

YOLO (You Only Look Once)

——リアルタイム物体検出の革新と進化の歩み

物体検出は、画像や映像の中から「どこに・何が」写っているのかを認識する技術であり、人工知能やコンピュータビジョン分野において長らく研究されてきた。その中で 2016 年に登場した YOLO (You Only Look Once) は、従来の常識を覆す高速な物体検出手法として世界に衝撃を与えた。

それまでの物体検出は、たとえば R-CNN や Fast R-CNN のように、画像の中から候補領域 (region proposals) を抽出し、それぞれに分類器を適用するというステップを踏んでいた。この方法は高精度だが計算量が多く、リアルタイム処理には不向きだった。

YOLO の登場により、状況は一変する。YOLO は「画像をグリッドに分割し、各セルに対して一度に物体のクラスと位置情報を予測する」という新しいアプローチを採用。これにより、1 枚の画像に対して 1 回のニューラルネットワークの実行で検出が完了する。初代 YOLO (v1) は速度重視であり、精度面では R-CNN に及ばなかったが、その革新性は多くの研究者や開発者にインパクトを与えた。

以降、YOLO はバージョンを重ねながら進化する。YOLOv2 (2017) では精度が大幅に向上し、YOLOv3 (2018) では小型物体検出にも対応。YOLOv4 (2020) は学術コミュニティ外 (Alexey Bochkovskiy) から登場し、軽量かつ高性能なモデルとして人気を集めた。YOLOv5 (2020) は Ultralytics 社による PyTorch 実装で、厳密には公式の継続ではないが、扱いやすさや速度から事実上の標準モデルとして広く使われるようになった。

さらに、YOLOv6・YOLOv7・YOLOv8 と続き、それぞれが推論速度・精度・使いやすさの面で進化してきた。YOLOv8 は 2023 年に登場し、OpenCV との親和性や ONNX 対応も強化され、産業用途やエッジ AI への実装が加速している。

YOLO はもはや単なる「1 つの物体検出手法」ではなく、リアルタイム AI の象徴的存在として、多くの応用領域 (監視カメラ、ドローン、AR、医療画像分析など) で活用されている。初代の「一度だけ見て検出する」という理念は、今なお YOLO シリーズの根幹に息づいており、その進化は止まる気配がない。

YOLO の進化：リアルタイム物体検出の革命史

YOLO のバージョンごとの概要

■YOLOv1 (2016 年)

開発：Joseph Redmon ら

特徴：画像をグリッドに分割し、それぞれのセルで物体の有無と位置を同時に予測

非常に高速だが、小さい物体の検出が苦手

リアルタイム処理に対応した初の物体検出手法

■YOLOv2 / YOLO9000 (2017 年)

改善点：バウンディングボックスの予測精度向上（Anchor Boxes の導入）

多クラス（9000 以上）を検出可能な YOLO9000 を発表

精度と速度のバランスが大きく向上

■YOLOv3 (2018 年)

特徴：Darknet-53 というより深いバックボーンを採用

複数スケールでの検出に対応（小型物体への対応強化）

より高精度で実用的な検出が可能に

■YOLOv4 (2020 年)

発表者：Alexey Bochkovskiy（Redmon の引退後）

特徴：より高速で高精度、GPU での学習や推論が容易

データ拡張や正則化技術（Mosaic augmentation、CSPDarknet）を多数導入

■YOLOv5 (2020 年)

開発：Ultralytics（非公式だが実質的な主流）

特徴：PyTorch ベースで扱いやすく、モデルサイズの実装が豊富（v5s～v5x）

軽量・高速な実装で多くの応用例に使用されている

■YOLOv6 (2022 年)

開発：Meituan（中国のテック企業）

特徴：産業応用を意識した設計、Edge 端末への展開を意識

精度と速度を両立するニューラルアーキテクチャを採用

■YOLOv7 (2022 年)

特徴：パラメータ効率が非常に高く、軽量モデルながら高精度
最新手法 (E-ELAN など) を活用し、速度・精度ともに最先端

■YOLOv8 (2023 年)

開発：Ultralytics

特徴：完全に再設計されたアーキテクチャ
インスタンスセグメンテーションや姿勢推定も統合的にサポート
ONNX や OpenVINO など多様なデプロイ形式に対応
用途に応じた軽量～大型モデル (n/s/m/l/x) を提供

■YOLOv9 (2024 年)

開発：Ultralytics

特徴：Programmable Gradient Information (PGI)：勾配の精度改善と情報損失の抑制
Generalized Efficient Layer Aggregation Network (GELAN)：層間情報の効率的統合
軽量構造の最適化：Depthwise 畳み込み + C3Ghost 構造により演算量削減
多様なモデルスケール：v9-S (小型) ～v9-E (大型) を提供し、用途に応じた選択可能

■YOLOv10 (2024 年)

開発：Ultralytics

特徴：NMS 不要の End to End 推論：Consistent Dual Assignments により
非最大抑制が不要に
空間・チャンネル分離ダウンサンプリング：情報損失を抑えつつ計算効率向上
Rank-Guided Block Design：構造冗長性を低減し、高効率化を実現
軽量分類ヘッド導入：末端処理の計算負荷を軽減
多段階モデル展開：n/s/m/b/l/x と、用途に応じたモデル選択が可能

YOLO の最大の強みは、「リアルタイムでの高精度な物体検出」に特化して進化してきた点にあります。現在では、監視カメラ、ロボティクス、自動運転、AR・VR、医療画像解析など、多くの分野で YOLO ベースの AI が活躍しています。

その進化は止まることを知らず、今後も軽量・高速・高精度という要求に応える次世代モデルの登場が期待されています。YOLO はまさに、「一度見るだけで見抜く」AI の理想形を体現する存在なのです。

以上