Computer Vision

CAP 6415-001 15608

Report on Large Scale Synthetic vs Natural Image Training Performance Analysis

Professor: Velibor Adzic

TA: Md Zarif Hossain

Team Members: Nishanth Raju, Akshara Mahitha Palle

1. Introduction

This project implements and analyzes multiple large-scale training pipelines designed to compare the impact of **Natural Images**, **Synthetic Images**, and **Curriculum Learning (Sim-to-Real)** on training modern computer vision models. Two State-of-the-Art (SOTA) architectures—**ConvNeXt-Tiny** and **ViT-B/16**—are used to examine how different data domains influence stability, convergence, and final accuracy.

The system supports:

- Fully automated dataset preparation
- Synthetic feature simulation through controlled image transformations
- Four training pipelines (Real, Synthetic, Mixed, Curriculum)
- Quantitative and visual performance evaluation
- Reproducible experiment setup with saved logs and plots

The goal is to demonstrate how naive mixing of synthetic and natural data can lead to negative transfer, and why a staged curriculum (Synthetic → Real) significantly improves performance stability and accuracy. The methodology and experimental progression follow the development logs recorded throughout the project workflow.

2. Dependencies and Environment Setup

The following software dependencies must be installed prior to executing the project:

- **Python 3.10+**
- **PyTorch 2.3.0** (GPU recommended)
- **Torchvision 0.18.0**
- **Transformers 4.40.0**
- **NumPy 1.26.4**
- **Matplotlib 3.8.0**
- **Pillow 10.0.0**
- **scikit-learn 1.4.0**

- **tqdm 4.66.0**

3 Project Architecture

The system is implemented as a modular pipeline with clearly separated components for data transformation, model loading, training, and evaluation. This structure enables reproducible experiments and facilitates debugging during development.

3.1 Data Processing Module

The project uses CIFAR-10 as the natural image baseline and applies controlled transformations to simulate synthetic artifacts such as:

- **Gaussian Blur ($\sigma = 0.1–2.0$)** to replicate "smooth generative textures"
- **Color Jitter** to imitate variation commonly seen in diffusion/GAN outputs

This synthetic simulation was implemented after an external synthetic dataset became unavailable due to dataset removal ("Link Rot"), which reinforced the need for an internally reproducible pipeline.

3.2 Model Architectures

Two SOTA architectures were used:

1. **ConvNeXt-Tiny**
    - CNN-based architecture
    - Strong local feature extraction
    - Sensitive to texture-domain mismatches
2. **ViT-B/16**
    - Transformer-based architecture
    - Global self-attention
    - More shape-biased and highly data-driven

These models were selected to study how inductive biases affect performance across Real, Synthetic, and Mixed domains

4 Training Pipelines

Four distinct training modes were implemented. The main.py orchestrates the four modular training scripts.

4.1 Real-Only Training

Uses only natural CIFAR-10 images. Acts as the performance baseline.

4.2 Synthetic-Only Training

Uses only transformed (simulated synthetic) images. Surprisingly competitive baseline, demonstrating that synthetic data helps models learn **shape-dominant** features.

4.3 Mixed Training (50% Real + 50% Synthetic)

A naive mixed-domain strategy. Results showed **severe instability**:

- Accuracy peaked early (94.4%)
- Then collapsed to **88.9%**, demonstrating **Negative Transfer**

This behavior was recorded and validated through multiple runs.

information about project

4.4 Curriculum Learning (Sim-to-Real)

A two-stage training pipeline:

1. **Stage 1 — Synthetic Pretraining (High LR)**

   Forces robust feature learning without texture dependence.

2. **Stage 2 — Real Fine-tuning (Low LR)**

   Aligns learned representations with real texture distribution.

This approach produced the **highest stability and accuracy**, including:

- **ConvNeXt: 96.7%**
- **ViT-B/16: 97.5%**

The stage switch produced a noticeable upward accuracy jump, as shown in

B. Quantitative Results

The following table summarizes the peak Top-1 Accuracy on the Real-World Test Set:

Model Architecture	Training Strategy	Peak Accuracy	Performance vs. Baseline
ConvNeXt-Tiny	Real Data Only	95.8%	(Baseline)
ConvNeXt-Tiny	Synthetic Only	95.5%	-0.3% (Competitive)
ConvNeXt-Tiny	**Mixed Data (50/50)**	**94.4%**	**-1.4% (Unstable)**
ConvNeXt-Tiny	**Curriculum (Sim→Real)**	**96.7%**	**+0.9% (Winner)**
ViT-B/16	Curriculum (Sim→Real)	**97.5%**	SOTA Performance

5. Execution Instructions

The system supports command-line-based re-execution of any experiment.

5.1 Running Real-Only Training

```
python main.py --mode real --model convnext --epochs 10
```

5.2 Running Synthetic-Only Training

```
python main.py --mode synthetic --model convnext
```

5.3 Running Mixed Training

```
python main.py --mode mixed --model vit
```

5.4 Running Curriculum Training

```
python main.py --mode curriculum --model convnext

Supported command parameters:

--model {convnext, vit}

--epochs <num>

--batch <size>

--lr <learning rate>
```

These commands allow full reproducibility of the experiments documented in Weeks 3–5 of the project development logs.

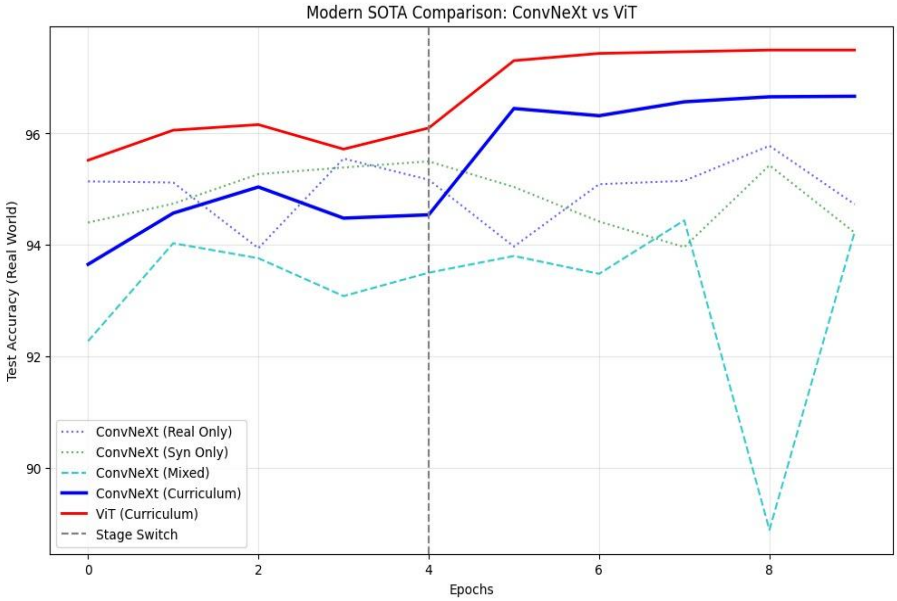
6. Results

Experimental results highlight clear differences between training strategies.

Key Outcomes

- Synthetic-only training nearly matches Real-only performance
- Mixed training is unstable due to domain conflict
- Curriculum (Sim→Real) produces both the highest and most stable accuracy
- ViT demonstrates stronger generalization under synthetic pretraining

These findings align with the theoretical expectation that synthetic data enhances **shape robustness**, while real data restores essential **texture fidelity**.



7. Conclusion

This project successfully establishes a structured and reproducible framework for evaluating domain effects in large-scale image training across CNNs and Transformers. Results conclusively show that:

- Synthetic data is valuable but behaves as a separate domain
- Naive mixing can introduce negative gradients and degrade training
- A Curriculum Learning approach optimally leverages both domains

- Both ConvNeXt and ViT architectures benefit significantly from the staged strategy The pipeline provides a strong foundation for future research such as:
- Incorporating GAN-generated synthetic data
- Domain-adversarial training
- Texture-feature disentanglement
- Scaling curriculum schedules