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 (ImageNet)
- EfficientNetB2 (ImageNet)
- VGG16 (ImageNet)

Model Comparison

Transfer Learning Models Comparison

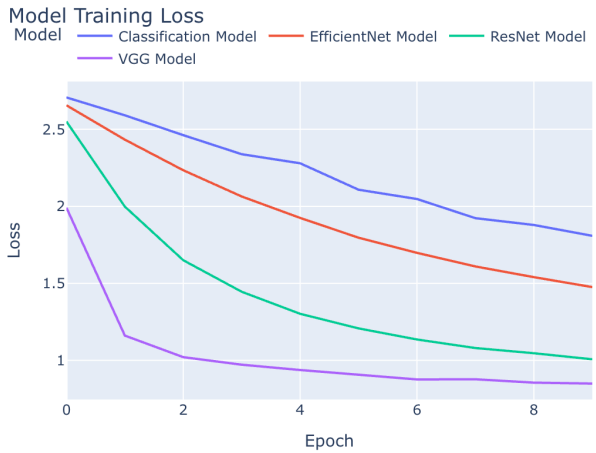
VGG16	ResNet18	EfficientNetB2
Parameters: 138.3M	**Parameters:** 11.7M	**Parameters:** 9.1M
Layers: 13 Conv, 3 FC	**Layers:** 17 Conv, 1 FC	**Layers:** 35 Conv, 1 FC
Key Features:	**Key Features:**	**Key Features:**
- Simple architecture	- Residual blocks	- Compound scaling
- 3x3 convolutions	- Skip connections	- MBConv blocks
- Max pooling	- Batch normalization	- Swish activation
- Dense layers	- Global average pooling	- Squeeze-and-excitation

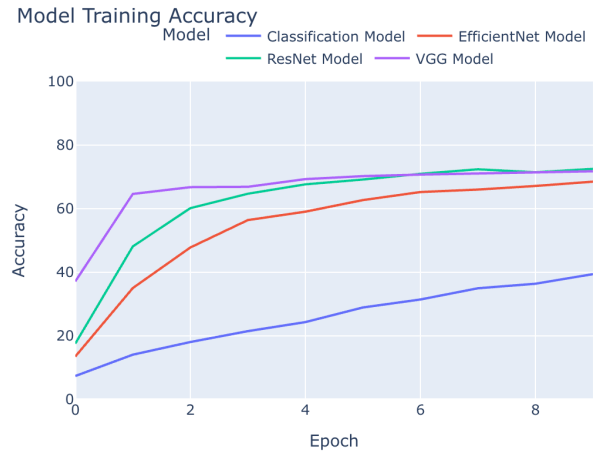
Training Methodology

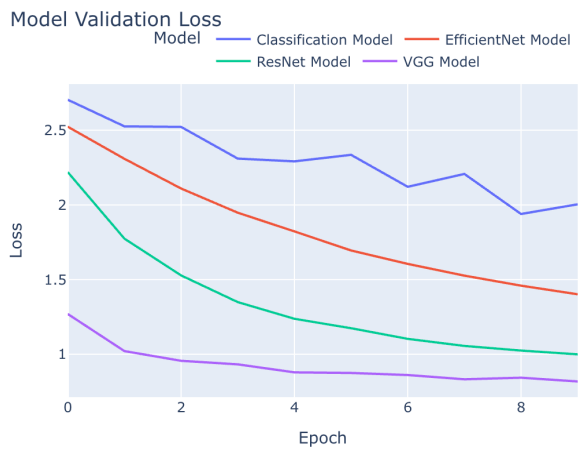
hyperparameters for all models:

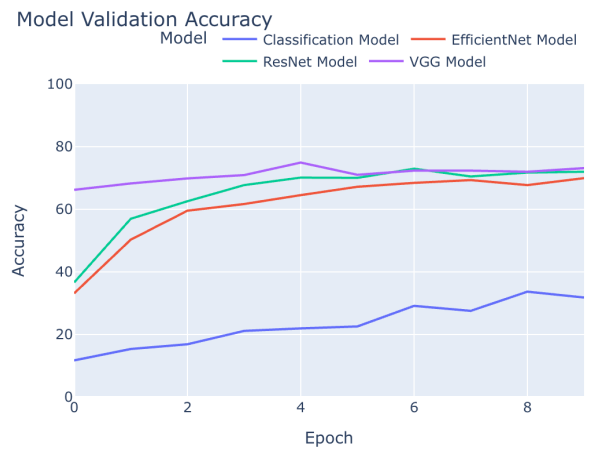
- Batch size: 128
- Momentum: 0.9
- Learning rate:
  - 0.01 (Custom CNN)
  - 0.001 (Other models)
- Loss function: Cross Entropy
- Optimizer: SGD
- Epochs:
  - 20 (Custom CNN)
  - 10 (Other models)

Results and Evaluation

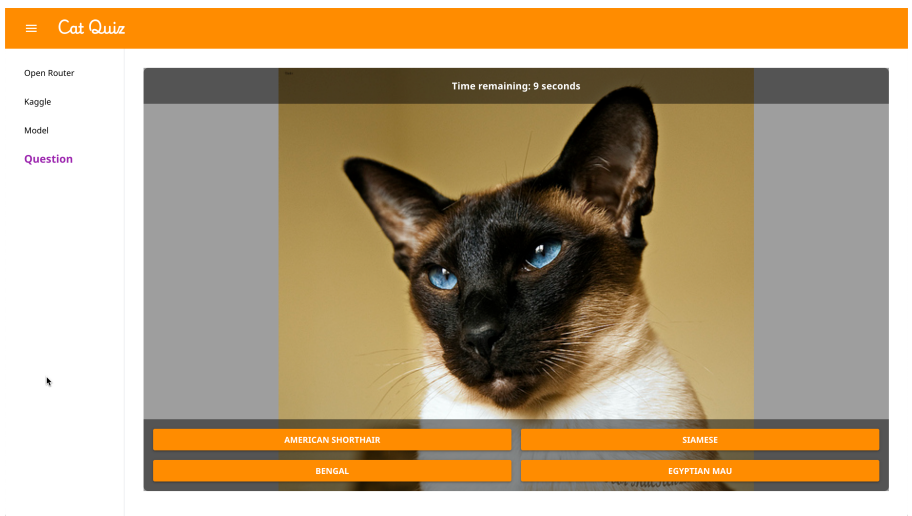








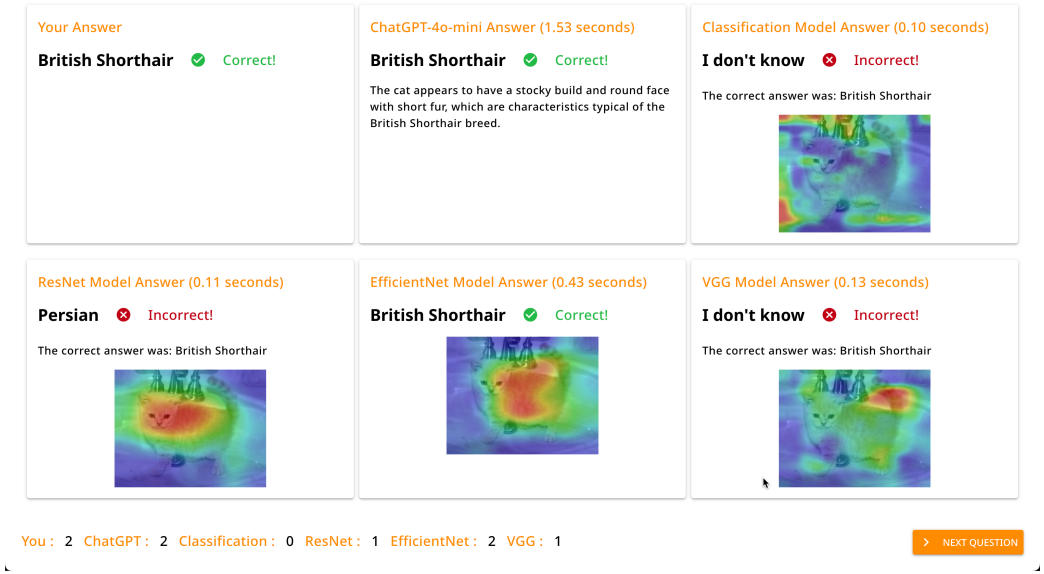
Visualization and Interpretability



- **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

Conclusion

Model	Parameters	Accuracy	Loss	Training Time (s)
Custom CNN Model	31.9M	42%	1.86	2,698
EfficientNet Model	9.1M	71%	1.40	1,256
ResNet Model	11.7M	73%	0.99	2,357
VGG Model	138.3M	76%	0.81	1,314

- **VGG16 Model** is the best model based on confusion matrix and accuracy score.
- **ResNet18 Model** the second best model, on the other hand, has a smaller number of parameters (11.7 Million parameters)
- **EfficientNet B2** has 9.1 Million parameters and slightly lower accuracy compared to VGG16 and ResNet18 on this dataset.
- **ChatGPT-4o-mini** which is a multi-modal model, has image recognition capabilities and has very good accuracy rate.