Multimodal Geolocation Estimation in News Documents

ESR 8: Golsa Tahmasebzadeh

golsa.tahmasebzadeh@tib.eu

TIB - Leibniz Information Centre for Science and Technology & L3S Research Center, Leibniz University of Hannover
Hannover, Germany

Motivation

- In news analysis, geographical content is essential (e.g., in fake news detection)
- News photos usually lack geographically representative content
- News body text indicates possible photo locations (Fig. 1)

We study geolocation estimation in two directions:

- Focus location of the news story [1]
- Geolocation of the news photo [2]

Photo by Terra Eclipse (CC BY 2.0)

Fig. 1: Image lacks geo-representative content

[...] **John Boehner** accused President **Barack Obama** [...] **Boehner** and the **House of Representatives** leadership [...] to give **Obama** power to raise **U.S.** [...] **U.S.** credit [...]

Washington, D.C., U.S, N.Amerika

We propose a demo that shows the applications of geolocation estimation of photos [3]

Related Work

Drawbacks of existing approaches and datasets for geolocation estimation

- Existing methods
  - Are either based on the visual content [4]
  - Do not use state-of-the-art methods for multimodal information extraction [5]
- Existing datasets
  - Provide only images and are not related to news [6]
  - Contain unreliable ground truth labels [7]

We propose **novel multimodal solutions for geolocation estimation** in news documents

Multimodal Geolocation Estimation of News Photos

We propose the **MMG-NewsPhoto dataset**

- Image-text pairs from news labeled for multimodal photo geolocation
- 617,920 data samples covering 14,331 cities, 241 countries, and 6 continents

Multimodal approach composed of three modules (Fig. 2):

- Image encoder: based on a powerful backbone CLIP [8]
- Text encoder: global contextual and entity-centric embeddings based on BERT [9]
- Granularity classifier: produces output probabilities for the city, country and continent

Experimental results demonstrate that

- Multimodal architecture outperforms all the unimodal and multimodal baselines
- Advanced representations needed for the concepts event, group of people & person (Fig. 5)
- Visual model succeeds for photos depicting a strong concept or multiple weak concepts (Fig. 3)
- Multimodal model succeeds if text provides rich information (entities/content) (Fig. 4)

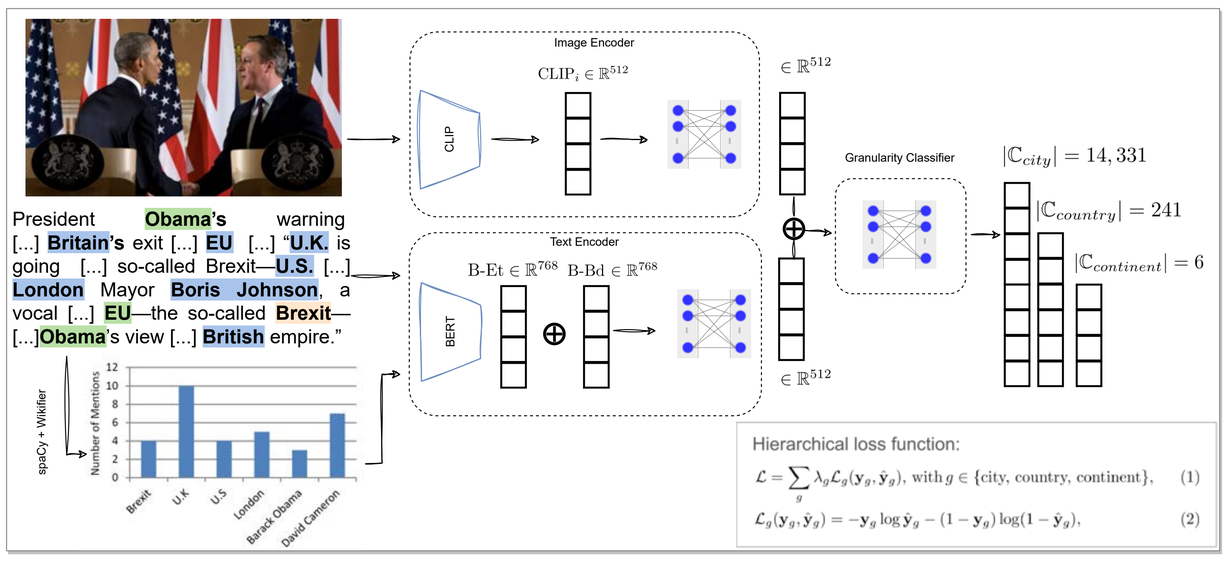

Fig. 2: Model architecture for multimodal geolocation estimation of photos

Photo by NATO (CC BY-SA 2.0)

[...] commanders in **Afghanistan** [...] my most recent visit to **Afghanistan** and [...] **Colonel Jane Crichton,** [...] **U.S**. forces [...] two **U.S.** generals one **Canadian** general [...] the **European** Union had withheld [...]

Afghanistan, Asia

Output predictions

Visual: **Afghanistan, Asia**
Textual: U.S, N.A
Multimodal: U.S, N.A

Fig. 3: Visual model performs better than the unimodal models

Photo by Tryfon Topalidis (CC BY-SA 3.0)

[...] across the **Philippines island** [...] the **Philippine Institute of Volcanology and Seismology (PHIVOLCS)** said lava eruptions[...] **Purita Araojo**, at **Vista Al Mayon Pensionne** [...]

Philippines, Asia

Output predictions

Visual: Indonesia, **Asia**
Textual: **Philippines, Asia**
Multimodal: **Philippines, Asia**

Fig. 4: Textual and multimodal models outperform

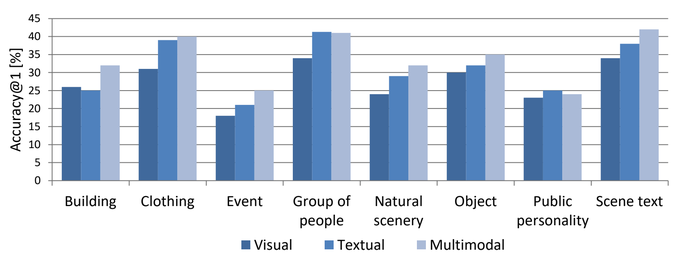

Fig. 5: Accuracy@1 [%] of the best performing models

MM-Locate-News: Multimodal Focus Location Estimation in News

We propose the **MM-Locate-News dataset**

- Which contains more than 6000 image-text pairs
- Labeled with the focus location of news documents

The proposed **multimodal approach**

- Integrates various visual and textual features based on powerful backbones such as BERT [9], CLIP [8], and ResNet [10]
- The output prediction relies on the most confident modality

Experimental results demonstrate that

- The multimodal architecture outperforms all the unimodal approaches
- Multimodal architecture is beneficial when image lacks geo-representative content (Fig. 7) or text mentions various named entities (e.g., location and person) (Fig. 8)

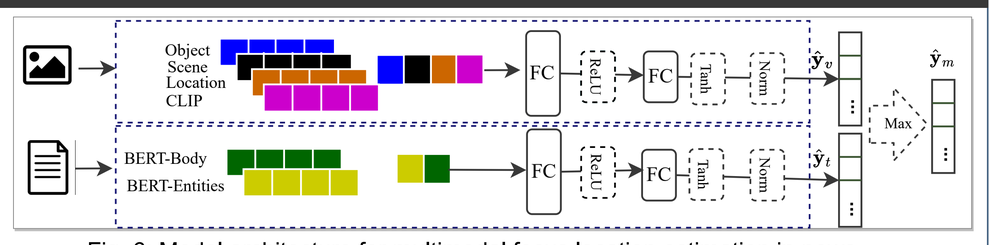

Fig. 6: Model architecture for multimodal focus location estimation in news

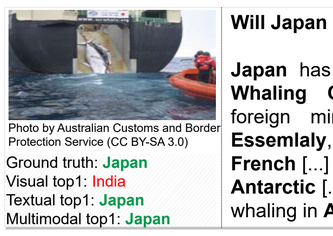

Photo by Australian Customs and Border Protection Service (CC BY-SA 3.0)

Ground truth: **Japan**
Visual top1: India
Textual top1: **Japan**
Multimodal top1: **Japan**

Will Japan ever stop hunting whales?

Japan has pulled out of the **International Whaling Commission**, [...] the **French** foreign ministry has warned [...] **Lamy Essemlaly**, president of **Sea Shepherd's French** [...] **Japan** has been using to whale in **Antarctic** [...] **Australia** and **New Zealand** by whaling in **Antarctic** waters [...] **Japan** [...]

Fig. 7: Image lacks geo-representative content

Photo by Bahnfrend (CC BY-SA 4.0)

Ground truth: **Sri Lanka**
Visual top1: **Sri Lanka**
Textual top1: Pakistan
Multimodal top1: **Sri Lanka**

Pakistan vs Sri Lanka, 2nd Test, Day 4, Live Cricket Score Updates: PAK in cruise control?

Pakistan ended [...] from **Abid Ali** and **Shan Masood**. [...] **Pakistan** will hope to get their lead [...] **Sri Lanka** with **Azhar Ali** [...] **Babar Azam** in **Karachi** on Sunday. [...] **Haris Sohail**, **Asad Shafiq**, **Mohammad Rizwan** (wk), **Yasir Shah** [...]

Fig. 8: Text mentions various named entities

GeoWINE: Geolocation based Wiki, Image, News and Event Retrieval

- GeoWINE closes the gaps between
  - Geolocation estimation
  - Information representation in knowledge graphs (Wikidata and OEKG)
  - Information retrieval
- GeoWINE is an event and news retrieval system based on image-based geolocation estimation
- Beneficial in many downstream tasks such as
  - Image verification
  - Places recommendation
  - Fact-checking

Demo Link

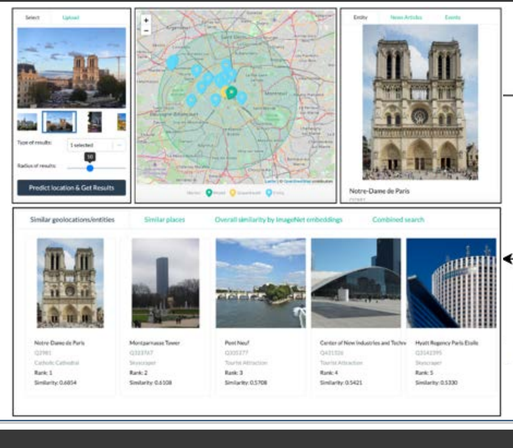

Fig. 9: Overview of the GeoWINE demonstrator

Conclusion & Future Work

Conclusion

- Novel datasets called MMG-NewsPhoto & MM-Locate-News
- Promising results using multimodal approaches for geolocalization of news documents
- A geolocation-based multimodal retrieval system

Future Work

- Extract individual visual concepts (e.g., events), including scene text (e.g., street signs)
- Multimodal geolocation explainability
- Study impact of photo geolocation on fake news detection and news recommendation

References

[1] Golsa Tahmasebzadeh, Sherzod Hakimov, Ralph Ewerth, Eric Müller-Budack: Multimodal Geolocation Estimation of News Photos. ECIR 2023.
[2] Golsa Tahmasebzadeh, Endri Kacupaj, Eric Müller-Budack, Sherzod Hakimov, Jens Lehmann, Ralph Ewerth. GeoWINE: Geolocation based Wiki, Image, News and Event Retrieval. SIGIR 2021.
[3] Golsa Tahmasebzadeh, Eric Müller-Budack, Sherzod Hakimov, Ralph Ewerth. MM-Locate-News: Multimodal Focus Location Estimation in News. MMM 2023.
[4] Eric Müller-Budack, Kader Pustu-Iren, Ralph Ewerth: Geolocation estimation of photos using a hierarchical model and scene classification. ECCV 2018.
[5] Giorgos Kordopatis-Zilos, Adrian Popescu, Symeon Papadopoulos, Yiannis Kompatsiaris: Placing images with refined language models and similarity search with pca-reduced VGG features. MediaEval Workshop 2016.
[6] Tobias Weyand, Andre Araujo, Bingyi Cao, Jack Sim: Google landmarks dataset v2 - A largescale benchmark for instance-level recognition and retrieval. CVPR 2020.
[7] Arnau Ramisa, Fei Yan, Francesc Moreno-Noguer, Krystian Mikolajczyk: Breakingnews: Article annotation by image and text processing. IEEE Trans. Pattern Anal. Mach. Intell. 2018.
[8] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever: Learning Transferable Visual Models From Natural Language Supervision. ICML 2021.
[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: Bert: Pre-training of deep bidirectional transformers for language understanding. NAACL-HLT 2019.
[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun: Identity Mappings in Deep Residual Networks. ECCV 2016.

Cleopatra ITN has received funding from the European Union's Horizon 2020 research and innovation programme (Marie Skłodowska-Curie grant agreement no. 812997)