments in this study were conducted using the Aerial Sheep dataset [15], which consists of RGB images captured from an aerial viewpoint containing instances of sheep. The dataset presents realistic challenges for object detection, including large variations in object scale, sparse and dense object distributions, complex natural backgrounds, and changes in illumination and terrain. Such characteristics make the dataset particularly suitable for evaluating small object detection performance in aerial imagery.

The dataset is divided into 1,203 training images, 350 validation images, and 174 test images, with all sheep instances annotated using bounding boxes. To systematically analyze the impact of input spatial resolution on detection performance, the entire dataset was resized into three fixed resolution versions:

320×320, 640×640, and 1280×1280. Resizing was performed using bilinear interpolation while preserving the original bounding box annotations.

These multi-resolution variants enable a controlled investigation of how image resolution affects detection accuracy under realistic aerial imaging conditions. Figure 1 illustrates representative samples from the Aerial Sheep dataset across varying environmental conditions and object scales.

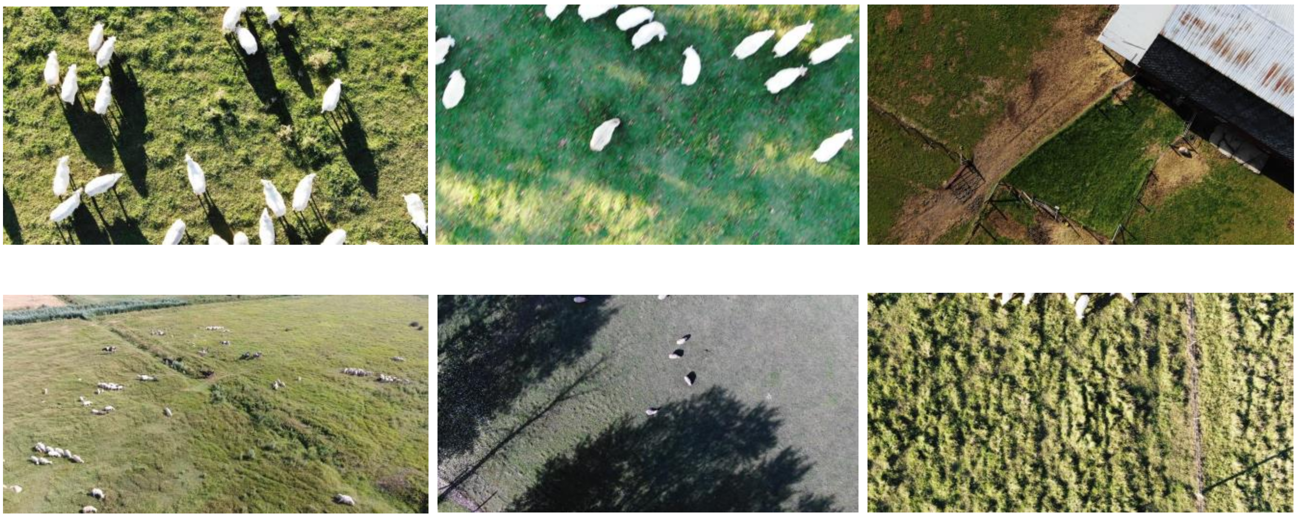


**Figure 1:** Sample images from the Aerial Sheep dataset

## 3. Methodology

This study investigates the impact of input image resolution and model scale on detection accuracy and inference efficiency using real-world aerial imagery. The Aerial Sheep dataset was resized to three fixed spatial resolutions: 320×320, 640×640, and 1280×1280, producing three resolution specific dataset variants while preserving original annotations.

Five YOLOv11 model variants Nano (N), Small (S), Medium (M), Large (L), and Extra Large (X) were independently fine-tuned on each resolution specific dataset, resulting in a total of 15 trained models. All models were initialized with publicly available COCO pretrained weights and trained using the Ultralytics framework under identical conditions, including learning rate schedules, data augmentation strategies, batch sizes, and optimization settings. Each model was trained for 100 epochs to ensure stable convergence and fair comparison across configurations.

Detection performance was evaluated on the corresponding test sets using standard object detection metrics, including mAP@50, mAP@50-95, precision, and recall. These metrics enable a systematic analysis of accuracy trends as a function of input resolution and model capacity.

To assess inference efficiency, 10 test images were randomly selected and resized to each of the three resolutions. All resized images were evaluated using the 15 trained models, allowing controlled comparison of inference latency across resolution and model combinations under identical input conditions.

Model training was performed on an NVIDIA GeForce RTX 4090 GPU with CUDA acceleration, while inference timing experiments were conducted on an Intel i5 CPU with 16 GB RAM. By maintaining consistent training and evaluation protocols, observed performance differences can be attributed primarily to variations in input resolution and YOLOv11 model scale.

## 4. Results and Discussions

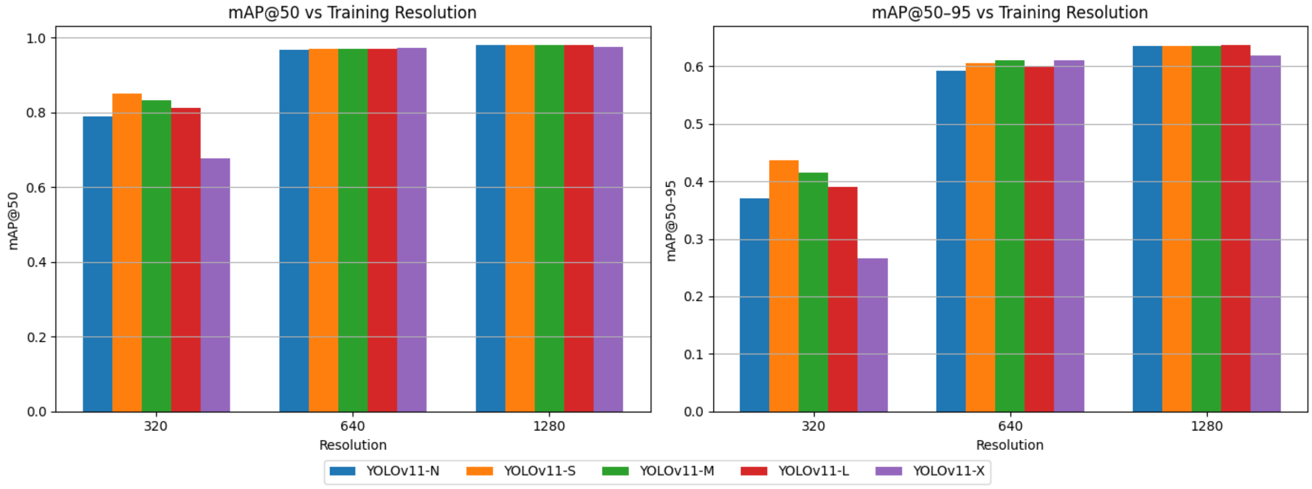
### 4.1 Detection Performance Across Resolutions and Model Variants

Table 1 summarizes the detection performance of YOLOv11 model variants trained at different input resolutions on the Aerial Sheep dataset. Across all variants, detection accuracy exhibits a strong and consistent dependence on training image resolution.

**Table 1:** Detection performance of YOLOv11 model variants trained at different input resolutions on the Aerial Sheep dataset.

| Model Variant | Resolution | mAP@50 | mAP@50-95 | Precision | Recall |
|---|---|---|---|---|---|
| YOLOv11-N | 320x320 | 0.790 | 0.370 | 0.823 | 0.724 |
| YOLOv11-N | 640x640 | 0.967 | 0.593 | 0.955 | 0.937 |
| YOLOv11-N | 1280x1280 | 0.979 | 0.635 | 0.974 | 0.963 |
| YOLOv11-S | 320x320 | 0.850 | 0.437 | 0.883 | 0.769 |
| YOLOv11-S | 640x640 | 0.970 | 0.605 | 0.948 | 0.940 |
| YOLOv11-S | 1280x1280 | 0.981 | 0.635 | 0.972 | 0.964 |
| YOLOv11-M | 320x320 | 0.833 | 0.415 | 0.862 | 0.755 |
| YOLOv11-M | 640x640 | 0.971 | 0.610 | 0.958 | 0.947 |
| YOLOv11-M | 1280x1280 | 0.979 | 0.636 | 0.971 | 0.964 |
| YOLOv11-L | 320x320 | 0.812 | 0.390 | 0.844 | 0.742 |
| YOLOv11-L | 640x640 | 0.969 | 0.601 | 0.948 | 0.942 |
| YOLOv11-L | 1280x1280 | 0.980 | 0.638 | 0.974 | 0.969 |
| YOLOv11-X | 320x320 | 0.676 | 0.266 | 0.759 | 0.614 |
| YOLOv11-X | 640x640 | 0.971 | 0.610 | 0.965 | 0.947 |
| YOLOv11-X | 1280x1280 | 0.975 | 0.620 | 0.970 | 0.956 |

To further illustrate these trends, Figures 2 and 3 present a graphical comparison of detection performance across training resolutions and model variants. These visualizations complement the numerical results in Table 1 by highlighting how accuracy scales with input resolution and model capacity.

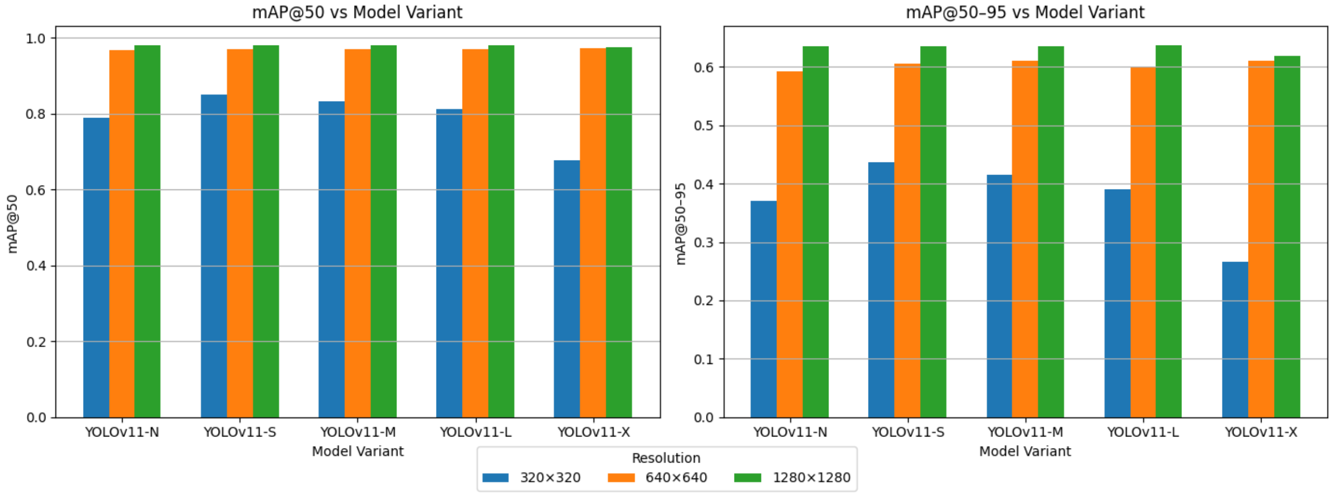


(a) (b)

**Figure 2:** Detection accuracy as a function of training resolution for different YOLOv11 model variants: (a) mAP@50 and (b) mAP@50-95.

### Strong Resolution Sensitivity Across All YOLOv11 Variants

Across all YOLOv11 variants, detection performance shows a strong and consistent dependence on training image resolution. Figures 2 and 3 provide complementary visual interpretations of the results in Table 1. Figure 2 illustrates the variation of mAP@50 and mAP@50-95 with training resolution for each YOLOv11 model variant, while Figure 3 presents performance trends across model variants at fixed resolutions. Increasing the resolution from 320×320 to 640×640 results in a substantial improvement in accuracy for every model size, with smaller additional improvements at 1280×1280.



(a) (b)

**Figure 3:** Detection accuracy across YOLOv11 model variants for different training resolutions: (a) mAP@50 and (b) mAP@50-95.

Similarly, mAP@50-95 exhibits even greater sensitivity to resolution, often increasing by nearly 60-100% when comparing 320×320 to 1280×1280 training resolutions. These improvements indicate that higher training resolutions preserve fine-grained spatial information that is essential for precise object localization, especially under stricter IoU thresholds. Overall, the results demonstrate that input resolution is a first-order determinant of YOLOv11 performance, frequently exerting a stronger influence than architectural scaling alone.

### Diminishing Returns Beyond 640×640 for mAP@50

Although increasing the resolution from 320×320 to 640×640 leads to marked performance improvements, further gains from 640×640 to 1280×1280 are modest for mAP@50. Across most variants, mAP@50 already exceeds 0.96 at 640×640, with absolute improvements below 1.5% at higher resolution. In contrast, mAP@50-95 continues to benefit from increased resolution, indicating that larger inputs primarily enhance localization precision rather than object presence detection. These results suggest that 640×640 offers an effective performance efficiency trade-off, while 1280×1280 is most beneficial for applications requiring high localization accuracy.

### Resolution Scaling Can Outperform Model Scaling

At higher training resolutions, smaller YOLOv11 variants can rival or even outperform larger models trained at lower resolutions. For example, YOLOv11-N trained at 1280×1280 achieves substantially higher performance than YOLOv11-L trained at 320×320 across all evaluated metrics, including mAP@50, mAP@50-95, precision, and recall. Similarly, YOLOv11-S trained at 640×640 matches or exceeds the performance of YOLOv11-M and YOLOv11-L trained at 320×320. These results indicate that input resolution contributes more strongly to detection performance than model parameter count, particularly for small object detection and precise localization. From a practical perspective, this finding implies that for compute constrained deployments, prioritizing higher input resolution can be more effective than increasing model size. Figure 3(a) and Figure 3(b) further demonstrate that resolution scaling often yields larger gains than increasing model size alone.

### Anomalous Behavior of YOLOv11-X at Low Resolution

YOLOv11-X demonstrates unexpectedly poor performance when trained at a resolution of 320×320. At this resolution, it records the lowest mAP@50 (0.676) and mAP@50-95 (0.266) among all evaluated YOLOv11 variants, with precision and recall also showing substantial degradation. This behavior suggests that the representational capacity of very large models is under utilized when spatial detail is insufficient, limiting their ability to learn discriminative features effectively. In contrast, at higher resolutions (640×640 and 1280×1280), YOLOv11-X becomes competitive and achieves strong detection performance. The key implication is that very large models require adequately high input resolution to fully exploit their capacity.

### Precision-Recall Balance Improves with Resolution

Precision and recall improve consistently with increasing training resolution across all YOLOv11 variants. Precision gains indicate reduced false positives, while recall improves markedly from 320×320 to 640×640, reflecting enhanced detection of small or challenging objects. At 1280×1280, most models achieve precision and recall values above 0.96, indicating stable and reliable detection behavior. Overall, higher training resolutions strengthen both classification confidence and object coverage, leading to more dependable real-world performance.

### 4.2 Inference Time Analysis Across Resolutions and Model Variants

To assess the inference time implications of model and resolution scaling, inference latency was evaluated for all fifteen fine-tuned YOLOv11 models. Ten images were randomly selected from the test set and independently resized to three inference resolutions: 320×320, 640×640, and 1280×1280. Each resized image was then processed by every YOLOv11 variant, irrespective of its training resolution. Table 2 reports the per image inference time (in milliseconds), where each column denotes a specific YOLOv11 model fine-tuned at a given training resolution (e.g., YOLOv11n_320, indicating the YOLOv11-N model fine-tuned on 320×320 images), and each row corresponds to a test image evaluated at a particular inference resolution (e.g., 1_640.jpg, where image 1.jpg was resized to 640×640 before inference).

**Table 2:** Image inference times (ms) for YOLOv11 variants trained at different resolutions and evaluated on images of various test resolutions.

| **Image Name** | **yolo11n_320** | **yolo11n_640** | **yolo11n_1280** | **yolo11s_320** | **yolo11s_640** | **yolo11s_1280** | **yolo11m_320** | **yolo11m_640** | **yolo11m_1280** | **yolo11l_320** | **yolo11l_640** | **yolo11l_1280** | **yolo11x_320** | **yolo11x_640** | **yolo11x_1280** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1_320.jpg | 61.84 | 182.37 | 340.31 | 79.47 | 314.6 | 984.92 | 183.34 | 603.99 | 2435.86 | 255.03 | 736.74 | 3133.08 | 394.52 | 1604.97 | 5358.06 |
| 1_640.jpg | 84.99 | 226.37 | 333.92 | 97.52 | 321 | 926.97 | 173.83 | 611.11 | 2443.52 | 225.86 | 762.37 | 2782.72 | 393.21 | 1547.05 | 5572.34 |
| 1_1280.jpg | 91.3 | 152.16 | 401.01 | 126.52 | 238.55 | 1049.96 | 234.15 | 598.83 | 2335.81 | 252.13 | 746.63 | 2815.33 | 440.34 | 1766.6 | 5368.98 |
| 2_320.jpg | 81.29 | 139.94 | 343.24 | 119.47 | 250.98 | 951.83 | 216.72 | 501.45 | 2315.21 | 263.93 | 818.19 | 2897.5 | 446.17 | 1447.98 | 5482.32 |
| 2_640.jpg | 70.98 | 123.08 | 358.49 | 111.07 | 228.38 | 981.13 | 199 | 514.24 | 2297.99 | 242.18 | 688.54 | 2797.35 | 412.9 | 1679.87 | 5514.9 |
| 2_1280.jpg | 126.23 | 164.11 | 402.7 | 127.25 | 307.68 | 1055.86 | 342.04 | 898.84 | 2468.11 | 215.29 | 725.02 | 2900.67 | 402.63 | 1656.29 | 5935.13 |
| 3_320.jpg | 58.9 | 116.84 | 338.72 | 115.21 | 206.52 | 956.58 | 206.13 | 511.88 | 2380.61 | 239.1 | 737.8 | 2876.31 | 425.88 | 1479.07 | 5563.52 |
| 3_640.jpg | 65.95 | 126.94 | 336.59 | 98.36 | 223.95 | 1005.18 | 187.76 | 505.14 | 2263.94 | 228.04 | 738.04 | 2823.12 | 420.64 | 1500.46 | 5474.09 |
| 3_1280.jpg | 75.36 | 136.16 | 354.95 | 103.52 | 352.4 | 951.75 | 203.29 | 541.35 | 2270.02 | 247.69 | 732.83 | 2829.92 | 415.91 | 1527.25 | 5573.36 |
| 4_320.jpg | 86.64 | 137.13 | 358.77 | 141.58 | 235.58 | 1104.57 | 218.63 | 581.92 | 2335.82 | 277.55 | 782.81 | 3040.61 | 491.09 | 1719.8 | 5691.83 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4_640.jpg | 75.16 | 136.23 | 369.01 | 116.16 | 262.72 | 1129.35 | 226.58 | 533.87 | 2346.47 | 250.19 | 716.57 | 2976.73 | 420.08 | 1365.51 | 5342.09 |
| 4_1280.jpg | 87.96 | 145.91 | 354.27 | 119.78 | 255.27 | 1187.3 | 200.84 | 590.73 | 2257.91 | 259.8 | 815.58 | 2861.25 | 419.86 | 1655.92 | 6045.97 |
| 5_320.jpg | 57.85 | 117.77 | 442.09 | 90.94 | 279.41 | 839 | 152.3 | 489.74 | 2223.09 | 214.29 | 714.85 | 2798.9 | 435.25 | 1241.24 | 5574.71 |
| 5_640.jpg | 124.46 | 166.63 | 369.95 | 153.12 | 240.91 | 868.78 | 165.51 | 574.12 | 2171.25 | 251.4 | 773.92 | 2686.94 | 453.33 | 1340.51 | 5434.25 |
| 5_1280.jpg | 97.66 | 154.79 | 401.52 | 103.58 | 237.52 | 974.38 | 213.99 | 543.73 | 2173.15 | 237.15 | 686.43 | 2645.52 | 464.48 | 1420.09 | 5399.15 |
| 6_320.jpg | 62.62 | 119.71 | 328.73 | 96.62 | 254.21 | 869.23 | 180.79 | 538.07 | 2131.9 | 206.83 | 640.45 | 2677.71 | 443.03 | 1330.56 | 5507.66 |
| 6_640.jpg | 65.18 | 118.45 | 341.33 | 87.29 | 224.9 | 802.79 | 182.02 | 512.16 | 2040.77 | 207.52 | 640.92 | 2600.59 | 370.1 | 1280.55 | 5183.13 |
| 6_1280.jpg | 84.68 | 180.3 | 366.55 | 117.62 | 264.21 | 847.75 | 194.09 | 549.79 | 2230.1 | 230.8 | 748.15 | 2734.75 | 383.3 | 1305.31 | 5780.85 |
| 7_320.jpg | 73.3 | 133.27 | 340.45 | 113.8 | 243.45 | 859.78 | 184.83 | 510.54 | 2166.44 | 214.02 | 650.85 | 2717.31 | 378.34 | 1338.59 | 5201.21 |
| 7_640.jpg | 79.99 | 148.81 | 399.2 | 111.06 | 245.5 | 831.23 | 183.36 | 532.76 | 2143.74 | 221.29 | 667.64 | 2688.67 | 417.28 | 1281.67 | 5325.51 |
| 7_1280.jpg | 129.4 | 279.24 | 382.74 | 170.14 | 248.01 | 1013.52 | 183.21 | 559.37 | 2112.38 | 233.32 | 663.17 | 2683.9 | 414.9 | 1325.18 | 5361.7 |
| 8_320.jpg | 74.14 | 157.1 | 345.6 | 97.9 | 226.07 | 830.46 | 174.6 | 535.13 | 2119.75 | 209.34 | 652.32 | 2672.68 | 406.51 | 1291.66 | 5359.19 |
| 8_640.jpg | 86.68 | 146.91 | 406.41 | 123.46 | 247.26 | 858.64 | 172.35 | 539.51 | 2185.38 | 218.46 | 645.29 | 2607.13 | 499.4 | 1337.7 | 5377.86 |
| 8_1280.jpg | 82.42 | 158.77 | 359.47 | 103.79 | 268.24 | 901.26 | 183.23 | 563.81 | 2125.33 | 224.15 | 716.84 | 2601 | 382.76 | 1313.75 | 5371.68 |
| 9_320.jpg | 64.14 | 109.48 | 319.46 | 84.82 | 213.31 | 830.45 | 173.25 | 514.77 | 2186.71 | 219.43 | 670.65 | 2626.28 | 392.86 | 1255.89 | 5272.2 |
| 9_640.jpg | 79.77 | 122.04 | 350.99 | 91.52 | 229.29 | 815.7 | 170.92 | 498.38 | 2048.5 | 225.99 | 635.48 | 2601.58 | 386.89 | 1258.78 | 5328.26 |
| 9_1280.jpg | 88.77 | 141.46 | 399.05 | 106.75 | 267.46 | 887.92 | 203.68 | 510.82 | 2100.79 | 220.97 | 670.98 | 2709.42 | 399.11 | 1277.2 | 5379.88 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10_320.jpg | 55.81 | 161.64 | 361.23 | 253.75 | 245.79 | 1148.25 | 177.14 | 510.97 | 2422.64 | 213.11 | 742.76 | 2744.06 | 534.85 | 1536.37 | 5375.94 |
| 10_640.jpg | 79.7 | 159.92 | 341.48 | 94.94 | 206.59 | 906.41 | 202.24 | 548.51 | 2289.55 | 247.98 | 793.43 | 2783.59 | 557.05 | 1535.75 | 5621.61 |
| 10_1280.jpg | 162.41 | 194.86 | 446.14 | 172.99 | 306.36 | 1099.48 | 342.54 | 706.05 | 2196.6 | 414.11 | 965.36 | 2881.93 | 799.99 | 1736.29 | 5466.74 |

### Influence of Model Size on Inference Time

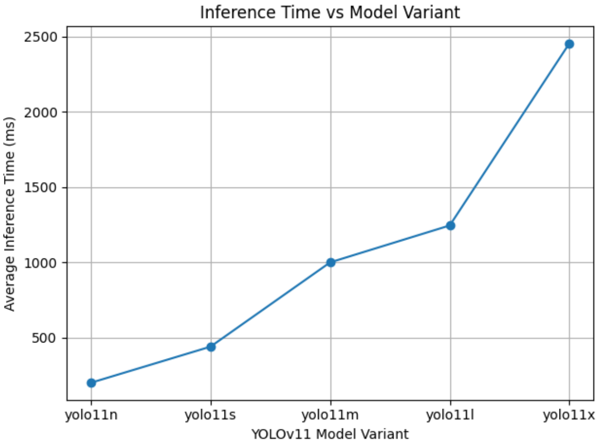


**Figure 4:** Average inference time across YOLOv11 model variants.

An analysis of the YOLOv11 model variants reveals a strong linear relationship between model size and inference latency, which can be observed in Figure 4. Across all evaluated input resolutions in Table 2, the speed hierarchy remains consistent: YOLOv11n is the fastest, followed sequentially by YOLOv11s, YOLOv11m, YOLOv11l, and finally YOLOv11x, which exhibits the highest latency. This trend confirms that architectural complexity driven by parameter count and network depth is a primary determinant of inference time.

### Influence of Image Resolution during Training on Inference Time

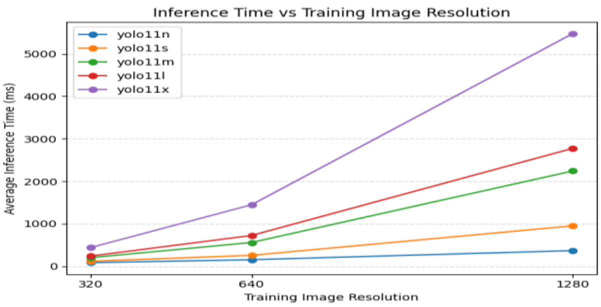


**Figure 5:** Effect of training image resolution on inference time for YOLOv11 models.

To evaluate the effect of training resolution on model performance, consider the example image 1_320.jpg. The recorded inference times for the three variants are as follows: YOLOv11n_320 achieves 61.84 ms, YOLOv11n_640 records 182.37 ms, and YOLOv11n_1280 reaches 340.31 ms. These results indicate a clear trend: the model trained at 320×320 resolution exhibits the lowest inference latency, whereas the model trained at 1280×1280 resolution incurs the highest latency, with the 640×640 model positioned between the two.

Overall, Figure 5 shows that inference time increases monotonically with higher training image resolutions, suggesting that models trained on larger input sizes require more computational effort during prediction. This behavior suggests that higher training resolutions lead to models with increased internal feature map sizes and computational pathways, which persist during inference and result in higher execution time.

**Effect of Inference-Time Image Resolution on Inference Latency**

Varying the inference image size while keeping model weights fixed shows that inference resolution has a comparatively weaker impact on latency. For smaller models (YOLOv11n and YOLOv11s), increasing inference resolution results in modest and sometimes non-monotonic latency changes, with similar runtimes observed at 640×640 and 1280×1280. For larger models (YOLOv11m, YOLOv11l, and YOLOv11x), inference time remains consistently high across all inference resolutions, suggesting that runtime is dominated by model depth rather than input size. Overall, inference resolution plays a secondary role compared to model scale and training resolution.

**Influence of Input Image Content on Inference Time**

An examination of inference times across different images of the same resolution, as shown in Table 2, indicates that input image content has minimal effect on prediction latency. For example, using the yolov11n_320 model, the recorded times for five distinct images are: 1_320.jpg with 61.84 ms, 2_320.jpg with 81.29 ms, 3_320.jpg with 58.90 ms, 4_320.jpg with 86.64 ms, and 5_320.jpg with 57.85 ms. The variation across these samples remains relatively small, generally within 30 ms.

These fluctuations appear to be random rather than systematic, indicating that YOLO inference time is largely independent of the visual content within the input image. The computational workload is driven primarily by model architecture and input resolution, not by the specific objects or scene complexity present in a given image.

**Effect of Image Resolution on Larger YOLO Models**

Large models, particularly YOLOv11x, exhibit poor scaling with high resolution inputs. As shown in Table 2, average inference time for YOLOv11x increases sharply from approximately 400-550 ms at 320×320 to 1250-1750 ms at 640×640, and further to 5300-6000 ms at 1280×1280. This nonlinear growth arises from the interaction between deep architectures and increasing spatial resolution, which amplifies computational cost through larger intermediate feature maps at every network stage.

In summary, inference latency in YOLOv11 is dominated by model scale and training resolution, while inference-time image resolution and input content exert comparatively secondary effects.

## Conclusion

This study presented a comprehensive evaluation of image resolution sensitivity in YOLOv11 across multiple model scales using a controlled experimental setup. The results clearly demonstrate that training image resolution is a first-order factor governing detection performance, frequently exerting a stronger influence than model scaling. Increasing resolution from 320×320 to 640×640 yields substantial gains in accuracy, particularly for small object detection and localization, while further scaling to 1280×1280 provides diminishing returns for coarse metrics such as mAP@50 but continues to improve localization precision measured by mAP@50-95. The analysis further reveals that resolution scaling can outperform architectural scaling, enabling smaller models trained at higher resolutions to rival or exceed larger models trained at lower resolutions. Inference experiments show that runtime is dominated by model size and training resolution, whereas inference-time resolution and image content have comparatively minor effects. Overall, these findings provide practical guidance for selecting appropriate resolution and model combinations, enabling more informed trade-offs between accuracy and computational efficiency in real world YOLOv11 deployments.

## References

1. F. Japar, H.R. Ramli, N.M.H. Norsahperi, W.Z.W. Hasan, “Oil palm loose fruit detection using YOLOv4 for an autonomous mobile robot collector”, IEEE Access, 2024, 12, 138582–138593.
2. S. Zhang, Y. Wu, C. Men, X. Li, “Tiny YOLO optimization oriented bus passenger object detection”, Chinese Journal of Electronics, 2019, 29 (1), 132–138.
3. Priadana, D.-L. Nguyen, X.-T. Vo, J. Choi, R. Ashraf, K. Jo, “HFD-YOLO: Improved YOLO network using efficient attention modules for real-time one-stage human fall detection”, IEEE Access, 2025, 13, 41248–41258.
4. Q. Zhang, K. Ahmed, M.I. Khan, H. Wang, Y. Qu, “YOLO-FCE: A feature and clustering enhanced object detection model for species classification”, Pattern Recognition, 2026, 171, 112218.
5. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, “You only look once: Unified, real-time object detection”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 2016, Las Vegas, Nevada, United States of America, 779–788.
6. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, “SSD: Single shot multibox detector”, Proceedings of the European Conference on Computer Vision, October 2016, Amsterdam, The Netherlands, 21–37.
7. R. Girshick, “Fast R-CNN”, Proceedings of the IEEE International Conference on Computer Vision, December 2015, Santiago, Chile, 1440–1448.
8. H. Zhu, W. Ling, H. Yan, X. Zhong, F. Liao, “YOLO-MP: A lightweight forest fire detection model”, Ecological Informatics, 2025, 92, 103516.
9. H. Bagherpour, N.F. Peyruo, “Enhanced YOLO-based framework for accurate detection and identification of common wheat impurities with distinct objects”, Scientific Reports, 2025, 15, 40436.
10. J. Yao, Y. Li, Z. Xia, P. Nie, X. Li, Z. Li, “WTAD-YOLO: A lightweight tomato leaf disease detection model based on YOLO11”, Smart Agricultural Technology, 2025, 12, 101349.

11. Z. Fu, F. Zhang, X. Ren, B. Hao, X. Zhang, C. Yin, G. Li, Y. Zhang, “LE-YOLO: Lightweight and efficient detection model for wind turbine blade defects based on improved YOLO”, IEEE Access, 2024, 12, 135985–135998.
12. Y. Meng, J. Zhan, K. Li, et al., “A rapid and precise algorithm for maize leaf disease detection based on YOLO MSM”, Scientific Reports, 2025, 15, 6016.
13. Peng, T.-K. Kim, “YOLO-HF: Early detection of home fires using YOLO”, IEEE Access, 2025, 13, 79451–79466.
14. Y. Hao, H. Pei, Y. Lyu, Z. Yuan, J.-R. Rizzo, Y. Wang, Y. Fang, “Understanding the impact of image quality and distance of objects to object detection performance”, Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, October 2023, Detroit, Michigan, United States of America, 11436–11442.
15. Riis, “Aerial Sheep Dataset”, Roboflow Universe Dataset, 2022. https://universe.roboflow.com/riis/aerial-sheep