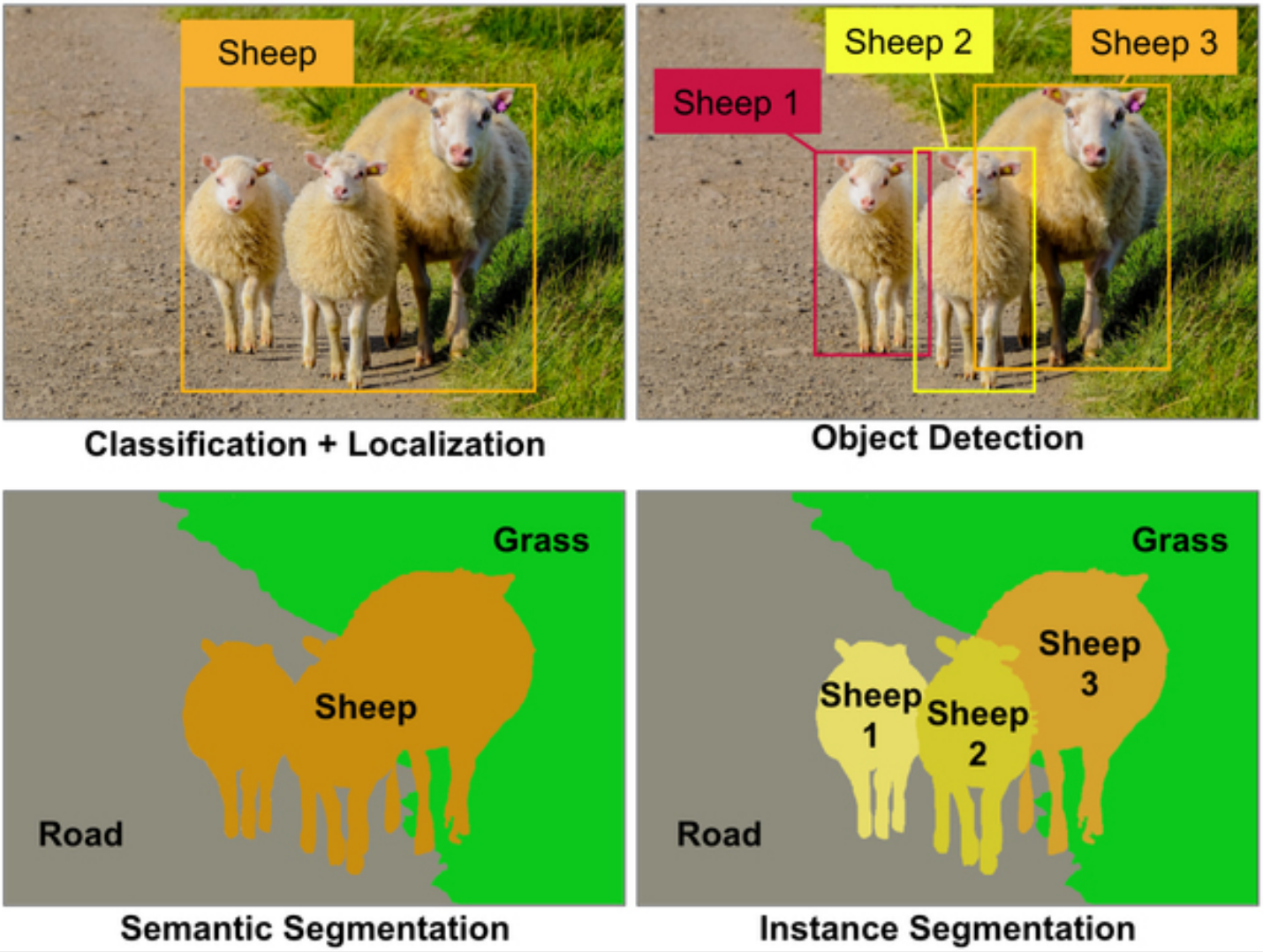
مدخل إلى الكبسولات الشبكية العصبية CapsNets



استطاعت الكبسولات الشبكية العصبية Capsule Networks (ويرمز لها اختصاراً بـ CapsNets) أنْ تتغلبَ على بعض أوجه القصور في الشبكات العصبية الملتفّة Convolutional Neural Networks (ويرمز لها اختصاراً بـ CNNs) من حيث حاجتها إلى تدريب أقل، وقدرتها على الاحتفاظ بتفاصيل الصورة وقدرتها على التعامل مع الغموض في الصور.

تعتبر الكبسولات الشبكية العصبية من التصاميم الجديدة والمهمة فيما يتعلق بالشبكات العصبية الاصطناعية، حيث يتوقع أن يكون لها أثر كبير على ما يسمى التعلم العميق **deep learning**، وبالتحديد الرؤية الحاسوبية **computer vision**، ولكن ألم تكن الرؤية الحاسوبية من المجالات التي حققنا فيها بالفعل تقدماً كبيراً؟ ألم تحقق الشبكات العصبية الملتفّة **CNN** نتائج رائعة في بعض مجالات الرؤية الحاسوبية مثل التصنيف **classification**، التموضع **localization**، التعرف على الأجسام **object detection**، التجزئة

الدلالية **semantic segmentation**، وحتى تجزئة الأجسام داخل الصورة **instance segmentation** (الموضحة في الشكل رقم **1**)



بعض أهم المهام المتعلقة بالرؤية الحاسوبية. كل واحدة من هذه المهام، اليوم، هي بحاجة الى بنية مختلفة جداً من الشبكات العصبية الملتفّة CNN. فمثلاً، يتم استعمال ResNet لأغراض التصنيف، وYOLO لغرض التعرف على الأجسام، وMask R-CNN لتجزئة الأجسام داخل الصورة، وهكذا. حقوق الصورة لـ أوريلين جيرون Aurelien Geron

بالفعل، كان أداءُ الشبكات العصبية الملتفة **CNN** ممتازاً. مع هذا، فقد كانت هناك بعض العيوب. فمثلاً:

1. تحتاج الى التدريب باستخدام عدد هائل من الصور (أو أنها تعيد استعمال أجزاء من الشبكات العصبية التي تم تدريبها بالفعل على عدد هائل من الصور). في المقابل، فإن الكبسولات الشبكية العصبية **CapsNets** قادرة على تعميم النموذج بشكل جيد وباستخدام بيانات تدريب أقل.
2. الشبكات العصبية الملتفة **CNN** لا تملك القدرة على التعامل مع الغموض الموجود في الصور بشكل جيد، بينما الكبسولات الشبكية العصبية **CapsNets** لها القدرة على الأداء بشكل جيد حتى لو استخدمنا مشاهد مزدحمة (مع ذلك، فهي ما تزال تواجه صعوبات في التعامل مع الخلفيات في الوقت الراهن).
3. تفقد الشبكات العصبية الملتفة **CNN** الكثير من المعلومات في طبقات التجميع، حيث تعمل هذه الطبقات على تقليل الدقة

المكانية (كما هو موضح في الشكل رقم **2**)، لذا فإنّ مخرجاتها تبقى ثابتة مهما حصل من تغيرات طفيفة في المدخلات. وتبرزُ هنا مشكلة تتعلق بضرورة الاحتفاظ بمعلومات تفصيلية في جميع أنحاء الشبكة، مثل المعلومات الخاصة بالتجزئة الدلالية. يتم التعامل مع هذه المشكلة اليوم من خلال بناء بُنى معقدة للشبكات العصبية الملتفة **CNN** وذلك لغرض استعادة بعض المعلومات المفقودة. في المقابل، ومع الكبسولات الشبكية العصبية **CapsNets**، يتم الاحتفاظ بالمعلومات المتعلقة بوضعية الجسم (مثل الموقع الدقيق للجسم، ومقدار التدوير، والسمك، والانحراف، والحجم، وما إلى ذلك) عبر الشبكة، بدلاً من فقدانها ثم محاولة استعادتها من جديد. التغيرات الطفيفة في المُدخلات تؤدي إلى تغيرات طفيفة في المخرجات ـ وهذا يعني الحفاظ على المعلومات (يطلق على هذه العملية تسمية الفروق المتساوية **equivariance**).ونتيجة لذلك، تستطيع الكبسولات الشبكية العصبية **CapsNets** استخدام نفس هذه البنية البسيطة والمتناسقة في مهام الرؤية المختلفة.

4. أخيراً، فإنّ الشبكات العصبية الملتفة **CNN** تحتاج إلى وجود مكوّنات إضافية لكي تصبح قادرة وبشكل تلقائي على تمييز الأجسام التي تنتمي إليها الأجزاء المختلفة (على سبيل المثال، تنتمي هذه الساق إلى ذلك الحيوان). في المقابل، فإنّ الكبسولات الشبكية العصبية **CapsNets** توفر تسلسل هرمي لتلك الأجزاء وبدون جهدٍ يُذكر.

مخطط سير منظومة التعلم العميق DeepLab2 الخاصة بتقطيع الصور، بواسطة Liang-Chieh Chen وآخرين. لاحظ أنّ ناتج الشبكة العصبية الملتفة CNN (الموجود في أعلى يمين الصورة) هو فضفاض للغاية مما يجعل مسألة إضافة خطوات أخرى لاستعادة بعض التفاصيل المفقودة ضرورية للغاية. الصورة مأخوذة من الورقة البحثية المعنونة DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs

تم طرح موضوع الكبسولات الشبكية العصبية **CapsNets** لأول مرة في العام **2011** من قبل جيفري هينتون **Geoffrey Hinton** وآخرين في الورقة البحثية التي كان عنوانها **Transforming Autoencoders**. مع هذا، فلم يتم إلّا قبل عام فقط، وبالتحديد في شهر تشرين الثاني/نوفمبر من عام **2017**، أنْ قام كل من سارة **صبورSara Sabour**، ونيكولاس فروست **Nicholas Frosst**، وجيفري هينتون بنشر ورقة بحثية بعنوان **Dynamic Routing between Capsules** شرحوا فيها بنية الكبسولات الشبكية العصبية **CapsNets** التي حققت الأداء الأفضل باستخدام مجموعة البيانات الشهيرة **MNIST** والخاصة بصور الأرقام المكتوبة بخط اليد، حيث تم الحصول على نتائج أفضل بكثير من تلك التي حققتها الشبكات العصبية الملتفّة **CNN** التي تم اختبارها على مجموعة البيانات المعروفة باسم **MultiMNIST** (وهي عبارة عن أزواج متداخلة من أرقام مختلفة).

الصور الخاصة بمجموعة بيانات MultiMNIST (باللون الأبيض) ويقابلها التمثيل المقابل الذي بنته الكبسولات الشبكية العصبية CapsNets (باللونين الأحمر والأخضر). يشير الحرف "R" لعملية بناء التمثيل المقابل، بينما يشير الحرف "L" إلى الرموز المستخدمة. فعلى سبيل المثال، التوقعات الخاصة بالأمثلة الموجودة في الجزء العلوي الأيسر صحيحة، وكذلك هو الحال مع التمثيلات المقابلة. ولكن، في المثال الخامس كان التوقع خاطئاً، ألا وهو (5,7) بدلاً من (5,0). لهذا السبب فقد تم إعادة تمثيل الرقم 5 بشكل صحيح بينما لم يكن الحال كذلك مع الرقم 0. الصورة مأخوذة من الورقة البحثية Dynamic routing between capsules

ولكن، وعلى الرغم من كل ما تمتاز به الكبسولات الشبكية العصبية **CapsNets**، فهي ما تزال تفتقر إلى الكمال. أولاً، وفي الوقت الحالي، فهي لا تقدم أداءً جيداً مع الصور كبيرة الحجم مثل مجموعة صور **CIFAR10** و **ImageNet** بالمقارنة مع الشبكات العصبية الملتفة **CNN**. علاوةً على ذلك، فإنها من ناحية الحوسبة تعتبر مكلفة وتفتقر إلى القدرة على التمييز بين اثنين من الأجسام من نفس النوع حين

يكونان قريبين جداً من بعضهما البعض (يطلق على هذه المشكلة تسمية "مشكلة الازدحام" وقد ثبت أنّ البشر يعانون منها أيضاً). مع هذا، فإنّ الأفكار الأساسية واعدة للغاية ويبدو من المرجح أنّها تحتاج فقط الى بضعة تعديلات لتكون قادرة على الوصول إلى الأداء الأمثل. وعلى أية حال، فإنّ الشبكات العصبية الملتفة **CNN** والتي تم ابتكارها في العام **1998** لم تصل إلى أداءها الأمثل من خلال اختبارها على مجموعة صور **ImageNet** إلا في العام **2012**، وبعد إجراء بضعة تعديلات.

والآن، ما هي بالضبط الكبسولات الشبكية العصبية CapsNets؟

باختصار، فإنّ الكبسولات الشبكية العصبية تتألف من كبسولات (عليبات) بدلاً من خلايا عصبية. كل كبسولة من هذه الكبسولات هي عبارة عن مجموعة صغيرة من الخلايا العصبية التي تتعلم لتصبح قادرة على تمييز جسم معين (على سبيل المثال، مستطيل) ضمن منطقة محددة من الصورة ومن ثم تقوم بإنتاج متجه **vector** (مثلاً متجه بثمانية أبعاد)، يمثل الطول فيه ماهية احتمالية كون الجسم موجوداً، بينما يرمز الانحناء **orientation** (على سبيل المثال في فراغ ذي ثمانية أبعاد) إلى وضعية الجسم (على سبيل المثال، الموضع الدقيق، التدوير، إلخ). في حال طرأ أي تغيير طفيف على الجسم (على سبيل المثال، إزاحة، تدوير، تغيير حجم، إلخ)، ستقوم الكبسولة حينها بإخراج متجه له نفس الطول ولكن بانحناء مختلف قليلاً. وهكذا يتحقق مبدأ الفروق المتساوية بين الكبسولات.

مثلما هو الحال مع الشبكات العصبية الاعتيادية، تنتظم الكبسولات الشبكية العصبية **CapsNets** على شكل طبقات متعددة (كما في الشكل رقم **4**). يطلق على الكبسولات في الطبقة المنخفضة تسمية "الكبسولات الرئيسية" حيث يتلقى كل منها منطقة صغيرة من الصورة كمُدخل (يطلق عليها منطقة الاستلام). وظيفة هذا النوع من الكبسولات هو محاولة الكشف عن وجود نمط معين وتحديد وضعيتها (على سبيل المثال، مستطيل). في المقابل، فإن وظيفة الكبسولات في الطبقات العلوية، والتي يطلق عليها "كبسولات التوجيه"، هو الكشف عن وجود الأجسام الأكبر والأكثر تعقيداً (مثل القوارب).

كبسولات شبكية عصبية CapsNets بطبقتين. في هذا المثال، تتألف طبقة الكبسولات الرئيسية من مستويين 5*5 من الكبسولات، بينما تتألف طبقة الكبسولات الثانية من مستويين 3*3 من الكبسولات. تقوم كل كبسولة بإنتاج متجه، بينما يمثل كل سهم المخرج لكبسولة مختلفة. فالأسهم الزرقاء تمثل مخرجات الكبسولات التي تحاول تمييز المثلثات، بينما الأسهم السوداء تمثل مخرجات الكبسولات التي تحاول تمييز المستطيلات، وهكذا. حقوق الصورة أوريلين جيرون

يتم تنفيذ طبقة الكبسولات الرئيسية باستخدام عدد قليل من الطبقات الملتفّة الاعتيادية. فعلى سبيل المثال، استخدم الباحثون في تلك الورقة العلمية طبقتين ملتفتين لها قدرة على انتاج **6*6*256** مستوي خصائص يحوي كميات عددية غير متجهة **scalars**. ثم يقوم الباحثون بإعادة تشكيل هذا المخرج ليصبح **6*6*32** مستوي خصائص وبمتجهات ثمانية الأبعاد. وأخيراً، يقوم الباحثون باستخدام أداة مبتكرة تسمى دالة السحق **squashing function** وظيفتها هو التأكد من أنّ هذه المتجهات تمتلك أطوال تتراوح بين الـ **0** وال **1** (لتمثيل الاحتمالية). وهكذا انتهى الأمر!! حيث أصبح الناتج النهائي هو مخرج الكبسولات الرئيسية.

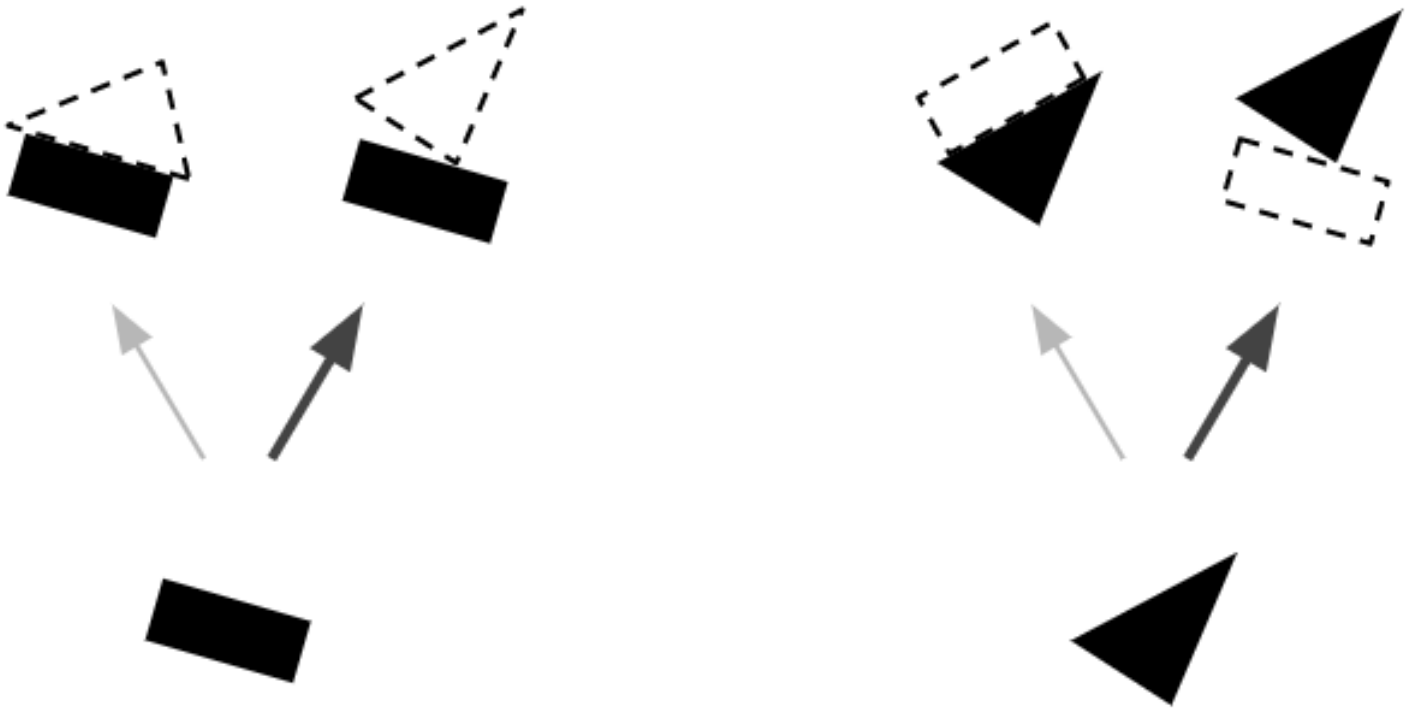
تقوم الكبسولات في الطبقات التالية أيضاً بمحاولة الكشف عن وجود الأجسام وعن أوضاعها، ولكنها تعمل بشكل مختلف تماماً، فهي تستخدم ما يطلق عليه اسم التوجيه عن طريق الاتفاق **routing** by **agreement**، حيث تبرز هنا قوة منظومة الكبسولات الشبكية العصبية **CapsNets**. لنأخذ المثال التالي:

لنفترض وجود كبسولتين أساسيتين فقط: إحداهما مستطيلة والأخرى مثلثة الشكل، ولنفترض أنّ كليهما استطاع الكشف عما يبحث عنه. في هذه الحالة، سيكون كل من المستطيل والمثلث جزءاً إما من بيت أو من قارب (موضح في الشكل رقم **5**). فلو نظرنا إلى الوضع

الذي يتخذه المستطيل، حيث تم تدويره قليلاً نحو جهة اليمين، سيتم تدويرُ كلٍ من المنزل والقارب قليلاً نحو اليمين أيضاً. ولو نظرنا إلى الوضع الذي يتخذه المثلث، فإنّ على المنزل أن يكون مقلوباً تقريباً نحو الاسفل، بينما يتم تدوير القارب قليلاً في هذه الحالة نحو اليمين. لاحظ أنّ كلا الشكلين، إضافة إلى العلاقات الجزئية والكلية، قد تم تعلمها خلال التدريب، ولاحظ أيضاً بأن المستطيل والمثلث قد اتفقا على الوضع الذي يتخذه القارب، بينما اختلفا بشكل كامل بخصوص الوضع الذي يجب أن يتخذه المنزل. لذا، فإن من المرجح أن يكون كل من المستطيل والمثلث جزءاً من نفس القارب، وليس هناك المنزل.

التوجيه عن طريق الاتفاق. في الخطوة رقم 1، يتم توقع وجود الجسم ووضعه بناءً على وجود أجزاء من ذلك الجسم وأوضاع تلك الأجزاء، ومن ثم يتم البحث عن التوافق بين التوقعات. حقوق الصورة أوريلين جيرون

نظراً لأننا أصبحنا واثقين الآن بأنّ المستطيل والمثلث هما أجزاء من القارب، فمن المنطقي إرسال مخرجات كبسولات المستطيل وكبسولات المثلث بشكل أكبر إلى كبسولة القارب، وبشكل أقل إلى كبسولة المنزل. فبهذه الطريقة، تصبح كبسولة القارب قادرة على استلام مدخلات أكثر فائدة، بينما سيصل إلى كبسولة المنزل ضوضاء أقل. في كل حالة اتصال، تقوم خوارزمية التوجيه بواسطة الاتفاق بتعديل أوزان التوجيه (الشكل رقم **6**) وذلك بزيادة تلك الأوزان في حال وجود اتفاق، وتقليلها في حال عدم وجود اتفاق.



التوجيه عن طريق الاتفاق. في الخطوة رقم 2، يتم تحديث أوزان التوجيه. حقوق الصورة لـ أوريلين جيرون

تتضمن خوارزمية التوجيه عن طريق الاتفاق تكرار للتحديثات المتعلقة بالكشف عن الاتفاقيات وتحديثات التوجيه (لاحظ بأنّ هذا الأمرَ يحدث لكل عملية توقع، ولا يحدث مرة واحدة فقط أو خلال وقت التدريب فقط). وهذا مفيد جداً في المشاهد المزدحمة. فمثلاً، وكما نرى في الشكل رقم 7، فإن المشهد يبدو غامضاً لأنك قد ترى منزلاً مقلوباً في المنتصف، تاركاً المستطيل العلوي والمثلث السفلي من غير تفسير. لذا، فمن المرجح أن تقوم خوارزمية الاتفاق عن طريق التوجيه بتعديل نفسها لكي تعطينا تفسيراً أفضل: قارب في الأسفل ومنزل في الأعلى. من هذا المنطلق، نستطيع القول بأنه تم القضاء على الغموض. فالمستطيل السفلي يمكن تفسيره على أنه قارب، والذي بدوره

أيضاً يفسر المثلث السفلي. وما أن يتم تفسير هاتين الجزئيتين، يصبح من السهل تفسير الأجزاء المتبقية على أنها منزل.

بإمكان خوارزمية التوجيه عن طريق الاتفاق تفسير المشاهد المزدحمة، مثل المشهد الموجود في الصورة، بدلاً من أن يساء تفسيرها على أنها تتكون من منزل مقلوب رأساً على عقب وبعض الأجزاء الأخرى التي لا يمكن تفسيرها. فبدلاً من ذلك، يتم تدوير المستطيل السفلي إلى القارب، وسيؤدي ذلك أيضاً إلى جذب المثلث السفلي الى القارب ايضاً. ما أنْ يتم تفسير القارب، يصبح من السهل أيضاً تفسير الجزء العلوي على أنه منزل. حقوق الصورة لـ أوريلين جيرون

وبذلك نكون شرحنا جميع الأفكار الأساسية الخاصة بالكبسولات الشبكية العصبية **CapsNets**.

يمكن الاطلاع على المزيد من المعلومات حول الكبسولات الشبكية العصبية **CapsNets** بالإضافة إلى كيفية بنائها وتطبيقها هنـا.

- التاريخ: 2019-02-04
- التصنيف: تكنولوجيا

#الرؤية الحاسوبية #الشبكات العصبية #الكبسولات الشبكية العصبية

المصطلحات

- **الرؤية الحاسوبية (Computer vision)**: هو أحد مجالات علوم الحاسب، ويعمل على تمكين الحاسوب من رؤية الصور ومعالجتها والتعرف عليها كالبشر تماماً، ومن ثم اعطاء المخرجات المناسبة.
- **التعلم العميق (deep learning)**: هو أحد ميزات الذكاء الاصطناعي التي تَعنى بمحاكاة نهج التعلم الذي يستخدمه البشر للحصول على أنواع معينة من المعرفة، كما يمكن اعتباره وسيلة لأتمتة التحليلات التنبؤية.

المصادر

- oreilly
- الصورة

المساهمون

- ترجمة
  - محمد زهير الطائي
- مُراجعة
  - فرح درويش
- تحرير
  - زين صالح
- تصميم
  - رنيم ديب
- نشر

- أمل أحمد