

ABSTRACT

This master's thesis addresses the development and evaluation of an end-to-end process for generating synthetic image data to train visual classifiers for automated slipsheet detection in modern warehouse logistics environments. As automated Order Picking Systems (OPSs) become more prevalent, intelligent perception systems capable of overcoming specific challenges, such as the detection and removal of separation layers (slipsheets) between product layers, are increasingly crucial. A primary focus is placed on the exploitation and potential of synthetic training data generated via the simulation platform *NVIDIA Omniverse*, specifically using the integrated *Omniverse Replicator* toolkit. Realistic pallet scenarios are modeled using scanned textures from over 400 real-world products to ensure high variability and fidelity with actual warehouse conditions.

The research is guided by three key questions: the effectiveness of synthetic data for fine-tuning pre-trained visual models in real-world tasks, optimal strategies for combining real and synthetic data to minimize the sim-to-real gap, and the impact of data augmentation on the generalization abilities of trained models. To answer these, both extensive synthetic datasets and real images collected from various domains (including live warehouses and laboratory environments) were created and systematically combined.

A total of 187 models employing different deep learning architectures, such as *ResNet*, *EfficientNet*, and *YOLO11s-cls*, were trained and evaluated. The results demonstrate that exclusive training on synthetic data is insufficient for real-world deployment. Significant accuracy improvements suitable of deployment in real-world scenarios were achieved either through targeted fine-tuning of models pre-trained on synthetic data with even small amounts of real data or mixing real data into the synthetic training dataset. Data augmentation further supported cross-domain generalization, though the optimal level of augmentation was found to depend on both the chosen architecture and investigated domain.

Overall, this thesis provides a comprehensive contribution to the practical use of synthetic training data and its optimal combination with real data for automation and deep learning in warehouse logistics.

Keywords - Synthetic Data Generation; Image Classification; Sim-To-Real Transfer Learning; Fine-tuning; Data Augmentation; Cross-Domain Predictions

ZUSAMMENFASSUNG

Die vorliegende Masterarbeit beschäftigt sich mit der Entwicklung und Evaluierung eines End-to-End-Prozesses zur Erzeugung synthetischer Bilddaten für das Training visueller Klassifikatoren zur automatisierten Slipsheet-Erkennung in modernen Lagerlogistikumgebungen. Mit dem zunehmenden Einsatz automatisierter Kommissioniersysteme (OPS) wächst die Bedeutung intelligenter Erkennungssysteme, die spezifische Herausforderungen wie das Erkennen und Entfernen von Trennlagen (Slipsheets) zwischen Produktlagen bewältigen können. Ein Schwerpunkt der Arbeit liegt auf der Nutzung und dem Potenzial synthetischer Trainingsdaten, welche mithilfe der Simulationsplattform *NVIDIA Omniverse* und dem darin integrierten *Omniverse Replicator* erzeugt werden. Dabei werden realistische Palettenszenarien unter Verwendung gescannter Texturen von über 400 Produkten generiert, um eine möglichst hohe Varianz und Nähe zu realen Lagerbedingungen zu gewährleisten.

Die Arbeit adressiert drei Forschungsfragen: die Effektivität synthetischer Daten beim Fine-tuning vortrainierter visueller Modelle für reale Aufgaben, die optimale Kombination von realen und synthetischen Daten zur Minimierung des Sim-to-Real-Gaps sowie den Einfluss von Datenaugmentierung auf die Generalisierungsfähigkeit der trainierten Modelle. Zur Beantwortung dieser Fragen wurden sowohl umfangreiche synthetische Datensätze als auch real erhobene Bilddaten aus unterschiedlichen Anwendungskontexten (u.a. Live-Lager und Testumgebungen) kombiniert und verglichen.

In systematischen Experimenten wurden unterschiedliche Deep-Learning-Architekturen, darunter *ResNet*, *EfficientNet* und *YOLO11s-cls*, mit insgesamt 187 Modellen trainiert und getestet. Die Resultate zeigen, dass das alleinige Training auf synthetischen Daten nicht ausreicht, um eine hinreichende Leistung im realen Einsatz zu erzielen. Erst durch gezieltes Fine-tuning von an synthetischen Daten vortrainierten Modellen mit selbst kleinen Mengen realer Daten oder dem Mischen von realen und synthetischen Datensätzen ließ sich die Erkennungsgenauigkeit signifikant steigern, sodass sie für den realen Einsatz ausreichend war. Datenaugmentierung trug insbesondere zur Domänenübertragung und Generalisierungsfähigkeit bei, wobei sich zeigte, dass der optimale Grad der Augmentierung von der verwendeten Architektur und den Eigenschaften der Trainingsdaten abhängt.

Die Arbeit liefert einen fundierten Beitrag zum praktischen Einsatz synthetischer Trainingsdaten und deren optimaler Kombination mit realen Daten im Bereich Automatisierung und Deep Learning für die Intralogistik.

Keywords - Synthetic Data Generation; Image Classification; Sim-To-Real Transfer Learning; Fine-tuning; Data Augmentation; Cross-Domain Predictions