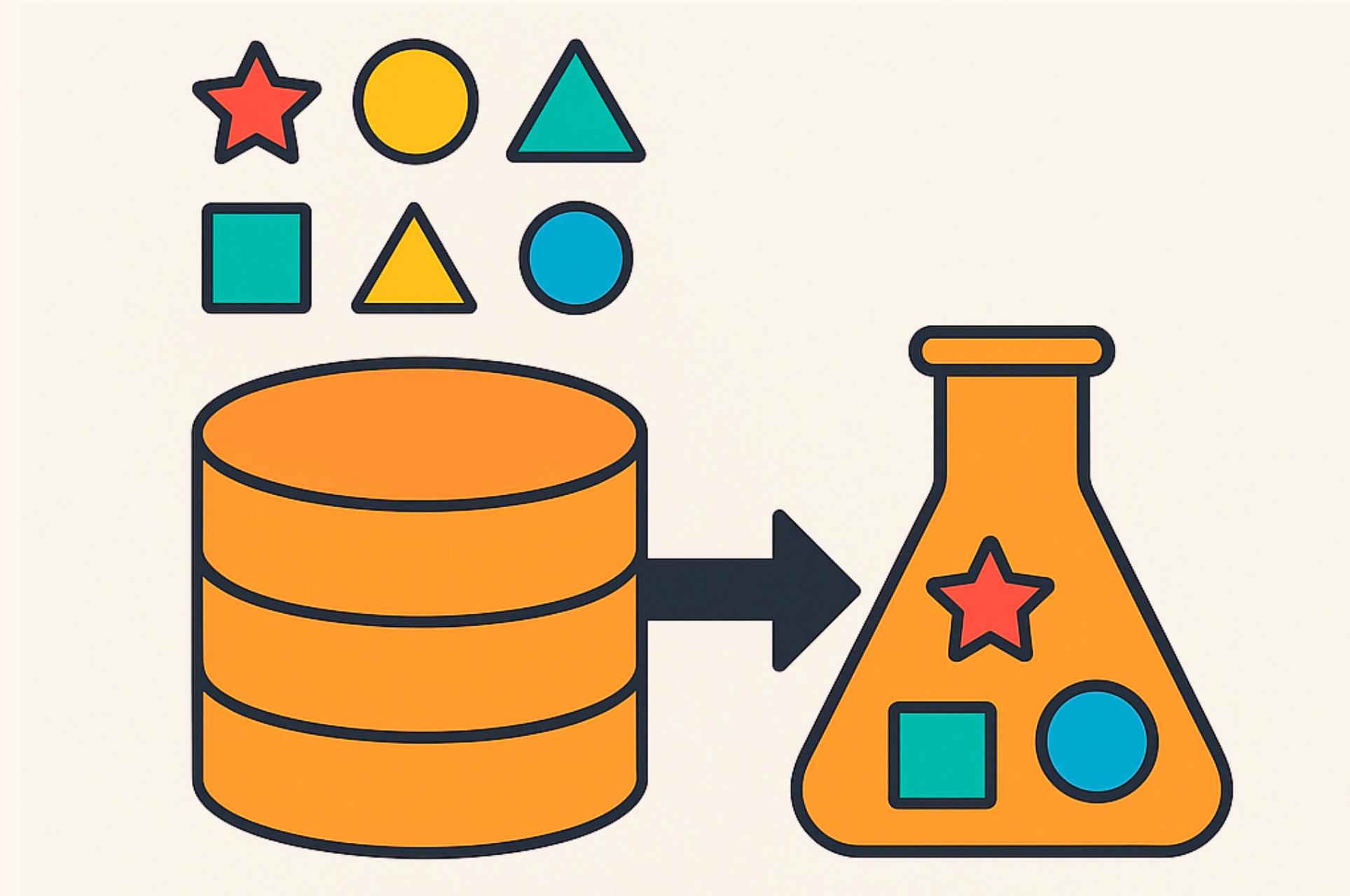


DATA DISTILLATION USING TRAJECTORY MATCHING



Ömer Faruk San
150220307

Mustafa Kerem
Bulut 150220303

The Challenge: Big Data, Big Cost

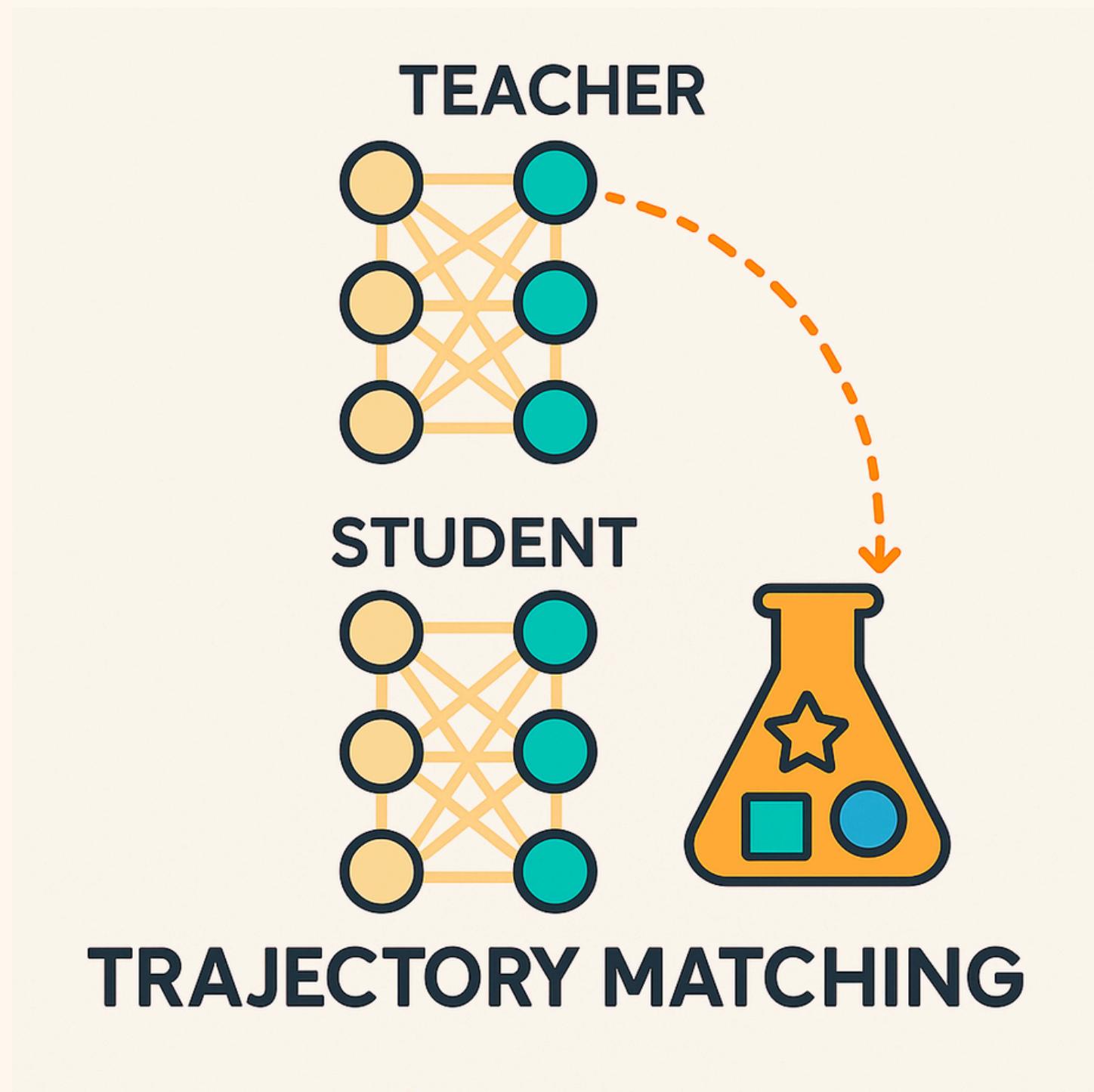
Why is this a problem?

- Deep learning needs LOTS of data.
- Training is slow & expensive.
- Our Goal: Can we create a tiny "distilled" dataset that acts like the big one?



Our Approach: Matching Learning Paths

- Teacher Model: Trained on full CIFAR-10 (50,000 images). We save its "learning path" (parameters over time).
- Synthetic Data: We create 100 "smart" synthetic images from scratch.
- Trajectory Matching:
- Train a "student" model on our 100 synthetic images.
- Optimize synthetic images so student's learning path mimics the teacher's path.



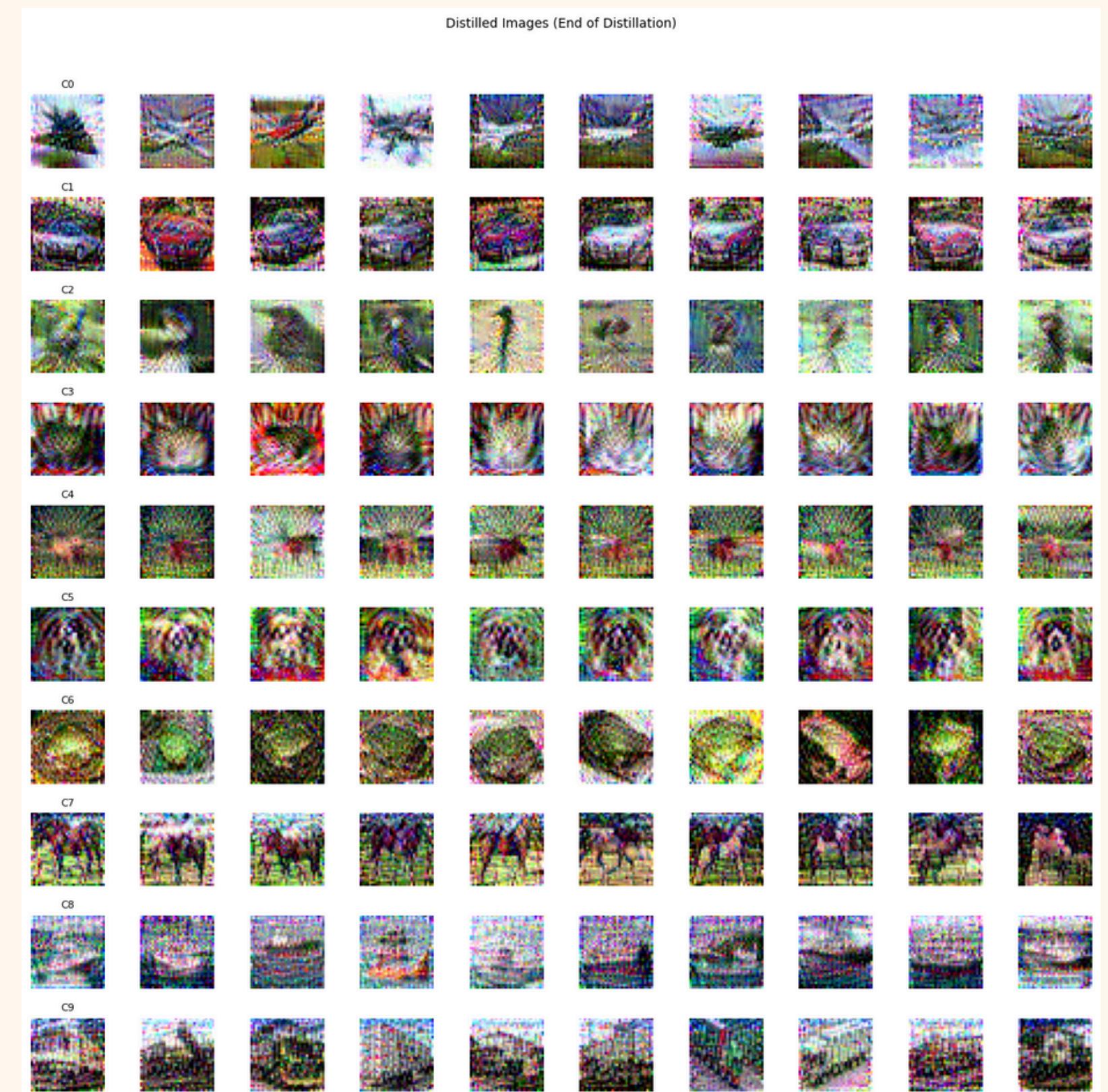


What We Found

Teacher (Full CIFAR-10): 83.31% Test Accuracy

Our Distilled Dataset (100 images):

- Generated 100 synthetic images (show the actual plot of your 100 images here!).
- Model trained only on these 100 images:
- Training Accuracy (on distilled set): 100% (perfectly learned the small set)
- Test Accuracy (on CIFAR-10 test): 10.28%



Distilled images

Key Takeaways & Next Steps



- **Learned:** Dataset distillation is complex! We successfully implemented the core idea.
- **Result:** Our 100 images learned by the model, but didn't generalize well to new data (10.28% vs 83% baseline).
- **Why?** Simplified matching, limited optimization time, trajectory granularity.
- **Future:** More advanced matching, longer training, finer teacher paths.

**THANKS
FOR
LISTENING**

