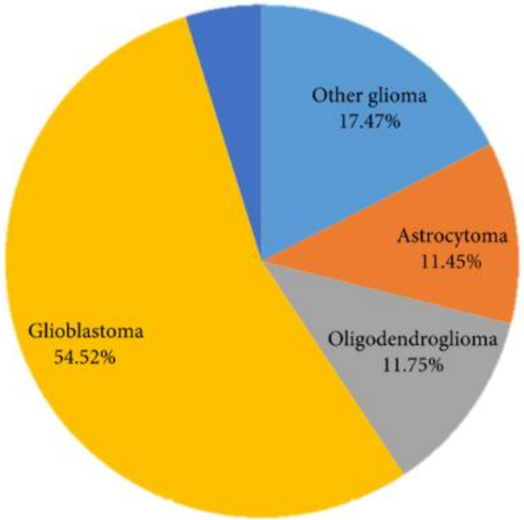
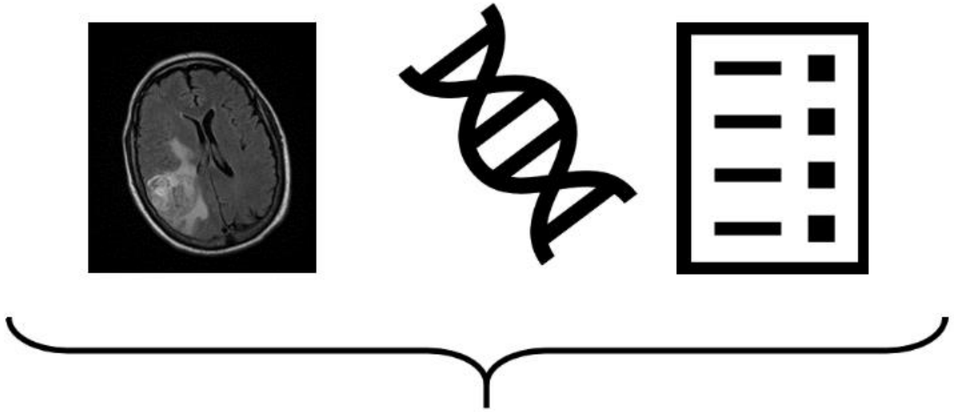
International Conference on Pattern Recognition (ICPR) 2020

SAGE: Sequential Attribute Generator for Analyzing Glioblastomas Using Limited Dataset

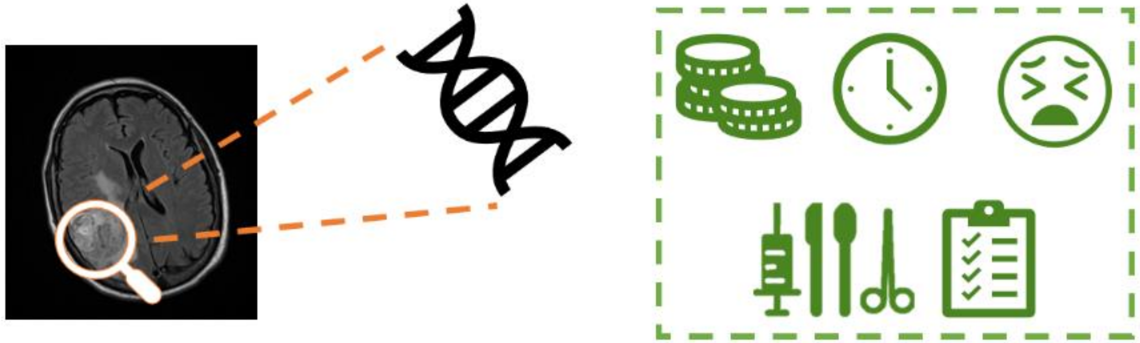
Padmaja Jonnalagedda, Brent Weinberg (MD, PhD), Jason Allen (MD, PhD), Taejin L. Min (MD), Shiv Bhanu (MD) and Bir Bhanu (PhD)*

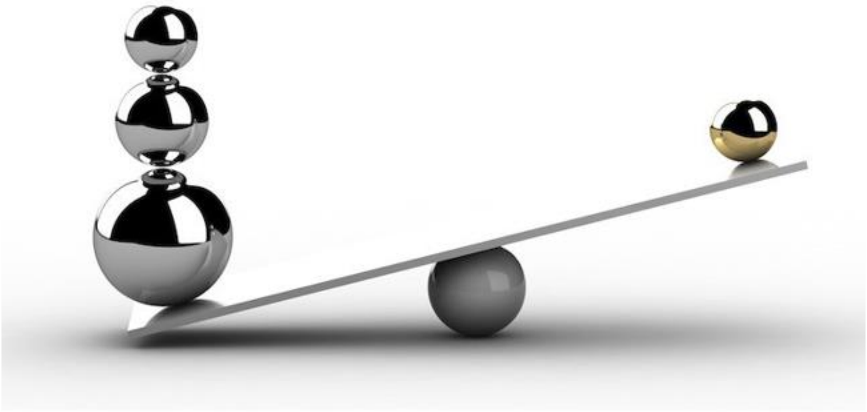
*Center for Research in Intelligent Systems
University of California, Riverside, CA, USA





Prognosis





Mutations are rare!

Problem and Significance

https://www.hindawi.com/journals/jce/2020/1429615/
https://images.app.goo.gl/iWz1UFf5yuYGYrom6

Related Work and Contributions

- Some research has been directed towards brain MR generation [1-3], however, they do not cater to the high diversity and low quantity of data
- Contributions:
  - Generate **diverse synthetic images** from very limited datasets
  - Feature disentanglement and sequential generation for **high resolution** images and **added control** over generated tumor properties,
  - Quantitative analysis of efficacy of proposed method in learning and **recreating visually unapparent data** distribution compared to naive GANs

1. Changhee Han, Hideaki Hayashi, Leonardo Rundo, Ryosuke Araki, Wataru Shimoda, Shinichi Muramatsu, Yujiro Furukawa, Giancarlo Mauri, and Hideki Nakayama, "Gan-based synthetic brain mr image generation," in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018, pp. 734–738.
2. Zhang, C., Song, Y., Liu, S., Lill, S., Wang, C., Tang, Z., You, Y., Gao, Y., Klistorner, A., Barnett, M. and Cai, W., 2018, December. MS-GAN: GAN-based semantic segmentation of multiple sclerosis lesions in brain magnetic resonance imaging. In 2018 Digital Image Computing: Techniques and Applications (DICTA) (pp. 1-8). IEEE.
3. Han, C., Murao, K., Noguchi, T., Kawata, Y., Uchiyama, F., Rundo, L., Nakayama, H. and Satoh, S.I., 2019, November. Learning more with less: Conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (pp. 119-127).

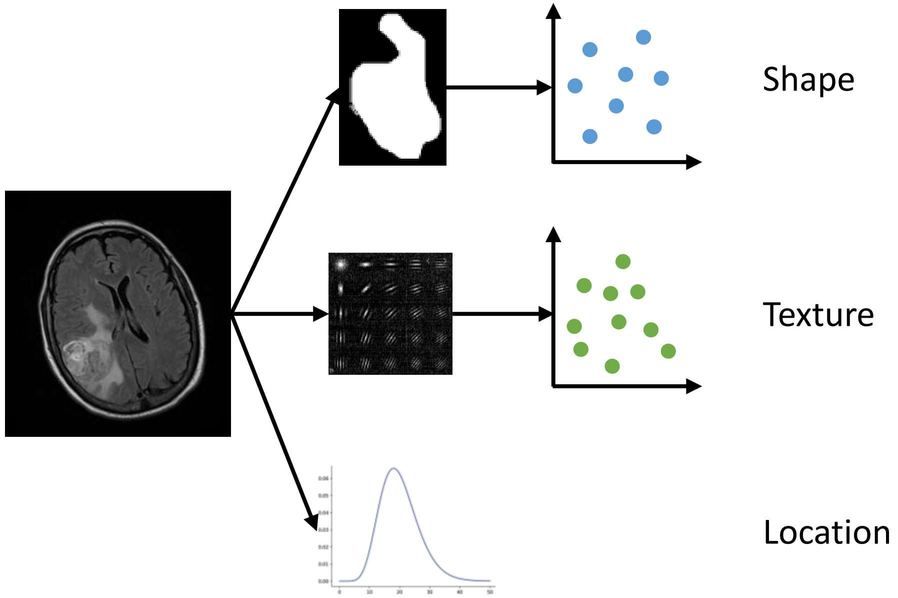
Proposed Approach

What we know*

- ✓ Visual indicators of mutation:
  - Exist
  - Reproducible
- ✓ Significant macro-features:
  - Shape, texture and location of tumor
- ✓ Data highly diverse and limited

What we do

Break down a tumor and recreate from the data distributions



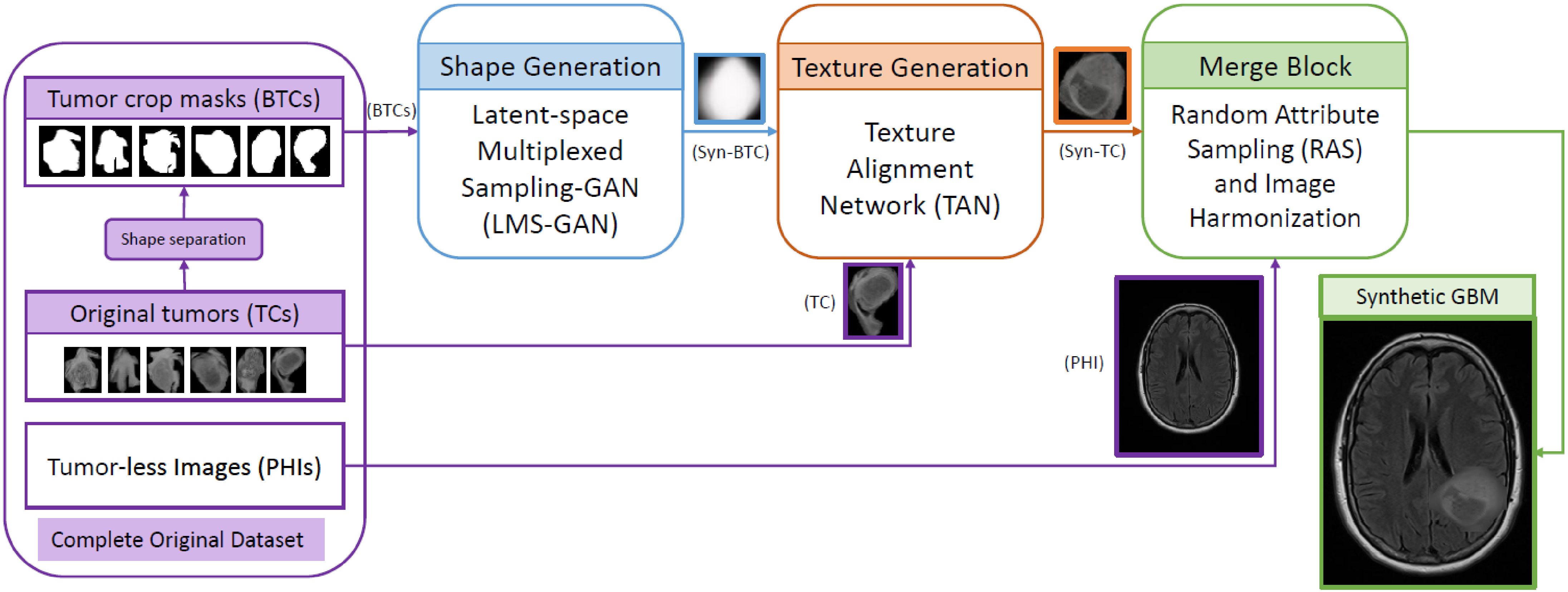
The result?

A model that can tackle the issues of:

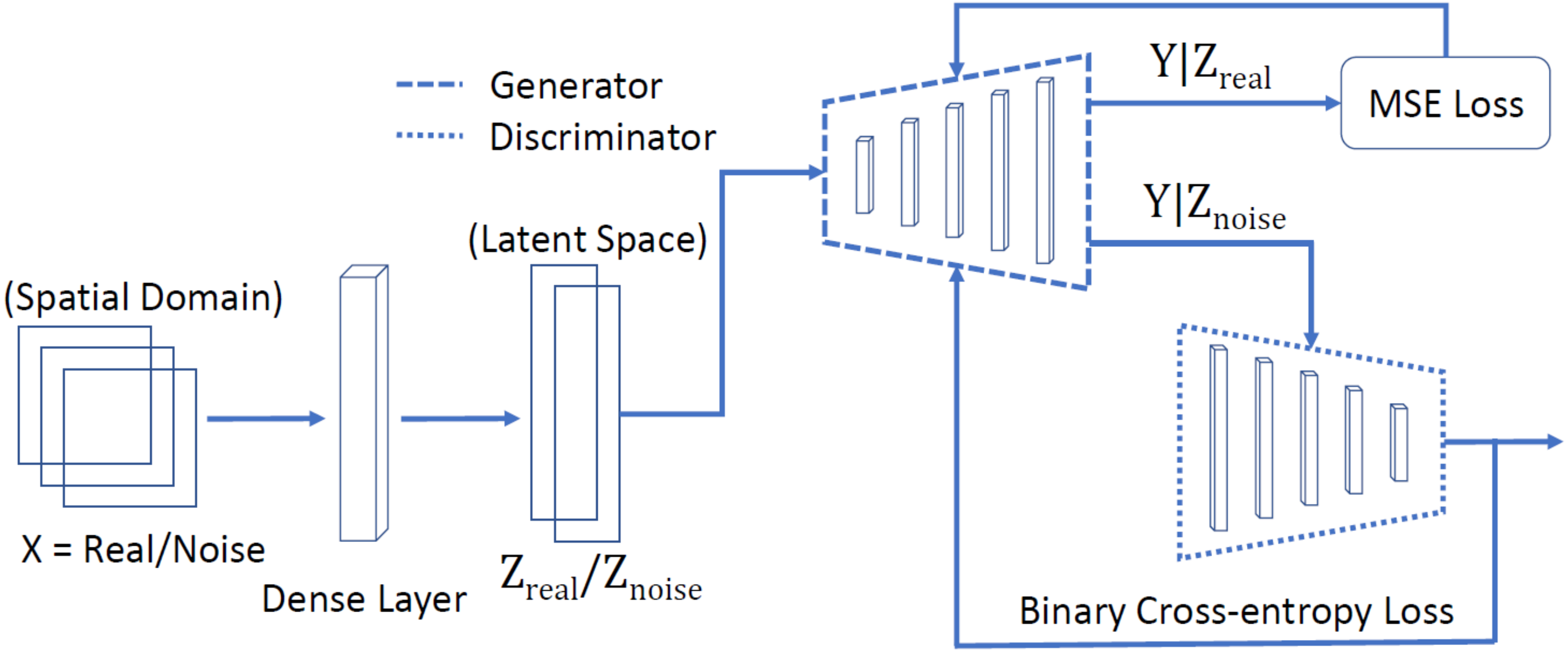
- ✓ Limited data
- ✓ High data diversity
- ✓ Learning visually unapparent features

*Jonnalagedda, Padmaja, Brent Weinberg, Jason Allen, and Bir Bhanu. "Feature Disentanglement to Aid Imaging Biomarker Characterization for Genetic Mutations." In *Medical Imaging with Deep Learning*, pp. 349-364. PMLR, 2020.

SAGE: Sequential Attribute Generator



Shape Generation

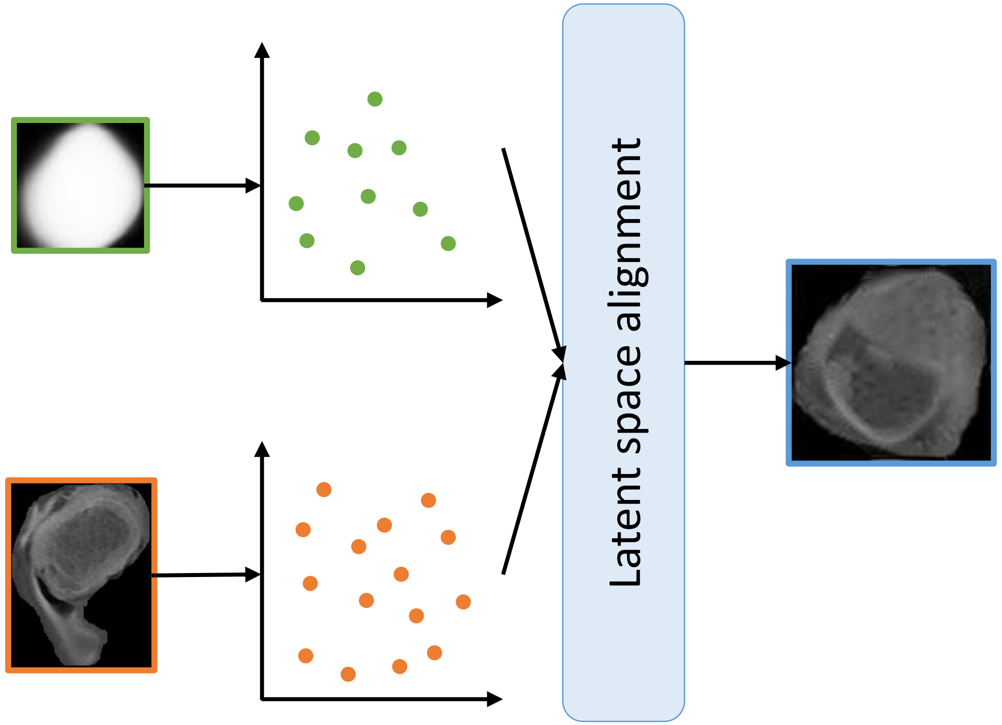


$$\mathcal{L}_{X\sim p(real)} : \mathbb{E}_{x\sim p_{real}(x)}||x_{real} - y||_2 \qquad \mathcal{L}_{X\sim p(noise)} : \mathbb{E}_{x\sim p_{noise}(x)}\left[\log\left(1 - \mathbf{D}(\mathbf{G}(z_{noise}))\right)\right]$$

$$\mathcal{L}_{LMS} = \mathbb{I}_x \times \mathcal{L}_{X\sim p(noise)} + (1 - \mathbb{I}_x) \times \mathcal{L}_{X\sim p(real)}$$

where I_x is the indicator function which is 0 when X is data and 1 when X is noise. The noise is sampled from random normal distribution. G and D are Generator and Discriminator, respectively. y is the generated image and x_{real} is the real image. $Z_{real=noise}$ are the latent space representations of real and noisy inputs.

Texture Generation

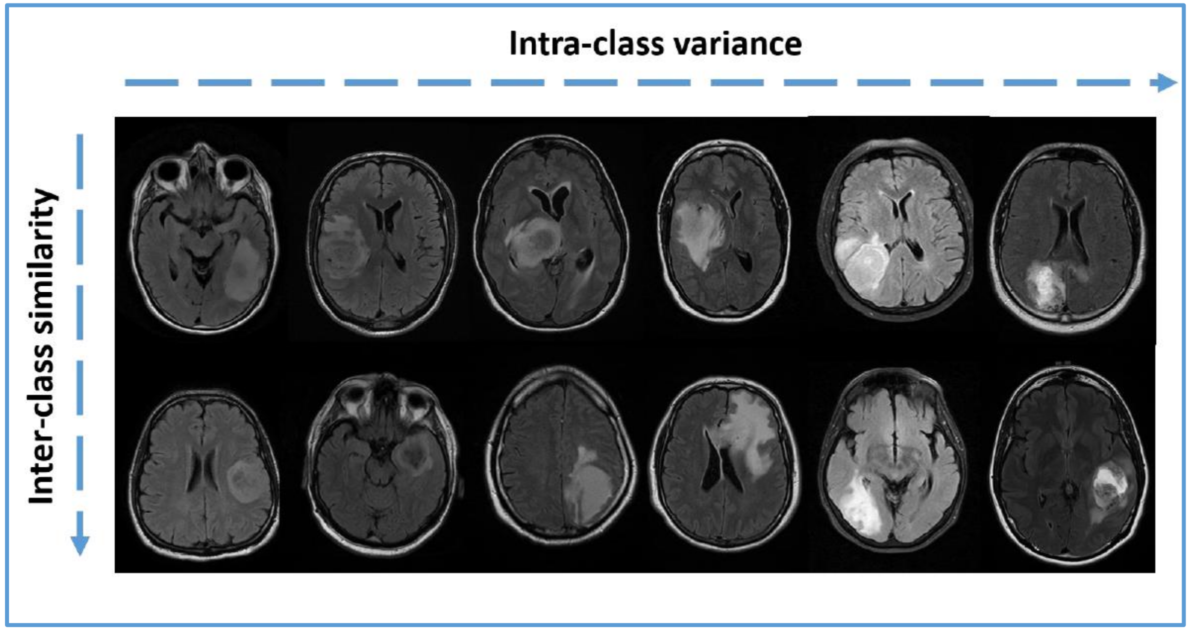


$$\mathcal{L}_{TAN} : \sum \omega_i \times ||\mathrm{g}_i(T) - \mathrm{g}_i(I)||_2 + ||\mathbb{F}_i(B) - \mathbb{F}_i(I)||_2$$

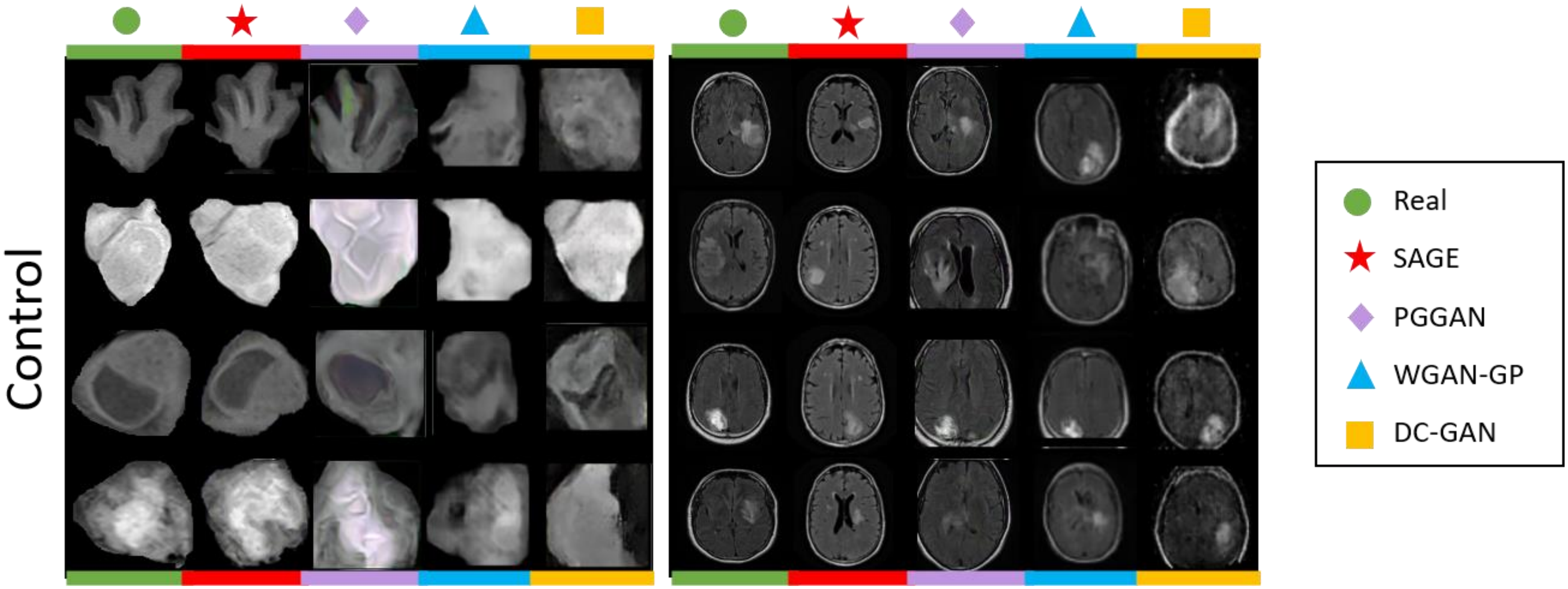
where: L_{TAN}: overall loss function, ω_i: weight of ith layer, T: sampled Tumor Crop (TC), B: input Syn-Binary TC, I: output Syn-TC, F_i: output of i^{th} layer and g_i: Gram matrix of i^{th} layer output

Data Used

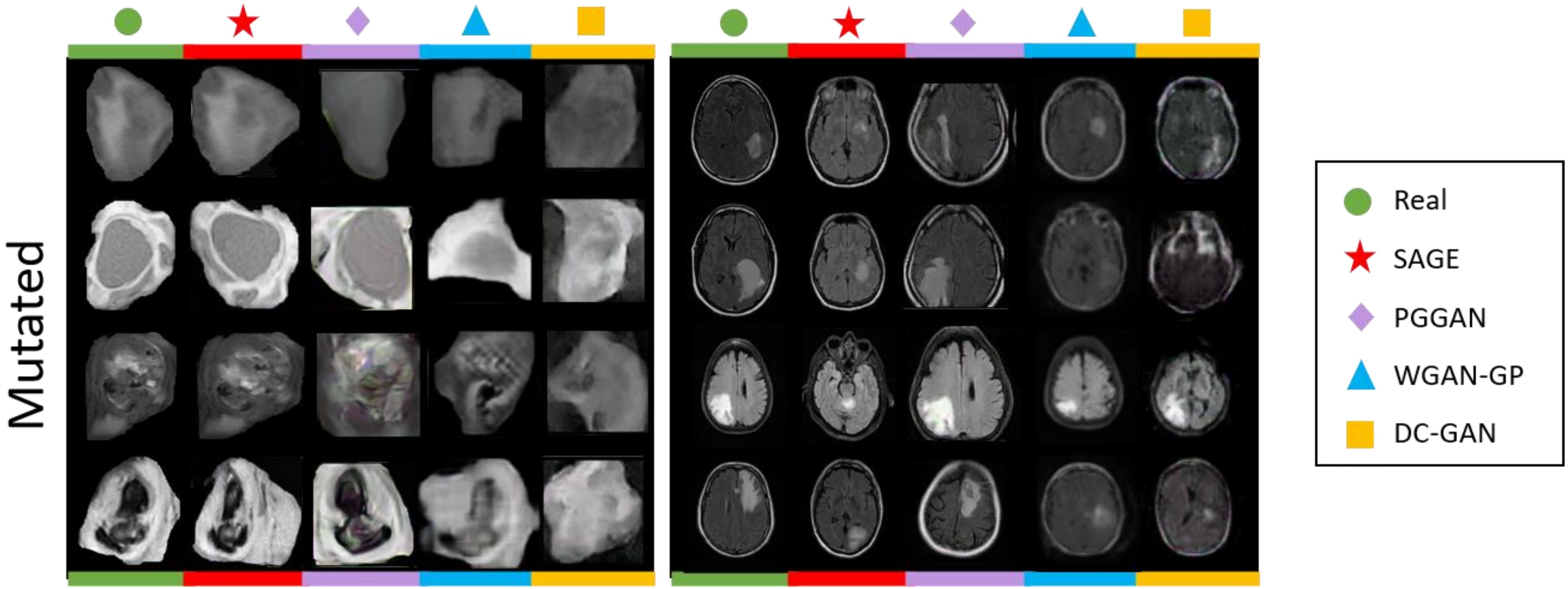
- Data provided by Emory University
- 38 patients: 18 mutated and 20 control
- Two classes: Mutated and control classes (with and without 19/20 co-gain)



Results



Results

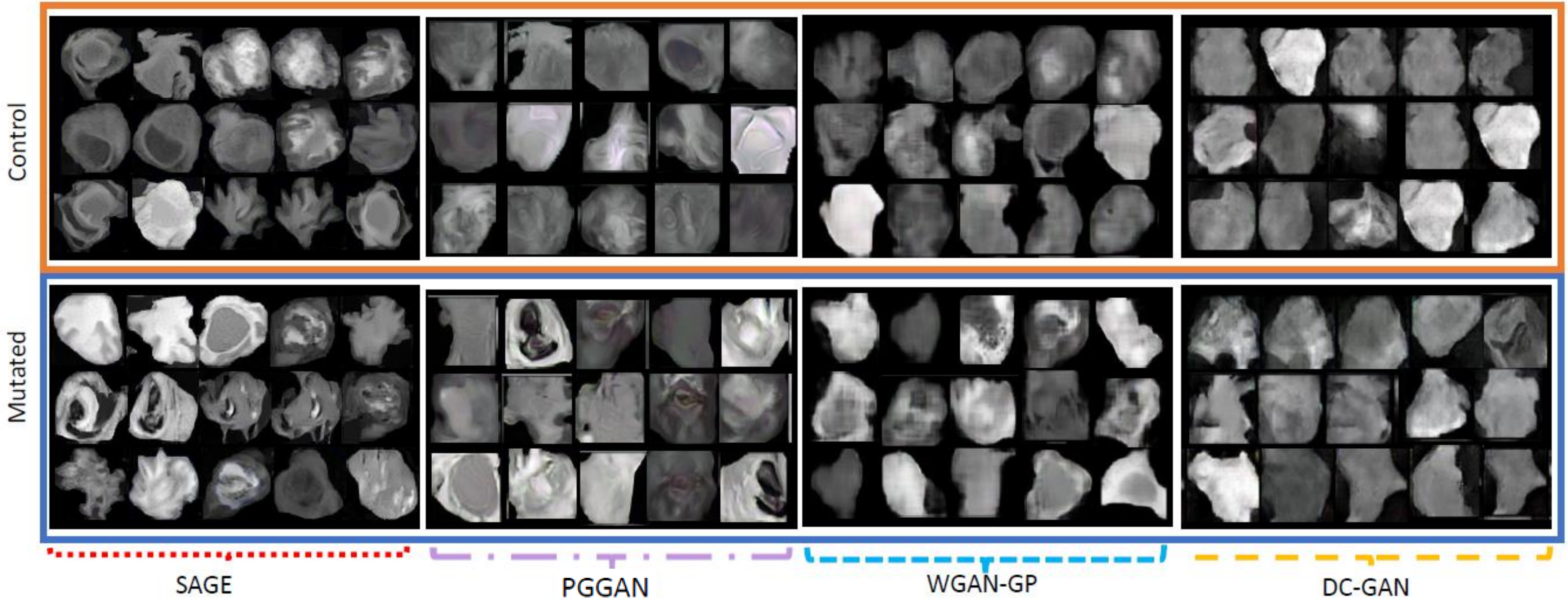


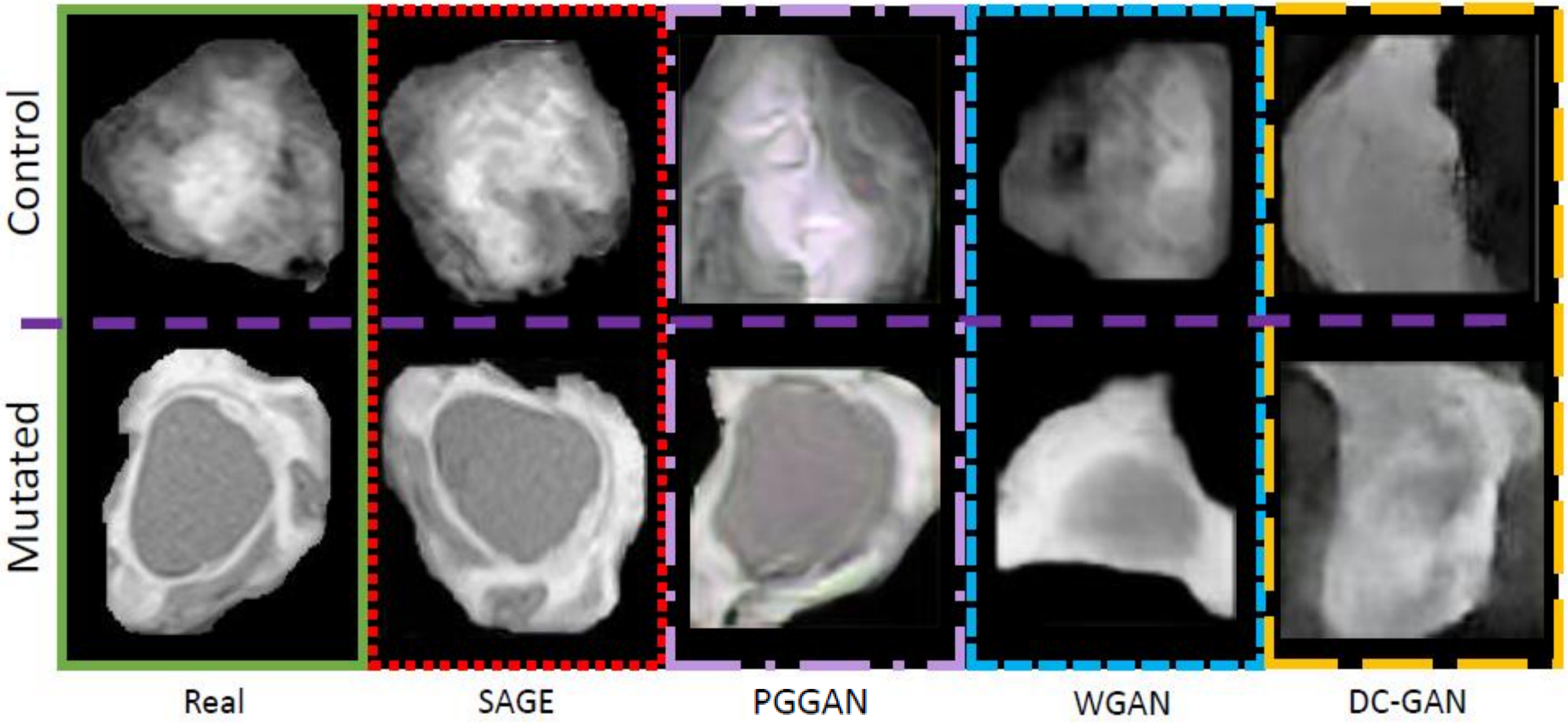
Blind Test by Radiologists

Blind test for radiologists to distinguish between real and synthetic tumor crops. The metrics: ACC - Accuracy, FPR - False Positive Rate, TNR - True Negative Rate, PR - Precision. Row (non-shaded) are values for each radiologist and row (shaded) is the mean value across all radiologists.

Data class	Radiologist	ACC	FPR	FNR	PR
Control	1	0.59	0.42	0.38	0.55
	2	0.67	0.34	0.31	0.66
	3	0.70	0.30	0.36	0.64
	Mean	0.66	**0.35**	0.33	0.64
Mutated	1	0.74	0.33	0.19	0.72
	2	0.82	0.19	0.18	0.87
	3	0.76	0.30	0.20	0.77
	Mean	0.77	**0.27**	0.19	0.79

Diversity





Generative Model	IS	SSIM
SAGE (ours)	**1.71**	**0.68**
PGGAN	1.55	0.66
WGAN-GP	1.35	0.57
DC-GAN	1.12	0.32

Inception score (IS) and Structural Similarity (SSIM)

Quality test

Conclusions

- Sequential generation of disentangled attributes can:
  1. Cater to limited datasets (as low as 60 training samples per class)
  2. Generate high resolution images
  3. Generate high diversity dataset

  compared to PG-GAN and standard GANs
- SAGE generates realistic images – thus, it captures data distribution accurately
- SAGE tackles problems of real-world datasets, rendering it very useful for data generation tasks

Thank You!

Acknowledgement

This research is supported by the Bourns Endowment Fund at University of California Riverside