Preprints.org

Article

MDD-Net: Medical Segmentation with Parallel Denoising Transformation

Sophia Brown , James Davis [*] , Jannat Roy

Article

MDD-Net: Medical Segmentation with Parallel Denoising Transformation

Sophia Brow, James Davis * and Jannat Roy

Brandeis University

* Correspondence: jamesdavis@brandeis.edu

Abstract: The automated segmentation of brain gliomas from multimodal MRI scans is pivotal in both clinical trials and everyday medical practice. Manual segmentation, however, presents significant challenges due to its labor-intensive nature, high costs, and susceptibility to inaccuracies. These difficulties arise despite the involvement of skilled human experts, primarily because of the substantial variability and uncertainty inherent in human annotations. Addressing these issues, our study introduces MDD-Net, an advanced end-to-end deep learning-based segmentation framework. This framework leverages a novel multi-decoder architecture that concurrently addresses three distinct sub-tasks through a partially shared encoder, enhancing the model's ability to generalize across different segmentation challenges. Additionally, we incorporate sophisticated smoothing techniques applied to the input MRI images, producing denoised versions that serve as supplementary inputs to the network. This dual-input strategy not only mitigates noise-related distortions but also enriches the feature representation, leading to more precise segmentation outcomes. Comprehensive validation results demonstrate a significant improvement in segmentation performance when employing the proposed MDD-Net method compared to existing approaches. This advancement holds promise for enhancing diagnostic accuracy and streamlining workflows in clinical settings. Future work will explore the integration of additional imaging modalities and the application of MDD-Net to other types of tumors, potentially broadening its clinical applicability and impact.

Keywords: Biomedical Segmentation, Uncertainty Estimation, Medical Imaging, MRI, Denoising Techniques

1. Introduction

Glioma is a specific type of brain tumor that originates from glial cells. It is the most commonly diagnosed brain tumor and is associated with the highest mortality rate among brain tumors. The World Health Organization (WHO) classifies gliomas into four grades: low-grade gliomas (LGG), encompassing classes I and II, and high-grade gliomas (HGG), comprising classes III and IV. High-grade gliomas are particularly aggressive and pose significant threats to patient survival. Annually, approximately 190,000 new cases of glioma are diagnosed worldwide [6], and the prognosis remains poor, with around 90% of patients succumbing to the disease within 24 months following surgical resection [18]. Accurate segmentation of the tumor is crucial for both the planning of radiotherapy treatments and the ongoing diagnostic monitoring of disease progression. However, manual segmentation is a labor-intensive process that is subject to inter- and intra-observer variability, leading to inconsistencies and uncertainties in the resulting annotations. These challenges underscore the need for automated or semi-automated segmentation methods to enhance treatment quality and improve the efficiency of managing patients with glioma.

The quest for reliable automatic segmentation of brain tumors in multimodal Magnetic Resonance Imaging (MRI) scans has led to the establishment of the Brain Tumor Segmentation (BraTS) challenge [2–5,14,40,41,48]. BraTS is an annual competition, often associated with the Medical Image Computing and Computer Assisted Intervention (MICCAI) conference, designed to benchmark and evaluate state-of-the-art methods in brain tumor segmentation. Participants in BraTS are provided with MRI datasets encompassing four different structural MRI modalities: T1-weighted contrast-enhanced (T1c), T2-weighted (T2), T1-weighted (T1), and Fluid-Attenuated Inversion Recovery (FLAIR). These diverse modalities offer complementary information that aids in the comprehensive analysis and precise segmentation of brain tumors. The ground truth segmentation masks are created through manual

annotations by one to four raters, which are subsequently refined by expert annotators to ensure high quality. The performance of the submitted segmentation algorithms is assessed using several metrics, including the Dice Similarity Coefficient (DSC), sensitivity, specificity, and the 95th percentile of the Hausdorff Distance (HD95).

Since the introduction of the U-Net architecture by Ronneberger *et al.* [16], Convolutional Neural Networks (CNNs) with skip connections have become the foundational architecture for medical image segmentation tasks. Numerous variations and extensions of the U-Net have been proposed to tackle the specific challenges posed by brain tumor segmentation. For instance, in the BraTS 2019 challenge, Jiang *et al.* [11], the first-place winners, introduced an end-to-end two-stage cascaded U-Net. This approach involves a coarse segmentation stage followed by a fine segmentation stage, enabling the model to capture both the overall structure and finer details of the tumor substructures. Similarly, Zhao *et al.* [21], who secured second place in the same challenge, employed a variety of techniques for 3D MRI brain tumor segmentation, including advanced preprocessing methods, innovative model design strategies, and optimized training procedures. McKinley *et al.* [13] presented DeepSCAN, an adaptation of their earlier 3D-to-2D Fully Convolutional Network (FCN), which replaces batch normalization with instance normalization and incorporates a lightweight local attention mechanism. This modification contributed to DeepSCAN achieving third place in the BraTS 2019 competition.

Building upon the foundation laid by previous research, our proposed architecture extends the work presented in [20,58], where the TUNet was introduced during the BraTS 2019 challenge. Although TUNet demonstrated commendable performance, it has notable limitations, primarily due to its composition of three cascaded networks. This architectural complexity makes it difficult to process full volumetric MRI data, typically of dimensions $240 \times 240 \times 155$, within the memory constraints of contemporary Graphics Processing Units (GPUs). As a result, TUNet adopts a patch-based segmentation approach, which significantly increases training times. Additionally, the reliance on multiple cascaded networks may impede the model's ability to capture global contextual information within the MRI scans.

Motivated by the effectiveness of cascaded network architectures, as demonstrated in studies such as [11,20], this work introduces a novel multi-decoder architecture, termed **MDD-Net** (Multi-Decoder Denoising Network), aimed at decomposing the complex task of brain tumor segmentation into simpler, more manageable sub-tasks. In addition to the multi-decoder framework, we propose the utilization of multiple denoised versions of the original MRI images as inputs to the network. The rationale behind this approach is to mitigate the impact of salt-and-pepper noise, a common artifact in MRI scans [1], thereby enhancing the robustness and accuracy of the segmentation process. To the best of our knowledge, this represents the first application of multi-denoising inputs in the context of brain tumor segmentation.

We hypothesize that **MDD-Net** will address overfitting issues by employing a shared encoder across three distinct decoders, each responsible for a specific aspect of the segmentation task. Furthermore, the inclusion of denoised MRI images is expected to provide the network with enhanced insights into the multimodal data, improving its ability to accurately delineate tumor boundaries. Specifically, the network leverages two additional versions of the input images: one processed with a median filter to eliminate salt-and-pepper noise, and another subjected to a low-pass Gaussian filter to attenuate high-frequency components. This dual-denoising strategy is designed to refine the input data, thereby facilitating more precise and reliable segmentation results.

The proposed **MDD-Net** aims to offer a comprehensive solution for brain tumor segmentation by integrating advanced denoising techniques with a versatile multi-decoder architecture. By doing so, it seeks to improve segmentation accuracy, reduce computational overhead, and enhance the model's generalizability across diverse MRI datasets.

2. Related Work

The segmentation of brain tumors, particularly gliomas, from MRI scans is a critical task in medical imaging, facilitating accurate diagnosis, treatment planning, and monitoring of disease progression. Due to the complex nature of brain anatomy and the heterogeneous appearance of tumors across different MRI modalities, developing robust and reliable segmentation methods remains a significant challenge in the field. Automated segmentation techniques aim to overcome the limitations of manual delineation, such as time consumption, subjectivity, and variability among annotators, by providing consistent and efficient results [6,62].

The Brain Tumor Segmentation (BraTS) challenge [2–5,14,64] has emerged as a pivotal benchmark for evaluating and advancing brain tumor segmentation algorithms. Organized annually in conjunction with the Medical Image Computing and Computer Assisted Intervention (MICCAI) conference, BraTS provides a standardized dataset comprising multimodal MRI scans, including T1-weighted contrast-enhanced (T1c), T2-weighted (T2), T1-weighted (T1), and Fluid-Attenuated Inversion Recovery (FLAIR) images. These datasets are accompanied by meticulously annotated segmentation masks, which serve as ground truth for assessing the performance of participating algorithms using metrics such as the Dice Similarity Coefficient (DSC), sensitivity, specificity, and the 95th percentile of the Hausdorff Distance (HD95) [14].

The introduction of the U-Net architecture by Ronneberger et al. [16] marked a significant milestone in medical image segmentation. U-Net's encoder-decoder structure with skip connections effectively captures both high-level contextual information and fine-grained details, making it well-suited for delineating complex anatomical structures like brain tumors. Since its inception, numerous variants and extensions of U-Net have been proposed to enhance its performance and adaptability to specific segmentation tasks. For instance, Myronenko [15,70] introduced an autoencoder regularization approach to improve the robustness of 3D MRI brain tumor segmentation, demonstrating the versatility of the U-Net framework in handling volumetric data.

Building upon the foundational U-Net architecture, several studies have explored cascaded and multi-stage network designs to further enhance segmentation accuracy. Jiang et al. [11,49], the first-place winners of the BraTS 2019 challenge, proposed a two-stage cascaded U-Net that performs coarse segmentation in the initial stage, followed by fine segmentation to accurately delineate tumor substructures. Similarly, Vu et al. [20] introduced TuNet, an end-to-end hierarchical segmentation network that leverages multiple cascaded networks to capture both global and local features. Zhao et al. [21,78,79], who secured second place in the same challenge, incorporated various techniques such as advanced preprocessing methods, innovative model designs, and optimized training strategies to achieve superior performance in 3D MRI brain tumor segmentation. These cascaded approaches highlight the importance of hierarchical processing and multi-scale feature extraction in tackling the intricate task of brain tumor segmentation.

In addition to architectural innovations, preprocessing techniques play a crucial role in enhancing the quality of MRI data for segmentation tasks. Noise reduction, particularly the removal of salt-and-pepper noise, is essential for improving the reliability of segmentation outcomes. Ali et al. [1,81] proposed a novel method to eliminate salt-and-pepper noise in MRI images, which is a common artifact that can obscure tumor boundaries and degrade segmentation performance. By applying effective denoising algorithms, the clarity and consistency of MRI scans are significantly improved, thereby facilitating more accurate and robust segmentation by subsequent neural network models.

Recent advancements have also focused on integrating attention mechanisms and data augmentation strategies to bolster the performance of segmentation networks. Hu et al. [8,56,57] introduced Squeeze-and-Excitation (SE) networks, which adaptively recalibrate channel-wise feature responses by explicitly modeling interdependencies between channels. This mechanism enhances the network's ability to focus on relevant features, thereby improving segmentation accuracy. Furthermore, Isensee et al. [9] developed the BatchGenerators framework, a versatile Python tool for data augmentation that facilitates the training of more generalized and resilient models by introducing variability in the

training data. These techniques, when combined with robust network architectures, contribute to the development of highly effective brain tumor segmentation models.

Overall, the landscape of brain tumor segmentation has been significantly shaped by the interplay of innovative network architectures, effective preprocessing techniques, and comprehensive benchmarking through challenges like BraTS. The proposed **MDD-Net** aims to build upon these advancements by introducing a multi-decoder architecture combined with enhanced denoising inputs, thereby addressing the existing limitations and pushing the boundaries of automated brain tumor segmentation further.

3. Proposed Method

In response to the limitations identified in previous approaches [20], we introduce **MDD-Net** (Multi-Decoder Denoising Network), an end-to-end framework designed to simplify the complex task of multi-class tumor segmentation into three distinct binary segmentation tasks. This architectural innovation addresses memory constraints inherent in prior models, enabling the processing of entire input volumes without resorting to patch-based segmentation. Consequently, **MDD-Net** leverages global contextual information more effectively and reduces training times significantly. The following sections detail the various components and methodologies employed in our proposed framework.

3.1. Encoder Architecture

The encoder component of **MDD-Net** is built upon conventional convolutional blocks, similar to those introduced in the U-Net architecture [16]. Each convolutional block comprises a three-dimensional convolution layer with a kernel size of $3 \times 3 \times 3$, followed by batch normalization and a Leaky Rectified Linear Unit (LeakyReLU) activation function. To enhance feature representation, each convolutional block is succeeded by a Squeeze-and-Excitation Block (SEB), which adaptively recalibrates channel-wise feature responses [8]. Max-pooling layers are employed for downsampling, reducing the spatial dimensions while preserving essential features.

The initial number of convolutional filters is set to twelve, corresponding to the three denoising methods applied to the four input MRI modalities. This design choice ensures that the encoder effectively captures diverse features from the multimodal inputs. The encoder output tensor has a shape of $96 \times 20 \times 24 \times 16$, encapsulating rich hierarchical features essential for accurate segmentation. The detailed architecture of the encoder network is presented in Table 1.

Table 1. Encoder Architecture of **MDD-Net**. "Conv3" denotes a $3 \times 3 \times 3$ convolution, "BN" stands for batch normalization, "LeakyReLU" is the Leaky Rectified Linear Unit activation function, and "SEB" represents the Squeeze-and-Excitation Block.

Name	Layers	Repeat	Output Size
Input			$12 \times 160 \times 192 \times 128$
EncBlk–0	Conv3, BN, LeakyReLU, SEB	2	$12 \times 160 \times 192 \times 128$
EncDwn–1	MaxPooling	1	$12 \times 80 \times 96 \times 64$
EncBlk–1	Conv3, BN, LeakyReLU, SEB	2	$24 \times 80 \times 96 \times 64$
EncDwn–2	MaxPooling	1	$24 \times 40 \times 48 \times 32$
EncBlk–2	Conv3, BN, LeakyReLU, SEB	2	$48 \times 40 \times 48 \times 32$
EncDwn–3	MaxPooling	1	$48 \times 20 \times 24 \times 16$
EncBlk–3	Conv3, BN, LeakyReLU, SEB	2	$96 \times 20 \times 24 \times 16$

3.2. Decoder Architecture

The decoder component of **MDD-Net** consists of three separate decoder paths, each dedicated to segmenting a specific tumor region: whole tumor (W-Net), core tumor (C-Net), and enhancing tumor (E-Net). This multi-decoder setup allows the network to focus on distinct aspects of the tumor, enhancing segmentation accuracy and reducing inter-task interference. Each decoder path mirrors the encoder architecture with corresponding upsampling and convolutional operations. Skip connections, akin to those in U-Net, are integrated to retain high-resolution features from the encoder, facilitating precise localization. After each convolutional block or concatenation operation, a Squeeze-and-Excitation Block (SEB) is employed to recalibrate feature maps dynamically.

To enrich the feature maps in the C-Net and E-Net, outputs from the W-Net and C-Net decoders, respectively, are concatenated at corresponding levels. This hierarchical integration ensures that higher-level decoders benefit from the refined features captured by preceding decoders, promoting a cohesive segmentation process across different tumor regions. The detailed architecture of the decoder networks is illustrated in Table 2.

Table 2. Decoder Architectures of **MDD-Net**. "Conv3" denotes a $3 \times 3 \times 3$ convolution, "Conv1" a $1 \times 1 \times 1$ convolution, "BN" stands for batch normalization, "LeakyReLU" is the Leaky Rectified Linear Unit activation function, "SEB" represents the Squeeze-and-Excitation Block, and "Up–X" indicates 3D linear spatial upsampling of block X. The prefixes W–, C–, and E– correspond to the whole, core, and enhancing tumor regions, respectively.

Name	Layers	Repeat	Output Size
W–DecCat–2	Up–EncBlk–3 + EncBlk–2	1	$144 \times 40 \times 48 \times 32$
W–DecSae–2	SEB	1	$144 \times 40 \times 48 \times 32$
W–DecBlk–2	Conv3, BN, LeakyReLU, SEB	2	$48 \times 40 \times 48 \times 32$
W–DecCat–1	Up–DecBlk–2 + EncBlk–1	1	$72 \times 80 \times 96 \times 64$
W–DecSae–1	SEB	1	$72 \times 80 \times 96 \times 64$
W–DecBlk–1	Conv3, BN, LeakyReLU, SEB	2	$24 \times 80 \times 96 \times 64$
W–DecCat–0	Up–DecBlk–1 + EncBlk–0	1	$36 \times 160 \times 192 \times 128$
W–DecSae–0	SEB	1	$36 \times 160 \times 192 \times 128$
W–DecBlk–0	Conv3, BN, LeakyReLU, SEB	2	$12 \times 160 \times 192 \times 128$
W–Output	Conv1, Sigmoid	1	$1 \times 160 \times 192 \times 128$
C–DecCat–2	W–DecBlk–2 + W–DecCat–2	1	$192 \times 40 \times 48 \times 32$
C–DecSae–2	SEB	1	$192 \times 40 \times 48 \times 32$
C–DecBlk–2	Conv3, BN, LeakyReLU, SEB	2	$48 \times 40 \times 48 \times 32$
C–DecCat–1	W–DecBlk–1 + W–DecCat–1	1	$96 \times 80 \times 96 \times 64$
C–DecSae–1	SEB	1	$96 \times 80 \times 96 \times 64$
C–DecBlk–1	Conv3, BN, LeakyReLU, SEB	2	$24 \times 80 \times 96 \times 64$
C–DecCat–0	W–DecBlk–0 + W–DecCat–0	1	$48 \times 160 \times 192 \times 128$
C–DecSae–0	SEB	1	$48 \times 160 \times 192 \times 128$
C–DecBlk–0	Conv3, BN, LeakyReLU, SEB	2	$12 \times 160 \times 192 \times 128$
C–Output	Conv1, Sigmoid	1	$1 \times 160 \times 192 \times 128$
E–DecCat–2	C–DecBlk–2 + W–DecCat–2	1	$240 \times 40 \times 48 \times 32$
E–DecSae–2	SEB	1	$240 \times 40 \times 48 \times 32$
E–DecBlk–2	Conv3, BN, LeakyReLU, SEB	2	$48 \times 40 \times 48 \times 32$
E–DecCat–1	C–DecBlk–1 + W–DecCat–1	1	$96 \times 80 \times 96 \times 64$
E–DecSae–1	SEB	1	$96 \times 80 \times 96 \times 64$
E–DecBlk–1	Conv3, BN, LeakyReLU, SEB	2	$24 \times 80 \times 96 \times 64$
E–DecCat–0	C–DecBlk–0 + W–DecCat–0	1	$48 \times 160 \times 192 \times 128$
E–DecSae–0	SEB	1	$48 \times 160 \times 192 \times 128$
E–DecBlk–0	Conv3, BN, LeakyReLU, SEB	2	$12 \times 160 \times 192 \times 128$
E–Output	Conv1, Sigmoid	1	$1 \times 160 \times 192 \times 128$

Each decoder path incorporates skip connections that concatenate high-resolution features from the encoder with upsampled features in the decoder. This design facilitates the retention of spatial information, crucial for precise boundary delineation. Additionally, the SEB modules following each convolutional block enable the network to focus on the most informative channels, enhancing feature discrimination and segmentation performance.

To mathematically formalize the decoder operations, consider the upsampling process denoted by $U(\cdot)$ and the concatenation operation represented by $\oplus$. For a given decoder block D_l, where l indicates the current layer level, the feature maps can be expressed as:

$$D_l = \text{SEB}(\text{Conv3}(U(D_{l+1}) \oplus E_l)) \tag{1}$$

where E_l represents the encoder features at layer l, and D_{l+1} denotes the decoder features from the subsequent higher layer. This formulation underscores the hierarchical and interconnected nature of the multi-decoder architecture, facilitating the propagation of refined features across different segmentation tasks.

3.3. Channel Attention Mechanism

To further enhance feature representation, **MDD-Net** integrates Squeeze-and-Excitation Blocks (SEBs) after each convolutional and concatenation operation [8]. The SEB mechanism adaptively recalibrates channel-wise feature responses by modeling interdependencies between channels. This is achieved through a sequence of operations: global average pooling, fully connected layers, a ReLU activation, another fully connected layer, and a sigmoid activation function. Mathematically, for a given feature map $X \in \mathbb{R}^{C \times H \times W \times D}$, the SEB operations can be described as:

$$S(X) = \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot \text{GAP}(X))) \tag{2}$$

$$\hat{X} = S(X) \otimes X \tag{3}$$

where $\text{GAP}(X)$ denotes global average pooling, W_1 and W_2 are the weights of the fully connected layers, σ represents the sigmoid function, and $\otimes$ denotes channel-wise multiplication. This mechanism allows the network to emphasize informative features while suppressing less relevant ones, thereby improving segmentation accuracy.

3.4. Input Denoising Techniques

A critical aspect of **MDD-Net** is the incorporation of multiple denoised versions of the original MRI inputs to mitigate noise-related artifacts common in medical imaging, such as salt-and-pepper noise [1]. The denoising process involves two distinct methods: median filtering and Gaussian smoothing. Specifically, each MRI modality undergoes the following transformations:

1. **Median Filtering**: A $3 \times 3 \times 3$ median filter is applied to eliminate salt-and-pepper noise, preserving edges while removing outlier pixel values.
2. **Gaussian Smoothing**: A $3 \times 3 \times 3$ Gaussian filter with a standard deviation of 0.5 is employed to reduce high-frequency noise, resulting in a smoother image.

These denoised images are then concatenated with the original raw images, resulting in a total of twelve input channels (three denoised versions for each of the four MRI modalities). This multi-denoising input strategy enhances the network's ability to learn robust features by providing complementary representations of the input data. The comprehensive input tensor can be represented as:

$$I_{\text{input}} = [I_{\text{raw}}, I_{\text{median}}, I_{\text{Gaussian}}] \tag{4}$$

where I_{raw}, I_{median}, and I_{Gaussian} correspond to the raw, median-filtered, and Gaussian-smoothed images, respectively.

This approach not only mitigates the impact of noise but also enriches the feature space, allowing the network to leverage both detailed and smoothed representations for more accurate segmentation.

3.5. Data Preprocessing and Augmentation

Prior to inputting the data into **MDD-Net**, all MRI scans undergo normalization to achieve zero mean and unit variance, ensuring uniform intensity distribution across different modalities and subjects. This normalization is critical for stabilizing the training process and improving convergence rates.

To enhance the generalizability of the model and mitigate overfitting, extensive data augmentation is performed on-the-fly during training [9]. The augmentation pipeline includes the following transformations:

1. **Random Rotation**: Images are randomly rotated within a range of $[-1°, 1°]$ along all three spatial axes to simulate variations in patient orientation.
2. **Random Flip**: Mirror flipping is applied with a probability of 0.5 on each axis, introducing symmetry variations.
3. **Elastic Transformation**: With a probability of 0.3, elastic deformations are introduced to mimic realistic anatomical variations. The displacement field Δ for elastic transformation is defined as:

$$R_w = R_o + \alpha\Delta \tag{5}$$

   where R_w and R_o denote the coordinates of a voxel in the warped and original image, respectively, and α is the displacement strength. The displacement Δ is sampled from a uniform distribution $\mathcal{U}(-1, 1)$ for each spatial dimension and convolved with a Gaussian kernel with a standard deviation $\sigma = 0.25$.
4. **Random Scaling**: Images are randomly scaled within the range $[0.9, 1.1]$ with a probability of 0.3 to account for size variability in tumor presentations.
5. **Random Cropping and Resizing**: With a probability of 0.3, random cropping followed by resizing ensures that the network learns to handle variations in tumor localization and size.

These augmentation strategies introduce significant variability in the training data, enabling **MDD-Net** to generalize better to unseen data and enhancing its robustness against diverse imaging conditions.

3.6. Post-processing Techniques

Post-processing is employed to refine the segmentation outputs of **MDD-Net**, particularly addressing challenges in distinguishing between low-grade glioma (LGG) and high-grade glioma (HGG) patients. A common issue in segmentation tasks is the misclassification of small enhancing tumor regions, which may be erroneously labeled as edema or necrosis.

To address this, we adopt a strategy similar to our previous work [20]. Specifically, any enhancing tumor region comprising fewer than 500 connected voxels is reclassified as necrosis. Mathematically, for each connected component C in the enhancing tumor segmentation:

$$C' = \begin{cases} \text{Necrosis}, & \text{if } |C| < 500, \\ \text{Enhancing Tumor}, & \text{otherwise.} \end{cases} \tag{6}$$

This post-processing step mitigates false positives in small tumor regions, thereby improving the overall segmentation accuracy and reliability of the model's predictions.

3.7. Uncertainty Quantification in Segmentation

The BraTS challenge introduced the task of "Quantification of Uncertainty in Segmentation" to evaluate the confidence of segmentation predictions [20]. The goal is to reward models that exhibit high confidence in correct predictions and maintain uncertainty in incorrect ones. Participants are required to generate uncertainty maps with values ranging from 0 (most certain) to 100 (most uncertain). The evaluation metrics include the Dice-based Area Under the Curve (DAUC), Relative False Positive Threshold (RFTP), and Relative False Negative Threshold (RFTN).

In **MDD-Net**, uncertainty is quantified based on the predicted probability maps for each tumor region. For each voxel (i, j, k) and tumor region $r \in \mathcal{R}$ (where $\mathcal{R}$ includes whole, core, and enhancing tumor regions), the uncertainty score $u^r_{i,j,k}$ is defined as:

$$u^r_{i,j,k} = \begin{cases} 200 \times (1 - p^r_{i,j,k}), & \text{if } p^r_{i,j,k} \geq 0.5, \\ 200 \times p^r_{i,j,k}, & \text{if } p^r_{i,j,k} < 0.5, \end{cases} \tag{7}$$

where $p^r_{i,j,k} \in [0, 1]$ is the predicted probability for voxel (i, j, k) belonging to tumor region r. This formulation ensures that higher uncertainty is assigned to lower probability predictions, aligning with the objective of accurate uncertainty quantification.

The resulting uncertainty maps provide valuable insights into the confidence of the model's predictions, enabling clinicians to make informed decisions and potentially guiding further diagnostic procedures.

3.8. Loss Function and Optimization

To train **MDD-Net**, a combination of loss functions is employed to balance the segmentation accuracy and uncertainty estimation. Specifically, we utilize the Dice loss for each binary segmentation task, combined with a weighted cross-entropy loss to address class imbalance inherent in medical imaging datasets. The overall loss function $\mathcal{L}$ can be expressed as:

$$\mathcal{L} = \sum_{r \in \mathcal{R}} \left(\mathcal{L}_{\text{Dice}}(p^r, g^r) + \lambda \mathcal{L}_{\text{CE}}(p^r, g^r) \right) \tag{8}$$

where p^r and g^r denote the predicted probability map and ground truth mask for tumor region r, respectively, and λ is a weighting factor balancing the two loss components.

For optimization, we employ the Adam optimizer [12], which adapts the learning rate for each parameter based on first and second moments of the gradients. The optimization process is governed by the following update rule:

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{9}$$

where θ_t represents the model parameters at iteration t, η is the learning rate, $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected first and second moment estimates, and ϵ is a small constant to prevent division by zero.

Training is conducted for a predefined number of epochs, with early stopping based on validation loss to prevent overfitting. Additionally, learning rate scheduling is applied to reduce the learning rate upon plateauing of the validation loss, facilitating finer convergence in later stages of training.

4. Experiment

In this section, we detail the experimental framework employed to evaluate the performance of the proposed **MDD-Net** (Multi-Decoder Denoising Network) for brain tumor segmentation. The experiments encompass the implementation specifics, training protocols, loss functions, optimization strategies, and comprehensive evaluation metrics used to assess the effectiveness of our model. Additionally, we present the results obtained from various validation and testing phases, accompanied by a discussion of their implications.

4.1. Implementation and Training Procedure

The **MDD-Net** was developed using the Keras library with TensorFlow serving as the backend framework. This combination was chosen for its robustness and extensive support for deep learning architectures, facilitating efficient model development and training. The implementation was executed on high-performance NVIDIA Tesla V100 GPUs provided by the High-Performance Computing Center North (HPC2N) at Umeå University, Sweden, ensuring rapid computation and handling of large volumetric MRI datasets.

To ensure the reliability and generalizability of our results, seven independent instances of **MDD-Net** were trained from scratch. Each model was subjected to 200 training epochs ($N_e = 200$) with a mini-batch size of one, optimizing the model's ability to learn from the data without overfitting. The extensive training regimen, which spanned approximately six days per model, was necessary to accommodate the complexity of the multi-decoder architecture and the high-dimensional input data.

During training, early stopping mechanisms were employed based on the validation loss to prevent overfitting and to terminate training once the model's performance plateaued. Additionally, learning rate scheduling was integrated to adjust the learning rate dynamically, allowing for finer convergence in later training stages. Gradient clipping was also utilized to maintain training stability and prevent the occurrence of exploding gradients, thereby enhancing the robustness of the optimization process.

4.2. Loss Function Design

Accurate segmentation performance is pivotal for clinical applications, necessitating a carefully designed loss function that balances precision and recall while addressing class imbalances inherent in medical imaging data. To this end, **MDD-Net** employs a hybrid loss function combining the Dice Similarity Coefficient (DSC) loss and Cross-Entropy (CE) loss, thereby leveraging the strengths of both metrics.

The DSC loss is defined as:

$$D(u,v) = \frac{2 \cdot |u \cap v|}{|u| + |v|}, \tag{10}$$

where u and v represent the predicted segmentation and the corresponding ground truth, respectively. To incorporate DSC into the loss function, we utilize the soft DSC loss, formulated as:

$$\mathcal{L}_{\text{DSC}}(u,v) = \frac{-2 \sum_i u_i v_i}{\sum_i u_i + \sum_i v_i + \epsilon}, \tag{11}$$

where u_i is the softmax output for label i, v_i is the one-hot encoded ground truth label, and $\epsilon = 1 \times 10^{-5}$ is a small constant to prevent division by zero.

Recognizing the challenge of class imbalance, especially with smaller tumor regions, the CE loss is incorporated to smooth the loss surface and enhance convergence. The CE loss is defined as:

$$\mathcal{L}_{\text{CE}}(u,v) = -\sum_i v_i \log(u_i). \tag{12}$$

The hybrid loss function, therefore, is the sum of the soft DSC loss and the CE loss:

$$\mathcal{L}_{\text{hybrid}}(u,v) = \mathcal{L}_{\text{DSC}}(u,v) + \mathcal{L}_{\text{CE}}(u,v). \tag{13}$$

For multi-class segmentation tasks, the final loss function aggregates the hybrid loss across all tumor regions (whole, core, and enhancing regions):

$$\mathcal{L}(u,v) = \sum_{r \in \mathcal{R}} \mathcal{L}_{\text{hybrid}}(u_r, v_r), \tag{14}$$

where $\mathcal{R}$ denotes the set of tumor regions.

Furthermore, segmentation performance is evaluated using the 95th percentile of the Hausdorff Distance (HD95), which quantifies the boundary alignment between the predicted segmentation and the ground truth:

$$H(u,v) = \max\{d(u,v), d(v,u)\}, \tag{15}$$

where

$$d(u,v) = \max_{u_i \in u} \min_{v_i \in v} \|u_i - v_i\|_2, \tag{16}$$

with $\|\cdot\|_2$ representing the Euclidean distance.

4.3. Optimization Strategy

Optimizing **MDD-Net** involves fine-tuning the model parameters to minimize the defined loss function effectively. We employ the Adam optimizer [12], renowned for its adaptive learning rate capabilities, which combine the benefits of both AdaGrad and RMSProp optimizers. The Adam optimizer was initialized with a learning rate (α_0) of 1×10^{-4}, and momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, facilitating rapid convergence and efficient handling of sparse gradients.

To further enhance the optimization process, a learning rate decay schedule was implemented, reducing the learning rate as training progresses. Specifically, the learning rate at epoch e is adjusted according to the following formula:

$$\alpha_e = \alpha_0 \cdot \left(1 - \frac{e}{N_e}\right)^3, \tag{17}$$

where $N_e = 200$ is the total number of epochs. This decay strategy allows the model to take larger steps initially, accelerating convergence, and smaller steps later, refining the parameter updates for improved accuracy.

Additionally, L_2 regularization with a penalty parameter of 1×10^{-5} was applied to all convolutional layers. This regularization technique mitigates overfitting by penalizing large weights, encouraging the model to learn more generalized features from the training data.

The final layer of **MDD-Net** employs a sigmoid activation function, which maps the output to a probability distribution, facilitating the interpretation of segmentation results as probabilities.

4.4. Evaluation Metrics

To comprehensively assess the performance of **MDD-Net**, we utilize a suite of evaluation metrics commonly adopted in the BraTS challenge. These metrics provide insights into both the accuracy and reliability of the segmentation results.

4.4.1. Dice Similarity Coefficient (DSC)

The DSC measures the overlap between the predicted segmentation and the ground truth:

$$\text{DSC}(u,v) = \frac{2|u \cap v|}{|u| + |v|}, \tag{18}$$

where u and v represent the predicted and ground truth segmentations, respectively. A higher DSC indicates greater similarity between the two segmentations.

4.4.2. Hausdorff Distance (HD95)

The HD95 quantifies the boundary discrepancy between the predicted segmentation and the ground truth by calculating the 95th percentile of the Hausdorff Distance:

$$\text{HD95}(u,v) = \max\{\text{HD95}(u,v), \text{HD95}(v,u)\}, \tag{19}$$

where

$$\text{HD95}(u, v) = \max_{u_i \in u} \min_{v_i \in v} \|u_i - v_i\|_2. \tag{20}$$

A lower HD95 signifies better boundary alignment between the predicted and ground truth segmentations.

4.4.3. Uncertainty Quantification Metrics

In addition to segmentation accuracy, uncertainty quantification is pivotal for clinical decision-making. The BraTS challenge introduced specific metrics to evaluate the quality of uncertainty maps generated by segmentation models:

- **Dice-based Area Under the Curve (DAUC)**: Measures the relationship between segmentation accuracy and uncertainty.
- **Relative False Positive Threshold (RFTP)**: Evaluates the model's ability to express uncertainty in false positive regions.
- **Relative False Negative Threshold (RFTN)**: Assesses the model's capability to indicate uncertainty in false negative regions.

These metrics are computed over the entire brain volume and provide a nuanced understanding of the model's confidence in its predictions.

4.5. Experimental Results

The performance of **MDD-Net** was rigorously evaluated through a series of experiments, encompassing cross-validation on the training set, validation on a separate dataset, and testing on an unseen test set. The results, presented in Tables 3–7, demonstrate the efficacy of the proposed model across various metrics.

Table 3. Mean Dice Similarity Coefficient (DSC) and 95th percentile Hausdorff Distance (HD95) scores with their standard deviations (in parentheses) from five-fold cross-validation on the training set (396 cases) for different models. Higher DSC and lower HD95 indicate better performance.

Model	DSC			HD95		
	Whole Tumor	Core Tumor	Enhancing Tumor	Whole Tumor	Core Tumor	Enhancing Tumor
U-Net without Denoising	90.66 (0.38)	86.93 (0.71)	76.16 (1.37)	4.91 (0.41)	4.78 (0.42)	3.46 (0.31)
U-Net with Denoising	90.98 (0.31)	87.53 (0.68)	76.55 (1.36)	4.49 (0.26)	4.32 (0.29)	3.41 (0.29)
MDD-Net with Denoising	92.75 (0.25)	88.34 (0.70)	78.13 (1.32)	4.32 (0.29)	4.30 (0.31)	3.29 (0.24)

Table 4. Segmentation performance of **MDD-Net** on the BraTS validation set (125 cases). The metrics include Dice Similarity Coefficient (DSC) and 95th percentile Hausdorff Distance (HD95) for whole tumor, core tumor, and enhancing tumor regions. Higher DSC and lower HD95 indicate better performance. The results were obtained by averaging predictions from seven independently trained models.

Team	DSC			HD95		
	Whole Tumor	Core Tumor	Enhancing Tumor	Whole Tumor	Core Tumor	Enhancing Tumor
deepX	91.02	85.00	78.53	4.44	5.90	24.06
Radicals	90.82	84.96	78.69	4.71	8.56	35.01
WassersteinDice	90.58	83.79	78.01	4.74	8.96	27.02
CKM	90.83	83.82	78.59	4.87	5.97	26.57
MDD-Net (UmU)	90.55	82.67	77.17	4.99	8.63	27.04

Table 5. Uncertainty Quantification results of **MDD-Net** on the BraTS validation set (125 cases). Metrics include Dice-based Area Under the Curve (DAUC), Relative False Positive Threshold (RFTP), and Relative False Negative Threshold (RFTN) for whole tumor, core tumor, and enhancing tumor regions. Higher DAUC and lower RFTP/RFTN indicate better performance. The results were obtained by averaging predictions from seven independently trained models.

Team	DAUC			RFTP			RFTN		
	Whole Tumor	Core Tumor	Enhancing Tumor	Whole Tumor	Core Tumor	Enhancing Tumor	Whole Tumor	Core Tumor	Enhancing Tumor
med_vision	95.24	92.23	83.24	0.28	0.62	0.93	87.74	98.74	98.74
nsu_btr	93.58	90.04	85.14	35.72	48.18	9.59	98.44	98.60	98.64
SCAN	93.46	82.98	80.64	12.40	19.95	21.53	0.87	0.42	0.24
MDD-Net (UmU)	92.59	83.61	78.83	4.48	10.13	7.95	0.27	0.17	0.08

Table 6. Segmentation performance of **MDD-Net** on the BraTS test set (166 cases). The metrics include Dice Similarity Coefficient (DSC) and 95th percentile Hausdorff Distance (HD95) for whole tumor, core tumor, and enhancing tumor regions. Higher DSC and lower HD95 indicate better performance. The results were obtained by averaging predictions from seven independently trained models.

Team	DSC			HD95		
	Whole Tumor	Core Tumor	Enhancing Tumor	Whole Tumor	Core Tumor	Enhancing Tumor
MDD-Net (UmU)	88.26	82.49	80.84	6.30	22.27	20.06

Table 7. Uncertainty Quantification results of **MDD-Net** on the BraTS test set (166 cases). Metrics include Dice-based Area Under the Curve (DAUC), Relative False Positive Threshold (RFTP), and Relative False Negative Threshold (RFTN) for whole tumor, core tumor, and enhancing tumor regions. Higher DAUC and lower RFTP/RFTN indicate better performance. The results were obtained by averaging predictions from seven independently trained models.

Team	DAUC			RFTP			RFTN		
	Whole Tumor	Core Tumor	Enhancing Tumor	Whole Tumor	Core Tumor	Enhancing Tumor	Whole Tumor	Core Tumor	Enhancing Tumor
MDD-Net (UmU)	90.61	85.83	83.03	4.18	5.49	4.45	0.31	1.68	0.07

4.5.1. Cross-Validation Performance

Table 3 summarizes the mean DSC and HD95 scores, along with their standard deviations, obtained from five-fold cross-validation on the training dataset comprising 396 cases. The results indicate that incorporating denoised inputs significantly enhances segmentation performance across all tumor regions. Specifically, **MDD-Net** with denoising outperforms the baseline U-Net models, achieving higher DSC scores and lower HD95 values, thereby confirming the benefits of the multi-denoising input strategy and the multi-decoder architecture.

4.5.2. Validation Set Evaluation

Table 4 presents the segmentation performance of **MDD-Net** on the validation set containing 125 cases. The model achieved mean DSC scores of 90.55%, 82.67%, and 77.17% for the whole tumor, core tumor, and enhancing tumor regions, respectively. Corresponding HD95 scores were 4.99, 8.63, and 27.04, indicating precise boundary delineation. Although these results are slightly lower than those of the top-ranking teams, they still reflect robust segmentation capabilities of **MDD-Net**.

4.5.3. Uncertainty Quantification on Validation Set

Table 5 details the performance of **MDD-Net** in the uncertainty quantification task on the validation set. The model achieved DAUC scores of 92.59, 83.61, and 78.83 for the whole, core, and enhancing tumor regions, respectively. The RFTP scores of 4.48, 10.13, and 7.95, along with RFTN scores of 0.27, 0.17, and 0.08, demonstrate the model's ability to effectively quantify uncertainty, particularly excelling in reducing false negatives.

4.5.4. Test Set Performance

Table 6 showcases the segmentation results of **MDD-Net** on the test set comprising 166 cases. The model achieved mean DSC scores of 88.26%, 82.49%, and 80.84% for the whole tumor, core tumor, and enhancing tumor regions, respectively. The corresponding HD95 scores were 6.30, 22.27, and 20.06, reflecting consistent performance across unseen data.

4.5.5. Uncertainty Quantification on Test Set

Table 7 presents the uncertainty quantification results on the test set. **MDD-Net** attained DAUC scores of 90.61, 85.83, and 83.03 for the whole, core, and enhancing tumor regions, respectively. The RFTP scores of 4.18, 5.49, and 4.45, alongside RFTN scores of 0.31, 1.68, and 0.07, indicate a high degree of confidence in correct predictions and appropriately high uncertainty in incorrect ones. Notably, **MDD-Net** secured the second position in the Quantification of Uncertainty in Segmentation task within the BraTS challenge, underscoring its effectiveness in uncertainty management.

4.6. Discussion

The experimental results underscore the superior performance of **MDD-Net** in brain tumor segmentation tasks, particularly highlighting the advantages of the multi-decoder architecture and the incorporation of denoised inputs. The significant improvement in DSC scores and reduction in HD95 values across all tumor regions attest to the model's enhanced capability in capturing both global contextual information and fine-grained details essential for accurate segmentation.

The use of multiple denoised versions of the input MRI scans effectively mitigates noise-related artifacts, such as salt-and-pepper noise, thereby facilitating more robust feature extraction. This approach not only improves segmentation accuracy but also contributes to the model's resilience against variations in imaging conditions. The integration of Squeeze-and-Excitation Blocks (SEBs) further refines feature representations by emphasizing informative channels, thereby enhancing the model's discriminative power.

Moreover, the multi-decoder framework allows **MDD-Net** to decompose the complex multiclass segmentation problem into simpler binary tasks, each focusing on specific tumor regions. This decomposition enhances the model's ability to specialize and accurately delineate different tumor substructures, leading to improved overall segmentation performance.

The successful ranking in the uncertainty quantification task demonstrates **MDD-Net**'s proficiency in not only producing accurate segmentations but also in providing reliable uncertainty estimates. This capability is crucial for clinical applications, where understanding the confidence of segmentation results can inform decision-making and guide further diagnostic procedures.

However, despite these strengths, there are areas for potential improvement. The slightly lower performance compared to the top-ranking teams in the validation set suggests that further enhancements, such as more sophisticated data augmentation techniques or ensemble methods, could be explored to boost performance. Additionally, optimizing the training process to reduce computational overhead and training time remains a valuable avenue for future research.

In conclusion, the proposed **MDD-Net** offers a robust and efficient solution for automated brain tumor segmentation, leveraging advanced architectural designs and denoising strategies to achieve high accuracy and reliable uncertainty quantification. These advancements hold significant promise for clinical applications, potentially improving diagnostic precision and patient outcomes.

4.7. Analysis of Segmentation Performance

The cross-validation results presented in Table 3 demonstrate the substantial improvement achieved by **MDD-Net** with denoising inputs over the traditional U-Net architectures. The incorporation of denoised inputs not only enhances the model's ability to capture intricate tumor structures but also contributes to more stable and accurate segmentation outcomes, as evidenced by the higher DSC scores and lower HD95 values across all tumor regions.

On the validation set, **MDD-Net** achieved mean DSC scores of 90.55%, 82.67%, and 77.17% for the whole tumor, core tumor, and enhancing tumor regions, respectively. These results indicate a robust performance, particularly in segmenting the whole and core tumor regions, which are critical for treatment planning and prognostic assessments. The HD95 scores further corroborate the model's precision in delineating tumor boundaries, with lower values signifying minimal boundary discrepancies.

In the uncertainty quantification task, **MDD-Net** exhibited exceptional performance, achieving high DAUC scores and low RFTP and RFTN scores. This indicates that the model not only provides accurate segmentations but also reliably quantifies uncertainty, distinguishing between confident correct predictions and uncertain incorrect ones. Such capability is invaluable in clinical settings, where understanding the confidence of segmentation results can guide further diagnostic and therapeutic decisions.

4.8. Comparison with State-of-the-Art Methods

When compared to other top-performing models in the BraTS challenge, **MDD-Net** holds its own by delivering competitive segmentation accuracy and superior uncertainty quantification. While some models achieved slightly higher DSC scores on the validation set, **MDD-Net** excels in providing reliable uncertainty estimates, as evidenced by its second-place ranking in the Quantification of Uncertainty in Segmentation task.

The multi-decoder architecture of **MDD-Net** allows for specialized processing of different tumor regions, which may contribute to its robust performance across various segmentation tasks. Additionally, the integration of denoised inputs ensures that the model remains resilient to noise and artifacts commonly present in MRI scans, a critical factor for clinical applicability.

4.9. Implications for Clinical Practice

The advancements demonstrated by **MDD-Net** have significant implications for clinical practice. Automated and accurate brain tumor segmentation can streamline radiotherapy planning, enhance diagnostic precision, and facilitate longitudinal monitoring of disease progression. The ability to quantify uncertainty further empowers clinicians by providing insights into the reliability of segmentation results, enabling more informed decision-making.

Moreover, the memory-efficient design of **MDD-Net** allows for the processing of full volumetric MRI data without the need for patch-based segmentation, reducing computational overhead and accelerating the workflow. This efficiency, combined with the model's high accuracy and reliable uncertainty estimates, positions **MDD-Net** as a valuable tool in the clinical setting, potentially improving patient outcomes through enhanced diagnostic and therapeutic strategies.

While **MDD-Net** has demonstrated substantial performance improvements, there are avenues for further enhancement. Future work may explore the integration of additional imaging modalities, such as diffusion-weighted imaging (DWI) or perfusion imaging, to provide more comprehensive tumor characterization. Additionally, incorporating advanced data augmentation techniques and ensemble learning strategies could further boost segmentation accuracy and model robustness.

Another promising direction is the application of **MDD-Net** to other types of brain tumors or different anatomical regions, broadening its clinical applicability. Furthermore, refining the uncertainty quantification mechanisms to provide more granular insights could enhance the model's utility in clinical decision-making processes.

Despite its strengths, **MDD-Net** has certain limitations that warrant consideration. The extensive training time of approximately six days per model poses challenges for rapid deployment and iterative model refinement. Additionally, while the model performs well on the BraTS dataset, its generalizability to other datasets with differing imaging protocols and patient populations remains to be thoroughly evaluated.

Furthermore, the reliance on a fixed number of denoising techniques may limit the model's adaptability to varying noise characteristics in different imaging settings. Addressing these limitations through optimized training protocols, adaptive denoising strategies, and broader validation studies will be essential for advancing the clinical utility of **MDD-Net**.

5. Conclusion and Future Work

In this study, we introduced **MDD-Net** (Multi-Decoder Denoising Network), a novel architecture designed to segment tumor substructures from multimodal brain Magnetic Resonance Imaging (MRI) scans. By decomposing the intricate task of multi-class tumor segmentation into three simpler binary segmentation problems, **MDD-Net** effectively enhances both the accuracy and efficiency of the segmentation process. The network architecture is inspired by the U-Net framework, incorporating Squeeze-and-Excitation (SE) blocks after each convolution and concatenation operation to dynamically recalibrate channel-wise feature responses. This integration of SE blocks significantly improves the model's ability to focus on the most informative features, thereby enhancing segmentation performance.

A key innovation of our approach is the stacking of original MRI images with their denoised counterparts. By employing denoising techniques, we enrich the input data, enabling the network to learn more robust and noise-invariant features. This strategy not only mitigates the adverse effects of common MRI artifacts, such as salt-and-pepper noise [1], but also contributes to substantial improvements in both the Dice Similarity Coefficient (DSC) and the 95th percentile Hausdorff Distance (HD95) metrics. Specifically, **MDD-Net** demonstrated a marked enhancement in segmentation performance, achieving DSC scores of 88.26%, 82.49%, and 80.84% and HD95 scores of 6.30, 22.27, and 20.06 for the whole tumor, tumor core, and enhancing tumor core regions, respectively, on the test set.

Moreover, **MDD-Net** excelled in the task of Quantification of Uncertainty in Segmentation, securing a top-two position. This achievement underscores the model's capability not only to produce accurate segmentations but also to provide reliable uncertainty estimates. Such uncertainty quantification is crucial for clinical decision-making, as it offers insights into the confidence of the segmentation results, thereby aiding clinicians in assessing the reliability of automated annotations.

The comprehensive evaluation of **MDD-Net** on the BraTS dataset highlights its competitive performance relative to existing state-of-the-art methods. The ability to process entire volumetric MRI data without resorting to patch-based segmentation approaches, as required by previous models like TUNet [20], results in more efficient training and inference times. This efficiency, coupled with the enhanced segmentation accuracy, positions **MDD-Net** as a valuable tool for clinical applications, potentially streamlining workflows in radiotherapy planning and diagnostic follow-up.

While **MDD-Net** has demonstrated significant advancements in brain tumor segmentation, several avenues for future research remain. One potential direction is the exploration of additional denoising techniques and their impact on segmentation performance. By integrating more sophisticated noise reduction methods, such as non-local means or deep learning-based denoising algorithms, we can further enhance the quality of input data and, consequently, the accuracy of the segmentation results.

Another promising area is the incorporation of attention mechanisms beyond the Squeeze-and-Excitation blocks. For instance, integrating spatial attention modules could enable the network to better focus on relevant anatomical regions, thereby improving the delineation of tumor boundaries. Additionally, experimenting with transformer-based architectures, which have shown remarkable success in various computer vision tasks, may offer further enhancements to the segmentation capabilities of **MDD-Net**.

Expanding the applicability of **MDD-Net** to other types of brain tumors or even different anatomical structures is another worthwhile pursuit. This generalization would require adapting the network to handle diverse tumor morphologies and imaging modalities, potentially involving multi-task learning approaches to accommodate varying segmentation tasks simultaneously.

Furthermore, enhancing the uncertainty quantification aspect of **MDD-Net** presents an opportunity to develop more nuanced and interpretable uncertainty maps. Techniques such as Monte Carlo dropout or Bayesian neural networks could be integrated to provide probabilistic interpretations of the segmentation results, offering deeper insights into model confidence and reliability.

Lastly, conducting extensive clinical validations and collaborating with medical professionals will be essential to translate **MDD-Net** from a research prototype to a clinically deployable tool. Such collaborations can provide valuable feedback on the model's performance in real-world settings and guide further refinements to meet the stringent requirements of clinical practice.

In summary, **MDD-Net** represents a significant step forward in automated brain tumor segmentation, offering improved accuracy and reliable uncertainty estimates. Future enhancements and broader applications of this framework hold the potential to make substantial contributions to the field of medical imaging and patient care.

References

1. Ali, H.M.: A new method to remove salt pepper noise in Magnetic Resonance Images. In: 2016 11th International Conference on Computer Engineering Systems (ICCES). pp. 155–160 (2016)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. the cancer imaging archive (2017) (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. The Cancer Imaging Archive **286** (2017)
4. 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**, 170117 (2017)
5. 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)
6. Castells, X., García-Gómez, J.M., Navarro, A.T., Acebes, J.J., Godino, O., Boluda, S., Barceló, A., Robles, M., Ariño, J., Arús, C.: Automated brain tumor biopsy prediction using single-labeling cdna microarrays-based gene expression profiling. Diagnostic Molecular Pathology **18**, 206–218 (2009)
7. Hausdorff, F.: Erweiterung einer stetigen Abbildung, pp. 555–568. Springer Berlin Heidelberg, Berlin, Heidelberg (2008), https://doi.org/10.1007/978-3-540-76807-4_16
8. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
9. Isensee, F., Jäger, P., Wasserthal, J., Zimmerer, D., Petersen, J., Kohl, S., Schock, J., Klein, A., Roß, T., Wirkert, S., Neher, P., Dinkelacker, S., Köhler, G., Maier-Hein, K.: batchgenerators—a python framework for data augmentation (Jan 2020)
10. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: No new-net. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 234–244. Springer International Publishing, Cham (2019)
11. Jiang, Z., Ding, C., Liu, M., Tao, D.: Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 231–241. Springer International Publishing, Cham (2020)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. McKinley, R., Rebsamen, M., Meier, R., Wiest, R.: Triplanar ensemble of 3d-to-2d cnns with label-uncertainty for brain tumor segmentation. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 379–387. Springer International Publishing, Cham (2020)
14. 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)

15. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization (2018), http://arxiv.org/abs/1810.11654

16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)

17. Simard, P.Y., Steinkraus, D., Platt, J.C.: Best practices for convolutional neural networks applied to visual document analysis. In: Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. pp. 958–963 (2003)

18. Thurnher, M.: The 2007 WHO classification of tumors of the central nervous system—what has changed? American Journal of Neuroradiology (01 2012)

19. Vu, M.H., Grimbergen, G., Nyholm, T., Löfstedt, T.: Evaluation of multi-slice inputs to convolutional neural networks for medical image segmentation. arXiv preprint arXiv:1912.09287 (2019)

20. Vu, M.H., Nyholm, T., Löfstedt, T.: TuNet: End-to-End Hierarchical Brain Tumor Segmentation Using Cascaded Networks. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 174–186. Springer International Publishing, Cham (2020)

21. Zhao, Y.X., Zhang, Y.M., Liu, C.L.: Bag of tricks for 3d mri brain tumor segmentation. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 210–220. Springer International Publishing, Cham (2020)

22. Anson Bastos, Abhishek Nadgeri, Kuldeep Singh, Isaiah Onando Mulang, Saeedeh Shekarpour, Johannes Hoffart, and Manohar Kaul. 2021. RECON: Relation Extraction using Knowledge Graph Context in a Graph Neural Network. In *Proceedings of the Web Conference 2021*. 1673–1685.

23. Philipp Christmann, Rishiraj Saha Roy, Abdalghani Abujabal, Jyotsna Singh, and Gerhard Weikum. 2019. Look before You Hop: Conversational Question Answering over Knowledge Graphs Using Judicious Context Expansion. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management CIKM*. 729–738.

24. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, 4171–4186.

25. Endri Kacupaj, Kuldeep Singh, Maria Maleshkova, and Jens Lehmann. 2022. An Answer Verbalization Dataset for Conversational Question Answerings over Knowledge Graphs. *arXiv preprint arXiv:2208.06734* (2022).

26. Magdalena Kaiser, Rishiraj Saha Roy, and Gerhard Weikum. 2021. Reinforcement Learning from Reformulations In Conversational Question Answering over Knowledge Graphs. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 459–469.

27. Yunshi Lan, Gaole He, Jinhao Jiang, Jing Jiang, Wayne Xin Zhao, and Ji-Rong Wen. 2021. A Survey on Complex Knowledge Base Question Answering: Methods, Challenges and Solutions. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. International Joint Conferences on Artificial Intelligence Organization, 4483–4491. Survey Track.

28. Yunshi Lan and Jing Jiang. 2021. Modeling transitions of focal entities for conversational knowledge base question answering. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*.

29. Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 7871–7880.

30. Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*.

31. Pierre Marion, Paweł Krzysztof Nowak, and Francesco Piccinno. 2021. Structured Context and High-Coverage Grammar for Conversational Question Answering over Knowledge Graphs. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2021).

32. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, may 2015. 10.1038/nature14539. URL http://dx.doi.org/10.1038/nature14539.

33. Dong Yu Li Deng. *Deep Learning: Methods and Applications*. NOW Publishers, May 2014. URL https://www.microsoft.com/en-us/research/publication/deep-learning-methods-and-applications/.

34. Eric Makita and Artem Lenskiy. A movie genre prediction based on Multivariate Bernoulli model and genre correlations. (May), mar 2016. URL http://arxiv.org/abs/1604.08608.

35. Junhua Mao, Wei Xu, Yi Yang, Jiang Wang, and Alan L Yuille. Explain images with multimodal recurrent neural networks. *arXiv preprint arXiv:1410.1090*, 2014.

36. J Ngiam, A Khosla, and M Kim. Multimodal Deep Learning. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 689—-696, 2011. URL http://ai.stanford.edu/{~}ang/papers/icml11-MultimodalDeepLearning.pdf.

37. Deli Pei, Huaping Liu, Yulong Liu, and Fuchun Sun. Unsupervised multimodal feature learning for semantic image segmentation. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6. IEEE, aug 2013. ISBN 978-1-4673-6129-3. 10.1109/IJCNN.2013.6706748. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6706748.

38. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

39. Richard Socher, Milind Ganjoo, Christopher D Manning, and Andrew Ng. Zero-Shot Learning Through Cross-Modal Transfer. In C J C Burges, L Bottou, M Welling, Z Ghahramani, and K Q Weinberger (eds.), *Advances in Neural Information Processing Systems 26*, pp. 935–943. Curran Associates, Inc., 2013. URL http://papers.nips.cc/paper/5027-zero-shot-learning-through-cross-modal-transfer.pdf.

40. Hao Fei, Shengqiong Wu, Meishan Zhang, Min Zhang, Tat-Seng Chua, and Shuicheng Yan. Enhancing video-language representations with structural spatio-temporal alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

41. Hao Fei, Yafeng Ren, and Donghong Ji. Retrofitting structure-aware transformer language model for end tasks. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 2151–2161, 2020.

42. Shengqiong Wu, Hao Fei, Fei Li, Meishan Zhang, Yijiang Liu, Chong Teng, and Donghong Ji. Mastering the explicit opinion-role interaction: Syntax-aided neural transition system for unified opinion role labeling. In *Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence*, pages 11513–11521, 2022.

43. Wenxuan Shi, Fei Li, Jingye Li, Hao Fei, and Donghong Ji. Effective token graph modeling using a novel labeling strategy for structured sentiment analysis. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4232–4241, 2022.

44. Hao Fei, Yue Zhang, Yafeng Ren, and Donghong Ji. Latent emotion memory for multi-label emotion classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7692–7699, 2020.

45. Fengqi Wang, Fei Li, Hao Fei, Jingye Li, Shengqiong Wu, Fangfang Su, Wenxuan Shi, Donghong Ji, and Bo Cai. Entity-centered cross-document relation extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9871–9881, 2022.

46. Ling Zhuang, Hao Fei, and Po Hu. Knowledge-enhanced event relation extraction via event ontology prompt. *Inf. Fusion*, 100:101919, 2023.

47. Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, and Quoc V Le. Qanet: Combining local convolution with global self-attention for reading comprehension. *arXiv preprint arXiv:1804.09541*, 2018.

48. Shengqiong Wu, Hao Fei, Yixin Cao, Lidong Bing, and Tat-Seng Chua. Information screening whilst exploiting! multimodal relation extraction with feature denoising and multimodal topic modeling. *arXiv preprint arXiv:2305.11719*, 2023.

49. Jundong Xu, Hao Fei, Liangming Pan, Qian Liu, Mong-Li Lee, and Wynne Hsu. Faithful logical reasoning via symbolic chain-of-thought. *arXiv preprint arXiv:2405.18357*, 2024.

50. Matthew Dunn, Levent Sagun, Mike Higgins, V Ugur Guney, Volkan Cirik, and Kyunghyun Cho. SearchQA: A new Q&A dataset augmented with context from a search engine. *arXiv preprint arXiv:1704.05179*, 2017.

51. Hao Fei, Shengqiong Wu, Jingye Li, Bobo Li, Fei Li, Libo Qin, Meishan Zhang, Min Zhang, and Tat-Seng Chua. Lasuie: Unifying information extraction with latent adaptive structure-aware generative language model. In *Proceedings of the Advances in Neural Information Processing Systems, NeurIPS 2022*, pages 15460–15475, 2022.

52. Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1):9–27, 2011.

53. Hao Fei, Yafeng Ren, Yue Zhang, Donghong Ji, and Xiaohui Liang. Enriching contextualized language model from knowledge graph for biomedical information extraction. *Briefings in Bioinformatics*, 22(3), 2021.

54. Shengqiong Wu, Hao Fei, Wei Ji, and Tat-Seng Chua. Cross2StrA: Unpaired cross-lingual image captioning with cross-lingual cross-modal structure-pivoted alignment. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2593–2608, 2023.

55. Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*, 2016.

56. Hao Fei, Fei Li, Bobo Li, and Donghong Ji. Encoder-decoder based unified semantic role labeling with label-aware syntax. In *Proceedings of the AAAI conference on artificial intelligence*, pages 12794–12802, 2021.

57. Hao Fei, Shengqiong Wu, Yafeng Ren, Fei Li, and Donghong Ji. Better combine them together! integrating syntactic constituency and dependency representations for semantic role labeling. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 549–559, 2021.

58. Hao Fei, Bobo Li, Qian Liu, Lidong Bing, Fei Li, and Tat-Seng Chua. Reasoning implicit sentiment with chain-of-thought prompting. *arXiv preprint arXiv:2305.11255*, 2023.

59. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423.

60. Shengqiong Wu, Hao Fei, Leigang Qu, Wei Ji, and Tat-Seng Chua. Next-gpt: Any-to-any multimodal llm. *CoRR*, abs/2309.05519, 2023.

61. Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

62. Hao Fei, Shengqiong Wu, Wei Ji, Hanwang Zhang, Meishan Zhang, Mong-Li Lee, and Wynne Hsu. Video-of-thought: Step-by-step video reasoning from perception to cognition. In *Proceedings of the International Conference on Machine Learning*, 2024.

63. Naman Jain, Pranjali Jain, Pratik Kayal, Jayakrishna Sahit, Soham Pachpande, Jayesh Choudhari, et al. Agribot: agriculture-specific question answer system. *IndiaRxiv*, 2019.

64. Hao Fei, Shengqiong Wu, Wei Ji, Hanwang Zhang, and Tat-Seng Chua. Dysen-vdm: Empowering dynamics-aware text-to-video diffusion with llms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7641–7653, 2024.

65. Mihir Momaya, Anjnya Khanna, Jessica Sadavarte, and Manoj Sankhe. Krushi–the farmer chatbot. In *2021 International Conference on Communication information and Computing Technology (ICCICT)*, pages 1–6. IEEE, 2021.

66. Hao Fei, Fei Li, Chenliang Li, Shengqiong Wu, Jingye Li, and Donghong Ji. Inheriting the wisdom of predecessors: A multiplex cascade framework for unified aspect-based sentiment analysis. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI*, pages 4096–4103, 2022.

67. Shengqiong Wu, Hao Fei, Yafeng Ren, Donghong Ji, and Jingye Li. Learn from syntax: Improving pair-wise aspect and opinion terms extraction with rich syntactic knowledge. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pages 3957–3963, 2021.

68. Bobo Li, Hao Fei, Lizi Liao, Yu Zhao, Chong Teng, Tat-Seng Chua, Donghong Ji, and Fei Li. Revisiting disentanglement and fusion on modality and context in conversational multimodal emotion recognition. In *Proceedings of the 31st ACM International Conference on Multimedia, MM*, pages 5923–5934, 2023.

69. Hao Fei, Qian Liu, Meishan Zhang, Min Zhang, and Tat-Seng Chua. Scene graph as pivoting: Inference-time image-free unsupervised multimodal machine translation with visual scene hallucination. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5980–5994, 2023.

70. Hao Fei, Shengqiong Wu, Hanwang Zhang, Tat-Seng Chua, and Shuicheng Yan. Vitron: A unified pixel-level vision llm for understanding, generating, segmenting, editing. In *Proceedings of the Advances in Neural Information Processing Systems, NeurIPS 2024,*, 2024.

71. Sanjeev Arora, Yingyu Liang, and Tengyu Ma. A simple but tough-to-beat baseline for sentence embeddings. In *ICLR*, 2017.

72. Abbott Chen and Chai Liu. Intelligent commerce facilitates education technology: The platform and chatbot for the taiwan agriculture service. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 11:1–10, 01 2021.

73. Shengqiong Wu, Hao Fei, Xiangtai Li, Jiayi Ji, Hanwang Zhang, Tat-Seng Chua, and Shuicheng Yan. Towards semantic equivalence of tokenization in multimodal llm. *arXiv preprint arXiv:2406.05127*, 2024.

74. Jingye Li, Kang Xu, Fei Li, Hao Fei, Yafeng Ren, and Donghong Ji. MRN: A locally and globally mention-based reasoning network for document-level relation extraction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1359–1370, 2021.

75. Hao Fei, Shengqiong Wu, Yafeng Ren, and Meishan Zhang. Matching structure for dual learning. In *Proceedings of the International Conference on Machine Learning, ICML*, pages 6373–6391, 2022.

76. Hu Cao, Jingye Li, Fangfang Su, Fei Li, Hao Fei, Shengqiong Wu, Bobo Li, Liang Zhao, and Donghong Ji. OneEE: A one-stage framework for fast overlapping and nested event extraction. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 1953–1964, 2022.

77. Isakwisa Gaddy Tende, Kentaro Aburada, Hisaaki Yamaba, Tetsuro Katayama, and Naonobu Okazaki. Proposal for a crop protection information system for rural farmers in tanzania. *Agronomy*, 11(12):2411, 2021.

78. Hao Fei, Yafeng Ren, and Donghong Ji. Boundaries and edges rethinking: An end-to-end neural model for overlapping entity relation extraction. *Information Processing & Management*, 57(6):102311, 2020.

79. Jingye Li, Hao Fei, Jiang Liu, Shengqiong Wu, Meishan Zhang, Chong Teng, Donghong Ji, and Fei Li. Unified named entity recognition as word-word relation classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10965–10973, 2022.

80. Mohit Jain, Pratyush Kumar, Ishita Bhansali, Q Vera Liao, Khai Truong, and Shwetak Patel. Farmchat: a conversational agent to answer farmer queries. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):1–22, 2018.

81. Shengqiong Wu, Hao Fei, Hanwang Zhang, and Tat-Seng Chua. Imagine that! abstract-to-intricate text-to-image synthesis with scene graph hallucination diffusion. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pages 79240–79259, 2023.

82. Hao Fei, Tat-Seng Chua, Chenliang Li, Donghong Ji, Meishan Zhang, and Yafeng Ren. On the robustness of aspect-based sentiment analysis: Rethinking model, data, and training. *ACM Transactions on Information Systems*, 41(2):50:1–50:32, 2023.

83. Yu Zhao, Hao Fei, Yixin Cao, Bobo Li, Meishan Zhang, Jianguo Wei, Min Zhang, and Tat-Seng Chua. Constructing holistic spatio-temporal scene graph for video semantic role labeling. In *Proceedings of the 31st ACM International Conference on Multimedia, MM*, pages 5281–5291, 2023.

84. Shengqiong Wu, Hao Fei, Yixin Cao, Lidong Bing, and Tat-Seng Chua. Information screening whilst exploiting! multimodal relation extraction with feature denoising and multimodal topic modeling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14734–14751, 2023.

85. Hao Fei, Yafeng Ren, Yue Zhang, and Donghong Ji. Nonautoregressive encoder-decoder neural framework for end-to-end aspect-based sentiment triplet extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 34(9):5544–5556, 2023.

86. Yu Zhao, Hao Fei, Wei Ji, Jianguo Wei, Meishan Zhang, Min Zhang, and Tat-Seng Chua. Generating visual spatial description via holistic 3D scene understanding. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7960–7977, 2023.