Deep Autoencoder Features for Registration of Histology Images

Ruqayya Awan[1] and Nasir Rajpoot[1,2,3]

[1]Department of Computer Science, University of Warwick, Coventry, UK
[2]The Alan Turing Institute, London, UK
[3]Department of Pathology, University Hospitals Coventry & Warwickshire, UK

Abstract. Registration of histology whole slide images of consecutive sections of a tissue block is mandatory for cross-slide analysis. Due to the stain variations, a feature-based method for deriving the transformation maps for these images is considered to be a reasonable choice as compared to the methods which work on image intensities. Autoencoders have been employed in a wide variety of applications due to their potential for representation learning and transfer learning for deep architectures. Representation learned by autoencoders has been used for a number of challenging problems including classification and regression. In this study, we analyze deep autoencoder features for the purpose of registering histology images by maximizing the feature similarities between the fixed and moving images. In this paper, we demonstrate the capability of autoencoder features for registration of histology images.

Keywords: Digital pathology, Autoencoders, Registration, Multiplex biomarkers, Immunohistochemistry.

1 Introduction

With the advent of digital scanners, the pathologist's practice for diagnosis is transforming from visual microscopic analysis to digitized tissue analysis. In diagnostic and research practice, cross-slide image analysis provides additional information by analyzing expression of different biomarkers as compared to a single slide image analysis. Slides stained with different biomarkers are analyzed side by side which can essentially provide some unknown relations between the different biomarkers. During the slide preparation, a tissue section may be placed at an arbitrary orientation as compared to other sections of the same tissue block. The problem is compounded by the facts that tissue contents are likely to change from one section to the next and there may be unique artefacts on some of the slides. This makes registration of each section with respect to a reference section of the same tissue block a mandatory task prior to any cross-slide analysis. Currently, this registration is done manually by the pathologists which is time-consuming due to the large number of sections taken from a single tissue block.

There are three main methods for image registration: control point registration, intensity based registration and feature-based registration [1] where the first method requires user input while the later two are fully automated methods. Handcrafted feature-based methods generally consist of four steps: 1) detection of salient features, 2) calculating descriptors from pixels around the detected features, 3) finding matching descriptors between the two images, 4) finding the transformation, mapping the matched features of a moving image with the corresponding features of a reference image. Registration based on matching hand-crafted features are not likely to perform well for all the image data. This limitation can be overcome by learning the latent feature representation from the data itself. Convolutional neural networks (CNNs) have been extensively used for various applications for learning the base functions from the data. These functions learn both low and high dimensional features from the data and are optimized by comparing the output with the ground truth for any particular task at hand. The only limitation in using CNN for medical applications is the availability of ground truth information. Unsupervised methods do not require ground truth data for feature extraction but their capability of learning complex representations depends on the type of method used. Linear unsupervised models such as PCA and ICA may not be suitable for a complex data representation since they are not able to learn the complex non-linear relationship upon reducing the dimensionality of the data. While autoencoder (AE), a deep unsupervised learning model, has the ability to learn the complex representation without using any ground truth information for feature learning and are more likely to perform better as compared to the handcrafted features. Recently, AEs have been used for a variety of tasks including transfer learning and getting the learned features and feeding them to separate model for a more challenging problems of classification and regression.

In this paper, we present a histology image registration framework based on unsupervised deep AE features. The motivation behind selection of AE for deep feature extraction is the fact that these networks do not require any ground truth for training. Our contribution in this paper is two-fold: 1) we propose to use convolutional AEs for learning features from the histology images which to the best of our knowledge have not been explored for histology image registration task, 2) we then perform registration by maximizing the mutual information between the learned features of the two images instead of using the images intensities. The AE learns the complex pattern among the training set for registration which have been shown to be useful for other complex tasks. Additionally, our adopted AE reduces the dimensionality by a factor of 16 which makes our proposed method fast making it tractable for its adoption for whole slide images (WSIs).

2 Previous Studies

There are many studies on medical image registration using intensity-based and feature-based methods including, both supervised and unsupervised learning.

Among intensity-based methods, mutual information has been widely used for medical images including pathology images [2, 3]. This method works by finding a transformation which maximises the mutual information between the two images in terms of pixel intensities. This criteria makes it less efficient for serial sections which are stained with different biomarkers since the tissue regions will be represented with different colors based on the type of biomarker used. This method would also be very time-consuming due to the large size of WSIs. Edge based registration methods are computationally very less expensive than the registration methods which are based on pixel intensities. Hierarchical Chamfer matching and registration based on curvature scale space (CSS) representation of the boundary points are the common examples of edge based registration. These methods are fast since the transformation is estimated based on the boundary point rather than the pixel intensities. In [4], Trahearn et al. employed curvature scale space (CSS) representation of the tissue boundaries for WSI registration. In another study [5], authors have used CSS based method for pre-alignment of serial sections stained with different biomarkers. The results of pre-alignment are further improved based on the nuclei clusters and fatty regions since these tissue structures are more likely to exist across several serial sections.

In [6], Mueller et al. demonstrated the feasibility of non-rigid spline-based registration for WSIs. The authors employed it for the alignment of serial tissue section WSIs with different staining characteristics such as tissues sections stained with immunohistochemical (IHC) biomarkers and haematoxylin and eosin (H&E) stain. Their proposed approach follows two-step multi-scale strategy to perform the transformation within a reasonable time constraint. In first stage, initial transformation map is estimated using a publicly available tool for registration, known as elastix. While in the second stage, the initial transformation is applied on the high resolution regions of WSI rather than the whole WSI. In another study [7], authors proposed a method for multi-stain and multimodal WSI registration, with the goal of 3D reconstruction. In this method, mutual information based strategy is used to construct the 3-dimensional stack of multi-stain and mulit-modal 2-dimensional images.

Different unsupervised feature learning methods have been used in previous studies for image registration. In [8], a stacked convolutional independent subspace analysis (ISA) network was proposed for extracting the features to be fed to the existing registration tool for deformable alignment of the MR images. In another study [9], the authors followed the same framework except that they replaced the ISA features with those learned using a sparse convolutional AE.

3 Materials and Methods

3.1 Experimental Dataset

In this study, all the experiments were conducted on a publicly available dataset [10]. This dataset comprises of 400 images of size 2048×1536 pixels captured at 20× magnification level with the pixel resolution of 0.42 μm. We split the dataset into training and validation sets, consisting of 249 and 151 images respectively.

This data split was used for AE training and to evaluate the registration, same validation set was used.

3.2 Deep Convolutional Autoencoders

A typical vanilla AE is structured to form three layers: an input layer, a dense hidden layer and an output layer. On giving an input image x_n, reshaped into a vector, to this AE, the hidden layer would map this image to vector h_n using an activation function, given as:

$$h_n = f(x_n) = f(Wx_n + b_1) \tag{1}$$

where $f(.)$, W and b_1 represents the activation function, weight matrix and bias vector respectively. The output of this activation function is then given to another mapping function $y_n = f(h_n) = f(Wh_n + b_2)$ which would map the h_n to a vector approximately equivalent to the input image vector x_n. The goal of the AE is to learn the feature representation by generating an output approximately equivalent to the input by using the following function:

$$W, b_1, b_2 = \underset{W,b_1,b_2}{\arg\min} \sum_{n=1}^{N} ||f(W^T(f(Wx_n + b_1)) + b_2) - x_n||_2^2 \tag{2}$$

where N represents the number of images. This type of AE learns feature representation from the input images without considering the spatial information which carries significant information and is also limited due to its shallow nature. These limitations can be avoided by adopting a sliding window based convolutional AE and by adding more hidden layers to the network. In a convolutional AE, the network is designed to form the U-shaped architecture, consisting of an encoder and a decoder. The encoder part consists of a number of convolutions, each followed by an activation function (ReLU) and a max-pooling layer. The decoder part also follows the same structure except that the max-pooling layer is replaced with the deconvolution layer. Here, in this study, the goal of AE is to extract features from the histology images which can be used for the registration purpose. For our experiments, we employed convolutional AE with three convolutional layers in both encoder and decoder part of the network. The network architecture of our AE is shown in Figure 1. For training, RGB patches of size 128×128 were used. During training, sparsity constraint was imposed on the last convolutional layer of the encoder and the features were extracted from the max-pooling layer right after this sparsity imposed convolutional layer. Rmsprop was used to optimize the objective function, with a batch size of 50 input images. Network training was carried out using a system with dual core i5-7500 CPU (3.40 GHz), NVIDIA GeForce GTX 1050 Ti and 32 GB RAM.

3.3 Registration

We posed registration as an optimization problem where the goal is to find a spatial transformation which gives the best correspondence between the two

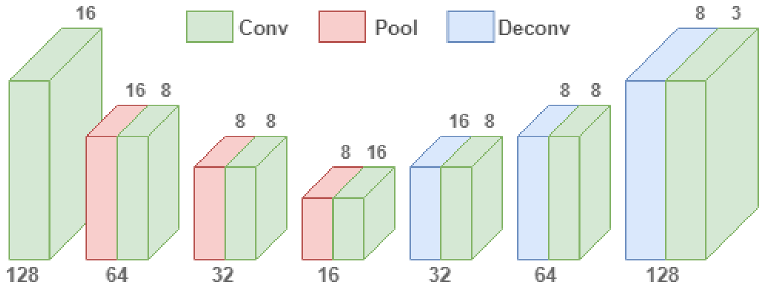

Fig. 1. Architecture of the AE used in this study where each block represents a layer. Each number mentioned above each block indicates the number of feature maps while the number listed below the block indicates the spatial resolution of the output of each layer/block.

images. This optimization problem is solved by using gradient descent. For registration, we use feature values of the images for finding the best transformation which aligns the moving image with respect to the reference image such that the mutual information between the features of two images is maximum. Among intensity-based methods, the use of mutual information has been widely used where the mutual information is computed for image intensities. One of the limitations of using this method when used with the image intensities is that it is often very time-consuming and computational expensive. Here, we adopt this method by replacing the intensity values of the images with deep AE features extracted from the images.

4 Experimental Results

In this section, we evaluate the efficacy of learned features by estimating the transformation of moving images using deep AE features. Figure 2 shows one feature map per image along with the reconstructed image generated by the AE. Since the network was trained on input patch of size 128×128 pixels, these feature maps and reconstructed images were generated by merging the output of small patches. Recall that we used 151 images for evaluating our proposed method. For each of these images, eleven rotated versions of the original image were generated. Features were extracted for the test set images (151) and their rotated versions (1661) using the trained AE. These features were then fed to the registration method based on mutual information. For comparison purpose, we also performed registration based on image intensities and compared its results with our results using root mean square error (RMSE) rate. RMSE is calculated between the ground truth angle and the angle predicted after registration and is formulated as:

$$RMSE = \sqrt{1/n \sum_{j=1}^{n} (y_j - \hat{y_j})^2} \tag{3}$$

where y_j and $\hat{y_j}$ represents ground truth angle and the predicted angle for all the test images. $n = 1661$ which represents the number of test images. The RMSE is calculated for angles in degrees. We found that the features learned are not only performing better in terms of mean square error rate but also outperform the intensity-based registration with a significant margin in terms of computational time. Table 1 gives results for both feature-based and intensity-based registration. For feature-based registration, the spatial correspondence of 75% of the moving images was exactly matched with the reference images while in the case of intensity-based method, only 55.75% of the moving images were exactly matched with their corresponding reference images. Figure 3 shows some good and bad examples of registered images along with the reference and moving images.

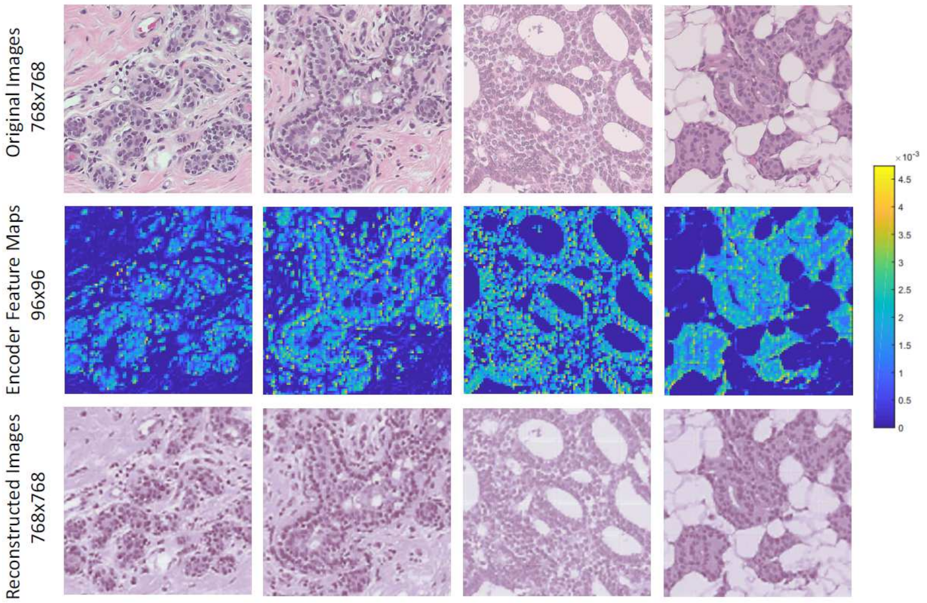

Fig. 2. Feature maps and reconstructed images generated by our trained AE. First, second and third rows show few examples of original images and their corresponding (one of the) feature maps and the reconstructed images respectively. The feature maps and reconstructed images are obtained by merging the output of small patches of size 128×128 pixels.

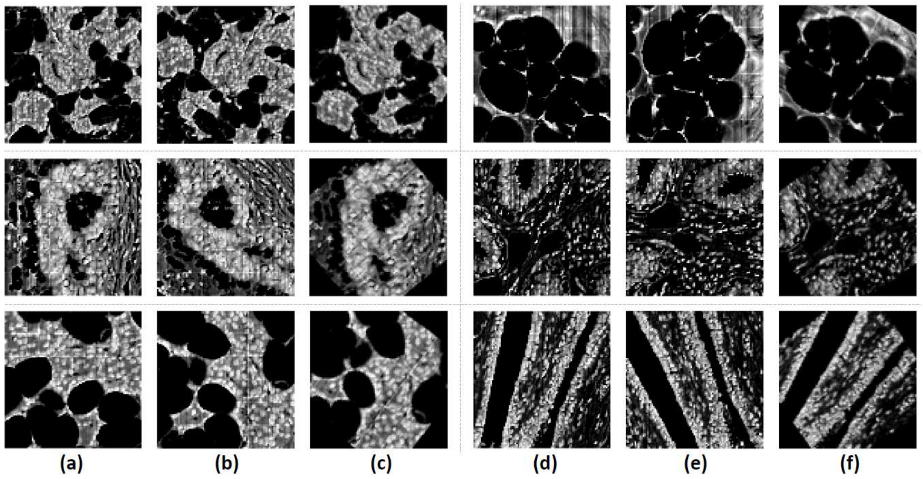

Fig. 3. Example image to show the visual results of some good and bad registration. (a) and (d) show reference images. (b) and (e) show moving images while (c) and (f) show registered moving images.

5 Conclusions

In this study, we have explored the efficacy of features learned in an unsupervised manner for the task of histology image registration. In particular, sparse convolutional AE is designed to extract high level features for this purpose. Using these features as an input to a registration method, we achieved promising accuracy in comparison to using intensity-based registration. In addition, the computational time is significantly lowered by factor of 16 which makes it encouraging for us to use these features for the WSI registration. The continuation of this study would comprise of WSI registration and evaluation of other unsupervised nonlinear models for learning better feature representation.

Method	RMSE	Computational Time (min)
Original Image Intensities	147.42	124.95
Autoencoder Features	104.86	3.05

Table 1. Root mean square error (RMSE) and times required for registration for the whole test set of 151 images. RMSE is calculated between the ground truth angle and the angle predicted by the registration method. The computational time for AE features includes both feature extraction and registration time. Feature extraction and registration for 151 images took 2.53 and 0.52 minutes respectively.

References

1. B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.
2. X. Moles Lopez, P. Barbot, Y.-R. Van Eycke, L. Verset, A.-L. Trépant, L. Larbanoix, I. Salmon, and C. Decaestecker, "Registration of whole immunohistochemical slide images: an efficient way to characterize biomarker colocalization," *Journal of the American Medical Informatics Association*, vol. 22, no. 1, pp. 86–99, 2014.
3. C. R. Meyer, B. A. Moffat, K. K. Kuszpit, P. L. Bland, P. E. Mckeever, T. D. Johnson, T. L. Chenevert, A. Rehemtulla, and B. D. Ross, "A methodology for registration of a histological slide and in vivo mri volume based on optimizing mutual information," *Molecular imaging*, vol. 5, no. 1, pp. 7290–2006, 2006.
4. N. Trahearn, D. Epstein, D. Snead, I. Cree, and N. Rajpoot, "A fast method for approximate registration of whole-slide images of serial sections using local curvature," in *Medical Imaging 2014: Digital Pathology*, vol. 9041, p. 90410E, International Society for Optics and Photonics, 2014.
5. N. Trahearn, D. Epstein, I. Cree, D. Snead, and N. Rajpoot, "Hyper-stain inspector: A framework for robust registration and localised co-expression analysis of multiple whole-slide images of serial histology sections," *Scientific Reports*, vol. 7, no. 1, p. 5641, 2017.
6. D. Mueller, D. Vossen, and B. Hulsken, "Real-time deformable registration of multi-modal whole slides for digital pathology," *Computerized Medical Imaging and Graphics*, vol. 35, no. 7-8, pp. 542–556, 2011.
7. D. Magee, Y. Song, S. Gilbert, N. Roberts, N. Wijayathunga, R. Wilcox, A. Bulpitt, and D. Treanor, "Histopathology in 3d: From three-dimensional reconstruction to multi-stain and multi-modal analysis," *Journal of pathology informatics*, vol. 6, 2015.
8. G. Wu, M. Kim, Q. Wang, Y. Gao, S. Liao, and D. Shen, "Unsupervised deep feature learning for deformable registration of mr brain images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 649–656, Springer, 2013.
9. S. Wang, M. Kim, G. Wu, and D. Shen, "Scalable high performance image registration framework by unsupervised deep feature representations learning," in *Deep Learning for Medical Image Analysis*, pp. 245–269, Elsevier, 2017.
10. T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, "Classification of breast cancer histology images using convolutional neural networks," *PloS one*, vol. 12, no. 6, p. e0177544, 2017.