Mutation is all you need

Determinants of BANANAS' performance on large search spaces

Lennart Schneider, Florian Pfisterer, Martin Binder, Bernd Bischl

Department of Statistics, LMU Munich, Germany

Main Ablation Study

We vary the main ingredients of BANANAS [2] in a factorial manner and benchmark the different configurations on NAS-Bench-301 [1].

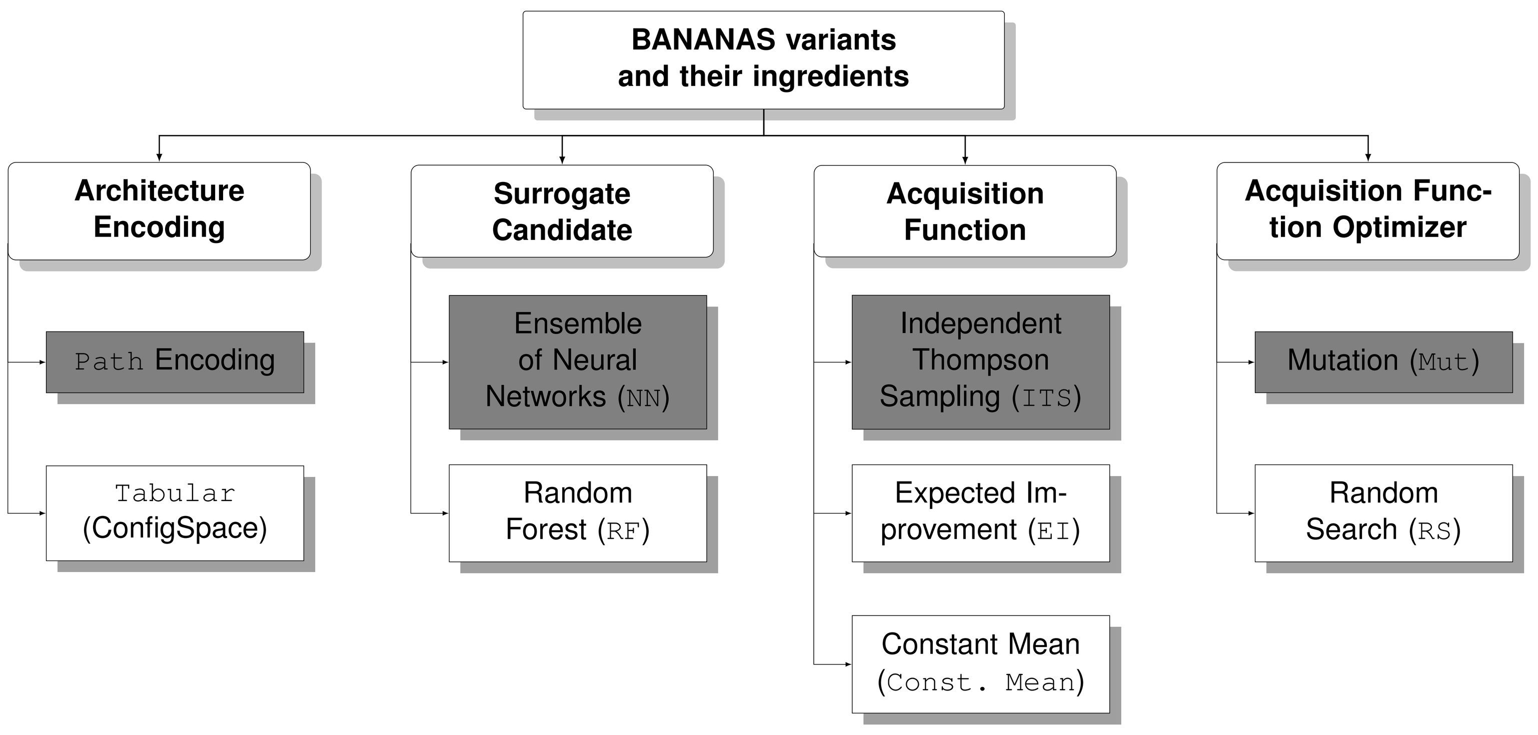

Figure 1: BANANAS variants and their ingredients architecture encoding, surrogate candidate, acquisition function and acquisition function optimizer. Default choices of BANANAS are colored in gray.

Results are given in Figure 2. The choice of acquisition function optimizer is by far the strongest determinant of performance as also indicated by a four-way ANOVA on the final performance (Table 1).

	Sum Sq	Df	F value	Pr(>F)
Architecture Encoding	0.41	1	19.57	0.0000
Surrogate Candidate	1.01	1	48.31	0.0000
Acquisition Function	0.56	2	13.49	0.0000
Acq. F. Optimizer	13.18	1	632.43	0.0000
Residuals	7.38	354		

Table 1: Results of a four-way ANOVA on the factors architecture encoding, surrogate candidate, acquisition function, and acquisition function optimizer. Type II sums of squares. Results are based on 20 replications.

Further Investigations

- Based on the `Tabular + RF + EI + Mut` configuration, the BO loop was run for 50 iterations (architecture evaluations)
- For edit distances ranging from 1 to 8, 100 test architectures were constructed each by mutating a fixed number of parameters (operations or edges) of the incumbent
- For these test architectures, Kendall's τ with respect to the predicted and true validation accuracy (Figure 3A), their true validation accuracy (Figure 3B) and their expected improvement and actual improvement (Figure 3C) was calculated

TL;DR

- We vary the main ingredients of BANANAS, architecture encoding, surrogate candidate, acquisition function and acquisition function optimizer in a factorial manner and benchmark the different configurations on NAS-Bench-301
- The choice of acquisition function optimizer (by default, BANANAS minimally mutates the incumbent architecture) is by the far the strongest determinant of performance
- Results hint that the surrogate model is not able to differentiate high performing architectures well
- Therefore, exploration of architectures or thorough optimization of the acquisition function may not be needed but minimally mutating the incumbent architecture is all you need

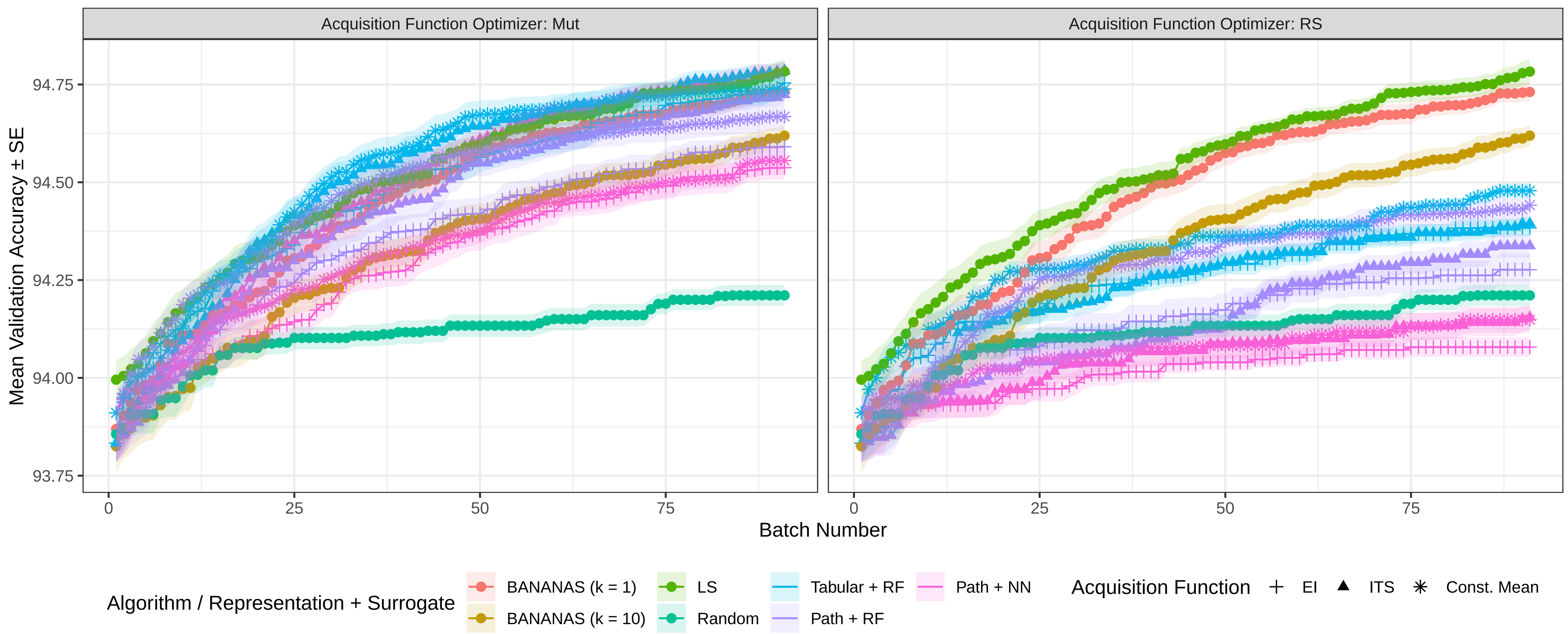

Figure 2: Different BANANAS configurations on NAS-Bench-301. Mean validation accuracy with standard error bands, higher is better. Color: optimization method and surrogate model. Facet: acquisition function optimizer, where applicable. Point shape: acquisition function, where applicable. The `ITS` acquisition function and `Mut` acquisition function optimizer is used for BANANAS methods, and local search [3] (`LS`) and random search (as NAS method; `Random`) do not use an acquisition function; their accuracy is therefore shown in both facets of the graph. Default BANANAS was run in two variants, updating its surrogate model every iteration ($k = 1$) or after $k = 10$ iterations. Results are based on 20 replications.

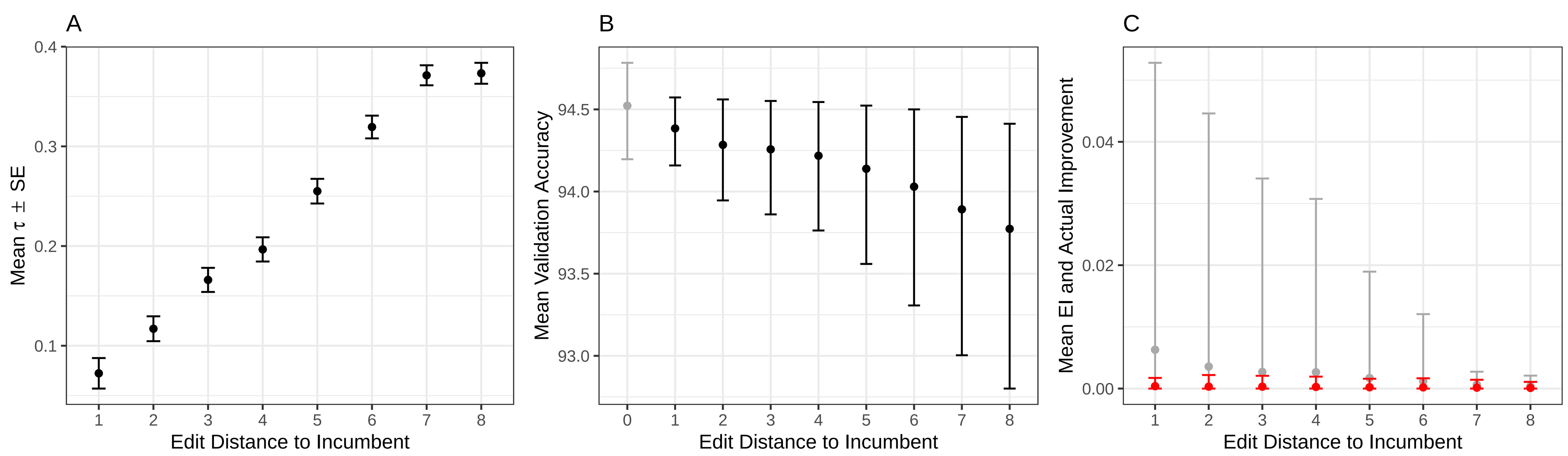

Figure 3: `Tabular + RF + EI + Mut` on NAS-Bench-301. A: Kendall's τ of the predicted and true validation accuracy of test architectures constructed to have different edit distances to the incumbent. B: True validation accuracy of these test architectures. Validation accuracy of the incumbent is given in gray. C: Expected Improvement (red) and actual improvement (gray) of these test architectures. Bars in B and C represent 2.5% and 97.5% quantiles. Results are based on 100 replications.

References

[1] J. Siems, L. Zimmer, A. Zela, J. Lukasik, M. Keuper, and F. Hutter. NAS-Bench-301 and the case for surrogate benchmarks for neural architecture search. arXiv:2008.09777 [cs.LG], 2020.

[2] C. White, W. Neiswanger, and Y. Savani. BANANAS: Bayesian optimization with neural architectures for neural architecture search. arXiv:1910.11858 [cs.LG], 2019.

[3] C. White, S. Nolen, and Y. Savani. Local search is state of the art for neural architecture search benchmarks. In *ICML Workshop on Automatic Machine Learning*, 2020.