# Inter-class separability loss for weakly supervised mutually exclusive multiclass segmentation of brain tumor lesions

Vivek Dhamale[1] and Vaanathi Sundaresan[1,2]

[1] Department of Computational and Data Sciences, Indian Institute of Science
Bengaluru, Karnataka 560012, India
[2] Corresponding author: vaanathi@iisc.ac.in

**Abstract.** Medical image segmentation is essential for diagnosis and treatment planning, however fully supervised deep learning methods require expensive pixel-level annotations. Weakly supervised semantic segmentation (WSSS) using class activation mapping (CAM) reduces this burden by utilizing image-level labels. While binary CAM has shown promising results, multiclass CAM remains under-explored and suffers from reduced accuracy due to weak localization signals. To address this, we propose a novel approach that improves multiclass WSSS by leveraging binary CAM to guide multiclass CAM, enhancing feature representation, inter-class boundary segmentation and prediction accuracy. Additionally, we introduce novel inter-class separability loss and agreement loss designed to enhance multiclass CAM learning by enforcing spatial consistency and class separability. Experimental results on brain tumor segmentation (BraTS) datasets demonstrate that our approach significantly enhances multiclass weakly supervised segmentation accuracy, outperforming existing methods. Our code is available at https://github.com/Vivek-Dhamale/WSS-Interclass-Sep.

**Keywords:** Weakly-Supervised Segmentation · Multiclass CAMs

## 1 Introduction

Accurate segmentation of structural anomalies on medical imaging data is an essential step in biomarker quantification and disease severity assessment. Recent advances in deep learning have produced promising results for several fully supervised binary and multiclass segmentation methods. However, these methods require pixel-level labels for training, which are time consuming, labor intensive, and costly to obtain because of the specialized medical knowledge required for accurate labeling [13].

Weakly supervised segmentation (WSS) methods address this issue by using partial (e.g. scribbles [23], points [5, 6], bounding boxes [12, 17, 19, 21] or lower-level labels (e.g., image-level labels [8]), thereby minimizing the need for detailed pixel-level annotations and making the segmentation process more efficient. Among the WSS methods, class activation mapping (CAM)-based approaches generate heat maps highlighting the most discriminative regions in an image and are particularly effective for identifying pathological structures using only image-level labels.

***Related work.*** Existing methods include pixel-wise approaches [1, 2] that employ pixel-wise loss functions or local patch similarity constraints, image-wise methods [11, 25] that enforce global consistency, and cross-image techniques [9, 24] that leverage inter-image relationships to improve segmentation performance [10]. In medical imaging, most CAM-based methods have been developed for single-class segmentation, with limited work extending them to multiclass settings [7, 16]. In a multiclass setting, the above CAM-based methods struggle with label co-occurrence, leading to similar activations for frequently coexisting classes and spatial adjacency (co-location) causing overlapping activations in closely located structures [16]. WS-MTST [7] introduces an aggregation loss to promote spatial compactness in weakly supervised multi-label tumor segmentation, leveraging transformer-based architectures. While effective for contiguous tumor regions, this design assumes spatial continuity within each class. Among the CAM-based methods, AME-CAM [8] employs a multi-exit classification network to extract activation maps at various resolutions, which are then hierarchically aggregated using an attention mechanism to generate high-resolution binary CAMs. However, when directly extended to multiclass settings, it fails to produce mutually exclusive activations, leading to significant overlap in segmentation because multiclass CAMs are generated independently, with the lack of mechanism to separate between classes. For instance, while foreground-background separability loss [9] has been proposed, it does not enforce separation between multiple foreground classes.

***Contributions.*** In this work, to address the limitations of CAM in generating mutually exclusive activations, we propose a multiclass CAM approach that leverages binary CAMs as guidance to enhance the generation of multiclass CAMs, improving class localization.To address the challenge of overlapping activations in multiclass lesions, we introduce two novel loss functions: inter-class separability loss, which reduces overlap in CAMs for these regions, and agreement loss, which ensures consistency between binary and multiclass CAMs. We also evaluate our method against state-of-the-art CAM techniques in extracting multiclass segmentation results from classification networks of medical imaging data with multiclass co-located, co-occurring labels.

## 2 Methodology

**Problem formulation:** Let $\mathcal{X} \subset \mathbb{R}$ be the image space $\forall\ x_i \in \mathcal{X}$ and the corresponding image-level label space be $\mathcal{Y}\ \forall\ y_i^c \in \mathcal{Y}$, thus making the training data $\{\mathcal{X}, \mathcal{Y}\}$. Let $c = 1...C$, where $C \geq 2$ be the number of co-occurring classes in anomalous lesions. Given the multiclass segmentation problem, the objective is to determine accurate multiclass CAMs for weakly supervised segmentation of lesions with co-occurring $C$ classes.

The proposed method involves training a classification model using image-level labels $y_i^c$ and multiclass CAM aggregation to combine CAM maps from multiple layers of the model. Here, we also propose a novel refinement strategy for multiclass CAMs with the guidance of the binary CAMs. Additionally, since using class-specific loss alone [24] only provides separability between foreground and background, we also introduce inter-class separability and agreement losses to enforce the mutual exclusion among foreground classes. The overall architecture of the proposed method is shown in Fig. 1.

### 2.1 Multiclass Classification using image-level labels

We used 2D ResNet-18 architecture to enable multiclass classification since the ResNet architecture has shown promising results [8]. We scale the model with multi-exit training to handle multiple classes, where we introduce internal classifiers after each residual block, with the number of classifiers equal to the number of classes $C$. These classifiers generate activation maps $\mathcal{F}^{(l,c)}(x)$ at different layers $l$ for class $c$. At each residual block, predictions are obtained by applying Global Average Pooling $(GAP)$ to the activation maps, as defined by:

$$\hat{y}^{(l,c)} = GAP\left(\mathcal{F}^{(l,c)}(x)\right) \tag{1}$$

where $\mathcal{F}^{(l,c)}(x)$ represents the feature map of class $c$ from the $l$-th residual block. We trained the classification model using the cross-entropy loss $\mathcal{L}_{cls}$,

$$\mathcal{L}_{cls} = \sum_l \sum_c w_i \cdot \mathcal{L}_{ce}\left(\hat{y}^{(l,c)}, y^{(c)}\right) \tag{2}$$

where $y^{(c)}$ is the image-level ground truth label for class $c$ and $w_l$ is a weight factor for loss at each block $l$. To enhance feature representation, we pretrain the classifier using a multi-label supervised contrastive learning method [26].

### 2.2 Multiclass CAM Aggregation Network

While we obtained the activation maps $\mathcal{F}^{(l,c)}(x)$ of different resolutions from different layers $l$, they need to be aggregated into mutually exclusive class-specific CAMs. Hence, we upsample $\mathcal{F}^{(l,c)}(x)$ from different layers $l$ to original dimensions so that we get $l$ activation maps for each class $c$, to be provided as input for the aggregation module (indicated in Fig. 1). Since our aim is not only to make the individual foreground classes mutually exclusive, but also to separate all foreground classes from background, we use the binary CAM to guide for the latter task. Hence, we generate binary CAMs $\mathcal{F}_{\text{ag}}^{bin}$ for the whole lesion (by training the aggregation model [8] for combined foreground classes vs background), to obtain a general lesion localization map. These binary CAMs serve as guidance for obtaining final aggregated mutually exclusive multiclass CAMs ($\mathcal{F}_{\text{ag}}^{c}$). The process of binary-guided multiclass CAM aggregation is explained below.

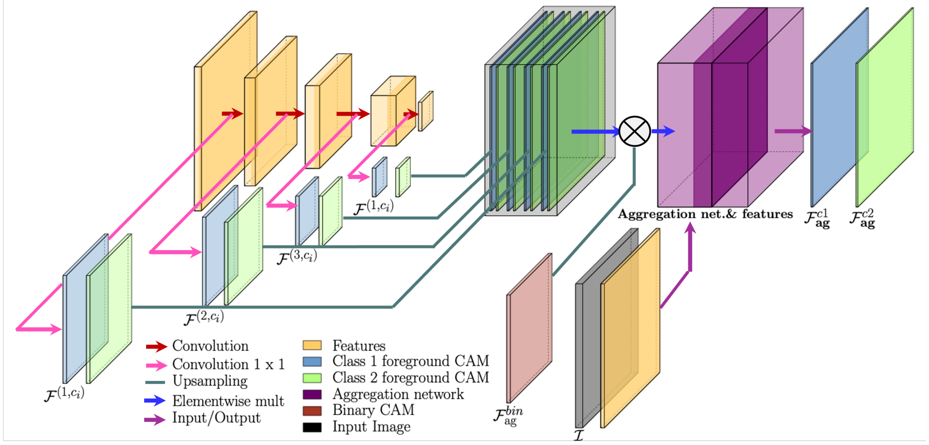

**Fig. 1.** The overall architecture consists of two main components: (1) a Multiclass Multi-Exit Classifier (left), which processes the input and generates intermediate feature representations, and (2) a Multiclass CAM Aggregation Network (right) with binary CAM guidance, which refines and combines class activation maps.

We refine the layer-wise multiclass activation maps $\mathcal{F}^{(l,c)}(x)$ using the binary CAM $\mathcal{F}^{bin}_{\text{ag}}$. Specifically, for each residual block $l$, the extracted multiclass activation map $\mathcal{F}^{l,c}$ is modulated using $\mathcal{F}^{bin}_{\text{ag}}$ as follows:

$$\tilde{\mathcal{F}}^{(l,c)} = \mathcal{F}^{(l,c)} \odot \mathcal{F}^{bin}_{\text{ag}} \tag{3}$$

where $\odot$ denotes element-wise multiplication. This ensures that the multiclass activation maps remain aligned with the overall lesion localization while improving the spatial consistency of class-specific regions using inter-class separability loss and agreement loss described later in Sec. 2.3. Unlike WS-MTST [7], this strategy avoids the spatial compactness constraint, potentially allowing flexibility in representing multiple disconnected regions of the same class.

To adapt the attention mechanism for the multiclass setting, we modify the attention network $A(\cdot)$ (proposed in [8]) to assign class-specific pixel-wise importance scores $\tilde{S}^{(l,c)}_{(jk)}$ to each activation map. Unlike the binary setting, where a single class attention score is learned, our modified attention network assigns separate attention scores to each class. Specifically, the network takes as input the image $x$, masked by the multiclass activation maps, and computes the corresponding importance scores as follows:

$$\tilde{S}^{(l,c)}_{(jk)} = A\left(\left[x \odot N\left(\tilde{\mathcal{F}}^{(l,c)}(x)\right)\right]_{l=1}^{L}\right) \tag{4}$$

where $[\cdot]_{l=1}^{L}$ denotes channel-wise concatenation across residual layers $l$, $N(\cdot)$ normalizes activation maps to the range $[0, 1]$, and $\odot$ represents element-wise multiplication. This modification ensures that the attention network learns distinct spatial importance patterns for each class, enabling improved localization and separation in the multiclass setting. Thus, the final aggregated class-specific activation map, $\mathcal{F}^c_{\text{ag}}$, is obtained as $\mathcal{F}^c_{\text{ag}} = \sum_l \tilde{S}^{(l,c)}_{(jk)} \odot \tilde{\mathcal{F}}^{(l,c)}$.

### 2.3 Loss Functions for Multiclass CAM Learning

To ensure effective multiclass activation map learning, we employ three loss functions as explained below.

*Class-Specific Separation Loss* : We extend the foreground-background separability loss [24] ($\mathcal{L}_{\text{pos}}$, $\mathcal{L}_{\text{neg}}$) for multiclass setting by applying it to each class separately, while maintaining class-specific foreground consistency. It is formulated as:

$$\mathcal{L}_{\text{c}} = \sum_{\text{c}} \Big( \mathcal{L}_{\text{neg}}(v^s_{f,c}, v^t_{b,c}) + \mathcal{L}_{\text{pos}}(v^s_{f,c}, v^t_{f,c}) + \mathcal{L}_{\text{pos}}(v^s_{b,c}, v^t_{b,c}) \Big) \tag{5}$$

Intuitively, the positive loss encourages consistency between similar regions (e.g., foreground–foreground or background–background), while the negative loss enforces separation between foreground and background features for each class.

Here $v^s_{f,c}$ and $v^t_{b,c}$ denote the foreground and background feature representations for class $c$ in the $s$-th and $t$-th instances (images), respectively and computed as below.

$$v^s_{f,c} = \mathcal{F}^c_{\text{ag}} \odot \mathcal{P}(x), \quad v^s_{b,c} = (1 - \mathcal{F}^c_{\text{ag}}) \odot \mathcal{P}(x) \tag{6}$$

where $\mathcal{P}(\cdot)$ represents the projection network (implemented as a $1 \times 1$ convolution), which maps the input three-channel image $x$ to a single-channel representation (indicated by features in Fig. 1) for downstream processing.

*Inter-Class Separability Loss* : To ensure class-wise distinctiveness, we minimize the similarity between the foreground features of different classes:

$$\mathcal{L}_{\text{sep}} = \sum_{c \neq c'} \mathcal{L}_{\text{neg}}(v^s_{f,c}, v^t_{f,c'}) \tag{7}$$

This constraint prevents overlapping feature representations between different classes, promoting better class separation.

*Agreement Loss* : To align the aggregated multiclass activation maps with the binary activation map, we introduce an agreement loss based on the Binary Cross-Entropy (BCE) function:

$$\mathcal{L}_{\text{agree}} = \mathcal{BCE}\left( \max_c \mathcal{F}^c_{\text{ag}}, \mathcal{F}^{bin}_{\text{ag}} \right) \tag{8}$$

The term $\max_c \mathcal{F}_{\text{ag}}^c$ selects the most confident class activation at each pixel, ensuring consistency between the multiclass and binary segmentation outputs.

Hence, the total loss function is defined as a weighted sum of the above components:

$$\mathcal{L} = \lambda_c \mathcal{L}_{\text{c}} + \lambda_{\text{sep}} \mathcal{L}_{\text{sep}} + \lambda_{\text{agree}} \mathcal{L}_{\text{agree}} \tag{9}$$

where $\lambda_c$, $\lambda_{\text{sep}}$, and $\lambda_{\text{agree}}$ are hyperparameters controlling the relative importance of each loss component. Once the CAMs are obtained, we apply a thresholding operation with binary guidance, ensuring that the multiclass segmentation mask remains constrained within the binary map.

## 3 Experiment Setup

***Dataset details.*** We used brain tumor segmentation (BraTS) 2020 dataset [3, 4, 18], with multimodal MRI scans from 369 patients across four modalities: FLAIR, T1, T1ce, and T2 being publicly available, of which we use T1ce, T2, and FLAIR. It provides ground-truth annotations for edema, enhancing tumor core, and nonenhancing tumor core. For our study, we merge the enhancing and non-enhancing tumor regions into a single core class. The training:validation:test split is 237:59:73 respectively. This dataset effectively captures the challenges of spatial co-location and co-occurrence of core and edema, making it well-suited for evaluating the proposed method.

***Implementation details.*** The multiclass classifier was pretrained with MulSupCon [26] on the training set. Both classification and aggregation models are optimized using Adam Optimizer [15], with learning rates of $5e^{-4}$ and $1e^{-3}$ respectively. Training is done for 50 epochs with model checkpointing, saving the best models based on validation loss. We threshold the binary and multiclass CAMs at 0.4 and 0.5 respectively. The loss weights used in the total loss function are set as $\lambda_c = 1$, $\lambda_{\text{sep}} = 1$, and $\lambda_{\text{agree}} = 5$. All the above parameters are set empirically. Experiments are conducted on an Intel i9-10980XE CPU, 128 GB RAM, and two NVIDIA RTX A6000 GPUs (each with 48 GB memory).

***Evaluation metrics.*** We evaluate the final segmentation performance using Dice score, IoU, and 95th percentile of Hausdorff distance (HD95).

### 3.1 Experiments and results

***Comparison with state-of-the-art methods.*** We compare our method with existing CAM-based methods: GradCAM [20], LayerCAM [14], ScoreCAM [22], and state-of-the-art method AME-CAM [8]. For AME-CAM, we train the classifier and the aggregation model separately for each class to adapt to the multiclass problem. Existing CAM-based methods result in poor segmentation quality with imprecise boundary localization, as reflected by their lower Dice (<0.700) and IoU scores (<0.700) and higher HD95 values (>100). Notably, these methods tend to produce similar CAMs for both classes, leading to significant overlap in the predicted class foregrounds. Although AME-CAM demonstrates improved spatial accuracy, CAM maps for edema still tend to propagate into the core region, as indicated in Fig. 2 (fourth row; magenta color shows false positive), leading to overlap between different class foregrounds. In contrast, our method generates mutually exclusive core and edema foreground CAMs, achieving the best segmentation performance across nearly all metrics, ensuring spatially consistent and mutually exclusive predictions.

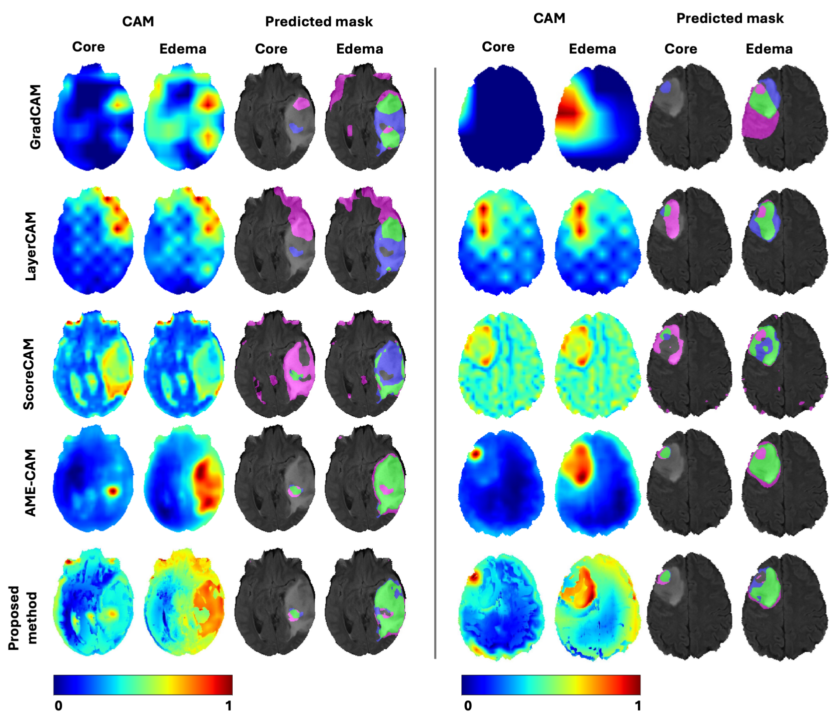

**Fig. 2.** Class activation maps and segmentation outputs for core and edema across different methods for two sample cases (left to right). Each row represents a different method, while columns depict the generated CAMs and segmentation masks. False negatives (Blue), true positives (Green), and false positives (Magenta) illustrate that our method achieving better separation between core and edema while reducing false predictions, compared to other methods including AME-CAM.

Also, a key drawback of AME-CAM is the need to train separate classifiers and aggregation models for each class, making it computationally expensive and less scalable for datasets with more classes. In contrast, our method uses only two multi-exit classifiers (multiclass and binary) with an additional 1×1 convolutional layer (each for additional class), enabling efficient handling of extra classes without significant computational overhead.

**Table 1.** Comparison of segmentation performance metrics across different methods for Core and Edema regions.

| **Method** | **Dice** | | **IoU** | | **HD95** | |
|---|---|---|---|---|---|---|
| | **Core** | **Edema** | **Core** | **Edema** | **Core** | **Edema** |
| GradCAM [20] | 0.623 ± 0.471 | 0.336 ± 0.444 | 0.616 ± 0.476 | 0.322 ± 0.450 | 126.080 ± 173.336 | 170.295 ± 158.429 |
| LayerCAM [14] | 0.617 ± 0.475 | 0.339 ± 0.441 | 0.611 ± 0.480 | 0.323 ± 0.449 | 127.209 ± 172.914 | 164.188 ± 160.511 |
| ScoreCAM [22] | 0.617 ± 0.475 | 0.337 ± 0.442 | 0.611 ± 0.480 | 0.322 ± 0.449 | 127.970 ± 172.970 | 163.568 ± 160.304 |
| AME-CAM [8] | 0.671 ± 0.421 | 0.711 ± 0.364 | 0.643 ± 0.438 | 0.663 ± 0.399 | **50.180 ± 116.223** | 34.647 ± 93.349 |
| Our Method | **0.757 ± 0.107** | **0.742 ± 0.088** | **0.732 ± 0.101** | **0.691 ± 0.085** | 55.429 ± 43.675 | **29.598 ± 16.961** |

***Ablation studies.*** Table 2 shows the effect of addition of individual loss components on the proposed method's performance. Paired two-tailed t-test were performed in the ablation study to determine the significance of improvement with addition of each loss component.

Adding $\mathcal{L}_{\text{sep}}$ to the baseline $\mathcal{L}_{\text{c}}$ helps in better separating the core and edema classes, as reflected in the statistically significant improvement in Dice and IoU for the edema class. However, this separation introduces slight boundary inconsistencies, leading to a significant increase in HD95 for the tumor core class, indicating reduced spatial precision. When $\mathcal{L}_{\text{agree}}$ is further incorporated, the segmentation quality improves across all metrics. This loss ensures that the predicted segmentation maps remain constrained within the binary whole tumor mask, leading to significantly higher Dice and IoU scores while also refining boundary localization, as shown by the reduction in HD95 for the tumor core and edema classes. Overall, the combined effect of separating the class regions ($\mathcal{L}_{\text{sep}}$) and constraining them within the whole tumor region ($\mathcal{L}_{\text{agree}}$) results in the most accurate and spatially consistent segmentation.

**Table 2.** Ablation study on different loss configurations, evaluating segmentation performance for Core and Edema regions. Statistical significance is assessed row-wise, comparing each loss configuration to the previous row using paired, two-tailed t-test, with */** indicating significant improvement with p-value $<$ 0.05/p-value $<$ 0.01.

| **Loss Config.** | **Dice** | | **IoU** | | **HD95** | |
|---|---|---|---|---|---|---|
| | **Core** | **Edema** | **Core** | **Edema** | **Core** | **Edema** |
| $\mathcal{L}_{\text{c}}$ | 0.732 ± 0.092 | 0.716 ± 0.084 | 0.712 ± 0.090 | 0.669 ± 0.081 | **55.248 ± 41.831** | 31.164 ± 22.943 |
| $\mathcal{L}_{\text{c}} + \mathcal{L}_{\text{sep}}$ | 0.740 ± 0.114 | 0.702 ± 0.085** | 0.719 ± 0.107 | 0.656 ± 0.082* | 61.204 ± 46.101** | 32.000 ± 23.010 |
| $\mathcal{L}_{\text{c}} + \mathcal{L}_{\text{sep}} + \mathcal{L}_{\text{agree}}$ | $\mathbf{0.757 \pm 0.107}$** | $\mathbf{0.742 \pm 0.088}$** | $\mathbf{0.732 \pm 0.101}$* | $\mathbf{0.691 \pm 0.085}$** | 55.429 ± 43.675** | $\mathbf{29.598 \pm 16.961}$* |

## 4 Conclusions

In this work, we propose an effective multiclass CAM method with binary CAM guidance for better separation of CAMs for the core and edema regions, ensuring spatially consistent and mutually exclusive weakly-supervised predictions. We introduce novel inter-class separability loss and agreement loss to improve spatial precision and ensure mutually exclusive segmentation of class foregrounds.

Compared to existing CAM-based methods, such as AME-CAM, our approach demonstrates superior segmentation performance with better spatial precision and reduced class overlap. Additionally, our method is computationally efficient, offering improved scalability for datasets with multiple classes. Future work includes extensive validation on additional datasets, analysis of performance with more foreground classes, and evaluation of computational efficiency. Other future directions involve adapting the proposed method across various domains (e.g., modalities, scanners) and lesion types, with potential applications in open-world anomaly detection problems.

***Acknowledgements.*** This work was supported by DBT/Wellcome Trust India Alliance Fellowship [IA/E/22/1/506763], the Council of Scientific & Industrial Research (CSIR) under its ASPIRE program [25WS(013)/2023-24/EMR-II/ASPIRE], the Science and Engineering Research Board Start-up Research Grant [SRG/2023/001406], Siemens Healthineers-CDS Collaborative Laboratory of Artificial Intelligence in Precision Medicine, India and Pratiksha Trust, Bangalore, India [FG/PTCH-23-1004].

***Disclosure of Interests.*** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Ahn, J., Cho, S., Kwak, S.: Weakly supervised learning of instance segmentation with inter-pixel relations. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 2204–2213 (2019)
2. Ahn, J., Kwak, S.: Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 4981–4990 (2018)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Scientific data **4**(1), 1–13 (2017)
4. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint arXiv:1811.02629 (2018)
5. Bearman, A., Russakovsky, O., Ferrari, V., Fei-Fei, L.: What's the point: Semantic segmentation with point supervision. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision – ECCV 2016. pp. 549–565. Springer International Publishing, Cham (2016)
6. Chen, H., Wang, J., Chen, H.C., Zhen, X., Zheng, F., Ji, R., Shao, L.: Seminar learning for click-level weakly supervised semantic segmentation. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 6900–6909 (2021)
7. Chen, H., An, J., Jiang, B., Xia, L., Bai, Y., Gao, Z.: WS-MTST: Weakly supervised multi-label brain tumor segmentation with transformers. IEEE Journal of Biomedical and Health Informatics **27**(12), 5914–5925 (2023)

8. Chen, Y.J., Hu, X., Shi, Y., Ho, T.Y.: Ame-cam: Attentive multiple-exit cam for weakly supervised segmentation on mri brain tumor. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 173–182. Springer Nature Switzerland, Cham (2023)
9. Chen, Z., Sun, Q.: Extracting class activation maps from non-discriminative features as well. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 3135–3144 (2023)
10. Chen, Z., Sun, Q.: Weakly-supervised semantic segmentation with image-level labels: from traditional models to foundation models (2024), `https://arxiv.org/abs/2310.13026`
11. Chen, Z., Wang, T., Wu, X., Hua, X., Zhang, H., Sun, Q.: Class re-activation maps for weakly-supervised semantic segmentation. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 959–968 (2022)
12. Dai, J., He, K., Sun, J.: BoxSup: Exploiting Bounding Boxes to Supervise Convolutional Networks for Semantic Segmentation . In: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 1635–1643. IEEE Computer Society, Los Alamitos, CA, USA (Dec 2015)
13. Ellis, R.J., Sander, R.M., Limon, A.: Twelve key challenges in medical machine learning and solutions. Intelligence-Based Medicine **6**, 100068 (2022)
14. Jiang, P.T., Zhang, C.B., Hou, Q., Cheng, M.M., Wei, Y.: Layercam: Exploring hierarchical class activation maps for localization. IEEE Transactions on Image Processing (2021)
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR **abs/1412.6980** (2014), `https://api.semanticscholar.org/CorpusID:6628106`
16. Kuang, Z., Yan, Z., Yu, L.: Weakly supervised learning for multi-class medical image segmentation via feature decomposition. Comput. Biol. Med. **171**(C) (Mar 2024)
17. Lee, J., Yi, J., Shin, C., Yoon, S.: BBAM: Bounding Box Attribution Map for Weakly Supervised Semantic and Instance Segmentation . In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2643–2651. IEEE Computer Society, Los Alamitos, CA, USA (Jun 2021)
18. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
19. Oh, Y., Kim, B., Ham, B.: Background-Aware Pooling and Noise-Aware Loss for Weakly-Supervised Semantic Segmentation . In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6909–6918. IEEE Computer Society, Los Alamitos, CA, USA (Jun 2021)
20. Selvaraju, R.R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., Batra, D.: Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization. CoRR **abs/1610.02391** (2016), `http://arxiv.org/abs/1610.02391`
21. Song, C., Huang, Y., Ouyang, W., Wang, L.: Box-Driven Class-Wise Region Masking and Filling Rate Guided Loss for Weakly Supervised Semantic Segmentation . In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3131–3140. IEEE Computer Society, Los Alamitos, CA, USA (Jun 2019)

22. Wang, H., Du, M., Yang, F., Zhang, Z.: Score-cam: Improved visual explanations via score-weighted class activation mapping. CoRR **abs/1910.01279** (2019), http://arxiv.org/abs/1910.01279
23. Wu, L., Zhong, Z., Fang, L., He, X., Liu, Q., Ma, J., Chen, H.: Sparsely annotated semantic segmentation with adaptive gaussian mixtures. In: CVPR. pp. 15454–15464 (2023), https://doi.org/10.1109/CVPR52729.2023.01483
24. Xie, J., Xiang, J., Chen, J., Hou, X., Zhao, X., Shen, L.: C2 am: Contrastive learning of class-agnostic activation map for weakly supervised object localization and semantic segmentation. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 979–988 (2022)
25. Xu, L., Ouyang, W., Bennamoun, Boussaid, F., Xu, D.: Multi-class token transformer for weakly supervised semantic segmentation. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 4300–4309 (2022)
26. Zhang, P., Wu, M.: Multi-label supervised contrastive learning. Proceedings of the AAAI Conference on Artificial Intelligence **38**(15), 16786–16793 (Mar 2024), https://ojs.aaai.org/index.php/AAAI/article/view/29619