DeepProjection: Specific and robust projection of curved 2D tissue sheets from 3D microscopy using deep learning

Daniel Haertter, Xiaolei Wang, Stephanie M. Fogerson, Nitya Ramkumar, Janice M. Crawford, Kenneth Poss, Stefano Di Talia, Daniel Kiehart and Christoph F Schmidt
DOI: 10.1242/dev.200621

Editor: Thomas Lecuit

Review timeline
Original submission: 10 February 2022
Editorial decision: 31 March 2022
First revision received: 6 July 2022
Accepted: 12 September 2022

Original submission

First decision letter

MS ID#: DEVELOP/2022/200621

MS TITLE: DeepProjection: Specific and robust projection of complex 2D tissue sheets from 3D microscopy using deep learning

AUTHORS: Daniel Haertter, Xiaolei Wang, Stephanie M Fogerson, Nitya Ramkumar, Janice M Crawford, Kenneth Poss, Stefano Di Talia, Daniel Kiehart, and Christoph F Schmidt

I have now received all the referees reports on the above manuscript, and have reached a decision. The referees' comments are appended below, or you can access them online: please go to BenchPress and click on the 'Manuscripts with Decisions' queue in the Author Area.

The overall evaluation is positive and we would like to publish a revised manuscript in Development, provided that the referees' comments can be satisfactorily addressed. Please attend to all of the reviewers' comments in your revised manuscript and detail them in your point-by-point response. If you do not agree with any of their criticisms or suggestions explain clearly why this is so. If it would be helpful, you are welcome to contact us to discuss your revision in greater detail. Please send us a point-by-point response indicating your plans for addressing the referee's comments, and we will look over this and provide further guidance.

Reviewer 1

Advance summary and potential significance to field

The manuscript "DeepProjection: Specific and robust projection of complex 2D tissue sheets from 3D microscopy using deep learning", uses a neural network to extract 2D surfaces from volumetric imaging.
Concretely, the authors carve out curved epithelial sheets from z-stacks acquired with a confocal microscope. They use a convolutional neural network to generate binary masks which mark the region of interest in the 3D data. These masks are used to block out-of-plane signal before computing z-projections to improve the quality of the resulting image. The neural network is

trained using manually labeled datasets. The authors illustrate their technique during Drosophila dorsal closure and zebrafish periderm development.
Extracting surfaces from volumetric data is a recurring task in quantitative developmental biology and the idea to use convolutional neural networks to assist in this process is sound. The figures shown by the authors demonstrate that their approach is working and show a convincing improvement of image quality after treatment.

Comments for the author

Major:
Although machine learning algorithms in general, and convolutional neural networks specifically have been proposed for image classification before, we still recognize the potential of these methods for use in developmental biology. The manuscript would likely benefit from clearly demonstrating the power of the chosen approach. One attempt is presented in the form of a comparison to a variety of surface detection approaches in the literature. We suggest to extended this list to include another machine learning algorithm called “Ilastik”, and discuss the result quality, as well as training time needed.
An example where the deep projection algorithm might come into its own is when processing very large datasets. Existing machine learning approaches such as Ilastik require some user labeling for each dataset to achieve optimal quality. Deep projection might have a competitive advantage in such situations. Perhaps the manuscript could reach a greater readership by including a discussion that ideally involves one additional dataset: Commonly in development tissues change shape. It would be interesting to see how deep projection performs on such time series data. It would also be interesting to speculate about how the algorithm handles a large number of fixed samples, as it occurs e.g. during a screen. In both cases, manual labeling might quickly become cumbersome, encouraging users to look for alternatives, which deep projection might very well be. In the current presentation, this might not be clear to a non-expert user.
On a more technical note, it appears 15 hours for manual training is extensive compared to other tools. It is not clear if training will have to be re-done for every different fluorescent marker, or embryo. Is it possible for the authors to reduce the amount of label data required by the model? For example, by reducing the complexity of their network, or by use of sparse labeled data? Is there potential to combine deep projection with Ilastik, by using user-assisted Ilastik segmentation to quicker generate a suitable training set, that then robustly predicts the rest of the data, without added manual training required?

Minor:
1) The detailed discussion about the network can be moved to an appendix (e.g. loss function, learning rate number of layers, …) since one would imagine it is more expert terrain than the typical readership of Development. Similarly, the mathematical discussion of “tissue unrolling” can be moved to the SI.
2) The authors should show the scaling of the model performance (preferably with a couple of representative
“projections”) with the size of the training set.
3) The authors note that the training set comprises data recorded by different scopes. Can they check how robustly transferable their model is? E.g. omit data from scope X from the training set, and then show how the model performs on images recorded by scope X, or even try the model on a slightly different fluorescent marker (e.g. different fluorescent markers labeling the same protein). Successfully demonstrating transferability would be a major plus.
4) For the end user, it would be useful if the authors supplied more detail about the procedures used to train and test their neural network. What size of the training and testing sets used, and should be targeted? Was a separate testing set, not used for model selection, used to generate the scores and images shown in the figures?

With kind regards,
Nikolas Claussen & Sebastian Streichan

Reviewer 2

Advance summary and potential significance to field

In their manuscript, the authors present DeepProjection, a new method to project 3D surfaces on a 2D plane using deep learning. As stated in their introduction, a good 2D projection can significantly ease the image analysis downstream. The need and interest for these kinds of methods is also reflected in the recent publications that the authors indeed mention.

The authors clearly explain their method and how they compared it to the state of the art. They successfully apply DeepProjection to drosophila and zebrafish embryos. In their manuscripts the authors also convincingly show the superiority of their method over the state of the art when it comes to quality results with a minor drawback on processing time which will probably get gained back during the downstream image analysis.

The method presented could be of significant interest to the community.

Comments for the author

Major comments:
The authors could, and maybe should, better show the superiority of their method by comparing the result of the extraction of biological information from their projection against other types of projection. For example by doing cell segmentation or by quantifying cell flow.
The authors do not comment on the continuity of their reconstructed projection over time. Though, when looking at the provided movie, it seems like there is some flickering indicating the fact that similar images might have significantly different projection (especially around the closing end of the tissue). This problem could prevent a good tracking of the cells. This could be avoided by, for example, averaging over time, the z values.

Minor comments:
The authors show images where the membranes are labelled, would their method work with nuclei data? If yes, it would strongly increase the impact of their work. Though, it would definitely be understandable if it does not transfer easily or at all.
L20: The authors mention that they create undistorted 2D projections from the original 3D shape. While the contextual meaning of undistorted is explained later on in the article as the fact that the created surface is continuous, without context it can be understood as the fact that the algorithm raises an undistorted 2D projection of an ellipsoidal shape.
L42: The authors state that it is “necessary” to create 2D projections to follow tissue sheet morphogenesis. While it is understandable that it can be of great help to do so, it might not be necessary. Or if indeed the authors think it is, they should justify their claim.
L79 and other places: the authors mention “crisp boundaries”. It is not completely clear what they mean by that. A definition of “crisp” in the context of this study would be welcome
L212: The authors briefly describe the correction of local distortion but do not fully disclose the method. Because it is a method paper, showing the method could be of great interest to the community.
Figure 3D.: It is unclear why the authors show MIP vs others rather than GT vs others, could they justify their choice?
Léo Guignard

First revision

Author response to reviewers' comments

Thank you very much for your supporting and helpful comments. Based on both referees’ feedback, we have carefully revised the manuscript and updated our algorithm. In summary:

- We implemented a class for mask post-processing to average masks over multiple frames to further improve time-consistency, offset masks in z-direction, and create unique tissue z-height maps.
- We demonstrate that averaging masks over several frames removes occasional flickering artifacts (Movie S1).
- We had previously accidentally omitted the unfolding algorithm in our Python package on GitHub. We have now added the unfolding to our package and added detailed instructions to the Jupyter notebook tutorial.
- We have moved all of the technical description to the methods section and added a short description that is aimed at non-experts to the result section.
- To better streamline the manuscript, we swapped Figure 1 and Figure 2 and adjusted the results section accordingly.
- To demonstrate the advantage of using DeepProjection as a pre-processing step prior to morphometric analysis, we added exemplary cell segmentation of the lateral epidermis tissue to Figure 2.
- Based on feedback from early users, we have improved the package API, fixed errors, and added support for all common tif-formats.

Responses to the reviewers' comments in detail:

Reviewer 1 (Nikolas Claussen & Sebastian Streichan) (reviewer comments in italics)**:**

Major:

*[We suggest to extend this list to include another machine learning algorithm called "**Ilastik**", and discuss the result quality, as well as training time needed. An example where the deep projection algorithm might come into its own is when processing very large datasets. Existing machine learning approaches such as Ilastik require some user labeling for each dataset to achieve optimal quality. Deep projection might have a competitive advantage in such situations.]*

Ilastik uses a random-forest classifier based on a set of features extracted by a pre-defined filter bank. We tested Ilastik (details see supplementary S1) – training on 5 images from our training data set. The resulting binary masks were very disconnected and did not have sharp edges. The resulting projections were of lesser quality than the DP result and had, e.g., holes and ragged edges. Increasing the number of training stacks did not improve performance.
We added a description of this comparison to the supplement (Figure S1).

Ilastik uses a very different approach than our DP deep learning algorithm and might be preferable for the quick analysis of medium-sized data sets with clear morphological boundaries. However, we show that the custom neural network used for DP performs much better in robustly classifying tissue regions and rejecting various fluorescence artifacts and background signals.

[Perhaps the manuscript could reach a greater readership by including a discussion that ideally involves one additional dataset: Commonly in development tissues change shape. It would be interesting to see how deep projection performs on such time series data.]

We have demonstrated that DP performs well with data from two experimental systems with vastly different geometries and cell morphologies, Drosophila dorsal closure and *Danio* periderm development. For dorsal closure, DP creates robust results throughout the whole closure process (see supplementary movie M1). We thus believe that DeepProjection also performs well for other systems with evolving geometries. Capturing and analyzing a complete data set of another experimental system would be beyond the scope of this paper.

[It would also be interesting to speculate about how the algorithm handles a large number of fixed samples, as it occurs e.g., during a screen. In both cases, manual labeling might quickly become cumbersome, encouraging users to look for alternatives, which deep projection might very well be. In the current presentation, this might not be clear to a non- expert user.]

For high-throughput screens with standardized samples, fixed and live, DP can definitely play out its strength, since the one-time effort to create a training data set is worthwhile the more the model is used. We have added a remark regarding high-throughput screenings to the manuscript.
[On a more technical note, it appears 15 hours for manual training is extensive compared to other tools. It is not clear if training will have to be re-done for every different fluorescent marker, or embryo. Is it possible for the authors to reduce the amount of label data required by the model? For example, by reducing the complexity of their network, or by use of sparse labeled data?]

The training data set used for *Drosophila* dorsal closure contained 160 stacks for various imaging conditions: Three microscopes using cameras with different pixel sizes and data qualities, three different fluorescent markers (E-cadherin ubicad-promotor and endogenous knock-in, E-cadherin-RFP), four fly lines (wt, mutants, …), different stages of DC (early, middle, late). The time for masking one stack is around 5 min, so we estimated a total of 15 h.

However, for just one fly line and one imaging condition, a much smaller data set will suffice. To demonstrate this feature, we trained DeepProjection on two distinct training data sets with each containing just five stacks (see supplementary information, Fig. S2) for (1) endogenous E-cadherin-GFP knock-in on a microscope with pixel size 0.16 µm (2300x1044), and (2) ubicad-E-cadherin-GFP on a different microscope with pixel size 0.36 µm (672x512). As shown in Figure S2, with a training data set of just five stacks, that required just 20 min to annotate, we obtained satisfactory results very similar to a network trained with the large training data set. We also applied the trained models to the other data sets and obtained slightly worse results than with the fully trained network with just some minor artifacts.

This demonstrates that DP shows good performance already with small data sets of 5-10 stacks that can be created in less than one hour. We acknowledge the user-advantage of sparse labeled data and immediate feedback as shown in Ilastik, but implementing this capacity is beyond the scope of the present project.

[Is there potential to combine deep projection with Ilastik, by using user-assisted Ilastik segmentation to quicker generate a suitable training set, that then robustly predicts the rest of the data, without added manual training required?]

As shown above, creating a training data set with Ilastik leads to masks with unwanted artifacts, no sharp edges, and sprinkled textures. Thus, Ilastik appears not well suited for training data generation.
We propose another procedure to accelerate training data set generation. To extend the model to new imaging conditions, e.g., a mutant, we suggest predicting a set of new training images with an existing model and then manually correct the predicted masks using Napari or Fiji. For the future, we consider implementing a Napari plugin for DeepProjection that allows user-friendly training and prediction within the Napari UI.

Minor:

1. *[The detailed discussion about the network can be moved to an appendix (e.g. loss function, learning rate, number of layers, …) since one would imagine it is more expert terrain than the typical readership of Development. Similarly, the mathematical discussion of "tissue unrolling" can be moved to the SI.]*

   We have moved the sections mentioned to the methods section and added an algorithm description to the main text that is aimed at non-experts.

2. *[The authors should show the scaling of the model performance (preferably with a couple of representative "projections") with the size of the training set*

   We have shown that DeepProjection already performs well with a training data set of just 5 stacks for a single fly line and microscope. By using extensive training data augmentation (adding noise, adjusting contrast and brightness), DP can correctly process data from different conditions, e.g., due to photobleaching.

3. *The authors note that the training set comprises data recorded by different scopes. Can they check how robustly transferable their model is? E.g., omit data from scope X from the training set, and then show how the model performs on images recorded by scope X, or even try the model on a slightly different fluorescent marker (e.g. different fluorescent markers labeling the same protein). Successfully demonstrating transferability would be a major plus.*

   We agree that transferability is a very important feature to lower the time investment needed to use our algorithm and to attract more users. As suggested, we tested how well a model performs on a data set with a different fluorescent marker and a different scope (see above, supplementary information S2).

4. *[For the end user, it would be useful if the authors supplied more detail about the procedures used to train and test their neural network. What size of the training and testing sets used, and should be targeted? Was a separate testing set, not used for model selection, used to generate the scores and images shown in the figures?]*

   We have updated the Jupyter notebook in the GitHub repository with detailed instructions on how to (1) select and annotate training data, (2) train a custom model, and (3) process data (https://github.com/danihae/DeepProjection/blob/main/Quickstart_training_and_prediction.ipynb). We feel that it is better to refer the user to GitHub rather than adding detailed instructions to the manuscript, since the package will potentially be changed and further improved after publication.

   For the validation of model performance and the figures, we used a disjunct set of images not included in the training data set (link see GitHub), as mentioned in the manuscript.

Reviewer 2 (Léo Guignard) (reviewer comments in italics):

Major comments:
[The authors could, and maybe should, better show the superiority of their method by comparing the result of the extraction of biological information from their projection against other types of projection. For example by doing cell segmentation or by quantifying cell flow.]

We entirely agree that it is useful to add some downstream cell segmentation to the manuscript. In fact, the motivation for developing DeepProjection was to improve subsequent cell shape segmentation. The method developed for the segmentation will be the subject of a separate publication, but we have now added some exemplary results to this manuscript. We segmented the lateral epidermis tissue of Drosophila embryos during dorsal closure. Lateral epidermis cell segmentation was challenging because cadherin- labeled cell boundaries were faint, and cells were strongly elongated. We compared (1) maximum intensity projection and (2) DeepProjection (Fig. 1), subsequently using a custom U-Net for cell boundary detection followed by watershed segmentation (details in a future publication). The segmentation results are significantly better following the application of DeepProjection than following MIP (almost no segmentation errors vs. vastly wrong segmentation).

[The authors do not comment on the continuity of their reconstructed projection over time. Though, when looking at the provided movie, it seems like there is some flickering indicating the fact that similar images might have significantly different projection (especially around the closing end of the tissue). This problem could prevent a good tracking of the cells. This could be avoided by, for example, averaging over time, the z values.]

We are very grateful for this suggestion and have now implemented the option to improve the time-consistency of masks over multiple frames by a mean filter with subsequent thresholding. We demonstrate that a moving-window averaging mask with a kernel size of 7 frames improves time-consistency and eliminates flickering at the canthi of the amnioserosa tissue (see supplementary movie M1). It is worth noting, though, that, apart from minor flickering artifacts, DeepProjection

is, even without time-filtering, well-suited to robustly process time-lapse recordings.

Minor comments:
[The authors show images where the membranes are labelled, would their method work with nuclei data? If yes, it would strongly increase the impact of their work. Though, it would definitely be understandable if it does not transfer easily or at all.]

We believe that DeepProjection will also work with fluorescently labeled cell nuclei and other labeling targets. However, creating a suitable data set with nuclear staining of a curved tissue, ideally by establishing a double-fluorescence (eCad, nucleus) *Drosophila* line, is unfortunately not possible in the timeframe of this revision.

[L20: The authors mention that they create undistorted 2D projections from the original 3D shape. While the contextual meaning of undistorted is explained later on in the article as the fact that the created surface is continuous, without context it can be understood as the fact that the algorithm raises an undistorted 2D projection of an ellipsoidal shape.]

The term “undistorted” in the abstract refers to the fluorescent gray values that are conserved (due to binary masking) and not to geometrical distortion. This is a major plus for quantitative imaging, e.g., quantification of protein signal. For example, DeepProjection allows quantitative tracking of the fluorescent signal of exclusively the apical surface of epithelial tissue in 3D volumetric data over long time periods. Later in the article, “distorted” refers to cell shapes. To avoid ambiguities, we have rephrased the respective section of the abstract.

[L42: The authors state that it is “necessary” to create 2D projections to follow tissue sheet morphogenesis. While it is understandable that it can be of great help to do so, it might not be necessary. Or if indeed the authors think it is, they should justify their claim.]

Excellent point. We fully agree that 2D projections are not absolutely necessary and that there are 3D methods. However, since 3D methods are computationally more complex and accordingly far less robust than 2D methods, it is **useful** to perform 2D projections prior to analyzing cell sheet morphogenesis. We have adjusted the wording in the manuscript accordingly.

[L79 and other places: the authors mention “crisp boundaries”. It is not completely clear what they mean by that. A definition of “crisp” in the context of this study would be welcome.]

By “crisp” we mean sharp, distinctive structures, as opposed to “blurred” or “opaque”. We have clarified this definition in the manuscript.

[L212: The authors briefly describe the correction of local distortion but do not fully disclose the method. Because it is a method paper, showing the method could be of great interest to the community.]

This was an omission on our side. We noticed that the unfolding algorithm was not yet included in the published Python code. We have now added the code into the Python package and added an instructional guide to the Jupyter notebook (https://github.com/danihae/DeepProjection/blob/main/Quickstart_training_and_prediction.ipynb).

[Figure 3D.: It is unclear why the authors show MIP vs others rather than GT vs others, could they justify their choice?]

Excellent point, we have changed the figure in the manuscript accordingly. We initially wanted to show how CSBD non-linearly distorts intensity values, and therefore plotting against MIP show this effect evidently. However, plotting against the GT is more informative, showing both the excellent performance of DP as well as the non-linear distortion of CSBD.

Second decision letter

MS ID#: DEVELOP/2022/200621

MS TITLE: DeepProjection: Specific and robust projection of curved 2D tissue sheets from 3D microscopy using deep learning

AUTHORS: Daniel Haertter, Xiaolei Wang, Stephanie M Fogerson, Nitya Ramkumar, Janice M Crawford, Kenneth Poss, Stefano Di Talia, Daniel Kiehart, and Christoph F Schmidt
ARTICLE TYPE: Techniques and Resources Article

I apologize for the long delay before coming back due to poor availabilities through the summer. I am happy to tell you that your manuscript has been accepted for publication in Development, pending our standard ethics checks.

Reviewer 1

Advance summary and potential significance to field

Thanks for the careful response. DP looks like a useful tool for the community. The comparison to alternative software is clear, and underscores the potential for use of this approach. In terms of significance, with quantitative biology on the rise, and few tools available for extracting image information in a user friendly way the one presented here is likely to have a good impact on the field.

Comments for the author

I am largely happy, but have one last question, or rather a clarification. This however does not require sending back to me.
Specifically, I am returning to your response about dynamic surfaces. If I understand the workings of DP correctly, there is an initial, tissue specific training phase. I also understand that you trained two distinct classifiers on the two samples shown. In some cases, tissues can change their shape radically, such that they hardly resemble the initial configuration. For example, think about a tube like the early heart coiling and writhing. If you had a membrane marker on the surface of this object, do you think it requires training multiple district classifiers to be trained, or can it be done in with a single training phase? Happy to talk about this offline.

Reviewer 2

Advance summary and potential significance to field

As mentioned in the previous review, the proposed work enables projection of 3D surfaces in a 2D plan using a deep learning model. Cell detection and tracking and cell dynamics analysis can be particularly complicated in 3D. Therefore, being able to faithfully project a tissue that spans a 3D surface onto a 2D plan would be useful for the community and potentially enable major advances.

Comments for the author

The authors have answered all the concerns that were raised. I do not think it needs more revisions.
Léo Guignard