CV2 Project - Eye Fixation Prediction

Imran Ibrahimli
Matriculation number: 7486484

Model

The model is based on the UNet architecture [1]. A ResNeXt50 [2] convolutional neural network is used as the backbone (encoder). The total number of trainable parameters is approx. 27M. The encoder passes feature maps from multiple levels to the decoder, where they are concatenated with the appropriately sized feature maps. In particular, 5 levels of feature representation are extracted from the backbone: 64, 128, 256, and 512-channel tensors with decreasing spatial resolution. The backbone can be “frozen” during training, which will prevent the optimizer from updating its weights.

Training

The binary cross-entropy loss function (BCEWithLogitsLoss in PyTorch) was used. The network was trained for 50 epochs and reached a mean validation loss of 0.16. Multiple augmentations were independently applied to training images using the Albumentations library:

- Horizontal flip *(probability of being applied = 0.3)*
- Random rotation with range [-45, 45] deg *(p = 0.3)*
- Random shift, scaling, and rotation *(p = 0.3)*
- Optical distortion *(p = 0.1)*
- Random brightness and contrast adjustment *(p = 0.4)*
- Random hue, saturation, and value adjustment *(p = 0.3)*

Final hyperparameters

Hyperparameter	Value
Optimizer	SGD
Learning rate	0.005
Batch size	64
Backbone frozen	False

Code

https://github.com/iibrahimli/cv2_project

References

[1] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: MICCAI 2015. vol 9351. Springer, Cham.

[2] Xie, S., Girshick, R., Dollar, P., Tu, Zhouwen., He, K. (2016). Aggregated Residual Transformations for Deep Neural Networks. Preprint, arXiv:1611.05431