01 Jan 2022

Chimeranet: U-Net for Hair Detection in Dermoscopic Skin Lesion Images

Norsang Lama

Reda Kasmi

Jason R. Hagerty

R. Joe Stanley
Missouri University of Science and Technology, stanleyj@mst.edu

et. al. For a complete list of authors, see https://scholarsmine.mst.edu/ele_comeng_facwork/4675

Recommended Citation

ChimeraNet: U-Net for Hair Detection in Dermoscopic Skin Lesion Images

Norsang Lama[1] · Reda Kasmi[2] · Jason R. Hagerty[3] · R. Joe Stanley[1] · Reagan Young[1] · Jessica Miinch[1] · Januka Nepal[3] · Anand Nambisan[1] · William V. Stoecker[3]

Abstract

Hair and ruler mark structures in dermoscopic images are an obstacle preventing accurate image segmentation and detection of critical network features. Recognition and removal of hairs from images can be challenging, especially for hairs that are thin, overlapping, faded, or of similar color as skin or overlaid on a textured lesion. This paper proposes a novel deep learning (DL) technique to detect hair and ruler marks in skin lesion images. Our proposed ChimeraNet is an encoder-decoder architecture that employs pretrained EfficientNet in the encoder and squeeze-and-excitation residual (SERes) structures in the decoder. We applied this approach at multiple image sizes and evaluated it using the publicly available HAM10000 (ISIC2018 Task 3) skin lesion dataset. Our test results show that the largest image size (448×448) gave the highest accuracy of 98.23 and Jaccard index of 0.65 on the HAM10000 (ISIC 2018 Task 3) skin lesion dataset, exhibiting better performance than for two well-known deep learning approaches, U-Net and ResUNet-a. We found the Dice loss function to give the best results for all measures. Further evaluated on 25 additional test images, the technique yields state-of-the-art accuracy compared to 8 previously reported classical techniques. We conclude that the proposed ChimeraNet architecture may enable improved detection of fine image structures. Further application of DL techniques to detect dermoscopy structures is warranted.

Introduction

An estimated 106,110 new cases of invasive melanoma and 101,280 in situ melanoma will be diagnosed in 2021 in the USA [1], in addition to 2 million epitheliomas [2]. Dermoscopy is an imaging modality that renders these cancers visible when they are fully curable. However, many cases of melanoma are missed by domain experts [3, 4].

Dermoscopy is a crucial tool in the early detection of melanoma, increasing the diagnostic accuracy over clinical visual inspection in the hands of experienced physicians [5–7]. Yet, dermatologists viewing dermoscopic images in recent studies have shown lower diagnostic accuracy than machine vision techniques [3, 4, 8, 9].

Pathan et al. published a recent review detailing both handcrafted and deep learning (DL) techniques for computer-aided diagnosis of skin lesions [10]. Recent studies have shown successful results for skin cancer diagnosis by fusing ensembles, in some cases handcrafted and DL techniques [11–15]. However, hair and ruler artifacts can interfere with handcrafted feature detection by mimicking the pigment network and interfering with accurate border detection; thus, detection of these structures is needed to maximize handcrafted feature detection accuracy.

Lee et al. reviewed hair detection in dermoscopic images [16]. The methods generally employ one of three types: mathematical morphology, edge detection, and matched filters. To detect hair, the same group [16] used the top-hat transform and modified Gaussian filter to enhance hair. Xie et al. [17] proposed an algorithm focused on dark hair, using a top-hat operator and an automatic threshold. Abbas et al. [18] implemented detection of both light and dark hairs using the first derivative of Gaussian followed by morphological techniques to remove non-hair noise. Nguyen et al. [19] also detected light and dark hair using a universal

✉ R. Joe Stanley
stanleyj@mst.edu

1   Missouri University of Science & Technology, Rolla, MO 65409, USA

2   University of Bejaia, Bejaia, Algeria

3   S&A Technologies, Rolla, MO 65401, USA

 Springer

matched filtering kernel, with the binary mask generated by local entropy thresholding.

Lee et al. [20] proposed Dullrazor, with dark hair identified by grayscale morphology. Subsequent operations verify long and narrow structures. Fiorese et al. [21] proposed VirtualShave, which uses top-hat filtering followed by morphological postprocessing. Koehoorn et al. [22] used a threshold-set model and a gap-detection algorithm and further postprocessing using skeletonizing. Toossi et al. [23] used Canny edge detection followed by refinement using morphologic operators. Abuzaghleh et al. [24] segment hair by a set of directional Gaussian filters. Huang et al. [25] sought to detect hairs often missed, thin hairs, and hairs in shaded backgrounds, using Gaussian matched filtering. Zhou et al. [26] used Steger's algorithm and a least-square method. Inpainting algorithms employed in these studies included interpolation, in either one or multiple directions, and a fast-marching technique.

Deep learning methods have also been applied to detect hair in skin lesion images. Attia et al. [27] used a hybrid model of convolution and recurrent layers for hair segmentation. However, due to the lack of a hair mask dataset, they trained using the weakly labeled data and tested on the simulated hair. Li et al. [28] created a new hair-mask dataset to train and test U-Net [29] based hair segmentation model. To create the hair-mask dataset, first, they applied the top-hat segmentation techniques [30] and then manually removed the over-segmented regions. Moreover, they excluded the under-segmented images from the dataset. The use of weakly labeled data in the first method and selecting only over-segmented images in the second method limit the quality of data, thereby weakening the robustness of the DL network.

The approach of this paper is as follows. First, we created a well-labeled hair mask dataset by manually annotating hair and ruler marks on the public HAM10000, also ISIC2018 Task3 lesion classification dataset [31, 32]. Many of the

reported hair removal algorithms were evaluated on small sets of images. Many of these methods find only dark hair, ignoring light hair and ruler marks. Second, we proposed a novel deep learning based hair detection method called ChimeraNet that detects light hair and ruler marks as well as dark hair in dermoscopic skin lesion images. Further, we compared the performance of our proposed method against two well-known DL approaches [29, 33] and eight conventional image processing approaches [16–18, 20–22, 25, 34].

The remainder of the paper is organized as follows. "Materials and Methods" explains the image datasets and proposed method. "Experimental Results and Discussion" presents the segmentation experiments and comparison. "Discussion" provides a discussion. "Conclusion and Future Work" gives the conclusion and possible future work.

Materials and Methods

Image Datasets

We used two datasets in this study. The first is the HAM10000 [31] dataset of Tschandl et al. a publicly available skin lesion dermoscopy dataset containing over 10,000 skin images for 7 diagnostic categories. All images are 8-bit RGB images of size 450×600, as shown in Fig. 1. These images comprised a training set for the ISIC2018 Task 3 lesion classification challenge [32]. Since the dataset lacks ground-truth hair segmentation masks, we manually drew hair masks for 1333 dermoscopy images in the dataset. The manually drawn hair masks include dark hair, white hair, and ruler marks (Fig. 2). As the width of hair and ruler marks differs within an image or from one image to other, the manual hair masks were drawn with varying widths.

The masks were evaluated and verified by a dermatologist. Some of the hair segments are very thin, in some

Fig. 1 Skin lesion dermoscopy images from the HAM10000 (ISIC2018 Task 3) dataset showing dark hair, light hair, and ruler marks
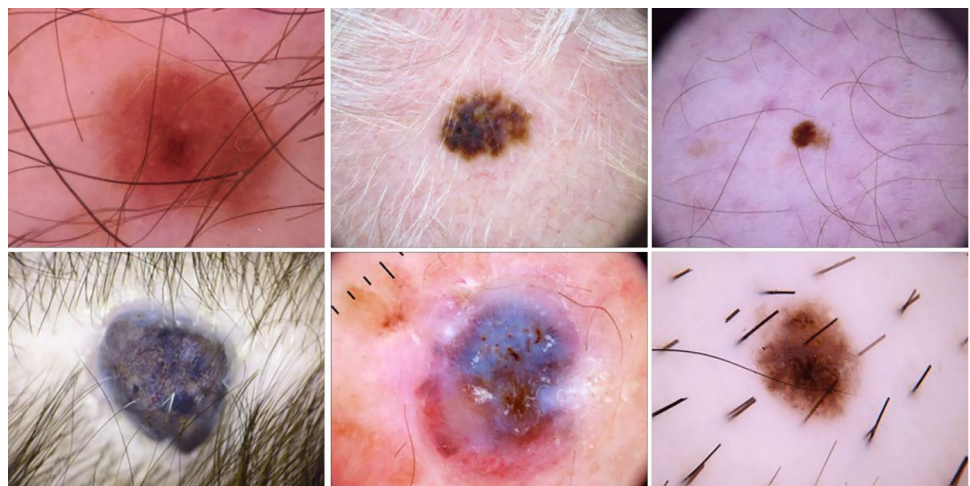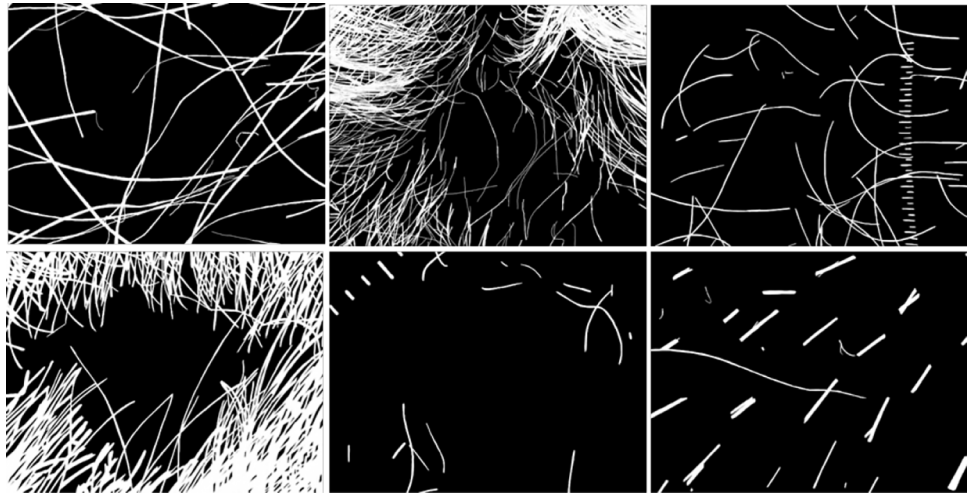
Fig. 2 Manually drawn hair masks corresponding to skin lesion images shown in Fig. 1. The hair masks include dark hair, white hair, and ruler marks

cases 1-pixel wide, so we dilated the binary masks with a 3×3 cross-shaped structuring element.

We randomly divided 1333 skin lesion images from the HAM10000 dataset (ISIC2018 Task 3 Lesion Classification dataset) into image subsets of 852 for training, 214 for validation, and 267 for testing. Both training and validation images combined with manually drawn masks were used to train the deep convolutional neural networks, and the holdout test images were used to evaluate the performance of the proposed model.

The second dataset comprised 25 dermoscopic skin lesion images with hair and calculated hair masks from [16]. The images are RGB images of size 768×1024. We resized all images into the same size as of the first dataset, i.e., 450×600, using the bilinear interpolation method. This set was small compared to the first set, so we only used as a test set in this study.

Data Augmentation

Data augmentation helps convolution neural networks to generalize better and reduce the overfitting problem by adding more training samples. It performs different image transform methods on original images to generate more examples and used them for training. In this study, we selected following image transforms for data augmentation:

- Height or width shift with a range of $(-0.15, +0.15)$
- Horizontal or vertical flip
- Rotation with range between $+90°$ and $-90°$
- Zoom with a range $(-0.15, +15)$
- Brightness with a range of $(0.9, 1)$

We performed online data augmentation to increase the number of training images by 5 times. Then, the augmented images' data range were rescaled between 0 and 1. Further, the images were normalized before feeding them to the deep network.

Proposed Network Architecture

U-Net [29] is a popular convolutional neural network architecture designed for biomedical image segmentation. It consists of encoder (left) and decoder (right) paths. The encoder path is a typical convolutional neural network (CNN) that extracts the high-level features from the input image. The feature maps are downsampled many times at different levels, thus reducing the spatial dimension of feature maps. The decoder path, which is symmetric to the encoder path, expands the low-resolution feature map to generate a segmentation map with spatial dimensions equal to those of the input. First, the decoder concatenates the feature map and high-resolution features from the encoder. Then, it upsamples the concatenated feature maps applying transpose convolution operations. The high-resolution features via skip connections from the encoder are helpful in recovering the fine-grained details in the image to generate the segmentation map.

The overall pipeline of our proposed ChimeraNet architecture is shown in Fig. 3. First, we used a pre-trained EfficientNet model as the encoder network in our model. Tanand Le [35] developed a family of CNN models, called EfficientNets, which achieved state-of-art top 1 accuracy in the ImageNet [36] image classification challenge. These models are composed of mobile inverted bottleneck convolution (MBConv) blocks. EfficientNet has seven blocks, from *Block1* to *Block7*, which are composed of multiple MBConv blocks. Based on the number of MBConv blocks in *Block1 to Block7*, they have 8 network variants from B0 to B7. EfficientNetB0 is the baseline architecture, and other variants are scaled-up versions by employing a compound scaling
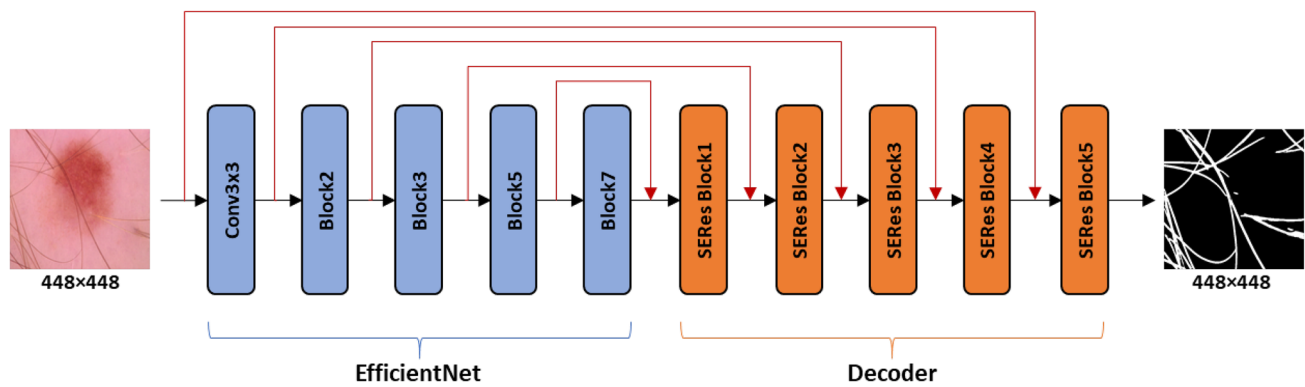
Fig. 3 Proposed ChimeraNet architecture for hair segmentation. An encoder-decoder architecture with pretrained EfficientNet model as the encoder network and the decoder network comprised of five squeeze-and-excitation residual blocks. Five skip-connections (red color) from the encoder to the decoder

method that uniformly scales network depth, width, and resolution with a fixed set of scaling factors. In our model, we used the EfficientNet-B5 variant of EfficientNet models trained on ImageNet. As the models were originally developed for image classification, we could not directly employ the EfficientNet-B5 [35] as the encoder network in our model. Thus, we removed some top layers of EfficientNet-B5 that are specifically designed for classification output: a final 3×3 convolution layer, a global average pooling layer, and a fully connected (FC) layer. And we used the remaining layers, all layers from input to Block7, as the encoder part without any modification on them. Also, the U-Net architecture uses the skip connections from the encoder network to the decoder network. Typically, the U-Net architecture has four skip connections, and we selected the outputs of *Conv3×3, Block2, Block3*, and *Block5* of EfficientNet-B5 as the sources of skip connections as shown in Fig. 3. Further, we added an extra skip connection from the input layer to the final decoder block. This adds more low-level spatial information to the decoder for precise localization of features and improves the segmentation of fine structures like hair and ruler marks which are very thin, being only a few pixels wide. The output shapes of the blocks corresponding to the skip connections and the final output of the encoder are shown in Table 1.

Second, we constructed a decoder network using a squeeze-and-excitation residual (SERes) structures developed by [37]. The squeeze-and-excitation (SE) block emphasizes the informative features and suppresses the weaker ones by modeling the interdependencies between channels of convolutional features. We used five SERes blocks named *SERes Block1* to *SERes Block5* as shown in Fig. 3. Each SERes block gets two feature maps as inputs — an output feature map from the previous block and a low-level feature map via skip connection from the encoder. For example, the first SERes block (*SERes Block1*) of the decoder gets

$14 \times 14 \times 512$ feature input from *Block7* and $28 \times 28 \times 176$ low-level feature input from *Block5* via skip connection; here, three dimensions of the feature map represent *width* (*W*), *height* (*H*), and *number of feature map* or *channel* (*C*). However, the dimensions of both inputs are not the same. To combine both inputs, first, the $14 \times 14 \times 512$ feature map from the previous layer is upsampled using a transposed convolution, also called deconvolution. The transposed convolution performs both 2×2 upsampling and 3×3 convolution operations. We selected the number of filters for transposed convolution as the half of the number of input channels, i.e., 256 ($= 512/2$), and thus generated a $28 \times 28 \times 256$ feature map. Then, the two inputs are concatenated along the *channel* axis to form a $28 \times 28 \times 432$ feature map before feeding to an SERes block. The SERes block combines an SE block with a residual structure [38], as shown in Fig. 4. The residual unit in the SERes block is a double convolution block, which applies two sets of 3×3 convolution, batch normalization [39], and rectified linear unit (ReLU) operations. Again, we selected the number of filters for two convolution layers in the residual unit as half of the input channels, i.e., 216 ($= 432/2$). The residual unit outputs a $28 \times 28 \times 216$ feature map, and then the squeeze-and-excitation operation is performed to scale the features along the *channel* axis. To

Table 1 Different blocks of the encoder showing the size and number of feature map

Block name	Size ($W \times H$)	#Feature map (C)
Input layer	448×448	3
Conv3×3	224×224	48
Block2	112×112	40
Block3	56×56	64
Block5	28×28	176
Block7	14×14	512
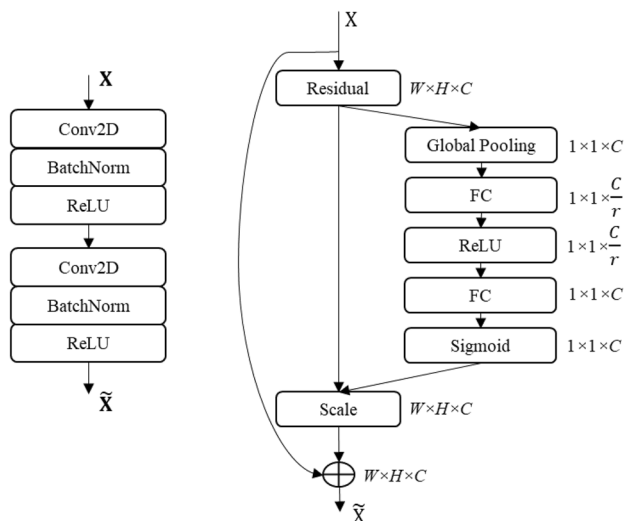
Fig. 4 Double convolution block (left) and SERes block (right)

Table 2 SERes blocks in the decoder and their output shapes

Block name	Size ($W \times H$)	Feature map (C)
SERes Block1	28×28	216
SERes Block2	56×56	86
SERes Block3	112×112	41
SERes Block4	224×224	34
SERes Block5	448×448	10

find the weights for each channel of the feature map, SE first applies global average pooling to reduce the feature map to $1 \times 1 \times 216$ and then applies non-linear operations like FC, ReLU, FC, and sigmoid. The number of neurons in two FC layers are *C/r* and *C* respectively, where *r* is a feature reduction factor and empirically selected as $r = 8$. The SE generates $1 \times 1 \times 216$ weight vector with each value in the range of 0 to 1. Then, a residual feature is multiplied with a weight vector to scale the features and generates a $28 \times 28 \times 216$ scaled feature map. Further, the SERes block combines this scaled residual feature map with the original input feature map. But the channels in the original input feature map ($28 \times 28 \times 256$) and residual feature output ($28 \times 28 \times 216$) are not the same so the 1×1 convolution operation with 216 filters followed by batch normalization operation is performed on the original input feature map. Then, *SERes Block1* adds two feature maps together and applies ReLU operation to generate the final $28 \times 28 \times 216$ feature output.

Similarly, the remaining SERes blocks (*SERes Block2* to *SERes Block5*) apply the same set of operations as *SERes Block1*. Only the size and the number of feature maps are different as shown in Table 2. The number of feature maps (*C*) corresponds to the number of convolutional filters applied in each SERes block. Finally, a dropout with a probability of 0.5, a 1×1 convolution, and a sigmoid function are applied to the output of *SERes Block5* to generate the final segmentation map of size $448 \times 448 \times 1$. The dropout operation is used to regularize the network. The 1×1 convolution reduces the number of channels to the desired number of classes, and the sigmoid operation converts all data values to the range between 0 and 1. Each pixel value in the segmentation map represents the probability score of that pixel belonging to the hair or ruler marks.

During inference, we give five different augmented versions of an input image to the trained deep network: an original image, a horizontally flipped image, a vertically flipped image, a 90° clockwise rotated image, and a 90° counterclockwise rotated image. The deep network generates segmentation output for each image, and the final segmentation mask is generated by aggregating these five outputs. The final aggregated mask is the unweighted average of the five predicted masks. Further, the mask is binarized using the threshold of 0.5 to generate a binary segmentation mask.

Training Details

All models were built using Keras with a TensorFlow backend in Python 3 and trained using a single 32 GB Nvidia V100 graphics card. The networks were trained for 200 epochs with a batch size of 16 and a constant learning rate of 0.0001 using the Adam optimization algorithm [40]. We set up an early stopping criterion with a patience of 30 epochs to stop the model from overtraining. All images are of equal size with dimensions 450×600 and were resized using bilinear interpolation into various sizes, with further details in "Hair Segmentation Performance for Different Image Sizes." We conducted experiments with five different loss functions; see "Hair Segmentation Performance for Different Loss Functions."

Experimental Results and Discussion

We evaluated the performance of the proposed algorithm by comparing the predicted hair masks with the manually drawn ground-truth hair masks on two test sets – 267 images from [31, 32] and 25 images from [16]. The evaluation metrics used are Jaccard index (Jac), Dice similarity coefficient (Dsc), precision (Prec), recall (Rec), accuracy (Acc), and inaccuracy (Inacc), which are defined as:

$$Dsc = \frac{2TP}{2TP + FP + FN}$$

Table 3 Segmentation test results with different input sizes

Image size	Acc	Inacc	Dsc	Jac	Prec	Rec
224×224	97.30	17.85	0.63	0.49	0.69	0.66
320×320	97.76	12.67	0.72	0.58	0.72	0.76
448×448	98.23	10.97	0.77	0.65	0.79	0.79

$$Prec = \frac{TP}{TP + FP}$$

$$Rec = \frac{TP}{TP + FN}$$

$$Acc = \frac{TP + TN}{TP + FP + FP + FN}$$

$$Inacc = \frac{FP}{FP + TN} + \frac{FN}{FN + TP}$$

Here, TP = total of true positive pixels (hair), TN = total of true negative pixels (background), FP = total of false-positive pixels, and FN = total of false-negative pixels.

Hair Segmentation Performance for Different Image Sizes

In this section, we compared the hair segmentation performance of the proposed method with different input sizes using 267 test images. We selected 3 different sizes to modify the original images in the HAM10000 dataset. When the image size was doubled from 224×224 to 448×448, the Jaccard score increased by 32.65% from 0.49 to 0.65, precision increased by 14.49% from 0.69 to 0.79, and recall increased by 19.7% from 0.66 to 0.79. Table 3 shows that larger image sizes have better segmentation performance. However, a larger image size requires more memory and more computational time.

Hair Segmentation Performance for Different Loss Functions

For highly imbalanced data sets, such as hair images where most pixels are true negative pixels, Jadon found that the Focal-Tversky and Log-Cosh Dice loss functions yield improved results [41].

We compared the performance of five widely used loss methods for image segmentation – Dice [42], Jaccard [43], Tversky loss [44], Log-Cosh Dice loss [41], and Focal-Tversky loss [45]. As early stopping with patience of 30 epochs was set up to avoid overfitting, none of the models trained up to 200 epochs. The Dice loss method was trained most with 149 epochs. Surprisingly, the remaining four methods were stopped exactly at same epochs of 95. On 267 test images with 448×448 input resolution, we found that the Dice loss had the best overall result with a 0.649 Jac and 0.775 Dsc. Table 4 shows that the four other loss methods had similar performance in Jac and Dsc measurements with differences of less than 0.5%.

Performance Comparison with Different U-Net Architectures

Next, we compared the segmentation performance of our proposed architecture with two state-of-the-art U-Net architectures using the same 448×448 input resolution and dice loss. Table 5 shows the segmentation performance of ChimeraNet compared to two DL approaches, U-Net [28, 29] and ResUNet-a [33], on 267 test images. The proposed ChimeraNet outperformed the other two methods in all the measurements. It improved the Jaccard score by 10% from 0.59 to 0.65 and the Dice score by 6.9% from 0.72 to 0.77 on U-Net even with the slightly smaller network size. Although our proposed model has a 74% smaller network size than ResUNet-a, it improved the Jaccard score by 3.2% from 0.63 to 0.65 and the precision by 2.6% from 0.77 to 0.79. Furthermore, we evaluated the performance of ChimeraNet with/out the residual squeeze-and-excitation (SERes) structure. ChimeraNet (basic) is the ChimeraNet without the SERes structure. The results show that the addition of the SERes structure improved the segmentation

Table 4 Segmentation test results with different loss functions

Loss method	Acc	Inacc	Dsc	Jac	Prec	Rec
Jaccard	98.081	10.350	0.762	0.634	0.751	0.808
Dice	**98.228**	10.963	**0.775**	**0.649**	**0.790**	0.792
Log-Cosh Dice	98.074	10.349	0.765	0.637	0.761	0.808
Tversky	98.065	**9.480**	0.764	0.635	0.737	**0.826**
Focal-Tversky	98.091	9.899	0.766	0.637	0.751	0.817

The bold entries in each column of Table 4 highlight the best result for the metric associated with the column

Table 5 Comparison of proposed architecture with existing U-Net architectures

Model	# Params	Acc	Inacc	Dsc	Jac	Prec	Rec
U-Net [29]	31.5 M	97.86	13.10	0.72	0.59	0.73	0.75
ResUNet-a [33]	52.8 M	97.79	11.94	0.76	0.63	0.77	0.78
ChimeraNet (basic)	30.3 M	98.18	11.27	**0.77**	0.64	0.78	0.79
ChimeraNet	30.4 M	**98.23**	**10.97**	**0.77**	**0.65**	**0.79**	**0.79**

The bold entries in each column of Table 5 highlight the best result for the metric associated with the column

performance without significantly increasing the number of model parameters. Nevertheless, ChimeraNet (basic) also achieved better performance compared to two other U-Net models.

In Fig. 5, we compare the segmentation results of the proposed model with two DL approaches. The examples in the first, second, and third rows show that U-Net performed well on long and dark hair but poorly on gel and bubble structures and falsely predicted them as hair and ruler objects. In contrast, ResUNet-a performed well on gel and bubble structures but missed many valid hair objects. In case of short hairs as shown in the fourth row, both U-Net and ResUNet-a had difficulty finding them. Furthermore, the proposed ChimeraNet performed well not only on long and short hair, but it was successful in separating the gel and bubble structures from hair and ruler objects. It was also able to find difficult hairs such as dark hairs inside a dark lesion as shown in the second and third rows compared to U-Net and ResUNet-a architectures.

Performance Comparison with Existing Hair Segmentation Algorithms

We compared the performance of our method with seven published hair segmentation algorithms and one under review [16–18, 20–22, 25, 34]. Table 6 shows different scoring metrics computed on the second dataset (25 dermoscopic images) for all algorithms. For methods except for SharpRazor, the hair masks reported in [16] were kindly provided by Ian Lee. Our method achieved better overall scores across all six measurements, as shown in Table 3. The most significant improvements with our method were 45% in Dice similarity coefficient, 36% in recall, and 68% in Jaccard scores from the existing best scores.

We compare reported hair detection methods with the proposed deep learning method in Fig. 6. The proposed method finds thin hairs that other methods fail to detect and finds less noise.

Fig. 5 Hair segmentation results of proposed ChimeraNet against U-Net and ResUNet-a for HAM10000 test images. The segmentation results show true positives (green), false positives (red), and false negatives (blue). The U-Net model finds more false positives (for example, gel bubbles), and ResUNet-a finds less hair. The proposed ChimeraNet model successfully detects hair with fewer false positives and false negatives
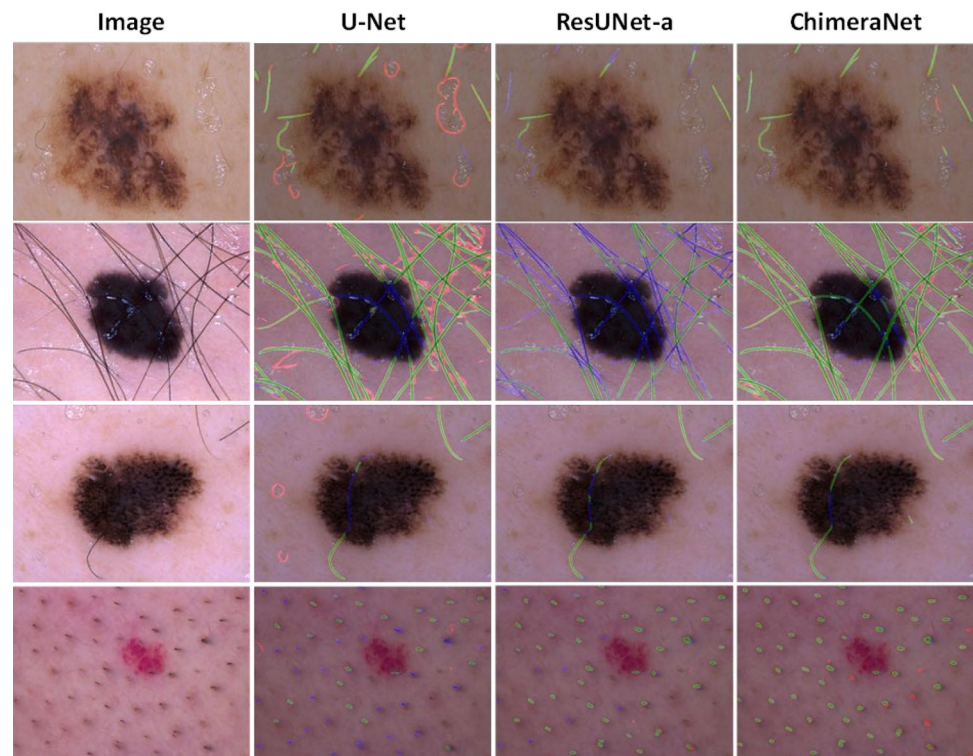
Table 6 Results, resolution 448×448, compared with existing hair segmentation algorithms on 25 test images

| Methods and year of publication	Acc	Inacc	Prec	Rec	Dsc	Jac
ChimeraNet	*96.06*	*11.59*	*0.79*	*0.79*	*0.77*	*0.64*
SharpRazor (2021) [34]	93.80	22.48	0.58	0.58	0.53	0.38
Lee et al. (2017) [16]	90.99	29.39	0.60	0.44	0.40	0.26
Xie et al. (2015) [17]	90.04	39.07	0.37	0.25	0.25	0.15
Koehoorn et al. (2015) [22]	80.13	49.11	0.08	0.14	0.07	0.03
Abbas et al. (2013) [18]	87.36	29.40	0.34	0.49	0.33	0.22
Huang et al. (2013) [25]	81.13	32.94	0.23	0.50	0.26	0.16
Fiorese et al. (2011) [21]	91.74	37.82	0.68	0.26	0.32	0.20
DullRazor (1997) [20]	93.15	34.87	0.66	0.31	0.38	0.25

Discussion

After Ronneberger et al. [29] proposed the U-Net architecture for biomedical image segmentation, many new variants of U-Net have been developed. One variant used in skin image segmentation is dense residual U-Net, which replaces some convolutions with dense convolutions and appends residual convolutions to the network [46]. This approach may offer better generalization capability and may be more robust on small dataset sizes, but at the cost of slightly decreased overall performance. Diakogiannis et al. [33] proposed ResUNet-a that employs residual connections, atrous (dilated) convolutions, pyramid scene parsing pooling, and multi-tasking inference to segment monotemporal very high-resolution aerial images. Baheti et al. [47] modified U-Net architecture by employing EfficientNet [35] in the encoder network and basic convolution layers in the decoder network for scene segmentation. Here, we propose a new hair segmentation method based on U-Net architecture by employing EfficientNet-B5 as the encoder network and squeeze-and-excitation structures as the building block of the decoder network. We used the squeeze-and-excitation structure, which has better generalization capability than basic convolution structures because it focuses on more critical channels of feature maps. In addition, we put an extra skip connection from an input to the final block of the decoder to add low-level spatial information for precise localization of features. The proposed architecture performed better than U-Net and ResUNet-a to detect fine structures like hair and ruler marks in skin lesion images.

We used several metrics to evaluate the performance of the DL hair detection approach. The accuracy metric counts true-negative pixels and true positive pixels equally. For most images, except on the scalp, most of the pixels are true negative pixels (not hair pixels). Thus, the accuracy metric overstates actual performance. Dice and Jaccard metrics give a better performance evaluation because they are better for scoring detection of structures with less overall representation in the images. Our DL approach gives nearly twice the Jaccard score of any classical method except for the SharpRazor method.

Although most reports consider hair pixels unwanted noise, hair and ruler structures may contribute to diagnostic accuracy. White hairs may indicate patient age. Shaved hairs

Fig. 6 Lesion image, ground-truth hair mask, and overlays of predicted hair mask on lesion image for nine hair detection methods. The proposed ChimeraNet method accurately detects more hair with less noise compared to other hair detection methods
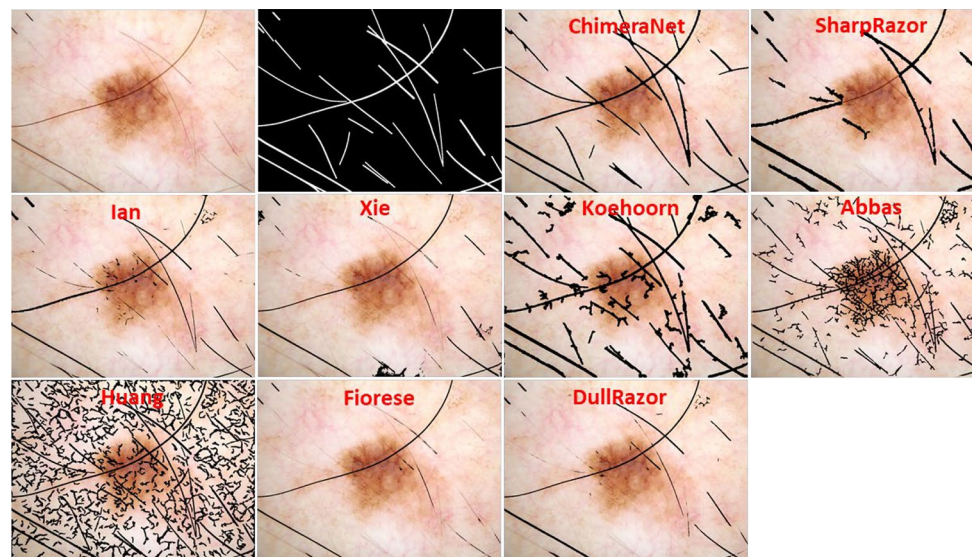
may determine body location and gender, such as a male face or a female leg. Ruler marks may give an indication of the clinic where the images originated. Although hair structures still constitute noise, their detection could contribute to diagnostic accuracy, and thus they could be considered useful noise. The deep learning approach reported here, which can detect ruler marks and white hair, may yield useful information from the automatically detected structures.

Conclusion and Future Work

In this study, we employ a novel deep learning technique to find hair and ruler structures. Using an image resolution on the order of 448×448, the reported method achieves state-of-the-art accuracy compared to other approaches. One disadvantage of the technique is that it requires many images, with the training, validation, and test sets totaling 1333 images. These training images, which required hundreds of hours to create, can nevertheless serve as training images for future hair detection methods. In future work, we will make the training hair masks available publicly. In addition, we will employ deep learning to detect other dermoscopic features.

Despite the fairly high accuracy achieved by this method, there is a need for future work to improve the results. New approaches could include fusion methods and postprocessing using modeling or other methods. Deep learning architectures have advanced rapidly, and as techniques evolve, other models can be employed.

Author Contribution We confirm that all authors contributed to the study conception and design. Data collection was performed by Reagan Young, Jessica Miinch, Januka Nepal, and Reagan Young. Data analysis as performed by all authors. The first draft of the manuscript was written by Norsang Lama, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, Cancer statistics, 2021, *CA. Cancer J. Clin.*, vol. 71, no. 1, pp. 7–33, 2021.

2. H. W. Rogers, M. A. Weinstock, S. R. Feldman, and B. M. Coldiron, Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012, *JAMA Dermatology*, vol. 151, no. 10, pp. 1081–1086, 2015, https://doi.org/10.1001/jamadermatol.2015.1187.

3. A. Esteva *et al.*, Dermatologist-level classification of skin cancer with deep neural networks, *Nature*, vol. 542, no. 7639, pp. 115–118, 2017, https://doi.org/10.1038/nature21056.

4. L. K. Ferris *et al.*, Computer-aided classification of melanocytic lesions using dermoscopic images, *J. Am. Acad. Dermatol.*, vol. 73, no. 5, pp. 769–776, Nov. 2015, https://doi.org/10.1016/J.JAAD.2015.07.028.

5. H. Pehamberger, M. Binder, A. Steiner, and K. Wolff, In vivo epiluminescence microscopy: improvement of early diagnosis of melanoma, *J. Invest. Dermatol.*, vol. 100, no. 3 SUPPL., pp. S356–S362, 1993, https://doi.org/10.1038/jid.1993.63.

6. H. P. Soyer, G. Argenziano, R. Talamini, and S. Chimenti, Is dermoscopy useful for the diagnosis of melanoma?, *Arch. Dermatol.*, vol. 137, no. 10, pp. 1361–1363, Oct. 2001, https://doi.org/10.1001/archderm.137.10.1361.

7. R. P. Braun, H. S. Rabinovitz, M. Oliviero, A. W. Kopf, and J. H. Saurat, Pattern analysis: a two-step procedure for the dermoscopic diagnosis of melanoma, *Clin. Dermatol.*, vol. 20, no. 3, pp. 236–239, May 2002, https://doi.org/10.1016/S0738-081X(02)00216-X.

8. M. A. Marchetti *et al.*, Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images, *J. Am. Acad. Dermatol.*, vol. 78, no. 2, pp. 270-277.e1, Feb. 2018, https://doi.org/10.1016/j.jaad.2017.08.016.

9. H. A. Haenssle *et al.*, Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists, *Ann. Oncol.*, vol. 29, no. 8, pp. 1836–1842, 2018, https://doi.org/10.1093/annonc/mdy166.

10. S. Pathan, K. G. Prabhu, and P. C. Siddalingaswamy, Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review, *Biomed. Signal Process. Control*, vol. 39, pp. 237–262, Jan. 2018, https://doi.org/10.1016/J.BSPC.2017.07.010.

11. T. Majtner, S. Yildirim-Yayilgan, and J. Y. Hardeberg, Combining deep learning and hand-crafted features for skin lesion classification, *2016 6th Int. Conf. Image Process. Theory, Tools Appl. IPTA 2016*, 2017, https://doi.org/10.1109/IPTA.2016.7821017.

12. N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images BT - machine learning in medical imaging, 2015, pp. 118–126.

13. N. C. F. Codella *et al.*, Deep learning ensembles for melanoma recognition in dermoscopy images, *IBM J. Res. Dev.*, vol. 61, no. 4–5, pp. 5:1–5:15, Jul. 2017, https://doi.org/10.1147/JRD.2017.2708299.

14. I. González-Díaz, DermaKNet: incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis, *IEEE J. Biomed. Heal. Informatics*, vol. 23, no. 2, pp. 547–559, 2019, https://doi.org/10.1109/JBHI.2018.2806962.

15. J. R. Hagerty *et al.*, Deep learning and handcrafted method fusion: higher diagnostic accuracy for melanoma dermoscopy images, *IEEE J. Biomed. Heal. Informatics*, vol. 23, no. 4, pp. 1385–1391, 2019, https://doi.org/10.1109/JBHI.2019.2891049.

16. I. Lee, X. Du, and B. Anthony, Hair segmentation using adaptive threshold from edge and branch length measures, *Comput. Biol. Med.*, vol. 89, no. August, pp. 314–324, 2017, https://doi.org/10.1016/j.compbiomed.2017.08.020.

17. F. Xie, Y. Li, R. Meng, and Z. Jiang, No-reference hair occlusion assessment for dermoscopy images based on distribution feature, *Comput. Biol. Med.*, vol. 59, pp. 106–115, 2015, https://doi.org/10.1016/j.compbiomed.2015.01.023.

18. Q. Abbas, I. F. Garcia, M. Emre Celebi, and W. Ahmad, A feature-preserving hair removal algorithm for dermoscopy images, *Ski. Res. Technol.*, vol. 19, no. 1, pp. e27–e36, 2013.

19. N. H. Nguyen, T. K. Lee, and M. S. Atkins, Segmentation of light and dark hair in dermoscopic images: a hybrid approach using a universal kernel, in *Medical Imaging 2010: Image Processing*, 2010, vol. 7623, pp. 1436–1443.

20. T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, Dullrazor®: a software approach to hair removal from images, *Comput. Biol. Med.*, vol. 27, no. 6, pp. 533–543, Nov. 1997, https://doi.org/10.1016/S0010-4825(97)00020-6.

21. M. Fiorese, E. Peserico, and A. Silletti, VirtualShave: automated hair removal from digital dermatoscopic images, *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 5145–5148, 2011, https://doi.org/10.1109/IEMBS.2011.6091274.

22. J. Koehoorn *et al.*, Automated digital hair removal by threshold decomposition and morphological analysis, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9082, pp. 15–26, 2015, https://doi.org/10.1007/978-3-319-18720-4_2.

23. M. T. B. Toossi, H. R. Pourreza, H. Zare, M.-H. Sigari, P. Layegh, and A. Azimi, An effective hair removal algorithm for dermoscopy images, *Ski. Res. Technol.*, vol. 19, no. 3, pp. 230–235, 2013.

24. O. Abuzaghleh, B. D. Barkana, and M. Faezipour, Noninvasive real-time automated skin lesion analysis system for melanoma early detection and prevention, *IEEE J. Transl. Eng. Heal. Med.*, vol. 3, pp. 1–12, 2015.

25. A. Huang, S. Y. Kwan, W. Y. Chang, M. Y. Liu, M. H. Chi, and G. S. Chen, A robust hair segmentation and removal approach for clinical images of skin lesions, *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 3315–3318, 2013, https://doi.org/10.1109/EMBC.2013.6610250.

26. H. Zhou *et al.*, Feature-preserving artifact removal from dermoscopy images, in *Medical Imaging 2008: Image Processing*, 2008, vol. 6914, p. 69141B.

27. M. Attia, M. Hossny, H. Zhou, S. Nahavandi, H. Asadi, and A. Yazdabadi, Digital hair segmentation using hybrid convolutional and recurrent neural networks architecture, *Comput. Methods Programs Biomed.*, vol. 177, pp. 17–30, 2019, https://doi.org/10.1016/j.cmpb.2019.05.010.

28. W. Li, A. N. Joseph Raj, T. Tjahjadi, and Z. Zhuang, Digital hair removal by deep learning for skin lesion segmentation, *Pattern Recognit.*, vol. 117, 2021, https://doi.org/10.1016/j.patcog.2021.107994.

29. O. Ronneberger, P. Fischer, and T. Brox, U-Net: convolutional networks for biomedical image segmentation. [Online]. Available: http://lmb.informatik.uni-freiburg.de/.

30. F. Rodriguez, E. Maire, P. Courjault-Radé, and J. Darrozes, The Black Top Hat function applied to a DEM: a tool to estimate recent incision in a mountainous watershed (Estibère Watershed, Central Pyrenees), *Geophys. Res. Lett.*, vol. 29, no. 6, pp. 1–9, 2002.

31. P. Tschandl, C. Rosendahl, and H. Kittler, Data Descriptor: the HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions Background & Summary, *Nat. Publ. Gr.*, 2018, https://doi.org/10.1038/sdata.2018.161.

32. N. Codella *et al.*, Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the International Skin Imaging Collaboration (ISIC).

33. F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data, *ISPRS J. Photogramm. Remote Sens.*, vol. 162, no. March 2019, pp. 94–114, 2020, https://doi.org/10.1016/j.isprsjprs.2020.01.013.

34. R. Kasmi et al., SharpRazor: automatic removal of hair and ruler marks from dermoscopy images, *Ski. Res. Technol.*, 2021. Accepted.

35. M. Tan and Q. Le, Efficientnet: rethinking model scaling for convolutional neural networks, in *International conference on machine learning*, 2019, pp. 6105–6114.

36. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, Imagenet: a large-scale hierarchical image database, in *2009 IEEE conference on computer vision and pattern recognition*, 2009, pp. 248–255.

37. J. Hu, L. Shen, and G. Sun, Squeeze-and-excitation networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.

38. K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

39. S. Ioffe and C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift.

40. D. P. Kingma and J. Ba, Adam: a method for stochastic optimization, *arXiv Prepr.* arXiv1412.6980, *2014*.

41. S. Jadon, A survey of loss functions for semantic segmentation, *2020 IEEE Conf. Comput. Intell. Bioinforma. Comput. Biol. CIBCB 2020*, 2020, https://doi.org/10.1109/CIBCB48159.2020.9277638.

42. C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2017, pp. 240–248.

43. M. A. Rahman and Y. Wang, Optimizing intersection-over-union in deep neural networks for image segmentation, in *International symposium on visual computing*, 2016, pp. 234–244.

44. S. S. M. Salehi, D. Erdogmus, and A. Gholipour, Tversky loss function for image segmentation using 3D fully convolutional deep networks, in *International Workshop on Machine Learning in Medical Imaging*, 2017, pp. 379–387.

45. N. Abraham and N. M. Khan, A novel focal tversky loss function with improved attention u-net for lesion segmentation, in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 683–687.

46. N. C. F. Codella *et al.*, Segmentation of both diseased and healthy skin from clinical photographs in a primary care setting, in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 3414–3417.

47. B. Baheti, S. Innani, S. Gajre, and S. Talbar, Eff-UNet: a novel architecture for semantic segmentation in unstructured environment, *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2020-June, pp. 1473–1481, 2020, https://doi.org/10.1109/CVPRW50498.2020.00187.