

# Camera Based Pose Estimation for Robotics

## Machbarkeitsstudie zur Schätzung von Gelenkwinkeln aus Synthetischen Bilddaten

### Diplomand



Sascha Scherrer

**Ausgangslage:** Schreitbagger neigen dazu, bei unsachgemässer Bedienung zu kippen, insbesondere wenn die Schaufel zu voll geladen wird, und der Baggerarm zu weit ausgefahren wird. Dadurch kann es vorkommen, dass die vertikale Projektion des Schwerpunkts ausserhalb des von den Stützen aufgespannten Polygons zu liegen kommt und der Bagger wird instabil. Um dies zu verhindern müssen für ein Warnsystem die Gelenkwinkel des Arms und die Position des Teleskoparms bekannt sein. Normalerweise werden dafür konventionelle Positions- und Winkelsensoren direkt am Baggerarm eingesetzt. Diese gehen aber durch die rauen Anwendungsbedingungen wie z.B. Schmutz, Vibration oder Schläge leicht kaputt. Ausserdem bedeutet die Installation ein hoher Kostenaufwand. Daher sind aktuell keine eingebaut. Ziel dieser Arbeit ist es, die Machbarkeit der Schätzung der Gelenkwinkel des Baggerarms mithilfe einer Bilderkennung in Form eines Neuronalen Netzwerkes zu zeigen. Dies würde es erlauben, ein Warn- oder Assistenzsystem zu integrieren.

**Vorgehen:** Die benötigten Bilddaten werden künstlich mit einer Roboter-Simulationsplattform auf der Basis von NVIDIA Omniverse generiert. Das hat den Vorteil, dass die Winkelwerte (Labels) bekannt sind und diese gleichzeitig mit den Bildern der zwei Kameras (Inputs) abgespeichert werden können. So lässt sich virtuell und effizient ein grosses, realitätsnahes Trainingsset erzeugen.

Als Feature-Extractor wird ein EfficientNet verwendet. Mittels Transferlearning wird ein Regression-Head trainiert, welcher die vier Winkelwerte und die Position des Teleskops des Baggerarms vorhersagt. In einem ersten Schritt werden zwei Bilder - linke und rechte Ansicht des Baggerarms - des gleichen Zeitschrittes verarbeitet. In einem zweiten Versuch werden drei Zeitschritte als Stream zusammengefasst, um eine zeitliche Information zu gewinnen. Diese verschiedenen Modellarchitekturen wurden implementiert und analysiert.

**Ergebnis:** Es hat sich gezeigt, dass die Grösse des EfficientNets keinen grossen Einfluss auf das Ergebnis hat und ein EfficientNetB2 ausreichend ist. Die Versuche haben auch gezeigt, dass Finetuning der letzten zwei bis drei Blöcke des EfficientNets in der Balance zwischen Ergebnis und Performance am besten funktioniert. Auch wurden andere Hyperparameter untersucht.

Mit dem besten Modell konnte ein Validierungs-MSE von unter 0.001 erreichen, was einem durchschnittlichen absoluten Fehler von  $< 2^\circ$  für die beiden ersten Gelenke des Baggerarms und 5 cm respektive  $5^\circ$  für die beiden hinteren Aktoren entspricht. Damit wurde die Machbarkeit der robusten Echtzeit-Schätzung der Freiheitsgrade eines Baggerarms mittels eines Kamerasystems gezeigt. Als nächsten Schritt ist vorgesehen, diese Modelle

mittels realen Bilddaten fein abzustimmen, um eine maximale Robustheit der Schätzung zu erreichen.

### Kaiser S12 Allroad

<https://www.kaiser.li/de/produkte/s12-allroad>



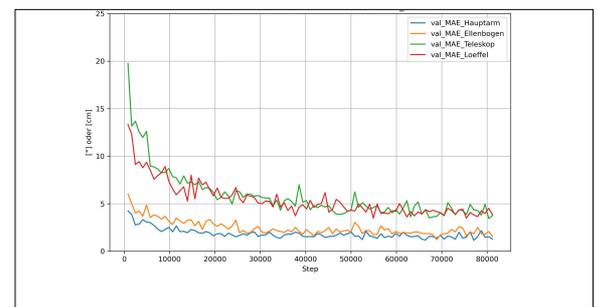
### Ansicht Boom-Kamera rechts

Eigene Darstellung



### Absoluter Fehler der Aktoren über die Trainingszeit: Snapshot-Modell

Eigene Darstellung



### Referent

Prof. Dr. Christoph Würsch

### Korreferent

M. Sc. Robin Vetsch

### Themengebiet

Computational Engineering