

PS07 Report

	Test 1 (EN & ZN)	Test 2 (EN & TH & ZN)	Test 3 (without RU & DA)	Test 4 (without DA) (with cropped image)
Number of epochs	5	5	5	5
Number of models	14	14	14	14
Number of convolutional filters(for two model)	(63, 128)	(100, 128)	(64, 128)	(12, 30)
kernel_size(for two model)	(3, 3)	(3, 1)	(3, 1)	(3, 3)
strides	2	2	2	2
Dense(relu)	512	520	520	120
Dense(softmax)	2	3	2	4
Trainable params	14,823,048	29,370,479	29,364,912	925,938
Accuracy	1.0	0.9935	0.6438	0.9855

After the first four tests, we can 100% accurately separate EN and ZN. When adding TH, we can still reach 99% accuracy. But for EN and DA, we only have 64.38% accuracy, because EN and DA are too similar. For test four, we used all of the language without the DA and got 98.55% accuracy.

According to the misclassified images above, we believe the reasons why it thinks the above Chinese images are English are as follows: first, there are numbers in some images, which can also be found in English images; second, there may be some characters that are too simple to be recognized as Chinese; and finally, "?" can appear in some English images, leading the model to believe it is English. We experimented with several layers, altering settings and image numbers. We received 100% of the finest outcomes. We encountered issues such as excessive loss of computers.

For our technical side, We got a Macbook Pro with an M1 pro chip, which runs super quickly after setting up GPU support for tensorflow. We did not cut down the dataset and we used all the images to do this project. I ran 5 epochs in total. Because of the lack of time, we could have done more epochs. Note: point(2a) is about making your best reasonable effort. This point is about explaining these efforts well.

Below is the final result.

Final results:

	Number of epochs	Number of models	Number of convolutional filters	Dense (relu)	Dense (softmax)	Trainable params	Accuracy
Final results	5	14	12, 24	128	5	1, 232,667	0.824

Confusion Matrix:

```
confusion matrix
predicted category
DA      EN      RU      TH      ZN
DA      821     198      0       0       0
EN      311    1739      0       0       0
RU      473     143    1025      1      51
TH       44      70       2    1783       0
ZN       80      20       2       7    1197
Accuracy 0.8240240994100665
```

Comments:

It took us 10 hours. The other tests' confusion matrix and accuracy are all in the code file. Thank you so much for correcting our homework throughout the quarter and have a great summer!