HyperNAS: Enhancing Architecture Representation for NAS Predictor via Hypernetwork

Supplementary Material

7. Implement Details

Training settings. During each training step, we randomly select an architecture-accuracy pair and a batch of 128 images for training HyperNAS. We implement a gradient accumulation strategy that updates the gradients every 10 steps. The batch of images is drawn from the auxiliary dataset $\mathcal{D}_{aux}$ and is used to assess the performance of the target network constructed based on the selected architecture. HyperNAS undergoes training for 200 epochs, employing the ADAM optimizer with an initial learning rate of 1e-3, which is halved at epochs 100 and 150. It's important to note that batch statistics in Batch Normalization (BN) operations are computed directly by the PyTorch framework, rather than being predicted by the hypernetwork. Given the fixed output dimension of the hypernetwork, the output of the hypernetwork is reshaped and sliced based on the parameter shapes in each node. Additionally, to maintain a balance between the dual tasks within HyperNAS, we set the preference factor q to 1.5.

GCN encoder settings. The shared GCN encoders in HyperNAS are configured differently depending on the search space. For the DARTS [30] search space, the encoder has 3 layers with a hidden dimension of 144. For the ViT search space, it also has 3 layers but with a hidden dimension of 256. For other search spaces [15, 19, 65], the encoder consists of 6 layers with a hidden dimension of 72.

Hypernetwork settings. In HyperNAS, the architecture of the shared hypernetworks is customized according to the specific search space. For the DARTS [30] search space, the hypernetwork is designed with 2 layers, each having a hidden dimension of 128. For ViT, the hypernetwork comprises 4 layers, each also featuring a hidden dimension of 128. Conversely, for other search spaces [15, 19, 65], the hypernetwork is configured with 3 layers, each with a hidden dimension of 96.

	Supernet-small	Supernet-base
Embed Dim	(320, 448, 64)	(528, 624, 48)
Q-K-V Dim	(320, 448, 64)	(512, 640, 64)
MLP Ratio	(3, 4, 0.5)	(3, 4, 0.5)
Head Num	(5, 7, 1)	(8, 10, 1)
Depth Num	(12, 14, 1)	(14, 16, 1)
Params Range	14–34M	42–75M

Table 5. Predefined ViT search space

8. Further Experiments

8.1. Comparisons with zero-cost NAS methods.

Table 6 summarizes the comparison results with recent popular zero-cost NAS methods [1, 14, 23, 24, 27, 39, 49, 52, 54]. The methods are grouped into two categories: those based on pruning-inspired proxies and those derived from theoretical proxies. Following the same experimental setup as ParZC [14], we find that HyperNAS consistently outperforms all existing approaches by a significant margin. These results confirm that HyperNAS achieves an excellent trade-off between efficiency and predictive accuracy.

8.2. Search Results on NAS-Bench-201

To further evaluate the effectiveness of HyperNAS, we conducted experiments to search for optimal architectures on NAS-Bench-201. Experiments are performed with three independent runs, with the predictor trained on 156 training samples. We use the current state-of-the-art predictor, PINAT, as our baseline for comparison. The search results, shown in Table 7, demonstrate that HyperNAS identifies superior architectures and exhibits strong robustness across all datasets.

8.3. Search Results on MobileNetV3

Search space. MobileNetV3 consists of 5 stages, with each stage containing 2 to 4 building blocks. This results in a maximum of 20 layers (5 stages × 4 blocks per stage). Each building block offers flexibility in its configuration: the kernel size is selected from $\{3, 5, 7\}$, while the expansion ratio is chosen from $\{3, 4, 6\}$.

For MobileNetV3, we represent each architecture with five normal cells, considering each stage as a cell. A shared GCN encoder is employed to encode the various cells. Moreover, we evaluate the searched architecture using the strategy from CTNAS [10] and restrict the number of FLOPs to less than 500M. As shown in Table 8, HyperNAS achieves the highest accuracy with just 200 queries, a query count that is orders of magnitude fewer than other methods. These outcomes further emphasize the robustness of HyperNAS in few-shot scenarios.

8.4. Ablation Studies

Ablation studies on NAS-Bench-201. Figure 6 presents the ablation results on NAS-Bench-201 [15], showing similar trends as those on NAS-Bench-101 [65]. As depicted in Figure 6a, HyperNAS-P significantly improves ranking

	NB101-CF10		NB201-CF10		NB201-CF100		NB201-IMG16	
	SP	KD	SP	KD	SP	KD	SP	KD
Fisher	-0.28	-0.20	0.50	0.37	0.54	0.40	0.48	0.36
GradNorm	-0.25	-0.17	0.58	0.42	-0.63	0.47	0.57	0.42
GraSP	0.27	0.18	0.51	0.35	0.54	0.38	0.55	0.39
L2Norm	0.50	0.35	0.68	0.49	0.72	0.52	0.69	0.5
SNIP	-0.19	-0.14	0.58	0.43	-0.63	0.47	0.57	0.42
Synflow	0.31	0.21	0.73	0.54	0.76	0.57	0.75	0.56
NWOT	0.31	0.21	0.77	0.58	0.80	0.62	0.77	0.59
Zen	0.59	0.42	0.35	0.27	0.35	0.28	0.39	0.29
ZiCo	0.63	0.46	0.74	0.54	0.78	0.58	0.79	0.60
EZNAS	0.68	0.45	0.83	0.65	0.82	0.65	0.78	0.61
ParZC	0.83	0.64	0.90	0.71	0.91	0.74	0.88	0.70
HyperNAS	**0.94**	**0.80**	**0.96**	**0.85**	**0.97**	**0.83**	**0.95**	**0.81**

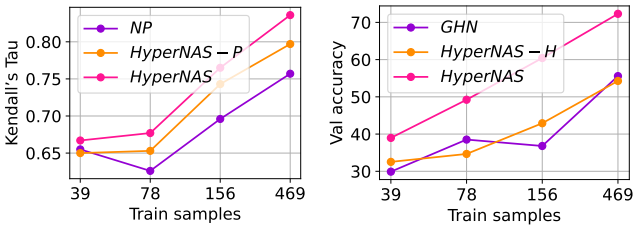
Table 6. Comparison results with zero-cost NAS methods. SP and KD denote Spearman and Kendall's Tau correlation coefficients, respectively.

Method	CIFAR-10		CIFAR-100		ImageNet-16-120	
	validation	test	validation	test	validation	test
PINAT	91.03±0.54	94.09±0.14	73.17±0.31	73.26±0.19	45.95±0.21	**46.55±0.72**
HyperNAS	**91.57±0.00**	**94.37±0.00**	**73.43±0.01**	**73.40±0.03**	**46.65±0.02**	46.50±0.00

Table 7. Search results on NAS-Bench-201. Three runs are conducted with different random seeds.

Method	Param (M)	FLOPs (M)	Top-1 (%)	Top-5 (%)	Queries (K)
MobileNetV3-like					
MobileNetV3	5.4	219	75.2	-	-
FBNet-C	5.5	375	74.9	-	11.5
MnasNet-A3	5.2	403	76.7	93.3	8
ProxylessNAS	7.1	465	75.1	92.3	-
OFA	-	230	76.0	-	16
SPOS	5.4	472	74.8	-	-
RLNAS	5.3	473	75.6	92.6	-
NP	6.4	536	74.8	-	-
FBNetV2	-	321	76.3	92.9	11.5
AtomNAS	-	367	75.9	92.0	78
CTNAS	-	482	77.3	93.4	1
PINAT	5.1	452	77.8	93.5	-
HyperNAS	5.7	428	**77.9**	**93.6**	**0.2**

Table 8. Performance comparison of architectures searched by SOTA methods on ImageNet in the MobileNetV3 search space.



(a) Ablation studies on ranking abilities of the neural predictor.

(b) Ablation studies on parameter prediction of the hypernetwork.

Figure 6. Ablation studies on NAS-Bench-201 search space.

Connectivity	50 (0.01%)	100 (0.02%)	172 (0.04)%)	424 (0.1)%
Concatenate	0.675	0.684	0.714	0.778
Sum	**0.678**	**0.688**	**0.736**	**0.779**

Table 9. Comparison results for cell connectivity strategy.

abilities compared to NP, while HyperNAS explicitly outperforms HyperNAS-P. These improvements demonstrate the effectiveness of the proposed global encoding scheme and the shared hypernetwork. When compared to single-task HyperNAS-P and HyperNAS-H, HyperNAS exhibits superior performance in both tasks, as illustrated in Figures 6a and 6b. These findings underscore the compatibility between neural predictors and the shared hypernetwork, as well as the effectiveness of our multi-task loss function.

Effect of GCN encoder on the hypernetwork. To assess the impact of the GCN encoder on the hypernetwork, we conducted experiments on GHN [66] and HyperNAS-H. GHN uses a GNN encoder with forward-backward propagation, whereas HyperNAS-H employs a specially designed GCN encoder. Figure 6b shows that HyperNAS-H consistently achieves higher validation accuracy than GHN, demonstrating the superior capability of our encoder in capturing architectural information.

Effect of cell connectivity. Table 9 presents the ablation studies of cell connectivity. It can be seen that the per-

formance of the concatenate strategy is slightly weaker than the sum strategy, showcasing the strength of the sum strategy in capturing the macro-structure of architectures. Additionally, the sum strategy is more parameter-efficient, making it a preferable choice.

8.5. Efficiency Evaluation

Training cost. Table 10 reports training costs across different search spaces. MobileNetV3 is trained on ImageNet-1K, while other search spaces are tested on the CIFAR10. The results show that resource consumption remains within an affordable range, including for ViT on ImageNet-1K.

Search space	NB101	NB201	DARTS	ViT
Training time	0.3	0.2	0.3	2

Table 10. Training cost across different search spaces, measured in GPU days. ViT is trained on **ImageNet**, while the others are trained on CIFAR-10.

Evaluation cost. Table 11 presents the time cost comparison for evaluating a single neural architecture in NAS-Bench-101. Notably, HyperNAS achieves a time cost 1000 times lower than GHN which directly employs hypernetworks for architecture evaluation, highlighting the high efficiency of our proposed paradigm. Moreover, HyperNAS achieves comparable efficiency to the current state-of-the-art predictor PINAT, indicating its practical feasibility.

Method	NP	GHN	PINAT	HyperNAS	Full training
Time cost	0.002s	14s	0.012s	0.014s	8h

Table 11. Time cost of evaluating a single architecture in NAS-Bench-101.

9. Visualization of Searched Architectures

DARTS. Figures 8 and 9 display architectures found by HyperNAS-P and HyperNAS on CIFAR-10, which achieve top-1 accuracy of 97.61% and 97.60% on CIFAR-10, respectively. Interestingly, in the architecture discovered by HyperNAS (Figure 9), all operations within the normal cell are "sep_conv_3×3", and all within the reduction cell are "sep_conv_5×5", with no parameter-free operations (e.g., skip connection or max pooling). By comparing with the architecture found by HyperNAS-P (Figure 8), this phenomenon can be attributed to the shared hypernetwork in HyperNAS, which is employed to generate weights for various architectures.

MobileNetV3. Figure 7 displays the architecture found by HyperNAS in the MobileNet-v3 search space on ImageNet [22].

ViT. Table 12 displays the architecture found by HyperNAS in the ViT search space on ImageNet [22].

method	depth	mlp ratio	head nums	embead dim
HyperNAS-s	13	[3.5, 4.0, 3.5, 4.0, 4.0, 3.5, 4.0, 4.0, 4.0, 4.0, 4.0, 3.5, 3.5]	[5, 5, 7, 6, 6, 5, 7, 6, 5, 7, 6, 5, 6]	384
HyperNAS-b	14	[3.0, 3.5, 3.5, 3.5, 4.0, 3.5, 4.0, 3.0, 3.5, 4.0, 4.0, 3.5, 4.0, 3.5]	[9, 9, 9, 10, 9, 9, 10, 10, 10, 10, 9, 9, 10, 10]	576

Table 12. The architectures found by HyperNAS in the ViT search space on ImageNet.

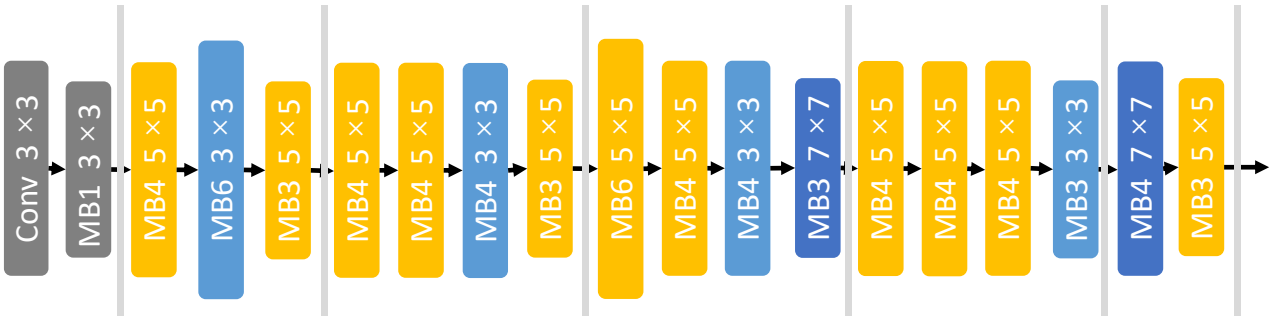
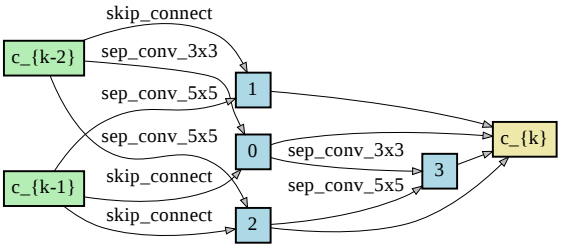
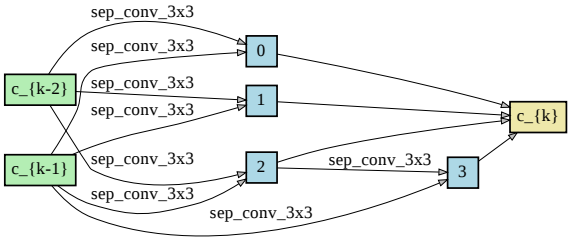


Figure 7. The architecture found by HyperNAS in the MobileNetV3 search space on ImageNet.

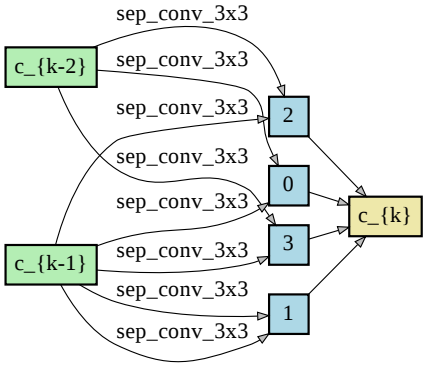


(a) Normal cell.

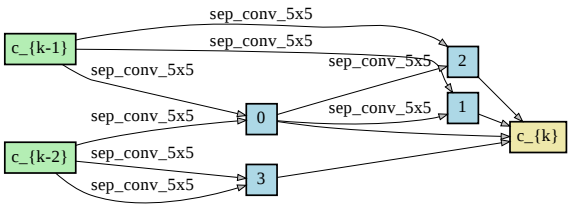


(b) Reduction cell.

Figure 8. The architecture of the convolutional cells found by HyperNAS-P in the DARTS search space.



(a) Normal cell.



(b) Reduction cell.

Figure 9. The architecture of the convolutional cells found by HyperNAS in the DARTS search space.