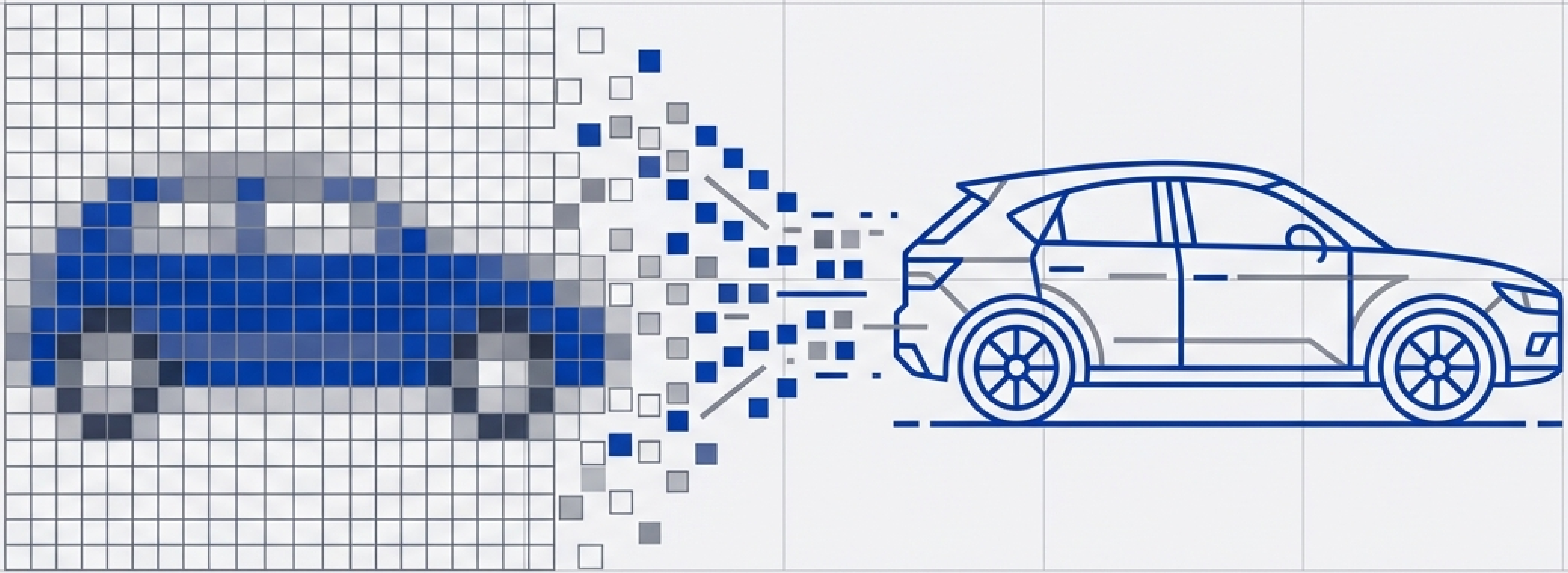
Mastering Low-Resolution Vision: CIFAR-10 Classification via ResNet50



A Transfer Learning Case Study in Optimization and Feature Extraction

Project: CIFAR-10 Computer Vision | **Architecture:** ResNet50 (ImageNet Weights) | **Method:** Two-Phase Transfer Learning

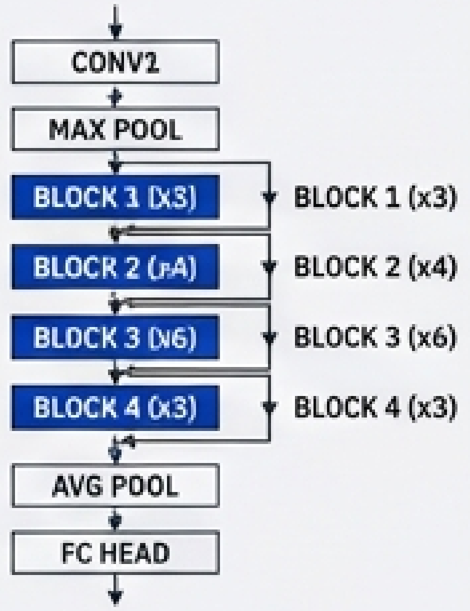
Project Dashboard: Achieving 92% Accuracy on Low-Fidelity Inputs

THE GOAL

Classify 10 distinct object classes in 32x32 RGB images using the CIFAR-10 dataset. The objective is to extract semantic meaning from low-pixel density inputs.

THE ARCHITECTURE

ResNet50 (ImageNet Weights) + Custom Classifier Head.



THE CONSTRAINT

Downsampled training set limited to 10,000 images to rigorously test efficiency vs. variance constraints.

THE RESULT

92% Accuracy

Stable convergence at 0.24 Loss.

Insight: Pre-trained ImageNet weights successfully compensated for low pixel density, 'inheriting' edge detection capabilities that would be impossible to train from scratch on this dataset size.

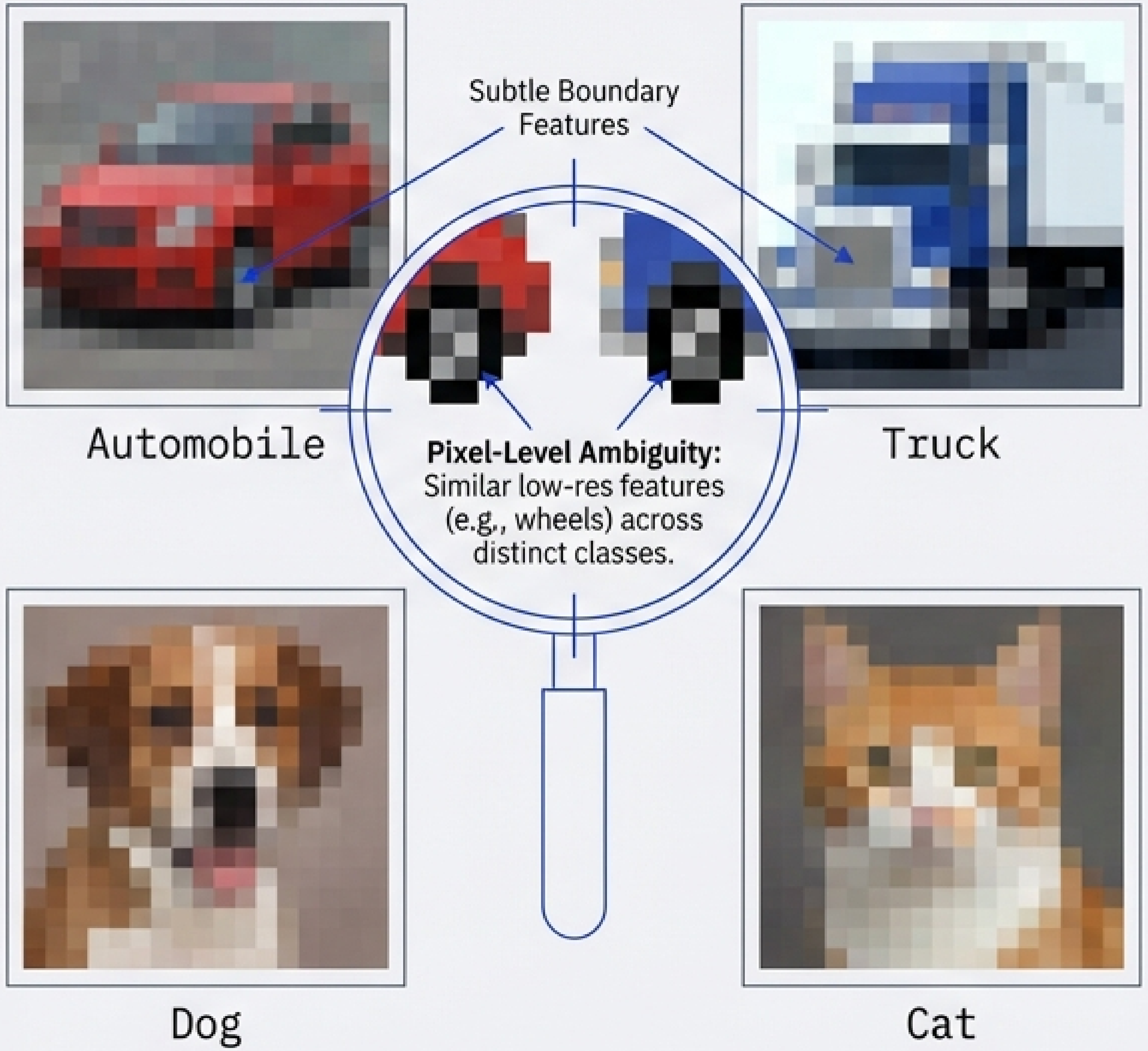
Complexity Hidden in Low Resolution: The CIFAR-10 Challenge

Dataset Specifications

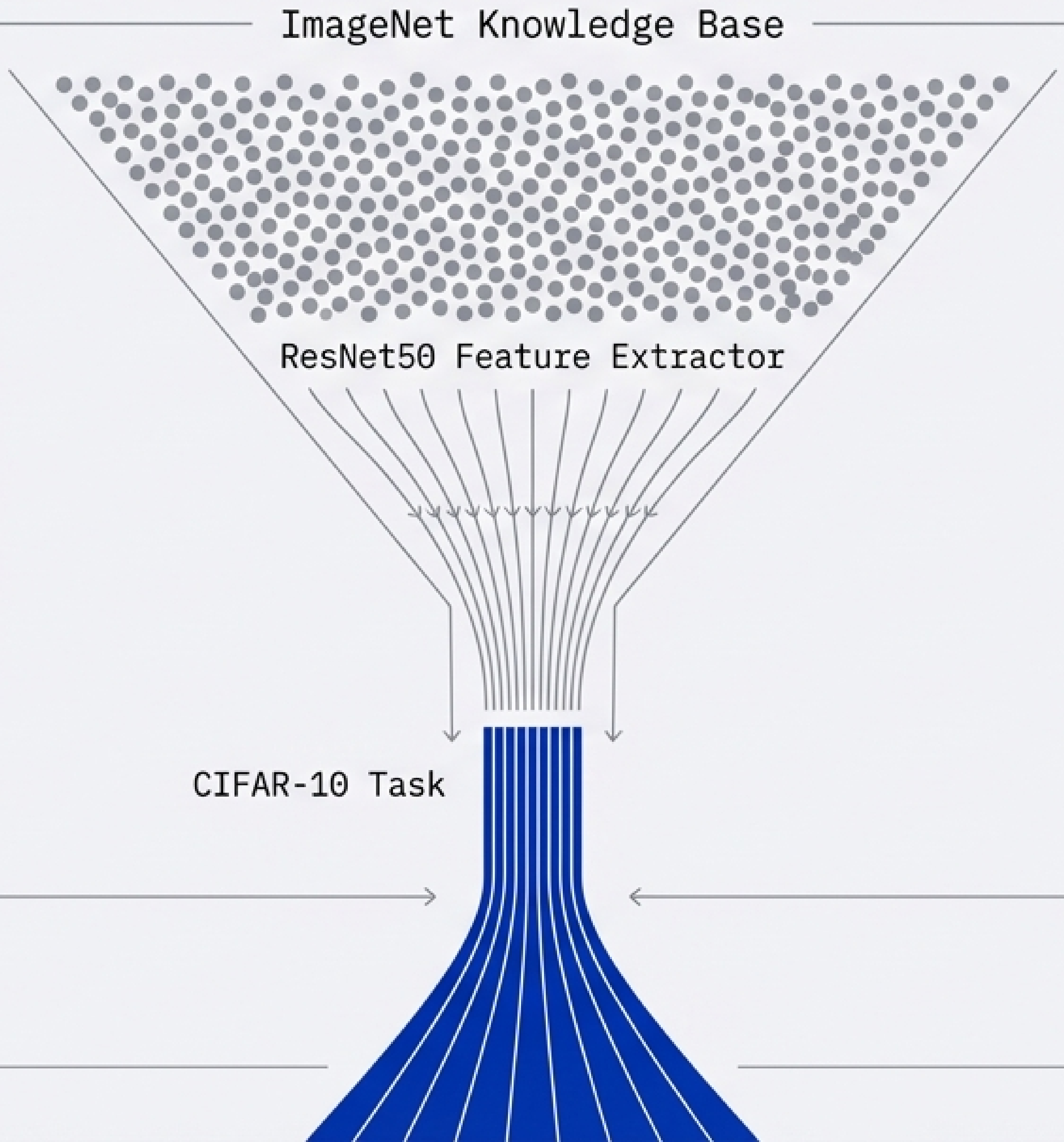
- **Volume:** 60,000 Total Images (Downsampled to 10k Training / 10k Testing)
- **Dimensions:** 32x32 pixels, 3 Channels (RGB)
- **Classes:** 10 mutually exclusive categories

The Semantic Challenge

The primary difficulty is nuance within low resolution. “Automobile” includes sedans and SUVs, while “Truck” is strictly heavy-duty vehicles. The model must learn subtle boundary features rather than just size or shape.



Leveraging 'Inherited' Vision to Prevent Overfitting



The Problem: Training from Scratch

Training a deep network on small 32x32 images causes massive overfitting or convergence failure. The model attempts to learn edge detection from too little data.

The Solution: Transfer Learning

1. **Inheritance**: We borrow pre-trained weights for edge/texture detection.
2. **Focus**: The model only learns how to combine these existing features to classify the 10 specific CIFAR classes.

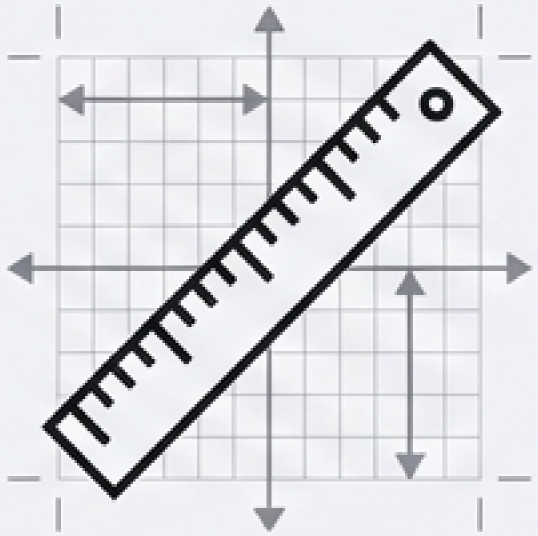
Technical Workflow: Preparing Data for Deep Residual Networks

Raw Input

CIFAR-10 Data

Normalization

Mean subtraction based on ImageNet distribution (centering data). Not simple /255 scaling.



Dimensional Integrity

Input shape fixed at (32, 32, 3). No resizing performed to avoid computational bloat.

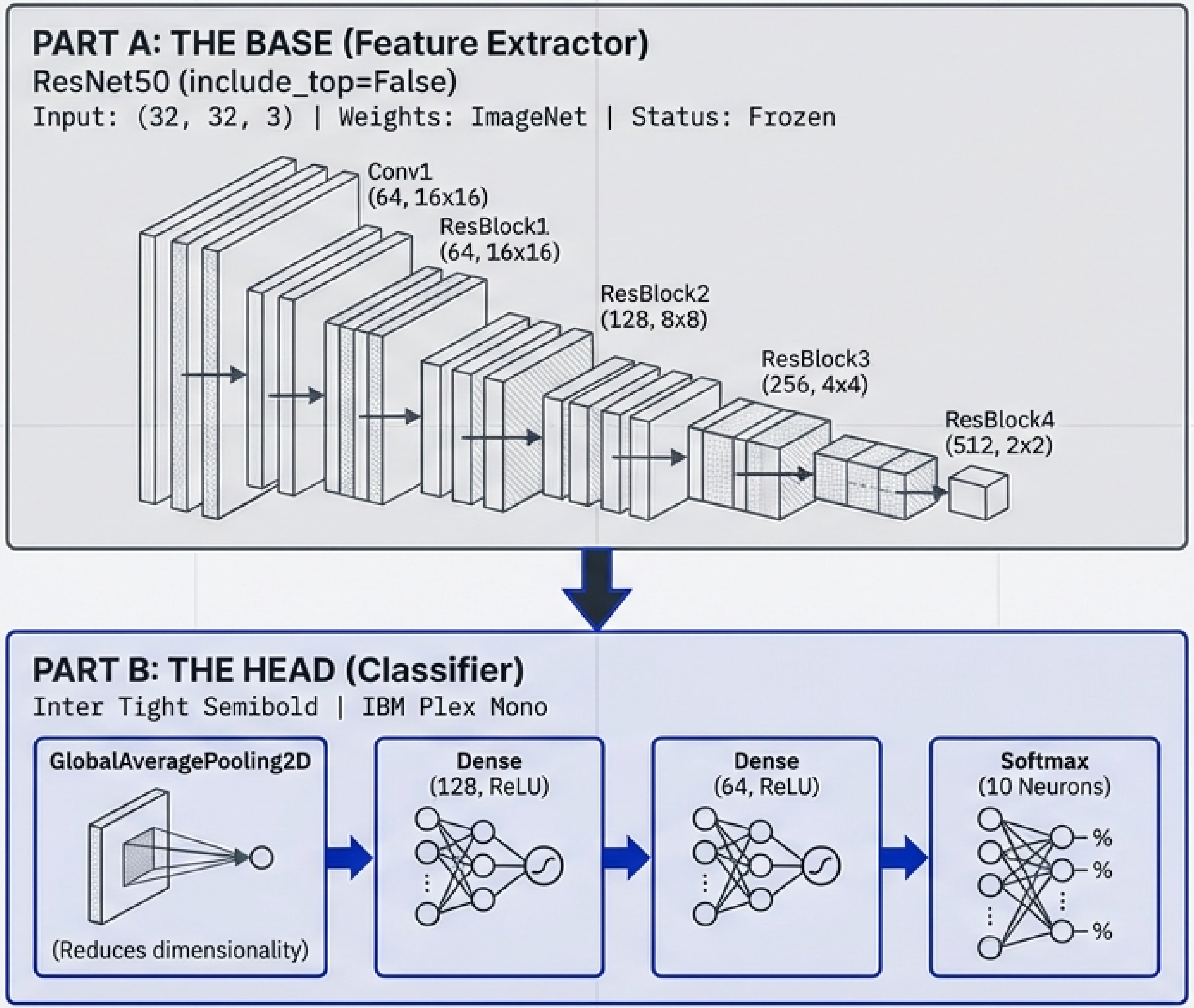
$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Encoding

One-Hot Encoding targets for 10-class output.

```
> input_shape = (32, 32, 3)
> preprocess_input(x) # Applies Caffe-style mean subtraction
> to_categorical(y, 10)
```

The Hybrid Design: Frozen Base + Custom Head



The Custom Head maps the complex ResNet features to the 10 specific classes of CIFAR-10.

A Two-Phase Training Strategy for Stability

Phase 1: Feature Extraction
10 Epochs

`Base layers are FROZEN.`

We train only the custom head. This prevents large, random gradients from destroying the delicate, pre-trained ImageNet weights.

Phase 2: Fine-Tuning
10 Epochs

`Base layers are UNFROZEN.`

We train the entire network with a very low learning rate. This allows the deep ResNet filters to adapt slightly to CIFAR-specific textures.

`Total Training Duration:` **`20 Epochs`**`. Strategy minimizes` **`"Catastrophic Forgetting"`**`.`

Reaching Convergence: 92% Accuracy

92%
Final Top-1 Accuracy

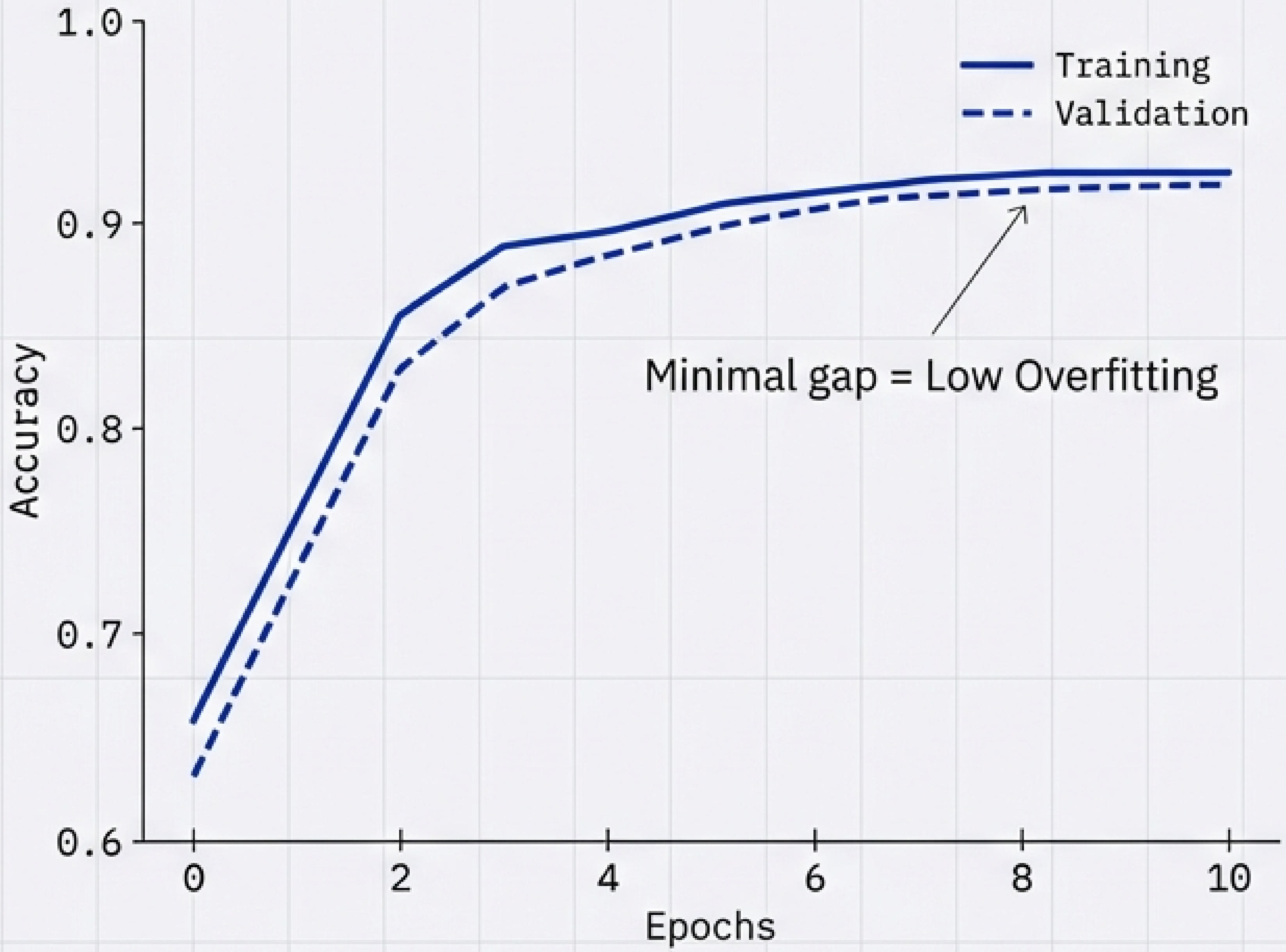
0.24
Final Loss

Training Stage	Accuracy	Loss	State
Epoch 1	75%	0.55	Initial Head Training
Epoch 10	92%	0.24	Post Fine-Tuning

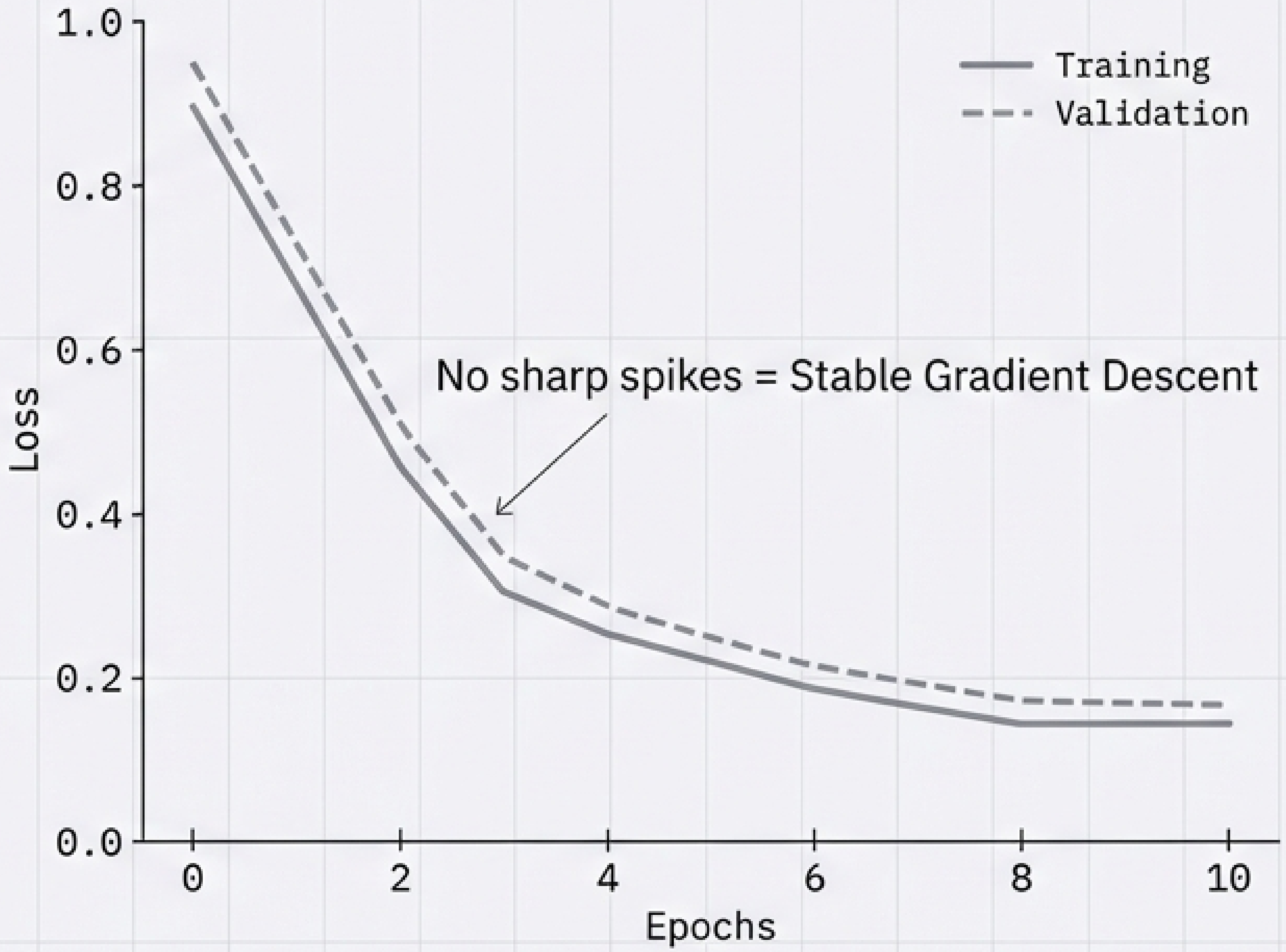
Despite the downsampled training set (10,000 images), the model achieved high accuracy, effectively validating the transfer learning hypothesis. The jump from 75% to 92% demonstrates the efficacy of the fine-tuning phase.

Smooth Optimization Indicates Robust Generalization

Model Accuracy

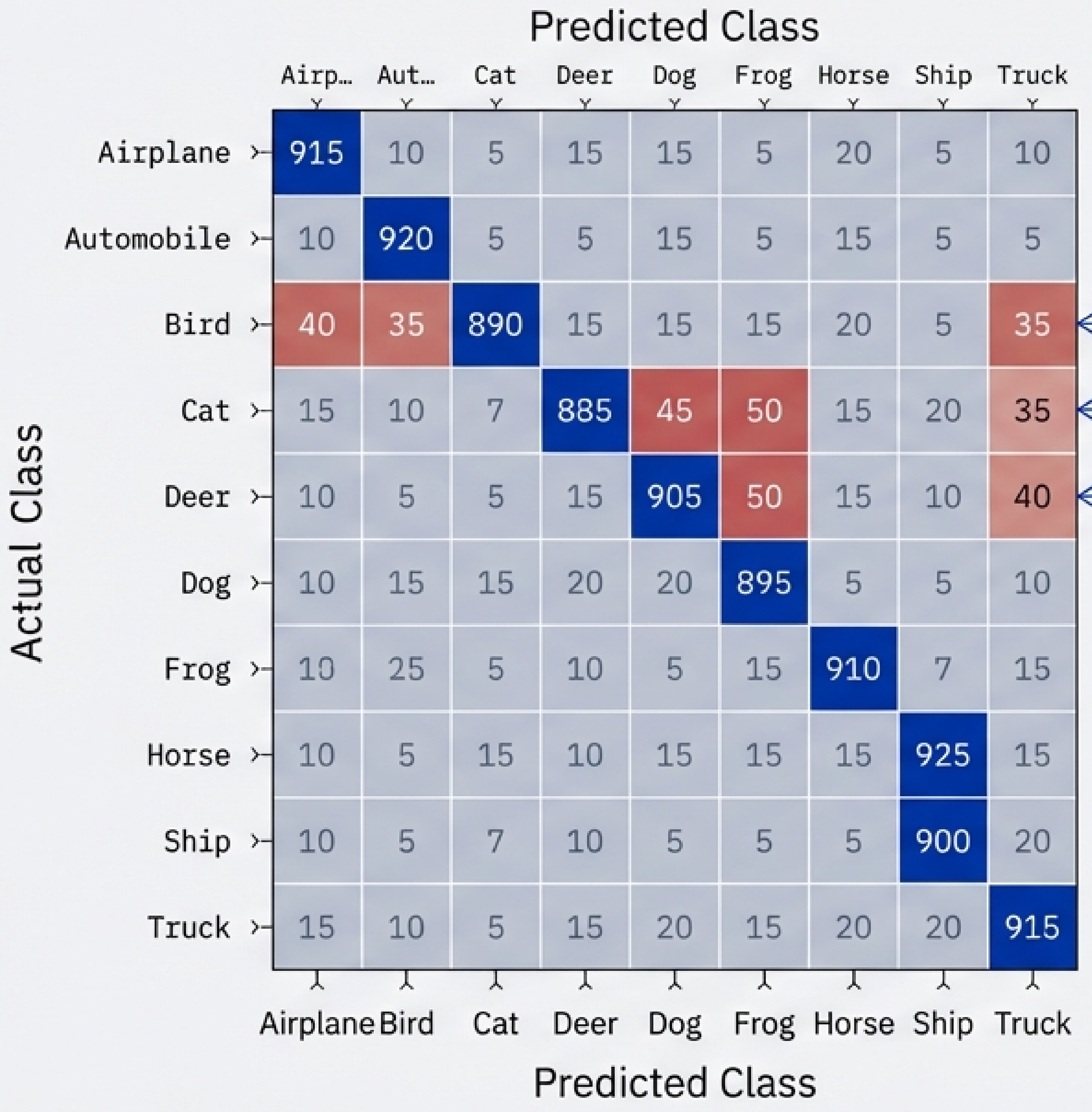


Model Loss



Convergence stabilizes around Epoch 8. The parallel nature of Train/Val lines proves the model generalizes well to unseen data.

Confusion Matrix Analysis: The Limits of Resolution



The Texture Leak (Cat vs. Dog)

High confusion. At 32x32 pixels, fur textures and facial features are indistinguishable. The model relies on shape, which is similar for both quadrupeds.

The Background Bias (Bird vs. Airplane)

Birds are often misclassified as Airplanes. **Reason:** Both classes frequently feature a dominant blue sky background, leading to context interference.

Performance Divergence: Geometric vs. Organic

Structured / Rigid

- Automobile: F1-Score **0.95**
- Ship: **High Precision**

Strong geometric features (metal edges, wheels, silhouettes) are preserved well even at low resolution.

Organic / Natural

- Cat: F1-Score **0.80**
- Dog: F1-Score **0.79**

Texture nuances (fur, ears, tails) blur together at 32x32 pixels, causing lower precision.

Insight: The model effectively transfers 'vision' for rigid structures but struggles with organic nuance without higher resolution input.

Strategic Project Evaluation (SWOT)

STRENGTHS ↗

- High Accuracy (92%)
- Efficient Feature Extraction (`ResNet50`)
- Stable Convergence

WEAKNESSES ↘

- Semantic Similarity struggles (`Cat vs Dog`)
- Sensitivity to background noise (`Bird vs Plane`)
- Computationally expensive relative to image size

OPPORTUNITIES ↗

- Data Augmentation (`Rotations/Flips`)
- API Deployment
- Unfreezing more layers

THREATS ↘

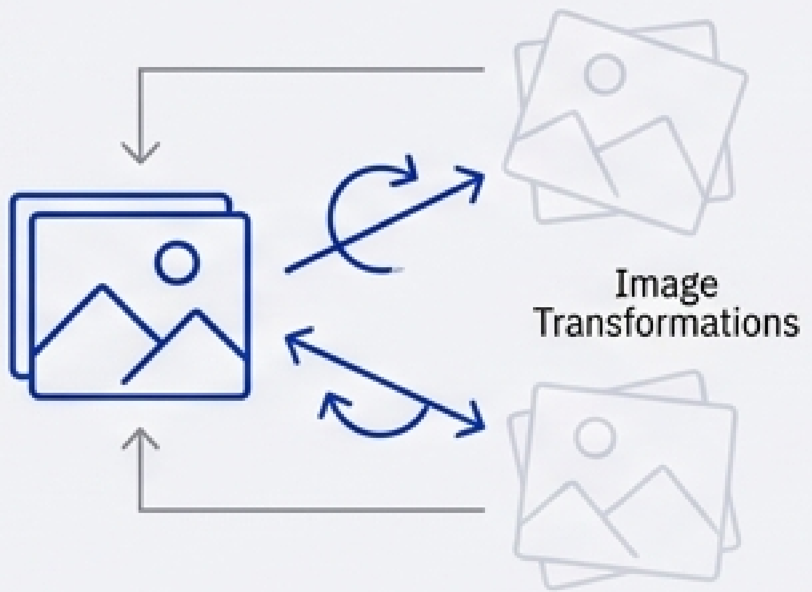
- Risk of Overfitting (needs regularization)
- Domain Shift on high-res real-world data

Optimization Roadmap: The Path to 99%

Action: Data Augmentation

Problem: Orientation Bias

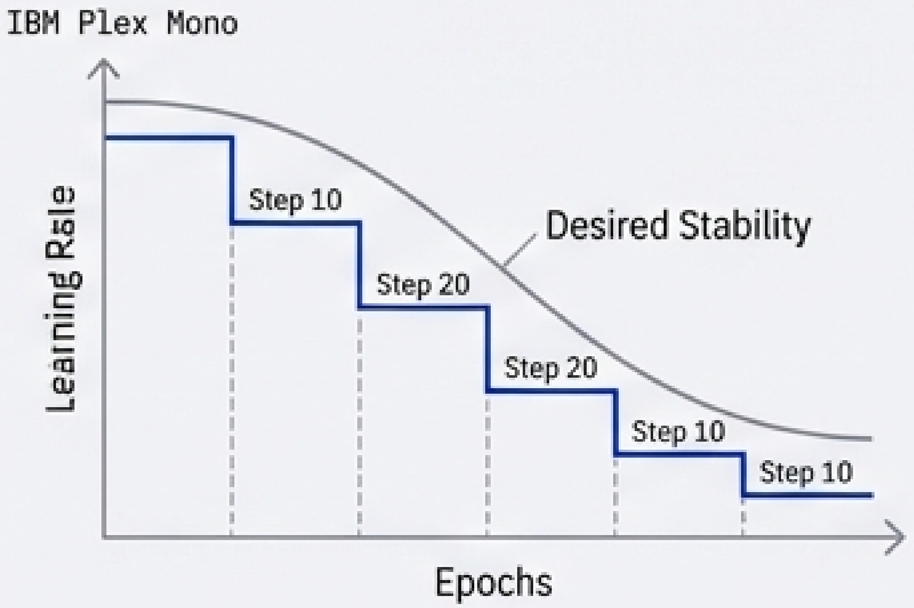
Introduce random rotations and horizontal flips to force the model to learn invariant features.



Action: Scheduler Implementation

Problem: Fine-tuning Instability

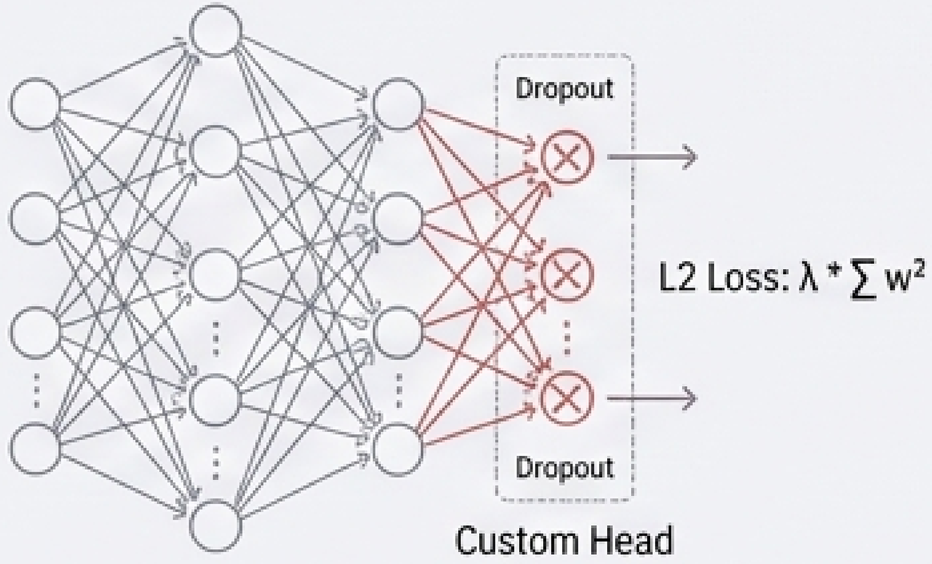
Implement automated learning rate decay to smooth the descent in later epochs.



Action: Regularization

Problem: Small Dataset Variance

Add Dropout layers or L2 regularization to the custom head to prevent overfitting on the 10k image set.



Final Verdict: A Solid Baseline for Computer Vision

ResNet50 has proven to be a formidable foundation, successfully bridging the gap between low-resolution inputs and high-performance classification.

Critical Takeaway: While the model mastered structural recognition (Vehicles), the ‘Cat vs. Dog’ confusion highlights that resolution constraints eventually hit a ‘semantic ceiling’ that requires data augmentation to break.

STATUS: READY FOR AUGMENTATION & DEPLOYMENT TESTING

Technical References & Stack

- Dataset:
  - CIFAR-10 (Canadian Institute for Advanced Research)
- Architecture:
  - ResNet50 (He et al., Microsoft Research)
  - Weights: ImageNet (Pre-training)
- Tools & Environment:
  - TensorFlow / Keras
  - Google Colab (GPU Runtime)