The Relation between Texts and Images in News: News Images in MediaEval 2023

Andreas Lommatzsch[1,†], Benjamin Kille[2], Özlem Özgöbek[2], Mehdi Elahi[3] and Duc-Tien Dang-Nguyen[3]

[1]*Technische Universität Berlin, Berlin, Germany*
[2]*Norwegian University of Science and Technology, Trondheim, Norway*
[3]*University of Bergen, Bergen, Norway*

Abstract

News articles typically consist of text and images. Images plays a crucial role in catching the user's attention and emphasising the article's message. For each news text, the editor must select the best photos from the available set of recent photos, archived photos or stock images to both attract user's attention and best fitting with the news article text. The NewsImages benchmark aims to shed a light on this real-world relation between news texts and the accompanying images. The task provides datasets and evaluation components for studying this relation. The datasets includes AI-generated images as an additional research challenge.

This paper describes the NewsImages task in detail, giving the explanations for the dataset and evaluation metrics. It also discusses the connections to existing research and the addressed challenges.

1. Introduction

In the fast-paced world of digital journalism, news articles are inherently multi-modal, seamlessly intertwining text and images to convey information. Among the various components of a news article, images occupy a pivotal role. They not only serve as a visual aid; they catch the readers' interest, compelling them to delve into the text. Furthermore, images reinforce the central message of the article, often providing context or offering a visual perspective that words alone might fail to capture. With the rise of generative artificial intelligence, there has been a shift towards automating news article creation. This automation includes the generation of text and images that align perfectly with the content.

NewsImages task aims to support the research in understanding the relationship between news texts and their accompanying images on news portals. This relationship is full of challenges. The vast expanse of news topics, the diversity in domains, the plethora of news portals, and the myriad styles of news articles, all culminate in a complex web of considerations when matching text with images. Delving deeper into the scenario, NewsImages is driven by several pertinent questions: How can the connection between texts and images in news articles be re-established?

Multimedia Evaluation Workshop, 1–2 Feb. 2024, Amsterdam, Netherlands
[†] Corresponding author.
andreas.lommatzsch@dai-labor.de (A. Lommatzsch); benjamin.u.kille@ntnu.no (B. Kille); ozlem.ozgobek@ntnu.no (Ö. Özgöbek); mehdi.elahi@uib.no (M. Elahi); ductien.dangnguyen@uib.no (D. Dang-Nguyen)

To what extent do generated images alter this re-establishment? Are there discernible patterns or principles that guide editors when they select images for news pieces? And, in the grand scheme of automated news generation, are there innovative methods to generate better-suited images for given news texts?

2. Background and Related Work

Deciphering the relationship between text and images in news articles is an important task for understanding both the creation and the perception of content in the news sector. The depiction gap between texts and images is a major problem [1]. Significant advances in image comprehension have recently been made through deep neural networks, enabling systems not only to detect intricate concepts within images but also to identify pertinent objects with high precision. The embedding of concepts extracted from images and texts within a unified vector space is central to this advancement, facilitating nuanced correlations. While there are multiple datasets tailored for optimizing learning strategies in image labeling (e.g. *MS COCO* [2]), a new frontier lies in generative AI's capability to produce high-resolution images from text descriptors. The landscape of news imagery, dominated by stock photos, portraits, and loosely related archival images, presents unique challenges, often accentuated by the absence of directly relevant visuals. This inspires the pivotal research question: How are images and text interconnected in the context of news? Furthermore, this opens a broader inquiry into AI's potential role in enhancing news article formulation, opening avenues for automated, contextually appropriate visual representation.

For the 5th time, the NewsImages challenge explores the aspects of multimedia content in news. The first editions (NewsREEL Multimedia [3, 4, 5]) focused on predicting the popularity of news items based on multimedia content. In 2021, the focus shifted to understanding the relationship between text and images [6]. In 2023, we extend the task by adding AI generated images to further explore the relation between news and AI generated images. The NewsImages task is related to several research topics, such as multi-modal recommender systems [7, 8, 9], the detection of fake news [10], and multi-modal embedding methods [11]. The task supports the research toward multi-modality in different news related domains.

3. Task Description

The NewsImages benchmark investigates the connection between textual news content and associated imagery. This year's task draws its data from two distinct news dissemination channels: official publishers' portals and RSS feeds. Participants are provided with a comprehensive training dataset, encompassing linked text-image pairs, complemented by a test dataset with disassociated pairs. The challenge mandates the development and critical evaluation of innovative methodologies to accurately re-associate news articles with corresponding images. The dataset represents a challenge with instances of images, such as conceptual stock photographs, potentially aligning with multiple articles. Participants are required to submit a prioritized list of plausible image matches, with the evaluation metric favoring early correct re-associations.

4. Dataset

NewsImages provides a dataset comprising three parts built on news from news portals and an RSS news feed. As source for the crawled web sites, we use the GDELT project (https://www.gdeltproject.org/) that aggregates news from all over the world. For the RT part the RSS feed `rtde`[1] has been used. The dataset has been created using the following three steps: **(1) Crawling:** Crawl news items from the selected sources and eliminate news articles that do not consist of an image and a suitable text. We use news items published in the period November 2022–August 2023. For the GDELT part the news title and the entities (extracted by GDELT from the news text for creating knowledge graphs) are used (http://data.gdeltproject.org/gkg/index.html. For the RT part, the news title and the snippet (both German originals and English machine translation of these fields) are used. **(2) Cleaning:** For ensuring the quality of images, we use different heuristics for removing duplicates, low quality images, and logos. In addition, we remove images mainly consist of text. **(3) Image generation:** For studying the problem of matching generated images we use Stable Diffusion. We use the news article's headline as the prompt. The generated images are used to replace some of the original images. In the three parts of the dataset, the fraction of generated images differs. Part GDELT-P1 does not contain any generated images; GDELT-P2 contains 80% generated images, and RT has 50% generated images. **(4) Splitting:** Each part of the dataset is split into a training and a test set as Table 1 illustrates.

Batch	Source	Purpose	No. Cases
GDELT-P1-a	Web sites	Training	8500
GDELT-P1-b	Web sites	Test	1500
GDELT-P2-a	Web sites	Training	12 041
GDELT-P2-b	Web sites	Test	1500
RT-a	RSS Feed	Training	9755
RT-b	RSS Feed	Test	3000

Table 1: Dataset Statistics. The dataset comes in six batches. The number of cases refers to the article-image pairs.

The data set contains information related to articles and images. Articles' metadata include the URL, title, and a text snippet (RT batch) or the entities extracted from the news text (GDELT batch). Image captions or image filenames must not be used in the task.

5. Evaluation

The NewsImages benchmark is designed to analyze the relation between news texts and the accompanying images. As a concrete task, the participants must assign a matching image to for each news text in the given test set. Concretely, for each news article an ordered list of 100 images must be submitted. The participants provide a text file that provides a tab separated list of 100 image IDs for each news article ID.

The participants' submissions are evaluated against a ground truth defined by the originally crawled connection between the images and the text. The ground truth ensures that a 1:1 relation between the images and the texts exists.

5.1. Evaluation Metric

The participants' submissions are evaluated using the Mean Reciprocal Rank (MRR) [12] as the main evaluation criteria. MRR is defined as $\mathrm{MRR} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\mathrm{rank}(x_n)}$, where $\mathrm{rank}(x_n)$ returns

[1] http://de.rt.com/feeds/news/

the rank at which the matching image was listed. The earlier the matching image appears on average, the higher the score. The Mean Reciprocal favors the top of the list and penalizes finding a match further down.

In addition to MRR, we also compute the Average Recall (AR) at ranks N for $N \in \{1, 5, 10, 20, 50, 100\}$. AR computes the average over the recall scores calculated for each news article. The evaluation scores are computed separately for each batch.

5.2. Run Description

Participants are encouraged to contribute working notes that elucidate their innovative concepts, fostering an in-depth exploration of the intricate relationship between textual content and images in news media. In this pursuit, participants have the opportunity to submit a maximum of five runs for each of the three test datasets. Each run entails a set of predictions tailored to these test datasets. We encourage participants to engage in a comprehensive comparative analysis of their various runs, encompassing assessments of quality, computational complexity, and resource utilization.

Furthermore, the discussion of results should be characterized by a nuanced consideration of the datasets' idiosyncrasies, illuminating how the discoveries made can be extrapolated to diverse scenarios. To culminate, participants are expected to articulate their insights and reflect on their potential contributions towards advancing cutting-edge research in this field.

6. Conclusion

The linking between news texts and images is still a complicated problem due to the news domain's diversity, editors' habits, and readers' expectations. The mixture of real photos, stock images, archived photos, and AI generated images makes it very challenging to extract not only concepts from images but also to understand the principles applying when selecting the images. The NewsImages challenge provides a medium-sized, real-world data set for investigating the existing principles for connecting images and texts. Participants can develop, optimize, and evaluate innovative re-matching methods for news texts and images.With the growing popularity and enhancement of AI methods for generating images, images that are more representative of the text could replace the partially matching images like stock photos. These artificial images could be used to reinforce the credibility of fake news but also avoid misinterpretation of news caused by ill-fitted stock images. Thus, understanding the relation between news texts and images remains a highly relevant and challenging research topic. News Images provides the foundation to foster the development and evaluation of innovative approaches.

Acknowledgments

We gratefully thank Marc Gallofré Ocaña and Sohail Ahmed Khan for supporting the data set creation. We acknowledge the contributions of the GDELT (https://www.gdeltproject.org/) project for providing the data which made the dataset creation possible.

References

[1] A. Lommatzsch, B. Kille, O. Özgöbek, Y. Zhou, J. Tešić, C. Bartolomeu, D. Semedo, L. Pivovarova, M. Liang, M. Larson, NewsImages: Addressing the Depiction Gap with an Online News Dataset for Text-Image Rematching, in: Proceedings of the 13th ACM Multimedia Systems Conference, MMSys '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 227–233. URL: https://doi.org/10.1145/3524273.3532891.

[2] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft COCO: Common Objects in Context, in: European Conference on Computer Vision, Springer, 2014, pp. 740–755. doi:10.1007/978-3-319-10602-1_48.

[3] A. Lommatzsch, B. Kille, F. Hopfgartner, L. Ramming, MediaEval 2018 - Overview on NewsREEL Multimedia, in: Proceedings of the MediaEval Benchmarking Initiative for Multimedia Evaluation 2018, CEUR Workshop Proceedings, 2018. URL: http://ceur-ws.org/Vol-2283/.

[4] Y. Deldjoo, B. Kille, M. Schedl, A. Lommatzsch, J. Shen, The 2019 Multimedia for Recommender System Task: MovieREC and NewsREEL at MediaEval, in: Procs. of the MediaEval Benchmarking Initiative for Multimedia Evaluation 2019, CEUR WS Procs., 2019. URL: http://ceur-ws.org/Vol-2670/.

[5] B. Kille, A. Lommatzsch, O. Özgöbek, NewsImages: The Role of Images in Online News, in: Proceedings of the MediaEval Benchmarking Initiative for Multimedia Evaluation 2020, CEUR Workshop Proceedings, 2020. URL: http://ceur-ws.org/Vol-2882/.

[6] B. Kille, A. Lommatzsch, Ö. Özgöbek, M. Elahi, D.-T. Dang-Nguyen, News Images in MediaEval 2021, in: Proceedings of the MediaEval Benchmarking Initiative for Multimedia Evaluation 2021, CEUR Workshop Proceedings, 2021. URL: http://ceur-ws.org/Vol-3181/paper2.pdf.

[7] A. Salah, Q.-T. Truong, H. W. Lauw, Cornac: A Comparative Framework for Multimodal Recommender Systems., J. Mach. Learn. Res. 21 (2020) 95–1.

[8] S. Oramas, O. Nieto, M. Sordo, X. Serra, A Deep Multimodal Approach for Cold-start Music Recommendation, in: Procs. of the WS on Deep Learning for Recommender Systems, 2017, pp. 32–37.

[9] Y. Deldjoo, M. Schedl, P. Cremonesi, G. Pasi, Recommender Systems Leveraging Multimedia Content, ACM Computing Surveys (CSUR) 53 (2020) 1–38.

[10] X. Zhou, R. Zafarani, A Survey of Fake News, ACM Computing Surveys 53 (2020) 1–40. URL: http://dx.doi.org/10.1145/3395046. doi:10.1145/3395046.

[11] L. Cui, S. Wang, D. Lee, SAME: Sentiment-Aware Multi-Modal Embedding for Detecting Fake News, in: Pros of the 2019 Intl. Con. on Advances in Social Networks Analysis and Mining, ASONAM '19, ACM, New York, NY, USA, 2020, p. 41–48. doi:10.1145/3341161.3342894.

[12] E. M. Voorhees, et al., The TREC-8 Question Answering Track Report., in: TREC, volume 99, 1999, pp. 77–82.