|  |  |
| --- | --- |
| Full name of the applicant | Victor Joos |
| Reference |  |

**SCIENTIFIC SECTION OF THE PROPOSAL**

Main language chosen = English

|  |
| --- |
| This part includes the following elements:   1. Description of the research project 2. Comments on changes made in the research project in case of resubmission (optional) 3. Activities report on the first year of doctorate (**ONLY** for 1st grant - **2nd year** applicants)**\*** 4. Potential interdisciplinary approach of the research project (optional) 5. Description of the work environment 6. Summary of the master’s thesis or equivalent 7. Additionnal comments (optional) 8. Ph.D. work calendar per month   **\* “1st grant - 2nd year” applicants have already worked on a full-time basis for one year full time equivalent on the Ph.D. project submitted to the FRIA.**  **The applicant must fill in the sections below and convert the file into an unprotected PDF before appending it to the online application form.**  The F.R.S.-FNRS insists on **strict compliance with the instructions given for each part of the proposal** (scientific section relevant to the instrument selected, number of pages allowed for documents to be enclosed with the application form…) and stresses again the sovereign consideration of the juries in case the file would exceed the applicable page limit. |

1. **Description of the research project**
   1. **Goals of the research**

Convolutional neural networks (CNNs) have recently emerged in the computer vision community as the most effective tools to segment natural images. In this project, we consider the use of CNNs for biomedical image segmentation and interpretation, with a focus on the X-ray CT scans of vascular and nervous networks encountered in applicative fields ranging from pulmonary embolism (PE) diagnosis to accurate and detailed characterization of tissues:

* PE is an extremely common and highly lethal disease that is a leading cause of death in all age groups. Diagnosis requires to detect and position emboli in pulmonary vessels. Segmenting the blood vessels in the lungs is a critical step for this task.
* Tissue engineering envisions the regeneration of a damaged body part through the manufacturing of appropriate constructs, but requires detailed knowledge about the 3D morphology and organization of the different types/layers of the extracellular matrix.

However, extension of CNN usage to the biomedical field is generally not trivial. This is partly due to the fact that deep (convolutional) neural nets are trickier to train than shallow multilayer perceptrons [5], but also to their failure to account for the specificities of biomedical content and imaging systems. Recent works, including in UCLouvain [1]–[3], have largely revealed that off-the-shelf CNN-based segmentation algorithms fail in exploiting common sense prior knowledge (e.g. related to the convexity of objects, or to the tree-like structure of blood vessels), making their high-performance dependent on the availability of a large training set of fully annotated samples (i.e. samples for which the model response is explicitly known, in most cases through manual data annotation). When dealing with biomedical images, annotation is especially tedious, due to the expert knowledge required to distinguish tissues and organs of interest. Moreover, the acquisition of X-ray image samples is constrained by the radiation toxicity or the availability of samples augmented with appropriate contrast agents [ref de Greet].

To circumvent the need for large sets of annotated training data, our project investigates training methodologies that mitigate the need for supervision when tailoring convolutional neural networks to specific tasks. In line with recent contributions from the CNN scientific community [6], our **research hypothesis** states that the image representations, also named features, manipulated by a neural network **trained for a given task might**, in some cases, **be relevant for another task**.This assumption raises a number of fundamental questions related to the conditions under which the transfer of features between CNNs becomes relevant. Those questions are central to the research methodology detailed in Section 1.4.

Overall, this project aims at bypassing the need for large annotated CT datasets by developing training methods that **derive image features from raw input data and prior knowledge about the branch-like structures of the vascular structures to segment**. As explained below, whilst challenging, our project sounds timely and realistic given the very recent progresses of the computer vision community in terms of clustering-based training methods.

* 1. **State of the art**

Less than five years ago, predominant methods for automatic image analysis and interpretation were based on the combination of ad-hoc hand-crafted representations, often chosen for their sparsity [7], [8], with graph-based processing [9], [10] and/or supervised machine learning approaches suited to low-dimensional feature spaces [11]. In the medical imaging community, altas-based methods [12] and statistical shape models [13] have also been intensively used as segmentation tools due to their ability to segment regions with few position, shape and appearance changes (as it is for example the case for many organs).

Since very recently, however, (2012 for image classification [14], 2014 for object detection [15], and 2015 for image segmentation [16]) a new class of cutting-edge machine learning algorithms, has emerged in the computer vision community, founded on artificial neural networks to deal with raw data, instead of sparse representations or manually engineered descriptors. Those algorithms, named convolutional neural networks (or CNN), have surpassed traditional methods on almost all computer vision tasks. In particular, the segmentation task is best served by networks adopting an auto-encoder structure, with skip-connections, as the U-Net [16] for 2D biomedical images, or the 3D U-Net [17] and the V-Net [18] for their 3D equivalent.

Despite their massive number of parameters, those CNN architectures have been shown to exhibit a remarkably small difference between training and test performance, as long as they are trained with a representative set of samples, and with proper regularization mechanisms (weight decay, dropout, normalization of activations, data augmentation). This observation shifts the problem towards the collection of data (image+annotations) to supervise CNN training. In many real-life cases, this raises a significant issue, especially for vascular segmentation in X-rays, where expert annotations are needed, and accurate segmentation is time-consuming ([19] extrapolates more than a year of annotation for the vascular structure of one mouse brain from the annotation of 0.02% of a single brain). Previous works, including from UCLouvain [ref simone], have studied the use of synthetic data [19], [20], but transferring the trained model from synthetic data towards real data still necessitate a significant subset of data from expert annotation (11% in [19]).

As an alternative to domain transfer, self-supervised methods [7]-[9] pre-learn the CNN internal features by using the content of the raw (without annotation) images to define pretext tasks. They allow training with reduced need for annotation and with no domain transfer issue (since inputs are the same for the pretext tasks and the task of interest). Very recently, self-supervised methods using using a contrastive loss to cluster the output vectors predicted by the CNN [10]-[16], have shown strong performance for image classification, with only a small need for model fine-tuning using annotated data. Those methods define the clusters independently of the task labels and based on input data augmentation. However, the extensive use of data augmentation makes such contrastive methods not directly suited for segmentation.

In addition to self-supervised methods, we aim at exploiting the branching topology of vascular and nervous systems to reduce the need for annotation during training. Some examples of use of intrinsic structure of data can be found in satellite imaging [17], or in works that combine segmentation with branching detection [18]. Our work plans to formulate self-supervised tasks leveraging the branch like structure of the data, while adopting the recently endorsed clustering-based paradigm.

* 1. **Research project**

Proposing strategies to train, with a limited amount of annotated data, CNN models that segment images exhibiting branching structure, is central to this research project.

To achieve this goal, we plan to investigate the two following questions:

* **How to define self-supervised mechanisms that mitigate the lack of accurate and reliable annotations in a segmentation context?** The definition of accurate annotation is generally unrealistic in the context of X-ray medical imaging due to the lack of expertise, and to the ambiguity inherent to the low X-ray image SNR. Different kinds of pretext tasks and contrastive losses will be developed for various segmentation tasks, as explained in Section 1.4.
* **How to account for the specificities of vascular or nervous structure as prior for the self-supervision task**? The specific (dis)similarity induced by vascular structures between neighboring CT voxels will be considered to define self-supervised contrastive losses that promote the learning of CNN features that capture the branching topology of the signal. We also plan to investigate the interaction between multiple tasks in self-supervised learning: joint learning of the main segmentation task and either branching detection or vascular orientation is expected to lead to better results in the different sub-tasks.
  1. **Work plan**

As explained in the previous section, this project addresses two main questions, respectively related to self-supervision and to the exploitation of the vascular structure as prior. Our work plan is thus organized in five tasks, ranging from data collection to understanding the emergence of CNN internal representations, while developing different kinds of (self-)supervised methods.

Task 1. Data Collection

In practice, to be properly fed in terms of data, our work will cover multiple applicative fields , each of them offering distinct specificities in terms of data access and data structure. In more details, the applicative fields of interest are:

* **Pulmonary Embolism Detection**: Computerized tomography (CT) scanners have gained acceptance as a minimally invasive method for diagnosing pulmonary embolisms. CNNs are here considered to detect embolism in images of low-contrast (to reduce the X-ray doses of CT image acquisition as low as reasonably achievable). The developed method will be used as a safeguard by radiologists, or even as preliminary computer-aided diagnosis (CAD) tool (for example for removing negative scanners from human diagnosis). This work will be done partly in the Radiology Department of the St Luc Clinics in Brussels, which will provide the required data sets. The models could be extended to better characterize pulmonary inflammation due to Covid-19.
* **Tissue Engineering**: Body parts (e.g. face, limbs, trunk) are made of composite tissues, which are organized in assembled tissue layers (e.g. skin, fat and cartilage for the ear) with an intertwining vascular and nervous network, shown in Fig. 1 (sharing the same shape ramification as the lung blood vessels considered for pulmonary embolism). Contrast-enhanced micro-CT now allows visualizing those structures, but accurate image segmentation tools are required to extract quantitative and statistically relevant information about them [3]. In this context, images are captured on dead mices, which gives the opportunity to consider multiple acquisitions of the same body part (typically a knee articulation or kidneys) with different settings and thus different SNR levels, allowing to properly assess the CNN performance (since high SNR images give access to a reliable ground truth). Access to images is guaranteed through our close interaction (a 4-year collaborative project started in 2019) with Prof. Greet Kerckhofs, expert in contrast enhanced micro-CT.

Task 2. Supervised CNN implementation for segmentation with shape regularization

State-of-the-art deep learning segmentation architectures are fully convolutional neural networks. For biomedical applications, the U-Net architecture (2D [16] or 3D [17] implementation) is particularly suited and is commonly used in the community. 2D and 3D U-Net have been implemented in UCLouvain to segment the bladder on CT and CBCT image slices [1], [2], but also the mineralized cartilage on micro-CT images [3]. Those two networks provide baseline references, to be compared with the alternatives studied in the rest of the project.This baseline will allow us to investigate the statistical characteristics of a model trained on the available data, which is the foundation of the following tasks, and will orient this work.

*Shape Prior:* When CNNs are only trained with the classical binary cross-entropy or the Dice loss, the network prediction and the ground truth (i.e. the manual annotations) are compared on a pixel-wise manner and do not necessarily incorporate local geometry such as smoothness and shape features (e.g. tree-based topology of vascular elements). For the integration of the shape prior, we consider the regularization of the image as pretext task for the training of a segmentation model.

The purpose is to favor the emergence of a segmentation mask that shares the structure exhibited by the ground truth output labels. Hence, the first proposed CNN regularization scheme relies on the assumption that the ground truth labels exhibit such an internal structure. This is the case in the considered datasets since the organs and vascular systems in CT/micro-CT images contain connected structures.

Our proposed regularization scheme works in two steps. First, an auto-encoder will be trained to model the internal structure of ground-truth segments. In a second phase, the actual segmentation network of interest (i.e. the U-Net model) is trained by including the auxiliary task of predicting the output via the decoder learned in the first phase [30]. This process is illustrated in Fig. 1. In our case, the auto-encoder will be trained on manual or synthetic segmentation labels embedding models for ramifications.

Alternative approaches to connect the segmentation and the structure encoding networks have been proposed [31], [32] as well. Hence, determining the appropriate way to connect them and to favor the joint training of multiple (i.e. auxiliary and main) tasks by a network remains open and is fully part of this research question [6], [29].

Task 3. Self-supervised segmentation using contrastive losses

*Pre-training Strategy:* We plan to investigate dictionary learning as pretext task. Unsupervised representation learning has been largely used in natural language processing, but has been less successful for vision tasks. Contrastive loss [21], illustrated in Fig. 3 (a), has very recently been envisioned to learn image representations that effectively transfer to a variety of natural vision tasks. Contrastive methods [22]–[24] are trained to predict output image vectors that cluster the sets of samples that are related (e.g. because they have been generated based on the same initial image, using various kinds of data augmentation), while maximizing distance between samples that are not related. Those contrastive methods have been designed for image classification, and are not directly suited for pixel-wise segmentation. To address this limitation, we plan to develop pixel-wise CNN feature clustering strategies, as depicted in Fig3(b). Instead of clustering or contrasting images instances, as in the classical paradigm, our scheme will cluster identical pixels from different instances of the same image, while contrasting unrelated pixels from the same image.

*Shape Prior:* In the previous statement, we can replace identical pixels with neighboring pixels, according to the branch-like structure prior. In order to cluster pixels according to the prior, we will use local information using classical algorithms like directional filters, or the graphcut [33] or watershed algorithms [34].

Task 4. Self-supervised segmentation with image reconstruction loss

*Pre-training Strategy:* In this task, we consider embedding the segmentation task into an image reconstruction task as pretext. As illustrated in Fig. X, a trained reconstruction network, will take an incomplete image and a vector field trained from the original image to recreate the given image. The network of interest will be trained to output directional information, in order to help the reconstruction task.

*Shape Prior:* The purpose is to favor the emergence of a vector field that shares the branch-like structure exhibited by the ground truth output labels. The direction of the vector field should become tangent to the vascular structure, while background pixels should contain no clear direction. Using this, we can separate the background from the vascular structure, while obtaining information about orientation.

Task 5. Understanding the emergence of CNN internal representations

In order to validate and understand the inner-working of the designed strategies, we will look at the weight and gradient characteristics during training, and compared to the baseline, prompted by work done inside our research team [29]. A possible technique would be to use dimensionality reduction (e.g. PCA) at different points in the network, in order to combine information at every layer. This might lead to unsupervised segmentation using clustering.

[1] J. Léger, E. Brion, U. Javaid, J. Lee, C. De Vleeschouwer, and B. Macq, “Contour Propagation in CT Scans with Convolutional Neural Networks,” in *Advanced Concepts for Intelligent Vision Systems*, Cham, 2018, pp. 380–391.

[2] E. Brion, J. Léger, U. Javaid, J. Lee, C. D. Vleeschouwer, and B. Macq, “Using planning CTs to enhance CNN-based bladder segmentation on cone beam CT,” in *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*, Mar. 2019, vol. 10951, p. 109511M, Accessed: Aug. 03, 2020.

[3] J. Léger, L. Leyssens, C. De Vleeschouwer, and G. Kerckhofs, “Deep Learning-Based Segmentation of Mineralized Cartilage and Bone in High-Resolution Micro-CT Images,” in *Computer Methods, Imaging and Visualization in Biomechanics and Biomedical Engineering*, Cham, 2020, pp. 158–170.

[4] V. Joos de ter Beerst, A. Vanderschueren, and C. De Vleeschouwer, “Utilisation conjointe d’information sémantique et géométrique pour la localisation d’objets par caméra temps-de-vol,” in *Online proceedings of RFIAP*, 2020, vol. 1

[5] G. Leclerc and A. Madry, “The Two Regimes of Deep Network Training,” *arXiv:2002.10376 [cs, stat]*, Feb. 2020, Accessed: Aug. 04, 2020.

[6] Y. Sun, X. Wang, Z. Liu, J. Miller, A. A. Efros, and M. Hardt, “Test-Time Training with Self-Supervision for Generalization under Distribution Shifts,” *arXiv:1909.13231 [cs, stat]*, Jul. 2020, Accessed: Jul. 28, 2020. [Online]. Available: http://arxiv.org/abs/1909.13231.

[7] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, Jun. 2005, vol. 1, pp. 886–893 vol. 1.

[8] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov, “Fast approximate energy minimization with label costs,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2010, pp. 2173–2180.

[9] Jianbo Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[10] Y. Boykov and G. Funka-Lea, “Graph Cuts and Efficient N-D Image Segmentation,” *Int J Comput Vision*, vol. 70, no. 2, pp. 109–131, Nov. 2006.

[11] S. Maji, A. C. Berg, and J. Malik, “Efficient Classification for Additive Kernel SVMs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 66–77, Jan. 2013.

[12] J. E. Iglesias and M. R. Sabuncu, “Multi-atlas segmentation of biomedical images: A survey,” *Medical Image Analysis*, vol. 24, no. 1, pp. 205–219, Aug. 2015.

[13] T. Heimann and H.-P. Meinzer, “Statistical shape models for 3D medical image segmentation: A review,” *Medical Image Analysis*, vol. 13, no. 4, pp. 543–563, Aug. 2009.

[14] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Neural Information Processing Systems*, vol. 25, Jan. 2012.

[15] P. Dollár, R. Appel, S. Belongie, and P. Perona, “Fast Feature Pyramids for Object Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1532–1545, Aug. 2014.

[16] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *MICCAI*, 2015, pp. 234–241.

[17] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, Cham, 2016, pp. 424–432.

[18] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” in *2016 Fourth International Conference on 3D Vision (3DV)*, Oct. 2016, pp. 565–571.

[19] M. I. Todorov *et al.*, “Machine learning analysis of whole mouse brain vasculature,” *Nature Methods*, vol. 17, no. 4, Art. no. 4, Apr. 2020.

[20] B. Ma, S. Liu, Y. Zhi, and Q. Song, “Flow Based Self-supervised Pixel Embedding for Image Segmentation,” Jan. 2019, [Online]. Available: http://arxiv.org/abs/1901.00520.

[21] M. Gutmann and A. Hyvärinen, “Noise-contrastive estimation: A new estimation principle for unnormalized statistical models,” in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Mar. 2010, pp. 297–304, Accessed: Jul. 28, 2020. [Online]. Available: http://proceedings.mlr.press/v9/gutmann10a.html.

[22] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A Simple Framework for Contrastive Learning of Visual Representations,” 2020, [Online]. Available: http://arxiv.org/abs/2002.05709.

[23] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, “Momentum Contrast for Unsupervised Visual Representation Learning,” 2019, [Online]. Available: http://arxiv.org/abs/1911.05722.

[24] Y. Tian, D. Krishnan, and P. Isola, “Contrastive Multiview Coding,” *arXiv:1906.05849 [cs]*, Mar. 2020, Accessed: Jul. 28, 2020. [Online]. Available: http://arxiv.org/abs/1906.05849.

[25] A. van den Oord, Y. Li, and O. Vinyals, “Representation Learning with Contrastive Predictive Coding,” *arXiv:1807.03748 [cs, stat]*, Jan. 2019, Accessed: Jul. 28, 2020. [Online]. Available: http://arxiv.org/abs/1807.03748.

[26] M. Noroozi and P. Favaro, “Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles,” *arXiv:1603.09246 [cs]*, Aug. 2017, Accessed: Jul. 28, 2020. [Online]. Available: http://arxiv.org/abs/1603.09246.

[27] S. Gidaris, P. Singh, and N. Komodakis, “Unsupervised Representation Learning by Predicting Image Rotations,” *arXiv:1803.07728 [cs]*, Mar. 2018, Accessed: Jul. 28, 2020. [Online]. Available: http://arxiv.org/abs/1803.07728.

[28] M. Blendowski, H. Nickisch, and M. P. Heinrich, “How to Learn from Unlabeled Volume Data: Self-supervised 3D Context Feature Learning,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Cham, 2019, pp. 649–657.

[29] S. Carbonnelle and C. De Vleeschouwer, “Layer rotation: a surprisingly simple indicator of generalization in deep networks?,” presented at the ICML, 2019.

[30] M. Mostajabi, M. Maire, and G. Shakhnarovich, “Regularizing Deep Networks by Modeling and Predicting Label Structure,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5629–5638.

[31] O. Oktay *et al.*, “Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 384–395, 2018.

[32] H. Ravishankar, R. B. Venkataramani, S. Thiruvenkadam, and P. Sudhakar, “Learning and Incorporating Shape Models,” *Miccai 2017*, vol. 10433, no. 2, pp. 203–211, 2017.

[33] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[34] L. Vincent and P. Soille, “Watersheds in digital spaces: an efficient algorithm based on immersion simulations,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 6, pp. 583–598, 1991.

[35] H. P. Ng, S. H. Ong, K. W. C. Foong, P. S. Goh, and W. L. Nowinski, “Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm,” in *2006 IEEE Southwest Symposium on Image Analysis and Interpretation*, Mar. 2006, pp. 61–65.

[36] Y. Boykov and M.-P. Jolly, “Interactive Organ Segmentation Using Graph Cuts,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000*, Berlin, Heidelberg, 2000, pp. 276–286.

[37] D. Freedman and Tao Zhang, “Interactive graph cut based segmentation with shape priors,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, Jun. 2005, vol. 1, pp. 755–762 vol. 1.

[38] G. Kerckhofs *et al.*, “Contrast-enhanced microCT to visualize and quantify the 3D vasculature in biological tissues without the need for perfusion,” 20170601, Accessed: Aug. 03, 2020. [Online]. Available: https://lirias.kuleuven.be/retrieve/451736.

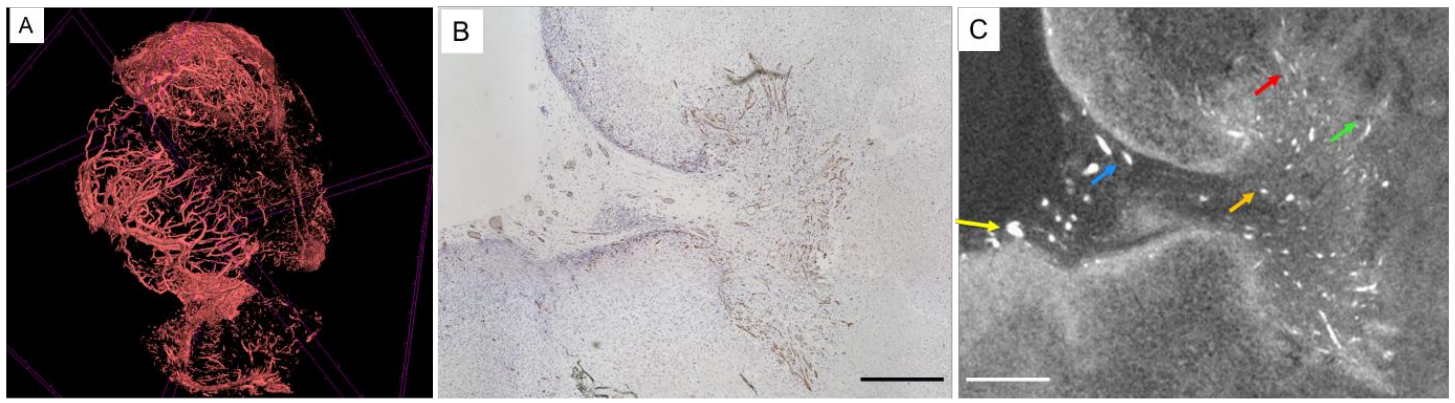


Figure 3: Illustration of contrastive loss for self-supervised learning. (a) Strategy for self-supervised classification. Each image undergoes 2 forms of data-augmentation (only shown for Im1). Once each are passed through a model the features from the same image are grouped, while isolated from other images. (b) Strategy for self-supervised segmentation. The same method as in (a) is used, pixel by pixel, taking zooming and cropping into account.

Figure 2: CE-microCT images of vascularization in a tumour xenograft sample, adapted from the work by Kerckhofs et al. [38] (A) 3D rendering of the vasculature in the xenograft, stained with Hf-WD POM; 3D scale bar = 100 μm. (B) The CD31 stained section and (C) the corresponding CE-microCT cross-section through the tumour xenograft. The brown colour in the histological section indicates CD31 positive blood vessels. The white colour in the CE-microCT image represents red blood cells in the blood vessels. Scale bars = 100 μm.

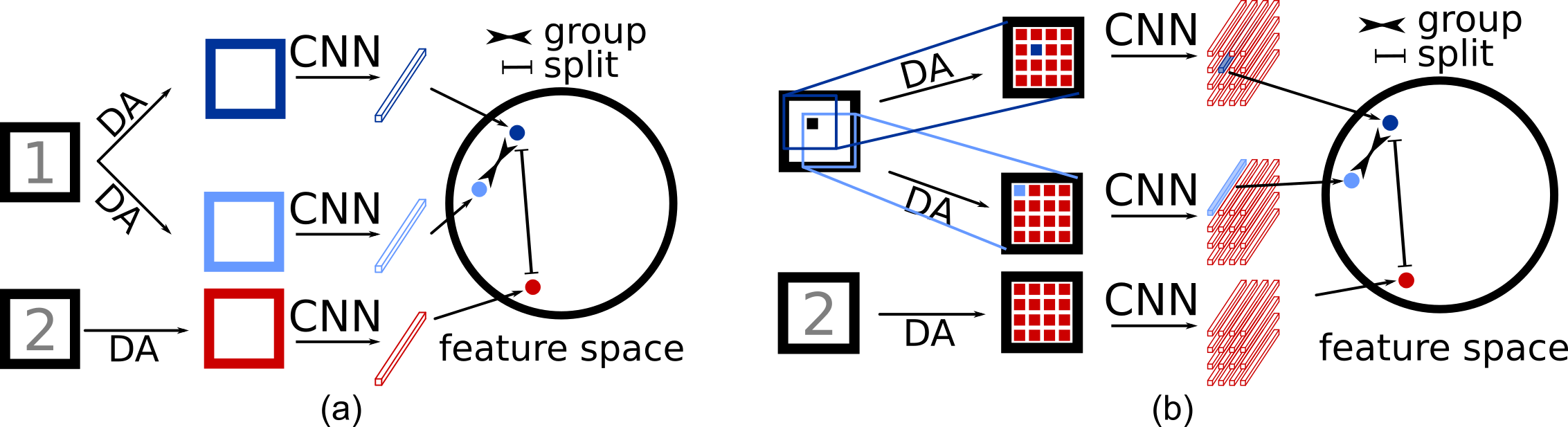
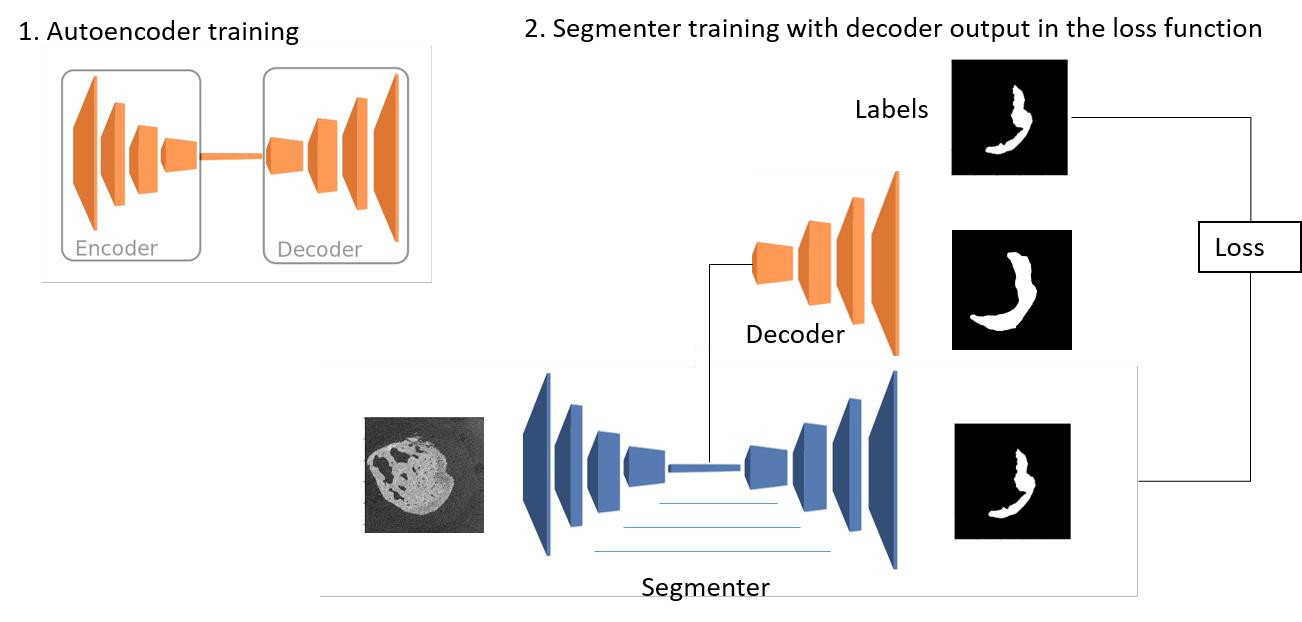


Figure 1: Proposed regularization framework. Step 1: an auto-encoder is trained on the ground truth binary masks. This encodes the internal structure (e.g. classical shape of the regions of interest) at the output of the encoder. Step 2: the trained and fixed decoder is connected to the output of the encoding path of the segmenter. The segmenter is trained by the minimization of a loss function penalizing inconsistencies between the labels and both the segmenter and the decoder outputs. Since the decoder is fixed, this forces the output of the segmenter encoding path to be consistent with the labels internal structure. If it is not the case, the fixed decoder will indeed predict a binary mask inconsistent with the ground truth.



1. **Comments on changes made in the research project in case of resubmission (optional)**

Not Applicable

1. **Activities report****on the first year of doctorate**

**ONLY for “1st grant - 2nd year” applicants**

Not Applicable

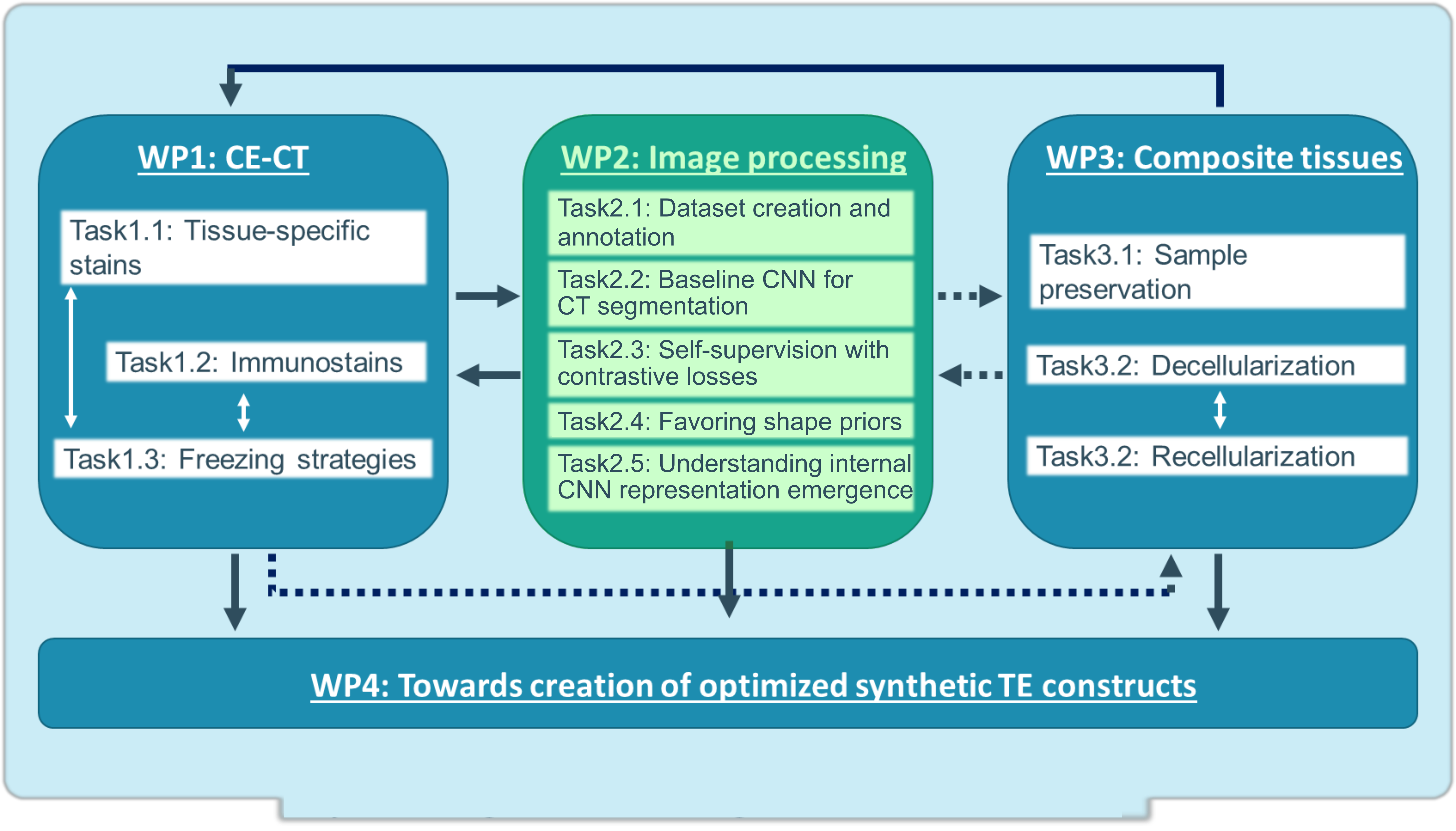
1. **Potential interdisciplinary approach of the research project (optional)**

This project will be done within a 4-year interuniversity project, called Bio-blueprints. This project will not only allow for the creation of a dataset without which this work would not be as thorough, but will also favor collaboration with a large number of researchers in different fields.

On the one hand will this work have a large impact on the creation of different staining protocols within the contrast-enhanced CT field, by validating the type of staining agent that are helpful in the segmentation process.

On the other hand, strong segmentation accuracy will enable advances in the creation of optimized synthetic tissue engineering constructs, by enabling the use of the characteristics of real tissues as stepping-stone.

Figure 4: Schematic overview of the Bio-blueprints organization, as presented by its work packages and their tasks. In green, my proposal and its tasks. The solid arrows indicate the input from one work package to another. The dashed arrows indicate feedback and potential reiteration based on the outcome of a WP or task.



1. **Description of the work environment**

*Please provide the information accounting for the adequacy of the environment (available intellectual and/or material means) to carry out the research as detailed in the submitted project. Please specify the assets of the research environment related to the project and the main publications of the laboratory/promoter (max. 1 page).*

My work is integrated in the activities of the image and signal processing group (ISPG) of UCLouvain within the ICTEAM institute. This research group is composed of Prof. Benoît Macq (Fellow IEEE), Prof. Laurent Jacques, Prof. Christophe De Vleeschouwer and about forty researchers. Their work is internationally appreciated through numerous scientific publications, involvements in Walloon and European collaborative projects and the creation of spin-off companies such as Keemotion, Telemis, Intuitim, ACIC, or Alterface. The seniors of the lab are editors or associated editors of major journals (IEEE or Elsevier).

Several doctoral or post-doctoral researchers of the group are working on topics related to this project. Their work include deep learning for adaptive protontherapy, segmentation of cells in microscopy and CT images, but also to more fundamental questions related to CNN generalization [Simon] and design [ordinal pooling].

Moreover I will be in close contact with the ContrasT team, under the supervision of Prof. Greet Kerckhofs. The ContrasT team has a highly interdisciplinary expertise (biology, chemistry, engineering, imaging), which will enable interaction and novel ideas in a very diverse setting. In addition, our lab has existing and future collaborations with Saint-Luc. Both partnerships will also enable the acquisition of extensive datasets, needed for self-supervision without annotations.

Our lab has access to several multi-gpu servers, including:

* 4 servers with 4-6 RTX2080ti 11Gb GPUs each
* 2 servers with 2 V100 32Gb GPUs each.

We can expect this number to grow as the need for powerful GPUs grows, both in our team and through collaboration with the CISM (“Calcul Intensif et Stockage de Masse”).

1. **Summary of the master’s thesis or equivalent**

*Please provide a summary of your master’s thesis or any equivalent, even if you have not been graduated yet (max. 1 page).*

My master’s thesis consisted of optimizing the speed of convolutional neural networks (CNNs) while running on programmable hardware (FPGA). This work stood on the edge between two domains : image processing, using deep learning for image classification and model optimization, and electronics with the optimization of convolutional networks for FPGA devices.

The core of the work concentrated on the one hand on building models for image classiﬁcation which took advantage of the speciﬁcs of FPGA devices, while being built from a generic CNN model, and on the other hand on the acceleration of convolutional layers. It should enable someone to take a model deﬁnition and retrain the model to be optimal for an FPGA implementation. To show that this is achievable, a standard residual network (ResNet) has been used, quantized and assembled from uniform building-blocks to allow better area utilization on the FPGA. The goal was to achieve no worse than 3% accuracy loss on the CIFAR-10 classiﬁcation task, compared to a ﬂoating-point model, while introducing highly quantized weights and activations.

Initially, the focus was put on training state-of-the-art (SOTA) residual networks with diﬀerent quantization approaches. Models with weight quantizations from 1 bit to 4 bits, and activation quantizations from 1 bit to 32 bits were developed. It was found by analyzing diﬀerent quantization strategies that, in general, quantization is a viable approach for neural networks, and in particular ternary networks (constrained to values -1, 0, 1) can achieve near state-of-the-art performance on the CIFAR-10 dataset. Secondly, the ResNet was modiﬁed to accommodate an easier design on FPGA. The network was restructured to only contain blocks with a convolution, a batch- normalization (BN), and an activation layer. All BN layers were changed to shift-based batch-normalization, to use fewer logic elements.

Like this project, my master’s thesis was both novel and timely, with multiple papers coming out during and after the work.

1. **Additional comments (optional)**

7.1 Ethical concerns

While this project uses biomedical images, with CT scans either on humans or mice, this project does not require to change standard treatment protocols, and will only use data from already planned imaging processes. Medical data will be anonymized.

7.2 Past year as a research assistant

During the past year I had the pleasure to work as research assistant in a project financed by Innoviris. The project, in partnership with a Brussels start-up working in the field of elderly care called “Kaspard”, consisted of developing Deep Learning tools for the automatic configuration of their fall detection device. This work has enabled me to familiarize myself more deeply with the domain of convolutional networks, and in particular segmentation models, which will be a cornerstone of my future research project. While at first glance the Kaspard project shows little similarity with biomedical imaging, there are some important similarities :

* an initial lack of annotated data, which led to a first foray in self-supervised methods.
* non-natural image data, due to the use of time-of-flight (ToF) sensors, which made the use of pre-trained networks infeasible, as they did not generalize to the available data
* Access to 2.5D/3D data. While CT scans are fully 3D, ToF data adds depth to “classical” intensity images. The use and transformation of this depth data was an integral part in producing a successful solution for the project.

In addition to this, some of the work performed for this project will be useful as a foundation for the investigation of my research questions (e.g. CNN segmentation implementation, design of annotation tools).

1. **Ph.D. work calendar per month**

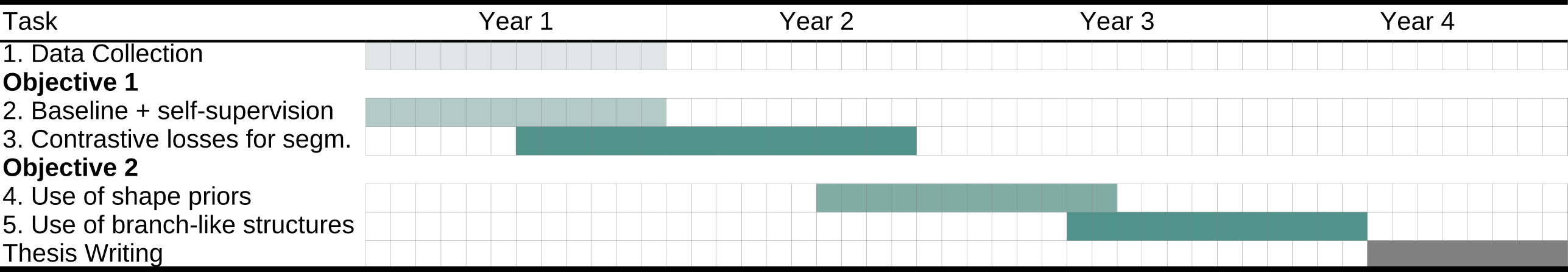


Table 1: Gantt diagram, Publication are expected at the end of each sub-task. Darkness of color represents advances compared to state-of-the-art.

Table 1 shows a tentative work calendar for this research project. Section 1.4 provides a thorough explanation of every task and their interactions. The first year of this project will be heavily focused on data collection and the creation of a baseline solution. Data collection consists of image annotation, analysis, and collaboration on the definition of segmentation needs and success metrics. In addition to this, we will develop several strategies for contrastive losses and their implementations. Once this endeavor is underway, we will attempt to use CNN regularization in order to utilize shape priors inherent in biomedical data. The gain will be measured both from scratch and in addition to the use of contrastive losses. This will bring us naturally to the use of the specificities of the segmentation of vascular and nervous systems.

Publications are expected for the different tasks apart from data collection. We will focus on high profile conferences (ICCV, CVPR, ECCV, MICCAI, ICIP, and BMVC), which have the clout of journal publications in terms of peer review. We expect to extend task 3 and task 5 to journal submissions for more long-form papers.

Table 2 shows the different risks associated with the project, and their possible mitigation.

|  |  |  |
| --- | --- | --- |
| Risk assessment & mitigation | Probability | Impact in case of failure |
| **Risk:** The CNN training does not converge or generalizes poorly  **Mitigation:** Previous research on baseline 3D U-Net with very limited amount of data has provided (perfectible but) encouraging results. Additionally, a significant part of the work plan aims at shaping the outputs and/or internal representations, so as to mitigate the black-box nature of deep learning. | Low | High |
| **Risk:** Lack of annotated data  **Mitigation:** Use of multiple datasets, including existing 3D CT medical scans available at UCLouvain. | Low | High |
| **Risk:** The methods fail in providing sufficiently accurate outputs  **Mitigation:** 3 different and interacting research tasks are considered, which reduces the impact of one method failing to provide decent results. Different datasets will also help finding contexts where each method can return meaningful results. | Medium | Medium |

Table 2 : Risk assessment and mitigation.