Fair comparison and ablation study. The results on CIFAR10 were listed in Table R1. **(1)** We increased the width of the backbone networks for SE and CBAM (denoted by "Wide") so that their parameters and FLOPs match those of AutoLA. It reveals that HOGA searched by AutoLA (k=4)) still outperforms SE and CBAM by a large margin. **(2)** We further customized SE and CBAM using the group split operation (denoted by "HOG"), resulting in a specific instantiation of HOGA (i.e., k=4). From Table R1, the performances of customized SE and CBAM (order=4) are improved significantly compared to the vanilla ones, attributing to the superiority of the proposed HOGA scheme. **(3)** The HOGA searched by AutoLA outperforms its randomly search counterparts (denoted by "Rand"). The results by random search exceed SE and CBAM. These results validate the superiority of the proposed AutoLA methods, including the new concept HOGA and the architecture search algorithm. We also presented the ablation study on the number of group split (i.e., the hyper-parameter k). Less groups mean lower order of attentions in HOGA, leading to inferior performance. We tested the generalization ability of HOGA searched on ResNet56 (denoted by "AutoLA_56") on a new backbone, i.e., ResNet20. Although the results are inferior to the ones searched directly on ResNet20, this HOGA still outperforms SE and CBAM.

Table R1: Experiments with fair settings of parameters and FLOPGs and ablation study results on CIFAR10.

	Acc (%)	Param (M)	FLOPS (G)		Acc (%)	Param (M)	FLOPS (G)
ResNet20 + SE	92.30	0.29	0.04	ResNet32 + SE	93.16	0.49	0.07
ResNet20 + CBAM	92.81	0.3	0.04	ResNet32 + CBAM	93.47	0.49	0.07
ResNet20_Wide + SE	93.16	0.36	0.05	ResNet32_Wide_SE	94.08	0.62	0.09
ResNet20_Wide + CBAM	93.13	0.37	0.05	ResNet32_Wide_CBAM	93.92	0.63	0.09
ResNet20 + HOG_SE (k=4)	92.87	0.32	0.05	ResNet32 + HOG_SE (k=4)	93.62	0.54	0.09
ResNet20 + HOG_CBAM (k=4)	93.07	0.35	0.05	ResNet32 + HOG_CBAM (k=4)	93.72	0.56	0.09
ResNet20 + AutoLA (k=2)	93.18	0.33	0.05	ResNet32 + AutoLA (k=2)	93.81	0.49	0.09
ResNet20 + AutoLA (k=4)	93.38	0.34	0.05	ResNet32 + AutoLA (k=4)	94.33	0.52	0.09
ResNet20 + AutoLA_56 (k=4)	93.31	0.35	0.05	ResNet32 + AutoLA_56 (k=4)	94.18	0.57	0.09
ResNet20 + Rand_HOGA (k=4)	93.28	0.35	0.05	ResNet32 + Rand_HOGA (k=4)	94.15	0.59	0.09

Comparison on different backbones. We presented the results for different backbones suggested by reviewer #4 and #5, including ResNeXt and the one searched by PNAS (Progressive Neural Architecture Search, ECCV2018) on CIFAR10 in Table R2. In Table 3 (in the submission), we reported the results on ImageNet with WiderResNet (denoted by WResNet18). Both tables show that the HOGA searched by AutoLA outperforms other attention modules on CIFAR10 and ImageNet when deployed on highly variable architectures including ResNet, ResNeXt, PNAS, and WiderResNet, indicating the consistent superiority of the HOGA searched by AutoLA over previous attention methods.

Table R2: Comparison of different attention modules on ResNeXt and PNAS.

	Acc (%)	Param (M)	FLOPS (G)		Acc (%)	Param (M)	FLOPS (G)
ResNext	94.76	1.71	0.28	PNAS	93.34	0.72	0.08
ResNext_SE	95.22	2.23	0.30	PNAS_SE	93.71	0.75	0.08
ResNext_CBAM	95.31	2.24	0.31	PNAS_CBAM	93.79	0.76	0.08
ResNext_AutoLA	95.67	2.35	0.41	PNAS_AutoLA	94.10	0.91	0.11

Comparison with other attention modules on larger backbones. As suggested by reviewer #5, we further compared the HOGA searched by AutoLA with other attention modules including **(1)** GCNet (Non-local Networks Meet Squeeze-Excitation Networks and Beyond, ICCV 2019); **(2)** AugAtt (Attention Augmented Convolutional Networks, ICCV 2019); and **(3)** GENet (Exploiting Feature Context in Convolutional Neural Networks, NeurIPS 2018). The results were listed in Table R3. With comparable or even less parameters and FLOPs, the proposed AutoLA outperforms other attention methods by a substantial margin. We also compared AutoLA with SE and CBAM on a larger backbone (e.g., ResNet101). The results in Table R3 suggest that AutoLA still outperforms other attention modules.

Table R3: Experiments results by different attentions on ImageNet.

	Top-1 Error (%)	Param (M)	FLOPS (G)		Top-1 Error (%)	Param (M)	FLOPS (G)
ResNet50 + GENet	22.00	31.20	3.87	ResNet101	23.38	44.55	7.57
ResNet50 + AugAtt	22.30	24.30	7.90	ResNet101 + SE	22.35	49.33	7.58
ResNet50 + GCNet	22.30	28.08	3.87	ResNet101 + CBAM	21.51	49.33	7.58
ResNet50 + AutoLA	21.77	29.39	4.73	ResNet101 + AutoLA	20.95	51.81	8.94

Large object detection model. We evaluated the performance of AutoLA and CBAM for object detection on the COCO dataset by equipping them with a powerful detection framework, i.e., Faster RCNN, as suggested by reviewers #2 and #3. We used ResNet50 as the backbone. The mAP values are ResNet50(29.1), CBAM(30.8), AutoLA(33.7).

Other details. (1) The search space of HOGA contains 10^8 potential structures although it can be further enlarged by including other types of attention operations such as non-local attention. Note that, since non-local is computationally heavy, an efficient implementation or approximation is necessary. **(2)** We termed the concept of high-order group attention as a series of cascaded attention operations. In this sense, the proposed HOGA has a higher order than CBAM. The effectiveness of designing or searching high-order attentions is also validated by the results of customized SE and CBAM (order=4)in Table R1, attributing to the strong representation capacity resulted from combining diverse nonlinear transformations (i.e.., different types of attention operations) among channel groups.