Dent.py: Training Dental Models from Zero to Hero

YOLO: You Only Look Once – Deep Dive

Title: YOLO: You Only Look Once - A Deep Dive

Slide 4: YOLO Network Architecture

- Input and feature extraction (CNN backbone)
- Prediction heads (bounding boxes, objectness, class probabilities)
- Grid-based detection

Slide 5: YOLO Evolution

- YOLOv1: Introduction
- YOLOv2: Better accuracy, YOLO9000
- YOLOv3: Multi-scale detection
- YOLOv4: Optimizations for speed and accuracy
- YOLOv5: PyTorch implementation, widely used
- YOLOv6, YOLOv7: Faster and more accurate
- YOLOv8: Latest improvements

Slide 6: Applications of YOLO

- Autonomous vehicles
- Security and surveillance
- Medical imaging
- Retail and logistics

Slide 7: Advantages and Limitations

- Pros: Fast, real-time capable, end-to-end training
- Cons: Struggles with small objects, trade-off between speed and accuracy

Slide 8: Hands-on with YOLO (Colab Demo)

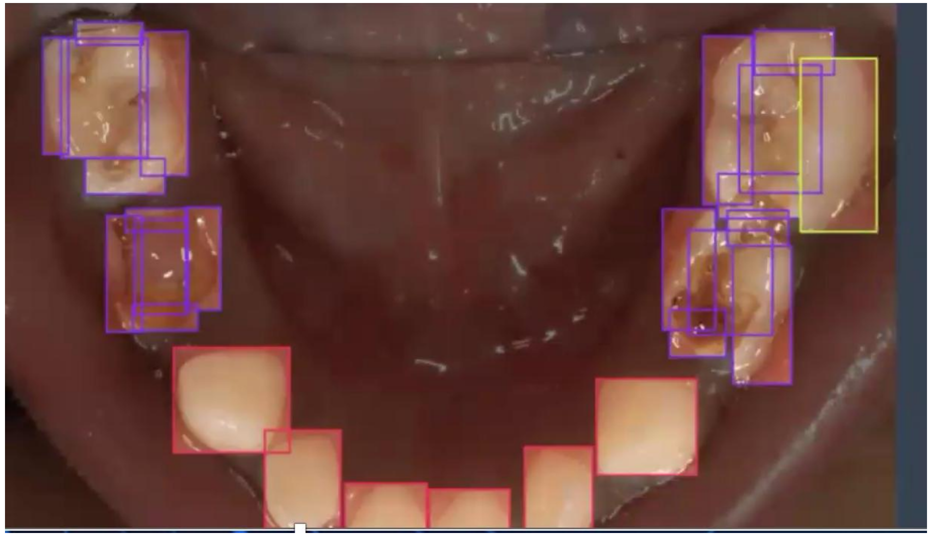
- Install YOLO using Ultralytics library
- Load a pre-trained model
- Run inference on an image
- (Optional) Train on a custom dataset

Slide 9: Summary and Q&A

- Recap of key points
- Open discussion

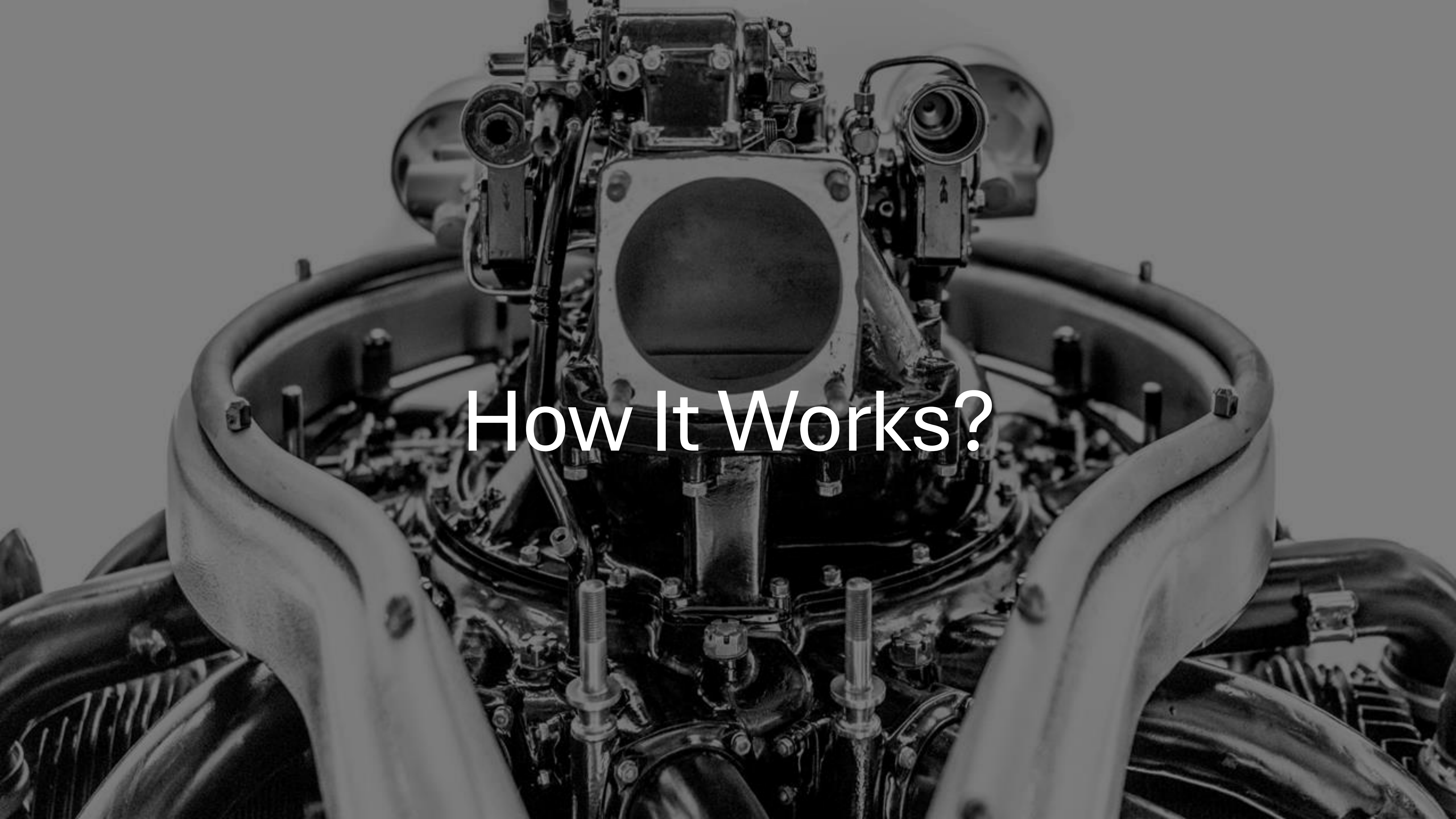
Object Detection

- What is object detection?
  - Identifying and locating objects in an image.
  - Used in various applications like autonomous driving, security, retail, etc...



YOLO

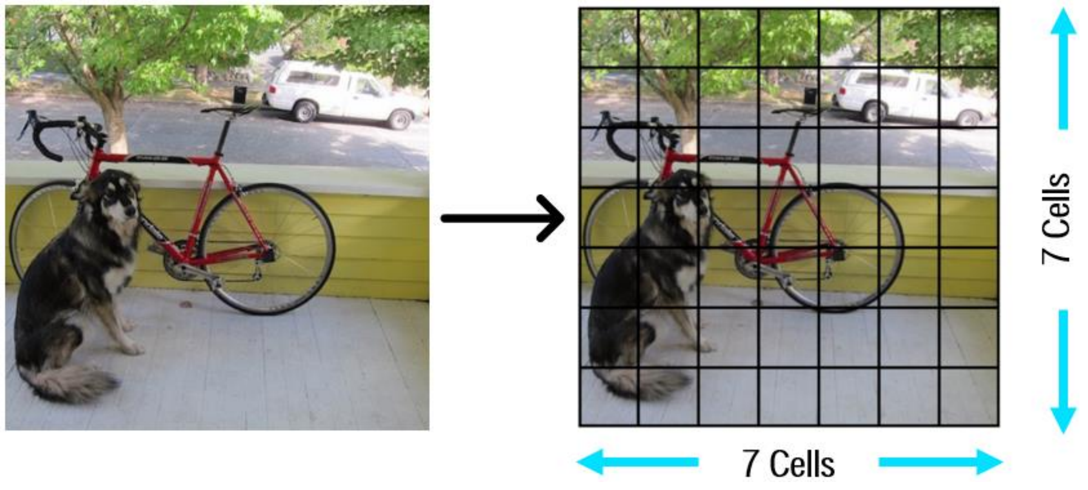
- YOLO
  - You Only Look Once
- Overview
  - A fast and efficient object detection algorithm.
  - Processes an image in a single pass, unlike traditional methods.
- Why YOLO is important
  - Real-time processing capability.
  - High accuracy and efficiency.



How It Works?

Grid-based prediction

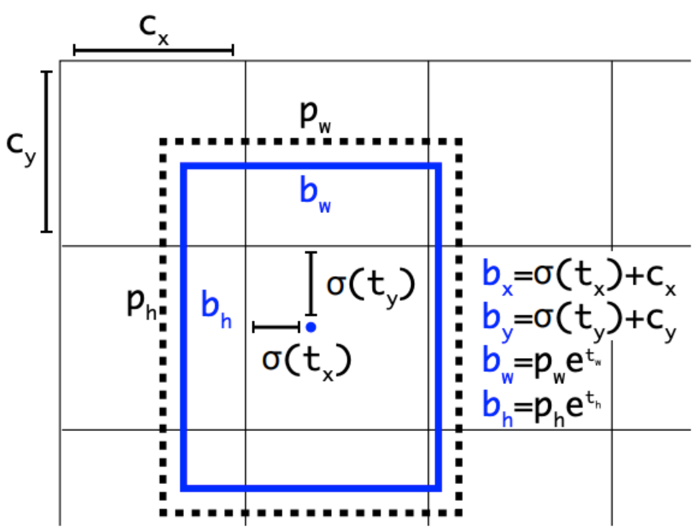
- The image is divided into an S × S grid (e.g., 7×7, 19×19).
- Each grid cell predicts bounding boxes, confidence scores, and class probabilities.
- A grid cell is responsible for detecting objects whose center falls within it.

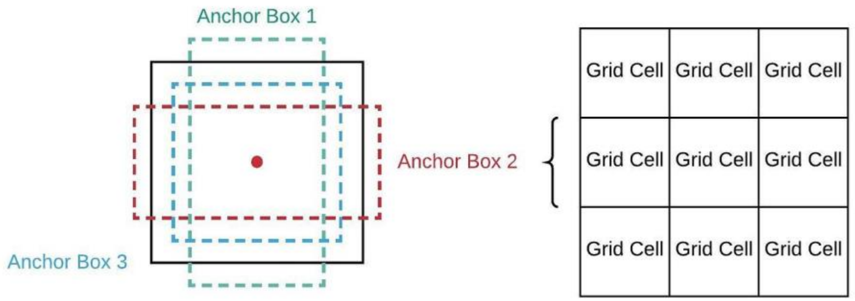


Yajnik, A. (2023, September 7). *Computer vision: Yolo: Grid cells and anchor boxes*. Medium. https://medium.com/@ayushyajnik2/computer-vision-yolo-grid-cells-and-anchor-boxes-57b8a33cb25b

Bounding box prediction

- Each grid cell predicts multiple bounding boxes with (x, y, w, h) coordinates.
- Predefined boxes of different aspect ratios and sizes improve detection.
- Helps in detecting objects of various scales and shapes.
- Each grid cell predicts adjustments to these anchor boxes rather than free-form bounding boxes.

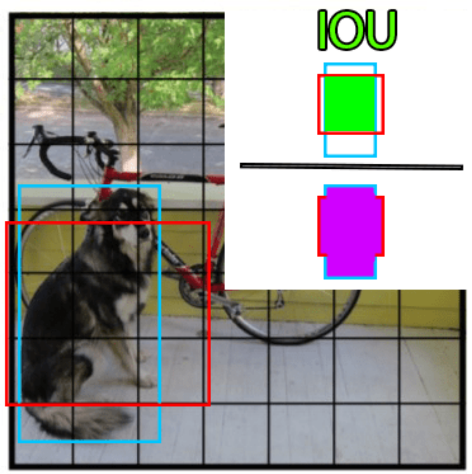




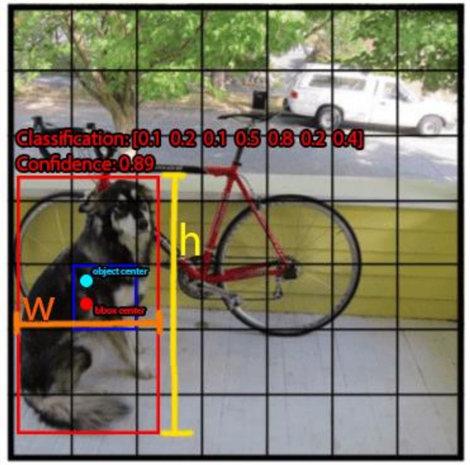
Li, E. Y. (2023, June 15). *Dive really deep into Yolo V3: A beginner's guide*. Medium. https://medium.com/data-science/dive-really-deep-into-yolo-v3-a-beginners-guide-9e3d2666280e

Confidence Scores

- Represents how likely an object is present and the accuracy of the bounding box.
- Confidence score = Object probability × IoU (Intersection over Union).
- Higher confidence scores indicate more reliable predictions.



S × S grid on input

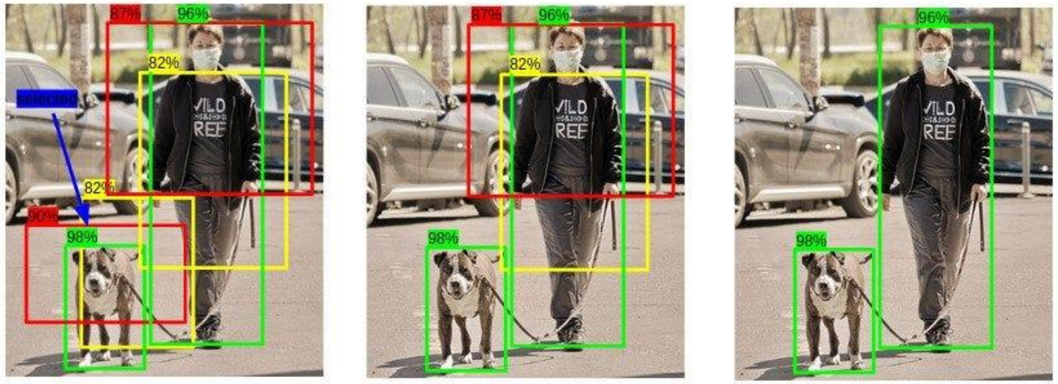


S × S grid on input

Yolo for Object Detection. GUVI Geeks Network. (n.d.-b). https://forum.guvi.in/posts/6991/yolo-for-object-detection

Non-Maximum Suppression

- Removes overlapping boxes to keep only the most relevant detections.
- Ensures that the best bounding box for each object is retained.



Step 1: Selecting Bounding box with highest score

Step 3: Delete Bounding box with high overlap

Step 5: Final Output

kamal_DS. (2023, March 14). *Non-max suppression*. Medium. https://korlakuntasaikamal10.medium.com/non-max-suppression-b0569e89e9e7

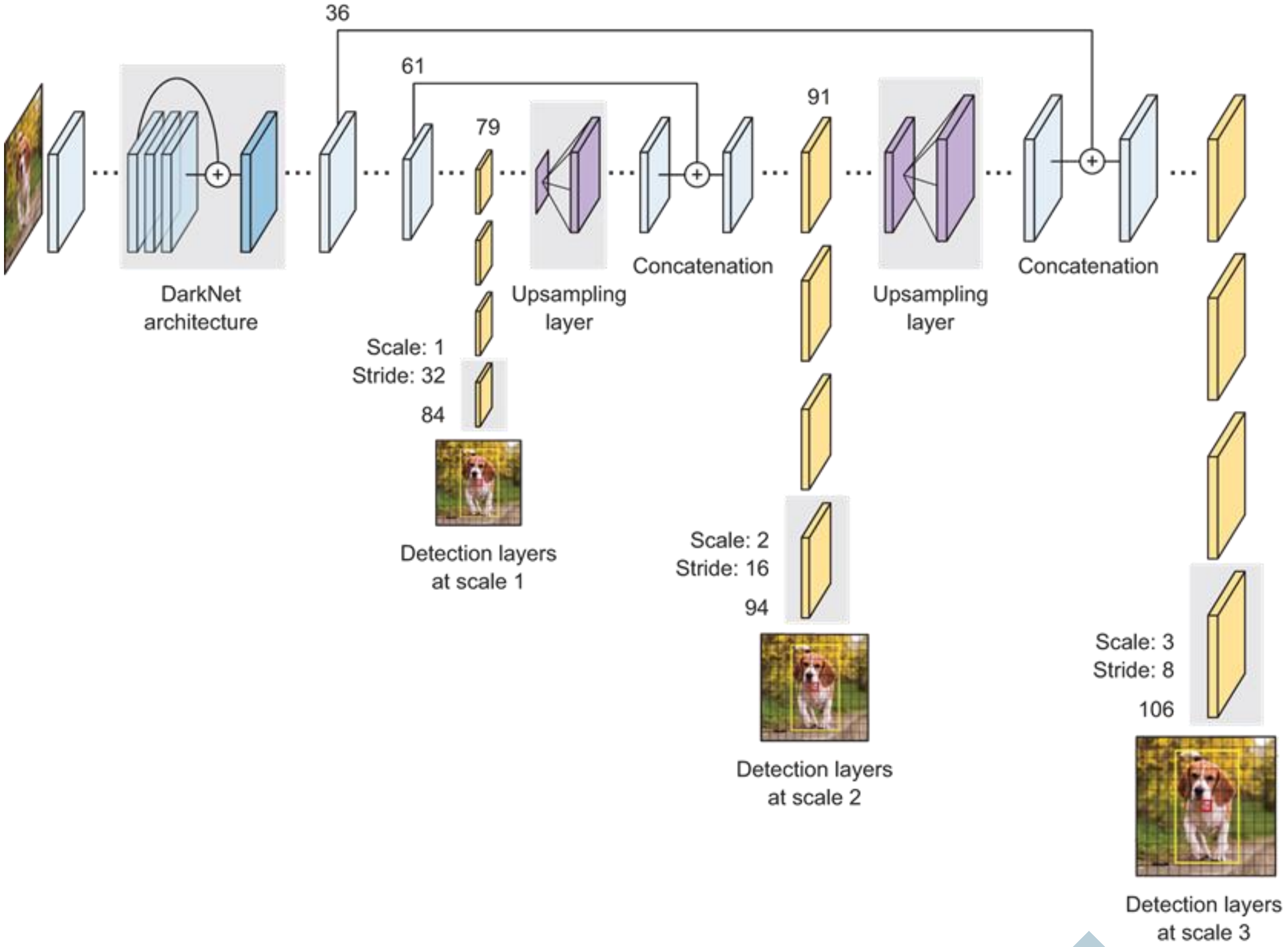

Architecture

Overview

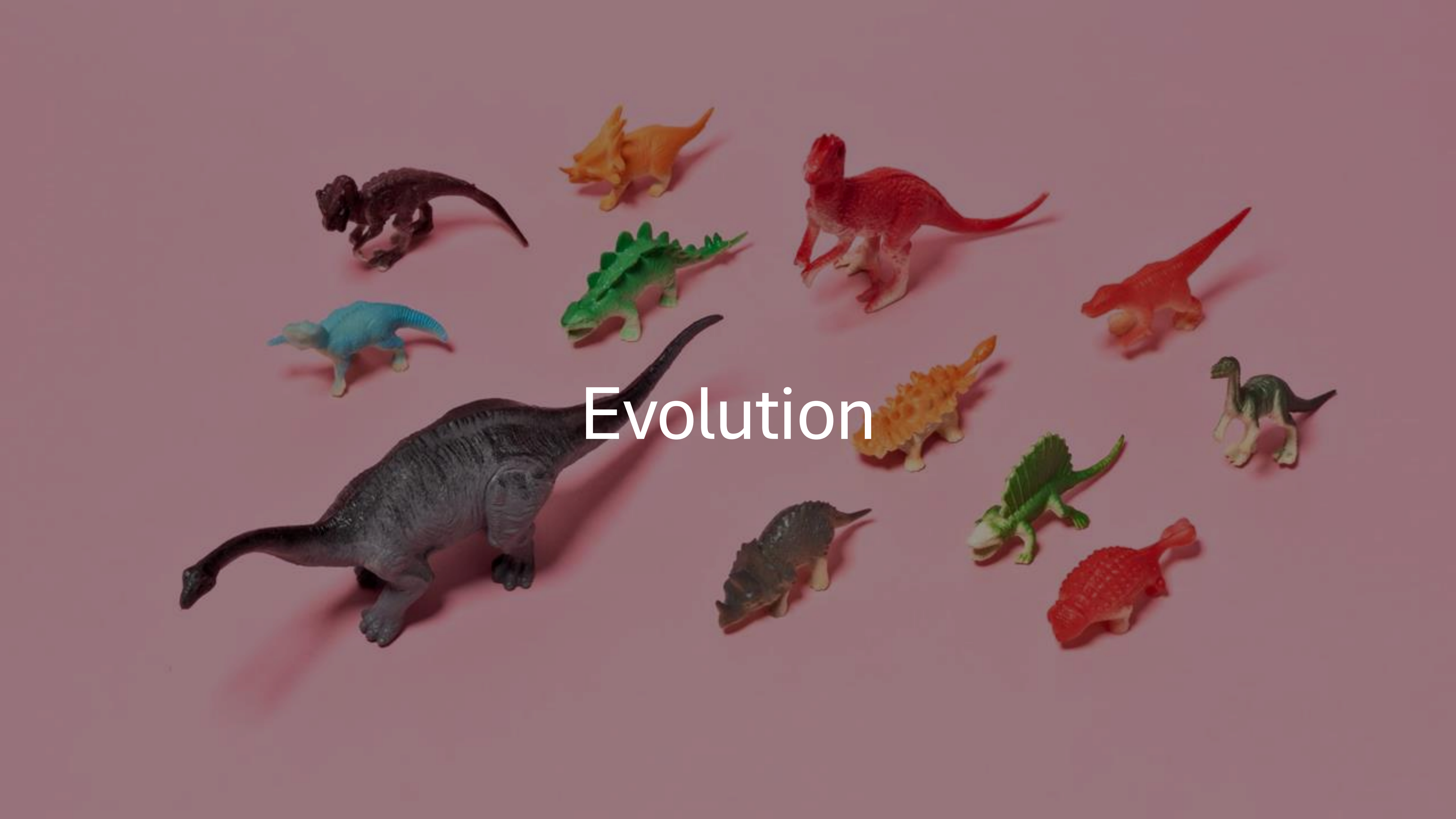
- Input and Feature Extraction
  - Uses a CNN backbone (e.g., Darknet-53 in YOLOv3) to extract hierarchical features.
  - Processes input images through convolutional layers with increasing depth.
- Detection Head
  - Splits into multiple scales for multi-scale detection (FPN-like structure in YOLOv3).
  - Outputs bounding box coordinates, objectness score, and class probabilities.

Building Blocks

- Convolutional Layers
  - Extract spatial features from images.
- Batch Normalization
  - Speeds up training and stabilizes learning.
- Residual Connections
  - Help gradient flow and improve learning efficiency.
- Leaky ReLU Activation
  - Prevents vanishing gradients and speeds up training.


36
61
79
91
DarkNet architecture
Upsampling layer
Concatenation
Upsampling layer
Concatenation
Scale: 1
Stride: 32
84
Detection layers at scale 1
Scale: 2
Stride: 16
94
Detection layers at scale 2
Scale: 3
Stride: 8
106
Detection layers at scale 3

Evolution



YOLO Greatest Hits

Jun 8, 2015
YOLO is introduced
Joseph Redmon et al / UW
[You Only Look Once: Unified, Real-Time Object Detection]

Dec 25, 2016
YOLOv2 aka YOLO9000
Joseph Redmon et al / UW
[YOLO9000: Better, Faster, Stronger]

Apr 8, 2018
YOLOv3
Joseph Redmon et al / UW
[YOLOv3: An Incremental Improvement]

Apr 23, 2020
YOLOv4
Alexey Bochkovskiy et al
[YOLOv4: Optimal Speed and Accuracy of Object Detection]

Jun 9, 2020†
YOLOv5
Glenn Jocher et al / Ultralytics
[YOLOv5 GitHub]

Jul 23, 2020*
PP-YOLO
X Long et al / Baidu
[An Effective and Efficient Implementation of Object Detector]

Nov 16, 2020*
Scaled-YOLOv4
Chuyi Li, Alexey Bochkovskiy et al
[Scaled-YOLOv4: Scaling Cross Stage Partial Network]

May 10, 2021
YOLOR
Chien-Yao Wang et al
[You Only Learn One Representation: Unified Network for Multiple Tasks]

Jun 1, 2021
YOLOS
Yuxin Fang et al / HUST
[You Only Look at One Sequence: Rethinking Transformer in Vision through Object Detection]

Jul 18, 2021*
YOLOX
Zheng Ge et al / Megvii
[YOLOX: Exceeding YOLO Series in 2021]

Jun 2022†
YOLOv6
Chuyi Li et al / Meituan
[YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications]

Jul 6, 2022
YOLOv7
Chien-Yao Wang, Alexey Bochkovskiy et al
[YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors]

Jan 10, 2023†
YOLOv8
Glenn Jocher et al / Ultralytics
[YOLOv8 GitHub]

Jan 13, 2023
YOLOv6 3.0
Chuyi Li et al / Meituan
[YOLOv6 v3.0: A Full-Scale Reloading]

May 2, 2023†
YOLO-NAS
Shay Aharon et al / Deci
[YOLO-NAS: A Next-Generation, Object Detection Foundational Model generated by Deci's Neural Architecture Search Technology]

Jan 30, 2024*
YOLO-World
Tianheng Cheng et al / Tencent
[YOLO-World: Real-Time Open-Vocabulary Object Detection]

Feb 21, 2024*
YOLOv9
Chien-Yao Wang et al
[YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information]

May 23, 2024
YOLOv10
Ao Wang et al / Tsinghua Univ
[YOLOv10: Real-Time End-to-End Object Detection]

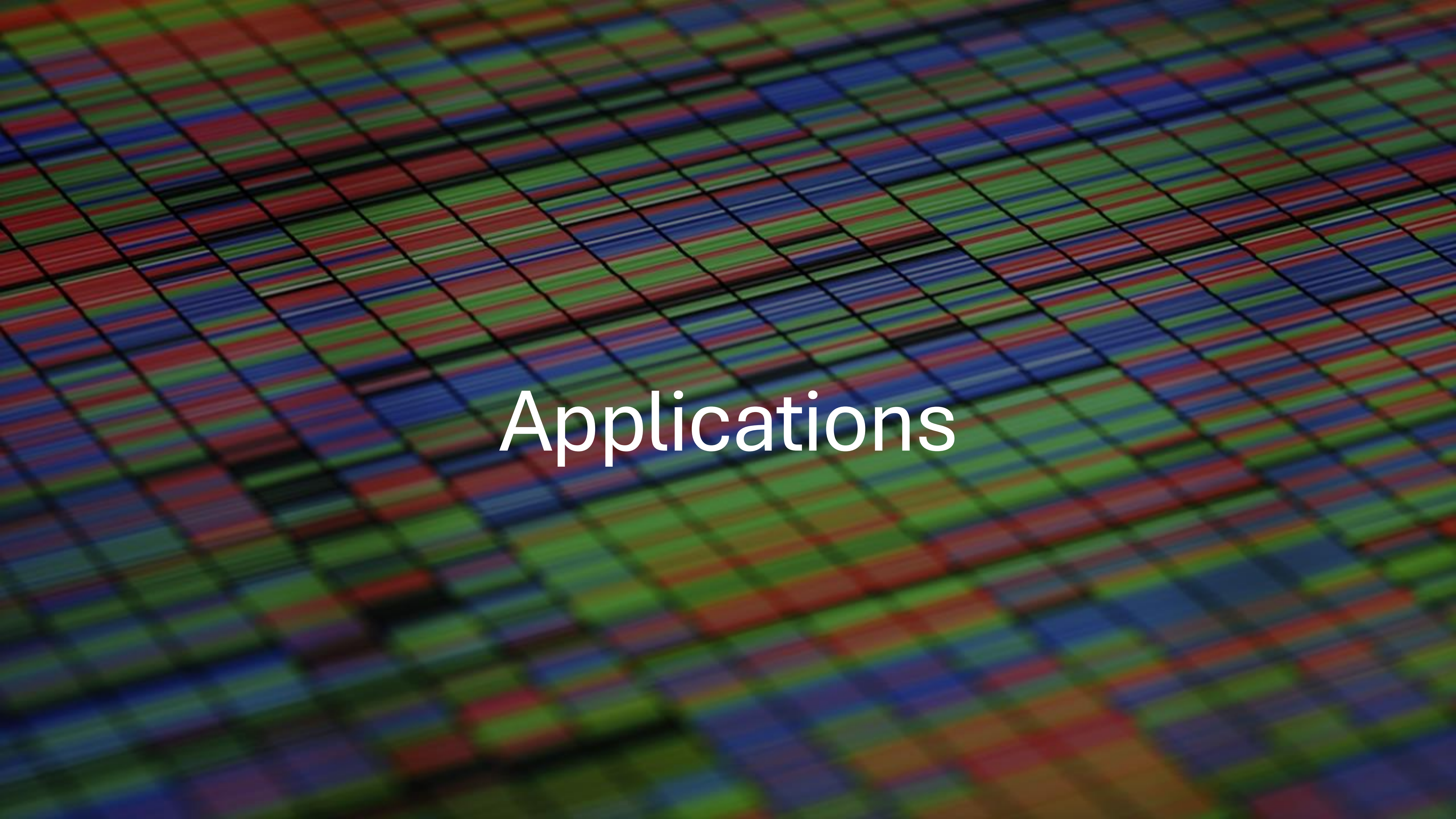
Sep 30, 2024†
YOLO 11
Glenn Jocher et al / Ultralytics
[YOLOv11 GitHub]

* Denotes paper updated after first publication date
† Denotes repository predates paper publication date

Joseph Nelson. (Jan 9, 2025). What is YOLO? The Ultimate Guide [2025]. Roboflow Blog: https://blog.roboflow.com/guide-to-yolo-models/

YOLOv12

- Released on February 18, 2025
- Introduced in the paper “YOLOv12: Attention-Centric Real-Time Object Detectors.”
- **Key Features**
  - Incorporates advanced attention mechanisms for improved detection accuracy.
  - Optimized for real-time applications with lower latency.
  - Open-source implementation available for fine-tuning and customization.
- **Performance**
  - Benchmarked on the Microsoft COCO dataset.
  - Achieves higher mean Average Precision (mAP) while reducing computational overhead.



Applications

General

Autonomous Vehicles

Detects pedestrians, vehicles, traffic signs in real-time.

Security & Surveillance

Identifies objects in CCTV footage.

Medical Imaging

Used for detecting tumors, anomalies in medical scans.

Retail & Logistics

Automated checkout systems, inventory tracking.

Dentistry

Tooth Segmentation

Identifying individual teeth for orthodontic planning.

Caries and Cavity Detection

Detecting early-stage cavities in dental X-rays.

Root Canal Detection

Assisting endodontists in visualizing canal morphology.

IAN (Inferior Alveolar Nerve) Localization

Preventing nerve damage during surgeries.

Pros and Cons

Pros	Cons
• Fast and efficient, capable of real-time detection. • Single-stage processing, end-to-end learning. • Works well in real-world applications.	• Struggles with detecting small objects. • May not be as accurate as two-stage methods like Faster R-CNN. • Limited interpretability due to end-to-end learning.

Questions?

Workshop Activity

- Notebooks Link
    - https://github.com/KnightsLab/EMRA-Workshop