

Problem Statement

Fabian Boni

Oliver Pejic

13. Mai 2025

Zuverlässige Lokalisierung von Hunden in Überwachungs- und Alltagsaufnahmen

Zielsetzung und Relevanz

Ziel dieses Projekts ist es, systematisch zu untersuchen, unter welchen Bedingungen Hunde in realitätsnahen Bildszenarien zuverlässig lokalisiert werden können. Der Fokus liegt nicht auf der Entwicklung einer neuen Architektur, sondern auf der Analyse der Leistungsfähigkeit etablierter Object Detection Modelle in verschiedenen Szenarien. Die Ergebnisse sollen praxisrelevant sein für Anwendungsbereiche wie Heimtierüberwachung, Smart-Home-Systeme, Tierschutz oder Such- und Rettungssysteme.

Forschungsfrage

Wie zuverlässig erkennen und lokalisieren gängige Object Detection Modelle Hunde in verschiedenen alltäglichen Szenarien, und welche Einflussfaktoren (z.B. Überlappung, Lichtverhältnisse) führen zu signifikanten Leistungseinbussen?

Spezifikation des Untersuchungsgegenstands

- Ziel ist die Erkennung und Lokalisierung der Klasse „Hund“, unabhängig von der Rasse.
- Die Modelle sollen mit Bildern aus Alltagssituationen umgehen können, einschliesslich kontrollierter (z.B. Überwachungskameras) und unkontrollierter (z.B. Smartphone-Fotos) Quellen.
- Relevanz: Hunde sind häufige Zielobjekte in häuslichen und öffentlichen Umgebungen.

Datengrundlage

Es werden vier öffentlich verfügbare Datensätze verwendet, die verschiedene Kontexte abdecken. Insgesamt stehen rund 5'300 Bilder mit über 6'200 Bounding-Box-Annotationen zur Verfügung. Die Annotationen stammen aus folgenden Quellen:

- **Kaggle – Dog and Cat Detection:** <https://www.kaggle.com/datasets/andrewmvd/dog-and-cat-detection>
- **Roboflow – Dogs OVDDC:** <https://universe.roboflow.com/wu-yuxuan-ovddc/dogs-5ulz9>
- **Roboflow – AIVLE5-WORF0 ASDF-T4TSD:** <https://universe.roboflow.com/aivle5-worf0/asdf-t4tsd>
- **Roboflow – Max-EVO5Q Dog UXSTE:** <https://universe.roboflow.com/max-evo5q/dog-uxste>

Nicht-Hunde-Klassen werden entfernt, sodass ein fokussiertes Detektionsproblem entsteht. Die Daten werden in ein einheitliches Format (z.B. COCO JSON) überführt und nach train/val/test aufgeteilt.

Untersuchte Einflussfaktoren

Die folgenden Faktoren werden gezielt untersucht, da sie im Alltag häufig auftreten und die Lokalisierung erschweren können. Die Reihenfolge entspricht der Priorität der geplanten Analyse. Weitere Szenarien werden behandelt, sofern es der zeitliche Rahmen erlaubt:

- Unterschiedliche Bildqualität (z.B. niedrige Auflösung, Rauschen)
- Variierende Lichtverhältnisse (z.B. Tag/Nacht, Schatten, Kunstlicht)
- Mehrere Hunde im Bild (Überlappung)

Diese Variablen werden auf Szenario-Ebene gelabelt und analysiert. Groundtruth-Annotationen stammen, soweit vorhanden, aus den Datensätzen.

Methodisches Vorgehen

1. **Setup und Datenvorbereitung:** Konvertierung der Annotationen in ein gemeinsames Format; Normalisierung und Resize der Bilder (z.B. 640×640 px); Entfernen beschädigter oder fehlerhafter Dateien
2. **Erste einfache Base-Line trainieren:** Verwendung eines einfachen, schnellen Object Detection Modells wie **YOLOv5n** oder **SSD MobileNet**; Training auf einem Teildatensatz zur Überprüfung der Datenpipeline
3. **Evaluation der Baseline:** Hauptmetrik: mean Average Precision (mAP) bei $\text{IoU} = 0.5$; Sekundärmetriken: Precision, Recall; visuelle Kontrolle von Erfolg und Fehlern
4. **Regelmässige Trainingsläufe mit Logging:** Einbindung von EarlyStopping, ModelCheckpoint, ggf. WandB für Logging und Visualisierung
5. **Gezielte Datenaugmentation:** Helligkeits-/Kontrastveränderung, Rauschen, horizontale Spiegelung, kleine geometrische Transformationen

Hinweis

Die technische Implementierung (Modellauswahl, Architekturdetails, Hyperparameter) sowie Zwischenergebnisse werden im Jupyter-Notebook dokumentiert.