CASCADED RANDOM FOREST CLASSIFIERS FOR SOMITE LOCALIZATION AND IDENTIFICATION IN ZEBRAFISH

Richmond D.L.[1], Kainmüller D.[1], Rother C.[2], Myers E.W.[1]

(1) Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, DE. (2) Technical University Dresden, Dresden, DE

Abstract

During embryonic development, vertebrates undergo a sequential patterning process that leads to their segmented body plan. Understanding this process is an important challenge for developmental biology, and will improve our understanding of diseases such as congenital scoliosis. This poster describes automated tools for analyzing timelapse movies of segmentation in zebrafish. We explore the use of cascaded Random Forest classifiers, following a scheme similar to auto-context.

Introduction

The regular segmented pattern of the vertebral column arises during early development, as individual segments (called somites) "bud" off rhythmically with a defined spatial and temporal period from growing mesoderm tissue (see Figure 1). Underlying this periodicity is an oscillating genetic regulatory network, called the segmentation clock, which sets the species specific period of somitogenesis (30min in fish, 2 hr in mice, ~8 hr in humans).

Zebrafish is a powerful model organism for studying somitogenesis by a combination of genetic perturbation and live imaging. E.g., the first mutant with an altered period, and consequently an altered anatomy, was recently discovered using this approach [1].

The aim of this work is to develop a tool for automatically tracking the process of somitogenesis in zebrafish, for use in large-scale genetic screens that aim to understand the complex genetic network underlying the segmentation clock.

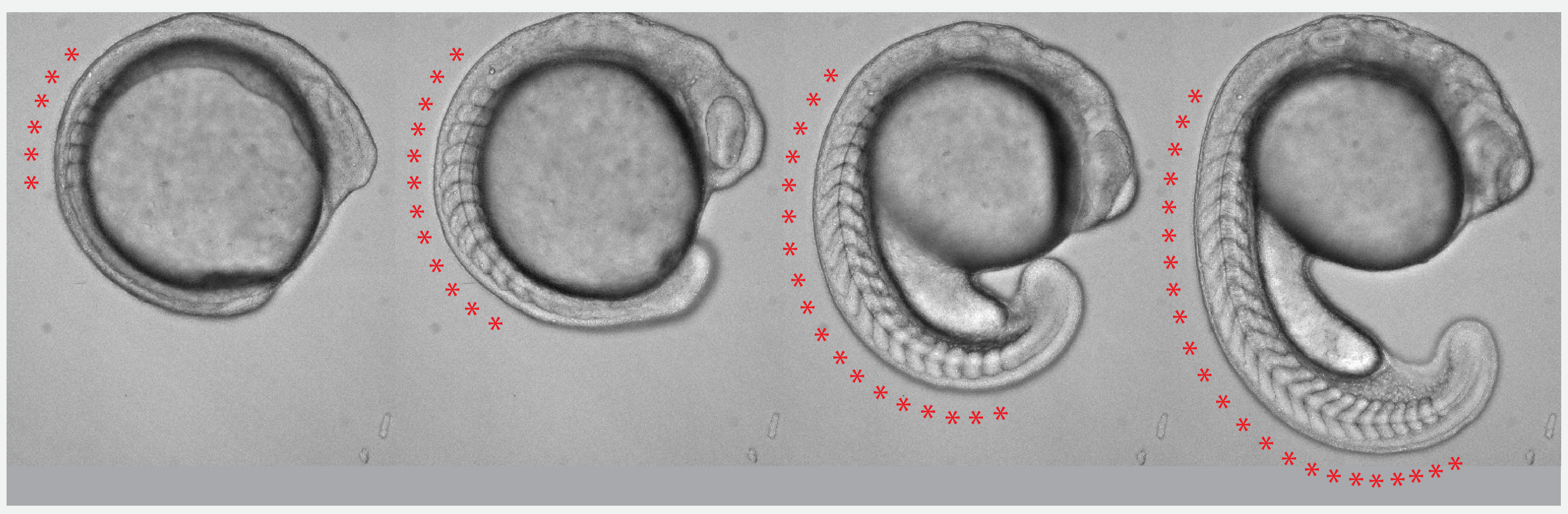


Figure 1. Zebrafish somitogenesis. Each asterisk (*) denotes a single somite.

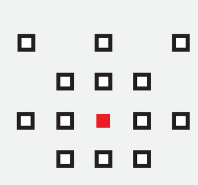
Methodology

A similar problem of spine detection arises in the medical imaging community, and we are inspired in particular by the approach of Glocker, *et al.* [2]. However, here we explore the use of Cascaded Random Forest (RF) Classifiers, similar to the approach of Tu and Bai [3], which we train according to the following scheme:

Train RF#1 on TS#1 → Classify TS#2 with RF#1 → Train RF#2 on TS#2 (incl. output of RF#1)* → Classify TS#3 with RF#1 → Classify TS#3 with RF#2 (incl. output of RF#1)* → Train RF#3 on TS#3 (incl. output of RF#2)* → ...

TS#1: Training Set #1
RF#1: Random Forest Classifier #1

* After each classification step, the probability maps are sampled over a lattice centered on each pixel (see image at right), in order to generate new contextual features to train the next Random Forest.



We acknolwedge D. Soroldoni and A.C. Oates for supplying timelapse images of the developing zebrafish, and for helpful discussions.

Results

In the first stage of the cascade, the RF sees the local response of a filter bank in addition to the XY position at every pixel. The classification result is reasonably accurate, but extremely noisy. In subsequent stages of the cascade, the RFs receive richer contextual features calculated from the posterior probability maps of previous RFs, sampled over a local neighborhood. As a result, the classification improves dramatically, and the relative positions and shapes of the somites starts to resemble the ground truth labeling.

Figure 2. A dense annotation of each somite is made from the grayscale image of a 21 somite stage zebrafish. The classification result at each of the first three levels of a cascade is shown at right. The asterisk (*) marks somite #7, which is explored further in Figure 3.

The improvements in the labeling map can be clearly seen in the posterior probability map of a single segment, with increasing levels in the cascade (Figure 3).

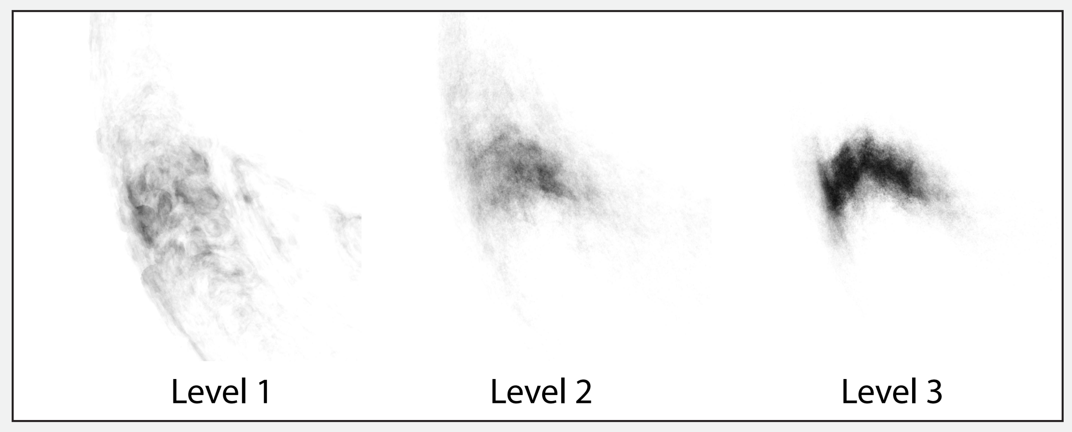


Figure 3. Posterior probability map for somite #7 (from Figure 2).

Discussion

Despite the promising preliminary results, we also observe cases where the detected somites actually extend well beyond the actual region of interest (Figure 4, curly brackets). Interestingly, the order of somites is maintained, as is the position of the posterior-most somite, which has quite clear contrast in the grayscale image. Consequently, one of the somites is dramatically elongated (Figure 4, arrow).

We speculate that this error is largely due to the short lengthscale over which contextual features are sampled. Extending this reach to the full image, e.g. using the technique of [4], should dramatically improve the robustness of this approach.

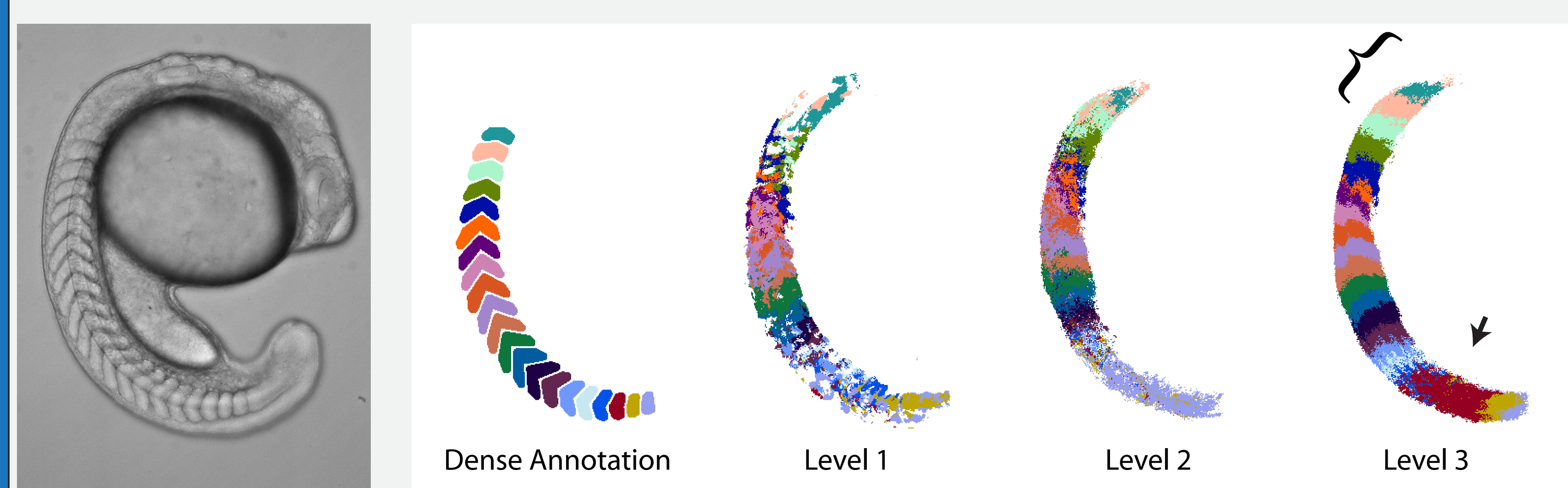


Figure 4. Dense annotation and output of Random Forest classification for different levels in a cascade, similar to Figure 2. Curly brackets highlight a region where the somites have extended past the ground truth labeling. An arrow points to a somite that is correspondingly elongated.

References

1. Schröter C and Oates AC. Current Biology. (2010). 20: 14, p. 1254-1258.
2. Glocker B, et al. MICCAI. (2013). LNCS 8150, Part II, p. 262-270.
3. Tu Z and Bai X. (2010). IEEE T Pattern Anal. 32: 10, p. 1744-1757.
4. Criminisi A, Shotton J and Bucciarelli S. (2009). MICCAI-PMMIA.