Individual Final Project Report

Dog Breed Recognition & Image Generation System

Junhua Deng

1. Introduction

This report outlines the individual contribution to Group 6's Deep Learning final project, which centered on building an end-to-end system for dog-breed classification and image generation. The project integrates a convolutional image classifier with a generative AI module, resulting in a unified application capable of both accurate breed identification and flexible artistic image synthesis.

1.1 Project Overview

The workflow consists of three integrated components designed to support dog-breed understanding and style-controlled image generation.

1) **Supervised Training of a Dog-Breed Classifier**
   A ResNet-50 model is trained on the original dog-breed dataset. After completing supervised training, the resulting .pth checkpoint is uploaded to the Hugging Face Hub to ensure reproducibility, model sharing, and seamless downstream integration.
2) **Style-Controlled Image Generation Using SDXL and LoRA**
   Stable Diffusion XL (SDXL) is then used to evaluate different LoRA adapters and prompt configurations for generating dog images in specific artistic or photographic styles. The chosen LoRA models and prompt templates are consolidated into a unified generator module.
3) **Deployment of an Integrated Streamlit Application**
   A Streamlit application was developed to integrate both the classifier and the generation modules within a unified interface.

1.2 Shared Work Outline

The project tasks were distributed among team members. My primary contributions included training the ResNet-50 model used in the final classifier, designing the generator framework integrated into the Streamlit application, and drafting the group report. Other team members focused on developing additional style modules for the generator and assisting with the Streamlit deployment. They also contributed to the classifier, and together we selected the model that demonstrated the best performance on external test images sourced from the internet.

2. Description of Individual Work

My primary contribution to this project involved developing the core machine learning components, including training the ResNet-50 model for dog-breed classification, constructing the inference-ready classifier, and building the manga-style image generation module. This section outlines the technical foundations, model design choices, and implementation details behind these components.

2.1 ResNet-50 Architecture and Training Framework

ResNet-50 was chosen for its stable residual architecture, which enables effective training of deep convolutional networks. The model is initialized with pretrained ImageNet weights and adapted to a 120-class dog-breed classification task. Training uses a standard cross-entropy objective with AdamW optimization and cosine-annealing learning-rate scheduling. To enhance generalization, the training pipeline includes resized cropping, horizontal flipping, and normalization, while the validation pipeline applies fixed preprocessing for consistent evaluation.

2.2 Classifier Construction for Inference

After training, a deployment-ready classifier is constructed by restoring the ResNet-50 architecture and loading the best checkpoint. The classifier applies the same validation preprocessing used during training to ensure consistency and reliability. It outputs the predicted breed label and confidence score and serves as the core model integrated into downstream applications.

2.3 Manga-Style Generator Framework

To complement the classifier, a manga-style generator is implemented using a Stable Diffusion–based pipeline. The generator accepts either the classifier's predicted breed or a user-selected breed and produces stylized dog images using breed-aware prompts. A specialized manga-style LoRA or style configuration is applied to achieve consistent visual aesthetics.

3. Detailed Work Description

3.1 Resnet-50 Training

Data Processing

My pipeline begins with standardized data preprocessing, including resized scaling, random resized cropping to 224×224, horizontal flipping, and ImageNet normalization. A deterministic

resize–center-crop transformation is applied to validation images to ensure consistent and unbiased evaluation.

Model Configuration

My model is constructed by initializing a ResNet-50 backbone with ImageNet pretrained weights and replacing the final fully connected layer with a 120-class output head. This configuration enables transfer learning while adapting the network to the dog-breed classification task.

Training Procedure

Training is conducted for 15 epochs using the AdamW optimizer with a learning rate of 1×10^{-4} and weight decay of 1×10^{-4}. A cosine-annealing learning-rate schedule progressively reduces the step size, promoting smoother and more stable convergence.

Checkpointing

Validation accuracy is monitored at every epoch, and an automatic checkpointing mechanism stores the best-performing model. This strategy ensures that the final classifier reflects the optimal training epoch rather than the final iteration.

Evaluation Metrics

After training, the best checkpoint is reloaded for comprehensive evaluation, including accuracy, macro/micro/weighted precision and recall, F1-scores, and a full per-class classification report with confusion matrix. All metrics are exported in structured form to support reproducibility and downstream analysis.

3.2 Classifier

Inference Module

The classifier is designed as a lightweight and reliable inference module built on top of the trained ResNet-50 model. It rebuilds the same architecture used during training, loads the saved checkpoint from the Hugging Face Hub, and imports the corresponding label mapping so that predicted class indices can be converted back into readable breed names.

Preprocessing Consistency

To make sure predictions remain consistent with training behavior, the classifier applies the same validation-style preprocessing: resizing, center cropping, converting images to tensors, and applying ImageNet normalization.

Prediction Logic

During inference, each image is processed and passed through the network inside a no-gradient context to keep computation efficient. The classifier outputs both the predicted breed and its confidence score.

3.3 Manga Style Generator

Model Setup

The generator is built on top of the Stable Diffusion XL (SDXL) base model, which provides a high-resolution latent diffusion backbone suitable for stylized image synthesis. Upon initialization, the module loads the SDXL base weights and prepares the pipeline on either GPU or CPU depending on availability.

LoRA Integration

To achieve a consistent manga aesthetic, the generator incorporates a LoRA module sourced directly from the Hugging Face Hub. The script supports automatic or explicit selection of LoRA weight files and applies a configurable scaling factor to control stylistic intensity.

Prompt Construction

The generator relies on detailed text prompts that blend anatomical accuracy with manga stylistic cues. Breed names are inserted directly into the prompts, enabling breed-specific conditioning. The negative prompt includes a curated list of common diffusion failure modes—such as extra limbs, duplicated tails, distorted anatomy, or inconsistent proportions—to reduce artifact frequency and improve output quality.

4. Results

4.1 Resnet-50 Model and Classifier

Evaluation

	Macro	Micro	Weighted
Precision	0.8508	0.8508	0.8601

Recall	0.8461	0.8508	0.8508
F1-score	0.8435	0.8508	0.8508
Accuracy			0.8508

The ResNet-50 classifier demonstrates strong overall performance across all evaluation metrics. Macro-averaged precision, recall, and F1-score are approximately 0.85, indicating that the model performs consistently across classes even when treating each breed equally, regardless of frequency. Micro metrics are identical to the overall accuracy (0.8508), showing that the model maintains stable performance when weighting predictions by sample count. The weighted precision is slightly higher (0.8601), suggesting that the model performs particularly well on more common breeds. Overall, the results show that ResNet-50 achieves balanced and reliable classification performance across the 120-dog-breed dataset.

Comparison:

My model prediction:
predicted breed: vizsla
Confidence: 0.9691827297210693

Teammate 1 model prediction:
Predicted breed: vizsla
Confidence: 0.46891164779663086

Teammate 2 model prediction:
Predicted breed: weimaraner
Confidence: 0.6107911467552185

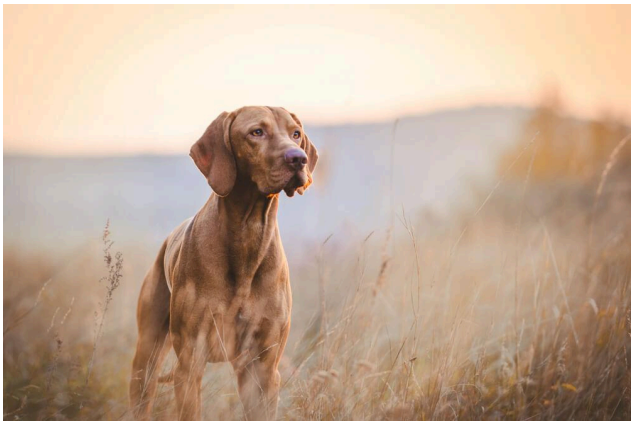

Fig. 1: Vizsla

4.2 Manga Style Generator

My SDXL + LoRA framework successfully produced high-quality manga-style dog images characterized by clean black-and-white line art, stable contours, and consistent anatomical structure. The LoRA module reinforced manga-specific visual elements such as screentone shading, dynamic motion lines, and high-contrast inked edges, allowing the generator to maintain both stylistic coherence and breed recognizability.

Prompt:
"Black and white side view of a **Shiba Inu** dog sprinting, accurate canine anatomy, single visible tail, one tail only, proper proportions, full body in frame, natural limb spacing, dynamic running pose, consistent perspective, shonen jump manga style, screentone shading, inked lineart, high contrast, speed lines, impact frame, dramatic action"

Output:

Fig.2 : Manga Output

5. Conclusions

5.1 Summary

Together, the ResNet-50 classifier, the inference module, and the manga-style generator form a cohesive and reliable end-to-end system for dog-breed understanding and stylized image synthesis. The ResNet-50 model provides a stable backbone for breed recognition, achieving strong performance through careful data processing, optimized training procedures, and comprehensive evaluation. The classifier module builds on this foundation by offering a consistent and deployment-ready interface that ensures accurate, reproducible predictions across diverse input conditions.

5.2 Improvements

Resnet: Training quality could be enhanced through stronger augmentation, balanced sampling strategies, or experimenting with more advanced architectures such as ConvNeXt or Vision Transformers to improve classification accuracy and robustness.

Generator: The manga-style generator may benefit from breed-specific LoRA training, refined negative prompts, or lightweight post-processing to further reduce anatomical artifacts and strengthen stylistic consistency.

6. References

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (pp. 770–778).

Podell, D., Tov, O., Alabdulmohsin, I., et al. (2023). *SDXL: Improving latent diffusion models for high-resolution image synthesis*. Stability AI Technical Report.

Hu, E. J., Shen, Y., Wallis, P., et al. (2021). *LoRA: Low-rank adaptation of large language models*. arXiv preprint arXiv:2106.09685.