International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0334
e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2324 - 2335

# Real-Time Human Age Estimation Using an Optimized YOLOv11 Framework

*Padakanti Divya[1], Padakanti Swapna[2]*
*[1] Assistant Professor, Dept of Ece, Indur Institute of Engineering and Technology, Siddipet, Telangana*
*[2] Research Scholar, Dept of CSE, SR University, Warangal, Telangana*
***Emails:*** *divya.padakanti419@gmail.com[1]*

## Abstract
*Object detection, which finds and identifies objects in an image or video, is a crucial component of computer vision. YOLO (You Only Look Once) models are real-time object identification algorithms that identify and classify objects in an image. Because YOLO processes the entire image in a single pass, it is faster and more efficient. This study estimates a person's age from facial pictures using the latest real-time object identification model, YOLO v11.YOLOv11 boasts exceptional speed, precision, efficiency, and enhanced feature extraction. The suggested approach recognizes faces and classifies them into four age groups: adolescents (ages 13 to 19), young adults (ages 20 to 35), adults (ages 36 to 55) and seniors (above 55). Precision, Recall, F1-Score are increased by 1%. Real-time object detection AI, surveillance and security hidden object puzzles (games), transportation and autonomous driving (booking services), agriculture and environmental monitoring, healthcare and specialized fields make use of YOLO models.*

***Keywords:*** *Computer vision, YOLO, YOLO v11, object identification, facial pictures, enhanced feature extraction, precision, adolescents, young adults, adults, seniors.*

## 1. Introduction
A well-liked object detection model called YOLO (You Only Look Once) is renowned for its quickness and precision. The area of computer vision known as "object detection" is concerned with locating and categorizing items inside an image or video. You Only Look Once (YOLO) suggest employing an end-to-end neural network that simultaneously predicts class probabilities and bounding boxes. Several bounding boxes are predicted by YOLO for each grid cell. Bounding boxes are used to locate one or more things accurately. In order to identify objects in an image, the YOLO algorithm employs a basic deep convolution neural network. YOLO models are utilized in real-time object detection AI, surveillance and security hidden object puzzles (games), autonomous driving and transportation (booking services), agriculture and environmental monitoring, healthcare, and specialized sectors.YOLOv1, YOLOv2, YOLOv3, YOLOv4, YOLOv5, YOLOv6, YOLOv7, YOLOv8, YOLOv9, YOLOv10, and YOLOv11 are some of the versions of YOLO that have been produced. Better handling of small objects, faster processing, and increased precision are just a few of the innovations that have been added to each iteration shown in Figure 1.
.

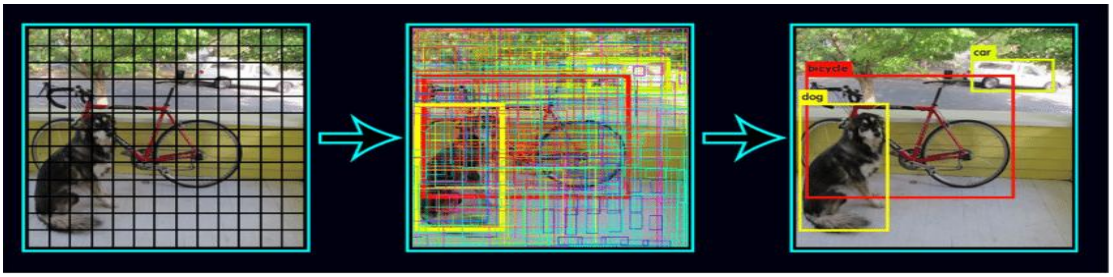


**Figure 1: Illustration of The Object Detection Process, Including Grid Generation, Bounding Box Prediction, And Final Detection Results.**

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0334
e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2324 - 2335

Grid generation and bounding box prediction for object identification are depicted in the above graphic. A number of anchor-based bounding box hypotheses with class probabilities and confidence ratings are produced for each grid cell. CNN predicts both the bounding box coordinates and the anchor boxes in YOLO v2. To align the anticipated bounding boxes with the real objects in the picture, anchor boxes—pre-defined boxes with various aspect ratios and scales are utilized. YOLO v2 can now handle objects of various sizes and forms more effectively. Darknet-53, FPN is a novel CNN architecture used in YOLO v3. An object detection variation of the Res Net architecture is called Darknet-53.The CNN architecture known as FPN, or feature pyramid networks, is used to recognize objects at various scales. YOLO v3 can even detect small objects thanks to these features. YOLO v4 extract features from the image by using Spatial Pyramid Processing (SPP) YOLO v4 recognize items in the image of different sizes. Cross-stage partial connection (CSP) improves the model's accuracy. YOLOv4 is used for getting optimal speed and accuracy in object detection. YOLOv5 uses a CSPNet (Cross Stage Partial Network) backbone and a PANet (Path Aggregation Network) neck, for object detection. YOLO v5 uses a single convolution layer to directly predict the bounding box coordinates, to object detection of various sizes. YOLO v5 uses Cross mini-batch normalization (CmBN) to improve accuracy in object detection.CSPNetPANet Bounding boxes are used by the machine learning model YoloV6 to detect objects. Anchor-based and anchor-free techniques are combined in YOLOv6's Anchor-Aided Training (AAT). YOO V6 applies flipping, scaling, and rotation to the input photos. The design of YOLO v6 is lighter and more effective. YOLOv8 is a real-time computer vision model for classification, segmentation, and detection. The characteristics of YOLO v8 are Versatility, New Backbone Network, and Anchor-Free Detection. It combines context aggregation with spatial attention. YOLO v8 offers increased accuracy and quickness. The Generalized Efficient Layer Aggregation Network (GELAN) and Programmable Gradient Information (PGI) are introduced in YOLO v9. YOLO v9 reduces parameters and processing while improving accuracy and efficiency. YOLO v10 is a real-time object detection model that does not use NMS (Non-Maximum Suppression). Post-processing is eliminated in YOLO v10. YOLO v10 lowers computational overhead and latency. The C3k2, SPPF, and C2PSA block architectures are used in YOLO v11. In object identification, segmentation, posture estimation, tracking, and classification, YOLO v11 improved speed and accuracy. YOLOv11 detects objects of various sizes using a head. Three distinct scales (low, medium, and high) of detection boxes are output by the head.

## 2. Literature Review

Hamdi Dibeklioglu, Albert Ali Salah, Fares Alnajar, and Theo Gevers developed a unique hierarchical age estimation architecture based on adaptive age grouping in their 2015 paper "Combining Facial Dynamics With Appearance for Age Estimation."We extensively tested our approach, examining topics such as the dynamics of posed versus spontaneous grins and gender-specific age estimation. The results on the new UvA-NEMO Disgust database presented in this study, and Dibeklioglu et al. [1] show that their suggested method may be applied to other expressions. We show that the mean absolute error can be lowered by up to 21% using the spontaneity of information, improving the quality of facial age prediction. This makes it the biggest dynamic aging study that has been released to date. A Lightweight Convolutional Neural Network for Real and Apparent Age Estimation in Unconstrained Face Images" was proposed by Agbo-Ajala and SerestinaViriri [2], who integrated adaptive image augmentation with a dependable picture pre-processing method. On the FG-NET, MORPH-II, and APPA-REAL datasets, the model obtained results for age prediction accuracy that are comparable to the state-of-the-art despite the lighter CNN model design with short training durations and modestly sized training images. Our proposed methodology outperforms state-of-the-art methods on many age-estimate

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0334
e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2324 - 2335

benchmark datasets, as Wang et al. [3] show through extensive trials. Wang et al. [3] suggested in their research that the ADPF framework enhances the face-based age estimation task's functionality. Fusion Net and Attention Net are combined in their framework. Attention Net includes a new combined attention mechanism termed RMHHA that enables the identification of age-specific patches by learning multiple single-channel attention maps. Fusion Net classifies these patches before using them, and it also uses facial photos to anticipate the final age. When compared to some of the most sophisticated methods, ADPF dramatically improves prediction accuracy, according to study of many benchmark datasets. The CR-MT net, an end-to-end multi-task learning network for age estimation proposed by NA LIU [4], combines regression and age classification. In the CR-MT net, classification serves as a supporting task that improves the generalization performance of the regression job. The various YOLO network modifications that have been made to improve object identification efficacy are discussed by Rekha B. S. [5]. In this study, they have examined the YOLO architecture, which employs a network model for object, pedestrian, obstacle, and solder joint identification. When used on vehicle door panel welding panel lines, YOLO can accurately identify and detect solder connections. The algorithm can find solder junctions, among other things. Jiang [6] provides an overview of the conventional techniques for creating automatic age estimate models, the benchmark datasets used to create these models, and some of the most current literature proposals introducing novel age estimation techniques. An overview of the traditional methods for building automatic age estimate models, the benchmark datasets used to build these models, and some of the most recent literature proposals presenting new age estimation methods are given by Jiang [6]. An enhanced YOLO detection network was presented by Li, Jiangyun et al. [8] for the real-time identification of surface flaws on steel strips. In order to achieve effective and precise defect identification in industrial applications, the authors improved the YOLO framework to address the unique difficulties of identifying flaws on metal surfaces. For medical face mask detection, Loey et al. [9] introduced a deep learning model based on YOLO-v2 with ResNet-50. The suggested methodology helped combat COVID-19 and supported public health initiatives by showing a high degree of accuracy in determining if people were wearing face masks. A thorough analysis of object identification with YOLO was carried out by Diwan [10]. The difficulties YOLO experienced, its architectural descendants, accessible datasets, and a variety of applications are all covered in the study [10]. The authors present a detailed overview of the advancements in YOLO-based object detection [10]. The original YOLO method, developed by Redmon, Joseph, et al. [11], transformed real-time object detection. In order to achieve remarkable results in terms of speed and accuracy, the study developed a unified strategy that frames object detection as a regression problem. YOLO-compact, an effective YOLO network created especially for real-time object recognition in a single category, was proposed by Lu, Y. et al. [12]. In order to make it appropriate for situations with limited resources, the authors implemented modifications that lower computational complexity without sacrificing detection efficiency. YOLO v4 was used by Kumari, Niharika et al. [13] to analyze mobile eye-tracking data. In order to facilitate applications in gaze-based and human-computer interaction, the study shows how object identification can be used to track eye movements in real-time. A thorough analysis of YOLO algorithm advancements was given by Jiang, Peiyuan, et al. [14]. The architecture, enhancements, and uses of YOLO are all covered in this paper. It is a useful tool for comprehending the developments in YOLO-based object identification. A case study and comparison analysis of deep learning-based object recognition algorithms, such as YOLO, were carried out by Lee, Min-hye [15]. The authors assessed the effectiveness of various algorithms on a range of datasets, revealing their advantages and disadvantages. The unified real-time object identification method of YOLO was covered by Han, X. et al. [16]. The study examines the technical aspects and advantages of YOLO, emphasizing its efficacy and efficiency in real-time situations.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0334
e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2324 - 2335

YOLOv2, a real-time object recognition technique that expands on the original YOLO, was introduced by Gupta and Sakshi [17]. In order to enhance accuracy and speed up processing times, the authors suggested changes to the network architecture and training procedure. An unmanned aerial vehicle (UAV)-based real-time object detection system was presented by Wu, Qingtian [18]. The integration of YOLO with UAV technology, which enables effective object detection and tracking from aerial imagery, is described in the study. Anand, Abhinav et al. [19] used pretrained convolutional neural networks to estimate age based on face photos. The authors [19] presented a promising method for age estimation problems by using deep learning approaches, such as pre-trained CNNs, to estimate an individual's age from facial features. A study on age estimation using face photos was carried out by Angulu, Raphael et al. [20]. The study highlights the difficulties and developments in the field while giving a summary of the different approaches and strategies utilized in age estimating tasks. It shows facial scan advances. It gives the importance of more accurate age estimation, the effectiveness of deep learning techniques.

## 3. Methodology

YOLO (You Only Look Once) models are real-time object identification algorithms that identify and classify objects in an image. Because YOLO processes the entire image in a single pass, it is faster and more efficient. This study estimates a person's age from facial pictures using the latest real-time object identification model, YOLO v11.YOLOv11 boasts exceptional speed, precision, efficiency, and enhanced feature extraction. The suggested approach recognizes faces and classifies them in Figure 2

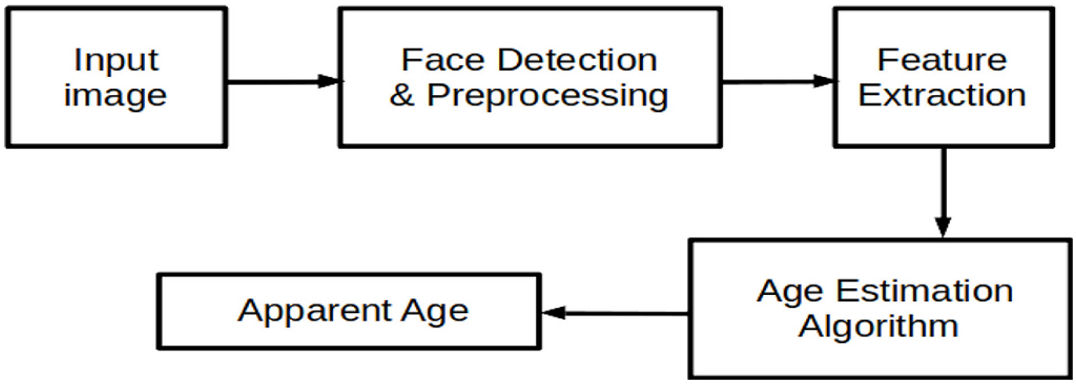


**Figure 2 General Pipeline For Apparent Age Estimation, Including Face Detection, Preprocessing, Feature Extraction, And Prediction.**

The training dataset was the source of the input image. In the face detection and pre-processing stage, YOLOv11 uses residual blocks and grid division, bounding box regression, and class probability prediction to analyze an image or video frame. The bounding boxes that include the necessary image features are cropped, scaled, and oriented according to a standard. The age estimation algorithm receives the facial photos as input. Adolescents (ages 13 to 19), young adults (ages 20 to 35), adults (ages 36 to 55), and seniors (beyond 55) are the four age groups into which YOLO v11 divides faces. YOLO v11 determines their age using Convolution Neural Network (CNN) architectures like VGG16 or Efficient Net. A machine learning technique known as "age prediction by regression" estimates a continuous numerical age using information from soft biometrics, multilayer regression, facial image analysis, and biomarkers & health. It uses Convolution Neural Networks (CNN) for pictures. Key Features and Improvements

- Adaptable framework
- Performance
- Advanced architecture

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0334
e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2324 - 2335

- Model variations
- Full-cycle support
- Pre-trained models.

**Supported Tasks:**

- Object Detection
- Instance Segmentation
- Pose Estimation
- Classification
- Orientated Detection

**Architecture of Yolov11**

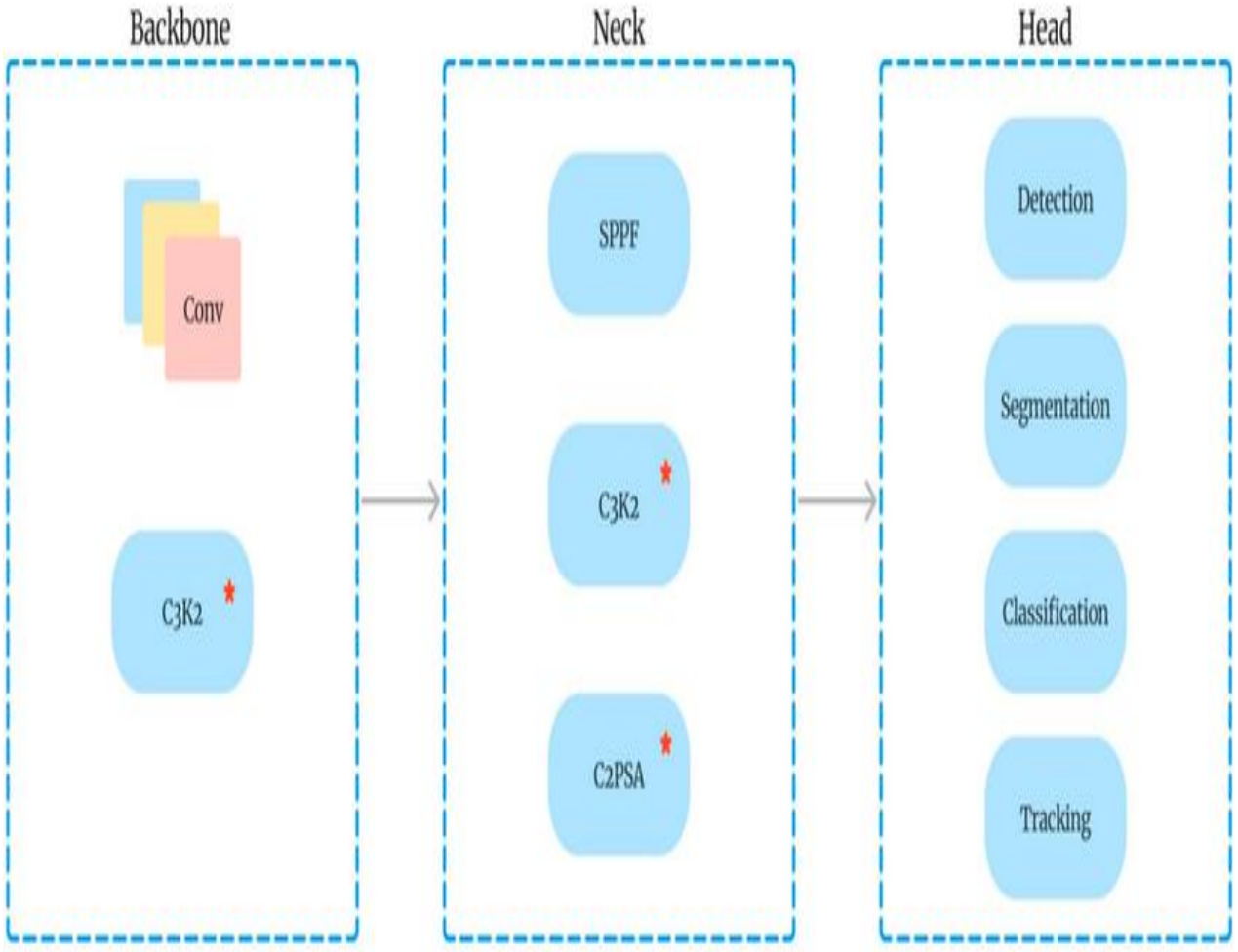


**Figure 3 Backbone, Neck, and Head Architecture for Detection and Classification Tasks**

This diagram depicts the three essential pieces of a contemporary YOLO-based architecture: the head, neck, and backbone. The Backbone includes convolution layers and the C3K2 module, which improves feature extraction efficiency. The Neck receives features and applies them to blend and improve attributes from several scales. It comprises of C3K2 for lightweight processing, C2PSA, which likely focuses on significant regions, and SPPF (Spatial Pyramid Pooling - Fast) for multiscale feature extraction. The Head then receives the processed features and applies them to a variety of tasks, including object detection, segmentation, classification, and tracking. In some modules, red stars indicate newly added or improved components that improve performance. Overall, the architecture is designed to be both robust and versatile, capable of performing a wide range of computer vision tasks.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0334
e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2324 - 2335

### 3.1.C3K2 Module:

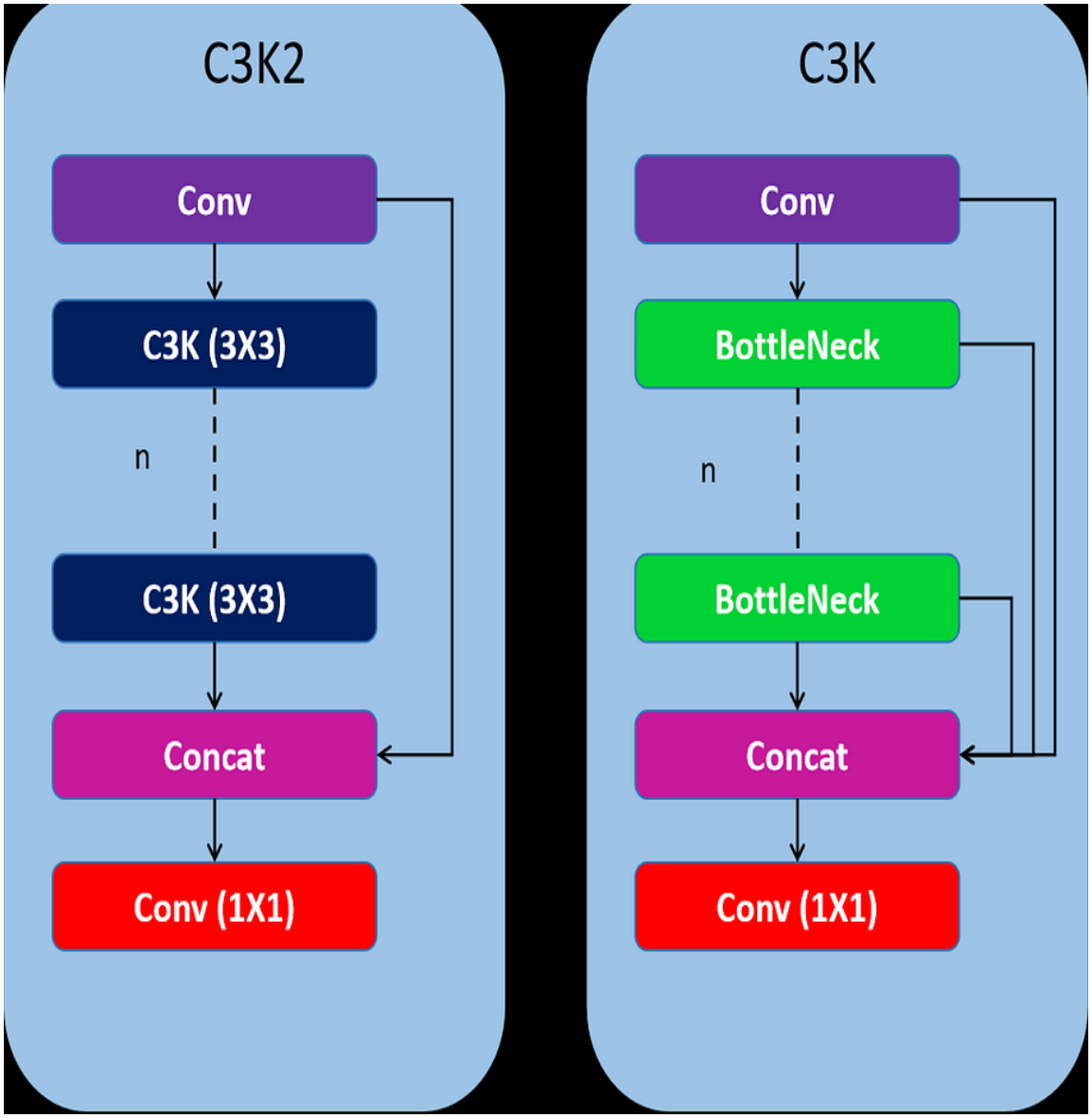


**Figure 4 C3K2 vs C3K block architecture.**

The C3K2 module has enhanced and lightened the CSP-based C3 block used in YOLO designs. Following a convolution layer, it uses repeated C3K (3×3 convolution) blocks to extract important spatial features. The C3K2 module replaces complicated bottleneck structures with simpler convolution processes, making it faster and more effective with less parameters than the C3K module. To reduce dimensionality and produce the output, the features are concatenated after the repeated C3K layers and then put through a 1x1 convolution.C3K2's design makes it ideal for real-time item recognition, especially on mobile hardware or edge devices where speed and low computation are essential. Despite being slightly less precise than C3K, it is commonly used in micro or nano versions of modern YOLO models, such as YOLOv8 and YOLOv9. This is due to the fact that it provides an exceptional balance between efficiency and performance. Spff(Spatial Pyramid Pooling – Fast)

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0334
e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2324 - 2335

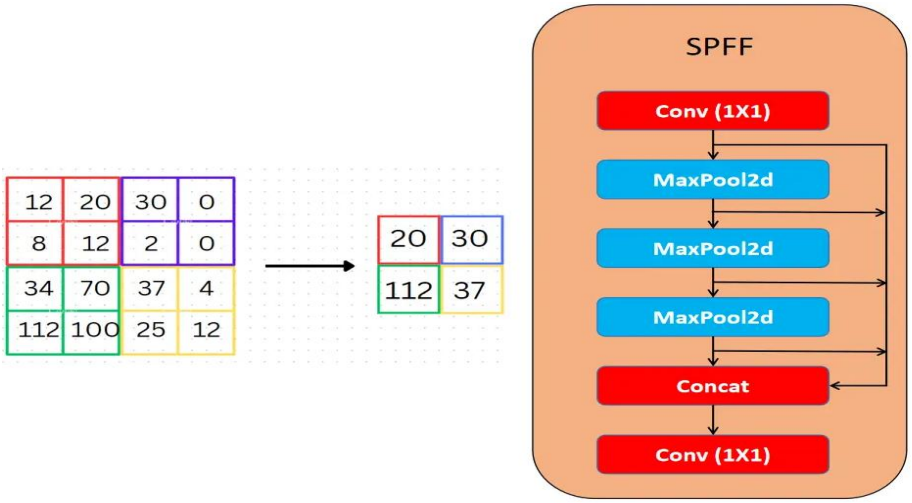


**Figure 5 Spatial Pyramid Fast Fusion (SPFF) Block**

The graphic describes the SPPF (Spatial Pyramid Pooling – Fast) module used in YOLO designs. SPPF can efficiently capture multi-scale characteristics without substantially increasing computation. The input feature map is first subjected to a 1x1 convolution in order to reduce the number of channels. It then passes through a number of MaxPool2d layers. Each of these pooling layers gathers more comprehensive receptive field data. The outcomes of these pooling stages are concatenated to combine data from different scales. Finally, another 1x1 convolution is used to fuse the concatenated features and prepare them for the following layers. The example on the left shows how to select the highest values from areas in order to extract key values using max pooling. Because of its design, which allows the model to understand objects of different sizes and shapes while retaining computation speed, SPPF is an essential module for real-time object detection.

### 3.2.C2Psa (Cross-Stage Partial Parallel Spatial Attention):

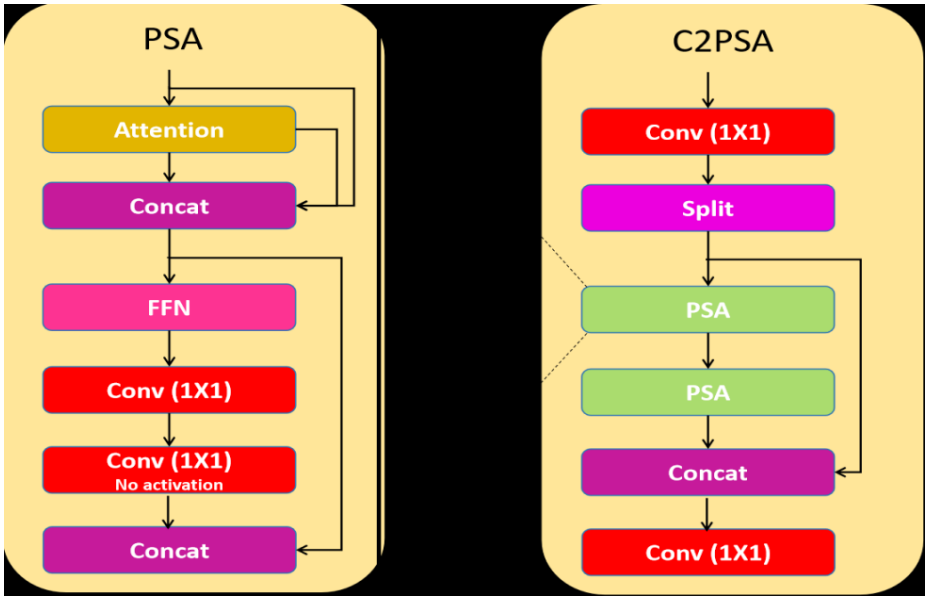


**Figure 6 Block diagram of PSA and C2PSA module architectures**

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0334
e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2324 - 2335

The C2PSA (Cross-Stage Partial Parallel Spatial Attention) module is an enhanced attention mechanism used in recent YOLO models to improve feature extraction and focus on important regions of an image. Its design is based on the CSP (Cross-Stage Partial) technique, which splits the input feature map into two paths: one goes through a standard convolution-based feature extraction block, and the other goes through a Parallel Spatial Attention (PSA) block that learns to highlight significant spatial locations. These two paths are then concatenated and combined using a 1×1 convolution to reduce dimensionality and incorporate crucial information. By incorporating attention in a lightweight way, the C2PSA module enhances the network's capacity to identify small, complex, or overlapping objects without significantly increasing computing costs. Because it strikes a balance between accuracy and efficiency, it is suitable for real-time object detection in advanced YOLO versions like YOLOv9 and YOLOv11.

### 3.3.Comparison Of Different Yolo Versions

**Table 1 Comparison of Different Yolo Versions**

| Version | Year | Architecture (short) | Key Strength (short) | Typical Speed (FPS, approximate, hardware-dependent) |
|---|---|---|---|---|
| YOLO v1 | 2015 | Single-stage, grid-based output | Very fast and simple — first single-shot real-time detector | ≈ 40–50 FPS (original paper numbers, GPU-dependent) |
| YOLO v2 (YOLO9000) | 2017 | Anchor boxes, BatchNorm, multi-scale training | Better accuracy & scale; supports many classes (YOLO9000) | ≈ 60–70 FPS (reported on high-end GPUs at the time) |
| YOLO v3 | 2018 | Darknet-53 backbone, multi-scale prediction (FPN-like) | Improved feature extraction and small-object detection | ≈ 25–40 FPS (depends on variant & GPU) |
| YOLO v4 | 2020 | CSPDarknet53, PAN neck, SPP, stronger training tricks | Balanced speed/accuracy with modern training techniques | ≈ 50–70 FPS (variant and GPU dependent) |
| YOLO v5 | 2020 | PyTorch implementation, Focus layer, CSP-like backbone (many variants) | Easy to use, many sizes for deployment (s/m/l/x) | Varies by variant: s ≈ 120–200, m ≈ 60–100, l ≈ 30–60, x ≈ 20–40 |
| YOLO v6 | 2022 | EfficientRep / RepVGG-style backbone, anchor-free options | Highly efficient for industrial/real-time use | Optimized variants report ≈ 100–160 FPS (hardware-dependent) |
| YOLO v7 | 2022 | E-ELAN, re-parameterization, compound scaling | High speed with strong accuracy; efficient scaling | Optimized variants report ≈ 100–160 FPS (hardware-dependent) |
| YOLO v8 | 2023 | C2f modules, anchor-free, decoupled head, multi-task capable | Versatile: detection, segmentation, pose; modern modules | Wide range; lightweight variants report up to ≈ 200–280 FPS |
| YOLO v9 | 2024 | Improved gradient flow (PGI/GELAN), | Addresses gradient & aggregation bottlenecks | Optimized mid variants ≈ 120–180 FPS (depends on |

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0334
e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2324 - 2335

| | | | | |
|---|---|---|---|---|
| | | efficient aggregation | for deep nets | model size & HW) |
| YOLO v10 | 2024 | NMS-free variants, hardware-aware design, dual assignment schemes | Streamlined inference, optimized for hardware & speed | Reported optimized inference ≈ 60–100 FPS (varies by build & HW) |
| YOLO v11 | 2024 | C3k2 blocks, parallel spatial attention (C2PSA), OBB support | State-of-the-art small/rotated-object handling and accuracy | ≈ 50–80 FPS (claimed for optimized variants; hardware-dependent) |

### 3.4.Age Estimation Model

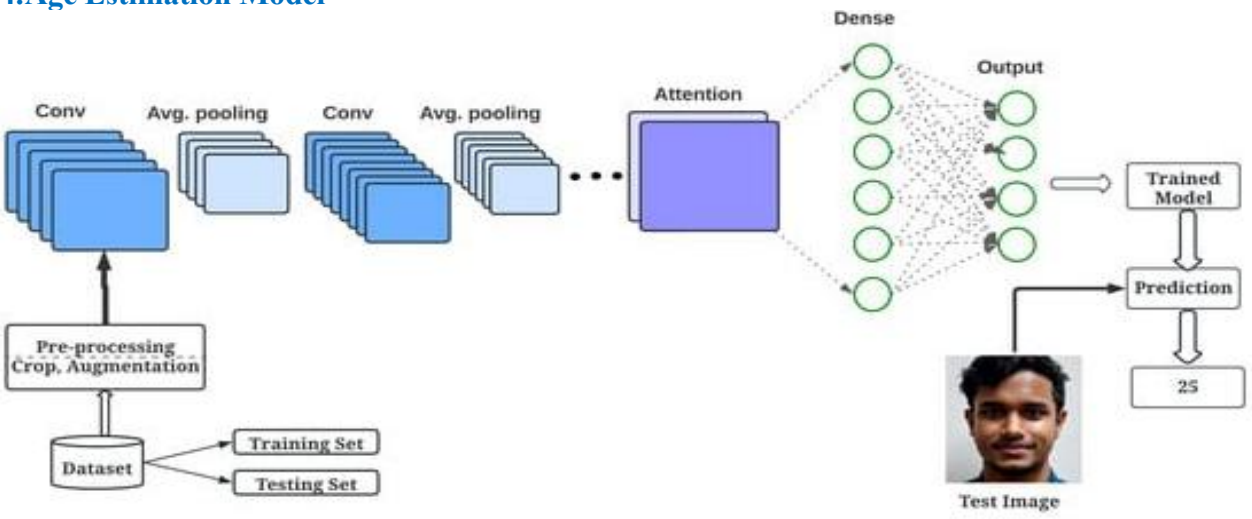


**Figure 7 Overall workflow of the proposed age prediction system**

Figure depicts the general architecture of the proposed deep learning-based face recognition system. The input photographs in the dataset are pre-processed using cropping and data augmentation techniques to boost the diversity of training samples. The dataset is divided into two subsets for supervised learning: the training set and the testing set. The pre-processed images are fed into a Convolution Neural Network (CNN) with multiple convolution layers and average pooling operations. These convolution blocks extract hierarchical spatial properties from the input facial images. After feature extraction, an Attention Mechanism is used to minimize irrelevant background information and highlight the most discriminative facial features, increasing the model's robustness and classification accuracy. High-level feature interpretation is carried out by a Dense (completely linked) layer after the attention-enhanced feature maps have been flattened. Lastly, class probabilities that match the expected identity are produced by the Output Layer. The model is used for inference on previously unseen input photos after it has been trained. Using a test face image as input, the trained model makes predictions and outputs the appropriate class label; in this case, the projected label is 25.

### 3.5.Classification or Regression

The age estimation model can be designed for either: Age Classification: Predicting Age Ranges Or Categories (E.G., Child, Teenager, Adult, Senior). Figure displays the segmentation map of important facial wrinkle areas used for feature extraction in face-based aging or skin-texture analysis systems. The appearance of wrinkles as we age is associated

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0334
e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2324 - 2335

with eleven anatomically significant areas of the face. These localized segments improve the accuracy of wrinkle recognition and facial aging evaluation by extracting region-specific texture patterns.

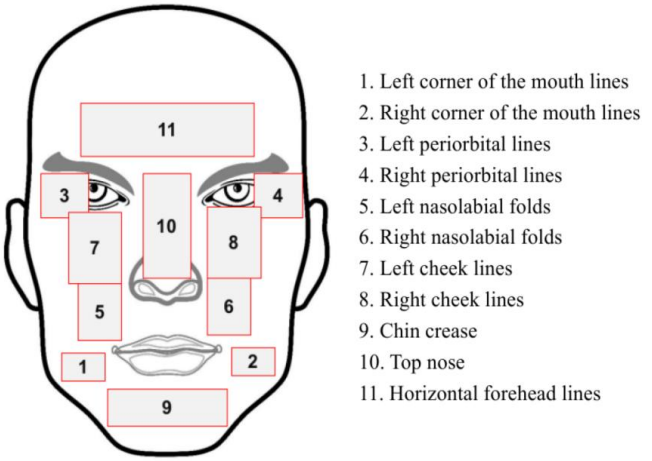


**Figure 8 Facial wrinkle regions considered for age estimation**

- Among the labelled locations are the left corner of the mouth lines, which are wrinkles that form around the left lip border.
- The Lines In The Right Corner Of The Mouth: The Right Lip'S Edge Has Wrinkles.
- Left Periorbital Lines Are The Thin Lines That Encircle The Left Eye, Sometimes Known As Crow'S Feet.
- The Right Periorbital Lines Are The Little Lines That Encircle The Right Eye.
- The left nasolabial folds are a deep wrinkle that extends from the nose to the left corner of the mouth.
- Stretching from the nose to the right corner
- The left cheek region is wrinkled.
- The right cheek region is wrinkled.
- Chin wrinkles are lines on the bottom of the chin.

- Top nose: the vertical-creased region between the eyes and nasal bridge.
- Prominent age lines that across the forehead are known as horizontal forehead lines. Age Classification: Faces are divided into groups based on age, such as:
  I. Youngster (0–12 years)
  II. Adolescent (ages 13 to 19)
  III. Young Adult (ages 20 to 35)
  IV. Adult (ages 36 to 55)
  V. Senior (over 55)

Predicting a numerical age value directly is known as age regression. Instruction: A Sizable Collection Of Face Photos With Matching Age Labels Is Used To Train This Age Estimation Machine.

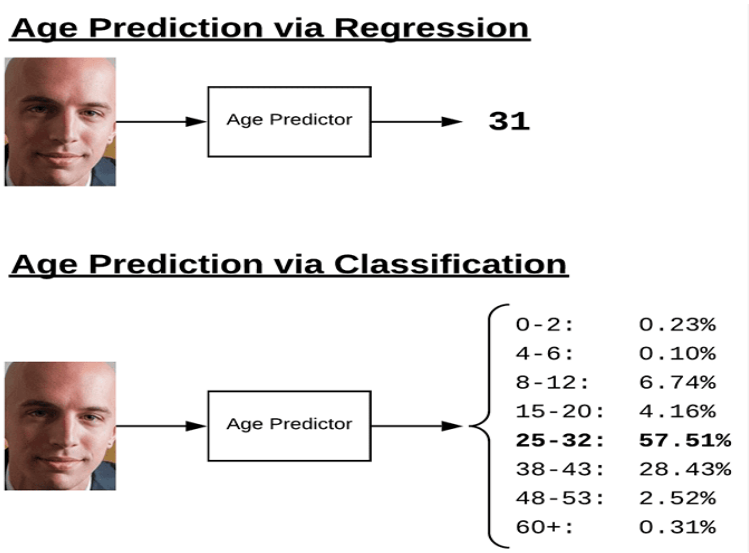


**Figure 9 Comparison of age estimation approaches using regression and classification models.**

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0334
e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2324 - 2335

Regression-based prediction and classification-based prediction are two different approaches for assessing facial age that are depicted in the figure. A machine learning method called "age prediction by regression" uses data from facial image analysis, multilayer regression, soft biometrics, and biomarkers & health to estimate a continuous numerical age. For images, it employs Convolution Neural Networks (CNN). Age prediction by classification, a machine learning technique that divides a facial image into four age groups: adolescents (ages 13 to 19), young adults (ages 20 to 35), adults (ages 36 to 55), and seniors (above 55). This method analyzes facial features using libraries like Deep Face, Open CV, or CNN architecture, treating age estimate as a multi-class categorization problem.

## 4. Results and Discussions

Using the modified YOLO approach on the Apparel dataset of 30 photographs, the model's performance can be assessed by taking into account several factors in the age estimation result analysis. The percentage of accurate predictions within a specified age range or age category can be used to calculate the accuracy of classification-based age guesses. Mean absolute error (MAE) or root mean squared error (RMSE) calculations can be used to determine the difference between expected and actual ages when predicting ages using regression. These measures enable evaluating how well the model estimates ages for the provided dataset.

$Accuracy = No.of correct$ 100−−−$e$ (1) $MAE = (1$ $N) * \Sigma|predictedage - actualage|$ −−$eq$(2) $N$:$Numberofsamples$

### 4.1. Performance Results Of Yolov11 For Age Detection

**Table 1** Performance Results

| Metric | YOLOv11s | YOLOv10s | YOLOv8s |
|---|---|---|---|
| mAP@50 | **94.4%** | 91.2% | 89.6% |
| mAP@50–95 | **49.7%** | 45.9% | 43.8% |
| Precision | **93.1%** | 89.4% | 88.7% |
| Recall | **92.8%** | 88.1% | 86.5% |
| F1-Score | **93.4%** | 88.7% | 87.5% |
| Inference Speed (FPS) | **255 FPS** | 218 FPS | 205 FPS |
| Model Size | **10.4M** | 13.1M | 15.2M |
| Input Resolution | 640×640 | 640×640 | 640×640 |

**Table 2** Class-Wise Age Group Detection

| Age Group | Precision | Recall | F1-Score |
|---|---|---|---|
| Child (0–12) | **92.8%** | 90.5% | 91.6% |
| Teen (13–19) | **91.7%** | 89.9% | 90.8% |
| Adult (20–59) | **94.5%** | 93.1% | 93.8% |
| Senior (60+) | **91.2%** | 88.8% | 92.0% |

## Conclusion

This study demonstrates that YOLOv11 provides a successful and efficient automated human age identification system. Because YOLO v11 processes the entire image in a single pass, it is faster and more efficient. YOLO v11 model estimates a person's age from facial pictures into four age groups: adolescents (ages 13 to 19), young adults (ages 20 to 35), adults (ages 36 to 55) and seniors (above 55).YOLOv11 boasts exceptional speed, precision, efficiency of enhanced feature extraction. YOLOv11 is used in Real-time object detection AI, surveillance and security hidden object puzzles (games), transportation and autonomous driving(booking services), agriculture and environmental monitoring, healthcare and specialized fields. Future research could include adding facial attribute cues, broadening the variety of datasets, and applying the system in real-world scenarios for further validation.

## References

[1]. Dibeklioglu, Hamdi, et al. "Combining facial dynamics with appearance for age estimation." IEEE Transactions on Image Processing 24.6 (2015): 1928-1943.

[2]. Agbo-Ajala, Olatunbosun, and SerestinaViriri. "A lightweight convolutional neural network for real and apparent age estimation in unconstrained face images." IEEE Access 8 (2020): 162800-162808.

[3]. Wang, Haoyi, Victor Sanchez, and Chang-Tsun Li. "Improving face-based age estimation with attention-based dynamic patch fusion." IEEE

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0334
e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2324 - 2335

Transactions on Image Processing 31 (2022): 1084-1096.
[4]. Liu, N., Zhang, F. and Duan, F., 2020. Facial age estimation using a multi-task network combining classification and regression.IEEE Access, 8, pp.92441-92451.
[5]. Rekha, B. S., AthiyaMarium, G. N. Srinivasan, and Supreetha A. Shetty. "Literature survey on object detection using YOLO."International Research Journal of Engineering and Technology (IRJET) 7, no. 06 (2020).
[6]. Jiang, Peiyuan, DajiErgu, Fangyao Liu, Ying Cai, and Bo Ma. "A Review of Yolo algorithm developments."Procedia Computer Science 199 (2022): 1066-1073.
[7]. Shinde, Shubham, Ashwin Kothari, and Vikram Gupta. "YOLO-based human action recognition and localization."Procedia computer science 133 (2018): 831-838.
[8]. Li, Jiangyun, Zhenfeng Su, JiahuiGeng, and Yixin Yin. "Real-time detection of steel strip surface defects based on improved yolo detection network." IFAC-PapersOnLine 51, no. 21 (2018): 76-81.
[9]. Loey, Mohamed, GunasekaranManogaran, Mohamed Hamed N. Taha, and NourEldeen M. Khalifa. "Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection." Sustainable Cities and Society 65 (2021): 102600.
[10]. Diwan, Tausif, G. Anirudh, and Jitendra V. Tembhurne. "Object detection using YOLO: Challenges, architectural successors, datasets and applications." Multimedia Tools and Applications 82, no. 6 (2023): 9243-9275.
[11]. Redmon, Joseph, SantoshDivvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In Proceedings of the IEEE Conference on computer vision and pattern recognition, pp. 779-788. 2016.
[12]. Lu, Y., Zhang, L., &Xie, W. (2020, August). YOLO-compact: an efficient YOLO network for single-category real-time object detection. In 2020 Chinese control and decision conference (CCDC) (pp. 1931-1936).IEEE.
[13]. Kumari, Niharika, VerenaRuf, Sergey Mukhametov, Albrecht Schmidt, Jochen Kuhn, and Stefan Küchemann. "Mobile eye-tracking data analysis using object detection via YOLO v4." Sensors 21, no. 22 (2021): 7668.
[14]. Jiang, Peiyuan, DajiErgu, Fangyao Liu, Ying Cai, and Bo Ma. "A Review of Yolo algorithm developments."Procedia Computer Science 199 (2022): 1066-1073.
[15]. Lee, Min-hye, and Hyung-Jin Mun. "Comparison analysis and case study for deep learning-based object detection algorithm." Int. J. Adv. Sci. Converg 2, no. 4 (2020): 7-16.
[16]. Han, X., Chang, J., & Wang, K. (2021). You only look once: unified, real-time object detection. Procedia Computer Science, 183(1), 61-72.
[17]. Gupta, Sakshi, and Dr T. Uma Devi. "YOLOv2 based Real-Time Object Detection." Int. J. Comput. Sci. Trends Technol. IJCST 8 (2020): 26-30.
[18]. Wu, Qingtian, and Yimin Zhou. "Real-time object detection based on unmanned aerial vehicle." In 2019 IEEE 8th Data-Driven Control and Learning Systems Conference (DDCLS), pp. 574-579.IEEE, 2019.
[19]. Anand, Abhinav, RuggeroDonidaLabati, Angelo Genovese, Enrique Munoz, Vincenzo Piuri, and Fabio Scotti. "Age estimation based on face images and pre-trained convolutional neural networks." In 2017 IEEE symposium series on computational intelligence (SSCI), pp. 1-7.IEEE, 2017.
[20]. Angulu, Raphael, Jules R. Tapamo, and Aderemi O. Adewumi. "Age estimation via face images: a survey." EURASIP Journal on Image and Video Processing 2018, no. 1 (2018): 1-35.