



**HAL**  
open science

# Utilisation d'algorithmes d'intelligence artificielle pour l'identification et la localisation de colis et leur saisie par un torse humanoïde dans des opérations de dépalettisation

Santheep Yesudasu, Jean-François Brethé, Wafae Sebbata, Patrick Bonnin

## ► To cite this version:

Santheep Yesudasu, Jean-François Brethé, Wafae Sebbata, Patrick Bonnin. Utilisation d'algorithmes d'intelligence artificielle pour l'identification et la localisation de colis et leur saisie par un torse humanoïde dans des opérations de dépalettisation. Cahiers de la logistique, 2024, La robotisation dans les entrepôts logistiques (5). hal-04854409

**HAL Id: hal-04854409**

**<https://normandie-univ.hal.science/hal-04854409v1>**

Submitted on 23 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



**La robotisation dans  
les entrepôts  
logistiques**

**PÔLE INGÉNIEUR  
LOGISTIQUE  
LE HAVRE NORMANDIE**

---

**CAHIERS DE LA  
LOGISTIQUE**



**CAHIERS DE LA  
LOGISTIQUE  
CL2024 – N° 05**

**La robotisation dans les entrepôts logistiques**

Décembre, 2024  
Copyright © 2024, ISEL



La présente publication est protégée par le code de la propriété intellectuelle et plus précisément ses articles relatifs au respect du droit d’auteur.

Les auteurs demeurent seuls responsables du contenu de leur œuvre. Ils garantissent disposer des autorisations nécessaires de la part des tierces parties titulaires de droits sur des œuvres partiellement ou globalement reproduites. Les droits moraux et patrimoniaux sur l’œuvre demeurent attachés à leurs auteurs.

Afin de faciliter le partage et l’utilisation de leur création, les auteurs autorisent les utilisateurs à :

- Télécharger et imprimer une copie de la présente publication à des fins d’études, de recherche, de diffusion de la culture scientifique et de partage des connaissances ;
- Distribuer gratuitement et sans aucune contrepartie l’URL identifiant la publication.

La présente publication, et l’URL associée ne peuvent en aucun cas être distribuées ou utilisées dans le cadre d’une activité à but lucratif ou à des fins commerciales. Aucune modification du contenu de l’œuvre n’est autorisée.

# UTILISATION D'ALGORITHMES D'INTELLIGENCE ARTIFICIELLE POUR L'IDENTIFICATION ET LA LOCALISATION DE COLIS ET LEUR SAISIE PAR UN TORSE HUMANOÏDE DANS DES OPÉRATIONS DE DÉPALETTISATION

Santheep YESUDASU<sup>1</sup>, Jean-François BRETHÉ<sup>1</sup>, Wafae SEBBATA<sup>1</sup>, Patrick BONNIN<sup>2</sup>

<sup>1</sup>GREAH, Université Le Havre Normandie, France  
santheep.yesudasu@univ-lehavre.fr  
jean-francois.brethe@univ-lehavre.fr

<sup>2</sup>ISTY, University of Versailles Mantes-la-Ville, France  
patrick.bonnin@gmail.com

## RESUME

Dans le contexte Logistique 4.0, la robotisation des opérations de manutention se développe mais doit encore résoudre certains verrous scientifiques pour réussir à remplacer la dextérité des opérateurs humains. L'utilisation d'un torse humanoïde s'est révélé une solution pertinente pour la saisie de colis de taille variable. Cependant la robotisation complète des opérations de dépalettisation par un torse humanoïde reste un problème complexe lorsque les cartons sont disposés en couche et juxtaposés. Il s'agit dans un premier temps d'identifier et localiser les cartons unité par unité pour pouvoir les manipuler par la suite. Les travaux présentés dans cet article montrent que l'approche usuelle par analyse d'image atteint ses limites, puis introduit l'utilisation de l'intelligence artificielle et de l'apprentissage profond sur les données capteurs fournies par une caméra RGD-D. La conception et la mise en œuvre de cette méthode sont détaillées et les résultats expérimentaux sont présentés.

**MOTS CLÉS:** Dépalettisation robotisée, Torse humanoïde, Apprentissage profond, Vision par ordinateur, Intelligence artificielle.

## ABSTRACT

In the context of Logistics 4.0, the robotisation of handling operations is developing but still needs to resolve specific scientific issues to succeed in replacing the dexterity of human operators. The use of a humanoid torso has proved to be a relevant solution for picking up parcels of varying sizes. However, the complete robotisation of depalletising operations by the humanoid torso remains a complex problem when the boxes are set up in layers and juxtaposed. First of all, the boxes have to be identified and located unit by unit so that they can then be handled. The work presented in this paper shows that the image analysis approach is reaching its limits and introduces the use of artificial intelligence and deep learning on the sensor data provided by an RGD-D camera. The design and implementation of this method are detailed and the experimental results are presented.

**KEYWORDS:** Depalletising Robot, Humanoid torso, Deep learning, Computer vision, Artificial intelligence.

# 1. Introduction

La logistique est en pleine mutation et intègre les concepts de l'Industrie 4.0 dans ses processus (Echelmeyer et al., 2008). Les nombreuses tâches de manutention nécessaires au fonctionnement des process de préparation de commandes sont aujourd'hui robotisées partiellement (taux estimé à 30 %). On estime que dans 10 ans, ce seront 70 % des tâches de manutention qui seront robotisées. Ceci permettra de diminuer le nombre d'accidents du travail et d'arrêts maladies, et soulagera les opérateurs victimes de Troubles Musculo-Squelettiques (TMS). La difficulté à recruter les ressources humaines dans le secteur est aussi un autre facteur qui pousse à une automatisation plus rapide des opérations de manutention.

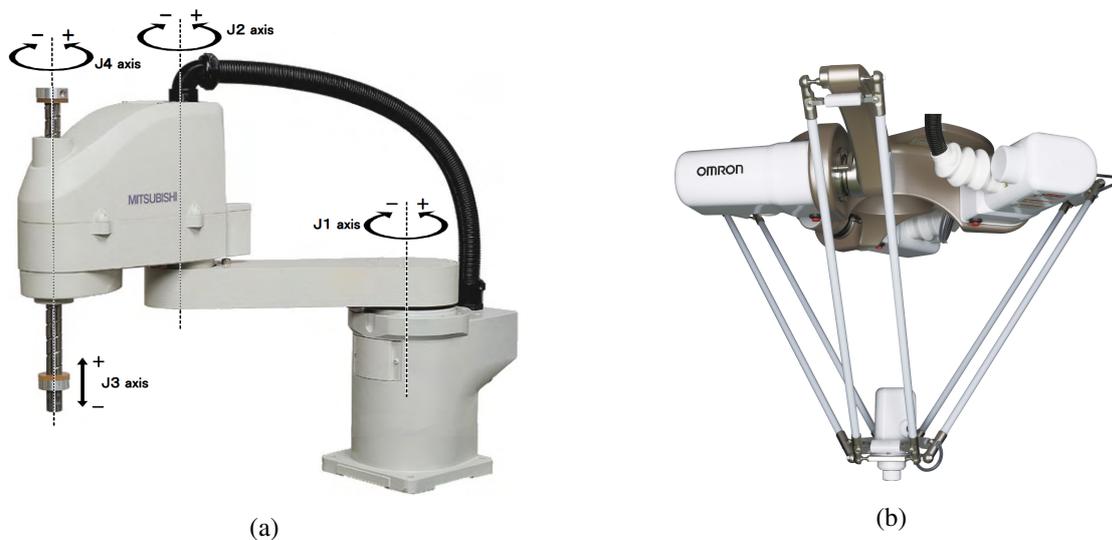


FIGURE 1 – (a) SCARA (Selective Compliant Assembly Robotic Arm) (b) bras robotiques parallèles (Delta)

Les opérations de saisie et de manutention robotisée de cartons sont majoritairement effectuées par des bras manipulateurs série de type anthropomorphes à 6 degrés de liberté, ou SCARA (Selective Compliant Assembly Robotic Arm) à 4 degrés de liberté, ou des bras robotiques parallèles (Delta, Quattro, etc) présentés sur la Figure 1b. Les préhenseurs utilisés sont généralement des systèmes de type ventouses (où la dépression est obtenue par utilisation d'une pompe à vide ou par effet Venturi à partir d'une source d'air comprimé), de type pinces à deux doigts ou des préhenseurs conçus spécifiquement pour un objet donné. Cependant dans le secteur de la préparation de commandes, ces bras sont confrontés à des objets de taille, de géométrie, de texture et de masse variable et ne peuvent pas toujours effectuer la manutention de cette grande variété d'objets. Or l'observation du comportement des opérateurs humains montre qu'ils sont souvent amenés à utiliser leurs deux bras simultanément pour saisir les objets. Ceci nous a amené à mettre en œuvre un torse humanoïde pour la saisie par les faces latérales de cartons de taille variable en mettant en œuvre un algorithme de commande spécifique (Benali et al., 2019) pour d'une part, exercer le juste effort empêchant le glissement sans écraser l'objet et d'autre part, assurer une coordination efficace des mouvements des deux bras.

Dans les travaux présentés ici, nous travaillons sur la dépalettisation automatisée effectuée par le torse humanoïde. Dans ce cas, les cartons présents sur la palette sont placés les uns à côté

des autres en couches superposées et ne sont plus isolés comme dans le cas de la thèse citée précédemment. Les géométries des cartons ne sont pas connues a priori. Nous sommes confrontés à deux verrous scientifiques : l'un concerne l'exploitation efficace des informations issues de la caméra RGB-D pour identifier les géométries et localiser unitairement les cartons ; l'autre la conception de préhenseurs adaptés et la mise en œuvre d'une série de mouvements de manipulation pour pouvoir saisir et transporter les cartons unitairement pour les déposer par exemple sur un convoyeur. Dans cet article, nous présentons les résultats obtenus relatifs à l'identification et la localisation. Après avoir dressé un état de l'art permettant de cerner les limites des techniques d'analyse d'images classique, nous nous intéressons aux potentialités offertes par les dernières méthodes d'intelligence artificielle de type apprentissage profond. Nous proposons par la suite une méthodologie de mise en œuvre de ces méthodes pour lever le verrou scientifique étudié. Enfin, nous présentons les résultats expérimentaux obtenus avec le torse humanoïde.

## 2. Etat de l'art

La détection d'objets basée sur la vision est un domaine largement étudié dans le domaine de la vision par ordinateur, et de nombreuses études existantes se sont concentrées sur la résolution des différents défis et limitations associés à ce type de détection. Ces défis peuvent inclure la sensibilité aux conditions d'éclairage et la difficulté à détecter les objets à faible texture. Les systèmes existants (Katsoulas et al., 2002), (Monica et al., 2020), (Prasse et al., 2011) ont ajouté un capteur supplémentaire tel que ToF (Time-of-Flight) et RFID (Radio-Frequency IDentification). Certains systèmes de vision 3D nécessitent toujours les dimensions de l'objet à détecter en entrée, bien que cette exigence diminue progressivement à mesure que la technologie évolue. En plus des systèmes de vision 3D, d'autres méthodologies (Caccavale et al., 2020) peuvent également être utilisées dans un système d'étiquetage et de dépalettisation. Une telle approche consiste à étiqueter les boîtes en carton cibles et à utiliser un système de dépalettisation hiérarchiquement structuré. Cette méthode implique d'appliquer des étiquettes aux boîtes contenant des informations sur le contenu, l'emplacement et la destination de la boîte. Le système d'étiquetage peut être intégré à un système de dépalettisation utilisant une approche hiérarchique pour retirer les boîtes de la palette et les trier dans leurs destinations finales.

Le système basé sur la vision mentionné dans (Nakamoto et al., 2016) peut utiliser diverses techniques de traitement d'image telles que l'extraction de contours et la détection de motifs à partir d'images lumineuses. Ces techniques sont utilisées pour extraire des informations pertinentes des images d'objets et les utiliser pour les identifier et les classer. L'extraction de contours 3D basée sur le nuage de points est une autre méthode bien connue qui peut être utilisée dans les systèmes basés sur la vision, comme décrit dans (Vohra et al., 2021). Cependant, cette approche ne fonctionne que lorsque les dimensions de l'objet sont connues. En d'autres termes, le système doit avoir une connaissance préalable de la taille et de la forme de l'objet pour extraire précisément ses contours 3D à partir des données de nuage de points. Cela peut limiter la polyvalence et la mise à l'échelle du système, car il peut ne pas être capable de manipuler une large gamme d'objets sans informations supplémentaires.

YOLO (You Only Look Once) (Redmon et al., 2016) est une méthode de détection d'objets en temps réel bien connue pour sa vitesse de traitement rapide et sa haute précision. En 2015, YOLOv1 a été introduit avec une approche matricielle performante pour la détection d'objets dans les images mais présentant des performances plus faibles lorsque les objets sont petits. YOLOv2 a amélioré les performances avec des boîtes d'ancrage, une normalisation par lots et une architecture de réseau neuronal plus puissante. En 2018, YOLOv3 a introduit une architecture de réseau plus grande et d'autres améliorations, atteignant une précision de pointe et une vitesse d'inférence plus rapide. YOLOv4 a ajouté une architecture de réseau neuronal plus puissante et un regroupement de boîtes basé sur des ancrages, atteignant une précision encore plus élevée et une vitesse d'inférence plus rapide. YOLOv5 est une réécriture complète avec une nouvelle architecture de réseau neuronal et une approche d'entraînement, atteignant une précision encore plus élevée et une vitesse d'inférence plus rapide que YOLOv4 (Jiang et al., 2022).

Dans la recherche de (Redmon et al., 2016) (Reyes et al., 2019) (Redmon and Farhadi, 2018) YOLOv3 utilise un seul réseau neuronal pour prédire plusieurs boîtes englobantes et les probabilités de classe associées pour chaque objet dans l'image. Le score d'objectivité pour chaque boîte englobante est déterminé par régression linéaire, et représente la probabilité que la boîte contienne un objet d'intérêt. Les scores de classification et les coordonnées de la boîte englobante sont combinés avec le score d'objectivité pour produire la prédiction finale pour chaque objet dans l'image. Cette approche permet à YOLOv3 d'effectuer la détection et la reconnaissance d'objets en temps réel avec une grande précision. Il utilise également une combinaison de convolutions 3x3 et 1x1 dans l'architecture Darknet-53 pour l'extraction de caractéristiques, qui est plus rapide que ResNet-101 et similaire en performance à ResNet-152. De plus, YOLOv3 surpasse les versions précédentes de YOLO et d'autres modèles de détection d'objets en termes de précision moyenne et de valeurs d'intersection sur l'union. Comme prouvé dans (Redmon et al., 2016) (Reyes et al., 2019) (Redmon and Farhadi, 2018) les résultats d'entraînement et de test basés sur le darknet montrent qu'ils ont une meilleure capacité dans YOLOv3, qui s'exécute de manière significativement plus rapide que d'autres méthodes de détection en temps réel avec des performances comparables. Dans la recherche (Sebbata et al., 2020) une application robotique utilisant YOLOv3 a été développée pour identifier des objets de même texture mais présentant différentes géométries. L'approche de (Schwarz et al., 2018) utilise l'apprentissage par transfert pour obtenir la puissance des réseaux neuronaux profonds dans la robotique avec des quantités minimales de données d'entraînement. En apprentissage profond, les couches de départ sont chargées d'identifier les caractéristiques significatives. Ensuite, seules les couches finales, qui effectuent des tâches avancées comme la détection ou la segmentation doivent être affinées.

Le travail présenté dans cet article est conçu pour surmonter les défis posés par les conditions d'éclairage pour la reconnaissance de boîtes polyvalentes sans texture (géométrie en forme de pavé droit) dans les opérations de dépalettisation en temps réel. En se concentrant sur la sensibilité des conditions d'éclairage standard et faible et des boîtes alignées étroitement, le système vise à fournir une détection rapide et précise des objets dans ces environnements difficiles. L'utilisation d'un torse humanoïde industriel avec des bras doubles UR10 est un choix intéressant car il offre un niveau élevé de flexibilité et de polyvalence dans la manipulation des objets. Il est important de noter que la détection d'objets en temps réel peut être une tâche complexe, car elle implique le traitement de grandes quantités de données en temps réel et la prise de décisions rapides basées sur l'apprentissage profond. Dans l'ensemble, la conception de ce système garantit

que le système peut répondre aux spécifications de performance requises.

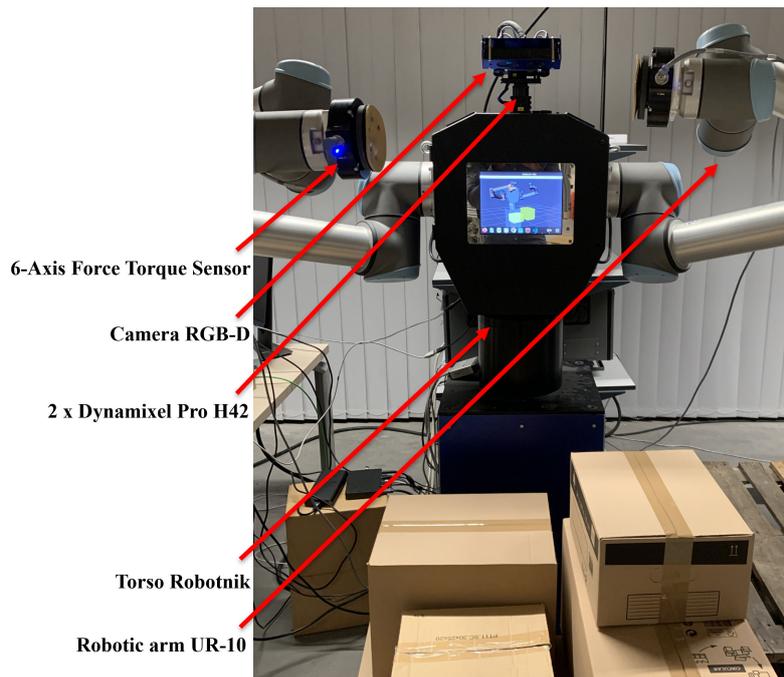


FIGURE 2 – Torse humanoïde pour la saisie et la manipulation.

### 3. Méthodologie

Cette recherche porte sur le développement d'un système automatisé de dépalettisation alimenté par l'IA pour des objets parallélépipédiques. Cet article est principalement axé sur le développement de la technologie de détection d'objets et d'apprentissage par transfert, en mettant particulièrement l'accent sur la résolution des défis précédemment abordés. De plus, l'article explore les avancées dans l'estimation de la pose et l'algorithme de dépalettisation pour les bras doubles.

La méthodologie décrite dans la Figure 3 utilise une seule caméra RGB-D et Convolutional Neural Network (CNN) basé sur YOLOv3 pour détecter les boîtes en carton sur une palette. Les boîtes englobantes détectées sont ensuite utilisées pour sélectionner une région d'intérêt dans les nuages de points, et Darknet 3D est utilisé pour estimer la pose de l'objet. Ce processus utilise les données de profondeur obtenues à partir de la boîte englobante 2D de YOLOv3. L'algorithme du système s'adapte au succès de la détection et continue de recueillir plusieurs positions de boîtes, ce qui permet une estimation précise de la pose. L'algorithme sélectionne la boîte avec le taux de précision de détection le plus élevé et la proximité la plus proche du bras double pour le processus de prélèvement, même lorsque les boîtes sont partiellement occultées par d'autres boîtes. Afin d'atteindre et de manipuler efficacement un objet, le manipulateur à bras double détermine d'abord la stratégie de préhension optimale pour l'objet, puis planifie le chemin le plus approprié pour l'atteindre. Après l'exécution du processus de dépalettisation, une base de données est générée, qui peut être utilisée pour gérer de manière plus efficace les poses complexes des boîtes et améliorer les processus de dépalettisation futurs.

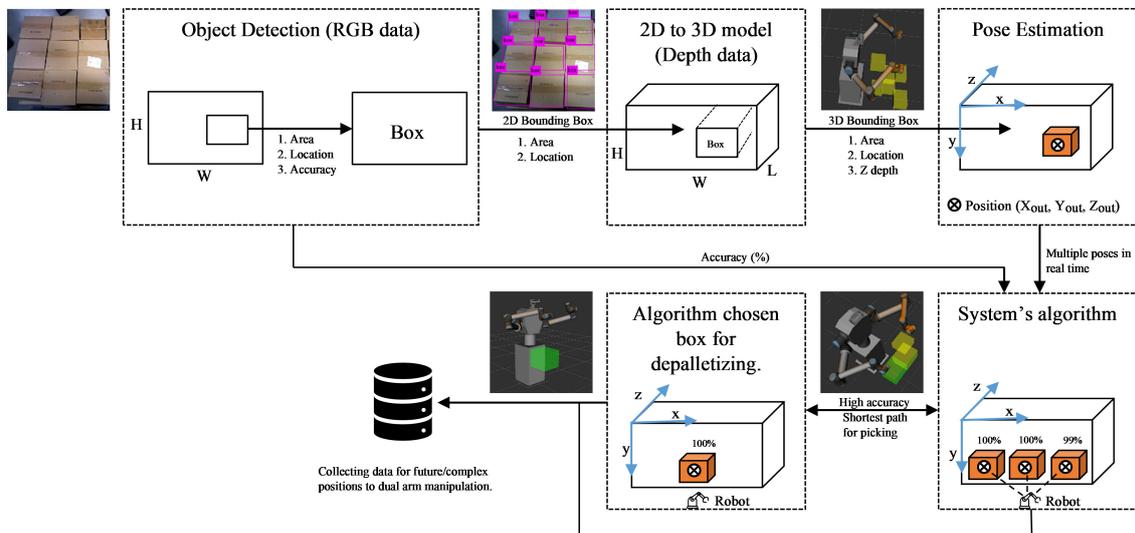


FIGURE 3 – Schéma de principe du processus global

### 3.1 Détection d'objets en 2D

Le modèle CNN YOLOv3, conçu avec une architecture d'entraînement à un seul stade, a été utilisé pour reconnaître avec précision les objets dans les images RGB. Au début, notre modèle a été entraîné en utilisant des images obtenues exclusivement dans notre environnement de dépalettisation. Cependant, afin d'améliorer son adaptabilité à des environnements divers, nous avons incorporé des images de boîtes en carton courantes dans l'ensemble de données. L'ensemble de données se compose de boîtes ("objets") étroitement annotées avec des étiquettes prioritaires pour des processus de prélèvement efficaces dans notre environnement. De plus, la surface supérieure des boîtes a été prise en compte pour détecter quand la couche supérieure de la palette était complètement remplie de boîtes en une couche. Pendant la tâche de dépalettisation, la caméra diffusait des images en temps réel, qui étaient traitées par le modèle CNN pour extraire la zone sous forme d'une boîte englobante en 2D.

### 3.2 Traitement des nuages de points 3D

La méthode proposée convertit les boîtes englobantes 2D en coordonnées 3D en cartographiant la relation géométrique entre l'image et le nuage de points. La région cartographiée est ensuite filtrée pour extraire la Region of Interest (ROI) optimale de l'objet. Cela est réalisé en calculant les valeurs minimales et maximales des coordonnées  $x$ ,  $y$  et  $z$  de ces points ( $x_{min}$ ,  $y_{min}$ ,  $z_{min}$ ,  $x_{max}$ ,  $y_{max}$ ,  $z_{max}$ ), formant une boîte englobante 3D dans le nuage de points. Avec ces valeurs, la surface, l'emplacement et la profondeur de l'objet dans l'espace 3D peuvent être estimés. L'étape de détection d'objet est essentielle pour garantir l'exactitude et la précision du traitement des nuages de points 3D, tout en réduisant simultanément l'impact du désordre causé par d'autres boîtes en carton dans l'environnement.

### 3.3 Estimation de la pose

Après avoir extrait la partie 3D du nuage de points, elle est traitée pour estimer la pose de la boîte englobante spécifique en vue de la manipulation et de la préhension. Une boîte englobante 3D est utilisée pour localiser l'objet dans le cadre de la palette, et le point central des coordonnées dédiées de l'objet est utilisé pour l'estimation de la pose. Ce processus permet la détection et l'estimation en temps réel de plusieurs objets dans le cadre considéré.

### 3.4 Algorithme du système

Cet algorithme a été créé pour prédire la priorité de la dépalettisation des boîtes en fonction de l'exactitude de la prédiction du CNN et de la boîte prédite la plus proche dans le tableau. Les coordonnées cartésiennes de chaque détection sont stockées dans le tableau avec une grille unique basée sur son taux de précision. Simultanément, l'algorithme traite ces coordonnées pour localiser la boîte la plus proche de la position réelle du Tool Center Point (TCP) des deux bras. Pour ce faire, nous utilisons un algorithme de force brute (Sandström et al., 2020), qui accélère la recherche de paires proches pour la planification de mouvement et l'optimisation de trajectoire. Cet algorithme utilise les coordonnées X, Y et Z du centre de chaque boîte pour identifier le point le plus proche du TCP en utilisant la distance euclidienne. Une fois que nous avons calculé les distances entre toutes les paires de points, l'Algorithme 1 peut trouver la plus petite distance entre un point et un autre. Cet algorithme utilise des planificateurs basés sur l'échantillonnage pour éviter le cas d'utilisation où le bras double a des contraintes et des limitations d'accessibilité pour effectuer les mouvements souhaités.

### 3.5 Dépalettisation

La planification de trajectoire des bras est réalisée par MoveIt (Sachin et al., 2012), progiciel de planification de mouvement basé sur ROS, comprenant Kinematics and Dynamics Library (KDL) et Open Motion Planning Library (OMPL) gérant aussi l'évitement d'obstacles. L'algorithme de dépalettisation utilise les résultats de détection d'objet et d'estimation de pose pour planifier la trajectoire du bras robotique et exécuter la tâche de dépalettisation.

---

**Algorithm 1** Optimize closest coordinate with Brute Force Algorithm

---

**Input :**  $P$  list of coordinates

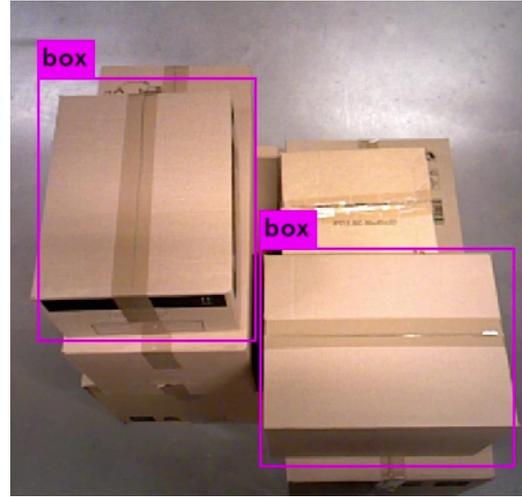
**Output :** Indices index1 and index2 of the closest pair of coordinates

```
 $dmin \leftarrow \infty$ 
1: for  $i \leftarrow 1$  to  $n-1$  do
2:   for  $j \leftarrow i+1$  to  $n$  do
3:      $d \leftarrow \text{sqrt}((x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2)$ 
4:     if ( $d < dmin$ ) then
5:        $dmin \leftarrow d; index1 \leftarrow i; index2 \leftarrow j$ 
6:     end if
7:   end for
8: end for
9: return  $index1, index2$ 
```

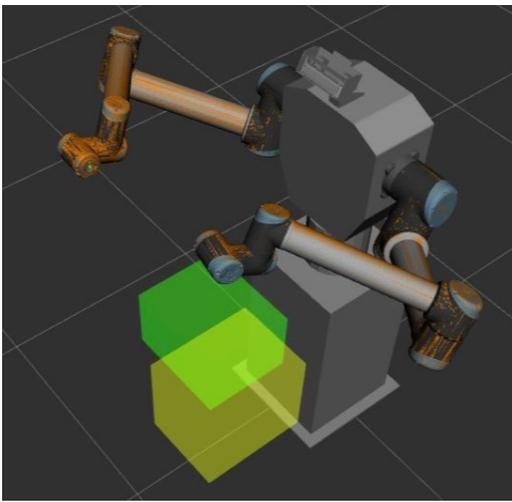
---



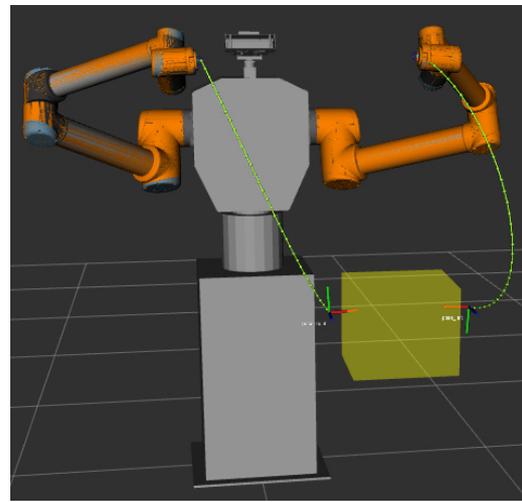
(a)



(b)



(c)



(d)

FIGURE 4 – (a) Boîtes à détecter (b) Boîtes englobantes 2D, (c) Nuage de points 3D, (d) Boîte à saisir en priorité.

## 4. Expériences

Le système robotique comporte deux bras UR10 montés sur un torse Robotnik, possédant des capteurs de force/couple à six axes et une caméra RGB-D à vision à 360 degrés (Asus XTION). Les mouvements du robot sont contrôlés par MoveIt, un package logiciel open source largement utilisé pour la planification de mouvement, la manipulation et le contrôle des systèmes robotiques. MoveIt fournit un contrôleur de haut niveau pour le robot, qui peut être programmé en utilisant C++ et Python via la distribution et les packages ROS Noetic. L'unité centrale est un processeur Intel Core i7-10875H haute performance, de 32 Go de RAM, d'un GPU NVIDIA Quadro RTX4000 avec 8 Go de mémoire, de Cuda 11.1, de cuDNN v8, et du système d'exploitation Ubuntu 20.04.

### 4.1 Création et entraînement de notre ensemble de données

Nous avons rassemblé une collection d'images de notre environnement en utilisant une caméra robotique, et avons également inclus des images de boîtes en carton courantes. Enfin, nous

avons annoté manuellement 807 images avec la classe 'boîte' en utilisant l'outil LabelImg. Plus précisément, nous nous sommes concentrés sur l'étiquetage des boîtes visibles en 3D et de la surface supérieure de ces boîtes. Notre priorité a été d'étiqueter les boîtes visibles en 3D tandis que l'étiquetage de la face supérieure minimale de la boîte a été de moindre priorité. Dans les cas où il existe deux possibilités pour la visibilité de la face supérieure et de la boîte en forme de 3D, nous avons constaté que la détection des boîtes de la face supérieure est moins précise que celle des boîtes en forme de 3D. Pour garantir une meilleure précision dans la détection et l'estimation de la pose, nous avons évité d'étiqueter à partir de l'autre possibilité de la portion visible de la boîte pendant l'annotation.

Notre ensemble de données comprend des images avec une résolution de  $608 \times 608$  pixels. Pour l'entraînement, nous avons fixé la taille du lot à 64 et spécifié le nombre maximum d'itérations à 4000, avec des étapes intermédiaires à 3200 et 3600. L'outil annoté utilisé pour l'étiquetage extrait des fichiers texte au format  $\langle classe - objet \rangle \langle x\_centre \rangle \langle y\_centre \rangle \langle largeur \rangle \langle hauteur \rangle$  à partir des images annotées. Nous avons utilisé un ratio de 80 :20 pour l'entraînement et le test. Notre formation a été réalisée en utilisant le GPU Colab avec les poids initiaux de darknet53.conv.74.

## 5. Résultats et discussions

Comme le montre la Figure 4, ce système de détection d'objets fonctionne sur deux boîtes inconnues situées en haut d'un ensemble, avec seulement deux faces visibles disponibles pour la détection. Un CNN en temps réel est utilisé pour prédire la classe des boîtes, ce qui donne une représentation visuelle avec une boîte englobante 2D. Ces boîtes englobantes 2D sont ensuite combinées avec des données de nuage de points et affichées dans le logiciel Rviz, fournissant une visualisation en 3D des objets prédits. L'algorithme du système choisit la boîte prioritaire, qui présente une précision plus élevée et est la plus proche de la boîte en fonction de l'Algorithme 1 : Optimisation de la coordonnée la plus proche avec l'algorithme de force brute. Comme le montre la Figure 4 (a), la boîte sélectionnée est au sommet de l'ensemble et la plus proche de la prise pour les bras doubles.

### 5.1 Test de détection et validation

Les performances vont être testées et validées sur la base des métriques de précision, de rappel et de score F (Goutte and Gaussier, 2005). Ces mesures sont utilisées pour évaluer l'efficacité de ce système dans la détection d'objets et comparer différentes méthodes.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

La précision mesure les prédictions vraies positives, tandis que le rappel mesure les vrais positifs trouvés. Une précision élevée est liée à un faible taux de faux positifs, tandis qu'un rappel

élevé est lié à un faible taux de faux négatifs. Un score de 1.0 pour la précision ou le rappel indique une classification ou une détection parfaite pour cette classe.

TABLE 1 – Score F1

Visible face of box to the detection(FDbox)	<i>F1-Score</i>
Single FDbox	0.81
Multi-FDbox	0.90
Single FDbox when Multi-FDbox	0.25
Multi-FDbox when Single FDbox	0.72
Mixed FDbox	0.82

Le Tableau 1 montre que lors de la phase de test en temps réel, à la fois les boîtes à une seule face et les boîtes avec plus d'une face ont obtenu des scores F1 de 0.81 et 0.90, respectivement. Ces scores sont restés constants quelles que soient les conditions d'éclairage normales et faibles. Le CNN a atteint une précision de détection maximale lors de la détection de boîtes avec au moins deux faces visibles (FDbox). Cependant, même lorsque seule une face de la boîte était visible dans l'ensemble complet de palettes, le CNN pouvait toujours détecter la boîte, bien que la précision soit plus faible que lorsque plusieurs faces étaient visibles. Cela s'explique par le fait que plusieurs faces visibles contiennent plus d'informations contextuelles. En revanche, lorsque seule une face de la boîte est visible, le CNN a plus de difficulté à la différencier du fond de la scène ou d'autres objets similaires. Sur la base des résultats globaux, il ressort que même lors de combinaison de différents scénarii (boîtes alignées étroitement, boîtes situées dans différentes couches et positions, conditions d'éclairage variables), le CNN donne des réponses précises, comme illustré dans la Figure 5.



FIGURE 5 – Résultats de la détection dans différentes situations

## 5.2 Validation des poses prédites

La précision de l'estimation de la pose est évaluée en réalisant une prédiction de boîtes englobantes 2D sur un ensemble de données annotées, ce qui implique de normaliser les valeurs de pixels et les coordonnées 3D. Les paramètres du modèle sont optimisés pour minimiser la différence entre les points clés prédits et les vérités terrain. L'évaluation en temps réel des boîtes détectées dans la couche supérieure de la palette montre que les coordonnées estimées ont une erreur de moins de  $1\text{cm}$  sur tous les axes. Cette erreur est due au chevauchement des boîtes englobantes causé par l'orientation angulaire de la caméra. Cependant, cette erreur est attendue et peut être corrigée en utilisant cette distance pour initier la préhension. Comme l'objet (les boîtes) a une hauteur minimale de  $20\text{cm}$ , l'erreur dans la direction  $z$  est négligeable. L'ensemble de la tâche est effectué dans un espace de travail non figé.

## 6. Conclusion

Cette recherche présente un système de dépalettisation automatisé à double bras humanoïde pour la saisie de boîtes en carton supervisé par l'IA. Le système surmonte les défis de la détection d'objets, de l'estimation de la pose et des algorithmes de dépalettisation pour les manipulateurs à double bras. Le système utilise une seule caméra RGB-D et YOLOv3 pour la détection d'objets et darknet 3D pour l'estimation de la pose. Le système peut détecter et estimer avec précision la pose des boîtes en carton, même dans des scénarios difficiles avec des boîtes partiellement visibles ou occultées et des conditions d'éclairage variables. Bien que la précision soit plus élevée pour les boîtes avec plusieurs faces visibles, le système est toujours capable de détecter les boîtes avec seulement une face visible. Le manipulateur à double bras peut saisir et manipuler efficacement les boîtes avec des stratégies de préhension optimales et des chemins appropriés. Le système génère également une base de données pour manipuler les poses complexes des boîtes de manière plus efficace dans les futurs processus de dépalettisation. Globalement, les performances du système sont fiables et précises, atteignant une précision de détection élevée même dans des conditions d'éclairage variables et avec des boîtes partiellement visibles. Cette recherche démontre le potentiel de l'automatisation alimentée par l'IA pour le processus de dépalettisation, offrant des avantages significatifs pour les industries qui traitent des marchandises palettisées.

## Références

- Benali, K., Brethé, J.-F., Guerin, F., Gorka, M., and Sayah, A. (2019). A humanoid torso for palletization in logistics warehouses. In *12th International Doctoral Student Workshop on Logistics*.
- Caccavale, R., Arpentì, P., Paduano, G., Fontanelli, A., Lippiello, V., Villani, L., and Siciliano, B. (2020). A flexible robotic depalletizing system for supermarket logistics. *IEEE Robotics and Automation Letters*, 5(3) :4471–4476.
- Echelmeyer, W., Kirchheim, A., and Wellbrock, E. (2008). Robotics-logistics : Challenges for automation of logistic processes. In *2008 IEEE International Conference on Automation and Logistics*, pages 2099–2103. IEEE.

- Goutte, C. and Gaussier, E. (2005). A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European conference on information retrieval*, pages 345–359. Springer.
- Jiang, P., Ergu, D., Liu, F., Cai, Y., and Ma, B. (2022). A review of yolo algorithm developments. *Procedia computer science*, 199 :1066–1073.
- Katsoulas, D., Bergen, L., and Tassakos, L. (2002). A versatile depalletizer of boxes based on range imagery. In *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*, volume 4, pages 4313–4319. IEEE.
- Monica, R., Aleotti, J., and Rizzini, D. L. (2020). Detection of parcel boxes for pallet unloading using a 3d time-of-flight industrial sensor. In *2020 Fourth IEEE International Conference on Robotic Computing (IRC)*, pages 314–318. IEEE.
- Nakamoto, H., Eto, H., Sonoura, T., Tanaka, J., and Ogawa, A. (2016). High-speed and compact depalletizing robot capable of handling packages stacked complicatedly. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 344–349. IEEE.
- Prasse, C., Skibinski, S., Weichert, F., Stenzel, J., Müller, H., and Ten Hompel, M. (2011). Concept of automated load detection for de-palletizing using depth images and rfid data. In *2011 IEEE International Conference on Control System, Computing and Engineering*, pages 249–254. IEEE.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once : Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.
- Redmon, J. and Farhadi, A. (2018). Yolov3 : An incremental improvement. *arXiv preprint arXiv :1804.02767*.
- Reyes, E., Gómez, C., Norambuena, E., and Ruiz-del Solar, J. (2019). Near real-time object recognition for pepper based on deep neural networks running on a backpack. In *RoboCup 2018 : Robot World Cup XXII 22*, pages 287–298. Springer.
- Sachin, C., Ioan, S., and Steve, C. (2012). Moveit![ros topics]. *IEEE Robotics and Automation Magazine*, 19(1) :18–19.
- Sandström, R., Denny, J., and Amato, N. M. (2020). Asymptotically-optimal topological nearest-neighbor filtering. *IEEE Robotics and Automation Letters*, 5(4) :6916–6923.
- Schwarz, M., Milan, A., Periyasamy, A. S., and Behnke, S. (2018). Rgb-d object detection and semantic segmentation for autonomous manipulation in clutter. *The International Journal of Robotics Research*, 37(4-5) :437–451.
- Sebbata, W., BRETHÉ, J.-F., et al. (2020). An adaptive robotic grasping with a 2-finger gripper based on deep learning network. In *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, volume 1, pages 620–627. IEEE.
- Vohra, M., Prakash, R., and Behera, L. (2021). Edge and corner detection in unorganized point clouds for robotic pick and place applications. *arXiv preprint arXiv :2104.09099*.

## Biographie des auteurs



Santheep Yesudasu est doctorant en robotique à l'Université Le Havre Normandie au laboratoire GREAH (Groupe de Recherche en Électrotechnique et Automatique du Havre). Ses recherches se concentrent sur le développement de l'automatisation basée sur l'IA pour améliorer la détection et l'estimation de pose des robots humanoïdes dans les applications logistiques. Depuis 2023, il travaille également comme ingénieur de recherche au laboratoire CRISTAL UMR CNRS 9189, associé à l'Université de Lille.



Jean-François Brethé est Professeur des Universités en robotique à l'Université Le Havre Normandie, Institut Supérieur d'Études Logistiques (ISEL) et membre du GREAH (Groupe de Recherche en Électrotechnique et Automatique du Havre). Ses travaux de recherche portent initialement sur la modélisation, la conception et l'étude des performances en précision des robots manipulateurs industriels, puis se focalisent aujourd'hui sur les verrous scientifiques posés par l'introduction de cobots dans les entrepôts logistiques (préparation de commandes robotisée par robot manipulateur mobile, saisie de colis de taille variable par torse humanoïde).



Wafae SEBBATA est docteure en robotique. Sa thèse de doctorat, soutenue en octobre 2021, intitulée "Étude et développement de tâches de manipulation mobile dans le contexte de l'industrie 4.0", a été réalisée au sein du laboratoire GREAH (Groupe de Recherche en Électrotechnique et Automatique du Havre), Université Le Havre Normandie, sous la direction du Pr Jean-François Brethé. Elle a travaillé comme ingénieure d'études en robotique dans le cadre du projet CoRoT de 2017 à 2022. Aujourd'hui, elle exerce comme enseignante à l'école d'ingénieurs ECE-Paris "École Centrale d'Électronique".



Patrick Bonnin est Professeur des Universités à l'ISTY (Institut des Sciences et Techniques des Yvelines), École d'Ingénieurs de l'Université de Versailles Saint Quentin, au Département Systèmes Electroniques Embarqués Il effectue sa recherche au LISV (Laboratoire d'Ingénierie des Systèmes de Versailles) dans l'équipe Robotique Interactive. Ses thématiques de recherche sont au cœur des problématiques de vision temps réel dans les systèmes embarqués.