

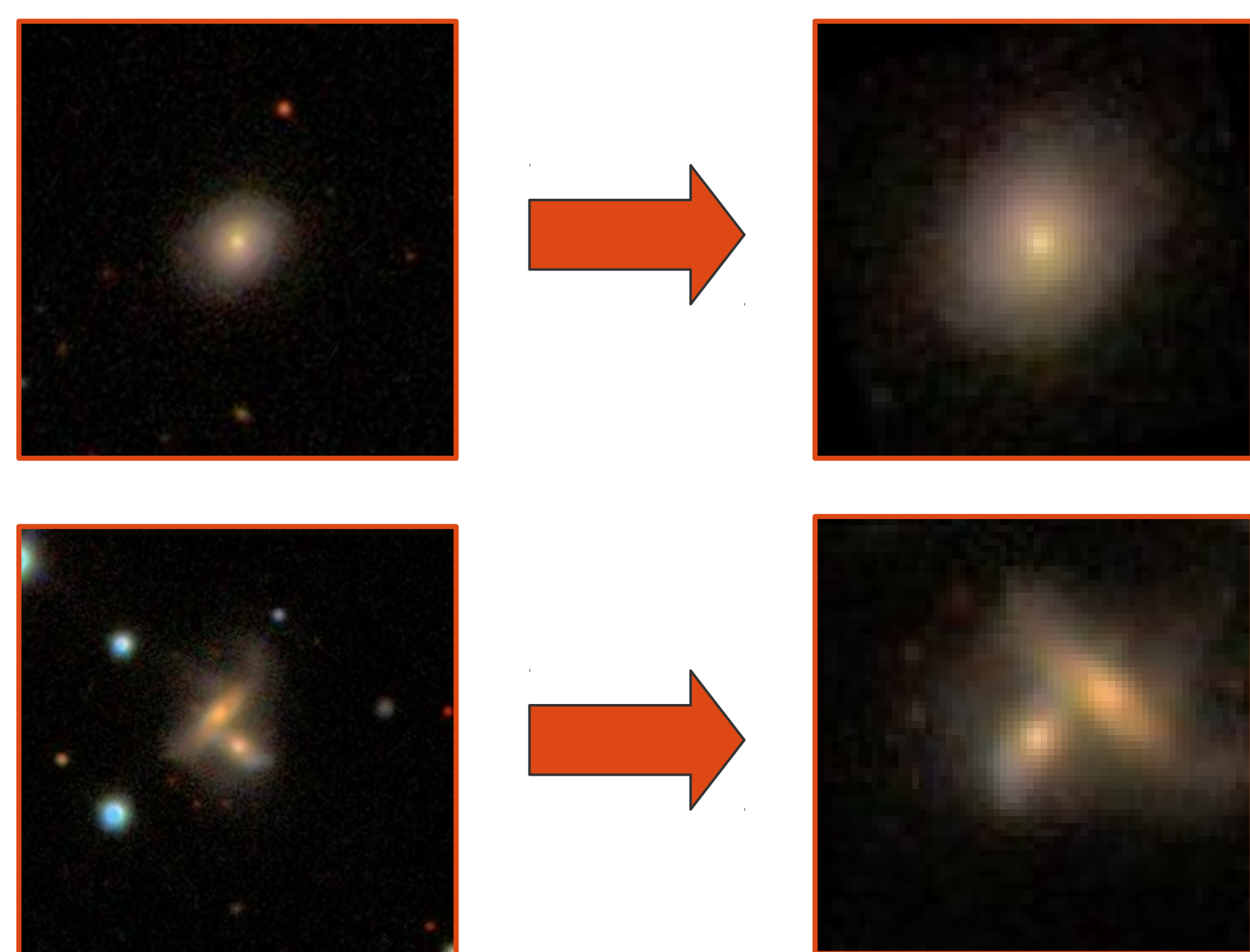
Overview

- Predict proportions of user classifications of galaxy morphologies
- Multiple regression problem
 - RMSE evaluation
- Leaderboard error: 0.12859
- Rank: 145/329

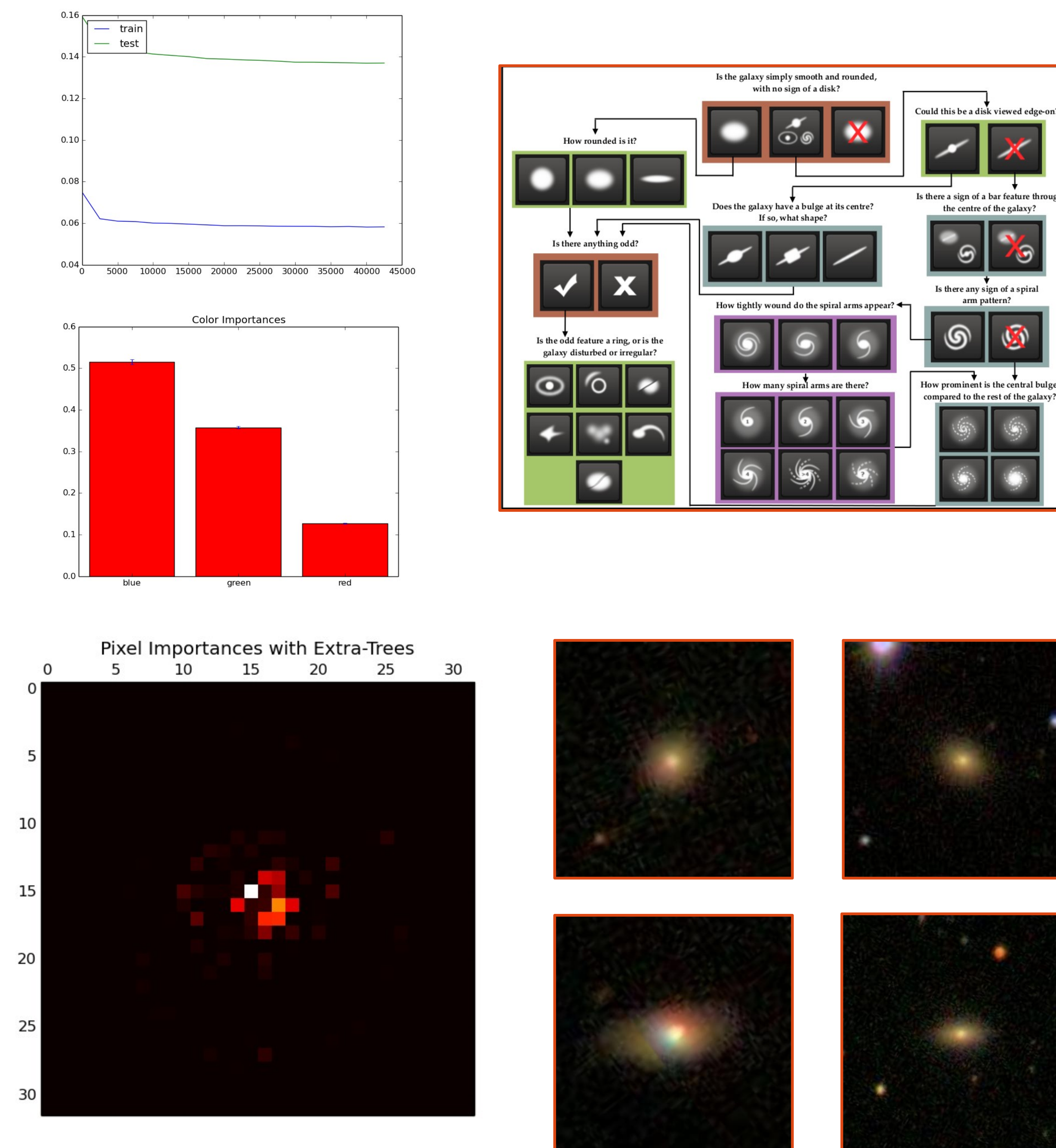
Preprocessing

All images underwent:

- (1) Cropping & downsampling
- (2) Data augmentation^[1]
 - random affine transform
- (3) Scaling & whitening



Results



What didn't work: (my) feature engineering

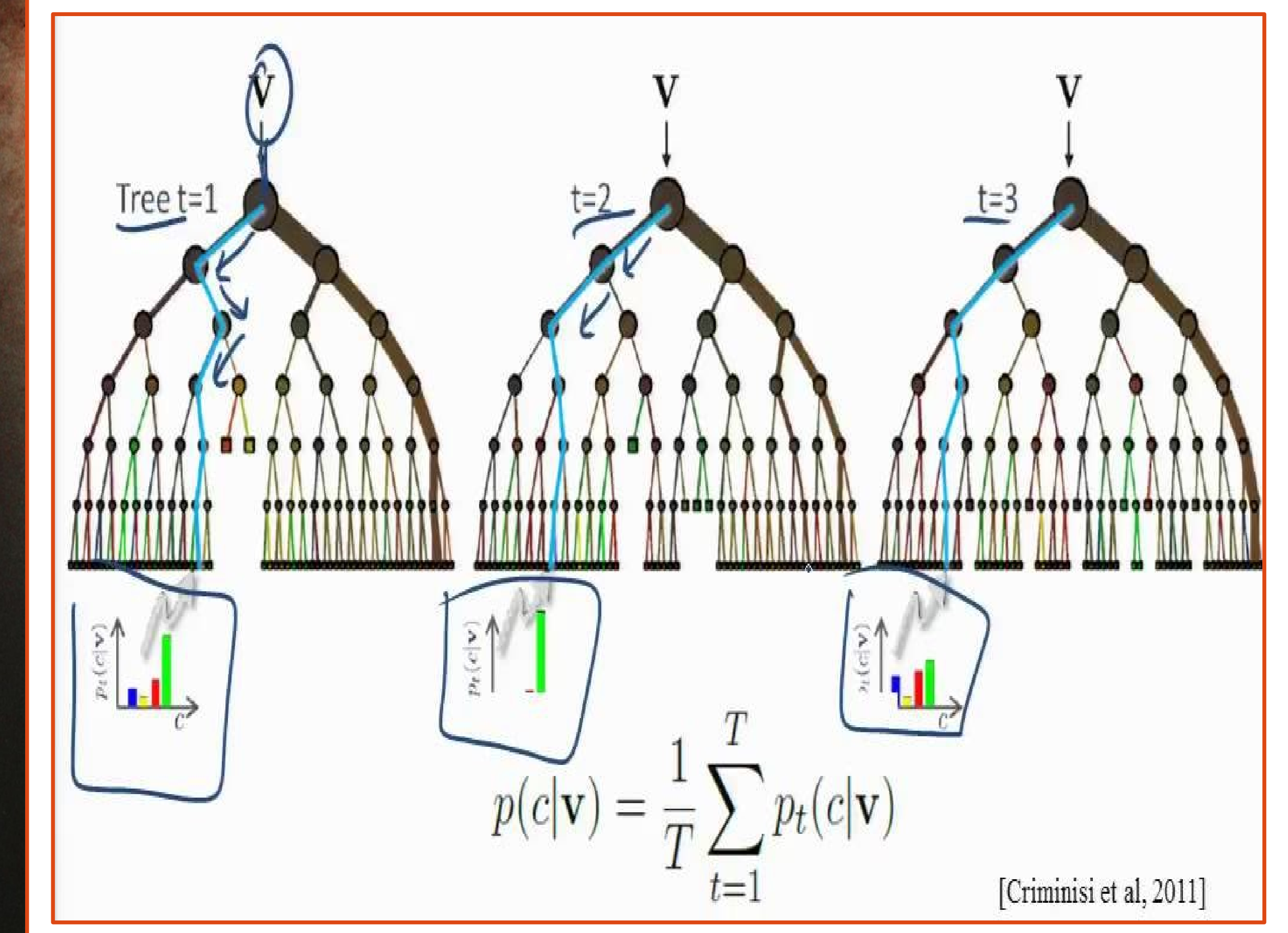
References

- [1] Adapted from sedielem's winning solution. See <http://benanne.github.io/2014/04/05/galaxy-zoo.html>.
- [2] Geurts, Pierre, Damien Ernst, and Louis Wehenkel. "Extremely randomized trees." *Machine learning* 63.1 (2006): 3-42.

Algorithm

Extra-Trees (ET)^[2]:
Ensemble of regression trees
(similar to Random Forests)

- more random
- no bagging



Conclusion

- Train large models with randomization to reduce over-fitting
- Feature selection if you don't have the computation