Deep Learning (COSC 2779) – Assignment 1 – 2021

Jakrapun Sangchan (s3808216)

1 Problem Definition and Analysis

Human activity recognition is a task that attempts to identify a person's actions based on the data. In this assignment, the human activity will be identified bases on a still image. The dataset has 3,030 images. This task is multi-classes classification task. However, the dataset has 2 target variables action (21 classes) and action class (5 classes). The aim of this assignment is creating a predictive model which is predict both action and action class at the same time.

2 Evaluation Framework

The task in this assignment is human activity recognition which is multi-class classification. For the classification problem, the two main evaluation metrics is accuracy and F1-score. In this problem, the **macro average F1-score** was use as an evaluation metrics of both tasks with two main reasons. First reason is the class distribution are imbalance, when the class imbalanced accuracy would over emphasise on majority class while F1 score will be less bias on majority class.

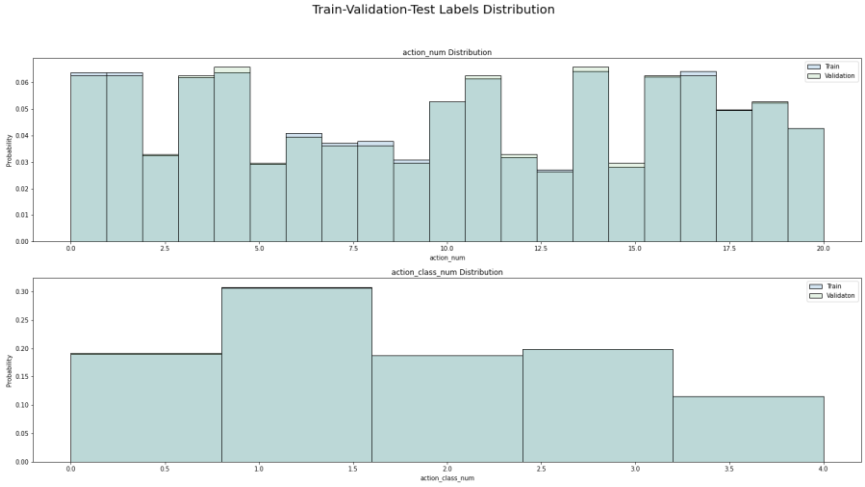

Fig 1. Train, validation and test labels distribution.

The second reason is the human activity recognition task all class is equally important and macro F1-score can represent each classes equally. However, for final evaluation the macro f1-score and confusion metrics were used to evaluate model performance.

The dataset was randomly split into train (2,424), validation (303) and test (303). The train and validation set were used to determine hyperparameter tuning such as regularisation and select the weight of the model from epoch. Test set was used to evaluate across the model and make ultimate judgement.

3 Approach & Justifications

The models that used for this project were base on resnet-34 and resnet-18 structure but replace fully connected part by 2 FC layers with 21-d (actions) and 5-d (action classes) to serve as multi-labels classification. The resnet-34 was use as a base model because the dataset considers as small compare to the ImageNet dataset. Another reason, resnet-34 is the smallest resnet architecture that still shows significant benefit of residual part in the paper [1]. However, due to the small training set we still considered resnet-18 with less model capacity and compare it with our base model. The third model (fig x.) was developed base on resnet-18 structure by added 3 residual blocks of [3*3, 512] on the action class classification task. Because the action class has more samples per label and more likely to increase performance when increase model capacity than action classification with less samples per label. The data augmentation and model regularization were used to reduce overfitting of the model. The random rotation between -10 to 10 degree and shift between -5 to 5 on both axes have been use to created augmentation dataset.

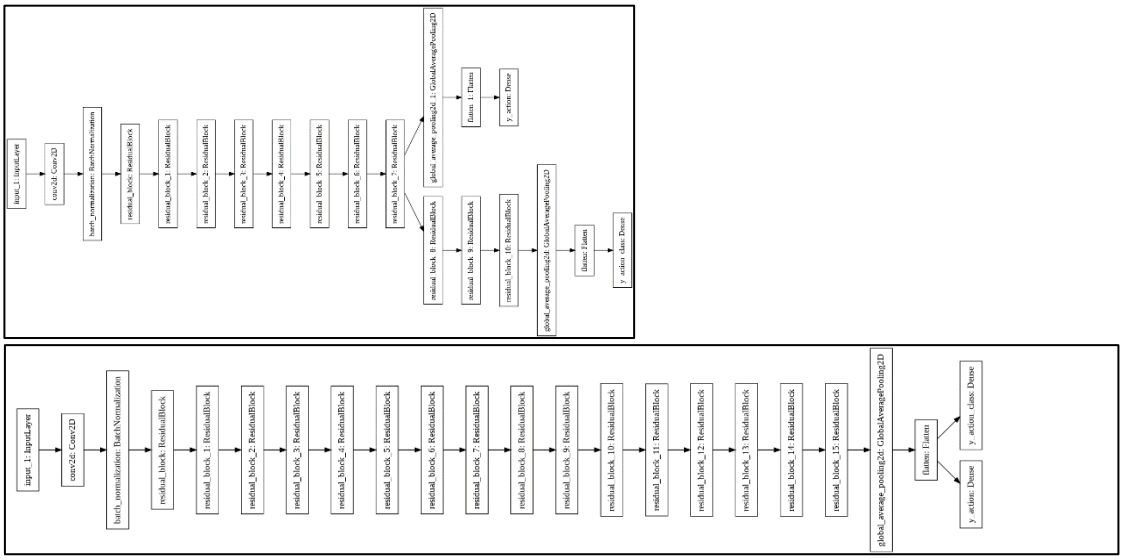

Fig 2. Model architecture Top: Resnet-18 with 3 additional residual blocks Bottom: resnet-34.

4 Experiments & Tuning

- Firstly, the resnet-34 was used for base model by training with only original training set. However, the result shows the significant overfitting.
- To reduce the overfitting the augmentation dataset has been combined with training set to increase the number of training set. The result still shows overfitting but it has better performance on test set.
- To make the model more generalize the regularization has been use on the augmentation dataset the lambda has been tuning start with 0.05, 0.001, 0.005 and 0.01 respectively.
- The result shows the 0.05 is too large and eliminate all the weight in the model so the model cannot reduce any loss. On the other hand, lambda 0.001 is too small it still has significant overfitting. Lambda equal 0.005 shows slightly overfitting. The best lambda is 0.01, the result show that after 25 epochs the model still not overfitting. However, resnet-34 with 0.01 lambda have worse performance on test set than resnet-34 with augmentation.
- Therefore, I decided to reduce model capacity by used resnet-18 with augmentation. The resnet-18 show better performance on action prediction than resnet-34 but worse performance on class prediction.
- Next step, I decided to increase complexity only for action class task because the resnet-34 which is higher complexity shows better performance than resnet-18 on action class classification by add 3 residual blocks on the action class classification part. However, the result on test set is still worse than resnet-34 and resnet-18.

Regularisation result

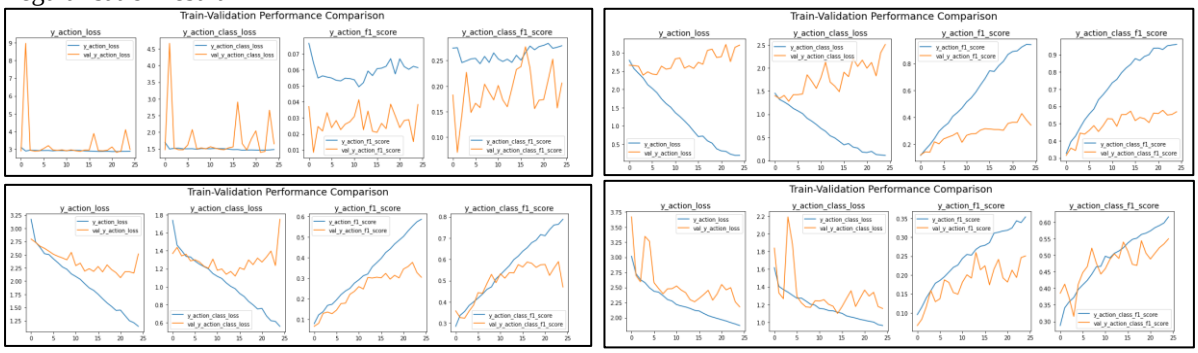

Fig 3. Training Resnet-34 with difference lambda. Blue line denotes training performance, yellow line denotes validation performance. Top left: lambda = 0.05, Top right: lambda = 0.001, Bottom right: lambda = 0.005 Bottom left: lambda = 0.01

5 Ultimate Judgment, Analysis & Limitations

According to the table1, the Resnet-34 with augmentation test set performance was 0.68 macro F1-score on task 1 and 0.42 macro-f1 score on task 2, while Resnet-18 with augmentation test set performance was 0.65 macro F1-score on task 1 and 0.43 macro-f1 score on task 2.

	Task 1 action class classification	Task 2 action classification
Resnet-34 (base model)	0.52	0.29
Resnet-34 with augmentation	0.68	0.42
Resnet-34 with regularisation and augmentation	0.61	0.38
Resnet-18 with augmentation	0.65	0.43
Extened resnet-18 with augmentation	0.65	0.42

Table 1. Test set performance

The confusion matrices were use to determine the final model. Fig 4 shows Resnet-34 with augmentation have better task 1 performance than Resnet-18 with augmentation on Interacting with animal and other activity and on tie with other classes.

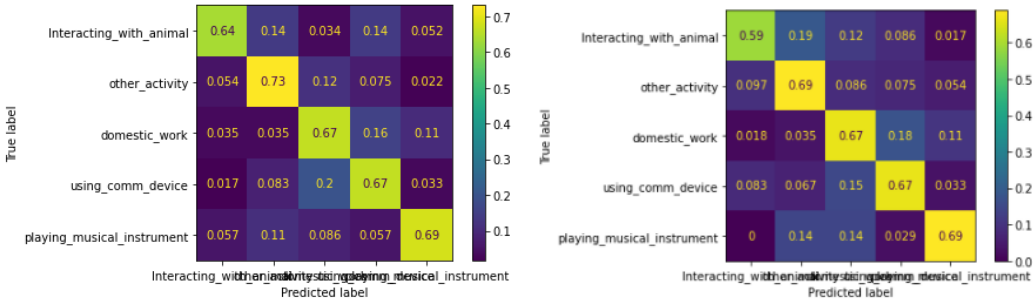

Fig 4. Confusion metric of action class classification task left: Resnet-34 with augmentation right: Resnet-18 with augmentation

On task 2, Even Resnet-18 can achieve a bit higher macro F1-score than Resnet-34 (0.43 and 0.42) but the confusion matrix shows Resnet-18 have 2 classes that have nearly 0 detection (shooting an arrow and texting message) while, Resnet-34 have only texting message that model cannot predict correct.

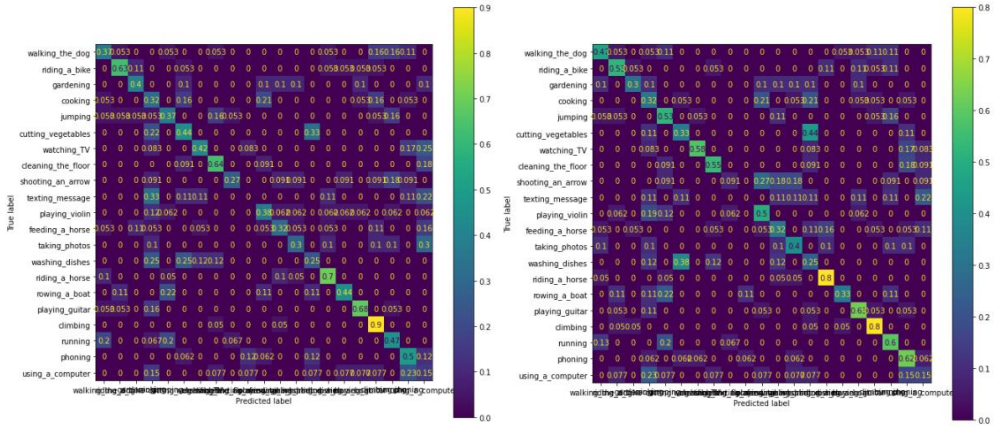

Fig 5. Confusion metric of action classification task left: Resnet-34 right: resnet-18

To conclude, the Resnet-34 with augmentation will be final model with 2 reasons Resnet-34 have higher macro F1-score in task 1 and have tie performance on task 2.

6 References:

[1] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition", *arXiv.org*, 2021. [Online]. Available: https://arxiv.org/abs/1512.03385. [Accessed: 10- Sep- 2021].