**Task 1**

**Judgement**

Regarding AlexNet, the validation accuracy saturates at 66% after training 100 epochs. Moving onto VGG, this model achieves the accuracy of 68.8%. It is better compared to AlexNet because VGG is deeper; therefore, it can learn more complex features of the dataset. Moving onto MobileNetV2, this model creates a trade-off between accuracy and training time. Hence, its accuracy, which are 60.9%, cannot as competitive as the number of other models. Turning to EfficientNetV2, this architecture gains an accuracy of 0.705 on the validation set. This situaition can be explained as this architecture aims to increase the network performance by enhacing the depth, witdth, and resolution. Additonally, Darknet-53 achieves the highest accuracy (78.1%). This is because this model is not as complicated as EfficientNetV2. Hence, it can decrease the affect of overfitting. Moreover, Darknet-53 implements the residual blocks; hence, it can increase its depth without any worry about gradient vanishing/explosing. Compareed to AlexNet and VGG, its higher depth allows the network to learn more complex feature and improve the accuracy. Finally, CSPDarknet-53, a lightweight version of Darknet-53, has the accuracy of 0.661. This situation happens because the CSP layer creates a trade-off between accuracy and training time. Because Darknet-53 has the highest accuracy, we choose this model as our final model for task 1.

More details about the training process of Darknet-53, it achieves an accuracy of 0.213 after training 25 epochs. However, we improve this situation by applying a harder augmentation (we call this as “hard augmentation”), and its performance grows remarkably to 66% after the next 25 epochs. Continuing training on this “hard augmentation”, our model achieves an accuracy of nearly 0.72. We recognize that the performance tends to saturate; thus, we increase the strength of the augmention (we refer this as “strong augmentation”), and this approach helps us achieves an accuracy of 78.1% after training more than 100 epochs. The reason why the model performance increases together with the diffculty of augmentation has been explained by Tan and Le [1].

**Evaluation**

Flowers102 is a popular dataset of flowers used to evaluate image classification neural networks. Different versions of one of the models that we are going to use, EfficientNet, have the accuracy of 98.8%, standing top 17 [2]. Different versions of ResNet also has their standings in the leaderboards. Vision transformers has been discussed that it can outperform structures without using pre-training or strong augumentation and therefore, many of the top models use vision transformers [3].

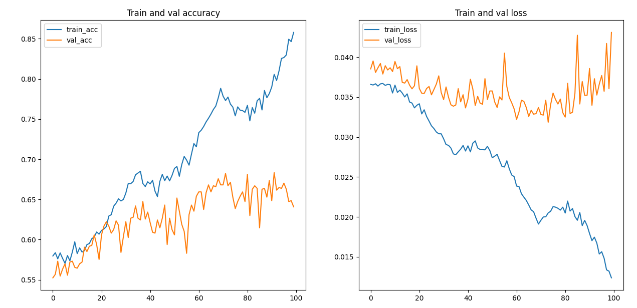


Figure 1: AlexNet next 100 epochs

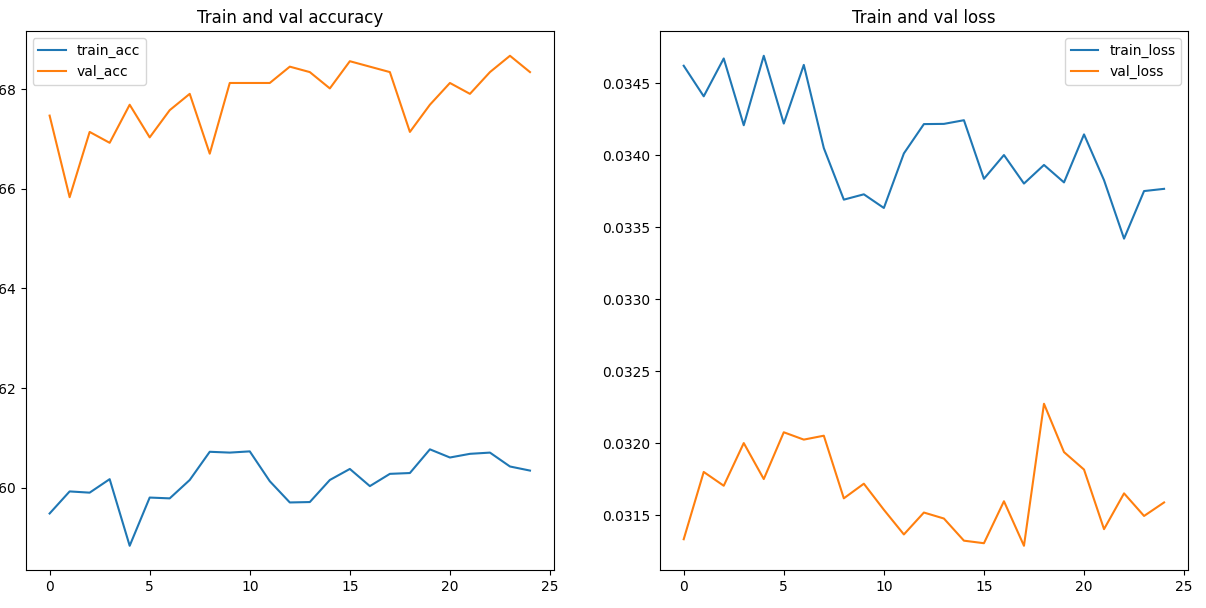


Figure 2: VGG11

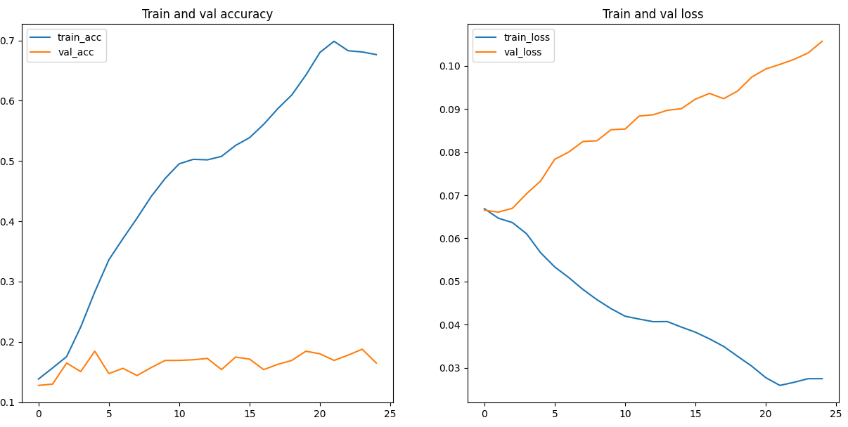


Figure 3: MobileNetV2

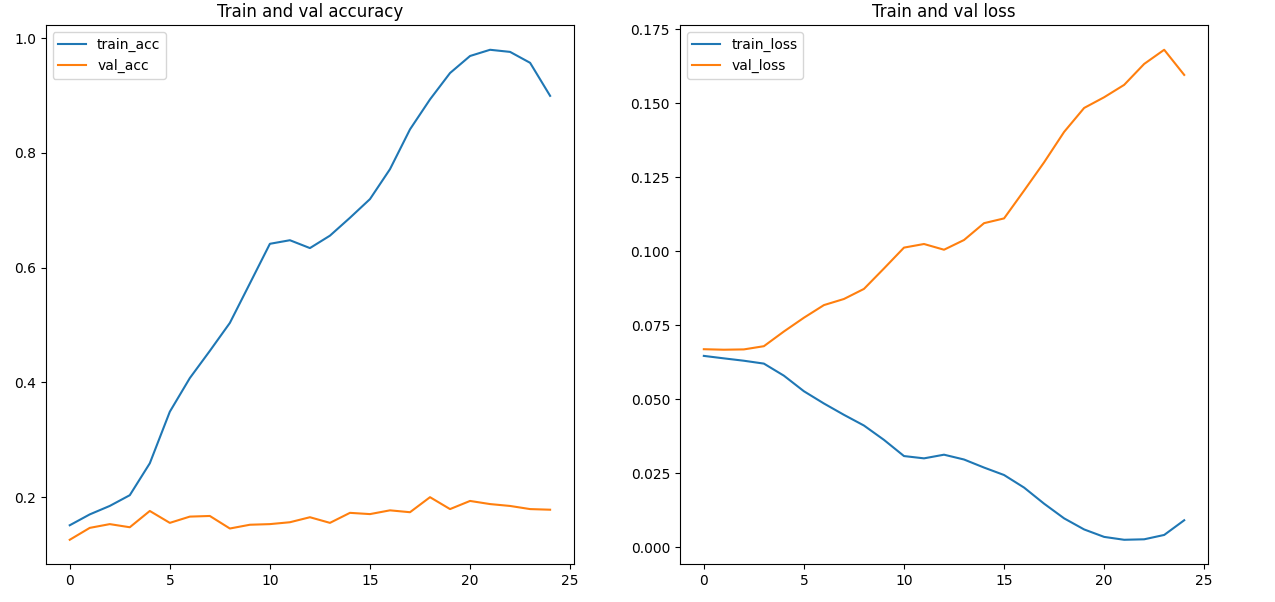


Figure 4: EfficentNetV2

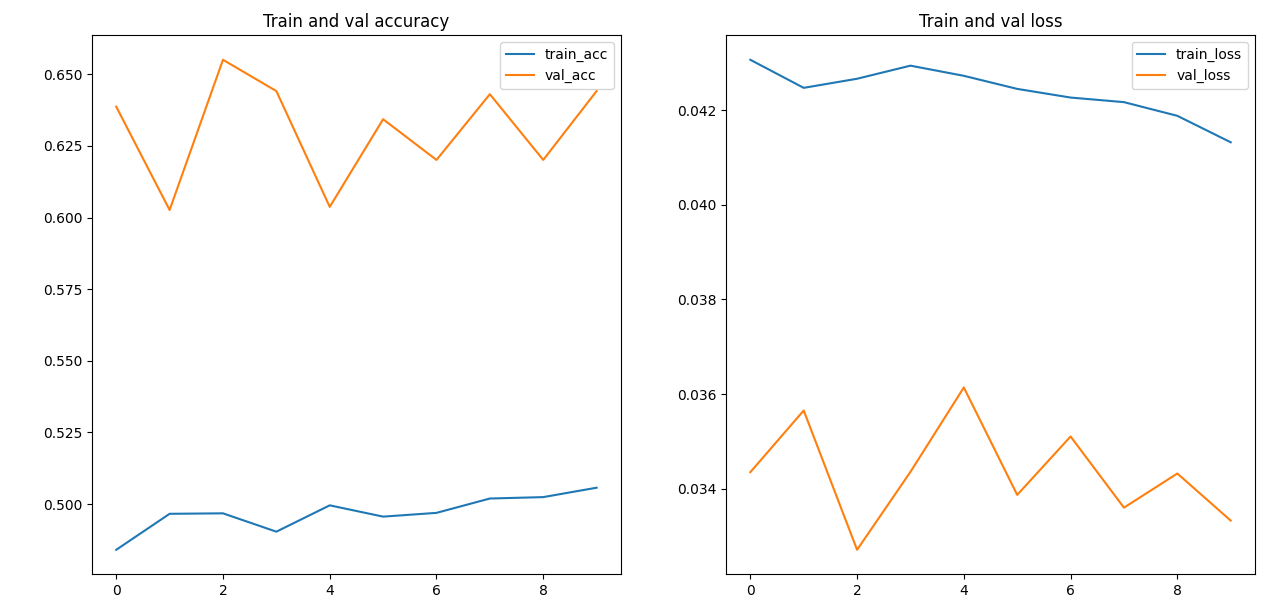


Figure 5: CSPDarkNet

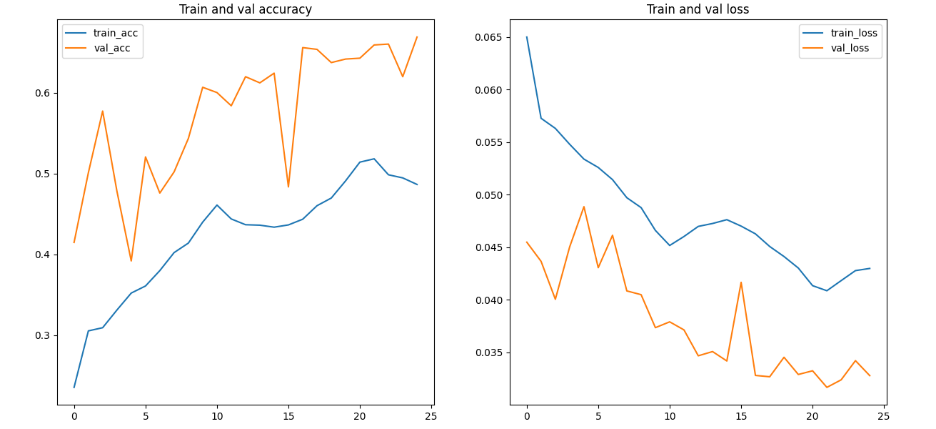


Figure 6: DarkNet after augumentation

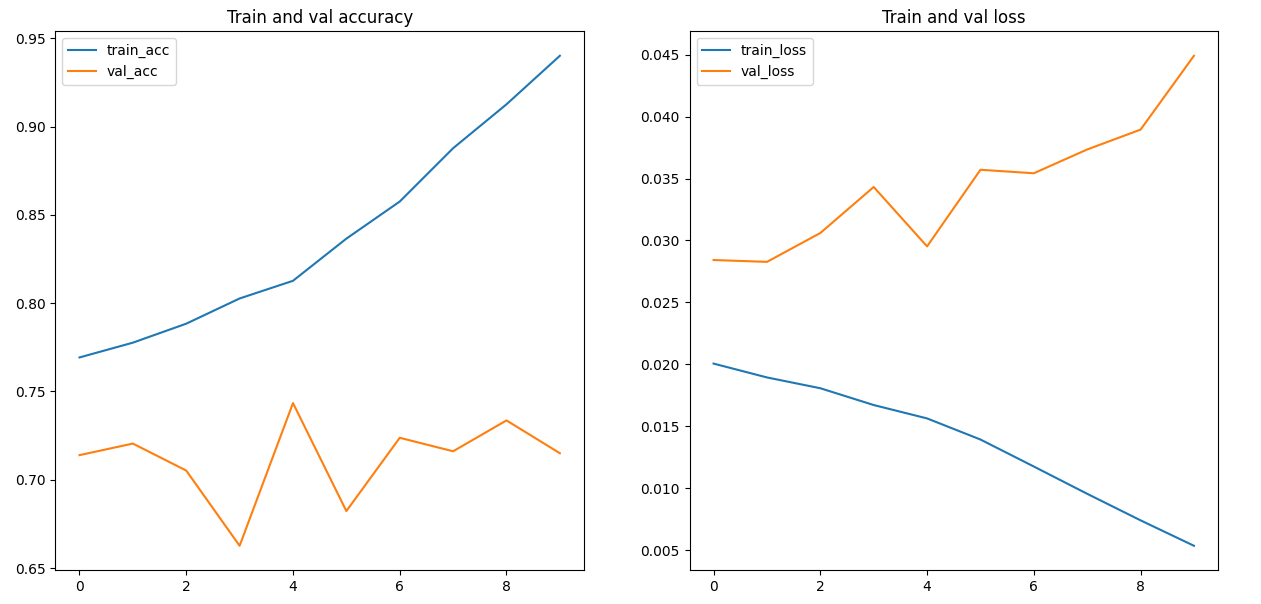


Figure 7: DarkNet with AdamW optimizer

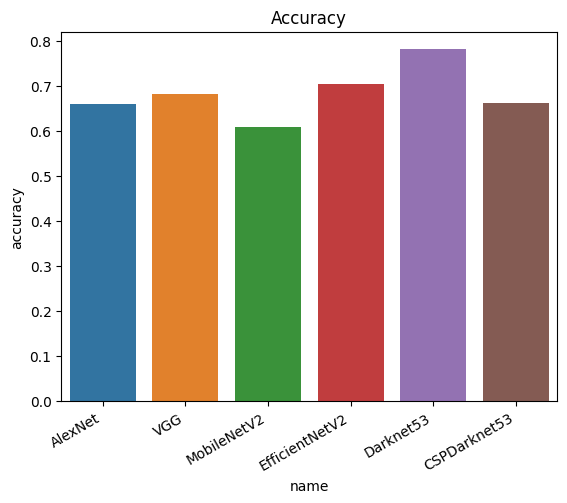


Figure 8: Models summary

References

[1] [2] [3] X. Chen, C. Hsieh, B. Gong. “Stabilizing molecule could pave way for lithium-air fuel cell.” paperswithcode.com. <https://paperswithcode.com/sota/image-classification-on-flowers-102> (accessed 18 May 2023)