APS360 PROJECT PROPOSAL - POKÉMON RECOGNITION

Teh Wei Sheng
1012872090
ws.teh@mail.utoronto.ca
GitHub Repository

Introduction

The primary motivation for this project lies in the global popularity of Pokémon. According to the National Pokédex [1], there are 1,025 distinct Pokémon species as of 2025, each with its own unique design, abilities, and description. Developing a deep learning system capable of recognizing these Pokémon from images is both an engaging and meaningful task, as it could serve as the foundation for applications such as automated Pokédex tools or educational games. Deep learning is a natural choice for this problem because Pokémon appear in diverse visual styles (e.g., sprites, anime artwork, and 3D renders). Convolutional Neural Networks (CNNs) are well suited to capture the complex shapes, colors, and textures that distinguish Pokémon species, making them an appropriate approach for this recognition task.

Illustration

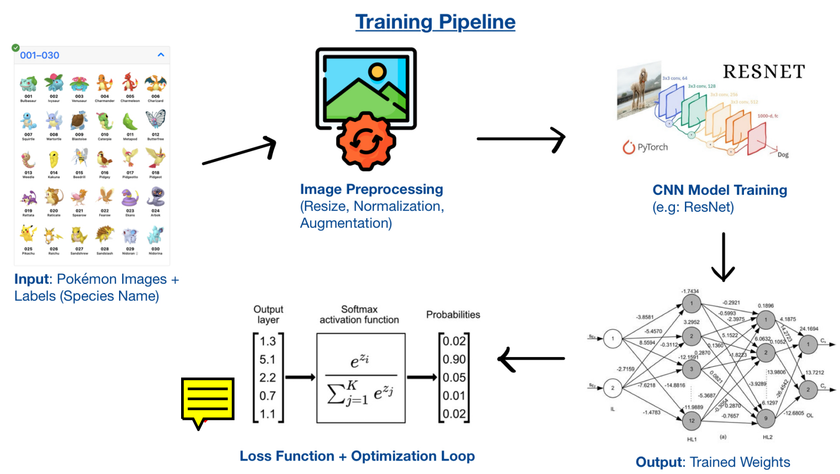

Figure 1: Training Pipeline of Pokémon Recognition Model

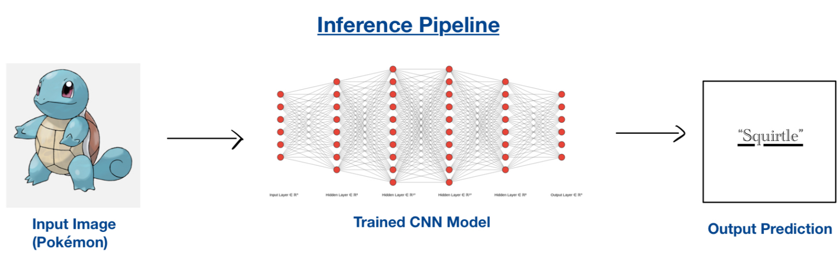

Figure 2: Inference Pipeline of Pokémon Recognition Model

Background & Related Work

- ImageNet [2]: large-scale dataset that established the foundation for training modern CNN architectures.
- Stanford Dogs Dataset [6]: example of fine-grained classification task involving visually similar classes.
- Publicly available Pokémon datasets (e.g., Kaggle "7,000 Labeled Pokémon") [9]: Contains 150 Pokémon image folders, with 25-50 images for each Pokémon.
- Transfer learning with pretrained networks such as ResNet [3] and VGG [10]: effective for adapting deep models to smaller, domain-specific datasets.
- CIFAR-100 [8]: a benchmark dataset containing 100 object categories, conceptually similar to Pokémon recognition where the goal is to distinguish between many visually distinct classes.

Data Processing

For this project, I will use the 7,000 Labeled Pokémon dataset from Kaggle [9], which contains images of 150 Pokémon species organized into folders by class. The dataset will be split into training (70%), validation (15%), and test (15%) sets to enable proper model evaluation. All images will be resized to a fixed resolution, normalized to standard pixel ranges, and augmented (e.g., random rotations, flips, and color jitter) to improve model robustness. If necessary, images will be cropped or padded to ensure consistent dimensions across the dataset. These preprocessing steps will help reduce overfitting and make the dataset suitable for training CNN models.

Architecture

The recognition model will be based on Convolutional Neural Networks (CNNs), as they are highly effective for image classification tasks. Specifically, I plan to use transfer learning with a pretrained ResNet-18 architecture [3], fine-tuned on the Pokémon dataset. The input images will be resized to a consistent resolution (e.g., 128×128 or 224×224 pixels) before being passed into the model. The final fully connected layer of ResNet will be replaced with a softmax output layer matching the number of Pokémon classes in the dataset. Training will be performed using cross-entropy loss and optimized with Adam [7]. Dropout [11] and data augmentation will be used to reduce overfitting.

Baseline Model

As a baseline, I will implement a simple logistic regression classifier [4] trained on raw pixel values and color histograms of the images. This approach does not leverage spatial information, but it provides a point of comparison to evaluate the effectiveness of CNNs. Alternatively, I may also compare against a k-Nearest Neighbors (k-NN) classifier on flattened image features. Both baseline models are easy to reproduce and highlight the advantages of deep learning in capturing complex visual patterns.

Ethical Considerations

There are several ethical considerations for this project. First, the dataset consists of copyrighted Pokémon images, which limits the use of the trained model to educational and research purposes only [9]. Second, class imbalance may occur if certain Pokémon have fewer images available than others, potentially biasing the model toward majority classes [5]. Third, the visual diversity of Pokémon (e.g., different art styles such as sprites, anime, or 3D renders) could cause the model to perform inconsistently across styles, limiting fairness and generalization [12]. These factors will be addressed by careful dataset preprocessing, augmentation, and transparent reporting of model limitations.

References

[1] Pokédex — the official pokémon website. `https://www.pokemon.com/us/pokedex`. Accessed: 2025-09-25.

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[4] David W Hosmer, Stanley Lemeshow, and Rodney X Sturdivant. *Applied Logistic Regression*, volume 398. John Wiley & Sons, 2013.

[5] Justin M Johnson and Taghi M Khoshgoftaar. A survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):1–54, 2019.

[6] Aditya Khosla, Nityananda Jayadevaprakash, Bangpeng Yao, and Li Fei-Fei. Novel dataset for fine-grained image categorization: Stanford dogs. In *IEEE CVPR Workshop on Fine-Grained Visual Categorization*, 2011.

[7] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[8] Alex Krizhevsky. Cifar-100 dataset. Technical Report TR-2009, University of Toronto, 2009.

[9] Lantian773030. 7,000 labeled pokémon dataset. `https://www.kaggle.com/datasets/lantian773030/pokemonclassification`. Accessed: 2025-09-25.

[10] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*, 2015.

[11] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[12] Antonio Torralba and Alexei A Efros. Unbiased look at dataset bias. In *CVPR 2011*, pages 1521–1528. IEEE, 2011.