# ABSTRACT

Turning a physical object into an editable digital model is a long-standing problem at the intersection of computer vision, computer-aided design (CAD), and intelligent manufacturing. Creating a CAD model usually requires time, effort, experience and access to specialized tools such as laser scanners or light devices. This thesis tries to tackle the problem in a more simple manner: how far can you get with just a phone camera? That is the question this thesis is trying to answer. More specifically, can a physical object be turned into a CAD-compatible digital file using only RGB images and a series of standard computer vision and geometry processing steps?

This work showcases a constrained image to CAD pipeline that is aimed at simple box-like objects. The system starts with multi-view image capture and image-based 3D construction in COLMAP. Then the constructed geometry turns into points clouds, which are cleaned and abstracted in Open3D. considering image-based reconstruction does not have the ability to determine real world dimensions the next step uses one known real-world dimension to determine the rest of the unknown dimensions. The recovered dimensions are then used in a script which generates both mesh based and parametric CAD output through FreeCAD, producing and editable local CAD file (.FCStd) and a cross-platform engineering exchange file (.STEP)

Past getting a proof of concept, in this thesis I also look closely at how different cuboid fitting strategies affect geometric accuracy. Four methods are evaluated: an axis-aligned bounding box (AABB), an oriented bounding box (OBB), a percentile-constrained axis aligned box, and lastly a hybrid method that combines both the percentile filtering and oriented fitting. In order to accurately gauge the relationship between reconstruction density and fitting accuracy each method is run on three reconstructions built from 50%, 75%, and 100% of the captured image set. The results show that no single fitting strategy is superior across all tests, OBB performs the best on the 50% and 75% subsets, with mean percentage dimension errors of 5.62% and 9.39% respectively, however the hybrid method performs best on the 100% reconstruction with 16.29%. a noteworthy aspect is that the hybrid method is also the only one of the three non-baseline methods that consistently beats the AABB baseline in every condition tested. Thus, it is concluded that fitting accuracy depends on the density, quality and structural stability of the points clouds rather than the amount of images you feed it.

This thesis also showcases a preliminary baseline experiment on the T-LESS dataset. During the experiment generic object detection and standard feature-based reconstruction failed on textureless industrial objects. Pretrained YOLOv8 produced semantically wrong class labels in nearly every frame, and COLMAP could not even initialise a sparse reconstruction from the available views. That failure ended up being the motivation for controlled capture, added surface texture, and a constrained geometric prior in the final prototype. All in all this thesis showcases that it is possible to use image based reconstruction to generate a CAD oriented model for simple man made objects, whilst still making it clear where the current method stops working.

**KEY WORDS:** Image-to-CAD, Photogrammetry, Point Cloud, Cuboid Fitting, Parametric CAD.

# 摘要

将物理对象转化为可编辑的数字模型，是计算机视觉、计算机辅助设计（CAD）与智能制造交叉领域中一个长期存在的难题。创建 CAD 模型通常需要耗费大量时间与精力，依赖丰富的专业经验，并借助激光扫描仪或光学设备等专用工具。本论文尝试以更简便的方式来解决这一问题：仅凭一部手机摄像头，究竟能走多远？这正是本论文试图回答的核心问题。具体而言，能否仅使用 RGB 图像，经由一系列标准计算机视觉与几何处理步骤，将物理对象转化为与 CAD 兼容的数字文件？

本研究展示了一套面向简单箱型物体的受约束图像转 CAD 流水线。系统首先在 COLMAP 中完成多视角图像采集与基于图像的三维重建，随后将重建几何体转换为点云，并在 Open3D 中进行清洗与抽象化处理。由于基于图像的重建无法确定真实世界的尺寸，下一步利用一个已知的真实尺寸推算其余未知尺寸。恢复的尺寸随后被输入脚本，通过 FreeCAD 生成基于网格的输出与参数化 CAD 输出，最终产生可编辑的本地 CAD 文件和跨平台工程交换文件。

在完成概念验证之外，本论文还深入考察了不同长方体拟合策略对几何精度的影响。研究评估了四种方法：轴对齐包围盒（AABB）、方向包围盒（OBB）、百分位约束轴对齐包围盒，以及融合百分位过滤与方向拟合的混合方法。为准确衡量重建密度与拟合精度之间的关系，每种方法分别在由 50%、75% 和 100% 图像子集构建的三组重建上运行。结果表明，没有任何一种拟合策略在所有测试条件下均占优势：OBB 在 50% 和 75% 子集上表现最佳，平均尺寸百分比误差分别为 5.62% 和 9.39%；而混合方法在 100% 重建上表现最优，误差为 16.29%。值得注意的是，混合方法也是三种非基线方法中唯一在所有测试条件下均优于 AABB 基线的方法。由此得出结论：拟合精度更多取决于点云的密度、质量与结构稳定性，而非输入图像的数量。

此外，本论文还展示了一项基于 T-LESS 数据集的初步基线实验。实验中，通用目标检测与标准特征点重建方法均在无纹理工业对象上失效：预训练 YOLOv8 在几乎每帧中均产生语义错误的类别标签，COLMAP 甚至无法从现有视角初始化稀疏重建。这一失败促使研究转向受控采集、添加表面纹理，并在最终原型中引入受约束的几何先验。总体而言，本论文表明，对于简单的人工制造对象，基于图像的重建与面向

CAD 的模型生成之间的衔接是切实可行的。同时也清晰地指出了当前方法的局限所在：重建鲁棒性、几何保真度，以及对更复杂形状的泛化能力。

关键词：图像转 CAD、摄影测量、点云、长方体拟合、参数化 CAD