

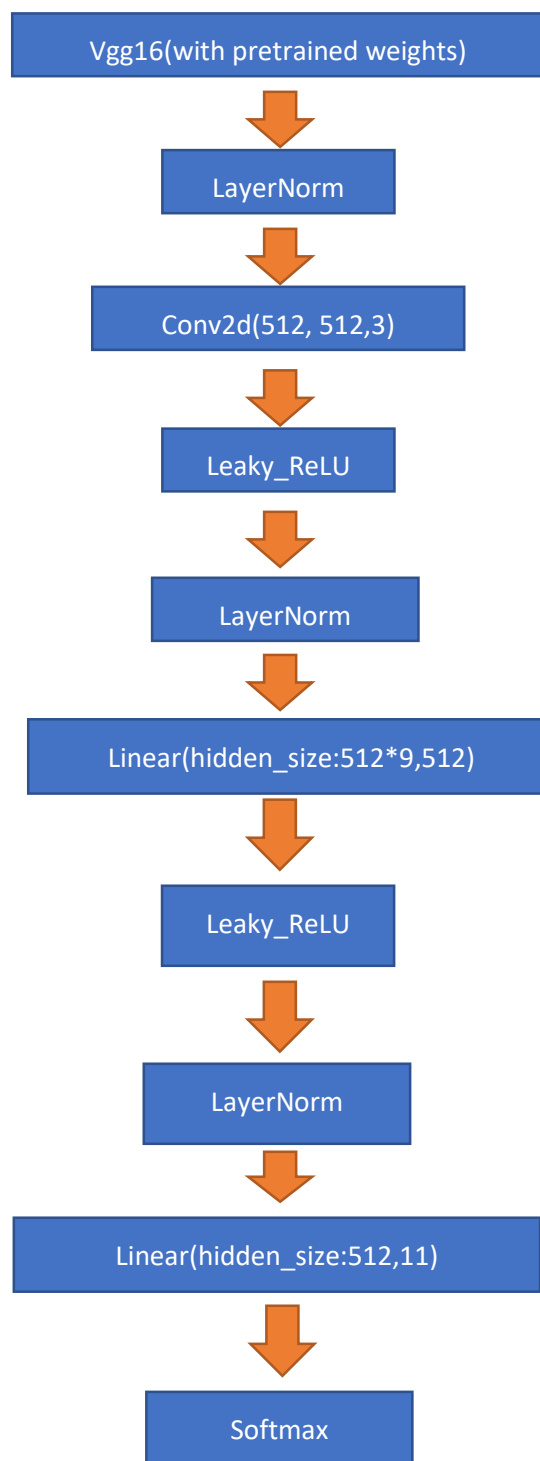
R06945055 王擎天

Problem 1 : Data preprocessing (20%)

Describe your strategies of extracting CNN-based video features, training the model and other implementation details (which pretrained model) and plot your learning curve (The loss curve of training set is needed, others are optional). (5%)

Strategies: CNN-based video feature 為利用第一個 frame、中間的 frame、和最後一個 frame，並設定 rescale_factor 為 0.5 製作而成。

Model:

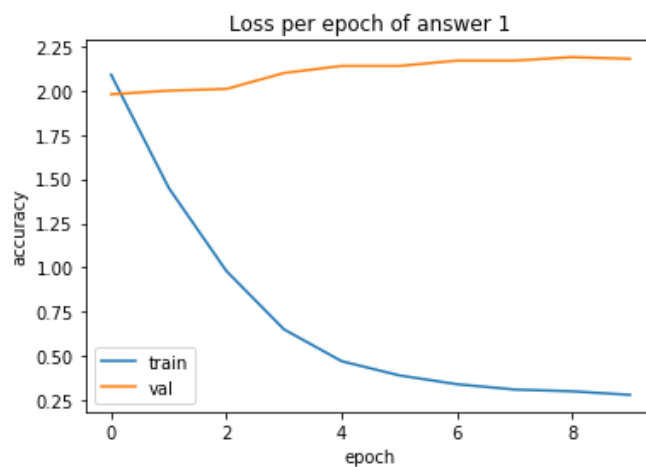
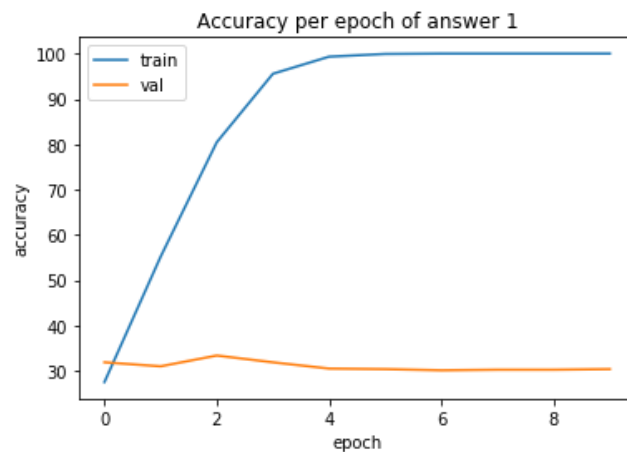


Hyperparameter:

Optimizer: Adam(lr:1e-5)

Batch size:32

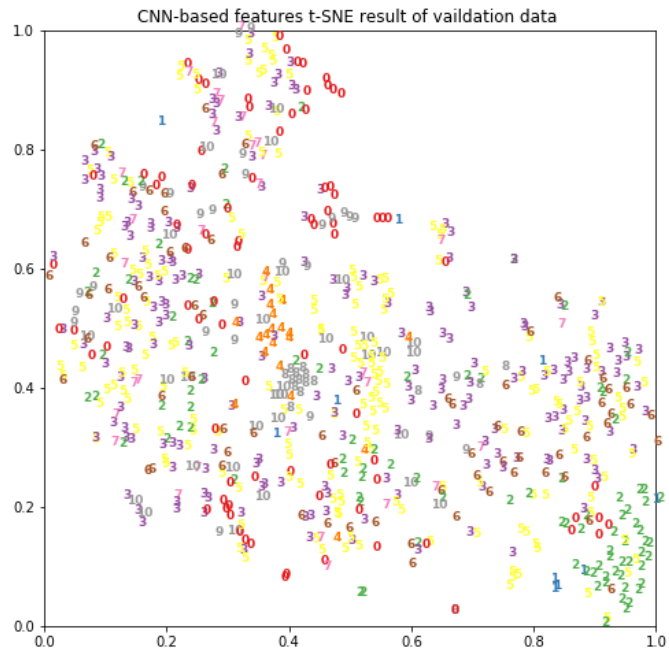
Epoch:10



Report your video recognition performance (valid) using **CNN-based video features** and make your code reproduce this result. (5%)

Valid accuracy: 31.6%

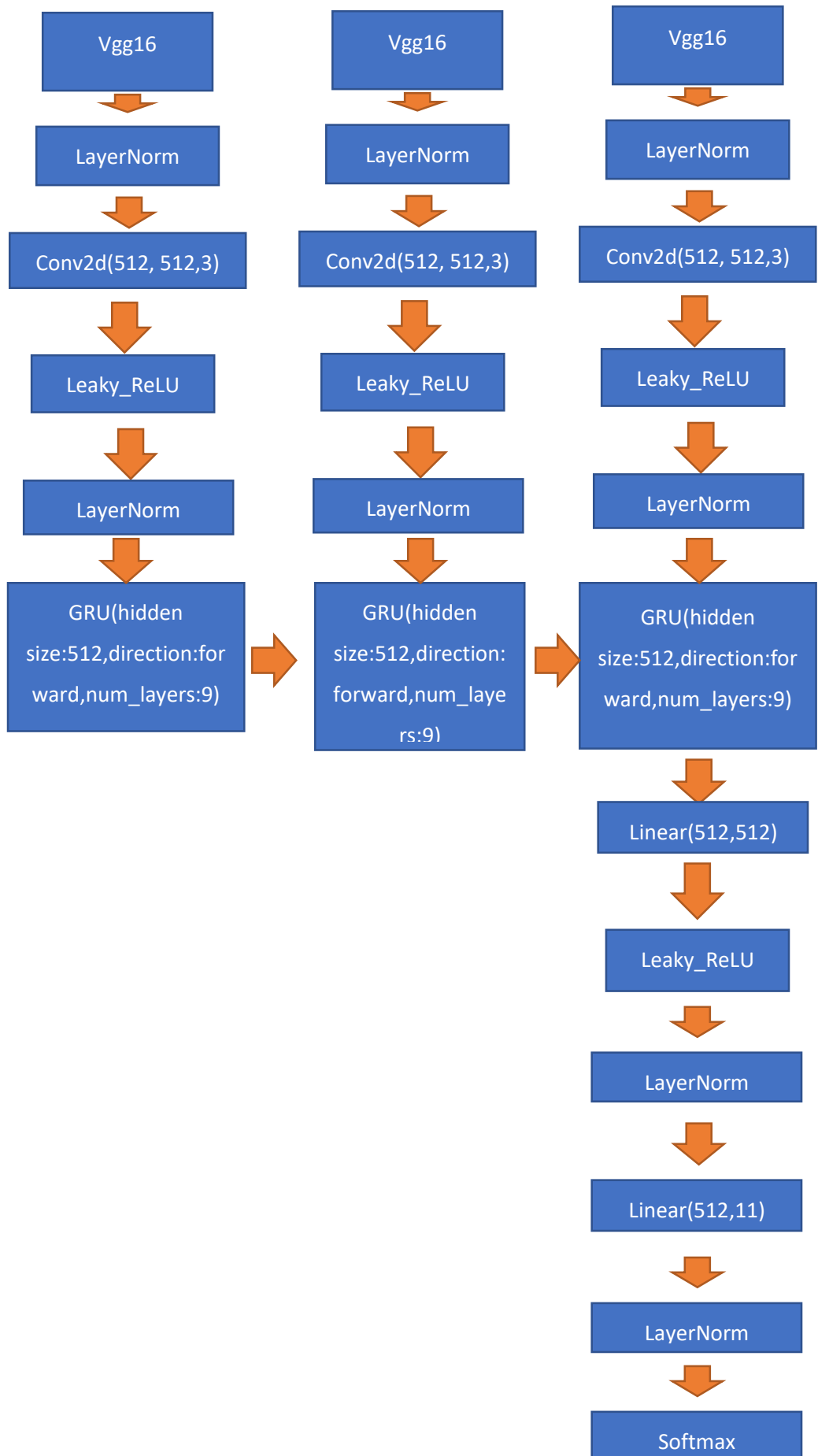
Visualize **CNN-based video features** to 2D space (with tSNE) in your report. You need to color them with respect to different action labels.(10%)



Problem 2 : Trimmed action recognition (40%)

Describe your RNN models and implementation details for action recognition and plot the learning curve of your model (The loss curve of training set is needed, others are optional). (5%)

Strategies: RNN-based video feature 為利用全部的 frame，並設定 `rescale_factor` 為 0.5，`downsample_factor` 為 12 製作而成。



Weights initialization:

Linear:Normal(0,0.02)

LayerNorm:Normal(1,0.02)

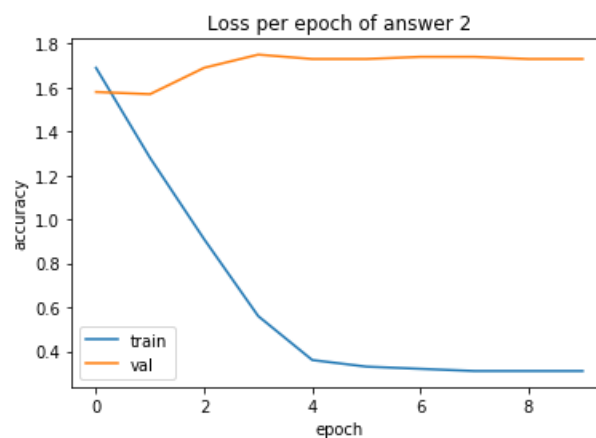
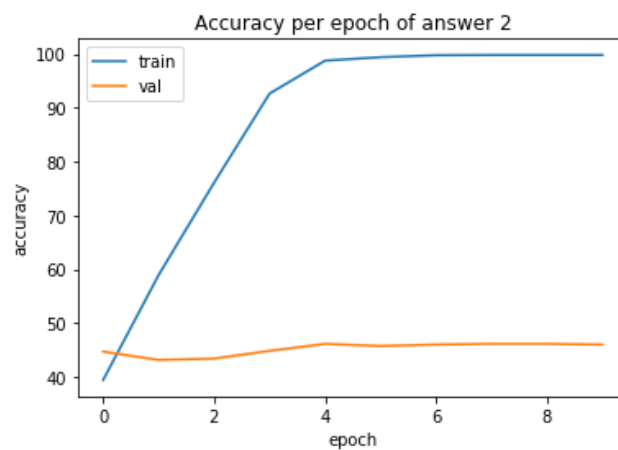
CNN:Normal(0,0.02)

Hyperparameter:

Optimizer: Adam(lr:1e-5, amsgrad=True))

Batch size:1(將影片以 256 time step 為單位切割，並將前一個 hidden states 保留下來給下個單位)

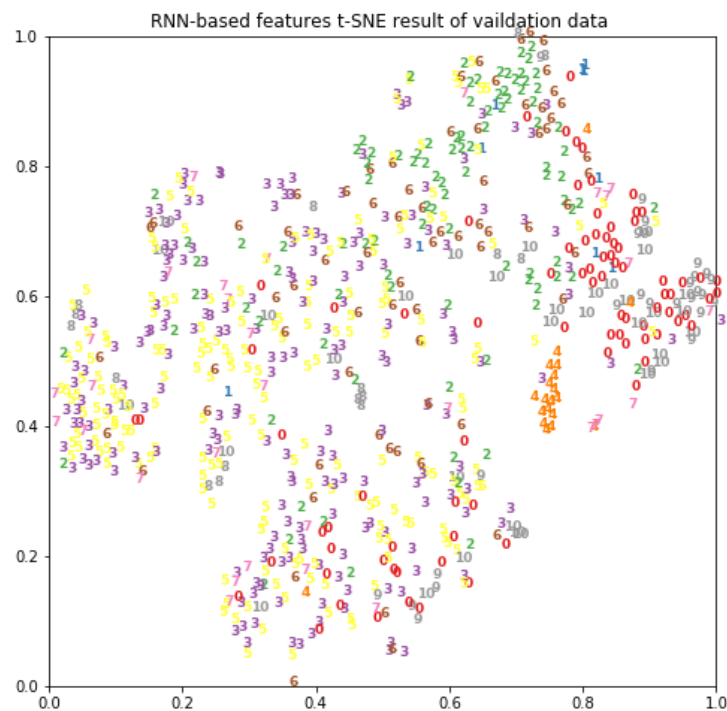
Epoch:10



Your model should pass the baseline (valid: 0.45 / test: 0.43) validation set (10%) / test set (15%, only TAs have the test set).

Valid accuracy:46.6%

Visualize **RNN-based video features** to 2D space (with tSNE) in your report. You need to color them with respect to different action labels.

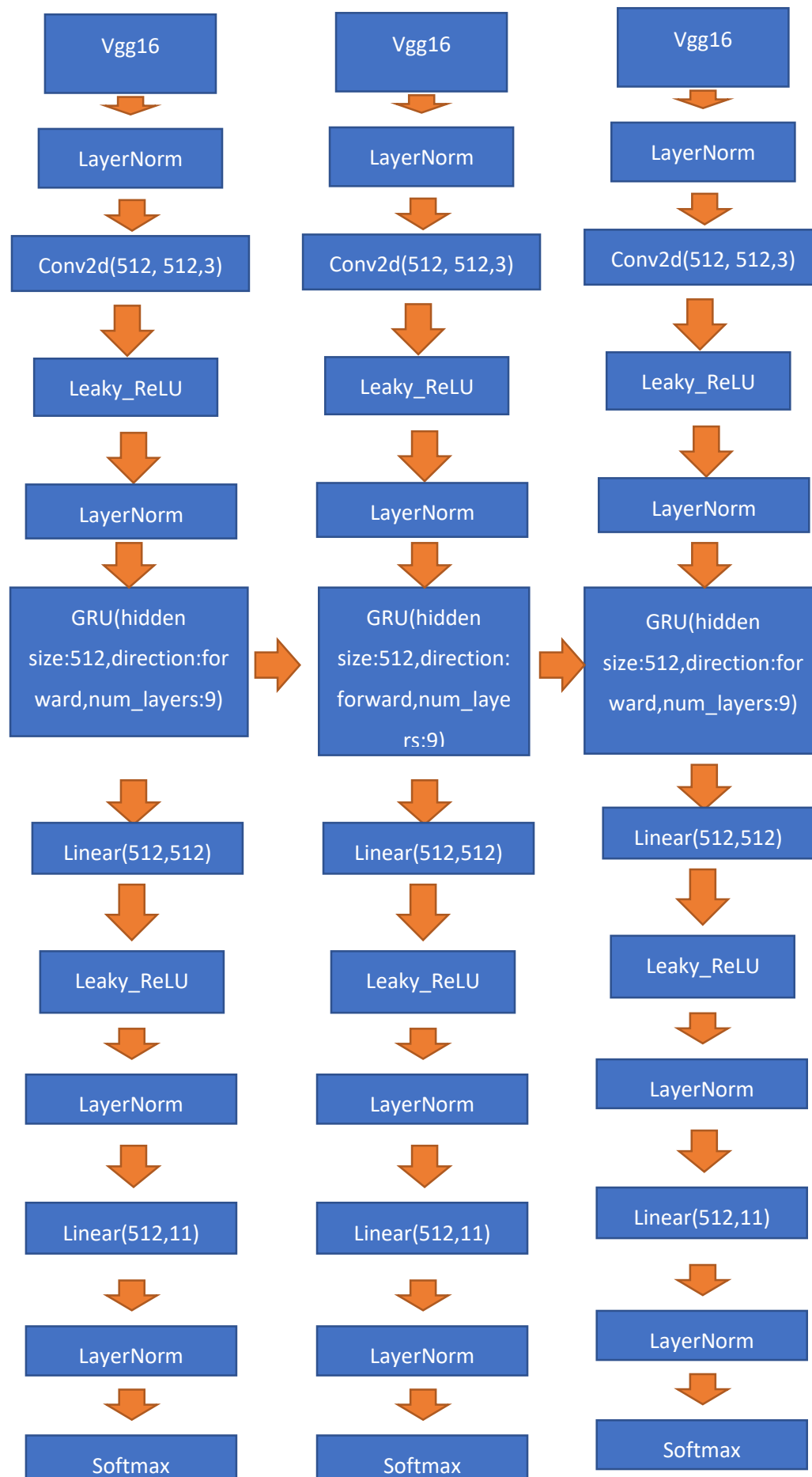


Do you see any improvement for action recognition compared to **CNN-based video features** ? Why? Please explain your observation (10%).

RNN-based feature:在 accuracy 上改善 15%，從 tSNE 圖可看出 RNN-based 可以將資料分布分得較開(相對 CNN-based)。

Problem 3 : Temporal action segmentation (40%)

Describe any extension of your RNN models, training tricks, and post-processing techniques you used for temporal action segmentation. (5%)



Weights initialization:

Linear:Normal(0,0.02)

LayerNorm:Normal(1,0.02)

CNN:Normal(0,0.02)

Hyperparameter:

Optimizer: Adam(lr:1e-5))

Batch size:1(將影片以 256 time step 為單位切割，

並將前一個 hidden states 保留下來給下個單位)

Epoch:10

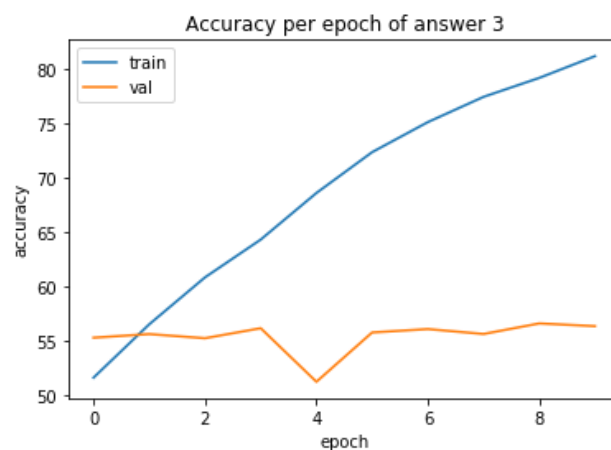
Post-processing techniques:

如果一個 label 的前面一個 label 和後面一個 label 依樣就將該 label 設為前一個 label，否則照舊。

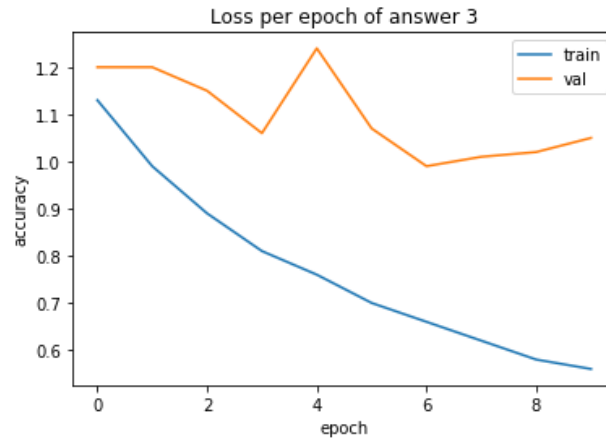
Report validation accuracy in your report and make your code reproduce this result.

(20%)

Accuracy (after post-processing) of all video accuracy: 57.84291787872007%



(without post-processing)

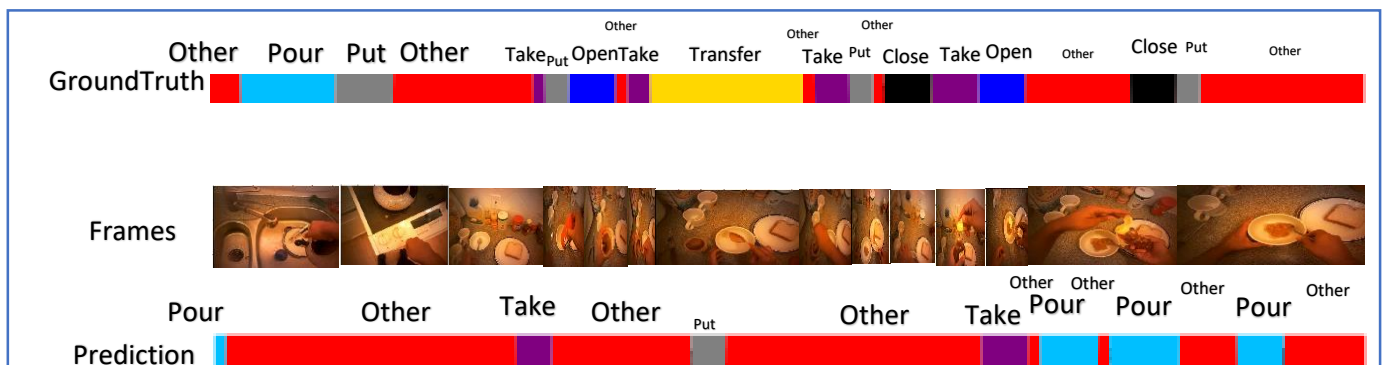


(without post-processing)

Choose one video from the 7 validation videos to visualize the best prediction result in comparison with the ground-truth scores in your report. Please make your figure clear and explain your visualization results (You need to plot at least 500 continuous frames). (15%)

Color from Label:0 to 10: ["red", "green", "blue", "purple", "pink", "gray", "black", "darkgreen", "tan", "deepskyblue", "gold"]

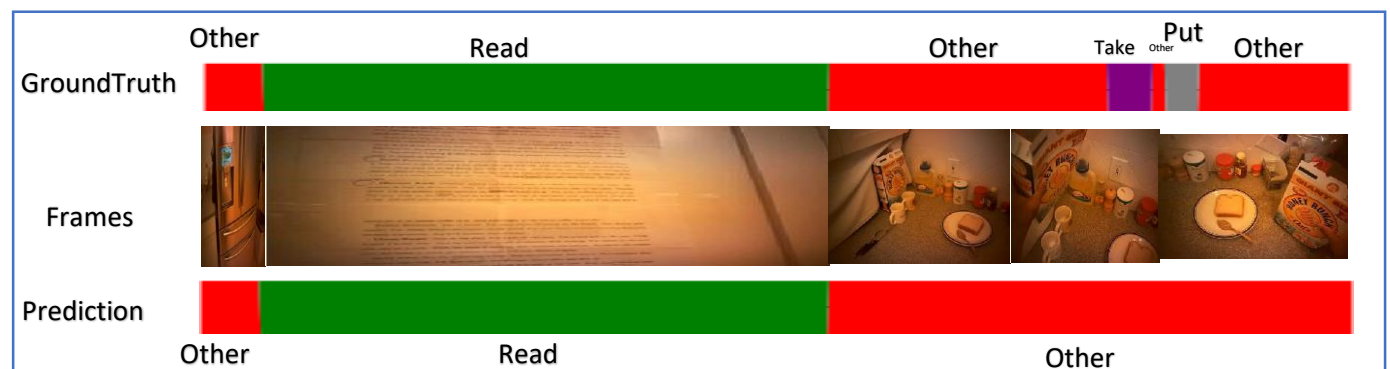
用 plt.stackplot 畫(由於解析度問題，太小的 time step 會有 alias 現象)



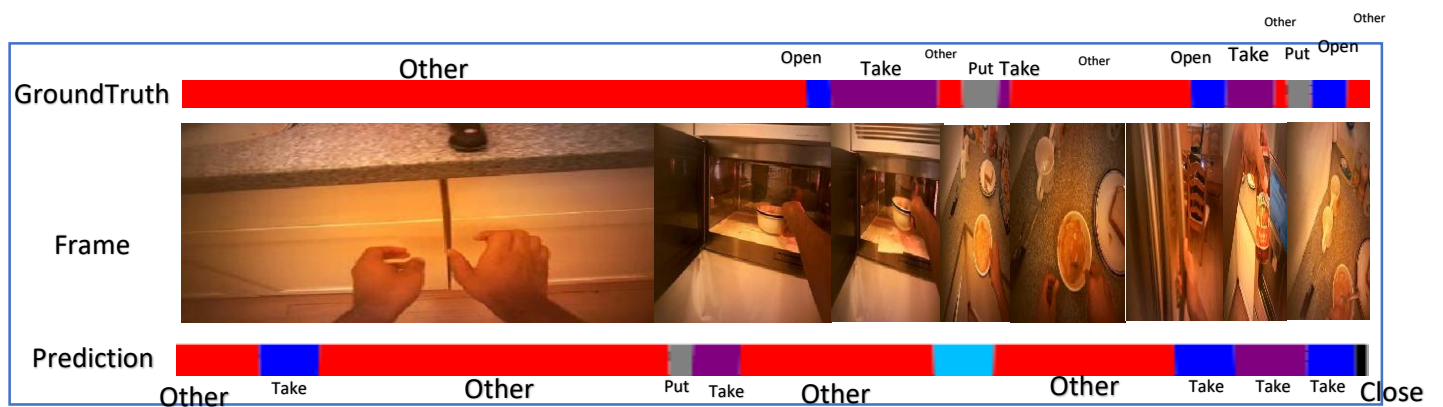
Result of OP04-R04-ContinentalBreakfast.txt (Timestep: 200 ~ 299)



Result of OP04-R04-ContinentalBreakfast.txt (Timestep: 300 ~ 399)



Result of OP04-R04-ContinentalBreakfast.txt (Timestep: 400 ~ 499)



Result of OP04-R04-ContinentalB reakfast.txt (Timestep: 500 ~ 599)



Result of OP04-R04-ContinentalB reakfast.txt (Timestep: 600 ~ 699)