Research on an Improved DeepLab V3+ Object Segmentation Method Based on Contextual Semantic Aggregation

HSU, Shih-Chang[1]* ,TSAI, Kuan-Chieh
Department of Electrical Engineering, National University of Tainan, Tainan, 70005, Taiwan
[1*]E-mail : scshei@mail.nutn.edu.tw

Abstract

This article primarily examines object segmentation methods based on contextual semantic aggregation, particularly focusing on comparing DeepLab V2 and DeepLab V3+ on the VOC2012 dataset, and enhancing DeepLab V3+ to optimize its parameter count while maintaining a certain segmentation accuracy.

Object segmentation is a crucial task in computer vision, aimed at distinguishing the target object from the background in an image. Recently, advancements in deep learning have led to significant improvements in object segmentation tasks, especially with convolutional neural network (CNN)-based methods such as DeepLab V2 and DeepLab V3+. These methods employ techniques like atrous convolution and multi-scale pooling to capture object information at various scales. Specifically, DeepLab V3+ introduced the Multi-scale atrous convolution module to further enhance segmentation accuracy.

Nevertheless, DeepLab V3+ has a higher parameter count, necessitating improvements to boost its efficiency. Therefore, we proposed an enhancement method based on MobileNetV2, which involves adjusting the sampling rate combination of the Atrous Spatial Pyramid Pooling (ASPP) module and incorporating the Convolutional Block Attention Module (CBAM) mixed attention mechanism. Additionally, we included parallel self-attention mechanisms in the decoder section of the DeepLab V3+ structure and optimized feature fusion into a multi-branch feature fusion module. Through these modifications, we successfully reduced the parameter count of DeepLab V3+ while maintaining a certain level of segmentation accuracy.

The experimental results indicate that the improved DeepLab V3+ method exhibits high efficiency and segmentation accuracy, making it highly effective in object segmentation tasks.

Keywords: Context semantic aggregation ；DeepLabv3 plus ；CBAM attention mechanism ；Self-attention mechanism ；Multi-branch feature fusion

---

Corresponding author:
scshei@mail.nutn.edu.tw
DOI:10.53106/222344892026041601001

基於上下文語義聚合的改進型 DeepLab V3+ 對象分割方法研究

許世昌, 蔡貫傑

國立臺南大學電機工程系
scshei@mail.nutn.edu.tw

摘要

本文主要研究基於上下文語義聚合的對象分割方法，重點關注了 DeepLab V2 和 DeepLab V3+兩種方法在 VOC2012 資料集上的比較，並對 DeepLab V3+方法進行了改進以優化其參數量並保持一定分割精度。

對象分割的目的是將圖像中的目標分為前景，其餘像素分為背景。近年來，深度學習技術的發展使得對象分割任務取得了重要進展。其中，基於卷積神經網絡

（Convolutional Neural Networks，CNN）的方法表現出了較好的效果，如 DeepLab V2 和基於其的 DeepLab V3+。這兩種方法都採用了空洞卷積和多尺度池化等技術，以捕捉不同尺度的對象信息。其中，DeepLab V3+還引入了多尺度空洞卷積模塊，以進一步提高分割精度。

在本次課題研究中，所做實驗使用 VOC2012 資料集進行了對比實驗，結果表明 DeepLab V3+在分割精度上優於 DeepLab V2。然而，DeepLab V3+的參數量也更大，因 此需要進行改進以提高其效率。為此，本文提出了一種基於 MobileNetV2 的改進方法，該方法主要包括兩個方面：一是改變空洞空間金字塔結構（Atrous Spatial Pyramid Pooling，ASPP）的採樣率組合併加入 CBAM 混合注意力機制，二是在 DeepLab V3+結構的解碼器部分加入並聯自注意力機制，並將特徵融合優化為多分支特徵融合模塊。通過這些改進，本文成功地減少了 DeepLab V3+的參數量，並保持了一定的分割精度。

實驗結果表明，改進後的 DeepLab V3+方法具有較高的效率和較好的分割精度，可在對象分割任務中發揮重要作用。本文的研究對於提高對象分割的效率和精度具有重要意義，同時也為其他基於 CNN 的對象分割任務提供了借鑒和啓示。

關鍵詞：上下文語義聚合；DeepLab V3+；CBAM 注意力機制；自注意力機制；多分支特徵融合

1. 緒論

1.1 研究背景及意義

隨著人工智慧的發展，語義分割已成為計算機視覺領域中的關鍵任務之一，應用場景包括醫學影像病理切片的分割和自動駕駛路況圖像判斷[1,2,3]。語義分割本質上是像素級分類，需充分利用圖像中的上下文語義信息，即像素之間的關聯。Biederman 將上下文信息劃分為語義、尺度、空間三類[4]，目前基於上下文語義聚合的分割方法在該領域有著重大研究價值[5]。然而，主流模型如 DeepLab V3+ 仍面臨感受野不夠大、信息不連續等問題。本文探討如何有效解決這些問題。

1.2 國內外研究現狀

早期的語義分割依賴傳統方法，如 Shi J 的 Normalized cuts[6] 和 Achanta R 的 SLIC 超像素分割[7]，但準確度和效率不佳。2014 年，Long J 等人提出基於深度學習的 FCN[8]，開啟語義分割的突破。後續方法如 ParseNet[9]、SegNet[10] 和 U-Net[11] 進一步提升效果。DeepLab 系列自 2015 年推出 V1[12]，通過空洞卷積和空間金字塔池化成為主流，DeepLab V3+[13] 引入編碼解碼結構和 ASPP 模塊提高精度。基於注意力的模型如 DFANet[14] 和 OCRNet[15] 顯示出優越性能，而 HRNet[16] 和 Transformer[17] 在融合上下文信息上有進展，但結構複雜。本文提出改進的 DeepLab V3+，結合 CBAM 混合注意力、自注意力和多分支特徵融合模塊，提升語義分割效果。

1.3 論文內容與結構

本文探討了上下文語義聚合在對象分割中的重要性，並基於 DeepLab 系列模型進行改進。由於圖像背景複雜且存在遮擋，語義分割變得困難。本文將通過改進 DeepLab V3+ 的 ASPP 結構，結合多尺度自注意力機制和多分支特徵融合模塊來提升模型性能。全文分為五章：第一章介紹語義分割的相關研究，第二章解釋卷積神經網絡基礎，第三章通過實驗優化 DeepLab 系列，第四章提出改進方案，最後第五章總結與未來方向。結果顯示，本文提出的方法有效提升了語義分割的準確性。

2. 對象分割理論概述

2.1 語意分割

語義分割的目的是將圖像中的每個像素分類到對應的目標標籤區域中，能夠精確描述圖像的內容與結構[18]。它在多個領域有重要應用，如自動駕駛中的道路目標分割與預測、醫學影像中的組織與器官分割等。傳統方法依賴手工設計特徵，雖計算快速，但受限於特徵設計；深度學習方法通過神經網絡自動學習特徵，表達和泛化能力更強，雖需要更多數據和計算資源[19]。

近年來，深度學習已逐漸取代傳統語義分割方法，主流模型多基於深層 CNN，並結合特殊設計提高準確性與效率，如 PSPNet、ENet、BiSeNet 等[20]。語義分割作為計算機視覺重要研究方向，隨著技術發展，其應用範圍與性能將不斷擴展。

上下文語義信息聚合在語義分割中至關重要，因為圖像中的每個像素點不會獨立存在，而是與周圍像素相互作用。根據區域大小，上下文特徵可分為局部與全局，通過卷積層、池化層、自注意力機制等方式進行聚合。這有助於提高分割精度、改善模型魯棒性及泛化能力，使模型能在複雜場景下依然表現優異[21]。

2.2 資料集介紹

VOC2012 資料集由牛津大學和卡爾斯魯厄理工學院共同發佈，是目標檢測和語義分割研究中具挑戰性的資料集之一。它包含 20 個物體類別和 1 個背景類別，圖像分辨率介於 300x300 至 500x500 之間，標註了物體邊界框和像素級語義分割信息。該資料集提供標準訓練/驗證集和測試集，用於演算法的訓練和評估。

VOC2012 推動了計算機視覺的發展，為研究人員提供了公正的測試平台，並廣泛應用於自動駕駛、機器人視覺等領域。許多經典深度學習模型如 FCN、U-Net 和 DeepLab 均在此資料集上進行訓練和測試。因此，本研究選擇使用 VOC2012 來驗證模型性能。

3. 基於上下文語義聚合的對象分割方法

3.1 引言

當前的深度學習模型如 U-Net、FCN、PSPNet 對圖像的上下文語義信息利用不足，難以有效融合多尺度信息，特別是在高分辨率和複雜場景下問題更加突出。為解決此問題，本章選取 DeepLab 系列進行性能驗證，擴展其 ASPP 結構至六分支以增大感受野，並引入 CBAM 混合注意力機制，提升模型對分割目標的關注，減少信息冗余，提升分割效果。

3.2 DeepLab 模型架構

DeepLab 系列模型在語義分割領域取得顯著成果。DeepLab V1 使用 VGG 網絡和空洞卷積提升性能，但計算代價較高。DeepLab V2 引入 ASPP 結構進行多尺度融合，提升精度並降低計算成本，但空洞卷積仍導致信息丟失和計算量增加。DeepLab V3 改進了 ASPP 結構，提升對多尺度對象的處理能力，解決了信息丟失問題。DeepLab V3+ 參考 FCN 和 Transformer 結構，進一步增強了解碼器和特征融合。然而，這些模型仍需大量訓練數據，難應對複雜場景，且計算代價高。

（a）傳統卷積　　（b）空洞卷積

圖 3.1 傳統卷積及空洞卷積

3.3 多尺度空洞空間金字塔池化

為解決空洞卷積的信息不連續性，通常選用多種採樣率進行重採樣。例如，採樣率為 1 和 3 的空洞卷積可以覆蓋 9×9 圖像，如圖 3.2 所示。紅點為採樣率 3 的位置，藍點為採樣率 1 的位置。本文在此基礎上加入採樣率為 24 的空洞卷積，形成 1：2：

3：4 的組合，擴大 ASPP 結構的感受野並減輕局部信息不連續性。該操作增加的參數量小，但有效提升了多尺度特徵提取，並通過 1×1 卷積進行降維，平衡了性能與訓練成本。

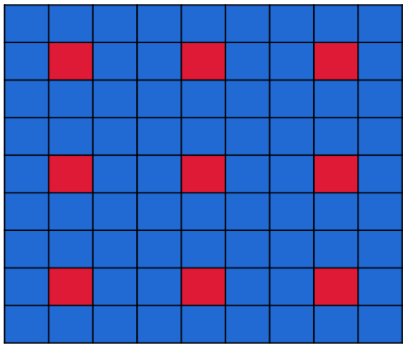

圖 3.2 多尺度空洞卷積

3.4 CBAM 混合注意力機制

CBAM 是一種基於混合注意力機制的 CNN 模塊，由通道注意力模塊（CAM）和空間注意力模塊（SAM）組成，其結構如圖 3.3 所示。

CAM 通過全局平均池化和最大池化獲得特征圖每個通道的重要性評分，減少冗余信息，使模型更專注於重要通道。

SAM 模塊關注特征圖中每個空間位置的重要性，通過學習空間權重矩陣來縮放特征圖。該過程可視為空間維度上的特征選擇，使用兩個卷積層學習空間權重矩陣，並通過矩陣乘法將特征圖與權重矩陣相乘以增強目標位置信息[22]。

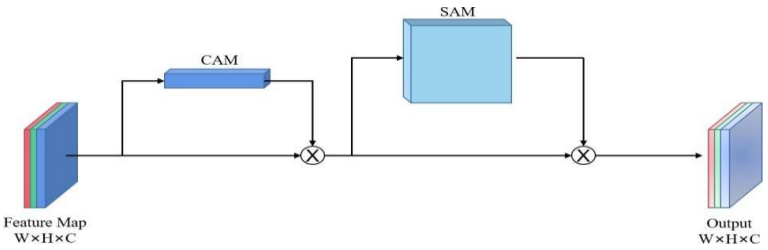

圖 3.3 CBAM 混合注意力機

3.5基於混合注意力機制的 DeepLab V3+

本文將 CBAM 混合注意力模塊嵌入 DeepLab V3+ 的 ASPP 結構中，進一步提取高級語義信息的位置信息及通道信息。同時，為防止梯度消失，參考 ResNet 系列模型的殘差結構，將 CBAM 模塊構建為跳躍連接結構，形成殘差 CBAM 模塊，以保持一定的恒等映射能力，其結構如圖 3.4 所示。

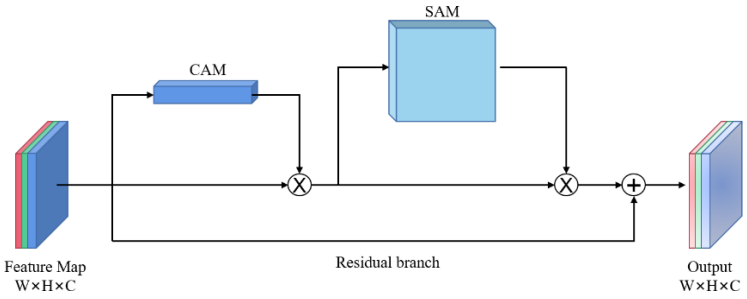

圖 3.4 殘差 CBAM 模塊

此外，本文采用 MobileNetV2 作為主幹網絡，綜合考慮訓練成本與準確率，認為犧牲少量準確率以減少大量參數是必要的。相較於傳統的 Xception 和 ResNet101，MobileNetV2 在保持極少參數的同時，性能仍保持較高水平。

MobileNetV2 通過深度可分離卷積減少網絡參數，並引入線性瓶頸、倒殘差模塊、空洞卷積和 SE 模塊等新結構。線性瓶頸在非線性變換之前執行線性變換，以更好地利用層間信息；倒殘差模塊在保持網絡深度的同時減少參數量；SE 模塊是一種基於通道的注意力機制，可以學習各通道之間的關系，從而更好地利用特征信息[23]。

綜上所述，本章使用的結構如圖 3.5 所示，DeepLab V3+ 其余部分與原網絡相同。

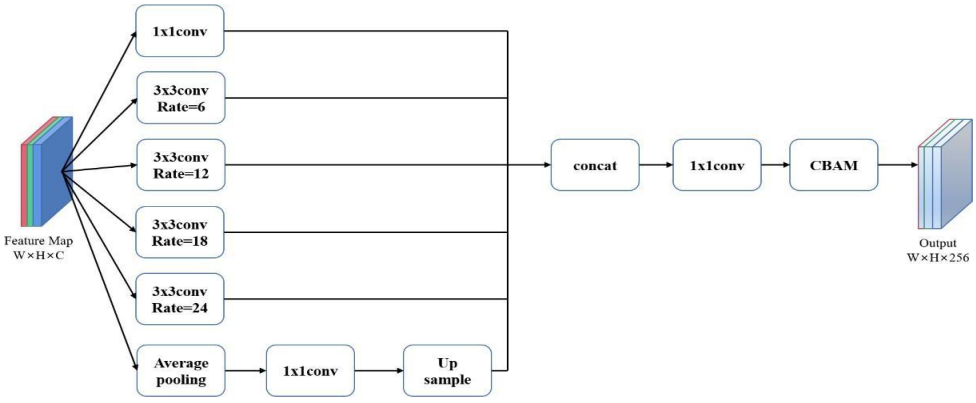

圖 3.5 基於 CBAM 混合注意力機制的 DeepLab V3+

3.6實驗結果及分析

本文實驗環境為 Ubuntu 16.04，使用 PyTorch 深度學習框架進行模型構建，實驗設備為 NVIDIA RTX 2080 Ti，顯存為 24GB。DeepLab V2 的主幹網絡選擇 ResNet101，DeepLab V3+ 選擇 MobileNet V2 作為默認 backbone。優化器為 Adam，batch size 設為 8，訓練輪數（epoch）為 150。模型性能評估指標為 mIoU、mPA 和 PA。

3.6.1對比實驗

在本本研究首先比較 DeepLab V2 和 DeepLab V3+ ，針對 VOC2012 數據集進行分割任務的訓練和測試。通過比較兩種模型在測試集上的性能指標來評估它們的優劣。

實驗結果顯示，在 VOC2012 數據集中，DeepLab V2 模型的 mIoU 為約 68.4%，表現一般，實驗測試結果如表 3-1 所示。相比之下，DeepLab V3+ 模型在同一數據集上的表現更優，mIoU 達到 71.48%，mPA 為 81.39%，PA 達到 93.13%。

表 3-1 對比實驗-結果

Methods	Backbone	mIoU	mPA	PA
DeepLab V2	ResNet101	68.4	78.24	89.86
DeepLab V3+	MobileNetV2	**71.48**	**81.39**	**93.13**

除此之外，DeepLab V2 由於採用 ResNet101 作為主幹網絡，其訓練速度慢，參數量大，分割精度相對較低，並不可作為語義分割模型首選方案。

將本章方法與其他主流模型效果對比結果如表 3-2 所示，所有模型的 backbone 選取都為 MobileNetV2，由實驗結果對比，本章方法相比於主流的語義分割模型在 mIoU、 mPA、PA 三項指標上都取得了更優秀的性能。

表　對比實驗二

3-2 結果

Methods	mIoU	mPA	PA
U-Net	68.32	78.12	89.56
FCN	66.73	76.94	84.36

續表 3-2 對比實驗二結果

PSPNet	69.32	80.32	91.04
基于 CBAM 的 DeepLab V3+	**72.11**	**82.6**	**93.53**

3.6.2 消融實驗

表 3-3 消融實驗結果

Methods	mIoU	mPA	PA
Baseline	71.48	81.39	93.13
ASPP 輸入處加入 CBAM	70.83	81.0	93.08
ASPP 分支 4 後加入 CBAM	71.59	82.21	93.26
ASPP 分支 2、4 後加入 CBAM	71.15	81.79	93.14
ASPP 輸出降維後加 CBAM	**72.11**	82.6	**93.53**
ASPP 所有分支後加 CBAM	71.49	81.73	93.17
ASPP 分支 5 加 CBAM	71.91	**82.68**	93.12
ASPP 輸出、分支 5 加 CBAM	70.76	81.45	93.03
ASPP 分支 2、5 加 CBAM	70.58	81.29	93.06
ASPP 分支 3、5 加 CBAM	70.81	81.71	93.08

本文通過改進 DeepLab V3+ 的 ASPP 結構和引入 CBAM 注意力機制，將 ASPP 采樣率從 6、12、18 擴展至 6、12、18、24，顯著提升模型性能。實驗結果顯示，改進後 mIoU 提升至 72.11%，mPA 為 82.6%，PA 為 93.53%，相較基模型，mIoU 增加 0.63%，mPA 增加 0.91%。在所有變種模型中，ASPP 輸出降維處加入 CBAM 的效果最佳。

3.6.3結果分析

本研究對比 DeepLab V2 和 DeepLab V3+ 在 VOC2012 數據集上的表現，發現 DeepLab V3+ 憑藉更先進的 backbone 和 ASPP 結構，mIoU 達到 71.48%。為提升性能，本文改進 ASPP 並加入 CBAM 注意力機制，將 mIoU 提升至 72.11%。同時，將 ASPP 的采樣率擴展為 6、12、18、24，有效提升對大物體的識別精度。儘管改進增加了計算覆雜度，但在細節識別上表現更佳。DeepLab V2 使用 ResNet101 需 20 小時訓練，而 DeepLab V3+ 採用 MobileNet V2，僅需 3 小時，且參數量更小。

3.1 本章小結

本章通過引入 CBAM 注意力機制和改進 ASPP 結構的採樣率，證明了改進的 DeepLab V3+ 模型提升了性能。CBAM 有助於自適應調整通道和空間位置的重要性，增強分類性能並過濾無用信息，提高模型的魯棒性和泛化能力。儘管增加了計算複雜度和訓練時間，改進後的模型參數量更少，且保持了較高精度。

4. 改進的 DeepLab V3+模型

4.1 引言

在目標分割任務中，複雜場景下準確分割是關鍵。常見問題是目標過多時容易出現過分割或分割缺失。為解決此問題，通常引入注意力機制進行目標學習。然而，普通通道和空間注意力未能全面反饋所有信息。為此，本文將自注意力機制並聯至 DeepLab V3+ 的 ASPP 結構，減少計算量並提高效率。此外，使用多分支特征融合模塊代替普通融合模塊，增加特征密集度和信息豐富度，提升上下文語義信息的聚合能力，改善網絡性能。

4.2 並聯自注意力模塊

自注意力機制源於注意力機制，廣泛應用於序列數據處理，尤其是自然語言處理。它通過計算序列中各部分的加權和，使每個部分能「自我關注」，更好地捕捉序列間的關聯。自注意力依賴 Q（Query）、K（Key）、V（Value）三個張量，由輸入張量通過線性變換生成。其加權過程通過相似度計算和 softmax 函數實現，進而根據關鍵信息調整輸出的表示。在圖像處理中，自注意力機制能靈活調整感受野大小，隨訓練次數增加，注意力逐漸集中在圖像的關鍵區域上，如圖 4-1 所示。

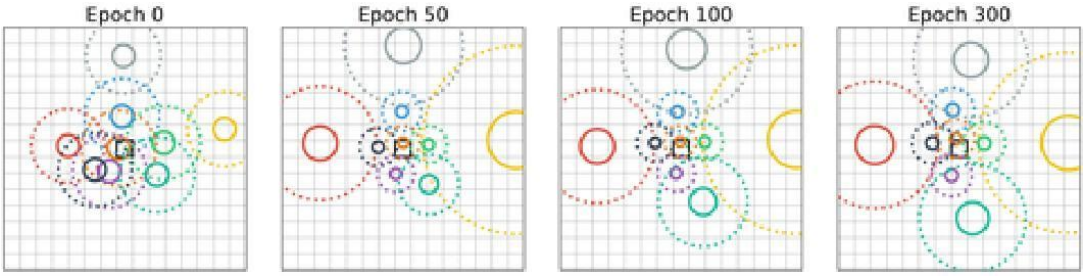

圖 4.1 自注意力訓練過程[27]

本文將自注意力機制獨立形成一個分支並將其並聯在 ASPP 兩側，將自注意力的與多尺度特徵的優勢融合，解決局部信息缺失、像素上下文語義信息不夠豐富的問題。

4.3多分支特徵融合模塊

本文引入基於多分支堆疊的特徵融合模塊（ELAN）改進並聯自注意力機制的 DeepLab V3+。傳統特徵融合通過拼接後的全連接或卷積層處理，而 ELAN 採用四個卷積標準化激活函數分支：全局特徵、局部特徵、交叉特徵、自適應特徵分支。這些分支分別提取全局、局部、交叉特徵並進行自適應調節，最終拼接產生更豐富的特徵圖。相比傳統方法，ELAN 提高了模型表現，且未增加顯著計算量。該方法借鑒 ResNet 的殘差結構，通過標準卷積層堆疊加深網絡，同時緩解梯度消失問題。更密集的殘差連接保證了特徵融合的信息豐富度。將 DeepLab V3+ 中的特徵拼接模塊替換為多分支特徵融合，提升模型的特徵表達能力。

4.4 DeepLab V3+

本文提出的基於並聯自注意力和多分支特徵融合的 Deeplab V3+ 結構如圖 4.2 所示。該方法解決了 ASPP 結構中空洞卷積導致的棋盤效應問題，通過集成全局上下文信息及自注意力機制，提升全局上下文感知與模型泛化能力。同時，ELAN 模塊的引入實現了更密集的特徵融合，增強了細節和邊緣的處理，提升語義分割性能，特別適合複雜場景。

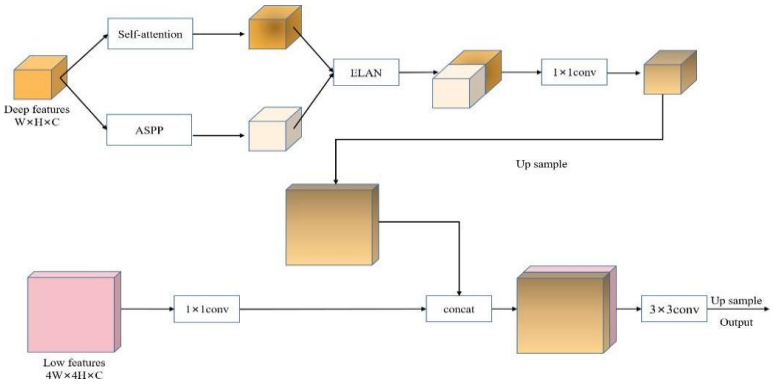

圖 4.2 基於並聯自注意力及多分支特徵融合的 Deeplab V3+

4.5 實驗結果及分析

本文實驗環境為 Ubuntu 16.04，使用 Pytorch 框架，設備為 NVIDIA RTX 2080ti(24GB 顯存)。主幹網絡選用 MobileNet V2，優化器為 Adam，batch size 設為 8，epoch 設為 100。模型性能評估指標為 mIoU、mPA 和 PA。

4.5.1 對比實驗

該實驗將改進模型與主流語義分割模型進行對比，結果如表 4-1 所示。實驗顯示，改進模型在 mIoU、mPA 和 PA 指標上均有顯著提升，且在大小目標的分割效果上優於主流模型，同時保持較少的模型參數量。

表 4-1 對比實驗結果

Methods	mIoU	mPA	PA
U-Net	68.32	78.12	89.56
FCN	66.73	76.94	84.36
PSPNet	69.32	80.32	91.04
SegNet	68.78	78.88	89.92
DeepLab V3plus+selfAttention+ELAN	**72.34**	**82.24**	**93.51**

4.5.2 消融實驗

本實驗在 VOC2012 資料集上比較了四個模型的性能：DeepLab V3+、基於並聯自注意力機制的 DeepLab V3+、基於多分支特徵融合模塊的 DeepLab V3+，以及基於並聯自注意力機制及 ELAN 的 DeepLab V3+。實驗結果如表 4-1 所示。結果顯示，基於並聯自注意力機制及 ELAN 的 DeepLab V3+ 在 mIoU 指標上達到 72.34%，相比傳統 DeepLab V3+ 有明顯提升，且因 ELAN 的密集特徵融合，模型性能進一步增強。此外，該方法在簡單和複雜場景的 VOC2012 資料集上的效果提升，表明其強大的魯棒性。

表 4-2 消融實驗結果

表 4-2 消融實驗結果

Methods	mIoU	mPA	PA
Baseline	71.48	81.39	93.13
DeepLab V3plus+selfAttention	72.13	81.91	93.11
DeepLab V3plus+ELAN	71.66	81.83	93.27
DeepLab V3plus+selfAttention+ELAN	**72.34**	**82.24**	**93.51**

4.5.3 結果分析

在 DeepLab V3+ 模型中加入並聯自注意力分支及 ELAN 模塊後，模型的 mIoU、mPA 和 PA 均有所提高，展現出優秀的準確度和魯棒性。然而，這些改進增加了網絡深度，導致模型參數量和時間複雜度上升。原始模型的權重文件約為 23M，而基於並聯自注意力及 ELAN 的模型約為 26M，僅小幅提升。訓練速度相較於原始模型略有降低。

4.6 本章小結

本章提出將自注意力機制並聯於 ASPP 結構，並用 ELAN 模塊替換傳統特徵融合，以實現更密集的特徵融合。自注意力模塊捕獲全局上下文信息，提升模型對物體結構和語義的理解，從而提高分割精度。此外，引入自注意力可減少模型參數數量，降低計算複雜度和存

儲開銷。ELAN 模塊以堆疊形式融合多尺度特徵圖，提供更豐富的信息。對比實驗、消融實驗和可視化實驗結果顯示，這些改進顯著增強了網絡性能。

5 結論與展望

本文主要比較 DeepLab V2 和 DeepLab V3+ 在 VOC2012 資料集上的表現，並對 DeepLab V3+ 進行改進以優化參數量並保持分割精度。實驗結果顯示，DeepLab V3+ 在分割精度上優於 DeepLab V2，為後續改進提供基礎。針對 DeepLab V3+ 在上下文信息和參數量上的問題，提出了一種基於 MobileNetV2 的改進方法，主要包括兩個方面：

（1）改變 ASPP 的採樣率，並加入 CBAM 混合注意力機制，以增強特徵並提高對象分割精度。消融和對比實驗均顯示性能小幅提升。

（2）在 ASPP 處並聯自注意力分支，並用多分支特徵融合模塊替換傳統特徵融合，以達到更密集的特徵融合，進一步提升對象分割精度。

在研究過程中遇到的問題包括資料集選擇、模型訓練挑戰及演算法改進的可行性。儘管如此，研究對基於上下文語義聚合的對象分割模型具有重要意義，並為其他基於 CNN 的任務提供了借鑒。

6.參考文獻

[1] 田萱, 王亮, 丁琪. 基于深度学习的图像语义分割方法综述[J]. 软件学报, 2019, 30(2): 440-468.

[2] 李璟, 朱善安. 医学图像分割技术[J]. 生物医学工程学杂志, 2006, 23(4): 891-894.

[3] 王中宇, 倪显扬, 尚振东. 利用卷积神经网络的自动驾驶场景语义分割[J]. Optics and Precision Engineering, 2019, 27(11): 2429-2438.

[4] Zou Z, Chen K, Shi Z, et al. Object detection in 20 years: A survey[J]. Proceedings of the IEEE, 2023.

[5] 姜枫, 顾庆, 郝慧珍, 等. 基于内容的图像分割方法综述[J]. 软件学报, 2016, 28(1): 160-183.

[6] Shi J, Malik J. Normalized cuts and image segmentation[J]. IEEE Transactions on pattern analysis and machine intelligence, 2000, 22(8):

888-905.

[7] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 34(11): 2274-2282.

[8] 陈娜,张荣芬,刘宇红,李丽,张雯雯.基于类特征注意力机制融合的语义分割演算法[J].液晶与显示,2023,38(02):236-244.

[9] Liu W, Rabinovich A, Berg A C. Parsenet: Looking wider to see better[J]. arXiv preprint arXiv:1506.04579, 2015

[10] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.

[11] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C].Medical Image Computing and Computer-Assisted Intervention– MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.

[12] Tang W, Zou D, Yang S, et al. A two-stage approach for automatic liver segmentation with Faster R-CNN and DeepLab[J]. Neural Computing and Applications, 2020, 32: 6769-6778.

[13] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C].Proceedings of the European conference on computer vision (ECCV). 2018: 801-818

[14] Li H, Xiong P, Fan H, et al. Dfanet: Deep feature aggregation for real-time semantic segmentation[C].Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 9522-9531.

[15] Yuan Y, Chen X, Wang J. Object-contextual representations for semantic segmentation[C].Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16. Springer International Publishing, 2020:

173-190.

[16] Wang J, Sun K, Cheng T, et al. Deep high-resolution representation learning for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2020, 43(10): 3349-3364.

[17] Khan S, Naseer M, Hayat M, et al. Transformers in vision: A survey[J]. ACM computing surveys (CSUR), 2022, 54(10s): 1-41.

[18] Guo Y, Liu Y, Georgiou T, et al. A review of semantic segmentation using deep neural networks[J]. International journal of multimedia information retrieval, 2018, 7: 87-93.

[19] Yu H, Yang Z, Tan L, et al. Methods and datasets on semantic segmentation: A review[J]. Neurocomputing, 2018, 304: 82-103.

[20] Yu C, Wang J, Peng C, et al. Bisenet: Bilateral segmentation network for real-time semantic segmentation[C].Proceedings of the European conference on computer vision (ECCV). 2018: 325-341.

[21] 温光玉, 唐雁, 吴梦蝶, 等. 基于图像上下文语义信息的场景分类方法[J]. 四川大学学报: 自然科学版, 2013 (6): 1223-1229.

[22] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C].Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[23] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C].Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.

[24] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[25] Cordonnier J B, Loukas A, Jaggi M. On the relationship between self-attention and convolutional layers[J]. arXiv preprint arXiv:1911.03584, 2019.

[26] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C].Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

2023: 7464-7475.

[27] Cordonnier J B, Loukas A, Jaggi M. On the relationship between self-attention and convolutional layers[J]. arXiv preprint arXiv:1911.03584, 2019.