INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 2, February 2026 | ISSN: 2456-4184 | IJNRD.ORG

Performance–Efficiency Trade-Off Analysis of YOLOv11 Variants for Face Anti-Spoofing

Satish Kumar Nath[1], Dr. Pallavi Pratap[2]
Research Scholar[1], Associate Professor[2]
Computer Sc. & Engg. Department, Maulana Azad University, Jodhpur, India
satishkumar.nath@gmail.com[1], pratappallavi@gmail.com[2]

ABSTRACT

Face anti-spoofing systems should be able to offer high detection rates with reduced computation requirements to be used as a biometric system operating in real-time. Deep learning models are highly capable of spoof detection; still their escalating complexity frequently makes them impractical on resource-constrained platforms. The current paper gives a systematic performance-efficiency trade-off analysis of several YOLOv11 architectural variants. Names of these variants are Nano, Small, Medium, Large, and Extra Large. Each variant is tested against a common dataset and an experimental setup based on a common dataset and experimental set up in terms of precision, recall, F1-score, accuracy, mean Average Precision, and inference speed. The results of the experiments indicate the existence of trade-offs between the model complexity and performance with lightweight variants of YOLOv11 being appropriate to run on a real-time and on an edge. On the other hand, larger variants are appropriate to offer higher accuracy to security-critical systems. The work offers viable recommendations on the use of suitable YOLOv11 variants depending on the deployment needs.

Index Terms
Face anti-spoofing, presentation attack detection, YOLOv11, model efficiency, real-time inference, deep learning.

I. INTRODUCTION

Use of face recognition is highly embraced in Biometric authentication systems because it is non intrusive and easily deployed [1]. Nevertheless, face recognition systems are susceptible to presentation attacks, such as printed photographs, replayed videos, and display-based spoofing. These attacks intimidate the security of the system [3], [29]. Presentation attack detection (PAD) which is also known as face anti-spoofing tries to overcome these vulnerabilities by authenticating the liveness of facial inputs [2].

The latest innovations in the field of deep learning have made high improvements in terms of face anti-spoofing performance, as automatic learning of discriminative features becomes feasible [14], [18]. However, real-time and edge-based deployments of modern deep neural networks can be very expensive in terms of computation and cannot be performed anywhere [24]. In real-world systems, to choose the right model, one has to balance between detection accuracy, inference speed and required hardware.

YOLO-based models have become popular by virtue of their ability to perform in real-time and scale [11], [12]. YOLOv11 model has several architectural variants with variation in depths and widths. It allow us to do a systematic study of the trade-offs between the desired accuracy and efficiency. This paper explores the variants of the YOLOv11 architecture in the face anti-spoofing to offer deployment-based information.

II. RELATED WORK

The initial work on face anti-spoofing involved handcrafted feature extraction like texture-based and frequency-domain features along with classical classifiers [4], [5], [9]. Although these methods showed fair performance in controlled conditions, they were not strong enough to withstand the changes in the environment [6].

Later, deep learning-based methods have proved to be predominant with CNN architectures including VNN, ResNet and DenseNet showing higher accuracy of spoof detection [11]-[13],[14]. The recent works examined auxiliary supervision, domain generalization and frequency-conscious learning to enhance robustness across datasets [16], [20], [25].

Even though lightweight CNNs to face anti-spoofing have been suggested to make mobile deployment feasible [24], none of the studies offer a comparative analysis at various levels of the architecture within the same model family. The literature has this gap since the paper compares variants of YOLOv11 to determine them under the same experimental conditions.

III. YOLOv11 Variant Architectures

YOLOv11 provides a scalable family of architectures designed for real-time vision tasks. Each variant shares the same core design but differs in depth, channel width, and parameter count:

- **YOLOv11-Nano:** Optimized for minimal latency and low memory usage.
- **YOLOv11-Small:** Balanced model offering real-time inference with improved accuracy.
- **YOLOv11-Medium:** Enhanced representational capacity with moderate computational cost.
- **YOLOv11-Large:** High-capacity model for improved spoof detection accuracy.
- **YOLOv11-Extra Large:** Maximum representational power for security-critical deployments.

These variants enable analysis of how architectural scaling influences face anti-spoofing performance [11], [12].

IV. EXPERIMENTAL SETUP

A. **Dataset:** The total amount of image samples in the dataset was 7092. This dataset was divided into two mutually exclusive classes, namely Real Faces (authentic live images) and Fake Faces (spoofed representations). The fake samples were the attacks on printed photos, the video replays and the display-

B. based spoofing with different illumination and pose conditions. Samples were collected in such manner that it covers different type of backgrounds, lighting levels and face orientations. [8], [17]. So dataset will be able to generalize widely. A stratified sampling method was used to divide the dataset into three subsets, i.e. training, validation and testing. In these subsets, classes (real and fake) are balanced equally.

C. **Experimental Environment:** All experiments were executed using Google Colab Pro with GPU acceleration. All the experiments were performed in the same hardware and software environment to ensure consistency in the experiments of all the YOLOv11 variants.

D. Training Configuration: All YOLOv11 models were trained using standardized hyper-parameters to make a fair comparison. : Models were initialized with pretrained weights to accelerate convergence [12]. Input images were resized to 640×640640 \times 640640×640 pixels, and data augmentation techniques such as flipping, rotation, brightness adjustment, and noise injection were applied [24]. During execution, concept of early stopping was initiated. It encourages convergence stability, when validation loss stopped gaining within 20 consecutive epochs. Checkpoints were model checkpoints that were saved at the epoch that had the largest value of validation mAP@0.5. The pre-trained weights were used to initialize all YOLO models using the COCO dataset. It allows quicker convergence and better performance with a small number of samples.

E. Evaluation Metrics: Performance was evaluated using Precision, Recall, F1-score, Accuracy, and mean Average Precision (mAP@0.5) [29]. Computational efficiency was assessed using inference speed measured in frames per second (FPS) and milliseconds per frame [30]. Model evaluation was performed using both **quantitative** and **qualitative** metrics. Some of the important matrices are follows:

- **Precision (P):** ratio of positive sample (correct perdition) to all predicted positives.
- **Recall (R):** ratio of correctly predicted positives to all actual positives.
- **F1-Score:** harmonic mean of precision and recall, balancing both error types.
- **mAP@0.5:** mean Average Precision at 0.5 IoU threshold. It measures localization and classification accuracy.
- **mAP@0.5–0.95:** averaged precision across multiple IoU thresholds (0.5 to 0.95 in 0.05 steps) It provide a detailed accuracy measure.
- **Loss Functions:** There are mainly Box Loss, Classification Loss and Distribution Focal Loss (DFL). These are validation losses to ensure stable convergence.
- **Inference Speed:** measured as frames per second (FPS) and milliseconds per frame (ms/frame). It reflects the model efficiency.

V. RESULTS AND PERFORMANCE ANALYSIS

Experimental results demonstrate clear performance–efficiency trade-offs across YOLOv11 variants. Lightweight models such as YOLOv11-Nano and YOLOv11-Small achieve high inference speed with ompetitive accuracy, making them suitable for real-time and edge-based applications [24].

Table 1: Comparison of all YOLOv11 Variants for Classification

Metric	N	S	M	L	XL
Precision	0.98	0.97	0.96	0.97	0.96
Recall	0.94	0.94	0.94	0.96	0.95
F1-score	0.96	0.95	0.95	0.96	0.96
mAP@0.5	0.97	0.97	0.98	0.97	0.96
mAP@0.5–0.95	0.89	0.89	0.89	0.89	0.88
Box Loss (val)	0.47	0.46	0.46	0.45	0.46
Cls Loss (val)	0.33	0.32	0.32	0.26	0.29
DFL Loss (val)	1.22	1.25	1.23	1.23	1.22

Table 1 shows the comparative study of results gathered from experiments done with all variant of YOLOv11 with classification problem. The findings reveal that the all variants of YOLOv11 achieve high accuracy in classification. The table shows that precision, recall and F1-scores are relatively high. These values are ranging between 0.96 and 0.97. It shows that all variant of YOLOv11 are stable and reliable to learn at all scales. With the increase of the model size (Nano, Medium and Large) the accuracy is increased up to 0.98. Parameter mAP@0.5 shows features representation and classification discrimination. This parameter improves marginally. In results highest value of mAP@0.5 is achieved with medium variant.

Table1 shows that extra large model has best results in Precision 0.96, Recall 0.96 and Accuracy 0.98. But it has slightly lower value of mAP@0.5 (0.96) than Medium model (0.98). Result shown in table demonstrates that larger architectures have significant improvements but it costs us a high price of increased computation. Smaller models (like Nano and Small) provided almost the same performance at significantly faster inference speed, and were suitable to lightweight deployment.

YOLOv11-Extra Large is a very stable classifier with high-quality accuracy. It has balanced performance with precision and recall. This is suitable in situations with a high stakes or complex classification where the computational resources are not limiting. Large and Medium models are not substantial with compare to extra large variant. Sometime the Extra Large model is most justified in case the maximum robustness and consistency are the priorities rather than efficiency. In most real-world classification applications, YOLOv11 medium or YOLOv11 large would be more balanced approach. YOLOv11 extra large is the last and most powerful when the main goal is to achieve performance.

Table2: Comparison of all YOLOv11 Variants for Detection

Metric	N	S	M	L	XL
Precision	0.98	0.97	0.96	0.96	0.96
Recall	0.94	0.94	0.94	0.94	0.95
F1-score	0.96	0.95	0.95	0.95	0.96
mAP@0.5	0.97	0.97	0.98	0.98	0.96
mAP@0.5–0.95	0.89	0.89	0.89	0.89	0.88
Box Loss (val)	0.47	0.46	0.46	0.46	0.46
Cls Loss (val)	0.33	0.32	0.32	0.32	0.29
DFL Loss (val)	1.22	1.25	1.23	1.23	1.22

Comparison analysis is shown in below in Table 2. This study determines the effect of various YOLOv11 models scales on detecting accuracy, recall and efficiency. Smaller models came at much faster inference times with very little accuracy loss. At other hand, larger models showed better recall and slightly higher F1-scores.

The comparison shown in above table, of the variants of YOLOv11 reveals that they all demonstrate high levels of performance consistently. The highest precision (0.98) belongs to YOLOv11-Nano. This is appropriate in the case when the false positives should be reduced to the minimum possible. At the same time, the overall accuracy is rather high. The best trade-off among the three is YOLOv11-Large with the highest recall (0.96) and F1-score (0.96), which is very essential to reduce the cases of missed spoof attacks.

YOLOv11-Medium has the best mAP@0.5 (0.98), as it has the best localization ability at larger values of IoU (standard). All the models show similar values of mAP@0.50.95 (0.88 to 0.89). It indicates consistency in the implementation of the models in strict demand areas. The final conclusion for detection task is YOLOv11-Large is the best model to use in face anti-spoofing detection with balanced precision and recall trade-off whereas, YOLOv11-Nano is

the best to be used in resource-constrained settings where it is lightweight and thus can be deployed.

YOLOv11-Medium and YOLOv11-Large provide improved detection accuracy due to enhanced feature representation, at the cost of increased computational complexity. YOLOv11-Extra Large achieves the highest accuracy across evaluation metrics but requires substantial computational resources, limiting its suitability to server-based deployments. These findings indicate that no single model is universally optimal; instead, variant selection should be guided by application requirements and hardware constraints [17], [29].

Trade-Offs and Deployment Considerations: The selection of the model for a particular application will be based on the intended deployment environment.

- YOLOv11Nano and Small variants are ideally suitable for mobile or embedded applications. In such application domain performance and power consumption are important factors.
- YOLOv1 Medium variant is ideal to use in real-time verification. Some of the examples are enterprise attendance, verification at ATM machine or online examination systems. Medium variant is suitable for such applications because of its moderate speed-accuracy profile.
- YOLOv11 Large and XLarge models are useful in certain application domains. Examples are centralized high-security verification, forensic analysis or multi-user real-time surveillance systems. Such larger models can utilize their high GPU resources.

Besides, the common training system for all variants guarantees linear scaling of model performance to obtain hardware accessibility. In case of a dynamic trade-off scenario (e.g., Nano-based assistance to live capture screening and XLarge-based assistance to post-verification), hybrid deployment (e.g., Nano and XLarge) can be used to provide efficient end-to-end anti-spoofing workflows.

VI. DISCUSSION

The experimental results enable to confirm that the YOLOv11 architecture is efficient in the recognition of real and spoofed facial inputs by visual and textual features. All the models performed well with the precision and recall value being higher than 0.94. The Medium type was found to have the best trade-off between computational efficiency and detection accuracy. It indicates that Medium model is to be the most balanced type of the face anti-spoofing system to be used in the real-world.

Curve, matrix and loss behavior analysis showed that YOLOv11 models are very consistent and have low overfitting with scales. Although, bigger models were better in recall and stability. Smaller versions had better real-time capabilities. These findings support the practicability of the use of YOLOv11-based frameworks to achieve dependable, quick and scalable face anti-spoofing recognition. This trade-off underscores the necessity of deployment-aware model selection in biometric security systems [24], [30].

VII. CONCLUSION

This paper presented a comprehensive performance–efficiency trade-off analysis of YOLOv11 variants for face anti-spoofing. The analysis of the entire experiment proves beyond a doubt that the YOLOv11 structure is a unified, effective and precise face anti-spoofing recognition system. In both detection and classification paradigms, all variants performed with high performance and steady convergence. This model is also feasible in real-time. The most balanced one was the YOLOv11-Medium model, which offered the highest possible accuracy with a moderate amount of computation. The YOLOv11-XLarge provided the highest accuracy in mission-critical verification.

All these results demonstrate that the YOLOv11-based models are more effective than conventional anti-spoofing approaches. These models also offer a scalable architecture (Nano to XLarge) that can be implemented in a variety of real-world scenarios. This effectiveness of such a method supports the possibility of the birth of modern deep convolutional detectors. These will be linkage between speedy visual recognition and solid liveness verification. This will eventually increases the credibility of biometric authentication methods.

REFERENCES:

[1] A. K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.
[2] S. Marcel, M. S. Nixon, and S. Z. Li, *Handbook of Biometric Anti-Spoofing*, 2nd ed. Cham, Switzerland: Springer, 2019.
[3] J. Galbally, S. Marcel, and J. Fierrez, “Biometric antispoofing methods: A survey,” *Pattern Recognit. Lett.*, vol. 32, no. 16, pp. 213–224, 2014.
[4] I. Chingovska, A. Anjos, and S. Marcel, “On the effectiveness of local binary patterns in face anti-spoofing,” in *Proc. IEEE BIOSIG*, 2012, pp. 1–7.
[5] Z. Boulkenafet, J. Komulainen, and A. Hadid, “Face spoofing detection using colour texture analysis,” *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1818–1830, 2016.
[6] T. de Freitas Pereira et al., “Face liveness detection using dynamic texture,” *EURASIP J. Image Video Process.*, vol. 2014, no. 1, 2014.
[7] J. Komulainen, A. Hadid, and M. Pietikäinen, “Context based face anti-spoofing,” in *Proc. IEEE BTAS*, 2013, pp. 1–8.
[8] Z. Zhang et al., “A face antispoofing database with diverse attacks,” in *Proc. IEEE BTAS*, 2012, pp. 1–8.
[9] J. Li, Y. Wang, T. Tan, and A. Jain, “Live face detection based on the analysis of Fourier spectra,” in *Proc. SPIE*, 2004, pp. 296–303.
[10] K. Kollreider et al., “Real-time face detection and motion analysis with application in liveness assessment,” *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 548–558, 2007.
[11] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv:1409.1556*, 2014.
[12] K. He et al., “Deep residual learning for image recognition,” in *Proc. IEEE CVPR*, 2016, pp. 770–778.
[13] G. Huang et al., “Densely connected convolutional networks,” in *Proc. IEEE CVPR*, 2017, pp. 4700–4708.
[14] Y. Liu et al., “Learning deep models for face anti-spoofing: Binary or auxiliary supervision,” in *Proc. IEEE CVPR*, 2018, pp. 389–398.
[15] A. George and S. Marcel, “Deep pixel-wise binary supervision for face presentation attack detection,” in *Proc. IEEE ICIP*, 2019, pp. 4542–4546.
[16] Z. Yu et al., “Searching central difference convolutional networks for face anti-spoofing,” in *Proc. IEEE CVPR*, 2020, pp. 5293–5303.
[17] J. Yang et al., “Face anti-spoofing: Model matters, so does data,” *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 358–372, 2021.
[18] Z. Yu et al., “Deep learning for face anti-spoofing: A survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6763–6781, 2022.
[19] Y. Atoum et al., “Face anti-spoofing using patch and depth-based CNNs,” in *Proc. IEEE IJCB*, 2017, pp. 319–328.
[20] S. Jia et al., “Single-side domain generalization for face anti-spoofing,” in *Proc. IEEE CVPR*, 2020, pp. 848–857.
[21] J. Sun et al., “Domain generalization via adversarial learning for face anti-spoofing,” *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 123–135, 2020.
[22] X. Tu et al., “Learning generalizable representations for face anti-spoofing,” in *Proc. IEEE ICCV*, 2021, pp. 9233–9242.
[23] A. Mohammadi et al., “Domain adaptation for face anti-spoofing,” *IEEE Access*, vol. 8, pp. 134590–134601, 2020.
[24] C. Benrabah et al., “Lightweight CNN for face anti-spoofing on mobile devices,” *IEEE Access*, vol. 9, pp. 16312–16324, 2021.
[25] Z. Wang et al., “Frequency-aware face anti-spoofing,” in *Proc. IEEE CVPR*, 2020, pp. 645–654.
[26] H. Wang et al., “Face anti-spoofing with supervised contrastive learning,” *IEEE Signal Process. Lett.*, vol. 29, pp. 1739–1743, 2022.

[27] S. Chen et al., "Transformer-based face anti-spoofing," *IEEE Access*, vol. 10, pp. 72118–72129, 2022.
[28] Y. Qin et al., "Deepfake face detection: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 1–22, 2023.
[29] J. Fierrez et al., "Biometric presentation attack detection: State of the art," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1270–1285, 2018.
[30] ISO/IEC 30107-3, "Information technology—Biometric presentation attack detection—Part 3: Testing and reporting," ISO/IEC Standard, 2017.