

we address the problem of predicting similarity between a pair of handwritten document images written by different individuals. This has applications related to matching and mining in image collections containing handwritten content. A similarity score is computed by detecting patterns of text re-usages between document images irrespective of the minor variations in word morphology, word ordering, layout and paraphrasing of the content. Our method does not depend on an accurate segmentation of words and lines. We formulate the document matching problem as a structured comparison of the word distributions across two document images. To match two word images, we propose a convolutional neural network (CNN) based feature descriptor. Performance of this representation surpasses the state-of-the-art on handwritten word spotting. Finally, we demonstrate the applicability of our method on a practical problem of matching handwritten assignments.

We address the problem of predicting similarity between a pair of handwritten document images written by different individuals. This has applications related to matching and mining in image collections containing handwritten content. A similarity score is computed by detecting patterns of text re-usages between document images irrespective of the minor variations in word morphology, word ordering, layout and paraphrasing of the content. Our method does not depend on an accurate segmentation of words and lines. We formulate the document matching problem as a structured comparison of the word distributions across two document images. To match two word images, we propose a convolutional neural network (cnn) based feature descriptor. Performance of this representation surpasses the state-of-the-art on handwritten word spotting. Finally, we demonstrate the applicability of our method on a practical problem of matching handwritten assignments.