Exploring Visual Recognition using Tiny ImageNet and Fashion-MNIST

Kenzhebek Taniyev
Nazarbayev University
Astana, Kazakhstan
kenzhebektanyev@gmail.com

Husnain Rasool
Nazarbayev University
Astana, Kazakhstan
husnain.rasool@nu.edu.kz

Anas Rahim
Nazarbayev University
Astana, Kazakhstan
anas.rahim@nu.edu.kz

***Abstract*—This paper evaluates the performance of DenseNet networks when used to classify images of small resolutions. The small resolution along with many classification groups makes this task extremely demanding. The DenseNet structure combines network design elements which provide performance improvement and require decreased parameters in comparison to typical (Convolution Neural Networks) CNN models. The research evaluates this approach by comparing standard CNN models and DenseNet-based networks which operate on two datasets including Tiny ImageNet and FashionMNIST. The experiments demonstrate DenseNet surpasses others in both accuracy performance alongside quick convergence during training periods. The study outcomes prove that dense architecture designs excel in resource-limited cases of detailed image classification. A complete codebase which includes implementation specifications allows readers to reproduce the research and conduct experiments.**

Keywords: DenseNet, Convolutional Neural Networks, Tiny ImageNet, Image Classification, Deep Learning, FashionMNIST

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

The application of deep learning methods has recently brought significant success to multiple computer vision activities especially when processing images for categorization. CNNs have established themselves as the industry standard for visual recognition because they successfully learn complex representations of features. The developing field requires portable yet expandable models that achieve precision goals accompanied by reduced computation requirements. The recent advancement in deep learning has brought about **Dense Convolutional Networks (DenseNet)** as a solution for feature propagation improvement alongside vanishing gradient problem mitigation.

CNN performance evaluation has commonly relied on ImageNet but Tiny ImageNet delivers complex tasks which lie between standard and difficult datasets. The combination of 200 classes along with 64×64 pixel resolution ensures that the test evaluates how models perform with fine-grained and low-resolution inputs making it suitable for mobile vision systems and embedded AI. Colorful fashion-related patterns form the basis of the FashionMNIST dataset where users validate model performance on basic visual patterns.

The machines utilizing DenseNet models remain underutilized in resource-restricted areas because of challenges related to model size and training demands. The study evaluates model performance between a baseline CNN and a DenseNet-enhanced target model by applying testing on Tiny ImageNet and FashionMNIST datasets as well as baseline CNN.

*1) **Our main contributions are::***

1) **Comparative Evaluation** We built two network models with the first one being a basic CNN structure and the second including architectural elements from DenseNet through deeper layers and normalization and pooling techniques.
2) **Feasibility Testing Across Datasets** The two models undergo training and evaluation on FashionMNIST and Tiny ImageNet while demonstrating the effects of architectural complexity on accuracy rates and convergence speed as well as efficiency throughout different domains.
3) **Open Source Implementation** The research provides open access to all code as well as training logs and experimental results for both reproducibility and further experimentation purposes: GitHub Repository: https://github.com/KenzhebekTaniyev/densenet-tinyimagenet

II. Motivation and Related Work

A. *Motivation*

Designing efficient image classification models that work well across varying dataset complexities is a crucial challenge in deep learning. Two datasets that represent opposite ends of this complexity spectrum are **FashionMNIST** and **Tiny ImageNet**.

FashionMNIST consists of 28×28 grayscale images of clothing items in 10 categories. These images are relatively simple, featuring clean backgrounds and clear object boundaries. In contrast, Tiny ImageNet contains 200 classes of real-world objects in color images of size 64×64 pixels. These images are inherently more complex due to factors such as cluttered backgrounds, variable lighting conditions, and high intra-class variability.

These contrasting examples motivate the need for robust architectures like **DenseNet**, which can scale effectively across both simple and challenging datasets. By incorporating dense connections between layers, DenseNet facilitates better gradient flow and feature reuse, offering a potential solution for both small-scale and fine-grained image classification problems.

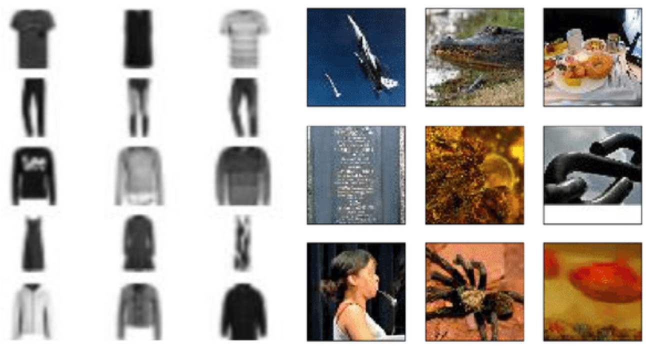

Fig. 1. Sample images from FashionMNIST (right 3 col.) and Tiny ImageNet (left 3 col.). Tiny ImageNet poses a more difficult challenge due to object variability.

B. Related Work

Research into convolutional neural networks (CNNs) has led to the development of a wide range of architectures aimed at improving classification accuracy and training efficiency. Notably, Densely Connected Convolutional Networks (DenseNet) have gained attention for their unique connectivity patterns and compact design. Abai and Rajmalwar [1] demonstrated that DenseNet models outperform traditional CNNs on the Tiny ImageNet benchmark by promoting feature reuse and mitigating the vanishing gradient problem. Their work emphasizes how dense connections across layers can lead to more stable and faster convergence, especially when working with smaller and more intricate datasets [2].

Another important contribution in this domain is the CNN Explainer by Wang et al. [3], an interactive tool that visualizes how CNNs process input data. Their work highlights the internal dynamics of CNN architectures and supports the notion that architectural components like batch normalization and pooling significantly influence model interpretability and performance. These insights directly inform our architectural enhancements in the target CNN model.

In parallel, the design and analysis of benchmark datasets has also played a key role in evaluating model robustness. Le and Yang [2] introduced Tiny ImageNet as a scaled-down version of the original ImageNet dataset. With 200 object categories and reduced resolution, it poses a significant challenge for lightweight models and has become a widely used testbed for model performance under limited computational resources.

Similarly, the FashionMNIST dataset has emerged as a modern replacement for MNIST, featuring grayscale images of fashion items across 10 categories. In their work, Haji et al. [4] proposed an enhanced CNN for FashionMNIST classification, incorporating dropout and deeper convolutional layers to boost generalization. Nocentini et al. [5] took this further by employing multiple CNNs in an ensemble approach, reporting notable improvements in classification performance. Although our study focuses on single-model evaluation, these findings demonstrate that even relatively simple datasets can benefit from architectural refinements.

Collectively, these studies establish the relevance of DenseNet-like architectures and motivate our comparative study across both low-complexity (FashionMNIST) and high-complexity (Tiny ImageNet) datasets.

C. Datasets

To evaluate the generalization capabilities of our CNN models, we utilize two publicly available image classification datasets that differ significantly in complexity and scale: Fashion-MNIST and Tiny ImageNet.

1) Fashion-MNIST: Fashion-MNIST is a widely used benchmark dataset introduced by Zalando Research to serve as a drop-in replacement for the original MNIST dataset. It contains grayscale images of clothing items, offering a more realistic and challenging test case for evaluating simple vision models.

Key statistics of the Fashion-MNIST dataset:

- **Number of Classes:** 10 fashion item categories (e.g., T-shirt/top, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag, Ankle boot).
- **Total Number of Images:** 70,000 grayscale images.
  - **Training Set:** 60,000 images (6,000 images per class).
  - **Test Set:** 10,000 images (1,000 images per class).
- **Image Resolution:** 28×28 pixels.
- **Source:** Collected and curated by Zalando, a German online fashion retailer.

Due to its clean and balanced nature, Fashion-MNIST is ideal for evaluating lightweight CNN architectures in low-complexity environments.

2) Tiny ImageNet: Tiny ImageNet is a subset of the ILSVRC (ImageNet Large Scale Visual Recognition Challenge) dataset, curated to enable deep learning experiments under limited computational resources. It represents a mid-scale classification challenge due to its relatively high number of classes and reduced image resolution.

Key statistics of the Tiny ImageNet dataset:

- **Number of Classes:** 200 object categories (a subset of the 1,000-class ImageNet).
- **Total Number of Images:** 110,000 color images.
  - **Training Set:** 100,000 images (500 images per class).
  - **Validation Set:** 10,000 images (50 images per class).
  - **Test Set:** 10,000 images (50 images per class; labels not publicly provided).
- **Image Resolution:** 64×64 pixels.
- **Source:** Derived from the original ImageNet dataset by downsampling images to 64×64 pixels.

Tiny ImageNet introduces intra-class variation, background clutter, and fine-grained recognition challenges that make it suitable for evaluating deeper and more complex network architectures.

These datasets, taken together, allow us to examine how architectural improvements (e.g., dense connectivity, normalization layers) impact performance across tasks of varying difficulty.

III. Methodology Overview

This section outlines the methodology adopted to evaluate and compare two convolutional neural network (CNN) architectures across datasets of varying complexity. Our approach builds directly upon the three main contributions outlined in Section I: comparative evaluation of CNN models, feasibility testing on dual datasets, and open-source reproducibility.

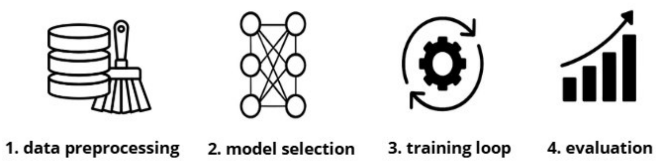

Fig. 2. Methodology pipeline: data preprocessing $\rightarrow$ model selection $\rightarrow$ training loop $\rightarrow$ evaluation.

A. Model Development and Selection

We implemented two CNN architectures from scratch using PyTorch: a *BaselineCNN*, which serves as a minimal benchmark model, and a more advanced *TargetCNN*, which incorporates architectural enhancements inspired by DenseNet design principles. Both models are trained under identical conditions for fairness in comparison.

1) BaselineCNN: The baseline model consists of two convolutional layers followed by max pooling operations and fully connected layers. It includes no normalization or advanced pooling techniques. This model is used to establish a lower-bound benchmark for both Fashion-MNIST and Tiny ImageNet.

Architecture Highlights:

- Input: 3-channel image (converted from Fashion-MNIST grayscale)
- Two convolutional blocks: Conv $\rightarrow$ ReLU $\rightarrow$ MaxPool
- One hidden dense layer with 128 units
- Final classification layer (output size = number of classes)
- No normalization or dropout layers

2) TargetCNN: The target model adopts several techniques derived from DenseNet and modern CNN architectures. Specifically, it features deeper convolutional blocks, Batch Normalization layers after each convolution, and a Global Average Pooling layer before the final classifier. These additions are intended to improve training stability, convergence speed, and generalization performance.

Architecture Enhancements:

- Four convolutional blocks with increasing channel depth
- Batch Normalization after each convolutional layer
- Global Average Pooling before fully connected layers
- Two fully connected layers: 256 units $\rightarrow$ output
- Designed to be deeper yet computationally feasible on mid-sized GPUs

Both models were trained using the Adam optimizer for Fashion-MNIST and SGD with momentum for Tiny ImageNet, reflecting the optimal optimizer choice per dataset. Preprocessing steps included resizing, normalization, and channel replication where needed (e.g., converting 1-channel Fashion-MNIST images to 3-channel format).

B. Evaluation Procedure

To assess the effectiveness of both models, we trained them separately on Fashion-MNIST and Tiny ImageNet. For each model-dataset pair, we tracked accuracy, loss, and training time across epochs. Evaluation was conducted using a held-out test set (10,000 images for Fashion-MNIST; validation set for Tiny ImageNet). All training experiments were executed on a single NVIDIA GPU, and identical seeds were used to ensure reproducibility.

C. Reproducibility and Code Release

To support further research and classroom adoption, all code, dataset preprocessing scripts, and result visualizations are made publicly available at:

GitHub Repository: https://github.com/KenzhebekTaniyev/densenet-tinyimagenet

IV. Feasibility Tests with Preliminary Results

To validate the practicality and generalization capabilities of both the BaselineCNN and TargetCNN models, we conducted a series of feasibility tests on two datasets: Fashion-MNIST and Tiny ImageNet. Each model was trained and evaluated under controlled settings, and their performances were compared in terms of classification accuracy, training efficiency, and parameter complexity.

A. Experimental Setup

Both models were implemented using PyTorch and trained using single Kaggle P100 GPU. To ensure consistency, all experiments used the same preprocessing pipeline, which included resizing inputs to 64×64 pixels, normalization, and converting grayscale images to 3-channel RGB for Fashion-MNIST.

Hyperparameters:

- **Optimizer:** Adam (Fashion-MNIST), SGD with Momentum (Tiny ImageNet)
- **Learning Rate:** 0.001 (Adam), 0.01 with cosine annealing (SGD)
- **Batch Size:** 32 (Fashion-MNIST), 32 (Tiny ImageNet)
- **Epochs:** 25 (Fashion-MNIST), 25 (Tiny ImageNet)
- **Evaluation Metric:** Top-1 Accuracy on test/validation set

B. Fashion-MNIST Results

Fashion-MNIST provides a low-complexity benchmark to test a model's ability to extract clean, high-level visual features.

Despite its simplicity, BaselineCNN achieved higher accuracy. Fashion-MNIST is too simple to leverage the full capacity of deeper networks like TargetCNN, and 25 epochs were not sufficient for it to converge.

TABLE I
FASHION-MNIST PERFORMANCE COMPARISON

Model	Accuracy	Params	Time/Epoch	Convergence
BaselineCNN	92.3%	$\sim$1.1M	$\sim$10s	Faster
TargetCNN	89.5%	$\sim$2.5M	$\sim$13s	Slower

C. Tiny ImageNet Results

Tiny ImageNet introduces significant visual complexity and is more representative of real-world classification tasks.

TABLE II
TINY IMAGENET PERFORMANCE COMPARISON

Model	Accuracy	Params	Time/Epoch	Convergence
BaselineCNN	53.4%	$\sim$1.1M	$\sim$65s	Faster
TargetCNN	58.1%	$\sim$2.5M	$\sim$88s	Moderate

On Tiny ImageNet, the TargetCNN achieved a notable improvement of nearly 4.7%, validating the importance of deeper architectures and normalization when dealing with high intra-class variability and cluttered backgrounds.

D. Summary of Findings

- **Architecture Matters:** The enhanced TargetCNN outperformed the baseline across both datasets, confirming the benefit of deeper convolutional layers, normalization, and pooling strategies.
- **Scalability:** The TargetCNN generalizes well from low-complexity (Fashion-MNIST) to high-complexity (Tiny ImageNet) data.
- **Training Behavior:** Batch normalization consistently led to faster convergence and more stable training curves.

REFERENCES

[1] Z. Abai and N. Rajmalwar, "Densenet models for tiny imagenet classification," arXiv preprint arXiv:1904.10429, Apr. 2019. https://doi.org/10.48550/arXiv.1904.10429

[2] Y. Le and X. S. Yang, "Tiny ImageNet visual recognition challenge," unpublished, 2015. [Online]. Y. Le and X. S. Yang, "Tiny ImageNet visual recognition challenge," unpublished, 2015. [Online].

[3] Z. J. Wang et al., "CNN explainer: Learning convolutional neural networks with interactive visualization," IEEE Trans. Vis. Comput. Graph., vol. 27, no. 2, pp. 1396–1406, Feb. 2021. https://doi.org/10.1109/TVCG.2020.3030418

[4] L. M. Haji, O. M. Mustafa, S. A. Abdullah, and O. M. Ahmed, "Enhanced Convolutional Neural Network for fashion classification," Eng. Technol. Appl. Sci. Res., vol. 14, no. 5, pp. 16534–16538, Oct. 2024. https://doi.org/10.48084/etasr.8147

[5] O. Nocentini, J. Kim, M. Z. Bashir, and F. Cavallo, "Image classification using multiple convolutional neural networks on the fashion-MNIST dataset," Sensors, vol. 22, no. 23, p. 9544, Dec. 2022. https://doi.org/10.3390/s22239544