Camera Based Pose Estimation for Robotics

Feasibility Study for Estimating Joint Angles from Synthetic Image Data:

Graduate

Sascha Scherrer

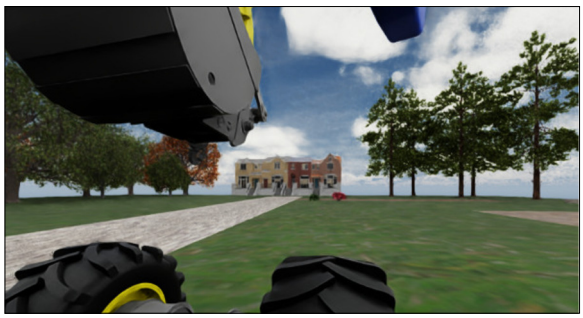
Problem: Walking excavators tend to tip over during improper operation if the bucket is overloaded and the boom is extended too far. This causes the center of gravity to move outside the support base. For a warning system, however, the angles of the arm's joints need to be known. Normally, this is achieved using conventional position sensors directly in the boom. These sensors, however, are prone to failure due to harsh conditions, such as dirt, and therefore are currently not installed.
The goal of this work is to conduct a feasibility study for estimating the joint angles of the boom using image recognition through a neural network. This would enable the integration of a warning or assistance system.

Approach / Technology: The required image data are generated artificially using a robotics simulation platform built on NVIDIA Omniverse. This has the advantage that the angle values (labels) are known and can be stored together with the images from two cameras. In addition, this approach allows for the creation of a large training dataset.
An EfficientNet is used as a feature extractor via transfer learning. A small regression head outputs the four angles and the displacement value of the telescopic boom. In the first approach, two images - left and right view of the boom - from the same time step are processed, while in a second approach, three consecutive time steps are combined to capture temporal information. These different model architectures were implemented and analyzed.

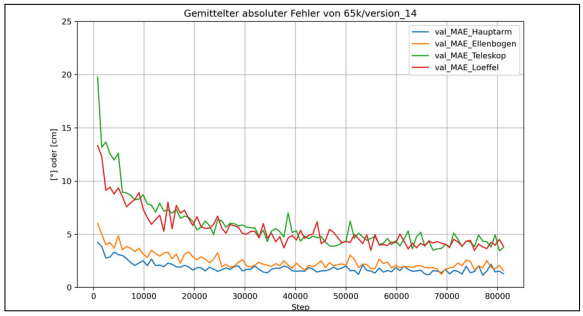
Result: It was found that the size of the EfficientNet has little impact on the results, and an EfficientNetB2 is sufficient. The experiments also showed that fine-tuning the last two or three blocks of the EfficientNet achieves the best balance between accuracy and performance. Other hyperparameters were also investigated.
The best model achieved a validation MSE of under 0.001, corresponding to an average absolute error of < 2° for the first two joints of the boom, and 5 cm and 5° for the two rear actuators, respectively. The work shows the feasibility for reliable boom pose estimation based on camera images. The next step will be the fine-tuning of the models on real data to achieve maximum robustness.

Kaiser S12 Allroad
https://www.kaiser.li/de/produkte/s12-allroad

Sight from the right boom camera
Own presentment



Absolute error of the actors over training time: snapshot-model
Own presentment



Advisor
Prof. Dr. Christoph Würsch

Co-Examiner
M. Sc. Robin Vetsch

Subject Area
Computational Engineering