

A ANALYSIS OF OUR METHOD

The influence of polynomial degree on the pruning process is illustrated in Figure 3, which shows the impact of different polynomial degrees on the accuracy of the VGG-16 model on the CIFAR-10 dataset and the corresponding remaining Floating Point Operations Per second (FLOPs). As the polynomial degree increases, we observe a gradual reduction in the remaining FLOPs, indicating more aggressive pruning. While this is desirable in terms of computational efficiency, it comes at the cost of model accuracy, which diminishes as the polynomial degree grows. This trend demonstrates that higher-degree polynomials cause over-pruning and notable accuracy degradation. Based on this finding, ACSP fixes the knee-finding polynomial to degree 2, which (as shown by our experiments) achieves aggressive pruning with minimal accuracy loss. We found degree 2 to consistently provide the best balance between FLOP reduction and accuracy retention across architectures and datasets, so we adopt it as the default. This standardized, automatic approach eliminates the need for a trial-and-error process in determining the pruning volume, ensuring that ACSP reliably and efficiently finds the optimal subset size and composition in each layer.

Table 3 illustrates a comparison across various datasets and architectures to explain the rationale behind choosing a polynomial degree of 2 within the knee-finding algorithm. This comparison is solely provided to illustrate how different polynomial degrees influence the balance between model accuracy and pruning volume, guiding our choice of degree 2 to achieve this optimal balance. It is important to note that the knee-point polynomial degree is not a component of the method itself but rather an external decision that supports ACSP’s effectiveness. The table further includes two selection methods: the “Regular” method, which selects the k medoids identified at the end of clustering, and the “Weighted” method, which selects the k components with the highest weights, one from each cluster.

Method. The weighted method consistently outperforms the regular method across all combinations of datasets, architectures, and polynomial degrees. This highlights the significance of incorporating network weights in the component selection process.

Polynomial degree. As the polynomial degree increases, the number of remaining FLOPs (%) decreases, reflecting more aggressive pruning. The polynomial degree influences the knee value the Kneedle algorithm returns: higher degrees result in smaller knee values, leading to fewer components remaining after pruning. However, higher polynomial degrees also result in greater compromises in the model’s accuracy after pruning. As evidenced by Table 3, and aligning with our discussion in Section 3.4.1, a polynomial degree of 2 yields the most pruning possible without hurting – and sometimes even improving – the model’s accuracy. Thus, ACSP uses degree 2 as the default setting for knee detection.

Dataset	Model	Method	Base		Polynum 2		Polynum 3		Polynum 4		Polynum 5	
			Accuracy	FLOPs	Accuracy	FLOPs	Accuracy	FLOPs	Accuracy	FLOPs	Accuracy	FLOPs
CIFAR-10	MobileNet-V2	Regular	94.48	-	94.51	51.81	90.12	39.62	87.87	30.80	86.34	24.07
		Weighted			94.98		90.71		88.82		87.95	
	VGG-16	Regular	93.55	-	93.55	38.50	89.44	29.28	86.98	22.89	85.33	17.76
		Weighted			93.92		90.01		87.83		86.97	
	ResNet-56	Regular	93.69	-	93.41	46.51	89.28	35.40	86.93	27.66	85.03	23.50
		Weighted			93.82		90.21		87.97		86.89	
CIFAR-100	VGG-16	Regular	73.70	-	73.95	49.75	70.69	36.14	68.67	30.38	67.57	24.10
		Weighted			74.31		71.08		69.68		68.79	
	VGG-19	Regular	73.38	-	73.36	47.39	69.95	38.91	68.30	33.63	67.16	27.34
		Weighted			73.90		70.89		68.83		68.00	
	DenseNet-40	Regular	74.30	-	73.51	52.35	70.23	39.62	69.13	30.84	67.91	23.98
		Weighted			73.94		70.85		69.34		68.75	
ImageNet-1K	MobileNet-V2	Regular	71.90	-	71.83	64.51	68.46	49.33	66.51	38.63	65.54	32.09
		Weighted			71.99		69.19		67.34		66.41	
	ResNet-50	Regular	76.32	-	76.40	44.44	73.29	36.77	71.36	28.51	69.61	24.56
		Weighted			76.98		73.98		72.02		71.15	

Table 3: Performance comparison of accuracy and remaining FLOPs (%) across different polynomial approximation levels, using regular (medoid selection) and weighted (highest-weight selection) component selection methods.

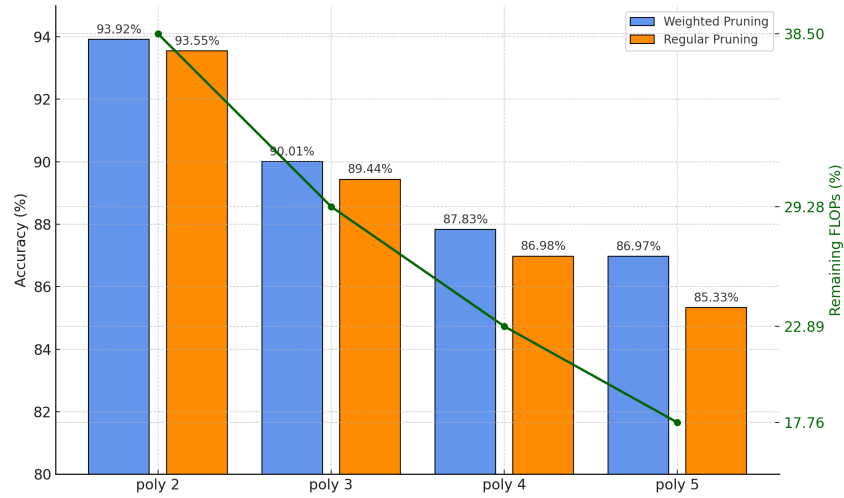


Figure 3: The effect of polynomial degree on the VGG-16 model on CIFAR-10 dataset. Higher polynomial degrees lead to fewer remaining FLOPs (%) but with a greater loss in accuracy.