

قراءة الشفاه المرئية للأبجدية العربية والكلمات القرآنية باستخدام التعلم العميق

ندأ فيصل الجهني

إشراف

د. عماد سامي جاها

المستخلص

مهبت التطورات المستمرة في التعلم العميق الطريق للعديد من الأفكار الصعبة. احدى هذه الأفكار هي قراءة الشفاه المرئية، والتي جذبت مؤخرا العديد من الاهتمامات البحثية. غالبا ما يشار الى قراءة الشفاه باسم التعرف على الكلام المرئي، وهي القدرة على فهم الكلام المنطوق والتنبؤ به بناءً على حركات الشفاه فقط دون استخدام الأصوات. بسبب نقص الدراسات البحثية حول التعرف على الكلام المرئي للغة العربية بشكل عام، وغياها في البحث القرآني، تهدف هذه الرسالة الى سد هذه الفجوة. يقدم العمل في هذه الأطروحة مجموعة بيانات جديدة لقراءة الشفاه باللغة العربية متاحة للباحثين تضم ١٠٤٩٠ مقطع فيديو تم التقاطها من زوايا متعددة وعينات بيانات تتضمن مستويات من الحروف ومستوى من الكلمات بناءً على محتوى وأحكام مساعدة للدراسة القرآنية وردت في كتاب القاعدة النورانية. من ناحية أخرى، يستخدم هذا العمل التعرف على الكلام المرئي للحروف العربية المنطوقة (الحروف الهجائية العربية)، والحروف القرآنية المقطعة، والكلمات القرآنية، بالشكل الصحيح، كما يتم تلاوتها في القرآن الكريم وفقاً للقاعدة النورانية. يمكن ان يتحقق كذلك من صحة النطق، وبالتالي يساعد الناس في تلاوة القرآن بشكل صحيح. تُستخدم مجموعة البيانات المقترحة الجديدة لتدريب نموذج CNN فعال للتعلم العميق مُدرَّب مسبقاً من خلال نقل التعلم لقراءة الشفاه، مما يحقق متوسط دقة يبلغ ٨٣٪ لجميع فئات مجموعات البيانات. يحاول هذا العمل استغلال الطبيعة المنهجية للتلاوة والترنيم القرآني (فن النطق الصحيح للحروف) من خلال نمذجة في مهمة التعرف على الكلام المرئي. أخيراً، تتم مقارنة نتائج زوايا الرؤية الأكثر شيوعاً لقراءة الشفاه من جوانب مختلفة، وتتم مناقشة تناسق مجموعة البيانات والتحديات واختتامها بالعديد من الاتجاهات الجديدة الواعدة للعمل المستقبلي .

Visual Lip-Reading for Arabic Alphabets and Quranic Words using Deep Learning

by

Nada Faisal Aljohani

Advisor

Dr. Emad Sami Jaha

Abstract

The continuing advances in deep learning have paved the way for several challenging ideas. One such idea is visual lip-reading, which has recently drawn many research interests. Lip-reading is often referred to as visual speech recognition, and it is the ability to understand and predict spoken speech based solely on lip movements without using sounds. Due to the lack of research studies on visual speech recognition for the Arabic language in general, and its absence in the Quranic research, this thesis aims to fill this gap. The work in this thesis introduces a new publicly available Arabic lip-reading dataset comprising 10490 videos captured from multiple viewpoints and data samples in two letter levels and in the word level based on the content and context of a Quranic study aid given in Al-Qaida Al-Noorania book. Furthermore, this work uses visual speech recognition of spoken Arabic letters (Arabic alphabets), Quranic disjoined letters, and Quranic words, mainly phonetic, as they are recited in the Holy Quran according to Al-Qaida Al-Noorania. It could further validate the correctness of pronunciation and, subsequently, assist people in correctly reciting the Quran. The new proposed dataset is used to train an effective pre-trained deep learning CNN model throughout transfer learning for lip-reading, achieving an average accuracy of 83% for all dataset categories. This work attempts to exploit the systematic nature of the Quranic recitation and intonation (the art of proper pronunciation of letters) by modeling them in the visual speech recognition task. Finally, the results of the most common view angles for lip-reading are compared from different aspects, and dataset collection consistency and challenges are discussed and concluded with several new promising trends for future work.