scientific reports

OPEN

Generative adversarial networks based skin lesion segmentation

Shubham Innani[1]✉, Prasad Dutande[1], Ujjwal Baid[1,2], Venu Pokuri[3], Spyridon Bakas[2], Sanjay Talbar[1], Bhakti Baheti[1,2,4] & Sharath Chandra Guntuku[3,4]

Skin cancer is a serious condition that requires accurate diagnosis and treatment. One way to assist clinicians in this task is using computer-aided diagnosis tools that automatically segment skin lesions from dermoscopic images. We propose a novel adversarial learning-based framework called Efficient-GAN (EGAN) that uses an unsupervised generative network to generate accurate lesion masks. It consists of a generator module with a top-down squeeze excitation-based compound scaled path, an asymmetric lateral connection-based bottom-up path, and a discriminator module that distinguishes between original and synthetic masks. A morphology-based smoothing loss is also implemented to encourage the network to create smooth semantic boundaries of lesions. The framework is evaluated on the International Skin Imaging Collaboration Lesion Dataset. It outperforms the current state-of-the-art skin lesion segmentation approaches with a Dice coefficient, Jaccard similarity, and accuracy of 90.1%, 83.6%, and 94.5%, respectively. We also design a lightweight segmentation framework called Mobile-GAN (MGAN) that achieves comparable performance as EGAN but with an order of magnitude lower number of training parameters, thus resulting in faster inference times for low compute resource settings.

Skin cancer results in approximately 91,000 deaths annually[1]. Early detection and regular monitoring are crucial in improving the quality of diagnosis, ensuring accurate treatment planning, and reducing skin cancer mortality rates[2]. A common detection method involves a dermatologist examining skin images to identify ambiguous clinical patterns of lesions that are often not visible to the naked eye. Dermoscopy, a widely used technique, helps dermatologists differentiate between malignant and benign lesions by eliminating surface reflections on the skin, thereby improving the accuracy of skin cancer diagnosis[3].

Skin lesion segmentation, a method to differentiate foreground lesions from the background, has received a lot of attention for over a decade due to its high clinical applicability. Computer-aided diagnostic algorithms for automated skin lesion segmentation could aid clinicians in precise treatment and diagnosis, strategic planning, and cost reduction. However, automated skin lesion segmentation is challenging due to several factors[7] such as (1) large variance in shape, texture, color, geographical conditions, and fuzzy boundaries, (2) the presence of artifacts such as hair and blood vessels, and (3) poor contrast between background skin and cancer lesions in addition to artifacts from image acquisition, as shown in Fig. 1.

Prior work. Pixel-level skin lesion segmentation algorithms can be divided into approaches built upon (a) classical image processing and (b) deep learning-based architectures. Deep learning-based methods can be further classified into Convolutional Neural Networks (CNN) and Adversarial Learning-based Generative Networks (GAN) based on the network topology. A brief review of a few prior works in these categories is presented in Table 1. The performance of classical image processing approaches heavily depends on post-processing, such as thresholding, clustering, and hole filling, tuning hyperparameters, and manual feature selection. Manually tuning these parameters can be expensive and could result in poor generalizability. Lately, deep learning-based approaches have surpassed several classical image processing-based approaches, mainly due to the wide availability of large labeled datasets and compute resources. Deep convolutional neural networks (DCNN) based methods gained a lot of popularity for skin lesion segmentation prior to the introduction of Transformer and GAN-based approaches in the field of medical imaging[23–27].

[1]Center of Excellence in Signal and Image Processing, Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, Maharashtra, India. [2]Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, USA. [3]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, USA. [4]These authors jointly supervised this work: Bhakti Baheti and Sharath Chandra Guntuku. ✉email: shubham.innani@gmail.com
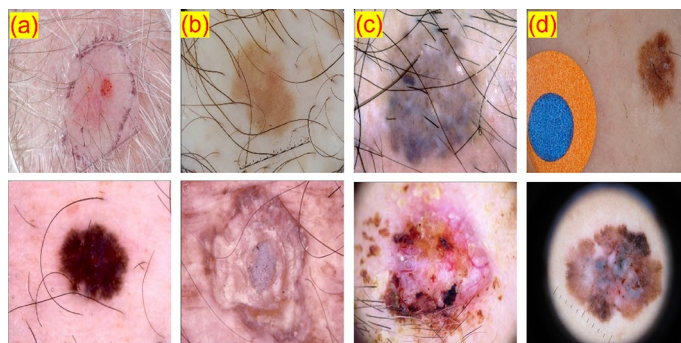
1

Figure 1. Challenges in skin lesion segmentation using dermoscopic images. First row: (**a**) minor variation in the lesion and skin color, (**b**) low contrast between wound and skin, (**c**) occlusion in lesions due to hair, and (**d**) artifacts from image acquisition. Second row: a few examples from the ISIC Lesion dataset[4] used in this paper.

Model name/citation	One phrase description	Architecture
Saliency Maps[5]	Segmentation based on Supervised Saliency Maps	Classical
UNet Segmentation[6]	Stochastic weight averaging using UNet	CNN
Deep CNN[7]	Full Resolution Networks	CNN
BLA-Net[8]	deformable convolution ResNet34 with auxiliary boundary learning network	CNN
ERU[9]	EfficientNetB4 with UNet based encoder-decoder	CNN
AS-Net[10]	Combines spatial and channel attention for learning	CNN
Attention Network[11]	Attention mechanism with high resolution features	CNN
SEACU-Net[12]	Squeeze and Excitation based Attentive ConvLSTM	CNN
Conditional Random Fields[13]	Deep Learning Approach with Pre and Post Processing	CNN
FAT-Net[14]	Feature Adaptive Transformers	Transformer
DFE-Net[15]	CNN and Transformer based Feature extraction	Transformer
SLT-Net[16]	CSwin Transformer replaced Conv module in UNet	Transformer
cGAN[17]	Conditional Generative Adversarial Network	GAN
Generative Network[18]	Decisive Generator for skin lesion segmentation	GAN
DCGAN[19]	Generating Synthetic Skin Images	GAN
FCA-Net[20]	Factorised channel attention and multi-scale features	GAN
DAGAN[21]	Deep Neural Network with generative Networks	GAN
UNet-SCDC GAN[21]	Leveraging power of discriminators	GAN
SLS-Net[22]	Lightweight device model with GAN	GAN

Table 1. Related work on skin lesion segmentation with CNN and GAN-based approaches.

The success of prior DCNN-based approaches in skin lesion segmentation is primarily based on supervised methods that rely on large labeled datasets to extract features related to the image's spatial characteristics and deep semantic maps. However, gathering a large dataset with finely annotated images is time-consuming and expensive. To address this challenge, Goodfellow *et al.*[28] introduced Generative Adversarial Networks (GANs), which have gained popularity in various applications, including medical image synthesis, due to the lack of widely available finely annotated data. Several recent and relevant GAN-based approaches in skin lesion analysis from the literature are listed in Table 1. Unsupervised learning-based algorithms that can handle large datasets with precision and high performance without requiring ground truth labels carry significant promise in addressing real-world problems such as computer-aided medical image analysis.

In our work, we address the challenges of skin lesion segmentation by utilizing generative adversarial networks (GANs)[28], which can generate accurate segmentation masks with minimal or no supervision. GANs work by training a generator and discriminator to compete against each other, where the generator tries to create realistic images, and the discriminator tries to differentiate between real and generated images (Fig. 2). However, designing an effective GAN for segmentation takes considerable time, as the performance is highly dependent on the architecture and choice of the loss function. Our study aims to optimize all three components (generator, discriminator, and loss function) for better segmentation results. The choice of the loss function is critical for the success of any deep learning architecture, and our approach takes this into account[29].

Proposed work. We propose two GAN frameworks for skin lesion segmentation. The first is Efficient-GAN (EGAN), which focuses on precision and learns in an unsupervised manner, making it data-efficient. It uses an
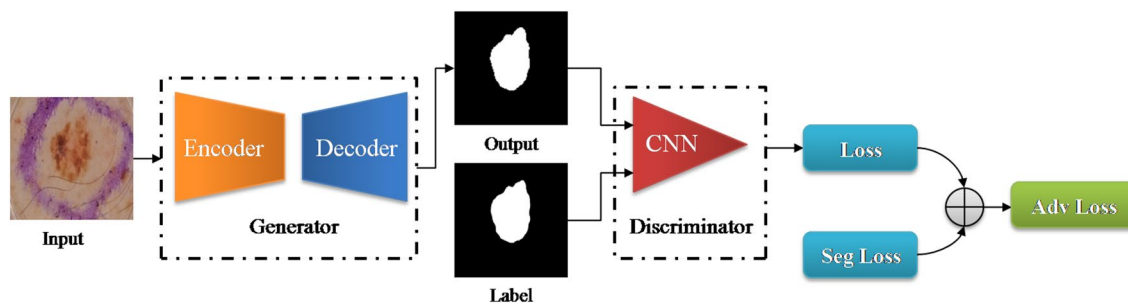
Figure 2. Flowchart of the proposed framework. The generator module is an encoder-decoder network. The discriminator classifies the segmentation result as real or fake.

encoder-decoder-based generator, patchGAN[30] based discriminator and smoothing-based loss function. The generator architecture uses a squeeze and excitation-based compound scaled encoder and a lateral connection-based asymmetric decoder. This architecture captures dense features to generate fine-grained segmentation maps, and the discriminator distinguishes between synthetic and original labels. We also implement a morphological-based smoothing loss function to capture fuzzy boundaries more effectively.

Although deep learning methods provide high precision for lesion segmentation, they are computationally expensive, making them impractical for real-world applications with limited resources like dermatoscopy machines. This presents a challenge in contexts where high-resource devices are unavailable to dermatologists. To address this issue, various devices like MoleScope II, DermLite, and HandyScope have been developed for lesion analysis and support low computational resources. These devices use a special lens with a smartphone. To create a more practical model for such real-time applications, we propose Mobile-GAN (MGAN), which is a lightweight unsupervised model consisting of an Inverted Residual block[31] with Atrous Spatial Pyramid Pooling[32]. With this model, we aim to achieve good segmentation performance in terms of the Jaccard score with lower resource strain. With only 2.2M parameters (as opposed to 27M parameters in EGAN), the model can run at 13 frames per second, increasing the potential impact of computer vision-based approaches in day-to-day clinical practice.

Results

Performance of CNN-based models.
We implemented and analyzed the results of several CNN and GAN-based approaches for this task. Table 2 summarizes the evaluation of CNN and GAN-based approaches on the unseen test dataset. We started with one of the most popular architectures in medical imaging segmentation-UNet[33]. Since this architecture is a simple stack of convolutional layers, the original UNet provided a baseline performance on ISIC 2018 dataset. We strategically conducted several experiments using deeper encoders like

Approach	Network	Dice coefficient	Accuracy	Jaccard Index	Sensitivity	Specificity
UNet*[33]	CNN	71.53	80.7	60.58	85.8	87.8
DeepLabV3+*[37]	CNN	76.3	91.4	77.3	87.8	88.1
FPN*[38]	CNN	84.46	92.4	73.76	88.2	87.8
Mobile-UNet*[39]	CNN	83.9	91.2	71.32	90	94.1
Res-UNet*[39]	CNN	86.3	90.8	76.77	90.9	93.7
Eff-UNet*[40]	CNN	89.56	92.2	81.42	90.7	92.6
CPFNet[36]	CNN	89.89	96.3	82.86	89.53	96.55
ERU[9]	CNN	88.12	94.35	80.56	90.32	96.92
FAT-Net[14]	Transformers	89.03	96.99	82	91	95.3
SEACU-Net[12]	CNN	87.58	93.60	78.12	–	–
SLT-Net[16]	Transformer	82.85	–	71.51	–	–
AS-Net[10]	CNN	88.07	94.66	80.51	89.92	95.72
cGAN[21]	GAN	83.8	90.2	74.8	89.2	92.9
DCGAN[19]	GAN	85.8	91.9	74.4	89.1	95.4
FCA-Net[20]	GAN	88.8	93.8	77.2	94.7	92.1
DAGAN[21]	GAN	88.5	92.9	82.5	95.3	91.1
MGAN (Ours)	GAN	88.3	93.4	75.0	93.8	92.1
EGAN (Ours)	GAN	90.1	94.5	83.6	93.6	95.5

Table 2. Results of CNN and GAN-based approaches including our proposed algorithms (MGAN and EGAN) on the ISIC 2018 test dataset. *Indicates the model was re-trained using the authors' source code. '–' indicates metrics not being reported.

ResNet, MobileNet, EfficientNet, and asymmetric decoders (described in the Methods section). The concatenation of low-level features is skewed, based on the number of output feature maps, rather than linking each block from the encoder, like in traditional UNet. Adding a batch normalization layer after each convolutional layer also helped achieve better performance. For detailed evaluation with CNN-based methods, we also experiment with DeepLabV3+[32] and Feature Pyramid Network (FPN)[34] decoders in combination with various encoders as described above, and the modification led to improved performance. These results on the ISIC 2018 test set from our experimentation, i.e., us running the authors' code to train the proposed models, are listed with ∗ in Table 2.

Performance of GAN-based models. Table 2 also lists several results from recent literature on this dataset for comparison completeness. Models trained by us are submitted to the evaluation server for a fair evaluation. We then compare the results of various GAN-based approaches, as shown in Table 2. We observe that a well-designed generative adversarial network (GAN) improves performance compared to techniques based on CNNs for medical image segmentation. This demonstrates GANs ability to overcome the main challenge in this domain of not having large labeled training data. Our proposed EGAN approach outperforms all other approaches in terms of Dice coefficient. A few works[8,9,35,36] report better performance compared to our results. But these works created and used an independent test split from ISIC training data and did not use the actual ISIC test data.

Performance of lightweight models. We designed a lightweight generator model called MGAN, based on DeepLabV3+ and MobileNetV2, which achieves results comparable to our EGAN model in terms of Dice Coefficient with significantly fewer parameters and faster inference times. Table 3 compares various mobile architectures based on the Jaccard Index, the number of parameters in a million, and inference speed on the test dataset for a patch size of 512×512. As we can see from the table 3, MGAN has 2.2M parameters providing the Inference Speed of 13FPS. Even though SLSNet reports a higher performance in terms of the Jaccard Index, this result is evaluated on the independent validation test set.

Visualization of the learned representations. One of the criticisms of deep neural networks, which can make valuable and skillful predictions, is that they are generally opaque, i.e., it is unclear how or why a particular prediction or decision is made. To address concerns about the opacity of deep neural networks, we utilized the internal structures of convolutional neural networks that work on 2D image data to investigate the representations learned by our unsupervised model. Figure 4 displays the segmentation results for visual interpretation. The proposed GAN framework also demonstrates better segmentation performance regardless of non-skin objects or artifacts in the image. We assessed and visualized the 2D filter weights of the model to explore the features learned by the model. Additionally, we investigated the activation layers of the model to understand precisely which features the model recognized for a given input image, and we visualized the results in Fig. 3. We selected the output of seven blocks of the encoder (Block1-Block7) and four output feature maps from the decoder (D1-D4) for visualization, as the model has numerous convolutional layers in each architecture block.

Method	Jaccard Index	Parameters (M)	Inference (FPS)
ENet[41]	72.7	0.36	27
SLSNet[22]	78.40	2.35	16
GAN FCN[42]	77.80	10.2	8
MGAN	**75**	**2.2**	**13**

Table 3. Comparison of various Mobile networks at the task of skin lesion segmentation on the ISIC 2018 dataset. The inference column indicates the Frames per Second (FPS) on original Images with a patch size of 512×512. Values of the proposed model MGAN are in bold.
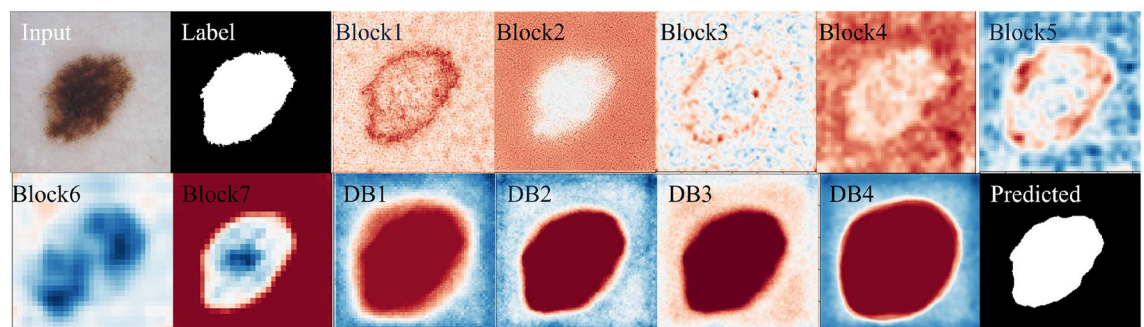

Figure 3. Visualization of the learned feature maps of the proposed EGAN architecture.
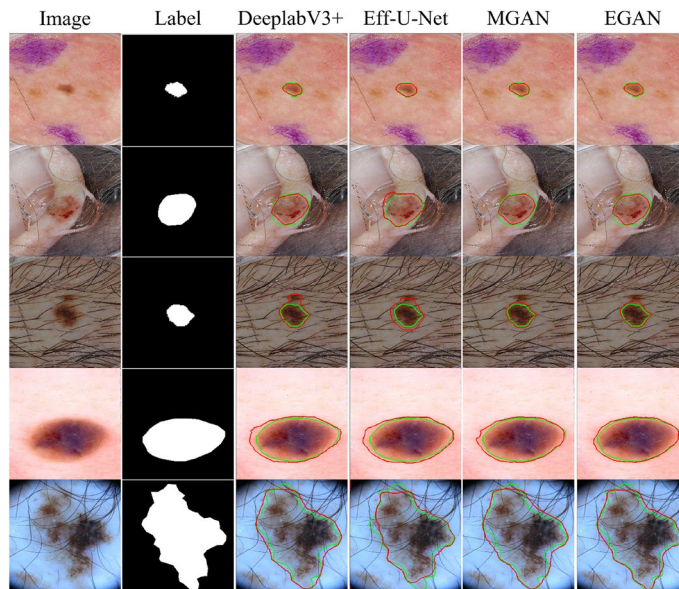
Figure 4. Comparison of the segmentation by various CNN and GAN-based approaches. Each column serially depicts the input image, label, output of various CNN-based approaches, and output of proposed MGAN and EGAN. Ground truth and segmented lesions are marked with green and red curves respectively.

Discussion

This paper has three main findings. First, we proposed a novel unsupervised adversarial learning-based framework (EGAN) based on Generative Adversarial Networks(GANs) to segment skin lesions in a fine-grained manner accurately. In data-scarce applications such as skin lesion segmentation, the success of GANs relies on the quality of the generator, discriminator, and loss function used. One of the main challenges in the field of medical imaging is the availability of large annotated data, collecting which is a tedious, consuming, and costly task. To address the data-efficiency challenge, we trained our model unsupervised, allowing the generator module to capture features effectively and segment the lesion without supervision. Our patchGAN-based discriminator penalized the adversarial network by differentiating between labels and predictions. As we do not backpropagate the error during training in the discriminator, no such advancement is needed as PatchGAN-based architecture is powerful to classify between real and fake. In skin lesion segmentation, capturing contextual information around the segmentation boundary is crucial for improving performance[8]. To address this, we implemented the morphological-based smoothing loss to capture fuzzy lesion boundaries, resulting in a highly discriminative GAN that considers contextual information and segmented boundaries. The performance-exclusive EGAN approach outperforms prior works achieving improved performance with a dice coefficient of 90.1% on the ISIC 2018 test dataset when trained with adversarial learning and morphology-based smoothing loss function compared to using the dice loss alone, which achieved a dice coefficient of 88.4%. Our evaluation of the ISIC 2018 dataset demonstrates significantly improved performance compared to existing models in the literature. Furthermore, the proposed framework's potential can be extended to other medical imaging applications.

Second, we proposed a lightweight segmentation framework (MGAN) that achieves comparable results while being much less computationally expensive – with an order of magnitude lower number of training parameters and significantly faster inference time. The MGAN approach is suitable for real-time applications, making it a viable solution for cutting-edge deployment, for instance, in low compute resource contexts. Our proposed framework includes two generative models: EGAN and MGAN, which are designed to balance performance and efficiency. Integrating models like MGAN with dermoscopy devices has the potential to revolutionize the future of dermatology, enabling more efficient, accurate, real-time segmentation and accessible care for patients with skin lesions.

Third, our approach enables visualizing the learned representations of the model to interpret the predictions. This is especially crucial for *clinical algorithms-in-the-loop* applications such as skin lesion segmentation, where the decisions of automated segmentation methods could be considered by clinicians in the context of the features learned by the model.

Limitations: Although our model achieved promising performance on ISIC 2018 dataset, the performance could not be evaluated on other datasets. We explored different datasets such as Derm7pt[43], Diverse Dermatology Images[44], and Fitzpatrick 17k[45], among others, to assess the generalizability of the proposed approach. However, we noticed that segmentation masks were not available at the time of writing this paper. While segmentation masks were available for the PH2 dataset[46], we could not access the dataset. Deep Learning models are computationally intensive and require significant resources. EGAN model is computationally heavy for deployment in real-time clinical applications. This can limit the use in resource constrained environments or devices with limited processing capabilities. In such scenarios, models such as MGAN could be utilized.

Methods

The skin lesion GAN-based segmentation framework we propose in this work is shown in Fig. 2. The framework contains three main components: (1) the generator, which consists of an encoder to extract feature maps and a decoder to generate segmentation maps without supervision and adapt to variations in contrast and artifacts; (2) the discriminator, which distinguishes between the reference label and the segmentation output; and (3) appropriate loss functions to prevent overfitting, achieve excellent convergence, and accurately capture fuzzy lesion boundaries.

Dataset.   The proposed segmentation approach was evaluated using the ISIC 2018 dataset, a standard skin lesion analysis dataset. This dataset contains 2594 images with corresponding ground truth, of which 20% (514 images) were used for validation. The images in the dataset vary in size and aspect ratio and contain lesions with different appearances in various skin areas. Some sample images from the dataset are shown in Fig. 1. To ensure a fair evaluation, the results of the test set were uploaded to the online server of the ISIC 2018[4] portal.

Generative adversarial network.   Goodfellow *et al.*[28] first introduced Generative adversarial networks (GAN) to generate synthetic data. Labeling clinical information is a tricky and time-consuming task requiring a specialist. Several medical imaging applications lack adequately annotated data. Inspired by this, the proposed work leverages unsupervised GAN for skin lesion segmentation. To begin with the methodology, we first briefly discuss generator and discriminator concepts. An adversarial network comprises a generator (G) and a discriminator (D). The generator maps a random vector γ from source domain space α to generate the desired output in the target domain β and tries to fool the discriminator. D learns to classify whether β is real (reference ground truth) or fake (generated by (G)). The generator's distribution p_G learns over α data, input noise distribution is defined as $P_\gamma(\gamma)$, which maps data space as $(G)(\gamma; \theta_G)$, where differentiable function (G) has parameters θ_G. (D) (α) is the probability distribution of α from the data instead of p_G.

The adversarial training is represented by following equation[28] which is minmax game between G and D :

$$\min_G \max_D V(D, G) = E_{\alpha \sim P_{data}(\alpha)}[logD_{\theta_D}(\alpha)]$$
$$+ E_{\gamma \sim P_\gamma(\gamma)}[log(1 - logD_{\theta_D}(G_{\theta_G}(\gamma)))] \tag{1}$$

where V is function of Discriminator (D), Generator (G),γ is from a input noise distribution $P_\gamma(\gamma)$, true samples are from $P_{data}(\alpha)$ and θ_G are generator paramaters and θ_D are discriminator paramaters.

Segmentation framework.   Generally, segmentation frameworks consist of encoder-decoder-based architecture. The encoder module is the block for feature extraction to capture spatial information within the image. It reduces the spatial size, i.e. the dimension of the input image, and decreases feature map resolution to catch significant level features. The decoder recuperates the spatial data by upsampling the feature map extracted by layers of the encoder and providing the output segmentation map. We propose to modify the architecture design of the encoder-decoder to capture the dense feature map rather than the traditional encoder and change the decoder appropriately, as shown in Fig. 5. Including squeeze and excitation-based compound scaled encoders significantly improves efficiency in terms of results.

Design of encoder.   Advancement of CNN designs is dependent on the accessibility of infrastructure and, afterward, the scaling of the model in terms of width (*w*), depth (*d*), or resolution (*r*) of the network to accomplish further significant improvement in performance when there is an expansion in the availability of resources. Instead of doing this scaling manually and arbitrarily, Tan *et al.*[47] proposed a novel systematic and automatic scaling approach by introducing a compound coefficient. The novel technique of compound coefficient ϕ to efficiently scale the network's depth, width, and resolution with a proper arrangement of scaling factors is per the following equation:

$$w : Network\ width = \beta^\phi$$
$$d : Network\ depth = \alpha^\phi$$
$$r : Input\ Resolution = \gamma^\phi \tag{2}$$
$$satisfying\ \alpha \times \beta^2 \times \gamma^2 \approx 2$$
$$also\ \alpha \geq 1 \beta \geq 1 \gamma \geq 1$$

The encoder is built using the above equation proposed by Baheti *et al.*[40], consisting of seven building blocks. Each basic building block for this encoder model is squeezing, and excitation functions[48] with mobile inverted bottleneck convolution (MBConv), as shown in Fig. 5b. Also, swish activation is used in each encoder block, enhancing performance.

Design of decoder.   The encoder downsamples the input image to a smaller resolution and captures contextual information. A decoder block likewise called an upsampling path, comprises many convolutional layers that progressively upsample the feature map obtained from the encoder. The conventional segmentation framework like UNet[33] has symmetric encoder and decoder architectures. The proposed architecture builds upon a compound scaled squeeze & excitation-based encoder and decoder as an asymmetric network. The output features from the encoder are expanded in the decoder blocks consisting of bilinear upsampling. The low-level features
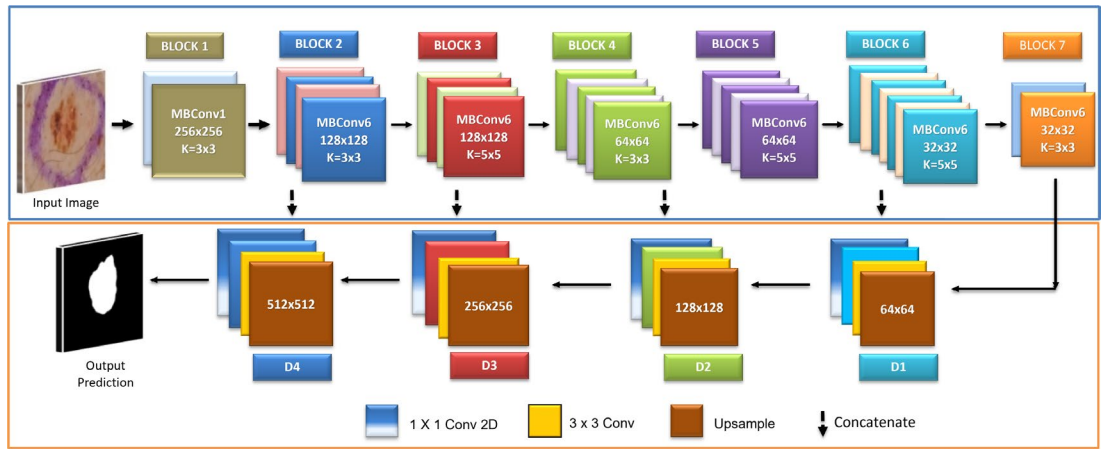
(a) Encoder has seven blocks and asymmetric concatenation to the decoder with lateral connections. Decoders made of four blocks D1-D4 of Conv2D.
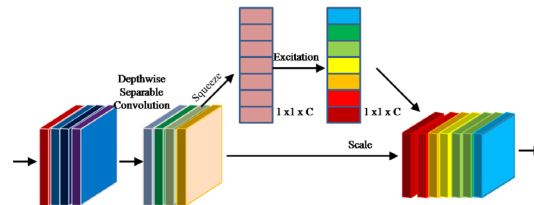
(b) MBConv Block of the Encoder

Figure 5. The architecture of the proposed generator in the EGAN architecture.

from the encoder are combined with the higher-level feature maps from the decoder of respective sizes to generate a more precise segmentation output.

Design of lightweight segmentation framework. To develop a lightweight segmentation architecture for the generator, we leverage the power of MobileNetV2[31] and DeepLabV3+[32] consisting of atrous spatial pyramid pooling module (ASPP) as shown in Fig. 6. MobileNetV2 uses depthwise separable convolution and inverted residual blocks as the basic building module, as shown in Fig. 6 above the encoder. MobileNetV2 is modified such that the output stride, i.e., the ratio of the input image to the output image, is 8. It has fewer computations and parameters and is thus suitable for real-time applications. The ASPP block has a variety of dilation rates, i.e., 1, 6, 12, and 18, to generate multi-scale feature maps and further integrate by concatenation. This feature map is upsampled and integrated with a low-level intermediate feature map from the contracting path, i.e., encoder, to generate fine-grained segmentation output. The feature extraction consisted of blocks of inverted residual
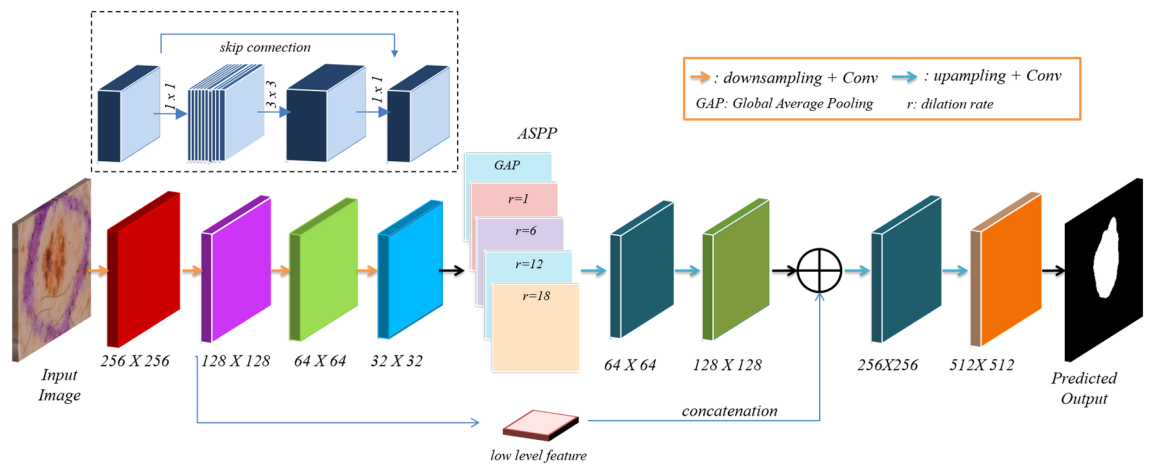

Figure 6. The architecture of the lightweight and efficient segmentation network MGAN. This architecture is based on an inverted residual network and atrous spatial pyramid pooling. The inverted residual block is shown above Encoder.

blocks, as shown in Fig. 6. The stride of the latter blocks is set as one. Images of size $512 \times 512 \times 3$ are fed as input to MGAN architecture.

Discriminator. In our architecture, we have a generator and a discriminator. The discriminator supervises the generator to produce precise masks that match the original ground truth. We have implemented a patch-GAN-based approach to achieve this, classifying each $m \times n$ mask as equivalent to the ground truth. The discriminator consists of five Conv2D layers with a kernel size of 4×4 and a stride of 2×2, with 64, 128, 256, 512, and 1 feature maps in each layer. LeakyReLU activation with an alpha value of 0.2 is used in each Conv2D layer, with the last layer using sigmoid activation. The patch-based discriminator has an output size ($m \times n$) of 16×16, where one pixel is linked to a patch of input probability maps with a size of 94×94. The discriminator classifies each patch as either fake or real. This learning strategy enforces the predicted label to be similar to the ground truth. The number of parameters is the same as proposed in patchGAN[30].

We practice the following adversarial technique for each generated label to align with the ground truth labels. A min-max two-step game alternatively renews the generator and discriminator network with adversarial learning. The discriminator function is given by:

$$L_D(x, y) = -\sum_{x,y} \gamma \, log(D(I_S)) + (1 - \gamma) log(1 - D(I_T)) \tag{3}$$

where x, y are the pixel locations of the input, $D(I_S)$ is the Discriminator function of Source Domain Images(I_S), i.e., Label Image, $D(I_T)$ is Discriminator function of Target Domain Images (I_T), i.e., Predicted Image and γ is the probability of the predicted pixel, $\gamma = 1$ when prediction is from ground truth, i.e., source domain, and $\gamma = 0$ when prediction is from generator segmented mask, i.e., target domain.

Loss function. We implement smoothing loss based on morphology to improve skin lesions segmentation and supervise the network that captures the lesion's smoothness and fuzzy boundaries. The network's loss function includes dice coefficient loss (L_{DL}) as well as the morphology-based smoothing loss (L_{SL}). The dice coefficient loss assesses the cross-over between the ground truth and prediction and is given by the condition:

$$L_{DL(\widehat{v}, v))} = 1 - \frac{2 \sum_{i \in \omega} \widehat{v}_i \cdot v_i}{\sum_{i \in \omega} \widehat{v}_i^2 + \sum_{i \in \omega} v_i^2} \tag{4}$$

where ω is the cumulative of pixels in the input image, v, and $\widehat{v}$ are the original mask and predicted mask probability map, respectively.

The morphology-based smoothing loss strengthens the network to allow smooth predictions within the nearest neighbor area[49]. It is pairwise interaction of binary labels written as:

$$L_{SL(\widehat{y}, y)} = \sum_{i \in \Omega} \sum_{j \in \mathbb{N}^i} B(i, j) \times y_i \times |\widehat{y}_i - \widehat{y}_j| \, where: B_{i,j} = \begin{cases} 1 & if \ \ y_i = y_j \\ 0 & otherwise \end{cases}$$

where $\mathbb{N}^i$ is four neighbor connection of pixels. y and $\widehat{y}$ denote the ground truth and prediction probability maps, respectively. The four connected neighbor algorithm-based smoothing loss encourages the surrounding area of pixel j with center pixel i to produce prediction probabilities similar to the original ground-truth class ($B_{i,j} = 1$).

The combined loss function is written as:

$$L_{\widehat{y}, y} = L_{DL}(\widehat{y}, y) + L_{SL}(\widehat{y}, y) \tag{5}$$

Thus, the complete framework works to optimize the loss function by training the network iteratively[49].

Data availibility
The dataset is available in ISIC archive publicly at https://challenge.isic-archive.com/data/#2018

Code availability
The source code is available at https://github.com/shubhaminnani/EGAN.

References
1. Society, C. Melanoma stats, facts, and figures. https://www.aimatmelanoma.org/about-melanoma/melanoma-stats-facts-and-figur es (2018). [Online accessed on 20-February-2020].
2. Rigel, D. S., Russak, J. & Friedman, R. The evolution of melanoma diagnosis: 25 years beyond the ABCDs. *CA Cancer J. Clin.* **60**, 301–316. https://doi.org/10.3322/caac.20074 (2010).
3. Centre, S. Dermoscopy and mole scans in perth and regional wa. https://myskincentre.com.au/service/dermoscopy/ (2018). [Online accessed on 20-February-2020].
4. Codella, N. C. F. *et al.* Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). arxiv:1902.03368 (2019).
5. Jahanifar, M., Zamani Tajeddin, N., Mohammadzadeh Asl, B. & Gooya, A. Supervised saliency map driven segmentation of lesions in dermoscopic images. *IEEE J. Biomed. Health Inf.* https://doi.org/10.1109/JBHI.2018.2839647 *(2019)*.
6. Tang, P. *et al.* Efficient skin lesion segmentation using separable-unet with stochastic weight averaging. *Comput. Methods Programs Biomed.* **178**, 289–301. https://doi.org/10.1016/j.cmpb.2019.07.005 (2019).

7. Al-masni, M. A., Al-antari, M. A., Choi, M.-T., Han, S.-M. & Kim, T.-S. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Comput. Methods Programs Biomed.* **162**, 221–231. https://doi.org/10.1016/j.cmpb.2018.05.027 (2018).

8. Feng, R., Zhuo, L., Li, X., Yin, H. & Wang, Z. Bla-net: Boundary learning assisted network for skin lesion segmentation. *Comput. Methods Programs Biomed.* **226**, 107190. https://doi.org/10.1016/j.cmpb.2022.107190 (2022).

9. Nguyen, D. K., Tran, T.-T., Nguyen, C. P. & Pham, V.-T. Skin lesion segmentation based on integrating efficientnet and residual block into u-net neural network. In *2020 5th International Conference on Green Technology and Sustainable Development (GTSD)* 366–371. https://doi.org/10.1109/GTSD50082.2020.9303084 (2020).

10. Hu, K., Lu, J., Lee, D., Xiong, D. & Chen, Z. As-net: Attention synergy network for skin lesion segmentation. *Expert Syst. Appl.* **201**, 117112. https://doi.org/10.1016/j.eswa.2022.117112 (2022).

11. Xie, F. *et al.* Skin lesion segmentation using high-resolution convolutional neural network. *Comput. Methods Programs Biomed.* **186**, 105241. https://doi.org/10.1016/j.cmpb.2019.105241 (2020).

12. Jiang, X., Jiang, J., Wang, B., Yu, J. & Wang, J. Seacu-net: Attentive convlstm u-net with squeeze-and-excitation layer for skin lesion segmentation. *Comput. Methods Programs Biomed.* https://doi.org/10.1016/j.cmpb.2022.107076 *(2022)*.

13. Ashraf, H., Waris, M., Ghafoor, M., Gilani, S. & Niazi, I. Melanoma segmentation using deep learning with test-time augmentations and conditional random fields. *Sci. Rep.* https://doi.org/10.1038/s41598-022-07885-y *(2022)*.

14. Wu, H. *et al.* Fat-net: Feature adaptive transformers for automated skin lesion segmentation. *Med. Image Anal.* **76**, 102327. https://doi.org/10.1016/j.media.2021.102327 (2022).

15. Fan, C., Yang, L., Lin, H. & Qiu, Y. Dfe-net: Dual-branch feature extraction network for enhanced segmentation in skin lesion. *Biomed. Signal Process. Control* **81**, 104423. https://doi.org/10.1016/j.bspc.2022.104423 (2023).

16. Feng, K., Ren, L., Wang, G., Wang, H. & Li, Y. Slt-net: A codec network for skin lesion segmentation. *Comput. Biol. Med.* **148**, 105942. https://doi.org/10.1016/j.compbiomed.2022.105942 (2022).

17. Mirza, M. & Osindero, S. Conditional generative adversarial nets. https://doi.org/10.48550/ARXIV.1411.1784 (2014).

18. Izadi, S., Mirikharaji, Z., Kawahara, J. & Hamarneh, G. Generative adversarial networks to segment skin lesions. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. https://doi.org/10.1109/ISBI.2018.8363712 (2018).

19. Bissoto, A., Perez, F., Valle, E. & Avila, S. Skin lesion synthesis with generative adversarial networks. In *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis* 294–302 (Springer, 2018).

20. Singh, V. *et al.* Fca-net: Adversarial learning for skin lesion segmentation based on multi-scale features and factorized channel attention. *IEEE Access* https://doi.org/10.1109/ACCESS.2019.2940418 *(2019)*.

21. Lei, B. *et al.* Skin lesion segmentation via generative adversarial networks with dual discriminators. *Med. Image Anal.* **64**, 101716. https://doi.org/10.1016/j.media.2020.101716 (2020).

22. Sarker, M. M. K. *et al.* Slsnet: Skin lesion segmentation using a lightweight generative adversarial network. *Expert Syst. Appl.* **183**, 115433. https://doi.org/10.1016/j.eswa.2021.115433 (2021).

23. Yi, X., Walia, E. & Babyn, P. Generative adversarial network in medical imaging: A review. *Med. Image Anal.* **58**, 101552 (2019).

24. Li, H., Zeng, N., Wu, P. & Clawson, K. Cov-net: A computer-aided diagnosis method for recognizing covid-19 from chest x-ray images via machine vision. *Expert Syst. Appl.* https://doi.org/10.1016/j.eswa.2022.118029 *(2022)*.

25. Yao, X., Wang, X., Wang, S.-H. & Zhang, Y.-D. A comprehensive survey on convolutional neural network in medical image analysis. *Multimed. Tools Appl.* **20**, 1–45 (2020).

26. Liu, M. *et al.* Aa-wgan: Attention augmented wasserstein generative adversarial network with application to fundus retinal vessel segmentation. *Comput. Biol. Med.* https://doi.org/10.1016/j.compbiomed.2023.106874 *(2023)*.

27. Wu, P. *et al.* Aggn: Attention-based glioma grading network with multi-scale feature extraction and multi-modal information fusion. *Comput. Biol. Med.* **152**, 106457. https://doi.org/10.1016/j.compbiomed.2022.106457 (2023).

28. Goodfellow, I. *et al.* Generative adversarial networks. *Adv. Neural Inf. Process. Syst.* **3**, 53–65 (2014).

29. Kervadec, H. *et al.* Boundary loss for highly unbalanced segmentation. In *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning*, vol. 102 of *Proceedings of Machine Learning Research* (eds Cardoso, M. J. *et al.*) 285–296 (PMLR, 2019).

30. Isola, P., Zhu, J., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 5967–5976 (2017).

31. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L. Mobilenetv2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2018.00474 (2018).

32. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Computer Vision—ECCV 2018* (eds Ferrari, V. *et al.*) 833–851 (Springer, 2018).

33. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (eds Navab, N. *et al.*) 234–241 (Springer, 2015).

34. Lin, T. *et al.* Feature pyramid networks for object detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 936–944 (2017).

35. Han, Q. *et al.* Hwa-segnet: Multi-channel skin lesion image segmentation network with hierarchical analysis and weight adjustment. *Comput. Biol. Med.* **152**, 106343. https://doi.org/10.1016/j.compbiomed.2022.106343 (2023).

36. Feng, S. *et al.* Cpfnet: Context pyramid fusion network for medical image segmentation. *IEEE Trans. Med. Imaging* **39**, 3008–3018. https://doi.org/10.1109/TMI.2020.2983721 (2020).

37. Baheti, B., Innani, S., Gajre, S. & Talbar, S. Semantic scene segmentation in unstructured environment with modified deeplabv3+. *Pattern Recogn. Lett.* **138**, 223–229. https://doi.org/10.1016/j.patrec.2020.07.029 (2020).

38. Innani, S., Dutande, P., Baheti, B., Talbar, S. & Baid, U. Fuse-pn: A novel architecture for anomaly pattern segmentation in aerial agricultural images. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 2954–2962. https://doi.org/10.1109/CVPRW53098.2021.00331 (2021).

39. Innani, S., Dutande, P., Baheti, B., Baid, U. & Talbar, S. Deep learning based novel cascaded approach for skin lesion analysis. **2301**, 06226 (2023).

40. Baheti, B., Innani, S., Gajre, S. & Talbar, S. Eff-unet: A novel architecture for semantic segmentation in unstructured environment. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 1473–1481. https://doi.org/10.1109/CVPRW50498.2020.00187 (2020).

41. Paszke, A., Chaurasia, A., Kim, S. & Culurciello, E. Enet: A deep neural network architecture for real-time semantic segmentation. https://doi.org/10.48550/ARXIV.1606.02147 (2016).

42. Bi, L., Feng, D. & Kim, J. Improving automatic skin lesion segmentation using adversarial learning based data augmentation. https://doi.org/10.48550/ARXIV.1807.08392 (2018).

43. Kawahara, J., Daneshvar, S., Argenziano, G. & Hamarneh, G. Seven-point checklist and skin lesion classification using multitask multimodal neural nets. *IEEE J. Biomed. Health Inform.* https://doi.org/10.1109/JBHI.2018.2824327 *(2019)*.

44. Daneshjou, R. *et al.* Disparities in dermatology ai performance on a diverse, curated clinical image set. *Sci. Adv.* **8**, 6147. https://doi.org/10.1126/sciadv.abq6147 (2022).

45. Groh, M. *et al.* Evaluating deep neural networks trained on clinical images in dermatology with the fitzpatrick 17k dataset. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 1820–1828. https://doi.org/10.1109/CVPRW53098.2021.00201 (IEEE Computer Society, 2021).
46. Mendonça, T., Ferreira, P. M., Marques, J. S., Marcal, A. R. S. & Rozeira, J. Ph2—A dermoscopic image database for research and benchmarking. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 5437–5440. https://doi.org/10.1109/EMBC.2013.6610779 (2013).
47. Tan, M. & Le, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. arxiv:1905.11946 (2019).
48. Hu, J., Shen, L. & Sun, G. Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 7132–7141 (2018).
49. Wang, S., Yu, L., Yang, X., Fu, C.-W. & Heng, P.-A. Patch-based output space adversarial learning for joint optic disc and cup segmentation. *IEEE Trans. Med. Imaging* **38**, 2485–2495. https://doi.org/10.1109/TMI.2019.2899910 (2019).

Acknowledgements

Author contributions

S.I. and P.D. conducted and analyzed the analyses, and S.I., B.B., and U.B. wrote the main manuscript text. V.P. helped in the preparation of the figures. B.B. and S.C.G. guided the complete work. S.T., B.B., S.B., and S.C.G. reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-39648-8.

Correspondence and requests for materials should be addressed to S.I.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.