

CHAPTER

5

HUMAN-OBJECT INTERACTION DETECTION

Jiali Zhang, Zuriahati Mohd Yunos, and Habibollah Haron

5.1 INTRODUCTION

The study intends to develop and empirically test a model explaining Industry 4.0. application capabilities, affect and influence the relationship between green supply chain management (GSCM) practices and sustainable performance through the moderating role of Industry 4.0 application. This section will introduce green practices and their impact on firm performance.

Human-object interaction (HOI) detection is crucial for understanding the scene, through which robots can more accurately perceive the environment using computer vision (Park et al., 2023). Unlike object and scene segmentation, HOI detection focuses on identifying interactions between people and objects in a scene, pinpointing the individuals and the objects involved, and the specific action category (Chen et al., 2023). This technology has broad applications, from human-computer interaction to service robotics and video surveillance, making precise HOI detection increasingly crucial (Li et al., 2022).

Figure 5.1 illustrates the object detection network's ability to identify individuals and objects in an image, emphasising a valid interaction between a person and a table. Nonetheless, current approaches link

humans with all objects, generating combinations such as human and plate, human and cup, human and wine, etc., without considering the interaction context. These non-interactive combinations outnumber the truly interactive <human, table> pairings. Figure 5.2 shows this imbalance creates issues during model training, hindering helpful learning. Gradient descent slows due to this imbalance, affecting model optimisation and increasing detection errors.

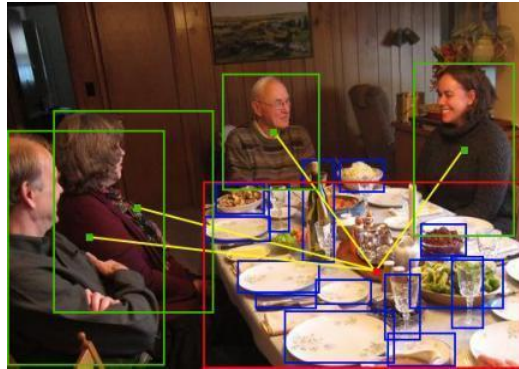


Figure 5.1 Illustrates network ability to current approaches detect all individuals and objects within an image



Figure 5.2 A substantial possibility of “ride” or “jump” interactions, as presented by the green bounding box rounding human above the red bounding box rounding the skateboard

A novel model called the interactivity recognition graph neural network (IRGNN) is proposed to enhance the effectiveness of recognising HOIs. Firstly, this approach will find valid human-object pairings before predicting interaction types. Section 5.2 presents relevant literature to organise this chapter. Section 5.3 describes the network model, while Section 5.4 discusses the results. Section 5.5 of this chapter provides the conclusion.

5.2 RELATED WORKS

Human-object interaction detection is vital in computer vision, enabling robots to perceive their environment in detail. Unlike essential object detection and similar tasks focusing on individual elements, it identifies interactions between people and objects, accurately pinpointing their positions and actions within a scene.

5.2.1 Human-object Interaction Detection

Computer vision performance has significantly increased by using deep learning (DL), and this method can automate feature extraction from massive datasets. Due to the availability of specific datasets for HOI detection, the challenge of HOI has progressed into a new development phase. To illustrate, Chao et al. (2018) have introduced the HO-RCNN model, a pivotal development in HOI. A structured output that incorporates attention mechanisms called instance-centric attention network (ICAN) was proposed by Gao et al. (2018) to enhance HOI identification utilising HO-RCNN by extracting contextual features, which refers to features related to the visual appearance of the local area. In addition, researchers are using graph models or graph convolutions to solve HOI problems. Liang et al. (2021) produced the VS-GATS, which is the two-graphic attentional network that takes in both conceptual and visual details and correctly separates them. The approach, nevertheless, exhausts all potential pairings of the image of humans and objects. Subsequently, the process faces many challenges in making calculations and obtaining accuracy.