U-Net: Ear segmentation

Assignment #2

Image Based Biometrics 2020/21, Faculty of Computer and Information Science, University of Ljubljana

Jana Řežábková

***Abstract*—In this report, we present results of ear segmentation on the AWE-W dataset. We created a segmentation pipeline with simple image augmentation preprocessing and a segmentation CNN model that utilizes transfer learning. The model employs U-Net architecture. We use EfficientNet-B0 pretrained on ImageNet as the encoder and custom deconvolution blocks for the decoder. We obtained 99.68 % accuracy and 76.88 % intersection over union (IoU) on the holdout test data.**

I. Introduction

The AWE-W dataset [1] consists of 1000 head shot photos of people. The dataset is split into train set of 750 images and test set of 250 images. Each image is already resized to 480x360 and comes with a same size binary segmentation mask of ears.

The U-Net architecture [2] is a classic architecture for segmentation tasks. We resort to transfer learning, because the dataset size is fairly small and thus build the encoding part of U-Net using an ImageNet pretrained EfficientNet-B0. We use the smallest of EfficientNet models [3] as our computing resources are limited.

II. Methodology

First we rescale the input images to 448x320 to fit our model architecture.

For model training we further split the training data to obtain a validation set (15 % of train data). To enhance the dataset, both in terms of image variety and dataset size, we augment the train and validation images. We augment them separately using random horizontal flips, HSV augmentation and converting ears to grayscale. Each time we keep both the original and the augmented image, thus effectively doubling each dataset size.

The EfficientNet backbone is frozen during training, so that the pre-trained weights are not updated. We use 5 deconvolutional blocks in decoder to bring the output back to input dimensions. Each block is concatenated with a same size layer in encoder to form skip connections. Each block consists of 2D deconvolution layer with 3x3 kernel that halves number of filters and doubles image width and height, followed by batch normalization and ReLU activation.

We don't put a final classification layer and rather calculate the categorical cross entropy loss from logits. We tried to experiment with Dice loss, since our classes are highly imbalanced, but the model did not converge to minima. We use Adam optimizer with default learning rate of 0.001. The batch size is set to 32 and we train over 15 epochs to loss convergence.

The model has total of 11.8M parameters, 4.1M of them belonging to the EfficientNet-B0.

To evaluate model performance throughout training we use IoU and accuracy metrics.

To boost model performance we finetune the converged model by defreezing the last third of layers of EfficientNet and training at 10 times smaller learning rate for another 10 epochs.

III. Results

Figure 1 shows evolution of model loss during training. The jump caused by finetuning is clearly visible. Because validation loss stops improving at epoch 20, we select model at that point to be final.

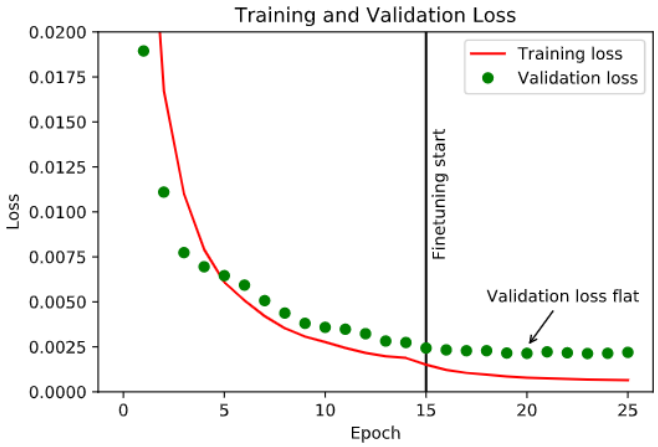

Figure 1. Loss during model training

We achieved 99.68 % accuracy and 76.88 % IoU on holdout test set. Figure 2 displays example test images along with true and predicted mask. These illustrate issues of constructed model caused by both its own inability and incorrect masks.

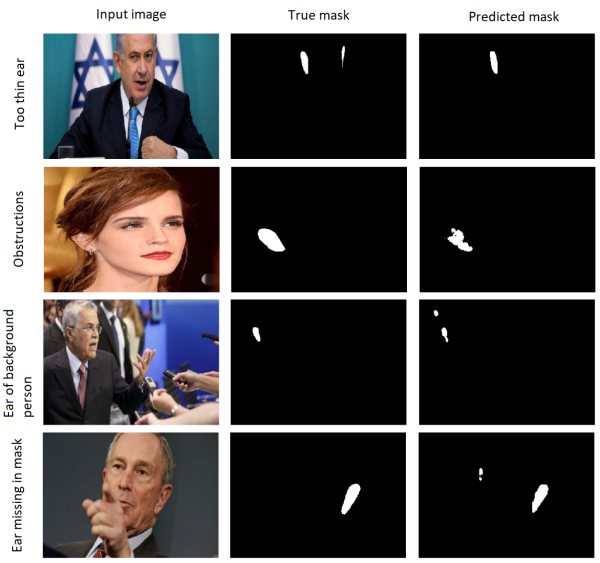

Figure 2. Segmentation issues

The results on validation data were much better. For selected checkpoint we obtained 99.83 % accuracy and 87.91 % IoU. This would suggest there is a difference between the original train and test set, since the validation set is a subset of the test

data and metrics on train and validation data during training correlate well.

IV. Conclusion

We successfuly created a U-Net based model for ear segmenation. Due to resource limitations we haven't explored hyperparameter space and architecture options for the deconvolutional blocks. Nevertheless, we obtained satisfactory results.

Apart from mentioned shortcomings, further investigation could also be done on the issue of generalization to test data.

All code with additional result plots is available in GitHub repository [4].

References

[1] "Annoted Web Ears." [Online]. Available: http://awe.fri.uni-lj.si/
[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015. [Online]. Available: http://arxiv.org/abs/1505.04597
[3] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *CoRR*, vol. abs/1905.11946, 2019. [Online]. Available: http://arxiv.org/abs/1905.11946
[4] "Repository with experiment code." [Online]. Available: https://github.com/janarez/AWE-W/