Exploring the Loss Landscape in Neural Architecture Search

Colin White
Abacus.AI

Sam Nolen
Abacus.AI

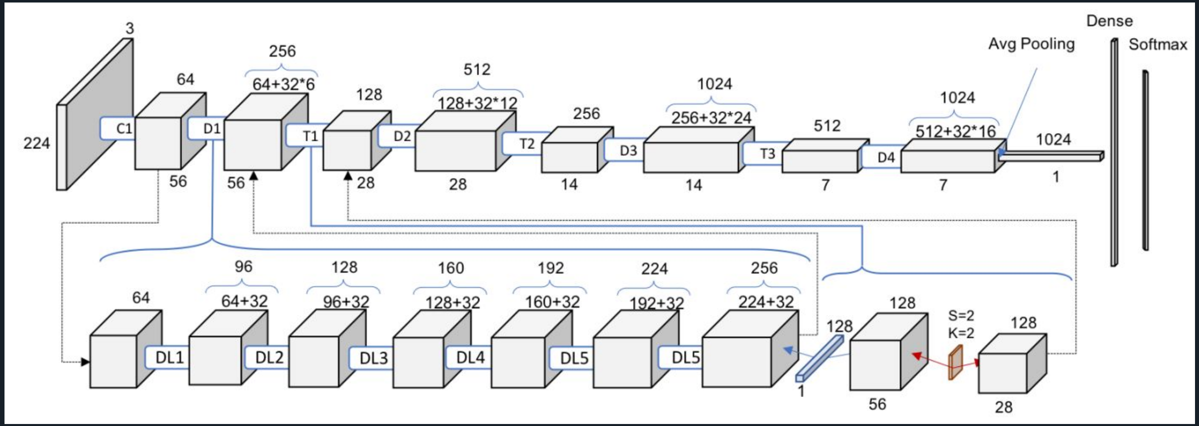
Yash Savani
Abacus.AI,
Carnegie Mellon University



Slides (with hyperlinks): https://crwhite.ml

Neural architecture search

Architectures are getting increasingly more specialized and complex

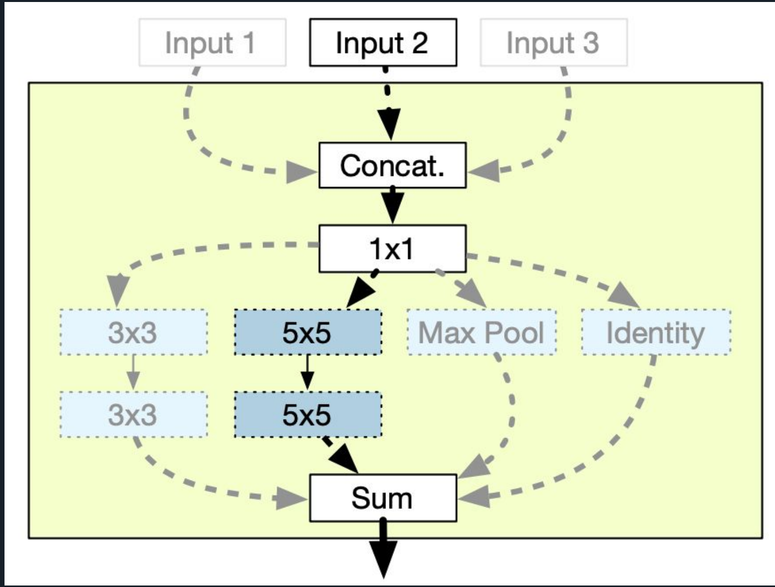


Source: https://towardsdatascience.com/understanding-and-visualizing-densenets-7f688092391a

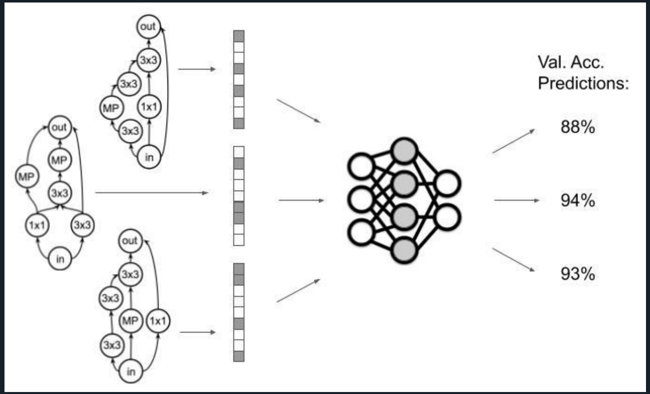
NAS Algorithms

NAS Algorithms

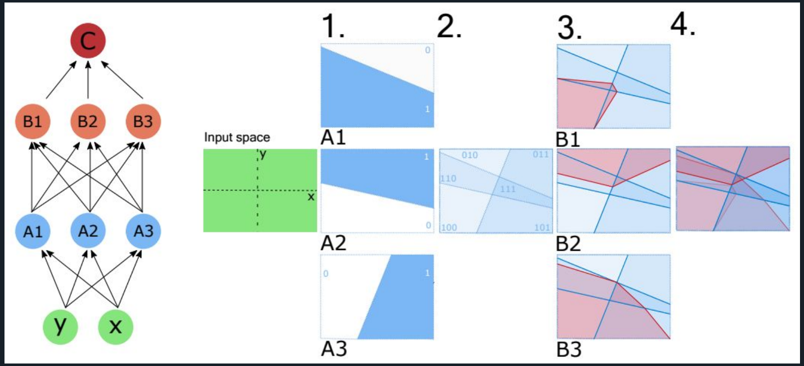
Algorithms are getting complex too



[Bender et al. 2018]



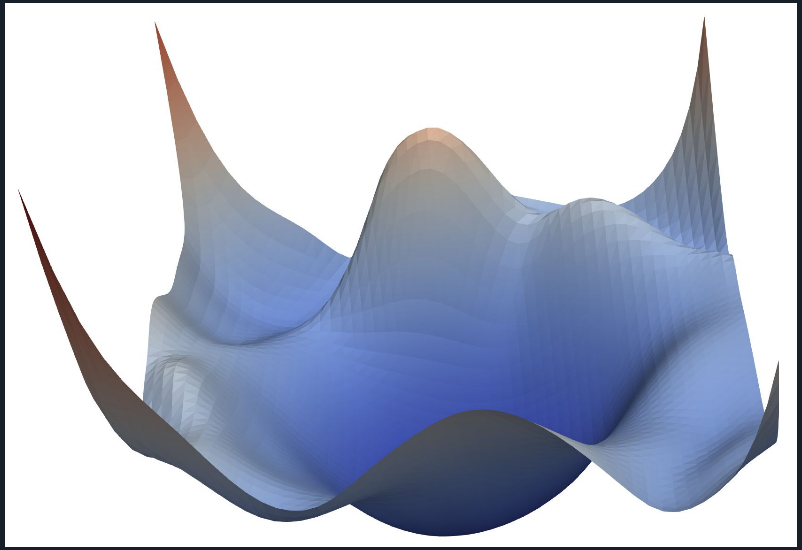
[White et al. 2021]



[Mellor et al. 2020]

Noise in the loss landscape

NAS: given a set *A*, find $max_{a \text{ in } A} f(a)$



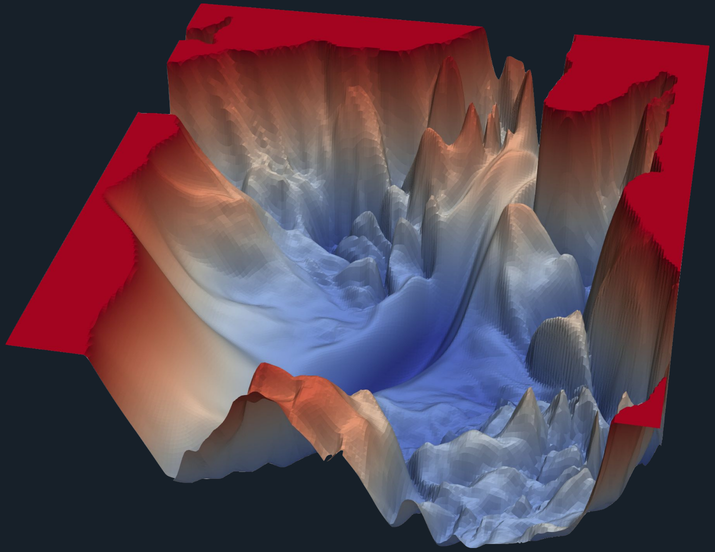
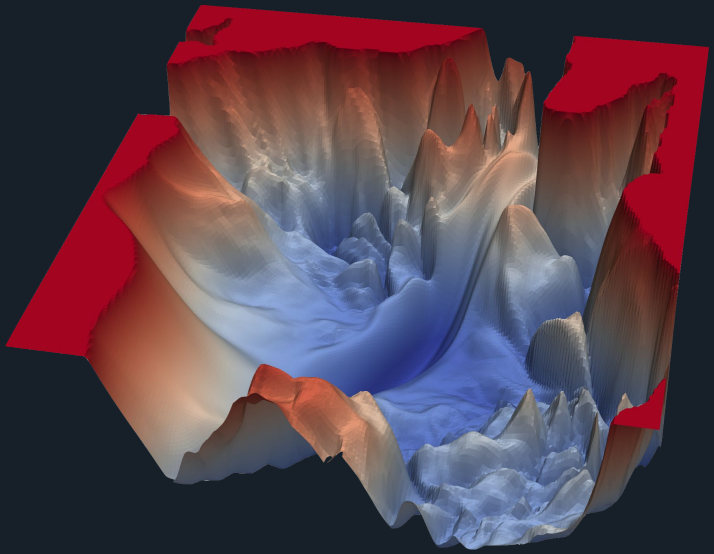


Image source: https://www.cs.umd.edu/~tomg/projects/landscapes/

Is NAS only hard because of noise?



Some recent NAS techniques have been designed specifically to deal with noisy architecture evaluations

Image source: https://www.cs.umd.edu/~tomg/projects/landscapes/

Outline

- Motivation
- Local Search
  - Strong baseline for NAS
  - SotA on de-noised search spaces
- Theoretical analysis
  - Characterization of local search
  - Characterization of number of local minima
  - Simulation results

ABACUS.AI

Local Search

- Five lines of code

Algorithm 1 Local search

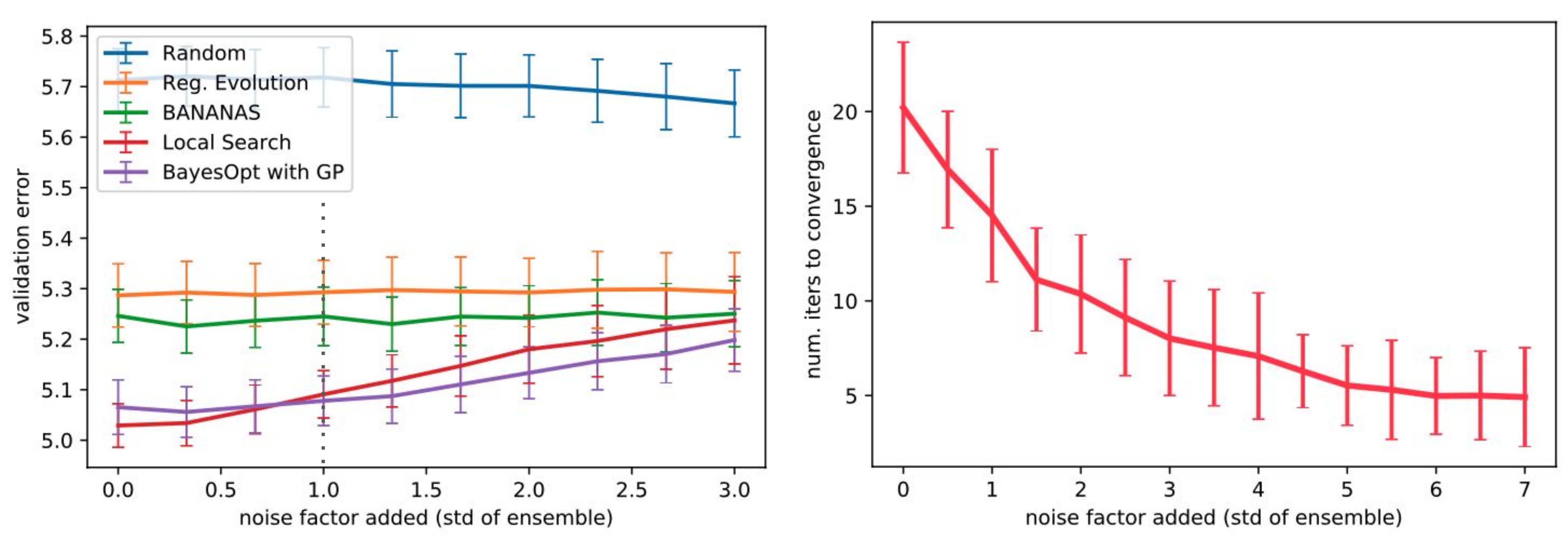
Input: Search space A, objective function ℓ, neighborhood function N

1. Pick an architecture $v_1 \in A$ uniformly at random
2. Evaluate $\ell(v_1)$; denote a dummy variable $\ell(v_0) = \infty$; set $i = 1$
3. While $\ell(v_i) < \ell(v_{i-1})$:
   i. Evaluate $\ell(u)$ for all $u \in N(v_i)$
   ii. Set $v_{i+1} = \text{argmin}_{u \in N(v_i)} \ell(u)$; set $i = i + 1$

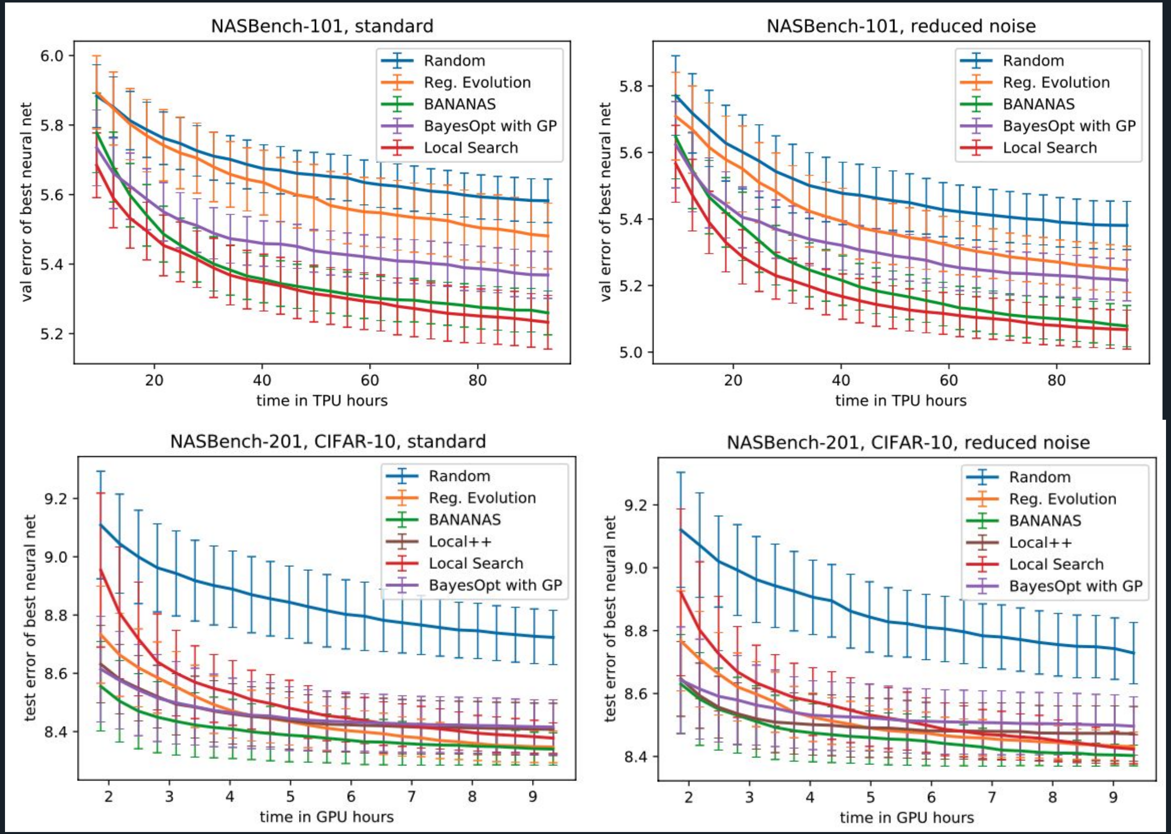
Output: Architecture v_i

Evaluate all architectures in the neighborhood

Performance vs. Noise on NAS-Bench-301



Performance vs. Noise on NASBench-101/201



NASBench-201: Computing #minima

Version	# iters	# local min.	% reached global min.
Denoised	5.36	21	47.4
Standard	4.97	55	6.71
Random	2.56	616	0.717

NAS-Bench-201 cifar10, 15,625 total architectures

Theoretical results

- Given the distribution of accuracies and neighborhood graph

Theorem 5.1. Given $|A| = n$, ℓ, s, ϵ, pdf_n, and pdf_e, we have

$$\mathbb{E}[|\{v \in A \mid \text{LS}^*(v) = v\}|] = n \int_{\ell(v^*)}^{\infty} \text{pdf}_n(x) \left(\int_x^{\infty} \text{pdf}_e(x,y)dy \right)^s dx, \text{ and}$$

$$\mathbb{E}[|\{v \in A \mid \ell(\text{LS}^*(v)) - \ell(v^*) \leq \epsilon\}|] = n \int_{\ell(v^*)}^{\ell(v^*)+\epsilon} \text{pdf}_n(x) \left(\int_x^{\infty} \text{pdf}_e(x,y)dy \right)^s \mathbb{E}[|\text{LS}^{-*}(x)|]dx.$$

Lemma 5.2. *Given A, ℓ, s, pdf_n, and pdf_e, then for all $v \in A$, we have the following equations.*

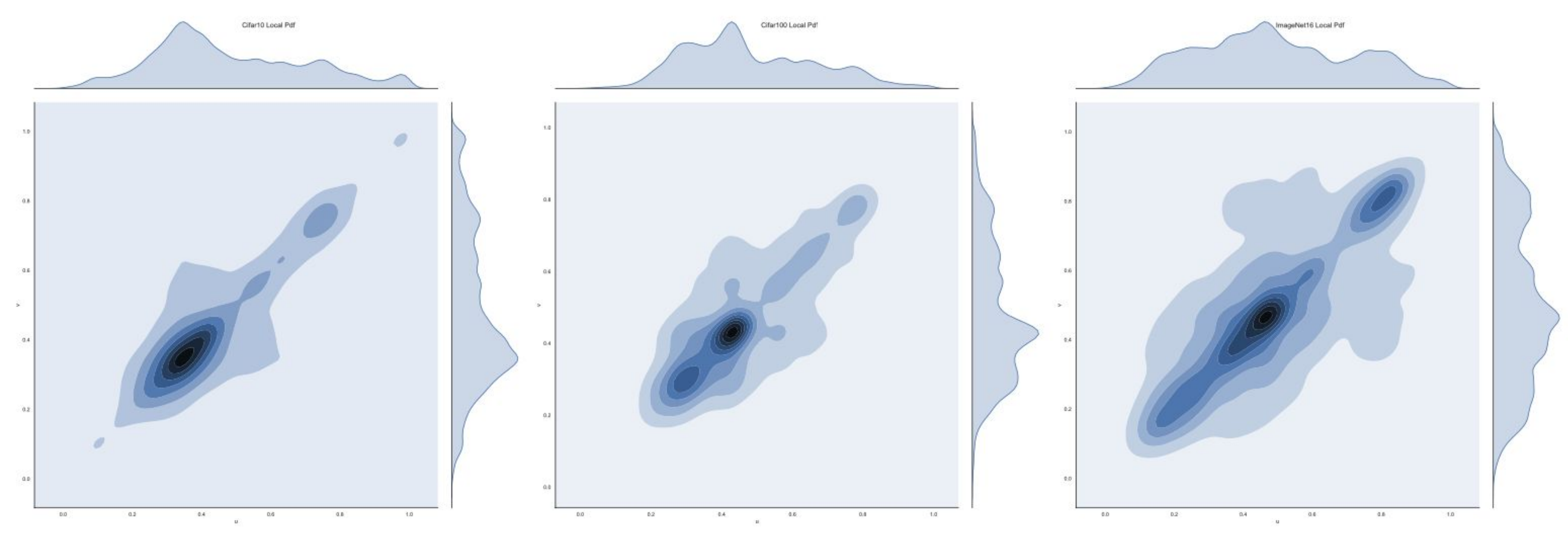
$$\mathbb{E}[|LS^{-1}(v)|] = s \int_{\ell(v)}^{\infty} pdf_e(\ell(v), y) \left(\int_{\ell(v)}^{\infty} pdf_e(y,z)dz \right)^{s-1} dy, \text{ and}$$

$$\mathbb{E}[|LS^{-k}(v)|] = b_{k-1} \cdot \mathbb{E}[|LS^{-1}(v)|] \left(\frac{\int_{\ell(v)}^{\infty} pdf_e(\ell(v),y)\mathbb{E}[|LS^{-(k-1)}(y)|]dy}{\int_{\ell(v)}^{\infty} pdf_e(\ell(v),y)dy} \right).$$

Example:
For uniform dist,
#local minima=
|search space|
/|neighbors|

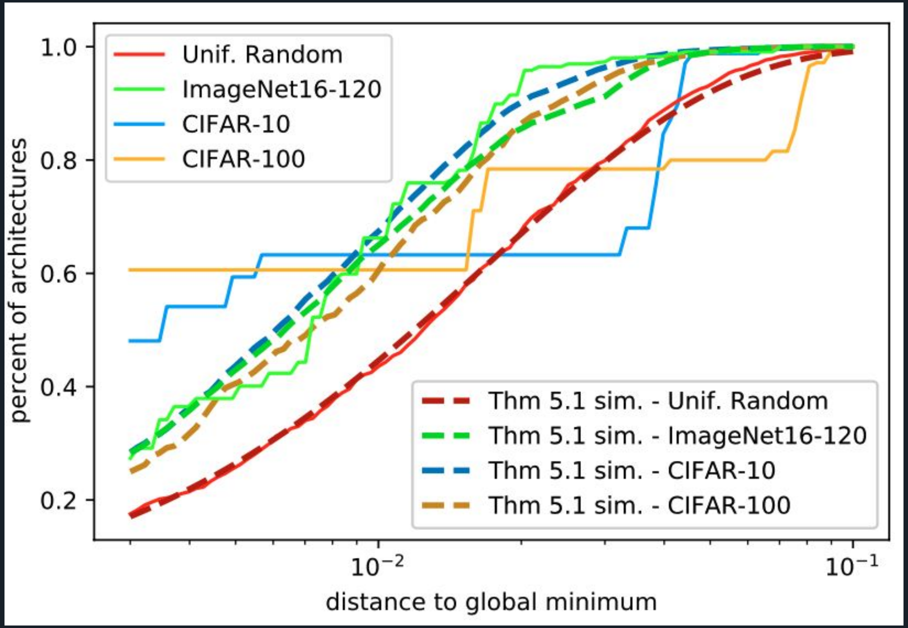
15625/24 = 651.04

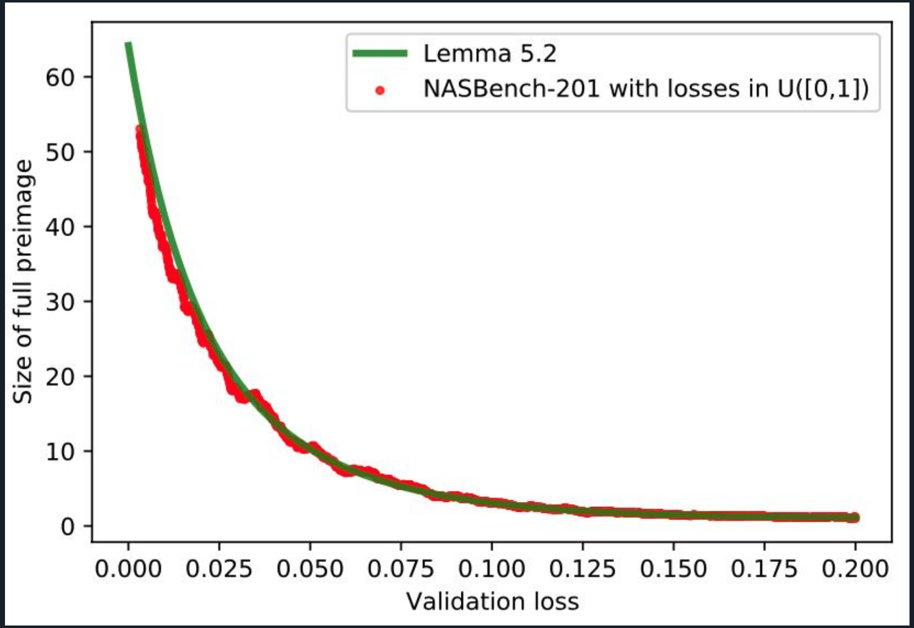
Neighborhood distributions



CIFAR-10 CIFAR-100 ImageNet16-120

Simulation Results





Conclusion

- Local search is a strong baseline for NAS
- Local search is SotA on de-noised search spaces
- Theoretical characterization of local search

https://github.com/naszilla/naszilla

Thanks!