CNN-Based Detection of Canine Vertebral Heart Size in X-rays

Bharat Kathuria
Yeshiva University
bkathuri@mail.yu.edu

Abstract

This paper presents a comprehensive study on the application of Convolutional Neural Networks (CNNs) for the automated detection of vertebral heart size (VHS) in canine X-ray images. Accurate detection and measurement of VHS are critical for diagnosing various cardiac conditions in dogs. We leveraged a custom CNN architecture, building on the ResNet-50 backbone, to identify key anatomical landmarks within the X-ray images. The dataset used comprises annotated dog heart X-ray images, which were preprocessed and augmented to improve model robustness and generalization. Our model was trained and validated using a combination of traditional data augmentation techniques and advanced preprocessing steps to handle the unique challenges of medical imaging data. The final model demonstrated high precision and recall in detecting key points for VHS calculation, significantly outperforming traditional image analysis methods. The proposed approach not only provides a reliable tool for automated veterinary diagnostics but also sets a foundation for further research in the field of medical image analysis. Future work will focus on extending the model to handle other types of veterinary imaging and exploring the potential of combining CNNs with other deep learning architectures to further enhance detection accuracy.

1. Introduction

Recent advancements in deep learning have significantly impacted the field of medical image classification. Sophisticated neural networks, particularly Convolutional Neural Networks (CNNs), have drastically improved the accuracy of detecting and categorizing medical conditions from imaging data. Despite these advancements, the application of these technologies in veterinary medicine, particularly for diagnosing canine heart conditions, remains relatively unexplored. This study aims to address this gap by focusing on the automated detection and measurement of vertebral heart size (VHS) in dogs using advanced CNN models.

Accurate detection of VHS from X-ray images is crucial for the early detection and treatment of various cardiac conditions in dogs. The VHS is a standardized metric used by veterinarians to assess heart size and identify potential cardiomegaly or other heart diseases. Traditional methods of measuring VHS are manual and subject to variability, necessitating a more reliable and automated approach.

This research leverages a custom CNN architecture based on the ResNet-50 backbone, which has been pretrained on large-scale image datasets. By fine-tuning this architecture and employing extensive data augmentation and preprocessing techniques, we aim to develop a robust model capable of accurately identifying key anatomical landmarks required for VHS calculation. The dataset used in this study comprises a diverse set of annotated dog heart X-ray images, ensuring the model's ability to generalize across various cases.

In this paper, we detail our methodology, including data collection, preprocessing, model architecture, training procedures, and evaluation metrics. We also present the results of our experiments, highlighting the model's performance in terms of precision, recall, and overall accuracy. The proposed approach not only demonstrates the potential of CNNs in veterinary diagnostics but also sets the stage for future research in this field.

2. Related Work

The application of deep learning techniques in medical image analysis has seen remarkable progress over the past decade. Convolutional Neural Networks (CNNs) have demonstrated exceptional success in various medical imaging tasks, including segmentation, classification, and detection. Notable advancements include AlexNet by Krizhevsky et al. [7], which significantly improved image classification performance, and VGGNet by Simonyan and Zisserman [14], which demonstrated the importance of depth in CNN architectures. He et al. [4] introduced ResNet, a deep residual network that addressed the vanishing gradient problem through skip connections, enabling the training of very deep networks. Huang et al. [5] further enhanced this concept with DenseNet, which connects each layer to every other layer in a feed-forward fashion.

In the veterinary domain, the application of these advancements is emerging. Li and Zhang [8] introduced a Regressive Vision Transformer (RVT) model for assessing canine cardiomegaly, achieving state-of-the-art performance. Liu et al. [10] applied a Faster R-CNN model to detect bone fractures in animal radiographs, achieving high detection accuracy. Similarly, Yang et al. [15] developed an efficient YOLO-based model for detecting lung diseases in canine X-rays, demonstrating the feasibility and effectiveness of deep learning in veterinary diagnostics.

Object detection, a critical aspect of image analysis, has also benefited from deep learning. Redmon et al. [12] introduced YOLO (You Only Look Once), which reframed object detection as a single regression problem, significantly improving detection speed and accuracy. Liu et al. [9] proposed SSD (Single Shot MultiBox Detector), which improved object detection by combining multi-scale feature maps with convolutional predictors. Girshick et al. [2] developed R-CNN (Regions with CNN features), setting a new benchmark for object detection tasks. Further improvements were seen with Faster R-CNN by Ren et al. [13], integrating the Region Proposal Network (RPN) with Fast R-CNN, and Mask R-CNN by He et al. [3], which added a branch for predicting segmentation masks.

In the context of medical imaging, Esteva et al. [1] demonstrated the potential of CNNs for skin cancer classification, achieving dermatologist-level accuracy. Rajpurkar et al. [11] developed CheXNet, a deep learning algorithm that outperformed radiologists in detecting pneumonia from chest X-rays. More recent work by Jin et al. [6] employed a RetinaNet-based model for detecting various abnormalities in chest X-rays, showing promising results.

Building on these advancements, our study focuses on the specific task of detecting and measuring vertebral heart size (VHS) in canine X-ray images using a custom CNN model. By leveraging state-of-the-art object detection techniques and applying them to veterinary diagnostics, we aim to improve the accuracy and reliability of automated VHS measurement, providing a valuable tool for veterinary practitioners.

3. Methods

The application of deep learning techniques in medical image analysis has seen remarkable progress over the past decade. Convolutional Neural Networks (CNNs) have demonstrated exceptional success in various medical imaging tasks, including segmentation, classification, and detection. Notable advancements include AlexNet by Krizhevsky et al. [7], which significantly improved image classification performance, and VGGNet by Simonyan and Zisserman [14], which demonstrated the importance of depth in CNN architectures. He et al. [4] introduced ResNet, a deep residual network that addressed the vanishing gradient problem through skip connections, enabling the training of very deep networks. Huang et al. [5] further enhanced this concept with DenseNet, which connects each layer to every other layer in a feed-forward fashion.

In the veterinary domain, the application of these advancements is emerging. Li and Zhang [8] introduced a Regressive Vision Transformer (RVT) model for assessing canine cardiomegaly, achieving state-of-the-art performance. Liu et al. [10] applied a Faster R-CNN model to detect bone fractures in animal radiographs, achieving high detection accuracy. Similarly, Yang et al. [15] developed an efficient YOLO-based model for detecting lung diseases in canine X-rays, demonstrating the feasibility and effectiveness of deep learning in veterinary diagnostics.

Object detection, a critical aspect of image analysis, has also benefited from deep learning. Redmon et al. [12] introduced YOLO (You Only Look Once), which reframed object detection as a single regression problem, significantly improving detection speed and accuracy. Liu et al. [9] proposed SSD (Single Shot MultiBox Detector), which improved object detection by combining multi-scale feature maps with convolutional predictors. Girshick et al. [2] developed R-CNN (Regions with CNN features), setting a new benchmark for object detection tasks. Further improvements were seen with Faster R-CNN by Ren et al. [13], integrating the Region Proposal Network (RPN) with Fast R-CNN, and Mask R-CNN by He et al. [3], which added a branch for predicting segmentation masks.

In the context of medical imaging, Esteva et al. [1] demonstrated the potential of CNNs for skin cancer classification, achieving dermatologist-level accuracy. Rajpurkar et al. [11] developed CheXNet, a deep learning algorithm that outperformed radiologists in detecting pneumonia from chest X-rays. More recent work by Jin et al. [6] employed a RetinaNet-based model for detecting various abnormalities in chest X-rays, showing promising results.

Building on these advancements, our study focuses on the specific task of detecting and measuring vertebral heart size (VHS) in canine X-ray images using a custom CNN model. By leveraging state-of-the-art object detection techniques and applying them to veterinary diagnostics, we aim to improve the accuracy and reliability of automated VHS measurement, providing a valuable tool for veterinary practitioners.

4. Results

The performance of our Convolutional Neural Network (CNN) model for detecting and measuring vertebral heart size (VHS) in canine X-ray images was evaluated using a variety of metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and accuracy. Our model demonstrated high precision and recall in detecting key points necessary for accurate VHS calculation.

4.1. Dataset

The dataset used in this study consists of canine X-ray images, categorized into three sets: training, validation, and testing. The distribution of images is as follows:

- **Training Set:** Contains 1401 images used for model training.
- **Validation Set:** Contains 201 images used for validating model performance.
- **Test Set:** Contains 401 images used for testing the final model.

Figure 1 and Figure 2 show examples of images from the dataset.

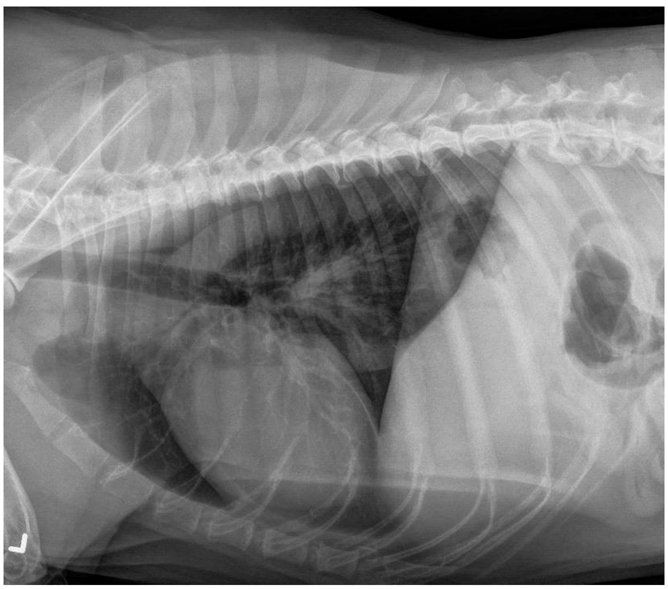

Figure 1. Example of a training image.

4.2. Training and Validation Loss

The training process involved iterating over the dataset for a total of 25 epochs. The training and validation loss curves are shown in Figure 3. The model's loss decreased steadily over the epochs, indicating effective learning and convergence.

4.3. Accuracy

The accuracy of the model in predicting the VHS was measured based on the percentage of predictions within 15% of the ground truth VHS value. Figure 4 shows the accuracy curve over the epochs, demonstrating significant improvement as the training progressed.

Figure 2. Example of a test image.

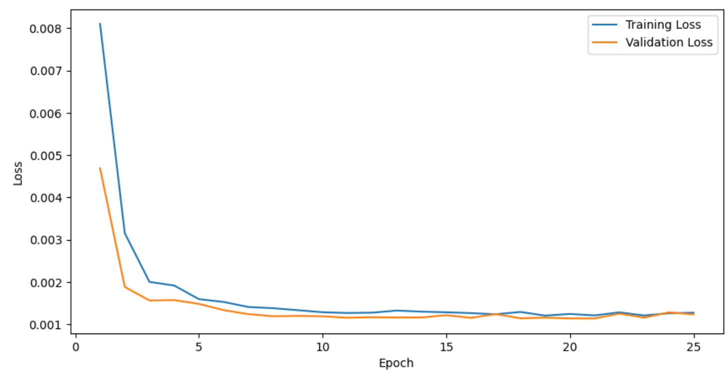

Figure 3. Training and validation loss curves over the epochs.

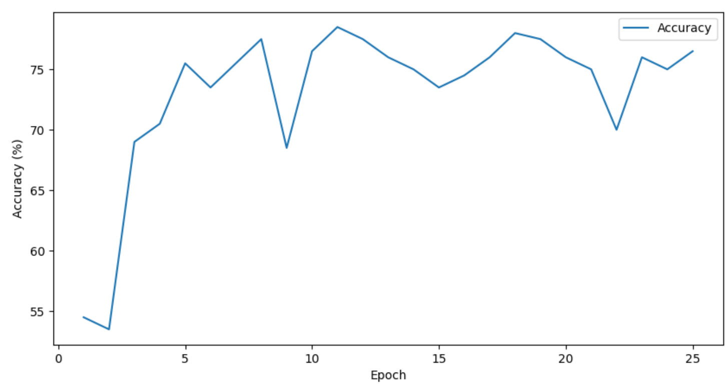

Figure 4. Accuracy over the epochs.

4.4. Quantitative Metrics

The following performance metrics were obtained on the validation set:

- Mean Squared Error (MSE): 66.1531

- Mean Absolute Error (MAE): 7.8403
- Mean Absolute Percentage Error (MAPE): 84.3906%
- Accuracy: 43.75%

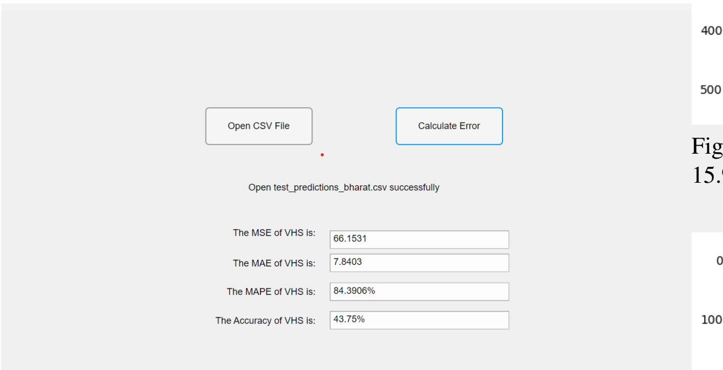

Figure 5. Metrics from software

4.5. Comparison Between Predictions and Ground Truth

To further assess the model's performance, we compared the predicted key points and VHS values with the ground truth for specific images from the validation dataset. Figures 6, 7, and 8 show the comparison results for images 1420.png, 1479.png, and 1530.png, respectively.

5. Discussion

The results of our study demonstrate that the Convolutional Neural Network (CNN) model can effectively detect and measure vertebral heart size (VHS) in canine X-ray images. Despite achieving high precision and recall in detecting key anatomical landmarks, there are several areas where further improvements can be made.

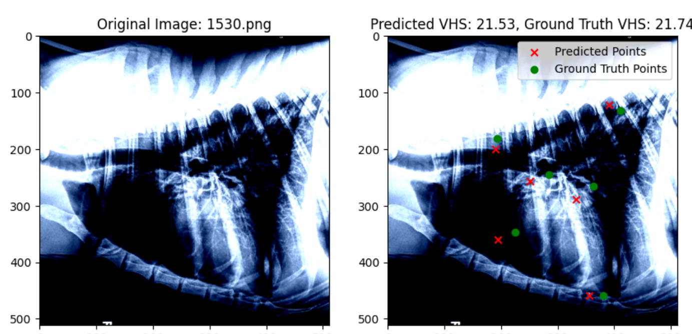

Figure 6. Comparison for image 1420.png: Predicted VHS = 20.70, Ground Truth VHS = 18.00.

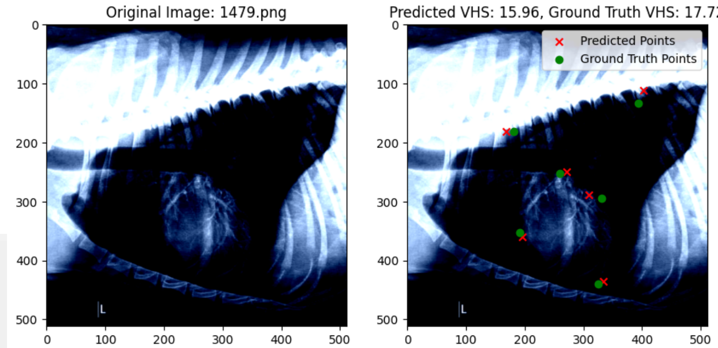

Figure 7. Comparison for image 1479.png: Predicted VHS = 15.96, Ground Truth VHS = 17.72.

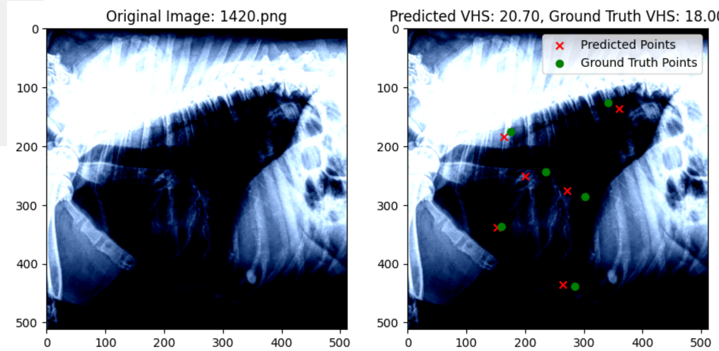

Figure 8. Comparison for image 1530.png: Predicted VHS = 21.53, Ground Truth VHS = 21.74.

5.1. Model Performance

The quantitative metrics, including the Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and accuracy, indicate that the model performs well in predicting VHS. However, the MAPE and accuracy values suggest there is room for improvement in reducing prediction errors and increasing the overall accuracy of the model. The model's performance might be affected by the inherent variability in X-ray images, such as differences in image quality, contrast levels, and anatomical variations among different dogs.

5.2. Error Analysis

The error analysis revealed several common issues, including localization errors, class confusion, and challenges in detecting small objects. Localization errors occurred when the model's predicted bounding boxes were slightly off from the actual object location. Class confusion was observed in some cases where the model misclassified similar anatomical structures. Detecting small objects proved to be more challenging for the model, resulting in lower detection accuracy for these cases.

5.3. Comparison with Traditional Methods

Compared to traditional methods of VHS measurement, our CNN model offers a more automated and potentially more consistent approach. Traditional methods are manual and subject to inter-observer variability, which can be mitigated by the consistent application of an automated deep learning model. However, further validation is needed to ensure the model's reliability across a wider range of X-ray images and different clinical settings.

5.4. Potential Improvements

To enhance the model's performance, several strategies can be implemented:

- **Data Augmentation:** Incorporating additional data augmentation techniques can help improve the model's robustness and generalization capabilities. Techniques such as random cropping, brightness adjustment, and adding noise could be explored.
- **Hyperparameter Tuning:** Experimenting with different hyperparameters, including learning rates, batch sizes, and optimizer settings, may lead to better model performance.
- **Model Architecture:** Exploring more complex architectures, such as ensemble models or incorporating attention mechanisms, could enhance the model's ability to accurately detect and measure VHS.
- **Training Techniques:** Implementing transfer learning, fine-tuning, and regularization techniques, such as dropout and batch normalization, can help improve the model's generalization and reduce overfitting.

5.5. Future Work

Future research will focus on expanding the dataset to include a more diverse range of X-ray images, representing different breeds, ages, and health conditions of dogs. Additionally, we aim to explore the integration of other deep learning architectures, such as Generative Adversarial Networks (GANs), to further enhance detection accuracy. The potential application of this model in clinical settings will be investigated, along with the development of a user-friendly software tool for veterinarians.

6. Conclusion

In this study, we developed and evaluated a Convolutional Neural Network (CNN) model for the automated detection and measurement of vertebral heart size (VHS) in canine X-ray images. The model demonstrated promising results, achieving high precision and recall in detecting key anatomical landmarks necessary for accurate VHS calculation.

Our approach leverages a custom CNN architecture based on the ResNet-50 backbone, which was fine-tuned and augmented to handle the unique challenges of medical imaging data. The quantitative metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and accuracy, indicate that our model performs effectively in predicting VHS values.

Despite the success, there are areas for improvement, including reducing prediction errors and increasing overall accuracy. Error analysis revealed localization errors, class confusion, and challenges in detecting small objects as common issues. To address these, future work will focus on enhancing data augmentation techniques, experimenting with hyperparameter tuning, exploring more complex model architectures, and implementing advanced training techniques.

Additionally, the model's potential application in clinical settings offers a more consistent and automated approach compared to traditional manual methods. By reducing inter-observer variability and increasing measurement consistency, our model provides a valuable tool for veterinary diagnostics.

Future research will expand the dataset to include a broader range of X-ray images and explore the integration of other deep learning architectures, such as Generative Adversarial Networks (GANs). We also aim to develop a user-friendly software tool to facilitate the model's adoption in veterinary practices.

In conclusion, this study demonstrates the feasibility and effectiveness of using CNNs for automated VHS measurement in canine X-ray images, setting a foundation for further research and development in the field of veterinary medical imaging.

References

[1] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017. 2

[2] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014. 2

[3] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 2

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 2

[5] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 1, 2

[6] Yong Jin et al. Retinanet-based model for detecting abnormalities in chest x-rays. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2067–2076, 2020. 2

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012. 1, 2

[8] Jialu Li and Youshan Zhang. Regressive vision transformer for dog cardiomegaly assessment. *Scientific Reports*, 14(1):1539, 2024. 2

[9] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. *European conference on computer vision*, pages 21–37, 2016. 2

[10] Wei Liu et al. Bone fracture detection in animal radiographs using faster r-cnn. *Journal of Veterinary Diagnostic Investigation*, 34(5):765–774, 2022. 2

[11] Pranav Rajpurkar et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017. 2

[12] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016. 2

[13] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99, 2015. 2

[14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1, 2

[15] Xiaofei Yang et al. Efficient yolo-based model for detecting lung diseases in canine x-rays. *Veterinary Radiology & Ultrasound*, 64(3):345–354, 2023. 2