FMD: Comprehensive Data Compression in Medical Domain via Fused Matching Distillation : Supplementary Material

Ju Heon Son* Jang-Hwan Choi†
Ewha Womans University
{thswngjs77, choij}@ewha.ac.kr

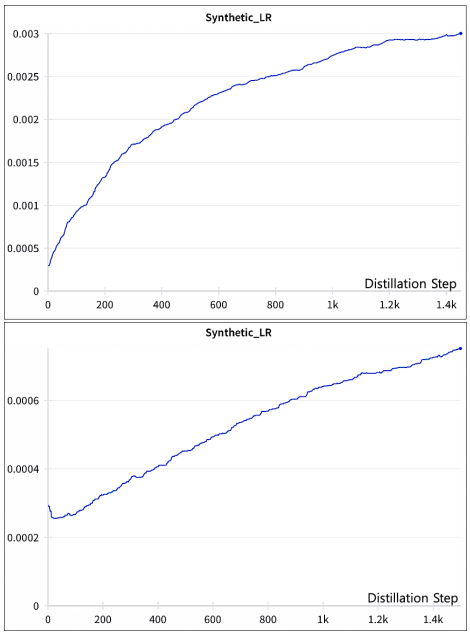

Figure 1. This figure illustrates the formation of the **learning rate** in the student network during a bifurcated distillation learning process based on parameter intervals in the **Pancreas cancer CT**. It aims to demonstrate how the learning rates differ based on different matching strategies applied to the student network according to parameter intervals, as shown in the upper and lower parts of the figure.

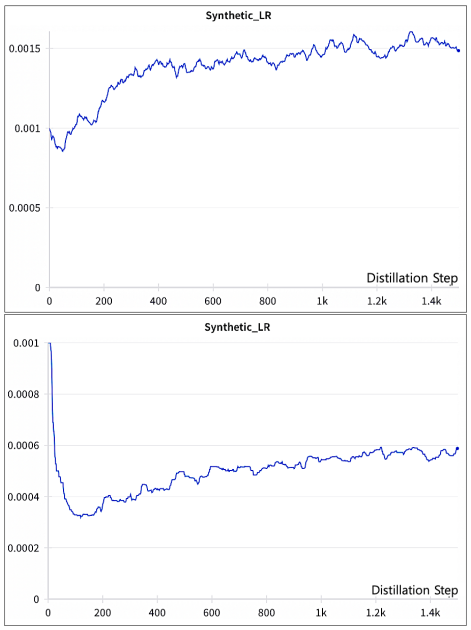

Figure 2. This figure demonstrates the development of the **learning rate** in the student network during a bifurcated distillation learning process based on parameter intervals in the **APTOS 2019** [1] dataset. Similar to Fig. 1, it demonstrates the distinct learning rate patterns of student networks in the upper and lower parts, which are trained under different parameter intervals.

1. Supplementary Details

1.1. Additional Experiments

In Fig. 1, the learning rate trends of a student network are depicted for cases where the parameter interval exceeds a criteria value, ω, in the Pancreas cancer CT dataset. The learning rate shows a fivefold increase at 1500 steps compared to when the student network is aptly trained with parameter intervals less than ω. Similarly, Fig. 2 examines the APTOS 2019 dataset, where the upper graph represents the learning rate trends for larger parameter intervals than ω, displaying a 2.5 times increase in learning rate at the same step count. These observations confirm that learning rates vary significantly with parameter intervals. Integrating this data through an appropriate weighted sum could better capture and reflect the distinct tendencies of each phase, thus enriching the information retained during the distillation process.

Fig. 3 and Fig. 4 illustrate the outcomes of feature distribution matching, demonstrating the evolving relationship between feature embeddings of real and synthetic datasets as the distillation steps progress. Both in the Pancreas can-

*First author
†Corresponding author

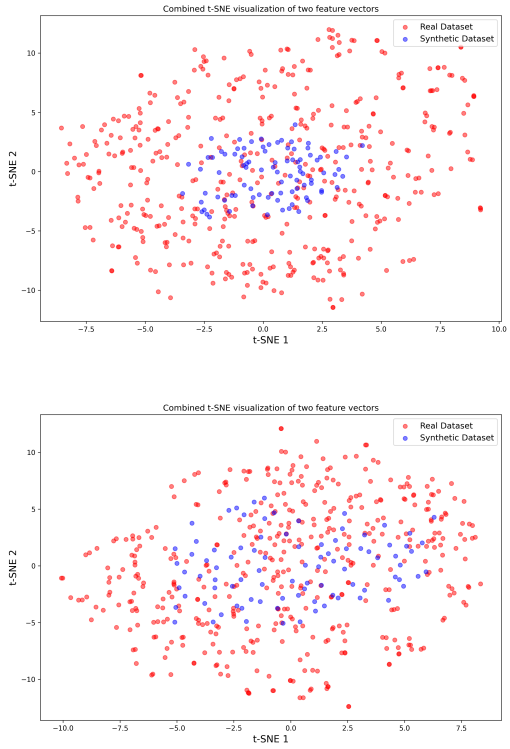

Figure 3. Visualization of **feature embeddings** for real and synthetic datasets of the **Pancreas Cancer CT Dataset**, achieved through feature distribution matching using t-SNE. Red dots represent the feature embeddings of the real dataset, while blue dots represent those of the synthetic dataset. The upper figure shows the embeddings at distillation step 0, and the lower figure at 800.

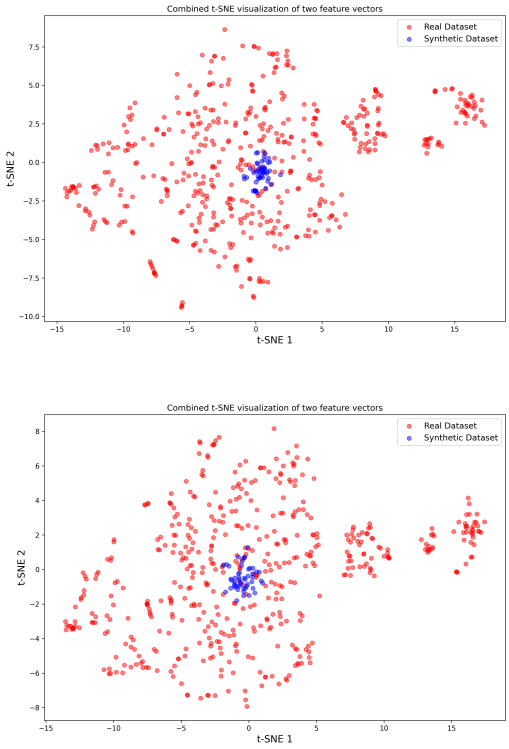

Figure 4. Visualization of **feature embeddings** for real and synthetic Datasets of the **APTOS 2019 Dataset** achieved through feature distribution matching using t-SNE. Similar to Fig. 3, the upper part of the figure displays the embeddings at distillation step 0, while the lower part shows them at step 2500.

Table 1. Experiment on the **criteria for dividing parameter intervals** on a COVID chest X-ray dataset. Here, ω denotes the criterion for segmenting parameter intervals, and IQR represents the Interquartile Range. The highest accuracy is indicated by bold text.

ω	**Acc**
Q1-1.5*IQR	0.8309
Q1	**0.8370**
Q3	0.8181
Q3+1.5*IQR	0.8286

cer CT and APTOS 2019 datasets, the synthetic dataset's feature embeddings, initially clustered at the center, gradually disperse to more closely resemble the feature distribution of the real dataset. However, the trend is less pronounced in the APTOS 2019 dataset, likely due to differences in the channels of the dataset, which suggests distinct behaviors between the two datasets.

An ablation experiment was conducted to determine the optimal criterion, ω, for dividing parameter intervals. It is common practice in statistics to use quartiles $Q1$ and $Q3$ to designate outliers [3]. Outliers are typically defined using the interquartile range (IQR), calculated as $Q3 - Q1$, where values below $Q1 - 1.5 \times$ IQR or above $Q3 + 1.5 \times$ IQR are considered outliers. According to Table 1, using $Q1$ as the criterion for outliers within parameter intervals results in better performance than using $Q3$. This suggests that differentiating between phases with extremely small parameter intervals and those without is effective. Further observations indicate that setting the boundary ω at $Q1$, thereby allowing a more generous threshold than when using the IQR, yields better performance. This appears to be due to the tendency of parameter updates to occur predominantly in the initial epochs, with only fine adjustments needed as training progresses. Therefore, the first quartile, $Q1$, has been set as the boundary value for dividing parameter intervals, denoted by ω. Table 2 facilitates the identification of the optimal hyperparameter values for the weighted summation process used to amalgamate multiple synthetic datasets into a single comprehensive synthetic dataset.

Table 2. **Performance comparison** based on hyperparameters during integration. The highest score is highlighted in bold.

	(α,β,γ)	**Acc**
$\alpha+\beta>\gamma$	(0.5,0.3,0.2)	0.8230
	(0.6,0.2,0.2)	**0.8412**
$\alpha+\beta=\gamma$	(0.4,0.1,0.5)	0.7840
$\alpha+\beta<\gamma$	(0.3,0.1,0.6)	0.7404

1.2. Experimental Details

We developed our framework using PyTorch and executed it on an NVIDIA RTX A6000 GPU. Table 3 displays the experimental conditions and hyperparameters used across all datasets. To ensure fair comparisons during benchmarking and experimentation, we fixed the seed.

2. Supplementary Visualizations

Fig. 5, Fig. 6, and Fig. 7 display synthesized datasets at the distillation step where the evaluation network achieves its peak performance. These datasets, which are visually anonymized, are derived from COVID chest X-ray [2], Pancreas cancer CT, and APTOS 2019, following the sampling of ten images per class from the entire dataset and subsequent distillation. In each figure, IPC stands for Images Per Class.

Table 3. This table displays the hyperparameters and experimental settings that yielded the highest performance across all datasets used in the experiments in this paper.

Dataset	Model	Distillation steps	lr_img	lr_init	lr_lr	DSA	M	Max start epoch
COVID chest X-ray	ConvNetD5	5000	100	0.001	1e-05	True	2	20
Pancreas cancer CT	ConvNetD5	5000	100	0.0003	1e-07	True	2	20
APTOS 2019	ConvNetD5	1500	1000	0.001	1e-07	True	2	10
CIFAR10	ConvNetD3	5000	1000	0.01	1e-05	True	2	20
Imagenette	ConvNetD5	5000	1000	0.01	1e-05	True	2	10
Imagewoof	ConvNetD5	5000	1000	0.01	1e-05	True	2	10

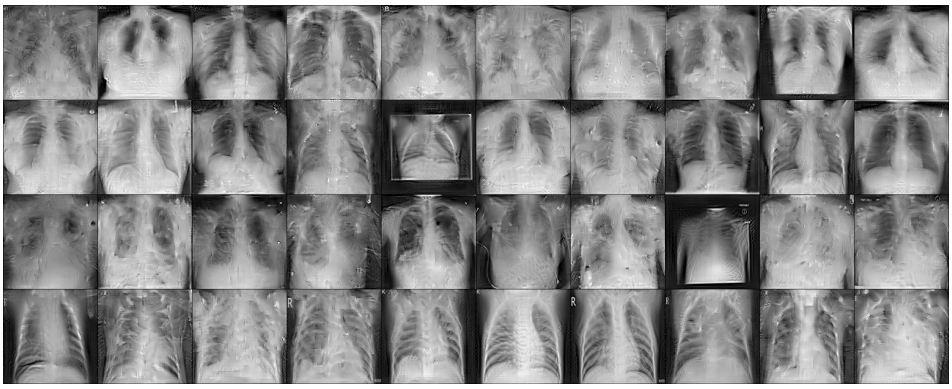

Figure 5. **COVID chest X-ray at IPC 10.**

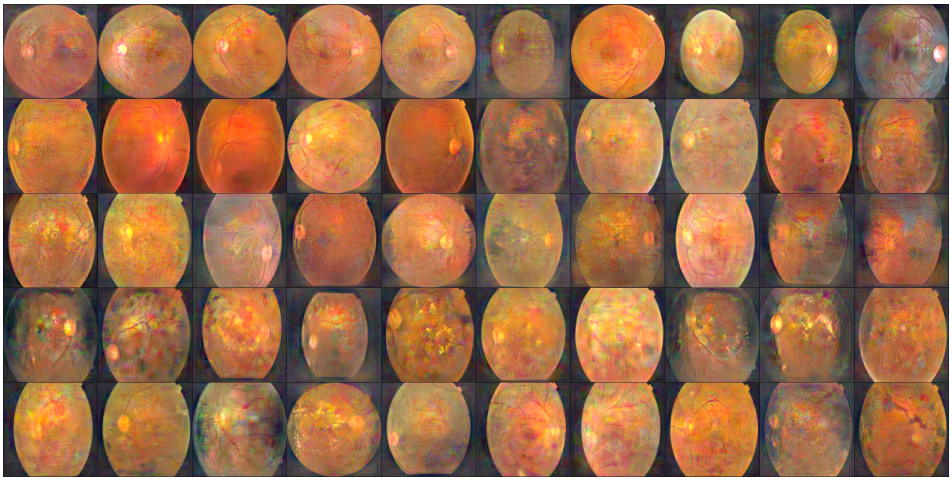

Figure 6. **APTOS 2019 at IPC 10.**

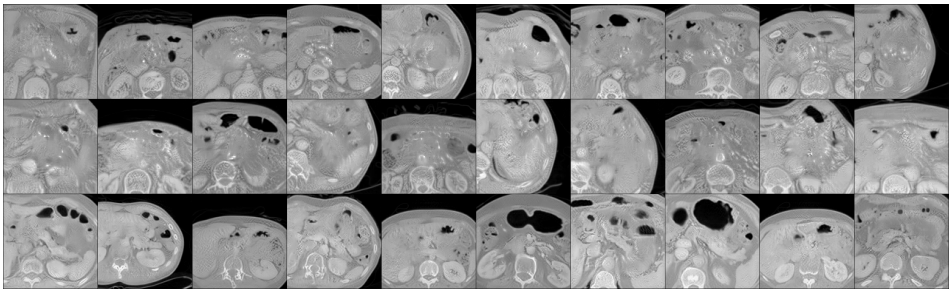

Figure 7. **Pancreas cancer CT at IPC 10.**

References

[1] Kaggle and APTOS. Aptos 2019 blindness detection, 2019. Available: https://www.kaggle.com/c/aptos2019-blindness-detection. 1

[2] Tawsifur Rahman, Amith Khandakar, Yazan Qiblawey, Anas Tahir, Serkan Kiranyaz, Saad Bin Abul Kashem, Mohammad Tariqul Islam, Somaya Al Maadeed, Susu M Zughaier, Muhammad Salman Khan, et al. Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images. *Computers in biology and medicine*, 132:104319, 2021. 3

[3] HP Vinutha, B Poornima, and BM Sagar. Detection of outliers using interquartile range technique from intrusion dataset. In *Information and decision sciences: Proceedings of the 6th international conference on ficta*, pages 511–518. Springer, 2018. 2