Optimizing Body Region Classification With Deep Convolutional Activation Features

Obioma Pelka[1,2] [0000-0001-5156-4429], Felix Nensa[3] [0000-0002-5811-7100], and Christoph M. Friedrich[1,4] [0000-0001-7906-0038]

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

Abstract. The goal of this work is to automatically apply generated image keywords as text representations, to optimize medical image classification accuracies of body regions. To create a keyword generative model, a Long Short-Term Memory (LSTM) based Recurrent Neural Network (RNN) is adopted, which is trained with preprocessed biomedical image captions as text representation and visual features extracted using Convolutional Neural Networks (CNN). For image representation, deep convolutional activation features and Bag-of-Keypoints (BoK) features are extracted for each radiograph and combined with the automatically generated keywords. Random Forest models and Support Vector Machines are trained with these multimodal image representations, as well as just visual representation, to predict body regions. Adopting multimodal image features proves to be the better approach, as the prediction accuracy for body regions is increased.

Keywords: Bag-of-Keypoints · DeCaf · Deep Learning · Multimodal Representation · Natural Language Processing · Radiographs

1 Introduction

To build classification systems capable of reliable performance, adequate image representation is necessary. Adopting multimodal image features presented in [10, 12, 13], proves to achieve higher classification accuracies for biomedical images, as this contributes towards sufficient image representation. However, some classification tasks such as ImageCLEF 2015 Medical Clustering Task [8], as well as real clinical cases, lack corresponding text representations.

Hence, this paper utilizes automatic generated keywords proposed in [14] to substitute as text representation for the classification of radiographs into body regions, focusing on a different feature extraction method. The obtained keywords are combined with visual features for multi modal image representation.

The generated text information can also be further applied for semantic tagging and image retrieval purposes.

We show that by adopting a multi-modal image representation and classification method described in subsections 2.2 and 2.3, the overall prediction accuracy is increased as shown in section 3, by evaluating the model performance on a dataset presented in subsection 2.1.

2 Material and Methods

2.1 Dataset

The Medical Clustering Task was held at ImageCLEF 2015, an evaluation campaign organized by the CLEF Initiative [1]. For this task, 750 high resolution x-ray images collected from a hospital in Dhaka, Bangladesh [1] were distributed. The training set included 500 images and test set 250 images, with annotations of the following classes: 'Body', 'Head-Neck', 'Upper-Limb, 'Lower-Limb' and 'True-Negative'. An excerpt of the x-rays is displayed in figure 1.

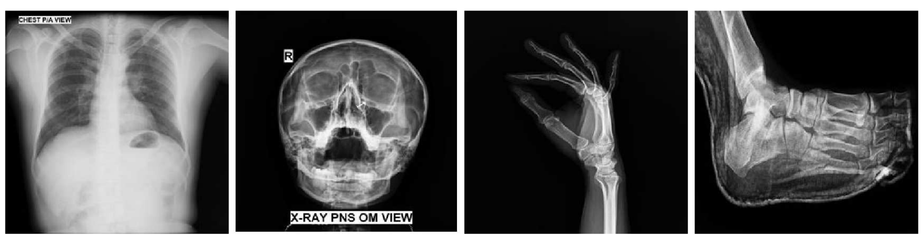

Fig. 1. An excerpt of images from the CVC digital x-ray dataset, Medical Clustering task, ImageCLEF 2015. Original data is available from www.cvcrbd.org.

For the creation of the keyword generative model, the dataset distributed for the ImageCLEF Caption Prediction Task [7] was applied and is presented in [14].

2.2 Image Representation

For visual representation, two methods are applied for comparison purposes: Deep convolutional activation features (DeCaf) [6] and Bag-of-Keypoints [5] computed with dense SIFT descriptors [11]. The deep visual features are the average pool layer of the deep learning system Inception_V3 [18], which is pre-trained on ImageNet [15]. The activation features were extracted using the neural

[1] http://www.clef-initiative.eu/

network API Keras 2.2.0 [4]. The Bag-of Keypoints visual features were created using the *VLFEAT* library [19].

To obtain multi-modal image representation, text information was created. The keyword generative model proposed in [14] was used to automatically create keywords for all 750 images, belonging to training and test sets. Furthermore, a compact text representation was achieved by applying vector quantization on a Bag-of-Words [17] codebook and Term Frequency-Inverse Document Frequency (Tf-IDF) [16].

2.3 Classification Models

Random forest (RF) [2] models with 1,000 trees were created as image classification models. These RF-models were trained using either visual or multi-modal image representations. Principal Component Analysis (PCA) [9] was applied to reduce computational time, feature dimension and noise. The vector size for visual features was reduced from 2,048 to 50, and from 150 to 50 for the text features. For comparison, multi-class Support Vector Machines (SVM) [3] using the same multi-modal image representations as the RF models, were modeled with the following parameters: kernel = radial basis function, cost parameter = 10 and gamma = 1/num_of_features.

3 Results

The achieved prediction accuracies using either visual or multi-modal image representation are listed in table 1. For comparison purposes, the different classifier setups used for training are shown in the first column.

Table 1. Prediction accuracies obtained using the different visual and text representations, as well as classifier setup. Evaluation was done on ImageCLEF Medical Clustering test set with 250 x-rays.

Classifier Setup	Accuracy	Image Representation
Random Forest + BoK	65.60 %	Visual
Support Vector Machines + BoK	66.40 %	Visual
Random Forest + DeCaf	74.00 %	Visual
Support Vector Machines + DeCaf	72.89 %	Visual
Random Forest + BoK + BoW (TF-IDF)	71.09 %	Visual + Text
Support Vector Machines + BoK + BoW (TF-IDF)	69.13 %	Visual + Text
Random Forest + DeCaf + BoW (TF-IDF)	**77.20 %**	Visual + Text
Support Vector Machines + DeCaf + BoW (TF-IDF)	76.35 %	Visual + Text
Best group ImageCLEF 2015 Med Clustering Task [1]	75.20 %	Visual

Figure 2 displays a word cloud created with the automatically generated keywords from the ImageCLEF Medical Clustering Training Set.

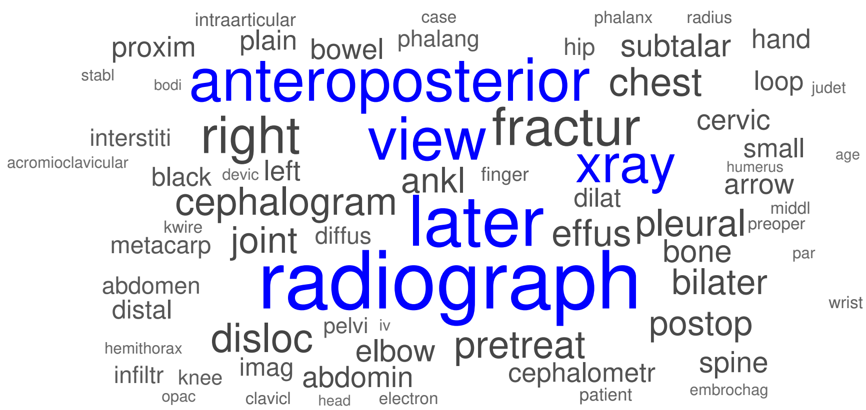

Fig. 2. Word cloud of automatically generated keywords for images in the ImageCLEF 2015 Medical Clustering Training Set.

4 Discussion

Adopting multi-modal representations for classification task proves to obtain higher prediction accuracies, as listed in Table 1. This is the case for both Random Forest and Support Vector Machines classification models. The prediction rate is optimized by applying DeCaf as visual representation, in comparison to the traditional Bag-of-Keypoints features. It can be seen from fig. 2, that the generated keywords contribute to a more adequate representation, as information on body regions achieved.

5 Conclusions

An approach for optimizing prediction accuracies using deep convolutional activation features combined with automatically generated keywords was presented. Following the results shown in Table 1, using multimodal image representations achieves higher classification accuracies than just visual features. This is observed for the different classification models and visual feature extraction method. As the prediction models trained with deep convolutional activation features outperform those trained with traditional Bag-of-Keypoints, continuous work can be based on evaluating several image enhancement techniques.

Acknowledgements

The work of Obioma Pelka was partially funded by a PhD grant from the University of Applied Sciences and Arts Dortmund (FHDO), Germany.

References

1. Amin, M.A., Mohammed, M.K.: Overview of the imageclef 2015 medical clustering task. 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/inv-pap1-CR.pdf
2. Breiman, L.: Random Forests. Machine Learning **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324
3. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Mining Knowledge Discovery **2**(2), 121–167 (Jun 1998). https://doi.org/10.1023/A:1009715923555
4. Chollet, F., et al.: Keras. https://keras.io (2015)
5. Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, European Conference on Computer Vision ECCV, Prague, Czech Republic, May 11-14, 2004. pp. 1–22 (2004)
6. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. In: Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014. pp. 647–655 (2014), http://jmlr.org/proceedings/papers/v32/donahue14.html
7. Eickhoff, C., Schwall, I., de Herrera, A.G.S., Müller, H.: Overview of ImageCLEFcaption 2017 - Image Caption Prediction and Concept Detection for Biomedical Images. In: Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017. (2017)
8. de Herrera, A.G.S., 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. pp. 219–232 (2016), http://ceur-ws.org/Vol-1609/16090219.pdf
9. Jolliffe, I.T.: Principal Component Analysis. In: International Encyclopedia of Statistical Science, pp. 1094–1096 (2011)
10. Kalpathy-Cramer, J., de Herrera, A.G.S., Demner-Fushman, D., Antani, S.K., Bedrick, S., Müller, H.: Evaluating performance of biomedical image retrieval systems - An overview of the medical image retrieval task at ImageCLEF 2004-2013. Computerized Medical Imaging and Graphics **39**, 55–61 (2015). https://doi.org/10.1016/j.compmedimag.2014.03.004
11. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision **60**, 91–110 (2004)
12. 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

13. Pelka, O., Friedrich, C.M.: Modality prediction of biomedical literature images using multimodal feature representation . GMS Medizinische Informatik, Biometrie und Epidemiologie **12**(2), 1345–1359 (2016). https://doi.org/10.3205/mibe000166
14. 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
15. 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 (IJCV) **115**(3), 211–252 (2015). https://doi.org/10.1007/s11263-015-0816-y
16. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Information Processing & Management **24**(5), 513 – 523 (1988). https://doi.org/10.1016/0306-4573(88)90021-0
17. Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. McGraw-Hill computer science series, McGraw-Hill, New York (1983)
18. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016. pp. 2818–2826 (2016). https://doi.org/10.1109/CVPR.2016.308
19. Vedaldi, A., Fulkerson, B.: VLFEAT: an open and portable library of computer vision algorithms. In: Proceedings of the 18th International Conference on Multimedia 2010, Firenze, Italy, October 25-29, 2010. pp. 1469–1472 (2010). https://doi.org/10.1145/1873951.1874249