Hochschule
München
University of
Applied Sciences

Fakultät 7

Anna Reiter, Chiara Perocco

Advanced Deep Learning

Team Project

OVERVIEW

Domain

Domain

Domain: informative text about American Sign Language (ASL)

Article Structure:

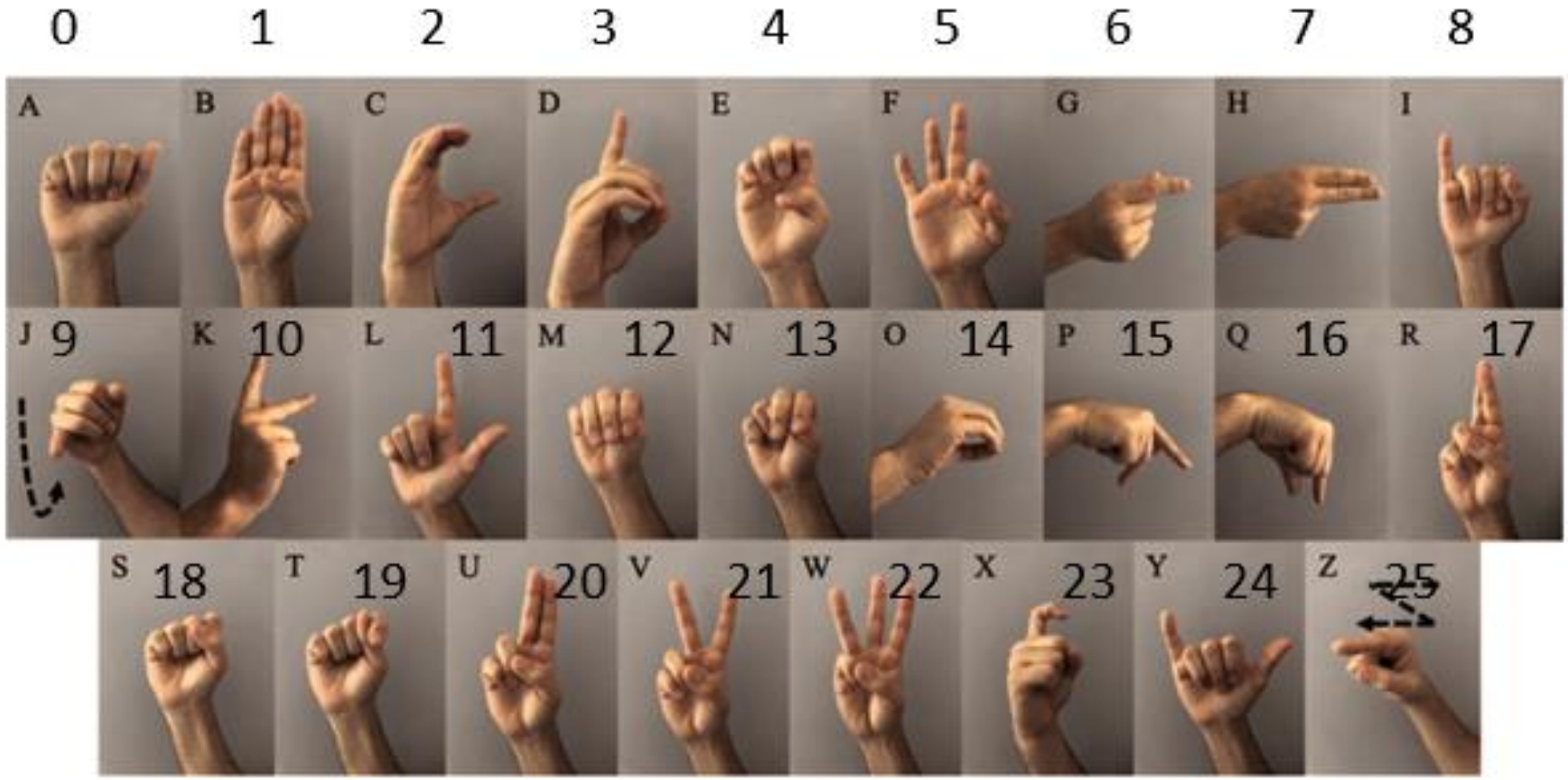
1. Introduction
2. The letter in written language
3. The letter in sign language
4. Conclusion

} Each paragraph is followed by an image

Dataset: ASL(American Sign Language) Alphabet Dataset → consists of 29 classes (only 26 used)

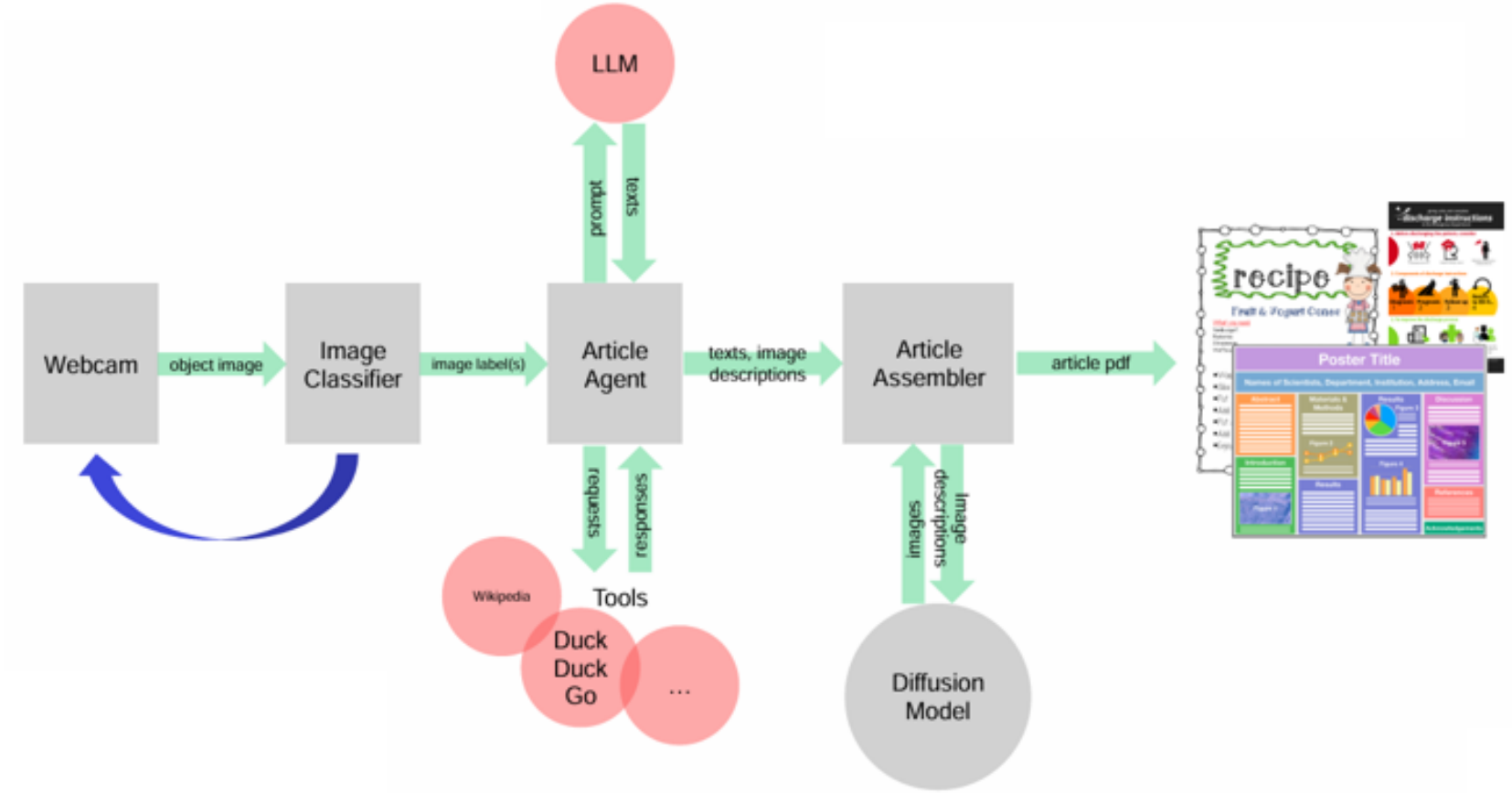
Source: Kaggle (ASL(American Sign Language) Alphabet Dataset)

Domain


0 1 2 3 4 5 6 7 8
A B C D E F G H I
J 9 K 10 L 11 M 12 N 13 O 14 P 15 Q 16 R 17
S 18 T 19 U 20 V 21 W 22 X 23 Y 24 Z 25

Architecture of the system

Architecture of the system



Object Detector

Object Detector

- Laptop camera
- Opencv-python
- Takes a photo by triggering the „y“ (yes) on keyboard
- Takes no photo by triggering the „n“ (no) on keyboard

Image Classifier

Image Classifier

- AlexNet from scratch
- VisionTransformer pretrained on ImageNet
- ResNet50 pretrained on ImageNet

Article Agent, Artcile Assembler, Diffusion Model

Article Agent, Artcile Assembler, Diffusion Model

- Tools: Wikipedia, DuckDuckGo
- LLM: Llama3.1 von Ollama
- Markdown and Pandoc
- HuggingFace Model: kakaobrain/karlo-v1-alpha

Results

1. Ds1
    - 1.1 AlexNet
    - 1.2 ResNet50
    - 1.3 VisionTransformer

2. Ds2
    - 2.1 AlexNet
    - 2.2 ResNet50
    - 2.3 VisionTransformer

Definition Ds1, Ds2

Ds1

300 images per class (80% train, 10% val, 10% test)
→ computing power

Data augmentation:

- horizontal flip
- crop (padding n = 4)

Hyperparameter tuning with optuna

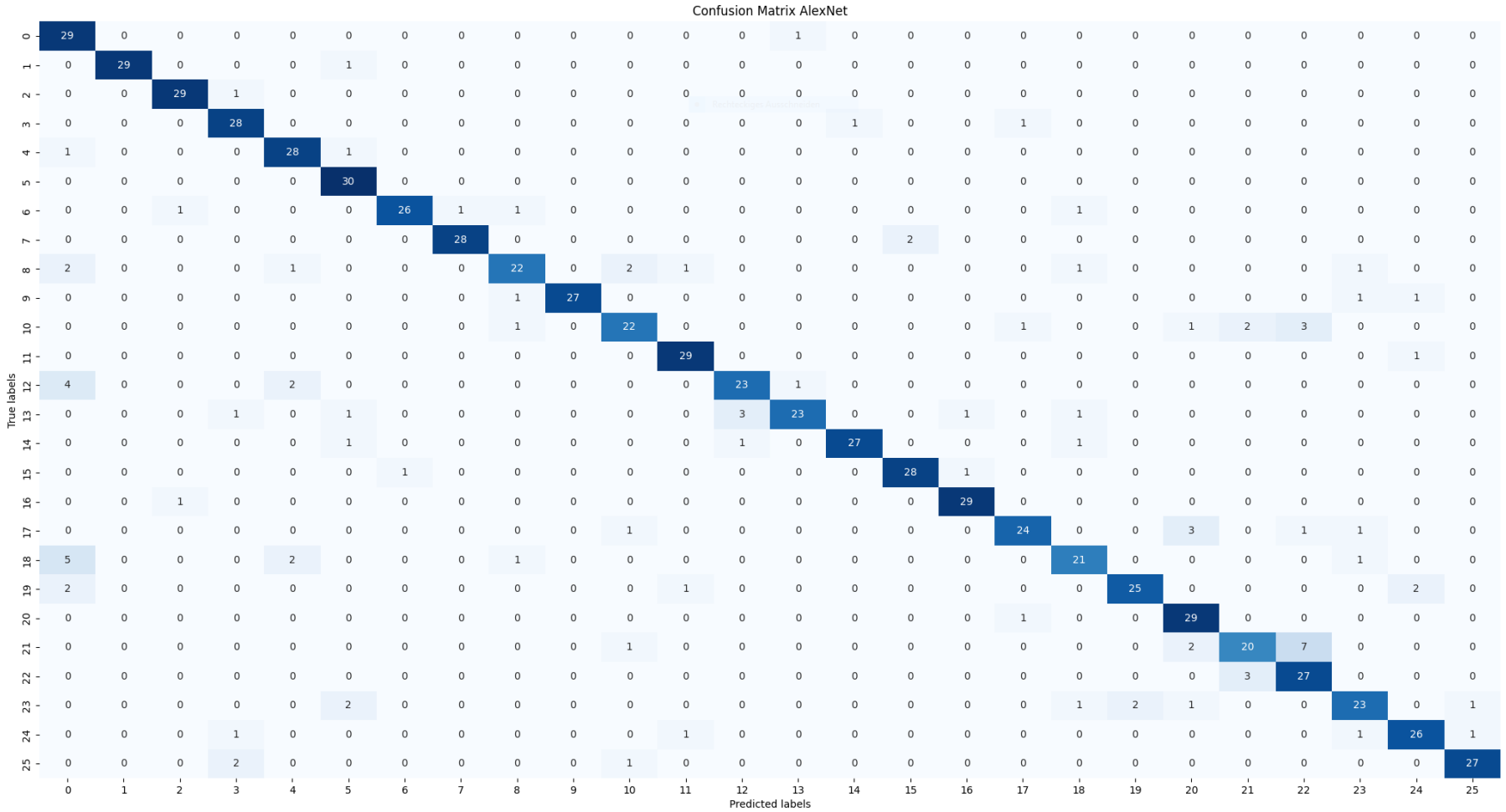
Ds2

Adjustment of ds1 with mediapipe; 1300 images per class (80% train, 10% val, 10% test)

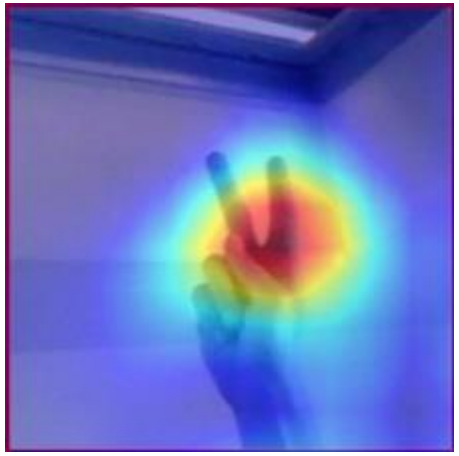
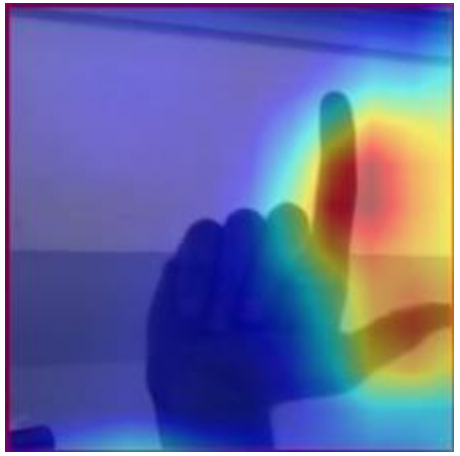
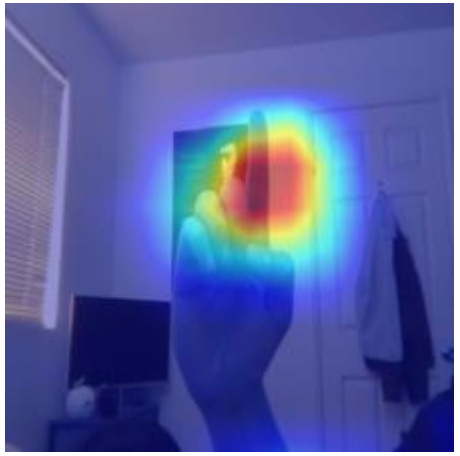
1. Data augmentation:
   - horizontal flip
   - crop (padding n = 4)
2. Data augmentation:
   - Rotation
   - Colour properties
   - Affine transformations
   - Grayscale conversion

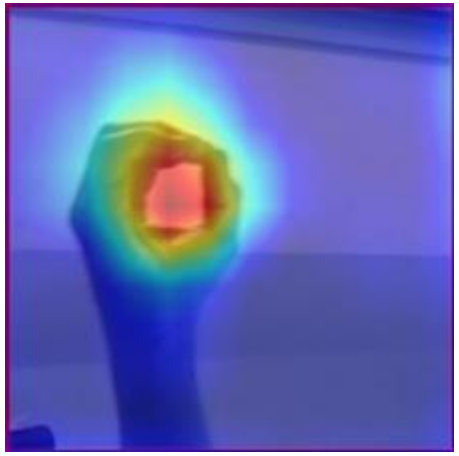
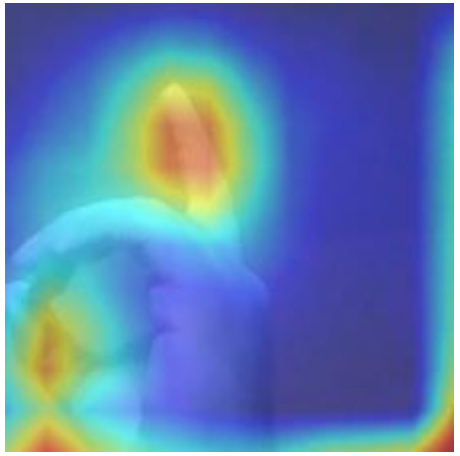
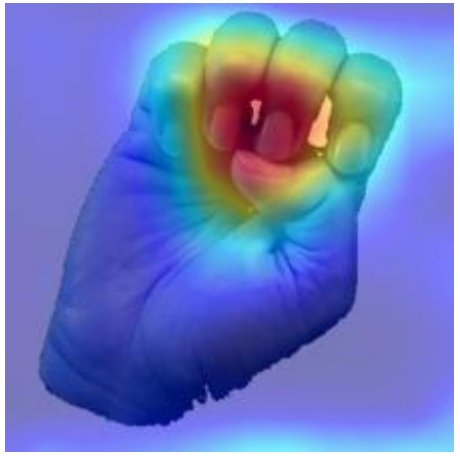
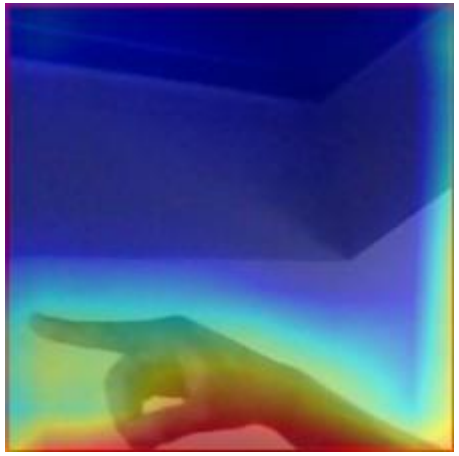
Ds1 – AlexNet

Learning rate: 3.99e-05; Batch size: 64; Epochs: 35

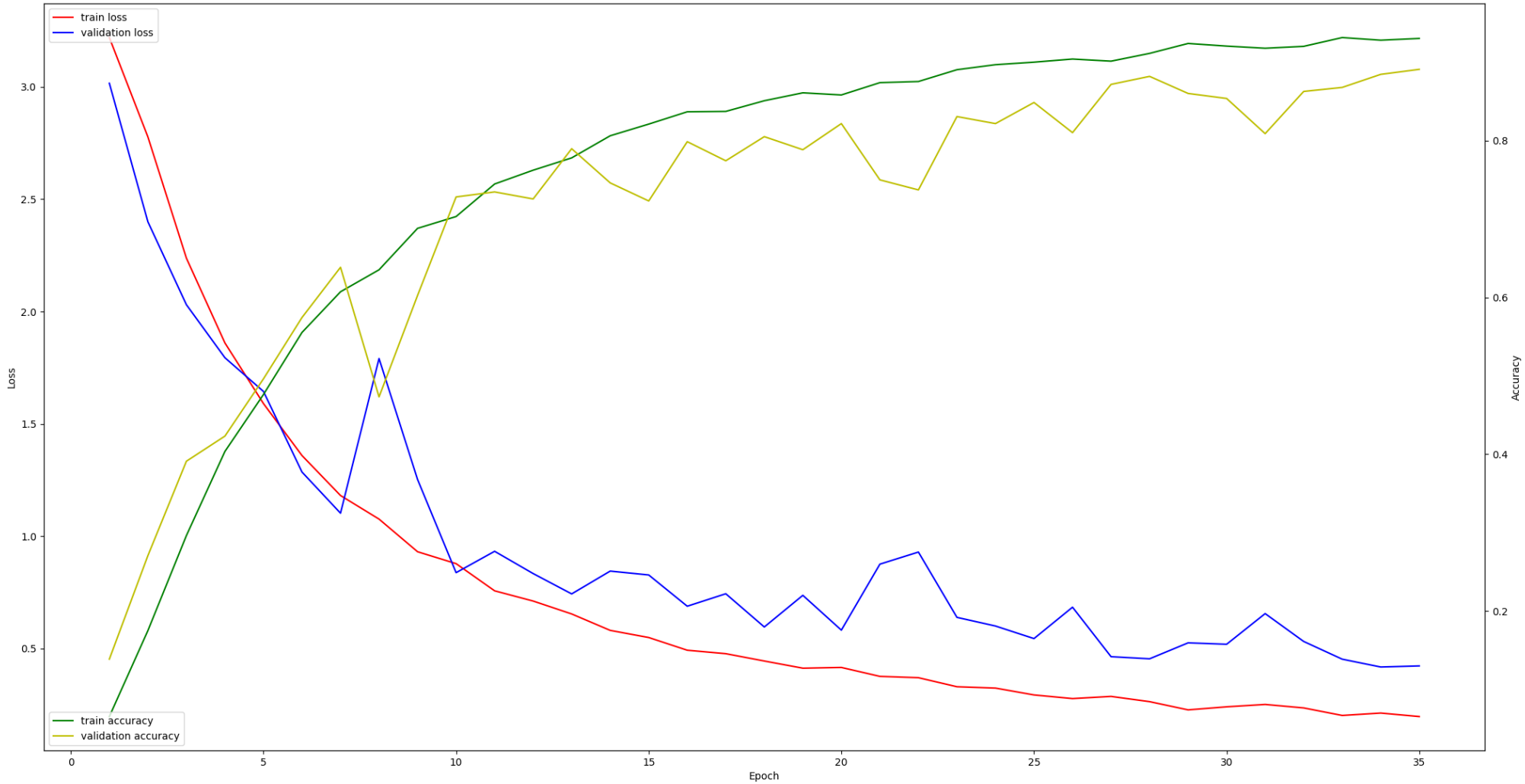


Ds1 - AlexNet

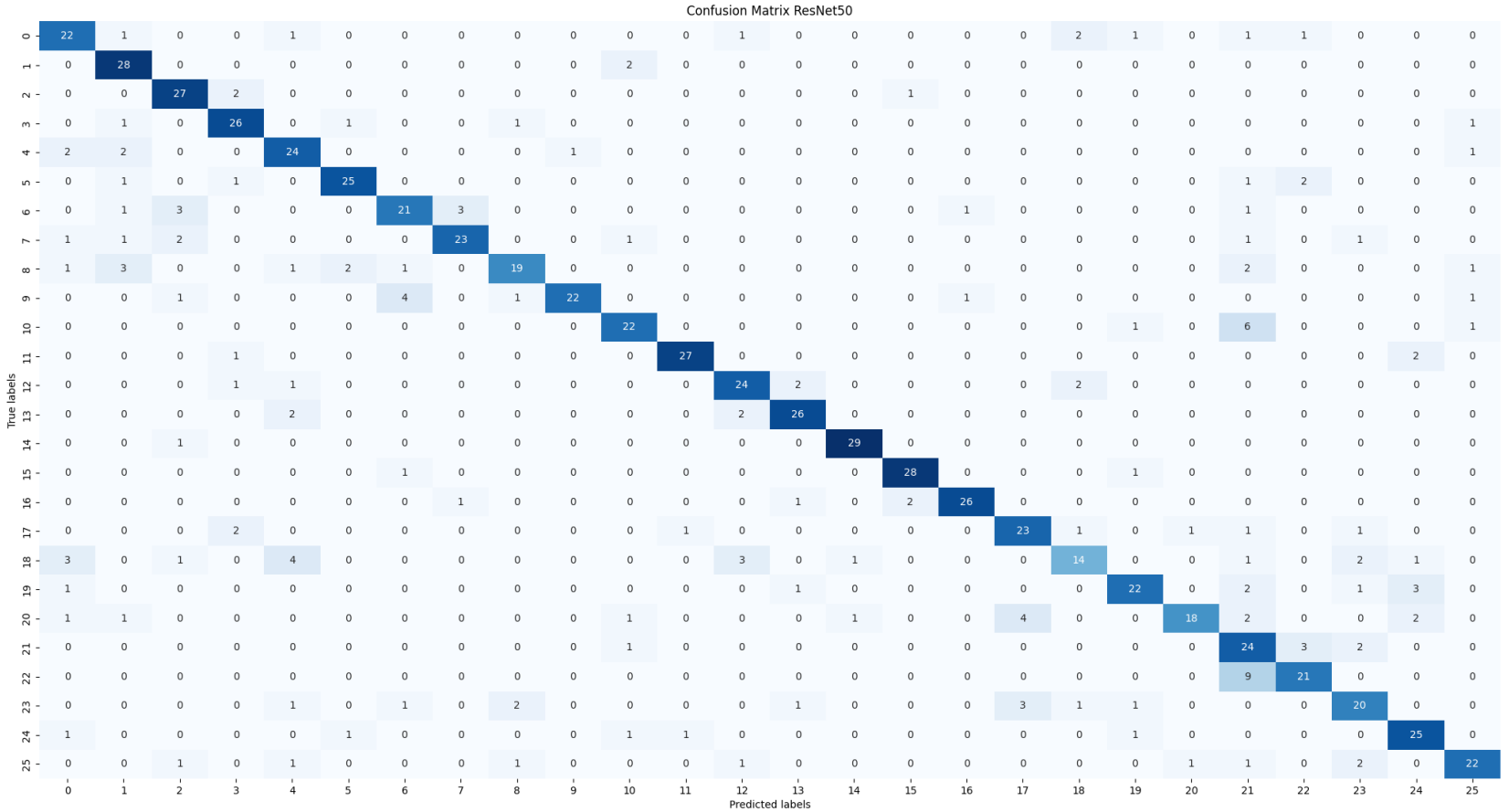





Ds1 - AlexNet

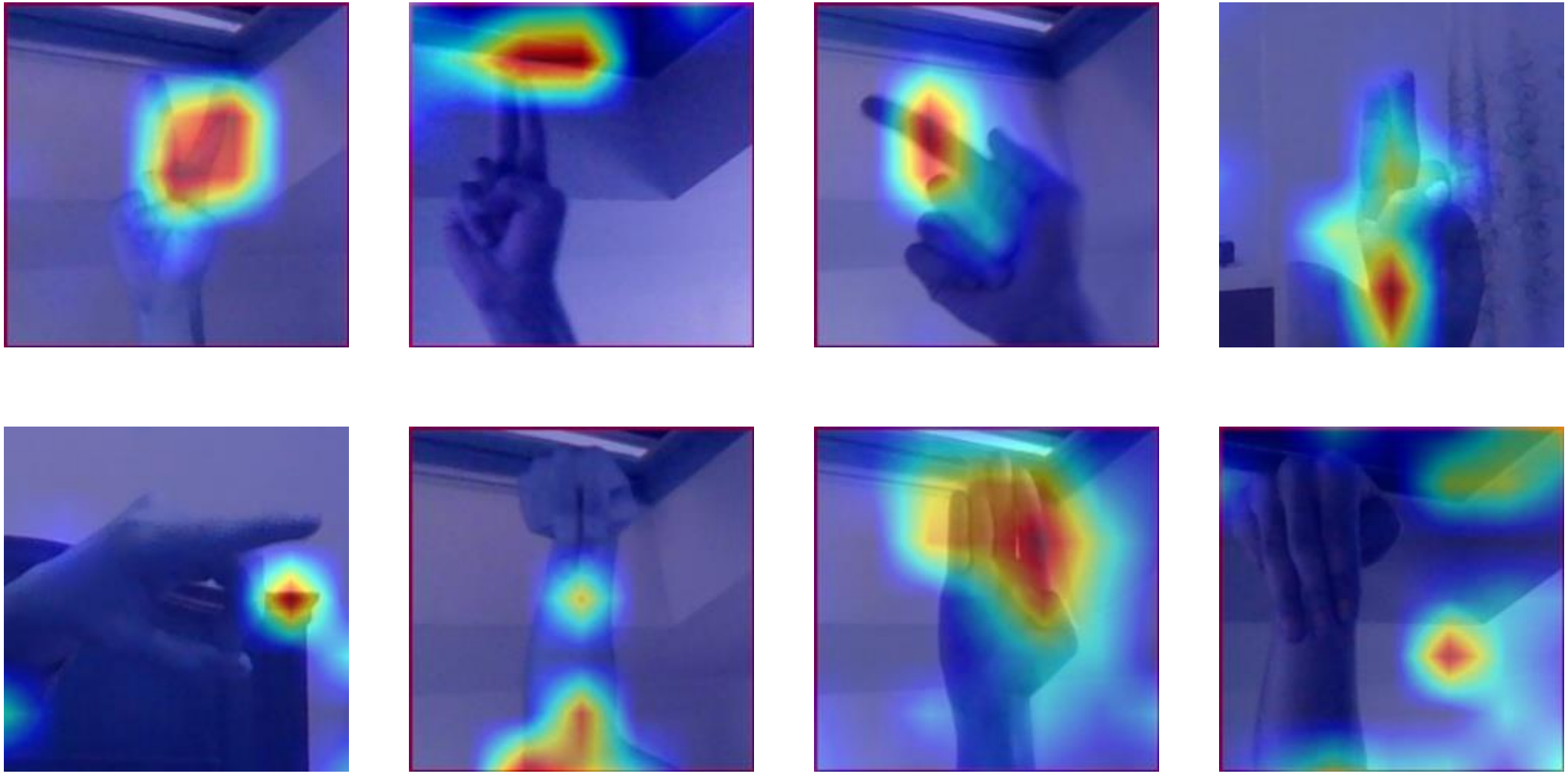

train loss
validation loss
train accuracy
validation accuracy
Loss
Accuracy
Epoch
0
5
10
15
20
25
30
35
0.5
1.0
1.5
2.0
2.5
3.0
0.2
0.4
0.6
0.8

Ds1 – ResNet50

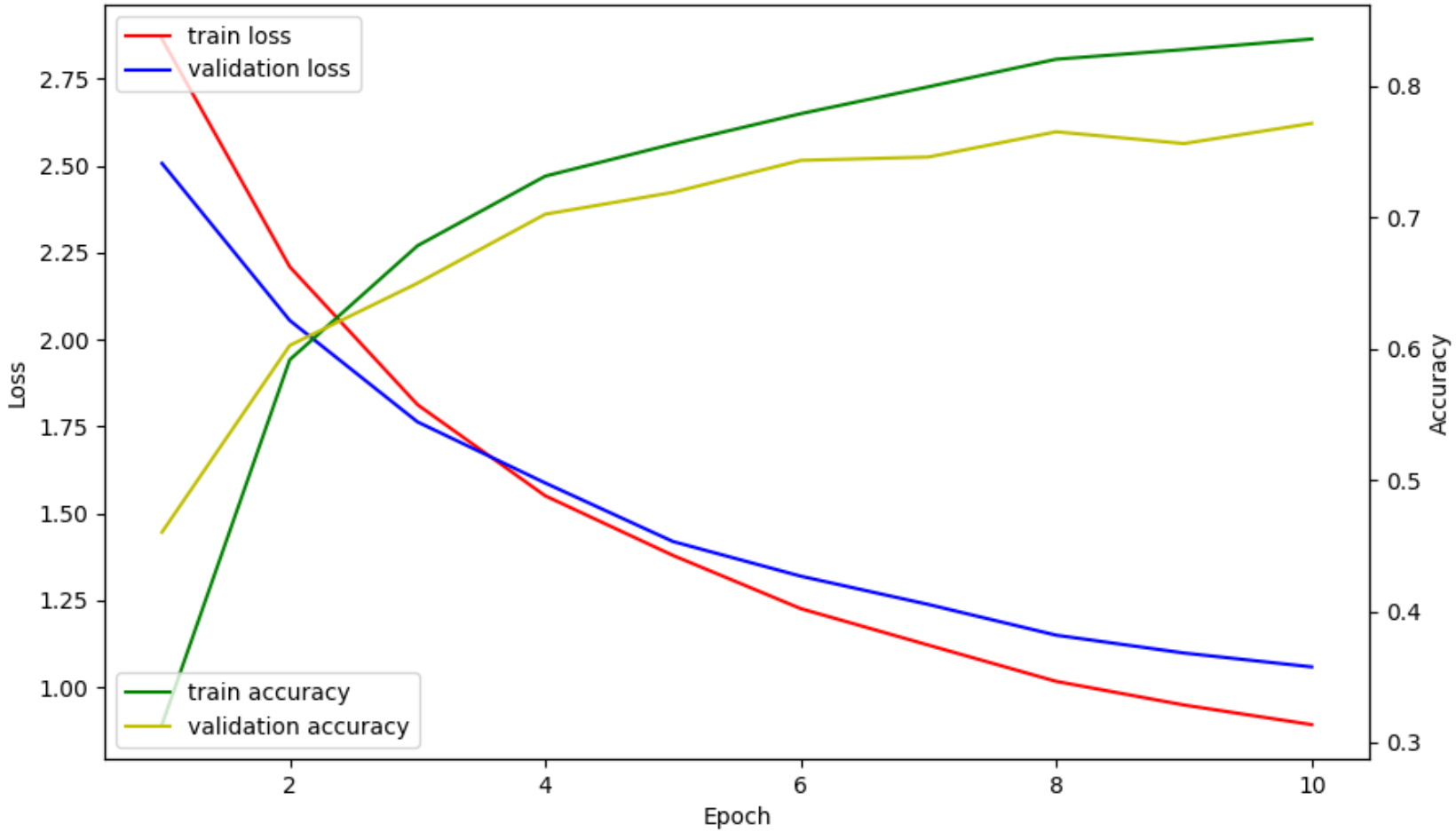
Learning rate: 0.0008; Batch size: 64; Epochs: 10



Ds1 – ResNet50

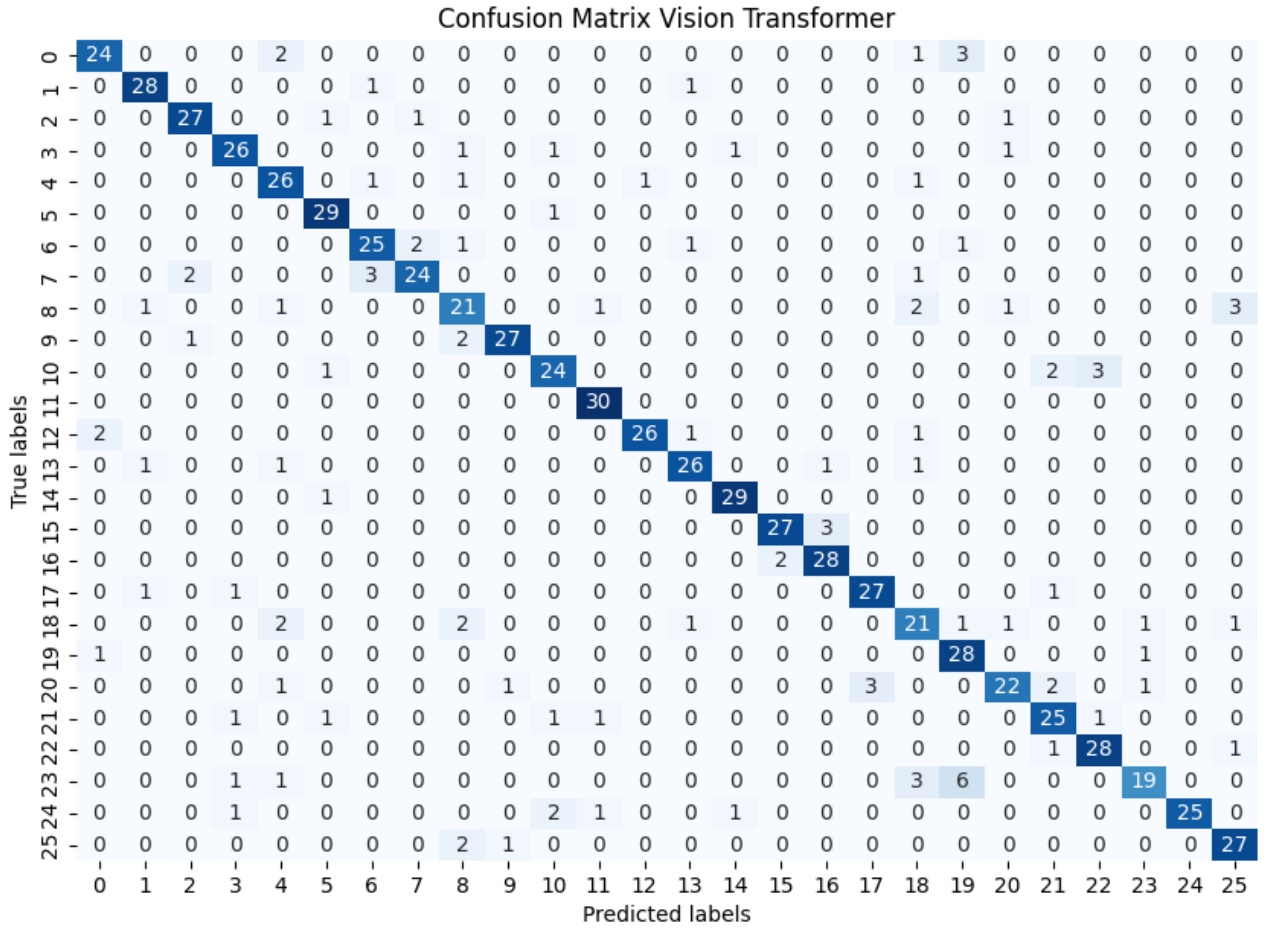


Ds1 – ResNet50

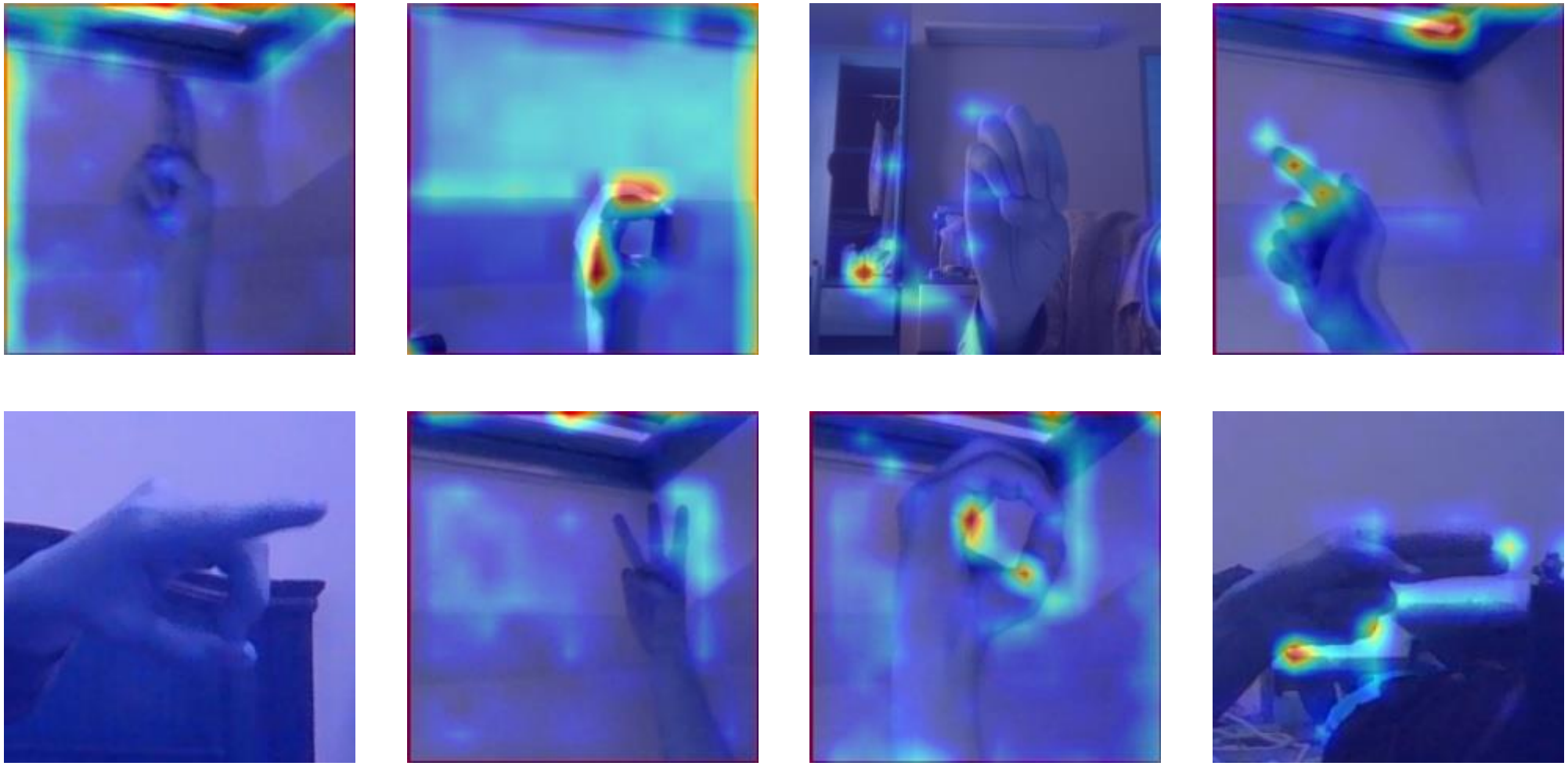


Ds1 – VisionTransformer

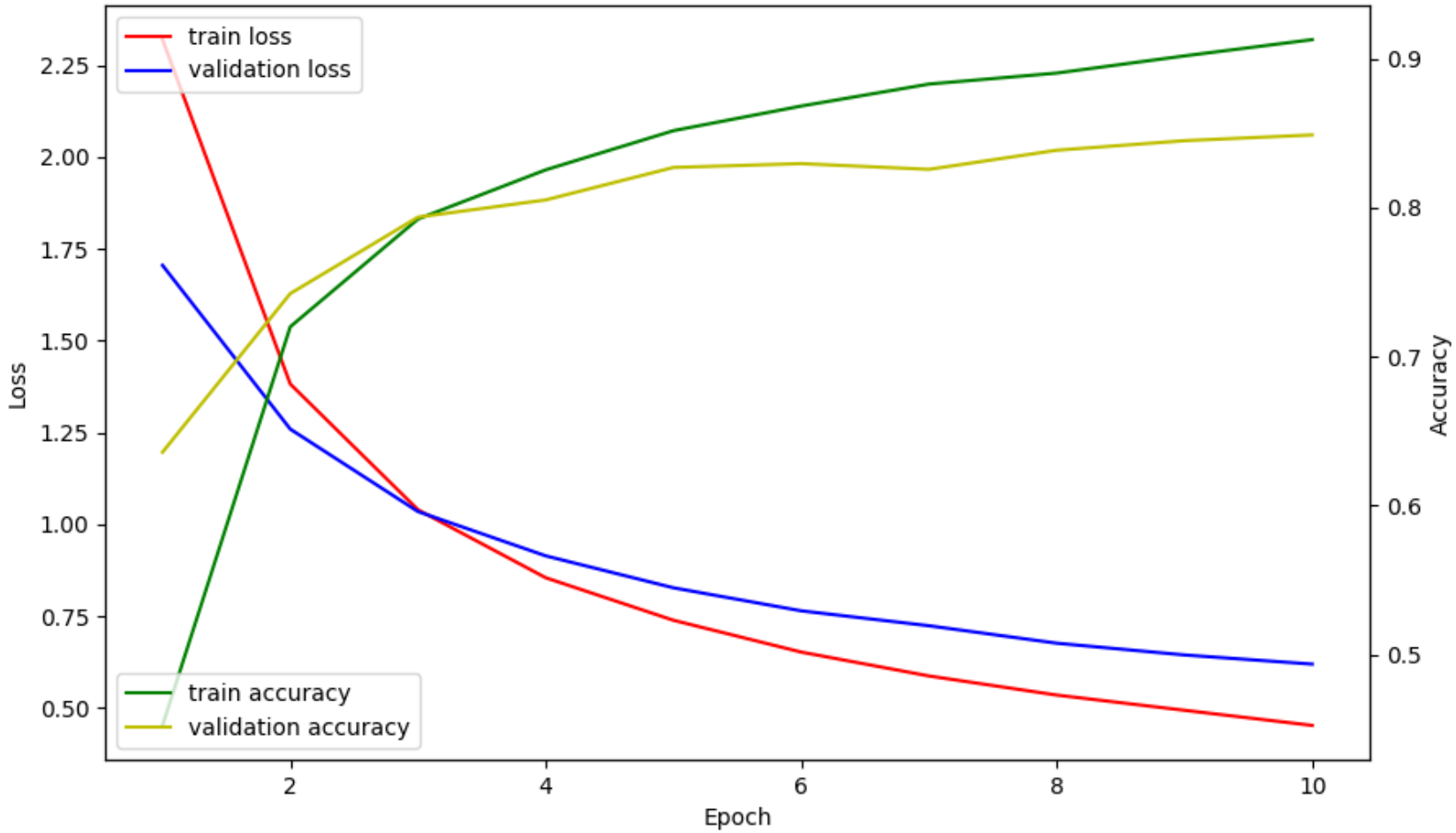
Learning rate: 0.0008; Batch size: 64; Epochs: 10



Ds1 – VisionTransformer

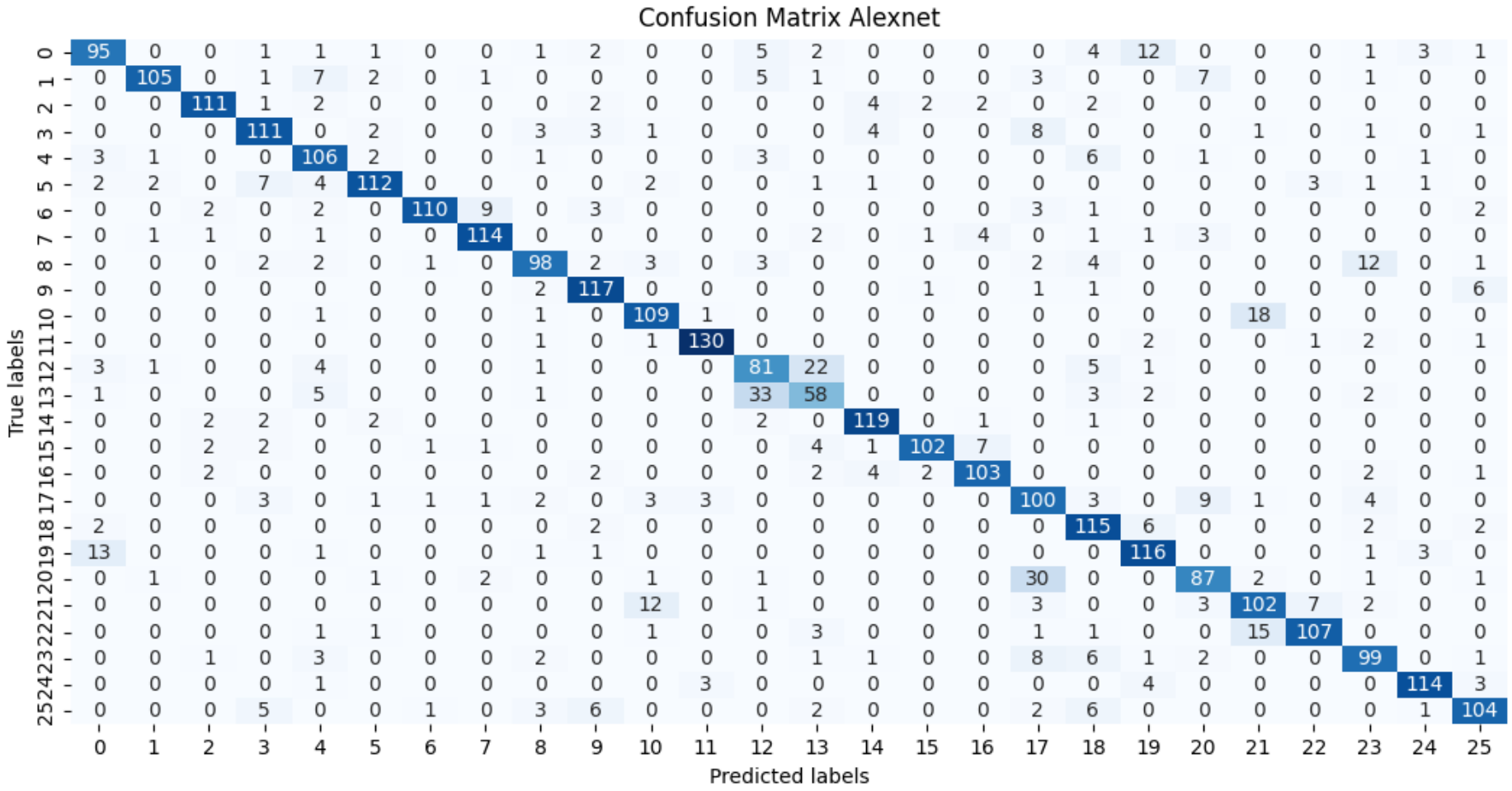


Ds1 – VisionTransformer

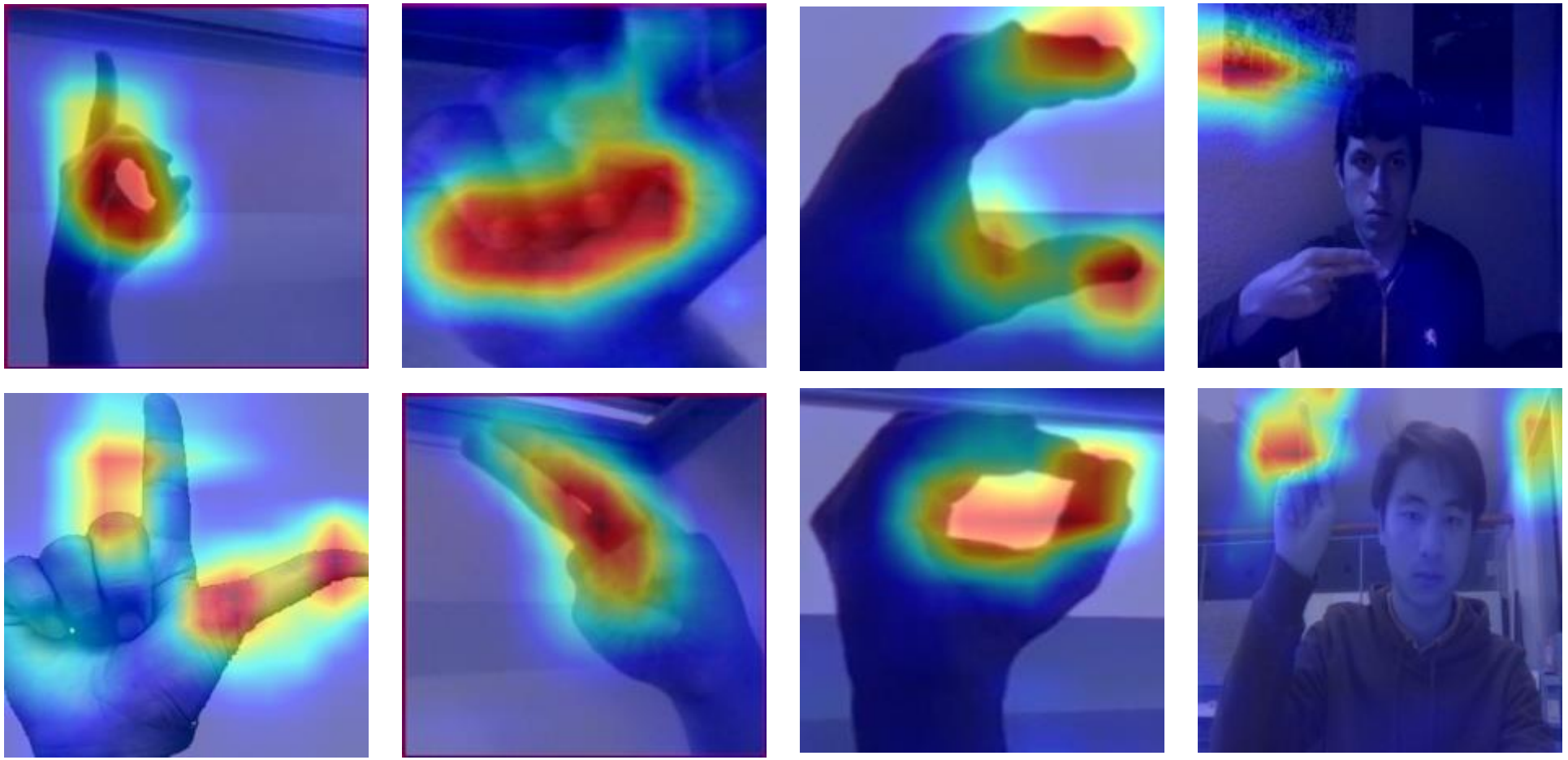


Ds2 – AlexNet

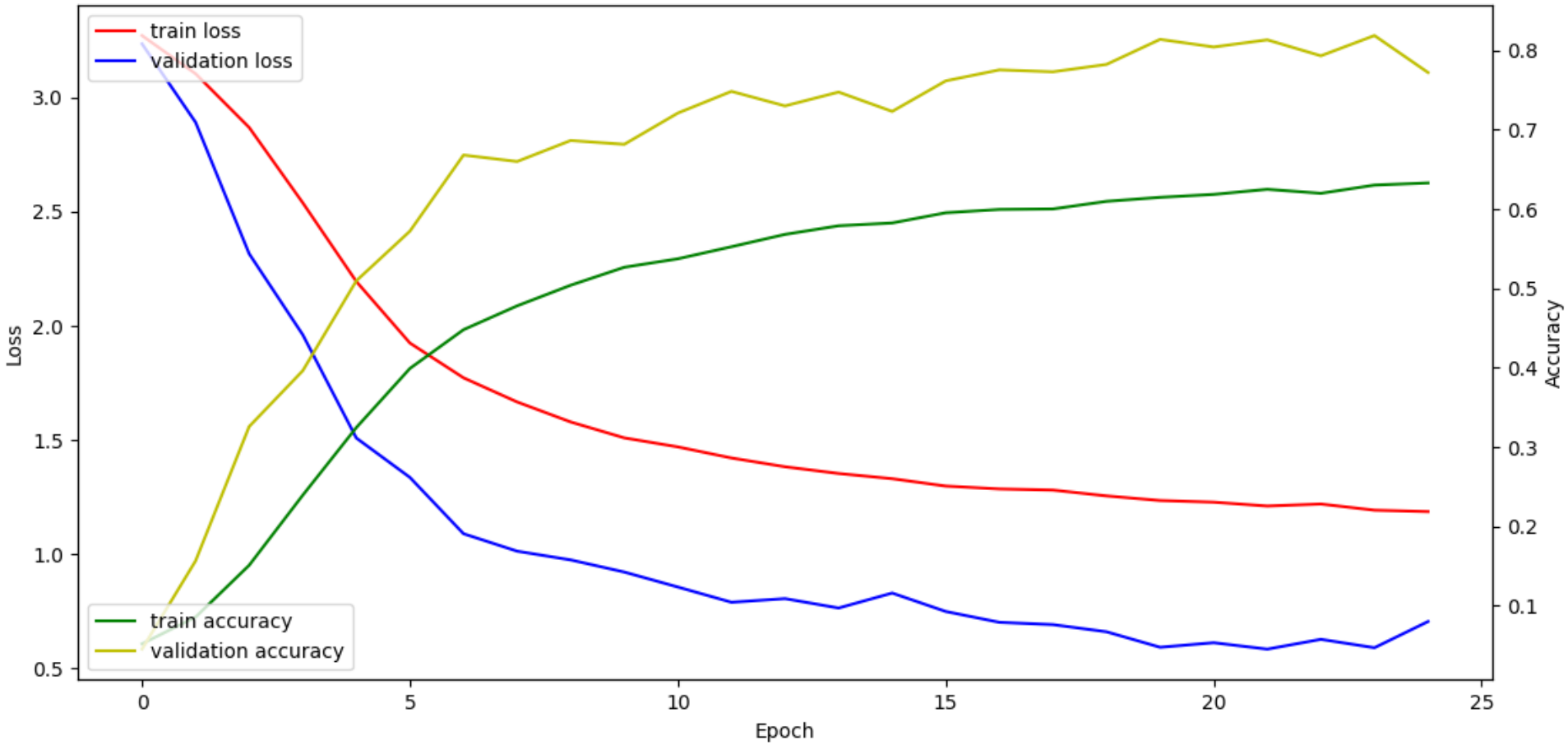
Batch size: 64, Drop out: 0.5, Learning rate: 0.0008, Epochs: 50 (early stopping)



Ds2 – AlexNet

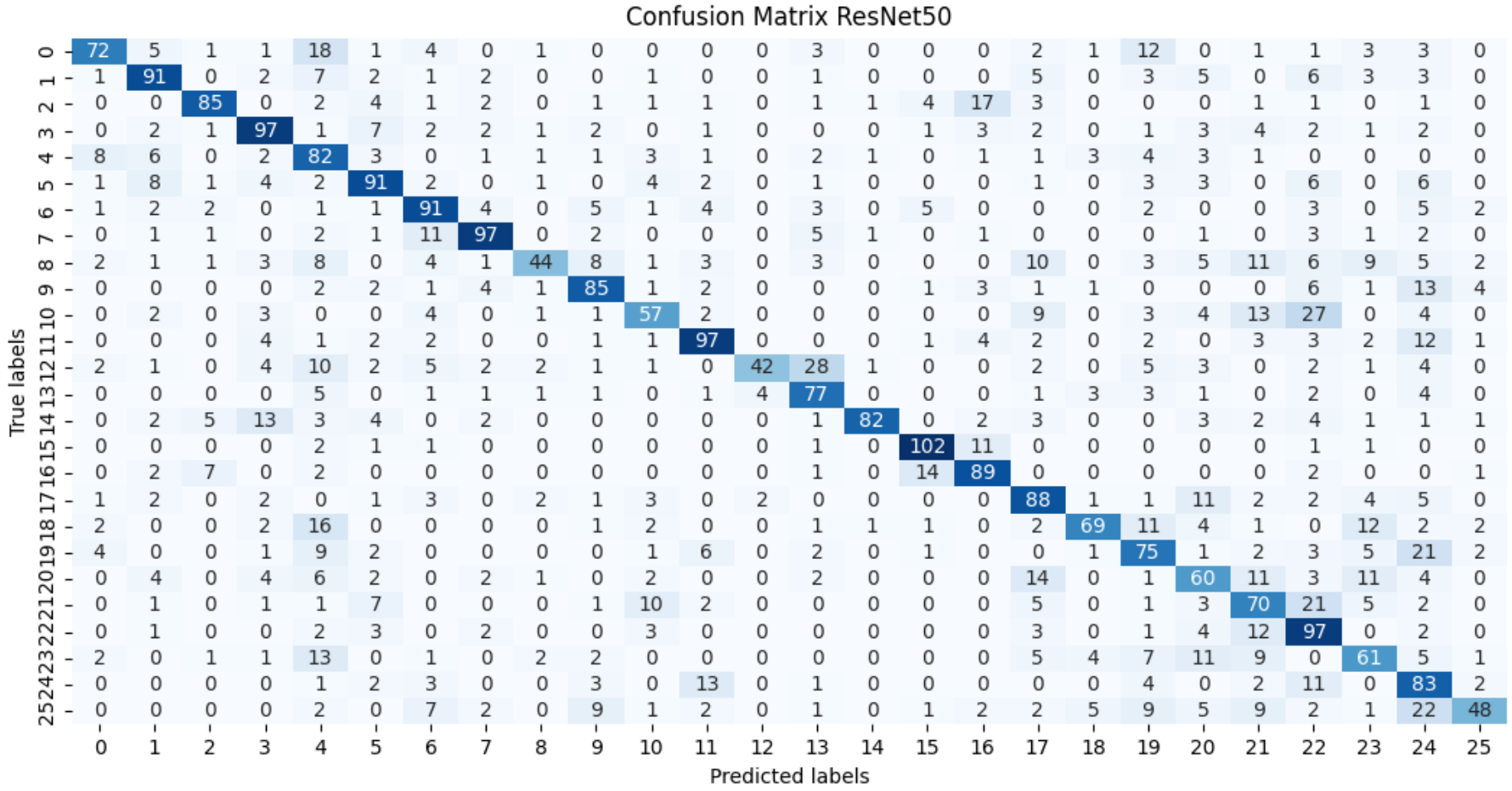


Ds2 – AlexNet

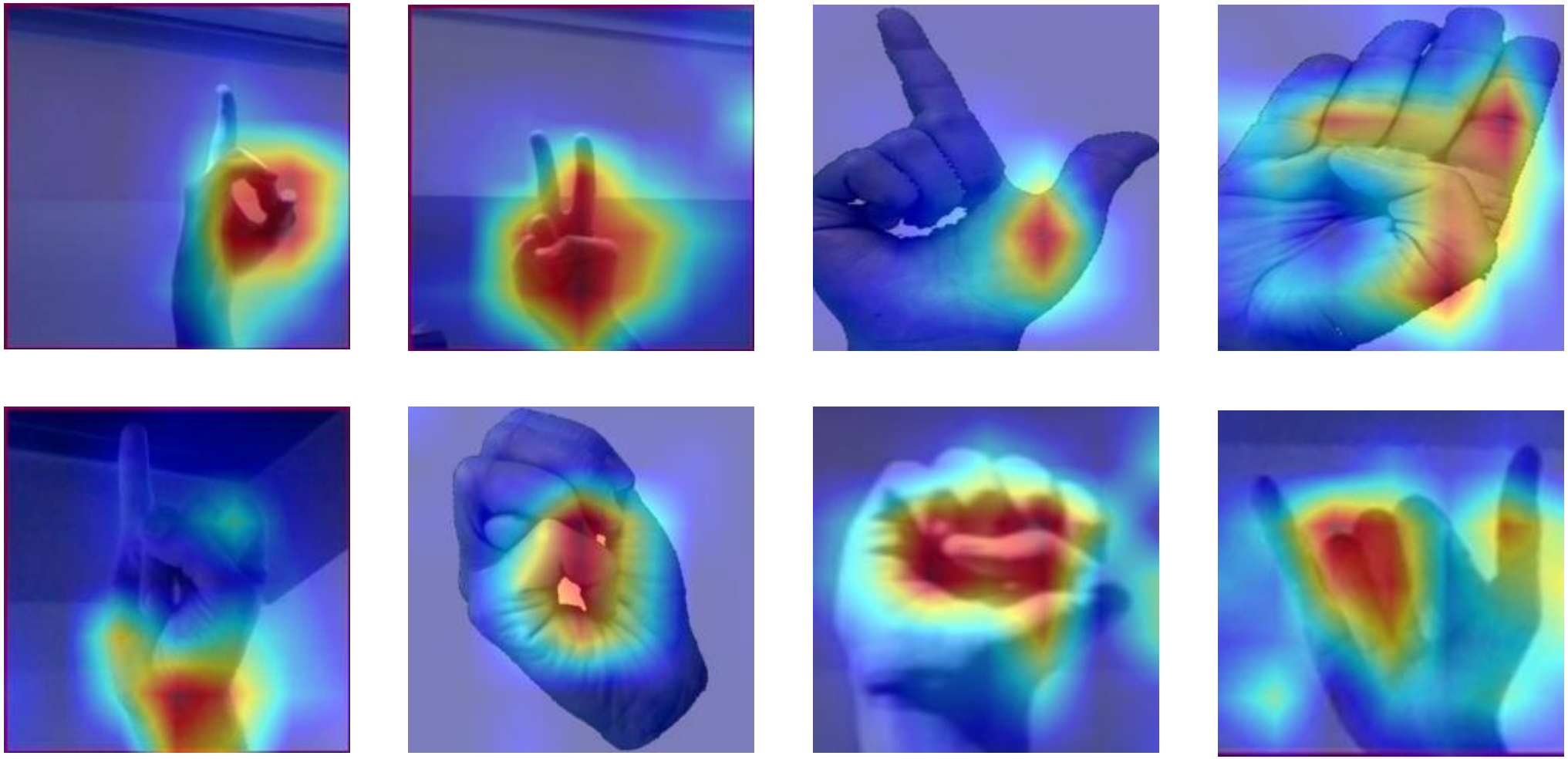

train loss
validation loss
train accuracy
validation accuracy
Loss
Accuracy
Epoch
3.0
2.5
2.0
1.5
1.0
0.5
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
15
20
25

Ds2 – ResNet50

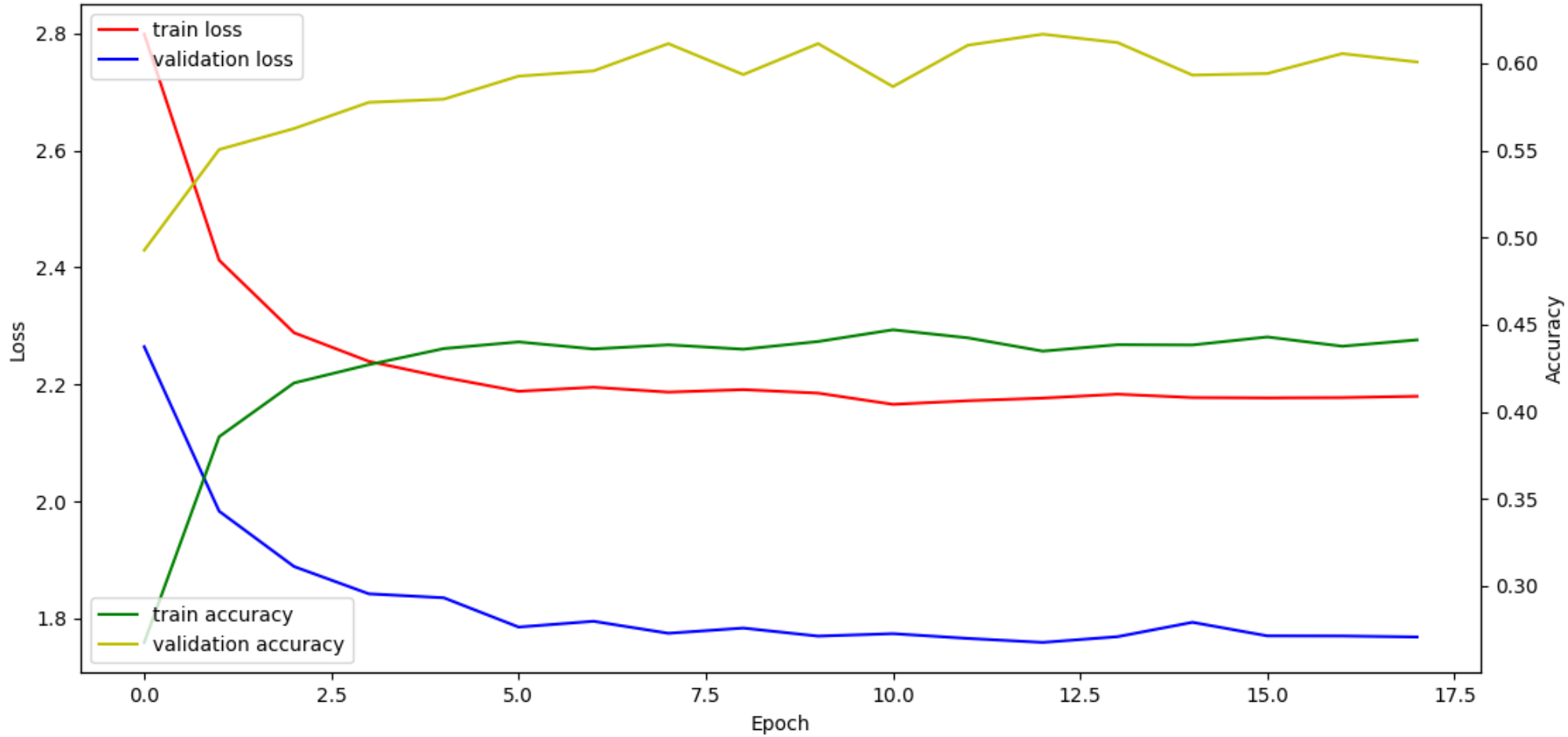
Batch size: 64, Drop out: 0.2, Learning rate: 0.001, Epochs: 50 (early stopping)



Ds2 – ResNet50

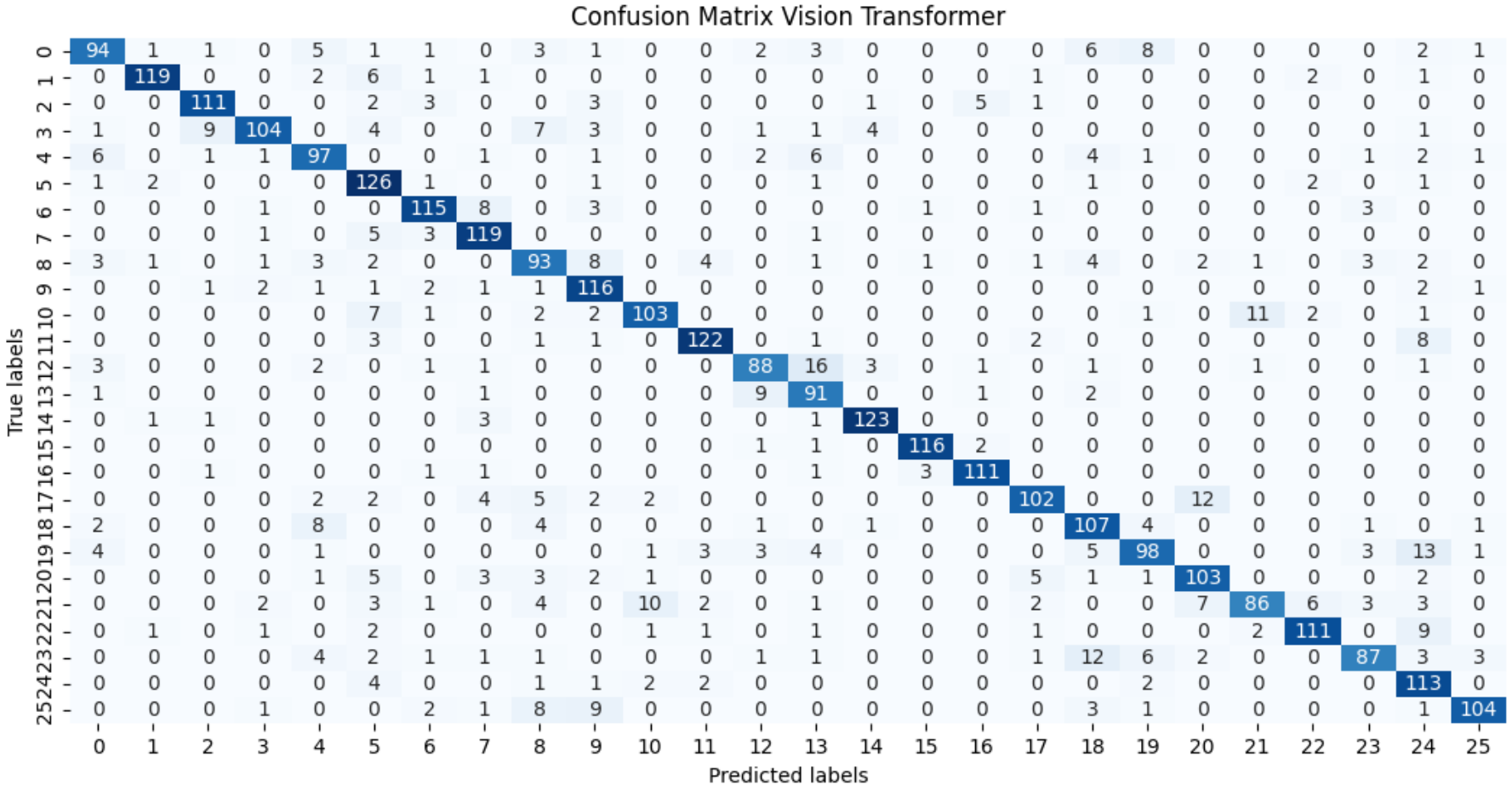


Ds2 – ResNet50

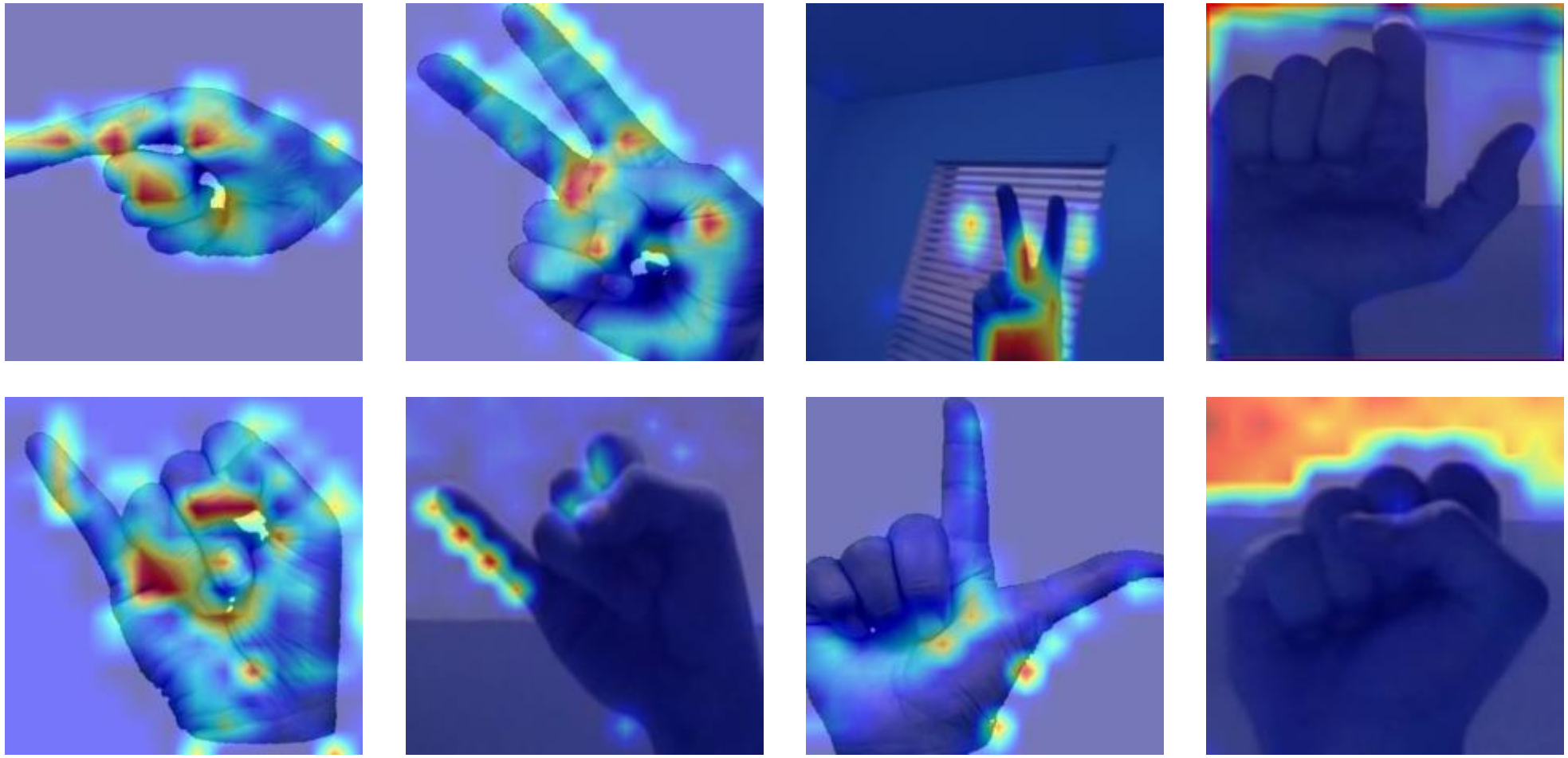


Ds2 – VisionTransformer

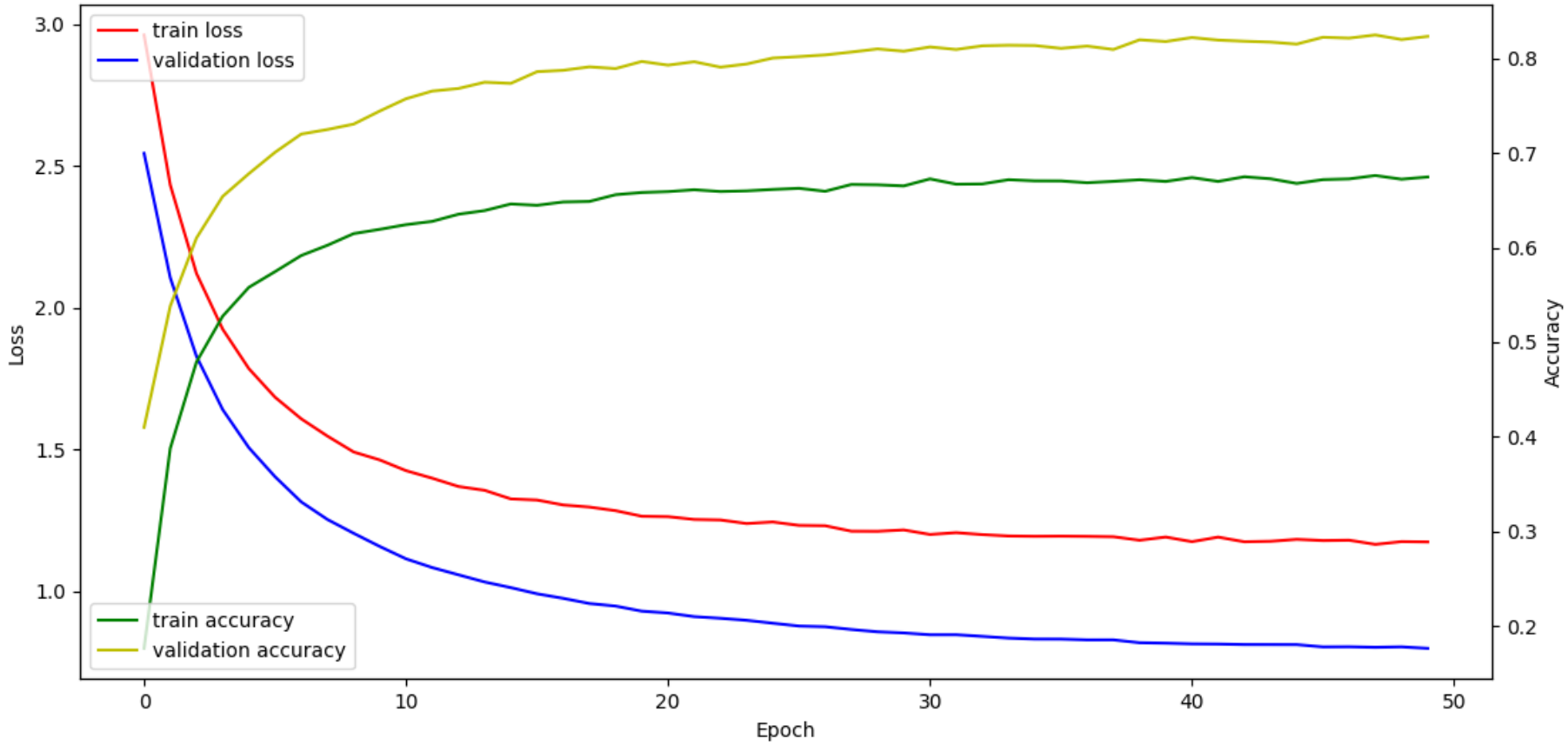
Batch size: 64, Drop out: 0.3, Learning rate: 0.0001, Epochs: 50



Ds2 – VisionTransformer



Ds2 – VisionTransformer



Conclusion

Conclusion

- Best Models: Vision Transformer, AlexNet
- Data Augmentation: More images and augmentation improve model performance
- Early Stopping: Adjusting patience prevents overfitting and enhances generalization
- Optimization: Dropout, weight decay, (learning rate scheduling) improve robustness

Outlook

Outlook

- Fine-tuning Diffusion model
- Optimize Prompt Engineering
- Improve data quality
- More experiments with hyperparameters