Master thesis topic (Masterproefonderwerp)	Multimodal image registration using deep reinforcement learning
Promoters (Promotoren)	prof. dr. ir. Aleksandra Pižurica – Department of Telecommunications and Information Processing
Mentors (Begeleiders)	Srđan Lazendić – Department of Electronics and Information Systems
Contact person (Contactpersoon)	Srđan Lazendić
Number of master theses (Aantal masterproeven)	1
Number of students (Aantal studenten)	1
Academic year (Academiejaar)	2021-2022
Keywords (Trefwoorden)	Deep learning, Convolutional neural networks, CNN, Image processing, Reinforcement learning, Deep reinforcement learning
Location – (Locatie)	Technicum, De Sterre, thuis

References (Referenties):

1. C. N. Matcha. *A 2021 guide to semantic segmentation*, 2021, https://nanonets.com/blog/semantic-image-segmentation-2020/
2. J. Hu, Z. Luo, X. Wang, S. Sun. *End-to-end multimodal image registration via reinforcement learning*, 2021.
3. Casanova, P. O. Pinheiro, N. Rostamzadeh, C. J. Pal. *Reinforced active learning for image segmentation*, 2020.
4. F. Ryosuke, N. Inoue, T. Yamasaki. *PixelRL: Fully convolutional network with reinforcement learning for image processing*, 2019.

Description of the problem (Probleemstelling)	Image registration is the process of aligning two or more images of the same object taken from different viewpoints, at different times or aggregated from multiple camera sources. In order to process the given data, it is often necessary to first register these images, i.e., to transform the images such that the content is aligned, and image processing techniques can simultaneously access the spatial information of the images. To diagnose their patients, medical experts daily use complementary information they extract from CT scans, MRI scans, etc. Hence, it is clear that medical image analysis is a task that benefits from multimodal data (Fig. 2). Recently, multimodal imaging is routinely employed in order to support the technical study of artwork and its conservation. The data obtained are images aggregated from multiple camera sources and are operating on different wavelengths (e.g., visible, infrared, X-ray) - see Fig. 1. For the analysis and restoration of paintings, multimodal data make it possible to access important information about the different layers of paint, without having to damage the painting. That way we can easily access the spatial and spectral information of the different modalities simultaneously. With an increased interest for image processing applications for art investigation (e.g., detection of damaged regions, virtual restoration, analyses of brush strokes), it would be advantageous to have access to the fully aligned multimodal data before processing. In other words, we need to perform image registration. Reinforcement learning is the training of machine learning models to make a sequence of decisions. The agent learns how to achieve a goal in a complex uncertain environment. While reinforcement learning is dynamically learning with a trial and error method to maximize the outcome, deep reinforcement learning is learning from existing knowledge and applying it to a new data set. Deep reinforcement learning (DRL) brought significant performance improvements in many areas including games, robotics and computer vision. Recently, DRL started to appear also in image processing applications such as image classification, face hallucination and face recognition, active object localization and visual object tracking, image segmentation, image enhancement, image recovery etc. Nonetheless, the applications of deep reinforcement learning for image processing are still limited and should be better investigated. In a recent article [3] authors proposed a novel method for a semantic (pixel based) segmentation based on deep reinforcement learning. An agent learns a policy to select a subset of small informative image regions – opposed to entire images – to be labeled, from a pool of unlabeled data (see Fig. 2). Motivated by the encouraging results obtained in [3], we want to extend this idea to solve the problem of multimodal image registration.

Goal of the thesis (Doelstelling)	The goal of this master thesis is to develop registration techniques that fulfil the requirements of typical art investigation or medical image processing tasks. The crucial components of a conventional image registration method are the choice of the right feature representations and similarity measures. These two components, although elaborately designed, are somewhat handcrafted using human knowledge. To avoid this, these two components could be tackled via reinforcement learning. Specifically, the student should build and train an artificial agent, which is composed of a combined policy and value network, in order to adjust the moving image toward the right direction. The student will start with a study on current state-of-the-art registration techniques, followed by extending these techniques with respect to the properties of the given dataset. Different modalities have different characteristics, so we can optimize by using DRL which features of the different modalities can be useful for registration. Moreover, images that need to be registered often have different spatial or/and gray level resolution, contain noise, they are blurred due to the movement of objects and/or cameras etc. Fortunately, the spatial resolution can be rescaled. If one modality has a significant lower resolution, one can make use of the other modalities to also include super-resolution technique for upscaling the image. Changing the spatial resolution makes it easier to separate and match certain features. At the beginning of the semester, a minicourse will be organized for a student to put him/her on the right track and to familiarize him/her with the topic. Furthermore, the existing code and literature will be made available to the student.

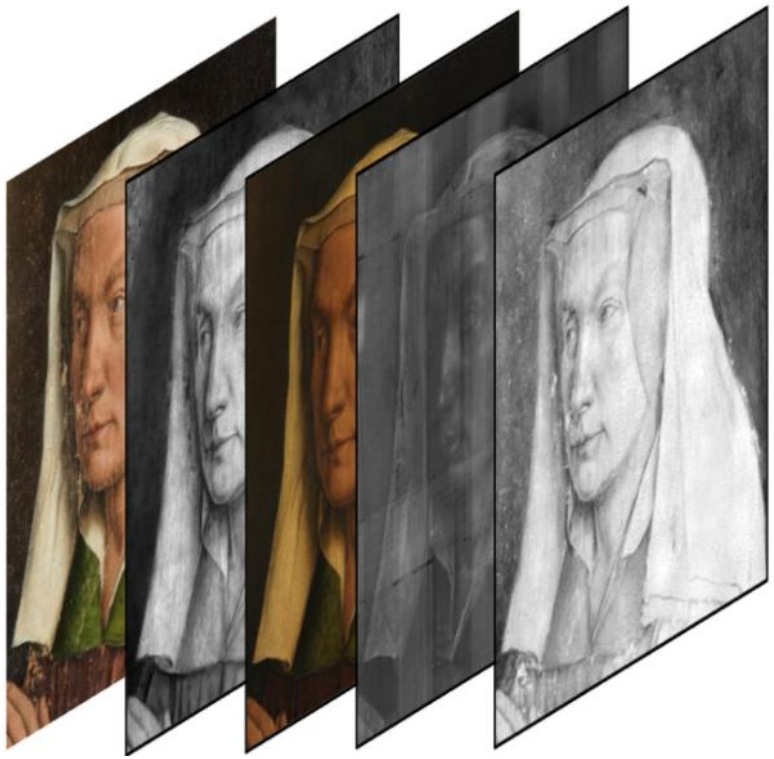

Figure 1: Multimodal image acquisitions.
Image copyright: Ghent, Kathedrale Kerkfabriek, Lukasweb; photo courtesy of KIK-IRPA, Brussels.

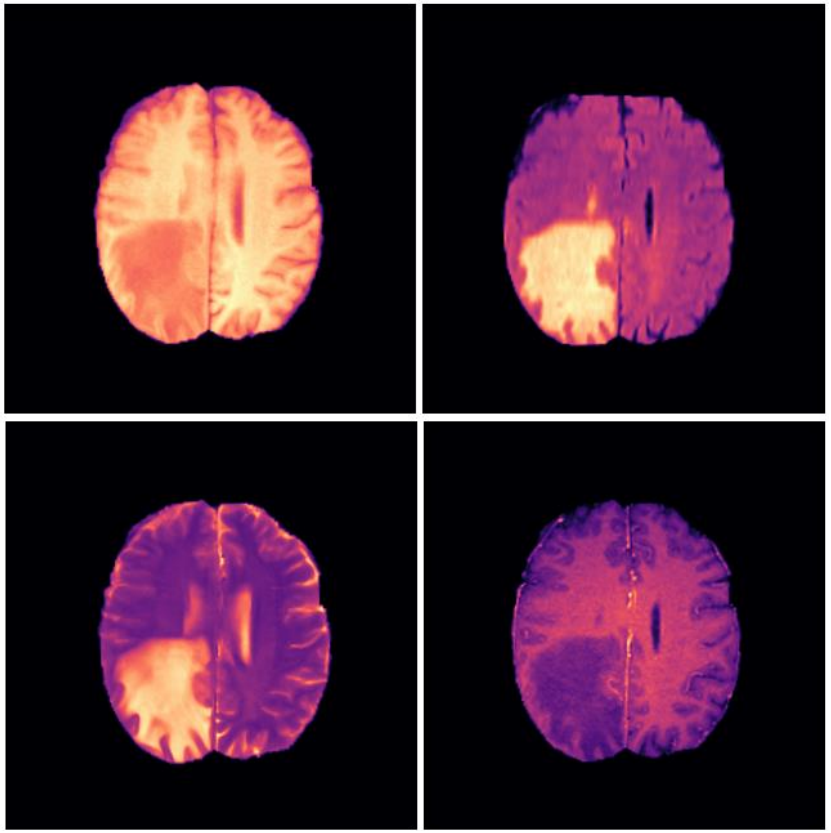

Figure 2: Four modalities of BraTS 2020 dataset

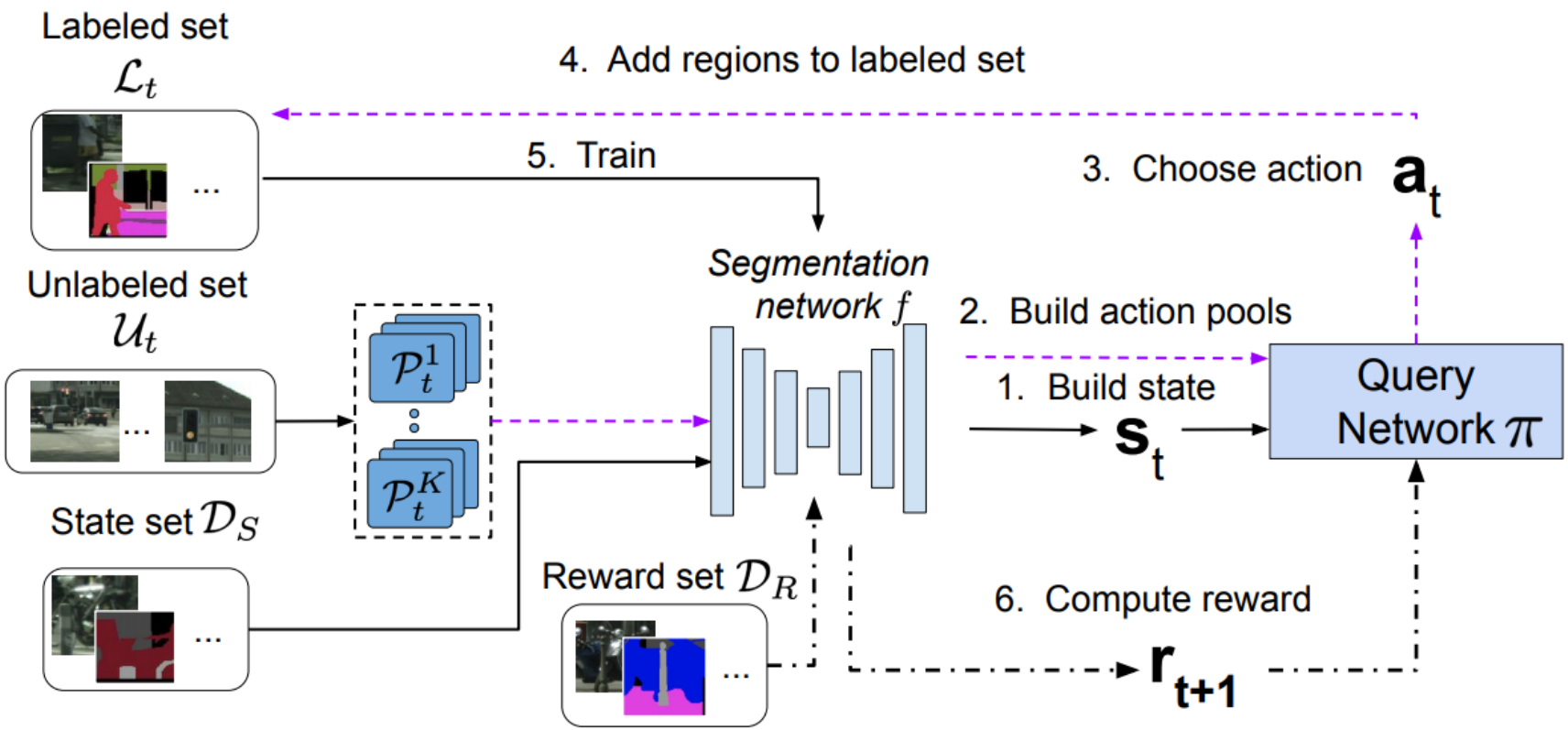

Figure 3: Reinforced active learning setting for semantic image segmentation (see [3]).