Molecular Vision 2023; 29:197-205 <http://www.molvis.org/molvis/v29/197>
Received 15 January 2023 | Accepted 15 October 2023 | Published 20 October 2023

Addressing bias in manual segmentation of spheroid sprouting assays with U-Net

Julian Rapp, Daniel Böhringer, Günther Schlunck, Hansjürgen Agostini, Thomas Reinhard, Felicitas Bucher

Eye Center, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Germany

Purpose: Angiogenesis research faces the issue of false-positive findings due to the manual analysis pipelines involved in many assays. For example, the spheroid sprouting assay, one of the most prominent in vitro angiogenesis models, is commonly based on manual segmentation of sprouts. In this study, we propose a method for mitigating subconscious or fraudulent bias caused by manual segmentation. This approach involves training a U-Net model on manual segmentations and using the readout of this U-Net model instead of the potentially biased original segmentations. Our hypothesis is that U-Net will mitigate any bias in the manual segmentations because this will impose only random noise during model training. We assessed this idea using a simulation study.
Methods: The training data comprised 1531 phase contrast images and manual segmentations from various spheroid sprouting assays. We randomly divided the images 1:1 into two groups: a fictitious intervention group and a control group. Bias was simulated exclusively in the intervention group. We simulated two adversarial scenarios: 1) removal of a single randomly selected sprout and 2) systematic shortening of all sprouts. For both scenarios, we compared the original segmentation, adversarial segmentation, and respective U-Net readouts. In the second step, we assessed the sensitivity of this approach to detect a true positive effect. We sampled multiple treatment and control groups with decreasing treatment effects based on unbiased ground truth segmentation.
Results: This approach was able to mitigate bias in both adversarial scenarios. However, in both scenarios, U-Net detected the real treatment effects based on a comparison to the ground truth.
Conclusions: This method may prove useful for verifying positive findings in angiogenesis experiments with a manual analysis pipeline when full investigator masking has been neglected or is not feasible.

Investigations into angiogenesis in retinal diseases are often hindered by manual analysis pipelines that risk inconsistent results [1]. Typical in vivo experiments studying vasoproliferative eye diseases, such as the mouse model of oxygen-induced retinopathy, require manual labeling of neovascularization [2]. This process can introduce bias. Furthermore, in vitro assays commonly used to examine angiogenesis, such as migration assays that require manual tracking of cell movement or sprouting assays that involve manually marking endothelial cell sprouts [3], face similar issues of potential biased analysis. Concerns about the reproducibility of results have been raised in the scientific community, leading to the concept of a "reproducibility crisis" [4]. According to an anonymous survey, more than 40% of researchers view poor data analysis, which can introduce unconscious bias, as an important contributing factor to irreproducibility [4]. Improved data analysis techniques that minimize bias have been identified as a major tool for enhancing reproducibility [5].

Correspondence to: Daniel Böhringer, Klinik für Augenheilkunde Killianstraße 5 79106 Freiburg; Phone: +49 761 270 40514; FAX: +49 761 270 40520; email: daniel.boehringer@uniklinik-freiburg.de

The spheroid sprouting assay is a widely used in vitro assay for investigating angiogenesis [6], but its manual analysis pipeline makes it prone to error. Currently, images of sprouts extending from cell spheroids in a collagen matrix are manually segmented using generic software, such as ImageJ [7,8]. For quantification, the median cumulative sprouting length per spheroid (CSLPS) is typically calculated manually. Validated software for automated segmentation and analysis is not commonly available. Although the analysis of medical images based on artificial intelligence (AI) has been applied in many areas, especially for clinical images [9-11], automated analysis of sprouting assays using the U-Net convolutional artificial neural network has only recently been reported [12,13]. Released in 2015, U-Net is a widely used network for image segmentation, especially phase contrast images [14,15]. However, to our knowledge, an automated analysis pipeline for sprouting assays is not publicly available. Moreover, generalizing such approaches across laboratories or experiments is challenging due to variations in cell morphology, imaging equipment, image processing, and protocols among laboratories [16,17]. Adapting an approach to individual laboratories would require training and calibration to achieve the needed quality, which is time-consuming and necessitates significant expertise. Scientists still rely primarily on manual

analysis of sprouting assays due to the challenges of proper automated segmentation. Masking sprout segmentation is difficult when the intervention is strong and is therefore directly evident from single images, for example, following vascular endothelial growth factor (VEGF) treatment [18,19]. Curved, faint, or bifurcating sprouts and numerous short sprouts also complicate analysis and introduce ambiguities [6,12,20]. Even experienced scientists may have difficulty ensuring consistent quantification [21,22].

We propose an automated method to improve the objectivity of manual segmentation in the spheroid sprouting assay, which can analyze whether studies report false-positive data based on systematic bias. This method could be used by auditors, supervisors, or journal editors to differentiate false positives from true positives with minimal effort. To achieve this, we propose training a disposable U-Net model on manual segmentations and using the U-Net output instead of the original manual segmentations. We hypothesize that U-Net will remove any systematic bias, provided that bias is present in less than every second image and that the bias is more or less subtle. From U-Net's perspective, any bias appears as random noise, as the network is unaware of group assignments during training. We believe that this circular U-Net segmentation will be of sufficient quality, even when the neural network is trained on partially erroneous segmentations. We evaluated this using actual sprouting data and two types of synthetic adversarial bias in a simulation.

METHODS

Image acquisition and processing: We obtained a training data set containing 1531 phase contrast images of spheroid sprouting assays, as well as manual segmentations. Images were acquired using a standard protocol [7,8,18].

Briefly, 200,000 human umbilical vein endothelial cells (HUVECs, Cat#: CC-2519, Lonza, Basel, Switzerland) of the sixth passage cultivated in endothelial cell growth medium (EGM, Cat#: CC-3124, Lonza) were suspended in 10 ml EGM containing 0.25% carboxy-methylcellulose (Methylcellulose, Cat#: M0512, Sigma-Aldrich, St. Louis, MO). Spheroids were formed in hanging drops of a volume of 25 µl incubated overnight and seeded on the following day in 0.5 ml of a three-dimensional collagen matrix consisting of 44.4% collagen (Collagen 1, Rat Tail, Cat#: 354,236, Corning, Corning, NY), 43.9% endothelial cell growth basal medium (EBM, Cat#: CC-3121, Lonza), 2.25% fetal bovine serum (FBS, Cat#: S0615, Lot#: 0453Z, Biochrome, Berlin, Germany), and 0.55% carboxy-methylcellulose in 24-well plates. The collagen was titrated to a physiologic pH by using NaOH (sodium hydroxide, Cat#: P031.2, Roth, Karlsruhe, Germany) and buffered at the final pH by using 1 µl of a 1M HEPES buffer (HEPES Buffer, Cat#: P05-P01100, PAN Biotech, Aidenbach, Germany). After the gel was solidified for 30 min at 37 °C and in 5% CO_2, it was layered with cytokines suspended in 0.1 ml EBM. Images of spheroids were taken the next day using an inverse microscope (Zeiss Axio Vert. A1, Oberkochen, Germany) and ProgPres CapturePro 2.10.0.1 imaging software (Jenoptik Optical Systems, Jena, Germany). All spheroids in each well were photographed. Sprout length was manually measured in all images by the same investigator with ImageJ Fiji and its measuring tool. Consistent guidelines were implemented. The median CSLPS was calculated by summing the lengths of all the sprouts of each spheroid and taking the mean across the spheroids for each condition. A higher CSLPS indicates greater angiogenic potential.

ImageMagick software was used to extract labels of the manually annotated spheroid sprouts using a global threshold. The binary labels were skeletonized to enable pixel-based quantification. This was performed with the thinning operator Skeleton:3 structuring element and five iterations (Figure 1A).

Creation of biased segmentation: All phase contrast images from the spheroids were randomly assigned to either group 1 (the control group; n = 766) or group 2 (the intervention group; n = 765; Figure 1A). To test our hypothesis that U-Net can remove adversarial bias, we generated systematic bias between the images of group 1 and group 2, as a biased experimenter might do. In adversarial approach 1, the segmentation of group 2 was altered by removing one randomly selected sprout, resulting in a systematically biased set of images. This was achieved with R using the bwlabel function of the EBImage package (https://bioconductor.org/packages/release/bioc/html/EBImage.html). To check for robustness against other kinds of bias, we also generated adversarial approach 2, in which we shortened each sprout segmentation by several pixels, but again exclusively in group 2. This was achieved via the morphologic operators of ImageMagick. We used a thinning operator in conjunction with a LineEnds hit-or-miss structuring element. In both approaches, we refer to the biased group as group 2b (n = 765; Figure 1A). The group 2b segmentation consequently comprises fewer pixels compared to group 2, mimicking subtle bias.

Training of U-Net: We used unmodified U-Net for all experiments. Training comprised 500 epochs, with 200 iterations in each epoch. We used adversarially modified segmentations for training (Figure 1A). We trained both adversarial scenarios separately.

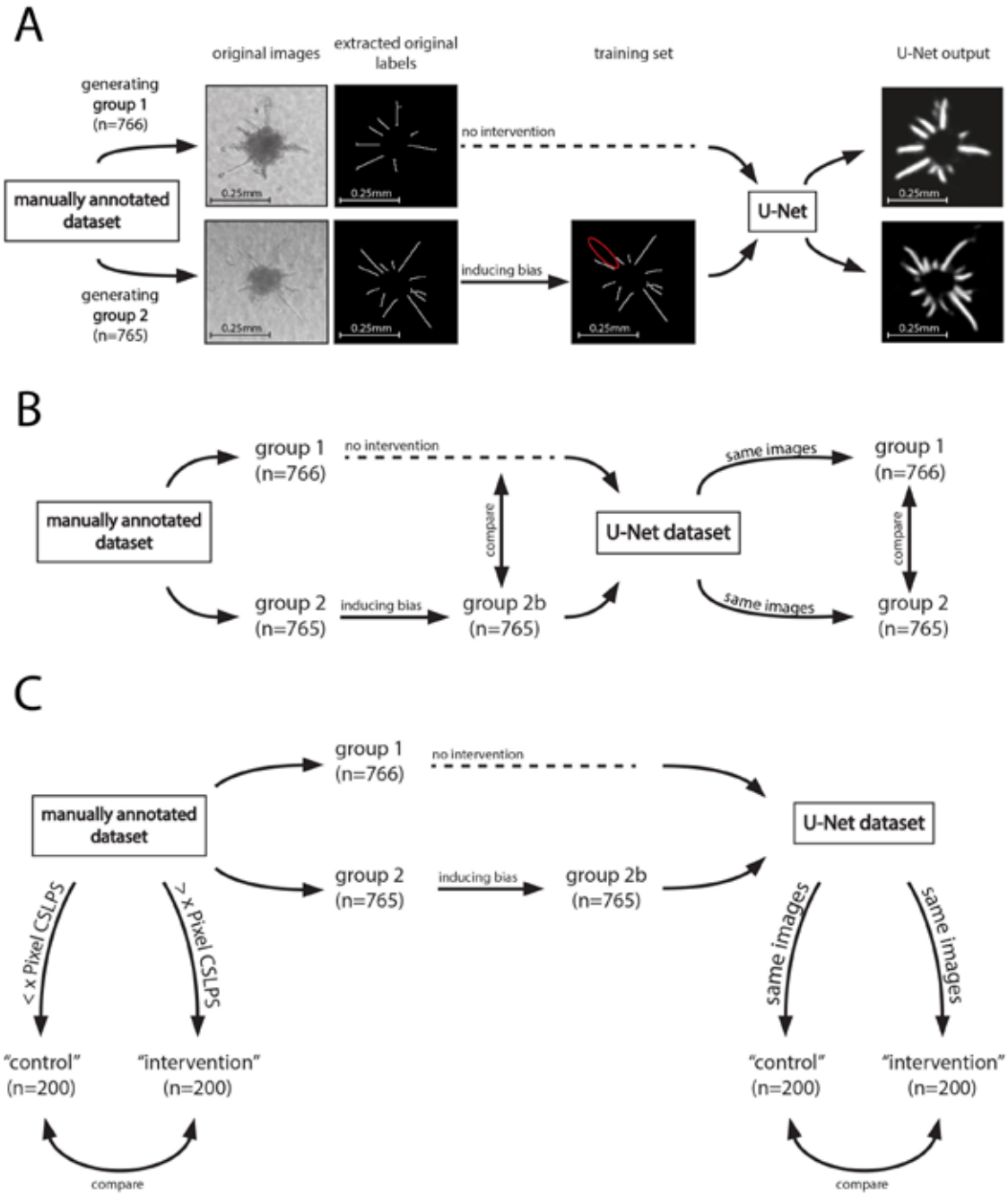

Figure 1. Visualization of the methods for evaluating U-Net's ability to remove systematic bias in a data set. **A**: A pipeline to train a U-Net model on a partially biased data set. Images were randomly assigned to group 1 or group 2. Systematic bias was induced in group 2 by removing one random sprout in every other image, resulting in group 2b. The U-Net model was trained on images from group 1 and group 2b and then used to reanalyze the same images. **B**: The approach to assess U-Net's ability to mitigate an adversarial bias. Images from group 1 and group 2b were compared to determine if there were statistically significant differences, indicating sufficient bias. The U-Net model was then trained on these images and used to reanalyze them. The results from the two groups were again compared to determine if there were statistically significant differences. **C**: The method for testing U-Net's ability to detect a true positive effect. Fictitious control and intervention groups were generated based on an unbiased, manually annotated ground truth data set. They had a statistically significant true positive difference. A biased data set was then generated, and the U-Net model was trained on the data set. The U-Net model was then used to reanalyze the same images to determine if it could also detect a statistically significant true positive difference despite being trained on the biased data set.

Inference from U-Net to calculate the CSLPS: The U-Net readouts were thresholded and skeletonized using image morphology operators from ImageMagick to enable comparisons with ground truth segmentation, as described above. The CSLPS was calculated by counting all white pixels. The CSLPS was compared between the groups to assess how far the adversarial bias was mitigated (Figure 1B).

Sensitivity of U-Net: The sensitivity of the U-Net model to detect differences in the CSLPS from interventions was evaluated by simulating several fictitious experiments. For this purpose, we sampled 200 images each from the total data set based on the ground truth segmentation (Figure 1C). These groups reflect fictitious controlled experiments with significant responses to spheroid sprouting. We sampled three control and intervention data sets, each with different effect sizes (Table 1). The first group comprised a strong intervention, with the threshold set to >200 pixels for the CSLPS for the control group, and the threshold set to <400 pixels for the CSLPS for the intervention group (Figure 1C). The second group comprised a medium response, with a threshold of >100 pixels for the CSLPS for the control group and a threshold of <500 pixels for the CSLPS for the intervention group. The third group comprised a weak experimental response, with a threshold of >50 pixels for the CSLPS for the control group and a threshold of <600 pixels for the CSLPS for the intervention group. We deliberately allowed overlap between the groups, as this is a key characteristic of sprouting experiments with the endpoint CSLPS. The sampling constraints are summarized in Table 1. We applied this approach to evaluate the sensitivity of the U-Net models that had been trained on labels with two different kinds of bias.

Statistical analysis: A total of 1531 images were included in the data and assigned to group 1 (n = 766) or group 2 (n = 765). Bars represent the mean, and error bars visualize the standard error of the mean. Unless stated otherwise, a Welch two-sample *t* test was used to evaluate the statistical significance. The alpha level was set at 0.05.

RESULTS

We first implemented an adversarial approach, in which one sprout was eliminated from every second image to systematically introduce bias (adversarial approach 1). This resulted in a statistically significant 15.5% decrease in the CSLPS between the group 1 and group 2b images (a difference of 47 pixels, $p<0.05$, Figure 2A). This result demonstrated that we successfully introduced bias, which resulted in false-positive results. We then trained a U-Net model on the biased segmentations and used it to reanalyze the group 1 and group 2 images. The U-Net model substantially mitigated the bias, eliminating the statistically significant difference in the CSLPS (difference of 12 pixels, $p = 0.12$, Figure 2A), demonstrating that U-Net was able to recover the adversarially deleted sprouts. Similarly, shortening all sprouts in every second image during adversarial approach 2 resulted in a 41% decrease in the CSLPS that was statistically significant (difference of 114 pixels, $p<0.05$, Figure 2B), again demonstrating a high potential for false-positive results due to bias. Again, U-Net successfully mitigated this bias, eliminating the statistical significance (difference of 13 pixels, $p = 0.0539$, Figure 2B). These results show that U-Net can generalize across different types of bias as long as the bias is present in every second image of the training data. This approach uncovered the adversarial bias in both approaches and performed well in recovering the manipulated labels. Nevertheless, it must be shown that U-Net also provides sufficiently good results, despite partially erroneous training data. This requires a sensitivity analysis.

Sensitivity of U-Net to detect true treatment effects: To evaluate the sensitivity of the U-Net models trained on biased data sets in detecting true treatment effects, we simulated data sets with varying magnitudes of true differences between the control and intervention groups. Specifically, we generated groups of 200 random images based on ground truth segmentation, which had known differences in their CSLPSs. The thresholds for the groups are summarized in Table 1. We then used the U-Net model trained on the biased data sets to determine whether it could detect differences between the control and intervention groups, thus detecting true treatment effects.

Table 1. Thresholds for subpopulations.

Strength of intervention	"Control" CSLPS [pixel]	"Intervention" CSLPS [pixel]
strong intervention (loose threshold)	<400	>200
medium intervention (medium threshold)	<500	>100
weak intervention (strict threshold)	<600	>50

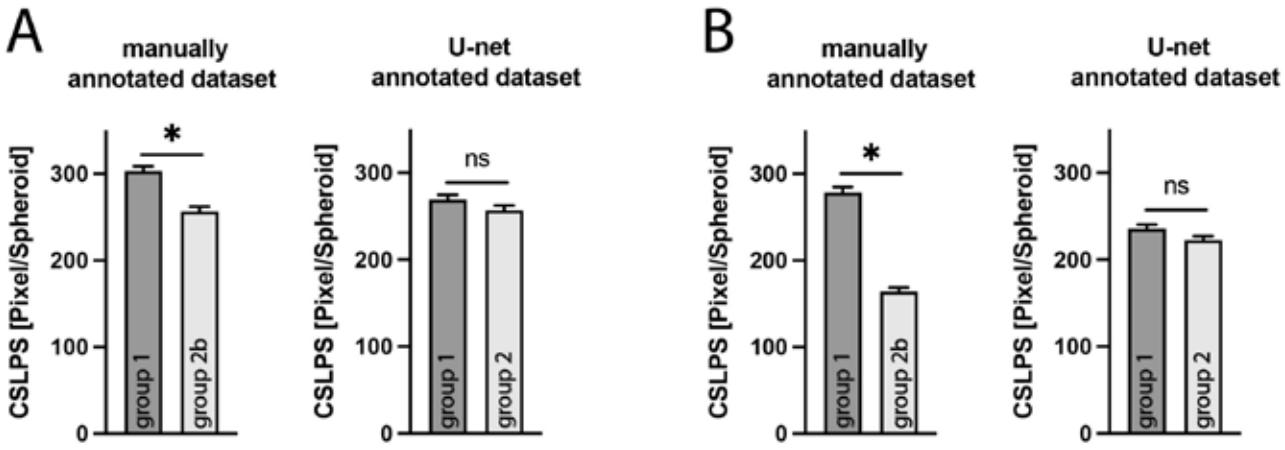

Figure 2. U-Net is able to mitigate adversarial bias. **A**: Evaluation of the statistical significance of the induced bias from removing one randomly selected sprout in all images of group 2, resulting in biased group 2b (adversarial approach 1) compared to the U-Net readout trained on this biased data set. **B**: Assessment of the difference between group 1 and group 2b after adversarial bias was induced by removing two pixels from every sprout of group 2, resulting in group 2b (adversarial approach 2). These results were compared to the readout from the U-Net model trained on this systematically biased data set.

We first evaluated the U-Net model trained on a biased data set, in which one random sprout was removed from every second image (adversarial approach 1). We simulated a large effect, resulting in a 199.4 pixel difference in the CSLPS between the control and intervention groups, which was statistically significant ($p<0.05$, Figure 3A). The U-Net model detected a similar large effect, yielding a 153.5 pixel difference ($p<0.05$, Figure 3A). For the moderate and small effect simulations, the U-Net model generated differences of 64.1 and 36.6 pixels, respectively (both $p<0.05$), comparable to the 83.8 and 52.8 pixel differences in the ground truth (both $p<0.05$, Figure 3A).

Similarly, the U-Net model trained on a biased data set, in which all sprouts were shortened in every second image (adversarial approach 2), detected the simulated interventions. For the strong intervention, U-Net reported a 125.1 pixel difference ($p<0.05$, Figure 3B), comparable to the 195.5 pixel difference in the ground truth ($p<0.05$). For the moderate and weak interventions, U-Net reported 51.1 and 35.8 pixel differences (both $p<0.05$), comparable to the 66.2 and 44.4 pixel differences in the ground truth.

These results suggest that although U-Net can mitigate bias, the neural network remains sensitive enough to detect true treatment effects despite not being trained on perfect ground truth labels. However, the U-Net model generally reported a lower CSLPS when compared to the ground truth. Nevertheless, the relative differences between the control and intervention groups were comparable between the U-Net model and ground truth, as summarized in Table 2. In total, the U-Net model trained on the segmentation from adversarial scenario 1 was able to detect an intervention with a relative difference of 16.1% between the control and intervention groups, which had approximately the same strength as the induced bias (15.5%, Figure 2A). Similarly, the U-Net model in adversarial scenario 2 yielded similar results by detecting a difference of 14.3%, while the strength of the bias was 41% (Figure 2B).

DISCUSSION

This method substantially mitigated two kinds of simulated systematic biases, resembling those of inexperienced or fraudulent investigators, to a large extent. At the same time, the method proved sensitive enough to detect real experimental responses robustly based on the CSLPS. Notably, no explicit manual input is required. Therefore, the proposed approach can be used in any laboratory without substantial up-front labor costs.

To our knowledge, this is the first time U-Net has been used solely to recreate the segmentation it was trained on to remove bias. The U-Net models were obviously unable to learn the artificial biases, most likely because they were present in only every second image. We assume that the few erroneous parts of the segmentation, present in only every second image, were overruled by most of the correct pixels. This aligns with the observation that U-Net can handle random noise in a training data set to some degree [23-25].

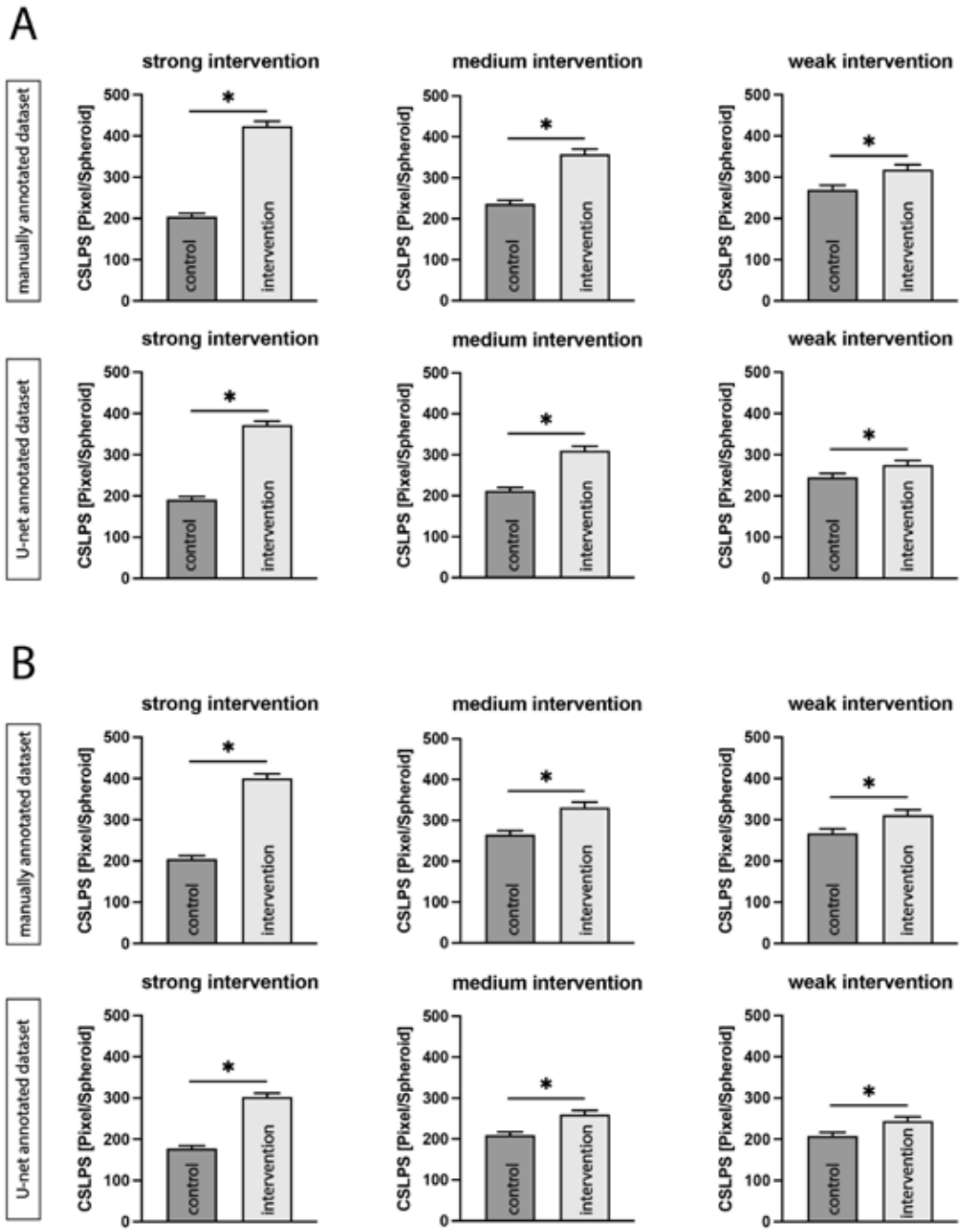

Figure 3. The U-Net model trained on biased data is sensitive enough to quantify the strength of interventions. **A**: Analysis of U-Net's ability to detect true positive differences in controlled sprouting experiments with different strengths of intervention. The U-Net model was trained on the biased segmentation generated by removing one random sprout in group 2 (adversarial approach 1). The data comprised 200 samples from the data set selected based on the ground truth cumulative sprouting length per spheroid (CSLPS; manually annotated data set). The U-Net readout for the same images is shown (the U-Net annotated data set). **B**: Evaluation of the sensitivity of the U-Net model trained on biased segmentation generated by removing two pixels from every sprout in group 2 (adversarial approach 2). Three different strengths of the intervention were simulated by selecting 200 random images for the control and intervention groups based on the ground truth CSLPS (the manually annotated data set). The U-Net readout for the same images was then compared (the U-Net annotated data set).

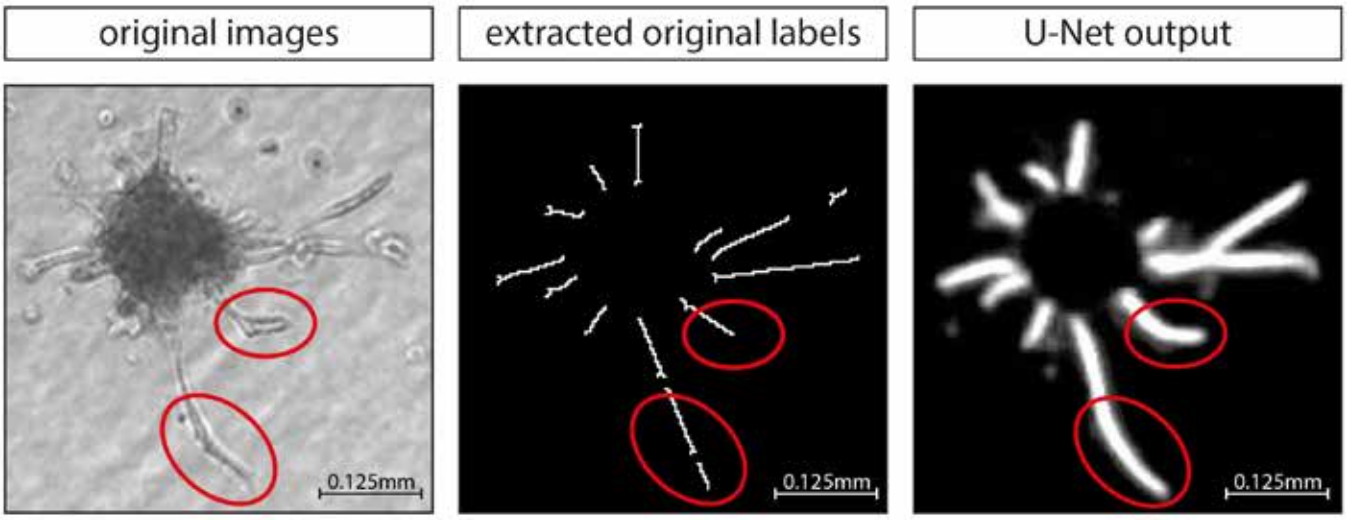

Figure 4. The U-Net model canonically segments bending sprouts, although it was trained on linear segmentation. The examples demonstrate U-Net's ability to detect bending sprouts and correct for systematic bias by using only straight lines in the manual analysis.

Interestingly, the present U-Net models were trained with imperfect sprout segmentations due to the technical limitations of the manual analysis pipeline. The original ground truth segmentations were based on straight lines pointing from the sprout base to the sprout tip, even if the sprouts were curved to some degree. The U-Net readout, however, always followed the actual sprout contours (Figure 4). This emphasizes the robustness of U-Net against imperfect training segmentations. However, U-Net-based CSLPS cannot be used as a direct replacement for the original CSLPS because of this and cannot be used as a fully compatible decrease in replacement. Instead, the proposed method should be applied to assess differences between the control and intervention groups from the data set on which the model is trained, that is, to check whether the reported finding is truly positive or caused by biased labels. When this check is passed, the interpretation can be performed based on the original manual segmentations.

The idea of using U-Net solely to recreate training segmentation makes this method applicable to more angiogenesis assay analysis pipelines, as it can work on any modality that yields images suitable for U-Net segmentation. The only requirements are that segmentation has already been performed, that the control-to-intervention ratio is close to 1:1, and that sufficient images are available for training. The simulated experiments indicated that the proposed approach can reveal and even sufficiently mitigate biases of up to 41% favoring the intervention group in a 1:1 experimental-to-control setting. However, this may not apply in settings with lower control-to-intervention ratios. It also remains unclear whether stronger biases can be sufficiently mitigated using this method. However, we believe bias can be strongly

Table 2. The performance by the U-Net to detect real intervention.

Strength of intervention	Data set where bias was induced by removing on sprout in "group 2"				Data set where bias was induced by removing 2 pixels per sprout in "group 2"			
	Ground truth segmentation		U-Net segmentation		Ground truth segmentation		U-Net segmentation	
	absolute Δ [pixels]	relative Δ	absolute Δ [pixels]	relative Δ	absolute Δ [pixels]	relative Δ	absolute Δ [pixels]	relative Δ
strong intervention	199.4	49%	153.5	43.3%	195,5	48.8%	125.1	41.3%
medium intervention	83.8	24.8%	64.1	21.4%	66.2	20%	51.1	19.6%
weak intervention	52.8	16.1%	36.6	12.9%	44.4	14.3%	35.8	14.7%

suspected when the difference in treatment response between the original CSLPS and the U-Net-based CSLPS is large.

Deep learning frameworks for scientific image segmentation have proliferated lately. However, almost all new architectures are derivatives of the original U-Net neural network tailored for specific tasks. In this study, therefore, we used the original U-Net neural network because of its generality and proven effectiveness, especially with phase contrast images [14,15]. In this context, AURA-Net seems to be a promising alternative because it is based on a pretrained encoder in combination with an Attention-U-Net decoder [26]. Due to extensive pretraining, AURA-Net could especially help to work around the most relevant limitation of the proposed method, that is, the availability of enough labeled images to train a U-Net model. This was not a problem in our proof-of-concept study, as we used images pooled from several real experiments in our laboratory. When small experiments do not yield sufficient training data to properly train a neural network, the recently published segment-anything model from Facebook research may be a novel option to objectivize manual readouts using few shot learning.

In summary, the proposed approach has the potential to automatically detect and correct for bias from manual segmentation in the analysis of spheroid sprouting experiments. This approach could be applied in other fields of research with manual analyses. The proposed approach can increase users' confidence that positive findings are not based on bias or even fraud. The method may be useful for auditors, supervisors, coauthors, and journal editors.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft [Bu3135/3-1to F.B.], the German Retina Society [Dr. Werner-Jackstädt-Nachwuchspreis to F.B.] and the Medizinische Fakultät der Albert-Ludwigs- Universität Freiburg [Berta-OttensteinProgram for Clinician Scientists and Advanced Clinician Scientists to F.B., MOTIVATE Program to J.R.].

REFERENCES

1. Gariano RF, Gardner TW. Retinal angiogenesis in development and disease. Nature 2005; 438:960-6. [PMID: 16355161].
2. Smith LE, Wesolowski E, McLellan A, Kostyk SK, D'Amato R, Sullivan R, D'Amore PA. Oxygen-induced retinopathy in the mouse. Invest Ophthalmol Vis Sci 1994; 35:101-11. [PMID: 7507904].
3. Nowak-Sliwinska P, Alitalo K, Allen E, Anisimov A, Aplin AC, Auerbach R, Augustin HG, Bates DO, van Beijnum JR, Bender RHF, Bergers G, Bikfalvi A, Bischoff J, Böck BC, Brooks PC, Bussolino F, Cakir B, Carmeliet P, Castranova D, Cimpean AM, Cleaver O, Coukos G, Davis GE, De Palma M, Dimberg A, Dings RPM, Djonov V, Dudley AC, Dufton NP, Fendt SM, Ferrara N, Fruttiger M, Fukumura D, Ghesquière B, Gong Y, Griffin RJ, Harris AL, Hughes CCW, Hultgren NW, Iruela-Arispe ML, Irving M, Jain RK, Kalluri R, Kalucka J, Kerbel RS, Kitajewski J, Klaassen I, Kleinmann HK, Koolwijk P, Kuczynski E, Kwak BR, Marien K, Melero-Martin JM, Munn LL, Nicosia RF, Noel A, Nurro J, Olsson AK, Petrova TV, Pietras K, Pili R, Pollard JW, Post MJ, Quax PHA, Rabinovich GA, Raica M, Randi AM, Ribatti D, Ruegg C, Schlingemann RO, Schulte-Merker S, Smith LEH, Song JW, Stacker SA, Stalin J, Stratman AN, Van de Velde M, van Hinsbergh VWM, Vermeulen PB, Waltenberger J, Weinstein BM, Xin H, Yetkin-Arik B, Yla-Herttuala S, Yoder MC, Griffioen AW. Consensus guidelines for the use and interpretation of angiogenesis assays. Angiogenesis 2018; 21:425-532. [PMID: 29766399].
4. Baker M. 1,500 scientists lift the lid on reproducibility. Nature 2016; 533:452-4. [PMID: 27225100].
5. Nuzzo R. How scientists fool themselves - and how they can stop. Nature 2015; 526:182-5. [PMID: 26450039].
6. Zahra FT, Choleva E, Sajib MS, Papadimitriou E, Mikelis CM. In Vitro Spheroid Sprouting Assay of Angiogenesis. Methods Mol Biol 2019; 1952:211-8. [PMID: 30825177].
7. Buehler A, Sitaras N, Favret S, Bucher F, Berger S, Pielen A, Joyal JS, Juan AM, Martin G, Schlunck G, Agostini HT, Klagsbrun M, Smith LE, Sapieha P, Stahl A. Semaphorin 3F forms an anti-angiogenic barrier in outer retina. FEBS Lett 2013; 587:1650-5. [PMID: 23603393].
8. Pfisterer L, Korff T. Spheroid-Based In Vitro Angiogenesis Model. Methods Mol Biol 2016; 1430:167-77. [PMID: 27172953].
9. Agn M, Munck Af Rosenschöld P, Puonti O, Lundemann MJ, Mancini L, Papadaki A, Thust S, Ashburner J, Law I, Van Leemput K. A modality-adaptive method for segmenting brain tumors and organs-at-risk in radiation therapy planning. Med Image Anal 2019; 54:220-37. [PMID: 30952038].
10. Chen Y, Li D, Zhang X, Jin J, Shen Y. Computer aided diagnosis of thyroid nodules based on the devised small-datasets multi-view ensemble learning. Med Image Anal 2021; 67:101819[PMID: 33049580].
11. Luo L, Yu L, Chen H, Liu Q, Wang X, Xu J, Heng PA. Deep Mining External Imperfect Data for Chest X-Ray Disease Screening. IEEE Trans Med Imaging 2020; 39:3583-94. [PMID: 32746106].
12. Chen Z, Ma N, Sun X, Li Q, Zeng Y, Chen F, Sun S, Xu J, Zhang J, Ye H, Ge J, Zhang Z, Cui X, Leong K, Chen Y, Gu Z. Automated evaluation of tumor spheroid behavior in 3D culture using deep learning-based recognition. Biomaterials 2021; 272:120770[PMID: 33798957].
13. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. in International Conference on Medical image computing and computer-assisted intervention. 2015. Springer.

14. Daniel MC. Automated segmentation of the corneal endothelium in a large set of 'real-world'specular microscopy images using the U-Net architecture. Sci Rep 2019; 9:1-7. .
15. Punn NS, Agarwal S. Inception u-net architecture for semantic segmentation to identify nuclei in microscopy cell images. ACM Trans Multimed Comput Commun Appl 2020; 16:1-15. TOMM.
16. Meijering E. Cell Segmentation: 50 Years Down the Road IEEE Signal Process Mag 2012; 29:140-5. Life Sciences.
17. Shuvaev SA, Lazutkin AA, Kedrov AV, Anokhin KV, Enikolopov GN, Koulakov AA. DALMATIAN: An Algorithm for Automatic Cell Detection and Counting in 3D. Front Neuroanat 2017; 11:117-[PMID: 29311849].
18. Bucher F, Walz JM, Bühler A, Aguilar E, Lange C, Diaz-Aguilar S, Martin G, Schlunck G, Agostini H, Friedlander M, Stahl A. CNTF Attenuates Vasoproliferative Changes Through Upregulation of SOCS3 in a Mouse-Model of Oxygen-Induced Retinopathy. Invest Ophthalmol Vis Sci 2016; 57:4017-26. [PMID: 27494343].
19. Shah S, Kang K-T. Two-Cell Spheroid Angiogenesis Assay System Using Both Endothelial Colony Forming Cells and Mesenchymal Stem Cells. Biomol Ther (Seoul) 2018; 26:474-80. [PMID: 30157615].
20. Kunz-Schughart LA, Kreutz M, Knuechel R. Multicellular spheroids: a three-dimensional in vitro culture system to study tumour biology. Int J Exp Pathol 1998; 79:1-23. [PMID: 9614346].
21. Millioni R, Sbrignadello S, Tura A, Iori E, Murphy E, Tessari P. The inter- and intra-operator variability in manual spot segmentation and its effect on spot quantitation in two-dimensional electrophoresis analysis. Electrophoresis 2010; 31:1739-42. [PMID: 20408132].
22. Rizwan I. Haque, I. and J. Neubert. Deep learning approaches to biomedical image segmentation. Inform Med Unlocked 2020; 18:100297.
23. Zhong T. Seismic random noise suppression by using deep residual U-Net. J Petrol Sci Eng 2022; 209:109901.
24. Heller N, Dean J, Papanikolopoulos N. Imperfect segmentation labels: How much do they matter? in Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. 2018, Springer. p. 112–120.
25. Karimi D, Dou H, Warfield SK, Gholipour A. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. Med Image Anal 2020; 65:101759[PMID: 32623277].
26. Cohen E, Uhlmann V. Aura-net: robust segmentation of phase-contrast microscopy images with few annotations. in 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). 2021. IEEE.

Articles are provided courtesy of Emory University and the Zhongshan Ophthalmic Center, Sun Yat-sen University, P.R. China. The print version of this article was created on 20 October 2023. This reflects all typographical corrections and errata to the article through that date. Details of any changes may be found in the online version of the article.