

# PyTorch Capstone Project Report

## Cat Breed Classification System

Wiwat Pholsomboon

April 8, 2025

## 1 Introduction

---

This project is to study on using PyTorch model to classify cat images. It implements Grad-CAM to visualized the attention of the model on one of CNN layers. ChatGPT-4o-mini is used to identify the image and given the rationale of result. Quiz is implemented to show the accuracy of each model along with human accuracy.

## 2 Dataset

---

- Dataset is from Kaggle from 'Geno Cat Breed Image Collection' dataset (<https://www.kaggle.com/datasets/shawngano/gano-cat-breed-image-collection>)
- Contains 15 cat breeds with 375 photos for each breed (total 5,625 photos)
- Preprocessing step
  - Resize to 256x256
  - Random Crop to 224x224
  - Random Horizontal Flip
  - Random Rotation
  - Convert to Tensor
  - Normalize with mean and std of ImageNet

## 3 Model Architecture

---

- Architecture of implemented models:
  - 5 Convolutional Layers (Conv2d)
  - 2 Fully Connected Layers (Linear)
  - 1 Dropout Layer (Dropout)
  - ReLU Activation Function
- Transfer Learning Models:
  - ResNet18 Model (Weights: ImageNet)
  - EfficientNetB2 Model (Weights: ImageNet)
  - VGG16 Model (Weights: ImageNet)
- GPT Model:
  - ChatGPT-4o-mini implemented from OpenRouter API
  - Evaluate the model accuracy use quiz.

## 4 Training Methodology

- All models are trained with the same hyperparameters for fair comparison.
  - Training Batch Size: 128
  - Learning Rate: 0.001
  - Momentum: 0.9
  - Epochs: 5
  - Loss Function: Cross Entropy Loss
  - Optimizer: SGD

## 5 Results and Evaluation

- Compare Training Accuracy and Loss of each model.
- Quiz is implemented to show the accuracy of each model along with human accuracy and ChatGPT-4o-mini accuracy.

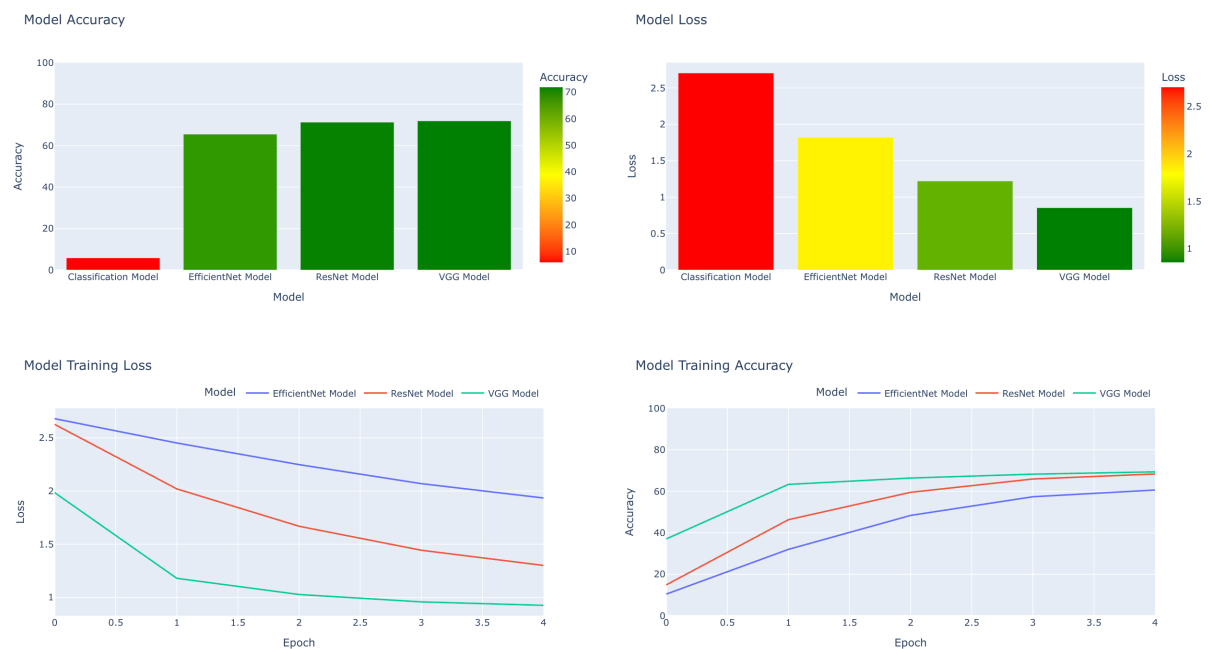


Figure 1: Comparison of model accuracy and loss

## 6 Visualization and Interpretability

- Implemented Grad-CAM to compare and visualize each model's attention on the image
- ChatGPT-4o-mini will give the rationale of its answer.

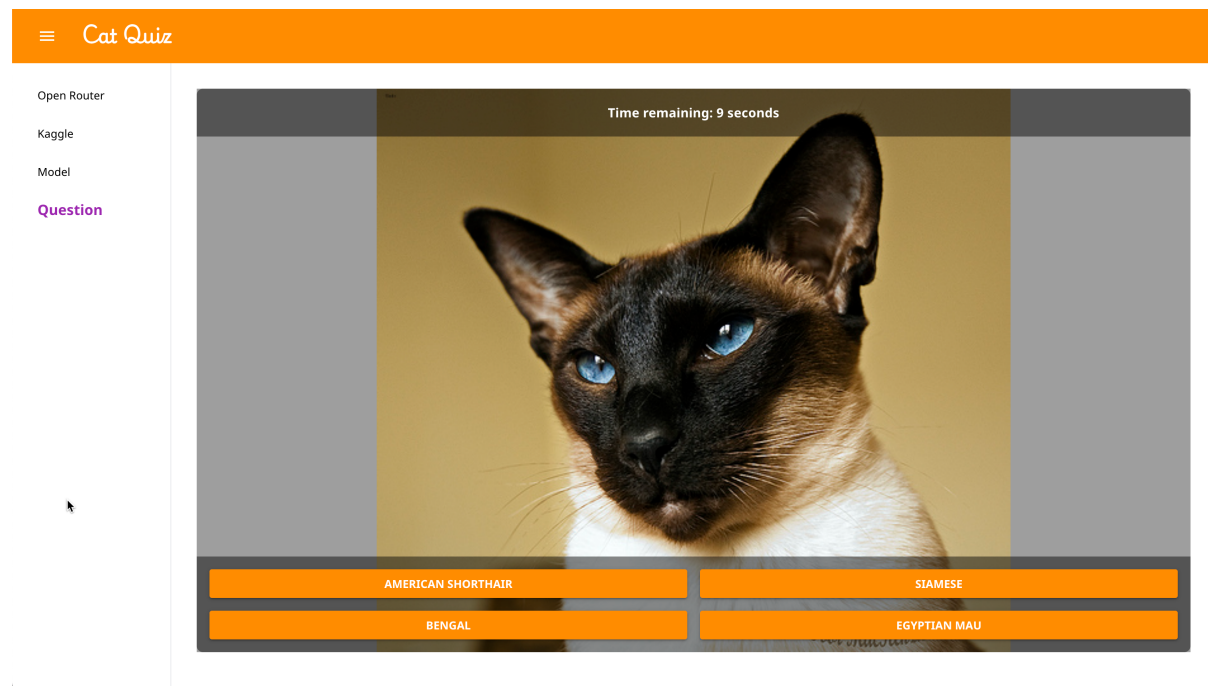


Figure 2: Quiz interface allow user to choose the answers

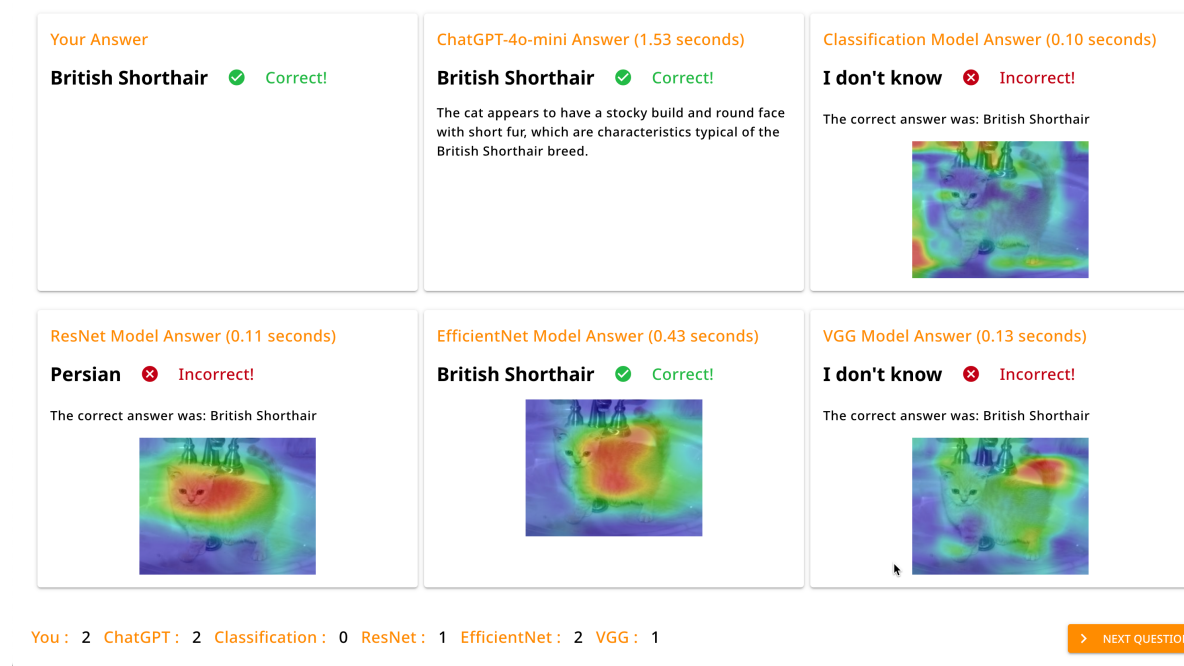


Figure 3: Evaluation results comparing human, ChatGPT-4o-mini and PyTorch Models accuracy

## 7 Challenges and Solutions

- Grad-CAM target layer for each model architecture is different and need to be selected manually.
- Solution is to print out model architecture and select the CNN layer at very last stage of the model.

---

## 8 Future Improvements

---

- Add users management to save the quiz result and compare with the previous quiz result.
- Add more data augmentation techniques.
- Add more models for comparison.
- Add more evaluation metrics.

## 9 Conclusion

---

- ChatGPT-4o-mini is able to classify the image with high accuracy due to its ability to understand the image and choices of quiz provided.
- PyTorch models from Transfer Learning are able to classify the image with moderate accuracy.
- CNN models are not able to classify the image with high accuracy due to very limited training epochs and architecture.