Identifying Carotid Plaque Composition in MRI with Convolutional Neural Networks

Yuxi Dong[1], Yuchao Pan[1], Xihai Zhao[2], Rui Li[2], Chun Yuan[2,3] and Wei Xu[1]
[1]Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China
[2]Center for Biomedical Imaging Research, Department of Biomedical Engineering,
Tsinghua University School of Medicine, Beijing, China
[3]Department of Radiology, University of Washington, Seattle, USA

***Abstract*—Carotid plaques may cause strokes. The composition of the plaque helps assessing the risk. Magnetic resonance imaging (MRI) is a powerful technology for analyzing the composition. It is both tedious and error-prone for a human radiologist to review such images. Traditional computer-aided diagnosis tools use manually crafted features that lack both generality and accuracy. We propose a novel approach using Deep convolutional neural networks (CNN) to classify these plaque tissues. In order to accommodate the multi-contrast MRI images, we modify state-of-the-art CNN models to support different number of input channels, and also adapt the models to do pixel-wise predictions. On a dataset with 1,098 human subjects, we show that we achieve significantly better accuracy than previous models. Our result also indicates interesting relations between contrast weightings and tissue types.**

I. Introduction

Carotid atherosclerosis is a common disease. It is caused by the building-up of cholesterol, fat, calcium, and other substances in the artery walls. The buildup is called a *plaque*. The plaque clogs the arteries, causing a local change in vessel size [1]. The narrowed vessel reduces blood supply. If a plaque suddenly ruptures, the blood clot may cause a stroke. Stroke keeps the second place in global death ranks from 1990 to 2010 [2], and rises from fifth to third in global disability-adjusted rank [3]. Of the different causes of stroke, atherosclerosis plaque has been identified as a major cause [4].

People have developed effective interventions / treatments for atherosclerosis, such as lipid-lowering drugs and calcium channel blockers, according to different stages and risks in the course of the disease development.

As most people do not develop symptom for atherosclerosis, it is essential that we can diagnose the condition early and reliably assess the risk of a certain plaque. Clinically, plaque composition, such as lipid core, calcification, and hemorrhage, are important indicators of their risks. For example, plaques with hemorrhage (bleeding) are vulnerable and are highly likely to cause strokes.

The most widely used techniques for plaque detection is B-mode ultrasound. However, ultrasound is highly operator dependent and thus not suitable for tracking the progression / regression of the disease for longer terms.

High-resolution multi contrast magnetic resonance imaging (MRI) has emerged as a promising tool for visualization of atherosclerotic plaques. Trained radiologist can identify plaque composition using MRI [5], [6].

However, it remains a challenge to review these images. Firstly, training experienced radiologists takes effort, resulting in the lack of qualified reviewers in hospitals in less developed regions. Secondly, the results may vary even with experienced reviewers on different readings. This is because the review guidelines are quite qualitative, based on the difference in the brightness between the plaques and the surrounding muscle tissues (see Section III-A). Thus, human factors still play an important role, making the results less repeatable [7]. Last but not least, the review process is time-consuming, as there are many locations per subject, and each location contains multiple images representing different contrast weightings. A reviewer needs to compare multiple images to identify the vessel walls and plaque composition.

All these problems call for an automated way to analyze the composition of atherosclerotic plaques. In other words, we need a way to *segment* the regions in an MRI representing different tissue types. Morphology-Enhanced Probabilistic Plaque Segmentation (MEPPS) [8] is a popular framework. MEPPS proposes a novel segmentation method using maximum a posteriori probability Bayesian theory. MEPPS models the likelihood of each pixel to be one of the tissue types. The model also includes morphologic information, such as local vessel wall thickness. MEPPS provides a general methodology for MR image analysis that many projects adopt [9], [10]. People have developed other Bayesian models demonstrating the feasibility of automatic MR image segmentation, such as [11], [12].

The limitation of the existing methods, however, is that they depend on hand-crafted features, and thus very specific to certain problems or sequences.

Recently, people have successfully applied convolutional neural networks (CNNs) in different applications including medical imaging processing [13]–[20]. There are several advantages with the CNN approach: 1) the CNN model is an end-to-end model that directly output the class labels for each pixel, eliminating most of the guessing work on different thresholds; 2) the CNN model does not depend on any domain-specific knowledge, such as the notion of vessel wall thickness. Of course, as we will see in the paper, data-specific tuning optimizes model performance, but even these optimizations

978-1-5090-6517-2/17/$31.00

do not require application-semantic-dependent features; 3) The CNN model improves with the amount of training data, which matches the current trend of data collection throughout the health care industry.

In this paper, we demonstrate a novel method that automatically segments the atherosclerotic plaques using deep convolutional neural networks and provides pixel-level classification of plaque tissue types. We build our model based on CNNs pre-trained on ImageNet [21] so we can obtain a good model with a limited number of training samples. To adapt the multi-contrast MRI images into existing models, we modify the network's input and output layers. Specifically, we expand the input to accept the four independent weightings. For the output, we modify the network to perform pixel-wise prediction. We also reassign the image downsampling rate to preserve high resolution.

We evaluate our method on a recent research dataset with over $70,000$ images from $1,098$ human subjects. In almost of the cases, we achieve significant accuracy improvements over the widely-used MEPPS approach.

II. Related Work

Automatic medical imaging analysis and diagnostics has been a hot area for many years. Traditional medical image analysis tools are based on hand-crafted features. Most of them use features about a region-of-interest (ROI) in an image, such as morphologic information in MEPPS [8], histogram of oriented gradients (HoG) [22] and scale-invariant feature transform (SIFT) [23]. These features are then used to train a classifier to differentiate normal anatomy from abnormal.

Researchers try to apply CNN on different medical images in recent years, such as MRI [13], [14], [16], [17], CT [18], [19] and chest X-rays [15]. Similar techniques are also used in classifying skin cancers [20]. Researchers use CNNs in medical imaging applications in the following three ways.

1) Combining off-the-shelf CNNs (pre-trained CNNs without fine-tuning) features with hand-crafted image features. The hope is that off-the-shelf CNNs complement the limited hand-crafted features with some general features like the lines and textures. Ginneken *et al.* [14] uses off-the-shelf CNN features in pulmonary nodule detection. Another application is chest pathology detection [15], and the best performance is achieved using CNN and GIST [24] features.

2) Training a CNN model from scratch. The technique is common for applications with abundant training data. For example, brain disease diagnosis [13], [16], [17] and detecting coronary artery calcifications [18].

3) Fine-tuning an ImageNet (or other database) pre-trained CNN model on medical image. *Shin et al.* [19] explores the possibility of fine-tuning a pre-trained CNN on CT images for lymph node detection and interstitial lung disease detection. An ImageNet pre-trained GoogLeNet is fine-tuned for skin cancer classification on a dataset of 129,450 skin lesions comprising of 2,032 different diseases [20].

All the above works are dependent on one image (in CT, MRI, X-ray or natural RGB images), where each image is taken in a very short period of time. While in our study, different contrast images are scanned separately. Different contrast weightings in the same slice may slightly differ in shape, scale and position due to patient movement. To our knowledge, our study is the first work on learning multi-contrast MRI in medical image area.

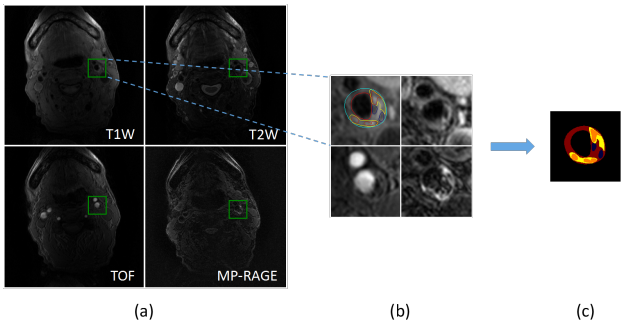


Fig. 1. An example of preprocessing of 4 contrast weightings, with necrotic core colored in yellow, hemorrhage in orange, calcification in dark blue and fibrous tissue in red.

III. Background

A. MRI and image contrast weighting

Magnetic resonance imaging (MRI), developed in the 1970s [25] and widely used today, is a tomographic imaging technology. It uses magnetic resonance phenomenon to form images of the human body anatomy.

In a nutshell, MRI generates images using pulses of radio waves to excite the nuclear spin energy transition so the atoms generate some detectable radio signal. Of all the atoms, hydrogen atoms, which are abundant in waters and fat in the body, are most often used in the detection. By varying the parameters of the pulse sequence and use a magnetic field to localize the signal in space, we can get different *contrasts* from different tissues, based on the hydrogen atoms within them.

In order to reveal different anatomical structures or pathologies, the MRI scans each position with different scan parameters, such as varying the repeat time (TR) or the echo time (TE). Under these scan parameters, as each type of tissue exhibits a unique way of returning to its equilibrium state after excitation (called *relaxation*), people can tell apart these tissues and their intrinsic properties such as blood flow speed. For example, there are T1 (spin-lattice) and T2 (spin-spin) relaxations.

Clinically, these contrasts are weighted in multiple ways to generate the final images. For example, *T1-weighted (T1W)* images put more weight on the tissues with more T1 relaxation, while *T2-weighted (T2W)* images emphasize the T2 relaxation process. Thus, tissues with different properties shows distinguishable intensity on T1W and T2W images.

Figure 1(a) provides an example of the four contrast weightings we use in this study. From image analysis point of view, we can treat these different contrast weightings as multiple channels of an image, like the RGB channels. However, there are two major differences: 1) these images contain much

TABLE I
TISSUE CLASSIFICATION GUIDELINE

	T1W	T2W	TOF
Necrotic/Lipid Core with			
A. No or little hemorrhage	o/+	−/o	o
B. Fresh hemorrhage	+	−/o	+
C. Recent hemorrhage	+	+	+
D. Old hemorrhage	−/o	−	−/o
Calcification	−	−	−
Hemorrhage			
A. Fresh	+	−	+
B. Recent	+	+	+
Loose Matrix	−/o	+	o
Fibrous Tissue	o	o	−

The symbols describes the signal intensities relative to adjacent muscle: + is hyper-intense, o is iso-intense and − is hypo-intense.

redundant information - a specific tissue can show some signal on one or more weightings; 2) different from a digital photo where all RGB channels are captured at the same time, the MRI equipment collects different contrast data at different time. During the collection, the subject (patient) may move, resulting in more noise.

B. Tissue classes

We focus on identifying the following tissues in a vessel.

- Lipid-rich/necrotic core (LR/NC) is an extracellular mass in the intima. Lipid cores may flow out mostly through a fissure or rupture of a fibrous cap into lumen, and it may cause severe cerebrovascular embolism.
- Calcification happens when calcium builds up in blood vessels. Calcium is transported through the bloodstream, and thus calcification can be found in almost any part of the body.
- Intraplaque hemorrhage is a liquid plaque component, and occurs frequently during the development of atherosclerotic lesions. Hemorrhage is prone to rupture, resulting in acute thrombosis.
- Loose matrix includes tissues that are loosely woven, such as proteoglycan-rich fibrous matrix and organizing thrombus.
- Fibrous tissue includes all other remaining tissues in the vessel except for four types above. It is considered the "normal" tissues in this study.

For the past years, people have developed guidelines to identify each tissue type [26], [27]. The reviewers of our dataset roughly follows the methods established in the literature [28], [29], and we summarize these guidelines in Table I.

C. Convolutional Neural Networks (CNN)

Comparing to a traditional three-layer neural network, a convolutional neural network (CNN) is stacked by many layers. Most of these layers are either convolutional layers or pooling layers. The CNN takes its input, i.e. images with multiple channels, at the lowest layer, and output the final results at the highest layer.

Most computation happens within the convolutional layers. These layers combine the information of adjacent pixels, as well as pixels from different input channels and produce a new output pixel. Assume that the input size of a convolutional layer is $d \times h \times w$, let $p_{i,x,y}$ be the pixel at channel i with position (x, y), and let $q_{j,x,y}$ be the pixel of the output channel j with position (x, y). Given a convolution filter $w_{i,j}$, we can compute the output pixel by

$$q_{j,x,y} = \sum_{i,s,t} p_{i,x+s,y+t} w_{j,i,s,t} + b_j$$

where b_j is the bias of channel j. To compute $q_{j,x,y}$, we enumerate i over all input channels, and s, t over the filter with typical size of 3×3 or 5×5.

The pooling layers reduce the size of their input, so that the input to the following convolutional layers can be smaller, allowing efficient computation. The size reduction is controlled by the window size w of a pooling layer. The output of a pooling layer is

$$q_{i,x,y} = g(p_{i,x*w+s,y*w+t})$$

where $s, t < w$, g is the pooling function, such as *max()* or *average()*.

To enable the network to capture nonlinear relationships in the input data, usually we add a non-linear function such as $ReLu$ or $sigmoid$ after layers with only linear transformations, such as the convolutional layer.

CNNs are widely used in classification tasks. In these tasks, we usually use a fully connected (FC) layer as the second-to-last layer. The FC layer receives the output from the previous pooling layer. It works in a similar way as to the hidden layer in traditional three-layer artificial neural networks.

IV. OUR APPROACH

A. Datasets

Data Acquisition. The dataset in this study comes from Chinese Atherosclerosis Risk Evaluation study (CARE II, [30]). This study consecutively recruits over 1000 patients, between ages 18 and 80, from 13 medical centers and hospitals all over China. All patients have stroke or transient ischemic attack within two weeks after onsets of symptoms. Also, B-mode ultrasound imaging has detected atherosclerotic plaques in carotid artery. This study is approved by institutional review board of each participating institution. All study participants have provided written informed consent.

These centers conduct the study in collaboration with the Vascular Imaging Laboratory (VIL) at the University of Washington, who has extensive experience in quantitative review of the carotid plaques. Also, Center for Biomolecular Imaging Research (CBIR) of Tsinghua University severed as a hub for the study, collecting all the data.

All MR imaging in this study is performed on state-of-the-art 3.0T MR scanners with 8-channel phase array coil. All centers adopt a multi-contrast high-resolution vessel wall imaging protocol for the carotid plaque imaging. The protocol includes the following imaging sequences (see Section III-A):

TABLE II
IMAGING PARAMETERS

	Standardized multicontrast imaging protocol			
	TOF	T1W	T2W	MP-RAGE
Sequence	FFE[1]	TSE[2]	TSE	FFE
Black blood	None	QIR	MDIR	
Repeat time (ms)	20	800	4800	8.8
Echo time (ms)	4.9	10	50	5.3
Flip angle	20°	90°	90°	15°
Field of view (cm)	14x14	14x14	14x14	14x14
Matrix	256x256	256x256	256x256	256x256
Scan plane	axial	axial	axial	axial
Slice thickness (mm)	1	2	2	1

[1] FFE: Fast Field Echo
[2] TSE: Turbo Spin Echo

- three-dimensional time-of-flight (TOF, TR=20ms, TE=4.9ms) [31];
- T1-weighted (T1W, TR=800ms, TE=10ms) quadruple inversion recovery [32];
- T2-weighted (T2W, TR=4800ms, TE=50ms) multislice double inversion recovery [33];
- magnetization-prepared rapid acquisition with gradient echo (MP-RAGE, TR=8.8ms, TE=5.3ms) [34].

Table II [30] details the imaging parameters of each imaging sequences for completeness. However, this detail is unimportant to the core method of this paper.

Data collection and reviewing. CBIR of Tsinghua University collects and archives all images in a centralized database. Trained reviewers with over three years' experience then review each image. The reviewers uses a custom-designed software, the *Computer-Assisted System for Cardiovascular Disease Evaluation (CASCADE)* [35]. CASCADE provides data storage as well as the labeling tool to draw the boundaries of different tissues. The software then stores the boundaries as contours in the database. Each image is reviewed by two reviewers with consensus.

Qualities of these images vary. Thus, the reviewers also grade these images on a 5-level image quality scale (1 being the worst). In this study, we only use the images with quality gradings of 2 or above.

The dataset contains 4 image sequences corresponding 4 contrast weightings for each subject. As different centers provide a different number of images per sequence, we align the different sequences using the bifurcation level and slice thickness. Finally, we keep 16 slice locations per sequence per subject, we call them the *study locations*. Each study location contains 4 images representing 4 different contrast weightings. With $1,098$ subjects, we have $1,098 \times 16 = 17,568$ study locations and $70,272$ images in total.

Data preparation. Before any analysis, all images goes through three preprocessing steps. 1) the reviewer identifies the regions of left and right carotid arteries; 2) she enlarges the artery region by 400% using bilinear interpolation, so people can see the details more clearly; and 3) she adjusts the regions with manual delineated shifts, so that the four images are co-registered (i.e. aligned). The co-registration is an approximation and it may be off by a few pixels. After the preprocessing, the reviewers draw contours to mark the vessel walls and each tissue type.

In addition, we add two steps before feeding the images to our model training. 1) we cut out from the image a square region of size from 256x256 to 480x480, containing the entire artery; and 2) we convert the contour labels into a pixel-level label, by assigning all pixels enclosed in a contour as the same class.

Figure 1 summarizes the preprocessing steps. After preprocessing, for each study location, we get 4 co-registered MR images of same size as the input (Fig. 1(b)) with pixel-wise labels (Fig. 1(c)).

B. Goal and challenges

Our goal is to identify the plaque tissue types, as well as the vessel walls, in each study location, using the 4 contrast weightings. In other words, we would like to classify each pixel with its tissue type. We treat the 4 images at each location as 4 separate input channels, much like the RGB input channels in a colored image.

Although we are using a fairly large dataset with over a thousand subjects, the data size is still orders of magnitude smaller than normal image classification training set, such as the ImageNet. In fact, the training fails to converge to a usable model using only the MRI dataset. Thus, we use pre-trained models as a starting point, like many existing projects do [19], [20].

There are two challenges adapting the pre-trained models in MRI images.

Firstly, we have four input channels, while all our pre-trained models take three channel (RGB) images.

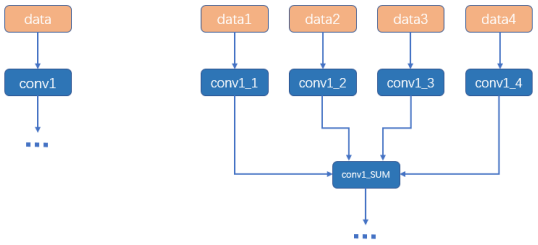
Secondly, unlike many image segmentation problems, where the object we want to recognize is significant in the image, the tissues we care about are small in the images. For example, a calcification we detect is only about 20 by 20 pixels in size. The output labels of most commonly used networks do not provide the pixel-level resolution enough for this problem.

Thus, we would like to design a model that provides enough resolution on our four channel MRIs, while still be able to use a pre-trained model to improve its quality.

C. Our Model

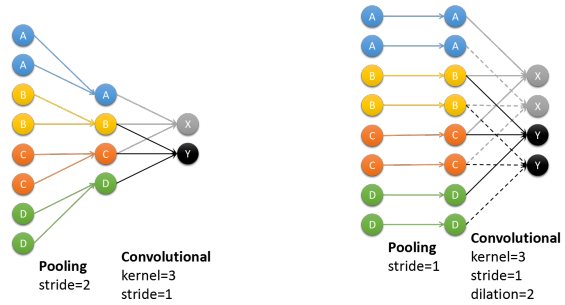
Base models. We build models based on three state-of-the-art CNN models: GoogLeNet [36], VGG-16 [37] and ResNet-101 [38].

VGG adopts a framework very similar to traditional networks such as LeNet [39] and AlexNet [21]. However, it stacks convolutional layers with kernel size of 3×3, the smallest kernel size required to capture neighboring information. The receptive field of 3 stacked convolutional layers with kernel size 3×3 is same as one convolutional layer with kernel size 7×7, while the former has fewer parameters and more nonlinearity. This structure allows VGG to have more layers, and thus improves performance over LeNet or AlexNet.



(a) Original input layers (b) Modified input layers

Fig. 2. Modification on input layers to accommodate four contrast weightings



(a) (b)

Fig. 3. Changing the stride of the last downsampling layers to retain resolution

GoogLeNet introduces the *inception module*, which combines the output of 1×1, 3×3, 5×5 convolutions and 3×3 max pooling on its input. The designers further propose 1×1 convolution before 3×3 and 5×5 convolutions to reduce dimension, in order to prevent blowing-up in computational complexity in large networks. This architecture actually generates a multi-scale output for a given input image, and it allows the next stage to process feature maps of different scales simultaneously. However, as the 3×3 max-pooling layers lead to spatial information loss, GoogLeNet usually performs worse than VGG in localization tasks.

Residual network is a recent work that allows the network to go much deeper, by addressing the degradation problem. The basic idea of ResNet is to learn the *residual* of a mapping $F(x)$, which is $F(x) - x$. Learning the residual instead of the mapping itself allows us to train a much deeper network. Thus it has outperformed VGG and GoogLeNet on ImageNet classification by a significant margin in many tasks. We use ResNet-101 with 100 layers.

We obtain the pre-trained model for each of the network from the Caffe Model Zoo [40]. We make the following three key modifications to each base model to adapt them to the MRI application.

Adapting 4-channel images into models pre-trained on RGB images. In order to adapt our data into a CNN model pre-trained on the RGB color images on ImageNet, we make 4 copies of the data layer in each model. Each data layer takes a single channel as input. We also make 4 copies of the first convolutional layer, and connect them to each of the data layer. Then we connect these 4 convolutional layers to a sum-up layer that merges several feature maps with the same size into a single feature map by simple summation. The output of this sum-up layer is then connected to the rest of the network. Figure 2 illustrates the modifications to the input part of the network, where Figure 2(a) shows the structure of original network, and Figure 2(b) shows the modified version.

Making the network to output pixel-level class labels. The base networks are designed to do classifications. Thus, they all have a fully connected layer near the end to produce the classification label. However, we want to classify *each pixel*, rather than the entire image.

Some methods [41], [42] approximate pixel-level classification by extracting patches from an image and predicting the label of the pixel at the center of each patch. With this approach, it is crucial to determine a reasonable patch size. Unfortunately, the sizes of tissues in our images vary dramatically, making the patch size hard to choose.

Instead, we adopt the fully convolutional neural network (FCN) [43] approach. With this design, we replace the last fully connected layer in classification models, such as AlexNet and VGG, to a fully convolution layer. This structure eliminates the need for extracting patches and thus improves the effectiveness and efficiency of pixel-wise prediction. We also adopt the skip layer fusion, in which we combine upsampling of lower layers with the final prediction layer to predict finer details while retaining semantic information.

Reduce the downsampling factor to retain resolution. The base model employs some downsampling layers in order to reduce the size of the images propagating through the network to save computation cost. This is usually enough for tasks like image classification. However, the downsampling causes too much information loss for pixel-wise prediction.

Thus, following the methods in DeepLab [44], we change the downsampling from $32\times$ to $8\times$ by changing the stride of the last two pooling layers of each model from 2 to 1 and apply the *atrous algorithm*.

Figure 3 shows an example of this algorithm in 1-dimention. Figure 3(a) shows the original structure with $2\times$ downsampling, and receptive field of neuron X is composed of A, B and C. After changing the stride, neuron X in Figure 3(b) should not convolute on any three continuous neurons, since this would make a wrong receptive field. Neuron X still need to convolute on neurons A, B and C, so we do the convolution by skipping neurons, and the number of skipped neurons is specified by *dilation* parameter in Caffe [45]. Thus the receptive field of neuron X stays unchanged.

V. Evaluation

We perform all model training and evaluations using a Ubuntu server with one Titan X GPU. We use Caffe [45] as our deep learning framework. We randomly select 20% subjects out of the $1,098$ as test set, while using the remaining for training. We count the percentage of different tissue types in both training and test set and Figure 4 shows the result [1]. We

[1]We omit the fibrous tissue as they are too common.

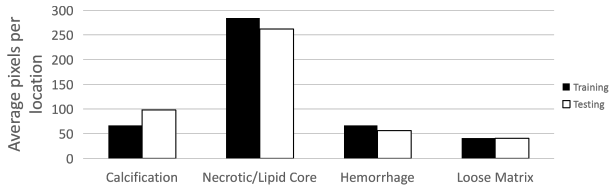


Fig. 4. Distribution of each tissue types, by number of pixels.

can see that both datasets have a similar distribution of tissue types, confirming the unbiased choice of the test set.

To provide the readers with some intuition before the formal evaluation, Figure 5 presents two concrete examples, showing the manual labels, and segmentation results using both MEPPS and ResNet-101. We can see that ResNet-101 detects all tissue types, even though the actual shape is a little different from the manual labels.

To formally evaluate our CNN models, we compare them to off-the-shelf MEPPS on two independent sets of accuracy metrics. We use reviewers' manual labels as the ground truth.

Firstly, we treat the problem as a pixel-level binary classification task for each tissue class, i.e. determining whether a pixel belongs to each of the classes. We present the recall, precision and F-measure for each tissue class. Let TP, FP, FN to be the true positive, false positive and false negative, respectively, we define the metrics as:

- precision = TP / (TP + FP)
- recall = TP / (TP + FN)
- F-measure = 2 (precision $\times$ recall) / (precision + recall)

Secondly, following the evaluation metrics in [43], [44], we present the problem as an image segmentation task and evaluate pixel-level accuracy between the manually drawn regions and predicted regions. Let a_{ij} be the number of pixels of tissue i predicted to be tissue j, $s_i = \sum_j a_{ij}$ be the total number of pixels of tissue i and n_t be the number of tissues ($n_t = 5$ in our case). We define the following three metrics:

- pixel accuracy $= \Sigma_i a_{ii} / \Sigma_i s_i$
- mean accuracy $= (1/n_t)\Sigma_i (a_{ii}/s_i)$
- mean IU = $(1/n_t)\Sigma a_{ii}/(s_i + \Sigma_j a_{ji} - a_{ii})$

The mean IU stands for the *region intersection over union*. Intuitively, a larger IU means the predicted tissue region overlaps more with the manually labeled regions.

A. Comparison Results

Table III presents the pixel-wise classification results for each tissue class, and Table IV shows the metrics on image segmentation. From both tables, we have the following observations:

1) CNNs outperform MEPPS in almost all tissue classes and metrics. The improvements mainly come from the elimination of manually crafted features, as well as the large amount of training data with consistent labels.

2) Among all the CNNs, ResNet-101 performs the best, as it does in many other image classification tasks, due to the depth it achieves.

TABLE III
PIXEL-WISE CLASSIFICATION ACCURACY

(a) Precision

Tissue	MEPPS	GoogLeNet	VGG-16	ResNet-101
Fibrous Tissue	0.947	0.922	0.944	**0.951**
Calcification	0.698	0.673	0.663	**0.704**
Necrotic/Lipid Core	0.373	0.533	0.536	**0.576**
Hemorrhage	0.526	0.710	0.717	**0.729**
Loose Matrix	0.103	0.422	**0.522**	0.488

(b) Recall

Tissue	MEPPS	GoogLeNet	VGG-16	ResNet-101
Fibrous Tissue	0.946	0.955	**0.978**	0.973
Calcification	0.457	0.446	0.481	**0.492**
Necrotic/Lipid Core	0.273	0.419	0.372	**0.474**
Hemorrhage	0.299	0.499	0.487	**0.622**
Loose Matrix	**0.253**	0.091	0.138	0.246

(c) F-measure

Tissue	MEPPS	GoogLeNet	VGG-16	ResNet-101
Fibrous Tissue	0.947	0.939	0.961	**0.962**
Calcification	0.552	0.536	0.557	**0.580**
Necrotic/Lipid Core	0.315	0.469	0.439	**0.520**
Hemorrhage	0.382	0.586	0.580	**0.671**
Loose Matrix	0.146	0.150	0.218	**0.327**

3) All methods identify the fibrous tissue class quite accurately. We believe the high accuracy is the results of abundant samples in the training data: fibrous tissue is the "normal" class that dominates the vessel walls.

4) We also get good accuracy on the calcification class. It is because this class is quite obvious: the calcified tissues produces hypo-intense relative to adjacent muscle in all weightings [29], which is easy to separate both by the reviewers and algorithms.

5) In contrast, although CNNs achieve several times improvements over MEPPS on the loose matrix cases, all metrics remain low comparing to other classes. The reason is that the review guidelines are vague for the loose matrix. This type of tissue is not crucial clinically and thus its labels are more noisy than other classes. Also, there is not enough training data for this class. For both reasons, the accuracy is lower than other classes.

6) All the pixel accuracy in Table IV is high because we can accurately classify the dominating fibrous tissue class. We also show the accuracy results excluding this class, which is much higher than MEPPS.

Table V reports the average running time for processing the 16 study locations of a subject. ResNet runs about 20% slower than other methods, but all the performance are acceptable considering that it takes minutes for a human reviewer to read these images.

B. Contribution of each contrast weighting

All our MRI results come as a bundle of four images with different contrast weightings, and our base model makes use of all these weightings. We want to see which contrast weighting contributes the most to the detection of each tissue class. To do so, we train four ResNet-101 models separately, using one

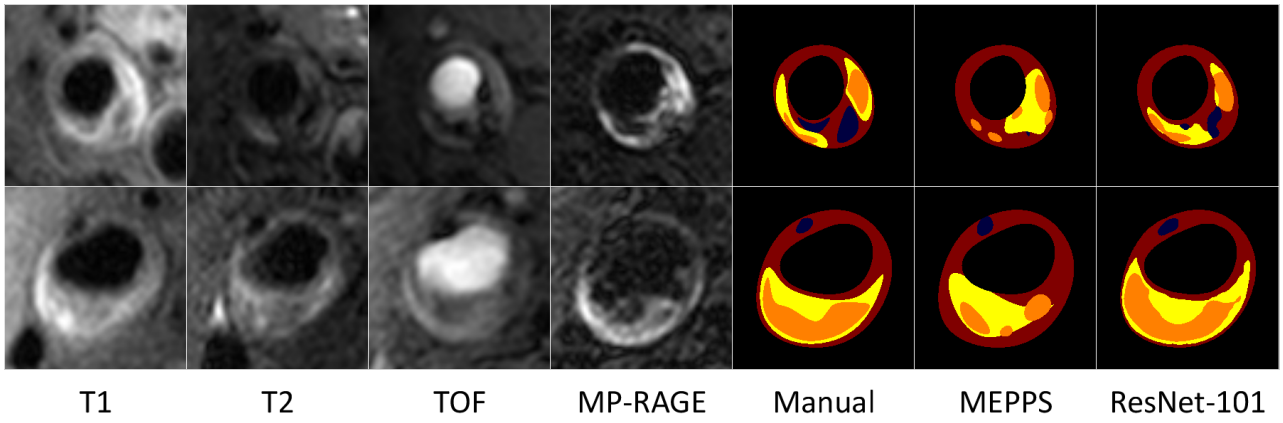


Fig. 5. Examples of composition analysis, with necrotic core colored in yellow, hemorrhage in orange, calcification in dark blue and fibrous tissue in red.

TABLE IV
SEGMENTATION ACCURACY

	pixel acc.	pixel acc. (no fibrous tissue)	mean acc.	mean IU
MEPPS	0.903	0.287	0.434	0.351
GoogLeNet	0.927	0.405	0.488	0.421
VGG-16	0.929	0.389	0.492	0.427
ResNet-101	**0.933**	**0.486**	**0.563**	**0.481**

TABLE V
PER-SUBJECT RUNNING TIME

	MEPPS	GoogLeNet	VGG-16	ResNet-101
Time (sec)	10.0	9.1	8.9	11.4

contrast weighting each, and the top half in Table VI reports the F-measure for each case.

We find that each contrast weighting does contribute to different tissues types in different ways. For example, T1W mainly contributes to calcification, while helping loose matrix little, and T2W helps lipid core and loose matrix mostly. The most extreme case is that MP-RAGE dominates the hemorrhage classification, achieving a higher F-measure than all weightings combined.

We also try to combine the most common weightings, T1W and T2W, and include the result in Table VI. We can see that there are some improvements over the single-weighting models, but not significant.

We confirm with the expert reviewers that the difference is consistent with their review guidelines - reviewers often emphasis on one or two weightings only for a specific tissue type. Reviewers also confirm that MP-RAGE can be understood as an enhanced T1, while hemorrhage is best detectable on T1s.

TABLE VI
CONTRIBUTIONS OF EACH CONTRAST WEIGHTING

Contrast Weighting	Fibrous Tissue	Calcification	Lipid Core	Hemorrhage	Loose Matrix
T1W	**0.961**	**0.536**	0.496	0.443	0.020
T2W	0.958	0.494	**0.515**	0.323	**0.387**
TOF	0.957	0.468	0.465	0.487	0.080
MP-RAGE	0.957	0.337	0.437	**0.681**	0.015
ALL	0.962	0.580	0.520	0.671	0.327
T1W_T2W	0.962	0.545	0.525	0.468	0.401

Note that our model training and evaluation all use the reviewer labels as the ground truth. Thus, our models emulate the reviewers' logic, e.g. using only one or two weightings to identify the tissue type. Though we can emulate the human behavior quite accurately, these labels prevent us from exploring the correlations among these different weightings and further improve the results. As an important future work, we will train models based on the actual pathological result.

VI. CONCLUSION AND FUTURE WORK

With the accumulation of high-quality medical imaging data, we believe it has come to the point that we can use CNNs to replace many traditional computer-aided diagnosis software using hand-crafted features. This paper provides an example application using MRI, with an emerging application of plaque composition analysis. The application is quite new that there is still no definite guideline how to review these images. CNN is suitable for this case as it learns on examples rather than rules. We show that the CNN achieves very good accuracy.

CNN works in an end-to-end way on any kind of labels. As future work, we would replace the human reviewer labels with pathological images, in order to learn about the fundamental correlations between the MRI image and the actual tissue. Also, we would like to eliminate the need for manual image co-registration, allowing the model to use information on different contrasts without first aligning them.

VII. ACKNOWLEDGEMENT

This research is supported in part by the National Natural Science Foundation of China (NSFC) grant 61532001, Tsinghua Initiative Research Program Grant 20151080475, MOE Online Education Research Center (Quantong Fund) grant 2017ZD203, and gift funds from Huawei and Ant Financial.

REFERENCES

[1] G. Pasterkamp and E. Falk, "Atherosclerotic plaque rupture: An overview," *Journal of Clinical & Basic Cardiology*, vol. 3, no. 2, pp. 705–713, 2000.

[2] R. Lozano, M. Naghavi, K. Foreman, and et al., "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010," *The Lancet*, vol. 380, no. 9859, pp. 2095–2128, 2013.

[3] C. J. Murray, T. Vos, R. Lozano, and et al., "Disability-adjusted life years (dalys) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the global burden of disease study 2010." *Lancet*, vol. 380, no. 9859, pp. 2197–223, 2012.

[4] C. Yuan, L. M. Mitsumori, K. W. Beach, and et al., "Carotid atherosclerotic plaque: noninvasive mr characterization and identification of vulnerable lesions." *Radiology*, vol. 221, no. 2, p. 285, 2001.

[5] N. Takaya, C. Yuan, B. Chu, and et al., "Presence of intraplaque hemorrhage stimulates progression of carotid atherosclerotic plaques," *Circulation*, vol. 111, no. 21, pp. 2768–75, 2005.

[6] C. Yuan, L. M. Mitsumori, M. S. Ferguson, and et al., "The in vivo accuracy of multispectral mr imaging for identifying lipid-rich necrotic cores and intraplaque hemorrhage in advanced human carotid plaques," *Acc Current Journal Review*, vol. 11, no. 2, p. 37, 2002.

[7] E. Touz, J. F. Toussaint, J. Coste, and et al., "Reproducibility of high-resolution MRI for the identification and the quantification of carotid atherosclerotic plaque components: consequences for prognosis studies and therapeutic trials." *Stroke*, vol. 38, no. 6, pp. 1812–9, 2007.

[8] F. Liu, D. Xu, M. S. Ferguson, and et al., "Automated in vivo segmentation of carotid plaque MRI with morphology-enhanced probability maps," *Magnetic Resonance in Medicine Official Journal of the Society of Magnetic Resonance in Medicine*, vol. 55, no. 3, pp. 659–68, 2006.

[9] T. Yoneyama, S. Jie, D. S. Hippe, and et al., "In vivo semi-automatic segmentation of multicontrast cardiovascular magnetic resonance for prospective cohort studies on plaque tissue composition: initial experience," *International Journal of Cardiovascular Imaging*, vol. 32, no. 1, pp. 1–9, 2015.

[10] B. E. Wenbo Liu, N. Balu, S. M. Jie, and et al., "Segmentation of carotid plaque using multicontrast 3d gradient echo mri," *Journal of Magnetic Resonance Imaging*, vol. 35, no. 4, p. 812C819, 2012.

[11] t. K. R. Van, O. Naggara, R. Marsico, and et al., "Automated versus manual in vivo segmentation of carotid plaque MRI." *Ajnr American Journal of Neuroradiology*, vol. 33, no. 8, pp. 1621–7, 2012.

[12] t. K. R. Van, A. J. Patterson, V. E. Young, and et al., "An objective method to optimize the MR sequence set for plaque classification in carotid vessel wall images using automated image segmentation." *Plos One*, vol. 8, no. 10, pp. e78 492–e78 492, 2013.

[13] R. Li, W. Zhang, H. I. Suk, and et al., "Deep learning based imaging data completion for improved brain disease diagnosis," *Med Image Comput Comput Assist Interv.*, vol. 17, no. 3, pp. 305–312, 2014.

[14] B. V. Ginneken, A. A. A. Setio, C. Jacobs, and et al., "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *IEEE International Symposium on Biomedical Imaging*, 2015, pp. 286–289.

[15] Y. Bar, I. Diamant, L. Wolf, and et al., "Chest pathology detection using deep learning with non-medical training," in *IEEE International Symposium on Biomedical Imaging*, 2015, pp. 294–297.

[16] Y. Pan, W. Huang, Z. Lin, and et al., "Brain tumor grading based on neural networks and convolutional neural networks." *Conf Proc IEEE Eng Med Biol Soc.*, vol. 2015, pp. 699–702, 2015.

[17] B. H. Menze, A. Jakab, S. Bauer, and et al., "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, p. 1993, 2015.

[18] J. M. Wolterink, T. Leiner, M. A. Viergever, and et al., *Automatic Coronary Calcium Scoring in Cardiac CT Angiography Using Convolutional Neural Networks*. Springer International Publishing, 2015.

[19] H. C. Shin, H. R. Roth, M. Gao, and et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[20] A. Esteva, B. Kuprel, R. A. Novoa, and et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, 2017.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, p. 2012, 2012.

[22] A. Seff, L. Lu, K. M. Cherry, and et al., "2d view aggregation for lymph node detection using a shallow hierarchy of linear classifiers," in *Medical Image Computing & Computer-assisted Intervention: Miccai International Conference on Medical Image Computing & Computer-assisted Intervention*, 2014, pp. 544–552.

[23] M. Toews and T. Arbel, "A statistical parts-based model of anatomical variability," *IEEE Transactions on Medical Imaging*, vol. 26, no. 4, pp. 497–508, 2007.

[24] A. Torralba, R. Fergus, and Y. Weiss, "Small codes and large image databases for recognition," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–8.

[25] P. C. Lauterbur, "Image formation by induced local interactions: Examples employing nuclear magnetic resonance," *Clinical Orthopaedics & Related Research*, vol. 244, no. 5394, p. 3, 1973.

[26] J. M. Cai, T. S. Hatsukami, M. S. Ferguson, and et al., "Classification of human carotid atherosclerotic lesions with in vivo multicontrast magnetic resonance imaging." *Circulation*, vol. 106, no. 11, pp. 1368–73, 2002.

[27] B. Chu, A. Kampschulte, M. S. Ferguson, and et al., "Hemorrhage in the atherosclerotic carotid plaque: A high-resolution mri study," *Stroke.*, vol. 35, no. 5, pp. 1079–1084, 2004.

[28] T. Saam, M. S. Ferguson, V. L. Yarnykh, and et al., "Quantitative evaluation of carotid plaque composition by in vivo mri." *Arteriosclerosis Thrombosis & Vascular Biology*, vol. 25, no. 1, p. 234, 2005.

[29] C. Yuan, W. S. Kerwin, V. L. Yarnykh, and et al., "Mri of atherosclerosis in clinical trials," *Nmr in Biomedicine*, vol. 19, no. 6, pp. 636–654, 2006.

[30] X. Zhao, R. Li, D. S. Hippe, and et al., "Chinese atherosclerosis risk evaluation (CARE II) study: a novel cross-sectional, multicentre study of the prevalence of high-risk atherosclerotic carotid plaque in chinese patients with ischaemic cerebrovascular events—design and rationale," *Stroke and Vascular Neurology*, vol. 2, no. 1, pp. 15–20, 2017.

[31] R. J. Nickles and H. O. Meyer, "Three-dimensional time-of-flight gamma camera system," 1976.

[32] V. L. Yarnykh and C. Yuan, "T1-insensitive flow suppression using quadruple inversion-recovery." *Magnetic Resonance in Medicine*, vol. 48, no. 5, pp. 899–905, 2002.

[33] ——, "Multislice double inversion-recovery black-blood imaging with simultaneous slice reinversion," *Journal of Magnetic Resonance Imaging*, vol. 17, no. 4, p. 478, 2003.

[34] M. Brant-Zawadzki, G. D. Gillan, and W. R. Nitz, "Mp rage: a three-dimensional, t1-weighted, gradient-echo sequence–initial experience in the brain." *Radiology*, vol. 182, no. 3, pp. 769–775, 1992.

[35] D. Xu, W. Kerwin, T. Saam, and et al., "Cascade: Computer aided system for cardiovascular disease evaluation," in *Proc ISMRM*, 2004, p. 1922.

[36] C. Szegedy, W. Liu, Y. Jia, and et al., "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1–9.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, 2015.

[38] K. He, X. Zhang, S. Ren, and et al., "Deep residual learning for image recognition," *Computer Science*, 2015.

[39] Y. Lcun, L. Bottou, Y. Bengio, and et al., "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[40] "Caffe Model Zoo," https://github.com/BVLC/caffe/wiki/Model-Zoo, 2014.

[41] F. Ning, D. Delhomme, Y. Lecun, and et al., "Toward automatic phenotyping of developing embryos from videos," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 14, no. 9, pp. 1360–1371, 2005.

[42] P. Pinheiro and R. Collobert, "Recurrent convolutional neural networks for scene labeling," in *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, 2014, pp. 82–90.

[43] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[44] L. C. Chen, G. Papandreou, I. Kokkinos, and et al., "Semantic image segmentation with deep convolutional nets and fully connected crfs," *Computer Science*, no. 4, pp. 357–361, 2014.

[45] Y. Jia, E. Shelhamer, J. Donahue, and et al., "Caffe: Convolutional architecture for fast feature embedding," *Eprint Arxiv*, pp. 675–678, 2014.