# MProtoNet: A Case-Based Interpretable Model for Brain Tumor Classification with 3D Multi-parametric Magnetic Resonance Imaging

Yuanyuan Wei[1,2], Roger Tam[2*], Xiaoying Tang[1*]

[1] Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen, China

[2] School of Biomedical Engineering, The University of British Columbia, Vancouver, Canada

## INTRODUCTION

### Motivation

- CNNs often take undesired shortcuts
- Medical settings emphasize interpretability
- Post-hoc explainers are unreliable
- Concept-based models need predefinitions
- Case-based models identify prototypes

### Related Work

- ProtoPNet: pioneering case-based model
  - IAIA-BL: require fine-grained annotations
- XProtoNet: no evaluation of interpretability

### Contributions

- Extend ProtoPNet to 3D mpMRIs
  - 3D ResNet as the backbone
  - Online data augmentation during training
- New attention module
  - Soft masking: sharpen attention maps
  - Online-CAM loss: assist localization
- Statistically significant improvements in interpretability metrics
  - Correctness and localization coherence
  - Without fine-grained annotations

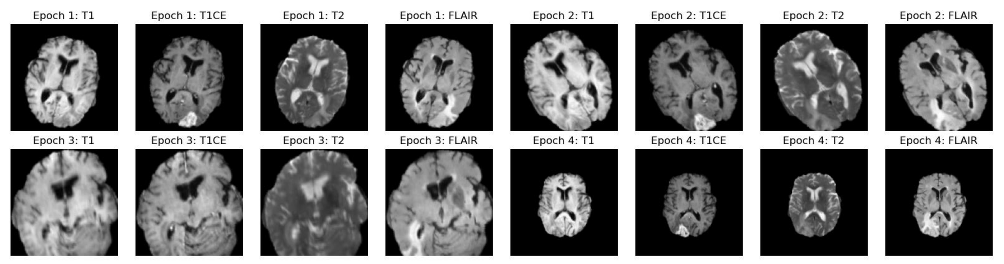
## DATASET

### BraTS 2020

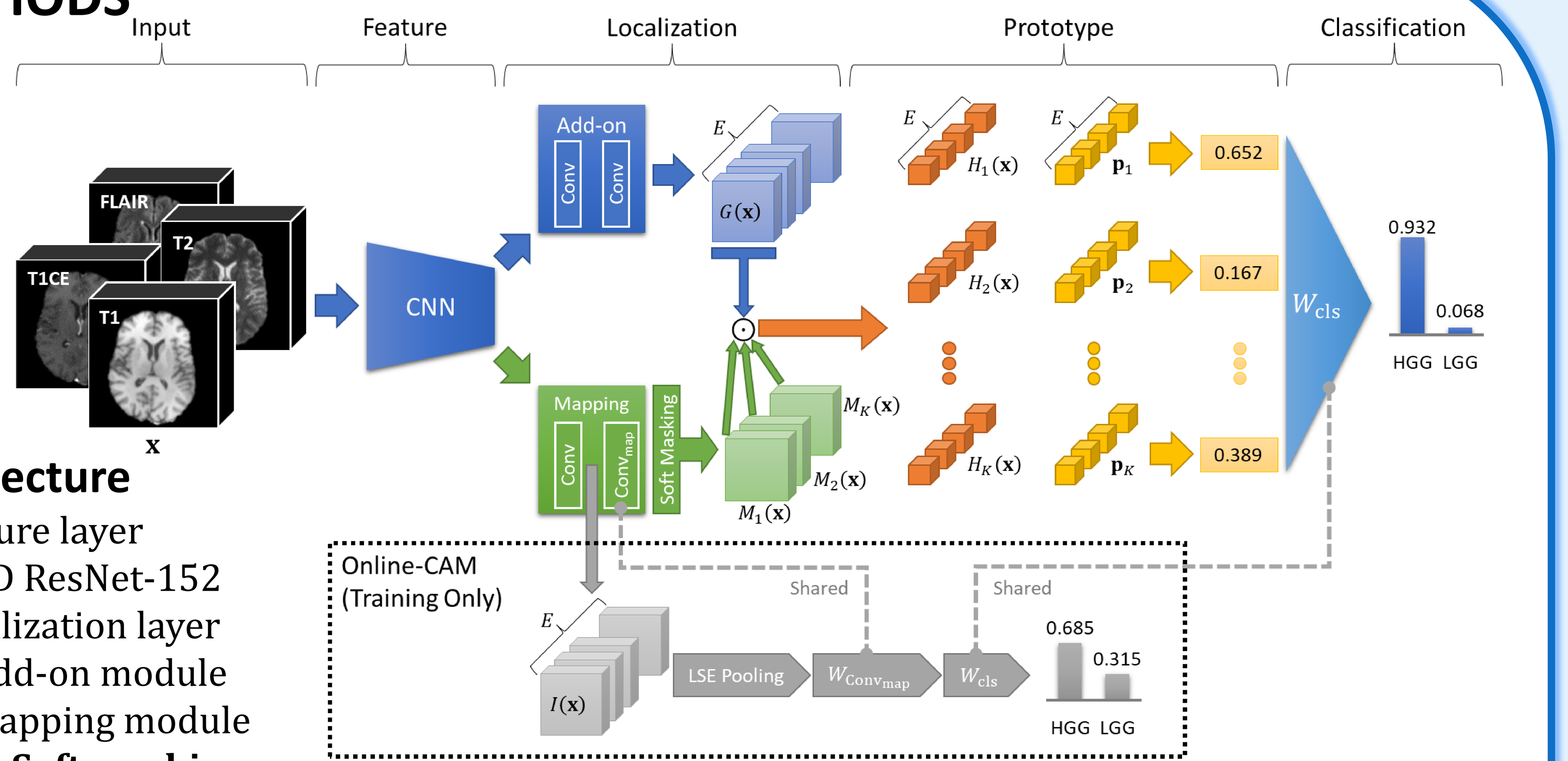
- 369 subjects: 293 with HGG, 76 with LGG
- Four modalities: T1, T1CE, T2, FLAIR
- Tumor sub-regions labeled
  - Unify as the whole tumor for evaluation

### Data Augmentation

- Online data augmentation follows nnU-Net



## METHODS




### Architecture

- Feature layer
  - 3D ResNet-152
- Localization layer
  - Add-on module
  - Mapping module
    - **Soft masking**
    - **Online-CAM loss**
- Prototype layer
- Classification layer

### Soft Masking

- Learn more accurate attention maps
- Reduce irrelevant background areas

$$M_k(\mathbf{x}) = \frac{1}{1+\exp\left(-\omega\left(M_k^0(\mathbf{x})-\sigma\right)\right)}$$

### Online-CAM Loss

- Directly utilize image-level labels
- Use LSE pooling to obtain class predictions directly from intermediate features

$$\mathrm{LSE}(I_e(\mathbf{x})) = \frac{1}{r}\log\left[\frac{1}{16\times16\times12}\sum_{h,w,d}\exp\left(r\cdot I_{e,h,w,d}(\mathbf{x})\right)\right], e\in\{1\ldots E\}$$

$$p_{\mathrm{OC}}(\mathbf{x}_i) = \mathrm{Softmax}\left(\mathrm{LSE}(\mathbf{I}(\mathbf{x}_i))W_{\mathrm{Conv_{map}}}W_{\mathrm{cls}}\right)$$

## EXPERIMENTS

### Evaluation Metrics

- Classification: balanced accuracy (BAC)
- Interpretability: incremental deletion score (IDS)
  - **Correctness** of reflecting the decision-making process
  - The normalized area under the incremental deletion curve
- Interpretability: activation precision (AP)
  - **Localization coherence** with the fine-annotated label

$$\mathrm{AP} = \frac{\left|H(\mathbf{x})\cap T\left(\mathrm{UpSample}(M(\mathbf{x}))\right)\right|}{\left|T\left(\mathrm{UpSample}(M(\mathbf{x}))\right)\right|}$$

$H(\cdot)$: fine-annotated label, $M(\cdot)$: activation map, $T(\cdot)$: threshold function

### Experimental Setup

- 5-fold cross-validation
- Training stages
  - Training of layers before the classification layer
  - Prototype reassignment
  - Training of the classification layer
- Models compared
  - CNN (with GradCAM): feature layer + add-on module + global average pooling + classification layer
  - ProtoPNet: without the mapping module and its losses
  - XProtoNet: without soft masking and the online-CAM loss

## RESULTS

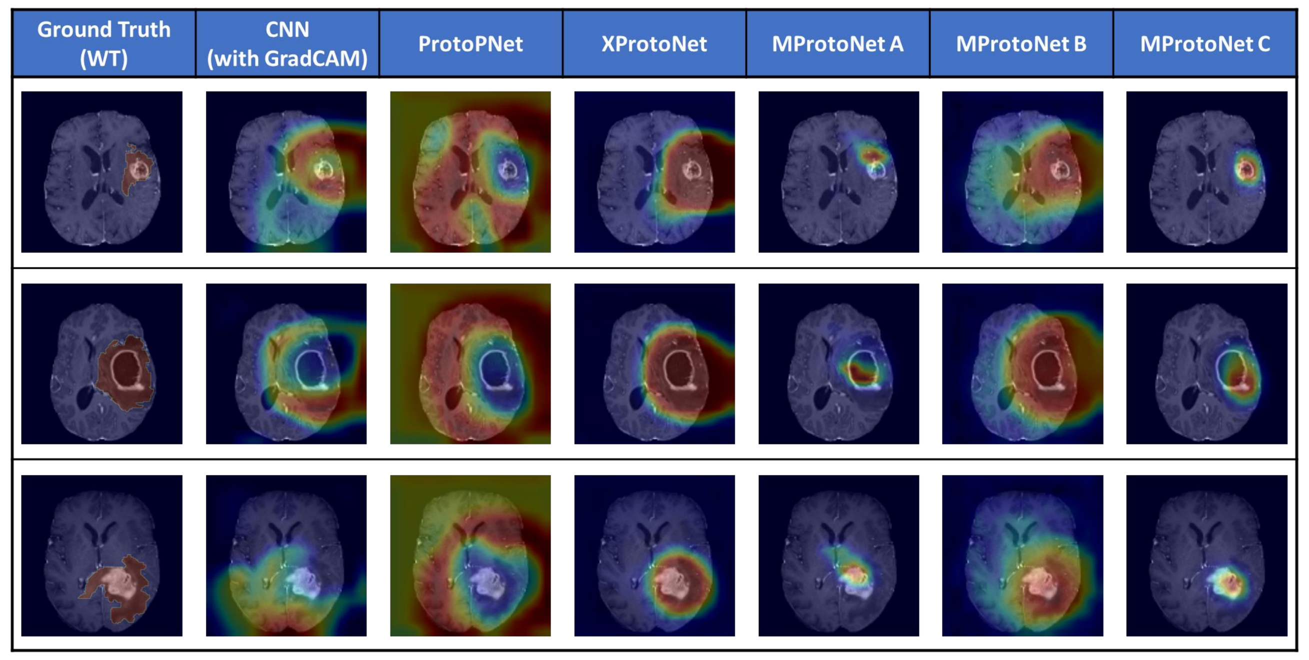
### Summary

- Classification performance: BAC
  - No statistically significant differences among all models
- Interpretability performance: IDS & AP
  - MProtoNet C achieves the best performance in the interpretability metrics of IDS & AP
  - Soft masking and the online-CAM loss are both important for statistically significant improvements over XProtoNet

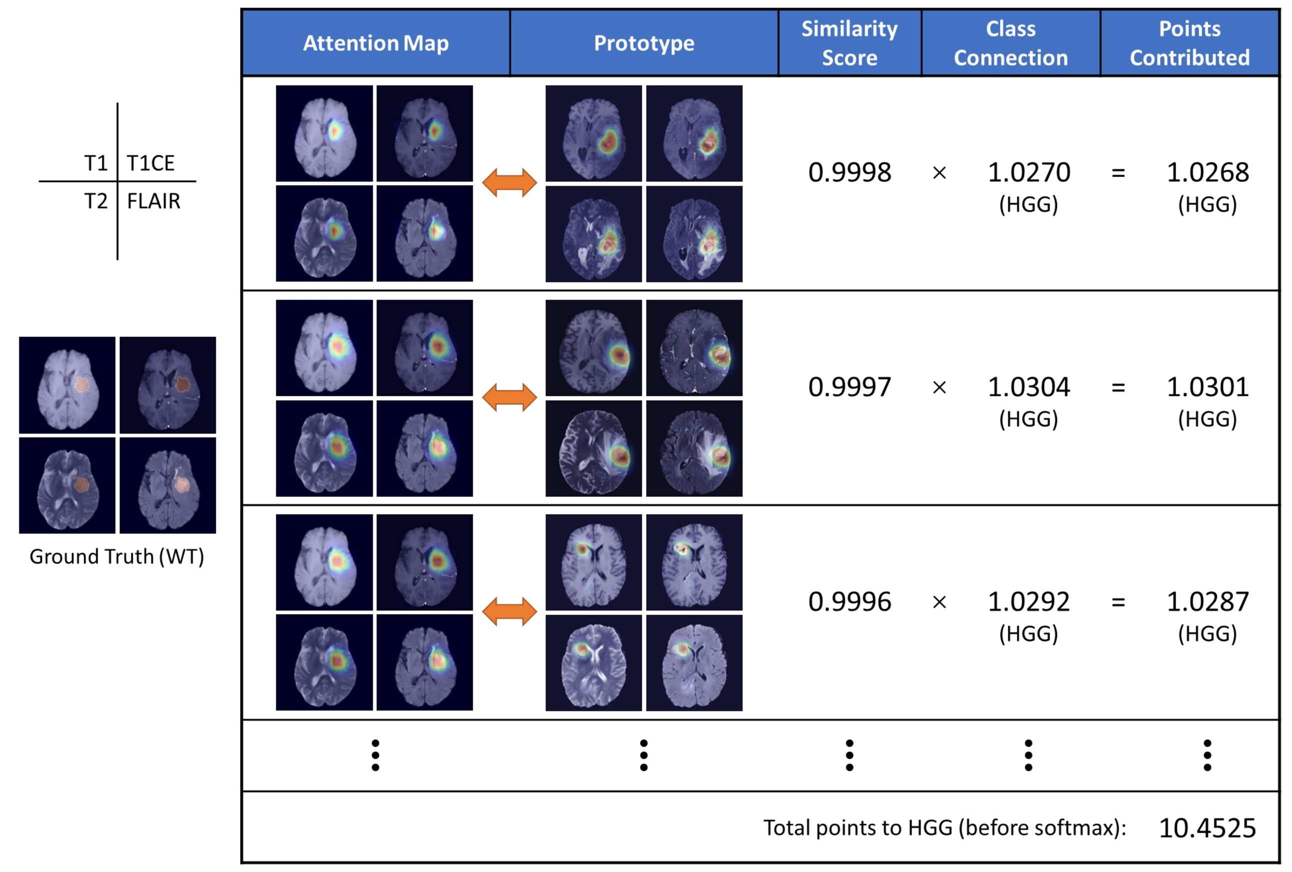
| Model | Condition | | | Classification | Interpretability | |
|---|---|---|---|---|---|---|
| | AM | SM | OC | BAC | IDS | AP |
| CNN (with GradCAM) | | | | 0.865±0.026 | 0.112±0.049 | 0.099±0.030 |
| ProtoPNet | | | | 0.868±0.032 | 0.609±0.164 | 0.007±0.001 |
| XProtoNet | √ | | | **0.870±0.021** | 0.170±0.041 | 0.203±0.030 |
| MProtoNet A | √ | √ | | 0.868±0.050 (p=0.929) | 0.150±0.088 (p=0.647) | 0.568±0.125 (p=0.004) |
| MProtoNet B | √ | | √ | 0.865±0.015 (p=0.360) | 0.103±0.020 (p=0.069) | 0.204±0.028 (p=0.963) |
| MProtoNet C | √ | √ | √ | 0.858±0.048 (p=0.516) | **0.079±0.034** (p=0.031) | **0.713±0.058** (p<0.001) |

AM: attention map, SM: soft masking, OC: online-CAM

### Visualization Examples of the Localization Coherence Results




### Demonstration of the Case-Based Reasoning




## REFERENCES

[1] C. Chen et al., "This looks like that: deep learning for interpretable image recognition," NeurIPS 2019.

[2] A. J. Barnett et al., "A case-based interpretable deep learning model for classification of mass lesions in digital mammography," Nature Machine Intelligence, Dec. 2021.

[3] E. Kim et al., "XProtoNet: diagnosis in chest radiography with global and local explanations," CVPR 2021.

* Corresponding authors: Dr. Xiaoying Tang (email: tangxy@sustech.edu.cn); Dr. Roger Tam (email: roger.tam@ubc.ca)

Source code: https://github.com/aywi/mprotonet