Ilastik Profiling Report

09-10-2915
Jaime I. Cervantes
cervantesj@janelia.hhmi.org

Cross-validation tests with varying numbers of trees and different features were conducted to optimize the running time of Ilastik. Five different data sets were used to run these tests, with each case containing 2 or 3 videos. Fifty frames were labeled for each video and used as ground-truth data for cross-validation. In figure 1 we can observe the classification results with different number of trees for the case of Virilis A larvae:

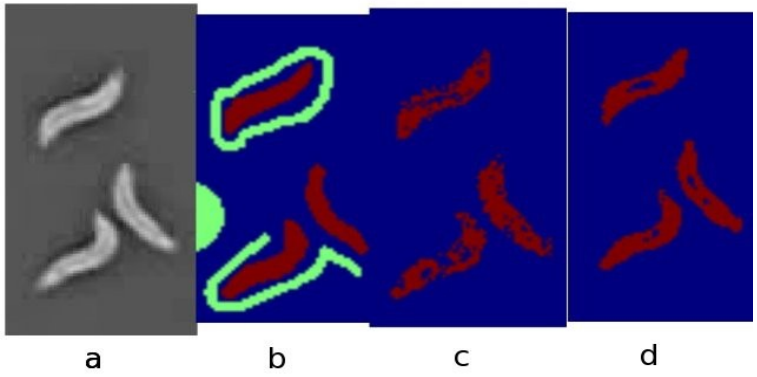


Figure 1: a) Frame screen-shot b) Ground-truth labels c) Prediction with 1 tree d) Prediction with 50 trees

In the following sections each feature is represented by an abbreviation, followed by sigma (eg. *GS(10.0)* means Gaussian smoothing with sigma 10.0):

Gaussian Smoothing	GS
Laplacian of Gaussian	LG
Gaussian Gradient Magnitude	GGM
Difference of Gaussians	DG
Structure Tensor Eigenvalues	STE
Hessian of Gaussian Eigenvalues	HGE

Feature Selection Test

The objective of this test is to select the optimal features that will result in the fastest running-time while maintaining an acceptable cross-validation error. For each case, features were removed gradually according to their importance.

Some notes on the results:

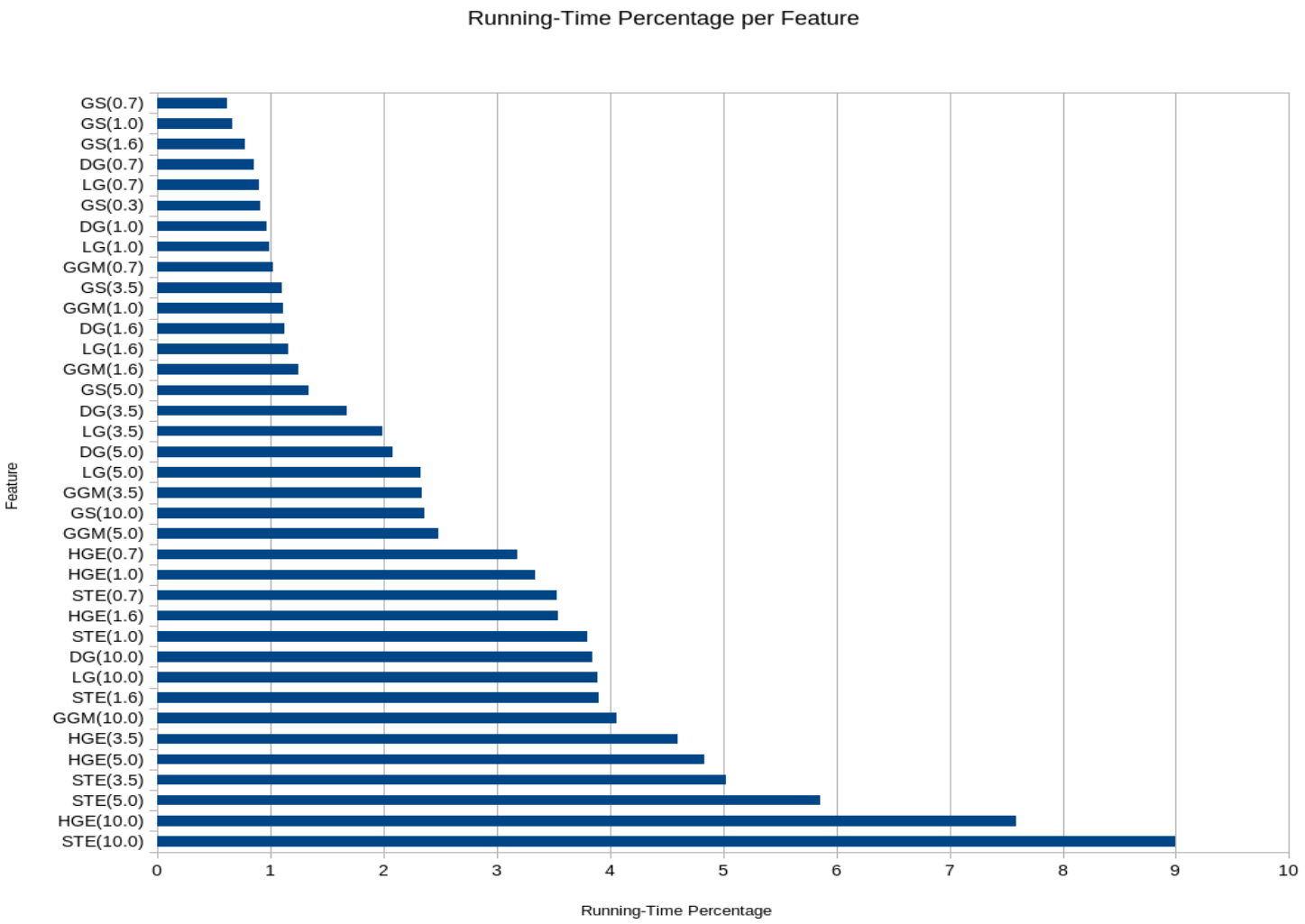
- The RF importance table is re-calculated every time a feature is removed.
- The total time corresponds to the time for feature computation plus prediction, and doesn't include training time.
- The time improvement is a comparison against the running time of the test case with all the features activated.

Feature Selection Test

Test Case	Resolution (px)	CV Error (%)	Features Removed	Features Remaining	Training Time (secs)	Feature Time (secs)	Prediction Time (secs)	Total Time (secs)	Time / Frame (secs)	Time Improvement (X Times)	Num of Feature Pxs (10%)	Foreground / Background Pixel Ratio
Alice's Courtship Bowl	1024x1024	0.124	GGM(5.0), GGM(3.5), DG(10.0) GGM(10.0), DG(0.7), HGE(0.7) STE(5.0), LG(0.7), STE(0.7) GS(3.5), DG(1.6), LG(10.0) DG(1.0), GGM(0.7), STE(10.0) GGM(1.0), STE(3.5), GS(1.0) LG(1.6), HGE(1.6), GS(5.0) HGE(1.0), HGE(10.0), LG(1.0) GS(10.0), GS(1.6), LG(5.0) GS(0.7), GGM(1.6), DG(3.5) STE(1.6), GS(0.3)	LG(3.5), HGE(3.5), DG(5.0) STE(1.0), HGE(5.0)	19.11	58.49	28.90	106.50	2.15	3.25	546459	113
Alice's Fly Bowl	1024x1024	0.075	GGM(10.0), LG(0.7) GGM(3.5), LG(10.0), GGM(5.0) STE(10.0), GS(1.6), DG(0.7) DG(10.0), GS(0.7), GGM(1.6) HGE(0.7), GS(3.5), GS(1.0) LG(1.0), STE(1.0), STE(0.7) GS(10.0), STE(5.0), GGM(0.7) DG(1.0), LG(1.6), GS(5.0) GGM(1.0), HGE(10.0), STE(1.6) DG(1.6), LG(5.0), GS(0.3) HGE(1.6), HGE(1.0), HGE(5.0) DG(5.0)	STE(3.5), HGE(3.5), LG(3.5) DG(3.5)	16.44	31.59	29.36	77.39	1.18	4.5	704711	141
Alice's Fly Bubble	1024x1024	0.164	GGM(5.0), LG(0.7), DG(0.7) GGM(1.0), GGM(3.5), DG(1.0) HGE(0.7), LG(1.0), HGE(1.6) DG(1.6), HGE(1.0), GGM(1.6) LG(1.6), GGM(10.0), GS(3.5) LG(3.5), GGM(0.7), STE(0.7) STE(1.0), DG(5.0), STE(1.6) STE(3.5), GS(5.0), STE(5.0) STE(10.0), HGE(3.5), DG(3.5) LG(5.0), HGE(5.0), GS(1.0) GS(0.3)	GS(0.7), GS(1.6), GS(10.0) LG(10.0), DG(10.0), HGE(10.0)	14.50	72.90	23.11	110.52	2.25	3.2	753950	44
Rivera's Larvae	480x640	0.435	DG(0.7), DG(10.0) LG(0.7), GS(0.3), GS(1.0) DG(5.0), LG(10.0), GGM(1.6) LG(5.0), GGM(3.5), GS(0.7) GS(1.6), GS(5.0), GS(3.5) GS(10.0), GGM(5.0), HGE(0.7) LG(3.5), DG(3.5), GGM(1.0) DG(1.0), LG(1.6), LG(1.0) DG(1.6), STE(3.5), STE(1.0) HGE(1.0), STE(5.0), HGE(1.6) GGM(0.7), STE(1.6)	STE(0.7), HGE(10.0), HGE(5.0) HGE(3.5), GGM(10.0), STE(10.0)	30.81	23.64	9.42	63.87	0.96	3.7	724720	149
Zlatic's Larvae	2816x2816	0.27	DG(0.7), LG(0.7), HGE(0.7) DG(1.0), LG(1.0), GGM(1.0) GGM(0.7), HGE(1.0), DG(1.6) LG(1.6), GGM(1.6), GGM(3.5) HGE(1.6), STE(0.7), STE(3.5) STE(1.6), GGM(5.0), GGM(10.0) DG(3.5), STE(5.0), STE(1.0) GS(0.3), DG(5.0), HGE(10.0) GS(10.0), STE(10.0)	LG(3.5), GS(0.7), GS(5.0) GS(3.5), HGE(3.5), LG(10.0) DG(10.0), LG(5.0), GS(1.0) HGE(5.0), GS(1.6)	299.32				14.50	2.8	3336278	97
Roian's Mice	840x840	0.21	DG(0.7), LG(0.7) HGE(0.7), STE(0.7), DG(1.6) DG(1.0), GGM(0.7), LG(1.0) GGM(1.6), GGM(1.0), HGE(1.0) LG(1.6), HGE(1.6), STE(1.0) STE(1.6), DG(3.5), LG(3.5) GGM(3.5), DG(5.0), HGE(3.5) HGE(5.0), LG(5.0), STE(3.5) GS(0.3), GGM(5.0), STE(5.0) GS(0.7), DG(10.0), STE(10.0) LG(10.0), GS(10.0)	HGE(10.0), GGM(10.0), GS(5.0) GS(1.0), GS(3.5)	56.37	57.41	22.63	136.42	1.19	3.9	1224050	32

Comparison of Running-Times for Each Feature

The following plot compares the running-times for each feature.



Random Forest Variable Importance and Highly Correlated Features

For the previous tests, the variable importance table was recalculated for every feature removed. One of the problems that can be observed from the results, is the selection of features that are highly correlated. There are efforts to solve this problem, and there has been some some previous work done on feature selection by Fabian Isensee. The following figure summarizes their results [Isensee et al.]:

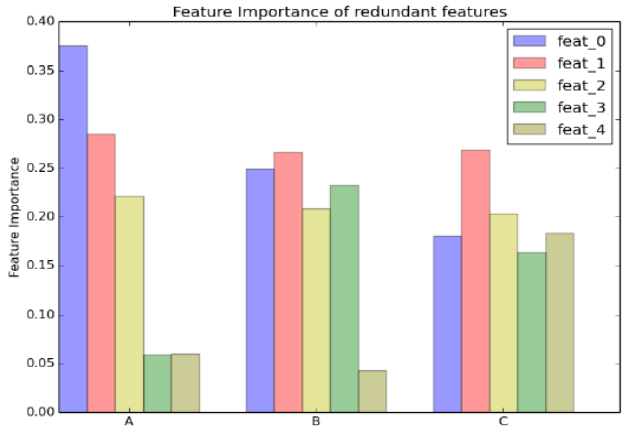


Figure 2.5: **Variable Importance of Redundant Features**. A shows the feature importances (Gini importance) for an artificial dataset. Feature 0 is by far the most important, followed by 1 and 2. Features 3 and 4 are noise features. In B feature 3 has been replaced by an exact copy of feature 0. They now share the importance previously assigned to feature 0. In C, feature 3 has been replaced by a noisy copy of feature 0. The importances behave accordingly. This toy example demonstrates that feature selection based on feature importance is problematic in the presence of redundant features.

Increasing Number of Trees Test

The objective of this test is to select the number of trees that will result in the the fastest running-time while maintaining an acceptable cross-validation error.

Some notes on these tests:

- Every test ran with all the features activated.
- The time improvement is a comparison against the running time of the test case with 100 trees.
- The total time corresponds to the running time for feature computation and prediction, and doesn't include training time.

Increasing Number of Trees Test

Test Case	*Resolution (px)*	*CV Error (%)*	*Trees Num.*	*Training Time (secs)*	*Feature Time (secs)*	*Prediction Time (secs)*	*Total Time (secs)*	*Time / Frame (secs)*	*Feature Time Gain*	*Prediction Time Gain*	*Total Time Gain*	*Foreground Pixels Num*	*Background Pixels Num*
Alice's Courtship Bowl	1024x1024	0.0854	20	67.03	116.4	46.26	162.66	3.25	1.02	1.60	1.19	4565	541894
Alice's Fly Bowl	1024x1024	0.049	17	74.32	118.04	31.63	149.67	2.99	1.03	2.18	1.27	5144	699567
Alice's Fly Bubble	1024x1024	0.14	11	76.46	118.06	31.35	149.41	2.99	1.01	2.52	1.32	16188	737762
Rivera's Larvae	480x640	0.337	20	97.45	70.94	27.72	98.66	1.97	0.99	1.52	1.14	4827	719893
Roian's Mice	840x840	0.28	15	161.58	78.93	24.94	103.87	2.08	1.01	2.27	1.31	37216	1186834