A Comparison of Deep Learning with Global Features for Gastrointestinal Disease Detection

Konstantin Pogorelov[1,2], Michael Riegler[1], Pål Halvorsen[1,2],
Carsten Griwodz[1,2], Thomas de Lange[3], Kristin Ranheim Randel[2,3], Sigrun Losada Eskeland[4],
Duc-Tien Dang-Nguyen[5], Olga Ostroukhova[8], Mathias Lux[6], Concetto Spampinato[7]
[1]Simula Research Laboratory, Norway [2]University of Oslo, Norway [3]Cancer Registry of Norway, Norway
[4]Vestre Viken Hospital Trust, Norway [5]Dublin City University, Ireland [6]University of Klagenfurt, Austria
[7]University of Catania, Italy [8]Research Institute of Multiprocessor Computation Systems n.a. A.V. Kalyaev, Russia
konstantin@simula.no,michael@simula.no

ABSTRACT

This paper presents our approach for the 2017 Multimedia for Medicine Medico Task of the MediaEval 2017 Benchmark. We propose a system based on global features and deep neural networks, and preliminary results comparing the approaches are presented.

1 INTRODUCTION

Following the initiative to investigate how multimedia can improve medical systems [15], the 2017 Multimedia for Medicine Medico Task [18] addresses the challenge of detecting diseases based on multimedia data collected in hospitals [13], i.e., the task focuses on detecting abnormalities, diseases and anatomical landmarks in images in the gastrointestinal (GI) tract. There do exist some proposals in this area using various approaches [20, 21], and in this paper, we describe our solutions, based on both our *global-features-based* and *neural-network-based* EIR prototypes [12, 14, 16, 17].

2 CLASSIFICATION APPROACHES

The proposed approaches are based on the hypothesis that GI tract diseases and findings can be recognized and classified based on color, shape and texture properties. In this challenge, there is no detailed ground truth ROIs provided for the training dataset, thus, already existing and well performing approaches to objects recognition are not suitable for this particular task. Moreover, a relatively low amount of training data is provided making it difficult to use modern convolutional neural network (CNN) image segmentation and region-based classification approaches. Furthermore, some objects like polyps and resection margins have a compact body and can be easily differentiated from the surrounding tissue, but other findings like ulcerative colitis have only tissue with a slightly different color properties. To address these different detection challenges, we present 17 different approaches that implement our idea of using visual properties of images for performing multi-class classification with the limited training set size. For the final classification step, we use the WEKA machine learning support library [7] which is an open source collection of algorithms for machine learning and data mining. For all the approaches based on global features (GFs), we use Lucene Image Retrieval (LIRE) [10], an open source implementation of global and local features extraction and comparison. For all the deep-learning-based approaches, we use Keras [3], an open source high-level neural networks API with Google Tensorflow [1] as a computational back-end.

MediaEval'17, 13-15 September 2017, Dublin, Ireland

2.1 Global-features-based

For the GF-based approaches, we use features that represent the overall image visual properties, they are easy and fast to calculate, and they can be used for image comparison, distance computing and image collection search. Here, we use the indexes of visual features extracted from training image set. A classifier is used to search the index for the image that is most similar to a given input image. The GFs we use are JCD, Tamura, Color Layout, Edge Histogram, Auto Color Correlogram and Pyramid Histogram of Oriented Gradients [10]. We decided for these combinations based on our previous findings and experiments in [14, 16]. Multi-class classification is implemented as an additional classification step to determine the final image class based on the the ranked lists of a search-based classifier for each class of findings. We use the random tree (RT), random forest (RF) and logistic model tree (LMT) classifiers [7] from WEKA.

2.2 Deep-features-based

For the deep-features-based approaches, we use a combined method with deep residual networks for image recognition as features extractor and machine-learning classifier with the input of extracted deep-features as a multi-class classifier. We use the Inception v3 [19] and ResNet50 [8] models pre-trained on a set of general images. The models were modified in order to produce numerical probability output for all recognized object classes. Then, we use the class (concept) probabilities (1000 values for both networks) directly in the *Concepts* runs. For the *Features* runs, we have used the same pre-trained models without including the fully-connected layer at the top of the network, which give us an output of high-level feature probabilities (16384 values for Inception v3 and 2048 for ResNet50). Finally, we combine the probabilities by simple early fusion in one big vector of floating point numbers and use it as an input for the same classifiers we used in the GF-based approaches.

2.3 CNN-based

For the CNN-based approach, we created and trained a custom CNN from scratch. Our CNN consist of six convolution layers. As an activation function, we used the rectified linear unit (ReLU) [6] and maxpooling for pooling. In all the layers, we also included a 0.5 dropout, and the final classification step was performed using two dense layers with first ReLU and then Sigmoid as activation functions. Both networks were trained for 200 epochs using the Adam optimizer [9].

2.4 Transfer-learning-based

For the transfer-learning-based (TFL) approach, we use the pre-trained Inception v3 [19] model and transfer learning technique [2] to train the network on our specific training set. We re-trained the base model and fine-tuned the last layers on the training set following the DeCAF approach [5]. We did not perform complex data augmentation and only relied on transfer learning. We froze all the basic convolutional layers of the network and only retrained the two top dense layers. The dense layers were retrained using the RMSprop [4] optimizer that allows an adaptive learning rate during the training process. After 1,000 epochs, we stopped the retraining of the dense layers and started fine tuning the convolutional layers. For that step, we did the analysis of the Inception v3 model layers structure and decided to apply the fine-tuning on the top two convolutional layers. For this training step, we used a stochastic gradient descent method with a low learning rate to achieve the best effect in terms of speed and accuracy [11].

3 EXPERIMENTAL RESULTS

First, we have performed an initial evaluation of the approaches using the development dataset only randomly splitting it into new training and test sets with the equal number of 2,000 images in each. We assessed 17 different methods executed in 17 internal runs using the new sets generated. An overview of the conducted internal runs can be found in table 1 where we provide the measured performance metrics [13]. We can see that not all our approaches can perform efficiently on the given dataset. In general, we can conclude that for all the machine-learning-based classification approaches, the *LMT* classifier is performing the best, the *RF* classifier is slightly worse, and the *RT* classifier performs the worst. The *6 Layers CNN* and *Inception v3 TFL* approaches performs with the comparable precision, but *Inception v3 TFL* have slightly better results. The *Inception v3 Concepts* and *ResNet50 Concepts* approaches performs with the comparable precision too, but all the *ResNet50 Concepts* approaches perform slightly better. The *Inception v3* Features approaches perform the worst compared to all other features-based approaches even for the efficient *LMT* classifier, which can be caused by the huge feature values vector generated by the Inception v3 network. Finally, the best performing approach is the *ResNet50 Features* approach with the *LMT* classifier showing the performance of 0.828 for R_K and 0.856 for F1 score.

Based on the initial evaluation, we have selected the five different approaches for the official competition submission. The approaches selected (see table 2) are the best performing in the internal runs while keeping as much diversity of the methods as possible. The official evaluation results provided by the organizers is presented in table 2. The best performing approach is again the *ResNet50 Features* approach with the *LMT* classifier (run #4) with the R_K value of 0.802 and F1 score of 0.826. The confusion matrix of this run is presented in table 3. The often miss-classified classes are *Esophagitis* and *Z-line* that is caused by the nature of the used visual features. Both of these classes consist of pictures of *Z-Line*, but *Esophagitis* is the inflammation of *Z-Line* area, thus local image characteristics should be used to distinguish between these classes more precisely. The same reason can explain some cases of miss-classification with *Dyed and Lifted Polyps*, *Dyed Resection Margins* and *Polyps* classes.

Table 1: Initial performance evaluation based on the random split of the task development dataset.

Method	PREC	REC	SPEC	ACC	F1	R_K	FPS
6 Layer CNN	0.659	0.642	0.947	0.900	0.640	0.600	43
Inception v3 TFL	0.700	0.695	0.961	0.925	0.704	0.661	53
Inception v3 Concepts RT	0.405	0.402	0.915	0.851	0.403	0.318	66
Inception v3 Concepts RF	0.704	0.701	0.957	0.925	0.699	0.659	50
Inception v3 Concepts LMT	0.771	0.763	0.970	0.940	0.745	0.721	37
Inception v3 Features RT	0.287	0.288	0.898	0.822	0.287	0.186	56
Inception v3 Features RF	0.436	0.447	0.921	0.862	0.436	0.362	43
Inception v3 Features LMT	0.444	0.438	0.920	0.859	0.438	0.360	30
ResNet50 Concepts RT	0.507	0.500	0.929	0.875	0.501	0.431	88
ResNet50 Concepts RF	0.762	0.753	0.965	0.938	0.751	0.720	78
ResNet50 Concepts LMT	0.781	0.799	0.983	0.970	0.797	0.750	53
ResNet50 Features RT	0.479	0.478	0.925	0.869	0.477	0.403	79
ResNet50 Features RF	0.790	0.782	0.980	0.928	0.769	0.763	70
ResNet50 Features LMT	**0.841**	**0.839**	**0.985**	**0.972**	**0.856**	**0.828**	**46**
6 Global Features RT	0.576	0.578	0.940	0.894	0.576	0.516	130
6 Global Features RF	0.744	0.734	0.981	0.951	0.784	0.705	105
6 Global Features LMT	0.800	0.785	0.980	0.964	0.781	0.748	80

Table 2: The official classification performance evaluation results (provided by the organizers) of the submitted runs.

Run #	Method	PREC	REC	SPEC	ACC	F1	R_K	FPS
1	Inception v3 TFL	0.735	0.715	0.963	0.725	0.725	0.686	53
2	Inception v3 Concepts LMT	0.742	0.738	0.963	0.934	0.737	0.701	37
3	ResNet50 Concepts LMT	0.766	0.763	0.966	0.941	0.761	0.729	53
4	**ResNet50 Features LMT**	**0.829**	**0.826**	**0.975**	**0.957**	**0.826**	**0.802**	**46**
5	6 Global Features LMT	0.766	0.760	0.966	0.940	0.757	0.727	80

Table 3: Confusion matrix for the ResNet50 Features LMT run #4.

		Detected class							
		A	B	C	D	E	F	G	H
Actual class	Esophagitis (A)	**319**	0	4	2	174	0	1	0
	Dyed and Lifted Polyps (B)	0	**385**	0	6	0	59	47	3
	Pylorus (C)	6	0	**460**	7	19	0	7	1
	Ulcerative colitis (D)	5	0	1	**460**	0	2	14	18
	Z-line (E)	104	0	8	0	**385**	0	3	0
	Dyed Resection Margins (F)	0	84	1	5	0	**403**	5	2
	Polyps (G)	1	3	1	19	1	1	**441**	33
	Cecum (H)	0	1	0	29	0	0	18	**452**

4 CONCLUSION

In this paper, we presented 17 different combined approaches designed for multi-class classification of medical imaging data with the limited training dataset. We presented a novel comparison of the performance of the various visual-features-based methods with traditional custom CNN and Inception v3 with transfer-learning-based approaches. We used modified Inception v3 and ResNet50 networks and the LIRE library for the features extraction, with machine-learning classification algorithms from WEKA. Despite the limited training dataset and a presence of visually similar image classes, we achieved a good multi-class classification performance with the R_K value of 0.802 and a classification speed of 46 frames per second. For our future research, we will investigate the combined approach with the fusion of multiple deep-network-based feature extractors for the initial coarse image classification together with the fine-tuned local-feature-based sub-classification for the efficient cross-class detection between visually similar images.

ACKNOWLEDGMENTS

This work is founded by the FRINATEK project "EONS" #231687.

REFERENCES

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, and others. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467* (2016).

[2] Souad Chaabouni, Jenny Benois-Pineau, and Chokri Ben Amar. 2016. Transfer learning with deep networks for saliency prediction in natural video. In *Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP)*. 1604–1608.

[3] François Chollet. 2015. Keras: Deep learning library for theano and tensorflow. (2015). https://keras.io/ Accessed: 2017-09-01.

[4] YN Dauphin, H De Vries, J Chung, and Y Bengio. 2015. RMSProp and equilibrated adaptive learning rates for non-convex optimization. *arXiv preprint arXiv:1502.04390* (2015).

[5] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. 2014. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, Vol. 32. 647–655.

[6] Richard HR Hahnloser, Rahul Sarpeshkar, Misha A Mahowald, Rodney J Douglas, and H Sebastian Seung. 2000. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature* 405, 6789 (2000), 947–951.

[7] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. 2009. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter* 11, 1 (2009), 10–18.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778.

[9] Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[10] Mathias Lux, Michael Riegler, Pål Halvorsen, Konstantin Pogorelov, and Nektarios Anagnostopoulos. 2016. LIRE: open source visual information retrieval. In *Proceedings of the 2016 ACM Conference on Multimedia Systems (MMSys)*. Article no. 30.

[11] Jiquan Ngiam, Adam Coates, Ahbik Lahiri, Bobby Prochnow, Quoc V Le, and Andrew Y Ng. 2011. On optimization methods for deep learning. In *Proceedings of the 28th International Conference on Machine Learning (ICML)*. 265–272.

[12] Konstantin Pogorelov, Sigrun Losada Eskeland, Thomas de Lange, Carsten Griwodz, Kristin Ranheim Randel, Håkon Kvale Stensland, Duc-Tien Dang-Nguyen, Concetto Spampinato, Dag Johansen, Michael Riegler, and others. 2017. A holistic multimedia system for gastrointestinal tract disease detection. In *Proceedings of the 8th ACM Conference on Multimedia Systems (MMSys)*. 112–123.

[13] Konstantin Pogorelov, Kristin Ranheim Randel, Carsten Griwodz, Sigrun Losada Eskeland, Thomas de Lange, Dag Johansen, Concetto Spampinato, Duc-Tien Dang-Nguyen, Mathias Lux, Peter Thelin Schmidt, Michael Riegler, and Pål Halvorsen. 2017. Kvasir: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys)*. 164–169.

[14] Konstantin Pogorelov, Michael Riegler, Sigrun Losada Eskeland, Thomas de Lange, Dag Johansen, Carsten Griwodz, Peter Thelin Schmidt, and Pål Halvorsen. 2017. Efficient disease detection in gastrointestinal videos – global features versus neural networks. *Multimedia Tools and Applications* (2017), 1–33. https://doi.org/10.1007/s11042-017-4989-y

[15] Michael Riegler, Mathias Lux, Carsten Gridwodz, Concetto Spampinato, Thomas de Lange, Sigrun L Eskeland, Konstantin Pogorelov, Wallapak Tavanapong, Peter T Schmidt, Cathal Gurrin, Dag Johansen, Håvard Johansen, and Pål Halvorsen. 2016. Multimedia and Medicine: Teammates for better disease detection and survival. In *Proceedings of the 2016 ACM Multimedia Conference (ACM MM)*. 968–977.

[16] Michael Riegler, Konstantin Pogorelov, Sigrun Losada Eskeland, Peter Thelin Schmidt, Zeno Albisser, Dag Johansen, Carsten Griwodz, Pål Halvorsen, and Thomas de Lange. 2017. From Annotation to Computer Aided Diagnosis: Detailed Evaluation of a Medical Multimedia System. *Transactions on Multimedia Computing, Communications and Applications* 9, 4 (2017).

[17] Michael Riegler, Konstantin Pogorelov, Pål Halvorsen, Thomas de Lange, Carsten Griwodz, Peter Thelin Schmidt, Sigrun Losada Eskeland, and Dag Johansen. 2016. EIR - Efficient Computer Aided Diagnosis Framework for Gastrointestinal endoscopies. In *Proceedings of the 14th International Workshop on Content-based Multimedia Indexing (CBMI)*. 1–6.

[18] Michael Riegler, Konstantin Pogorelov, Pål Halvorsen, Kristin Ranheim Randel, Sigrun Losada Eskeland, Duc-Tien Dang-Nguyen, Mathias Lux, Carsten Griwodz, Concetto Spampinato, and Thomas de Lange. 2017. Multimedia for Medicine: The Medico Task at MediaEval 2017. In *Proceedings of the 2017 MediaEval Benchmarking Initiative for Multimedia Evaluation*.

[19] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. 2015. Rethinking the inception architecture for computer vision. *arXiv preprint arXiv:1512.00567* (2015).

[20] Yi Wang, Wallapak Tavanapong, Johnny Wong, JungHwan Oh, and Piet C De Groen. 2011. Computer-aided detection of retroflexion in colonoscopy. In *Proceeding of the 24th International Symposium on Computer-Based Medical Systems (CBMS)*. 1–6.

[21] Yi Wang, Wallapak Tavanapong, Johnny Wong, Jung Hwan Oh, and Piet C De Groen. 2015. Polyp-alert: Near real-time feedback during colonoscopy. *Computer methods and programs in biomedicine* 120, 3 (2015), 164–179.