A Study on Automatic Intensity-Modulated Radiation Therapy for Nasopharyngeal Carcinoma Using Deep Learning

ビレル, ダウド

A Study on Automatic Intensity-Modulated Radiation Therapy for Nasopharyngeal Carcinoma Using Deep Learning

Bilel Daoud

Graduate School of Information Science and Electrical Engineering
at Kyushu University

December 2020

Abstract

Nasopharyngeal carcinoma(NPC) arises from epithelial cells that cover the surface and line the nasopharynx. One of the treatments for NPC is an intensity modulated radiation therapy (IMRT) which is one type of radiation therapies. Generally, the radiation therapy is to destroy or damage cancer cells by irradiating X-rays. Unlike the general radiation therapy, the advantage of IMRT is that radiation oncologists can control the intensity of the radiation beam according to the size, shape and location of a target tumor. This control allows IMRT to destroy tumors with complex shapes by delivering more focused radiation to the tumor or specific areas within the tumor. At the same time, IMRT can avoid the exposure of healthy tissues by reducing the radiation dose to the healthy tissues.

However, the current IMRT treatment requires the radiation oncologists the three manual tasks: 1) detection of the NPC and organs-at-risks (OARs) regions; 2) determination of the treatment planing; 3) re-planning the current treatment during the course of radiation therapy (RT). The three tasks are difficult and time-consuming for the radiation oncologists.

In this thesis, I propose three fundamental techniques to construct an automatic system for IMRT. Firstly, I develop a new CNN-based method for segmenting NPC from CT images. The proposed method is based on the concept of cascade classifiers. The first phase in the proposed method is to eliminate from a given CT image the regions which the NPC never invades. In the second phase, I detect NPC regions from the remaining regions in the CT image. Secondly, I propose a new method for predicting the dose distribution images of NPC from contoured CT images that contain the contours and regions of tumors and OARs. The proposed method consists of

two phases: 1) extracting the features of each beam from contoured CT images of a patient, and 2) predicting the dose distribution images by integrating the obtained features of all beams. Finally, I present a new system, called Tumor Evolution Prediction (TEP-Net), for predicting the future evolution of NPC and OARs from three orthogonal CT images. From the experiments, the three proposed systems achieve the best performance compared with their conventional methods.

Acknowledgements

The work reported in this thesis was carried out in Morooka Laboratory, Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan, during the years 2017-2020. Undertaking this PhD has been a truly life-changing experience for me and it would not have been possible to do without the support and guidance that I received from many people. Primarily, I would like to express my special appreciation and thanks to my mentor Professor Ken'ichi Morooka for all the support and encouragement he gave me, during my thesis. Without his guidance and constant feedback this PhD would not have been achievable. Also, I am grateful to my committee members, Professor Ryo Kurazume, Professor Hidetaka Arimura and Professor Ryoma Bise for giving me valuable advices and comments about my work. Moreover, I deeply appreciate Doctor Jamel Daoud, Professor Leila Farhat, Doctor Wafa Mnejja and Doctor Tarek Sahnoun at Radiotherapy Department of Habib Bourguiba Hospital, Sfax, Tunisia for providing medical data, medical knowledge, and valuable feedback to my work. Next, I would like to extend my gratitude to my dear friend, Assistant Professor Shoko Miyauchi at Kyushu University, for all her venerable guidance in my PhD study and in my life in Japan. Also, this PhD study would not have been possible without the supporting of Otsuka foundation Scholarship on 2017 and JIUN corporation from 2018 to 2020. I am especially grateful to them for their support to my life and my research. I also would like to thank all members in Morooka and Kurazume Laboratory. Finally, I am indebted to all my family and friends in Tunisia and Japan for their support and encouragement with this work.

Contents

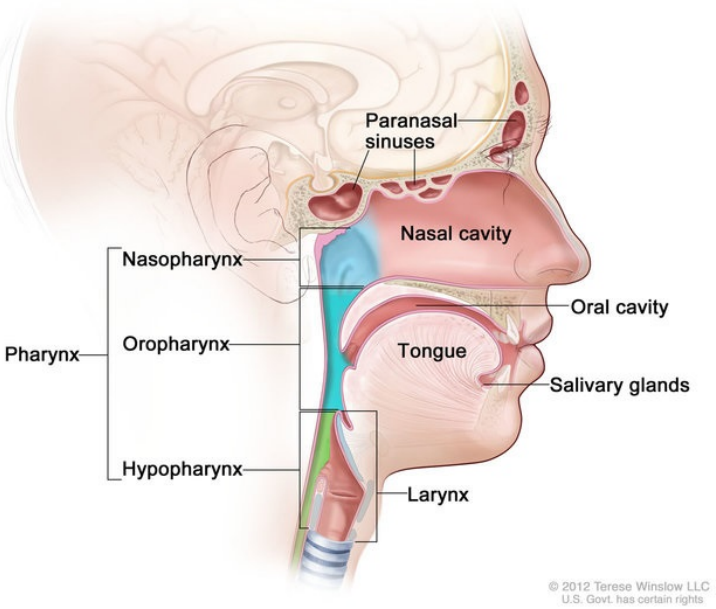
1

Introduction

1.1 Nasopharyngeal carcinoma

Cancer describes a group of diseases caused by abnormal cell growth in a part of the body such as breast, prostate and cervical. There are 100 different type of cancers. According to the world health organization (WHO)[1], cancer is the second leading of death in the world. Practically, there were 9.6 million deaths from cancer worldwide in 2018.

Among them, head and neck cancers (HNCs) are solid tumors located in head and neck areas including larynx, throat, lips, mouth, nose, and salivary glands. HNC is the sixth leading cancer by incidence worldwide with more than 550,000 cases and around 300,000 deaths every year[2]. According to the location where the NHC originates, there are seven categories of the NHCs: hypopharynx, larynx, lip and oral cavity, nasopharynx, oropharynx, paranasal sinus and nasal cavity, and salivary gland (Figure1.1).



Source: https://www.teresewinslow.com/head

Figure 1.1: Head and neck anatomy[3].

Nasopharyngeal carcinoma(NPC), one of the seven HNC, arises from epithelial cells that cover the surface and line the nasopharynx[4]. NPC is a rare disease and ranks as the 24^{th} most frequent cancer in the world. However, in Southwest Asia, Southern part of China and Northern Africa, the annual NPC incidence and mortality are higher. It was reported that in 2018, there were 129,079 new NPC cases diagnosed and 72,987 cancer deaths worldwide[5]. The factors, which causes the NPC, include food, tobacco, heavy alcohol use, human papillomavirus and Epstein-Barr Virus(EBV) [6; 7; 8]. Here, EBV is a virus in the mouth and pharynx.

According to the TNM Classification of Malignant Tumors of the Union of International Cancer Control(UICC)[9], there are four stages of NPC, T1, T2, T3 and T4, depending on the extension of the tumor in head and neck regions or in other regions of the body. In the earliest stage T1, the NPC is found only in the nasopharynx. In stage T2, the NPC grows into the soft tissue of the middle throat. Moreover, the NPC with stage 2 have spread to lymph nodes on one side of the neck. The size of the lymph nodes is less than 6 [cm]. In advanced stage T3, the NPC invades the bone structure of the skull base while the NPC in stage 4 extends inside the head to, cranial nerves, the eye or its nearby tissues. In the case of NPC patients in the early stages (T1and T2),

the overall 5-year survival rates of NPC are 72% for stage T1 and 64% for stage T2. However, in the advanced stage (T3 and T4), the survival rate drops significantly to 62% for stage T3 and 38% for stage T4. Therefore, detecting and treating the patients with early stage NPC play an important role for destroying the tumors completely[10]. However, the majority of NPC patients are only diagnosed when the tumor has reached advanced stages[11].

According to the spread of the tumor, several symptoms appear in NPC patients such as trouble breathing, speaking, or hearing, pain or ringing in ears. If a patient has these symptoms, in order to determine whether a patient has NPC, generally, radiation oncologists apply a biopsy test which is used to know if there are malignant cells in a small sample of tissue taken from the nasopharynx tissue of the patient. Moreover, to confirm the status of the patient, additional diagnostic tests are available such as physical examination, endoscopic examination, pathological diagnosis, general blood analyses, EBV-DNA testing, nutritional status assessment and medical imaging. From the testing results, the oncologists diagnose NPC, and its stage according to its spread that is called staging[12; 13].

Currently, among the NPC diagnostic tests, the medical imaging test plays a crucial role for therapy and diagnosis of the NPC patients. The performance improvement of medical imaging devices, including cranial-cervical computed tomography (CT) scan and magnetic resonance (MR) scan, enables to obtain high-resolution medical images. From CT images, radiation oncologist determine the extension of the NPC from the paranasopharyngeal extension, the bony structures and the presence of cervical lymph nodes in the CT images. On the contrary, MR images are useful to specify a better assessment of the primary tumor location, intracranial structures and retro-pharyngeal spaces.

Using medical images of a patient, radiation oncologists identify NPC by manually drawing the contours of the tumor and its neighboring organs, called organ-at-risks (OARs), on the images (Figure 1.2). Once the size, location and shape of the NPC are known, the radiation oncologist determine the stage of the NPC to choose the appropriate treatment from three ways: radiation therapy (RT), chemotherapy and surgery. RT is used as the primary and only curative treatment for NPC. NPC patients with stages 2, 3 and 4 get chemotherapy during 1 month in order to stop the spread of the tumor. If the patients still have the lymph nodes after the treatment, the surgery is done to remove the lymph nodes from the patient.

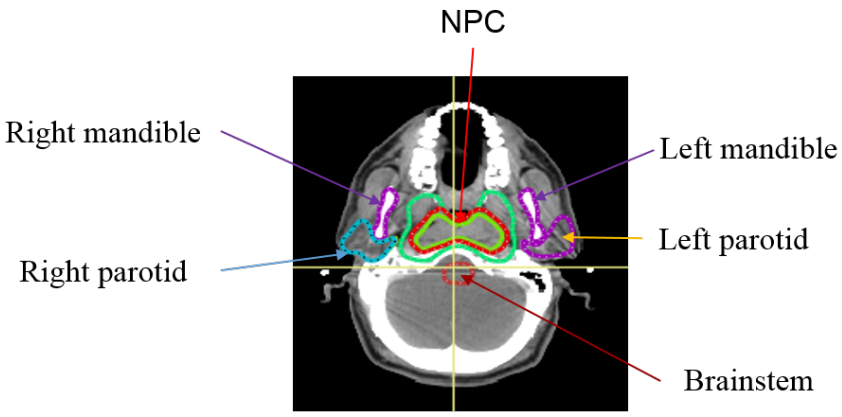
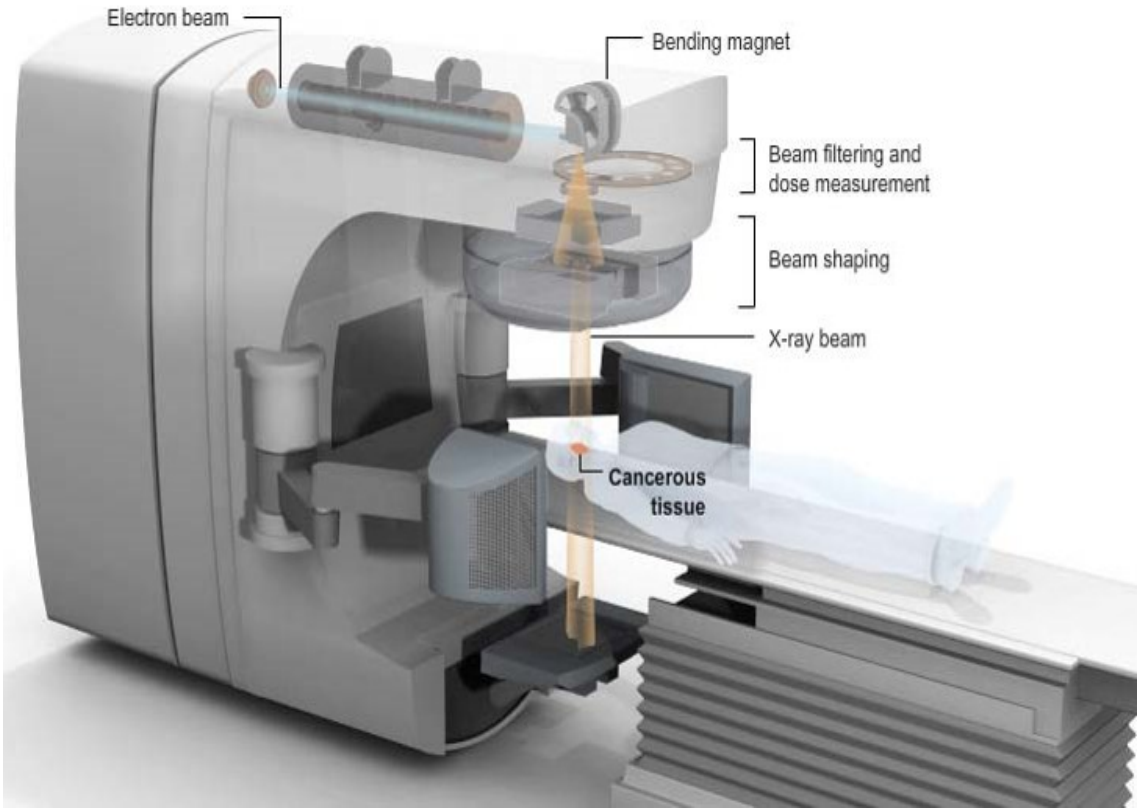


Figure 1.2: CT images with NPC and non target organs.

One main approach of RT for NPC is an intensity modulated radiation therapy (IMRT)[4] because of its effective NPC treatment (Figure 1.3). To irradiate the tumor cells, IMRT uses multiple high energy X-rays (4-20 MV) from multiple beam directions. Here, since the beam damages not only cancerous but also healthy tissues, the prescribed dose of radiation (70 Gy) is usually delivered to a target tumor for 33 days with 2.12 Gy per day to minimize the effect of damaging the healthy tissues. The advantage of IMRT is that radiation oncologists can control the intensity of the radiation beam according to the size, shape and location of a target tumor. This control allows IMRT to destroy tumors with complex shapes by delivering more focused radiation to the tumor or specific areas within the tumor. At the same time, IMRT can avoid the exposure of healthy tissues by reducing the radiation dose to the healthy tissues. Owing to these benefits, compared with traditional RT, IMRT contributes to the improvement of the Quality-of-Life (QOL) for NPC patients and is recommended for over 50% of all cancer to eradicate cancer cells[15].

1.2 Problems in IMRT

Although IMRT has a great advantage described above, the current IMRT treatment requires the radiation oncologists the three manual tasks which are difficult and time-consuming.

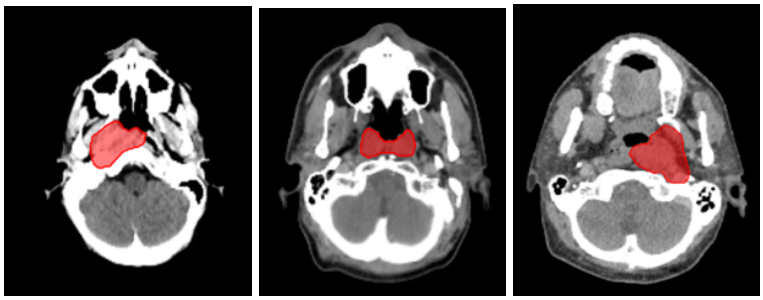


Source: https://danwin.com/tag/imrt

Figure 1.3: System for intensity modulated radiation therapy (IMRT)[14].

1.2.1 NPC segmentation from medical images

The task of segmenting NPC from medical images is a quite challenging for radiation oncologists because of the following reasons. In general, NPC develops in various shapes, sizes and locations in the body of the patient[16]. Figure1.4 shows the examples of NPC regions in three different patients. As shown in these figures, the NPC regions appear at various locations. Moreover, in the medical images, the intensity of the NPC is similar to those of its neighboring organs and tissues. Therefore, the accuracy and time of detecting the NPC tumors from medical images highly depend on the knowledge and experience of the radiation oncologists. Practically, it has been reported in [17] that it takes around 2.7 hours on average to manually draw NPCs for a single head and neck case.

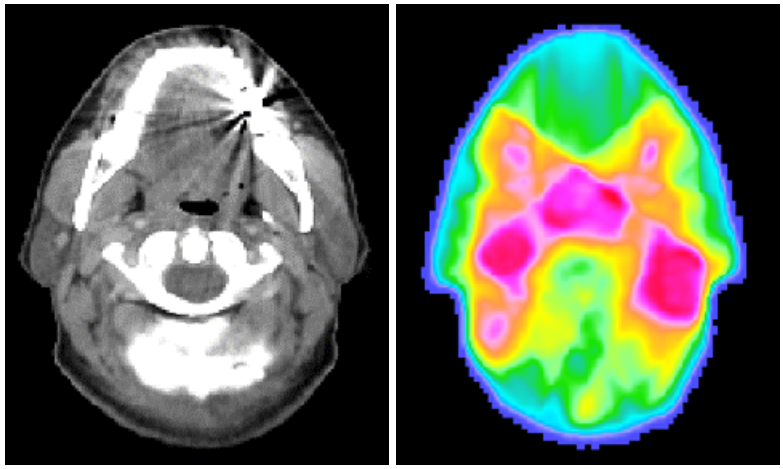


(a) Patient 1 (b) Patient 2 (c) Patient 3

Figure 1.4: Examples of NPC regions (red regions) in three different patients with red color.

1.2.2 Determination of dose distribution

When applying IMRT, radiation oncologists plan a IMRT treatment by using treatment planning systems which optimize manually the treatment by adjusting all beam parameters, such as the number of beams, beam directions, shapes, weights, OARs, etc., and calculate dose distribution[18] (Figure 1.5). Here, the dose distribution is a 2D image that visualizes the total dose delivered to the patient from all beams. The value of each pixel in the dose distribution is calculated based on all beam doses and the anatomical structure of the patient body. However, the dose distribution provided by the planning system is not always approved by radiation oncologists. IMRT delivers a high radiation to the tumor and reduce the radiation dose to the healthy tissues. However, since the NPC in the advanced stage spreads into other organs and has complex shapes, these makes the planning systems difficult to estimate the suitable dose distribution considering the characteristics of the advanced NPC. Therefore, the radiation oncologists need to tune the parameters repeatedly to obtain a desired dose distribution. This tuning process is time-consuming, from several hours to a week. Moreover, the quality of the determined treatment plan highly depends on the knowledge and experiences of the radiation oncologists[19]. Figure1.6 shows the process that the treatment planning system calculates the dose distribution of a patient. Here, the dose constraints means the range of the delivered dose to OARs and tumor regions without any damages used to optimize the beam dose. In the current treatment process, firstly, the beams dose is determinate from CT axial image, NPC, OARs regions and the dose constraints by using an optimization system. However, the optimization system does not always pro-



(a) CT image (b) Dose distribution image

Figure 1.5: An example of (Right) the planned dose distribution image from (Left) its original CT image.

vide good beams dose. Therefore, radiation oncologists need to modify the parameters repeatedly. Once the beams dose is approved, the dose distribution is calculated by the sum of all beams dose.

1.2.3 IMRT based on the condition of a patient before starting the therapy

In the current IMRT, once the dose distribution is calculated, the initial plan is determined, and not changed during 33 days. On the contrary, during the 33 days, the anatomical structures of the patient have changed such as the shrinkage of the tumor by IMRT, the weight loss, and the displacement or sizes of the normal structures. Wang et al.[20] prove that after three weeks of RT, a significant variation in a patient body volume was found in with an average decrease of 29% for the NPC. The anatomical change means the difference between the initial planned dose and the optimal dose for the actual patient structure. If the difference is large, the IMRT may cause the overdose to OARs, the risk of the exposure to healthy organs and a second chance for recurrent patients with less unrecoverable tissue damage. These factors introduce the side effects which influence on the patient QOL such as xerostomia and myelitis[21]. Therefore, the re-planning procedure offers better protection for the OARs.

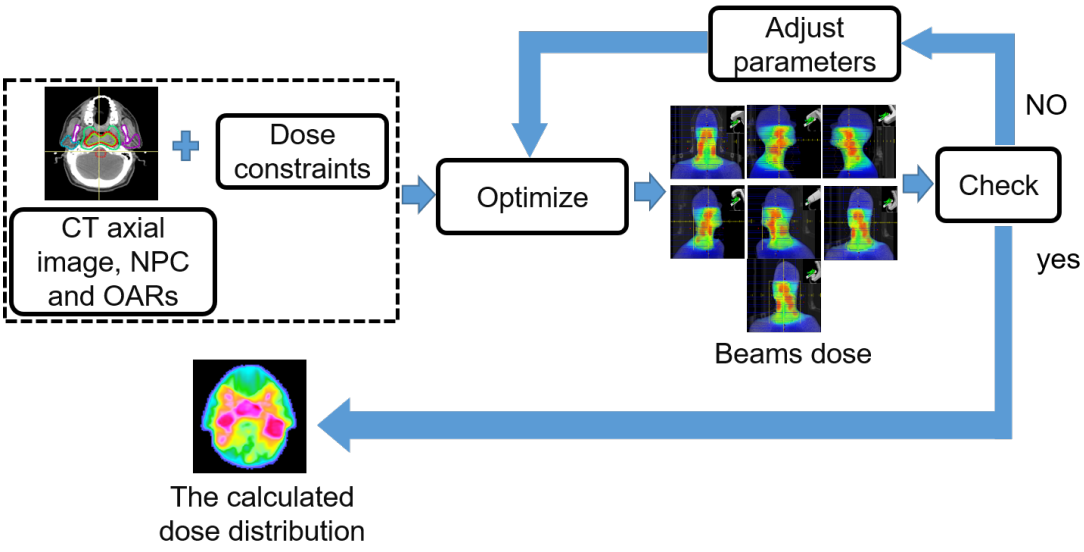


Figure 1.6: The process that the treatment planning system calculates the dose distribution of a patient.

As one approach for this problem, the initial planning in the pre-treatment should be modified according to the current patient condition such as the tumor response and the anatomical changes of OARs during the course of RT. This approach is called adaptive radiation therapy (ART). Firstly, the ART needs new CT images of the patient after receiving the specific dose determined by the radiation oncologists. In the previous studies in ART, different specific doses from the initial planning are introduced to modify the current treatment. Moreover, the timing, when the treatment planning is changed, is also important for ART. Here, the best timing for re-planning the treatment plan is the time when the planned dose distribution can be delivered faithfully to the patient. From the report by Heukelom et al.[22], if the re-planning is made at not the best timing, the 2-year survival rates of several patients were low. However, the anatomical changes of NPC and OARs are observed every day or within a few days between week 4 and 6. Therefore, the determination of the best timing for ART is the hard task for radiation oncologist and still under debate[23]. For this problem, the report in [24] suggested the remedy if the future shrinkage of NPC and OARs can be predicted previously, re-planning the current treatment should be made during the RT course.

Through the current workflow of radiation oncologists, three processes, that is, detection of NPC and OARs, determining the treatment planing and re-planning the

current treatment during ART, are time consuming and quite challenging as mentioned above. One remedy for these problems is to automate the treatment planning system for NPC to accelerate the workflow of radiation oncologist and to improve the accuracy and efficiency of IMRT, which is the final goal of this thesis.

1.3 Related works

1.3.1 NPC segmentation

In the beginning, researchers have developed NPC segmentation methods by using simple image processing [25; 26; 27; 28]. The conventional NPC segmentation methods are classified into two categories: region-based([25],[26]) and contour-based methods([27],[28]).

The region-based methods employ region growing methods to detect NPC regions by growing initial seeds. Therefore, the accuracy of detecting NPC regions depends on the positions of the initial seeds. In order to determine the initial seeds, tatanun et al.[25] and chanapai et al.[26] introduced a probability map and a self-organizing map, respectively. However, the methods in [25] and [26] detect NPC with a low performance. NPC and other regions in CT images have a similar contrast. In other words, the contours of the NPC are unclear. In this case, generally, a region growing method can not find such unclear contour accurately.

The contour-based methods([27],[28]) segment the NPC region by evolving the shape of an active contour. Here, as mentioned in Chapter1, NPC develops in various shapes, sizes and locations. However, generally, it is difficult to fit the active contour to the contours with complex shapes, especially, concave shapes. Accordingly, it is difficult to automatically detect NPC using simple image processing techniques introduced in [25], [26], [27], and [28].

Recently, deep learning methods using convolution neural networks(CNNs) have achieved a great success in medical image analysis([29],[30]). Especially, CNNs become a powerful tool for segmenting organs and tumors such as skin cancer[31], liver[32], nuclei[33], brain[34] and prostate gland[35]. Therefore, several NPC segmentation methods using CNNs have been developed([36],[37],[38]).

Men et al.[36] constructed a deep deconvolutional neural network(DDNN) composed of two networks. The first network has a VGG-16-based architecture to extract

features of NPC from whole CT images. The second network outputs the segmented NPC regions from the extracted features obtained in the first phase. The method in [36] directly segments the NPC. However, NPC and its neighboring regions have similar intensities. This makes it difficult to segment the NPC from the CT images. Wang et al.[37] presented the CNN for NPC detection by using small overlapping patches obtained from axial CT images. The proposed method is tested on four patients to evaluate the detection performance. However, considering the high variety of the sizes, shapes and locations of NPC, the experiments using such small number of patient images provide no guarantee that the method in [37] is applicable to detect NPC. Ma et al.[38] introduced a method composed of two-stages to detect the NPC. In the first stage, CNN is applied to roughly segment NPC from multiple orthogonal CT images. The second stage is to refine the segmented NPC regions by using a 3D graph cut. The method in [38] achieved a high performance detection of NPC when the intensity difference between NPC and its surroundings is larger. However, NPC has a low contrast and a small difference to its surroundings. Such segmentation of the NPC by the graph cut method a difficult task. Although one solution for this is to tune the parameters in the graph cut, the parameter tuning does not always lead to obtain the reliable segmentation.

Unlike the CNN-based methods for detecting NPC regions ([36],[37] and [38]), Ibragimov et al.[39] introduced CNN to segment non-target organs from CT images of NPC cases. Here, non-target regions are OARs regions which NPC never invades. However, the method in [39] does not consider the detection of NPC regions.

1.3.2 Dose distribution prediction

Recent researches have been developing IMRT support systems which predict the dose distribution. These methods can be categorized into non-deep learning and deep-learning methods.

In non-deep learning methods, objective-based planning (OBP)[40] determines desirable IMRT treatment by using pre-set objectives of dose. As another IMRT planning method, knowledge-based planning (KBP)[41] uses a collection of previous plans of treated patients to predict the dose distribution of a new patient. However, the non-deep learning methods obtain unacceptable dose distributions so that the use of the dose distribution may damage some OARs. Therefore, radiation oncologists need to manually tune the parameters several times to obtain a good dose distribution with

less damage such as spatial information of OARs and tumors, number of beams and distance to the tumor.

On the contrary, the deep-learning based methods use deep neural networks to predict the dose distribution by using 2D shapes of tumors and OARs in CT images. Mahmood et al.[42] introduced a generative adversarial network (GAN) model to construct the U-net-based generator which predicts the dose distributions by the 2D contours of tumor and OARs in CT images. Chen et al.[43] constructed the dose distribution estimation system by re-training ResNet101 using CT images containing tumors and OARs regions. Fan et al.[44] developed an encoder-decoder network for predicting the dose distribution by using the features extracting from 2D contours of tumor and OARs independently.

The previous deep learning-based methods ([42],[43],[44]) predict directly the dose distribution from the contoured CT images of 217, 80 and 270 NPC patients, respectively. Although the previous deep learning-based methods use many images, the accuracy of the previous deep leaning-based methods for predicting the dose distribution is still low because of the following reason. The NPCs have complex shapes, and are located close to OARs that may influence on the direct prediction of the dose distribution. Therefore, the estimation of the desired dose distribution is complex and difficult problem to deliver more focused radiation to the NPCs while avoiding the exposure of OARs. Considering this, the direct prediction of the dose distribution requires a huge number of dataset. As mentioned before, according to the NPC incidence, it is quiet difficult to collect such huge number of NPC patients.

1.3.3 NPC volumes estimation

CNNs have achieved successes in medical image analysis including the extraction of tumors[32] and OARs[39]. Moreover, a recurrent neural network (RNN) is applicable to predict the future state of a target object. For example, in the medical research field, RNN is applied to cell growth[45].

Considering the characteristic of the CNN and RNN, Wang et al.[46] combined CNN with RNN to predict the future evolution of lung tumors in week 4 to 6 from 3D CT images obtained in week 1 to week 3. The method in [46] used 430 of weekly patients images in the experiments. Here, there is high change of the lung tumor in two successive weekly RT. Moreover, generally, the anatomical structures of tumors and OARs highly change between week 5 and 6 of RT. However, the proposed method in

[46] use only weekly CT images obtained in week 1 to week 3. Therefore, The method in [46] results a low accuracy of predicting the NPC volumes in week 5 and 6.

1.4 Purpose

In this thesis, to construct an automatic system for IMRT, I propose three fundamental techniques, for segmenting NPC from CT images, predicting the dose distribution and predicting the future evolution of NPC and OARs in response to RT, respectively.

Firstly, I develop a new CNN-based method for segmenting NPC in the early stage, that is, T1 and T2. The proposed method is based on the concept of cascade classifiers. Here, NPC never invades the non-target organ regions. Moreover, the non-target organs can be easily determinate in CT images. Considering these, the first phase in the proposed method is to extract and eliminate non-target organ regions from a given CT image. In the second phase, we detect NPC regions from the remaining regions in the CT image. CNN in each phase is constructed to detect the target regions in the phase: non-target organs and NPCs in the first and second phases, respectively. We use as an input of CNN two types of patches: overlapping patches with a fixed size and with different sizes. Therefore, we develop two different systems for detecting NPC tumor according to the type of patches.

Secondly, I propose a new system for predicting the dose distribution from contoured CT images that contain the contours and regions of tumors and OARs. The proposed system consists of two main steps. In the first step, for each beam, the proposed system extracts the feature from the corresponding beam orientation. Next, the system predicts the dose distribution by using all the extracted features from the corresponding beam. For each phase, a different CNN is devoloped to predict the target that is the feature of the beam or the dose distribution.

Thirdly, I present a system for predicting the future evolution of NPC and OARs in *n*-th week in response to RT from previous weekly CT images one to (*n-1*)-th weeks ago. The proposed system consists of two neural networks: one is CNN, and the other is a Gated Recurrent Unit (GRU)[47] which is one type of RNN networks. The former uses two types of input data : a whole CT image and overlapping patches with three different sizes obtained from the CT image. Therefore, four sub-CNNs extract relevant features from their corresponding input data by a weekly CT image. Using the three section of CT images, our proposed method learns the 3D evolution of the NPC and

OARs by different features extracted from the three views in the CT images.

1.5 Contribution

The three proposed methods have the following contributions:

- NPC segmentation:

  Our main contribution for NPC segmentation is to introduce a new cascade strategy for detecting NPC. As mentioned above, NPC has a similar intensity value to its surrounding tissues in the CT images. This results in the low NPC detection accuracy of the previous works that focused on the direct detection of the NPC from CT images. On the contrary, in the early stage, NPC and non-target organs are separately in the CT images. Practically, the segmentation of non-target organs from CT images have already achieved by Ibragimov et al.[39]. In addition, non-target organs never contain NPC. Considering these, I introduce two phases: the first phase is to detect non-target organ regions from CT images while the second phase is to identify NPC regions from the remained CT images after eliminating the detected non-target regions in the first phase. The elimination of the non-target organ regions from the CT image enables us to easily detect the NPC regions with high accuracy compared with the direct NPC detection from the original CT images. Owing to this, the use of the cascade strategy leads to the improvement of NPC detection accuracy.

  In addition, most of the related works for NPC segmentation have focused on only axial section of CT image as input data. On the contrary, the proposed method trains the CNNs by using three orthogonal images: axial, coronal and sagittal images. The use of the three orthogonal images enables the CNNs to learn the NPCs through different features extracted from three orthogonal views in the CT images.

- Dose distribution estimation:

  The dose distribution is determined by the sum of all beams dose. In other words, the dose of each beam is one of the most important factors to estimate the dose distribution. As mentioned above, the direct prediction of the dose distribution from contoured CT images is a complex and difficult problem. Considering these, I introduce two phases: the first phase is to extract the features of

beams dose from contoured CT images while the second phase predict the dose distribution from the extracted features. Owing to my strategy, the problem for predicting the dose distribution becomes easier. Because of this advantage, the proposed method can improve the dose prediction accuracy compared with the previous works.

- Future change prediction of NPC and OARs:

  The previous works for estimating the tumor evolution use the CT images acquired one to three weeks ago. Compared with the previous works, the proposed method use richer information, that is, all CT images acquired one to (*n-1*)-th weeks ago. Moreover, the proposed method determines the reliable prediction by integrating the three predictions obtained from three directional images: axial, coronal and sagittal images. Owing to these, the proposed method achieves prediction of the future evolution of the NPC and OARs regions.

1.6 Organization of this thesis

This thesis is organized into the following chapters.

- **Chapter 2** explains the details of the dataset and the proposed strategy for segmenting NPC. Moreover, the experimental results with discussion are described.
- **Chapter 3** describes the proposed system for predicting dose distribution by the two different integration methods. Moreover, the experimental results with discussion are described.
- **Chapter 4** states the proposed method for predicting the evolution of NPC and OARs during ART treatment. Moreover, the experimental results with discussion are described.
- **Chapter 5** makes the conclusion and future works of the proposed methods.

2

3D NPC segmentation

As mentioned in Chapter 1, the previous CNN-based methods[36],[37],[38] for segmenting the NPC directly from CT images result in a low detection accuracy of NPC in early stages T1 and T2. Moreover, most of NPC patients are diagnosed at advanced stages. Therefore, it is quite difficult to collect a huge number of CT images of NPC cases at early stages.

To solve these problems, in this chapter, a new system, CNN-based NPC detector (CND) system, is proposed for segmenting NPC from CT images. Considering that non target organs never include NPC, the proposed CND system introduces a cascade strategy composed of two-phase manners: 1) segmenting non-target organs from CT images and 2) Segmenting NPC from the remained regions in CT images.

In addition, the CNN-based methods in [36],[37],[38] use only axial section of CT images as input data. On the contrary, 3D shape of NPC is useful for radiation oncologists to destroy the NPC completely. Therefore, the proposed CND system is a 3D segmentation of NPC by integrating the obtained results from the three orthogonal:

axial, coronal and sagittal images.

2.1 Material

A total of 70 patients with NPC stage T1 and T2 were used to train and test the proposed system. All CT images of the 70 patients were collected from the same hospital, Radiotherapy Department of Habib Bourguiba Hospital of Sfax-Tunisia, and had been acquired in axial section in the helical mode by using a Philips Brilliance 16 Big Bore CT Scanner. The CT images from a Philips Brilliance 16 Big Bore CT Scanner had in-plane spatial resolution of 1.16×1.16 mm, 3 mm slice thickness with no intersection gap and tilted gantry. Therefore, the size of each voxel in the constructed CT images is $1.16 \times 1.16 \times 3mm^3$, while the size of an axial CT image is 512×512 [pixel]. During the CT scanning process, each patient is immobilized with a thermoplastic mask (head, neck, shoulder) in head first supine position. Figure2.1 shows the examples of the CT images used to train the proposed method.

In Philips Brilliance 16 Big Bore CT Scanner, the total number of axial images obtained for each patient is between 165 and 178 according to the patients as shown in Figure2.2. The horizontal and vertical axis in Figure2.2 shows the patient ID and the number of the CT images of the patient, respectively.

In each CT image, two radiation oncologists manually extract the NPC and OARs regions, that used as non-target regions in the proposed system. Here, each pixel of the contoured regions in the CT image is assigned to one unique label that is NPC or non-target organ region. Therefore, two labels are used as teaching labels to train and evaluate the proposed CND.

2.2 Methods

2.2.1 Overview

The proposed system, CND, focuses on the segmentation of NPC in stage T1 and T2. As mentioned in Chapter1, NPC regions are never invade into the non-target organ regions. Therefore, the proposed CND consists of two main phases as shown in Figure2.3. In the CND, the first phase outputs the segmented non-target organ regions from the CT images. Then, the segmented regions are eliminated from CT images

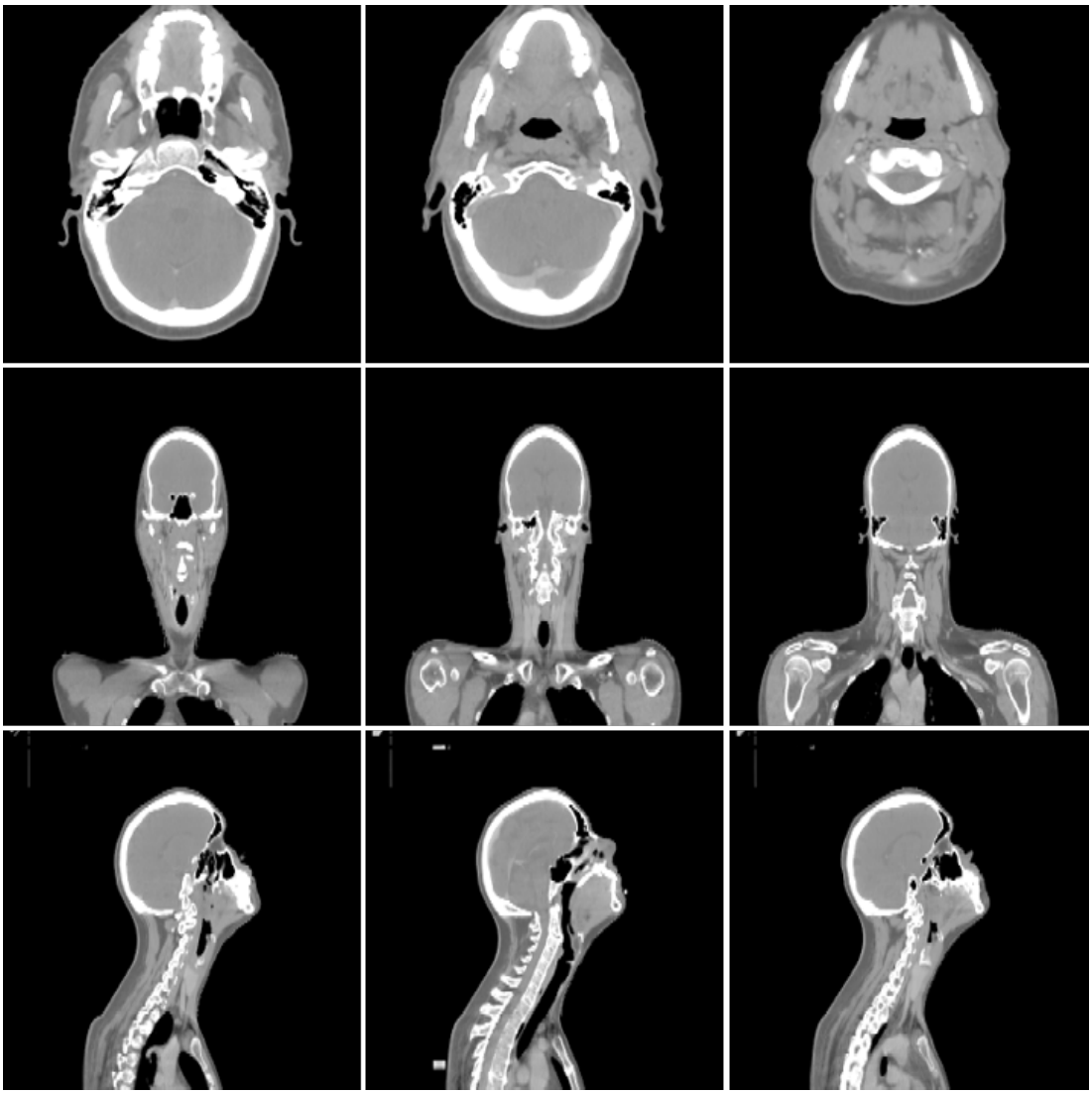


Figure 2.1: Some examples of the employed images in my database. The top row corresponds to CT images in the axial section, while the second and bottom row are CT images from coronal and sagittal section, respectively.

to be used in the second phase. The second phase in the proposed CND trains the remained regions in the obtained image to segment the NPC regions. Through the two phases, two CNNs are applied to segment the non-target and NPC regions in the first and second phase, respectively. In addition, since NCP appears in different locations and same intensity to non-target organ regions, overlapping patches are obtained from the original size of CT image by sliding a window from left to right and top to bottom to train each CNN. Here, the overlapping patches aims to train the CNN with local

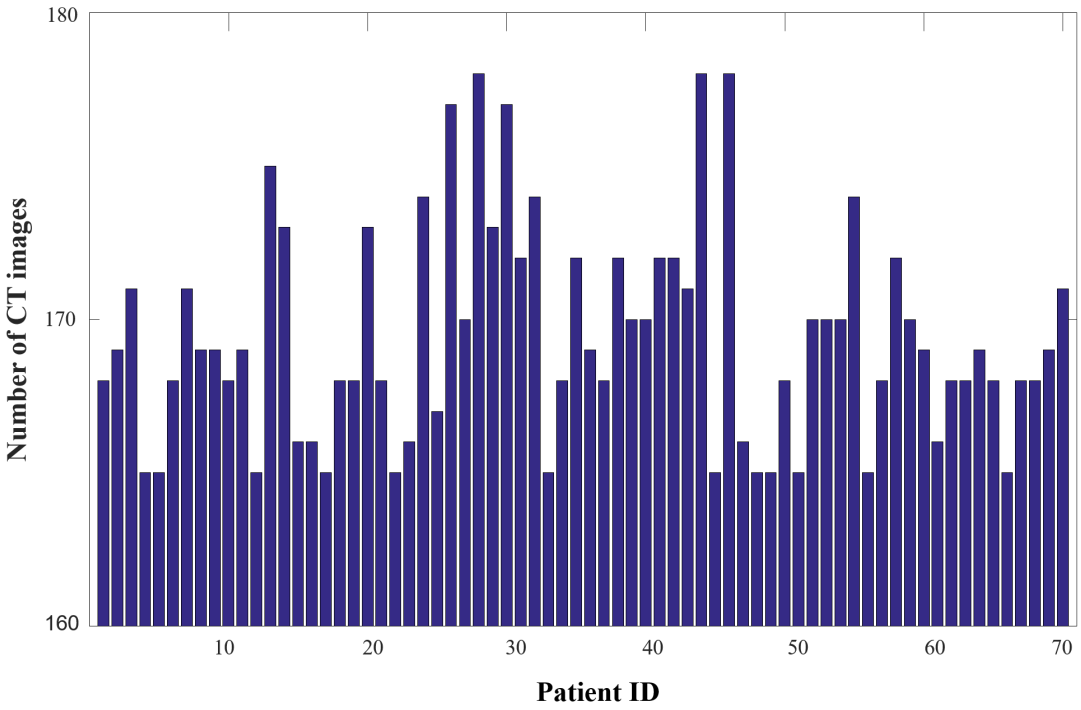


Figure 2.2: Number of CT images of each NPC patient in my database.

information. Therefore, according to the segmented target regions, the two CNNs are called, CNN_n^o and CNN_n^t, where CNN_n^o and CNN_n^t segment non-target organs and tumors from $n \times n$ overlapping patches, respectively.

Moreover, the proposed system CND trains three orthogonal CT images, axial, coronal and sagittal images, separately. Therefore, three types of CNNs are used in the first and second phase of the proposed CND to segment the target regions from the three orthogonal CT images. The first three CNNs used in the first phase are called $CNN_n^{o,a}$, $CNN_n^{o,c}$ and $CNN_n^{o,s}$, while those are used in the second phase are called $CNN_n^{t,a}$, $CNN_n^{t,c}$ and $CNN_n^{t,s}$ for axial, coronal and sagittal sections, respectively. All CNNs used in the first and the second phase, have the same architecture. As a result, CND system outputs different NPC regions from three sections images. To obtain the final NPC regions, an integration process is introduced to integrate all NPC regions obtained from the three sections. In the segmented NPC regions by the three sections, the pixel has a value 1 corresponds to the NPC region. In addition, the same pixel can also have a value 0 which means other region obtained from one and/or two sections. Therefore, in the integration process, I count the number of each label using all the

sections including the pixel. The label with the maximum vote is regarded as the final segmentation result of the pixel.

As mentioned above, the CNNs in the proposed CND system train $n \times n$ overlapping patches obtained from the CT images. Here, two types of the size of patches is introduced, that are overlapping patches with a fixed and different sizes. Therefore, two types of CNDs, called CND1 and CND2, are constructed to train overlapping patches with a fixed and different sizes, respectively. The details of each CND are described in the following.

2.2.2 CND using fixed-size overlapping patches

The first proposed method, CND1, trains a fixed size of overlapping patches obtained by sliding a window from left to right and top to bottom from the CT image. A preliminary experiments are made to determine the size of the patches. Here, through the preliminary experiments, it is observed that the proposed method CND1 is applicable to any size of patches. Therefore, 16×16 window is chosen as the fixed size of the patch in CND1 to contain a small and local information of the target region since the NPC and non-target organs have similar intensities as mentioned in Chapter 1. The overview of the CND1 using overlapping patches with a fixed size is shown in Figure 2.4.

As shown in Figure2.4, CND1 outputs in the first and second phase the segmentation result of the patch. Since CND1 trains multiple patches obtained from CT image, multiple segmentation results for the pixel are obtained. Therefore, a voting method is introduced in both phases of CND1 to decide the final segmentation result of each pixel. Here, for the same pixel in the multiple segmentation results of all patches, the voting method counts the labels of the pixel, that are ”target region” and ”other”. The label with maximum vote is regarded as the label of the final segmentation result.

2.2.2.1 CNN construction

As shown in Figure2.5, CNN^{o}_{16} and CNN^{t}_{16} composed of three convolutional layers, three max pooling and one fully connected layer. The three convolutional layers consists of a window size of 3x3, a stride of 1 and the same padding, where their outputs are 32, 32 and 64 dimensional feature vectors, respectively. In addition, each convolution layer uses a rectified linear unit (ReLU) as an activation function. The three max

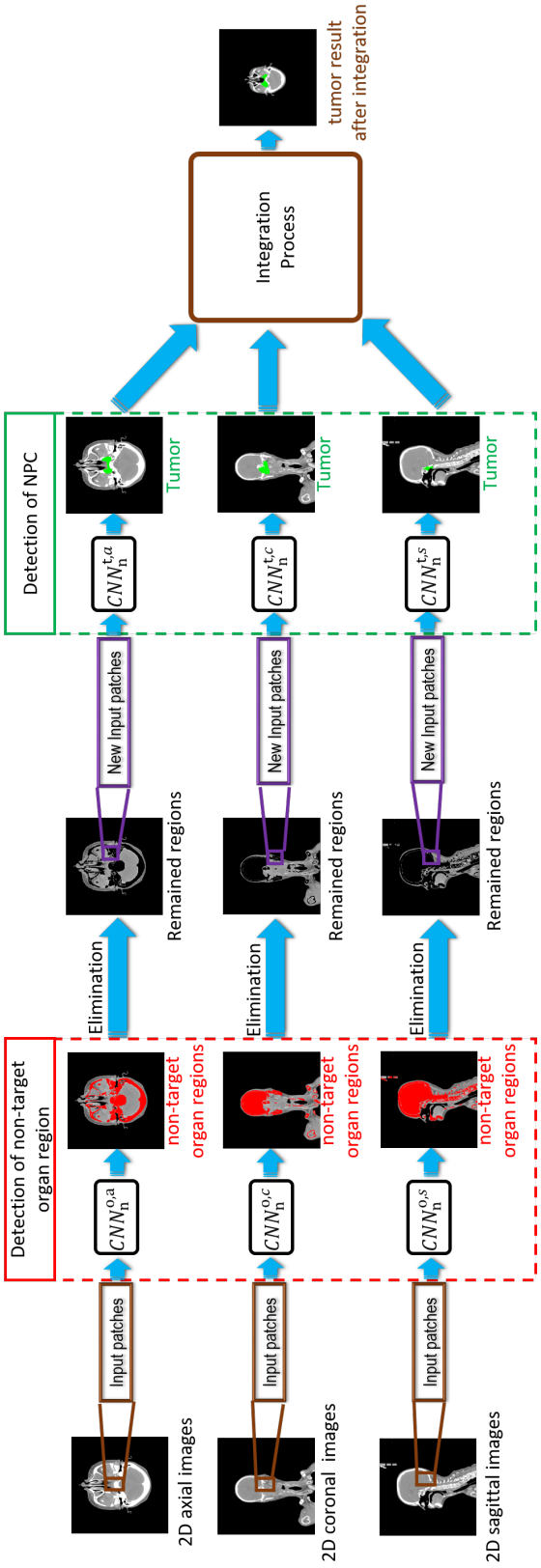


Figure 2.3: The schematic of the proposed CNN-based NPC detector (CND) system.

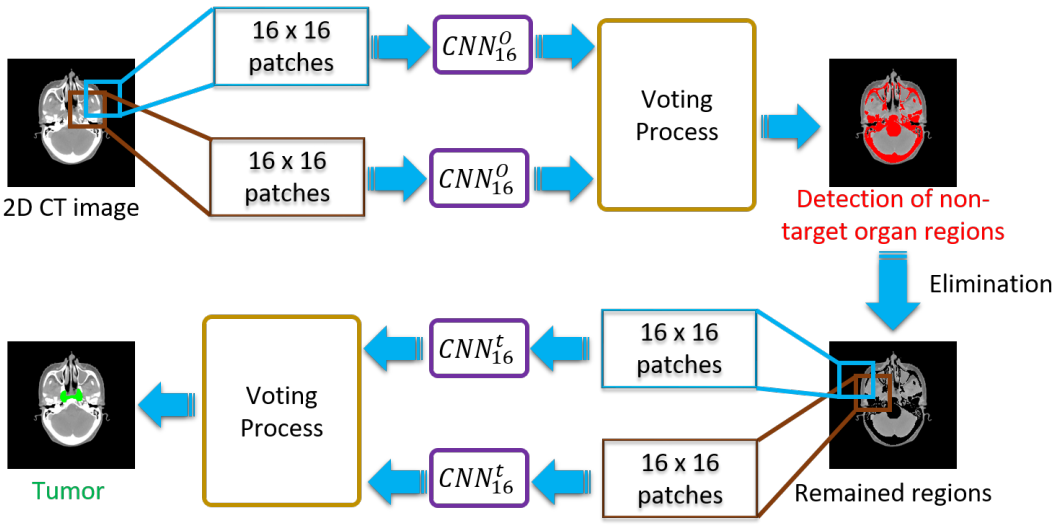


Figure 2.4: CND using fixed-size overlapping patches.

pooling have a window size of 3x3 and same padding that output a feature vector with 256 dimensions.

After the last max pooling, the proposed CNNs in CND1 is ended with fully connected network (FCN) that segments the target regions. FCN contains one hidden and output layer with 256 nodes. The activation functions in the hidden and output layers are ReLu and sigmoid functions, respectively. The FCN outputs the segmented patch in which each pixel is assigned to a label "target region" or "others".

The training of CNNs is achieved by binary cross entropy function that maximizes the probability of detecting the target region F described in Eq.2.1:

$$F = -\left(y_i log\left(\widehat{y_i}\right) + \left(1 - y_i\right) log\left(1 - \widehat{y_i}\right)\right) \ , \tag{2.1}$$

where i is the number of classes, and y and $\hat{y}$ are the value of the teaching label and the output of the CNN, respectively.

2.2.3 CND using various-size overlapping patches

Unlike CND1, the proposed method, CND2, trains overlapping patches with various sizes (Figure2.6). As mentioned above, through the preliminary experiments, there are no limitation about the size of the patches. Therefore, 16x16, 32x32 and 64x64 [pixel] are used as input patch sizes in the proposed CND2. By using three different

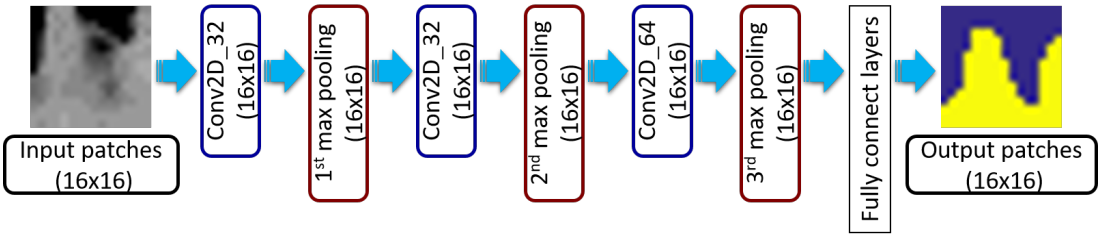


Figure 2.5: The architecture of the CNN using 16x16 patches, CNN^{o}_{16} and CNN^{t}_{16}.

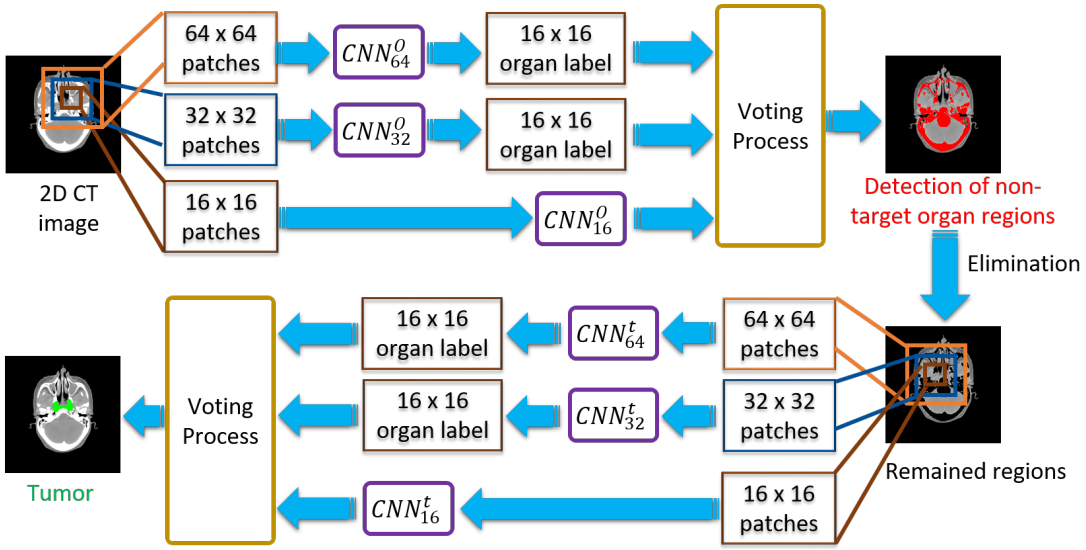


Figure 2.6: CND using various-size overlapping patches.

size of patches, CND2 aims to predict the label of each pixel in the multiple patches by considering both local and global information of the pixel.

According to the three different size of patches, 16×16, 32×32 and 64×64 patches, three CNNs are introduced in the first and second phase of CND2 to predict the target region from the three orthogonal images. Here, the CNNs used in CND1 are incorporated into CND2 to train the 16×16 patches, while the other two CNNs are the modified CNN^{o}_{16} and CNN^{t}_{16} to suit with the 32×32 and 64×64 patches. Both CNNs used 32×32 and 64×64 patches as input data have the same architecture as CNNs introduced in CND1. However, the modification is represented in the stride size of the max pooling layers to obtain segmented patches with size 16×16. In the case of the CNNs using 32×32 patches, CNN^{o}_{32} and CNN^{t}_{32}, the third max pooling is modified by changing a stride size of 2. For CNNs using 64×64 patches, CNN^{o}_{64} and CNN^{t}_{64}, the stride sizes in the 2nd and 3rd max pooling are replaced into 2. The CNNs using

32x32 and 64x64 patches are shown in Figure2.7 and 2.8.

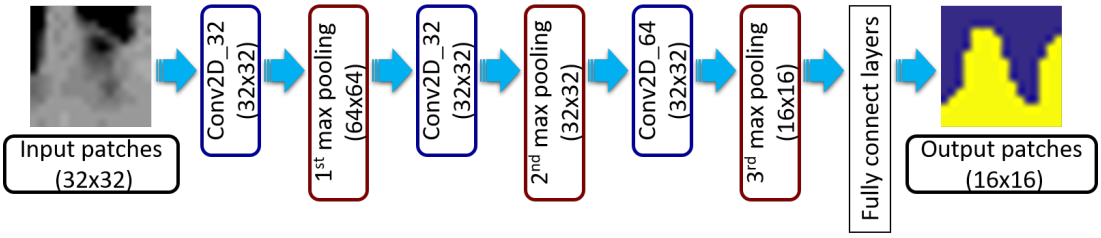


Figure 2.7: The architecture of the CNN using 32x32 patches, CNN^o_{32} and CNN^t_{32}.

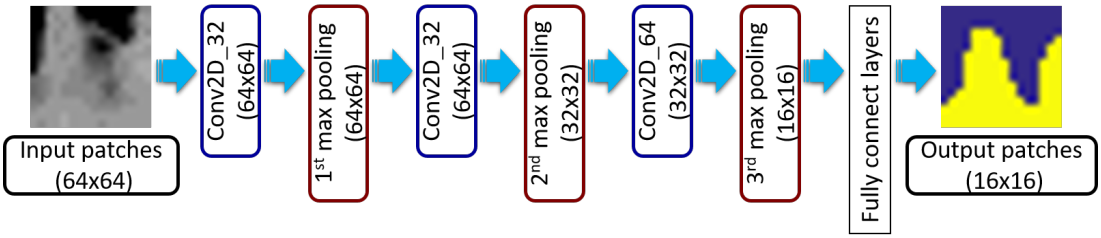


Figure 2.8: The architecture of the CNN using 64x64 patches, CNN^o_{64} and CNN^t_{64}.

Through the three CNN, CND2 outputs multiple segmentation results with size 16x16 from 16×16, 32×32 and 64×64 patches. Same as CND1, a voting method is applied in CND2 to determinate the final segmentation result from the results obtained by three different patches. The voting method has the same process as those used in CND1 to decide the label of the pixel in the final segmentation result of the target region.

2.2.4 Integration of segmentation results from three directional sections

The proposed system trains the dataset from three orthogonal CT images independently and outputs three different NPC results. To decide the final NPC regions, an integration process is introduced after the second phase of CND1 and CND2 by using the three segmented regions of NPC. In each segmented region, each pixel has a value 1 or 0 that corresponds to NPC or other region, respectively. Here, the integration process counts the number of labels of each pixel with label=1. Through the integration process, three voting numbers 1, 2 or 3 can be collected from one, two or three sections, respectively. In the end, the integration process outputs the final segmentation

result of NPC region that contains all pixels with a voting number 2 or 3, while each pixel with voting number 1 is not considered as NPC region.

2.3 Experimental results

2.3.1 Overview

In the experiments, CT images of 70 patients with NPC were trained and tested the proposed systems, CND1 and CND2, to verify its applicability. The 70 patient images are divided into seven sub-datasets of 10 patients images. Therefore, seven fold cross validation is used to train and evaluate the proposed system, where six sub-datasets are the training dataset and the rest of the sub-dataset is employed as testing dataset.

To compare the applicability of the proposed systems with the related works, the method proposed by Wang et al.[37] (WCNN) is selected as conventional method since the method uses overlapping patches obtained from the CT images as input to identify NPC directly without the phase of detecting non-target organ regions. In addition, Men et al. [36] (DDNN), and U-net([48]) are used as comparison methods and are constructed based on the proposed strategy. Another type of U-net, U-net 2, is introduced as additional comparison method. Here, U-net 2 segments non-target organ and NPC regions simultaneously from the whole CT image. Therefore, WCNN, DDNN, U-net and U-net 2 are applied as comparison methods that use CNN. As a non-deep learning method, the proposed method by Huang et al.[48] called AC, is used to detect NPC directly from a given CT image by active contours. Therefore, seven methods, performed on a workstation with Intel Core i7-3960X CPU and NVIDIA Quadro GP100 GPU, are applied to detect NPC in given CT images in the experiments.

In the experiment, four measurements, precision(Eq.2.2), recall(Eq.2.3), dice similarity coefficient(DSC)(Eq.2.4) and average symmetric surface distance (ASSD)(Eq.2.5), were used as quantitative evaluations to evaluate the segmentation performance of the seven methods with the manual contouring of the target region obtained by the two radiation oncologists and used as a ground -truth. Here, the DSC in Eq.2.4 is used to measure the overlapping area between the ground truth and the segmented region by each method. Otherwise, ASSD in Eq.2.5 is used to evaluate the shape difference between the boundary contour of the ground truth and the boundary contour obtained by each method. The four measurements are defined as fellow:

$$Precision = \frac{TP}{TP+FP}\ ; \tag{2.2}$$

$$Recall = \frac{TP}{TP+FN}\ ; \tag{2.3}$$

$$DSC = \frac{2TP}{2TP+FN+FP} \tag{2.4}$$

where *TP*(true positive) is the total number of pixels in the overlapping part between the output of each method and the ground-truth while *FP*(false positive) and *FN* (false negative) are the numbers of pixels rejected by the manual contouring and the method, respectively.

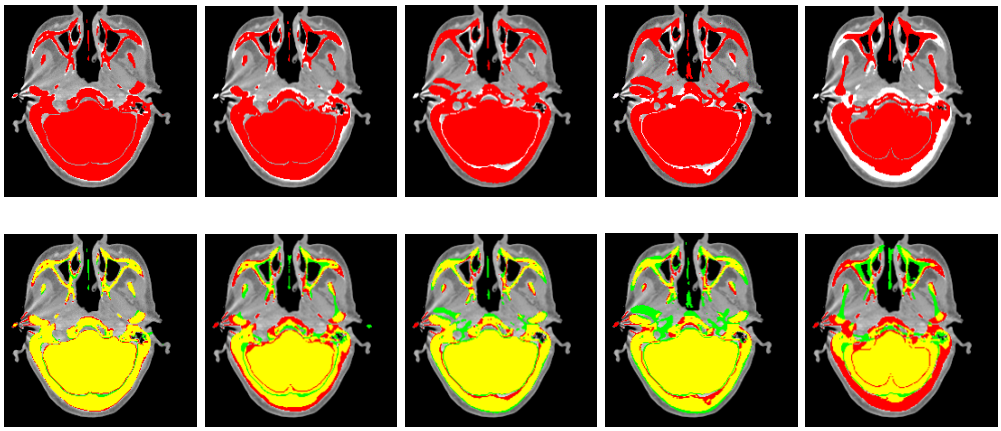
$$ASSD = \frac{1}{|S|}\sum_{\mathbf{p}\in S} d_1(\mathbf{p},\mathbf{q}) + \frac{1}{|\widetilde{S}|}\sum_{\mathbf{q}\in \widetilde{S}} d_2(\mathbf{q},\mathbf{p}) \tag{2.5}$$

where S and $\widetilde{S}$ are the set of the points on the contour of the segmented region by the method and its ground truth, respectively. Moreover, $|S|$ (or $|\widetilde{S}|$) is the total number of the points in S (or $\widetilde{S}$). The function $d_1(\mathbf{p},\mathbf{q})$ returns the distance between the point $\mathbf{p}$ in S and its closest point $\mathbf{q}$ in $\widetilde{S}$ while $d_2(\mathbf{q},\mathbf{p})$ calculates the distance between the point $\mathbf{q}$ in $\widetilde{S}$ and its closest point $\mathbf{p}$ in S.

2.3.2 Results

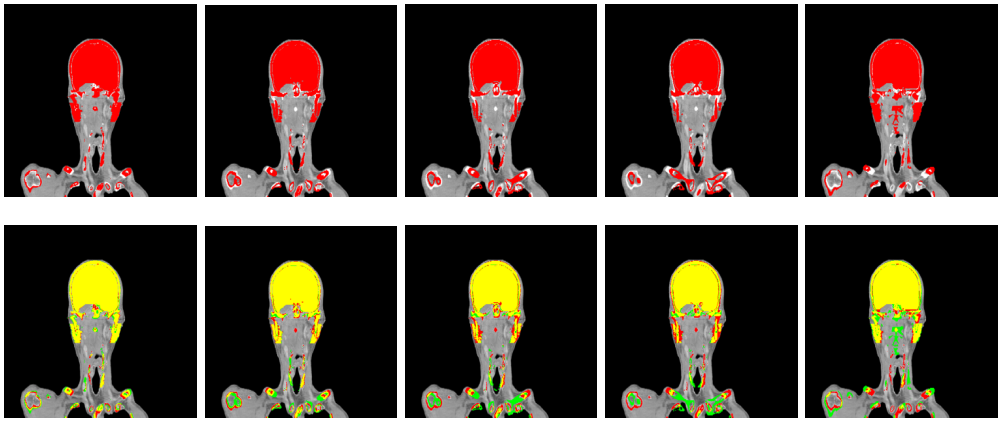
The proposed system, CND, outputs non-target organ and NPC regions in the first and second phase, respectively. Therefore, two steps of evaluation are applied to evaluate the proposed methods that identify the target region.

In the first evaluation, the quantitative values of the proposed methods that segment the non-target regions, that are CND1, CND2, DDNN, U-net and U-net2, are calculated. In addition, since all comparison methods use axial, coronal and sagittal images as training data, the segmented results of non-target organ regions by the five methods are calculated and shown in Table.2.1(a)–(c). Examples of the segmented results from the axial, coronal and sagittal images by the five methods are illustrated in Figures. 2.9, 2.10 and 2.11. In these figures, the segmented target region by the corresponding method is represented in the top row with red color. On the contrary, the matching and non-matching regions between the segmented result of the target region and the



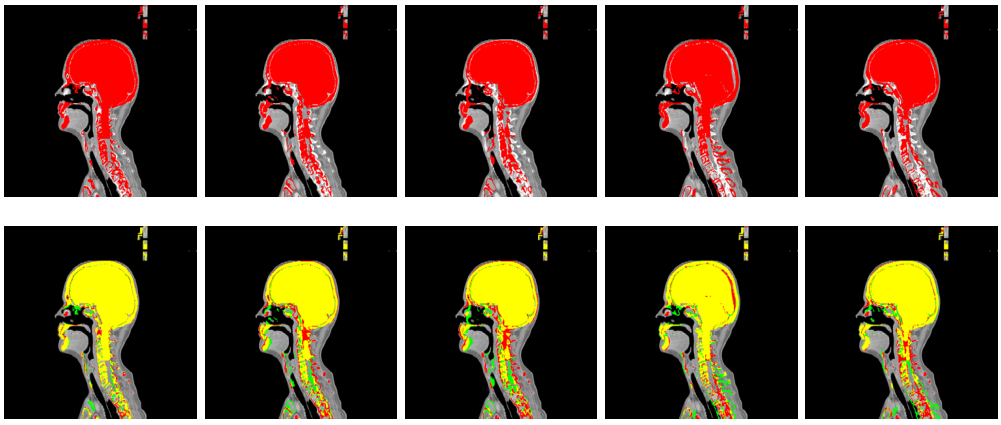
(a) (b) (c) (d) (e)

Figure 2.9: Non-target organ segmentation results (red) from axial section obtained by (a,b) my method using (a)CND2 and (b)CND1, (c)DDNN, (d)U-net and (e)U-net2. (Red) the segmented non-target organ; (Green and Yellow) the matching (green) and non-matching (yellow) regions between the segmented non-target organ and the ground-truth.



(a) (b) (c) (d) (e)

Figure 2.10: Non-target organ segmentation results (red) from coronal section obtained by (a,b) my method using (a)CND2 and (b)CND1, (c)DDNN, (d)U-net and (e)U-net2. (Red) the segmented non-target organ; (Green and Yellow) the matching (green) and non-matching (yellow) regions between the segmented non-target organ and the ground-truth.



(a) (b) (c) (d) (e)

Figure 2.11: Non-target organ segmentation results (red) from sagittal section obtained by (a,b) my method using (a)CND2 and (b)CND1, (c)DDNN, (d)U-net and (e)U-net2. (Red) the segmented non-target organ; (Green and Yellow) the matching (green) and non-matching (yellow) regions between the segmented non-target organ and the ground-truth

Table 2.1: Average and standard deviation of the accuracy of detecting non-target organ regions from (a) axial, (b) coronal and (c) sagittal sections by my two methods, DDNN, U-net and U-net2.

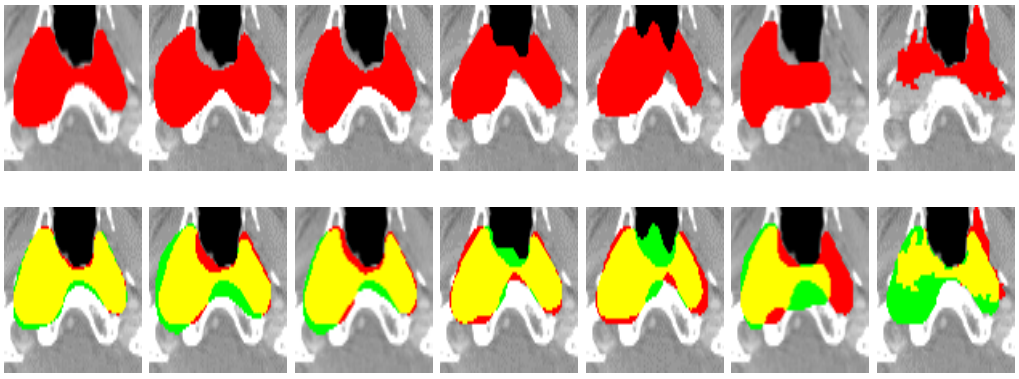
Method	Precision	Recall	DSC		ASSD	
			Values	p-value	Values	p-value
(a)Axial sections						
CND2	**0.88 ± 0.029**	**0.87± 0.041**	**0.87± 0.034**	—	**1.19± 0.177**	—
CND1	0.83 ± 0.037	0.83± 0.033	0.83± 0.35	N.S	1.52± 0.226	N.S.
DDNN	0.79± 0.015	0.80± 0.023	0.79± 0.018	< 0.05	1.89± 0.684	< 0.05
U-net	0.77± 0.029	0.78± 0.047	0.77± 0.036	< 0.05	2.09± 0.322	< 0.05
U-net2	0.63± 0.018	0.61± 0.021	0.62± 0.019	< 0.05	3.93± 0.713	< 0.05
(b)Coronal sections						
CND2	**0.85 ± 0.037**	**0.85± 0.012**	**0.85± 0.18**	—	**1.46± 0.174**	—
CND1	0.83 ± 0.051	0.83± 0.027	0.83± 0.035	N.S	1.55± 0.531	N.S.
DDNN	0.75± 0.053	0.74± 0.034	0.74± 0.041	< 0.05	2.19± 0.722	< 0.05
U-net	0.77± 0.044	0.78± 0.022	0.77± 0.029	< 0.05	2.01± 0.146	< 0.05
U-net2	0.67± 0.026	0.65± 0.041	0.66± 0.032	< 0.05	3.63± 0.467	< 0.05
(c)Sagittal sections						
CND2	**0.92 ± 0.057**	**0.91± 0.045**	**0.91± 0.05**	—	**0.58± 0.613**	—
CND1	0.87 ± 0.023	0.88± 0.043	0.87± 0.03	N.S	1.17± 0.422	< 0.05
DDNN	0.83± 0.025	0.83± 0.028	0.83± 0.026	< 0.05	1.56± 0.336	< 0.05
U-net	0.79± 0.031	0.78± 0.047	0.78± 0.037	< 0.05	1.92± 0.221	< 0.05
U-net2	0.71± 0.021	0.71± 0.034	0.71± 0.026	< 0.05	2.34± 0.411	< 0.05

ground-truth have yellow and green regions, respectively.

As a second evaluation, the accuracy of detecting the NPC regions is evaluated by CND1, CND2 and the five comparison methods from different sections. The quantitative results of segmenting NPCs from the three orthogonal images by the seven methods are presented in Table 2.2(a)–(c). Furthermore, Figures.2.12, 2.13 and 2.14 show examples of the detected NPCs from the axial, coronal and sagittal images obtained by the seven methods.

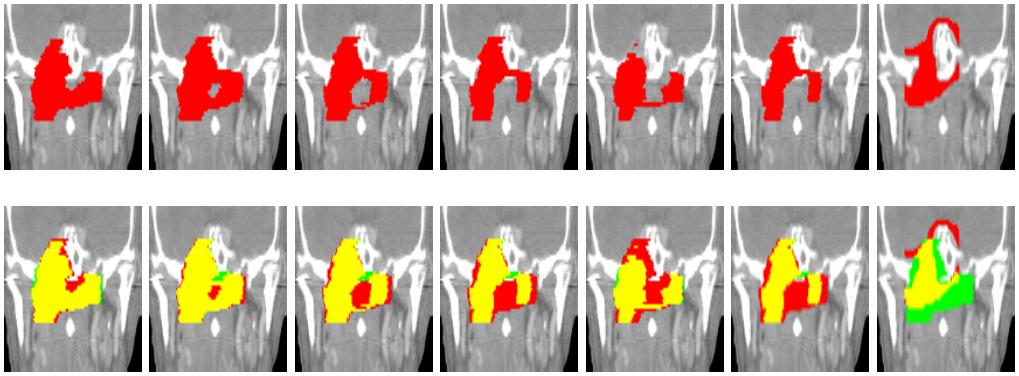
In all Figures.2.9 to 2.14, the segmented target region by the corresponding method is represented in the top row with red color. On the contrary, in the bottom row of Figures.2.9 to 2.14, the matching and non-matching regions between the segmented result of the target region and the ground-truth have yellow and green regions, respectively.

As mentioned above, the final segmentation of NPC regions is obtained by integrating the three different segmentation results obtained from the three orthogonal sections. Table.2.3 shows the quantitative results of the final estimated NPCs. An example of the final segmented NPCs obtained by the seven methods are illustrated in Figure.2.15. The yellow and green regions in Figures.2.15 have the same meanings as in Figures.2.9 to 2.14.



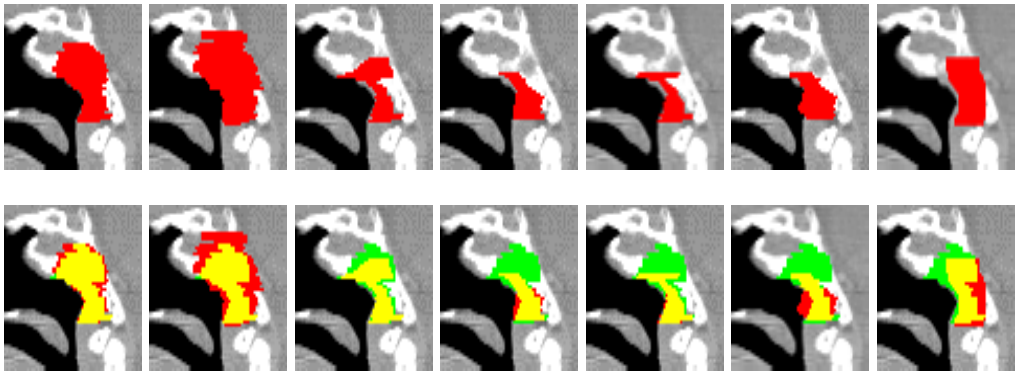
(a) (b) (c) (d) (e) (f) (g)

Figure 2.12: NPC segmentation results (red) from axial section obtained by (a,b) my method using (a) CND2 and (b) CND1, (c) DDNN, (d) WCNN, (e) U-net, (f) U-net2 and (g) AC. (Red) the segmented non-target organ; (Green and Yellow) the matching (green) and non-matching (yellow) regions between the segmented non-target organ and the ground-truth.



(a) (b) (c) (d) (e) (f) (g)

Figure 2.13: NPC segmentation results (red) from coronal section obtained by (a,b) my method using (a) CND2 and (b) CND1, (c) DDNN, (d) WCNN, (e) U-net, (f) U-net2 and (g)AC. (Red) the segmented non-target organ; (Green and Yellow) the matching (green) and non-matching (yellow) regions between the segmented non-target organ and the ground-truth.



(a) (b) (c) (d) (e) (f) (g)

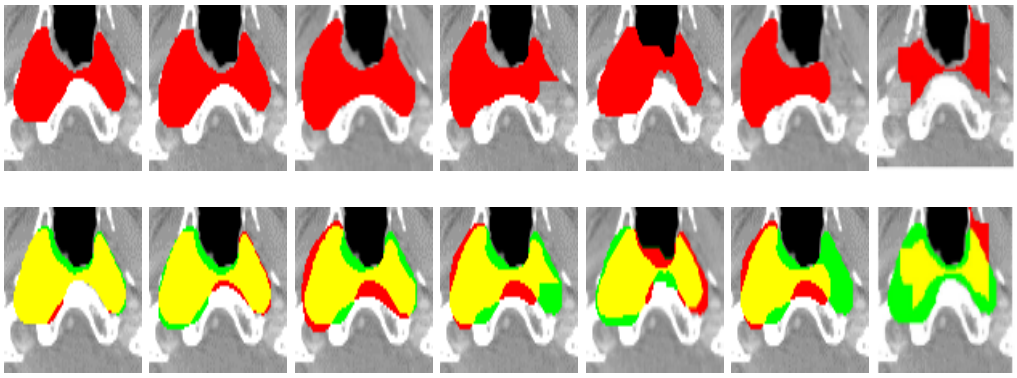
Figure 2.14: NPC segmentation results (red) from sagittal section obtained by (a,b) my method using (a) CND2 and (b) CND1, (c) DDNN, (d) WCNN, (e) U-net, (f) U-net2 and (g)AC. (Red) the segmented non-target organ; (Green and Yellow) the matching (green) and non-matching (yellow) regions between the segmented non-target organ and the ground-truth.

Table 2.2: Average and standard deviation of the accuracy of detecting NPC regions from (a) axial, (b) coronal and (c) sagittal sections by my two methods, DDNN, WCNN, U-net, U-net2 and AC.

Method	Precision	Recall	DSC		ASSD	
			Values	p-value	Values	p-value
(a)Axial sections						
CND2	**0.87 ± 0.012**	**0.83± 0.041**	**0.85± 0.019**	—	**1.47± 0.392**	—
CND1	0.84 ± 0.023	0.83± 0.027	0.83± 0.025	N.S	1.50± 0.926	N.S.
DDNN	0.79± 0.025	0.79± 0.011	0.79± 0.015	< 0.05	1.87± 0.312	< 0.05
WCNN	0.71± 0.013	0.72± 0.014	0.71± 0.013	< 0.05	2.13± 0.644	< 0.05
U-net	0.64± 0.036	0.61± 0.047	0.62± 0.041	< 0.05	3.96± 0.551	< 0.05
U-net2	0.59± 0.021	0.57± 0.014	0.58± 0.017	< 0.05	4.48± 0.192	< 0.05
AC	0.54± 0.091	0.76±0.011	0.63± 0.044	< 0.05	3.91± 0.474	< 0.05
(b)Coronal sections						
CND2	**0.85 ± 0.025**	**0.81± 0.031**	**0.83± 0.028**	—	**1.49± 0.142**	—
CND1	0.81 ± 0.033	0.80± 0.043	0.80± 0.037	N.S	1.76± 0.522	N.S.
DDNN	0.76± 0.030	0.74± 0.021	0.75± 0.025	< 0.05	2.11± 0.153	< 0.05
WCNN	0.71± 0.023	0.72± 0.047	0.71± 0.031	< 0.05	2.59± 0.711	< 0.05
U-net	0.75± 0.013	0.72± 0.041	0.73± 0.020	< 0.05	2.39± 0.645	< 0.05
U-net2	0.70± 0.031	0.68± 0.030	0.69± 0.030	< 0.05	2.88± 0.418	< 0.05
AC	0.51± 0.082	0.72±0.031	0.60± 0.014	< 0.05	4.19±0.361	< 0.05
(c)Sagittal sections						
CND2	**0.89 ± 0.017**	**0.89± 0.022**	**0.89± 0.019**	—	**1.03± 0.270**	—
CND1	0.86± 0.021	0.84± 0.013	0.85± 0.016	N.S	1.39± 0.311	N.S.
DDNN	0.84± 0.025	0.83± 0.029	0.83± 0.027	< 0.05	1.50± 0.631	< 0.05
WCNN	0.74± 0.025	0.75± 0.048	0.74± 0.033	< 0.05	2.21± 0.229	< 0.05
U-net	0.80± 0.013	0.79± 0.014	0.79± 0.013	< 0.05	1.88± 0.412	< 0.05
U-net2	0.73± 0.021	0.74± 0.043	0.73± 0.028	< 0.05	2.41± 0.446	< 0.05
AC	0.64±0.054	0.71±0.037	0.67± 0.024	< 0.05	3.49± 0.153	< 0.05

Table 2.3: Average and standard deviation of the accuracy of detecting the final NPC regions after the integration by my two methods, DDNN, WCNN, U-net, U-net2 and AC.

Method	Precision	Recall	DSC		ASSD	
			Values	p-value	Values	p-value
CND2	**0.89 ± 0.027**	**0.93± 0.028**	**0.91± 0.027**	—	**0.59± 0.346**	—
CND1	0.87± 0.017	0.89± 0.013	0.88± 0.015	< 0.05	1.12± 0.115	< 0.05
DDNN	0.81± 0.032	0.82± 0.041	0.82± 0.036	< 0.05	1.57± 0.391	< 0.05
WCNN	0.72± 0.017	0.72± 0.028	0.72± 0.021	< 0.05	2.63± 0.412	< 0.05
U-net	0.77± 0.031	0.81± 0.041	0.79± 0.035	< 0.05	1.88± 0.323	< 0.05
U-net2	0.66± 0.012	0.67± 0.019	0.67± 0.015	< 0.05	3.68± 0.118	< 0.05
AC	0.58± 0.054	0.76± 0.037	0.66± 0.024	< 0.05	3.71± 0.913	< 0.05



(a) (b) (c) (d) (e) (f) (g)

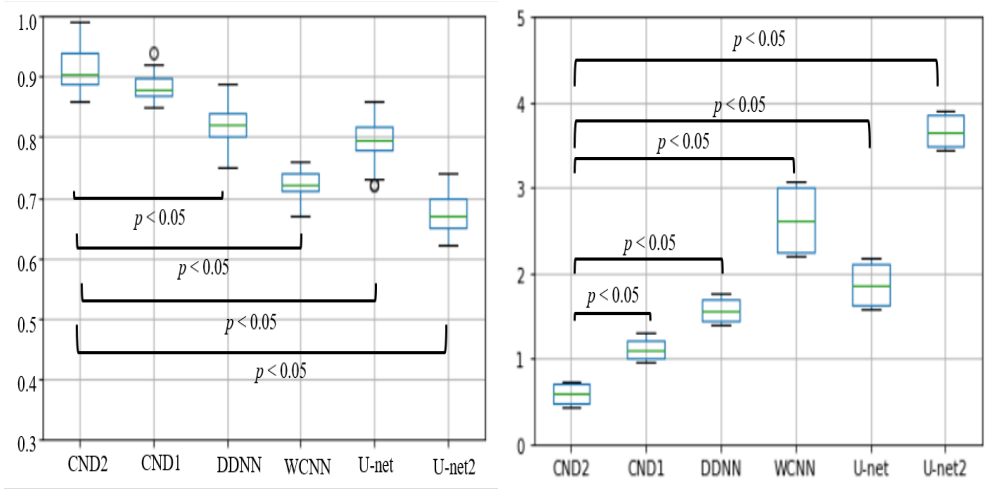
Figure 2.15: Final NPC results (red) obtained by (a,b) my method using (a) CND2 and (b) CND1, (c) DDNN, (d) WCNN, (e) U-net, (f) U-net2 and (g) AC. (Red) the segmented non-target organ; (Green and Yellow) the matching (green) and non-matching (yellow) regions between the segmented non-target organ and the ground-truth.

In addition, to compare the proposed method, CND2, with the other six methods, statistical analysis is performed with paired t-tests (p-value < 0.05) for DSC and ASSD. The calculated values of p-value are represented in the fifth and seventh columns in Tables.2.1–2.3. Furthermore, the boxplots of DSC and ASSD measurements for NPC after the integration process are shown in Figure.2.16.

2.4 Discussion

Generally, to evaluate the segmentation of NPC and non-target organ regions, DSC is the most common and the important quantitative evaluation in all conventional methods. It has been reported in [49]and [50] that the proposed method considered as a good method where its performance achieves a DSC with more than a threshold 0.75. Here, the accuracy of NPC regions in [49]and [50] are calculated by using the manual NPC contour delineation by two radiologists with DSC of 0.68 and 0.75, respectively. Referring these, the proposed system including CND1 and CND2 achieves a good performance where the minimum DSC value among all results was 0.83 obtained in the segmented NPC regions from coronal section by CND1.

In the first evaluation, that is the segmented non-target organ regions, the four methods, CND1, CND2, DDNN and U-net, except U-net 2 achieve a good segmentation. In the experiments, U-net2 detects NPC and non-target organ simultaneously. As



(a) DSC (b) ASSD

Figure 2.16: Boxplots of Integration NPC

mentioned in Chapter 1, the non-target organs and NPC regions have similar intensities in CT images. This is the reason why U-net 2 results a lower performance compared with the other four methods.

In the second evaluation, as shown in Table.2.2, the accuracy of the segmented NPC regions by the two proposed methods, CND1 and CND2, are higher than the other methods, especially, CND1. In addition, WCNN, U-net 2 and AC results a DSC lower than the threshold 0.75. Here, WCNN, U-net 2 and AC try to directly identify NPCs from the CT images. However, since NPC in CT images appear in different locations and has almost the same intensity value as those of non-target organ regions, the direct detection of NPCs from the CT images is a difficult task. Furthermore, AC results the lowest accuracy among all methods. the AC method uses active contour method to extract the NPC regions by evolving the shape of the active contour. However, NPC regions appear with complex shapes that makes the fitting to the contour is difficult. On the contrary, CND1, CND2, DDNN and U-net, segment NPC regions from the remaining regions obtained by eliminating non-target organs from the CT images. Therefore, the proposed strategy leads to obtain reliable NPC segmentation results.

Among the seven methods, the two proposed methods, CND1 and CND2, achieve a higher performance of the NPC segmentation. The high performance of CND1 and

CND2 is due to the following reasons. The proposed methods, CND1 and CND2, train patches obtained from a given image as input data while DDNN and U-net directly employ a whole image. By using patches, CND1 and CND2 can extract useful features that include the local differences between NPC and its surroundings compared with using whole CT images. Furthermore, overlapping patches can extract features from the local region around a certain pixel from different view points.

In the statistical analysis, the paired t-test indicates that CND2 is significantly different from all five comparison methods for DSC and ASSD. On the contrary, there is no significant difference between CND2 and CND1 for DSC while for ASSD, CND2 is significantly different from CND1.

As mentioned above, the seven methods output the final NPC region by integrating three segmentation results obtained from axial, coronal and sagittal images. In Table.2.3, It is observed that the final NPC segmentation by using the four methods, expect CND1 and CND2, results a lower accuracy compared with the best segmentation result among the three section images. On the contrary, the accuracy of the final NPC segmentation obtained by CND1 and CND2 is improved compared with those of the three segmentation results obtained from the three section. To check this tendency,

for each pixel, I count the number of the segmentation results in which the pixel is included in the estimated NPC regions. Here, set Φ of the pixels assigned to the label "NPC region" in at least one of the three segmentation results. The average rates of the pixels according to the number of the section segmentation in which the pixels are assigned the label "NPC region" is represented in Table.2.4. From Table.2.4, many

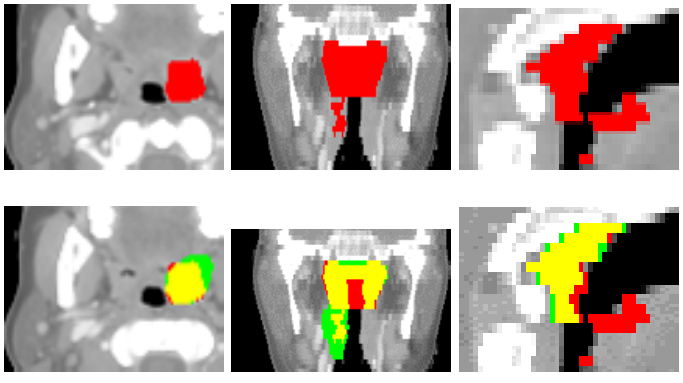
Table 2.4: Average percentage of the total amount of pixels voting in the integration process, where the obtained vote number of the pixel is =1 (from one section), =2 (from two sections) or =3 (from three sections).

Method	=1	=2	=3
CND2	8.02%	12.62%	79.36%
CND1	10.50%	20.11%	69.39%
DDNN	20.45%	17.31%	62.24%
WCNN	24.76%	25.14%	50.10%
U-net	32.61%	29.09%	38.30%
U-net2	45.16%	18.71%	36.13%
AC	34.17%	28.92%	36.91%

pixels among the three segmentation results obtained from DDNN, WCNN, U-net,U-net2 and AC are assigned as “NPC region” from only one orthogonal section. On the contrary, CND1 and CND2 obtain many pixels assigned to “NPC region” by the two or three segmentation. In other words, CND1 and CND2 can stably identify the NPC regions from CT images regardless of the section directions. Therefore, this stability of segmenting NPC by CND1 and CND2 results an improvement of the accuracy of the final NPC segmentation by using the integration process.

In addition, according to the qualitative values in all tables, there is a significant difference for ASSD between CND2 and CND1. By combining local and global structures, CND2 facilitates the detection of NPC and outputs more robust results that fit with the low contrast or the small difference between NPC and its surroundings compared to CND1.

Despite the high accuracy of the obtained results by CND2, it is observed that the proposed method results NPC regions with low accuracy in some cases (Figure.2.17). This problem is appeared when the NPC regions reach lymph nodes on a neck or invade a part of oropharynx that is the case of stage T3. However, CND1 and CND2 focus only on the segmentation of NPCs in early stage, T1 and T2. To avoid such problem, by increasing the number of CT images in stage T2, where the NPC is similar to stage T3 is one solution.



(a) (b) (c)

Figure 2.17: Failed NPC results (red) obtained by CND2 from (a) axial section, (b) coronal section and (c) sagittal section. (Red) the segmented non-target organ; (Green and Yellow) the matching (green) and non-matching (yellow) regions between the segmented non-target organ and the ground-truth.

2.5 Conclusion

In this chapter, I proposed a novel CNN-based system, CND, for segmenting the NPC regions from CT axial, coronal and sagittal images. The proposed CND consists of two phases. The first phase is to segment and eliminate the non-target organ regions from CT images. The second phase is to extract the NPCs from the remained regions in the CT images. The final segmented NPC regions are obtained by integrating the NPCs extracted from three orthogonal section images.

In the experiments, two CNDs, CND1 and CND2, are trained by using overlapping patches with fixed and different sizes, respectively. From the experimental results using the five comparison methods, CND2 leads to segment the NPC regions with high performance compared with CND1 and conventional methods.

3

Dose distribution prediction

As mentioned in Chapter1, the second step in the workflow of radiation oncologists is the determination of RT treatment. Although the previous deep learning-based methods [42], [43], [44] use many images, the accuracy of the previous deep leaning-based methods for predicting the dose distribution is still low because of the following reason. The NPCs have complex shapes, and are located close to OARs that may influence on the direct prediction of the dose distribution. Therefore, the estimation of the desired dose distribution is complex and difficult problem to deliver more focused radiation to the NPCs while avoiding the exposure of OARs.

Here, in the workflow of radiation oncologist, the dose distribution is obtained by the sum of all beams dose. Here, the dose of each beam is calculated from NPC and few OARs regions. Considering this, a new system introduces a cascade strategy composed of two-phase manners: 1) predicting the feature maps of each beam from contoured CT images and 2) predicting the dose distribution by integrating the obtained feature maps of all beams.

3.1 Materials

I used the data of 80 NPC patients undergoing IMRT at Radiotherapy department of Habib Bourguiba Hospital of Sfax-Tunisia. For all patients, treatment plans were generated a similar clinical protocol by two radiation oncologists and two dosimetrists. There are two data of each patient. One is 2D images (Figure.1.5(b)) of delivered dose determined by the two dosimetrists and approved by the two radiation oncologists. This image is called 2D real dose distribution image. The other is an axial section CT image with 256×256[pixels] trimmed from 512×512[pixels] original CT image. From each CT image, the two radiation oncologists manually trace all contours of NPC (gross tumor volume (GTV), clinical target volume (CTV), planning target volume (PTV)) and OARs. Here, GTV corresponds to the real NPC region while PTV and CTV are a margin regions (between 0.5 to 1 cm) that contain some tissues close to GTV to avoid any error by IMRT during the RT treatment. Each region is assigned with a unique teaching label. Table.3.1 presents the teaching label and the range of the delivered dose, called dose constraints, of each region. Here, there are six dose constraints: the maximum dose value(D_{max}), 1 cubic centimeter(D_{1cc}) of the dose, and 2%, 50% , 95% and 98% ($D_{2\%}$, D_{50}, $D_{95\%}$ and $D_{98\%}$) of the dose. Furthermore, two delineation images were constructed and used as training data. The first delineation image contains GTV, CTV and PTV regions, while the second delineation image corresponds to OARs regions. In the training, each region in the two delineation images is treated with a different channel of the input layer when fed into the learning network.

In the clinical protocol and planning template proposed by radiation oncologists, NPC patients are irradiated from 7 dynamic beams whose orientations are 0, 51, 102, 153, 204, 255 and 306 [deg] around a patient (Figure.3.1). Therefore, 2D real dose distribution of NPC patients, with a spatial resolution 0.25×0.25[cm], is constructed

Table 3.1: Detailed information of contoured regions and its dose constraint.

Contours	Dose constraint	Label	Contours	Dose constraint	Label
L/R temporomandibular joint(TMJ)	$D_{max} \leq$55 Gy	1/2	L/R parotid gland	$D_{50} <$26 Gy	12/13
Optic chiasma	$D_{max} \leq$54 Gy	3	Brainstem	$D_{1cc} <$54 Gy	14
L/R Lens	$D_{max} \leq$7 Gy	4/5	Spinal cord	$D_{1cc} \leq$40 Gy	15
Larynx	$D_{max} \leq$66 Gy	6	PTV(70 Gy)	$D_{95\%} \geq$66.5 Gy	16
Temporal lobe	$D_{2\%} <$60 Gy	7	GTV(70 Gy)	$D_{98\%} \geq$70 Gy	17
L/R Mandible	$D_{max} \leq$65 Gy	8/9	CTV(70 Gy)	$D_{95\%} \geq$70 Gy	18
L/R optic nerve	$D_{max} \leq$55 Gy	10/11			

L/R: Left/Right; D_{max}: maximum dose; D_{50}: 50% of the dose, $D_{2\%}$, $D_{95\%}$, $D_{98\%}$: 2%, 95%, 98% of the dose, respectively; D_{1cc}: 1 cubic centimeter of the dose.

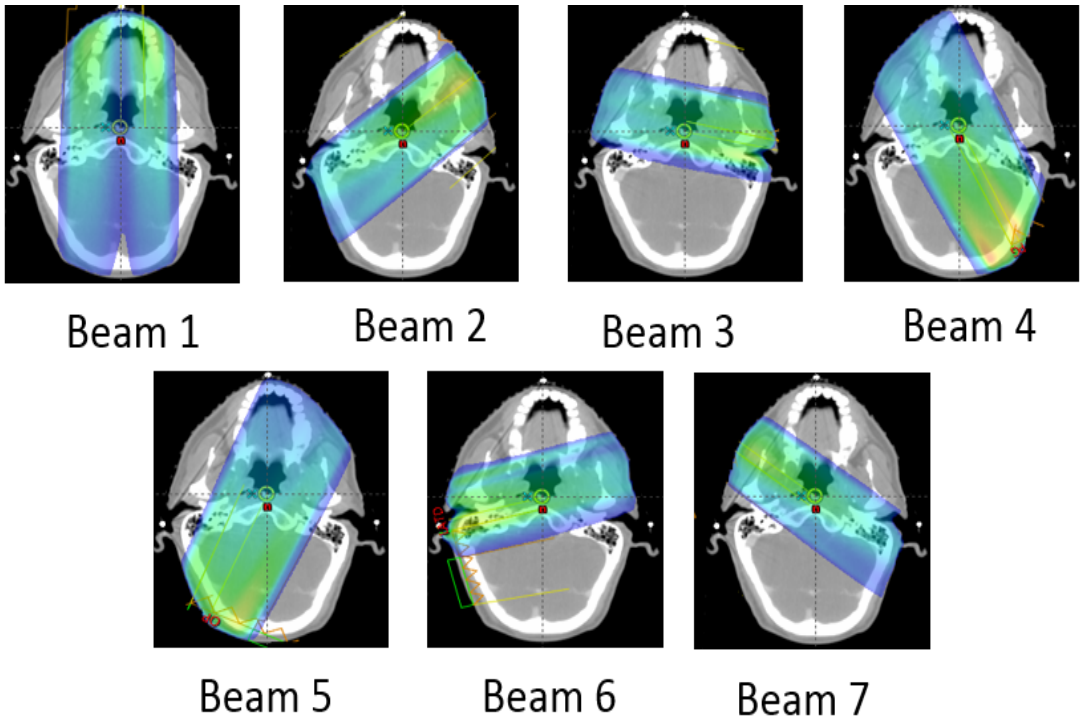


Figure 3.1: An example of the dose of each beam.

by integrating the 2D real dose images of the 7 beams. Here, each beam has a 6 [MeV] energy with a delivered dose between 0.27 and 0.34 Gy in one RT session. Through the 33 days of RT course, each NPC patient received 70 Gy as a total dose delivered to the GTV. An example of the real dose distribution of each beam is shown in Figure.3.1. The 2D real dose image of each beam and the 2D real dose distribution are constructed by radiation oncologists and used as teaching labels.

3.2 Methods

The overview of the proposed system is shown in Figure.3.2. The proposed system composed of two-phase manners: 1) extracting the feature map of each beam from contoured CT images of a patient, and 2) integrating the extracted features of all the beams to predict the dose distribution for the patient.

In the first phase, I used the encoder network of my constructed CNN, called beam dose prediction network (BDP-Net), to extract the feature map of each beam separately from contoured CT images of a patient. Here, BDP-Net is is an encoder-decoder

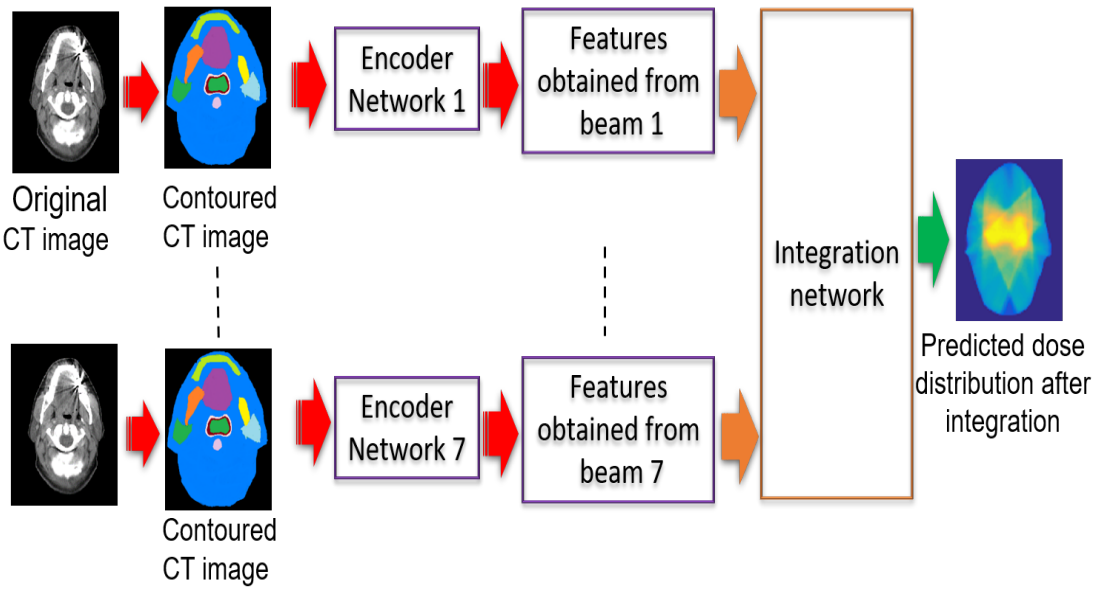


Figure 3.2: The architecture of the proposed dose distribution prediction.

network to estimate the dose irradiated by one beam when a contour CT image is given (Figure.3.3). The architecture of the BDP-Net is based on the network proposed by Fan et al.[44]. Unlike Fan et al.[44] method, the BDP-Net contains four types of blocks with less convolutional layers; dilated convolution block, identity block, convolution block and deconvolution block as shown in Figure.3.4. The output of the block is also connected with its own input through an identity mapping path. The architecture have two advantages: one is to avoid the gradient vanishing problem, and the other is to improve the gradient backward flow in the CNN. Moreover, The BDP-Net uses ReLu as the activation function of each convolutional layer and binary cross entropy function in Eq.2.1 as loss function.

The inputs of the BDP-Net are two images: 1) the image includes the contours of NPC, that is, PTV, CTV and GTV, and 2) the image contains the contours of all OARs. Therefore, the BDP-Net has two encoder networks for the NPC and the OARs. The aim of training the NPC and OARs regions separately is to obtain more detailed features of the target regions. Each of the two encoders contains 42 convolutional layers for extracting the features of its target regions started with dilated convolution block. The four dilated convolutional layers in the encoders are implemented with four rate values, r=8, 16, 24 and 32, respectively. A preliminary experiments are made to determine the rate values. The using of artous convolution layers aims to obtain

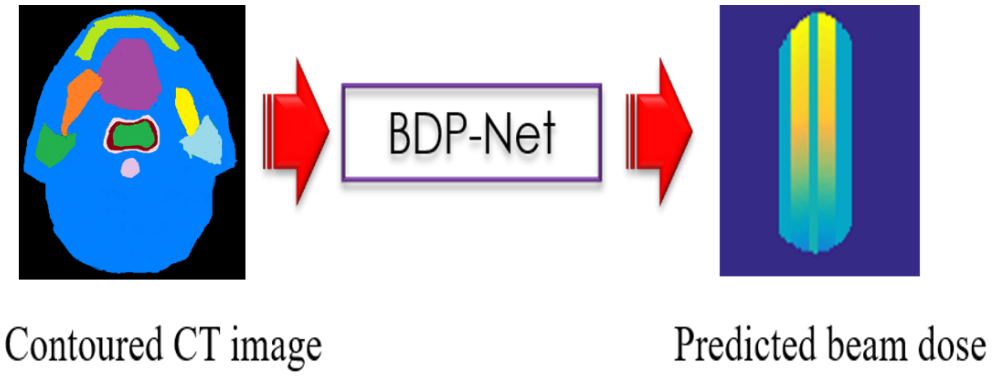


Figure 3.3: The overview of BDP-Net

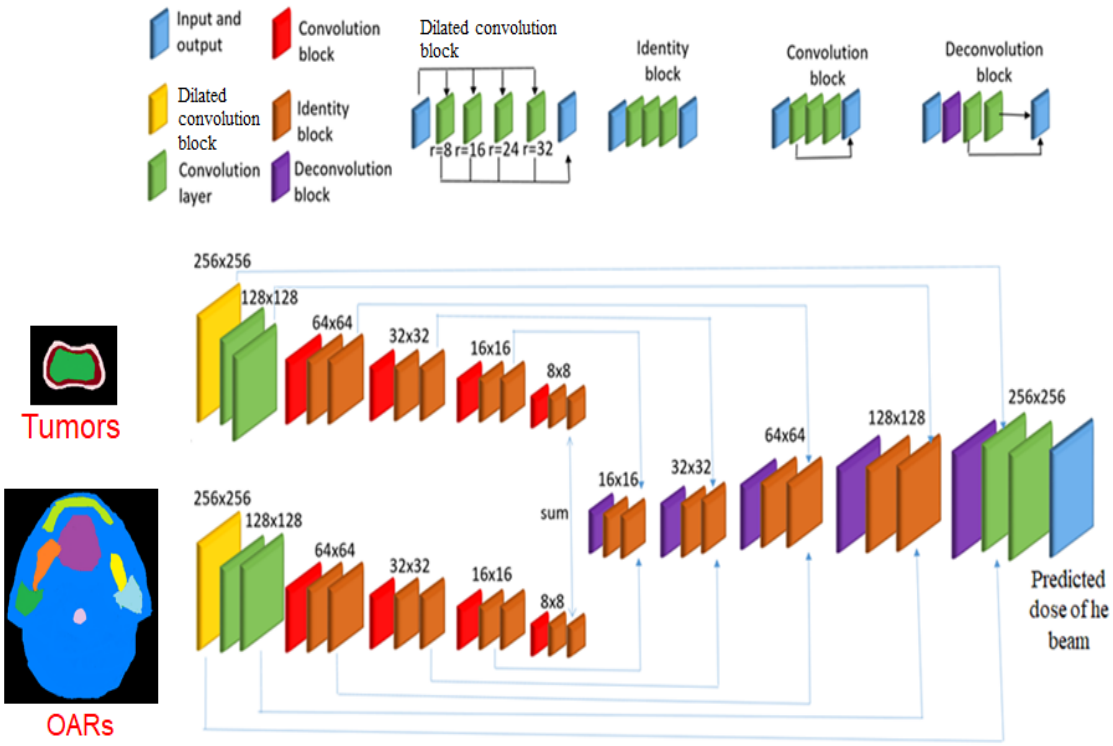
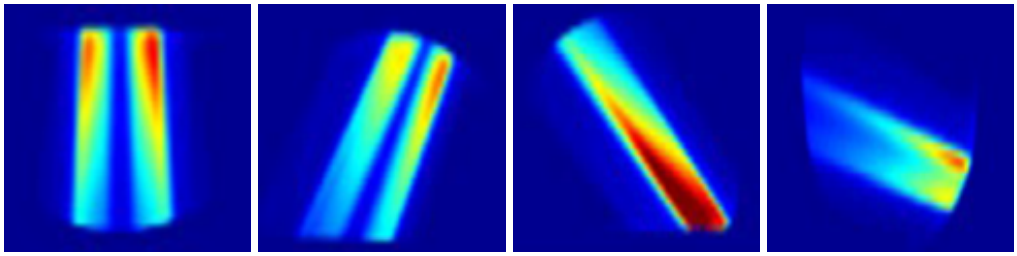
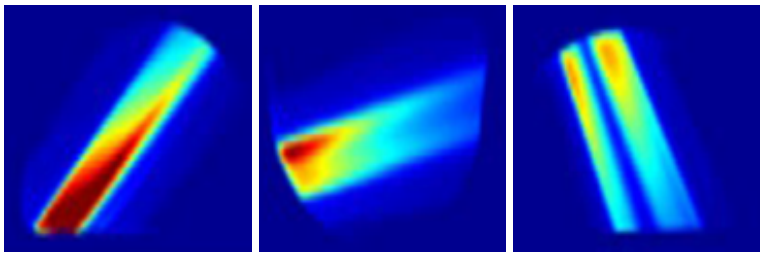


Figure 3.4: The architecture of BDP-Net.



(a) Feature map 1 (b) Feature map 2 (c) Feature map 3 (d) Feature map 4



(e) Feature map 5 (f) Feature map 6 (g) Feature map 7

Figure 3.5: Examples of the feature maps obtained from the dose distributions of beam 1 to beam 7 by using the contoured NPC and OARs images.

additional information of the corresponding region and to distinguish between high and low dose values of overlapping regions in the contoured CT images. The output of the encoder is the feature map of the input data. Using the two feature maps of the tumors and OARs, the decoder network of the BDP-Net generates the 2D image of the dose distribution delivered from the beam. In the decoder network, the convolutional and deconvolutional layers connected with multiple skip-layer connections. The skip connections tackle the problem of gradient vanishing and the recovery of the clean image. Also, it benefits from the image details passed from convolutional layers to deconvolutional layers. The decoder network outputs the predicted dose of the beam.

The second phase of the proposed system is to estimate the 2D dose distribution image by integrating the extracted feature maps from the contoured CT images of all the seven beams. The integration process introduced in the second phase is a simple CNN network that generates all the features of the seven beams, as input data, obtained in the first phase. The integration network (Figure.3.6) composed of 5 convolutional layers and one fully connected layer. The first convolution layer generates all the features of the seven beams after its concatenation. In the end of the integration network,

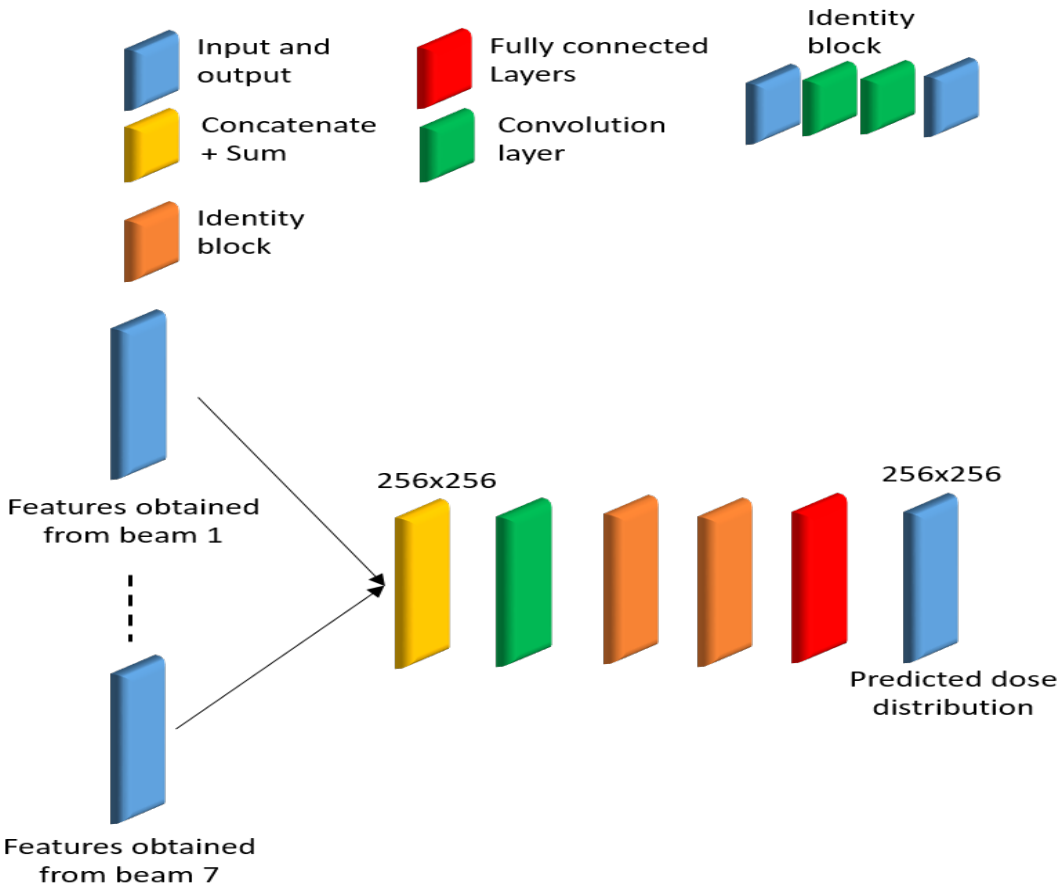


Figure 3.6: The architecture of the integration network.

the fully connected layer outputs the predicted 2D dose distribution.

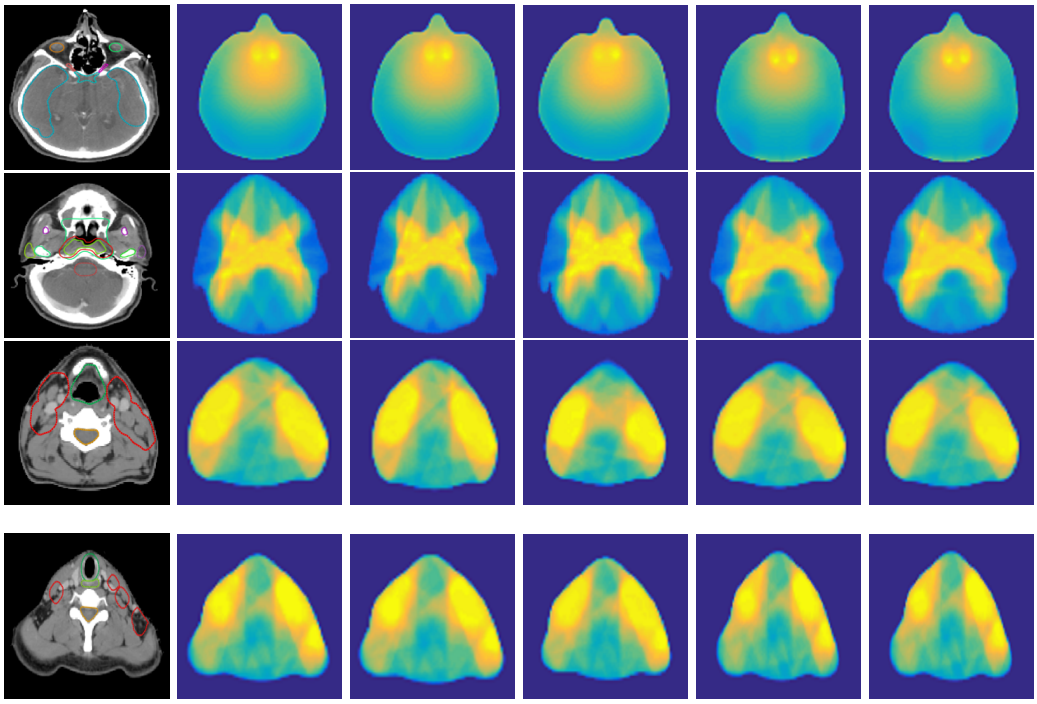
The training process of the proposed method consists of two steps. The first step is to train the BDP-Net for extracting the feature map of each beam dose. Using the encoder networks of the seven trained BDP-Nets, the second step is to train the integration network while fixing the parameters of all the encoder networks.

3.3 Experimental results

To verify the applicability of the proposed system using the integration network, the obtained results are compared with the manual dose distribution by radiation oncologists (MD) used as a ground-truth. In the experiments, 80 NPC patients is used for training and testing the proposed system. The dataset is divided into 8 sub-dataset where seventy of the eighty NPC patients were chosen as a training set to adjust the

parameters of the dose distributions prediction model. The remaining ten NPC patients were used as the test set to evaluate its performance. Therefore, 8-fold cross validation is applied to evaluate the proposed system. The proposed system is implemented by using keras and were performed on a workstation with an Intel Core i7-3960 CPU and NIVIDIA Quadro GP100 GPU. In addition, the proposed system is compared with with Mahmood et al.[42] (GAN), Chen et al.[43] (ResNet) and Fan et al.[44] (Fan) methods. Therefore, four methods are applied to predict the dose distribution. The two methods in [42] and [43] are downloaded, while I implemented the method in [44]. Here, the three comparison methods, GAN, ResNet and Fan, replaced the proposed CNN in the first phase to predict the features of each beam and trained by using my dataset.

The predicted result of dose distribution obtained by the proposed method, GAN, ResNet and Fan methods are shown in Figure.3.7(c)-(f), while Figure.3.7(a)-(b)



(a) (b) (c) (d) (e) (f)

Figure 3.7: Examples of (a)contoured CT image and its dose distribution obtained by (b) Ground-truth, (c)proposed method, (d)GAN, (e)ResNet and (f)Fan.

Table 3.2: Average of number of pixels that its dose values obtained from the proposed method, GAN, ResNet and Fan satisfy the dose constraints of tumors and OARs regions.

Region	Proposed method	GAN	ResNet	Fan
Tumors	**81.9%**	76.8%	75.6%	75.5%
OARS	**86.1%**	81.2%	82.4%	80.4%

represents the contoured CT image and the ground-truth. Generally, to irradiate the NPC patient, radiation oncologists use the dose constraint as a metric to evaluate plan quality and approve the dose distribution after it satisfies a sufficient number of the dose constraint. Therefore, to evaluate the predicted results of the dose distribution obtained from the four methods, I calculate the number of pixels of each OARs and tumor region of each patient, that the dose values of the pixels satisfy the given threshold of the region dose constraint described in Table.3.1. Table.3.2 presents the average ratio of the number of pixels to the total number of pixels of OARs and tumors regions of all patients obtained by the proposed method and the three comparison methods.

In addition, another quantitative evaluation method is used to evaluate the accuracy of the proposed network, GAN, ResNet and Fan methods, called 3D gamma analyses. A global 3D gamma analysis with 3%3mm and 4%4mm for $\gamma \leq 1$ is applied to evaluate the obtained results of dose distribution by the proposed method with the three comparison methods. The calculated values of the mean pass of 3 D gamma analysis with 3%3mm and 4%4mm are represented in Tables.3.3 and 3.4, respectively.

Table 3.3: Mean pass rates of 3D gamma analysis with 3%3mm of ours, GAN, ResNet and Fan.

Region	Proposed method		GAN		ResNet		Fan	
	Values	p-value	Values	p-value	Values	p-value	Values	p-value
L/R TMJ	**91.8 ± 0.5**	—	89.5 ± 0.2	0.316	89.9 ± 1.0	0.241	89.1 ± 1.1	0.188
Optic chiasma	**88.9 ± 0.1**	—	86.1 ± 1.2	0.213	88.2 ± 1.9	0.112	88.2 ± 0.5	0.178
L/R Lens	**92.9 ± 3.0**	—	88.2 ± 0.9	0.401	89.3 ± 1.0	0.319	88.1 ± 1.4	0.351
Larynx	**82.1 ± 0.4**	—	77.4 ± 0.6	0.344	77.9 ± 2.3	0.291	77.5 ± 2.1	0.279
Temporal lobe	**86.9 ± 1.1**	—	83.8 ± 0.9	0.316	84.4 ± 1.0	0.181	83.7 ± 0.4	0.128
L/R Mandible	**82.9 ± 0.7**	—	79.1 ± 1.4	0.288	79.7 ± 0.7	0.240	78.9 ± 0.8	0.129
L/R optic nerve	**94.5 ± 2.1**	—	92.3 ± 2.7	0.226	91.9 ± 0.9	0.386	91.9 ± 0.4	0.443
L/R parotid gland	**89.9 ± 1.4**	—	87.2 ± 1.1	0.219	87.9 ± 0.4	0.164	87.1 ± 0.9	0.127
Brainstem	**84.7 ± 1.3**	—	83.2 ± 0.4	0.113	81.1 ± 0.8	0.247	81.4 ± 1.0	0.283
Spinal cord	**87.9 ± 2.3**	—	82.1 ± 1.9	0.397	83.8 ± 0.6	0.344	82.0 ± 3.1	0.241
PTV	**93.1 ± 4.0**	—	92.2 ± 1.6	0.112	92.3 ± 0.1	0.149	92.1 ± 3.2	0.145
GTV	**97.2 ± 2.4**	—	95.9 ± 0.7	0.219	96.1 ± 1.5	0.154	95.9 ± 1.4	0.158
CTV	**96.8 ± 2.4**	—	95.0 ± 1.4	0.187	94.3 ± 0.8	0.233	94.1 ± 0.5	0.211

Table 3.4: Mean pass rates of 3D gamma analysis with 4%4mm of ours, GAN, ResNet and Fan.

Region	Proposed method		GAN		ResNet		Fan	
	Values	p-value	Values	p-value	Values	p-value	Values	p-value
L/R TMJ	**97.9 ± 0.4**	—	95.0 ± 0.4	0.233	96.1 ± 1.1	0.188	95.2 ± 0.6	0.117
Optic chiasma	**97.1 ± 0.1**	—	95.9 ± 1.1	0.214	96.2 ± 0.5	0.187	95.8 ± 1.5	0.181
L/R Lens	**98.0 ± 0.6**	—	95.1 ± 1.2	0.294	95.4 ± 1.7	0.305	95.0 ± 0.9	0.227
Larynx	**89.4 ± 1.1**	—	86.7 ± 1.0	0.209	87.5 ± 0.3	0.197	86.2 ± 1.2	0.334
Temporal lobe	**95.0 ± 0.7**	—	93.1 ± 1.2	0.183	93.9 ± 0.1	0.167	99.1 ± 0.4	0.212
L/R Mandible	**93.9 ± 0.6**	—	88.0 ± 1.9	0.462	87.5 ± 0.6	0.497	97.9 ± 2.1	0.414
L/R optic nerve	**98.8 ± 1.0**	—	95.6 ± 2.1	0.274	95.9 ± 0.7	0.228	95.2 ± 1.3	0.133
L/R parotid gland	**96.0 ± 1.4**	—	94.7 ± 0.7	0.182	94.6 ± 0.9	0.197	94.3 ± 0.7	0.112
Brainstem	**92.7 ± 1.8**	—	91.4 ± 0.7	0.157	89.4 ± 1.2	0.183	89.4 ± 0.6	0.182
Spinal cord	**95.8 ± 2.7**	—	92.8 ± 1.4	0.279	93.7 ± 0.9	0.211	92.6 ± 0.9	0.119
PTV	**99.7 ± 1.3**	—	98.2 ± 3.7	0.110	98.9 ± 3.2	0.109	98.4 ± 0.1	0.136
GTV	**98.3 ± 0.1**	—	96.3 ± 2.7	0.284	96.9 ± 1.3	0.156	96.2 ± 1.4	0.374
CTV	**99.0 ± 1.2**	—	98.6 ± 1.3	0.197	98.9 ± 3.5	0.249	98.8 ± 0.8	0.283

Furthermore, to compare the predicted results for OARs and tumor regions by the four methods with those obtained by radiation oncologist, MD, cumulative dose volume histograms (DVH) is used as a third quantitative evaluation to evaluate their differences. DVH is the dose value per fractional volume of the target region. To evaluate the accuracy of the predicted DVH, mean absolute error, MAE_{DVH}, of the difference between two DVHs of each tumors and OARs regions are calculated and presented in 3.5. MAE_{DVH} is formulated in 3.1 by:

$$MAE_{DVH} = \frac{1}{m}\sum_{k=1}^{m}\frac{1}{n}\sum_{j=1}^{n}|D_p(j)_{k\%} - D_g(j)_{k\%}|, \tag{3.1}$$

where m is the number of DVH bins separated by 0.01 Gy, k is the dose volume index of DVH, j is the 2D dose distribution image of the target region, n is the total number of images. $D_p(j)_{k\%}$ and $D_g(j)_{k\%}$ are the predicted and the ground-truth doses value at $k\%$ volume of the pixel of the j^{th} image, respectively.

An example of the predicted DVH of the four methods and MD is shown in Figures.3.8 and .3.9. Here, Figures.3.8 and .3.9 show the minimum and maximum differences between the predicted region the the four methods and MD obtained from PTV and Larynx, respectively.

Table 3.5: MAE_{DVH} of each region by ours, GAN, ResNet and Fan.

Contours	Proposed method	GAN	ResNet	Fan
L/R TMJ	**0.9** ± **0.9**	1.1± 0.2	1.2± 0.4	1.3± 0.5
Optic chiasma	**1.2** ± **0.1**	1.7± 0.6	1.3± 0.7	1.6± 0.2
L/R Lens	**1.3** ± **0.8**	1.7± 1.2	1.6± 0.3	1.8± 0.9
Larynx	**1.5** ± **0.4**	2.3± 0.8	1.8± 0.7	2.2± 0.7
Temporal lobe	**0.9** ± **0.7**	1.5± 0.3	1.2± 1.1	1.5± 0.8
L/R Mandible	**1.0** ± **0.7**	1.2± 0.9	1.3± 0.7	1.2± 1.1
L/R optic nerve	**0.8** ± **1.6**	1.3± 0.3	1.2± 0.6	1.2± 0.7
L/R parotid gland	**1.2** ± **0.8**	1.7± 0.9	1.6± 1.0	1.7± 0.1
Brainstem	**1.1** ± **0.9**	1.2± 0.4	1.6± 0.1	1.5± 0.6
Spinal cord	**1.3** ± **0.4**	1.4± 0.7	1.4± 0.6	1.4± 0.8
PTV	**0.8** ± **0.4**	1.0± 0.6	1.1± 1.7	1.1± 0.7
GTV	**0.8** ± **0.6**	1.1± 0.1	0.9± 1.6	1.1± 1.3
CTV	**0.7** ± **0.1**	1.1± 0.7	0.9± 0.5	1.1± 1.0

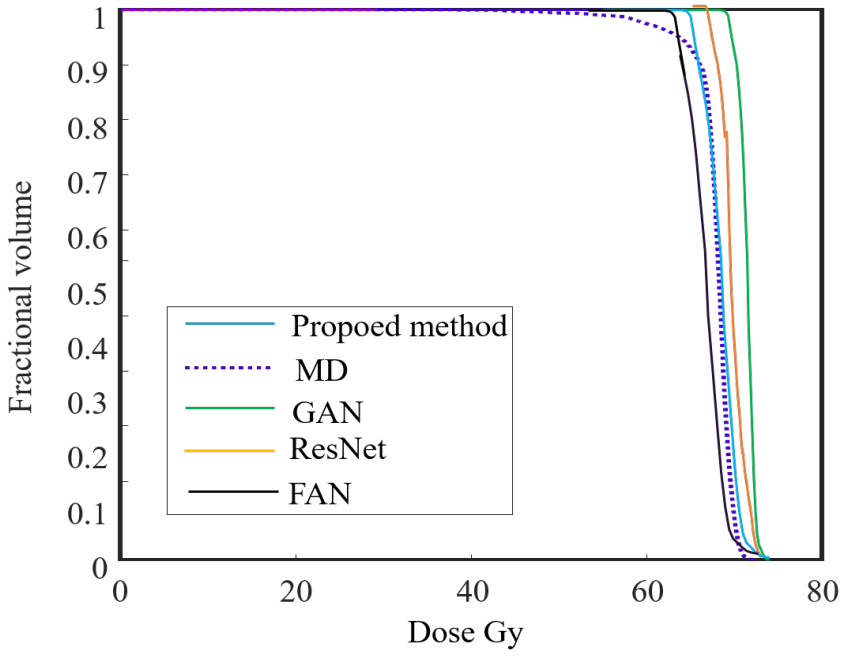


Figure 3.8: DVH of PTV by the four methods vs the ground-truth.

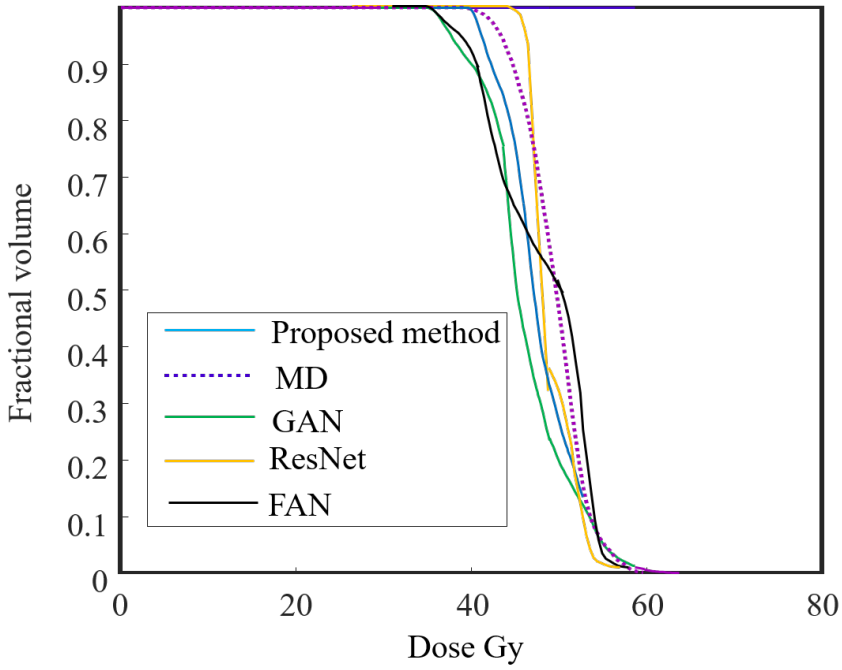


Figure 3.9: DVH of larynx by the four methods vs the ground-truth.

3.4 Discussion

According to the first experiment, as shown in Table.3.2, the accuracy of the dose values of the predicted OARs and tumors regions obtained by the proposed method is higher compared to those obtained by GAN, ResNet and Fan methods. However, the proposed method fails to satisfy all the dose constraints for tumors and OARs with 18.1%, and 13.9%, respectively. Generally, the dose distribution cannot satisfy all criteria simultaneously because the proximity of the tumors regions to the OARs regions and the complexity of the NPC site. Therefore, the obtained results by the proposed method can be acceptable.

In addition, as shown in Tables.3.3 and 3.4, the mean pass rates of 3D gamma of the proposed method is higher than the three comparison methods for all tumors and OARs regions. To compare the proposed method with GAN, ResNet and Fan methods, the statistical analysis is performed with paired t-tests (p-value(P)< 0.05). From the paired t-test, there are no significantly difference for all regions. Furthermore, the obtained values by MAE of DVH, MAE_{DVH}, in Table.3.5 show small differences between the proposed method and the ground-truth compared with comparison methods.

From these results, it is observed that the proposed method results a high accuracy,

especially for small volume of OARs, such as left and right Lens. This may due to the strategy that used in the proposed system. By training tumors and OARs regions separately, the network trains the dataset with more detailed information. Then, by using the relevant features obtained from the seven beams, the proposed system achieves a good prediction of the dose distribution.

3.5 Conclusion

In this chapter, I proposed a new method for predicting the dose distribution for NPC patients from contoured CT images considering the delivered dose from the beam. Firstly, the proposed method extracts the features of seven beams from contoured CT images. Using the obtained features from all the beam, the proposed method outputs the predicted dose distribution.

In the experiments, the proposed system achieves the best performance of predicting the dose distribution for NPC patients. In addition, the accuracy of the predicted results obtained by the proposed method is the highest compared with those obtained by the conventional methods.

4

NPC and OARs evolution prediction for ART treatment

In the current IMRT, once the initial RT planning is obtained, the initial plan is not changed during 33 days. On the contrary, during the 33 days, the anatomical structures of the patient have changed such as the shrinkage of the tumor by IMRT, the weight loss, and the displacement or sizes of the normal structures. Therefore, the re-planning of RT is needed according to the current patient condition.

To achieve this, in this chapter, I propose a new system, called Tumor Evolution Prediction(TEP)-Net, for predicting the evolution of NPC and OARs of *n*-th week in response to RT from previous weekly CT images from *1*st to (*n-1*)-th week. The proposed TEP-Net is composed of CNN and GRUs. In addition, TEP-Net is applied on three section images: axial, coronal and sagittal sections, separately. By using the three different sections, the proposed method learns the 3D structure of the NPC and OARs by different features extracted from the three views in the CT images. Therefore,

the use of the 3D structure leads to the improvement of the accuracy of predicting the evolution of NPC and OARs.

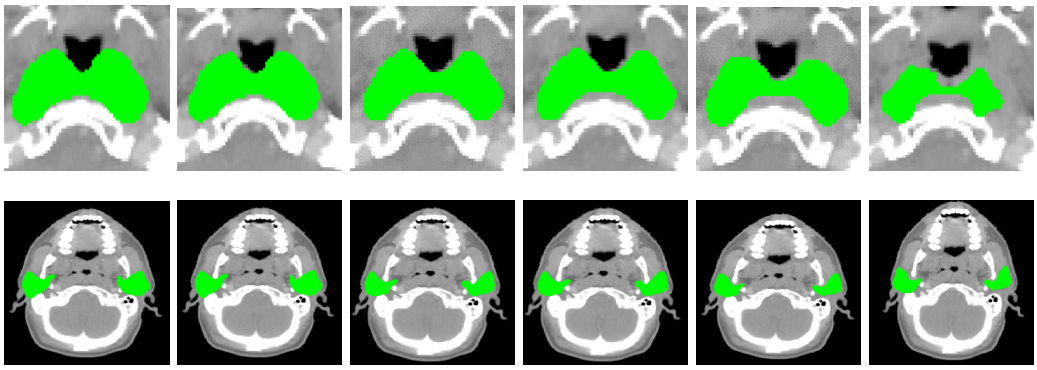
4.1 Dataset

The proposed method is based on weekly CT images, with size 512×512[pixels], constructed by using Philips Brilliance 16 Big Bore CT Scanner at Radiotherapy Department of Habib Bourguiba Hospitalof Sfax-Tunisia. A total of 140 patients with NPC is collected from 2017 to 2019 and used as dataset. Practically, to destroy the tumor cells, all NPC patients received 2.12 Gy/fraction during 33 fraction RT course of 7 weeks after one month of chemotherapy to avoid the growth of the NPC. Through this methodology, the characteristic of NPCs, such as size, shape and location, extracted one month before are similar to those in the first day of RT course. Each axial CT image is constructed with 3 mm slice thickness, no intersection gap and 180 slices were weekly acquired with a resolution of $1.16 \times 1.16 \times 3mm^3$. In addition, the weekly CT images were rigidly registered to the first CT image obtained before the RT course by respecting the same parameters of the gantry isocenter and the degree of flexion of the head. This aims to improve the quality of the fusion between the weekly CT images. In each weekly CT image, two radiation oncologists manually extract the NPC and OARs regions. Here, only five OARs, spinal cord, optic nerve, parotid gland, brainstem and lens, are used as training and testing data. In addition, each region is trained separately as a teaching label to predict its evolution in week n.

In the 140 NPC patients, the weekly shrinkage of NPC occurred between weeks 2 and 5. This is observed from the quantitative evaluation of the difference between the initial NPC volume before RT course and the weekly shrinkage of the NPC volume in Table.4.1. In addition, an example of the evolution of NPC and OARs change during RT treatment is illustrated in Figure.4.1. In Figure.4.1, the top row represents the NPC regions from week 1 to 6, while the button row is the weekly change of OARs regions.

Table 4.1: The tumor's volume change ratio of the NPC volume at week n to the initial NPC volume before RT course (week 0) (mean ± std).

Contours	**Week 1**	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**
NPC	14% ± 9%	23% ± 10%	31% ± 6%	41% ± 14%	52% ± 22%	58% ± 19%



(a) (b) (c) (d) (e) (f)

Figure 4.1: Examples of the evolution of NPC (Top row) and OARs (button row) during, (a)week 1, (b)week 2, (c)week 3, (d)week 4, (e)week 5 and (f)week 6, in response to RT.

4.2 Methods

4.2.1 Overview

The proposed system, TEP-Net, was specifically designed to predict the future evolution of NPC and OARs in week n of RT course. TEP-Net consists of the following three major components: CNN, GRU and global attention model. The proposed system is illustrated in Figure.4.2. As shown in Figure.4.2, the first component, CNN is composed of *n-1*-CNNs that train the weekly CT images obtained from week 1 to week *n-1*, respectively, as input data. This aims to extract the relevant features on the weekly CT images. Through the obtained features from the *n-1* CNNs, the second component of TEP-Net is GRU network. Here, GRU can simple store and process the trajectories of NPC and OARs evolution in response to RT treatment with less training data. In addition, GRU outputs the useful features for predicting the evolution of NPC and OARs in week n after removing the irrelevant features obtained from the *n-1*-CNNs. To do that, the proposed GRU is used as an encoder-decoder network. For each k-th (k=1,..,n-1) weekly CT image, the first GRU, $GRU_1{}^k$, is the encoder network, while the second GRU, $GRU_2{}^k$, is the decoder network. In the end of TEP-Net, the last component, global attention model is used to weight the importance of weekly CT images and to produce the final predictions of the evolution of NPC and OARs at the n -th week. The details of the three components are described in the following.

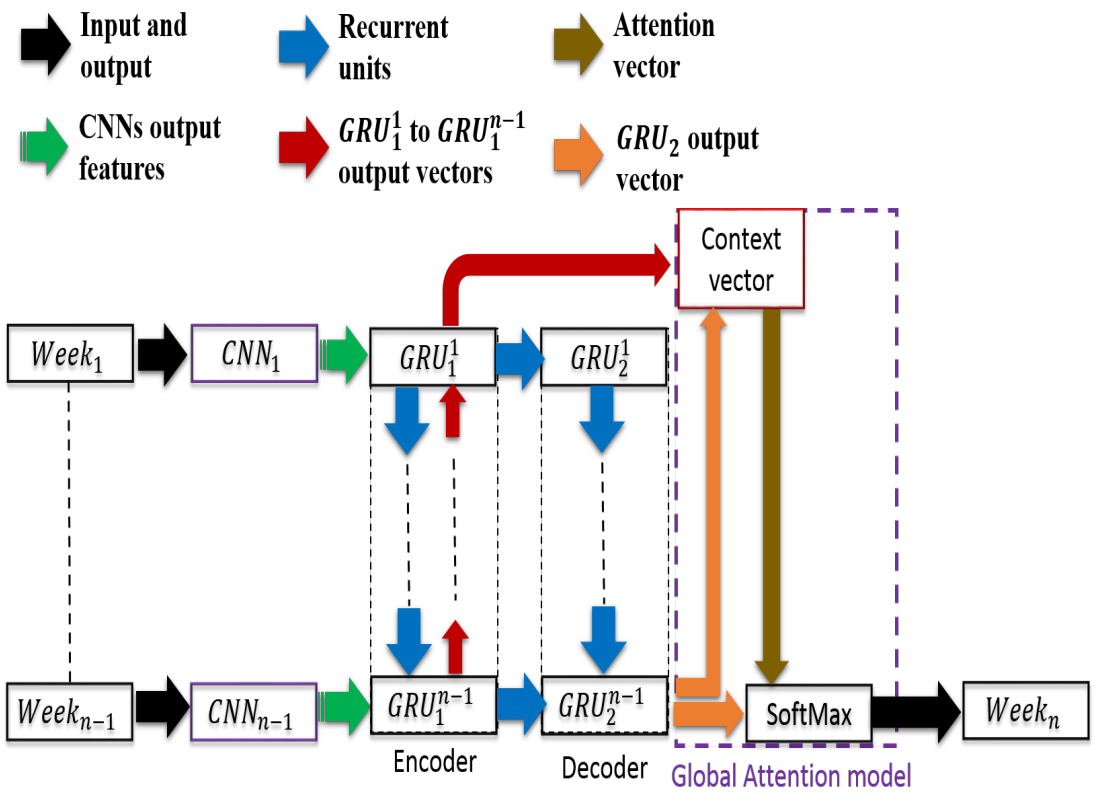


Figure 4.2: The architecture of the proposed method, TEP-Net, using one of three directional images.

4.2.2 CNN construction

As mentioned above, the first component of TEP-Net is CNN network. In Chapter.2, the proposed CNNs show a powerful tool for segmenting NPC and OARs regions. In addition, the use of overlapping patches as input data aims to improve the accuracy of the segmented results of the target region compared with the direct segmentation by using the whole CT image. Based on that, the constructed CNN in TEP-Net uses overlapping patches with various sizes as input data. Same as Chapter.2, 16×16, 32×32 and 64×64[pixels] are the size of the patches, extracted from the CT image by sliding a window from left to right and top to bottom, used as input data. Moreover, to obtain more global information about the weekly NPC and OARs regions, CNN trains the whole CT image as additional information. Therefore, the proposed CNN is composed of four CNNs with different architectures to train the corresponding input data. Through the proposed CNN, the four sub-CNNs learn the label of a pixel by

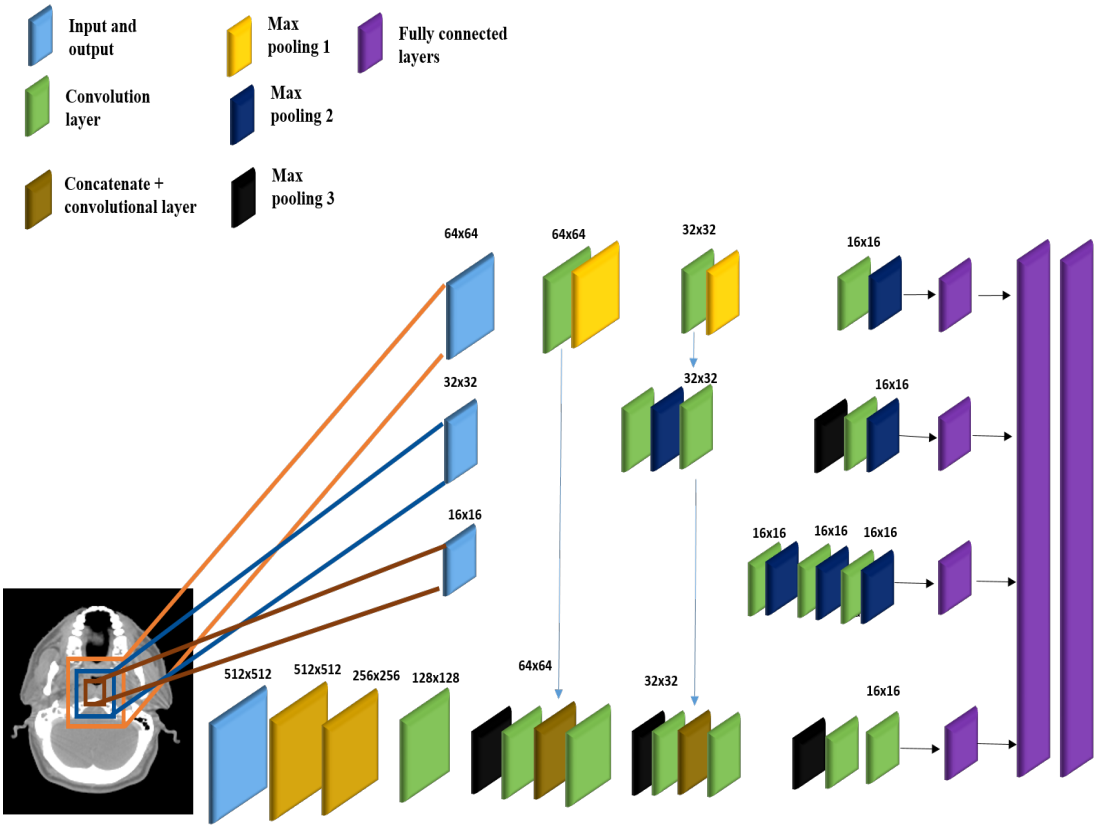


Figure 4.3: CNN construction.

considering both local and global structures. This aims to facilitate the work of the next components, GRUs and global attention model. The overview of the CNN is shown in Figure.4.3.

In Figure.4.3, the three CNNs used to train the three types of overlapping patches, 16×16, 32×32 and 64×64. The three CNNs have the same architecture as those CNNs in Figures.2.5, 2.7 and 2.8, used to segment the NPC and OARs regions in Chapter.2. The ouput of three CNNs is feature maps with size 16×16. On the other hand, the CNN, that its input is the whole CT image, is composed of 12 convolutional layers, 3 max pooling and 1 fully connected layer. All convolutional layers consist of a window size of 3×3, a stride of 1 and the same padding, while the 3 max pooling were performed over a 2×2 kernel with stride size of 2. Furthermore, the obtained features with output size 64×64, 32×32 in the three sub-CNNs, that their inputs are 64×64, 32×32 patches respectively, are added to the 5^{th} and 9^{th} convolutional layers in the fourth sub-CNN. This aims to get more useful features. The output of

four-CNN is also feature maps with size 16×16. In the end, to output the relevant features that contain temporal and spatial information for predicting the evolution of NPC and OARs, the obtained features from the four sub-CNNs are integrated into two fully connected layers.

4.2.3 Gated recurrent unit for learning the sequential features

After the extraction of high-level features in the CNNs from the weekly CT images, two GRUs, GRU_1 and GRU_2 as an encoder-decoder network, are implemented to store and process the evolution of NPC and OARs in response to RT treatment. Figure.4.4 shows the architecture of the original GRU. Practically, GRU includes the smaller number of training parameters and smaller memory. Therefore, GRU is useful to train a network by using a small number of dataset and train more faster.

In order to analyze the status of the target region between the obtained features from the weekly CT images, I modify the original GRU (Figure.4.5) to output only the new hidden state and to be used it in recursive route. Here, the two GRUs contain (*n-1*) sub-GRUs with one layer of 512 hidden units to strengthen the processing power. As shown in Figure.4.2, the encoder GRU trains the features obtained from the weekly CT images. The decoder GRU trains the output of the encoder GRU. The encoder and decoder GRUs have the same architecture. In the encoder GRU, at week *i*, the individual obtained feature passed through the encoder GRU to output the encoder hidden state, h_1^i, obtained from the corresponding week. The obtained encoder hidden state, h_1^i, is used as the input of the previous hidden state of the encoder GRU at week *i+1*. In the decoder GRU, at week i, the obtained encoder hidden state, h_1^i, passed through the decoder GRU to output the decoder hidden state, h_2^i. The decoder hidden state, h_2^i is used as the input of the previous hidden state of the decoder GRU at week *i+1*. Therefore, each (*n-1*) sub-GRUs trains two inputs: the obtained features with the encoder hidden state in the encoder GRU or the encoder hidden state with the decoder hidden state in the decoder GRU. Here, the two inputs of each (*n-1*) sub-GRUs in the encoder and decoder GRUs are modulated via two gates, namely a reset gate r and an update gate z. The reset gate r determines the portion of the past information to be forgotten. Otherwise, the update gate z decides the portion of the past information needs to be passed along to the future. Here, at a given time t, the inputs of the reset gate r and the update gate z are the given features in $week_t$, x_t, and the previous hidden state, h_{t-1}. The update gate z, the reset gate r and the candidate state of the hidden unit

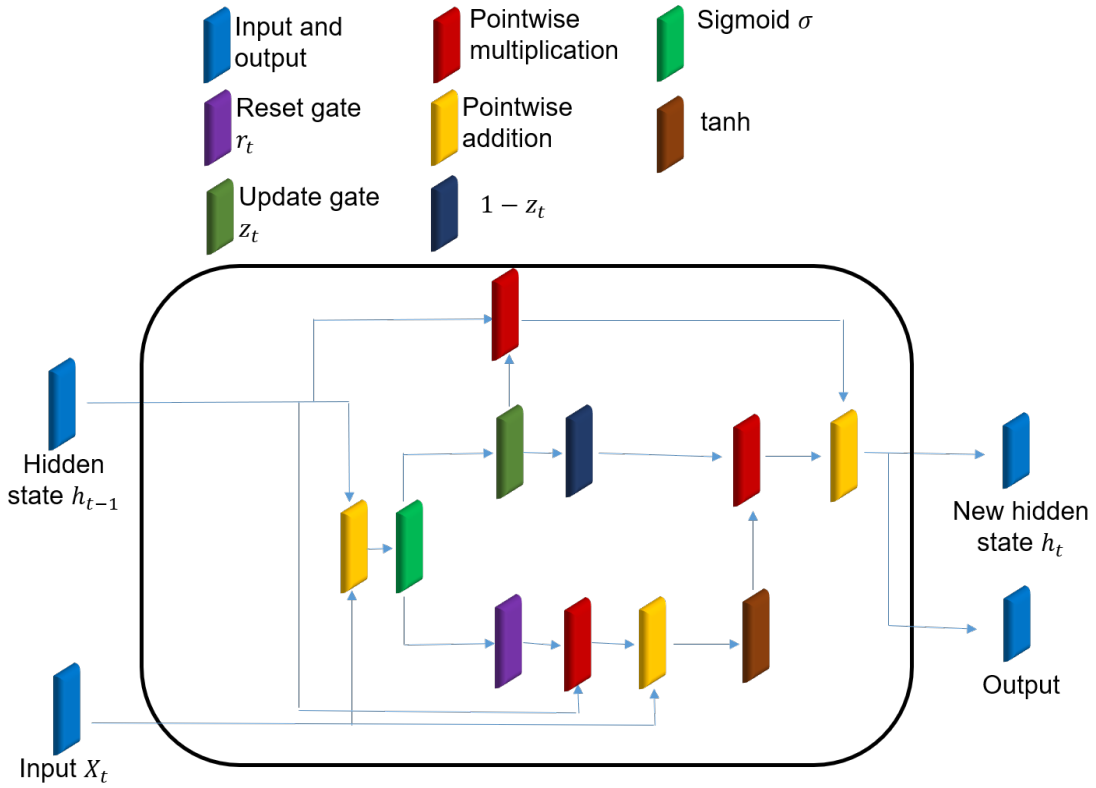


Figure 4.4: The architecture of the original GRU.

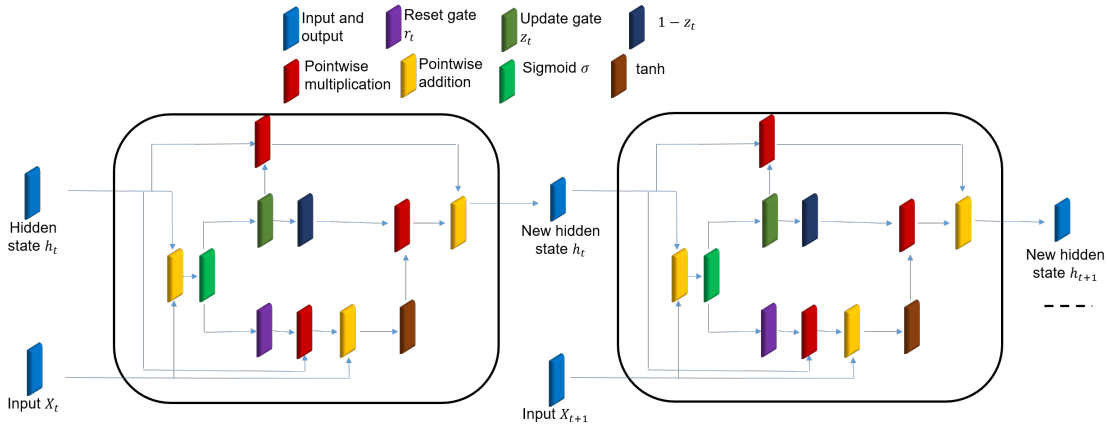


Figure 4.5: The architecture of the proposed encoder or decoder GRU

$\widetilde{h_t}$ are computed in Eq.4.1, 4.2 and 4.3, respectively at a given time t. The output of hidden activation state h_t is represented in Eq.4.4. Through the two GRUs, the second GRU outputs the refined feature vector for predicting the evolution of NPC and OARs after removing the irrelevant information.

$$z_t = \sigma(W^{xz}x_t + W^{hz}h_{t-1}); \tag{4.1}$$

$$r_t = \sigma(W^{xy}x_t + W^{hr}h_{t-1}); \tag{4.2}$$

$$\widetilde{h_t} = \tan(W^{xh}x_t + W^{hh}(h_{t-1} \odot r_t)), \tag{4.3}$$

$$h_t = (1 - z_t) \odot \widetilde{h_t} + z_t \odot h_{t-1}. \tag{4.4}$$

where W^{xz}, W^{hz}, W^{xr} and W^{hh} are the corresponding weight metrics, σ is the sigmoid function and $\odot$ is an element wise multiplication.

4.2.4 Global attention model

After the obtained of a refined feature vector from the two GRUs, the global attention model is introduced as the last component of the proposed TEP-Net. Generally, an attention model is used to parse the weights of the inputs which were used to calculate the prediction. Therefore, the global attention model is applied to approximate all the attention distribution weights, $a_{t,i}$, of the obtained weights of all the hidden states from the first GRU, that is the encoder network, and generate a context vector c_t that is calculated in Eq.4.5 by :

$$c_t = \sum_{i=1}^{i} a_{t,i}h_i, \tag{4.5}$$

Where $a_{t,i}$ is an alignment vector derived by comparing the current target hidden state h_t with each source hidden state h_i calculated by :

$$a_{t,i} = \frac{exp(score(h_t, h_i))}{\sum_{i=1}^{T} exp(score(h_t, h_i))}, \tag{4.6}$$

$$score(h_t, h_i) = \upsilon_a{}^t \tanh(W_a[h_t, h_i]) \tag{4.7}$$

Subsequently, the decoder network, 2^{nd} GRU, ouputs the probability of the target region in the future week, week n, according to the previous outputs and hidden states

via a SoftMax layer calculated in Eq.4.8 by :

$$p(y_t|y<t,x) = sotmax(W_{out}o_t); \tag{4.8}$$

$$o_t = \tanh(W_c[h_t;c_t]), \tag{4.9}$$

where W_{out} and W_c are weighting parameters.

4.2.5 Integration process for 3D prediction

In order to predict the evolution of NPC and ORAs at n-th-week, TEP-Net is trained from three directional images: axial, coronal and sagittal images. Therefore, three different results are obtained from the corresponding directional image. To estimate the final prediction, an integration process is used to incorporate the three predictions obtained from the three directional images. In this integration process, two different methods of integrating the three predictions is implemented. The first method uses weighted voting method, called TEP-Net-WV, while the second method predict the final result by using fully connected networks, called TEP-Net-FC. Here, the final prediction result by using TEP-Net-WV is the same method as the integration of segmentation results from three directional sections mentioned in Chapter.2, where the pixel is considered as target region if the pixel is obtained at least from two sections. However, TEP-Net-FC predicts the target region by incorporating the obtained features of the three directional images by two fully connected layers and one Softmax.

4.2.6 Implementation

Thr proposed method is implemented with keras library in Python. The training of my network is achieved by cross entropy loss function which is optimized by ADAM, Eq.4.10 to 4.13, with a learning rate 10^{-4} for both CNN and GRU.

$$w_{t+1} = w_t + \Delta w_t, \tag{4.10}$$

$$\Delta w_t = -\eta \frac{v_t}{\sqrt{s_t+\varepsilon}} * g_t, \tag{4.11}$$

$$v_t = \beta_1 * v_{t-1} - (1-\beta_1) * g_t, \tag{4.12}$$

$$s_t = \beta_2 * s_{t-1} - (1-\beta_2) * g_t^2; \tag{4.13}$$

where w_{t+1} and w_t are weighting parameters, η is initial learning rate, β_1 and β_2 are hyperparameters and g_t indicates the element wise. The v_t corresponds to the exponential average of g_t along w_t, while s_t is the exponential average of squares of g_t along w_t.

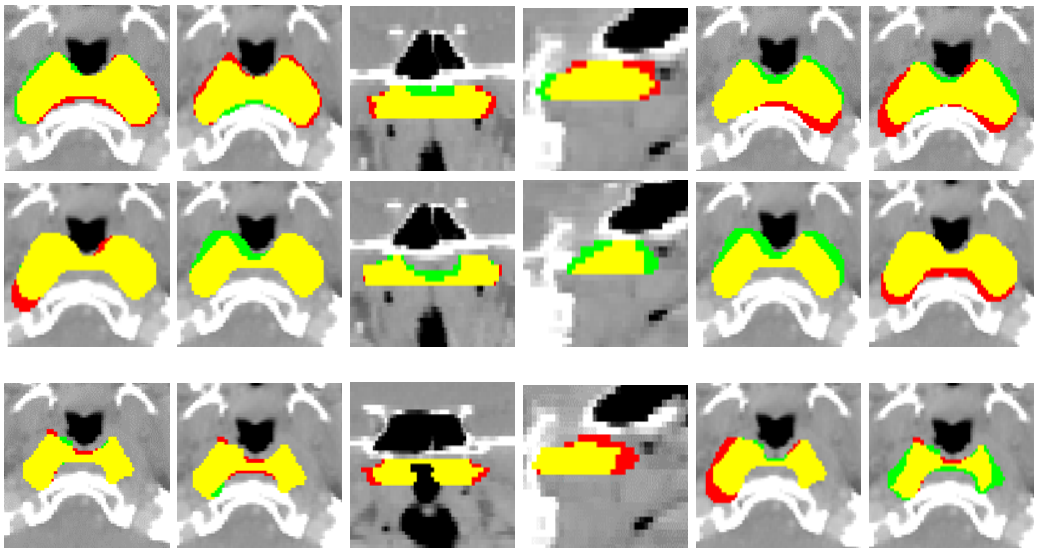
To avoid over-fitting problem, L1 and L2 regulations and dropout are used in the CNNs that its inputs are three different size of overlapping patches. The proposed method becomes stable after iteration of 100 epochs. After each convolutional layer in the four CNNs, a rectified linear unit (ReLU) is used as an activation function, while Tanh activation is used for the GRU hidden unit. All experiments were performed on a workstation with Intel Core i7-3960X CPU and NVIDIA Quadro GP100 GPU with 12 GB memory for each GPU.

4.3 Experimental results

4.3.1 Results

Table 4.2: Average and standard deviation of the accuracy of predicting NPC in (a) week 4, (b) week 5 and (c) week 6 by my five methods and P-net.

Method	Precision	Recall	DSC		RMSSD	
			Values	p-value	Values	p-value
(a)Week 4						
TEP-Net-FC	**0.86 ± 0.33**	**0.87 ± 0.25**	**0.86 ± 0.39**	-	**1.39 ± 0.10**	-
TEP-Net-WV	0.85± 0.16	0.85 ± 0.28	0.85 ± 0.24	N.S.	1.42 ± 0.32	N.S.
TEP-Net sagittal	0.84 ± 0.36	0.83 ± 0.17	0.83 ± 0.41	< 0.05	1.49 ± 0.50	< 0.05
TEP-Net axial	0.82 ± 0.21	0.83 ± 0.32	0.82 ± 0.14	< 0.05	1.56 ± 0.31	< 0.05
TEP-Net coronal	0.83 ± 0.13	0.82 ± 0.54	0.82 ± 0.92	< 0.05	1.57 ± 0.17	< 0.05
P-net	0.77 ± 0.11	0.78 ± 0.48	0.77 ± 0.31	< 0.05	2.01 ± 0.14	< 0.05
(b)Week 5						
TEP-Net-FC	**0.89 ± 0.037**	**0.88 ± 0.024**	**0.89± 0.044**	-	**1.18 ± 0.20**	-
TEP-Net-WV	0.88 ± 0.08	0.87 ± 0.011	0.88 ± 0.025	N.S.	1.23 ± 0.18	N.S.
TEP-Net sagittal	0.86 ± 0.07	0.86 ± 0.41	0.86 ± 0.012	< 0.05	1.38 ± 0.40	< 0.05
TEP-Net axial	0.86 ± 0.022	0.84 ± 0.27	0.86 ± 0.12	< 0.05	1.4 ± 0.15	< 0.05
TEP-Net coronal	0.84 ± 0.11	0.84 ± 0.07	0.84 ± 0.13	< 0.05	1.5 ± 0.09	< 0.05
P-net	0.82 ± 0.43	0.81 ± 0.04	0.82 ± 0.61	< 0.05	1.7 ± 0.70	< 0.05
(c)Week 6						
TEP-Net-FC	**0.82 ± 0.27**	**0.83 ± 0.51**	**0.82 ± 0.44**	-	**1.57 ± 0.41**	-
TEP-Net-WV	0.82 ± 0.16	0.81 ± 0.32	0.81 ± 0.42	N.S.	1.58 ± 0.17	N.S.
TEP-Net sagittal	0.79 ± 0.27	0.80 ± 0.11	0.79 ± 0.61	< 0.05	2.2 ± 0.24	< 0.05
TEP-Net axial	0.78 ± 0.43	0.78 ± 0.20	0.78 ± 0.63	< 0.05	2.4 ± 0.22	< 0.05
TEP-Net coronal	0.77 ± 0.54	0.76 ± 0.71	0.77 ± 0.22	< 0.05	2.5 ± 0.08	< 0.05
P-net	0.73 ± 0.14	0.72 ± 0.55	0.72 ± 0.39	< 0.05	2.9 ± 0.14	< 0.05



(a) (b) (c) (d) (e) (f)

Figure 4.6: NPC predicted results in week 4 (Top row), week 5 (middle row) and week 6 (button row) obtained from (a)TEP-Net-FC, (b)TEP-Net-WV, (c)TEP-Net sagittal, (d)TEP-Net coronal, (e)TEP-Net axial and (f)P-net. (Red) the segmented non-target organ; (Green and Yellow) the matching (green) and non-matching (yellow) regions between the segmented non-target organ and the ground-truth.

In the experiments, the proposed method, TEP-Net, is applied to predict the evolution of NPC and OARs in week 4, 5 and 6 from weekly CT images obtained in week 1 to 3, week 1 to 4 and week 1 to 5, respectively. The weekly CT images are collected from 140 patients and are divided into training, testing and validating dataset. Here, 120 patients are used as training dataset where 10 patients are used as testing and validating dataset. In the experiments, 14 fold-cross validation are used: 12 folds for training, 1 fold for validation and 1 fold for testing. The validation set is used for the optimization of the training process. As mentioned before, TEP-Net predicts the evolution of NPC and OARs from three orthogonal sections. Therefore, TEP-Net trains a dataset that contains 151,200 axial, 210,150 coronal, and 294,500 sagittal images.

The proposed system, TEP-Net, is only compared with the method proposed by Wang et al. [46], P-net, since ART is a future approach. Here, P-net trains 3D overlapping patches with size $45 \times 45 \times 3$[pixels] obtained from CT images in week 1 to 3. The 3D overlapping patches are constructed by cropping out three consecutive two-dimensional axial section from above and below neighbor images with window size 45×45. The output of P-Net is the evolution of NPC and OARs in week 4 to 6.

Table 4.3: Average and standard deviation of the accuracy of predicting spinal cord in (a) week 4, (b) week 5 and (c) week 6 by our five methods and P-net.

Method	Precision	Recall	DSC		RMSSD	
			Values	p-value	Values	p-value
(a)Week 4						
TEP-Net-FC	$\mathbf{0.87 \pm 0.31}$	$\mathbf{0.85 \pm 0.12}$	$\mathbf{0.86 \pm 0.51}$	-	$\mathbf{1.4 \pm 0.7}$	-
TEP-Net-WV	0.86 ± 0.22	0.82 ± 0.18	0.84 ± 0.07	N.S.	1.5 ± 1.1	N.S.
TEP-Net sagittal	0.84 ± 0.091	0.84 ± 0.10	0.84 ± 0.09	< 0.05	1.4 ± 0.7	< 0.05
TEP-Net axial	0.82 ± 0.17	0.81 ± 0.15	0.81 ± 0.22	< 0.05	1.6 ± 0.4	< 0.05
TEP-Net coronal	0.80 ± 0.18	0.80 ± 0.23	0.80 ± 0.17	< 0.05	1.8 ± 0.9	< 0.05
P-net	0.78 ± 0.14	0.76 ± 0.21	0.77 ± 0.14	< 0.05	2.6 ± 1.1	< 0.05
(b)Week 5						
TEP-Net-FC	$\mathbf{0.88 \pm 0.23}$	$\mathbf{0.86 \pm 0.25}$	$\mathbf{0.87 \pm 0.15}$	-	$\mathbf{1.3 \pm 0.7}$	-
TEP-Net-WV	0.85 ± 0.15	0.86 ± 0.16	0.85 ± 0.41	N.S.	1.4 ± 0.8	N.S.
TEP-Net sagittal	0.83 ± 0.34	0.83 ± 0.16	0.83 ± 0.12	< 0.05	1.6 ± 1.4	< 0.05
TEP-Net axial	0.82 ± 0.09	0.83 ± 0.15	0.82 ± 0.31	< 0.05	1.7 ± 1.5	< 0.05
TEP-Net coronal	0.82 ± 0.12	0.82 ± 0.41	0.82 ± 0.17	< 0.05	1.7 ± 0.9	< 0.05
P-net	0.80 ± 0.12	0.80 ± 0.14	0.80 ± 0.17	< 0.05	1.8 ± 0.9	< 0.05
(c)Week 6						
TEP-Net-FC	$\mathbf{0.85 \pm 0.08}$	$\mathbf{0.81 \pm 0.27}$	$\mathbf{0.83 \pm 0.13}$	-	$\mathbf{1.4 \pm 0.7}$	-
TEP-Net-WV	0.82 ± 0.17	0.81 ± 0.24	0.81 ± 0.19	N.S.	1.8 ± 1.8	N.S.
TEP-Net sagittal	0.80 ± 0.32	0.80 ± 0.12	0.80 ± 0.13	< 0.05	1.8 ± 0.6	< 0.05
TEP-Net axial	0.78 ± 0.33	0.80 ± 0.01	0.79 ± 0.29	< 0.05	2.2 ± 1.4	< 0.05
TEP-Net coronal	0.78 ± 0.11	0.78 ± 0.09	0.78 ± 0.14	< 0.05	2.5 ± 0.2	< 0.05
P-net	0.76 ± 0.32	0.74 ± 0.21	0.75 ± 0.07	< 0.05	2.8 ± 1.2	< 0.05

To verify the prediction of the target region, in week 4, 5 and 6, against the manual contouring of radiation oncologists, the six methods, TEP-Net axial, TEP-Net coronal, TEP-Net sagittal, TEP-Net-WV, TEP-Net-FC and P-net, are evaluated by using four measurements: precision(Eq.2.2), recall(Eq.2.3), DSC(Eq.2.4) and root mean square symmetric surface distance (RMSSD)(Eq.4.14).

$$RMSSD = \sqrt{\frac{1}{|S|}\sum_{\mathbf{p}\in S} d_1(\mathbf{p},\mathbf{q})^2} + \sqrt{\frac{1}{|\widetilde{S}|}\sum_{\mathbf{q}\in \widetilde{S}} d_2(\mathbf{q},\mathbf{p})^2}; \tag{4.14}$$

where the parameters of Eq.4.14 are described in Eq.2.5

In the experiments, firstly, I calculate the quantitative results of the predicted shrinkage of NPC in response to RT at week 4, 5 and 6 by the six methods when the inputs of the five methods based on TEP-Net are week 1 to 3, week 1 to 4 and week 1 to 5, respectively and week 1 to 3 by P-net method. The calculated values of week 4 to 6 are presented in Table.4.2 (a)-(c), respectively. In addition, an example

Table 4.4: Average and standard deviation of the accuracy of predicting optic nerve in (a) week 4, (b) week 5 and (c) week 6 by our five methods and P-net.

Method	Precision	Recall	DSC		RMSSD	
			Values	p-value	Values	p-value
(a)Week 4						
TEP-Net-FC	$\mathbf{0.83 \pm 0.13}$	$\mathbf{0.81 \pm 0.09}$	$\mathbf{0.82 \pm 0.11}$	-	$\mathbf{1.7 \pm 1.5}$	-
TEP-Net-WV	0.82 ± 0.19	0.80 ± 0.27	0.81 ± 0.13	N.S.	1.7 ± 0.8	N.S.
TEP-Net sagittal	0.80 ± 0.41	0.79 ± 0.29	0.79 ± 0.21	< 0.05	1.8 ± 0.9	< 0.05
TEP-Net axial	0.80 ± 0.40	0.77 ± 0.18	0.78 ± 0.34	< 0.05	2.0 ± 0.7	< 0.05
TEP-Net coronal	0.76 ± 0.14	0.77 ± 0.05	0.76 ± 0.25	< 0.05	2.7 ± 1.8	< 0.05
P-net	0.73 ± 0.31	0.72 ± 0.12	0.72 ± 0.13	< 0.05	3.1 ± 1.4	< 0.05
(b)Week 5						
TEP-Net-FC	$\mathbf{0.84 \pm 0.11}$	$\mathbf{0.83 \pm 0.29}$	$\mathbf{0.83 \pm 0.27}$	-	$\mathbf{1.6 \pm 0.7}$	-
TEP-Net-WV	0.82 ± 0.22	0.82 ± 0.17	0.82 ± 0.14	N.S.	1.8 ± 0.3	N.S.
TEP-Net sagittal	0.80 ± 0.41	0.80 ± 0.26	0.80 ± 0.18	< 0.05	1.8 ± 1.2	< 0.05
TEP-Net axial	0.77 ± 0.41	0.78 ± 0.19	0.77 ± 0.14	< 0.05	2.2 ± 1.4	< 0.05
TEP-Net coronal	0.78 ± 0.17	0.78 ± 0.25	0.78 ± 0.06	< 0.05	2.2 ± 0.7	< 0.05
P-net	0.76 ± 0.29	0.76 ± 0.14	0.76 ± 0.18	< 0.05	2.7 ± 0.5	< 0.05
(c)Week 6						
TEP-Net-FC	$\mathbf{0.81 \pm 0.22}$	$\mathbf{0.80 \pm 0.16}$	$\mathbf{0.80 \pm 0.23}$	-	$\mathbf{1.7 \pm 0.8}$	-
TEP-Net-WV	0.78 ± 0.18	0.78 ± 0.84	0.78 ± 0.91	N.S.	2.4 ± 0.8	N.S.
TEP-Net sagittal	0.77 ± 0.18	0.76 ± 0.41	0.76 ± 0.23	< 0.05	2.7 ± 0.9	< 0.05
TEP-Net axial	0.73 ± 0.15	0.77 ± 0.11	0.75 ± 0.25	< 0.05	2.8 ± 1.1	< 0.05
TEP-Net coronal	0.73 ± 0.21	0.74 ± 0.19	0.73 ± 0.19	< 0.05	3.3 ± 0.1	< 0.05
P-net	0.71 ± 0.33	0.70 ± 0.17	0.70 ± 0.04	< 0.05	3.7 ± 2.1	< 0.05

of NPC regions in week 4, 5 and 6 is shown in Figures.4.6. Secondly, the predicted evolution of the five OARs in week 4, 5 and 6 obtained by the six methods are evaluated with the manual contouring of radiation oncologists. The quantitative results of the five OARs in week 4 to 6 are presented in Tables. 4.3 to 4.7 (a)-(c), respectively. Example of OARs regions predicted in week 4, 5 and 6 by the six methods is shown in Figures.4.7. In Figures 4.6 and 4.7, the red regions are the predicted regions by the six methods. On the contrary, the yellow and green regions in Figures.4.6 and 4.7 correspond to the matching and non-matching regions between the predicted result and the ground truth, respectively.

4.3.2 Discussion

From the results in Table.4.2 and Table 4.3 to 4.7, practically, it is observed that the six methods, TEP-Net-FC, TEP-Net-WV, TEP-Net sagittal, TEP-Net axial, TEP-Net coronal and P-net, can predict the future changes of NPC and OARs, week 4 to week

Table 4.5: Average and standard deviation of the accuracy of predicting parotid gland in (a) week 4, (b) week 5 and (c) week 6 by our five methods and P-net.

Method	Precision	Recall	DSC		RMSSD	
			Values	p-value	Values	p-value
(a)Week 4						
TEP-Net-FC	$\mathbf{0.87 \pm 0.23}$	$\mathbf{0.86 \pm 0.07}$	$\mathbf{0.86 \pm 0.14}$	-	$\mathbf{1.4 \pm 0.9}$	-
TEP-Net-WV	0.78 ± 0.28	0.78 ± 0.31	0.78 ± 0.22	N.S.	2.4 ± 0.7	N.S.
TEP-Net sagittal	0.84 ± 0.17	0.83 ± 0.32	0.83 ± 0.07	< 0.05	1.4 ± 0.5	< 0.05
TEP-Net axial	0.84 ± 0.12	0.82 ± 0.08	0.82 ± 0.13	< 0.05	1.5 ± 0.9	< 0.05
TEP-Net coronal	0.80 ± 0.23	0.80 ± 0.16	0.80 ± 0.11	< 0.05	1.8 ± 1.7	< 0.05
P-net	0.79 ± 0.44	0.79 ± 0.16	0.79 ± 0.17	< 0.05	2.2 ± 0.9	< 0.05
(b)Week 5						
TEP-Net-FC	$\mathbf{0.88 \pm 0.21}$	$\mathbf{0.87 \pm 0.31}$	$\mathbf{0.87 \pm 0.12}$	-	$\mathbf{1.3 \pm 0.5}$	-
TEP-Net-WV	0.85 ± 0.34	0.86 ± 0.22	0.85 ± 0.40	N.S.	1.4 ± 0.4	N.S.
TEP-Net sagittal	0.84 ± 0.04	0.83 ± 0.09	0.83 ± 0.19	< 0.05	1.6 ± 1.4	< 0.05
TEP-Net axial	0.83 ± 0.29	0.83 ± 0.45	0.83 ± 0.16	< 0.05	1.6 ± 0.8	< 0.05
TEP-Net coronal	0.83 ± 0.13	0.83 ± 0.51	0.83 ± 0.09	< 0.05	1.6 ± 0.5	< 0.05
P-net	0.81 ± 0.04	0.81 ± 0.08	0.81 ± 0.19	< 0.05	1.8 ± 0.8	< 0.05
(c)Week 6						
TEP-Net-FC	$\mathbf{0.86 \pm 0.21}$	$\mathbf{0.82 \pm 0.47}$	$\mathbf{0.84 \pm 0.24}$	-	$\mathbf{1.5 \pm 1.4}$	-
TEP-Net-WV	0.83 ± 0.18	0.82 ± 0.28	0.82 ± 0.19	N.S.	1.7 ± 0.5	N.S.
TEP-Net sagittal	0.81 ± 0.38	0.80 ± 0.14	0.80 ± 0.23	< 0.05	1.8 ± 1.1	< 0.05
TEP-Net axial	0.78 ± 0.35	0.80 ± 0.29	0.79 ± 0.15	< 0.05	2.1 ± 0.8	< 0.05
TEP-Net coronal	0.78 ± 0.17	0.77 ± 0.21	0.77 ± 0.19	< 0.05	2.6 ± 0.8	< 0.05
P-net	0.76 ± 0.17	0.74 ± 0.23	0.75 ± 0.25	< 0.05	2.8 ± 1.2	< 0.05

6, during RT course. This demonstrates that using GRU in the six methods determines the relevant weights of time series obtained from previous weeks for prediction. In addition, the immobility of NPC tumors and the rigid registration of the weekly CT images during the RT course may achieve and improve the prediction of NPC and OARs, respectively.

The five proposed methods are higher than the comparison method, P-net in Tables.4.2 and 4.3. This is because the proposed CNN in TEP-Net can capture more information that includes local and global information about the NPC and OARs obtained from three different size of patches and the original image size. The additional obtained features from the proposed CNN in week 1 to week n aid GRU to process the trajectories of NPC and OARs evolution in response to RT well. Moreover, P-net outputs NPC and OARs evolution from week 4 to week 6 simultaneously by using only CT images in week 1 to week 3. On the contrary, the proposed five methods based on TEP-Net use all CT images in week 1 to week *n-1* to predict the evolution in week n.

Table 4.6: Average and standard deviation of the accuracy of predicting brainstem in (a) week 4, (b) week 5 and (c) week 6 by our five methods and P-net.

Method	Precision	Recall	DSC		RMSSD	
			Values	p-value	Values	p-value
(a)Week 4						
TEP-Net-FC	$\mathbf{0.86 \pm 0.17}$	$\mathbf{0.83 \pm 0.10}$	$\mathbf{0.84 \pm 0.31}$	-	$\mathbf{1.5 \pm 1.2}$	-
TEP-Net-WV	0.85 ± 0.26	0.82 ± 0.32	0.83 ± 0.44	N.S.	1.5 ± 0.4	N.S.
TEP-Net sagittal	0.83 ± 0.31	0.83 ± 0.11	0.83 ± 0.14	< 0.05	1.4 ± 1.3	< 0.05
TEP-Net axial	0.82 ± 0.36	0.80 ± 0.15	0.81 ± 0.02	< 0.05	1.7 ± 0.3	< 0.05
TEP-Net coronal	0.79 ± 0.17	0.79 ± 0.41	0.79 ± 0.23	< 0.05	2.2 ± 1.7	< 0.05
P-net	0.78 ± 0.35	0.77 ± 0.48	0.77 ± 0.27	< 0.05	2.6 ± 1.8	< 0.05
(b)Week 5						
TEP-Net-FC	$\mathbf{0.87 \pm 0.33}$	$\mathbf{0.84 \pm 0.24}$	$\mathbf{0.85 \pm 0.19}$	-	$\mathbf{1.4 \pm 1.2}$	-
TEP-Net-WV	0.86 ± 0.18	0.82 ± 0.46	0.84 ± 0.11	N.S.	1.4 ± 0.8	N.S.
TEP-Net sagittal	0.85 ± 0.32	0.82 ± 0.07	0.83 ± 0.14	< 0.05	1.7 ± 1.4	< 0.05
TEP-Net axial	0.82 ± 0.51	0.83 ± 0.14	0.82 ± 0.46	< 0.05	1.7 ± 1.3	< 0.05
TEP-Net coronal	0.82 ± 0.17	0.83 ± 0.15	0.82 ± 0.24	< 0.05	1.7 ± 1.2	< 0.05
P-net	0.81 ± 0.32	0.81 ± 0.12	0.81 ± 0.41	< 0.05	1.9 ± 0.8	< 0.05
(c)Week 6						
TEP-Net-FC	$\mathbf{0.84 \pm 0.12}$	$\mathbf{0.81 \pm 0.33}$	$\mathbf{0.82 \pm 0.23}$	-	$\mathbf{1.4 \pm 0.5}$	-
TEP-Net-WV	0.81 ± 0.10	0.81 ± 0.32	0.81 ± 0.41	N.S.	1.8 ± 0.5	N.S.
TEP-Net sagittal	0.80 ± 0.26	0.79 ± 0.31	0.79 ± 0.33	< 0.05	2.2 ± 0.5	< 0.05
TEP-Net axial	0.78 ± 0.41	0.80 ± 0.07	0.79 ± 0.09	< 0.05	2.2 ± 1.4	< 0.05
TEP-Net coronal	0.78 ± 0.05	0.78 ± 0.27	0.78 ± 0.12	< 0.05	2.5 ± 0.1	< 0.05
P-net	0.77 ± 0.24	0.75 ± 0.28	0.76 ± 0.35	< 0.05	2.5 ± 1.4	< 0.05

Therefore, the prediction of NPC and OARs by P-net depends only about the obtained information in week 1 to week 3. With a less information in P-net method, one week prediction of NPC and OARs by using my five methods is a better strategy than three consecutive predictions.

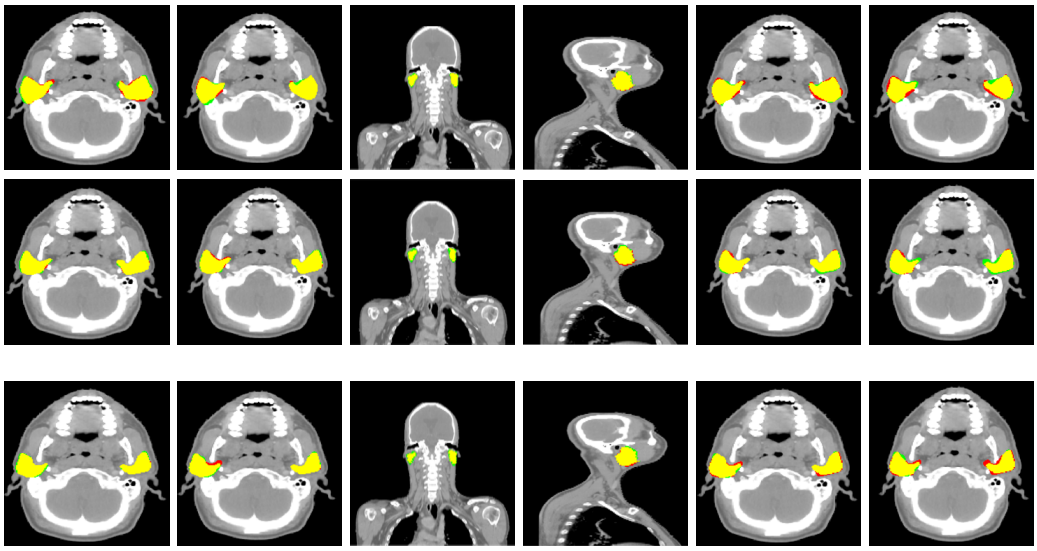
Furthermore, TEP-Net-FC can predict the future evolution of NPC and OARs in response to RT robustly compared to the other five methods. The main reasons why TEP-Net-FC is superior to the other five methods is that TEP-Net-FC determines the final prediction by integrating the three predictions obtained from three directional images. In addition, compared with a simple voting method, FCN used in the integration process attempts to estimate the refined prediction by the three predictions.

To compare TEP-Net-FC with the other methods, the statistical analysis is performed with paired t-tests (p-value < 0.05). The calculated p-values are shown in Table.4.2 and Table 4.3 to 4.7. From the paired t-test, TEP-Net-FC is significantly different from all comparison method for DSC and RMSSD except TEP-Net-WV.

Table 4.7: Average and standard deviation of the accuracy of predicting lens in (a) week 4, (b) week 5 and (c) week 6 by our five methods and P-net.

Method	Precision	Recall	DSC		RMSSD	
			Values	p-value	Values	p-value
(a)Week 4						
TEP-Net-FC	$\mathbf{0.84 \pm 0.22}$	$\mathbf{0.82 \pm 0.17}$	$\mathbf{0.83 \pm 0.05}$	-	$\mathbf{1.6 \pm 0.7}$	-
TEP-Net-WV	0.84 ± 0.30	0.80 ± 0.11	0.82 ± 0.21	N.S.	1.6 ± 1.7	N.S.
TEP-Net sagittal	0.81 ± 0.30	0.80 ± 0.18	0.80 ± 0.05	< 0.05	1.7 ± 0.6	< 0.05
TEP-Net axial	0.79 ± 0.22	0.78 ± 0.11	0.78 ± 0.09	< 0.05	2.0 ± 1.7	< 0.05
TEP-Net coronal	0.78 ± 0.14	0.79 ± 0.21	0.78 ± 0.05	< 0.05	2.3 ± 1.4	< 0.05
P-net	0.76 ± 0.15	0.74 ± 0.09	0.75 ± 0.27	< 0.05	2.5 ± 1.3	< 0.05
(b)Week 5						
TEP-Net-FC	$\mathbf{0.85 \pm 0.30}$	$\mathbf{0.84 \pm 0.21}$	$\mathbf{0.84 \pm 0.19}$	-	$\mathbf{1.6 \pm 0.6}$	-
TEP-Net-WV	0.85 ± 0.18	0.82 ± 0.33	0.83 ± 0.21	N.S.	1.5 ± 0.9	N.S.
TEP-Net sagittal	0.80 ± 0.14	0.80 ± 0.67	0.80 ± 0.14	< 0.05	1.8 ± 0.8	< 0.05
TEP-Net axial	0.80 ± 0.12	0.81 ± 0.05	0.80 ± 0.31	< 0.05	1.8 ± 1.4	< 0.05
TEP-Net coronal	0.81 ± 0.17	0.80 ± 0.21	0.80 ± 0.16	< 0.05	1.8 ± 1.4	< 0.05
P-net	0.79 ± 0.33	0.79 ± 0.14	0.79 ± 0.16	< 0.05	2.4 ± 0.6	< 0.05
(c)Week 6						
TEP-Net-FC	$\mathbf{0.83 \pm 0.22}$	$\mathbf{0.80 \pm 0.14}$	$\mathbf{0.81 \pm 0.33}$	-	$\mathbf{1.7 \pm 1.2}$	-
TEP-Net-WV	0.80 ± 0.28	0.79 ± 0.14	0.79 ± 0.13	N.S.	2.2 ± 0.6	N.S.
TEP-Net sagittal	0.80 ± 0.09	0.79 ± 0.17	0.79 ± 0.03	< 0.05	2.3 ± 0.7	< 0.05
TEP-Net axial	0.77 ± 0.14	0.82 ± 0.37	0.78 ± 0.29	< 0.05	2.5 ± 2.9	< 0.05
TEP-Net coronal	0.77 ± 0.14	0.75 ± 0.47	0.76 ± 0.19	< 0.05	2.9 ± 0.2	< 0.05
P-net	0.74 ± 0.21	0.70 ± 0.27	0.72 ± 0.14	< 0.05	2.5 ± 0.8	< 0.05

Although the success of TEP-Net-FC, the predicted results obtained on week 6 is lower than those obtained on week 4 and week 5. Here, TEP-Net-FC predicts the evolution of NPC and OARs in week 6 from weekly CT images obtained in week 1 to 5. Theoretically, by training 5 weekly CT images, TEP-Net-FC aims to increase the performance of the predicted results in week 6 as observed in week 5. On the contrary, TEP-Net-FC fails to predict the evolution of NPC and OARs in week 6 and ouputs a lower performance compared with those obtained in week 4 and 5. As mentioned before, there is a high variation of NPC changes in the used dataset occurred between week 5 and week 6 which influences on the performance of its prediction in the end of week 6. To solve this problem, collecting more training data can improve the accuracy of the prediction of my method.



(a) (b) (c) (d) (e) (f)

Figure 4.7: OARs predicted results in week 4 (Top row), week 5 (middle row) and week 6 (button row) obtained from (a)TEP-Net-FC, (b)TEP-Net-WV, (c)TEP-Net sagittal, (d)TEP-Net coronal, (e)TEP-Net axial and (f)P-net. (Red) the segmented non-target organ; (Green and Yellow) the matching (green) and non-matching (yellow) regions between the segmented non-target organ and the ground-truth.

4.3.3 Conclusion

In this chapter, I proposed a new system for predicting the evolution of NPC and OARs in week *n* during RT course from CT images obtained in previous weeks, week 1 to week *n-1*. The proposed method, TEP-Net, is the combination of three components, CNN, GRU and global attention model, to achieve the prediction of the target region.

In the experiments, TEP-Net trains the three orthogonal CT images separately as input data to predict the evolution of NPC and OARs in week 4 to week 6 from week 1 to 3, week 1 to 4 and week 1 to 5, respectively. The final NPC and OARs prediction among the three predicted results are decided by using an integration process. Here, two integration methods are applied. From the predicted results, TEP-Net predicts accurately the evolution of NPC and OARs regions in response to RT compared with conventional method, especially those obtained by using weighted fully connected layers, TEP-Net-FC, as integration process.

5

Conclusion

The automatic support system for IMRT is useful for not only radiation oncologists but also patients. The radiation oncologists provides the suitable therapy to a patient according to the patient conditions. Such therapy can improve the QOL of the patient.

In this thesis, to construct an automatic system for IMRT, I propose three fundamental techniques, for segmenting NPC from CT images, predicting the dose distribution and predicting the future evolution of NPC and OARs in response to RT, respectively.

Firstly, I proposed the CNN-based method for segmenting NPC and OARs from three orthogonal, axial, coronal and sagittal, CT images. The proposed method, CND, is a cascade system composed of two phases. In the first phase, non-target organ regions are detected and eliminated from a given CT image. The second phase segments NPC regions from the remained region in the CT image after the first phase. The proposed system outputs the final NPCs by integrating the segmentation results obtained from the three orthogonal images. From the experimental results using 11900 NPC pa-

tient images, the proposed CND using various overlapping patches with different sizes detects NPC with high performance compared with other conventional NPC detection methods.

Secondly, I presented a new CNN which predicts the dose distribution for a patient from contoured CT images. In the proposed method, features maps of dose images of each beam are estimated from contoured CT images. The dose distribution for a patient is predicted by integrating the obtained features maps from all the beams. The experimental results using 13600 NPC patient images prove that the proposed method can predict the dose distribution images reliably and accurately compared with the conventional methods.

Thirdly, I proposed a new system for predicting the evolution of the NPC and OARs in response to RT treatment by using previous weekly changes of target region, week 1 to week *n-1*. The proposed method, TEP-Net-FC, consists of CNN and two GRUs. Using weekly CT axial, coronal and sagittal images, the CNN is used to extract relevant features from CT images. The combination of the two GRUs remove the irrelevant features and outputs useful features for predicting the evolution of NPC and OARs. Using fully connected layers, the final NPC and OARs prediction is obtained by integrating the obtained weights from the three sections. From the experimental results using 71400 NPC patient images, the proposed method using weighted fully connected layers as integration process, TEP-Net-FC, achieves a reliable prediction of NPC and OARs compared with other conventional prediction methods.

5.1 Future works

The construction of the three methods use the CT images of 70, 80 and 140 patients, respectively. Practically, from the experimental results, the number and quality of CT images used as training dataset influence on the accuracy of the proposed method. In this thesis, the regions of NPC and OARs are determined manually by two radiation oncologists. In order to guarantee that the manual contouring of NPC and OARs are used as dataset by the proposed methods, it is necessary to obtain the contours of NPC and OARs of the same patient regions by more radiation oncologists.

The current CND system for segmenting NPC and OARs regions focuses on stage T1 and stage T2 that non-target regions have no NPC. However, in stage T3 and stage T4, NPC spreads to non-target organs that influence on the performance of the CND.

Therefore, one of my future works is to extend the CND system to detect NPC in advanced stage. In addition, the CT image resolution of a patient is related to the spatial resolution of scanning the patient. Therefore, if the resolution of a given CT image is lower than those of the CT images in my database, the accuracy of the proposed method may become worse because the NPC contours in the low resolution images become blurred and unclear. One solution for this is to introduce image super resolution techniques [51] to obtain its high resolution. This is one of my future works.

When predicting the 2D dose distribution images from contoured 2D CT images, the proposed method dose not consider dose constraints of OARs and tumor regions. As mentioned in Chapter 3, the dose constraints means the range of the delivered dose to OARs and tumor regions without any damages. If they receive a higher dose than their dose constraints, the OARs regions may get damaged. Therefore, by introducing the dose constraints in the proposed system, the predicted dose distribution can further minimize the damage of OARs. Furthermore, generally, some OARs such as spinal cord may move due to the patient's breathing during the RT course. In this case, a more focused radiation may deliver to some OARs close to the NPC during their movement. Therefore, the therapy may affects the QOL of the patient. One solution for this is to track the movement of OARs regions by adjusting the predicted dose distribution in real-time to improve the predicted dose distribution.

Bibliography

[1] Ahmedin Jemal, Freddie Bray, Melissa M Center, Jacques Ferlay, Elizabeth Ward, and David Forman. Global cancer statistics. *CA: a cancer journal for clinicians*, 61(2):69–90, 2011.

[2] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2015. *CA: a cancer journal for clinicians*, 65(1):5, 2015.

[3] Terese Winslow. `https://www.teresewinslow.com/head/ny7jwz4w6ox02f2o75h5shuk586z33`, Retrieved 20 September 2020.

[4] William I Wei and Jonathan ST Sham. Nasopharyngeal carcinoma. *The Lancet*, 365(9476):2041–2054, 2005.

[5] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L Siegel, Lindsey A Torre, and Ahmedin Jemal. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6):394–424, 2018.

[6] F De-Vathaire, H Sancho-Garner, H De-The, C Pieddeloup, G Schwaab, JHC Ho, R Ellouz, C Micheau, M Cammoun, Y Cachin, et al. Prognostic value of ebv markers in the clinical management of nasopharyngeal carcinoma (npc): a multicenter follow-up study. *International journal of cancer*, 42(2):176–181, 1988.

[7] C Yu Mimi, John HC Ho, Shiu-Hung Lai, and Brian E Henderson. Cantonese-style salted fish as a cause of nasopharyngeal carcinoma: report of a case-control study in hong kong. *Cancer research*, 46(2):956–961, 1986.

[8] Dolly P Huang, Kwok-Wai Lo, C Andrew Van Hasselt, John KS Woo, Peter HK Choi, Sing-Fai Leung, Siu-Tim Cheung, Paul Cairns, David Sidransky, and Joseph CK Lee. A region of homozygous deletion on chromosome 9p21–22 in primary nasopharyngeal carcinoma. *Cancer research*, 54(15):4003–4006, 1994.

[9] James D Brierley, Mary K Gospodarowicz, and Christian Wittekind. *TNM classification of malignant tumours*. John Wiley & Sons, 2017.

[10] Tingting Xu, Chunying Shen, Guopei Zhu, and Chaosu Hu. Omission of chemotherapy in early stage nasopharyngeal carcinoma treated with imrt: a paired cohort study. *Medicine*, 94(39), 2015.

[11] Melvin LK Chua, Joseph TS Wee, Edwin P Hui, and Anthony TC Chan. Nasopharyngeal carcinoma. *The Lancet*, 387(10022):1012–1024, 2016.

[12] Catherine Kubrak, Kärin Olson, Naresh Jha, Louise Jensen, Linda McCargar, Hadi Seikaly, Jeffery Harris, Rufus Scrimger, Matthew Parliament, and Vickie E Baracos. Nutrition impact symptoms: key determinants of reduced dietary intake, weight loss, and reduced functional capacity of patients with head and neck cancer before treatment. *Head & Neck: Journal for the Sciences and Specialties of the Head and Neck*, 32(3):290–300, 2010.

[13] Xiao-Kang Zheng, Long-Hua Chen, Quan-Shi Wang, and Fu-Bing Wu. Influence of [18f] fluorodeoxyglucose positron emission tomography on salvage treatment decision making for locally persistent nasopharyngeal carcinoma. *International Journal of Radiation Oncology* Biology* Physics*, 65(4):1020–1025, 2006.

[14] Dan Nguyen. `https://danwin.com/tag/imrt`, Retrieved 20 September 2020.

[15] William M Mendenhall, Robert J Amdur, and Jatinder R Palta. Intensity-modulated radiotherapy in the standard management of head and neck cancer: promises and pitfalls. *Journal of clinical oncology*, 24(17):2618–2623, 2006.

[16] George Fountzilas, Amanda Psyrri, Eleni Giannoulatou, Ioannis Tikas, Kyriaki Manousou, Dimitra Rontogianni, Elisabeta Ciuleanu, Tudor Ciuleanu, Liliana Resiga, Thomas Zaramboukas, et al. Prevalent somatic brca1 mutations shape

clinically relevant genomic patterns of nasopharyngeal carcinoma in southeast europe. *International journal of cancer*, 142(1):66–80, 2018.

[17] Paul M Harari, Shiyu Song, and Wolfgang A Tomé. Emphasizing conformal avoidance versus target definition for imrt planning in head-and-neck cancer. *International Journal of Radiation Oncology* Biology* Physics*, 77(3):950–958, 2010.

[18] U Oelfke and T Bortfeld. Inverse planning for photon and proton beams. *Medical Dosimetry*, 26(2):113–124, 2001.

[19] Benjamin E Nelms, Greg Robinson, Jay Markham, Kyle Velasco, Steve Boyd, Sharath Narayan, James Wheeler, and Mark L Sobczak. Variation in external beam treatment plan quality: an inter-institutional study of planners and planning systems. *Practical radiation oncology*, 2(4):296–305, 2012.

[20] Rui-hao Wang, Shu-xu Zhang, Ling-hong Zhou, Guo-qian Zhang, Hui Yu, Xiao-dan Lin, and Shengqu Lin. Volume and dosimetric variations during two-phase adaptive intensity-modulated radiotherapy for locally advanced nasopharyngeal carcinoma. *Bio-medical materials and engineering*, 24(1):1217–1225, 2014.

[21] S Marzi, P Pinnarò, D D'Alessio, L Strigari, V Bruzzaniti, C Giordano, G Giovinazzo, and L Marucci. Anatomical and dose changes of gross tumour volume and parotid glands for head and neck cancer patients during intensity-modulated radiotherapy: effect on the probability of xerostomia incidence. *Clinical Oncology*, 24(3):e54–e62, 2012.

[22] Jolien Heukelom and Clifton David Fuller. Head and neck cancer adaptive radiation therapy (art): conceptual considerations for the informed clinician. In *Seminars in radiation oncology*, volume 29, pages 258–273. Elsevier, 2019.

[23] Xue-Song Sun, Xiao-Yun Li, Qiu-Yan Chen, Lin-Quan Tang, and Hai-Qiang Mai. Future of radiotherapy in nasopharyngeal carcinoma. *The British journal of radiology*, 92(1102):20190209, 2019.

[24] Jie Lu, Yidong Ma, Jinhu Chen, Liming Wang, Guifang Zhang, Mukun Zhao, and Yong Yin. Assessment of anatomical and dosimetric changes by a deformable registration method during the course of intensity-modulated radiotherapy for nasopharyngeal carcinoma. *Journal of radiation research*, 55(1):97–104, 2014.

[25] Chanon Tatanun, Panrasee Ritthipravat, Thongchai Bhongmakapat, and Lojana Tuntiyatorn. Automatic segmentation of nasopharyngeal carcinoma from ct images: Region growing based technique. In *2010 2nd International Conference on Signal Processing Systems*, volume 2, pages V2–537. IEEE, 2010.

[26] Weerayuth Chanapai, Thongchai Bhongmakapat, Lojana Tuntiyatorn, and Panrasee Ritthipravat. Nasopharyngeal carcinoma segmentation using a region growing technique. *International journal of computer assisted radiology and surgery*, 7(3):413–422, 2012.

[27] Isabelle Fitton, SAP Cornelissen, Joop C Duppen, RJHM Steenbakkers, STH Peeters, FJP Hoebers, Johannes HAM Kaanders, PJCM Nowak, CRN Rasch, and Marcel van Herk. Semi-automatic delineation using weighted ct-mri registered images for radiotherapy of nasopharyngeal cancer. *Medical physics*, 38(8):4662–4666, 2011.

[28] Kai-Wei Huang, Zhe-Yi Zhao, Qian Gong, Juan Zha, Liu Chen, and Ran Yang. Nasopharyngeal carcinoma segmentation via hmrf-em with maximum entropy. In *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 2968–2972. IEEE, 2015.

[29] Liang Chen, Leitao Cui, Rong Huang, and Zhengyun Ren. Bio-inspired neural network with application to license plate recognition: hysteretic elm approach. *Assembly Automation*, 2016.

[30] Jun Shi, Xiao Zheng, Yan Li, Qi Zhang, and Shihui Ying. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of alzheimer's disease. *IEEE journal of biomedical and health informatics*, 22(1):173–183, 2017.

[31] Jin Qi, Miao Le, Chunming Li, and Ping Zhou. Global and local information based deep network for skin lesion segmentation. *arXiv preprint arXiv:1703.05467*, 2017.

[32] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Modeling the intra-class variability for liver lesion detection using a multi-class patch-based cnn. In *International Workshop on Patch-based Techniques in Medical Imaging*, pages 129–137. Springer, 2017.

[33] Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter. Nuclei segmentation in histopathology images using deep neural networks. In *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*, pages 933–936. IEEE, 2017.

[34] Sérgio Pereira, Adriano Pinto, Victor Alves, and Carlos A Silva. Brain tumor segmentation using convolutional neural networks in mri images. *IEEE transactions on medical imaging*, 35(5):1240–1251, 2016.

[35] Tyler Clark, Alexander Wong, Masoom A Haider, and Farzad Khalvati. Fully deep convolutional neural networks for segmentation of the prostate gland in diffusion-weighted mr images. In *International Conference Image Analysis and Recognition*, pages 97–104. Springer, 2017.

[36] Kuo Men, Xinyuan Chen, Ye Zhang, Tao Zhang, Jianrong Dai, Junlin Yi, and Yexiong Li. Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. *Frontiers in oncology*, 7:315, 2017.

[37] Yan Wang, Chen Zu, Guangliang Hu, Yong Luo, Zongqing Ma, Kun He, Xi Wu, and Jiliu Zhou. Automatic tumor segmentation with deep convolutional neural networks for radiotherapy applications. *Neural Processing Letters*, 48(3):1323–1334, 2018.

[38] Zongqing Ma, Xi Wu, Qi Song, Yong Luo, Yan Wang, and Jiliu Zhou. Automated nasopharyngeal carcinoma segmentation in magnetic resonance images by combination of convolutional neural networks and graph cut. *Experimental and therapeutic medicine*, 16(3):2511–2521, 2018.

[39] Bulat Ibragimov and Lei Xing. Segmentation of organs-at-risks in head and neck ct images using convolutional neural networks. *Medical physics*, 44(2):547–557, 2017.

[40] Ilma Xhaferllari, Eugene Wong, Karl Bzdusek, Michael Lock, and Jeff Z Chen. Automated imrt planning with regional optimization using planning scripts. *Journal of applied clinical medical physics*, 14(1):176–191, 2013.

[41] M Folkerts, T Long, R Radke, W Lu, X Jia, X Gu, M Chen, and S Jiang. Knowledge-based automatic treatment planning for prostate imrt using 3-dimensional dose prediction and threshold-based optimization: Su-e-fs2-06. *Medical Physics*, 44(6), 2017.

[42] Rafid Mahmood, Aaron Babier, Andrea McNiven, Adam Diamant, and Timothy CY Chan. Automated treatment planning in radiation therapy using generative adversarial networks. *arXiv preprint arXiv:1807.06489*, 2018.

[43] Xinyuan Chen, Kuo Men, Yexiong Li, Junlin Yi, and Jianrong Dai. A feasibility study on an automated method to generate patient-specific dose distributions for radiotherapy using deep learning. *Medical physics*, 46(1):56–64, 2019.

[44] Jiawei Fan, Jiazhou Wang, Zhi Chen, Chaosu Hu, Zhen Zhang, and Weigang Hu. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. *Medical physics*, 46(1):370–381, 2019.

[45] Ling Zhang, Le Lu, Ronald M Summers, Electron Kebebew, and Jianhua Yao. Convolutional invasion and expansion networks for tumor growth prediction. *IEEE transactions on medical imaging*, 37(2):638–648, 2017.

[46] Chuang Wang, Andreas Rimner, Yu-Chi Hu, Neelam Tyagi, Jue Jiang, Ellen Yorke, Sadegh Riyahi, Gig Mageras, Joseph O Deasy, and Pengpeng Zhang. Toward predicting the evolution of lung tumors during radiotherapy observed on a longitudinal mr imaging study via a deep learning algorithm. *Medical physics*, 46(10):4699–4707, 2019.

[47] Rahul Dey and Fathi M Salemt. Gate-variants of gated recurrent unit (gru) neural networks. In *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, pages 1597–1600. IEEE, 2017.

[48] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[49] Luciana Caravatta, Gabriella Macchia, Gian Carlo Mattiucci, Aldo Sainato, Nunzia LV Cernusco, Giovanna Mantello, Monica Di Tommaso, Marianna Trignani,

Antonino De Paoli, Gianni Boz, et al. Inter-observer variability of clinical target volume delineation in radiotherapy treatment of pancreatic cancer: a multi-institutional contouring experience. *Radiation oncology*, 9(1):198, 2014.

[50] L Lu, L Cuttino, I Barani, S Song, M Fatyga, M Murphy, P Keall, J Siebers, and J Williamson. Su-ff-j-85: Inter-observer variation in the planning of head/neck radiotherapy. *Medical Physics*, 33(6Part6):2040–2040, 2006.

[51] Yukai Wang, Qizhi Teng, Xiaohai He, Junxi Feng, and Tingrong Zhang. Ct-image super resolution using 3d convolutional neural network. *arXiv preprint arXiv:1806.09074*, 2018.