FEW SHOT LEARNING IN HISTOPATHOLOGICAL IMAGES: REDUCING THE NEED OF LABELED DATA ON BIOLOGICAL DATASETS

Alfonso Medela[1*]*, Artzai Picon*[1]*,*
Cristina L. Saratxaga[1]*, Oihana Belar*[2]*, Virginia Cabezón*[2]*,*
Riccardo Cicchi[3,4]*, Roberto Bilbao*[2]*, Ben Glover*[5]

[1] Computer Vision, Tecnalia Research & Innovation, Derio, Spain
[2] BIOEF, Fundación Vasca de Innovación e Investigación Sanitarias, Spain
[3] National Institute of Optics, National Research Council, Sesto Fiorentino, Italy
[4] LENS, European Laboratory for Non-Linear Spectroscopy, Sesto Fiorentino, Italy
[5] Imperial College, London, UK
[*] corresponding author

ABSTRACT

Although deep learning pathology diagnostic algorithms are proving comparable results with human experts in a wide variety of tasks, they still require a huge amount of well annotated data for training. Generating such extensive and well labelled datasets is time consuming and is not feasible for certain tasks and so, most of the medical datasets available are scarce in images and therefore, not enough for training. In this work we validate that the use of few shot learning techniques can transfer knowledge from a well defined source domain from Colon tissue into a more generic domain composed by Colon, Lung and Breast tissue by using very few training images.

Our results show that our few-shot approach is able to obtain a balanced accuracy (BAC) of 90 % with just 60 training images, even for the Lung and Breast tissues that were not present on the training set. This outperforms the fine-tune transfer learning approach that obtains 73 % BAC with 60 images and requires 600 images to get up to 81 % BAC.

Index Terms— Histopathology analysis, few shot learning, convolutional neural network, domain adaptation, optical biopsy

1. INTRODUCTION

The generalization of Convolutional Neural Networks [1] have revolutionized the field of biomedical imaging analysis [2] demonstrating diagnostic accuracy at expert level. Examples can be found in many different tasks, such as diabetic retinopathy screening [3], skin lesion classification [4], lymph node metastasis detection [5] or histopathology analysis [6, 7] among others.

Although these deep learning methods are capable of obtaining unprecedented performance not achievable by classical machine learning algorithms, they normally require several thousands of well-labeled images to generate robust methods with the appropriate generalization capabilities. Small datasets are usually obtained at clinical settings. Creating these datasets is time consuming and might not be feasible for early clinical studies, for rare diseases or for new imaging modalities.

In this work we take the *few shot learning* approach from a metric learning perspective where new concepts and new representation are learned from few samples. A Deep Siamese Neural Network [8] is adapted to learn class distances from a source domain D_s composed by a multi-class dataset of *colorectal tissue textures* [9]. Then, a few-shot learning scheme is applied into a target domain D_t composed by a dataset of healthy and tumoral samples of *colon, breast* and *lung* tissue where a comparison on the classification performance among the number of training samples is performed.

2. RELATED WORK

Transfer learning problem has been traditionally covered by the use of fine-tuning methods [10] that incorporate the knowledge from a *domain source* D_s into a *target domain* D_t by using a pre-trained network at D_s where its final layers are adapted to perform the task at D_t. These weights are trained for the task at D_t while maintaining the pre-trained weights from D_s followed by a final training over all the network layers. Deep Siamese Neural Networks were adapted for one shot learning by Koch [8] where two identical classification neural networks are joined by their final feature extraction layer. This network is fed by two different images and is optimized by a loss function that tends to pull apart the input images belonging to different classes and pushing together images from identical classes acting as a metric learning loss that maps the distance between similar and different classes.

Euclidean-distance-based loss functions like contrastive loss [11] has played an important role in class distance learning, while contrastive loss uses the absolute distances of the matching pairs and non-matching pairs. Triplet loss [12] considers also the relative difference of the distances by minimizing the distances between an anchor and a positive sample of the same class and maximizing the distance between the anchor and a negative sample of different class being more prone to convergence as the loss performs simultaneous minimization for the positive and negative matching samples.

3. DESCRIPTION OF THE DATASETS

In order to validate the few-shot approach, we have selected two different datasets: an extensive dataset from a source domain D_s that will be used as support dataset for metric learning and a second dataset from target domain D_t that will be used for few-shot learning validation.

3.1. Dataset - D_s

Public dataset [9] from the University Medical Center Mannheim (Germany). Contains tissue samples obtained from low-grade and high-grade primary tumours of digitalized *colorectal cancer* tissue slides. The database is divided into eight different types of textures that are present on the tumours samples: 1. tumour epithelium, 2. simple stroma, 3. complex stroma, 4. immune cells, 5. debris and mucus, 6. mucosal glands, 7. adipose tissue and 8. background, as depicted in Fig. 1. There are 625 image samples per class, producing a total dataset of 5000 image tiles of dimension 150 px x 150 px (74 μm x 74 μm).
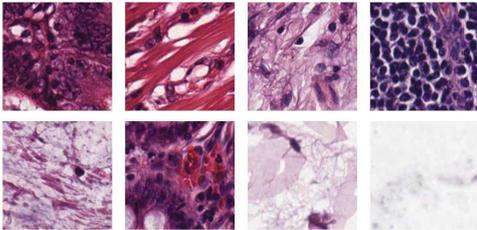

Fig. 1: Sample images from dataset D_s [9]. First row: Tumour epithelium, stroma, complex, immune cells. Second row: debris, mucosa, adipose and empty tile samples are depicted.

3.2. Dataset -D_t

Extensive private dataset with data from the Basque Biobank (BIOEF – Spain) obtained at five local hospitals of the Basque Public Health system-Osakidetza with the collaboration of 10 specialized pathologists. It contains *healthy* and *tumoral* samples from digitized tissue slides of three different organs: *colon, breast* and *lung*. Various types of tumour

are considered: colon adenocarcinoma, breast ductal carcinoma, breast lobular carcinoma, lung adenocarcinoma and lung squamous cell carcinoma. Colon samples are classified into low or high grade, breast samples into grade I, II or III, whereas lung samples are organized into stage IA, IIA, IB and IIB. The database contains a total of 259,425 images tiles of dimension 1300px x 1300px (162μm x 162μm).

The original tissue slides have been annotated by a panel of expert pathologists. According to their criteria, the diagnosis assigned to each slide is the most significant diagnosis encountered within. Since original digitized slides are huge, the same slice can contain parts with healthy, low-grade tumour and high-grade tumour tissue at the same time (plus background, adipose tissue, etc). For this reason, the database includes a reviewed subset of 1755 image tiles which are representative of its associated annotated diagnosis.In order to validate the few-shot approach and ensure the integrity of the data used as input, the experiments performed in this work make use of this subset.

The Biopool Colon, Breast and Lung Hematoxylin-Eosin dataset (B-CBL-HE) request form is available from the Basque Biobank's catalogue web-page [13].
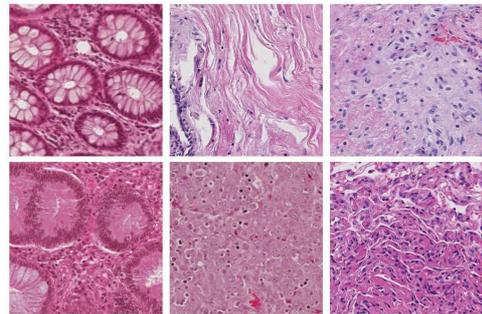

Fig. 2: Sample images from dataset D_t [14]. Upper row are healthy samples and lower row tumoral samples.From left to right: colon, breast and lung tissue are depicted.

4. PROPOSED METHOD

Based on previous research in siamese neural networks [8], we propose to adapt such an architecture to transfer the knowledge gained on a complete and well-labeled dataset (dataset D_s) from a *source domain* to the *target domain* D_t by using just a few images from D_t for training. The hypothesis is that the siamese network will learn feature and distance representations from the eight textures present in tumoral tissues and will transfer that knowledge to classify tumoral and healthy tissues on a more generic domain D_t.

4.1. *Histology Siamese Network*

We train a Deep Siamese Neural Network over dataset D_s to learn its inherent characteristics that map the class distances from the source domain D_s, and to apply it as a feature extractor in a target domain D_t, and use those low-dimensional representations of the images to train a shallow classifier that can perform the classification task at D_t with very few samples.

4.1.1. *Network architecture*

As depicted in Fig. 3, we use a three-headed siamese network, in which all the weights and biases are tied. Network symmetry assures network input commutativity. Three images X^p, X^a and X^n are used as input during training stage. X^a represents an image from a certain class, X^p an image of same class as X^a and X^n an image of a different class than X^a.

A VGG16 [15] network with some modifications is used as base-network. The fully connected layers containing 4096 nodes were removed in order to reduce the number of parameters from 139M to less than 18M to prevent over-fitting by. As suggested by[12], we replace this by adding a single 128 neuron layer to get a embedding layer f_i that acts as a low-dimensional representation of the input images.
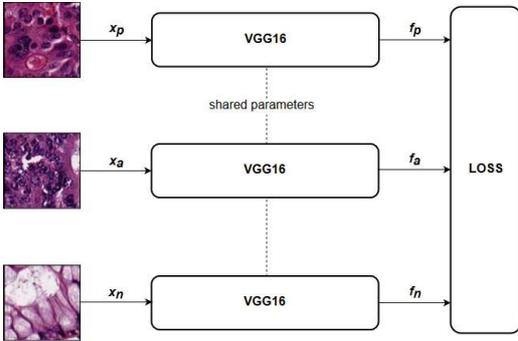

Fig. 3: *Histology Siamese Network* training phase.

As depicted in Fig. 3 , the VGG16 network is able to generate embeddings for each of the input images. During the training stage, we force the embeddings f^p and f^a from X^p and X^a images to be as close as possible while pulling apart f^n and f^a representations from X^n and X^a images. This is achieved by using a non-linear version (1) of the triplet loss function [12] (2).

The reason to use the *Lossless* version (1) of triplet loss is that the original triplet loss annuls the gradient for losses below zero, making more difficult to reach local minima. The *Lossless triplet*, however, captures the information for negative values, and additionally, its non-linearity boosts penalization as the distance grows.

$$\mathcal{L}_{lossless-triplet} = \sum_{i=1}^{N}[-ln(-\frac{||f_i^a - f_i^p||_2^2}{\beta} + 1 + \epsilon) \\ -ln(-\frac{D - ||f_i^a - f_i^n||_2^2}{\beta} + 1 + \epsilon)] \tag{1}$$

$$\mathcal{L}_{triplet} = \sum_{i=1}^{N} max(0, ||f_i^a - f_i^p||_2^2 - ||f_i^a - f_i^n||_2^2 + \alpha) \tag{2}$$

Equations (1) and (2) show these triplet and lossless triplet functions where f_i corresponds to the embedding vector of the anchor,positive or negative sample on the ith triplet on a batch, β and D are parameters set up to 128 (the dimension of the embedding vector) and ϵ is a small number.

4.1.2. *Training procedure*

The siamese network is trained over the source domain D_s dataset by minimizing (1). To enhance convergence during the siamese network training stage, we enforce triplet generation in a way that a triplet is composed by three samples that present high difficulty on classification. This is done by solving a classification problem at the source dataset D_s in first instance and analyzing which categories are confused by computing a confusion matrix where a triplet sampling probability function is obtained.

4.1.3. *Shallow Classifier*

The previously trained siamese neural network at source domain D_s is able to extract, for a image X_i an image embedding f_i. This embedding vector is a low-dimensional representation of X_i that minimizes (1) and thus, is designed to estimate the distance among D_s classes.
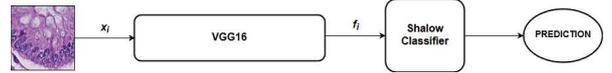

Fig. 4: *Histology Siamese Network* prediction phase.

This shallow classifier is trained from a reduced number of sample images X_s from the target domain D_t. These X_s sample images are described by their feature vector f_i that is obtained from the trained siamese network as depicted in Fig. 4 . f_i set is then used to train a classification Support Vector Machine (SVM) [16] where a linear kernel has been selected.

Given n points of the form

$$(\vec{f}_1, y_1), ..., (\vec{f}_n, y_n)$$

where the y_i are either 1 or -1, each indicating the class to which the point $\vec{f}_i$ belongs. Each $\vec{f}_i$ is a 128-dimensional real vector which is the output of the siamese network. The aim is to find the *maximum-margin hyperplane* that divides the group of points $\vec{f}_i$ for which $y_i = 1$ from the group of points for which $y_i = -1$.

5. RESULTS

5.1. Network training at source domain D_s

In order to validate our proposed few-shot approach, both D_s and D_t datasets were divided into train (80 %), validation (10 %) and test (10 %) sets.

To work on similar scaling, dataset D_s was re-scaled by a factor of 2 (300px x 300px) to match D_t. In the case of dataset D_t, images are cropped into 300px x 300px. At each training interaction, data augmentation such as spatial transformations, Gaussian blur and color transformations [17] have been applied randomly and the resulting image is cropped into 224px x 224px.

The proposed *Histology Siamese Network* is trained over the training set at D_s by minimizing equation (1).

5.2. Learning at target domain D_t with few samples

D_t task is learned sequentially from a different number N of D_t images per class that are obtained from the training subset of the D_t dataset. The selected N images per class are described by the *Histology Siamese network* and their internal representation f_i vectors are used to train a linear SVM that classifies the image. The obtained classifier was tested against the full testing set from the D_t domain. To provide statistically significant results, this process was repeated 50 times for each N value.

5.3. Fine-tuned VGG16 network

For comparative purposes, an identical architecture VGG16 classification network was generated. This network matches with the base network of the Siamese approach by a single fully connected layer of 128 nodes. Pre-trained weights from Imagenet are used for initialization and the last convolutional block and fully connected layers are trained over the source dataset D_s to adjust the domain representation to the histology domain. This fine tuning is followed by a full weight training over the source domain [18].

This baseline VGG16 network last layer is then replaced and fine-tuned over dataset D_t. Following the same procedure as with *Histology Siamese network*, N images per class are used when fine tuning. As training process requires significant time, we trained the network 5 times for each N.

5.4. Discussion

The relation between the obtained accuracy over the testing subset of D_t and the number of images per class are depicted in Fig. 5. It is worth noting that the obtained accuracy of the *Histology Siamese Network* was around 80 % for one single training image per class and it very rapidly increased almost reaching its maximum at around twenty samples per

class. Then reaches a plateau in the upper border of $> 90\,\%$ accuracy.

It is of interest that although source domain D_s was composed only by colon tissue images we have been able to perform few shot learning within a more generic target domain where tissues from colon, breast and lung tissues were present.
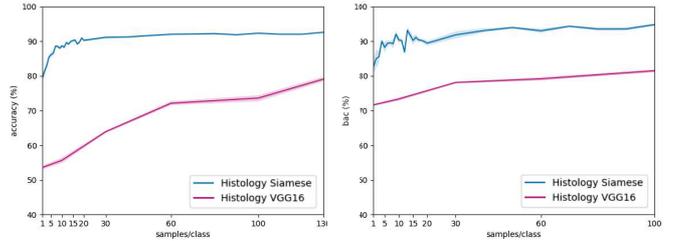

Fig. 5: Accuracy (left) and balanced accuracy (right) of both networks over domain D_t. (*x*-axis: samples per class).

6. CONCLUSIONS

A common problem when developing machine learning-based biomedical applications is the lack of data. In this paper, it has been demonstrated that a *Histology Siamese Network* clearly outperforms traditional fine-tuned neural network when extensive databases are not available. At least 130 samples per class are needed for the fine-tuned network to achieve accuracy 15 % lower than the equivalent siamese. It is notable how the *Histology Siamese Network* surpasses a 90 % accuracy with just 20 samples/class.

We have also validated that it is possible to transfer knowledge from a more restrictive source domain D_s where only colon images were present into a more generic domain D_t where other tissues such as colon, breast and lung are present.

Novel imaging technologies (e.g. OCT, MPT, etc) can provide microscopic information comparable to in-vivo histopathology with a label-free modality. The development of a decision support system for real-time diagnosis via "optical biopsy" is one of the aims of the proposed work. However, in these domains, no extensive datasets are available or can be produced without extensive clinical trials. It is therefore desirable to explore techniques that may reduce the reliance on large datasets for training of deep learning algorithms. Additionally, transfer learning solutions moving from histopathology images to images from these new modalities are feasible with the approach proposed in this article and aimed to be developed in the near future.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Yann LeCun, Patrick Haffner, Léon Bottou, and Yoshua Bengio, "Object recognition with gradient-based learning," in *Shape, Contour and Grouping in Computer Vision*, London, UK, UK, 1999, pp. 319–, Springer-Verlag.

[2] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram Van Ginneken, and Clara I Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88.

[3] Gulshan V, Peng L, Coram M, and et al, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.

[4] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115, 2017.

[5] Ehteshami Bejnordi B, Veta M, Johannes van Diest P, and et al, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, 2017.

[6] Korsuk Sirinukunwattana, Shan E Ahmed Raza, Yee-Wah Tsang, David RJ Snead, Ian A Cree, and Nasir M Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[7] Angel Alfonso Cruz-Roa, John Edison Arevalo Ovalle, Anant Madabhushi, and Fabio Augusto González Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 403–410.

[8] Gregory R. Koch, "Siamese neural networks for one-shot image recognition," 2015.

[9] Jakob Nikolas Kather, Cleo-Aron Weis, Francesco Bianconi, Susanne M Melchers, Lothar R Schad, Timo Gaiser, Alexander Marx, and Frank Gerrit Zöllner, "Multiclass texture analysis in colorectal cancer histology," *Scientific reports*, vol. 6, pp. 27988, 2016.

[10] Geoffrey E Hinton and Ruslan R Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[11] Raia Hadsell, Sumit Chopra, and Yann LeCun, "Dimensionality reduction by learning an invariant mapping," *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 1735–1742, 2006.

[12] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," *CoRR*, vol. abs/1503.03832, 2015.

[13] "Biopool colon, breast and lung hematoxylin-eosin dataset (b-cbl-he)," 2019, http://www.biobancovasco.org/en/CA128-sample-catalog.html.

[14] Bas de Jong, Oihana Belar, Arantza Bereciartua, Artzai Picon, Elena Muñoz, D Sevilla, F Moscone, Fabienne Gandon, E Tosseti, S García, et al., "Services associated to digitalised contents of tissues in biobanks across europe: A proof of concept–biopool," ISBER Annual Meeting, 2013.

[15] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[16] Stefan Knerr, Léon Personnaz, and Gérard Dreyfus, "Single-layer learning revisited: a stepwise procedure for building and training a neural network," in *Neurocomputing*, pp. 41–50. Springer, 1990.

[17] Adrian Galdran, Aitor Alvarez-Gila, Maria Inês Meyer, Cristina López Saratxaga, Teresa Araujo, Estíbaliz Garrote, Guilherme Aresta, Pedro Costa, Ana Maria Mendonça, and Aurélio J. C. Campilho, "Data-driven color augmentation techniques for deep skin image analysis," *CoRR*, vol. abs/1703.03702, 2017.

[18] Brady Kieffer, Morteza Babaie, Shivam Kalra, and Hamid R. Tizhoosh, "Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks," *CoRR*, vol. abs/1710.05726, 2017.