A APPENDIX

A.1 VISUALIZATION OF BIASES INTRODUCED INTO DM

Below is a visualization of how biases are introduced into DM-synthesized images through the distillation process. The process shows that the initial synthetic dataset is initialized with random noise in order not to introduce any biases at the very beginning. Then DM is applied on the randomly initialized synthetic set to iteratively update the learned images. As shown in fig. 5, the learned images also reflect the bias in the original biased dataset.

A.2 EXPERIMENTAL SETUP

To avoid introducing biases during initialization, all synthetic images are initialized with random noise instead of randomly selected real images. For KDE, we fix kernel variance and temperature to be 0.1 across all datasets. All experiments are run on a single 48GB NVIDIA RTX A6000 GPU.

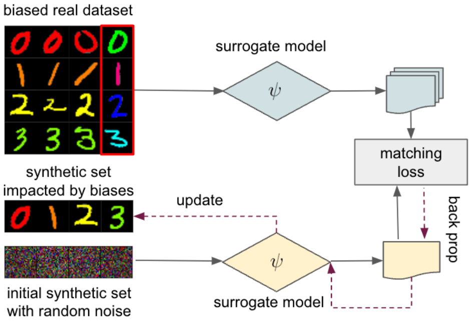

Figure 5: The distillation process of DM on a biased dataset.

A.3 MORE QUALITATIVE RESULTS

First of all, we show the results under IPC 1 in appendix A.3. It can be seen that images synthesized by the vanilla DM is dominated by the biased samples such as red 0s and green 4s. On the contrary, the ones produced by our method is a fusion of both biased samples and unbiased samples, thus mitigating the biases in synthetic datasets. Similar results can also be seen from gradient based method in appendix A.3 where the single image synthesized by DSA is also a fusion of both biased and unbiased samples.

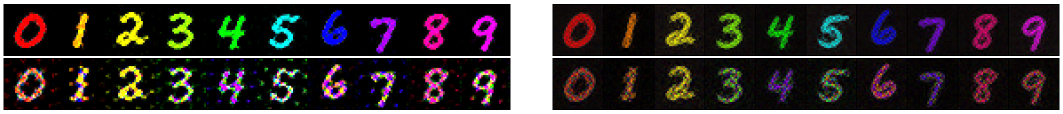

Figure 6: Synthetic images from original DM (top left) vs DM+Ours (bottom left) and original DSA (top right) vs DSA+Ours (bottom right). Results are generated from CMNIST with 5% bias-conflicting samples and IPC 1.

A.4 MATCHING TRAINING TRAJECTORIES

The objective function of MTT can be described as below

$$\mathcal{L} = \|\hat{\theta}_{t+T} - \theta^*_{t+M}\|_2^2 / \|\theta^*_t - \theta^*_{t+M}\|_2^2. \tag{5}$$

Where T is the number of synthetic data training steps and M is the expert model training steps using the whole dataset. As MTT tries to match the training trajectory of expert models and models trained using synthetic data, the real dataset is not needed during matching phase. Therefore, our reweighting scheme doesn't work directly in MTT. However, as shown in table 5, we should still be able to mitigate bias for MTT through fixing the expert models. We leave this to future work.

A.5 VISUALIZATION OF KDE

See Figure 7 for a visualization of KDE applied on a normal distribution. The dotted line is a true normal distribution. The histogram represents the observed data points and the red curve shows the density function estimated using KDE.

A.6 ABLATION STUDY ON CUTTING-OFF SCORE

Here we conduct an ablation study on the cutoff score used when computing the sample weights, the results are shown in Figure 8. It can be seen that choosing a cutoff score that's too large or too small will hinder the de-biasing performance. Also our algorithm is not very sensitive to cutoff scores and the optimal performances can be achived with a wide range of cutoff scores.

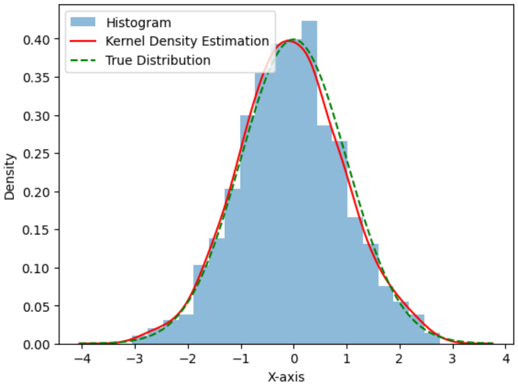

Figure 7: KDE applied on a normal distribution.

A.7 IMPLEMENTATION DETAILS OF APPLYING DE-BIASING METHODS TO SYNTHETIC DATASET

As mentioned before, one natural idea is to directly apply de-biasing methods on distilled synthetic dataset. Here we describe the implementation details. For the results of DM, it's acquired by directly running the vanilla DM on CMNIST dataset. For DM+DFA, we apply the DFA de-biasing methods on the distilled synthetic dataset from the original DM. For DM+SelecMix, we test the two methods proposed in Hwang et al. (2022) which corresponds to directly applying SelecMix methods and appling it on top of LfF (Nam et al., 2020).

A.8 IMPLEMENTATION DETAILS OF APPLYING DE-BIASING METHOD TO SURROGATE MODEL

For applying de-biasing methods to surrogate models, we mainly test two combinations. The first one is DM plus DFA. The reason this can potentially work is that DM tries to match the distribution of real data and synthetic data in the embedding space. DFA is a perfect fit for DM because DFA tries to separate the embeddings into the intrinsic parts and the bias parts. DM can choose to match only the intrinsic parts, thus getting rid of biases. For MTT, since it doesn't rely on real data during matching phase but the expert training trajectories. Therefore, the best way to mitigate bias in MTT synthesized datasets is to debias the expert training trajectories. Thus we choose to apply the most recent SOTA model de-biasing method to acquire the expert trajectories first. Then we have MTT match these de-biased trajectories.

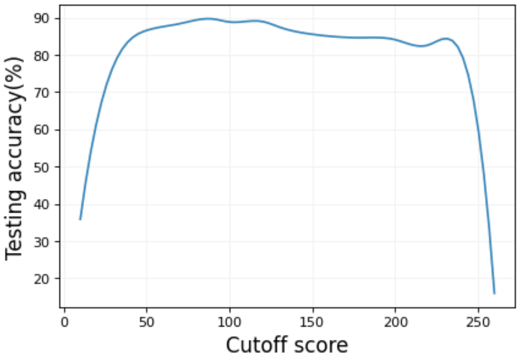

Figure 8: Ablation study on cutoff score. The experiment is conducted on CMNIST with 5% bias-conflicting samples and IPC 10.

A.9 EXAMPLE SYNTHETIC IMAGES

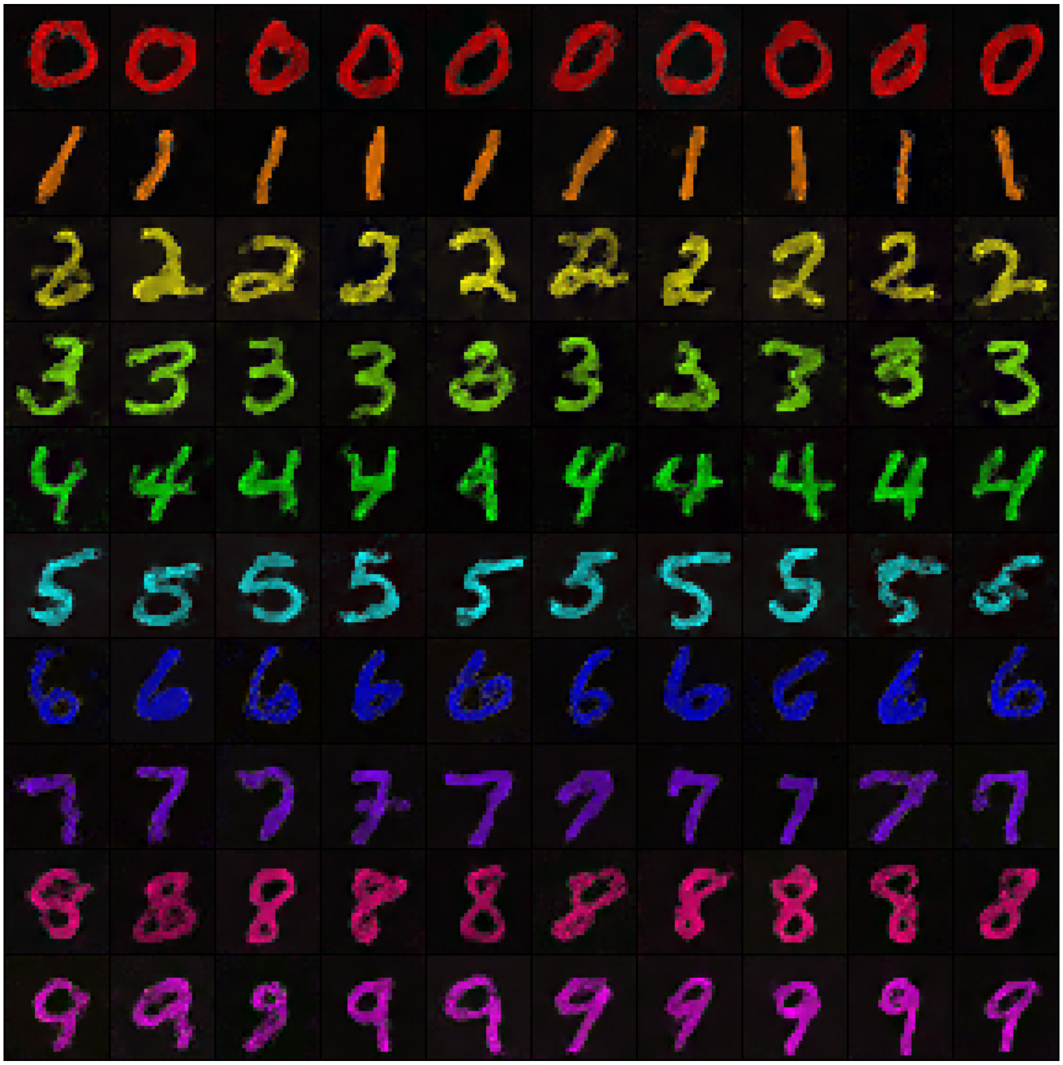

Figure 9: Synthesized dataset by DM on CMNIST with 5% bias-conflicting samples.

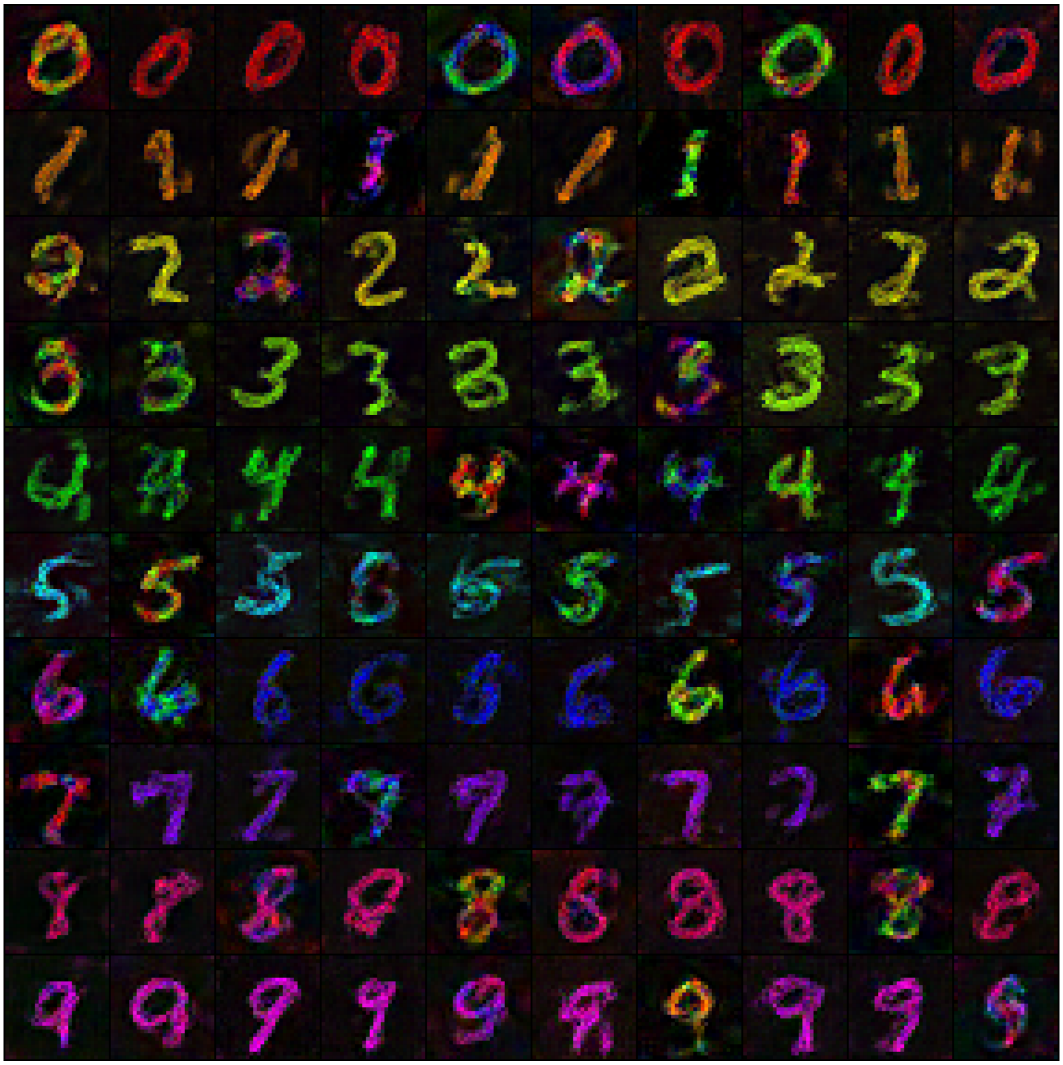

Figure 10: Synthesized dataset by DSA on CMNIST with 5% bias-conflicting samples.

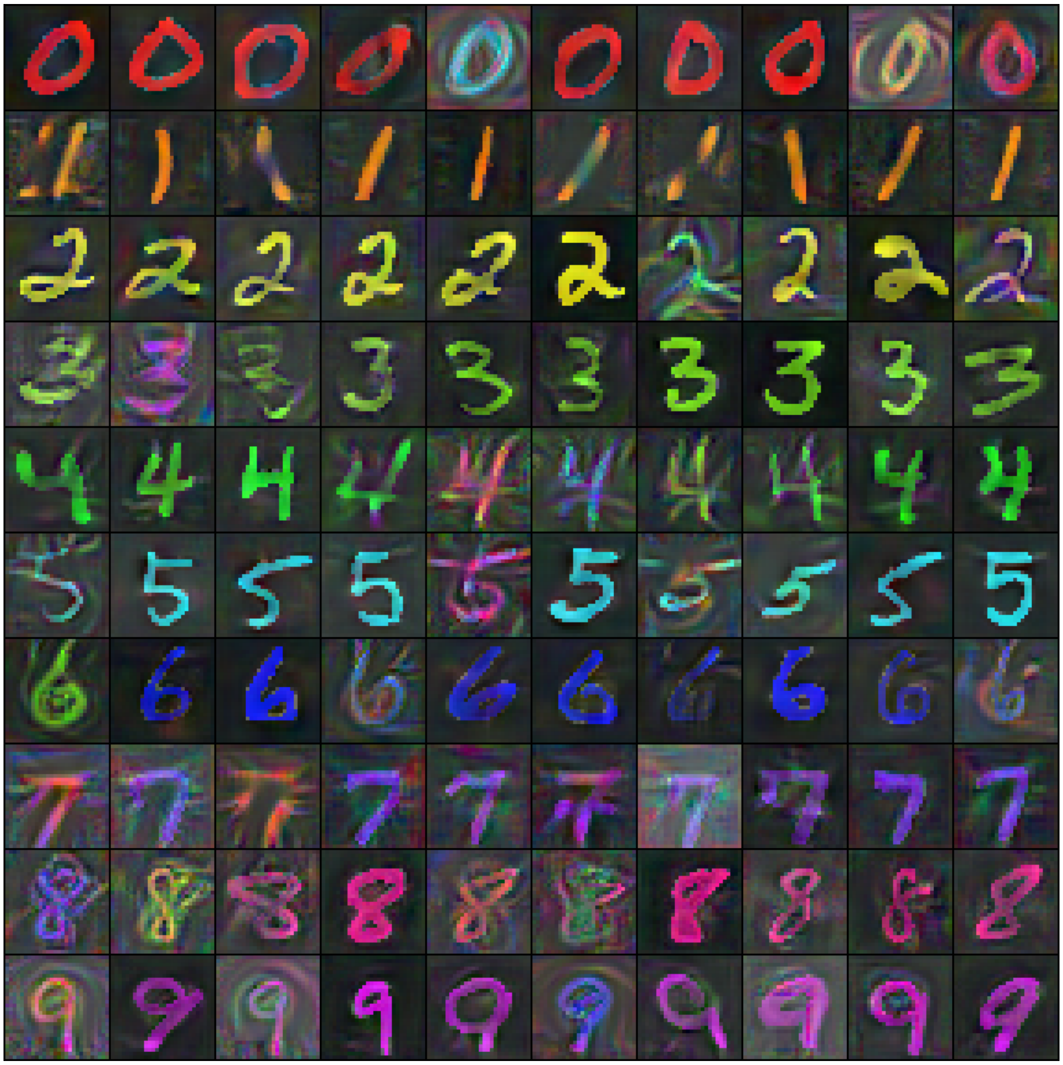

Figure 11: Synthesized dataset by MTT on CMNIST with 5% bias-conflicting samples.

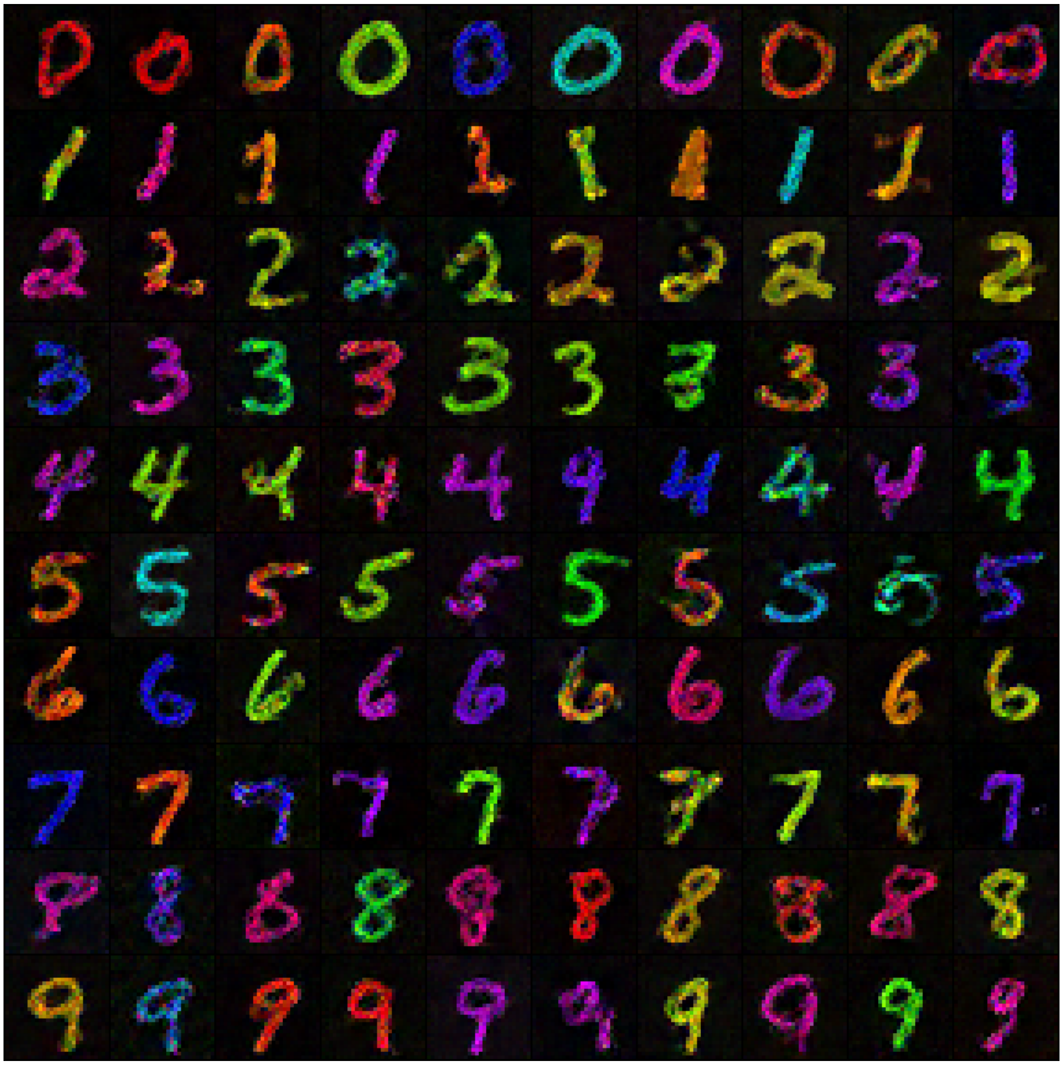

Figure 12: Synthesized dataset by DM+Ours on CMNIST with 5% bias-conflicting samples.

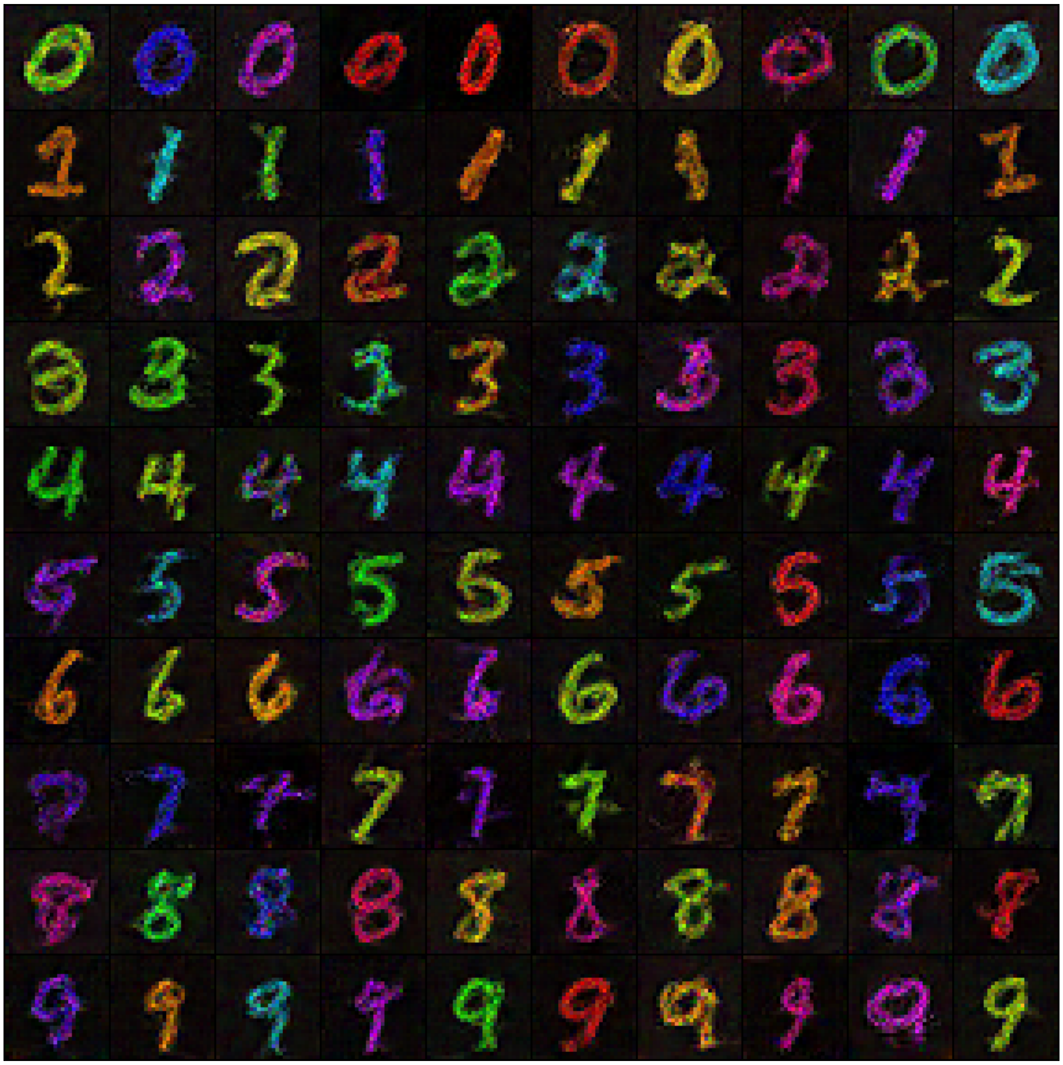

Figure 13: Synthesized dataset by DSA+Ours on CMNIST with 5% bias-conflicting samples.

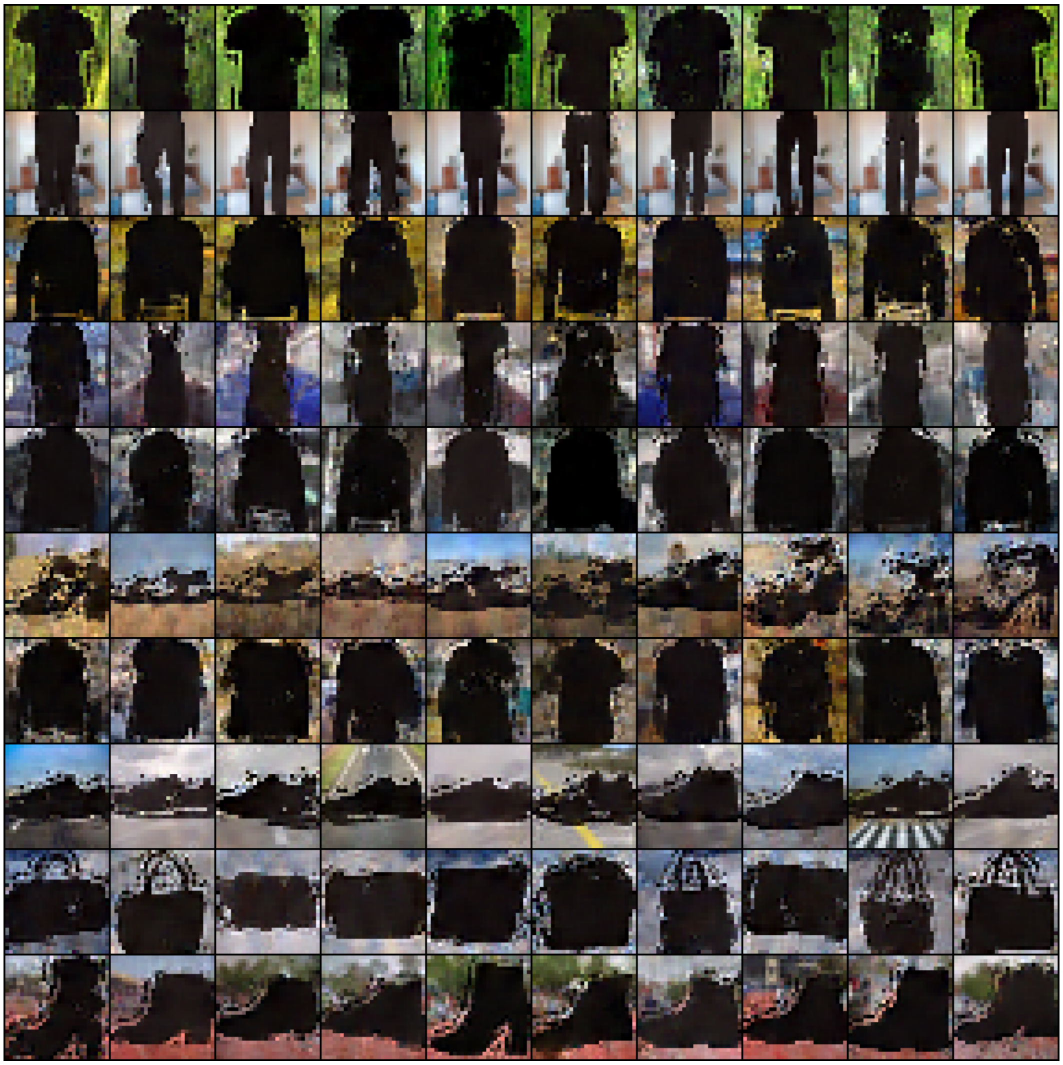

Figure 14: Synthesized dataset by DM on BG FMNIST with 5% bias-conflicting samples.

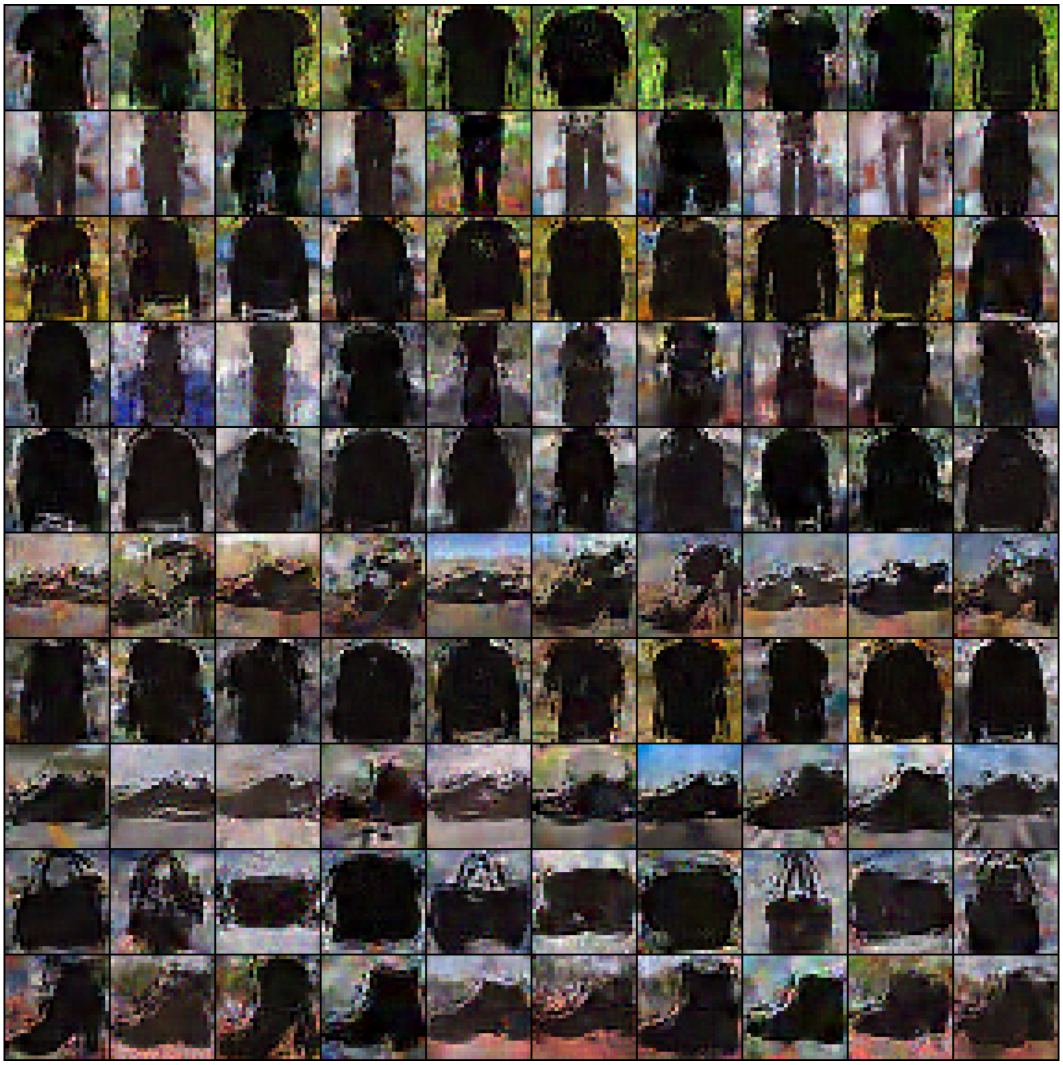

Figure 15: Synthesized dataset by DSA on BG FMNIST with 5% bias-conflicting samples.

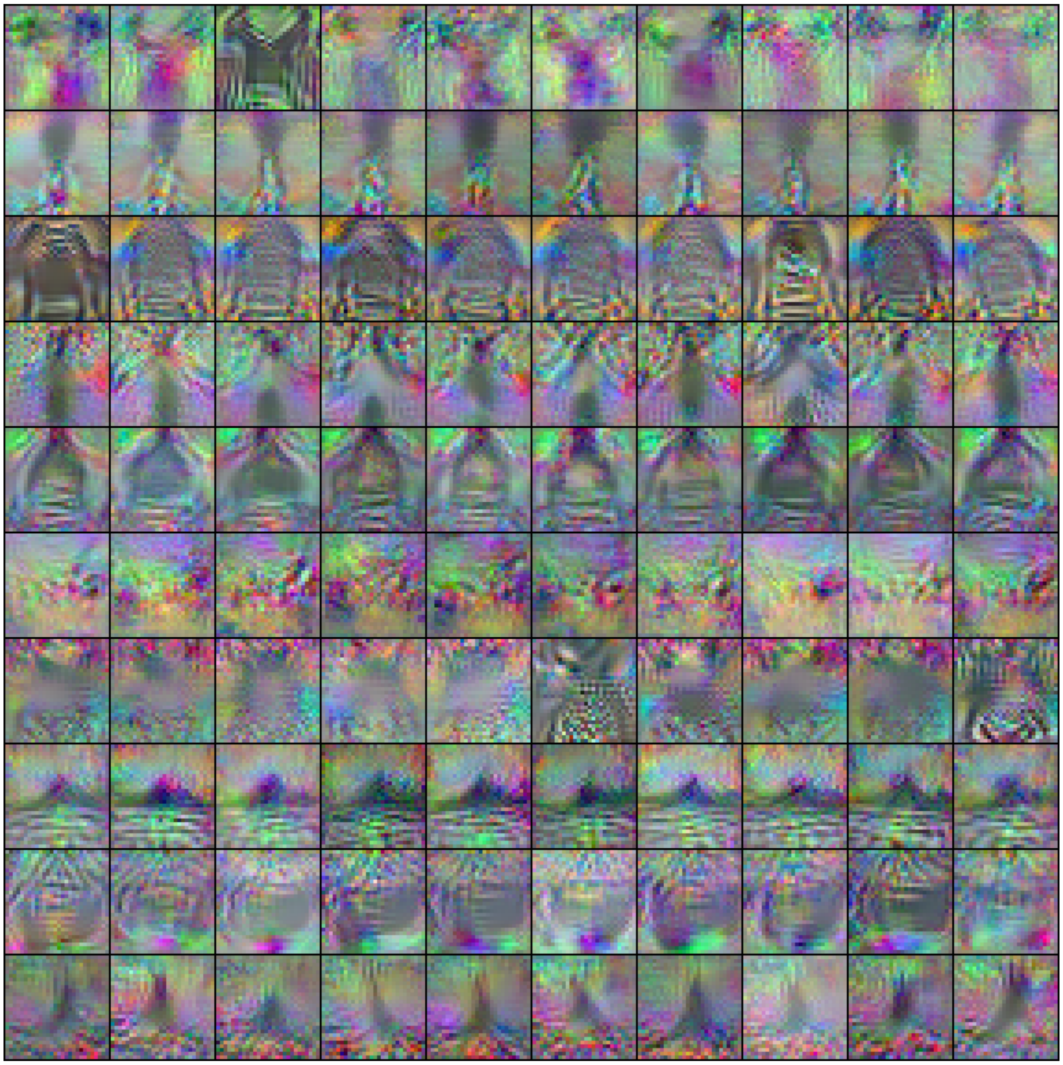

Figure 16: Synthesized dataset by MTT on BG FMNIST with 5% bias-conflicting samples.

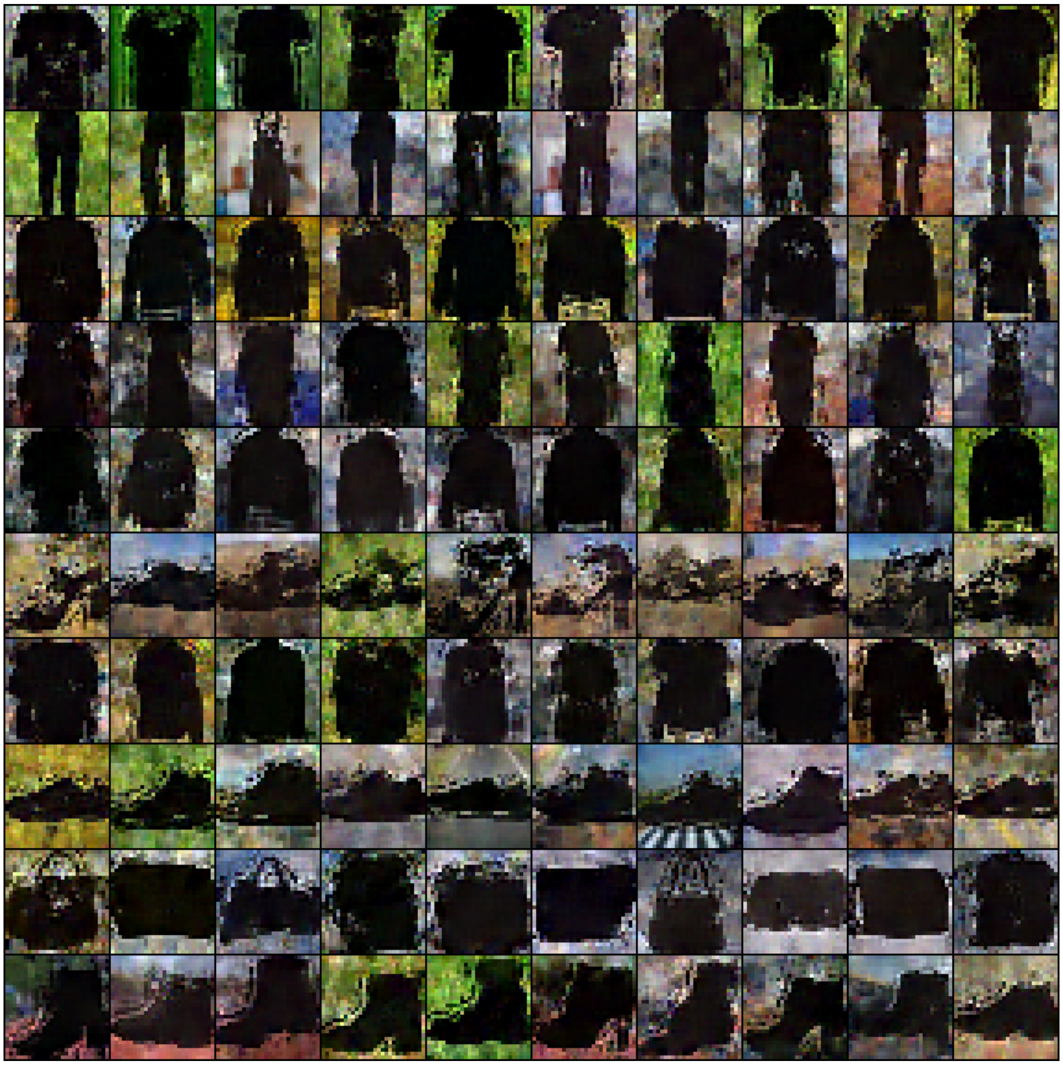

Figure 17: Synthesized dataset by DM+Ours on BG FMNIST with 5% bias-conflicting samples.

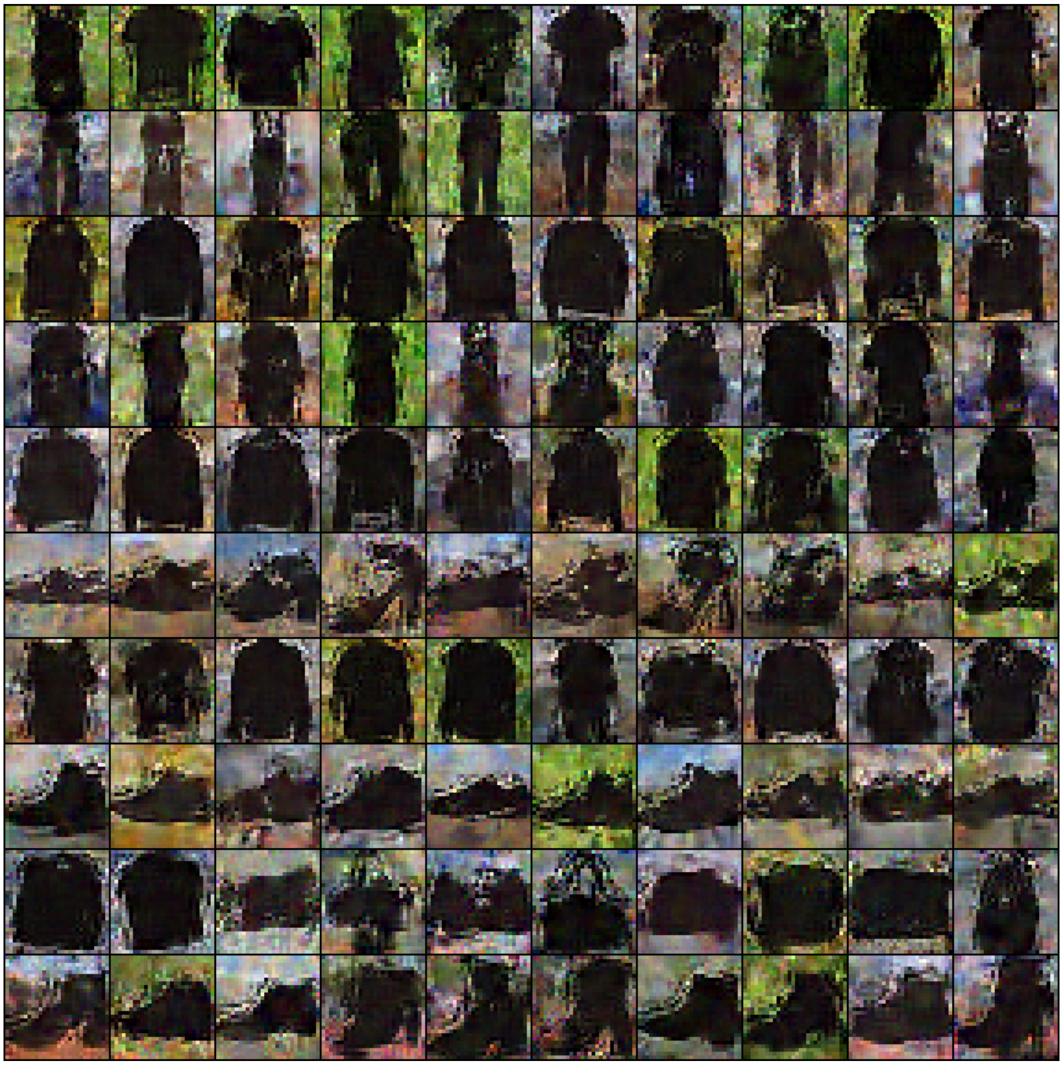

Figure 18: Synthesized dataset by DSA+Ours on BG FMNIST with 5% bias-conflicting samples.