

## Umsetzung eines Verfahrens zur gleichzeitigen Objekterkennung und Schätzung menschlicher Posen aus Einzelbildern [Bachelor/Master]

Die Detektion von objektumschließenden Bounding Boxes (BBs) (*Objekterkennung*) sowie die Detektion von skelettbasierten Personenmodellen (*Schätzung menschlicher Posen*) sind zwei Problemstellungen, bei denen in den vergangenen Jahren enorme Fortschritte unter Einsatz neuronaler Netze erzielt wurden. Dies betrifft sowohl die Detektionsqualitäten als auch die Detektionsgeschwindigkeiten (z.B. One-Shot-Ansätze wie YOLOv4 [1] für die Objekterkennung sowie OpenPose [2] für die Schätzung menschlicher Posen). Dennoch ist der Leistungsbedarf derartiger Detektoren insbesondere für mobile Anwendungen mit hohen zeitlichen Verfügbarkeiten noch immer ein limitierender Faktor für den Einsatz.

Gerade die gemeinsame Betrachtung von Personenbeschreibungen und Szenenobjekten liefert durch die Möglichkeit zur Kontextualisierung wertvolle Hinweise beispielsweise zur nachfolgenden Erkennung der Handlung von Personen (vgl. Abbildung 1). Zudem ergibt sich die Möglichkeit, Neben gegenseitigen (Teil)Verdeckung von Personen auch Aspekte der (Teil)Verdeckung durch Gegenstände zu untersuchen.

Nutzbare Objektdetektoren, wie YOLOv4, basieren auf der Regression von objektbezogenen BB-Parametern (vier Parameter). Auch für eine Keypoint-Beschreibung von Personen besteht die Möglichkeit der Regression von Keypoint-Koordinaten (je zwei Parameter). Sind in einem Bild mit mehreren Personen lediglich Keypoint-Koordinaten vorhanden, ist die Zuordnung zu einzelnen Personen allerdings schwierig – insbesondere dann, wenn die Abstände zwischen den Personen gering sind. Hilfestellung zur Auflösung der Mehrdeutigkeiten können die BB-Detektionen der Personen bieten, die sich aber auch überlappen können. Zum anderen sind für die Disambiguierung Hinweise über Nachbarschaftsbeziehungen von Nutzen. Derartige Hinweise werden beispielsweise von OpenPose generiert (sog. *Part Affinity Fields*) [2]. Einen ähnlichen Nutzen hätte die Detektion von Keypoint-Paaren (vgl. Abbildung 2), die wie BBs von vier Koordinaten beschrieben werden.

In dieser Arbeit soll diese Analogie zur gleichzeitigen Objekterkennung (z.B. 80 Objektklassen aus dem Microsoft Coco-Datensatz [3]) und Schätzung der Posen sämtlicher Personen aus Einzelbildern mit nur einer Netzwerk-Inferenz und nachgelagerter Heuristik ausgenutzt werden.

### Die Bearbeitung sollte die folgenden Schritte umfassen:

- Adaptierung/Optimierung vorhandener Detektor-Netzwerkstrukturen zur Detektion mehrerer Objektklassen sowie personenbezogener Keypoint-Paare.

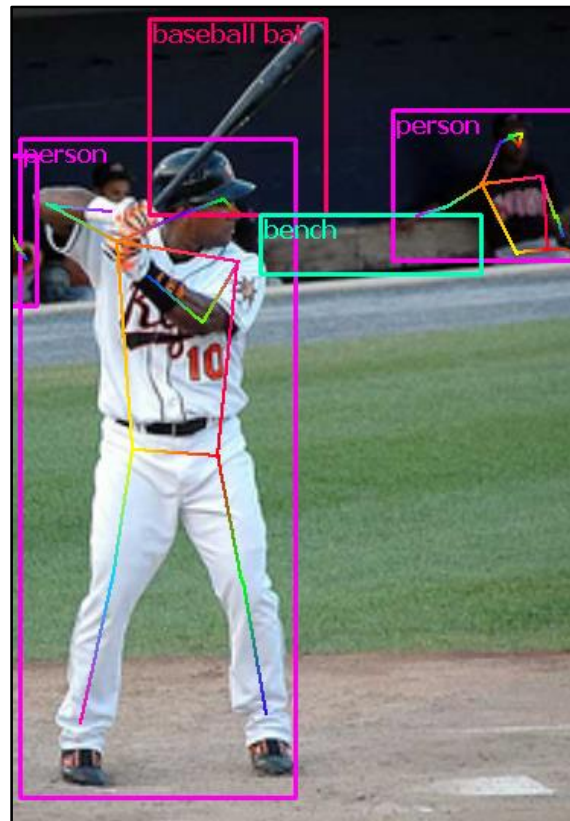


Abbildung 1: Bildausschnitt aus Coco-Datensatz mit Bounding-Box-Annotationen und zu Skeletten zusammengefassten Keypoint-Annotationen.

- Entwicklung eines Algorithmus, welcher die einzelnen Detektionen von Keypoint-Paaren zu optimierten skelett-basierten Personenmodellen (Abbildung 2) zusammenfasst.
- Untersuchung und Bewertung von Aspekten der (Teil)Verdeckung, sowohl von Personen durch weitere Personen als auch durch Gegenstände.
- Training und Evaluierung unter Verwendung des *Microsoft Coco*-Datensatzes [3].
- Dokumentation der Ergebnisse.

**Ansprechpartner am ZBS e. V.:** David Reese ([david.reese@zbs-ilmenau.de](mailto:david.reese@zbs-ilmenau.de))



**Abbildung 2:** Veranschaulichung für das gewünschte Zusammenfassen von Keypoint-Paar-Detektionen zu optimierten skelett-basierten Personenmodellen (die rechts dargestellten Personenmodelle entsprechen hier den Keypoint-Annotationen des *Coco*-Datensatzes).

### Literaturhinweise

- [1] A. Bochkovskiy, C.-Y. Wang und H.-Y. M. Liao, „YOLOv4: Optimal Speed and Accuracy of Object Detection,“ arXiv preprint arXiv:2004.10934, 2020.
- [2] „OpenPose: A Real-Time Multi-Person Keypoint Detection And Multi-Threading C++ Library,“ CMU Perceptual Computing Lab, 2017. [Online]. Available: <https://github.com/CMU-Perceptual-Computing-Lab/openpose>. [Zugriff am 8. August 2017].
- [3] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár und C. L. Zitnick, „Microsoft coco: Common objects in context,“ Springer, 2014.