Surgery-aware Generative AI for Intraoperative MRI Generation

Shuhan Wang 2025.9.26

UWinnipeg Applied Computer Science department

Advisor : Dr. Qian Liu

By transferring the high-fidelity appearance and tissue style of preoperative MRI to compensate for and correct degradations and artifacts in intraoperative MRI.

Overview

1. **Project Background and Related Work / Gap**
   - Clinical motivation, prior approaches, and the unmet need this work addresses.
2. **Proposed Method**
   - Surgery-aware masking and attention-augmented generator; training strategy.
3. **Results and Metrics Analysis**
   - Datasets, evaluation protocol, quantitative/qualitative results, ablations.
4. **Conclusion**
5. **Future Work**

Project Background and Related Work / Gap

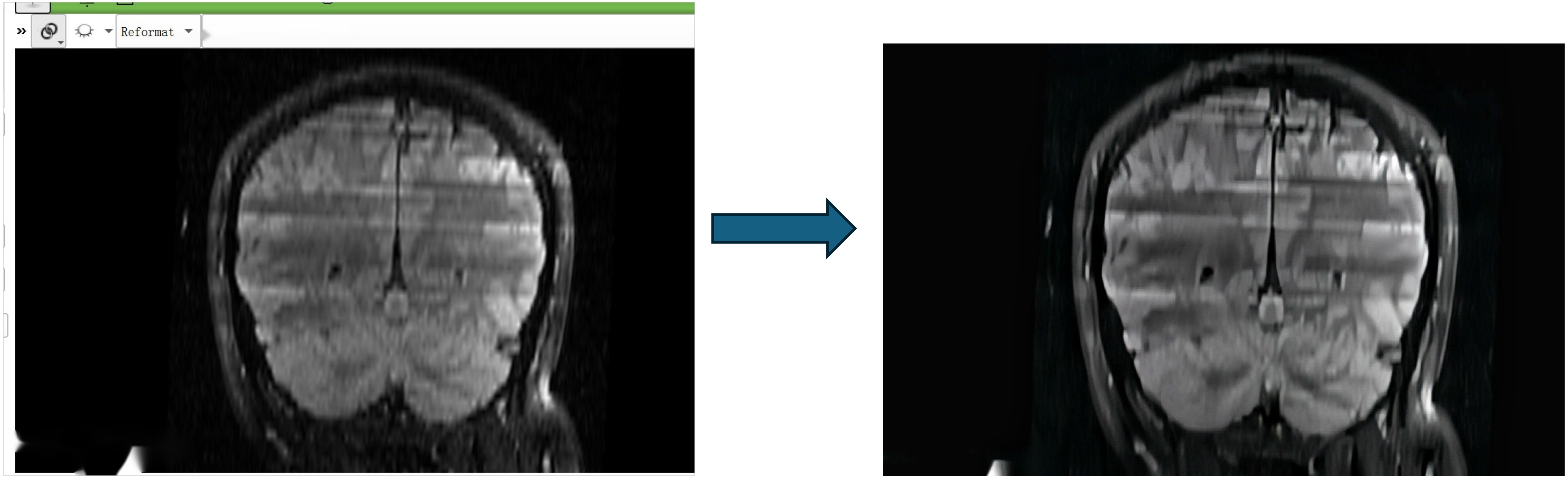
Preoperative MRI

Intraoperative MRI

The absence of clear tumor boundaries necessitates additional scanning sessions to ensure accurate localization and resection.

Poor intraoperative MRI image quality can affect the doctor's judgment of residue, prolong the time for re-scanning and prolong anesthesia.

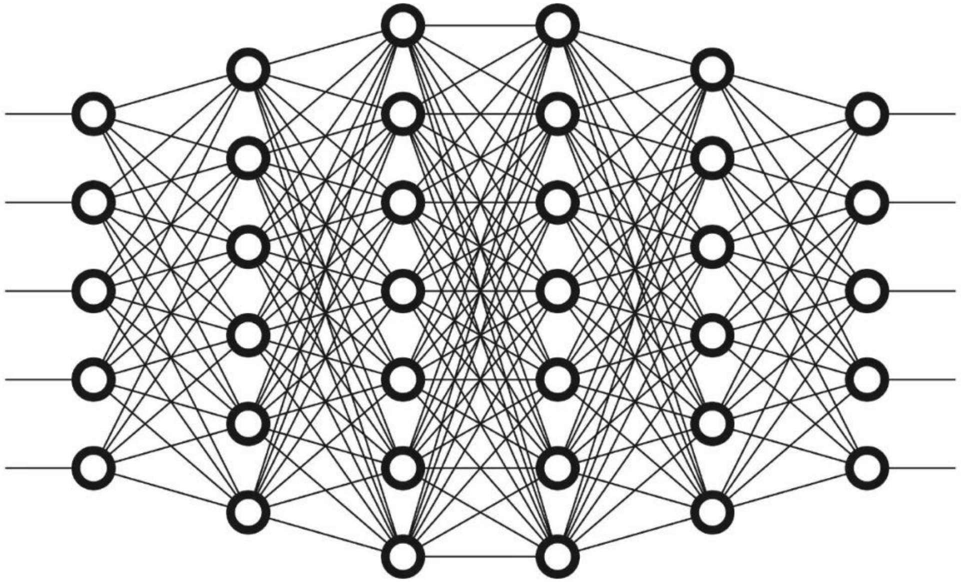
Disadvantages of traditional methods



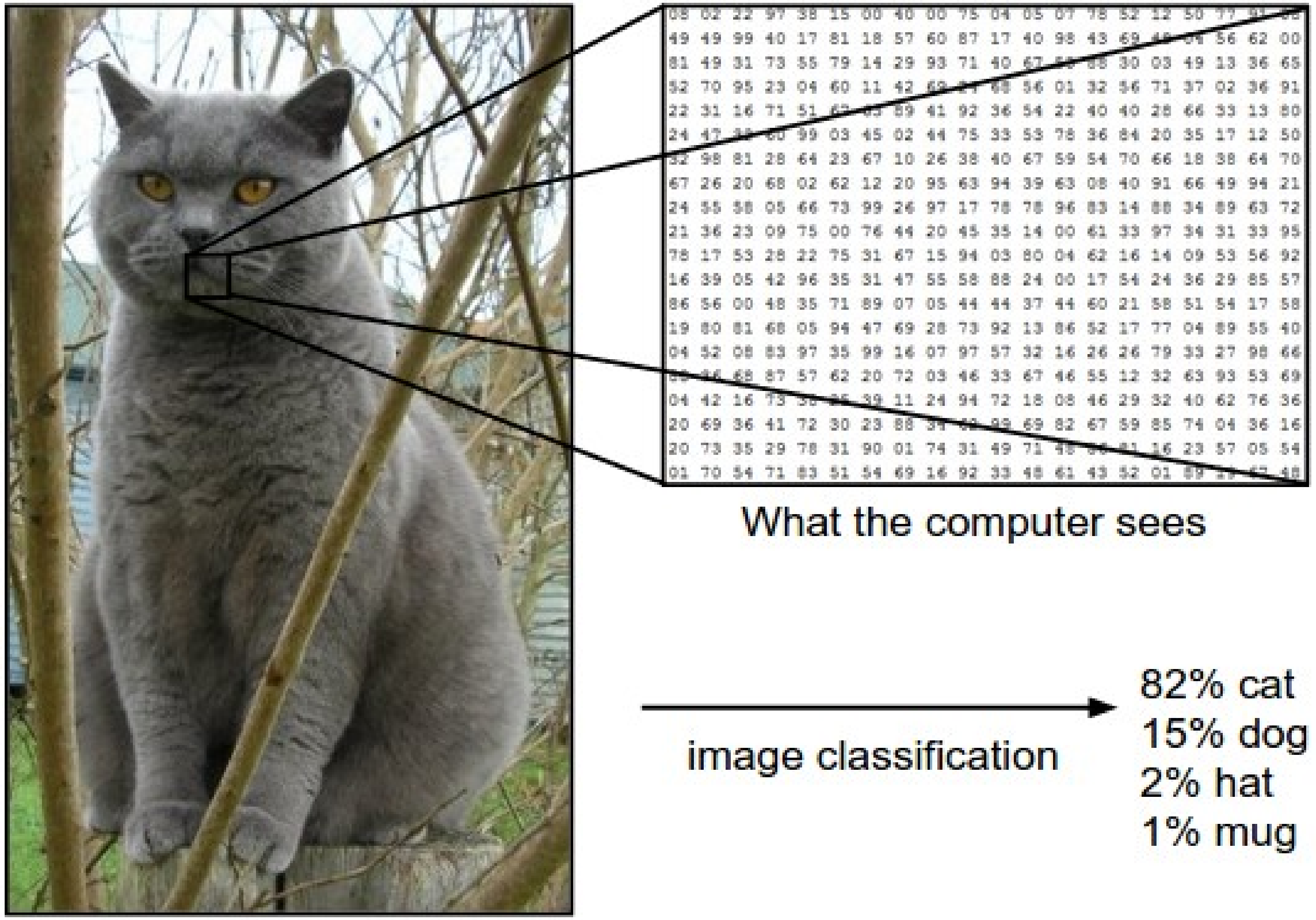
Traditional enhancement/noise reduction:
such as smoothing and brightening the entire photo, it is easy to erase the small boundaries or paint the holes.

What is deep learning?

- Deep learning is a technology that "teaches computers to think like humans". By imitating the working mode of the human brain, computers can learn to automatically recognize, analyze and judge from a large amount of data.
- The core tool of deep learning is the neural network, which is like a "universal formula". Input data, after layer upon layer of processing, outputs the result.

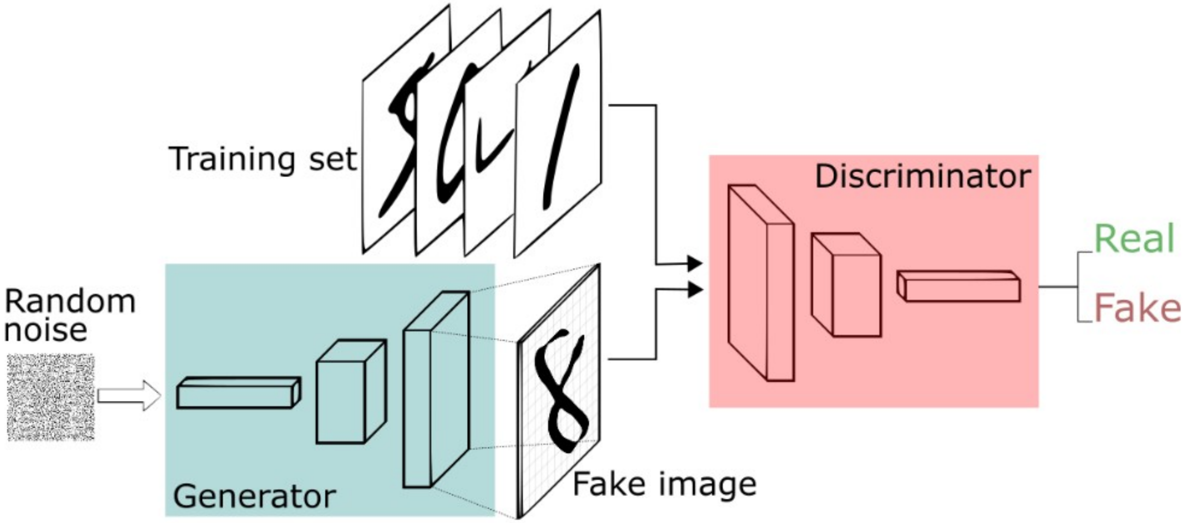


Example


What the computer sees
image classification
82% cat
15% dog
2% hat
1% mug

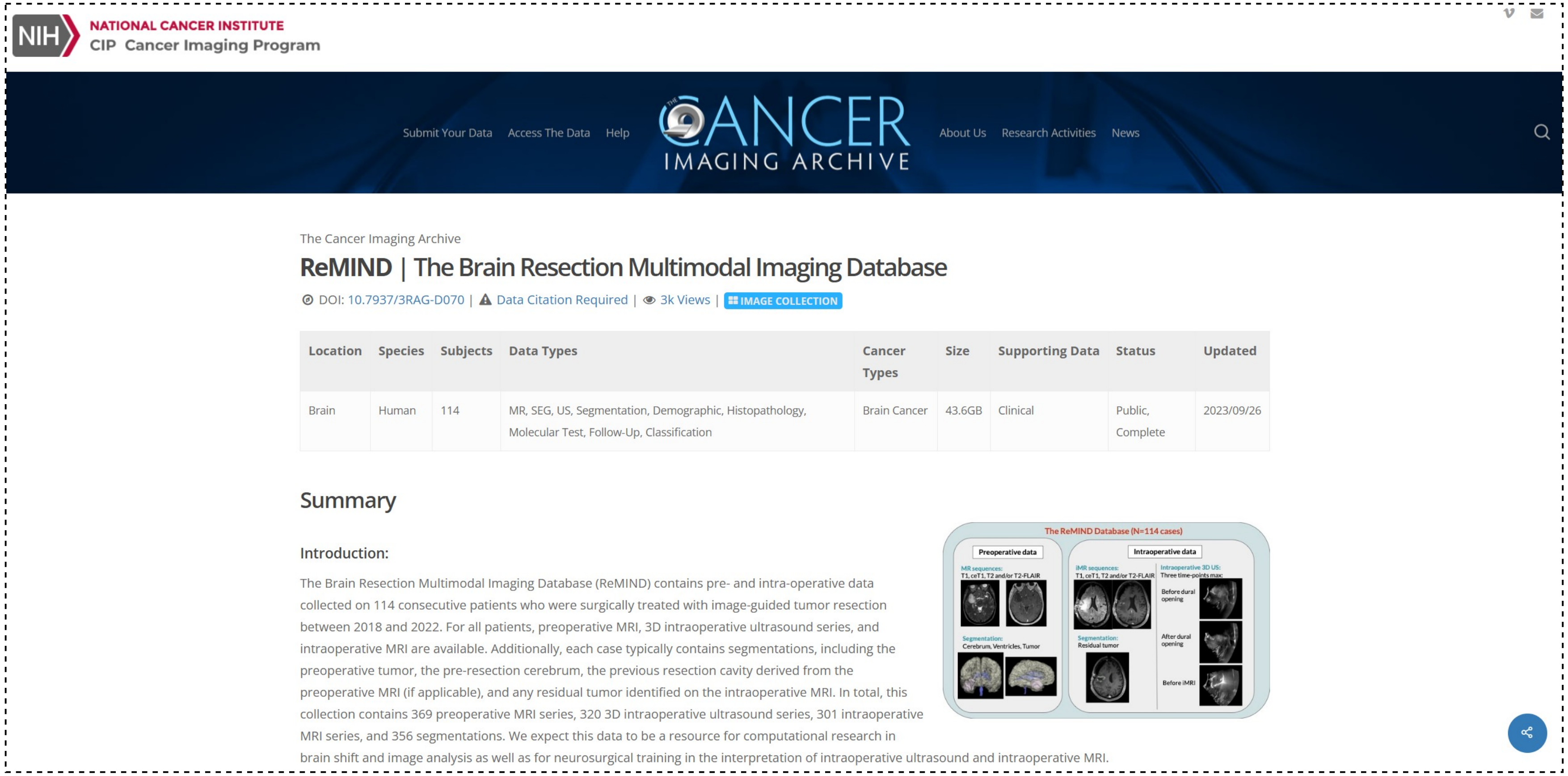
GAN (Generative Adversarial Network)

- ***Generator***: Responsible for "creating things," such as generating a fake image of a cat.
- ***Discriminator***: Responsible for "picking faults," judging whether the image is of a real or fake cat.
- ***Training Process***:
    - The generator repeatedly tries to "fool" the discriminator, improving its ability to create images that look increasingly like real cats.
    - The discriminator, on the other hand, keeps getting smarter, becoming better at distinguishing between real and fake images.
    - Eventually, the generator learns to create images of "cats" that are indistinguishable from real ones.



Proposed Methods

ReMIND Dataset Introduction


NATIONAL CANCER INSTITUTE
CIP Cancer Imaging Program

Submit Your Data   Access The Data   Help   THE CANCER IMAGING ARCHIVE   About Us   Research Activities   News

The Cancer Imaging Archive

ReMIND | The Brain Resection Multimodal Imaging Database

DOI: 10.7937/3RAG-D070 | Data Citation Required | 3k Views | IMAGE COLLECTION

Location	Species	Subjects	Data Types	Cancer Types	Size	Supporting Data	Status	Updated
Brain	Human	114	MR, SEG, US, Segmentation, Demographic, Histopathology, Molecular Test, Follow-Up, Classification	Brain Cancer	43.6GB	Clinical	Public, Complete	2023/09/26

Summary

Introduction:

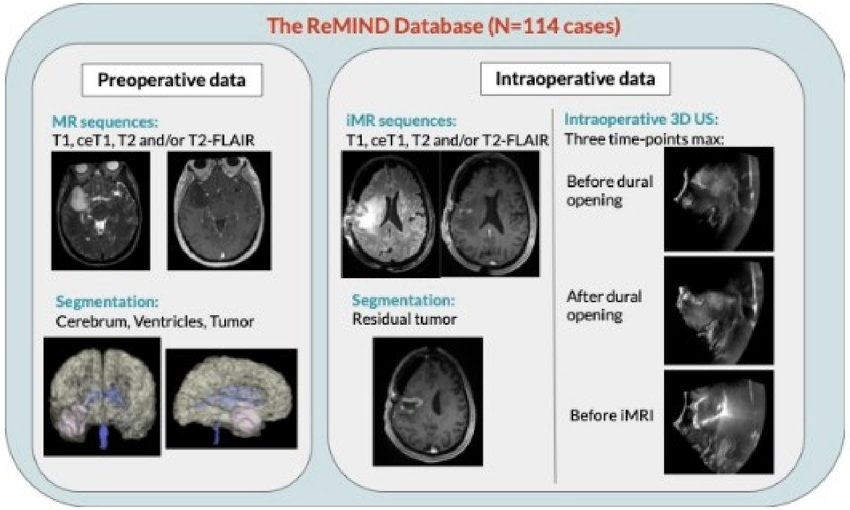
The Brain Resection Multimodal Imaging Database (ReMIND) contains pre- and intra-operative data collected on 114 consecutive patients who were surgically treated with image-guided tumor resection between 2018 and 2022. For all patients, preoperative MRI, 3D intraoperative ultrasound series, and intraoperative MRI are available. Additionally, each case typically contains segmentations, including the preoperative tumor, the pre-resection cerebrum, the previous resection cavity derived from the preoperative MRI (if applicable), and any residual tumor identified on the intraoperative MRI. In total, this collection contains 369 preoperative MRI series, 320 3D intraoperative ultrasound series, 301 intraoperative MRI series, and 356 segmentations. We expect this data to be a resource for computational research in brain shift and image analysis as well as for neurosurgical training in the interpretation of intraoperative ultrasound and intraoperative MRI.

ReMIND dataset (The Brain Resection Multimodal Imaging Database)[1] official website address: https://www.cancerimagingarchive.net/collection/remind/#citations

1. Juvekar, P., Dorent, R., Kögl, F., Torio, E., Barr, C., Rigolo, L., Galvin, C., Jowkar, N., Kazi, A., Haouchine, N., Cheema, H., Navab, N., Pieper, S., Wells, W. M., Bi, W. L., Golby, A., Frisken, S., & Kapur, T. (2023). The Brain Resection Multimodal Imaging Database (ReMIND) (Version 1) [dataset]. The Cancer Imaging Archive. https://doi.org/10.7937/3RAG-D070

ReMIND Dataset Detailed

Data source and time: The ReMIND dataset collected data on **114 consecutive patients** treated with image-guided tumor resection before and during surgery between **2018 and 2022**.



Preoperative data

- MR sequences: Includes 369 MRI series, such as T1-weighted (T1), Contrast-enhanced T1-weighted (ceT1), T2-weighted (T2), T2-FLAIR
- Segmentation: Whole preoperative tumor, Preoperative tumor target (specific for resection), Preoperative brain and ventricles, Previous resection cavity (if applicable)

Intraoperative data

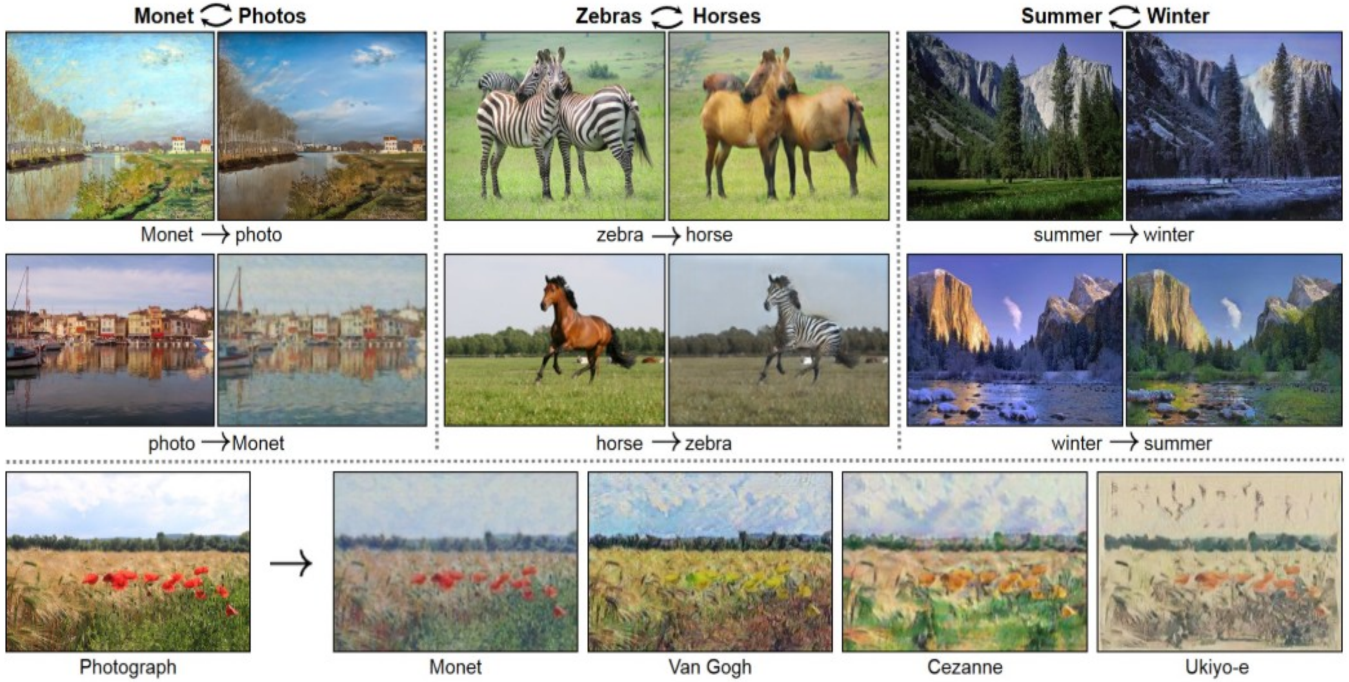
- iMR sequences: Includes 301 MRI series
- Intraoperative 3D US: Before dural opening, After dural opening, Before intraoperative MRI (iMRI)
- Segmentation: Residual tumor (from intraoperative MRI), Brain and ventricles (automatically segmented)

CycleGAN

"Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks" by Jun-Yan Zhu, Taesung Park, Phillip Isola and Alexei A. Efros proposed to solve the Image-to-Image Translation problem of unpaired images.

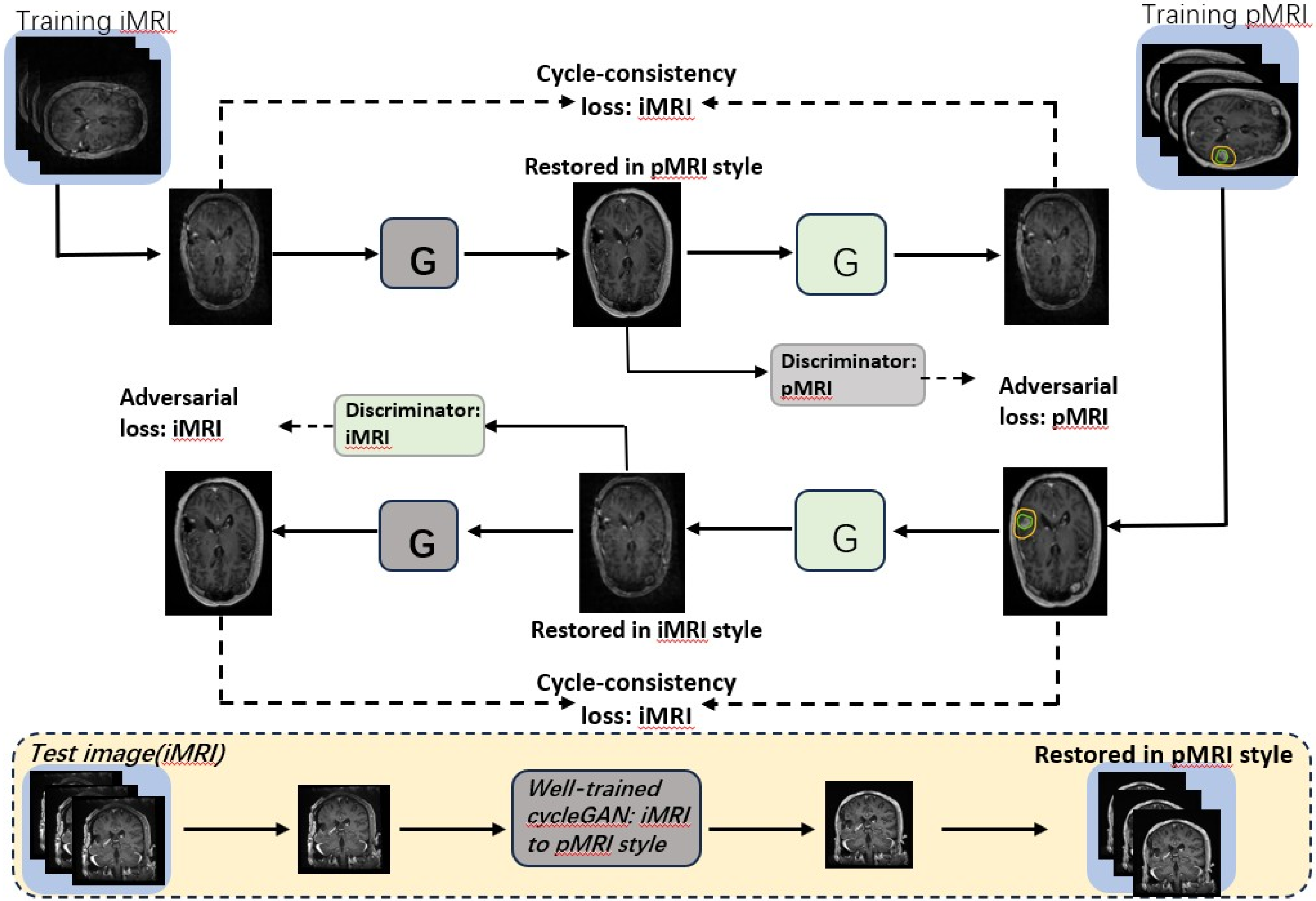
Main contribution:

- Unpaired Image-to-Image Translation
- Cycle Consistency



https://github.com/junyanz/CycleGAN

CycleGAN



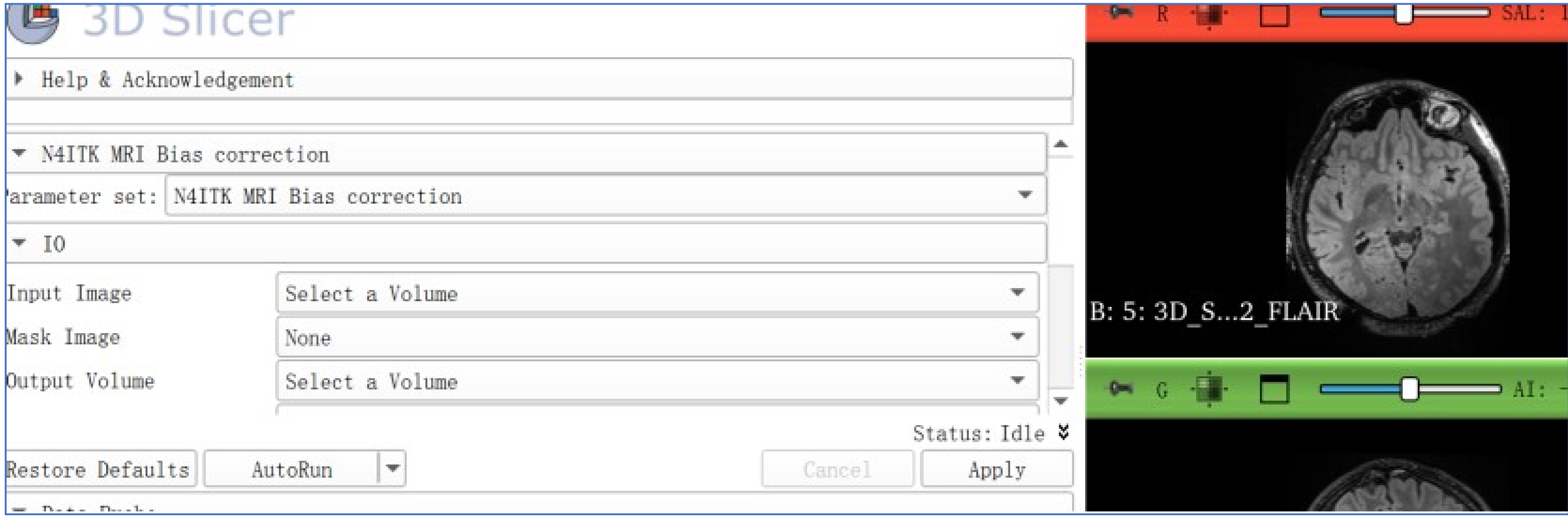
Training iMRI
Training pMRI
Cycle-consistency
loss: iMRI
Restored in pMRI style
G
G
Discriminator:
pMRI
Adversarial
loss: pMRI
Adversarial
loss: iMRI
Discriminator:
iMRI
G
G
Restored in iMRI style
Cycle-consistency
loss: iMRI
Test image(iMRI)
Well-trained
cycleGAN: iMRI
to pMRI style
Restored in pMRI style

Tools and Motivation

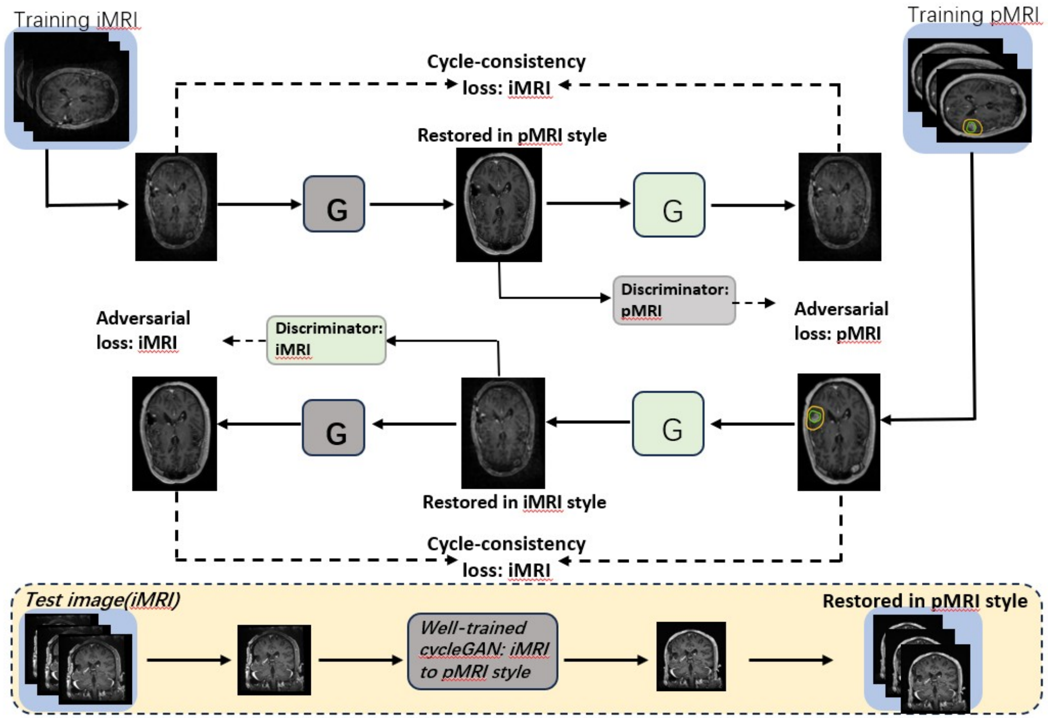
- Tools: 3D Slicer (N4 Bias Field Correction, Gaussian Smooth, General Registration).
- Motivation: Although CycleGAN does not require paired registration, we perform registration and standardization before training to (i) improve cross-case visualization consistency, (ii) ensure strict spatial correspondence between masks and evaluation regions, and (iii) facilitate slice-level display and analysis.

Procedure

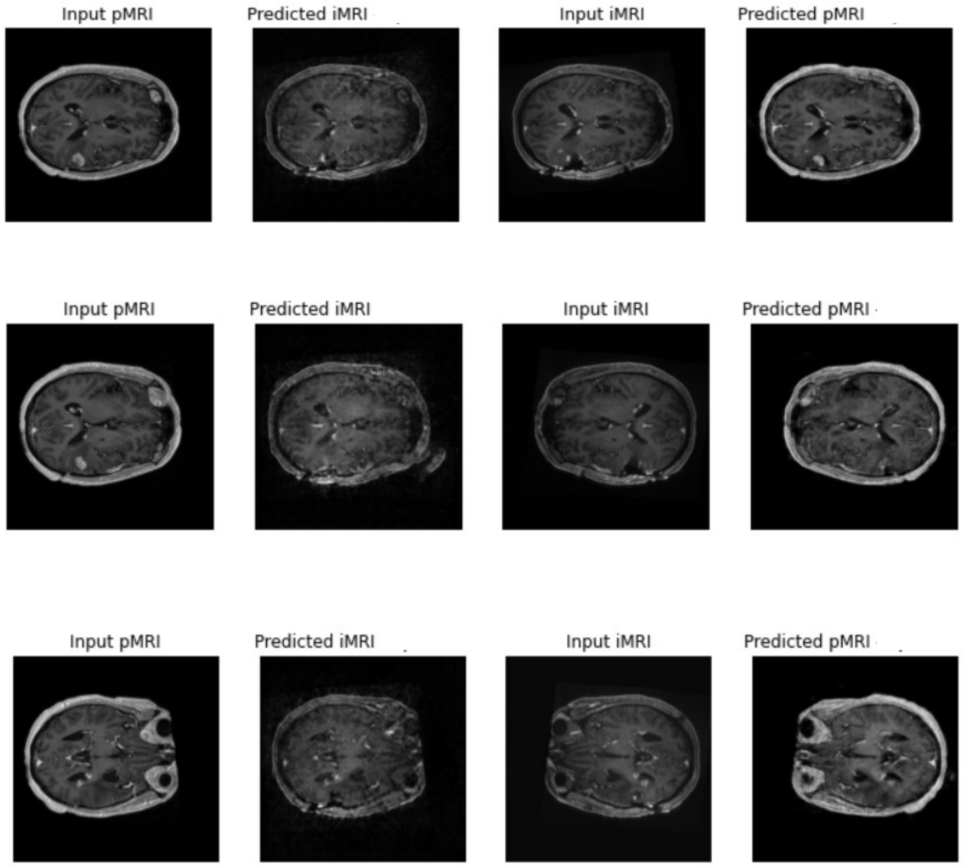
1. Bias Field Correction (N4 Bias Field Correction)
2. Noise Suppression (Gaussian Smooth)
3. Anatomical Registration
4. Mask Creation (Masking)
5. Intensity Normalization (Code)
6. Data Organization



CycleGAN



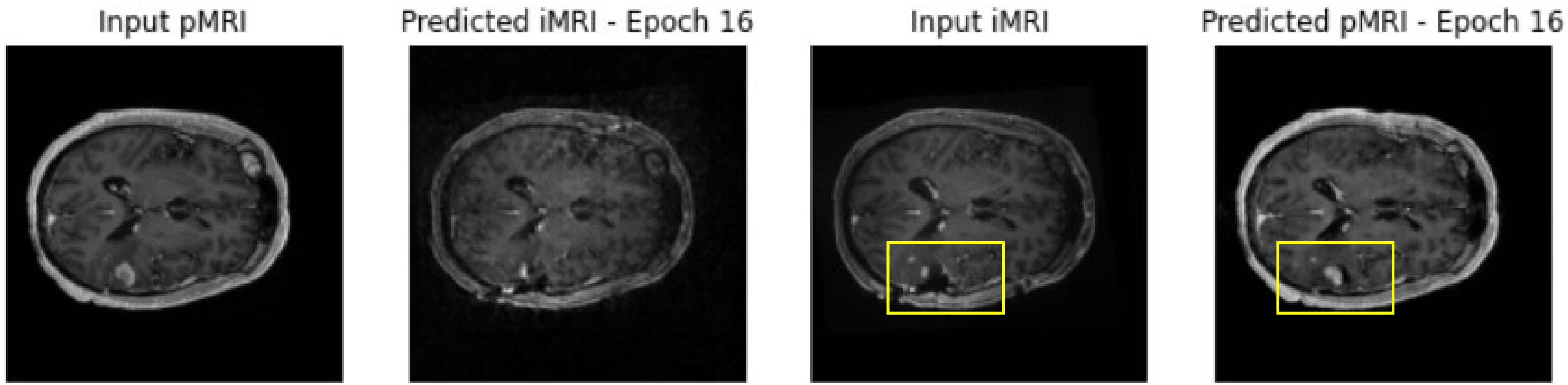
Experimental results during training



1. https://github.com/junyanz/CycleGAN

Problem!!

The tumor removed by surgery was restored □□□□



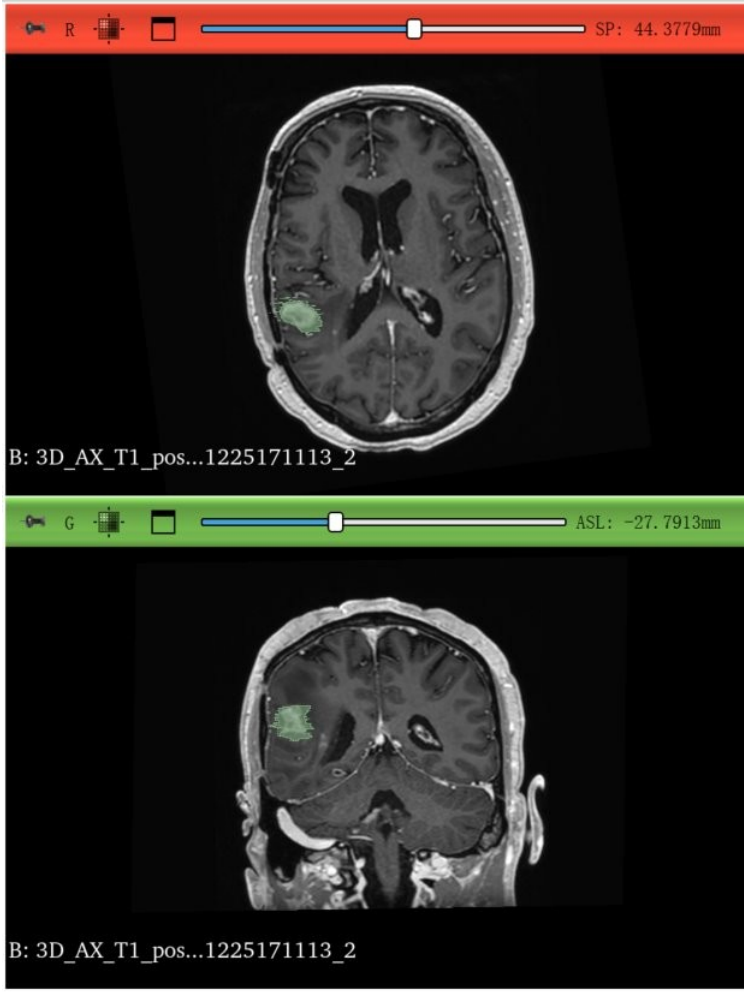
Because the essence of cycleGAN is style transfer, we need to make improvements based on this model and ignore the repaired tumor.

Method 1 — Add "Rules" (Mask-Guided)

- Intuition
  - The model doesn't know where to learn more or less; we explicitly tell it.

- Approach
  - Use pre-op segmentation (registered to intra-op) to create two weight maps:
    - Yellow ring (~5 mm peri-tumoral "critical zone"): high weight to boost sharpness/contrast/boundaries.
    - Green region (resection cavity/tumor area): suppression $\lambda g \approx 0.3$ to prevent hallucinating tumor.
  - Integrate weights into all key losses and critics: adversarial, cycle/reconstruction, perceptual/style, and discriminator scoring.

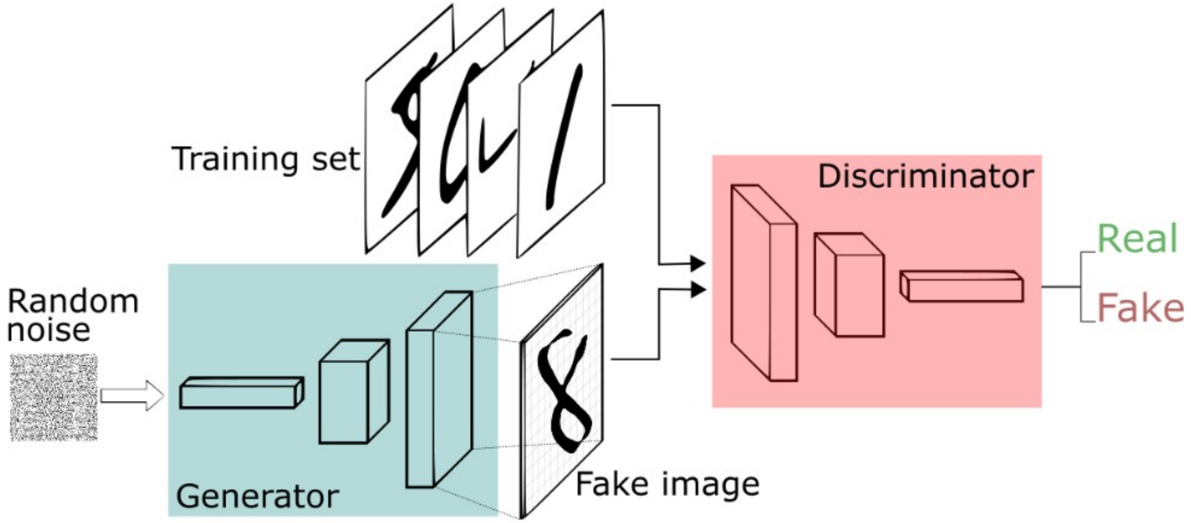
- Outcome
  - Optimization focuses on clinically critical anatomy; false tumor regeneration in the green region is strongly reduced.



GAN (Generative Adversarial Network)

- ***Generator***: Responsible for "creating things," such as generating a fake image of a cat.
- ***Discriminator***: Responsible for "picking faults," judging whether the image is of a real or fake cat.

LOSS ?



LOSS

1. ***Adversarial Loss***

$$\mathcal{L}_{\text{adv}}^{\text{std}} = -\log D(G(x_{\text{intra}}))$$

$$\mathcal{L}_{\text{adv}}^{\text{mask}} = -(\text{mask}_{\text{yellow}} \cdot \log D(G(x_{\text{intra}})) + \lambda_{\text{green}} \cdot \text{mask}_{\text{green}} \cdot \log D(G(x_{\text{intra}})))$$

LOSS

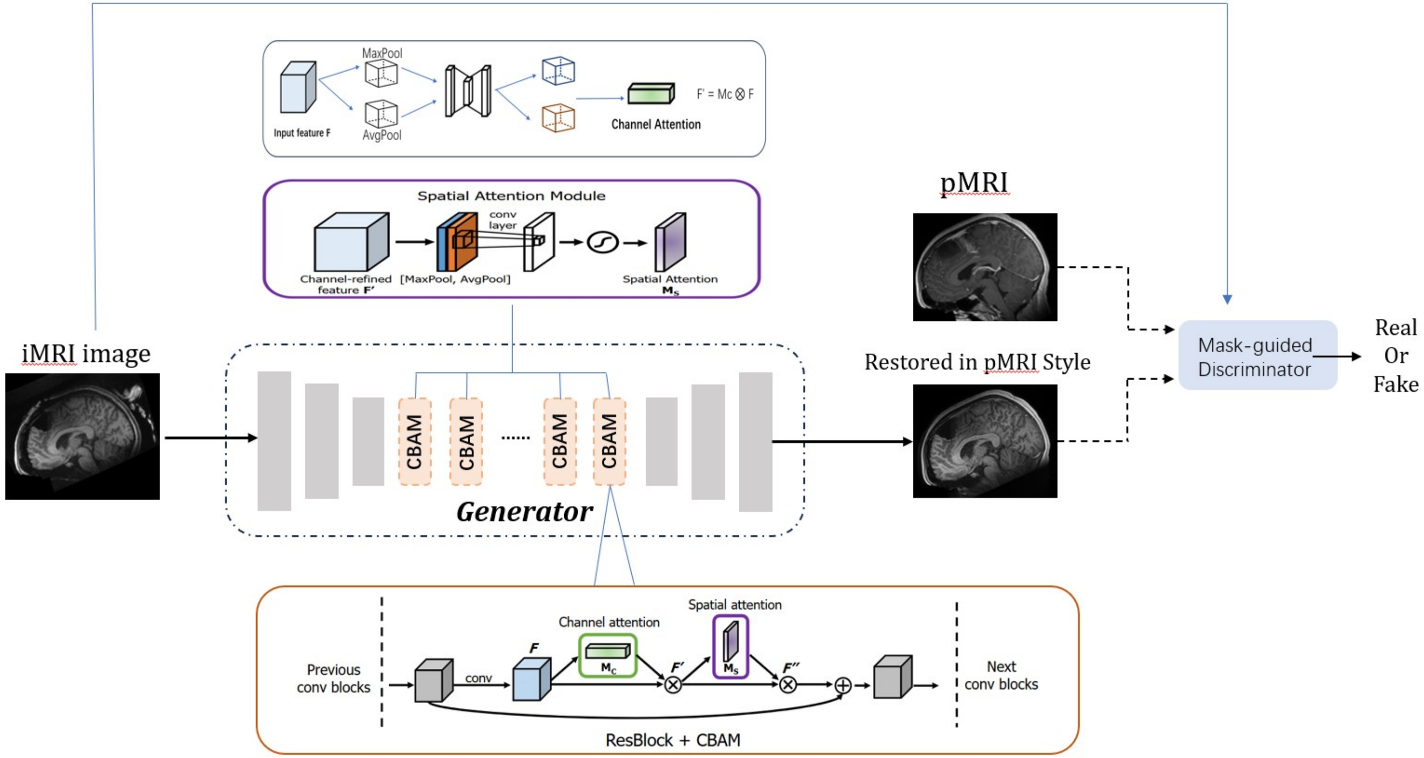
2. Reconstruction Loss

$$\mathcal{L}_{\text{recon}}^{\text{std}} = \|G(x_{\text{intra}}) - x_{\text{intra}}\|$$

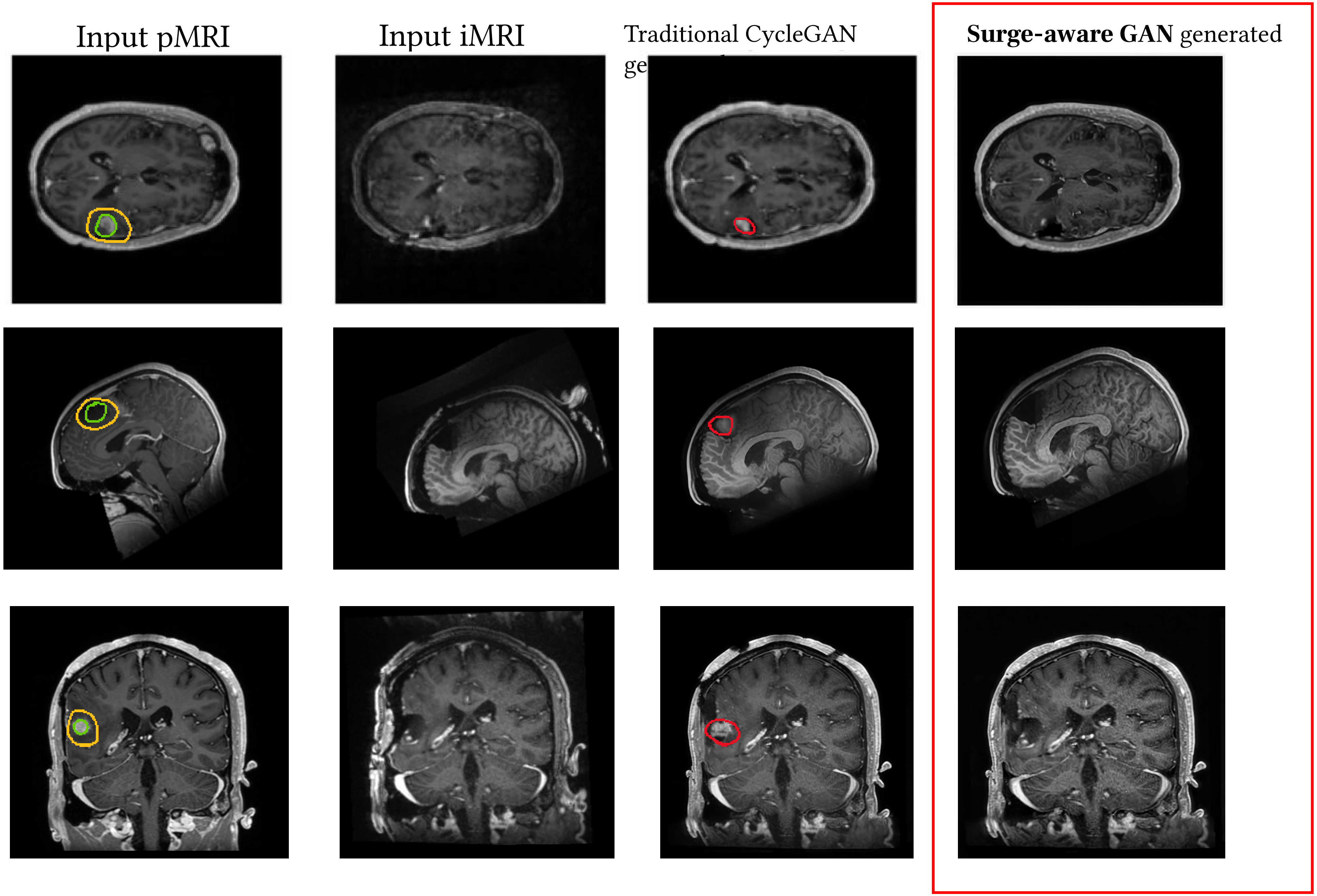
$$\mathcal{L}_{\text{recon}}^{\text{mask}} = \text{mask}_{\text{yellow}} \cdot \|G(x_{\text{intra}}) - x_{\text{intra}}\|_1 + \lambda_{\text{green}} \cdot \text{mask}_{\text{green}} \cdot \|G(x_{\text{intra}}) - x_{\text{intra}}\|$$

Method 2 — Add “Attention” (CBAM Learns Where to Look)

- Rules help, but the network must also allocate attention efficiently.
- Insert CBAM after each encoder/decoder residual block (channel + spatial attention).
- Feed the masks as hints to attention: more focus in the yellow ring, conservative in the green region.
- Smarter feature extraction; more localized quality gains; smoother, more natural boundaries.



Results and Metrics Analysis


Input pMRI
Input iMRI
Traditional CycleGAN
Surge-aware GAN generated

Performance

Table 3: Performance Metrics Across Training Epochs

Metric	Epoch 10	Epoch 50	Epoch 100	Optimal Value
PSNR (dB)	33.0	35.5	37.2	>37.0
SSIM	0.85	0.89	0.91	>0.90
NT-SSIM	0.86	0.89	0.90	>0.90
TAPI	0.045	0.018	0.002	<0.001
LPIPS-NT	0.04	0.04	0.04	<0.05
BNS	6.5	5.8	5.4	<3.0

- PSNR (↑): "Signal-to-noise" for images. Higher PSNR means the enhanced image is closer to the clean reference in brightness and details, with less noise. Above ~37 dB is typically considered very clean.

- SSIM (↑): "Structure preservation" score. It checks not only brightness but also contrast and spatial relationships of textures. Values >0.90 indicate that major structures and texture relations are well preserved.

- NT-SSIM (↑): Non-Tumor SSIM. The same idea as SSIM but computed only in healthy tissue. Think of it as "do surgeons' normal brain structures remain unchanged?" Values >0.90 mean normal anatomy is well preserved.

Performance

Table 3: Performance Metrics Across Training Epochs

Metric	Epoch 10	Epoch 50	Epoch 100	Optimal Value
PSNR (dB)	33.0	35.5	37.2	>37.0
SSIM	0.85	0.89	0.91	>0.90
NT-SSIM	0.86	0.89	0.90	>0.90
TAPI	0.045	0.018	0.002	<0.001
LPIPS-NT	0.04	0.04	0.04	<0.05
BNS	6.5	5.8	5.4	<3.0

- PSNR / SSIM / NT-SSIM (higher is better): Do we preserve clarity and anatomical structure well, especially in non-tumor regions?
- LPIPS-NT (lower is better): Does it look perceptually similar to the reference to a human observer in non-tumor regions?
- TAPI (lower is better): Are tumor/cavity areas kept intact without being mistakenly "restored" or filled in?
- BNS (lower is better): Are boundaries natural, without halos or over-sharpening around tumor/cavity edges?

Performance

Table 3: Performance Metrics Across Training Epochs

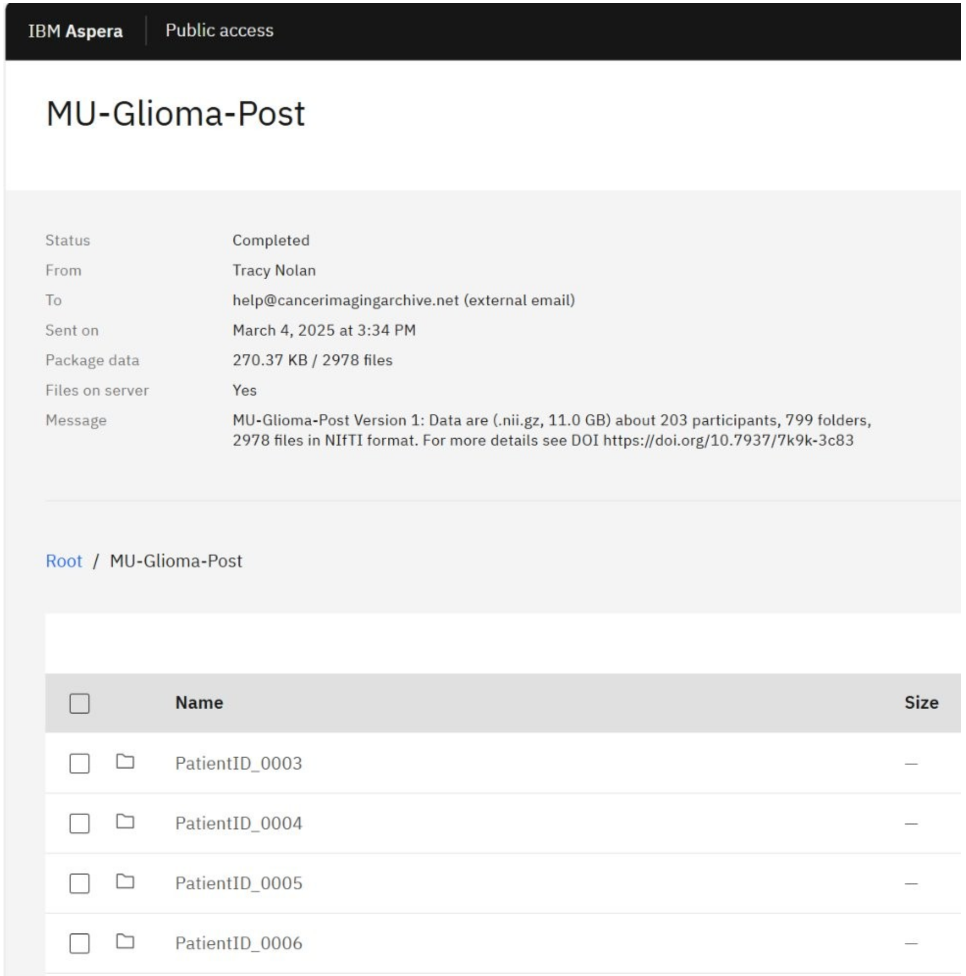
Metric	Epoch 10	Epoch 50	Epoch 100	Optimal Value
PSNR (dB)	33.0	35.5	37.2	>37.0
SSIM	0.85	0.89	0.91	>0.90
NT-SSIM	0.86	0.89	0.90	>0.90
TAPI	0.045	0.018	0.002	<0.001
LPIPS-NT	0.04	0.04	0.04	<0.05
BNS	6.5	5.8	5.4	<3.0

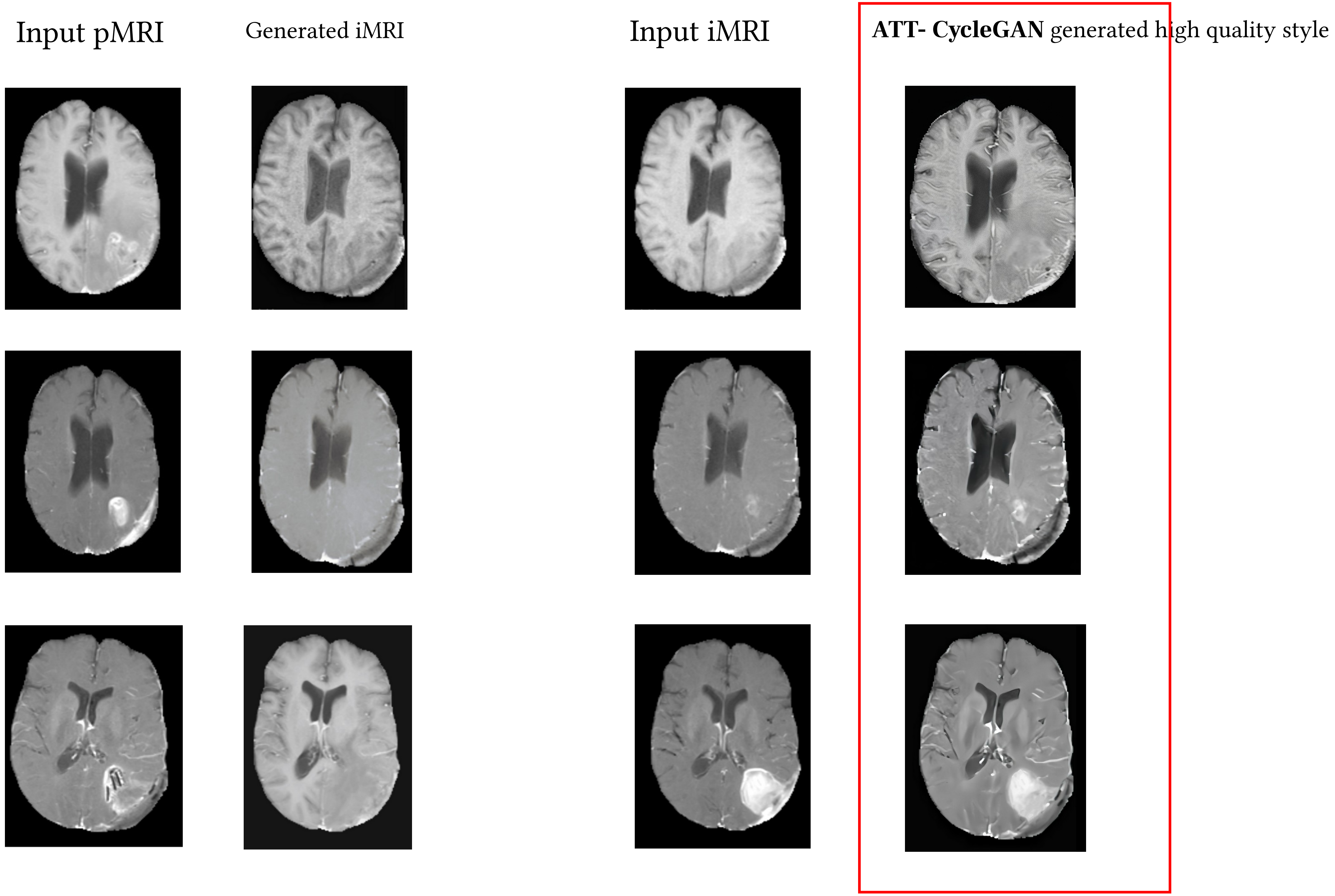
Table 4: Main Test Results (patient-level mean ± SD; N=21). ↑ higher is better; ↓ lower is better.

Method	NT-SSIM ↑	LPIPS-NT ↓	TFP ↓	TAPI ↓	BNS ↓
Full (ours)	0.905 ± 0.012	0.040 ± 0.006	0.022 ± 0.004	0.0022 ± 0.0010	3.2 ± 0.5
UNet Denoise (Ii only, L1)	0.892 ± 0.015	0.052 ± 0.008	0.030 ± 0.005	0.0110 ± 0.0040	4.1 ± 0.6
Self-supervised N2V/N2S	0.887 ± 0.017	0.058 ± 0.009	0.033 ± 0.006	0.0130 ± 0.0050	4.4 ± 0.7

Testing ---- MU-Glioma-Post Dataset

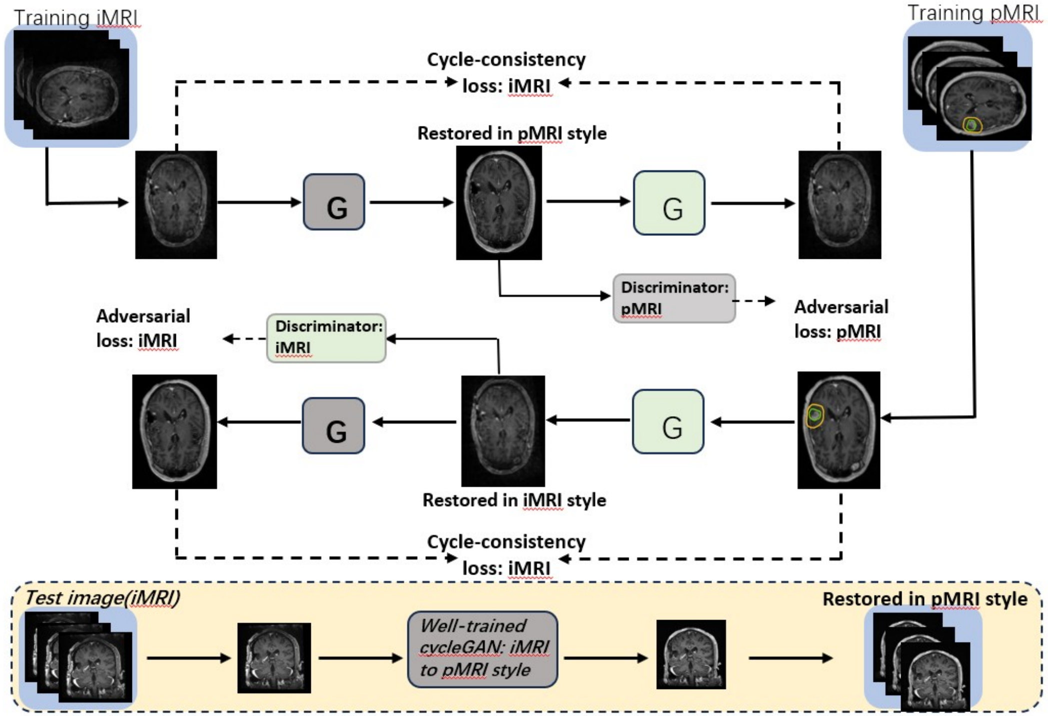
- Data source and size
  - Hosted on TCIA (NIIfTI, .nii.gz)
  - Version 1: ~203 participants, 799 folders, 2,978 files, ~11.0 GB
  - DOI: https://doi.org/10.7937/7k9k-3c83
- Imaging and labels
  - Sequences: T1, T1ce, T2, FLAIR (some cases ± DWI/ADC, SWI)
  - Annotations: residual/recurrent tumor; postoperative changes (resection cavity, hemorrhage, nonspecific enhancement)
- Key characteristic
  - Post-resection MRI with low clarity/SNR
  - Common artifacts: motion, metal-clip, thicker slices → fuzzy small lesions, unstable registration
- Intended use
  - Segmentation of residual/recurrent tumor
  - Recurrence vs pseudoprogression classification
  - Baseline: bias correction + normalization + light denoising; 3D nnU-Net with coarse-to-fine and cleanup
- Access
  - Research use; de-identified data via TCIA/Aspera download




Input pMRI
Generated iMRI
Input iMRI
ATT- CycleGAN generated high quality style

Conclusion

Our surgery-aware enhancement framework improves global image quality (PSNR/SSIM/NT-SSIM) while markedly reducing safety-critical errors (TAPI, BNS, LPIPS-NT). It maintains consistent advantages on the external MU-Glioma-Post test set, demonstrating clinical robustness and scalability.



Future Work

Replace fixed λ with learnable/adaptive weights, possibly driven by uncertainty or case difficulty, to reduce hyperparameter sensitivity.