Zebrafish Skeleton Measurements using Image Analysis and Machine Learning Methods

Stern O.[1], **Marée R.**[1,2], **Aceto J.**[3], **Jeanray N.**[3], **Muller M.**[3], **Wehenkel L.**[1] **and Geurts P.**[1]
{OLIVIER.STERN, P.GEURTS}@ULG.AC.BE
[1]GIGA-Systems Biology and Chemical Biology, Dept. EE & CS, University of Liège; [2]GIGA Bioinformatics Core Facility, University of Liège; [3]GIGA-Development, Stem cells and regenerative medicine, Molecular Biology and Genetic Engineering, University of Liège

Keywords: machine learning, image segmentation, pixel annotation, zebrafish morphometry

Abstract

The zebrafish is a model organism for biological studies on development and gene function. Our work aims at automating the detection of the cartilage skeleton and measuring several distances and angles to quantify its development following different experimental conditions.

1. Introduction

In this work, we address two subproblems: 1) quantifying the surface of the cartilage skeleton and 2) detecting several points of interest in zebrafish images. These two problems are not trivial from a machine learning point of view. These are essentially structured output problems for which there exist no straightforward solution. In addition, labels have to be indicated manually and thus training sets are rather small. Finally, these annotation problems are not easy even for human.

2. Image Generation

As we focus on the evolution of the cartilage skeleton, zebrafish larvae were grown fom 4 to 10 days post-fertilization, sacrificed by tricaine treatment and fixed in 4% para-formaldehyde. Alcian blue staining of the cartilage extracellular matrix was performed according to (Kimmel et al., 1998) and photographs were obtained using a Olympus dissection microscope by observing single larvae under identical conditions. Ventral and lateral views were recorded, but in this ongoing work, we will only use 15 ventral views.

3. Segmentation of Cartilage Skeleton

Our goal is to isolate the cartilage skeleton to quantify its surface. This approach uses supervised learning methods to partition the elements of the embryo within a new, unseen image by predicting the class of every pixel. Our starting point is a technique developed in (Dumont et al., 2009) which extracts a random sample of subwindows in a set of images with pixel-wise labelling (i.e. every pixel is labelled with one class among a finite set of predefined classes) and exploits extremely randomized trees to build a classifier.

Training Stage. We assume a partial annotation by an expert of the image pixels into three classes: Eye, Skeleton and Others. For each image of the learning sample, we construct subwindows of size w x h centered on the labelled pixels. Subwindows are then tagged by the class of the central pixel (output) and described by the color (i.e. the values in HSV and RGB color spaces), the edges (i.e. the gradient of the Sobel operator) and the texture (i.e. the histogram of local binary pattern (Ojala et al., 1996)) of the pixels in the subwindows (inputs). From this learning set, we construct a classification model using extremely randomized trees.

Prediction Stage. A subwindow centered on each pixel of every test image is extracted and its class is predicted using tree ensembles.

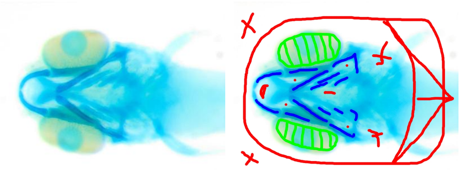
Parameters. The approach depends on several parameters that are related to the machine learning method (number of trees, number of random test splits) or to the subwindow descriptors (type of attributes, size of the subwindow). We only report our best results below.

Results. We manually and partially labelled 3 images out of the 15 (see Figures 1(a) and 1(b) for one ex-

---

Appearing in *Proceedings of the 20th Machine Learning conference of Belgium and The Netherlands.*

ample), from which the classification model was learnt. The 12 remaining images were then automatically annotated. Figures 1(c) and 1(d) show a test image and its annotation. The assessment of our model is difficult in the absence of a complete ground truth annotation but results seem visually very satisfying.



(a) Training Image (b) Annotation

(c) Test Image (d) Segmentation

Figure 1. Segmentation of Cartilage Skeleton

4. Detection of Points of Interest

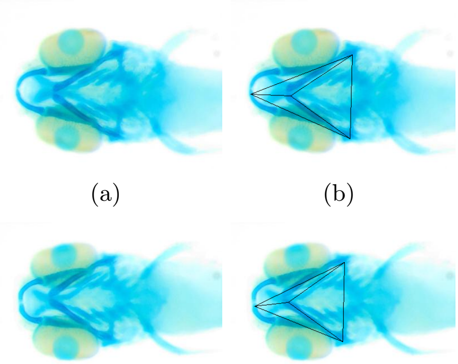
The objective of our second contribution is to automatically detect particular points of interest in the image, corresponding to particular position in the cartilage. By using several of such detectors, we can then automatically measure various lengths and angles defined by the corresponding points. As for pixel classification, we describe our pixel with a subwindow centered on the points of interest and we exploit extremely randomized trees to build a classifier.

Learning Stage. For a given point of interest, we indicate its coordinates in all training images. The learning set is then composed, on one hand, of subwindows centered on pixels in a radius r around the pixel of interest (positive class) and, on the other hand, of subwindows taken randomly in the rest of the image (negative class). Subwindows features are the same as in Section 3. Each particular point of interest thus requires the construction of a separate binary classification model using extremely randomized trees.

Prediction Stage. We predict the class of every subwindow centered on the pixels of the test images using tree ensembles. To determine the final coordinates of the points, we rank the prediction on their predicted probability and take either the mean, the median or a weighted mean of the coordinates of the best predicted pixels (i.e. above a given theshold).

Parameters. In addition to the parameters of the learning algorithm and subwindow descriptors, this method introduces several new parameters such as the radius around the points and the threshold on predicted probabilities. We tested several values of these parameters and only report the best results below.

Results. We apply this idea to detect four different points of interest in zebrafish images. These four points were manually identified in all images and then a leave-one-out was performed to assess the four classifiers. Figure 2 shows the prediction of these four points in two test images. Once these points have been detected several measures of lengths and angles can be done (represented by straight lines in Figures 2(b) and 2(d)).



(a) (b)

(c) (d)

Figure 2. Detection of points of interest

5. Future Work

We will further investigate ways to improve point detection, using multi-class classifiers and/or regression trees, and evaluate the approach on larger image sets of larvae presenting various defects in cartilage formation to obtain automatic determination of the different morphometric measures.

References

Dumont, M., Marée, R., Wehenkel, L., & Geurts, P. (2009). Fast multi-class image annotation with random subwindows and multiple output randomized trees. *Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP)*.

Kimmel, C. B., Miller, C. T., Kruze, G., Ullmann, B., BreMiller, R. A., Larison, K. D., & Snyder, H. C. (1998). The shaping of pharyngeal cartilages during early development of the zebrafish. *Dev Biol*, *203*, 245–263.

Ojala, T., Pietikainen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, *29*, 51–59.