

**CHENNAI – PONDICHERRY**

**BreakingNews: Article Annotation by Image and Text Processing**

**Abstract**

Building upon recent Deep Neural Network architectures, current approaches lying in the intersection of Computer Vision and Natural Language Processing have achieved unprecedented breakthroughs in tasks like automatic captioning or image retrieval. Most of these learning methods, though, rely on large training sets of images associated with human annotations that specifically describe the visual content. In this paper we propose to go a step further and explore the more complex cases where textual descriptions are loosely related to the images. We focus on the particular domain of news articles in which the textual content often expresses connotative and ambiguous relations that are only suggested but not directly inferred from images. We introduce an adaptive CNN architecture that shares most of the structure for multiple tasks including source detection, article illustration and geolocation of articles. Deep Canonical Correlation Analysis is deployed for article illustration, and a new loss function based on Great Circle Distance is proposed for geolocation. Furthermore, we present BreakingNews, a novel dataset with approximately 100K news articles including images, text and captions, and enriched with heterogeneous meta-data (such as GPS coordinates and user comments). We show this dataset to be appropriate to explore all aforementioned problems, for which we provide a baseline performance using various Deep Learning architectures, and different representations of the textual and visual features. We report very promising results and bring to light several limitations of current state-of-the-art in this kind of domain, which we hope will help spur progress in the field.