Project Report

Skin Lesion Classification Project

Made by

Ferdinánd André

Barkaszi Richárd Miklós

Cserna Bálint

Team Name

DeepVision

Table of Contents

Documentation....2

Data and Initial Challenges....2

Data Splitting....2

Addressing Class Imbalance Through Augmentation....3

Model Architecture....3

Regularization Techniques....3

Training Process....4

Results and Observations....4

References....5

Documentation

This project aimed to classify skin lesions into benign (non-cancerous) and malignant (cancerous) categories using dermatoscopic images. The data was sourced from the Kaggle ISIC2024 competition, which provides high-quality labeled datasets for dermatological research (Nicholas Kurtansky, 2024). Below, we will provide a detailed overview of the project, highlighting the key challenges and the steps taken to address them.

Data and Initial Challenges

The dataset consisted of high-resolution JPG images, each labeled as either benign or malignant. However, a significant challenge arose from the severe class imbalance in the dataset. Out of the total images, only 393 were labeled as malignant, while the majority represented benign lesions. This imbalance posed a risk of the model being biased towards the dominant benign class, leading to poor performance on the minority malignant class.

To address this issue, we devised a strategy that involved careful data splitting, targeted augmentation for the minority class, and ensuring data safety throughout the process.

Data Splitting

We split the dataset into training, validation, and test sets in a 70-15-15 ratio (Gyires-Tóth, Bálint, 2021). For the malignant images, this meant:

- 275 images were allocated to the *training* set
- 59 images were reserved for *validation*
- 59 images were used for *testing*

The same split ratio was applied to the benign images. Importantly, we ensured there was no data leakage—no image appeared in more than one of these subsets. This careful splitting was essential to ensure a fair and reliable evaluation of the model's performance.

To safeguard the data and maintain clear boundaries between subsets, we stored the images in separate directories. Each subset (train, validation, and test) was organized into subdirectories for the two classes (benign and malignant). This structure not only improved data organization but also allowed for seamless integration with PyTorch's Dataset and DataLoader utilities.

Addressing Class Imbalance Through Augmentation

To tackle the imbalance, we focused exclusively on augmenting the malignant training images. For each of the 275 malignant images in the training set, we generated 9 augmented copies, increasing the malignant training dataset size to 2,750 images. This significantly improved the balance between the two classes, giving the model more opportunities to learn from the minority class.

The augmentations applied included:

- *Random rotation* to simulate varied angles of imaging
- *Horizontal* and *vertical flipping* to introduce variations in orientation
- *Random cropping* and *zooming* to mimic changes in image scale

These augmentations were generated dynamically during training to prevent the model from memorizing specific augmented versions of the images. This approach also introduced additional variability, enhancing the model's generalization ability.

Model Architecture

For the classification task, we decided to use transfer learning techniques because they are more efficient and less computationally demanding than building a convolutional neural network from scratch (Gyires-Tóth, Bálint, 2024). We employed the Vision Transformer (ViT) model (Alexey Dosovitskiy, n. d.), specifically the vit-base-patch16-224 variant. ViT is a state-of-the-art architecture, developed by Google, that leverages transformer-based attention mechanisms to capture both local and global image features. It processes image patches as sequences, which makes it well-suited for handling visual data. Additionally, ViT was pre-trained on ImageNet (Stanford Vision Lab, 2020), a large-scale dataset with millions of labeled images, which further enhanced its ability to generalize to new tasks. This pre-training on ImageNet made ViT an ideal candidate for our task.

We made minimal modifications to the base model, replacing the original classification head with a single-neuron classifier to output a logit for binary classification. This straightforward approach allowed us to utilize the pre-trained features of the ViT model while adapting it to the specific requirements of our dataset.

Regularization Techniques

To improve the model's robustness and prevent overfitting, we incorporated several regularization techniques:

- *Dropout*: We added a dropout layer before the classifier head. This layer randomly deactivates neurons during training, reducing the risk of the model becoming overly reliant on specific features.
- *Layer Freezing*: To focus the training on the most relevant layers, we froze all but the last six layers of the ViT model. This allowed us to fine-tune only the higher-level features while preserving the pre-trained knowledge in the earlier layers.
- *Learning Rate Scheduling*: We used a ReduceLROnPlateau scheduler, which monitored the validation loss and reduced the learning rate by a factor of 0.25 after 2 epochs of no

improvement. This helped the model converge more effectively in later stages of training.

- *Early Stopping*: To prevent overfitting, we stopped training if the validation loss did not improve for 3 consecutive epochs, with a minimum improvement threshold (delta) of 0.001.

Training Process

The training configuration was carefully designed to balance computational efficiency and model performance:

- *Batch Size*: 32
- *Optimizer*: We used the Adam optimizer with:
- *Learning Rate*: 3e-7
- *Weight Decay*: 2e-2
- *Loss Function*: Binary Cross-Entropy with Logits (BCEWithLogitsLoss), which is well-suited for binary classification tasks.

Training was performed on an *NVIDIA A100 GPU* (if available), which significantly accelerated the process, especially given the augmented dataset size. The augmented malignant images were dynamically generated during training, while benign images were loaded directly from their respective directories. The validation set was used to monitor the model's performance after each epoch, guiding learning rate adjustments and determining early stopping.

Results and Observations

The combination of data augmentation, regularization techniques, and the Vision Transformer architecture proved to be highly effective. Augmenting the malignant images and balancing the training data greatly improved the model's ability to detect malignant lesions, addressing the initial class imbalance challenge.

Storing the train, validation, and test data in separate directories ensured that the evaluation process remained unbiased and reliable. The model demonstrated strong performance on both classes, with the malignant class benefitting significantly from the targeted augmentations.

In conclusion, this approach successfully tackled the challenges of class imbalance and generalization, providing a robust solution for skin lesion classification.

References

Alexey Dosovitskiy, L. B. (n. d.). *Vision Transformer*. Retrieved from Hugging Face: https://huggingface.co/docs/transformers/main/model_doc/vit

Gyires-Tóth, Bálint. (2021). *Hatékony Tanítás*. Retrieved from Moodle: https://edu.vik.bme.hu/pluginfile.php/499128/mod_resource/content/1/vitmav45-3sum-hatekony-tanitas-pub.pdf

Gyires-Tóth, Bálint. (2024). *Konvolúciós Neurális Hálózatok*. Retrieved from Moodle: https://edu.vik.bme.hu/pluginfile.php/500770/mod_resource/content/1/vitmav45-7-ConvNets_1D_trans-PUB.pdf

Nicholas Kurtansky, V. R. (2024). *ISIC 2024 - Skin Cancer Detection with 3D-TBP*. Retrieved from Kaggle: https://www.kaggle.com/competitions/isic-2024-challenge

Stanford Vision Lab, S. U. (2020). *ImageNet*. Retrieved from https://www.image-net.org/