Tell Me If It Heals: Future Wound Stage Prediction Through Latent Space Extrapolation

anonymous

anonymous
anonymous@anonymous

Abstract. Prediction of the healing trajectory of a wound is a challenging problem both in the medical domain and in machine learning. This paper tackles the problem using a conditional generative adversarial network. The proposed framework contains three main components: 1) an image encoder to transform wound images into temporally-relevant 16-dimensional embeddings, 2) linear extrapolation of future wound image in the latent space given the two images of the wound on preceding days, 3) a conditional image generation for generating future wound images that are realistic, represent the corresponding healing stage, and provide coherency for the appearance of the wound throughout the healing process. The proposed framework can be used for the prediction of the healing trajectory of a wound, including an analysis of the wound healing stage to identify the need for early intervention.

Keywords: Generative adversarial network · Wound stage prediction · Trajectory prediction.

1 Introduction

Wounds and their healing process are intensely studied in the medical domain. Whether or not a wound will heal, how long the healing process takes, or how much intervention is needed are all aspects of the wound healing process. Recent works on wound image analysis using Machine Learning (ML) methods aim to classify wound types, identify chronic conditions, or segment wound area using Deep Learning (DL) architectures, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN) [1–3]. This paper intends to take a step further by leveraging Generative Adversarial Network (GAN) [4] to predict the healing trajectory of a wound. A wound can be assessed more effectively, and treatment plan can be adjusted sooner if the healing trajectory information is known.

GANs are used to generate different types of images and videos [6–8]. In this work, we generate what a wound will look like in the future. We use a variation of a conditional Deep Convolutional GAN (DCGAN) [9, 10] to generate wound images. The condition used for the GAN is a set of 16-dimensional latent embeddings extracted from a trained wound stage classifier. The GAN framework is used to generate future wound images with the capability of accurately identifying the associated wound stage in the future (Fig. 1). To the best of our

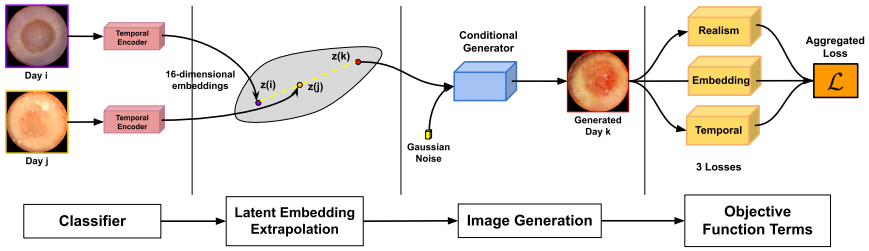


Fig. 1. Our model takes input images i and j, feeds them into the temporal image encoder [5], and extracts the latent embeddings. The model interpolates over wound image embeddings on days i and j to find wound image embedding on day k with $i < j < k$. Our generator takes in embedding k and Gaussian Noise to produce a generated wound image of day k. The GAN's objective function uses 3 loss functions (realism for real vs. fake classification, embedding as conditional loss, and temporal for temporal coherency between i, j, and k)

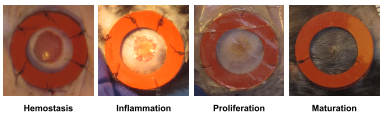
knowledge, we are the first to predict the trajectory of wounds. This work has excellent potential for impacting the analysis of human wounds and identifying the trajectory of wounds, especially in diabetic wound care.

2 Data

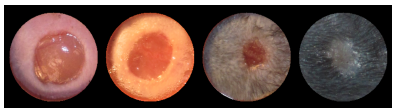
The data for this project is taken from the 16-day wound closure progress in C57BL/6J mice dataset [11, 12]. The dataset includes images of mice that have been administered small wounds, which are photographed for 16 days to observe their healing process. Each mouse has two wounds, one on the left and one on the right. Wound images from eight mice (four young and four old) exist in the dataset, providing us with 256 ($= 8 \times 2 \times 16$) data points. However, one image was removed due to camera failure, bringing the total to 255 data points.

During their healing process, a wound can be classified into one of the four healing stages: Hemostasis, Inflammation, Proliferation, and Maturation, where Proliferation and Maturation (also called *Remodeling*) stages span the most prolonged duration and have much more visual varieties. Based on a guideline provided by a group of medical professionals, the labeling task was carried out by ten non-expert annotators.

The raw images are cropped around the surgical splints to be centered and emphasized. Fig. 2a shows examples of wound images for each wound stage. We further process images to automatically detect splints and crop images to contain only the area enclosed by the splint. This allows the model to focus on the wound region instead of attending to the presence and color of the splint. Fig. 2b shows examples of circular cropped data without the splints.



(a)



(b)

Fig. 2. (a) Cropped images of the wound region from the raw dataset. (b) Circular cropped images of the wound region with the splints removed to eliminate interference.

3 Models and Algorithms

Our proposed framework consists of three components. 1) a temporal image encoder to provide temporally-relevant 16-dimensional embeddings, 2) extrapolation of the future embedding of a wound using linear extrapolation from two prior wound embeddings, and 3) a conditional image generation using the extrapolated 16-dimensional embedding. The conditional image generation module uses a modified loss function to ensure temporal coherency and similarity between the sequences' images. Section 3.1 describes the two elements of the wound classifier model. Section 3.2 outlines the conditional GAN model and associated loss function, and section 3.3 explains how trajectory projections are performed.

In the first stage, the temporal image encoder serves as a backbone to initially create 16-dimensional embedding [5]. The downstream Wound Stage Classifier is used to classify the wound stages based on the latent space embedding. The temporal image encoder is lightweight and utilizes the pre-trained DenseNet-121 [13] as its backbone. The conditional GAN was trained using the classifier-generated embedding in the second stage. In the third stage, we used the trained GAN to generate wound images and feed them back to the classifier to predict the stage of the generated wound images.

3.1 Temporal Image Encoder

In [5], the authors found that wound embeddings can be clustered into four healing stages using a self-supervised temporal image encoder. The temporal image encoder consists of two encoder modules trained in a Siamese tuple configuration [14]. The encoder architecture is DenseNet-121 and is pre-trained on ImageNet [15]. The temporal image encoder takes a pair of wound images (from the same wound). The two embeddings from DenseNet-121 are further reduced to size 16 and concatenated before being sent to a classifier layer where the pair of images are classified as positive temporal order (forward in time) or negative temporal order (backward in time). We use the same architecture as [5] to generate latent embedding of size 16 for the input conditions of our GAN model.

3.2 Conditional GAN Model

The GAN architecture consists of two components: the Generator $\mathcal{G}$ that is trained to generate realistic wound images, and the Discriminator $\mathcal{D}$ that aims to distinguish real and generated data. We use condition c to train a conditional GAN, with c being either 4-dimensional class labels or the 16-dimensional embedding of a wound image. For simplicity, we define the aggregated loss $\mathcal{L} = \mathcal{L}_{\mathcal{G}} + \mathcal{L}_{\mathcal{D}}$. We train the model in a GAN fashion meaning that the $\mathcal{L}_{\mathcal{G}}$ and $\mathcal{L}_{\mathcal{D}}$ each only affect the generator and the discriminator parameters, respectively. Fig. 3 illustrated the outline of our wound generation framework.

The loss function for the Generator $\mathcal{G}$ and Discriminator $\mathcal{D}$ are as follows.

$$\begin{aligned} \mathcal{L}_{\mathcal{D}} &= \mathcal{L}_R(\hat{x}_k) + \mathcal{L}(\hat{x}_k|x_i, x_j) \\ \mathcal{L}_{\mathcal{G}} &= \mathcal{L}_{z|c}(\hat{x}_k) - [\mathcal{L}_R(\hat{x}_k) + \mathcal{L}(\hat{x}_k|x_i, x_j)] \end{aligned} \tag{1}$$

- $\mathcal{L}_R(\hat{x})$ reflects how realistic the generated images $\hat{x}$ are compared to real wound images. The loss is measured in Binary Cross Entropy (BCE).
- $\mathcal{L}_{z|c}(\hat{x})$ measures how well the generated $\hat{x}$ adheres to the *desired* condition c (class labels or embedding). We use the temporal image encoder to obtain the 16 dimensional embedding of the generated image, z. The Mean Squared Error (MSE) loss is used to calculate the distance between the desired embedding c and the embedding of generated sample z.
- $\mathcal{L}(\hat{x}_k|x_i, x_j) \mid i < j < k$ measures if $\hat{x}$ at day k appears coherent with respect to the ground truth wound images from day i and day j. This loss is used to enforce temporal coherence of the generated image with the wound images in the preceding days. A combination of a pairwise Siamese Convolutional Neural Network (CNN) and feed-forward layers are used to identify whether a generated image is coherent with the real images from preceding days. To train such discriminator, we generate positive pairs (Day i, Day k real) or (Day j, Day k real) and negative pairs (Day i, Day k generated) or (Day j, Day k generated). By doing so, the model can enforce the temporal validity (coherence) of image pairs and thus encode temporal dynamics. The temporal coherency loss is measured in BCE.

3.3 Trajectory Prediction through Extrapolation

The final step is to generate future wound images and predict their stage based on existing trajectories. To generate a future image, we first take two embeddings of day i and j and linearly extrapolate these embeddings to obtain a requested embedding of day k (see Eq. 2). The extrapolated embedding for day k is then sent to the conditional GAN to generate a wound image (see Fig. 3 top row). Finally, the generated image will be sent to the wound stage classifier.

$$z_k = z_i + \frac{z_j - z_i}{j - i} \times (k - i), \quad i < j < k \tag{2}$$

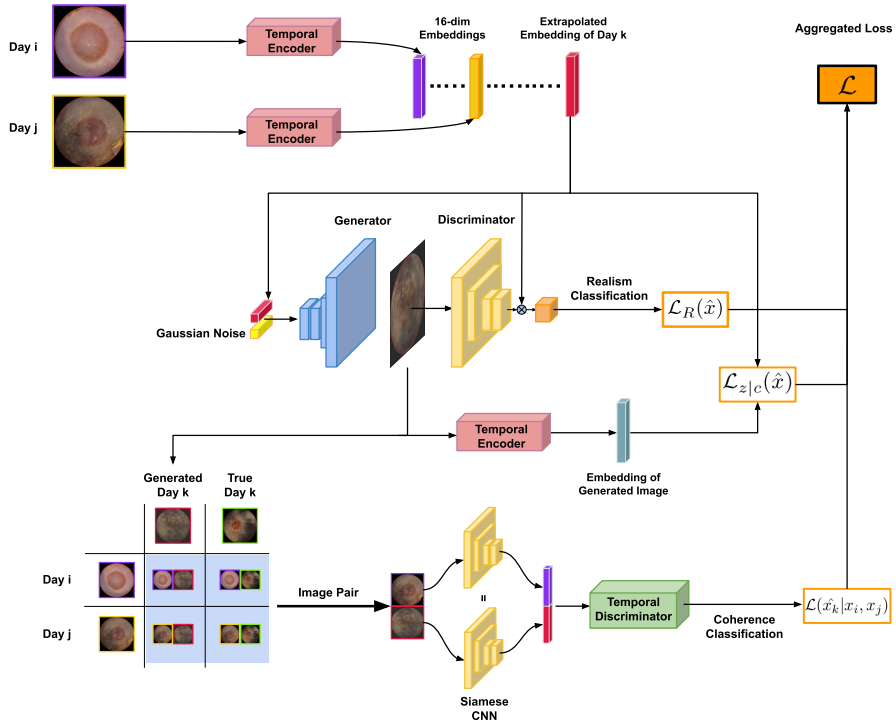


Fig. 3. The diagram of our model. The pink module is the temporal image encoder for generating latent embeddings. The blue module is the Generator $\mathcal{G}$, and the yellow module is the Discriminator $\mathcal{D}$ for classifying real and fake wound images. Finally, the green module is the *Temporal Discriminator* that identifies whether the generated image belongs to the wound images in preceding days. For the loss terms in orange boxes, $\mathcal{L}_R$ represents realism, $\mathcal{L}_{z|c}$ is the conditional loss for the given condition c, and $\mathcal{L}(\hat{x}_k|x_i, x_j)$ is the loss for temporal coherence.

4 Results and Discussion

This section outlines quantitative and qualitative results from wound generator and extrapolation modules. Sections 4.1 and 4.2 outline four incremental stages of designing the wound extrapolation framework. **Stage 1** is when the model is initially trained using 4-dimensional class labels as the condition to the GAN. In **stage 2**, richer conditions were provided to the generator, and hence the 4-dimensional class labels are subsequently replaced by 16-dimensional embeddings generated from temporal image encoder. **Stage 3**, illustrates the impact of including the temporal coherency loss, $\mathcal{L}(\hat{x}_k|x_i, x_j) \mid i < j < k$, in improving the quality of the generated wound images. Finally, **stage 4**, discusses the quality of extrapolated wound images, and associated agreement percentage between ground-truth wound stage and the predicted stage for the extrapolated wound image (Section 4.2).

4.1 Image Generation Results

When 4-class conditions (wound stage probability distribution) are used to train the model, the network is trained for 500 epochs and a learning rate of 3×10^{-4}. $\mathcal{L}_R$ is calculated with BCE loss, and $\mathcal{L}_{z|c}$ is calculated with MSE loss, and temporal coherence loss is not incorporated. Fig 4 shows a sample of images generated using 4-dimensional class labels. Each image label is synthesized using a random mean and a random standard deviation within custom ranges, so that the label mimic human uncertainty.

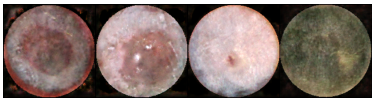


Fig. 4. The images generated using synthesized 4-dimensional class labels. The columns are arranged in the order of Hemostasis, Inflammation, Proliferation, and Maturation.

In the second stage, using 16-dimensional embedding, the network is trained with identical parameters as before: 500 epochs and a learning rate of 3×10^{-4}. The images shown in Fig. 5 are generated with 16-dimensional embeddings, without the temporal coherency loss ($\mathcal{L}(\hat{x}_k|x_i, x_j)$).

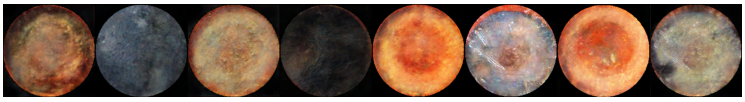


Fig. 5. Sample wound images generated with 16-dimensional latent embeddings. The columns are not ordered based on the wound stage due to the nature of the embeddings. There were more visual details in mice fur and wound color.

Finally, the temporal coherency loss, $\mathcal{L}(\hat{x}_k|x_i, x_j)$, is added to see how the quality of the generated images will change through enforcing temporal coherency between the generated image and the real images in the preceding days. We trained the network for 80 epochs and set the learning rate to 5×10^{-4}. BCE is used to compute temporal coherency loss. Fig. 6 shows a sample of generated wounds for a given mouse in our dataset, and the associated generated wounds over a 14-day period using all loss terms.

4.2 Quality and Accuracy of Extrapolated Wound Images

Extrapolations are done based on a pair of images for days i and j with $i < j$. The extrapolated embedding for day k with $k > j$ is done through linear extrapolation, which then is passed onto the conditional GAN for wound image

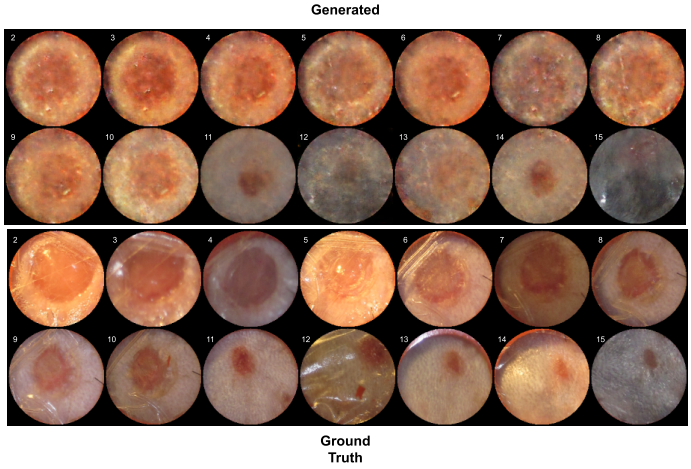


Fig. 6. Generated vs. real wound images for a mouse in the test set from day 2 to day 15. Top: generated; bottom: ground truth. The images are ordered by the number of days healed from left to right and top to bottom.

generation. Fig. 7 illustrates examples from our test set to show how the framework is able to generate realistic-looking wound images that also demonstrate temporal coherency between the wound images on days i, j, and the generated wound image for day k.

To quantify the capability of the proposed framework in predicting the future wound stage, we trained a classifier on top of the temporal image encoder. We add a dense layer of size 8 followed by the classification layer of dimension 4. The wound classifier is trained using the labels from non-expert annotators. The training of the classifier is done with 120 epochs, a learning rate of 4×10^{-5}, weight decay of 0.02, utilizing Adam [16] optimizer, and Cross Entropy (CE) as loss function. Our dataset is split using two mice for validation and testing and the other six for training. The accuracy of the classifier is 90.6%, 84.4%, and 62.5% for training, validation, and test sets.

We demonstrate the agreement percentage between the ground-truth wound stage for a given day k and the predicted stage for the extrapolated wound image of day k. To look into the impact of the the difference in days between i and j as well as j and k, we break down the agreement assessments into 2-day intervals. Table 1 contains the agreement percentages for two wounds in our test set. As outlined, the framework is able to predict the correct wound stage in future through observing two images in the past more accurately when the delta in days between j and k is small. This is also expected due to the limitation of linear extrapolation approach. Future work can explore the accurate extrapolation order given the current information about the wound and hence use higher order extrapolation function to achieve a higher agreement percentage.

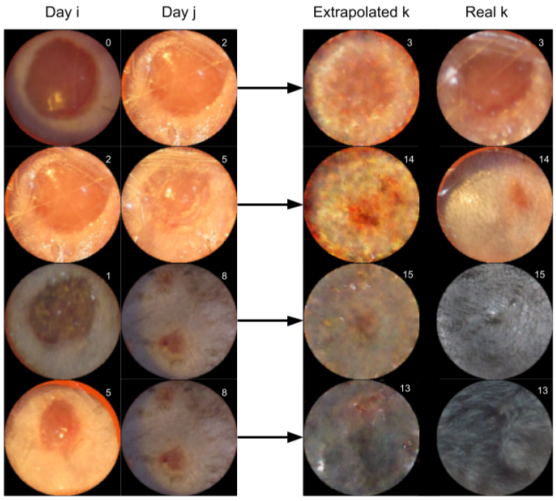


Fig. 7. Examples of generated wound images on days i, j, k, and the associated generated image on day k.

Table 1. Results for the extrapolated generation for two test wounds (one young and one aged mice). The results are split up into sections based on i, j and k, where i is the first day, j is the second day, and k is the future day used for extrapolation.

Young-ID4-Left	$k-j \leq 2$	$2 < k-j \leq 4$	$4 < k-j \leq 6$	$6 < k-j$
$j-i \leq 2$	57.8%	59.1%	63.9%	36%
$2 < j-i \leq 4$	65.9%	50%	32.1%	47.2%
$4 < j-i \leq 6$	77.8%	64.3%	55%	31.3%
$6 < j-i$	84.4%	72.2%	43.8%	50%
Aged-ID1-Right	$k-j \leq 2$	$2 < k-j \leq 4$	$4 < k-j \leq 6$	$6 < k-j$
$j-i \leq 2$	52%	25%	11.1%	26.6%
$2 < j-i \leq 4$	68.2%	50%	46.4%	25%
$4 < j-i \leq 6$	75%	60.7%	55%	50%
$6 < j-i$	68.8%	55.6%	62.5%	50%

5 Conclusion

We propose a novel framework for predicting the healing trajectory of mice wounds through linear extrapolation of the latent embedding and a conditional generative adversarial network. Our proposed model utilizes the potential of rich latent embedding and a temporal coherency loss to generate future wound images that are realistic, represent the corresponding healing stage, and provide coherency for the appearance of the wound throughout the healing process. The proposed framework is shown to provide accurate predictions for the wound stage in the future. The approach can potentially improve the quality of wound care as proper interventions can be identified ahead of time.

References

1. DM Anisuzzaman, Yash Patel, Jeffrey Niezgoda, Sandeep Gopalakrishnan, and Zeyun Yu. Wound severity classification using deep neural network. *arXiv preprint arXiv:2204.07942*, 2022.
2. DM Anisuzzaman, Chuanbo Wang, Behrouz Rostami, Sandeep Gopalakrishnan, Jeffrey Niezgoda, and Zeyun Yu. Image-based artificial intelligence in wound assessment: A systematic review. *Advances in Wound Care*, 2021.
3. Lei Wang, Peder C Pedersen, Emmanuel Agu, Diane M Strong, and Bengisu Tulu. Area determination of diabetic foot ulcer images using a cascaded two-stage svm-based classification. *IEEE Transactions on Biomedical Engineering*, 64(9):2098–2109, 2016.
4. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
5. Héctor Carrión, Mohammad Jafari, Hsin-Ya Yang, Roslyn Rivkah, Marco Rolandi, Marcella Gomez, and Narges Norouzi. Healnet – self-supervised acute wound healstage classification, 2022.
6. Amir Mazaheri and Mubarak Shah. Video generation from text employing latent path construction for temporal modeling. *arXiv preprint arXiv:2107.13766*, 2021.
7. Ivan Skorokhodov, Savva Ignatyev, and Mohamed Elhoseiny. Adversarial generation of continuous images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10753–10764, 2021.
8. Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1316–1324, 2018.
9. Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
10. Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
11. Hsin ya Yang, Michelle Bagood, Héctor Carrión, and Rivkah Isseroff. Photographs of 15-day wound closure progress in c57bl/6j mice, 2022.
12. Michelle D Bagood, Anthony C Gallegos, Andrea I Medina Lopez, Vincent X Pham, Daniel J Yoon, Daniel R Fregoso, Hsin-ya Yang, William J Murphy, and R Rivkah Isseroff. Re-examining the paradigm of impaired healing in the aged murine excision wound model. *Journal of Investigative Dermatology*, 141(4):1071–1075, 2021.
13. Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
14. Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, page 0. Lille, 2015.
15. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
16. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.