Table 4: Evaluation of Filtered Data Augmentation on CIFAR-100 (20 Classes). Baseline: original samples; None: mean of Wasserstein, TV, MMD at 100% tolerance; Wass, TV, MMD: filtered data at 40%, 60%, 80% tolerance.

Size	Model	Acc	Prec	\mathbf{Rec}	$\mathbf{F1}$	Size	Model	Acc	Prec	\mathbf{Rec}	$\mathbf{F1}$
500	Baseline	0.692	0.695	0.692	0.692	1500	Baseline	0.792	0.789	0.789	0.788
	None	0.840	0.840	0.837	0.837		None	0.882	0.877	0.874	0.874
	Wass-40	0.813	0.814	0.810	0.810		Wass-40	0.867	0.866	0.864	0.864
	Wass-60	0.825	0.824	0.821	0.821		Wass-60	0.872	0.868	0.866	0.865
	Wass-80	0.840	0.837	0.834	0.834		Wass-80	0.875	0.872	0.870	0.869
	TV-40	0.818	0.813	0.809	0.808		TV-40	0.861	0.858	0.856	0.854
	TV-60	0.820	0.817	0.815	0.814		TV-60	0.872	0.869	0.867	0.866
	TV-80	0.834	0.829	0.821	0.821		TV-80	0.873	0.875	0.873	0.873
	MMD-40	0.815	0.812	0.808	0.807		MMD-40	0.867	0.866	0.864	0.864
	MMD-60	0.824	0.820	0.816	0.816		MMD-60	0.872	0.868	0.866	0.865
	$\operatorname{MMD-80}$	0.832	0.829	0.825	0.824		MMD-80	0.875	0.872	0.870	0.869
1000	Baseline	0.763	0.766	0.763	0.762	2000	Baseline	0.815	0.815	0.813	0.812
	None	0.874	0.877	0.872	0.872	2000	None	0.888	0.886	0.884	0.884
	Wass-40	0.851	0.851	0.848	0.847		Wass-40	0.874	0.872	0.870	0.870
	Wass-60	0.863	0.865	0.860	0.861		Wass-60	0.876	0.877	0.876	0.876
	Wass-80	0.866	0.855	0.851	0.851		Wass-80	0.883	0.882	0.880	0.880
	TV-40	0.848	0.843	0.836	0.836		TV-40	0.869	0.869	0.866	0.865
	TV-60	0.860	0.858	0.856	0.855		TV-60	0.884	0.876	0.874	0.874
	TV-80	0.872	0.873	0.870	0.869		TV-80	0.884	0.877	0.875	0.874
	MMD-40	0.851	0.851	0.848	0.847		MMD-40	0.873	0.871	0.870	0.869
	MMD-60	0.863	0.865	0.860	0.861		MMD-60	0.878	0.876	0.875	0.874
	MMD-80	0.866	0.855	0.851	0.851		MMD-80	0.882	0.880	0.879	0.879

Performance metrics (%) on a 20-class CIFAR-100 dataset [1] with varying training sizes (Size). For each original image, 10 images are generated using Stable Diffusion XL, with strength=0.15 (5 images) and strength=0.8 (5 images). Models use pretrained ResNet-18 (ImageNet weights), with conv1 and layer1-3 frozen, and layer4 and classifier trained using Adam optimizer (learning rate 1e-4 for Baseline, 5e-5 for augmented, batch size 32, dropout 0.5). Filtering yields performance nearly identical to unfiltered augmentation, with minimal differences, as CIFAR-100 is well-represented in Stable Diffusion's pretraining, reducing generation anomalies. However, for fine-grained classification tasks, we recommend filtering to enhance robustness.

[1] The CIFAR-100 Dataset