DLMI HW

ID: r12922075

Name: 吳韋論

Data

Breast Ultrasound Images Dataset

Introduction

Detecting breast cancer early is crucial for saving lives. Combining ultrasound imaging with deep learning significantly enhances the accuracy of identifying breast cancer stages.

The dataset used in this study is available on Kaggle, and collected in 2018. This dataset consists of ultrasound images from 600 women aged between 25 and 75. It's designed to enhance learning techniques for segmentation, classification, and detection breast cancer in ultrasound images.

Dataset

This dataset comprises 780 images, each with a resolution of 500×500 pixels in PNG format. The images are classified into three categories and each image has corresponding mask:

- normal
- benign
- malignant

Project Objective

Segmentation

Medical segmentation is a important process in analyzing medical images, focusing on lesion areas pixel by pixel. The accuracy of segmentation can help doctors to diagnose diseases and make well-informed treatment decisions.

Segmentation in 2D medical imaging data has many methods, ranging from traditional CNN-based models to more recent Transformer-based models. We aim to experiment with different state-of-the-art models. Finally, we will compare their performances in terms of Dice loss, and visualize the segmentation results for a comprehensive evaluation.

Methodology

I chose 9 different types of popular models: Unet, Unet++, MAnet, Linknet, FPN, PSPNet, PAN. DeepLabV3, and DeepLabV3+. Each model has unique characteristics explained below.

1. Unet

- **Architecture**: Consists of a contracting path (encoder) to capture context and a symmetric expanding path (decoder) to enable precise localization.
- **Key Features**: Uses skip connections to concatenate feature maps from the encoder to the decoder, improving the flow of information across the network and aiding in precise segmentation.

2. Unet++

- **Architecture**: Extends the Unet by introducing nested, dense skip pathways.
- **Key Features**: These nested connections aim to bridge the semantic gap between the feature maps of the encoder and decoder, providing more refined features to the decoder and enhancing the model's ability to capture fine details.

3. MAnet (Multi-Scale Attention Network)

- **Architecture**: Integrates an attention mechanism into a standard segmentation framework to focus on relevant features.
- **Key Features**: Employs multiple attention modules to refine the feature map at different stages of the network, enhancing the model's focus on salient regions of the input image.

4. Linknet

- **Architecture**: Designed for efficiency, features an encoder-decoder structure with lightweight residual blocks.
- **Key Features**: Prioritizes fast inference speeds while maintaining good segmentation performance, making it suitable for applications requiring real-time processing.

5. FPN (Feature Pyramid Network)

- **Architecture**: Builds a multi-scale pyramid of features by combining low-resolution, semantically strong features with high-resolution, semantically weak features through a top-down pathway and lateral connections.
- **Key Features**: Improves performance in detecting and segmenting objects at various scales by effectively utilizing multi-scale feature representations.

6. PSPNet (Pyramid Scene Parsing Network)

- **Architecture**: Utilizes a pyramid pooling module at different grid scales to aggregate global context information effectively.
- **Key Features**: By capturing the scene context at multiple levels, PSPNet can handle complex scene parsing tasks, making it robust to varying object sizes and scene layouts.

7. PAN (Pyramid Attention Network)

- **Architecture**: Combines pyramid feature fusion with attention mechanisms to focus on informative features across different scales.
- **Key Features**: Enhances the model's ability to segment objects with fine detail by directing the network's focus towards relevant features and improving multi-scale feature integration.

8. / 9. DeepLabV3/DeepLabV3+

- **Architecture**: Uses atrous (dilated) convolutions to capture context at multiple scales without losing resolution. DeepLabV3+ adds an encoder-decoder structure to refine segmentation results further.
- **Key Features**: The atrous convolution allows the model to control the field-of-view and capture multi-scale information effectively. The addition of the

encoder-decoder setup in DeepLabV3+ improves segmentation boundaries.

Each of these models brings distinct advantages to the table, from Unet's foundational approach to segmentation to DeepLabV3+'s advanced use of atrous convolutions for capturing multi-scale information.

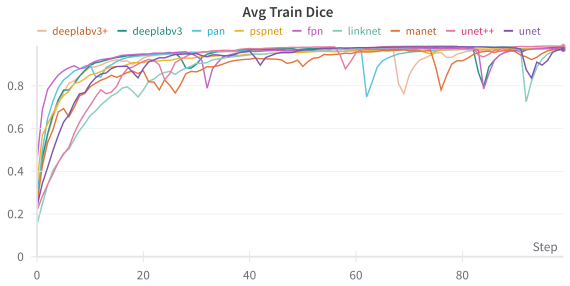
Experiments

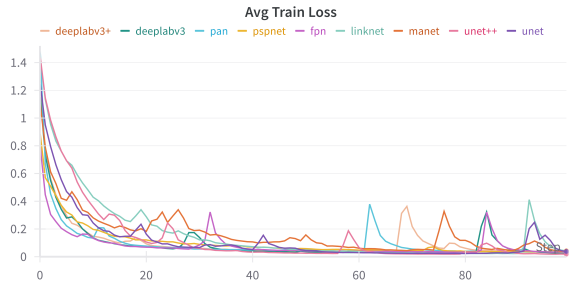
1. Implementation Details

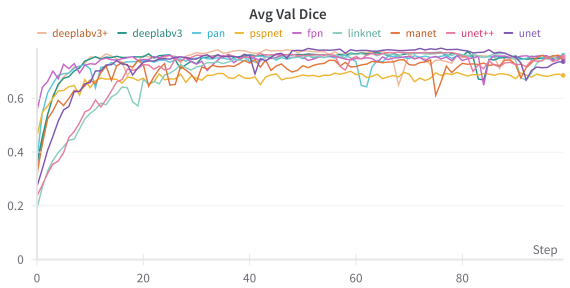
- Backbone: Resnet50, Weights: imagenet
- Split ratio is 0.8, 0.2 for the training set, the validation set.
- Epoch: 100
- Optimizer: Adam
- Learning rate: 1e-4
- Weight Decay: 1e-6
- Weights & Biases to track dice score and loss

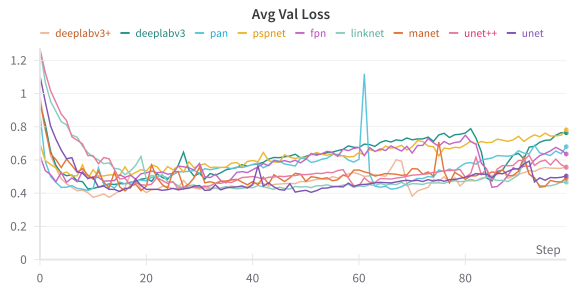
2. Dice Score & Loss

	Train Dice	Val Dice	Train loss	Val loss
Unet	0.9544	0.7713	0.02077	0.4951
UNet++	0.9425	0.787	0.0237	0.4715
MAnet	0.9418	0.7607	0.03968	0.4671
Linknet	0.9278	0.7602	0.02913	0.4455
FPN	0.9234	0.7693	0.06205	0.5184
PSPNet	0.9328	0.7008	0.04972	0.6204
PAN	0.9452	0.7605	0.03251	0.6799
DeepLabV3	0.9481	0.7741	0.03	0.7885
DeepLabV3+	0.9351	0.7819	0.046	0.4419









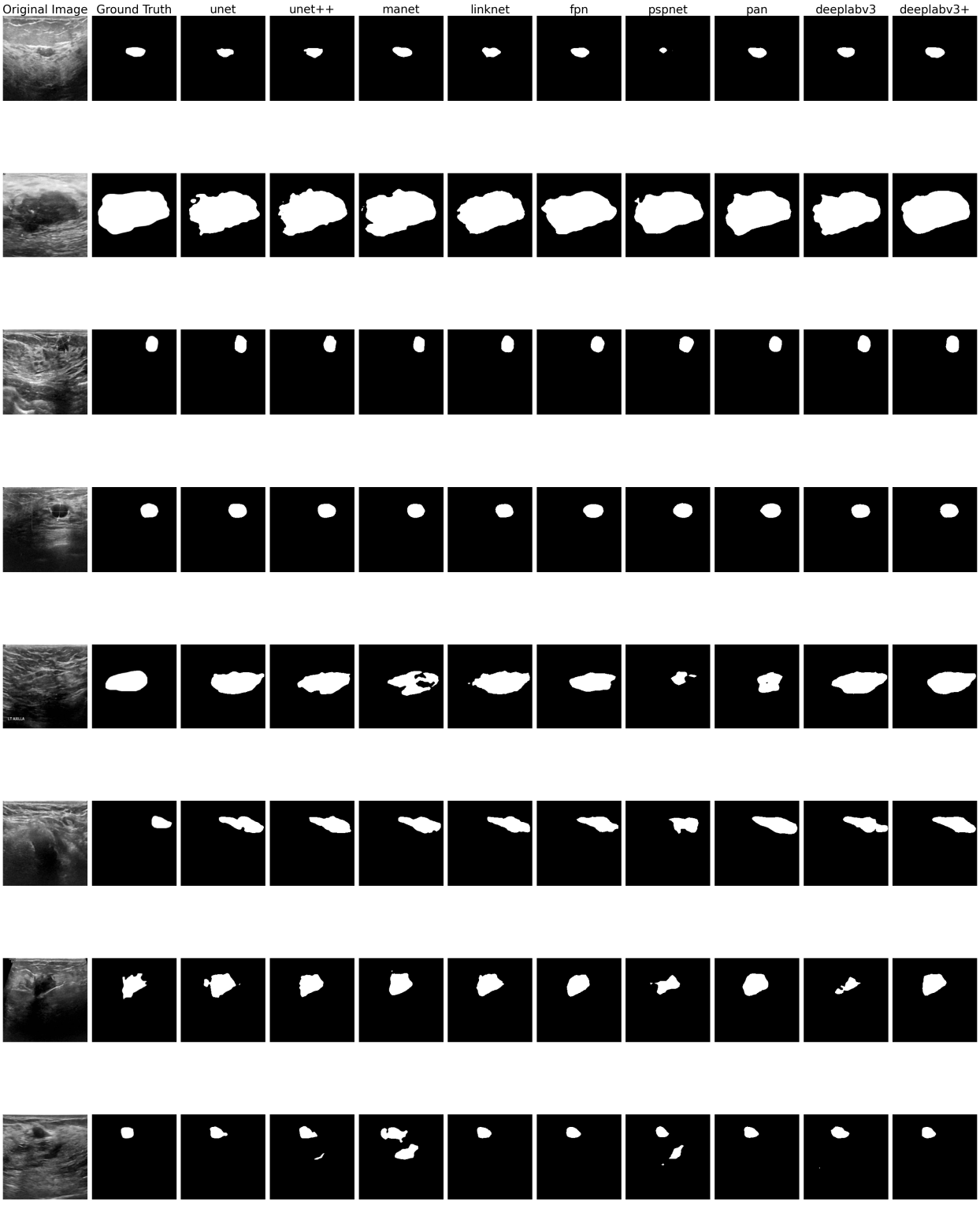
Dice score analysis:

- **UNet++** and **DeepLabV3+** has high Dice coefficients on the validation set while maintaining relatively low validation losses, demonstrating their excellent performance and good generalization ability on this dataset.
- **DeepLabV3** also performs well in terms of validation Dice but its higher validation loss may indicate some instability in performance on certain samples.
- Although **Unet** performs exceptionally on the training set, its performance drops on the validation set, which means overfitting.

Visualization & Analysis

Below, There are some visual results of the segmentation obtained by each model.

It seems **DeepLabV3+, UNet++, FPN** achieve better segmentation results compared to the other models presented.


Original Image
Ground Truth
unet
unet++
manet
linknet
fpn
pspnet
pan
deeplabv3
deeplabv3+

DeepLabV3+

- Sharper and more defined object boundaries, closely matching the ground truth.
- Better handling of textures and structures within the images.
- More consistent shape and contour representation of the objects being segmented.

UNet++

- Fine-grained segmentation with nested skip connections aiding in capturing detailed spatial information.
- Robust segmentation performance, especially around the edges and contours of objects, leading to more accurate object delineation.

FPN

- Effective representation of different object scales, capturing both small details and larger shapes accurately.
- Improved edge detection which helps in distinguishing the foreground objects from the background.

Conclusion

In this homework, I try different deep learning models for segmenting breast ultrasound images from a Kaggle dataset to identify cancerous tissues. Notable models like Unet, Unet++, DeepLabV3+, and so on.

In experiments, it reveal that UNet++ and DeepLabV3+ showed exceptional performance on the validation set, indicating strong generalization capabilities.

These models stand out for their precise boundary and attention to detail, essential for accurate medical diagnosis.

Visual comparisons further demonstrated that DeepLabV3+, UNet++, and FPN produced more accurate and refined segmentation results than their counterparts. This suggests the potential of these models in clinical settings where precise segmentation is crucial for early cancer detection and treatment planning.

Reference

https://github.com/qubvel/segmentation_models.pytorch

kaggle dataset