Image Classification on the Animals-10 Dataset: CNN Baselines vs. ResNet-18 Transfer Learning

DATA 370 Final Project (Autumn 2025)
Qianrun Chen

Abstract

This project investigates image classification on the Animals-10 dataset by comparing (i) a small convolutional neural network (SmallCNN) trained from scratch and (ii) a transfer-learning approach based on an ImageNet, pret-rained ResNet-18 fine-tuned on the target task. To control class imbalance and computational cost, a balanced subset of 1,200 images (10 classes × 120 images/class) was constructed and split stratified into train/validation/test (70%/15%/15%). On this split, the 10-epoch SmallCNN run achieves 35.0% best validation accuracy and 36.1% test accuracy, while an extended 25-epoch SmallCNN run improves to 41.7% best validation accuracy and 38.9% test accuracy. The fine-tuned ResNet-18 reaches 91.1% best validation accuracy and 93.0% test accuracy. Learning curves show that the scratch-trained CNNs exhibit a large generalization gap, whereas ResNet-18 converges rapidly with minimal overfitting. These results support the standard hypothesis that pretrained representations substantially reduce sample complexity for small real-world image datasets.

1. Introduction

Animal image classification is a representative real-world computer vision task because images differ in pose, background, scale, illumination, and degree of occlusion. When only limited labeled data are available, training a deep network from scratch can suffer from poor generalization. Transfer learning addresses this by reusing features learned on a large source dataset (e.g., ImageNet) and adapting the model to a target task with fewer labeled examples.

2. Background and Related Work

Convolutional neural networks learn hierarchical representations: early layers capture edges and textures, while deeper layers encode object parts and global semantics. Residual networks (ResNets) introduce skip connections that ease optimization of deep

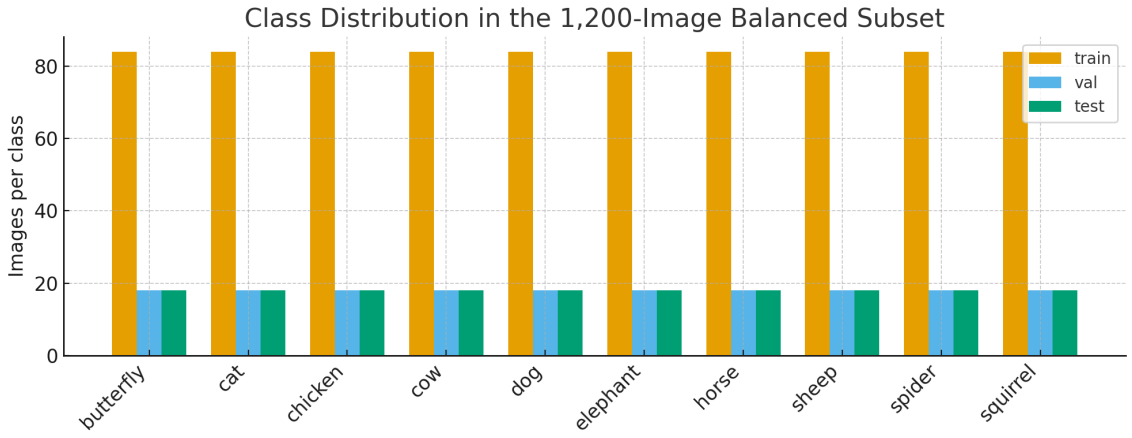
models by learning residual functions, enabling deeper architectures with strong empirical performance.

3. Dataset and Exploratory Data Analysis

The Animals-10 dataset (Kaggle) contains ~28k medium-quality web images in 10 animal categories: butterfly, cat, chicken, cow, dog, elephant, horse, sheep, spider, and squirrel. The raw dataset is organized by Italian folder names; a deterministic mapping to English labels is applied (e.g., cane→dog, cavallo→horse).

A class-balanced subset was created by sampling 120 images per class (N = 1,200). A stratified 70%/15%/15% split produces 840 training images, 180 validation images, and 180 test images (84/18/18 per class).

Figure 1. Class distribution (train/val/test) in the balanced subset.



4. Preprocessing and Data Pipeline

For the scratch-trained CNN runs, images are loaded with PIL, converted to RGB, resized to 128×128, and converted to float tensors scaled to [0, 1]. Training-time augmentation includes random horizontal flips (p=0.5) and a simple transpose-based augmentation (p=0.3) applied to the cached tensor.

For the ResNet-18 run, torchvision transforms are used: resize to 224×224, random horizontal flips for training, conversion to tensor, and normalization with ImageNet mean=[0.485, 0.456, 0.406] and std=[0.229, 0.224, 0.225].

5. Methods

5.1 Problem formulation

Let (x, y) denote an image x and its class label $y \in \{1,\ldots,10\}$. The model $f_\theta(x)$ outputs logits $z \in \mathbb{R}^{10}$. Training minimizes the empirical risk with the cross-entropy loss: $L(\theta) = -(1/N)\,\Sigma_i \log \mathrm{softmax}(z_i)[y_i]$.

5.2 Baseline: SmallCNN (10 epochs)

Architecture. The SmallCNN contains three convolutional blocks (3×3 conv + ReLU + batch normalization + 2×2 max-pooling), reducing the spatial resolution to 16×16, followed by a classifier (Flatten → Linear(128·16·16→256) → ReLU → Dropout(0.4) → Linear(256→10)).

Optimization. The model is trained with Adam (learning rate 1e−3, weight decay 1e−4), batch size 32, for 10 epochs. Model selection uses the best validation accuracy across epochs; the selected checkpoint is evaluated once on the test set.

5.3 Enhanced baseline: SmallCNN (25 epochs)

To provide a stronger scratch baseline, the same architecture is trained for 25 epochs under the same optimizer settings. This improves validation and test accuracy, but remains substantially below transfer learning performance.

5.4 ResNet-18 transfer learning (fine-tuning)

The transfer-learning model uses an ImageNet-pretrained ResNet-18. The final fully connected layer is replaced with a new 10-class classifier and the entire network is fine-tuned (all parameters trainable). Training uses Adam with a smaller learning rate (1e−4) and weight decay (1e−5) for 6 epochs.

6. Evaluation Protocol

All models are evaluated on a fixed stratified split of the balanced Animals-10 subset (N=1,200 images; 10 classes). Model selection is performed using validation accuracy when logged. The selected checkpoint is evaluated once on the held-out test set.

Primary metric: overall accuracy (correct predictions / total). We also report confusion matrices and per-class precision/recall/F1 for the final ResNet-18 model.

7. Results

7.1 Summary Metrics

Model	Training / selection	Best Val Acc	Test Acc
SmallCNN (scratch, 10 epochs)	Best validation epoch	35.00%	36.11%
SmallCNN (scratch, 25 epochs)	Best validation epoch	41.67%	38.89%
ResNet-18 (pretrained, fine-tuned)	Best validation epoch	91.11%	93.00%

Compared to the 10-epoch SmallCNN baseline, ResNet-18 improves test accuracy by 56.89 percentage points.

7.2 Learning Curves and Confusion Matrices

Baseline SmallCNN (10 epochs). Best validation accuracy: 35.00%. Test accuracy: 36.11%.

Figure 2. Baseline SmallCNN — confusion matrix on the test split.

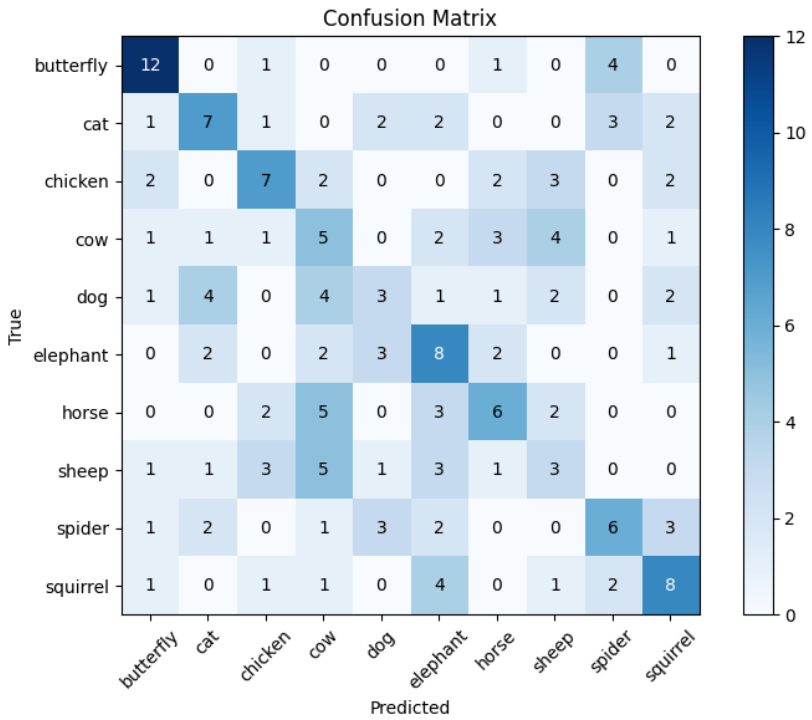


Figure 3. Baseline SmallCNN — train/validation accuracy curve.

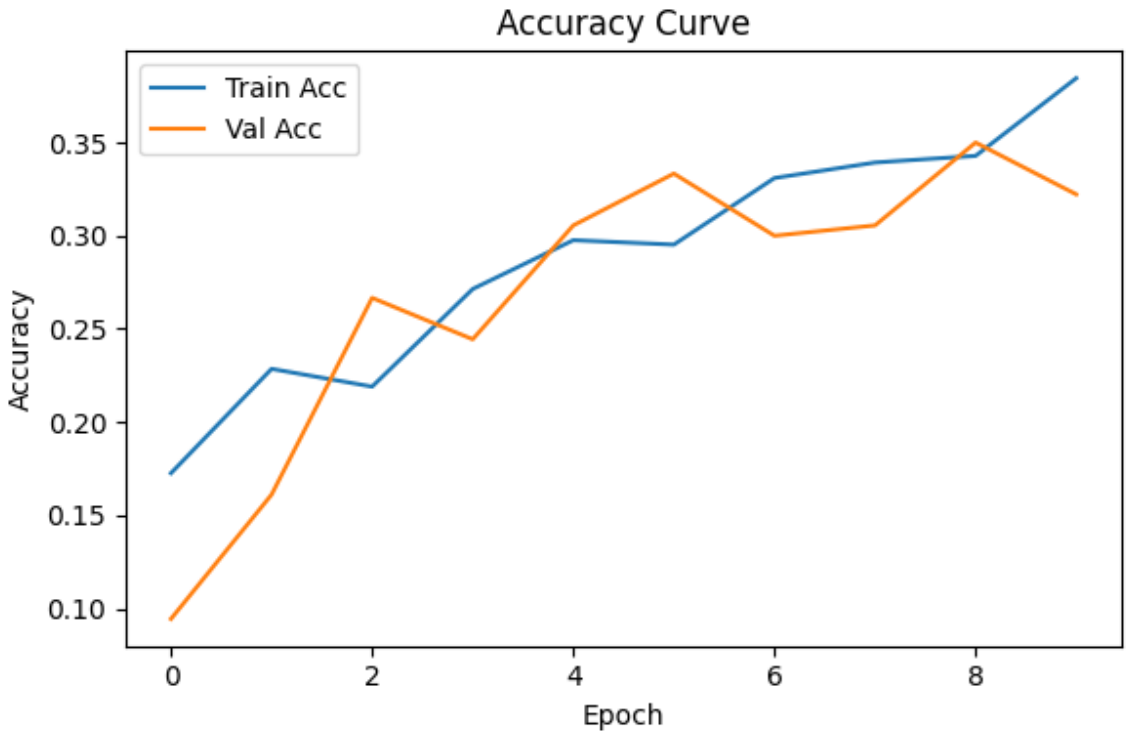
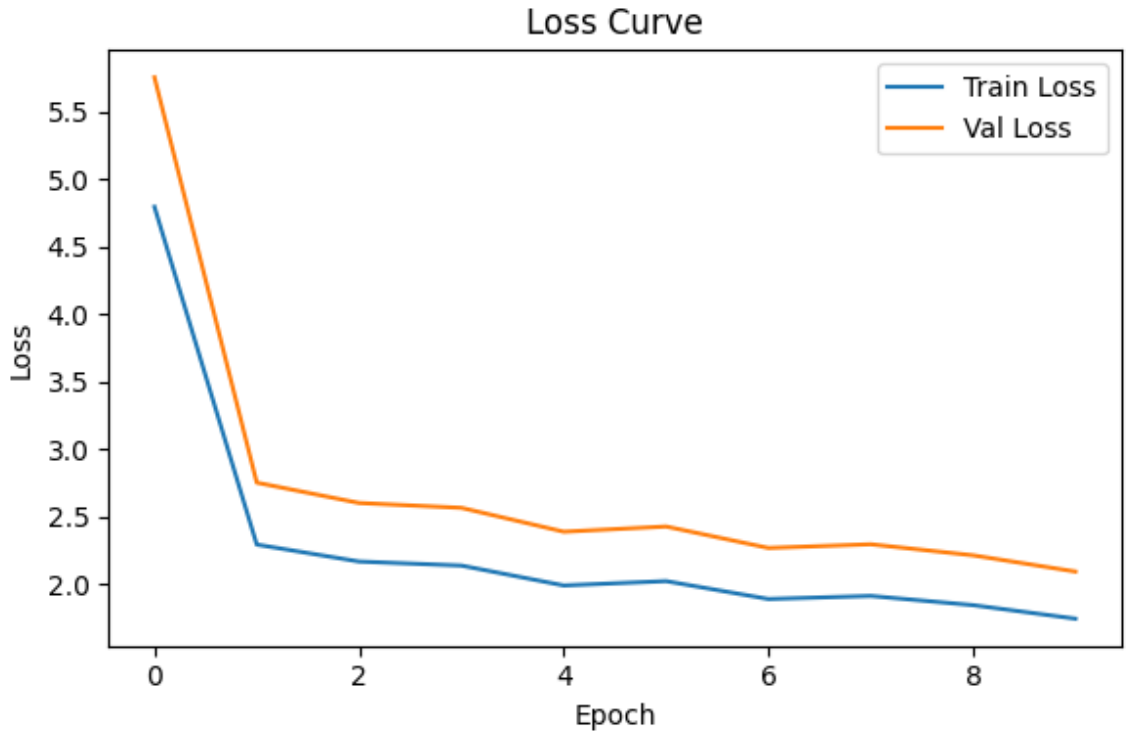


Figure 4. Baseline SmallCNN — train/validation loss curve.



Enhanced SmallCNN (25 epochs). Best validation accuracy: 41.67%. Test accuracy: 38.89%.

Figure 5. Enhanced SmallCNN — confusion matrix on the test split.

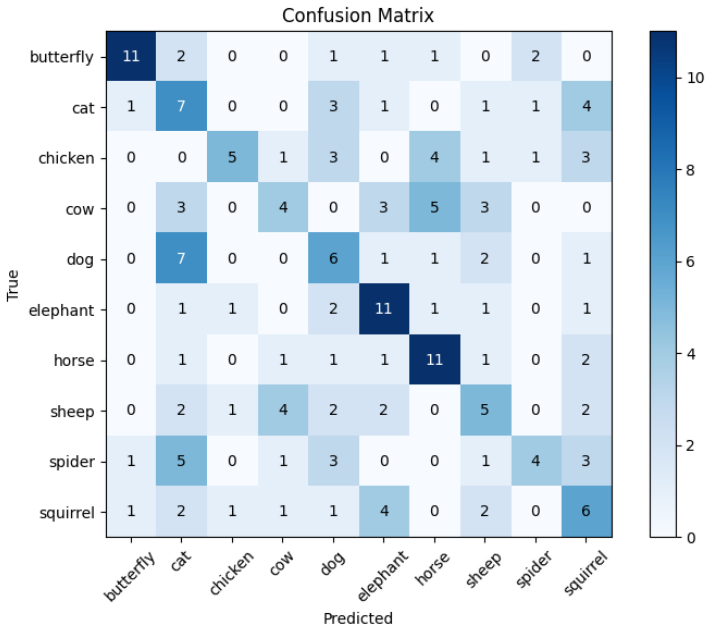


Figure 6. Enhanced SmallCNN — train/validation accuracy curve.

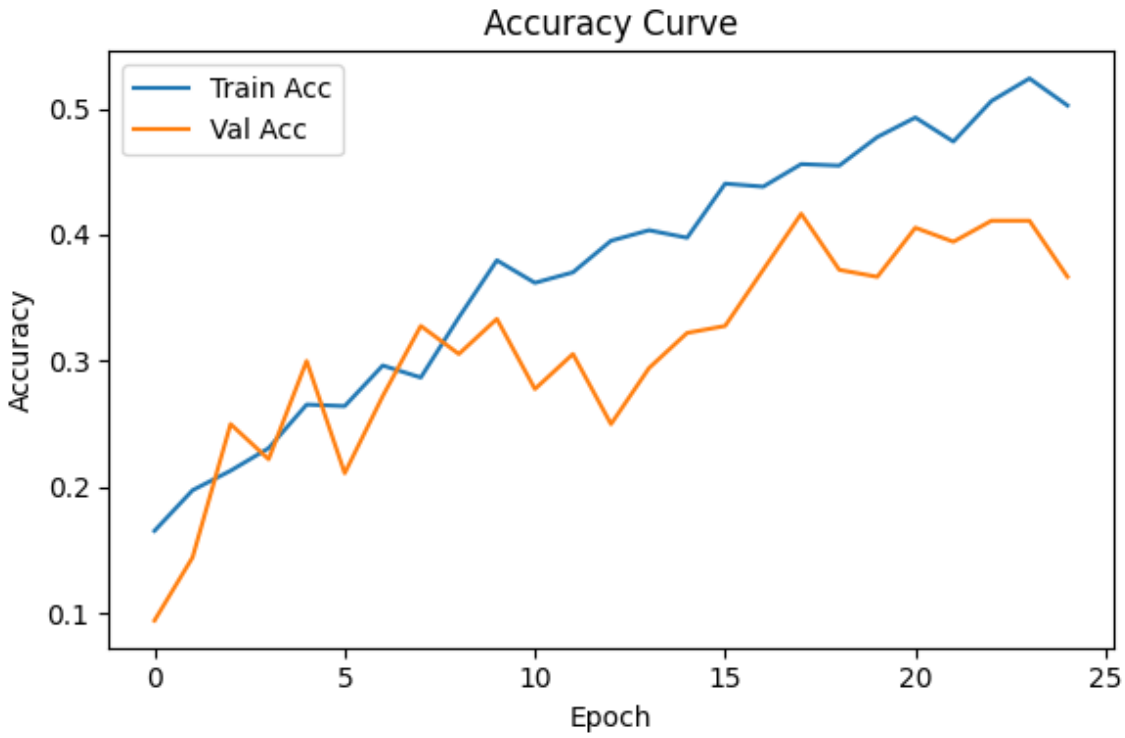
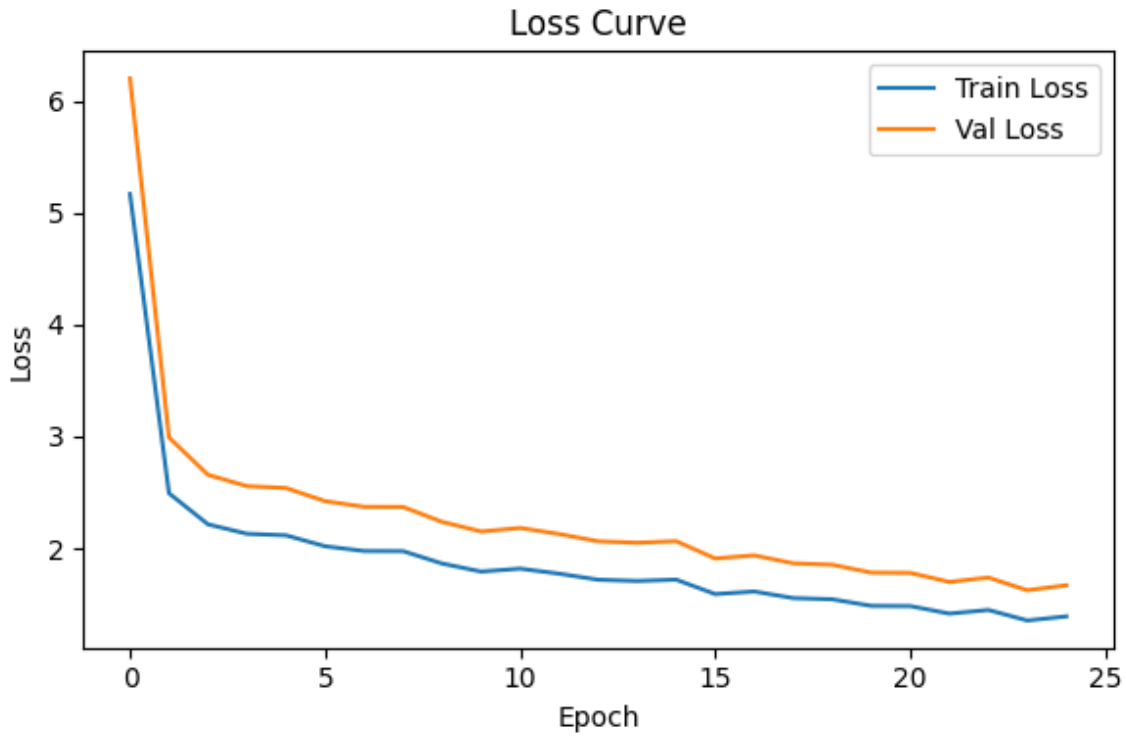


Figure 7. Enhanced SmallCNN — train/validation loss curve.



ResNet-18 (fine-tuning). Best validation accuracy: 91.11%. Test accuracy: 93.00%.

Figure 8. ResNet-18 — confusion matrix on the test split.

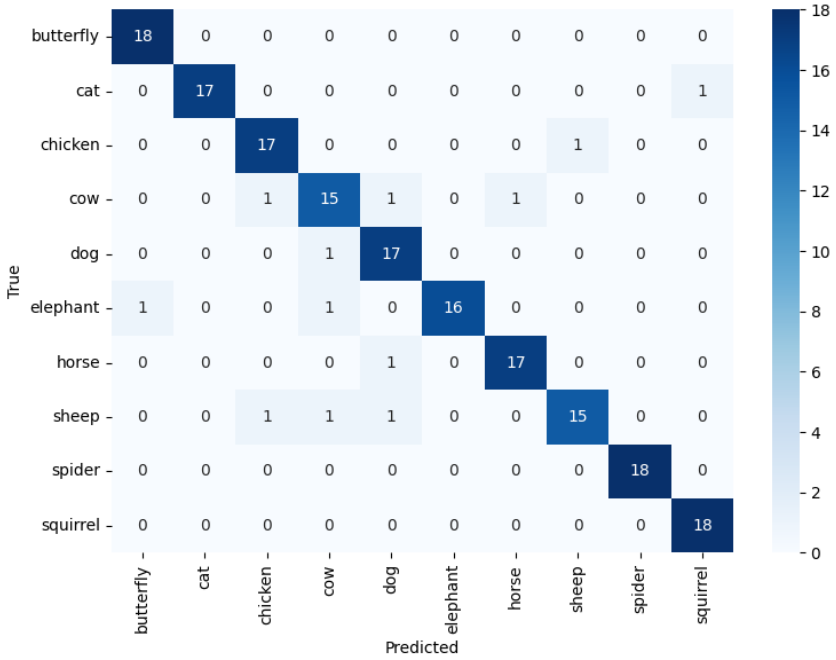


Figure 9. ResNet-18 — train/validation accuracy curve.

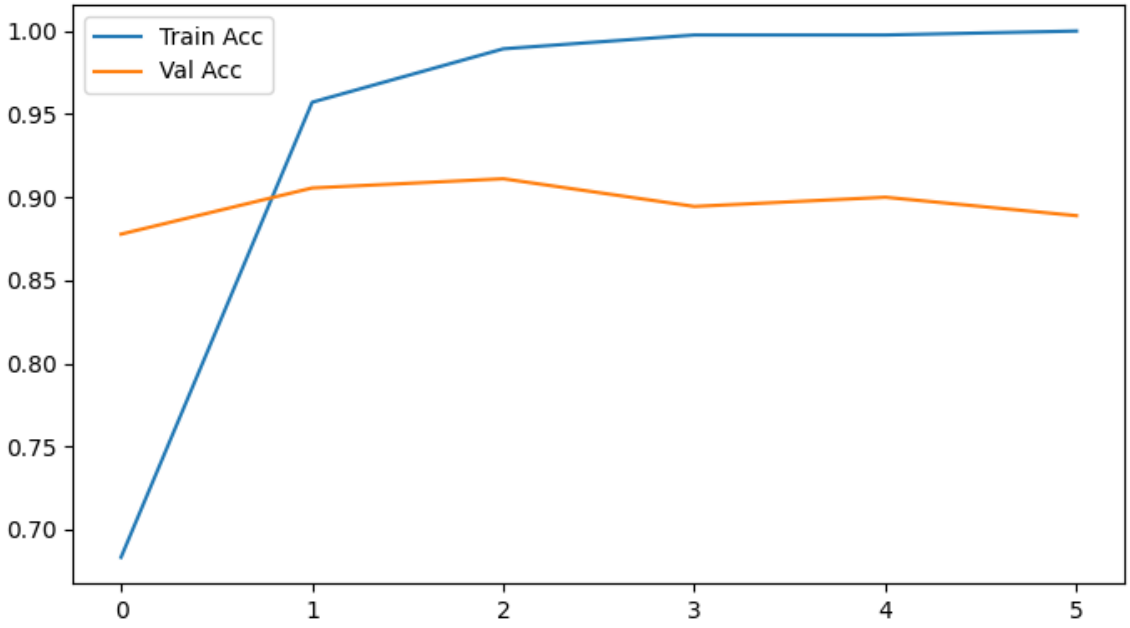
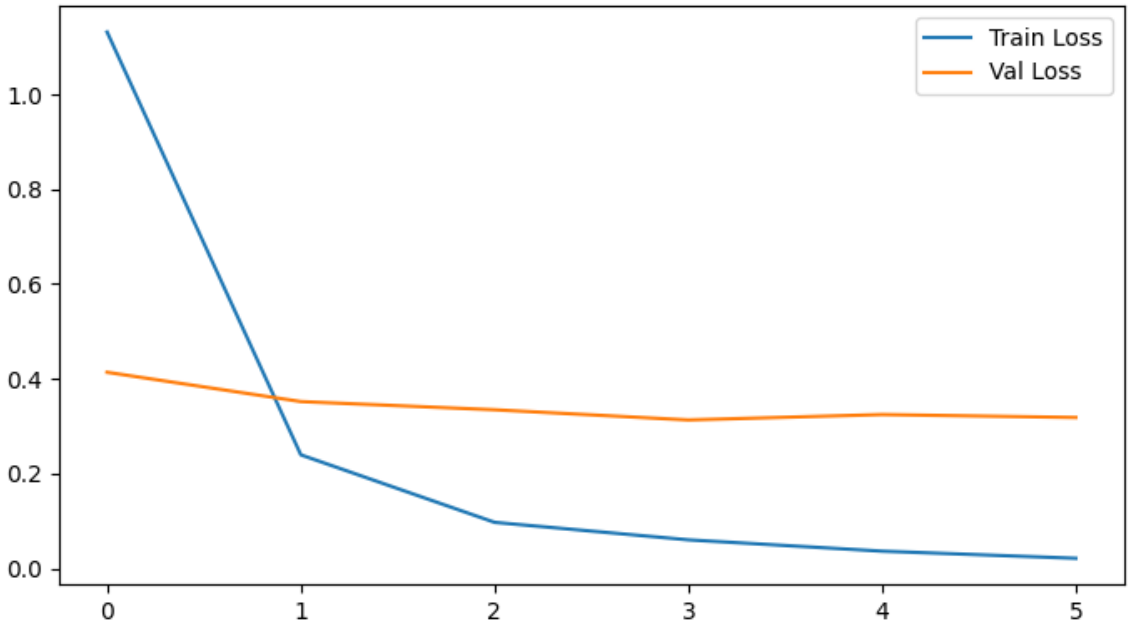


Figure 10. ResNet-18 — train/validation loss curve.



7.3 Per-Class Metrics (ResNet-18, Test Set)

Per-class precision/recall/F1 for ResNet-18 on the held-out test split (18 images per class).

Class	Precision	Recall	F1-score	Support
butterfly	0.95	1.00	0.97	18
cat	1.00	0.94	0.97	18
chicken	0.89	0.94	0.92	18
cow	0.83	0.83	0.83	18
dog	0.85	0.94	0.89	18
elephant	1.00	0.89	0.94	18
horse	0.94	0.94	0.94	18
sheep	0.94	0.83	0.88	18
spider	1.00	1.00	1.00	18
squirrel	0.95	1.00	0.97	18

8. Discussion

Across all runs, transfer learning delivers the greatest performance gains. Scratch-trained CNNs remain constrained by the need to learn robust low- and mid-level visual features from limited data, resulting in lower test accuracy and a noticeable generalization gap. In contrast, the pretrained ResNet-18 benefits from ImageNet features, achieving strong validation performance and high test accuracy after only a few epochs.

From the confusion matrices, most classes are predicted correctly by ResNet-18, with remaining errors concentrated on visually similar animals (e.g., cow vs. sheep) and images with complex backgrounds.

A limitation of this work is that results are reported on a single train/validation/test split of a 1,200-image subset. Future work could evaluate multiple random seeds/splits, apply stronger augmentation, and explore alternative fine-tuning strategies.

9. Conclusion

Using the balanced Animals-10 subset, the best performing approach is ImageNet-pre-trained ResNet-18 fine-tuning. ResNet-18 achieves 91.11% best validation accuracy and 93.00% test accuracy, substantially outperforming scratch baselines (10-epoch SmallCNN: 36.11% test; 25-epoch SmallCNN: 38.89% test). Overall, pre-trained representations greatly reduce sample complexity and deliver better generalization under limited labeled data.