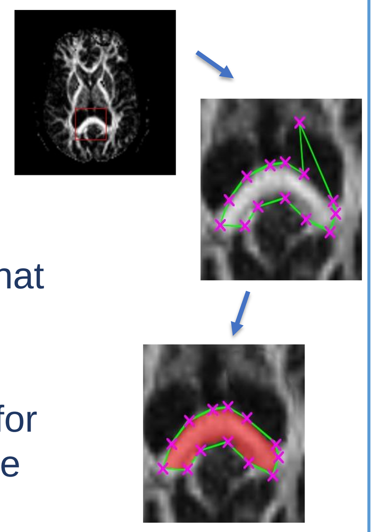
Biomedical Image Region of Interests Detection Based Transfer Learning and Deep Feature Extraction and Classification

Jose Dixon, Mahmudur Rahman, Ph.D (Mentor)

Computer Science Department, Morgan State University, Baltimore, Maryland

Introduction

- Regions of interest (ROIs) that are pointed to by overlaid markers (arrows, asterisks, etc.) in biomedical images are expected to contain more important and relevant information than other regions for biomedical article indexing and retrieval.
- Region of interests are an image region of samples within a data set that are identified by a person to focus on for a particular purpose.
- Techniques used for ROI detection are associated with learning-based algorithms. Learning-based algorithms utilize the pattern of the region for detection on the basis of a large quantity of image data pertaining to the learning target.



Process of ROI drawing

Objectives

- In deep learning (DL), Convolutional Neural Networks have emerged as a particularly powerful tool by both providing outstanding performances in conventional tasks and allowing a wide variety of unprecedented applications in computer vision.
- CNNs are designed to recognize visual patterns directly from pixel images with minimal preprocessing. Using a pre-trained CNN as a feature extractor also provides an alternative to the handcrafted features that are not manually engineered from raw pixel data in general machine learning classifiers.

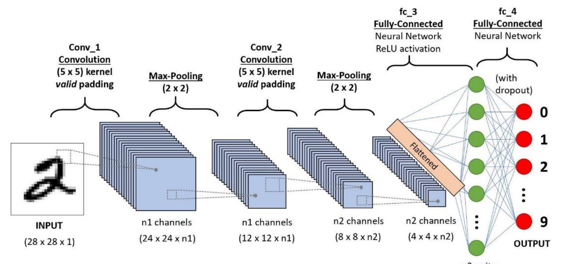
Convolutional Neural Networks - Background and Building Blocks

Convolutional Neural Network (CNN) typically has five convolutional layers and then optionally preceded by some fully connected layers in a standard multilayer neural network. The processing is possible with local connections and tied weights followed by some form of pooling which results in translation invariant features. CNNs thare easier to train by having fewer parameters than fully connected networks with a similar number of hidden units.

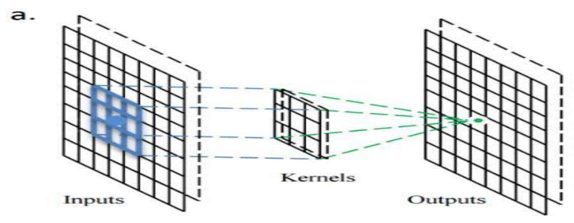
- **Input Layer** - hold the raw pixel values of the image.
- **Convoultional Layer** - compute the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume.
- **Non-Linerarly Layer** - apply an elementwise activation function, such as the max(0,x) thresholding at zero. This leaves the size of the volume unchanged
- **Pooling Layer** - will perform a downsampling operation along the spatial dimensions (width, height), resulting in volume such as width, height, and depth
- **Fully Connected Layer** - will compute the class scores, resulting in volume of the depth of an image correspond to a class score, for categories. As with ordinary Neural Networks and as the name implies, each neuron in this layer will be connected to all the numbers in the previous volume.

The essential parts of a convolutional neural networks are convolutions, matrices or kernels.

- A **convolution** is an element-wise multiplication of two matrices followed by a sum.
- A **matrix** is from an image that is multi dimensional. Unlike as CNN are arranged in a 3D volume in three dimensions: width, height, and depth. Our image has a width (# of columns) and height (# of rows), just like a matrix. An image as big matrix and a kernel or convolutional matrix as a tiny matrix that is used for blurring, sharpening, edge detection, and other processing functions.
- The **kernel** apply shift-based multiplication by moving in the input, from left to right and from top to bottom, and each one of the values on the kernel is multiplied by the value on the input on the same position.



Layers of a CNN



A stride one 3x3 convolutional kernel acting on a 8x8 input image, outputting an 8x8 filter/channel

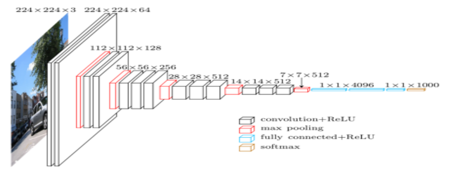
Architectures of Convolutional Neural Networks

VGGNet: is the name of a pre-trained convolutional neural network VGG 16 and VGG 19, having 16 and 19 weight layers, respectively, have been used for object recognition. VGGNet takes input of RGB images and passes them through a stack of convolutional layers with the fixed filter size of 3×3 and the stride of 1.

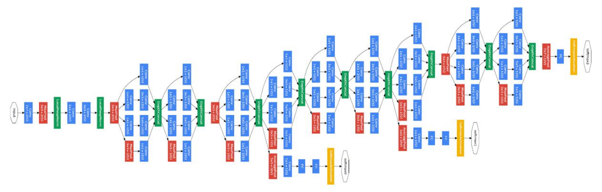
GoogleNet: The first branch in the Inception module simply learns a series of 1 ×1 local features from the input. The second batch first applies 1 ×1 convolution, not only as a form of learning local features, but instead as dimensionality reduction. Larger convolutions (i.e., 3 × 3 and 5 × 5) by definition take more computation to perform.

AlexNet: consists of 5 convolutional layers and 3 fully connected layers. Multiple Convolutional Kernels extract interesting features in an image. In a single convolutional layer, there are usually many kernels of the same size. The first convolutional layer of AlexNet contains 96 kernels of size 11x11x3. Note the width and height of the kernel are usually the same and the depth is the same as the number of channels.

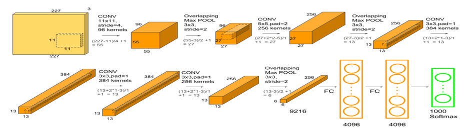
ResNet: is available in two types of plain network and residual network. ResNet as a plain network is inspired by the philosophy of VGGNets. The convolutional layers mostly have 3×3 filters have the same output feature map of the number of filters. The features map is halved so the time complexity of each layer is halved ResNet has fewer filters and lower complexity than VGG nets.



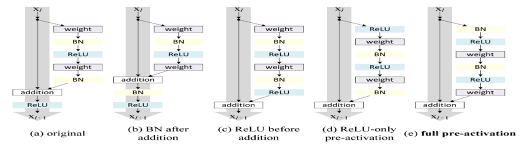
VGGNet



GoogleNet



AlexNet



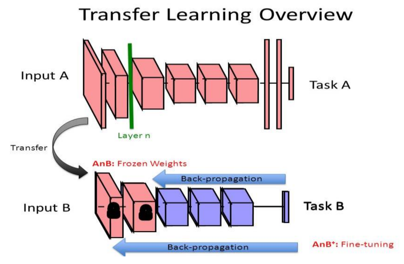
ResNet

Deep Learning Methods and Algorithms

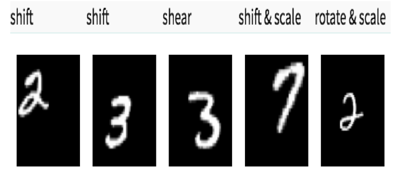
Transfer Learning: learns in one task to improve generalization in another. We transfer the weights that a network has learned at "task A" to a new "task B." The general idea is to use the knowledge a model has learned from a task with a lot of available labeled training data in a new task that doesn't have much data.

Data augmentation: is about accumulating more data than what is already available. 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).

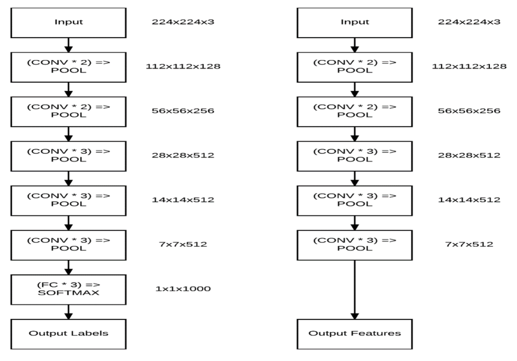
Deep feature extraction: begins by propagating the data of image through the layers in a convolutional neural network. The final classification probabilities are obtained at the end of the network. Deep feature extraction works through the various pooling layers that each implement a different filters or sizes for each layer.



Transfer Learning



Data augmentation



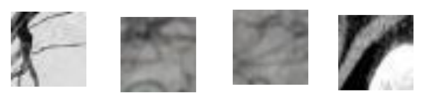
Deep feature extraction

Dataset and Experiment

•The experiment is conducted with the CT_Chest_Patch dataset having 2645 images with 12 different segmented CT lung patches and the MedPatch60 dataset with 18854 images with 60 different modalities. 80% images was used for the testing set and 20% for training set.

•Accuracies are compared using different classification techniques by using VGGNet, AlexNet, GoogleNet, and ResNet convolutional neural network architectures.

Images of CT_Chest_Patch Dataset



Images of MedPatch60 Dataset

Results

Results are compared with other four CNNs. These CNNs are tested with two datasets.

VGGNet Results for CT Chest Dataset

	precision	recall	f1-score	support
background_black	0.66	0.91	0.77	34
bronchi	0.72	0.78	0.75	69
corner_black_grey	0.88	0.85	0.86	75
cyst	0.71	0.72	0.72	72
fat_tissue_corner	0.93	0.92	0.92	73
ftissue_white	0.95	0.90	0.93	42
groundglass	0.91	0.80	0.85	60
honeycomb	0.72	0.88	0.79	41
nod_micro	0.77	0.82	0.80	62
nodule_macro	0.89	0.65	0.75	60
tissue_normal	0.91	0.85	0.88	74
avg / total	0.83	0.82	0.82	662

VGGNet Results for MedPatch60 Dataset

angio_coronary	0.81	0.75	0.78	99
background_black	0.81	0.74	0.77	65
background_blue	0.99	0.99	0.99	109
background_brown	0.99	1.00	1.00	144
background_green_blue	0.88	0.81	0.84	78
background_grey	0.97	0.91	0.94	75
background_white	0.98	0.78	0.87	67
background_yellow	0.92	0.92	0.92	39
charts_bar_color	0.87	0.92	0.89	36
charts_bar_grey	0.90	0.93	0.91	56
chemical_structure_bw	0.88	0.92	0.90	38
chemical_structure_color	0.93	1.00	0.97	28
ct_abdomen_liver	0.66	0.77	0.71	95
ct_abdomen_spine	0.79	0.52	0.62	29
ct_bronchi	0.68	0.71	0.70	73
ct_corner_black_grey	0.81	0.90	0.85	136
ct_fat_tissue_corner	0.84	0.88	0.86	83
ct_ftissue_white	0.80	0.80	0.80	90
ct_grey_fattissue	0.74	0.83	0.78	105
ct_groundglass	0.80	0.66	0.72	65
ct_honeycomb	0.65	0.68	0.67	44
ct_lungcyst	0.72	0.74	0.73	74
ct_nodules	0.72	0.74	0.73	120
ct_tissue_normal	0.91	0.71	0.80	75
dental_white	0.93	0.78	0.85	64
ecg_signal	0.99	0.98	0.98	150
endoscopy_red	0.88	0.84	0.86	133
gel	0.85	1.00	0.92	47
handdrawn_color	0.97	0.76	0.85	37
handdrawn_grey	0.93	0.95	0.94	119
microscopy_cell_blue	0.86	0.93	0.89	45
microscopy_cell_pink	0.95	0.84	0.89	44
microscopy_electron_grey	0.75	0.77	0.76	111
microscopy_fluorescence	0.98	0.94	0.96	126
microscopy_histo_blue	0.92	0.93	0.92	152
microscopy_light_green	0.97	1.00	0.99	71
microscopy_light_pink	0.87	0.91	0.89	195
microscopy_transmission_grey1	0.85	0.97	0.91	77
microscopy_transmission_grey2	0.67	0.78	0.72	64
microscopy_violet	0.96	0.90	0.93	182
mri_fetus	0.47	0.48	0.48	52
mri_head_brain	0.79	0.80	0.79	106
mri_leg	0.79	0.74	0.76	81
organ_tissue_other	0.78	0.79	0.78	39
organ_tissue_red	0.80	0.89	0.85	170
pet_ct_color	0.83	0.97	0.90	40
pet_lung_cancer	0.88	0.47	0.61	15
photo_eye	0.89	0.42	0.57	19
photo_tissue_cardiac	0.80	0.87	0.83	86
print_letters_black_bkg	0.88	0.94	0.91	31
print_letters_white_bkg	0.92	0.98	0.95	57
skin_abnormal	0.93	0.91	0.92	89
skin_normal_dark	0.93	0.86	0.89	72
skin_normal_light	0.89	0.79	0.84	62
us_color	0.87	0.97	0.92	62
us_grey	0.88	0.89	0.89	104
xray_chest_bone	0.84	0.89	0.86	61
xray_finger_bone	0.99	0.82	0.89	87
xray_knee	0.71	0.85	0.77	20
xray_vertebra	0.64	0.50	0.56	18
avg / total	0.86	0.86	0.85	4711

CNN	VGGNet	ResNet	AlexNet	GoogleNet / Inception v4
Precision	0.83 (CTC) 0.86 (MP60)	0.95 (CTC) 0.97 (MP60)		
Recall	0.82 (CTC) 0.86 (MP60)	0.95 (CTC) 0.97 (MP60)		
F1-Score	0.82 (CTC) 0.85 (MP60)	0.95 (CTC) 0.97 (MP60)		
Accuracy (with crops)			50.89% (CTC) 50.14% (MP60)	
Accuracy (no crops)			44.64% (CTC) 45.17% (MP60)	
Rank-1 and Rank-5				CTC R-1: 77.5% R-5: 100% MP60 R-1-: 72% R=5: 93%

Conclusions and Future Work

The surge of deep learning over the last years is to a great extent due to the strides it has enabled in the field of computer vision.

We will build content based retrieval system that will act as a image search engine that perform classification from a search query and perform

References and Acknowledgements

•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.

•Simonyan, K., & Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition,Abs/1409.1556*. http://arxiv.org/abs/1409.1556 (accessed 26 August 2018)