# Template-Based Automatic Search of Compact Semantic Segmentation Architectures

Vladimir Nekrasov  Chunhua Shen  Ian Reid
The University of Adelaide, Australia
E-mail: {vladimir.nekrasov, chunhua.shen, ian.reid}@adelaide.edu.au

## Supplementary Material

## 1. Analysis of Search Results

In addition to the discussion in the main text, we are following up on two more questions: i.) whether there appears to be any correlation between the number of parameters of a sampled architecture and its performance, and ii.) what templates lead to larger rewards.

### Number of Parameters

First, we consider the distribution of rewards based on the number of parameters (Fig. 1). From it, the size of the architecture appears to have no connection with its reward. A more detailed plot tells a different story, though (Fig. 2): while for small architectures ($\leq$ 250K) the rewards are almost identically distributed, a negative trend can be seen when the number of parameters is growing. It is possible that this effect occurs due to the utilised training strategy favouring compact architectures.

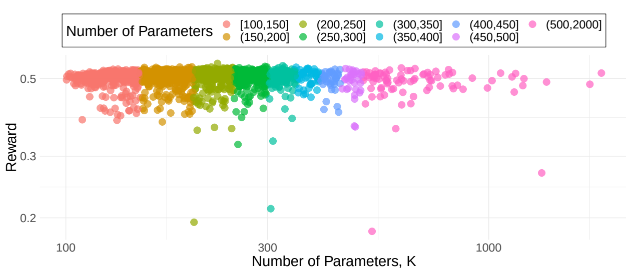


Figure 1. Reward as a function of the size of the architectures.

### Templates

We remind the reader that during the search process the controller samples 3 template structures that can be applied at any of 7 blocks 1−4 times recursively. Each template consists of two individual operations and one aggregation operation. In total, there are 6 unique operations and 2 unique aggregation operations, which leads to

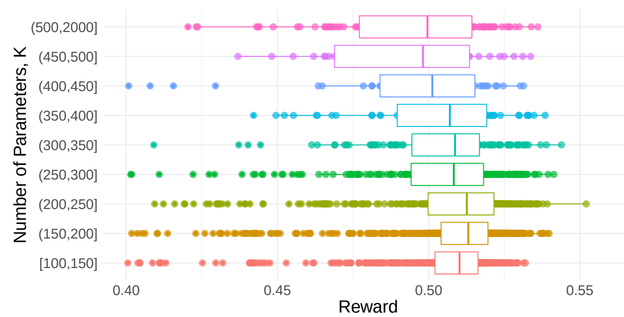


Figure 2. Distribution of rewards attained by architectures with varying size. For compactness of the plot, we only visualise rewards greater than or equal to 0.40.

$C^2_{6+1} = \frac{7!}{5!2!} = 42$ unique templates taking into account the symmetry in the order of individual operations.

We consider a particular template to be sampled during the search process, if its structure was sampled and it was chosen by the controller at least once. We visualise the distribution of rewards for each of 42 templates sampled during the search process in Fig. 3 - note that several templates can share the same reward as they might belong to a single architecture. Overall, around half of all the templates steadily achieve rewards higher than 0.5.

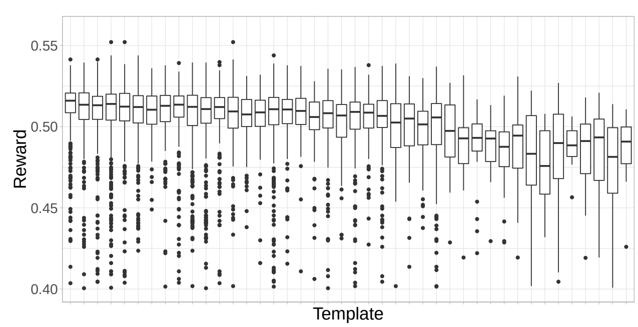


Figure 3. Distribution of rewards for each of 42 unique templates. For compactness of the plot, we only visualise rewards greater than or equal to 0.40.

We further depict each of top-5 templates with highest rewards in Fig. 4. Interestingly, all top-performing templates rely on separable $5 \times 5$ convolution, and some of them differ only in the aggregation operation used (*e.g*. *Template 0* and *Template 4*, or *Template 1* and *Template 3*).

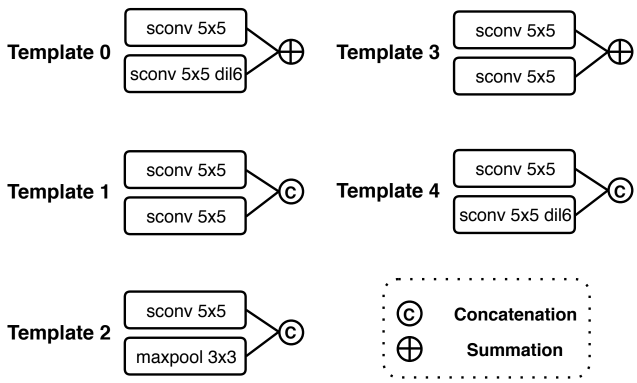


Figure 4. Top-5 templates with highest average reward.

## 2. Architecture Characteristics

We visualise another of discovered architectures - *arch1* - in Fig. 5 (please refer to the main text for the visualisation of *arch0*). In contrast to *arch0*, this architecture fully relies on the summation as its aggregation operation. Notably, *arch1* tends to duplicate the templates more often, which, thanks to the light-weight layers in each template, does not lead to a significant growth in the number of parameters.

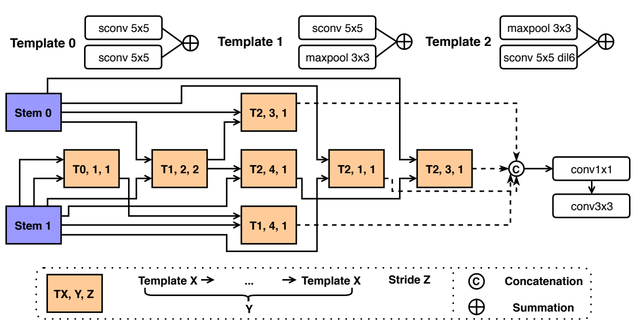


Figure 5. Depiction of *arch1*.