IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 3, 2021, accepted February 18, 2021, date of publication February 22, 2021, date of current version March 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3060926

Autofluorescence Image Reconstruction and Virtual Staining for In-Vivo Optical Biopsying

ARTZAI PICON[1,2], ALFONSO MEDELA[1], LUISA F. SÁNCHEZ-PERALTA[3], RICCARDO CICCHI[4,5], ROBERTO BILBAO[6], DOMENICO ALFIERI[7], ANDONI ELOLA[8], BEN GLOVER[9], AND CRISTINA L. SARATXAGA[1]

[1]Tecnalia, Basque Technology Research Alliance (BRTA), 48160 Derio, Spain
[2]Department of Automatic Control and Systems Engineering, University of the Basque Country UPV/EHU, 48940 Leioa, Spain
[3]Jesús Usón Minimally Invasive Surgery Centre (JUMISC), 10071 Cáceres, Spain
[4]National Institute of Optics, National Research Council, 50019 Sesto Fiorentino, Italy
[5]European Laboratory for Non-Linear Spectroscopy (LENS), 50019 Sesto Fiorentino, Italy
[6]Basque Foundation for Health Innovation and Research, 48902 Barakaldo, Spain
[7]L4T-Light4Tech s.r.l., 50019 Sesto Fiorentino, Italy
[8]Department of Communications Engineering, University of the Basque Country UPV/EHU, 48940 Leioa, Spain
[9]Imperial College London, London SW7 2BU, U.K.

Corresponding author: Artzai Picon (artzai.picon@tecnalia.com)

This work was supported in part by the European Union's Horizon 2020 Research and Innovation Programme under Grant 732111 (PICCOLO project), and in part by the Basque Government's Industry Department through the ELKARTEK Program's Project 3KIA under Grant KK-2020/00049. The work of Andoni Elola was supported by his pre-doctoral research from the Basque Government under Grant PRE_2019_2_0100 and Grant IT1229-19.

ABSTRACT Modern photonic technologies are emerging, allowing the acquisition of in-vivo endoscopic tissue imaging at a microscopic scale, with characteristics comparable to traditional histological slides, and with a label-free modality. This raises the possibility of an 'optical biopsy' to aid clinical decision making. This approach faces barriers for being incorporated into clinical practice, including the lack of existing images for training, unfamiliarity of clinicians with the novel image domains and the uncertainty of trusting 'black-box' machine learned image analysis, where the decision making remains inscrutable. In this paper, we propose a new method to transform images from novel photonics techniques (e.g. autofluorescence microscopy) into already established domains such as Hematoxilyn-Eosin (H-E) microscopy through virtual reconstruction and staining. We introduce three main innovations: 1) we propose a transformation method based on a Siamese structure that simultaneously learns the direct and inverse transformation ensuring domain back-transformation quality of the transformed data. 2) We also introduced an embedding loss term that ensures similarity not only at pixel level, but also at the image embedding description level. This drastically reduces the perception distortion trade-off problem existing in common domain transfer based on generative adversarial networks. These virtually stained images can serve as reference standard images for comparison with the already known H-E images. 3) We also incorporate an uncertainty margin concept that allows the network to measure its own confidence, and demonstrate that these reconstructed and virtually stained images can be used on previously-studied classification models of H-E images that have been computationally degraded and de-stained. The three proposed methods can be seamlessly incorporated on any existing architectures. We obtained balanced accuracies of 0.95 and negative predictive values of 1.00 over the reconstructed and virtually stained image-set on the detection of color-rectal tumoral tissue. This is of great importance as we reduce the need for extensive labeled datasets for training, which are normally not available on the early studies of a new imaging technology.

INDEX TERMS Histopathology analysis, convolutional neural network, domain adaptation, optical biopsy, virtual staining, Siamese semantic regression networks.

I. INTRODUCTION

White light endoscopy and biopsy are performed respectively for clinical and histological assessment of the gastrointestinal tract [1]. Lesion histology, resection level and infiltration level can be assessed, but even with expert assessment there may be significant intraobserver and interobserver variability [2]. During recent years, there has been an increasing interest in developing optical methods to visualize and evaluate the histological parameters on living tissue. Optical coherence

The associate editor coordinating the review of this manuscript and approving it for publication was Ali Shariq Imran.

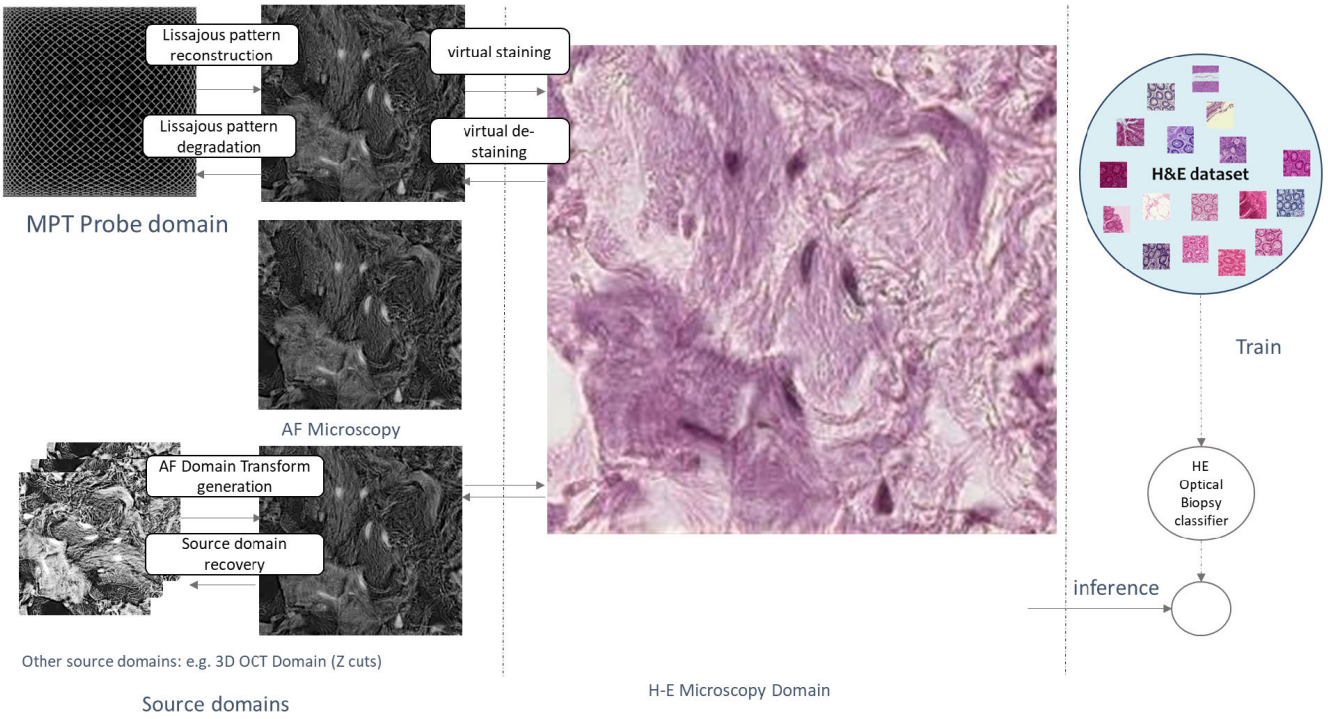

FIGURE 1. **Proposed domain transfer methodology: left) Source domain images are transferred to target domain H-E microscopy imaging right) The transformed image is analyzed over an H-E domain native inference model.**

tomography methods [3]–[6], confocal microendoscopy [7]–[9], light scattering spectroscopy [10]–[12], Raman spectroscopy [13], [14], multi photon tomography [15], [16] among others have been studied.

Although promising, major issues for these technologies precludes their entrance into real clinical practice [17]: endoscopists are not familiar with the images provided by the different optical biopsy systems. In this sense, we performed a group of interviews [18] that described their need for a system providing diagnostic decision making support, rather than simply displaying a new type of image from an unfamiliar established new modality, for which clinical experience is limited. Other studies have also reported concerns on such a system providing just a 'black-box' prediction of lesion histology and expressed their desire to have the information to understand the diagnostic decision process [19], [20].

Based on these studies, we summarize that optical biopsy models:

- Present a strong entrance barrier in the clinical community as clinicians are unfamiliar with the images provided by these modalities [17], [18].
- Should be developed to produce inference models that provides clinicians, a clinically useful assessment of the lesions.
- Should accommodate the comparison of virtually generated histology with known modalities such as Hematoxilyn-Eosin (H-E) staining. [19], [20]

These factors hinder the development and establishment of new imaging modalities for optical biopsy. One of the difficulties with the generation of an appropriate inference model is the necessity of large labeled and highly varied datasets of the targeted imaging modality. This is normally impractical and it is not possible to get the same variability and image quantity as existing datasets of traditional microscopic slides. Even in the case this inference model is developed, it will serve as a black-boxed algorithm where doctors cannot verify its assessment due to the complexity and novelty of the image interpretation.

In this work, we tackle these problems simultaneously by generating an algorithm, which translates an image from a source domain D_s of a novel image modality, into a H-E histopathological microscopy image domain target D_t, which is current gold standard image modality. The novel image modalities taken into account as source domain D_s for validating this work are Lissajous scanning probe acquired multiphoton tomography images $D_{s_{afp}}$ and AF microscopy images $D_{s_{afm}}$. First, the source domain images need reconstruction due to the degradation caused from the acquisition process; secondly they require 'virtual staining' to obtain appropriate H-E images. Figure 1 details the whole process. The main benefits for the proposed method are:

- The virtually stained images serve also as a virtual H-E biopsy as they can be used by clinicians for the verification of the assessment on an already established domain.
- This cross-domain image translation eliminates the need for a large, varied and annotated dataset on the novel images to generate an inference model as optical biopsy can be performed over target domain models.

We hypothesise that these cross-modal translated images (reconstructed and virtually stained) can be successfully used with existing cancer classification models trained with existing H-E images. This approach benefits from H-E images and already existing datasets, avoiding the need for extensive annotated datasets for the novel source domain.

This paper is organized as follows: Section II details existing work on domain adaptation while Section III details the datasets used both for the target domain and the source domains. Section IV details the proposed semantic regression network topologies and the training architectures. We incorporate domain cycle similarity loss and high level image embedding similarity. This last step allows mapping the imaging between domains taking into account also the high level features and not only the pixel level ones. Section V presents the optical biopsy algorithm that uses the reconstructed and virtually stained images whereas section V details the performance of the proposed algorithms. Conclusions are summarized in Section VII.

II. RELATED WORK

As stated by [21], the generalization of convolutional neural networks [22] have revolutionized the field of biomedical imaging analysis [23] demonstrating diagnostic accuracies comparable to experts. Examples can be found in many different tasks, such as diabetic retinopathy screening [24], skin lesion classification [25], lymph node metastasis detection [26], electro-cardiogram classification [27], [28], polyp detection [29]–[31] or histopathology analysis [32]–[34], among others. Although these deep learning methods are capable of obtaining unprecedented performance not affordable by classical machine learning algorithms, they normally require several thousands of well-labeled images to generate robust methods with the appropriate generalization capabilities. Unfortunately, only small datasets are usually obtained in clinical settings. Creating large and well annotated datasets is time consuming [35] and might not be feasible for early clinical studies, for rare diseases or for new imaging modalities.

This problem is also present when working with new diagnosis imaging modalities (e.g: optical coherence tomography, multiphoton tomography, Raman spectroscopy, reflectance confocal microscopy, etc.), where the necessary number of well annotated images to design and validate a disease diagnosis algorithm require scheduling large clinical trials. Furthermore, annotating and interpreting these images acquired with novel techniques is challenging for histopathologists, since they have little experience with those images, compared to H-E images. In order to tackle this problem, alternative approaches have been proposed. For example, *domain transfer* techniques aim to leverage the existing labels of a different task or domain, while *domain adaptation* aims to map the statistical distribution among the different domains of interest either at pixel or at representation level. Furthermore, in the case of *few shot learning* [21], [36], [37], new concepts and new representation are learned from few samples.

In this work, we follow the *domain transfer* approach to map existing images from novel imaging domains D_s into target H-E microscopy domain D_t at pixel level. This has been done traditionally by training a convolutional neural network that minimizes a mean squared error function between target and source domains [38], [39]. However, the use of this loss leads to blurred images. More advanced approaches use generative adversarial networks (GANs) [40] in co-registered inter-domain images to translate the image among the two domains, making it impossible for a discriminator network to distinguish between the mapped image and the original one in the target domain. More advanced networks such as cycle GANs [41], [42] allow domain mapping with no need of pixel co-registration among domains. Most GAN reconstruction methods suffer from the perception-distortion trade-off problem [43]. They prioritize the generation of perceptually realistic images by penalizing fulfilling the similarity loss, which minimizes the level of distortion of the reconstructed image. This makes the generated images look perceptually realistic and plausible images that, however, can reflect a disease condition different from the original real sample. This is being tackled in most recent work. For example, DiCyc GAN architecture can handle domain specific deformations [44] whereas other approaches include higher level information [45] to incorporate visual embeddings into the reconstruction.

III. DESCRIPTION OF THE DATASETS

A. HEMATOXYLIN-EOSIN MICROSCOPY TARGET DATASET-D_t

We used an extensive 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 tumours 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 image tiles of dimension 1300px $\times$ 1300px ($162\mu m \times 162\mu m$). 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 might contain parts with healthy, low-grade tumour and high-grade tumour tissue (plus background, adipose tissue, etc.). For this reason, the database includes a reviewed subset of 1,755 image tiles which are representative of its associated annotated diagnosis (Figure 2). In order to validate the proposed approach and ensure the integrity of the data used as input, when referring to this dataset we will refer as $D_{t_{sure}}$. The Biopool [46] Colon, Breast and Lung Hematoxylin-Eosin dataset (B-CBL-HE) request form

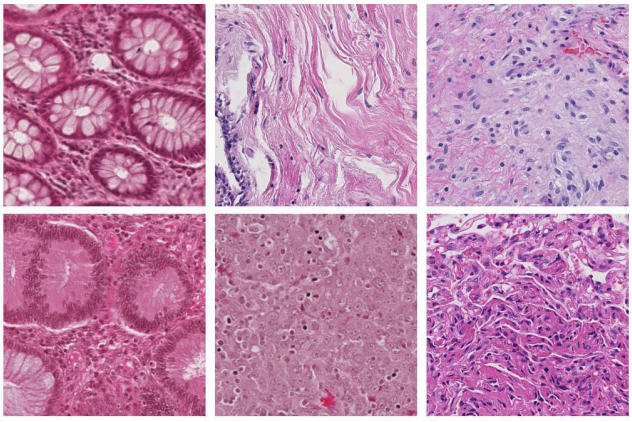

FIGURE 2. Sample images from dataset D_t [46]. Upper row depicts healthy samples and the lower row depicts tumoral samples. From left to right: colon, breast and lung tissue are depicted.

is available from the Basque Biobank's catalogue web-page [47]. The dataset is divided into train (80%), validation (10%) and testing sets (10%) according to the anonymized patient ID to assure patient independency between sets.

B. HUMAN DEGRADED AUTOFLUORESCENCE SOURCE DATASET -D_S

This dataset simulates autofluorescence (AF) image acquisition from a miniaturized scanning probe similar to the one described at [48]. Our probe consists of a PZT tube that moves a fiber tip resonantly in X-Y axes providing a Lissajous scanning pattern. In order to simulate images produced by this system, we virtually de-stain histopathological microscopy images [49] to simulate autofluorescence. These AF images are then degraded by simulating a sub-sampling that resembles the Lissajous pattern from the acquisition probe. Figure 3 shows the de-staining and degradation process. This simulated dataset serves as a conceptual dataset that allows us to focus on the reconstruction and domain transfer problem for histology image classification.

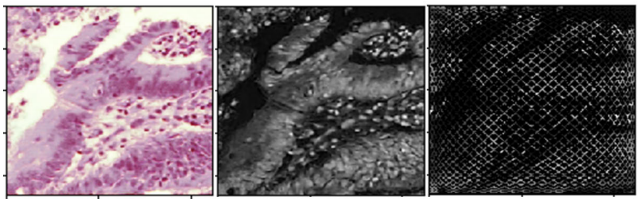

FIGURE 3. Generation of simulated AF dataset: left) Color H-E Image, middle) AF image, right) degraded AF image.

Virtual image de-staining is defined as an injective application from R^3 from the H-E image into R^1 on the AF domain. This transformation is projected into a lower dimension domain. The de-staining to generate the D_s dataset is calculated from a pixelwise color transformation from the H-E domain to the AF domain [49]. We use this function to generate an autofluorescence dataset from the Biopool dataset (described at section III-A). The database contains a total of 259,425 image tiles of dimension 1300px $\times$ 1300px ($162\mu m \times 162\mu m$). The dataset is divided into train (80%), validation (10%) and testing sets (10%) according to the anonymized patient ID identically to D_t dataset. It is worth to remark that the inverse problem transform for the staining problem (from R^1 to R^3) is not straightforward and the proposed methodology for its calculation is proposed on next section.

IV. CROSS-DOMAIN IMAGE TRANSLATION ARCHITECTURES

As a design constrain, we have avoided the use of the well-known GANs. This is because they incorporate a perceptual factor that is in opposition to the similarity factor [43] and thus, tends to generate less similar but more plausible images. We propose the use of fully convolutional regression networks that we present in section IV-A. In order to add regularization capabilities to the network, we propose to generate a double Siamese network that simultaneously assures that the generated transformed images both from the source and target domain can be transformed back to their original domain in a pixel-wise manner, therefore enforcing the invertibility of the learnt transforms, which is introduced in section IV-B. We also impose not only pixel-wise similarity on the reconstruction but also higher-level similarity by integrating image embedding vector to the loss function. It is noteworthy that all the proposed modifications on this paper can be incorporated on any existing architecture such as GANs by including the proposed loss function terms on the GAN generator output.

A. NETWORK TOPOLOGY

In our approach, we propose the use of a semantic regression densenet neural network to perform the image translation. The baseline network consists of a fully convolutional densenet architecture [50]. This network combines the descriptive power from traditional segmentation networks based on fully convolutional versions of classification ones such as SegNet [51] with the accuracy on the border detection provided by the skip connections on the U-Net segmentation network [51]. The last layer of this network has been substituted by a linear activation function and the loss function has been substituted by the mean absolute error in order to learn a pixel-wise regression transformation that translates the image from the source to the target domain. We apply this topology to the different use cases with specific modifications, as described in the following subsections.

1) AUTOFLUORESCENCE PROBE ACQUIRED IMAGE RECONSTRUCTION NETWORK

The architecture used for Lissajous pattern reconstruction is fed with a two channel 448×448 image. The first channel corresponds to the image degraded by the probe's Lissajous pattern. The pixel value corresponds to the fluorescence intensity information on the pixels in a Lissajous pattern position and

it is zero on the pixels where signal has not been acquired (i.e. not in the path of the scanning probe Lissajous pattern). The second channel includes a mask with the pixels of the Lissajous pattern position of each of the acquired pixels. This second channel serves as a mask to inform the network about the validity of the data stored in the channel. This input signal is fed to the fully convolutional densenet network.

The image passes through a subsequent set of convolutional operations and max-pooling that reduces the spatial dimension of the signal while gradually increasing the descriptive part. This results in a descriptor image. The second stage is composed of a set of convolutional filters and up-sampling layers that recover the spatial resolution of the input image on the desired domain. To be able to recover the input image high level details, the network makes use of skip connections [52] that transfer the low level features and spatial information from the source domain into the detailed reconstruction of the target domain. A final sigmoid layer performs the final reconstruction of the AF image. Figure 4 shows the network topology.

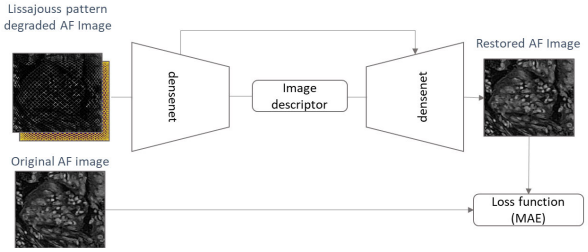

FIGURE 4. Fully convolutional DenseNet regression network for Lissajouss pattern reconstruction from AF images.

2) VIRTUAL H-E STAINING NETWORK

The architecture used for virtual staining receives a single channel 448 × 448 AF image. This input signal is connected to a fully convolutional densenet network similar to the one described in section IV-A1. The image passes through a subsequent set of convolutional operations and max-pooling that reduces the spatial dimension of the signal while gradually increasing the descriptive part. This results in a descriptor image. The second stage is composed of a set of convolutional filters and up-sampling layers that recover the spatial resolution of the input image on the desired domain. To be able to recover the input image high level details, the network make use of skip connections [52] that transfer the low level features and spatial information from the source domain into the detailed reconstruction of the target domain. A final 3 channel layer with sigmoid layer activation performs the final reconstruction of the virtual staining image. Figure 5 shows the network topology.

B. SIAMESE NETWORK ARCHITECTURES FOR CYCLE DOMAIN TRANSFORMATION LEARNING

In this section, we enhance the learning capabilities of the semantic regression networks by imposing the invertibility property to the learnt solution. In the case that the transform

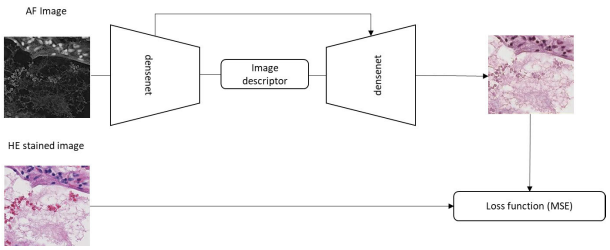

FIGURE 5. Fully convolutional DenseNet regression network for Virtual Staining of H-E images.

function between a source domain and a target domain is not bijective, the inverse transformation between the target and source domains is ill-posed and presents multiple solutions. In this case, a regression network tends to infer an average non-real intermediate solution [53].

In order to circumvent this, we propose a network that simultaneously learns both the inverse and direct transformations between the source domain and the target domain. This imposes the restriction for any network solution to be invertible to its original domain acting as a network learning regularizer. Three different architectures are proposed: 1) a baseline learning architecture consisting of independent regression networks, 2) a cycle network architecture that learns the transformations between the source and target domain and remaps this transformation into the original domain, and 3) a Siammese cycled network that simultaneously, learns the transformation from source and target domains and also remaps the transformation back to its original domain.

1) BASELINE LEARNING ARCHITECTURE

As a baseline, we train, in an independent way, two different domain transformation networks that follow the regression densenet topology defined in section IV-A. The first transformation network (direct domain transform) maps images from the source domain into the target domain, while the second one (inverse domain transform) maps the inverse transform from the target domain into the source domain. Each network (direct and inverse problem transformation) is optimized to minimize a mean absolute error loss function, as shown in equations 1 and 2. This topology allows training both direct and inverse domain transformation in an independent way. Figures 6 and 7 show the training procedure: we draw a pair of co-registered images $\{X_i, Y_i\}$, being X_i an image from the source dataset and Y_i its correspondence in the target domain dataset. The first network receives a sample from the source dataset, X_i, and estimates its correspondence on the target dataset, Y_i', by optimizing equation 1, while the second network receives a sample Y_i from the target domain and estimates its correspondence on the source domain, X_i', by minimizing the loss function defined in equation 2.

$$\mathcal{L}_{s2t} = \sum_{i=1}^{N} |Y_i - Y_i'| \tag{1}$$

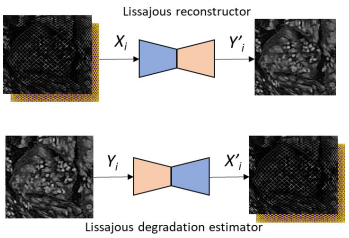

FIGURE 6. Independent training of direct and inverse transform learning used as baseline. for the Lissajous pattern reconstruction / degradation use case.

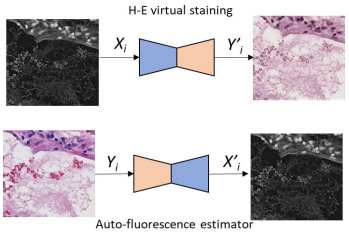

FIGURE 7. Independent training of direct and inverse transform learning used as baseline for the virtual staining/ virtual de-staining use case.

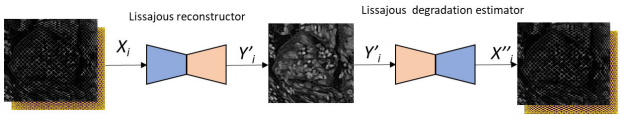

FIGURE 8. Cycle network for domain translation for the Lissajous pattern reconstruction / degradation use case.

$$\mathcal{L}_{t2s} = \sum_{i=1}^{N} |X_i - X'_i| \tag{2}$$

2) CYCLE NETWORK ARCHITECTURE

We extend the previous architecture by enforcing that any sample from the domain dataset X_s transformed into its target domain estimation Y'_s can be transformed back into its original source domain X''_s. This is enforced by generating a concatenated network that first performs the transform between the source image into the target domain and after that, the resulting estimation is transformed back into the original domain. A combined loss function defined in equation 3 ensures that both the direct and the inverse networks are trained simultaneously, while implicitly imposing bijectivity properties on the learnt transformations. Figures 8 and 9 depict this architecture for the two use cases.

$$\mathcal{L}_{s2t} = \sum_{i=1}^{N} |Y_i - Y'_i| + \sum_{i=1}^{N} |X_i - X''_i| \tag{3}$$

3) SIAMESE CYCLE NETWORK ARCHITECTURE

The previous approach enforces bijectivity property for the estimated direct and inverse transforms. However, it focuses only on the direct transform from the source domain into the target domain. To balance the importance of both transforms, we extend the previous architecture by generating a mirrored Siamese architecture that trains the direct and inverse transform networks by simultaneously transforming a sample from source dataset X_i into a source domain Y'_i and the target domain sample Y_i into its source domain counterpart X'_i. Both estimated transformations are then projected back into their original domain obtaining Y''_i and Y''_i as depicted in figures 10 and 11. Both networks performing the direct transform and both networks performing the inverse transform present shared weights. A combined loss function (equation 4) that incorporates the reconstruction similarity for the different transform is minimized to simultaneously learn the direct and inverse transforms that also acts as a regularizer to impose plausible solutions on the transformed domain samples.

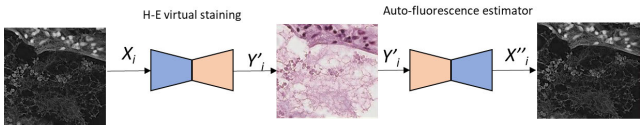

FIGURE 9. Cycle network for domain translation for the virtual staining/ virtual de-staining use case.

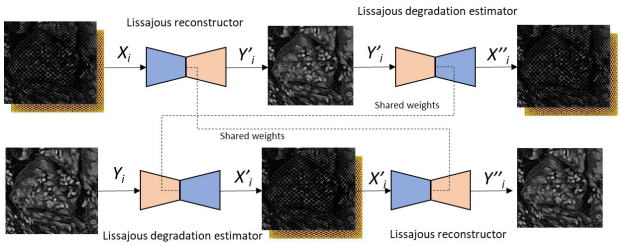

FIGURE 10. Siamese cycled network for domain translation for the Lissajous pattern reconstruction / degradation use case.

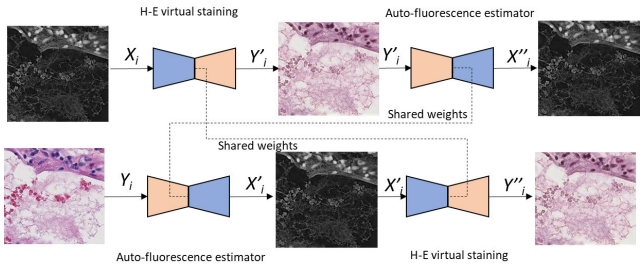

FIGURE 11. Siamese cycled network for domain translation for the virtual staining/ virtual de-staining use case.

$$\mathcal{L}_{s2t} = \sum_{i=1}^{N} |Y_i - Y'_i| + \sum_{i=1}^{N} |X_i - X''_i| + \sum_{i=1}^{N} |X_i - X'_i| + \sum_{i=1}^{N} |Y_i - Y''_i| \tag{4}$$

4) INTEGRATING FEATURE LEVEL RECONSTRUCTION CONSTRAINT

The reconstruction performed by previous networks is enforced at pixel level with the minimization of the mean

absolute error loss function. However, small changes on the pixel reconstruction can provoke changes on the response of an optical biopsy classifier or unaccurate clinical assessment. To minimize this, an additional term is added to the previous loss functions to minimize the differences on the reconstruction at image embeddings representation level. To this end, a Resnet50 classification network [21] trained over the colon training subset of the D_t dataset is used (Figure 12). This network classifies between healthy and cancerous colon tissues. The last convolutional layer of the network is removed and a global average pooling is performed over the last layer to incorporate all spatial information. The image embeddings f_i that represent the high level description of the image are included in the loss function (see Figure 13) to minimize also the reconstruction differences of the image at feature level that is extracted after the last average pooling.

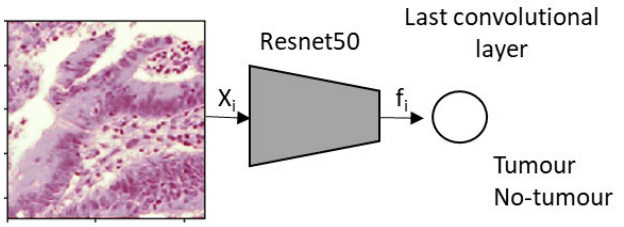

FIGURE 12. Siamese cycled network for domain translation for the Virtual staining/ virtual de-staining use case.

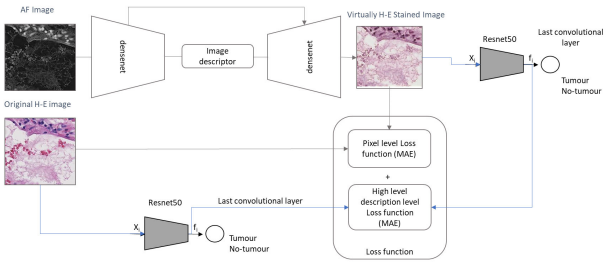

FIGURE 13. Integration of the feature vector similarity into the model loss function.

Thus, the differences between the image embeddings from the source images and the reconstructions are added into the loss function. A β parameter controls the influence of the embedding part in the loss function. In our implementation, β was normalized by the number of embedding vectors 512 to be in the same range as the other loss function terms. Equations 5 and 6 add the image embedding similarity constraint to the baseline independent model, equation 7 adds it to the cycle network whereas equation 8 adds it to the Siamese cycle network.

$$\mathcal{L}_{s2t} = \sum_{i=1}^{N} |Y_i - Y_i'| + \beta * |f_{Y_i} - f_{Y_i'}| \quad (5)$$

$$\mathcal{L}_{t2s} = \sum_{i=1}^{N} |X_i - X_i'| \quad (6)$$

$$\mathcal{L}_{s2t} = \sum_{i=1}^{N} |Y_i - Y_i'| + \sum_{i=1}^{N} |X_i - X_i''| + \beta * |f_{Y_i} - f_{Y_i'}| \quad (7)$$

$$\begin{aligned}\mathcal{L}_{s2t} = &\sum_{i=1}^{N} |Y_i - Y_i'| + \sum_{i=1}^{N} |X_i - X_i''| \\ &+ \sum_{i=1}^{N} |X_i - X_i'| + \sum_{i=1}^{N} |Y_i - Y_i''| \\ &+ \beta * |f_{Y_i} - f_{Y_i'|}| + \beta * |f_{Y_i} - f_{Y_i''}| \end{aligned} \quad (8)$$

5) IMPLEMENTATION DETAILS

The proposed architectures have been developed by using tensorflow/keras framework [54]. Baseline architecture consisted on a standard fully convolutional densenet network [50] with the modifications described on section IV-A. Both cycle network architecture, Siamese cycle network and the high level embedding architectures were constructed from the previous method by directly adding the different loss function terms by using the low-level API training capabilities from Tensorflow on the eager mode. This makes straightforward applying both methods in any existing network architecture and allowing easy integration of the subsequent model outputs on the loss function calculation. All experiments were trained for 200 epochs with an Adam optimizer with a learning rate $= 1E - 5$ and a decay of $1E - 8$ where a learning rate reduction of a factor equivalent to 0.5 was additionally included when validation loss stagnates during 5 epochs. Inference time on CPU for a frame is lower than 100ms. This time can be increased by using a GPU processor or dedicated hardware for inference. In order to simulate the effect of fluorescence fluctuation that can be caused by the non-equilibrium metabolic state of the cells [55], non-uniform optical excitation and unknown sources of noise images, training process includes a random intensity level augmentation on the fluorescence signal.

V. OPTICAL BIOPSY CLASSIFICATION ALGORITHM

A. HEMATOXYLIN-EOSIN DOMAIN CLASSIFIER

The final goal for the learnt transformations is to virtually generate H-E images from a degraded AF images. The objective of this restoration is twofold: 1) to transform the source domain image into an image understandable by the medical community and 2) to use the transformed image to perform an optical biopsy on a target domain with richer and more extensive image datasets as depicted in Figure 14.

To this end, the Lissajous pattern restoration transform learnt on the previous sections is used to reconstruct an AF image degraded by a Lissajous pattern due to acquisition probe optics. This reconstructed AF image is then virtually stained by the virtual staining network resulting into a H-E image. This virtually stained image is then classified as cancerous or non-cancerous by a residual neural network [56], where the last layer is substituted by a (1,1) sized convolution of 2 filters convolution with a softmax activation. We trained this classification network over the training set from the H-E

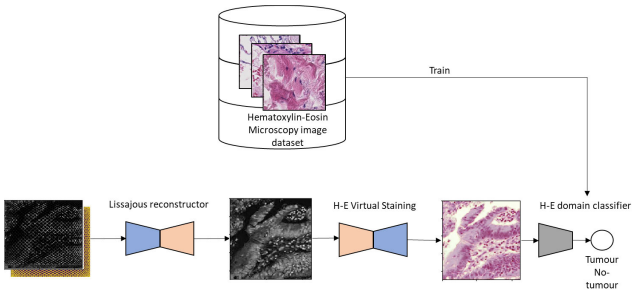

FIGURE 14. H-E domain optical biopsy classifier.

Biopool dataset [21] described in section III-A by minimizing the *categorical cross-entropy* loss function.

Test time augmentation is performed during prediction to estimate the posterior distribution of the network answer [57]. This is performed by adding noise and altering the input image. Lastly, the mean and standard deviation of the network decisions are computed, which serve as an estimation on the network confidence.

We also introduce the concept of virtual histology visual report. The virtually stained H-E image serves as a visual report where clinicians can validate the algorithm error based on the appearance of already established imaging techniques such as H-E. This report is not based on perceptually plausible reconstructions such as [41] or [58], but on minimizing the reconstruction error both on the image domain and in the image embedding domain.

B. CROSS-DOMAIN IMAGE TRANSLATION

1) AF PROBE IMAGE TO H-E DOMAIN RECONSTRUCTION NETWORK

In this section, we validate our proposed Lissajous pattern reconstruction network. For this, we assimilate this problem to a classical inpainting method where the source region is defined by the Lissajous pattern mask. We use as baseline the state-of-the-art Navier-Stokes based image restoration method [59]. We compare it against the proposed fully convolutional regression densenet defined in section IV-A1. We train two translation networks following the three approaches described in section IV-B. One learnt network covers the direct problem where an AF image is degraded while the other approximates to the inverse problem that reconstructs the degraded image. We train these networks over the training subset of the human simulated AF dataset D_s. Table 1 shows the obtained results over the testing set of this dataset. We can appreciate that the three neural networks outperform the Navier-Stokes reconstruction baseline. It can be also appreciated that, when using the simultaneous training of the inverse and direct transform with the cycle network, the back-projected image into its original domain degradation error is greatly reduced, especially when using the Siamese cycle network architecture.

Figure 15 shows the reconstruction obtained both with the Navier-Stokes approach and with the fully convolutional

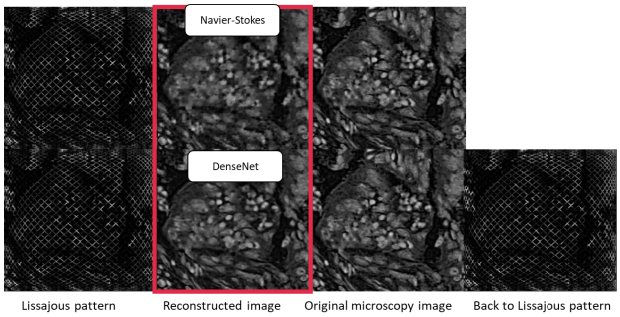

FIGURE 15. Lissajous pattern reconstruction with Navier-Stokes reconstruction baseline and with the proposed fully convolutional regression densenet.

TABLE 1. Lissajous pattern reconstruction.

Reconstruction method	MAE (reconstruction)	MAE (back-degradation)
Navier-Stokes based inpainting	0.017	N.A.
Baseline Densenet	0.014	0.22
Cycle-Network	0.013	0.010
Siamese Cycle-Network	0.014	0.006

densenet regression network. It can be appreciated that the densenet reconstruction restores more similar and sharper edges.

2) H-E VIRTUAL STAINING NETWORK

In order to validate our proposed virtual staining network, we compare the three approaches described in section IV-B by training the proposed fully convolutional regression densenet defined in section IV-A2. We train two translation networks, one covering the direct problem (virtual de-staining) and another one covering the inverse problem (virtual staining) We train these networks over the training subset of the human simulated AF dataset D_s and its original H-E dataset. Table 2 shows the obtained results over the testing set of this dataset. We can appreciate that the three neural networks obtain similar virtual staining errors. However, when de-staining the virtually stained image, we can appreciate that the proposed cycle network and symmetric cycle network architecture obtain better reconstruction errors.

Figure 16 shows that the structure and colorizing of the network is appropriately recovered.

TABLE 2. Virtual staining reconstruction.

Reconstruction method	MAE (staining)	MAE (back-destaining)
Baseline Densenet	0.031	0.030
Cycle-Network	0.030	0.024
Siamese Cycle-Network	0.031	0.018

C. OPTICAL BIOPSY

In this section, we compare the results of the two proposed algorithms for optical biopsy under the different configurations proposed on this paper.

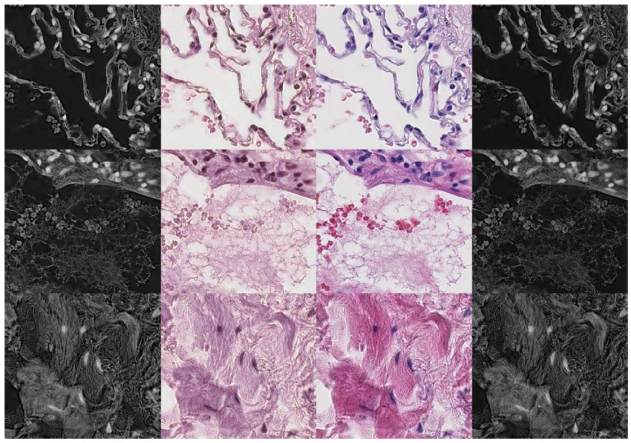

FIGURE 16. Virtual staining (inverse) and de-staining (direct) of autofluorescence microscopy image.

1) H-E CLASSIFIER

We validated the three training architectures against the testing set of the D_t H-E dataset. Detailed results are shown in table 3. It can be observed that the combination of the Siamese cycle network with the high level visual embeddings reconstruction constraint offers the best results. The regularization of the Siamese cycle network avoids generating reconstructions or transformations that alters the high level representation of the image.

TABLE 3. H-E optical biopsy classifier metrics.

Method	f_i	BAC	Sen	Spe	NPV	PPV
Original HE	-	0.89	0.88	0.90	0.89	0.89
Baseline Densenet	No	0.86	0.87	0.86	0.88	0.86
Cycle-Network	No	0.84	0.84	0.85	0.84	0.85
Siamese Cycle-Network	No	0.86	0.88	0.85	0.86	0.85
Baseline Densenet	Yes	0.86	0.90	0.82	0.89	0.83
Cycle-Network	Yes	0.84	0.84	0.84	0.84	0.84
Siamese Cycle-Network	Yes	0.88	0.88	0.88	0.88	0.88

This is also beneficial from the clinical point of view as all transformations on the target domain D_t are enforcing indirectly that the transformation does not alter the visual perception of the sample from a diagnostics point of view. This overcomes one of the issues when using other methods such as GANs, which create a distortion to make the image visually appealing while penalizing the similarity factor. [43].

As shown in Table 3, we demonstrate that virtually reconstructed and stained images perform with BAC greater than 0.85 when applied to H-E domain native classifiers which is not far from the BAC $= 0.89$ that we obtain when we perform the same experiment using the original undegraded H-E images. This is of tremendous importance as no extensive datasets are required for generating a classifier, and just a reduced number image pairs from both modalities are required. When applying the proposed Siamese cycle network together with the imposed visual embedding similarity constraint, balanced accuracy (BAC) increases up to 0.88. The accuracy of this method was statistically validated against the baseline densenet model using the McNemar [60] test at the 95% confidence interval obtaining a $p < 0.001$ (significant).

Figure 17 shows reconstruction and classification results from several images.

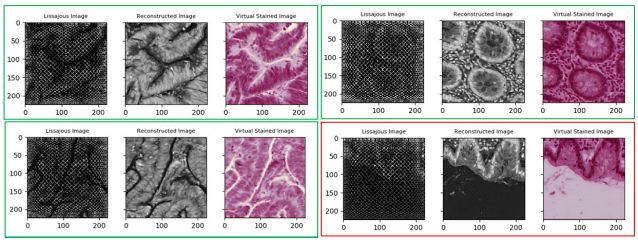

FIGURE 17. Examples of optical biopsy algorithm outputs. Top-Left) Tumoral sample classified as tumoral, Top-Right) Healthy sample classified as healthy, Bottom-Left) Tumoral sample classified as tumoral, Bottom-Right) Healthy sample classified as tumoral.

However, as explained in section III-A, D_t dataset presents tiles that are not representative of their labeled class due to the extension of the microscopy H-E images presenting lesions of different degrees along the full tissue image. Because of this, we perform the same analysis over $D_{t_{sure}}$ dataset that presents only tiles that are representative of the targeted class. Results on this validation are showed in table 4, where the obtained BAC reaches 0.95 with a negative predictive value of 1.00 when including the high level f_i term with the Siamese Cycle-Network architecture. The accuracy of this method was statistically validated against the baseline densenet model using the McNemar [60] test at the 95% confidence interval obtaining a $p = 0.031$ (significant).

TABLE 4. H-E optical biopsy classifier metrics over D_{tsure} (Representative tiles).

Method	f_i	BAC	Sen	Spe	NPV	PPV
Original HE	-	0.95	1.00	0.90	1.0	0.91
Baseline Densenet	No	0.93	0.97	0.90	0.96	0.91
Cycle-Network	No	0.90	0.93	0.87	0.93	0.87
Siamese Cycle-Network	No	0.93	0.97	0.90	0.96	0.90,
Baseline Densenet	Yes	0.88	0.87	0.90	0.87	0.90
Cycle-Network	Yes	0.91	1.00	0.84	1.00	0.86
Siamese Cycle-Network	Yes	0.95	1.00	0.90	1.00	0.91

These metrics similar (BAC $= 0.95$) to the ones we obtain when using the original undegraded H-E images., the obtained BAC reaches same value than when using the original H-E images.

2) ANALYSIS OF THE PREDICTION SELF-CONFIDENCE ESTIMATION

We applied test time augmentation [57] to regularize network response and to map a posterior probability estimation of the network predictions. We use the validation subset of the human simulated AF dataset D_s to establish a threshold that maximizes the balanced accuracy metric. We define also a confidence margin value around this threshold value that

makes network predictions falling on that range to be marked as uncertain. We apply this threshold and uncertainty margin to the testing set of the dataset. In figure 18, we show the effect of this margin for all the testing subset of D_s where we have applied the H-E classifier as explained in section V-A. For this configuration, we use Siamese cycle network for Lissajous pattern reconstruction and Siamese cycle network with visual embeddings for the virtual H-E staining with the class embedding regularization. We can observe a performance increment as we increase the uncertain margin, getting close to BAC of 0.91 just considering 10% of the images as uncertain and reaching 0.93 when we classify as uncertain 20% of the input images.

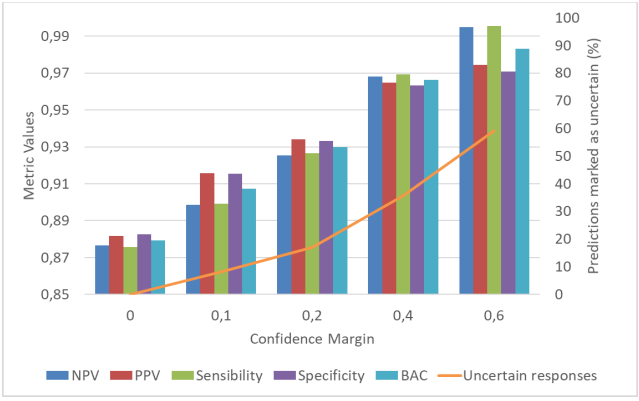

FIGURE 18. Effect of confidence margin on performance metrics on the BIOPOOL H-E dataset.

When applying the confidence margin to the testing part of the $D_{t_{sure}}$ subset, we can observe in figure 19 that even considering only a 10% of the images as uncertain, the BAC reaches 0.98 which validates the confidence margin approach.

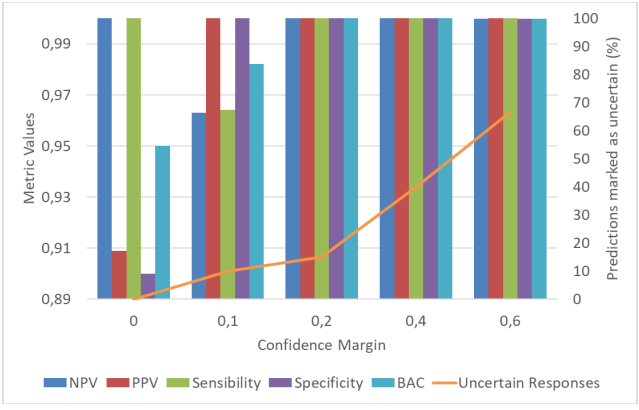

FIGURE 19. Effect of confidence margin on performance metrics on the BIOPOOL H-E dataset over the $D_{t_{sure}}$ subset.

VI. DISCUSSION

Results showed that proposed algorithm can accurately translate images from one imaging modality D_s into another D_t. Besides this, when employing the Siamese cycle-network we are able to obtain the same classification performance than when using the chemically stained H-E data showing that the proposed method can recover inherent features incorporated on the image. This avoids the need for manual labeling on the novel modality domain D_s as only co-registered images in both modalities are required to generate the translation model.

However, fully corregistered images are still required for training this domain translation model. Although this is an affordable problem for ex-vivo applications where we can scan and corregister the data among different images modalities, it is not feasible in the case of in-vivo fluorescence as we cannot H-E stain and process the in-vivo tissue. For that, it makes necessary to include the effect of fluorescence intensity variations that occur on the ex-vivo case. On one hand, convolutional neural networks learnt filters directly deal with intensity increments rather than with absolute values [61] that partially overcome this problem. On the other hand, we have incorporated during the training process a random augmentation over the intensity images to help the algorithm model this variability. However, it is still necessary to validate real performance with real in-vivo fluorescence data where movement artifacts, non-uniform optical excitation and autofluorescence intensity fluctuation effects have to be tackled [55]. Future work should focus on removing the need for strict full co-registering among modalities.

VII. CONCLUSION

In this paper we have proposed a baseline method based on regression fully convolutional densenet that is able to translate images from AF domains, virtually reconstruct and stain the image. Over this baseline model we have proposed the following innovations: - In order to minimize the perception distortion trade-off problem [43], we enforce the minimization of reconstruction error not only at pixel-level through enforcing minimization of the mean absolute error at pixel-level but also at the feature-level by enforcing similarity between the original and reconstructed image embeddings. - In order to stabilize this method, we propose a Siamese cycle network that is able to enforce bijectivity property of the learned transformation functions.

These improvements can be implemented as a modification on the training loss function. This implies that can be build on top of any existing network architecture for domain translation including GANs.

We also demonstrate that this reconstructed and virtually stained images can be used on classification models that previously existed on the target domain. When employing the proposed Siamese Cycle-Network architecture together with high-level features term, we obtain balanced accuracies of 0.95 and negative predictive values of 1.00. These values are in agreement to the results obtained with the original H-E images. This is of great importance as we eradicate the need for extensive and labeled datasets on the source domain, which are normally non affordable on the early stage clinical studies.

As the reconstruction models enforce transformed image similarity not only at pixel level but also at the image descriptive level (visual embeddings), the generated images can be used by clinicians as a histology virtual report to verify the algorithm assessment on already known target domain (optical microscopy with H-E stained slides). The developed algorithm requires paired training data among the different domains. Future work will focus on dealing with not perfectly or non-aligned domains. We have also introduced the concept of uncertainty range where the network is not confident on its own prediction. Increasing this uncertainty range leads to higher performance on the network output which allows also to determine if additional confirmation analysis is needed.

ACKNOWLEDGMENT

The authors would like to thank all pathologists that generated the BIOPOOL dataset (FP7-ICT-296162) that has been used for this work and specially to M. Saiz, A. Gaafar, S. Fernandez, A. Saiz, E. de Miguel, B. Catón, J. J. Aguirre, R. Ruiz, M^a A. Viguri, and R. Rezola. All experiments carried out in this work have been approved by the Basque Clinical Research Ethics Committee (CEIC-E).

REFERENCES

[1] T. D. Wang and J. Van Dam, "Optical biopsy: A new frontier in endoscopic detection and diagnosis," *Clin. Gastroenterol. Hepatol.*, vol. 2, no. 9, pp. 744–753, Sep. 2004.

[2] R. H. Riddell, H. Goldman, D. F. Ransohoff, H. D. Appelman, C. M. Fenoglio, R. C. Haggitt, C. Hren, P. Correa, S. R. Hamilton, B. C. Morson, S. C. Sommers, and J. H. Yardley, "Dysplasia in inflammatory bowel disease: Standardized classification with provisional clinical applications," *Hum. Pathol.*, vol. 14, no. 11, pp. 931–968, Nov. 1983.

[3] M. V. Sivak, K. Kobayashi, J. A. Izatt, A. M. Rollins, R. Ung-Runyawee, A. Chak, R. C. K. Wong, G. A. Isenberg, and J. Willis, "High-resolution endoscopic imaging of the GI tract using optical coherence tomography," *Gastrointestinal Endoscopy*, vol. 51, no. 4, pp. 474–479, Apr. 2000.

[4] P. R. Pfau, M. V. Sivak, A. Chak, M. Kinnard, R. C. K. Wong, G. A. Isenberg, J. A. Izatt, A. Rollins, and V. Westphal, "Criteria for the diagnosis of dysplasia by endoscopic optical coherence tomography," *Gastrointestinal Endoscopy*, vol. 58, no. 2, pp. 196–202, Aug. 2003.

[5] L. P. Hariri, A. R. Tumlinson, D. G. Besselsen, U. Utzinger, E. W. Gerner, and J. K. Barton, "Endoscopic optical coherence tomography and laser-induced fluorescence spectroscopy in a murine colon cancer model," *Lasers Surg. Med.*, vol. 38, no. 4, pp. 305–313, 2006.

[6] D. C. Adler, Y. Chen, R. Huber, J. Schmitt, J. Connolly, and J. G. Fujimoto, "Three-dimensional endomicroscopy using optical coherence tomography," *Nature Photon.*, vol. 1, no. 12, p. 709, 2007.

[7] O. Watanabe, T. Ando, O. Maeda, M. Hasegawa, D. Ishikawa, K. Ishiguro, N. Ohmiya, Y. Niwa, and H. Goto, "Confocal endomicroscopy in patients with ulcerative colitis," *J. Gastroenterol. Hepatol.*, vol. 23, pp. S286–S290, Dec. 2008.

[8] T. D. Wang, S. Friedland, P. Sahbaie, R. Soetikno, P. Hsiung, J. T. C. Liu, J. M. Crawford, and C. H. Contag, "Functional imaging of colonic mucosa with a fibered confocal microscope for real-time *in vivo* pathology," *Clin. Gastroenterol. Hepatol.*, vol. 5, no. 11, pp. 1300–1305, Nov. 2007.

[9] W. Piyawattanametha, H. Ra, K. E. Loewke, M. J. Mandella, C. H. Contag, Z. Qiu, T. D. Wang, S. Friedland, J. T. C. Liu, G. S. Kino, and O. Solgaard, "*In vivo* near-infrared dual-axis confocal microendoscopy in the human lower gastrointestinal tract," *J. Biomed. Opt.*, vol. 17, no. 2, 2012, Art. no. 021102.

[10] G. Zonios, L. T. Perelman, V. Backman, R. Manoharan, M. Fitzmaurice, J. V. Dam, and M. S. Feld, "Diffuse reflectance spectroscopy of human adenomatous colon polyps *in vivo*," *Appl. Opt.*, vol. 38, no. 31, pp. 6628–6637, 1999.

[11] M. B. Wallace, L. T. Perelman, V. Backman, J. M. Crawford, M. Fitzmaurice, M. Seiler, K. Badizadegan, S. J. Shields, I. Itzkan, R. R. Dasari, J. Van Dam, and M. S. Feld, "Endoscopic detection of dysplasia in patients with Barrett's esophagus using light-scattering spectroscopy," *Gastroenterology*, vol. 119, no. 3, pp. 677–682, Sep. 2000.

[12] R. S. Gurjar, V. Backman, L. T. Perelman, I. Georgakoudi, K. Badizadegan, I. Itzkan, R. R. Dasari, and M. S. Feld, "Imaging human epithelial properties with polarized light-scattering spectroscopy," *Nature Med.*, vol. 7, no. 11, p. 1245, 2001.

[13] E. Widjaja, W. Zheng, and Z. Huang, "Classification of colonic tissues using near-infrared Raman spectroscopy and support vector machines," *Int. J. Oncol.*, vol. 32, no. 3, pp. 653–662, 2008.

[14] A. Molckovsky, L.-M.-W. K. Song, M. G. Shim, N. E. Marcon, and B. C. Wilson, "Diagnostic potential of near-infrared Raman spectroscopy in the colon: Differentiating adenomatous from hyperplastic polyps," *Gastrointestinal Endoscopy*, vol. 57, no. 3, pp. 396–402, Mar. 2003.

[15] R. Cicchi, A. Sturiale, G. Nesi, D. Kapsokalyvas, G. Alemanno, F. Tonelli, and F. S. Pavone, "Multiphoton morpho-functional imaging of healthy colon mucosa, adenomatous polyp and adenocarcinoma," *Biomed. Opt. Exp.*, vol. 4, no. 7, pp. 1204–1213, 2013.

[16] R. Cicchi, A. Sturiale, G. Nesi, F. Tonelli, and F. Pavone, "Two-photon imaging and spectroscopy of fresh human colon biopsies," *Proc. SPIE*, vol. 8226, Feb. 2012, Art. no. 82263S.

[17] S. Coda and A. Thillainayagam, "State of the art in advanced endoscopic imaging for the detection and evaluation of dysplasia and early cancer of the gastrointestinal tract," *Clin. Exp. Gastroenterol.*, vol. 7, p. 133, May 2014.

[18] L. Bote-Curiel, J. F. O. Morán, J. B. Pagador, F. M. S. Margallo, B. Glover, J. Teare, F. Polo, N. Arbide, C. L. Saratxaga, P. Solleder, D. Alfieri, F. di Noia, B. Roycroft, J. Bain, R. Cicchi, F. S. Pavone, and A. Picon, "Innovative multiphotonic endoscope to address technological challenges in current colonoscopy procedure," in *Proc. CASEIB*, 2018, pp. 1–4.

[19] H. K. Roy and V. Backman, "Spectroscopic applications in gastrointestinal endoscopy," *Clin. Gastroenterol. Hepatol.*, vol. 10, no. 12, pp. 1335–1341, Dec. 2012.

[20] M. Panjehpour, B. Overholt, T. Vo-Dinh, R. Haggitt, D. Edwards, and F. Buckley, "Endoscopic fluorescence detection of high-grade dysplasia in Barrett's esophagus," *Gastroenterology*, vol. 111, no. 1, pp. 93–101, Jul. 1996.

[21] A. Medela, A. Picon, L. C. Saratxaga, O. Belar, V. Cabezon, R. Cicchi, R. Bilbao, and G. Ben, "Few shot learning in histopathological images: Reducing the need of labeled data on biological datasets," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Apr. 2019, pp. 1860–1864.

[22] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," in *Shape, Contour and Grouping in Computer Vision*. London, U.K.: Springer-Verlag, 1999, p. 319.

[23] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[24] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *J. Amer. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, 2016.

[25] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, p. 115, 2017.

[26] B. E. Bejnordi *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *J. Amer. Med. Assoc.*, vol. 318, no. 22, pp. 2199–2210, Dec. 2017.

[27] A. Elola, E. Aramendi, U. Irusta, A. Picón, E. Alonso, P. Owens, and A. Idris, "Deep neural networks for ECG-based pulse detection during out-of-hospital cardiac arrest," *Entropy*, vol. 21, no. 3, p. 305, Mar. 2019.

[28] A. Picon, U. Irusta, A. Álvarez-Gila, E. Aramendi, F. Alonso-Atienza, C. Figuera, U. Ayala, E. Garrote, L. Wik, J. Kramer-Johansen, and T. Eftestøl, "Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia," *PLoS ONE*, vol. 14, no. 5, May 2019, Art. no. e0216756.

[29] L. Sanchez-Peralta, A. Calderon, V. Cabezon, J. Ortega-Moran, F. Sanchez-Margallo, F. Polo, C. Saratxaga, and A. Picon, "Systematic acquisition and annotation of clinical cases for the generation of a medical image database," *Brit. J. Surg.*, vol. 106, p. 16, Feb. 2019.

[30] L. F. Sánchez-Peralta, L. Bote-Curiel, A. Picón, F. M. Sánchez-Margallo, and J. B. Pagador, "Deep learning to find colorectal polyps in colonoscopy: A systematic literature review," *Artif. Intell. Med.*, vol. 108, Aug. 2020, Art. no. 101923.
[31] L. F. Sánchez-Peralta, A. Picón, J. A. Antequera-Barroso, J. F. Ortega-Morán, F. M. Sánchez-Margallo, and J. B. Pagador, "Eigenloss: Combined PCA-based loss function for polyp segmentation," *Mathematics*, vol. 8, no. 8, p. 1316, Aug. 2020.
[32] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.
[33] A. A. Cruz-Roa, J. E. A. Ovalle, A. Madabhushi, and F. A. G. Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Berlin, Germany: Springer, 2013, pp. 403–410.
[34] J. P. S. García, A. P. Ruiz, M. A. B. Pérez, and U. I. Zarandona, "Contaje de mitosis en imágenes histológicas mediante redes neuronales convolucionales," in *Proc. 35th Congreso anual Sociedad Española Ingeniería Biomédica (CASEIB)*, Bilbao, Spain, 2018, pp. 199–202.
[35] J. Bernal, A. Histace, M. Masana, Q. Angermann, C. Sánchez-Montes, C. R. de Miguel, M. Hammami, A. García-Rodríguez, H. Córdova, O. Romain, G. Fernández-Esparrach, X. Dray, and F. J. Sánchez, "GTCreator: A flexible annotation tool for image-based datasets," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 2, pp. 191–201, Feb. 2019.
[36] A. Picon and A. Medela, "Constellation loss: Improving the efficiency of deep metric learning loss functions for the optimal embedding of histopathological images," *J. Pathol. Informat.*, vol. 11, no. 1, p. 38, 2020.
[37] D. Argüeso, A. Picon, U. Irusta, A. Medela, M. G. San-Emeterio, A. Bereciartua, and A. Alvarez-Gila, "Few-shot learning approach for plant disease classification using images taken in the field," *Comput. Electron. Agricult.*, vol. 175, Aug. 2020, Art. no. 105542.
[38] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2536–2544.
[39] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 649–666.
[40] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.
[41] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.
[42] H. Uzunova, J. Ehrhardt, and H. Handels, "Memory-efficient GAN-based domain translation of high resolution 3D medical images," *Comput. Med. Imag. Graph.*, vol. 86, Dec. 2020, Art. no. 101801.
[43] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6228–6237.
[44] C. Wang, G. Yang, G. Papanastasiou, S. A. Tsaftaris, D. E. Newby, C. Gray, G. Macnaught, and T. J. MacGillivray, "DiCyc: GAN-based deformation invariant cross-domain information fusion for medical image synthesis," *Inf. Fusion*, vol. 67, pp. 147–160, Mar. 2021.
[45] S. Yao, J. Tan, Y. Chen, and Y. Gu, "A weighted feature transfer GAN for medical image synthesis," *Mach. Vis. Appl.*, vol. 32, no. 1, pp. 1–11, Jan. 2021.
[46] B. de Jong, "Services associated to digitalised contents of tissues in biobanks across europe: A proof of concept–biopool," in *Proc. ISBER Annu. Meeting*, 2013.
[47] (2015). *Basque Biobank's Catalogue*. [Online]. Available: http://www.biobancovasco.org/en/CA128-sample-catalog.html
[48] W. Liang, G. Hall, B. Messerschmidt, M.-J. Li, and X. Li, "Nonlinear optical endomicroscopy for label-free functional histology *in vivo*," *Light, Sci. Appl.*, vol. 6, no. 11, pp. e17082–e17082, Nov. 2017.
[49] A. H. Fischer, K. A. Jacobson, J. Rose, and R. Zeller, "Hematoxylin and eosin staining of tissue and cell sections," *Cold Spring Harbor Protocols*, vol. 2008, no. 6, p. 4986, May 2008.
[50] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 11–19.
[51] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[52] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2015, pp. 234–241.
[53] M. Shahriari, D. Pardo, A. Picón, A. Galdran, J. D. Ser, and C. Torres-Verdín, "A deep learning approach to the inversion of borehole resistivity measurements," *Comput. Geosci.*, vol. 24, no. 3, pp. 1–24, 2020.
[54] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Oper. Syst. Design Implement. (OSDI)*, 2016, pp. 265–283.
[55] J. Surre, C. Saint-Ruf, V. Collin, S. Orenga, M. Ramjeet, and I. Matic, "Strong increase in the autofluorescence of cells signals struggle for survival," *Sci. Rep.*, vol. 8, no. 1, pp. 1–14, Dec. 2018.
[56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[57] A. Picon, A. Alvarez-Gila, M. Seitz, A. Ortiz-Barredo, J. Echazarra, and A. Johannes, "Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild," *Comput. Electron. Agricult.*, vol. 161, pp. 280–290, Jun. 2019.
[58] Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, "PhaseStain: The digital staining of label-free quantitative phase microscopy images using deep learning," *Light, Sci. Appl.*, vol. 8, no. 1, p. 23, Dec. 2019.
[59] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-Stokes, fluid dynamics, and image and video inpainting," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Dec. 2001, p. 1.
[60] M. W. Fagerland, S. Lydersen, and P. Laake, "The McNemar test for binary matched-pairs data: Mid-p and asymptotic are better than exact conditional," *BMC Med. Res. Methodol.*, vol. 13, no. 1, p. 91, Dec. 2013.
[61] A. P. Ruiz, A. A. Gila, U. Irusta, and J. E. Huguet, "Why deep learning performs better than classical machine learning?" *Dyna Ingenieria E Industria*, vol. 95, no. 1, pp. 119–122, Mar. 2020.

ARTZAI PICON received the M.Eng. degree in industrial engineering in 2002, the M.Res. degree in 2005, the European Ph.D. degree in 2009, and the M.Sc. degree in biomedical engineering in 2012. He currently serves as a Principal Researcher for the Tecnalia Research & Innovation Computer Vision group. He was one of the coordinators of PICCOLO project. He also contributes by developing industry level deep learning models for transnational companies under private research contracts. He has published more than 40 articles, holds 13 patents, and he is a reviewer of different prestigious journals related to machine vision and artificial intelligence. His research interests include deep learning techniques for life sciences applications, especially focusing on domain translation, reliability assurance, and robustness. In 2002, he received the Best M.Eng. Final Project Award from Accenture Company for the implementation of a facial recognition system. In 2006, he also received the ONCE International Research and Development Award in New Technologies for The Blind and Visually Impaired. In 2011, the industrial application of his thesis developments received the 3rd EARTO's award on innovation. In 2014, he forms part of the team granted in 1st EARTO's Award on Innovation for the development for a histopathology image retrieval system for biopsies among biobanks.

ALFONSO MEDELA was born in Bilbao, Spain, in 1995. He received the B.Sc. degree in physics from the University of the Basque Country (UPV-EHU) and the University of Groningen (RUG), and the master's degree in big data and business intelligence from the University of Deusto. He was selected to take part in the Talentia programme, promoted by Bizkaia Talent, which is targeted at the students living in Biscay with the highest potential at the three Basque universities. He works applying deep learning and other computer vision and machine learning techniques to solve complex problems. He is specialized in training algorithms with few data, also known as few-shot learning, and in medical data. In 2019, he co-founded a Startup called LEGIT, where he works as the CTO.

LUISA F. SÁNCHEZ-PERALTA received the Telecommunication Engineering degree from the University of Seville, in 2008, and the master's degree from the University of Extremadura, in 2010. She is currently pursuing the Ph.D. degree with the Information Technology programme, University of Extremadura, and her thesis is focused on deep learning methods applied to colorectal polyps. She got a Research Fellowship with the University of Seville. She developed her degree thesis in the Medical Imaging Center, Katholieke Universiteit Leuven, Belgium, in 2008. Since 2008, she has been working as a Researcher with the Bioengineering and Health Technologies Unit, Centro de Cirugía de Mínima Invasión Jesús Usón. She has participated in 23 research projects, is the author or coauthor of 18 scientific articles; 87 communications to national and international congresses; and has presented four invited lectures. She has organized six international congresses and symposia related to the field of bioengineering. She is the author of two patents and an industrial design and a reviewer for several journals and conferences. Her research interests include new approaches to training in minimally invasive surgery, including technologies ICTs in the training process (e-learning, virtual reality, and serious games) and medical imaging processing using deep learning methods.

RICCARDO CICCHI received the master's degree in physics from the University of Florence, in 2003, working on skeletal muscle myosin mechanics using optical trapping and high-resolution imaging techniques, and the Ph.D. degree in physics from the University of Florence, in 2007, with a thesis titled "Non-linear imaging of human skin." He is currently a Researcher with the National Institute of Optics–National Research Council of Italy, where he coordinates the Biomedical Optics group, and he is also associated with the European Laboratory for Non-linear Spectroscopy (LENS), where he has served as a Reference Person for the EU-H2020 project PICCOLO. He is also a Lecturer of "Optical methods in biology" within the Master Course on Molecular Biotechnology, University of Florence. His research interests include multiphoton microscopy imaging of human skin tissue, in particular functional imaging of skin cancer and the study of collagen using SHG microscopy. His research interests also include non-linear imaging and spectroscopy of human tissues such as cornea, bladder, colon, and brain. He has authored over 120 scientific publications (source: ISI-Web of Science) that have received more than 1600 citations, and holds three patents. His research team develops, in collaboration with various medical experts, optical methods for diagnostics on biological tissues, using a microscopic approach for tissue diagnostics and pathological evaluation, and a spectroscopic approach with fiber optic sensors for clinical applications.

ROBERTO BILBAO is currently pursuing the Ph.D. degree in cell biology. He is also the Director of the Basque Biobank, the Basque Health System's tool for the management of biological samples for research purposes. He has been the Coordinator of the BIOBOOL project and has participated in the MIDAS project. His research interests include the advance in the development of biobanking techniques of human samples and big data.

DOMENICO ALFIERI received the M.Sc. degree in telecommunication engineering from the University of Naples, "Federico II," Naples, Italy, in 2002. He worked as a Researcher with the National Institute of Optics in Naples dealing with developing optical methods for monitoring and diagnostic non-invasive f-MEMS and for characterization of materials by means of digital holography, until the 2006. In the 2006, he joins Light4Tech s.r.l. and focus on activity related to research, design and development of equipment in the field of optics, photonics, spectroscopy, microscopy, and optoelectronics. In 2017, he became a Shareholder and the Head of the Research and Development area of Light4Tech. He deals with the design of prototypes, systems, and products based on the photonics and physics mainly applied in the fields of medical devices and microscopy, he busy on the CE mark and regulatory of medical devices, write and manage European and regional projects, he manages external collaborations with research centres. He have published 41 articles, 15 conference papers, holds one patent, and has collaborated in the drafting of three books on the non-destructive characterization of materials based on interferometric techniques.

ANDONI ELOLA was born in Donostia, in 1992. He received the bachelor's and master's degrees from the University of the Basque Country, in 2015 and 2017, respectively, where he is currently pursuing the Ph.D. degree. He has published ten scientific articles and made more than 15 contributions to scientific conferences. His research interests include applications of biomedical signal processing and advance machine learning techniques specially focused on cardiac arrest.

BEN GLOVER received the B.Sc. degree in 2006, and the B.M. degree in 2009. He is currently continuing work on a research M.D. at Imperial College, London, while working clinically. He is also a Physician and Specialist Trainee in gastroenterology, with an interest in advanced diagnostic endoscopy and emerging endoscopic imaging modalities. His research interests include advanced optical techniques and optical diagnosis of colorectal polyps, and prediction of gastric Helicobacter Pylori infection.

CRISTINA L. SARATXAGA received the M.Sc. degree in systems and control engineering in 2014. In 2012, she joined the Computer Vision group of Tecnalia. She has been a Computer Science Engineer since 2006. Since then, she has been working in the development of advanced image processing algorithms for the extraction of clinically relevant image features and characteristics in medical images for the development of computer aided diagnosis systems (CAD). In this respect, she forms part of the team granted in 2014 with the 1st EARTO Award on Innovation for an Image Retrieval Systems of histopathological images in Biobanks (BioSimil). She is the current coordinator of the European project PICCOLO.