Dog Cardiomegaly Assessment Using a Custom CNN Model

Mohammed Uvais Poozhithara

1. Introduction

Cardiomegaly, or the enlargement of the heart, is a serious clinical condition that affects both humans and animals. Early detection is critical in veterinary care, as late-stage heart enlargement often correlates with reduced life expectancy and quality of life. Thoracic radiographs (chest X-rays) are commonly used for assessing heart size in dogs. However, manual evaluation is prone to variability and subjectivity.

With the advancement of deep learning, automated image analysis offers a powerful solution for consistent and scalable diagnosis. In this project, we aim to design a Convolutional Neural Network (CNN) from scratch using PyTorch to classify dog chest X-ray images into three categories: Small, Normal, and Large heart size. The goal is to evaluate model performance and compare it with the Regressive Vision Transformer (RVT) and VGG16 as reported in prior work [1].

We chose to build a custom CNN without pretraining to better understand the training pipeline and experiment with deep learning components such as normalization, activation functions, regularization, and optimization strategies. Our work is intended to offer a lightweight yet effective alternative to larger models like RVT.

2. Dataset and Preprocessing

The dataset used includes X-ray images of dogs, organized into structured folders:

- **Train:** 3 folders representing Small, Normal, and Large classes

- **Validation:** Same structure as Train
- **Test:** 400 PNG images with filenames like 100.png, 1621.png, etc.

All images were resized to 128×128 pixels. Data augmentation was applied using PyTorch transforms, including random horizontal flip, random rotation, and color jitter. The pixel values were normalized using a mean and standard deviation of 0.5. This ensured consistency across image batches and improved convergence during training.

Table 1: Sample Class Distribution (Train)

Class	Small	Normal	Large
Number of Images	250	300	250

3. Model Architecture and Training

A custom CNN was developed with the following architecture:

- **Convolutional Layers:** Three layers with 32, 64, and 128 filters respectively
- **Normalization:** BatchNorm after each convolution
- **Activation:** ReLU
- **Pooling:** MaxPooling with kernel size 2
- **Classifier:** Two fully connected layers ($512 \to 256 \to 3$) with dropout

Diagram of CNN with 3 conv layers, maxpool, and FC layers (to be inserted)

Figure 1: Planned architecture of the CNN model.

The model was trained for 60 epochs using cross-entropy loss and Adam optimizer (learning rate = 0.00005, weight decay = $1e^{-5}$). The training was done on Google Colab with GPU support.

Plot showing loss curve over 60 epochs (to be inserted)

Figure 2: Training loss curve (placeholder).

Our training process included shuffling, batch loading, and checkpoint saving. Each epoch processed over 700+ samples and took approximately 25–35 seconds on Colab GPU (T4).

4. Evaluation and Results

Evaluation was done on 400 test images using the software provided by the course team. The model predictions were exported to CSV format. The final test accuracy was **71.5%**.

Table 2: Accuracy Comparison

Model	Reported Accuracy
VGG16 (baseline)	75.0%
Custom CNN (ours)	**71.5%**
RVT (paper)	83.4%

Confusion Matrix (conceptual)

	Predicted Small	Normal	Large
Actual Small	72	18	10
Actual Normal	20	78	12
Actual Large	8	15	67

Table 3: Example confusion matrix (simulated).

Our model displayed higher confidence for Normal-sized heart images but occasionally confused Small with Large due to visual similarity in radiographs. This highlights the challenge of feature learning in smaller networks.

5. Conclusion

We built a CNN model to classify heart size in dog X-rays and achieved 71.5% accuracy on the official test set. Although this falls slightly short of the 75% threshold, the results demonstrate the capability of custom CNNs in veterinary radiology applications. Further improvements may include the use of transfer learning, more robust augmentation techniques, or ensemble modeling to enhance accuracy.

This work forms a stepping stone for integrating AI in animal healthcare, where fast and accurate predictions are essential in real-time clinical workflows.

References

[1] Jialu Li and Youshan Zhang. Regressive vision transformer for dog cardiomegaly assessment. *Scientific Reports*, 14(1):1539, 2024.