ELL888 Assignment 1 Report

Siddharth Dangwal
2017EE10497

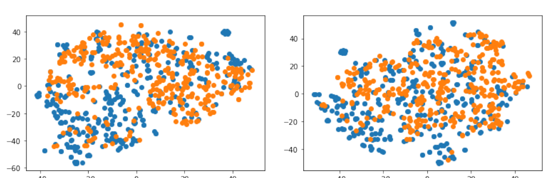
Ajay Baldev Sailopal
2017MT10726

***Abstract*—We try to benchmark the spine dataset (courtesy Prof. Prathosh A.P.) for semantic segmentation and classification tasks. We use various state of the art methods to come up with what we perceive as reasonably good estimates of the metrics for the tasks mentioned above and try to present reasoning for the obtained results.**

***Index Terms*—**

I. Introduction

The aim of this assignment is to provide bench-marking results for the "Spine" data-set which was given to us. We start by describing and visualizing the data. The data is divided into two classes, namely "Normal" and "Damaged". The "Normal" subclass, as the name suggests contains x-rays which do not contain any sorts of injuries/ abnormalities. The "Damaged" subclass consists of images which are blemished. The Damaged subclass contains 328 patient IDs, each of which contains the following: an AP view image of the x-ray, segmentation masks for the vertebra, pedicle and spinous process of the AP view of the x-ray, an LAT image, segmentation masks for the vertebra, anterior vertebral line, disk height, posterior vertebral line and spinous process of the LAT image. The Normal subclass contains 350 patient IDs each composed of the same contents as mentioned above.



(a) AP images (b) LAT images

Fig. 1: The t-SNE visualization of AP and LAT images of both the damaged and normal classes. The damaged samples are shown with blue and the normal ones with orange.

For the segmentation task we use a Unet with an EfficientNet-B3 encoder. For the classification part we use a ResNet that takes in multiple views (both AP and LAT) of a patient ID to determine whether the x-ray is damaged or normal.

II. Classification

The objective of this exercise is to obtain parameters for a model which can predict whether a given patient's x-rays are "Normal" or "Damaged" (as defined in the Introduction).

Identify applicable funding agency here. If none, delete this.

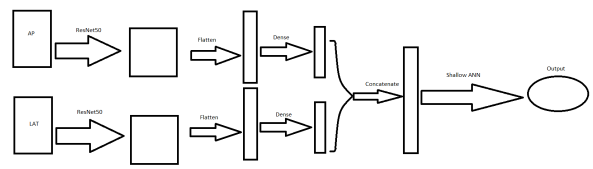


Fig. 2: The LAT and AP images corresponding to each ID are independently passed through a ResNet. The outputs are flattened, passed through a dense layer and then concatenated. The concatenated outputs are passed through a dense layer to obtain the final answer.

A. Model Description

Fig 2 shows the broad structure of the model. An elaborate description can be obtained by calling the $model.summary()$ function of the classification model. The loss used is binary crossentropy

$$J(w) = -(y_t * log(1 - y_p) + y_p * log(1 - y_t)) \quad (1)$$

Here, y_t is the ground truth and y_p is the prediction. The optimizer used is adam.

The intuition behind using the aforementioned structure is to utilize the features of both the LAT and AP views of a particular patient's x-rays to make a decision about it being "Damaged" or "Normal". It may so happen that the defect is occluded in one of the views but is visible in the other one. In such cases, the model that uses both the views to make a decision is bound to give a better result.

B. Results

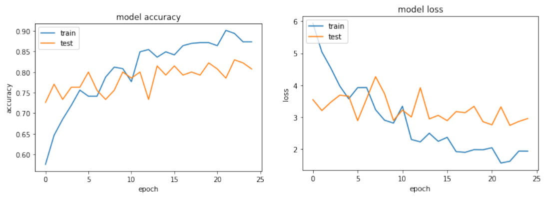
The final train accuracy after 25 epochs is equal to 90.88%. The validation accuracy is 82.22%. The test accuracy reported on the unseen data is

.

Fig. 3 shows the variation of training and validation loss and accuracy with epochs.

C. Observations while conducting classification experiments

Adding Dropout layers increased the validation accuracy. This happened because without the dropout layers, the model



(a) Accuracy v/s Epochs (b) Loss v/s Epochs

Fig. 3: Variation of Training and Validation accuracy and loss with epochs.

was overfitting to the training data. Adding dropout layers caused the model to generalise better as it is equivalent to regularisation, thus increasing validation accuracy. Adam optimiser was used while compiling the model. However, as can be seen in the graph of validation accuracy, the Adam optimiser provides high variation in the training and validation accuracy. SGD optimiser was also used, though it gave less noisy accuracy, the overall accuracy was less than that of Adam, hence Adam was preferred over SGD.

III. Segmentation

The objective of this exercise is to segment the AP view of an x-ray into 'Pedicle', 'Vertebra' and 'Spinous Process' and to segment the LAT view of an x-ray image into 'Anterior Vertebral Line', 'Disk Height', 'Posterior Vertebral Line', 'Spinous Process' and 'Vertebra'.

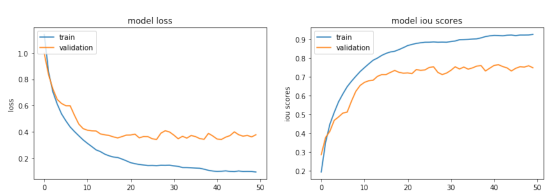
A. Model Description

For all the semantic segmentation tasks, we use the keras library 'Segmentation Models' [1]. We used the inbuilt Unet model with EfficientNet-B3 backbone. An elaborate description can be obtained by calling the $model.summary()$ function of the model.

B. Results

We present the metrics for obtained for various segmentation tasks in TABLE I.

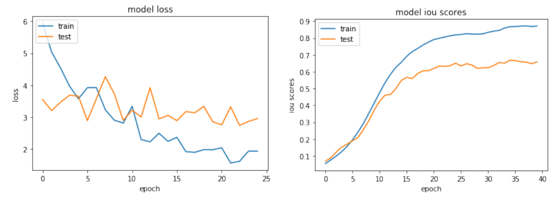
Fig. 4, 6, **??** are the 'loss v/s epochs' and 'IOU score v/s epochs' curves for some of the segmentation tasks:



(a) Loss v/s Epochs (b) IOU score v/s epochs

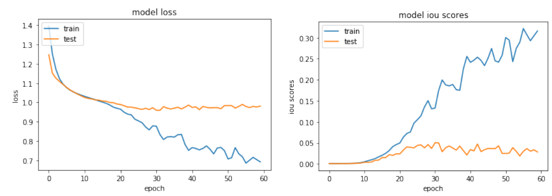
Fig. 4: Curves showing IOU score v/s epochs and loss v/s epochs for the segmentation of Vertebra in AP images.

We observe a wide variety of IOU values corresponding to the eight segmentation tasks given to us. For the given data-set, the IOU values depend on the area of the white pixels present in the ground truth masks of the training data. For the case of



(a) Loss v/s Epochs (b) IOU score v/s epochs

Fig. 5: Curves showing IOU score v/s epochs and loss v/s epochs for the segmentation of Spinous Process in LAT images.



(a) Loss v/s Epochs (b) IOU score v/s epochs

Fig. 6: Curves showing IOU score v/s epochs and loss v/s epochs for the segmentation of Anterior Vertebral Line in LAT images.

"Anterior Vertebral Line" that is visible in the LAT images, we see that the IOU score is rather low (0.02). This is because the ground truth lines are very thin for this segmentation task and the Model is hardly able to learn such fine variations.

The Vertebra segmentation task gives consistently good results for both the AP and LAT images. This is primarily because the area covered by "white pixels" (i.e. the area where instances of a vertebra lie) is the maximum of all the other types of segmentations.

We also provide results for the Spinonus Proccess and see that the values obtained using the IOU metric in this case is somewhat intermediate of the two aforementioned cases. This is becasue the area covered by the spinous process is intermediate of that of the Anterior Vertebra Line and Vertebra.

IV. Conclusion

Thus we have succesfully benchmarked the Spine dataset. The benchmarking data presented might not be reflective of the best possible achievable metric numbers, however these are values that can be comfortably reached with using state of the art methods.

References

[1] P. Yakubovskiy, "Segmentation models," https://github.com/qubvel/segmentation$_{m}odels$, 2019.

Sr. No.	Mask	(Train IOU, Train Loss)	(Val IOU, Val Loss)	(Test IOU, Test Loss) (split from training data)	(Test IOU, Test Loss) (data released recently)
1	AP Pedicle	(0.6201, 0.4101)	(0.5912, 0.4413)	(0.5602, 0.4828)	(0.5413, 0.5092)
2	AP Spinous Process	(0.3958, 0.6083)	(0.2347, 0.7809)	(0.2559, 0.7638)	(0.2236, 0.7961)
3	AP Vertebra	(0.9264, 0.0931)	(0.7488, 0.3773)	(0.763064, 0.34893)	(0.7542, 0.3803)
4	LAT Anterior Vertebral Line	(0.3147, 0.6926)	(0.0284, 0.9810)	(0.0202, 0.989)	(0.0219, 0.9874)
5	LAT Disk Height	(0.8634, 0.1482)	(0.6375, 0.4174)	(0.6240949, 0.435335)	(0.5924, 0.4800)
6	LAT Posterior Vertebral Line	(0.2030, 0.8038)	(0.0276, 0.9804)	(0.04017173, 0.968718)	(0.0251, 0.9822)
7	LAT Spinous Process	(0.8718, 0.1426)	(0.6582, 0.4182)	(0.6572, 0.4169)	(0.62414, 0.4673)
8	LAT Vertebra	(0.8975, 0.1266)	(0.7787, 0.2968)	(0.7757, 0.3019)	(0.7444, 0.3618)

TABLE I: Semantic Segmentation Results