Gender Detection from Spine X-ray Images Using Deep Learning

Zhiyun Xue, Sivaramakrishnan Rajaraman, Rodney Long, Sameer Antani, and George R. Thoma

Lister Hill National Center for Biomedical Communications
National Library of Medicine
Bethesda, USA
{xuez, sivaramakrishnan.rajaraman, rlong, santani, gthoma}@mail.nih.gov

***Abstract*—The algorithm described in this paper aims to classify the spine x-ray images according to image characteristics that exhibit gender. We developed a customized sequential CNN model which is trained from scratch using the spine images first and tested it on the NHANES II dataset hosted by the U.S. National Library of Medicine (NLM). Aiming to improve the performance, we then developed a method for extracting the region-of-interest (ROI) in the cervical spine images using a content-based image retrieval (CBIR) method and compared the results of using the original images vs. the ROI images. Later, we applied/tested the method of fine-tuning a DenseNet model that was pre-trained with the ImageNet dataset with the spine images, and this approach gets the best result, achieving classification accuracy of 99% for cervical spine image set and 98% for the lumbar spine image set.**

Keywords-spine x-ray; deep learning; gender classification; convolutional neural network; content-based image retrieval

I. INTRODUCTION

In recent years, deep learning has become a popular and effective technique in the field of computer vision (especially for natural image classification) due to the availability of large annotated datasets, affordability of GPU cards, and collaboration within the research community for providing open access to source codes and trained models. In this paper, we attempt to detect the gender of the imaged person based on their characteristics imaged in the spine x-rays. This was motivated by the observation that gender information is sometimes missing due to lack of proper acquisition procedures or aggressive de-identification processes. According to [1, 2], there are anatomical gender differences in cervical/lumbar spines (such as vertebral geometry and spinal curvature). There are also sex differences in characteristics associated with spinal diseases (such as clinical and radiological manifestations) [3, 4]. As shown in Figure 1 and Figure 2, in addition to the spinal area that may exhibit gender differences, other regions such as the head-neck segment (in the cervical spine image) or the breast region (in the lumbar spine image) may also present characteristics that are gender identifiable. It is very challenging to hand-engineer features that are effective for capturing gender-differentiating morphological characteristics. Therefore, we examine the use of the deep convolutional neural network (CNN) which learns feature representation automatically from raw image data.

There has been some work in the research literature on classifying gender from images, with the majority being applications to visible light or infrared face images [5, 6] or NIR Iris images [7, 8]. Very few works use medical images to identify gender. Such efforts that have been made include [9], based on hand x-rays, and [10], based on x-ray images of femur bone. Both of [9, 10] use conventional methods that are based on hand-crafted features. Our group has applied deep learning, specifically, transfer learning (using CNN-based feature extractor plus conventional classifier), to identify gender from the patient's frontal chest x-ray [11]. In [11], we compared six CNN feature extractors (AlexNet [12], VggNet-16 [13], VggNet-19 [13], GoogleNet [14], ResNet-50 [15], and ResNet-152 [15], all pre-trained using ImageNet [16] data) plus two conventional classifiers (SVM and Random Forest).

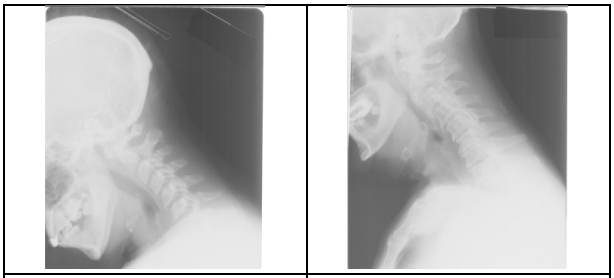

Figure 1. Examples of cervical spine images

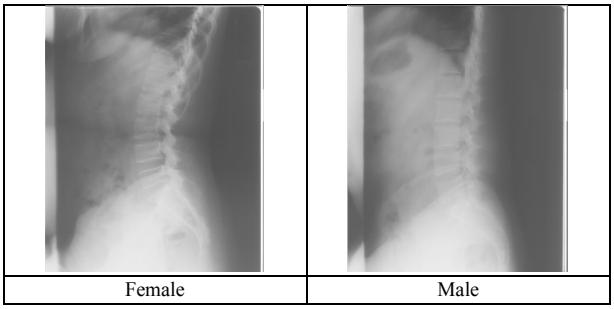

Figure 2. Examples of lumbar spine images

2372-9198/18/$31.00 ©2018 IEEE
DOI 10.1109/CBMS.2018.00017

In this paper, we work on the spine images obtained by the second National Health and Nutrition Examination Surveys (NHANES II dataset) in which the gender information was provided with each image. For developing a CNN-based classifier to identify gender information from spine x-ray images, we evaluated a customized sequential CNN model which is trained from scratch using the spine images at first. In order to improve the performance, we then developed a method for extracting the region-of-interest (ROI) in the cervical spine images using a content-based image retrieval method (CBIR, similarity-searching). This method aims to alleviate the negative effect on the classifier of the irrelevant information in the image border regions. We compared the results from using the cropped ROI images as input with the results from using the original, uncropped images. Later, we test the method of fine-tuning a DenseNet [17] model that was pre-trained with the ImageNet dataset with our spine images and this approach obtains the best result.

The rest of the paper is organized as follows. In Section II, we present the image data. In Section III, we describe the sequential CNN model, the method of ROI image generation, and the corresponding experimental results and comparison. In Section IV, we present the method of DenseNet fine-tuning and the corresponding results analysis and discussions. Section V concludes the paper and describes future work.

II. IMAGE DATA

The U.S. National Library of Medicine (NLM) has a collection of 17,100 spine (cervical and lumbar) x-ray images. The images were collected during the second National Health and Nutrition Examination Surveys (NHANES II), carried out by the National Center for Health Statistics, Centers for Disease Control (NCHS/CDC) during the years 1976-1980. Besides images, related text data such as demographic information, anthropometric data, and health and medical history were also collected. Detailed information on this dataset can be found on this NLM web site [18]. This large dataset is a valuable resource which has been previously used by researchers in various studies such as bone morphometry education, vertebra segmentation, and content-based image retrieval [19, 20]. The data in NHANES II contains 9667 cervical spine images and 7428 lumbar spine images. The cervical spine images have size 1462 x 1755, and the lumbar spine images, 2048 x 2487. Among cervical spine images, 5053 are female and 4614 are male. For lumbar images, there are 2833 female images and 4595 male images. The images were acquired as a part of a public health survey and have been de-identified.

III. APPROACH I: CUSTOMIZED SEQUENTIAL CNN MODEL

A. Customized Sequential CNN Model

The architecture of the CNN model tested in this study is shown as in Figure 3. Since each layer connects only to the previous and following layers, we refer to this conventional CNN as a "sequential CNN model". This model is trained from scratch using the spine images. The proposed model encompasses five convolutional and three fully-connected layers. Input images of dimension 227 × 227 × 3 are fed into the input layer. There are 96 filters in the first convolutional layer, each of dimension 7×7 and stride 2. A Rectified Linear unit (ReLU) activation follows each convolutional layer to enhance learning [21]. All the other convolutional layers have filters of dimension 3×3. Weights are initialized from a zero-mean Gaussian distribution. A local response normalization (LRN) layer is included after the first and second convolutional layers to aid in generalization, motivated by the lateral inhibition process of biological neural networks [12]. Max-pooling layers with a pooling window of 3×3 and stride 2 follow the LRN layers and the fifth convolutional layer. There are three fully-connected layers, the first two fully connected layers having 4096 neurons each; the third fully connected neurons has two neurons which feed into the Softmax classifier. Dropout regularization is achieved by dropping 50% of the neurons in the first and second fully-connected layers during the process of training, to alleviate over-fitting issues. The proposed model is trained by optimizing the multinomial logistic regression objective using stochastic gradient descent (SGD) with momentum. The model is optimized for its hyper-parameters by a randomized grid search method [22]. L2-regularization is used with a weight penalty of 5×10-4. The learning rate is initialized to 0.001 and is reduced three times before convergence. The mini-batch size is 10 and the training is stopped after 60 epochs. The proposed model achieves faster convergence due to implicit regularization imposed by smaller convolutional filter dimensions, greater depth, usage of L2-regularization parameter, and dropouts in the fully-connected layers.

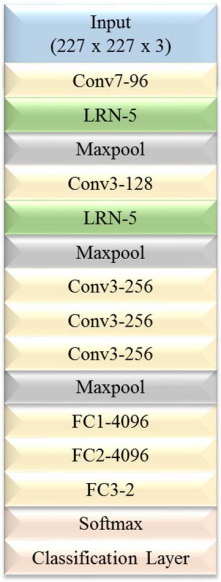

Figure 3. Architecture of our CNN (conventional sequential model)

B. ROI Image Cropping

As shown in Figure 4, there are images that contain unwanted border regions (Figure 4a and 4b) or notably larger regions of background or shoulder (Figure 4c and 4d). To analyze the effect of these unwanted regions on the classification accuracy of CNN models, we propose a CBIR-based method to

remove them (at least partially) and focus more on the ROI. In our previous work on cervical vertebra segmentation [19], we collected 136 images in which the rectangular regions containing the spine were marked. Several examples of these spinal-ROI marked images are shown in Figure 5. We use these 136 images as the retrieval database. For each cervical spine image, we first compare it with all the images in the retrieval database to find the most similar one. The feature we use to represent the image content is the PHOG (Pyramid Histogram of Oriented Gradients) feature [23]. The similarity measure is Euclidean distance. We then use the coordinates of the center point of the marked spinal ROI in the image returned from the database to extract a square ROI image of size 1400 x 1400 from the query image; this square ROI is centered on the point obtained from the database image. We chose not to use the ROI from the database image "as-is" for two reasons: 1) Since gender differences are not only exhibited on the spine but also on nearby structures (such as the base of the skull), our ROI needs to include these important areas as well as the spine; 2) As shown in Figure 5, the size of the marked spinal ROI is not the same across the 136-image dataset. If the variable size spinal ROI images are used as input to the CNN models, since each image needs to be resized to a constant value, the anatomical regions in the image will be resized disproportionally across the dataset. Since the relative size of certain anatomical regions (such as vertebra width and disc-facet depth) may indicate gender difference [2], the size of the cropped images should be consistent across the dataset. Figure 6 shows examples of images with ROI extraction results.

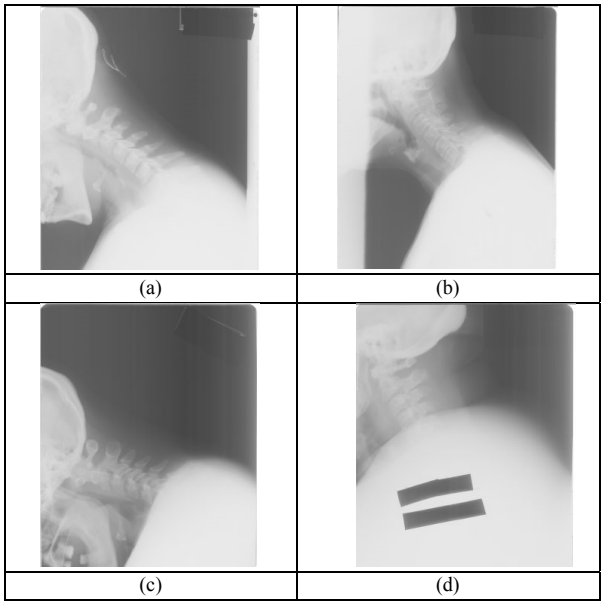

Figure 4. Examples of cervical spine images

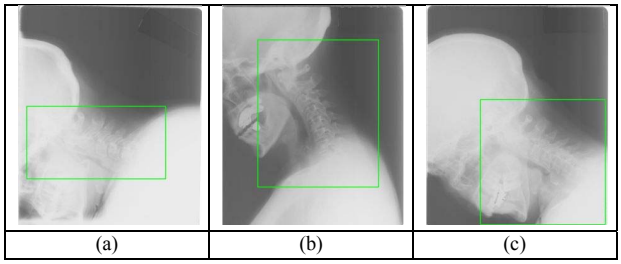

Figure 5. Examples of collected images with spinal ROI marked

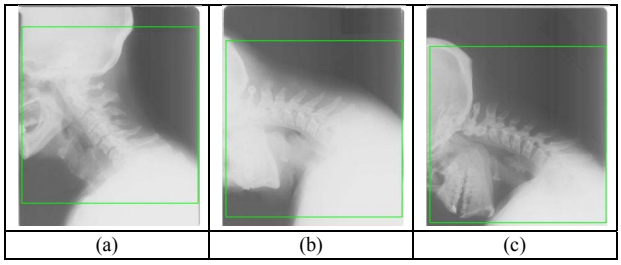

Figure 6. Examples of ROI extraction results

C. *Experimental Tests*

We evaluate the performance of this customized sequential CNN model on the cervical spine image set. For the experiments, we randomly split the images into training set, validation set, and test set (in approximate proportions of 80%/10%/10%). The number of cervical spine images in each set is shown in Table 1. The confusion matrix and classification accuracy for the customized sequential CNN model on the original cervical spine images are shown in Table 2. It obtains accuracy of 88.9%. We then use the cropped ROI images as the input to the model to see if it improves the performance compared to using the original cervical spine images, however, as shown in Table 3, the result of accuracy is almost the same. This suggests that the network is able to learn to ignore irrelevant, noisy information in the images and focus on the key features.

Table 1. Cervical Spine Images

	Training	**Validation**	**Test**	**Total**
Female	4000	500	553	5053
Male	3680	460	474	4614

Table 2. Results For Cervical Images (the customized sequential CNN Model on the original images)

Original Images			
Confusion matrix			*Accuracy*
Predict ->	Female	Male	0. 889
F	506	47	
Male	65	409	

Table 3. Results For Cervical Images (the customized sequential CNN Model on the original images)

Cropped ROI Images			
Confusion matrix			*Accuracy*
Predict ->	Female	Male	0. 890
Female	486	67	
Male	47	427	

IV. Approach II: Fine-tuning DenseNet

Later, we apply/test the method of fine-tuning the ImageNet-pre-trained DenseNet with the spine dataset to see if it gets better classification result than that of the customized sequential CNN model.

A. DenseNet Fine-Tuning

Compared to conventional CNNs in which layers only connect to adjacent layers, the DenseNet includes densely-connected blocks named *dense blocks*. Within a dense block, each layer is connected to all the preceding layers and all the subsequent layers. (To contrast the DenseNet architecture with conventional CNNs, we refer to it as an example of a "non-sequential CNN model".) Therefore, instead of using only the feature maps output from the closest preceding layer as the inputs, each layer in the dense block uses the feature maps from all the preceding layers as the inputs and its own feature maps are used as inputs to all the subsequent layers in the block [17]. Typically, each dense block is followed by transition layers which contain a pooling layer for reducing the feature map size. Figure 7 shows an example of a DenseNet with four dense blocks. The integration of the dense connections in the architecture enables DenseNet to promote feature reuse and feature propagation. As a result, it achieves high performance while being computationally efficient [17, 24]. Implementations of the DenseNet and pre-trained models have been made available by the authors at [25]. For our application, we use the DenseNet-121 model that is pre-trained with ImageNet data. It consists of 4 dense blocks, 4 transition layers and 1 classification layer. The 4 dense blocks contain 6, 12, 24, and 16 pairs of convolutional layers respectively. Each transition layer consists of a convolutional layer and a pooling layer. The classification layer consists of a pooling layer, a fully connected layer and a Softmax layer. The size of the images is 224 × 224 × 3. For the detailed DenseNet-121 architecture and parameters, please refer to [17]. To fine-tune the pre-trained DenseNet-121 model using our spine image data, we replace the original 1000-D fully-connected layer (for ImageNet classes) with a 2-D fully-connected layer. The learning rate is set as 0.001 initially and decays over time. The momentum is set as 0.9. The mini-batch size is 4 and the number of epoch is set to be 60. The optimization method is SGD and the cross-entropy loss function is used. Both batch normalization and dropout are applied.

B. Experimental Tests

We evaluate the performance of this approach on the same training/validation/test cervical image set used by the approach I. Figure 8 shows the training and validation loss of the model over the number of epochs and Figure 9 shows the training and validation accuracy of the model over the number of epochs. Table 4 lists the confusion matrix and classification accuracy for the test set. It obtains accuracy of 0.99 for the test set which is much higher than that of the first approach, which demonstrates the advantages of the network architecture of DenseNet (with significant feature reuse and propagation) and the transfer learning (pre-trained with large image set) over our customized sequential CNN model trained from scratch using a relatively small dataset. We apply the method of fine-tuning DenseNet to the lumbar spine original images as well. The number of lumbar spine images randomly split into three sets (in approximate proportions of 80%/10%/10% for training/validation/test) is shown in Table 5. The corresponding confusion matrix and accuracy for the test set is shown in Table 6. For lumbar spine images, the test set accuracy is 0.98.

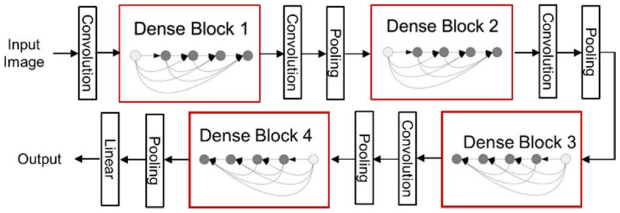

Figure 7. Diagram of a DenseNet with four dense blocks. DenseNet is a non-sequential model.

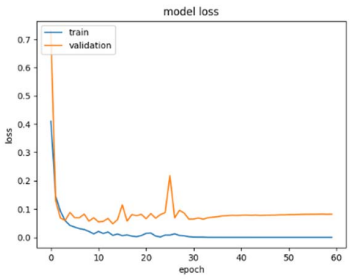

Figure 8. The model loss over the number of epochs

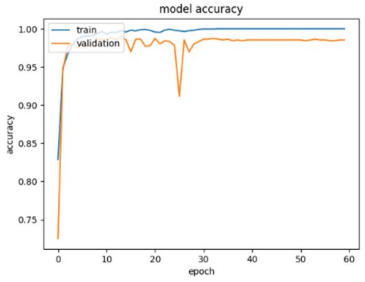

Figure 9. The model accuracy over the number of epochs

Table 4. Results For Cervical Images (DenseNet/Original Images)

Confusion matrix			Accuracy
Predict ->	Female	Male	0.99
Female	546	7	
Male	2	472	

Table 5. Lumbar Spine Images

	Training	Validation	Test	Total
Female	2260	280	293	2833
Male	3670	450	475	4595

Table 6. Results For Lumbar Images (DenseNet/Original Images)

Confusion Matrix			Accuracy
Predict ->	Female	Male	0.98
Female	283	10	
Male	7	468	

V. CONCLUSION

In this paper, we process spinal x-ray images to detect the gender of the imaged person. We first developed a customized sequential CNN model which was trained from scratch using the spine images in the NHANES II. We then examined if using the ROI images as the input to the CNN model yields better performance compared to using the original cervical spine images. However, it obtained almost the same results as using the original images, which indicates that the CNN model is quite effective in learning effective features while ignoring irrelevant information. We then applied the method of fine-tuning the ImageNet-pretrained DenseNet. The fine-tuned DenseNet achieves much better performance, obtaining 99% accuracy for the cervical set and 98% accuracy for the lumbar set. In the future, we will work on applying deep learning techniques for processing other tasks on the NHANES II dataset such as vertebra segmentation, and comparing with the traditional method we developed.

ACKNOWLEDGMENT

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC).

REFERENCES

[1] B.D. Stemper, N. Yoganandan, F.A. Pintar, D. J. Maiman, M.A. Meyer, J. DeRosia, B.S. Shender, G. Paskoff, "Anatomical gender differences in cervical vertebrae of size-matched volunteers," Spine, vol.33, no. 2, 2008, pp. E44-49.

[2] O. Hay, G. Dar, J. Abbas, D. Stein, H. May, Y. Masharawi, N. Peled, I. Hershkovitz, "The lumbar lordosis in males and females, revisited," PLoS One, vol. 10, no. 8, 2015, doi: 10.1371/journal.pone.0133685

[3] A.A.J. Miller, C. Schmatz, A.b. Schultz, "Lumbar disc degeneration: correlation with age, sex, and spine level in 600 autopsy specimens," Spine, February 1988, pp. 173-178.

[4] M. Landi, H. Maldonado-Ficco, R. Perez-Alamino, J. A. Maldonado-Cocco, G. Citera, etc. "Gender differences among patients with primary ankylosing spondylitis and spondylitis associated with psoriasis and inflammatory bowel disease in an iberoamerican spondyloarthritis cohort," Medicine, vol. 95, no. 95, 2016, doi: 10.1097/MD.0000000000005652

[5] A. Dehghan, E.G. Ortiz, G. Shu, S. Zain Masood, "DAGER: deep age, gender and emotion recognition using convolutional neural network," arXiv eprint arXiv:1702.04280, 2017

[6] E. Mohammady Ardehaly, A. Culotta, "Co-training for demographic classification using deep learning from label proportions," arXiv eprint arXiv:1709.04108, 2017

[7] J. Tapia, C. Aravena, "Gender classification from NIR iris images using deep learning", Deep Learning for Biometrics, 2017, pp. 219-239.

[8] D. Bobeldyk, A. Ross, "Iris or periocular? Exploring sex prediction from near infrared ocular images," Proceedings of the 15th International Conference of the Biometrics Special Interest Group (BIOSIG), 2016, doi: 10.1109/BIOSIG.2016.7736928.

[9] Y. Kabbara, A. M. Shahin, A. Naït-ali, M.A. Khalil, "Gender Classification by X-Ray Images of Hand," The 20th LAAS International Science Conference Advanced Research for Better Tomorrow, 2014.

[10] C. S. Raghavendra, D.C. Shubhangi, G. Karnatka, "Gender identification in digital x-ray images of femur bone," International Journal Of Technological Exploration And Learning, 2014.

[11] Z. Xue, S.K. Antani, L. R. Long, G. R. Thoma, "Using deep learning for detecting gender in adult chest radiographs," Proceedings of SPIE Medical Imaging, vol. 10579, 2018.

[12] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classification with deep convolutional neural networks", Conf. of Advances in Neural Information Processing Systems, 2012, pp. 1106-1114.

[13] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv eprint arXiv:1409.1556v6, 2014

[14] C. Szegedy et al., "Going deeper with convolutions," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1-9.

[15] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778

[16] http://www.image-net.org/challenges/LSVRC/, accessed at 02/25/2018.

[17] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, "Densely connected convolutional networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2261-2269.

[18] https://ceb.nlm.nih.gov/repositories/nhanes/, accessed at 02/25/2018.

[19] A. Gururajan, S. Kamalakannan, H. Sari-Sarraf, M. Shahriar, R. Long, S. Antani, "On the creation of a segmentation library for digitized cervical and lumbar spine radiographs," Computerized Medical Imaging and Graphics, vol. 35, 2011, pp. 251-265.

[20] S. Antani, L. R. Long, G.R. Thoma, "Content-based image retrieval for large biomedical image archives", Studies in Health Technology and Informatics, vol. 107, 2004, pp.829-833.

[21] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, "Deep networks with stochastic depth," Computer Vision – ECCV, 2016, pp 646-661.

[22] J. Bergstra and Y. Bengio, "Random search for hyperparameter optimization," Journal of Machine Learning Research, vol. 3, 2012, pp. 281–305

[23] A. Bosch, A. Zisserman, X. Munoz, "Representing shape with a spatial pyramid kernel," Proceedings of the International Conference on Image and Video Retrieval, 2007, pp. 401-408.

[24] G. Pleiss, D. Chen, G. Huang, T. Li, L. van der Maaten, K.Q. Weinberger, "Memory-efficient implementation of densenets," arXiv preprint arXiv:1707.06990, 2017

[25] https://github.com/liuzhuang13/DenseNet, accessed at 02/25/2018.