

## An end to end system for subtitle text extraction from movie videos

**Hossam Elshahaby, Mohsen Rashwan**

Cairo University, Egypt

A new technique for text detection inside a complex graphical background, its extraction, and enhancement to be easily recognized using the optical character recognition (OCR). The technique uses a deep neural network for feature extraction and classifying the text as containing text or not. An Error Handling and Correction (EHC) technique is used to resolve classification errors. A Multiple Frame Integration (MFI) algorithm is introduced to extract the graphical text from its background. Text enhancement is done by adjusting the contrast, minimize noise, and increasing the pixels resolution. A standalone software Component-Off-The-Shelf (COTS) is used to recognize the text characters and qualify the system performance. Generalization for multilingual text is done with the proposed solution. A newly created dataset containing videos with different languages is collected for this purpose to be used as a benchmark. A new HMVGG16 Convolutional Neural Network (CNN) is used for frame classification as text containing or non-text containing, has accuracy equals to 98%. The introduced system weighted average caption extraction accuracy equals to 96.15%. The Correctly Detected Characters (CDC) average recognition accuracy using the Abbyy SDK OCR engine equals 97.75%.

### Biography

Hossam Elshahaby is affiliated to Cairo University, Egypt. He is a recipient of many awards and grants for his valuable contributions and discoveries in major area of Artificial Intelligence. His international experience includes various programs, contributions and participation in different countries for diverse fields of study. His research interests reflect in his wide range of publications in various national and international journals in Big Data.