Radiology Objects in COntext (ROCO): A Multimodal Image Dataset

Obioma Pelka[1,3], Sven Koitka[1,2,4], Johannes Rückert[1],
Felix Nensa[4], and Christoph M. Friedrich[1,5]

[1] University of Applied Sciences and Arts Dortmund (FHDO)
Department of Computer Science
{obioma.pelka,johannes.rueckert,christoph.friedrich}@fh-dortmund.de
[2] TU Dortmund University, Department of Computer Science
[3] University of Duisburg-Essen, University Hospital Essen
[4] Department of Diagnostic and Interventional Radiology and Neuroradiology
{sven.koitka,felix.nensa}@uk-essen.de
[5] Institute for Medical Informatics, Biometry and Epidemiology (IMIBE)

Abstract. This work introduces a new multimodal image dataset, with the aim of detecting the interplay between visual elements and semantic relations present in radiology images. The objective is accomplished by retrieving all image-caption pairs from the open-access biomedical literature database PubMedCentral, as these captions describe the visual content in their semantic context. All compound, multi-pane, and non-radiology images were eliminated using an automatic binary classifier fine-tuned with a deep convolutional neural network system. Radiology Objects in COntext (ROCO) dataset contains over 81k radiology images with several medical imaging modalities including Computer Tomography, Ultrasound, X-Ray, Fluoroscopy, Positron Emission Tomography, Mammography, Magnetic Resonance Imaging, Angiography. All images in ROCO have corresponding caption, keywords, Unified Medical Language Systems Concept Unique Identifiers and Semantic Type. An out-of-class set with 6k images ranging from synthetic radiology figures to digital arts is provided, to improve prediction and classification performance. Adopting ROCO, systems for caption and keywords generation can be modeled, which allows multimodal representation for datasets lacking text representation. Systems with the goal of image structuring and semantic information tagging can be created using ROCO, which is beneficial and of assistance for image and information retrieval purposes.

Keywords: Deep Learning · Image Retrieval · Image Captioning · Multimodal Representation · Natural Language Processing · Radiology

1 Introduction

Given the growing complexity of the information radiologists are faced with interpreting, automatic image interpretation is becoming inevitable. However, the more knowledge of the image characteristics, the more structured is the radiology report and hence, the more efficient are the radiologists regarding interpretation. We present in this work a new, free and accessible dataset that concentrates on the detection of contextual object interplay solely in radiology images. Knowledge on semantic relations supplemented with visual elements improves the image understanding and interpretation.

The Radiology Objects in Context (ROCO) dataset has two classes: Radiology and Out-Of-Class. The first contains 81,825 radiology images with several medical imaging modalities including, Computer Tomography (CT), Ultrasound, X-Ray, Fluoroscopy, Positron Emission Tomography (PET), Mammography, Magnetic Resonance Imaging (MRI), Angiography and PET-CT. The latter contains 6,127 out-of-class samples, including synthetic radiology figures, digital art and portraits. The corresponding captions, keywords, UMLS (Unified Medical Language System) Semantic Types (SemTypes) [3], UMLS Concept Unique Identifiers (CUIs) [3] and download link is distributed for each image. Generative models trained on ROCO image - caption pairs can be used to automatically create natural sentences describing radiology images, as proposed in [14, 23]. The keywords distributed can be adopted for multi-class classification tasks, semantic tagging and multi-modal image representation, as this has proven to obtain higher prediction results [15]. Each ROCO image has a ftp-download link, containing the figure name and PMC identifier, which can be used to extract the image and corresponding article. Information on how to download the ROCO dataset as well as details on baseline creation are provided on the project website[1].

ROCO was created using datasets listed in Section 3 with the dataset gathering approach explained in Section 4 and displayed in Fig. 1. Example of images and corresponding text information are displayed in Section 5.

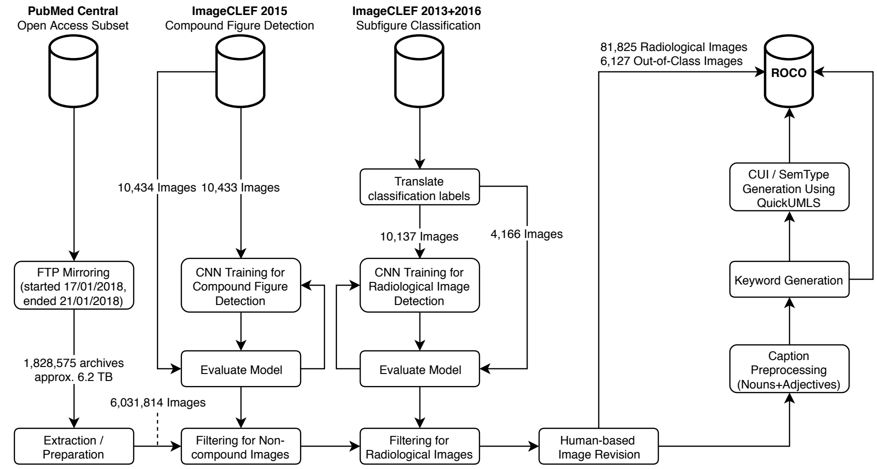

PubMed Central
Open Access Subset
ImageCLEF 2015
Compound Figure Detection
ImageCLEF 2013+2016
Subfigure Classification
81,825 Radiological Images
6,127 Out-of-Class Images
ROCO
10,434 Images
10,433 Images
Translate classification labels
CUI / SemType Generation Using QuickUMLS
10,137 Images
4,166 Images
FTP Mirroring (started 17/01/2018, ended 21/01/2018)
CNN Training for Compound Figure Detection
CNN Training for Radiological Image Detection
Keyword Generation
1,828,575 archives approx. 6.2 TB
Evaluate Model
Evaluate Model
Caption Preprocessing (Nouns+Adjectives)
6,031,814 Images
Extraction / Preparation
Filtering for Non-compound Images
Filtering for Radiological Images
Human-based Image Revision

Fig. 1. Overview of the workflow used for the ROCO dataset development.

[1] https://github.com/razorx89/roco-dataset

2 Related Work

Research datasets aid the evaluation of model algorithms as well as create new research focus topics. Hence, there is need for extensive, large-scale and easily accessible datasets. ImageNet [19] is a popular and often applied dataset for image classification tasks. It contains 500-1,000 images for over 22 thousand categories each. The Microsoft Common Objects in Context (MSCOCO), a dataset for detailed object model learning of 2D localization, was presented in [10]. It contains 91 common object categories with each category having over 5,000 labeled instances [10].

In the medical domain, several datasets have been introduced. Since 2003, ImageCLEF organizes image retrieval evaluation campaigns and with each year several computer vision task with accompanying datasets. In ImageCLEF 2009 [22] and ImageCLEF 2010 [12], a dataset with 77,506 images was made accessible by the Radiological Society of North America (RSNA). The recently presented ChestX-ray14 dataset released by [17] contains 112,120 frontal-view X-ray images of 30,805 unique patients. Each image is annotated with approximately 14 different thoracic pathology labels and is labeled positive when pneumonia is present and negative when not [17].

The proposed ROCO dataset provides medical knowledge, originated from peer-reviewed scientific biomedical literature with different textual annotations and a broad scope of medical imaging techniques. The dataset is easily accessible, as the originating source is open access. Image and information retrieval tasks with textual and content-/context-based searching can be initiated as many health informatics systems struggle with unstructured medical entities.

3 Datasets

3.1 PubMedCentral Database

The PMC archive contains of over 4.8 million articles which are provided by 2,110 full participants, 329 NIH portfolio and 4,597 selective deposit journals, and was initiated in 1999 [18]. This electronic archive offers free access to full-text journal articles with corresponding PubMed Central identifiers (PMCID). As database development method, the PubMedCentral Open Access Subset was chosen, due to the free access and trustworthy source. This subset was extracted between 2018-01-17 - 2018-01-21 and contains 1,828,575 archives/documents made available under a Creative Commons or similar license. From these archives, a total number of 6,031,814 image - caption pairs were extracted, as some articles contain no or several images.

3.2 ImageCLEF 2015 Medical Classification

This dataset was distributed for the compound detection task at the ImageCLEF 2015 Medical Classification Tasks and contains 20,867 figures, split into 10,433 and 10,434 images for training and test sets, respectively. The training set has 6,144 compound and 4,289 non-compound images [5]. This dataset set was used for training and classification of the extracted PubMed figures to compound and non-compound.

3.3 ImageCLEF 2013/2016 Medical Task

The images distributed at the subfigure classification task of ImageCLEF 2013 [4] and ImageCLEF 2016 [6] are annotated using the classification scheme shown in Fig. 2. The images from both datasets are used for training the image classifier used to detect radiology / non-radiology figures.

Following [7], the ImageCLEF 2016 training set of 6,775 images was extended with images from the ImageCLEF 2015 dataset, which resulted in 10,137 training images in total.

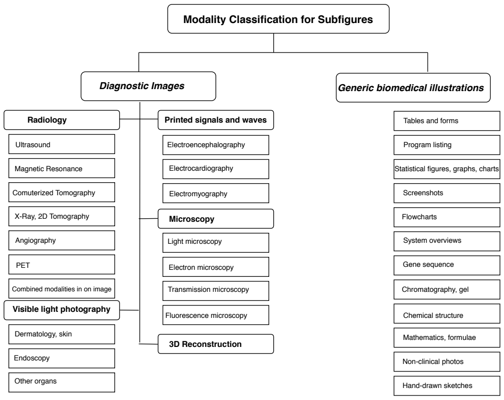


Fig. 2. Classification scheme adopted for the subfigure classification task at ImageCLEF 2013 and ImageCLEF 2016. Illustration adopted from [11].

4 Methodology

As deep learning techniques [9] have improved prediction accuracies in image classification [8] and in applications such as medical imaging [24], a deep learning architecture is used to filter the extracted PMC subset and is described in Subsection 4.1. The procedure of extracting keywords, UMLS CUIs and Semantic Types is explained in Subsection 4.2.

4.1 Filtering PubMedCentral Database

For filtering the 6,031,814 image - caption pairs to just radiology and non-compound figures, annotated images from the medical tasks of ImageCLEF 2013, 2015 and 2016 were utilized, which is explained below.

Removing Compound Figures First, compound and non-compound images need to be separated in order to isolate images of interest. The authors in [5, 6] also published a dataset with annotations to slice compound figures into corresponding non-compound image panes. However, this process including the human examination is time consuming. In order to detect compound or non-compound images automatically, neural networks were trained on images from the compound figure detection task of ImageCLEF 2015 [5]. The performance of the trained model was evaluated on the official test set including 10,434 images, resulting in approximately 90% accuracy.

For all neural network models, the popular Tensorflow [1] framework was utilized in conjunction with the slim training framework[2]. As underlying feature extractor network, the Inception-ResNet-V2 [21] was adopted. Since the datasets are too small to train a generalizable model, the pre-trained ImageNet model weights [19] were used for initialization. The last layer was replaced with two random initialized units outputs. The training process was split into two steps. First, only the last layer was trained using the RMSProp optimizer with *epoch*=1, *learning rate*=1e−3, *weight decay*=1e−4 and *mini batch*=32. Afterwards, the whole network was trained with *epoch*=30, *learning rate*=[1e−2,1e−3,1e−4], *weight decay*=4e−5 and *mini batch*=32.

All images were preprocessed following the proposals of [7]. Additionally, all images were resized to 360 pixel square images, preserving the original aspect ratio and filling areas with white as background color. The network was trained on 299 pixel square images, which were randomly cropped from the preprocessed images at run-time. For data augmentation, the standard Inception preprocessing implemented in Slim was adopted.

Filtering for Radiological Images All non-compound images were further processed by a second neural network to predict if it is a radiological image or not. The model performance was evaluated using the ImageCLEF 2016 test set, which includes 4,166 images labeled with the classification scheme shown in Fig. 2. For this task, all seven categories of ”DRxx Radiology” are fused into a radiological label, the remaining 23 labels are fused into a non-radiological label.

For training the same schedule as for the compound figure detection dataset was used. The final model achieved an accuracy of 98.6% for the binary classification of radiological vs non-radiological image. Please note, the ImageCLEF subfigure classification datasets are highly unbalanced and only 1,500 images of the extended training set of 10,137 images were labeled as radiological.

Revision of Final Images After filtering for non-compound and radiology images, the extracted PMC subset was reduced to 87,952 figures. A final revision of these images was done and false positives were manually detected. These include synthetic radiology images, illustrations, portraits, digital artwork, compound radiology images, and make up the out-of-class set, as shown in Fig. 4.

[2] https://github.com/tensorflow/models/research/slim [last access: 09.04.2018]

4.2 Caption Preprocessing

The captions of all 87,952 images from the filtered final subset were preprocessed in two steps and are described as follows:

Caption to Keywords Focusing on image and information retrieval purposes, certain contents in biomedical figure captions are undesirable and should be omitted. These are the performed preprocessing steps: ***C*ompound Figure Delimiter:** 67.26% of biomedical figures in PubMed Central are compound figures. These captions most likely address the subfigures using delimiters. Such delimiters were detected and removed. An excerpt of delimiters removed is listed in [13]. ***E*nglish Stopwords:** Using the NLTK Stopword corpus, present stopwords in the captions were omitted. This corpus contains 2,400 stopwords for 11 languages [2]. ***S*pecial Characters and Single Digits:** Special characters such as symbols, punctuations, metrics, etc. and words which consist of just numbers were removed. ***W*ord Stemming:** To reduce complexity, the captions are stemmed using Snowball Stemming [16]. ***N*oun and Adjectives:** The remaining words from the processed captions were trimmed down to nouns and adjectives, as these content the important contextual information. The trimmed captions are the proposed keywords.

Keywords to UMLS CUIs and SemTypes With the derived keywords, UMLS Concept Unique Identifiers and Semantic types are extracted. The transformation from keywords to CUIs and SemTypes was achieved using QuickUMLS, which is a fast, unsupervised, and approximate dictionary matching algorithm [20]. The parameters used are: *overlapping criteria=score*, *similarity name=jaccard*, *threshold*=0.7, *window*=5, *ngram length*=3.

5 Results

An excerpt of images included in the radiology subset is displayed in Fig. 3, showing the various medical imaging modalities. Some images of the additional out-of-class, such as portraits, digital arts, are displayed in Fig. 4. Figure 5 shows a ROCO image with all textual information extracted from the caption, with PMCID needed for downloading the articles.

ROCO consists of the subsets 'radiology' and 'out-of-class', representing the 81,825 true positives and 6,127 false positives, respectively. The figures in the 'out-of-class' set includes synthetic radiology images, clinical photos, portraits, compound radiology images as well as digital art.

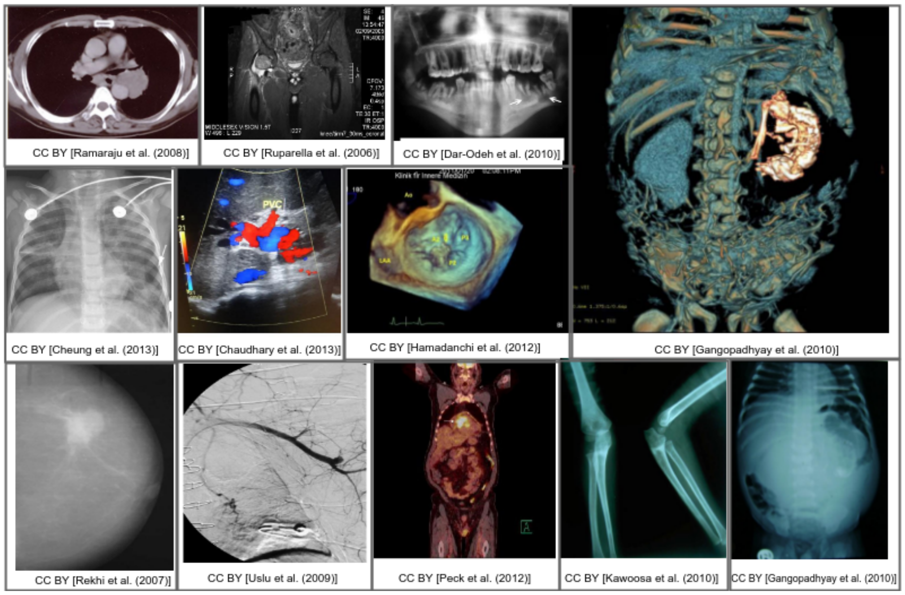

CC BY [Ramaraju et al. (2008)]
CC BY [Ruparella et al. (2006)]
CC BY [Dar-Odeh et al. (2010)]
CC BY [Cheung et al. (2013)]
PVC
CC BY [Chaudhary et al. (2013)]
CC BY [Hamadanchi et al. (2012)]
CC BY [Gangopadhyay et al. (2010)]
CC BY [Rekhi et al. (2007)]
CC BY [Uslu et al. (2009)]
CC BY [Peck et al. (2012)]
CC BY [Kawoosa et al. (2010)]
CC BY [Gangopadhyay et al. (2010)]

Fig. 3. Examples of images contained in the ROCO dataset, illustrating the variety of medical imaging modalities. All images belong to the 'Radiology' subset.

For reproducible and evaluation purposes, a random 80/10/10 split was applied on the ROCO dataset. Following this split, a training set with (65,460 'radiology' and 4,902 'out-of-class') images, a validation set with (8,183 'radiology' and 612 'out-of-class') images and a test set with (8,182 'radiology' and 613 'out-of-class') images, was created.

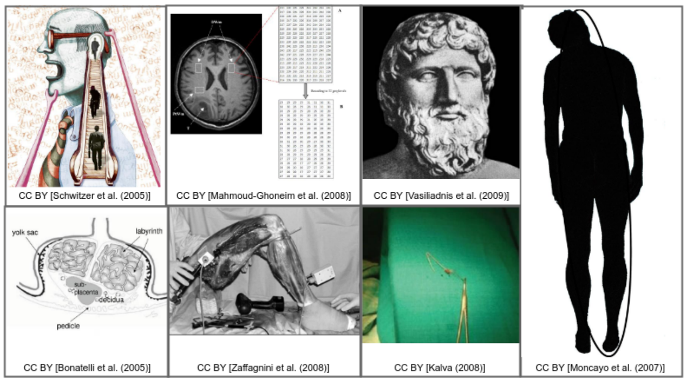
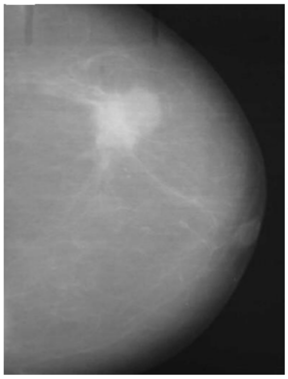

CC BY [Schwitzer et al. (2005)]
CC BY [Mahmoud-Ghoneim et al. (2008)]
CC BY [Vasiliadnis et al. (2009)]
yolk sac
labyrinth
sub-placenta
decidua
pedicle
CC BY [Bonatelli et al. (2005)]
CC BY [Zaffagnini et al. (2008)]
CC BY [Kalva (2008)]
CC BY [Moncayo et al. (2007)]

Fig. 4. Examples of images contained in the ROCO dataset, illustrating contents of the 'Out-Of-Class' subset. All figures were randomly chosen.



Caption: Mammographic findings. A bilobed soft tissue lesion measuring about 3.2 × 2.5 cms, suspicious for malignancy, seen in the upper and outer quadrant of the left breast.

CUI:

C0006141	Breast
C0024671	Mammography
C0006826	Malignant Neoplasms
C0772294	Epinastine Hydrochloride
C0242114	Suspicion
C4282132	Malignancy
C0221198	Lesion

Keywords: upper, lesion, cm, malignancy, suspicious, quadrant, tissue, breast, finding, outer, soft, mammographic, left

SemTypes:
T060 Diagnostic Procedure
T041 Mental Process
T121 Pharmacologic Substance
T191 Neoplastic Process
T033 Finding
T023 Body Part, Organ, or Organ Component
T109 Organic Chemical

Download:

```
wget -r ftp://ftp.ncbi.nlm.nih.gov/pub/pmc/oa_package/38/bc/PMC1808459.tar.gz
 -P /path/to/Dir    Figure_name: 1477-7819-5-24-1.jpg
```

CC BY [Rekhi et al. (2007)]

Fig. 5. An example of a ROCO image, showing corresponding caption, keywords, UMLS CUIs, UMLS Semantic Types and PMCID download link.

6 Conclusion

To create modeling approaches regarding the detection of contextual object interplay in radiology images, the Radiology Objects in COntext (ROCO) dataset is introduced in this paper. ROCO does not focus on a single specific disease or anatomical structure but addresses several medical imaging modalities. The database development method was to extract all articles available in the PubMed Central Open Access Subset.

To filter the 6,031,814 images to radiology and non-compound figures, two automatic binary classifiers fine-tuned with a deep convolutional neural network system were trained. For data standardization and additional image interrelations, the textual annotation per image is extended with the Unified Medical Language System (UMLS) Concept Unique Identifiers (CUIs) and Semantic Types (SemType). These were achieved by processing the image captions to keywords and using QuickUMLS to transform these keywords to CUIs and SemTypes.

The keywords can be adopted as textual features for multi-modal image representations in classification tasks, as well as for multi-class image classification and labeling. Automatic keyword generation models can be designed using ROCO image - keyword pairs, enabling perceivable order for unstructured and unlabeled radiology images and for datasets lacking textual representations. Natural sentences describing radiology images can be created using generative models trained with ROCO image - caption pairs. This will offer additional knowledge of the images and not be limited to solely visual representations.

In future work, an extensive evaluation on ROCO will be performed. This will include baselines for specific applications such as, generative models for image captioning and keywords, image classification using multi-modal image representation and information and image retrieval using semantic labeling.

References

1. Abadi, Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: Tensorflow: A system for large-scale machine learning. In: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. USENIX Association, Berkeley, CA, USA (2016)
2. Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python. O'Reilly (2009), http://www.oreilly.de/catalog/9780596516499/index.html
3. Bodenreider, O.: The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Research **32**(Database-Issue), 267–270 (2004). https://doi.org/10.1093/nar/gkh061
4. García Seco de Herrera, A., Kalpathy-Cramer, J., Demner Fushman, D., Antani, S., Müller, H.: Overview of the ImageCLEF 2013 medical tasks. In: Working Notes of CLEF 2013 - Conference and Labs of the Evaluation forum, Valencia, Spain, 23-26 September, 2013. CEUR-WS Proceedings Notes, Volume 1179
5. García Seco de Herrera, A., Müller, H., Bromuri, S.: Overview of the ImageCLEF 2015 medical classification task. In: Working Notes of CLEF 2015 - Conference and Labs of the Evaluation Forum, Toulouse, France, September 8-11, 2015. CEUR-WS Proceedings Notes, Volume 1391 (September 2015)
6. García Seco de Herrera, A., Schaer, R., Bromuri, S., Müller, H.: Overview of the ImageCLEF 2016 medical task. In: Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum, Évora, Portugal, 5-8 September, 2016. CEUR-WS Proceedings Notes, Volume 1609 (September 2016)
7. Koitka, S., Friedrich, C.M.: Optimized convolutional neural network ensembles for medical subfigure classification. In: Jones, G.J., Lawless, S., Gonzalo, J., Kelly, L., Goeuriot, L., Mandl, T., Cappellato, L., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction at the 8th International Conference of the CLEF Association, Dublin, Ireland, September 11-14, 2017, Lecture Notes in Computer Science (LNCS) 10456. pp. 57–68. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-65813-1_5
8. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. pp. 1097–1105. Curran Associates Inc., USA (2012)
9. LeCun, Y., Bengio, Y., Hinton, G.E.: Deep Learning. Nature **521**(7553), 436–444 (2015). https://doi.org/10.1038/nature14539
10. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common Objects in Context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Proceedings of Computer Vision – ECCV 2014, 13th European Conference, Zurich, Switzerland, September 6-12, 2014. pp. 740–755. Springer International Publishing, Cham (2014)
11. Müller, H., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S.: Creating a classification of image types in the medical literature for visual categorization. In: Proc. SPIE 8319, Medical Imaging 2012: Advanced PACS-based Imaging Informatics and Therapeutic Applications, 83190P (February 23, 2012). vol. 8425, pp. 194– (02 2012). https://doi.org/10.1117/12.911186

12. Müller, H., Kalpathy-Cramer, J., Eggel, I., Bedrick, S., Reisetter, J., Jr., C.E.K., Hersh, W.R.: Overview of the CLEF 2010 medical image retrieval track. In: CLEF 2010 LABs and Workshops, Notebook Papers, 22-23 September 2010, Padua, Italy. CEUR-WS Proceedings Notes, Volume 1176 (2010), http://ceur-ws.org/Vol-1176/CLEF2010wn-ImageCLEF-MullerEt2010.pdf
13. Pelka, O., Friedrich, C.M.: FHDO Biomedical Computer Science Group at Medical Classification Task of ImageCLEF 2015. In: Working Notes of CLEF 2015 - Conference and Labs of the Evaluation forum, Toulouse, France, September 8-11, 2015. (2015), http://ceur-ws.org/Vol-1391/14-CR.pdf
14. Pelka, O., Friedrich, C.M.: Keyword Generation for Biomedical Image Retrieval with Recurrent Neural Networks. In: Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017. CEUR-WS Proceedings Notes, Volume 1866 (2017)
15. Pelka, O., Nensa, F., Friedrich, C.M.: Adopting semantic information of grayscale radiographs for image classification and retrieval. In: Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018) - Volume 2: BIOIMAGING, Funchal, Madeira, Portugal, January 19-21, 2018. pp. 179–187 (2018). https://doi.org/10.5220/0006732301790187
16. Porter, M.F.: Snowball: A language for stemming algorithms (2001), http://www.webcitation.org/6yci04ExR
17. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M.P., Ng, A.Y.: Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. CoRR **abs/1711.05225** (2017), https://arxiv.org/abs/1711.05225
18. Roberts, R.J.: PubMed Central: The GenBank of the published literature. Proceedings of the National Academy of Sciences of the United States of America **98**(2), 381–382 (Jan 2001). https://doi.org/10.1073/pnas.98.2.381
19. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: Imagenet large scale visual recognition challenge. International Journal of Computer Vision **115**(3), 211–252 (Dec 2015). https://doi.org/10.1007/s11263-015-0816-y
20. Soldaini, L., Goharian, N.: QuickUMLS: a fast, unsupervised approach for medical concept extraction. In: MedIR Workshop, SIGIR (2016)
21. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-First AAAI Conference on Artificial Intelligence (2017), https://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14806
22. Tommasi, T., Caputo, B., Welter, P., Güld, M.O., Deserno, T.M.: Overview of the CLEF 2009 Medical Image Annotation Track. In: Multilingual Information Access Evaluation II. Multimedia Experiments - 10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009, Corfu, Greece, Sept 30 - Octr 2, 2009, pp. 85–93 (2009), https://doi.org/10.1007/978-3-642-15751-6_9
23. Vinyals, O., Toshev, A., Bengio, S., Erhan, D.: Show and Tell: Lessons Learned from the 2015 MSCOCO Image Captioning Challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(4), 652–663 (2017). https://doi.org/10.1109/TPAMI.2016.2587640
24. Xu, Y., Mo, T., Feng, Q., Zhong, P., Lai, M., Chang, E.I.: Deep learning of feature representation with multiple instance learning for medical image analysis. In: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2014, Florence, Italy, May 4-9, 2014. pp. 1626–1630 (2014). https://doi.org/10.1109/ICASSP.2014.6853873