A Novel Method For Unsupervised Scanner-Invariance With DCAE Model

Andrew Moyes[1]
amoyes01@qub.ac.uk
Kun Zhang[1,2*]
zhangkun_nt@163.com
Liping Wang[1]
liping.wang@qub.ac.uk
Ming Ji[1]
j.ming@qub.ac.uk
Danny Crookes[1]
d.crookes@qub.ac.uk
Huiyu Zhou[3]
hz143@leicester.ac.uk

[1] ECIT, Queen's University, Belfast, UK
[2] School of Electrical Engineering, Nantong University, China (* Corresponding Author)
[3] Department of Informatics, University of Leicester, UK

Abstract

Automated analysis of histopathology whole-slide images is impeded by the scanner-dependent variance introduced in the slide scanning process. This work presents a novel dual-channel auto-encoder based model with a multi-component loss which learns a scanner-invariant representation of histopathology images. The learned representation can be used for a number of histopathology-related applications where images are captured from different scanners such as nuclei detection and cancer segmentation. The approach is validated on a set of lung tissue sub-images extracted from whole slide images. This method achieves a 50% improvement in SSIM score on tissue masks derived from the learned representation compared to related methods. To the best of the author's knowledge, this is the first work which explicitly learns a scanner-invariant representation of histopathology images from multiple domains simultaneously without labelled data or expensive preprocessing techniques.

1 Introduction

Histopathology has long been the gold standard for the diagnosis of different types of cancer [23][14]. This effect has been further compounded by the introduction of computational pathology thanks to advancements in whole-slide image scanners. Computational pathology is a state-of-the-art technology that aims to distinguish and diagnose tissue components (e.g. nuclei, stroma, tumour) which has seen great advancements in recent years due to the development of artificially intelligent diagnosis systems [13].

Whole slide images (WSIs) are produced by colouring tissue sections with staining agents; the most commonly used staining agents are Haematoxylin and Eosin [20]. Haematoxylin binds to genetic material, staining nuclei a deep blue whilst Eosin binds to proteins and stains cytoplasm, collagen and muscle fibres a pinkish hue[8].

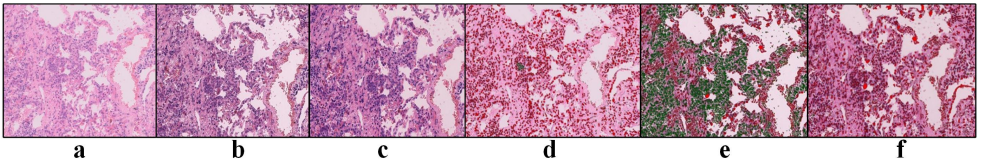

Figure 1: Examples of cell segmentation on images from different scanners. (a,b,c): original data from Philips, Aperio, Hamamatsu. (d,e,f): multi-cell segmentations from a deep learning model

Due to the chemical nature of the staining process, many variables can alter the visual appearance of a given tissue section [3]. The inter-scanner variation is introduced by the type of slide digitiser used which can cause large discrepancies in the illumination, contrast, blur and resolution of the digitised images as well as introduce certain artefacts [10]. Whilst these digital scanners do have parameters that can be modified, they typically remain static for routine scanning procedures and so we do not account for these in this work.

Figure 1 shows examples of inter-scanner variations introduced by different digital scanners for the same tissue specimen. We are using lung tissue specimens scanned by Aperio, Hamamatsu and Philips branded scanners. Take the Aperio image (b) as a reference, the Philips image (a) seems to be brighter and have lower contrast between nuclei and cytoplasm, whilst the Hamamatsu (c) image looks more similar to Aperio, but the pinkish hue structures are less obvious than in Aperio. Images (d,e,f) demonstrate the same multi-nuclei segmentation algorithm applied to the three images. The aim of the algorithm is to segment the nuclei and classify them as benign (in green) or malignant (in red). We found that the three scanners provide very different results, which shows that the state of art automated classification systems struggle with diagnostic decisions under various staining conditions. For instance, training a nuclei segmentation model on images from the Philips scanner would not generalise well when applied to images from the other two scanners.

The existing methods to overcome the variation in histopathology images typically involve attempting to match the input images to some target image as a preprocessing step. The aim of stain normalisation is to find and match the colour space distribution among the different scanners [19]. Our approach is based off the idea that the feature representation of images will be variable under the colour distribution transform from one domain to another even with the same tissue. If we can learn a scanner invariant feature representation we will overcome this issue without requiring domain specific knowledge. Our algorithm employs a dual-channel auto encoder framework to learn filters that are optimized by a novel multi-component loss and are learned to map input data to a stain-and-scanner invariant feature space which means the same features will be obtained from the different types of scanners for the same tissue.

The main contributions of this work are as follows:

- We propose a novel dual-channel model based on auto-encoders to learn a normalised feature representation of samples from different types of scanners which does not require labelled data or expensive pre-processing techniques such as ZCA whitening.
- We propose a novel multi-component loss which forces our dual-channel architecture to learn similar representations for similar tissue regardless of the scanner used to capture the tissue.

The rest of the paper is organized as follows. Section 2 gives a brief introduction about the related works. Section 3 describes the framework and methodology. Section 4 evaluates the method using a lung dataset and compares the algorithm with a state of art method. Finally, in section 5 we present our concluding remarks.

2 Related Works

Existing slide normalisation methods fall into one of three categories: 1) colour matching, 2) stain separation and 3) learning-based techniques. These methods often frame slide normalisation as a preprocessing step to be applied before some subsequent process such as classification or segmentation and usually require the selection of a target image on which the colour distributions of other images (the source images) will be matched. Reinhard et al [15] proposed a colour histogram statistics feature to represent the whole target and source image which is used to match the colour distributions across the images globally. Magee at al [12] extend Reinhard's method by grouping similar pixels with a Gaussian Mixture Model and normalising these groups in the LAB colourspace. Due to the requirement for manual selection of a target image these methods do not perform well when applied to a larger dataset. The selection of target image is troublesome as it must contain a majority of both staining variations and tissue types in order to perform well[21].

To achieve better stain representation, some works utilise the stain separation method which maps images from the RGB space to the stain space (e.g. Haematoxylin, Eosin, DAB) after which stain normalisation can be applied on a channel-to-channel basis. The traditional approach is to find a stain vector by calculating the RGB to optical density space transformation. Khan et al developed this idea to estimate the stain matrix which can map every pixel to the appropriate stain component [6]. Vahadane et al [21] use a sparse stain-separation strategy with colour normalisation to match colour distributions whilst preserving image structure. Despite good results in a number of scenarios, the methods based on stain separation are required to learn a transfer matrix from the template image; however this matrix may not be applicable to the rest of the dataset meaning one may need to calculate a matrix for each image depending on the level of variation found in the dataset.

Learning based methods have been blossoming due to the development of deep learning algorithms. These methods use the feature representation of images instead of pixel-level colour matching. BenTaiab et. al use a GAN based stain transfer model to predict the two domain spaces of the target and source datasets [2]. The main contribution of this idea is to model an entire domain of images instead of a single reference image, however this method relies on annotated training images [1] which can be difficult and expensive to acquire.

3 Method

3.1 Dual Channel Auto-Encoder Architecture

Given two stained and digitally scanned images of the same tissue section, a trained pathologist can easily make decisions about the images regardless of the variance introduced during the scanning process because the general structures within each image remain relatively unaffected. Automated image analysis algorithms do not currently have the same ability to intuitively separate the structure of the tissue from the scanner-specific artefacts.

To overcome this limitation of existing methods we define a dual-channel model which is guided by a clustering mechanism. An overview of our system is shown in Figure 2. Each channel of our system consists of an auto-encoder with identical architecture. Each auto-encoder consists of two modules - an encoder and a decoder. The encoders have two fully

connected layers which map from the flattened input image $x_{i,S}$ or $x_{i,T}$ to a low-dimensional feature vector. We choose two layers as it allows our models to learn higher level abstractions of our data in a low-dimensional space without causing problems with convergence or over-fitting [4]. The decoders map from the low-dimensional feature vector to an reconstruction of the input image. Our choice of fully-connected layers over convolutional layers is motivated by the fact that our input sizes are small, meaning the translational-invariance provided by convolutional layers yielded little benefit to performance whilst increasing the time required to train our model.

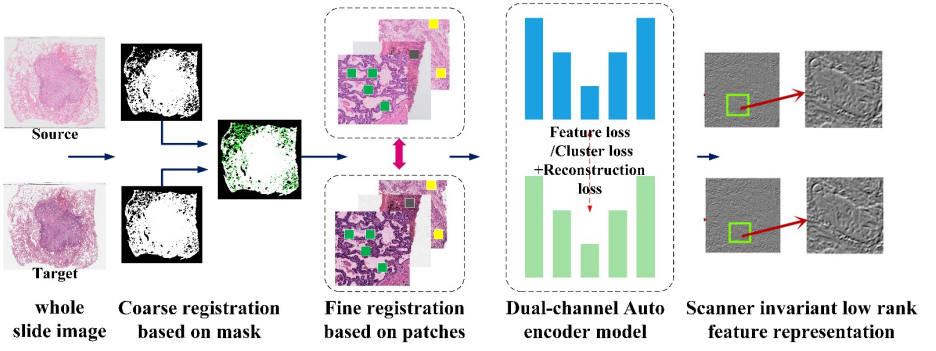

Figure 2: Overview of our alignment, patch extraction and representation learning pipeline

3.1.1 Model Training

Given a dataset of aligned tissue image-patch pairs X where $X_S \in X$ and $X_T \in X$ represent the set of source and target patches respectively, we train both auto-encoders in tandem in such a way that the input $x_{i,S}$ to $g_S(\cdot)$ represents the same tissue as the input $x_{i,T}$ to $g_T(\cdot)$ but captured by a different scanner. These patches are simultaneously fed through the auto-encoders AE-S and AE-T respectively during which we capture the feature representations $z_{i,S}$, $z_{i,T}$ of each input from the encoders as well as the reconstructions $\hat{x}_{i,S}$, $\hat{x}_{i,T}$ from the decoders. We calculate a multi-component loss function of the outputs from each encoder and decoder followed by back-propagation [9] to compute the gradients of our loss with respect to the parameters of each auto-encoder. We then optimise the parameters of our model using the ADAM algorithm [6].

For each epoch of training, we capture the encoded representation $z_{i,T} \in Z_T$ of each i^{th} target patch $x_{i,T} \in X_T$. We then apply the standard k-means method to Z_T, splitting it into separate clusters of similar data points which we use to define the pseudo-labels C_T. These labels are crucial to the success of our system as they allow us to apply the centre loss to the source and target data points which, along with the feature loss increases the similarity of learned representations between the source and target datasets.

3.2 Problem Formulation

Given a set of N pairs of aligned tissue patches, our goal is to learn a compact representation of these patches which is invariant to the scanner each patch was captured from without the use of tissue labels. Auto-encoders are suitable for this task as they are able to produce meaningful representations of input data without labels by learning to reconstruct each input.

The goal of a traditional auto-encoder can be defined like so:

$$\arg\min_{W_E,W_D} \sum_{i=1}^{N} L(x_i, f(g(x_i, W_E), W_D)) \tag{1}$$

Where $g(x_i, W_E)$ is the encoding function parameterised by W_E which maps the input data x_i to a feature vector z_i, $f(\cdot, W_D)$ is the decoder function parameterised by W_D which maps from the feature vector to a reconstruction $\hat{x}_i$ of x_i and L denotes a loss function. The optimisation of this formula therefore involves finding the optimal weights W_E^*, W_D^* such that the loss function L (typically the mean square error) is minimised.

To achieve a scanner-invariant representation of images from two domains, we must ensure that any two images of the same tissue are mapped to the same feature vector. Within the Dual-Channel Auto-Encoder (DCAE) architecture we have two auto-encoders and therefore two encoding functions, denoted $g_s(\cdot)$ and $g_t(\cdot)$ for the source and target domains respectively. Experimentally we observe that in the standard auto-encoder scenario, any pair of aligned patches $\{x_{i,S}, x_{i,T}\}$ from separate domains are mapped to very different feature vectors due to the variance introduced in the slide-scanning process. To enforce similarity between the respective feature vectors $\{z_{i,S}, z_{i,T}\}$ of the inputs, we introduce a regularising term which we call the feature loss ($L_{feature}$) which quantifies the L2-distance between the feature vectors.

Each patch in our dataset may contain any number of different tissue types which leads to blurry boundaries between the clusters in our feature space. This is problematic when the representations $\{z_{i,S}, z_{i,T}\}$ of a pair of aligned images $\{x_{i,S}, x_{i,T}\}$ sit on a boundary between clusters. Discrepancies between their feature vectors can result in them being assigned to separate classes or clusters in some subsequent classification task. Our solution to this is a second regularisation term called the cluster loss ($L_{cluster}$) which is a modified version of the center loss [18]. Different from the traditional center loss which typically requires predefined labels, we sum the L2 distance between the feature vectors of two aligned patches and the cluster centroid which is nearest to $z_{i,T}$. We also achieve this without the use of predefined labels by deriving pseudo-labels through the application of the k-means algorithm to the set of all target feature vectors Z_T. We detail each objective of our system below.

$$\begin{aligned} L_r(x_i) = L_{reconstruction}(x_{i,S}, x_{i,T}) &= \|x_{i,S} - f(g(x_{i,S}, W_{E,S}), W_{D,S})\|_2^2 \\ &+ \|x_{i,T} - f(g(x_{i,T}, W_{E,T}), W_{D,T}))\|_2^2 \qquad (2) \\ L_f(x_i) = L_{feature}(x_{i,S}, x_{i,T}) &= \|g_S(x_{i,S}) - g_T(x_{i,T})\|_2^2 \qquad (3) \\ L_c(x_i) = L_{cluster}(x_{i,S}, x_{i,T}) &= \|g_S(x_{i,S}) - c_{i,T}\|_2^2 + \|g_T(x_{i,S}) - c_{i,T}\|_2^2 \qquad (4) \end{aligned}$$

We then define the overall objective of our system below.

$$\arg\min_{W} \sum_{i=1}^{N} \begin{cases} (1-\lambda)(L_r(x_i)) + \lambda(\max(L_f(x_i), L_c(x_i))) & \text{if } L_r(x_i) < \alpha \\ L_r(x_i) & \text{otherwise} \end{cases} \tag{5}$$

Where W represents the weights of both auto-encoders ($\{W_{E,S}, W_{D,S}, W_{E,T}, W_{D,T}\}$), L_r is the reconstruction loss, L_f is the feature loss, L_c is the cluster loss and α is a threshold hyper-parameter.

The reconstruction loss $L_r(\cdot)$ is the primary supervising signal guiding our model towards learning meaningful representations of each input image. In the early stages of training we can assume that the representations produced by each encoding function will change

drastically in order to produce better reconstructions and therefore reduce the associated penalty. Applying the feature loss and/or cluster loss when the reconstruction loss is high encourages our network to converge towards representations that hold little semantic value; this is undesirable as it can result in non-discriminative representations of our images.

In our example the feature loss may produce stronger gradients than the reconstruction loss and so our network will produce representations which are similar across domains but may hold no semantic value. To counteract this, we employ the loss-balancing approach described in [11] In our experiments we use $\lambda = 0.25$; this means our objective is weighted as 75% reconstruction loss and 25% the other terms.

The feature loss and cluster loss are both defined as the L_2-norm between two points in our proposed feature space defined by the encoding functions; this can be seen in figure 3 where it is visualised as a triangle. Our losses residing in the same space means we can apply them without normalisation, however in our work we balance each feature loss and cluster losses using the $\max(\cdot,\cdot)$. Similar to the λ parameter, this method of combining the losses prevents one term from taking priority over another which can be interpreted geometrically as keeping the triangle in figure 3 as equilateral as possible.

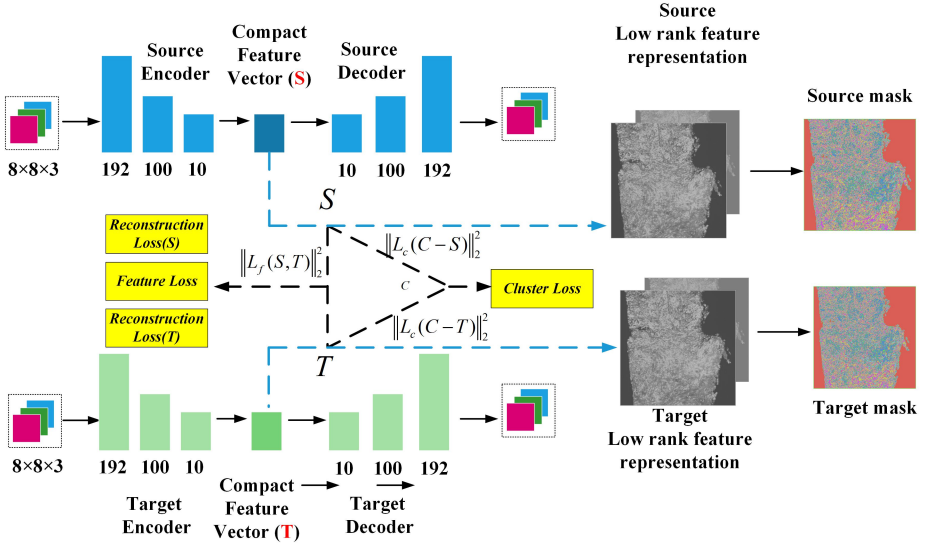

Figure 3: Overview of our DCAE architecture. S and T represent the encoded representation of the source and target images respectively. C represents the cluster centroid closest to T. The triangle depicts how our losses are calculated.

4 Experimental Evaluation

We evaluate the efficacy of our proposed DCAE model for the objective of learning a compact, discriminative, scanner-invariant representation of histopathology images on a privately sourced dataset of 263 sub-image pairs of size 960x960 which are extracted from whole slide images. We compare our method to the most similar state-of-the-art technique called StaNoSA [4] as well as a basic clustering example based on pixel RGB values in the source and target images.

To evaluate the efficacy of our system, we calculate tissue masks from the source and target domains using our DCAE model and then calculate a number of similarity metrics across each mask.

This is achieved by splitting each image into a set of overlapping patches such that every pixel in the tissue mask is a function of the corresponding patch surrounding this pixel in the original image. Given two sets of image patches $\{x_S, x_T\}$, we map each patch to the feature space using the corresponding encoder function, i.e $z_S = g_S(x_S), z_T = g_T(x_T)$. We fit a k-means model on z_T and assign labels to the source and target feature sets based on maximal similarity to one of the k-means cluster centroids.

4.1 Experimental Setup

We implement a bespoke auto-encoder architecture in the PyTorch framework which is used for both the source encoder and the target encoder. Each fully connected layer in our model is followed by the SELU activation function [7] which improves the flow of gradients through our auto-encoders and allows us to omit popular regularisation strategies such as batch normalisation. We optimise the parameters of our model using the ADAM algorithm [8] with a learning rate of 2e-04 for both networks. As no code was available at the time of writing we also implemented the StaNoSA model in PyTorch as faithfully as possible. For each experiment on the DCAE model we use $\alpha = 0.1$ and $\lambda = 0.25$. We set the α parameter based on the experimental observation that when $L_r \leq 0.1$ the reconstructed images are of relatively high quality and the feature vectors Z_S and Z_T change very little at this point,

We run a representation learning experiment with our custom dataset with multiple models and input sizes. For each experiment we train the model from scratch for 250 epochs for approximately 16 hours. We train each model on a machine with an NVidia M60 GPU with 8GB of on-board memory, 126GB RAM and an Intel Xeon CPU.

Each scanner is calibrated using a reference slide in accordance with the manufacturer's specification and image variance was reduced by only digitising slides after a burn-in period of 2 days.

When applying the K-Means algorithm at any point in this paper we select $K = 10$. This choice is based on the resolution of our images; StaNoSA uses images captured at 40x magnification and use $K = 50$ for their experiments whereas our images are captured at 10x magnification and therefore will not contain as much detail; we reflect this change in magnification in our choice of K.

4.2 Dataset

We start with a set of 60 whole slide images representing 30 different lung tissue specimens. Each specimen has been stained once and scanned twice using either Aperio or Philips brand scanners. We match each WSI from the Philips set to the corresponding WSI in the Aperio set such that each pair represents the same tissue specimen. These images are not guaranteed to be aligned. Due to the size of the WSIs (typically 2-3GB) our first challenge is to split these WSI pairs into a set of smaller, coarsely-aligned images to facilitate more convenient processing of the data.

The standard U-Net model [16] is applied to each WSI pair to segment the background and tissue classes which then form binary tissue masks. Affine registration is then applied to align each pair of masks and the discovered transformation is applied to each original image to form a set of coarsely aligned WSI pairs. Each pair is then split into smaller sub-images of size 960x960x3.

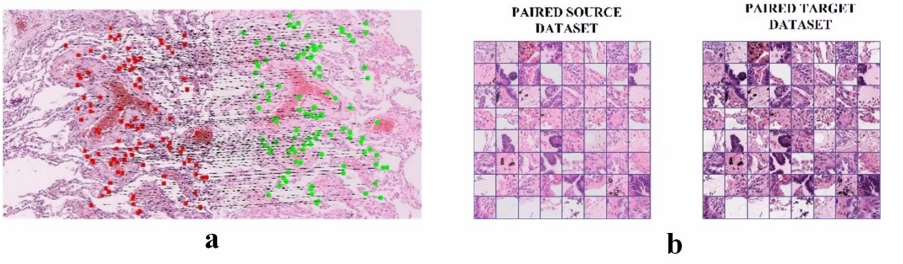

Figure 4: Overview of our ORB-based patch alignment and extraction process. Left: ORB features are calculated matched across two images of the same tissue. Right: a patch is extracted from the corresponding points of each match and added to the source and target datasets respectively.

We then derive a set of aligned image patch pairs X using ORB feature matching [17]. We compute and match ORB features across each sub-image pair and extract a small patch around each corresponding point. Applying the patch-extraction process to each sub-image in our dataset we produce approximately 300,000 pairs of finely aligned patches for the training dataset. See Figure 4.

4.3 Comparison

Given two masks of the same tissue, we wish to calculate the similarity between them. This is made difficult by the misalignment in our custom dataset. Therefore, to make comparisons between masks we use a similar approach to our patch extraction method; we compute and match ORB features between the source and target images and extract patches from each mask corresponding to the matched points. We are then able to measure similarity between each matched patch without having to perfectly align each image globally. For every matched patch we calculate mean squared error (MSE), structural similarity (SSIM) [22] and a histogram-based measure of similarity called SSD (see [4] for a definition of SSD).

MSE gives us an absolute, pixel-wise measure of similarity between the two patches. SSIM gives us a similar measure of similarity except that it is more robust to the effects of noise and translation whilst SSD provides a global measure of similarity based on histograms.

4.4 Experiment: Evaluation of Scanner Invariance

Our aim in this experiment is to evaluate whether the DCAE model can represent two images of the same tissue captured by different scanners with greater similarity than the existing StaNoSA method. To examine the efficacy of our proposed system over raw colour information we provide a baseline called 'pixel' which is simply the K-Means algorithm applied to raw pixel values. We apply the DCAE, StaNoSA and pixel methods to the custom dataset.

4.4.1 Results

We first discuss the quantitative results of our experiment with respect to each model. Table 1 shows the averaged performance achieved by each model with respect to the ORB-aligned MSE, SSIM and SSD scores. When discussing our models, we use 'F' and 'FC' to refer our model using just the feature loss and the combined feature-cluster loss respectively. We see

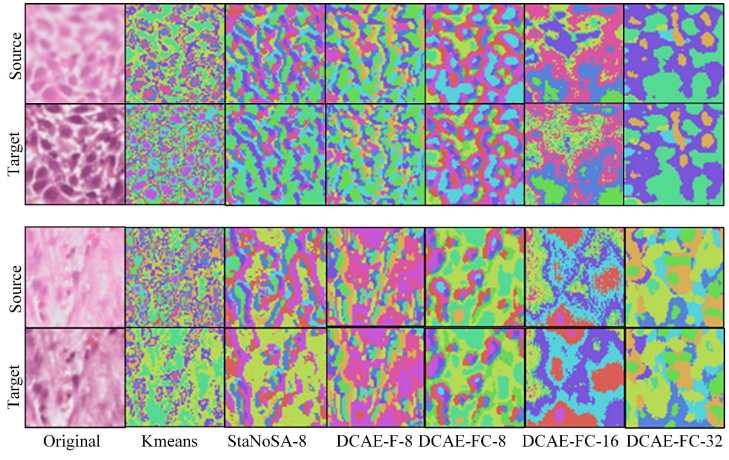

Figure 5: Two pairs of matched patches and their corresponding learned representations with each method. The top row in each group represents the source patch and each representation. The bottom row of each group represents the target image and each representation.

that the FC model is able to achieve a 0.097 increase in SSIM score in the custom dataset, a 16% improvement in MSE score and a 23% reduction in SSD score compared to StaNoSA. Our F model improves on these results further with a 0.29 increase in SSIM score, a 51% improvement on MSE and a 84% improvement in SSD. Our ORB-based metric calculations provide a measure of similarity across the two images but they are biased towards areas of the images which are deemed 'interesting' due to the nature of the ORB feature extraction algorithm; this means that areas with relatively little variation are not well represented in our calculations. We observe that the DCAE model performs significantly better in these low-interest image areas such as the background but this is not reflected in our results.

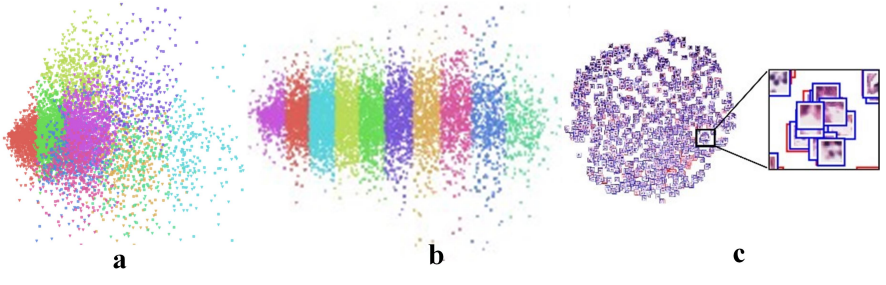

Figure 6: Visualisations our DCAE feature space with (a) feature loss only and (b) feature loss and cluster loss. (c) TSNE visualisation with image patches. Each colour in (a,b) represent a psuedo-label defined by the k-means algorithm. Source and Target data points are represented by squares and triangles respectively. (Best viewed in colour).

Next we discuss the qualitative, visual results of the various models. A visual example of the representation produced by each model for the same aligned image patches can be seen in Figure 5. We see a notable improvement in tissue mask similarity when using the DCAE model over StaNoSA. There is a clear increase in structural similarity as well as an increase in the level of detail available in the image. For example, the DCAE-FC model appears to

	F-8	FC-8	FC-16	FC-32	Pixel	STANOSA-8
MSE	0.030979	0.052535	0.047547	0.028316	0.086439	0.062901
SSIM	0.611349	0.438763	0.421607	0.611243	0.178944	0.324174
SSD	0.000085	0.000454	0.000218	0.000125	0.001637	0.000590

Table 1: Tissue-mask similarity results for the different model. F=DCAE with feature loss only, FC=DCAE with feature-cluster combination. Number after name indicates patch size.

segment nuclei with a high degree of accuracy whereas many of these details are lost in the StaNoSA and RGB-pixel methods. We hypothesise that the increased similarity between the masks using the DCAE-FC model is due to the combined feature-cluster loss which forces our system to produce consistent features regardless of the image domain whilst the increased level of detail is because we use two auto-encoders instead of just one. Allowing each network to specialise in representing a single domain of images, we are able to learn more detailed representations of each domain as the models do not have to account for scanner-related variances as they would with a single model.

To further evaluate the effect our combined feature-cluster loss has on our learned representation, we apply principal component analysis (PCA) to a batch of 4096 image patches for the DCAE-F (a) and DCAE-FC (b) models respectively which can be seen in Figure 6. We observe a significant level of overlap between the clusters when using only the feature loss (a) which is not seen when using the combined feature-cluster loss which exhibits clear separation between each cluster.

5 Conclusion

The performance of automated diagnostic systems for histopathology images is hindered by the significant variation across data sources due to differing staining and scanning methods. Existing works tackle this problem with colour normalisation techniques. In this paper we show that colour normalisation can be avoided entirely with an unsupervised dual-channel representation learning model that learns similar representations of the same tissue across different domains that can easily be adapted to include classification and segmentation type tasks. Although our proposal demonstrates promising results, it should be noted that an auto-encoder is required for each scanner which could lead to problems when running the system at scale. Future work will entail learning a representation that is invariant to both stain and scanner variation with greater robustness and testing our scanner-invariant representation on the publicly available MITOS-ATYPIA and GLAS datasets for representation learning and classification tasks.

6 Acknowledgements

This work was financially supported by Invest NI, the Natural Science Foundation of Jiangsu Province, China under Grant No. BK20170443 and the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province, China under Grant No. 17KJB520030. H. Zhou has been supported by UK EPSRC under grant EP/N011074/1 and the Royal Society-Newton Advanced Fellowship under Grant NA160342.

References

[1] Neslihan Bayramoglu, Mika Kaakinen, Lauri Eklund, and Janne Heikkila. Towards Virtual H&E Staining of Hyperspectral Lung Histology Images Using Conditional Generative Adversarial Networks. In *2017 Ieee International Conference on Computer Vision Workshops (iccvw 2017)*, page IEEE; IEEE Comp Soc, 2017. doi: 10.1109/ICCVW.2017.15.

[2] Aïcha BenTaieb and Ghassan Hamarneh. Adversarial Stain Transfer for Histopathology Image Analysis. *IEEE Transactions on Medical Imaging*, 37(3):792–802, 2018. ISSN 1558254X. doi: 10.1109/TMI.2017.2781228.

[3] Francesco Ciompi, Oscar Geessink, Babak Ehteshami Bejnordi, Gabriel Silva De Souza, Alexi Baidoshvili, Geert Litjens, Bram Van Ginneken, Iris Nagtegaal, and Jeroen Van Der Laak. The importance of stain normalization in colorectal tissue classification with convolutional networks. In *Proceedings - International Symposium on Biomedical Imaging*, pages 160–163. IEEE, 2017. ISBN 9781509011711. doi: 10.1109/ISBI.2017.7950492.

[4] Andrew Janowczyk, Ajay Basavanhally, and Anant Madabhushi. Stain Normalization using Sparse AutoEncoders (StaNoSA): Application to digital pathology. *Computerized Medical Imaging and Graphics*, 57:50–61, 2017. ISSN 18790771. doi: 10.1016/j.compmedimag.2016.05.003.

[5] Adnan Mujahid Khan, Nasir Rajpoot, Darren Treanor, and Derek Magee. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Transactions on Biomedical Engineering*, 61(6):1729–1738, 2014. ISSN 15582531. doi: 10.1109/TBME.2014.2303294. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6727397.

[6] Diederik P Kingma and Jimmy Ba. Adam: {A} Method for Stochastic Optimization. *CoRR*, abs/1412.6980, 2014. URL http://arxiv.org/abs/1412.6980.

[7] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-Normalizing Neural Networks. In *Advances in Neural Information Processing Systems*, pages 972–981, 2017. doi: 1706.02515. URL http://arxiv.org/abs/1706.02515.

[8] Amal Lahiani, Eldad Klaiman, and Oliver Grimm. Enabling Histopathological Annotations on Immunofluorescent Images through Virtualization of Hematoxylin and Eosin. *Journal of pathology informatics*, 9(1):1, 2018. ISSN 2229-5089. doi: 10.4103/jpi.jpi_61_17. URL http://www.ncbi.nlm.nih.gov/pubmed/29531846{%}0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5841016.

[9] Yann LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–50. Springer, 1998.

[10] Patrick Leo, George Lee, Natalie N. C. Shih, Robin Elliott, Michael D. Feldman, and Anant Madabhushi. Evaluating stability of histomorphometric features across scanner and staining variations: prostate cancer diagnosis from whole slide images. *Journal of Medical Imaging*, 3(4):047502, 2016. ISSN 2329-4302. doi: 10.1117/1.JMI.

3.4.047502. URL http://medicalimaging.spiedigitallibrary.org/article.aspx?doi=10.1117/1.JMI.3.4.047502.

[11] William J. Mack, Andrew F. Ducruet, Peter D. Angevine, Ricardo J. Komotar, Debra B. Shrebnick, Niloo M. Edwards, Craig R. Smith, Eric J. Heyer, Linda Monyero, E. Sander Connolly, and Robert A. Solomon. Deep hypothermic circulatory arrest for complex cerebral aneurysms: Lessons learned. *Neurosurgery*, 60(5):815–824, 2007. ISSN 0148396X. doi: 10.1227/01.NEU.0000255452.20602.C9.

[12] Magee D., Treanor D., Crellin D., Shires M., Smith K., Mohee K. and Quirke P. Colour Normalisation in Digital Histopathology Images. In *Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop)*, pages 100–111. Citeseer, 2009. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.157.5405.

[13] Sidra Nawaz and Yinyin Yuan. Computational pathology: Exploring the spatial dimension of tumor ecology, 2016. ISSN 18727980.

[14] Celeste N Powers. Cancer Cytopathology Enters Its Third Decade. *Cancer Cytopathology*, 125(1):9–10, 2017. doi: 10.1002/cncy.21810.

[15] Erik Reinhard, Michael Ashikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41, 2001. ISSN 02721716. doi: 10.1109/38.946629.

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351:234–241, 2015. ISSN 16113349. doi: 10.1007/978-3-319-24574-4_28. URL http://lmb.informatik.uni-freiburg.de/http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net.

[17] E Rublee, V Rabaud, and K Konolige. ORB : an efficient alternative to SIFT or SURF About local feature and matching Motivation oFAST – Oriented FAST BRIEF (Calonder et al . 2010). In *Intl. Conf. Computer Vision*, pages 1–5. IEEE, 2011. ISBN 9781457711022.

[18] Li Shen, Zhouchen Lin, and Qingming Huang. Relay Backpropagation for Effective Learning of Deep Convolutional Neural Networks. In *European Conference on Computer Vision*, volume 1, pages 499–515. Springer, 2015. ISBN 9783319464787. doi: 10.1007/978-3-319-46478-7. URL http://arxiv.org/abs/1512.05830.

[19] Allison Tam, Jocelyn Barker, and Daniel Rubin. A method for normalizing pathology images to improve feature extraction for quantitative pathology. *Medical Physics*, 43 (1):528–537, 2016. ISSN 00942405. doi: 10.1118/1.4939130.

[20] Riku Turkki, Nina Linder, PanuE Kovanen, Teijo Pellinen, and Johan Lundin. Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin stained breast cancer samples. *Journal of Pathology Informatics*, 7(1):38, 2016. ISSN 2153-3539. doi: 10.4103/2153-3539.189703. URL http://www.jpathinformatics.org/text.asp?2016/7/1/38/189703.

[21] Abhishek Vahadane, Tingying Peng, Amit Sethi, Shadi Albarqouni, Lichao Wang, Maximilian Baust, Katja Steiger, Anna Melissa Schlitter, Irene Esposito, and Nassir Navab. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. *IEEE Transactions on Medical Imaging*, 35(8):1962–1971, 2016. ISSN 1558254X. doi: 10.1109/TMI.2016.2529665.

[22] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[23] Kun Zhang, Danny Crookes, Jim Diamond, Minrui Fei, Jianguo Wu, Peijian Zhang, and Huiyu Zhou. Multi-scale Colorectal Tumour Segmentation Using a Novel Coarse to Fine Strategy. In *Bmvc*, number 1, page 2010, jan 2016. doi: 10.5244/C.30.97.