

Behavior Cloning Imitation Learning For Vision Transformer To Predict Steering Angle And Speed Of Self-Driving Vehicles

Dr. Bilal Shiha *

Dr. Majed Ali **

Hydar Hasan ***

(Received 15 / 10 / 2024. Accepted 22 / 12 / 2024)

□ ABSTRACT □

Controlling a self-driving vehicle on highways is a complex and essential task that involves numerous challenges. Rule-based expert systems and traditional self-driving systems offer limited solutions when dealing with complex real-world scenarios. This has led to a shift towards using learning-based planning, supported by big data, particularly through behavior cloning for End-to-End Architectures. This study proposes the use of transformers that rely on self-attention mechanisms in computer vision, which have proven successful in natural language processing tasks and effective in patch-embedding colored images. The proposed model is trained on a comprehensive dataset that includes driving behaviors, environmental conditions, and vehicle dynamics, with the goal of predicting the steering angle and speed of self-driving vehicles on highways. The results show that the model achieves a significantly lower Mean Absolute Error (MAE) compared to traditional deep learning techniques, by using vision transformers. The paper introduces two models: the first uses a single Multi-Layer Perceptron Head with two output neurons for both the steering angle and vehicle speed, while the second uses two separate prediction heads with one output neuron for each. The mean absolute error for the first model was 0.175 and for the second model 0.37, while the lowest mean absolute error in studies using traditional self-driving systems was 0.491.

Keywords: Vision Transformers, Steering Angle and Speed Prediction, Self-Driving Vehicles, Behavior Cloning.

Copyright



:Tishreen University journal-Syria, The authors retain the copyright under a CC BY-NC-SA 04

* Associate Professor - Department of Computer and Automatic Control Engineering - Faculty of Mechanical and Electrical Engineering - Tishreen University - Latakia - 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.

التعلم باستنساخ السلوك لمحاولات الرؤية في مجال التنبؤ بزوايا التوجيه والسرعة للمركبات ذاتية القيادة

د. بلال شيجا *

د. مجد علي **

حيدر حسن ***

(تاريخ الإيداع 15 / 10 / 2024. قُبِلَ للنشر في 22 / 12 / 2024)

□ ملخص □

تعتبر مهمة التحكم بالمركبة ذاتية القيادة على الطرق السريعة من المهام المعقدة والأساسية والتي تتضمن الكثير من المعوقات، حيث قدمت النظم الخبيرة القائمة على القواعد وأنظمة القيادة الذاتية التقليدية حلاً قد تكون محدودة في التعامل مع سيناريوهات معقدة في العالم الحقيقي، الأمر الذي أدى إلى التوجه نحو استخدام التخطيط القائم على التعلم المدعوم ببيانات ضخمة ولا سيما التعلم باستنساخ السلوك للأنظمة الشاملة End-to-End Architecture، اقترحت هذه الدراسة استخدام المحاولات التي تعتمد على آلية الانتباه الذاتي في مجال الرؤية الحاسوبية و التي أثبتت نجاحها في مهام معالجة اللغات الطبيعية. حيث تبين فعاليتها في تضمين الجزء Patch Embedding للصور الملونة ومن ثم التنبؤ بزوايا التوجيه وسرعة المركبة ذاتية القيادة على الطرق السريعة في الحالة الدراسية المطروحة في هذا البحث، وذلك بعد تدريبها على مجموعة بيانات شاملة لسلوكيات القيادة والظروف البيئية وديناميكيات المركبات على طول الطريق السريع.

أظهرت النتائج أن مقياس تقييم الأداء المعتمد متوسط الخطأ المطلق أقل بشكل ملحوظ من بقية الدراسات التي اعتمدت تقنيات التعلم العميق التقليدية وذلك باستخدام محاولات الرؤية بعد اختبار نموذجي تنبؤ مختلفين من حيث استخدام رأس تنبؤ Multi-Layer Perceptron Head بعصبوني خرج لكلا زوايا التوجيه وسرعة المركبة أو رأسي تنبؤ منفصلين بعصبون خرج لكل على حدا.

حيث بلغ متوسط الخطأ المطلق في النموذج الأول 0.175 وفي النموذج الثاني 0.37 في حين كانت أقل قيمة لمتوسط الخطأ المطلق في الدراسات التي اعتمدت على أنظمة القيادة الذاتية التقليدية 0.491.

الكلمات المفتاحية: محاولات الرؤية الحاسوبية، تقدير زوايا التوجيه والسرعة، المركبات ذاتية القيادة، استنساخ السلوك.



حقوق النشر : مجلة جامعة تشرين- سورية، يحتفظ المؤلفون بحقوق النشر بموجب الترخيص

CC BY-NC-SA 04

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

**مدرس - قسم الذكاء الصناعي - 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:

تكمن الأهمية البالغة للسيارات الذاتية القيادة بأثرها في السلامة المرورية، فحوادث السيارات هي السبب الثامن للوفيات على مستوى العالم، و95٪ من الحوادث المرورية ناجمة عن أخطاء بشرية. لذلك هناك توقع منطقي بأن تُحدث أتمتة النقل انخفاضاً كبيراً في عدد الحوادث المرورية، وبالتالي في عدد الضحايا. حيث تأخذ السيارات ذاتية القيادة أهمية متزايدة في الوقت الحالي، ويمكن أن تُغيّر نظام النقل مستقبلاً تغييراً جذرياً [1].

وقد ساهم التطور المتسارع في نماذج تعلم الآلة (Machine Learning (ML والتعلم العميق (Deep Learning (DL) إلى تحقيق إنجازات غير مسبوقة لا سيما في مجال تطوير المركبات ذاتية القيادة التي تهدف إلى زيادة كفاءة الحركة المرورية وتخفيض معدلات الحوادث على الطرقات [2,3].

أظهرت العديد من الشركات الكبرى التي تعمل على تطوير مركبات ذاتية القيادة اهتماماً كبيراً في الاستفادة من قدرات الذكاء الاصطناعي لتطوير أنظمة ذكية قادرة على قيادة المركبات بكفاءة تتفوق على البشر في بعض الحالات [4]، حيث تعتمد المنهجية التقليدية المتبعة على استراتيجية تصميم يتم فيها تقسيم النظام إلى مجموعة مكونات مستقلة مسؤولة عن إدراك البيئة وتخطيط المسار والتحكم بالمركبة [5]، وعلى الرغم من أن هذه الاستراتيجية مفيدة من حيث إمكانية تفسير القرارات وسهولة تصحيحها، إلا أن حدوث خلل ما في أي مكون من المكونات المذكورة قد يؤدي إلى تعريض حياة السائقين والمشاة إلى الخطر [7,8]، ولذلك هناك توجه كبير اليوم إلى تطوير أنظمة القيادة الذاتية الشاملة End-to-End Autonomous Driving Systems.

1- الدراسات المرجعية:

اقترح Bayarov وآخرون [10] 5 نماذج مختلفة من الشبكات العصبونية المستوحاة من الدماغ البشري، والتي تجمع بين الميزات المرئية المستخرجة باستخدام الشبكات العصبونية الالتفافية ومخرجات شبكات التكرار الزمني Temporal Recursion Networks (NCP) لمحاكاة سلوكيات القيادة البشرية. تم تقييم البنى المقترحة على مجموعتي بيانات للقيادة الذاتية هما مجموعة بيانات محاكي Udacity ومجموعة بيانات comma.ai من العالم الحقيقي. حقق أفضل النماذج متوسط خطأ مطلق قدره 0.691.

قدمت الدراسات [11,12,13] بنى شبكات عصبونية التفافية للتنبؤ بزاوية توجيه والسرعة من خلال الصور، مع اختلافات طفيفة في بنية CNN المستخدمة من حيث عدد الطبقات والتفاصيل الأخرى، وقد حققت هذه النماذج متوسط خطأ مطلق قدره 0.527 و 0.593 و 0.491 على الترتيب عند اختبارها على مجموعة البيانات Comma.ai.

اقترح Valiente وآخرون [14] نهجاً قائماً على مشاركة الصور بين المركبات ذاتية القيادة التعاونية لتحسين دقة التنبؤ بزاوية التوجيه، حيث تم تدريب نموذج شامل مؤلف من شبكات عصبونية التفافية وشبكات الذاكرة الطويلة قصيرة الأمد Long Short-Term Memory (LSTM) والطبقات كاملة الاتصال. تم تقييم النموذج على مجموعة بيانات العالم الحقيقي comma.ai، وقد حقق متوسط خطأ مطلق قدره 0.510.

اقترح Mohammadi وآخرون [15] نظاماً للتنبؤ بزاوية التوجيه يعتمد على استخدام خوارزميات معالجة صورة تقليدية مثل Harris و SIFT و SURF و FAST و BRIEF للعثور على المناطق التي تحتوي المعلومات الأكثر أهمية ضمن الصورة، وبعد استخراج المناطق الهامة يتم تمريرها إلى شبكة عصبونية التفافية لاستخراج الميزات المفيدة منها، ثم التنبؤ بزاوية التوجيه بناءً على تلك الميزات. تم اختبار النموذج على مجموعة البيانات comma.ai وحقق أفضل النماذج متوسط خطأ مطلق قدره 0.896.

اقترح Morgia-Bonilla وآخرون [16] شبكة عصبونية التفاضلية للتنبؤ بزوايا التوجيه للمركبة ذاتية القيادة من خلال استنساخ السلوك. تتألف الشبكة المقترحة من طبقات التفاضلية لاستخراج الميزات وطبقات تجميع لتقليل الأبعاد وطبقات كاملة الاتصالية لاتخاذ القرار بشكل يحاكي أفعال الخبير بناءً على مدخلات مرئية. تمت تجربة عدد من البنى المختلفة من خلال تعديل عدد المرشحات في الطبقات الالتفافية وتوابع التفعيل المختلفة مثل RELU و Swish و Tanh، بالإضافة إلى عدد العصبونات في الطبقات كاملة الاتصالية. تم تدريب 20 نموذج للعمل في محاكي Udacity بحيث تعبر الخسارة عن مدى الانحراف عن المسار الأصلي، وقد حقق أفضل النماذج خسارة قدرها 0.013 على مسار Lake و 0.011 على مسار Mountain.

اقترح Huang وآخرون [17] نظاماً متعدد النماذج تحت اسم (MFE-SSNet). يتألف هذا النظام من ثلاثة نماذج تعتمد على أنواع مختلفة من بيانات الدخل (صور الطريق وسرعة المركبة وزوايا التوجيه)، ويتم دمج مخرجات هذه النماذج باستخدام نموذج دمج ميزات قائم على الانتباه المحلي (Local Attention (LA-FFM) يشكل خرج دخالاً لشبكة عصبونية تؤدي مهمة التنبؤ بزوايا التوجيه والسرعة للمركبة ذاتية القيادة. يتألف النموذج الخاص بالصور من شبكة عصبونية التفاضلية (HRNet) تليها شبكة LSTM، فيما يتألف نموذج سرعة المركبة وزوايا التوجيه من شبكة LSTM تليها شبكة MLP، وأخيراً يتم دمج الميزات المستخرجة من كل نموذج. تم اختبار النظام على مجموعة البيانات Udacity Challenge 2 وتقييم أدائه باستخدام معيار RMSE فحقق عند استخدام الصور كمدخلات فقط خطأ قدره 0.118 للسرعة و 0.090 لزوايا التوجيه. [2024]

أهمية البحث وأهدافه:

تهدف هذه الدراسة إلى تطوير نظام قيادة ذاتية شامل للتنبؤ بزوايا توجيه عجلة القيادة وسرعة المركبة من خلال مدخلات مرئية فقط (تسلسل إطارات فيديو) بالاعتماد على نماذج التعلم العميق المعروفة باسم محولات الرؤية Vision Transformers (ViT) التي أظهرت نتائجاً ممتازة مقارنة بأفضل الشبكات العصبونية الالتفافية Convolutional Neural Networks (CNNs) في مهام الرؤية الحاسوبية على الرغم من حاجتها لموارد حاسوبية أقل للتدريب [18].

تتجلى أهمية البحث في القدرة للتوصل إلى نظام تحكم دقيق وناجح على الطرق الحضرية السريعة خاص بالمركبات ذاتية القيادة.

2- منهجية وحدود البحث:

تتقسم منهجية البحث إلى خطوات تتعلق باقتراح نموذج معتمد على محولات الرؤية الحاسوبية لاستخراج الميزات من المدخلات المرئية الملتقطة من كاميرا أحادية ومن ثم استخدام تلك الميزات للتنبؤ بزوايا توجيه عجلة القيادة وسرعة المركبة بالاعتماد على نهج التعلم باستنساخ السلوك. أما بالنسبة لحدود البحث فهي القيادة الآلية على الطرق الحضرية السريعة متعددة المسارب.

طرائق البحث ومواده:

1- الطرق السريعة Highway:

وهي الطرق المصممة خصيصاً لحركة السير السريعة وغالباً ما تكون للربط بين المدن والأماكن المتباعدة وتحقق ما يلي [19]:

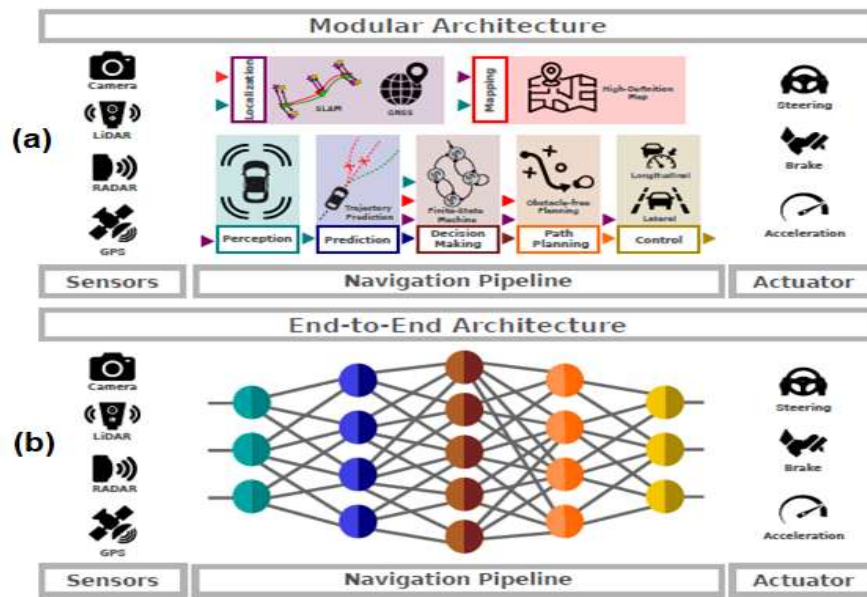
- تتكون من مسربين على الأقل وبحيث يكون لكل مسرب حد أقصى محدد للسرعة ومنفصلة عن بعضها، إما بواسطة مساحة فاصلة غير معدة لحركة السير أو بوسائل أخرى.
- لا تنتفرج إلى أماكن تقع على جوانبها
- لا تتقاطع مع أي طريق آخر أو سكة حديد أو طريق ترام، أو معبر للمشاة
- مزود بلافئات تشير إلى أنه طريق سريع ومخصص لفئات معينة من المركبات الآلية البرية
- مجهزة بإشارات المرور المعلنة الخاصة بالطرق السريعة ومخصصة لفئات محددة من المركبات الآلية.
- يُحظر توقف المركبات أو ركنها على مسارها.

2- مقارنة بين أنظمة القيادة الذاتية التقليدية والشاملة:

يتم عادةً تصميم أنظمة القيادة الذاتية التقليدية بالاعتماد على نهج معياري يتم من خلاله تطوير ودمج وظائف مختلفة مثل الإدراك والتنبؤ والتخطيط (Planning) بشكل منفصل، ويعتبر نموذج التخطيط جزءاً أساسياً في التحكم بالتوجيه والسرعة [21]. تعتمد عمليات التخطيط في الأنظمة التقليدية غالباً على نظم خبيرة قائمة على القواعد، والتي قد تكون محدودة في التعامل مع سيناريوهات معقدة في العالم الحقيقي، الأمر الذي أدى إلى التوجه نحو استخدام التخطيط القائم على التعلم المدعوم ببيانات ضخمة.

تُعرف أنظمة القيادة الذاتية الشاملة (التي تستخدم البنية من النوع End-to-End Architecture) بأنها بُنى ذكية قادرة على التعلم تأخذ بيانات المستشعرات كمدخلات وتعطي على الخرج أوامر تحكم منخفضة المستوى [22]. وعلى عكس الطرق التقليدية (التي تستخدم البنية من النوع Navigation Pipeline)، تعمل الأنظمة الشاملة من خلال تحويل تمثيلات الميزات لبيانات الدخل (Feature Representations) إلى أوامر تحكمية من خلال عملية تعلم موجهة مما يعزز أداء التخطيط [21].

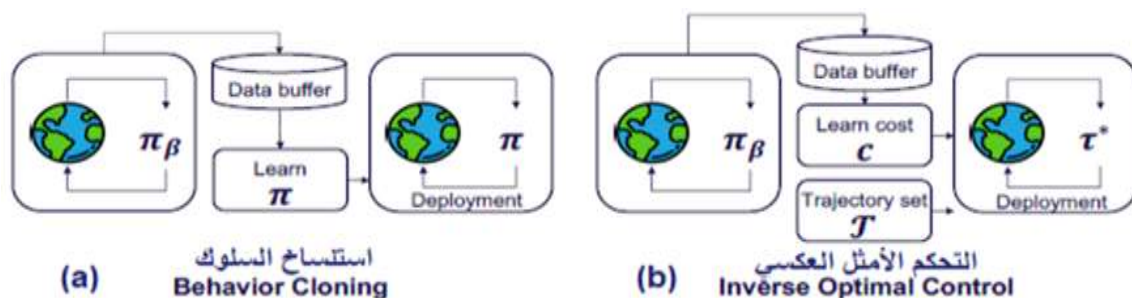
تتميز أنظمة القيادة الذاتية التقليدية بإمكانية تفسير قراراتها وسهولة تصحيح أخطاءها [23]، لكن عمليات التدريب المستقلة للوحدات المختلفة قد يؤدي إلى عدم توافق النظام بشكل عام وحدث أخطاء تراكمية [25]. أما بالنسبة لأنظمة القيادة الذاتية الشاملة فتتميز بالبساطة إذ يتم دمج عمليات الإدراك والتنبؤ والتخطيط في نموذج واحد يخضع لتدريب موحد موجه نحو أداء المهمة النهائية، مما يحسن كفاءة استهلاك الموارد الحاسوبية نتيجة الاعتماد على بنية واحدة مشتركة [27].



الشكل (1) منهجية تصميم الأنظمة في المركبات ذاتية القيادة (a) المنهجية التقليدية، (b) المنهجية الشاملة.

3- التعلم بالتقليد:

يمكن تعريف التعلم بالتقليد (IL) Imitation Learning أو التعلم من التجارب العملية بأنه عملية تقوم على تدريب وكيل ذكي (Agent) لتعلم سياسة ما عن طريق تقليد سلوك خبير. يعد الهدف الأساسي من التعلم بالتقليد في مجال المركبات ذاتية القيادة إنتاج مخططات مسار أو إشارات تحكم [28]. يتعلم الوكيل الذكي ربط ميزات الدخل بالمخرجات المطلوبة مع تقليل أخطاء التنبؤ إلى أدنى حد ممكن. يمكن لتقنية التعلم بالتقليد تحقيق نتائج جيدة في مراحل مبكرة من عملية التدريب، لكن للحصول على أفضل نتيجة ممكنة يجب استخدام مجموعة بيانات ضخمة وتدريب النماذج لوقت طويل. يعد استنساخ السلوك Behavior Cloning أحد أهم أشكال تقنية التعلم بالتقليد في مجال القيادة الذاتية، حيث يقوم الوكيل الذكي باستنساخ سلوك سائق بشري أو متحكم تقليدي من خلال تعلم ميزات محددة من مجموعة بيانات التدريب المستخدمة، ومن أشكال التعلم بالتقليد الأخرى التحكم الأمثل العكسي Inverse Optimal Control المعروف أيضاً باسم التعلم المعزز العكسي Inverse Reinforcement Learning [28].



الشكل (2) منهجيات التعلم بالتقليد (a): منهجية استنساخ السلوك، (b): منهجية التحكم الأمثل العكسي (التعلم المعزز العكسي).

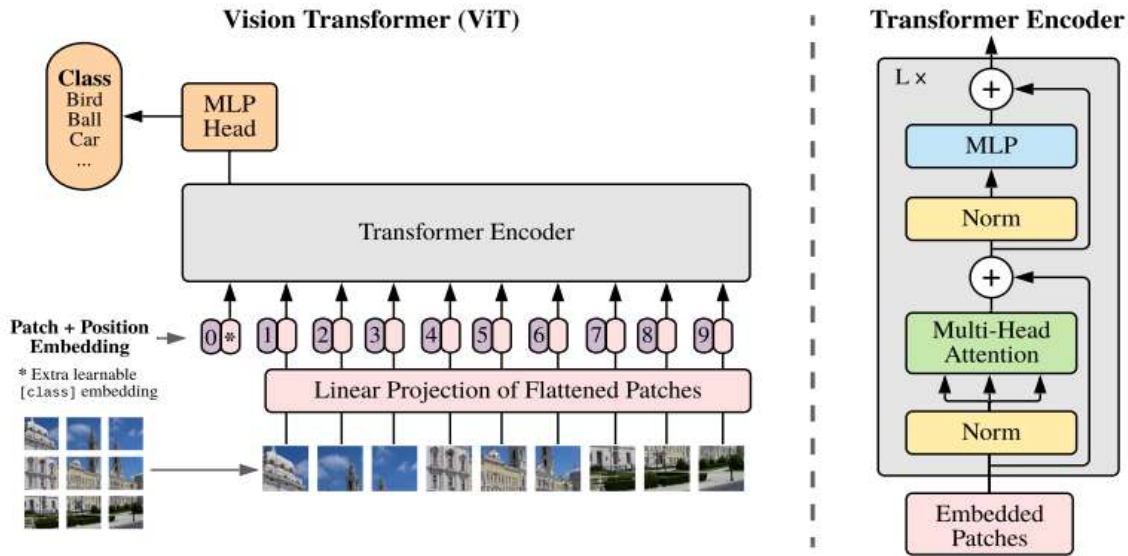
4- محاولات الرؤية:

تعتبر نماذج المحولات Transformers التي تعتمد على آلية الانتباه الذاتي (Self-Attention) الخيار المعتمد في الكثير من مهام معالجة اللغة الطبيعية (Natural Language Processing (NLP. أمّا في مجال الرؤية

الحاسوبية Computer Vision فتعتبر البنى القائمة على العمليات الالتفافية هي الأكثر استخداماً، لكن على الرغم من ذلك أثبتت الدراسات أن استخدام المحولات وآلية الانتباه يعد فعالاً للغاية ويمكن أن يستبدل حتى الشبكات العصبونية الالتفافية (Convolutional Neural Networks (CNNs) في بعض الحالات [31].

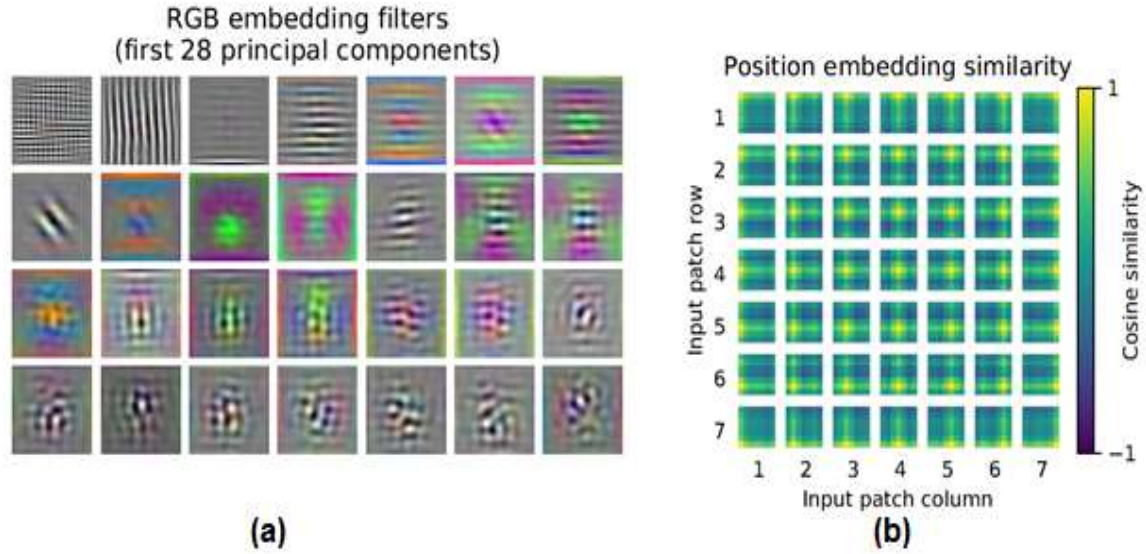
يمكن تطبيق آلية الانتباه الذاتي على الصور بطريقتين: الأولى تعتمد على إيجاد العلاقة بين البيكسل الواحد وكل بكسلات الصورة الأخرى، الأمر الذي يتطلب قدرات معالجة هائلة بالنسبة للصور ذات العدد الكبير من البكسلات، أما الطريقة الثانية تعتمد على تقسيم الصورة وتطبيق آلية الانتباه على المناطق المحلية. يعد الهدف الرئيسي لمحول الرؤية Vision Transformer (ViT) تطبيق مبدأ المحولات على الصور مباشرة مع أقل قدر من التعديلات. لتحقيق ذلك يتم تقسيم الصورة إلى أجزاء ومن ثم إدخالها على شكل تسلسل من التمثيلات الخطية (Linear Embeddings) بحيث تعامل هذه الأجزاء معاملة الكلمات (Tokens) في تطبيقات معالجة اللغات الطبيعية [31].

تتعامل نماذج المحولات عادةً مع تسلسل أحادي البعد من تمثيلات الكلمات كدخل لها، ولمعالجة الصور ثنائية الأبعاد بطريقة مشابهة يتم تحويل الصورة x من الشكل $H \times W \times C$ إلى تسلسل أحادي الأبعاد من البكسلات x_p لكل جزء (Patch) من الأجزاء ثنائية الأبعاد بعد التقسيم كما هو موضح في الشكل 3. يشير H و W إلى الطول Hight والعرض Width فيما يشير C إلى عدد القنوات. يقوم المحول بعد ذلك باستخدام شعاع ميزات (Latent Vector) ثابت الحجم D ، عبر جميع طبقاته لتسطيح كل جزء من أجزاء الصورة وتمثيله بإسقاط خطي Linear Projection مؤلف من D بُعد قابل للتدريب يطلق عليه اسم تضمين الجزء Patch Embedding [31].



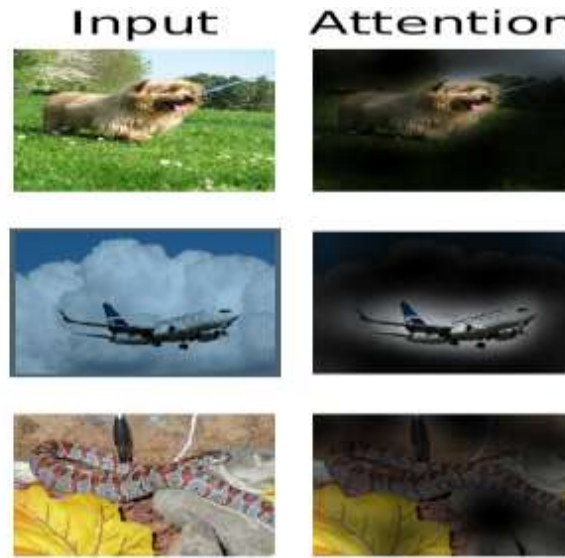
الشكل (3) معمارية محول الرؤية ViT [31].

تقوم الطبقة الأولى من محول الرؤية بأخذ إسقاط خطي لأجزاء الصورة المسطحة إلى فضاء ميزات بأبعاد أقل، ثم يتم بعد ذلك إضافة تضمين مكاني (Positional Embedding) إلى تمثيلات أجزاء الصورة، حيث تم تدريب النموذج على ترميز المسافات ضمن الصورة. يوضح الشكل 4 المرشحات المستخدمة في عملية التضمين الخطي الأولى لقيم RGB الخاصة بالصورة، ومقدار التشابه بين التضمين المكاني للأجزاء المختلفة تبعاً للأسطر والأعمدة [25]، حيث تملك الأجزاء القريبة من بعضها تمثيلات متشابهة، كما تملك الأجزاء الواقعة في نفس السطر أو العمود تمثيلات متشابهة.



الشكل (4): (a) مرشحات التمثيل الخطي الأولي لقيم RGB لمحول الرؤية ViT. (b) تشابه مواضع التمثيلات المضمنة [31].

تسمح بعد ذلك طبقات الانتباه الذاتي بدمج المعلومات عبر كامل الصورة حتى في الطبقات الدنيا، بحيث يكون الخرج كما هو موضح في الشكل 5.



الشكل (5) عينات لخرج طبقة الانتباه الذاتي [31].

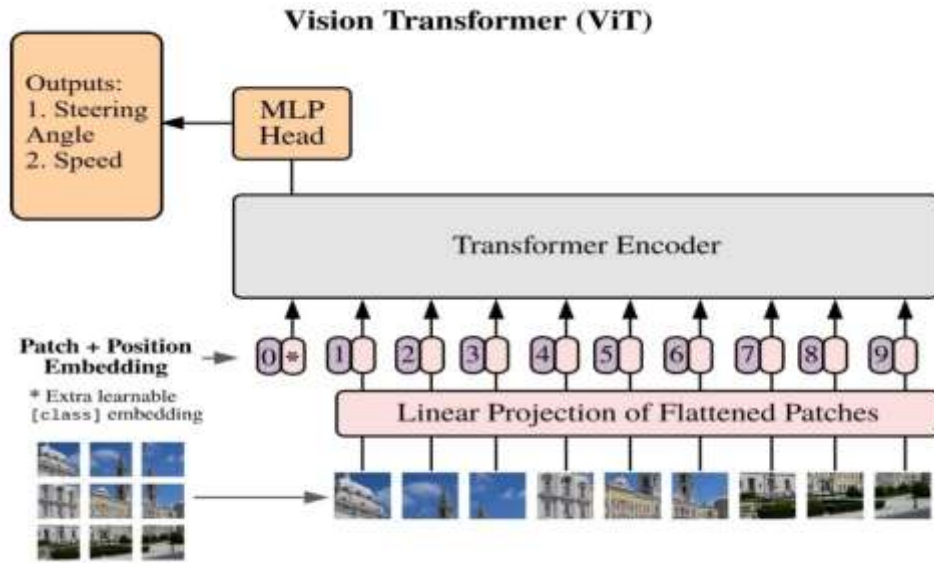
5- البنى المقترحة بالاعتماد على محول الرؤية:

البنية المقترحة الأولى: تتألف البنية الأولى المقترحة من محول الرؤية ViT كمستخرج ميزات، وهو نموذج مدرب مسبقاً على 14 مليون صورة بحجم 224×224 من مجموعة البيانات ImageNet-21k، وقد تم تعديل خرج المحول الذي يستخدم عادةً في مهام التصنيف ليصبح قادراً على أداء مهمة التنبؤ، حيث تم استبدال الطبقات الأخيرة بشبكة عصبونية Multi-Layer Perceptron Head (MLP) مؤلفة من ثلاث طبقات كاملة الاتصالية Fully Connected Layers (FC)، بحيث تحتوي الطبقة الأخيرة على عصبونين للتنبؤ بثنائية (زاوية توجيه، سرعة).

الجدول (1): البنية المقترحة الأولى.

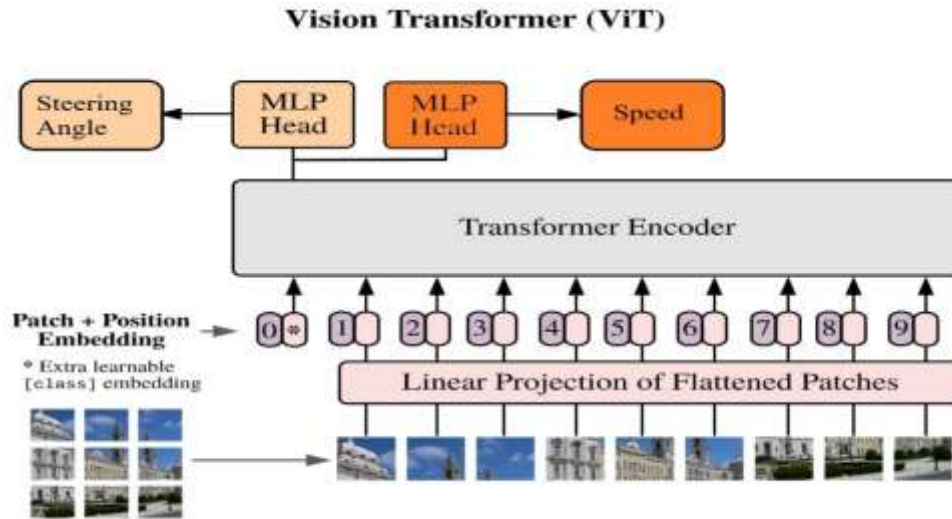
Input	Operator	Output
224×224×3	ViT	Last Hidden State 1×768
1×768	FC	1×512
1×512	FC	1×256
1×256	FC (2 Neurons)	1×2

يوضح الشكل (6) البنية المقترحة.



الشكل (6) معمارية النموذج المقترح الأول.

البنية المقترحة الثانية: تشبه هذه البنية سابقتها من حيث المبدأ، حيث تتألف من محول الرؤية ViT كمستخرج ميزات يتعامل مع صور دخل بحجم 224×224، ويعطي على الخرج شعاع ميزات بحجم 768×1. لكن يكمن الاختلاف بين البنيتين في آلية التنبؤ بزاوية التوجيه والسرعة، حيث تم اعتماد بنية متعددة المهام Multi-Tasking Architecture تتألف من رأسي تنبؤ MLP Heads، يتألف كل منهما ثلاث طبقات كاملة الاتصالية (FC)، لكن طبقة الخرج تحتوي عصبون واحد فقط. يتم تدريب الرأس الأول على التنبؤ بزاوية التوجيه، فيما يتم تدريب الثاني على التنبؤ بالسرعة. تم استخدام تابع خسارة L1 مستقل لكل رأس، وخسارة الشبكة المقترحة هي مجموع الخسارتين. يوضح الشكل (7) البنية المقترحة الثانية.



الشكل (7): معمارية النموذج المقترح الثاني.

6- مجموعة البيانات المستخدمة لتدريب البنى المقترحة:

يستخدم البحث مجموعة البيانات Comma 2K19 [34]، تتضمن هذه المجموعة أكثر من 33 ساعة من بيانات القيادة على الطريق السريع 280 في ولاية كاليفورنيا في الولايات المتحدة الأمريكية، وهي مقسمة إلى مجموعة فيديو هات ملتقطه خلال النهار والليل.

تم جمع البيانات بدقة باستخدام جهاز Comma ENOS، وهو عبارة عن جهاز مزود بمستشعرات مماثلة لتلك الموجودة في الهواتف الذكية الحديثة، مثل الكاميرا المواجهة للطريق ونظام تحديد المواقع العالمي (GPS) ومقاييس الحرارة ووحدة قياس ممانعة (IMU) ذات 9 محاور، بالإضافة إلى بيانات منظومة التحكم التي نحصل عليها من شبكة منطقة المتحكم (CAN) المضمنة بالمركبة [34]. تقدم مجموعة البيانات الغنية هذه رؤية شاملة لسلوكيات القيادة والظروف البيئية وديناميكيات المركبات على طول الطريق السريع المحدد، وبالتالي يمكن الاستفادة منها في تدريب نماذج القيادة الذاتية الشاملة كتلك المستهدفة في البحث.



الشكل (8) آلية جمع البيانات باستخدام جهاز Comma ENOS.

7- سيناريو التدريب:

تم تدريب النماذج المقترحة باستخدام 74752 إطاراً مستخرجة من مقاطع الفيديو التي توفرها مجموعة البيانات Comma 2k19، حيث تمت معالجة الفيديوهات بحيث يتم الحصول على 5 إطارات لكل ثانية، كما تمت إعادة تحجيم هذه الإطارات لتصبح بحجم 224×224، أما بالنسبة لقيم زاوية التوجيه والسرعة الموافقة لكل إطار، فقد تم أخذها من بيانات شبكة منظومة التحكم (CAN) الذي يعمل بتردد 10 Hz، وللمطابقة تم حساب التغير بين القيمة عند بداية زمن الإطار والقيمة عند نهايته (بداية الإطار التالي).

تم تدريب كلا النموذجين بنفس البارامترات التالية وهي: خمس دورات تدريب لكل نموذج باستخدام معالج رسومات Nvidia Tesla T4 بذاكرة وصول عشوائية 16 GB، مع معدل تعلم (Learning Rate) قدره 0.0001، وتوابع خسارة الموضحة تالياً لحساب الفرق بين القيم الحقيقية والقيم التي تم التنبؤ بها. تابع الخسارة للنموذج الأول:

$$L1\ Loss = \sum_{i=1}^n |y_{true} - y_{predicted}|$$

تابع الخسارة للنموذج الثاني:

$$Loss = L1_{Speed} + L1_{Angle}$$

8- معايير تقييم الأداء:

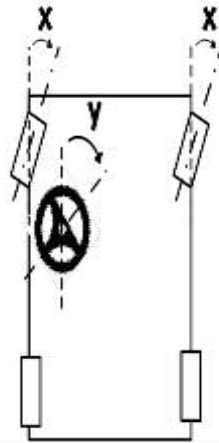
تم تقييم دقة نماذج التنبؤ المقترحة باستخدام مقياس تقييم الأداء متوسط الخطأ المطلق Mean Absolute Error (MAE)، والجذر التربيعي لمتوسط مربع الخطأ Root Mean Square Error (RMSE)، ومعامل التحديد R^2 . يعتبر معيار MAE مقياس كمي سهل الفهم للأخطاء في حل مشكلات التنبؤ والانحدار، في سياق التعلم الآلي، يشير الخطأ المطلق إلى حجم الاختلاف بين التنبؤ بملاحظة والقيمة الحقيقية لتلك الملاحظة، حيث يأخذ متوسط الأخطاء المطلقة لمجموعة من التنبؤات والملاحظات كمقياس لحجم الأخطاء للمجموعة بأكملها. يمكن أيضاً الإشارة إلى MAE باسم دالة الخسارة L1.

يقوم معيار RMSE بحساب متوسط الفرق بين القيم الحقيقية والقيم التي تم التنبؤ بها بواسطة النموذج، ومن خلال ذلك يقيس شدة الأخطاء ويوضح مقدار الانحراف عن القيم الحقيقية، لذلك يعتبر من المقاييس الجيدة في مجال تقييم أداء النماذج التنبؤية.

أخيراً يستخدم معيار معامل التحديد بشكل شائع في النماذج التنبؤية لتقييم مدى قدرة النموذج على تفسير التباين في المتغير التابع بناءً على المتغيرات المستقلة، ويعبر هذا المعيار عن النسبة المئوية للتباين الكلي في البيانات التي تم التنبؤ بها بواسطة النموذج، أي قدرته على تكرار النتائج.

9- نظام التوجيه:

إن تحقيق نظام التوجيه هي خطوة هامة من أجل التحقق من صحة عمل نظام التحكم، والذي يتضمن إيجاد العلاقة الحقيقية ما بين زاوية توجيه عجلة القيادة (Steering Angle) وزاوية عجلات السيارة الأمامية [30] كما هو موضح بالشكل 9.



الشكل (9): توضيح العلاقة الزاوية بين دوران عجلة القيادة والعجلتين الأماميتين للسيارة.

إن البحث الحالي يختص باستخراج قيمة خرج منظومة القرار (الزاوية Y) دون الخوض في التحقيق الدارتي لدارات القيادة اللازمة للوصول إلى الزاوية X أو سرعة المركبة.

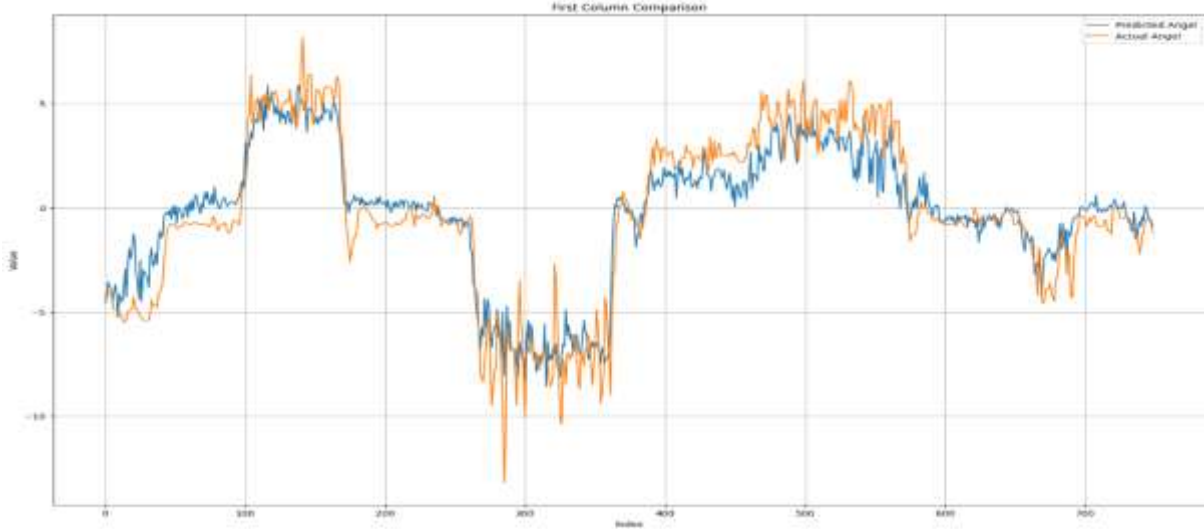
النتائج والمناقشة:

تم تقييم النماذج المقترحة باستخدام مجموعة اختبار مؤلفة من 18736 إطار، حيث تغطي هذه المجموعة حالات مختلفة للقيادة النهائية والليلية في مختلف الظروف الجوية، وقد تم تكرار التجربة ثلاث مرات فحققت النماذج النتائج التالية:

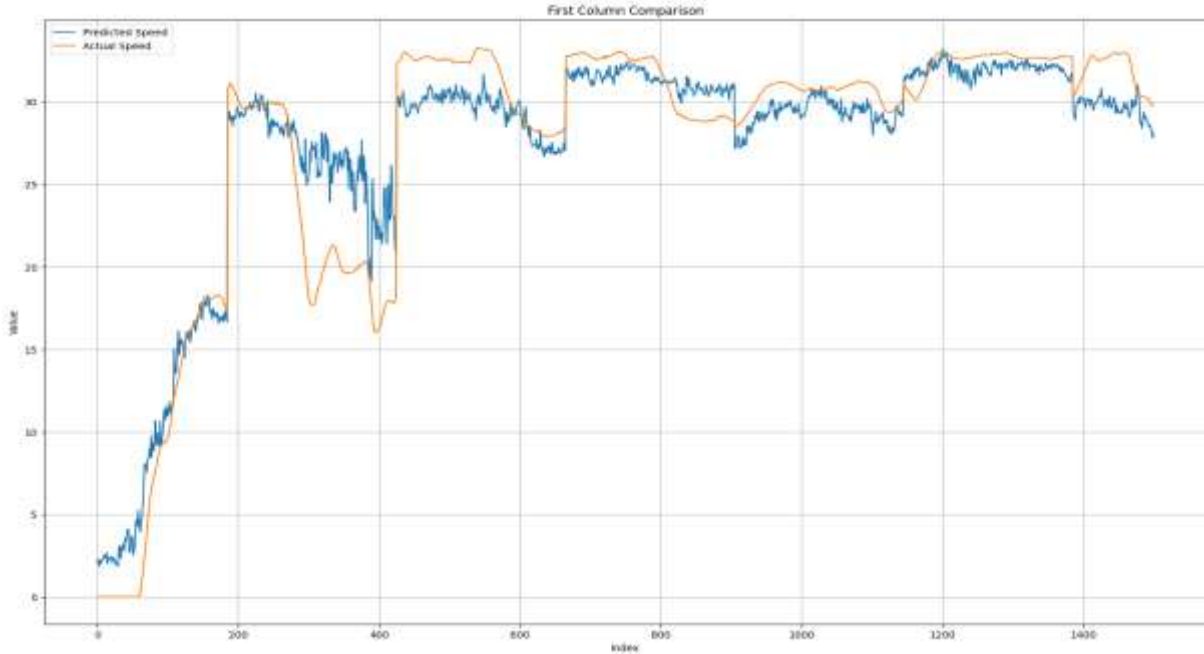
الجدول (2): مقارنة نتائج النماذج المقترحة بعد تكرار التجارب ثلاث مرات متتالية.

المتوسط التجارب	التجربة الثالثة	التجربة الثانية	التجربة الأولى	المعيار	
0.198	0.201	0.218	0.175	MAE	النموذج المقترح الأول
0.328	0.3308	0.3355	0.3167	RMSE	
0.815	0.8145	0.7980	0.8321	R^2	
0.368	0.3708	0.3632	0.3699	MAE	النموذج المقترح الثاني
0.6765	0.6776	0.6532	0.6985	RMSE	
0.7762	0.7769	0.743	0.8085	R^2	

تظهر النتائج الموضحة في الجدول (2) أن استخدام شبكة عصبونية وحيدة MLP على خرج محول الرؤية للحصول على قيمة التنبؤ لكلا زاويتي التوجيه والسرعة يعطي نتائج أفضل من استخدام شبكتين عصبونيتين منفصلتين على أساس اعتماد بنية متعددة المهام Multi-Tasking Architecture. يوضح الشكل (10) و(11) الفرق بين القيم الحقيقية والقيم التي تتنبأ بها النموذج الأول لكل من زاوية التوجيه والسرعة على الترتيب.



الشكل (10): الفرق بين قيم زاوية التوجيه الحقيقية والقيم التي تنبأ بها النموذج



الشكل (11): الفرق بين قيم السرعة الحقيقية والقيم التي تنبأ بها النموذج

يمكن الملاحظة من الشكل (10) وجود مناطق تكون فيها زاوية التوجيه المتوقعة قريبة جداً من القيم الفعلية، حيث يتم في هذه الأجزاء تتبع المسار فقط دون أداء أي مناورة، فيما تظهر بعض الانحرافات الملحوظة بين القيم المتوقعة والفعلية عند وجود تغيرات مفاجئة أو حادة في زاوية التوجيه، أي عند أداء مناورات مثل التجاوز. نظراً لكون الهدف من البحث هو البقاء ضمن المسار، فإن الأداء الحالي للنموذج يبدو معقولاً في معظم الحالات. يمكن تقليل الضجيج باستخدام مرشحات تنعيم.

أما بالنسبة لقيم السرعة الموضحة في الشكل (11) يمكن ملاحظة أن النموذج يعمل بشكل جيد في الحالات التي تكون فيها السرعة مستقرة أو تتغير بشكل تدريجي، أما عند التغيرات المفاجئة يتأخر النموذج في الاستجابة، مما يؤدي إلى فجوات بين القيم الفعلية والمتوقعة، وهو أمر طبيعي نظراً لكون الصور وحدها غير كافية للتنبؤ بسرعة المركبة.

ظهرت معظم الأخطاء بين القيم الحقيقية والمنتبئة لزاوية التوجيه والسرعة عند استخدام مجموعات اختبار جزئية تغطي حالات القيادة الليلية والظروف الجوية السيئة وأداء المناورات. بينما كان أداء النموذج جيداً في معظم الحالات التي تتطلبها ضمن حدود البحث.

الجدول (3): مقارنة نتائج نماذجنا المقترحة مع الدراسات المرجعية.

الدراسة المعيار	مجموعة البيانات	النموذج المقترح الأول	النموذج المقترح الثاني	الدراسة Bayarov et al. [10]	الدراسة Valiente et al. [14]	الدراسة Mohammad i et al. [15]	الدراسة Kalim et al. [13]	الدراسة Ali et al. [11]	الدراسة Garg et al. [12]
MAE	Comma.ai	0.198	0.328	0.691	0.510	0.896	0.491	0.527	0.593
RMSE		0.368	0.676	0.985	0.789	0.644	0.78	0.706	0.798

استناداً إلى البيانات الواردة في الجدول (2)، يمكن ملاحظة أن كلا النموذجي المقترحين حققا دقة تتبؤ عالية بالمقارنة مع بقية الدراسات المرجعية التي استخدمت نفس مجموعة البيانات ولكن بمنهجيات معالجة مختلفة وهذا يثبت فعالية استخدام التعليم بالتقليد لمحاولات الرؤية في مهمة القيادة الآلية على الطرق السريعة.

الاستنتاجات والتوصيات:

تم في هذا البحث دراسة جدوى استخدام المحولات التي تعتمد على آلية الانتباه الذاتي في مجال الرؤية الحاسوبية والتي أثبتت نجاحاً جيداً في مهام معالجة اللغات الطبيعية. حيث تبين فعاليتها في تضمين الجزء Patch Embedding للصور الملونة ومن ثم التنبؤ بالقيم المطلوبة والتي هي زاوية التوجيه وسرعة المركبة ذاتية القيادة على الطرق السريعة في الحالة الدراسية المطروحة في هذا البحث، وذلك بعد تدريبها على مجموعة بيانات تكون رؤية شاملة لسلوكيات القيادة والظروف البيئية وديناميكيات المركبات على طول الطريق السريع المحدد باستخدام نمط التعليم المعروف باسم استنساخ السلوك Behavior Cloning الذي يعتبر أحد أهم أشكال تقنية التعلم بالتقليد في مجال القيادة الذاتية.

المحددات: تظهر أخطاء النموذج بشكل أساسي في الحالات التي تمثل مناورات مختلفة عن متابعة المسار مثل التجاوز والتباطؤ والتسارع المفاجئين، لذلك يجب تطوير النموذج للتعامل مع هذه الحالات، أو إزالة الإطارات التي تمثل حالة مناورة وإعادة تدريب النموذج. كما أن التغيرات في ظروف الإضاءة مثل القيادة ليلاً أو في الظروف الجوية السيئة تمثل تحدياً كبيراً نظراً لكون النموذج يعتمد على الميزات المستخرجة من الصور الملتقطة باستخدام كاميرا أحادية فقط، حيث يستهدف البحث تحقيق المهمة المطلوبة بأقل تكلفة ممكنة، ومن المؤكد أن زيادة الميزات من خلال إضافة مستشعرات أخرى سيؤدي إلى الحصول على نتائج أفضل لاسيما بالنسبة للتنبؤ بسرعة المركبة.

تشكل نتائج هذا البحث حجراً أساساً لتصميم منظومة تحكم موثوق للسيارات ذاتية القيادة على الطرق السريعة والتي تعتبر نظاماً فرعياً بالغ الأهمية في منظومة التحكم الكاملة.

يمكن تطوير هذا البحث من حيث إيجاد آلية فعالة وفي الزمن الحقيقي للتحقيق الدارتي والبرمجي لنظام القرار المقترح من ناحية ومن ناحية أخرى إيجاد طريقة للتكامل بينه وبين بقية الأنظمة الفرعية الأخرى لمنظومة التحكم الشاملة. حيث من الممكن تدريب نظام قيادة ذاتية شامل End-to-End Architecture من أجل عدة مهام مجتمعة مثل اكتشاف الكائنات وتقدير البعد والتنبؤ بزاوية التوجيه وسرعة المركبة معاً وعلى مختلف أنواع الطرق والذي يتطلب

مجموعة بيانات ضخمة جداً وموارد حاسوبية هائلة وهذا ما تسعى إليه الشركات المنتجة للسيارات ذاتية القيادة في الوقت الحالي وتعمل من أجله.

References:

- [1] Akshitha K, Devanshi M, Manu K S, Artificial Intelligence in Self Driving Cars: Applications, Implications and Challenges. Ushus - Journal of Business Management 2022.
- [2] Kuutti S, Bowden R, Jin Y, Barber P, Fallah S. A survey of deep learning applications to autonomous vehicle control. IEEE Transactions on Intelligent Transportation Systems. 2021;22(2):712-733. doi:10.1109 /tits.
- [3] Yurtsever E, Lambert J, Carballo A, Takeda K. A survey of Autonomous Driving: Common practices and Emerging technologies. IEEE . 2020; 8:58443-58469. doi:10.1109/access.2020.2983149
- [4] Ryan M. The Future of Transportation: Ethical, legal, social and economic impacts of self-driving vehicles in the year 2025. Science and Engineering Ethics. 2019;26(3):1185-1208. doi:10.1007/s11948-019-00130-2.
- [5] Mohammadi A, Jamshidi K, Shahbazi H, Rezaei M. Efficient deep steering control method for self-driving cars through feature density metric. Neurocomputing. 2023; 515:107-120. doi:10.1016/ j.neucom.
- [6] Sharma R, Sungheetha A. An Efficient Dimension Reduction based Fusion of CNN and SVM Model for Detection of Abnormal Incident in Video Surveillance. Journal of Soft Computing Paradigm. 2021;3(2):55-69. doi:10.36548/jscp.2021.2.001
- [7] Khan LU, Yaqoob I, Tran NH, Han Z, Hong CS. Network Slicing: recent advances, taxonomy, requirements, and open research challenges. IEEE Access. 2020; 8:36009-36028. doi:10.1109/access.2020.2975072
- [8] Garcia P, Darroch F, West L, BrooksCleator L. Ethical Applications of Big Data-Driven AI on Social Systems: literature analysis and example Deployment use case. Information. 2020;11(5):235. doi:10.3390/ info11050235.
- [9] Nishime T, Endo S, Toma N, Yamada K, Akamine Y. Feature acquisition and analysis for facial expression recognition using convolutional neural networks. Transactions of the Japanese Society for Artificial Intelligence. 2017;32(5): F-8. doi:10.1527/tjsai. f-h34
- [10] Bayarov HB, Yi D, Sui J. Application of a brain-inspired deep imitation learning algorithm in autonomous driving. Software Impacts. 2021; 10:100165. doi: 10.1016/j.simpa.2021.100165.
- [11] S. Ali, “Self-driving cars: Automation testing using Udacity simulator,” Int. Res. J. Eng. Technol., vol. 8, no. 4, pp. 1–7, 2021.
- [12] N. Garg, K. S. Ashrith, G. S. Parveen, K. G. Sai, A. Chintamaneni, and F. Hasan, “Self-driving car to drive autonomously using image processing and deep learning,” Int. J. Res. Eng., Sci. Manage., vol. 5, no. 1, pp. 125–132, 2022.
- [13] M. K. A. Chy, A. K. M. Masum, K. A. M. Sayeed, and M. Z. Uddin, “Delicar: A smart deep learning based self driving product delivery car in perspective of Bangladesh,” Sensors, vol. 22, no. 1, p. 126, Dec. 2021.
- [14] Valiente, R., Zaman, M., Ozer, S., & Fallah, Y. P. (2019). Controlling Steering Angle for Cooperative Self-driving Vehicles utilizing CNN and LSTM-based Deep Networks. IV, 2423–2428.
- [15] A. Mohammadi, K. Jamshidi, H. Shahbazi, and M. Rezaei, “Efficient deep steering control method for self-driving cars through feature density metric,” Neurocomputing, vol. 515, pp. 107–120, Jan. 2023

- [16] Morga-Bonilla, S. I., Rivas-Camero, I., Torres-Jiménez, J., Téllez-Cuevas, P., Núñez-Cruz, R. S., & Perez-Arista, O. V. Behavioral cloning strategies in steering Angle Prediction: applications in mobile robotics and autonomous driving. *World Electric Vehicle Journal*, 15(11), 486.2024.
- [17] Huang, Y., Liu, W., Li, Y., Yang, L., Jiang, H., Li, Z., & Li, J. MFE-SSNET: Multi-Modal Fusion-Based End-to-End Steering Angle and Vehicle Speed Prediction Network. *Automotive Innovation* (2024).
- [18] Kale C. Autonomous Vehicles: applications of deep reinforcement learning. *International Journal for Multidisciplinary Research*. 2024;6(1). doi:10.36948/ijfmr.2024.v06i01.13792.
- [19] ESCWA. OECD Glossary (Transport Statistics) <https://stage-unescwa.unescwa.org/ar/sd-glossary/> Accessed 7-2024
- [20] Schwarting W, Alonso-Mora J, Rus D. Planning and Decision-Making for autonomous vehicles. *Annual Review of Control Robotics and Autonomous Systems*. 2018;1(1):187-210. doi:10.1146/annurev-control-060117-105157
- [21] Yao Y, Wang H, Li S, et al. End-To-End Convolutional Neural Network model for gear fault diagnosis based on sound signals. *Applied Sciences*. 2018;8(9):1584. doi:10.3390/app8091584
- [22] Ardi T, Maksym S, Muhammad N, Dmytro F, Tambet M, A survey of End-to-End Driving: Architectures and training methods, *Artificial Intelligence Robotics (cs.RO)*, Published April 1, 2022.
- [23] Liu S, Li L, Tang J, Wu S, Gaudiot JL. Creating autonomous vehicle systems.; 2020. doi:10.1007/978-3-031-01805-3
- [24] Kalra N, Paddock SM. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation Research Part A Policy and Practice*. 2016; 94:182-193. doi: 10.1016/j.tra.2016.09.010
- [25] Koopman P, Wagner M. Challenges in autonomous vehicle testing and validation. *SAE International Journal of Transportation Safety*. 2016;4(1):15-24. doi:10.4271/2016-01-0128
- [26] Batsch F, Kanarachos S, Cheah M, Ponticelli R, Blundell M. A taxonomy of validation strategies to ensure the safe operation of highly automated vehicles. *Journal of Intelligent Transportation Systems*. 2020;26(1):14-33. doi:10.1080/15472450.2020.1738231
- [27] Park SM, Kim YG. A metaverse: taxonomy, components, applications, and open challenges. *IEEE Access*. 2022; 10:4209-4251. doi:10.1109/access.2021.3140175
- [28] Gronauer S, Diepold K. Multi-agent deep reinforcement learning: a survey. *Artificial Intelligence Review*. 2021;55(2):895-943. doi:10.1007/s10462-021-09996-w
- [29] Ziebart BD, Maas A, Bagnell JA, Dey AK. Maximum entropy inverse reinforcement learning. *National Conference on Artificial Intelligence*. July 2008:1433-1438.
- [30] Brunke L, Greeff M, Hall AW, et al. Safe learning in Robotics: From Learning-Based Control to Safe Reinforcement Learning. *Annual Review of Control Robotics and Autonomous Systems*. 2022;5(1):411-444. doi:10.1146/annurev-control-042920-020211
- [31] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *International Conference on Learning Representations*. May 2021
- [32] Y Xiao, Felipe C, Akhil G, Onay U, Antonio M. L, Multimodal End-to-End autonomous driving. arXiv:1906.03199v2 [cs.CV] 25 Oct 2020.

- [33] Zhang Q, Tang M, Geng R, Chen F, Xin R, Wang L. MMFN: Multi-Modal-Fusion-Net for End-to-End driving. 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). October 2022. doi:10.1109/iros47612.2022.9981775
- [34] Schafer, H., Santana, E., Haden, A., & Biasini, R. (2018). A Commute in Data: The comma2k19 Dataset. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.1812.05752>.
- [35] Rosero LA, Gomes IP, Da Silva JAR, Przewodowski CA, Wolf DF, Osório FS. Integrating Modular Pipelines with End-to-End Learning: A Hybrid Approach for Robust and Reliable Autonomous Driving Systems. Sensors. 2024;24(7):2097. doi:10.3390/s24072097
- [36] Joukhadar A, Issa H, Kalaji Y. Design and implementation of auto car driving system with collision avoidance. Cogent Engineering. 2018;5(1):1485458. doi:10.1080/23311916.2018.1485458
- [37] A survey of Motion Planning and Control Techniques for Self-Driving Urban Vehicles. IEEE Journals & Magazine | IEEE Xplore. <https://ieeexplore.ieee.org/document/7490340>. Published March 1, 2016.
- [38] Katrakazas C, Quddus M, Chen WH, Deka L. Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions. Transportation Research Part C Emerging Technologies. 2015; 60:416-442. doi:10.1016/j.trc.2015.09.011
- [39] Yuan T, Da Rocha Neto WB, Rothenberg CE, Obraczka K, Barakat C, Turletti T. Machine learning for next-generation intelligent transportation systems: A survey. Transactions on Emerging Telecommunications Technologies. 2021;33(4). doi:10.1002/ett.4427
- [40] Hu Y, Abu-Dakka FJ, Chen F, et al. Fusion dynamical systems with machine learning in imitation learning: A comprehensive overview. Information Fusion. 2024; 108:102379. doi:10.1016/j.inffus.

