World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:16, No:11, 2022

Efficient Layout-Aware Pretraining for Multimodal Form Understanding

Authors : Armineh Nourbakhsh, Sameena Shah, Carolyn Rose

Abstract : Layout-aware language models have been used to create multimodal representations for documents that are in image form, achieving relatively high accuracy in document understanding tasks. However, the large number of parameters in the resulting models makes building and using them prohibitive without access to high-performing processing units with large memory capacity. We propose an alternative approach that can create efficient representations without the need for a neural visual backbone. This leads to an 80% reduction in the number of parameters compared to the smallest SOTA model, widely expanding applicability. In addition, our layout embeddings are pre-trained on spatial and visual cues alone and only fused with text embeddings in downstream tasks, which can facilitate applicability to low-resource of multi-lingual domains. Despite using 2.5% of training data, we show competitive performance on two form understanding tasks: semantic labeling and link prediction.

Keywords : layout understanding, form understanding, multimodal document understanding, bias-augmented attention
Conference Title : ICDAR 2022 : International Conference on Document Analysis and Recognition
Conference Location : Jerusalem, Israel
Conference Dates : November 29-30, 2022