

國立中正大電機工程研究所

碩士論文

基於先驗知識引導之多器官切割

Prior guiding based multiple organ segmentation

研究生：林敬翔

指導教授：黃敬群 博士

賴文能 博士

中華民國一零九年十月

致謝

在碩士生涯裡跌跌撞撞，邊學邊跌倒邊成長。雖然兩年間獲得了許多新的體悟以及成長，仍有許多不足之處，需要持續地精進自己。一篇論文的完成需要許多人的幫助，絕非一人的努力所能完成。在此感謝求學路上給过我幫助的人們。感謝黃敬群老師兩年來的帶領以及指導，老師不只在學術上給予我指導以及幫助，我也從老師身上學到了許多為人處世的道理，老師的包容以及鼓勵也是我前進的動力之一。

感謝國立中正大學資訊工程學系的林維暘老師擔任學生的口試委員召集人，感謝國立雲林科技大學資訊工程系的林建州老師以及國立中正大學的電機工程學系的賴文能老師、江瑞秋老師擔任學生的口試委員，並給予許多寶貴的建議，得學生的論文研究更加完善而嚴謹，在此致上最深的謝意與敬意。

我也非常幸運能夠身處於 ACMLAB 的環境中，感謝吳侑峰同學、劉育麟同學、Su Huynh Van 同學，他們是非常優秀的同儕，我從他們身上學到了許多，他們的砥礪以及幫助也激勵我不斷往前。也感謝趙梓吟同學、賴震東同學、鄭惟中同學以及廖志翔同學。他們是優秀、上進的學弟，感謝他們在許多事情上可靠的幫助。最後感謝我親愛的家人以及女朋友家珍，他們不要求回報的支持是我的最大後盾。希望，我能夠持續前進不辜負他人、自己所望，也希望所有這些支持、幫助過我的人都能夠得以完成自身的理想、充實自身的人生。

摘要

電腦斷層掃描 (Computed Tomography) 中正確且穩健的腹部器官切割在許多臨床應用都是重要且必須的工作，例如：電腦輔助診斷、電腦輔助手術等等。然而這項任務也往往面臨了許多挑戰，包括了器官間的相似性、斷層掃描影像缺乏紋理變化等等。本研究針對這些挑戰設計了一種基於多任務學習、注意力機制以及時序資訊的模型。我們注意到雖然個體間存有差異，但是器官的形狀、位置變化仍有一模式可遵循，包括了器官總是位於固定的區域以及個體間有著相似的器官相對位置等等，藉由這些先驗知識，我們的模型在執行器官切割任務的同時也執行了人體縱向分類任務(longitudinal classification)，模型將同時學習如何進行多器官的分割以及每張影像與人體縱向的對應關係。另外我們也透過引用注意力機制於我們的演算法當中，相較於過往的注意力機制讓模型自動學習特徵需要特別關注的區域，我們的方法提供了由縱向的分類結果與先驗知識(prior)共同結合所產生的導引圖(guidance)。最後，電腦斷層掃描所獲得的三維資訊可以轉換為連續序列，這些序列影像彼此存在關聯性，我們透過使用雙向門控循環單元(Bidirectional Gated Recurrent Unit) 學習這樣的關聯性，期望演算法透過學習影像間的序列關係獲得性能的增強。本研究所提出之方法在 2015 MICCAI Multi-Atlas Abdomen Labeling Challenge 上驗證集取得了 86.2 的 Average Dice Score。

關鍵字：多重器官切割、注意力模型、循環神經網路、導引圖

Abstract

For many clinical applications such as computer-aided diagnosis and computer-aided surgery, robust and accurate segmentation of abdominal organs on CT is essential. However, a lot of different challenges will be faced, including high similarity between organs, lack of texture in CT scan images. To solve these challenges, this research designs a deep learning model based on multi-task learning, attention mechanism, and temporal information. We noticed that although there are differences among individuals, there is still a pattern that can be followed in the shape and positional changes of organs, including the fact that organs are always located in some specific regions and that there are similar relative positions of organs among individuals, etc. Through this prior knowledge, our model performs the organ segmentation task while performing the longitudinal classification of the human body. The model will simultaneously learn how to segment multiple organs and the correspondence between each image and the human longitudinal axis. Besides, we also apply the attention mechanism in our algorithm. Our method provides guidance that is produced by combining longitudinal classification results with prior knowledge. Finally, the three-dimensional information obtained by computer tomography can be converted into a continuous sequence. These sequential images are related to each other. We apply the Bidirectional Gated Recurrent Unit (Bi-GRU) to learn such correlation and expect the algorithm improved through additional temporal information. Our method achieved the Average Dice Score of 86.2 on the validation set of the 2015 MICCAI Multi-Atlas Abdomen Labeling Challenge.

Keywords: multi-organ segmentation, attention model, recurrent neural network, guidance

目錄

致謝.....	i
摘要.....	ii
Abstract.....	iii
圖目錄.....	vi
表目錄.....	vii
第一章、緒論.....	1
1.1 研究背景與動機.....	1
1.2 研究目標與困難.....	1
1.3 論文架構.....	4
第二章、文獻回顧.....	5
2.1 多重器官切割.....	5
2.1.1 基於二維全卷積網路之切割演算法.....	5
2.1.2 基於三維全卷積網路之切割演算法.....	7
2.2 注意力模型應用於切割演算法.....	9
2.2.1 注意力機制.....	9
2.2.2 基於注意力模型之切割演算法.....	10
第三章、基於先驗知識引導式之多器官切割方法.....	13
3.1 系統架構.....	13
3.2 基於先驗知識之導引圖生成.....	15
3.3 基於導引圖之全卷積網路.....	17
3.4 雙向卷積門控循環單元.....	19
3.5 目標函數.....	20
第四章、實驗結果與討論.....	22
4.1 實驗配置.....	22

4.1.1 實作細節及相關配置.....	22
4.1.2 資料集.....	23
4.1.3 評估指標.....	25
4.2 實驗結果.....	27
4.2.1 消融研究.....	27
4.2.2 與其他方法之比較.....	28
4.3 可視化結果.....	30
4.3.1 二維器官切割可視化結果.....	30
4.3.2 注意力可視化結果.....	31
第五章、結論.....	34
5.1 結果討論.....	34
5.2 未來工作.....	34
參考文獻.....	35

圖目錄

圖 1 、腹部器官電腦斷層掃描影像.....	2
圖 2 、進行腹部器官醫學影像處理時所面臨的挑戰.....	3
圖 3 、同個實驗對象中不同人體縱向位置的脾臟標註結果.....	4
圖 4 、U-net 之模型架構	6
圖 5 、MultiRes U-net 中提出了 MultiRes block.....	6
圖 6 、DeepLab v3 架構以及 ASPP 模塊	7
圖 7 、nnUnet 提出之訓練架構.....	9
圖 9 、殘差注意力網路之模型架構。.....	11
圖 10 、串級精化網路之模型架構.....	12
圖 11 、本研究模型架構圖。.....	14
圖 12 、透過訓練標籤的分群及正則化所預先產生之先驗知識.....	16
圖 13 、縱向位置分類器之詳細架構.....	17
圖 14 、本研究工作編碼器及解碼器架構之一，PGN-v1	18
圖 15 、本研究工作編碼器及解碼器架構之二，PGN-v2	19
圖 16 、腹部多器官切割結果.....	31
圖 17 、注意力可視化結果.....	32
圖 18 、反向注意力之可視化結果.....	33

表目錄

表 1、實驗環境配置.....	22
表 2、主要使用函式庫。.....	23
表 3、2015 MICCAI Multi-Atlas Abdomen Labeling Challenge 資料集	24
表 4、2019 Combined (CT-MR) Healthy Abdominal Organ Segmentation 資料集..	25
表 5、各模組使用對於 PGN-v1 整體性能影響的比較	27
表 6、各模組對於 PGN-v2 整體性能影響的比較	28
表 7、本研究架構於 2015 MICCAI Multi-Atlas Abdomen Labeling Challenge 之評估結果以及其他方法的比較.....	29
表 8、本研究架構於 CHAOS dataset Task5: MRI 腹部器官切割之評估結果以及其他方法之比較.....	30

第一章、緒論

1.1 研究背景與動機

影像切割一直都是從影像中萃取語意資訊的一種重要方式，這項技術也被廣泛地使用於各個領域，包括了自動駕駛車、機器人領域以及醫學影像等等。其中，在醫學影像領域，影像切割往往用於取得人類、動物體內的相關資訊，這些對象包括了細胞、腫瘤、器官的位置、形狀、大小等等。其中，腹部器官切割尤其地重要。正確且穩健的腹部器官切割在許多臨床應用都是重要且必須的工作，例如：電腦輔助診斷、電腦輔助手術等等。

近年來，由於硬體的資源的進步帶動了深度學習的快速發展，深度學習也在許多視覺相關任務取得了許多重大的突破，包括了影像分類、目標偵測、影像生成等等。在影像切割任務中，深度學習也是現今使用頻率最高以及表現最佳的方法。本研究工作希望透過引進自然影像切割中常用之注意力模型、循環神經網路，能夠獲得穩健的切割結果。

1.2 研究目標與困難

相較於主流電腦視覺領域探討的自然影像相關的切割工作，多器官切割面臨了不同的挑戰。在醫學領域中多使用了電腦斷層掃描(Computed Tomography Scan, CT)、磁共振成像(Magnetic Resonance Imaging, MRI)等成像方式生成三維資料再將其轉換為連續序列影像。這樣的影像大多為灰階影像，相比於一般語意切割方法所使用的自然影像較缺乏紋理的多樣性，如圖 1。

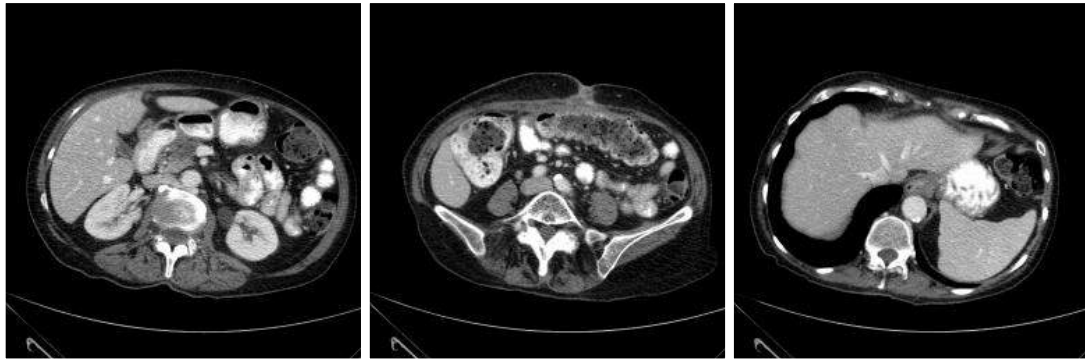


圖 1、腹部器官電腦斷層掃描影像，醫學影像相比於自然影像往往缺乏紋理、特徵的變化。

此外，模糊的器官邊界、雜亂背景等問題都造成了多器官切割的困難，我們將這些處理腹部器官影像上的所碰到的挑戰細分為以下四個問題，並整理如圖 2 所示。

1. 模糊的器官邊界 (Weak boundaries of organs)
2. 雜亂背景 (Clustering background)
3. 器官與軟組織間的高度相似性 (High appearance similarity between organ and tissue)
4. 外部因素所造成的外觀變化 (Appearance variation caused by external factors)

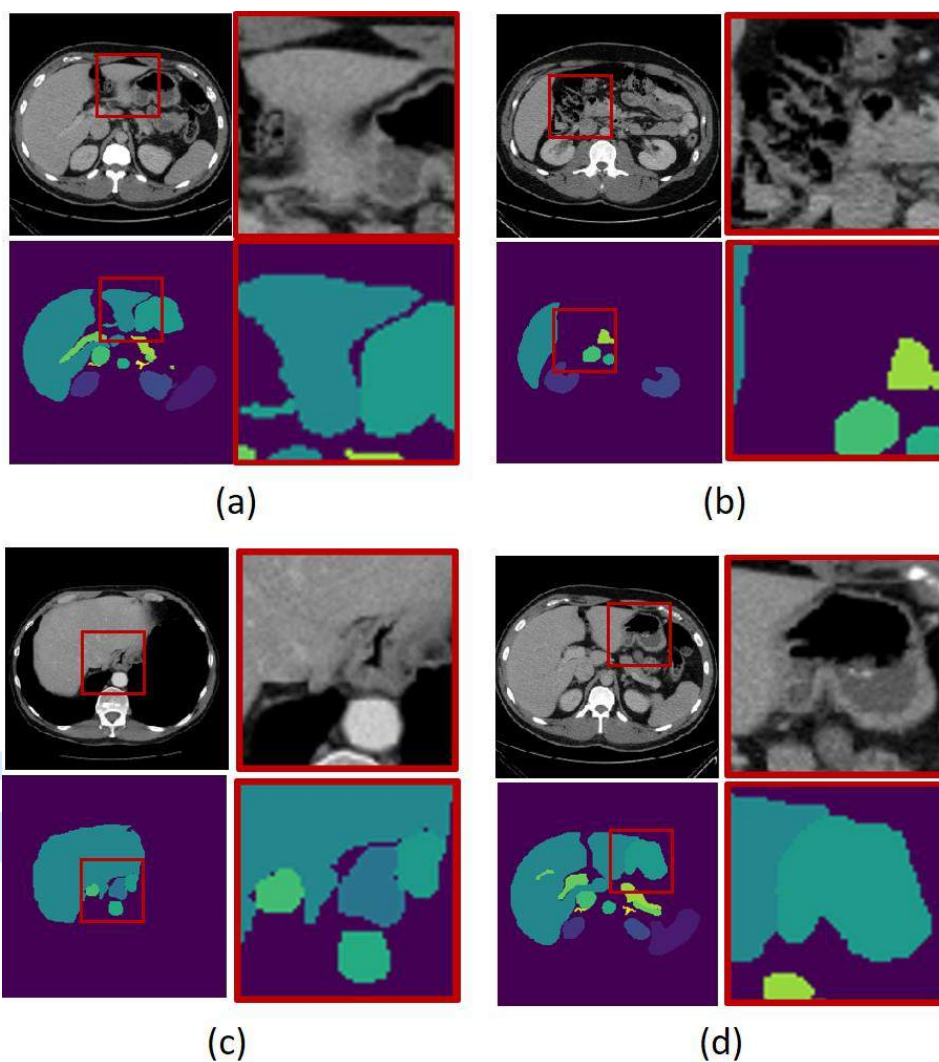


圖 2、進行腹部器官醫學影像處理時所面臨的挑戰。包含但不限於 (a) 模糊的器官邊界、(b) 雜亂背景、(c) 器官與軟組織間的高度相似性、(d) 外部因素所造成的外觀變化

多器官切割還面臨了另一項挑戰，便是相同器官間的變化以及不同器官間的高度相似。在二維的序列影像中，相同的器官隨著在人體縱向(longitudinal) 位置的改變在尺寸及形狀可能會有相當大的變化，如圖 3 為實驗者在不同人體縱向位置所拍攝的脾臟，可以發現器官的尺寸呈現極端地變化。相反地，不同器官間也有可能在外觀上十分相近，許多器官由於尺寸較小且沒有明顯的紋理變化，容易與其他器官、一般的軟組織混淆造成切割上的困難。

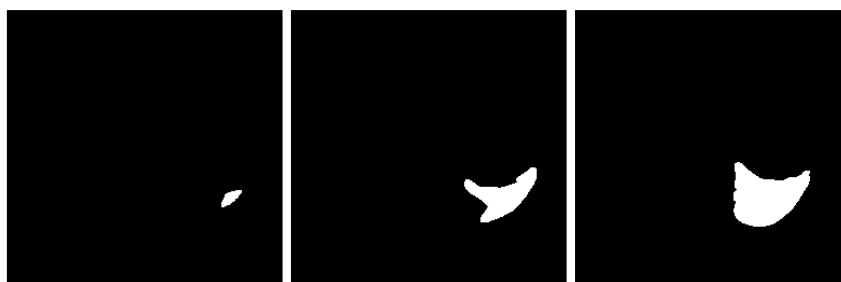


圖 3、同個實驗對象中不同人體縱向位置的脾臟標註結果

1.3 論文架構

本論文共分為五個章節，各章節內容概述如下：

第一章、緒論

此章節概略地介紹了何謂多器官切割，以及多器官切割在醫學領域之重要性。

此外也介紹了研究現況、研究所期望達成的目標以及遭遇的挑戰。

第二章、文獻回顧

此章節基於語意切割、多器官切割對於文獻做了完整地回顧，介紹了包含基於全卷積網路的方法、基於注意力機制的方法等等。

第三章、基於先驗知識引導式之多器官切割方法

此章節詳細地介紹了本研究工作主要使用的演算法，描述了本研究工作如何針對多器官切割所面臨的挑戰設計相關算法及模型，以及這些模型背後所代表之物理意義及設計理念。

第四章、實驗結果與討論

此章節介紹了本研究工作所使用的資料集、評估方法、評估結果等等，也透過了不同實驗參數以及不同方法之間的比較對演算法做出更詳細、深度的討論。

第五章、結論

此章節針對本研究工作進行最後的總結，描述了本研究工作以及所帶來之貢獻。此外也針對未來可能持續進行的方向以及相關工作做出討論。

第二章、文獻回顧

2.1 多重器官切割

2.1.1 基於二維全卷積網路之切割演算法

基於全卷積網路 (Fully Convolutional Network) 之方法為現今影像切割以及許多其他影像相關任務的主流方法。全卷積網路得名於在網路中全面地使用了卷積層，因此網路可以接受任何大小的輸入並且可以保有影像的空間特徵。在醫學影像切割中，全卷積網路又可概分為基於二維及三維之方法，因為醫學領域中使用的電腦斷層掃描以及核磁共振成像所取得之資料多為三維體積 (3D Volume)，因此透過三維卷積層直接對於三維體積進行運算的網路便稱為三維全卷積網路，如[1, 4, 14, 24, 25, 32, 36]。相反地，若先將三維體積轉為序列影像，再個別對於影像進行二維卷積操作則稱為二維全卷積網路，如[5, 17, 27, 28]。

U-net[3]由於其精簡的網路設計以及優異的表現，其架構後來被大量地使用於醫學切割方法中，模型如圖 4 所示。U-net 的取名原因來自於成對的編解碼架構，使得網路呈現一”U”形狀。編碼器由一系列的卷積層以及池化層(Max Pooling Layer)所構成，卷積層用於從影像中擷取特徵，池化層則同時達到減少資料量以及擴大接受野(Receptive Field)的目的。經過編碼器獲得的特徵再經由一系列的卷積層以及上採樣層(Upsampling Layer)或是反卷積層(Deconvolution Layer)恢復成原始影像大小。雖然池化層透過擴大網路的接受野可以有效提升網路的辨識效能，但是由於空間資訊的丟失，將使得最後的切割結果無法很好對物體作定位(Localization)。因此 U-net 提出了跳級連結(Skip-connection)的架構，將編碼器中的淺層特徵與解碼器中的深層特徵串聯(concatenate)後再共同進行卷積，這些淺層特徵更為接近原始影像的特徵以及空間資訊，解碼器透過同時處理兩種特徵，

能夠有效地提升定位的效能，進而最終得到較佳的切割結果。

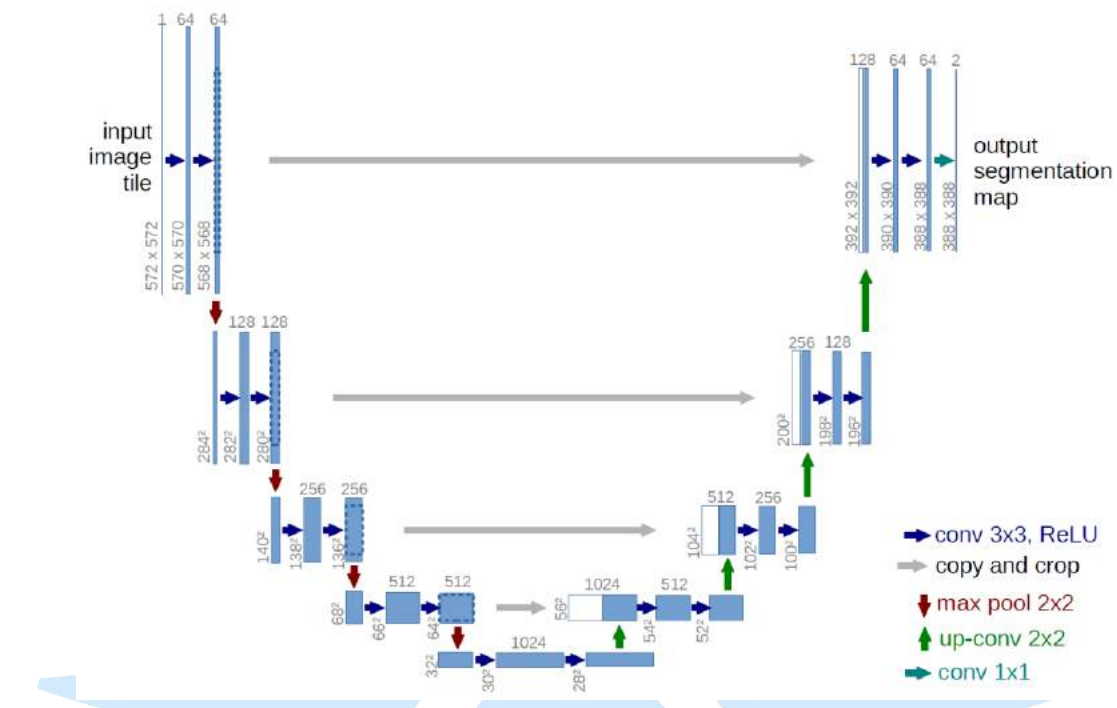


圖 4、U-net 之模型架構

由於醫學影像大多面臨切割對象的尺寸有巨大落差的情況，同樣的組織、器官在不同的個體、區域、病症會有極大的差異。所以許多方法嘗試改善 U-net，企圖讓網路能夠更有效地處理多尺度的問題。許多過去的方法透過使用不同大小的卷積核擷取不同尺度下的特徵 [27,28]，並透過不同的方式去聚合 (aggregate) 這些特徵，使得切割物體在不同尺度的情況都能夠較好地被處理，如圖 5。

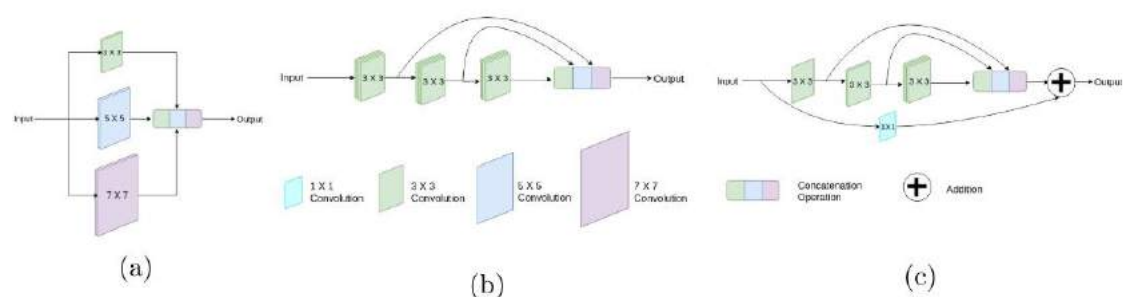


圖 5、MultiRes U-net 中提出了 MultiRes block

PSP (Pyramid Spatial Pooling) [7]提出了金字塔池化模組(Pyramid Pooling Module)使用了多個不同尺寸的池化層加上不同大小卷積層取得的特徵，並將這些新擷取的特徵與原始特徵串聯進入分類器得到最後切割結果。DeepLab [8,9]則針對了PSP 進行改良，提出了(Atrous Spatial Pyramid Pooling, ASPP)透過空洞卷積(Atrous Convolution)，也被稱作擴張卷積 (Dilated Convolution)，如圖 6，在不增加運算效能的前提下將接受野更大程度地進行擴張，使得 DeepLab 有著更加優秀的效能。[29, 30]同樣使用了擴張卷積，透過串接多個不同擴張率(dilate rate)的卷積核，試圖放大接受野增強器官以及器官邊界的辨識效能。

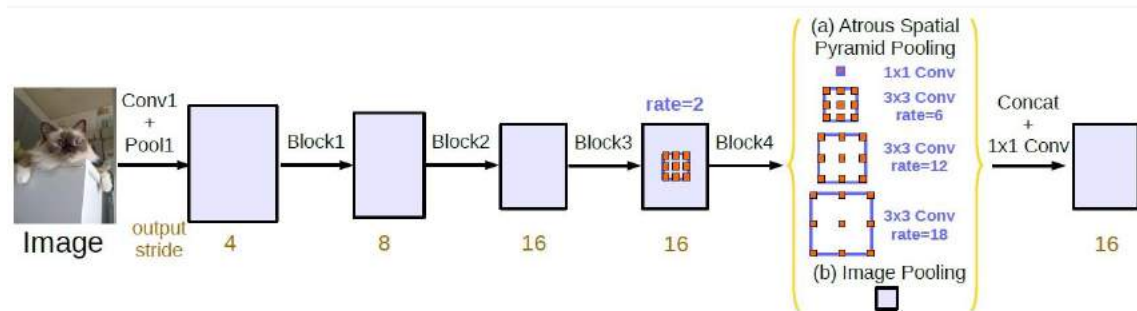


圖 6、DeepLab v3 架構以及 ASPP 模塊

2.1.2 基於三維全卷積網路之切割演算法

無論是透過電腦斷層或是核磁共振，大多數成像的醫學影像皆為三維數據。因此透過二維影像進行切割在直覺上是一種次優(suboptimal)的方法。因為在第三維度上，資料之間的相關性被忽略了，因此在醫學影像切割領域，許多方法也透過三維的卷積直接對三維資料進行處理。這樣做的優點就是能夠直接透過網路參數去學習第三維度之間的相關性，這對於器官、軟組織的切割、定位是很重要的資訊。然而，相較於二維全卷積網路，缺點也是顯而易見的，便是參數量的增加。不僅是卷積核增加了一個維度，經過處理所獲得的特徵也多了一個維度，這大幅

地增加了訓練時所需要占用的硬體資源。因此大多數基於三維卷積的方法會對於資料進行預處理，將長寬比先進行縮放，或是對資料進行裁切，透過降低輸入的資料量去限制訓練模型的大小。另外，基於三維卷積的方法大多要求資料為各向同性(isotropic)，希望資料在每個維度的變化兼保有相同性質。這取決於掃描參數的調整，也是使用基於三維卷積方法所必須注意的。

以下我們對於基於三維卷積的器官切割演算法進行介紹。首先，3D U-net [25]為早期對於三維 U-net 的重要研究，論文中探討了 2D U-net 和 3D U-net 對於 dense volumetric segmentation 的影響，以及透過二維標籤進行訓練對於三維資料切割的影響。其後，3D U-net 也開始被應用於多器官切割任務，Cascaded 3D U-net [4]則透過了第一階段先透過對輸入影像進行形態運算(morphological operations)用於提取候選區域 C1，此舉可以減少非必要計算的體素數量，提升運算效率。第二階段透過對測試階段的標籤進行擴張運算(image dilation)獲得候選區域 C2 最終獲得切割結果。nnUnet 架構[32, 36]提出了一種新的訓練架構，如圖 7。首先，經由過往的經驗與知識制定一系列規則用於決定訓練參數以及預處理方式。接著透過三個不同的 U-net 架構，包含二維 U-net、三維 U-net 以及使用了兩個三維 U-net 進行串接的串級 U-net (Cascaded U-net)，分別對資料集進行訓練以及交叉驗證，最後透過結合三個架構的結果以及後處理方式對測試資料進行預測。

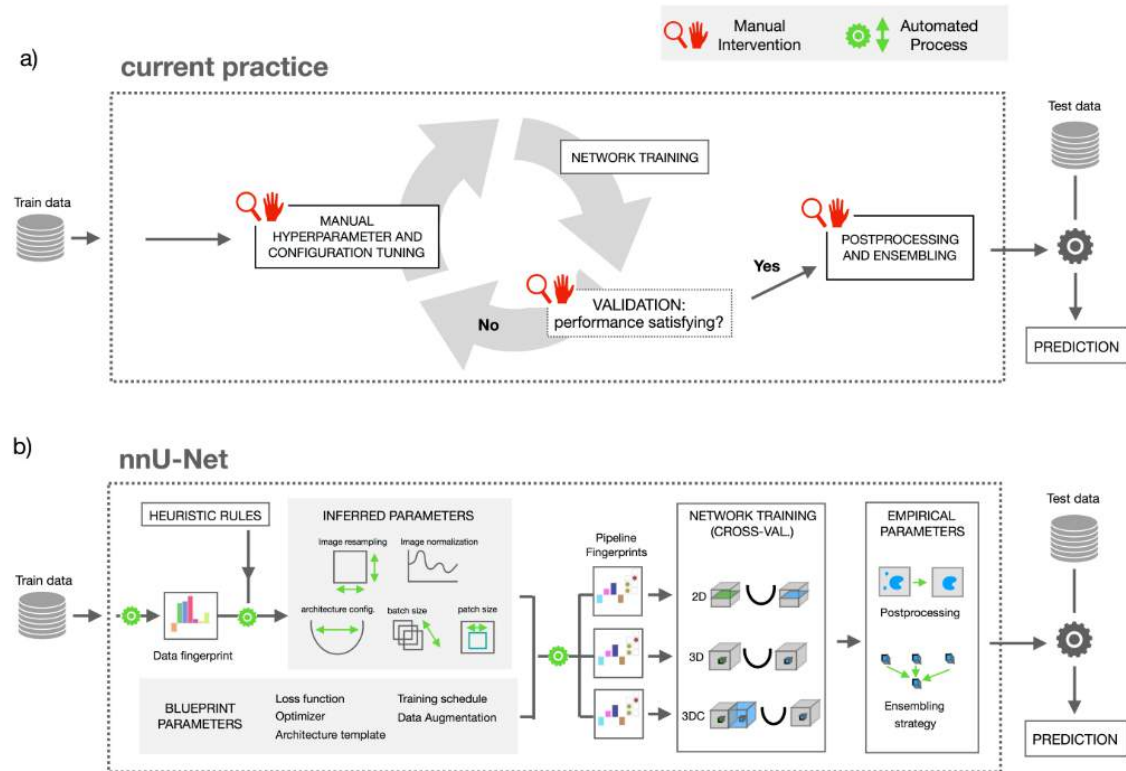


圖 7、nnUnet 提出之訓練架構

2.2 注意力模型應用於切割演算法

2.2.1 注意力機制

雖然基於全卷積網路的方法早在 2015 就已被提出[11]，但是至今仍然被廣泛應用於切割演算法。然而，全卷積網路仍有幾項缺點在近期被人提出。一、全卷積網路難以捕捉長期相依性(long-range dependence)。其名詞來自於自然語言處理領域，探討長序列中詞語間的依賴關係。若在影像相關任務則可以視作影像的全局語意資訊(Global Semantic)，若能在進行視覺相關任務時建立起長期依賴性，勢必對於模型效能有大幅地提升。因全卷積網路多仰賴卷積層進行特徵的擷取，多局限於局部的接受野。因此長期依賴性難以建立。二、近期的方法，如 DeepLab[8,

9]提出了透過空洞卷積的方式增加接受野的大小，試圖捕捉全局語意資訊。然而這樣的方式缺點在於無法建立密集的上下文資訊(Dense Context Information)，空洞化、零散的資訊難以建立起穩健的長期依賴性。

由於注意力機制近年來在圖像識別[21]、自然語音處理 [12]、醫學影像處理 [24, 34, 35]、視訊處理 [10]都有著良好的表現，且如章節 2.1.2 所描述，單純使用全卷積的方法開始逐漸出現侷限性。近年來開始有大量的影像切割方法開始引入注意力機制，並且取得相當卓越的成果。注意力機制有著多種不同的形式，包含了空間注意力 (spatial attention), 通道注意力 (channel attention)以及殘差注意力 (residual attention)等。這些方法目前尚未有一個統一定義的名詞，以及嚴格的數學定，總體來說，我們可以將注意力機制大致定義為「在原有的特徵上額外學習特徵間的相關性，並對特徵重新進行權重分配」。

2.2.2 基於注意力模型之切割演算法

注意力模型有許多種不同運作方式，[21]便提出了一種稱作殘差注意力網路 (Residual Attention Network)，也有論文稱為相加性注意力(additive attention)。這些論文透過額外的卷積層以及 sigmoid、softmax 等非線性函數對原始特徵進行運算獲得權重，這些權重便可視為原始特徵的重要程度，也可以透過這些權重了解網路在進行最佳化的過程中「注意」了那些特徵，因此也將這些權重稱為注意力。將這些注意力與原始特徵相乘後再與原始特徵相加在一起。不僅透過注意力重新對原始特徵進行重新分配，且在回傳梯度時，這樣的做法也能有效剔除無效的梯度，如圖 9。現在這樣的方法也大量被應用於腹部的器官切割，如[24, 35]。

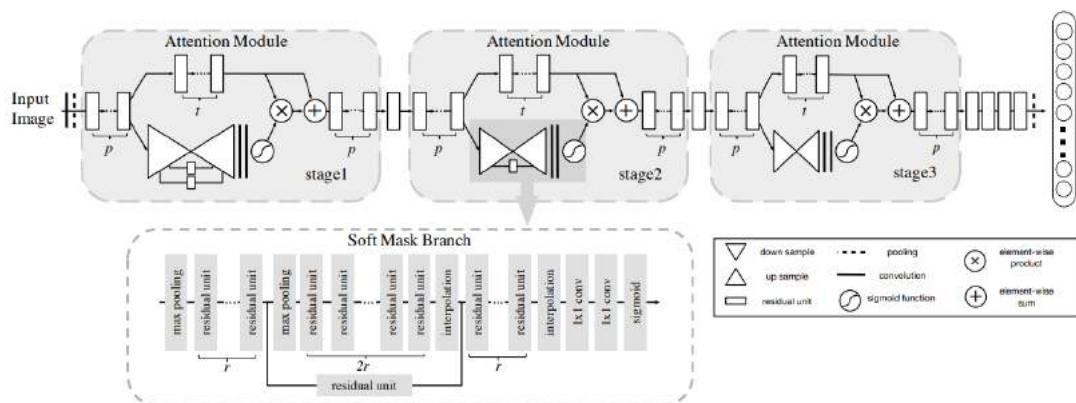


圖 9、殘差注意力網路之模型架構。

串級精化網路(Cascaded Refinement Network, CRN) [10]也應用了類似的方式於視訊物體切割任務，差別在於額外使用了基於光流及機率模型的方法計算了用於導引模型的注意力，稱作導引圖(Guidance)。此外，串級精化網路也通過低階特徵的使用更有效地進行特徵聚合(feature aggregation)，在與最先進模型(state of the art, SOTA)有相近的表現下能達到更快速的運算速度，如圖 10。Attention-gated Network [39] 提出了一種新的切割演算法結合 U-net 結構以及 Attention Gate (AG)。同時使用淺層特徵以及深層特徵計算注意力，並將此注意力用於深層特徵的空間權重化。Deep Attentional Features (DAF) [34] 提出了一種基於多尺度特徵的注意力機制。透過融合多個單一尺度特徵 (Single-layer features) 取得多尺度特徵 (Multi-layers features) 用於基於超聲波影像的前列腺切割。類似於 DAF，另外一篇論文也提出了相似的架構 [35]。不只使用了多尺度的注意力機制，更結合由 DANet [38] 提出的 Position self-attention 以及 Channel self-attention。對特徵進行更深入、複雜的注意力提取。也另外探討多次進行注意力提取對於整體效能的增進。

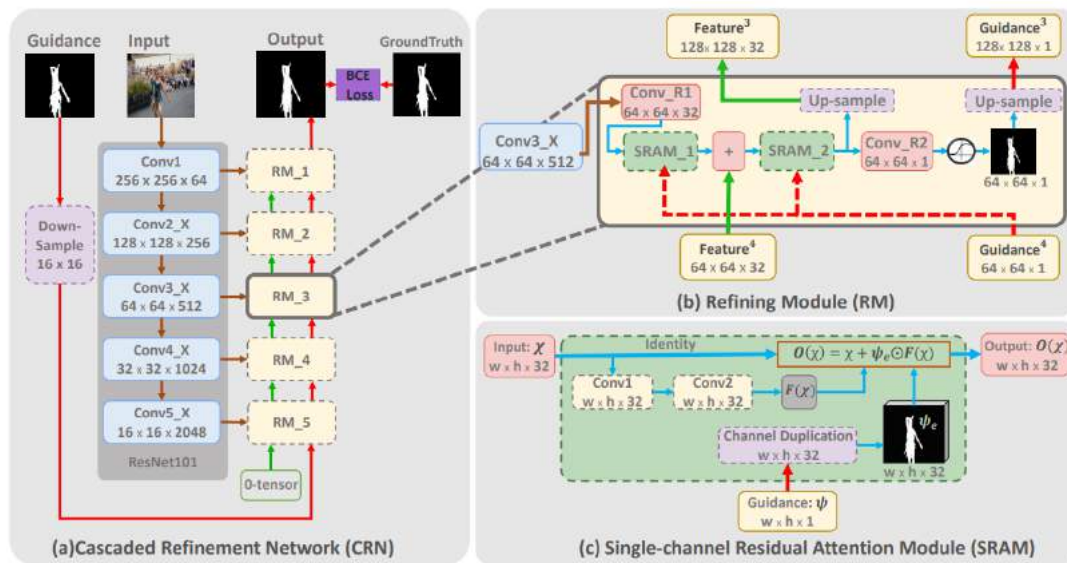


圖 10、串級精化網路之模型架構

第三章、基於先驗知識引導式之多器官切割方法

3.1 系統架構

人體是個相當複雜的構造，單一個體中包含了許多不同的器官及組織。透過電腦斷層掃描、核磁共振成像等方法我們可以針對這些資訊產生影像。然而這些影像往往為灰階影像並且缺乏紋理變化。因此本研究透過以下觀察針對多器官的切割問題設計了一系列的演算法。一、器官的形狀在個體間雖存有差異，但是其尺寸以及位於人體的位置是相似的。藉由產生導引圖，我們對於器官的分布進行模擬。二、器官為三維物體，電腦斷層掃描等技術可將其轉化為一序列影像，這些影像彼此相關且可互相參照。透過時序性網路的引入，我們針對其序列影像之間的關係進行學習。我們的研究架構稱為先驗知識引導網路 (Prior Guiding Network) 包含了一組用於提取特徵的編碼器 (Encoder)、解碼器 (Decoder)，用於進行多任務學習的人體縱向分類器 (Longitudinal classifier)、用於透過先驗知識 (Prior)產生導引圖 (Guidance)的導引圖產生模組 (Guidance generator)以及雙向卷積門控循環單元 (Convolutional Gated Recurrent Unit)用於捕捉序列影像間的時序性關係，如圖 11。

以下我們對於系統架構做初步的闡述。首先，醫學影像經過編碼器運算後可以得到初步的影像特徵，可於輸入解碼器後獲得切割結果，並引入人體縱向分類任務，使得模型在學習對器官進行切割的同時學習辨識輸入影像位於人體縱向位置，並按照其分類結果以及先驗知識取得導引圖。我們將會在章節 3.2 中進一步詳述多任務學習在此研究工作中的重要性。當模型學習到了輸入影像與人體層級的關聯性，就能透過此關聯性以及預先獲得的先驗知識產生出導引圖，我們將在

章節 3.3 具體說明如何取得先驗知識以及產生導引圖的方法。接著我們將此導引圖應用於解碼器當中，並在解碼器中引入了注意力機制，模型將根據導引圖對特徵給予不同的注意力（權重），我們將在章節 3.4 詳細說明解碼器架構以及注意力模型的運作機制，我們也將在第四章的實驗結果中對這些權重進行可視化，提供對於注意力的直觀理解。最後，我們嘗試使用循環神經網路學習序列影像之間的關係，並將每個時間所獲得的分類結果進行進一步的融合獲得最後的切割結果，我們將在 3.5 章節中說明我們所使用的循環神經網路—雙向卷積門控循環單元。

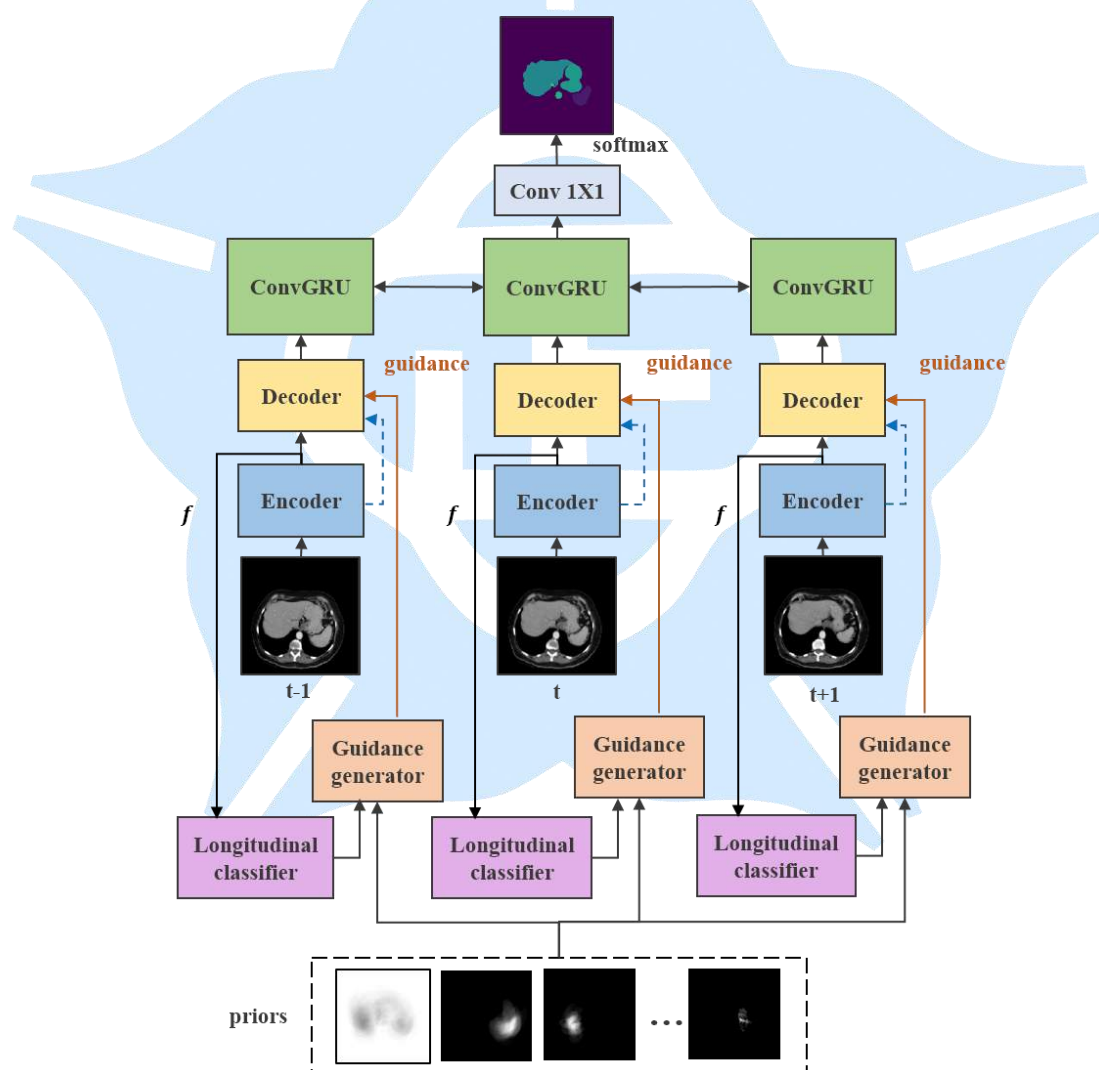


圖 11、本研究模型架構圖。

3.2 基於先驗知識之導引圖生成

多器官切割相對於自然影像切割有一個特性，在於輸入影像的相似性。儘管個體間多少存在差異，器官的位置、形狀分布卻遵循一固定模式。當人類專家在觀察某張腹部器官的電腦斷層掃描影像時，可以很輕易地推斷其來自於人體大致的縱向位置，反之亦然。我們可以從這樣的事實中了解到腹部器官影像與人體縱向位置間有著對應的關係。我們透過產生導引圖 (guidance)，用於對解碼器的特徵進行權重化，使得模型對於靠近器官的區域有更高的權重，提高辨識以及定位的精確程度。關於導引圖的應用，我們將在章節 3.3 更詳細地進行說明。此章節中我們將詳細說明如何生成導引圖。我們主要透過兩種方式產生導引圖，一、透過人體縱向位置的分類結果結合預先定義的先驗知識，如圖 14 (a)。二、在編碼器的最後一層額外添加一分類器，透過二元切割結果及額外的相加運算作為導引圖，如圖 15 (a)。

我們首先對第一種導引圖生成方式進行說明。如我們於章節 3.1 中所描述，我們認為相同器官在不同個體中存在的位置是大致相似的。因此我們透過將不同個體中的訓練影像按照器官、人體縱向位置進行分群，並將同一群的訓練影像進行平均運算以及正規化以獲得先驗知識。由於模型需要動態地根據輸入決定所使用的先驗知識，我們引入額外的人體縱向位置分類任務，在取得經過編碼器所處理的影像特徵 f 後，使用全局平均池化 (Global Average Pooling, GAP) 對影像特徵進行壓縮，之後經過兩層的全連接層 (Fully Connected Layer) 獲得縱向位置的分類結果 f_z 。最後透過交叉熵 (cross entropy) 計算分類結果與預先定義的標籤間的損失。經由縱向位置分類器 (Longitudinal classifier)，如圖 13，我們可以推估各器官約略位於人體縱向的位置，透過這樣的資訊結合先驗知識，我們可以得到所需之導引圖。另外，模型透過同時進行影像資訊的編碼以及判別輸入的縱向位置進行多任務學習，我們認為這樣的方式可以有效地提升最後的切割結果，我們

也將在章節 4.2 中進行相關的驗證。

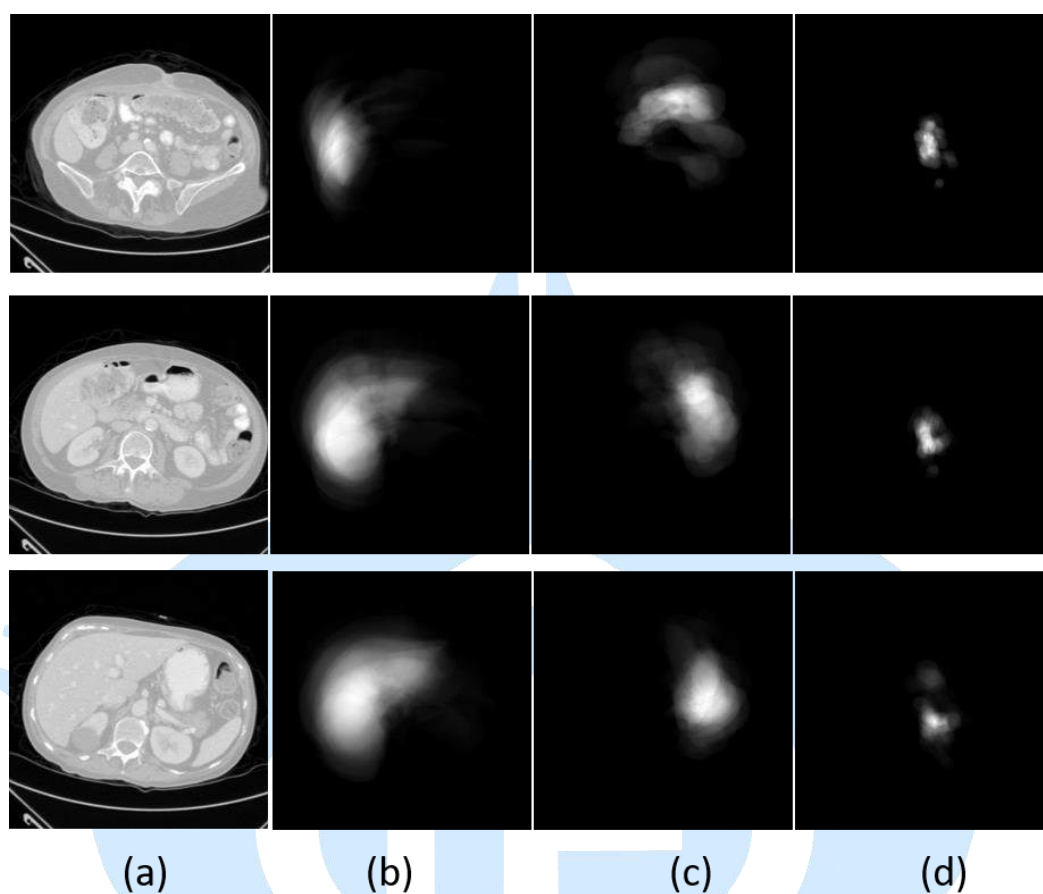


圖 12、透過訓練標籤的分群及正則化所預先產生之先驗知識。(a)為輸入之 CT 掃描影像。(b)-(d)則分別代表了肝臟、胃部以及主動脈的先驗知識

另外，我們也嘗試透過第二種方式進行導引圖的生成，我們在編碼器的最後一層額外添加分類器，獲得每個像素的機率分布，接著對每個像素的分類結果進行平均求得導引圖。我們希望透過這樣的操作，可以凸顯容易被分類器混淆的像素，如圖 18。這些像素大多出現在器官的邊緣、組織與器官的交接處等。如我們在章節 1.2 所描述，這些像素通常不易被分類正確。因此透過提高權重使得模型能夠更好地處理這些像素。

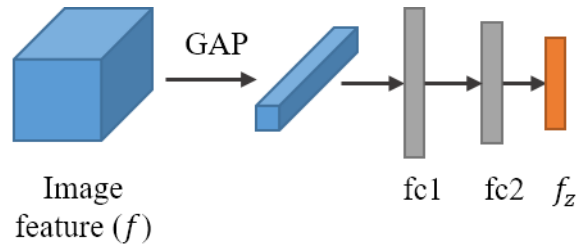


圖 13、縱向位置分類器之詳細架構

3.3 基於導引圖之全卷積網路

首先，我們使用 ResNet-50 [15] 作為編碼器的架構，並參考了 DeepLab [8,9] 中對於編碼器的設計，將最後兩階層(stage)中的卷積運算替換為空洞卷積 (Dilated convolution)，目的在於在擴大接受野的同時保留空間資訊，另外，由於人體中某些器官相當小若下採樣過大將導致導引圖失真嚴重，空洞卷積也很好地解決了這樣的問題。我們的解碼器參考了串級精化網路(Cascaded Refinement Network, CRN) [10]，對其加以改良使其符合我們的任務，使用的模組分別為 (a) 基於導引圖之全卷積網路 (Guidance based FCN) (b) 精化模組 (Refining Module, RM) (c) 單通道殘差注意力模組 (Single Residual Attention Module, SRAM)，如圖 14。首先我們沿用了單通道殘差注意力模組中的設計，透過章節 3.2 中所描述之方法產生導引圖並對特徵進行權重分配以及殘差運算。而在精化模組中，我們針對章節 3.2 中不同的導引圖生成方式使用了兩種不同的架構，並將其分別稱之為 PGN-v1、PGN-v2。

一、我們同時使用了多個單通道殘差注意力模組用於關注不同的類別（器官）。我們也配合了多器官切割任務，將原始進行二元分類的 Sigmoid 函數替換成進行多分類的 Softmax 函數。最後，多器官切割需要同時考慮多個類別以及小物體的切割，因此額外添加損失函數於每層的導引圖結果，以確保導引圖運作的正確性，如圖 14。

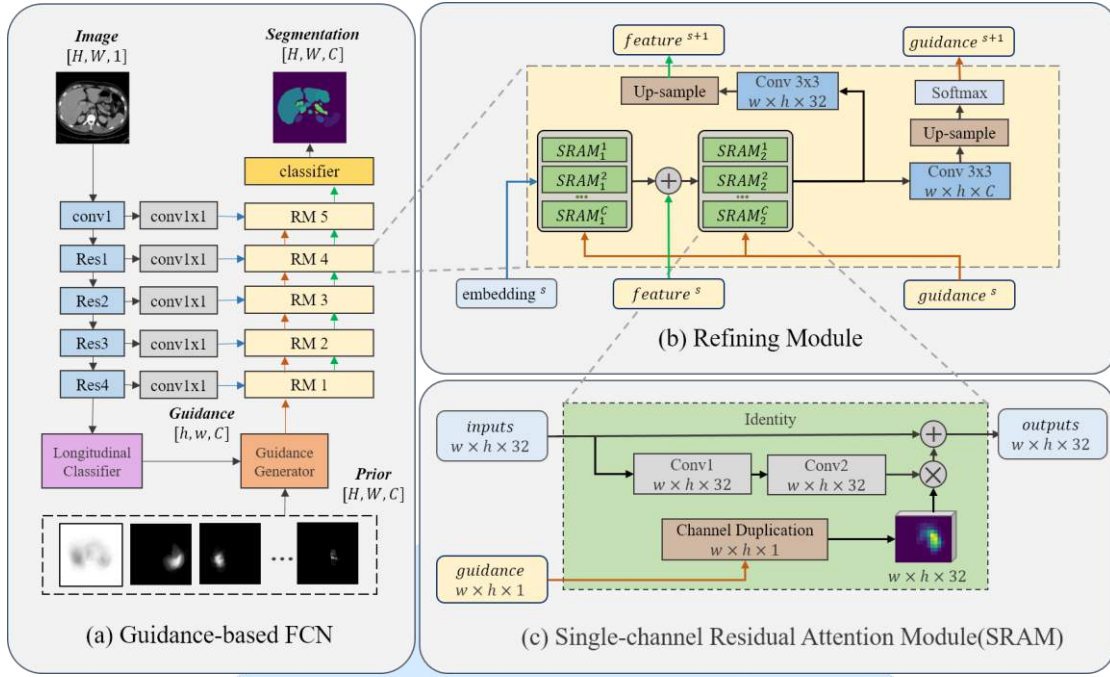


圖 14、本研究工作編碼器及解碼器架構之一，PGN-v1

二、由於每個層級的 SRAM 模組都需要針對每個類別使用不同的卷積層，用以萃取不同類別的注意力。這樣的方式雖然能夠針對不同類別的特徵進行編碼。但是會使得參數量較大，訓練較不容易。因此我們嘗試融合各類別的切割結果去凸顯易被分類錯誤的像素，稱之為反向注意力 (Reverse Attention)，如圖 15 (a)。此外，我們在編碼器的一開始就引入先驗知識，與影像串連後進入編碼器。這麼做的目的在於透過各類別的先驗知識引入語意概念於編碼器中，則我們將能在淺層的切割就能獲得較佳的結果。接著我們透過編碼器進行切割任務，嘗試利用粗略的切割結果 (Coarse Segmentation) 作為解碼器的導引圖。最後，我們觀察到透過單通道殘差注意力模組傳遞特徵，容易會有累積錯誤的問題，也就是在低解析度切割的結果所造成的錯誤無法隨著特徵的傳遞而被消除，反而會持續傳播造成高解析度切割的錯誤。因此引入額外的卷積層解決此問題，如圖 15 (b)。

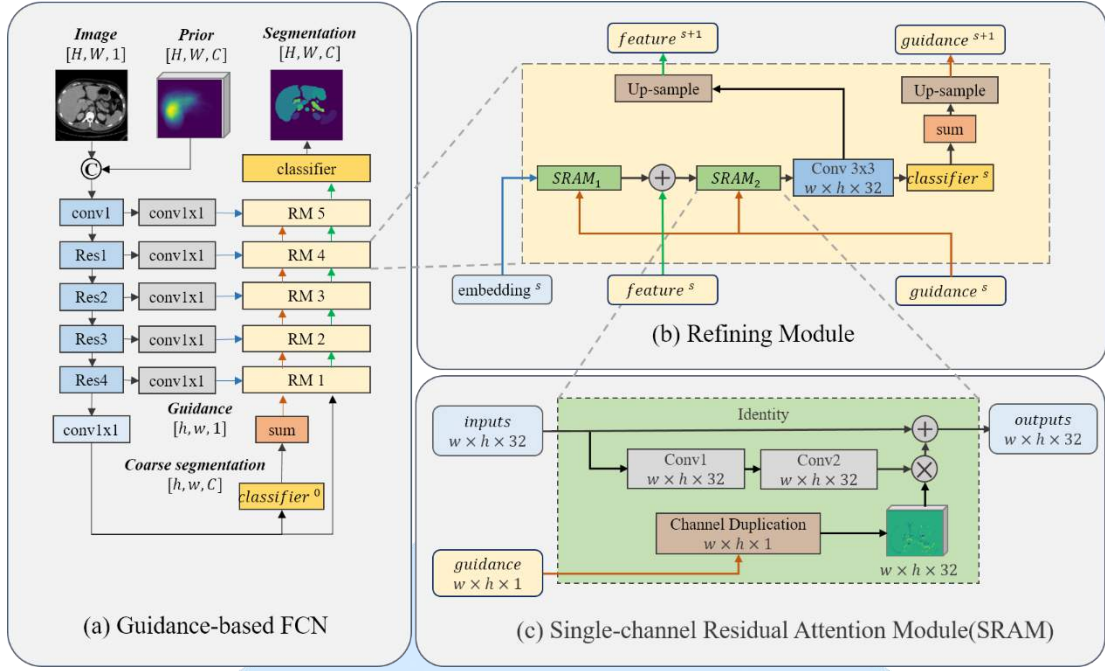


圖 15、本研究工作編碼器及解碼器架構之二，PGN-v2

3.4 雙向卷積門控循環單元

最後，我們在模型的最後使用了雙向卷積門控循環單元(Bidirectional Convolutional Gated Recurrent Unit, Bidirectional GRU)用於學習時序性的特徵。GRU[14]為長短期神經網路(Long Short Term Memory, LSTM) [16]的一種變體，簡化了LSTM對於閥的設計。使用重置閥 (Reset Gate)決定使用多少過去隱藏狀態 h_{t-1} 來決定候選狀態 \tilde{h}_t 。使用更新閥(Update Gate) z_t 將 LSTM 中的遺忘閥(forget gate) 與輸入閥 (input gate) 進行替代，透過單一的更新閥決定過去隱藏狀態 h_{t-1} 以及候選狀態 \tilde{h}_t 的比重。在我們的應用裡每個時刻的 x_t 即是每個時刻網路的最後一層特徵，透過時序模型學習到了序列間的依賴關係，透過這樣的關係獲得最後的輸出隱藏狀態 h_t ，並用於預測最後的分割結果：

$$r_t = \sigma(W_{xr} * x_t + W_{hr} * h_{t-1}) \quad (1)$$

$$\tilde{h}_t = \tanh(W_{xh} * x_t + W_{hh} * (r_t \cdot h_{t-1})) \quad (2)$$

$$z_t = \sigma(W_{xz} * x_t + W_{hz} * h_{t-1}) \quad (3)$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \quad (4)$$

3.5 目標函數

我們的研究工作中使用了兩種不同的編碼器架構，兩者所使用的目標函數雖然大致相同，但是在細節上有著細微的差異，以下我們分別進行描述。在第一種架構中，我們總共使用了三個目標函數於我們的研究工作中，如式(5)。包含了切割損失 L_{seg} 、導引圖損失 L_{guid}^s 以及縱向位置分類損失 L_z ， K 代表了總共產生的引導圖數量。我們依據函數的重要性給予了不同的權重 λ_{guid} 、 λ_z 並透過梯度下降方法共同最佳化了目標函數 L_{total} 。

$$L_{total} = L_{seg} + \lambda_{guid} \sum_{s=1}^K L_{guid}^s + \lambda_z L_z \quad (5)$$

在多器官切割任務中，我們關注的器官相較於背景通常只佔了影像中非常小的一部分，考慮我們的資料面臨了類別不平衡的情況。我們的 L_{seg} 、 L_{guid} 皆參考了 V-net [22] 所提出的 Dice Loss，如式(6)。 N 代表了像素數量， C 代表了切割任務的類別數量、 p 、 q 則分別代表了切割結果以及真實的切割標籤。由於分母使用了預測以及標準答案的相加，器官的大小自然地成為權重使用，因此能夠較好的處理類別不平衡的問題。除了在多器官切割任務添加切割損失 L_{seg} ，我們也對於所有網路預測產生的導引圖加上導引圖損失 L_{guid} 。透過引入真實切割標籤的監督，導引圖的品質得以提升並更好地權重化特徵。

$$L_{seg} = L_{guid}^s = -\frac{1}{C} \sum_{i=1}^C \frac{2 \sum_j^N p_{i,j} q_{i,j}}{\sum_j^N p_{i,j} + \sum_j^N q_{i,j}} \quad (6)$$

縱向位置分類損失 L_z 則使用了一般常用於分類任務的交叉熵損失 (Cross Entropy Loss) 做為目標函數，如式(7)。 Z 代表了縱向分類任務的類別數量、 p 、 q 則分別代表了切割結果以及真實的切割標籤。

$$L_z = -\frac{1}{Z} \sum_{i=1}^Z q_i \log(p_i) \quad (7)$$

在第二種架構中所使用的目標函數略有不同，如式(8)。同樣透過 Dice Loss 作為切割損失 L_{seg} 。但是如章節 3.2 中所提到的，我們使用了另一種導引圖產生方式，因此導引圖損失 L_{guid} 則由多分類任務轉化為多個二分類任務，對各個類別加上二元交叉熵損失 (Binary Cross Entropy Loss)，如式(9)。

$$L_{total} = L_{seg} + \lambda_{guid} \sum_{s=1}^K L_{guid}^s \quad (8)$$

$$L_{guid} = -\frac{1}{NC} \sum_{k=1}^C \sum_{i=1}^N q_{i,k} \log(p_{i,k}) \quad (9)$$

第四章、實驗結果與討論

4.1 實驗配置

4.1.1 實作細節及相關配置

我們根據 [6] 中的設定，使用了 $[-125, 275]$ Hounsfield Unit (HU) 作為 CT 影像的值域範圍。接著不論是 CT 或是 MRI 都將重新縮放值域到 $[0, 1]$ 。為了降低模型過擬合的可能，我們也使用了資料增強 (data augmentation) 的方式增加資料的多樣性，包含了隨機影像縮放、裁剪。我們使用了 Adam 作為最佳化的方法，並且使用了 0.001 的權重衰減 (weight decay)。我們分別對於單張影像以及影像序列進行實驗，若輸入只使用單張影像，則批次大小 (batch size) 為 16 張影像。當使用影像序列作為輸入，我們使用 3 張連續影像作為一個序列，每個批次大小則為 4 個序列。我們參考了[8, 9]的學習率 (learning rate) 設定方法。使用 0.0075 作為起始的學習率，之後隨著訓練步階增加而逐步衰減，衰減率設定為 0.9。模型在訓練十二萬步階後紀錄最佳模型並結束訓練。最後，縱向分類損失權重 λ_z 以及引導圖損失權重 λ_{guid} 分別設置為 1.0 及 1.0。

本研究以 Python 3.6 作為程式語言，並以 Tensorflow 作為深度學習框架，詳細實驗相關設備表列如下：

表 1、實驗環境配置

電腦硬體規格	作業系統	Ubuntu 16.04 LTS
	處理器	Intel® Xeon® W-2125 Processor @4.5GHz
	記憶體	64GB
	顯示卡	GeForce GTX 1080 Ti
整合式開發環境		Visual Studio Code 1.48.2
程式語言		Python 3.6.2

主要使用函式庫表列如下：

表 2、主要使用函式庫。

Library	Version
Numpy	1.16.3
Opencv-python	3.1.0
Matplotlib	2.2.3
Nibabel	3.0.1
Simple ITK	1.2.4
Tensorflow	1.14.0
Cuda	10.2

4.1.2 資料集

2015 MICCAI Multi-Atlas Abdomen Labeling Challenge

我們使用了由 2015 MICCAI Multi-Atlas Abdomen Labeling Challenge 所釋出之 CT 腹部器官資料集[26]作為本研究訓練及測試使用。此資料集為已知現有最大的 CT 腹部器官公開資料集，訓練集提供了 30 組腹部器官掃描，由 3,779 張軸向對比增強腹部臨床 CT 圖像 (axial contrast-enhanced abdominal clinical CT images) 所構成，測試集提供了 20 組腹部器官掃描，由 2,387 張軸向對比增強腹部臨床 CT 圖像所構成。標註了人體腹部的十三種器官，包含了脾，右腎，左腎，膽囊，食道，肝，胃，主動脈，下腔靜脈 (IVC)，靜脈和脾靜脈，胰腺，左腎上腺，右腎上腺。每個 Volume 由 85~198 張 512×512 影像所組成，體素空間解析度 (voxel spatial resolution) 則為 $[0.54 \sim 0.54] \times [0.98 \sim 0.98] \times [2.5 \sim 5.0] \text{ mm}^3$ ，我們將關於資料集的重要資訊整理為下表，如表 3。

表 3、2015 MICCAI Multi-Atlas Abdomen Labeling Challenge 資料集

Multi-Atlas Abdomen Labeling Challenge [31]	
Training images	30 cases
Testing images	20 cases
Modality	CT
Resolution	512 × 512
Number of Organ categories	13
Annotated organs	(1) Spleen, (2) right kidney, (3) left kidney, (4) gallbladder, (5) esophagus, (6) liver, (7) stomach, (8) aorta, (9) inferior vena cava, (10) portal vein and splenic vein, (11) pancreas, (12) right adrenal gland, (13) left adrenal gland

2019 Combined (CT-MR) Healthy Abdominal Organ Segmentation (CHAOS challenge)

為了測試架構的泛化性，我們也使用了由 The IEEE International Symposium on Biomedical Imaging (ISBI) 於 2019 年舉辦的挑戰 Combined (CT-MR) Healthy Abdominal Organ Segmentation (CHAOS challenge 中所提供的資料集 [31]，此資料集提供了兩個資料庫。第一個資料庫提供了 40 組測試者的腹部電腦斷層掃描，針對所有測試者的肺臟進行了標註。每個三維體積由 77~105 張 512×512 影像所組成，合計共 1,367 張影像提供訓練，1408 張影像提供測試。x-y spacing 為 0.7-0.8 mm、inter-slice distance (ISD) 為 3.0-3.2 mm。第二個資料庫提供了 120 組測試者的腹部核磁共振造影，針對所有測試者的四個器官進行標註，分別為肺臟、左腎臟、右腎臟以及脾臟。每個三維體積由 26~50 張 256×256 影像所組成，合計共 1,594 張影像提供訓練，1,537 張影像提供測試。x-y spacing 為 1.36-1.89 mm、inter-slice distance (ISD) 為 3.0-3.2 mm，我們將關於資料集的重要資訊整理為下表，如表 4。

表 4、2019 Combined (CT-MR) Healthy Abdominal Organ Segmentation 資料集

Combined (CT-MR) Healthy Abdominal Organ Segmentation [26]		
Database	Database 1	Database 2
Training images	20 cases	60 cases
Testing images	20 cases	60 cases
Modality	CT	T1-DUAL in phase 40 data sets), out phase (40 data sets) and T2- SPIR (40 data sets)
Resolution	256 x 256	512 x 512
Number of Organ categories	1	4
Annotated organs	Liver	Liver, Left Kidney, Right Kidney, Spleen

4.1.3 評估指標

1. Dice 係數 (Dice Similarity Coefficient, DSC)

Dice 係數最常被使用於醫學切割領域，用以評分切割的品質。 A 為切割結果， B 則為參考標準。分母為切割結果以及參考標準的總和，包含了 A 以及 B 的交集以及聯集。分子則為切割結果與參考標準之間的交集，乘上常數二則可使 Dice 係數的值域範圍落在 0 到 1 之間。其物理意義在於透過以 A 、 B 兩者共同涵蓋的範圍為條件下，去計算 A 、 B 兩者重疊的程度。重疊的程度愈高則代表切割結果 A 愈契合參考標準 B 。

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

2. 最大表面距離 (Maximum Symmetric Surface Distance)

也稱作郝斯多夫距離 (Hausdorff distance)，相較於透過重疊程度進行切割品質的評估，另一種方式則透過 A 、 B 兩者間距離的量測。將 $S(B)$ 定義為 B 的所有表

面體素(surface voxels)，而從任意體素到 $S(B)$ 的最短距離 $d(v, S(B))$ 定義為：

$$d(v, S(B)) = \min_{s_B \in S(B)} \|v - s_B\|$$

將此定義擴展為 A 之中的所有體素，則可以得到從 A 到 B 的所有最短距離，這個統計數據代表了 A 相對於 B 的不匹配程度。進一步求得之中的最大距離，則為 A 相對於 B 的最大不匹配程度：

$$\max_{s_A \in S(A)} d(s_A, S(B))$$

反之，也可以求得 B 相對於 A 的最大不匹配程度。郝斯多夫距離便是兩者間的最大值。透過量測郝斯多夫距離，我們得以理解兩個點集間的最大不匹配程度：

$$MSSD(A, B) = \max \left\{ \max_{s_A \in S(A)} d(s_A, S(B)), \max_{s_B \in S(B)} d(s_B, S(A)) \right\}$$

3. 平均表面距離 (Average Symmetric Surface Distance)

平均表面距離與郝斯多夫距離有著相似的概念，相較於郝斯多夫距離所取得的最大不匹配程度，平均表面距離由於計算了所有距離的平均，較可以有效地抑制離群值(outlier)對於距離估計的影響。

$$ASSD(A, B) = \frac{1}{|S(A)| + |S(B)|} \left(\sum_{s_A \in S(A)} d(s_A, S(B)) + \sum_{s_B \in S(B)} d(s_B, S(A)) \right)$$

4. 相對體積差異 (Relative Absolute Volume Difference)

若 A 為切割結果， B 為參考標準，相對體積差異為 A 、 B 兩者之間的尺寸絕對值差異 (Absolute size difference)相對於 B 的尺寸絕對值。針對切割結果進行尺寸差異的估計，無法直接地評估切割品質，因此在醫學影像切割中通常必須與其他評估方法搭配使用。

$$RAVD(A, B) = \frac{|B| - |A|}{|B|}$$

4.2 實驗結果

4.2.1 消融研究

本研究首先對於 PGN-v1 不同模組的使用進行了消融研究，如表 5。使用了訓練資料集中的 25 組腹部器官掃描，由 3,111 張影像所組成作為訓練，使用了訓練資料集中的另外 5 組腹部器官掃描，由 6,68 張影像所組成作為測試所得到的數據。透過觀察可以發現，首先我們不額外增加任何模組於模型上，則模型退化成為 U-net，可以發現切割結果並不是太理想，其 DSC 只有 0.729。接著我們加入了多任務學習，也就是透過縱向分類器 (Longitudinal Classifier) 於深層特徵中同步學習縱向位置的分類，我們可以發現透過這樣的方式切割的精度有了大幅度地提升。接著我們引入了基於引導圖之全卷積網路(Guidance-based FCN) 透過引導圖對特徵進行權重化，並在每個尺度進行引導圖的學習。最後我們使用了雙向卷積門控循環單元 (Bidirectional Convolutional Gated Recurrent Unit, Bidirectional GRU) 作為 RNN 模型，我們可以發現時序性的資訊也能夠改善最後的切割效果。

表 5、各模組使用對於 PGN-v1 整體性能影響的比較，LC 表示 Longitudinal Classifier 的使用，**Guid-FCN** 則表示 Guidance-based FCN，**RNN** 則表示 Recurrent Neural Network，在我們的研究工作所使用的是 Bidirectional Convolutional GRU。

	LC	Guid-FCN	RNN	DSC (↑)
1				0.729
2	✓			0.803
3	✓	✓		0.826
4	✓	✓	✓	0.831

另外，為了理解各模組對於 PGN-v2 所造成的影響，我們對於模組的使用進行消融研究並取得了表 6。首先可以發現，整體來說加入 prior 能夠有效地提升切割結果的品質，我們推測是引入語意資訊於編碼器的成果。另外，相比於單純使用 U-net，SRAM 1 所帶來的效果提升也是較為明顯的，SRAM 2 雖對於結果有所提升，但是差異性較小。最後，實驗證實引入在精化模組的最後引入額外的卷積運算是關鍵的，即使只有一層的卷積，對於效果的提升卻是十分有幫助的。

表 6、各模組對於 PGN-v2 整體性能影響的比較。**prior** 表示有無共同輸入先驗知識於網路當中，**SRAM 2** 代表每個 RM 中是否使用第二次的 SRAM 以及 **conv** 則表示在 RM 的最後有無額外引入卷積運算。

	prior	SRAM 2	conv	DSC (↑)	Precision	Recall
1				0.8311	0.8454	0.8157
2		✓		0.8357	0.8560	0.8243
3		✓	✓	0.8430	0.8632	<u>0.8268</u>
4	✓			0.8419	0.8653	0.8246
5	✓	✓		<u>0.8433</u>	0.8759	0.8177
6	✓	✓	✓	0.8601	<u>0.8687</u>	0.8501

4.2.2 與其他方法之比較

我們也將本研究之方法與同樣使用了 2015 MICCAI Multi-Atlas Abdomen Labeling Challenge 資料集的其他方法進行比較，如表 7。雖然我們的方法贏過了大多數的過去方法，但是整體結果仍然略差於目前的最佳模型。我們認為改善以下幾點，將能帶來更好的成果。一、雖然我們在實驗中證明透過注意力機制以及引導圖對於特徵進行權重化是重要的，但是的做法似乎仍然不夠泛化，如何針對腹部器官切割資料，學習更為泛化有力的特徵將會是改善最後結果的

關鍵之一。二、目前最佳模型 nnUnet 使用了 3D U-net 最為其主要架構，足以可見序列資訊的重要性。目前我們的研究只針對最後的切割結果進行序列關係的學習，若擴大序列資訊的利用，或許能夠再提升切割結果。

表 7、本研究架構於 2015 MICCAI Multi-Atlas Abdomen Labeling Challenge 之評估結果以及其他方法的比較

	DSC (↑)	ASSD (↓)	MSSD (↓)
AutoContext3DFCN[6]	0.782	1.936	26.095
deedsJointCL[19]	0.790	2.262	25.504
Dlk0.1_unet_sub2[13]	0.815	1.861	62.872
PaNN[20]	0.850	1.450	18.648
nnUnet[32,36]	0.881	1.3894	17.2583
Ours (PGN-v1)	0.860	1.849	36.116
Ours (PGN-v1-BiConvGRU)	0.862	1.7395	32.984

我們也將本研究之方法與同樣使用了 2019 Combined (CT-MR) Healthy Abdominal Organ Segmentation 資料集的其他方法進行比較。如章節 4.1.2 所描述此資料集包括了兩種不同的資料。透過這些資料的組合構成了不同的任務 (Task)。本研究工作主要專注於 MRI 腹部器官切割的 Task5。我們將結果整理如表 8。透過結果我們可以發現我們的方法即使是使用於不同資料以及模態 (Modality) 仍然是有效的且具有競爭力的。CHAOS 資料集透過所有即時排行榜的資訊，將所有評估方法的數據換算成百分比並且取得平均成績為表格中的 Mean Score。我們的成績取得了第三順位，透過觀察和分析發現我們的方法在 ASSD、MSSD 等距離的量測表現較差，因我們的模型在預測時會出現較為嚴重的假陽性(False Positive) 或者是假陰性 (False Negative)的問題。當量測

ASSD、MSSD 等需要對整個三維切割結果進行評估時，就會對於最後的距離量測結果有較大的影響。這也是我們日後需要進行加強及改善的部分。

表 8、本研究架構於 CHAOS dataset Task5: MRI 腹部器官切割之評估結果以及其他方法之比較

Team Name	Mean Score	DSC (\uparrow)	RAVD (\downarrow)	ASSD (\downarrow)	MSSD (\downarrow)
nnU-Net	72.44 ± 5.05	0.95 ± 0.02	5.07 ± 2.57	1.05 ± 0.55	14.87 ± 5.88
PKDIA	66.46 ± 5.81	0.93 ± 0.02	6.91 ± 3.27	1.43 ± 0.59	20.1 ± 5.90
Mountain	60.2 ± 8.69	0.90 ± 0.03	8.04 ± 3.97	2.27 ± 0.92	25.57 ± 8.42
ISDUE	56.25 ± 19.63	0.83 ± 0.23	18.33 ± 27.58	5.82 ± 11.72	32.88 ± 33.38
METU_MMLAB	56.01 ± 6.79	0.89 ± 0.03	12.44 ± 4.99	3.21 ± 1.39	32.70 ± 9.65
Ours (PGN-v2)	60.86 ± 17.97	0.88 ± 0.10	12.33 ± 11.12	2.85 ± 2.82	41.16 ± 35.54
Ours (PGN-v2-BiConvGRU)	63.58 ± 9.97	0.91 ± 0.05	9.89 ± 6.95	2.16 ± 1.52	25.88 ± 14.23

4.3 可視化結果

4.3.1 二維器官切割可視化結果

我們將最後的切割結果透過二維影像進行可視化。如章節 1.2 所描述，透過觀察可以發現我們的方法可以有效地對於腹部器官進行切割。即使面對較小的器官，我們的模型也能夠給予正確的位置及分類。然而我們也可以發現模型在面對一些非器官的軟組織時容易產生錯誤。另外當器官出現較不尋常的紋理變化時，模型也容易造成判讀。以及如圖 16。

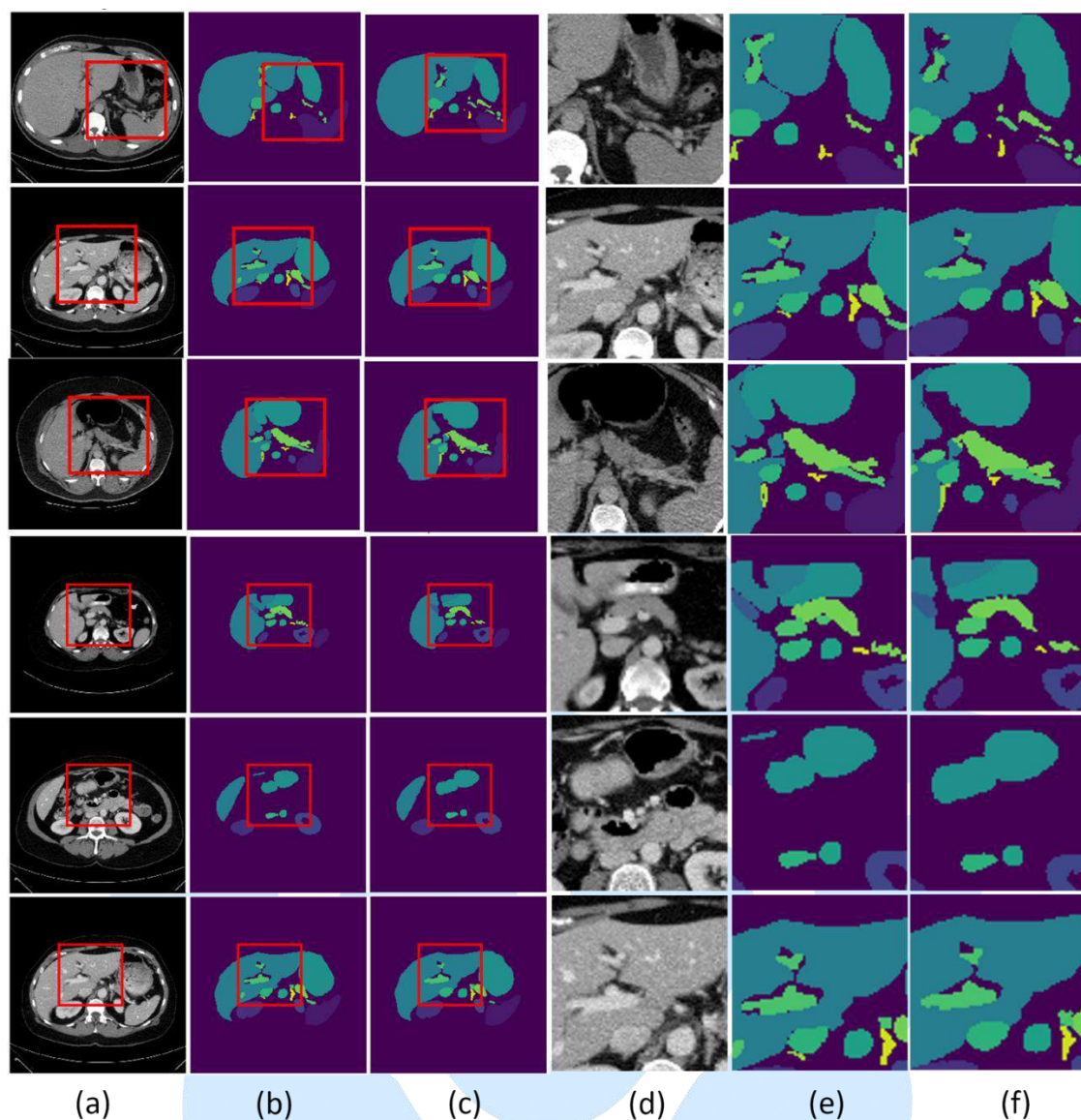


圖 16、腹部多器官切割結果。(a)輸入腹部 CT 影像、(b)參考標準、(c)模型預測切割結果。(d)-(f)則分別為(a)-(c)的紅色方框區域放大結果。

4.3.2 注意力可視化結果

我們先針對 PGN-v1 的注意力進行可視化，如圖 16。特別注意此處之導引圖為模型自行產生，針對不同尺寸大小的特徵進行切割任務的預測，與圖 12 之輸入先驗知識有所不同。由於我們在每一層導引圖都透過了不同尺度大小的真實切割結果計算了導引損失做為導引圖的限制，可以看見每層導引圖都因此關注了與

真實器官所在位置相近的區域。

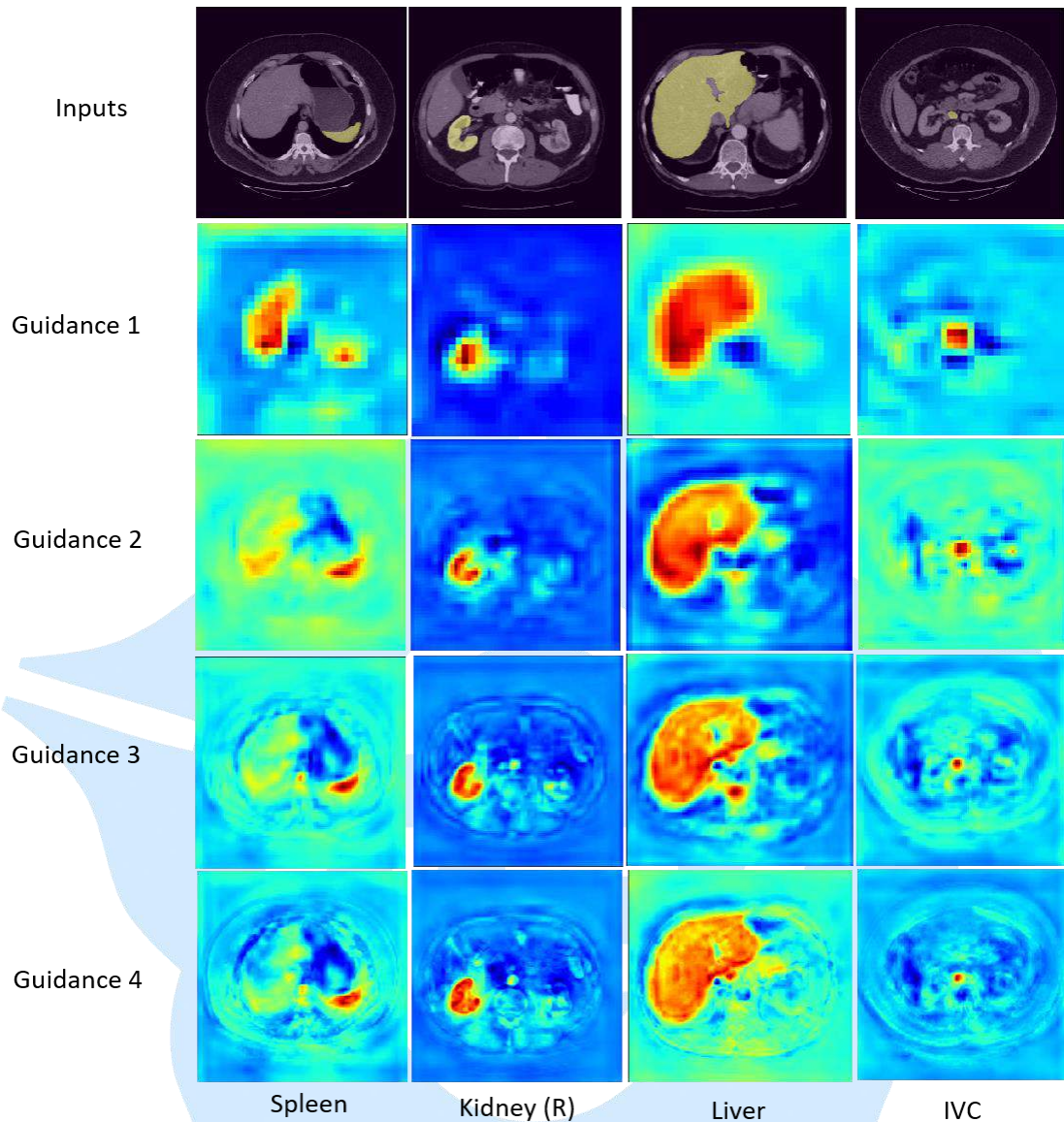


圖 17、注意力可視化結果。由上而下分別代表了輸入影像以及各層的導引圖，由左至右分別代表了不同樣本中不同器官的導引圖，依序為脾臟、右腎臟、肝臟以及下腔靜脈。

我們也針對提出的反向注意力 (Reverse Attention) 做出可視化，如圖 18。我們發現反向注意力在低解析度時，如 Reverse Attention 1~ Reverse Attention 3。大多關注於器官間的交界處，這些區域也是進行切割時最容易出現錯誤的。而後隨著解析度的提高，關注的對象逐漸轉移到部分特定器官的邊界，而後至整體器官。

可以看見到 Reverse Attention 5 時，由於切割的精度提高，所有類別都有不錯的分割效果，關注特定區域的程度也逐漸趨於平緩。

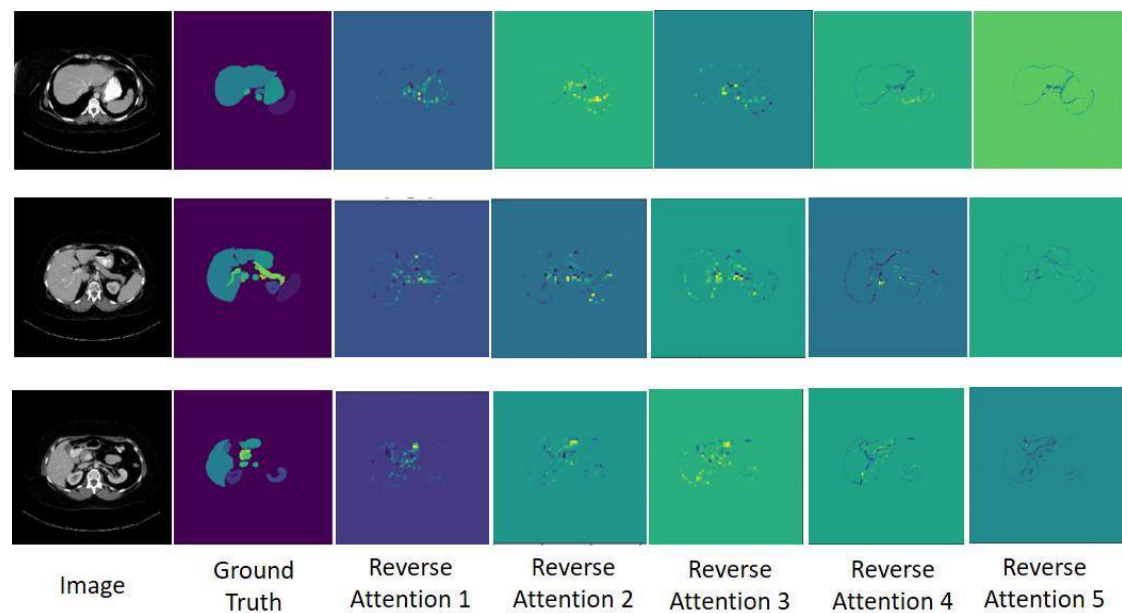


圖 18、反向注意力之可視化結果

第五章、結論

5.1 結果討論

本研究工作針對腹部多器官分割提出一個新的深度學習模型，架構中包含了多任務學習、注意力機制、先驗知識導引以及循環神經網路。我們在實驗中驗證了縱向分類任務與器官切割任務的關聯性。透過結合縱向分類的結果以及先驗知識產生導引圖。我們也另外提出了透過反向注意力產生導引圖的方式，結合二元切割果用於凸顯難以被正確分類的特徵。我們將這些導引圖應用於解碼器作為特徵的導引，也經由實驗證明了透過引入導引圖能夠確實地達到權重化特徵的效果並且增進整體效能。最後，我們也引入了循環神經網路用於學習連續影像的時序性表徵，並透過這樣的表徵方式對於前後時間的切割結果進行融合，達到更良好、有效的多器官切割結果。

5.2 未來工作

導引圖的品質對於整體模型的效能是有舉足輕重之影響。因此未來期望透過兩個方式嘗試增進導引圖品質。一、更精細、有效地獲得先驗知識，如我們在章節 3.3 中所描述的，現今先驗知識的產生只透過概略的人體縱向位置判斷，這樣的作法雖然足以有效地導引解碼器進行分割，但是仍缺乏通用性，導引圖品質也有加強空間。我們計畫在未來透過對人體進行更加精細的分類。我們也將避免預先產生靜態的先驗知識，轉而動態地產生先驗知識以增進導引圖的通用性。二、我們也希望在不增加太多運算的情況下增強縱向分類結果的性能，或是透過額外的模組設計、訓練方法減少人體縱向分類結果對於最後切割結果的影響。最後，我們也期望改良時序性訊號的導入以及應用，並探討前後時刻輸入的分割結果作為當下時刻輸入的導引圖或是額外限制的可能性。

參考文獻

- [1] Roth, H.R., Oda, H., Hayashi, Y., Oda, M., Shimizu, N., Fujiwara, M., Misawa, K., Mori, K.: Hierarchical 3D fully convolutional networks for multi-organ segmentation. arXiv preprint arXiv:1704.06382 (2017)
- [2] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz et al., “Attention U-Net: Learning Where to Look for the Pancreas,” arXiv preprint arXiv:1804.03999, 2018.
- [3] Ronneberger, O., P. Fischer, and T. Brox (2015). U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pp. 234–241. Springer.
- [4] H. R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayashi, M. Oda, M. Fujiwara, K. Misawa, K. Mori, "An application of cascaded 3d fully convolutional networks for medical image segmentation", *Computerized Medical Imaging and Graphics*, vol. 66, pp. 90-99, jun 2018.
- [5] Wang, Y., Zhou, Y., Shen, W., Park, S., Fishman, E.K., Yuille, A.L.: Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. CoRR abs/1804.08414 (2018)
- [6] preprint arXiv:1704.06382, 2017. 1, 2 [34] H. R. Roth, C. Shen, H. Oda, T. Sugino, M. Oda, Y. Hayashi, K. Misawa, and K. Mori. A multi-scale pyramid of 3d fully convolutional networks for abdominal multi-organ segmentation. arXiv preprint arXiv:1806.02237, 2018.
- [7] Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: CVPR. (2017)
- [8] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. CoRR, abs/1606.00915, 2016. 1
- [9] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587, 2017.
- [10] P. Hu, G. Wang, X. Kong, J. Kuen, and Y.-P. Tan. Motion-guided cascaded refinement network for video object segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1400–1409, 2018.

- [11] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR, 2015.
- [12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Advances in Neural Information Processing Systems, pages 6000–6010.
- [13] N. Pawlowski, S. I. Ktena, M. C. Lee, B. Kainz, D. Rueckert, B. Glocker, and M. Rajchl. Dltk: State of the art reference implementations for deep learning on medical images. arXiv preprint arXiv:1711.06853, 2017.
- [14] Junyoung Chung, Çağlar Gülçehre, Kyunghyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. CoRR, abs/1412.3555, 2014.
- [15] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. (2016)
- [16] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural computation, 9(8):1735–1780, 1997
- [17] Cai, J., Lu, L., Xing, F., Yang, L.: Pancreas segmentation in CT and MRI images via domain specific network designing and recurrent neural contextual learning. arXiv:1803.11303 (2018)
- [18] H. R. Roth, C. Shen, H. Oda, T. Sugino, M. Oda, Y. Hayashi, K. Misawa, and K. Mori. A multi-scale pyramid of 3d fully convolutional networks for abdominal multi-organ segmentation. In MICCAI, 2018.
- [19] M. P. Heinrich. Multi-organ segmentation using deeds, self- similarity context and joint fusion. 2015.
- [20] Y. Zhou, Z. Li, S. Bai, C. Wang, X. Chen, M. Han, E. Fishman and A. Yuille. Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation. arXiv:1904.06346, 2019
- [21] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang. Residual attention network for image classification. In CVPR, 2017.
- [22] Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Proceedings of the Fourth International Conference on 3D Vision (3DV), IEEE. pp. 565-571.
- [23] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu. Ccnet: Criss-cross attention for semantic segmentation. arXiv preprint arXiv:1811.11721, 2018.

- [24] Q. Jin, Z. Meng, C. Sun, L. Wei, and R. Su, "RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans," no. October, pp. 1–13, 2018.
- [25] Çiçek, Özgün, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. "3D U-Net: learning dense volumetric segmentation from sparse annotation." In International conference on medical image computing and computer-assisted intervention, pp. 424-432. Springer, Cham, 2016.
- [26] B. Landman, Z. Xu, J. E. Igelsias, M. Styner, T. R. Langerak, and A. Klein. (2015). Multi-Atlas Labeling Beyond the Cranial Vault—Workshop and Challenge. Accessed: Jul. 2017. [Online]. Available: <https://www.synapse.org/#!/Synapse:syn3193805>, doi: 10.7303/syn3193805.
- [27] Ibtehaz, N., & Rahman, M. S. (2020). MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121, 74-87.
- [28] Ange Lou , Shuyue Guan, and Murray Loew DC-UNet: Rethinking the U-Net Architecture with Dual Channel Efficient CNN for Medical Images Segmentation. arXiv:2006.00414, 2020
- [29] Vesal, S. , Ravikumar, N. , Maier, A.K. , 2019. A 2d dilated residual u-net for multi-or- gan segmentation in thoracic CT. In: Proceedings of the Challenge on Segmen- tation of THoracic Organs at Risk in CT Images
- [30] L. Folle, S. Vesal, N. Ravikumar, and A. Maier, "Dilated deeply supervised networks for hippocampus segmenta- tion in mri," in *Bildverarbeitung für die Medizin 2019*, 2019, pp. 68–73.
- [31] A. E. Kavur, N. S. Gezer, M. Barı,s, P.-H. Conze, V. Groza, D. D. Pham, S. Chatterjee, P. Ernst, S. Özkan, B. Baydar, et al. Chaos challenge—combined (ct-mr) healthy abdominal organ segmentation. arXiv preprint arXiv:2001.06535, 2020.
- [32] F. Isensee, P. F. Jager, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, " "Automated Design of Deep Learning Methods for Biomedical Image Segmentation," arXiv, 2020
- [33] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR, 2015
- [34] Y. Wang et al., "Deep attentional features for prostate segmentation in ultrasound," in MICCAI, 2018.
- [35] A. Sinha and J. Dolz, "Multi-scale self-guided attention for medical image segmentation," ArXiv190602849 Cs, Feb. 2020.

- [36] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert et al., “nnu-net: Selfadapting framework for u-net-based medical image segmentation,” arXiv preprint arXiv:1809.10486, 2018.
- [37] Cai, J., Lu, L., Xing, F., Yang, L.: Pancreas segmentation in CT and MRI images via domain specific network designing and recurrent neural contextual learning. arXiv:1803.11303 (2018)
- [38] J. Fu et al., “Dual attention network for scene segmentation,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
- [39] Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D.: Attention gated networks: Learning to leverage salient regions in medical images. Medical Image Analysis (2019)

