

MRI Explainer for Pediatric Brain Tumors

Yinuo Zhang, Dan Ding

MSML 640 – Project

December 10, 2025

Motivation

- Pediatric brain tumor MRIs are visually complex and difficult to interpret for non-experts.
- Typical ML work focuses on segmentation accuracy only.
- I wanted a pipeline that:
 - Segments suspected tumor regions,
 - Shows *where* the model is focusing (saliency),
 - Generates a short, human-readable explanation per slice.
- Personal goal: create something I can show to a pediatric oncology doctor (my girlfriend) to support communication about MRI images.

Task Definition

Input

3D brain MRI volume from BraTS-PED (e.g., T2-F / T2-FLAIR, NIfTI format).

Outputs (per 2D slice)

- ① Binary segmentation mask (tumor vs. background).
- ② Saliency / Grad-CAM heatmap over the slice.
- ③ Short text caption describing size and laterality of the highlighted region.

Data: BraTS-PED Overview

- Dataset: ASNR-MICCAI BraTS 2023 Pediatric (BraTS-PED).
- Each subject has multiple modalities:
 - T2-F (T2-FLAIR): *-t2f.nii.gz
 - T2-weighted: *-t2w.nii.gz
 - T1 native / contrast: *-t1n.nii.gz, *-t1c.nii.gz
 - Segmentation labels: *-seg.nii.gz
- In this project I primarily use **T2-F** as input and the whole-tumor mask as target.
- Training split:
 - Train slices: 3945 (after preprocessing)
 - Val slices: 1377 from a subject-level split

Preprocessing: 3D → 2D Slices

- Convert 3D NIfTI volumes to 2D axial slices.
- Implementation: `src/data/prepare_slices.py`
- Steps:
 - ① Load T2-F volume and segmentation (`*-t2f.nii.gz`, `*-seg.nii.gz`).
 - ② Normalize volume per-subject (zero-mean, unit-variance, then rescale to [0, 1]).
 - ③ Slice along axis 2 (axial).
 - ④ Resize to 256×256 (bilinear for image, nearest for mask).
 - ⑤ Filter slices with very tiny masks (`--min-area 1` in final version).
- Resulting dataset:
 - PNG images in `images/`, masks in `masks/`.
 - Convenient for standard PyTorch Dataset/DataLoader.

Model: 2D U-Net

- Architecture: lightweight 2D U-Net implemented from scratch.
- Input: $1 \times 256 \times 256$ grayscale slice (T2-F).
- Output: $1 \times 256 \times 256$ logit map (tumor vs. background).
- Channel configuration:
 - Encoder: $32 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 512$.
 - Decoder: symmetric upsampling with skip connections.
- Loss:
 - BCE with logits + soft Dice loss.
- Optimization:
 - Adam, learning rate $1e-3$, batch size 4 on CPU.
 - Training epochs: 10.

Interpretability: Grad-CAM

- Use Grad-CAM on a late encoder block to highlight regions that contribute most to the predicted mask.
- Implementation: `src/interpret/gradcam.py`.
- For each slice:
 - ① Forward pass to get segmentation logits.
 - ② Compute Grad-CAM over encoder feature maps for the mean predicted probability.
 - ③ Normalize and overlay heatmap on top of the original slice.
- This gives a rough sense of *where* the model is focusing compared to its predicted mask.

Caption Generation

- Simple rule-based captioning in `src/caption/templates.py`.
- Inputs:
 - Predicted binary mask per slice.
 - Optional saliency map.
- Extract features from the mask:
 - Relative area of highlighted region.
 - Approximate centroid (left vs. right vs. center).
 - Whether saliency is concentrated.
- Example caption pattern:

"This slice shows a moderate highlighted area on the right side. The overlay marks tissue the model considers unusual."
- Goal: short, intuitive description that matches the visual overlay.

Training Setup and Metric

- Train/validation split:
 - Subject-level split from BraTS-PED training set.
 - No subject overlap between train (5322 slices) and val (1377 slices).
- Training configuration (final run):
 - 2D U-Net with base channels 32.
 - Optimizer: Adam, learning rate $1e-3$.
 - Batch size: 4, epochs: 10, device: CPU.
 - Loss: BCE with logits + soft Dice loss.
- Metric: Dice similarity coefficient on validation slices

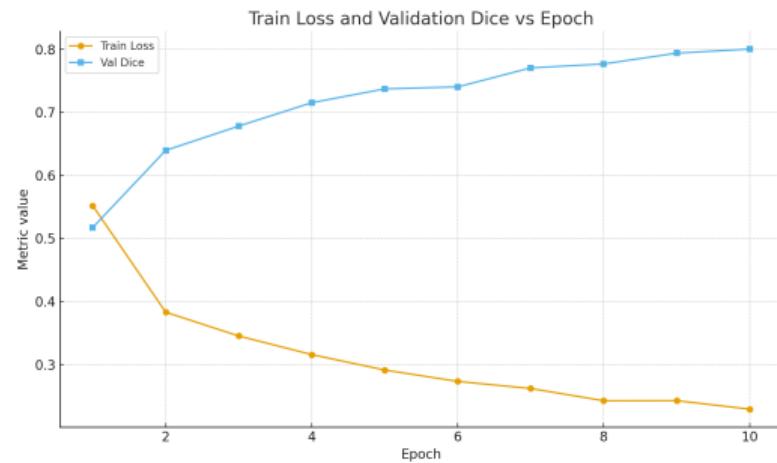
$$\text{Dice}(P, G) = \frac{2|P \cap G|}{|P| + |G|}.$$

- Best model:
 - Achieved mean validation Dice of **0.8002** on 1377 slices.
 - Train loss decreased from 0.55 → 0.23 over 10 epochs.

Quantitative Results

- Training dynamics:
 - Train loss: $0.55 \rightarrow 0.23$ (epochs 1–10).
 - Val Dice: $0.52 \rightarrow 0.80$.
- Final performance:
 - Mean slice-wise val Dice: **0.8002**.
 - Stable by epoch 9–10, indicating limited overfitting.
- Interpretation:
 - For a small 2D U-Net trained on T2-F slices only, a Dice of ≈ 0.80 suggests reasonably accurate whole-tumor segmentation on the held-out subjects.

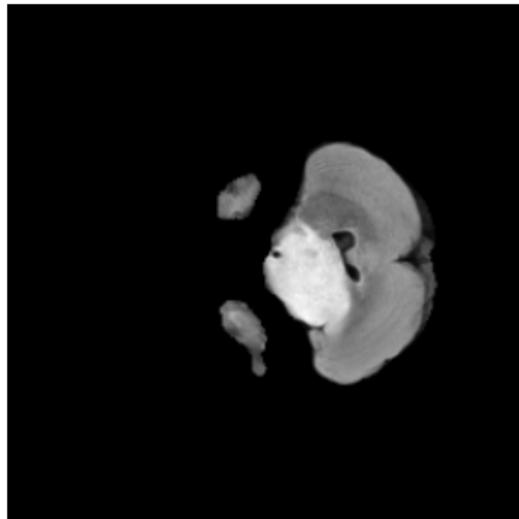
Train/Val Curves



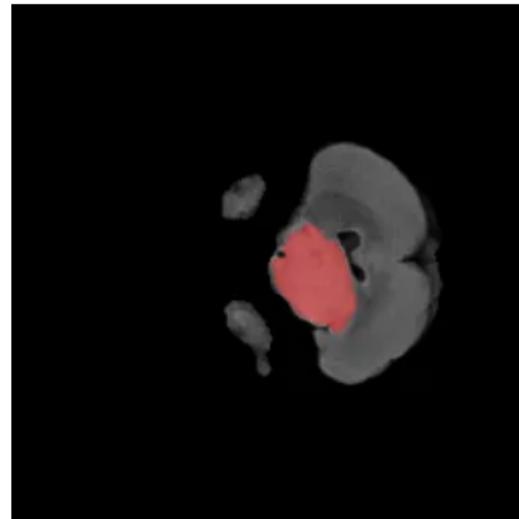
(Train loss and validation Dice vs. epoch from eval/train_log.csv)

Qualitative Examples (1)

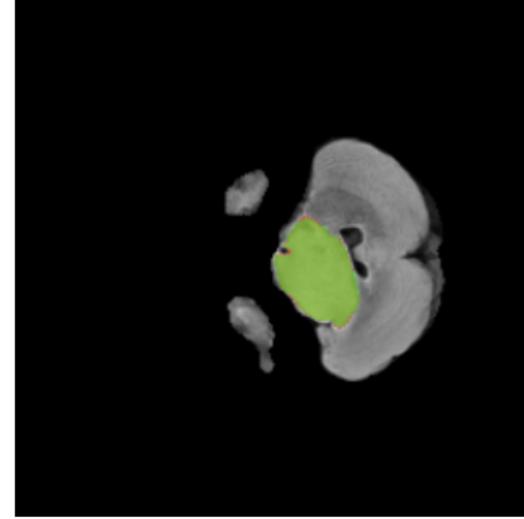
Input slice



Segmentation



Mask + Grad-CAM

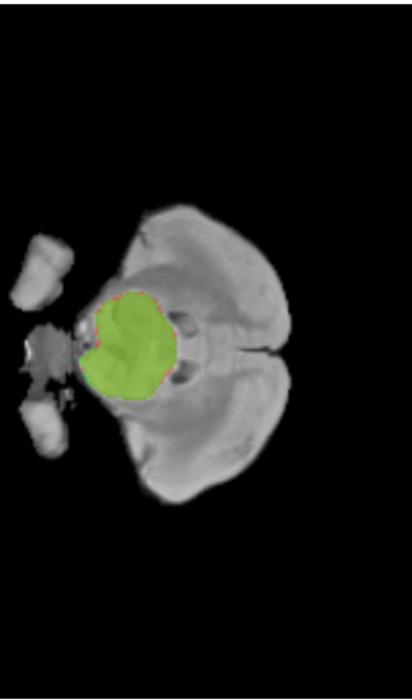


Caption example:

"This slice shows a moderate highlighted area on the right side. The overlay marks tissue the model considers unusual."

Qualitative Examples (2)

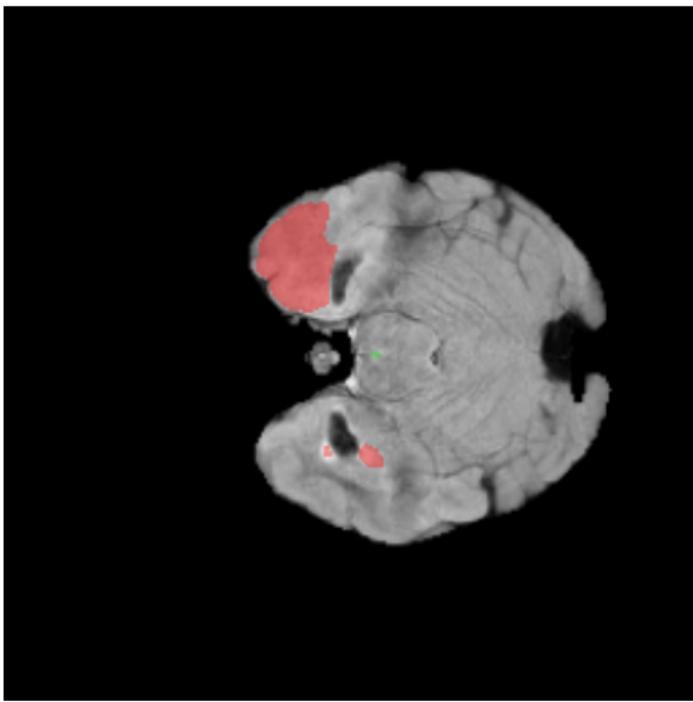
Good example: BraTS-PED-00079-000,
slice 29



- Predicted mask closely overlaps the ground-truth tumor.
- Grad-CAM heatmap is concentrated over the tumor core.
- Generated caption correctly describes size and side of the highlighted region.
- Dice ≈ 0.981 .

Qualitative Examples (3)

Failure example: BraTS-PED-00108-000,
slice 42



- Clear mismatch between prediction and ground truth: under-segmentation, over-segmentation, or shape error.
- Grad-CAM heatmap partially focuses on non-tumor structures.
- Caption is less trustworthy here because the underlying mask is inaccurate.

Interactive Demo

- Implemented a Gradio web UI (`demo/app.py`):
 - Upload a `.nii` / `.nii.gz` MRI volume.
 - Optional: upload a trained checkpoint.
 - Slider to browse axial slices.
 - Display:
 - Original slice.
 - Segmentation mask overlay.
 - Saliency (Grad-CAM) overlay.
 - Generated caption.
- This makes it easy to sit together with someone (e.g., a pediatric oncologist) and look at images interactively.

Limitations

- 2D model only:
 - Ignores 3D context across slices.
- Simple preprocessing:
 - No advanced intensity harmonization or skull stripping.
- Grad-CAM is coarse:
 - Heatmaps are approximate and sometimes highlight non-tumor regions.
- Captions are template-based:
 - Limited expressiveness; not personalized to specific clinical questions.

Future Directions

- Modeling:
 - 3D U-Net or 2.5D context (stack of slices).
 - Multi-modal input (T2-F + T1c, etc.).
- Interpretability:
 - Try alternative methods (e.g., integrated gradients, occlusion).
- Captioning:
 - Replace templates with a small vision-language model.
 - Condition on user role (doctor vs. parent).
- Human feedback:
 - Collect structured feedback from pediatric oncologists on which visualizations and captions are most helpful.

Takeaways

- Built an end-to-end MRI explainer:
 - Data preprocessing → U-Net segmentation → Grad-CAM → captions → Gradio demo.
- Achieved reasonable segmentation performance on BraTS-PED with a simple 2D model.
- Combined visual overlays and short text explanations to make results easier to understand for non-experts.
- This serves as a starting point for richer, clinician-facing tools for explaining pediatric brain tumor MRIs.

Questions?