Comparative Analysis of Dimensionality Reduction Methods for CNN Features

Sahil Dadhwal (917598320), Purva Khadke (924112595), Sarvesh Halbe (925581073)

Introduction & Motivation

Convolutional Neural Networks have fundamentally changed how machines interpret images by learning to extract meaningful features automatically. Models like ResNet50 [1], trained on millions of images, can represent any picture as a 2048-dimensional vector that captures its essential characteristics. However, these high-dimensional features create significant practical challenges. Storing features for a million images consumes around 8GB of memory, which causes real-time similarity searches to be slow, and visualizing anything beyond 3D is nearly impossible for human interpretation.

Traditional dimensionality reduction approaches have notable limitations. PCA struggles with non-linear patterns in the data. UMAP was not designed primarily for classification tasks. Autoencoders [2] compress features without understanding which dimensions matter most for discrimination. None of these methods address the challenges caused by high-dimensional feature representations in production systems.

Recent advances in transformer architectures [5] suggest a viable solution.Transformers with their self-attention mechanisms [5] can adaptively identify and prioritize the discriminative features, causing compression of 2048 dimensions down to just 128 without losing critical information [3]. This project investigates whether transformer self-attention enables important feature compression compared to traditional methods.

Companies like Pinterest and Shopify process billions of visual searches daily, these smarter compression could reduce infrastructure costs by millions while enabling real-time performance. Mobile apps, autonomous vehicles, and edge devices with strict memory constraints all require lightweight yet accurate feature representations. In this project, we are comparing PCA, UMAP, autoencoders, and transformers on CIFAR-10 [4] to understand their relative strengths and weaknesses for feature compression. Our goal is to provide practical insights into method selection based on accuracy, efficiency, and interpretability trade-offs.

Research Questions

RQ1 Which dimensionality reduction method best preserves classification accuracy?

RQ2 Do transformers outperform autoencoders for feature compression in terms of accuracy and reconstruction quality?

RQ3 Which method produces the most interpretable low dimensional representations (measured by cluster separation, class cohesion in 2D projections)?

RQ4 What compression ratios (32D, 64D, 128D, 256D) provide optimal trade-offs between accuracy and efficiency?

Technical Approach

Pipeline: Images → ResNet50 → 2048-D features → Reduction (128-D) → MLP Classifier → Evaluation
Methods:

- **PCA:** Baseline, sklearn implementation
- **UMAP:** Manifold learning, n_neighbors=15
- **Autoencoder:** [2048 → 1024 → 512 → 128 → 512 → 1024 → 2048], MSE loss
- **Transformer:** Novel approach which treats 2048-D vector as sequence of 16 tokens (128-D each), 4 layers, 8 heads, positional encoding, learned projection to 128-D

Evaluation: Classification accuracy, reconstruction error, training time, 2D visualizations

Datasets & Experiments

Primary: CIFAR-10 (60k images, 10 classes). **Secondary:** CIFAR-100 (100 classes)

Experiments: **(1)** Dimension sweep [32,64,128,256] for optimal compression (RQ4), **(2)** Method comparison at 128-D (RQ1 and RQ2), **(3)** Interpretability analysis with visualizations and attention weights (RQ3)

Tech Stack

PyTorch 2.0+, scikit-learn, umap-learn, matplotlib, NumPy, pandas

Timeline & Milestones

Week	Key Milestones
1 to 2	Setup, feature extraction (ResNet50), implement PCA/UMAP, baseline results
3 to 4	Autoencoder & Transformer implementation, training, hyperparameter tuning
5 to 6	Classification experiments, dimension sweep, interpretability analysis, report writing

Expected Outcomes

- **Performance:** Transformers expected to achieve 2-4% higher accuracy than traditional methods at 128D compression
- **Optimal Compression:** 128D identified as best accuracy-efficiency trade-off
- **Interpretability:** Attention weights reveal discriminative feature patterns
- **Guidelines:** Decision framework for practitioners selecting methods
- **Deliverables:** Trained models, comparison tables, visualizations, technical report, open-source code

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[2] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[3] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.

[4] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Tech. Rep., University of Toronto, 2009.

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.