Contributions to medical image segmentation with artificial intelligence for decision support

Pierre-Henri Conze

COLLEGE DOCTORAL BRETAGNE — SCIENCES DE LA VIE ET DE LA SANTE

Université de Bretagne Occidentale

HABILITATION À DIRIGER DES RECHERCHES DE

L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE

Par

Pierre-Henri CONZE

Contributions to medical image segmentation with artificial intelligence for decision support

Habilitation à Diriger des Recherches présentée et soutenue à Brest, le 5 juillet 2024

Unité de recherche : IMT Atlantique, LaTIM UMR 1101, Inserm

Rapporteurs avant soutenance :

Elsa ANGELINI	Professeur - Telecom Paris, Imperial College London, Columbia University
Antoine VACAVANT	Professeur des Universités – Université Clermont d'Auvergne
Maria A. ZULUAGA	Maître de Conférences HDR – EURECOM, King's College London

Composition du Jury :

Présidente :	Émilie CORNEC-LE GALL	PU-PH – Université de Bretagne Occidentale, CHRU de Brest
Rapporteurs :	Elsa ANGELINI	Professeur - Telecom Paris, Imperial College London, Columbia University
	Antoine VACAVANT	Professeur des Universités – Université Clermont d'Auvergne
	Maria A. ZULUAGA	Maître de Conférences HDR – EURECOM, King's College London
Examinateurs :	Émilie CORNEC-LE GALL	PU-PH – Université de Bretagne Occidentale, CHRU de Brest
	Franck VERMET	Maître de Conférences HDR – Université de Bretagne Occidentale
	Vincent NOBLET	Ingénieur de Recherche HDR – ICube, CNRS, Université de Strasbourg

Table of Contents

List of Figures

Chapter 1

Foreword

General context. This manuscript presents an overview of the research activities I conducted from 2015 to 2024, and outlines research perspectives for the upcoming years. Choices and compromises were necessary to deliver a digest and concise summary. Thus, I modestly strive to synthesize my past contributions and anticipate future prospects on medical image analysis with artificial intelligence, towards better clinical decision support. At the interface between image processing, artificial intelligence and applied mathematics, this manuscript mainly emphases on medical image segmentation, multi-modal information fusion, medically-sound computational modeling and longitudinal follow-up.

In current clinical practice, the accuracy of disease detection, diagnosis, characterization and monitoring from medical images hinges on the proficiency of individual clinicians. However, this dependence can result in considerable inter-reader variability in the interpretation of medical images. To address this challenge, numerous computer-aided medical image analysis techniques have been developed in the past decades to benefit from precise, reproducible and objective measurements performed by computers. By transitioning from grids of millions of voxels to the understanding of the intricacies of the human body, these tools strive to support clinicians in effectively interpreting medical images and enhancing clinical decision-making. The underlying rationale is that employing computer-aided medical image analysis systems can mitigate the difficulties faced in clinical settings, including variations in clinical skills, potential fatigue of human experts or insufficient medical resources.

In recent years, supervised methodologies have become increasingly popular in the field of medical image analysis by involving the use of training data to develop computer-aided systems. Pattern recognition and machine learning approaches enabled a shift from systems that were completely designed by humans to systems that are trained from training data from which feature vectors are extracted. These algorithms ascertain the optimal decision boundary in high-dimensional feature space. A crucial step in the design of such systems is the extraction of discriminant features from medical images. Instead of designing handcrafted features, a logical next step has been to let computers learn the features that optimally represent the available data, for the problem at hand. This concept lies at the basis of many deep learning algorithms. Resulting models are composed of many layers that transform input data to the targeted output while learning increasingly higher level features, in an automated fashion.

Past research activities. In this context, my research work has focused, starting from 2014, on the analysis of medical images through machine and deep learning techniques. The integration of

artificial intelligence technologies has revolutionized the landscape, offering unprecedented insights into human anatomy understanding and pathology characterization. A key breakthrough in this field was the application of deep learning techniques for medical image segmentation purposes (Chap.2). The ability to accurately delineate and classify anatomical or pathological structures within medical images is substantial to achieve anatomically accurate representations and enhance clinical decision-making processes in diagnosis, treatment planning or disease monitoring.

Many medical imaging protocols often involve the use of various imaging information (e.g. multi-view, multi-domain, multi-modal). However, while deep learning techniques based on convolutional neural networks have become very popular in medical image analysis, a common practice is to develop modality-specific (i.e. marginal) computational anatomical models, without explicitly taking into account the multi-modal nature of the underlying imaging data. Marginal processing may leave potentially valuable cross-modality information unused and as a consequence hamper the performance in downstream applications. By leveraging transfer learning and multi-modal analysis (Chap.3), my contributions aimed at exploiting cross-domain or cross-modal information for improved image processing performance, especially in medical image detection and segmentation.

By providing a summary of my past research activities into two distinct parts (Chap.2 and 3), Part.I highlights the progression from the foundational aspects of medical image segmentation to more advanced techniques that involve leveraging the transferability of knowledge across different modalities and integrating information from various sources. While Chap.2 focuses on achieving precise segmentation of medical images using artificial intelligence, Chap.3 explores how insights gained from one modality or domain can be effectively transferred and fused with information from others, leading to enhanced analysis and interpretation of medical imaging data.

On-going and future research activities. The manuscript continues in Part.II with the presentation of my current projects, perspectives and future directions which will guide my research activities over the coming years. Thus, Chap.4 delves into the development of computational tools and algorithms to aid surgeons in planning surgical procedures, optimizing surgical outcomes and minimizing risks. It also addresses the need for decision support systems that facilitate the monitoring during the post-treatment phase, assisting clinicians in assessing treatment efficacy and making informed decisions regarding further interventions or adjustments to the treatment plan. The strong hypothesis is that the development of tools based on artificial intelligence and dedicated to the characterization and the prediction of pathological tissue evolution can provide further guidance for clinicians, towards early biomarker extraction as well as improved surgical and therapy planning and management.

Finally, the ending part of the manuscript provides a curriculum, a list of publications and patents as well as a summary of the conducted supervision and teaching activities.

Part I

Overview of past research activities

Chapter 2

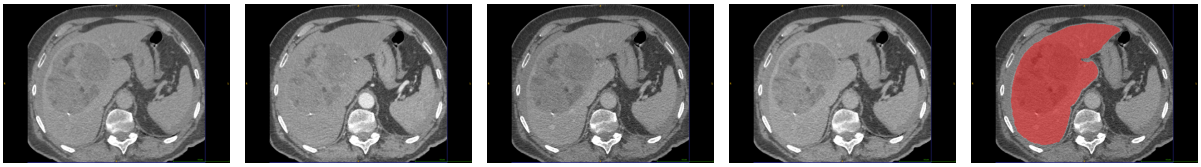
Anatomically-sound medical image segmentation

2.1 Introduction

The increased volume of medical data to be interpreted by clinicians for diagnosis, therapeutic or surgical planning purposes has encouraged the development of computer-aided image analysis tools to leverage accurate, fast, and repeatable measurements facilitated by computational resources. Among existing analysis tasks, medical image segmentation whose goal is to extract the boundaries of anatomical or pathological structures from medical images is crucial. Also commonly used in computer vision [1], semantic segmentation is a key step for many medical imaging workflows since the information arising from the resulting voxel-wise delineations can help clinicians to diagnose disorders, assess disease progression, plan therapeutic interventions or monitor treatment effects. Core feature of many computer-aided detection and diagnosis systems, segmentation is involved in the analysis of many modalities including computed tomography (CT) and magnetic resonance (MR) imaging.

Delineating anatomical or pathological structures from medical images is traditionally performed manually. This task is time-consuming and requires suitable clinical expertise to get clinically-relevant contours. This is therefore not applicable to large volumes of data produced in clinical routine. Given the potential fatigue of human experts and the wide variations in expertise, manual segmentation is prone to strong intra- and inter-expert variability. Irregularities of the targeted structures, morphological variations or pathological deformities between patients as well as the potential lack of clearly visible boundaries with the surrounding anatomy further affect the non-agreement between operators.

Mathematical models and low-level image processing have been extensively exploited for segmentation before the rise of machine and deep learning techniques. In particular, model-based segmentation incorporating statistical shape models has been used in various clinical contexts [2]. These models have been further improved by exploiting prior knowledge of shape information, e.g. by relying on shape fitting to guide the delineation process [3]. Conversely, aligning and merging manually segmented images into a specific atlas coordinate space have been developed as a reliable alternative to statistical shape models. In this context, various single- and multi-atlas methods have been proposed relying on non-linear registration [4]. Some hybrid methods relying on statistical shape models constrained with



$I_{\texttt{WI}}$ $I_{\texttt{AR}}$ $I_{\texttt{EV}}$ $I_{\texttt{LV}}$ liver mask

Figure 2.1 – Registered dynamic contrast-enhanced CT scans with ground truth liver segmentation.

probabilistic atlases have also been investigated. Medical image segmentation has also been performed through graph-based methods [5], clustering [6], region growing [7] or active contours [8].

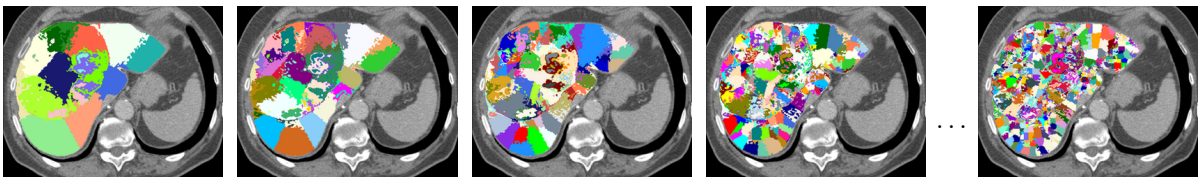
However, these methodologies are not perfectly suited for high inter-subject shape variability, weak boundaries and significant differences in tissue appearance. In most cases, their robustness does not meet the intrinsic challenges of medical images (e.g. noise, non-uniform contrasts, motion artifacts). Moreover, many of these methods are semi-automatic and require prior knowledge, associated with high computational costs. This has motivated the development of novel paradigms based on machine and deep learning techniques to exploit image characteristics in a data-driven manner. The remaining part of this chapter describes some of my contributions in this context, in a mono-modal setting.

2.2 Machine learning and hierarchical supervoxels

My post-doctoral work at ICube laboratory[1] mainly focused on machine learning approaches for the segmentation of healthy and pathological liver tissues from dynamic contrast-enhanced (DCE) images (Fig.2.1). This study looked at pre-operative locoregional treatments (PLT) for hepatocellular carcinoma (HCC), a primary form of liver cancer, which delay tumor progression through necrosis. Standard evaluation scores (e.g. RECIST [9]) used to predict the PLT response do not provide sufficient results [10]. A more efficient HCC patient follow-up can be reached through tumor necrosis (TN) rate (i.e. percentage of necrotic tissue in tumors) which provides more significant correlation with survival rates, as shown after transplantation [11] or resection [12]. However, calculating the TN rate requires distinguishing tumors from healthy liver tissues and segmenting tumoral active and necrotic areas.

Medical image segmentation covered in an interactive perspective through random forests (RF) [13] has grown in popularity from the 2010s. Nevertheless, one key aspect was how to accurately capture long-range spatial context by means of conditional random fields regularization, context-rich features [14], entangled decision forests or auto-context [15]. Alternatively, the idea came in [16], [17] to extend the use of RF from voxels to supervoxels, visual primitives generated by aggregating neighboring voxels sharing similar characteristics [18]. The aim was to improve sample representativeness, regularize features over consistent areas and reduce the computational complexity. Performed in a single-scale fashion, methodologies from [16], [17] were limited in their ability to deal with spatial

1. `https://icube.unistra.fr`



scale $k = 0$ $k = 1$ $k = 2$ $k = 3$ K

Figure 2.2 – Hierarchical multi-scale supervoxel decomposition [**19**] of the liver area in CT modality.

adaptivity. Indeed, the spatial extent of two different structures may highly differ and intrinsic tissue properties may emerge at some specific scales only. This justified the need for robust multi-scale feature representations in RF-based supervoxel classification.

In this regard, we proposed two main contributions [**19**][2] with liver tissue segmentation as application. First, we proposed to combine RF and supervoxel decomposition by exploiting multi-phase supervoxel-based features extracted from registered DCE-CT scans (Fig.2.1) acquired before contrast agent injection (`WI`) and after, at arterial (`AR`), early (`EV`) and late venous (`LV`) phases [**20**]. Second, to deal with multiple spatial extents and appearance heterogeneity of tumoral tissues, we extended this baseline by combining RF and a hierarchical multi-scale tree resulting from a recursive supervoxel decomposition [**21**]. Our motivation was to describe each leaf supervoxel as a sequence of supervoxels belonging to its ascendant hierarchy. By concatenating features across the hierarchical tree, RF can find itself the best data sampling without explicitly choosing how to combine the different scales.

More precisely and contrary to supervoxel decomposition defined on one single scale, we relied on a hierarchical multi-scale supervoxel representation. The liver area Ω_l was decomposed into a set of $K+1$ partitions $\mathcal{P}^k = \{r_i^k\}$ defined at scale $k \in \{0, \ldots, K\}$ where 0 and K denote respectively the coarsest and finest scales (Fig.2.2). Each partition $\{\mathcal{P}^k\}$ was a collection of 3D connected SLIC [18] supervoxels $\{r_i^k\}$ built at scale k such that $r_i^k \cap r_{j\neq i}^k = \emptyset$ and $\bigcup_i r_i^k = \Omega_l$. As illustrated in Fig.2.3, the resulting sequence of partitions $\{\mathcal{P}^k\}$ was encoded in the layers of a multi-resolution tree $\mathcal{M} = \{\mathcal{M}^k\}$ where layer $\mathcal{M}^k$ maps each supervoxel $r_i^k \in \mathcal{P}^k$ to a set of child supervoxels $\{r_j^{k+1}\} \subset \mathcal{P}^{k+1}$ such that $r_i^k = \bigcup_j r_j^{k+1}$. This representation was obtained through a recursive supervoxel generation process. Instead of merging supervoxels in a fine-to-coarse strategy starting from individual voxels, we first generated an initial coarse partition $\mathcal{P}^0$ by dividing Ω_l into a small set of large supervoxels. Each of these supervoxels r_i^0 of $\mathcal{P}^0$ was then split into K_s supervoxels $\{r_j^1\} \subset \mathcal{P}^1$ using SLIC applied on the area formed by voxels of r_i^0. Once built, each r_j^1 was split into K_s children and this procedure was repeated iteratively in a coarse-to-fine fashion.

Once $\mathcal{M}$ built, we started by assigning multi-phase supervoxel-based visual features $\theta(r^k)$ to all supervoxels r^k in each partition $\mathcal{P}^k$. Visual features were multi-phase supervoxel-based, relying on spatial intensity, spatial gradient and tissue dynamic. Then, we associated to each supervoxel r^K (i.e. at the finest scale K) all the supervoxels of decreasing scale belonging to its ascendant hierarchy including

2. references in **bold** highlight co-authored publications

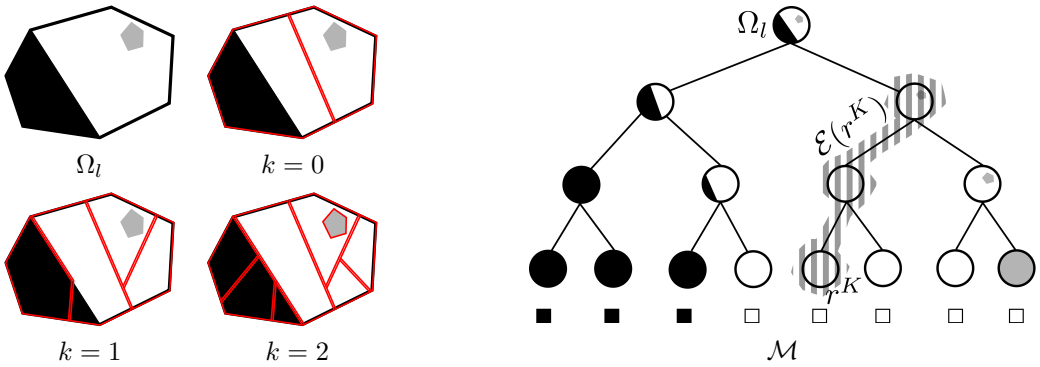


Figure 2.3 – Hierarchical multi-scale supervoxel representation **[19]** on a binary classification example with $K = 2$ and $K_s = 2$. Expected labels are displayed into squared boxes.

itself: $\mathcal{E}(r^K) = \{r^k\}_{k\in[0,\dots,K]}$ (Fig.2.3). Finally, we defined a new feature vector $\gamma(r^K)$ associated to each r^K of $\mathcal{P}^K$ as the concatenation of all visual features θ assigned to supervoxels of $\mathcal{E}(r^K)$. Concatenating all visual features across $\mathcal{M}$ reached a powerful multi-scale description of finest scale supervoxels. A tissue classification based on supervoxels of scale K was carried out using RF based on the features stacked into $\gamma(r^K)$. Internal node parameters were optimized with respect to $\mathcal{S}^K = \{r_i^K, c(r_i^K)\}_{i\in\{1,\dots,K_u\}}$, a set of K_u finest scale training supervoxels $\{r_i^K\}$ manually labeled by the clinician. Label prediction for $r^K \in \mathcal{P}^K \backslash \mathcal{S}^K$ was performed as follows:

$$c(r^K) = \arg\max_{c_l} \frac{1}{T}\sum_{t=1}^{T} \frac{|\{r_i^K, c(r_i^K)\} \in \mathcal{S}_{l_t} \mid c(r_i^K) = c_l|}{|\mathcal{S}_{l_t}|} \tag{2.1}$$

where $\mathcal{S}_{l_t}$ is the partition of $\mathcal{S}^K$ received by leaf node l_t of the t^{th} tree, $c(r_i^K)$ the ground truth label of supervoxel r_i^K among classes c_l (i.e. parenchyma, tumoral active and necrotic tissues).

As suggested by qualitative results (Fig.2.4), our multi-phase hierarchical multi-scale supervoxel-based classification method (`hSLIC`) outperformed both voxel-wise (`Vx`) and single-scale supervoxel-wise (`sSLIC`) RF for delineating parenchyma, active and necrotic areas. In addition, we outlined in **[19]** the benefits reached by multi-phase features exploiting the temporal dynamics of tissues, HCC being characterized by arterial enhancement followed by venous washout. While ensuring weak interaction efforts, our methodology reached a TN rate error of 4.08% on a collected in-house DCE-CT dataset.

2.3 Deep learning: a paradigm shift

A crucial step to determine the optimal decision boundary in high-dimensional feature space is the extraction of discriminant features from images. In a machine learning set-up (Sect.2.2), this process is performed by humans building so-called handcrafted features. In recent years, the logical next step has been to let computers learn the features that optimally represent the data under investigation [22]. This concept lies at the basis of deep learning whose foundations are described in this section.

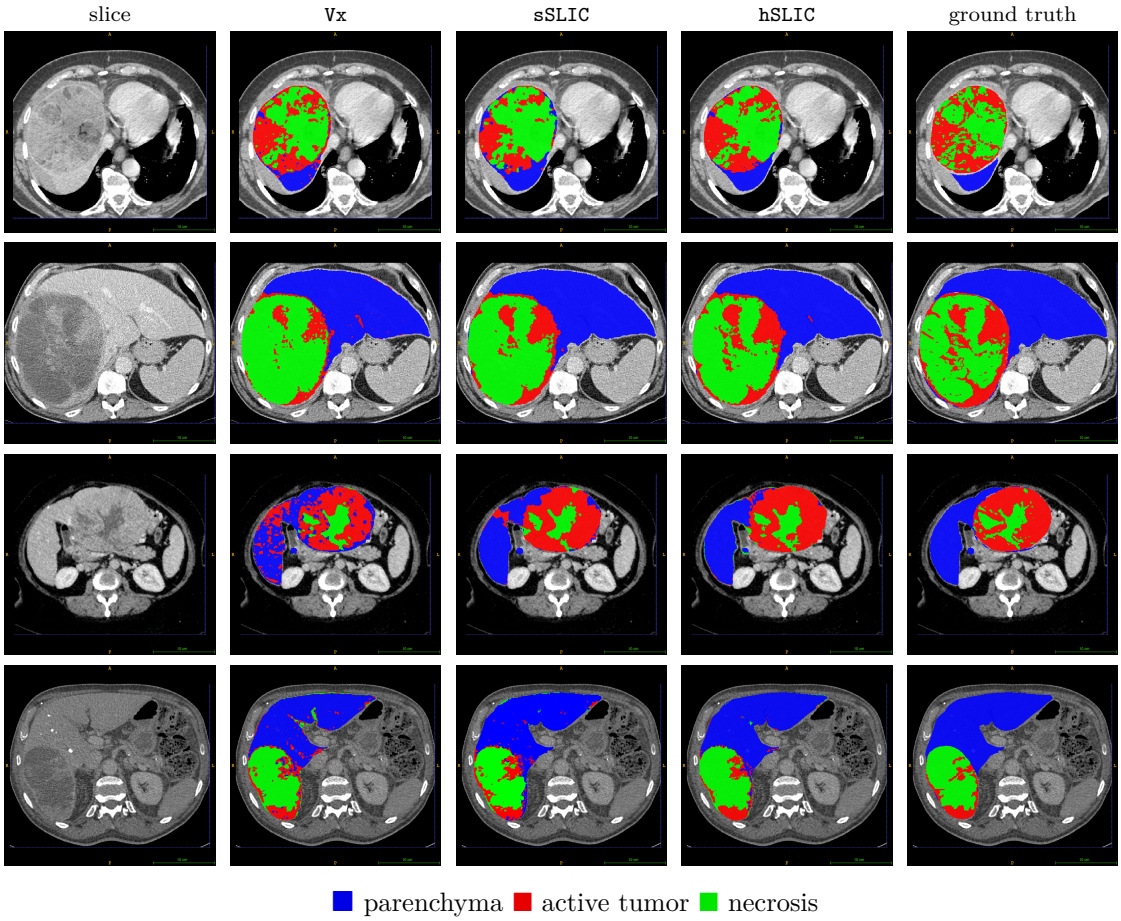


Figure 2.4 – Liver tumor segmentation results from **[19]** via multi-phase voxel-wise (`Vx`), single-scale supervoxel-wise (`sSLIC`) and hierarchical multi-scale supervoxel-wise (`hSLIC`) random forest.

2.3.1 Problem formulation

Let $\mathcal{X}$ be a set of images $\boldsymbol{x} \in \mathbb{R}^{H \times W \times D}$ where H, W and D are the image dimensions in x-, y- and z- axis while the annotation set $\mathcal{Y} \subset [0,1]^{H \times W \times D \times C}$ contains for each $\boldsymbol{x} \in \mathcal{X}$ a map $\boldsymbol{y}$ of $H \times W \times D$ one-hot vectors indicating the ground truth class for all voxels. In a fully-supervised setting, a deep segmentation network ϕ aims at approximating a mapping function $\phi : \boldsymbol{x} \rightarrow \phi(\boldsymbol{x}; \Theta_\phi) = \hat{\boldsymbol{y}}$ between intensity $\boldsymbol{x}$ and class labels $\boldsymbol{y}$ images from N training samples $\{\boldsymbol{x}_n, \boldsymbol{y}_n\}_{1 \leq n < N}$ by optimizing a loss function $\mathcal{L}_\phi(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \dfrac{1}{N} \displaystyle\sum_{n=1}^{N} \ell_\phi(\boldsymbol{y}_n, \hat{\boldsymbol{y}}_n)$ with $\hat{\boldsymbol{y}}_n = \phi(\boldsymbol{x}_n)$ through an optimizer. The parameters of ϕ, namely Θ_ϕ, are optimized during training. A stochastic gradient descent scheme aims at finding the optimal weights Θ_ϕ^* such that $\Theta_\phi^* = \arg\min_{\Theta_\phi} \mathcal{L}_\phi(\boldsymbol{y}, \hat{\boldsymbol{y}})$. Network weights are iteratively updated in the direction of the steepest descent to reach the local minimum:

$$\Theta_\phi \leftarrow \Theta_\phi - \alpha \nabla_{\Theta_\phi} \mathcal{L}_\phi \tag{2.2}$$

where the learning rate α is a hyper-parameter controlling the step size at each iteration. Tuning α is of paramount importance to find a good trade-off between convergence speed and stable optimization. Back-propagation deals with gradient computation, while the gradient descent algorithm, based on this gradient, aims at performing the learning procedure. ℓ_ϕ is a per-image loss function which is usually the cross-entropy loss defined, in a multi-class scenario with C classes, following:

$$\ell_{CE}(\boldsymbol{y}_n, \hat{\boldsymbol{y}}_n) = \frac{1}{|\mathscr{C}||\Omega|} \sum_{c \in \mathscr{C}} \sum_{u \in \Omega} -\boldsymbol{y}_{n,c,u} \log(\hat{\boldsymbol{y}}_{n,c,u}) \tag{2.3}$$

where Ω is the image grid and $c \in \mathscr{C} = \{0, ..., C\}$ indexing the structures of interest and the background. A wide variety of loss functions exist [23]: distribution-based (e.g. cross-entropy), region-based (e.g. Dice), compound (e.g. DiceCE) or boundary-based (e.g. Hausdorff distance). To date, the most widely used architectures deal with convolutional neural networks (CNN). Mainly composed of convolutional and pooling layers, CNN have become the state-of-the-art in numerous image analysis tasks due to their ability to learn hierarchical representations of image features in a purely data-driven way.

2.3.2 From seminal works to UNet

The simplest and early attempts to perform segmentation using CNN consisted in classifying each pixel individually in a patch-based manner [24]. Since input patches from neighboring pixels have large overlaps, the same convolutions were computed many times. By replacing fully-connected layers with convolutional layers, fully convolutional networks (FCN) gave the opportunity to take entire images as inputs and produce likelihood maps instead of single-pixel outputs. It removed the need to select representative patches and eliminated redundant calculations. To avoid outputs with far lower resolution than input shapes, FCN were applied to shifted versions of images [25]. The multiple outputs were stitched together to get results at full resolution. Further improvements were proposed with architectures comprising a regular FCN to extract features, followed by an up-sampling part that enables the recover the input resolution using up-convolutions. Compared to patch or shift-and-stitch methods, a precise localization was possible in a single pass while considering the full image context. This motivated the interest for convolutional encoder-decoders among which UNet [26] is the most widely used thank to its ability to output detailed contours using a quite low amount of training data.

Among existing convolutional encoder-decoders architecture, most deep learning-based medical image segmentation models rely on UNet [26] or its 3D counterpart VNet [27]. UNet is a symmetrical architecture comprising an encoder that gradually reduces spatial dimensions using pooling layers, a decoder progressively recovering object details and initial resolution as well as skip-connections (i.e. long-range shortcuts) which concatenate features between contracting and expanding paths to help in improving localization accuracy and convergence speed. The contracting path encoder of a standard UNet consists of sequential layers including 3×3 convolutional layers followed by batch normalization (BN) and rectified linear unit (ReLU) activations (Fig.2.5). Spatial size is reduced using 2×2 max-

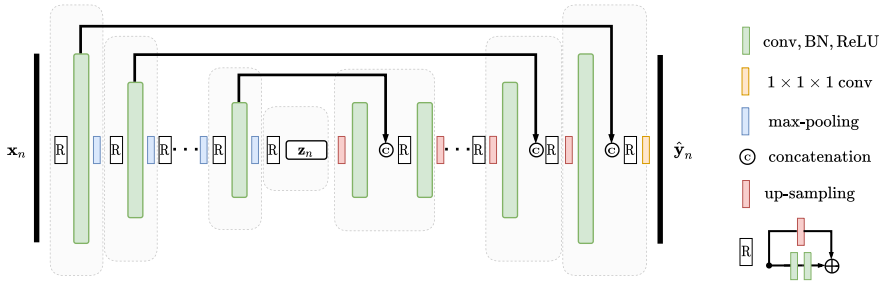


Figure 2.5 – VNet inspired [27] convolutional encoder-decoder for medical image segmentation.

pooling layers. The first convolutional layer typically generates 32 or 64 channels. This number doubles after each pooling as the network deepens. The encoder finally projects each input image $\boldsymbol{x}_n$ to a latent representation ($\boldsymbol{z}_n$ in Fig.2.5). On its turn, the decoder is built symmetrically with respect to the encoder, except that max-pooling layers are replaced by up-sampling operations (e.g. transpose convolution). In a binary (multi-class) problem, a final 1×1 convolutional layer with sigmoid (softmax) activation achieves pixel-wise segmentation $\hat{\boldsymbol{y}}_n = \phi(\boldsymbol{x}_n)$, at native resolution. VNet inspired models may more suffer from higher computational cost and GPU memory usage than their 2D counterparts.

2.4 Multi-organ segmentation using deep learning

The development of non-invasive imaging technologies over the last decades has opened new horizons in studying the abdominal anatomy. Computational abdominal analysis from CT or MR scans has become a crucial task for various applications: computer-assisted diagnosis, surgery planning (e.g. organ pre-evaluation for resection) or image-guided interventions [28]. However, despite intensive developments in deep learning, it remained difficult to judge the effectiveness of deep networks for abdominal multi-organ segmentation since they are mainly assessed on one single organ (liver most often), one single modality (usually CT) and relatively small or private datasets. Their robustness to delineate multiple abdominal structures from different modalities and to manage inter-subject variability was therefore under-investigated. In this direction and in conjunction with the organization of both Combined Healthy Abdominal Organ Segmentation (CHAOS) **[29]** and Fast and Low GPU memory Abdominal oRgan sEgmentation (FLARE) **[30]** challenges, my research activities have focused on abdominal multi-organ segmentation (e.g. liver, kidneys, spleen) with deep learning. Two methodological avenues were mainly studied: adversarial training with cascaded generators using auto-context (Sect.2.4.1) as well as deep supervision towards fast and low-GPU-memory algorithms (Sect.2.4.2).

2.4.1 Cascaded convolutional and adversarial networks

Conditional generative adversarial networks (GAN) are known to be a general-purpose solution for image-to-image translation [31]. Applied to segmentation purposes, conditional GAN architectures

comprise a generator aiming at providing segmentation masks through encoding and decoding layers as well as a discriminator which assesses if a given segmentation mask is synthetic or real. The adversarial network learns to discriminate real from synthetic delineations, i.e. ground truth masks versus those arising from the generator. This enforces the generative part to create increasingly plausible segmentation masks. During the training process, the generated delineations are gradually close to the ground truth, to the point of being able to deceive the discriminator.

As generator ϕ, conditional GAN pipelines may use any type of UNet [26] inspired architecture. The inputs of the discriminator D are the concatenation of source images and ground truth or predicted masks to be evaluated. Defined between 0 (i.e. fake) and 1 (i.e. plausible or real), the output of D is an array where each value corresponds to the degree of segmentation likelihood for a given image crop and its associated segmentation mask. Let $\phi(\boldsymbol{x})$ and $D(\boldsymbol{x}, \phi(\boldsymbol{x}))$ be the outputs of ϕ and D respectively. The loss function $\mathcal{L}_\phi(\Theta_\phi, \Theta_D)$ for the generator ϕ can be defined as the following combination:

$$\mathcal{L}_\phi(\Theta_\phi, \Theta_D) = \frac{1}{N}\sum_{n=1}^{N} \ell_{CE}(\phi(\boldsymbol{x}_n), \boldsymbol{y}_n) + \lambda \times \ell_{adv}(\phi(\boldsymbol{x}_n), \boldsymbol{y}_n) \tag{2.4}$$

where λ is an empirically set weighting factor, Θ_ϕ and Θ_D the trainable parameters of ϕ and D, respectively. The adversarial term $\ell_{adv}(\phi(\boldsymbol{x}_n), \boldsymbol{y}_n)$ equals to $-\log(D(\boldsymbol{x}_n, \phi(\boldsymbol{x}_n)))$. Minimizing ℓ_{CE} tends to provide rough predictions whereas maximizing $\log D(\boldsymbol{x}_n, \phi(\boldsymbol{x}_n))$ aims at improving contour delineations. Conversely, the optimizer typically fits D through cross-entropy using estimated and ground truth masks. The loss function $\mathcal{L}_D(\Theta_\phi, \Theta_D)$ for D is defined as:

$$\mathcal{L}_D(\Theta_\phi, \Theta_D) = \frac{1}{N}\sum_{n=1}^{N} -\log(D(\boldsymbol{x}_n, \boldsymbol{y}_n)) - \log(1 - D(\boldsymbol{x}_n, \phi(\boldsymbol{x}_n))) \tag{2.5}$$

Optimizing $\mathcal{L}_D(\Theta_\phi, \Theta_D)$ means maximizing loss values for ground truth, $\log(D(\boldsymbol{x}, \boldsymbol{y}))$, and minimizing loss values for generated masks, $-\log(1 - D(\boldsymbol{x}, \phi(\boldsymbol{x})))$. The optimization process is performed sequentially by alternating gradient descents on ϕ and D, at each batch.

To further improve the ability of conditional GAN architectures to extract the targeted structures, we performed investigations on more robust generators than the traditional UNet [26]. In particular, we extended in **[32]** standard segmentation networks to a cascade of partially pre-trained deep convolutional encoder-decoders exploiting multi-level contextual information through auto-context and end-to-end training. Such model was used as generator ϕ in a conditional GAN to further encourage the generative part to provide plausible organ delineations. More precisely, instead of increasing *ad-infinitum* the network depth to exploit larger receptive fields which is not suitable for memory and computational issues, we combined two scale-specific and partially pre-trained networks with auto-context [15], i.e. using posterior probabilities resulting from the first network as features for the second one. The sigmoid activation of the first networks used in the last convolution layer was replaced by a linear function to generate continuous output maps. These maps were normalized, concatenated to source images and given as inputs of the second network which was trained to give final

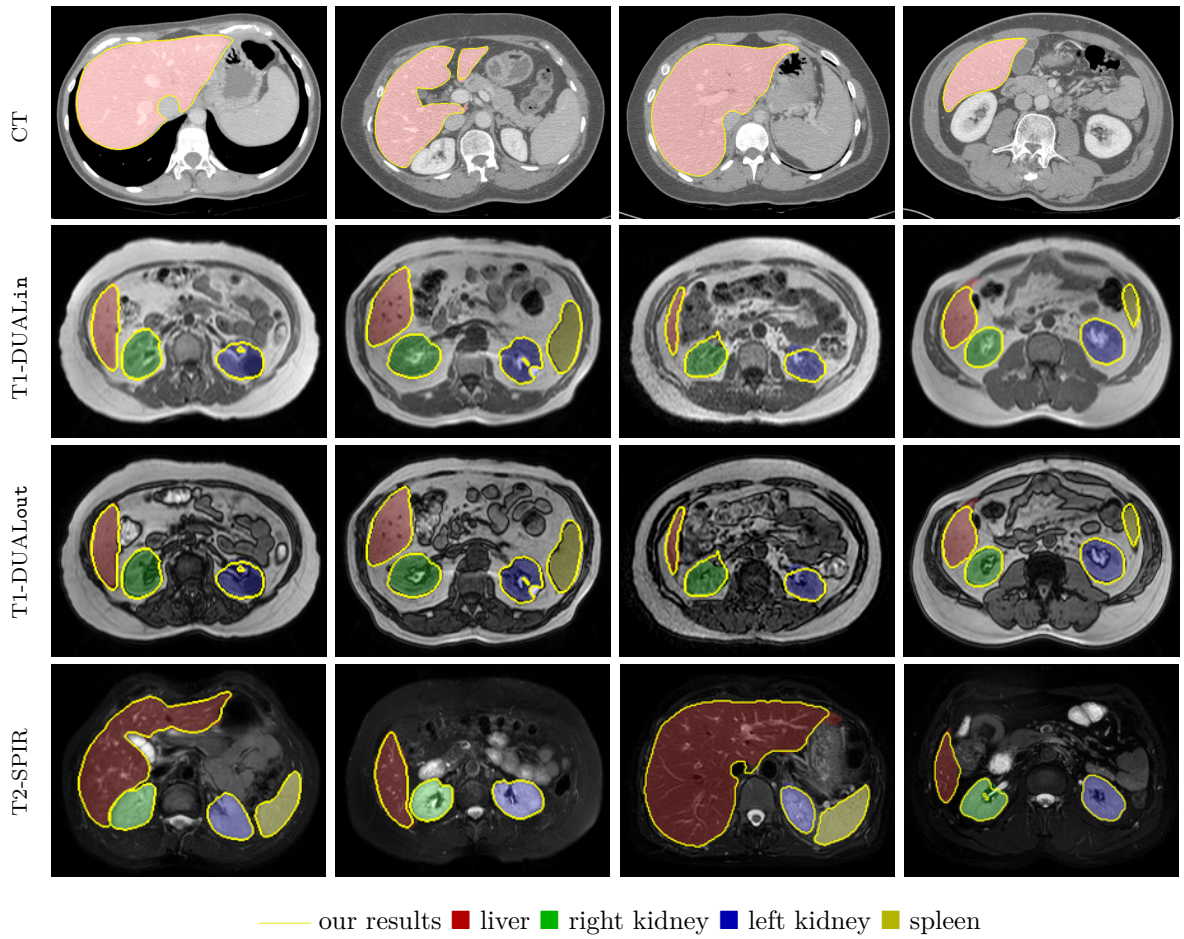


Figure 2.6 – Liver CT and abdominal multi-organ MR (T1-DUAL`in/out`, T2-SPIR) segmentation using the methodology proposed in **[32]**. Liver, right kidney, left kidney and spleen ground truth delineations are superimposed in red, green, blue and yellow colors, respectively.

organ delineations. Making the first network generating continuous instead of binary outputs propagated pixel-wise confidence information to the second network and postponed the final decision to the pipeline ending part. Instead of training both models separately, our pipeline was trained end-to-end to exploit simultaneous multi-level segmentation refinements. Each network of the cascade was partially pre-trained on ImageNet [33] since encoder fine-tuning from a large amount of non-medical images is known **[34]** to improve predictive performance while alleviating data scarcity limitations.

Employed for healthy liver, kidneys and spleen segmentation, our pipeline provided promising results (Fig.2.6) by outperforming state-of-the-art encoder-decoder schemes **[32]**. Followed for the CHAOS[3] challenge organized in conjunction with the IEEE International Symposium on Biomedical Imaging (ISBI) 2019, it gave us the first rank for three competition categories: liver CT, liver MR and multi-organ MR segmentation **[29]**. This achievement further proved that combining cascaded

3. `https://chaos.grand-challenge.org`

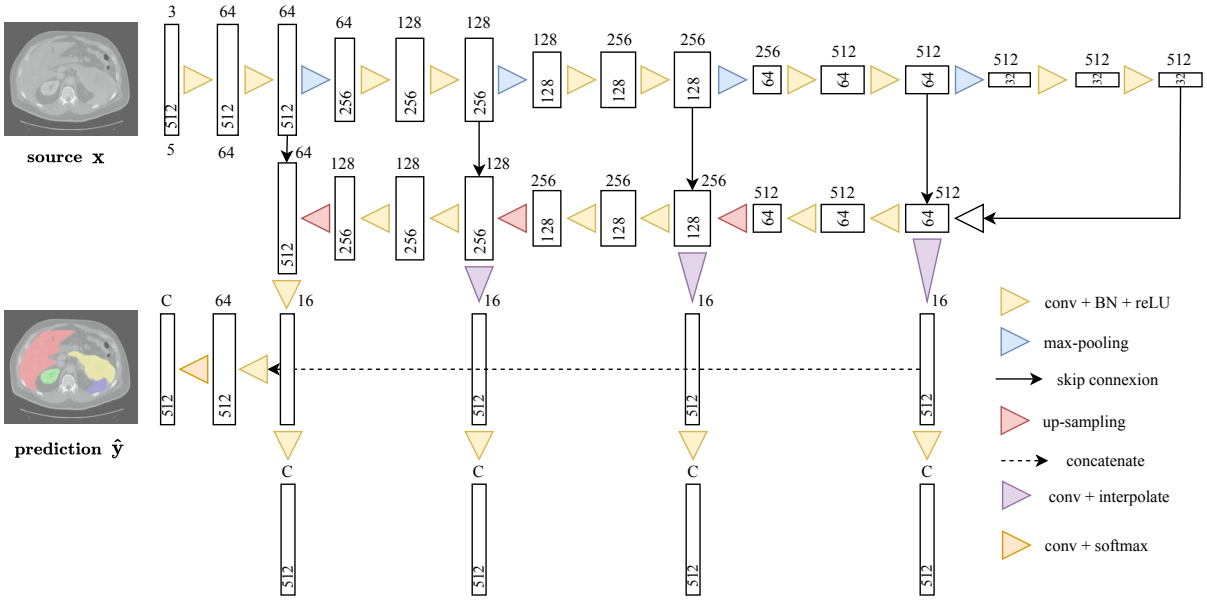


Figure 2.7 – Convolutional encoder-decoder with deep supervision used for the FLARE challenge **[30]**. The overall loss combines losses estimated at different decoder levels. C is the number of classes.

convolutional and adversarial networks strengthens the ability of deep learning segmentation pipelines to automatically delineate multiple abdominal organs, with good generalization capability.

2.4.2 Fast and low-GPU segmentation using deep supervision

My participation in the FLARE[4] challenge, organized as part of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2021, led to address multi-organ segmentation from abdominal multi-center, multi-phase, multi-vendor and multi-disease CT examinations using deep learning **[30]**. To meet clinical requirements while obtaining a deep model at low computational cost, the deep architecture was built around a lightweight encoder from the VGG family [35] and a symmetrically-designed decoder. To improve the gradient flow, promote the extraction of discriminative features and reach a good compromise between computational complexity and performance, our model (Fig.2.7) was based on multi-level deep supervision.

Introduced in the context of holistically-nested edge detection [36], additional convolutional operations can be applied at different levels of the decoder branch to exploit a deep supervision mechanism (Fig.2.7) able to boost the segmentation performance. Companion objective functions were estimated at some hidden layers of the network and added to the output loss. In practice, feature maps as outputs of each intermediate decoder blocks were up-sampled to the size of the input image, similarly to [37]. These maps then went through deep supervision modules to encourage learning more useful represen-

4. `https://flare.grand-challenge.org`

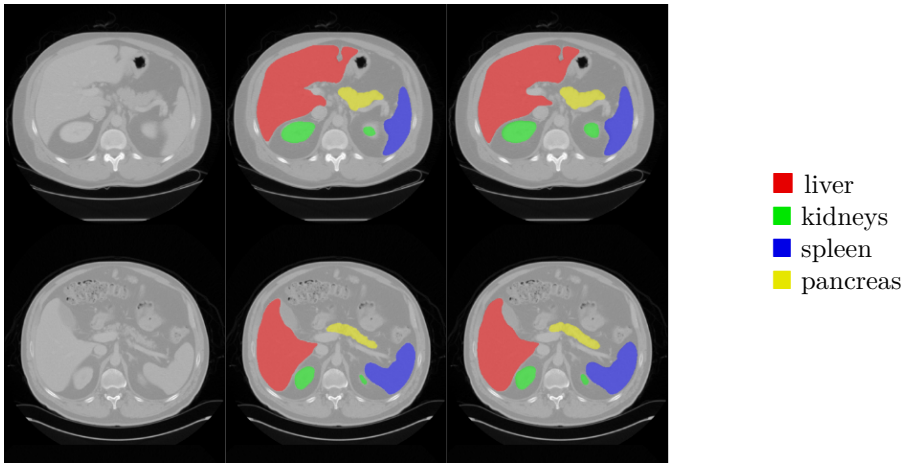


Figure 2.8 – Abdominal multi-organ CT segmentation using the approach developed for the FLARE challenge **[30]**. Results display, from left to right, source images, ground truth and predictions.

tations. After having performed the concatenation of these intermediate outputs, convolutional layers followed by softmax activation finally achieved multi-label segmentation to delineate liver, kidneys, spleen and pancreas organs. The overall loss function $\mathcal{L}_\phi$ was defined as the weighted sum of the cross-entropy losses estimated at different decoder levels involving supervision:

$$\mathcal{L}_\phi(\Theta_\phi) = \sum_{j=1}^{M} w_j \times \mathcal{L}_{CE}^j + w_f \times \mathcal{L}_{CE}^f \tag{2.6}$$

where w_j and $\mathcal{L}_{CE}^j$ denote the weight and loss for the points of supervision at level j of the decoder, w_f and $\mathcal{L}_{CE}^f$ the weight and loss computed at the final network output. By using VGG-13 [35] and $M = 4$ intermediate decoder levels, we used $w_1 = 0.8$, $w_2 = 0.7$, $w_3 = 0.6$, $w_4 = 0.5$ and $w_f = 1$ where level $j = 1$ is closer to the network ending part than level $j > 1$. Our approach enabled us to reach accurate delineations (Fig.2.7) and to take the 8th place of the challenge, with honorable mention.

2.5 Prior knowledge embedding

Regularization plays a key role in deep learning since it tends to increase robustness and generalizability of deep models when applied to unseen data. One common strategy consists in adding a regularization term into the loss function to get more accurate and plausible results [38]. The regularization term $\mathcal{R}_\phi$ deals with adding some prior knowledge to the model ϕ and its regularization effect is achieved by incorporating the scaled regularizer $\lambda \times \mathcal{R}_\phi$ to the loss function $\mathcal{L}_\phi$ to ensure further consistency between predictions $\phi(\boldsymbol{x})$ and targets $\boldsymbol{y}$. The resulting loss function is expressed as follows:

$$\mathcal{L}_\phi(\Theta_\phi) = \frac{1}{N} \sum_{n=1}^{N} \ell_\phi(\phi(\boldsymbol{x}_n), \boldsymbol{y}_n) + \lambda \times \mathcal{R}_\phi \tag{2.7}$$

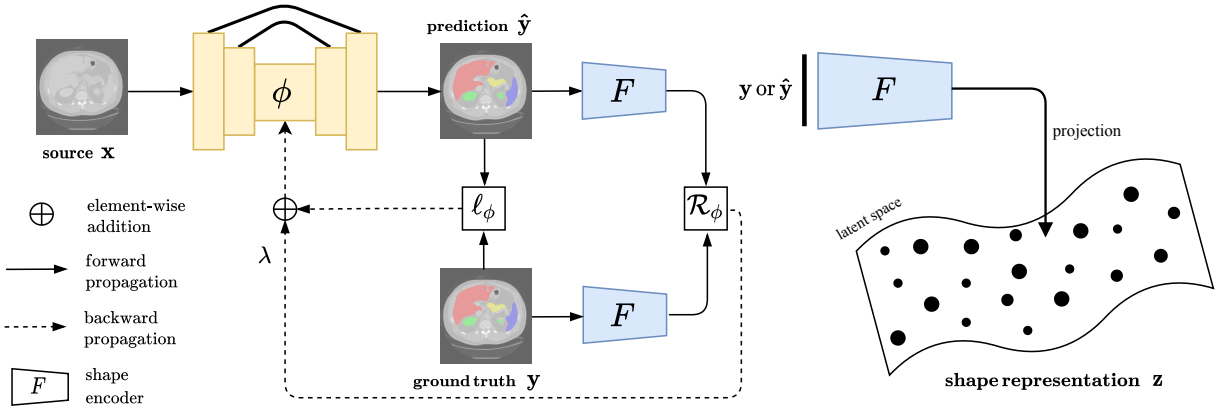


Figure 2.9 – Integration of shape priors into a deep segmentation pipeline [39]. Shape priors-based regularization is performed using a shape encoder F arising from a convolutional auto-encoder.

Many different types of information can be incorporated as prior knowledge into deep frameworks: shape constraints [39], topology specifications [40], adjacency rules between regions [41]... Nevertheless, integrating shape priors remains one of the most commonly used strategies towards anatomically meaningful predictions. Especially, the relevancy of using a convolutional auto-encoder to learn anatomical shape variations from medical images has been demonstrated in multiple applications [39], [42]. A convolutional auto-encoder is a deep network made of an encoder $F : \boldsymbol{y} \mapsto F(\boldsymbol{y}; \Theta_F)$ and a decoder $G : F(\boldsymbol{y}; \Theta_F) \mapsto G(F(\boldsymbol{y}; \Theta_F); \Theta_G)$ where Θ_F and Θ_G correspond to the learnable parameters of F and G. F maps the input to a low-dimensional feature space whereas G reconstructs the original input from the compact representation. To avoid the auto-encoder to copy the input, F is usually designed to be under-complete such that the latent space is much smaller than the input dimension. By penalizing the reconstruction $G \circ F(\boldsymbol{y})$, the cross-entropy loss can be used to optimize the auto-encoder:

$$\Theta_F^*, \Theta_G^* = \underset{\Theta_F, \Theta_G}{\arg\min} \frac{1}{N} \sum_{n=1}^{N} \ell_{CE}((G \circ F)(\boldsymbol{y}_n), \boldsymbol{y}_n) \tag{2.8}$$

with Θ_F^* and Θ_G^* the optimal weights for the encoder and decoder. By training the auto-encoder on ground truth segmentation masks, its encoder finally acts as a non-linear shape model which can project any predicted ($\phi(\boldsymbol{x})$) or ground truth ($\mathbf{y}$) segmentation to a shape manifold space. Once the auto-encoder trained, its encoder component (i.e. F) can be integrated into the segmentation pipeline (Fig.2.9). A regularizer $\mathcal{R}_\phi$ that penalizes the deviation between predicted and ground truth segmentation masks fed as inputs of the learned shape model F is included into the global loss (Eq.2.7). A Euclidean distance between both latent shape representations [39] is usually employed, following:

$$\mathcal{R}_\phi = \ell_{shape}(\hat{\boldsymbol{y}}, \boldsymbol{y}) = \frac{1}{N} \sum_{n=1}^{N} \|F(\boldsymbol{y}_n; \Theta_F^*) - F(\phi(\boldsymbol{x}_n); \Theta_F^*)\|_2^2 \tag{2.9}$$

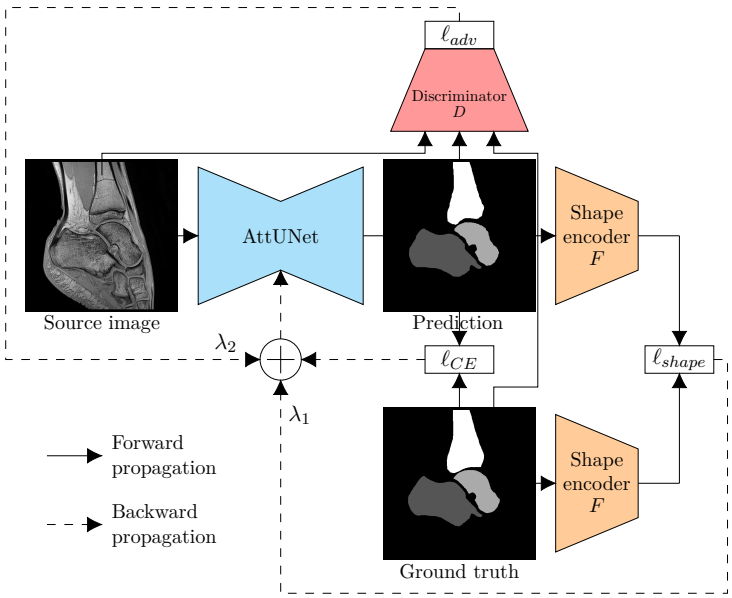


Figure 2.10 – Segmentation framework built in **[43]** around AttUNet [44] and exploiting cross-entropy loss ℓ_{CE}, shape priors-based ℓ_{shape} and adversarial ℓ_{adv} regularizations.

2.5.1 Shape priors and adversarial regularization

In this direction and as part of the PhD thesis work of A. Boutillon, we have conceived a novel optimization scheme to train a segmentation network with two regularization terms into the loss function **[43]**. First, in order to obtain globally consistent predictions, we incorporated a shape priors-based regularization, derived from a non-linear shape representation learnt by an auto-encoder [39]. Second, an adversarial regularization computed by a discriminator (Sect.2.4.1) was integrated to encourage precise delineations. Thus, the framework comprised a segmentation network ϕ, an auto-encoder whose encoder component F was employed to benefit from shape priors as well as a discriminator D for adversarial training. The proposed regularized segmentation framework (Fig.2.10) was based on attention UNet [44] (AttUNet) and exploited cross-entropy loss ℓ_{CE} (Eq.2.3), shape priors-based ℓ_{shape} (Eq.2.9) and adversarial ℓ_{adv} regularizations respectively computed by a shape encoder F with fixed weights and a discriminator D trained in competition with AttUNet. The shape encoder corresponds to the encoder of an auto-encoder previously optimized on ground truth delineations, while the discriminator learns the plausibility of segmentation masks conditioned by their corresponding intensity image. Combining shape priors-based and adversarial regularizations led to the following loss formulation:

$$\mathcal{L}_\phi(\Theta_\phi) = \frac{1}{N}\sum_{n=1}^{N} \ell_{CE}(\phi(\boldsymbol{x}_n), \boldsymbol{y}_n) + \lambda_1 \times \ell_{shape}(\phi(\boldsymbol{x}_n), \boldsymbol{y}_n) + \lambda_2 \times \ell_{adv}(\phi(\boldsymbol{x}_n), \boldsymbol{y}_n) \tag{2.10}$$

where the adversarial term $\ell_{adv}(\phi(\boldsymbol{x}_n), \boldsymbol{y}_n)$ equals to $-\log(D(\boldsymbol{x}_n, \phi(\boldsymbol{x}_n)))$ whereas λ_1 and λ_2 are two empirical weighting hyper-parameters. Our method was evaluated for multi-bone segmentation on two scarce pediatric MR imaging datasets from ankle and shoulder joints, comprising pathological and

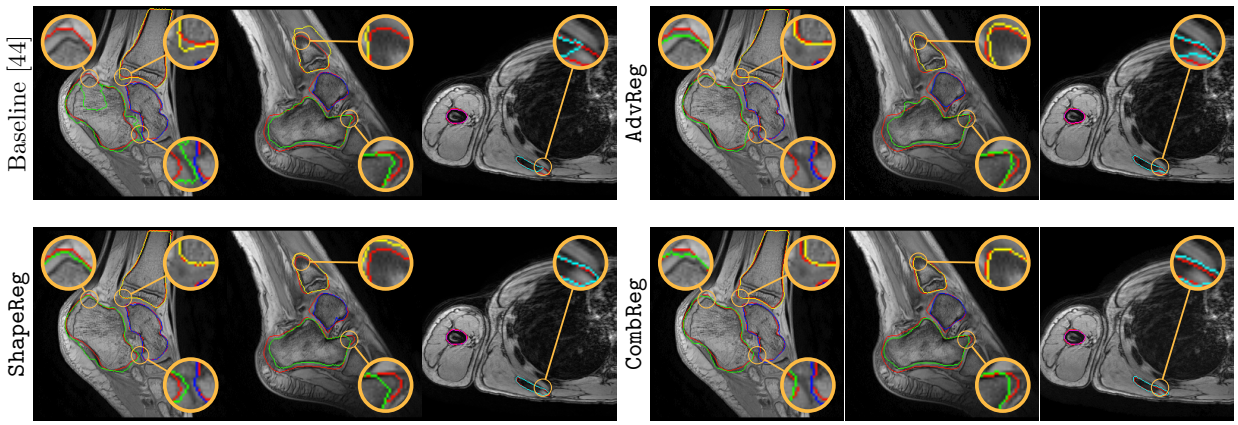


Figure 2.11 – Segmentation of ankle/shoulder bones via regularized segmentation networks based on AttUNet [44] exploiting cross-entropy ℓ_{CE} (`Baseline`), shape priors ℓ_{shape} (`ShapReg`) [39], adversarial ℓ_{adv} **[32]** (`AdvReg`) and combined **[43]** (`ComReg`) regularizations. Ground truth in red (–), calcaneus, talus, tibia, humerus and scapula in green (–), blue (–), yellow (–), magenta (–) and cyan (–).

healthy examinations. Comparisons between regularization approaches (Fig.2.11) proved the relevancy of combining shape priors-based and adversarial regularizations.

2.5.2 Adversarial shape priors

To go further, the PhD thesis of A. Boutillon led to the design of a shape code discriminator trained in an adversarial fashion against the segmentation network **[45]**. Our shape priors-based adversarial contribution encouraged the segmentation network to follow global anatomical properties of the shape representation by guiding prediction masks closer to ground truth segmentation in latent shape space. Thus, we adapted the standard adversarial regularization (Sect.2.5.1) to the latent shape representation arising from segmentation masks instead of segmentation masks themselves. In particular, we designed a novel shape code discriminator $D : \boldsymbol{z} \mapsto \{0, 1\}$ which assesses if an input latent shape code $\boldsymbol{z}$ (i.e. $F(\boldsymbol{y})$ or $F(\phi(\boldsymbol{x}))$) corresponds to a synthetic or real delineation mask. The architecture of such discriminator consisted of a succession of convolutional filters followed by BN, max-pooling and ReLU layers with a final sigmoid activation layer computing the likelihood of $\boldsymbol{z}$ being fake (0) or real (1).

The training scheme optimized the shape code discriminator and the segmentation network alternatively (Fig.2.12). The discriminator was trained to differentiate latent codes using binary cross-entropy $\ell_D = -\log(1 - D(\hat{\boldsymbol{z}})) - \log(D(\boldsymbol{z}))$ with $\hat{\boldsymbol{z}} = F(\phi(\boldsymbol{x}))$ and $\boldsymbol{z} = F(\boldsymbol{y})$. At the segmentation network optimization step, the shape code discriminator computed the shape priors based adversarial regularization term $\ell_{SPAR}(\hat{\boldsymbol{y}}) = -\log(D(\hat{\boldsymbol{z}}))$ which represents the probability that the network considers the generated shape codes to be ground truth latent codes. Thus, the segmentation training strategy was modified as follows: $\ell_\phi = \ell_{CE}(\hat{\boldsymbol{y}}, \boldsymbol{y}) + \lambda \times \ell_{SPAR}(\hat{\boldsymbol{y}})$ with λ a weighting hyper-parameter. The optimization of ℓ_{SPAR} enforced ϕ to fool the discriminator and generate delineations whose latent representation is close to the ground truth shape representation.

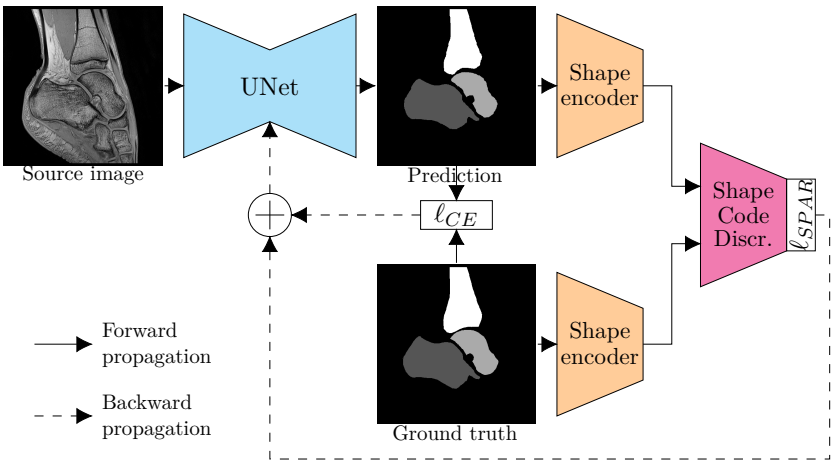


Figure 2.12 – Segmentation framework proposed in **[45]** based on UNet, a shape representation learnt from a convolutional auto-encoder and a shape code discriminator. The segmentation network exploits cross-entropy ℓ_{CE} and shape priors-based adversarial regularization ℓ_{SPAR}.

Promising results were obtained in **[45]** for multi-structure ankle bone MR segmentation, with gradual improvements from the UNet baseline to the shape priors-based adversarial regularization.

2.5.3 Semi-overcomplete shape priors

Several approaches had recently focused on the exploration of more sophisticated deep architectures than the standard UNet [26]. Especially, overcomplete architectures [48] have appeared with the goal of projecting data onto higher dimensions to constrain the receptive field to be small and therefore capture finer low-level features details. The ability of over-complete architectures to encode small anatomical structures appeared of high interest for vessel extraction purposes. Since modeling small vascular structures with standard convolution auto-encoders [39] **[43], [45]** does not guarantee a fully efficient shape representation in latent space, we proposed in the framework of the PhD thesis of A. Sadikine to integrate into the segmentation framework the encoder of a novel semi-overcomplete convolutional auto-encoder (S-OCAE, Fig.2.13) with a multi-path encoder leveraging both non-linear under and overcomplete representations of the multi-scale vascular tree geometry **[46]**.

Compared to standard auto-encoder, an overcomplete convolutional auto-encoder is obtained by replacing max-pooling by up-sampling layers and vice-versa. In this scenario, intermediate layers are projected to a higher dimensionality than the input, giving to the model a better flexibility to encode small details. In practice, we designed a multi-path encoder composed of both undercomplete and overcomplete branches (Fig.2.13). Since an overcomplete encoder would have been too expensive in terms of memory, we proposed a semi-overcomplete branch which does not take the input image as input but features from an intermediate hidden layer. A specific block named communication block (CB) was designed to communicate features between both branches. To combine the outputs

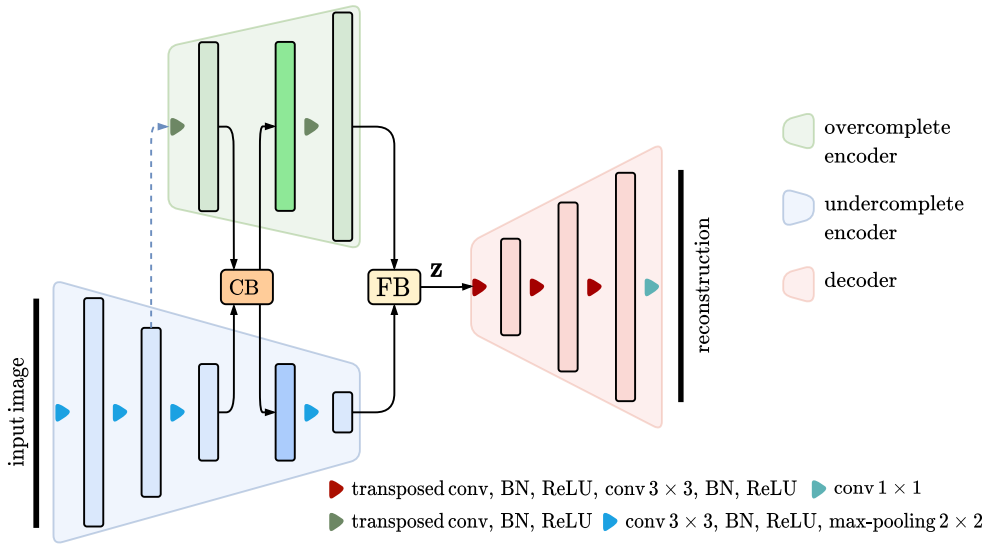


Figure 2.13 – Semi-overcomplete convolutional auto-encoder network proposed in **[46]** with multi-path encoder made of undercomplete and overcomplete branches.

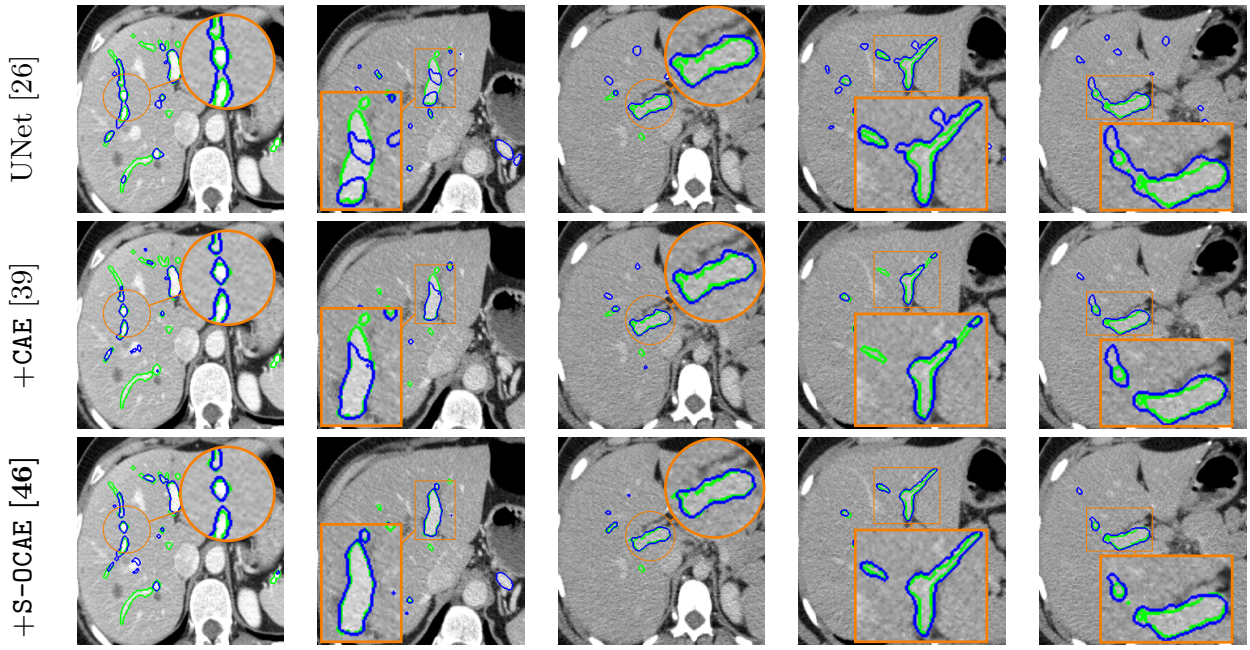


Figure 2.14 – Hepatic portal veins segmentation on 3D-IRCADb [47] using UNet [26] with standard [39] and semi-overcomplete **[46]** shape priors. Ground truth in green, predictions in blue.

from undercomplete and overcomplete encoding branches, a fusion block (FB) was introduced at the bottleneck of the network to compute a latent code $\boldsymbol{z}$. The encoder of the S-OCAE architecture was integrated in the segmentation pipeline, as in Fig.2.9, by replacing the standard encoder F by the designed multi-path encoder. The effectiveness of our semi-overcomplete shape priors was illustrated for liver vessel extraction (Fig.2.14). A better ability to extract small structures was revealed.

2.5.4 Joint shape and topological priors

Most existing works related to prior knowledge embedding focus on incorporating prior knowledge of a single type (e.g. shape priors [39] **[45], [46]** or topology constraints [40] only). Conversely, simultaneously incorporating multiple priors in medical imaging segmentation has not received much attention to date. Incorporating multiple anatomical prior-based loss functions into the segmentation pipeline typically requires the use of multiple individual non-linear encodings. The viability of such strategy may pose challenges as it necessitates training multiple auto-encoders and tuning multiple hyper-parameters for the prior penalty terms in the loss function. In addition, it can lead to higher memory consumption during training. To address these drawbacks, we developed in the framework of the PhD thesis of A. Sadikine a novel approach referred to as joint multi-prior encoding (JMPE) that involved learning both shape and topological connectivity priors within a unified manifold **[49]**.

Topological connectivity refers to the arrangement and connection of components within the structure to segment. It involves analyzing how different parts of the structure are connected (e.g. bifurcations or endpoints in vessel delineation). This connectivity can be effectively captured by using the Euclidean distance transform (EDT) to convert any ground truth mask $\boldsymbol{y}_n$ into a grayscale image $\boldsymbol{T}_n$ where the voxel values represent the distance from that voxel to the nearest boundary voxel.

The pursuit of learning multiple priors in a unified compact representation $\boldsymbol{z}$ stands as a more efficient alternative than employing separate encodings. This challenge was addressed through a convolutional auto-encoder ξ trained in a multi-task learning fashion (Fig.2.15). ξ comprised a single encoder F and multiple decoders D_p, all sharing the same latent code representation $\boldsymbol{z}$ and offering a streamlined approach to jointly managing multiple priors representation in a single latent space. The optimization of the multi-task auto-encoder ξ was performed following:

$$\xi(\boldsymbol{y}) = \{D_s(\boldsymbol{z}) = \tilde{\boldsymbol{y}}, D_t(\boldsymbol{z}) = \tilde{\boldsymbol{T}} \mid \boldsymbol{z} = F(\boldsymbol{y})\} \tag{2.11}$$

where D_s and D_t are dedicated to the tasks of reconstruction and regression, respectively. The optimal model ξ was achieved by minimizing the following loss across both training tasks:

$$\mathcal{L}_{JMPE}(\boldsymbol{y}, \tilde{\boldsymbol{y}}) = \alpha_s \sum_{\boldsymbol{y}} \ell_{L_1}(\boldsymbol{y}_n, \tilde{\boldsymbol{y}}_n) + \alpha_t \sum_{\boldsymbol{T}} \ell_{L_1}(\boldsymbol{T}_n, \tilde{\boldsymbol{T}}_n) \tag{2.12}$$

where α_p weighting factors balance both tasks during training with $p \in \{s, p\}$. ℓ_{L_1} was a smooth l_1 distance function. Once ξ trained, F was integrated in the segmentation pipeline (Fig.2.9). Applied for

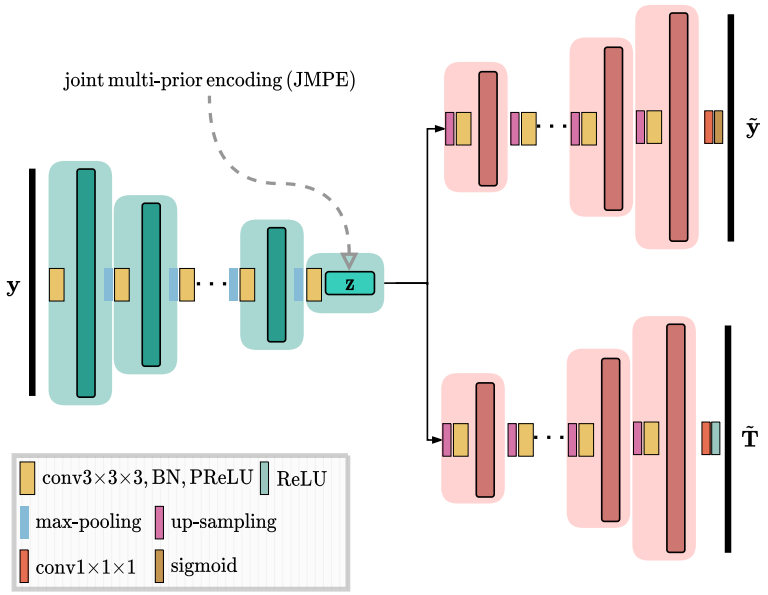


Figure 2.15 – Multi-task convolutional auto-encoder ξ for joint multi-prior encoding (JMPE) **[49]**.

liver vessel segmentation on 3D-IRCADb [47], it delivered robust performance in topological assessment through the `clDice` metric [50], positioning it as a promising topology-aware segmentation model.

2.6 Multi-task segmentation networks

Multi-task learning deals with jointly training a single model to perform multiple related tasks simultaneously. By sharing representations across tasks, it can effectively exploit commonalities among various objectives, thus facilitating improved generalization and reducing overfitting. As suggested in Sect.2.5.4, joint training of multiple tasks can also mitigate issues related to data scarcity since sharing network parameters can enable the model to learn a more efficient and compact representation of the data. This section present two of my past research works involving multi-task segmentation networks.

2.6.1 Scale-specific auxiliary multi-task contrastive learning

Despite a good ability to extract visceral organ contours **[32]**, the ability of deep network to delineate vascular systems from medical images still remains a challenge due to their complex multi-scale geometry. Additionally to contributions related to prior knowledge embedding **[46]**, **[49]** (Sect.2.5), the PhD of A. Sadikine also investigated in **[51]** how to integrate information arising from multiple scales (Fig.2.16) to enable the extraction of relevant information from different levels of granularity and ensure a more nuanced representation of complex structures such as vessel trees.

Typically, blood vessel extraction with deep learning is tackled without taking advantage of multi-task learning whose aim is to improve generalization, learning efficiency and prediction accuracy by

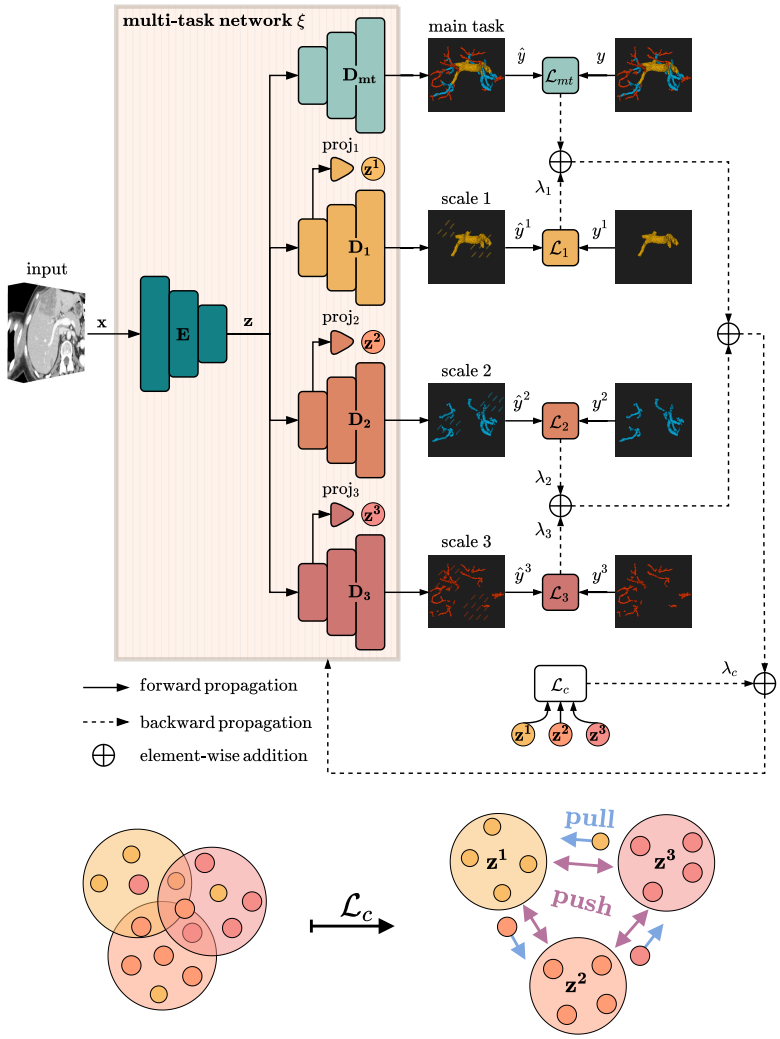
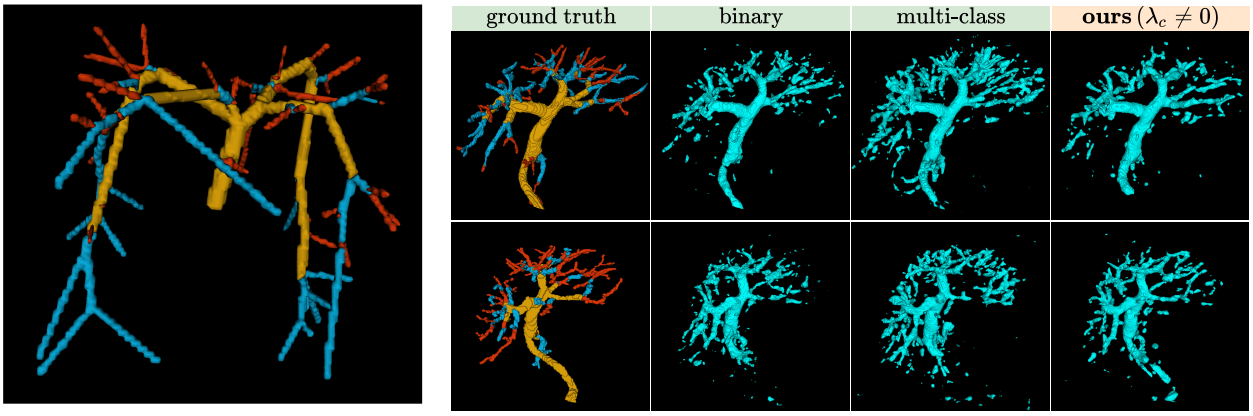


Figure 2.16 – Scale-specific auxiliary multi-task contrastive learning [**51**]. To increase the discriminative power of the shared representation, a contrastive loss $\mathcal{L}_c$ is applied to $\boldsymbol{z}^s$ from each decoder D_s.

leveraging domain-specific information from related tasks. Further, the network architectures related to multi-task learning are diverse and adjusted depending on the specific concerns under investigation [53]. In the setting of single-input and multi-output network, the encoding stage shares its layers to learn a generic representation which is then used as input of independent task-specific decoders. Nevertheless, multi-scale geometry, as an intrinsic characteristic of vascular trees, has been weakly investigated in the literature. This motivated us to design in [**51**] a novel approach considering multiple sub-tasks and providing a strong focus on the different scales arising from vascular networks. Thus, we proposed a multi-scale clustering methodology allowing vasculature decomposition into size-related classes (e.g. small, medium and large vessels) by relying on inherent statistics and estimated branch radii of the vascular tree (Fig.2.17a). Based on such multi-scale vessel clustering, we designed an end-to-end deep



(a) (b)

Figure 2.17 – (a) Example of 3-scale vasculature clustering applied on a synthetic vasculature [52]. (b) Qualitative liver vessel CT segmentation results on 3D-IRCADb [47] dataset using 3D ResUNet in binary and multi-class settings as well as the proposed approach **[51]** with contrastive learning.

multi-task segmentation framework (Fig.2.16) involving multi-scale contrastive learning [54], with the ability to discriminate intra-scale compactness and inter-scale separability.

We denoted $\mathcal{D} = \{(\boldsymbol{x}_n, \{\boldsymbol{y}_n^s\}_s)\}$ as the training dataset where $s \in \{0, ..., S\}$ indexes task referring to different scales, $\boldsymbol{x}_n$ a greyscale volume and $\{\boldsymbol{y}_n^s\}_{s=0}^S$ the multiple sets of ground truth vasculature segmentation masks. In our case, we defined s as a specific vascular scale that belongs to the binary mask $\boldsymbol{y} = \cup_{s=1}^S \boldsymbol{y}^s$ where $\boldsymbol{y}^0$ reflects the initial scale (i.e. $\boldsymbol{y}^0 = \boldsymbol{y}$). In this setting, supervised multi-task segmentation with a single input, multiple outputs and S learning tasks $\{\mathcal{T}_s\}$ consisted of approximating a mapping function $\phi : \boldsymbol{x} \rightarrow \phi(\boldsymbol{x}) = \{\hat{\boldsymbol{y}}^s\}_{s=0}^S$ from $\mathcal{D}$, through a deep multi-task convolutional network ϕ (Fig.2.16) comprising a contracting path encoder E projecting $\boldsymbol{x}$ to a latent representation $\boldsymbol{z}$ and $S+1$ decoders D_s, such that:

$$\phi(\boldsymbol{x}) = \{D_s(\boldsymbol{z}) = \hat{\boldsymbol{y}}^s \mid \boldsymbol{z} = E(\boldsymbol{x})\}_{s=0}^S \tag{2.13}$$

where $D_0(\boldsymbol{z}) = D_{\texttt{mt}}(\boldsymbol{z})$ and $\{D_s(\boldsymbol{z})\}_{s=1}^S$ refer respectively to the main (`mt` stands for main task) and the auxiliary tasks decoders. Furthermore, the model parameters were estimated by learning both main and auxiliary tasks jointly through the optimization of the following objective function:

$$\mathcal{L} = \mathcal{L}_{\texttt{mt}}(\boldsymbol{y}, \hat{\boldsymbol{y}}) + \sum_{s=1}^{S} \lambda_s \times \mathcal{L}_s(\boldsymbol{y}^s, \hat{\boldsymbol{y}}^s) + \lambda_c \times \mathcal{L}_c \tag{2.14}$$

where $\mathcal{L}_{\texttt{mt}}$ and $\mathcal{L}_s$ combine Dice and weighted cross-entropy. λ_s and λ_c are hyper-parameters that regulates the strength of each loss. $\mathcal{L}_s$ is a contrastive loss which encourages features representation $\boldsymbol{z}_n^s$, belonging to an anchor set $\mathcal{A}$ from the same scale to be aligned and separate features from different scales apart in order to ensure uniformity. Since it is difficult to make the distinction between the latent representation of each auxiliary task at $\boldsymbol{z}$, we define the compact representation derived from the first

layer of the decoder D_s, for each task $\mathcal{T}_s$. The representations $\boldsymbol{z}_i^s$ from the same scale (resp. different scale) belong to the set $\mathcal{P}_i$ (resp. $\mathcal{N}_i$). From this, the contrastive loss was defined as:

$$\mathcal{L}_c = \frac{1}{|\mathcal{A}|} \sum_{i \in \mathcal{A}} \frac{1}{|\mathcal{P}_i|} \sum_{j \in \mathcal{P}_i} \log \frac{\exp(\boldsymbol{z}_i^{s\top} \boldsymbol{z}_j^s / \tau)}{\exp(\boldsymbol{z}_i^{s\top} \boldsymbol{z}_j^s / \tau) + \sum_{k \in \mathcal{N}_i} \exp(\boldsymbol{z}_i^{s\top} \boldsymbol{z}_k^s / \tau)} \tag{2.15}$$

where $\boldsymbol{z}^s$ is a l_2-normalized vector and $\tau > 0$ a scalar temperature hyper-parameter. The effectiveness of our pipeline based on scale-specific auxiliary multi-task contrastive learning was illustrated for liver vessel extraction with $S = 3$, on the 3D-IRCADb [47] dataset. Especially, we compared 3D ResUNet in binary and multi-class settings as well as the proposed approach without and with contrastive learning. The ablation study demonstrates the effectiveness of integrating multi-scale contrastive learning, especially in `clDSC`, illustrating the connectivity improvement (Fig.2.17*b*) reached by our method.

2.6.2 Dual-task segmentation with Transformers

Another facet of my research activities involving semantic segmentation has focused on the analysis of MR images acquired from patients with autosomal-dominant polycystic kidney disease (ADPKD), a systemic genetic disorder affecting 12.5 million people worldwide [55]. This pathology is characterized by the enlargement of kidneys due to the progressive development of renal cysts. Fourth leading cause of kidney failure, it requires dialysis or kidney transplantation for the majority of patients [55]. ADPKD, whose degree of phenotypic variability among affected individuals is extremely broad, can also manifest with extra-renal symptoms (e.g. presence of cysts in the liver). The continuous growth of cysts in ADPKD leads to a progressive increase in total kidney volume (TKV). TKV is the most important imaging biomarker for quantifying the severity of ADPKD and predicting future renal function decline [56]. TKV is used in clinical care to assess the risk of individual disease progression or as primary or secondary end-point to assess treatment effects [57].

Polycystic kidney delineation from MR data has predominantly relied on CNN architectures so far [59]. However, recent developments in medical image analysis have demonstrated the potential of Transformer-based models, which have shown superior performance in various computer vision applications. Vision Transformer (ViT) models have especially gained significant attention in medical image segmentation [60]. Unlike CNN models, Transformers do not require any convolution or pooling operations but instead rely on self-attention mechanisms to model the relationships between different image areas. This approach has shown to be effective for capturing global context information in medical images, which can be critical to reach sufficiently efficient delineations. In addition to model architecture, multi-task learning has also become increasingly important in image analysis, with a great potential to improve dense prediction tasks such as semantic segmentation [53]. However, in the context of polycystic kidney segmentation, the potential benefits of multi-task strategies had not been fully explored, representing a significant gap in current studies.

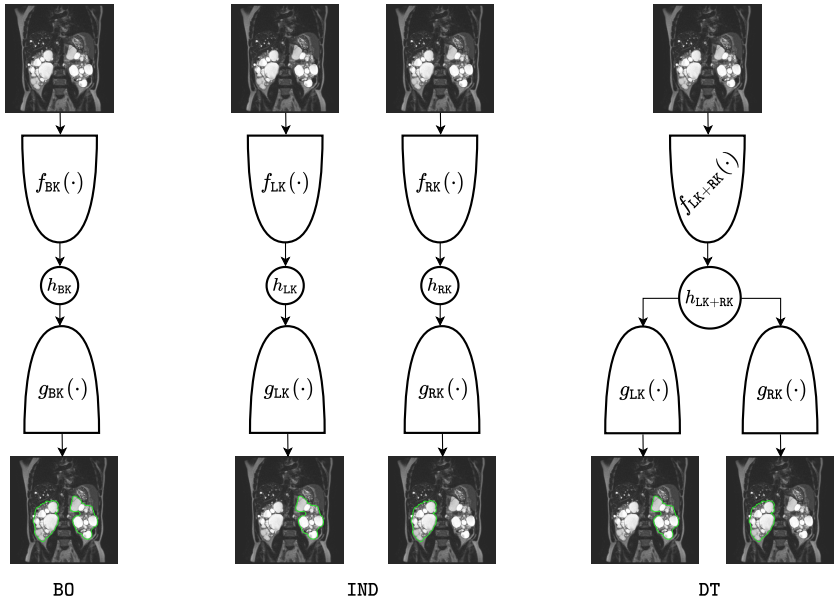


Figure 2.18 – `BO`, `IND` and `DT` learning schemes compared in [**58**] for kidney segmentation in patients with ADPKD. For sake of clarity, skip-connections are not displayed.

With the aim of designing the best possible polycystic kidney delineation system from MR scans, the stakes of the collaboration with the nephrology department of University Hospital of Brest were two-fold. First, evaluate and compare various purely CNN-based, Transformers-based, and hybrid CNN/Transformers-based networks in the context of polycystic kidney segmentation. Second, extend these backbones with an effective dual-task learning scheme involving a common feature extractor followed by per-kidney decoders. Indeed, each left or right kidney is present in a different spatial context, close to different anatomical structures (e.g. right kidney in interaction with the liver, left kidney in the vicinity of the spleen). In addition, the spatial distribution and heterogeneity of cysts can greatly vary from one kidney to another. In this context, three learning schemes were considered (Fig.2.18) towards polycystic kidney segmentation with deep learning. First, the « both organs » (`BO`) configuration exploited a single deep network segmenting both left and right kidneys, without any distinction between them. The network made of one single encoder $f_{\texttt{BK}}(\cdot)$ followed by one single decoder $g_{\texttt{BK}}(\cdot)$ performed a binary segmentation task distinguishing between renal and non-renal tissues, whatever the laterality. Second, the « independent » (`IND`) strategy involved two separate encoder-decoder networks: $f_{\texttt{LK}}(\cdot)$ followed by $g_{\texttt{LK}}(\cdot)$ for left kidney and $f_{\texttt{RK}}(\cdot)$ followed by $g_{\texttt{RK}}(\cdot)$ for right kidney. Each of them performed a binary segmentation task without any weight sharing. Last, the « dual-task » (`DT`) scheme made use of a single network comprising one single encoder $f_{\texttt{LK+RK}}(\cdot)$ and two task-specific decoders $g_{\texttt{LK}}(\cdot)$ and $g_{\texttt{RK}}(\cdot)$, one for each kidney. Features arising from the encoder were common to both task. In `IND` and `DT` configurations, results were fused through a simple union operator.

In practice, we explored the use of different deep models for polycystic kidney segmentation: CNN-based (v19pUNet [**34**]), hybrid CNN/Transformer-based (TransUNet [62], MedT [63], SwinUNetV2

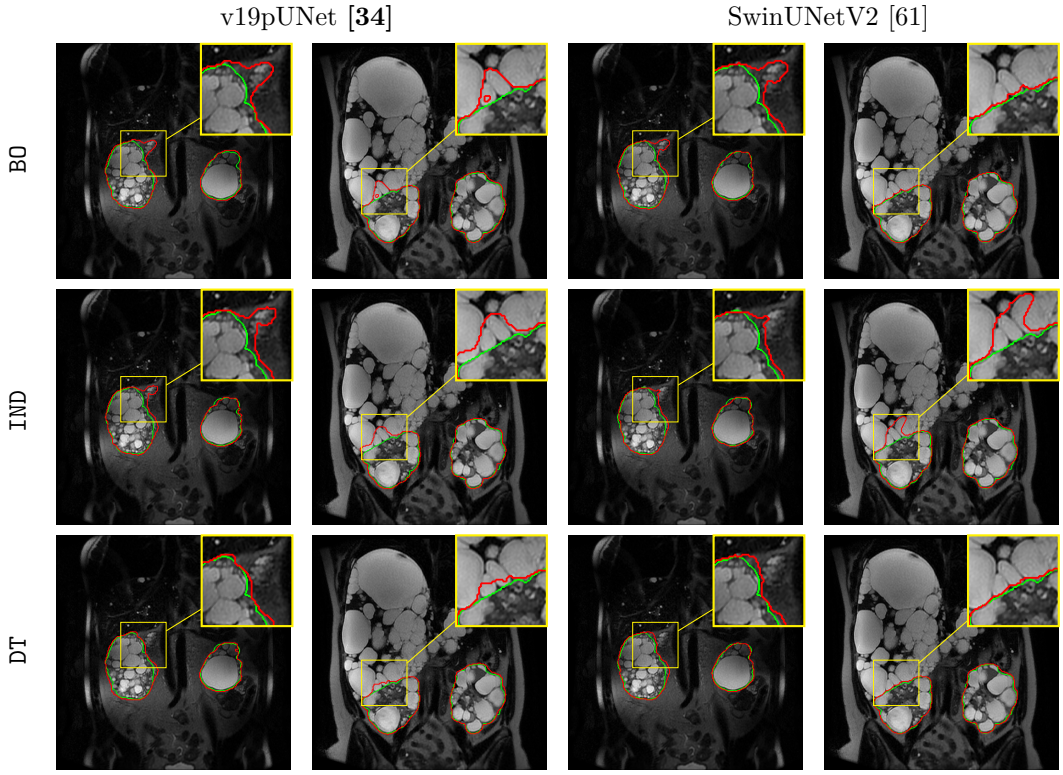


Figure 2.19 – Polycystic kidney segmentation using v19pUNet **[34]** and SwinUNetV2 [61] trained with `BO`, `IND` and `DT` learning strategies **[58]**. Ground truth in green, predictions in red.

[61]) and Transformer-based (Segmenter [64]). Each architecture was employed in a single- (`BO`, `IND`) and dual-task (`DT`) learning fashion. Three main findings arised from experiments (Fig.2.19). First, task-specific decoders were better able to process features from a single joint encoder $f_{\texttt{LK+RK}}(\cdot)$ (`DT`) than from kidney-specific encoders (`IND`), i.e. $f_{\texttt{LK}}(\cdot)$ and $f_{\texttt{RK}}(\cdot)$. Second, exploiting decoding branches, i.e. $g_{\texttt{LK}}(\cdot)$ and $g_{\texttt{RK}}(\cdot)$, respectively targeting left and right kidneys (`DT`), provided better results than employing a single joint decoder $g_{\texttt{BK}}(\cdot)$ (`BO`). Third, results underscored the substantial advancements made by hierarchical Transformers (SwinUNetV2) whose ability to capture multi-scale information and leverage hierarchical representations significantly contributed to their superior performance. The hybrid TransUNet and the pure convolutional network (v19pUNet) also demonstrated competitive performance, indicating the continued relevance and effectiveness of CNN-based approaches.

2.7 Semi-supervised learning with label propagation

While fully-supervised segmentation methods based on deep learning have been investigated with sparse annotated 2D slices **[34]**, a relatively significant number of training examples is required. Conversely, learning from few examples (i.e. few-shot learning) remains an open issue. Especially,

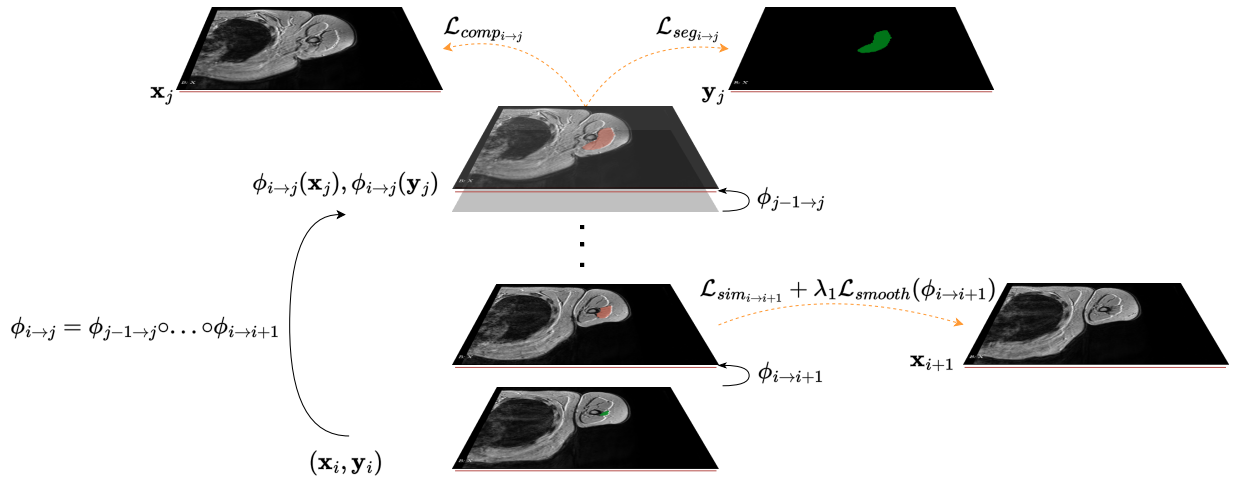


Figure 2.20 – Deep registration-based label propagation from an annotated slice to the next one, toward 3D muscle segmentation. Refer to **[65]** for the description of notations.

interactive approaches have been investigated in the field of deep learning for image segmentation purposes (e.g. point or scribble-based techniques) but may suffer from a lack of reproducibility and require complex optimization procedures. As an alternative, intra-subject semi-automatic segmentation, which involves spreading annotated slices throughout a volume, enables to deal with the limited amount of annotations. This technique is more commonly seen in the field of video segmentation, when the first annotated frame is propagated throughout the sequence [66]. In medical image analysis, morphological-based interpolation of distant annotated slices provides a fast strategy to propagate labels over a 3D volume, but may require some interactions and fail on multi-structure objects [67]. **[58]**elies solely on segmentation contours, such strategy does not capture local structure variations.

In this context and to complement the fully-supervised segmentation approaches described so far, the PhD thesis of N. Decaux enabled the development of a segmentation method relying on registration-based label propagation to provide 3D muscle MR delineations from a limited number of annotated 2D slices. The objective was to develop an intra-subject 3D segmentation method for pediatric muscles based on very few manually annotated slices. To this end, we investigated in **[65]** the use of an unsupervised deep learning-based registration framework to propagate 2D labels through the full 3D volume (Fig.2.20). The registration approach relied on intensity similarity between successive slices and on muscle shapes from annotated slices. A regularization term was introduced through the definition of a dedicated loss from combined deformation fields. Propagated masks from different manual segmentations were merged through a weighting technique based on image similarity measures.

Among various experiments, a challenging intra-subject segmentation scenario was considered with only three annotated slices for each MR volume to segment (Fig.2.21). The middle slice allowed a fine delineation of the muscle while the two other slices provided its spatial extent. It appeared that the proposed label propagation was less sensitive to distance from the nearest annotated slice, compared

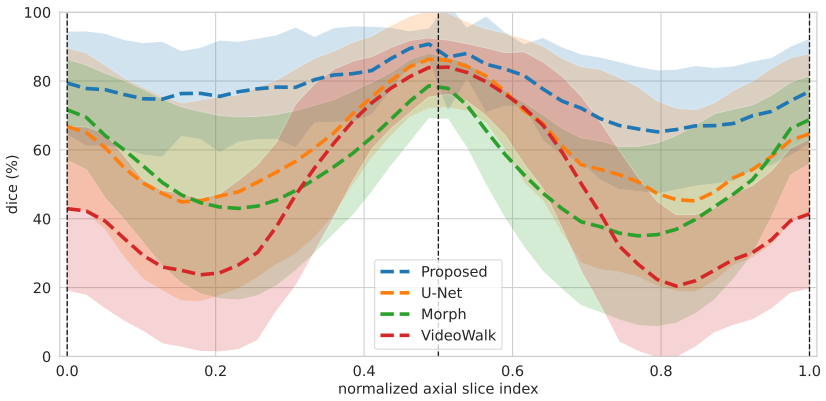


Figure 2.21 – Averaged Dice over the NIH pediatric shoulder dataset [68], in a minimal supervision setting **[65]**. Dice is displayed with respect to the normalized axial slice number. Vertical black lines represents the location of annotated slices used for training. Colored areas deal with standard deviation.

to interpolation-based approaches [69], UNet [26] and VideoWalk [70]. Results also demonstrated a higher overall robustness with the proposed method, as evoked in Fig.2.21 by standard deviations.

2.8 Conclusion

Deep learning has proven to be a powerful tool for medical image segmentation. Compared to standard techniques including atlas-based, active contours, or machine learning (Sect.2.2), its ability to automatically learn complex and hierarchical representations from data enables to achieve a high level of robustness in segmentation tasks tackling various diseases, populations, anatomies and imaging modalities. The availability of large datasets and open-source frameworks has facilitated the development and deployment of deep learning-based segmentation algorithms, making them more accessible to both researchers and clinicians. A tangible impact on diagnosis, surgery planning, therapeutic follow-up (Chap.4), prognostic, dosimetric or radiomics applications at large is starting to witness.

Since the introduction of UNet (Sect.2.3) and U-shaped convolutional encoder-decoder derivatives trained with data augmentation, various developments and methodological breakthrough have emerged in the medical image analysis community **[71]**. Among current trends, the relevancy of conditional generative adversarial networks, cascaded networks (Sect.2.4.1) and deep supervision (Sect.2.4.2) have been proven to enable the improvement of segmentation accuracy for large (e.g. liver) and smaller structures (e.g. kidneys). Regularization techniques embedding prior knowledge (Sect.2.5) such as shape, topological or adjacency constraints tend to be democratized towards greater generalizability of deep segmentation models. In particular, employing a convolutional auto-encoder to exploit geometrical and/or topological constraints aims at encouraging models to follow the global anatomical properties of the underlying anatomy via learnt non-linear representations. Additionally to novel ar-

chitecture designs and the use of contraints during training, a strong focus has been recently devoted to contrastive (Sect.2.5.4), multi-task (Sect.2.6) and semi-supervised (Sect.2.7) learning paradigms, especially to alleviate issues related to the lack of annotated imaging data.

Despite recent successes, there are still challenges to the use of deep learning for medical image segmentation. These challenges include the need for large amounts of labeled data for training, the sensitivity of deep models to noise, non-uniform contrast and artifacts in medical images, the needed incorporation of local and global context to benefit from both short- and long-range spatial dependencies, the management of small structures and weak boundaries as well as the robustness to inter-subject variability and various multi-center, multi-scanner intensity domains. Given these challenges, the use of deep learning has shown great promise in line with the emergence of vision Transformers whose ability to model long-range dependencies from 3D medical images appears similar or even better than standard convolutional only architectures (Sect.2.6.2). Either hybrid when used in conjunction with convolutional layers or purely Transformers-based, these approaches are still at an early stage. More works in this direction are expected, especially in the context of information fusion (Chap.3) comprising multi-view (Sect.3.3), multi-domain (Sect.3.4) and multi-modal (Sect.3.5) analysis.

Chapter 3

From knowledge transfer to information fusion

3.1 Introduction

The core of medical image analysis revolves around information, which is inherently complex, multi-modal and heterogeneous. The challenge lies not only in harnessing the wealth of information arising from medical imaging data but also in effectively transferring knowledge across different domains. To bridge the gap between learning paradigms and clinical needs, recent investigations have struggle with novel transfer learning scenarios and information fusion strategies towards more efficient medical image analysis models. The increasing popularity of knowledge transfer and information fusion is driven by the need to capitalize on both complementary and redundancy across various conditions, viewpoints, modalities, or anatomies when managing paired or unpaired multi-domain datasets.

As prime example, multi-modal segmentation with deep learning offers a powerful playground for improving robustness and generalization across a wide range of applications. Robustness is improved as the models can rely on complementary information, especially when the modalities provide information of a different nature. It also enables models to better generalize to diverse datasets since robust representations can be learned across different data distributions. Going further, the synergy of information from different modalities can lead to a more nuanced representation, enabling models to capture intricate patterns and relationships within the data.

A special attention has been also paid in recent years to multi-domain segmentation strategies which are far more relevant than focusing on multiple intensity domains separately. In this direction, multi-task and multi-domain techniques with multiple anatomies as targets tend to overcome the inherent scarcity of imaging data while leveraging shared features between imaging datasets.

This chapter presents some of my contributions in this context, by covering various architectures, clinical contexts and learning paradigms including active learning, multi-task and contrastive learning.

3.2 Learning transferability

Learning transferability refers to the ability of a given model to generalize its knowledge across different domains. In medical imaging, where diverse imaging technologies and anatomical variations

exist, achieving robust and adaptable segmentation models is crucial. Transfer learning techniques aim to leverage knowledge acquired from one dataset to enhance the performance on a target dataset with potentially distinct characteristics. This concept is particularly significant in medical image segmentation tasks as it addresses challenges related to data scarcity, domain shifts, and the need for models able of handling various clinical settings. In this context, my research activities mainly addressed healthy versus pathological transferability (Sect.3.2.1), knowledge transfer from off-the-shelf to specific lesions (Sect.3.2.2) as well as cross-dimensional transfer learning (Sect.3.2.3).

3.2.1 Healthy versus pathological transferability

The automatic segmentation of pathological shoulder muscles from MR scans collected from children with musculoskeletal disorders is a context in which data scarcity raise huge challenges (Sect.2.7). The purpose of a study made in collaboration with ILDYS[5], University Hospital of Brest and NIH[6] was to develop an automatic muscle segmentation pipeline, able to support new insights into the evaluation and management of musculo-skeletal diseases [**34**]. In particular, our work addressed the learning transferability from healthy to pathological data, focusing on how available data from both healthy and pathological muscles can be jointly exploited for pathological shoulder muscle delineation.

Studying the learning transferability from healthy to pathological structures is key in musculoskeletal pathologies for two reasons. First, despite different shapes and sizes due to growth and atrophy, healthy and pathological muscles may share common characteristics (e.g. anatomic locations, overall aspects). Second, combining healthy and pathological data for deep learning-based segmentation can act as a smart data augmentation strategy when faced with limited annoted data. In exploring the combined use of healthy and pathological data for pathological muscle segmentation, determining the optimal learning scheme was crucial. In practice, three different learning schemes were considered (Fig.3.1). First, the « pathological only » (`P`) strategy consisted in exploiting ground truth annotations made on impaired shoulder muscles only, making the hypothesis that features extracted from healthy examinations are not suited enough for pathological anatomies. Second, the « transfer from healthy to pathological » (`HP`) strategy dealt with transfer learning and fine tuning from healthy to pathological muscles. In this context, a first UNet was trained using ground truth segmentations from unaffected shoulders only. The weights of the resulting model were then used as initialization for a second UNet network which was trained using pathological inputs only. Third, the « simultaneous healthy and pathological » (`A`) configuration consisted in training a UNet with annotations made on both healthy and pathological shoulder muscles to benefit from a more consequent dataset.

Built around a UNet model comprising a VGG-16 encoder pre-trained on ImageNet, the `P`, `HP` and `A` learning schemes were evaluated on four shoulder muscles: deltoid, infraspinatus, supraspinatus, subscapularis. Results showed that features extracted from unimpaired limbs were suited enough for

5. `https://www.ildys.org/`
6. `https://www.nih.gov/`

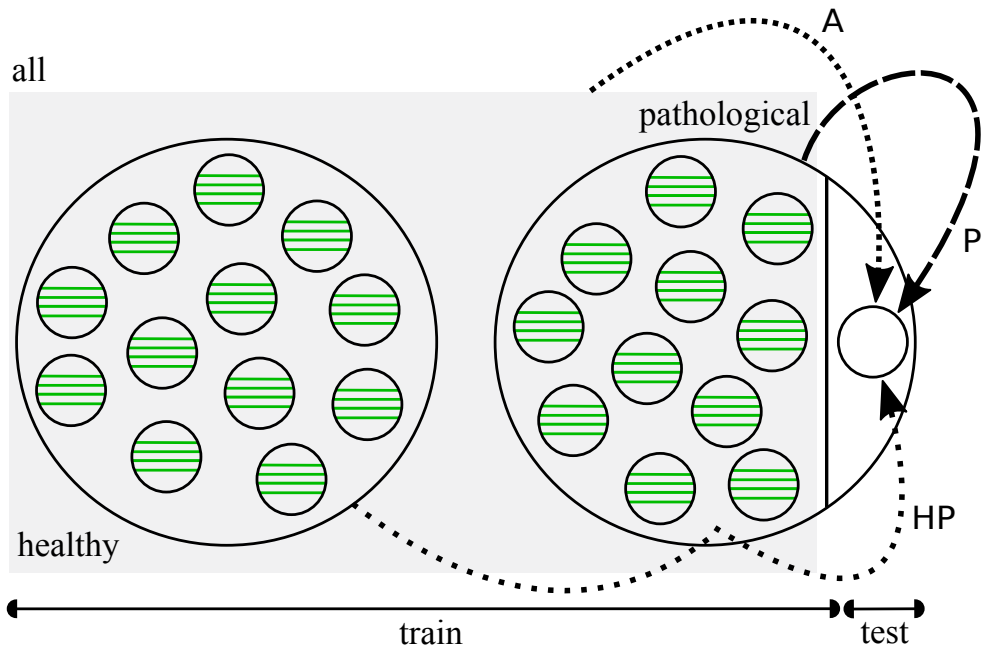


Figure 3.1 – Different learning transferability schemes (`P`, `HP`, `A`) evaluated in [**34**] for deep learning-based pathological shoulder muscle segmentation.

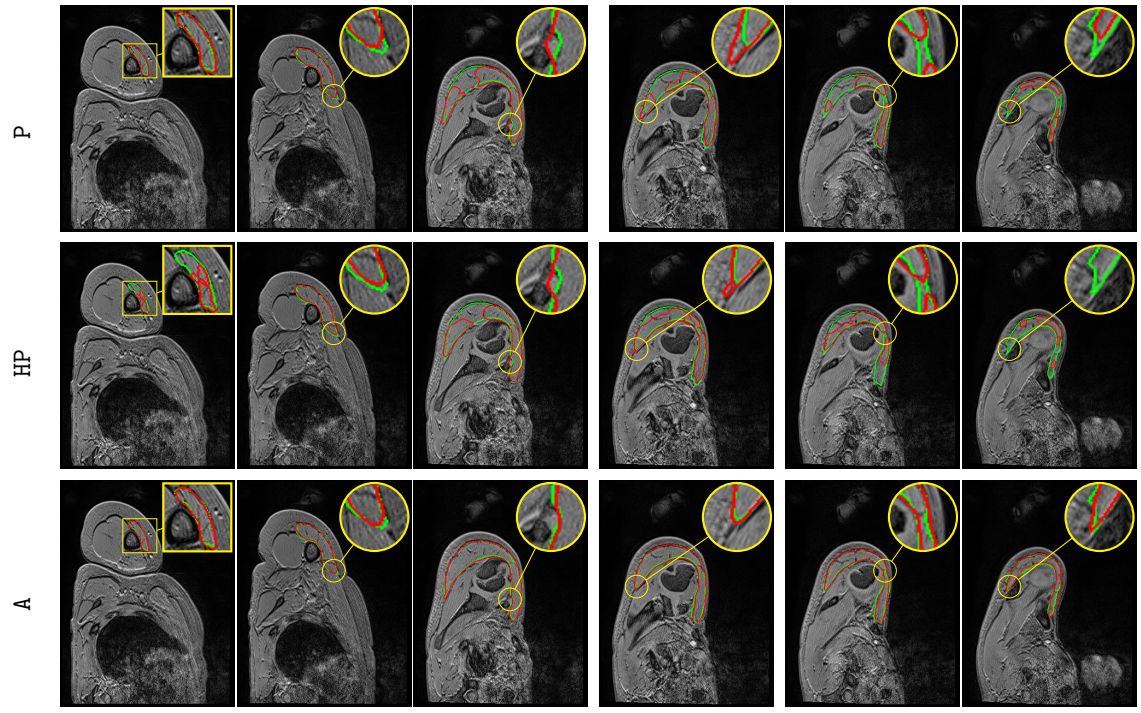


Figure 3.2 – Pathological deltoid segmentation using UNet [26] comprising a VGG-16 encoder pre-trained on ImageNet and embedded with learning schemes `P`, `HP` and `A` [**34**]. Groundtruth and estimated muscle delineations are respectively in green and red.

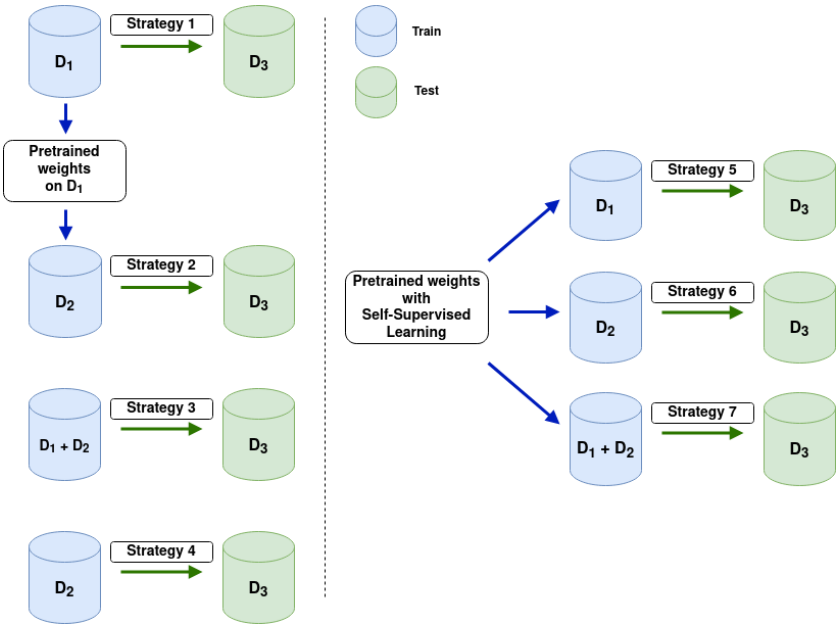
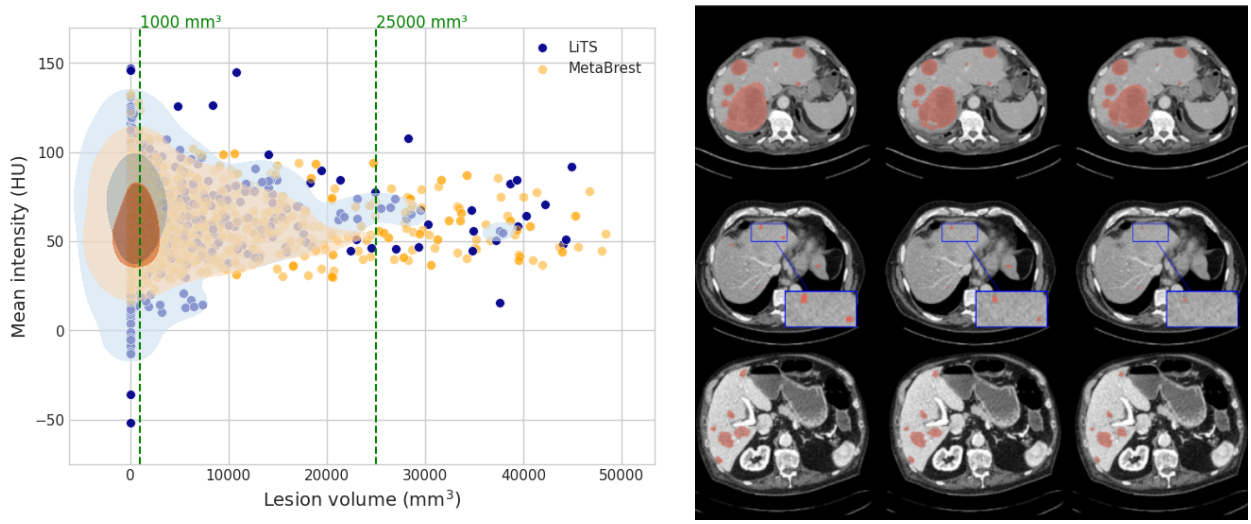


Figure 3.3 – Various training strategies employed in **[72]** for liver metastasis segmentation purposes.

pathological anatomies while acting as an efficient data augmentation scheme. Compared to transfer learning, combining healthy and pathological data for training provided the best segmentation accuracy (Fig.3.2), with outstanding delineation performance for muscle boundaries including insertion areas.

3.2.2 Knowledge transfer from off-the-shelf to specific lesions

Despite the progress of deep learning algorithms using Transformers (Sect.2.6.2), automatically segmenting small hepatic metastases remains a persistent challenge, with key implications in surgery planning and treatment follow-up (Chap.4). This can be attributed to the degradation of small structures due to the process of feature down-sampling inherent to many architectures as well as class imbalance. While similar challenges have been observed for liver tumors originated from HCC [73], their manifestation in the context of liver metastasis delineation remained under-investigated. Hence, a study **[72]** conducted in the framework of the PhD thesis of M. Abbas revolved around integrating a heterogeneous dataset (`LiTS` [74], referred to as `D1`) encompassing a broad spectrum of tumor types with another dataset (`MetaBrest`, collected from University Hospital of Brest) solely concentrating on hepatic metastases from colorectal cancer. `MetaBrest` was divided into two subsets (`D2` for training, `D3` for test) while maintaining an even distribution of lesion sizes. Our objective was to pinpoint the most effective training strategies that can effectively combine these datasets, confronting the complexities of dataset diversity and aiming to leverage the innate potential of transfer learning for the target task of segmenting hepatic metastases from CT scans. Various approaches were designed to explore different facets of model training, including the impact of domain-specific training and pre-training, the use of extensive pre-training on a mega-dataset and the efficacy of training from scratch.



(*a*) (*b*)

Figure 3.4 – (*a*) Scatter plot of tumor volume (mm^3) versus mean intensity (HU) for `LiTS` [74] and `MetaBrest` datasets. Each dot corresponds to a single lesion. (*b*) CT scans with overlapping metastasis delineations. From left to right, ground truth is compared against `strategy 1` and `strategy 4` predictions, showcasing the superior performance of `strategy 1` in segmenting small metastases [**72**].

Using SwinUNETR [75] as backbone and `D3` as test set, seven training scenarios were designed (Fig.3.3). As `strategy 1`, SwinUNETR was exclusively trained on `D1`. In `strategy 2`, we used the pre-trained weights acquired from `strategy 1` and subsequently trained them on `D2` in a fine-tuning fashion with the underlying goal to leverage the knowledge gained from `D1` to `D2`. `Strategy 3` consisted in training SwinUNETR using `D1` and `D2` datasets to explore the benefits of joint training and its impact on generalization to a known dataset. `Strategy 4` was a straightforward training approach using only `D2` to investigate the direct knowledge transfer between data arising from `MetaBrest` only. For the last three training scenarios, we aimed at assessing the impact of weight initialization through self-supervised learning. Hence, we employed SwinUNETR weights arising from a training performed on a mega-dataset of roughly 50,000 images encompassing MR, CT and histology images through self-supervised learning [76]. The main objective was to investigate the advantages of weight initialization in an ImageNet-like manner. Thus, we initialized the SwinUNETR architecture with the pre-trained weights obtained from self-supervised learning and fine-tuned it using `D1` for `strategy 5`. In `strategy 6`, the self-supervised pre-trained weights were used to train the model on `D2`. Finally, `strategy 7` involved combined training on `D1` and `D2` datasets using the self-supervised pre-trained weights.

The conducted comparative analysis revealed a complex picture of the efficacy of different training strategies towards liver metastasis CT segmentation. Contrary to the broader trends in deep learning where pre-training often improves performance, our findings indicated that models trained from scratch or with domain-specific pre-training reach greater proficiency. While `strategy 3` reached the best average Dice score, `strategy 1` and 2 stood out as the top performers for both small ($<1000mm^3$) and

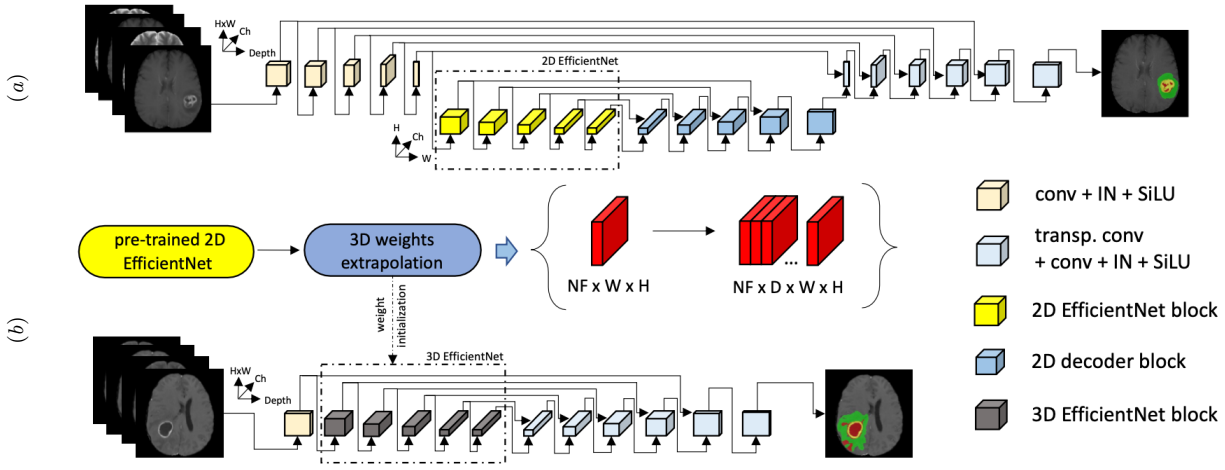


Figure 3.5 – Cross-dimensional transfer learning strategies proposed in [**77**]: (*a*) weight transfer to embed a 2D pre-trained encoder into a higher dimensional UNet, (*b*) dimensional transfer to expand a 2D segmentation UNet into a higher dimension one.

medium (from 1000 to 25000mm^3) liver metastases (Fig.3.4). Strategies employing self-supervised pre-training did not perform as anticipated. This less satisfactory performance underscored the limitations of transfer learning in scenarios where domain-specific features are crucial, emphasizing the significance of dataset and task specificity in developing deep segmentation models.

3.2.3 Cross-dimensional transfer learning

The performance of 2D image classification networks, being trained on databases made of millions of natural images, is constantly improving. Conversely, in the field of medical image analysis, the progress is also remarkable but has mainly slowed down due to the relative lack of annotated data and besides, the inherent constraints related to the acquisition process. These limitations are even more pronounced given the volumetry of medical imaging data. In a research work conducted in collaboration with University of Bejaia[7], Algeria and University Hospital of Brest, we developed an efficient way to transfer the efficiency of a 2D classification network trained on natural images for 2D, 3D uni- and multi-modal medical image segmentation purposes. In this direction, we designed in [**77**] novel architectures (Fig.3.5) based on two key principles: weight transfer by embedding a 2D pre-trained encoder into a higher dimensional UNet and dimensional transfer by expanding a 2D segmentation network into a higher dimension one. The proposed networks were tested on benchmarks comprising different modalities including CT, MR and ultrasound. Through the strategic utilization of pre-trained 2D image classification networks and the facilitation of inter-slice relationship extraction within the depth dimension, both transfer learning paradigms outperformed state-of-the-art methods on CHAOS [**29**], BraTS [78] and CAMUS [79] publicly-available datasets.

7. `http://univ-bejaia.dz/en/`

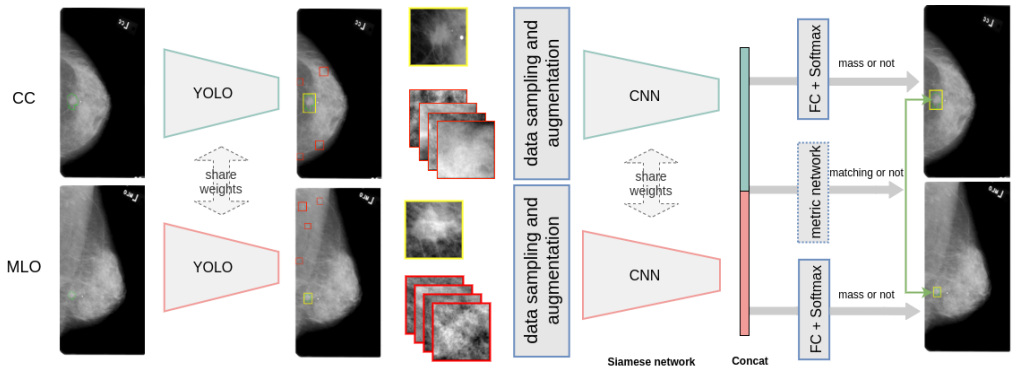


Figure 3.6 – Multi-task, multi-view mammography analysis for improved breast mass detection **[80]**. Green contours indicate ground truth delineations, red (yellow) boxes indicate false (true) detections.

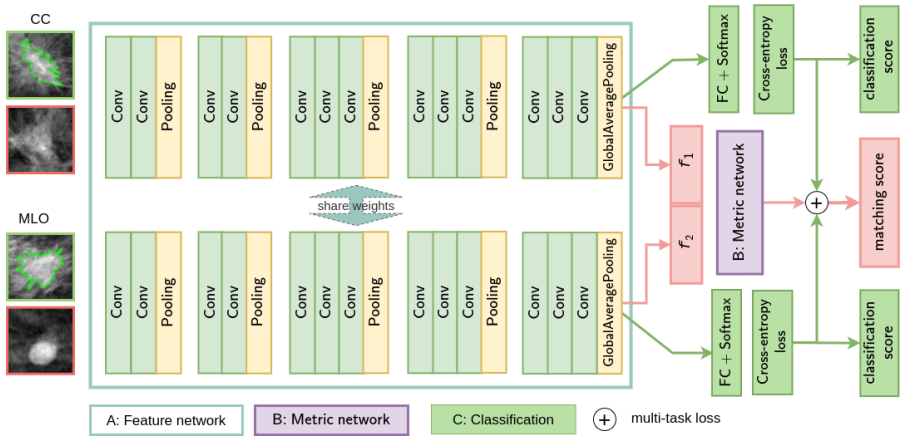


Figure 3.7 – Combined matching and classification network proposed in **[80]**. Green and red patches respectively correspond to positive and negative samples.

3.3 Multi-view analysis

Breast cancer screening routinely benefits from visual analysis of mammograms acquired from different viewpoints to improve decision-making. Similarly, the performance of computer-aided diagnosis systems can be improved by the fusion of multi-view information. Based on this finding, contributions were carried out within the SePEMeD joint laboratory[8], as part of the PhD thesis of Y. Yan.

3.3.1 Better lesion detection with dual-view matching

To address the limitations of single-view processing, we took advantage of multi-view information by developing in **[80]** a deep architecture exploiting both craniocaudal (CC) and mediolateral-oblique (MLO) views from the same mammographic examination, towards fully-automated detection of breast masses **[80]**. Rather than being limited to a single-view mass identification task [81], we exploited the

8. IMT Atlantique, LaTIM and Medecom (`https://medecom.fr/en/`)

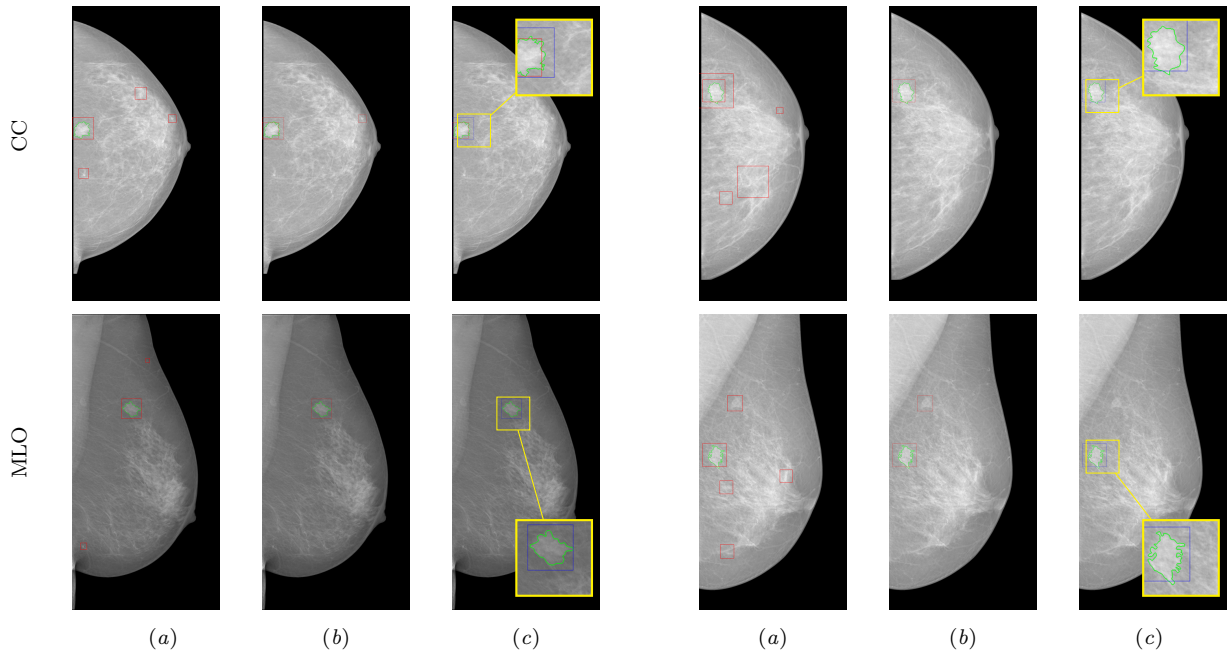


Figure 3.8 – Breast mass detection with YOLO (a), YOLO followed by (b) a classification-only model or (c) our matching and classification network **[80]**. Red boxes are detected mass candidates. Green represents ground truth annotations. Blue boxes show the pairs matched through dual-view matching.

multi-tasking properties of deep networks to jointly learn multi-view matching and classification of suspicious areas (Fig.3.6) from candidate mass patches extracted with a You Only Look Once (YOLO) detection model [82]. Patch level classification and matching were performed through a novel multi-task Siamese network (Fig.3.7). Positive and negative patch samples of CC/MLO views were fed into a two-branch feature extractor to compute robust patch representations. Going further, we not only jointly learned representations from the two views but also simultaneously learned both patch matching and classification to exploit the potential relationships between viewpoints. Compared to classification-only schemes, the underlying goal was to make the matching task improving the robustness of the classification task. Thus, the overall loss $\mathcal{L}$ was designed as the sum of three losses:

$$\mathcal{L} = \alpha \times \mathcal{L}_{cls,CC} + \beta \times \mathcal{L}_{cls,MLO} + \gamma \times \mathcal{L}_{mat} \tag{3.1}$$

where $\mathcal{L}_{cls,CC}$ and $\mathcal{L}_{cls,MLO}$ are the classification losses for CC and MLO views. $\mathcal{L}_{mat}$ is the matching loss which can be cross-entropy or contrastive loss. α, β and γ are coefficients balancing the loss terms.

We carried out experiments to highlight the contribution of dual-view matching for both patch-level classification and examination-level detection scenarios. Results demonstrated that mass matching highly improves the detection performance by outperforming conventional single-task models (Fig.3.8). Our system further guides clinicians by providing accurate dual-view mass correspondences, acting as a relevant second opinion to guide mammogram interpretation and breast cancer diagnosis.

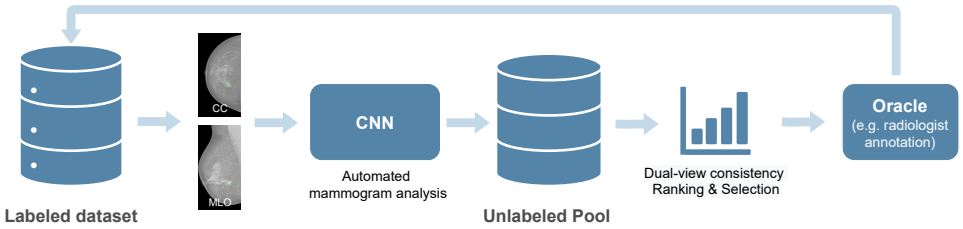


Figure 3.9 – Deep active learning workflow for dual-view mammogram analysis [**83**].

3.3.2 Active learning for dual-view analysis

Supervised learning requires a great deal of manual annotation, which is time-consuming and requires considerable medical expertise. Active learning, on the other hand, aims at reducing annotation efforts by adaptively selecting the most informative samples for labelling. The PhD thesis of Y. Yan also led to the development of a new active learning approach for analyzing multi-view mammograms [**83**], with use cases dedicated to breast mass detection and segmentation. Intuitively, if mass detection or segmentation is carried out robustly, predictions achieved on the different views should be consistent. Hence, exploiting this inter-view consistency appeared as a relevant way to guide the sampling mechanism that iteratively selects the next pairs of images to be labeled by an oracle (Fig.3.9).

In this context, many latent relationships can be exploited as query factors including the number of masses detected on both views, their size, position, shape, texture... In our work, we considered the first two factors as consistency criteria since their correlation is more interpretable. In particular, the number of identified masses from both views $\{N_{CC}, N_{MLO}\}$ should be identical and their sizes $\{S_{CC}, S_{MLO}\}$ (i.e. number of pixels) should be similar. Thus, we defined the two following scores:

$$S_{num} = \frac{\texttt{min}(N_{CC}, N_{MLO})}{\texttt{max}(N_{CC}, N_{MLO})}, S_{size} = \frac{\texttt{min}(S_{CC}, S_{MLO})}{\texttt{max}(S_{CC}, S_{MLO})} \tag{3.2}$$

where S_{num} and S_{size} varies from 0 (low) to 1 (high consistency). Correct predictions should meet the above conditions simultaneously. The final consistency score S was calculated as $S = \texttt{min}(S_{num}, S_{size})$ to provide a rough estimation of the prediction quality. For comparison purposes, we implemented three active learning strategies: random (`rand`), best (`bestC`) and worst consistency (`worstC`) selections. For each cycle, `rand` strategy randomly selects mammogram pairs from the unlabeled set, while `bestC` (`worstC`) selects pairs with the highest (lowest) consistency score S. Results revealed that `worstC` was not superior to `rand`. In addition, `bestC` was consistently better than other strategies, suggesting that picking examples with good prediction results helps to consolidate what has been learned while avoiding corner cases. Experiments (Fig.3.10) also highlighted that achieving performance similar to fully-supervised models is possible by employing only 6.83% (9.56%) of the labeled data for segmentation (detection). The combination of multi-view image analysis and active learning showed promise for contributing to the development of systems to help interpret mammograms.

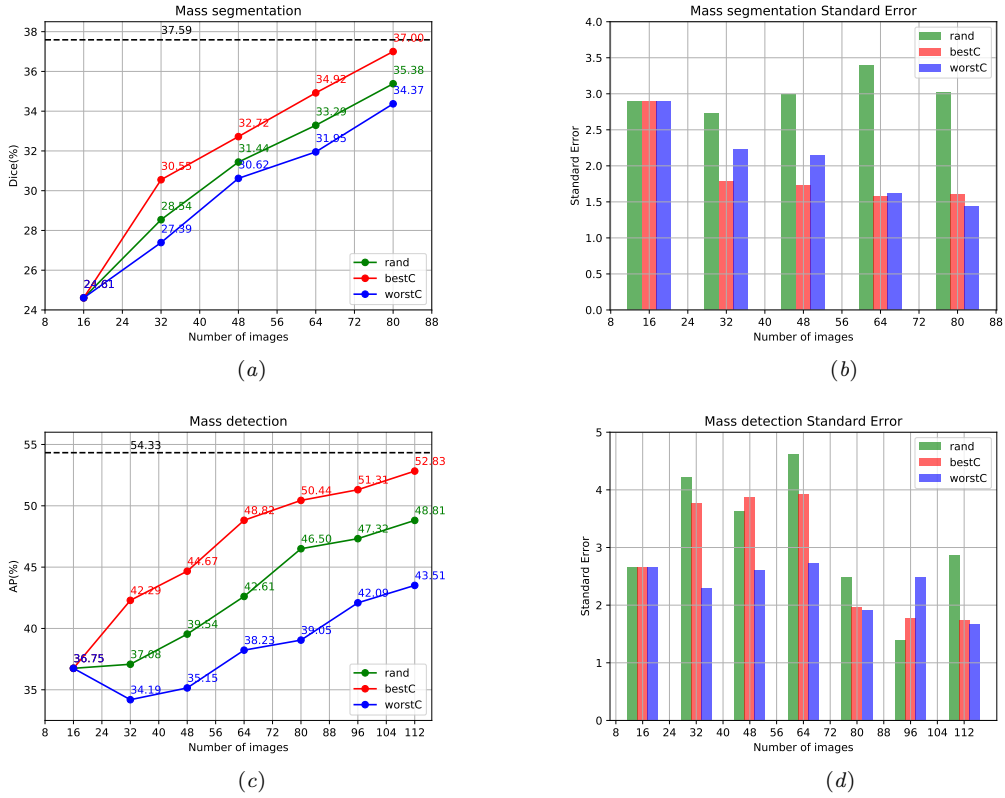


Figure 3.10 – Breast mass segmentation/detection performance with `rand` (–), `bestC` (–) and `worstC` (–) active learning strategies [**83**]. Black dashed lines indicate results using the complete training set. We report average Dice scores for segmentation (a,b) and average precision scores for detection (c,d).

3.4 Multi-task, multi-domain segmentation

In recent years, progress has been made in training models across multiple intensity domains (e.g. multi-modal, multi-scanner, multi-center, multi-protocol). The objective is to capitalize on a larger volume of training data by applying the same image analysis task [85]–[87]. These novel architectures aim at exploiting inter-dependencies between intensity domains, enabling the acquisition of more robust, domain-invariant feature representations. Following the trend to re-use and share an increasing number of parameters, the resulting compact architectures [86], [87] reached superior performance for multi-modal segmentation. However, such methodology was specific to a given anatomical region of interest and the segmentation task involved the same structures across various intensity domains.

Furthermore, multi-task, multi-domain learning frameworks have been concurrently developed for natural image analysis. In the context of semantic scene labeling, a single network was trained in [88] over the union of multiple datasets to address the limited amount of annotated data. In this approach, each dataset was characterized by its own task and domain. In the same direction, studies on universal representations in computer vision proposed to employ a single model with agnostic kernels, as visual primitives may be shared across tasks and domains, and dataset-specific layers which enable

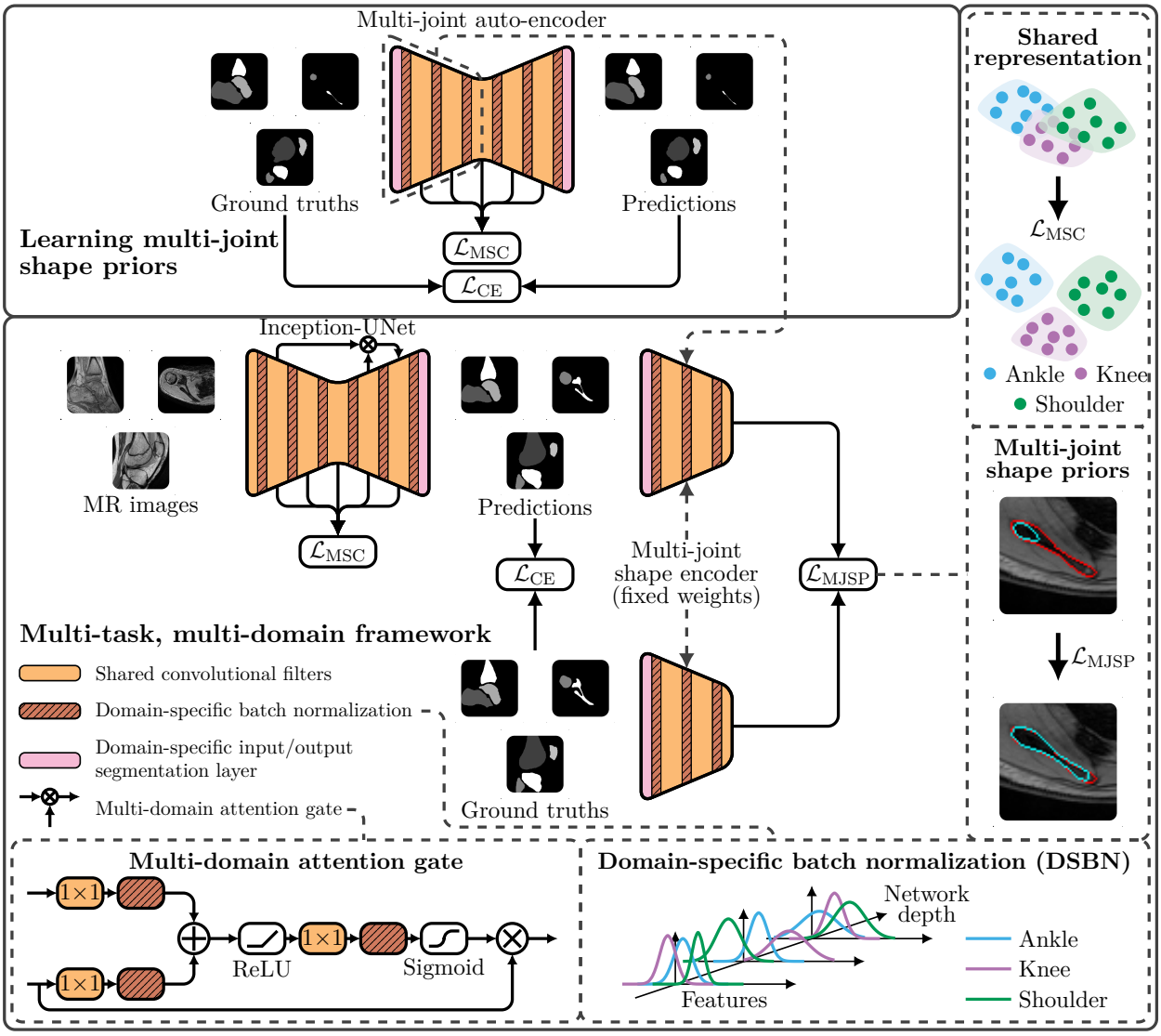


Figure 3.11 – Multi-task, multi-domain segmentation pipeline **[84]** with Inception UNet using shared convolutional filters along with domain-specific batch normalization and multi-domain attention gates.

task and domain specialization [89]. Based on shared representations, these approaches performed at par or better than independent models. However, multi-task, multi-domain learning has rarely been applied to medical image analysis, with the exception of [90] where a single neural network was developed to simultaneously segment multiple anatomies. Nevertheless, instead of generating pixel-wise segmentation masks, the model relied on a computationally expensive tri-planar patches-based approach predicting the class of single pixels. It also failed to account for the difference in intensity distribution between domains, as evidenced by the absence of domain-specific feature normalization.

In the framework of the PhD thesis of A. Boutillon, we proposed to implement and optimize a single segmentation network over the union of multiple pediatric imaging datasets arising from separate

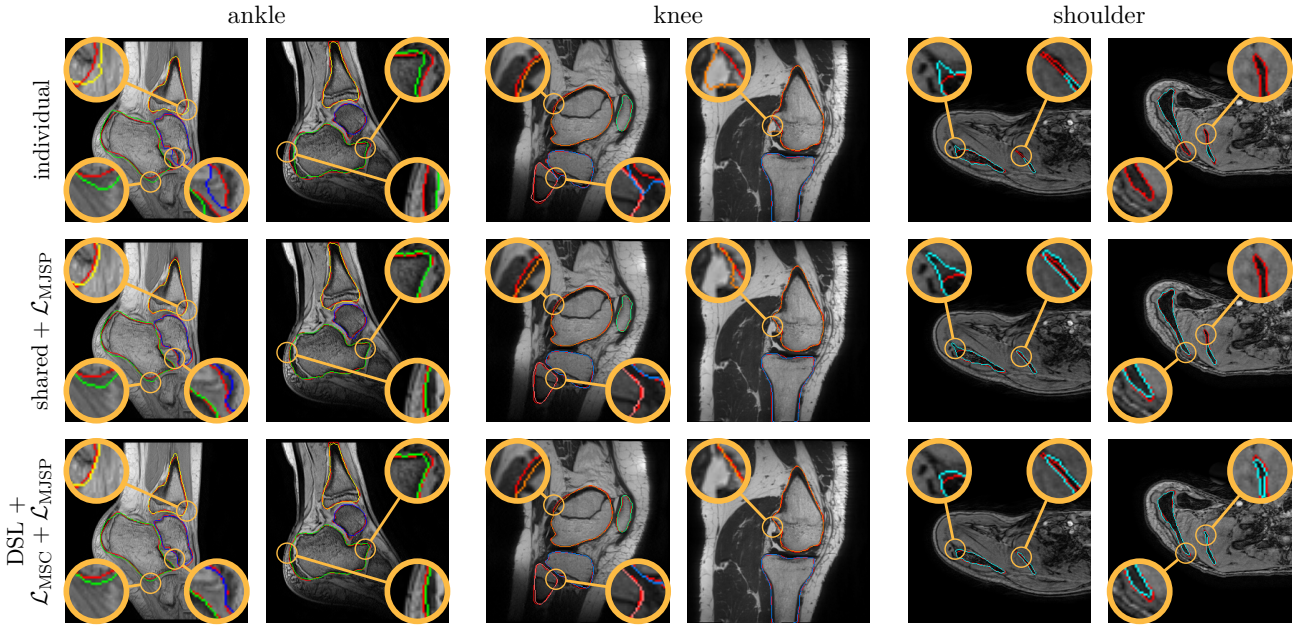


Figure 3.12 – Segmentation of ankle, knee, and shoulder bones employed in individual, shared+$\mathcal{L}_{\text{MJSP}}$, and domain-specific layers (DSL) +$\mathcal{L}_{\text{MSC}}$ + $\mathcal{L}_{\text{MJSP}}$ **[84]** strategies. Ground truth delineations are in red (–) while predicted bones appear in green (–) for calcaneus, blue (–) for talus, yellow (–) for tibia (distal), orange (–) for femur (distal), pink (–) for fibula (proximal), light green (–) for patella, light blue (–) for tibia (proximal), magenta (–) for humerus, and cyan (–) for scapula.

anatomical regions **[84]**. Unlike previous methods that operate on individual musculoskeletal joint, our framework simultaneously learned multiple intensity domains and segmentation tasks emerging from distinct anatomical joints (Fig.3.11). This approach allows to overcome the inherent scarcity of pediatric data while benefiting from more robust shared representations. More precisely, we formalized a segmentation model which incorporates a pre-trained encoder, shared convolutional filters, multi-domain attention gates, domain-specific batch normalization, and domain-specific output layers (Fig.3.11). A multi-scale contrastive regularization was integrated during optimization to improve the generalization capabilities. As opposed to classical contrastive approaches that operate on image classes, we leveraged dataset label information to enhance intra-domain similarity and impose inter-domain margins. Moreover, we extended the multi-task, multi-domain segmentation framework by employing multi-joint shape priors (Sect.2.5) to encode the anatomical characteristics of multiple joints and further constrain the delineation task. Overall, the training procedure integrated the cross-entropy loss function $\mathcal{L}_{\text{CE}}$ defined in a multi-task, multi-domain setting, a multi-scale contrastive regularization $\mathcal{L}_{\text{MSC}}$ to promote inter-domain separation in the shared representations and multi-joint shape priors $\mathcal{L}_{\text{MJSP}}$ previously learnt using a multi-joint auto-encoder on ground truth delineations.

These contributions were assessed for bone segmentation using three datasets of the ankle, knee, and shoulder joints. Results demonstrated that the proposed approach outperforms individual, transfer, and shared segmentation schemes with statistically sufficient margins. While the shared + $\mathcal{L}_{\text{MJSP}}$

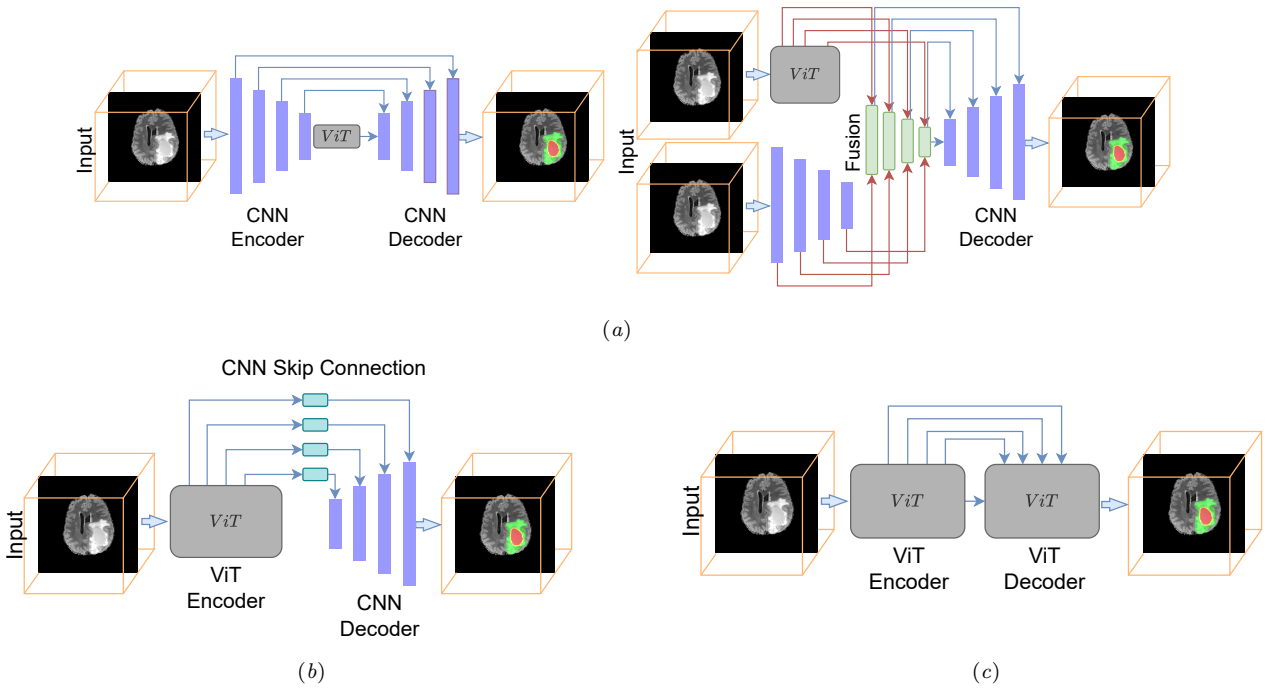
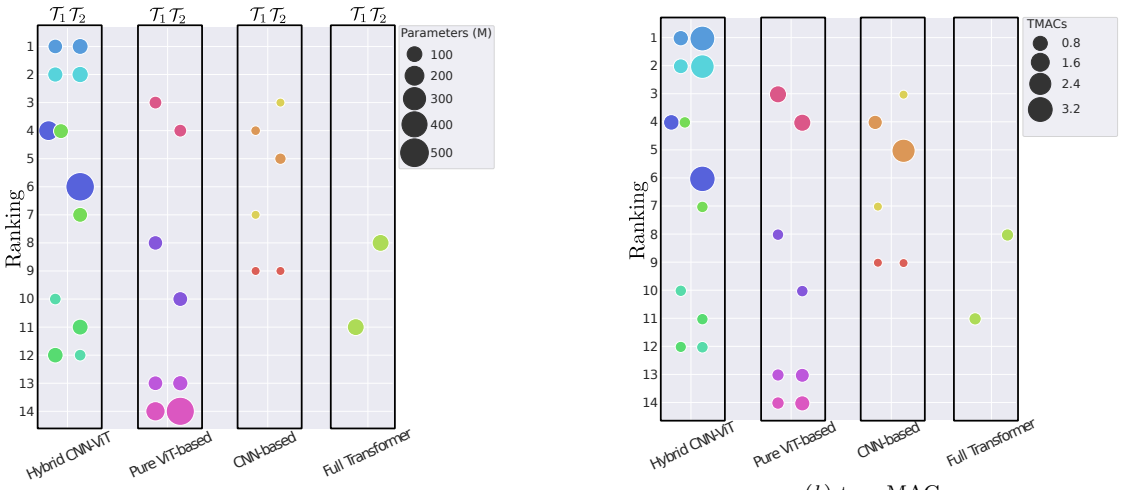


Figure 3.13 – Transformer-based segmentation models used in **[91]** in a multi-modal setting: (a) hybrid CNN-Transformer encoder, (b) pure Transformer-based encoder, (c) full Transformer-based network.

model produced segmentation improvements over its individual counterpart, employing domain-specific layers with $\mathcal{L}_{\text{MSC}}$ and $\mathcal{L}_{\text{MJSP}}$ regularizers was key to learn robust shared representations and achieve precise bone shape predictions (Fig.3.12) on unseen images.

3.5 Multi-modal segmentation with Transformers

Multi-modal learning approaches have the potential to enhance overall segmentation performance by leveraging both complementary and redundant information across modalities [92]. Unfortunately, the majority of multi-modal medical image segmentation methods are usually limited to a straightforward modality concatenation at the input level [62]. It becomes challenging to uncover intricate non-linear relationships between the low-level features of different modalities, particularly when these modalities exhibit significantly different statistical properties. As an alternative, some studies opted to employ multiple encoders to handle multi-modal information separately [93]. Then, the multi-pathways are mid or late-fused to take advantage of latent correlations between modalities. Despite the success of vision Transformers [60] (Sect.2.6.2), only a few works have been inspired by their intrinsic advantages and scalability in modeling different modalities [94]. Concretely, they could encompass multiple sequences of tokens where each sequence's attribute represents a different modality, thus allowing a multi-modal learning framework without architectural modification. Furthermore, learning per-modal specificity and inter-modal correlation can be achieved by manipulating self-attention input patterns.



(*a*) parameters (M) (*b*) tera MACs

Figure 3.14 – Numbers of parameters (a) and multiply-accumulate operations (MACs) (b) for various convolutional, hybrid CNN-Transformers and pure Transformer segmentation networks compared in a multi-modal setting [**91**]. Multi-parametric segmentation of brain tumors and multi-modal segmentation of head and neck tumors respectivement correspond to Task 1 ($\mathcal{T}_1$) and Task 2 ($\mathcal{T}_2$).

In this context, the benefits of using one or multi-pathway CNN encoders, along with the integration with Transformers to leverage robust multi-modal features were unclear. Furthermore, the question of whether using or not a pure Transformer-based encoder [95] to solve multi-modal image segmentation tasks remained unanswered. To address these concerns, the post-doctoral work of G. Andrade-Miranda focused on the implementation of multi-modal image segmentation networks based on Transformers. The conducted work [**91**] aimed at comparing various convolutional, hybrid CNN-Transformers and purely Transformers-based models (Fig.3.13) in a controlled environment using well-labelled cross-modal datasets. We specifically looked into whether differences in performance can be explained by convolutional or Transformer stems, one-path or multi-path CNN-based encoders, and Transformer modality interaction schemes. Thus, we introduced a unified mathematical framework to describe current Transformer-based architectures and conducted a series of experiments to analyze different multi-modal segmentation networks. We additionally explored the gain variation in performance depending on one-path versus multi-path CNN-based encoder. Finally, the impact of single- and multi-stream (through cross-Transformer blocks) joint representations was also investigated.

Experiments on both multi-parametric segmentation of brain tumors and multi-modal segmentation of head and neck tumors showed that multi-path hybrid models combining convolutional layers and Transformers improve delineation accuracy over traditional methods, but at the cost of increased computation time and larger model size (Fig.3.14). Using 3D CNN feature tokenization (hybrid CNN-Transformers encoder) appeared better than employing a 3D patchify tokenization (pure Transformers-based encoder) strategy. Nevertheless, the performance of hybrid networks were primarily due to the multi-path CNN encoders rather than the Transformer itself, in agreement with a similar finding

recently made in [96]. In addition, we suggested caution while designing hybrid CNN-Transformers networks, particularly given the influence of spatial downsampling on the capacity of Transformers to describe long-range dependencies. No notable differences were found in the modality interaction scheme between single- and multi-stream models, either for hybrid or pure Transformer-based networks.

3.6 Conclusion

To bridge the gap between learning paradigms and clinical needs, recent investigations have struggle with novel and concrete emerging applications related to knowledge transfer and information fusion across heterogeneous domains. Multi-domain segmentation strategies have gained in popularity in order to fully exploit both complementary and redundancy across conditions (Sect.3.2), viewpoints (Sect.3.3), anatomies (Sect.3.4) or modalities (Sect.3.5). Multi-task and multi-domain techniques with multiple anatomies as targets appear as one promising avenue to be further investigated. Advances in multi-modal segmentation with Transformers pave the way for effective cross-modal transfer learning, where knowledge gained from one modality could be transferred to another. Given the complexity of collecting and annotating a large amount of medical images, transfer learning (Sect.3.2), self-supervised, semi-supervised (Sect.2.7) and active learning (Sect.3.3.2) are sub-fields of clear progress. However, more research efforts are needed to maximize or avoid the time-consuming and costly manual efforts made by clinical experts. Further research beyond the application of off-the-shelf solutions are needed to enable a wider adoption of image segmentation with deep learning into clinical routine, especially in surgery planning and longitudinal follow-up (Chap.4).

Overall, the potential for bias in deep approaches is a common concern across medical image analysis tasks including segmentation. In this context, encouraging the collection of large and diverse datasets through collective work with various experts is highly recommended. The development of challenges with publicly-available imaging data is an effort in this direction. Since medical data is often sensitive and subject to strict regulations on sharing, federated learning also appears as a promising field to develop since it offers the possibility for multiple hospitals and research institutions to collaborate by training a shared model on their own local data while keeping the data private and secure. Demonstrating a better reproducibility when designing deep learning pipelines could increase the trust and confidence of clinicians and make them more suitable for large-scale clinical applications. Finally, the development of lightweight models with few memory and computational resource requirements could be beneficial to ease the deployment of deep learning-based solutions on computationally-limited platforms.

Part II

On-going and future research activities

Chapter 4

Decision support in surgery planning and therapeutic follow-up

4.1 Introduction

Pre-operative planning and therapeutic follow-up are integral components of patient care, crucial for ensuring optimal outcomes and long-term survival. With advancements in medical technology and computational tools, decision support systems are emerging as invaluable aids for healthcare professionals. Surgery planning involves decision-making processes that require careful consideration of numerous factors, including patient-specific anatomy, disease features, surgical techniques, and potential risks. In this area, medically-sound computational models can benefit from imaging and clinical data to provide interpretable decision support and steadily push forward the integration of realistic computer-aided systems into clinical practice. When surgery is not an option, therapeutic follow-up entails monitoring patient longitudinally, assessing treatment efficacy, and making informed decisions regarding further interventions or adjustments to the treatment plan. Despite the enthusiasm generated by deep learning techniques for characterizing anatomical and pathological structures, the longitudinal follow-up of patients through the analysis of sequences of consecutive examinations has not been widely studied to date. However, modeling pathological evolution over time could enable a better therapeutic management, tailored to each individual patient.

This chapter summarizes my on-going and future research activities by addressing various facets of decision support systems in surgery planning and longitudinal follow-up, with a main focus on abdominal imaging. We start by exploring the intricate task of leveraging imperfect data in uncertainty-aware patient-specific modeling (Sect.4.2) before delving into the field of computer-assisted pre-operative planning, where we examine how computational tools can be employed to aid clinicians in optimizing surgical approaches such as liver resection (Sect.4.3). On its turn, therapeutic follow-up is first described by shedding light on medical image analysis methodologies for automatically predicting treatment response (Sect.4.4) and assessing immunotherapy eligibility (Sect.4.5). We lastly explore disease progression modeling (Sect.4.6), discussing the use of computational models to predict disease severity and model disease trajectories, towards improved patient outcomes.

4.2 Leveraging imperfect data in patient-specific models

A fully-automated patient-specific cartography arising from sophisticated deep models is of interest in surgical planning and therapeutic follow-up to guide image interpretation, ease decision making and improve patient care. By harnessing cutting-edge deep learning methodologies, my on-going and future developments aim at building accurate and thorough 3D maps of the patient anatomy from CT or MR imaging data. In this perspective, making multi-structure segmentation more robust and able to provide interpretable guidance tools is expected to gain the trust and benefit for both clinicians and patients. This is especially true when managing patients with hepatic metastases from colorectal cancer [97] where reaching a patient-specific liver cartography can assist clinicians in surgery planning (Sect.4.4.3) or chemotherapy response assessment (Sect.4.4.4). Another clinical use-case deals with both cystic liver and renal cartography to guide the follow-up of patients with polycystic kidney disease [55], thereby expanding upon past research activities (Sect.2.6.2).

The efforts to reach patient-specific thorough cartographies are driven by the inherent challenges of abdominal imaging, where variations in acquisition protocols, contrast agents, and lesion characteristics present obstacles for conventional approaches. Notably, dealing with abnormality detection given varying contrast enhancement levels, dissimilar resolutions and a large diversity in lesion type, shape, size and texture still need to be overcome through the deep learning paradigm. Modeling healthy organ tissues and lesions (e.g. tumors, metastases, cysts) is not always enough to provide a full patient-specific cartography. Blood vessel modeling is also essential for its completion. As an example, extracting and differentiating arterial, supra-hepatic and portal venous systems can divide the liver into height functionally independent areas [98], known as the Couinaud scheme. In this way, we transition from an intensity-based representation to a geometric and semantic description of the organ sub-structures. However, as highlighted in Chap.2, vascular segmentation faces its own limitations including class imbalance and appearance similarity with non-vascular tissues, complex multi-scale geometry with decreasing diameter along tree-like networks or variability in branching patterns.

To ensure sufficient robustness with respect to clinical requirements, many of the methodologies described in Chap.2 and 3 can be implemented. In particular, preserving both geometry and topology of the targeted anatomy (e.g. metastatic liver, cystic kidneys) is key. In this direction, medically-driven (Sect.2.5) or multi-scale (Sect.2.6.1) constraints within extended hybrid architectures leveraging high-resolution spatial information from convolutional features and global context encoded by Transformers (Sect.2.6.2) is one of the possible targeted avenues.

Leveraging external datasets. While such supervised learning methods may exhibit satisfactory performance when provided with sufficient labeled data, their effectiveness could be further enhanced by leveraging external but related labeled datasets. This brings us back to the issue of knowledge transfer (Sect.3.2), where learning transferability strategies tend to generalize the predictive capability of the network across different domains. A perspective in this area is to extend from MR to CT imaging

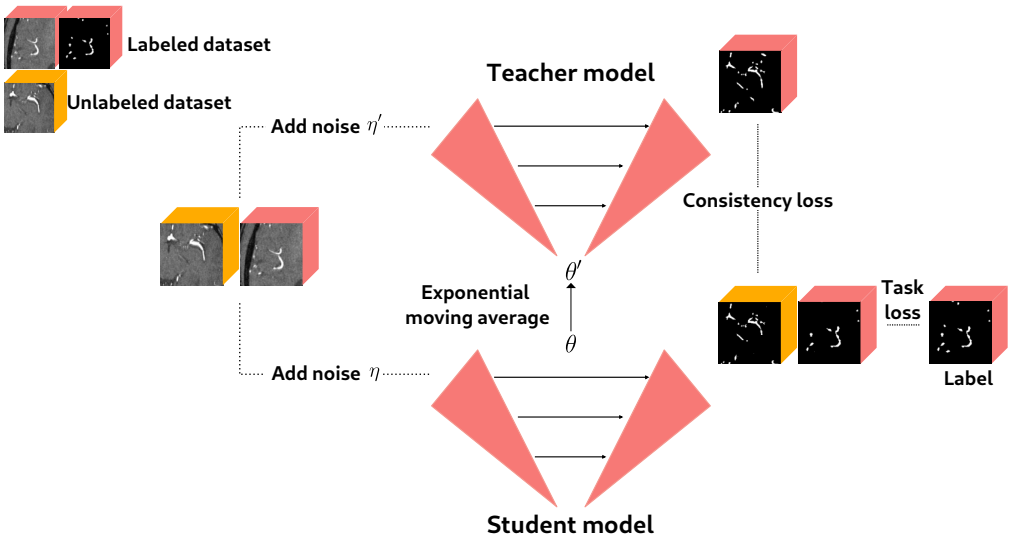


Figure 4.1 – Mean-teacher model leveraging labeled and unlabeled data for vessel segmentation [**100**].

modality the dual-task Transformers-based segmentation pipeline (Sect.2.6.2) tailored for polycystic kidney delineation. In particular, the idea would be to exploit CT imaging data from the related but different task of tumor kidney CT segmentation from the KiTS23 dataset [99]. Extending the dual-task architecture to a multi-task model comprising one single encoder and three task-specific decoders, one for each kidney and a third one for the liver, represents another pathway for improvement.

Foundation models, pre-trained neural network architectures that are typically trained in a self-supervised manner on massive datasets and designed to capture rich and general-purpose representations, can particularly benefit from external but related data sources [101]. Although vision foundation models trained on large-scale natural image datasets can provide a strong starting point (e.g. prompt-driven segment anything model [102]), organ or task-specific foundation model may provide better accuracy as well as significantly reduce the amount of labeled data required for a given target modeling task, as it has already learned more specific features. An optimal trade-off between developmental effort and practical efficacy must be found given the spectrum of emerging general [103] versus specialized [104] foundation models in medical image analysis. It is worth noting that future directions include multi-modality foundation models, combining various data types and scales, from genome to anatomy.

Leveraging both labeled and unlabeled data. Another scenario deals with changing the training paradigm and leveraging both labeled and unlabeled data through semi-supervised learning. Unlabeled data, being easier to obtain as it does not necessitate annotations, offers a valuable resource. According to [105], three distinct strategies can be followed: generating pseudo-labels to train a supervised model [106], using unlabeled data to perform unsupervised regularization [107] or employing unlabeled data to learn prior knowledge through self-supervised tasks [108]. The second category, which is the most commonly used, typically employs a mean-teacher [107] framework and incorporates various schemes,

whether it leverages consistency learning (Fig.4.1), co-training, adversarial learning or entropy minimization. By exploiting the advantages of both labeled and unlabeled data, semi-supervised segmentation techniques appear attractive in addressing the scarcity of labeled data. Such approach can not only enhance model performance but also induce regularization to mitigate overfitting. Nevertheless, it is crucial to include an adequate amount of labeled data to ensure the network comprehensively learns anatomical concepts, with the sufficiency being contingent upon annotation quality. Consequently, when faced with a new segmentation task, striking a balance between labeled data quantity, quality, and the selected semi-supervised approach becomes paramount.

In this direction, an on-going collaboration with CREATIS[9] and CReSTIC[10] laboratories aims at investigating the data dependency of deep learning methods within the context of imperfect data and semi-supervised learning for cerebrovascular segmentation **[100]** (Fig.4.1). The goal of this study is to compare various state-of-the-art semi-supervised methods based on unsupervised regularization and to evaluate their performance in diverse quantity and quality data scenarios to provide guidelines for the annotation and training of cerebrovascular segmentation models. We plan to benchmark the same semi-supervised learning methodologies to improve polycystic kidney MR segmentation by taking advantage of unlabeled images provided by NIH and arising from CRISP and HALT clinical studies.

Leveraging imperfect annotations. The issue of imperfect annotations is intricately connected with the problem of annotation scarcity. When faced with a shortage of labeled data, the lack of diversity in annotations can lead deep learning models to overfit to specific concepts or noise present in the training labels and introduce biases. Consequently, there is an interest in investigating these issues together. In blood vessel segmentation, the obstacles posed by data scarcity, concept shift, and noisy labels are more prominent than in other segmentation applications. As a result, these challenges stand out as significant performance bottlenecks for supervised methods. Thus, we also aim at investigating the impact of several common imperfection types found in ground truth labels such as missing sub-structures, over- or under-delineation issues.

Leveraging partially-labelled datasets. Collectively, existing open datasets in medical image segmentation offer extensive information on different target structures. However, each dataset is incompletely labeled, in the sense that it contains annotations for certain structures only. Consequently, achieving a thorough segmentation of all anatomical entities necessitates training numerous independent models, one for each dataset. A multi-class approach able of simultaneously learning from multiple partially-labeled datasets would offered several advantages: shared representation over datasets, enhanced robustness to changes in the field of view and significant reductions in both training and inference time compared to multiple individual networks. Managing such avenue using dataset- and class-adaptive loss functions would allow us to obtain an holistic whole-abdominal segmentation model.

9. `https://www.creatis.insa-lyon.fr/`
10. `https://crestic.univ-reims.fr/`

Leveraging uncertainty estimation. When employing deep learning for image segmentation, the forward pass operates a deterministic process that assigns a unique label to each voxel. This apparent determinism however fails to take into account the various sources of uncertainty that affect neural network predictions. Understanding these uncertainties would be useful to detect potential segmentation errors. Therefore, there is a clear imperative to understand the limitations of segmentation models through the evaluation of voxel-wise confidence measures. This objective underpins the application of epistemic (model-related) and aleatoric (data-related) uncertainty quantification to segmentation tasks [109]. Among the most widely used methods for measuring epistemic uncertainty is Monte-Carlo dropout [110], which involves conducting multiple stochastic dropout forward passes of a model equipped with dropout weights during training. Uncertainty modeling may also importantly be used to directly improve segmentation performance, as with test-time data augmentation [111], an aleatoric uncertainty estimation method where multiple forward passes are executed on inputs altered through basic data augmentations. Uncertainty quantification being an emerging trend, more contributions are expected to provide clinicians with results and associated confidence information.

4.3 Computer-assisted pre-operative planning

The management of patients with colorectal cancer, second leading cause of cancer death [97], is a major public health issue. Half of the patients with colorectal cancer develop a distant recurrence. The liver, through the development of liver metastases, is the most common spread site, accounting for 15-25% of patients at diagnosis and a further 18-25% of patients within 5 years [112]. The objective is to find a management adapted to each patient by integrating individual (e.g. age, comorbidity), tumoral (e.g. number, size) and collective data. With an estimated 5-year survival rate from 37% to 58%, hepatic resection consists of the complete removal of lesions, leaving at least 30% of the parenchyma.

In this context, a full patient-specific liver cartography including metastasis and vessel tree delineations could make it possible to automatically report the number and size of metastases, their location in relation to the vascular systems as well as their distribution for each functionally independent Couinaud segment [98] derived from the vascular segmentation. This information can greatly improve image interpretation guidance and ease the decision making process, especially to evaluate the resectability status of each patient. Indeed, resectable patients are those whose metastases can be completely removed, leaving at least 30% of healthy liver parenchyma to limit the risks of post-operative failure. However, determining the feasibility of liver resection is prone to variability between experts. Exploiting both automatic liver metastasis and vessel contours to predict with deep learning the respectability status as a classification problem could act as a relevant second-opinion.

Apart from resectability status prediction, my future works aim at building a pre-operative liver resection planning tool benefiting from a fully-automated liver cartography comprising precise metastasis (Sect.3.2.2) and vessel (Sect.2.5.3, 2.5.4) delineations. Despite advances in chemotherapy and

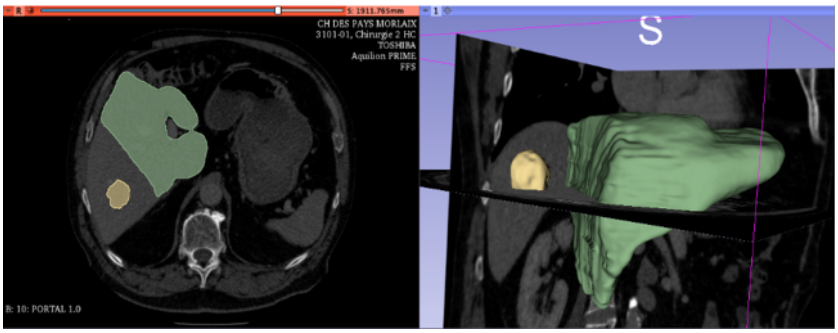


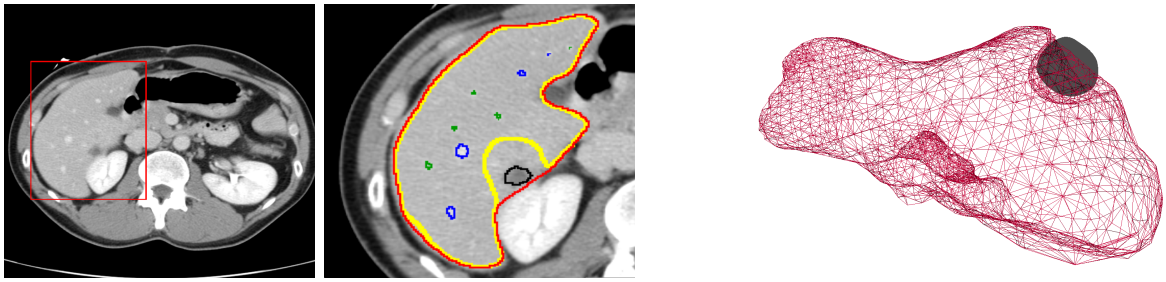
Figure 4.2 – Simulation of right hepatectomy by an experienced surgeon. Liver area remaining after surgery is displayed in green, metastasis to be resected in yellow.

targeted therapies, only resection or destruction of lesions by local treatment can achieve complete remission. Full recovery from major hepatic resection requires a healthy, well-perfused liver remnant able to regenerate [113]. Besides, the percentage of functional parenchyma remaining after major resection is an important predictor of post-operative dysfunction and morbidity [114]. A precise pre-operative planning of the future liver remnant is hence crucial since accurately simulating the resection plan (Fig.4.2) enables to predict the post-operative volume and optimize the remaining healthy areas [113].

Two virtual pre-operative resection methodologies are currently investigated, application which has never been addressed with deep learning before, to our knowledge. The first strategy aims at defining the resection plan around predicted metastases using simple geometric rules, taking into account both identified vein and artery branches as well as appropriate safety margins [115]. As an alternative, encoder-decoders derived from UNet [26] and integrating as inputs predicted metastatic and vascular tissue location information can be trained to automatically infer the area to be resected.

To evaluate the feasibility of this latter strategy, a preliminary study conducted in collaboration with University of Bejaia delved into the application of convolutional networks for the delineation of future liver remnants in patients with colorectal liver metastases undergoing resection. Our approach utilized pre-operative segmentation masks, thereby presenting a paradigm shift in predicting the resection area compared to conventional methodologies which often rely on Couinaud segments only [116]. Our experiments involved a pre-operative CT scan along with binary ground truth masks representing the liver, hepatic veins, portal veins, and metastases (Fig.4.3*a*) as inputs of a DX-Net [**77**] segmentation model. Various combinations of input sets were explored to determine the most accurate strategy through experiments performed on available simulations of hepatic surgical resections collected from University Hospital of Brest and NIH[11]. The outcomes underscored the substantial impact of the chosen input set. In particular, the input set encompassing liver, hepatic veins, portal veins and tumoral masks achieved the highest Dice score. The exclusion of the CT scan did not hinder but rather contributed to superior performance, underlying the robustness of the binary masks in capturing the essential information for precise remnant delineation (Fig.4.3*b*).

11. `https://www.cancerimagingarchive.net/collection/colorectal-liver-metastases/`



(a) (b)

Figure 4.3 – (*a*) CT scan and delineation of liver (red), liver post-resection (yellow), metastasis (black), hepatic (blue) and portal (green) veins. (*b*) Visualization of metastatic tissue (black) and futur liver remnant (red) predictions using DX-Net [**77**].

Although the feasibility of such an approach has now been demonstrated, additional experiments are required, especially by testing other deep networks (e.g. Transformers-based) and by using fully-automated (instead of ground truth) metastasis and vessel segmentation masks. It is worth noting that the accuracy of metastasis and vessel delineations are of critical importance to ensure an appropriate guidance for pre-operative resection planning. This challenge need be overcome by relying on robust segmentation methodologies (e.g. multi-task and adversarial training with Transformers, geometric and topological constraint integration) to exploit discriminating contextual information, towards better medically-driven representations of hepatic structures.

4.4 Treatment response assessment and prediction

When surgery is not an option due to the tumor burden, the therapeutic regimen for the management of patients with cancer usually consists of palliative oncological treatments. In this context, CT image analysis is a critical step in assessing the response to treatments. Most evaluation methods are based on measures related to lesion size. In particular, the response evaluation criteria in solid tumours (RECIST) 1.1 [9] criterion deals with a unidimensional assessment of lesions which are classified into 3 categories: target, non-target or new lesions. The existence of progression can be affirmed by a progression of the diameter of target lesions of 20% or more, in case of indisputable increase of non-target lesions or when at least one new lesion has appeared. However, morphologic criteria are not suited for treatments that give tumor necrosis (anti-angiogenic agents) or that provide an immune response (immunotherapy) since density information is not taken into account. Moreover, these criteria require identifying and segmenting tumoral areas (e.g. hepatic metastases) before measurement. Still largely performed manually by clinicians, these time-consuming tasks are prone to strong intra and inter-expert variability [117]. This finding also limits the option of extracting quantitative indices from CT scans to determine the treatment response group in a radiomic fashion [118]. Fully automating both assessment and prediction of the treatment response could allow clinicians to stop an unnecessary and potentially toxic treatment and to substitute it for a more efficient therapeutic alternative.

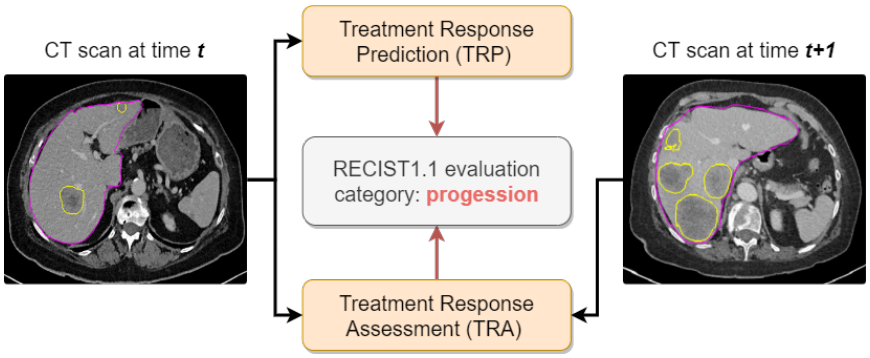


Figure 4.4 – Treatment response assessment and prediction for patients with liver metastases from colorectal cancer **[119]**. Liver and liver metastases are respectively with pink and yellow boundaries.

Deep learning has been widely used in medical image analysis for a variety of diagnosis purposes but fairly few in the domain of treatment response assessment. A first idea would be to validate a measure of disease progression using tumor delineations performed on baseline and follow-up CT scans through an automated deep learning-based segmentation pipeline (Chap.2, 3) and therefore standardize both RECIST 1.1 [9] and radiomic [118] evaluations which may differ due to manual contouring pre-requisites. Another key advantage of this avenue would be to consider all the detected lesions (i.e. not just the main ones) and to fully exploit their 3D spatial extent.

Furthermore, the majority of studies concentrate on disease identification and evaluation via the analysis of images obtained at a single time point. For example, a two-step approach was proposed in [120] for predicting treatment response of liver metastases from CT scans. Untreated liver lesions were identified and segmented before an Inception network was applied to predict the treatment response. This method is fundamentally constrained in terms of evolution estimation since it does not account for therapy-induced changes. Given the complexity of 3D volumetric data, there remains a great demand for an approach that effectively captures the dynamic information from baseline and follow-up images.

More recent studies are taking the longitudinal fact into consideration. Pre- and post-treatment MR images were exploited in [121] to predict the tumor response to chemotherapy in patients with colorectal liver metastases through a 3D multi-stream deep convolutional network. Response prediction was treated as a binary classification task by creating two synthetic groups (response versus non-response), out of the 4 groups (complete response, partial response, progressive disease, stable disease) defined by RECIST 1.1 [9]. This is prohibitive since in this scenario, clinicians would not have the fine-grained division into multiple response groups, which is essential for patient-specific stratification. None of existing studies fully address treatment response assessment which aims to evaluate the response (considering all response groups from RECIST 1.1) of a certain time frame given a specific chemotherapy regimen with pre- and post-treatment images. Further, treatment response prediction whose goal is to predict the response after a certain time frame given a specific chemotherapy regimen with the pre-treatment CT scan only is an issue that has not been widely addressed so far.

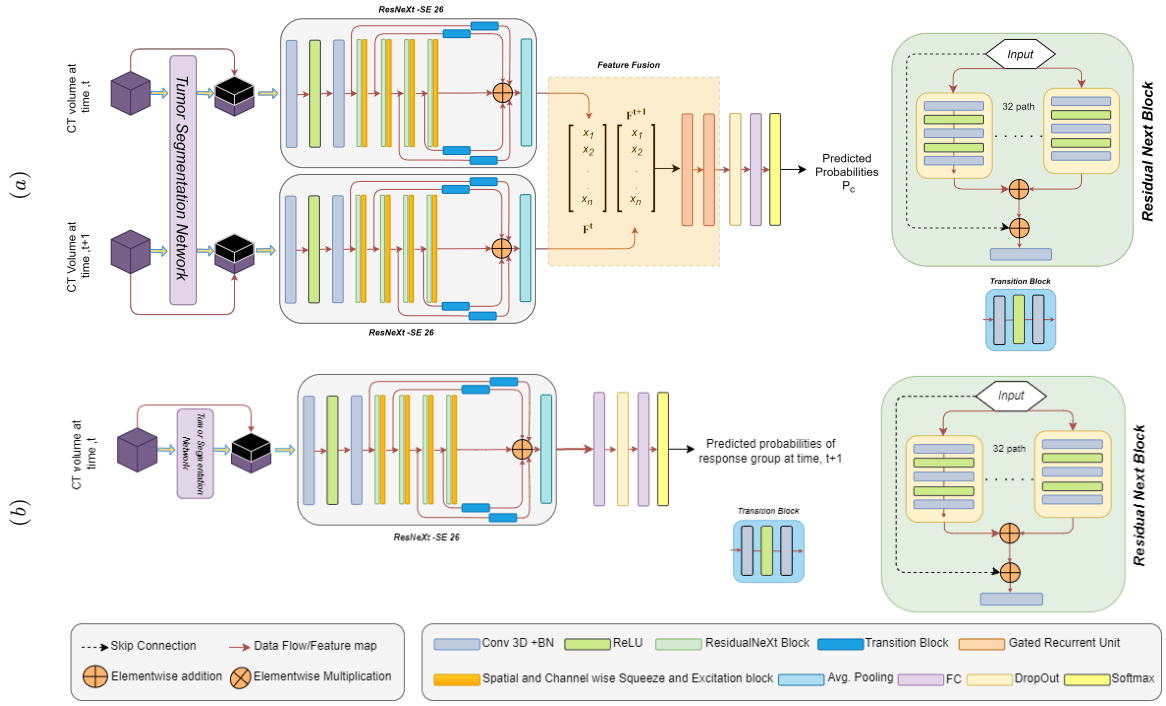


Figure 4.5 – Chemotherapy response assessment (a) and prediction (b) pipelines from **[119]**.

As a first attempt, we developed in **[119]** a deep learning-based treatment response assessment pipeline and its extension for prediction purposes based on a newly designed 3D Siamese classification network (Fig.4.5). Treatment response assessment was performed through a Siamese network built upon a modified 3D ResNeXt-26 [122] architecture. The features extracted by each stream of the Siamese network from CT scans acquired at both time stamps were fused through sequences of gated recurrent units (GRU) and fully-connected layers. To predict the treatment response at time $t+1$, we adapted the previously defined pipeline by only feeding the model with the CT scan acquired at time t and by stripped off the feature fusion mechanism. A single 3D ResNeXt branch was thus required, instead of a Siamese-like architecture. By achieving 94.94% and 86.86% overall accuracy scores respectively, the effectiveness of both treatment assessment and prediction frameworks was illustrated on the PRODIGE 20 dataset[12] [123] collected from a phase-II multi-center clinical trial in metastatic colorectal cancer that evaluated chemotherapy alone or combined with Bevacizumab in elderly patient during follow-up. GradCAM [124] was used to see what the model is focusing on to make the decision (Fig.4.6). We noticed that the model learned to focus on the region of the largest metastasis cluster inside the liver. The size of the primary metastasis cluster, as well as darker regions (i.e. necrotic tissues), seemed to play a significant role in producing higher levels of activation.

As perspectives, new multi-stream Siamese networks based on Transformers could be proposed to improve both assessment and prediction tasks. In particular, we plan to develop, in the same

12. provided by Fédération Francophone de Cancérologie Digestive (FFCD), `https://www.ffcd.fr/`

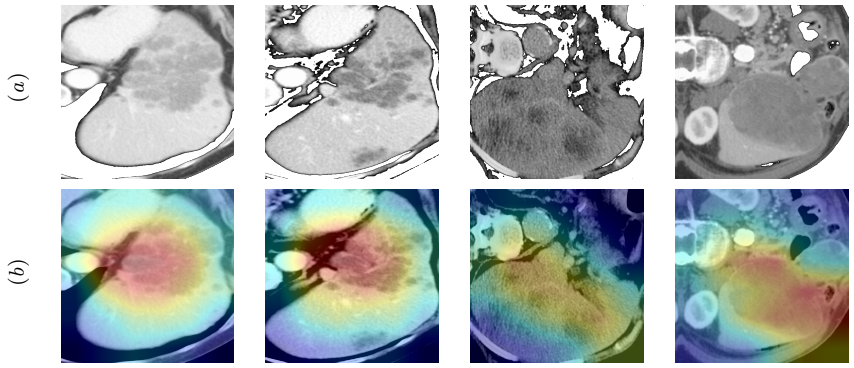


Figure 4.6 – CT scans (a) and gradient-weighted class activation heat maps (b) with GradCAM [124] for treatment response prediction **[119]**. Higher activation corresponds to reddish areas.

spirit as [125], a multi-task deep learning approach that allows simultaneous tumor segmentation and response prediction. A longitudinal data augmentation strategy will be considered by taking into account non-consecutive pairs of CT examinations. These developments will not only be evaluated on PRODIGE 20 [123] (237 CT scans) but will also benefit from the availability of data[12] arising from the PRODIGE 9 [126] phase-III clinical trial (576 CT). It is worth noting that a parallel on-going study in collaboration with University Hospital of Reims tends to confirm that organ atrophy is a marker of therapeutic response to Bevacizumab in colorectal cancer patients. On a broader scale, typical patterns of pathology evolution will be extracted from both PRODIGE datasets to find early markers of chemotherapy response with, in fine, a possible gain in survival. The ultimate goal would be to build a complete chemotherapy regimen recommendation system able to predict the best treatment for each patient and which would include other organ assessment such as lungs, lymph nodes and bones.

4.5 Immunotherapy eligibility assessment

Immunotherapy is a treatment that is increasingly involved to stimulate the immune system. Cancer cells express negative molecules, called immune checkpoints, which block the action of the immune system, especially lymphocytes. Immunotherapy works by freezing these negative immune checkpoints. In oncology (e.g. in lung or gastric cancers), the expression of immune checkpoint inhibitors such as programmed cell death-ligand 1 (PD-L1), protein determined by immunohistochemistry (IHC), is considered as a key factor to trigger the use of immunotherapy. However, obtaining the PD-L1 expression status requires surgical or biopsied tumor specimens collected through invasive procedures, with associated risks of morbidities [127]. In addition, the process takes a long time and requires a strong expertise from pathologists. Such analysis can be further affected by the tumor heterogeneity, which does not fully guarantee reproducibility. Conversely, an alternative non-invasive method able to evaluate the PD-L1 status would have an important impact in clinical decision support, especially when tissues are not available or when IHC fails [128]. Given the established correlation between imaging

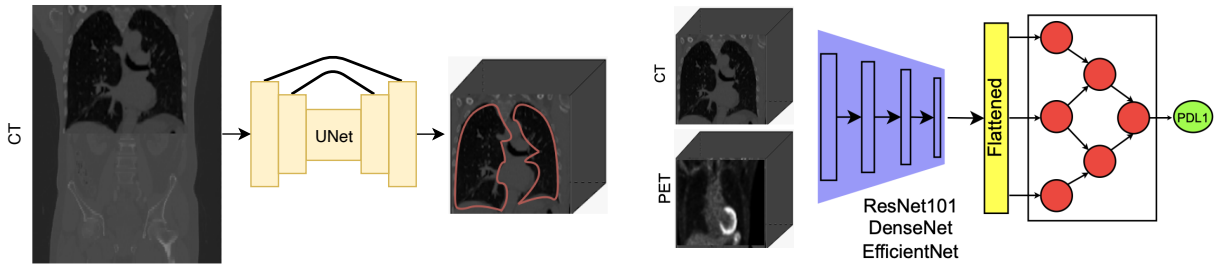


Figure 4.7 – PD-L1 status prediction based on CT and PET imaging data for lung cancer **[132]**.

and lymphocyte infiltrate [129], we hypothesize that the analysis of CT and positron emission tomography (PET) images can provide an alternative surrogate for reaching the PD-L1 expression status in clinical practice. This represents a promising prospect for the management of poor prognosis diseases.

An emerging avenue deals with radiomics whose potential to uncover tumoral patterns and characteristics that fail to be appreciated by the naked eye is encouraging. Standard radiomics approaches consist in providing a radiomic signature derived from multi-parametric image analysis exploiting homogeneity, contours, texture, or density information from the investigator-defined primary lesion area [130]. Additionally to the design of handcrafted imaging features which necessitates expert knowledge, such assessment requires the manual contouring of the primary lesion which is time-consuming and prone to expert variability between clinicians. Recently, deep convolutional models have been recently integrated into radiomics frameworks. This integration can be done by automating the lesion delineation task **[32]** or by extracting deep features from intermediate hidden layers. Compared to handcrafted features, deep features contain more representative and high-level medical image information and provide more predictive patterns to fully address the target task. In this context and with the aim of extending current deep radiomics frameworks [131], my current research activities aim at developing medical image analysis pipelines acting as a digital biopsy and able to predict clinical variables (e.g. PD-L1 expression) from PET/CT or CT-only images acquired before any systemic treatment, with applications for immunotherapy eligibility assessment.

In the continuity of early studies [133], [134], a first step in this direction was to investigate a variety of deep learning-based architectures to perform PD-L1 status prediction based on CT and PET imaging data for non-small cell lung cancer (NSCLC), a significant source of disease-related mortality. In **[132]**, a study made in the framework of the post-doctoral position of R. Da-Ano, we thus compared ResNet, DenseNet and EfficientNet models using different settings : CT only, PET only and PET/CT fusion performed in an early fashion. The dataset comprised a cohort of 189 patients collected from University Hospital of Poitiers, with available PET and CT images along with corresponding lung tumor delineations. Since it was essential to exclusively focus on the lung area, we used a UNet model [26] trained on the LIDC-IDRI dataset [135] to create a bounding box around the lungs. The resulting bounding boxes were given as inputs of the deep architectures (Fig.4.7) to estimate the PD-L1 expression. Models were evaluated using the areas under the receiver operating

characteristic curves (AUC) based from their 95% confidence intervals. Experiments showed that the fusion architectures consistently outperformed the standalone modalities (CT only, PET only), highlighting the benefit of integrating multi-modal information from both imaging modalities. Further, early PET/CT fusion with ResNet and DenseNet as our first efforts to better categorize the PD-L1 status for lung cancer demonstrated encouraging results with 0.81 in mean AUC.

Regarding PET/CT data analysis, a strong focus is currently given to the way both modalities can be fused. Apart from CT only, PET only and early PET/CT fusion, considering different schemes comprising late PET/CT fusion without as well as with partially and fully shared weights is currently under investigation to determine the best fusion strategy. We also aim at comparing different image configurations as inputs of the deep architectures, from broadly extended bounding boxes that incorporate lung regions and surrounding tissues to low-range patches centered around the main lesion.

In the near future, both lesion delineation and PD-L1 expression assessment tasks will be simultaneously handled through multi-task learning. In addition, since Transformers are likely to become the new horizon for image analysis applications beyond CNN architectures which do not fully explicitly model long-range relations [60], hybrid models combining the complementary strengths of CNN and Transformers will be investigated. Apart from non-invasive PD-L1 status measurement, our strategies tend to be enough generic to be easily extended to the automated prediction of other clinical variables (e.g. EGFR mutation status) as well as treatment response, progression-free and overall survivals, in a wide range of clinical applications including gastric cancers. With the ultimate goal of identifying a universal biomarker able to select patients eligible for immunotherapy, one can imagine building a model considering images from different anatomies, pathologies or grades of severity.

4.6 Disease progression modeling

Longitudinal image analysis deals with capturing both static anatomical structures and dynamic changes in disease progression, towards earlier and better patient-specific pathology management. In clinical routine, radiologists often compare the current examination to prior studies from the same patient. Although several studies [136], [137] attempted to automate the analysis of longitudinal images, conventional computer-aided diagnosis approaches typically do not effectively utilize longitudinal information for decision-making purposes. However, the visual patterns predicting disease progression that exist in follow-up images could be extracted through deep learning to improve the guidance in clinical tasks (e.g. diagnosis, grading, therapeutic management) with respect to single-image strategies.

In my research activities, the topic of longitudinal follow-up has mainly taken place in the field of ophthalmology, with a focus on diabetic retinopathy (DR), common and high-risk complication of diabetes and leading cause of visual impairment and blindness worldwide [139]. Although regular screening is crucial for preventing blindness, the expected increase in the number of patients with diabetes means that the burden of screening and follow-up represents a substantial challenge [140]. In

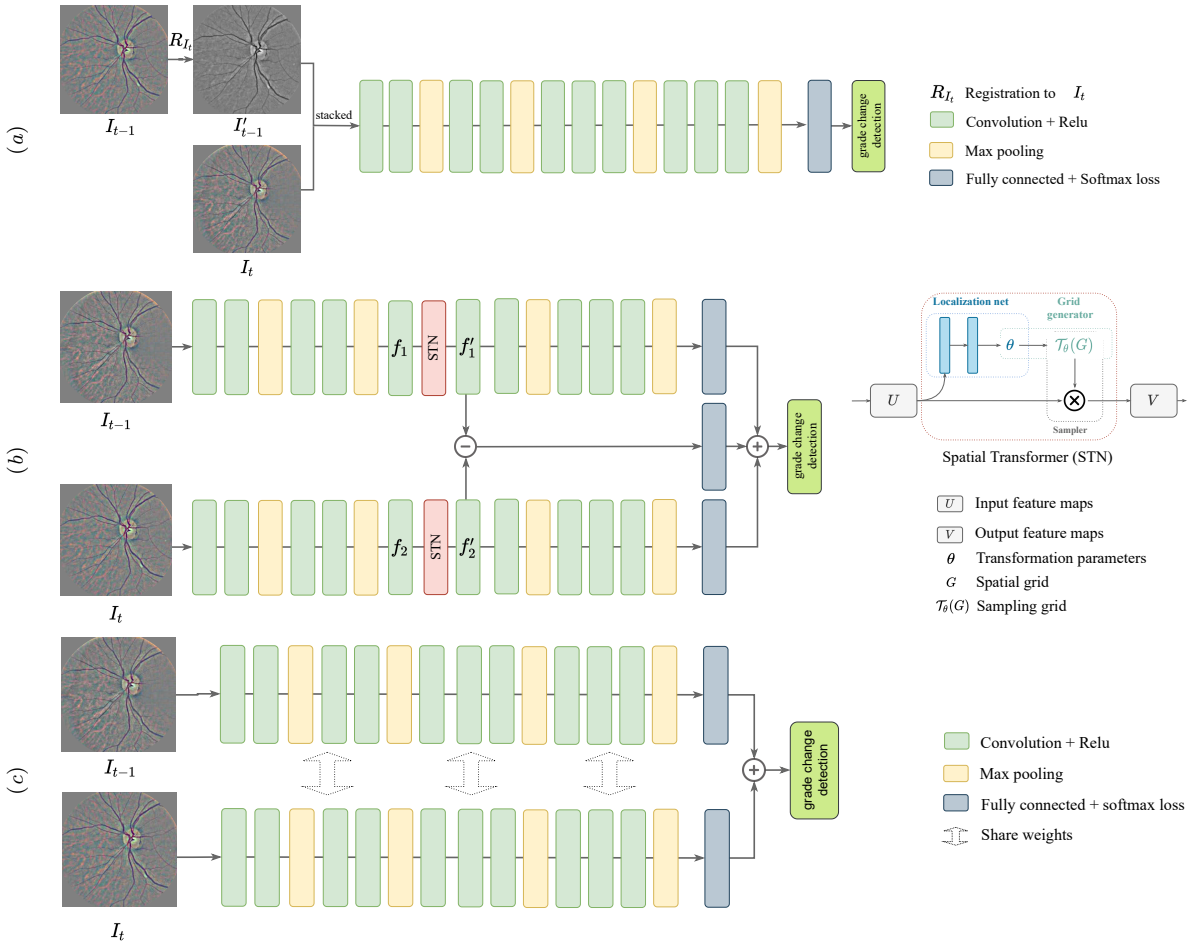


Figure 4.8 – Early (a), intermediate (b) and late (c) fusion strategies compared in **[138]** for diabetic retinopathy severity grade change detection.

this context, exploiting follow-up color fundus photographs (CFP) to improve severity and progression assessment by taking advantage of both past and current examinations is highly required.

Longitudinal fusion. A first step in this direction was reached during the PhD thesis of Y. Yan with the development of deep information fusion frameworks exploiting two consecutive longitudinal studies for the assessment of early DR severity changes **[138]**. In this regard, we aimed at integrating longitudinal information of CFP images to help in predicting referable DR (moderate to severe) severity changes. Specifically, we targeted the change detection between no DR/mild non-proliferative DR and more severe DR by analyzing two consecutive follow-up examinations. To this end, we explored three fusion methods that exploit current and prior studies: early fusion of input images, intermediate fusion of feature vectors incorporating spatial transformer networks (STN) and late fusion of feature vectors (Fig.4.8). We conducted a comprehensive evaluation by comparing these pipelines on the longitudinal DR screening OPHDIAT dataset [141], a massive CFP database made of examinations acquired from

101,383 patients between 2004 and 2017. To our knowledge, this work was the first to automatically assess the early DR severity changes between consecutive images. Exhaustive experiments compared with respect to no-fusion baselines validated that incorporating prior DR studies can improve referable DR severity classification performance, especially through the late fusion scheme.

Longitudinal self-supervised learning. More recently, self-supervised learning hold great promise to learn robust high-level representations by training on pre-text tasks [142] before solving a supervised downstream task. Current self-supervised models are largely based on contrastive learning [54]. However, the choice of the pre-text task to learn a good representation is not straightforward, and the application of contrastive learning to medical images is relatively limited. To go further, the PhD thesis of R. Zeghlache is currently investigating the benefit of exploiting self-supervised learning with a longitudinal nature for DR diagnosis and progression prediction purposes. Longitudinal self-supervised learning was initially introduced in the context of disease progression through pre-text tasks. Thus, by considering consecutive image pairs as inputs, a Siamese-like model was trained in [143] to predict the difference in time between the two examinations. Latter, additional efforts aimed at conceptualizing longitudinal pre-text tasks as a way to comprehend the progression of diseases. In [144], longitudinal self-supervised learning was integrated into an auto-encoder using pairs of consecutive scans as inputs. A cosine alignment term was introduced alongside the traditional reconstruction loss to encourage the latent space topology to adapt in accordance with longitudinal changes. The advantages are to avoid any registration requirements, to leverage population-level data to capture longitudinal changes that are shared across individuals and to offer the ability to visualize individual-level changes.

The first contribution of R. Zeghlache in this direction was to compare different longitudinal self-supervised learning methods including a Siamese-like model trained in a self-supervised learning fashion (with time prediction between examinations), an auto-encoder trained with a loss forcing the trajectory vector in latent space to be aligned with the disease progression as well as longitudinal neighborhood embedding to model the disease progression from longitudinal retinal CFP **[145]**. The results obtained on the OPHDIAT dataset [141] suggested that the latent space arising from longitudinal self-supervised learning enables to encode the dynamic of DR progression.

Neural ordinary differential equations. Since many temporal systems can be described by means of ordinary differential equations (ODE), we made the connection between longitudinal self-supervised learning and neural ordinary differential equations (NODE) **[146]**, neural network architectures that learn the dynamics of ODE through the use of neural networks. The approximation of an unknown ODE is performed through a neural network ϕ that parameterizes the continuous dynamics of hidden units $\mathbf{z} \in \mathbb{R}^n$ over time with $t \in \mathbb{R}$ [147]. NODEs are able to model the instantaneous rate of change of $\mathbf{z}$ with respect to t using ϕ with trainable parameters Θ_ϕ:

$$\lim_{h \to 0} \frac{\mathbf{z}_{t+h} - \mathbf{z}_t}{h} = \frac{\partial \mathbf{z}}{\partial t} = \phi(t, \mathbf{z}, \Theta_\phi) \tag{4.1}$$

The analytical solution of Eq.4.1 is given by:

$$\mathbf{z}_{t_1} = \mathbf{z}_{t_0} + \int_{t_0}^{t_1} \phi(t, \mathbf{z}, \Theta_\phi)\mathrm{d}t = \texttt{ODESolve}(\mathbf{z}(t_0), \phi, t_0, t_1, \Theta_\phi) \tag{4.2}$$

where $[t_0, t_1]$ represents the time horizon for solving the ODE. By using a black-box ODE solver introduced in [147], we can solve the initial value problem (IVP) and calculate the hidden state at any desired time (Eq.4.2). We can differentiate the solutions of the ODE solver with respect to the parameters Θ_ϕ, the initial state $\mathbf{z}_{t_0}$ at initial time t_0 and the solution at time t. This can be achieved by using the adjoint sensitivity method [147]. Through the latent representation of a given image, we define an IVP that aims to solve the ODE from t_i to a terminal time t_{i+1}:

$$\dot{\mathbf{z}}(t) = \phi(\mathbf{z}(t), t, \Theta_\phi) \text{ with initial value } \mathbf{z}(t_i) = \mathbf{z}_{t_i} \tag{4.3}$$

NODEs are ideally suited for solving time-related problems. In particular, irregular time series [148], an inherent aspect of the disease progression context, can be leveraged by NODEs thanks to their ability to deal with continuous time.

Longitudinal mixing training. Based on the NODE formalism, we extended in **[149]** mix-up [150] and manifold mix-up [151] techniques to a longitudinal context referred to as longitudinal mixing training. Mix-up was introduced in [150] as a simple regularization method to minimize overfitting in deep neural networks. It linearly interpolates a mini-batch of random examples and their labels to transform the training set. Manifold mix-up is an extension of mix-up to hidden representations [151].

Let $\mathcal{V}$ be a set of consecutive patient-specific image pairs. $\mathcal{V}$ contains all $(\mathbf{x}_{t_i}, \mathbf{x}_{t_{i+1}})$ that are from the same patient where $\mathbf{x}_{t_i}$ is scanned before $\mathbf{x}_{t_{i+1}}$ with $i \in [0, m-2]$, m being the number of acquired examinations. Let $g_{1:n}$ be a backbone with n layers, where $g_{1:k}$ denotes the part of the neural network mapping the input data to the hidden representation at layer k. h_l represents a classification or regression head with index l, $(\mathbf{y}, \mathbf{y}')$ one-hot labels, $\text{Beta}\,(\alpha, \alpha)$ the Beta distribution and $\ell(.)$ the cross-entropy loss. The mixing operator is defined by: $\texttt{Mix}_\lambda(a, b) = \lambda \cdot a + (1-\lambda) \cdot b$ with $\lambda \sim \text{Beta}\,(\alpha, \alpha)$ $\in [0, 1]$. With manifold mix-up, a random layer k from a set of eligible layers in the neural network is selected. It processes two random data mini-batches $(\mathbf{x}, \mathbf{y})$ and $(\mathbf{x}', \mathbf{y}')$, until reaching layer k. The mix-up is then performed on these two intermediate mini-batches $(g_k(\mathbf{x}), \mathbf{y})$ and $(g_k(\mathbf{x}'), \mathbf{y}')$, continuing the forward pass with the mixed representation until the ending layer n. These mixed representations are then fed to a classification head h_l and projected to the number of classes.

Conversely, the proposed longitudinal mixing training was set up as follows. Let us denote $s(t_i)$ the severity grade of image $\mathbf{x}_{t_i}$. The function $s(t)$ is the severity interpolation function between two consecutive longitudinal pairs $\mathbf{x}_{t_i}$ and $\mathbf{x}_{t_{i+1}}$, defined in the linear case as: $s(t) = \frac{t-t_i}{t_{i+1}-t_i} \cdot (s_{t_{i+1}} - s_{t_i}) + s_{t_i}$. In conventional mix-up training, the labels are mixed. Instead, we proposed in **[146]** to mix the time between consecutive pairs $t_{\texttt{mix}} = \texttt{Mix}_\lambda(t_i, t_{i+1})$, then employ this $t_{\texttt{mix}}$ to evaluate $s(t)$ and use this

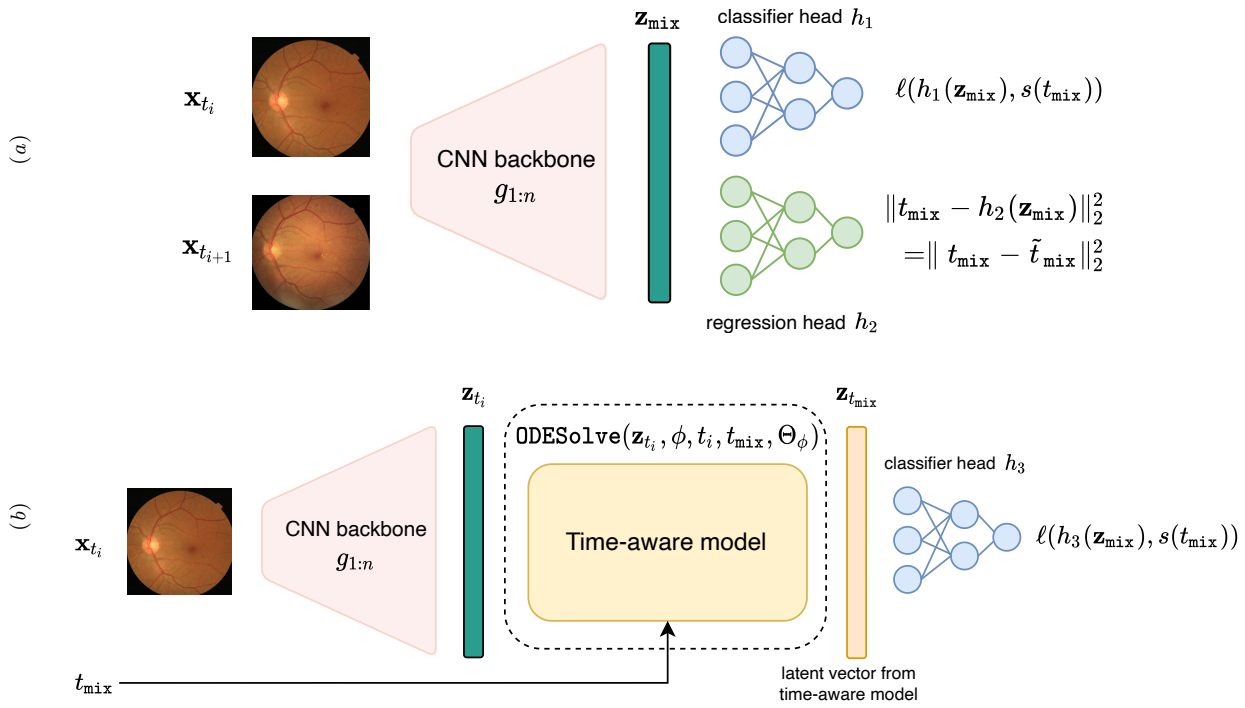


Figure 4.9 – Illustration of (a) longitudinal mixing training and (b) time-aware model training using $t_{\texttt{mix}}$ **[149]**. (a) and (b) can be trained simultaneously or independently.

signal as supervision. During training, we one-hot encoded the value of interpolation at $t_{\texttt{mix}}$ to get our soft label. Depending on the mix-up method, we obtained as latent representation of the mixed pair:

$$\mathbf{z}_{\texttt{mix}} = \begin{cases} g_{1:n}(\texttt{Mix}_\lambda(\mathbf{x}_{t_i}, \mathbf{x}_{t_{i+1}})) & \text{for mix-up} \\ g_{k+1:n}(\texttt{Mix}_\lambda(g_{1:k}(\mathbf{x}_{t_i}), g_{1:k}(\mathbf{x}_{t_{i+1}}))) & \text{for manifold mix-up} \end{cases} \tag{4.4}$$

In this context, $g_{1:n}$ was trained by considering a standard classification loss $\ell(h_1(\mathbf{z}_{\texttt{mix}}), s(t_{\texttt{mix}}))$ with h_1 a classification head (Fig.4.9a) as well as an additional time consistency loss following:

$$\mathcal{L}_{t_{\texttt{mix}}} =\| t_{\texttt{mix}} - \tilde{t}_{\texttt{mix}} \|_2^2 =\| t_{\texttt{mix}} - h_2(\mathbf{z}_{\texttt{mix}}) \|_2^2 \tag{4.5}$$

where h_2 is a regression head that predicts the value of the current $t_{\texttt{mix}}$ for a given pair (Fig.4.9a). Inspired by [143], this term was motivated by [152] where manifold mix-up coupled with self-supervised learning showed to enhance the quality of feature extraction. Longitudinal mixing training can be seen both as a regularizer and as a pre-text task that encodes the disease progression in latent space.

To accurately predict the disease progression from a single image, we trained a NODE as time-aware model using $t_{\texttt{mix}}$. Instead of solving the ODE from t_i to a terminal time t_{i+1} (Eq.4.3), we solved it to the intermediate time $t_{\texttt{mix}}$ (Fig.4.9b). Then, we used this $t_{\texttt{mix}}$ to evaluate $s(t)$ and take this signal as supervision for training. In practice, the NODE was trained using $\mathbf{x}_t$ and $t_{\texttt{mix}}$ as inputs through

$\ell(h_3(\mathbf{z}_{\texttt{mix}}), s(t_{\texttt{mix}}))$ with h_3 being a classification head (Fig.4.9*b*). It is worth noting that this approach could be applied with any time-aware model. Using OPHDIAT [141], we demonstrated that we can predict whether a patient would develop a severe DR in the following visit using a single image, with an AUC of 0.798. Our results also indicated that our longitudinal pre-text task can learn the progression of DR disease and that introducing $t_{\texttt{mix}}$ augmentation is beneficial for time-aware models.

Current investigations in the framework of the PhD thesis of R. Zeghlache deals with longitudinal representation learning in continuous-time models, especially by employing NODEs in extended self-supervised learning SimCLR [54] and BYOL [153] paradigms or by making use of longitudinal masked auto-encoder with time-aware position embedding and disease progression-aware masking.

Further investigations are needed to enable the creation of pathological trajectories in latent space by taking into account the full temporal dynamic as well as multi-modal data (e.g. various imaging modalities, clinical variables) to capture the complexity of disease progression comprehensively. Research perspectives in disease progression modeling extend beyond the current clinical focus. I aim to delve into other pathologies such as autosomal dominant polycystic kidney disease (ADPKD), known for its significant clinical variability regarding the rate of renal function decline. By leveraging such complexity into multi-modal modeling frameworks, we could enhance our understanding of disease progression dynamics and potentially pave the way for more effective interventions and treatments.

4.7 Conclusion

In conclusion, the challenges and advancements discussed in this chapter represent significant strides towards enhancing decision support in surgery planning and therapeutic follow-up. By leveraging various types of imaging data (e.g. external, unlabeled, imperfectly-annotated), we raised the potential of fully-automated patient-specific thorough cartography to improve the effectiveness of pre-operative planning. Uncertainty modeling is another important path to study as it may improve the learning process and provide clinicians with locally-estimated confidence information.

Moving forward, continued research and innovation are essential for addressing the evolving challenges of therapeutic follow-up. As patient care continues to advance, there is an opportunity to further refine deep learning methodologies to reach complete treatment regimen recommendation systems, leading to improved patient outcomes. In this direction, developing stronger artificial intelligence methodologies able to integrate longitudinal data from various modalities while providing information that can be understood by clinicians and complying with confidentiality standards remains key.

Beyond the technical aspects, it is essential to engage in a wider reflection aimed at defining the place to be given to the machine in clinical routine. Two directions are emerging: one aimed at facilitating the day-to-day work of radiologists, the other involving the automatic sorting of images by level of interpretation complexity. Whatever the scenario, the design of novel evaluation metrics reflecting the clinical applicability of artificial intelligence algorithms is a prospect to be developed.

BIBLIOGRAPHY

[1] A. Saporta, T.-H. Vu, M. Cord, and P. Pérez, « Multi-target adversarial frameworks for domain adaptation in semantic segmentation », *in IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9072–9081.

[2] T. Heimann and H.-P. Meinzer, « Statistical shape models for 3D medical image segmentation: a review », *Medical Image Analysis*, vol. 13, *4*, pp. 543–563, 2009.

[3] S. Kim, D. Lee, S. Park, K.-S. Oh, S. W. Chung, and Y. Kim, « Automatic segmentation of supraspinatus from MRI by internal shape fitting and autocorrection », *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 165–174, 2017.

[4] J. E. Iglesias and M. R. Sabuncu, « Multi-atlas segmentation of biomedical images: A survey », *Medical Image Analysis*, vol. 24, *1*, pp. 205–219, 2015.

[5] X. Chen and L. Pan, « A survey of graph cuts/graph search based medical image segmentation », *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 112–124, 2018.

[6] V. Barra and J.-Y. Boire, « Segmentation of fat and muscle from MR images of the thigh by a possibilistic clustering algorithm », *Computer Methods and Programs in Biomedicine*, vol. 68, *3*, pp. 185–193, 2002.

[7] S.-Y. Wan and W. E. Higgins, « Symmetric region growing », *IEEE Transactions on Image Processing*, vol. 12, *9*, pp. 1007–1015, 2003.

[8] S. Zhou, J. Wang, S. Zhang, Y. Liang, and Y. Gong, « Active contour model based on local and global intensity information for medical image segmentation », *Neurocomputing*, vol. 186, pp. 107–118, 2016.

[9] E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, *et al.*, « New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1) », *European Journal of Cancer*, vol. 45, *2*, pp. 228–247, 2009.

[10] M. Ronot, M. Bouattour, J. Wassermann, O. Bruno, C. Dreyer, B. Larroque, L. Castera, V. Vilgrain, J. Belghiti, E. Raymond, *et al.*, « Alternative response criteria versus RECIST 1.1 in patients with advanced hepatocellular carcinoma treated with Sorafenib », *The Oncologist*, vol. 19, *4*, pp. 394–402, 2014.

[11] M.-H. Ho, C.-Y. Yu, K.-P. Chung, T.-W. Chen, H.-C. Chu, C.-K. Lin, and C.-B. Hsieh, « Locoregional therapy-induced tumor necrosis as a predictor of recurrence after liver transplant in patients with hepatocellular carcinoma », *Annals of Surgical Oncology*, vol. 18, pp. 3632–3639, 2011.

[12] J. H. Shim, K. M. Kim, Y.-J. Lee, G.-Y. Ko, H.-K. Yoon, K.-B. Sung, K.-M. Park, S.-G. Lee, Y.-S. Lim, H. C. Lee, *et al.*, « Complete necrosis after transarterial chemoembolization could predict prolonged survival in patients with recurrent intrahepatic hepatocellular carcinoma after curative resection », *Annals of Surgical Oncology*, vol. 17, pp. 869–877, 2010.

[13] L. Breiman, « Random forests », *Machine Learning*, vol. 45, pp. 5–32, 2001.

[14] E. Geremia, O. Clatz, B. H. Menze, E. Konukoglu, A. Criminisi, and N. Ayache, « Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images », *NeuroImage*, vol. 57, *2*, pp. 378–390, 2011.

[15] Z. Tu and X. Bai, « Auto-context and its application to high-level vision tasks and 3D brain image segmentation », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, *10*, pp. 1744–1757, 2009.

[16] W. Wu, A. Y. Chen, L. Zhao, and J. J. Corso, « Brain tumor detection and segmentation in a CRF (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features », *International Journal of Computer Assisted Radiology and Surgery*, vol. 9, pp. 241–253, 2014.

[17] B. Irving, J. M. Franklin, B. W. Papież, E. M. Anderson, R. A. Sharma, F. V. Gleeson, M. Brady, and J. A. Schnabel, « Pieces-of-parts for supervoxel segmentation with global context: application to DCE-MRI tumour delineation », *Medical Image Analysis*, vol. 32, pp. 69–83, 2016.

[18] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, « SLIC superpixels compared to state-of-the-art superpixel methods », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, *11*, pp. 2274–2282, 2012.

[19] P.-H. Conze, V. Noblet, F. Rousseau, F. Heitz, V. de Blasi, R. Memeo, and P. Pessaux, « Scale-adaptive supervoxel-based random forests for liver tumor segmentation in dynamic contrast-enhanced CT scans », *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, *2*, pp. 223–233, 2017.

[20] P.-H. Conze, F. Rousseau, V. Noblet, F. Heitz, R. Memeo, and P. Pessaux, « Semi-automatic liver tumor segmentation in DCE-CT scans using random forests and supervoxels », *in International Workshop on Machine Learning in Medical Imaging*, 2015.

[21] P.-H. Conze, V. Noblet, F. Rousseau, F. Heitz, R. Memeo, and P. Pessaux, « Random forests on hierarchical multi-scale supervoxels for liver tumor segmentation in dynamic contrast-enhanced CT scans », *in IEEE International Symposium on Biomedical Imaging*, 2016, pp. 416–419.

[22] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, « A survey on deep learning in medical image analysis », *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[23] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel, « Loss odyssey in medical image segmentation », *Medical Image Analysis*, vol. 71, p. 102 035, 2021.

[24] D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber, « Deep neural networks segment neuronal membranes in electron microscopy images », *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[25] J. Long, E. Shelhamer, and T. Darrell, « Fully convolutional networks for semantic segmentation », *in IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[26] O. Ronneberger, P. Fischer, and T. Brox, « U-Net: convolutional networks for biomedical image segmentation », *in International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[27] F. Milletari, N. Navab, and S.-A. Ahmadi, « V-Net: fully convolutional neural networks for volumetric medical image segmentation », *in International Conference on 3D Vision*, 2016, pp. 565–571.

[28] R. M. Summers, « Progress in fully automated abdominal CT interpretation », *American Journal of Roentgenology*, vol. 207, *1*, p. 67, 2016.

[29] A. E. Kavur, N. S. Gezer, M. Barış, S. Aslan, P.-H. Conze, V. Groza, D. D. Pham, S. Chatterjee, P. Ernst, S. Özkan, *et al.*, « CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation », *Medical Image Analysis*, vol. 69, p. 101 950, 2021.

[30] J. Ma, [...], P.-H. Conze, Z. Huang, Z. Zhout, D.-P. Fan, H. Xiong, G. Dong, Q. Zhu, J. He, and X. Yang, « Fast and low-GPU-memory abdomen CT organ segmentation: the FLARE challenge », *Medical Image Analysis*, vol. 82, p. 102 616, 2022.

[31] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, « Image-to-image translation with conditional adversarial networks », *in IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134.

[32] P.-H. Conze, A. E. Kavur, E. Cornec-Le Gall, N. S. Gezer, Y. Le Meur, M. A. Selver, and F. Rousseau, « Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks », *Artificial Intelligence in Medicine*, vol. 117, p. 102 109, 2021.

[33] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, « ImageNet large scale visual recognition challenge », *International Journal of Computer Vision*, vol. 115, pp. 211–252, 2015.

[34] P.-H. Conze, S. Brochard, V. Burdin, F. T. Sheehan, and C. Pons, « Healthy versus pathological learning transferability in shoulder muscle MRI segmentation using deep convolutional encoder-decoders », *Computerized Medical Imaging and Graphics*, vol. 83, p. 101 733, 2020.

[35] K. Simonyan and A. Zisserman, « Very deep convolutional networks for large-scale image recognition », *arXiv preprint 1409.1556*, 2014.

[36] S. Xie and Z. Tu, « Holistically-nested edge detection », *in IEEE International Conference on Computer Vision*, 2015, pp. 1395–1403.

[37] H. Dou, D. Karimi, C. K. Rollins, C. M. Ortinau, L. Vasung, C. Velasco-Annis, A. Ouaalam, X. Yang, D. Ni, and A. Gholipour, « A deep attentive convolutional neural network for automatic cortical plate segmentation in fetal MRI », *IEEE Transactions on Medical Imaging*, vol. 40, *4*, pp. 1123–1133, 2020.

[38] M. S. Nosrati and G. Hamarneh, « Incorporating prior knowledge in medical image segmentation: A survey », *arXiv preprint 1607.01092*, 2016.

[39] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. De Marvao, T. Dawes, D. P. O'Regan, *et al.*, « Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation », *IEEE Transactions on Medical Imaging*, vol. 37, *2*, pp. 384–395, 2017.

[40] D. Keshwani, Y. Kitamura, S. Ihara, S. Iizuka, and E. Simo-Serra, « TopNet: topology preserving metric learning for vessel tree reconstruction and labelling », *in Medical Image Computing and Computer Assisted Intervention*, 2020, pp. 14–23.

[41] P.-A. Ganaye, M. Sdika, B. Triggs, and H. Benoit-Cattin, « Removing segmentation inconsistencies with semi-supervised non-adjacency constraint », *Medical Image Analysis*, vol. 58, p. 101 551, 2019.

[42] A. Boutillon, B. Borotikar, V. Burdin, and P.-H. Conze, « Multi-structure bone segmentation in pediatric MR images with combined regularization from shape priors and adversarial network », *Artificial Intelligence in Medicine*, vol. 132, p. 102 364, 2022.

[43] A. Boutillon, B. Borotikar, V. Burdin, and P.-H. Conze, « Combining shape priors with conditional adversarial networks for improved scapula segmentation in MR images », *in IEEE International Symposium on Biomedical Imaging*, 2020, pp. 1164–1167.

[44] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, *et al.*, « Attention U-Net: learning where to look for the pancreas », *arXiv preprint 1804.03999*, 2018.

[45] A. Boutillon, B. Borotikar, C. Pons, V. Burdin, and P.-H. Conze, « Multi-structure deep segmentation with shape priors and latent adversarial regularization », *in IEEE International Symposium on Biomedical Imaging*, 2021, pp. 999–1002.

[46] A. Sadikine, B. Badic, J.-P. Tasu, V. Noblet, P. Ballet, D. Visvikis, and P.-H. Conze, « Improving abdominal image segmentation with overcomplete shape priors », *Computerized Medical Imaging and Graphics*, 2024.

[47] L. Soler, A. Hostettler, V. Agnus, A. Charnoz, J.-B. Fasquel, J. Moreau, A.-B. Osswald, M. Bouhadjar, and J. Marescaux, « 3D image reconstruction for comparison of algorithm database: a patient specific anatomical and medical image database », *IRCAD Technical Report*, 2010.

[48] J. M. J. Valanarasu, V. A. Sindagi, I. Hacihaliloglu, and V. M. Patel, « KiU-Net: overcomplete convolutional architectures for biomedical image and volumetric segmentation », *IEEE Transactions on Medical Imaging*, vol. 41, *4*, pp. 965–976, 2021.

[49] A. Sadikine, B. Badic, E. Ferrante, V. Noblet, P. Ballet, D. Visvikis, and P.-H. Conze, « Deep vessel segmentation with joint multi-prior encoding », *in IEEE International Symposium on Biomedical Imaging*, 2024.

[50] S. Shit, J. C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, A. Zhylka, J. P. Pluim, U. Bauer, and B. H. Menze, « ClDice – A novel topology-preserving loss function for tubular structure segmentation », *in IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 16 560–16 569.

[51] A. Sadikine, B. Badic, J.-P. Tasu, V. Noblet, P. Ballet, D. Visvikis, and P.-H. Conze, « Scale-specific auxiliary multi-task contrastive learning for deep liver vessel segmentation », *in IEEE International Symposium on Biomedical Imaging*, 2023.

[52] P. Jassi and G. Hamarneh, « VascuSynth: vascular tree synthesis software », *Insight Journal*, 2011.

[53] S. Vandenhende, S. Georgoulis, W. Van Gansbeke, M. Proesmans, D. Dai, and L. Van Gool, « Multi-task learning for dense prediction tasks: A survey », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[54] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, « A simple framework for contrastive learning of visual representations », *in International Conference on Machine Learning*, 2020.

[55] E. Cornec-Le Gall, A. Alam, and R. D. Perrone, « Autosomal dominant polycystic kidney disease », *The Lancet*, vol. 393, *10174*, pp. 919–935, 2019.

[56] E. Higashihara, K. Nutahara, T. Okegawa, T. Shishido, M. Tanbo, K. Kobayasi, and T. Nitadori, « Kidney volume and function in autosomal dominant polycystic kidney disease », *Clinical and Experimental Nephrology*, vol. 18, pp. 157–165, 2014.

[57] M. D. van Gastel, M. E. Edwards, V. E. Torres, B. J. Erickson, R. T. Gansevoort, and T. L. Kline, « Automatic measurement of kidney and liver volumes from MR images of patients affected by autosomal dominant polycystic kidney disease », *Journal of the American Society of Nephrology*, 2019.

[58] P.-H. Conze, G. Andrade-Miranda, Y. Le Meur, E. Cornec-Le Gall, and F. Rousseau, « Dual-task kidney MR segmentation with transformers in autosomal-dominant polycystic kidney disease », *Computerized Medical Imaging and Graphics*, 2024.

[59] T. L. Kline, P. Korfiatis, M. E. Edwards, J. D. Blais, F. S. Czerwiec, P. C. Harris, B. F. King, V. E. Torres, and B. J. Erickson, « Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys », *Journal of Digital Imaging*, vol. 30, *4*, pp. 442–448, 2017.

[60] F. Shamshad, S. Khan, S. W. Zamir, M. H. Khan, M. Hayat, F. S. Khan, and H. Fu, « Transformers in medical imaging: A survey », *Medical Image Analysis*, p. 102 802, 2023.

[61] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, *et al.*, « Swin Transformer v2: scaling up capacity and resolution », *in IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12 009–12 019.

[62] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, « TransuNet: Transformers make strong encoders for medical image segmentation », *arXiv preprint 2102.04306*, 2021.

[63] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, « Medical Transformer: gated axial-attention for medical image segmentation », *in International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2021, pp. 36–46.

[64] R. Strudel, R. Garcia, I. Laptev, and C. Schmid, « Segmenter: Transformer for semantic segmentation », *in IEEE/CVF International Conference on Computer Vision*, 2021, pp. 7262–7272.

[65] N. Decaux, P.-H. Conze, J. Ropars, X. He, F. T. Sheehan, C. Pons, D. Ben Salem, S. Brochard, and F. Rousseau, « Semi-automatic muscle segmentation in MR images using deep registration-based label propagation », *Pattern Recognition*, vol. 140, p. 109 529, 2023.

[66] J. Sun, Y. Mao, Y. Dai, Y. Zhong, and J. Wang, « MUNet: Motion uncertainty-aware semi-supervised video object segmentation », *Pattern Recognition*, vol. 138, p. 109 399, 2023.

[67] R. Feng, X. Zheng, T. Gao, J. Chen, W. Wang, D. Z. Chen, and J. Wu, « Interactive few-shot learning: limited supervision, better medical image segmentation », *IEEE Transactions on Medical Imaging*, vol. 40, *10*, pp. 2575–2588, 2021.

[68] C. Pons, F. T. Sheehan, H. S. Im, S. Brochard, and K. E. Alter, « Shoulder muscle atrophy and its relation to strength loss in obstetrical brachial plexus palsy », *Clinical Biomechanics*, vol. 48, pp. 80–87, 2017.

[69] A. B. Albu, T. Beugeling, and D. Laurendeau, « A morphology-based approach for interslice interpolation of anatomical slices from volumetric images », *IEEE Transactions on Biomedical Engineering*, vol. 55, *8*, pp. 2022–2038, 2008.

[70] A. Jabri, A. Owens, and A. Efros, « Space-time correspondence as a contrastive random walk », *Advances in Neural Information Processing Systems*, vol. 33, pp. 19 545–19 560, 2020.

[71] P.-H. Conze, G. Andrade-Miranda, V. K. Singh, V. Jaouen, and D. Visvikis, « Current and emerging trends in medical image segmentation with deep learning », *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 7, *6*, pp. 545–569, 2023.

[72] M. Abbas, G. Adrade-Miranda, V. Bourbonne, D. Visvikis, B. Badic, and P.-H. Conze, « Learning transferability in deep segmentation of liver metastases », *in IEEE International Symposium on Biomedical Imaging*, 2024.

[73] R. Zheng, L. Wang, C. Wang, X. Yu, W. Chen, Y. Li, W. Li, F. Yan, H. Wang, and R. Li, « Feasibility of automatic detection of small hepatocellular carcinoma ($\leq$ 2 cm) in cirrhotic liver based on pattern matching and deep learning », *Physics in Medicine & Biology*, vol. 66, *8*, p. 085 014, 2021.

[74] P. Bilic, P. Christ, H. B. Li, E. Vorontsov, A. Ben-Cohen, G. Kaissis, A. Szeskin, C. Jacobs, G. E. H. Mamani, G. Chartrand, *et al.*, « The liver tumor segmentation benchmark (LiTS) », *Medical Image Analysis*, vol. 84, p. 102 680, 2023.

[75] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, « Swin UNETR: Swin Transformers for semantic segmentation of brain tumors in MRI images », *in International MICCAI Brainlesion Workshop*, 2021, pp. 272–284.

[76] Y. Tang, D. Yang, W. Li, H. R. Roth, B. Landman, D. Xu, V. Nath, and A. Hatamizadeh, « Self-supervised pre-training of Swin Transformers for 3D medical image analysis », *in IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 20 730–20 740.

[77] H. Messaoudi, A. Belaid, D. Ben Salem, and P.-H. Conze, « Cross-dimensional transfer learning in medical image segmentation with deep learning », *Medical Image Analysis*, p. 102 868, 2023.

[78] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, « The multimodal brain tumor image segmentation benchmark (BRATS) », *IEEE Transactions on Medical Imaging*, vol. 34, *10*, pp. 1993–2024, 2014.

[79] S. Leclerc, E. Smistad, J. Pedrosa, A. Østvik, F. Cervenansky, F. Espinosa, T. Espeland, E. A. R. Berg, P.-M. Jodoin, T. Grenier, *et al.*, « Deep learning for segmentation using an open large-scale dataset in 2D echocardiography », *IEEE Transactions on Medical Imaging*, vol. 38, *9*, pp. 2198–2210, 2019.

[80] Y. Yan, P.-H. Conze, M. Lamard, G. Quellec, B. Cochener, and G. Coatrieux, « Towards improved breast mass detection using dual-view mammogram matching », *Medical Image Analysis*, vol. 71, p. 102 083, 2021.

[81] N. Dhungel, G. Carneiro, and A. P. Bradley, « A deep learning approach for the analysis of masses in mammograms with minimal user intervention », *Medical Image Analysis*, vol. 37, pp. 114–128, 2017.

[82] J. Redmon and A. Farhadi, « Yolov3: an incremental improvement », *arXiv preprint 1804.02767*, 2018.

[83] Y. Yan, P.-H. Conze, M. Lamard, H. Zhang, G. Quellec, B. Cochener, and G. Coatrieux, « Deep active learning for dual-view mammogram analysis », *in International Workshop on Machine Learning in Medical Imaging*, 2021, pp. 180–189.

[84] A. Boutillon, P.-H. Conze, C. Pons, V. Burdin, and B. Borotikar, « Generalizable multi-task, multi-domain deep segmentation of sparse pediatric imaging datasets via multi-scale contrastive regularization and multi-joint anatomical priors », *Medical Image Analysis*, vol. 81, 2022.

[85] N. Karani, K. Chaitanya, C. Baumgartner, and E. Konukoglu, « A lifelong learning approach to brain MR segmentation across scanners and protocols », *in International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 476–484.

[86] V. V. Valindria, N. Pawlowski, M. Rajchl, I. Lavdas, E. O. Aboagye, A. G. Rockall, D. Rueckert, and B. Glocker, « Multi-modal learning from unpaired images: application to multi-organ segmentation in CT and MRI », *in IEEE Winter Conference on Applications of Computer Vision*, IEEE, 2018, pp. 547–556.

[87] Q. Dou, Q. Liu, P. A. Heng, and B. Glocker, « Unpaired multi-modal segmentation via knowledge distillation », *IEEE Transactions on Medical Imaging*, vol. 39, *7*, pp. 2415–2425, 2020.

[88] D. Fourure, R. Emonet, E. Fromont, D. Muselet, N. Neverova, A. Trémeau, and C. Wolf, « Multi-task, multi-domain learning: application to semantic segmentation and pose regression », *Neurocomputing*, vol. 251, pp. 68–80, 2017.

[89] S.-A. Rebuffi, H. Bilen, and A. Vedaldi, « Learning multiple visual domains with residual adapters », *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[90] P. Moeskops, J. M. Wolterink, B. H. Van Der Velden, K. G. Gilhuijs, T. Leiner, M. A. Viergever, and I. Išgum, « Deep learning for multi-task medical image segmentation in multiple modalities », *in Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 478–486.

[91] G. Andrade-Miranda, V. Jaouen, O. Tankyevych, C. C. Le Rest, D. Visvikis, and P.-H. Conze, « Multi-modal medical Transformers: A meta-analysis for medical image segmentation in oncology », *Computerized Medical Imaging and Graphics*, vol. 110, p. 102 308, 2023.

[92] Y. Zhang, J. Yang, J. Tian, Z. Shi, C. Zhong, Y. Zhang, and Z. He, « Modality-aware mutual learning for multi-modal medical image segmentation », *in Medical Image Computing and Computer Assisted Intervention*, 2021, pp. 589–599.

[93] D. Zhang, G. Huang, Q. Zhang, J. Han, J. Han, and Y. Yu, « Cross-modality deep feature learning for brain tumor segmentation », *Pattern Recognition*, vol. 110, p. 107 562, 2021.

[94] Y. Zhang, N. He, J. Yang, Y. Li, D. Wei, Y. Huang, Y. Zhang, Z. He, and Y. Zheng, « mmFormer: multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation », *in International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2022, pp. 107–117.

[95] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, « Unetr: transformers for 3D medical image segmentation », *in IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 574–584.

[96] S. Roy, G. Koehler, M. Baumgartner, C. Ulrich, J. Petersen, F. Isensee, and K. Maier-Hein, « Transformer utilization in medical image segmentation networks », *arXiv preprint 2304.04225*, 2023.

[97] A. W. C. Kow, « Hepatic metastasis from colorectal cancer », *Journal of Gastrointestinal Oncology*, vol. 10, *6*, p. 1274, 2019.

[98] M.-A. Lebre, A. Vacavant, M. Grand-Brochier, H. Rositi, A. Abergel, P. Chabrot, and B. Magnin, « Automatic segmentation methods for liver and hepatic vessels from CT and MRI volumes, applied to the Couinaud scheme », *Computers in Biology and Medicine*, vol. 110, pp. 42–51, 2019.

[99] N. Heller, F. Isensee, K. H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han, *et al.*, « The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: results of the KiTS19 challenge », *Medical Image Analysis*, vol. 67, p. 101 821, 2021.

[100] P. Rougé, P.-H. Conze, N. Passat, and O. Merveille, « Guidelines for cerebrovascular segmentation: managing imperfect annotations in the context of semi-supervised learning », *arXiv preprint arXiv:2404.01765*, 2024.

[101] S. Zhang and D. Metaxas, « On the challenges and perspectives of foundation models for medical image analysis », *Medical Image Analysis*, vol. 91, p. 102 996, 2024.

[102] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, *et al.*, « Segment anything », *in IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4015–4026.

[103] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, « Segment anything in medical images », *Nature Communications*, vol. 15, *1*, p. 654, 2024.

[104] Y. Zhou, M. A. Chia, S. K. Wagner, M. S. Ayhan, D. J. Williamson, R. R. Struyven, T. Liu, M. Xu, M. G. Lozano, P. Woodward-Court, *et al.*, « A foundation model for generalizable disease detection from retinal images », *Nature*, vol. 622, *7981*, pp. 156–163, 2023.

[105] R. Jiao, Y. Zhang, L. Ding, B. Xue, J. Zhang, R. Cai, and C. Jin, « Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation », *Computers in Biology and Medicine*, p. 107 840, 2023.

[106] Y. Shi, J. Zhang, T. Ling, J. Lu, Y. Zheng, Q. Yu, L. Qi, and Y. Gao, « Inconsistency-aware uncertainty estimation for semi-supervised medical image segmentation », *IEEE Transactions on Medical Imaging*, vol. 41, *3*, pp. 608–620, 2021.

[107] A. Tarvainen and H. Valpola, « Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results », *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[108] C. You, Y. Zhou, R. Zhao, L. Staib, and J. S. Duncan, « SimCVD: simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation », *IEEE Transactions on Medical Imaging*, vol. 41, *9*, pp. 2228–2237, 2022.

[109] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, *et al.*, « A review of uncertainty quantification in deep learning: Techniques, applications and challenges », *Information Fusion*, vol. 76, pp. 243–297, 2021.

[110] K. Wickstrøm, M. Kampffmeyer, and R. Jenssen, « Uncertainty and interpretability in convolutional neural networks for semantic segmentation of colorectal polyps », *Medical Image Analysis*, vol. 60, p. 101 619, 2020.

[111] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, and T. Vercauteren, « Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks », *Neurocomputing*, vol. 338, pp. 34–45, 2019.

[112] Y. van Gestel, I. de Hingh, M. van Herk-Sukel, F. van Erning, L. Beerepoot, J. Wijsman, *et al.*, « Patterns of metachronous metastases after curative treatment of colorectal cancer », *Cancer Epidemiology*, vol. 38, *4*, pp. 448–454, 2014.

[113] A. L. Simpson, D. A. Geller, A. W. Hemming, W. R. Jarnagin, L. W. Clements, M. I. D'Angelica, P. Dumpuri, M. Gönen, I. Zendejas, M. I. Miga, *et al.*, « Liver planning software accurately predicts postoperative liver volume and measures early regeneration », *Journal of the American College of Surgeons*, vol. 219, *2*, pp. 199–207, 2014.

[114] M. Schindl, D. Redhead, K. Fearon, O. Garden, and S. Wigmore, « The value of residual liver volume as a predictor of hepatic dysfunction and infection after major liver resection », *GUT*, vol. 54, *2*, pp. 289–296, 2005.

[115] B. Christ, U. Dahmen, K.-H. Herrmann, M. König, J. R. Reichenbach, L. Ole Schwen, and N. Waschinsky, « Computational modeling in liver surgery », *Frontiers in Physiology*, vol. 8, p. 307 495, 2017.

[116] D. C. Le, J. Chansangrat, N. Keeratibharat, and P. Horkaew, « Functional segmentation for preoperative liver resection based on hepatic vascular networks », *IEEE Access*, vol. 9, pp. 15 485–15 498, 2021.

[117] C. K. Kuhl, Y. Alparslan, J. Schmoee, B. Sequeira, A. Keulers, T. H. Brümmendorf, and S. Keil, « Validity of RECIST version 1.1 for response assessment in metastatic cancer: a prospective, multireader study », *Radiology*, vol. 290, *2*, pp. 349–356, 2019.

[118] A. Dohan, B. Gallix, B. Guiu, K. Le Malicot, C. Reinhold, P. Soyer, J. Bennouna, F. Ghiringhelli, E. Barbier, V. Boige, *et al.*, « Early evaluation using a radiomic signature of unresectable hepatic metastases to predict outcome in patients with colorectal cancer treated with FOLFIRI and Bevacizumab », *GUT*, vol. 69, *3*, pp. 531–539, 2020.

[119] M. M. Islam, B. Badic, T. Aparicio, D. Tougeron, J.-P. Tasu, D. Visvikis, and P.-H. Conze, « Deep treatment response assessment and prediction of colorectal cancer liver metastases », *in International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2022, pp. 482–491.

[120] A. Maaref, F. P. Romero, E. Montagnon, M. Cerny, B. Nguyen, F. Vandenbroucke, G. Soucy, S. Turcotte, A. Tang, and S. Kadoury, « Predicting the response to FOLFOX-based chemotherapy regimen from untreated liver metastases on baseline CT: a deep neural network approach », *Journal of Digital Imaging*, pp. 1–9, 2020.

[121] H.-B. Zhu, D. Xu, M. Ye, L. Sun, X.-Y. Zhang, X.-T. Li, P. Nie, B.-C. Xing, and Y.-S. Sun, « Deep learning-assisted magnetic resonance imaging prediction of tumor response to chemotherapy in patients with colorectal liver metastases », *International Journal of Cancer*, vol. 148, *7*, pp. 1717–1730, 2021.

[122] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, « Aggregated residual transformations for deep neural networks », *in IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1492–1500.

[123] T. Aparicio, O. Bouché, J. Taieb, E. Maillard, S. Kirscher, P.-L. Etienne, R. Faroux, *et al.*, « Bevacizumab+chemotherapy versus chemotherapy alone in elderly patients with untreated metastatic colorectal cancer: a randomized phase II trial—PRODIGE 20 study results », *Annals of Oncology*, vol. 29, *1*, pp. 133–138, 2018.

[124] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, « GradCAM: visual explanations from deep networks via gradient-based localization », *in IEEE International Conference on Computer Vision*, 2017, pp. 618–626.

[125] C. Jin, H. Yu, J. Ke, P. Ding, Y. Yi, X. Jiang, X. Duan, J. Tang, D. T. Chang, X. Wu, *et al.*, « Predicting treatment response from longitudinal images using multi-task deep learning », *Nature Communications*, vol. 12, *1*, p. 1851, 2021.

[126] T. Aparicio, B. Linot, K. Le Malicot, O. Bouché, V. Boige, E. François, F. Ghiringhelli, J.-L. Legoux, M. B. Abdelghani, J.-M. Phelip, *et al.*, « FOLFIRI+Bevacizumab induction chemotherapy followed by Bevacizumab or observation in metastatic colorectal cancer, a phase III trial », *Digestive and Liver Disease*, vol. 47, *4*, pp. 271–272, 2015.

[127] H. Yu, T. A. Boyle, C. Zhou, D. L. Rimm, and F. R. Hirsch, « PD-L1 expression in lung cancer », *Journal of Thoracic Oncology*, vol. 11, *7*, pp. 964–975, 2016.

[128] S. Rossi, L. Toschi, A. Castello, F. Grizzi, L. Mansi, and E. Lopci, « Clinical characteristics of patient selection and imaging predictors of outcome in solid tumors treated with checkpoint-inhibitors », *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 44, pp. 2310–2325, 2017.

[129] R. Sun, E. J. Limkin, M. Vakalopoulou, L. Dercle, S. Champiat, S. R. Han, L. Verlingue, D. Brandao, A. Lancia, S. Ammari, *et al.*, « A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study », *The Lancet Oncology*, vol. 19, *9*, pp. 1180–1191, 2018.

[130] X. Gu, X. Yu, G. Shi, Y. Li, and L. Yang, « Can pd-l1 expression be predicted by contrast-enhanced ct in patients with gastric adenocarcinoma? a preliminary retrospective study », *Abdominal Radiology*, vol. 48, *1*, pp. 220–228, 2023.

[131] A. Chaddad, P. Daniel, M. Zhang, S. Rathore, P. Sargos, C. Desrosiers, and T. Niazi, « Deep radiomic signature with immune cell markers predicts the survival of glioma patients », *Neurocomputing*, vol. 469, pp. 366–375, 2022.

[132] R. Da-Ano, G. Andrade-Miranda, O. Tankyevych, D. Visvikis, P.-H. Conze, and C. Cheze Le Rest, « Automated PD-L1 status prediction in lung cancer with multi-modal PET/CT fusion », *Scientific Reports*, 2024.

[133] W. Mu, L. Jiang, Y. Shi, I. Tunali, J. E. Gray, E. Katsoulakis, J. Tian, R. J. Gillies, and M. B. Schabath, « Non-invasive measurement of PD-L1 status and prediction of immunotherapy response using deep learning of PET/CT images », *Journal for Immunotherapy of Cancer*, vol. 9, *6*, 2021.

[134] C. Wang, J. Ma, J. Shao, S. Zhang, J. Li, J. Yan, Z. Zhao, C. Bai, Y. Yu, and W. Li, « Non-invasive measurement using deep learning algorithm based on multi-source features fusion to predict PD-L1 expression and survival in NSCLC », *Frontiers in Immunology*, vol. 13, p. 828 560, 2022.

[135] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, *et al.*, « The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans », *Medical Physics*, vol. 38, *2*, pp. 915–931, 2011.

[136] R. Santeramo, S. Withey, and G. Montana, « Longitudinal detection of radiological abnormalities with time-modulated lstm », *in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018, pp. 326–333.

[137] S. Perek, L. Ness, M. Amit, E. Barkan, and G. Amit, « Learning from longitudinal mammography studies », *in International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 712–720.

[138] Y. Yan, P.-H. Conze, G. Quellec, P. Massin, M. Lamard, G. Coatrieux, and B. Cochener, « Longitudinal detection of diabetic retinopathy early severity grade changes using deep learning », *in International Workshop on Ophthalmic Medical Image Analysis*, 2021, pp. 11–20.

[139] K. Ogurtsova, J. da Rocha Fernandes, Y Huang, U. Linnenkamp, L Guariguata, N. H. Cho, D. Cavan, J. Shaw, and L. Makaroff, « IDF diabetes atlas: global estimates for the prevalence of diabetes for 2015 and 2040 », *Diabetes Research and Clinical Practice*, vol. 128, pp. 40–50, 2017.

[140] A. Bora, S. Balasubramanian, B. Babenko, S. Virmani, S. Venugopalan, A. Mitani, G. de Oliveira Marinho, J. Cuadros, P. Ruamviboonsuk, G. S. Corrado, *et al.*, « Predicting the risk of developing diabetic retinopathy using deep learning », *The Lancet Digital Health*, vol. 3, *1*, 2021.

[141] P Massin, A Chabouis, A Erginay, C Viens-Bitker, A Lecleire-Collet, T Meas, P.-J. Guillausseau, G Choupot, B André, and P Denormandie, « OPHDIAT: a telemedical network screening system for diabetic retinopathy in the île-de-france », *Diabetes & Metabolism*, vol. 34, *3*, 2008.

[142] S. Albelwi, « Survey on self-supervised learning: auxiliary pretext tasks and contrastive learning methods in imaging », *Entropy*, vol. 24, *4*, p. 551, 2022.

[143] A. Rivail, U. Schmidt-Erfurth, W.-D. Vogl, S. M. Waldstein, S. Riedl, C. Grechenig, Z. Wu, and H. Bogunovic, « Modeling disease progression in retinal OCTs with longitudinal self-supervised learning », *in Predictive Intelligence in Medicine*, 2019, pp. 44–52.

[144] Q. Zhao, Z. Liu, E. Adeli, and K. M. Pohl, « Longitudinal self-supervised learning », *Medical Image Analysis*, vol. 71, p. 102 051, 2021.

[145] R. Zeghlache, P.-H. Conze, M. E. H. Daho, R. Tadayoni, P. Massin, B. Cochener, G. Quellec, and M. Lamard, « Detection of diabetic retinopathy using longitudinal self-supervised learning », *in International Workshop on Ophthalmic Medical Image Analysis*, 2022, pp. 43–52.

[146] R. Zeghlache, P.-H. Conze, M. E. H. Daho, Y. Li, H. Le Boité, R. Tadayoni, P. Massin, B. Cochener, I. Brahim, G. Quellec, and M. Lamard, « Longitudinal self-supervised learning using neural ordinary differential equation », *in International Workshop on Predictive Intelligence in Medicine*, 2023.

[147] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, « Neural ordinary differential equations », *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[148] Y. Rubanova, R. T. Chen, and D. K. Duvenaud, « Latent ordinary differential equations for irregularly-sampled time series », *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[149] R. Zeglache, P.-H. Conze, M. El Habib Daho, L. Yihao, H. Le Boité, P. Massin, R. Tadayoni, B. Cochener, I. Brahim, G. Quellec, and M. Lamard, « LMT: longitudinal mixing training a framework for the prediction of disease progression using a single image », *in International Workshop on Machine Learning in Medical Imaging*, 2023.

[150] S. Thulasidasan, G. Chennupati, J. A. Bilmes, T. Bhattacharya, and S. Michalak, « On mixup training: improved calibration and predictive uncertainty for deep neural networks », *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[151] V. Verma, A. Lamb, C. Beckham, A. Najafi, I. Mitliagkas, D. Lopez-Paz, and Y. Bengio, « Manifold mixup: better representations by interpolating hidden states », *in International Conference on Machine Learning*, 2019, pp. 6438–6447.

[152] P. Mangla, N. Kumari, A. Sinha, M. Singh, B. Krishnamurthy, and V. N. Balasubramanian, « Charting the right manifold: manifold mixup for few-shot learning », *in IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 2218–2227.

[153] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, *et al.*, « Bootstrap your own latent-a new approach to self-supervised learning », *Advances in Neural Information Processing Systems*, vol. 33, pp. 21 271–21 284, 2020.

CURRICULUM

Academic background

- 2017: "Qualifications aux fonctions de Maître de Conférences" in CNU sections 61 (Computer sciences, Automatic Control and Signal Processing) and 27 (Informatics)
- Dec. 2010 - Apr. 2014: Industrial PhD thesis at Technicolor and INSA de Rennes, IETR UMR 6164, Rennes, France. *Long-term dense motion estimation and view synthesis quality assessment with application to joint stereo and motion processing.* Defended on April 16, 2014.
- 2009 - 2010: Master EEA APS Signal & Image, University Bordeaux I, Bordeaux, France
- 2007 - 2010: ENSEIRB engineering school, specialization in Signal & Image, Bordeaux, France
- 2005 - 2007: 2 year intensive Maths and Physics courses at Lycée Chateaubriand, Rennes, France, in preparation for the selective entrance examination to French engineering schools

Professional experience

- From Dec. 2016: Associate Professor in medical image analysis, computer vision and artificial intelligence at IMT Atlantique and LaTIM UMR 1101 Inserm, Brest, France
- Sept. 2015 - Nov. 2016: Post-doctoral researcher at ICube UMR 7357 CNRS, University of Strasbourg, Strasbourg, France. *Multi-modal CT/MR abdominal image registration for patient follow-up.* 3D-Surg project with financial support from BPI France.
- Sept. 2014 - Aug. 2015: Post-doctoral researcher at ICube UMR 7357 CNRS, University of Strasbourg. *Liver tumor segmentation in dynamic contrast-enhanced CT for computer-aided diagnosis.* TheraHCC project with financial support from Fondation ARC.
- Dec. 2010 - Nov. 2013: R&D engineer at Technicolor, Rennes, France on disparity estimation, view synthesis, image quality assessment and long-term dense motion estimation.
- Feb. 2010 - Aug. 2010: M2 intern at Technicolor on color look grabbing for video production.

Collective activities

- From 2024: Member of Advisory Committee of Data Science department at IMT Atlantique
- From 2021: Associate Editor for *IEEE Transactions on Radiation and Plasma Medical Sciences*
- Co-organisation of *Recherche en Imagerie et Technologies pour la Santé* (RITS) 2022 conference
- 2021, 2023: Participation in selection committees for Associate Professor positions
- Regular reviewer for international journals: *Medical Image Analysis*, *IEEE Transactions on Medical Imaging*, *IEEE Transactions on Image Processing*, *IEEE Transactions on Radiation and Plasma Medical Sciences*, *Pattern Recognition*, *Pattern Recognition Letters*, *International*

Journal of Computer Assisted Radiology and Surgery, *Medical Physics*...
- Regular reviewer for international conferences: MICCAI, IEEE ISBI, EMBC, MIC, ICPR...
- Oral session chair at IEEE MIC 2022 and IEEE ISBI 2024
- 2015 - 2016 : Elected member of the board of ICube UMR 7357

Research projects as principal investigator

- 2025 - 2026: Longitudinal follow-up of ADPKD using medical image analysis and AI, granted by Société Francophone de Néphrologie, Dialyse et Transplantation (20k€)
- 2022 - 2025: Therapeutic follow-up of liver metastases from colorectal cancer, granted by Ligue Contre le Cancer (1/2 PhD)
- 2022 - 2024: Multi-modal fusion and weakly supervised learning for prostate cancer segmentation, granted by France Life Imaging, in collaboration with CREATIS (24k€)
- 2021 - 2023: Computational models with cross-modality learning, granted by Région Bretagne (2 years post-doctoral position)
- 2020 - 2022: Deep vessel segmentation with geometric and topological constraints, granted by France Life Imaging, in collaboration with CREATIS (20k€)
- 2019: Convolutional neural networks on superpixel graphs for medical image segmentation, granted by IMT Atlantique (1 year post-doctoral position)
- 2018 - 2020: Supervoxel matching and spectral analysis for non-rigid registration of multi-modal medical images, granted by France Life Imaging, in collaboration with ICube (10,8k€)
- 2017 - 2018: Multi-view mammography processing using superpixel decomposition and deep learning for breast cancer diagnosis, granted by IMT, in collaboration with Mines Paris (10k€)

Participation in research projects

- 2022 - 2026: RaReTiA, a pilot AI project on retinitis pigmentosa, funded by ANR
- 2020 - 2025: RHU EviRed - Intelligent evaluation of diabetic retinopathy, funded by ANR
- 2020 - 2024: Chaire AI4CHILD - AI for pediatric neurorehabilitation, funded by ANR
- 2020 - 2021: Automatic segmentation of metastatic liver, funded by Ligue contre le Cancer
- 2019 - 2022: Intelligent geometrical models for automatic segmentation, funded by IMT
- 2018 - 2022: RHU FollowKnee - Improve follow-up of knee surgery, funded by ANR

Awards

- Outstanding reviewer award at MICCAI 2023
- IEEE TMI distinguished reviewer award (bronze level 2022 - 2023)
- Reviewer awards at MICCAI 2021 and MICCAI 2022 (honorable mention)
- Honorable mention at the international FLARE challenge (MICCAI 2021)
- 3 tasks out of 5 won at the international CHAOS challenge (IEEE ISBI 2019)
- ENSEIRB internship award in 2010 for M2 internship at Technicolor

PUBLICATIONS

The list of publications comprises 1 book chapter, 36 publications in international journals, 42 publications in international conferences and workshops, 7 patents, 11 abstracts in international conferences and additional ressources (publications in national conferences, pre-prints, article for general public, PhD manuscrit).

Book chapter

- S. Matta, M. Lamard, **P.-H. Conze**, A. Le Guilcher, V. Ricquebourg, A.-A. Benyoussef, P. Massin, J.-B. Rottier, B. Cochener, and G. Quellec, « Meta learning for anomaly detection in fundus photographs », *in Meta-Learning with Medical Imaging and Health Informatics Applications*, 2023.

Publications in international journals

- R. Da-Ano, G. Andrade-Miranda, O. Tankyevych, D. Visvikis, **P.-H. Conze**, and C. Cheze Le Rest, « Automated PD-L1 status prediction in lung cancer with multi-modal PET/CT fusion », *Scientific Reports*, 2024.
- Y. Li, M. El Habib Daho, **P.-H. Conze**, R. Zeghlache, H. Le Boité, R. Tadayoni, B. Cochener, M. Lamard, and G. Quellec, « A review of deep learning-based information fusion techniques for multimodal medical image classification », *Computers in Biology and Medicine*, 2024.
- G. Sallé, G. Andrade-Miranda, **P.-H. Conze**, N. Boussion, J. Bert, D. Visvikis, and V. Jaouen, « Cross-modal tumor segmentation using generative blending augmentation and self-training », *IEEE Transactions on Biomedical Engineering*, 2024.
- A. Sadikine, B. Badic, J.-P. Tasu, V. Noblet, P. Ballet, D. Visvikis, and **P.-H. Conze**, « Improving abdominal image segmentation with overcomplete shape priors », *Computerized Medical Imaging and Graphics*, 2024.
- **P.-H. Conze**, G. Andrade-Miranda, Y. Le Meur, E. Cornec-Le Gall, and F. Rousseau, « Dual-task kidney MR segmentation with transformers in autosomal-dominant polycystic kidney disease », *Computerized Medical Imaging and Graphics*, 2024.

- M. El Habib Daho, Y. Li, R. Zeghlache, H. Le Boité, P. Deman, L. Borderie, H. Ren, [...], B. Cochener, A. Couturier, R. Tadayoni, **P.-H. Conze**, M. Lamard, and G. Quellec, « DISCOVER: 2-D multiview summarization of optical coherence tomography angiography for automatic diabetic retinopathy diagnosis », *Artificial Intelligence in Medicine*, 2024.
- N. Ben Chaabane, **P.-H. Conze**, M. Lempereur, G. Quellec, O. Rémy-Néris, S. Brochard, B. Cochener, and M. Lamard, « Quantitative gait analysis and prediction using artificial intelligence for patients with gait disorders », *Scientific Reports*, 2023.
- C. Hognon, **P.-H. Conze**, V. Bourbonne, O. Gallinato, T. Colin, V. Jaouen, and D. Visvikis, « Contrastive image adaptation for acquisition shift reduction in medical imaging », *Artificial Intelligence in Medicine*, 2023.
- Y. Li, M. El Habib Daho, **P.-H. Conze**, R. Zeghlache, H. Le Boité, S. Bonnin, D. Cosette, S. Magazzeni, B. Lay, A. Le Guilcher, R. Tadayoni, B. Cochener, M. Lamard, and G. Quellec, « Hybrid fusion of high-resolution and ultra-widefield OCTA acquisitions for the automatic diagnosis of diabetic retinopathy », *Diagnostics*, 2023.
- G. Andrade-Miranda, V. Jaouen, O. Tankyevych, C. C. Le Rest, D. Visvikis, and **P.-H. Conze**, « Multi-modal medical Transformers: A meta-analysis for medical image segmentation in oncology », *Computerized Medical Imaging and Graphics*, vol. 110, p. 102 308, 2023.
- S. Matta, M. Lamard, **P.-H. Conze**, A. Le Guilcher, C. Lecat, R. Carette, F. Basset, P. Massin, J.-B. Rottier, B. Cochener, and G. Quellec, « Towards population-independent, multi-disease detection in fundus photographs », *Scientific Reports*, 2023.
- H. Messaoudi, A. Belaid, D. Ben Salem, and **P.-H. Conze**, « Cross-dimensional transfer learning in medical image segmentation with deep learning », *Medical Image Analysis*, 2023.
- **P.-H. Conze**, G. Andrade-Miranda, V. K. Singh, V. Jaouen, and D. Visvikis, « Current and emerging trends in medical image segmentation with deep learning », *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 7, *6*, pp. 545–569, 2023.
- N. Decaux, **P.-H. Conze**, J. Ropars, X. He, F. T. Sheehan, C. Pons, D. Ben Salem, S. Brochard, and F. Rousseau, « Semi-automatic muscle segmentation in MR images using deep registration-based label propagation », *Pattern Recognition*, vol. 140, p. 109 529, 2023.
- J. Ma, [...], **P.-H. Conze**, Z. Huang, Z. Zhout, D.-P. Fan, H. Xiong, G. Dong, Q. Zhu, J. He, and X. Yang, « Fast and low-GPU-memory abdomen CT organ segmentation: the FLARE challenge », *Medical Image Analysis*, vol. 82, p. 102 616, 2022.
- A. Boutillon, **P.-H. Conze**, C. Pons, V. Burdin, and B. Borotikar, « Generalizable multi-task, multi-domain deep segmentation of sparse pediatric imaging datasets via multi-scale contrastive regularization and multi-joint anatomical priors », *Medical Image Analysis*, vol. 81, p. 102 556, 2022.

- A. Boutillon, B. Borotikar, V. Burdin, and **P.-H. Conze**, « Multi-structure bone segmentation in pediatric MR images with combined regularization from shape priors and adversarial network », *Artificial Intelligence in Medicine*, vol. 132, p. 102 364, 2022.
- A. Touil, K. Kalti, **P.-H. Conze**, B. Solaiman, and M. Mahjoub, « A new collaborative classification process for microcalcification detection based on graphs and knowledge propagation », *Journal of Digital Imaging*, vol. 35, *6*, pp. 1560–1575, 2022.
- S. Matta, M. Lamard, **P.-H. Conze**, A. Le Guilcher, V. Ricquebourg, A.-A. Benyoussef, P. Massin, J.-B. Rottier, B. Cochener, and G. Quellec, « Automatic screening for ocular anomalies using fundus photographs », *Optometry and Vision Science*, vol. 99, *3*, pp. 281–291, 2022.
- G. Quellec, H. Al Hajj, M. Lamard, **P.-H. Conze**, P. Massin, and B. Cochener, « ExplAIn: explanatory artificial intelligence for diabetic retinopathy diagnosis », *Medical Image Analysis*, vol. 72, p. 102 118, 2021.
- **P.-H. Conze**, A. E. Kavur, E. Cornec-Le Gall, N. S. Gezer, Y. Le Meur, M. A. Selver, and F. Rousseau, « Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks », *Artificial Intelligence in Medicine*, vol. 117, p. 102 109, 2021.
- A. Touil, K. Kalti, **P.-H. Conze**, B. Solaiman, and M. A. Mahjoub, « A new conditional region growing approach for microcalcification delineation in mammograms », *Medical & Biological Engineering & Computing*, vol. 59, pp. 1795–1814, 2021.
- Y. Yan, **P.-H. Conze**, M. Lamard, G. Quellec, B. Cochener, and G. Coatrieux, « Towards improved breast mass detection using dual-view mammogram matching », *Medical Image Analysis*, vol. 71, p. 102 083, 2021.
- Y. Yan, **P.-H. Conze**, G. Quellec, M. Lamard, B. Cochener, and G. Coatrieux, « Two-stage multi-scale breast mass segmentation for full mammogram analysis without user intervention », *Biocybernetics and Biomedical Engineering*, vol. 41, *2*, pp. 746–757, 2021.
- A. E. Kavur, N. S. Gezer, M. Barış, S. Aslan, **P.-H. Conze**, V. Groza, D. D. Pham, S. Chatterjee, P. Ernst, S. Özkan, *et al.*, « CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation », *Medical Image Analysis*, vol. 69, p. 101 950, 2021.
- A. Touil, K. Kalti, **P.-H. Conze**, B. Solaiman, and M. Mahjoub, « Automatic detection of microcalcification based on morphological operations and structural similarity indices », *Biocybernetics and Biomedical Engineering*, vol. 40, *3*, pp. 1155–1173, 2020.
- **P.-H. Conze**, S. Brochard, V. Burdin, F. T. Sheehan, and C. Pons, « Healthy versus pathological learning transferability in shoulder muscle MRI segmentation using deep convolutional encoder-decoders », *Computerized Medical Imaging and Graphics*, vol. 83, p. 101 733, 2020.

- G. Quellec, M. Lamard, **P.-H. Conze**, P. Massin, and B. Cochener, « Automatic detection of multiple pathologies in fundus photographs using spin-off learning », *Medical Image Analysis*, vol. 61, 2020.
- M. S. Chaibou, **P.-H. Conze**, K. Kalti, M. A. Mahjoub, and B. Solaiman, « Learning contextual superpixel similarity for consistent image segmentation », *Multimedia Tools and Applications*, vol. 79, *3*, pp. 2601–2627, 2020.
- K. Souadih, A. Belaid, D. Ben Salem, and **P.-H. Conze**, « Automatic forensic identification using 3D sphenoid sinus segmentation and deep characterization », *Medical & Biological Engineering & Computing*, vol. 58, pp. 291–306, 2020.
- **P.-H. Conze**, F. Tilquin, M. Lamard, F. Heitz, and G. Quellec, « Unsupervised learning-based long-term superpixel tracking », *Image and Vision Computing*, vol. 89, pp. 289–301, 2019.
- H. Al Hajj, M. Lamard, **P.-H. Conze**, S. Roychowdhury, X. Hu, G. Maršalkaitė, O. Zisimopoulos, M. A. Dedmari, F. Zhao, J. Prellberg, *et al.*, « CATARACTS: challenge on automatic tool annotation for cataract surgery », *Medical Image Analysis*, vol. 52, pp. 24–41, 2019.
- H. Al Hajj, M. Lamard, **P.-H. Conze**, B. Cochener, and G. Quellec, « Monitoring tool usage in surgery videos using boosted convolutional and recurrent neural networks », *Medical Image Analysis*, vol. 47, pp. 203–218, 2018.
- M. S. Chaibou, **P.-H. Conze**, K. Kalti, B. Solaiman, and M. A. Mahjoub, « Adaptive strategy for superpixel-based region-growing image segmentation », *Journal of Electronic Imaging*, vol. 26, *6*, p. 61 605, 2017.
- **P.-H. Conze**, V. Noblet, F. Rousseau, F. Heitz, V. de Blasi, R. Memeo, and P. Pessaux, « Scale-adaptive supervoxel-based random forests for liver tumor segmentation in dynamic contrast-enhanced CT scans », *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, *2*, pp. 223–233, 2017.
- **P.-H. Conze**, P. Robert, T. Crivelli, and L. Morin, « Multi-reference combinatorial strategy towards longer long-term dense motion estimation », *Computer Vision and Image Understanding*, vol. 150, pp. 66–80, 2016.
- T. Crivelli, M. Fradet, **P.-H. Conze**, P. Robert, and P. Pérez, « Robust optical flow integration », *IEEE Transactions on Image Processing*, vol. 24, *1*, pp. 484–498, 2014.

Publications in international conferences and workshops

- R. Zeghlache, **P.-H. Conze**, M. El Habib Daho, Y. Li, H. Le Boité, R. Tadayoni, P. Massin, B. Cochener, A. Rezaei, I. Brahim, G. Quellec, and M. Lamard, « LaTiM: Longitudinal representation learning in continuous-time models to predict disease progression », *in International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2024.

- V. Jaouen, **P.-H. Conze**, J. Bert, and D. Visvikis, « Self super-resolution of anisotropic volumes in prostate MRI with normalized edge priors », *in International Conference on the use of Computers in Radiation therapy*, 2024.
- P. Zhang, Y. Li, J. Zhang, W. Jiang, **P.-H. Conze**, M. Lamard, G. Quellec, and M. El Habib Daho, « Detection and classification of glaucoma in the JustRAIGS challenge: achievements in binary and multilabel classification », *in Justified Referral in AI Glaucoma Screening challenge, in conjunction with IEEE International Symposium on Biomedical Imaging*, 2024.
- M. Abbas, G. Adrade-Miranda, V. Bourbonne, D. Visvikis, B. Badic, and **P.-H. Conze**, « Learning transferability in deep segmentation of liver metastases », *in IEEE International Symposium on Biomedical Imaging*, 2024.
- R. Da-Ano, O. Tankyevych, G. Andrade-Miranda, **P.-H. Conze**, C. Cheze Le Rest, and D. Visvikis, « Multi-modal PET/CT fusion for automated PD-L1 status prediction in lung cancer », *in IEEE International Symposium on Biomedical Imaging*, 2024.
- A. Sadikine, B. Badic, E. Ferrante, V. Noblet, P. Ballet, D. Visvikis, and **P.-H. Conze**, « Deep vessel segmentation with joint multi-prior encoding », *in IEEE International Symposium on Biomedical Imaging*, 2024.
- V. Jaouen, **P.-H. Conze**, and D. Visvikis, « One-sided unsupervised medical image synthesis with normalized edge consistency », *in IEEE International Symposium on Biomedical Imaging*, 2024.
- R. Zeglache, **P.-H. Conze**, M. El Habib Daho, L. Yihao, H. Le Boité, P. Massin, R. Tadayoni, B. Cochener, I. Brahim, G. Quellec, and M. Lamard, « LMT: longitudinal mixing training a framework for the prediction of disease progression using a single image », *in International Workshop on Machine Learning in Medical Imaging*, 2023.
- R. Zeghlache, **P.-H. Conze**, M. E. H. Daho, Y. Li, H. Le Boité, R. Tadayoni, P. Massin, B. Cochener, I. Brahim, G. Quellec, and M. Lamard, « Longitudinal self-supervised learning using neural ordinary differential equation », *in International Workshop on Predictive Intelligence in Medicine*, 2023.
- M. E. H. Daho, Y. Li, R. Zeghlache, Y. C. Atse, H. Le Boité, S. Bonnin, D. Cosette, P. Deman, L. Borderie, C. Lepicard, R. Tadayoni, B. Cochener, **P.-H. Conze**, M. Lamard, and G. Quellec, « Improved automatic diabetic retinopathy severity classification using deep multimodal fusion of UWF-CFP and OCTA images », *in International Workshop on Ophthalmic Medical Image Analysis*, 2023.
- A. Sadikine, B. Badic, J.-P. Tasu, V. Noblet, P. Ballet, D. Visvikis, and **P.-H. Conze**, « Scale-specific auxiliary multi-task contrastive learning for deep liver vessel segmentation », *in IEEE International Symposium on Biomedical Imaging*, 2023.

- Y. Li, R. Zeghlache, I. Brahim, H. Xu, Y. Tan, **P.-H. Conze**, M. Lamard, G. Quellec, and M. El Habib Daho, « Segmentation, classification, and quality assessment of UW-OCTA images for the diagnosis of diabetic retinopathy », *in Mitosis Domain Generalization and Diabetic Retinopathy Analysis*, 2022, pp. 146–160.
- R. Zeghlache, **P.-H. Conze**, M. El Habib Daho, R. Tadayoni, P. Massin, B. Cochener, G. Quellec, and M. Lamard, « Detection of diabetic retinopathy using longitudinal self-supervised learning », *in International Workshop on Ophthalmic Medical Image Analysis*, 2022, pp. 43–52.
- Y. Li, M. El Habib Daho, **P.-H. Conze**, H. Al Hajj, S. Bonnin, H. Ren, N. Manivannan, S. Magazzeni, R. Tadayoni, B. Cochener, *et al.*, « Multimodal information fusion for glaucoma and diabetic retinopathy classification », *in International Workshop on Ophthalmic Medical Image Analysis*, 2022, pp. 53–62.
- I. Brahim, M. Lamard, A.-A. Benyoussef, **P.-H. Conze**, B. Cochener, D. Cornec, and G. Quellec, « Mapping the ocular surface from monocular videos with an application to dry eye disease grading », *in International Workshop on Ophthalmic Medical Image Analysis*, 2022.
- A. Sadikine, B. Badic, J.-P. Tasu, V. Noblet, D. Visvikis, and **P.-H. Conze**, « Semi-overcomplete convolutional auto-encoder embedding as shape priors for deep vessel segmentation », *in IEEE International Conference on Image Processing*, 2022, pp. 586–590.
- G. Andrade-Miranda, V. Jaouen, V. Bourbonne, F. Lucia, D. Visvikis, and **P.-H. Conze**, « Pure versus hybrid transformers for multi-modal brain tumor segmentation: A comparative study », *in IEEE International Conference on Image Processing*, 2022, pp. 1336–1340.
- M. M. Islam, B. Badic, T. Aparicio, D. Tougeron, J.-P. Tasu, D. Visvikis, and **P.-H. Conze**, « Deep treatment response assessment and prediction of colorectal cancer liver metastases », *in International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2022, pp. 482–491.
- A. Boutillon, **P.-H. Conze**, C. Pons, V. Burdin, and B. Borotikar, « Multi-task, multi-domain deep segmentation with shared representations and contrastive regularization for sparse pediatric datasets », *in International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2021, pp. 239–249.
- Y. Yan, **P.-H. Conze**, M. Lamard, H. Zhang, G. Quellec, B. Cochener, and G. Coatrieux, « Deep active learning for dual-view mammogram analysis », *in International Workshop on Machine Learning in Medical Imaging*, 2021, pp. 180–189.
- Y. Yan, **P.-H. Conze**, G. Quellec, P. Massin, M. Lamard, G. Coatrieux, and B. Cochener, « Longitudinal detection of diabetic retinopathy early severity grade changes using deep learning », *in International Workshop on Ophthalmic Medical Image Analysis*, 2021, pp. 11–20.

- A. Boutillon, B. Borotikar, C. Pons, V. Burdin, and **P.-H. Conze**, « Multi-structure deep segmentation with shape priors and latent adversarial regularization », *in IEEE International Symposium on Biomedical Imaging*, 2021, pp. 999–1002.
- Y. Yan, **P.-H. Conze**, G. Quellec, M. Lamard, B. Cochener, and G. Coatrieux, « Two-stage multi-scale mass segmentation from full mammograms », *in IEEE International Symposium on Biomedical Imaging*, 2021, pp. 1628–1631.
- H. Messaoudi, A. Belaid, M. L. Allaoui, A. Zetout, M. S. Allili, S. Tliba, D. Ben Salem, and **P.-H. Conze**, « Efficient embedding network for 3D brain tumor segmentation », *in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 2020, pp. 252–262.
- A. Touil, K. Kalti, **P.-H. Conze**, B. Solaiman, and M. A. Mahjoub, « A new conditional region growing approach for an accurate detection of microcalcifications from mammographic images », *in IEEE International Conference on Bioinformatics and Bioengineering*, 2020.
- Y. Yan, **P.-H. Conze**, M. Lamard, G. Quellec, B. Cochener, and G. Coatrieux, « Multi-tasking siamese networks for breast mass detection using dual-view mammogram matching », *in International Workshop on Machine Learning in Medical Imaging*, 2020, pp. 312–321.
- A. Touil, K. Kalti, **P.-H. Conze**, B. Solaiman, and M. A. Mahjoub, « Morphological-based microcalcification detection using adaptive thresholding and structural similarity indices », *in International Conference on Advanced Technologies for Signal and Image Processing*, 2020.
- A. Boutillon, B. Borotikar, V. Burdin, and **P.-H. Conze**, « Combining shape priors with conditional adversarial networks for improved scapula segmentation in MR images », *in IEEE International Symposium on Biomedical Imaging*, 2020, pp. 1164–1167.
- Y. Yan, **P.-H. Conze**, E. Decencière, M. Lamard, G. Quellec, B. Cochener, and G. Coatrieux, « Cascaded multi-scale convolutional encoder-decoders for breast mass segmentation in high-resolution mammograms », *in Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2019, pp. 6738–6741.
- **P.-H. Conze**, C. Ponse, V. Burdin, F. T. Sheehan, and S. Brochard, « Deep convolutional encoder-decoders for deltoid segmentation using healthy versus pathological learning transferability », *in IEEE International Symposium on Biomedical Imaging*, 2019, pp. 36–39.
- F. Tilquin, **P.-H. Conze**, P. Pessaux, M. Lamard, G. Quellec, V. Noblet, and F. Heitz, « Robust supervoxel matching combining mid-level spectral and context-rich features », *in Interional Workshop on Patch-based Techniques in Medical Imaging*, 2018, pp. 39–47.
- A. Guerre, M. Lamard, **P.-H. Conze**, B. Cochener, and G. Quellec, « Optical flow estimation in ocular endoscopy videos using Flownet on simulated endoscopy data », *in IEEE International Symposium on Biomedical Imaging*, 2018, pp. 1463–1466.

- **P.-H. Conze**, F. Tilquin, M. Lamard, F. Heitz, and G. Quellec, « Long-term superpixel tracking using unsupervised learning and multi-step integration », *in International Conference on Advanced Technologies for Signal and Image Processing*, 2018.
- **P.-H. Conze**, F. Tilquin, V. Noblet, F. Rousseau, F. Heitz, and P. Pessaux, « Hierarchical multi-scale supervoxel matching using random forests for automatic semi-dense abdominal image registration », *in IEEE International Symposium on Biomedical Imaging*, 2017, pp. 490–493.
- **P.-H. Conze**, V. Noblet, F. Rousseau, F. Heitz, R. Memeo, and P. Pessaux, « Random forests on hierarchical multi-scale supervoxels for liver tumor segmentation in dynamic contrast-enhanced CT scans », *in IEEE International Symposium on Biomedical Imaging*, 2016, pp. 416–419.
- **P.-H. Conze**, F. Rousseau, V. Noblet, F. Heitz, R. Memeo, and P. Pessaux, « Semi-automatic liver tumor segmentation in DCE-CT scans using random forests and supervoxels », *in International Workshop on Machine Learning in Medical Imaging*, 2015.
- **P.-H. Conze**, P. Robert, T. Crivelli, and L. Morin, « Dense long-term motion estimation via statistical multi-step flow », *in International Conference on Computer Vision Theory and Applications*, 2014, pp. 545–554.
- **P.-H. Conze**, T. Crivelli, P. Robert, and L. Morin, « Dense motion estimation between distant frames: combinatorial multi-step integration and statistical selection », *in IEEE International Conference on Image Processing*, 2013, pp. 3860–3864.
- T. Crivelli, **P.-H. Conze**, P. Robert, M. Fradet, and P. Pérez, « Multi-step flow fusion: towards accurate and dense correspondences in long video shots », *in British Machine Vision Conference*, 2012.
- T. Crivelli, **P.-H. Conze**, P. Robert, and P. Pérez, « From optical flow to dense long term correspondences », *in IEEE International Conference on Image Processing*, 2012, pp. 61–64.
- P. Robert, C. Thébault, and **P.-H. Conze**, « Disparity-compensated view synthesis for s3D content correction », *in Stereoscopic Displays and Applications*, 2012, pp. 843–856.
- **P.-H. Conze**, P. Robert, and L. Morin, « Objective view synthesis quality assessment », *in Stereoscopic Displays and Applications*, 2012, pp. 557–570.

Patents

- R. Zeghlache, **P.-H. Conze**, M. El Habib Daho, M. Lamard, and G. Quellec, *Method and apparatus for predicting progression of a pathology*, EP 23306730.5, 2023.
- P. Robert, V. Drazic, **P.-H. Conze**, and T. Viellard, *Disparity maps in uniform areas*, US 20130176300, 2013.

- P. Robert, M. Fradet, **P.-H. Conze**, and T. Viellard, *Method and apparatus for processing occlusions in motion estimation*, US 20130148730, EP 2602997, 2013.
- P. Robert, T. Crivelli, **P.-H. Conze**, M. Fradet, and P. Pérez, *Filtering a displacement field between video frames*, WO 2013131819, EP 2823467, US 20150015792, 2013.
- M. Fradet, **P.-H. Conze**, *et al.*, *Procédé d'édition d'un plan dans une séquence vidéo*, FR 2998399, 2013.
- **P.-H. Conze**, P. Robert, T. Crivelli, and L. Morin, *Method for generating a motion field for a video sequence*, WO 2014122131, 2014.
- P. Robert, T. Crivelli, and **P.-H. Conze**, *Method and device for generating a motion field for a video sequence*, WO 2013107833, US 20140363053, EP 2805306, 2014.

Abstracts in international conferences

- M. El Habib Daho, Y. Li, R. Zeghlache, A. Rezaei, H. Le Boité, A. Couturier, S. Magazzeni, A. Le Guilcher, F. Potevin, M. Gallardo, R. Tadayoni, B. Cochener, **P.-H. Conze**, M. Lamard, and G. Quellec, « Cross-device AI fusion: enhancing diabetic retinopathy diagnosis with combined Clarus and Optos images », *Investigative Ophthalmology & Visual Science*, 2024.
- M. Faure, **P.-H. Conze**, B. Cochener, M. Lamard, and G. Quellec, « Predicting cataract surgery errors via artificial intelligence alert generation », *Investigative Ophthalmology & Visual Science*, 2024.
- V. Jaouen, Z. Wang, **P.-H. Conze**, and D. Visvikis, « Self super-resolution for hepatic vessel CT segmentation », *in IEEE Medical Imaging Conference*, 2023.
- R. Zeghlache, **P.-H. Conze**, M. El Habib Daho, Y. Li, I. Brahim, H. Le Boité, P. Massin, R. Tadayoni, B. Cochener, G. Quellec, and M. Lamard, « Time-aware deep models for predicting diabetic retinopathy progression », *Investigative Ophthalmology & Visual Science*, 2023.
- M. El Habib Daho, R. Zeghlache, Y. Li, H. Le Boité, S. Bonnin, S. Magazzeni, L. Borderie, B. Lay, R. Tadayoni, B. Cochener, **P.-H. Conze**, M. Larmard, and G. Quellec, « Performance of two ultra-widefield retinal imaging systems for the automatic diagnosis of diabetic retinopathy », *Investigative Ophthalmology & Visual Science*, 2023.
- G. Quellec, Y. Li, H. Al Hajj, S. Bonnin, H. Ren, N. Manivannan, S. Magazzeni, R. Tadayoni, **P.-H. Conze**, and M. Lamard, « 3D style transfer between structure and flow channels in OCT angiography », *Investigative Ophthalmology & Visual Science*, vol. 63, 7, 2022.
- Y. Li, H. Al Hajj, **P.-H. Conze**, S. Bonnin, H. Ren, N. Manivannan, S. Magazzeni, R. Tadayoni, M. Lamard, and G. Quellec, « Multimodal information fusion for the diagnosis of diabetic retinopathy », *Investigative Ophthalmology & Visual Science*, vol. 63, 7, 2022.

- G. Andrade-Miranda, V. Jaouen, D. Visvikis, and **P.-H. Conze**, « Comparing modern segmentation architectures under low data regime for PET-CT tumor segmentation », *in IEEE Medical Imaging Conference*, 2022.
- G. Sallé, V. Bourbonne, **P.-H. Conze**, N. Boussion, J. Bert, D. Visvikis, and V. Jaouen, « Tumor blending augmentation using one-shot generative learning for brain CT tumor segmentation », *in IEEE Medical Imaging Conference*, 2022.
- G. Sallé, **P.-H. Conze**, N. Boussion, J. Bert, D. Visvikis, and V. Jaouen, « Synthetic tumor insertion using one-shot generative learning for cross-modal image segmentation », *in IEEE Medical Imaging Conference*, 2021.
- G. Quellec, M. Lamard, **P.-H. Conze**, P. Massin, and B. Cochener, « Automatic detection of multiple pathologies in fundus photographs », *Investigative Ophthalmology & Visual Science*, vol. 61, 7, 2020.

Publications in national conferences

- **P.-H. Conze**, F. Rousseau, V. Noblet, F. Heitz, R. Memeo, and P. Pessaux, « Segmentation semi-automatique de tumeurs du foie en TDM dynamique pour l'estimation du taux de nécrose », *in Colloque GRETSI Traitement du Signal & des Images*, 2015.
- **P.-H. Conze**, T. Crivelli, P. Robert, and L. Morin, « Estimation de mouvement entre images distantes: intégration combinatoire et sélection statistique », *in Colloque GRETSI Traitement du Signal & des Images*, 2013.

Pre-prints

- P. Rougé, **P.-H. Conze**, N. Passat, and O. Merveille, « Guidelines for cerebrovascular segmentation: managing imperfect annotations in the context of semi-supervised learning », *arXiv preprint arXiv:2404.01765*, 2024.
- V. Jaouen, **P.-H. Conze**, G. Dardenne, J. Bert, and D. Visvikis, « Regularized directional representations for medical image registration », *arXiv preprint arXiv:2111.15509*, 2021.

Article for general public

- **P.-H. Conze**, « Les promesses et enjeux de l'IA pour l'interprétation d'images médicales », *in Télécom*, 2024.

PhD thesis

- **P.-H. Conze**, « Long-term dense motion estimation and view synthesis quality assessment with app. to joint stereo and motion processing », Ph.D. dissertation, INSA Rennes, 2014.

SUPERVISION

Post-doctoral fellowships

- A. Rezaei, from December 2023. *Artificial intelligence for diabetic retinopathy management.*
- R. Da-Ano, from October 2023. *Deep radiomics for treatment response assessment.*
- M. El Habib Daho, from Jan. 2022. *Predicting diabetic retinopathy using artificial intelligence.*
- G. Andrade-Miranda, from April 2021 to March 2023. *Computational models with cross-modality learning.*

PhD theses

Defended

- N. Decaux, defended in June 2024. *Structure and function analysis from muscle MRI data*, supervised by F. Rousseau and S. Brochard.
- Y. Li, defended in December 2023. *Multi-modal information fusion for the diagnosis of diabetic retinopathy*, supervised by G. Quellec.
- N. Ben Chaabane, defended in December 2023. *Analysis of physiological signals for the prediction of pathological evolution using artificial intelligence. Application to Parkinson disease and quantified gait analysis*, supervised by M. Lamard.
- I. Brahim, defended in December 2022. *Automatic dry eye quantification using artificial intelligence*, supervised by G. Quellec and D. Cornec.
- A. Boutillon, defended in November 2022. *Regularized deep learning models for multi-anatomy segmentation in pediatric imaging*, supervised by V. Burdin.
- Y. Yan, defended in October 2021. *Medical image analysis with deep learning for computer-aided diagnosis in screening*, supervised by G. Coatrieux.
- A. Touil, defended in November 2021. *Collaborative combination of classifiers: Application to microcalcification detection in mammograms*, supervised by B. Solaiman and M. A Mahjoub.
- S. Matta, defended in April 2021. *Automatic recognition of retinal pathologies using deep learning for mass screening*, supervised by G. Quellec.
- A. Guerre, defended in December 2019. *Enhanced field of view for ocular endoscopic surgery*, supervised by G. Quellec and B. Cochener.
- M. S Chaibou, defended in July 2019. *Interpretation of images by iterative knowledge integration*, supervised by B. Solaiman and M. A Mahjoub.

Ready to defend

– M. A Sadikine, started in October 2020, defense planed in July 2024. *Deep vessel segmentation with geometric and topological constraints*, supervised by P. Ballet.

On going

– S. Dedeken, started in September 2023. *Robustness of multi-modal machine learning models to heterogeneous imaging data*, supervised by D. Visvikis, in collaboration with Sophia Genetics[13].
– M. Abbas, started in November 2022. *Longitudinal follow-up of metastases from colorectal cancer using artificial intelligence*, supervised by B. Badic.
– M. Faure, starded in November 2021. *Surgical assistance based on surgical simulators and artificial intelligence*, supervised by G. Quellec.
– R. Zeghlache, started in October 2021. *Longitudinal follow-up of diabetic retinopathy with deep learning*, supervised by M. Lamard.

Master theses (M2)

– E. Kharroubi, from March to Aug. 2024. *Deep characterization of prostate cancer.*
– C. Martin, from February to July 2022. *Towards a generic 3D vascular segmentation network.*
– A. Ruiz Guijosa, from Jan. to Jun. 2022. *Topological constraints for liver vascular segmentation.*
– L. Wang, from March to Sept. 2022. *Longitudinal follow-up of metastases from colorectal cancer.*
– M. Islam, from Feb. to June 2021. *Predicting treatment response of liver metastases.*
– M. Riera i Marin, from May to Oct. 2020. *Segmentation of fetal brain MRI by neural networks.*
– M. Koralewski, from April to Sept. 2020. *Deep hepatic vascular system segmentation.*
– G. Sallé, from April to September 2020. *Multi-modal segmentation of polycystic kidneys.*
– A. Boutillon, from Mars to September 2019. *MR image segmentation using deep learning.*
– Y. Yan, from Mars to August 2018. *Deep mammogram analysis for breast cancer diagnosis.*
– F. Tilquin, from Mars to September 2016. *Multi-modal registration of abdominal images.*

Research internships (M1)

– M. Lin, 2024. *Self super-resolution for prostate cancer segmentation and detection.*
– I. Hamoud, 2018. *Convolutional neural networks and transfer learning for liver segmentation.*
– A. Heitz, 2016. *Deep learning for CT and MRI segmentation of liver tumors.*
– F. Allender, 2016. *Random forests and multi-scale supervoxels for glioma classification.*
– A. Krebs, 2015. *Segmentation of liver tumors for necrosis rate estimation.*

13. `https://www.sophiagenetics.com`

TEACHING

Teaching responsibilities

- From Sept. 2023: co-responsible for Health Engineering option (M1, M2) at IMT Atlantique
- From Sept. 2020: co-responsible for research projects (M2) at IMT Atlantique
- From Sept. 2019: responsible for the Digital Patient[14] courses (M1, M2) at IMT Atlantique
- From Sept. 2019: responsible for the Computer Vision[15] courses (M1, M2) at IMT Atlantique

Provided courses (∼600h)

- Main courses provided at IMT Atlantique

Module	Level	Year	Hours/year
Deep learning	M1, M2	2019 -	12h
Introduction to artificial intelligence	M1, M2	2020 - 2022	21h
Digital patient[14]	M1, M2	2019 -	32h
Computer vision[15]	M1, M2	2019 -	25h
Digital image processing	M2	2018 - 2019	10h
Information coding	L3, M1	2017 - 2018	3h
Multimedia technologies	L3, M1	2017 - 2018	3h

- Courses provided at ENSSAT

Module	Level	Year	Hours/year
Medical image analysis	M2	2023 -	6h
Machine and deep learning	M2	2023 -	4h

- Courses provided at Télécom Physique Strasbourg

Module	Level	Year	Hours/year
Probability and stochastic processes	L3	2015 - 2016	35h
Statistics	M1	2015 - 2016	14h
Numerical analysis	L3	2015 - 2016	16h

- Lecturer at European School for Medical Physics Experts (ESMPE) in 2022 and 2024

14. medical signal processing, computer-assisted medical image analysis, medical image segmentation, medical image registration, time: the fourth dimension in medical image analysis, statistical shape models, morphometry

15. object detection and recognition, semantic segmentation, image interpretation, dense motion estimation, optical flow, video object tracking, single-camera geometry, stereo and 3D structure, two-view geometry

COLLEGE SCIENCES
DOCTORAL DE LA VIE
BRETAGNE ET DE LA SANTE

Université de Bretagne Occidentale

Titre : Contributions à la segmentation d'images médicales par intelligence artificielle pour l'aide à la décision

Mot clés : analyse d'images médicales, apprentissage profond, segmentation sémantique, fusion multi-modale, suivi longitudinal, imagerie abdominale

Résumé : L'analyse d'images médicales assistée par ordinateur offre un avantage notable par rapport aux analyses manuelles subjectives et chronophages, en s'appuyant sur des techniques informatiques reproductibles pour interpréter les images médicales. Le développement de modèles computationnels fait de plus en plus appel à des techniques d'intelligence artificielle, ouvrant la voie à une médecine personnalisée. Ce manuscrit synthétise mes contributions et perspectives de recherche en analyse d'images médicales par intelligence artificielle, pour l'aide à la décision. À l'interface entre traitement d'images, science des données et mathématiques appliquées, il met en lumière mes contributions en segmentation sémantique, fusion multi-modale et suivi longitudinal par apprentissage profond. L'enthousiasme pour l'intelligence artificielle cache un aspect plus nuancé, lié aux nombreux défis d'un déploiement plus large en routine clinique. Étant donnée la complexité de la collecte et de l'annotation d'images médicales, divers paradigmes d'apprentissage ont émergé : transfert d'apprentissage, intégration de connaissances a-priori, apprentissage semi-supervisé, multi-tâche ou multi-domaine... De plus, les modèles computationnels visent à fournir une aide décisionnelle interprétable pour les médecins. Divers aspects de ces systèmes sont décrits dans les contextes de la planification pré-opératoire et du suivi thérapeutique, notamment en imagerie abdominale.

Title: Contributions to medical image segmentation with artificial intelligence for decision support

Keywords: medical image analysis, deep learning, semantic segmentation, multi-modal fusion, longitudinal follow-up, abdominal imaging

Abstract: Computer-aided medical image analysis provides a significant advantage over labor-intensive and subjective manual analyses by leveraging reproducible and objective computational techniques to interpret medical images. With the rapid advancement of artificial intelligence, the field has witnessed a shift towards deep learning methodologies able to perform a wide range of interpretation tasks, paving the way for personalized medicine. This manuscript synthesizes my past contributions, describes ongoing works and anticipate future prospects on medical image analysis with artificial intelligence for better clinical decision support. At the interface between image processing, data science and applied mathematics, it highlights my contributions on semantic segmentation, multi-modal information fusion and longitudinal follow-up with deep learning. Enthusiasm for artificial intelligence hides a more nuanced side, linked to the challenges of a wider deployment in clinical routine. Given the complexity of collecting and annotating a large number of medical images, various learning paradigms have emerged comprising transfer learning, prior knowledge embedding, semi-supervised learning, multi-task or multi-domain analysis. Further, medically-sound computational models aim at providing an interpretable guidance for clinicians. Various facets of decision support systems are described in pre-operative planning and therapeutic follow-up, with a focus on abdominal imaging.