

GI Image Classification with ResNet-18

BY: SAMIR TISCAREÑO



Kvadir Dataset

Data was collected from the gastroenterology department in the Bærum Hospital, the images are carefully annotated by one or more medical experts.

The multiclass dataset consists of images with different sizes from 720x576 to 1920x1072 pixels that show anatomical landmarks, pathological findings or endoscopic procedures in the GI tract.



Preprocessing

- **Resize:** Training images were resized to 224 x random zoom between 0.8x to 1.2x, validation and testing images were resized to 224 X 224.
- **NumPy array to PyTorch tensor:** Transforms the image into a format suitable for PyTorch, a tensor of shape [C, H, W], where C is channels, H is height, W is width, for example [3, 224, 224] for RGB using ToTensor() function.
- **Normalization:** Image statistics were aligned with the ones of ImageNet dataset, the same dataset which ResNet-18 was pretrained by default. The tensor values were normalized using the mean and standard deviation for each RGB channel.

Preprocessing

- **Flipping images:** Randomly flipped the images horizontally with a 50% probability, to increase variety and helping the model generalize better.
- **Histogram equalization:** Implemented a custom histogram equalization function using OpenCV, converted the images to YUV color space, equalized the Y (luminance) channel, and converted them back to RGB to enhance contrast.
- **Zoom:** Randomly cropped and resized the images to 224x224 with a scale factor between 0.8x and 1.2x. with the intention to act as a zoom augmentation, simulating different magnifications and improving robustness.



ResNet-18

Is a convolutional neural network to classify images, feature extraction and transfer learning. In this case it is applied to detect and analyze the GI diseases, it is useful for object detection or segmenting regions in images.

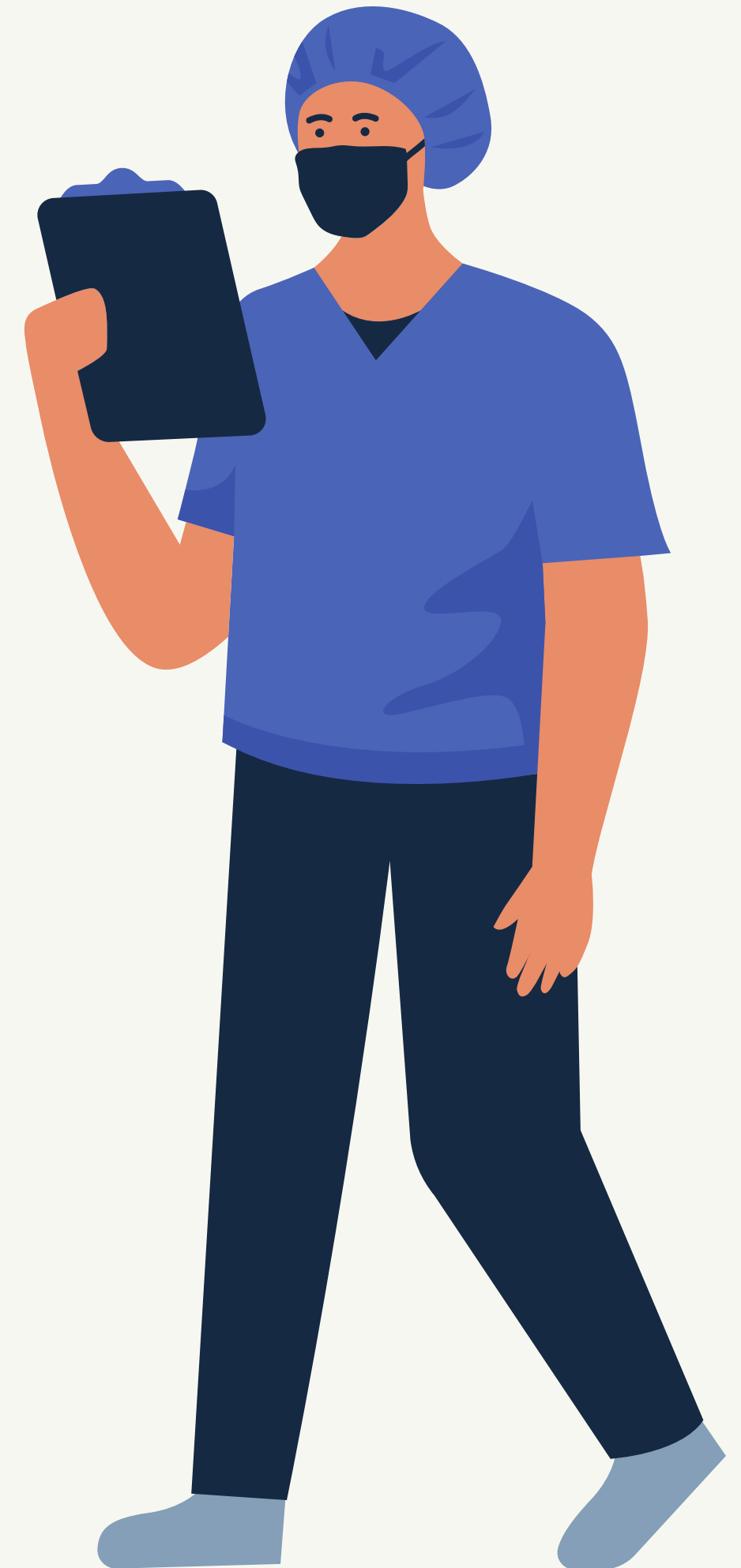


Methodology

The model is trained for 4 epochs, meaning it goes through the entire training set 4 times. It processes the data in batches of 8 images, calculating predictions, computing loss, and updating weights for each batch. During each epoch, all batches are processed once, allowing the model to gradually improve its predictions.

Metrics

- True positive (TP)
- True negative (TN)
- False positive (FP)
- False negative (FN)
- Recall (REC)
- Precision (PREC)
- Specificity (SPEC)
- Accuracy (ACC)
- Matthews correlation coefficient (MCC)
- F1 score (F1)
- Loss (per epoch)





Results

Validation Metric Table

Class	TP	TN	FP	FN	Precision	Recall	Specificity	Accuracy	F1 Score	MCC
dyed-lifted-polyps	101	1049	1	49	0.9902	0.6733	0.9990	0.9583	0.8016	0.7974
dyed-resection-margins	149	1006	44	1	0.7720	0.9933	0.9581	0.9625	0.8688	0.8565
esophagitis	124	1028	22	26	0.8493	0.8267	0.9790	0.9600	0.8378	0.8151
normal-cecum	146	1043	7	4	0.9542	0.9733	0.9933	0.9908	0.9637	0.9585
normal-pylorus	149	1048	2	1	0.9868	0.9933	0.9981	0.9975	0.9900	0.9886
normal-z-line	130	1025	25	20	0.8387	0.8667	0.9762	0.9625	0.8525	0.8311
polyps	139	1043	7	11	0.9521	0.9267	0.9933	0.9850	0.9392	0.9307
ulcerative-colitis	143	1039	11	7	0.9286	0.9533	0.9895	0.9850	0.9408	0.9323
Macro-Averaged	-	-	-	-	0.9090	0.9008	0.9090	0.9008	0.8993	0.8882

Test Metric Table

Class	TP	TN	FP	FN	Precision	Recall	Specificity	Accuracy	F1 Score	MCC
dyed-lifted-polyps	113	1044	6	37	0.9496	0.7533	0.9943	0.9642	0.8401	0.8272
dyed-resection-margins	146	1014	36	4	0.8022	0.9733	0.9657	0.9667	0.8795	0.8658
esophagitis	110	1022	28	40	0.7971	0.7333	0.9733	0.9433	0.7639	0.7326
normal-cecum	149	1043	7	1	0.9551	0.9933	0.9933	0.9933	0.9739	0.9703
normal-pylorus	149	1050	0	1	1.0000	0.9933	1.0000	0.9992	0.9967	0.9962
normal-z-line	123	1009	41	27	0.7500	0.8200	0.9610	0.9433	0.7834	0.7519
polyps	137	1046	4	13	0.9716	0.9133	0.9962	0.9858	0.9416	0.9341
ulcerative-colitis	144	1043	7	6	0.9536	0.9600	0.9933	0.9892	0.9568	0.9506
Macro-Averaged	-	-	-	-	0.8974	0.8925	0.8974	0.8925	0.8920	0.8780



Why does the accuracy is a bit lower?

SUM OF ACCURACIES = $0.9642 + 0.9667 + 0.9433 + 0.9933 + 0.9992 + 0.9433 + 0.9858 + 0.9892 = 7.7850$

MACRO-AVERAGED ACCURACY = $7.7850 / 8 = 0.973125$

$7.7850 / 8 = 0.973125$ 87.7850=0.973125 (97.31%)

THIS STILL DOESN'T MATCH 0.8925. THE ACTUAL MACRO-AVERAGED ACCURACY FROM THE OUTPUT IS 0.8925, INDICATING A POTENTIAL ISSUE.

LET'S INVESTIGATE WHY!

Why does the accuury is a bit lower?

True \ Predicted	dyed-lifted-polyps	dyed-resection-margins	esophagitis	normal-cecum	normal-pylorus	normal-z-line	polyps	ulcerative-colitis
dyed-lifted-polyps	113	36	0	0	0	0	1	0
dyed-resection-margins	4	146	0	0	0	0	0	0
esophagitis	0	0	110	0	0	40	0	0
normal-cecum	0	0	0	149	0	0	1	0
normal-pylorus	0	0	0	0	149	1	0	0
normal-z-line	0	0	27	0	0	123	0	0
polyps	2	0	0	4	0	0	137	7
ulcerative-colitis	0	0	1	3	0	0	2	144

Why does the accuracy is a bit lower?

Test Confusion Matrix								
True \ Predicted	dyed-lifted-polyps	dyed-resection-margins	esophagitis	normal-cecum	normal-pylorus	normal-z-line	polyps	ulcerative-colitis
dyed-lifted-polyps	113	36	0	0	0	0	1	0
dyed-resection-margins	4	146	0	0	0	0	0	0
esophagitis	0	0	110	0	0	40	0	0
normal-cecum	0	0	0	149	0	0	1	0
normal-pylorus	0	0	0	0	149	1	0	0
normal-z-line	0	0	27	0	0	123	0	0
polyps	2	0	0	4	0	0	137	7
ulcerative-colitis	0	0	1	3	0	0	2	144

- Total TP = Sum of diagonal = $113 + 146 + 110 + 149 + 149 + 123 + 137 + 144 = 1071$
- Total samples = 1200 (150 per class \times 8)
- Global accuracy = $\frac{1071}{1200} = 0.8925$
 $\frac{1071}{1200} = 0.8925$ (89.25%)

The "Macro-Averaged" accuracy is the global accuracy (overall correct predictions), not the average of per-class accuracies.

Summary

- 1st epoch -0.6422 loss
- 2nd epoch -0.3585 loss
- 3rd epoch - 0.2879 loss
- 4th epoch - 0.2316 loss





Summary

- A loss of 0.2316 suggesting 76.84% "correctness"
- On average, the model avoids 76.84% of the potential error, leaving 23.16% as the error contribution.

Thanks for
your
attention!

