



Development and qualification of a machine learning algorithm for automated hair counting

Jarek P. Sacha¹  | Tamara L. Caterino¹ | Brian K. Fisher¹ | Gregory J. Carr¹ |
R. Scott Youngquist² | Brian M. D'Alessandro³ | Anthony Melione³ |
Douglas Canfield³ | Wilma F. Bergfeld^{4,5} | Melissa P. Piliang^{4,5} |
Raghu Kainkaryam² | Michael G. Davis¹ 

¹The Procter and Gamble Company,
Mason, Ohio, USA

²Former P&G employee, Mason, Ohio,
USA

³Canfield Scientific, Parsippany, New
Jersey, USA

⁴Department of Dermatology,
Cleveland Clinic, Cleveland, Ohio, USA

⁵Department of Pathology, Cleveland
Clinic, Cleveland, Ohio, USA

Correspondence

Jarek P. Sacha, P&G, Mason Business
Center, 8700 Mason-Montgomery Rd,
Mason 45040, OH, USA.
Email: sachajp@pg.com

Funding information

P&G Beauty

Abstract

Objective: Determining the amount of hair on the scalp has always been an important metric of patient satisfaction for hair growth and hair retention technologies. While simple in concept, this measurement is a difficult, resource intensive task for the dermatologist and the research scientist. Specifically, counting and measuring hair in phototrichogram images is very time consuming and labour intensive. Due to cost, often only a fraction of available images is manually analysed. There is a need for an automated method that can significantly increase speed and throughput while reducing the cost of counting and measuring hair in phototrichogram images.

Methods: Recent advances in machine learning and deep convolutional neural networks (deep learning) have led to a revolution in the analysis of image, video, speech, text and other sensor data. Image diagnostics have seen remarkable improvements with completely automated methods outperforming both human experts and human-engineered analysis methods. Deep learning methods can also provide speed and cost benefits. To enable use of a deep learning, we created a data set of 288 manually annotated phototrichogram images with marked location and length of each hair (the training dataset). We designed a custom neural network architecture and custom image processing algorithms to best utilize the available training data and to maximize performance for hair counting and length measurement. The performance of the algorithm was qualified by comparing hair count and length measurements to an independent ground truth method, the semi-manual Canfield's Hair Metrix method.

Results: Leveraging deep neural networks, we have developed capability to apply machine learning to reduce the time needed to acquire data from phototrichograms of patients' scalp from months to seconds. Our algorithm enables fast and fully automated hair counting and length measurement. The algorithm shows high agreement with human manually assisted analysis (ground truth).

Conclusions: We have trained and deployed an algorithm utilizing this technology and have demonstrated the reproducibility, accuracy and speed of this algorithm that, once deployed, requires little to no recurring cost or manual intervention for its operation. The method allows fast analysis of large number of images, reducing study cost and significantly reducing study analysis time.

KEYWORDS

computer modelling, hair growth, hair treatment

INTRODUCTION

Society has placed a great deal of social and cultural importance on hair and hairstyles. Consequently, when hair begins to appear thinner, both women and men become concerned. A statistically significant increase in hair count with treatment relative to vehicle control is a desirable feature for new ingredients and products. Traditionally, the evaluation of new hair growth and hair retention technologies have relied on hair quantification methodology such as Canfield Scientific's Hair Metrix systems [1]. Effectively quantifying these treatment benefits requires an accurate and robust counting method for clinical scalp images. Additionally, due to relatively small changes in hair count, clinical studies have large subject base-sizes to observe statistically significant changes and this requires the counting of hundreds of images for each study. Though the current method has the accuracy and robustness required for the task, it has a long turn-around time and is costly. In response to these challenges, we developed novel machine learning algorithms that can accurately and reliably count hairs in an image in under a second at no recurring cost and limited human intervention post-image capture.

Our method performs detection, counting and length measurement of hairs in phototrichogram images [2, 3]. Some method in the past attempted to automate analysis of similar images. TrichoScan [4] performs detection and measurement of clipped hair in phototrichogram images. It cannot detect light, grey or translucent hair directly as they exhibit low contrast compared to the scalp skin and to artefacts such as bubbles, reflections or flakes. It requires dyeing of the hair before taking images to help in successful detection. TrichoScan implements traditional image analysis approaches consisting of multiple phases such as automated thresholding, connected objects detection and object filtering based on size and shape to discard tattoos and other artefacts. A method based on a modified Hough transform to detect hair in the microscopy images was proposed in [5]. Like TrichoScan, it requires significant contrast between hair and scalp (hair is assumed to

be dark and scalp light) and relies on traditional image analysis techniques. A deep learning approach to hair count estimation is presented in [6]. The application is focused to counting short shaved hair to assess performance of laser hair removal. It is reapplying some of well-established deep learning classification models the task of count estimation. In contrast to TrichoScan [4] and Hough-based method [5], their approach does not perform hair detection, rather it 'weights' appearance of the hair in the image to estimate hair count (similar to estimating number of jelly beans in closed glass jar). This is similar to the approach for estimating object counts presented in [7]. They conclude that their error count is significant compared with expert human counting but is sufficient for their hair removal application.

Compared to TrichoScan [4] and Hough-based method [5], our hair detection method does not require dyeing of hair before image capture. Neither does it require additional image preprocessing (like some contrast enhancement or artefact suppression) prior to analysis as that processing could introduce its own artefacts. Detecting grey and translucent hair in phototrichogram images is a challenging task even for human graders. With our approach, we were able to address that challenge.

MATERIAL AND METHODS

Study design and statistical analysis

This was a 10-week randomized, double-blind, vehicle controlled split-scalp study designed to test the efficacy of the treatment products on the appearance of hair density. Qualifying female subjects ($N = 170$) ages 45 to 65 inclusive with self-perceived thinning hair participated in a 2-week pretreatment phase and were randomized into one of three (with one group receiving Minoxidil) treatment groups for an 8-week treatment phase. Subjects came onsite for twice daily split-scalp application by technician. One side of the subject's scalp was randomized to the treatment product while the other side was randomized to

a vehicle control. Scalp images for hair count were taken at Baseline and Week 8. Scalp images for anagen and telogen related metrics were taken at Baseline+24 h and Week 8+24 h.

The statistical method used to analyse the results of this study is a linear mixed effects model [8] that uses baseline hair count as a covariate to adjust for the wide variation in individual subject's hair density, and a treatment factor to estimate the magnitude of treatment benefit. Because subjects are exposed to both treatments on the two sides of their heads, a random effect for subject is also part of the model. The identical model is used to analyse data from both the ground truth method and the novel algorithm.

Imaging area preparation

Prior to Baseline, to ensure image repositioning of the image areas at Baseline and after treatment, a micro-pigmentation tattoo is placed in the middle of each site (located as described below) by a licensed tattoo artist (MCN International equipment; Model number: MC-8880-SP). At Baseline, two (one on the left and one on the right side of the scalp) approximately 1.9 x 1.9 cm square sites are prepared on the subject's scalp approximately 2" lateral to the mid-scalp line at the bisection of the mid-eyebrow arch line and the middle of the upper ear. Long hairs within each site are removed with scissors. Short hairs within the marked sites are clipped to approximately 0.5 mm in length using a #2041 Doodle Blade® attached to a corded Wahl® hair clipper. Using a magnifying loupe, each clipped site was prepped to ensure the area was free of hair clippings and stray hairs. The sites are prepared for image capture by ensuring that hair around each clipped site had been taped back and away from the clipped area to be imaged.

Image capture

Ensuring that the tattoo is in the centre of the image, phototrichogram images [3] are captured at Baseline/Day 1 (0 h) and Day 2, (~24 h later). At each time point, images are captured for each clipped site. Hairs for analysis are considered within a circular area of interest fixed to be 1 cm² near the centre of the image.

Images are captured with two different generations of an imaging system from Canfield. The newer system used in more recent studies was the Canfield VEOS SLR hand-held digital imaging system [9]. The system consists of a Canon EOS Rebel T6i 24.2-megapixel camera with a specially designed macro lens with 36 LEDs illumination.



FIGURE 1 Manual labelling representing hair as line segments: one point at tip (green) and one at base (red). Notice a highly translucent hair, second from the left

It captures images within 22 mm by 15 mm field of view. Each image is 6000 by 4000 pixels. The system used in earlier studies was the Canfield EpiFlash imaging system [10]. The system consists of a Nikon D80 10.2-megapixel camera with Canfield EpiFlash and 60 mm Nikkor lens. It captures images within 22 mm by 15 mm field of view. Each image is 4288 by 2848 pixels.

Dataset selection

The automated method was developed and tested using phototrichogram images acquired from a clinical study captured with the Canfield VEOS imaging system. From this study, a subset of images was manually analysed using Canfield Scientific's Hair Metrix [1]. The Metrix analysis involves a manual process by trained and certified technicians to identify and measure all hairs within the area of interest (AOI). Several rounds of quality control (QC) by multiple technicians ensure a reliable output, which includes hair count and hair length data. The Metrix data were used as the ground-truth for assessing the automated method performance. The Metrix data were not used in the method's training as it did not provide individual hair annotations.

An additional study captured with the older Canfield EpiFlash imaging system was used for qualification of the method described in the Results section.

Creation of training labels & training approach

Images from a clinical study utilizing the Canfield VEOS imaging system were selected as training images. 288 images were randomly selected from the Baseline visit, at both 0 h and 24 h. This set was split into 227 training images and 61 training validation images. The split was done in such a way that the training and validation sets contained different subjects.

Creation of training data was performed by manually marking every hair within a 1 cm^2 circular area of interest within the phototrichogram images. The total of 86103 hairs were marked. On average, there were 300 hairs in each image's area of interest. The objective of the method is to detect individual hairs and to measure their length. Post-clipping, the vast majority of the hairs in the phototrichogram are short and straight. Therefore, we determined that it is sufficient to mark (label) each hair with a two-point line segment with one point at the tip of a hair and the second at the base of the hair; see example in Figure 1. This significantly simplifies the process of creating the training dataset. All labels in the training images were reviewed by an expert.

Automated hair detection method

Automated hair detection is complicated by the close proximity of distinct hairs, low contrast, and varying appearance of hair and scalp. Current off-the-shelf automated object detection approaches, like Mask-RCNN [11], have difficulty detecting and resolving intersecting, closely spaced, elongated objects. To overcome this difficulty, we developed a custom approach. Conceptually, the

algorithm is performing two operations: detection of hair tips, and for each tip, detection of the hair's base.

The tip detection approach is inspired by a concept similar to that proposed previously [7] in which objects to detect are represented by dots. In our case, a dot represents the centre of the hair tip. In order to account for a certain level of ambiguity, image resolution and hair tip appearance, a Gaussian distribution is centred at the dot. In [7], the authors sum the individual distributions to estimate total object count. Here, we use the Gaussians maxima to explicitly detect each hair tip location. The Gaussian estimates of tip locations are created by a U-Net-like convolutional neural network [12]. An example of an output produced by the tip detection model is shown in Figure 2.

One of the key contributions of our approach is the efficient detection of the bases for each detected hair tip. We encode each base location as an offset (dx , dy) from the location of the tip. To make the process efficient, both during training and prediction, we combine the tip and base detection into a single step. We encode the probability of tip location and the base offset as a triple (p , dx , dy). The network is set up to take a colour phototrichogram image as an input and produce a matrix of triples as an output. The size of the output matrix is the same as the

FIGURE 2 Example of tip localization. Note that the model only detects hair that is fully contained in the image

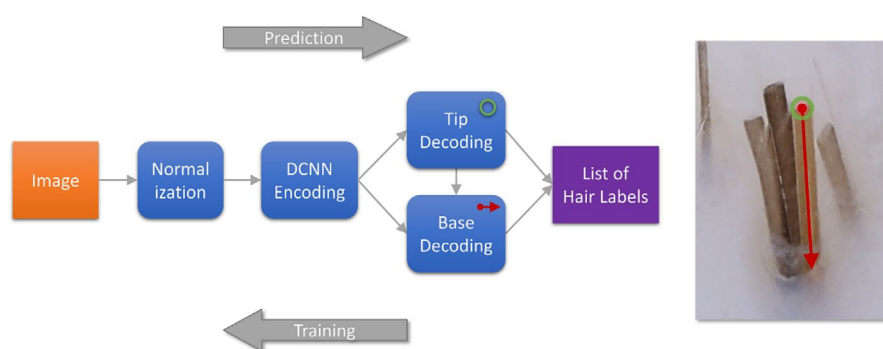
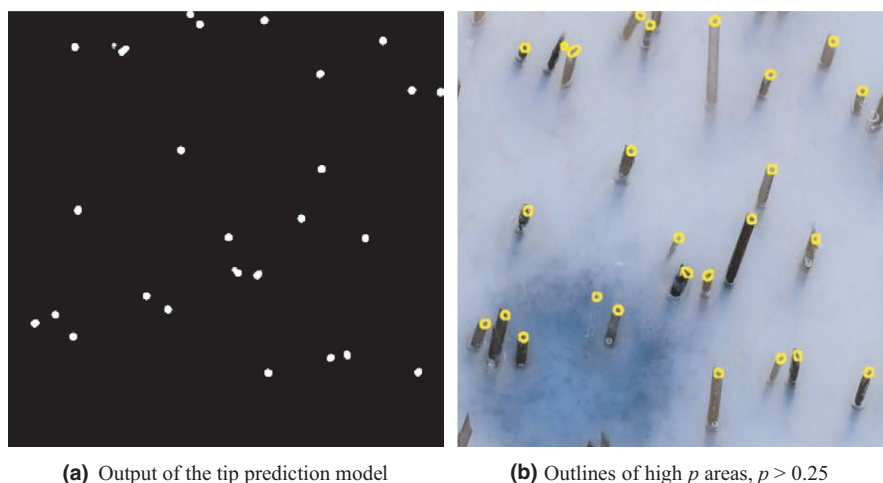


FIGURE 3 Hair detection method – training and prediction process

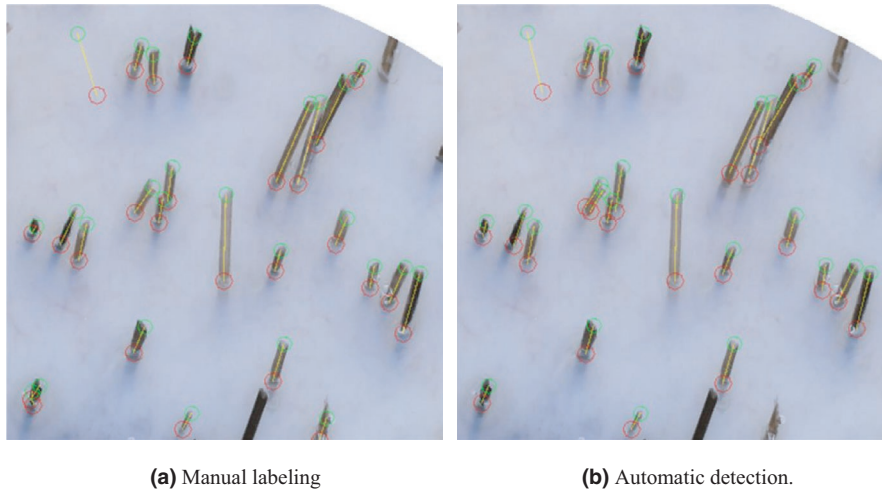


FIGURE 4 Manual labelling (left) and automated detection using the proposed algorithm (right). Note: by design the algorithm only detects hair that are fully contained in the image and the area of interest. The white area near the top right corner indicates the edge of the area of interest. Highly translucent, low contrast, hairs (top left corner) are also detected, see Figure 5 for a close-up view



FIGURE 5 Close-up of the top left of the image in Figure 4 showing example of a difficult to detect highly translucent, low contrast hair. Tip and base of a hair on the left are marked with arrows: green tip, red base

size of the input image. The training and the prediction processes are illustrated in Figure 3.

During training, a list of hair labels corresponding to an input image is encoded into a form of a matrix of triples (p , dx , dy). Consider that matrix to be three image bands. In the p -band, at the location of the hair tip, we create a Gaussian of a unit height (pixel intensity value) and fixed width. In the dx -band, we create a dot at the location of the hair tip. The dot has fixed radius, and its pixel intensity value is proportional to value of the offset dx . The dy -band is created likewise.

The training of the auto-decoder is a regression problem. The cost function is a weighted average squared distance between the predicted and the extracted (p , dx , dy) matrix. The cost function has two elements: tip localization cost and base localization cost. Base localization cost is larger than 0 only within proximity of the corresponding tip.

During prediction of hair location, the (p , dx , dy) matrix is decoded to extract the location of hair labels. We use a maximum compression algorithm to determine location of the hair tips from the p -band. For each detected hair tip location, we decode the offset (dx , dy) to the hair base. To decode dx , we compute a weighted average of values in the dx -band at the location of the tip and likewise for the dy offset. An example of manual hair labelling and results of automated hair detection are shown in Figures 4 and 5.

Image capture for treatment benefit confirmation

Phototrichogram images were captured from a second clinical study at Baseline/Day 1 (0 h) and Week 8/Day 1 (0 h) with a Canfield Epiflash SLR hand-held digital imaging system described earlier. At each time point, images were captured for clipped sites as described previously. Hair for analysis are considered within a circular area of interest fixed to be 1 cm^2 near the centre of the image. Hairs which cross over the border of the AOI are included for analysis only if the base of the hair is inside the AOI.

Qualification / acceptance criteria

The following criteria were devised to demonstrate algorithm qualification and suitability for use in analysing images for total hair count:

1. Deming regression [13] to visualize and quantify the agreement between the automated and manual methods.
2. Bland–Altman visualizations of the repeatability were performed to assess the agreement between the two

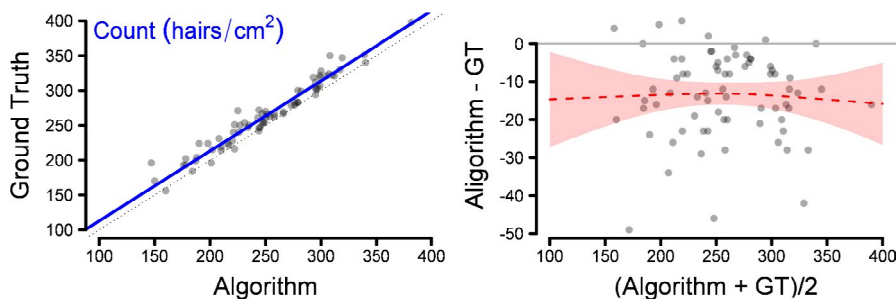
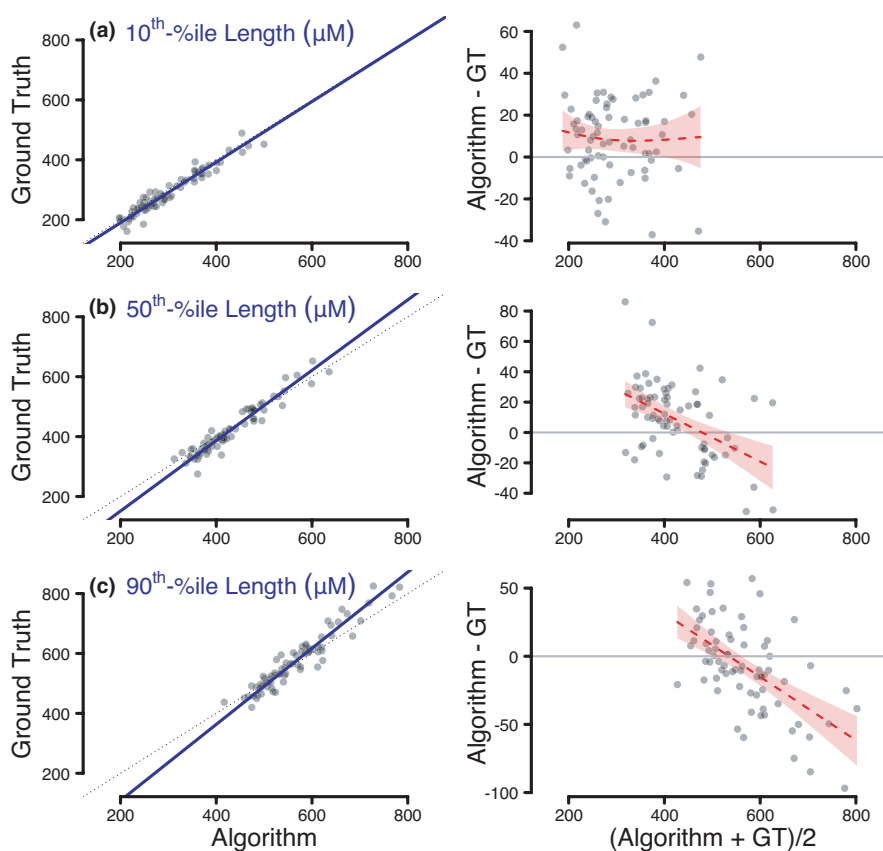


FIGURE 6 Visualizations of the agreement between the algorithm and ground truth methods for hair count. The left graph shows the data used with a Deming regression superimposed (blue) and the line of equality is the dotted diagonal line. The right graph contains Bland–Altman visualizations, which help to quantify biases, estimated by red regression line and confidence region, as a function of the average measurement level

FIGURE 7 Visualizations of the agreement between the algorithm and ground truth methods for hair length. The left-hand column shows the data used with a Deming regression superimposed (blue) and the line of equality is the dotted diagonal line. The right-hand column contains Bland–Altman visualizations, which help to quantify biases, estimated by red regression line and confidence region, as a function of the average measurement level



methods. Various publications describe their use [14]. The regression model through these data, analogous to a paired t-test on the difference between the two methods, should be close to zero.

3. Sources of variability related to the measurement methods themselves, and the variability between the samples to which they are applied are summarized in an intra-class correlation coefficient (ICC) for which values >0.75 are considered to be excellent by various external sources [15, 16]. ICC is a measure of measurement noise (differences between the methods) versus

subject-to-subject noise. When the measurement variability is very small relative to the expected range of the measurements made, the ICC will be close to 1 and goes to 0 if the measurement variability is very large. Another way to think of the ICC is that it measures the fraction of total variation that is due to systematic differences between the samples being measured (here, subjects and sites within a subject).

4. The ability of the algorithm to detect hair count treatment benefits was confirmed to reproduce results of the Canfield Hair Metrix gold standard method. Statistical

TABLE 1 Confirmation of ability to detect treatment benefits for hair count

Ground Truth method								
	Mean levels		Change from baseline			Benefit (Difference of CFBs)		
	Week 0	Week8	Mean	SE	<i>p</i> value	Mean	SE	<i>p</i> value
Minoxidil	201	214	12.98	2.11	<0.0001	5.82	2.25	0.016
Minoxidil Vehicle	205	212	7.16	2.10	<0.001			
Algorithm method								
	Mean levels		Change from baseline			Difference of CFBs		
	Week 0	Week8	Mean	SE	<i>p</i> value	Mean	SE	<i>p</i> value
Minoxidil	193	195	2.53	1.77	0.16	5.58	2.43	0.022
Minoxidil Vehicle	194	190	−3.04	1.77	0.09			

significance as well as the estimate of proven benefit were compared between the methods.

RESULTS

Seventy-two phototrichogram images were selected to investigate the performance of the hair analysis algorithm. Care was taken to ensure that the 72 images have no subjects in common with the training images. Manual counts by the standard Canfield process were provided as ground truth numbers. Deming regression analysis (Figure 6 left) demonstrates a high level of agreement between the two methods across the range of hair counts. There is a slight tendency for the manual method to detect more hairs on average, with about 13.5 extra hairs detected. The slope of the line is nearly equal to 1, so the two methods seem to differ primarily by this offset which may be related to the presence of fine vellus hairs, hairs the algorithm was not trained to detect. The Bland-Altman visualization (Figure 6 right) is consistent with a similar interpretation and shows that the difference between the two methods is statistically significant because the confidence region does not cover the reference line at zero. This regression analysis of bias is a variation on Bland-Altman commonly used in gene expression normalization [17], as it provides a visualization of the evidence for bias present. The ICC value is 0.97. This very high value suggests that the algorithm tracks the ground truth very well and that the rank-ordering of measurements will be in close agreement.

For length, a slightly different approach to the validation is taken, because there is no way to pair the measurements at the hair level. Instead, the hair lengths are summarized into a single number for each image and the 10th, 50th and 90th percentiles of length measurements are compared by the two methods (Figure 7). Like Figure

6, the left column of Figure 7 shows the Deming regression analysis, the right column shows Bland-Altman visualization. There is some evidence of relatively small biases. For example, the middle row for median length (Figure 7B) shows that the algorithm might differ by ± 20 units, but when the average measurement is 400 or more, this is a 5% difference. Most importantly, the ICC across these different percentiles is very high (96%, 95% and 93% for the 10th, 50th and 90th percentiles of hair length, respectively) and demonstrates that conclusions from statistical analyses for effects on these end points should be consistent.

Finally, the ability to detect hair count treatment benefits utilizing the algorithm was confirmed to reproduce results of Canfield Hair Metrix gold standard/ground truth method in showing a statistically significant difference between Minoxidil and control after 8 weeks of treatment (Table 1).

DISCUSSION

Our algorithm effectively detects and counts hairs and determines hair length. The quality of these measures has been confirmed by replicating the results from a clinical treatment study. Furthermore, we have demonstrated that our algorithm is robust to changing imaging conditions. The method was developed using training images from the high resolution VEOS imaging system. We also evaluated the method using the older, lower resolution EpiFlash system, with different camera and illumination. Despite being trained only on VEOS, the method was able to satisfy qualification criteria on both VEOS and EpiFlash imaging systems.

The algorithm can be extended to measure additional end points, for example, hair width, because our method is built on a flexible convolutional neural network

architecture. The implementation presented in this paper focused two end points: dx and dy. They are used to locate hair base and compute hair length. The end points are represented by features in the network output layer. Additional end points can be added by incorporating additional output features, similar to dx and dy features. The most straight forward end-point to add is hair *width*. We will need to add hair width to the training data (hair labels). The same neural network architecture can be used with the only change being the addition of one more output feature for hair width. Updated network weight can be quickly computed using transfer learning, or if more or different training data are available, with full retraining of the weights.

The utilization of machine learning algorithms has the potential to transform patient care, research and companies selling hair growth and hair retention products. Automated algorithms which can immediately assess hair count taken from phototrichogram images offers the potential for clinicians to provide immediate information to their patients and assess their hair condition or treatment progress over time. Additionally, algorithms such as those described here have the potential to revolutionize product development within hair care companies through the streamlining of analysis and increasing accuracy thus informing better decision making, as well as increasing the pace of innovation.

ACKNOWLEDGEMENTS

This work was funded by The Procter & Gamble Company (Cincinnati, OH, USA).

ORCID

Jarek P. Sacha  <https://orcid.org/0000-0001-6705-4669>

Michael G. Davis  <https://orcid.org/0000-0002-4756-3295>

REFERENCES

1. J. A. Harness, B. Kohut, J. Garner, W. Canfield, D. Canfield & A. P. Bertolino Evaluation of Hair Count and Thickness Measurements in Male and Female Pattern Hair Loss (Androgenetic Alopecia) Using a Computer-Assisted Technique. in European Hair Research Society, 11th Annual Meeting, Zurich, Switzerland, July 7-9, (2005).
2. M. Saitoh, M. Uzuka, M. Sakamoto, Human hair cycle. *J. Investig. Dermatol.* 54(1), 65–81(1970).
3. U. Blume-Peytavi, D. A. Whiting, R. M. Trüeb, *Hair Growth and Disorders*, (Springer-Verlag, Berlin Heidelberg, 2008).
4. R. Hoffmann, TrichoScan: A novel tool for the analysis of hair growth in vivo. *J. Investig. Dermatol. Symp. Proc.* 8(1), 109–115(2003).
5. H. Shih, A precise automatic system for the hair assessment in hair-care diagnosis applications. *Skin Res. Technol.* 21, 500–507(2015).
6. A. Gallucci, D. Znamenskiy, N. Pezzotti, M. Petkovic, *Hair counting with deep learning*, (2020). in International Conference on Biomedical Innovations and Applications (BIA).
7. V. Lempitsky, A. Zisserman, *Learning To Count Objects in Images*, (2010). in Advances in Neural Information Processing Systems.
8. D. Bates, M. Mächler, B. Bolker, S. Walker. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Soft.* 67(1) (2015).
9. "VEOS SLR | Canfield Scientific," [Online]. Available: <https://www.canfieldsci.com/imaging-systems/veos-slr/>. Accessed 27 8 2019.
10. "Canfield EpiFlash," Canfield Scientific, [Online]. Available: <http://www.canfieldsci.com/FileLibrary/TBP-Brochure.pdf>. Accessed 18 11 2019.
11. K. He, G. Gkioxari, P. Dollár, R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, (Venice, Italy, 2017).
12. O. Ronneberger, P. Fischer, T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*, (2015).
13. B. D. Ripley, M. Thompson, Regression techniques for the detection of analytical bias. *Analyst.* 112, 377–383(1987).
14. J. M. Bland, D. G. Altman. Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* 8(2), 135–160(1999).
15. S. Nakagawa, P. C. D. Johnson, H. Schielzeth, The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J. R. Soc. Interface.* 14(134), 20170213(2017).
16. D. Liljequist, B. Elfving, K. S. Roaldsen. Intraclass correlation – A discussion and demonstration of basic features. *PLoS ONE.* 14(7), e0219854(2019).
17. B. M. Bolstad, R. A. Irizarry, M. Åstrand, T. P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics.* 19(2), 185–193(2003).

How to cite this article: J.P. Sacha, T.L.

Caterino, B.K. Fisher, G.J. Carr, R.S. Youngquist, B.M. D'Alessandro, A. Melione, D. Canfield, W.F. Bergfeld, M.P. Piliang, et al, Development and qualification of a machine learning algorithm for automated hair counting. *Int. J. Cosmet. Sci.* 43(Suppl. 1), 34–41 (2021). <https://doi.org/10.1111/ics.12735>