Performance of Objects Classification System in an Image using Convolutional Neural Networks

Dr. Mohammad Mazen Mahyry*
Dr. Qosai Kanafani**
Raneem H Kiwan***

(Received 21 / 11 / 2018. Accepted 10 / 6 / 2019)

□ ABSTRACT □

In recent years, the problem of classifying objects in images has increased by using deep learning as a result of the industrial sector requirements. Despite of many algorithms used in this field, such as Deep Learning Neural Network DNN and Convolutional Neural Network CNN, the proposed systems to address this problem Lack of comprehensive solution to the difficulties of long training time and floating memory during the training process, low rating classification.

Convolutional Neural Networks (CNNs), which are the most used algorithms for this task, were a mathematical pattern for analyzing images data. A new deep-traversal network pattern was proposed to solve the above problems. The aim of the research is to demonstrate the performance of the recognition system using CNNs networks on the available memory and training time by adapting appropriate variables for the bypass network. The database used in this research is CIFAR10, which consists of 60000 colorful images belonging to ten categories, as every 6,000 images are for a class of these items. Where there are 50,000 training images and 10,000 test tubes. When tested on a sample of selected images from the CIFAR10 database, the model achieved a rating classification of 98.87%.

Key words: Convolutional Neural Network, Deep learning, classification, convolutional layer, fully connected Layer, classification layer, Activation function.

* **Professor, Department of Computer Engineering and Automation, Faculty of Mechanical and electrical Engineering, Damascus University, Damascus, Syria.**

** **Associate Professor, Department of Basic Sciences, Faculty of Mechanical and electrical Engineering, Damascus University, Damascus, Syria.**

*** **Postgraduate Student in computer Engineering, Department of Computer Engineering and Automation, Faculty of Mechanical and electrical Engineering, Damascus University, Damascus, Syria.**

أداء نظام تصنيف الأشكال في الصورة باستخدام الشبكات العصبونية الالتفافية.

الدكتور محمد مازن المحايري*

الدكتور قصي كنفاني**

رنيم حافظ كيوان***

(تاريخ الإيداع 21 / 11 / 2018. قُبِل للنشر في 10/ 6 / 2019)

□ ملخّص □

في السنوات الأخيرة نمت مشكلة تصنيف الكائنات في الصّور نتيجة لمتطلبات القطاع الصناعي.على الرّغم من تعدد التقنيات المستخدمة للمساعدة في عملية التصنيف SIFT Scale Invariant Feature Transforms، ORB Oriented Fast And Rotated Brief , SURF Speed Up Robust Features، إضافة لشبكات التعلم العميق Deep Learning Neural Network DNN والشبكات العصبونية الالتفافية Convolutional Neural Network CNN، فإن الأنظمة المقترحة لمعالجة هذه المشكلة تفتقر للحل الشّامل للصعوبات المتمثلة بوقت التّدريب الطّويل والذاكرة العائمة أثناء عملية التدريب، وانخفاض معدّل التصنيف.

تعتبر الشبكات العصبونية الالتفافية Convolutional Neural Networks (CNNs) من أكثر الخوارزميات استخداما لهذه المهمة، فقد كانت نموذجا حسابيا لتحليل البيانات الموجودة في الصور. تم اقتراح نموذج شبكة التفافية عميقة جديد لحل المشاكل المذكورة أعلاه. يهدف البحث إلى إظهار أداء نظام التّعرف باستخدام شبكات CNNs على الذّاكرة المتاحة وزمن التدريب وذلك من خلال منهجة متغيرات مناسبة للشبكة العصبونية الالتفافية. قاعدة البيانات المستخدمة في هذا البحث هي CIFAR10 المكونة من60000 صورة ملونة تنتسب لعشرة أصناف، حيث أن كل 6000 صورة تكون لصنف من هذه الأصناف. يوجد 50000 صورة للتدريب و 10000 صورة للاختبار. حقق النموذج لدى اختباره على عينة من الصور المنتقاة من قاعدة البيانات CIFAR10 معدل تصنيف 98.87%، .

الكلمات المفتاحية: الشبكات العصبونية الالتفافية CNN، التعلم العميق، التصنيف، الطبقة الالتفافية، طبقة الاتصال الكامل، طبقة التصنيف، تابع التفعيل.

*أستاذ – قسم هندسة الحواسيب والأتمتة– كلية الهندسة الميكانيكية الكهربائية– جامعة دمشق– دمشق– سورية.

**أستاذ مساعد – قسم العلوم الأساسية – كلية الهندسة الميكانيكية والكهربائية– جامعة دمشق– دمشق– سورية.

***طالبة دراسات عليا(دكتوراه) – قسم هندسة الحواسيب والأتمتة– كلية الهندسة الميكانيكية والكهربائية– جامعة دمشق– دمشق– سورية.

مقدمة:

يعتبر التّعرف على الكائنات إحدى تقنيات الرّؤية الحاسوبية التي تعمل على إيجاد وتحديد الكائنات في الصّورة أو تسلسل الفيديو[7]. وبالرغم من اختلاف القياسات والأحجام واقتطاع أجزاء من الكائنات، يمكن للآلة التّعرف على هذه الكائنات. لكن مازالت هذه المهمة تعاني الكثير من الصعوبات ضمن أنظمة الرؤية الحاسوبية، مما يتطلب تقنيات ذات كفاءة عالية للمشاكل المتزايدة.[7]

تبيّن للباحثين في مجال الشبكات العصبونية العميقة، تفوّق مثل هذا النوع من الشبكات عن باقي الابتكارات في مجال الرؤية الحاسوبية، مولدة أداء مثالي في مجال تصنيف الصور. وقد أثارت الشبكات العصبونية الالتفافية Convolutional Neural Network CNN اهتماماً كبيرا بكونها أداة لدراسة الرؤية البيولوجية. منذ ذلك الحين، أخذت هذه الفئة من أنظمة الرؤية الاصطناعية تعرض قدرات التّعرف البصري القابلة للمقارنة مع ما يقابلها من المقدرات البشرية.[9]

تقوم النماذج المبتكرة في تحسين أداء عملية التّعرف، إضافة إلى فعاليتها في مجال التنبؤ. وقد ظهرت الشبكات CNN كأساليب ممتازة للتّعرف على الكائنات، حتّى أن التّطورات التكنولوجية سمحت باستخدام وحدات المعالجة الرسومية للأغراض العامة GPUs (Graphical Processing Units) لتسريع حل المشكلة الرقمية باستخدام هذا النهج.[10] وقد لجأ الباحثون إلى تقليل زمن الحسابات، والنظر في شبكات أكبر. وبالتالي أصبحت أجهزة الحاسب قادرة على التحرك بشكل أعمق وأوسع وبنماذج أكثر رصانة. كما حققت الشبكات CNNs أداء أقرب ما يكون لأداء الإنسان في عدة مهام للتعرف، مثل التعرف على الحروف المكتوبة بخط اليد[11]، التعرف على الوجوه[12]، بحث الصور[13] وغير ذلك الكثير.

في السنوات القليلة الماضية، تمت دراسة طرائق التعلم العميق deep learning على نطاق واسع من أجل تصنيف الصور image classification ومهام معالجة الصور. تستخدم الشبكات العصبونية العميقة Deep Neural Network (DNN) تصميم الطبقة العميقة لاستخلاص السمات الكامنة من بيانات الصورة وبالتالي تحقيق إمكانية تصنيف النماذج بشكل مناسب. تمتلك الطبقات في الشبكات العصبونية الالتفافية ConvNet Convolutional Neural Network (CNN) عصبونات مرتبة بشكل ثلاثي البعد: عرض، طول، عمق.(يشير العمق هنا إلى البعد الثالث لحجم التفعيل،وليس إلى عمق الشبكة العصبونية الكاملة). تتصل العصبونات في طبقة ما إلى منطقة صغيرة من عصبونات الطبقة التي قبلها، بدلا من الاتصال بشكل كامل fully connected مع جميع الخلايا العصبونية في الطبقة السابقة، كما هو الحال في الشبكات العصبونية العادية. تتكون البنية العامة للشبكات العصبونية من قسمين: قسم لاستخلاص السمات(الطبقات الالتفافية)، والقسم الآخر للتصنيف (طبقة الاتصال الكامل).

في العام 2018 [1] كان الهدف من الدراسة التي قام بها الباحثون Sudipta و Mahtab و Muhammad هو التحقق من نظام تصنيف الصور ذات الضجيج العالي بالاعتماد على DNN، وقد اعتمد الباحثون على تقنيات إزالة الضجيج باستخدام مرمز أوتوماتيكي Autoencoder (AE) وذلك لإعادة بناء الصورة الأصلية من صورة الدخل المشوشة. لتقوم بعدها الشبكة العصبونية الالتفافية CNN بتصنيف الصور المعاد بناؤها، حيث أن شبكة CNN هي مثل الشبكة DNN مع القدرة على الحفاظ على تمثيل أفضل للهيكل الداخلي لبيانات الصورة. تم الاعتماد على مجموعة متنوعة من المرمزات الأوتوماتيكية وذلك ضمن خطوة إزالة الضجيج ومنها المرمز الأوتوماتيكي لإزالة

الضجيج denoising autoencoder (DAE)، والمرمز الأوتوماتيكي الإلتفافي لإزالة الضجيج convolutional denoising autoencoder (CDAE).

في العام 2017 [2] قام Miroslav وباحثون آخرون بدراسة التعرف على العواطف من المسارات الموسيقية، مقترحين خوارزمية تعتمد على الشبكات العصبونية الالتفافية CNN والشبكات العصبونية التكرارية (Recurrent Neural Network) RNN ، تمتلك هذه الخوارزمية برامترات أقل بشكل واضح مقارنة مع خوارزميات أخرى لنفس المهمة. تم في هذه الخوارزمية استخدام طبقة واحدة من الشبكة CNN متبوعة بفرعين من الشبكة RNN. تم تقييم هذه الخوارزمية باستخدام مجموعة البيانات "MediaEval2015 emotion in music"، محققين أفضل نتيجة تم الحصول عليها في مجموعة البيانات هذه.

عمل Robert و Semcon في العام 2016 [3] على تقييم أداء عمل الشبكات العصبونية الالتفافية المدربة مباشرة على بيانات ثلاثية البعد، متجاوزين فقدان البيانات من خلال استخلاص البيانات أو التحويلات transformation والسماح للكثافة بتحديد النقاط التي ستستخدم. تم تقييم فعالية هذه الخوارزمية على مجموعة بيانات تم إنشاؤها من قبل KITTI Vision Benchmarking Suite. أظهرت النتائج أن مجموع نقاط الدقة 96.35% ومتوسط الدقة 95.67% على مجموعة بيانات تدريب لسبعة أصناف.

اقترح Ming وآخرون في العام 2015 [4] شبكة عصبونية التفافية متكررة Recurrent CNN (RCNN) من أجل التعرف على الكائنات، حيث قاموا بدمج الروابط المتكررة ضمن كل طبقة تلافيفية. على الرغم من الدخل الثابت، فإن عمل وحدات RCNN يتطور بمرور الوقت لدرجة أن نشاط كل وحدة يتعدل بنشاط الوحدات المجاورة لها. وجد الباحثون أن هذه الخاصية تعزز قدرة النموذج على دمج المعلومات مما يفيد عملية التعرف على الكائنات. تم اختبار النموذج على أربعة قواعد بيانات للتعرف على الكائنات وهي: CIFAR-10, CIFAR-100, MNIST و SVHN، فوجدوا أن النموذج RCNN يتفوق على أحدث النماذج المطبقة على قواعد البيانات هذه. وبزيادة عدد البرامترات يتم الوصول إلى أداء أفضل.

في العام 2018 [5]، تمكنت التّطورات الحديثة في مجال التعلم العميق لتحديد واكتشاف الكائنات من أتمتة تحليل الصور. قام الباحثان Stefan و Graham بشرح قدرة التعلم العميق عن طريق تدريب ومقارنة مصنفي تعلم عميق لاكتشاف الكائنات وتحديد موقعها في الصور وهما Faster R-CNN و YOLO v2.0. أظهرت خوارزمية التعرف على الكائنات نجاحا كبيرا عند تدريبها على قاعدة بيانات كبيرة. بينت التجارب أن المصنِّف Faster R-CNN يتفوق على المصنف YOLO v2.0 بمتوسط دقة 93.0% و 76.7% على التوالي.

في العام 2017 [6] عمل الباحثون S. Ben Driss, M. Soua, R. Kachouri, and M. Akil على المقارنة بين عمل الشبكات العصبونية الالتفافية CNN وشبكات البيرسيبترون متعددة الطبقات Multilayer perceptron (MLP)، حيث قاموا بتشكيل مناسب للشبكة مستخدمين واصفات UCD. صممت شبكة CNN من أجل التعرف على المحارف المطبوعة والمكتوبة بخط اليد وتم تكييفها لتتوافق مع 62 صنف تشمل كلا من الأرقام والمحارف. وقد بينت التجارب أن شبكات CNN الوقت الحقيقي أكثر ملائمة بمرتين من شبكات MLP عند تصنيف المحارف.

أهمية البحث وأهدافه:

هناك العديد من الطّرائق المقترحة لعملية التعرف على الكائنات وقد تمّ تنفيذها عمليا خلال العقود الماضية. بالرّغم من ذلك، لا يزال هناك ضعف بالحلول العامّة والشّاملة لصعوبات التّعرف الحديثة، كما في مجال المراقبة الأمنية حيث يزداد عدد الكاميرات Closed-Circuit Television (CCTV) باطراد لنتمكن من التركيز على أشكال بمواصفات معينة بحال وجود اشتباه بشيء ما، أو ملاحظة بقاء أشياء لفترات زمنية مريبة. وكذلك الأجهزة الرقمية التي تتطلب تقنيات كشف فعالة.

أصبحت الأجهزة المحمولة قوية بما يكفي للتعامل مع الحسابات المطلوبة لتنفيذ نماذج CNN بأقرب زمن حقيقي. مع أخذ ذلك بعين الاعتبار، تم تطوير البحث بالاعتماد على أبسط وأقوى الطرق في مجال التعرف الآلي بالاعتماد على نموذج CNN ذات الأوزان الخفيفة وبشكل خاص لمهام التعرف على الكائنات باستخدام أقل ما يمكن من الموارد للأجهزة الطرفية النهائية.

يهدف البحث الحالي إلى :

1- التوصل لنظام يقوم بالتوصيف الآلي لشكل موجود في صورة باستخدام الشبكات العصبونية الالتفافية CNN.

2- تقليل كمية الحسابات الضخمة التي تتطلبها CNN، من خلال انتقاء الحجوم المناسبة للفلاتر ضمن الطبقات وحجم الخطوة Stride وتخفيض حجم الخرج لطبقات Max pooling.

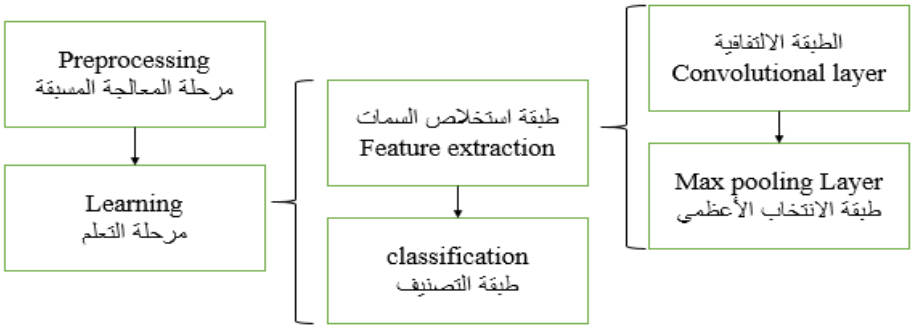
3- استخدام الحد الأدنى من الذاكرة للحسابات التي تتطلبها مثل هذه الأبحاث.

طرائق البحث ومواده:

تم اقتراح نسخة بسيطة وفعالة من شبكات CNN لمعالجة مهمة التعرف على الأشياء بالزمن الحقيقي، حيث تمكنت هذه الشبكة المقترحة من تحقيق أداء ممتاز بموارد حاسوبية ومادية قليلة التكاليف.

تم تطبيق عملية التدريب على حاسوب شخصي يمتلك وحدة معالجة مركزية dual-core Intel Core i3-5010U بسرعة 2.1 GHz ، وذاكرة نظام 4GB. استُخدِم لتدريب النموذج برنامج ماتلاب R2017a 64-bit February 23,2017 ضمن مكتبات التعلم العميق، باستخدام وحدة المعالجة الرسومية Graphical Processing Units(GPUs).

يتضمن البحث عدة مراحل بدءا من مرحلة المعالجة المسبقة ومن ثم التدريب وصولا إلى مرحلة التنبؤ. يبين الشكل (1) المخطط الصندوقي المعبر عن التسلسل العام للمراحل المنفذة.



الشكل(1) المخطط الصندوقي لمراحل العمل

1- مرحلة المعالجة المسبقة:

تتضمن قاعدة البيانات المستخدمة مجموعة كبيرة من الصور لكائنات معروفة ومتنوعة مثل الطائرات والأسماك والطيور والقطط وغيرها وذلك لجعل الشبكة واسعة الطيف وقادرة على التعرف على عدد كبير من الأنواع. الصور ملونة بمكونات الأحمر والأخضر والأزرق RGB، ما يجعلها متناسبة مع دخل الشبكة CNN ثلاثي البعد. تختلف أبعاد الصور في قاعدة البيانات إلا أنه ولضرورة العمل تم جعلها جميعا بنفس الأبعاد 132x132x3 بكسل وذلك لأن دخل الشبكة محدد بأبعاد ثابتة. يبين الشكل(1) بعضا من صور قاعدة البيانات.

الشكل(2) صور من قاعدة البيانات.

إن نوع الضغط لهذه الصور هو jpeg، يؤدي هذا النوع من الضغط إلى خسارة بعض البيانات بحيث أنه لا يمكن استرداد هذه البيانات عند فك الضغط. تصل معدلات الضغط في هذا النوع إلى أكثر من 40% ، حيث تقوم هذه المنهجيات بالاستفادة من الخصائص الإدراكية للعين والأذن البشرية.

تم اقتراح استخدام تقنية تسوية الهيستوغرام لتحسين التباين في الصورة وذلك بتنفيذ تابعي كثافة الاحتمال والكثافة التراكمية على الصورة. يتضمن هذان التابعان حساب عدد البكسلات من كل لون في الصورة وإيجاد جمع هذا العدد. ثم وببساطة وبقياس الخرج scaling ، يتم إنجاز تسوية الهيستوغرام. يمكن ملاحظة أنه بالإمكان إنجاز تسوية الهيستوغرام للصورة دون الحاجة لمساحة مؤقتة (حيث يمكن الكتابة فوق قيم البكسل).

لإنجاز العمليات السابقة باستخدام برنامج الماتلاب يجب استخدام التعمليات الأساسية التالية:

1- قراءة الصورة بالأبعاد الثلاثة (x,y,color). حيث يمكن لبرنامج الماتلاب قراءة الصور بتنسيقات مختلفة.

```
img= imread (image_name);
```

2- تحويل الصورة الملونة RGB إلى صورة بتدرج الرمادي:

```
imgGray = rgb2gray(img);
```

3- إجراء الهيستوغرام للصورة حيث تعيد التعليمة التالية النتيجة ضمن المصفوفة myHist. يمكن لهذه المصفوفة أن تُستخدم لتحليل الرسم البياني للصورة.

```
myHist = imhist(imGray);
```

4- إجراء تسوية الهيستوغرام للصورة، حيث يمكن إجراء تسوية الهيستوغرام بدون الخطوة 3.

```
eqImage = imhisteq(imGray);
```

2- مرحلة التعلم Learning:

بيّنت الأعمال السابقة أن النموذج البسيط لشبكات التعلم العميق له تأثير سلبي على المهمة المعتبرة المراد تنفيذها، لذا فقد كان التوجه مُلحّ لزيادة حجم وتعقيد النموذج المستخدم[9].

تتضمن الشبكة العصبونية الالتفافية العميقة عددا كبيرا من الطبقات. تتضمن كل طبقة عددا كبيرا من العصبونات ترتبط إلى عدد من عقد الطبقة السابقة، أي أن خرج كل عقدة من الطبقة L هو عبارة عن تابع لمخرجات العُقد في الطبقة L-1. [7]

تتقسم الطبقات إلى جزأين أساسيين[8] : طبقات لاستخلاص السمات وطبقات للتصنيف. كل طبقة من طبقات استخلاص السمات تستقبل خرج الطبقة السابقة لها مباشرة كدخل، وتقوم بتمرير خرجها كدخل للطبقة اللاحقة. وتكون العمارة الأساسية للشبكة المقترحة عبارة عن دمج ثلاثة أنواع من الطبقات: الطبقة الالتفافية وطبقة الانتخاب الأعظمي max-pooling وطبقة التصنيف classification. حيث أن الطبقات بالمستويات المتوسطة والمنخفضة للشبكة هي الطبقات الالتفافية ذات الترتيب بالأرقام الزوجية وطبقات الانتخاب الأعظمي ذات الترتيب بالأرقام الفردية.

تتميز الطبقة الالتفافية بأوزانها (قيمة فلاترها)[7]. تقوم كل عقدة من عقد الطبقة الالتفافية باستخلاص السمات باستخدام العمليات الالتفافية على العُقد المدخلة. حيث يتم تطبيق عدة التفافات بحجم ثابت لكل طبقة، وبخطوات ثابتة Stride على كامل الدخل. يتم تجميع عُقد الخرج ضمن مخطط ثنائي البعد يدعى خريطة السمات feature mapping يُشتق عادة من مخطط أو أكثر من الطبقات السابقة. حيث يتم دمج convolved مخططات السمات من الطّبقات السابقة مع الفلاتر القابلة للتعلم. يذهب خرج هذه الفلاتر عبر توابع تفعيل خطية أو لا خطية مثل sigmoid و hyperbolic tangent و Softmax و rectified linear وتوابع المطابقة identity functions. ينتج عن التعلم في الطبقة الالتفافية الأولى خريطة السمات ذات المستوى المنخفض كالحواف والخطوط والزوايا. بينما ينتج عن التعلم في الطبقات التالية تمثيلات أكثر تعقيدا كالأجزاء parts والنماذج models. وكلما ازدادت الشبكة عمقا وازداد عدد مخططات السمات، اكتسبت مستويات أعلى من الميزات وحصلنا على تمثيل أفضل لسمات صور الدخل مما يضمن دقة في التصنيف .

كل واحدة من خرائط السمات على الخرج يمكن أن تُدمج بأكثر من خريطة سمة على الدخل. تتناقص أبعاد خريطة السمات تبعا لحجم الفلاتر المستخدمة في العمليات الالتفافية وعمليات التجميع الأعظمي. وبشكل عام لدينا العلاقة (1):

$$x_j^l = f\left(\sum_{i \,\epsilon\, M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \qquad (1)$$

حيث إن x_j^l خرج الطبقة الحالية، x_i^{l-1} خرج الطبقة السابقة، k_{ij}^l فلتر الطبقة الحالية، b_j^l انحياز الطبقة الحالية. يمثل M_j واحدا من مخططات الدخل. يعطى انحياز إضافي b لكل مخطط دخل.

تشترك مع الطبقة الالتفافية في استخلاص السمات طبقة أخرى تقوم بعملية تخفيض أبعاد مخططات الدخل وعادة ما تُعرف بطبقة pooling. في هذا النوع من الطبقات فإن عدد مخططات السمات لكل من الدخل والخرج لا يتغير. أي، في حال وجود N مخطط على الدخل سيكون هناك N مخطط على الخرج. ولكن وكنتيجة لعملية تخفيض الأبعاد فإن حجم كل بعد لمخططات الخرج سيقل بالاعتماد على حجم قناع تخفيض الأبعاد. على سبيل المثال، إذا كان حجم

الفلتر المستخدم لتخفيض الأبعاد 2x2 فستكون أبعاد الخرج نصف أبعاد الدخل لكل الصور.يمكن صياغة هذه العملية بالعلاقة (2):

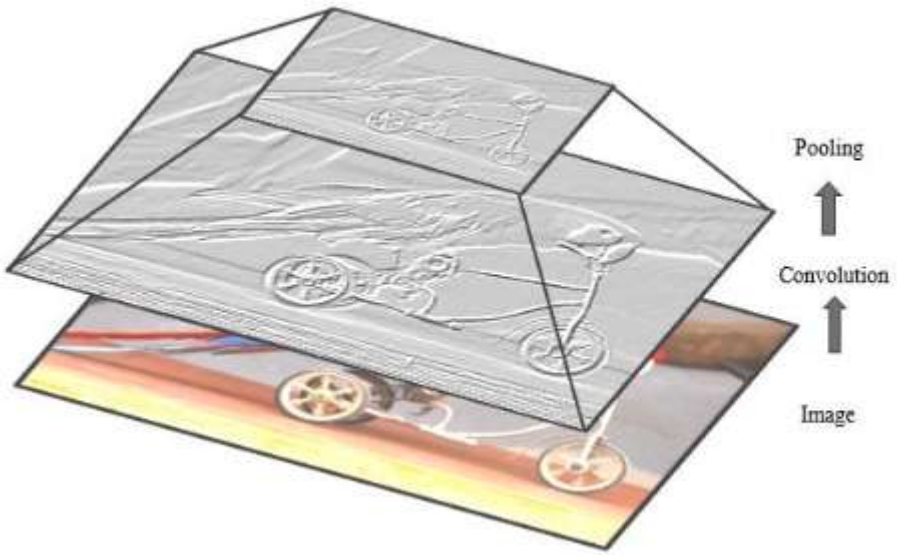
$$x_j^l = down\ (\ x_j^{l-1}\ ) \qquad (2)$$

يمثل ()down تابع تخفيض الأبعاد. يتم تنفيذ نوعين من العمليات في هذه الطبقة وهما average pooling و max pooling. في average pooling يقوم التابع بالجمع عبر NxN من خريطة السمات للطبقة السابقة وتختار القيمة المتوسطة. أما في حالة max pooling، يتم اختيار أكبر القيم في NxN من خريطة السمات التي يقوم الفلتر بالالتفاف عندها. لذا سيتم تخفيض أبعاد خريطة السمات لا خطيا بمقدار n مرة.

إضافة لما سبق، هناك طبقة التفعيل Activation Layer، وفيها توابع التفعيل التي يحاكي سلوكها سلوك الحبل العصبي الذي يرسل الإشارة عند وجود المحفزات. من توابع التفعيل الأكثر شيوعا Sigmoid, Rectified Linear Units (ReLU), Hyperbolic Tangent. يعتبر التابع ReLU أحد السمات المرافقة للشبكات CNN ويعرف بالعلاقة (3):

$$F(x) = \max(0, x) \qquad (3)$$

يوضح الشكل (3) نتيجة تنفيذ العمليات الالتفافية والانتخاب الأعظمي على الصورة المدخلة.



الشكل (3) مثال عن العمليات الالتفافية والانتخاب الأعظمي.

يُستخدم خرج آخر طبقة من طبقات استخلاص السمات كدخل لطبقة التصنيف classification layer وهي طبقة الاتصال الكامل(Fully Connected (FC والتي تقوم بحساب الهدف لكل صنف من السمات المستخلصة من الطبقة الالتفافية في الخطوات السابقة. يُمثل مخطط السمات للطبقة الأخيرة بشعاع ذي قيم قياسية تُمرر للطبقة FC. حيث يتم اختيار العدد المرغوب من السمات كدخل مع الأخذ بعين الاعتبار أبعاد مصفوفة الأوزان للشبكة العصبونية النهائية. تختلف هذه الطبقة عن سابقاتها بأن جميع مخرجات الطبقة السابقة مرتبطة بجميع مدخلات الطبقة FC.

لا يوجد هناك قاعدة من خلالها يتم تحديد عدد الطبقات FC المدرجة في الشبكة. إجمالا، تتراوح عدد الطبقات في أغلب الشبكات بين اثنيتين وأربع كما في شبكات LeNet، Alex Net و VGG Net. تعتبر طبقات الاتصال

الكامل FC مكلفة من ناحية الحسابات، لذا فقد لجأت العديد من الاقتراحات الأخرى لبنى الشبكات باستبدالها بطبقات global average pooling و طبقة average pooling ما يمكِن من تقليل برامترات الشبكة بشكل ملحوظ.

يتم أيضا إسقاط مجموعة عشوائية من العناصر في الصورة بإسناد قيمتها إلى الصفر باحتمالية معينة وغالبا ما تكون 0.5، ما يجعل الشبكة قادرة على تحقيق التصنيف الصحيح أو إيجاد الخرج لمثال محدد. ويكون ذلك باستخدام طبقة التسريب Dropout Layer والتي تشترك بعملها مع طبقة التصنيف.

يتم حساب درجة الصنف المرغوب ضمن طبقة soft-max. بالاعتماد على الدرجة الأعلى يقوم المصنف بإعطاء الخرج للصنف المناسب. حيث استُخدم تابع التفعيل Softmax في طبقة الخرج لتحويل الخرج إلى قيم احتمالية وبالتالي سيتم اختيار صنف واحد من بين الأصناف المعتبرة.

استُخدم المحسن Efficient Stochastic Gradient Descent (SGD) لتعليم الأوزان، كما استُخدم التابع اللوغاريتمي cross-entropy error الذي يقيس أداء نموذج التصنيف الذي خرجه عباره عن قيمة احتمالية تقع بين 0 و 1. يزيد قيمة هذا التابع والمعطى بالعلاقة (4) مع زيادة الاحتمالية للصنف المعتبر:

$$L(X,Y) = -\frac{1}{n}\sum_{i=1,n} y^i \; ln\, ax^i + \left(1 - y^i\right)\ln\left(1 - ax^i\right) \qquad (4)$$

حيث أن X={ x(1), …,x(n)} هي مجموعة صور الدخل في قاعدة بيانات التدريب، و Y={y(1), …,y(n)} هي المجموعة المقابلة من العناوين من أجل أمثلة الدخل. تمثل $a(x)$ تمثل خرج الشبكة العصبونية التي يكون دخلها x.

البرامترات ومتطلبات الذاكرة في الشبكات العصبونية الالتفافية:

يمثل عدد البرامترات الحسابية في الشبكة العصبونية الالتفافية مقياسا مهما لقياس مدى تعقيد نموذج الشبكة. يُصاغ حجم خريطة السمات على الخرج بالصيغة (5):

$$M = \frac{(N-F)}{S} + 1 \qquad (5)$$

تمثل N أبعاد خريطة السمات على الدخل، تشير F إلى أبعاد الفلتر، أما M فهي أبعاد خريطة السمات على الخرج، بينما تشير S إلى طول الخطوة. يتم تنفيذ الحشو Padding خلال العمليات الالتفافية لضمان امتلاك خرائط السمات الأبعاد نفسها على الدخل والخرج. يمكن تحديد كمية الحشو بالاعتماد على حجم الفلتر. تُستخدم الصيغة (6) لتحديد عدد الأعمدة والأسطر للحشو.

$$P = (F-1)/2 \qquad (6)$$

حيث إن P هو عدد الأسطر أو الأعمدة على حدى المراد حشوها بقيمة الصفر، و F أبعاد الفلتر.

هناك عدة معايير لمقارنة نماذج الشبكات المقترحة. عادة ما يستخدم في المقارنة عدد برامترات الشبكة ، والحجم الكلي المستخدم لذاكرة الحاسب. يمكن تحديد عدد برامترت الشبكة $(param_l)$ للطبقة l^{th} بالحساب باستخدام المعادلة (7):

$$param_l = (F \times F \times FM_{l-1}) \times FM_l \qquad (7)$$

في حال إضافة الانحياز إلى الأوزان، عندئذ ستصبح المعادلة السابقة كما يلي:

$$param_l = (F \times (F+1) \times FM_{l-1}) \times FM_l \qquad (8)$$

يمثل العدد الكلي للبرامترات بالنسبة للطبقة l^{th} من خلال p_l ، FM_l يمثل العدد الكلي لخرائط السمات على الخرج، و FM_{l-1} يمثل العدد الكلي لخرائط السمات على الدخل أو القنوات. على سبيل المثال، لو فرضنا أن الطبقة l^{th} تمتلك $FM_{l-1} = 32$ و $FM_l = 64$ وحجم الفلتر $F = 5$، عندئذ يكون العدد الكللي للبرامترات مع انحياز لهذه الطبقة هو:

$$param_l = (5 \times (5+1) \times 32) \times 64 = 61440$$

لذا فإن حجم الذاكرة المطلوب للعمليات في الطبقة l^{th} يُعبر عنه بالعلاقة (9):

$$Mem_l = (N_l \times N_l \times FM_l) \qquad (9)$$

بنية الشبكة المقترحة:

يعتبر اختيار البنية الصحيحة للشبكة العصبونية الالتفافية من الصعوبات التي تواجه هذا النوع من الشبكات، حيث تتطلب العملية التعامل بكثير من الدقة مع عدد من البرامترات، كتحديد عدد الطبقات اللازمة وحجم الفلاتر وعددها ، وكذلك القيم الصحيحة لحجم الخطوة Stride. حيث إنه لا يوجد قواعد ومعايير ثابتة لجميع هذه البرامترات. وذلك لأن هذه الشبكة تختلف حسب نوع البيانات وتعقيدها وحجم الصورة وموارد الأجهزة المتاحة وغير ذلك الكثير.[9]

البنية المقترحة لهذا البحث تتألف من تتابع للطبقات (convolutional, Dropout, convolutional, max-pooling) الموضوعة بشكل متسلسل. يتم تكرار هذا النموذج ثلاث مرات وحجم الفلاتر المختارة التي تعبر عن عدد خرائط السمات في خرج كل طبقة هي 32، 64، 128 على التوالي لكل تكرار. يؤدي ذلك إلى زيادة عدد خرائط السمات ولكن بحجوم تصغر بعد كل طبقة max pooling. إن عملية الحشو padding تُنفذ في الطبقات الالتفافية، حيث يتم الحشو عند الحاجة بالتساوي يمنة ويسرة ; في حال كان عدد الأعمدة المراد إضافتها فردي سيتم إضافة عمود إضافي إلى اليمين. ويطبق نفس المنطق عموديا حيث يكون هناك صف إضافي من الأصفار في الأسفل. وأخيرا تم إضافة عدد من الطبقات عالية الكثافة في نهاية خرج الشبكة للحصول على ترجمة أفضل لخرائط السمات إلى الصنف المرغوب. يوضح الجدول التالي البنية الأساسية للشبكة CNN المقترحة بطبقاتها مبينا معلومات كل طبقة على حدى:

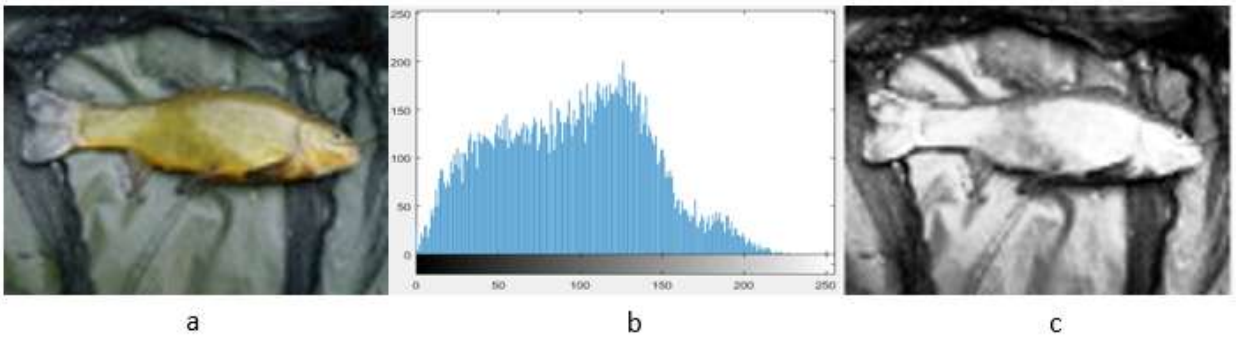
الجدول(1) البنية الأساسية للشبكة CNN المقترحة.

الطبقة الاسم/الرقم	نواة/ عصبونات	تابع التفعيل	عدد القنوات	حجم الخرج	البرامترات
Convolution1	3x3	ReLU	32	32x32	896
Dropout 1(20%)			32	32x32	0
Convolution2	3x3	ReLU	32	32x32	9248
Max-pooling 1	2x2		32	16x16	0
Convolution 3	3x3	ReLU	64	16X16	18496
Dropout 2 (20%)			64	16x16	0
Convolution 4	3X3	ReLU	64	16X16	36928
Max pooling 2	2x2		64	8x8	0
Convolution 5	3x3	ReLU	128	8x8	73856
Dropout 3 (20%)			128	8x8	0
Convolution 6	3x3	ReLU	128	8X8	147584
Max pooling 3	2x2		128	4x4	0
Dropout 4 (20%)	2048				0
Fully connected 2	1024	ReLU			2098176

Dropout 5 (20%)	1024				0
Fully connected 2	512	ReLU			524800
Dropout 6 (20%)	512				0
Fully connected 3	10	softmax			5130

التدريب Training :

تمت معالجة الصور المنتقاة من قاعدة البيانات مسبقا قبل إدخالها على الشبكة باستخدام عملية تسوية الهيستوغرام، يبين الشكل (3) نتائج عملية تسوية الهيستوغرام لإحدى الصور في قاعدة البيانات.



الشكل (3): a الصورة الأصلية، b هيستوغرام الصورة، c تسوية الهيستوغرام للصورة.

تم تدريب النموذج الذي تمت مناقشته باستخدام التابع Loss function مع خوارزمية التحسين stochastic optimization descent ، وقد بدأت بمعدل تعلّم 0.01. نُظِّمت الشبكة على 50 دورة epochs بحجم رقعة كبير 64 ، تم التوصل إليها من خلال بعض التجارب الثانوية. تأخذ كل دورة متوسط زمني للحسابات 145 ثانية، مما يجعل مرحلة التدريب تحتاج تقريبا لساعتين من الزمن. تمثل كل دورة جلسة تدريب مصغرة تعمل على جميع البيانات الواردة في مجموعة التدريب. أثناء هذا التشغيل، تم ضبط قيم الفلاتر (أو الأوزان) من خلال ما يسمى بالانتشار العكسي. حيث كان ذلك جزءا تناغمت فيه أوزان الطبقات آخذة بعين الاعتبار التابع loss function أثناء القيام بالمرور الأمامي والمرور العكسي. والهدف النهائي هو تحقيق مجموعة من البرامترات التي من الممكن تعميمها على بيانات جديدة. انعكس هذا الأمر على دقة التعرف.

مناقشة نتائج الاختبار Testing Result:

إن تشغيل النموذج المقترح على مجموعة بيانات التدريب أعطى عدة قيم لدقة التصنيف تبعا لقيم التابع loss function خلال كل دورة. وبعد اختبار نموذج الشبكة المعتبر على مجموعة بيانات التقييم validation dataset للتمكّن من تقييم قدرته على التعميم على بيانات جديدة، نظرا لاحتواء مجموعة البيانات هذه على بيانات لم تتعرف عليها هذه الشبكة من قبل. حقق هذا النموذج دقة تعرف وصلت بداية إلى 81.46% مع القيمة 0.7199 لتابع الخسارة. وكانت أفضل دقة تصنيف تم التوصل إليها 98.87% مع القيمة 0.1146 لتابع الخسارة loss function.

بإلقاء نظرة فاحصة على الفرق بين دقة التصنيف و قيم تابع الخسارة التي حصلنا عليها من قواعد بيانات التدريب والتقييم قبل وبعد إجراء تكبير بيانات الصورة(زيادة الأبعاد)، سنجد أن دقة بيانات التدريب قبل التكبير حافظت على

التحسين فكان النسبة 95.87% بينما حافظت على قيمة تابع الخسارة كأقل من 0.1166، دقة بيانات التقييم كانت أسوأ بكثير إذ انخفضت إلى 81.46% بينما وصل تابع الخسارة إلى القيمة 0.7099 .

في الطريقة الأخرى بعد تحقيق تكبير في البيانات، فإن دقة بيانات التدريب وصلت إلى 75.13% مع الحفاظ على دقة تابع الضياع 0.7016 بينما تصل دقة بيانات التقييم إلى 77.56% بينما انخفضت قيمة تابع الضياع إلأى 0.6297.

على الرغم من الدقة التي تم الحصول عليها من قاعدة البيانات الأصلية والتي تعتبر أكبر من قاعدة البيانات المعززة، إلاّ أن الأخيرة تمتلك إمكانات أكبر لتحقيق دقة تصنيف أعلى بعدد أقل من دورات التدريب.

يبين الجدول التالي أداء التصنيف المنجز باتباع بعض الخوارزميات على صور قاعدة البيانات المستخدمة CIFAR−10.

الجدول (2) أداء التصنيف المنجز على صور قاعدة البيانات CIFAR-10 باستخدام أحدث الخوارزميات.

الخوارزمية	دقة التصنيف
Fractional Max-Pooling	%96.53
Striving for Simplicity: The All Convolutional Net	%95.59
All you need is a good init	%94.16
Fast and Accurate Deep Network Learning by Exponential Linear Units	%93.45
Training Very Deep Networks	%92.40

الخلاصة:

يعتبر التّعرف على الكائنات من المهام الصعبة والمهمة، ومن المشاكل المطروحة نظرا لتعقيد أصناف الكائنات من جهة وقلة الموارد الحسابية المطلوبة من جهة أخرى. ساعد استخدام مفهوم التعلم العميق Deep learning على شرح مدى الدقة في مجال التعرف على الكائنات بالاعتماد على نماذج الشبكات العصبونية الالتفافية لدرجة أنه يمكن أن نقترب من التعرف ضمن الزمن الحقيقي إضافة إلى إمكانية انتشار هذه النماذج تجاريا باستخدام الحد الأدنى من الأجهزة.

أعطت النتائج تفوق النموذج المقترح من خلال قدرته على استخدام الحد الأدنى من الذاكرة 60%، وخفض التعقيد الحسابي من خلال استخدام وحدات المعالجة الرسومية Graphical Processing Units(GPUs)، وأداء التعرف مقارنة مع نماذج الشبكات العصبونية الالتفافية الموجودة في أجهزة الاستخدام النهائي. حقق النموذج المقترح معدل خطأ وصل إلى 18.54% والناتج من العلاقة:

$$\text{error rate\%} = \frac{|\text{approximations value} - \text{exact value}|}{exact\ value} * 100 = \frac{|81.46-100|}{100} * 100 = 18.54\%$$

الاستنتاجات والتوصيات:

تم في البحث الحالي تطوير نظام جديد للتعرف على الكائنات باستخدام الشبكات العصبونية الالتفافية، وذلك بتحقيق تسلسل مناسب للطبقات واختيار مناسب لحجوم الفلاتر والأوزان. تم التوصل لمعدل تصنيف مناسب لهذه المهمة. يوصي البحث الحالي باستخدام هذا النظام في مجال رعاية الأشخاص المكفوفين من خلال ابتكار أجهزة يحملونها للتّعرف على الكائنات الملتقطة من كاميرا يحملونها في أيديهم.

المراجع:

[1] Roy, S. S., Ahmed, M., & Akhand, M. A. H.," Noisy image classification using hybrid deep learning methods", Journal of ICT, 18, No. 2 (April) 2018, pp: 233–269.
[2] Malik, M., Adavanne, S., Drossos, K., Virtanen, T., Ticha, D., Jarina R., "Stacked Convolutional and Recurrent Neural Networks for music emotion recognition", Department of Multimedia and Information-Communication Technologies, University of Zilina, Slovakia, asXiv: 1706.02292v1 [cs.SD], p:2-4.
[3]Bender, A., Porsteinsson, E. E., "Object Classification using 3D Convolutional Neural Networks", Chalmers University of technology, Master's thesis in Systems, Control and Mechatronics, Sweden 2016, p: 5-8.
[4] Liang, M., Hu, X., "Recurrent Convolutional Neural Network for Object Recognition", Tsinghua National Laboratory for Information Science and Technology (TNList) Department of Computer Science and Technology Center for Brain-Inspired Computing Research (CBICR) Tsinghua University, Beijing 100084, China, 2015, p: 3376-3378.
[5]Schneider, S., Taylor, G. W., Kremer, S. C.," Deep Learning Object Detection Methods for Ecological Camera Trap Data", arXiv:1803.10842v1 [cs.CV] 28 Mar 2018, P:1-4.
[6]Driss, S. B., Soua, M., Kachouri, R., Akil, M., "A comparison study between MLP and Convolutional Neural Network models for character recognition", ESIEE Paris, IGM, A3SI, 2 Bd Blaise Pascal, BP 99, 93162 Noisy-Le-Grand, France Submitted on 21 May 2017, P: 301- 303.
[7]Ahmad, S., "Visual Object Recognition Using Deep Convolutional Neural Network", Bachelor of science in computer science Brack university, Dhaka springm2017.
[8]Alom, M.Z., Taha, T.M., Yakopcic, C., "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches", 2018, p:8-10.
[9] Gauthier, I., & Tarr, M. J. (2016). Visual Object Recognition: Do We (Finally) Know More Now Than We Did?. Annual Review of Vision Science, 2, 377-396.
[10] Tobias, L., Ducournau, A., Rosseau, F., Fablet, R., & Mercier, G. (2016). Convolutional Neural Networks for Object Recognition on Mobile Devices: a Case Study. International Conference on Pattern Recognition (pp. 2-7). Cancun: ResearchGate.
[11] Mohamed Mhiri, Christian Desrosiers, Mohamed Cheriet,",Convolutional pyramid of bidirectional character sequences for the recognition of handwritten words, Pattern Recognition Letters, Volume 111, 1 August 2018, Pages 87-93.
[12] Amit Dhomne, Ranjit Kumar, Vijay Bhan, "Gender Recognition Through Face Using Deep Learning", Procedia Computer Science, Volume 132, 2018, Pages 2-10.
[13] Yanfen Chang, "Research on De-motion Blur Image Processing Based on Deep Learning", Journal of Visual Communication and Image Representation, In press, accepted manuscript, Available online 25 February 2019.
[14] AnuPriya George, X.Felix Joseph, "Object Recognition Algorithms for Computer Vision System: A Survey", International Journal of Pure and Applied Mathematicsm Volume 117 No. 21 2017, 69-74.