Generative Bone Lesions Synthesis for Data Augmentation in X-ray

Anant Gupta*
New York University
New York, NY
anantgupta@nyu.edu

Sumit Chopra
Imagen Technologies
New York, NY
sumit@imagen.ai

Christian Ledig
Imagen Technologies
New York, NY
christian@imagen.ai

Abstract

Insufficient training data and severe class imbalance are often the main limiting factors when developing machine learning models for the classification of rare diseases. In this work, we address this problem by augmenting the training set with synthesized images. We pose the generative task as an unsupervised image-patch translation problem with the aim to generate bone lesions on images without pathology. In experimental results, we show that this can enable the training of superior classifiers achieving better performance on a held-out test set in the binary classification task of bone lesion detection. Additionally, we demonstrate the feasibility of transfer learning and apply a generative model that was trained on one bone to another.

1 Introduction

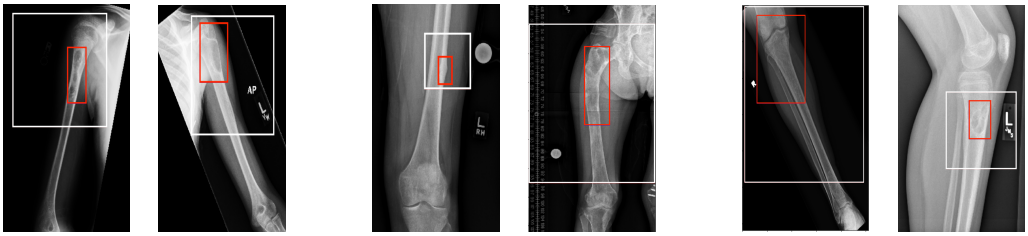
Deep neural networks have demonstrated their potential to reach human-level performance for image classification, however, their performance generally correlates with the amount of available training samples. When focusing on rare medical conditions, the limited availability of pathological training images can cause severe class imbalance and limits the accuracy of machine learning models. One example of a pathology that is both of high interest but also rare is bone lesion [1]. The binary classification of the presence of bone lesion in X-ray images is the subject of our work.

Several methods have been proposed to address the class-imbalance problem, including image transformations [2] and sampling strategies [3], such as under- or over-sampling. Often those approaches are of limited benefit as they do not address the inherent problem of dealing with a small training set not fully representing the underlying data distribution. Recent works have proposed the use of generative adversarial networks (GANs) [4] to generate synthetic data from scratch in order to augment and increase diversity in the training set [5, 6]. However, learning to generate high-resolution images from random noise requires an often prohibitively large training data set too.

In this work, we aim to synthesize bone lesions by translating spatially-constrained patches extracted from non-pathological X-rays rather than translating whole images or generating from scratch. The model is trained on patches extracted from full images to ensure localized generation of pathology (c.f. Figure 1). A blending approach is described that merges the translated patches back into full images. A standard classifier is subsequently trained for bone lesion detection for the following bones

*Work done when author was at Imagen Technologies.

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.



(a) Humerus (b) Femur (c) Tibia

Figure 1: Bone lesion with expert-annotated red bounding box. Random sized patches, cropped around lesions, used for training the generative model is marked by the white bounding box.

(body parts): humerus, tibia and femur. We observed performance gains for individual body parts by using an augmented training set. We further show that transfer learning is achieved by using an image-translation model trained on one body part to generate bone lesions on other similar-looking body parts, for which training data is too noisy or insufficient in size to train a model with good performance.

Related Work GANs are being used in the medical domain to accomplish tasks like image translation [7, 8, 9, 10], image segmentation [11, 12, 13] and data augmentation in controlled settings [14, 15]. Multiple works have addressed lesion classification [16, 17, 18], but to our knowledge, there has been no substantial work focused on developing machine learning models for bone lesion classification from X-ray images.

2 Methodology

Generative Model The generation of bone lesion pathology is posed as an unsupervised image-to-image translation task [19], where $P\chi_l$ and $P\chi_h$ are the two distributions from which X-ray image patches showing bone anatomy with lesion x_l, and bone anatomy without lesion x_h are drawn respectively. The model then maps these unpaired samples to a shared latent space Z, using encoders for respective distributions: $E_l(x_l) = E_h(x_h) = z \in Z$. The generators then decode back the input sample from this latent vector: $G_l(z) = x_l, G_h(z) = x_h$. The model is trained to optimize the following objectives: i) ELBO [20]: in order to learn an approximate posterior, ii) adversarial [4]: to ensure the generated images resemble samples from the target distribution, and iii) cycle-consistency [21]: to ensure the cross-domain translators are inverses of each other.

The weighted losses are optimized jointly. We closely follow the hyperparameter settings from [19]. Lesion-like properties are generated with the following translation operation: $x_{h\to l} = G_l(E_h(x_h))$.

Patch-making Bone lesions tend to cause local alterations in bone anatomy without substantially affecting the remaining, global visual appearance of the image. We therefore aim to translate localized image patches rather than training a translation model for the complete images. Patches allow the training of a generative model that is both computationally more efficient to optimize and that also leads to higher quality generated samples.

Blending The translated patches do not only consist of generated bone-lesion-like pathology, but also exhibit subtle changes in the overall image characteristics, such as contrast and brightness. We employ alpha-blending to smoothly blend the translated patch in the original image as: $\alpha x_{h\to l} + (1-\alpha)x_h$. Specifically, we define a locally varying blending factor α as: $\alpha = \cos(|i|^n * \frac{\pi}{2}) \cos(|j|^n * \frac{\pi}{2})$, where i and j are interpolations in the interval [-1, 1] and n is a hyper-parameter.

Binary Classifier We used a standard dilated residual net (DRN) [22] for the binary classification task. In contrast to the image translation model the binary classifier was trained on full, uncropped images. See Figure 2 for an example of a synthesized training image.

3 Results

Dataset A set of X-ray images (c.f. Table 1) showing bone anatomy with and without lesion are sourced from various U.S. hospitals and assessed by expert, board-certified radiologists (c.f. Figure 1). Images acquired from pediatric patients or showing confounding image features (e.g., congenital,

Table 1: Data Splits for each model. Left: classification. Right: generation; the classes are kept balanced for training. The source samples are only non-lesion and used for creating the augmented sets. In both tables, the ratio denotes lesion:non-lesion class split.

Classification Model			
Body part	**Train**	**Validation**	**Test**
Humerus	268:2295	41:305	50:500
Tibia	214:14482	22:1628	50:500
Femur	32:4558	14:573	50:500

Generative Model		
Body part	**Train**	**Source**
Humerus	536:536	4643
Tibia	515:515	4680
Femur	285:285	9171

Table 2: Test-set evaluation of bone lesion detection models.

Body part	**Type**	**Threshold t**	**Augm. Samples**	**ROC AUC (CI 95%)**
Humerus	Baseline	0	0	0.876 (0.817-0.926)
Humerus	Augmented	0.9	401	0.924 (0.889-0.955)
Tibia	Baseline	0	0	0.618 (0.532-0.705)
Tibia	Augmented	0.9	124	0.640 (0.547-0.732)
Tibia	TL	0.9	1264	0.698 (0.610-0.785)
Femur	Baseline	0	0	0.533 (0.441-0.627)
Femur	TL	0.95	1342	0.682 (0.594-0.764)

fixation hardware) are removed from the data set for the classification task. We do not exclude those images when training the generative model as it is trained on cropped image patches. A test data set was held out ensuring sufficient positive samples and used at no point to train or fine-tune any model.

Augmented Classification We trained a baseline model for each body part on the given training set to predict the probability of the X-ray showing a bone lesion. To augment the training set with few samples but those that likely resemble characteristic bone lesion pathology, the baseline classifier is used to filter the generated images that achieve a score greater than a certain threshold (t), which is chosen on the validation set. In both the models, regularization is performed through augmentation procedures consisting of linear transformations, along with L2 weight decay. Table 2 summarizes the results for training on the baseline and augmented training set measured as Area Under the ROC-Curve (AUC) with bootstrapped confidence intervals (CI).

Transfer Learning In comparison to the available humerus X-rays, the available tibia and femur data sets were highly heterogeneous in terms of radiographic view and often showed confounding image content such as presence of external objects. This made it particularly challenging to train a valuable generative model for tibia and practically impossible for femur. We explored the potential of straightforward transfer learning and applied the generative model trained on humerus to generate lesions on both tibia and femur images. The humerus lesion classifier was used to filter the generated samples. On both tibia and femur test sets we observed substantially increased AUC scores indicating that transfer learning can be a powerful approach to enrich low-data training sets.

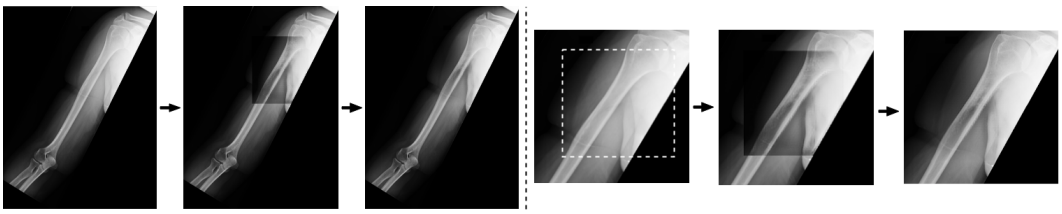


Figure 2: Stages of patch translation for full image (left) and selected patch with surrounding context (right): i) original, ii) translated and iii) blended. The white dotted box highlights the translated patch.

4 Conclusion

We trained a generative model that can represent some properties of the target pathology (bone lesions in X-ray) and synthesize those into sample patches drawn from another distribution (normal anatomy). When employing generative models for augmenting medical imaging data sets, great care needs to be taken to avoid and control for possibly introduced bias. Future work should be concerned with the exploration of those limitations and explore the method's potential on both a more diverse set of disease pathology and other imaging modalities.

Acknowledgements

The project is funded by Imagen Technologies. The work presented in this manuscript is for research purposes only and is not for sale within the United States.

References

[1] Alessandro Franchi. Epidemiology and classification of bone tumors. *Clinical Cases in mineral and bone metabolism*, 9(2):92, 2012.

[2] Zeshan Hussain, Francisco Gimenez, Darvin Yi, and Daniel Rubin. Differential data augmentation techniques for medical imaging classification tasks. In *AMIA Annual Symposium Proceedings*, volume 2017, page 979. American Medical Informatics Association, 2017.

[3] Rashmi Dubey, Jiayu Zhou, Yalin Wang, Paul M Thompson, Jieping Ye, Alzheimer's Disease Neuroimaging Initiative, et al. Analysis of sampling techniques for imbalanced data: An n= 648 adni study. *NeuroImage*, 87:220–241, 2014.

[4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[5] Antreas Antoniou, Amos Storkey, and Harrison Edwards. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*, 2017.

[6] Giovanni Mariani, Florian Scheidegger, Roxana Istrate, Costas Bekas, and Cristiano Malossi. Bagan: Data augmentation with balancing gan. *arXiv preprint arXiv:1803.09655*, 2018.

[7] Karim Armanious, Chenming Yang, Marc Fischer, Thomas Küstner, Konstantin Nikolaou, Sergios Gatidis, and Bin Yang. Medgan: Medical image translation using gans. *arXiv preprint arXiv:1806.06397*, 2018.

[8] Jelmer M Wolterink, Anna M Dinkla, Mark HF Savenije, Peter R Seevinck, Cornelis AT van den Berg, and Ivana Išgum. Deep mr to ct synthesis using unpaired data. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 14–23. Springer, 2017.

[9] Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 417–425. Springer, 2017.

[10] Qianye Yang, Nannan Li, Zixu Zhao, Xingyu Fan, Eric I Chang, Yan Xu, et al. Mri image-to-image translation for cross-modality image registration and segmentation. *arXiv preprint arXiv:1801.06940*, 2018.

[11] Yuan Xue, Tao Xu, Han Zhang, L Rodney Long, and Xiaolei Huang. Segan: Adversarial network with multi-scale l 1 loss for medical image segmentation. *Neuroinformatics*, pages 1–10, 2018.

[12] Avisek Lahiri, Kumar Ayush, Prabir Kumar Biswas, and Pabitra Mitra. Generative adversarial learning for reducing manual annotation in semantic segmentation on large scale miscroscopy images: Automated vessel segmentation in retinal fundus image as test case. In *Conference on Computer Vision and Pattern Recognition Workshops*, pages 42–48, 2017.

[13] Konstantinos Kamnitsas, Christian Baumgartner, Christian Ledig, Virginia Newcombe, Joanna Simpson, Andrew Kane, David Menon, Aditya Nori, Antonio Criminisi, Daniel Rueckert, et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In *International Conference on Information Processing in Medical Imaging*, pages 597–609. Springer, 2017.

[14] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *arXiv preprint arXiv:1803.01229*, 2018.

[15] Hojjat Salehinejad, Shahrokh Valaee, Tim Dowdell, Errol Colak, and Joseph Barfett. Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks. *arXiv preprint arXiv:1712.01636*, 2017.

[16] Tyler Bradshaw, Timothy Perk, Song Chen, Hyung-Jun Im, Steve Cho, Scott Perlman, and Robert Jeraj. Deep learning for classification of benign and malignant bone lesions in [f-18] naf pet/ct images. *Journal of Nuclear Medicine*, 59(supplement 1):327–327, 2018.

[17] Jie-Zhi Cheng, Dong Ni, Yi-Hong Chou, Jing Qin, Chui-Mei Tiu, Yeun-Chung Chang, Chiun-Sheng Huang, Dinggang Shen, and Chung-Ming Chen. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans. *Scientific reports*, 6:24454, 2016.

[18] 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, 2017.

[19] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In *Advances in Neural Information Processing Systems*, pages 700–708, 2017.

[20] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[21] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint*, 2017.

[22] Fisher Yu, Vladlen Koltun, and Thomas A Funkhouser. Dilated residual networks. In *CVPR*, volume 2, page 3, 2017.