# طريقة سريعة للكشف عن الأجسام في مقاطع الفيديو




**على تتالي من الإطارات تمت معالجتها بآلية ROI. حقوق الصورة: Athindran et al**

تعاون باحثون من مركز روبرت بوش المتخصص بعلوم البيانات والذكاء الاصطناعي **Robert Bosch Center for Data Science** ومركز أبحاث الدماغ الحاسوبية **Center for Computational Brain Research**، والمعهد الهندي للتكنولوجيا مادراس **Indian Center of Technology Madras**، وجامعة بوردو **Purdue University**، حيث تمكنوا مؤخراً من تطوير طريقةٍ جديدة قادرة على تقليل المتطلبات الحسابية الضرورية للكشف عن الأجسام التي تظهر في مقاطع الفيديو وذلك باستخدام الشبكات العصبية، حيث تم نشر تفاصيل هذه الطريقة، والتي أطلق عليها اسم **Pack and Detect** أو اختصاراً بـ (**PaD**) في النسخة الأولى من البحث على موقع أركايف **arXiv**.

يُعتبر الكشف عن الأجسام أحد العناصر الأساسية في كثير من التطبيقات المتعلقة بمجال رؤية الحاسوب **computer vision**، مثل تتبع الأجسام، وتلخيص الفيديوهات، والبحث داخل مقاطع الفيديو. وعلى الرغم من أنّ الإنجازات التي تم تحقيقها في مجال تعلم الآلة **machine learning** كان لها دوراً كبيراً في بناء أدواتٍ ذات دقةٍ عاليةٍ قادرة على أداء هذه المهمة، إلا أنّ الأساليب المستخدمة حاليّاً ما تزال تعتبر حسابياً مكلفةً للغاية. فعلى سبيل المثال، يتطلب معالجة فيديو بوضوع صورة يبلغ **300** * **300** وباستخدام شبكة تتبع الأجسام من نوع **SSD300**، ودعامة من نوع **VGG16** وسرعة تبلغ **30** إطار في الثانية، يتطلب استخدام **1.87** ترليون عملية من عمليات الفاصلة العائمة (تسمى أيضاً **FLOPS**).

مع ذلك، فقد لاحظ الباحثون أنّه وفي بعض الحالات، تكون معظم المناطق الموجودة في مقطع الفيديو هي فقط عبارة عن خلفية تحوي على عدد من الأجسام البارزة التي تشغل جزءاً محدوداً من مساحة الصورة. إضافةً إلى ذلك، وجدوا أنّ هناك علاقة زمنية قوية تربط بين الإطارات المتعاقبة. وبناءً عليه، قاموا بالاستفادة من تلك الملاحظات واقترحوا طريقةً جديدة للكشف عن الأجسام الموجودة في مقاطع الفيديو بإمكانها تقليل المتطلبات الحسابية لمهمة تتبع تلك الأجسام.

وفي هذا السياق، تحدّث أثيندران راميش كومار **Athindran Ramesh Kumar** وهو أحد المساهمين في هذا البحث إلى موقع **TechXplore** عن هذه التجربة قائلاً: "لقد استلهمنا الفكرة الحالية من آلية الإبصار المركزي الموجودة لدى أنظمة الرؤية الطبيعية والاصطناعية. في الماضي، تركزت الجهود السابقة المتعلقة بآليات الانتباه البصري المركزي في أنظمة الرؤية الاصطناعية على منطقة واحدة فقط من الصورة أو على جسمٍ واحدٍ في الوقت الواحد. وهذا دفعنا إلى التساؤل: كيف سيكون شكل نظام الرؤية إذا كان بالإمكان التركيز على جميع المناطق البارزة الموجودة في الصورة دفعةً واحدة؟".

بناءً عليه، يمكنُ اعتبارُ الطريقة الحالية للكشف عن الأجسام والتي توصل إليها هؤلاء الباحثون قد استُلهمت من أنظمة الرؤية البيولوجية البصرية. ولكن، وبخلاف المحاولات السابقة، فإنّ الطريقة الحالية تقوم بجمع كل المناطق ذات الأهمية في إطارٍ واحد، بدلاً من معالجتها جميعاً وبشكلٍ متعاقب.

من جانبه، تحدّث بالارامان رافيندران **Balaraman Ravindran** وهو باحثٌ آخر في نفس الدراسة قائلاً: "كان الهدف من الدراسة هو تسريع عملية الكشف عن الأجسام في مقاطع الفيديو من خلال التركيز بشكل حصري على المناطق البارزة الموجودة في الصورة والتخلص من الفوضى الموجودة في الخلفية. ولغرض التخلص من تلك الفوضى في الخلفية، قمنا بالاستفادة من الارتباط الزمني الموجود بين الإطارات المتقاربة في مقطع الفيديو. فبينما تستفيد تقنيات الضغط من هذه الطريقة من أجل تقليل متطلبات الخزن وعرض النطاق الترددي **bandwidth**، فنحن نستخدمها هنا من أجل تسريع عمليات الحوسبة."

تعمل تقنية **PaD** الجديدة عن طريق معالجة الإطارات على فتراتٍ منتظمة وبحجمها الكامل، حيث يشارُ إلى هذه الإطارات بـ "الإطارات الرّئيسة"، وتتعرف الأداة على المناطق المهمة في بقية الإطارات اعتماداً على الموضع الذي تكون فيه الأجسام موجودة في الإطار السابق.

ويضيف الباحث أناند راغوناثان **Anand Raghunathan**: "تتجمّع تلك المناطق ذات الأهمية مع بعضها في صورةٍ مجمّعة يتم استخدامها كدخل لأداة الكشف عن الأجسام، حيث يتمُّ بعدها مطابقةُ الكشوفات مع المواقع الموجودة في الصورة الأصلية. وهذه الطريقة الجديدة تعتبر أسرع لأنّ الصور المجمّعة تكونُ أصغر حجماً من الإطارات الكاملة، مستفيدين من المرونة التي تمنحها بعضُ النظم الشائعة في الكشف عن الأجسام مثل **SSD300** لمعالجة الصور بكلا الحجمين: الحجم الكامل والأصغر حجماً."

استخدم الباحثون نوعاً من البيانات اسمها **ImageNet VID** لتقييم أداء الطريقة الجديدة، حيث كانت النتيجة أن تلك الطريقة حققت

سرعةً أفضل بمقدار **1.25** مرة، بينما صاحَبها انخفاض في الدقة بلغ أقل من **1.6** في المائة. إضافةً إلى ذلك، وجد الباحثون بأنّ الوقت المستغرق من أجل معالجة الصور ذات الحجم الصغير كان أقل بثلاث مراتٍ تقريباً، رافقه انخفاض في عدد عمليات **FLOPS** بمقدار أربع مرات.

علاوةً على ذلك، أشارت تلك الدراسة إلى أمرين اثنين لهما أهميّة كبيرة حول كيفية تطوير طرق تكون أسرع وأقل كلفة من الناحية الحسابية للكشف عن الأجسام في مقاطع الفيديو. الأمر الأول هو أنّ الأجسام موضع الاهتمام عادةً ما تحتل عدداً صغيراً من البكسلات pixels الموجودة في الإطار، والأمر الثاني هو وجود علاقة تربط بين الإطارات المتتالية داخل مقطع الفيديو.

وفي هذا السياق، يقول اثندران: "باستطاعة العمل الذي قدمناه أن يجعل تحليل الفيديو باستخدام أجهزة محدودة الموارد وباستخدام أجهزة إنترنت الأشياء **Internet of Things** أمراً ممكناً وذلك من خلال تقليل متطلبات الحوسبة، كما يمكنه تحسين عدد الفيديوهات المتدفقة التي يمكن معالجتها بواسطة الخوادم الموجودة في المنصة السحابية."

تعتبر تلك الدراسة التي قام بها هذا الفريق من الباحثين خطوة أوّلية باتجاه تطوير أدوات أكثر فاعلية للكشف عن الأجسام، حيث من المرجح القيام بمزيد من البحوث والتجارب التي من شأنها أن تجعل تلك الطريقة ذات فاعلية أكبر.

فعلى سبيل المثال، فإنّ طريقة **PaD** بشكلها الحالي تقوم باختيار الصور الرئيسية على فواصل منتظمة، ومع هذا فإنّ بإمكان الباحثين تطوير آلية تستطيع التعرف بشكل تلقائي على تلك الإطارات الرئيسية. إضافةً إلى ذلك، يتم العمل حالياً من أجل اختبار تلك الطريقة في بيئات تكون فيها الموارد الفيزيائية محدودة مثل مجال الهواتف الذكية والأجهزة القابلة للارتداء والأجهزة المنزلية الذكية.

وهنا يضيف رافندران: "قمنا بصياغة خوارزمية قادرة على تحديد المناطق ذات الأهمية من أجل تشكيل صورة مجمّعة. ولكن، طموحنا المستقبلي يتمثل في استخدام نظام عصبي متكامل يحتوي على شبكات عصبية بإمكانها تشكيل صورة مجمّعة اعتماداً على الإطار السابق".

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

#الذكاء الاصطناعي #مقاطع الفيديو

---

المصطلحات

- **تعليم الآلة (machine learning)**: تعلم الآلة هو أحد أنواع الذكاء الاصطناعي، يمكّن التطبيقات البرمجية من التنبؤ بنتائج أكثر دقة دون برمجتها بشكل صريح. ويتم ذلك عن طريق بناء خوارزميات تتلقى بيانات الإدخال وتستخدم التحليل الإحصائي للتنبؤ بقيمة المخرجات ضمن نطاق مقبول.
- **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.

## المصادر

- techxplore

## المساهمون

- ترجمة
  - محمد زهير الطائي
- مُراجعة
  - فرح درويش
- تحرير
  - زين صالح
- تصميم
  - محمد نور حماده
- نشر
  - أمل أحمد