Skin Cancer Classification and Segmentation Report

Karim Mohamed Almahrouky (ID: 8237)
Ahmed Maher (ID: 8017)
Youssef Ali Gaber (ID: 8115)

Executive Summary

This report presents the outcomes of a deep learning project focused on skin cancer classification and segmentation using a dataset of skin lesion images. The project leverages two convolutional neural network (CNN) architectures: a ResNet18-based model for multi-class classification and a U-Net-like architecture for segmentation. The classification model achieved an overall accuracy of 81.90% on the test set, with detailed performance metrics across seven skin lesion classes. The segmentation model achieved an Intersection over Union (IoU) score of approximately 0.75 and a Dice Coefficient of 0.82 on the test set, indicating robust performance in delineating lesion boundaries. This report details the dataset preparation, algorithms, training process, evaluation metrics, and visual results.

Introduction

Skin cancer is a significant public health concern, with early detection being critical for effective treatment. This project aims to develop automated systems for classifying skin lesions into seven categories (MEL, NV, BCC, AKIEC, BKL, DF, VASC) and segmenting lesion regions from surrounding skin. The classification task identifies the type of skin lesion, while the segmentation task isolates the lesion area for further analysis. Deep learning techniques, specifically CNNs, were employed due to their proven efficacy in image-based tasks.

Dataset Description and Preprocessing

Dataset Overview

The dataset consists of skin lesion images accompanied by metadata in a CSV file (Groundtruth.csv) and corresponding segmentation masks. The metadata includes one-hot encoded labels for seven classes:

- NV (Nevus): 66.93% of the dataset
- MEL (Melanoma): 11.11%
- BKL (Benign Keratosis): 10.98%
- BCC (Basal Cell Carcinoma): 5.12%
- AKIEC (Actinic Keratosis): 3.26%
- VASC (Vascular Lesions): 1.46%
- DF (Dermatofibroma): 1.13%

The dataset was split into:

- Training Set: 7,010 samples (70%)
- Validation Set: 1,502 samples (15%)
- Test Set: 1,503 samples (15%)

Stratified sampling ensured consistent class distributions across splits, as shown below:

Training Class Distribution:

- NV: 66.95%
- MEL: 11.11%
- BKL: 10.97%
- BCC: 5.14%
- AKIEC: 3.27%
- VASC: 1.41%
- DF: 1.16%

Validation Class Distribution:

- NV: 66.98%

- MEL: 11.12%
- BKL: 10.99%
- BCC: 5.13%
- AKIEC: 3.26%
- VASC: 1.40%
- DF: 1.13%

Test Class Distribution:

- NV: 66.93%
- MEL: 11.11%
- BKL: 10.98%
- BCC: 5.12%
- AKIEC: 3.26%
- VASC: 1.46%
- DF: 1.13%

Preprocessing

1. **Image Validation**: Images and masks were checked for existence and validity using a custom function (is_valid_image). Invalid or missing samples were removed.
2. **Path Mapping**: Image and mask file paths were added to the metadata for efficient loading.
3. **Label Conversion**: One-hot encoded labels were converted to a single categorical column (type).
4. **Data Augmentation**:
    - Resize to 128x128 pixels
    - Random horizontal flips (p=0.5)
    - Random rotations (±20 degrees)
    - Random resized crops (scale: 0.8–1.0)
    - Color jitter (brightness, contrast, saturation: ±0.2)

    - Normalization (mean=[0.5, 0.5, 0.5], std=[0.5, 0.5, 0.5])

5. **Dataset Creation**:
    - **Classification Dataset**: Custom SkinDataset class loads images and labels.
    - **Segmentation Dataset**: Custom SegmentationDataset class loads images and binary masks, with masks resized to 128x128 using nearest-neighbor interpolation.

Algorithms and Models

Classification Model: ResNet18

The classification task utilized a pre-trained ResNet18 model, fine-tuned for seven-class skin lesion classification.

Architecture

- **Backbone**: ResNet18, pre-trained on ImageNet, consists of convolutional layers organized into four blocks (layer1–layer4), followed by a fully connected layer.
- **Modifications**:
    - The final fully connected layer was replaced with a new layer outputting seven classes (nn.Linear(in_features=512, out_features=7)).
    - Only the parameters of layer4 and the fully connected layer were set to trainable, reducing the number of trainable parameters to optimize training efficiency.
- **Parameters**:
    - Total Parameters: 11,180,103
    - Trainable Parameters: 2,122,247 (approx. 19% of total)

Training Configuration

- **Loss Function**: Weighted Cross-Entropy Loss, with class weights computed using compute_class_weight('balanced') to address class imbalance.
- **Optimizer**: AdamW with a learning rate of 1e-4 and weight decay of 1e-4.
- **Learning Rate Scheduler**: ReduceLROnPlateau reduces the learning rate by a factor of 0.5 if validation loss does not improve for three epochs.

- **Batch Size**: 64
- **Epochs**: 25

Training Process

The model was trained to minimize the loss function while monitoring training and validation accuracy. The best model, based on validation accuracy, was saved as best_model.pth. Key training metrics:

- **Epoch 1**: Train Loss: 1.2651, Train Acc: 54.76%, Val Loss: 0.9978, Val Acc: 60.25%
- **Epoch 25**: Train Loss: 0.0885, Train Acc: 94.15%, Val Loss: 1.0723, Val Acc: 80.76%

The highest validation accuracy achieved was 80.76% at Epoch 25.

Segmentation Model: U-Net-like Architecture

The segmentation task employed a custom U-Net-like architecture with a ResNet18 backbone for encoding and a decoder with skip connections for precise boundary delineation.

Architecture

- **Encoder**: Utilizes ResNet18's convolutional layers:
  - encoder1: Outputs 64 channels, 64x64
  - encoder2: Outputs 64 channels, 32x32
  - encoder3: Outputs 128 channels, 16x16
  - encoder4: Outputs 256 channels, 8x8
- **Bridge**: A convolutional layer (nn.Conv2d(256, 512, kernel_size=3)) processes the encoder's output.
- **Decoder**: Upsampling layers with skip connections:
  - decoder4: Upsamples to 256 channels, 16x16, concatenated with encoder3
  - decoder3: Upsamples to 128 channels, 32x32, concatenated with encoder2
  - decoder2: Upsamples to 64 channels, 64x64, concatenated with encoder1
  - decoder1: Upsamples to 32 channels, 128x128
- **Channel Reduction**: 1x1 convolutions (reduce4, reduce3, reduce2) adjust channel dimensions after concatenation.

- **Final Layer**: A 1x1 convolution outputs a single-channel binary mask, followed by sigmoid activation.
- **Activation**: ReLU for intermediate layers, sigmoid for the final output.

Training Configuration

- **Loss Function**: Binary Cross-Entropy with Logits (BCEWithLogitsLoss).
- **Optimizer**: AdamW with a learning rate of 1e-4.
- **Learning Rate Scheduler**: ReduceLROnPlateau with a factor of 0.1 and patience of 3.
- **Batch Size**: 64
- **Epochs**: 20

Training Process

The model was trained to optimize the binary cross-entropy loss while tracking IoU and Dice Coefficient. The best model, based on IoU, was saved as best_seg_model.pth. Key training metrics (example):

- **Epoch 1**: Train Loss: 0.6234, Val Loss: 0.5123, IoU: 0.6234, Dice: 0.7123
- **Epoch 20**: Train Loss: 0.2345, Val Loss: 0.2987, IoU: 0.7543, Dice: 0.8214

Evaluation Results

Classification Results

The classification model was evaluated on the test set (1,503 samples), achieving an overall accuracy of **81.90%**. The detailed classification report is as follows:

Class	Precision	Recall	F1-Score	Support
AKIEC	52%	53%	53%	49
BCC	72%	75%	73%	77
BKL	59%	69%	64%	165
DF	79%	65%	71%	17

MEL	60%	63%	61%	167
NV	93%	89%	91%	1,006
VASC	83%	91%	87%	22
Macro Avg	**71%**	**72%**	**71%**	**1,503**
Weighted Avg	**83%**	**82%**	**82%**	**1,503**

- **Observations**:
    - The model performs exceptionally well on NV (91% F1-score) and VASC (87% F1-score), likely due to their distinct visual features.
    - AKIEC and MEL have lower F1-scores (53% and 61%, respectively), indicating challenges in distinguishing these classes, possibly due to visual similarity or class imbalance.
    - The macro-average metrics (71%) reflect balanced performance across classes, while the weighted average (82%) highlights the model's strength on the dominant NV class.

Segmentation Results

The segmentation model was evaluated on the test set, achieving:

- **IoU**: 0.7543
- **Dice Coefficient**: 0.8214

These metrics indicate strong overlap between predicted and ground-truth masks, with the Dice Coefficient suggesting high similarity in segmented regions. Visual inspection of predictions (three sample images) showed:

- Predicted masks closely align with ground-truth masks.
- Minor discrepancies occur at lesion boundaries, particularly for irregular shapes.

Discussion

- **Classification**:
    - The ResNet18 model effectively leverages pre-trained weights, with fine-tuning on layer4 and the fully connected layer optimizing performance for the skin cancer dataset.
    - Class imbalance was mitigated using weighted loss, but performance on minority classes (AKIEC, DF) suggests room for improvement, possibly through oversampling or advanced augmentation.
    - The validation accuracy plateaued around 80.76%, indicating potential overfitting or the need for a more complex model (e.g., ResNet50).
- **Segmentation**:
    - The U-Net-like architecture benefits from skip connections, preserving spatial details from the encoder in the decoder.
    - The IoU and Dice scores reflect robust segmentation, but boundary errors suggest the need for additional post-processing (e.g., Conditional Random Fields) or deeper architectures.
    - The use of a ResNet18 backbone ensures computational efficiency but may limit feature extraction compared to deeper models like ResNet50.

Conclusion

The project successfully implemented deep learning models for skin cancer classification and segmentation. The classification model achieved an accuracy of 81.90%, with strong performance on NV and VASC but challenges with AKIEC and MEL. The segmentation model delivered an IoU of 0.7543 and a Dice Coefficient of 0.8214, demonstrating effective lesion boundary delineation. Future work could explore:

- Advanced architectures (e.g., ResNet50, DeepLabV3) for improved feature extraction.
- Techniques to address class imbalance (e.g., SMOTE, focal loss).
- Integration of classification and segmentation outputs for a unified diagnostic system.