IRIT & MISA at Image CLEF 2017 - Multi label classification

Nomena Ny Hoavy[1], Josiane Mothe[2], Mamitiana Ignace Randrianarivony[1]

[1] MISA,Univ. Antananarivo ,Madagascar
[2] IRIT, UMR5505, CNRS & Univ. Toulouse, France,

Abstract. In this paper, we describe the participation of the Mami team at ImageCLEF 2017 for the Image Caption task. We participated to the concept detection subtask which aims at assigning a set of concept labels to a medical image. We used transfer learning method with VGG19 model for feature extraction to solve this task, and apply those features as input of a new neural network.

Keywords: Information systems, Information retrieval, multi-label classification,convolutional neural network, transfer learning, fine-tuning

1 Introduction

In this paper, we describe the participation of the Mami team to ImageCLEF 2017 for the Image Caption task. This team results from a collaboration between IRIT SIG team from the Université de Toulouse (France) and MISA from the Universit d'Antananarivo (Madagascar).

The Image Caption task consists in two subtasks: concept detection and caption prediction. The main goal of the concept detection task is to identify the existence of relevant biomedical concepts in the medical images delivered by the organizers (see [7] for a general overview and [4] for the Caption task overview as well as its web page [10]).

This problem can be see as a classification problem where each class corresponds to a concept label. As each image may contain multiple concepts, we tackled this task as a multi-label classification problem. The multi-label classification consists to associate a given instance $x_i \in X$ to a set of labels $Y_i = y_{i1}, .., y_{il_i} \in Y$. Where x_i is the instance of the i-st image and Y_i the set of concepts x_i belongs to. l_i is the number of concepts in Y_i.

We used deep learning method to solve this problem. Deep convolutional neural networks are considered among the best classifiers for single-label image classification [12,14].

In this work, we adapt a convolutional neural network with transfers learning to multi-label classification task.

In the next section we present in Section 2 a brief analysis of the dataset used in the concept detection subtask. We then present in Section 3 the method we developed. Section 4 presents the results while Section 4 concludes the paper.

2 Dataset and data exploration

The dataset we used was delivered by the image CLEF Lab and contains 3 subsets: the training set is composed of 164,614 images and 20,463 labels; the validation set consists of 10,000 images and 7,070 labels from which 309 are not in the training set nor in the test set. The test set contains 10,000 images.

Table 2 presents some characteristics of the training set.

	Number of labels per image
mean	5.58
std	4.47
min	0.00
25%	3.00
50%	4.00
75%	7.00
max	75.00

	Number of images per label
mean	44.95
std	320.06
min	1.00
25%	1.00
50%	3.00
75%	13.00
max	17,998.00

Table 1. Number of labels per image and number of images per label in the training set.

We found that 3.9% of the training images and 3.79% of the validation images have no labels.

3 Method we developed

As said previously, we used a convolutional neural network. For the transfer learning [16], we used the Oxford VGG19 model from Simonyan *et al.* [14]. This model performed well at ImageNet Large Scale Visual Recognition Challenge (ILVRC 2012 [11]) for classifying images over 1,000 classes [14].

As illustrated in Figure 1, the Oxford VGG19 model consists in 19 layers. To adapt the pre-trained model [3] to the imageCLEF Concept detection task, we froze the parameters of the 18 first layers and add 2 new layers above it. Our learning consists to train the 2 new layers to map the images to the corresponding concepts. The goal of the training is to assign a positive score to concepts the given image belongs to. The method consists in 2 steps that are described in the following sub-sections.

3.1 Processing

The first step of the process we developed is to forward each image to the VGG19 model and extract the output of the second to last layer of the VGG19 CNN

[3] The VGG19 model trained on ILSRVC-2014 http://www.robots.ox.ac.uk/%7Evgg/research/very_deep/

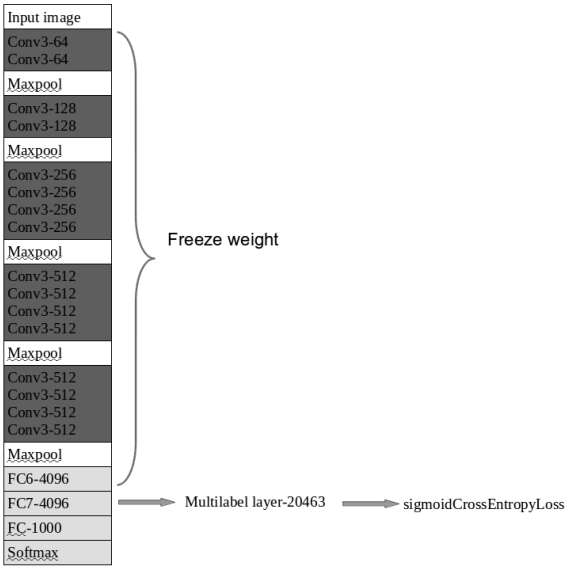

Fig. 1. Transfer learning with VGG19 CNN. The top 18 layers (conv3-64 to FC7-4096) are frozen and the second to last and last are replaced with a multilabel layer with 20,483 nodes (number of concepts or labels). More specifically, conv3-64 is transformed into FC7-4096 and FC-1000 into FC-20483. This deep neural network model is then trained using sigmoid cross entropy loss function.

(fully connected layer 7(fc7)). This output, named feature vector, is a 4,096 dimensional vector. This feature vector summarizes the representation of the forwarded image. This representation helps to gain time and space on disk, as suggested by Sharif *et al.* [13]. Moreover, Bengio *et al.* suggested that learning from such a feature vector gives a great potential in various vision recognition tasks [2,3].

Then, the feature vector is sent to the first new layer we add for multi-label classification ; this process results in scores associated to each concept or label. To train our model for assigning positive scores to the relevant concepts, we adopted the sigmoid cross entropy as loss function.

Let us define:
$\hat{p}_n = \sigma(x_n) \in [0, 1]$,
σ the sigmoid function,
x_n the predicted score for the concept n,
$p_n = 1$ if the image belongs to the concept n and 0 otherwise.
Formally, the sigmoid cross entropy loss [1] E is :

$$E = \frac{-1}{n} \sum_{n=1}^{N} [p_n \log \hat{p}_n + (1 - p_n) \log(1 - \hat{p}_n)]. \quad (1)$$

3.2 Training

We used the test/validation method to train our model performance and its capacity to generalize on new data. For that, we kept the data sets as split by the organizers. We trained our model using the training set. The validation set serves then as a new data set to validate the model performance and to tune the hyper parameters of the model. We trained our model with a stochastic gradient descent method using 0,0001 as learning rate (which corresponds to 1/10 of the pre-trained model initial learning rate) and a batch size of 128.

We preserved the momentum default value: 0.9 and set the weight decay 5.10^{-4}. After 150 epochs, training stopped and the resulting model was stored although it converges to E=42.547 . The delivered model is applied to predict the concepts in the validation and test data sets; the results are presented in the following section.

4 Results and Discussion

The task organizers suggested F1-score [4] to evaluate the results. F1-score is the average of F1-scores. It represents the weighted average of precision and recall.

We also added 2 other measures to check the performance of our model on the validation dataset : exact matching ratio measure, Hamming loss. The results are detailed in Table 3. As we predicted a vector, the exact matching ratio corresponds to the percentage of correctly predicted vector elements. This measure does not take into account partially correct predictions. The formula is:

$$MR = \frac{1}{N}\sum_{i=1}^{N} I(y_i = x_i). \tag{2}$$

where N is the number of samples, y_i the ground truth vector, and x_i is the prediction vector. The Hamming loss (HamLoss) computes the percentage of labels that are misclassified in a multi-label classification, i.e., relevant labels that are not predicted or irrelevant labels that are predicted [8]. The formula is given by:

$$HamLoss = \frac{1}{N}\sum_{i=1}^{N} \frac{xor(x_i, y_i)}{L}. \tag{3}$$

where N is the number of samples, L the number of labels, y_i the ground truth, and x_i is the prediction.

Table 2 presents the results for our model when evaluated on the validation data set with the different measures as mentioned before. The same configuration of the model was used for our unique run named DET_CORRECTED_mami_resulat.txt that we submitted to ImageCLEF and that was evaluated by the organizers on the test set (see Table 3 for the results).

Measure Names	Measure Values
F1-score	0.047
Exact matching ratio	0.041
Hamming loss	0.0002

Table 2. Evaluation on the validation dataset

F1-score	Run name
0.046	DET_CORRECTED_mami_resulat.txt

Table 3. F1-score evaluated by the organizers on the test set.

5 Conclusion

The model we developed to participate to ImageCLEF 2017 uses a transfer learning based on VGG19 and that we adapt for multi-label classification. Using this model, we obtain a low F1-score (0.0462) on the test set. Our result is far from the other teams. We expect to improve this score by using more appropriate machine that will make it possible to make a better training. We would like to make a finer tuning and deeper network. Alternatively, we could experiment with another classifier like SVM to classify the feature vectors that we extracted. Also, one of the main issue we faced was to handle the medical compound images (like formula, graphs,) that is more complex to process. We envisage to tackle this problem using Region based CNN (R-CNN) [5] [9]. On other hand, previous works [6] [15] can be used to extract relevant regions that can be used for a MIML (Multi-Instance Multi-Label) learning [17]. Unfortunately, our current facilities in terms of computing did not allow us to do more while the competition was still open.

References

1. Beckham, C., Pal, C.: A simple squared-error reformulation for ordinal classification. arXiv preprint arXiv:1612.00775 (2016)
2. Bengio, Y., Courville, A., Vincent, P.: Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence 35(8), 1798–1828 (2013)
3. Coates, A., Lee, H., Ng, A.Y.: An analysis of single-layer networks in unsupervised feature learning. Ann Arbor 1001(48109), 2 (2010)
4. Eickhoff, C., Schwall, I., García Seco de Herrera, A., Müller, H.: Overview of ImageCLEFcaption 2017 - image caption prediction and concept detection for biomedical images. CLEF working notes, CEUR (2017)
5. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 580–587 (2014)
6. Hosang, J., Benenson, R., Schiele, B.: How good are detection proposals, really? arXiv preprint arXiv:1406.6962 (2014)

7. Ionescu, B., Müller, H., Villegas, M., Arenas, H., Boato, G., Dang-Nguyen, D.T., Dicente Cid, Y., Eickhoff, C., Garcia Seco de Herrera, A., Gurrin, C., Islam, B., Kovalev, V., Liauchuk, V., Mothe, J., Piras, L., Riegler, M., Schwall, I.: Overview of ImageCLEF 2017: Information extraction from images. In: CLEF 2017 Proceedings. Lecture Notes in Computer Science, vol. 10456. Springer, Dublin, Ireland (September 11-14 2017)
8. Mencía, E.L., Fürnkranz, J.: Efficient multilabel classification algorithms for large-scale problems in the legal domain. In: Semantic Processing of Legal Texts, pp. 192–215. Springer (2010)
9. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. pp. 91–99 (2015)
10. roger: ImageCLEFcaption. http://imageclef.org/2017/caption/ (2017), [Online; accessed 17-May-2017]
11. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International Journal of Computer Vision 115(3), 211–252 (2015)
12. Shankar, S., Garg, V.K., Cipolla, R.: Deep-carving: Discovering visual attributes by carving deep neural nets. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3403–3412 (2015)
13. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: Cnn features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 806–813 (2014)
14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
15. Uijlings, J.R., Van De Sande, K.E., Gevers, T., Smeulders, A.W.: Selective search for object recognition. International journal of computer vision 104(2), 154–171 (2013)
16. 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)
17. Zhou, Z.H., Zhang, M.L., Huang, S.J., Li, Y.F.: Multi-instance multi-label learning. Artificial Intelligence 176(1), 2291–2320 (2012)