

Daniel Blase*, Mengying Wang, Steffen Leonhardt, and Markus Lueken

Generation of Artificial Pressure Ulcer Images Using Deep Convolutional GANs

<https://doi.org/10.1515/cdbme-2025-0171>

Abstract: Pressure ulcers pose a threat, especially for older people and ICU patients. Data-driven approaches could allow for early detection and continuous evaluation. In this work, a deep convolutional generative adversarial network is presented which is able to generate new images of pressure ulcers. The cavities concerning the training dataset and the transfer-learning strategy are analyzed to optimally supplement small clinical datasets for more robust algorithm development.

Keywords: pressure ulcer, deep convolutional GAN, artificial image generation, deep learning

1 Introduction

Prolonged pressure and shear stress onto soft tissue can destroy cells, cause inflammation, and ischemia-reperfusion injuries, which jointly lead to localized damage of the tissue. Especially in patients with reduced mobility or older people, the skin and underlying tissues are affected mostly above bony prominences, resulting in the formation of pressure ulcers (PUs), also called pressure sores, pressure injuries, or decubitus. Typically, the wound healing process stagnates in the inflammatory phase leading to extensive hospital stays, medical attention, personal burden, high costs, risk of infection, and increased mortality [1]. An analysis of the Global Burden of Disease Study 2021 shows that the incidence of PUs has globally increased by 116% and the associated number of deaths by 122% in the years from 1990 to 2021 [2]. In German hospitals, 395,980 cases of having at least one PU were recorded in 2021, mostly of stage 2 or higher [3]. Depending on the depth of the affected tissue, PUs are classified into four stages (1-4), unstageable and assumed deep tissue injuries, each exposing a different layer of the skin with a different appearance [1].

The continuous improvement of AI algorithms for image processing have also been used to aid in the tasks of pressure ulcer segmentation and classification [4]. However, the small amount of PU images that is publicly available, is a common problem. Generative AI networks can help to increase datasets

or correct imbalances by creating and adding artificial images [5]. Generative adversarial networks (GANs) have been shown to achieve good results in unsupervised learning by using the concept of adversarial learning between a generator and a discriminator [6]. In 2015, Radford et al. proposed a GAN version focusing more on 4 de-/convolutional layers instead of linear or pooling layers, the deep convolutional GAN (DCGAN), for more refined features and colors in images of 64 by 64 pixels [7]. Subsequently, DCGANs with a maximum of 5 de-/convolutional layers, small datasets and a low image resolution (smaller than 128 by 128 pixels) have been employed to generate chronic wounds and skin lesions with moderate success [8, 9]. Diffusion models are a different modern technique for image generation using denoising but will not be discussed here [10]. This work contributes another variant of the DCGAN trained on a bigger dataset resulting in images with a higher resolution and clearer wound properties. Furthermore, different transfer learning approaches are discussed.

2 Methods

The dataset for the training process contains PU images from multiple publicly available datasets as well as images from a clinical study conducted in the Geriatric Department (Med. Klinik VI) of University Hospital RWTH Aachen (UKA), which additionally provided more images of stage 1 and stage 2 PUs, which had been taken during regular treatments for obligatory documentation and by patient agreement. The famous Medetec Wound Database (Medetec) [11], the AZH Wound Database (AZH) [12], the database of the foot ulcer segmentation challenge (FUSeg) [13] and the WoundsDB database [14] compose the publicly available part of the dataset. Table 1 gives an overview of the distribution of images per class in the different data sources.

In order to assess the influence of the background, two more datasets have been created based on the original one (OD) using the skin mask of an existing segmentation network for cropping (cropped dataset with less background, CD) and masking (cropped and masked dataset without any background, MD) as can be seen in Figure 1. Figure 2 shows target examples taken from the PU datasets (blue) in comparison to those used for pre-training with successive transfer-learning in this work (orange), which are the ImageNet dataset (more

*Corresponding author: Daniel Blase, Mengying Wang, Steffen Leonhardt, Markus Lueken, Medical Information Technology (MedIT), RWTH Aachen University, Pauwelsstraße 20, Aachen, Germany, e-mail: blase@hia.rwth-aachen.de

Tab. 1: Dataset overview.

Sources	1	2	3	4	unst.	Total
AZH	-	-	26	6	31	63
FUSeg	-	2	415	53	416	886
WoundsDB	-	-	-	-	26	26
Medetec	-	-	7	19	48	74
UKA	56	103	3	-	-	162
Total per class	56	105	451	78	521	1211

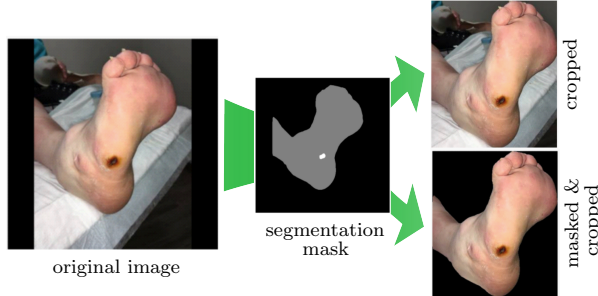


Fig. 1: Description of the three different datasets used: original, cropped or masked and cropped with the help of a segmentation network (skin is grey, wound is white).

than 14 mio. images) [15] and the Kaggle skin lesion dataset (25331 images) [16–18].

The experiments have been conducted on a server with GPUs using Visual Studio Code (Microsoft, Redmond, USA), Python 3.9, the supervision platform by Weights&Biases (Weights&Biases, San Francisco, USA), and the Pytorch framework. Even though the basic characteristics of a DCGAN such as strided convolutions in the discriminator, BatchNorm, ReLU and Tanh activation in the generator were met, some adaptations have been made. As applying transposed convolutional layers lead to stability issues, upsampling (mode: nearest) has been used instead. The Leaky ReLU



Fig. 2: Examples of images from the PU source datasets (blue) and the datasets used for pre-training (orange).

activation has been replaced with the similar but differentiable SELU activation to help with faster convergence, and the depth has been increased to a symmetrical pair of 6 convolutional layers in the generator and discriminator. Figure 3 shows the final structure of both networks. The input is a 100-dimensional noise vector following a standard normal distribution $Z \sim \mathcal{N}(0, 1)$. The convoluting kernels had a size of 3x3, stride was 2 (in the discriminator), and padding was set to 1. The final layer of the discriminator had a sigmoid activation to classify images into "real" or "fake" (binary).

Experiments without pre-trained networks (P-0), with a pre-trained generator (P-G), and with both networks having been pre-trained (P-GD) have been conducted. Kaggle's skin lesion dataset has been used for a 120-steps pre-training to let the network learn the concepts of skin, features on the skin and objects in the background. Furthermore, the decoder network of a pre-trained (ImageNet) VGG autoencoder using 5 transposed convolutions, BatchNorm and ReLU from [19] together with an untrained DCGAN discriminator with 5 strided convolutions to match the reduced depth and image dimension of 224 by 224 pixels has been tested. All pairs of generator and discriminator have been trained with the classic binary-cross-entropy loss function (BCE) and the Adam optimizer. During one training step, the discriminator's weights are updated twice (real and fake image), while the generator's weights are only updated once.

Pre-processing included resizing to 256 by 256 pixels, flip, rotation by 90 or -90 degrees and brightness and contrast adjustments, all with a probability of 70%. For quantification of the authenticity and diversity of the generated images, the Fréchet Inception Distance (FID) score has been used, which compares the mean and covariance of feature vectors generated by an Inception-v3 network between a set of real and a set of generated images [20]. To achieve a stable training process, a batch size of 4, a similar learning rate of smaller than 0.0008 in both antagonists, and using SELU and ReLU activations were necessary. Adding the convolutional layer to transform the 32 channels after the last upsampling layer first into 16 and subsequently into 3 color channels has proven useful for training stability.

3 Results

The FID score is a relative score that optimally should approach 0. For baselining, the original dataset has been split into two halves and the FID score of comparing real images with real images resulted in an FID score of 53.29. Comparing 25,331 acceptable images of the pre-trained generator with the real skin lesion dataset gave a FID score of 129.91. Con-

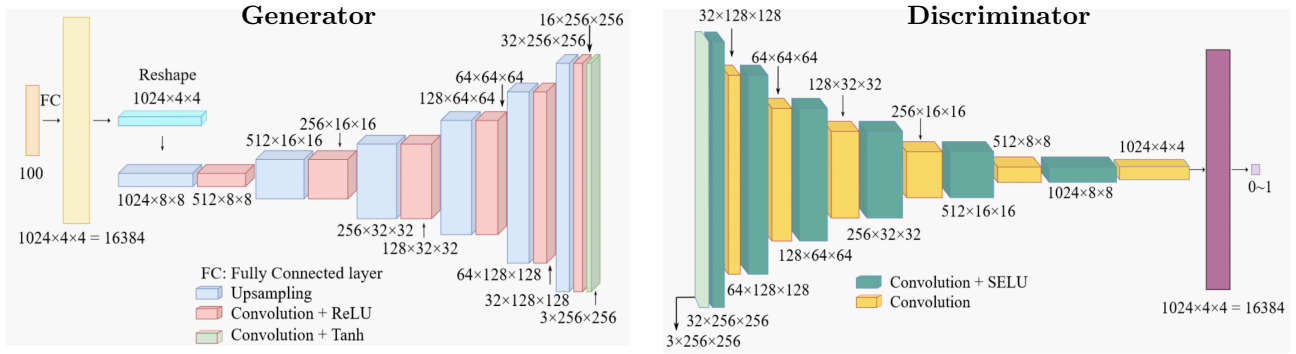


Fig. 3: Network structure of the DCGAN using upsampling and the SELU activation function.

sequently, FID scores between 50 and 130 have been considered optimal to satisfactory. In order to evaluate the generator's performance qualitatively, five criteria have been evaluated in addition to the FID score: presence of artifacts, realistic skin color, realistic background, meaningful shape of the body part, and diversity of the results. Figure 4 presents exemplary results of the most successful combinations of dataset, architecture and transfer-learning strategy noted as Gx ("generator x").

4 Discussion

The high FID score among the original dataset itself has shown the large diversity of images, correlating with the difficulty of the generation task, justifying deeper GAN structures. This is also supported by the fact that the pre-trained VGG-16 generator did not perform as well as the DCGAN variants (G6 vs. G1-G4). However, the FID score alone has been found to be insufficient to guarantee the generation of good-quality images. It only allows for a rough quality assessment (see G4 vs. G5). Selecting networks according to FID and quality metrics resulted in generators that were able to generate sufficiently accurate PU images with different shapes of wounds and body parts. Reconstructing the background, however, turned out to be a difficult task for all generators. Training on the masked dataset showed the best performance in skin and wound features (G3). If only the wound is of interest, this training strategy is recommended. If the background is of relevance, the two-stage approach of first training without background and after that training on the same dataset with background achieved good results (G4). The partially disjunct skin parts due to the cropping and masking technique also cause disjunct skin areas in generated images. Apart from that, pre-training and transfer-learning on the skin lesion dataset reduced the quality of the images, at least in the number of epochs trained. As the dataset size is relatively small compared to the number of trainable parameters,

it has originally been assumed that pre-training would improve the results. Another general problem that has been identified, is that sometimes the generated images contain no wound at all (e.g. G6 right), even though all training images have displayed wounds. A possible cause is that in most images, the wound covers fewer pixels than the skin.

5 Conclusion

The combination of the dataset and the DCGAN structure resulted in images with better skin texture, wound features, and colors compared to previous publications, but showed problems with background creation and identifying the concepts of wound, skin, and background individually. The generated images are expected to be sufficient to contribute to enhancements in algorithms focusing on the wound part. Reducing the depth of the discriminator seems promising to support a more similar learning progress in the adversarial network pair. Furthermore, being able to generate PU images of a specific class or to generate background, skin, and wounds separately and adjusting the cropping and masking technique is expected to improve the image quality. Finally, a validation by clinical experts should be performed.

Author Statement

This work has been supported by the project "NeuroSys", which as part of the initiative "Clusters4Future" is funded by the Federal Ministry of Education and Research BMBF (03ZU1106DA). The authors state no conflict of interest. The study for obtaining the UKA Dataset includes informed consent from all individuals. It complies with all the relevant national regulations and institutional policies, and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the responsible ethics committee at the medical faculty of RWTH Aachen (EK 326/19).

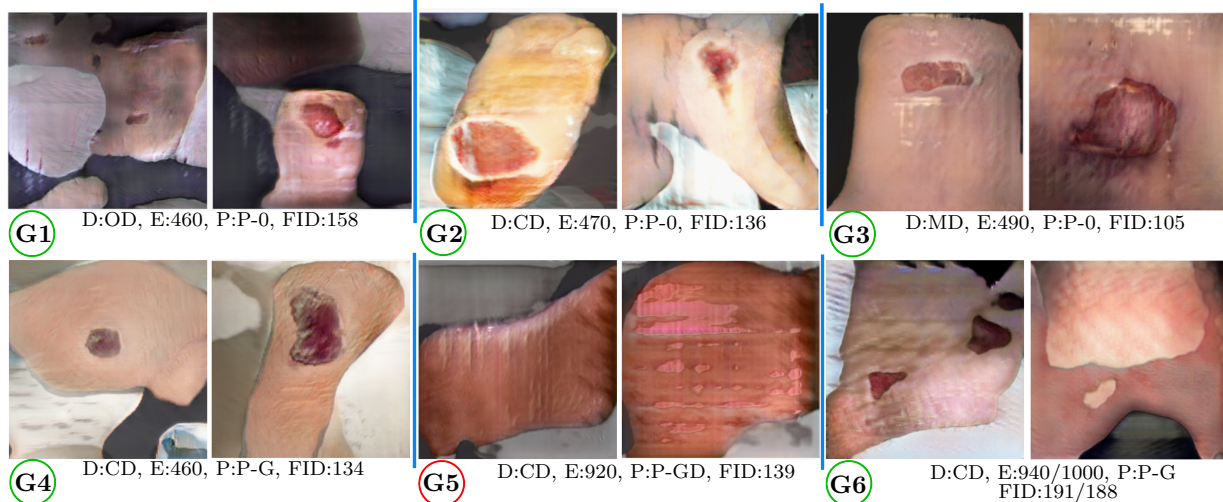


Fig. 4: Exemplary results of the best tests regarding DCGANs without pre-training (P, G1-G3), with a generator trained on the MD (G4), a fully pre-trained DCGAN (G5) and a pre-trained VGG16 as generator (G6). The epoch (E) and the dataset (D) are noted.

References

- [1] European Pressure Ulcer Advisory Panel, National Pressure Injury Advisory Panel and Pan Pacific. Prevention and Treatment of Pressure Ulcers/Injuries: Clinical Practice Guideline: The International Guideline. EPUAP/NPIAP/PPPIA; 2019.
- [2] Wang X, Jiang Y, Zhang W, Zhao W, Xu Y, Wang N. Global, Regional, and National Burden of Decubitus ulcers from 1990 to 2021: Findings from the 2021 Global Burden of Disease Study. 2024.
- [3] Lahmann N, Mayer M F, Posnett J. Pressure ulcers in German hospitals: Analysis of reimbursement and length of stay. *Open medicine* 2024;19:20230839. 10.1515/med-2023-0839.
- [4] Swerdlow M, Guler O, Yaakov R, Armstrong DG. Simultaneous Segmentation and Classification of Pressure Injury Image Data Using Mask-R-CNN. *Computational and mathematical methods in medicine* 2023; 2023:3858997.
- [5] Yang Y, Khorshidi HA, Aickelin U. A review on over-sampling techniques in classification of multi-class imbalanced datasets: insights for medical problems. *Frontiers in digital health* 2024; 6:1430245.
- [6] Hong Y, Hwang U, Yoo J, Yoon S. How Generative Adversarial Networks and Their Variants Work. *ACM Comput. Surv.* 2020; 52:1–43. Available: <https://arxiv.org/abs/1711.05914>.
- [7] Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks; 2015 Nov 19. Available: <http://arxiv.org/pdf/1511.06434>.
- [8] Zhang J, Zhu E, Guo X, Chen H, Yin J. Chronic Wounds Image Generator Based on Deep Convolutional Generative Adversarial Networks. *Theoretical computer science: 36th National Conference*. 2018.
- [9] Behara K, Bhero E, Agee JT. Skin Lesion Synthesis and Classification Using an Improved DCGAN Classifier. *Diagnostics* 2023; 13:2635.
- [10] Ho J, Jain A, Abbeel P. Denoising Diffusion Probabilistic Models. Canada; 2020 Jun 19. *NeurIPS* 2020.
- [11] Thomas S, Medetec wound database: stock pictures of wounds. MEDETEC™. Available: medetec.co.uk/files/medetec-image-databases.html? Accessed 2023.
- [12] Anisuzzaman D M, Patel Y, Rostami B, Niezgoda J, Gopalakrishnan S, Yu Z, Multi-modal wound classification using wound image and location by deep neural network. *Sci Rep* 12 2022; 20057. doi: 10.1038/s41598-022-21813-0 .
- [13] Wang C, Anisuzzaman D M, Williamson V, Dhar M K, Rostami B, Niezgoda J et al. Fully automatic wound segmentation with deep convolutional neural networks. *Sci Rep* 10 2020; 21897. doi:10.1038/s41598-020-78799-w .
- [14] Kręćichwost M, Czajkowska J, Wijata A, Juszczak J, Pyćński B, Biesok M, et al. Chronic wounds multimodal image database. *Comput Med Imaging Graph.* 2021 Mar;88:101844. doi: 10.1016/j.compmedimag.2020.101844.
- [15] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. 2014, Available: [arXiv:1409.0575](http://arxiv.org/abs/1409.0575).
- [16] Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* 5, 2018:180161.
- [17] Combalia M, Codella N C F, Rotemberg V, Helba B, Vilaplana V, Reiter O, et al. BCN20000: Dermoscopic Lesions in the Wild. 2019; Available: [arXiv:1908.02288](http://arxiv.org/abs/1908.02288).
- [18] Codella N C F, Gutman D, Celebi M E, Helba B, Marchetti M A, Dusza S W, et al. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging. 2017; Available: [arXiv:1710.05006](http://arxiv.org/abs/1710.05006).
- [19] Horizon2333: GitHub - Horizon2333/imagenet-autoencoder: AutoEncoder trained on ImageNet. Version 2021. Available: <https://github.com/Horizon2333/imagenet-autoencoder>.
- [20] Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Advances in Neural Information Processing Systems* 30.