

[SUPPLEMENTARY MATERIAL]

FASTER AND SMARTER AUTOAUGMENT: AUGMENTATION POLICY SEARCH BASED ON DYNAMIC DATA-CLUSTERING

Anonymous authors

Paper under double-blind review

A OPERATION DISTRIBUTION

Figure 1 is a heatmap showing the distribution of 15 types of augmentation operations, each found in $D_{high,i}$ in FSAA, and data fold in the FAA. *CutMix* is included in all policies’ last operation, so it is omitted from the figure. The closer to the blue, the more operations are centralized, and the closer to white, the less there are operations. As shown in the Figure 1(a), the FAA operation are centralized to the same operation (*Cutout*) for all data folds. On the other hand, FSAA shows the different results about the distribution of operations for each $D_{high,i}$, as shown in Figure 1(b). The representative operations for each $D_{high,i}$ are all searched differently (e.g., *AutoContrast*, *Cutout*, *Sharpness*, *ShearY*, *Brightness*, *TranslateY*).

Figure 2 is the example when the images are converted based on each dataset’s policies shown in Figure 1. As shown in Figure 2(a), since most operations of FAA consist of *Cutout*, the result images to which *Cutout* operation is applied are commonly sampled. On the other hand, in FSAA, each dataset consists of varied policy pattern, more diverse types of images are transformed than FAA.

	D ₁	D ₂	D ₃	D ₄	D ₅			D _{high,1}	D _{high,2}	D _{high,3}	D _{high,4}	D _{high,5}	D _{high,6}	D _{high,7}
Cutout	12%	14%	13%	13%	15%	16-18%	Cutout	9%	16%	8%	7%	14%	6%	5%
AutoContrast	11%	11%	10%	11%	12%		ShearY	3%	11%	9%	15%	8%	8%	8%
TranslateX	10%	8%	9%	8%	7%	13-15%	Sharpness	7%	14%	15%	9%	6%	8%	3%
Equalize	10%	8%	7%	6%	9%		Brightness	6%	8%	5%	9%	6%	16%	9%
Posterize	10%	9%	9%	5%	2%		TranslateY	10%	6%	7%	4%	4%	8%	17%
Color	7%	6%	8%	8%	4%	10-12%	Color	11%	5%	8%	6%	9%	9%	7%
ShearY	2%	6%	4%	7%	14%		AutoContrast	16%	2%	8%	10%	6%	4%	5%
Brightness	5%	4%	6%	8%	8%	7-9%	Contrast	5%	7%	5%	4%	7%	8%	9%
ShearX	6%	7%	8%	4%	5%		Posterize	6%	3%	5%	6%	7%	7%	9%
Sharpness	4%	7%	5%	5%	8%		TranslateX	4%	11%	8%	9%	5%	1%	4%
TranslateY	4%	5%	7%	6%	3%	4-6%	ShearX	3%	2%	4%	7%	4%	7%	8%
Contrast	5%	4%	3%	6%	4%		Solarize	4%	5%	4%	5%	8%	3%	5%
Invert	2%	5%	6%	4%	4%	1-3%	Invert	8%	2%	7%	4%	4%	8%	1%
Solarize	3%	3%	2%	7%	4%		Rotate	4%	3%	6%	2%	7%	5%	4%
Rotate	9%	3%	3%	2%	1%		Equalize	4%	5%	1%	3%	6%	2%	6%

(a) FAA

(b) FSAA (ours)

Figure 1: Heatmap of operation distribution found in FAA and FSAA on Wide-ResNet-40×2 with CIFAR-10

B ROBUSTNESS ON CIFAR-10

Figure 3 shows the analysis for attacked image samples about FGSM, center occlusion, and boundary occlusion on the CIFAR-10 test set. In Figure 3(a), as the epsilon value increase, both CM, FAA, and FSAA show robust top-1 accuracy compared to the base. However, from epsilon 0.05 and 0.1, the difference between FSAA and CM, FAA increases, respectively. From epsilon 0.25, CM and FAA show an accuracy difference of within 3% from the base. On the other hand, FSAA performed a top-1 accuracy higher than FAA until 0.45. Figures 3(b) and (c) show robustness in the case of occlusion in the center and boundary, respectively. FSAA shows a similar accuracy trend compared

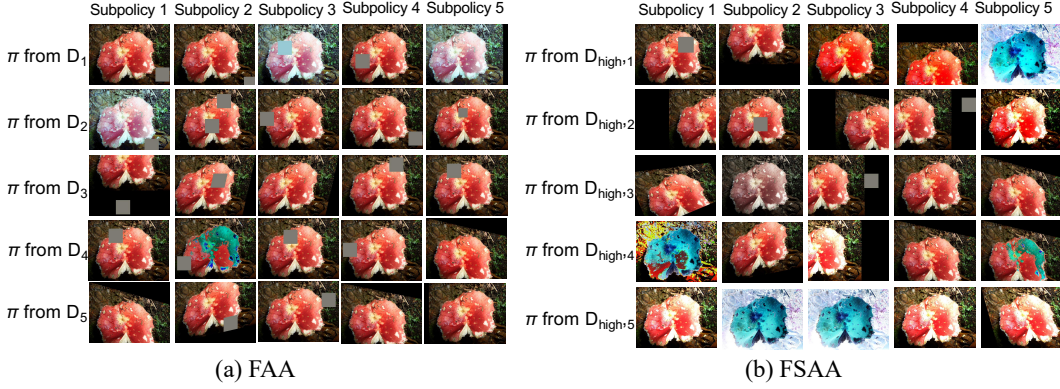


Figure 2: Example augmented images which are randomly sampled from each dataset in Figure 1 (a) FAA and (b) FSAA

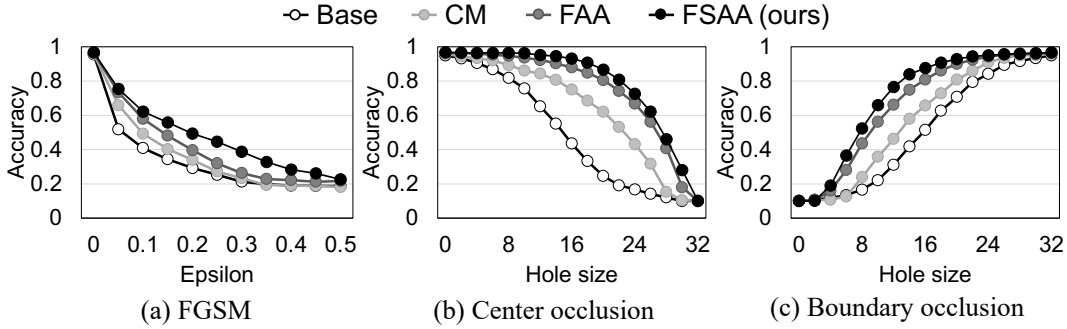


Figure 3: Robustness experiments on Wide-ResNet 40 \times 2 with CIFAR-10 test set (a) FGSM (b) Center occlusion (c) Boundary occlusion

to the FAA, but FSAA has higher accuracy than the FAA in both center occlusion and boundary occlusion.

C TOP-1 ERROR RATE OF FSAA WITHOUT CUTMIX

To evaluate the performance of the augmentation method strategy by the data split stage and policy search stage, excluding the effect of CutMix, the policy exploration is conducted with only 15 operations excluding CutMix. Table 1 shows the top-1 error rate of FAA, FSAA, and FSAA without CutMix on CIFAR-10, CIFAR-100, and SVHN. As shown in the table, all results of FSAA without CutMix show a higher error rate than FSAA. However, compared with the FAA and DADA, FSAA without CM shows a lower error rate. This means that although it is orthogonal to FSAA like CutMix, higher accuracy can be expected when a high-performance data augmentation operation is added as an operation of FSAA.

D CHARACTERISTICS OF CLUSTERED DATA

As data diverges by the data split stage, we check the class distribution of each clustered data to evaluate the data clustering effect. Figure 4 shows a heatmap representing the class distribution for each clustered data using the CIFAR-10 dataset in the Wide-ResNet-40 \times 2 model. In most of the clustered data, the proportion of a specific class is highly concentrated, and in particular, the class 9 images occupy a high proportion of all clustered data. On the other hand, class 2 and 3 show a low ratio for all clustered data, and these singularities are planned to analyze, such as feature map analysis or image shape as future work.

Table 1: Top-1 error rate (%) with FAA, DADA, FSAA, and FSAA without CutMix on CIFAR-10, CIFAR-100, and SVHN.

	CIFAR-10				CIFAR-100				SVHN			
	FAA	DADA	FSAA w/o CM (ours)	FSAA (ours)	FAA	DADA	FSAA w/o CM (ours)	FSAA (ours)	FAA	DADA	FSAA w/o CM (ours)	FSAA (ours)
WRN (40×2)	3.56	3.60	3.42	3.43	21.70	20.90	21.03	20.66	1.44	–	1.50	1.46
WRN (28×10)	2.71	2.70	2.65	2.44	17.24	17.50	16.52	16.12	1.20	1.20	1.23	1.11
SS (2×32d)	2.66	2.70	2.66	2.63	18.58	–	17.85	17.37	1.33	–	1.38	1.32
SS (2×96d)	2.59	2.00	1.99	1.86	14.98	15.30	15.06	15.02	1.14	1.10	1.14	1.11
SS (2×112d)	1.98	2.00	1.92	1.89	16.05	–	15.47	15.11	–	–	–	–
RXN	8.87	–	9.07	8.88	34.65	–	32.26	30.59	2.75	–	2.75	2.71

	D _{high,1}	D _{high,2}	D _{high,3}	D _{high,4}	D _{high,5}	D _{high,6}	D _{high,7}	
Class 0	10.3%	13.5%	9.3%	10.8%	10.5%	11.9%	7.5%	13-15%
Class 1	12.8%	12.1%	14.6%	9.5%	11.0%	9.6%	13.0%	10-12%
Class 2	5.1%	8.7%	9.3%	6.9%	7.4%	8.6%	6.7%	7-9%
Class 3	5.2%	5.2%	6.3%	8.3%	6.8%	6.3%	5.9%	
Class 4	8.3%	8.9%	9.5%	12.5%	9.6%	9.2%	9.2%	
Class 5	9.4%	7.3%	7.0%	9.8%	11.2%	12.4%	8.4%	
Class 6	11.3%	10.0%	11.2%	10.5%	11.0%	10.6%	13.0%	4-6%
Class 7	13.8%	10.0%	11.2%	8.6%	10.9%	10.1%	12.6%	
Class 8	11.1%	10.7%	7.4%	11.8%	9.7%	11.0%	13.0%	1-3%
Class 9	12.7%	13.6%	14.2%	11.3%	12.0%	10.4%	10.9%	

Figure 4: Heatmap of dataset class distribution clustered by data split stage of FSAA on Wide-ResNet-40×2 with CIFAR-10.

Table 2: Top-1 error rate(%) with FAA, FSAA, and FSAA without policy search stage

	CIFAR-10			CIFAR-100		
	FAA	FSAA w/o policy search stage (ours)	FSAA (ours)	FAA	FSAA w/o policy search stage (ours)	FSAA (ours)
WRN (40×2)	3.56	3.54	3.43	21.70	21.45	20.66
RXN	8.87	9.52	8.88	34.65	33.95	30.59

Table 3: Policy transferable between CIFAR-10 and CIFAR-100.

	CIFAR-10			CIFAR-100		
	FAA	FSAA with CIFAR-100 policy (ours)	FSAA (ours)	FAA	FSAA with CIFAR-10 policy (ours)	FSAA (ours)
WRN (40×2)	3.56	3.49	3.43	21.70	20.76	20.66
RXN	8.87	9.06	8.88	34.65	30.73	30.59

E PERFORMANCE COMPARED TO RANDOM SEARCH POLICY

If the DNN model is trained only with a random augmentation policy not tuned without the policy search stage, suboptimal accuracy is obtained. Table 2 shows the top-1 error rate using only the policies explored in the data split stage without the policy search stage compared to FSAA. As shown in the table, the untuned random policy can also show a slight improvement in accuracy, but performance as much as the augmentation method searched by the policy search stage cannot be guaranteed.

F POLICY TRANSFERABLE BETWEEN DATASET

To identify the transferability of the augmentation policy found in similar datasets, the data augmentation methods found in CIFAR-10 and CIFAR-100 are applied to CIFAR-100 and CIFAR-10

dataset training, respectively. As shown in the Table 3, even if training with a policy searched in different datasets can guarantee a lower error rate than FAA. This proves that policies discovered through CIFAR-10 and CIFAR-100, which are composed of similar images, are transferable.