Generative Style Transfer for MR Image Segmentation: A Case of Glioma Segmentation in Sub-Saharan Africa

Rancy Chepchirchir, & Jill Sunday, & Raymond Confidence, & Dong Zhang, & Udunna C Anazodo, & Kendi Muchungi & Yujing Zou [2]

Objective

Brain MR images acquired in Sub-Saharan Africa experience suboptimal image contrast and resolution which may be due to the use of low-field MRI scanners. Magnetic resonance images obtained using low magnetic field strength, typically below 1 Tesla (T), usually have impaired resolution along the slice direction and lower contrast.

Data & Methods

The dataset employed represents the largest publicly-available retrospective cohort of adult Africans with pre-operative glioma, covering both low-grade glioma (LGG) and high-grade glioma (GBM/HGG). It comprises routine multi-parametric MRI (mpMRI) scans acquired during standard clinical care, encompassing T1, post-contrast T1 (T1Gd), T2, and T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) MRI.

Here, the model's performance is evaluated using Dice scores for three different classes: Dice_ET, Dice_TC, and Dice_WT. Each fold represents a different split of the dataset used for training and testing. The results are recorded for five different epochs (2, 5, 10, 30, and 300). Overall, the model achieves good performance across the folds, with Dice scores ranging from approximately 0.77 to 0.94 for the different classes.

Figure 1. Style transfer between the content (SSA image) and style (GLI image)

Figure 2. Graphical Representation of training progress for 300 epochs

Model	Dice Score			Epochs
	Dice_ET	Dice_TC	Dice_WT	
Fold 0	0.8689	0.8205	0.8082	2
	0.9131	0.8534	0.8304	5
	0.8963	0.8231	0.8415	10
	0.9388	0.9031	0.8991	30
	0.9471	**0.9179**	**0.9179**	300
Fold 1	0.8745	0.7932	0.776	2
	0.9234	0.866	0.8509	5
	0.9327	0.8839	0.8922	10
	0.9294	0.8949	0.8958	30
	0.9488	**0.9051**	**0.8890**	300
Fold 2	0.8805	0.8024	0.8212	2
	0.9178	0.8568	0.8352	5
	0.9388	0.9062	0.8977	10
	0.9469	0.9145	0.902	30
	0.9369	**0.8955**	**0.8989**	300
Fold 3	0.8197	0.7734	0.7608	2
	0.9129	0.837	0.8247	5
	0.9303	0.8709	0.8414	10
	0.9395	0.9096	0.9009	30
	0.9441	**0.9101**	**0.9011**	300
Fold 4	0.8706	0.8028	0.7922	2
	0.8914	0.8247	0.8004	5
	0.9248	0.8637	0.8592	10
	0.9294	0.8949	0.8958	30
	0.9459	**0.9222**	**0.9211**	300

Table 1. A comparison of performance across the folds

Results

Dataset	Epoch	Learning Rate	Train Loss	Val Loss	Pseudo Dice	Epoch Time (s)
2021	2	0.00536	-0.6746	-0.7091	0.7833	169.28
	5	0.00235	-0.8625	-0.8687	0.8629	165.65
	10	0.00126	-0.7959	-0.7784	0.8305	195.28
	30	0.00047	-0.8341	-0.8523	**0.8965**	190.9
	300	6e-05	-0.8844	-0.8664	**0.9471**	201.65
GLI + SSA	2	0.00536	-0.4167	-0.6398	0.7762	449.16
	5	0.00235	-0.7226	-0.7168	0.7945	467.69
	10	0.00126	-0.777	-0.7911	0.8385	427.12
	30	0.00047	-0.841	-0.85	**0.895**	427.3
	300	6e-05	-0.8891	-0.8561	**0.9256**	496.42

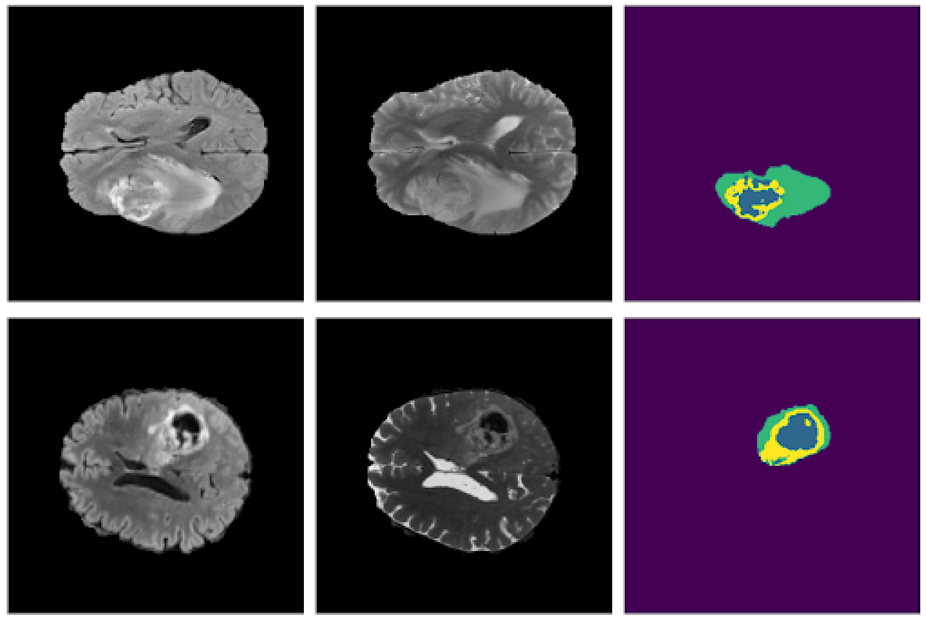

Figure 3. Predicted Masks for BraTS-GLI-00001-001, and BraTS-GLI-00013-000 cases respectively

Conclusion

Based on our experiments, training a model after proper data augmentation with a higher epoch size, say 300, yields better results as opposed to a smaller epoch size.

References

[1] Maruf Adewole, Jeffrey D Rudie, Anu Gbadamosi, Oluyemisi Toyobo, Confidence Raymond, Dong Zhang, Ol-ubukola Omidiji, Rachel Akinola, Mohammad Abba Suwaid, Adaobi Emegoakor, et al. The brain tumor segmentation