

Phenotype Classification of Zebrafish Embryos by Supervised Learning

Nathalie Jeanray¹, Raphaël Marée^{2,3}, Benoist Pruvot¹, Olivier Stern², Pierre Geurts²,
Louis Wehenkel², Marc Muller¹

¹GIGA-Development, Stem cells and regenerative medicine, Molecular Biology and Genetic Engineering,
University of Liège, Belgium

²GIGA-Systems Biology and Chemical Biology, Dept. EE & CS, University of Liège, Belgium

³GIGA Bioinformatics Core Facility, University of Liège, Belgium

Keywords: Phenotype classification, supervised learning, zebrafish

Abstract

Zebrafish, or “*Danio rerio*”, is increasingly used as a vertebrate model organism in the field of developmental biology, gene function and also in toxicology and pharmacology. Indeed, due to several advantages such as fast growth, *ex vivo* development, larvae transparency, low cost and permeability to small molecules, zebrafish appears as a powerful model to assess toxic activities on vertebrates. In particular, it is used to assess biological properties of chemical substances on embryo development. In such studies, phenotypical effects are generally evaluated manually, one by one, through microscopic observation by an expert to detect potential developmental or morphological modifications (e.g. edemas, curved tails, growth delay, etc.), and documented by several typical photographs. This method is tedious, time-consuming and prone to appreciation subjectivity. The large number of substances to be tested and the need for accuracy and reproducibility of the results call for methods allowing automation of both data acquisition and classification of the images.

This presentation will focus on the automatic phenotype recognition pipeline. It includes standard image processing methods to first standardize images, followed by recent variants of a supervised image classification method [1]. This method is based on dense random subwindows extraction in images, their description by normalized raw pixel values and the use of ensembles of extremely randomized trees [2] either to classify these subwindows hence images, or to build visual features subsequently classified by linear SVM.

In this work, we evaluated systematically the approach on batches of microscopy images acquired during independent experiments across several days. Systematic evaluation of parameter values and classification schemes allow us to obtain roughly 90% recognition rates for two different defects, edema and curved tail, compared to ground-truth provided by experts. In the future, we will focus on rendering the acquisition procedure fully automatic, and the approach will be evaluated on larger image sets including other phenotypes.

References

- [1] Marée, R., Geurts, P., Wehenkel, L. “*Random subwindows and extremely randomized trees for images classification in cell biology.*”, in BMC Cell Biology, 2007, vol. 8 (Suppl 1):S2
- [2] Geurts P, Ernst D, Wehenkel L. “*Extremely Randomized Trees.*” in Machine Learning, 2006, vol. 63, pp.3-42