Comparison of Pre-trained CNN Models on CIFAR-100 (20 Classes)

May 24, 2025

1 Introduction

The performance of 10 pre-trained Convolutional Neural Networks (CNNs)—ResNet50, VGG16, VGG19, MobileNetV2, DenseNet121, EfficientNetB0, InceptionV3, Xception, ResNet101, and ResNet152—fine-tuned on a subset of the CIFAR-100 dataset (20 classes, 4,000 training images, 1,000 test images). The evaluation was conducted on Google Colab with a Tesla T4 GPU (16 GB VRAM), using mixed precision training to optimize memory and runtime.

2 Results

Summary of the performance and characteristics of the 10 models.

Table 1: Comparison of Pre-trained CNN Models

Model	Accuracy	Loss	Param(M)	Size (MB)	VRAM (GB)	Inference
ResNet50	0.7780	0.6834	25.6	98	4–5	100–120
VGG16	0.6560	1.0900	138.4	528	5–6	120–140
VGG19	0.6990	0.9597	143.7	548	5–6	130–150
MobileNetV2	0.6790	0.9604	3.5	14	2–3	60–80
DenseNet121	0.7510	0.7679	8.0	33	3–4	90–110
EfficientNetB0	0.7760	0.6801	5.3	29	2–3	70–90
InceptionV3	0.7130	0.9125	23.9	92	4–5	110–130
Xception	0.7480	0.8021	22.9	88	4–5	110–130
ResNet101	0.8110	0.5804	44.7	171	5–6	120–140
ResNet152	0.8160	0.5843	60.4	232	6–7	140–160

2.1 Test Accuracy

ResNet152 (0.8160) and ResNet101 (0.8110) achieved the highest accuracy, benefiting from deep architectures with residual connections. ResNet50 (0.7780) and EfficientNetB0 (0.7760) followed closely. VGG16 (0.6560) and MobileNetV2

(0.6790) performed worst, likely due to simpler designs (VGG16) or insufficient fine-tuning (MobileNetV2). DenseNet121 (0.7510), Xception (0.7480), and InceptionV3 (0.7130) showed moderate performance.

2.2 Hardware Requirements

MobileNetV2 and EfficientNetB0 require the least VRAM (2–3 GB), making them ideal for Colab's T4 GPU (16 GB). VGG16/19 and ResNet152 demand 5–7 GB, posing risks on lower-end GPUs. System RAM usage was 1–2 GB, well within Colab's 12–15 GB. Mixed precision training ensured memory efficiency, preventing crashes observed in prior runs with 12 GB RAM.

2.3 Model Size

MobileNetV2 (3.5M parameters, 14 MB) and EfficientNetB0 (5.3M, 29 MB) are the smallest, suitable for resource-constrained deployment. VGG16/19 (138–143M, 528–548 MB) and ResNet152 (60.4M, 232 MB) are significantly larger, increasing storage and loading overhead.

2.4 Inference Time

MobileNetV2 (60–80 ms/batch) and EfficientNetB0 (70–90 ms/batch) are the fastest, benefiting from lightweight architectures. ResNet152 (140–160 ms/batch) and VGG19 (130–150 ms/batch) are the slowest due to high parameter counts and depth.

2.5 Architecture

- **ResNet50/101/152**: Residual connections mitigate vanishing gradients, enabling deep networks. Bottleneck blocks reduce computation, with deeper variants (ResNet152) excelling in accuracy.
- **VGG16/19**: Simple 3x3 convolutions, high parameter count, no skip connections, leading to poor performance on CIFAR-100.
- **MobileNetV2**: Depthwise separable convolutions and inverted residuals for lightweight design, but underperformed in this task.
- **DenseNet121**: Dense connectivity reuses features, balancing efficiency and performance.
- **EfficientNetB0**: Compound scaling optimizes depth, width, and resolution, but slightly underperformed expectations.
- **InceptionV3/Xception**: Multi-scale feature extraction (Inception modules) and depthwise separable convolutions (Xception) offer moderate performance.