

# News Images in MediaEval 2022

Benjamin Kille<sup>1,\*</sup>, Andreas Lommatzsch<sup>2</sup>, Özlem Özgöbek<sup>1</sup>, Mehdi Elahi<sup>3</sup> and Duc-Tien Dang-Nguyen<sup>3,4</sup>

<sup>1</sup>Norwegian University of Science and Technology, Norway

<sup>2</sup>Berlin Institute of Technology, Germany

<sup>3</sup>University of Bergen, Norway

<sup>4</sup>Kristiania University College, Norway

## Abstract

Images play a crucial role in online news perception. Images catch users' attention and strongly affect how they interpret the news. Images serve different roles, e.g., visualizing a scene discussed in the text, highlight certain aspects by a stock photo, or showing archived footage of a relevant person or organization. News Images as part of MediaEval 2022 aims to gain more insight into the interplay of images and texts in different news domains. Participant access a large set of articles and accompanying images collected from general online news portals and an RSS-based news stream. In contrast to NewsImages 2021 data come from different news sources. Thus, this year's task facilitates comparing image usage on different portals and analyzing transfer learning strategies. This paper describes the NewsImages task, explains the dataset and evaluation metrics. It draws connections to existing research.

## 1. Introduction

Publishers present news as a multimodal mix: images, video clips, and soundbites accompany the textual content. Imagery attracts readers' attention and illustrates aspects of the textual body. Research both on multimedia and personalization has previously assumed a simple relationship between the modalities. For instance, image captioning [1] models the task as quite literally describing the image's scenery. Contrarily, Oostdijk et al. [2] finds that the relationship is much more complicated. The News Images task at MediaEval 2022 investigates the relationship with real-world data seeking to better understand its implications for journalism and news personalization.

Participants have access to news from three different sources: Publishers' websites, RSS feeds, and social media. For each kind of source, the data presents a training portion and an evaluation set for which the link between text and image has been removed. Participants have to re-match the articles and images. Thereby, the task addresses questions such as: What makes an image appealing as depiction of news events? How do editors select images? What do readers find most relevant for news images? News Images seeks to surpass conventional research on image concept detection.

---

*MediaEval'22: Multimedia Evaluation Workshop, January 13–15, 2023, Bergen, Norway and Online*


\*Corresponding author.

✉ benjamin.u.kille@ntnu.no (B. Kille); andreas.lommatzsch@dai-labor.de (A. Lommatzsch); ozlem.ozgobek@ntnu.no (Özgöbek); mehdi.elahi@uib.no (M. Elahi); ductien.dangnguyen@uib.no (D. Dang-Nguyen)

ORCID 0000-0002-3206-5154 (B. Kille)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

## 2. Background and Related Work

The *Multimedia Evaluation Benchmark* task *NewsImages* researches the interesting aspects of multi-media content in the news domain for the fifth time in 2022. In the years 2018-2020, the *NewsREEL Multimedia* [3, 4, 5] focused on predicting the popularity of news items based on multi-media content. In 2021 the focus shifted on understanding the connection between texts and images [6].

The annotation and the understanding of images has been an active research topic in recent years. Several datasets (e.g. *Flickr30k* [7], *MS COCO* [8]) exist, providing textual annotation of images. These datasets are designed for learning methods for automated image labeling and image retrieval. The dataset focus on describing the image content but do neither consider the role of the image in the context nor relation between the image and the surrounding text. The *NewsImages* task focus on researching the relation of images and news articles. *NewsImages* is also strongly related to the news recommendation problem. CLEF *NewsREEL* researched methods for identifying trends in news streams and for providing news recommendations. News recommender systems extract features from news articles and the user behavior for computing highly relevant recommendations. Most existing recommender systems only consider the textual content, making use of traditional Information Retrieval methods and advanced Language Model-based approaches. Text-based recommender systems can efficiently provide good recommendations. Besides, recent years have seen an upward trend concerning research on multimodal recommender systems. For instance, Truong and Lauw [9] investigate how to leverage multimodal user feedback, Salah et al. [10] prepare a framework for multimodal recommender systems and Oramas et al. [11] examine the use of multimodal data for music recommendation. For a comprehensive review, we refer to [12]. The *News Images* task supports the research toward multimodality.

Strongly related with news recommendations is the detection of ‘fake news’ [13]. ‘Fake news’ often put data or images in an misleading context in order to attract users or to achieve an intended perception. Thus, a fine-grained analysis of images and text helps to get a better understanding of this phenomenon. The *NewsImages* task allows to research the use of images in news articles with respect to different news domains.

## 3. Task Description

*News Images* explores the relation between news texts and images. Concretely, the task’s data set considers three news channels: publishers’ news portals, RSS feeds, and social media. For each channel, participants obtain a training set that includes the link between text and image. Besides, they receive an evaluation set with the link removed. Participants’ task is to develop and evaluate ways to re-match news articles and images. The set of images contains some instances that could be related to more than one article. For instance, the editor uses a stock photo capturing the happening more conceptually. Thus, participants can submit an ordered list of image candidates. The evaluation protocol checks the position of the actually linked image and rewards submission with the match early on.

## 4. Dataset

We provide a dataset comprising three different news sources: Online News portals, Twitter, and an RSS news feed. The heterogeneity of the data sources allows us to analyze the relation of images and texts in three different yet similar news domains.

Creating the dataset follows three steps: (1) We crawled news articles from the source. The crawling stretched from March to August 2022. (2) We applied two filters to guarantee data quality. We removed articles with more than 30% non-ASCII characters as well as articles with fewer than 20 characters. (3) We apply a set of rules to determine the best image in cases with multiple images. An expressive image has a reasonable size, sufficient entropy in the color spectrum—which helps us to filter out logos—, contains little text if at all, and ideally is unique. Having filtered the images automatically, an annotator checked the output and assured that inadequate images and logos had been excluded. The filtering succeeded for the analyzed RSS feed. Still, the high variety of images on major news portals necessitated a manual post-filtering.

The data contains information related to articles and images. Articles’ metadata include the URL, title, and a text snippet. The image data consist of a URL and an image hashcode. We do not provide image captions and ask participants not to make use of the image filename.

The data set comprises three batches each consisting of a training and a test set. The test sub batches provide 1,500 elements to simplify the comparison of the results obtained for the three batches. Table 1 illustrates the data set.

**Table 1**

Dataset Statistics. The dataset comes in six batches. The number of cases refers to both articles and images. The test batches have the same size to ensure the comparability of the evaluation results.

Batch	Source	Purpose	No. Cases
<b>Batch 1a</b>	Web sites	Training	2539
<b>Batch 1b</b>	Web sites	Test	1500
<b>Batch 2a</b>	Tweets (Twitter)	Training	2387
<b>Batch 2b</b>	Tweets (Twitter)	Test	1500
<b>Batch 3a</b>	RSS Feed	Training	3292
<b>Batch 3b</b>	RSS Feed	Test	1500

## 5. Evaluation

We want to better understand the interplay between news texts and images. As a proxy, we task participants to re-match texts and images.

Participants obtain a set of unlinked news articles and images. For each article, they have to provide a list of images sorted by match likelihood. We cap the lists at 100 to simplify computation and account for the expectation that editors will not spend time browsing to long lists of images. Each of the three evaluation sets contains 1500 articles and images. Consequently, participants provide a tab-separated file with one column for the article reference followed by 100 columns with image identifiers.

### 5.1. Evaluation Metric

We use Mean Reciprocal Rank (MRR) [14] as the main evaluation criteria. MRR is defined as

$$\text{MRR} = \frac{1}{N} \sum_{n=1}^N \frac{1}{\text{rank}(x_n)},$$

where  $\text{rank}(x_n)$  returns the rank at which the matching image was listed or a very large number such as  $10^{12}$  if the list excludes the matching image. The earlier the matching image appears on

average, the higher the score. The MRR strongly favors the top of the list and penalizes finding a match further down.

Besides, we compute the Average Precision (AP) at ranks  $N$  for  $N \in \{1, 5, 10, 20, 50, 100\}$ . AP lets us investigate whether the predictions are more accurate in some ranges.

## 5.2. Run Description

Participants' working notes inform about their ideas. We encourage participants to explore different ideas to help us better understand the interplay between texts and images in news. Consequently, participants can submit up to five runs for each of the three test sets. Each run consists of predictions for each of the three test sets. We further motivate participants to compare the results of different runs and analyze the findings with respect to quality, computational complexity, and used resources. The discussion of the results should take into consideration the dataset's particularities and explain how findings translate to other scenarios. Finally, participants should describe what they have learned and contemplate how their insights can help to advance the research.

## 6. Conclusion

Understanding the relation between text content and images in news remains a tough challenge. Images have different roles in news such as attracting users, highlighting specific aspects of a message or providing additional context for a news article. Dependent of the concrete news event, the image may depict the news event, shows an old similar scene or depicts persons or objects related to the news text. Understanding the user preferences in consuming images and the publishers policies for selecting images helps to research to induced news perception and intentions.

Consequently, understanding the relation between news texts and images can give publishers an competitive advantages, help to detect fake news and click bating, as well as pave to way for the penalization of news.

**Acknowledgements** We would like to thank Marc Gallofré Ocaña for kindly supporting the providing real world data. Further, we thank Martha Larson for her support.

## References

- [1] M. Z. Hossain, F. Sohel, M. F. Shiratuddin, H. Laga, A comprehensive survey of deep learning for image captioning, *ACM Comput. Surv.* 51 (2019). URL: <https://doi.org/10.1145/3295748>. doi:10.1145/3295748.
- [2] N. Oostdijk, H. van Halteren, E. Basar, M. Larson, The connection between the text and images of news articles: New insights for multimedia analysis, in: *Proceedings of The 12th Language Resources and Evaluation Conference*, 2020, pp. 4343–4351.
- [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: *Proceedings of the MediaEval Benchmarking Initiative for Multimedia Evaluation 2019*, CEUR Workshop Proceedings, 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] P. Young, A. Lai, M. Hodosh, J. Hockenmaier, From Image Descriptions to Visual Denotations: New Similarity Metrics for Semantic Inference over Event Descriptions, Transactions of the Association for Computational Linguistics 2 (2014) 67–78. doi:10.1162/tacl\_a\_00166.
- [8] 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.
- [9] Q.-T. Truong, H. Lauw, Multimodal review generation for recommender systems, in: The World Wide Web Conference, 2019, pp. 1864–1874.
- [10] A. Salah, Q.-T. Truong, H. W. Lauw, Cornac: A comparative framework for multimodal recommender systems., J. Mach. Learn. Res. 21 (2020) 95–1.
- [11] S. Oramas, O. Nieto, M. Sordo, X. Serra, A deep multimodal approach for cold-start music recommendation, in: Proceedings of the 2nd workshop on deep learning for recommender systems, 2017, pp. 32–37.
- [12] Y. Deldjoo, M. Schedl, P. Cremonesi, G. Pasi, Recommender systems leveraging multimedia content, ACM Computing Surveys (CSUR) 53 (2020) 1–38.
- [13] 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.
- [14] E. M. Voorhees, et al., The TREC-8 Question Answering Track Report., in: TREC, volume 99, 1999, pp. 77–82.
- [15] F. Corsini, M. Larson, CLEF NewsREEL 2016: Image based Recommendation, in: Working Notes of the 7th International Conference of the CLEF Initiative, Evora, Portugal, CEUR Workshop Proceedings, 2016.
- [16] A. Lommatzsch, B. Kille, F. Hopfgartner, M. Larson, T. Brodt, J. Seiler, Ö. Özgöbek, CLEF 2017 NewsREEL overview: A stream-based recommender task for evaluation and education, in: 8th International Conference of the CLEF Association: Experimental IR Meets Multilinguality, Multimodality, and Interaction (CLEF 2017), Springer, 2017.
- [17] M. Ge, F. Persia, A survey of multimedia recommender systems: Challenges and opportunities, International Journal of Semantic Computing 11 (2017) 411–428. URL: <https://doi.org/10.1142/S1793351X17500039>. doi:10.1142/S1793351X17500039.
- [18] M. Karimi, D. Jannach, M. Jugovac, News recommender systems—survey and roads ahead, Information Processing & Management 54 (2018) 1203–1227.
- [19] Ö. Özgöbek, B. Kille, J. A. Gulla, A. Lommatzsch, The 7th international workshop on news recommendation and analytics (inra 2019), in: Proceedings of the 13th ACM Conference on Recommender Systems, 2019, pp. 558–559.
- [20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 248–255.
- [21] 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/>.