Yolo - Klassifikation

### Überblick:

YOLOv8 (You Only Look Once Version 8) ist die neueste Version einer Serie von Echtzeit-Objekterkennungsmodellen. Diese Modelle sind für ihre Geschwindigkeit und Genauigkeit bekannt, was sie ideal für Anwendungen wie Videoüberwachung, autonomes Fahren und viele weitere Bereiche macht.

Training: Die Trainingsdaten für YOLOv8 umfassen eine Vielzahl von Objekten und Szenarien, um das Modell darauf vorzubereiten, verschiedene Objekte in unterschiedlichen Kontexten und Ansichten zu erkennen. Jede Trainingsinstanz enthält ein Bild zusammen mit den Koordinaten und Klassenlabels der enthaltenen Objekte. Während des Trainings lernt das Modell, die Merkmale in den Bildern zu erkennen und die Begrenzungsrahmen sowie die zugehörigen Klassenwahrscheinlichkeiten präzise vorherzusagen.

* Objekterkennung
* Objektsegmentierung
* Posenschätzung
* Orientierte Objekterkennung
* Klassifizierung

### Geschichte:

YOLO (You Only Look Once) ist ein Computer-Vision-Algorithmus, der entwickelt wurde, um Objekte in Echtzeit zu detektieren und zu verfolgen. Es wurde von Joseph Redmon 2016 an der University of Washington veröffentlicht.

YOLOv8 ist die neueste Version des YOLO-Objekterkennungsmodells und wurde von Ultralytics entwickelt. Ultralytics bietet YOLOv8 als Open-Source-Projekt unter der AGPL-3.0-Lizenz an, was es ermöglicht, dass die Community an der Weiterentwicklung mitwirkt.

### Versionen:

* **YOLOv1:** Einführung der Idee eines einzigen neuronalen Netzwerks für die gleichzeitige Objekterkennung und Klassifizierung.
* **YOLOv2 (YOLO9000):** Verbesserungen bei der Genauigkeit und Einführung von Multi-Scale-Training.
* **YOLOv3:** Einführung von Feature Pyramid Networks zur Verbesserung der Erkennung kleiner Objekte.
* **YOLOv4:** Optimierung der Geschwindigkeit und Genauigkeit durch die Einführung von CSPDarknet53 und anderen Techniken.
* **YOLOv5:** Weiterentwicklungen im Bereich Benutzerfreundlichkeit und Anpassung.
* **YOLOv8:** Die neueste Version, die auf den Fortschritten der vorherigen Versionen aufbaut und weitere Optimierungen bietet.

### Funktionsweise von YOLOv8

1. **Eingabeverarbeitung:** YOLO v8 verwendet ein Bild als Eingabe und unterteilt es in ein Raster, das je nach Typ normalerweise eine Größe von 13 x 13 oder 26 x 26 hat. Jede Rasterzelle ist für die Vorhersage von Objekten innerhalb ihres räumlichen Bereichs verantwortlich.

2. **Merkmalsextraktion:** Das Netzwerk verwendet ein spezielles Modell, das auf Bilder spezialisiert ist und durch mehrere Schichten schrittweise Merkmale wie Kanten und Texturen aus dem Eingabebild lernt. Diese Merkmale werden immer abstrakter, beginnend mit einfachen Formen und Texturen bis hin zu komplexen Objekten oder spezifischen Teilen davon. Diese schrittweise Erfassung hilft YOLO v8, die wichtigen visuellen Eigenschaften des Bildes zu verstehen, die für die spätere Erkennung und Identifizierung von Objekten entscheidend sind.

Das Netzwerk von YOLOv8 verwendet ein spezialisiertes Modell, das darauf ausgelegt ist, visuelle Merkmale aus Eingabebildern zu extrahieren. Dies geschieht durch mehrere Schichten von Faltungen, bei denen Filter über das Bild bewegt werden, um lokale Merkmale wie Kanten, Texturen oder Muster zu identifizieren. Diese Faltungen ermöglichen es dem Netzwerk, schrittweise abstraktere Merkmale zu erfassen, beginnend mit grundlegenden Strukturen und sich hin zu komplexeren Objektteilen entwickelnd. Anschließend wird Pooling angewendet, eine Operation, die darauf abzielt, die Dimensionalität der entstehenden Feature Maps zu reduzieren, indem sie die Informationen innerhalb kleiner Patches aggregiert. Eine Feature Map ist eine zweidimensionale Darstellung aktivierter Merkmale, wobei jeder Wert in der Map ein spezifisches erkanntes Merkmal oder eine Kombination davon aus dem Eingabebild darstellt.

3. **Vorhersage von Begrenzungsrahmen:** YOLO v8 sagt den Begrenzungsrahmen für Objekte vorher, indem es die Koordinaten der oberen linken Ecke, die Breite und die Höhe des Rahmens regressiert. Zusätzlich berechnet es einen Konfidenzwert, der die Wahrscheinlichkeit angibt, dass der vorhergesagte Rahmen das Objekt enthält.

4. **Klassenvorhersage:** Erfolgt in YOLOv8 durch eine spezielle Schicht, die parallel zu den Begrenzungsrahmenvorhersagen arbeitet. Für jede Rasterzelle im vordefinierten Raster (z. B. 13x13 oder 26x26 Zellen) sagt das Modell die Wahrscheinlichkeiten für verschiedene Objektklassen vorher. Jede Zelle im Raster von YOLOv8 gibt an, wie wahrscheinlich es ist, dass ein bestimmtes Objekt dort zu finden ist. Diese Wahrscheinlichkeiten werden so angepasst, dass sie gut vergleichbar sind und sich insgesamt zu 1 addieren, was dem Modell hilft, Objekte im Bild genau und schnell zu identifizieren.

5. **Nachbearbeitung:** Sobald die Vorhersagen getroffen wurden, wird ein Vertrauensschwellenwert angewendet, um Erkennungen mit geringer Vertrauenswürdigkeit herauszufiltern. Danach wird eine nicht maximale Unterdrückung durchgeführt, um redundante oder überlappende Begrenzungsrahmen zu eliminieren und sicherzustellen, dass nur die präzisesten Ergebnisse weiterverfolgt werden.

### Varianten:

YOLO v8 umfasst mehrere Varianten, die jeweils sorgfältig auf bestimmte Anwendungsfälle und Bedürfnisse zugeschnitten sind.

1**. YOLO v8-Tiny:** YOLO v8-Tiny geht einen Kompromiss zwischen Genauigkeit und Vorhersagegeschwindigkeit ein. Erreicht wird dies durch eine kleine Netzwerkarchitektur und eine kleine Gittergröße (z. B. 13x13), die eine Echtzeitleistung auch auf Geräten mit eingeschränkten Ressourcen ermöglicht.

2**. YOLO v8-SPP:** Der SPP-Typ (Spatial Pyramid Pooling) enthält ein SPP-Modul im Netzwerk. Dieser Zusatz erleichtert die effektive Erfassung von Merkmalen in mehreren Maßstäben (unterschiedliche Skalierung im Bild), was zu einer verbesserten Genauigkeit führt, insbesondere bei Objekten unterschiedlicher Größe.

3. **YOLO v8-CSPDarknet:** YOLOv8-CSPDarknet kombiniert die YOLOv8-Architektur mit dem CSPDarknet-Backbone. Die CSP-Verbindung (Cross-Stage Partial) ist eine spezielle Architektureigenschaft, die die Effizienz der Merkmalsextraktion und -darstellung verbessert. Diese Architektur ist besonders gut geeignet für Anwendungen, die eine hohe Leistung und Genauigkeit erfordern, da sie den Informationsfluss zwischen den Netzwerkschichten optimiert und die Repräsentationsfähigkeit des Modells erhöht.

4. **YOLO v8-Panet:** YOLO v8-Panet integriert die PANET-Architektur (Path Aggregation Network), die die Genauigkeit der Merkmalsfusion und der Objekterkennung verbessert. Diese Architektur ist besonders geeignet für Anwendungen, bei denen eine präzise Lokalisierung von Objekten von entscheidender Bedeutung ist. PANET optimiert die Zusammenführung von Merkmalen aus verschiedenen Teilen des Netzwerks, um eine verbesserte und zuverlässige Erkennung von Objekten zu gewährleisten, unabhängig von ihrer Größe oder Position im Bild

### Vor & Nachteile:

Vorteile:

* **Hohe Geschwindigkeit:** YOLOv8 ist für seine Fähigkeit bekannt, Objekte in Echtzeit zu erkennen, was es ideal für Anwendungen macht, die schnelle Reaktionszeiten erfordern, wie zum Beispiel Videoüberwachung oder autonome Fahrzeuge.
* **Effiziente Nutzung von Ressourcen:** Das Modell ist darauf optimiert, mit vergleichsweise wenig Rechenleistung und Speicher auszukommen, was seine Implementierung auf verschiedenen Plattformen und Geräten erleichtert.
* **Gute Balance zwischen Geschwindigkeit und Genauigkeit:** YOLOv8 bietet eine gute Balance zwischen Geschwindigkeit und Genauigkeit in der Objekterkennung, was es zu einer praktischen Wahl für viele Echtzeitanwendungen macht.
* **Vielseitigkeit:** YOLOv8 kann eine Vielzahl von Objekten und Szenarien gut verarbeiten, da es auf einem umfassenden und vielfältigen Trainingsdatensatz basiert.

Nachteile:

* **Geringere Detailgenauigkeit:** Aufgrund seiner schnellen Verarbeitung kann YOLOv8 in der Detailerkennung im Vergleich zu tieferen, rechenintensiveren Modellen etwas weniger präzise sein.
* **Abhängigkeit von Trainingsdaten:** Die Leistung von YOLOv8 hängt stark von der Qualität und der Vielfalt der Trainingsdaten ab. Ein unzureichender oder unrepräsentativer Trainingsdatensatz kann die Leistung beeinträchtigen.

### Anwendungsgebiete:

* **Autonomes Fahren:** Erkennung von Fahrzeugen, Fußgängern, Verkehrsschildern und anderen Verkehrsteilnehmern für die Umgebungswahrnehmung autonomer Fahrzeuge.
* **Videoüberwachung und Sicherheit:** Echtzeit-Erkennung von Personen, Fahrzeugen oder verdächtigen Aktivitäten in Überwachungsvideos zur Sicherheitsüberwachung.
* **Objektverfolgung:** Kontinuierliche Verfolgung und Identifizierung von Objekten über verschiedene Frames in Videos, zum Beispiel zur Überwachung von Warenbeständen oder Menschenmassen.
* **Industrielle Anwendungen**: Erkennung von Fehlern oder Defekten in Produktionsprozessen, Überwachung von Inventarbeständen und Unterstützung bei der Robotik in Fabriken.