

逢 甲 大 學

資 訊 工 程 學 系

專 題 研 究 報 告

輕量化卷積神經網路進行車道偵測

指導教授：張哲誠 老師

學生：葉秋彤 D1050957

林郁倫 D1051089

王岳華 D1051415

中 華 民 國 一 百 一 十 三 年 十 二 月

致謝

本次專題研究能夠順利完成，得益於師長們的悉心指導、親友的支持與學校的鼎力相助，在此致上最誠摯的感謝。

衷心感謝張哲誠副教授在本專題研究中的悉心指導與不懈支持。研究主題的選定、研究方向的指引和成果的評估，老師在每個環節都提供了寶貴的建議和啟發。研究過程中，當我們遇到技術瓶頸或理論難題時，總有老師的耐心解答，幫助我們加深對專業知識的理解，並引導我們找到可行的解決方案，使我們能夠快速突破並持續前行。老師的支持和肯定，不僅讓我們在學術路上更加自信，也讓本專題研究得以如期完成。這段經驗將成為我們未來學術發展的堅實基礎，深深影響著我們在人工智慧領域的學習與探索。

同時，也感謝在課程中悉心指導的各位師長，特別是在人工智慧領域授課的老師們，讓我們在專題研究中能夠靈活運用課堂中所學的知識，並提升了我們分析與解決問題的能力，讓我們能夠很好的面對挑戰。這些能力得以讓我們奠定了良好的基礎，也促使我們能夠更加順利地完成整個專題研究。

接著，感謝系辦助教們協助處理專題研究的申請流程與其他繁瑣事務，並提供良好的研究環境，使我們得以全心投入研究。此外，也感謝逢甲大學在整個大學期間提供優良的學術環境與研究空間，並配置完善的設備，讓我們能在舒適的環境中進行討論與研究，提升了學習的效率與品質。

最後，也要感謝所有在研究過程中給予幫助的親友。父母的無私支持是我們能夠全力以赴的強大後盾，他們的理解與陪伴讓我們能夠專心地投入研究中，並在面對壓力時能夠感受到溫暖和力量。再者，感謝朋友在我們書寫書面報告時分享經驗，提供我們格式參考的依據，不必多花時間自己慢慢嘗試。朋友們無論是在學術上的討論，抑或是情感上的支援，都給了我們堅持與前進的動力。這些支持與鼓勵都讓我們在艱難時刻能夠堅定前行。

再次感謝所有在研究和成長路上陪伴我們、支持我們的人，感謝他們無私的幫助與支持，這一切使我們的研究歷程變得更加豐富且意義非凡。

摘要

隨著自駕車技術的逐漸崛起，車輛能否適應不同環境中的自動導航至關重要，其中車道線偵測作為自駕車系統的核心任務之一，直接影響行車安全和準確性。在使用傳統的卷積神經網路(CNN)進行車道偵測時，往往需要大量的計算資源，這對於嵌入式系統和低階設備來說，是個需要克服的挑戰，因此，我們計畫實現一個輕量化模型，使其在推理階段僅需少量的浮點數運算量(FLOPs)即可進行車道線偵測。

在執行階段時，為了加快模型學習特徵的速度與效率，並減輕模型運算量，通過文獻探討和實驗測試，我們採用 Robust Lane Detection from Continuous Driving Scenes Using Deep Neural Networks[1]這篇論文的模型架構作為原模型。

在模型選定好之後，我們保留其 U-Net 的架構，因為 U-Net 在語義分割任務上的表現十分不錯。其編碼器(Encoder)-解碼器(Decoder)架構中的下採樣層(如池化層)能有效減少空間分辨率並增加特徵數量，上採樣層(如反卷積)則用於恢復空間分辨率。另外，U-Net 的對稱結構通過跳躍連接將編碼器層的特徵直接傳遞到對應的解碼器層，並在解碼過程中融合不同層次的特徵，做到保留原有特徵的同時也結合細節特徵。我們在相同的 CNN 的設計下，嘗試用不同的時序性輸入方式進一步提高模型在偵測車道線的準確度，我們測試了長短期記憶(LSTM)和時間偏移模組(TSM)，從實驗結果顯示，TSM 搭配 4 層的 encoder CNN、decoder CNN 的準確度最高，效能也最好。

關鍵字：車道線偵測、卷積神經網路、語義分割、LSTM、TSM

目錄

第一章 緒論.....	1
1.1 研究動機.....	1
1.2 研究目的.....	1
1.3 研究方法.....	1
1.3.1 研究工具.....	1
1.3.2 模型訓練.....	2
1.3.3 結果驗證與數據比較.....	2
1.4 研究流程.....	2
1.4.1 閱讀論文與決定研究方向.....	2
1.4.2 研究 CNN 與 RNN 的模型架構.....	2
1.4.3 優化並訓練模型.....	2
1.4.4 數據比較.....	3
1.4.5 決定模型.....	3
1.4.6 撰寫報告.....	3
第二章 文獻回顧.....	4
2.1 Robust Lane Detection[1]	4
2.2 資料集.....	4
2.2.1 nuScenes	4
2.2.2 Tusimple.....	5
2.3 結構相似性(structural similarity, SSIM)	5
2.4 LSTM (Long Short-Term Memory).....	6
2.5 TSM (Temporal Shift Module)	6

第三章 研究方法.....	7
3.1 資料蒐集.....	7
3.2 資料集預處理.....	7
3.3 評估指標說明.....	7
3.3.1 PSNR.....	8
3.3.2 SSIM	8
3.4 模型訓練.....	8
3.5 模型比較.....	10
第四章 結果分析.....	11
第五章 結論.....	14
參考文獻.....	15

圖目錄

圖 3.1 模型架構圖.....	9
圖 4.1 Ground Truth-影像一	12
圖 4.2 TSM_4 結果-影像一	12
圖 4.3 Ground Truth-影像二	12
圖 4.4 TSM_4 結果-影像二	12
圖 4.5 Ground Truth-影像三	12
圖 4.6 TSM_4 結果-影像三	12
圖 4.7 Ground Truth-影像四	12
圖 4.8 TSM_4 結果-影像四	12
圖 4.9 Ground Truth-影像五	12
圖 4.10 TSM_4 結果-影像五	12

表目錄

表 3.1 LSTM 與 TSM 搭配不同層數 encoder 及 decoder 之結果比較表	10
--	----

第一章 緒論

1.1 研究動機

身在這個交通便利、科技發達的時代，我們的駕駛體驗正在經歷顯著的轉變。傳統上，駕駛者主要依賴自身的記憶與判斷駕駛車輛。如今，因為科技的進步，我們可以透過先進技術的輔助，獲得到更多元的功能及行車建議，進而有效降低駕駛的危險性。而輔助駕駛技術的快速發展在這一轉變中扮演了關鍵角色，它不僅能提升行車安全，還具備降低交通事故發生率。

在台灣，交通事故是備受關注的社會議題。根據內政部警政署的統計資料，在 A1 類道路交通事故中，「駕駛人未注意車前狀態」是多數人的肇事原因。這說明在行車過程中，有許多人未能充分留意道路狀況。此種未注意車前狀態的行為可能是由於駕駛人分心、疲勞、使用手機或其他影響造成的，因此突顯了提高駕駛人對路況注意力和專注力的重要性。然而，長時間保持高度專注並非易事，自動駕駛技術的崛起將會減輕駕駛人的負擔並提高行車安全。

1.2 研究目的

使用傳統的卷積神經網路(CNN)進行車道偵測時，雖然能提供良好的準確性，但通常伴隨著大量的計算資源，這對於嵌入式系統或低階設備來說是一個巨大的挑戰。這些設備通常計算能力有限，因此在此類環境下部署高效且準確的模型成為一個必須解決的問題。

因此，本研究決定設計一個輕量化模型，使其在推理階段僅需使用較少的浮點數運算量(FLOPs)，便能有效完成車道線偵測工作。我們計畫透過結合多種技術，來大幅降低模型的可訓練參數量與計算需求，同時保持或提升其偵測準確性和效能並透過實驗比較不同方法的效果。我們預期該輕量化模型不僅適用於嵌入式環境，還能有效應用於各類計算資源受限的場景，為自動駕駛技術提供一個高效且實用的解決方案。

1.3 研究方法

1.3.1 研究工具

使用 Python 程式語言和 Google Colab 平台進程式編寫，而模型開發則使用 PyTorch 開源機器學習庫。

1.3.2 模型訓練

- A. 模型選擇：依研究需求研究、分析後挑選合適的模型。
- B. 資料準備：對資料進行預處理，包括格式轉換、正規化、增強等。
- C. 模型訓練：更改超參數調整模型的訓練。
- D. 模型評估：以不同評估指標針對模型的性能進行評估。
- E. 模型優化：根據評估結果對模型進行調整及優化。

1.3.3 結果驗證與數據比較

確保研究成果的準確度與可用性，對資料進行測試和比較。以下是一些結果驗證與應用的測試方式：

- A. 使用未被訓練的資料：同樣使用原模型之資料集進行測試。
- B. 文獻回顧和比較：若前一點的結果皆符合預期，則與原模型之測試結果進行比較。

1.4 研究流程

1.4.1 閱讀論文與決定研究方向

在研究初期，首先需要確認可行的研究方法，因此我們閱讀了以不同技術偵側車道線的論文，最終決定重點關注 CNN 模型、RNN 模型、輕量化技術的設計方法、深度學習在車道線偵測中的應用等方向，在多方比較與嘗試之後，確定研究方向並著手研究不同的模型架構。

1.4.2 研究 CNN 與 RNN 的模型架構

進行文獻探討並分析現有關於 CNN 和 RNN 模型的優缺點。接著，針對 CNN 模型探究卷積層設計、池化層配置與激活函數的選擇，衡量在減少計算量時對準確性的影響。在 RNN 方面，則會著重於分析長短期記憶網路 (LSTM) 的時序處理能力，尤其是時間上連續的車道線影像的處理能力。

1.4.3 優化並訓練模型

在模型設計方面，本研究以輕量化作為核心目標，設計一個高效、資源節省且準確度良好的車道偵測模型。本研究將嘗試構建並訓練多種不同的模型架構組合，從原模型開始，除了增加或減少 encoder CNN 及 decoder CNN 層數外，也更換 RNN 神經網路的種類、技術，並透過不斷調整模型的結構，持續優化模型性能，達成平衡準確度與計算效能的目標。

1.4.4 數據比較

分析使用 RNN 神經網路建構的模型成效，根據 FLOPs、可訓練參數量 (trainable parameters)、準確度等指標，量化並比較各模型的效率與效能。最後，選擇最符合輕量化需求且能維持高準確度的模型，並計算不同模型預測結果的結構相似性。

1.4.5 決定模型

綜合上述提到的 FLOPs、trainable parameters、準確性等指標，選擇表現最佳的模型架構作為最終成果。

1.4.6 撰寫報告

將研究過程、數據、成果寫成書面報告，製作介紹短片與海報。

第二章 文獻回顧

2.1 Robust Lane Detection[1]

傳統的車道偵測方法在特定環境條件下可以達到良好的準確度，但在遇到陰影、霧霾或障礙物遮蔽等複雜場景時，準確度往往不如預期，使得在實際應用中穩定性和可靠性大打折扣。根據 Zou 等人提出了一種結合 CNN 與 RNN 的車道偵測方法，解決傳統方法在複雜駕駛環境下的不足，這種方法利用 U-net 深度學習模型架構提取車道影像的空間特徵，再引入 LSTM 模型從多幀連續影像中提取時序資訊，提高了車道偵測的準確性和穩定性。從他們的實驗結果中，驗證了該方法在影像重疊區域中有效提取車道線資訊的能力，特別是在駕駛場景的時間序列預測框架中取得了良好的效果。

此外，Zou 等人也提出多種車道偵測技術，如 encoder-decoder CNN、FCN、CNN+RNN 以及 GAN 模型，而最後選定以 CNN+RNN 的方法進行車道偵測模型的設計。在 CNN 的部分選用 U-net 模型，因其 encoder-decoder 的架構可使輸入影像特徵大小與輸出影像特徵大小相同，其中 encoder 部分負責將每一幀影像抽象為低維度的特徵圖，保留了重要的空間資訊，接著，這些連續影像的特徵圖會被傳入 RNN 的模型中，也就是在 LSTM 進行時間序列特徵的提取，從而捕捉影像幀之間的時序關聯性，最後 decoder 會將這些提取出的特徵進行上採樣，生成與輸入影像特徵相同尺寸的車道線預測結果。

2.2 資料集

2.2.1 nuScenes

在自動駕駛領域，資料集的多樣性對於模型的訓練和性能提升來說至關重要。而近年來，隨著人工智慧技術的進步，相關領域的研究不再單純依賴單一類型 sensor 的資料，而是開始使用多種 sensor 的資料來提升自動駕駛技術的準確性。

2020 年由 Holger Caesar 等人提出的 nuScenes 資料集是一個用於自動駕駛系統的大型資料集。該資料集提供豐富的 sensor 資料，共涵蓋一千個分布於不同行駛環境及天氣條件下的場景。此外，nuScenes 使用多種 sensors 紀錄場景資料，包括 6 個攝影機、5 個雷達(RADAR)及 1 個光學雷達(LIDAR)等，並使用 23 個的類別和 8 個 3D 屬性框對所有資料標註特徵。[2][3]此外，nuScenes 的一大特色為高頻率的資料記錄率，這使得該資料集對於動態場景理解具有明顯的優勢。

2.2.2 Tusimple

TuSimple 資料集作為專門針對車道線偵測的公開資料集之一，為研究人員提供了豐富且高品質的數據來源，成為許多自動駕駛研究中的首選。

2017 年由 TuSimple 公司釋出的 TuSimple 資料集主要蒐集高速公路上的車道影像，並涵蓋了多種駕駛場景。該資料集的最大特色是提供了大量帶有精確標註的影像，特別是車道線的位置資訊，使模型能夠針對車道線的結構進行精確地學習與預測。TuSimple 資料集包含 6000 多個帶有標註的行車影像片段，這些影像片段涵蓋不同的光照條件（例如晴天、陰天）和車道情境（例如車道線清晰或磨損的狀態），提供多樣化的訓練素材。[4]

此外，TuSimple 資料集所提供的素材皆為 1280x720 的高解析度影像[4]，這種高解析度使模型能夠捕捉更細緻的車道線特徵，並為模型的訓練提供詳細的空間細節。而車道線的標註精度高，更使模型能夠學習到更精確的車道線位置和形狀特徵，甚至在模糊、部分遮擋或退化的情況下仍能保持穩定的偵測效果。

2.3 結構相似性(structural similarity, SSIM)

數位影像在其生命週期中，從採集、處理、壓縮、儲存、傳輸到再現的過程中，常常受到多種因素的影響，導致視覺效果的下降。而量化影像的視覺效果最可靠的方法是透過主觀評估(Mean Opinion Score, MOS)，然而，這種方法在實際應用中通常耗時且難以標準化。[5]

在以往的研究中，均方誤差(Mean Squared Error, MSE)和峰值訊噪比(Peak Signal-to-Noise Ratio, PSNR)是最常用的影像品質評估指標。這些方法之所以受到廣泛應用，主要是因為其計算簡單且具備明確的物理意義。然而，這些傳統方法存在根本性的缺陷。當兩張影像的 MSE 或 PSNR 數值相同時，人眼所感受到的差異可能會顯著不同。[6]這一現象主要是因為這些傳統方法僅關注像素之間的差異，而忽略了人類視覺系統(Human Visual System, HVS)的特性。

為了解決影像品質評估的相關問題，Zhou Wang 等人在 2004 年提出了一種基於結構相似性的影像品質評估方法，即結構相似性指標(structural similarity index, SSIM index)。該技術通過分析亮度、對比度及結構，來評估兩張影像之間的相似程度，這些因素正是人眼在觀察影像時所重視的要素。SSIM 不僅能夠對局部區域的細節進行評估，還能夠適應影像在壓縮及失真情況下的變化。[7]

近年來，SSIM 在影像處理領域得到了廣泛應用，特別是在影像壓縮品質分析和傳輸失真評估等方面。其創新之處在於能夠有效捕捉影像的結構特徵，從而提供比傳統方法更為準確的評估結果。

2.4 LSTM (Long Short-Term Memory)

循環神經網路(Recurrent Neural Network, RNN)是專門用於處理序列數據的神經網路，每個步驟的輸入都會影響之後的輸出，但是傳統的作法隨著時間的推移，梯度的大小會受到權重矩陣大小的指數影響，當權重在訓練過程中頻繁且不穩定的大幅變動中，可能會造成梯度消失或梯度爆炸，導致需要花時間的修正或者可能根本沒辦法起作用。[8]

因此 Sepp Hochreiter 跟 Jürgen Schmidhuber 提出了 LSTM 來解決這些問題，這是一種新的架構，結合適當的學習演算法，使其在有雜訊、不可壓縮的輸入的情況下，它也能連接超過 1000 步的時間間隔，而不損失短時間延遲的能力[8]，因此在自然語言處理、語音識別和其他應用中表現出色。

2.5 TSM (Temporal Shift Module)

由於科技的快速發展，影片的品質也隨其提升，許多地方都會需要準確且有效率的影片處理和影像辨識，而深度學習也成為一個標準，因此，高準確度以及低運算成本下進行影片處理也成為挑戰。2D CNN 沒辦法很好地對時間資訊進行建模，3D CNN 雖然可以處理學習空間和時間特徵，但運算成本較大。當然也有人嘗試兼具高準確度跟低運算成本，將 2D CNN 跟 3D CNN 混合使用，或是將 3D 分解成 2D 的空間卷積和 1D 的時間卷積，但是仍需要計算時間建模的成本。還有一些方法可以在時間建模和計算之間進行權衡，但可能會犧牲低階時間建模以達到提高效率，在時間融合發生之前的特徵提取過程中，會遺失許多有用的資料。[9]

因此 Ji Lin 等人提出了一種時間移位模組(Temporal Shift Module, TSM)。傳統的 2D CNN 在時間維度上獨立運作，因此沒有時間模型，而他們的時間移動模組會沿著時間維度向前和向後移動通道。然而，全部位移會導致效率不高，且需消耗大量記憶體，因此他們使用部分位移，也就是只移動了少部分通道，以減少移動成本。最終他們達到在具備高準確度的同時，也有較低的運算成本，也就是說，它能有 3D CNN 的性能，且保持 2D CNN 的計算複雜度。

第三章 研究方法

3.1 資料蒐集

Zou 等人在車道偵測模型上的研究中，選擇使用 LSTM 作為處理連續影像幀的方式，降低硬體設備的消耗並提高偵測性能。[1]其設計的核心理念是利用 LSTM 處理多幀影像的時間序列特徵，來提升車道偵測的穩定性，與我們的研究目標不謀而合。我們同樣希望在資源有限的嵌入式系統中實現高效的車道偵測，因此參考其設計方式進行模型的構建。

我們在此基礎上，將 LSTM 部分改用其他不同的技術進行研究，並比較不同模型和技術在資源消耗與偵測準確度上的表現，再依評估指標的數據進行篩選，找出最好的方法。本研究包含多組模型的設計對比，實驗不同技術在輕量化與準確度的平衡，最終再找出最符合預期研究成果的方法。

3.2 資料集預處理

為了與 Zou 等人的研究成果進行公平且精確的比較[1]，我們在資料集預處理階段採用了相同的基礎設計，並將原先使用的 LSTM 替換為其他技術來處理時序資料，使研究成果能夠更聚焦在不同技術所帶來的效果差異，確保比較的嚴謹性。

首先，影像的輸入大小設定為 128x256 像素，與原論文保持一致，這樣的設定在降低計算資源消耗的同時，也能確保模型捕捉到車道線的關鍵特徵。此外，影像數據經過正規化的處理，將像素值縮放至 [0,1] 範圍內，確保數據一致性，使模型在學習過程中不受數據範圍的影響，保持穩定的學習效果。

在訓練過程中，我們同樣採用相同的超參數設置，例如 batch size 和學習率，以避免不同實驗條件對比較結果造成影響。這樣的資料集預處理設計，使我們能夠更直觀地評估不同模型在車道偵測任務中的性能，並與原論文模型的輸出結果進行直接比較，從而驗證我們改進方法的有效性。

3.3 評估指標說明

因為本研究之主題是在維持甚至提升模型效能的前提下輕量化模型的架構，使得模型的運算成本降低。因此，除了使用可訓練參數量及 FLOPs 做為評估模型效能的指標外。在視覺化結果方面，本研究選擇一種可以計算兩張結果影像相似度的計算方式，作為評斷是否有比原模型表現更好的參考依據。

3.3.1 PSNR

PSNR 主要用於影像處理領域上，並作為量化兩張影像之間差異的指標。PSNR 基於 MSE 的概念，定義兩張影像每個像素誤差平方的平均值，測量訊號最大可能功率與破壞性雜訊功率的比值，並以對數分貝(dB)為單位。以下為 MSE 的計算方式：

$$(1) \quad MSE = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} [O(x,y) - C(x,y)]^2$$

其中 $O(x,y)$ 及 $C(x,y)$ 分別是原始影像與比較影像在位置 (x,y) 處的像素值，而 m 及 n 則是代表影像的寬度與高度。計算完 MSE 後，將其值帶入到 PSNR 的公式中。以下為 PSNR 的公式：

$$(2) \quad PSNR = 20 \times \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right)$$

其中， MAX_I 是影像的最大像素值。從以上公式中可以了解到，若 MSE 的數值越大，即兩張影像越相似，PSNR 的數值就越大。

3.3.2 SSIM

SSIM 與 PSNR 類似，主要應用於影像的相似性評估及分類上，是一種用來衡量兩張影像相似度的指標。透過分析兩張影像整體的亮度、對比度與結構來計算相似性，並聚焦於模擬人類視覺系統對於影像的感受，是常用的影像評估工具。而 SSIM 的數值範圍為 0~1，1 表示兩張影像完全相同。以下為 SSIM 的計算公式：

$$(3) \quad SSIM(x,y) = [L(x,y)]^\alpha \times [C(x,y)]^\beta \times [S(x,y)]^\gamma$$

其中， x 與 y 代表兩張影像的位置。而 $L(x,y)$ 是亮度比較值， $C(x,y)$ 是對比度比較值， $S(x,y)$ 是結構比較值。

PSNR 是一種專注於計算像素誤差值的指標，簡單且直觀，但它對人類視覺系統的模擬有限。而 SSIM 則是專注於影像的結構訊息，並考慮影像中的多種特徵，更適合作為反映人類的視覺系統對影像判斷的客觀指標。因此，相較於 PSNR，SSIM 對於影像的評估較為精確。所以本研究最終選擇 SSIM 作為研究中評斷模型表現的參考指標。

3.4 模型訓練

首先，本研究先使用原模型架構進行訓練，但在經過實驗後發現，模型在訓

練 5 次後的結果準確度已無明顯上升，因此最終將 epochs 定為 5。訓練好匯出 .pth 檔後，算出它的 FLOPs 及 trainable parameters，分別為 69,084,774,400 及 51,148,226，作為最後比較模型的依據，接著就開始進行優化的實驗。

根據 trainable parameters 的資料，我們發現 LSTM 占了模型運算成本中最大的一部份。因此，首先我們先將 encoder 及 decoder 各減少了一層，使其輸入至 LSTM 的尺寸變為原本的一半，從而降低 trainable parameters。而最終成功減少了運算成本，trainable parameters 降到 12,801,474，但 FLOPs 相較起來就沒有減少很多。

當前的研究目標是顯著降低模型的 trainable parameters 及 FLOPs，為此，首先嘗試將 encoder 及 decoder 中的卷積層改為深度可分離卷積(Depthwise separable convolution, DSConv)。DSConv 是由深度卷積(Depthwise convolution)及逐點卷積(Pointwise convolution)組成。在 Depthwise convolution 階段，channels 會先被分組處理，這使得其運算成本相較於傳統的標準卷積層來說可以有顯著的降低。

在指導老師的建議下，本研究進一步嘗試結合自注意力技術(Self-Attention)，期望模型可以自己適應不同尺寸的輸入。然而，根據實驗結果顯示，模型的特徵捕捉能力不符預期，甚至大幅降低。因此，基於以上的嘗試結果，本研究選擇放棄 Self-Attention 以保證模型的準確度。

後來我們發現了另一種技術——TSM，它主要是用於針對與時間相關且具有連續性的影像，通過將部分通道按照時間維度進行移位，進而促進相鄰兩幀之間的訊息交換。其具備高準確度的同時，也有較低的運算成本。因此，我們認為 TSM 適合用於車道偵測。綜合上述所提到的所有有效的技術，我們將原本的模型經過刪減層數、更改成 DSConv，以及將 LSTM 換成 TSM 後，進行訓練所得出的結果在 trainable parameters 及 FLOPs 都降低了 90%以上。

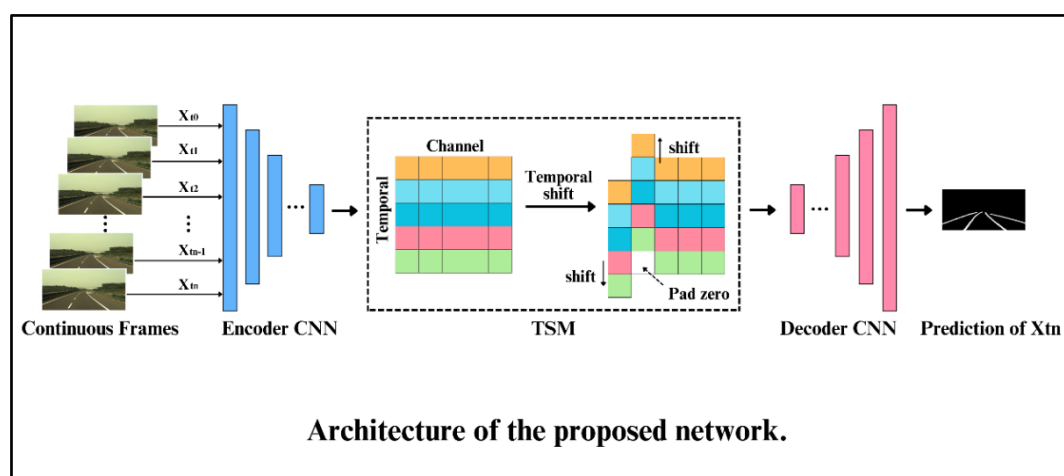


圖 3.1 模型架構圖

3.5 模型比較

雖然最終 FLOPs 跟 trainable parameters 都降低了非常多，但是也因此犧牲了準確度，這跟我們所期望的不符合。並且我們發現，可能是因為 DSConv 在 channels 分離後無法有效捕捉其特徵之間的關聯性，進而導致其準確度下降。因此我們決定採用成效較好的 TSM 搭配不同層數的 encoder 跟 decoder 來進行比較，觀察在哪種情況下，綜合的表現會比較好。

表 3.1 LSTM 與 TSM 搭配不同層數 encoder 及 decoder 之結果比較表

	Trainable parameters	FLOPs	SSIM
LSTM_3	12,801,474	58,502,938,624	0.8567
LSTM_4	51,148,226	69,084,774,400	0.8803
LSTM_5	204,518,850	79,660,515,328	0.8795
TSM_3	3,362,242	34,343,747,584	0.8743
TSM_4	13,395,394	44,925,583,360	0.8828
TSM_5	53,515,714	55,501,324,288	0.8870

LSTM_3 代表 LSTM 搭配 3 層 encoder 及 decoder，TSM_3 代表 TSM 搭配 3 層 encoder 及 decoder，以此類推。此外，為了明顯看出 SSIM 數值間的差距，其值皆四捨五入取到小數點後第四位。因此，根據表 3.1 的結果，trainable parameters 及 FLOPs 數值最低的模型為 TSM 搭配 3 層 encoder 及 decoder，其 trainable parameters 為 3,362,242、FLOPs 為 34,343,747,584。而在 SSIM 中，TSM_5 的數值是最高的，為 0.8870。

第四章 結果分析

根據上述的實驗設計，本研究最終選定以 trainable parameters、FLOPs 及 SSIM 作為評估模型效能的三項核心指標。這三項標準能夠有效地衡量模型的輕量化程度、運算效率及準確度，以找到最佳的效能平衡點。此外，由於本研究的硬體資源有限，無法如原論文中那樣進行多次的模型訓練以獲得充分優化的預訓練模型。因此，在模型比較方面，本研究選擇直接使用原論文中的模型作為基準，與最終成果模型進行運算成本的比較。並將資料集中用於驗證的影像作為基準真相 (Ground Truth)，計算其與最終模型結果影像的 SSIM，以評估模型的準確度。

根據表 3.1 的實驗結果，在六個候選模型中，TSM_4 模型在運算成本與準確度方面均展現了優異的表現。該模型不僅顯著降低了計算需求，還保持了較高的 SSIM，表現出對特徵的保留能力與穩定的輸出表現。這意味著 TSM_4 在保持準確度的同時，成功減少了計算資源的消耗，特別適合資源受限的應用場景。因此，綜合多方因素，最終將 TSM_4 選定為本研究的最終成果模型。

接下來，將分別從準確度與運算成本兩個方面進行詳細分析，將以客觀數據比較原始模型與最終成果模型 TSM_4 之間的差異，並說明最終結果模型的選定原因。

首先，在準確度方面，TSM_4 及 TSM_5 模型皆展現出極佳的特徵保留能力，在與 Ground Truth 的 SSIM 值上達到 0.8823 及 0.8870。相比之下，與原模型結構相同的 LSTM_4 模型僅達到 0.8803 的 SSIM 值。因此，代表 TSM_4 及 TSM_5 在捕捉影像細節與保持特徵完整性方面具有顯著優勢。由於 SSIM 值的提升空間隨著訓練次數的增加而增大，這兩個模型的性能有望在更多訓練下進一步優化，達到更高的準確度和特徵保留效果。

接著，在 TSM_4 及 TSM_5 的運算成本上，TSM_4 模型則展現了顯著的優勢。相較於原模型的數據，TSM_4 的 trainable parameters 減少了約 73%，而 FLOPs 則下降了約 35%。至於 TSM_5 的 FLOPs 雖然也有減少了約 20%，但 trainable parameters 卻上升了約 4%。關於 TSM_5 的這一劣勢主要來自於 encoder 及 decoder 層數的增加，雖然提升了準確度，但也同時增加了運算成本。因此，並不符合本專題的研究目標——輕量化。

TSM_4 透過模型架構的設計，使其能夠大幅減少運算資源。這種設計對於硬體資源受限的環境尤為重要，因為降低運算成本不僅提升了訓練效率，還使得 TSM_4 更適合於嵌入式設備等計算能力有限的平台運行。以下為 Ground Truth 以及最終的 TSM_4 模型的結果輸出影像。



圖 4.1 Ground Truth-影像一

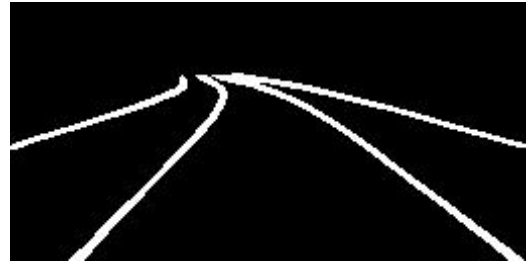


圖 4.2 TSM_4 結果-影像一

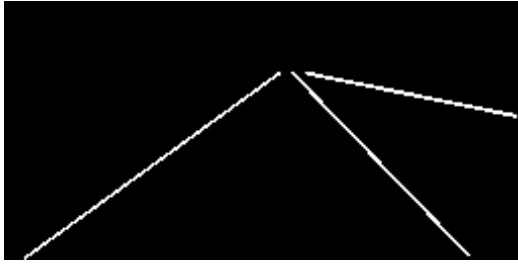


圖 4.3 Ground Truth-影像二

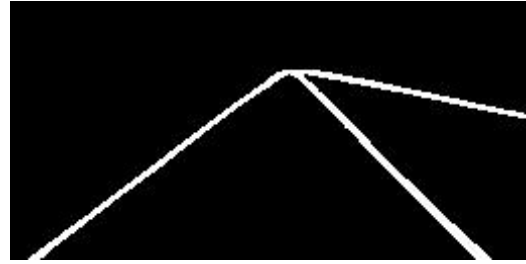


圖 4.4 TSM_4 結果-影像二



圖 4.5 Ground Truth-影像三



圖 4.6 TSM_4 結果-影像三



圖 4.7 Ground Truth-影像四

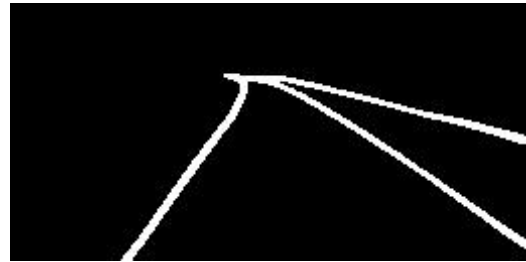


圖 4.8 TSM_4 結果-影像四



圖 4.9 Ground Truth-影像五

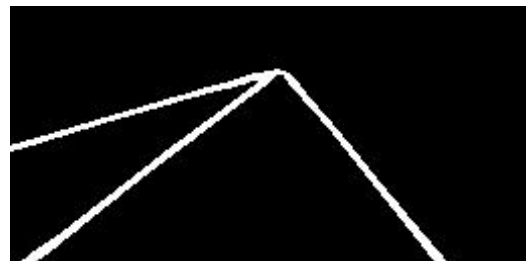


圖 4.10 TSM_4 結果-影像五

透過與原模型的對比，本研究可以在資源有限條件下驗證所提出模型的有效性，並從數據上探討不同模型架構在運算成本、準確度和資源消耗間的差異。這樣的比較也將為後續研究提供參考，幫助識別在硬體資源受限情況下的最佳模型選擇策略。此外，TSM_4 模型的選定也意味著該結構在達成輕量化與準確度之間的平衡上具有潛力，未來的研究可基於此模型進行進一步的改良與優化，以應用於更廣泛的實際場景中。

第五章 結論

本專題研究的重點在於實現輕量化的車道偵測模型，目標在於透過不同技術的結合應用與模型架構的設計，做到減輕模型的運算成本，期望在嵌入式系統或計算資源有限的設備上依然能夠保持高偵測準確度，以便應用於自動駕駛輔助系統中，並擴大其應用範圍並提升穩定性。後續說明研究結果、研究限制以及未來可研究方向。

在經過了多次的實驗後，我們發現將原模型的 LSTM 模型換成專門用於影片分析的 TSM Module，因為此技術能夠較高效的處理時序變化，並將 encoder 及 decoder 的層數維持原本的 4 層，這些改變可以使得 trainable parameters 及 FLOPs 的數值大幅下降，而特徵捕捉能力也可以維持。不過，此模型雖然經過了輕量化，但仍然無法實現實時偵測，因此要實際運用於自駕系統中還需再減少 FLOPs 以提升計算速度。此外，在複雜的環境中，如十字路口，此模型仍無法準確地偵測出此類場景中的車道線。

關於未來的展望，十字路口的車道線判斷一直是自動駕駛領域的一項挑戰，因為十字路口的環境複雜、路徑多樣，並且車道線的方向在此區域變得不明確。此外，不同的十字路口設計可能包含多種道路標記、箭頭或特殊標線，導致標記的形狀和位置較難識別和預測。因此，要準確地判斷十字路口的車道線成為目前技術的一大限制。

參考文獻

- [1] Zou, Q., Jiang, H., Dai, Q., Yue, Y., Chen, L., & Wang, Q. (2019). Robust lane detection from continuous driving scenes using deep neural networks. *IEEE transactions on vehicular technology*, 69(1), 41-54.
- [2] Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., ... & Beijbom, O. (2020). nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11621-11631).
- [3] nuScenes, URL:<https://www.nuscenes.org/>
- [4] TuSimple. *TuSimple Benchmark Dataset*. 2019, URL:<https://github.com/TuSimple/tusimple-benchmark>
- [5] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), 600-612.
- [6] Chandler, D. M., & Hemami, S. S. (2007). VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE transactions on image processing*, 16(9), 2284-2298.
- [7] Hore, A., & Ziou, D. (2010, August). Image quality metrics: PSNR vs. SSIM. In *2010 20th international conference on pattern recognition* (pp. 2366-2369). IEEE.
- [8] Hochreiter, S. (1997). Long Short-term Memory. *Neural Computation MIT-Press*.
- [9] Lin, J., Gan, C., & Han, S. (2019). Tsm: Temporal shift module for efficient video understanding. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 7083-7093).
- [10] Sato, T., & Chen, Q. A. (2022). Towards driving-oriented metric for lane detection models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 17153-17162).
- [11] Li, Q., Wang, Y., Wang, Y., & Zhao, H. (2022, May). Hdmapnet: An online hd map construction and evaluation framework. In *2022 International Conference on Robotics and Automation (ICRA)* (pp. 4628-4634). IEEE.