Medical image retrieval using Resnet-18

Ayyachamy, Swarnambiga, Alex, Varghese, Khened, Mahendra, Krishnamurthi, Ganapathy

Swarnambiga Ayyachamy, Varghese Alex, Mahendra Khened, Ganapathy Krishnamurthi, "Medical image retrieval using Resnet-18," Proc. SPIE 10954, Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications, 1095410 (15 March 2019); doi: 10.1117/12.2515588

SPIE.

Event: SPIE Medical Imaging, 2019, San Diego, California, United States

Medical image retrieval using ResNet-18

Swarnambiga Ayyachamy, Varghese Alex, Mahendra Khened, and Ganapathy Krishnamurthi

Medical Imaging and Reconstruction Lab, Department of Engineering Design,
Indian Institute of Technology Madras, Chennai-600036, India

ABSTRACT

Advances in medical imaging technologies have led to the generation of large databases with high-resolution image volumes. To retrieve images with pathology similar to the one under examination, we propose a content-based image retrieval framework (CBIR) for medical image retrieval using deep Convolutional Neural Network (CNN). We present retrieval results for medical images using a pre-trained neural network, ResNet-18. A multi-modality dataset that contains twenty-three classes and four modalities including (Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Mammogram (MG), and Positron Emission Tomograph (PET)) are used for demonstrating our method. We obtain an average classification accuracy of 92% and the mean average precision of 0.90 for retrieval. The proposed method can assist in clinical diagnosis and training radiologist.

Keywords: Classification, Convolutional neural network, Retrieval, Class, Anatomy, Mean average precision

1. INTRODUCTION

Medical images reveal internal structures (anatomy and physiology) to diagnose and treat disease. The storage and usage of electronic image data are problematic due to large dataset sizes. Content-Based Image Retrieval (CBIR) is the preferred approach to query large medical imaging databases. Suitable querying, i.e, Query by image content(QBIC), matching, indexing and searching techniques are required for efficient CBIR. In machine learning, the convolutional neural network has successfully been applied to analyze images. This article focuses on retrieval of 2D slices of medical images based on an input query image. The deep CNN model ResNet-18 is trained for classifying medical images. For Content-based medical image retrieval (CBMIR) the learned feature representations are used. The major contributions of this work are,

- A multimodal retrospective dataset is created based on disease and anatomy.
- A comprehensive and thorough analysis of the proposed system with retrieval efficacy and reasoning is presented.

2. RELATED WORK

2.1 Deep CNN in Medical Image Retrieval

Camlica et al.[1] suggest that by reducing the feature dimensionality the retrieval process could be done faster. Yan et al. proposed retrieval by examining the histogram of errors of image blocks for each image class[2]. A neural network model is trained using Stochastic Gradient Descent (SGD) with back propagation by Qayyum et al.[3]. This is considered as the main related work to our proposed model. Other than CNN, Bag of visual word by Foncubiuta et.al.[4] approach is a widely used technique to learn meaningful features from the dataset and describing documents or images in terms of the histogram of these features. Qayyum et al. implemented retrieval by classifying images using DCNN model with eight layers, five are convolutional layers and three are fully connected layers. From last three fully connected layers(FCL) of the trained model i.e, from FCL1 - FCL3 feature representations are extracted. Then for retrieval task comparative analysis is performed from the feature

Further author information: (Send correspondence to Swarnambiga Ayyachamy)
Swarnambiga Ayyachamy: E-mail:aswarnambiga@gmail.com, Telephone:+91-7904392157
Ganapathy Krishnamurthi: E-mail:gankrish@iitm.ac.in, Telephone:+91 44 2257 4745

Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications,
edited by Po-Hao Chen, Peter R. Bak, Proc. of SPIE Vol. 10954, 1095410

Table 1: Dataset for classification and retrieval task

Class name	Modality	Total set	Training Set	Validation set	Test Set
Brain	MR	234(11346)	164(4904)	47(4084)	23(2358)
Breast	MG	3198(3198)	2239(2239)	640(640)	319(319)
Cervical	MR	53(399)	37(278)	11(85)	5(36)
Chest	CT	54(239)	38(124)	11(66)	5(49)
Colon	CT	18(316)	13(206)	3(87)	2(46)
Esophagus	CT	16(526)	11(341)	3(138)	2(74)
Retina	PNG	89(267)	62(188)	8(54)	9(27)
Mandible	CT and MR	45(1308)	31(871)	9(266)	5(17)
Kidney	CT and MR	27(108)	19(85)	5(14)	3(18)
Lung	CT	58(330)	41(224)	11(65)	6(41)
Liver	CT	28(448)	20(320)	5(80)	3(96)
Lymph	CT		95(1900)	27(540)	14(280)
Ovarian	CT and MR	108(759)	76(456)	21(209)	11(188)
Pancreas	CT	82(7059)	57(4855)	17(1483)	8(721)
Phantom	CT	4(2814)	2(1780)	1(517)	1(517)
Prostate	MR	50(778)	35(581)	10(130)	5(67)
Rectum	CT and MR	3(82)	1(73)	1(3)	1(6)
Renal	CT and MR	199(971)	139(679)	40(192)	20(100)
Soft Tissue	CT and MR	3(1602)	1(534)	1(534)	1(534)
Stomach	CT	44(277)	31(211)	9(44)	4(44)
Bladder	CT and MR	47(2290)	33(1602)	9(391)	5(277)
Thyroid	CT	6(158)	4(135)	1(11)	1(24)
Uterine	CT and MR	18(477)	13(406)	3(45)	2(54)
Total		**4479(37652)**	**3094(21632)**	**880(9418)**	**441(5613)**

extracted. In contrast, we are training a different network (ResNet-18) and to implement retrieval we have used the hidden representation of classes from classification task. This is described in detail in later sections.

3. MATERIALS AND METHODS

In this work, a residual network[5] pre-trained on natural images is fine-tuned to classify medical images into 23 different classes. The trained classification network is then utilized to extract features so as to retrieve similar images from the training database.

3.1 Data

The network is trained on medical images (n=37652) which is acquired from 4479 patients. The images are acquired from 23 different anatomical regions. The dataset comprises of both magnetic resonance images and computed tomography images. Images arising from structures such as brain, cervical & prostate are acquired using MR while images from the chest, colon, esophagus, lung, liver to name a few are acquired using CT. The images from the brain, liver, prostate & mandible are obtained from the BRATS[6]; LITS[7]; PROMISE[8], Head and neck dataset[9] respectively. The eye dataset is created from Cornell university database[10]. Data from the remaining classes are downloaded from TCIA repository. The images from dataset are resized to the dimension of 256 $\times$ 256 and intensity is normalized to the range 0-1. The detailed list of different classes of images are given in Tab.1. For each volume of a patient, we select the slice which depicts the tumor clearly and a total of 5 adjacent slices to form the image database.

3.2 Training

A deep residual network (ResNet-18)[11] is initialized with pre-trained weights. The number of neurons in the classification layer is equivalent to the number of classes in the dataset (n=23). The parameters of networks are learned by minimizing the cross- entropy loss with SGD as the optimizer and learning rate set to 0.001. The learning rate is annealed by a factor of 0.01 at the end of every 3 epochs and the network is trained for a total of 25 epochs. The architecture of the network is depicted in Fig. 1.

3.3 Extraction of Hidden Representation of the training data

The penultimate layer in the classification is utilized to extract features from an input image. This leads encoding an image of size 256×256 to dimension of 512. Images from different classes are sequentially fed into the trained network so as to form hidden representation associated with each class. The process stated above leads to the formation of as many hidden representations as the number of classes.

3.4 Image retrieval

During the testing phase, the images are resized and normalized similar to the training data. The normalized data is then fed to the trained network. The trained network generates 2 outputs viz; the class associated to the image and the deep hidden representation generated by the network [3].

Based on the class generated by the network, the hidden representation generated for the test image is compared with hidden representation of all images in the training database with the same class. Mean squared error is the measure used to compute the similarity between the test images and the images in the training database. The top 10 images in the training database with highest similarity are then displayed to the user. The pipeline utilized during the training phase and testing phase is given in Fig.2. A hidden representation for all 23 classes are attained by feeding all images in the training database under a particular class to the network Fig.3.

4. EXPERIMENTATION RESULTS

In this paper, we have used deep learning tool pytorch for developing and training the pretrained ResNet-18. In terms of classification and retrieval results the proposed method is evaluated.

4.1 Classification Performance

During inference, for a given test image, the trained network predicts the class of it and also generates the deep hidden representation. On the test set, the network achieves an overall accuracy of 92%.

The misclassification is higher in the classes such as kidney, rectum, and uterus when compared to other classes. However,the classification performance of Bladder, colon, chest, esophagus, and ovary is lesser when compared with other remaining classes.

4.2 Retrieval Performance

For the task retrieval of an image, the often used metrics to calculate the performance of the technique are Precision(P), Recall(R), Accuracy(A), Average precision(AP) and Mean average precision(MAP). Precision is defined as the number of relevant images in the retrieved images, while recall or sensitivity is defined as the fraction of relevant images from the entire training database and is given in Eqs. (1), (2), Eqs. (3) and (4) .

The average precision (AP) is the average of the precision value obtained for the set of Top K images for every query image after each retrieval. This value is then averaged for all queries in the category.

$$\text{Precision} = \frac{\text{No. of relevant images retrieved}}{\text{No. of images retrieved}} \tag{1}$$

$$\text{Recall} = \frac{\text{No. of relevant images retrieved}}{\text{No. of relevant images}} \tag{2}$$

$$\text{Accuracy} = \frac{\text{Images classified correctly}}{\text{All images classified}} \tag{3}$$

$$\text{Average precision} = \frac{1}{|R|} \cdot \sum_{i=1}^{n} \text{Precision}(i).\text{relevance}(i) \tag{4}$$

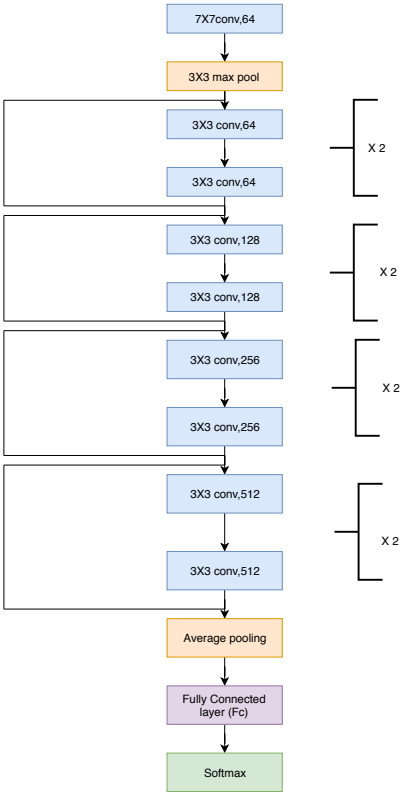

Figure 1: Residual network structure

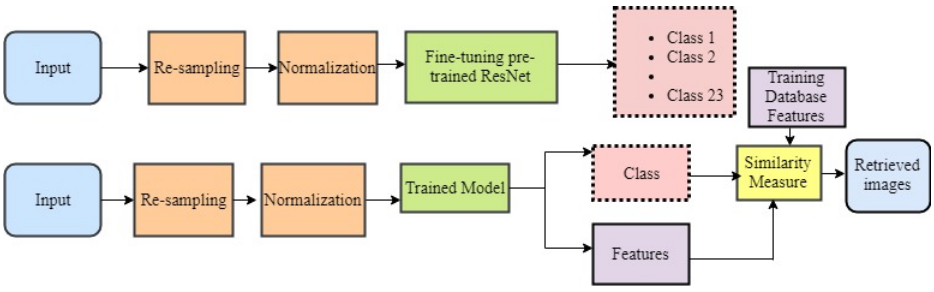

Figure 2: The proposed pipeline for retrieval of medical images using CNN

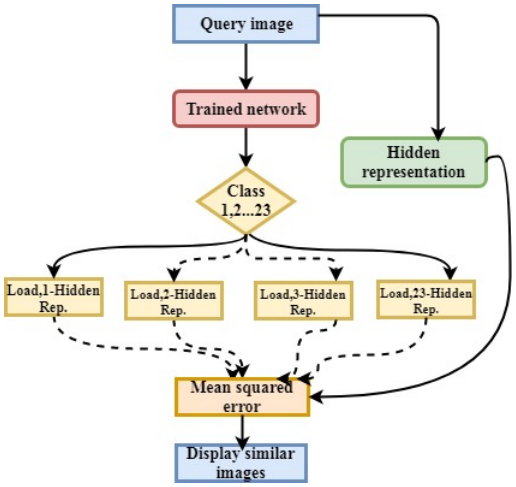

Figure 3: Pipeline for retrieval of images. Solid lines in the image represent input or output while dashed lines in the image represent inactive pathways as certain conditions for example class of the query image is not met

For 23 different classes, 5 possible queries are given for each classes. This results in a set of 115 queries over which the retrieval system can be tested and evaluated for mean average precision (MAP) [12, 3, 13]. The Average precision for retrieval is 95% and average recall is 90%.

4.3 Discussion

On the test set, the network achieved a classification accuracy of 92%. Among 23 classes the interclass variation in classification is higher in Bladder, kidney, Rectum, chest, esophagus, and uterus. The best performance in classification is given by Brain, Breast, Eye, Liver, Lung, Mandible, Prostate, Pancreas, Thyroid, cervical, lymph, Phantom, soft tissue, and stomach. The other classes like Colon, Ovarian and Renal accuracy are slightly lesser than the above-said classes. The intra-class variation can't be systematically "distinguished" by the raters in a fair way without the help of physicians. But for interclass variation, we can distinguish the differences relatively easily.

Based on the study on each slice of each class, it is observed that due to the presence of same organs in one or more classes i.e. overlapping regions or organs in anatomy, the classification accuracy varies. Because each slice has both distinctly referred and associated organs even for different anatomy. Brain, Thyroid, and mandible have overlapping organs (same organ present in both the classes)which we label as overlapping region 1. Similarly different sets of organs have overlapping regions which are presented in Tab.2.

From the image retrieval point of view, overlapping organs or regions do not affect performance. The retrieval results have a mean average precision of 0.90. Except for Bladder, kidney, Rectum, chest, esophagus and uterus the network is able to retrieve top 11 relevant images. These relevant retrieved images can be used as reference by the physician and radiologist to assist in critical cases in their day to day clinical practice. This helps them to avoid their direct visit to workstation to access the archive. The comparison of our network with similar retrieval is given in Tab. 3.

This approach aids in minimizing the memory requirement and computational overhead, as typically the number of images in the database is huge in a hospital scenario. Fig. 4, 5, 6, 7, illustrates the top 11 images that have been retrieved based on a query image.

In this article, we make use of pre-trained deep convolution neural network for the task of medical image retrieval. The proposed technique is unique in certain number of ways:

- The use of pre-trained network helps in reducing the need of extremely large database of labelled volumes.

Table 2: Overlapping Classes

Classes	**Overlapping Region**
Brain	Region 1
Thyroid	
Mandible	
Breast	Region 2
Chest	Region 3
Lung	
Lymph	
Stomach	Region 4
Esophagus	
Bladder	
Bladder	Region 5
Rectum	
Uterus	
Cervix	
Ovary	
Liver	Region 6
Stomach	
Pancreas	
Kidney	
Colon	
Phantom	Region 7
Soft Tissue Sarcoma	Region 8
Kidney	Region 9
Renal	
Uterus	Region 10
Ovary	
Rectum	
Prostate	Region 11

Table 3: Comparison of the ResNet-18 with the contemporary CBMIR systems

Method	**Image slices**	**Train Set**	**Test Set**	**Modalities**	**Classes**	**Mean average precision**
ResNet-18	38452	22,990	5778	MR,CT,PET,MG	23	0.90
Adnan et.al	7200	7200	5040	PT,MR, CT,PET,OPT(optical projection tomography)	24	0.69
Camlica et.al	14410	14410	12677	X-Ray	57	0.86

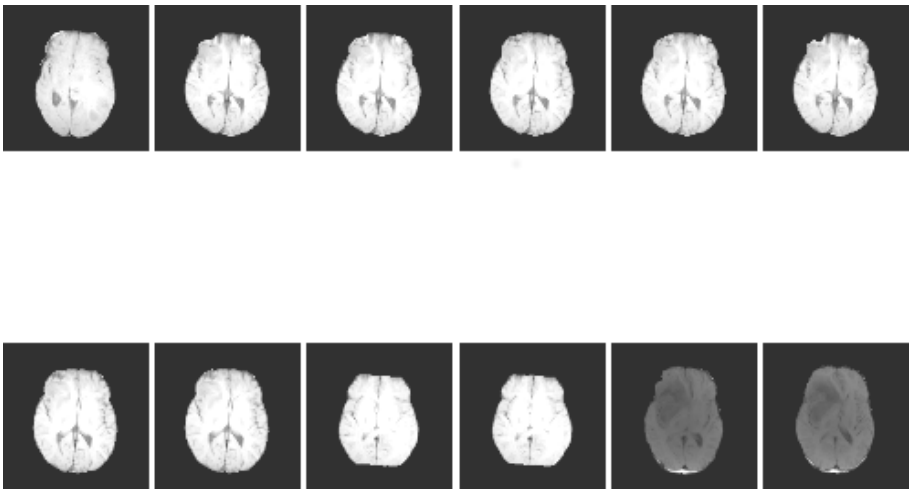

Figure 4: Retrieval results for Brain class. Left: Query image and Remaining images are the retrieved images

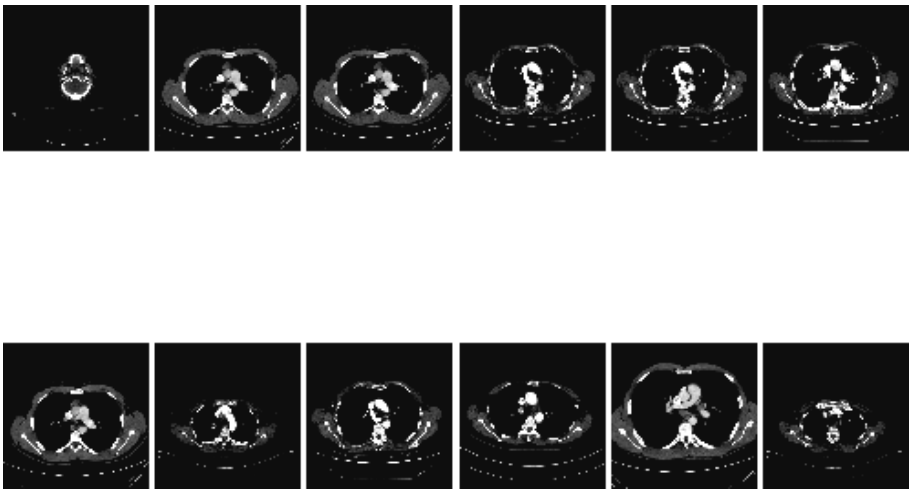

Figure 5: Retrieval results for Mandible class. Left: Query image and Remaining images are the retrieved images

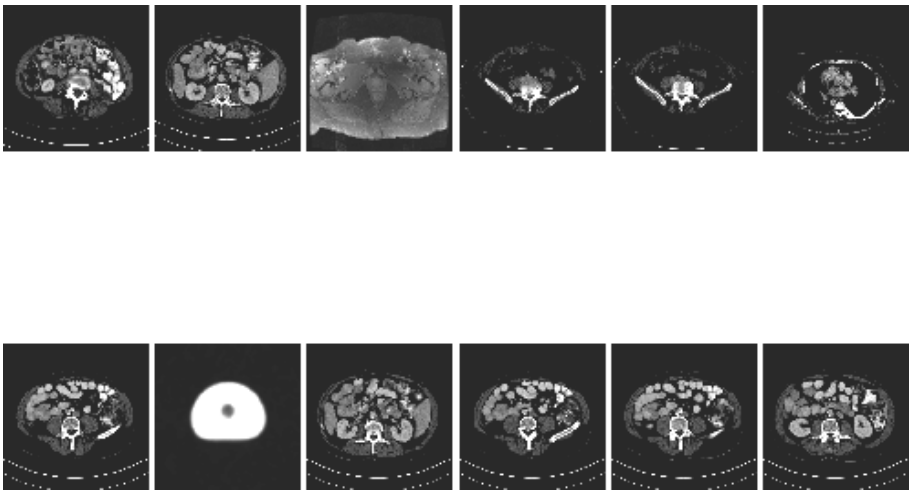

Figure 6: Retrieval results for Ovarian class. Left: Query image and Remaining images are the retrieved images

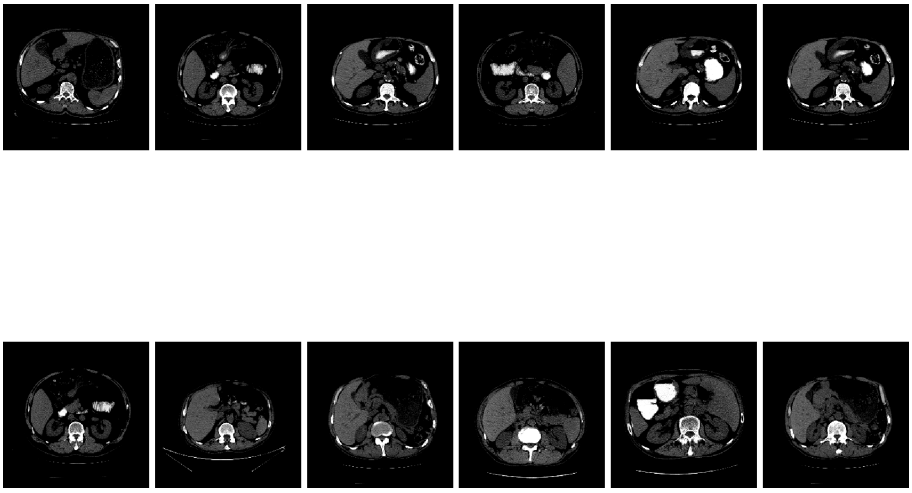

Figure 7: Retrieval results for stomach class. Left: Query image and Remaining images are the retrieved images

- The proposed network achieved an mean average precision for retrieval of 0.90 and classification accuracy of 92 % .
- The use of class information for data retrieval helps in the reducing computational overhead and will assist in clinical diagnosis even for a large dataset.

5. CONCLUSION

A pre-trained deep learning based framework for content based medical image retrieval using image classification and feature extraction is successfully implemented. The penultimate layer of the network is used to extract features for the retrieval task. Retrieval of medical images is performed based on the class prediction. The network achieves an average classification accuracy of 92% and mean average precision of 0.90 for retrieval. Future work is focused on including the patients' reports and adapting the network for combined text and medical image retrieval.

References

[1] Camlica, Z., Tizhoosh, H. R., and Khalvati, F., "Autoencoding the retrieval relevance of medical images," in [*Image Processing Theory, Tools and Applications (IPTA), 2015 International Conference on*], 550–555, IEEE (2015).

[2] Yan, Z., Zhan, Y., Peng, Z., Liao, S., Shinagawa, Y., Zhang, S., Metaxas, D. N., and Zhou, X. S., "Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition," *IEEE transactions on medical imaging* **35**(5), 1332–1343 (2016).

[3] Qayyum, A., Anwar, S. M., Awais, M., and Majid, M., "Medical image retrieval using deep convolutional neural network," *Neurocomputing* **266**, 8–20 (2017).

[4] Foncubierta and et.al, "Medical image retrieval using bag of meaningful visual words: unsupervised visual vocabulary pruning with plsa," *In Proceedings of the 1st ACM international workshop on Multimedia indexing and information retrieval for healthcare 2013*, 75–82 (2013).

[5] Girshick.R, Donahue, J., Darrell, T., and Malik, J., "Rich feature hierarchies for accurate object detection and semantic segmentation;Image net." http://www.image-net.org/. (Accessed: 15 September 2018).

[6] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al., "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging* **34**(10), 1993 (2015).

[7] Han, X., "Automatic liver lesion segmentation using a deep convolutional neural network method," *arXiv preprint arXiv:1704.07239* (2017).

[8] "Promise Challenge." https://promise12.grand-challenge.org/Organizers/. (Accessed: 15 September 2018).

[9] Raudaschl, P. and et.al, "Evaluation of segmentation methods on head and neck ct: Auto-segmentation challenge 2015," *Medical physics* **44**, 2020–2036 (2017).

[10] Kauppi, T. and et.al, "DIARETDB1 - Standard Diabetic Retinopathy Database Calibration level 1." http://www.it.lut.fi/project/imageret/diaretdb1/#DOWNLOAD. (Accessed: 15 September 2018).

[11] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," *http://image-net.org/challenges/LSVRC/2015/ and http://mscoco.org/dataset/#detections-challenge2015* **2015**, 1–12 (2015).

[12] Mller H and, M. N., D, B., and A., G., "A review of content-based image retrieval systems in medical applicationsclinical benefits and future directions," *International journal of medical informatics* **73**, 1–23 (2004).

[13] A Alzubi, A Amira, N. R., "Content-based image retrieval with compact deep convolutional features," *Neurocomputing* **249**, 95–105 (2017).