A. Supplementary Material

In this supplementary section, we discuss the following details, which could not be included in the paper owing to space constraints.

- Details of the randomly sampled architectures used in Section 4.4.
- A quantitative and qualitative comparison of searched, randomly sampled architectures used in Section 4.4.
- Comparison of NAS and hand-crafted architectures for black-box attacks on CIFAR-10 dataset.

A.1. Details of Randomly sampled architectures from DARTS search space

In Section 4.4, we report results over four random runs, details of each of these runs are shown in Table 7. '# Networks' denotes the number of sub-networks in a given ensemble, and '# Cells' denote the number of cells in each sub-networks. '# Epochs' denotes the number of epochs each sub-network is trained for.

Run #	# Networks	# Cells	# Epochs
1	3	{12, 6, 2}	{360, 180, 40}
2	5	{4, 4, 4, 4, 4}	{120, 120, 120, 120, 120}
3	5	{6, 5, 4, 3, 2}	{180, 150, 120, 90, 60}
4	3	{16, 2, 2}	{480, 60, 60}

Table 7. Details of sub-networks in each ensemble across four runs

A.2. Comparison of Searched and Randomly sampled architectures

In this section, we make a qualitative and quantitative comparison between the randomly sampled architectures (used in Section 4.4) with standard SoTA DARTS based architectures like DARTS [21], P-DARTS [5], and PC-DARTS [44]. For this study, we randomly choose 3 of the 16 randomly sampled sub-networks shown in Table 7.

	Max Pool	Avg Pool	Skip connection	Separable Conv	Dilated Conv	# unique operations
DARTS	0	0	2	5	1	3
P-DARTS	0	0	2	4	2	3
PC-DARTS	0	1	1	4	2	4
R1	1	1	2	1	3	**5**
R2	1	1	3	2	1	**5**
R3	1	1	1	2	3	**5**

Table 8. Comparison on usage of different operations in **normal** cell of micros from DARTS search space; R1, R2, R3 denote three randomly sampled micros.

DARTS search space consists of 5 operations: Max Pooling, Average Pooling, Skip Connections, Separable and Dilated convolutions. We report the number of times these operations are used in the normal and reduce cells of different DARTS architectures/micros in Tables 8 and 9. When compared with searched cells, randomly sampled ones, in general, have more number of unique operations. In a searched micro, the maximum number of unique operations for the normal, reduce cells in three, four, respectively. For a randomly sampled architecture, the count is five for both normal and reduce cells. Searched cells have many occurrences of a single operation (Separable convolution), which is not the case in randomly sampled architectures. We hypothesize that the presence of diverse set operations is a plausible for improved adversarial accuracy of randomly sampled DARTS architectures. While we only show 3-randomly chosen sub-networks in Tables 8, 9, we observe similar inferences with other random choices for sub-networks too. A qualitative comparison of these operations is shown in Figures 5 and 6.

	Max Pool	Avg Pool	Skip connection	Separable Conv	Dilated Conv	# unique operations
DARTS	5	0	3	0	0	2
P-DARTS	1	1	0	2	4	3
PC-DARTS	1	0	0	7	0	2
R1	1	1	1	2	3	**5**
R2	1	1	2	3	1	**5**
R3	1	1	2	3	1	**5**

Table 9. Comparison on usage of different operations in **reduce** cell of micros from DARTS search space; R1, R2, R3 denote three randomly sampled micros;

A.3. More Results: Adversarial Accuracy on Black-box Attacks

Table 10 shows the robustness of different architectures for black-box attacks on CIFAR-10. In a black-box setting, the robustness of a model is tested on adversarial examples generated using a source model. For the source model, we use two variants of hand-crafted models and two variants of NAS models. Since we use no adversarial training in our experiments, the accuracy numbers are relatively lower. In general, the average adversarial accuracy on hand-crafted architectures is higher than NAS methods, having complex operations in the topology (DenseNet, ProxylessNAS) or a large number of parameters (VGG-16) provide the best adversarial accuracy in a black-box setting. Either way, our hypothesis that the complexity of the architecture (in terms of operations or parameters) plays a major role is corroborated in this study too. Understanding specific details of the nature and origin of such complexity in topology may be an interesting direction of future work.

Target ↓ Source →	ResNet-18	DesneNet-169	DARTS	NSGA Net
ResNet-18	9.59	9.51	9.59	9.60
ResNet-50	10.96	10.98	10.97	10.92
DenseNet-121	11.69	11.67	11.66	11.66
DenseNet-169	9.74	9.71	9.74	9.71
VGG16 BN	**13.96**	**13.64**	**13.85**	**13.94**
DARTS	10.05	10.04	10.04	10.05
PDARTS	9.95	9.98	10.07	10.11
NSGA Net	9.85	9.86	9.90	9.88
Proxyless-NAS	11.97	11.89	11.95	11.97
PC-DARTS	7.96	8.02	8.1	7.94

Table 10. Comparison of adversarial accuracy for different Black-Box attacks on CIFAR-10 dataset

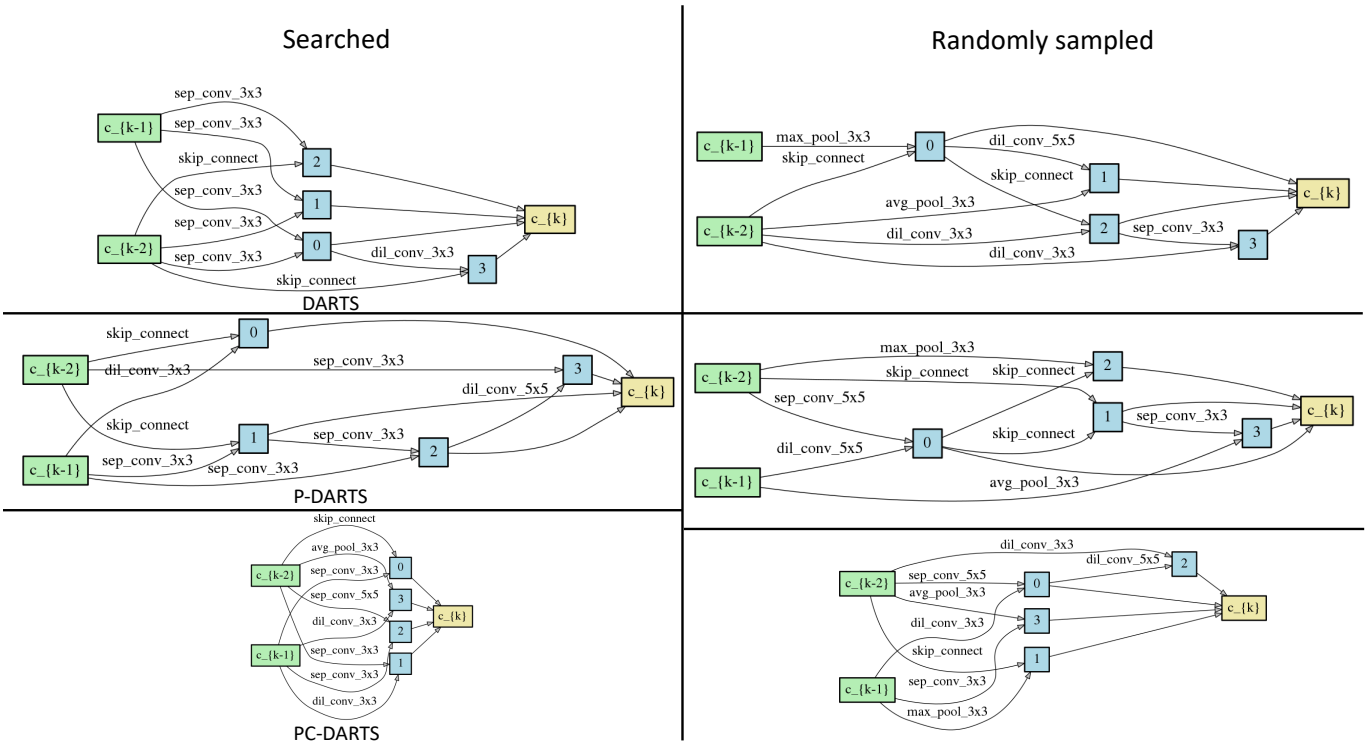

Figure 5. Qualitative comparison of different operations in **normal** cells of searched, randomly sampled micros from DARTS search space; *Left:* Searched micros, *Right:* Randomly sampled micros; Randomly sampled ones, in general, have more unique operations than searched ones

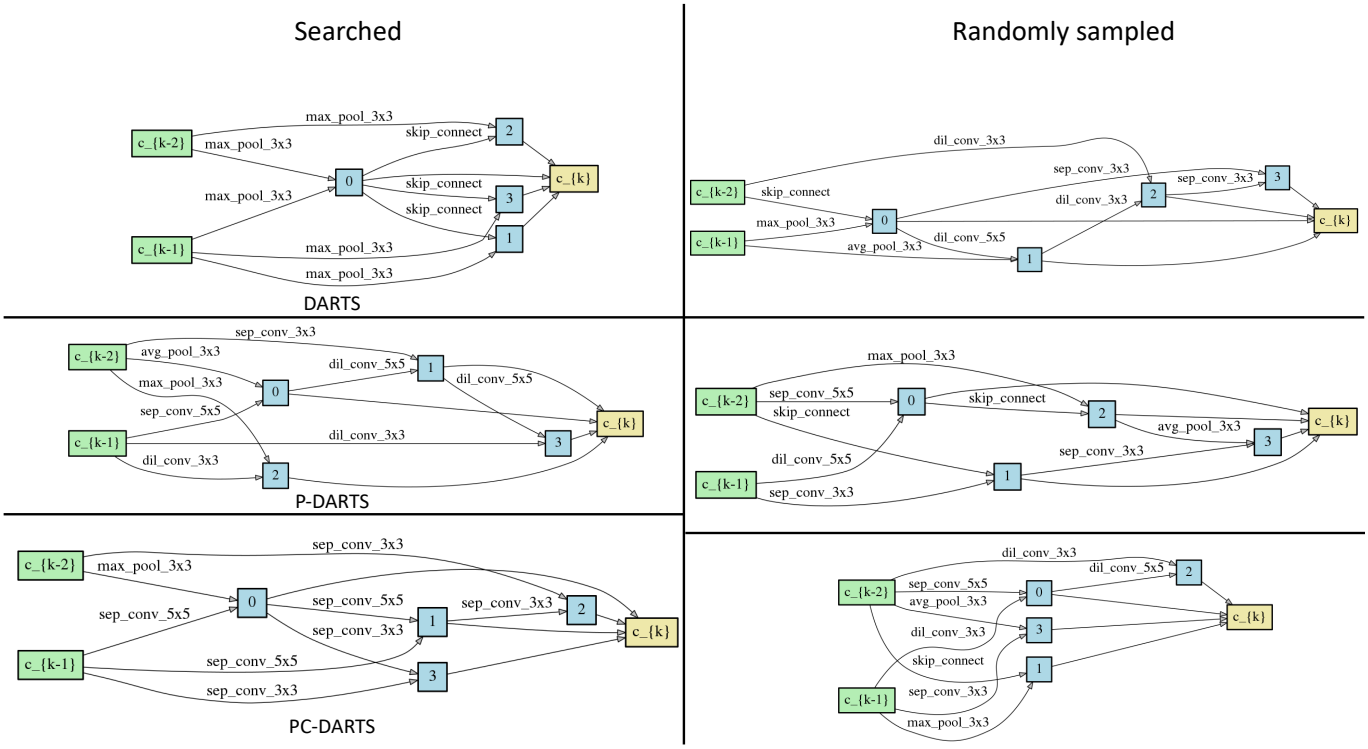

Figure 6. Qualitative comparison of different operations in **reduce** cells of searched, randomly sampled micros from DARTS search space; *Left:* Searched micros, *Right:* Randomly sampled micros; Randomly sampled ones, in general, have more unique operations than searched ones