

استرجاع صور المخطوطات العربية باستخدام التعلم العميق

منال محمود علي خياط

باشراف
د. لمياء عبدالله الرفاعي

المستخلص

توفر المخطوطات العربية القديمة قيمة من المعلومات التاريخية التي تعكس التعليم وثقافة المجتمع وتقاليدده خلال فترات زمنية محددة. لذلك نظراً لأهميتها ودورها الأساسي في إثراء المعلومات التاريخية القيمة، تهدف هذه الدراسة إلى جمع المخطوطات العربية القديمة في مجموعة بيانات. ثم، تصنيف صورها حتى تتمكن من استرداد أكثر الصور مشابهة لصورة الاستعلام بدقة وبشكل لحظي. يمكن بناء نظام الاسترجاع وفقاً لمعايير بحثية مختلفة. استوفينا في هذه الدراسة الاسترجاع حسب ثلاث معايير وهي استرجاع الصور حسب المخطوطه وحسب المؤلف وحسب نوع الخط. يعد الاستخراج والتصنيف التلقائي وفقاً لأهم السمات المميزة للصور، خطوة حاسمة لاكتشاف أوجه التشابه بين الصور. مع العلم أن المخطوطات العربية التاريخية هي صور نصية. وبالتالي، من المهم استخراج النص من الصور واسترجاع الصور وفقاً لخصائصها النصية. يتم تنفيذ هذه الخطوه من خلال تطوير نموذج التعلم العميق LSTM ثنائي الاتجاه والمحسن. بعد ذلك، يتم قياس أوجه التشابه بين النصوص باستخدام ثلاثة مقاييس مختلفة للمسافة. صور المخطوطات ليست نصية بحتة، عوضاً عن ذلك فإنها تشمل توقيعات ورسومات وأشكال وجدول وملاحظات جانبية، إلخ. وبالتالي، من الضروري الإهتمام بالأجزاء غير النصية في الصور واستعادة الصور وفقاً لخصائصها المرئية.

لتحقيق ذلك، نقلنا التعلم من أربع شبكات عصبية تلافيفية مدربة مسبقاً تسمى MobileNetV1 و ResNet50 و VGG19 كما تم اختبار نموذج التعلم العميق السيامي إلى جانب مقاييس المسافة الثلاثة لقياس أوجه التشابه بين الصور واسترجاعها.

وأخيراً، تم دمج نماذج التعلم العميق المرئيه والنصيه الأكثر دقه بإستخدام ثلاثة نماذج مختلفة للانصهار تسمى: مستوى القرار ومستوى الميزات ومستوى النتيجة . حسن نموذج الانصهار على مستوى النتيجة من كل نموذج مستخدم بشكل فردي .

كما وتم تطوير واختبار أداء نموذج تعليم معزز لاسترجاع صور المخطوطات العربية . ومن ثم، تطرح هذه الدراسه نهجاً جديداً يوصي بتطبيق نموذج انصهار لكلا النماذج المرئيه والنصيه لتصنيف واسترجاع صورالمخطوطات العربية بدقه ونجاح .

Arabic Manuscript Images Retrieval Using Deep Learning

By
Manal Mahmoud Ali Khayyat

Supervised By
Dr. Lamiaa Elrefaei

ABSTRACT

The ancient Arabic manuscripts are valuable pieces of historical information that reflect the education, culture, society, and tradition during specific time periods. Therefore, due to their significance and substantial role in enriching the valuable historical information, this study aims to collect the ancient Arabic manuscripts in a dataset. Then, classify its images to be able to retrieve the ranked top similar images to a user query image accurately and instantaneously. The retrieval system can be according to different search criteria. We satisfied the retrieval according to the manuscript, author, and the calligraphy.

The automatic extraction and classification according to the most distinguishable features, is a crucial step to detect the similarities among images successfully. Considering that, the historical Arabic manuscripts are text-based images. Thus, it is important to extract the text from the images and to retrieve them according to their textual features. This step is performed through developing an optimized bidirectional LSTM deep learning model including attention and batch normalization layers. Then, the similarities among the textual contents measured using three different distance metrics named: Manhattan, Euclidean, and Cosine.

The manuscripts' images are also not purely textual. Instead, they include handwritten signatures, drawings, figures, tables, side-notes, ...etc. Thereby, it is necessary to consider the non-textual parts within the images and to retrieve them according to their visual features. To accomplish this, we transferred learning from four pre-trained convolutional neural networks named: MobileNetV1, DenseNet201, ResNet50, and VGG19. The Siamese deep learning model along with the three distance metrics tested for measuring the similarities among the images and retrieve them.

The most accurate visual and textual deep learning models were fused at three different fusion-levels named: decision-level, features-level, and score-level. The score-level fusion model resulted in a considerable improvement of each model used individually.

Moreover, we experimented an alternative solution for the images' retrieval utilizing an enhanced deep reinforcement learning technique. The model uses the features-level fusion of both the VGG19 and the BiLSTM deep learning models. Then, it hashes the fused features to reduce the dimensionalities among them.

Thus, the study introduces a novel approach that utilizes a fusion model of both the visual and the textual models to classify and retrieve the historical Arabic manuscripts' images successfully.