HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

GRADUATION THESIS

DANN-Based Depth-RGB Fusion For Cross-Domain Learning

MAI HÀ ĐẠT
dat.mh200135@sis.hust.edu.vn

Major: Data Science And Artificial Intelligent
Specialization: Data Science And Artificial Intelligent

Supervisor: Doctor Trịnh Văn Chiến _____

Signature

Department: Computer Engineering

School: School of Information and Communications Technology

HANOI, 07/2024

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my family, and my friends for their unwavering support throughout my graduate thesis journey. I am deeply grateful to my professors for their invaluable guidance and encouragement. Finally, I acknowledge my own dedication and perseverance, which have been crucial in achieving this accomplishment. Thank you all for contributing to this significant milestone in my academic career.

PLEDGE

Student's Full Name: Mai Hà Đạt
Student ID: 20200135
Contact Phone Number: 0354598619
Email: dat.mh200135@sis.hust.edu.vn
Class: Data Science And Artificial Intelligent Class 01
Program: Data Science And Artificial Intelligent

I, **Mai Hà Đạt**, commit that the Graduation Thesis (GT) is my own research work under the guidance of **Dr.Trịnh Văn Chiến**. The results presented in the GT are truthful and are my own achievements, not copied from any other works. All references in the GT, including images, tables, data, and quotations, are clearly and fully cited in the bibliography. I take full responsibility for any violations of the school's regulations concerning plagiarism.

Hà Nội, 07/2024

Student

ABSTRACT

In the dynamic landscape of industrial logistics, the precise identification of pallet conditions is pivotal for efficient supply chain operations. With recent advancements in computer vision and the affordability of cameras and depth sensors, many researchers have addressed the challenge of automated pallet classification in factory settings using machine learning and deep learning models with depth or RGB images as input. However, the limited datasets in this field have led to the use of synthetic data in the model training processes. This approach can lead to domain shift problems, causing the model to make incorrect predictions in real-world data cases. Therefore, this research primarily focus on building and training models using domain adaptation techniques to overcome domain shift problems. Additionally, the research utilize depth images as additional features to enhance the model's accuracy. The thesis proposes an approach that combines domain adaptation with depth and RGB image fusion techniques on a semi-supervised dataset. Specifically, the approach integrates domain adversarial training with combined fusion features after both RGB and depth feature extractors on a dataset consisting of labeled synthetic and unlabeled real-world depth images. This model overcomes challenges related to data scarcity and domain shift. Experiments have been conducted to demonstrate the effectiveness of this approach in classifying pallet states. The results showcase the model's adaptability to real-world variations, achieving increased accuracy compared to other benchmarks. The proposed method is expected to provide a vision understanding brain for automatic processes, helping to optimize resource allocation in warehouses. Additionally, it offers a solution for the integration of computer vision into logistics applications.

Student

(Signature and full name)

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION 1

1.1 Problem Statement 1

1.2 Background and Problems of Research 4

1.3 Research Objectives and Conceptual Framework 4

1.4 Contributions 5

1.5 Organization of Thesis 6

CHAPTER 2. LITERATURE REVIEW 7

2.1 Scope of Research 8

2.2 Theoretical Framework 9

2.2.1 Image Classification 9

2.2.2 Digital Image (Specify RGB Image) 10

2.2.3 Depth Image 11

2.2.4 Domain Adaptation 12

2.2.5 Depth and RGB Image Fusion 13

2.2.6 Pont Cloud Data 14

2.3 Related Works 15

2.3.1 Domain-Adversarial Training of Neural Networks 15

2.3.2 MiDaS Depth Estimation Models 17

2.3.3 Residual Network-ResNet 17

2.3.4 PointNet 18

2.4 Summary 19

CHAPTER 3. METHODOLOGY 22

3.1 DANN(Domain-Adversarial Training of Neural Networks) Model Structural Details 22

3.2 Modifications in the ResNet Backbone DANN Model 24
3.3 Depth-RGB Fusion DANN Model 27
3.3.1 Overall Model Structural 27
3.3.2 Depth Branch of Model. 28
3.4 Summary 30
CHAPTER 4. NUMERICAL RESULTS 31
4.1 Simulation Method 31
4.1.1 Evaluation Metrics 31
4.1.2 Dataset Annotation 33
4.1.3 Experiments Setup 36
4.2 Training and Testing the Models. 38
4.3 Summary 42
CHAPTER 5. CONCLUSIONS 43
5.1 Summary 43
5.2 Suggestion for Future Works 43
REFERENCE 48

LIST OF FIGURES

Figure 1.1 Pallet-The backbone of warehouse [4]. 2
Figure 1.2 Pallet helping warehouse operations [5]. 2
Figure 1.3 The predicted market in the Autonomous and Sensor Technology market [7]. 3
Figure 1.4 Pallet datasets in roboflows, which mostly about detection task [17]. 5

Figure 2.1 Color space of original images of a duck [29]. 10
Figure 2.2 Depth images representation of some color images [30]. . . . 11
Figure 2.3 The RGB-D features fusion after the feature extractors from both depth and RGB images [36]. 14
Figure 2.4 The example of a point cloud data of a donut [14]. 15
Figure 2.5 The proposed architecture for the DANN in the Domain-Adversarial Training of Neural Networks paper [39]. 16
Figure 2.6 The residual block architecture 18
Figure 2.7 The structure of the Pointnet [13]. 19

Figure 3.1 The ResNet backbone DANN structure. 25
Figure 3.2 The proposed DANN structure using both PointNet and ResNet backbone. 28

Figure 4.1 The formula between F1, Precision and Recall [41]. 32
Figure 4.2 The examples of three labels images data. 33
Figure 4.3 The proposed datasets. 35
Figure 4.4 The hard example of dataset. 36
Figure 4.5 The datasets distributions. 37
Figure 4.6 Class losses over epochs of proposed methods. 41
Figure 4.7 Domain losses over epochs of proposed methods. 41

LIST OF TABLES

Table 4.1 Data distribution table. 34
Table 4.2 Testing on test target dataset result table. 39
Table 4.3 Testing on hard examples dataset result table. 39

LIST OF ABBREVIATIONS

Abbreviation	Definition
ADDA	Adversarial Discriminative Domain Adaptation
CDAN	Conditional Adversarial Domain Adaptation
DANN	Domain Adversarial Neural Networks
SOTA	State of the art

CHAPTER 1. INTRODUCTION

This chapter walks through the background and problems in the warehouse pallet field. It discusses the most up-to-date solutions that have been applied, their issues, and how the thesis proposals solve them. Finally, a summary of the contributions of the thesis is presented.

1.1 Problem Statement

In today's industrial landscape, logistics plays a crucial role in ensuring smooth supply chain operations by facilitating the flow of goods from origin to destination. Fundamental tasks such as stacking, unloading, and intra-warehouse movements are essential in this process, with the precise identification of warehouse object's conditions being of utmost importance. Recent technological advancements, including the widespread availability of cost-effective cameras and the successful implementation of computer vision, revolutionize logistics operations [1]–[3]. The affordability of cameras and sensors leads to their widespread use in many warehouses. These computer vision advancements enable accurate identification and tracking of objects, particularly pallets, within the logistics environment.

Pallets play a crucial role in the warehouse industry. They are wooden structures, as shown in Figure 1.1. These flat structures are used to support goods in a stable manner while being lifted by forklifts, pallet jacks, or other handling equipment. Pallets make it easier to move, store, and organize products efficiently. Overall, pallets are essential for ensuring smooth and efficient operations in warehouses, leading to cost savings and improved productivity. Therefore, recognizing pallet conditions becomes an essential task in industrial automation.

The adoption of pallets revolutionized how warehouses operate. Before their invention, goods were typically stored in barrels, crates, or boxes, making them difficult to handle and organize. Pallets allowed for better organization, increased accessibility, and faster inventory turnover. Warehouses equipped with pallets can easily load, unload, and rearrange products, ensuring that items are readily available for shipment and reducing valuable downtime. In the storage and warehousing sector, space optimization is critical for cost-effectiveness. Pallets facilitate efficient use of vertical space through stacking, maximizing warehouse capacity by installing the racks as per the available space in both horizontally and vertically. Pallets also simplify inventory management, as they provide a standardized unit to count and track items. As shown in Figure 1.2, pallets help forklift to stabilize the goods while they are lifted.