التعليق التلقائي للصور باستخدام الشبكة الخصومية التوليدية

اسم الطالبة/ عبير محمد الشهري

اسم المشرف على الرسالة
د/ منيرة طيلب ، د/ ريم العتيبي

المستخلص

في الوقت الحاضر، يستخدم التعليق التلقائي للصور في تطبيقات مختلفة مثل استرجاع الصور وتصنيفها. حيث يلعب التعليق التلقائي للصور دورًا مهمًا في تحسين عملية استرجاع الصور. حيث يستخدم التعلم العميق في التعليق التلقائي للصور لاستخراج ميزات الصورة ثم تحويلها إلى أوصاف نصية. وهذا يعتبر حاليًا أحد أكثر مجالات البحث جاذبية. ومع ذلك، لا تزال النماذج التقليدية للتعليق التلقائي للصور التي تستخدم أساليب التعلم العميق تعاني من أوجه قصور مختلفة، كالأداء الضعيف في التعليق على الصور. لذا، هدف هذه الأطروحة إلى اقتراح نموذج تعليق تلقائي للصور بناءً على الشبكات العصبونية الالتفافية والشبكات الخصومية التوليدية وعلى نقل المعرفة. يهدف النموذج المقترح إلى التغلب على المشاكل الحالية مثل التحدي الناتج عن تدريب نموذج الشبكات العصبونية الالتفافية باستخدام بيانات صغيرة الحجم والتي تؤثر على فعالية التعليق التلقائي للصور. على وجه الخصوص، جذبت الشبكات الخصومية التوليدية الاهتمام في مجال رؤية الكمبيوتر نظرًا لقدرتها على إنشاء البيانات دون استخدام كثافة الاحتمالات بشكل صريح. فقد أثبتت فائدتها في شرح الصورة وزيادة عددها. في هذا البحث، تم استخدام المصنف الإضافي للشبكات الخصومية التوليدية، حيث يتنبأ المميّز بفئة الصورة بدلاً من اعتبارها مدخلات معينة؛ لذلك، يتم ضمان استقرار مرحلة التدريب مع توفير صور عالية الجودة. على الرغم من التطورات الحالية في التعليق التلقائي للصور باستخدام التعلم العميق، إلا أنها تحتاج إلى تحسينات من حيث الأداء والموارد الحسابية. لذلك تم استخدام نقل المعرفة في النموذج المقترح لحل هذه المشكلات؛ ولإثبات فاعلية النموذج المقترح، تم إجراء التجارب على أربع مجموعات من بيانات الصور المختلفة، وتشمل (ImageClef2011, ImageClef2012, ESPGame, Iaprtc12). حيث أشارت النتائج إلى أن النموذج المقترح قد تفوق في الأداء على غالبية نماذج التعليقات التلقائية للصور الحالية المشابهة من حيث: متوسط الدقة المحرف، والقياس F، والمساحة الواقعة تحت المنحنى، ومعدل الخطأ المتساوي، والدقة، والاستدعاء.

Automatic Image Annotation based on Generative Adversarial Networks

By
Abeer Mohammed Faiz Alshehri

Supervised by
Dr.Mounira Taileb and Dr.Reem Alotaibi

Abstract

Nowadays, Automatic Image Annotation (AIA) has been adopted in different applications such as image retrieval and classification. Indeed, AIA plays a significant role in enhancing image retrieval. Deep Learning is used in AIA to extract image features and then convert these features into text descriptions (i.e., labels). It is currently considered one of the most attractive research areas. However, Conventional AIA models that employ deep learning methods suffer from various shortcomings, such as poor annotation performance. Accordingly, this thesis proposes an AIA model based on convolutional neural networks (CNNs), Generative Adversarial Networks (GANs), and transfer learning. The proposed model aims to overcome the limitations of the existing models such as the challenge of training a CNN model using small-sized datasets that affect the effectiveness of AIA.

GANs have attracted a lot of interest in the computer vision field because of its ability to generate data without explicitly using probability density. Thus, it has proven its usefulness in image annotation, and image augmentation. In this research, an Auxiliary classifier-GAN (ACGAN) has been used, where the discriminator predicts the class of an image rather than taking it as a given input; therefore, the stabilization of the training stage is ensured, and the generation of high-quality images is provided. Also, transfer learning is used to enhance the performance of the classification.

The proposed model enhanced the best-scored state-of-art model, using ImageClef dataset, by 5.61% in terms of MiAP, a gain of 16.15% in terms of F-measure, and achieved a 69.8% reduction in terms of EER. Also, using ESPGame dataset, the proposed model enhanced the best-scored state-of-art model by 130% in terms of F-measure. Last, using the IAPR-TC12 dataset, the proposed model improved to the best-scored state-of-art model, by 119% in terms of F-measure.