مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (5) العدد(2) 2021
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (5) No. (2) 2021

دراسة الفرق بين الشبكات العصبونية الالتفافية وشبكات الكبسولة

د. محمد مازن المحايري*

د. قصي كنفاني**

رنيم حافظ كيوان***

(تاريخ الإيداع 14 /10 /2020 . قُبِل للنشر في 3 /3 /2021)

□ ملخّص □

أصبحت مهمة تحديد الكائنات object detection وتصنيفها classification من مهام شبكات التعلم العميق بأنواعها المختلفة: الشبكات العصبونية الالتفافية التقليدية (Convolutional Neural Network (CNNs، والشبكات الالتفافية الكاملة (Fully Convolutional Network (FCN، وشبكات الكبسولة Capsule Network (Caps Net). تستخلص هذه الشبكات السمات من الصور باعتماد المنهج الالتفافي كما أنها تحدد الكائنات وتصنفها عبر تعلم السمات Feature Learning. سيتم في هذا البحث عرض نموذجين من نماذج الشبكات العصبونية لتحديد الكائنات وتصنيفها: الأول هو النموذج VGG16 كشبكة عصبونية التفافية تنجز مهمة استخلاص السمات باستخدام الطبقات الالتفافية ومهمة التصنيف باستخدام طبقات الاتصال الكامل الثلاثة في نهاية النموذج. تتميز هذه البنية ببنيتها المتجانسة، من حيث حجم مرشحات الطبقات الالتفافية وحجم مرشحات طبقات الاتصال الكامل. والنموذج الآخر يعتمد على شبكة الكبسولة Caps Net مستخدما الشبكة VGG16 كشبكة أساسية تؤدي مهمة استخلاص سمات من صور الدخل لتلافي الضعف الذي تعانيه شبكة الكبسولة. تم اختبار النموذجين على قاعدة بيانات Pascal Voc 2007، فكان معدل متوسط دقة التصنيف mAP 67.8% للنموذج VGG16، و 67.3% لنموذج شبكة الكبسولة كنتيجة لهذا الاختبار من قبل الباحثين.

الكلمات المفتاحية: الشبكات العصبونية الالتفافية، شبكة الكبسولة، التوجيه الديناميكي، VGG16.

*أستاذ – قسم هندسة الحواسيب والأتمتة– كلية الهمندسة 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.

مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (5) العدد(2) 2021
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (5) No. (2) 2021

A study of the difference between Convolutional Neural Network and the Capsule Network

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

(Received 14/ 10/ 2020 . Accepted 3 / 3/ 2021)

□ ABSTRACT □

Object detection and classification has become an important task using deep learning networks of various types: Convolutional Neural Networks (CNNs), Fully Convolutional Networks (FCN), and Capsule Networks (CapsNet). These networks extract features from the images by using convolutional approach. It also detects and recognizes objects through feature learning. In this article, two neural network models for object detection will be presented: the first is VGG16 that performs the feature extraction task using convolutional layers and the classification task using three fully connected layers at the end of the model. This structure is distinguished by its homogeneous architecture, in terms of the size of the convolutional layer filters and the fully connected layer filters. The other model is based on the capsule network, using the VGG16 network as the backbone that performs the task of extracting features from the input images to avoid the vulnerability of the capsule network. The two models were tested on the Pascal Voc 2007 database. Mean Average Precision mAP was 67.8% for the VGG16 model, and 67.3% for the capsule model as a result of this test by researchers.

Keywords: Convolutional Neural Network (CNN), Capsule network (Caps Net), Dynamic Routing, VGG16.

*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.

1- مقدمة:

يعمل نظام الرؤية البشرية على كشف وتصنيف الكائنات بشكل سريع جدا ودقيق ، إذ يستطيع الإنسان اكتشاف أكثر من 30000 صنف كائن بسهولة في غضون ميلي ثانية. أما نظم الرؤية الحاسوبية التي تقوم بمهمة تحديد الكائنات وتصنيفها بالاعتماد على الشبكات العصبونية الالتفافية القليدية Convolutional Neural Network (CNN)، لا تملك التصور الحقيقي لنظام الرؤية البشرية، إلا أنها تمتاز بكونها دقيقة وسريعة، على الرغم من عيوبها العديدة ومنها [1]:

1.الشبكات العصبونية الالتفافية هي شبكات ثابتة الإزاحة، أي أن نموذج CNN يحتاج إلى أكثر من صورة لكائن التدريب المعتبر تمثل مواقعه وإزاحاته المفترضة ووجهات نظره المختلفة ليتعلم مهمة تحديده وتصنيفه وبالتالي قاعدة بيانات التدريب يجب أن تكون بحجم كبير [1].

2. المرشحات المستخدمة ضم هذه الشبكات بحجوم مختلفة إضافة إلى عدم تركيزها على المعلومات النسبية والهرمية (المكانية) بين الكائنات مما لا يمكنها من تحديد موضع كائن نسبة إلى كائن آخر أو موضع جزء من كائن نسبة لباقي أجزائه، كموقع العين نسبة للفم في صورة وجه [1] .

3. تؤدي عملية التجميع الأعظمي Max Pooling في نماذج CNN إلى خسارة معلومات مهمة عن موقع ووضعية الكائن في الصورة [1]. حيث يتمثل دور عملية التجيمع الأعظمي في تقليل الحجم المكاني وبالتالي إنقاص كمية البرامترات والحسابات في الشبكة للتحكم بمشكلة تعويم الذاكرة.

نظرا لقيود نماذج CNN، فإن نتائج التصنيف لم تكن دقيقة بما يكفي في هذه النماذج. لذا، تم اقتراح بنية أخرى من الشبكات الالتفافية، تحتوي طبقات التفافية فقط بدون طبقات الاتصال الكامل المعمول بها في CNN، تعمل هذه الشبكات على التصنيف على مستوى البكسل في الصورة. ظهرت فعالية هذه الشبكات بشكل تقريبي أو بشكل مطلق، وعلى الرغم من قيامها بتحسين الدقة كما في نماذج SSD وYOLO، فهي لم تأخذ بعين الاعتبار معلومات السياق الشاملة، كما أن تجزأتها ليست على مستوى المثيل instance-level، ولمعالجة هذه النقاط، اقترح Geoffrey Hinton نوع أخر من الشبكات الالتفافية يدعى شبكة الكبسولة Capsule network (Caps Net) [9] وذلك في العام 2017 [8].

تعتبر شبكات الكبسولة باستخدامها مفهوم الرؤية العكسية Inverse graphics [10] تمثيل حقيقي لنظام الرؤية البشري. مما سمح للنموذج المعتبر بالتعلم بدون الحاجة للتدريب على كل مواقع وإزاحات الكائن [1].

استخدم الباحثون في [2] شبكة الكبسولة لدراسة وتحليل وتصنيف الأنواع الفرعية من خلايا الدم البيضاء ضمن خمسة أصناف مشكلة لمجموعة بيانات صغيرة. أظهر معدل الدقة العالي الذي حققه النموذج (96.86%) قدرة النظام على الاقتراب من مستوى أخصائي أمراض الدم بالرغم من محدودية عدد الصور المستخدمة. ناقش الباحثون في هذه الدراسة قدرات ومعدلات النجاح لشبكة الكبسولة في التدريب الناجح على قواعد بيانات صغيرة مقارنة مع طرق التعلم العميق الأخرى على نفس مجموعة التدريب، فأثبتت شبكة الكبسولة تفوقها في التخلص من مشكلة تعويم الذاكرة over-fitting (أي امتلائها نتيجة كمية البارامترات الكبير عند التدريب) مع معدل دقة عال، وكذلك عدم حاجتها إلى استخدام أي من تقنيات تعزيز البيانات أو المعالجة المسبقة في حالة مجموعة بيانات التدريب الصغيرة [2] .

عمل الباحثون في [3] على تدريب أربعة أنواع من الشبكات العصبونية ثلاثة منها شبكات CNN بينما تمثلت الرابعة بشبكة كبسولة بطبقات أقل وبنية أقل تعقيدا من شبكات CNNs. تم تقسيم قاعدة البيانات التي تحتوي على

مشاهد البيت الداخلية والمكونة من 20000 صورة إلى أربع أقسام أي 5000 صورة لكل نوع من أنواع الشبكات المقترحة، وقد توزعت الصور بالتساوي على خمسة أصناف. أشارت الاختبارات أن نتائج الاختبار على قاعدة بيانات الاختبار متقاربة بفارق 8٪ تقريبًا بين أفضل وأسوأ العروض. تم تدريب نفس الشبكات أيضًا على مجموعة بيانات ذات حجم أقل (أقل من 5000) لاختبار قدرتها على التعلم باستخدام مجموعات البيانات الأصغر. لوحظ انخفاض أداء الشبكات الثلاث القائمة على CNN بشكل كبير أي انخفاض معدل دقتها ، بينما احتفظت شبكة الكبسولة بنفس مستوى الأداء مما يثبت قدرة شبكة الكبسولة في إنجاز مهمتها حتى في حالة قواعد البيانات الصغيرة. إلا أنه عند تخفيض مجموعة البيانات إلى 1500 صورة ، فإن أداء شبكة الكبسولة انخفض إلى مستويات غير مقبولة ؛ هذا يعزى إلى انخفاض الحد الأدنى من المعلومات المطلوبة. بلغت دقة التحقق (على قاعدة بيانات التحقق validation dataset لشبكة الكبسولة 71٪ ودقة الاختبار على قاعدة بيانات الاختبار 70% [3].

اقترح الباحثون في [4] تقنية تسمى تصور مسار التوجيه لشبكات الكبسولة Routing Path Visualization والذي يكشف عن حدود المنطقة في الصورة التي يتم توجيهها إلى كل كبسولة. تم تدريب النموذج على قاعدة البيانات MNIST وهي قاعدة بيانات تم تشكيلها من دمج الأرقام المكتوبة يدويا من قبل طلاب المدرسة NIST الأمريكية وموظفو مكتب تعداد الولايات المتحدةUnited States Census Bureau وعلى قاعدة بيانات الصور الفلكية(Canada–France– Hawaii Telescope) CFHT. تشير النتائج التجريبية إلى إمكانية استخدام هذه التقنية " للحصول على موقع الصنف المتنبأ به في الصورة وتفسير الكيانات التي تحددها كبسولات الطبقات الوسيطة والنهائية وتحديد الحالات التي قد تفشل فيها الشبكة في تصنيف الكائن إلى صنفه الصحيح أو عدم قدرتها على تحديد موقع الصنف المتوقع من الصورة. يسمح تقاسم الوزن Weight sharing لشبكات الكبسولة على إنشاء عدد صغير من الأوزان ، مما يسهل تدريب شبكات الكبسولة على مجموعات بيانات صغيرة دون الوقوع في مشكلة تعويم الذاكرة Out of Memory (OOM) الناتج عن كمية البيانات الكبيرة من صور قاعدة البيانات الضخمة. يبين الجدول (1) دقة النظام المعتبر على قاعدة بيانات التحقق من الصحة Validation dataset وقاعدة بيانات الاختبار test dataset لكل من قواعد البيانات المستخدمة MNIST و CFHT [4]:

الجدول (1) دقة التحقق والاختبار للنظام المعتبر على قاعدتي البيانات MNIST و CFHT.

قاعدة البيانات	دقة التحققValidation accuracy	دقة الاختبار Testing accuracy
MNIST	92%	91%
CFHT	88%	92%

وجد الباحثون في [5] أن التعلم العميق التقليدي ينخفض أداؤه عندما تتغير العلاقة النسبية المكانية بين الكائنات المحددة، أي تنخفض قدرته على إنجاز المهمة التي يقوم بها سواء أكانت مهمة تحديد أو تصنيف أو تعرف، فعند غياب هذه العلاقة النسبية يمكن أن يقوم النظام بتصنيف الوجه المشوه (تغيير ملامحه بشكل يدوي) والتي تكون أجزائه بغير موقعها الأساسي على أنه وجه، وقد بدى ذلك من خلال قيمة معدل متوسط الدقة mean Average precision (mAP) المتخذ كمقياس في هذا المجال. لذلك، تم تطوير شبكة جديدة هي شبكة الكبسولة، يمكن لها أن تحقق معدل كشف عال مقارنة بالتقنيات المتقدمة الأخرى. تم الحصول على

أعلى معدل نجاح لشبكات الكبسولة مع مجموعة البيانات MNIST بالرغم من صغر حجمها. إذ أن قدرة الكبسولة على تعلم التغيرات المميزة يسمح لها باستنتاج المتغيرات المحتملة بفعالية وببيانات تدريب أقل. تبين أيضا أن خوارزمية التوجيه الديناميكي تمكنهم من التمييز بين الكائنات المتداخلة في الصور– كما سيتم عرض ذلك في وقت لاحق من هذا البحث– وهي الكائنات التي تقف أمام بعضها عند أخذ الصورة فقد لا يظهر من كائن سوى جزء منه. كذلك أثبتت شبكة الكبسولة قدرتها على الاحتفاظ بمعلومات مثل التجانس، والتدرج، والوضع، والملمس، والتشوه، والسرعة وموضع الكائن وهي قيم مكونات شعاع الخرج الناتج عن كبسولة التصنيف.

2- أهمية البحث وأهدافه:

ظهرت في السنوات الأخيرة العديد من النماذج المستخدمة في تحديد وتصنيف الكائنات ببنى شكلت نقلات نوعية في هذا المجال. حيث تدرجت هذه البنى من الشبكات العصبونية الالتفافية التقليدية CNN والشبكات الالتفافية الكاملة FCN ، وشبكات الكبسولة CapsNet والتي تعتبر البنية الأحدث الذي يحاول المطورون معرفة ما يمكن أن تقدمه في مجال رؤية الحاسب.

يهدف البحث الحالي إلى دراسة الفرق بين البنيتين المطورتين لأغراض التحديد والتصنيف. البنية الأولى هي نموذج الشبكة العصبونية الالتفافية التقليدية VGG16 التي تتكون من طبقات التفافية بعمق 16 طبقة لاستخلاص سمات أكثر تخصصا تحتوي معلومات أكثر فائدة، حيث أن الطبقات الالتفافية الضحلة غالبا ما تحتوي معلومات أقل عن الكائنات الموجودة في الصورة وكلما زادت الطبقات الالتفافية تصبح المعلومات أكثر تخصصا وفائدة، تنتهي هذه البنية بثلاث طبقات اتصال كامل والتي تمتاز بأن كل عصبون في الطبقة يرتبط مع كامل عصبونات الطبقة التالية. البنية الثانية هي النموذج الهجين المكون من شبكة الكبسولة مضاف لها الشبكة VGG16 المستخدمة لاستخلاص السمات من صور الدخل حيث ستتم مناقشة عمل كلتا الشبكتان من حيث استخلاص السمات والطريقة المتبعة في تحديد الكائنات، ومقارنة دقة التصنيف mean Average Precision mAP لكتا البنيتين وذلك لبيان آخر ما توصل له مجال الذكاء الصنعي من آليات في التحديد ليكون منطلقا وتطويرا للأبحاث التي ستتم في هذا المجال في كل من النقاط التالية:

1- فهم آلية عمل شبكة الكبسولة للتوصل لنظام يقوم بتحديد الكائنات وتصنيفها آليا باستخدام إحدى أنواع شبكات التعلم العميق.

2- تقليل كمية البيانات المستخدمة في عملية التدريب مقابل رفع دقة التحديد والتصنيف.

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

3- طرائق البحث ومواده:

هناك العديد من الطّرائق المقترحة لعملية تحديد وتصنيف الكائنات في صورة تمّ تنفيذها عمليا ببنى مختلفة خلال العقود الماضية باستخدام شبكات التعلم العميق، حققت هذه البنى نتائج جيدة من ناحية الدقة تارة وزادت من سرعة المعالجة على حساب الدقة تارة أخرى. اعتمدت الشبكات العصبونية الالتفافية التقليدية CNNs على المنهج الالتفافي في طبقاتها لاستخلاص السمات كما استخدمت في التصنيف طبقات الاتصال الكامل. بينما اعتمدت شبكات

الكبسولة على العديد من المفاهيم كانت أساسا في عملها كالتوجيه الديناميكي Dynamic Routing، الرؤية العكسية Inverse Graphics، نقل التعلم Transfer Learning.

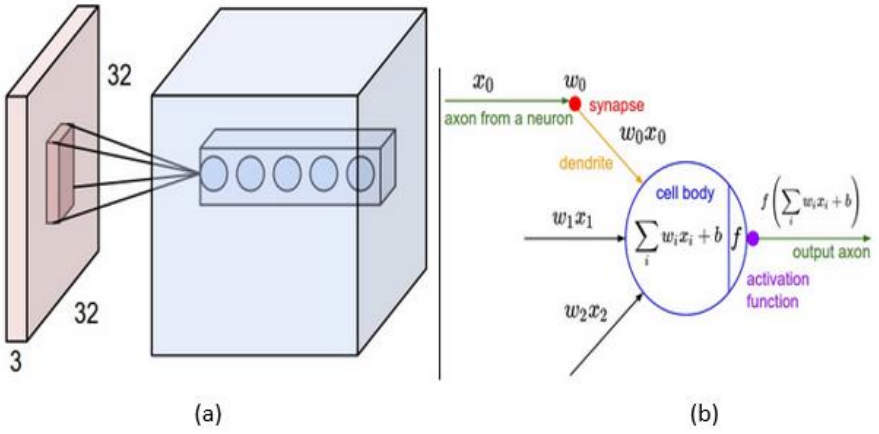
3-1 المنهج الالتفافي Convolutional Approch [13]:

تعتبر الطبقة الالتفافية حجر البناء الأساسي للشبكات العصبونية الالتفافية إذ إنها تقوم بمعظم العمليات الحسابية الأساسية. يتم تعريف الطبقة الالتفافية بحجم صورة الدخل ومجموعة من المرشحات المستخدمة في العملية الالتفافية القابلة للتعلم، تكون الأبعاد المكانية لكل مرشح (الطول والعرض) صغيرة، ويمتد هذا المرشح عبر كامل عمق حجم الدخل. مثلا: الحجم النموذجي لمرشح الطبقة الالتفافية الأولى للشبكة CNN يمتلك حجما 5x5x3 (أي 5 بكسل عرض و 5 بكسل طول و3 لقنوات اللون). خلال المسار الأمامي، يتم تمرير كل مرشح عبر العرض والطول والعمق لحجم الدخل وحساب الجداءات النقطية بين مداخل المرشح ودخل الطبقة. أثناء تمرير المرشح عبر العرض والطول لحجم الدخل سيتم انجاز خريطة التفعيل ثنائية البعد والتي تعطي استجابات هذا المرشح عند كل موقع مكاني. تقوم الشبكة بتعليم هذه المرشحات التي تنشَط عندما ترى بعض أنواع السمات المرئية كالحواف وغيرها. تمتلك كل طبقة التفافية مجموعة محددة من المرشحات (مثلا 12 مرشح)، يقوم كل مرشح بمفرده بتشكيل خريطة تفعيل activation map منفصلة ثنائية البعد ليتم في نهاية عمل الطبقة الالتفافية تكديس خرائط التفعيل (التنشيط) هذه على طول بُعد العمق (البعد الثالث للصورة) لينتج حجم الخرج وفقا للعمليات المنجزة عبر الطبقات الالتفافية وهو الحجم المكون من الأبعاد المكانية وبُعد العمق.

3-1-1ارتباطات العصبون [13]:

عند التعامل مع صور بأبعاد كبيرة على الدخل، سيكون من غير العملي ربط العصبون إلى كل العصبونات في الحجم السابق. سيتم بدلا عن ذلك ربط العصبون إلى منطقة محلية صغيرة من حجم الدخل كما هو موضح في الشكل (1-a). الحجم المكاني لهذا الاتصال هو برامتر يسمى بالحقل المستقبل للعصبون (يكافئ حجم المرشح). يمثل امتداد الاتصال على طول محور العمق ما يسمى عمق حجم الدخل.

يُظهر الشكل (1-a) مثالا يعبر عن صورة ملونة من قاعدة البيانات CIFAR-10 حجمها [32x32x3]، يتصل كل عصبون في الطبقة الالتفافية إلى منطقة محلية صغيرة فقط من الحجم المكاني للدخل، ولكن على امتداد بُعد العمق (أي كل قنوات الألوان). إذا كان الحقل المستقبل (أي حجم المرشح للطبقة الالتفافية) هو 5x5 عندئذ فإن أوزان كل عصبون في الطبقة الالتفافية الأولى ستكون إلى المنطقة ذات الأبعاد [5x5x3] من حجم صورة الدخل، أي : إجمالا 5*5*3 = 75+1 (الواحد المضاف هو قيمة برامتر الانحياز ويعتبر برامترا مضافا).مع ملاحظة أن مدى الاتصال على امتداد بُعد العمق يجب أن يكون 3، لأنه يعبر عن عمق حجم الدخل. أما الشكل (1-b) يوضح مفهوم العصبونات في الشبكات العصبونية العادية، حيث يتم تنفيذ عملية الضرب النقطي بين أوزانها والدخل لينتج عن ذلك قيمة يتم تمريرها إلى تابع التنشيط المعتبر للعصبون (cell body) ليكون خرج تابع التنشيط هو خرج ذلك العصبون [13].



الشكل (1) : (a) اتصال العصبون في الطبقة الالتفافية الأولى إلى جزء من حجم صورة الدخل، (b) آلية حساب خرج العصبون.

3-1-2 برامترات حجم الخرج المكاني للطبقات الالتفافية [13] :

تمت مناقشة الترتيب المكاني، وشرح الترابط لكل عصبون في الطبقة الالتفافية إلى حجم الدخل، لكن لم يتم التطرق بعد لمناقشة كم عدد العصبونات على حجم الخرج أو كيف يتم ترتيبهم. هناك ثلاثة بارامترات تتحكم بحجم الخرج: العمق depth، حجم الخطوة Stride، ومقدار الحشو الصفري zero-padding.

1- العمق: يمثل مجموعة العصبونات التي ترتبط إلى نفس المنطقة من الدخل. يعتبر العمق في حجم الخرج (أي الحجم الناتج عن طبقة الالتفافية والمكون من الأبعاد المكانية وبُعد العمق الناتج من تكديس خرائط السمات الثنائية والتي يتوافق عددها مع عدد المرشحات المستخدمة في الطبقة) برامتر يتوافق مع عدد المرشحات المستخدمة في الطبقة، والمرشحات هي مصفوفة صغيرة الحجم 2x2 أو 3X3 تعمل على البحث عن شيء مختلف في الدخل. عند إدخال الصورة الأصلية إلى الطبقة الالتفافية الأولى في الشبكة، عندئذ ستنشط العصبونات الممتدة عبر بُعد العمق لدى وجود حواف مختلفة أو مناطق لونية.

2- حجم الخطوة Stride: يجب تحديد حجم الخطوة التي سيخطو المرشح بها. فعندما تكون الخطوة stride=1، سيتم تحريك المرشح في كل مرة بمقدار بكسل واحد فقط. وعندما يكون حجم الخطوة stride=2 (أو بشكل غير مألوف يمكن أن تساوي 3) عندئذ سيخطو المرشح بكسلين في كل مرة، فينتج عن ذلك أحجام خرج أصغر مكانيا.

3- الحشو padding: يمثل الحشو بالأصفار برامتر يستخدم في بناء الشبكات العصبونية الالتفافية. قد يكون في بعض الأحيان من المناسب حشو حجم الدخل بالأصفار حول حدود الدخل، مما يسمح بالتحكم بالحجم المكاني لخرج الطبقات. غالبا ما يتم استخدامه للمحافظة على الحجم المكاني لأحجام المدخلات بحيث يكون عرض وطول المدخلات والمخرجات نفسه.

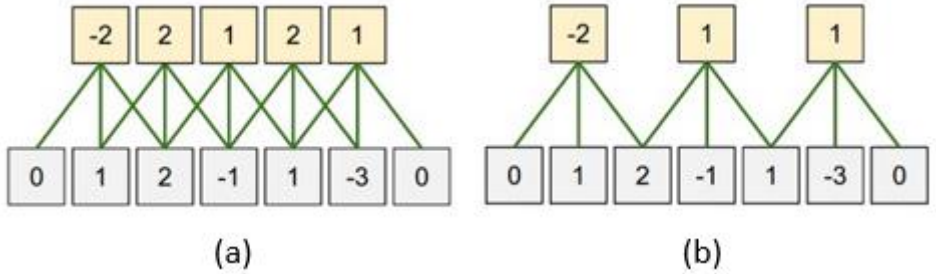
يمكن حساب حجم الخرج المكاني (خرج الطبقة الالتفافية) وفق العلاقة (1) [13]:

$$\text{حجم خرج الطبقة الالتفافية} = \frac{W-F+2P}{S} + 1 \qquad (1)$$

حيث أن حجم الخرج المكاني هو عبارة عن تابع لحجم الدخل (W)، وحجم الحقل المستقبِل لعصبونات الطبقة الالتفافية (F) "حجم المرشح"، وحجم الخطوة المقترح (S)، وكمية الحشو الصفري المستخدم (P) على الحدود.

فمثلا لو كان الحجم المكاني للدخل 7x7 وحجم الفلتر 3x3 وطول الخطوة S=1 ومقدار الحشو الصفري على حواف الصورة P=0 فإننا سنحصل على الحجم المكاني 5x5 على الخرج. بينما عندما يكون S=2 سنحصل على الحجم المكاني 3x3 على الخرج.

يُظهر الشكل (2) الترتيب المكاني، فلو فرضنا جدلا أنه لدينا بُعد مكاني واحد (x–axis) وحجم الحقل المستقبل F=3، وحجم الدخل W=5، وكمية الحشو الصفري P=1. يبين الشكل (a–2) أن حجم الخطوة التي تم تمرير الفلاتر بها S=1 فكان الحجم على الخرج $\frac{5-3+2}{1}+1=5$. بينما يبين الشكل (b–2) أنه تم استخدام حجم خطوة S=2 فكان الحجم على الخرج $\frac{5-3+2}{2}+1=3$. مع ملاحظة أنه لا يمكننا استخدام حجم خطوة S=3 فلو طبّقنا الحد الأول من الصيغة السابقة سنحصل على 4=(2+3–5) لا يمكن تقسيمة على 3. يوضح الشكل أيضا أوزان العصبونات وهي [1– ,0 ,1] وانحياز هذه العصبونات مساو للصفر.



الشكل (2) حجم الخرج المكاني طبقا لبارامترات حجم الخرج، (a) حجم الخطوة S=1، (b) حجم الخطوة S=2.[13]

يتضح من الشكل (a–2) أن بُعد الدخل 5 وبُعد الخرج أيضا مساويا 5. تم إنجاز ذلك لأن الحقول المستقبلة كانت 3 واستخدمنا الحشو الصفري بمقدار بكسل واحد على الحواف. إذا لم يتم القيام بحشو الحواف بالأصفار سيكون حجم الخرج المكاني 3 فقط، لأن ذلك هو عدد العصبونات المناسب المتوافق مع الدخل الأصلي كما هو موضح في الشكل (b–2). بشكل عام يتم تحديد مقدار الحشو الصفري *P* لكل بعد مكاني بالمعادلة (2)، حيث أن *F* حجم الحقل المستقبِل لعصبونات الطبقة الالتفافية "حجم المرشح" [13]:

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

عندما يكون حجم الخطوة S=1 عندئذ سنضمن أن حجم الخرج لن يتغير عن حجم الدخل المكاني.

3–2 شبكة الكبسولة Capsule Network:

3–2–1 مكونات شبكة الكبسولة: [9]

لفهم كيفية عمل شبكة الكبسولة ، يجب فهم تصميمها المعماري الموضح في الشكل (3) . تتكون شبكة الكبسولة من جزأين: وحدة التشفير ووحدة فك التشفير.

<u>الجزء الأول:</u> وحدة التشفير Encoder: تعمل على تحليل صورة الدخل بواسطة طبقاتها الثلاث، كما يبين الشكل (a–3):

1. الطبقة الالتفافية: تتمثل مهمة هذه الطبقة في تحديد السمات الأساسية من صورة الدخل باستخدام المنهج الالتفافي عبر الطبقات الالتفافية المشكلة لهذه الطبقة، وتشكيل خرائط السمات ثنائية البعد التي يتم تكديسها لتشكيل حجم الخرج Feature maps.

2. طبقة الكبسولات الأولية Primary Caps: (مستوى أدنى) تحتوي هذه الطبقة على عدة كبسولات أولية تتمثل مهمتها في أخذ السمات الأساسية التي تم تحديدها بواسطة الطبقة الالتفافية، وإعادة تشكيلها على شكل أشعة. حيث أن السمات الأساسية تكون ضمن خرائط السمات ثنائية البعد المكدسة ضمن بُعد العمق لحجم الخرج الناتج عن الطبقة الالتفافية، يتم تجميع عدد من خرائط سمات ثنائية البعد لتشكل كبسولة بحيث أن بُعدا الكبسولة يتطابقان الأبعاد مع المكانية لحجم الخرج، وعناصر الكبسولة هي أشعة وبُعد كل شعاع متطابق مع عدد خرائط السمات الثنائية التي شكلت الكبسولة.

3. طبقة class capsule: (مستوى أعلى) تحتوي هذه الطبقة على كبسولة يتطابق عدد أشعتها مع عدد الأصناف المكوِّنة لقاعدة البيانات، وبُعد كل شعاع يتوافق مع بُعد شعاع الخرج الذي يحتوي سمات الكائن المحدد (كل بعد من شعاع الخرج يمثل سمة معينة للكائن المعتبر (سماكة – استدارة...)، يبين الشكل (3–a) طبقة كبسولة الصنف المكونة من عدد من الأشعة يتطابق مع عدد الأصناف، وبُعد كل شعاع 16 الذي يمثل بعد الكبسولة.

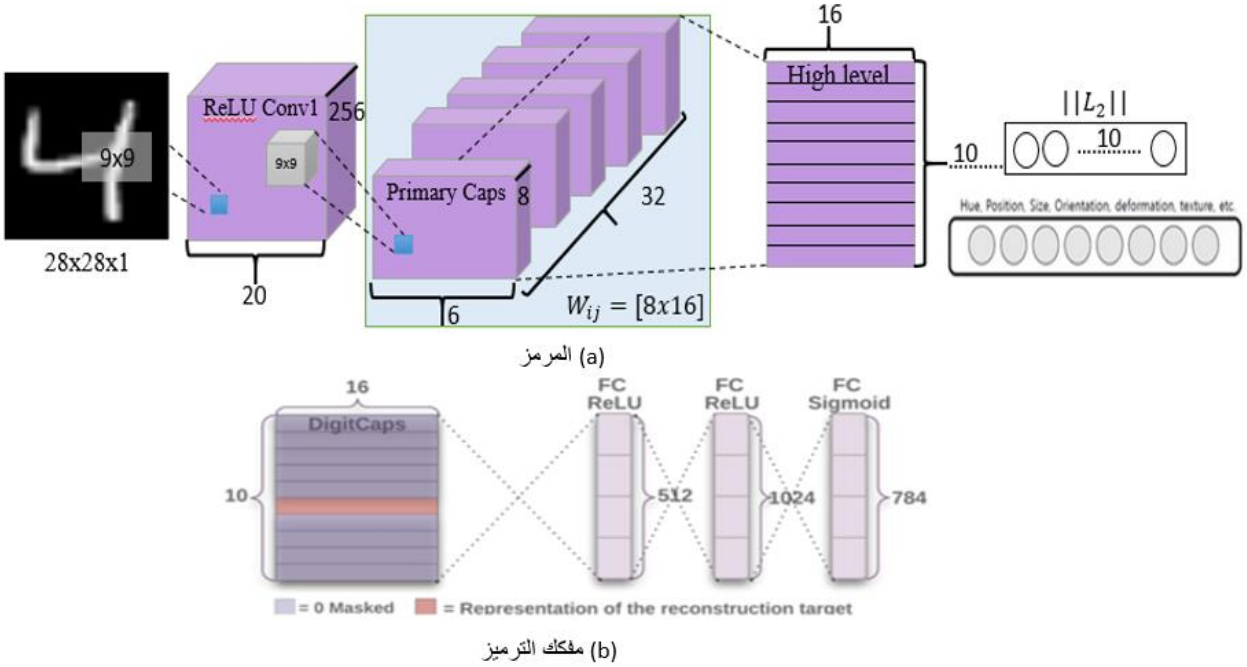
<u>الجزء الثاني:</u> وحدة فك التشفير Decoder: يأخذ مفكك التشفير Decoder شعاع الخرج الفعال المقترح من خوارزمية التوجيه الديناميكي والمنفذة ضمن الطبقة class capsule، لإعادة بناء الصورة الأصلية. الشكل (3–b):

4. طبقة الاتصال الكامل الأولى التي تحوي 512 خلية عصبونية، يتم تحديث أوزان كل خرج من المستوى الأدنى وتوجيهه إلى كل خلية عصبونية للطبقة المتصلة بالكامل كمدخلات. كل خلية عصبونية لها أيضًا انحياز.

5. طبقة الاتصال الكامل الثانية والتي تحوي 1024 خلية عصبونية.

6. طبقة الاتصال الكامل الثالثة والتي تحوي 784 أي (28x28) خلية عصبونية.

تختلف طبقات الاتصال الكامل بالحجوم ولكنها تعمل نفس العمل وتهدف لإعادة بناء صورة الدخل.

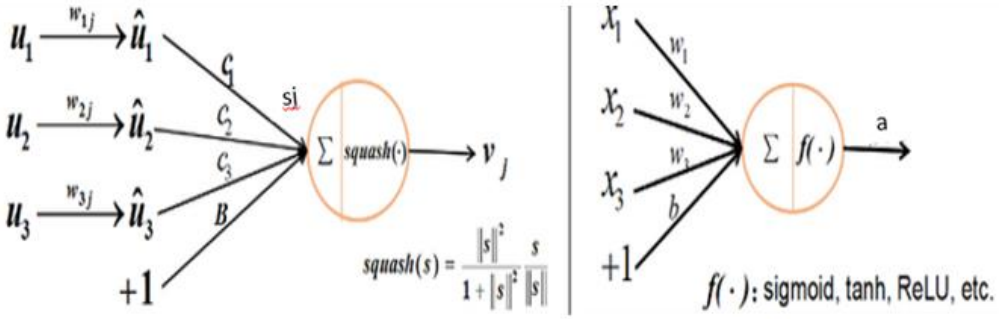


الشكل (3) مكونات شبكة الكبسولة بشكل عام (a) المرمز، (b) مفكك الترميز.

3-2-2 منهجية عمل شبكة الكبسولة:

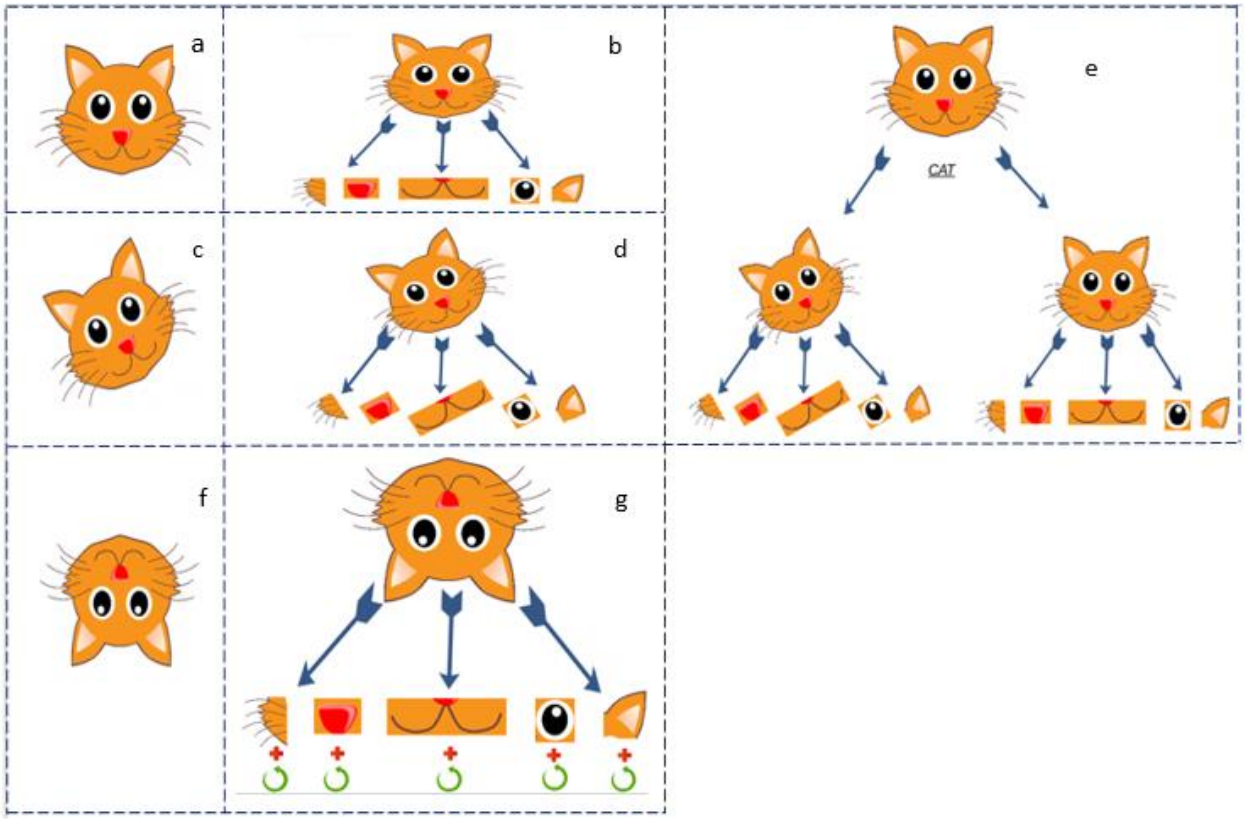
يستقبل العصبون في الشبكات العصبونية القيمة العددية x من العصبونات السابقة ويقوم بجدائها بالوزن المقابل w للحصول على قيمة عددية، يتم إدخالها إلى إحدى توابع التنشيط اللاخطية (sigmoid, tanh ReLU), لخلق قيمة الخرج كقيمة دخل للعصبون التالي كما هو موضح في الشكل (4-b) [15] . ظهرت لهذه الشبكات العديد من العيوب كفقدان المعلومات المهمة الناتج عن طبقات التجميع، وكذلك فهي لا تأخذ بعين الاعتبار العلاقات المكانية بين أجزاء الصورة تلك العلاقات الهامة في التعرف على الهوية، وحساسية هذا النوع من الشبكات للصورة الأصلية نفسها من أجل تصنيف الصور على أنها نفس الصنف، [1] لذا لجأ الباحثون إلى ابتكار شبكة الكبسولة.

تُوصف الكبسولة بأنها مجموعة من الخلايا العصبونية التي تخزن على شكل شعاع معلومات الشبكة وسماتها المختلفة المعبرة عن الكائن المحدد في الصورة الناتجة عن العملية الالتفافية في الطبقة الالتفافية إضافة إلى البارامترات المتعلقة بإنشاء مثيل لسمة من السمات المذكورة، ويعتبر تخزين السمات على شكل شعاع وليس ضمن قيم عددية مفردة من أبرز ما تمتاز به شبكة الكبسولة. علما أن الكبسولة قادرة على ترميز ليس فقط تواجد سمات مرئية لعينة وإنما ترميز التحويلات (إزاحة – دوران) لهذه السمة مما يجعلها حساسة لتحولات الكائن في الصورة. يشير طول الشعاع في النموذج إلى قيمة احتمالية تواجد الكائن في الصورة والذي يحدد من تابع squash من خوارزمية التوجيه الديناميكي، أما اتجاه هذا الشعاع فهو يمثل وضعية المعلومات attitude information، حيث أن هذه المعلومات (الموضع والمقياس والدوران ...) هي التي تحدد اتجاه الشعاع [16].



الشكل (4): (a) آلية حساب خرج الكبسولة j (شعاع) في شبكة الكبسولة، (b) آلية حساب خرج العصبون (قيمة عددية) في الشبكات العصبونية.

لبيان آلية التفكير التي تتبعه شبكة الكبسولة يوضح الشكل (5–a) وجه قطة، تسعى شبكة الكبسولة لاكتشاف أن ما يوجد في الصورة هو وجه قطة بتقسيمها لسمات فردية مثل العينين والأنف والأذن وما هنالك كما هو موضح في (5–b)، أي تحليل السمات عالية المستوى إلى سمات منخفضة المستوى، وكذلك يمكن التحليل إلى المزيد من السمات منخفضة المستوى كالأشكال والحواف.



الشكل (5): (a) وجه قطة، (b) تقسيم وجه القطة لسمات فردية، (c) وجه قطة مدار بزاوية 30 درجة، (d) سمات وجه القطة المدار، (f) وجه القطة المقلوب، (g) سمات وجه القطة المقلوب مع إضافة سمة الدوران لسمات الوجه، (e) أداة التعرف على وجه قطة من وجهة نظر كبسولة.

في حال كان وجه القطة الذي تم تدريب الشبكة عليه مدارا بزاوية مقدراها 30 درجة مثلا كما في الشكل (5–c) أو كحالة خاصة أن يكون الوجه مقلوبا كما في (5–f) فلن نتمكن من خلال السمات التي تم تحديدها مسبقا من قبل

الطبقات الالتفافية في الشبكة والموضحة في الشكل (5–b) من تصنيف الوجه كوجه قطة – وهو الضعف الذي تعاني منه الشبكات العصبونية التقليدية CNN، عندئذ كان الحل في جعل الكبسولة تعمل على تضمين خصائص إضافية مثل زاوية الدوران إضافة إلى سمات المستوى المنخفض المعتبرة في CNNs والمحددة مسبقا كما في الشكل (5–g) مما يجعلها قادرة على تصنيف الوجه مهما بلغت درجة استدارته في الصور الحاوية عليه – أي تمكنت الشبكة من تحديد السمة وزاوية دورانها– حتى وإن لم تتدرب على جميع هذا الزوايا.

مكنت هذه الميزة شبكة الكبسولة من التدريب على قواعد بيانات أصغر مما هو عليه في الشبكات العصبونية الالتفافية التقليدية CNN. وبالتالي من المحتمل أن أداة التعرف على القطط بشكل عام تبدو كما هو موضح في الشكل (5–e). [17]إضافة إلى ذلك تستبدل شبكة الكبسولة عملية التجميع الأعظمي المستخدمة بين الطبقات الالتفافية في شبكات CNN والتي تؤدي لفقدان الكثير من البيانات الهامة بخوارزمية التوجيه الديناميكي، مما يسمح بتكرار ما تم تعلمه عبر فضاء الصور حتى التي لم يتم التدرب عليها ضمن شبكة الكبسولة [16] .

يلخص الجدول (2) الاختلافات بين الكبسولة والعصبون التقليدي [9] :

الجدول(2) الاختلافات الوظيفية بين العصبون والكبسولة.

الفرق بين الكبسولة والعصبون التقليدي			
الدخل من المستوى المنخفض كبسولة / عصبون		الشعاع (u_i)	القيمة العددية (x_i)
العملية	التحويل اللاخطي Affine	$\hat{u}_{j\|i} = w_{ij}u_i + B_j$	–
	الوزن	$s_j = \sum_i c_{ij}\hat{u}_{j\|i}$	$a_j = \sum_i w_i x_i + b$
	المجموع		
	التنشيط اللاخطي	$v_j = \frac{\|\|s_j\|\|^2}{1+\|\|s_j\|\|^2}\frac{s_j}{\|\|s_j\|\|}$ v_j = squash (s_j)	$h_j = f(a_j)$
الخرج		الشعاع (v_j)	القيمة العددية (h_j)

حيث أن c_{ij} : هي معاملات مزدوجة coupling coefficients يتم تحديدها من خلال تكرارات خوارزمية التوجيه الديناميكي[9] ، s_j: شعاع الدخل الموزون بمعاملات التوجيه c_{ij} للكبسولة j، v_j: شعاع خرج الكبسولة j، w_{ij}: مصفوفة الأوزان الناتجة عن خوارزمية التوجيه الديناميكي، B_{ij} قيمة مؤقتة يتم تهيئتها إلى الصفر في بداية التدريب وسيتم تحديثها بشكل متكرر، $\widehat{u_{j|i}}$ مجموعة خطية من أشعة الدخل للكبسولة j، a_i الجمع الموزون للعصبون، وb هو الانحياز للعصبون، h_j خرج العصبون وهو ناتج تابع التنشيط لهذا العصبون.

كما ذكرنا بأن طول شعاع خرج الكبسولة v_j يمثل احتمال أن الكائن المعتبر موجود. لذا فقد تم استخدام تابع التنشيط اللاخطي Squashing لضمان أن الأشعة القصيرة تميل لأن تكون بطول صفري " أي لا يوجد

كائن" والأشعة الطويلة تميل لأن تكون بطول الواحدة " أي يوجد كائن".في المعادلة (3) التعبير الرياضي لتابع squash:[9]

$$v_j = \frac{||s_j||^2}{1+||s_j||^2} \frac{s_j}{||s_j||} \quad (3)$$

يبين الشكل (4–a) مخطط يوضح الآلية المقابلة لحساب شعاع خرج الكبسولة، حيث أن s_j هو الدخل الكلي للكبسولة **j**. يتم حساب الدخل الكلي s_j من خلال إجراء الجمع الموزون لأشعة التنبؤ $\hat{u}_{j|i}$ والتي تقابل مفاهيميا خرج العصبون التقليدي والتي تحتوي على المعلومات حول الكائن والناتجة عن جداء شعاع الخرج للكبسولة i في المستوى الأدنى u_i مع مصفوفة الوزن w_{ij} وفق المعادلتين (4)، (5) والتي تساعد في إتمام المهمة المطلوبة من الشبكة بما تحتويه من معلومات عن الكائن الموجود في الصورة: [9]

$$s_j = \sum_i c_{ij} \hat{u}_{j|i} \quad (4)$$

$$\hat{u}_{j|i} = w_{ij} u_i + B_{ij} \quad (5)$$

3–3 التوجيه الديناميكي Dynamic routing:[9]

تحتاج الكبسولة في طبقة الكبسولات الأولية (المستوى الأدنى) لمعرفة إلى أي كبسولة من كبسولات المستوى الأعلى سيتم إرسال شعاعها –بمعنى أي كبسولة في المستوى الأعلى تتوافق معها. يتم اتخاذ القرار عن طريق تغيير الوزن القياسي cij الذي إما سيضاعف خرج الكبسولة (الشعاع) وبالتالي فإن احتمال هذا الشعاع الذي ينتجه تابع Squash سيكون أعلى من احتمال باقي الأشعة الواردة من الكبسولات الأخرى فيتم التعامل معه كدخل لكبسولة ذات مستوى أعلى أو بالعكس. ويتم ذلك باتباع خوارزمية التوجيه الديناميكي. حيث أن جوهر خوارزمية التوجيه الديناميكي هو

"**سوف ترسل الكبسولة ذات المستوى الأدنى مدخلاتها إلى الكبسولة ذات المستوى الأعلى والتي "تتوافق" مع مدخلاتها**".

قبل الخوض في معرفة آلية عمل هذه الخوارزمية يجب أن نعلم أن الأوزان المحددة من خلال خوارزمية التوجيه الديناميكي تحقق ما يلي:

1. كل وزن cij هو عدد غير سالب.
2. لكل كبسولة منخفضة المستوى i ، مجموع كل الأوزان cij يساوي 1.
3. لكل كبسولة منخفضة المستوى i ، عدد أوزان يساوي عدد الكبسولات ذات المستوى الأعلى.
4. يتم تحديد هذه الأوزان بواسطة خوارزمية التوجيه الديناميكي التكراري iterative dynamic routing.

توضح التعليمات التالية المكتوبة باللغة الزائفة خطوات عمل خوارزمية التوجيه الديناميكي، أي آلية عمل المرور الأمامي للشبكة: [9]

1: **procedure** ROUTING($\hat{u}_{j|i}, r, l$)
2: for all capsule i in layer l and capsule j in layer $(l+1)$: $b_{ij} \leftarrow 0$.
3: **for** r iterations **do**
4: for all capsule i in layer l: $\mathbf{c}_i \leftarrow \mathtt{softmax}(\mathbf{b}_i)$ ▷ softmax computes
5: for all capsule j in layer $(l+1)$: $\mathbf{s}_j \leftarrow \sum_i c_{ij}\hat{\mathbf{u}}_{j|i}$
6: for all capsule j in layer $(l+1)$: $\mathbf{v}_j \leftarrow \mathtt{squash}(\mathbf{s}_j)$ ▷ squash computes
7: for all capsule i in layer l and capsule j in layer $(l+1)$: $b_{ij} \leftarrow b_{ij} + \hat{\mathbf{u}}_{j|i}.\mathbf{v}_j$
return $\mathbf{v}_j$

السطر الأول: يشير إلى اعتبار جميع الكبسولات في مستوى أدنى *l* ومخرجاتها $\hat{u}_{ij}$ بالإضافة إلى عدد مرات تكرار التوجيه *r*. ويتوضح عمل هذه البرامترات في الأسطر التالية.

السطر الثاني: المعامل b_{ij} هو قيمة مؤقتة يتم تهيئتها إلى الصفر في بداية التدريب وسيتم تحديثها بشكل متكرر، وبعد انتهاء الإجراء ، سيتم تخزين قيمتها في c_{ij}.

السطر الثالث: يبين أن الخطوات من 4–7 سيتم تكرارها *r* مرة (عدد تكرارات التوجيه).

السطر الرابع: يحسب قيمة الشعاع ci وهي جميع أوزان التوجيه لكبسولة المستوى الأدنى i. يتم ذلك لجميع كبسولات المستوى الأدنى. سيعمل تابع التنشيط Softmax على التأكد من أن كل وزن cij هو رقم غير سالب وأن مجموعها يساوي 1،أي سيفرض softmax الطبيعة الاحتمالية للمعاملات cij المذكورة أعلاه.

في التكرار الأول ، ستكون قيمة جميع المعاملات c_{ij} متساوية، وذلك لأنه يتم في السطر الثاني ضبط قيم كل البارامترات bij إلى الصفر. على سبيل المثال ، إذا كان لدينا 3 كبسولات منخفضة المستوى وكبسولتان بمستوى أعلى ، فستكون كل cij تساوي 0.5. تشير حالة أن قيم cij متساوية –عند تهيئة الخوارزمية – حالة الحد الأقصى من عدم الإدراك: أي لا تملك الكبسولات ذات المستوى الأدنى أي فكرة عن الكبسولات ذات المستوى الأعلى التي تناسب خرجها، ولكن مع تكرار العملية، ستتغير هذه التوزيعات المنتظمة. بعد حساب جميع الأوزان cij لجميع الكبسولات ذات المستوى الأدنى ، يمكننا الانتقال إلى السطر الخامس.

السطر الخامس: يتعامل مع كبسولات المستوى الأعلى. تحسب هذه الخطوة مجموعة خطية من أشعة الدخل، الموزونة بمعاملات التوجيه cij، المحددة في الخطوة السابقة. مما ينتج شعاع خرج sj، ويتم ذلك لجميع الكبسولات ذات المستوى الأعلى.

السطر السادس: بعد حساب أشعة الخرج من الخطوة الأخيرة يتم تمريرها عبر تابع Squash اللاخطي، وهذا يضمن الحفاظ على اتجاه الشعاع، ولكن يتم فرض طوله بحيث لا يزيد عن 1. هذه الخطوة تنتج شعاع الخرج vj لجميع مستويات الكبسولات الأعلى.

السطر السابع: يتم فيه تحديث الوزن.

حتى الآن فإن الخطوات 4–6 ببساطة تحسب ناتج الكبسولات ذات المستوى الأعلى. الخطوة في السطر 7 هي المكان الذي يتم فيه تحديث الوزن. تلتقط هذه الخطوة جوهر خوارزمية التوجيه. تبحث هذه الخطوات في كل كبسولة مستوى أعلى j ثم تفحص كل دخل وتقوم بتحديث الوزن المقابل bij وفقًا للصيغة المذكورة. تنص المعادلة على أن قيمة الوزن الجديدة تساوي القيمة القديمة بالإضافة إلى الجداء النقطي للخرج الحالي للكبسولة j والمدخلات إلى هذه الكبسولة من كبسولة المستوى الأدنى i. يوافق الجداء النقطي بين تشابه الدخل إلى الكبسولة وخرجها. وبما أن كبسولة المستوى الأدنى سترسل ناتجها إلى كبسولة المستوى الأعلى التي

يكون ناتجها مشابهًا. يتم التقاط هذا التشابه بواسطة الجداء النقطي dot product. بعد هذه الخطوة ، تبدأ الخوارزمية من الخطوة 3 وتكرر العملية r مرة.

السطر الأخير: يبين أن الخوارزمية ستعطي ناتج الكبسولة j ذات المستوى الأعلى vj.

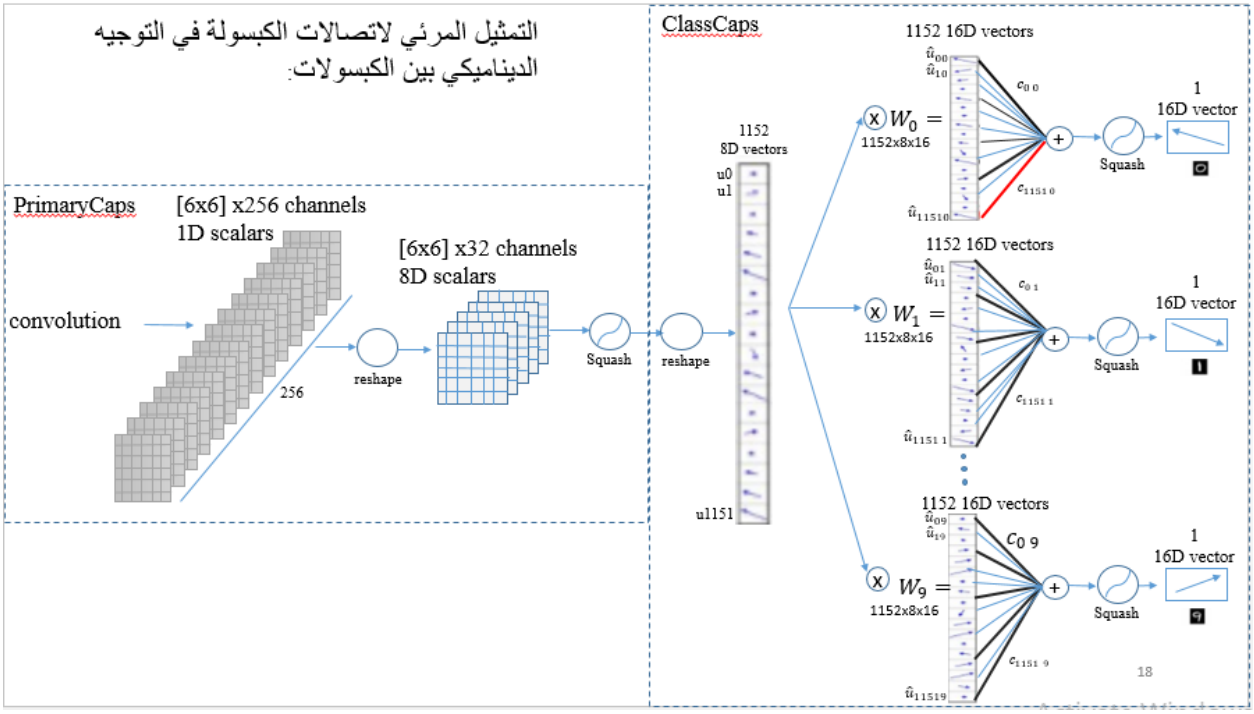
بعد r مرة، تم حساب جميع النواتج للكبسولات ذات المستوى الأعلى وتم إنشاء أوزان التوجيه. يمكن أن يستمر التمرير الأمامي إلى المستوى التالي من الشبكة. يبين الشكل (6)التمثيل المرئي لاتصالات الكبسولة في التوجيه الديناميكي بين الكبسولات.

3-4 الرؤية العكسية Inverse Graphics [16]:

في مجال الرسوم الحاسوبية، يتم مراعاة التمثيلات الداخلية المختلفة لكائن ما مثل موضعه وتدويره ومقياسه وتحويلها إلى صورة على الشاشة. يعمل دماغنا في الاتجاه المعاكس لهذا النهج، يسمى الرسوم العكسية inverse graphics. عندما ننظر إلى أي شيء، فإننا نقوم بتفكيكه داخليا إلى أجزاء فرعية هرمية مختلفة، ونميل إلى تطوير علاقة الأجزاء الداخلية من الجسم بأكمله. هذه هي الطريقة التي نتعرف بها على الأشياء ، يعتبر هذا المفهوم اللبنة الأساسية لبناء شبكات الكبسولات. إذ توفر شبكات الكبسولة فرصة للاستفادة الكاملة من العلاقات المكانية الهرمية ومحاكاة القدرة على فهم التغييرات في الصورة. بالنسبة للكبسولات في المستوى الأدنى، تكون معلومات الموقع بواسطة الكبسولة النشطة (الكبسولة التي يمتلك شعاعها مقدار الحتمال الأكبر). وكلما تم الارتفاع بمستوى الكبسولات، سيكون هنالك معلومات موضعية أكثر ضمن مكونات القيم الفعلية لشعاع خرج الكبسولة. هذا يعني أنه كلما ارتقينا بالتسلسل الهرمي يجب أن تزيد أبعاد الكبسولة.

3-5 نقل التعلم Transfer Learning [14]:

تعد عملية نقل التعلم transfer learning طريقة شائعة في مجال رؤية الحاسب، يتم التعبير عنها عادة من خلال استخدام نماذج شبكات عصبونية التفافية مدربة مسبقًا على مجموعة بيانات مرجعية كبيرة لحل مشكلة مشابهة لتلك التي نريد حلها، فنتجاوز بذلك التكلفة الحسابية والزمن الطويل لتدريب النماذج الحديثة. من الأمثلة على هذه النماذج المدربة مسبقا VGG و Inception و MobileNet.



الشكل (6) التمثيل المرئي لاتصالات الكبسولة في التوجيه الديناميكي بين الكبسولات.

تحتوي الشبكة العصبونية الالتفافية النموذجية CNN على جزأين، كما هو موضح في الشكل (7):

1- القاعدة الالتفافية convolutional base: والتي تتكدس فيها مجموعة من الطبقات الالتفافية تتخللها عدة طبقات تجميع pooling، ويكون عملها الأساسي استخلاص السمات من الصورة الدخل.

2- المصنف classifier: وهو سلسة من الطبقات المتصلة بشكل كامل Fully connected Layer. الهدف الرئيسي منه تصنيف الصورة بناءً على السمات المستخلصة.

الشكل(7): بنية نموذج الشبكة العصبونية الالتفافية النموذجي.[14]

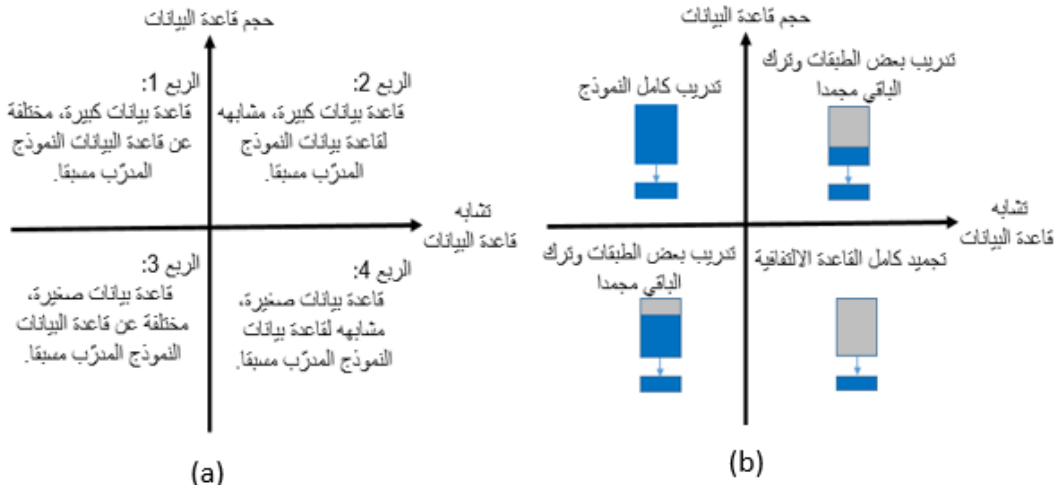
أحد الجوانب المهمة في نماذج التعلم العميق قدرتها على التعلم التلقائي للسمات هرميا. أي أن السمات المحسوبة من قبل الطبقة الأولى عامة (الملامح العامة في الصورة) ويمكن إعادة استخدامها لحل مشكلات مختلفة ، في حين أن السمات المحسوبة من قبل الطبقة الأخيرة محددة (سمات متخصصة) وتعتمد على مجموعة البيانات والمهمة المختارة.

عندما يتم استيراد نموذج شبكة عصبونية التفافية تم تدريبه مسبقًا ليُستخدم في حل مشكلة أخرى، يجب ضبط آلية التدريب للنموذج الجديد وفقًا لإحدى الاستراتيجيات الثلاث، كماه موضح في الشكل (8-ب):

1–تدريب النموذج بأكمله.

2– تدريب بعض الطبقات الالتفافية (سفلية) إضافة للمصنف ، وترك الطبقات الالتفافية العليا مجمدة.

3–تجميد القاعدة الالتفافية.



الشكل(8): مصفوفة حجم التشابه التي تساعد في تحديد استراتيجية بناء نموذج شبكة عصبونية التفافية جديد بالاعتماد على مفهوم نقل التعلم [14] .

بالاعتماد على نقل التعلم يتم تنفيذ النموذج الجديد بتحديد نموذج الشبكة العصبونية الالتفافية المدرب مسبقا على نفس منصة العمل التي سنقوم بتدريب نموذجنا الجديد عليها، فمثلا لو كانت منصة العمل Keras عندئذ يمكن اختيار النماذج التالية : VGG (2014) و InceptionV3 (2015) و ResNet5 (2015). ثم وبالاعتماد على مصفوفة حجم التشابهSize–Similarity Matrix الموضحة في الشكل (8–a) يقوم النظام بإنجاز مهمة الرؤية الحاسوبية قيد المعالجة آخذا بعين الاعتبار حجم قاعدة البيانات وتشابهها مع قاعدة البيانات التي تم تدريب النموذج المختار عليها. على سبيل المثال ، إذا كانت المهمة المعتبرة هي تحديد القطط والكلاب في الصورة وتصنيفها، فيمكن اعتبار أن ImageNet مجموعة بيانات مماثلة لأنها تحتوي على صور للقطط والكلاب، في حين إذا كانت المهمة هي تحديد الخلايا السرطانية ، فلا يمكن اعتبار ImageNet مجموعة بيانات مماثلة. بعد ذلك، يتم بناء النموذج الجديد والاعتماد على مصفوفة حجم التشابه الموضحة في الشكل (8–b) والتي نستخلص منها آلية التدريب للنموذج الجديد وفقًا لإحدى الاستراتيجيات الثلاث المذكورة آنفا [14] . خلال نقل التدريب، يتم نسخ الأوزان من الطبقات الأولى للنموذج المدرب مسبقا إلى النموذج الجديد والتي تتضمن معلومات حول السمات الأساسية الموجودة في الكائنات مثل اللون، الشكل، الحواف، الخطوط.أما طبقة التصنيف الأخيرة، المسؤولة عن التصنيف عالي المستوى للكائن ضمن مجموعات الأصناف فلا يمكن لها أن تُنقل. يتم تدريب النموذج الجديد لاحقا لمهمة تحديد وتصنيف الأصناف الجديدة.

4– النتائج العملية:

تمت دراسة بنية الشبكات العصبونية الالتفافية التقليدية CNN لتحديد وتصنيف الكائنات في الصور، و بنية شبكة الكبسولة Capsule net. حيث سيتم دراسة الفروقات البنيوية لكلتا الشبكتان والمنهج المعتمد في العمل، ومن ثم سيتم عرض أهم النقاط السلبية في الشبكات CNN والتي عملت شبكة الكبسولة على تلافيها في بنيتها الجديدة.

سنتخذ مثالا عن شبكات CNN النموذج VGG16 و عن شبكات الكبسولة نموذجا يتخذ VGG16 شبكة أساسية له لاستخلاص السمات.

4-1 بنية النموذج VGG16 [7] [6] :

يستقبل نموذج الشبكة خلال التدريب كدخل صورة ملونة بحجم مكاني ثابت مثلا 224 × 224. حيث تتم معالجة حجوم الصور في قاعدة البيانات المستخدمة لتصبح بحجم واحد، كما تم العمل على طرح متوسط قيمة RGB المحسوبة على مجموعة التدريب من كل بكسل في ضوء المعالجة المسبقة لصور قاعدة البيانات.

يبين الشكل (9) بنية الشبكة VGG16، حيث يتم تمرير الصورة عبر مجموعة من الطبقات الالتفافية التي تستخدم مرشحات بحقل استقبال صغير جدا بحجم 3x3 (وهو أصغر حجم لالتقاط فكرة اليسار / اليمين ، أعلى / أسفل ، مركز). حجم خطوة الالتفاف Stride ثابتة تساوي بكسلا واحدا، كما أن الحشو المكاني Padding يبلغ بكسلا واحدا. تم إجراء التجميع المكاني Spatial pooling باستخدام خمس طبقات تجميع أعظمي max-pooling، والتي تلي بعض الطبقات الالتفافية (لا تلي كل الطبقات الالتفافية طبقات تجميع). يتم إنجاز التجميع عبر نافذة بحجم 2x2 بكسل بحجم خطوة S=2. تلي مجموعة الطبقات الالتفافية -والتي يختلف عددها في الشبكات العصبونية الالتفافية بحسب النموذج المقترح -هنا لدينا 13 طبقة التفافية، ثلاث طبقات اتصال كامل Fully Connected (FC): تمتلك أول طبقتين 4096 قناة، بينما تمتلك الطبقة الثالثة 1000 قناة بحيث تقابل كل قناة صنفا من أصناف مجموعة البيانات المعتبرة ImageNet. تنتهي هذه الهيكلية بطبقة softmax. علما أن كل الطبقات المخفية تستخدم تابع التنشيط اللاخطي ReLU.

ويتضح من الشكل (9) أيضا أن عدد قنوات الطبقات الالتفافية يبدأ 64 قناة في الطبقة الالتفافية الأولى ويزداد بمقدار الضعف بعد كل طبقة تجميع أعظمي حتى يصل عددها إلى 512. كذلك الأمر تمتاز الشبكة ببنية متجانسة للغاية بخلاف النماذج الأخرى من شبكات CNN، حيث أن حجم المرشحات لجميع الطبقات الالتفافية واحد وهو3x3 وحجم مرشحات طبقات الاتصال الكامل 2x2 من البداية حتى النهاية.

يمكن حساب عدد البارامترات وحجم الحسابات لهذا النموذج وفق المعادلتين (6) (7): [6]

$$\#\ parameter = kernel\_size^2 \times channel_{in} \times channel_{out} \qquad (6)$$

$$\#\ computation = \#\ parameter \times X \times Y \times batch\_size \qquad (7)$$

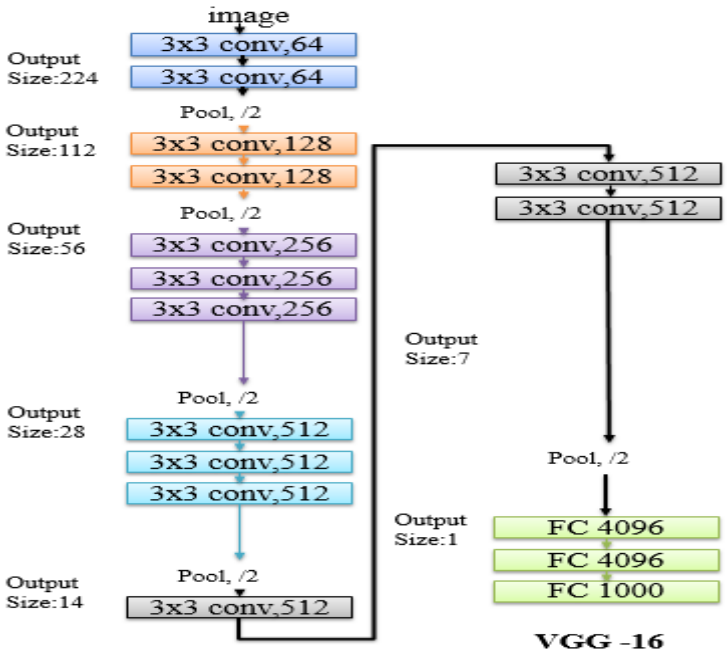
حيث أن Kernel_size تمثل حجم النواة لمرشح الأوزان، $channel_{in}$ عدد قنوات الدخل في كل طبقة التفافية، $channel_{out}$ عدد قنوات الخرج لكل طبقة التفافية، X، Y الأبعاد الأفقية والرأسية لخرائط السمات في كل طبقة التفاف، و batch_size حجم الدفعة والذي يشير لعدد الصور في كل تكرار.

من المعادلة (7) يتضح أن حجم الحسابات يتناسب طردا مع عدد البارامترات في الشبكة وبالتالي فإن تقليل عدد البارامترات في كل من الطبقات الالتفافية وطبقات الاتصال الكامل يمكن أن يقلل حجم الحسابات.

4-2 بنية نموذج شبكة الكبسولة Capsnet [1]:

يبين الشكل (10) هيكلية نموذج شبكة الكبسولة المعتبر، يأخذ النموذج كامل الصورة كدخل ليعيد تشكيلها إلى حجم جديد مثلا 300x300 كخطوة معالجة مسبقة. تحتاج شبكة الكبسولة إلى طبقات التفافية

تقوم باستخلاص خرائط السمات وذلك نظرا لما تبديه هذه الشبكة من ضعف في استخلاص السمات الأساسية في الصورة. تعتبر الشبكة VGG16 والمدربة بشكل مسبق على قاعدة بيانات ImageNet شبكة التفافية مضافة إلى نموذج شبكة الكبسولة بطريقة نقل التعلم Transfer learning كشبكة أساسية تعمل على استخلاص السمات من الصورة وتشكيل خرائط السمات يلي ذلك، طبقة الكبسولات الأولية primary capsules التي تعمل على إعادة تشكيل الخرج (حجم الخرج الناتج عن عملية استخلاص السمات من الشبكة الأساسية) على شكل أشعة بثمانية أبعاد 8D (أي من بُعد العمق لحجم الخرج والمكون من عدة خرائط سمات مكدسة إلى بعضها البعض، يتم جعل كل 8 خرائط سمات كبسولة عناصرها أشعة وكل شعاع هو تتالي المواقع المتقابلة للأبعاد المكانية في خرائط السمات هذه)، تمرر هذه الأشعة إلى تابع squash اللاخطي المبين بالمعادلة (3) والذي ينتج عنه خرج الكبسولات الأساسية. تلي هذه الطبقة طبقة caps pooling التي تنفذ خوارزمية التوجيه الديناميكي بين الكبسولات والتي تضمن ثبات الشبكة أمام تغيرات السمة (دوران – إضاءة – إزاحة وغير ذلك) تساعد هذه الطبقة على تدريب النموذج المعتبر، إذ تستخلص الطبقة caps pooling سمات الكائن من خرائط السمات وتعطي على الخرج أشعة سمات تمتلك معلومات مفيدة للطبقة التالية. يتم ضبط النموذج بإضافة طبقتي كبسولة ذات اتصال كامل وتوابع تنشيط لا خطية ReLU وتتجه أشعة السمات في هذا النموذج إلى فرعين: أحدهما يعطي احتمال سيغموند sigmoid (أي قيمة احتمال الصنف الناتجة عن تابع التنشيط sigmoid المستخدم في حالات التصنيف الثنائي) لصنف الكائن، والآخر يعطي أربعة نقاط تحدد موقع المربعات المحيطية التي ستحيط بالكائن المحدد بالصورة.



الشكل (9): بنية النموذج VGG16.

4-3 تدريب النماذج:

تم تدريب كلا النموذجين باستخدام قاعدة البيانات ImageNet، أثناء التدريب يتم تمرير الصورة ذات الحجم الموحد إلى نموذج الشبكة VGG16، بينما تسمح شبكة الكبسولة باستقبال صور بحجوم مختلفة كدخل لها لتقوم فيما

بعد بتوحيد الحجم ضمن طبقاتها.[1] يمرر النموذج VGG16 الصورة عبر مجموعة من الطبقات الالتفافية لتقوم باستخلاص السمات بحيث تتخلل هذه الطبقات طبقات التجميع الأعظمي max pooling التي تعمل على تخفيض أبعاد خرائط السمات الناتجة عن الطبقات الالتفافية بغية تقليل الحسابات مما يفقد هذا النوع من النماذج معلومات مهمة قد تكون هذه المعلومات حرجة في مجال الصور الطبية، بالمقابل تستبدل شبكة الكبسولة عملية التجميع الأعظمي بخوارزمية التوجيه الديناميكي عبر إضافة طبقات caps pooling التي تعمل على اختيار سمات من طبقة الكبسولات الأساسية ليتم تمريرها إلى الكبسولات في المستوى الأعلى، أي بدلا من إرسال كل أشعة الخرج الناتجة عن الكبسولات الأولية، سنحتاج إرسال شعاع خرج واحد يحمل صفات صنف الكائن الهدف، وبالتالي إخفاء كل أشعة الخرج الأخرى، الأمر الذي يقلل من الحسابات غير المفيدة في شبكة تحديد الكائنات. يتم استخدام الشبكة VGG16 كشبكة أساسية في نموذج شبكة الكبسولة للحصول على ترميزات الصور الأكثر تعقيدا قبل أن يتم إدخالهاعبر طبقات الكبسولات الأولية. مما يحسن من دقة النموذج.

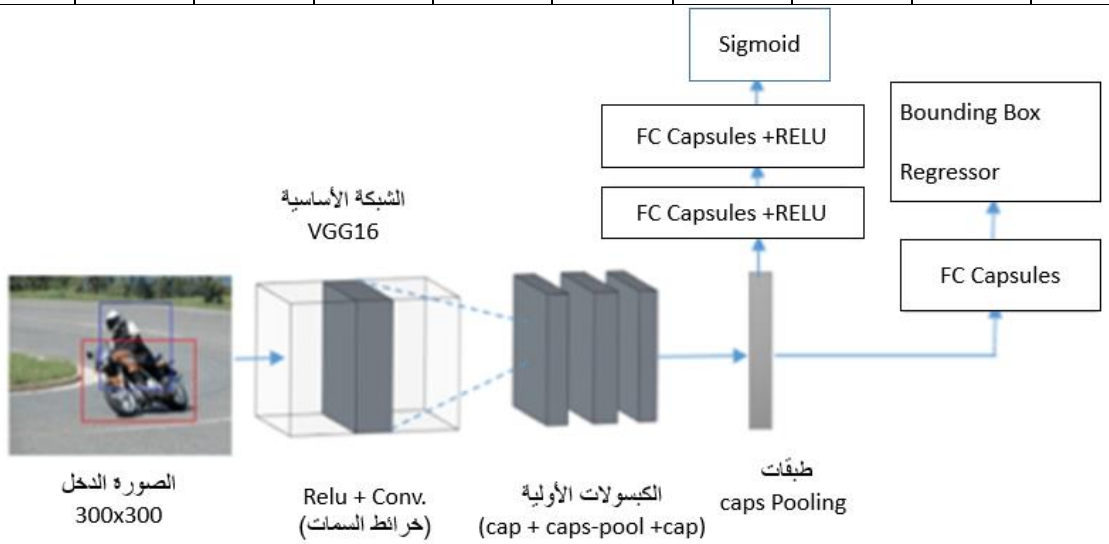
4-4 الاختبار ومناقشة النتائج:

تم اختبار كلا النموذجين (نموذج الشبكة العصبونية الالتفافية VGG16، ونموذج شبكة الكبسولة) على قاعدة البيانات Pascal Voc 2007، وتمت مقارنة نتائج النموذجين باستخدام نفس بيانات التدريب والاختبار. أحرز نموذج VGG16 معدل متوسط دقة التصنيف mean Average Precision (mAP) 67.8%، بينما أحرزت شبكة الكبسولة 67.3%. أظهرت النتائج تنافسية بين النموذجين على الرغم من الفروقات الكبيرة في الأداء، إذ لم يقدم نموذج الكبسولة دقة أعلى مما تم التوصل إليه في VGG16 إلا أنها بداية لمفهوم شبكات الكبسولة في مجال تحديد الكائن.

يبين الجدول (3) دقة تصنيف نموذج VGG16 لعدة أصناف من قاعدة البيانات Pascal Voc 2007:

الجدول (3) دقة تصنيف النموذج VGG16 لعدة أصناف من قاعدة البيانات Pascal Voc 2007:

mAP%	table	cat	car	bus	bottle	boat	bird	bike	النموذج
68.7	65.7	79.7	79.5	75.0	52.1	56.6	56.6	78.6	VGG16



الشكل (10): هيكلية شبكة الكبسولة المعتبر.[1]

بلغ عدد بارامترات نموذج VGG16 138 مليون بارامتر،حيث أن عدد البارامترات في طبقات FC أكثر من تلك الموجودة في الطبقات الالتفافية. وعلى الرغم من العمق الكبير إلا أن عدد الأوزان ليس أكبر بكثير من أوزان نماذج شبكات عصبونية التفافية تقليدية تمتلك طبقات التفافية أقل. كما أثبتت التجارب أن العمق (عدد الطبقات الالتفافية) مكون أساسي للأداء الجيد إذ يزيد من التمثيلات المرئية وبالتالي كلما زاد العمق يزداد معه دقة التصنيف، وبالمقابل يزيد زمن التدريب واستهلاك الذاكرة.

الجانب السلبي لهذه الشبكة هو كلفتها الكبيرة حيث أنها تحتاج المزيد من الذاكرة والبرمترات (M 140)، معظم هذه البارامترات موجودة في الطبقة FC الأولى، وقد تبين أنه يمكن إزالة طبقات FC بدون خفض مستوى الأداء مما يقلل بشكل كبير من عدد البارامترات.

على الرغم من ضحالة عدد الطبقات في شبكة الكبسولة إلا أنها تقدم أداءا جيدا في إنجازعملية التصنيف وقد تم قياس ذلك من خلال دقة التصنيف العالية المنجزة مقارنة مع بنى مختلفة تم اختبارها على نفس قاعدة البيانات Pascal Voc 2007 كما هو موضح في الجدول (4)، حيث أن السعي حثيث لبناء نموذج شبيه بـ YOLO أو SSD وذلك لأنهم أسرع نماذج تحديد كائنات على الإطلاق.

الجدول (4) مقارنة طرق تحديد الكائنات مع نموذج شبكة الكبسولة على قاعدة البيانات Pascal Voc 2007 [1]:

نموذج تحديد الكائنات	الدقة mAP%
YOLOv1	78.6
SSD 300	74.3
VGG16 +capsnet	67.3

تم استخدام Caps–Pooling في هذا النموذج وهي تقابل الطبقة max–pooling في الشبكات العصبونية الالتفافية التقليدية CNN، لتقوم بمهمة التوجيه الديناميكي بين الكبسولات، وبالتالي الحصول على نتائج أفضل خلال وقت التدريب إذ أنها تحافظ على ثبات الشبكة أمام تغيرات الكائن في الصورة (إزاحة، دوران، تغيرات الإضاءة، وغيرها...). فتزداد القدرة على تصنيف وتحديد الكائنات التي لم يشاهدها بعد وبالتالي ليس هناك حاجة لزيادة بيانات التدريب. أظهر الاختبار في كثير من الأحيان فشل هذا النموذج في تحديد واكتشاف كائن في صورة تم اكتشاف مثيل له في نفس الصورة، كما لم يحرز دقة مثل دقة نماذج SSD، YOLO والتي تعرف بأنها أفضل أنظمة تحديد واكتشاف في الزمن الحقيقي.

لتحسين قدرة النموذج في التمثيل الجيد لا بد من تكديس المزيد من طبقات الكبسولة لزيادة التعقيد والحصول على السمات ذات التفاصيل الأكثر تعقيدا، ولكن سيزداد مقابل ذلك استهلاك الذاكرة والحسابات. كما أن زيادة تكرارات التوجيه يزيد من سعة الشبكة ويناسب مجموعة بيانات التدريب. لذا يتم استخدام ثلاثة تكرارات للتوجيه.

بزيادة أبعاد البيانات، نحاول أن نزيد عدد طبقات الكبسولات الأولية. فينتج دقة أفضل لتعلم السمات الأكثرغنى حتى الطبقة1024 ولكن بعد 1024 كبسولة لا تزيد الدقة وبالتالي يتم اعتماد 1024 كبسولة. يتم استخدام الشبكة VGG16 كشبكة أساسية للحصول على ترميزات الصور الأكثر تعقيدا قبل أن يتم إدخالهاعبر طبقات الكبسولات الأولية. فينتج ترميزات أكثر تعقيدا في الصورة، وتحسن من دقة النموذج.

تعتبر شبكة الكبسولة طريقة جديدة في معالجة الصور ولا يمكن الهروب من قيودها في العديد من التطبيقات العملية. على الرغم من ذلك فقد ساهمت في معالجة العديد من القضايا التي كانت تمثل ضعفا في الشبكات العصبونية الالتفافية التقليدية:

1- فقدان المعلومات المهمة الناتج عن طبقات التجميع، فاعتمدت شبكة الكبسولة خوارزمية التوجيه الديناميكي: إذ عملت طبقات التجميع على اختصار الحجم المكاني للصورة بغية تقليل الحسابات والحفاظ على موارد الحوسبة وقد سبب ذلك فقدان معلومات مهمة عن الكائنات وتمثيلها في الصورة، بينما عملت خوارزمية التوجيه الديناميكي لتقليل حجم الحسابات على تقليل كمية البيانات الممررة إلى الطبقة التالية عن طريق تحديد البيانات "سمات الكائن" المفيدة في التمثيل وتقويتها (بالأوزان) والعمل على إضعاف باقي البيانات أي جعلها تقترب من الصفر حتى يكون تأثيرها معدوم في حجز موارد الحوسبة وبهذا المنطق لم تتخلى الشبكة عن البيانات وإنما عملت على انتقائها.

2- لا تأخذ الشبكات CNNs بعين الاعتبار العلاقات المكانية بين أجزاء الصورة الهامة في التعرف على الهوية، لذا فقد عانت من حساسية الدوران . [2] على العكس من ذلك ، تقوم شبكات الكبسولة عن طريق كبسولاتها بتمثيل العلاقات المكانية والاحتفاظ بمعلومات الموضع، فمثلا، عند تدريب CNNs على مجموعة بيانات الصور الحاوية على الباندا فإنها تتعلم سمات مثل "العين اليسرى" و "العين اليمنى" و "الأنف" وغير ذلك، كما هو مبين بالشكل (11):

الشكل (11): السمات المدربة في قاعدة البيانات، والنتيجة المقابلة.

عندما يتم تغذية صورة من قاعدة البيانات إلى CNN من أجل التصنيف، فإنه ستكتشف السمات التي تم تعلمها في الدخل المعتبر. ومن ثم ، فإنها ستصنفها بشكل صحيح على أنها "باندا." كما هو مبين في الشكل (12):

الشكل (12): التصنيف الصحيح للمصنف CNNعند إدخال صورة من قاعدة البيانات.

عندما يتم تغذية صورة مشوهه على CNN من أجل التصنيف ، فإنه يكتشف السمات التي تم تعلمها في الإدخال المحدد. ومن ثم ، فإنها ستصنفه بشكل غير صحيح على أنه "باندا." كما هو مبين في الشكل (13):

الشكل (13): التصنيف الخاطئ للمصنف CNNعند إدخال صورة مشوهة.

عندما يتم تغذية الصورة المدارة إلى CNN من أجل التصنيف ، فإنها ستفشل في اكتشاف السمات المكتسبة في الدخل المحدد .ومن ثم ، فإنها ستصنفه بشكل غير صحيح على أنه "ليس باندا." كما هو مبين في الشكل (14):

الشكل (14): التصنيف الخاطئ للمصنف CNNعند إدخال صورة مدارة.

الحل البديل لجعل مصنف CNN يقوم بالتصنيف الصحيح للصور المدارة هو إضافة صور مماثلة (صور ذات اتجاه وحجم مماثل ، ..) في مجموعة بيانات التدريب ووصفها بأنها "باندا ."سيؤدي هذا إلى شبكة CNN بسمات تعلم لتوجيهات أكثر للأنف والعينين .هذا يتطلب المزيد من البيانات في أوضاع مختلفة.

3- تعتبر CNNs أكثر حساسية للصورة الأصلية نفسها من أجل تصنيف الصور على أنها نفس الصنف، فهي قد تخطئ في حال تغيير ملامح الصورة الأصلية"، أي: يمكن للشبكة تصنيف الصور أو الكائنات الموجودة والتي تكون قريبة جدًا من الكائنات التي شاهدتها من قبل " أثناء التدريب". ولكن إذا تم تدوير الكائن قليلاً كما في الشكل (14)، أو تصويره من زاوية مختلفة قليلاً ، خاصةً في الأبعاد الثلاثية ، أو في اتجاه آخر غير ما تم تدريب – CNN عليه لن تتعرف الشبكة عليه بشكل جيد. لذا تعد CNN شبكات رائعة لحل المشكلات المتعلقة بالبيانات المشابهة لما تم تدريبها عليهم .أحد الحلول ضمن CNNs هو إنشاء تمثيل مائل للصورة بشكل مصطنع وإضافتها إلى مجموعة "التدريب ."ومع ذلك ، لا يزال هذا يفتقر بالأساس إلى هيكل أكثر قوة وثبات. عالجت شبكة الكبسولة هذا الضعف عن طريق الاحتفاظ بمعلومات العلاقات المكانية باستخدام علاقات الوضعية بين أجزاء الكائنات ؛ أي قياس التدوير النسبي والإزاحات بين الأجسام.

5- الاستنتاجات والتوصيات:

تعتبر شبكات caps تمثيل حقيقي لنظام الرؤية البشرية. إذ إنها تستخدم Inverse graphics وبالتالي لن تحتاج الشبكة لتعلم كل موقع وتغير للكائن.[1] يستخدم نموذج شبكة الكبسولة شبكات التفافية عصبونية بالاعتماد على مفهوم نقل التعلم لتستفيد من قدرتها على استخلاص السمات الأساسية من صور الدخل. وبدلا من استخدام التجميع الأعظمي تم إضافة caps–poolingوالتي تعمل على الحفاظ على معلومات مفصلة عن الموضع والموقع للبيانات بحيث يحتاج نموذجنا إلى بيانات تدريب أقل من النماذج الأخرى. للحصول على سمات أكثر تعقيدا يتم العمل على تكديس طبقات كبسولة أكثر الأمر الذي يحسن من قدرة النموذج التمثيلية.

كذلك فإن زيادة أبعاد البيانات، ستزيد عدد طبقات الكبسولات الأولية. فينتج دقة أفضل لتعلم السمات الأكثرغنى. حتى الآن لم تتمكن شبكات الكبسولة من محاكاة الزمن الحقيقي والحصول على دقة عالية في التحديد والتصنيف كما في نماذج SSD و YOLO. يمكن أن يتم مستقبلا بناء شبكة كبسولة بطبقات أكبر واعتماد إحدى نماذج التحديد والتصنيف بالزمن الحقيقي كشبكة أساسية لها ودراسة تأثير هذه النماذج مع القدرات التي حسنتها شبكة الكبسولة.

المراجع:

[1] B. Amit, D. Swati, "Capsule-Networks: Towards Object-Detection Capsule Object-Detector (COD)", International Journal of Computer Sciences and Engineering, Vol.-7, Issue-2, Feb 2019 E-ISSN: 2347-2693.

[2]Y. Yusuf, B. Yargi, A. Umit, "Classification of white blood cells using capsule networks", Computerized Medical Imaging and Graphics, 3 January 2020, S0895-6111(20)30002-1.

, https://doi.org/10.1016/j.compmedimag.2020.101699.

[3] B. Amlan, P. Lykourgos, D.C. ،Gaetano, S. John, "Indoor Home Scene Recognition Using Capsule Neural Networks", International Conference on Computational Intelligence and Data Science (ICCIDS 2019), Procedia Computer Science 167 (2020) 440–448.

[4] B. Amanjot, A. R. Ayesha, L. W. Douglas, "Interpreting Capsule Networks for Image Classification by Routing Path Visualization", A Thesis presented to The University of Guelph, March, 2020.

[5] Z. Ping, W. Ping, H. SHuhuan, "CapsNets algorithm", Journal of Physics: Conference Series, 1544 (2020) 012030, doi:10.1088/1742-6596/1544/1/012030.

[6] S. Kuan-Hung, C. Ching-Te, L. Jiou-Ai, and B. Yen-Yu, "Real-Time Object Detection With Reduced Region Proposal Network via Multi-Feature Concatenation", 2162-237X © 2019 IEEE.

[7] S. Karen, Z. Andrew, "visual object recognition using deep convolutional neural network",Bachelor of science in computer science Brack university, Dhaka springm2017.

[8] J. Xuefeng, W. Yikun, L. Wenbo, L. Shuying, and L. Junrui, "CapsNet, CNN, FCN: Comparative Performance Evaluation for Image Classification", International Journal of Machine Learning and Computing, Vol. 9, No. 6, December 2019.

[9] S. Sara, F. Nicholas, H.Geoffrey E., "Dynamic Routing Between Capsules", arXiv:1710.09829v2 [cs.CV] 7 Nov 2017.

[10] K. D. Tejas, F. W. William, K. Pushmeet, B. T. Joshua, "Deep Convolutional Inverse Graphics Network", December 19, 2015.

[11] Redmon, J., Farhadi, A., YOLOv3: An Incremental Improvement, arXiv:1804.02767v1 [cs.CV] 8 Apr 2018.

[12] S. Karen, Z. Andrew, "very deep convolutional networks for large-scale Image recognition", Visual Geometry Group, Department of Engineering Science, University of Oxford, arXiv:1409.1556v6 [cs.CV] 10 Apr 2015.

[13]"Convolutional Neural Networks (CNNs / ConvNets)", CS231n: Convolutional Neural Networks for Visual Recognition, 17 Jan. 2017.

[14]https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751, 25/1/2020, 12:19AM.

[15]https://towardsdatascience.com/capsule-neural-networks-are-here-to-finally-recognize-spatial-relationships-693b7c99b12, 17/4/2020 ،9:33PM.

[16]https://software.intel.com/en-us/articles/understanding-capsule-network-architecture

2020/4/191:10 AM

[17] https://www.analyticsvidhya.com/blog/2018/04/essentials-of-deep-learning-getting-to-know-capsulenets/11:42 ،2020/3/9 PM