A deep autoencoder method for image quality assessment

Andre Maximo1, Chitresh Bhushan2, Desmond T.B. Yeo2, and Thomas K Foo2

1GE Healthcare, Rio de Janeiro, Brazil, 2GE Global Research, Niskayuna, NY, United States

Synopsis
We demonstrate a classification approach for MRI image-quality based on deep auto-encoder that can be trained with samples coming from only one class (e.g., only good image-quality). This approach is helpful in situations where class-imbalance is unavoidable (i.e., it is easy to obtain a large number of image samples from one class but very difficult to obtain similar number of samples from other class). Our approach shows excellent accuracy in binary classification with AUC of 0.975 in identifying MRI images of good & bad quality in clinical practice from several sites.

Introduction
Classification of MRI images into two or more classes has several clinical applications. Most classification approaches require knowledge of each class in the dataset and need a good balance of number of samples in each class. However, in many scenarios, it is easy to obtain a large number of image samples from a homogeneous group of subjects (e.g., normal or control), while it is difficult to obtain similar number of samples from diseased population (either because there is shortage of samples from other class or because it is extremely expensive to acquire such data). This imbalance in sample count across different classes can severely affect data-driven approaches to able to perform well in practical scenarios.

In this work we explore an approach using deep-autoencoder for binary classification for image-quality that can be trained to work with MR images coming from the first class and images from other classes are not needed for training. Specifically, given a MRI image we want to classify if it's image-quality is acceptable or not for a particular task. Our framework is generic and can be potentially also extended into multi-class classification problem.

Method
We use deep autoencoder to obtain a representation or encoding that is useful for classification of images. Autoencoders has been used for compression, denoising or dimensionality reduction1. However in this work we use a statistic of the internal encoded parameters for classification purposes, similar to anomaly detection2.3. Given a set of training images from homogenous class (images of acceptable image quality), we first train a deep autoencoder model. As shown in Fig.1, our deep-autoencoder model uses convolutional layers with successive max-pooling for the encoder part and successive up-sampling for decoder part. For the training we use mean square error (MSE) as our loss-function.

After training, we use the trained encoder model to generate 256 encoded features for each training sample (16x16 output from the middle layer of trained auto-encoder). We use these encoded features to compute a cluster statistic across all of the training data. For a new input image, we compare the encoded features to the computed cluster statistic to determine if the new sample is similar to the pool of data from homogenous training set. In particular, we compute true feature count (TFC), which is number of features that are similar to training-set based on mean and std. deviation of training pool (see Fig.2a). TFC can be used a distance metric and we use a TFC threshold based on the 5th percentile on training set for classification.

Data: For training autoencoder, we used T2-weighted MR image of brain from 750 clinical exams from 6 different sites (GE 3T scanner) that were known to be of good quality (dataset A). All studies were approved by an appropriate IRB. All images were resized to matrix size of 128x128x8 and were augmented 25-fold for the training before training. For testing purposes, we use two independent sets of 50 cases each: (dataset B) with good quality images (i.e., good quality images not included in the training data set); and (dataset C) that were rejected as bad-quality image in clinical practice.

Results and Discussion
Fig.2 shows the box-plots of TFC across the three datasets. We can see that box-plots dataset-B aligns well with the training set (dataset-A), indicating that features learnt by auto-encoder are consistent across two independent dataset of good-quality. Further, dataset-C (bad-quality images) shows a TFC distribution that is very distinct from both datasets with good-quality. A simple TFC threshold based on the 5th percentile yields excellent classification for testing set (B & C combined) with AUC of 0.975.

We also investigated effect of using structural similarity index (SSIM) loss-function while training the auto-encoder and found that classification accuracy was very similar with AUC of 0.972. Further, we also investigated effect of resizing input images and found to have minimal effect on classification with AUC of 0.975 for 128x128x8 and AUC of 0.974 for 256x256x8 input.

One limitation of our approach is that the encoder model is dependent on anatomy i.e. a model trained with only brain-image cannot be used other anatomy like knee and would require a separate model.
Conclusion
We demonstrated a classification approach for MRI image-quality based on deep auto-encoder that can be trained with samples coming from only one class (e.g., only good image-quality). This approach is helpful in situations where class-imbalance is unavoidable. Our approach shows excellent classification accuracy with AUC of 0.975 in identifying images of good & bad quality in clinical practice from several sites.

Acknowledgements
No acknowledgement found.

References


Figures
Details of our deep auto-encoder network, which is used to encode image from only one homogeneous class. After training the output from middle layer is used as encoded feature, that is used to compare with the training set for classification.

(a) Step of computing true feature count (TFC) for an input image. TFC is number of features that are less than one std. dev away from mean (as computed from training set of ) (b) Box plot of TFC for (A) training set with only good-quality images; (B) independent testing set with good-quality image; and (C) independent testing set with bad-quality image. A simple TFC threshold based on the 5th percentile of training-set yields excellent classification for testing set (B & C combined) with AUC of 0.975.