Patch-Based Single-Lesion Segmentation for Diabetic Retinopathy Detection

By: Alyssa Abogado

Abstract:

Diabetic Retinopathy (DR) is an eye disease related to diabetes that causes lesions in the retina, damaging blood vessels. To prevent vision loss, early detection for DR is important for effective treatment. In this study, patch-based, single-lesion segmentation targets the four primary lesion types of DR (microaneurysms, hemorrhages, hard exudates, and soft exudates). For single lesion segmentation, four models are trained, each corresponding to one of the lesion categories. Each model takes 128 x 128 patches that are processed using CLAHE and green-channel inputs. The dataset is randomly separated into a 70%/15%/15% train/validation/test split, with the train split aiming for a 60%/40% lesion to healthy sampling ratio. A ResNet-34 encoder pretrained on ImageNet alongside a custom decoder is utilized to construct lesion boundaries. During training, optimization combines weighted BCE and Focal-Tversky losses to combat the strong class imbalance of the dataset. Each model outputs a single-channel probability map indicating confidence level for the target lesion for each patch. Performance analysis shows consistent single-lesion performance with hard exudates and hemorrhages achieving the highest scores, and microaneurysms and soft exudates with the weakest. Detection of tiny and diffuse lesions remains a notable weakness in patch-based pixel-wise segmentation.

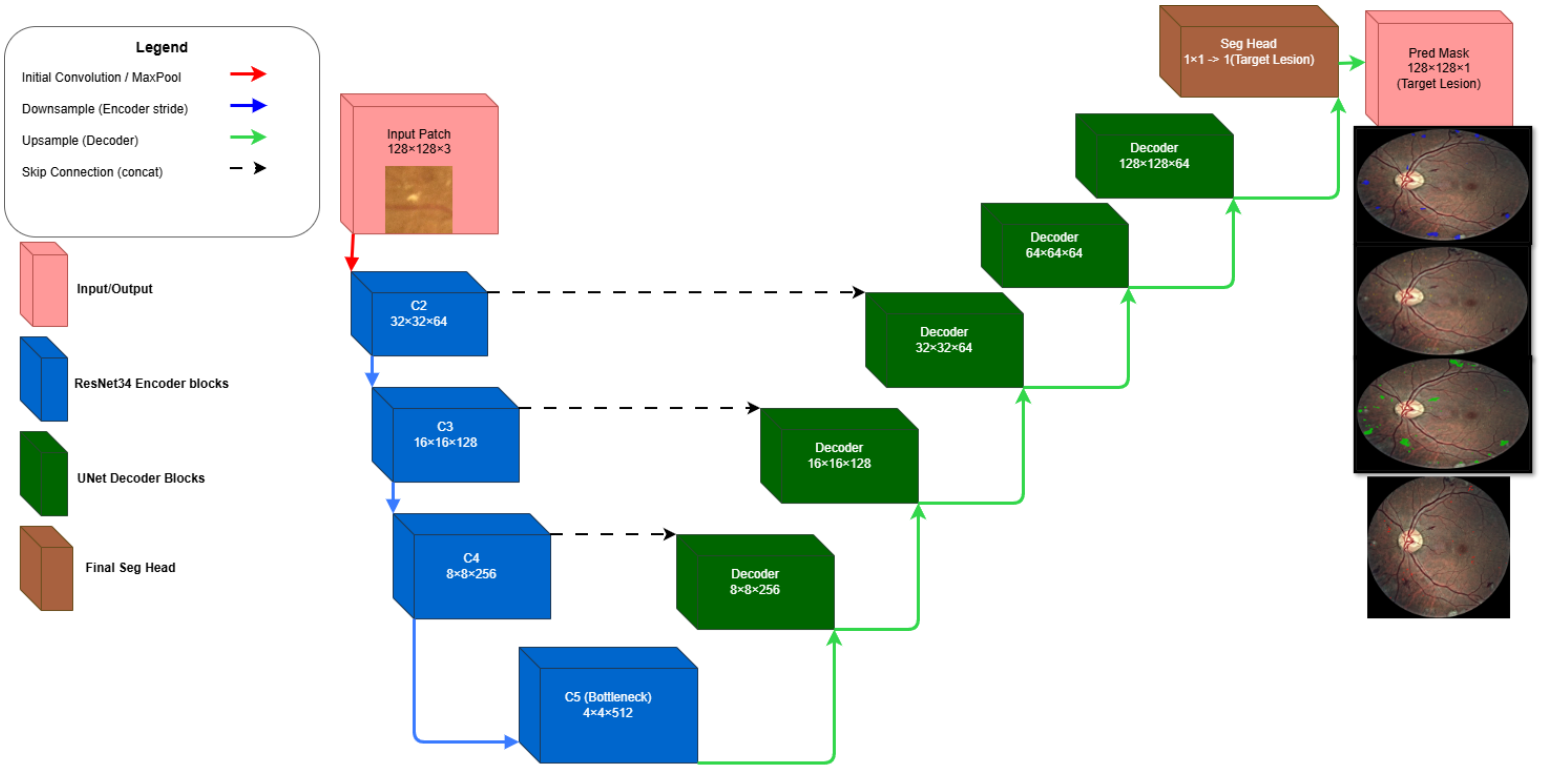
Image Preprocessing:

The dataset includes four lesion-specific binary masks for each fundus image: microaneurysms (MA), hemorrhages (HE), hard exudates (EX), and soft exudates (SE). After each image and their respective masks are split into 128x128 patches and padded for alignment, each patch is enhanced using Contrast Limited Adaptive Histogram Equalization (CLAHE) with a clip limit of 2.0 and an 8 x 8 tile grid. The green channel was chosen as the main representation for best contrast between lesion details and retinal veins. Data augmentation was applied to increase generalization: Horizontal flips ($p = 0.5$), vertical flips ($p = 0.1$), rotations 10°, and padding if needed. ImageNet mean and standard deviation values were used to normalize patches. When sampling at the patch level, each image contributed up to 48 patches in the training set and 96 patches in the validation set. In the training set, sampling was stratified to aim for a 60%/40% lesion to healthy ratio of patches.

Model Architecture:

The model is based on a ResNet34 backbone with ImageNet-pretrained weights with features stages (C2-C5) at special resolutions of ¼, 1/8, 1/16, and 1/32 of the input. A custom encoder-decoder framework with output stride of 8 upsamples these features through residual convolutional blocks with skip connections to progressively upsample feature maps while

preserving spatial detail. The final convolution head (1 x 1) outputs a single-channel probability map for each DR lesion type.



Training Configuration:

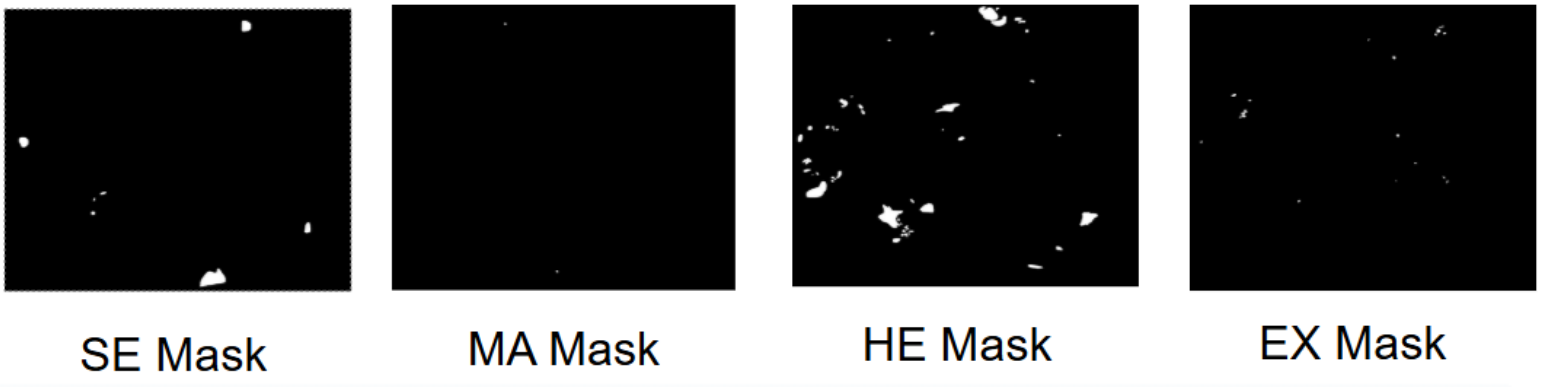
Training is capped at 55 epochs with early stopping (patience = 12). Dynamic thresholding is triggered every 5 epochs. The AdamW algorithm is utilized during optimization with parameters: weight decay = 1 x 10^{-4}, learning rate = 1 x 10^{-4}, gradient clipping at norm = 1.0, Exponential Moving Average (EMA) with decay = 0.999, and Automatic Mixed Precision (AMP). AMP was used to accelerate training and reduce memory consumption, and EMA weights were used during validation to reduce noise and yield smoother generalization estimates. A ReduceLROnPlateau scheduler monitored validation loss, with a lower bound of 1 x 10^{-6} and a learning rate reduction factor of 0.5 upon plateau. This configuration balances convergence speed and generalization stability across all lesion classes.

When deciding losses, Focal-Tversky was chosen for its focus on geometric overlap under class imbalances. BCE-with-logits was chosen to provide dense gradients and probability calibration that Focal-Tversky might struggle with, BCE is especially important for classes with low prevalence. A per-class positive weight vector (capped at 12) was used to further balance rare lesions.

The model was evaluated using Dice and Intersection over Union (IoU) scores. Both metrics measure the overlap between the predicted and ground-truth lesion masks. IoU penalizes predicted lesions that are too large, and Dice is sensitive to false negatives.

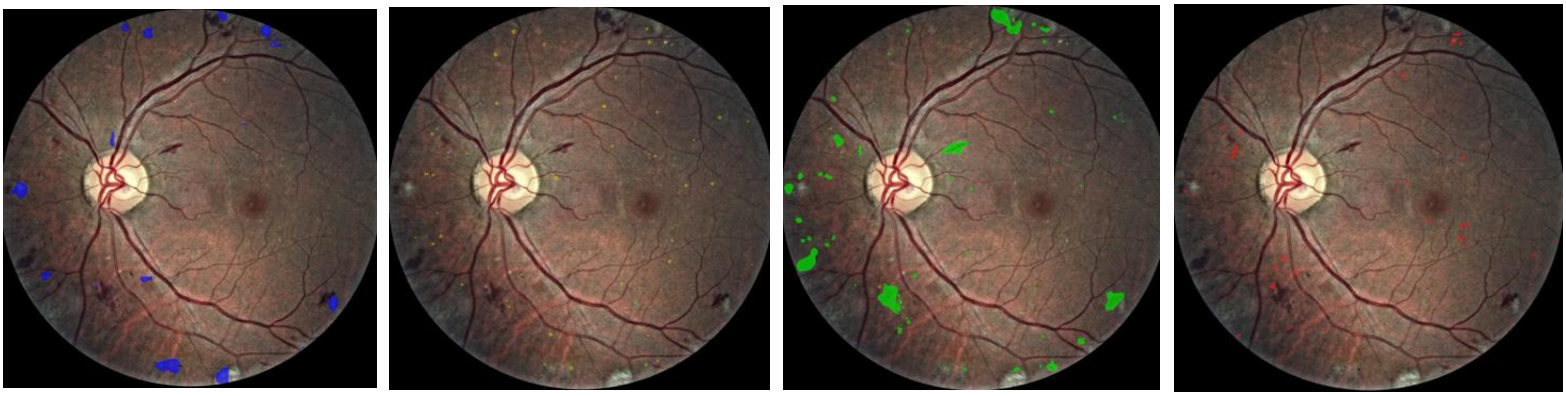
Qualitative results and Error Analysis:

The patch-based segmentation approach achieved varying levels of Dice performance, reflecting lesion shape and prevalence.



Each image (see above) represents the ground truth masks of each lesion within the given original image. The following figures are all four prediction mask overlays on the original fundus image in their respective colors.

SE Mask MA Mask HE Mask EX Mask



Lesion Type	Dice Score	IoU Score
MA	0.10	0.05
HE	0.37	0.23
EX	0.59	0.42
SE	0.25	0.15

As displayed in the predicted overlay comparison vs the ground truth masks, the model achieves relatively accurate hard exudates and hemorrhage region predictions, while suffering with microaneurysm and soft exudates which reflects in their respective Dice and IoU scores. Hard exudates are much easier for the model to predict because of their brightness, sharp borders, and small size. The model performed best on these localized lesions because their small yet contrasting morphology against the background of the patch lets them fit neatly within patches. This allows the model to focus on finer details and texture without background variations that larger lesions would have.

In contrast, hemorrhages and especially soft exudates are much larger than hard exudates and have the disadvantage of spanning over multiple patches. Hemorrhage segmentation performed significantly better than soft exudate segmentation, likely due to the fact hemorrhages are more distinct in shape and boundary than soft exudates. Soft exudates appear faint with lower contrast, lack clear edges, and appear in significantly less patches than hemorrhages, all these factors heavily contribute to the low model scores compared to hemorrhages.

Microaneurysms achieved the lowest scores across all lesion types. This score is consistent with their rarity, size (only a few pixels wide), lack of contrast, and red color that easily blends with the background and any surrounding vessels. Their miniscule size means that any difference by a pixel in predictions can heavily drop Dice and IoU. Additionally, due to their extremely small size, the visual signal is overshadowed by the sheer amount of background pixels, effectively washing them out during downsampling.

Error Analysis continued:

Despite any relative visual accuracy in segmentation masks, the model struggles with class confusion, especially between hard and soft exudates. Both appear brighter than their surroundings, and are classified as one another based on that brightness, revealing that the model favors segmentation by brightness over texture (sharpness of hard exudates versus diffuse borders of soft exudates. This notion is further solidified by the fact hemorrhages and microaneurysms are confused for each other as well, but never with hard or soft exudates. Lesion continuity between patches also suffers because of lack of global image context.

Single-lesion segmentation yielded worse or similar results as multi-lesion. Since each model targets one lesion at a time, specific details such as shared vascular context, inter-lesion relationships, and co-occurrence cues are lost. Without each model learning how different lesions behave around each other, overfitting is especially prevalent with more sparse lesions. Single lesion models inferring their classes in isolation is a major factor in why single-lesion segmentation performs weaker than multi-lesion segmentation.

Future Work:

Further improvements to single-lesion segmentation will focus on incorporating limited context between patches or lightweight cross-lesion attention to mitigate overfitting and improve

contextual awareness. Class-specific augmentations and training on a higher resolution dataset are expected to improve tiny or faint lesions. Additionally, implementing vessel-aware context that considers lesion proximity to retinal vessels could further improve localization accuracy.