التعرف على معالم أندلسية باستخدام نموذج VGGNet للتعلم العميق

رضوان علي بلقاسم حسين، عبد الحميد الفلاح الواعر
كلية تقنية المعلومات – جامعة طرابلس
r.husain@uot.edu.ly

الملخص

تقنيات الذكاء الاصطناعي تُسخر الآن لخدمة كافة مجالات الحياة، اقتصادية كانت أو طبية، أو تعليمية، أو عسكرية أو سياحية، وهي تقنيات تتميز باستمرارية تطورها وتستوحي بناء نماذج خوارزميات ذكائها من خلال الطبيعة التي نحيا فيها، في أسلوب التعامل مع المعضلات وحلها، وهي متعددة المنهجيات في الذكاء الاصطناعي، وأشهرها في هذه الحقبة، منهجية تعلم الآلة (Machine Learning) التي يتفرع منها أسلوب حديث يعرف بالتعليم العميق (Deep Learning)، وهو الذي بناؤه مستوحى من مفهوم شبكة الخلايا العصبية الدماغية (Artificial Neural Networks).

إن هذا المجال المتطور يبشر بحل مشاكل كانت ضربا من الخيال يوماً ما، وانتشرت تطبيقاته المبتكرة الجديدة بشكل كبير جداً مؤخراً، وفي هذه الورقة سيتم بناء نموذج تعلم عميق يعمل على التعرف على بعض المَعالم الأندلسية الشهيرة، والنموذج سيكون بمثابة العقل المفكر في تطبيق الهاتف المحمول الذي يلتقط صورة المَعلم الأندلسي، فيحلل جزئيات الصورة محاولاً التعرف عليها وذكر اسم ذلك المَعلم، والنظام المتطور لهذا التطبيق الذكي سيستخدم تقنية خدمات الويب(Web Services) للتواصل مع قاعدة بيانات النظام، والرد بالمعلومات التي يحتاجها المستخدم، كما يعتبر هذا المجال من بصريات الحاسوب(Computer Vision) التي تعنى بقدرة الحواسيب على تمييز الصور والأشكال.

الكلمات الدالة: تعلم الآلة، تعلم عميق، بصريات الحاسوب، معالم أندلسية.

Recognition of Andalusian Landmarks Using VGGNet Deep Learning Model

Rudwan Husain, Abdelhamid Elwaer

Faculty of Information Technology, University of Tripoli

r.husain@uot.edu.ly

Abstract

Artificial Intelligence technologies are now being harnessed to serve various aspects of life, whether economic, medical, educational, military, or tourism-related. These technologies continuously evolve and draw inspiration from nature, mimicking the building of intelligent algorithms based on the way we live and deal with problems. They encompass diverse approaches in Artificial Intelligence, with one of the most prominent methods in this era being Machine Learning, from which a modern approach known as Deep Learning has emerged. Deep Learning is inspired by the concept of Artificial Neural Networks, mimicking the neural networks in the human brain. Indeed, this evolving field holds the promise of solving problems that were once considered purely imaginative. Recently, innovative applications of Artificial Intelligence have seen significant expansion. In this paper, a Deep Learning model will be constructed to recognize some famous landmarks of Andalusia. The model will act as the cognitive mind within a mobile application. When the user captures an image of an Andalusian landmark, the model will analyze the image's details, attempting to recognize the landmark's name. The advanced system of this intelligent application will employ Web Services technology to communicate with the system's database and provide the user with the required information. Moreover, this field falls within the domain of Computer Vision, which focuses on the ability of computers to distinguish images and shapes accurately.

Keywords: Machine Learning, Deep Learning, Computer Vision, Andalusian landmarks.

1. المقدمة

إن قدرة الأجهزة الحاسوبية، المختلفة الحجم والقدرة، والثابتة والمتنقلة، في التعرف على الأشكال والصور وتمييزها، هي خاصية تميز مدى تطور تقنيات الحواسيب (Magic وMagic، 2019)، ولكن دعم هذه القدرة البصرية بالذكاء الاصطناعي هو ما يفتح الآفاق أمام مجال فسيح من التطبيقات الحديثة التي تخدم الإنسان حيثما كان. وفي عصر ازدهار البيانات، وعلى وجه الخصوص البيانات البصرية، التي تشهد نمواً هائلاً مع تنامي مقدرة تحليل هذه البيانات باستخدام الأساليب الحديثة المادية والبرمجية، يمنحنا تقدم التكنولوجيا وتوافر مثل هذا الكم الهائل من البيانات فرصاً لاستخدامها بطرق يمكن أن تغير حياتنا وتسهم في تيسيرها (Termritthikun ورفاقه، 2012)، (Yap ورفاقه، 2012). ومن أهم تقنيات الاستفادة من البيانات الهائلة هو مجال تعلم الآلة. فهو يشكل الآلية الأساسية لاستخراج المعلومات من البيانات ثم استخلاص الحكمة وراء تلك المعلومات، فإن تطبيقات هذه التقنيات عندما تتكامل مع بعضها تصبح غير محدودة، وآفاقها تزداد اتساعاً (Bulkunde ورفاقه، 2017) (Chen ورفاقه، 2014).

تقليديا كان التعلم الآلي يقتصر على معالجة مجموعات البيانات الصغيرة فقط، وهذا يعني أنه كان من العسير تطبيق أفكار البصريات من مجال معالجة الصور الرقمية لحل مشاكل العالم الحقيقي، ولم يكن ممكنا أساساً لمحدودية القدرة الحاسوبية مقارنة بكم المعلومات التي تحتويها البيانات التصويرية، إضافةً إلى ذلك بيانات الصور الرقمية التي لم تكن متاحة بدقة كافية لاستخلاص معلوماتها، فتلكم الدقة ضرورية لتدريب الآلات الحاسوبية، ولم تتوفر قوة حسابية كافية لتشغيل خوارزميات التعلم (Schmidhuber، (2015، Krizhevsky) ورفاقه، 2017). أما في السنوات الأخيرة نمت قدرة الحوسبة بشكل كبير ، وأصبحت كميات هائلة من البيانات متاحة، وتم اكتشاف خوارزميات جديدة لمعالجة البيانات. كل هذه التطورات مهدت الطريق لنشر البرامج والأجهزة ذات التعقيد الذي يصعب جداً تخيله.

نتيجة للتقدم في التعلم الآلي، والتعلم العميق، شهد مجال رؤية الكمبيوتر تقدمًا كبيرًا مؤخرًا، أدى إلى تطوير خوارزميات جديدة، منها التعرف على بعض المهام الواقعية التي كان البشر يقومون بها يدوياً في السابق، ورؤية الكمبيوتر لديها إمكانات هائلة ليتم تطبيقها في سياقات مختلفة، حيث يمكن للنظام البصري البشري فقط القيام بهذه المهام في الماضي، كذلك التعرف على الصور مثلا بالنسبة للإنسان، وتبدو المهمة بسيطة؛ لأن النظام البصري البشري قد تطور لملايين السنين، وتعود على هذا النوع من المهام الأساسية، وأيضاً تطور فهم الإنسان لآليات البصر وأساليب تعامل العقل البشري مع بياناته مما أسهم في فهم عمل النظام البصري البشري جيداً، وهو الأمر الذي أسهم في بناء خوارزميات حسابية لمعالجة الصور، وتحديد تطور أدوات التقاط الصور والرفع من كفاءتها.

الصورة الرقمية التي تخزن في الحواسيب هي مجرد مجموعة من النقاط المصفوفة (pixels) غير المترابطة من حيث المعنى الذي تحمله أي نقطة، وإن كانت النقاط متجاورة في المكان، ومن وجهة نظر الكمبيوتر كل نقطة مستقلة بذاتها، وكان الإنسان وحده القادر على استنباط الترابط بين نقاط الصورة، وأصبح من الصعب جعل الآلات " تفهم" التغييرات التي تحدث فيما بين نقاط الصور لتشكل معالم الأشياء المصورة، ما يؤدى إلى تغير فهم الكائنات المجسدة في الصور، وأي تغير بين نقاط الصور تعتبر تغييرات لها تأثير على ما ترويه بيانات الصورة. فإدراك الآلات لهذا التباين بين مكونات الصور يتطلب براعة هائلة، وقوة حساب ضخمة عالية، لتتمكن الآلة من أداء هذه المهمة بنجاح، ولم يتحقق ذلك إلا حين تمكنت الحواسيب من هذا النوع من الإدراك، عندها تفوقت على البشرية في استيعاب مكونات الصور، وتنافسوا في التمييز بين تفاصيلها التي قد لا تدركها العين البشرية، ولم يتمرس العقل البشري على إدراك تفاصيلها.

2. الأعمال المناظرة

في بحث (Banlupholsakul ورفاقه، 2014) تم استخدام أسلوب يعرف بآلة دعم المتجه (Support Vector Machine) كخوارزمية تستعمل للتعرف على بعض المعالم

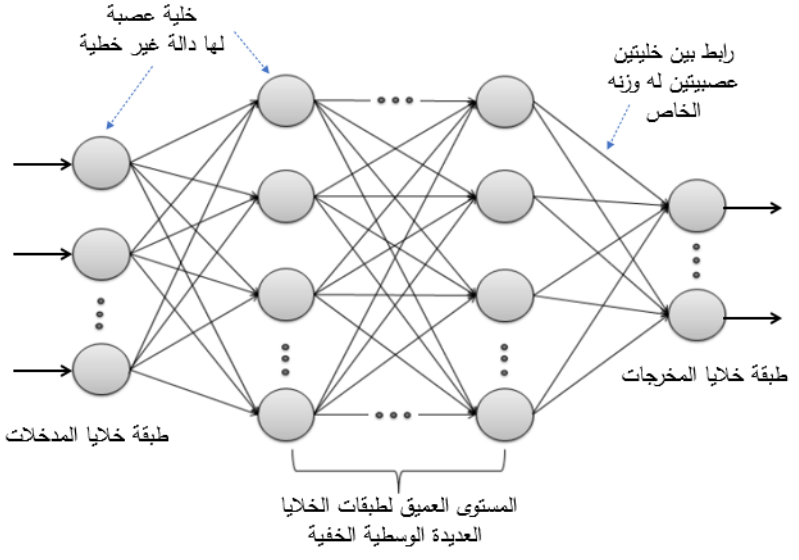
الأندلسية. هذا الأسلوب يتعرف على الصور بطريقة توافق المحتوى حيث يتم استخلاص المعلومات من الصور بناءً على حقيبة من نماذج الكلمات وذلك لتقليص التكلفة الحاسوبية (Yap ورفاقه، 2012). بالرغم من ذلك، نموذج حقيبة الكلمات لديها عيبان كبيران، أولهما: الحقيبة التي تتجاهل التوزيع المكاني للمعلومات في صورة المعالم. وثانيهما: أن هذا الأسلوب لا يستطيع التمييز بين الواجهات المتعددة التي قد تحتويها الصورة.

العمل المنجز في المرجع (Lowe، 1999)، لكي يتم التعرف على المعالم المختلفة ويستخلص معلومات من الصور، فالصورة تصنف إلى معالم شاملة، ومعالم جزئية، متعددة، محلية، وهذا للتفرقة بين مكونات الصورة، من حيث اللون، والاتجاه النسيجي لنقاط الصورة. إن خوارزمية ما يعرف بثبات المقياس لتحول المعالم (Scale–Invariant Feature Transform) هي طريقة لاستخلاص المعالم من الصور، وهي التي حسنت دقة استخلاص المعالم كما حسنت في فترة الحوسبة. ولقد استخدم مبدأ آلة دعم المتجه (SVM) في (Chen ورفاقه، 2014)، لأجل التدريب على بيانات الصور، ولمعرفة حزمة البيانات التدريبية لابد أن تحتوي على بيانات متجهة مضافة لبيانات الصورة الأصلية؛ لكي تدعم الآلية لتمييز ملامحها، ولاستخلاص معالم الصورة يجب بناء مخطط هيستقرام توزيع ألوان الصورة (Dalal و Triggs، 2005). وهذا عيبه أنه لا يتكيف مع ثبات توجه الصور.

في هذه الورقة البحثية ، سيتم تطوير نظام لتمييز بعض المعالم الأندلسية باستحدام منهجية التعلم العميق وهذا الأسلوب مبني على تقنية الشبكات العصبية الدماغية الالتفافية (Convolutional Neural Networks)لكي يتم استخلاص ملامح الصور وبالتالي التعرف على المعالم التي تجسدها الصورة (Benkaddour و Bounoua، 2017). باستخدام هذه التقنية متمثلة في شبكة رزنت (ResNet)، سيكون ممكناً الوصول إلى أفضل اختيارات لتصنيف المعالم بدقة تصل إلى 96.53% من خلال مختلف فئات الصور التي سيتتدرب عليها النظام ويتعلم ملامحها فيتعرف عليها كلما التقط صورة مقاربة لتلك العينة (Hembury ورفاقه، 2003).

3. تقنية الشبكات العصبية الدماغية الالتفافية للتعلم العميق

مفهوم الشبكة العصبية الدماغية هو مستوحاً من تركيبة شبكة الخلاية العصبية بمخ الإنسان. من خلال تواصل تلكم الخلايا العصبية مع بعضها واستمرار تقوي ترابطها مع دوام التواصل بينها لاستيعاب المعلومات وتحليلها، بُنيَ نظام حاسوبي يتكون من مجموعة من العقد التي تسمى الخلايا العصبية (Neurons) يتم تنظيمها في طبقات. تشكل تلك الوحدات العصبية المتصلة شبكة واسعة ذات طبقات عديدة يعطيها اسمها الذي يصفها بالعميقة. وكل اتصال بين الخلايا العصبية ينقل إشارة من خلية عصبية إلى أخرى. تقوم الخلايا العصبية المستقبلة بمعالجة الإشارة وترسل إشارة إلى الخلايا العصبية المتصلة بها داخل شبكة الخلايا والروابط بينها تتجسد في مصفوفات رقمية تحوي أرقام تمثل قوة الترابط بين الخلايا العصبية المتمثلة في صيغتها الرياضية غير الخطية. فكلما تعلمت الشبكة معلومات جديدة كلما ضبطت أوزان الروابط بين خلاياها العصبية. الشكل (1) يوضح مخطط لشبكة الخلايا العصبية.



شكل (1): مخطط لشبكة الخلايا العصبية الذكية

مثل هذه الأنظمة تتعلم لأداء المهام من خلال النظر في الأمثلة المقدمة لها، وبشكل عام، دون أن تكون مبرمجة مسبقاً مع أي قواعد خاصة بالمهمة الموكلة إليها، فإن المهمة الرئيسية لها أنها تستقبل مجموعة من المدخلات، ثم أداء الحسابات المعقدة، ثم إنتاج المخرجات لحل مشكلة، وأما طبقة المدخلات فتأتيها البيانات الأولية لتدخل إلى النظام، مع مزيد من المعالجة، بواسطة طبقات لاحقة متعددة، من الخلايا العصبية الاصطناعية، وأم الطبقات الوسطية المخفية المتعددة، فتقع بين طبقة المدخلات، وطبقة المخرجات،

حيث تأخذ خلاياها العصبية الاصطناعية مجموعة ما، وتعمل على معالجتها في مستوى المدخلات، وتنتج مخرجات من معالجة البيانات، في دوال غير خطية تفعيلية، وأما طبقة خلايا المخرجات فتشكل الطبقة الأخيرة من الخلايا العصبية الاصطناعية التي تنتج مخرجات نموذج النظام الذكي العميق (Krizhevsky و Hinton، (2017، Szegedy) ورفاقه، 2014)، (Simonyan و Zisserman، (2014، Hembury) ورفاقه، 2003).

تتميز هذه الشبكات عن الشبكات العصبية ذات الطبقة المخفية المفردة من حيث عمقها، وهنا يكمن سر التعليم العميق المتمثل في تعدد الطبقات، وكثرة الخلايا فيها، إلى عدة طبقات تمر خلالها البيانات في عملية متعددة الخطوات، فهي قادرة على حل مهام التصنيف، والتعرف على الأنماط الأكثر تعقيداً، بفضل الطبقات المخفية المتعددة، وكانت الإصدارات السابقة من الشبكات العصبية مؤلفة من طبقة إدخال واحدة، وطبقة إخراج واحدة، مع طبقة واحدة مخفية بينهما.

4. نظرية تعلم الشبكة العصبية

في شبكات التعلم العميق يتم تدريب كل طبقة من العقد، على مجموعة مميزة من الملامح، استناداً إلى مخرجات الطبقة السابقة، وكلما تقدمت في الشبكة العصبية، كلما ازدادت تعقيدات الملامح التي يمكن للعقد التعرف عليها؛ وذلك نظرًا لأنها تقوم بتجميع وإعادة تجميع ميزات الملامح من الطبقات السابقة، ومن يتدرب على نموذج لشبكة

عصبية عميقة، فهو يحاول تحسين أوزان النماذج المتجسدة في الروابط بيت عقد الخلايا العصبية، والهدف من التدريب هو العثور على الأوزان التي تحدد بدقة بيانات المدخلات، إلى فئة الإخراج الصحيحة، وهذا التعيين هو ما يجب أن تتعلمه الشبكة ويمثل المعرفة التي تكتسبها (Schmidhuber، 2015.).

بداية التعلم تبدأ بما يعرف بالانتشار الأمامي وهو عملية تغذية الشبكة العصبية بمجموعة من المدخلات، مع أوزان القوة بين الروابط البينية للخلايا، ثم تغذيتها عبر دوال غير خطية، تمثل عقدة الخلية العصبية، لأن الشبكة العصبية تولد جاهلةً في البداية، أي لا تعرف ما هي الأوزان التي ستقوم بترجمة المدخلات لإجراء التنبؤات الصحيحة، فيجب أن تبدأ بتخمين قيم عشوائية للقوة الرابطة بين أي خليتين، ومن ثم محاولة إجراء تنبؤات أفضل مع الوقت، لأنها تتعلم باستمرار ؛ لتحسين قوة وزن روابط الخلايا من خلال أخطائها السابقة، واختبار الفرضيات والمحاولة مرة أخرى، وهنا يكمن مبدأ التعلم والتحسن المستمر (Simonyanو Zisserman، 2014). يقاس فاقد المعرفة للشبكة من خلال التنبؤات غير الصحيحة التي تحاولها الشبكة، أي أن الفاقد هو رقم يشير إلى مدى سوء تنبؤ النموذج، وهذا الفاقد هو الذي يقرر به زيادة، أو نقصان قوة وزن روابط الخلايا، فإذا كان تنبؤ النموذج مثالياً، فيكون الفاقد صفراً، وخلاف ذلك إن كان الفاقد أكبر أو أصغر من الإجابة الصحيحة، والهدف من تدريب نموذج هو العثور على مجموعة من الأوزان التي تصف قوة ترابط الخلايا التي تحسب من دقة تعلم الشبكة، ويحتاج النظام لتقليل الفاقد في التعلم المحسوب في شبكة عصبية، يتم ذلك باستخدام مفهوم الانتشار الخلفي، الذي يعمل عن طريق تحديد الفاقد في المخرجات ثم نشرها مرة أخرى إلى الخلف، نحو طبقة سابقة في الشبكة، حيث يتم استخدامه لتعديل الأوزان بطريقة تقلل من الخطأ في الناتج.

شبكات التعلم العميق مفيدة للغاية للتعرف على الأشياء في الصور، والوجوه، والكائنات، وما إلى ذلك، فبالنظر إلى مجموعة بيانات الصور أحادية اللون الرمادية ذات الحجم القياسي 32 × 32 نقطة، فإن الشبكة العصبية التقليدية تتطلب 1024 مدخلاً في طبقة المدخلات، وهذا يفقد الميزة المكانية للملامح، أما منهجية الشبكة الالتفافية (CNN) فهي

استخراج ميزات ملامح الصورة كمدخلات للشبكة العصبية، وليس مجرد نقاط منعزلة من الصورة. هذا الأسلوب يحافظ على العلاقة المكانية بين نقاط الصورة، مما يفيد في تعلم ميزات الصورة وملامحها، ويتم ذلك بتجزئة الصورة المصدرية إلى مربعات صغيرة، تؤلف مصفيات الطبقات الالتفافية وهي مدخلات الخلايا العصبية للطبقة الأولى، وحجم المدخلات هو مصفٍ بمربع ثابت، إذا كانت الطبقة عبارة عن طبقة الإدخال، فسيكون الإدخال هو قيم من نقاط الصورة، وإذا كان القيم داخلة للطبقات العميقة في بنية الشبكة، فإن الطبقة ستأخذ المدخلات من الطبقة السابقة لها. من أهم الميزات التي يمكن للمربعات المصفيات اكتشافها في الصورة، هي تلكم الحواف، أو قد تكتشف الزوايا، أو قد يكتشف البعض من الدوائر، أو مربعات، وكلما زاد عمق الشبكة أصبحت المصفيات أكثر تطوراً، وبالتالي تزداد تعلماً في الطبقات التالية، فيتطور التعرف بدلاً من الحواف والأشكال البسيطة في الطبقات اللاحقة، فقد تتمكن من اكتشاف كائنات محددة، مثل العينين، والأذنين، والشعر وتفاصيل الوجه، في الطبقات الأعمق، وتكون قادرة على اكتشاف كائنات أكثر تطوراً.

5. المعالم الأندلسية بيانات تدريبية

فئة بيانات المعالم الأندلسية التي تستخدم لتعليم هذا النظام، تتألف من مئات الصور التي تجسد مناطق، ومشاهد مختلفة، للقصور، والمباني، والحدائق، والكثير من الآثار الأندلسية، وهذه المعالم يتم تجميعها من خلال محركات البحث عبر الإنترنت، والشكل (2) يبين عينة من صور المعالم الأندلسية، الصور مجمعة من مصادر مختلفة مثل Wikipedia، Twitter، و Flicker. إن تجميع بعض صور المعالم الأندلسية وحفظها كبيانات يتعلمها النظام يمر بمراحل مختلفة، حيث تجمع الصور بطريقة آلية من خلال تعليمات برمجية تعمل على البحث الذاتي عن صور بصفات معينة، وتحدد هذه الصفات من خلال كلمات مفتاحية تساق مع التعليمات البرمجية (Flickr، 2018). بعد الحصول على الصور، تصنف الصور إلى مجموعات حسب اسم المعلم الذي تصوره. كل صنف من الصور يقسم إلى ثلاثة أقسام بنسبة 60% لتدريب نموذج خوارزمية التعلم العميق،

وبنسبة 20% لاختبار صحة تعلم الخوارزمية، و20% من صنف صور المعالم بمصادقة تعلم الخوارزمية.

شكل (2): عينات من المعالم الأندلسية يتعلم النظام الذكي للتعرف عليها، مصدرها: ar.wikipedia.org

6. تعزيز البيانات لتعويض الصور المستهدفة المفقودة

لكي تتدرب الشبكة العصبية بكفاءة أعلى، على نموذج التعلم العميق، فتحتاج للكثير من الصور المختلفة عن المعلم الواحد، وهذا يجنب نموذج الشبكة العصبية مما يعرف بتخمة الأوزان (Overfitting) التي تحدث نتيجة تكرار تدريب النموذج على فئة محدودة من البيانات، فيتشبع النظام بها ويضبط أوزان قوة الترابط بين الخلايا، بناءً على فئة قليلة من المعلومات، فلا يميز غيرها، ظنا منه بأنه تعلم تمييز تلك الملامح، ولتعزيز البيانات استُخدم من قبل لزيادة فئة الصور في حزمة بيانات التدريب (Shorten و Khoshgoftaar، 2019)، والاختبار والمصادقة، وهو أساس لتعويض الفاقد من بيانات الصور، وتستعمل تقنيات مختلفة لتعزيز البيانات التصويرية، فيقوم النظام بتحويرات على الصور الرقمية، لإنتاج صور جديدة، وهذه التحويرات تشمل الدوران العشوائي، من قلب الصور، وتكبيرها أو تصغيرها، وتغيير الإضاءة، وتباين الألوان، وإزاحة محتويات الصورة،

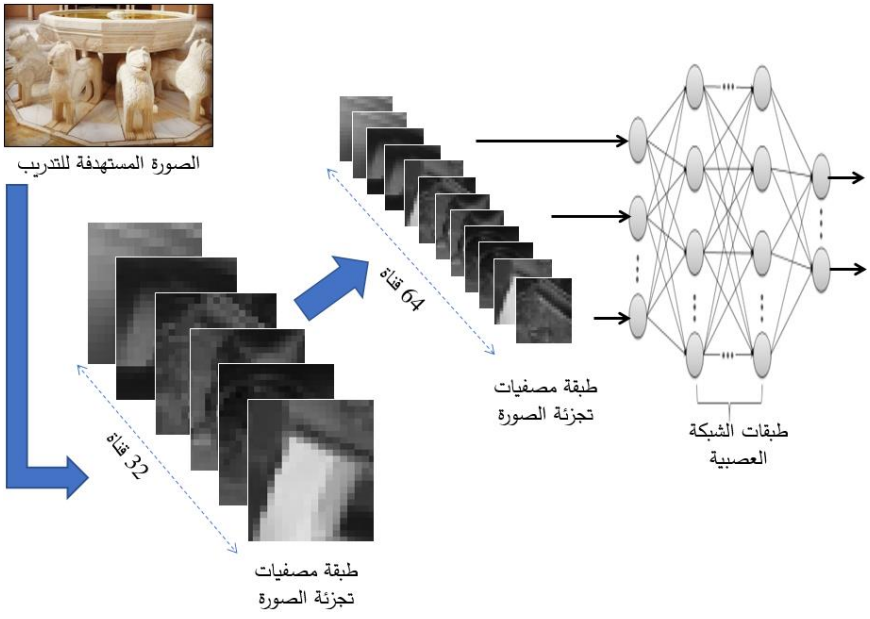
وغيره كثير. الشكل التالي (4) يعطي لمحة عن تعزيز البيانات التصويرية لإنتاج آلاف الصور الجديدة التي تدعم عملية التعلم في الشبكة العصبية.

شكل (3): صور متعددة الإضاءات والإسقاطات لنافورة السباع

شكل (4): إنتاج صور جديدة لتعزيز بيانات تدريب نموذج الشبكة العصبية العميقة

7. بناء الشبكة العصبية وتدريبها

هذه الورقة البحثية تبني شبكة للتعلم العميق، بنموذج لشبكة عصبية التفافية، مستوحاة من شبكة المجموعة الافتراضية الفراغية (Virtual Geometry Group)، وتعرف بفجيجي نت (VGGNet). والشبكة المتعددة الطبقات العصبية، وتسبقها طبقات عمليات تصفية للصور الداخلة في مرحلة الالتفاف، التي تصفي صور المعالم الأندلسية من خلال تجزئتها لقطع متساوية، تحوي ملامح جزئيات تلك المعالم المصورة، وبهذا الأسلوب تُمكِّن خوارزمية النظام المطور الشبكة العصبية العميقة من التعلم أكثر فأكثر من الملامح الغنية بالمعلومات، عن المعالم المستهدفة، الشكل التالي (5) يعرض البنية العامة لهيكلية شبكة التعلم العميق، التي تتدرب على التعرف على ملامح معلم نافورة السباع.



شكل (5): هيكلية نموذج الشبكة العصبية الالتفافية للتعلم العميق المستخدم للتعرف على المعالم الأندلسية

الشبكة تحتوي على فئتين من طبقات مصفيات جزئيات الصورة، تتبعها فئة من الطبقات المتعددة للخلايا العصبية العميقة، مع طبقة المصفيات الأولى، وستتعلم من 32 مصفيا، كل بحجم 3×3 نقاط. والطبقة الثانية للمصفيات سوف تتعلم بمصفيات عددها 64

مصفيا، كل بحجم 3×3 نقاط. يتلوها مصفيات بنافذة منزلقة بحجم 2×2 نقاط. والتفاصيل التقنية لطبقات الشبكة موضحة في الجدول (1).

8. التجارب والنتائج

صُممت الشبكة، ودُربت في جهاز حاسوب على الأداء، ومواصفات الحاسوب تتلخص في وجود معالج من نوع إنتل آر آي 7 (Intel R i7 CPU) وبسرعة 2.2 قيقا هيرتز، ويحتوي على 8 أقطاب، بذاكرة 16 قيقا بايت. والحاسوب يعمل بنظام تشغيل ويندوز 10 بهيئة 64 بت. والصور الداخلة لخوارزمية نموذج الشبكة تم تجهيزها لتكون بحجم 64×64 نقطة لكل صورة من صور المعالم الستة المستهدف تعلمها وموضحة في الشكل (2) أعلاه. كل معلم منح رمزاً رقمياً خاصا به، وبقيمة بين 0 و 1 ليكون عدد مخارج الشبكة العصبية 6 مخارج، كل منها مخصص لمعلم معين. كما أن بيانات الصور المجمعة والمعززة، تم تقسيمها لثلاث فئات. 60% من صور أي معلم تستخدم لتدريب الشبكة، 20% من صور المعلم تستخدم لاختبار الشبكة، و20% الباقية تستخدم للمصادقة على دقة تعلم الشبكة لمعلم أندلسي.

جدول (1): مواصفات تقنية للشبكة العصبية متعددة الطبقات

نوع الطبقة	حجم المخرجات	حجم نافذة المصفى
INPUT IMAGE	64 × 64 × 3	
CONV	64 × 64 × 32	3 × 3, K = 32
ACT	64 × 64 × 32	
BN	64 × 64 × 32	128
CONV	64 × 64 × 32	3 × 3, K = 32
ACT	64 × 64 × 32	
BN	64 × 64 × 32	128
MAX POOL	32 × 32 × 32	2 × 2
DROPOUT	32 × 32 × 32	
CONV	32 × 32 × 64	3 × 3, K = 64

ACT	32 × 32 × 64	
CONV	32 × 32 × 64	3 × 3, K = 64
ACT	32 × 32 × 64	
MAX POOL	16 × 16 × 64	2 × 2
DROPOUT	16 × 16 × 64	
FLATTERN	16384	
FC	512	
ACT	512	
BN	512	
DROPOUT	512	
FC	5	
SOFTMAX	5	

مبدأ تبادل القصور (Cross−Entropy) حسب المرجع (Janocha، 2017) تم استخدامه كوظيفة دالة حساب فاقد التعلم، ومعدل تزايد التعلم (Learning Rate) وكان بقيم ابتدائية 0.006 ثم يتولى النظام ضبطها حسب صحة وخطأ التعرف، وبها يتم ضبط قوة أوزان الروابط بين الخلايا العصبية، وإن الشبكة دُربت لعدد 100 جولة باستخدام حزم بيانات بحجم 32. والجدول (2) يعرض تقرير دقة تعلم نموذج الشبكة في تصنيف صور المعالم بدقة تصل إلى 99%.

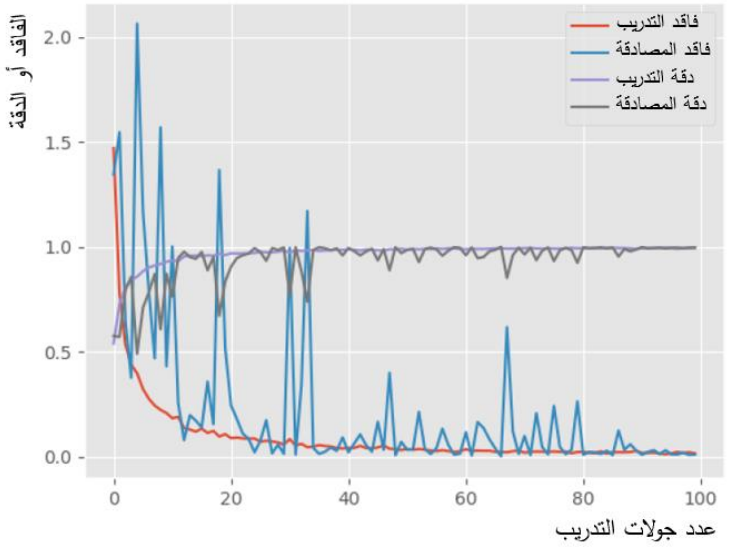
جدول (2): تقييم أداء الشبكة العصبية العميقة

المعلم	الدقة	تذكر المعلم	تكرار التعلم
نافورة السباع	1.00	0.99	270
قصر الحمراء	0.98	0.97	260
عمود لا غالب إلا الله	1.00	1.00	300
المحراب	0.99	1.00	255

باب ونقوش	0.97	0.96	284
قصر إشبيلية	0.96	1.00	265
الدقة العامة للشبكة	0.98	0.99	1634

فمعدل فاقد التعلم، ودقة تدرب الشبكة، مع مرور الزمن، موضحة في منحنيات جولات تدريب الشبكة المرسومة في الشكل (6). على المحور الأفقي السيني، يوجد عدد جولات التعلم تتزايد مع مرور الزمن لتصل إلى 100 جولة. على المحور العمودي التصاعدي يوجد الفاقد ودقة التعلم. بالنظر لسلوك المنحنيات يظهر جلياً تعلم الشبكة لتلك المعالم المستهدفة. بعد حوالي 20 جولة تعلم الشبكة بدأت تصل لدقة تصنيف للمعالم الأندلسية الستة بنسبة 0.98% وهذه دقة عالية ممتازة. والفاقد في التدريب والمصادقة على دقة التمييز يتنازل باستمرار مع فترة تذبذبات محدودة تبينها نبضات المنحنى. وهذا يرجع لمعدل التعلم الابتدائي المفترض بقيمة 0.006 ثم يتحسن الأداء مع ضبط معدل التعلم الذاتي. تقريباً مع الجولة 22 أو 23 يصل النظام لدقة 0.99 أو 1.00 أي تقريبا! نسبة دقة 100%.

الشكل (7) التالي يبين مكونات النظام النهائي الذي يتألف من تطبيق نقال أندرويد، ويعتمد على نموذج الشبكة العصبية العميقة التي تم تدريبها للتعرف على بعض المعالم الأندلسية، ويمكن لنظام العمل بشكل مستقل ما دام حجم البيانات يتلاءم مع إمكانيات الهواتف النقالة، والنظام يمكنه التواصل مع خدمة ويب، التي تحفظ نموذجا معقدا ضخم البيانات، يحافظ على خادم شبكة إنترنت، وليقوم بدور التحليل والتعرف على المعالم، ثم إرسال نتيجتها للمستخدم على هاتفه المحمول، الصور التقطت من الهاتف المحمول لمشاهد حقيقية للمعالم أم لصور ورقية أو ضوئية على شاشات الحواسيب أو التلفاز.



شكل (6): منحنيات جولات تدريب الشبكة العصبية لتتعلم تمييز المعالم الأندلسية

شكل (7): النظام العام لتمييز المعالم الأندلسية

9. الخلاصة

يزداد انتشار استخدام تقنيات الذكاء الاصطناعي مع تطور برمجيات وأعداد الحواسيب، وأيضاً مع إدراك الناس للفوائد الجمة التي تدرك بتوظيف هذه التقنيات الحديثة المتجددة، وشبكات التعلم العميق من أحدث تطورات أدوات الذكاء الاصطناعي، وقد يعود سبب تزايد انتشارها إلى المزايا الرئيسية التي تجعلها أكثر ملاءمة لحل المشاكل التي كانت عصية على البشر، والخوارزميات، والمعدات التقليدية القديمة، وأيضاً يعود تزايد استخدامها إلى القدرة على تعلم ونمذجة العلاقات غير الخطية والمعقدة، وهو أمر مهم للغاية؛ لأن في الحياة الواقعية تكون أغلب العلاقات بين المدخلات والمخرجات، وهي علاقات غير خطية، وكذلك معقدة، كما يمكن القول إن بعد التعلم العميق من المدخلات الأولية وعلاقاتها ببعض، ويكون النموذج قادرا على التنبؤ بماهية البيانات الجديدة التي تمر عليه، ولعل أكبر قيود على نماذج التعلم العميق هي أنها تتعلم من خلال الملاحظات، وهذا يعني أنهم يعرفون فقط ما كان في البيانات التي تم تدريبهم عليها. إذا كان لدى المستخدم كمية صغيرة من البيانات فلن تتعلم النماذج بطريقة يمكن تعميمها.

في هذه الورقة تم تطوير نموذج شبكة عصبية عميقة التعلم من خلال أدوات لغة بايثون (Python) ومكتبة كيراس (Keras API) الذي يمكنه التعرف على بعض المعالم الأندلسية، وتم تدريبه واختباره على مجموعة بيانات الصور الرقمية وتعزيز نقص الصور بإجراء تحويرات برمجية على الصور المتاحة لإنتاج صور جديدة. تم تدريب النموذج عل حزمة البيانات، وأظهرت خوارزمية شبكة التعلم العميق نموذجاً يتمكن من التعرف على المعالم الأندلسية بدقة تصل إلى 99%. يمكن استمرار تطوير الخوارزمية لإنتاج نموذج يتعرف على أكبر قدر من المعالم ولكن بزيادة المعالم تزداد الحاجة لإمكانيات حاسوبية أكبر قدرة من حيث سعة الذاكرة وسرعة المعالجة.

10. المراجع

[1] J. Magic and M. Magic, *Image Classification: Step-By-step Classifying Images with Python and Techniques of Computer Vision and Machine Learning (2; Python*. Amazon Digital Services LLC - Kdp Print Us, 2019.

[2] C. Termritthikun, S. Kanprachar, and P. Muneesawang, "NU-LiteNet : Mobile Landmark Recognition using Convolutional Neural Networks," no. 1, pp. 3–8, 2012.
[3] K. Yap, Z. Li, D. Zhang, and Z. Ng, "Efficient mobile landmark recognition based on saliency-aware scalable vocabulary tree," 2012, p. 1001.
[4] K. Bulkunde, A. Chakraborty, S. Kazi, and K. Dhumal, "Landmark Recognition using Image processing with MQTT protocol," pp. 2405–2407, 2017.
[5] T. Chen, K. Yap, and D. Zhang, "Discriminative Soft Bag-of-Visual Phrase for Mobile Landmark Recognition", IEEE Transactions on Multimedia, 2014.
[6] J. Schmidhuber, "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.
[7] A. Krizhevsky and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks,", Communications of ACM, 2017.
[8] C. Szegedy *et al.*, "Going Deeper with Convolutions,", Computer Vision Foundation, 2014.
[9] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," pp. 1–14, 2014.
[10] G. A. Hembury, V. V. Borovkov, J. M. Lintuluoto, and Y. Inoue, "Deep Residual Learning for Image Recognition Kaiming," *Cvpr*, vol. 32, no. 5, pp. 428–429, 2003.
[11] K. Banlupholsakul, J. Ieamsaard, and P. Muneesawang, "Re-ranking approach to mobile landmark recognition," in *2014 International Computer Science and Engineering Conference (ICSEC)*, 2014, pp. 251–254.
[12] K.-H. Yap, Z. Li, D.-J. Zhang, and Z.-K. Ng, "Efficient Mobile Landmark Recognition Based on Saliency-aware Scalable Vocabulary Tree," in *Proceedings of the 20th ACM International Conference on Multimedia*, 2012, pp. 1001–1004.
[13] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International*

Conference on Computer Vision, 1999, vol. 2, pp. 1150–1157 vol.2.

[14] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, 2005, pp. 886–893.

[15] M. K. Benkaddour and A. Bounoua, "Feature extraction and classification using deep convolutional neural networks, PCA and SVC for face recognition," *Trait. du Signal*, vol. 34, no. 1–2, pp. 77–91, 2017.

[16] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-fei, "ImageNet : A Large-Scale Hierarchical Image Database," pp. 2–9.

[17] Flickr, "Flickr Services. The App Garden. API Documentation." [Online]. Available: https://www.flickr.com/services/api/. [Accessed: 01-Dec-2018].

[18] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, p. 60, 2019.

[19] K. Janocha and W. M. Czarnecki, "On Loss Functions for Deep Neural Networks in Classification," *CoRR*, vol. abs/1702.0, 2017.

[20] https://ar.wikipedia.org/wiki/بوابة:الأندلس.