Automatic classification of medical X-ray images with convolutional neural networks

Xolisani Nkwentsha
dept. Electrical Engineering
University of Cape Town
Cape Town, South Africa
nkwxol003@myuct.ac.za

Anicet Hounkanrin
dept. Electrical Engineering
University of Cape Town
Cape Town, South Africa
hnkmah001@myuct.ac.za

Fred Nicolls
dept. Electrical Engineering
University of Cape Town
Cape Town, South Africa
fred.nicolls@uct.ac.za

***Abstract*—The classification of medical images is an important step for image-based clinical decision support systems. With the number of images taken per patient scan rapidly increasing, there is a need for automatic medical image classification systems that are accurate because manual classification and annotation is time-consuming and prone to errors. This paper focuses on automatic classification of X-ray image from the ImageCLEF 2009 dataset based on anatomical and biological information using the InceptionV3 model. The X-ray images are prepared and preprocessed with two different padding techniques, two image enhancement techniques and layering to convert the grey-scale images to 3-channel images to prepare them for InceptionV3. In terms of classification loss, constant padding with no enhancements had the best performance with an accuracy of 68.67% and a loss of 1.442. In terms of classification accuracy, constant padding with enhancement had the best performance with an accuracy of 71.34% and a loss of 1.608.**

***Index Terms*—Convolutional Neural Networks, X-ray, Classification, Transfer Learning, CBMIR.**

I. Introduction

With the advances in digital technology, medical facilities are producing large amounts of medical imaging data resulting in an exponential increase in medical image repositories. Over the years, medical imaging has and will continue to play an important role in modern healthcare aiding healthcare professionals by providing relevant information about anatomical and biological structures, which improve analysis and diagnosis. The rapid advancements in medical imaging, producing more data in different modalities has introduced a problem. Radiologists have to analyse more data and maintain quality and efficiency. Therefore the development of systems that can automatically interpret, analyse, and categorise medical images are needed to aid effective diagnosis.

Convolutional neural networks (CNNs) are driving major advances in many computer vision tasks, such as image classification [1], object detection [2], and image segmentation [3]. This motivates us to apply CNNs to perform automatic classification on medical images.

This paper focuses on automatic classification of X-ray image from the ImageCLEF 2009 dataset based on anatomical and biological information. The rest of the paper is organized as follows. Related work is presented in Section II, the proposed method is discussed in Section III, experimental setup is discussed in Section IV, results are shown and discussed in Section V and a conclusion is presented in Section VI.

II. Related Works

The aim of content-based medical image retrieval (CBMIR) is to aid healthcare professionals with diagnosis by searching through a medical image database to find images that are perceptually similar to a query image. CBMIR requires accurate and efficient search algorithms for the vast databases of medical images. X-ray imaging is the most widely used medical imaging modality today as many healthcare facilities are equipped with X-ray scanners and maintain their own database of images. Annotation and feature extraction is an important component in a CBMIR system.

A variety of methods exist for medical image feature extraction and classification. Ganesan and Subashini [4], automatically classify X-rays at the macro level (coarse level) using a support vector machine (SVM) classifier with six classes of X-ray images. Using the ImageCLEF 2007 dataset [5] their classification task started with extracting local invariant features from all images. A generative model such as probabilistic latent semantic analysis (PLSA) was applied on extracted features in order to provide more stable representation of the images. Subsequently, this representation was used as input to a discriminative SVM classifier to construct a classification model. Aboud et al [6], presents the results of an experimental evaluation of X-ray images classification in the ImageCLEF-2015 challenge. They found best classification results were obtained using the intensity, texture and HoG features and the KNN classifier. Pelka et al [7], used enhancement techniques to improve classification accuracy. To evaluate the image enhancement techniques, five classification schemes including the complete IRMA code were adopted.

Link to ImageCLEF 2009 dataset

978-1-7281-4162-6/20/$31.00 ©2020 IEEE

Two pretrained models (Inception-v3 and Inception-ResNet-v2, and Random Forest models) were trained using extracted Bag-of-Key points visual representations. The classification model performances were evaluated using the ImageCLEF 2009 Medical Annotation Task test set. The applied visual enhancement techniques proved to achieve better annotation accuracy in all classification schemes.

III. Method

In this work an automatic classification of X-ray images for medical image retrieval with convolutional neural networks is implemented.

A. ImageCLEF 2009 Dataset

ImageCLEF 2009 is a medical dataset that consists of X-ray images taken randomly from medical routine for a medical annotation task. The dataset represents multiple cases with respect to patient's age and gender, viewing position and pathologies. The training set consists of 12,677 grayscale images and the evaluation set has 1,733 grayscale images. The dataset uses a Image Retrieval in Medical Applications (IRMA) coding system to annotate the X-ray images. The dataset has 193 classes.

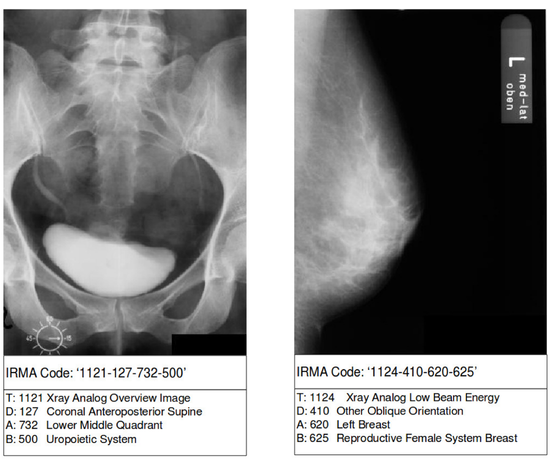

Fig. 1. Sample images from ImageCLEF 2009 dataset with their respective IRMA codes [7].

The IRMA coding system consists of four axes with three to four positions:

- T (technical): image modality
- D (directional): body orientation
- A (anatomical): body region examined
- B (biological): biological system examined

These axes create a short and unambiguous 13-character string notation (TTTT–DDD–AAA–BBB) [8]. The T-axis consists of a 4-character string which denote physical source, modality position, techniques and sub-techniques. The D-axis consists of a 3-character string which denote the orientation plane of the radiographs (e.g. coronal, sagittal, transversal, other) and has a more detailed specification in the second position (e.g. posteroanterior, anteroposterior). The A-axis consists of a 3-character string and denotes body regions and 2 hierarchical sub-regions. The B-axis consists of a 3-character string which denote organ system [8].

Fig 1 shows two radiographs with the IRMA codes 1121-127-732-500 and 1121-410-620-625, which represent "Xray Analog Overview Image; Coronal Anteroposterior Supine; Lower Middle Quadrant; Uropoietic System" and "Xray Analog Low Beam Energy; Other Oblique Orientation; Left Breast; Reproductive Female System Breast" [7]. A complete list of the IRMA code representation is given by Lehmann et al. [8]. High class imbalance was added to the ImageCLEF dataset to promote the function of prior knowledge encoded into the hierarchy [9]. The images in the test set were mainly from classes which had only a few examples in the training data, making annotation significantly harder [10].

B. Image Preprocessing

The X-ray images in the IRMA dataset are prepared and preprocessed with two padding techniques, Contrast Limited Adaptive Histogram Equalization (CLAHE), Non Local Means (NL-MEANS) and layering to convert the grey-scale images to 3-channel images. Fig 2 displays an X-ray image from the dataset that will be used to illustrate the effect of each technique.

Fig. 2. Original X-ray image with from ImageCLEF 2009 Dataset.

1) Padding: Images in the IRMA dataset have various sizes, hence padding is performed to resize the images to 512×512. Firstly the images are padded with a constant padding (using OpenCV's BORDER_CONSTANT) that fills surrounding pixels with zeros, as seen in Fig 3. In the second padding seen in Fig 4 also implemented Pelka et al, images are padded with their repetition (using OpenCV's BORDER_WRAP).

2) Image Resizing: In order to bring images of various sizes to the same size, all the images have been resized to 512×512 using nearest neighbor interpolation. The result of applying this preprocessing technique to a grayscale X-ray image of random size is illustrated in Fig 7.

3) CLAHE: CLAHE is a contrast enhancement method, modified from the Adaptive Histogram Equalization (AHE). It

Fig. 3. Constant padding

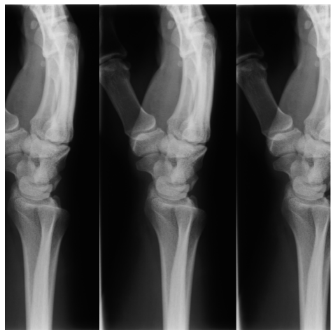

Fig. 4. Repetitive padding

Fig. 5. Constant padding

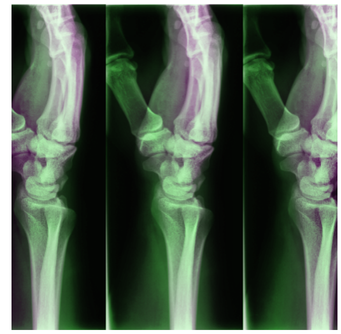

Fig. 6. Repetitive padding

is designed to be broadly applicable and having demonstrated effectiveness, especially for medical images [7]. The CLAHE output images were obtained using the OpenCV's implementation with the following parameters:

- clipLimit : 2.55
- tileGridSize : (8,8)

4) NL-MEANS: This is a digital image denoising method, based on a non local averaging of all present pixels in an image [7]. The NL-MEANS output images were obtained using the OpenCV's implementation with the following parameters:

- filter strength : 2
- templateWindowSize : 4
- searchWindowSize : 4

5) layering: Images in ImageCLEF dataset are grey-scale and need to be converted to 3-channel RBG images for the InceptionV3 pretrained model, we followed a simlar process to [7] to create the RBG images. The first layer was obtained from the CLAHE output of the same image and the second layer from the NL-MEANs output. The RGB image is obtained from adding the two layers to the original X-ray image. The resulting image is displayed in Fig 5 and Fig 6 for Constant padding and reflective padding respectively. Fig 8 shows the result of applying the Layering technique to a resized image.

C. Network Architecture and Implementation

Considering the number of images in the ImageCLEF 2009 Medical Annotation Task, the adaptation of transfer learning with the pre-trained InceptionV3 model was done. InceptionV3 is a convolutional neural network that is trained on more than a million images from the ImageNet database [11]. The network is 48 layers deep and can classify images into 1000 object categories. In our implementation, We remove the last layer and add a global average pooling layer followed by a 1024 kernel Dense layer with a relu activation and for the last layer we have a Dense layer with a softmax activation.

Fig. 7. Resizing

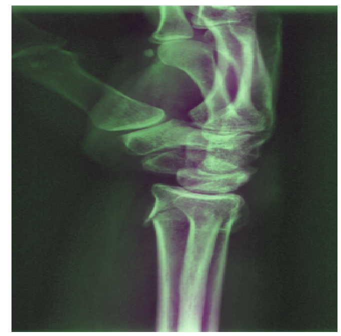

Fig. 8. Resizing & Layering

IV. Experimental Setup

For the InceptionV3 pretrained model, we used the Keras framework with a TensorFlow back-end. The training was run on 2 NVIDIA KeplerK40M and RAM of 12GB DDR5 per GPU.

A. Training

We used InceptionV3 with weights from ImageNet. The model was fine-tuned with ImageCLEF dataset with the following hyper-parameters: The model was trained using a stochastic gradient descent (SGD) optimizer with a learning rate of 0.01 and decay of 1e-6. We trained with a batch size of 16 for 25 epochs. The categorical crossentropy loss function was used during training.

The performance of the InceptionV3 model was evaluated on four different experiments with the prepossessing techniques detailed in Section III-B. We use two padding techniques, and each technique is ran twice without enhancements and with enhancements (Contrast Limited Adaptive Histogram Equalization and Non Local Means).

V. Results & Discussion

Evaluation of the performance of the InceptionV3 model for each experiment was performed on the official test set of 1,732 images from the ImageCLEF 2009 Medical Annotation Task. The classification Accuracy of the model with the different image processing techniques with and without the random rotation are listed in Table I. The image inputs are

Repetitive (with the original images being padded with their repetitions and 3-channels created by placing the same image in all 3 channels), Repetitive-Enhanced (with the repetitive padding and 3 channels created with the NL-Means and CLAHE outputs). Constant and Constant-Enhanced image input are similar to the repetitive inputs but have a constant padding.

TABLE I
CLASSIFICATION ACCURACIES

Image Input	No Enhancement	Enhanced
Resize	68.90%	64.87%
Repetitive	67.92%	67.80%
Constant	68.67%	**71.34%**

The corresponding classification losses of the model with the different image processing techniques with and without the random rotation are listed in Table II.

TABLE II
CLASSIFICATION LOSSES

Image Input	No Enhancement	Enhanced
Resize	1.520	1.611
Repetitive	1.612	1.508
Constant	**1.442**	1.608

With no Enhancements (ie. CLAHE and NL-Means), the model performed the best in terms of loss had an input with constant padding with a loss of 1.442, and in terms of accuracy the model performed the best with resized inputs and no padding with an accuracy of 68.90%.

With Enhancements (ie. CLAHE and NL-Means), the model performed the best in terms of loss with a repetitive padding with a loss of 1.508, and in terms of accuracy the model performed the best with constant padding with an accuracy of 71.34%.

For the repetitive padding, enhancements improved the loss but also reduced the accuracy. For the constant padding, enhancements also increased the loss but also improved the accuracy. For the resized inputs, enhancements did not improve performance in terms of loss and accuracy, the model decreased in performance.

Using an stochastic gradient descent (SGD) optimizer, instead of the root mean square propagation optimizer used by Pelka et al. improved the accuracy by approximately 20% compared to the accuracy reported by Pelka et al. [7]. Although Pelka et al. did not report on the loss, we replicated their results and got a loss of 3.71. Therefore the SGD optimizer also improved the loss.

VI. Conclusion

This paper focuses on automatic classification of X-ray image from the ImageCLEF 2009 dataset based on anatomical and biological information used pretrained model InceptionV3. The X-ray images in the ImageCLEF 2009 dataset are prepared and preprocessed for training InceptionV3. In terms of classification loss, constant padding with no enhancements during training had the best performance. In terms of classification accuracy, constant padding with enhancement had the best performance. overall inputs with constant padding have the best performance.

Acknowledgment

Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: http://hpc.uct.ac.za

References

[1] Sultana, Sufian, Dutta, "Advancements in Image Classification using Convolutional Neural Network", 2019.
[2] Zhao, Zheng, Xu, Wu, "Object Detection with Deep Learning: A Review", 2019.
[3] Liu, Deng, Yang, "Recent progress in semantic image segmentation", 2018.
[4] Ganesan, Subashini, "Classification of medical X-ray images for automated annotation", 2014.
[5] Zare, Mueen, Awedh, Seng, "Automatic classification of medical X-ray images:hybrid generative-discriminative approach", 2013.
[6] Aboud, Spanier, Joskowicz, "Automatic Classification of Body Parts X-ray Images", 2015.
[7] Pelka, Nensa, Friedrich, "Annotation of enhanced radiographs for medical image retrieval with deep convolutional neural networks", 2018.
[8] Lehmann, Schubert, Keysers, Kohnen, Wein, "The IRMA code for unique classification of medical image. Proceedings of SPIE - The International Society for Optical Engineering", 2003.
[9] Tommasi, Caputo, Welter, Guld, Desern, "Overview of the CLEF 2009 Medical Image Annotation Track", 2009.
[10] Tommasi, Deselaers, "The Medical Image Classification Task", 2010.
[11] Szegedy, Vanhoucke, Ioffe, Shlens, Wojna, "Rethinking the Inception Architecture for Computer Vision", 2015.