2021 Masters Projects*

Combining visual and textual information for enhancing radiological practices

Guillaume SERIEYS[1], Camille KURTZ[1], Laure FOURNIER[2], Florence CLOPPET[1]

[1]LIPADE, Université de Paris [2]Hôpital Européen Georges Pompidou, AP-HP

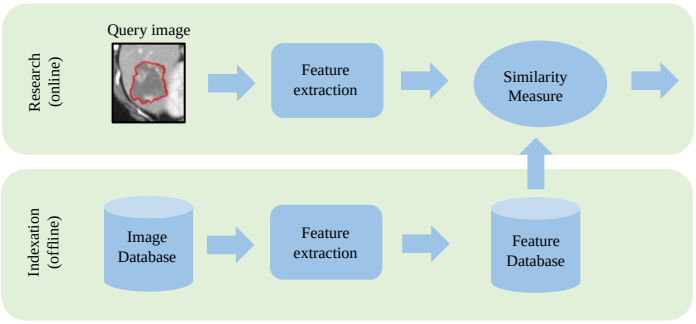
*Work supported by diiP, IdEx Université de Paris, ANR-18-IDEX-0001

Contact: guillaume.serieys@espci.fr

Background and Motivation

When facing complex cases, radiologists tend to look at known cases to establish a diagnosis. To do so, they use a PACS (Picture and Archiving Communication System) which stores all clinical and imaging data produced by the hospital. However, such systems are built for archiving purposes only. To search for a specific case, radiologists can only search by keywords which is suboptimal.

General aim: Integration of CBIR (Content-Based Image Retrieval) in PACS so that radiologists can search known cases using images.



Content-Based Image Retrieval

- **Medical images**: Challenging in CBIR due to the **scarcity of annotated data** and the complex nature of medical images which use **fine-grained visual features** compared to natural images.
- **Specific aim**: Build a better visual representation for medical images for enhancing CBMIR (Content-Based Medical Image Retrieval).

Materials and Methods

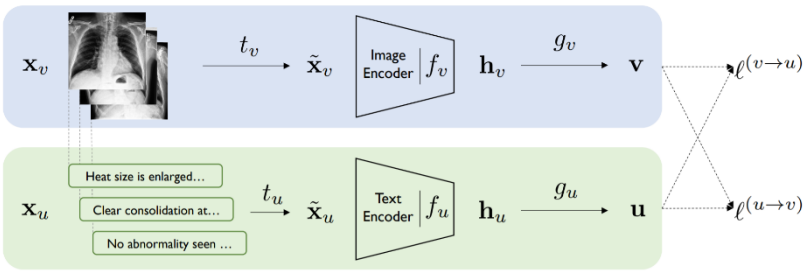


Illustration from [2]

Contrastive Learning

- **SimCLR [1]**: Unsupervised learning of a visual representation by maximizing the agreement between positive pairs of images.
- **Used method**: Unsupervised learning of a visual representation by maximizing the agreement between positive image-text pairs (c.f. left illustration).

ROCO [3]

- **81,825 radiology images** with corresponding captions.
- **Multimodal** image dataset (CT, X-Ray, etc…).

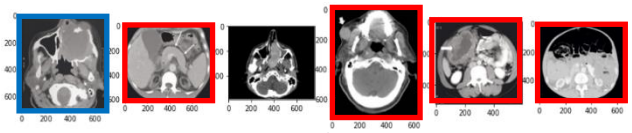
MedICaT [4]

- **217,060 figures** from 131,410 open access papers.
- Inline references for ~25k figures in the ROCO dataset.

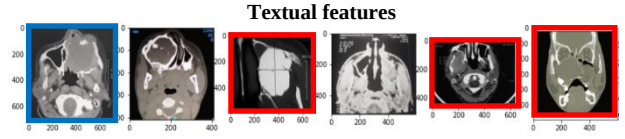
Preliminary Qualitative Results

« Computed tomography scan in axial view showing obliteration of the left maxillary sinus »

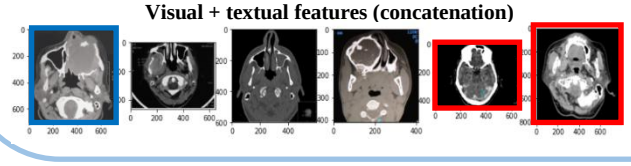
Visual features



Textual features



Visual + textual features (concatenation)



- **Visual features**: We retrieve abdominal CT scans instead of CT scans of sinus.
- **Textual features**: We retrieve CT scans of sinus but with dissimilar views.
- **Visual + textual features**: We retrieve CT scans of sinus with similar views.

Perspectives

- **Pathologic case retrieval**: This would imply to integrate multimodal and multiparametric CBIR by specialising neural networks for each sequence or modality, by using variational encoders to project the different modalities in a same latent space, etc…
- **Interpretability**: It is crucial for practicians. We could explore approaches such as attention mechanisms to identify ROIs, etc…
- **Evaluation**: It is a major issue when it comes to CBIR due to the scarcity of labelled data. We are currently working on creating reference queries with radiologists to have an unbiased evaluation.

References

[1] T. Chen *et al.*, 'A Simple Framework for Contrastive Learning of Visual Representations', in *International Conference on Machine Learning*, Nov. 2020, pp. 1597–1607.
[2] Y. Zhang *et al.*, 'Contrastive Learning of Medical Visual Representations from Paired Images and Text', *arXiv:2010.00747 [cs]*, Oct. 2020.
[3] O. Pelka *et al.*, 'Radiology Objects in COntext (ROCO): A Multimodal Image Dataset', in *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, vol. 11043, D. Stoyanov *et al.* Cham: Springer International Publishing, 2018, pp. 180–189.
[4] S. Subramanian *et al.*, 'MedICaT: A Dataset of Medical Images, Captions, and Textual References', *arXiv:2010.06000 [cs]*, Oct. 2020.