Essex at ImageCLEFcaption 2020 task

Alba G. Seco de Herrera, Francisco Parrilla Andrade,
Luke Bentley and Arely Aceves Compean

School of Computer Science and Electronic Engineering,
University of Essex, Colchester, UK
Corresponding author: alba.garcia@essex.ac.uk

Abstract. The University of Essex participated in the fourth edition of the ImageCLEFcaption task which aims to detect concepts on radiology images as an approach to medical image understanding. In this paper, the University of Essex team presents its participation in the ImageCLEF 2020 caption task based on a retrieval based approach for concept detection. A Densely Connected Convolutional Network is used to encode the images. This paper explores compares several modification of the baseline considering several aspects such as the image modality or the selection of concepts among the top retrieved images. The University of Essex was third best team participating in the task achieving a 0.381 mean F1 score, very close to the results obtained by the top two teams. Code and pre-trained models are available at https://github.com/fjpa121197/ImageCLEFmedEssex2020.

Keywords: ImageCLEF, image understanding, concept detection, medical image retrieval, Densely Connected Convolutional Network

1 Introduction

This paper describes the participation of the School of Computer Science and Electronic Engineering (CSEE) at the University of Essex at ImageCLEFcaption 2020 task [9]. ImageCLEF [6] is an evaluation campaign organised as part of the CLEF[1] initiative labs. The ImageCLEFcaption task aims to interpret and summarise the insights gained from medical images. The 2020 edition, similar to 2019, focused on concept detection in a large corpus of radiology images. This task provides tools for radiology image understanding. A detailed description of the data and the task is presented in Pelka et al. [9].

ImageCLEFcaption 2020 task is the forth edition of this successful task. In previous editions [8, 4, 3] multiple approaches have been explored by the participants and retrieval approaches achieved best results [7, 13, 1]. Following past year experience, in this paper we proposed a retrieval-based approach where the images are encoded by a Densely Connected Convolutional Network, DenseNets [5].

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

Several experiments are presented to select the most relevant concepts based on the concepts associated to the top ranked images retrieved. Code and pre-trained models are publicly available[2].

The rest of the paper is organised as follows. Section 2 presents collection and the evaluation methodology used in this work. Section 3 explains the techniques proposed in this paper including a detail description of the runs submitted to the ImageCLEFcaption task. The results are presented in Section 4. Finally, the conclusions are given in Section 5.

2 Collection & evaluation

In this work we used the ImageCLEFmed caption 2020 collection [9]. It consists on three subsets:

- training set including 64,753 images;
- validation set including 15,970 images;
- test set including 3,534 images.

The images originate from biomedical journal articles extracted from the PubMed Central® (PMC)[3] repository [10]. Each image is associated to multiple Unified Medical Language System® (UMLS) Concept Unique Identifiers (CUIs) [2]. The UMLS CUIs associated to the images in the training and validation sets were distributed and include 3,047.

The UMLS CUIs from the test set were not distributed and, therefore, not used to build the model. The ImageCLEFcaption task [9] organisers evaluated the submitted runs computing the F1-scores (see Section 4).

In 2020, the ImageCLEFmed caption collection is classified in seven medical image modalities (Angiography, Computer Tomography, Magnetic Resonance, Positron Emissions Tomography, Ultrasound, X-ray and combined modalities in one image).

3 Methodology

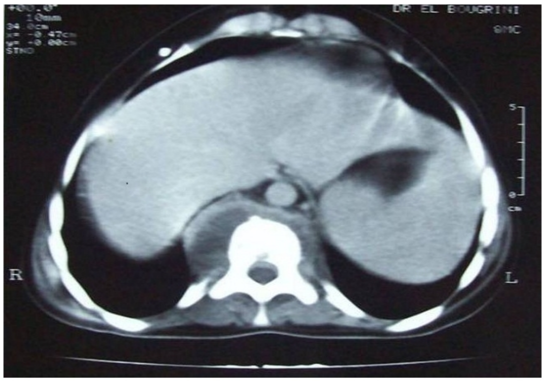
The proposed approach is based in a content-based image retrieval model, where DenseNets are used for feature extraction (see Section 3.1). A similarity comparison is done between the query image and the images in the training and validation test sets (see Section 3.2). Finally, concept selection is performed to predict the medical concepts for the query image (see Section 3.3).

Figure 2 shows an overview of the approach.

[2] https://github.com/fjpa121197/ImageCLEFmedEssex2020

[3] https://www.ncbi.nlm.nih.gov/pmc/

Image: ROCO2_CLEF_26437



Concepts:

- C0040398: Tomography, Emission-Computed
- C0224338: Structure of sternal muscle
- C0040405: X-Ray Computed Tomography

Fig. 1: Example of an image and the associated UMLS CUIs the validation set of the ImageCLEFcaption 2020 task.

3.1 Feature extraction

Following the success of the AUEB NLP Group at ImageCLEFmed Caption 2019 [7], this approach also uses a pre-trained DenseNet model (DenseNet-121) to encode the images, i.e, to extract their relevant features bases on this model. The existing DenseNet-121 has many parameters which require immense computing power and very large scale datasets to be trained from scratch. Hence, transfer learning is used in this work to mitigate this problem as its power in computer vision has been extensively study in the literature [12].

DenseNet models are Convolutional Neural Networks (CNN) models where each layer is connected directly to other layers [5]. DenseNet models have been recognised for their ability to reach similar performance to ResNet models, which use double the amount of layers [11]. DenseNet-121 has 121 layers with trainable weights. The model uses the weights from the ImageNet dataset, which consists of 1.2 million images, and it has 1,000 classes.

The input image is resized to 64×64 and transformed to an array, then a preprocessing module from DenseNet Keras is used. This module is in charge of transforming the pixel values into a 0-1 range, and also to normalise the values based on the ImageNet dataset. The DenseNet-121 model is then used to encode each image representing it as a vector of 4,096 dimensions excluding the classification layer.

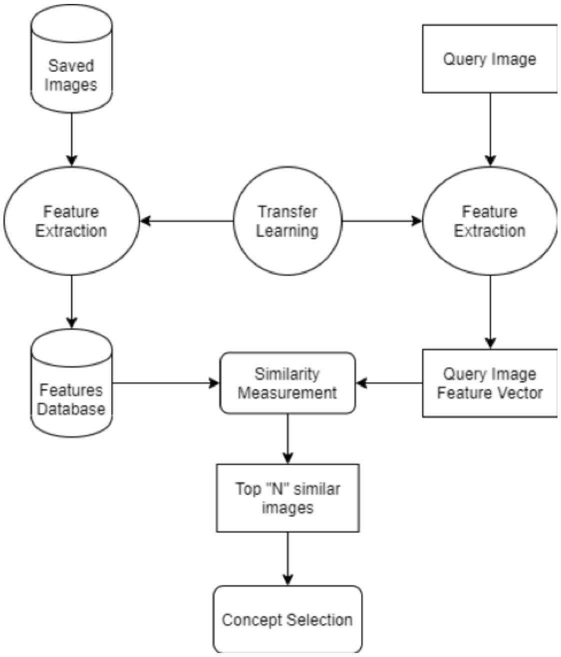


Fig. 2: Overview of the concept prediction approach.

Fine-tuning. In this work, a fine-tuning strategy is also explored to transfer learned recognition capabilities to the specific challenge of concept detection. The fine-tuning has been done specifically for each image modality, where a fully connected layer has been added to the DenseNet-121 model transforming it into a multi-label classification model. The last fully connected layer was trained for 10 rounds.

In particular the following parameters were used:

- *Optimizer*: RMSProp
- *Learning rate*: 0.0001
- *Batch size*: 32
- *Momemtum*: 0.0

The model was trained in two phases:

- *1st phase*: Only training the classification layers.
- *2st phase*: Training a portion of the feature learning layers and the classification layer.

Each phase consisted on 10 epochs (each epoch consisted of 100 steps, of which 10 steps were for validation).

3.2 Image retrieval

In this work, the image modality is used to improve the system performance. Each image in the test set is compared to all the images in the training or validation sets belonging to the same image modality as the query image. In the case of *run 64104* the images were retrieved from all the training set without considering the modality.

In order to perform the comparison, Canberra and Manhattan distances are computed given the encoded features (see Section 3.1). This metrics were chosen based on their accuracy and speed of their computational performance. The 10 most similar images to the given query were selected and their associated concepts extracted. Each of the extracted concept is tagged with a score based on its ranked position or the computed distance value (see next Section 3.3 for more details).

3.3 Concepts selection

In order to assign the concepts to the query images in the test set two methodologies were tested:

Ranking based selection. Each concept is assigned with a score based on the ranking of the 10 retrieved image which they were associated to. If the concept is associated to more than one image, then the value is added to it. For example, the highest ranked image has all its concepts given a value of 10 and the second highest has all its concepts given a value of 9. If the final score (after the addition) given to a concept is equal or over the threshold 20, then the concept is considered relevant to the query image and assigned to it.

Distance based selection. Each concept is assigned a scored based on the distance value computed of the 10 retrieved image which they were associated to. Similar to the ranked based selection, if the concept is associated to more than one image, then the value is added to it. For each concept final score (after the addition), the mean or percentile (99 or 95) is set as a threshold to select the concept. If the score was equal or over the threshold, then the concept is considered relevant to the query image and assigned to it. During the experimental set up other thresholds were tested such as percentiles 75 and 98 or a normalisation process, however there were no finally submitted to the challenge since mean and percentile 95 and 99 achieved better F1 score on the validation set.

3.4 Runs

This section provides a detailed description of the runs submitted to ImageCLEFcaption 2020 task. The methods used to implement these runs are described in Section 3. Table 1 summarises the techniques used by each run.

Table 1: Description and performance of the runs submitted to ImageCLEF 2020 Concept Detection Task and their ranks compared with all the 57 runs submitted by the 7 participating teams.

Run ID	Training	Per modality	Fine-tuning	Similarity measure	Threshold	F1 Score	Ranking
64104	T	No	No	Canberra	20	0.345	26
67416	T	Yes	No	Canberra	20	0.380	9
63804	T + V	Yes	No	Canberra	20	0.380	8
64394	T	Yes	Yes	Canberra	20	**0.381**	**7**
68019	T	Yes	Yes	Canberra	mean	0.280	34
68026	T	Yes	Yes	Canberra	95th perc.	0.246	36
68025	T	Yes	Yes	Canberra	98th perc.	0.337	31
68022	T	Yes	Yes	Canberra	99th perc.	0.379	10
68027	T	Yes	Yes	Manhattan	99th perc.	0.378	11
Best ImageCLEF2020	-	-	-	-	-	0.394	1

- *Run 64104 - baseline*: In this run, DenseNet-121 is used to encode the images. The top 10 images are retrieved from the training set using Canberra distance. Ranking based selection is used to select the relevant concepts from the retrieved images.
- *Run 67416*: This run is similar to the baseline. In this case the image modality information is used in the retrieval step. The top 10 images from the same modality as the query image are retrieved from the training set.
- *Run 63804*: This run is similar to the *Run 67416*. In this cases, the images are retrieved from both training and validation sets. The modality information is also considered.
- *Run 64394*: This run is similar to the *Run 67416*. In this run, fine-tuning is applied.
- *Run 68019*: This run is similar to the *Run 64394*. For this run, distance based selection is used to select the relevant concepts from the retrieved images using the mean of the scores as a threshold.
- *Run 68026*: This run is similar to the *Run 64394*. For this run, distance based selection is used to select the relevant concepts from the retrieved images setting 95th percentile as a threshold.
- *Run 68025*: This run is similar to the *Run 64394*. For this run, distance based selection is used to select the relevant concepts from the retrieved images setting 98th percentile as a threshold.
- *Run 68022*: This run is similar to the *Run 64394*. For this run, distance based selection is used to select the relevant concepts from the retrieved images setting 99th percentile as a threshold.
- *Run 68022*: This run is similar to the *Run 68027* but using the Manhattan distance in the retrieval step.

4 Results

Table 1 presents the official results achieved in the ImageCLEF 2020 Concept Detection Task and their ranks compared with all the 57 runs submitted by the 7 participating teams.

This year, our team was the third team with best results. Best results was achieved with *run 64394* with F1 score of 0.381, very close to the results of the second and first team which achieved a F1 score of 0.392 and 0.394, respectively. In particular, our best submitted used fine-tuning and Canberra distance to retrieved the top 10 images from the training set considering only the images from the same modality. Ranking based selection was also used to select the relevant concepts from the retrieved images.

Based on the the results achieved, it is clear that the used of the modality improve the results. Interestingly, we did not find difference when augmenting the set of images in the collection by including the validation set. It might be due of the nature of the images, since the retrieved images belonged to the same modality, including the validation set did not include many new concepts to retrieve.

Beside the possible advantages that fine-tuning can bring, in the official results, only a small improved is noticed when applying it. Similar when comparing Canberra and Manhattan distances, slightly better results were achieved when using Canberra distance.

Finally, the method used to select the concepts has a bigger impact on the overall results, achieving the best results when using the ranking based methodology.

5 Conclusions

This paper describes the participation of CSEE at the University of Essex at ImageCLEFcaption 2020 task. CSEE proposes a retrieval-based approach using a DenseNet-121 model to encode the images in the collection. CSEE compares different modifications in the baseline to study their effects on the final performance. Best submitted run used fine-tuning per image modality and Canberra distance in the retrieval step. Concepts were selected based on the top 10 ranked images. CSEE was the third best team at the benchmark achieving a F1 score of 0.381, very close to the results obtained by the top two teams. In 2020, the image modality was provided and future improvements can tackle an initial modality classification step as well as training the retrieval step per each modality. Further work is also needed to better understand the effects of the concept selection step.

References

1. Abacha, A.B., García Seco de Herrera, A., Gayen, S., Demner-Fushman, D., Antani, S.: Nlm at imageclef 2017 caption task. In: CLEF2017 Working Notes.

CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Dublin, Ireland (September 11-14 2017)

2. 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
3. Eickhoff, C., Schwall, I., García Seco de Herrera, A., Müller, H.: Overview of ImageCLEFcaption 2017 - the image caption prediction and concept extraction tasks to understand biomedical images. In: CLEF2017 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Dublin, Ireland (September 11-14 2017)
4. García Seco de Herrera, A., Eickhoff, C., Andrearczyk, V., , Müller, H.: Overview of the ImageCLEF 2018 caption prediction tasks. In: CLEF2018 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Avignon, France (September 10-14 2018)
5. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
6. Ionescu, B., Müller, H., Péteri, R., Abacha, A.B., Datla, V., Hasan, S.A., Demner-Fushman, D., Kozlovski, S., Liauchuk, V., Cid, Y.D., Kovalev, V., Pelka, O., Friedrich, C.M., de Herrera, A.G.S., Ninh, V.T., Le, T.K., Zhou, L., Piras, L., Riegler, M., l Halvorsen, P., Tran, M.T., Lux, M., Gurrin, C., Dang-Nguyen, D.T., Chamberlain, J., Clark, A., Campello, A., Fichou, D., Berari, R., Brie, P., Dogariu, M., Ştefan, L.D., Constantin, M.G.: Overview of the ImageCLEF 2020: Multimedia retrieval in lifelogging, medical, nature, and internet applications. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the 11th International Conference of the CLEF Association (CLEF 2020), vol. 12260. LNCS Lecture Notes in Computer Science, Springer, Thessaloniki, Greece (September 22-25 2020)
7. Kougia, V., Pavlopoulos, J., Androutsopoulos, I.: AUEB NLP group at ImageCLEFmed Caption 2019. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, vol. 2380. CEUR-WS.org, Lugano, Switzerland (September 09-12 2019)
8. Pelka, O., Friedrich, C.M., García Seco de Herrera, A., Müller, H.: Overview of the ImageCLEFmed 2019 concept prediction task. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, vol. 2380. CEUR-WS.org, Lugano, Switzerland (September 09-12 2019)
9. Pelka, O., Friedrich, C.M., García Seco de Herrera, A., Müller, H.: Medical image understanding: Overview of the ImageCLEFmed 2020 concept prediction task. In: CLEF2020 Working Notes. CEUR Workshop Proceedings, vol. 12260. CEUR-WS.org, Thessaloniki, Greece (September 22-25 2020)
10. 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
11. Tan, T., Li, Z., Liu, H., Zanjani, F.G., Ouyang, Q., Tang, Y., Hu, Z., Li, Q.: Optimize transfer learning for lung diseases in bronchoscopy using a new concept: sequential fine-tuning. IEEE journal of translational engineering in health and medicine **6**, 1–8 (2018)
12. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems. pp. 3320–3328 (2014)

13. Zhang, Y., Wang, X., Guo, Z., Li, J.: Imagesem at imageclef 2018 caption task: Image retrieval and transfer learning. In: CLEF2018 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Avignon, France (September 10-14 2018)