Journal of Education and Science (ISSN 1812-125X), Vol: 31, No: 03, 2022 (27-41)

Implementation of OCR using Convolutional Neural Network (CNN): A Survey

Ahmed A. Idris[1*], Dujan B. Taha[2]

[1*,2]Softwares Department, College of Computer Sciences & Mathematics, University of Mosul, Mosul, Iraq
E-mail: [1*]ahmed.20csp4@student.uomosul.edu.iq, [2]dujan_taha@uomosul.edu.iq

(Received May 02, 2022; Accepted June 28, 2022; Available online September 01, 2022)

DOI: 10.33899/edusj.2022.133711.1236, © 2022, College of Education for Pure Science, University of Mosul.
This is an open access article under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/)

Abstract

Recently, character recognition and deep learning have caught the attention of many researchers. Optical Character Recognition (OCR) usually takes an image of the character as input and generates the identical character as output. The important role that OCR does is to transform printed materials into digital text files. Convolutional Neural Network (CNN) is an influential model that is generous with bright results in optical character recognition (OCR). The state-of-the-art performance which exists in deep neural networks is usually used to handle frequently recognition and classification problems. Many applications are using it, for instance, robotics, traffic monitoring, articles digitization, etc. CNN is designed to adaptively and automatically learn features by using many kinds of layers (convolution layers, pooling layers, and fully connected layers). In this paper we will go through the advantages and recent usage of CNN in OCR and why it's important to use it in handwritten and printed text recognition and what subjects we can use this technique for. Researchers are progressively using CNN for the machine-printed characters and recognition of handwritten, that is because CNN architectures are suitable for recognition tasks by inputting some images.

Keywords: Convolutional Neural Network (CNN), Optical Character Recognition (OCR), Deep learning, Pre-Processing, Feature Extraction, and Classification.

تنفيذ التعرف الضوئي على الحروف باستخدام الشبكة العصبية التلافيفية (CNN): دراسة

احمد عبد الرحمن ادريس[1*]، دجان بشير طه[2]

[1*،2] قسم البرامجيات، كلية علوم الحاسوب والرياضيات، جامعة الموصل، الموصل، العراق

الملخص

في الآونة الأخيرة، جذب التعرف الضوئي على الاحرف والتعلم العميق انتباه العديد من الباحثين، إذ ان المنافسة قائمة للحصول على اعلى دقة، ناهيك عن التحديات الحاصلة في بعض اللغات لتقارب وتشابه الاحرف فيما بينها. عادةً ما يأخذ التعرف الضوئي على الأحرف (OCR) صورة الاحرف كمدخلات وبالاعتماد على هذه الادخالات يقوم بتوليد الحروف المتطابق كإخراج. الدور المهم الذي يقوم به OCR هو تحويل الأنماط المطبوعة إلى ملفات نصية رقمية. عادةً ما يتم استخدام أحدث الطرق الموجودة في الشبكات العصبية العميقة للتعامل مع مشاكل التعرف على الأنماط والتصنيف، وتعد الشبكة العصبية التلافيفية (CNN) نموذجًا مؤثرًا سخيًا بنتائج ساطعة في التعرف الضوئي على الأحرف (OCR). ومن التطبيقات التي تستخدمه في مجالاتها، على سبيل المثال، الروبوتات، ومراقبة حركة المرور، ورقمنه المقالات، وما إلى ذلك. تم تصميم CNN لتتعلم الميزات بشكل تكيفي وتلقائي باستخدام أنواع عديدة من الطبقات. في هذه الدراسة، نتعرف على مزايا الـ CNN واستخداماته الحديثة في التعرف الضوئي على الحروف ولماذا يكون من المهم استخدامها في هذا المجال.

الكلمات المفتاحية: الشبكة العصبية التلافيفية (CNN)، التعرف الضوئي على الحروف (OCR)، التعلم العميق، المعالجة المسبقة، استخلاص الميزات، التصنيف.

1. مقدمة

إن قارئ الأحرف البصري أو التعرف الضوئي على الحروف هو التحويل الميكانيكي أو الإلكتروني للصور المكتوبة أو النص المطبوع أو المكتوبة بخط اليد إلى نص مشفر آليًا (آلة ترميز النصوص). التغيير الحاصل في الشبكة العصبية متعدد الطبقات (MLP) هو الشبكة العصبية التلافيفية (CNN) يكون من كانت المرة الأولى المستخدمة في عام 1980 [1]. مصدر إلهام الشبكة العصبية التلافيفية هو الدماغ البشري حيث يدرك البشر الأشياء بصريًا وهكذا هو عمل الCNN. كما ولدينا مصطلح التصنيف (Classification) وهو عملية لتعلم الآلة على بيانات إدخال معينة وتعيينه أو تسميته في فئات محددة مسبقًا (Labels). تعد الشبكات العصبية التلافيفية فعالة للغاية في إدراك بنية الأحرف / الكلمات المكتوبة بخط اليد بطرق تساعد في استخراج للسمات المميزة وتجعل CNN الطريقة الأنسب لحل مشاكل التعرف على خط اليد [2]. في المراجعة المنهجية [3] تم البحث في جميع المقالات المتعلقة بالموضوع، حيث تم اختيار 176 مقالة، ونوقشت تقنيات التصنيف الأكثر شيوعًا في الدراسات البحثية للتعرف الضوئي على الحروف بدءًا من عام 2000 حتى عام 2019. في الآونة الأخيرة، حققت CNN نجاحًا كبيرًا في التعرف الضوئي على الحروف [4] إذ أنشئت تم إنشاء أسس الشبكات العصبية العميقة (طبقات عديدة)، مثل الشبكات العصبية التلافيفية والشبكات العصبية المتكررة، واختيرت الشبكات العصبية كواحدة من أكثر تقنيات التصنيف المعقولة للتمييز (Recognition) بما في ذلك التعرف الضوئي على الحروف (OCR) [5] [6]. التغيير المبتكر في ANN هي الشبكات العصبية التلافيفية (CNN)، حيث أحدثت CNN طفرة في التعرف على الأنماط ورؤية الكمبيوتر (Computer vision). اول من قدمها كان Yann LeCun عام 1998، وبعدها تحسنت الدقة لـ CNN بنسبة تزيد عن 15٪ مقارنة بالفائز في عام 2011 وعندها بدأت بالمنافسة عام 2012، ومنذ ذلك الحين، يستخدم CNNفي العديد من مشكلات التعرف على الأنماط، لا سيما في OCR. في كثير من الحالات والاستخدامات، أظهرت CNN دقة أفضل من حيث الأداء في التعرف والتمييز، إذ ظهرت بشكل خاص في التعرف على الأنماط المعقدة داخل الصور، وكذلك يميزها تصميم المعمارية والخطوات التي تمد بها الادخالات اثناء مرحلة التدريب [7]. بالإضافة الى استخداماتها في التصنيف، يمكن استخدام CNN في استخلاص الميزات (Feature extraction)، حيث تم تنفيذ كلتا المهمتين في هذا العمل [8] إذ طبق استخلاص الميزات والتصنيف في CNN. تحتوي CNN على طبقات مختلفة مثل الطبقة التلافيفية وطبقة التجميع وطبقة التنشيط وطبقة الارتباط الكامل وطبقة التصنيف [8] وسيتم مناقشة كل طبقة من هذه الطبقات على حدى في هذه الدراسة، بالإضافة الى المرشحات (filters) التي تستخدم لكل صورة. وهذا ما يجعل CNN مفهومًا ومنهجياً ومن أفضل الحلول لمهام التعرف على الصور والأنماط المعقدة، ويقدم أفضل فعالية وكفاءة في الاستخدام [9]، [10]، [11]، [12]. اما فيما يخص ANN، حيث تواجه ANN الكلاسيكية قيودًا شديدة، فإن التعقيد الزمني للحساب في التعرف على الأنماط لمعالجة الصور كبير جدًا بحيث لا يمكن تحمله [7]. استخدمت شبكة CNN أحادية البعد (one dimension) للتعرف على الحروف الصينية المكتوبة بخط اليد على الإنترنت [13]. يبدو أن CNN أحادي البعد قد حقق أداءً أعلى إذ إن دقة التعرف [13] هي 98.1٪ مقارنة بـ [14] حيث كانت الدقة الخاصة بها 83٪.

الهدف الرئيسي من هذه الدراسة:

١) تلخيص العمل البحثي الحالي حول الشبكات العصبية التلافيفية (CNN) في التعرف الضوئي على الحروف (OCR) عن طريق جمع البحوث ذات الصلة.

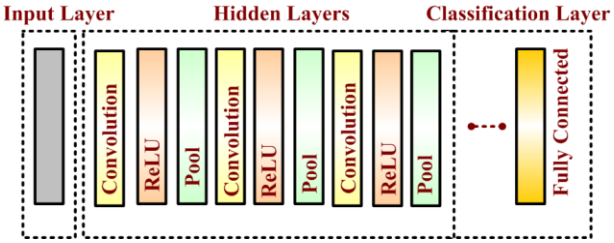
٢) تحديد بعض المجالات الحديثة التي تم استخدام التعرف الضوئي على الحروف فيها.

بنية الشبكة العصبية التلافيفية CNN

تعتمد بنية CNN بشكل أساسي على العديد من المراحل حيث تعتمد كل مرحلة بشكل أساسي على العديد من الطبقات. الطبقة التلافيفية (convolutional layer) وطبقة التجميع (pooling layer) وتعتبر اختيارية في بعض الأحيان، وطبقة الإخراج (output layer). تعتبر بنية الشبكة العصبية التلافيفية ذات الثلاث طبقات جيدًا لتصنيف الصور المكتوبة بخط اليد كما هو موضح في الشكل 1. وتتكون من طبقة الإدخال وطبقات متعددة مخفية (تكرار تلافيفي، وتطبيع، وتجميع)، وطبقة الارتباط الكامل (fully connected) وطبقة الإخراج (output layer) [2].

1.2 طبقة الإدخال (Input layer)

في هذه الطبقة يتم تحميل البيانات وتخزينها. تتكون هذه الطبقة من الارتفاع والعرض وعدة قنوات (معلومات RGB) لصورة الإدخال [2]. كما هو مبين في الشكل 1.



الشكل 1. بنية الشبكة العصبية التلافيفية (CNN)

2.2 الطبقة المخفية (Hidden Layer)

الطبقات المخفية هي العمود الفقري لمعمارية الـ CNN إذ نفذت عملية استخراج واستخلاص الميزات ويتم استخدام سلسلة من وظائف الالتفاف(convolution) والتجميع (pooling) والتنشيط (activation functions)، حيث يتم الكشف عن السمات المميزة للأنماط المكتوبة بخط اليد في هذه المرحلة [2].

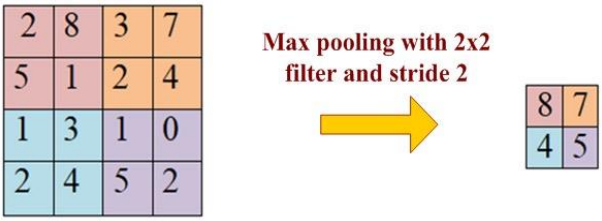
3.2 الطبقة التلافيفية (Convolutional Layer)

الطبقة الأولى الموضوعة فوق صورة الإدخال هي الطبقة التلافيفية. حيث يتم استخدامها لاستخلاص ميزات الصورة المدخلة. المساهمون الرئيسيون في الطبقة التلافيفية هم حقل الاستقبال (receptive field) والخطوة (stride) والمرشح (filter) وفي المقابل يتم انتاج المخرجات (outputs) [2]. receptive field هو مصطلح مستخدم في حساب المناطق الفرعية الفعالة في الشبكة. Stride هي معلمة أخرى مستخدمة في بنية CNN. يتم تعريفه على أنه حجم الخطوة التي يتحرك بها المرشح في كل مرة. تشير قيمة الخطوة 1 إلى ازاحة المرشح بكسلًا واحداً في كل قراءة.

2.4 طبقة التجميع (Pooling Layer)

تتم إضافة طبقة التجميع بين طبقتين تلافيفيتين لتقليل أبعاد الإدخال وبالتالي تقليل العمليات الحسابية في كل مرحلة متقدمة بين طبقات الCNN. تسمح هذه الطبقة بتمرير القيم المحددة إلى الطبقة التالية مع إهمال القيم غير المرغوب فيها كما في الشكل 2. تساعد طبقة

التجميع أيضًا في الطبقة التالية بنفس العمل وذلك بإهمال القيم غير المرغوب فيها. حيث تساعد طبقة التجميع أيضًا في تحديد المعالم [2].



الشكل 2. مرشح (filter) 2 × 2

5.2 طبقة التنشيط (Activation Layer)

تتضمن بنية CNN أيضًا وظيفة التنشيط لتقديم وتفعيل اللاخطية في النظام. ReLu و Softmax بعض الخيارات البارزة بين وظائف التنشيط المختلفة المستغلة على نطاق واسع في نماذج التعلم العميق. ولقد لوحظ أن sigmoid قد تقوم بإضعاف نموذج CNN بسبب فقدان المعلومات في بيانات الإدخال اما وظيفة التنشيط المستخدمة في العمل الحالي هي وظيفة ReLu والتي تقوم بتعيين وإعطاء قيمة 0 للإدخال أقل من 0. تتمثل بعض مزايا وظيفة التنشيط ReLu في تشابهها مع نظام العصب البشري، والوضوح في الاستخدام، والقدرة على تحقيق تدريب أسرع لشبكات أكبر [2].

6.2 طبقة التصنيف (Classification Layer)

طبقة التصنيف هي الطبقة النهائية في بنية CNN. إنها طبقة الارتباط الكامل (Fully connected)، واعتمدت بشكل أساسي كمصنف للبيانات. إذ ترتبط الخلايا العصبية الموجودة في طبقة الارتباط الكامل بجميع الخلايا العصبية في الطبقة السابقة. تحسب هذه الطبقة الفئات المتوقعة من خلال التعرف على صورة الإدخال، ويتم ذلك من خلال دمج جميع الميزات التي تم الحصول عليها من الطبقات السابقة [2].

7.2 خوارزمية تحسين (Optimization Algorithm)

تُستخدم خوارزميات التحسين لتطوير الشبكات العصبية ولإحداث أداء أفضل ونتائج أسرع. تدعم الخوارزمية تقليل أو تعظيم دالة التكلفة عن طريق تحديث قيم الوزن / التحيز (weight/bias)، والتي تُعرف باسم معلمات التعلم للشبكة (parameters)، ويطلق على الخوارزمية التي تقوم بتحديث هذه القيم اسم خوارزمية التعلم التكيفي (algorithm adaptive learning). تؤثر معلمات التعلم (parameters) هذه بشكل مباشر على عملية التعلم للشبكة ولها دور مهم في إنتاج نموذج فعال للشبكة. تهدف جميع خوارزميات التحسين إلى إيجاد القيم المثلى لمعلمات التعلم هذه.

8.2 الإعداد التجريبي

لإنجاز مهمة التعرف على الأرقام المكتوبة بخط اليد، تم تطوير نموذج للشبكة العصبية التلافيفية وتحليلها لمعلمات التعلم المختلفة المناسبة لتحسين دقة التعرف ووقت المعالجة.

تتكون عملية التعرف على الأرقام المكتوبة بخط اليد من الخطوات التالية:

1. الحصول على البيانات أو مجموعة من الصور الرقمية المكتوبة بخط اليد (مثل MNIST).

2. تقسيم الصور المدخلة إلى صور تدريب واختبار.
3. عمل معالجة مسبقة في كل من مجموعة بيانات التدريب ومجموعة بيانات الاختبار.
4. تطبيع البيانات بحيث يتراوح من 0 إلى 1.
5. تقسيم مجموعة بيانات التدريب إلى مجموعات ذات حجم مناسب.
6. تدريب نموذج CNN ومتغيراته باستخدام البيانات المصنفة.
7. تشغيل نموذج مدرب للتصنيف.
8. تحليل دقة التعرف ووقت المعالجة لجميع المتغيرات [2].

3. اتجاهات البحث

في هذا الجزء من البحث تم تحليلCNN حسب اتجاهات البحوث في التعرف البصري على الأحرف. بدءًا من المفاهيم والدوافع الأولية لـ CNN، ونوقشت المزايا الرئيسية لـ CNN مقارنة بـ ANN الكلاسيكية [7]. أحد التغييرات الحديثة في ANN هو ظهور CNN، حيث تم استخدام تقنيات اكتساب المعرفة بشكل كبير في مجالات متنوعة وواسعة [15]. يتم تطبيق الشبكات العصبية التلافيفية (CNN) للحصول على أكثر النتائج دقة في حل مشاكل العالم الحقيقي، وخاصة في التعرف على الصور [16]. حيث تم عمل جهاز يمكن ارتداؤه بالأصابع لقراءة النص الصيني المطبوع باستخدام خوارزمية CNN [15] (سيتم الحديث عنه بشكل موسع لاحقاً). تم تدريب واختبار CNNs (VGG و ResNet و CapsNet) على مجموعة بيانات من الأحرف الصينية الكبيرة والحصول على دقة ووقت الاختبار للمقارنة. قمنا أيضًا بتشذيب شبكة CNN عن طريق حذف الخلايا العصبية المعطلة والطبقات الكثيفة. ونتيجة لذلك ، انخفض وزن الشبكة بنسبة 96.5٪ بدقة 97.70٪. ومن الجدير بالذكر أن شبكة ResNet و CapsNet حققتا دقة بلغت 99.17٪ و 99.38٪ على مجموعة بيانات الاختبار. كان وقت الاختبار هو متوسط الوقت الذي يتم فيه تشغيل كل شبكة 20 مرة على CPU وGPU. يمكن أن يؤدي تطبيق التعلم العميق على التعرف الضوئي على الحروف إلى تحقيق دقة عالية ووقت معالجة منخفض، كما تؤدي الى تقليل وقت اختبار النظام بشكل أكبر [17]. تم ايضاً تقديم نظام التعرف على الحروف لنص اللغة الإنجليزية على جريد النخيل (سعف النخيل)، حيث قام هذا النظام بالتعرف على هذه الكلمات بواسطة نظام التعرف الضوئي على الحروف باستخدام CNN بمعدل تعرف عالي [18]. تم تنفيذ خوارزمية CNN من خلال تطبيق TensorFlow من Google، وهي مكتبة مفتوحة المصدر تم توفيرها من خلال Google brain ويتم أخذها في الاعتبار لتكون مناسبة لأغراض البحث [19]. ناقش نموذجًا هجينًا باستخدام مصنفين (نموذج متكامل) وهما CNNوXGBoost، إذ تعمل CNN كمستخرج للميزات قابل للتدريب باستخدام الصور الخام، وXGBoost يأخذ الميزات المستخرجة كمدخلات للتعرف والتصنيف. في هذا النموذج المتكامل ، تعمل CNN كمستخرج تلقائي للميزات يمكن تدريبه من الصور الأولية ، بينما يقوم XGBoost بجزء التعرف. يتم تقييم كفاءة النموذج المقترح واستمراريته في جانبين: دقة التعرف وصلابة الاداء. وتظهر النتائج التجريبية في مجموعة بيانات HECR تحسينات كبيرة حققها النموذج المقترح [20].

يمكن لإصدار CNN أثناء تدريبه بشكل صحيح في مرحلة ادخال مجموعة بيانات (Dataset)، اكتشاف ميزات الصورة بشكل صحيح دون عناء، وبالتالي يؤدي أداءً جيدًا أعلى من MLP في مجالات مشكلات التعرف على الصور التقليدية [21]. في هذا البحث [21]، تم اقتراح خوارزمية جديدة تعتمد على الشبكات العصبية للتعلم العميق باستخدام وظيفة التنشيط المناسبة (activation function) وطبقة التنظيم (regularization layer)، والتي تظهر دقة محسنة بشكل ملحوظ مقارنة بأساليب التعرف على الأرقام العربية الحالية. يوفر النموذج المقترح دقة تبلغ 97.4 بالمائة ، وهي أعلى دقة مسجلة لمجموعة البيانات المستخدمة. حيث تم اقتراح خوارزمية جديدة للتعرف

على الأرقام المكتوبة بخط اليد العربية بمساعدة CNN. يستخدم نموذج CNN المقترح لدينا عدة طبقات تلافيفية جنبًا إلى جنب مع تنشيط ReLU. هذه الطريقة الجديدة المقترحة تحقق دقة عالية جداً. يلخص الجدول 1 اتجاهات البحث مع الاستخدام الكبير لـ CNN في أنظمة مختلفة.

جدول (1): اتجاهات بحث CNN

اتجاهات الأبحاث	الاستخدام	المصادر
التغيير الثوري في ANN هو شبكة الالتفاف العصبية CNN	مزايا CNN مقارنة بالـ ANN التقليدية.	[7]
نتائج دقيقة في حل مشاكل العالم الحقيقي، لا سيما في التعرف على الصور	قارئ النص التركي.	[16]
جهاز يمكن ارتداؤه بالأصابع لاستكشاف النص الصيني المطبوع باستخدام خوارزمية CNN	جهاز محمول للتعرف على اللغة الصينية والإنجليزية وأرقام.	[15]
يمكن أن تؤدي شبكة التعلم العميق إلى تقليل وقت اختبار النظام وحجم الشبكة بشكل كبير	تدريب واختبار CNN لمجموعة بيانات الأحرف الصينية الكبيرة.	[17]
التعرف على النص باللغة الإنجليزية حيث يتم كتابته على سعف النخيل	التعرف على النص من المخطوطة القديمة لغة إنجليزية	[18]
تم تنفيذ خوارزمية CNN من خلال تطبيق TensorFlow من Google وهو برنامج مفتوح المصدر مقدم من Google Brain	مع TensorFlow	[19]
نموذج هجين يستخدم مصنفين (نموذج متكامل): الأول هو الشبكة العصبية التلافيفية (CNN) والثاني تعزيز التدرج الأقصى (XGBoost)	ميزة مستخرج من الصور (خام)	[20]
تستخدم CNN الالتواء في طبقتها لاكتشاف ميزات الصور بطرق أكثر قوة	تستخدم CNN للتعرف على الأرقام المكتوبة بخط اليد العربية	[21]

4. بعض نطاقات التعرف الضوئي على الحروف

1.4 اللغات

استخدمت الشبكة العصبية التلافيفية على نطاق واسع لتصنيف جميع اللغات التي تمت مراجعتها والتعرف عليها تقريبًا [3]. وفيما يلي بعض اللغات التي استخدمت CNN في تدريبها:

1.1.4 المخطوطات الديفاناغارية القديمة (Devanagari ancient manuscripts)

لم يتم استخدام CNN قط للتعرف على المخطوطات الديفاناغارية القديمة. إذ تم في هذا البحث استخدام قوة CNN لاستخراج ومعرفة النتائج من المخطوطات القديمة المكتوبة بخط اليد الديفاناغارية [8] حيث تم تحقيق أقصى قدر من الدقة بنسبة 93.73 ٪ مع 75 ٪ من بيانات المدربة و25 ٪ من البيانات للاختبار بعد 30 حقبة (epochs). وايضاً في هذا العمل تم استخدم CNN كمستخرج للميزات بالإضافة إلى مصنف ووجد أنه أفضل من تقنيات استخراج السمات الموجودة وتقنيات التصنيف للتعرف على المخطوطات الديفاناغارية القديمة [8]. النموذج القائم على CNN تم تطبيقه أيضًا على الوثائق القديمة المكتوبة بالخط الديفاناغاري أو السنسكريتي في دراسة أخرى والنتائج عند مقارنتها مع رؤية Google أعطى OCR دقة 93.32٪ [3].

2.1.4 البنغالية (Bengali)

التحدي الواضح لمجموعة الأحرف البنغالية هو التقارب في أشكالها. اتضح أنه يمكن استخراج المزيد من الميزات عن طريق زيادة سعة الشبكات [22]. في تجربة مجموعة بيانات متعددة (Datasets)، طبق الباحثون نموذجًا خفيف الوزن يعتمد على 13 طبقة من CNN مع طبقتين فرعيتين على أربع مجموعات بيانات للغة البنغالية. تم تحقيق دقة تبلغ 98٪ و96.81٪ و95.71٪ و96.40٪ حيث تم تطبيق النموذج على مجموعة بيانات CMATERdb و ISI و BanglaLekha المعزولة ومجموعات البيانات المختلطة على التوالي [3].

3.1.4 التيلجو (Telugu)

لغة التيلجو هي لغة درافيدية يتحدث بها سكان التيلجو، وهم يعيشون في ولايتي تيلانجانا وأندرا براديش الهندية، حيث تعتبر لغة رسمية. تتم مناقشة العديد من تقنيات CNN ويتم استخدام بنياتها لتصنيف التعرف على الأحرف المكتوبة بخط اليد. في الطريقة المقترحة، تمت مناقشة معمارية VGG-16 وتدريبها باستخدام مجموعة بيانات الأحرف التيلجو للتعرف بخط اليد وكان هناك دقة في نتائج التدريب والاختبار لمجموعات البيانات [23].

4.1.4 العربية (Arabic)

تعتبر اللغة العربية من اللغات التي فيها تحد كبير لتقنيات الCNN وذلك لاختلاف رسم اشكال الحروف. نظام التعلم العميق القائم على شبكة عصبية تلافيفية قادر على تصنيف الحروف العربية المكتوبة بخط اليد بدقة تصل إلى 94.8٪ و97.6٪ على مجموعتي بيانات (Datasets) AIA9k و AHDC على التوالي [24] وتم الحصول على هذه الدقة العالية عن طريق استخدام معمارية الCNN. تم تدريب واختبار CNN باستخدام 168000 حرف مكتوب بخط اليد باللغة العربية وبعدها قسمت مجموعة البيانات إلى مجموعتين: (1) مجموعة تدريب مكونة من 13440 حرفًا لـ 48 صورة لكل حرف و (2) مجموعة اختبار مكونة من 3360 حرفًا إلى 120 صورة لكل حرف في اختبار تجريبي، أظهرت النتائج تفوق في الأداء على معدل دقة التصنيف بنسبة 100٪ في اختبار الصور [25]. في دراسة أجريت في

عام 2016 بمساعدة Elleuch et al [26] اكتشفت شبكة عصبية تلافيفية (CNN) تعتمد أساسًا على آلة ناقلات الدعم (SVM) للتعرف على اللغة العربية المكتوبة بخط اليد دون اتصال بالإنترنت، حيث يستخرج النموذج تلقائيًا الميزات من المدخلات الأولية ويحقق التصنيف. في عام 2018، نفذ الباحثون نهج DCNN للتعرف على الأحرف العربية المكتوبة بخط اليد وغير المتصلة بالإنترنت، وتم الحصول على دقة قدرها 98.86٪ [27].

5.1.4 الإنجليزية (English)

في الخمسينيات من القرن الماضي، تم اقتراح العديد من الاستراتيجيات للتعرف على الأحرف المكتوبة بخط اليد باللغة الإنجليزية. ومع ذلك، لا يوجد نظام التعرف الضوئي على الحروف (OCR) ليتعرف على الحروف المكتوبة بخط اليد [28]. أصبحت عندها محط اهتمام الكثير من الباحثين ولكن يبقى من الصعب إدراك تلك المواقف الصعبة من حيث إدراك أنماط الكتابة البشرية المختلفة، والأشكال المميزة، والزوايا المحددة من الحروف، وأحجام الحروف، ومعدل الدقة المنخفض في التعرف، وجودة المدخلات الطبيعية، وما إلى ذلك. نتيجة لذلك، يجب إجراء العديد من الدراسات لحل هذه المشاكل. تم تطوير قاعدة البيانات هنا بواسطة [29] وهي تتكون من نصوص إنجليزية ويونانية، وحقق النظام دقة التعرف على مجموعة البيانات هذه بنسبة 98.186٪.

في الجدول 2، نجمع اللغات التي ذكرت أعلاه بأعلى دقة توصلت لها.

جدول (2): أعلى دقة لبعض اللغات المستخدمة في الـCNN

الدقة (٪)	اللغة	المصدر
93.32	الديفاناغارية القديمة	[3]
98	البنغالية	[3]
92	التيلجو	[23]
97.6	العربية	[24]
98.186	الإنجليزية	[29]

علاوة على ذلك، هناك لغات أخرى كثيرة استخدمت CNN، لكننا ذكرنا هذه اللغات لأن هذه اللغات واجهت تحديًا في استخداماتها لـ CNN. وأيضاً تحتوي بعض هذه اللغات على بيانات قليل في مجموعات البيانات (Dataset) الخاصة بها لتدريب النموذج، وعلى الجانب الآخر، يتم استخدام بعض هذه اللغات بشكل متكرر في الأنظمة الحالية، لذلك نحتاج إلى الحفاظ على دقتها العالية قدر الإمكان. استخدم الباحثون على نطاق واسع شبكة عصبية تلافيفية (CNN) جنبًا إلى جنب مع تقنيات مختلفة للتعرف على الحروف. تم تنفيذ هذه التقنيات على مجموعات بيانات متميزة للتحقق من دقة التقنيات [3]. في تجربة مجموعة بيانات متعددة، طبق الباحثون نموذجًا خفيف الحجم يعتمد

على 13 طبقة من CNN مع طبقتين فرعيتين على أربع مجموعات بيانات للغة البنغالية. أربع شبكات CNN لها بنى مشابهة مثل AlexNet [30] وVGG-16 [31] وVGG-19 [31] وMobileNet2 حيث تم إنشاء مجموعة البيانات واختبارها بواسطة CNN [32].

2.4 معماريات (CNN architectures)

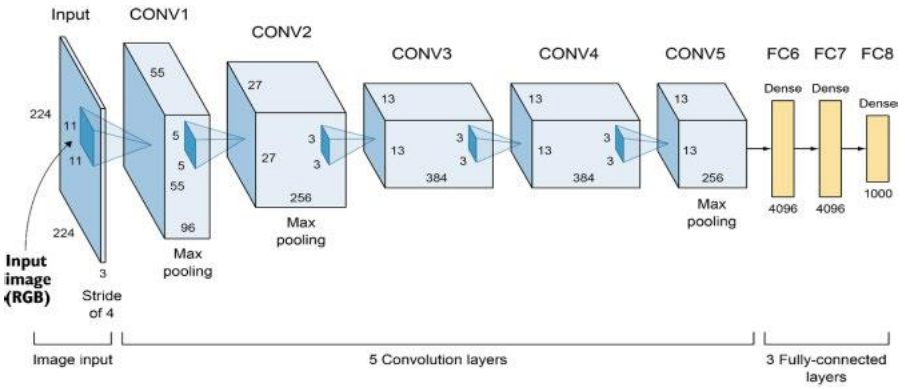
تم تطبيق العديد من التغييرات على هندسة CNN من عام 1989 حتى اليوم. يلخص الجدول 3 معماريات CNN [33]. في هذا الجزء سنتطرق لواحده من المعماريات التي نفذت التعرف الضوئي على الحروف. معمارية AlexNet تعد من احده الأمثلة الشائعة لهندسة CNN الشكل 3. تم اقتراح AlexNet في البداية بواسطة Krizhevsky et al [30]، حيث تتكون من 8 طبقات؛ الطبقات الخمس الأولى هي طبقات تلافيفية، وقد لوحظ عدد منها عن طريق طبقات التجميع، والطبقات الثلاث المتبقية تحتوي طبقات الارتباط الكامل. إنه يستخدم ReLU التي أظهرت أداءً شاملاً للتدريب المتقدم على tanh وsigmoid [34]. تم اختيار AlexNet [35] لاستخلاص المميزات من صور LP المقروءة وغير المقروءة، واظهرت كفاءة عالية في الاستخدام.

الجدول (3): نظرة عامة موجزة عن معماريات CNN [33]

النموذج Model	عمق Depth	مجموعة البيانات Dataset	معدل الخطأ Error rate	حجم الإدخال Input size	السنة Year
AlexNet	8	ImageNet	16.4	227 × 227 × 3	2012
NIN	3	CIFAR -10، CIFAR- 100، MNIST	10.41، 35.68، 0.45	32 × 32 × 3	2013
ZfNet	8	ImageNet	11.7	224 × 224 × 3	2014
VGG	16، 19	ImageNet	7.3	224 × 224 3	2014
GoogLeNet	22	ImageNet	6.7	224 × 224 × 3	2015

Inception-V3	48	ImageNet	3.5	229 × 229 × 3	2015
Highway	32 ،19	CIFAR - 10	7.76	32 × 32 × 3	2015
Inception-V4	70	ImageNet	3.08	229 × 229 × 3	2016
ResNet	152	ImageNet	3.57224	الزائد× 224 × 3	2016
Inception-ResNet- v2	164	ImageNet	3.52	229 × 229 × 3	2016
FractalNet	كتنظيم 40،80 CIFAR	،- 10 CIFAR- 100	4.6 18.85	32 × 32 × 3	2016
WideResNet	28	،CIFAR - 10 CIFAR- 100	3.89 18.85	32 × 32 × 3	2016
Xception	71	ImageNet	0.055	229 × 229 × 3	2017
Residual attention neural network	452	،CIFAR - 10 CIFAR- 100	20.4 ،3.90	40 × 40 × 3	2017
Squeeze-and-exci- tation networks	152	ImageNet	2.25	229 × 229 × 224 × × 320 × 3	2017

DenseNet	201	CIFAR - 10، CIFAR- 100، ImageNet	3.46، 17.18، 5.54	224 × 224 × 3	2017
Competitive squeeze and excitation net- work	152	CIFAR - 10 CIFAR -100	3.58 18.47	32 × 32 × 3	2018
MobileNet-v2	53	ImageNet	-	224 × 224 × 3	2018
CapsuleNet	3	MNIST	0.00855	28 × 28 × 1	2018
HRNetV2	-	ImageNet	5.4	224 × 224 × 3	2020



الشكل٣: معمارية AlexNet

فحص تأثير زيادة عدد الطبقات التلافيفية في بنية CNN على أداء التعرف على الأرقام المكتوبة بخط اليد من خلال التجارب حيث حققوا معدل التعرف بنسبة 99.89 ٪ مع مُحسِّن Adam لقاعدة بيانات MNIST وهو أفضل من جميع النتائج التي تم الاعلان عنها سابقًا [2]. يتم تغيير الطبقات داخل نموذج معمارية CNN والهندسة المعمارية والأوزان في نموذج ResNet50 المدرب مسبقًا، إذ تمت إزالة الطبقة العليا من طراز ResNet50 وإدراج طبقة جديدة في الشبكة، وتدريبها (ضبطها بدقة) على صور الاحرف البصرية. يُظهر هذا

النظام المقترح أداءً جيدًا على مجموعة بيانات حرف CHAR74K وقد حقق دقة تبلغ 87.56٪ ودرجة النظام الذي تم التعديل عليه تبلغ 88٪ [36]. وهذا يدل على انه يمكننا تطبيق بعض التعديلات على بنية ال CNN لجعلها أكثر صلابة وزيادة في الدقة.

3.4 الشبكة العصبية التلافيفية (CNN) مع LSTM

تصنيف الكلمات وتجزئة الأحرف، تُستخدم الشبكة العصبية التلافيفية (CNN) مع العديد من البنى لتدريب نموذج يمكنه تصنيف الكلمات بدقة. بعد ذلك، يتم استخدام شبكات الذاكرة طويلة المدى (LSTM) مع الالتفاف لإنشاء مربعات محيط لكل حرف. يتم تمرير الأحرف المجزأة إلى CNN لتصنيفها ثم إعادة بناء كل كلمة وفقًا لنتائج التصنيف والتجزئة [37]. ولتحديد الأحرف داخل الحروف المركبة ذات المقاييس المختلفة (أحجام الخطوط)، يتم استخدام شبكة الالتفاف العصبية (CNN) وشبكة الذاكرة طويلة المدى (LSTM). في هذا البحث [38]، تم استخدام مجموعتين، أي صور نصية مطبوعة باللغة الأردية (UPTI) وصور نصية من مركز هندسة اللغة (CLE). يتراوح الأداء العام للشبكات بين 90٪ و99.8٪. يبلغ متوسط الدقة في ميزات Meta 98.08٪ أثناء استخدام الصور الأولية، ويتم تحقيق متوسط الدقة 97.07٪.

5. أنظمة دراسات خوارزميةCNN

في موضوع التعرف على خط اليد تقنيات الدراسة العميقة وقد نفذت أداءً متقدمًا في السنوات القليلة الماضية. ومع ذلك، فإن التطور السريع في كمية البيانات المكتوبة بخط اليد وتوفير قوة معالجة كبيرة يتطلب تطويرًا في دقة التعرف ويكسب بالإضافة إلى البحث. تعد الشبكات العصبية التلافيفية (CNN) قوية جدًا في إدراك شكل الأحرف / الكلمات المكتوبة بخط اليد في الأساليب التي تساعد في الاستخراج التلقائي للسمات المميزة وتجعل من CNN الطريقة المناسبة القصوى لإصلاح مشكلات التعرف على خط اليد [2].

1.5 الكشف الرقمي باستخدام مستشعرات الراديو

قد أثبتت CNN دقتها العالية في كشف وتصنيف الارقام. حيث تم التحقق من دقة التعرف على الأرقام باستخدام البيانات الأولية كمدخلات لشبكة CNN. كانت الدقة التي تم الحصول عليها باستخدام البيانات الأولية عالية عندما كانت بيانات التدريب والاختبار لها نفس الاتجاه والمسافة؛ ومع ذلك، بعد تغيير الاتجاه أثناء التقييم، كانت الدقة الناتجة أقل بكثير. في المقابل، باستخدام الطريقة المقترحة في هذا البحث [39]، والتي تستخدم CNN لتحويل بيانات إلى صورة للتصنيف، وقمنا اختبار النظام حتى بعد كتابة الأرقام في الهواء في اتجاهات مختلفة، كانت الدقة الناتجة لا تزال عالية جدًا.

2.5 جهاز يمكن ارتداؤه بالأصابع لاستكشاف النص الصيني المطبوع

في الآونة الأخيرة، تم استخدام تقنيات التعلم العميق على نطاق واسع في مجموعة واسعة من المجالات. وفي هذه الدراسة [15]، تم تطوير قارئ Finger Reader الصيني للأشخاص ضعاف البصر للتعرف على النص الصيني المطبوع. الجهاز محمول وسهل التشغيل ومصمم ليتم ارتداؤه على السبابة. يعمل معالج Micro IoT للتعرف على الأحرف الصينية والحروف الإنجليزية والأرقام. حيث تم تصميم هذا الجهاز باستخدام تقنيات ال الشبكات العصبية التلافيفية (CNNs) للحصول على أكثر النتائج دقة في حل مشاكل العالم الحقيقي، وخاصة في التعرف على الصور.

الخلاصة

إن الشبكة العصبية التلافيفية (CNN) هي نموذج فعال يعطي نتائج عالية الدقة في مجال التعرف على الكلمات. في هذه الدراسة قمنا بجمع المصادر التي تساعد الباحثين للحصول على فهم كامل لشبكة CNN واهميتها في التعرف الضوئي على الحروف. علاوة على ذلك، يساعد هذا المسح في العثور على النتائج والدقة لبعض الأنظمة والبيئات التي تستخدم الCNN . وفي الوقت الحاضر، يتم استخدام منهجيات مختلفة على نطاق واسع للتعرف على الأحرف، وهناك أيضًا مجالات تطبيق لمعالجة المستندات الرقمية، مثل التحقق من الملفات وإدخال البيانات والمكتبات الرقمية وقراءة العناوين البريدية واستخراج البيانات من الشيكات والأوراق الضريبية وغيرها الكثير. هذا ما يجعل الباحثين يسلطون الضوء على هذا المجال بعمق ويختارون الحلول المثالية لتحقيق الهدف بأدق نسبة ممكنة. وفي هذا المسح قد تم جمع المصادر الحديثة لاظهار فاعلية الCNN وامكانياتها في مجال التعرف على الحروف.

REFERENCES

[1] K. N. Fukushima, "A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," vol. 36, p. 193–202., 1980.

[2] S. A. C. A. N. S. S. a. B. Y. Ahlawat, "Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN)," *Sensors,* vol. 20, no. 12, p. 3344, 2020.

[3] J. M. S. R. A. K. a. M. U. Memon, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," *IEEE Access 8* , vol. 8, pp. 142642-142668, 2020.

[4] C. L. &. F. H. Liu, "Classification and learning for character recognition: Comparison of methods and remaining problems,," *In Int. Workshop on Neural Networks and Learning in Document Analysis and Recognition.* , 2005.

[5] S. N. M. S. A. Z. a. W. G. A.-K. Nawaz, "An approach to offline Arabic character recognition using neural networks,," *In 10th IEEE International Conference on Electronics, Circuits and Systems, 2003. ICECS 2003. Proceedings of the 2003,* vol. 3, pp. 1328-1331, 2003.

[6] P. a. S. B. Singh, "Feature extraction and classification techniques in OCR systems for handwritten Gurmukhi Script–a survey," *International Journal of Engineering Research and Applications (IJERA)* , vol. 1, no. 4, pp. 1736-1739, 2011.

[7] G. Sarker, "A Survey on Convolution Neural Networks," *In 2020 IEEE REGION 10 CONFERENCE (TENCON),* pp. 923-928, 2020.

[8] S. R. M. K. a. M. K. J. Narang, "DeepNetDevanagari: a deep learning model for Devanagari ancient character recognition," *Multimedia Tools and Applications* , vol. 80, no. 13, pp. 20671-20686, 2021.

[9] C. Nebaaer, "Evaluation of convolution neural networks for visual recognition. Neural Networks," *IEEE Transactions on* , vol. 9, no. 4, pp. 685-696 , 1998.

[10] P. Y. D. S. a. J. C. P. Simard, "Best practices for convolutional neural networks applied to visual document analysis.," *In Icdar,* vol. 3, no. 2003, 2003.

[11] M. D. a. R. F. Zeiler, "Stochastic Pooling for Regularization of Deep Convolutional Neural Networks.," *arXiv preprint arXiv:,* p. 1301.3557, 2013.

[12] M. D. a. R. F. Zeiler, "Visualizing and Understanding Convolution Networks.," *In European conference on computer vision,*, pp. 818-833, 2014.

[13] J. W. W. a. K. L. Gan, "A new perspective: Recognizing online handwritten Chinese characters via 1-dimensional CNN,'," *Information Sciences*, vol. 478, pp. 375-390, 2019.

[14] X. a. K. Y. Zhang, "An algorithm of bidirectional RNN for offline handwritten Chinese text recognition,," *In International Conference on Intelligent Computing*, pp. 423-431, 2019.

[15] Y.-S. C.-H. C. Y.-L. C. a. Z.-Y. Y. Su, "A Finger-Worn Device for Exploring Chinese Printed Text With Using CNN Algorithm on a Micro IoT Processor," *Ieee Access*, vol. 7, pp. 116529-116541, 2019.

[16] B. C. a. M. B. H. Kandemir, "Mobile reader: Turkish scene text reader for the visually impaired," *In 2016 24th Signal Processing and Communication Application Conference (SIU)*, pp. 1857-1860, 2016.

[17] Y. W. Z. S. H. J. Y. J. X. a. G. G. Yin, "Deep Learning-Aided OCR Techniques for Chinese Uppercase Characters in the Application of Internet of Things," *IEEE Access*, vol. 7, pp. 47043-47049, 2019.

[18] J. P. Premi, "CNN based Digital alphanumeric archaeolinguistics apprehension for ancient script detection," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* , vol. 12, no. 6, pp. 5320-5326, 2021.

[19] M. A. A. B. P. B. E. C. Z. C. C. C. G. D. A. D. J. D. M. a. G. S. Abadi, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv*, p. 1603.04467, 2016.

[20] H. T. H. L. A. U. H. E. B. a. D. Z. Weldegebriel, "A New Hybrid Convolutional Neural Network and eXtreme Gradient Boosting Classifier for Recognizing Handwritten Ethiopian Characters," *IEEE Access 8* , pp. 17804-17818, 2019.

[21] A. K. T. a. A. Ashiquzzaman, "Handwritten Arabic numeral recognition using deep learning neural networks," *IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, pp. 1-4, 2017.

[22] B. T. D. a. M. S. I. Purkaystha, "Bengali Handwritten Character Recognition Using Deep Convolutional Neural Network," *In 2017 20th International conference of computer and information technology (ICCIT)*, pp. pp. 1-5. IEEE, 2017.

[23] N. N. S. a. M. S. V. Sarika, "CNN based Optical Character Recognition and Applications," *In 2021 6th International Conference on Inventive Computation Technologies (ICICT)*, pp. 666-672, 2021.

[24] K. S. Younis, "Arabic handwritten character recognition based on deep convolutional neural networks," *Jordanian Journal of Computers and Information Technology (JJCIT)* , vol. 3, no. 3, pp. 186-200, 2017.

[25] H. M. K. M. J. S. M. A. a. M. A. R. Al-Barhamtoshy, "Arabic Documents Information Retrieval for Printed, Handwritten, and Calligraphy Image," *IEEE Access 9*, pp. 51242-51257, 2021.

[26] M. R. M. a. M. K. Elleuch, "A new design based-SVM of the CNN classifier architecture with dropout for offline Arabic hand- written recognition,," *Procedia Computer Science* , vol. 80, pp. 1712-1723, 2016.

[27] A. K. a. M. B. C. Boufenar, "Investigation on deep learning for off-line handwritten Arabic character recognition," *Cognitive Systems Research 50,* p. 180–195, 2018.

[28] T. P. B. K. Nisha Sharma, "Recognition for Handwritten English Letters: A Review," vol. 2, 2013.

[29] A. M. Z. A. F. S. M. L. a. T. M. B. Ul-Hasan, "A sequence learning approach for multiple script identification.," *In 2015 13th International Conference on Document Analysis and Recognition (ICDAR) ,* pp. 1046-1050, 2015.

[30] I. S. a. G. E. H. A. Krizhevsky, "ImageNet classification with deep convolutional neural networks,," *in Proc. Adv. Neural Inf. Pro- cess. Syst.,,* p. 1097–1105, 2012.

[31] K. a. A. Z. Simonyan, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv,* p. 1409.1556, 2014.

[32] P. S. B. a. R. P. Dansena, "Generation of Synthetic Data for Handwritten Word Alteration Detection.," *IEEE Access,* vol. 9, pp. 38979-38990, 2021.

[33] L. Z. J. H. A. A.-D. A. D. Y. A.-S. O. S. J. F. M. A.-A. M. a. F. Alzubaidi, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions.," *Journal of big Data,* pp. 1-74, 2021.

[34] M. A. Z. a. I. A. Eltay, "Exploring Deep Learning Approaches to Recognize Handwritten Arabic Texts," *IEEE Access ,* vol. 8, pp. 89882-89898, 2020.

[35] O. V. K. P. R. a. M. S. Bulan, "Segmentation- and Annotation-Free License Plate Recognition With Deep Localization and Failure Identification," *IEEE Transactions on Intelligent Transportation Systems ,* vol. 18, no. 9, pp. 2351-2363, 2017.

[36] H. a. Z. F. Bhiwapurkar, "Optical Character Recognition using Natural Images based on ResNet50 Pre-trained model with Data Augmentation," 2021.

[37] D. S. a. D. S. B. Balci, "Handwritten Text Recognition Using Deep Learning," *CS231n: Convolutional Neural Networks for Visual Recognition, Stanford University, Course Project Report, Spring,,* pp. 752-759, 2017.

[38] Naseer, Asma, and Kashif Zafar. "Comparative analysis of raw images and meta feature based Urdu OCR using CNN and LSTM." *International Journal of Advanced Computer Science and Applications* 9, no. 1 (2018): 419-424.

[39] S. K. F. K. a. S. H. C. Leem, "Detecting Mid-Air Gestures for Digit Writing With Radio Sensors and a CNN," *IEEE Transactions on Instrumentation and Measurement ,* vol. 69, no. 4, pp. 1066-1081, 2019.