PyTorch Capstone Project

Cat Breed Classification System

Wiwat Pholsomboon

April 8, 2025

Deep Learning with Pytorch - INFO-6147-(01)-25W

Introduction

- Study on using PyTorch models to classify cat breeds
- Implementation of Grad-CAM for visual explanation
- Integration with ChatGPT-4o-mini for image identification
- Interactive quiz system to compare model accuracy with human performance
- Github: https://github.com/innozent/pytorch_capstone

Dataset

- **Source:** Kaggle 'Geno Cat Breed Image Collection'
- **Size:** 15 cat breeds with 375 photos each (5,625 total)
- **Preprocessing pipeline:**
  - Resize to 256×256 pixels
  - Random crop to 224×224 pixels
  - Random horizontal flip for augmentation
  - Random rotation for variety
  - Conversion to PyTorch tensors
  - Normalization using ImageNet statistics

Model Architecture

Custom CNN:

- 5 Convolutional layers
- 2 Fully connected layers
- Dropout for regularization
- ReLU activation functions

Transfer Learning Models:

- ResNet18
- EfficientNetB2
- VGG16
- All pre-trained with ImageNet weights

Model Comparison

Transfer Learning Models Comparison

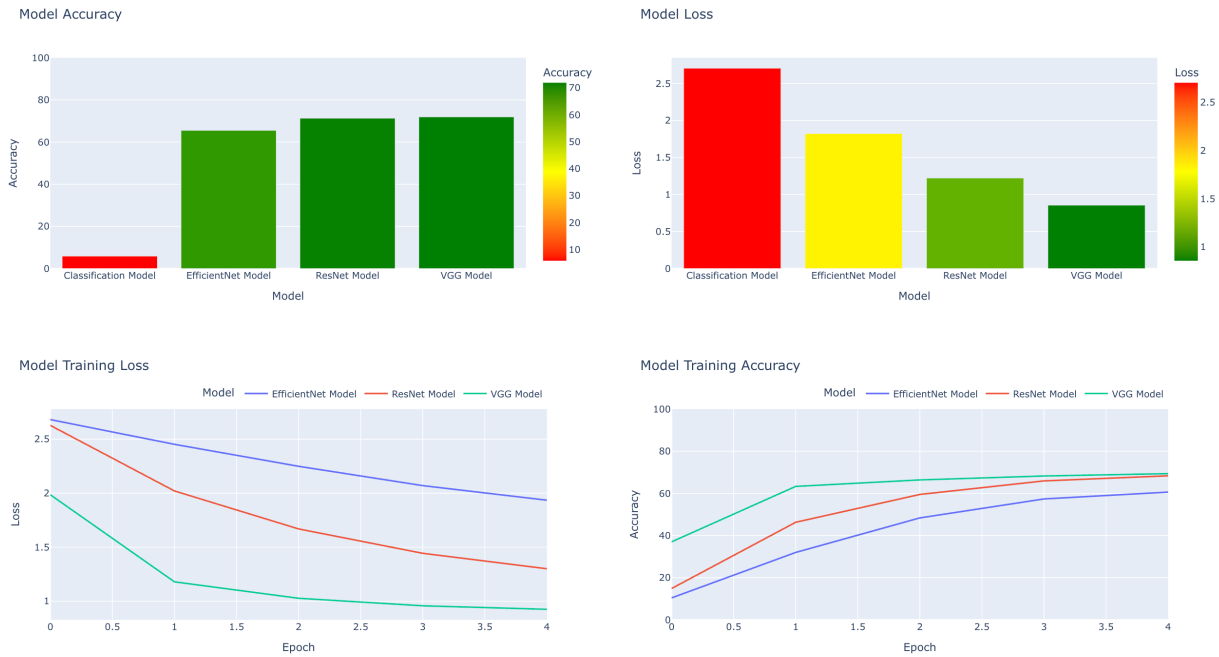
VGG16	ResNet18	EfficientNetB2
Parameters: 138M	**Parameters:** 11.7M	**Parameters:** 9.2M
Layers: 16	**Layers:** 18	**Layers:** 82
Key Features: - Simple architecture - 3x3 convolutions - Max pooling - Dense layers	**Key Features:** - Residual blocks - Skip connections - Batch normalization - Global average pooling	**Key Features:** - Compound scaling - MBConv blocks - Swish activation - Squeeze-and-excitation

Training Methodology

Consistent hyperparameters across all models:

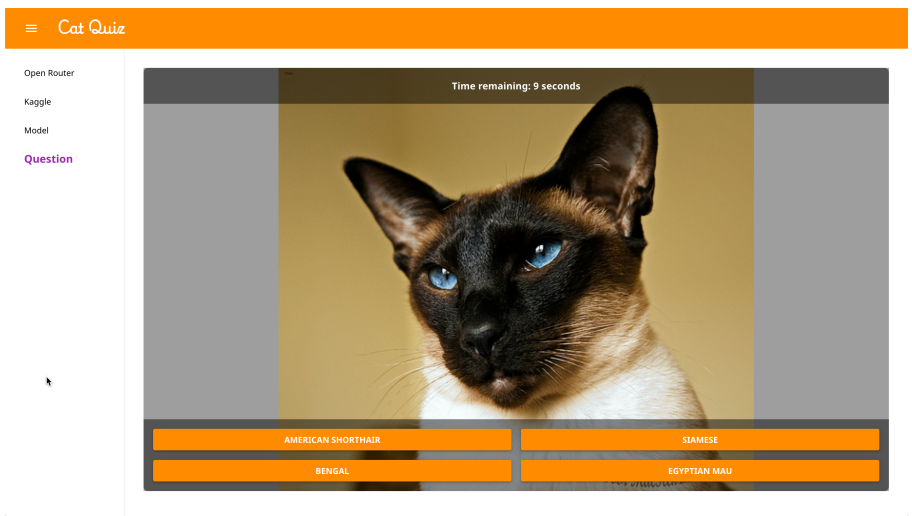
- Batch size: 128
- Learning rate: 0.001
- Momentum: 0.9
- Epochs: 5
- Loss function: Cross Entropy
- Optimizer: SGD

Results and Evaluation



- Comparison of model accuracy and loss
- Interactive quiz implementation for human vs AI performance analysis

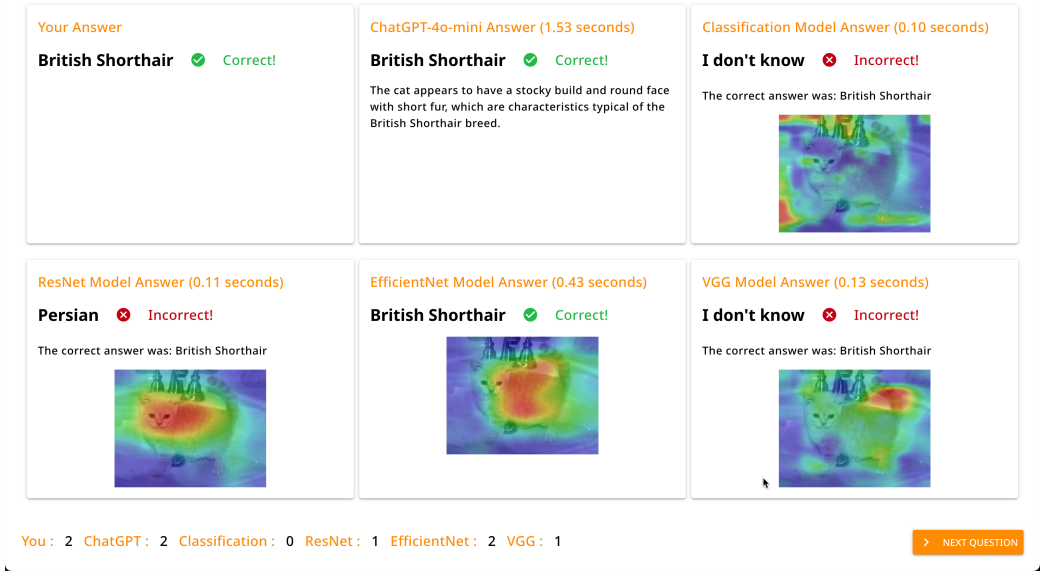
Visualization and Interpretability



- **Grad-CAM visualization** highlights regions of interest
- **ChatGPT-4o-mini** provides reasoning for predictions
- **Interactive interface** for user engagement

Evaluation Results

- Side-by-side comparison of accuracy:
  - Human participants
  - ChatGPT-4o-mini
  - PyTorch models (CNN, ResNet, EfficientNet, VGG)



Challenges and Solutions

Challenges:

- Different target layers for Grad-CAM across models
- Low accuracy with custom CNN model

Solutions:

- Manual selection of appropriate CNN layers
- Transfer learning to leverage pre-trained weights

Future Improvements

Potential Enhancements

- User management system to track quiz performance over time
- Enhanced data augmentation techniques for better generalization
- Additional model architectures for comparison
- More comprehensive evaluation metrics beyond accuracy
- Fine-tuning of transfer learning models

Conclusion

- **ChatGPT-4o-mini** demonstrates superior accuracy due to advanced image understanding capabilities
- **Transfer learning models** perform moderately well with limited training
- **Custom CNN model** shows potential but requires more extensive training
- **Visualization tools** provide valuable insights into model decision-making

Thank you!