AutomaticationAutomaticationConstruction <th>GUILIN UNIT 1956</th> <th>14/18- 150 00 00 00 00 00 00 00 00 00 00 00 00 0</th> <th></th> <th>G</th> <th>l0</th> <th>ba</th> <th></th> <th>7il</th> <th>te</th> <th>r]</th> <th>Pri</th> <th>unin</th>	GUILIN UNIT 1956	14/18- 150 00 00 00 00 00 00 00 00 00 00 00 00 0		G	l 0	ba		7il	te	r]	Pri	unin
Dur contributions can be summarized as follows:(1)Our work provides a fresh perspective 1 improve efficiency and precision of UAV tracking by developing DL-based trackers with filter pruning method, which has not been well explored before. (2)We proposed a method of global filter pruning with self-attention for real-time UAV tracking, with which the propose B-SimaFC++ can globally compress the baseline SiamFC++ to about 60% of its original model size, meanwhile maintaining and even significantly boosting precision simultaneously.(3)We evaluate our PS-SiamFC++ to no four public UAV benchmarks, i.e., UAVI34@10fps, DTB70, UAVDT and VisDrone2018. Experimental results show that the proposed PS-SiamFC++ a chieves state-of-the-art performance.Table 1: Comparison of average precision and speed (FPS) between PS-SiamFC++ are hand-crafted based trackers on the four benchmarks. The reported FPSs are evaluated on single CPU. Red, blue and green respectively show the first, second and third places.Table 2: Precision and speed (FPS) (evaluated on a single GPU) comparison between PS SiamFC++ and deep-based trackers on UAVDT 115].Table 2: Precision and speed (FPS) (evaluated on a single GPU) comparison between PS SiamFC++ and deep-based trackers on UAVDT 115].Table 3: Comparison of the proposed PS-SiamFC++ tracker with two baseline trackers terms of model size (Parma Num	RSITY OF T					/(22					
Our contributions can be summarized as follows:(1)/Our work provides a fresh perspective 1 improve efficiency and precision of UAV tracking by developing DL-based trackers with filter pruning method, which has not been well explored before. (2)/We proposed a method of global filter pruning with self-attention for real-time UAV tracking, with which the propose PS-SimaPC++ can globally compress the baseline SimaPC++ to about G0% of its original model size, meanwhile maintaining and even significantly boosting precision simultaneously:(3)/We evaluate our PS-SimaPC++ on four public UAV benchmarks, i.e., UAV1236(10fps, DTB70, UAVDT and VisDrone2018. Experimental results show that the propose PS-SimaPC++ 1 achieves state-of-the-art performance. Table 1: Comparision of average precision and speed (FPS) between PS-SimaPC++ at and-crafted based trackers on the four benchmarks. The reported FPSs are evaluated on single CPU. Red, blue and green respectively show the first, second and third places. <u>Experimental wave to be added to</u>				Ir	ntr	00	luc	cti	011			
<section-header> Precision of the proposed PS-SiamFC++ tracker with two baseline trackers terms of model size (Parameters), precision (DP), and tracking speed (FPS) on the PC CPI CPI CPI StamFC++ tracker with two baseline trackers terms of model size (Parameters), precision on DTB70 of PS-SiamFC++ varies with the global state of the proposed PS-SiamFC++ tracker with two baseline trackers terms of model size (Parameters), precision on DTB70 of PS-SiamFC++ varies with the global state of the precision on DTB70 of PS-SiamFC++ varies with the global state of the precision on DTB70 of PS-SiamFC++ varies with the global state of the precision on DTB70 of PS-SiamFC++ varies with the global state of the precision on DTB70 of PS-SiamFC++ varies with the global state of the precision on DTB70 of PS-SiamFC++ varies with the global state of the precision of the precision on DTB70 of PS-SiamFC++ varies with the global state of the precision on DTB70 of PS-SiamFC++ varies with the global state of the precision on DTB70 of PS-SiamFC++ varies with the global state of the precision on DTB70 of PS-SiamFC++ varies with the global state of the precision on DTB70 of PS-SiamFC++ varies with the global state of the precision on DTB70 of PS-SiamFC++ varies with the global state of the precision on DTB70 of PS-SiamFC++ varies with the global state of the precision on DTB70 of PS-SiamFC++ varies with the global statement of the precision on DTB70 of PS-SiamFC++ varies with the global statement of the precision on DTB70 of PS-SiamFC++ varies with the global statement of the the precision on DTB70 of PS-SiamFC++ varies with the global statement of the the precision on DTB70 of PS-SiamFC++ varies with the global statement of the the precision on DTB70 of PS-SiamFC++ varies with the global statement of the the the precision on DTB70 of PS-SiamFC++ varies with the global statement of the the the the the the the the the the</section-header>	Our contr improve e filter prun global filt PS-SimaF model siz simultane UAV1230 proposed	ibution efficient ing me er prun C++ caously.(3)@ 10fps.PS-Sian	s can be cy and j thod, w ing with an globa while i 3)We ev , DTB7 mFC++	e summ precisio /hich ha h self-a ally con maintai valuate '0, UA - achiev	narized on of U as not attention mpress ning a our PS VDT a ves stat	l as fol JAV tra been w on for 1 the ba nd eve S-Siam nd Vis te-of-th	lows:(1 acking l vell exp ceal-tim aseline s n signif FC++ o Drone2 he-art p)Our w by deve lored b e UAV SiamF(SiamF(Cicantly on four 018 . E erform	vork pro eloping efore. (trackin C++ to a boostin public Experim ance.	ovides DL-ba 2)We 1g, with about of ng pre- UAV f nental f	a fresh j ased trac propose h which 60% of i cision benchma results si	perspective to ckers with d a method of the proposed its original arks, i.e., how that the
Table 1: Comparision of average precision and speed (FPS) between PS-SiamFC++ ar hand-crafted based trackers on the four benchmarks. The reported FPSs are evaluated on single CPU. Red, blue and green respectively show the first, second and third places. $\frac{1}{10000000000000000000000000000000000$				E	Xſ)e1	•111	101	nts			
Table 2: Precision and speed (FPS) (evaluated on a single GPU) comparison between P: SiamFC++ and deep-based trackers on UAVDT [15].Table 2: Precision and speed (FPS) (evaluated on a single GPU) comparison between P: SiamFC++ and deep-based trackers on UAVDT [15].Trevision 038 Provision 665PripingleKYS PCV 20SiamGAT CVPR 21LightTrack CVPR 21TransT CVPR 21HET CCV 21SOAT CCV 21AutoMatch PS-SiamFC++ OursPS-SiamFC++ OursTable 3: Comparison of the proposed PS-SiamFC++ tracker with two baseline trackers terms of model size (Parameters), precision (DP), and tracking speed (FPS) on the PC CPUMethodsParameters Puning Self-attentionOut V123@10fpsDTB70 VLV12@10fpsOut VT VisDrone2018 Avg. Precision Avg. Precision <b< td=""><td>Table 1: C hand-crafte single CPU</td><td>Compared base J. Red, KCF TPAMI 15 53.3 655.6 572.1</td><td>ision o d track blue ar fDSST CVPR 16 60.4 203.6 177.6</td><td>of avera ers on nd gree Staple-CA CVPR 17 64.2 67.7 59.4</td><td>age pro the for n response BACF ICCV 17 65.3 57.0 50.1</td><td>ecision ur ben ectivel ECO-HG CVPR 1 68.8 88.9 77.8</td><td>n and s chmark y show C MCCT-H 7 CVPR 18 66.8 66.7 58.4</td><td>peed (s. The the fir STRCF CVPR 1 67.1 29.9 26.3</td><td>FPS) b report st, seco ARCF-H 8 ICCV 19 71.9 36.0 31.2</td><td>etween ed FP ond and C AutoT 9 CVPF 72. 61. 54</td><td>n PS-Si Ss are e d third p Track RACE 20 3DV 2 3 75.7 8 37.5 2 32 9</td><td>amFC++ and valuated on a places. PS-SiamFC++ 1 Ours 79.1 71.3 62.4</br></td></b<>	Table 1: C hand-crafte single CPU	Compared base J. Red, KCF TPAMI 15 53.3 655.6 572.1	ision o d track blue ar fDSST CVPR 16 60.4 203.6 177.6	of avera ers on nd gree Staple-CA CVPR 17 64.2 67.7 59.4	age pro the for n response BACF ICCV 17 65.3 57.0 50.1	ecision ur ben ectivel ECO-HG CVPR 1 68.8 88.9 77.8	n and s chmark y show C MCCT-H 7 CVPR 18 66.8 66.7 58.4	peed (s. The the fir STRCF CVPR 1 67.1 29.9 26.3	FPS) b report st, seco ARCF-H 8 ICCV 19 71.9 36.0 31.2	etween ed FP ond and C AutoT 9 CVPF 72. 61. 54	n PS-Si Ss are e d third p Track RACE 20 3DV 2 3 75.7 8 37.5 2 32 9	amFC++ and valuated on a places.
$\frac{ \mathbf{x} _{CVPR 20} \mathbf{x} _{CVPR 20} \mathbf{x} _{CVPR 20} \mathbf{x} _{CVPR 20} \mathbf{x} _{CVPR 21} \mathbf{x} _{CVPR 21} \mathbf{x} _{CVPR 21} \mathbf{x} _{CV 21} x$	Table 2: P SiamFC++	recision and de	n and sep-bas	peed (1 ed trac	FPS) (kers of	evalua n UAV	ited on DT [15	a sing	le GPU	() com	parison	between PS-
$\frac{Precision}{FPS(GPU)} = \frac{66.5}{7.2} = \frac{72.2}{44.6} = \frac{73.2}{44.5} = \frac{79.8}{30.2} = \frac{76.4}{74.8} = \frac{80.4}{84.8} = \frac{82.6}{42.1} = \frac{65.2}{135.3} = \frac{82.1}{29.4} = \frac{82.1}{50.4} = \frac{80.5}{291.9}$ $Table 3: Comparison of the proposed PS-SiamFC++ tracker with two baseline trackers terms of model size (Parameters), precision (DP), and tracking speed (FPS) on the PC CPI \frac{Methods}{SiamCAR} = \frac{Parameters}{Pruning} = \frac{Puning}{Self-attention} = \frac{VAV123@10fps}{VAV123@10fps} = \frac{DTB70}{DTB70} = \frac{VAVDT}{VisDrone2018} = \frac{Visp}{Visp} = \frac{Visp}{Visp}$	S	SiamR-CNN CVPR 20	D3S CVPR 20	PrDimp18 CVPR 20	KYS ECCV 20	SiamGAT CVPR 21	LightTrank CVPR 21	TransT CVPR 21	HiFT ICCV 21	SOAT ICCV 21	AutoMatch ICCV 21	PS-SiamFC++ Ours
Table 3: Comparison of the proposed PS-SiamFC++ tracker with two baseline trackers terms of model size (Parameters), precision (DP), and tracking speed (FPS) on the PC CPU $\frac{Methods Parameters}{N} \frac{Praining Self-attention UAV123@10fps DTB70 UAVDT VisDrone2018 Avg. Precision Avg. FPS (CPU)}{SiamCAR [18] 8.5M 7 70.9 68.2 72.7 74.6 71.6 79.4 40.7 PS-SiamCAR 5.1M 7 70.9 68.2 72.7 74.6 71.6 79.4 179.3 SiamCAR 5.1M 7 70.9 68.2 72.7 74.6 71.6 79.4 179.3 SiamFC++ [47] 9.7M 72.8 80.5 76.2 72.5 75.5 36.5 PS-SiamFC++ 5.8M 7 74.3 79.9 80.5 81.6 79.1 71.1 75.8 74.1 79.3 PS-SiamFC++ 5.8M 7 74.3 79.9 80.5 81.6 79.1 71.1 71.3 PS-SiamFC++ 5.8M 7 74.3 79.9 80.5 81.6 79.1 71.1 71.1 Table 4: Illustration of how the precision on DTB70 of PS-SiamFC++ varies with the global pruning ratio, with or without the self-attention module. The precisions that have been improved by the self-attention component are marked in bold. The precisions that have been improved by the self-attention component are marked in bold. The precision of precision that have been improved by the self-attention component are marked in bold. The precision of the$	Precision FPS (GPU)	66.5 7.2	72.2 44.6	73.2 48.5	79.8 30.2	76.4 74.8	80.4 84.8	82.6 42.1	65.2 135.3	82.1 29.4	82.1 50.4	80.5 291.9
Table 4: Illustration of how the precision on DTB70 of PS-SiamFC++ varies with the global pruning ratio, with or without the self-attention module. The precisions that have been improved by the self-attention component are marked in bold. $ \frac{\rho \frac{UAV123@10fps}{w/o} \frac{DTB70}{w/} \frac{UAVDT}{visDrone2018}}{\frac{100}{0.1} 70.8 72.2 79.6 79.4 81.4 81.3 79.6 83.1}{0.2 71.6 72.3 80.0 80.1 76.9 77.2 80.2 77.7}{0.3 71.3 72.4 81.0 81.5 83.9 80.3 75.6 79.3}} $	Table 3: C terms of m Methods SiamCAR [18 P-SiamCAR PS-SiamCAR SiamFC++ [47 P-SiamFC++ PS-SiamFC++	Compari odel siz Paramet 3] 8.5M 5.1M 5.1M 5.1M 7] 9.7M 5.8M + 5.8M	son of ze (Para ers Prunir	the pr ameter ng Self-att	oposec s), prec tention U	1 PS-S cision AV123@10 73.7 70.9 71.0 72.8 71.9 74.3	biamFC (DP), a Ofps DTB7 76.6 68.2 73.2 80.5 79.5 79.9	++ trac nd trac 0 UAVD 76.1 72.7 77.1 76.2 78.8 80.5	cker wi cking sp r VisDrone 80.3 74.6 75.8 72.5 79.3 81.6	th two peed (1 2018 A 3 5 3 5 5 3 5 5 5 5 5 5 5 5 5 5 5 5 5	b baselin FPS) on vg. Precision 76.7 71.6 74.1 75.5 77.4 79.1	ne trackers in the PC CPU Avg. FPS (CPU) 40.7 79.4 79.3 36.5 71.3 71.1
	Table 4: Ill pruning rat proved by	ustratio tio, with the self	on of ho h or wi -attenti	$\rho = \frac{UAV}{w/o}$ $\rho = \frac{UAV}{100}$ $\rho = \frac{UAV}{100}$ $\rho = \frac{UAV}{100}$	precision he self nponer 123@10fp w/ 72.2 72.3 72.4	ion on f-atten nt are 1 s DTE w/o 79.6 80.0 81.0 70.5	DTB7(tion mc marked 370 U. w/ w/c 79.4 81.4 80.1 76.9 81.5 83.9 70 0 70 () of PS odule. in bolo AVDT w/ 81.3 77.2 80.3	5-SiamF The pred. d. VisDrone20 w/o w 79.6 83 80.2 77 75.6 79	FC++ v ecision $\overline{018}$ $\frac{1}{7}$	varies w is that h	ith the global ave been im-

66.0

12 2

ng with Self-Attention for Real-Time UAV Tracking

Mengyuan Liu¹

¹ Guilin University of Technology, China

³ Guangxi Key Laboratory of Embedded Technology and Intelligent System, China







Illustration of the Multi-Head Self-Attention layer applied to the backbone output feature f_Z . Note that f_Z is encoded in a pixel-wise manner, i.e., the spatial coordinates of f_Z index the tokens, and the query, key, and value are initially the same.

78.6 77.1 76.6 **78.3** 75.2 **76.9 77.9** 74.9 77.8 **80.7** 76.7 **79.4**

76.4 70.1 74.6 **77.0** 71.8 **74.9**

Yuelong Wang¹ Qiangyu Sun² Shuiwang Li^{1,3}

² Hubei Enshi College, China



SP-SiamFC++ (Ours) RACF

Qualitative evaluation on 4 sequences from UAV123@10fps, DTB70, UAVDT and VisDrone2018 (i.e. truck1, BMX5, S0309 and uav0000294_00000_s), respectively. The results of different methods have been shown with different colors.

Conclusions

In this work, we present a method of global filter pruning with self-attention for realtime UAV tracking and achieve state-of-the-art performance on four public UAV tracking benchmarks. When using the proposed method to improve UAV tracking efficiency, experimental results reveal that the proposed method is quite effective at maintaining and even improving precision. Surprisingly, the proposed PS-SiamFC++ not only outperforms the baseline SiamFC++ in terms of efficiency (PS-SiamFC++ can run at and more than 62 FPS on a single CPU of a mini PC, i.e., Intel NUC), but it also outperforms the baseline in terms of precision on UAVDT and VisDrone2018, well combating the adverse effects of filter pruning.

