Pi-NAS: Improving Neural Architecture Search by Reducing Supernet Training Consistency Shift

Supplementary Material

Jiefeng Peng[1,2]*, Jiqi Zhang[1]*, Changlin Li[3], Guangrun Wang[4]†, Xiaodan Liang[1], Liang Lin[1]
[1]Sun Yat-sen University [2]DarkMatter AI Research
[3]GORSE Lab, Dept. of DSAI, Monash University [4]University of Oxford
{jiefengpeng,wanggrun,xdliang328}@gmail.com,
zhangjq49@mail2.sysu.edu.cn, changlin.li@monash.edu, linliang@ieee.org

A. Appendix

A.1. Extensive Experiments on NAS-Bench-201

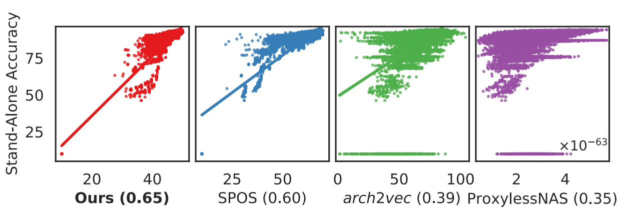


Figure 6: Ranking correlations on all 15625 architectures on NAS-Bench-201 [18] on CIFAR-10 compared to SPOS [20], *arch2vec* [59] and ProxylessNAS [9].

We also validate our Π-NAS's ranking correlation on all 15625 architectures on NAS-Bench-201 [18]. As shown in Figure 6, our method still outperforms SPOS [20], *arch2vec* [59] and ProxylessNAS [9]. However, our Π-NAS's advantage over SPOS [20] (0.65 *vs*. 0.60) shrinks compared to the submission's Figure 3 (NOT Figure 8 in this appendix) (0.70 *vs*. 0.57).

We assume the ranking correlation degradation might be attributed to *skip connection* operation in the search space. To justify this assumption, we provide some visualization in Figure 7. Specifically, we replace a non-skip-connection operation (*i.e*., *zero,* 1×1 *convolution,* 3×3 *convolution* or 3×3 *average pooling*) with a *skip connection* for all architectures in the search space. Then, we compare the estimated accuracy change *vs*. actual accuracy change before and after such a replacement. When an architecture's estimated accuracy change is smaller than 0.01, we plot its actual accuracy change (absolute value) in Figure 7 for Π-NAS and SPOS, respectively. As shown, there is a significant gap between the estimated accuracy change and actual accuracy change before and after a *skip connection* replacement (*i.e*., **before:** < 0.01; **after:** usually > 1). This visualization indicates *skip connection* operation does hurt ranking correlation for both Π-NAS and SPOS, verifying our assumption.

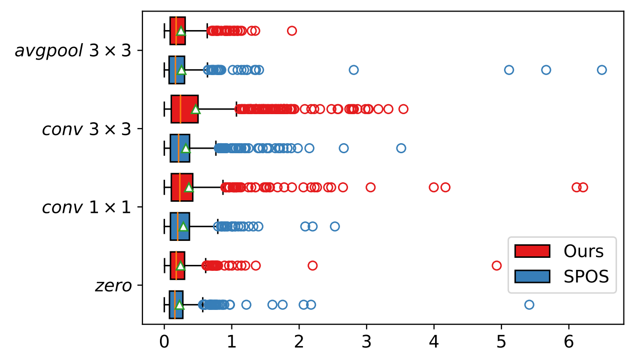


Figure 7: The actual accuracy change (absolute value) of architecture satisfying that the estimated accuracy change is smaller than 0.01 after replacing one operation (*zero,* 1×1 *convolution,* $3{\times}3$ *convolution* or $3{\times}3$ *average pooling*) with *skip connection* for Π-NAS and SPOS [20], respectively.

Previous works have also observed this ranking correlation degradation. As pointed out in [15, 30], skip connection can increase supernet scalability in the depth dimension, but it can lead to convergence difficulty of the supernet and unfair comparison of subnets. In addition, our cross-path learning is more prone to this problem since we directly reduce the feature consistency shift between different paths whether with *skip connection* operation or not without special treatment, which can cause the overestimated performances of the architectures containing *skip connection*.

*Jiefeng Peng and Jiqi Zhang are co-first authors and share equal contributions. Their names are listed in alphabetical order.
†Corresponding Author.

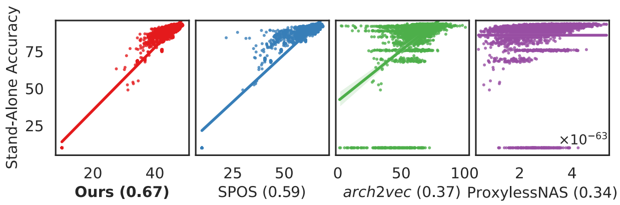
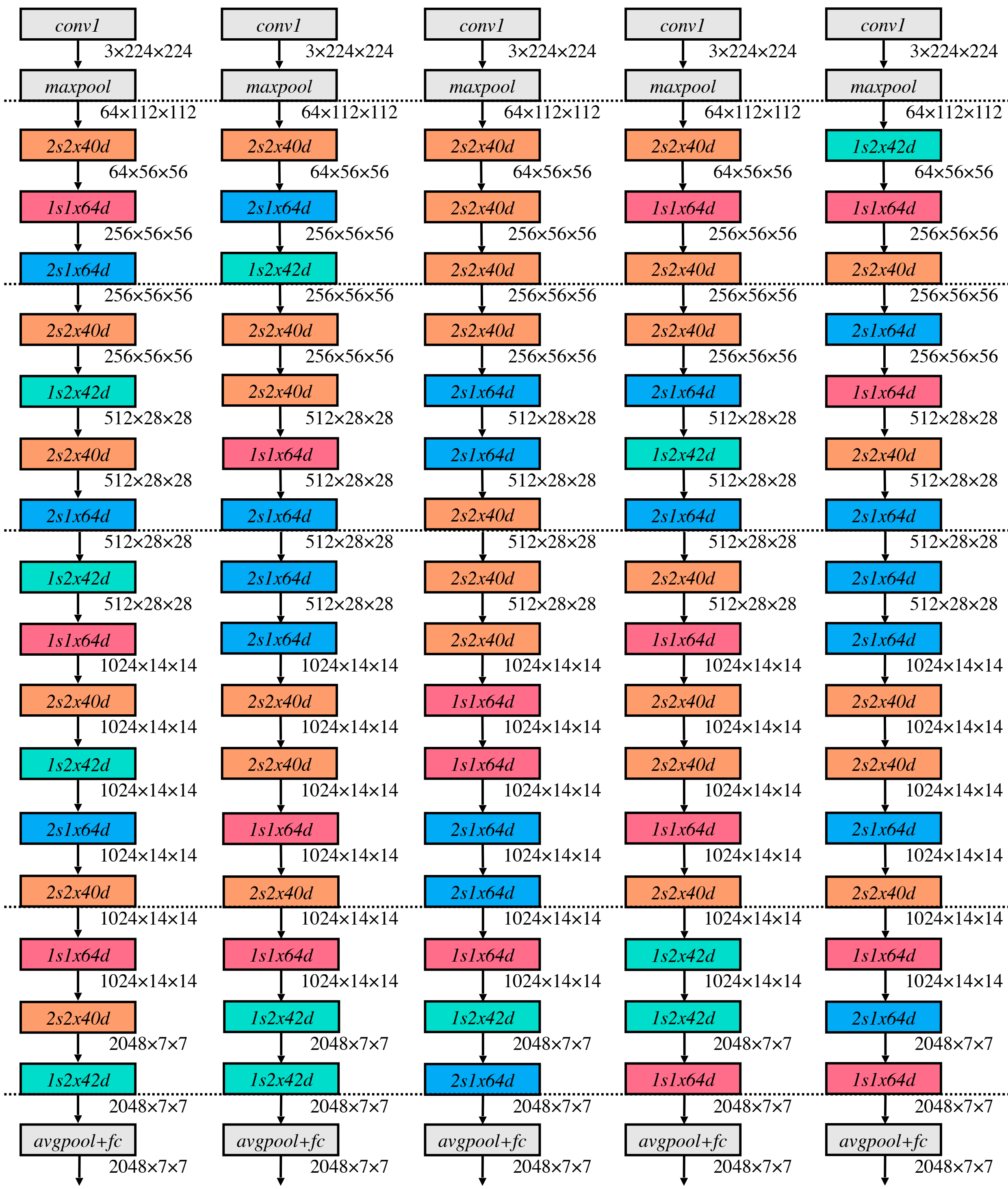


Figure 8: Ranking correlations on 4096 architectures on NAS-Bench-201 [18] on CIFAR-10 without *skip connection* operation compared to SPOS [20], *arch2vec* [59] and ProxylessNAS [9].

As Figure 8 shows, after leaving out the architectures with *skip connection* operation, Π-NAS's advantage recovers. For future work, we will try to develop the scalability of Π-NAS to solve such limitation.

There is one more page showing our searched architectures. Don't hesitate to scroll your mouse.

A.2. Model Architectures



(a) Π-NAS-α *(cls)* (b) Π-NAS-β (c) Π-NAS-γ *(trans)* (d) Π-NAS-δ (e) Π-NAS-ϵ

Figure 9: Architectures of Π-NAS-α (*i.e.*, Π-NAS-*cls*), Π-NAS-β, Π-NAS-γ (*i.e.*, Π-NAS-*trans*), Π-NAS-δ and Π-NAS-ϵ. They are combinations of *Split-Attention* block with radix s, cardinality x and width d.