Diffusion Data Augmentation for Enhancing Norberg Hip Angle Estimation

Katz School of Science and Health

Sheng-Han Yueh, M.S. in Artificial Intelligence

Faculty Advisor: Youshan Zhang, Ph.D.

ABSTRACT

Medical imaging is fundamental to disease diagnosis and treatment, yet the availability of diverse, high-quality images remains constrained, hindering critical medical applications. Our project addresses this challenge by leveraging state-of-the-art computer vision techniques to generate high-quality medical images. The research focuses on utilizing radiographic images of dog hips to explore the potential of applying generative Artificial Intelligence (GenAI) techniques, specifically diffusion models, to medical images. We aim to augment the dataset for training a Norberg angle (NA) prediction model, a crucial metric for evaluating hip joint quality and diagnosing Canine Hip Dysplasia (CHD), the most prevalent hereditary orthopedic disorder in dogs. Controlling CHD prevalence involves excluding affected dogs from breeding programs, making the automated estimation of NA vital. Our study generated 1,274 images from a training set of 219, revealing an average improvement of 35.3% in loss across 18 models predicting NA. The results underscore the potential of GenAI techniques to address the challenges posed by data scarcity in the medical field.

INTRODUCTION

Motivation

- Access to diverse medical images is crucial for effective disease diagnosis & treatment.
- Limited access, especially for rare conditions, hinders medical applications.
- GenAI augments the dataset, improving model performance (Moon et al., 2024).
- The diffusion model surpasses the image sample quality of existing GenAI models (Dhariwal et al., 2021).

Diffusion-based Systems

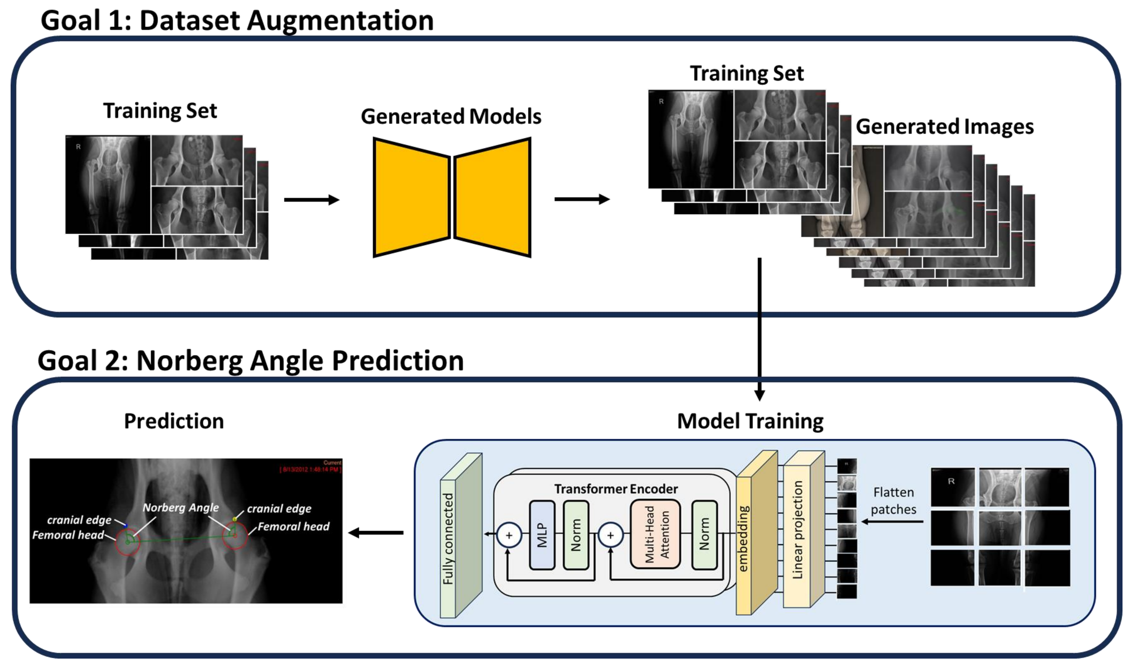
- Forward process: disrupts the distribution of an image by introducing noise
- Backward process: restores the image by eliminating the noise from a heavily distorted image. A neural network is trained to predict noise using the loss function defined by Eq. (1).

$$\text{Loss} = \left\| \epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t\right) \right\|^2 \qquad (1)$$

Aim

- To generate more images and to further improve accuracy and efficacy of our algorithms in NA estimation, we used advanced image generation models like Denoising Diffusion Probabilistic Models, DDPM (Ho et al., 2020) and DreamBooth (Ruiz et al., 2023).

METHODOLOGY



Goal 1: Dataset Augmentation

1. Train DDPM and DreamBooth models with training set (x).
2. Generate images with trained DDPM and DreamBooth models.
3. Combine training set and generated images as an augmented dataset (x')

Goal 2: Norberg Angle Prediction

1. Screen 18 pre-trained models based on the following training process:

```
for epoch in range(epochs):
    for img in Dataloader(images):
        Predict = model(img)
        Loss = MSE(ground truth - Predict)
        Loss.backward()
```

2. Select the best model by comparing validation losses from models trained with x and models trained with x'.
3. Fine-tune model.
4. Predict key points, Norberg angles, and radii in the testing set using fined-tuned model.

RESULTS

Quantity of Generated Images

1,274 images were generated using the DDPM and DreamBooth models.

Table 1. Generated Images in DDPM and DreamBooth

	# Images
Training	219
DDPM	307
DreamBooth	967

Quality of Generated Images

- DDPM model: low resolution and reproduce similar positions
- DreamBooth model: higher-resolution and diversified positions cartoon-like quality

Figure 1. Examples of Generated Images

Quantitative Results

Inclusion of generated images has the potential to enhance overall model performance.

Table 2. MAPE Comparison on the Testing Set

Model	w/o	w/	Improvement
SqueezeNet	48.71	12.33	74.69 %
EfficientNet	20.48	6.87	66.40 %
DenseNet161	34.77	14.88	57.20 %

Qualitative Results

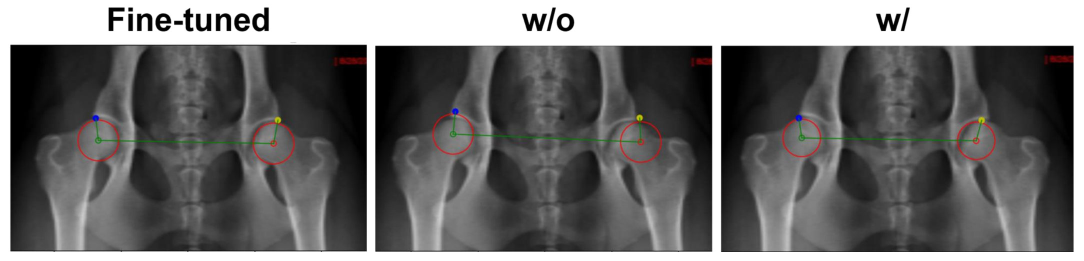


Figure 2. Predicted Results from EfficientNet for Validation Set

DISCUSSION & CONCLUSIONS

- The study has shown improved NA prediction with generated images using 18 different ImageNet models.
- Integrating diffusion models improved radiology image predictions, reducing MAPE by 35.3%.
- The study is crucial for addressing imbalances in medical image datasets, showcasing the effectiveness of synthetic data in enhancing model robustness and accuracy.
- For future work:
  - Implement various text-to-image generation methods, like CLIP (Radford et al., 2021) and VQGAN (Crowson et al., 2022), creating diverse imagery.
  - Enhance angle estimation model by expanding the dataset and leveraging state-of-the-art transformers, aiming for more accurate predictions.

REFERENCES

Crowson, K., Biderman, S., Kornis, D., Stander, D., Hallahan, E., Castricato, L., & Raff, E. (2022). Vqgan-clip: Open domain image generation and editing with natural language guidance. In European Conference on Computer Vision (pp. 88–105). Springer.

Dhariwal, P., & Nichol, A. (2021). Diffusion models beat GANs on image synthesis. In Advances in neural information processing systems, 34, 8780-8794.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. In Advances in neural information processing systems, 33, 6840–6851.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700–4708).

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. arXiv preprint arXiv:1602.07360.

Moon, H. H., Jeong, J., Park, J. E., Kim, N., Choi, C., Kim, Y. H., ..., & Kim, H. S. (2024). Generative AI in glioma: ensuring diversity in training image phenotypes to improve diagnostic performance for IDH mutation prediction. Neuro-oncology, noae012.

Radford, A., Kim, J. W., Hallacy, C., et al. (2021). Learning transferable visual models from natural language supervision. In International conference on machine learning (pp. 8748–8763).

Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2023). Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In Proceedings of CVPR (pp. 22 500–22 510).

Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8697–8710).