Trusted Artificial Intelligence

Homework 1

Spring 2024

Your Name

ID: StudentID | Department of Electrical and Computer Engineering, University of Tehran

Instructor: Dr. Mostafa Tavasolipour

Submitted: February 12, 2026

Abstract. A short summary (max 200 words) describing the goal, main methods, and key results. Replace this with a concise overview of your experiments and findings.

Contents

1 Introduction

Brief problem statement, datasets, and objectives.

2 Methods

Describe model architecture, preprocessing, augmentations, and training protocol.

2.1 Model

Explain the architecture (e.g., custom ResNet18).

2.2 Training setup

Optimizers, lr schedule, batch size, epochs, and any regularization.

3 Experiments

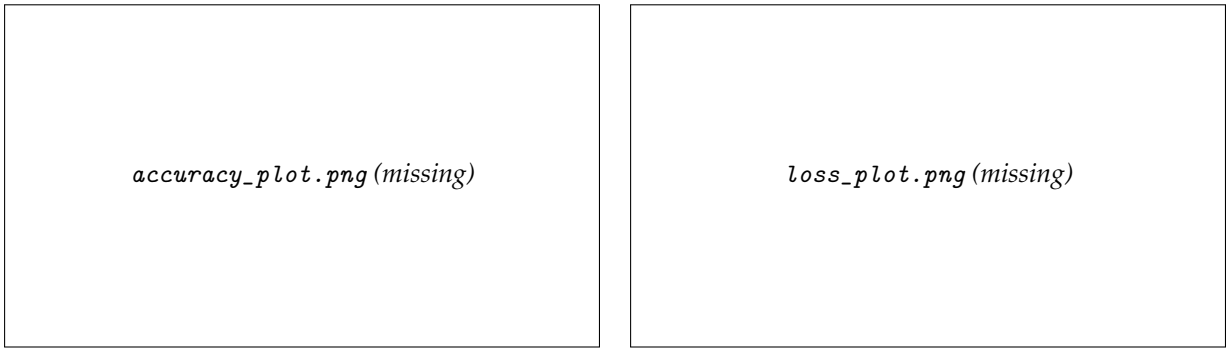
Dataset splits, metrics, and evaluation protocol.

3.1 Evaluation metrics

Accuracy, loss, robustness, UMAP visualizations, etc.

4 Results

Quantitative tables and qualitative figures.



(a) Train / validation accuracy (b) Train / validation loss

Figure 1: Training curves (export your notebook plots to 'template/figures/').

Table 1: Summary of main results

Model	Test accuracy (SVHN)	Test accuracy (MNIST)
Baseline ResNet18	0.00	0.00
+Augmentations	0.00	0.00
Pretrained Feat. Extractor	0.00	0.00

5 Discussion

Interpret the results and propose next steps.

6 Conclusion

Concise takeaways.

A Hyperparameters and Implementation Details

Full hyperparameters and experimental settings.

B Selected Code

Important snippets or reference to script files.

```
for epoch in range(epochs):
    model.train()
    for x,y in train_loader:
        # training step
        pass
```

Listing 1: Training loop (example)

C Additional Figures

Extra visualizations, UMAP plots, adversarial examples, etc.

References