Quest Journals
Journal of Software Engineering and Simulation
Volume 11 ~ Issue 7 (July 2025) pp: 09-12
ISSN(Online) :2321-3795 ISSN (Print):2321-3809
www.questjournals.org

Research Paper

Application of DISCOV on U-Net Detection of Skin Lesion Segmentation

Eric Yang

Abstract
Skin lesions differ from surrounding skin in color, texture, or appearance. While some are benign (e.g., freckles), others may indicate serious conditions like infections, allergies, or cancer. Current detection mainly relies on visual inspection or dermatoscopy, which depends on a physician's expertise. Genetic tests can identify mutations linked to diseases like melanoma, but visual assessment remains essential before testing.This research utilizes DImensionless Shunting COlor Vision (DISCOV), an image processing model, and trains it to recognize images of skin lesions from the International Skin Imaging Collaboration (ISIC) 2018 dataset and classify them based on visually identifiable features. This effort leads to potential improvements in convolutional neural networks (ConvNets), which specialize in examining images and processing data with a grid-like topology, allowing computers to visualize and process the data. With the same amount of epochs, DISCOV + U-Net performed relatively similarly to U-Net but given more epochs, the U-Net adapts to DISCOV features and produces a higher accuracy. There are still many improvements to be done with DISCOV, decreasing running time for example, but applying DISCOV to machine learning could lead to a positive impact on diagnostic accuracy and therefore outperform ConvNets by themselves.

***R**eceived 08 July, 2025; **R**evised 18 July, 2025; **A**ccepted 20 July, 2025 © The author(s) 2025.*
P**ublished with open access at **www.questjournas.org

DISCOV

DISCOV (DImensionless Shunting COlor Vision) model mimics the primate color vision cells (retinal ganglion, thalamic single opponent, and two classes of cortical double opponents). It aims to mimic human perception of color by processing a red-green-blue image into retinal, single-opponent, and double-opponent features that the code would use as input for image recognition. DISCOV models multiple cells, each with different processing of color information, which allows it to have a richer feature representation when processing an image with multiple colors.

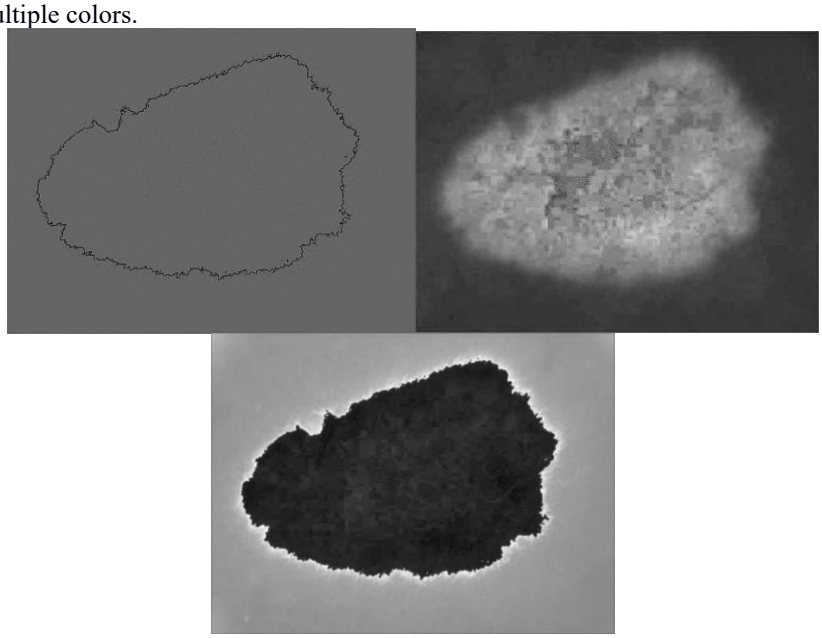

Fig 1. Examples of DISCOV output done on image ISIC_0000000 (left for retinal ganglion cells channel four, middle and right for thalamic single opponent channel one and three respectively)

Each cell DISCOV model has an ideal stimulus, with the retinal cells having the ideal stimulus of a red center (excitatory red center) surrounded by anything but red (inhibitory red surround). The thalamic single opponent cells have an ideal stimulus of an excitatory red center with an inhibitory green surround, meaning a red center with anything blue-green surrounding it. Double opponent I cells exhibit both center-surround spatial antagonism and chromatic antagonism, meaning that some surrounding colors can inhibit the center's response and specific colors can be influenced by other colors. In this case, the ideal stimulus is a red center surrounded by green. Lastly, double opponent II cells also exhibit chromatic antagonism within the center and the surround of the cells, but the surround is also broad-band suppressive, meaning the cell can inhibit on a wide scale. The ideal stimulus for this is a red center surrounded by black. From this, DISCOV can create "features" that use two contrasting colors like red-green or blue-yellow, and model each type of vision cell based on those colors.

DISCOV operates on the equation $\frac{d}{dt}z = -Bz(1-z)Cx - (1+zD)y$. Here, 'z' represents the cell's activation level, 'x' is the excitatory input from the center, and 'y' is the inhibitory input from the surrounding area. Parameters like A, B, C, and D which means the ratio of the surround to the center, passive decay rate– the rate at which 'z' decreases over time– of the cell's activation, the ratio between the strength of the excitatory to the inhibitory input, and the ratio of excitatory potential to inhibitory potential respectively, were used. The excitatory and the inhibitory potential refers to a comparison between the excitatory input relative to the inhibitory input. When both input, x and y, are 0, the activation 'z' is 0. The larger 'x' is, or the excitatory input, the closer z reaches its maximum value of 1 while the larger 'y' is, or the inhibitory input, the closer z reaches its minimum. From this, a value of 'z' can be determined using the equation $z = \frac{Cx-y}{Cx+Dy}$. This shows a ratio between the excitatory and inhibitory inputs.

In short, DISCOV uses retinal cells and how they see colors differently and process information from the excitatory and inhibitory inputs, analyzing images using a center-surround antagonism.

II. Methods And Materials

The ISIC 2018 dataset supplies both ground truth masks as well as training images, serving as the dataset used and tested for the model. There are a total of 1000 test images and 2594 ground truth masks.

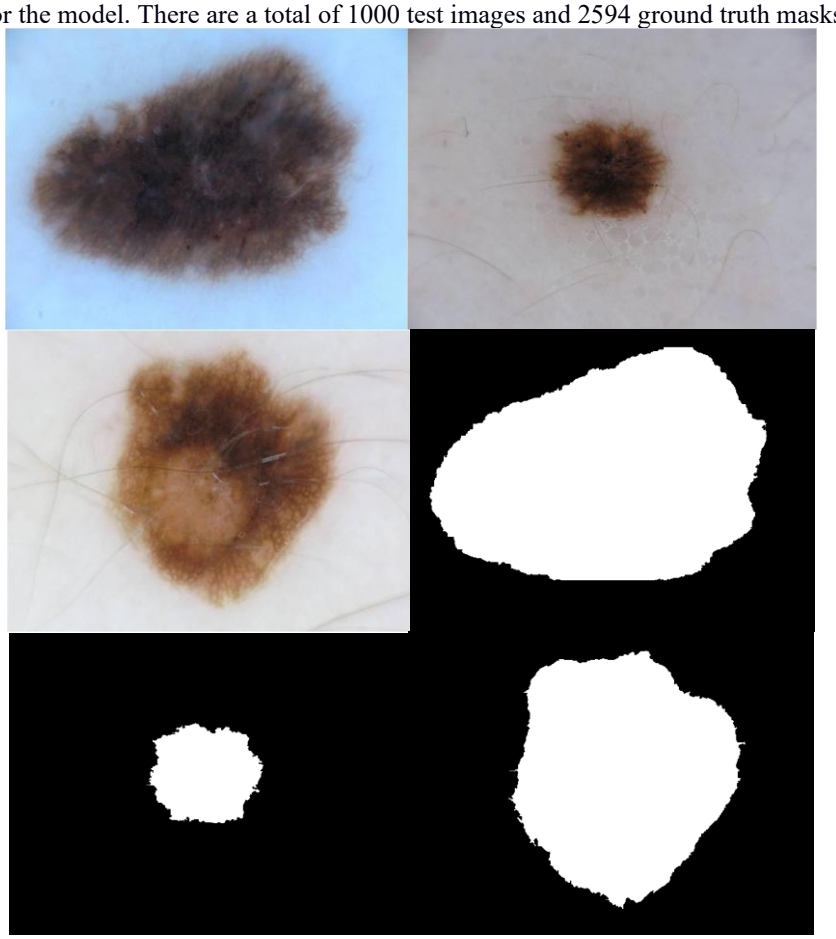

Fig.2: Images of training input, ISIC_0000000, ISIC_0000001, and ISIC_ 0000003 with their ground truth respectively

ConvNets, also known as U-Nets, operate using three main operations: encoder blocks, decoder blocks, and skip connections. The U-Net is symmetrical and follows a "U" shape.

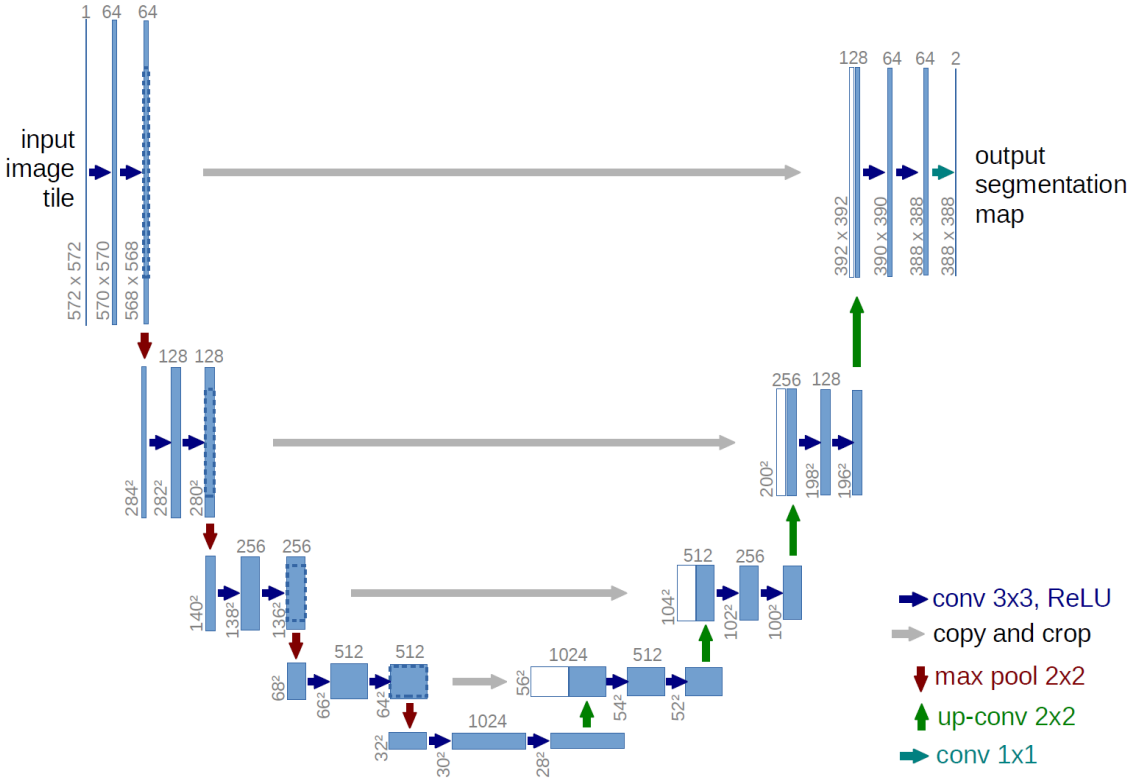

Fig.3: Image of U-Net with Encoding and decoding layers, as well as skip connections.

The process of a ConvNet starts from the encoder layer, which contains encoder blocks that use the original input image and perform convolutional operations that reduce the spatial resolutions while increasing their depths, capturing the relevant information about the image. Each encoder block performs 2 convolution layers, each followed by a rectified linear set, or ReLU for short. At the very first encoder block, the image's channel is increased to 64 while the spatial dimension should be halved. Each encoder block repeats the action described above, with the only difference being the number of channels increasing to be doubling starting from 64. There are a total of 5 encoder blocks for the encoder layer. At the last encoder block, the same operations are performed but only include one layer instead of two. After the features have been extracted, the U-Net moves on to the decoding path. Lastly, skip connections send images directly from the encoding layer to the decoding layer to reduce information loss during the ConvNet process.

The decoding layer consists of decoder blocks that perform convolutions as well as up-convolutions to combine features and upsample the image, generating a segmentation map. From the fifth block from the encoding layer, the second convolution is performed with a ReLU layer following it. The image is then up-scaled twofold and halves its channels. Using the skip connections in the sixth decoder block, the map from the encoding layer is concatenated, doubling the number of channels. The two convolutions are applied again with ReLU layers following each, reducing the number of channels by half again. Lastly, the decoder block performs the up-convolution to further reduce the amount of channels by half and doubling the spatial dimension. This process is performed at each of the decoding blocks. There are 5 decoding blocks in total. In the final decoding block, the same two convolutions and ReLU are applied but following the convolutions is another convolution block which is used to reduce the channels to the amount desired.

The data analysis used were several, including F1 (harmonic mean of precision and recall), Jaccard (measure the similarities between the output masks and provided masks), Recall (ratio between the number of objects found and the total objects), and Precision (ratio between the amount of correctly predicted masks and all of the masks).

III. Results

With the same amount of epochs, DISCOV + U-Net performed relatively similarly to U-Net but given more epochs, the U-Net adapts to DISCOV features and produces a higher accuracy.

The array "keep_chnl" is referring to the type of DISCOV images used. '1,' '2,' and '3' were used to refer to BGR(blue, green, red). In total, there are 19 images from DISCOV(RGB and each type of vision cell and channel modeled), meaning any number like "3+4+3" or "3+4" refers to the type of DISCOV vision cell and channel used. For example, "3+4+3" refers to the vision cell and channel "rgl" and "chnl2" respectively. After BGR, f1-f4(feature 1-4) represented rgl and channel 1 - 4, f5-f8 represented tso and channel 1 - 4, f9-12 represented do1 and channel 1 - 4, and f13-f16 represented do2 and channel 1 - 4 respectively.

RGB	RGB + DISCOV	RGB + DISCOV
Subset =1000, epoch = 5 **Accuracy: 0.91667** **F1: 0.83951** **Jaccard: 0.75048** **Recall: 0.86030** **Precision: 0.87490**	Subset = 1000, epoch = 5, keep_chnl = [1,2,3,3+4+3] Running time = ~45 minutes **Accuracy: 0.89327** **F1: 0.76653** **Jaccard: 0.66551** **Recall: 0.74595** **Precision: 0.89352**	Subset = 1000, epoch = 10, keep_chnl = [1,2,3,3+4+3] Running time = ~99 minutes **Accuracy: 0.92573** **F1: 0.84805** **Jaccard: 0.76160** **Recall: 0.86656** **Precision: 0.88296**

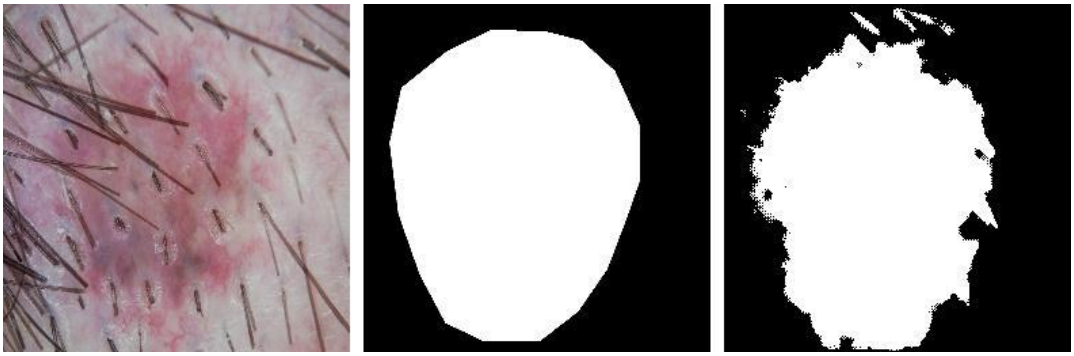

Fig4. example of DISCOV + U-NET output. Original image(ISIC_0010361) on the left, the ground truth mask from the dataset in the middle, and the predicted ground truth on the right from DISCOV

IV. Discussion

There are a few possible improvements to DISCOV + U-Net. For example, it could be hyperparameter-tuned to see higher accuracy. This could mean experimenting with different DISCOV images. During the experiment conducted, using RGB + one type of DISCOV images allowed for realistic running time, and attempting to include all DISCOV images (do1, do2, rgl, tso) is not realistic for running time. Previous attempts at running all 16 DISCOV images and RGB resulted in spending one hour on less than half an epoch. Because of that, the accuracy might be impacted because the amount of images used is less, but even so, DISCOV and U-Net combined allowed for growth in accuracy.

Another possible improvement that could be made using HSV (Hue, Saturation, Value). Along with RGB, HSV is another parameter that can improve the recognition of skin pixels in an image. Using HSV instead of RGB might yield better accuracy but future testing is required.

References

[1]. Chelian, S., & Carpenter, G.. (2005). DISCOV: A Neural Model of Colour Vision, with Applications to Image Processing and Classification.

[2]. Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, Allan Halpern: "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)", 2018; https://arxiv.org/abs/1902.03368

[3]. Onsoi, Witchuwan & Chaiyarit, Jitjira & Techasatian, Leelawadee. (2019). Common misdiagnoses and prevalence of dermatological disorders at a pediatric tertiary care center. Journal of International Medical Research. 48. 030006051987349. 10.1177/0300060519873490.

[4]. Ronneberger, O. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Uni-Freiburg.de. https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/

[5]. Zhang, J., Zhong, F., He, K., Ji, M., Li, S., & Li, C. (2023). Recent Advancements and Perspectives in the Diagnosis of Skin Diseases Using Machine Learning and Deep Learning: A Review. *Diagnostics (Basel, Switzerland)*, *13*(23), 3506. https://doi.org/10.3390/diagnostics13233506