

الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

المدرسة العليا للإعلام الآلي - 08 ماي 1945 - بسيدي بلعباس
Ecole Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès



Mémoire de Fin d'étude

Pour l'obtention du diplôme d'ingénieur d'état

Filière : **Informatique**

Spécialité : **Ingénierie des Systèmes Informatiques (ISI)**

Thème

Mise en œuvre d'un modèle de détection d'objets en temps réel
intégrant les connaissances antérieures

Présenté par :

- Mr Omari Souhil

Soutenu le : **13/09/2023**

Devant le jury composé de :

- | | |
|---------------------------|--------------|
| - M Chaib Souleyman | Président |
| - M Belfedhal Alaa Eddine | Encadreur |
| - Mme Taouli Amina | Examinatrice |

Année Universitaire : 2022 / 2023

Remerciements

En premier lieu, j'exprime ma gratitude envers mon DIEU qui m'a accordé le courage et la volonté nécessaires pour mener à bien ce travail.

Je tiens à exprimer mes remerciements particuliers et ma gratitude à ma famille et en particulier à mes chers parents pour leur soutien et leur aide pour devenir un ingénieur.

Je tiens à remercier spécialement Monsieur BELFEDHAL Alaa Eddine, Mr ACHOUR Abdessalam et Mme ALASSAAD Hiba, qui ont supervisé et suivi mon travail.

Mes remerciements vont également aux professeurs et aux enseignants de l'ESI qui m'ont guidé tout au long de mon parcours académique.

Je souhaite également exprimer ma reconnaissance envers Monsieur le président du jury et les membres du jury qui ont consacré leur temps pour évaluer mon travail.

Mes amis et collègues de la promotion méritent également mes remerciements.

Enfin, je suis reconnaissant envers tous ceux qui, de près ou de loin, m'ont apporté leur aide et leur soutien pour concrétiser ce projet.

Résumé

Le secteur industriel est l'un des nombreux domaines dans lesquels l'automatisation de plusieurs tâches est possible grâce à la vision par ordinateur. Cependant, quelle que soit la tâche, par exemple la maintenance ou la manipulation par des robots mobiles, un système de détection précis est nécessaire. La détection d'objets consiste à localiser les différents objets dans une image (via une boîte englobante) et à associer une catégorie à chacun de ces objets. Cette tâche peut être une classification à l'échelle d'une image entière, et une segmentation à l'échelle d'un pixel. Les nombreux défis proposés au fil des années, consistant à résoudre l'une ou l'autre de ces tâches sur une base de données spécifique, ont permis de grandes avancées dans les méthodes de traitement d'images. Parmi ces méthodes, les réseaux de neurones convolutifs (CNN) ont montré une capacité supérieure à résoudre ces tâches, notamment avec l'amélioration continue des capacités de calcul des ordinateurs et la disponibilité de bases de données à grande échelle. Ces bases de données, qui comprennent des images et leurs annotations associées, permettent d'analyser le contenu des images avec différents niveaux de précision.

Malgré toutes ces avancées, la détection d'objets dans un environnement industriel reste une tâche difficile, notamment en raison du manque d'images réelles ou de bases de données pour l'entraînement, et de la contrainte de temps réel requise pour l'algorithme, qui est généralement embarqué dans un dispositif mobile. Ces dernières années, plusieurs chercheurs se sont intéressés à proposer des détecteurs d'objets légers pour les appareils mobiles, tels que PP-PicoDet, MobileNet SSD et Yolo. Ces modèles ont des performances intéressantes qui leur permettent d'être utilisés dans des applications industrielles nécessitant une réponse en temps réel, mais il est encore possible d'optimiser leur fonctionnement pour les rendre plus légers. En effet, la plupart de ces modèles traitent l'image entière pour extraire les régions d'intérêt et ensuite déterminer les détections. Cependant, ils ne prennent pas en compte les connaissances préalables pour optimiser et réduire le temps d'exécution du modèle de détection. En effet, l'utilisation de ces connaissances peut permettre de ne traiter que certaines parties de l'image d'entrée, et de s'appuyer sur les détections précédentes pour réduire le nombre de régions à examiner.

Dans ce travail, nous proposons une méthode qui permet d'optimiser un modèle en intégrant les connaissances préalables afin que le modèle soit léger en raison des ressources de calcul limitées des robots mobiles.

Abstract

The industrial sector is one of the many areas in which the automation of several tasks is possible thanks to computer vision. However, whatever the task, for example maintenance or handling by mobile robots, an accurate detection system is required. Object detection consists in locating the different objects in an image (via a bounding box) and associating a category to each of these objects. This task can be a classification at the scale of an entire image, and a segmentation at the scale of a pixel. The numerous challenges proposed over the years, consisting in solving one or the other of these tasks on a specific database, have allowed great advances in image processing methods. Among these methods, convolutional neural networks (CNNs) have shown a superior ability to solve these tasks, especially with the continuous improvement of computer computing capabilities and the availability of large-scale databases. These databases, which include images and their associated annotations, allow the analysis of image content with different levels of accuracy.

Despite all these advances, object detection in an industrial environment remains a challenging task, especially due to the lack of real images or databases for training, and the real-time constraint required for the algorithm, which is usually embedded in a mobile device. In recent years, several researchers have been interested in proposing lightweight object detectors for mobile devices, such as PP-PicoDet, MobileNet SSD and Yolo. These models have interesting performances that allow them to be used in industrial applications requiring real-time response, but there is still an opportunity to further optimize their operation to make them more lightweight. Indeed, most of these models process the whole image to extract the regions of interest and then determine the detections. However, they do not take into account prior knowledge to optimize and reduce the execution time of the detection model. Actually, the use of this knowledge can allow to process only some parts of the input image, and to rely on the previous detections to reduce the number of regions to be examined.

In this work, we propose a method that allows for optimizing a model by incorporating prior knowledge so that the model is lightweight due to the limited computing resources of mobile robots.

ملخص

يعد القطاع الصناعي واحدًا من العديد من المجالات التي يمكن فيها أتمتة العديد من المهام بفضل رؤية الكمبيوتر. ومع ذلك، بغض النظر عن المهمة، مثل الصيانة أو التعامل بواسطة الروبوتات المتحركة، يتطلب نظام كشف دقيق. يتكون كشف الكائنات من تحديد المواقع المختلفة للكائنات في صورة (عن طريق مربع محدد) وربط فئة بكل من هذه الكائنات. يمكن أن تكون هذه المهمة تصنيفًا على مستوى صورة كاملة، وتجزئةً على مستوى بكسل. لقد أتاحت التحديات العديدة التي تم اقتراحها على مر السنين، والتي تتمثل في حل إحدى هاتين المهمتين في قاعدة بيانات محددة، قدرة (CNNs) تحقيق تقدم كبير في طرق معالجة الصور. من بين هذه الطرق، أظهرت الشبكات العصبية التكرارية متفوقة على حل هذه المهام، خاصة مع تحسين مستمر لقدرات الحوسبة الكمبيوترية وتوفير قواعد البيانات الضخمة. تسمح هذه القواعد، التي تتضمن الصور وتوضيحاتها المرتبطة، بتحليل محتوى الصورة بمستويات مختلفة من الدقة.

على الرغم من كل هذه التقدّمات، يظل كشف الكائنات في بيئة صناعية مهمة صعبة، خاصة بسبب نقص الصور الحقيقية أو قواعد البيانات التدريبية، والقيود الزمني الفعلي المطلوب للخوارزمية، والتي عادة ما تكون مضمنة في جهاز محمول. في السنوات الأخيرة، تطلب العديد من الباحثين اهتمامهم بتقديم كاشفات كائنات خفيفة الوزن للأجهزة المحمولة، تتميز هذه النماذج بأداء مثير للاهتمام يسمح باستخدامها في التطبيقات Yolo و MobileNet SSD و PP-PicoDet مثل الصناعية التي تتطلب استجابة فورية، ولكن لا يزال هناك فرصة لتحسين عملها بشكل أكبر لجعلها أكثر خفة. في الواقع، يقوم معظم هذه النماذج بمعالجة الصورة بأكملها لاستخراج مناطق الاهتمام ثم تحديد الكشفات. ومع ذلك، فإنها لا تأخذ في الاعتبار المعرفة المسبقة لتحسين وتقليل وقت تنفيذ نموذج الكشف. في الواقع، يمكن استخدام هذه المعرفة لمعالجة بعض أجزاء الصورة المدخلة فقط، والاعتماد على الكشفات السابقة لتقليل عدد المناطق التي يجب فحصها.

في هذا العمل، نقترح طريقة تسمح بتحسين النموذج من خلال دمج المعرفة السابقة بحيث يكون النموذج خفيفًا بسبب الموارد الحسابية المحدودة للروبوتات المتحركة.

Liste des Abréviations

AI : Artificial Intelligence.

COCO: Common Objects in Context.

CNN : Convolution Neural Network.

DL : Deep Learning.

GT : Ground Truth.

IoU : Intersection over Union.

mAP : Mean Average Precision.

mAR : Mean Average Recall.

ML : Machine Learning

NMS : Non-Maximum Suppression.

R-CNN : Region-based Convolutional Network.

R-FCN : Region-based Fully Convolutional Network.

RoI : Region of Interest.

SSD : Single Shot Multibox Detection.

SVM : Support Vector machine.

YOLO : You Only Look Once.