A Hybrid System for Object Detection and Distance Estimate Using a Monocular Camera Designed for Autonomous Vehicles

Dr. Bilal shiha [*]
Dr. Majed Ali [**]
Hydar hasan [***]

(Received 1 / 4 / 2024. Accepted 3 / 9 / 2024)

□ ABSTRACT □

This paper proposes a sensing system for an autonomous vehicle for detecting surrounding objects and for identifying their type. This system uses depth maps estimated from RGB images captured by a monocular camera to calculate their distance from the vehicle to minimize cost. The system uses a pre-trained deep learning model called YOLO v8, which has been fine-tuned on the Kitti dataset to detect and recognize objects obstructing the vehicle's path. The Depth maps for images captured from the monocular camera can be estimated using a deep-learning model called AdaBins. Finally, the system combines the resulting information from both previous models to estimate the distances of objects in real-time. The proposed system achieves a mean Average Precision (mAP) of 74%, an average object detection time of 8.5 milliseconds whaen tested on a video clip with a frame rate of 30 fps, and a relative error in estimating the object's distance ranging between 4% and 20% depending on its position within the camera's field of view.

Keywords: Deep learning, Depth map, Distant Estimation, Avoid collision, Machine learning.

[*] **Associate Professor - Department of Computer and Automatic Control Engineering - Faculty of Mechanical and Electrical Engineering - Tishreen University - Lattakia - Syria.**
[**] **Assistant Professor - Department of Artificial Intelligence - Faculty of Information Engineering - Tishreen University - Latakia - Syria.**
[***] **Postgraduate student (PhD) - Department of Computer Engineering and Automatic Control, Faculty of Mechanical and Electrical Engineering - Tishreen University - Latakia, Syria.**

نظام هجين لاكتشاف الكائنات وتقدير مسافتها باستخدام كاميرا أحادية مصمم للمركبات ذاتية القيادة

د. بلال شيحا *

د. مجد علي **

حيدر حسن ***

(تاريخ الإيداع 1 / 4 / 2024. قُبِل للنشر في 3/ 9 / 2024)

□ ملخّص □

يقترح هذا البحث نظام استشعار لمركبة ذاتية القيادة حيث يقوم باكتشاف الكائنات المحيطة وتحديد نوعها ثمَ تقدير مسافتها عن المركبة بالاعتماد على خرائط العمق المقدرة من صور ملونة RGB لكاميرا واحدة فقط بهدف تقليل التكلفة إلى أقل حد ممكن. يتم اكتشاف الكائنات التي تعترض مسار المركبة والتعرف عليها باستخدام نموذج التعلم العميق YOLO v8 بعد إعادة تدريبه على مجموعة بيانات Kitti التي تلبي متطلبات النظام المقترح. أما تقدير خرائط العمق للصور الملتقطة من الكاميرا الأحادية فيتم باستخدام نموذج التعلم العميق AdaBins. تستخدم المعلومات الناتجة عن النموذجين السابقين لتقدير مسافة كل كائن مكتشف في الزمن الحقيقي. يحقق النظام المقترح متوسط دقة تعرف على الكائنات mAP= 74%، ومتوسط زمن اكتشاف كائن قدره 8.5 ميلي ثانية عند تجربته على مقطع فيديو ذو معدل إطارات 30 إطار في الثانية، وخطأ نسبي Relative Error في تقدير مسافة الكائن تراوح بين 4% و20% حسب موضعه ضمن حقل الرؤية للكاميرا.

الكلمات المفتاحية: التعلم العميق، تقدير العمق، تقدير المسافة، تجنب التصادم، التعلم الآلي.

*أستاذ مساعد- قسم هندسة الحاسبات والتحكم آلي- كلية الهندسة الميكانيكية والكهربائية -جامعة تشرين-اللاذقية - سورية.

**مدرس- قسم الذكاء الصنعي- كلية الهندسة المعلوماتية -جامعة تشرين-اللاذقية - سورية.

***طالب دراسات عليا(دكتوراه)-قسم هندسة الحاسبات والتحكم آلي، كلية الهندسة الميكانيكية والكهربائية-جامعة تشرين-اللاذقية سورية.

مقدمة:

التعلم العميق هو أحد طرائق الذكاء الاصطناعي Artificial Intelligence (AI) الذي يمكّن أجهزة الكمبيوتر من تعلم كيفية معالجة البيانات بطريقة مستوحاة من آلية التعلم والتفكير في الدماغ البشري. يمكن لنماذج التعلم العميق التعرف على الأنماط المعقدة في الصور والنصوص والأصوات وغيرها من البيانات لإنتاج رؤى وتنبؤات دقيقة. يمكن استخدام التعلم العميق لأتمتة المهام التي تتطلب عادةً ذكاءً بشرياً، مثل وصف الصور أو تحويل ملف صوتي إلى نص. حيث يمكن إيجاد تطبيقات واسعة له في قطاعات مختلفة مثل الروبوتات والأمن والطب والسيارات ذاتية القيادة وغيرها.

مع التطور المتسارع للتكنولوجيا، وتطبيقات الإنترنت التي أدت إلى توفر كميات هائلة من البيانات، و تطوير أجهزة حاسوب عالية الأداء، وتطور خوارزميات التعلم العميق، حقق الباحثون تقدماً ملحوظاً في مهام الرؤية الحاسوبية مثل اكتشاف الكائنات، والتجزئة الدلالية، وإعادة بناء المشهد [1، 2، 3]، وقد قطعت هذه الأبحاث شوطاً طويلاً و بدأنا نشهد استخداماتها الواسعة في مجالات مختلفة، إلا أن اكتشاف الكائنات (مثل المركبات والمشاة وراكبي الدراجات) في الزمن الحقيقي بتكلفة منخفضة و باستخدام أجهزة استشعار رخيصة و سهلة الاستخدام (كاميرا واحدة) وتقدير مسافتها لا يزال يمثل تحدياً [4] .

يركز هذا البحث على كيفية استخدام تقنيات التعلم العميق والرؤية الحاسوبية للتعرف على الكائنات وتقدير مسافتها باستخدام كاميرا واحدة وإطلاق إنذار إذا كانت هذه المسافة ضمن نطاق الخطر، حيث يمكن أن يساعد مثل هذا النظام في تعزيز اتخاذ القرار في السيارة ذاتية القيادة لتجنب الاصطدام.

أهمية البحث وأهدافه:

وفقاً لتقرير السلامة الطرقيّة الصادر عن منظمة الصحّة العالميّة World Health Organization (WHO) لعام 2023 يتزايد عدد حالات الوفاة على الطرقات نتيجة الحوادث بشكلٍ مستمر ليتجاوز 1.35 مليون ضحيّة كل عام [5]. يشير التقرير أيضاً إلى أنه من المتوقع أن تكون الحوادث المرورية السبب الخامس لوفاة الشباب بحلول عام 2030.

تفيد بيانات أسواق السيارات إلى توجه عام من قبل مصنعيها إلى زيادة متسارعة في تصنيع وتسويق السيارات ذاتية القيادة ما يجعل من انتشارها على نطاق واسع مسألة وقت حيث من المتوقع أن يصل عدد السيارات ذاتية القيادة العاملة على الطرقات الى 127 ألف سيارة بحلول عام 2030 [6] وبناء عليه كان واجباً تطوير أنظمة اكتشاف الكائنات وتقدير مسافتها بهدف تعزيز أمان القيادة الذاتية.

يعد الهدف الرئيسي لهذا البحث صياغة حل تكنولوجي سهل التنفيذ ومنخفض التكلفة وقادر على العمل في الزمن الحقيقي لاكتشاف الكائنات وتقدير مسافتها تمهيدا لربط هذه المنظومة مع منظومة التحكم بالمركبات ذاتية القيادة.

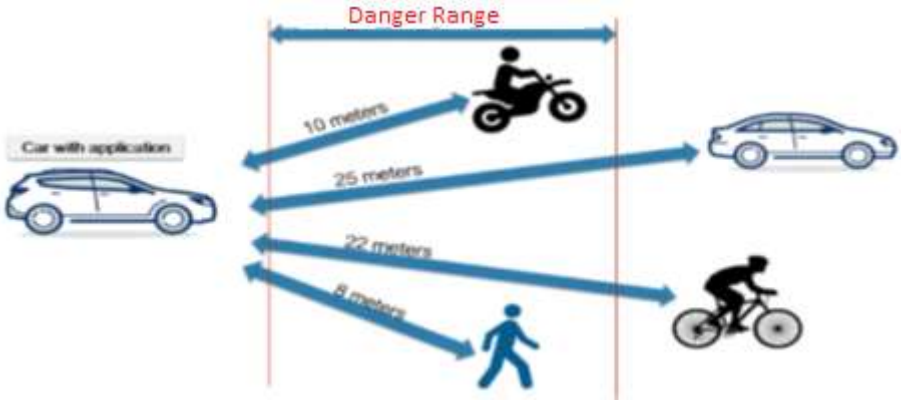
طرائق البحث ومواده:

تعتمد معظم الأبحاث في هذا المجال على استخدام كاميرتين (Stereo camera) أو استخدام مستشعرات مرتفعة التكلفة مثل الرادار وحساسات المسافة التقليدية. يسعى هذا البحث إلى تطوير نظام برمجي يعمل في الزمن الحقيقي بالاعتماد على كاميرا أحادية (Monocular camera) ونماذج التعلم الآلي للكشف عن الكائن وتقدير مسافته من

الكاميرا وتزويد نظام التحكم في المركبة ذاتية القيادة بهذه البيانات على شكل واصفة دخل منظومة التحكم تتضمن نوع الكائن ومسافته عن المركبة وسرعة المركبة لحظة اكتشاف الكائن ليصار الى اتخاذ القرار السليم.

تفترض الدراسة الحالية أن مجال الخطر يكون على مسافة وسطية أقل من 15 متر(بناءً على استقصاء رأي الخبراء في مجال الحركة المرورية) كما هو موضح في الشكل (1).

يستهدف البحث الكائنات التي تندرج تحت الفئات التالية: المركبات والشاحنات والحافلات والمشاة والدراجات. بهدف الوصول الى النتيجة المرجوة قمنا بتقسيم البحث إلى مرحلتين: المرحلة الأولى وهي مرحلة الكشف عن الكائنات، حيث سنقوم بإعداد مجموعة البيانات المناسبة، واختيار إطار التعلم الآلي وبنية النموذج المناسب. المرحلة الثانية وهي مرحلة تقدير المسافة، حيث سنبحث عن تقنية الرؤية الحاسوبية وطريقة التعلم العميق المناسبين لإنشاء خريطة العمق وتقدير المسافة من العمق الناتج.



الشكل رقم (1): نطاق الخطر للكائنات المستهدفة في نظامنا المقترح.

1- تقنيات التعرف على الكائنات باستخدام التعلم العميق:

يُستخدم التعلم العميق على نطاق واسع في أبحاث الرؤية الحاسوبية (Computer Vision) في الوقت الحاضر. حيث تركز أساليب التعلم العميق الحديثة على استخدام نموذج الشبكة العصبونية متعددة الطبقات باستخدام خوارزمية الانتشار الخلفي Back propagation، ومن أكثر تقنيات التعلم العميق شيوعاً واستخداماً هي الشبكات العصبونية الالتفافية Convolution Neural Network (CNN) [7]. وبغض النظر عن نوع خوارزمية التعلم العميق فإن التعلم إما أن يكون خاضع للإشراف Supervised learning حيث تكون المدخلات والمخرجات معروفة، أو غير خاضع للإشراف Unsupervised learning حيث تكون المدخلات معروفة ولكن المخرجات غير معروفة، أو شبه خاضع للإشراف حيث نستفيد من ميزات طريقتي التعلم السابقتين. تستخدم طرق التعلم المذكورة أعلاه في نماذج التعرف على الكائنات وفق طريقتين هما:

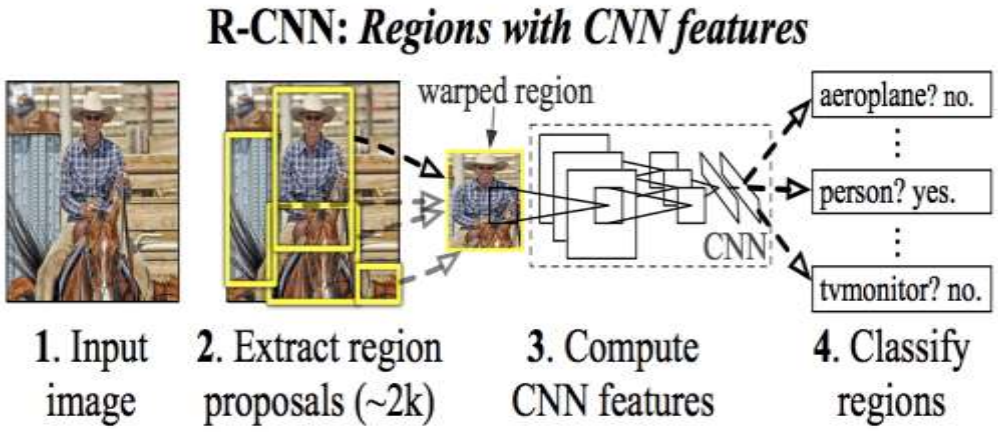
- أسلوب التدريب من أوزان بدائية تأخذ قيم عشوائية حيث تستغرق الشبكة العصبونية الكثير من الوقت في التدريب وتتطلب الكثير من البيانات لتحقيق الأداء الفعال للنموذج.
- أسلوب نقل التعلم (حيث يتم إعادة تدريب شبكة تم تدريبها مسبقاً على آلاف العينات على مجموعة جديدة من البيانات .هذا هو أسلوب التعلم الأكثر استخداماً بين الباحثين لأنه يعطي نتائج ذات جودة أفضل في وقت أقل مقارنة بالأسلوب السابق.

أظهرت تقنيات التعلم الآلي والتعلم العميق أنها تقنيات واعدة بسبب نتائجها الجيدة في مجال واسع من التطبيقات [8]. نلاحظ أنه في الوقت الحالي يتم تفضيل استخدام تقنيات وخوارزميات التعلم العميق على تقنيات التعلم الآلي نظراً، لتوفر كميات ضخمة من البيانات وتقنيات قادرة على استخلاص الميزات والأنماط الخفية بشكل تلقائي. تم تقسيم حالياً خوارزميات التعلم العميق لاكتشاف الكائنات إلى صنفين حسب آلية العمل، يعتمد الصنف الأول على شبكة واحدة لإنتاج مواقع الكشف عن الكائنات والتنبؤ بالفئة في وقت واحد. أما الصنف الثاني فيقسم العمل إلى مرحلتين متمايزتين، حيث يتم اقتراح المناطق ذات الأهمية أولاً، ثم يتم تصنيفها بواسطة شبكات تصنيف منفصلة ثانياً، وفيما يلي سنسلط الضوء على أشهر الخوارزميات وطريقة عملها.

1-1 الشبكات العصبونية الالتفافية المعتمدة على المنطقة (Region-Convolutional Neural Network):

تعتبر من أولى خوارزميات التعلم العميق التي استخدمت للتعرف على الكائنات، يقوم مبدأ عملها على تمرير النوافذ المنزلقة للبحث في جميع أنحاء الصورة عن الكائنات المستهدفة باستخدام خوارزمية البحث الانتقائي، التي تقوم باستخراج 2000 منطقة تقريباً من الصورة والتي نسميها مقترحات المنطقة [9]، بعد ذلك يتم إرسال المناطق المستخرجة إلى شبكة عصبونية التفافية كما هو موضح بالشكل رقم (2). تقوم الطبقات الالتفافية في CNN باستخراج الميزات التي يتم تمريرها بعد ذلك إلى مصنف آلة متجهات الدعم Support Vector Machine (SVM) الذي يتم استخدامه للتنبؤ بوجود الكائنات ضمن المناطق المقترحة.

تحقق هذه الخوارزمية دقة عالية في التعرف على الكائنات، لكنها تستغرق وقت طويل في الكشف كونها تقترح 2000 منطقة لكل صورة، بالتالي لا يمكن استخدامها في أنظمة الزمن الحقيقي.

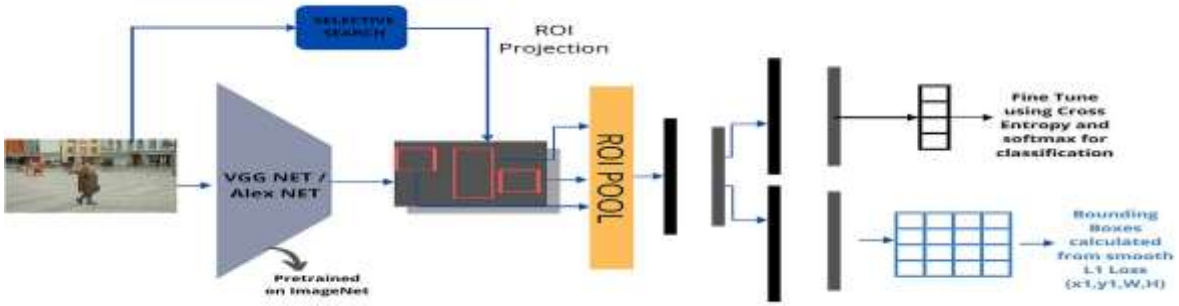


الشكل رقم (2): معمارية خوارزمية R-CNN.

2-1 الشبكات العصبونية الالتفافية المعتمدة على المنطقة السريعة (Fast R-CNN):

تعتمد آلية عملها على تحسين طريقة الخوارزمية السابقة R-CNN عن طريق استخدام شبكة CNN لتحليل الصورة بأكملها لاستخراج الميزات اللازمة للكشف عن الكائنات بدلاً من تقسيمها إلى أجزاء، ثم توليد اقتراحات لمواقع الكائنات المحتملة في الصورة باستخدام خوارزمية Region Proposal Network (RPN) ، وبعد ذلك تستخدم خوارزمية Region of Interest Pooling (RoIPool) لتحديد مواقع الكائنات المحتملة واستخراج المعالم اللازمة للكشف وبعد

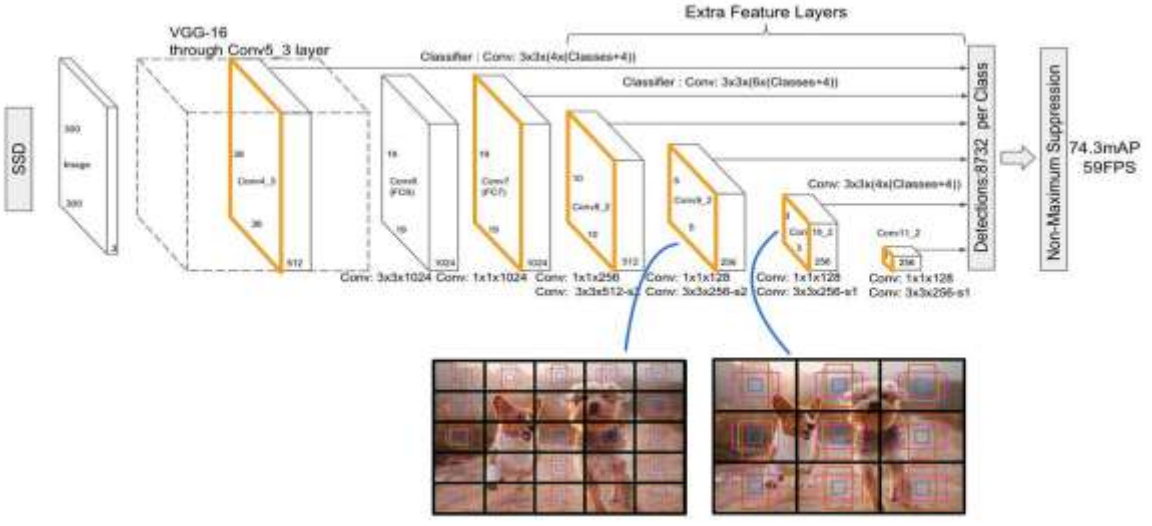
ذلك تمرر المخرجات المفيدة إلى طبقات كاملة الاتصالية (fully connected layer). يليها بعد ذلك تابع تفعيل SoftMaxالذي يستخدم للتنبؤ بفئة كل منطقة وطبقة تنتج إحداثيات المربع المحيط كما هو مبين بالشكل رقم (3). بالتالي من خلال تغير ترتيب مراحل عملية الكشف زادت الدقة وسرعة الكشف عن الخوارزمية السابقة [10].



الشكل رقم (3): معمارية خوارزمية fast R-CNN.

3-1 خوارزمية كاشف اللقطة الواحدة متعدد الصناديق ((Single Shot MultiBox Detector (SSD):

تعتبر هذه التقنية من أحداث التقنيات في مجال الكشف عن الكائنات في الصور، حيث يعتمد مبدأ عملها على اكتشاف الكائن وتصنيفه من خلال مسار واحد للشبكة العصبونية العميقة عن طريق تحليل الصورة المدخلة بأكملها واستخراج المميزات اللازمة للكشف باستخدام شبكة (CNN)، ويتم توليد اقتراحات لمواقع الكائنات المحتملة باستخدام شبكة (CNN)، وبعد ذلك يتم استخدام خوارزمية (NMS)(Non-Maximum Suppression) لتصفية اقتراحات مواقع الكائنات وإزالة التداخل بينها، ليتم تصنيف الكائنات باستخدام المميزات المستخرجة وإخراج مواقع الكائنات وأحجامها المحتملة بالصورة كما هو موضح في الشكل رقم (4).



الشكل رقم (4): معمارية خوارزمية (SSD).

تحقق (SSD) توازناً أفضل بين السرعة والدقة عن طريق توليد اقتراحات متعددة للكائنات المحتملة في الصورة باستخدام مجموعة من الاحجام والأشكال المختلفة للمربعات المحيطة بالكائنات، واستخدام شبكات (CNN) ذات أحجام مختلفة لتحليل الصورة مما يزيد من دقة الكشف ويقلل من نسبة الخطأ وحجم الذاكرة المطلوبة لتشغيل النظام[11].

4-1 خوارزمية الكشف بنظرة واحدة ((You Only Look Once (YOLO)):

تعتبر خوارزمية كاشف النظرة الواحدة YOLO واحدة من أسرع خوارزميات الكشف عن الكائنات في الصور، حيث تم تطويرها عام 2016، وتعمل بطريقة مشابهة لخوارزمية (SSD)، حيث تستخدم YOLO خطوة واحدة فقط لمعالجة الصورة بأكملها. وهذا يجعلها أسرع وأكثر كفاءة، حيث تقسَّم كل صورة إلى شبكة من الخلايا، ويتم تحديد الكائنات في كل خلية وتمرير كل خلية عبر شبكة (CNN) لاستخراج المميزات، وتستخدم دالة SoftMax للتنبؤ بكل صنف من الكائنات في كل خلية، ويتم توقع المربع المحيط بالنسبة لحجم الخلية وتمثَّل بإحداثيات المركز والعرض والارتفاع. تستخدم تقنية (NMS) لإزالة المربعات المحيطة الزائدة والمتداخلة بدرجة ثقة أقل وتحسين دقة الاكتشاف. بالرغم من إنها ليست الأكثر دقة ولكنها من أفضل الخيارات لمهام الكشف عن الكائنات في الصور في الزمن الحقيقي مع الحفاظ على مستوى مقبول من الدقة[12] .

تم العمل على تطوير نموذج YOLO بأكثر من إصدار لكشف الكائنات وتجزئة الصور بالزمن الحقيقي. حيث كان أخر اصدار YOLO v8 الذي قدم أداءً فائقاً من حيث السرعة والدقة، فالتصميم البسيط للنموذج يعد مناسبًا لمختلف التطبيقات وقابلاً للتكيف بسهولة مع منصات الأجهزة المختلفة [12].

2- تقنيات تقدير المسافة:

يعد تحديد بعد الكائن منذ اكتشافه واحدة من أهم المشكلات التي تظهر غالباً في أنظمة الرؤية الحاسوبية، وهو محل اهتمام العديد من الباحثين الذين يعملون على تقدير المسافة أو تقدير عمق الجسم من المصدر باستخدام الكاميرا، وقد طورت مجموعة متنوعة من التقنيات والخوارزميات التي تعمل في الزمن الحقيقي. حيث تستخدم هذه التطبيقات في مجالات عديدة منها: أنظمة النقل الذكية، والروبوتات، وصناعة المركبات ذاتية القيادة [11] وغيرها. يمكن تصنيف طرق تقدير المسافة للمركبات ذاتية القيادة إلى تصنيفين حسب طريقة العمل، وهما:

- الطرق الفعالة (Active Methods): تعتمد على إرسال إشارة معينة من السيارة واستقبالها بواسطة جهاز استشعار خارجي، ومن ثم يتم حساب المسافة بناءً على الزمن اللازم للإشارة للذهاب والعودة. وتشمل هذه الطرق تقنيات الرادار والأشعة تحت الحمراء والليزر، عادة تكون هذه الطرق مكلفة للغاية [13].
- الطرق السلبية (Passive Methods): وتعتمد على استقبال إشارات خارجية مثل الضوء والحرارة، ومن ثم تحوَّل هذه الإشارات إلى معلومات مفيدة باستخدام تقنيات الرؤية الحاسوبية لتحديد المسافة. وتشمل هذه الطرق تقنيات الكاميرات [12] وغالباً ما تكون مثل هذه الطرق منخفضة التكلفة.

تشمل الطرق المعتمدة على الكاميرات تقنيتين:

تستخدم تقنية الرؤية المجسمة (Binocular vision) زوج من الكاميرات الموضوعة على بعد معين لتحديد المسافة بين الكائنات. حيث يُصوَّر نفس المشهد من زوايا مختلفة وبعدها تُحسَب المسافة بين الكائنات باستخدام تقنية التصوير الثلاثي الأبعاد (3D)، حيث يتم حساب الإحداثيات الثلاثية لكل نقطة في الصورة بالاستناد إلى الفرق في المواقع الإحداثية للنقط المتماثلة في كل صورة، ويُعرف الاختلاف في موقع الصورة بالتباين بينما تسمى المسافة بين الكاميرتين بخط الأساس.

تقنية الرؤية الأحادية (Monocular vision) حيث يتم استخدام كاميرا واحدة لتحديد المسافة بين الكائنات. ولكن، يتطلب ذلك تقنيات معالجة الصور المتقدمة لتحديد المسافة بالاستناد إلى عوامل مثل حجم الكائن في الصورة وزاوية

الرؤية والتباين البصري والتفاصيل الهيكلية للكائنات. ويتم تحديد المسافة بالاستناد إلى معلومات العمق التي تم استخلاصها من الصورة باستخدام تقنيات التعلم العميق.

قدم العديد من الباحثين أنظمة معتمدة على الرؤية المجسمة لتقدير المسافة، إلا أننا مهتمون بالطريقة الأحادية في هذه البحث. تنطلق العديد من الأبحاث من مبدأ الرؤية المجسمة وتحاول تنفيذها باستخدام كاميرا واحدة. حيث يمكن استخدام استراتيجيات تقدير المسافة بين الهدف وموضع الكاميرا الأحادية، مثل الطريقة الزمنية، التي تحسب المسافة بناءً على التسلسل الزمني لنسخ الكائنات (مثل قياس المسافة البصرية) أو الطريقة الأخرى التي يتم بها التنبؤ بالمسافة في الإطار الفعلي بشكل مستقل عن الإطارات السابقة مثل الخوارزميات المعتمدة على نظرية المثلثات أو تقدير خرائط العمق [13].

يوضح الجدول رقم (1) أهم التقنيات المستخدمة لتقدير مسافة الكائنات باستخدام أجهزة استشعار مختلفة.

الجدول رقم (1): تقنيات تقدير المسافة في المركبات ذاتية القيادة.

الحساس	التقنية	القيود
LIDAR	آلية المسح بالليزر	– أجهزة ضخمة وثقيلة وغالية الثمن. تتأثر بالتشويش الإلكتروني ومضرة بالصحة.
RADAR	أمواج راديوية	– تتأثر بالتشويش الإلكتروني والطرق المنحنية. يتعذر استخدامها في المدن ذات الأبنية الكثيفة.
كاميرا Stereo	مستشعر سلبي للرؤية (كاميرتين)	– تحتاج تزامن للصور وبيئة متناظرة – تحتاج عمليات حسابية معقدة ومتعددة الخطوات – تتطلب حجم ذاكرة ضخم للعمل.
كاميرا Monocular	مستشعر سلبي للرؤية (كاميرا)	– تحتاج مجموعة تدريب كبيرة. – تتطلب عمليات معقدة حسابياً.

3- تقنيات تقدير العمق:

خريطة العمق هي صورة تحتوي على معلومات حول كثافة كل بكسل في الصورة من محور الرؤية المركزي للكاميرا (المحور z على اعتبار الصورة الأساسية تأخذ المحورين x & y) أي مسافة تلك النقطة من الكاميرا. تكون صور خريطة العمق ذات تدرج رمادي وتكون النقطة أقل سطوعاً كلما كانت أبعد. في معظم الأحيان، يُخزَّن العمق عند كل نقطة كرقم 16 بت/مم. يعتبر تقدير العمق من الصور الأحادية موضوعا مهماً يمكن تنفيذه بالعديد من الطرق، حيث يمكن صياغة المشكلة على أنها انحدار لخريطة العمق من صورة RGB واحدة وهنا يمكن الاستفادة من قدرة الشبكات العصبية العميقة على تعلم العلاقات المعقدة بين ميزات الصورة وإشارات العمق لإنتاج خرائط عمق دقيقة من صور ثنائية الأبعاد [15]،[14]. بفضل أداء الشبكات العصبية الالتفافية (CNN) الفائق في معالجة الصور يمكن تحديد الترابط في خرائط الكثافة بشكل صحيح من الصور الأحادية [7].

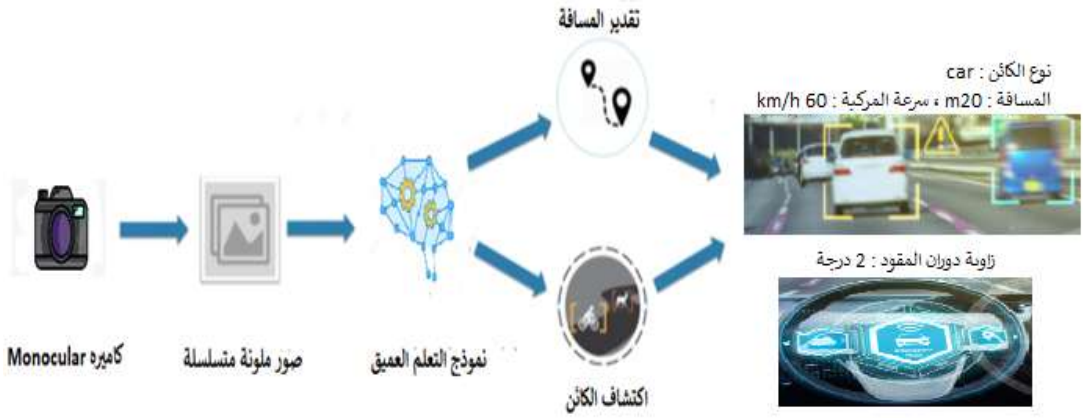
تعتمد بعض تقنيات تقدير العمق على صور متعددة كمدخلات (Stereo Camera) لتتمكن من الحصول على عمق كائن بنجاح ومن الأمثلة على هذه التقنيات العمق من الحركة والعمق من التركيز [16]. حيث يمكن تطوير تقنيات تقدير العمق باستخدام الصور الأحادية من خلال استخدام نماذج التعلم العميق المدرب على زوج من الصور مع

خرائط العمق الخاصة بها والتي حققت دقة مرضية [17]، ويمكن القيام بذلك إما عن طريق استخدام التعلم تحت الإشراف أو التعلم غير الخاضع للإشراف. هناك العديد من مجموعات البيانات الشائعة الاستخدام والتي تحتوي على خرائط عمق للصور والبيانات اللازمة للتدريب مثل (NYU Depth) تحوي خرائط عمق RGB−D للصور الداخلية (الملتقطة داخل المنازل)، ومجموعة بيانات (Kitti) تحوي خرائط عمق لصور الطرق.

4- النظام المقترح:

يقترح البحث نظام يقوم باكتشاف الكائنات من تسلسل صور من كاميرا أحادية وتقدير مسافة الكائنات باستخدام نماذج التعلم العميق مع بعض تقنيات الرؤية الحاسوبية للمعالجة وإعطاء النتيجة لمنظومة التحكم بالمركبة ذاتية القيادة.

يتضمن مخطط البحث إجراء مجموعة من الخطوات تبدأ بتحليل المتطلبات للنظام وصولاً لتصميم النظام كما هو موضح في الشكل رقم (5).



الشكل رقم (5): خطوات تصميم النظام المقترح.

يتم جمع صور متسلسلة من كاميرا أحادية (أو ملف فيديو) وتمريرها إلى نموذج التعلم العميق المكون من نموذجين فرعيين، أحدهما لاكتشاف الكائنات والآخر لتقدير خريطة العمق. يقوم نموذج كاشف الكائن بتصنيف وتحديد موضع الكائن المستهدف في الصورة، بينما يقوم نموذج مولد خريطة العمق بإنشاء خريطة العمق من الصورة، حيث تُستخَدم خريطة العمق هذه مع إحداثيات الكائن من الكاشف، لتقدير مسافة هذا الكائن من الكاميرا. بمجرد اكتشاف الجسم ومعرفة موضعه وحساب المسافة من الكاميرا، يمكن معرفة فيما إذا كانت هذه المسافة ضمن منطقة الخطر وبناء عليه سيتم اتخاذ القرار المناسب في منظومة التحكم. بهدف تنفيذ هذا النظام تم تقسيمه إلى المراحل التالية:

1. تدريب وبناء نموذج كاشف الكائنات.
2. تقدير خرائط العمق من الصور الملونة لكاميرا أحادية العين (صور ثنائية الأبعاد).
3. دمج خريطة العمق مع اكتشاف الكائنات.
4. تزويد نظام التحكم بالمعلومات اللازمة (كائن – المسافة عن المركبة – درجة الخطورة – سرعة المركبة – زاوية دوران المقود).

4−1 تدريب وبناء نموذج كاشف الكائنات:

يعتمد النظام المقترح على كشف كائنات من فئات محددة تناسب بيئة العمل للمركبات ذاتية القيادة، إضافةً على ذلك يجب أن يعمل الكاشف بالزمن الحقيقي وبدقة عالية. نلاحظ من مقارنة نماذج الكشف عن الكائنات (YOLOv3،

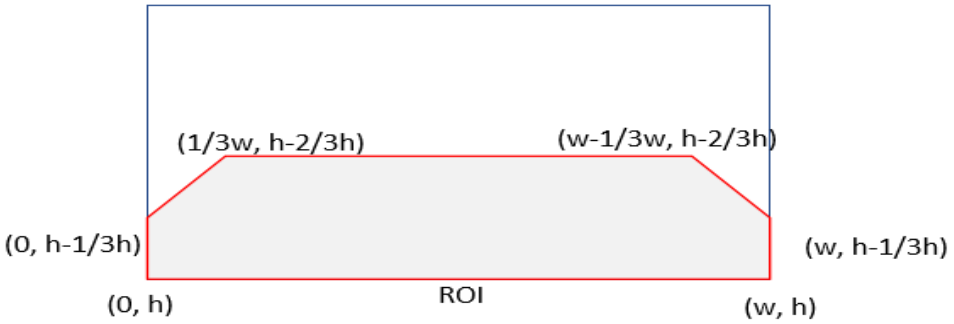
YOLO ، SDD،Faster R-CNN) [9]، أن النموذج YOLO هي الخيار الأفضل بالنسبة للنظام المقترح ولا سيما v8 الذي يتمتع بدقة وسرعة تمكنه من العمل في الزمن الحقيقي بالإضافة لقدرته على التقاط الميزات عالية المستوى ودمج الميزات من مقاييس متعددة.

تم اختيار مجموعة البيانات [18] Kitti (7481 صورة تدريب و7518 صورة اختبار، 12 جيجابايت) التي تلبي متطلبات النظام المقترح من حيث صور عالية الجودة لفئات الكائنات (مركبة، شاحنة صغيرة، حافلة، شاحنة، مشاة، دراجة نارية) والتي سيتم استخدامها لتدريب المصنف المقترح.

تم تدريب المصنف YOLO v8 على مجموع بيانات Kitti والمدرب مسبقاً pretrained على مجموعة بيانات MS COCO بهدف عدم تدريبه من قيم أوزان بدائية عشوائية لتسريع التدريب وزيادة دقته.

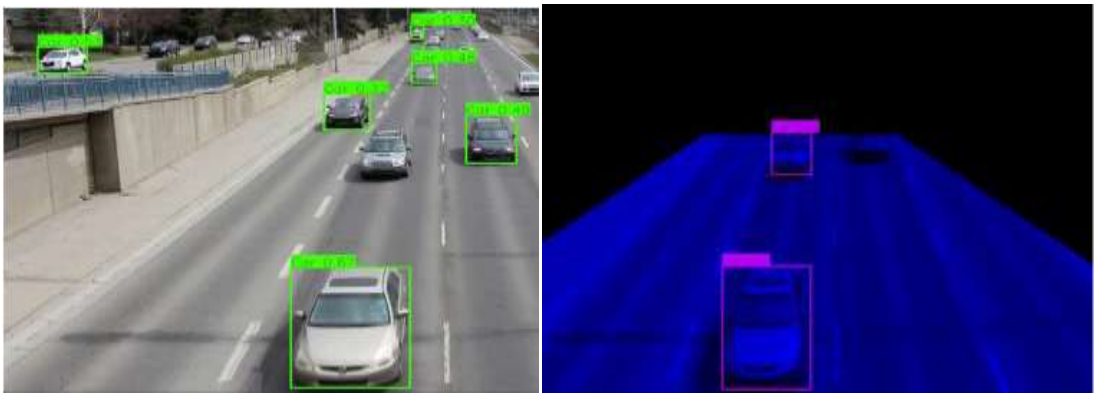
تم اختبار دقة وأداء النموذج في الزمن الحقيقي على الفيديو والصور من خلال بارامتر قياس أداء التقاطع عبر الاتحاد Intersection over Union (IoU) وهو معيار يتم استخدامه لقياس مدى قدرة النموذج على التنبؤ بالمربع المحيط بالكائن المكتشف بالزمن الحقيقي، يكون اختبار متوسط الدقة mean Average Precision (mAP) عند الحد الأدنى من (IoU) وعادةً ما يتم استخدام قيمة عتبة 0.5 لمعرفة ما إذا كان التنبؤ إيجابيًا أم لا.

من الضروري بمكان تحديد منطقة الاهتمام في الصور الملتقطة والتي تتضمن الكائنات الأمامية فقط بالنسبة للكاميرا الأحادية وبما ينسجم مع التطبيق المقترح للنظام في المركبات ذاتية القيادة كما هو موضح في الشكل رقم(6).



الشكل رقم (6): منطقة الاهتمام في النظام المقترح.

يوضح الشكل (7) عينة اختبار لصورة ملتقطة قبل وبعد تحديد منطقة الاهتمام.



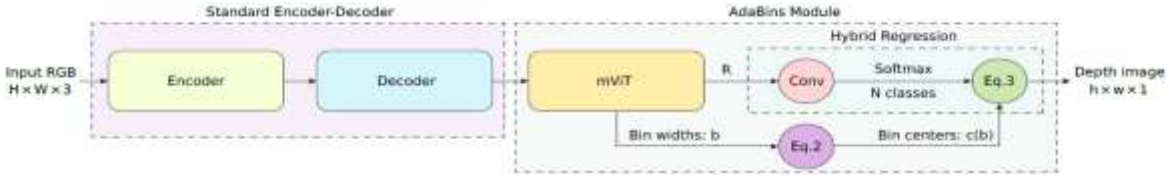
(a) **(b)**

الشكل رقم (7): عينة اختبار منطقة الاهتمام في النظام المقترح (a) قبل تطبيق منطقة الاهتمام (b) بعد تطبيق منطقة الاهتمام.

4-2 تقدير خرائط العمق من الصور الملونة:

تهدف هذه المرحلة الى توليد خرائط العمق من الصور الملونة RGB لاستخدامها في تقدير مسافة الجسم المكتشف وذلك بعد تحويل الصور ثنائية البعد الملتقطة من كاميرا أحادية الى صورة عميقة ذات تمثيل ثلاثي الأبعاد.

تقوم المعمارية المقترحة في [19] للشبكة العميقة AdaBins الموضحة في الشكل (8) بمعالجة المشهد بهدف تقسيم نطاق العمق المتوقع للصورة إلى صناديق تكيفية، ومن ثم تقدير العمق النهائي كمجموعات خطية من قيم مركز الصندوق.



الشكل (8): معمارية الشبكة العميقة AdaBins

تم تدريب النموذج على مجموعة بيانات Kitti التي تنسجم مع متطلبات النظام المقترح وتم تقييم أداء النظام المقترح باستخدام بارامترات تقييم الأداء التالية:

- الجذر التربيعي لمتوسط مربعات الخطأ Root Mean Square Error (RMS): يقيس مقدار الفرق بين القيم الحقيقة والقيم التي تم التنبؤ بها من قبل النموذج، ويعطى بالعلاقة:

$$RMS = \sqrt{\frac{1}{n}\sum_{p}^{n}(y_p - \hat{y}_p)^2}$$

- متوسط الخطأ النسبي Average Relative Error (REL): وهو النسبة بين الخطأ المطلق لكمية مقاسة والكمية المقاسة نفسها، ويعطى بالعلاقة:

$$REL = \frac{1}{n}\sum_{p}^{n}\frac{|y_p - \hat{y}_p|}{y}$$

حقق نموذج AdaBins النتائج الموضحة في الجدول (2) التالي:

الجدول رقم (2): تقييم نتائج نموذج AdaBins.

Root Mean Square Error (RMS)	Average Relative Error (REL)
2.360	0.058

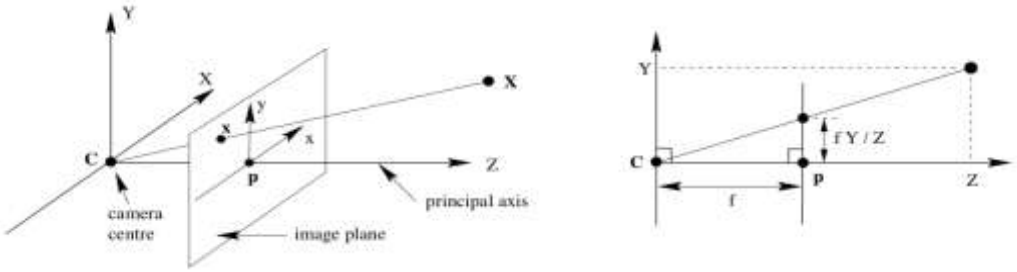
تعتبر هذه النتائج أفضل بنسبة 13.5% من حيث RMS و22.4% من حيث REL مقارنة بأفضل نموذج (state-of-the art) تم تطويره في مجال تقدير خرائط العمق، وبناء عليه يمكن اعتماد نموذج الشبكة العميقة AdaBins في نظام تقدير المسافة للمركبات ذاتية القيادة [18].

4-3 نظام تقدير مسافة الكائنات الهجين:

استخدمت هذه المرحلة إحداثيات الجسم المكتشف باستخدام نموذج YOLO مع خريطة العمق التي تم توليدها باستخدام AdaBins على شكل نظام هجين يمكن من تقدير مسافة الكائنات المكتشفة من الكاميرا كما هو موضح في الشكل (9).

الشكل رقم (9): دمج اكتشاف الكائن مع تقدير العمق في النظام المقترح.

يتم استخراج قيمة العمق لجميع البكسلات الواقعة ضمن صندوق الإحاطة بالكائن المكتشف للكائن المكتشف ومن ثم يتم حساب مسافة هذا الكائن عن الكاميرا بالاعتماد على قيمة البيكسل المركزي ضمن صندوق الإحاطة وبارامترات الكاميرا المستخدمة كما هو مبين في الشكل (10).



الشكل رقم (10): هندسة الاسقاط للكاميرا ذات الثقب.

بهدف الحصول على القيمة Z التي تعبر عن المسافة، يمكننا تحويل نقطة (u, v) على المستوي ثنائي الأبعاد إلى نقطة (X, Y, Z) على المستوى ثلاثي الأبعاد وفق النموذج الرياضي التالي

$$P = K[R \mid t] * Q$$

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

حيث:

- P [u, v, 1]: النقطة المتوقعة.
- K: بارامترات الكاميرا، وهي عادة ما تكون البعد البؤري (أو النقطة البؤرية) (f_x, f_y) والنقطة الرئيسية أو المركز البصري (v_0, u_0) ويتم ذكر هذه الخصائص في توصيف الكاميرا.
- [R|t]: تحويل هجين للدوران والانسحاب بالنسبة للمحاور الإحداثية في جملة ديكارتية متعامدة منتظمة.
- Q, [X, Y, Z, 1]: النقطة ثلاثية الأبعاد المقابلة معبراً عنها باستخدام نظام الإحداثيات الديكارتية.
- S: مقياس يوضح كيفية تغيير حجم البيكسل مع التغيير في البعد البؤري.

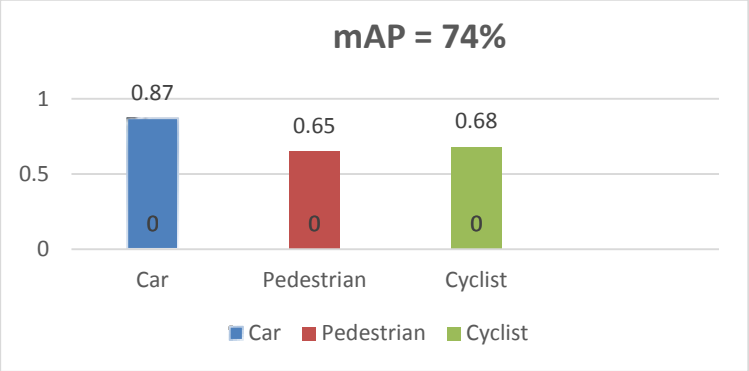
بالنتيجة تم إنشاء سحابة نقطية ثلاثية الأبعاد تمكن من الحصول على المسافة من مركز صندوق الإحاطة للكائن المكتشف.

النتائج والمناقشة:

يعطي النظام المقترح على الخرج واصفة تتضمن المعلومات التالية (نوع الكائن المكتشف، مسافة الكائن عن الكاميرا، سرعة المركبة اللحظية، زاوية دوران المقود اللحظية) حيث يمكن استخدام الواصفة الناتجة عن نظام الاستشعار المصمم في نظام التحكم في المركبات ذاتية القيادة.

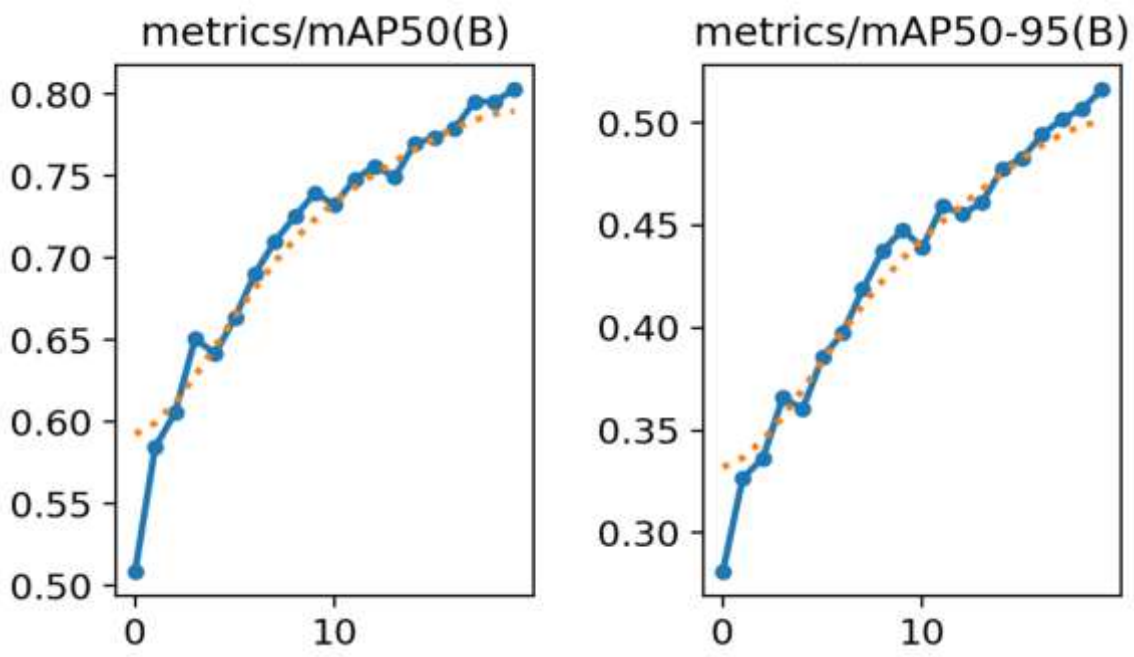
تقسم عملية تقييم أداء النظام المقترح إلى مرحلتين، المرحلة الأولى تقييم أداء نموذج اكتشاف الكائنات، المرحلة الثانية تقييم أداء نظام تقدير المسافة ككل.

يتم اختبار أداء نماذج اكتشاف الكائنات باستخدام مقياس متوسط معدل الدقة mean Average Precision (mAP) الذي يعبر عن متوسط دقة النموذج بالنسبة لجميع الكائنات المكتشفة. حقق نموذج اكتشاف الكائنات الذي تم اعادة تدريبه على مجموعة بيانات Kitti متوسط دقة mAP= 74%، ومتوسط زمن اكتشاف كائن قدره 8.5 ميلي ثانية عند تجربته على مقطع فيديو ذو معدل إطارات 30 إطار في الثانية.



الشكل رقم (11): دقة اكتشاف كل صنف للنموذج المقترح.

يوضح الشكل رقم (12) منحنيات متوسط الدقة mAP50 من أجل عتبة IoU تساوي 0.5 على اليسار، وmAP50-95 من أجل عتبة IoU متغيرة ضمن المجال 0.5 و0.95 بخطوة 0.05 على اليمين.



الشكل رقم (12): دقة اكتشاف النموذج المقترح .

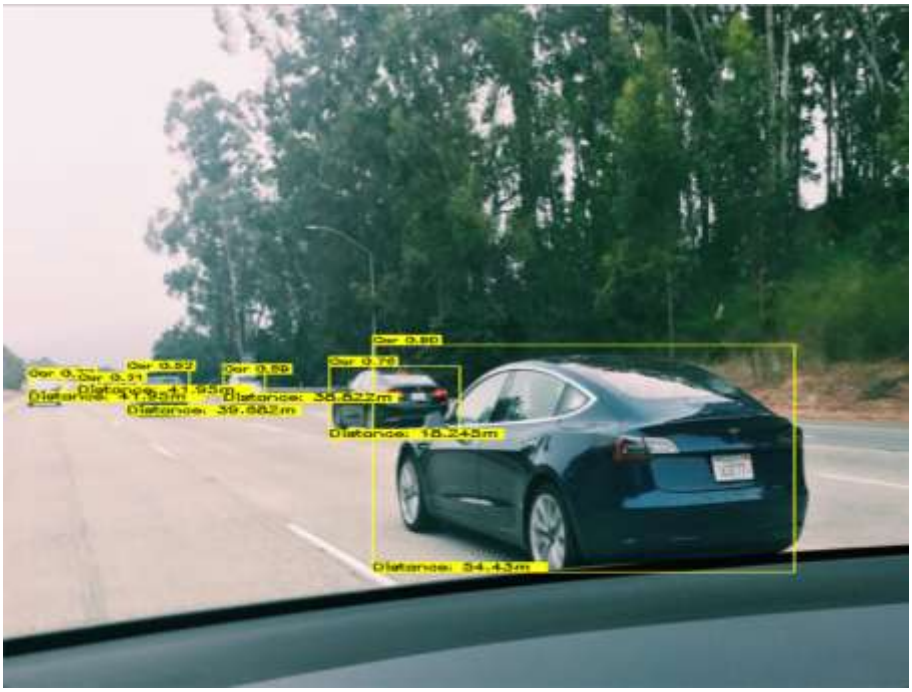
لتقييم أداء نظام تقدير المسافة تم إجراء مجموعة من التجارب في العالم الحقيقي في المحاكي حيث تم تقدير مسافة أكثر من مئة كائن ذات مسافات معلومة في مجموعة من الصور الملتقطة باستخدام كاميرا ذات موقع ثابت ضمن المركبة، ومن ثم تم حساب الخطأ النسبي Relative Error بين المسافات الحقيقية والمسافات المُقدرة كما هو موضح في الجدول (3).

الجدول رقم (3): عينات اختبار لتقدير المسافة لصور ملتقطة من مركبة.

عينة الاختبار	المسافة الحقيقة	المسافة المقدرة	نسبة الخطأ %
1	8.5	7.2	15.29
2	10	10.2	1.96
3	18	19.37	7.07
4	20	21	4.76
5	35	33.48	4.34

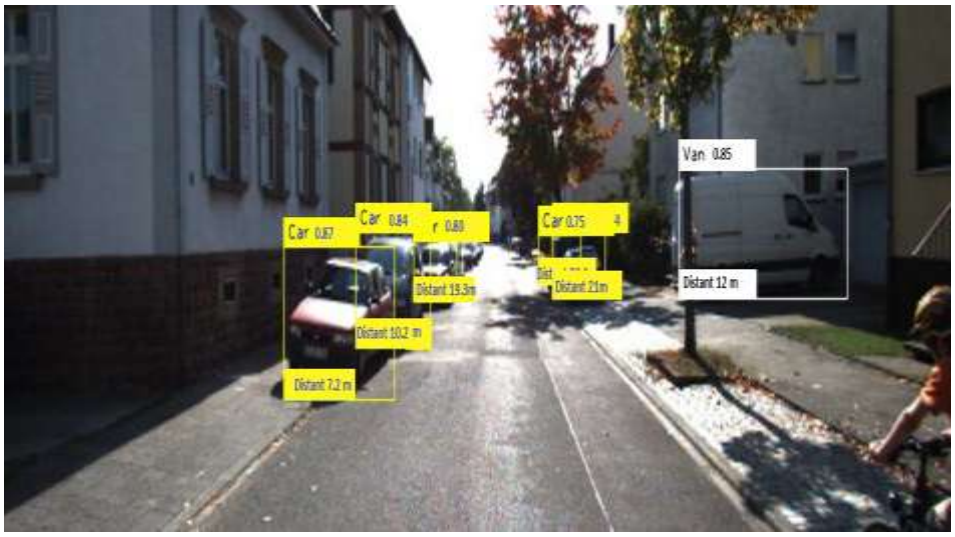
أظهرت النتائج أن النظام يعطي أفضل أداء مع الكائنات الواقعة ضمن منطقة الاهتمام للنظام المقترح، ومع الكائنات غير القريبة جداً من الكاميرا.

يظهر الشكل (13) عينة من التجارب التي تم إجرائها حيث يمكن ملاحظة أن المسافة المقدرة للسيارة الأولى المكتشفة غير صحيحة بسبب قربها وزاويتها بالنسبة إلى الكاميرا، أما بالنسبة للسيارات اللاحقة فالمسافات التقديرية قريبة إلى حد ما من المسافات الفعلية.



الشكل رقم (13): صورة من عينات اختبار النظام.

يظهر الشكل (14) أن الصور الملتقطة أفقياً أدت إلى دقة تقدير مسافة أعلى من الصور الملتقطة بشكل عمودي والواقعة ضمن منطقة الاهتمام لنظامنا المقترح.



الشكل رقم (14): صورة من عينات اختبار النظام.

الاستنتاجات والتوصيات:

يمكن من خلال المناقشة السابقة اقتراح نظام استشعار يلبي وبدرجة مقبولة مجال التطبيق في المركبات ذاتية القيادة بالتكامل مع منظومة التحكم. حيث يمكن للنظام المقترح تمييز الكائنات وبدقة عالية كما أنه يستطيع تقدير مسافة هذه الكائنات عن الكاميرا المستخدمة وبدقة جيدة ولا سيما الكائنات الواقعة ضمن منطقة الاهتمام للمركبات ذاتية القيادة.

يوصي البحث بتطوير نظام تحكم ذكي تكون مدخلاته هي خرج منظومة الاستشعار المدروسة في هذا البحث (نوع الكائن المكتشف، مسافة الكائن عن الكاميرا، سرعة المركبة اللحظية، زاوية دوران المقود اللحظية) وخرجه زاوية دوران المقود والعزم المطبق على كل من دواستي الوقود والفرامل.

يمكن تطبيق النظام المقترح في بيئة المرور السورية على شكل نظام منفصل عن النظام المضمن المصنع من قبل الشركة الأم بهدف استخدامه كنظام تنبيه مساعد للسائق لتقليل حوادث السير وزيادة درجة الأمان.

References:

[1]P. Viola and M. J. Jones, "Robust Real-time Object Detection," International journal of computer vision, Springer, Volume 57 ,2004.

[2]G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006, doi: 10.1162/neco.2006.18.7.1527.

[3] J.-J. Hernandez-Lopez, A.-L. Quintanilla-Olvera, J.-L. Lopez-Ramirez, F.-J.Rangel-Butanda, M.-A. Ibarra-Manzano, and D.-L. Almanza-Ojeda, "Detecting objects using color and depth segmentation with Kinect sensor," *Procedia Technol.*, vol. 3, pp. 196–204, 2012, doi:10.1016/j.protcy.2012.03. 21.

[4]Y. Bengio, "Deep Learning of Representations for Unsupervised and Transfer Learning," JMLR: Workshop and Conference Proceedings 27:17-37, 2012.

[5]"WHO | Global status report on road safety 2023," *WHO*,2023

[6]Martin placek, https://www.statista.com/statistics/1230664/projected-number-autonomous-cars-worldwide/ . [Accessed: 1-Feb-2024].

[7]M. D. Zeiler and R. Fergus, Visualizing and Understanding Convolutional Networks, arXiv preprint arXiv:1311.2901v3, 2013. https://arxiv.org/abs/1311.2901

[8] "A Beginner's Guide to Multilayer Perceptrons (MLP) | Pathmind." [Online]. Available: https://pathmind.com/wiki/multilayer- perceptron. [Accessed: 15-Feb-2020].

[9]J. Hui, "Object detection: speed and accuracy comparison (Faster R-CNN, RFCN,SSD, FPN, RetinaNet and YOLOv3)." [Online].Available: https://medium.com/ @jonathan_hui/ object-detectionn -speed-and -accuracy comparison-faster-r-cnn-r-fcn-ssd-and-yolo-5425656ae359. [Accessed: 20-Jan-2024].

[10] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9905 LNCS, pp. 21–37, Dec. 2015, doi: 10.1007/978-3-319-46448-0_2.

[11] "A Sensor-Fusion Drivable-Region and Lane-Detection System for Autonomous Vehicle Navigation in Challenging Road Scenarios.[Online] .Available: https:// www.researchgate.net/publication/260522418_A_Sensor-Fusion_Drivable Region_and Lane Detection_System_for Autonomous_Vehicle_Navigation_in_ Challenging_ Road_ Scenarios. [Accessed: 22-Nov-2023].

[12] J. Redmon and A. Farhadi, "YOLOv8: An Incremental Improvement."Python Lessons." [Online]. Available: https://pylessons.com/YOLOv8-introduction/. [Accessed: 26-DEC-2023].

[13] A. Yadav and T. B. Mohite-Patil, "Distance Measurement with Active & Passive Method Distance Measurement with Active & Passive Method," *Int. J. Comput. Sci. Netw.*, vol. 1, no. 4, pp. 2277–5420, 2012.

[14] H. Kim, C.-S. Lin, J. Song, and H. Chae, "Distance Measurement Using Single Camera with a Rotating Mirror," 2005.

[15] T. Ophoff, K. Van Beeck, and T. Goedeme, "Exploring RGB+depth fusion for real-time object detection,"*Sensors (Switzerland)*,vol.19, no. 4, 2019, doi: 10.3390/s19040866

[16] S.-S. Tung and W.-L. Hwang, "Depth Extraction from a Single Image and Its Application," in *Pattern Recognition - Selected Methods and Applications*, IntechOpen, 2019.

[17] C. Holzmann and M. Hochgatterer, "Measuring distance with mobile phones using single-camera stereo vision," in *Proceedings - 32nd IEEE International Conference on Distributed Computing Systems Workshops, ICDCSW 2012*,2012, pp. 88–93, doi: 10.1109/ICDCSW.2012.22.

[18] "The KITTI Vision Benchmark Suite." [Online]. Available: http://www.cvlibs.net/ datasets/kitti/eval_object.php?obj_benchmark=2d. [Accessed: 16-Dec-2023].

[19] Shariq Farooq Bhat, Ibraheem Alhashim, Peter Wonka "AdaBins: Depth Estimation using Adaptive Bins" , KAUST, arXiv:2011.14141v1 [cs.CV] 28 Nov 2020.