Multi-channel auto-encoders for learning domain invariant representations enabling superior classification of histopathology images

Andrew Moyes [a,*], Richard Gault [a], Kun Zhang [b,*], Ji Ming [a], Danny Crookes [a], Jing Wang [c]

[a] School of Electronics, Electrical Engineering and Computer Science, Queen's University, Belfast, 18 Malone Road, BT9 6RT, UK
[b] School of Electrical Engineering, Nantong University, Nantong, China
[c] Second People's Hospital of Nantong, China

ARTICLE INFO

ABSTRACT

Domain shift is a problem commonly encountered when developing automated histopathology pipelines. The performance of machine learning models such as convolutional neural networks within automated histopathology pipelines is often diminished when applying them to novel data domains due to factors arising from differing staining and scanning protocols. The Dual-Channel Auto-Encoder (DCAE) model was previously shown to produce feature representations that are less sensitive to appearance variation introduced by different digital slide scanners. In this work, the Multi-Channel Auto-Encoder (MCAE) model is presented as an extension to DCAE which learns from more than two domains of data. Experimental results show that the MCAE model produces feature representations that are less sensitive to inter-domain variations than the comparative StaNoSA method when tested on a novel synthetic dataset. This was apparent when applying the MCAE, DCAE, and StaNoSA models to three different classification tasks from unseen domains. The results of this experiment show the MCAE model out performs the other models. These results show that the MCAE model is able to generalise better to novel data, including data from unseen domains, than existing approaches by actively learning normalised feature representations.

1. Introduction

The introduction of digital whole-slide scanners has given rise to the field of digital histopathology. Alongside digital histopathology, the volume of digital tissue images has grown enormously. In order to process this wealth of data, numerous image analysis algorithms have been produced (Gurcan et al., 2009). However, many of these algorithms are dataset specific, meaning they may not perform well if they are applied to datasets that are significantly different to those for which they were created. Image preprocessing is very important in histopathology due to the inter-domain differences, such as inter-scanner differences and differences in staining protocols.

Although digital whole slide scanners have enabled the production of enormous datasets that are useful for data mining and machine learning, the differences in the underlying capture technology can result in very different colour distributions in the final images even when digitising the same tissue specimen. This can lead to a requirement for specific models to be trained for each scanner. Modern machine learning methodologies such as convolutional neural networks typically capture colour and texture information in the initial layers of the network (Krizhevsky et al., 2012). Different slide scanners can introduce significant variance in colour, therefore a neural network that has been trained only on data from one scanner will likely suffer from performance degradation if applied to data captured using a different type of scanner (Ciompi et al., 2017). Global colour transformations have been used to normalise the colour distributions of images from different scanners but these often introduce artefacts due different proportions of anatomical structures (Janowczyk et al., 2017). This problem could be avoided by calibrating a model based upon data obtained by the same tissue specimen being digitally captured by the different scanners, however this type of data is very rare. Even when this data can be acquired, tissue registration (or image alignment) between scanners can be difficult due to differences in the underlying capture technology that can mean a simple linear translation in the x and y axes is not sufficient to achieve a whole slide alignment.

In this work these challenges are tackled by extending an earlier method developed by the authors called the Dual-Channel Auto-Encoder model (DCAE) (Moyes et al., 2019) to model data that, theoretically, could derive from N domains through the development of the Multi-Channel Auto-Encoder (MCAE) model. The aim of this paper

---

is to describe a model that can learn features from multiple domains so that it can have domain-invariant knowledge that is applicable in unseen domains. The major contributions of this work are:

- The development of a novel auto-encoder model based on dual-channel auto-encoders to learn domain-invariant feature representations from N domains of histopathology images. This methodology enhances the expanding area of classification modelling within digital histopathology to account for inter-scanner (or inter-laboratory) variations that might restrict the generalisation of automated digital histopathology analysis.
- The development and publication of a novel, synthetic dataset in which images of the same tissue have been transformed such that they appear to have been produced under different scanning and staining conditions. This dataset will be useful for the development and evaluation of novel methodologies for analysing histopathology image data that are better able to generalise across image domains.

The rest of this paper is organised as follows: Section 2 provides a review of the relevant literature and an outline of the context in which this work is proposed. Section 3 outlines the methods used to develop the synthetic dataset, which is the first contribution of this paper. The second contribution of this work is the proposed MCAE which is outlined in Section 4 and utilises the synthetic dataset from Section 3. In Section 5 the quality of the synthetic dataset is evaluated in terms of its suitability for use in further experiments. Additionally, the MCAE is evaluated and compared with the StaNoSA model (Janowczyk et al., 2017). The paper concludes with a discussion in Section 6.

2. Background

During the staining process, the individual standards and protocols of each laboratory can lead to deviations in average colour intensity of stains and overall image appearance (Ciompi et al., 2017). Additionally, factors such as ambient temperature, time in staining solution and improper clearing can all affect the end result with respect to tissue staining. In a similar fashion to the inter-scanner variances, these differences in staining conditions can result in dramatically different colour distributions across domains and therefore cause problems for image processing algorithms. A common approach to normalising the stain colours of an image is to first apply stain separation. Given a tissue section stained with hematoxylin and eosin, stain separation aims to decompose the single image stained with hematoxylin and eosin into two images where each contains only a single stain. Global colour transformations can then be applied to each single-stain image separately to normalise the colours to that of another domain Magee et al. (2009). Some stain separation algorithms are limited by the need for expert parameter definition (Macenko et al., 2009a). The more pressing problem is that there are very few datasets available to obtain such information because the process of producing a dataset where each specimen has been stained with hematoxylin and eosin as well as each individual stain is time consuming, meaning normalisation methods based on stain separation can be unreliable.

Deep learning has provided radical improvements in the achievable results of many histopathology-related tasks including cell segmentation, tumour detection, mitosis detection and colour normalisation (Janowczyk and Madabhushi, 2016; Sirinukunwattana et al., 2016; Komura and Ishikawa, 2018). Generative adversarial networks (GANs) (Goodfellow et al., 2014) are a type of deeply-learned model that enable the sampling of novel data points given a dataset of real examples. GANs are comprised of two differentiable models — the generator and the discriminator. The discriminator learns a function that maps observed data points (e.g. images) to scalar values and updates its parameters such that the values assigned to observations sampled from the set of real images are maximised whilst the values

assigned to observations drawn from the generator function are minimised. The generator function approximates the real data distribution by conditioning the observations x on latent variables z. Producing artificial or "fake" samples is achieved by sampling from this conditional distribution $p_{model} = p(x|z)$. The parameters of the generator are updated in the direction which maximises the score assigned to the fake samples; thus an adversarial minimax game is defined where the discriminator tries to minimise the score assigned to samples drawn from the generator and the generator is trying to maximise the score indirectly. Conditional generative adversarial networks (cGANs) (Isola et al., 2017) are a modification of the traditional GAN formulation where the generator receives additional information about the sample to be generated. For example, given a task of generating hand-written digits from the MNIST dataset, the observations are conditioned on some representation of the class y (e.g. a one-hot vector) as well as the latent variable z, i.e. the generator is modelling $p(x|y, z)$. A method for mapping colour appearances between domains of histopathology images called Stain Style Transfer (Cho et al., 2017) used cGANs with an additional feature-preserving loss to enable mapping of stain colour appearance without corrupting the underlying tissue structure of transferred images.

Cycle-consistent generative adversarial networks, or CycleGANs (Zhu et al., 2017) are a type of cGAN but rather than conditioning the generative process on a class label, it is conditioned on an image. These models are useful as they are able to learn transformations between domains of images without requiring the datasets to be aligned or corresponding in any way. CycleGANs have proven highly effective in a number of colour normalisation scenarios within the scope of histopathology images. The StainGAN model (Shaban et al., 2019) uses a CycleGAN to map between domains in the MITOS-ATYPIA dataset (Roux et al., 2014) containing pairs of tissue images where each pair corresponds to images of the same tissue specimen that have been captured using different digital slide scanners. When measuring structural similarity (SSIM) between synthetically transferred images and the corresponding ground truth image, StainGAN was found to be more effective than many traditional methods such as those presented in Reinhard et al. (2001), Macenko et al. (2009b), Khan et al. (2014), Vahadane et al. (2016). This demonstrates that CycleGANs are effective at transferring the colour appearance of images from different scanners. Another CycleGAN-based model called the Transitive Adversarial Network (TAN) (Cai et al., 2019) proposed an extension of StainGAN that used a more efficient generator and was able to map between domains in the MITOS-ATYPIA dataset more accurately whilst using much less time to process each sample. CycleGANs were augmented with self-attention in Zhang et al. (2019) and a structural consistency loss in the Self-Attentive Adversarial Stain Normalisation (SAASN) (Shrivastava et al., 2019) model was found to outperform StainGAN on a large custom dataset of duodenal biopsy slides which originated from one of three sites with significant variation in stain colour distribution across sites.

The DCAE (Moyes et al., 2019) was shown to be effective at learning a normalised representation of digital images of tissue slides that had been captured using two different scanners. However, this learning process required the same tissue specimen be captured using both scanners and then spatially aligned. Such a dataset is expensive and time consuming to produce. In this work, the DCAE model is adapted to model three domains, thereby becoming the MCAE model and its efficacy is evaluated on a novel synthetic dataset. This dataset is generated using a CycleGAN with self-attention; however, unlike many existing methods that use data with relatively low colour variance between domains, the source domains for this dataset are selected to have very high variance between domains. This is beneficial because it enables the evaluation of the MCAE model across a wider variety of tissue image appearances and thus provides insight into the performance of the model on novel data.
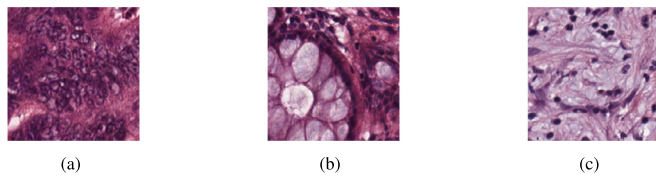
(a) (b) (c)

Fig. 1. Example 128 × 128 images from the NCT-CRC dataset.

(a) (b) (c)

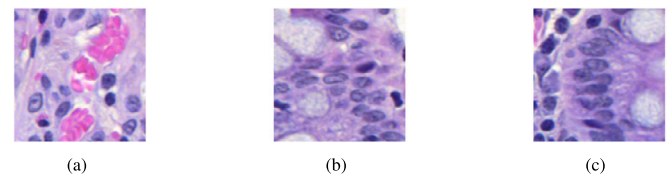**Fig. 2.** Example 128 × 128 images from the TCGA-COAD dataset.

(a) (b) (c)

Fig. 3. Example 128 × 128 images from the NTSPH dataset.

3. CycleGAN generated synthetic dataset

Training and evaluating models that are invariant to the effects of staining and scanning becomes much easier when the problem of image alignment can be mitigated. By using cycle-consistent generative adversarial networks (CycleGANs), tissue images from one domain can be mapped to another whilst the underlying tissue structure remains approximately the same.

This section discusses a method for generating a dataset that is aligned by default using self-attentive, cycle-consistent, generative adversarial networks. A brief introduction of CycleGANs will be presented in Section 3.1, followed by a description of the different data domains used to train the models in Section 3.2. The architecture and training procedure of the models will be discussed in Section 3.3 and, finally, the approach used to generate the synthetic data points is discussed in Section 3.4. The code and dataset discussed in this section has been made publicly available (Andrew et al., 2022a).

3.1. Cycle-consistent adversarial networks

The goal of a CycleGAN is to map an image from one domain to another, for example: horses to zebras, summer to winter or art to photos (Zhu et al., 2017). In the context of this work, a domain will refer to images that have similar appearance characteristics, such as those created using the same staining and scanning protocols and methods. Therefore, mapping from domain A to domain B refers to the process of adjusting the appearance of images in domain A to match the appearance of those in domain B without affecting the underlying tissue structure.

3.2. Data domains

The proposed generated dataset aims to depict the same tissue from a variety of staining and scanning conditions. Therefore the input data used to train the CycleGAN models should come from a variety of sources. In this section, a brief description of these data sources is presented.

NCT-CRC. This dataset (Kather et al., 2018) contains 100,000 non-overlapping tissue image patches stained with hematoxylin and eosin. The dataset is derived from 86 colorectal tissue slides captured at 20× resolution that have been split into 224 × 224 pixel patches and contains a variety of healthy and cancerous tissue. Patches have been divided into a number of anatomical classes, for example, adipose, background, mucosa, smooth muscle and tumour. There are two versions of this dataset, one in which the slides have been normalised using Macenko's slide colour normalisation method (Macenko et al., 2009b) and the other which has not. In this study the dataset with normalisation is omitted in order to preserve as much of the natural variance as possible. Fig. 1 shows example patches from this dataset. The desired output size of the generated dataset is 128 × 128 pixels, thus to transform this dataset to the correct size without affecting the magnification; a 128 × 128 pixel patch is extracted from the centre of each image.
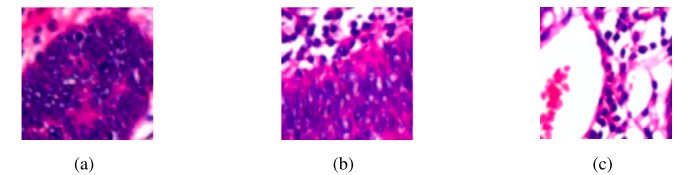
TCGA-COAD. The Cancer-Genome Atlas (TCGA) is a repository of cancer-related data (NIH, 2008). This includes thousands of digital tissue slides from a variety of anatomical locations and of varying diseases. In order to match the approximate anatomical location of the NCT-CRC dataset, the TCGA-COAD (Kirk et al., 2016) project is selected for these purposes as it contains colorectal tissue images under a variety of staining and scanning conditions as well as varying diagnostic status. A single slide was selected that contained a large proportion of clean, undamaged tissue. This slide was downloaded and divided into approximately 20,000 non-overlapping patches, each 128 × 128 pixels in size. The slide was captured at 20× resolution and was selected by manually comparing the colour appearances with the NCT-CRC dataset images and attempting to maximise the colour differences. Fig. 2 shows examples of patches derived from the selected TCGA-COAD slides.

NTSPH dataset. The third dataset used in this study is sourced from the Second People's Hospital of Nantong (NTSPH), China. It contains 18 hematoxylin and eosin stained tissue images of colorectal tissue captured at 20× resolution. The tissue in this dataset demonstrates very strong staining characteristics as can be seen in Fig. 3 alongside high average brightness. Although the tissue type is the same as the NCT-CRC and TCGA-COAD datasets, the variance in staining and capture appearances is large, making this dataset very useful for the purposes of capturing and normalising tissue-image colour variation.

Figs. 1–3 highlight the diverse colour spectra found across the hematoxylin and eosin stained images from different centres. The NCT-CRC dataset heavily features dark purples and reds/purple whilst in contrast the NTSPH dataset is primarily comprised of bright pinks and blue/purple pixels. The colour spectra of the TCGA-COAD data are in stark contrast again featuring more light blues and pinks. Despite all three datasets consisting of colorectal tissue, stained with hematoxylin and eosin, and extracted at the same magnification resolution, the resulting images are distinct to each dataset.

3.3. CycleGAN architecture

The architecture used to generate the synthetic dataset follows the structure defined in Self-Attentive Adversarial Stain Transfer model (Shrivastava et al., 2019). This model is a CycleGAN structure similar to that of StainGAN (Shaban et al., 2019); however each convolutional layer in the generator and discriminator networks are followed by a self-attention layer (Zhang et al., 2019). The self-attention element provides less reliance on local receptive field of neurons than traditional GANs and enables long-range dependencies to be learnt. The self-attention creates a balance between the model's ability to capture long-range dependencies whilst maintaining computational and statistical efficiency. For this work, the generator and discriminator
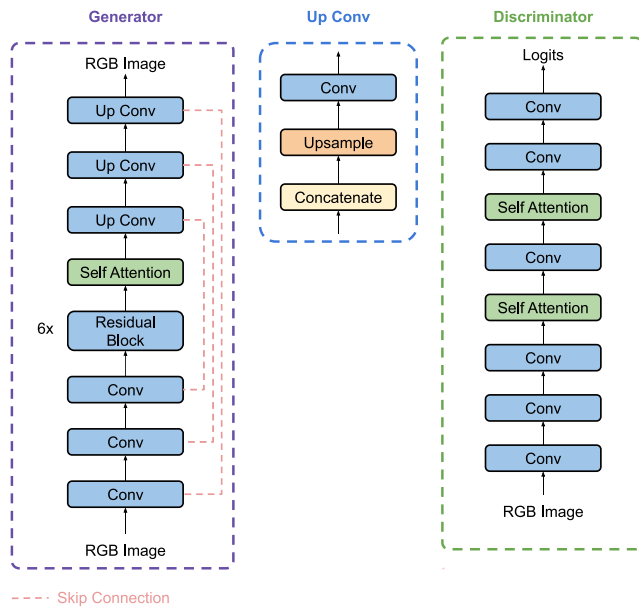
Fig. 4. CycleGAN model architectures.

Fig. 5. Triplet generation: Appropriate generators from each CycleGAN model are selected (depicted as purple circles) and used to map real patches from TCGA-COAD domain (A) to domains for the NCT-CRC dataset (B) and NTSPH dataset (C).

Fig. 6. Example triplets from the synthetic dataset. (a) real images from TCGA-COAD, (b) synthetic images (TCGA-COAD ↦ NCT-CRC), (c) synthetic images (TCGA ↦ NTSPH).

architectures are adjusted to improve performance but in general resemble those defined in Shrivastava et al. (2019). These adjustments include reducing the number of convolutional and self-attention layers as well as changing the number of features learned at each layer. Fig. 4 depicts the architectures used for the generator and discriminator portions of the CycleGAN.

The generator architecture follows the U-Net encoder–decoder pattern with skip connections between corresponding layers. Input images of size 128×128 pixels are downsampled by three convolutional layers, each of which are followed by spectral normalisation and the ReLU activation function. This results in a $16 \times 16 \times 256$ feature map that is subsequently processed by 6 residual layers (He et al., 2016), leaving the spatial and channel dimensionality of the feature maps the same. Finally, three transpose convolutional layers upsample the feature maps, doubling the spatial dimensionality at each layer. The final output of the generator is an RGB image of 128×128 pixels in size.

The discriminator model receives 128×128 pixel RGB images as input. The discriminator is comprised of 6 convolutional layers each with the ReLU activation function and spectral normalisation. Self-attention layers are added after the third and fourth convolutional layers as these have relatively low spatial dimensionality (16×16 pixels).

For the work in this paper, these models are slightly adjusted to improve memory efficiency by only adding self-attention layers after convolutional layers with relatively small spatial dimensions. This still allows for the benefits of capturing the long-range spatial dependencies via the self-attention layers but avoids the large memory costs incurred when applying self-attention to feature maps with large spatial dimensionality.

3.4. Data generation

The final dataset X produced by the CycleGAN is comprised of 20,000 triplets of tissue-image patches. Each triplet represents the same tissue structure but with the colour appearance of the domains discussed in Section 3.2. To achieve this, two CycleGAN models are trained independently; the first mapping between the domains of the TCGA-COAD and NCT-CRC datasets and the second mapping between the domains of the TCGA-COAD and NTSPH datasets. Given the trained
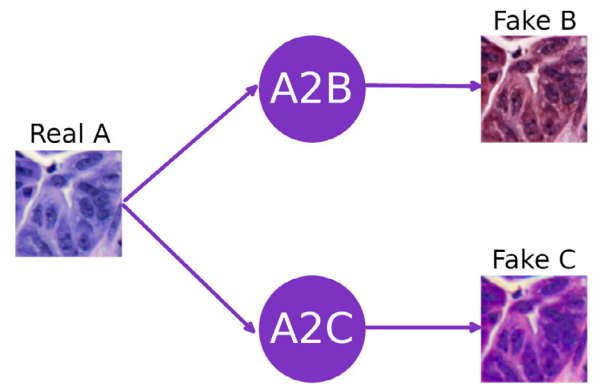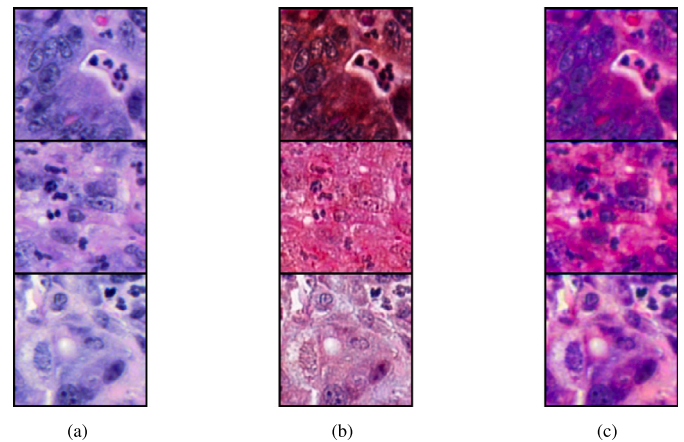
models, the dataset is generated by mapping every real image patch in the TCGA-COAD dataset onto the domains of the NCT-CRC and NTSPH datasets using the appropriate generator from each CycleGAN model. Fig. 5 illustrates the process of generating the dataset. Fig. 6 shows example triplets in the dataset generated through the process illustrated in Fig. 5. Any image patch in X can be referenced using a triplet index i and a domain index $d \in \{A, B, C\}$. For example, $X_{i,A}$ refers to the image patch for domain A in the ith triplet.

4. Multi-Channel Auto-Encoder

This section will outline the methodology behind the proposed Multi-Channel Auto-Encoder model and how it relates to the synthetic dataset established in Section 3. The core idea of the Multi-Channel Auto-Encoder model and Dual-Channel Auto-Encoder model (Moyes et al., 2019) is that there exists some representation of histopathology images where some change in the staining or scanning conditions of an image does not correspond to a change in the representation of the images, i.e. the representation is invariant to scanning and staining parameters. It was shown in Moyes et al. (2019) that the DCAE model is able to learn a representation of histopathology images that is more invariant to scanner-induced colour variance compared to a similar method called StaNoSA (Janowczyk et al., 2017). This was achieved by training a series of auto-encoders, one for each domain and using a combination of a feature similarity loss and a cluster-based loss to maximise similarity between the features of image-patches representing

the same tissue from each of the encoders. In this section, a brief outline of the Multi-Channel Auto-Encoder model is given but more details are available in Moyes et al. (2019). The MCAE model can be decomposed into a set of encoders $F = \{f_A, f_B, f_C\}$ and a set of decoders $G = \{g_A, g_B, g_C\}$.

Following the same process as the DCAE model, the encoding process for the MCAE model involves dividing the ith image patch for domain d, $X_{i,d}$ into a set of overlapping 8×8 sub-patches and applying the encoding function f_d to each sub-patch. The encoding function maps each 8×8 sub-patch into a 10-dimensional feature vector. For the ith triplet $X_{i,:}$, the tensor $\mathbf{Z} \in \mathbb{R}^{3 \times J \times 10}$ represents the encoded feature representation of each image the triplet, where 3 corresponds to domain A, B or C, J is the total number of 8×8 patches that can be extracted from each image and 10 is the size of the feature vectors. A reconstruction of $X_{i,d}$ is formed by applying the decoder function g_d to every feature vector in $\mathbf{Z}_{d,:,:}$.

4.1. Model architecture

Given N disjoint data domains, the MCAE model requires N auto-encoders. These auto-encoders process an image in a patch-by-patch fashion and each of these patches are typically 8×8 pixels. Each 8×8 pixel patch is mapped onto a 10-dimensional feature vector by the encoder and this feature vector is mapped to a reconstruction of the 8×8 pixel patch by the decoder. Each encoder is a multi-layer perceptron with 192 nodes in the input layer (8×8 pixels and 3 colour channels), 100 neurons in the hidden layer and 10 neurons in the output layer. Each neuron in the hidden and output layers use the hyperbolic tangent activation function. The decoders have a symmetrical architecture, with 10 nodes in the input layers, 100 neurons in the hidden layer and 192 neurons in the output layer; however the sigmoid activation function is used on the output of the decoder in order to map back to the range of the RGB space. The specific architecture used for the MCAE model is chosen so as to remain as close as possible to previous work in Moyes et al. (2019), Janowczyk et al. (2017).

4.2. Reconstruction loss

The primary training signal of the MCAE model is the reconstruction loss. For any 8×8 pixel sub-patch, the reconstruction loss quantifies the error between the original sub-patch and the reconstruction of it formed by the auto-encoder. Minimising this reconstruction loss ensures that the features produced by each of the encoding functions in F are information-rich. This is because the encoder functions act as a form of compression while the decoders act as decompression. The only way for the decoders to accurately reconstruct the original sub-patch is if the feature vector has captured key information about the sub-patch. The reconstruction loss for the ith triplet is defined as follows:

$$L_{reconstruction} = \frac{1}{3 \times J \times 192} \sum_{d \in D} \sum_{j=1}^{J} \sum_{k=1}^{192} (x_{d,j,k} - \hat{x}_{d,j,k})^2 \qquad (1)$$

where x is the real image sub-patch and $\hat{x}$ is the reconstructed sub-patch of $X_{i,:}$, $D = \{A, B, C\}$ is the set of domains, j indexes each of the $8 \times 8 \times 3$ sub-patches in each image and k indexes each pixel in the sub-patches, and recalling the number of neurons in the output layer is 192.

4.3. Feature loss

Given that each image patch in the ith triplet represents the same tissue with different colour characteristics, it follows that this is also true for each of the 8×8 sub-patches. In order to learn a representation of images that is invariant to changes in colour characteristics, a feature loss is used. This feature loss quantifies the differences between the encoded feature representations of sub-patches using the mean-squared

error. For the ith triplet, this is calculated as per Eq. (2) (recalling that the size of the feature vectors is 10).

$$L_{feature} = \frac{1}{2 \times 10 \times J} \sum_{j=1}^{J} \sum_{z=1}^{10} (Z_{A,j,z} - Z_{B,j,z})^2 + (Z_{A,j,z} - Z_{C,j,z})^2 \qquad (2)$$

Minimising this loss function alongside the reconstruction loss ensures that the encoded representations in $\mathbf{Z}$ are feature rich (due to minimisation of the reconstruction loss) and colour appearance invariant (due to the feature loss).

4.4. Cluster loss

The centre loss (Wen et al., 2016) has been shown to improve the performance of classification algorithms by bringing the feature representations of separate instances of the same class closer to each other than to the feature representations of instances of other classes. This is problematic in this work as there are no class labels to utilise; therefore these labels are approximated through pseudo-labels that are derived from a K-Means clustering algorithm that has been fitted on randomly sampled feature vectors from patches of images in domain A (TCGA-COAD) and is re-fitted at the end of each training epoch. The cluster loss can be calculated as per Eq. (3)

$$\begin{aligned} L_{cluster} &= \|Z_{A,i,j} - \mu_{i,j}\|_2^2 \\ &+ \|Z_{B,i,j} - \mu_{i,j}\|_2^2 \\ &+ \|Z_{C,i,j} - \mu_{i,j}\|_2^2 \end{aligned} \qquad (3)$$

where $\mu_{i,j}$ is the centroid corresponding to the K-Means determined pseudo-label for $Z_{A,i,j}$ that represents the encoded feature representation of the jth patch of the image for domain A in the ith triplet, and so on for $Z_{B,i,j}$ and $Z_{C,i,j}$.

Both the feature loss defined in Section 4.3 and the cluster loss defined in this section behave similarly, in that they are both explicitly influencing the feature representations of certain patches. The feature loss brings the feature representations of patches that are spatially equivalent within triplets closer together. The cluster loss brings the feature representations of patches that are categorically equivalent closer together. Due to the cluster loss being based on the average of many similar features vectors of patches belonging to the same class, it presents a slower moving, more robust direction in which to move when compared to the feature loss. Following the previous work in Moyes et al. (2019), $k = 10$ for all experiments in this work.

4.5. Combined objectives

The overall objective function of the MCAE model is defined through a summation of the reconstruction loss defined in Section 4.2, the feature loss defined in Section 4.3 and the cluster loss defined in Section 4.4.

$$\arg \min_{W} L = L_{reconstruction} + L_{feature} + L_{cluster} \qquad (4)$$

where W is the set of all model parameters for all auto-encoders.

5. Experimental evaluation

The experimental evaluation in this work is split into three parts. In Section 5.1, the quality of the CycleGAN-based colour normalisation approach is evaluated in order to determine its validity for use in subsequent experiments. Next, in Section 5.2, the similarity between learned feature representations is compared for the MCAE and StaNoSA models. Finally, in Section 5.3, the quality of these learned feature representations is evaluated on three different classification tasks (Kather et al., 2018; Borkowski et al., 2019).

5.1. Synthetic dataset evaluation

In this section, the CycleGAN-generated, synthetic dataset will be evaluated with respect to the quality of colour normalisation between the domains discussed in Section 3.2.

Many existing methods that make use of CycleGAN-based models evaluate their approaches on the MITOS-ATYPIA dataset because it contains images of the same tissue specimens captured by different digital slide scanners. However, the colour distributions between domains in the MITOS-ATYPIA dataset are relatively similar. In this work, the effectiveness of CycleGAN-based models is demonstrated on data with much larger variance in colour distributions due to significant variations in the staining and scanning processes. Previous work (Bejnordi et al., 2015) has used the normalised median intensity (NMI) to measure colour similarity before and after colour normalisation, however this approach requires the use of nuclei segmentation masks; which are not available in this work. Instead, similarity measures based on the hue–saturation–density (HSD) space are used. The HSD transform is the hue–saturation-intensity (HSI) colour transformation applied to optical density images rather than RGB images. Much like the HSI colour space, the HSD colour space transforms tissue images into two colour components — hue and saturation, as well as a density component which represents the underlying tissue structure. The chromatic components of the HSD space are defined by the c_x and c_y attributes, from which the hue and saturation values are derived. The c_x component for each pixel is calculated as per Eq. (5).

$$c_x = \frac{D_R}{D} - 1 \qquad (5)$$

where D_R is the red optical density value at the pixel and D is the average optical density value of the pixels across each colour channel (R, G, and B). The c_y component for each pixel can then be calculated as per Eq. (6).

$$c_y = \frac{D_G - D_B}{\sqrt{3} \cdot D} \qquad (6)$$

where D_G is the pixel-wise green optical density value, D_B is the pixel-wise blue optical density value and D is the pixel-wise average optical density value.

The HSD colour space (Van der Laak et al., 2000) has been used in existing colour normalisation algorithms (Geijs et al., 2018; Bejnordi et al., 2014) where the density channel is fixed and the hue and saturation values are transformed across domains. Therefore the HSD space provides the opportunity for evaluating colour normalisation quality. The HSD-based evaluation strategy is based on the following two criteria:

1. Colour normalisation: a mapping from domain A to domain B will be deemed as successful if the c_x and c_y components (corresponding to hue and saturation) of the synthetic images of domain B (i.e. images from domain A that have been normalised using the CycleGAN generator) match the components of the real images from domain B.
2. Density similarity: throughout the mapping process, the underlying tissue structure of images should remain the same. If the density components of synthetic images roughly match those of the real images, then it can be assumed that the tissue structure has been preserved.

5.1.1. Colour normalisation

The quality of colour normalisation achieved by the CycleGAN is determined through analysis of the c_x and c_y components of the HSD space corresponding to hue and saturation respectively. As discussed in Section 5.1, the CycleGAN-based colour normalisation from a source domain (e.g. the domain of the TCGA-COAD dataset) to a target domain (e.g. the domain of the NCT-CRC dataset) will be deemed successful if the distribution of hue and saturation components for the synthetic
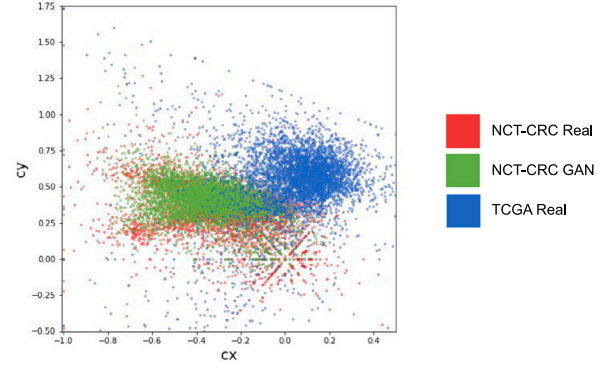
(a) c_x and c_y components of pixels randomly sampled from real images in the from the TCGA-COAD dataset (blue), real images in the NCT-CRC dataset (red) and synthetic images that have been mapped from the domain of the TCGA-COAD dataset to the domain of the NCT-CRC dataset using a CycleGAN (green).
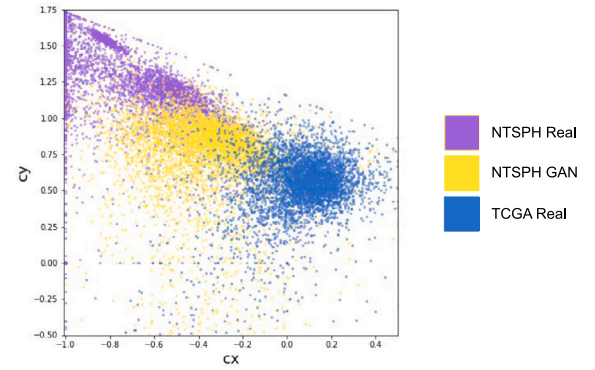
(b) c_x and c_y components of pixels randomly sampled from real images in the TCGA-COAD dataset (blue), real images in the NTSPH dataset (purple) images and synthetic images that been mapped from the domain of the TCGA-COAD dataset to the domain of the NTSPH dataset using a CycleGAN (yellow).

Fig. 7. Distributions of c_x and c_y components of the real and synthetic data distributions.

output images (e.g. the domain of the TCGA-COAD dataset $\mapsto$ the domain of the NCT-CRC dataset via CycleGAN generator) matches the real target domain more than the real source domain.

In Fig. 7(a), pixels are randomly sampled from images in the real domains of the NCT-CRC and TCGA datasets and finally images from the synthetic 'NCT-CRC GAN' domain. The 'NCT-CRC GAN' domain refers to real images from the TCGA domain that have had their appearance normalised to match the domain of the NCT-CRC dataset using the CycleGAN generator. These randomly sampled pixels are plotted in Fig. 7(a) according to their corresponding c_x and c_y components. The significant overlap of the red and green data-points in Fig. 7(a) demonstrate that the distribution of c_x and c_y components in the 'NCT-CRC GAN' domain match the 'NCT-CRC Real' domain much more closely than the original TCGA-COAD domain. This suggests that the CycleGAN architecture described in Section 3.3, is able to successfully map the colour appearance of images between domains with very different staining conditions.

A similar approach has been taken to produce Fig. 7(b). However, in this figure, the pixels are randomly sampled from the real NTSPH domain, the real TCGA domain and finally the synthetic 'NTSPH GAN' domain, which similarly refers to images from the real TCGA-COAD dataset that have had their colour appearance mapped using a CycleGAN generator to match the colour appearance of the real

Table 1
Mean and standard deviation of SSIM values calculated between the density components of source and target domains within triplets.

| Source	Target	Mean	Std
A	B	0.818682	0.119560
A	C	0.852628	0.047245
B	C	0.865696	0.055421

NTSPH domain. It can be seen in Fig. 7(b) that the transformation of component distributions from the TCGA to NTSPH domains has not been as successful as with the NCT-CRC domain in Fig. 7(a). Examining Fig. 7(b) shows that the real NTSPH domain is extreme in both c_x and c_y values. This observation concurs with the visual properties of the NTSPH dataset seen in Fig. 6. For example, the images from the NTSPH domain are largely pink and purple which reside near an extreme of the unravelled hue line in the HSD space, and therefore have extreme values in the c_x axis. Additionally, the colours observed are 'strong' and so have high saturation values seen on the c_y axis. It appears the CycleGAN has struggled to fully translate these values which suggests the model may not be totally sufficient when it comes to the task of normalising tissue image appearance between domains with very large discrepancies in colour distribution. Despite the imperfect mapping of appearance from the TCGA to NTSPH domains, the purpose of this dataset is to present images of the same underlying tissue but with significant variation in their colour appearance. With this in mind, the colour appearance normalisation results appear to be adequate with respect to the hue and saturation components of the HSD space. In Section 5.1.2, the quality of the normalisation process will be evaluated with respect to the similarity of the density component.

5.1.2. Density similarity

Within the synthetic dataset, images of the same tissue should differ only by their colour appearance. The degree to which the CycleGAN-based normalisation procedure has altered the underlying tissue structure can be measured by evaluating the similarity of the density components across domains for images of the same tissue. Ideally, the density component would be the same across domains.

Table 1 shows the average similarity scores of the density channels between domains for structural similarity index (SSIM). In general, the differences in the density channels are relatively low. For example, the SSIM score of 0.852628 between domains A and C suggests a high degree of similarity between the density channels. The SSIM score of 0.818682 between domain A and B suggests there is less similarity between the two domains. These observations are also reflected in the images in Fig. 8 where it can be seen the density images are more similar between domains B (NCT-CRC) and C (NTSPH) than between domains A (TCGA-COAD) and B.

Fig. 8 depicts two real images from the TCGA domain followed by corresponding synthetic images. Each image is accompanied by a visualisation of its density channel. Visually, it can be seen that the density channels have a very similar structure, implying the overall structure is relatively consistent. However, there are some obvious differences in the overall intensity, for example, the synthetic images from the NCT-CRC domain appear brighter than the TCGA and NTSPH domains; especially in areas containing nuclei. This suggests the overall generative process has not entirely preserved the tissue structure. Despite these differences, the overall structure of the density channels remain similar. Combining this with the good colour normalisation results shown in Fig. 7 and the example triplets in Fig. 6 leads to the conclusion that the CycleGAN-generated synthetic dataset has achieved its primary purpose of depicting the same tissue specimens from a variety of stain colour appearances.
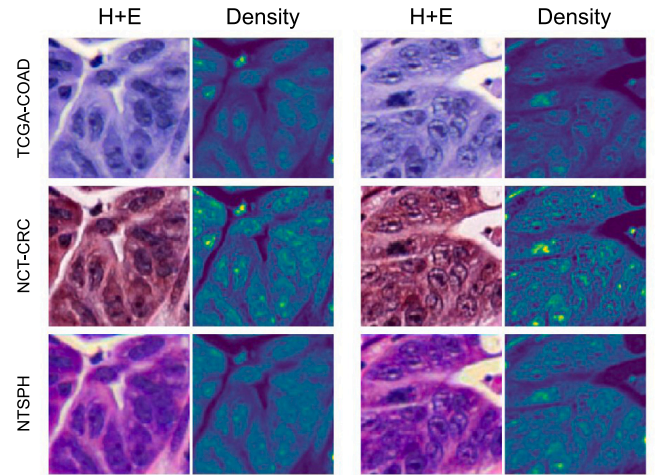
Fig. 8. Example triplets from the synthetic dataset ('H+E' columns) and their corresponding density components ('Density' columns). The 'TCGA-COAD' images are real, the 'NCT-CRC' and 'NTSPH' images are synthetically generated using the CycleGAN and the real 'TCGA-COAD' image. Visually, the tissue structure appears to have been mostly preserved in the 'H+E' images, but the 'Density' images suggest that some degradation of tissue structure has occurred during the normalisation process.

5.2. Domain invariance

This section explores the efficacy of the MCAE and StaNoSA models on the CycleGAN-generated synthetic dataset. The performance of the MCAE model will be evaluated with respect to how much the feature maps vary across the domains, where a low variance is desirable. For these experiments, the synthetic dataset is divided into train and test sets following an 80%–20% split, resulting in approximately 16,000 and 4,000 image-patch triplets for each set respectively. In this section, domain A refers to the TCGA-COAD dataset, domain B refers to the synthetic NCT-CRC dataset and domain C refers to the synthetic NTSPH dataset.

5.2.1. Experimental setup

Both the MCAE and StaNoSA models are trained for 300 epochs using the Adam optimiser (Kingma and Ba, 2014) using a learning rate of 0.0002. For the MCAE model, three auto-encoders are defined and each is trained in tandem using image triplets from the training subset of the synthetic dataset described in Section 3. This results in three trained auto-encoders, one for each domain in the synthetic dataset. For the StaNoSA model, each domain in the synthetic dataset is preprocessed according to the steps outlined in Janowczyk et al. (2017) that includes global contrast normalisation and ZCA whitening of each 8×8 RGB patch. The StaNoSA model is trained on the training subset of the synthetic dataset using the TCGA-COAD domain only. This is in line with the training instructions outlined in Janowczyk et al. (2017) where the auto-encoder is trained on a single domain.

5.2.2. Feature similarity results

Once the training process is complete, the triplets contained in the testing set are mapped to the feature space using the appropriate encoding function for each domain. For each triplet, the Mahalanobis distance is then calculated between each pair-configuration (A and B, A and C, B and C). The distributions of these distances are shown in Fig. 9 which shows that the MCAE model produces feature vectors which are more similar across domains than the StaNoSA model. When comparing features from domain A with those from domain B, the MCAE model achieved an average distance of 2.71 compared to 4.38 for StaNoSA. Comparing domains A and C yields average distances of 2.91 and 4.30 for MCAE and StaNoSA respectively. Finally, comparing domains
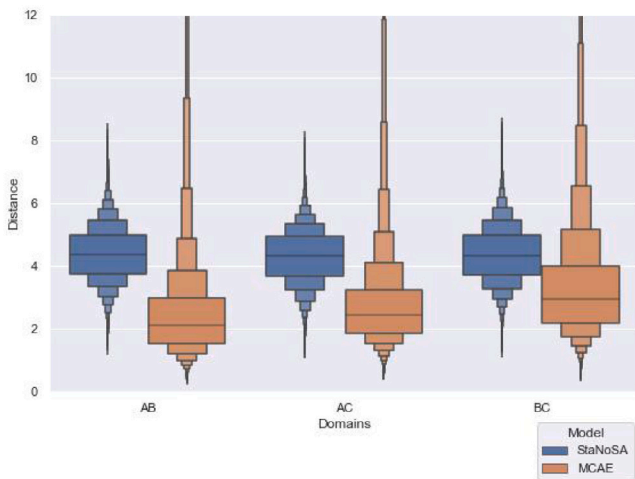
Fig. 9. Distributions of Mahalanobis distances between feature vectors in different domains for the StaNoSA and MCAE models.
Note: the MCAE distributions contain more outliers than the StaNoSA distributions. This figure has been cropped for improved visualisation but all quantities stated regarding these distributions include these extreme values.

B and C gives 3.48 for MCAE and 4.35 for StaNoSA. These results demonstrate a significant improvement in feature similarity when using the MCAE model over StaNoSA. It is likely that this is due to the active-feature normalisation approach taken in the MCAE model where each auto-encoder is actively learning feature representations that are domain-invariant whereas the StaNoSA model relies on global contrast normalisation and ZCA whitening to reduce the impact of the inter-domain variances. It is worth noting that although the average distance between features is lower for the MCAE model, it does create more outlier values in the range [0, 120]. This suggests that the active feature normalisation may introduce some degree of instability in features for certain images.

5.3. Tissue classification

In Section 5.2, it was shown that the MCAE model produces feature representations that are statistically more similar across domains than the StaNoSA model. In this section, the usefulness of the feature representations is evaluated with respect to three separate tissue classification tasks; 1. NCT-CRC dataset with 9 distinct tissue classes (approximately 100,000 images), 2. a lung tissue dataset (Borkowski et al., 2019) with 3 classes (15,000 images), 3. a colon cancer dataset (Borkowski et al., 2019) with 2 classes (10,000 images). These three tasks were chosen as they vary by pathology, origin, and class number. Furthermore, the MCAE, StaNoSA and the DCAE, have never seen data from the colon and lung dataset. This will examine whether a basic domain-invariant understanding can enable the encoders to elicit meaningful features from new, unseen domains.

Three standard classifiers are considered in this experiment, multi-layer perceptron (MLP), support vector classifier (SVC), and XGBoost. These classifiers are chosen as they vary by fundamental methodology and are well established in a variety of classification tasks. Experiments were conducted in Python 3.9 and sklearn's implementation of the classifiers. MLflow was used to manage experimental runs and an overview of the full experimental runs and analysis are available: (Andrew et al., 2022b). Datasets were randomly divided by a 80%–20% training-testing split and repeated through 5-fold cross-validation.

The efficacy of the MCAE, DCAE, and StaNoSA models will be determined by mapping images from the relevant dataset into feature maps using the encoder portion of the MCAE, DCAE, or StaNoSA models as feature extractors. The feature extractor models will be frozen and identically structured classifiers will be trained on top of them using the training subset. Given that the feature extraction models are frozen for this experiment, their performance on this task can be used to estimate how well each model generalises to new data.

5.3.1. Training process

MCAE, DCAE and StaNoSA models are trained on each of the domains in the synthetic dataset described in Section 3. An MCAE model is produced by training on all 3 domains at once. For DCAE, 3 models are produced by training on each possible pair of domains (AB, AC, BC) and for StaNoSA, 3 models are produced by training on one domain at a time. For each trained model, the decoder portions are discarded and the encoders are selected for use as feature extractors. Each model is trained for 100 epochs using the Adam optimiser and a learning rate of 0.0002.

Once the encoders have been trained as per their respective protocols, 3 classification datasets are transformed into each of the respective feature spaces and 3 classifiers are trained on each of the transformed datasets. Datasets are transformed by first dividing each image in the dataset into 8×8 sub-patches, each of which is mapped to the feature space using the relevant encoder. Any preprocessing such as ZCA whitening is applied prior to feature extraction with the encoder. The feature vectors for each sub-patch are summed over each image in the dataset such that each image is represented by a single 10-dimensional feature vector. The classifiers are then trained using these image-level feature vectors as input alongside the relevant ground truth labels.

Note that every available encoder is tested as part of the following experiments. For the MCAE model, although a single model was trained, this model contains 3 encoders (one for each of the domains) and as such, each of these encoders are evaluated separately. Similarly for the DCAE model there are 6 encoders (3 models and 2 encoders per model). Because there is no inter-domain training for the StaNoSA model, only 3 encoders are available. In the results in Section 5.3.2, f1 scores are averaged over the encoders for a given model. This is important as certain encoders may be better or worse suited to novel domains due to correlations in their colour distributions thus giving less robust results. By averaging over each encoder, the results indicate the overall ability of the given model to generalise to new domains.

Given the 12 encoders (3 for MCAE, 6 for DCAE and 3 for StaNoSA), the 3 classification datasets and the 3 classifiers, 108 experiments are done in total, with each experimented being repeated via 5-fold cross-validation.

5.3.2. Results

The full experimental results with extended granularity to each run and additional metrics can be found in the GitHub repository (Andrew et al., 2022b). However, mindful of space, the key findings are presented in the main body of this paper. Tables 2–4 present the average f1 score for each classifier starting with the relevant encoder for feature elicitation. MCAE generally performs the best in most of the experiments (with the exception of the MLP model for the lung and colon datasets). The results of this experiment suggest that the features produced by the MCAE model are indeed useful and able to extract useful information about the tissue structures contained in histopathology images. The methodology underpinning in the DCAE and MCAE significantly outperforms the StaNoSA model across all dataset and models.

Given the results from Tables 2–4 are taken as the average over 5-fold cross-validation (and the DCAE further over different domain pairs), the standard deviation for each classifier, encoder, and dataset are presented in Tables 5–7. The results show that in general the MLP, as a classifier, is significantly less consistent in its performance compared with SVC and XGBoost. It is notable that this was the only classifier where either DCAE or StaNoSA were able to outperform MCAE based models (Tables 3–4). In general, the classifiers generally

Table 2
Summary of the f1 score for each classification model different encoders for the NCT-CRC dataset.

Classifier/model type	MLP	SVC	XGBoost
StaNoSA	0.331804	0.377023	0.269767
DCAE	0.616478	0.599723	0.535119
MCAE	**0.620738**	**0.599852**	**0.565263**

Table 3
Summary of the f1 score for each classification model starting with the different encoders for the lung dataset.

Classifier/model type	MLP	SVC	XGBoost
StaNoSA	0.333136	0.528870	0.455735
DCAE	**0.723412**	0.806611	0.799957
MCAE	0.622465	**0.811241**	**0.802298**

Table 4
Summary of the f1 score for each classification model different encoders for the colon dataset.

Classifier/model type	MLP	SVC	XGBoost
StaNoSA	0.442983	0.641454	0.589402
DCAE	**0.614127**	0.797639	0.800078
MCAE	0.580587	**0.800511**	**0.804332**

Table 5
Summary of the standard-deviation of the f1 scores for each classification model over all iterations of the experiment different encoders for the NCT-CRC dataset.

Classifier/model type	MLP	SVC	XGBoost
StaNoSA	0.025390	0.027815	0.032394
DCAE	0.024818	0.032710	0.037174
MCAE	**0.022305**	**0.029236**	**0.007385**

Table 6
Summary of the standard-deviation of the f1 scores for each classification model over all iterations of the experiment different encoders for the lung dataset.

Classifier/model type	MLP	SVC	XGBoost
StaNoSA	**0.053328**	0.012291	**0.003095**
DCAE	0.103813	**0.005103**	0.011062
MCAE	0.193919	0.007058	0.009452

Table 7
Summary of the standard-deviation of the f1 scores for each classification model over all iterations of the experiment different encoders for the colon dataset.

Classifier/model type	MLP	SVC	XGBoost
StaNoSA	**0.114713**	0.032827	0.035709
DCAE	0.122355	0.008490	0.022511
MCAE	0.231736	**0.005873**	**0.004423**

performed the most consistent (and with greatest performance) when starting with the MCAE encoder. This suggests that the MCAE has been able to provide domain invariant feature representation due to its multi-domain learning.

It is important to reminder ourselves that although Table 6 shows that StaNoSA was quite consistent, the modelling performance was generally quite poor (f1 score between 0.33 and 0.53) whilst the MCAE was relatively more consistent than DCAE whilst achieving very high performance (f1 score between 0.62–0.81) as shown in Table 3.

6. Conclusion

This work presented the Multi-Channel Auto-Encoder which is a novel extension of the Dual-Channel Auto-Encoder model that makes use of more than two data domains. Additionally, this work demonstrated that the active feature normalisation approach used by the MCAE model is superior to other statistical normalisation methods like ZCA whitening. The multi-domain knowledge garnered by the

MCAE enables strong feature representation even from unseen domains. The DCAE has less interdomain knowledge after training and would primarily be useful if the two domains were known and well established *a priori*. However, in the context of enabling mass digital image analysis on histopathology images across multiple centres world-wide, the MCAE provides great stability and generalisability.

A limitation of the MCAE and DCAE models is that they require aligned data from multiple domains in order to learn normalised feature representations. In order to alleviate the impact of this limitation, a novel synthetic dataset was generated using a self-attentive CycleGAN based on the work outlined in Shrivastava et al. (2019). This dataset was generated by fitting 2 CycleGAN models to 3 domains of unaligned tissue image data. The trained CycleGANs were then used to produce a novel dataset of aligned tissue image triplets where each triplet represents the same tissue patch but with the staining appearance of other domains.

In Section 5.1, the synthetic dataset was evaluated with respect to the quality of appearance normalisation and suitability for use in further experiments. It was shown in Section 5.1 (Figs. 7, 8 and Table 1) that the CycleGAN models are able to capture and transfer the colour appearance of different tissue domains.

In Section 5.2, the abilities of the MCAE and StaNoSA models to produce domain-invariant feature representations were evaluated on the synthetic dataset. Using the Mahalanobis distance metric, the MCAE model is able to produce features that vary much less across domains than the StaNoSA model. This is likely due to the active feature normalisation approach in the MCAE model that uses feature-based loss functions to actively encourage the learning of normalised feature representations of images as opposed to the StaNoSA model, which makes use of preprocessing techniques such as global contrast normalisation and ZCA whitening in order to reduce the impact of inter-domain variances.

The MCAE's primary focus is directed towards the development of a stain invariant system. The tissue structures within the histopathology images could additionally be characterised by textural information in addition to the colour information. Unlike the MCAE model, CycleGANs use convolutions and therefore account for texture as well as colour. Although the MCAE model can learn textural features, it is likely to be less efficient in doing so compared with a CNN. In the present work the MLP architecture of the MCAE was chosen to allow for a better comparison to the StaNoSA method, allowing us to highlight the benefits of active feature normalisation over traditional approaches like ZCA. Future work may consider adjusting the MCAE model to include convolutions that would likely improve the model's performance and speed.

In Section 5.3, the ability of the MCAE, DCAE, and StaNoSA models to generalise to new datasets and new tasks was evaluated using three different classification tasks. The MCAE, DCAE, and StaNoSA models were used as pre-trained feature extractors, the outputs of which were fed into separate classification modules where MLP, SVC and XGBoost were considered in this work. The results showed that the classifiers that used MCAE as a feature extractor generally outperformed the models using StaNoSA and DCAE with respect to f1-score.

These results suggest that the active feature normalisation approach used in the MCAE model results in features that generalise better to novel tasks. This in particular is present in the fact that the lung and colon datasets did not derive from a domain on which the MCAE was trained. Furthermore, the MCAE provided the foundation for more consistent classification over multiple folds whilst achieving high performance in comparison with the DCAE and StaNoSA. Despite promising results, the MCAE model is limited by the lack of convolutional layers. Explicitly extracting patches and processing them with fully connected layers is slower than using standard convolutions and so the exact architecture used in this paper is not recommended. Instead, the concept of active feature normalisation should be applied to convolutional feature extractors.

Future work will involve adapting the MCAE model to utilise convolutional layers which would allow for greater computational efficiency when processing image data and also present the opportunity to include textural information in the normalisation process. This would allow an improved representation of tissue to be learned that would better account for spatial dependencies. Additionally, it would be beneficial to develop a novel stain augmentation method that can be used to simulate various staining and scanning conditions on-the-fly, thus reducing the need to use CycleGANs to synthesise the data beforehand.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code to produce the models, experiments and synthetic dataset as well as the synthetic dataset itself has been made freely available at the GitHub repositories cited in the work.

Acknowledgements

References

Andrew, M., Richard, G., Kun, Z., Ji, M., Danny, C., Jing, W., 2022a. MCAE dataset. URL https://github.com/admoyes/MCAE_Dataset.

Andrew, M., Richard, G., Kun, Z., Ji, M., Danny, C., Jing, W., 2022b. Multi channel auto encoder. URL https://github.com/admoyes/MCAE.

Bejnordi, B.E., Litjens, G., Timofeeva, N., Otte-Höller, I., Homeyer, A., Karssemeijer, N., van der Laak, J.A., 2015. Stain specific standardization of whole-slide histopathological images. IEEE Trans. Med. Imaging 35 (2), 404–415.

Bejnordi, B.E., Timofeeva, N., Otte-Höller, I., Karssemeijer, N., van der Laak, J.A., 2014. Quantitative analysis of stain variability in histology slides and an algorithm for standardization. In: Medical Imaging 2014: Digital Pathology, vol 9041. International Society for Optics and Photonics, 904108.

Borkowski, A.A., Bui, M.M., Thomas, L.B., Wilson, C.P., DeLand, L.A., Mastorides, S.M., 2019. Lung and Colon Cancer Histopathological Image Dataset (LC25000), https://arxiv.org/abs/1912.12142, arXiv:10.48550/ARXIV.1912.12142.

Cai, S., Xue, Y., Gao, Q., Du, M., Chen, G., Zhang, H., Tong, T., 2019. Stain style transfer using transitive adversarial networks. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 1905 LNCS. pp. 163–172. http://dx.doi.org/10.1007/978-3-030-33843-5_15, arXiv:1910.10330.

Cho, H., Lim, S., Choi, G., Min, H., 2017. Neural stain-style transfer learning using gan for histopathological images. arXiv preprint arXiv:1710.08543.

Ciompi, F., Geessink, O., Bejnordi, B.E., De Souza, G.S., Baidoshvili, A., Litjens, G., Van Ginneken, B., Nagtegaal, I., Van Der Laak, J., 2017. The importance of stain normalization in colorectal tissue classification with convolutional networks. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, pp. 160–163.

Geijs, D.J., Intezar, M., van der Laak, J.A., Litjens, G.J., 2018. Automatic color unmixing of IHC stained whole slide images. In: Medical Imaging 2018: Digital Pathology, vol. 10581. International Society for Optics and Photonics, p. 105810L.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in Neural Information Processing Systems. pp. 2672–2680.

Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., Yener, B., 2009. Histopathological image analysis: A review. IEEE Rev. Biomed. Eng. 2, 147–171.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1125–1134.

Janowczyk, A., Basavanhally, A., Madabhushi, A., 2017. Stain normalization using sparse autoencoders (StaNoSA): application to digital pathology. Comput. Med. Imaging Graph. 57, 50–61.

Janowczyk, A., Madabhushi, A., 2016. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. J. Pathol. Inform. 7.

Kather, J.N., Halama, N., Marx, A., 2018. 100,000 Histological images of human colorectal cancer and healthy tissue. http://dx.doi.org/10.5281/zenodo.1214456, URL http://dx.doi.org/10.5281/zenodo.1214456.

Khan, A.M., Rajpoot, N., Treanor, D., Magee, D., 2014. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. IEEE Trans. Biomed. Eng. 61 (6), 1729–1738.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kirk, S., Lee, Y., Sadow, C.A., Levine, S., Roche, C., Bonaccio, E., Filiippini, J., 2016. The Cancer Genome Atlas Colon Adenocarcinoma Collection (TCGA-COAD). The Cancer Imaging Archive, http://dx.doi.org/10.7937/K9/TCIA.2016.HJJHBOXZ, URL https://wiki.cancerimagingarchive.net/x/YQH-/.

Komura, D., Ishikawa, S., 2018. Machine learning methods for histopathological image analysis. Comput. Struct. Biotechnol. J. 16, 34–42.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1097–1105.

Van der Laak, J.A., Pahlplatz, M.M., Hanselaar, A.G., de Wilde, P.C., 2000. Hue-saturation-density (HSD) model for stain recognition in digital images from transmitted light microscopy. Cytometry: J. Int. Soc. Analyt. Cytol. 39 (4), 275–284.

Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E., 2009a. A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano To Macro. IEEE, pp. 1107–1110.

Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Xiaojun, G., Schmitt, C., Thomas, N.E., Guan, X., Schmitt, C., Thomas, N.E., 2009b. A method for normalizing histology slides for quantitative analysis. In: Proceedings - 2009 IEEE International Symposium on Biomedical Imaging: From Nano To Macro, ISBI 2009. pp. 1107–1110. http://dx.doi.org/10.1109/ISBI.2009.5193250.

Magee, D., Treanor, D., Crellin, D., Shires, M., Smith, K., Mohee, K., Quirke, P., 2009. Colour normalisation in digital histopathology images. In: Proc Optical Tissue Image Analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop), vol. 100. Citeseer, pp. 100–111.

Moyes, A., Zhang, K., Wang, L., Ji, M., Crookes, D., Zhou, H., 2019. A novel method for unsupervised scanner-invariance with DCAE model. In: British Machine Vision Conference 2018, BMVC 2018.

NIH, National Cancer Institute, 2008. The cancer genome atlas. https://www.cancer.gov/tcga/.

Reinhard, E., Ashikhmin, M., Gooch, B., Shirley, P., 2001. Color transfer between images. IEEE Comput. Graph. Appl. 21 (5), 34–41. http://dx.doi.org/10.1109/38.946629.

Roux, L., Racoceanu, D., Capron, F., Calvo, J., Attieh, E., Le Naour, G., Gloaguen, A., 2014. Mitos & atypia. In: Image Pervasive Access Lab (IPAL), Agency Sci., Technol. & Res. Inst. Infocom Res., Singapore. Tech. Rep, 1, pp. 1–8.

Shaban, M.T., Baur, C., Navab, N., Albarqouni, S., 2019. Staingan: Stain style transfer for digital histological images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, pp. 953–956.

Shrivastava, A., Adorno, W., Ehsan, L., Ali, S.A., Moore, S.R., Amadi, B.C., Kelly, P., Syed, S., Brown, D.E., 2019. Self-attentive adversarial stain normalization. arXiv preprint arXiv:1909.01963.

Sirinukunwattana, K., Raza, S.E.A., Tsang, Y.-W., Snead, D.R., Cree, I.A., Rajpoot, N.M., 2016. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Trans. Med. Imaging 35 (5), 1196–1206.

Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A.M., Esposito, I., Navab, N., 2016. Structure-preserving color normalization and sparse stain separation for histological images. IEEE Trans. Med. Imaging 35 (8), 1962–1971.

Wen, Y., Zhang, K., Li, Z., Qiao, Y., 2016. A discriminative feature learning approach for deep face recognition. In: European Conference on Computer Vision. Springer, pp. 499–515.

Zhang, H., Goodfellow, I., Metaxas, D., Odena, A., 2019. Self-attention generative adversarial networks. In: International Conference on Machine Learning. PMLR, pp. 7354–7363.

Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2223–2232.