Multi-scale Superpixel based Hierarchical Attention Model for Brain CT Classification

Xiao Song[a], Xiaodan Zhang[a,*], Junzhong Ji[a], Ying Liu[b,*]

[a] Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

[b] Department of Radiology, Peking University Third Hospital, Beijing 100191, China

[*] Email: zhangxiaodan@bjut.edu.cn(Xiaodan Zhang)

Introduction

- Brain CT image classification is critical for assisting brain disease diagnosis.
- However, the brain CT images contain much noisy information and the lesions are unstable in shapes and locations, making the classification task more difficult using conventional CNN models representing uniform grid-level features of neural receptive fields with equal sizes and shapes (Fig. 1 (a)).
- In this study, we propose a novel Multi-scale Superpixel based Hierarchical Attention (MSHA) model for brain CT classification, introducing the multi-scale superpixels to a hierarchical fusion structure to remove noise and help the model focus on the lesion areas.

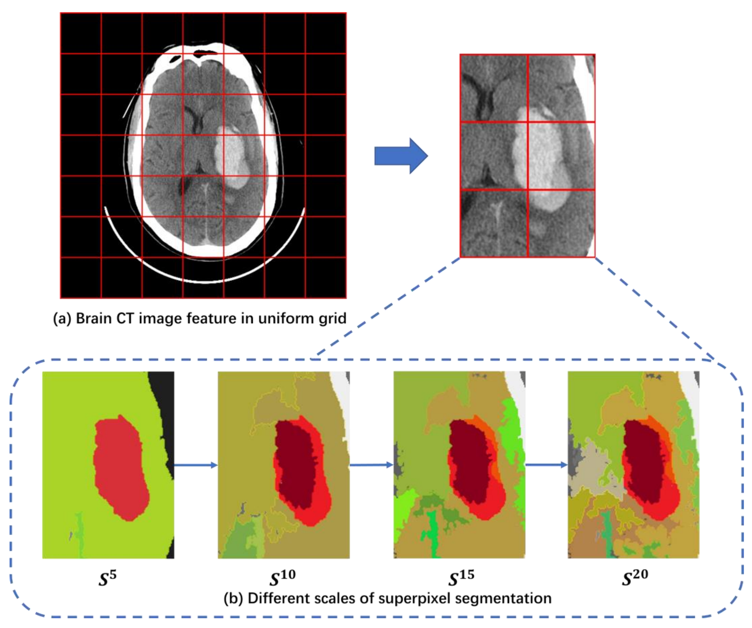

Fig 1. One example of a Brain CT image with its cerebral hemorrhage lesion.

Methodology

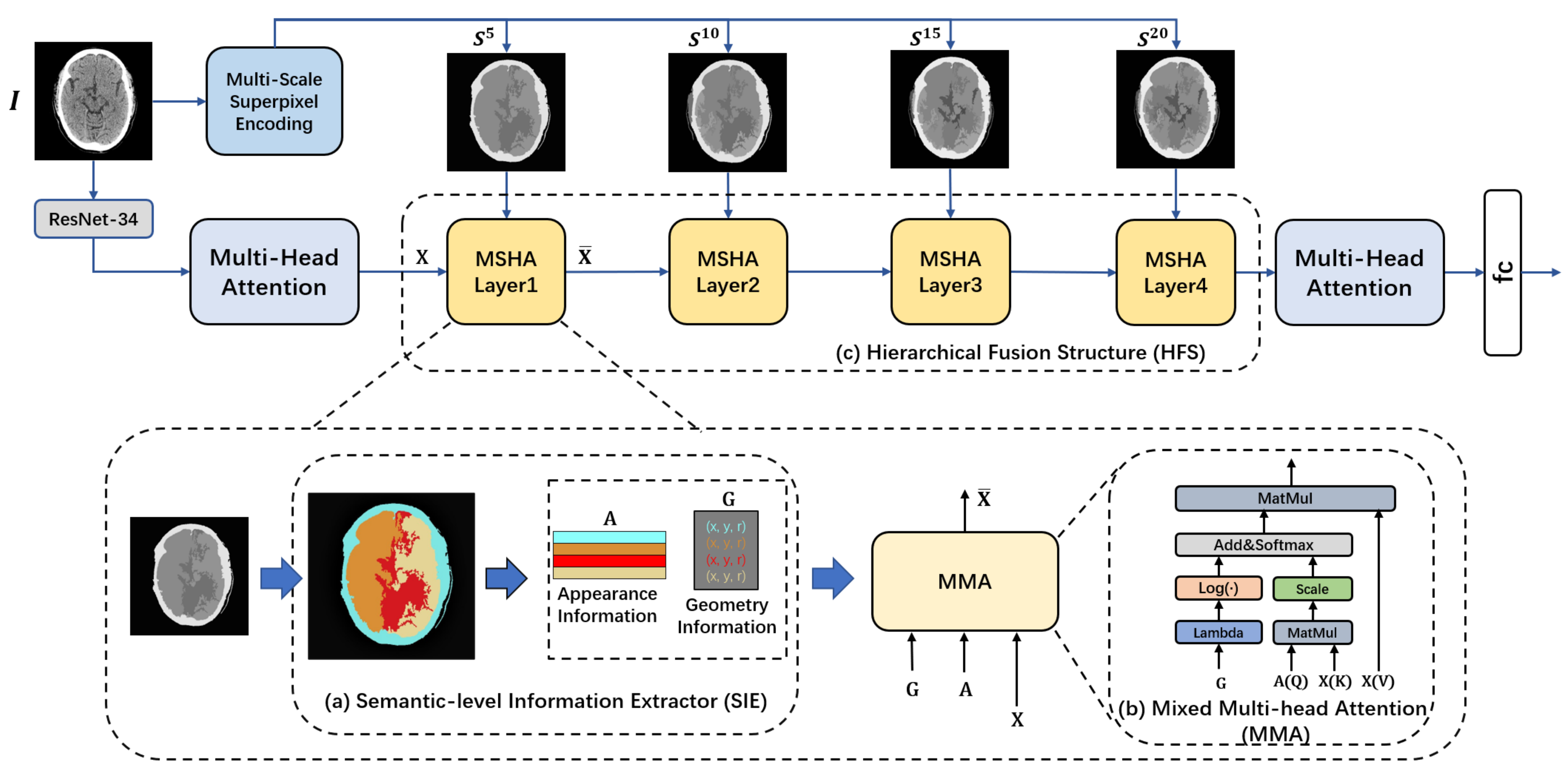

Fig 2. The overall pipeline of the proposed Multi-scale Superpixel based Hierarchical Attention model for Brain CT classification.

- MSHA introduces the multi-scale superpixels to a hierarchical fusion structure to remove noise and help the model attend to lesion areas from coarse to fine (Fig. 2).
- In MSHA, a semantic-level information extractor (Fig. 2 (a) and Fig. 3) is proposed to extract appearance and geometry information for lesions, which are then fused through a mixed multi-head attention model (Fig. 2 (3)) to enrich the feature representation of superpixel regions.

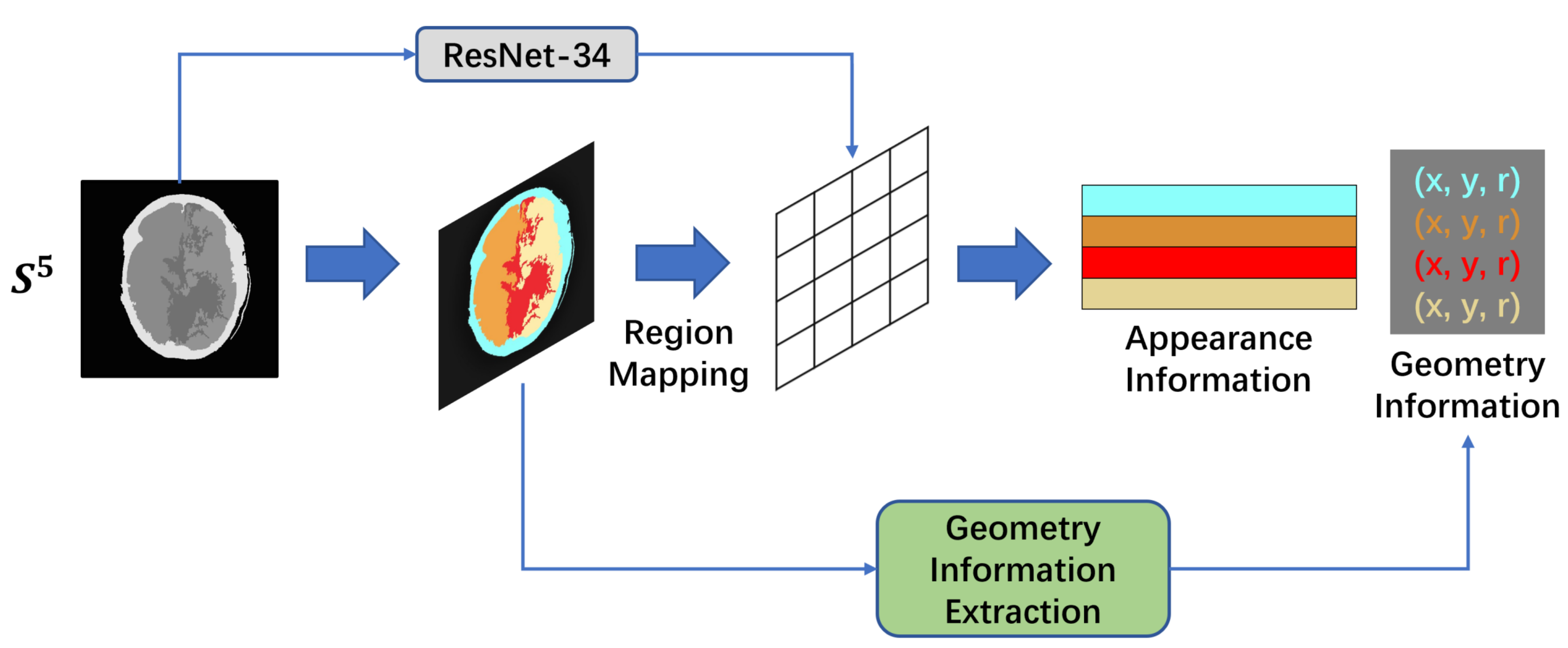

Fig 3. Structure of the proposed Semantic-level Information Extractor.

Experiments

The comparison and ablation experimental results are listed in the following three tables.

Table 1. Comparison of the MASH and other conventional image classification methods.

Methods	ACC	SEN	SPE	PPV	NPV	F
ResNet-18 [3]	86.40	73.56	98.20	24.42	87.23	36.15
ResNet-34 [3]	90.24	76.09	96.67	56.16	92.11	65.53
ResNet-50 [3]	87.39	71.96	97.44	34.41	88.68	46.47
ResNet-101 [3]	87.83	73.28	97.55	37.38	89.13	49.48
Inception-v4 [4]	91.41	74.34	94.55	71.99	94.26	72.89
Inception-ResNet-v2 [4]	90.52	64.18	95.37	72.67	93.56	67.69
Residual Attn [27]	90.52	78.15	97.05	55.91	92.12	65.10
SENet [28]	86.17	71.25	**98.40**	21.66	86.91	32.79
MSHA	**92.46**	**78.54**	96.20	**72.79**	**94.92**	**75.46**

Table 2. Comparison of the baseline model with the MSHA method. Scale means the number of superpixels used in different scales.

Methods	Scale	ACC(%)	SEN(%)	SPE(%)	PPV(%)	NPV(%)	F(%)
Baseline	–	90.24	76.09	**96.67**	56.16	92.11	65.53
MSHA(S^5)	S^5	91.45	76.45	96.07	68.24	94.15	71.40
MSHA(S^{10})	S^{10}	91.47	75.11	95.60	69.70	94.34	72.22
MSHA(S^{15})	S^{15}	91.49	75.68	95.75	68.99	94.23	72.08
MSHA(S^{20})	S^{20}	91.69	77.68	96.33	67.18	93.95	71.93
MSHA(S^{5-10})	S^5, S^{10}	91.46	75.26	95.69	69.06	94.26	71.94
MSHA(S^{5-15})	S^5, S^{10}, S^{15}	92.03	78.08	96.24	69.75	94.40	73.52
MSHA(S^{5-20})	$S^5, S^{10}, S^{15}, S^{20}$	**92.46**	**78.54**	96.20	**72.79**	**94.92**	**75.46**

Table 3. Comparison of the baseline model with the semantic-level information method. AppInfo and GeoInfo mean whether the MSHA model utilize appearance and geometry information.

Methods	AppInfo	GeoInfo	ACC(%)	SEN(%)	SPE(%)	PPV(%)	NPV(%)	F(%)
Baseline	–	–	90.24	76.09	**96.67**	56.16	92.11	65.53
(a)	✓		92.10	79.65	95.71	67.99	94.07	73.52
(b)	✓	✓	**92.46**	**78.54**	96.20	**72.79**	**94.92**	**75.46**

Analyses

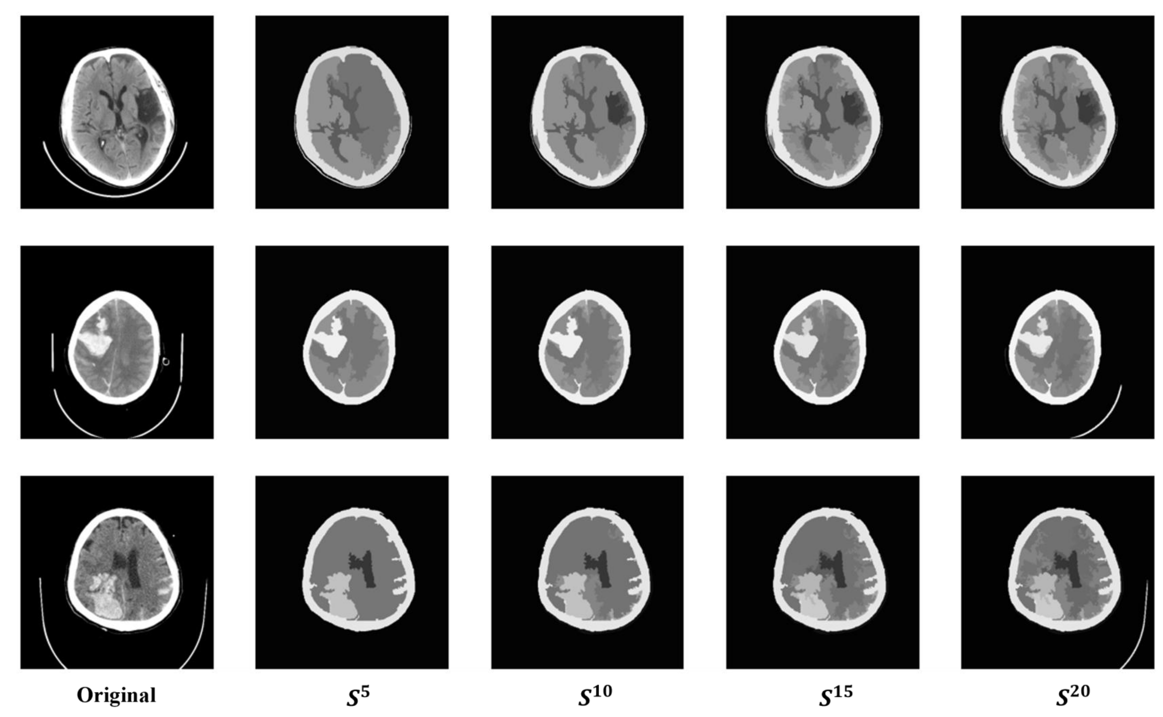

Fig 4. Visualization of the original images and their superpixel maps with different scales.

- The MSHA boosts the performance of the brain CT classification task, which validates the effectiveness of the proposed modules. (Tab. 1)
- Effect of Superpixel: The original images contain much noise, and the normal physiological structures and lesion areas are mixed, chaotic, and complex Superpixel can effectively plot the lesion regions in the raw brain CT images. (Fig. 4)
- Effect of Hierarchical Fusion Structure: The hierarchical fusion structure is able to effectively fuse different scales of superpixel attention features (Tab. 2)
- Effect of Semantic-level Information: It can be seen that the model employed with appearance information outperforms the model without appearance information in all evaluation metrics, and the model with geometry information performs better, which indicates the effectiveness of introducing semantic-level information. (Tab. 3)

Conclusion

MSHA introduces the multi-scale superpixel into a hierarchical attention structure to help the model focus on lesion areas and mines more semantic (appearance and geometry) information for classifying Brain CT images. Experimental results CQ500-based0 based Brain CT dataset show that our proposed method can more precisely classify brain diseases, which demonstrates the effectiveness of our MSHA model.