The 6D pose estimation is an important computer vision task with applications in robotics, *e.g.* for manipulation or grasping, but also in computer graphics and augmented reality. Given an image, the task is to estimate the 3D rotation and 3D translation of the known object with respect to the camera. The task is even more challenging in an uncontrolled environment, *e.g.* when we do not have proper camera calibration. In that case, the focal length also needs to be estimated with the 6D pose. In this work, we address the issues of methods that work in such uncontrolled environments.

First, we focus on FocalPose, a state-of-the-art method for joint estimation of object 6D pose and camera focal length. We review the method and propose several improvements. These include (i) re-deriving and improving the 6D pose and focal length update rule, (ii) replacing the model retrieval method, and (iii) changing the distribution of 6D poses and focal lengths used for synthetic training data rendering. These changes lead to improved results compared to the state-of-the-art FocalPose method.

Second, to avoid often costly retraining of models for 6D pose estaimation, it is beneficial to consider methods with the ability to generalize to novel objects that have not been seen during training. These methods require a 2D bounding box and an object identity from a 3D object database to be known at inference time. Thus, a model-based method for novel object detection is also required. In particular, we investigate CNOS, a method for model-based segmentation of novel objects, that is widely used to solve the detection for 6D pose estimation. We provide an evaluation of average recall and average precision in its two stages: the proposal stage, which uses the Segment Anything Model to generate candidate object locations, and the matching stage, which uses DINOv2 feature similarity to assign confidence scores and object identities. We analyze the strengths and weaknesses of CNOS, and discuss the potential space for improvements.