

二、研究計畫內容：

(一) 摘要

現今心臟核磁共振是檢查心臟疾病的重要指標之一，醫生除了可以藉由影像判斷組織有無病變，亦可從連續跳動的行為模式獲得重要的生理檢驗指標，如各心房心室容積、射血分數（Ejection fraction）等等。但是若想從影像上求出這些參數需要專家花費大量時間標記外框與計算。因此本計畫希望透過深度學習的方式幫助醫生有效率的計算出各時序下的心臟容積，並協助求出各項重要的醫學指標參數，同時期望以此技術做為基石朝不同造影方法的醫學圖像邁進。

(二) 研究動機與研究問題

研究動機

我的外公在中年時被醫生判定為心臟瓣膜脫垂，但是他隨後就以對子女負擔過大為由拒絕深入看診，因為他認為多次就醫會拖累家人生活品質與經濟並造成家人的額外心理與生理負擔。事實上有很多病人抱持著跟我外公一樣的想法，擔心勞煩家人再加上本人認為輕微病情不需深入看診。但心臟疾病對患者而言非常危險，稍有不慎就可能造成無法挽救的傷害。而現今市面上雖有多種智慧手環可供使用者參考心臟健康程度，但更深入的資訊我們卻無從得知。因此醫生還是會建議患者進行各項深入的檢驗如超音波、CT、MRI 等等，超音波雖可以看出心臟的粗淺輪廓但無法判斷微血管的受損程度[1]；CT 雖可看出微血管和鈣化程度但仍無法精確判斷心肌纖維化的程度[1]；不過 MRI 可精確看出心肌病變情形也可看出心臟血液流動狀況。由以上所述可知進行深入檢查 MRI 是一項很好的檢驗方式，因此國際上認為心臟 MRI 是判斷心臟疾病的黃金指標[1]。

臨床上心臟 MRI 有不同成像模式來偵測病灶，其也導致會有不同類型的圖像格式：分別為 Cine imaging、Late gadolinium enhancement (LGE)、First-pass enhancement (FPE) perfusion imaging 等等。Cine 就是我們平常熟悉的影像，主要可由此判斷心臟的活躍程度並計算射血分數[2]；LGE 照影下可以評斷心臟細胞水腫的區域，而水腫意味著心肌已面臨凋亡階段[2]；FPE 可著重於心臟缺血部分，是做心臟微血管重建手術的評估標準之一[2]。

而現今深度學習正熱，國外也有很多學者致力於心臟方面的切割，同時更有許多公開的競賽如 MICCAI、Kaggle 的挑戰等，但大多競賽提供影像多為 cine 圖像格式；其他的如 LGE 等因邊界十分模糊因此鮮少被納入挑戰之列[3][4]。話雖如此，cine 圖像格式在各項挑戰中心尖邊界的正確率卻依然仍不甚理想[5]，因此針對心臟 cine 圖精準定位以及精準邊界切割仍是值得探討的技術問題。該問題若能獲得充分解決，則在臨床研究以及臨床判斷上能替醫生或技師提供更為豐富且有意義的資訊。

研究問題

在本研究計畫中，我們將醫學影像切割常用的深度學習技術 U-Net、可以細微調整深度學習模型的 Selective Kernel Networks (SK-Net)，和 SK-Net 同為注意力模型架構的 Convolution Block Attention Module (CBAM) 等三大影像分割之深度學習技

表 C802

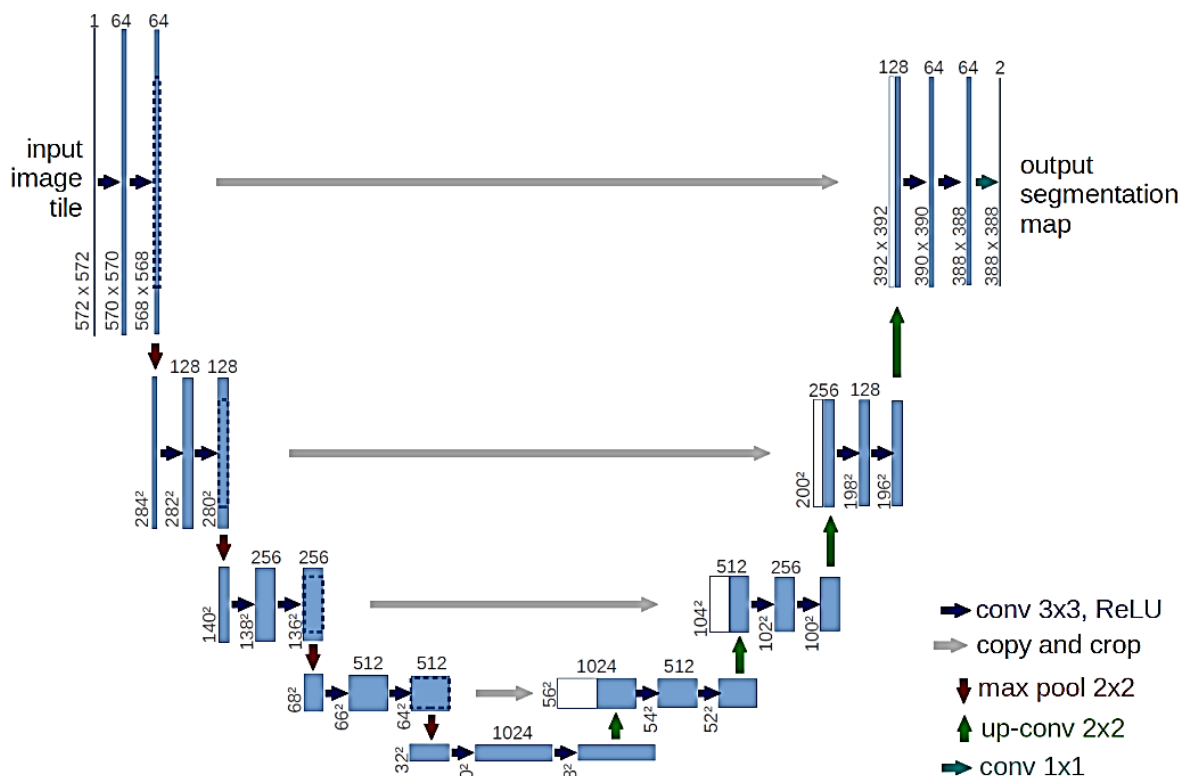
術，系統性的融合成本次計畫所使用的方法 **SKU-Net**。希望擷取各自的優點能在心臟 MRI 的 Cine 圖下更精確算出心房心室容積量、心輸出量與射血分數，作為臨床上判斷心臟疾病嚴重與否的指標，以及為未來的各種特殊照影鋪路。

(三) 文獻回顧與探討

現今學界醫學影像分割主要為 Fully Convolutional Network(FCN)和 U-Net 為技術主流。FCN 無疑是語意分割領域中代表作，除了可以彌補傳統上 CNN 分割的效率不高與預測不準確，同時十分有創意地將 CNN 中的全連接層更改為卷積層以及提出反卷積層的概念以利於圖形的分割[6]。而 U-net 則是在 FCN 的基礎上進行改良的深度學習網路，同時其特點在於可以使用較少的圖形樣本來達成不差的訓練效果，因此相當適合用於醫學影像的訓練與分割[7]。SK-Net 與 CBAM 則分別有利於降低計算複雜度以及捕捉影像的邊緣特徵[8][9]。其各結構將分別如下所述：

1. U-Net：

U-Net 是一種醫學影像分割時常用到的深度學習方式，因其特殊模型架構可運用較少的參數達成和多參數網路一樣的結果，所以可以提升運算效率。而且在資料量少的時候也可保有不錯的分割正確率，因此非常適合運用在資料圖形大但圖形數量極少的醫學影像上。其網路架構如圖一：



圖一、U-Net 網路架構 [7]

(1) conv：卷積(convolution)的簡寫。使用卷積核的目的在於可以強化或減弱某些區域像素的加權，使得目標特徵可以被截取出來。右下 conv 3 × 3 意指捲積核為 3 × 3

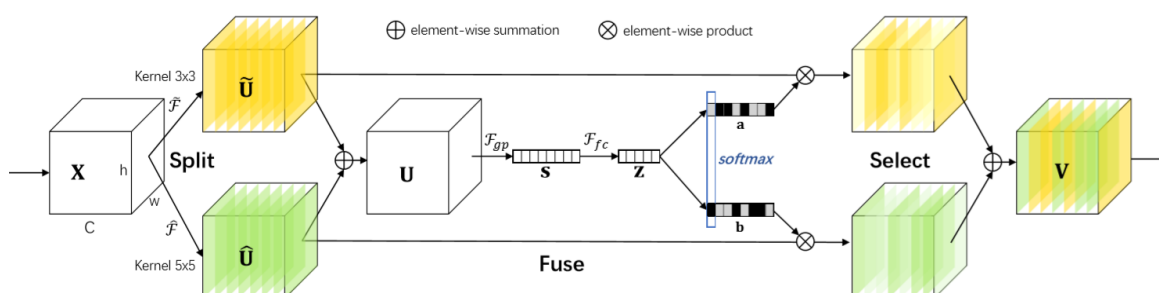
的矩陣。

- (2) relu：一種激活函數，激活函數可以設定卷積過後值的傳送方式。若輸入不大於零通過 relu 函數後輸出為零，大於零則輸入和輸出相同。通常激活函數會加在卷積之後作為輸出。
- (3) Max pool：最大池化。池化可以將資料量降低但同時可保有局部的重要資訊，一般有平均池化法、最大池化、最小池化等方式，U-Net 採用的為最大池化法且 kernel 為 2×2 。
- (4) Up-conv：上採樣的一種方式又稱反卷積(Deconvolution)，目的在於將高維度但小尺寸的特徵圖恢復成一開始原圖輸入的大小，同時也將特徵保留即可得知分類結果。

U-Net 和 FCN 的結構大致相同，但 U-Net 在上採樣後會將同階並壓縮後的特徵圖和上採樣的結果連接在一起。這種讓結果維度增厚的方式和 FCN 融合時使用的相加方式不同，且 FCN 只有在最後做一次融合而不像 U-Net 多次融合。因此就結構穩定性而言 U-Net 會比 FCN 好，因為它同時考慮到淺層的訊息，而 FCN 雖有考慮但淺層訊息融合後訊息有可能消失。但即使 U-Net 已有不錯的輪廓也仍無法處理精細的邊界，結果和人工標記的 ground truth 還有些誤差，因此仍需加入其他元素來改良 U-Net 的結構。

2. Selective Kernel Networks (SK-Net)：

普通的深度神經網路一層只會有一個卷積核去擷取特徵，而 SK-Net 特點即是它在每層可有多個卷積核因此可得出不同的特徵圖。最後多張特徵圖相加得出總特徵圖，因此訊息得以保存下來；再加上各層的捲積核可以透過輸入尺寸大小而調整，所以可以自動調整到最適合的大小得出最優解。其網路架構如圖二：



圖二、SK-Net 原理圖[8]

由圖二可知此網路主要可分為三個計算區塊 split、fuse 和 select：

- (1) split：當資料進入兩個以上卷積核(圖二架構以使用兩個卷積核為例，分別是 5×5 和 3×3 的 kernel)會降維成相應數量個特徵圖，如圖二中所示之 \tilde{U} 和 \hat{U} 。且為了加快運行速度並減少計算量，這裡的 5×5 的矩陣實為空洞卷積核，也就是普通的 3×3 矩陣在間隔中補零使其擴展成 5×5 的矩陣，因此除了保留重要的局部訊息也可以保留部分的全局訊息。
- (2) fuse：將不同的特徵圖相加並做平均池化法得出特徵 S，S 後連接一個全連階層 Z，

其意義是為了讓資料壓縮使其減少維度並提高運算效率。

- (3) select: 為了從壓縮後的 Z 挑選出不同維度特徵，因此使用 soft attention 的方式判斷其不同維度的加權權重。值得強調，attention 是一種可以在眾多訊息中挑選出關鍵訊息而忽略掉無關訊息，就像人只會注意重要訊息而忽略不重要的訊息一樣，而其方法通常由三個步驟組成：

- I. 資料 $X = [X_1, X_2, \dots, X_N]$ 輸入，其代表有 N 個輸入訊息。
- II. 為了降低運算時間，通常會從 N 個訊息中挑選出和分割有關的訊息並計算出注意力分布(近似於機率分布)，一般會使用 soft max 做為激活函數。
- III. 運用注意力分布計算輸入訊息的加權。

Soft attention 即是採用這種程序，但本方法會採納 N 個特徵而不挑選其中幾種，而各維度的注意力分布可由前段求得得 Z 算出。其計算方法如下方程式：

$$a_c = \frac{e^{A_c z}}{e^{A_c z} + e^{B_c z}}, \quad b_c = \frac{e^{B_c z}}{e^{A_c z} + e^{B_c z}}$$

$$V_c = a_c \cdot \tilde{U}_c + b_c \cdot \hat{U}_c, \quad a_c + b_c = 1.$$

其中 A_c 與 B_c 分別為 \tilde{U}_c 與 \hat{U}_c 的注意力分佈，而 z 為前段輸出 Z ，因此可求出 Z 中 \tilde{U}_c 與 \hat{U}_c 各自所佔的注意力分佈，同時也可得到加權的權重。最後 V_c 則為 \tilde{U}_c 與 \hat{U}_c 加權後得到最後的總特徵圖。

由以上方法可知我們經過兩個不同 kernel size 所產生出的特徵圖可以經由 fuse 和 select 的方式算出不同尺度在同一通道下各自的權重，因此可以適當的挑選出最佳的線性組合，解出最優的邊界。

3. Convolution Block Attention Module(CBAM):

CBAM 和 SK-Net 一樣都是由注意力模型演化來而來的的方法，不過兩者注重的技術發展則不太一致。SK-Net 強調的同一通道的不同特徵融合，CBAM 則是較類比於注意力模型的始祖 SE-Net 一樣強調的是各個通道加權的權重，而其和 SE-Net 除了各通道加權的權重之外又更加强空間分布，因此可以避免像 SE-Net 將所有通道全部擠在同個模塊而損失空間訊息的狀況產生。CBAM 的主要原理架構圖如下頁的圖三與圖四所示。如圖四所描述，我們將假設輸入圖像 $F \in R^{C \times H \times W}$ ，其中 H 與 W 為輸入圖像的尺寸而 C 為通道數。在此架構下，CBAM 會將其通過一個一維的通道架構和二維的空間架構得到結果，而其中圖四內所描述各類模組意義則如下所示：

(1) Channel Attention Module (通道注意模組):

如圖四上所描述，通道注意模組利用圖形內部的特徵關係取得注意力分佈。為了減少運算性 CBAM 將特徵圖壓縮至一維，並運 max pool 和 average pool 來獲取訊息。以往專家都認為 average pool 才能有效的擷取物體訊息，但 CBAM 的作者卻認為 max pool 可以將重點擷取出來，因此他將兩者皆運用，實驗也證實用此方法比起單一使用更有效。

其運算經過 pooling 後會得到 F_{avg}^c 和 F_{max}^c 兩種不同空間分布的特徵圖，爾後這兩圖會進入由多層感知器(multi-layer perceptron, MLP)和一個隱藏層所組成的網路並產生其通道注意力分佈。為了減少參數設定，隱藏層的 sigmoid 激活函數將設定為 $R^{C/r \times 1 \times 1}$ ，其中 r 為壓縮率，經過激活函數後各通道相加即為通道注意

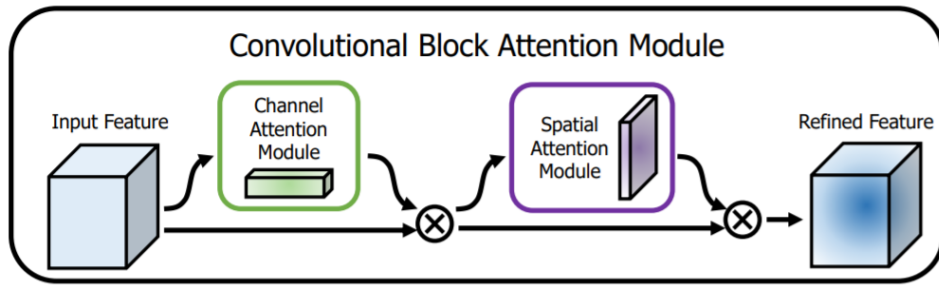
力分佈。其通道注意力圖的公式如下所示，其中 σ 為 sigmoid 函數、 $W_0 \in R^{C/r \times C}$ 和 $W_1 \in R^{C \times C/r}$ 皆為 MLP 的權重：

$$\begin{aligned} \mathbf{M}_c(\mathbf{F}) &= \sigma \left(\text{MLP}(\text{AvgPool}(\mathbf{F})) + \text{MaxPool}(\mathbf{F}) \right) \\ &= \sigma \left(\mathbf{W}_1 \left(\mathbf{W}_0(\mathbf{F}_{avg}^c) \right) + \mathbf{W}_1(\mathbf{W}_0(\mathbf{F}_{max}^c)) \right) \end{aligned}$$

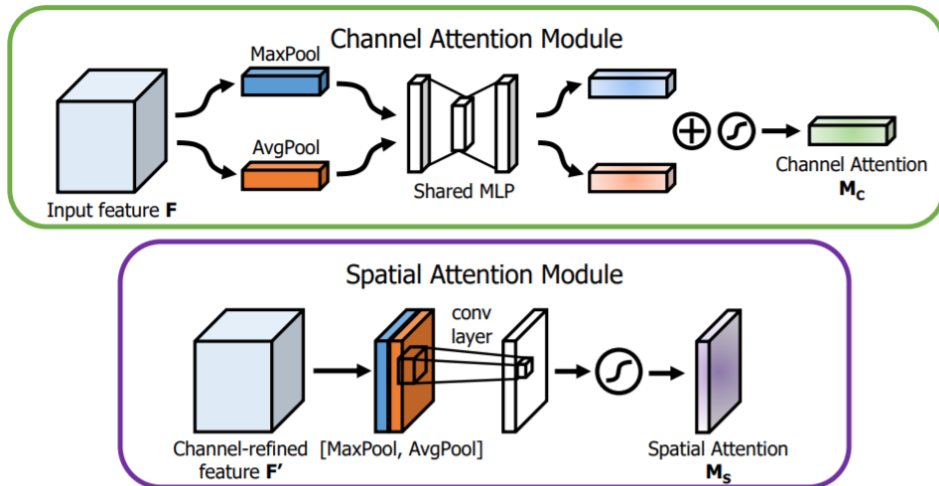
(2) Spatial attention module (空間注意力模組)

空間注意力模組則是用於尋找特徵間的空間關係。但本模組和上述的通道注意模組最大的差別就是其強調訊息位於何處而非擷取訊息，以做為彌補通道模型的不足。為了做出空間模型，通道模型的輸出會先經過 average pool 和 max pool 來標註空間中強調的區域；同時為連接兩者訊息，他們會經過卷積層和 sigmoid 函數求出最後的輸出，其運算方程式如下所示，其中 $f^{7 \times 7}$ 表示其為 7×7 kernel：

$$\mathbf{M}_s(\mathbf{F}) = \sigma(f^{7 \times 7}([\text{AvgPool}(\mathbf{F}); \text{MaxPool}(\mathbf{F})]))$$



圖三、CBAM 主架構圖[9]

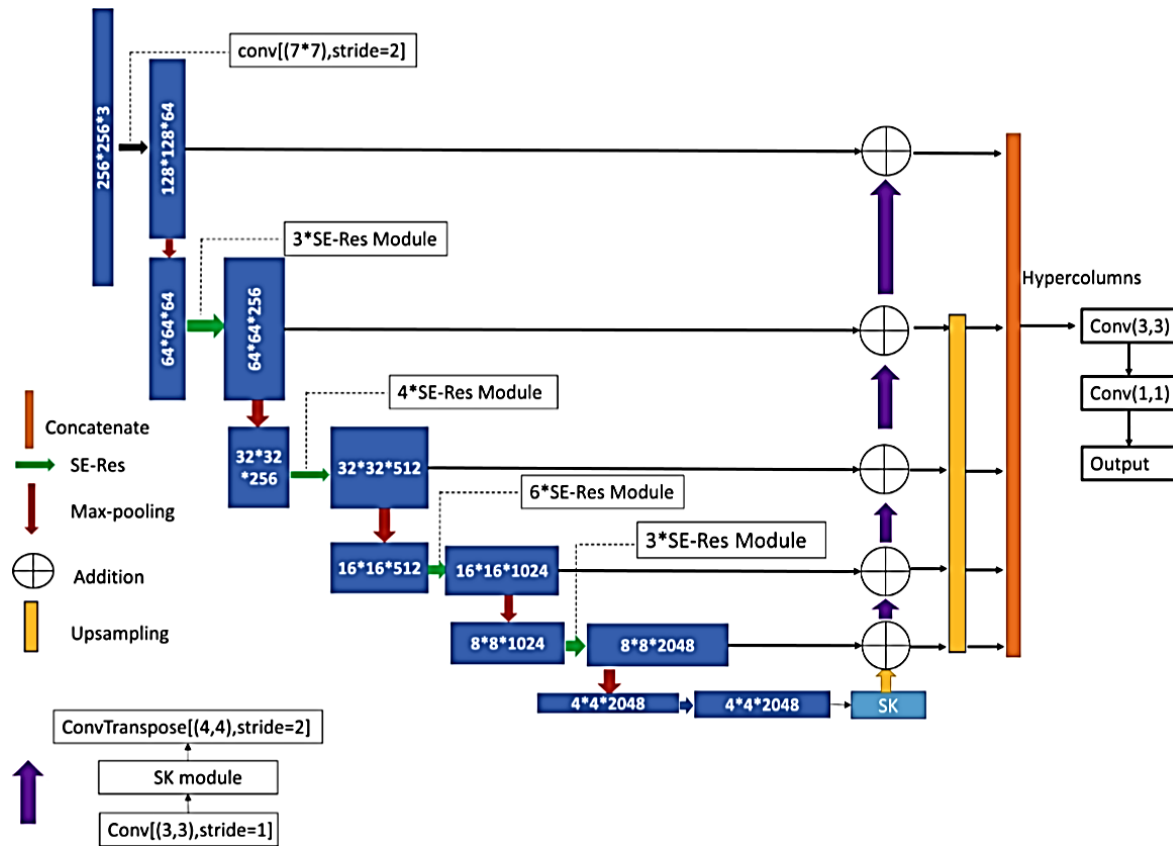


圖四、CBAM Channel 和 spatial 的架構圖[9]

4. 其他研究團隊運用 SKU-Net 在心臟分割挑戰的結果：

在 MICCAI 2019 挑戰時亦有研究團隊提出使用 SKU-Net[3]的深度學習方法。雖然該挑戰主要任務是在切割心臟的 LGE 圖，但由於 LGE 圖的邊界比起普通的心臟表 C802

cine 圖更模糊不清，因此該參賽之研究團隊在切割邊緣的方法將有益於本計畫中的分割模型之發展，而其結構如下頁之圖五所示：



圖五、參賽隊伍所使用的 SKU-Net[3]

研究團隊最經典之舉在於大膽的將全連接層加入以往為全捲積層的 U-Net 裡，同時運用通道權重融合的 SE-Net 和計算各通道內各特徵權重的 SK-Net 組合而成他們所提出之方法。令人驚豔的是該方法在邊界非常不明顯的圖上有很好的切割結果，因此在本研究計畫中將納入此架構做為邊緣細節的切割並加入自己的元素。

(四) 研究方法及步驟

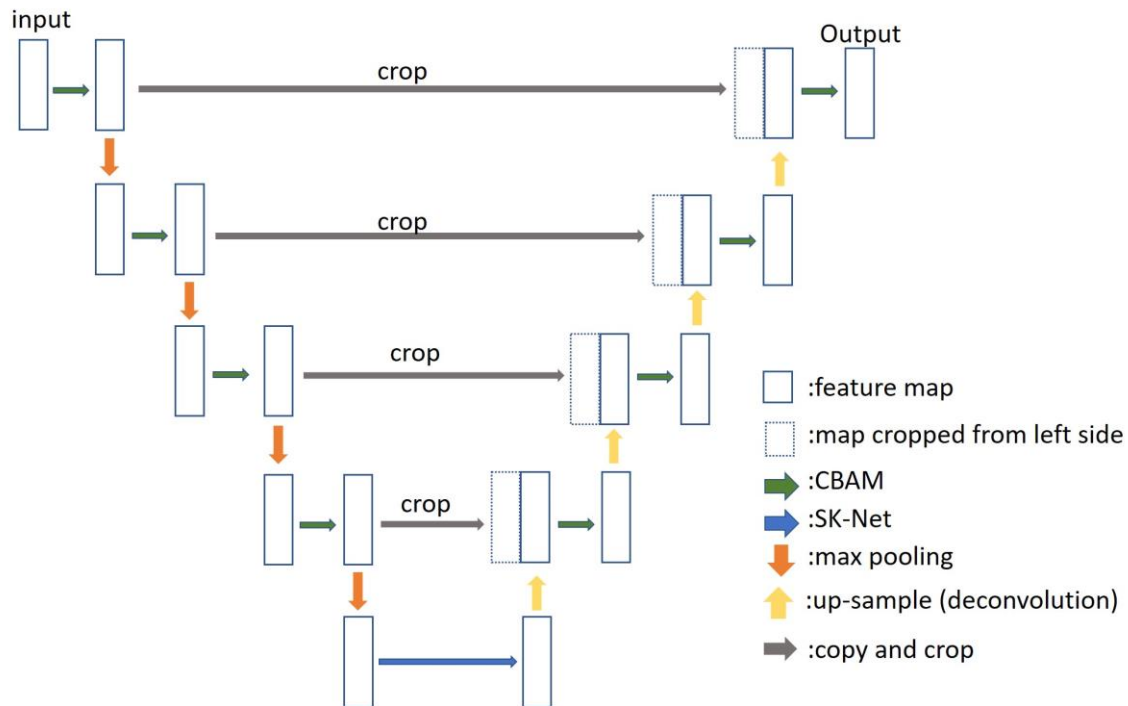
Dataset 與硬體

本研究計畫中將採用 Kaggle2016[10]競賽的心臟 cine MRI 圖資料集。該資料集中每張影像大小為 192pixel*256pixel，而在每位病人的一個心動週期中擷取 30 張影像。此資料集提供共有訓練組 500 人、測試組 440 人，以及驗證組 200 人。在本研究計畫中，我們使用的伺服器硬體效能為 Intel i9-9980XE 18 核 CPU(3GHz)搭載兩張 GPU (NVIDIA GeForce RTX 2080Ti)並有 128 GB 記憶體，因此相當適合在 python 的 TensorFlow 或是 PyTorch 框架下作為深度學習參數訓練以及驗證之用。

研究方法

本研究方法將上述 MICCAI 2019[11]參賽之研究團隊所提出之 SKU-Net 深度學習表 C802

方法加以修正，其中我們將捨棄 SE-Net 改為更全面的 CBAM 且保留 U-Net 全捲積層的特性，進而實現本研究計畫所提 SKU-Net 之特性。其結構如下頁之圖六所示：



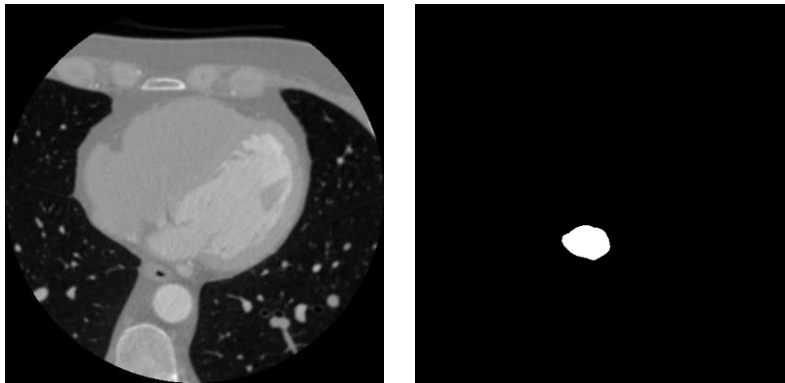
圖六、本計畫使用的 SKU-Net 架構圖

其中，在本架構採用之 SK-Net 和 CBAM 的架構在文獻中已提過，故在本架構中不再重複說明。

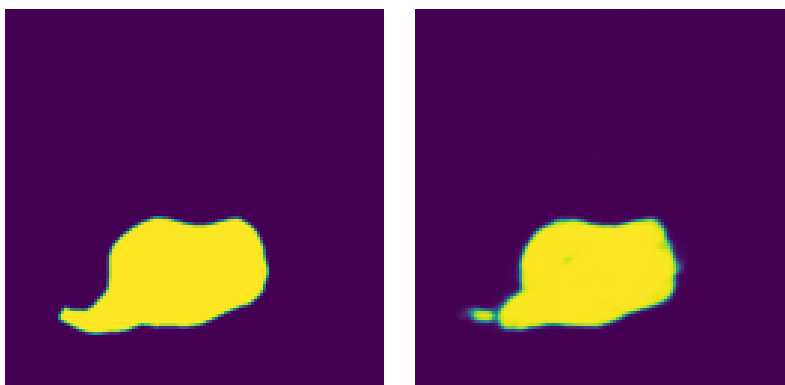
先期研究成果

由於 Kaggle 競賽提供的資料集中的 ground truth 並非是傳統的遮罩圖(mask)，因此在前期研究成果方面我們已先從有遮罩圖的 stacom 2013[13]年左心房切割挑戰開始。該挑戰所提供之左心房影像資料集共有 1895 組的原本影像及 ground truth(其影像大小皆為 512×512)。圖七(左)所示為其中一頁之左心房圖像，以及經由相關醫學專家們鑑定過之 ground truth，如圖七(右)所示。

在先期研究成果的演算法實踐上，我們以 PyTorch 為框架並以 stacom 2013 為資料集嘗試進行傳統 U-Net，其結果如圖八所示。由圖八(右)的分割結果可知用傳統的 U-Net 方法所得的組織邊界精細度仍不夠完美，從圖中可以明顯發現與 ground truth 相比在左下角之處產生分割不完全的地方。這些分割不完全的影像將導致做影像的三維結構化或是手術導引時，容易對細緻處或是轉折處產生解剖結構上的誤解而造成判斷失誤。因此本研究期待能獲得計畫並在我們提出的 SKU-Net 深度學習架構中望達到更好的細緻分割效能。



圖七、(左)cmr 下的心臟 cine 圖，(右)左心房的 ground truth 圖[13]



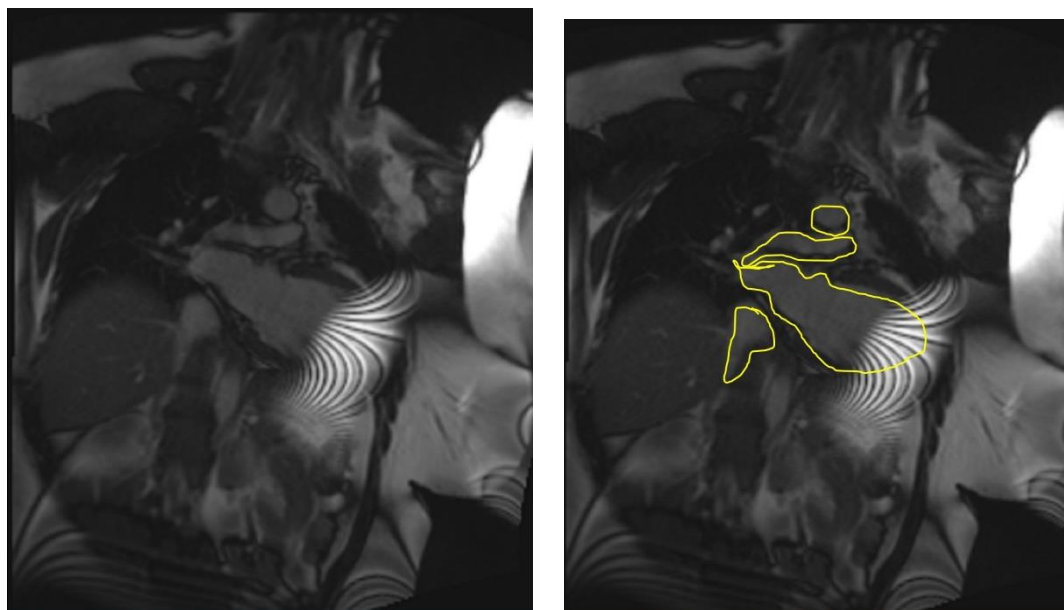
圖八、(左)原本資料集提供的 ground truth、(右)以傳統 U-Net 切割的結果

未來規劃

項目	109 年						110 年	
	7 月	8 月	9 月	10 月	11 月	12 月	1 月	2 月
論文文獻探討								
確認研究方法								
嘗試其餘注意力模型的演算法								
改善模型架構								
精進資料處理方式								
與指導教授討論								

（五）預期結果

如圖九(左)為 Kaggle 的 test 中的一組資料，其資料集提供的為心臟的收縮與舒張容積。未來會以本計畫提出的 SKU-net 切割出如圖九(右)的圖：



圖九、(左)Kaggle 原圖，(右) SK-Unet 預期分割結果圖

未來目標

如圖八所示的影像分割結果，目前為止雖已抓到輪廓但細節部分仍不甚理想，後續會加以修正參數與架構令輪廓更接近實際情況，讓切割結果更加完美，才有機會挑戰到程度更高的 LGE 問題分割。

其他的 Kaggle 競賽隊伍[12]也有提其他更高深的技巧如來切割心臟圖，但實際上亦有研究團隊使用傳統的 U-Net 就能達到和前者一樣的結果。這證實了計算複雜度更高的演算法不一定會得到比較好的結果，重點是在於圖形的表徵比較適合用哪種演算法。但在這方面仍是我目前所需要學習的方向，希望未來能對資料的敏感度更高，研究出最適合心臟的切割方式。

（六）參考文獻

- [1] 心臟 MRI 為何是檢查心臟疾病的金指標?
<https://www1.cgmh.org.tw/intr/intr2/c33d00/characteristic.asp?La=pno=4&pno=9>
- [2] W .Chan, A .H .Ellims, S. J. Duffy, D. M. Kaye, A. J. Taylor, Principles, current status and clinical implications of ischaemic heart disease assessment by cardiac magnetic resonance imaging(2011)
<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1445-5994.2011.02606.x>
- [3] Xyiue Wang, Sen Yang, Mingxuan Tang, Yunpeng Wei, Xiao Han, Ling He, Jing Zhang, SK-Unet: An Improved U-Net Model with Selective Kernel for the Segmentation of Multi-sequence Cardiac MR(2020)

- https://link.springer.com/chapter/10.1007/978-3-030-39074-7_26
- [4] Tim Leiner, Daniel Rueckert, Avan Suinesiaputra, Bettina Baeßler, Reza Nezafat, Ivana Išgum, Alistair A. Young, Machine learning in cardiovascular magnetic resonance: basic concepts and applications(2019)
<https://link.springer.com/article/10.1186/s12968-019-0575-y>
- [5] Tuan Anh Ngo, Zhi Lu, Gustavo Carneiro, Combining Deep Learning and Level Set for the Automated Segmentation of the Left Ventricle of the Heart from Cardiac Cine Magnetic Resonance(2017)
<https://www.sciencedirect.com/science/article/abs/pii/S136184151630038X>
- [6] Jonathan Long, Evan Shelhamer, Trevor Darrell, Fully Convolutional Networks for Semantic Segmentation(2015)
<https://arxiv.org/abs/1411.4038>
- [7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation(2015)
https://link.springer.com/content/pdf/10.1007%2F978-3-319-24574-4_28.pdf
- [8] Xiang Li, Wenhai Wang, Xiaolin Hu, Jian Yang, Selective Kernel Networks (2019)
http://openaccess.thecvf.com/content_CVPR_2019/html/Li_Selective_Kernel_Networs_CVPR_2019_paper.html
- [9] Sanghyun Woo, Jongchan Park, Joon-Young Lee, In So Kweon, CBAM: Convolutional Block Attention Module(2018)
https://link.springer.com/chapter/10.1007/978-3-030-01234-2_1
- [10] Kaggle Challenge
<https://www.kaggle.com/c/second-annual-data-science-bowl>
- [11] MICCAI Challenge
<http://www.miccai.org/events/challenges/>
- [12] Multi-sequence Cardiac MR Segmentation Challenge (MS-CMRSeg 2019)
<https://zmiclab.github.io/mscmrseg19/>
- [13] stacom 2013 dataset
<http://www.cardiacatlas.org/challenges/left-atrium-segmentation-challenge/>

(七) 需要指導教授指導內容

1. 需要從學理以及技術層面上指導如何融合多種各類型的深度學習模型架構方法之優勢，並在以 python 為基礎的架構下實踐模型視覺化。
2. 需要從技術上指導如何精簡化演算法、降低演算法的計算複雜度，並能更快速的除去冗長的資料儲存與處理。
3. 需要從學理上糾正學生學習多種 attention 模型的謬誤，加強各種模型的機率統計模型或是 Layer 之間的關聯性。
4. 需要從臨床的角度上指導圖形分割的準確性，以及如何與醫生專家溝通並將結果實際貢獻在醫學服務上。
5. 需要指導論文寫作、文章發表之能力。