

淡江大學電機工程學系

指導教授：江正雄博士

基於卡爾曼濾波的多解析度特徵融合於
三維腎臟腫瘤分割

**Kalman Filter-Based Multi-Resolution
Feature Fusion for 3D Kidney Tumor
Segmentation**

學生：詹盛惟
潘信瑞
沈群博

中華民國 115 年 10 月

目錄

摘要	1
Abstract.....	2
第一章 緒論	3
1.1 背景介紹	3
1.2 實驗動機	3
1.3 研究問題與評估安排	4
1.4 章節安排	4
第二章 相關文獻	5
2.1 U-Net 與 nnU-Net	5
2.2 下採樣方法的改進	5
2.3 狀態空間模型與 Mamba	6
2.4 Mamba 於醫學影像分割之應用	6
2.5 卡爾曼濾波及其與深度學習的結合	7
2.6 KiTS 資料集與既有方法	7
2.7 本作品與既有架構的關係	8
2.8 稀疏卷積與區域式分割	8
2.9 注意力與預訓練分割模型	9
2.10 卡爾曼啟發式分割方法	9
2.11 三維分割方法的比較依據	10
2.12 融合增益與門控機制的關係	10
第三章 作品介紹	12
3.1 整體架構與符號定義	12
3.2 標準下採樣的限制	14
3.3 卡爾曼下採樣	14
3.4 不確定度代理量的跨層遞迴	16
3.5 受後向平滑啟發的解碼器融合	16
3.6 複雜度分析	17
3.7 張量尺寸與模組介面	17
3.8 融合公式的代數性質	18
3.9 初始設定與局部更新範例	19

3.10 解析度配置與實體視野	20
3.11 共享增益與通道殘差的分工	20
3.12 融合權重與誤差變異數的區別	21
3.13 代理量儲存與整體顯存需求	22
3.14 空間對齊對代理量傳遞的影響	22
第四章 成果展示	24
4.1 資料集	24
4.2 評估指標	24
4.3 實驗環境與訓練設定	25
4.4 設計歷程與消融實驗	26
4.5 5-fold 交叉驗證結果	27
4.6 計算成本比較	28
4.7 與既有方法之比較	28
4.8 分割結果視覺化	30
4.9 限制與討論	31
4.10 既有結果的整合分析	31
4.11 訓練設定與計算資源比較	33
4.12 相近卡爾曼方法之設計對照	36
4.13 文獻內部對照與研究定位	37
4.14 標籤混淆與階層評估的關係	37
4.15 推論流程與成本比較的範圍	38
4.16 訓練預算與比較單位	39
第五章 結論與未來發展	41
5.1 結論	41
5.2 未來發展	41
參考文獻	42

圖目錄

圖 1. 基準模型 UMambaBot 架構	12
圖 2. KalmanDown 編碼器雙分支下採樣與各層獨立的 P	13
圖 3. KalMed-UNet 跨層 P 遞迴與解碼器後向融合	14
圖 4. 卡爾曼下採樣模組之內部資料流	15
圖 5. case_00268 分割結果比較（第 79 張軸狀切片）	30
圖 6. case_00555 分割結果比較（第 185 張軸狀切片）	30
圖 7. case_00551 分割結果比較（第 34 張軸狀切片，本作品小幅領先）	30
圖 8. case_00472 分割結果比較（第 66 張軸狀切片，本作品小幅領先）	31
圖 9. 五折整體 Dice 與配對增益（依表 6 繪製）	32
圖 10. 設計版本的效能與計算量（依表 7 繪製）	33

表目錄

表 1. 卡爾曼濾波要素與網路組件之對應	15
表 2. KiTS23 資料集概要	24
表 3. 實驗環境與訓練設定	25
表 4. 設計歷程與消融實驗結果 (250 epochs、單一 fold, 數值為 Dice)	26
表 5. 5-fold 交叉驗證之三類 Dice (1000 epochs, 平均 $\pm$ 標準差)	27
表 6. 各 fold 之整體 Dice 與配對差值 (1000 epochs)	28
表 7. 參數量與計算量比較 (Dice 為 250 epochs、單一 fold 之結果)	28
表 8. KiTS23 公開集內驗證之文獻結果	29
表 9. KiTS23 獨立保留集之文獻結果 (89 例)	29
表 10. 公開集內驗證方法之訓練設定	33
表 11. 獨立保留集方法之共同設定與架構差異	34
表 12. 外部方法與本作品之參數量及硬體資源	34
表 13. 卡爾曼啟發式分割方法之設計與實驗條件	36
表 14. 標籤轉換對三個階層評估集合的影響	38

摘要

三維腎臟腫瘤分割在術前規劃與療效評估上具有臨床價值，然而腫瘤形態與體積差異大，部分病灶邊界模糊且與周圍組織對比度低，自動分割仍具挑戰性。以 U-Net 為基礎的架構普遍透過步幅卷積或池化降低特徵解析度。此過程可能損失局部細節，因此本作品探討在下採樣階段引入自適應特徵融合的效果。本作品提出 KalMed-UNet，以 UMambaBot 為基準模型，將受卡爾曼濾波啟發的預測—更新結構引入編碼器下採樣階段。模組結合步幅卷積與池化投影兩條分支，利用殘差與不確定度代理量 P 決定融合增益，並將 P 跨解析度傳遞至解碼器，進行受後向平滑啟發的特徵融合。此設計以 P 作為網路內部的遞迴輔助量，串接編碼器與解碼器的融合過程，並以分割實驗評估整體架構的效益。本作品於 KiTS23 公開訓練集進行 1000 epochs、5-fold 交叉驗證，整體 Dice 由基準模型的 0.8700 提升至 0.8813，其中腎臟區域由 0.9661 提升至 0.9807、腫塊由 0.8378 提升至 0.8467、腫瘤由 0.8059 提升至 0.8165，五個 fold 全數改善；參數量與計算量則分別減少 1.31% 與 4.48%。另以 250 epochs、單一 fold 的設計比較呈現模組演進，並整理既有方法的分數、訓練條件及硬體資源，交代本作品在三維分割研究中的位置。相同資料切分與訓練流程下的基準對照顯示，所提融合設計可在既定模型規模與推論流程內，同時改善區域重疊表現及架構成本。

關鍵字：醫學影像分割、卡爾曼濾波、狀態空間模型、Mamba、腎臟腫瘤、KiTS23

Abstract

Accurate three-dimensional segmentation of kidneys and renal tumors is important for quantitative image analysis. Variations in lesion size, shape, and contrast make the task challenging, while repeated downsampling can limit the representation of local structures. This study proposes KalMed-UNet, a multi-resolution feature fusion architecture built on UMambaBot. A Kalman-inspired prediction–update module combines strided convolution with a max-pooling and projection branch at two early encoder levels. The fusion gain depends on the branch residual and a nonnegative auxiliary state, P . The normalized residual is reduced using the mean absolute value across channels before spatial standardization. The auxiliary state is propagated between resolutions and reused in the corresponding decoder levels for fusion inspired by backward smoothing. Here, P serves as a feature-fusion proxy rather than a calibrated error covariance. Experiments were conducted on the 489 publicly available KiTS23 training cases using five-fold cross-validation and 1,000 training epochs per fold. Compared with UMambaBot under the same data splits and training procedure, KalMed-UNet increased the overall Dice score from 0.8700 to 0.8813. Dice scores for the kidney region, kidney masses, and tumors increased from 0.9661, 0.8378, and 0.8059 to 0.9807, 0.8467, and 0.8165, respectively. Improvements were observed in all five folds, while parameter count and computational cost decreased by 1.31% and 4.48%. Design variants were additionally examined under a reduced budget of 250 epochs on one fold. Published methods are reviewed with their evaluation protocols, training settings, and computational resources to contextualize the results. The controlled baseline comparison supports the effectiveness of the integrated fusion architecture within the evaluated setting, without test-time augmentation, model ensembling, or post-processing.

Keywords: medical image segmentation; Kalman-inspired fusion; Mamba; kidney tumor; KiTS23

第一章 緒論

1.1 背景介紹

腎細胞癌為泌尿系統中常見的惡性腫瘤，臨床上多以對比劑增強電腦斷層（contrast-enhanced CT）影像作為診斷、分期與術前規劃的依據。醫師需在影像中判讀腎臟本體、腫瘤與囊腫的範圍，以評估腫瘤大小、位置及其與周邊組織的關係，作為決定冷凍消融、部分腎切除或根治性腎切除的參考。然而三維 CT 影像的切片數量龐大，單一病例往往包含數百張切片，人工逐層描繪輪廓相當耗時，且不同判讀者之間存在主觀差異。此一標註瓶頸限制了放射組學（radiomics）等量化分析在臨床上的推展，因此自動化的三維分割技術具有實際價值。U-Net 透過降低特徵解析度取得較大範圍的影像脈絡，再由解碼器恢復空間資訊。nnU-Net 將前處理、網路配置與訓練納入自動化流程，U-Mamba 則以狀態空間模型補充長距離依賴建模[1][4][13]。深層表徵能力與解析度轉換共同影響輸出；即使瓶頸層能整合廣泛資訊，仍需適當的特徵傳遞將其連結至高解析度位置。本作品由此切入，在維持既有分割流程下，研究下採樣與解碼器融合的局部改動。

1.2 實驗動機

本作品的基準編碼器以步幅卷積進行下採樣，將 128^3 的輸入特徵逐層降至 4^3 ，各軸解析度共縮減 32 倍。共享卷積核仍會依局部輸入產生不同輸出，但此操作沒有顯式結合另一條下採樣分支，也沒有維持可供後續融合使用的輔助狀態。對三維病灶而言，降低解析度雖能擴大感受範圍，卻也使細小結構的表示更依賴通道編碼及 skip connection 的補充。因此，本作品將下採樣視為可改善的環節之一，以步幅卷積與池化投影形成兩種同源特徵，再依兩分支差異決定修正幅度。兩條分支均由同一輸入產生，設計目的在於建立互補的特徵轉換，而非模擬彼此獨立的感測器。卡爾曼濾波的預測—更新結構提供了組織此類融合的方式：先建立預測，再利用觀測差值修正，並維持可遞迴傳遞的輔助量。本作品據此提出雙分支下採樣、跨解析度的不確定度代理量 P ，以及對應解碼器的特徵融合，將改動集中於前兩個降解析度層級。方法的價值由相同設定下的分割結果與成本共同評估，而非直接套用線性高斯模型的最適性。研究完成 KiTS23 五折基準對照，並以縮減預算的設計比較說明架構演進；最終模型在五個 fold 均取得正向增益，同時降低參數量及

計算量。這些結果構成本作品的主要貢獻，也使研究重點落在可實作、可追溯的融合設計及其整體效益。

1.3 研究問題與評估安排

本作品主要評估三個相互關聯的問題：雙分支下採樣能否補充分割所需的特徵，跨層 P 與解碼器融合的整合是否有助於效能與成本取捨，以及最終模型的增益能否在不同資料切分下維持。步幅卷積以可學習權重組合鄰域資訊，最大池化投影則先選取局部響應再調整通道，KalmanDown 與基準的對照用於觀察這項下採樣改動。完整 KalMed-UNet 另加入跨層遞迴與解碼器上採樣／融合，因此該版本的差值代表整合配置的效果。研究同時報告三類 Dice、參數量及 GFLOPs，使準確度的變化能與其架構成本一併理解。評估分為設計探索與最終驗證兩個層次。250 epochs、單一 fold 的比較用於整理六個版本在相同縮減預算下的結果；1000 epochs、五折交叉驗證則作為主要效能證據。基準與完整模型採相同五組切分，逐折列出配對差值，並以三個階層評估類別檢視平均提升的組成。外部文獻另依資料切分、訓練方式與推論流程整理，提供方法定位及資源需求的參考。如此可分別回答「本作品的改動在共同設定下是否有效」與「既有研究採用哪些方式處理相同任務」，避免將跨研究的分數差距直接歸因於網路模組。

1.4 章節安排

第二章回顧相關方法並說明研究定位。第三章介紹架構、張量介面、融合公式及成本來源。第四章呈現實驗設定、設計歷程、五折結果與視覺案例，並整理外部方法的分數、訓練及硬體條件。第五章歸納成果與延伸方向。

第二章 相關文獻

2.1 U-Net 與 nnU-Net

U-Net[1]於 2015 年提出，採用對稱的編碼器—解碼器結構，並以 skip connection 將編碼器各層級的特徵直接串接至解碼器對應層級，使解碼過程能同時利用深層的語意資訊與淺層的空間細節。此設計在標註資料有限的醫學影像領域表現突出，後續 3D U-Net[2]與 V-Net[3]分別將其擴展至三維體積資料。Skip connection 可將較高解析度的編碼器特徵傳至解碼器，以補充深層表示的空間細節。本作品在保留此結構的基礎上，進一步探討下採樣分支融合及跨層輔助量的作用。nnU-Net[4]的核心在於依據資料集特性安排前處理、網路配置與訓練策略，包括重取樣間距、patch size、batch size、網路深度及排程，而非只提出固定的一個模型。這使同樣稱為 nnU-Net 的實驗，仍可能因資料規劃與資源條件而具有不同尺寸及計算需求。因此，本文沿用 nnU-Net 作為底層框架，並以共同前處理、資料切分及訓練流程建立 UMambaBot 與 KalMed-UNet 的直接對照。外部方法若僅報告其使用 nnU-Net，仍需進一步確認實際配置，才能判斷分數及參數量所對應的實驗條件。

2.2 下採樣方法的改進

下採樣研究包含反混疊濾波、取樣格點選擇及可學習步幅等方向。Zhang[5]提出 BlurPool，在降取樣前加入低通濾波以改善位移穩定性；Azulay 與 Weiss[6]分析深度網路對微小影像變換的敏感性，Zou 等人[7]則依空間位置及通道群組預測濾波權重。另一類方法調整取樣規則，Chaman 與 Dokmanić[8]的自適應多相取樣依分量範數選擇格點，Rojas-Gomez 等人[9]使此選擇可端到端學習，Riad 等人[10]的 DiffStride 則在傅立葉域學習裁切遮罩尺寸。這些研究顯示，內容自適應及可學習取樣已是既有方向，因此本作品的新穎性應由具體的融合結構說明，而非概括為首次依位置調整下採樣。本作品的切入點由三項相連的設計構成。首先，以步幅卷積形成預測分支，以最大池化及 $1 \times 1 \times 1$ 卷積形成觀測分支，在相同解析度與通道空間計算殘差。其次，殘差依序經 InstanceNorm、通道平均絕對值與空間標準化，形成局部調變項，再與 $P^-/(P^- + 1)$ 相乘，以決定各位置的融合比例。最後，完整模型將 P 從 stage 1 傳至 stage 2，並提供給對應解碼器，使前向融合及

後向修正使用相連的輔助狀態。此設計關注的是不同轉換之間如何融合，以及融合狀態如何跨解析度利用；與改變取樣格點或低通濾波器的研究具有不同的操作重點。P 在本文中作為融合代理量，其效果由分割實驗評估。

2.3 狀態空間模型與 Mamba

狀態空間模型 (State Space Model, SSM) 以線性系統描述序列，經離散化後可作為神經網路層使用。S4[11]藉由結構化的參數設計解決了長序列訓練的數值問題，使 SSM 得以有效建模長距離依賴。Mamba[12]在此基礎上提出選擇性掃描機制，使狀態轉移參數隨輸入內容而變化，並搭配硬體感知的平行掃描演算法，在保持線性複雜度的同時取得與 Transformer 相當的建模能力。相較於自注意力機制隨序列長度呈二次成長的計算量，此一特性對三維醫學影像特別有利：一個 128^3 的體積展平後即為超過兩百萬個元素的序列。在本作品的配置中，Mamba 位於 4^3 的瓶頸特徵上，並非直接對完整 128^3 體積執行序列掃描，因此不能以原始體素數推算其實際序列長度。瓶頸層的序列建模與前段的解析度轉換分別處理不同環節：前者組織深層特徵之間的關係，後者決定高解析度表示如何被組合後傳向深層。本作品保留瓶頸層 Mamba，將主要設計變更放在前兩次下採樣及相對應的解碼器融合，使實驗能在共同深層建模方式下，觀察這些局部改動的整體影響。

2.4 Mamba 於醫學影像分割之應用

U-Mamba[13] 將 Mamba 整合至 nnU-Net，提出僅於瓶頸層使用的 UMambaBot，以及在編碼器多個層級使用的 UMambaEnc。SegMamba[14]針對三維醫學影像設計長距離序列建模架構，Swin-UMamba[15]研究預訓練與 U 形架構的結合，VM-UNet[16]則以視覺狀態空間模組建立分割網路。這些研究提供了不同的放置位置與特徵建模策略，也說明「使用 Mamba」並不代表所有模型具有相同的資料流、計算成本及訓練條件。比較時仍須回到模組所在的層級，以及它與卷積和 skip connection 的連接方式。本作品選用 UMambaBot 作為基準，因其將 Mamba 集中在瓶頸層，便於在固定深層序列建模方式的條件下研究下採樣改動。KalmanDown 先調整編碼器，完整 KalMed-UNet 再串接 P 的跨層傳遞與解碼器融合。如此安排使方法描述能逐步對照共同骨幹，也使成本變化可以連結至新增觀測投影及替換上採樣等操作。最終模型與基準的差距反映整合設計的效益；本研究

並未將所有 Mamba 變體視為同一基準，亦不由單一骨幹的結果推論其他配置必然取得相同增益。

2.5 卡爾曼濾波及其與深度學習的結合

卡爾曼濾波[17]於 1960 年提出，在線性高斯系統的假設下提供最小均方誤差意義上的最佳狀態估計。其運作由預測與更新兩個步驟遞迴構成：預測步驟依系統動態推估下一時刻的狀態與其協方差；更新步驟則在取得新觀測後，以卡爾曼增益加權地修正該預測。卡爾曼增益的形式使其具備自適應特性：當先驗協方差較大、亦即對預測的信心較低時，增益提高，估計結果較倚重觀測；反之當觀測雜訊較大時，增益降低，估計結果較倚重預測。此一由不確定度自動決定權重的性質，正是本作品欲引入下採樣過程的核心。在濾波之外，Rauch、Tung 與 Striebel 提出的後向平滑（RTS 平滑）[18]可在取得全部觀測後回頭修正較早的估計，本作品的解碼器設計即受其啟發。深度學習與濾波結構的結合可有不同目的。KalmanNet[19]利用神經網路學習增益，以處理系統模型部分已知的動態估計；醫學影像中的卡爾曼啟發式方法，亦可把遞迴結構用於特徵或取樣位置的調整。本作品將索引由時間改為解析度層級，借用預測、殘差修正及輔助狀態傳遞的組織方式。兩分支皆由同一輸入轉換而來，且 P 未建立與真實特徵誤差一致的機率模型，因此本文使用「卡爾曼啟發式融合」描述其性質。此定位保留理論來源與實作之間的清楚對應，也讓實驗結論聚焦於網路的分割表現，不依賴尚未驗證的協方差或最適性假設。

2.6 KiTS 資料集與既有方法

KiTS (Kidney and Kidney Tumor Segmentation) 系列競賽自 2019 年起舉辦[20]，本作品採用其 2023 年版本 (KiTS23) [21]。該資料集共 599 例，其中 489 例為公開訓練集、110 例為標註未公開的測試集。病例來自 2010 至 2022 年間因疑似腎臟惡性腫瘤而接受冷凍消融、部分腎切除或根治性腎切除的患者，影像涵蓋皮髓質期與腎實質期兩種對比劑相位。KiTS23 的官方評估採用階層評估類別 (Hierarchical Evaluation Classes, HECs)，分別為腎臟區域 (腎臟、腫瘤與囊腫之聯集)、腫塊 (腫瘤與囊腫之聯集) 與腫瘤 (僅腫瘤本身)，並同時計算體積 Dice 與表面 Dice。KiTS23 冠軍流程[22] 以 MONAI Auto3DSeg 訓練 SegResNet 與 DiNTS，最終十五模型集成在官方測試集取得平均 Dice 0.835。公開訓練集上的研究則採用不同評估安排：稀疏卷積研究[23]以五折交叉驗證比較多種模型，Uhm 等

人[24]採 4:1 的單次訓練／驗證切分，GASA-UNet[27]另保留 89 例作為獨立測試集。這些結果均有參考價值，但其統計單位與推論方式並不相同。第四章因此將公開集內驗證與獨立保留集分表整理，並保留官方 110 例測試集與自訂切分的區別，使分數能連結至正確的資料來源。

2.7 本作品與既有架構的關係

前述文獻可依其處理的網路環節理解。U-Net 與 nnU-Net 提供特徵編碼、解碼及訓練配置的共同框架；下採樣研究關心解析度轉換時的資訊處理；Mamba 類方法則著重特徵間的序列依賴建模。本作品把 UMambaBot 作為共同基底，使方法改動主要落在解析度轉換及對應特徵融合，而保留原有瓶頸層的 Mamba。在此安排下，Mamba 與所提融合模組具有不同職責。瓶頸層負責在最深層表示上組織特徵關係；前段下採樣模組則決定較高解析度特徵如何被組合並傳向深層。完整模型再利用對應層級的 P 參與解碼器融合。三者共同形成端到端分割流程，但本文的基準對照只檢驗加入後兩項改動後的整體效果。與直接增加通道數或延伸網路深度相比，本作品保留既有的六個解析度層級與通道配置，將改動集中於部分層級的資料流。此方式便於對照輸入輸出尺寸，也能把計算成本的增減連結至具體操作：編碼器增加觀測投影，解碼器則以插值與通道投影替換部分轉置卷積。卡爾曼相關概念在本文中提供的是結構化設計語言：先建立預測，再利用分支差值修正，同時維持可遞迴傳遞的輔助量。其價值在於讓融合規則及跨層連結可以明確寫成公式。模型的任務效益則由同設定實驗判斷，理論來源、網路實作與實驗證據各自對應不同層次的論述。

2.8 稀疏卷積與區域式分割

一般三維卷積會在整個規則網格上計算特徵，即使其中相當部分屬於背景，仍須配置中間特徵圖。Alonso-Monsalve 等人[23]將子流形稀疏卷積用於腎臟與腫瘤分割，透過兩階段處理，先取得較粗略的區域資訊，再在較高解析度下分析相關體積。此路線的關鍵在於決定哪些空間位置需要被表示及計算，與在所有位置執行相同密集卷積有所不同。對三維影像而言，活躍位置的數量與分布會影響實際工作量，因此參數量相同也不代表不同病例的運算需求完全相同。原文另將 Sparse U-Net 與 nnU-Net v2、SegVol 置於共同評估背景中，使稀疏處理、任務專門訓練及零樣本推論的差異更容易辨識。區域式分割也可由密集網路實現。Auto3DSeg[22]在訓

練與推論中結合腎臟區域處理，Uhm 等人[24]則比較全解析度、低解析度、串接式及多尺度集成等配置。降低解析度使固定尺寸區塊能覆蓋較大的物理範圍，高解析度路徑則提供更細的局部表示；串接或集成設計利用兩者，但同時改變訓練與推論流程。KalMed-UNet 保留單一網路內的多層級表示，改動的是相鄰解析度之間的特徵融合，而非增加第二階段的影像區域模型。因此，兩種路線可由「計算哪些位置」及「如何傳遞已計算的特徵」加以區分。這個觀點也說明，外部研究的較低顯存或較高分數，可能來自完整流程的多項安排，須配合輸入尺寸、階段數及推論處理理解。

2.9 注意力與預訓練分割模型

除卷積與狀態空間模型之外，注意力方法透過特徵之間的關係建立較廣範圍的資訊交換。GASA-UNet[27]在 U 形網路中使用全域軸向自注意力，沿三維影像的軸向及平面組織特徵，並以位置資訊協助建立空間關係。其 KiTS23 實驗同時列出 VNet、UNETR、nnMamba 與 nnU-Net，使不同建模方式能在同一篇研究的資料安排下對照。對本作品而言，這類方法與瓶頸層 Mamba 均涉及較長距離的特徵關係，而卡爾曼下採樣主要處理局部解析度轉換。兩者的作用位置不同，故不能僅以「具備全域建模」或「使用自適應權重」判定方法等價；仍須比較權重的來源、運算對象及所連接的網路層級。預訓練分割模型則把部分能力建立在目標資料集以外的學習過程。SegVol[25]支援以提示引導體積分割，文獻[23]使用其公開權重，以腎臟、腎臟腫瘤及腎臟囊腫的文字提示進行 KiTS23 零樣本評估。此結果回答的是預訓練表示能否直接轉移至指定標籤，與使用 KiTS23 標註從頭訓練或微調後的結果具有不同意義。其模型參數較多，也不能據此預期必然取得更高的任務分數，因提示方式、預訓練涵蓋範圍及標籤定義皆會影響輸出。本文將其納入比較，是為呈現另一種取得分割能力的路線；KalMed-UNet 的重點仍是固定任務訓練流程下的架構改動，並以自行重現的 UMambaBot 作為主要對照。

2.10 卡爾曼啟發式分割方法

與本作品在設計概念上較接近的研究，包括 KaLDeX[28]及 KLDD[29]。兩者皆以視網膜血管為主要對象，透過卡爾曼結構調整線性可變形卷積的取樣位置，使感受範圍能沿細長的血管形態變化。KaLDeX 將此特徵擷取方式與交叉注意力結合於 U-Net++ 類架構；KLDD 則將擷取出的血管特徵作為擴散式分割的條件資訊。

它們說明卡爾曼概念可以用於神經網路內部的結構調整，而不侷限於影像序列的時間濾波；其公開結果及訓練條件另列於第四章。

三者的差異在於遞迴狀態與修正對象。本作品調整兩條下採樣分支的融合比例， P 隨解析度層級傳遞，並在解碼器參與淺、深層特徵融合；KaLDeX 與 KLDD 的重點則是取樣座標及血管特徵擷取。因此，不能僅因名稱中包含 Kalman，即把這些方法視為同一個模組，或以二維血管資料上的 Dice 直接衡量三維腎臟分割的優劣。對本作品而言，這些相關研究提供了明確的設計對照：卡爾曼啟發式結構的效益需要由其實際作用的位置、所使用的狀態，以及任務內的受控比較來說明。本文據此將貢獻界定為跨解析度融合的整合設計，並以既有五折實驗支持其效用。

2.11 三維分割方法的比較依據

三維分割方法的效能由多個環節共同決定，單一架構名稱不足以描述完整實驗。重取樣間距決定體素所代表的物理範圍，patch size 決定每次可見的脈絡，batch size 與梯度累積則影響參數更新方式；同樣的訓練 epoch 數，亦可能對應不同的每輪迭代次數。推論端的區域裁切、滑動視窗、鏡射增強及模型集成，還會改變最後預測所使用的資訊與計算需求。因此，本文整理外部方法時，同時記錄資料切分、解析度、輸入尺寸、最佳化器、損失函數、訓練長度與推論處理，讓分數具有可解讀的實驗背景。硬體與模型規模需要另行辨識。參數量描述可學習權重的數量，顯存還受輸入體積、特徵圖、最佳化器狀態及數值精度影響；多階段模型或模型集成，也不宜只用其中一個網路的參數量代表完整流程。GPU 型號若出現在推論效能測試中，並不能據此認定訓練亦使用相同設備。本文依原文分別記錄訓練硬體及推論測試硬體，未報告的欄位以 NR 表示，不以框架預設值推定。此整理方式使文獻比較既能涵蓋不同設計路線，也能保留現有公開資訊的界限；本作品本身的優勢則以共同設定下的 Dice 改善及參數量、GFLOPs 下降來呈現。

2.12 融合增益與門控機制的關係

本作品的增益可從門控融合的角度理解。對齊後的兩分支若直接相加，輸出尺度會同時受到兩者幅值影響；若取算術平均，則在所有位置固定分配相同權重。公式(10)使用非負且和為 1 的兩個係數，讓輸出沿兩分支之間變化，並保留以預測特徵作為起點的殘差形式。這種加權結構本身並不是卡爾曼濾波專有，區分方法時更重要的是權重由哪些資訊產生，以及更新後還保留哪些狀態。在 KalMed-UNet

中，權重同時使用當前分支的標準化差異及由前一解析度傳入的 P ，更新後的 P 又參與下一層與解碼器融合。因此，方法的辨識重點是增益生成、輔助狀態更新和跨層使用三者的連接方式，而非僅在公式中出現 K 或 sigmoid。這個觀點也使其與內容自適應取樣[7]及神經網路輔助濾波[19]之間的關係更清楚：它們皆可能依輸入改變運算，但權重所控制的對象及狀態所代表的資訊不同。與直接學習一個多通道門控張量相比，本文採每個位置一個增益，再廣播到該位置的所有特徵通道。這讓跨層狀態不必隨通道數重新定義，也使公式能用逐元素運算實作；相對地，它並未替每個通道分別估計應信任哪一條分支。注意力權重若描述不同位置間的關聯，通常涉及跨位置的資訊聚合，而本文增益是在兩條已對齊路徑間調整比重，不能直接解讀為一張器官重要度或病灶機率圖。相關概念的共同點可以用於說明設計動機，但本作品的五折對照只支持完整架構相對基準的任務效益，並未單獨比較平均融合、一般門控與本增益的高下。本文將可由公式確認的結構性質與由實驗觀察的效能分開呈現，使方法定位集中於實際完成的整合設計，也避免由名稱相近推論模型功能相同。

第三章 作品介紹

3.1 整體架構與符號定義

本作品以 UMambaBot 作為基準模型。其編碼器將輸入的三維 CT 影像逐層降維，於最深層以 Mamba 模組進行全域建模，再由解碼器逐層上採樣還原至原始解析度，並透過 skip connection 串接對應層級的編碼器特徵。基準模型架構如圖 1 所示。依 nnU-Net 對本資料集的自動配置，網路共有六個解析度層級。令 k 為層級索引、 F_k 為該層的特徵圖、 C_k 為其通道數，各層的空間尺寸與通道數如下： F_0 為 128^3 (32 通道)、 F_1 為 64^3 (64 通道)、 F_2 為 32^3 (128 通道)、 F_3 為 16^3 (256 通道)、 F_4 為 8^3 (320 通道)、 F_5 為 4^3 (320 通道)。其中第一層的步幅為 $1 \times 1 \times 1$ ，即不進行降維，故實際的降維次數為五次，空間解析度共縮減 32 倍。通道數上限設定為 320。解碼器對應層級的特徵以 D_k 表示。

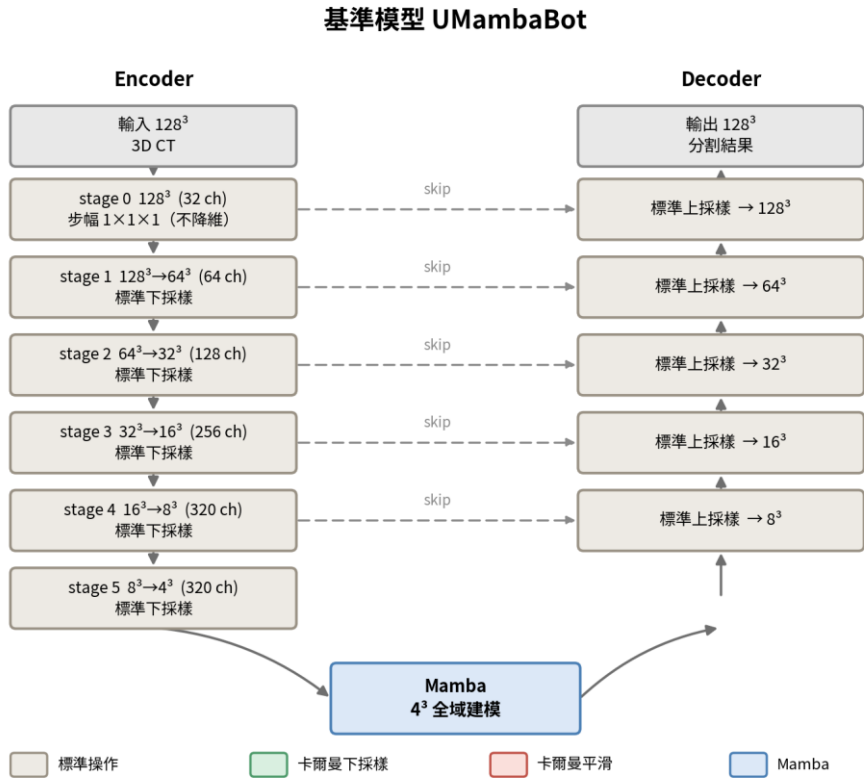


圖 1.基準模型 UMambaBot 架構

本作品的第一階段改動為 KalmanDown，將編碼器 stage 1 ($128^3 \rightarrow 64^3$) 與 stage 2 ($64^3 \rightarrow 32^3$) 的標準下採樣替換為卡爾曼啟發式下採樣，其餘層級維持標準操作。各層的 P 各自初始化、不跨層傳遞，如圖 2 所示。選擇前兩個降解析度層

級的動機，是在較高解析度處進行分支融合；此配置是否優於其他放置方式，仍需層級對照實驗驗證。第二階段的完整版本 KalMed-UNet 在相同的兩個層級使用卡爾曼啟發式下採樣，並將不確定度代理量 P 由 stage 1 傳至 stage 2。解碼器於對應的 64^3 與 32^3 層級進行受後向平滑啟發的融合，其餘層級的上採樣維持標準操作，如圖 3 所示。

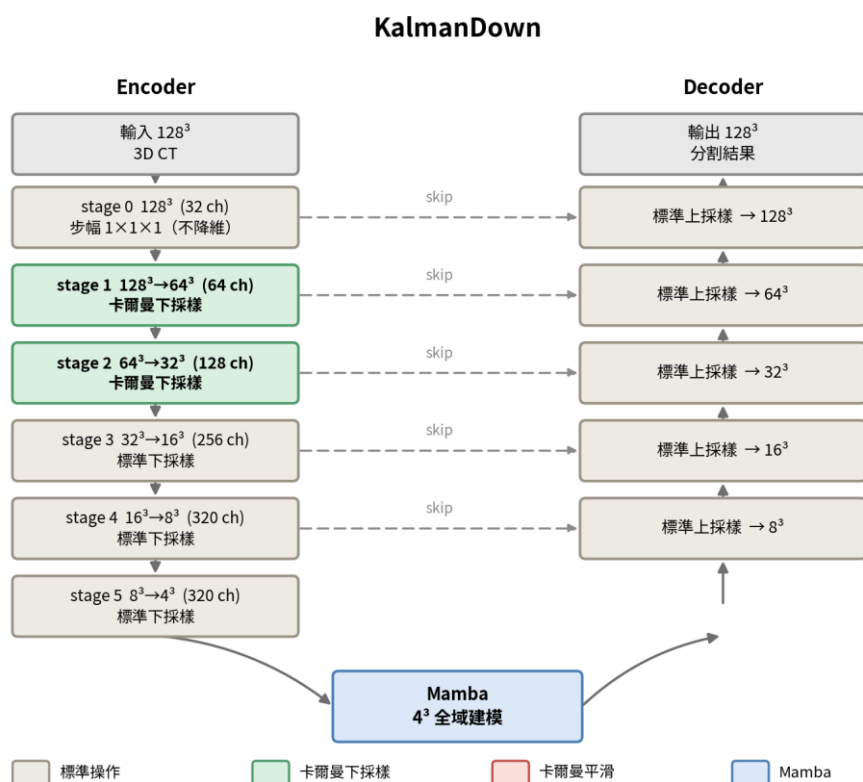


圖 2. KalmanDown 編碼器雙分支下採樣與各層獨立的 P

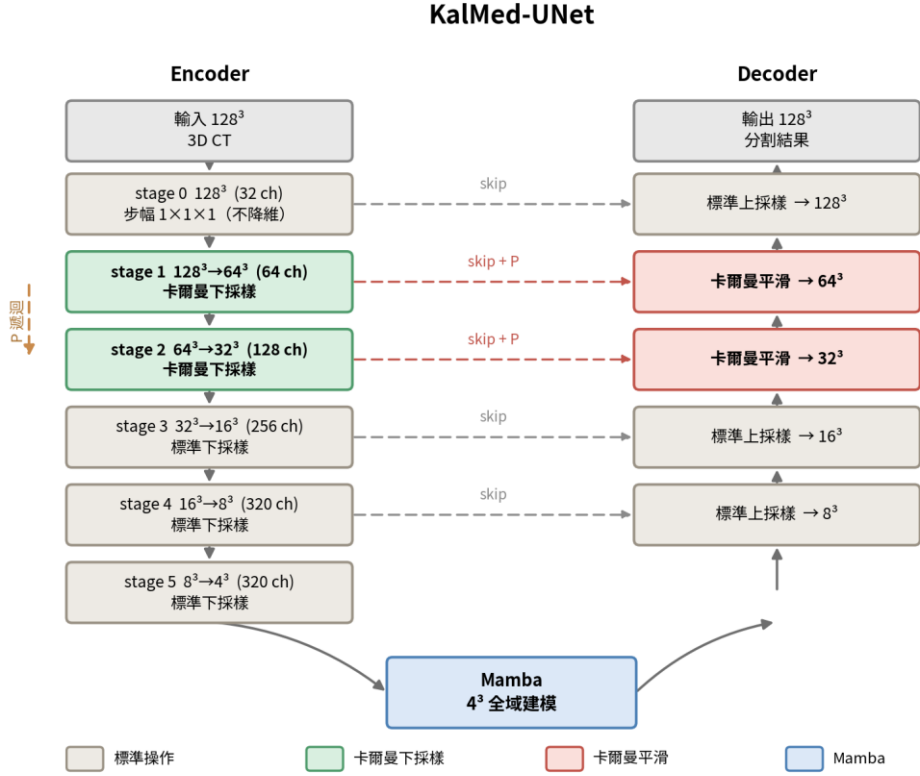


圖 3. KalMed-UNet 跨層 P 遞迴與解碼器後向融合

3.2 標準下採樣的限制

本作品的標準步幅卷積下採樣如公式(1)所示，其中 W_k 為卷積核， $\ast_s$ 表示步幅為 2 的卷積，Norm 泛指該層正規化操作， φ 為非線性激活函數。卷積核於各位置共享，但輸出仍依局部特徵而變化；本研究關注的是此單分支操作未顯式利用另一種降採樣表示，也未維持跨層融合的輔助狀態。

$$F_{k+1} = \varphi(\text{Norm}(W_k \ast F_k)) \quad (1)$$

為引入另一種局部資訊組合方式，本作品在步幅卷積之外增加最大池化與通道投影分支。兩條路徑先完成空間與通道對齊，再由其差值及 P 決定修正比例。這使融合規則能明確表示為「預測加上加權殘差」，並提供可傳至下一層的單通道狀態。此處不預設最大池化一定保留較正確的結構；兩分支如何共同服務分割任務，仍由端到端訓練與第四章的基準對照檢驗。

3.3 卡爾曼下採樣

本作品以解析度層級 k 作為遞迴索引，將步幅卷積輸出稱為預測特徵 x^- ，池化投影輸出稱為觀測特徵 z ，兩者的差值為殘差 r 。傳入的 P 先以自適應最大

值池化對齊解析度，再加上可學習的非負項 Q ，首個啟用模組則從 $P^- = 1 + Q$ 起始。最大值池化在此保留區域內較大的代理量數值，作為跨尺度傳遞規則；此操作不等同於標準協方差預測 $FPF^T + Q$ 。表 1 整理記號與網路組件，圖 4 呈現完整資料流，公式(2)至公式(7)則依序列出預測、增益及更新步驟。

表 1.卡爾曼濾波要素與網路組件之對應

卡爾曼濾波要素	本作品之對應	維度／實作說明
遞迴索引 k	解析度層級	$k = 0, \dots, 5$
狀態轉移 F	步幅為 2 的 $3 \times 3 \times 3$ 卷積	即基準模型原有的下採樣層
先驗估計 x^-	步幅卷積的輸出	$C_k \times D \times H \times W$
觀測 z	MaxPool $2 \times 2 \times 2$ 後接 $1 \times 1 \times 1$ 卷積	$C_k \times D \times H \times W$
觀測矩陣 H	恆等映射	觀測已投影至輸出通道空間
殘差 r 與調變量 $\tilde{r}$	$r = z - x^-$ ； $\tilde{r}$ 由 InstanceNorm 後的 r 經通道平均絕對值與空間標準化得到	$r: C_k \times D \times H \times W$ ； $\tilde{r}: 1 \times D \times H \times W$
過程雜訊形式的項 Q	softplus(Q_{raw})，可學習非負純量	初值 -2.0 ，即 $Q \approx 0.127$
觀測雜訊形式的尺度 R	固定為 1，未經雜訊校準	隱含於增益式的分母
協方差形式的代理量 P	每體素一個非負純量，跨通道共用	$1 \times D \times H \times W$ ；未校準為真實誤差協方差

卡爾曼下採樣模組內部資料流

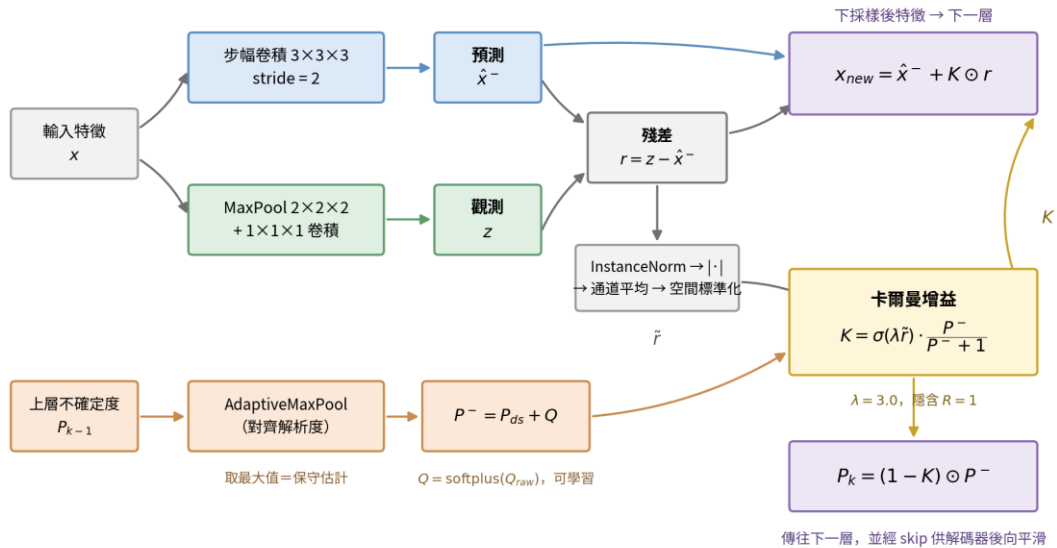


圖 4.卡爾曼下採樣模組之內部資料流

$$x_k^- = \text{Conv}_{\text{stride } 2}(x_{k-1}) \quad (2)$$

$$P_k^- = \text{MaxPool}(P_{k-1}) + Q, \quad Q = \text{softplus}(Q_{\text{raw}}) \quad (3)$$

$$r_k = z_k - x_k^-, \quad z_k = \text{Conv}_{1 \times 1}(\text{MaxPool}(x_{k-1})) \quad (4)$$

$$K_k = \sigma(\lambda \cdot \tilde{r}_k) \cdot \frac{P_k^-}{P_k^- + 1} \quad (5)$$

$$x_k = x_k^- + K_k \odot r_k \quad (6)$$

$$P_k = (1 - K_k) \odot P_k^- \quad (7)$$

增益式中的 σ 為 sigmoid， λ 固定為 3.0， $\odot$ 表示逐元素相乘。殘差先經 InstanceNorm，再逐通道取絕對值並平均，最後沿空間維度標準化得到 $\tilde{r}$ ；因此使用的是通道平均絕對值，而特徵修正仍採保留符號及通道維度的原始殘差 r 。 P 、 $\tilde{r}$ 與 K 均為單通道空間場，更新時沿特徵通道廣播，不涉及完整協方差矩陣求逆。 $P^-/(P^- + 1)$ 借用純量卡爾曼增益在 $R = 1$ 時的形式，局部調變項 $\sigma(\lambda\tilde{r})$ 則使其隨分支差異變化。較大的殘差表示分支較不一致，但不能單獨判定哪一支較正確； P 的更新亦只是協方差式的遞迴外形。由於兩分支同源，且未顯式建立誤差相關性與雜訊校準，本文將 P 視為輔助融合的代理量，並以任務結果評估此設計 [17][26]。這使公式中的每項運算都有清楚的功能定位，同時保留其與標準濾波理論的適用範圍差異。

3.4 不確定度代理量的跨層遞迴

在 KalmanDown 中，stage 1 與 stage 2 各自從均勻初始 P 起算，不傳遞前一層的結果。KalMed-UNet 則將 stage 1 更新後的 P 對齊至 stage 2，作為公式(3)的輸入，使後一層的融合權重受到前一層輔助狀態的影響。此為跨解析度的遞迴資訊傳遞，不代表已量測或校準各層的真实估計誤差。相鄰層級的空間尺寸不同，傳遞前以自適應最大值池化對齊。未啟用卡爾曼模組的層級若沿用 P ，僅執行尺寸對齊，不進行公式(7)的更新。編碼器保留對應層級的 P ，供解碼器在 64^3 與 32^3 兩處融合時使用；其餘上採樣層維持標準操作。跨層 P 遞迴使編碼器各層的融合具有前後連結，並為解碼器提供對應的輔助狀態。表 4 中，整合此傳遞機制、解碼器融合與上採樣改動的完整模型，整體 Dice 由 KalmanDown 的 0.8313 提升至 0.8323，且計算成本下降。因此，該對照呈現的是完整整合設計的效益，各元件的獨立貢獻則屬後續可延伸分析的範圍。

3.5 受後向平滑啟發的解碼器融合

標準 U-Net 的 skip connection 傳遞編碼器特徵；本作品另將對應層級的 P 提供給解碼器，以調節淺層特徵與深層上採樣特徵的比例。此設計受 RTS 後向平

滑[18]利用後續資訊修正較早估計的思路啟發，但本文所使用的 P 並非經校準的誤差協方差。解碼器於 64^3 與 32^3 兩個層級設置平滑層，恰與編碼器啟用卡爾曼下採樣的層級相對應。每一個平滑層先將深層結果上採樣至當前解析度，再依公式(8)與公式(9)進行融合。

$$G_k = \frac{P_k}{P_k + 1} \quad (8)$$

$$x_{\text{smooth}} = x_{\text{filt}} + G_k \odot (x_{\text{up}} - x_{\text{filt}}) \quad (9)$$

其中 x_{filt} 為編碼器 skip connection 的特徵， x_{up} 為較深層特徵經上採樣與通道對齊後的結果； $G_k = P_k / (P_k + 1)$ 為解碼器融合增益。 P_k 增加時，融合較偏向深層分支， P_k 較小時則保留較多淺層特徵。此規則把編碼器保留的單通道狀態連結至解碼器，但未建立標準 RTS 增益 $P_k F^T (P_{k+1}^-)^{-1}$ 所需的跨層協方差，因此稱為受後向平滑啟發的融合[18]。實作中，深層結果以最近鄰插值搭配 $1 \times 1 \times 1$ 卷積完成對齊，對應 P 缺失時保留標準上採樣路徑。

3.6 複雜度分析

KalMed-UNet 的參數量為 41.57 M、計算量為 1852.71 GFLOPs，較基準分別減少 1.31% 與 4.48%。此結果由編碼器與解碼器的成本共同形成：卡爾曼下採樣增加觀測分支的 $1 \times 1 \times 1$ 投影、正規化及可學習純量，本身並不減少參數；表 7 中僅調整下採樣的 KalmanDown 因而較基準增加約 0.03% 參數。完整架構則在相對應的解碼器層級替換上採樣操作，使局部新增的成本可由另一部分的減量抵銷。若只計卷積權重並忽略偏置，核尺寸 $2 \times 2 \times 2$ 的轉置卷積具有 $8C_{\text{in}}C_{\text{out}}$ 個參數，而最近鄰插值本身沒有可學習權重，後接的 $1 \times 1 \times 1$ 卷積則為 $C_{\text{in}}C_{\text{out}}$ 。這項比較說明相同通道對應下的局部參數差異；整體 GFLOPs 仍須包含插值、投影、融合及其他網路層的實際運算，不能直接由八倍的權重比例推定。第四章量測顯示，完整配置在抵銷編碼器增量後，仍有參數與計算量的淨減少；於 250 epochs 的設計比較中，整體 Dice 也由 KalmanDown 的 0.8313 增至 0.8323。這些結果共同說明整合架構的取捨，而非將所有效率改善歸因於卡爾曼下採樣本身。

3.7 張量尺寸與模組介面

為便於對照圖 4 與公式，以下以批次大小 B 、通道數 C 及空間尺寸 D 、 H 、 W 表示特徵張量。輸入 x 的形狀為 $B \times C_{\text{in}} \times D \times H \times W$ 。下採樣後，預測分支

與觀測分支均須得到 $B \times C_{\text{out}} \times D' \times H' \times W'$ ，才能進行逐元素殘差計算。本文所使用的前兩個降解析度層級，各軸尺寸均減半。以 stage 1 為例，輸入特徵為 $B \times 32 \times 128 \times 128 \times 128$ ，兩分支對齊後皆為 $B \times 64 \times 64 \times 64 \times 64$ ；stage 2 的輸入則為 $B \times 64 \times 64 \times 64 \times 64$ ，輸出為 $B \times 128 \times 32 \times 32 \times 32$ 。觀測路徑中的 $1 \times 1 \times 1$ 卷積負責通道投影，使其能與步幅卷積輸出比較。最大池化負責空間降採樣，兩項操作分別處理空間與通道的對齊。殘差 r 保留與兩分支相同的通道數。計算增益時，先對正規化殘差取絕對值，再沿通道平均，因此局部調變量的通道數縮為 1。 P 與 K 也以單通道空間場表示，在特徵更新時沿通道維度廣播。每個位置的融合比例由多個通道的分支差異共同形成，而實際修正方向仍由各通道帶正負號的原始殘差決定。

通道平均絕對值與平均後取絕對值具有不同意義。若某一位置只有兩個通道，其正規化殘差為 2 與 -2，先取絕對值再平均的結果為 2；先平均再取絕對值則為 0。本文採前者，使不同通道的相反符號不會在增益輸入中互相抵銷。此例僅說明運算差異，並非對實際特徵分布的量測。跨層傳遞時， P 需要的是空間尺寸對齊，而不必跟隨特徵通道由 64 擴增至 128。這是單通道代理量的介面優點：不同層級可以使用同一形式的輔助狀態，並在各自的特徵通道上廣播。stage 1 的 P 經尺寸對齊後提供給 stage 2，編碼器同時保留相應層級的 P ，供解碼器在相同解析度使用。解碼器融合也須先完成尺寸與通道對齊。 x_{up} 與 skip 特徵 x_{filt} 必須具有相同形狀，才能計算差值；對應 P 所形成的增益再沿通道廣播。由此可見，圖 3 中的 skip 與 P 是兩種不同的傳遞內容：skip 提供多通道特徵， P 提供逐位置融合係數的來源。兩者分工使完整資料流可以從編碼器一路追蹤至解碼器。

3.8 融合公式的代數性質

當 P^- 為有限的非負值時， $P^-/(P^- + 1)$ 介於 0 與 1 之間，sigmoid 輸出亦位於此範圍，故融合增益滿足 $0 \leq K < 1$ 。將 $r = z - x^-$ 代入公式(6)，可得到公式(10)的凸組合形式。對已計算出的 K 而言，每個位置、每個通道的結果位於兩分支對應值之間； K 接近 0 時偏向預測分支，增加時則提高觀測分支比重。固定 $a = \sigma(\lambda \tilde{r})$ 時， $K = aP^-/(P^- + 1)$ 對 P^- 的導數為 $a/(P^- + 1)^2$ ，顯示此分式的敏感度隨 P^- 增加而下降。這些性質說明單次融合所允許的行為，不等同於完整網路的輸出界限或收斂保證。

$$x_k = (1 - K_k) \odot x_k^- + K_k \odot z_k \quad (10)$$

由公式(7)可知，非負的 P^- 經更新後仍保持非負，且 P 不大於該次的 P^- ；然而下一層會在尺寸對齊後加入非負 Q ，因此不能將單次更新的不增性解讀為 P 沿整個網路必然遞減。解碼器的 $G = P/(P + 1)$ 具有相同的權重範圍，故公式(9)亦可視為淺、深層特徵的凸組合，但它沒有編碼器 K 的殘差調變項。 $Q = \text{softplus}(Q_{\text{raw}})$ 確保可學習純量非負，分母中的固定值 1 則避免非負 P 下的零分母。另一方面， $\tilde{r}$ 已經空間標準化， $\lambda = 3.0$ 調整的是相對位置差異經 sigmoid 後的反應，不能把其數值直接當成原始殘差的信心分數。以上分析釐清代理量及增益的代數範圍，也提供核對實作的依據；分割效能仍由第四章的受控比較判斷。

3.9 初始設定與局部更新範例

為具體說明公式的作用，以下以初始設定計算一個局部位置的更新，不涉及新增訓練或量測。表 1 設定 $Q_{\text{raw}} = -2.0$ ，因此 $Q = \text{softplus}(-2.0)$ 約為 0.1269；首個啟用模組的 $P^- = 1 + Q$ 約為 1.1269。若某位置的標準化調變量 $\tilde{r}$ 恰為 0，sigmoid 項即為 0.5，代入公式(5)可得 K 約為 0.2649。此時預測分支與觀測分支的權重分別約為 0.7351 與 0.2649。若僅考慮該位置的一個通道，預測值為 2、觀測值為 4，則原始殘差為 2，更新結果約為 2.5298；修正方向由原始殘差決定，標準化後的調變量則決定其幅度。相同 K 會沿通道廣播，但各通道的殘差可以不同，故各通道並不會得到相同的更新值。這個範例也說明， $\tilde{r} = 0$ 不代表原始殘差必為 0，兩者分別位於經正規化的調變路徑與原始特徵修正路徑。在同一組條件下，更新後的 P 約為 0.8284，若提供給對應解碼器， $G = P/(P + 1)$ 約為 0.4531，表示淺層特徵與深層上採樣特徵的比重約為 0.5469 與 0.4531。這些數字來自指定的初始狀態及局部調變量，不代表訓練後整個特徵場的平均增益。若兩分支原始特徵相同，殘差為零，則無論 K 如何變化，特徵更新皆保持原值；若 P^- 接近零，增益也接近零，結果則較接近預測分支。另一方面，即使 $\tilde{r}$ 很大， K 仍受到 $P^-/(P^- + 1)$ 的限制，並非直接把觀測分支完全覆蓋到輸出。這些邊界情況將兩項調節因素的分工具體化：殘差路徑反映當前分支差異， P 路徑則限制融合的尺度並保留跨層狀態。其實際分布與任務效益仍以訓練後模型及分割結果為準，不能由初始數值推定模型的可靠程度。

3.10 解析度配置與實體視野

三維網路的輸入尺寸需與體素間距一併理解。本作品採 $128 \times 128 \times 128$ 的 patch，重取樣間距為 $1.0 \times 0.78125 \times 0.78125$ mm；以體素數乘上間距估算名目覆蓋範圍，三軸約為 $128 \times 100 \times 100$ mm。因此，張量雖為立方體，其涵蓋的物理範圍並非三軸等長。文獻[24]的低解析度配置同樣使用 128^3 patch，但間距約為 $1.84 \times 1.84 \times 2.36$ mm，對應範圍約為 $235.5 \times 235.5 \times 302.1$ mm。兩者的軸向排列依各自配置列示，數值僅用於說明間距與覆蓋範圍的關係。這項直接計算顯示，相同 patch size 不代表模型看到相同尺度的脈絡：粗解析度可納入較大的區域，細解析度則用更多體素表示局部結構。單看體素邊長或「全解析度」名稱，因而不足以判斷不同模型的輸入條件。網路內部的下採樣又與影像重取樣具有不同角色。重取樣先決定送入網路的體素尺度，編碼器則在此輸入上逐層建立較粗的特徵表示；本作品前兩個改動層級將 128^3 降至 64^3 ，再降至 32^3 ，並將通道數由 32 增至 64、128。僅以張量元素數計算，stage 1 輸出為 64×64^3 ，約為輸入 32×128^3 的四分之一；stage 2 亦有相同的元素數比例。通道增加雖提供新的特徵表示空間，總元素數仍明顯下降，這正是解析度轉換需要仔細安排的原因。上述元素數不能直接換算成峰值顯存，因訓練還需保存多個中間張量及梯度；也不能只由下採樣倍率推定感受野，因卷積堆疊與其他模組皆有影響。KalMed-UNet 在此過程中增加第二種同源轉換及融合狀態，讓降低解析度時的特徵組合具有明確規則，同時維持與基準相同的六層尺寸及通道配置。

3.11 共享增益與通道殘差的分工

編碼器中的增益計算與特徵修正使用不同形式的殘差，這個區別直接影響模組的作用。計算增益時，正規化後的各通道殘差先取絕對值再平均，使通道間符號相反的差異不會相互抵銷；實際更新則使用原始的多通道殘差，所以每個通道仍保有自己的修正方向。以某位置的兩個通道為例，若原始殘差分別為 2 與 -1，已計算出的共享增益為 0.25，兩通道的修正即為 0.5 與 -0.25，而非對兩者加入同一個數值。此為公式(6)的直接代入，用於說明廣播操作，並非另一次特徵量測。共享的是分支的混合比例，並不是把特徵通道合成一個值；只有產生局部調變量的路徑縮減了通道維度，主特徵路徑的通道數仍維持不變。因此，單通道 P 與多通道特徵可以同時存在，各自負責狀態傳遞及內容表示。

空間標準化使局部調變量反映同一特徵場內的相對差異，故某個位置的增益不能只依原始殘差大小獨立判斷。若一個區域的兩分支差異較周圍明顯，該區域在調變路徑中的相對位置可能提高；但這種差異也可能來自背景紋理、局部雜訊或兩種轉換的響應方式，並不等同於病灶邊界。再乘上 $P^-/(P^- + 1)$ 後，增益才取得跨層狀態所限制的尺度。這個分工使局部內容與前一層狀態共同參與融合，同時保留明確的權重範圍。若未來將增益視覺化，也應先區分它是編碼器 K、解碼器 G，或尚未經分式轉換的 P，因三者使用的資訊及數值意義不同。本文的分割案例展示最終輸出，並未把其中的改善直接等同於某張增益圖的效果；對目前模型的解讀仍以此處可核對的資料流，以及第四章的整體對照為依據。

3.12 融合權重與誤差變異數的區別

標準卡爾曼增益與誤差協方差之間的關係，可用純量估計的特例說明[17][26]。假設兩個估計量針對同一真值，皆為無偏估計，其誤差變異數分別為 p 與 v ，且誤差互不相關；若以固定權重 K 組合兩者，融合誤差變異數便是 $J(K) = (1 - K)^2p + K^2v$ 。在 $p + v > 0$ 時，對 K 微分並令其為零，可得 $K = p/(p + v)$ ，這才是依兩種誤差大小決定權重的由來。當 $v = 1$ 時，式子與本作品的 $P^-/(P^- + 1)$ 有相同外形，但 p 的定義是估計誤差的變異數，並非任意非負特徵場。上述推導把兩估計量、真值及誤差關係事先定義清楚，所得最小化結果也只在這些條件下成立。相較之下，本作品的兩分支由同一特徵轉換而來，沒有一組已知乾淨特徵作為共同真值， P 也不是由實際誤差平方的期望值建立，因此本文借用的是更新結構與增益形狀。若兩個無偏估計的誤差另有共變異數 c ，固定權重下的變異數須增加交叉項，成為 $J(K) = (1 - K)^2p + K^2v + 2K(1 - K)c$ ；當 $p + v - 2c$ 大於零時，無限制最小化的權重則為 $(p - c)/(p + v - 2c)$ 。這項由同一平方展開得到的關係顯示，兩條路徑是否具有相關誤差，會改變權重與最小誤差之間的對應。本作品沒有估計此交叉項，而且 K 還會依當前殘差的調變量改變，故不能把純量特例的最適性直接套用到網路。相同理由也適用於 $P = (1 - K)P^-$ ：它在標準假設及相應最適增益下可具有後驗協方差的意義，在本文則是一個定義明確的遞迴規則。保留這個區分並不影響其作為特徵融合模組的可訓練性；它使研究的理論主張限於可由公式確認的範圍，而方法是否有助於分割，由相同資料切分及訓練程序下的結果回答。

3.13 代理量儲存與整體顯存需求

P 的單通道設計可由張量大小估算其儲存規模。若僅為計算而假設所有元素使用 32 位元浮點數，每個元素占 4 bytes，張量資料量便等於元素總數乘上 4；以下以 MiB 表示 2^{20} bytes，與十進位 MB 區分。在 batch size 2 下，stage 1 的 P 形狀為 $2 \times 1 \times 64 \times 64 \times 64$ ，單一張量約占 2 MiB；stage 2 的 P 為 $2 \times 1 \times 32 \times 32 \times 32$ ，約占 0.25 MiB，兩者合計 2.25 MiB。同樣條件下，兩個層級的單一多通道特徵張量分別具有 64 及 128 個通道，資料量約為 128 MiB 與 32 MiB。這些數字由既有張量尺寸推算，用於比較狀態與特徵的相對規模，並非 GPU 峰值顯存的量測。若單一張量改用每元素 2 bytes 的表示，其資料量會減半，但不能因此假設完整訓練流程的記憶體需求也恰好減半。

實際顯存需求還包含兩分支輸出、殘差、增益、卷積中間結果及反向傳播所需資料，此外有模型權重、梯度、最佳化器狀態與運算暫存空間。哪些張量同時存活、哪些可被重用，以及不同算子的數值精度，都會影響峰值，因此不能把上述兩個 P 張量直接加到模型參數量上，便宣稱得到完整訓練成本。本文在此可確認的設計優點是輔助狀態的通道數固定，傳至下一層時只需對齊空間尺寸，不會跟隨主特徵由 64 通道擴增至 128 通道。這與第四章參數量及 GFLOPs 的減少屬不同層次的資訊：前者說明單一資料結構的規模，後者是整體架構的量測。兩者共同交代設計的工程背景，但 GPU 延遲及峰值顯存仍須在指定執行條件下另行量測，本文不由尺寸估算生成尚未實際取得的效能數據。

3.14 空間對齊對代理量傳遞的影響

跨解析度傳遞 P 時，除了維持單通道介面，還必須決定高解析度的多個位置如何對應到下一層。本作品使用自適應最大值池化；在 64^3 對齊至 32^3 的規則尺寸下，可將其理解為每個 $2 \times 2 \times 2$ 區域保留最大數值。若一個區域內的 P 都相同，對齊前後的代表值不變；若只有部分位置較大，輸出則保留這些位置的較大值，而不因區域內其他較小值而被平均稀釋。例如八個數值中，七個為 0.2、一個為 0.8，最大值池化的結果為 0.8，算術平均則為 0.275。這個算例只比較聚合規則，並不是額外訓練一個平均池化版本。完成空間對齊後，公式(3)再加入非負 Q，使下一層增益使用的是該層的 P^- 。固定殘差調變量時，較大的 P^- 會提高分式增益，故空間聚合方式實際參與了後一層的融合尺度，而不只是調整張量形狀。

此規則同時帶來一項清楚的空間取捨：高解析度的一個較大值可能影響下一層整個對齊位置，但傳遞後無法只由 P 還原它原先位於區域內何處。因此，P 的角色是提供較粗解析度的融合狀態，細部位置資訊仍由多通道特徵及相應 skip connection 承載。解碼器在 64^3 與 32^3 使用的是編碼器保留的對應層級 P，不應把較深層 P 經任意上採樣後視為所有淺層狀態的替代品。這樣的介面使每個融合位置都有可追溯的來源：主特徵沿編碼器與解碼器流動，P 則依解析度對齊、更新並在指定層級被讀取。其設計目的在於讓狀態傳遞保持規則且可實作；最大值所保留的是代理量的相對大小，而不是經統計校準的最壞誤差界限。對讀者而言，將空間聚合、非負增量及對應層級的保存分開理解，便能由圖 3 與圖 4 追蹤完整模型的輔助資料流。

第四章 成果展示

4.1 資料集

本作品採用 KiTS23 公開訓練集，共 489 例三維對比劑增強 CT 影像，標註類別包含腎臟、腫瘤與囊腫。該競賽另有 110 例測試集，惟其標註未公開釋出，本作品無法取得，故所有評估均在公開訓練集上以交叉驗證進行。資料集概要如表 2 所示。每個病例只屬於一個驗證 fold，其餘四組提供該折的訓練資料，五組驗證例數分別為 98、98、98、98 與 97。以病例而非切片為切分單位，可使同一病例的體積資料保持完整，並與三維分割的評估單位一致。

表 2. KiTS23 資料集概要

項目	內容
資料來源	KiTS23 (2023 Kidney and Kidney Tumor Segmentation Challenge)
公開訓練集	489 例
測試集	110 例 (標註未公開，本作品未使用)
影像類型	對比劑增強 CT，皮髓質期與腎實質期
標註類別	腎臟、腫瘤、囊腫
交叉驗證切分	5-fold，驗證集分別為 98/98/98/98/97 例
重取樣間距	$1.0 \times 0.78125 \times 0.78125$ mm
中位影像尺寸	$417 \times 512 \times 512$ voxels
強度正規化	CTNormalization，裁切至 $[-58, 302]$ HU 後標準化
前景強度統計	平均 103.12、標準差 73.37 HU

表 2 中的重取樣間距與強度統計為本作品採用的資料配置。HU 裁切限制輸入的強度範圍，後續標準化則調整數值尺度；重取樣用於統一體素間距，使相同尺寸的 patch 對應較一致的物理範圍。 128^3 的訓練 patch 並不等於整個病例體積，推論仍需以滑動視窗覆蓋完整影像。這些前處理及推論安排在基準與完整模型間保持一致，使主要對照集中於網路內部的下採樣與融合方式。

4.2 評估指標

本文以 Dice 相似係數衡量預測區域與標準答案的重疊，如公式(11)所示，數值越接近 1 代表兩者越一致。式中 S_{pred} 與 S_{gt} 分別表示預測集合與標準答案集合，以避免與第三章的不確定度代理量 P 混淆。評估依 KiTS23 階層類別規則進行，而非只對三個互斥標籤各自計分。若以 TP、FP 與 FN 分別表示正確前景、誤判前景及遺漏前景的體素數，同一公式亦可寫為 $2TP/(2TP + FP + FN)$ 。因此，漏分與多分都會降低 Dice，且同樣數量的錯誤對不同大小目標具有不同的相對影響。作為指標的算例，若標準答案含 100 個前景體素，預測完整涵蓋它們但多出

10 個體素，Dice 約為 0.9524；若標準答案只有 10 個體素而同樣多出 10 個，Dice 則約為 0.6667。此例由集合公式直接計算，用於說明尺度敏感性，不是本作品新增的病灶分組結果。

$$\text{Dice} = \frac{2|S_{\text{pred}} \cap S_{\text{gt}}|}{|S_{\text{pred}}| + |S_{\text{gt}}|} \quad (11)$$

三個評估類別依序為腎臟區域（腎臟、腫瘤及囊腫的聯集）、腫塊（腫瘤與囊腫的聯集），以及腫瘤本身，整體分數為三者的算術平均。這個階層關係有助於解讀錯誤：例如將腫瘤錯分為囊腫時，病灶聯集可能仍被保留，但腫瘤指標會下降。因此，本文同時列出三類及整體 Dice，使改善不只由單一平均值呈現。五折結果另報告平均與標準差，逐折差值則用於觀察相同切分下的變化。KiTS23 官方亦採表面 Dice，本作品主要評估體積重疊，故外部比較僅使用可對應的體積 Dice，不以視覺上較平滑的輪廓替代邊界指標。三個 HEC 具有包含關係，並非彼此獨立的三組病例或標籤。若某處腫瘤被誤分為背景，可能同時影響三個聯集的重疊；若被分為囊腫，影響則主要落在腫瘤項目。因此，三類皆改善代表本次評估目標的整體表現提升，不能進一步等同於每種病灶尺寸皆改善。另一方面，病例間平均、折內平均與跨折平均各有不同的彙整層級，本文保留既有紀錄的分數及標準差，並列出逐折結果供追溯。外部文獻的分數亦依來源精度呈現，不由表內四捨五入後的類別數值重新生成另一個整體分數。

4.3 實驗環境與訓練設定

本作品以 PyTorch 2.7.0、CUDA 12.8 與 nnU-Net v2 實作，沿用 U-Mamba 的訓練流程，於單張 NVIDIA GeForce RTX 5080（16 GB）進行訓練，設定如表 3。基準及完整模型皆採 128³ patch、batch size 2、SGD 與 Nesterov momentum 0.99，初始學習率為 0.01 並使用 poly decay；損失函數結合 Dice 與 Cross-Entropy，啟用 batch Dice 及深度監督。固定這些設定的目的，是在共同資源及最佳化流程下比較架構改動，避免將額外訓練預算混入模組的效能差異。

表 3.實驗環境與訓練設定

項目	設定
GPU	NVIDIA GeForce RTX 5080（16 GB）
深度學習框架	PyTorch 2.7.0 + CUDA 12.8 / nnU-Net v2
網路配置	3d_fullres
Patch size	128 × 128 × 128

項目	設定
Batch size	2
最佳化器	SGD (Nesterov momentum = 0.99)
初始學習率與排程	0.01, poly decay
損失函數	Dice + Cross-Entropy (batch dice)
網路基礎通道數	32, 上限 320
Stage 數與降維次數	6 個 stage, 各軸降維 5 次
深度監督	啟用
訓練長度 (主要結果)	1000 epochs/fold
訓練長度 (消融實驗)	250 epochs, 僅單一 fold
單次訓練時間	約 10 小時/fold (1000 epochs)
推論方式	滑動視窗, 不使用測試時增強

主要結果各訓練 1000 epochs，單折訓練約需 10 小時；設計探索使用 250 epochs、單一 fold，以較低預算比較不同配置。推論採滑動視窗，未使用測試時增強、多模型集成或後處理，每個驗證病例由該 fold 的單一模型預測。五折平均因此是不同驗證組結果的彙整，並非把五個模型共同套用於同一病例。這個區別對外比較尤其重要，因部分文獻採五折模型集成於獨立測試集，其預測成本與評估方式均不同於本作品。

4.4 設計歷程與消融實驗

本節於固定的縮減訓練預算（250 epochs、單一 fold）下比較六個版本，呈現設計歷程與模組改動的相對效果，結果如表 4 所示。Kalman (S3)、Kalman + B 與 UK 為設計探索版本；基準模型、KalmanDown 與 KalMed-UNet 則呈現由下採樣改動至完整架構的演進。各版本在本節的相同預算下比較，4.5 節另以完整訓練預算檢驗最終模型。

表 4.設計歷程與消融實驗結果（250 epochs、單一 fold，數值為 Dice）

版本	整體	腎臟區域	腫塊	腫瘤
基準模型 (UMambaBot)	0.8184	0.9593	0.7813	0.7147
Kalman (S3)	0.8222	0.9611	0.7849	0.7205
Kalman + B	0.8272	0.9635	0.7913	0.7268
UK (S3 + S4)	0.8241	0.9635	0.7826	0.7264
KalmanDown	0.8313	0.9670	0.7958	0.7312
KalMed-UNet (本作品)	0.8323	0.9625	0.8030	0.7315

由表 4 可知，KalmanDown 相較於基準模型，在腎臟區域、腫塊與腫瘤三個階層評估類別上均取得提升，整體 Dice 由 0.8184 增至 0.8313。此結果顯示，下採樣模組的效益涵蓋三個評估目標，且所需額外參數與計算量有限，為後續完整架構的設計提供實驗依據。完整 KalMed-UNet 將整體 Dice 進一步提升至 0.8323，其中腫塊與腫瘤類別較 KalmanDown 改善，腎臟區域則略有下降，呈現不同評估

目標間的取捨。結合表 7 可見，完整模型在取得本組最高整體 Dice 的同時，參數量與計算量亦低於基準。此比較支持所提出整合配置的效能與成本表現，個別元件的效果未於本組實驗中分離。本作品將 UMambaBot 的 bottleneck 固定為共同基底，使設計比較聚焦於編碼器下採樣與解碼器融合。此安排保留相同的深層特徵建模方式，便於觀察不同融合配置在既定架構下的效果；跨骨幹的適用性則列為後續延伸方向。

4.5 5-fold 交叉驗證結果

本節於完整訓練預算（1000 epochs）下進行 5-fold 交叉驗證，將 489 例訓練資料切分為五組，依序以四組訓練、一組驗證。基準模型與 KalMed-UNet 使用相同的五組切分及訓練流程，使各 fold 可進行對應比較。此設計可觀察效能增益在不同資料切分下是否一致，降低結論對單次切分結果的依賴。三個階層評估類別的結果如表 5 所示。

表 5. 5-fold 交叉驗證之三類 Dice（1000 epochs，平均 $\pm$ 標準差）

版本	腎臟區域	腫塊	腫瘤	整體
UMambaBot（基準）	0.9661 $\pm$ 0.0076	0.8378 $\pm$ 0.0231	0.8059 $\pm$ 0.0188	0.8700 $\pm$ 0.0146
KalMed-UNet（本作品）	0.9807 $\pm$ 0.0074	0.8467 $\pm$ 0.0239	0.8165 $\pm$ 0.0242	0.8813 $\pm$ 0.0130
提升幅度	+0.0146	+0.0089	+0.0106	+0.0113

三個階層評估類別均獲得提升，腎臟區域、腫塊與腫瘤 Dice 分別增加 0.0146、0.0089 與 0.0106。此結果與 4.4 節下採樣模組在三類指標均改善的觀察相呼應，顯示所提設計的效益涵蓋腎臟整體區域及病灶相關目標，而非集中於單一類別。表 6 進一步列出逐 fold 的整體 Dice 與配對差值，五個 fold 的增益介於 0.0067 至 0.0149，平均為 0.0113，顯示平均提升並非由單一高分切分支撐。以五組差值計算的配對 t 檢定為 $t(4) = 7.50$ 、 $p = 0.0017$ ，作為本次結果的輔助描述；由於交叉驗證的訓練資料互有重疊，主要判斷仍建立在相同切分下的逐折正向差值及三類平均表現。這些結果支持最終整合架構在 KiTS23 與既定訓練流程下的效益，尚不延伸為每個病例或其他資料來源都必然改善。

表 6.各 fold 之整體 Dice 與配對差值（1000 epochs）

版本	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	平均
UMambaBot	0.8700	0.8693	0.8712	0.8927	0.8466	0.8700
KalMed-UNet（本作品）	0.8849	0.8828	0.8802	0.8994	0.8590	0.8813
差值	+0.0149	+0.0135	+0.0090	+0.0067	+0.0124	+0.0113

4.6 計算成本比較

表 7 比較各版本的參數量與計算量。其中僅有 KalMed-UNet 在參數量與計算量同時下降的情況下取得最佳效能，其餘卡爾曼變體皆是以增加計算成本換取效能提升。

表 7.參數量與計算量比較（Dice 為 250 epochs、單一 fold 之結果）

版本	參數量 (M)	GFLOPs	參數量變化	GFLOPs 變化	整體 Dice
基準模型（UMambaBot）	42.12	1939.68	—	—	0.8184
Kalman（S3）	42.23	1997.70	+0.26%	+2.99%	0.8222
Kalman + B	42.29	2026.71	+0.39%	+4.49%	0.8272
UK（S3 + S4）	42.26	2113.80	+0.33%	+8.98%	0.8241
KalmanDown	42.13	1941.30	+0.03%	+0.08%	0.8313
KalMed-UNet（本作品）	41.57	1852.71	-1.31%	-4.48%	0.8323

此一結果與 4.4 節的觀察相互呼應：效能提升並非來自模型容量的增加。相反地，計算量增加最多的 UK 版本（+8.98%）其效能反而低於幾乎不增加計算量的 KalmanDown（+0.08%）。從列示值計算，完整模型比基準少 0.55 M 參數及 86.97 GFLOPs，分別對應約 1.31% 與 4.48% 的減量。參數量反映模型權重規模，GFLOPs 反映既定輸入下的運算量；實際延遲及峰值顯存還與核心實作、記憶體存取及硬體有關。本文因此將效率結論限定於已量測的架構成本，並與相同設定的 Dice 一併呈現，不直接換算成加速倍數。

4.7 與既有方法之比較

本節整理 16 項外部方法與配置，涵蓋 U-Net、殘差網路、Transformer、Mamba、稀疏卷積及分割基礎模型；同一論文的不同配置以獨立列呈現，不視為不同研究。表 8 列出公開訓練集內的驗證結果，表 9 另列 GASA-UNet 研究的 89 例獨立保留集結果。編號 A01 - A10 與 B01 - B06 可對照 4.11 節的訓練及資源資料。外部數值均引自原文，並非本作品重新訓練；NR 表示原文未報告該欄位，表內保留來源的有效位數。由於切分、訓練預算及推論方式不同，以下比較用於說明效能背景與方法取捨，直接的架構增益仍由表 5 及表 6 的同設定對照支持。

表 8. KiTS23 公開集內驗證之文獻結果

方法與來源	腎臟區域	腫塊	腫瘤	整體	切分
A01 nnU-Net v2 [23]	0.9669	0.8366	0.7965	0.8666	隨機五折
A02 SegVol (零樣本) [23]	0.9140	0.2184	0.1953	0.4426	公開集評估
A03 Sparse U-Net [23]	0.9581	0.8568	0.8033	0.8728	隨機五折
A04 3D U-Net (全解析度) [24]	0.977	0.840	0.790	0.869	98 例驗證
A05 3D U-Net (低解析度) [24]	0.973	0.858	0.794	0.875	98 例驗證
A06 殘差 3D U-Net (全解析度) [24]	0.979	0.858	0.803	0.880	98 例驗證
A07 串接式 3D U-Net [24]	0.979	0.858	0.804	0.880	98 例驗證
A08 多尺度集成與後處理 [24]	0.979	0.857	0.826	0.887	98 例驗證
A09 SegResNet A [22]	NR	NR	NR	0.88924	隨機五折
A10 DiNTS [22]	NR	NR	NR	0.8767	隨機五折
UMambaBot (本作品重現)	0.9661	0.8378	0.8059	0.8700	相同五折
KalMed-UNet (本作品)	0.9807	0.8467	0.8165	0.8813	相同五折

表 9. KiTS23 獨立保留集之文獻結果 (89 例)

方法與來源 [27]	腎臟區域	腫塊	腫瘤	整體
B01 VNet	0.8280	0.5782	0.4723	0.6262
B02 UNETR	0.7301	0.5533	0.4978	0.5938
B03 nnMamba	0.8054	0.6363	0.4900	0.6439
B04 nnU-Net	0.8994	0.7279	0.6223	0.7499
B05 GASA-UNet	0.9121	0.7306	0.6363	0.7597
B06 GASA-UNet-L	0.9276	0.7423	0.6248	0.7649

表 8 中，本作品的 0.8813 位於殘差及串接式 U-Net 所報告的 0.880，與多尺度集成的 0.887 之間；這可說明其數值水準，卻不能據以宣稱跨研究勝出。文獻 [24] 的驗證僅含單次切分的 98 例，文獻 [22][23] 使用各自的隨機五折；SegVol 的結果則反映預訓練模型直接以文字提示進行零樣本分割，不是 KiTS23 專門訓練模型的上限。表 9 的研究先保留 89 例，其餘 400 例再作五折訓練並集成測試預測，因此其整體分數不能與本作品的五折驗證平均直接排名。另須區分，SegResNet A 為單組五折結果，並非冠軍十五模型集成的官方測試成績。就本作品而言，較具體的成果是在單模型、無 TTA 及後處理的既定流程中，三類 Dice 與五折整體表現均優於自行重現的基準，且模型成本同步降低。

4.8 分割結果視覺化

本節選取四個驗證病例，依序呈現 CT、標準答案、基準與本作品的分割，綠、紅、藍色分別表示腎實質、腫瘤及囊腫；影像裁切至前景附近，並採 $[-58, 302]$ HU 視窗。圖 5 與圖 6 呈現較明顯的病灶類別改善：case_00268 中，基準將病灶主體標為囊腫，本作品的腫瘤區域則較接近標準答案，病例三類平均 Dice 由 0.752 提升至 0.928；case_00555 中，基準的腫瘤標註較破碎且混有囊腫判斷，本作品形成較連續的腫瘤區域，分數由 0.771 提升至 0.929。這些切片展示了改善所在的位置，並與三類指標的變化相互補充；僅由輸出影像尚不能分離各融合元件的作用，或直接證明特定特徵在網路內被保留。

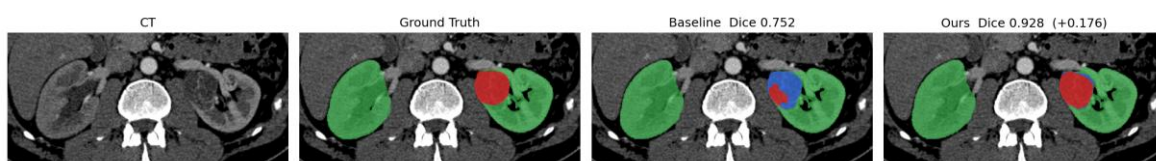


圖 5. case_00268 分割結果比較（第 79 張軸狀切片）

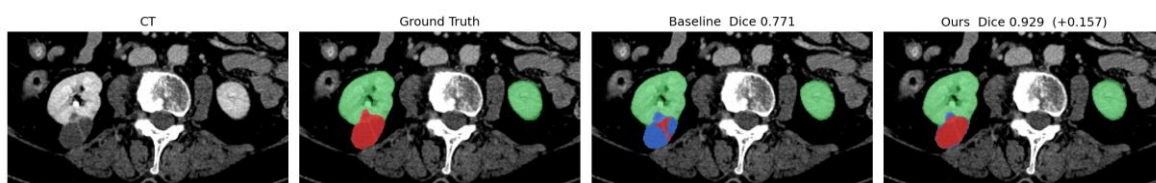


圖 6. case_00555 分割結果比較（第 185 張軸狀切片）

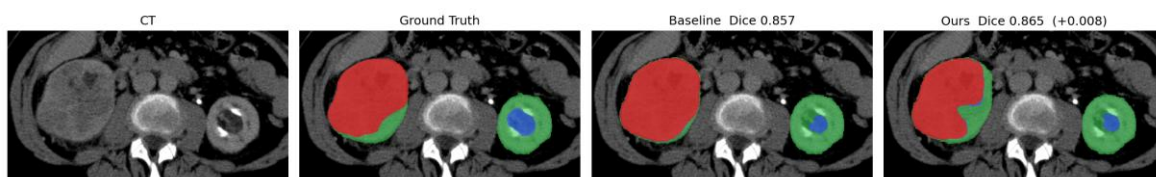


圖 7. case_00551 分割結果比較（第 34 張軸狀切片，本作品小幅領先）

圖 7 與圖 8 的兩模型表現較接近，更能呈現目前的誤差型態。case_00551 的病例三類平均 Dice 由 0.857 提升至 0.865，本作品保留了部分腫瘤與腎實質交界的正常組織，對側囊腫的範圍也較接近標準答案，但局部交界仍有偏差。case_00472 的分數由 0.911 提升至 0.920，兩模型皆能定位大型病灶，本作品使其中一處腎實質誤判區域縮小，仍未完全消除。這四例用於說明類別混淆、局部輪廓及病灶內部不一致等現象，並不代表每一病例皆有同等改善；整體效益仍以完整五折結果為依據。

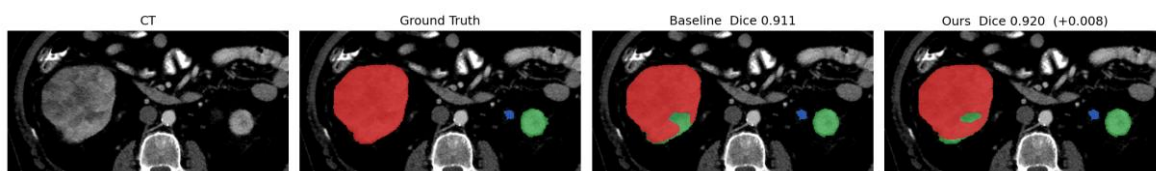


圖 8. case_00472 分割結果比較（第 66 張軸狀切片，本作品小幅領先）

4.9 限制與討論

本作品以 KiTS23 的同設定五折對照驗證最終整合架構，主要證據涵蓋三類區域重疊、逐折增益及模型成本。這些結果顯示改善在不同切分下維持一致方向，但病例中的局部組織混淆仍存在，如圖 7 與圖 8 所示。設計歷程使用單一 fold、250 epochs，適合呈現相同探索預算下的配置差異；各元件的獨立效果及版本排序在其他切分下是否相同，則未由本組結果確認。因此，本作品將結論集中於完整 KalMed-UNet 相對共同基準的效益，避免把整合改動拆解為尚未單獨量測的因果貢獻。在適用範圍上，目前結果來自單一公開資料集，主要指標為體積 Dice，尚未涵蓋跨中心影像或表面距離的量化評估。P 亦作為未經校準的融合代理量，Q 為可學習純量，增益式中的 R 固定為 1；這些設定提供明確且成本有限的實作，但不能直接解讀為對特徵誤差的機率估計。相對應地，本文已藉由公式範圍分析、張量介面與文獻對照說明其設計意義，並以實際五折對照支撐任務效用。未來的延伸可分別針對泛化、邊界品質或代理量的對應關係展開，使新增證據能回答具體問題。

4.10 既有結果的整合分析

本節以既有結果分析資料切分及評估目標的共同變化。圖 9 上半部列出五折整體 Dice，下半部呈現配對差值；兩模型最高分皆出現在 fold 4，最低分皆出現在 fold 5，顯示病例組成會共同影響其表現，而 KalMed-UNet 在這些起伏之上仍維持正向增益。各折改善介於 0.0067 至 0.0149，說明平均提升並非只來自單一容易的切分，也不表示所有病例的改善幅度相同。三類 Dice 則區分腎臟與病灶聯集、病灶聯集及腫瘤本身的表現；完整訓練下三類皆改善，與縮減預算中 KalMed-UNet 相對 KalmanDown 的類別取捨，屬於不同對照及訓練條件，應分別解讀。

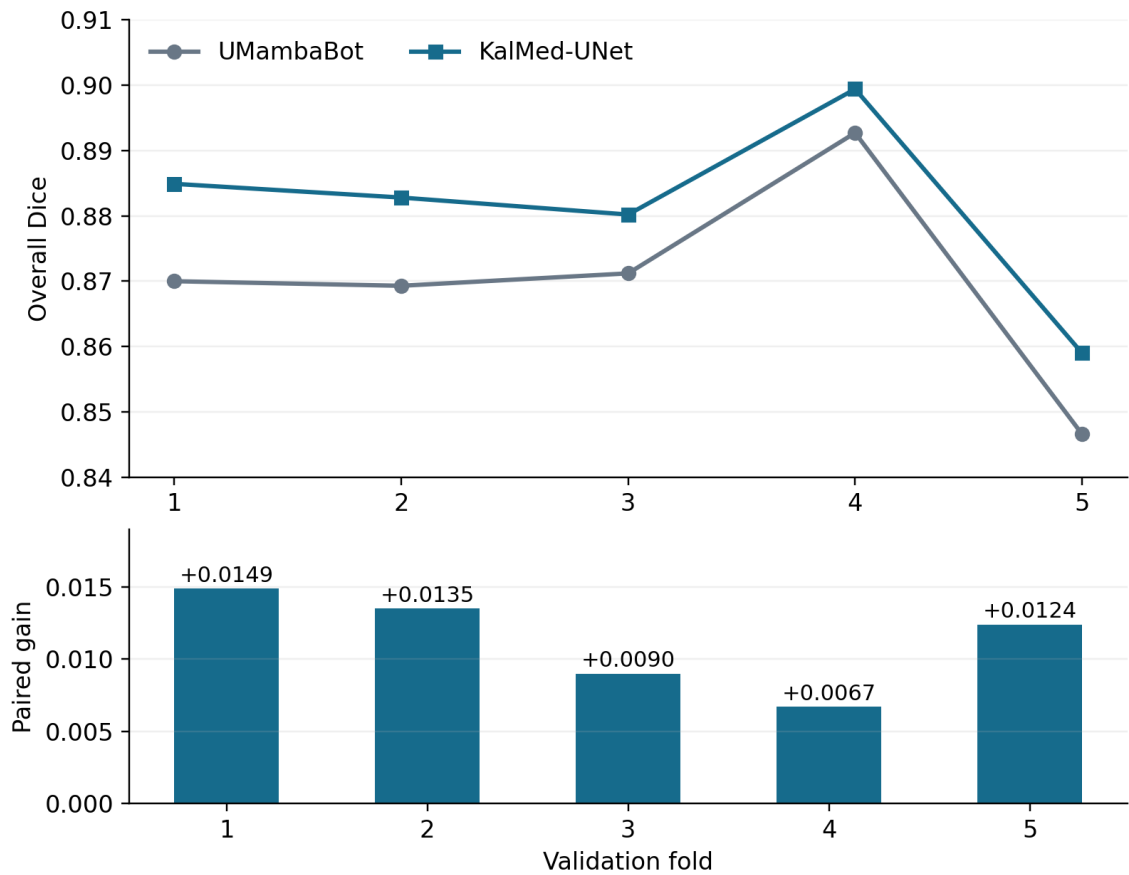


圖 9.五折整體 Dice 與配對增益（依表 6 繪製）

圖 10 將六個設計版本的 GFLOPs 與整體 Dice 同時呈現，數值均取自表 7。KalmanDown 以接近基準的計算量取得主要增益，完整 KalMed-UNet 則在本組比較中具有較低計算量與較高整體分數；增加較多計算的探索版本未得到最高分，顯示本次設計的收益不能僅以模型容量解釋。結合第三章的成本分析，可將完整模型的減量連結至解碼器上採樣的替換，而視覺案例則補充分數無法描述的類別誤判位置。設計歷程、五折對照、三類指標及成本量測因而形成不同面向的證據，共同支持本文對整合架構的定位：在維持既有深層建模及推論流程的條件下，以局部融合改動改善準確度與架構成本。

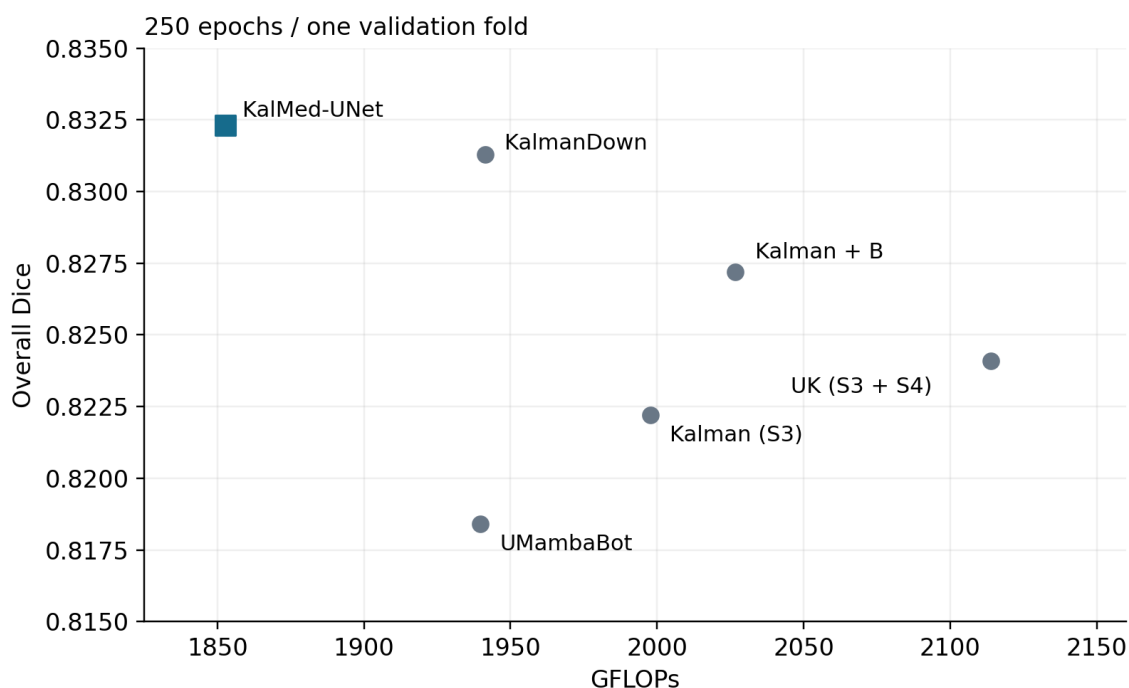


圖 10.設計版本的效能與計算量（依表 7 繪製）

4.11 訓練設定與計算資源比較

表 10 與表 11 分別整理公開集內驗證方法及獨立保留集方法的訓練設定，表 12 則彙整參數量與 GPU。同一來源中明確共用的設定合併列示，避免重複占用欄位；架構或解析度不同的配置仍保留編號，與分數表相互對照。這些資料顯示，各方法的訓練預算並不只由 epoch 數決定：Sparse U-Net 使用兩階段稀疏體積處理與梯度累積，Auto3DSeg 使用八張 A40 及較大的輸入區塊，GASA-UNet 研究則在推論使用鏡射及五模型集成。相較之下，本作品以單張 16 GB GPU、 128^3 patch 與單模型推論建立完整對照，提供了資源條件明確的實作結果。

表 10.公開集內驗證方法之訓練設定

方法編號	輸入與訓練預算	最佳化、損失與推論
A01 nnU-Net v2 [23]	128^3 ；batch 2；1000 epochs；間距 $1.0 \times 0.78 \times 0.78$ mm。	SGD／Nesterov；學習率及損失細項 NR；滑動視窗，無 TTA。
A02 SegVol [23][25]	使用公開預訓練權重；不在 KiTS23 微調；滑窗 $32 \times 256 \times 256$ 。	文字提示進行零樣本分割；zoom-out／zoom-in；無 KiTS23 訓練排程。
A03 Sparse U-Net [23]	兩階段全體積／ROI；間距 1.99^3 及 0.78^3 mm；batch 1，累積 8 次；500 epochs。	AdamW，lr 5×10^{-4} ，wd 10^{-5} ， $\beta=(0.9, 0.95)$ ；1 epoch 暖身、99 固定、400 cosine；Dice＋深度監督。
A04 全解析度 U-Net [24]	128^3 ；間距約 $0.78 \times 0.78 \times 1.0$ mm；batch、epochs NR。	SGD；CE＋Dice；旋轉、縮放與鏡射增強；lr NR。

方法編號	輸入與訓練預算	最佳化、損失與推論
A05 低解析度 U-Net [24]	128 ³ ；間距約 1.84×1.84×2.36 mm；batch、epochs NR。	最佳化、損失及增強同 A04；較粗間距使固定 patch 覆蓋更大物理範圍。
A06 殘差 U-Net [24]	全解析度殘差編碼器；128 ³ ；batch、epochs NR。	最佳化、損失及增強同 A04；以殘差結構調整特徵編碼。
A07 串接式 U-Net [24]	結合低解析度與高解析度階段；128 ³ ；batch、epochs NR。	最佳化、損失及增強同 A04；由前段結果提供後段分割資訊。
A08 多尺度集成 [24]	低解析度、低解析度殘差、全解析度 batch 4 三種模型；訓練 epochs NR。	結合三模型與多尺度預測，再進行連通域後處理；單一組件訓練原則同 A04。
A09 SegResNet A [22]	256 ³ ；5 層，基礎通道 32；每卡 batch 1，8 卡總 batch 8；完整 epochs NR。	AdamW，lr 2×10^{-4} ，wd 10^{-5} ，cosine；Dice+深度監督；ROI 裁切及後處理。
A10 DiNTS [22]	每卡 batch 1，8 卡；起始 patch 96 ³ ，後以 192 ³ 至 192×192×288 微調 25 epochs；初始 epochs NR。	SGD，lr 0.2；Dice+focal；混合精度；ROI 裁切及後處理。
本作品兩模型	128 ³ ；batch 2；1000 epochs×5 folds；探索為 250 epochs×1 fold。	SGD/Nesterov 0.99，lr 0.01，poly；Dice+CE，深度監督；無 TTA、集成及後處理。

表 11.獨立保留集方法之共同設定與架構差異

方法編號	架構與最佳化差異	全組共同設定 [27]
B01 VNet	三維卷積分割網路；初始 lr 0.01。	保留 89 例測試；其餘 400 例五折訓練；batch 2；1000 epochs，每 epoch 250 iterations。
B02 UNETR	Transformer 編碼器；初始 lr 0.01。	SGD/Nesterov，momentum 0.99；poly 排程指數 0.9；Dice+soft CE。
B03 nnMamba	Mamba 與卷積整合；初始 lr 0.0001。	nnMamba 的學習率為例外；其餘訓練程序依原文共同設定；未使用預訓練。
B04 nnU-Net	自動配置的 U 形網路；初始 lr 0.01。	前處理及資料增強沿用 nnU-Net 流程；KiTS23 patch 具體尺寸 NR。
B05 GASA-UNet	nnU-Net 加全域軸向自注意力；初始 lr 0.01。	滑動視窗重疊 50%，Gaussian 加權；各軸鏡射 TTA。
B06 GASA-UNet-L	較大版本的 GASA 架構；初始 lr 0.01。	五折模型機率平均後評估 89 例；不使用後處理。

表 12.外部方法與本作品之參數量及硬體資源

方法編號	參數量	訓練 GPU／時間	另行報告之推論硬體
A01 nnU-Net v2	31.2 M	NR	V100 16 GB；A100 40／80 GB [23]
A02 SegVol	180.9 M	KiTS23 零樣本，無此項訓練	V100 16 GB；A100 40／80 GB [23]

方法編號	參數量	訓練 GPU／時間	另行報告之推論硬體
A03 Sparse U-Net	27.47 M／網路； 兩階段	NR	V100 16 GB；A100 40／80 GB [23]
A04 全解析度 U-Net	NR	NR	NR
A05 低解析度 U-Net	NR	NR	NR
A06 殘差 U-Net	NR	NR	NR
A07 串接式 U-Net	NR	NR	NR
A08 多尺度集成	NR；三種配置組 成	NR	NR
A09 SegResNet A	NR	8×A40 48 GB；時間 NR [22]	未另列
A10 DiNTS	NR	8×A40 48 GB；時間 NR [22]	未另列
B01 VNet	KiTS23 配置 NR	A100 40 GB [27]	未另列
B02 UNETR	KiTS23 配置 NR	A100 40 GB [27]	未另列
B03 nnMamba	KiTS23 配置 NR	A100 40 GB [27]	未另列
B04 nnU-Net	KiTS23 配置 NR	A100 40 GB [27]	未另列
B05 GASA-UNet	KiTS23 配置 NR	A100 40 GB [27]	未另列
B06 GASA-UNet-L	KiTS23 配置 NR	A100 40 GB [27]	未另列
UMambaBot（本作品）	42.12 M	RTX 5080 16 GB；約 10 小時／fold	沿用本作品環境
KalMed-UNet（本作品）	41.57 M	RTX 5080 16 GB；約 10 小時／fold	沿用本作品環境

表 12 的 NR 表示來源未提供可對應至該 KiTS23 配置的數值。文獻[27]另列 nnU-Net 與 GASA-UNet 的 29.54 M、30.55 M 參數，但該表針對 BTCV 配置，故未移入 KiTS23 欄位；其約 20 小時／fold 亦為文中整體訓練敘述，不用來推算各架構相對速度。文獻[23]的 V100 及 A100 為推論基準測試硬體，並未明確指定為訓練 GPU。這些區別提醒讀者，單卡或多卡、模型大小、訓練時間與推論成本是相關但不同的量。對本作品最有依據的成本結論，仍是相同輸入與量測條件下的參數減少 1.31% 及 GFLOPs 減少 4.48%；跨論文資源表則用於說明實作條件與資訊完整程度。

4.12 相近卡爾曼方法之設計對照

表 13 進一步列出與本作品理論來源相近的 KaLDeX 及 KLDD，補充分數表以 KiTS23 為主的比較範圍。兩者以二維視網膜血管分割驗證卡爾曼啟發式取樣設計，本作品則處理三維 CT 的跨解析度特徵融合。這項對照的目的在於具體區分「濾波結構用在何處」：前兩者修正可變形卷積的空間取樣，KalMed-UNet 則對已對齊的分支特徵作殘差更新，再把 P 提供給後續層級及解碼器。相同理論來源因而可形成不同的網路介面、訓練需求及應用對象。

表 13.卡爾曼啟發式分割方法之設計與實驗條件

比較項目	KaLDeX [28]	KLDD [29]	KalMed-UNet
處理對象	二維視網膜血管	二維視網膜血管	三維腎臟、腫瘤及囊腫
主要機制	線性可變形卷積＋交叉注意力	線性可變形特徵擷取＋條件擴散	雙分支下採樣＋跨層 P＋解碼器融合
卡爾曼修正對象	取樣座標	取樣座標	特徵與融合代理量
GPU	RTX A5000 24 GB	RTX A5000 24 GB	RTX 5080 16 GB
最佳化／學習率	最佳化器 NR；lr 0.0025	Adam；lr 0.0001；wd 10^{-5}	SGD／Nesterov；lr 0.01；poly
訓練長度	10 epochs	100 epochs	1000 epochs／fold
輸入與 batch	具體尺寸及 batch NR	256×256；batch NR	128 ³ ；batch 2
損失	clDice＋BCE	去雜訊損失＋clDice	Dice＋CE；深度監督
參數量	NR	NR	41.57 M
原文 Dice	DRIVE：0.8287；CHASE：0.8172	DRIVE：0.8365；CHASE：0.8385	KiTS23 五折：0.8813

表 13 的分數各自對應其資料集，不能由數值大小推論三者的分割能力高低。較有意義的共同點，是把一般的特徵轉換加上明確的狀態更新規則，使模組不只依固定卷積產生輸出；差異則在更新所使用的資訊與後續用途。KalMed-UNet 的 P 為單通道空間場，可在不同通道數的層級間傳遞，並與 skip 特徵一同參與解碼，形成本文特有的跨解析度介面。結合五折對照可知，這個整合配置在本作品的訓練與推論條件下帶來穩定方向的增益。本文因此以模組的具體作用及可重現的實驗條件界定成果，不以卡爾曼名稱本身或跨資料集的高分作為優越性依據。

4.13 文獻內部對照與研究定位

在相同來源內觀察配置差異，比跨論文直接排序更能提供設計上的參考。文獻[24]的全解析度 U-Net 整體 Dice 為 0.869，低解析度配置為 0.875，顯示提高解析度本身並不保證整體分數較高；全解析度加入殘差及串接式配置皆為 0.880，多尺度集成與後處理則達 0.887。這些差異反映解析度、架構及推論組合的共同作用，並不能拆成單一因素的因果效果。文獻[27]中，nnU-Net、GASA-UNet 與 GASA-UNet-L 的整體分數依序為 0.7499、0.7597 及 0.7649，但較大版本的腫瘤 Dice 為 0.6248，低於 GASA-UNet 的 0.6363。此例說明整體平均與單類目標可能有不同的最佳配置，與本作品在縮減預算下觀察到的類別取舍相呼應。設計評估因此需要同時保留三類指標，才能看出改善發生在哪個目標。從工程條件來看，集成與多階段流程可以綜合不同模型或解析度的預測，但通常需要額外的模型管理與推論步驟；預訓練模型減少目標資料集的訓練需求，其輸出則取決於原有表示與提示是否適合任務。KalMed-UNet 選擇在單一骨幹內調整前段下採樣及對應解碼器，讓額外資料流限制在少數解析度層級，並以相同五折安排檢驗最終效益。其 0.8813 不代表已超越各研究的完整系統，但同設定下增加 1.13 個百分點、五折皆改善，以及參數量和 GFLOPs 同時降低，具有直接且一致的實驗支持。文獻比較因而補足兩個面向：一方面說明本作品所處的分數與資源條件，另一方面指出不同研究透過架構、解析度、預訓練或推論組合取得效益的方式。本文的特色在於以跨解析度融合為主要改動單位，在既定訓練與推論流程內取得可量測的改善。

4.14 標籤混淆與階層評估的關係

KiTS23 的三個階層集合具有包含關係，因而同一次標籤誤判可能在不同指標中呈現不同影響。表 14 依標準答案與預測標籤的轉換，整理該體素在腎臟區域、腫塊及腫瘤集合中的分類狀態；FP 表示原本不屬於該集合卻被納入，FN 表示原本屬於該集合卻被排除，「不變」則表示兩種標籤在該集合中具有相同的成員身分。此表由 HEC 定義推得，並非本作品驗證集的混淆矩陣，也不代表各種錯誤的實際出現頻率。以腫瘤被判為囊腫為例，腎臟區域與腫塊的聯集仍包含該位置，只有腫瘤集合出現遺漏；若腫瘤被判為背景，三個集合便同時遺漏。因而只看整體聯集輪廓，可能無法發現病灶內部的類別混淆，而三類指標一起報告，才較能辨識輪廓涵蓋與病灶分類的差別。

表 14.標籤轉換對三個階層評估集合的影響

標準答案 → 預測	腎臟區域	腫塊	腫瘤
腫瘤 → 囊腫	不變	不變	FN
囊腫 → 腫瘤	不變	不變	FP
腫瘤 → 腎實質	不變	FN	FN
腎實質 → 腫瘤	不變	FP	FP
囊腫 → 腎實質	不變	FN	不變
腎實質 → 囊腫	不變	FP	不變
腫瘤 → 背景	FN	FN	FN
背景 → 腫瘤	FP	FP	FP
囊腫 → 背景	FN	FN	不變
背景 → 囊腫	FP	FP	不變
腎實質 → 背景	FN	不變	不變
背景 → 腎實質	FP	不變	不變

這個集合關係也提供了閱讀圖 5 至圖 8 的方式。圖 5 與圖 6 中，病灶主體在腫瘤與囊腫之間的判斷差異，應連結至腫瘤項目的變化；圖 7 與圖 8 則還涉及腎實質與病灶之間的局部混淆，可能同時改變腫塊及腫瘤集合。由於圖中只有選定切片，且其上方標示的是病例三類平均分數，不能由單張切片重新算出該病例的完整體積指標，也不能只靠顏色面積推定三類改善的比例。本文採用的互補安排，是由五折結果說明整體效益，再以案例呈現錯誤的位置與型態。表 14 補充兩者之間的邏輯連結，使讀者能理解何以某些類別修正對整體腎臟輪廓的影響不大，卻仍可能對病灶辨識具有重要意義。對本作品三類平均皆提升的結果，也應依此理解為多個階層目標的改善，而非三份彼此獨立的驗證證據。

4.15 推論流程與成本比較的範圍

模型的單次前向計算量與整個病例的推論成本並不是同一個量。使用滑動視窗時，一個病例通常需要多個區塊的預測，再將重疊位置的結果整合；即使兩個網路使用相同 patch，其病例尺寸、滑窗重疊及執行方式也可能使前向次數不同。若另外採鏡射增強，每個區塊會接受不同方向的變換，再把結果轉回共同座標進行彙整；若使用多模型集成，還需要執行多套權重。作為流程計數的示例，假設所有條件相同，每個區塊使用 M 個模型及 T 種變換，則需要 $M \times T$ 組模型與變換的前向評估。這只是評估次數的乘法關係，並不是實測時間的加速或減速倍率，因批

次化、資料傳輸與硬體利用率都會改變實際耗時。本文在 4.6 節報告的 GFLOPs 對應固定輸入下的網路成本，不能直接取代完整病例的延遲。不同文獻的效率路線也因此需要放回完整流程理解。Sparse U-Net[23]先找出區域，再於高解析度區域內分割，計算需求受到區域大小及活躍體素影響；Auto3DSeg[22]與多尺度流程[24]則使用區域處理、模型組合或後處理，GASA-UNet 研究[27]在獨立保留集採用鏡射及五折模型集成。本作品的五折平均是各折驗證病例的結果彙整，每個病例使用其對應的單一模型，與五模型共同預測同一病例不同。因此，本文所強調的實作特點是在既定單模型流程內取得基準以上的表現，而不是以目前資料宣稱完整系統的推論速度領先所有文獻。將訓練 GPU、輸入尺寸、模型數量與推論處理一併列出，可使比較的條件清楚，也讓參數量減少 1.31% 及 GFLOPs 減少 4.48% 的意義保持精確：它們是相同設定下的架構減量，並未包含尚未量測的部署收益。

4.16 訓練預算與比較單位

訓練長度除了 epoch 數，還需考慮每輪迭代、batch size 及梯度累積。若一個訓練流程共有 E 個 epochs，每輪固定 I 次參數更新，則更新次數為 $E \times I$ ；文獻[27]所列 1000 epochs、每輪 250 iterations，即對應單折 250000 次更新。若另一流程以固定數量的區塊取樣定義每輪，便不能直接把 epoch 解讀為每個病例恰好出現一次。表 10 中的 Sparse U-Net 採 batch 1、累積 8 次梯度，其一次更新整合的樣本數與單步 batch 1 不同；這仍不保證等同於一次放入 batch 8 的所有運算，因損失的彙整與正規化方式亦有影響。因此，本文在外部比較中保留來源明列的設定，並將缺少的項目標示為 NR。相同的 epoch 名稱可以作為訓練長度的描述，但不足以單獨代表不同網路所使用的總運算資源。本作品的兩層評估安排也應依比較單位解讀。設計探索在單一 fold 使用 250 epochs，各版本面對共同的縮減預算，適合整理配置演進；主要結果則對每個最終模型訓練五折，每折 1000 epochs。若將各 fold 的訓練輪次加總，單一最終模型配置的五折排程合計為 5000 輪，相對單一探索配置的 250 輪為二十倍；其中同一折的訓練長度增加四倍，另外擴大至五組資料切分。這個計數不等於實測時間倍率，也不代表五個模型會在推論時同時使用。它說明兩組結果回答的問題不同：探索結果提供設計選擇的依據，完整訓練則檢查基準與最終架構的差異能否在多個切分下維持。將兩組分數分表呈現，既保

留設計歷程，也使 0.8813 的主要結果明確對應完整五折程序，避免把較高訓練預算本身誤列為某個模組的獨立效益。

第五章 結論與未來發展

5.1 結論

本作品提出 KalMed-UNet，以卡爾曼濾波的預測—更新結構啟發三維分割網路的雙分支下採樣融合。模組利用殘差與不確定度代理量 P 調節融合比例，並將 P 跨解析度傳遞至解碼器，進行受後向平滑啟發的特徵融合。研究將編碼與解碼階段的局部改動整合為可訓練的架構，並以共同基底與一致的實驗流程評估其整體效益。於 KiTS23 進行 1000 epochs、五折交叉驗證，整體 Dice 由 0.8700 提升至 0.8813，增加 1.13 個百分點；腎臟區域、腫塊與腫瘤分別由 0.9661、0.8378、0.8059 提升至 0.9807、0.8467、0.8165，五個 fold 全數改善。參數量與計算量分別減少 1.31% 及 4.48%，顯示準確度與架構成本可同時改善。縮減預算的設計比較交代由 KalmanDown 至完整模型的演進，文獻整理則提供不同方法的分數、訓練與資源背景；主要結論仍以本作品相同設定下的五折基準對照為依據。

5.2 未來發展

後續研究可沿資料適用範圍與融合機制兩個方向延伸。前者將方法應用於其他器官、影像模態或跨中心資料，以檢視同一融合設計在不同條件下的表現；後者則分析放置層級、 Q 的輸入相依設計及增益尺度，進一步釐清代理量與分割結果的關係。這些方向各自回答不同問題：跨資料驗證關心方法是否能轉移，機制分析則關心哪些設計形成觀察到的增益。現有結果已完成共同骨幹下的整合架構評估，可作為此類延伸的實驗基礎。在評估與推論方面，可加入表面 Dice 等邊界指標，並分析測試時增強、集成及後處理的額外成本與效益。本作品已在單模型推論及既有深層建模方式下，透過下採樣與解碼器融合改動，同時改善 Dice 與架構成本。以資料流與解析度轉換為改進單位，因而提供一個可控制實作範圍、兼顧分割品質及資源需求的研究方向。

參考文獻

- [1]. O. Ronneberger、P. Fischer、T. Brox，「U-Net: Convolutional Networks for Biomedical Image Segmentation」，MICCAI，2015 年。
- [2]. Ö. Çiçek、A. Abdulkadir、S. S. Lienkamp、T. Brox、O. Ronneberger，「3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation」，MICCAI，2016 年。
- [3]. F. Milletari、N. Navab、S.-A. Ahmadi，「V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation」，3DV，2016 年。
- [4]. F. Isensee、P. F. Jaeger、S. A. A. Kohl、J. Petersen、K. H. Maier-Hein，「nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation」，Nature Methods，第 18 卷，第 203 – 211 頁，2021 年。
- [5]. R. Zhang，「Making Convolutional Networks Shift-Invariant Again」，ICML，2019 年。
- [6]. A. Azulay、Y. Weiss，「Why Do Deep Convolutional Networks Generalize so Poorly to Small Image Transformations?」，Journal of Machine Learning Research，第 20 卷，2019 年。
- [7]. X. Zou、F. Xiao、Z. Yu、Y. J. Lee，「Delving Deeper into Anti-aliasing in ConvNets」，BMVC，2020 年。
- [8]. A. Chaman、I. Dokmanić，「Truly Shift-Invariant Convolutional Neural Networks」，CVPR，2021 年。
- [9]. R. A. Rojas-Gomez、T.-Y. Lim、A. Schwing、M. Do、R. Yeh，「Learnable Polyphase Sampling for Shift Invariant and Equivariant Convolutional Networks」，NeurIPS，2022 年。
- [10]. R. Riad、O. Teboul、D. Grangier、N. Zeghidour，「Learning Strides in Convolutional Neural Networks」，ICLR，2022 年。
- [11]. A. Gu、K. Goel、C. Ré，「Efficiently Modeling Long Sequences with Structured State Spaces」，ICLR，2022 年。
- [12]. A. Gu、T. Dao，「Mamba: Linear-Time Sequence Modeling with Selective State Spaces」，arXiv:2312.00752，2023 年 12 月。
- [13]. J. Ma、F. Li、B. Wang，「U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation」，arXiv:2401.04722，2024 年 1 月。
- [14]. Z. Xing、T. Ye、Y. Yang、G. Liu、L. Zhu，「SegMamba: Long-range Sequential Modeling Mamba for 3D Medical Image Segmentation」，MICCAI，2024 年。
- [15]. J. Liu、H. Yang、H.-Y. Zhou 等，「Swin-UMamba: Mamba-based UNet with ImageNet-based Pretraining」，MICCAI，第 615 – 625 頁，2024 年。
- [16]. J. Ruan、J. Li、S. Xiang，「VM-UNet: Vision Mamba UNet for Medical Image Segmentation」，arXiv:2402.02491，2024 年 2 月。

- [17]. R. E. Kalman, 「A New Approach to Linear Filtering and Prediction Problems」, Journal of Basic Engineering, 第 82 卷, 第 35 – 45 頁, 1960 年 3 月。
- [18]. H. E. Rauch、F. Tung、C. T. Striebel, 「Maximum Likelihood Estimates of Linear Dynamic Systems」, AIAA Journal, 第 3 卷第 8 期, 第 1445 – 1450 頁, 1965 年 8 月。
- [19]. G. Revach、N. Shlezinger、X. Ni、A. L. Escoriza、R. J. G. van Sloun、Y. C. Eldar, 「KalmanNet: Neural Network Aided Kalman Filtering for Partially Known Dynamics」, IEEE Transactions on Signal Processing, 第 70 卷, 第 1532 – 1547 頁, 2022 年。
- [20]. N. Heller、F. Isensee、K. H. Maier-Hein 等, 「The State of the Art in Kidney and Kidney Tumor Segmentation in Contrast-Enhanced CT Imaging: Results of the KiTS19 Challenge」, Medical Image Analysis, 第 67 卷, 101821, 2021 年。
- [21]. N. Heller、A. Wood、F. Isensee 等, 「The 2023 Kidney and Kidney Tumor Segmentation Challenge (KiTS23)」, <https://kits-challenge.org/kits23/>, 2023 年。
- [22]. A. Myronenko、D. Yang、Y. He、D. Xu, 「Automated 3D Segmentation of Kidneys and Tumors in MICCAI KiTS 2023 Challenge」, Lecture Notes in Computer Science, 第 14540 卷, 2024 年。 arXiv:2310.04110。
<https://arxiv.org/abs/2310.04110>
- [23]. S. Alonso-Monsalve、L. H. Whitehead、A. Aurisano、L. Escudero Sanchez, 「Submanifold Sparse Convolutional Networks for Automated 3D Segmentation of Kidneys and Kidney Tumours in Computed Tomography」, Scientific Reports, 第 16 卷, 23089, 2026 年 6 月。 doi:10.1038/s41598-026-51801-7。
- [24]. K.-H. Uhm、H. Cho、Z. Xu、S. Lim、S.-W. Jung、S.-H. Hong、S.-J. Ko, 「Exploring 3D U-Net Training Configurations and Post-processing Strategies for the MICCAI 2023 Kidney and Tumor Segmentation Challenge」, Lecture Notes in Computer Science, 第 14540 卷, 2024 年。 arXiv:2312.05528。
<https://arxiv.org/abs/2312.05528>
- [25]. Y. Du、F. Bai、T. Huang、B. Zhao, 「SegVol: Universal and Interactive Volumetric Medical Image Segmentation」, NeurIPS, 2024 年。
- [26]. E. Matsinos, 「The Kalman Filter: a didactical overview」, arXiv:1607.05590, 2016 年。
- [27]. C. Sun、R. S. Terry、J. Bian、J. Xu, 「GASA-UNet: Global Axial Self-Attention U-Net for 3D Medical Image Segmentation」, arXiv:2409.13146, 2024 年。本文比較採原文及 Technical Appendix。
<https://arxiv.org/abs/2409.13146>
- [28]. Z. Zhao、Y. Zhao、J. Yang、Q. Liang、D. Zapp、K. Huang、M. A. Nasseri, 「Kalman Filter Based Linear Deformable for Retinal Vessel Segmentation」, arXiv:2410.21160v2, 2026 年 4 月。
<https://arxiv.org/abs/2410.21160>
- [29]. Z. Zhao、Y. Zhao、J. Yang、K. Huang、N. Navab、M. A. Nasseri, 「KLDD: Kalman Filter based Linear Deformable Diffusion Model in Retinal Image Segmentation」, BIBM, 2024 年。 arXiv:2410.02808。
<https://arxiv.org/abs/2410.02808>