Modality Detection and Classification of Biomedical Images with Deep Transfer Learning and Feature Extraction

Jose Dixon
Computer Science Department
Morgan State University
Baltimore, Maryland, USA
jodix5@morgan.edu

Mahmudur Rahman
Computer Science Department
Morgan State University
Baltimore, Maryland, USA
md.rahman@morgan.edu

***Abstract*—Classification of medical images of diverse modalities is an important first step for many applications in biomedical domain, such as computer aided diagnosis (CAD) and retrieval. However, the descriptiveness and discriminative power of features extracted from medical images are critical to achieve good classification and retrieval performances. Recently, Deep learning algorithms, such as convolutional neural networks (CNN) seem to be effective in supplying better accuracy. CNN trained on large-scale data sets such as ImageNet have demonstrated to be excellent at the task of transfer learning. These networks learn a set of rich, discriminating features to recognize 1,000 separate object classes. CNNs not only give state-of-the-art results when trained for a specific task, but experiments have shown that the filters learned over the ImageNet dataset are generic and useful for other image tasks that the CNN was not originally trained for. Using a pre-trained CNN as a feature extractor also provides an alternative to the hand-crafted features based on learning a transformation of raw data input to a representation that can be effectively exploited in general machine learning tasks, such as Random Forest or Support Vector Machine (SVM) Classifiers. This work is focusing on extracting such features from medical images by applying deep transfer learning with pre-trained VGG16 network as feature extractors, and also exploring data augmentation to address the limitation of small training set. The experimental results in a medical data set of 5000 images of 30 different categories based on modalities, body parts, orientations, and specific visual features showed that this approach obtained the best accuracy (97%).**

Keywords: Classification, Modality Detection, Convolutional Neural Networks, Deep learning, Feature extraction, Deep feature extraction.

I. INTRODUCTION

The Deep learning is the subfield of machine learning that is devoted to building algorithms that explain and learn a high and low level of abstractions of data that traditional machine learning algorithms often cannot. Convolutional neural networks (also called ConvNet) leverage spatial information and are therefore very well suited for classifying images [1]. These nets use an ad hoc architecture inspired by biological data taken from physiological experiments done on the visual cortex. In the last 20 years, computer vision research has focused on manually defined pipelines for extracting good image features. For a while, image feature extractors such as LBP and HOG were the standard. Recent developments in deep learning research have extended the reach of traditional machine learning models by incorporating automatic feature extraction in the base layers. They replace manually defined feature image extractors with manually defined models that automatically learn and extract features.

II. TRANSFER LEARING METHODS

Transfer learning reuses knowledge from past related tasks to ease the process of learning to perform a new task (Donges, 2018). The goal of transfer learning is to leverage previous learning and experience to more efficiently learn novel, but related, concepts, compared to what would be possible without this prior experience. The utility of transfer learning is typically measured by a reduction in the number of training examples required to achieve a target performance on a sequence of related learning problems, compared to the number required for unrelated problems [2]

Convolutional neural networks (CNNs) trained on large-scale datasets such as ImageNet have demonstrated to be excellent at the task of transfer learning. These networks learn a set of rich, discriminating features to recognize 1,000 separate object classes. CNNs not only give state-of-the-art results when trained for a specific task, but experiments have shown that the filters learned over the ImageNet dataset are generic and useful for other image tasks that the CNN was not originally trained for [3,4]

Data augmentation is about accumulating more data than what is already available [5]. The process of performing these transformations on existing training images to generate new images is called data augmentation. Another advantage of

using data augmentation is that you are able to increase the size of your training dataset (when used with data generators, we can get infinite images).

Each type CNN utilizes its own layers, kernels, training, and feature extraction techniques. Deep learning architectures can be composed of several types of layers (Casari & Zheng, 2018). Fully connected layers are so named because every input can be used in every output. In contrast to fully connected layers, a convolutional layer uses only a subset of inputs for each output. A pooling layer combines multiple inputs into a single output.

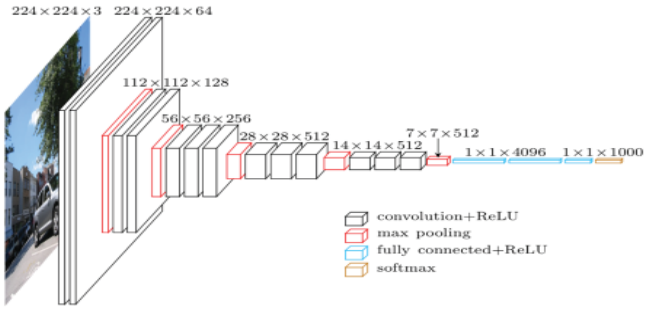

Figure 1: A representation of VGG16 architecture

VGGNet is a network based 3x3 convolutional layers stacked and alternated with max pooling, two 4096-connected layers stacked and alternated with max pooling, two 4096 fully-connected layers followed by a softmax classifier [4]. Prior to VGG the initial convolutional layers of a network used filters with large receptive fields, such as 7 x 7. Additionally, the networks usually had alternating single convolutional and pooling layers. Figure 1 shows how VGGNet architecture looks like of a convolutional layer in VGGNet with a large filter size can be replaced with a stack of two or more convolutional layers with smaller filters (factorized convolution). VGGNet architecture with very small (3 × 3) filters can be trained to increasingly higher depths (16-19 layers) and obtain state-of-the-art classification [7].

III. Deep Feature Extractor

Using a pretrained CNN as a feature extractor rather than training a CNN from scratch is attractive as it transfers learning (i.e. filters) from other domains where more training data is available, and avoids a time consuming training process.

In this work, we loaded all the images from the training and test sets, extract their features using a pre-trained 16-layer VGG mode [4]l, and store the extracted features keyed on the image id to a new file in HDF5 dataset format that is later loaded and used as input for training with few general machine learning classifiers, such as Logistic Regression, Random Forest, SVM, etc. Here, we at first pre-processed the images (e.g. 3 channel 224 x 224 pixel image) for the VGG model (without the output layer) trained on ImageNet dataset [3] and used the extracted features predicted by this model as input.

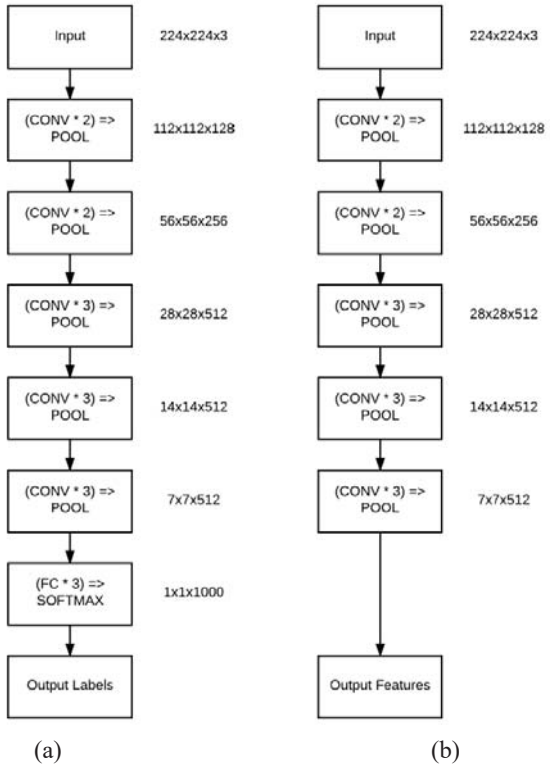

Figure 2 (a) The original VGG16 network architecture (b) VGG16 network as feature extractor [7]

When treating the VGG16 networks (Fig. 2) as a feature extractor, we essentially "chop off" the network prior to the fully-connected layers. The last layer of the network is a max pooling layer (Figure 2(a)), which will have the output shape of 7 x 7 x 512 implying there are 512 filters each of size 7 x 7. If we were to forward propagate an image through this network with its FC head removed, we would be left with 512, 7x7 activations that have either activated or not based on the image contents. Therefore, we can actually take these 7 x 7 x 512 = 25,088 values and treat them as a feature vector that quantifies the contents of an image [7].

After repeating this process for the entire dataset of images (including datasets that VGG16 was not trained on), we are left with a design matrix of N images, each with 25,088 columns used to quantify their contents (i.e., feature vectors). Given our feature vectors, we can train an off-the-shelf machine learning model such SVM, Logistic Regression classifier, Random Forest, etc. on top of these features to

obtain a classifier that recognizes new classes of images. Hence, the VGG16 network is used here as an intermediary feature extractor. The downstream machine learning classifier will take care of learning the underlying patterns of the features extracted from the CNN.

IV. EXPERIMENTS

For experiments, we used a subset (5000 images) of the ImageCLEFMed dataset [9] with 30 different medical modalities (Figure 3). 50% of the 5,000 images are used for testing and training vice versa. The dataset comprises of different classes of medical modalities of compound or multi pane images, diagnostic images, and generic biomedical illustrations. The diagnostic images consists of radiology, visible light photography, microscopy, 3D reconstructions, and printed signals or waves. The generic biomedical illustrations are not related to medical imaging.

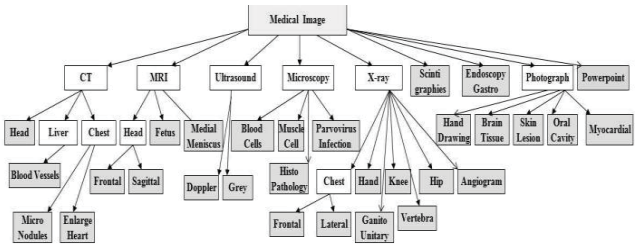

Figure 3: Different classes and categories used for ImageCLEFMed dataset.

The medical modalities are organized into different classes and subclasses that are relevant to medical imaging hierarchy map as shown in Figure 3 [6].

V. RESULTS AND DISCUSSION

Figure 4 and 5 show VGGNet based classification results of different categories and histogram graph of accuracy results based on applying data augmentation technique. It is observed that data augmentation feature can help significantly to improve the feature extraction processing for a convolutional neural network such as VGGNet.

	precision	recall	f1-score	support
Angiogram_Budd_Chiari	0.98	0.81	0.89	58
Brain_Limb_Alzheimer	0.97	0.88	0.92	34
CT_Chest	1.00	1.00	1.00	36
CT_Head	0.89	0.89	0.89	27
CT_Liver_BloodVessels	0.90	0.94	0.92	48
CT_Mediastinal	0.94	1.00	0.97	32
Endoscopy_Gastro	1.00	1.00	1.00	62
Handdrawn_Illustration	0.81	0.97	0.88	35
MRI_Fetus	1.00	0.69	0.82	13
MRI_Frontal_Head	0.98	0.87	0.92	53
MRI_Medial_Meniscus	0.88	1.00	0.94	60
MRI_Sagittal_Head	0.86	1.00	0.92	36
Microscopic_Bacterial_Meningitis	0.48	1.00	0.65	32
Microscopic_BloodSmears	0.94	0.87	0.90	97
Microscopic_HistoPathology	0.75	0.63	0.69	38
Microscopic_MuscleCell	0.83	0.91	0.87	32
Microscopic_Parvovirus_Infection	0.88	0.66	0.76	68
Patho_Gross_Miocardial_Infarction	0.91	0.84	0.87	25
Patho_Oral_Cavity	0.90	1.00	0.95	28
Powerpoint	0.91	0.91	0.91	22
Scintigraphies	0.95	0.69	0.80	29
Skin_Lesion	1.00	0.91	0.96	47
US_Doplar	0.89	0.83	0.86	30
US_Grey	0.94	0.96	0.95	53
Xray_Chest_Frontal	0.99	0.97	0.98	71
Xray_Chest_Lateral	0.93	1.00	0.96	26
Xray_GenitoUnitary	0.94	0.98	0.96	47
Xray_Hand	1.00	0.96	0.98	25
Xray_Head	0.92	1.00	0.96	11
Xray_Hip	0.97	0.92	0.95	38
Xray_Knee	0.89	0.96	0.93	26
Xray_Vertebral_Osteophytes	0.95	0.95	0.95	22
avg / total	0.92	0.90	0.91	1261

Figure 4: Results of VGGNet (with Data Augmentation) Classification

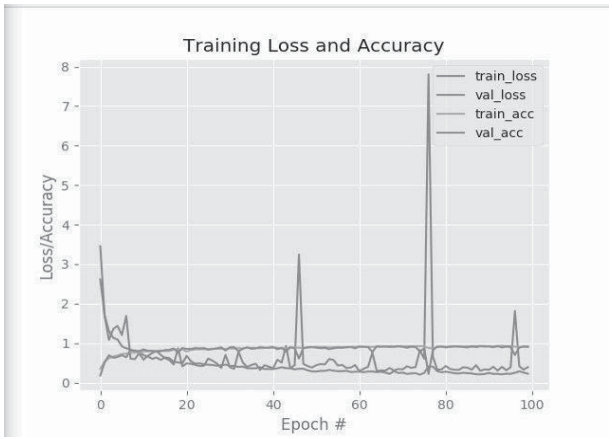

Figure 5: Histogram graph of Accuracy Results from the VGGNet (with Data Augmentation) Classification

Table 1: Classification Results (Test set)

Classifier	Precision	Recall	F1-Score
VGGNet w/oDA	0.63	0.57	0.64
VGGNet w/DA	0.92	0.90	0.91
VGGNet (Deep Feature)	0.97	0.97	0.97

From Table 1, it is observed that without the data augmentation feature, VGGNet has the worst results for the deep learning classification in the experiment. With data augmentation, it achieved around 92% accuracy. However, we obtained the best accuracy (97%) after extracting the deep

features based on using VGG16 model as transfer learning and performing a logistic linear regression based classification on this feature space. Hence, it proves to be effective to extract deep features at first and perform ordinary classification, such as linear regression, support vector machine (SVM) at a later stage.

VI. CONCLUSIONS

Modality detection is an essential task to enhance the performance of a medical image retrieval system. This paper present modality detection/classification of medical images based on applying deep transfer learning (VGGNet) with different approaches. It is found that the best classification accuracy is obtained when the deep feature is extracted by applying a transfer leaning approach and later general machine learning based classification (linear regression) is performed in that deep feature space instead of fine tuning the classifier. In future, we will try to explore other deep learning models and use ensemble of classifiers as classifier combination approach.

ACKNOWLEDGMENT

This research was supported by an NSF HBCU-UP Research Initiation Award (RIA) grant (Award Id: 1601044).

REFERENCES

[1] Mishra, C., & Gupta, D. L. (2017). Deep machine learning and neural networks: An overview. IAES International Journal of Artificial Intelligence, 6(2), 66-73. Retrieved from https://search-proquest-com.proxy-ms.researchport.umd.edu/docview/1924958782?accountid=12557
"Modern Machine Learning Algorithms: Strengths and Weaknesses." (2018, May 20). https://elitedatascience.com/machine-learning-algorithms (accessed 6 August 2018)

[2] Donges, N., & Donges, N. (2018, April 23). Transfer Learning. Retrieved from https://towardsdatascience.com/transfer-learning-946518f95666

[3] O. Russakovsky et al., "ImageNet large scale visual recognition challenge," IJCV, 2015.

[4] Karen Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large- Scale Image Recognition". In: CoRR abs/1409.1556 (2014). URL: http://arxiv.org/abs/1409.1556 (cited on

[5] Guillou, P., & Guillou, P. (2019, January 23). Data Augmentation by fastai v1. Retrieved May 8, 2019, from https://medium.com/@pierre_guillou/data-augmentation-by-fastai-v1-84ca04bea302

[6] Rahman, M., Antani, S. K., & Thoma, G. R. (2010). A classification-driven similarity matching framework for retrieval of biomedical images. 2010 Proceedings of the International Conference on Multimedia Information Retrieval, 147-154.

[7] Rosebrock, A. Deep Learning for Computer Vision Practitioner Bundle https://www.pyimagesearch.com/deep-learning-computer-visionpython- book/ , pg. 31-47 (accessed 26 August 2018)

[8] Zocca, V., Roelants, P., Spacagna, G., Slater, D., & Vasilev, I. (n.d.). Python Deep Learning - Second Edition. Retrieved January, 2019, from https://learning.oreilly.com/library/view/python-deep-learning/9781789348460/96fa5814-e23e-4b62-8103-e1c097f3eaa7.xhtml

[9] H. M¨uller, J. Kalpathy-Cramer, I. Eggel, S. Bedrick, Jr. J. Reisetter, C. E. K., and W. R. Hersh, W. R. "Overview of the CLEF 2010 Medical Image Retrieval Track", CLEF 2010 Evaluation Labs and Workshop, On-line Working Notes. Padua, Italy, Sep. 2010.